



HAL
open science

A NOTE ON THE MEAN-FIELD LIMIT FOR THE PARTICLE SWARM OPTIMIZATION

Hui Huang

► **To cite this version:**

Hui Huang. A NOTE ON THE MEAN-FIELD LIMIT FOR THE PARTICLE SWARM OPTIMIZATION. Applied Mathematics Letters, In press, pp.107133. 10.1016/j.aml.2021.107133 . hal-03152987v1

HAL Id: hal-03152987

<https://hal.science/hal-03152987v1>

Submitted on 26 Feb 2021 (v1), last revised 10 Apr 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A NOTE ON THE MEAN-FIELD LIMIT FOR THE PARTICLE SWARM OPTIMIZATION

HUI HUANG

ABSTRACT. Recently a continuous description of particle swarm optimization (PSO) based on a system of stochastic differential equations was proposed in [10] for solving global optimization problems. This note is devoted to solving the open problem suggested in [10] by providing a rigorous proof of the large particle limit under the assumption of finite second moment.

Keywords: Swarm optimization, propagation of chaos, consensus based optimization, Laplace's principle.

1. INTRODUCTION

Particle Swarm Optimization (PSO) has fascinated the scientific community ever since it was initially introduced to model the intelligent collective behavior of complex biological systems such as flocks of birds or schools of fish [14, 15, 19], and it is now widely recognized as an efficient method for tackling complex optimization problems [18]. However the mathematical understanding of PSO is still in its infancy. Recently Grassi and Pareschi [10] 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

$$\begin{cases} dX_t^i = V_t^i dt, \\ dV_t^i = -\frac{\gamma}{m} V_t^i dt + \frac{\lambda}{m} (X_t^\alpha(\rho^N) - X_t^i) dt + \frac{\sigma}{m} D(X_t^\alpha(\rho^N) - X_t^i) dB_t^i, \quad i = 1, \dots, N, \end{cases} \quad (1.1)$$

where $X_t^i, V_t^i \in \mathbb{R}^d$ denote the position and velocity of the i -th particle at time t , m 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. Moreover, we use the notations $D(X_t) := \text{diag}\{(X_t)_1, \dots, (X_t)_d\} \in \mathbb{R}^{d \times d}$, and the weighted average $X_t^\alpha(\rho^N) := \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\mathcal{E}(x) \rho^N(t, x) dx}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^N(t, x) dx}$ with the empirical measure $\rho^N(t, x) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}(x)$. Here the initial data $(X_0^i, V_0^i)_{i=1}^N$ is i.i.d. with common distribution $f_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$, where $\mathcal{P}_2(\mathbb{R}^{2d})$ denotes the set of probability measures with finite second moment, which is endowed with the 2-Wasserstein distance W_2 [1]. The choice of the weight function $\omega_\alpha^\mathcal{E}(x) := \exp(-\alpha \mathcal{E}(x))$ comes from the well-known Laplace's principle [6, 16], a classical asymptotic method for integrals, which states that for any probability measure $\rho \in \mathcal{P}(\mathbb{R}^d)$, it holds

$$\lim_{\alpha \rightarrow \infty} \left(-\frac{1}{\alpha} \log \left(\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) d\rho(x) \right) \right) = \inf_{x \in \text{supp}(\rho)} \mathcal{E}(x). \quad (1.2)$$

Thus for α large enough, one expects that $X_t^\alpha(\rho^N) \approx \text{argmin} \{\mathcal{E}(X_t^1), \dots, \mathcal{E}(X_t^N)\}$. Due to the nonlinearity coming from the local best position $X_t^\alpha(\rho^N)$, the authors in [10] could only formally derive the mean-field Vlasov-Fokker-Plank equation in the limit of a large number of particles. This note works as a completion for a theory gap suggested in [10] by providing the rigorous derivation of the mean-field limit under the assumption of finite second moment. Before proving the mean-field limit result (see Section 3), we need to

The author would like to thank Prof. Lorenzo Pareschi for suggesting this problem to him, and thank Prof. Massimo Fornasier for his encouragement and support during the author's stay at TUM. This work is partially supported by the Pacific Institute for the Mathematical Sciences (PIMS) postdoctoral fellowship.

establish the well-posedness for both the particle system (1.1) and its limiting equation (1.3) (see Section 2), which was missing in [10].

In the following we give the assumption on our given cost function \mathcal{E} .

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

1. $|\mathcal{E}(v) - \mathcal{E}(u)| \leq L|v - u|$ for all $u, v \in \mathbb{R}^d$, where $L > 0$ is the global Lipschitz constant ;
2. \mathcal{E} is bounded, i.e. $-\infty < \underline{\mathcal{E}} := \inf \mathcal{E} \leq \mathcal{E} \leq \sup \mathcal{E} =: \bar{\mathcal{E}} < +\infty$ and define $\delta_{\mathcal{E}} := \bar{\mathcal{E}} - \underline{\mathcal{E}}$.

For later use let us denote $[N] := \{1, \dots, N\}$. As $N \rightarrow \infty$, the mean-field limit result [2, 3, 12, 13, 20] shall show that our PSO dynamic (1.1) well approximates solutions of the following mean-field kinetic McKean-Vlasov type equations

$$\begin{cases} d\bar{X}_t^i = \bar{V}_t^i dt, \\ d\bar{V}_t^i = -\frac{\gamma}{m}\bar{V}_t^i dt + \frac{\lambda}{m}(X_t^\alpha(\rho) - \bar{X}_t^i)dt + \frac{\sigma}{m}D(X_t^\alpha - \bar{X}_t^i)dB_t^i, \quad i \in [N], \end{cases} \quad (1.3)$$

where

$$X_t^\alpha(\rho) = \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\mathcal{E}(x) \rho(t, x) dx}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho(t, x) dx}, \quad \rho(t, x) = \int_{\mathbb{R}^d} f(t, x, v) dv, \quad (1.4)$$

and the initial data $(\bar{X}_0^i, \bar{V}_0^i)$ is the same as in (1.1). Then the processes $(\bar{X}_t^i, \bar{V}_t^i)_{t \geq 0}$ are independent since the initial conditions and driving Brownian motions are independent. Here $f(t, x, v)$ is the common distribution of $(\bar{X}_t^i, \bar{V}_t^i)$ at time t for all $i \in [N]$, which makes the set of equations (1.3) nonlinear. As a direct application of Itô's formula, the law $f_t := f(t, \cdot, \cdot)$ at time t is a weak solution to the following nonlinear Vlasov-Fokker-Plank equation

$$\partial_t f_t + v \cdot \nabla_x f_t = \nabla_v \cdot \left(\frac{\gamma}{m} v f_t + \frac{\lambda}{m} (x - X_t^\alpha(\rho)) f_t + \frac{\sigma^2}{2m^2} D(x - X_t^\alpha(\rho))^2 \nabla_v f_t \right) \quad (1.5)$$

with the initial data $f_0(x, v) = \mathcal{L}(\bar{X}_0^i, \bar{V}_0^i)$. It was formally showed in [10] that, in the zero inertia limit ($m \rightarrow 0$), one expects the macroscopic density in the mean-field PSO equation (1.5) to be well approximated by the mean-field Consensus Based Optimization (CBO) [4, 5, 8, 9, 11, 17] equation of the form

$$\partial_t \rho_t + \lambda \nabla_x \cdot (\rho_t (X_t^\alpha(\rho) - x)) = \frac{\sigma^2}{2} \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} \left(\rho_t (x_j - (X_t^\alpha(\rho))_j)^2 \right). \quad (1.6)$$

In the sequel we shall use the notation $m_{f_t}^2 := \iint_{\mathbb{R}^{2d}} (|x|^2 + |v|^2) f(t, x, v) dx dv$ for the second moment of f_t .

2. WELL-POSEDNESS FOR THE PARTICLE SYSTEM(1.1) AND THE MEAN-FIELD DYNAMIC (1.3)

To start with, we recall a lemma from [8, Lemma 2.1], which implies that the coefficients in (1.1) are locally Lipschitz and have linear growth.

Lemma 2.1. [8, Lemma 2.1] *Let $N \in \mathbb{N}$, $\alpha > 0$ be arbitrary and \mathcal{E} satisfy Assumption 1. Then for any $\rho^N = \frac{1}{N} \sum_{i=1}^N \delta_{X^i}$ and $\hat{\rho}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\hat{X}^i}$, it holds*

$$|X_t^\alpha(\rho^N)| \leq \frac{C_{\alpha, \mathcal{E}}}{N} \|\mathbf{X}^N\|_1 \quad (2.1)$$

and

$$|X_t^\alpha(\rho^N) - X_t^\alpha(\hat{\rho}^N)| \leq \left(\frac{C_{\alpha, \mathcal{E}}}{N} + \frac{2\alpha C_{\alpha, \mathcal{E}} L}{N} \|\hat{\mathbf{X}}^N\|_\infty \right) \|\mathbf{X}^N - \hat{\mathbf{X}}^N\|_1, \quad (2.2)$$

where $C_{\alpha,\varepsilon} = e^{\alpha\delta\varepsilon}$ with $\delta\varepsilon = \bar{\varepsilon} - \varepsilon$. Here we used the notations for norms of vectors $\|\mathbf{X}^N\|_\infty = \sup_{i \in [N]} |X^i|$ and $\|\mathbf{X}^N\|_1 = \sum_{i=1}^N |X^i|$. Let $X_t^\alpha(\rho)$ be as in (1.4), then it holds

$$|X_t^\alpha(\rho)| \leq C_{\alpha,\varepsilon} \int_{\mathbb{R}^d} |x| \rho(t, x) dx = C_{\alpha,\varepsilon} \iint_{\mathbb{R}^d} |x| f(t, x, v) dx dv \leq C(C_{\alpha,\varepsilon}, m_{f_t}^2). \quad (2.3)$$

Theorem 2.1 (Well-posedness of (1.1)). *Let $f_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$, and assume that $(X_0^i, V_0^i)_{i=1}^N$ are i.i.d. with the common law f_0 . Then the particle system (1.1) admits a unique global solution with the initial data $(X_0^i, V_0^i)_{i=1}^N$. Moreover one has*

$$\mathbb{E} \left[\sum_{i=1}^N (|X_t^i|^2 + |V_t^i|^2) \right] \leq \mathbb{E} \left[\sum_{i=1}^N (|X_0^i|^2 + |V_0^i|^2) \right] e^{Ct}, \quad (2.4)$$

for all $t \geq 0$, where C depends only on λ, m, σ and $C_{\alpha,\varepsilon}$.

Proof. We can rewrite (1.1) as the SDE

$$d\mathbf{Z}_t^N = \sigma^N(\mathbf{Z}_t^N) d\mathbf{B}_t^N + \mathbf{b}(\mathbf{Z}_t^N) dt, \quad (2.5)$$

in \mathbb{R}^{2dN} , where $\mathbf{Z}_t^N = (X_t^1, V_t^1, \dots, X_t^N, V_t^N)$. Then it holds

$$\begin{aligned} \|\sigma^N(\mathbf{Z}^N)\| &\leq \frac{\sigma}{m} |X_t^\alpha(\rho^N) - X_t^i|^2 \leq \frac{2\sigma}{m} (|X^i|^2 + |X^\alpha(\rho^N)|^2) \\ &\leq \frac{2\sigma}{m} (\|\mathbf{Z}^N\|^2 + \frac{C_{\alpha,\varepsilon}^2}{N^2} (\sum_{i=1}^N |X^i|^2)) \leq \frac{2\sigma}{m} (\|\mathbf{Z}^N\|^2 + \frac{C_{\alpha,\varepsilon}^2}{N} \|\mathbf{Z}^N\|^2). \end{aligned} \quad (2.6)$$

Moreover one has

$$\begin{aligned} \langle \mathbf{Z}^N, \mathbf{b}(\mathbf{Z}^N) \rangle &= \sum_{i=1}^N X^i \cdot V^i - \sum_{i=1}^N \frac{\gamma}{m} |V^i|^2 + \sum_{i=1}^N \frac{\lambda}{m} V^i \cdot (X^\alpha(\rho^N) - X^i) \\ &\leq \left(\frac{1}{2} + \frac{\gamma}{2m}\right) \|\mathbf{Z}^N\|^2 + \frac{\lambda}{2m} \left(\sum_{i=1}^N |V^i|^2 + N |X^\alpha(\rho^N)|^2\right) \\ &\leq \left(\frac{1}{2} + \frac{\gamma}{m}\right) \|\mathbf{Z}^N\|^2 + \frac{N\lambda}{2m} \frac{C_{\alpha,\varepsilon}^2}{N^2} (\sum_{i=1}^N |X^i|^2) \leq \left(\frac{1}{2} + \frac{\gamma}{m} + \frac{\lambda C_{\alpha,\varepsilon}^2}{2m}\right) \|\mathbf{Z}^N\|^2. \end{aligned} \quad (2.7)$$

This is a sufficient condition for the global existence and path-wise uniqueness [7, Chapter 5, Theorems 3.7 and 3.11]. Moreover for any $i \in [N]$ we compute

$$\frac{d}{dt} \mathbb{E}[|X_t^i|^2 + |V_t^i|^2] = 2X^i \cdot V^i - \frac{2\gamma}{m} |V^i|^2 + \frac{2\lambda}{m} V^i \cdot (X^\alpha(\rho^N) - X^i) + \frac{\sigma^2}{m^2} \sum_{k=1}^d (X^\alpha(\rho^N) - X^i)_k^2. \quad (2.8)$$

Summation of the previous equality over $i \in [N]$ yields

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[\|\mathbf{Z}_t^N\|^2] &\leq 2\mathbb{E}[\langle \mathbf{Z}_t^N, \mathbf{b}(\mathbf{Z}_t^N) \rangle] + \frac{2N\sigma^2}{m^2} \mathbb{E}[|X_t^\alpha(\rho^N)|^2] + \frac{2\sigma^2}{m^2} \mathbb{E}[\|\mathbf{Z}_t^N\|^2] \\ &\leq \left(1 + \frac{2\gamma}{m} + \frac{2\lambda C_{\alpha,\varepsilon}^2}{2m} + \frac{2\sigma^2}{m^2} + \frac{2\sigma^2 C_{\alpha,\varepsilon}^2}{m^2}\right) \|\mathbf{Z}_t^N\|^2, \end{aligned} \quad (2.9)$$

which leads to (2.4) by Gronwall's inequality. \square

To prove the well-posedness of (1.3), let us first introduce a related linear problem:

$$\begin{cases} d\bar{X}_t = \bar{V}_t dt, \\ d\bar{V}_t = -\frac{\gamma}{m} \bar{V}_t dt + \frac{\lambda}{m} (X_t^\alpha(\tilde{\rho}) - \bar{X}_t) dt + \frac{\sigma}{m} D(X_t^\alpha(\tilde{\rho}) - \bar{X}_t) dB_t, \end{cases} \quad (2.10)$$

where $X_t^\alpha(\tilde{\rho}) = \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\varepsilon(x) \tilde{\rho}(t, x) dx}{\int_{\mathbb{R}^d} \omega_\alpha^\varepsilon(x) \tilde{\rho}(t, x) dx}$, $\tilde{\rho}(t, x) = \int_{\mathbb{R}^d} \tilde{f}(t, x, v) dv$ for any given $\tilde{f} : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^{2d})$.

Proposition 2.2 (Well-posedness of (2.10)). *For any $T > 0$, let $f_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$ and $\tilde{f} : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^{2d})$ be a given continuous curve in the W_2 topology. Assume that (\bar{X}_0, \bar{V}_0) is distributed according to f_0 . Then the linear SDE (2.10) admits a unique solution up to time T with the initial data (\bar{X}_0, \bar{V}_0) . Moreover one has*

$$\sup_{t \in [0, T]} \mathbb{E}[|\bar{X}_t|^2 + |\bar{V}_t|^2] \leq (\mathbb{E}[|\bar{X}_0|^2 + |\bar{V}_0|^2] + C) e^{CT}, \quad (2.11)$$

where C depends only on $\lambda, m, \sigma, C_{\alpha, \varepsilon}$ and $\sup_{t \in [0, T]} m_{\tilde{f}_t}^2$. Here we have used the second moment notation $m_{\tilde{f}_t}^2 := \iint_{\mathbb{R}^{2d}} |x|^2 \tilde{f}(t, x, v) dx dv$.

Proof. We rewrite (2.10) as

$$dZ_t = \sigma(Z_t) dB_t + b(Z_t) dt. \quad (2.12)$$

It easy to check that

$$\|\sigma(Z)\| \leq \frac{2\sigma}{m} \left(C_{\alpha, \varepsilon}^2 \iint_{\mathbb{R}^{2d}} |x|^2 \tilde{f}(t, x, v) dx dv + |Z|^2 \right) \quad (2.13)$$

and

$$\langle Z, b(Z) \rangle = \bar{X} \cdot \bar{V} - \frac{\gamma}{m} |\bar{V}|^2 + \frac{\lambda}{m} \bar{V} \cdot (X^\alpha(\tilde{\rho}) - \bar{X}) \leq \left(\frac{1}{2} + \frac{\lambda}{m}\right) |Z|^2 + \frac{\lambda}{2m} C_{\alpha, \varepsilon}^2 \iint_{\mathbb{R}^{2d}} |x|^2 \tilde{f}(t, x, v) dx dv.$$

Notice that $\tilde{f} : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^{2d})$, so we again obtain a sufficient condition for the global existence and path-wise uniqueness.

Moreover by Itô's formula one has

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[|\bar{X}_t|^2 + |\bar{V}_t|^2] &= 2\mathbb{E}[\langle Z_t, b(Z_t) \rangle] + \frac{\sigma^2}{m^2} \sum_{k=1}^d (X_t^\alpha(\tilde{\rho}) - \bar{X}_t)_k^2 \leq 2\mathbb{E}[\langle Z_t, b(Z_t) \rangle] + \frac{2\sigma^2}{m^2} \mathbb{E}[|\bar{X}_t|^2 + |X_t^\alpha(\tilde{\rho})|^2] \\ &\leq C(\lambda, m, \sigma) \mathbb{E}[|\bar{X}_t|^2 + |\bar{V}_t|^2] + C(\lambda, m, \sigma, C_{\alpha, \varepsilon}) \iint_{\mathbb{R}^{2d}} |x|^2 \tilde{f}(t, x, v) dx dv. \end{aligned} \quad (2.14)$$

So Gronwall's inequality implies that the second moment $\mathbb{E}[|\bar{X}_t|^2 + |\bar{V}_t|^2]$ is bounded on $[0, T]$ as desired in (2.11). \square

Before we state our theorem on the well-posedness of (1.3), let us introduce the following lemma.

Lemma 2.2. [8, Lemma 2.2] *Assume that $f, \hat{f} \in \mathcal{P}_2(\mathbb{R}^{2d})$. Then the following stability estimate holds*

$$|X^\alpha(\rho) - X^\alpha(\hat{\rho})| \leq CW_2(\rho, \hat{\rho}), \quad (2.15)$$

where C depends only on $C_{\alpha, \varepsilon}, \alpha, L, m_f^2$ and $m_{\hat{f}}^2$.

Theorem 2.3 (Well-posedness of (1.3)). *For any $T > 0$, let $f_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$, and assume that (\bar{X}_0, \bar{V}_0) is distributed according to f_0 . Then the following nonlinear SDE*

$$\begin{cases} d\bar{X}_t = \bar{V}_t dt, \\ d\bar{V}_t = -\frac{\gamma}{m} \bar{V}_t dt + \frac{\lambda}{m} (X_t^\alpha(\rho) - \bar{X}_t) dt + \frac{\sigma}{m} D(X_t^\alpha(\rho) - \bar{X}_t) dB_t. \end{cases} \quad (2.16)$$

has a unique solution up to time T with the initial data (\bar{X}_0, \bar{V}_0) . Here the nonlinearity comes from the requirement that (\bar{X}_t, \bar{V}_t) has the law f_t , and $\rho_t(x) = \int_{\mathbb{R}^d} f_t(x, v) dv$.

Proof. Take the random variable (\bar{X}_0, \bar{V}_0) with distribution $f_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$. We define the stochastic process $(\bar{X}_t^n, \bar{V}_t^n)$ recursively by

$$\begin{cases} d\bar{X}_t^n = \bar{V}_t^n dt, \\ d\bar{V}_t^n = -\frac{\gamma}{m} \bar{V}_t^n dt + \frac{\lambda}{m} (X_t^\alpha(\rho^{n-1}) - \bar{X}_t^n) dt + \frac{\sigma}{m} D(X_t^\alpha(\rho^{n-1}) - \bar{X}_t^n) dB_t, \end{cases} \quad (2.17)$$

for $n \geq 1$, where

$$X_t^\alpha(\rho^{n-1}) = \frac{\int_{\mathbb{R}^d} x \omega_\alpha^\mathcal{E}(x) \rho^{n-1}(t, x) dx}{\int_{\mathbb{R}^d} \omega_\alpha^\mathcal{E}(x) \rho^{n-1}(t, x) dx}, \quad \rho^{n-1}(t, x) = \int_{\mathbb{R}^d} f^{n-1}(t, x, v) dv. \quad (2.18)$$

Here $f_t^n (n \geq 1)$ is the distribution function of $(\bar{X}_t^n, \bar{V}_t^n)$ and $f_t^0 := f_0$ for all $t \in [0, T]$. Below we will also use the alternate notation $f^n(t, x, v)$ for $f_t^n(x, v)$. According to Proposition 2.2, for any given f^{n-1} , the linear SDE (2.17) is well-posed up to time T , and the second moment $m_{f_t^n}^2$ of $(\bar{X}_t^n, \bar{V}_t^n)$ is bounded on $[0, T]$.

Step 1: Cauchy estimate. Denote

$$y_t^n := \bar{X}_t^{n+1} - \bar{X}_t^n, \quad u_t^n := \bar{V}_t^{n+1} - \bar{V}_t^n, \quad e_t^n := |y_t^n|^2 + |u_t^n|^2. \quad (2.19)$$

We compute by Itô's formula and for any $n \geq 1$,

$$\frac{d}{dt} \mathbb{E}[|y_t^n|^2] = 2\mathbb{E}[\langle y_t^n, u_t^n \rangle] \leq \mathbb{E}[|y_t^n|^2 + |u_t^n|^2] = \mathbb{E}[e_t^n], \quad (2.20)$$

and

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[|u_t^n|^2] &= 2\mathbb{E} \left[\left\langle u_t^n, -\frac{\gamma}{m} u_t^n - \frac{\gamma}{m} y_t^n + \frac{\gamma}{m} (X_t^\alpha(\rho^n) - X_t^\alpha(\rho^{n-1})) \right\rangle \right] + \frac{\sigma^2}{m^2} \sum_{k=1}^d \mathbb{E}[(X_t^\alpha(\rho^n) - X_t^\alpha(\rho^{n-1}) - y_t^n)_k^2] \\ &\leq \left(\frac{\gamma}{m} + \frac{2\sigma^2}{m^2} \right) \mathbb{E}[e_t^n] + 2\mathbb{E} \left[\left\langle u_t^n, \frac{\gamma}{m} (X_t^\alpha(\rho^n) - X_t^\alpha(\rho^{n-1})) \right\rangle \right] + \frac{2\sigma^2}{m^2} \mathbb{E}[|X_t^\alpha(\rho^n) - X_t^\alpha(\rho^{n-1})|^2]. \end{aligned}$$

It follows from Lemma 2.2 and the definition of Wasserstein metric that one has

$$|X_t^\alpha(\rho^n) - X_t^\alpha(\rho^{n-1})|^2 \leq CW_2^2(\rho_t^n, \rho_t^{n-1}) \leq C\mathbb{E}[|y_t^{n-1}|^2], \quad (2.21)$$

where C depends only on $C_{\alpha, \mathcal{E}}, \alpha, L, \sup_{t \in [0, T]} m_{f_t^{n-1}}^2$ and $\sup_{t \in [0, T]} m_{f_t^n}^2$. Thus we have

$$2\langle u_t^n, X_t^\alpha(\rho^n) - X_t^\alpha(\rho^{n-1}) \rangle \leq |u_t^n|^2 + |X_t^\alpha(\rho^n) - X_t^\alpha(\rho^{n-1})|^2 \leq |u_t^n|^2 + C\mathbb{E}[|y_t^{n-1}|^2]. \quad (2.22)$$

This yields

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[|u_t^n|^2] &\leq \left(\frac{\gamma}{m} + \frac{\sigma^2}{m^2} \right) \mathbb{E}[e_t^n] + \frac{\gamma}{m} \mathbb{E}[|u_t^n|^2] + C\mathbb{E}[|y_t^{n-1}|^2] + \frac{\sigma^2}{m^2} C\mathbb{E}[|y_t^{n-1}|^2] \\ &\leq C\mathbb{E}[e_t^n] + \mathbb{E}[e_t^{n-1}]. \end{aligned} \quad (2.23)$$

Collecting estimates (2.20) and (2.23), we conclude that

$$\frac{d}{dt} \mathbb{E}[e_t^n] \leq C\mathbb{E}[e_t^n] + C\mathbb{E}[e_t^{n-1}]. \quad (2.24)$$

Let us denote $\eta^n(t) := \mathbb{E}[e_t^n]$, then one has

$$\frac{d}{dt} \eta^n(t) \leq C\eta^n(t) + C\eta^{n-1}(t). \quad (2.25)$$

Applying Gronwall's inequality leads to

$$\eta^n(t) \leq C \int_0^t e^{C(t-s)} \eta^{n-1}(s) ds, \quad (2.26)$$

and iterating this inequality gives

$$\eta^n(t) \leq C^n \int_0^t e^{C(t-s)} \eta^0(s) \frac{(t-s)^{n-1}}{(n-1)!} ds \leq C^n e^{Ct} t^n \sup_{s \in [0, t]} \eta^0(s) \leq C^n e^{Cnt} t^n \sup_{s \in [0, t]} \eta^0(s). \quad (2.27)$$

Hence we obtain

$$\eta^n(t) \leq \exp(n(Ct + \ln(Ct))) \sup_{s \in [0, t]} \eta^0(s), \quad (2.28)$$

where C depends only on $C_{\alpha, \varepsilon}, \alpha, L, \sup_{t \in [0, T]} m_{f_t}^{2n-1}$ and $\sup_{t \in [0, T]} m_{f_t}^{2n}$. Note here that $\sup_{t \in [0, T]} m_{f_t}^{2n-1}$ and $\sup_{t \in [0, T]} m_{f_t}^{2n}$ can be uniformly bounded in n . Choosing $0 < T_* < T$ small enough such that $\beta := CT_* + \ln(CT_*) < 0$, then

$$\sup_{t \in [0, T_*]} \eta^n(t) \leq e^{\beta n} \sup_{s \in [0, T_*]} \eta^0(s) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (2.29)$$

Step 2: Passing the limit. The last step proves that $\{(\bar{X}_t^n, \bar{V}_t^n)_{t \in [0, T_*]}\}_{n \geq 1}$ is a Cauchy sequence. Denote by $(\bar{X}_t, \bar{V}_t)_{t \in [0, T_*]}$ the limit point such that

$$\sup_{t \in [0, T_*]} \mathbb{E}[|\bar{X}_t^n - \bar{X}_t|^2 + |\bar{V}_t^n - \bar{V}_t|^2] \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (2.30)$$

Assume that $(\bar{X}_t, \bar{V}_t)_{t \in [0, T_*]}$ has the distribution $(f_t)_{t \in [0, T_*]}$, then by the definition of the Wasserstein metric one has

$$\mathcal{W}_2^{T_*}(f^n, f) := \sup_{t \in [0, T_*]} W_2(f_t^n, f_t) \leq \sup_{t \in [0, T_*]} (\mathbb{E}[|\bar{X}_t^n - \bar{X}_t|^2 + |\bar{V}_t^n - \bar{V}_t|^2])^{\frac{1}{2}} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad (2.31)$$

and $f \in C(0, T_*; \mathcal{P}_2(\mathbb{R}^{2d}))$ since $C(0, T_*; \mathcal{P}_2(\mathbb{R}^{2d}))$ is complete w.r.t. the metric $\mathcal{W}_2^{T_*}$. Lastly, taking $n \rightarrow \infty$ (up to a subsequence) in (2.17), one obtains that the limit process $(\bar{X}_t, \bar{V}_t)_{t \in [0, T_*]}$ is a solution to the Nonlinear SDE (2.16) up to time T_* . Next, we can easily extend the existence to the whole interval $[0, T]$ by iterating this procedure starting from T_* .

Step 3: Uniqueness. Assume that $(\bar{X}_t^1, \bar{V}_t^1)_{t \in [0, T]}$ and $(\bar{X}_t^2, \bar{V}_t^2)_{t \in [0, T]}$ are two solutions to SDE (2.16) with the same Brownian motion. Similar to *Step 1*, it is easy to derive

$$\frac{d}{dt} \mathbb{E}[|\bar{X}_t^1 - \bar{X}_t^2|^2] \leq \mathbb{E}[|\bar{X}_t^1 - \bar{X}_t^2|^2 + |\bar{V}_t^1 - \bar{V}_t^2|^2],$$

and

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}[|\bar{V}_t^1 - \bar{V}_t^2|^2] \\ & \leq \left(\frac{\gamma}{m} + \frac{\sigma^2}{m^2}\right) \mathbb{E}[|\bar{X}_t^1 - \bar{X}_t^2|^2 + |\bar{V}_t^1 - \bar{V}_t^2|^2] + \frac{\gamma}{m} \mathbb{E}[|\bar{V}_t^1 - \bar{V}_t^2|^2] + C \mathbb{E}[|\bar{X}_t^1 - \bar{X}_t^2|^2] + \frac{\sigma^2}{m^2} C \mathbb{E}[|\bar{X}_t^1 - \bar{X}_t^2|^2]. \end{aligned}$$

Applying Gronwall's inequality yields

$$\sup_{t \in [0, T]} \mathbb{E}[|\bar{X}_t^1 - \bar{X}_t^2|^2 + |\bar{V}_t^1 - \bar{V}_t^2|^2] \leq \mathbb{E}[|\bar{X}_0^1 - \bar{X}_0^2|^2 + |\bar{V}_0^1 - \bar{V}_0^2|^2] e^{CT}, \quad (2.32)$$

showing uniqueness of solutions to (2.16). \square

3. MEAN-FIELD LIMIT

Let us start this section with introducing the following lemma on a large deviation bound.

Lemma 3.1. [8, Lemma 3.1] *For any $T > 0$, let $\{(\bar{X}_t^i, \bar{V}_t^i)_{t \in [0, T]}\}_{i=1}^N$ be the solution to the mean-field dynamics (1.3), which are i.i.d. with common distribution $f \in C([0, T], \mathcal{P}_2(\mathbb{R}^{2d}))$. Then it holds*

$$\sup_{t \in [0, T]} \mathbb{E}[|X_t^\alpha(\bar{\rho}^N) - X_t^\alpha(\rho)|^2] \leq \left(\sup_{t \in [0, T]} m_{f_t}^2\right) C_{\alpha, \varepsilon}^3 N^{-1}, \quad (3.1)$$

where we have used the notation $\bar{\rho}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{X}^i}$.

We now state our main result on the mean-field limit, which states that the intermediate dynamics (1.3) well approximates the interacting particle system (1.1).

Theorem 3.1. For any $T > 0$, let $f_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$, and assume that $\{(X_t^i, V_t^i)_{t \in [0, T]}\}_{i=1}^N$ and $\{(\bar{X}_t^i, \bar{V}_t^i)_{t \in [0, T]}\}_{i=1}^N$ are solutions to (1.1) and (1.3) respectively with the same i.i.d. initial data distributed according to f_0 . Then it holds that

$$\sup_{t \in [0, T]} \sup_{i=1, \dots, N} \mathbb{E}[|X_t^i - \bar{X}_t^i|^2 + |V_t^i - \bar{V}_t^i|^2] \leq CN^{-1}, \quad (3.2)$$

where C depends only on $\gamma, \lambda, \sigma, m, T, m_{f_0}^2$ and $C_{\alpha, \varepsilon}$.

Remark 3.1. By the assumption $f_0 \in \mathcal{P}_2(\mathbb{R}^{2d})$ and the computation similar to (2.14), one can easily obtain a uniform bound of the second moment: $\sup_{t \in [0, T]} m_{f_t}^2 \leq m_{f_0}^2 e^{C(\lambda, m, \sigma, C_{\alpha, \varepsilon})T}$.

Proof. For all $i = 1, \dots, N$, one can compute

$$\frac{d}{dt} \mathbb{E}[|X_t^i - \bar{X}_t^i|^2] = 2\mathbb{E}[\langle X_t^i - \bar{X}_t^i, V_t^i - \bar{V}_t^i \rangle] \leq \mathbb{E}[|X_t^i - \bar{X}_t^i|^2 + |V_t^i - \bar{V}_t^i|^2]. \quad (3.3)$$

Applying the Itô formula to $|V_t^i - \bar{V}_t^i|^2$ we get

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}[|V_t^i - \bar{V}_t^i|^2] \\ &= 2\mathbb{E} \left[\left\langle V_t^i - \bar{V}_t^i, -\frac{\gamma}{m}(V_t^i - \bar{V}_t^i) - \frac{\gamma}{m}(X_t^i - \bar{X}_t^i) + \frac{\gamma}{m}(X_t^\alpha(\rho^N) - X_t^\alpha(\rho)) \right\rangle \right] + \frac{\sigma^2}{m^2} \sum_{k=1}^d \mathbb{E}[(X_t^\alpha(\rho^N) - X_t^\alpha(\rho) - (X_t^i - \bar{X}_t^i))_k^2] \\ &\leq \left(\frac{\gamma}{m} + \frac{2\sigma^2}{m^2} \right) \mathbb{E}[|X_t^i - \bar{X}_t^i|^2 + |V_t^i - \bar{V}_t^i|^2] + 2\mathbb{E} \left[\left\langle V_t^i - \bar{V}_t^i, \frac{\gamma}{m}(X_t^\alpha(\rho^N) - X_t^\alpha(\rho)) \right\rangle \right] + \frac{2\sigma^2}{m^2} \mathbb{E}[|X_t^\alpha(\rho^N) - X_t^\alpha(\rho)|^2] \\ &\leq \left(\frac{2\gamma}{m} + \frac{2\sigma^2}{m^2} \right) \mathbb{E}[|X_t^i - \bar{X}_t^i|^2 + |V_t^i - \bar{V}_t^i|^2] + \left(\frac{\gamma}{m} + \frac{2\sigma^2}{m^2} \right) \mathbb{E}[|X_t^\alpha(\rho^N) - X_t^\alpha(\rho)|^2]. \end{aligned}$$

It follows from Lemma 2.2, Lemma 3.1 and the definition of Wasserstein distance that

$$\begin{aligned} & \mathbb{E}[|X_t^\alpha(\rho^N) - X_t^\alpha(\rho)|^2] \leq 2\mathbb{E}[|X_t^\alpha(\rho^N) - X_t^\alpha(\bar{\rho}^N)|^2] + 2\mathbb{E}[|X_t^\alpha(\bar{\rho}^N) - X_t^\alpha(\rho)|^2] \\ &\leq 2C\mathbb{E}[W_2^2(\rho^N, \bar{\rho}^N)] + 2\left(\sup_{t \in [0, T]} m_{f_t}^2\right) C_{\alpha, \varepsilon}^3 N^{-1} \leq 2C \sup_{i \in [N]} \mathbb{E}[|X_t^i - \bar{X}_t^i|^2] + 2\left(\sup_{t \in [0, T]} m_{f_t}^2\right) C_{\alpha, \varepsilon}^3 N^{-1}, \end{aligned}$$

where C depends only on $C_{\alpha, \varepsilon}, \alpha, L, T$ and $m_{f_0}^2$. This implies that

$$\frac{d}{dt} \mathbb{E}[|V_t^i - \bar{V}_t^i|^2] \leq C \sup_{i \in [N]} \mathbb{E}[|X_t^i - \bar{X}_t^i|^2 + |V_t^i - \bar{V}_t^i|^2] + CN^{-1}. \quad (3.4)$$

where C depends only on $\gamma, \lambda, \sigma, m, T, m_{f_0}^2$ and $C_{\alpha, \varepsilon}$. Collecting (3.3) and (3.4), one has

$$\sup_{i \in [N]} \mathbb{E}[|X_t^i - \bar{X}_t^i|^2 + |V_t^i - \bar{V}_t^i|^2] \leq C \int_0^t \sup_{i \in [N]} \mathbb{E}[|X_s^i - \bar{X}_s^i|^2 + |V_s^i - \bar{V}_s^i|^2] ds + CN^{-1}t,$$

which leads to

$$\sup_{i \in [N]} \mathbb{E}[|X_t^i - \bar{X}_t^i|^2 + |V_t^i - \bar{V}_t^i|^2] \leq CN^{-1}te^{Ct}, \quad (3.5)$$

for all $t \in [0, T]$, where C depends only on $\gamma, \lambda, T, \sigma, m, m_{f_0}^2$ and $C_{\alpha, \varepsilon}$. \square

Remark 3.2. We point out here that the proof in this note can be easily extended to the case of the PSO dynamic with memory and local best [10, Equation (3.6)] since the only difference there is an additional variable Y_t^i , which satisfies an ODE with globally Lipschitz coefficients.

REFERENCES

- [1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows: In metric spaces and in the space of probability measures*, Springer Science & Business Media, 2008.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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).
- [6] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, Springer-Verlag Berlin Heidelberg, 2010.
- [7] Stewart N Ethier and Thomas G Kurtz, *Markov processes: characterization and convergence*, Vol. 282, John Wiley & Sons, 2009.
- [8] 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.
- [9] Massimo Fornasier, Hui Huang, Lorenzo Pareschi, and Philippe Sünnen, *Consensus-based optimization on the sphere II: Convergence to global minimizers and machine learning*, arXiv:2001.11988v3 (2020).
- [10] Sara Grassi and Lorenzo Pareschi, *From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit*, arXiv preprint arXiv:2012.05613 (2020).
- [11] 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.
- [12] 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.
- [13] Pierre-Emmanuel Jabin and Zhenfu Wang, *Mean field limit for stochastic particle systems*, *Active particles*, volume 1, 2017, pp. 379–402.
- [14] James Kennedy, *The particle swarm: social adaptation of knowledge*, *Proceedings of 1997 IEEE International Conference on Evolutionary Computation*, 1997, pp. 303–308.
- [15] James Kennedy and Russell Eberhart, *Particle swarm optimization*, *Proceedings of 1995 IEEE International Conference on Neural Networks*, 1995, pp. 1942–1948.
- [16] Peter David Miller, *Applied asymptotic analysis*, Vol. 75, American Mathematical Soc., 2006.
- [17] 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.
- [18] Riccardo Poli, James Kennedy, and Tim Blackwell, *Particle swarm optimization*, *Swarm intelligence* **1** (2007), no. 1, 33–57.
- [19] Yuhui Shi and Russell Eberhart, *A modified particle swarm optimizer*, *Proceedings of 1998 IEEE international conference on evolutionary computation*, 1998, pp. 69–73.
- [20] Alain-Sol Sznitman, *Topics in propagation of chaos*, *Ecole d’été de probabilités de Saint-Flour XIX1989*, 1991, pp. 165–251.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF CALGARY, 2500 UNIVERSITY DR NW, CALGARY, AB T2N 1N4, CANADA

E-mail address: hui.huang1@ucalgary.ca