

# Long run convergence of discrete-time interacting particle systems of the McKean-Vlasov type

Pascal Bianchi<sup>1</sup>, Walid Hachem<sup>2</sup>, and Victor Priser<sup>1</sup>

<sup>1</sup>LTCI, Télécom Paris

<sup>2</sup>Laboratoire d'informatique Gaspard Monge (LIGM / UMR 8049), Université Gustave Eiffel

February 10, 2025

## Abstract

We consider a discrete time system of  $n$  coupled random vectors, a.k.a. interacting particles. The dynamics involves a vanishing step size, some random centered perturbations, and a mean vector field which induces the coupling between the particles. We study the doubly asymptotic regime where both the number of iterations and the number  $n$  of particles tend to infinity, without any constraint on the relative rates of convergence of these two parameters. We establish that the empirical measure of the interpolated trajectories of the particles converges in probability, in an ergodic sense, to the set of recurrent Mc-Kean-Vlasov distributions. A first application example is the granular media equation, where the particles are shown to converge to a critical point of the Helmholtz energy. A second example is the convergence of stochastic gradient descent to the global minimizer of the risk, in a wide two-layer neural networks using random features.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The setting</b>	<b>5</b>
2.1	Notations . . . . .	5
2.1.1	General notations . . . . .	5
2.1.2	Random variables . . . . .	5
2.2	Spaces of probability measures . . . . .	6
2.3	Spaces of McKean-Vlasov measures . . . . .	7
2.4	Dynamical systems . . . . .	8
<b>3</b>	<b>Main results</b>	<b>8</b>
3.1	Interpolated process and weak $\star$ limits . . . . .	8
3.2	Ergodic convergence . . . . .	9
3.3	Pointwise convergence to a global attractor . . . . .	10
<b>4</b>	<b>Granular media</b>	<b>10</b>
<b>5</b>	<b>Proofs of Section 3</b>	<b>12</b>
5.1	Proof of Proposition 4 . . . . .	12
5.2	Proof of Proposition 5 . . . . .	15
5.3	Proof of Theorem 1 . . . . .	20
5.4	Proof of Corollary 1 . . . . .	22
5.5	Proof of Theorem 2 . . . . .	24

<b>6</b>	<b>Proofs of Section 4</b>	<b>25</b>
6.1	Proof of Prop. 6	25
6.2	Proof of Prop. 7	29
6.3	Proof of Prop. 8	30
6.4	Proof of Prop. 9	30
6.5	Proof of Th. 4	33
<b>A</b>	<b>Technical proofs</b>	<b>33</b>
A.1	Proof of Proposition 1	33
A.2	Proof of Lemma 2	33
A.3	Proof of Lemma 3	34

## 1 Introduction

Given two integers  $n, d > 0$ , consider the iterative algorithm defined as follows. Starting with the  $n$ -uple  $(X_0^{1,n}, \dots, X_0^{n,n})$  of random variables  $X_0^{i,n} \in \mathbb{R}^d$ , the algorithm generates at the iteration  $k+1$  for  $k \in \mathbb{N}$  the  $n$ -uple of  $\mathbb{R}^d$ -valued random variables  $(X_k^{1,n}, \dots, X_k^{n,n})$ , referred to as the *particles*, according to the dynamics:

$$X_{k+1}^{i,n} = X_k^{i,n} + \frac{\gamma_{k+1}}{n} \sum_{j=1}^n b(X_k^{i,n}, X_k^{j,n}) + \sqrt{2\gamma_{k+1}} \xi_{k+1}^{i,n} + \gamma_{k+1} \zeta_{k+1}^{i,n}, \quad (1)$$

for each  $i \in [n]$  where  $[n] := \{1, \dots, n\}$ . In this equation, the function  $b : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a continuous vector field,  $(\gamma_k)_k$  is a vanishing sequence of deterministic positive step sizes, and  $((\xi_k^{i,n})_{i \in [n]})_{k \in \mathbb{N}^*}$  and  $((\zeta_k^{i,n})_{i \in [n]})_{k \in \mathbb{N}^*}$  are  $\mathbb{R}^{d \times n}$ -valued random noise sequences in the time parameter  $k$ . We assume that for each  $n$ , the  $n$ -uple  $(X_0^{1,n}, \dots, X_0^{n,n})$  is exchangeable, and that the same holds for the  $n$ -uple of sequences  $((\xi_k^{1,n})_{k \in \mathbb{N}^*}, \dots, (\xi_k^{n,n})_{k \in \mathbb{N}^*})$  and  $((\zeta_k^{1,n})_{k \in \mathbb{N}^*}, \dots, (\zeta_k^{n,n})_{k \in \mathbb{N}^*})$ . Defining, for each  $n > 0$ , the filtration  $(\mathcal{F}_k^n)_{k \in \mathbb{N}}$  as

$$\mathcal{F}_k^n := \sigma((X_0^{i,n})_{i \in [n]}, ((\xi_\ell^{i,n})_{i \in [n]})_{\ell \leq k}, ((\zeta_\ell^{i,n})_{i \in [n]})_{\ell \leq k}), \quad (2)$$

we furthermore assume that for each  $n$ , the sequence  $((\xi_k^{i,n})_{i \in [n]})_k$  is a  $(\mathcal{F}_k^n)_k$ -martingale increment sequence *i.e.*,  $\mathbb{E}(\xi_{k+1}^{i,n} | \mathcal{F}_k^n) = 0$ . Finally, we assume the following  $\mathbb{E}(\xi_{k+1}^{i,n} (\xi_{k+1}^{j,n})^T | \mathcal{F}_k^n) = \sum_{q, \ell \in [n]} \sigma(X_k^{i,n}, X_k^{j,n}) \sigma(X_k^{\ell,n}, X_k^{\ell,n})^T \mathbf{1}_{i=j}$  for some  $\sigma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d'}$ , with  $d' > 0$ .

The aim of the paper is to characterize the asymptotic behavior of the empirical measure of the particles

$$\mu_k^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_k^{i,n}} \quad (3)$$

in the regime where both the time index  $k$  and the number of particles  $n$  tend to infinity (denoted hereinafter as  $(k, n) \rightarrow (\infty, \infty)$ ), without any constraint on the relative rates of convergence of these two parameters. To this end, we consider for each  $i \in [n]$  the random continuous process  $\bar{X}^{i,n} : [0, \infty) \rightarrow \mathbb{R}^d, t \mapsto \bar{X}_t^{i,n}$  defined as the piecewise linear interpolation of the particles  $(X_k^{i,n})_k$ . Specifically, writing

$$\tau_k := \sum_{j=1}^k \gamma_j \quad (4)$$

for each  $k \in \mathbb{N}$ , we define:

$$\forall t \in [\tau_k, \tau_{k+1}), \quad \bar{X}_t^{i,n} := X_k^{i,n} + \frac{t - \tau_k}{\gamma_{k+1}} (X_{k+1}^{i,n} - X_k^{i,n}). \quad (5)$$

The interpolated processes  $\bar{X}^{i,n}$ , for  $i \in [n]$ , are elements of the set  $\mathcal{C}$  of the  $[0, \infty) \rightarrow \mathbb{R}^d$  continuous functions, equipped with the topology of uniform convergence on compact intervals. This paper studies the empirical measure of these processes:

$$m^n := \frac{1}{n} \sum_{i=1}^n \delta_{\bar{X}^{i,n}}. \quad (6)$$

For each  $n$  and each  $p \in [1, 2]$ ,  $m^n$  is a random variable on the space  $\mathcal{P}_p(\mathcal{C})$  of probability measures on  $\mathcal{C}$  with a finite  $p$ -moment, equipped with the  $p$ -Wasserstein metric  $W_p$  (precise definitions of these notions provided below). Our aim is to analyze the convergence in probability, of the shifted random measures

$$\Phi_t(m^n) = \frac{1}{n} \sum_{i=1}^n \delta_{\bar{X}_{t+}^{i,n}},$$

when both  $n$  and  $t$  converge to infinity with arbitrary relative rates, where for every  $m \in \mathcal{P}_p(\mathcal{C})$ ,  $\Phi_t(m) \in \mathcal{P}_p(\mathcal{C})$  is defined by  $\Phi_t(m)(f) = \int f(x(t + \cdot)) dm(x)$  for every bounded continuous function  $f$  on  $\mathcal{C}$ . Under mild assumptions on the vector field  $b$ , and some moment assumptions on the iterates and on the noise sequence  $((\zeta_k^{i,n})_{i \in [n]})_k$ , ensuring that the effect of the latter becomes negligible in our asymptotic regime, we establish the following result, which we explain hereafter.

**Main theorem (informal).** The sequence  $(\Phi_t(m^n))$  ergodically converges in probability as  $(t, n) \rightarrow (\infty, \infty)$  to the set of *recurrent McKean-Vlasov distributions*.

Let us explain what the terms *McKean-Vlasov distribution*, *recurrent*, and *ergodic convergence* mean in this paper. Here, a McKean-Vlasov distribution  $\rho$  is defined as the law of a  $\mathbb{R}^d$ -valued process  $(X_t : t \in \mathbb{R})$  satisfying the following condition: for every smooth enough compactly supported function  $\phi$ , the process

$$\phi(X_t) - \int_0^t L(\rho_s)(\phi)(X_s) ds$$

is a martingale, where  $\rho_t$  the marginal law of  $X_t$ , and where the linear operator  $L(\rho_t)$  associates to  $\phi$  the function  $L(\rho_t)(\phi)$  given by:

$$x \mapsto \langle b(x, \rho_t), \nabla \phi(x) \rangle + \text{tr}(\sigma(x, \rho_t)^T H_\phi(x) \sigma(x, \rho_t)),$$

where  $H_\phi$  is the Hessian matrix of  $\phi$ ,  $\text{tr}$  denotes the Trace operator, and where we use the slightly abusive notation  $b(x, \rho_t) := \int b(x, y) d\rho_t(y)$ , and  $\sigma(x, \rho_t) := \int \sigma(x, y) d\rho_t(y)$ .

A McKean-Vlasov distribution  $\rho$  is said recurrent if, for some sequence  $(t_k) \rightarrow \infty$ ,  $\rho = \lim_{k \rightarrow \infty} \Phi_{t_k}(\rho)$ . The  $W_p$ -closure of the set of recurrent McKean-Vlasov distributions will be referred to as the *Birkhoff center*, and denoted by  $\text{BC}_p$ , following the terminology used for general dynamical systems.

By *ergodic convergence*, we refer to the fact that the time averaged Wasserstein distance between the measures  $\Phi_t(m^n)$  and the Birkhoff center converges to zero. Our main theorem can thus be written more precisely:

$$\frac{1}{t} \int_0^t W_p(\Phi_s(m^n), \text{BC}_p) ds \xrightarrow{(t,n) \rightarrow (\infty, \infty)} 0, \text{ in probability.}$$

The Birkhoff center can be characterized in a useful way, provided that one is able to show the existence of a *Lyapunov function*, namely a function  $F$  on  $\mathcal{P}_p(\mathcal{C})$  such that, for every McKean-Vlasov distribution  $\rho$ ,  $F(\Phi_t(\rho))$  is non-increasing in the variable  $t$ . Indeed, in such a situation, the Birkhoff center is included in the subset  $\Lambda$  of McKean-Vlasov distributions which satisfy the property that  $t \mapsto F(\Phi_t(\rho))$  is constant whenever  $\rho \in \Lambda$ .

Finally, in the case where the McKean-Vlasov dynamics can be cast in the form of a gradient flow in the space of measures  $\mathcal{P}_p(\mathbb{R}^d)$ , and in case this gradient flow has a global attractor  $A_p$ , we show that

$$W_p(\mu_k^n, A_p) \xrightarrow{(k,n) \rightarrow (\infty, \infty)} 0 \text{ in probability.}$$

To illustrate our results, we provide an important example of a McKean-Vlasov distribution where these results can be applied: the granular media equation. Additionally, our results can also be applied in several machine learning applications, such as two-layer neural networks or the Stein Variational Gradient Descent (SVGD) algorithm.

**Granular media.** Our example is in  $\mathcal{P}_2(\mathcal{C})$  and corresponds to the scenario where  $\sigma(x, y) = \sigma I_d$  for some real constant  $\sigma \geq 0$ , and the vector field  $b$  takes the form:

$$b(x, y) = -\nabla V(x) - \nabla U(x - y),$$

where the *confinement potential*  $V$  and the *interaction potential*  $U$  denote two real differentiable functions on  $\mathbb{R}^d$ , whose gradients satisfy some linear growth condition. In this case, a Lyapunov

function exists, which can be expressed as a function of the so-called *Helmholtz energy*. As a consequence of our main result, we establish that, when  $\sigma > 0$ , the empirical measures  $(\mu_k^n)$  converge ergodically in probability as  $(k, n) \rightarrow (\infty, \infty)$  to the set  $\mathcal{S}$  of critical points of the Helmholtz energy, namely:

$$\frac{\sum_{l=1}^k \gamma_l W_2(\mu_k^n, \mathcal{S})}{\sum_{l=1}^k \gamma_l} \xrightarrow{(n,k) \rightarrow (\infty, \infty)} 0, \quad \text{in probability.}$$

where, this time,  $W_2$  represents the classical Wasserstein distance, and where  $\mathcal{S}$  is the set of probability measures  $\mu$  on  $\mathbb{R}^d$  which admit a second order moment and a density  $d\mu/d\mathcal{L}^d$  w.r.t. the Lebesgue measure, and such that:

$$\nabla V(x) + \int \nabla U(x - y) d\mu(y) + \sigma^2 \nabla \log \frac{d\mu}{d\mathcal{L}^d}(x) = 0,$$

for  $\mu$ -almost every  $x$ . Our result holds under mild assumptions, and does not require the rather classical strong convexity or doubling conditions on  $U$  and/or  $V$ .

**Contributions.** Compared to existing works, our contributions are threefold. First, our results hold under mild assumptions on the vector field  $b$  aside from continuity and linear growth, whereas most of the existing works (see below) rely on stronger conditions, such as Lipschitz, doubling or even global boundedness conditions. Second, we address the case of discrete-time systems with a step size vanishing arbitrarily slowly towards 0, whereas the continuous time model is more often considered in the literature. Discrete-time algorithms are important in applications, such as neural networks, transformers, Monte Carlo simulations or numerical solvers. In particular, stability results are more difficult to establish in this setting. Finally, our result focuses on a double limit  $(k, n) \rightarrow (\infty, \infty)$ . At the exception of some papers listed below, the results of the same kind generally consider the case, where the time window is fixed, while the number of particles grows to infinity, ignoring long time convergence, or assume certain constraints on the relative rate of convergence of the two variables.

**About the literature.** The first results addressing the limiting behavior of a finite system of particles are provided in the context of the propagation of chaos. These findings are discussed in detail in [CD22]. Such results have broad applicability across a variety of particle systems, where the interacting term  $b$  can manifest in various forms [MRC87, Oel84, Szn84, ELL21]. In our case, if we set aside the transition from continuous to discrete time, such results typically establish the convergence to zero of the expectation of the squared Wasserstein distance between the empirical measure of the particles, over some fixed time interval  $[0, T]$ , and a McKean-Vlasov distribution with the same initial measure. Under classical assumptions, this convergence occurs at a rate of  $1/n$ , where  $n$  is the number of particles, but with a constant that grows exponentially with  $T$ . This type of result performs poorly in the long run, making the achievement of the double limit in both time and the number of particles unattainable.

By imposing additional assumptions, one can derive a bound that is uniform in time, thereby explicitly addressing the double asymptotic regime. However, these uniform-in-time propagation of chaos results are typically established in continuous time. The paper [KJHK24] bridges the gap between continuous and discrete time in the specific context where uniform-in-time propagation of chaos holds for the continuous-time particle system, allowing for the recovery of our results. They demonstrate that the limiting distribution of the discrete-time particle system coincides with that of the continuous-time particle system. When uniform-in-time propagation of chaos holds, the limiting distributions of the continuous-time particle system converge to the unique stationary distribution of the associated McKean-Vlasov system as time grows. This, in turn, implies the convergence of the discrete-time particle system to the McKean-Vlasov stationary distribution in the doubly asymptotic regime. However, it should be noted that when applying the results of [KJHK24], we lose the convergence rate provided by uniform-in-time propagation of chaos, and the resulting result is no better than ours in the restrictive case where it is applicable.

Our contribution lies in the fact that our assumptions are weaker than those requiring uniform-in-time propagation of chaos, which are generally too strong for practical applications. Specifically, the first paper to address uniform-in-time propagation of chaos in the granular media setting is [Mal01], which requires the strong convexity of the confinement potential and the convexity of the interaction potential. Later, [CGM08] relaxed the strong convexity assumption on the confinement potential. [DEGZ20] proposed a uniform-in-time propagation of chaos result when the confinement potential is strongly convex outside a ball, and the interaction potential has a sufficiently small

Lipschitz constant. More recently, [MRW24, CLRW24, LLF23] provide sharp uniform-in-time propagation of chaos results under a Log-Sobolev inequality on the vector field  $b$  and a noise with variance large enough.

As highlighted in [DMT19], achieving uniform propagation of chaos over time is only possible when a unique McKean-Vlasov stationary distribution exists. A condition that [Tug13] has demonstrated is not always met. In this regard, our assumptions are weaker, allowing for the existence of multiple stationary distributions. It is noteworthy that the study of McKean-Vlasov stationary distributions in cases where the uniqueness of such distributions does not hold remains an open area of research. For instance, [Cor23] explores the stability of stationary distributions. Additionally, [BDFR15] explores a general class of non-linear Markov processes in finite-dimensional space and proposes a method to obtain Lyapunov functions for these processes.

Among papers that address the long-run convergence of discrete-time particle systems, [Mal03] employs an implicit Euler scheme for the granular media case, assuming a zero potential function and strongly convex interaction. The work in [BCEM24] studies a Jordan–Kinderlehrer–Otto (JKO) scheme for granular media, assuming a strongly convex confinement potential. The contribution of [Ver06] is the closest to the present work, considering an equation similar to Eq. (1), but assumes that  $b$  is globally bounded and only addresses the convergence of the expectation of the empirical measure, not convergence in probability. Lastly, [BS00] is closely related but not specific to McKean-Vlasov processes, as it does not consider particle systems or double limits. However, it establishes ergodic convergence of the empirical measure of a weak asymptotic pseudotrajectory to the Birkhoff center of a flow on a metric space, similar in spirit to our approach.

Finally, let us review some applications of our model. Particle systems have historically been motivated by statistical physics. However, in recent decades, they have found utility in various models including neural networks, Markov Chain Monte Carlo theory, mathematical biology, and mean fields game, among others. A well-known model in statistical physics is granular media [Vil06]. This model has been extensively studied due to its property of being a gradient system, and the uniform propagation of chaos over time works well within this model. It can also be described by a gradient flow [AGS08]. In Markov Chain Monte Carlo theory, the Stein Variational Gradient Descent estimates a target distribution using a particle system [LW16, SSR22], and the convergence of this algorithm remains an open question. Wide Neural Networks can also be represented by particle systems. A convergence result to the minimizers of the risk is attainable when both time and the number of particles tend to infinity [CB18]. Here, the authors establish convergence to gradient descent in continuous time and in the double asymptotic regime. The paper [MMN18] establishes the convergence of noisy stochastic gradient descent when the number of iterations depends on the number of particles. See also [RVE22, SS20, HRŠS21, Chi22, NWS22] for related works.

## 2 The setting

We begin by introducing some notations and by recalling some definitions.

### 2.1 Notations

#### 2.1.1 General notations

We denote by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$  the inner product and the corresponding norm in a Euclidean space. We use the same notation in an infinite dimensional space, to denote the standard dual pairing and the operator norm.

For  $k \in \mathbb{N} \cup \{\infty\}$ , we denote by  $C^k(\mathbb{R}^d, \mathbb{R}^q)$  the set of functions which are continuously differentiable up to the order  $k$ . We denote by  $C_c(\mathbb{R}^d, \mathbb{R})$  the set of  $\mathbb{R}^d \rightarrow \mathbb{R}$  continuous functions with compact support. Given  $p \in \mathbb{N}^* \cup \{\infty\}$ , we denote as  $C_c^p(\mathbb{R}^d, \mathbb{R})$  the set of compactly supported  $\mathbb{R}^d \rightarrow \mathbb{R}$  functions which are continuously differentiable up to the order  $p$ .

We denote by  $\mathcal{C}$  the set of the  $[0, \infty) \rightarrow \mathbb{R}^d$  continuous functions. It is well-known that the space  $\mathcal{C}$  endowed with the topology of the uniform convergence on the compact intervals of  $[0, \infty)$  is a Polish space.

#### 2.1.2 Random variables

The notation  $f_{\#}\mu$  stands for the pushforward of the measure  $\mu$  by the map  $f$ , that is,  $f_{\#}\mu = \mu \circ f^{-1}$ .

For  $t \geq 0$ , we define the projections  $\pi_t$  and  $\pi_{[0,t]}$  as  $\pi_t : (\mathbb{R}^d)^{[0,\infty)} \rightarrow \mathbb{R}^d, x \mapsto x_t$  and  $\pi_{[0,t]} : (\mathbb{R}^d)^{[0,\infty)} \rightarrow (\mathbb{R}^d)^{[0,t]}, x \mapsto (x_u : u \in [0, t])$ .

Let  $p \geq 1$ . For  $\rho \in \mathcal{P}_p(\mathcal{C})$ , we denote

$$\rho_t := (\pi_t)_\# \rho.$$

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. We say that a collection  $A$  of random variables on  $\Omega \rightarrow E$  is *tight* in  $E$ , if the family  $\{X_\# \mathbb{P} : X \in A\}$  is weak $\star$ -relatively compact in  $\mathcal{P}(E)$  *i.e.*, has a weak $\star$  compact closure in  $\mathcal{P}(E)$ .

We say that a  $n$ -uple of random variables  $(X_1, \dots, X_n)$  is *exchangeable*, if its distribution is invariant by any permutation on  $[n]$ .

Let  $\mathbb{T}$  represent either  $\mathbb{N}$  or  $[0, +\infty)$ . Let  $(U_t^n : t \in \mathbb{T}, n \in \mathbb{N})$  be a collection of random variables on a metric space  $(E, d)$ . We say that  $(U_t^n)$  converges in probability to  $U$  as  $(t, n) \rightarrow (\infty, \infty)$  if, for every  $\epsilon > 0$ , the net  $(\mathbb{P}(d(U_t^n, U) > \epsilon) : t \in \mathbb{T}, n \in \mathbb{N})$  converges to zero as  $t$  and  $n$  both converge to  $\infty$ . We denote this by  $U_t^n \xrightarrow[(t,n) \rightarrow (\infty, \infty)]{\mathbb{P}} U$ . Moreover, assuming that the collection of random variables  $(U_t^n : t \in \mathbb{T}, n \in \mathbb{N})$  are real valued, we say that the latter collection is *uniformly integrable* if:

$$\lim_{a \rightarrow \infty} \sup_{t \in \mathbb{T}, n \in \mathbb{N}^*} \mathbb{E} [|U_t^n| \mathbb{1}_{|U_t^n| > a}] = 0.$$

Finally, for any  $d \in \mathbb{N}^*$ ,  $\mathcal{L}^d$  stands for the Lebesgue measure on  $\mathbb{R}^d$ .

## 2.2 Spaces of probability measures

Let  $(E, d)$  denote a Polish space. If  $\mathcal{A} \subset E$  is a subset, we define  $d(x, \mathcal{A}) := \inf\{d(x, y) : y \in \mathcal{A}\}$ , with  $\inf \emptyset = \infty$ . We say that a net  $(\mu_\alpha)$  converges to  $\mathcal{A}$  if  $d(x_\alpha, \mathcal{A}) \rightarrow_\alpha 0$ .

We denote by  $\mathcal{P}(E)$  the set of probability measures on the Borel  $\sigma$ -algebra  $\mathcal{B}(E)$ . We equip  $\mathcal{P}(E)$  with the weak $\star$  topology. Note that  $\mathcal{P}(E)$  is a Polish space. We denote by  $d_L$  the Levy-Prokhorov distance on  $\mathcal{P}(E)$ , which is compatible with the weak $\star$  topology. We define the *intensity* of a random variable  $\rho : \Omega \rightarrow \mathcal{P}(E)$ , as the measure  $\mathbb{I}(\rho) \in \mathcal{P}(E)$  that satisfies

$$\forall A \in \mathcal{F}, \quad \mathbb{I}(\rho)(A) := \mathbb{E}(\rho(A)).$$

**Lemma 1** ([MRC87]). *A sequence  $(\rho_n)$  of random variables on  $\mathcal{P}(E)$  is tight if and only if the sequence  $(\mathbb{I}(\rho_n))$  is weak $\star$ -relatively compact.*

Let  $p \geq 1$ . If  $E$  is a Banach space, we define

$$\mathcal{P}_p(E) := \{\mu \in \mathcal{P}(E) : \int \|x\|^p d\mu(x) < \infty\}.$$

We define the Wasserstein distance of order  $p$  on  $\mathcal{P}_p(E)$  by

$$W_p(\mu, \nu) := \left( \inf_{\varsigma \in \Pi(\mu, \nu)} \int \|x - y\|^p d\varsigma(x, y) \right)^{1/p}, \quad (7)$$

where  $\Pi(\mu, \nu)$  is the set of measures  $\varsigma \in \mathcal{P}(E \times E)$ , such that  $\varsigma(\cdot \times E) = \mu$  and  $\varsigma(E \times \cdot) = \nu$ . We denote by  $\Pi_p^0(\mu, \nu)$  the set of optimal transport plans *i.e.*, the set of measures  $\varsigma \in \Pi(\mu, \nu)$  achieving the infimum in Eq. (7). The set  $\mathcal{P}_p(E)$  is endowed with the distance  $W_p$ . Define:

$$\mathcal{P}_p(\mathcal{C}) := \{\rho \in \mathcal{P}(\mathcal{C}) : \forall T > 0, \int \sup_{t \in [0, T]} \|x_t\|^p d\rho(x) < \infty\}.$$

For every  $\rho, \rho' \in \mathcal{P}_p(\mathcal{C})$ , we define:

$$W_p(\rho, \rho') := \sum_{n=1}^{\infty} 2^{-n} (1 \wedge W_p((\pi_{[0,n]})_\# \rho, (\pi_{[0,n]})_\# \rho')).$$

We equip  $\mathcal{P}_p(\mathcal{C})$  with the distance  $W_p$ . We say that a subset  $\mathcal{A} \subset \mathcal{P}_p(\mathcal{C})$  has *uniformly integrable  $p$ -moments* if the following condition holds:

$$\forall T > 0, \quad \lim_{a \rightarrow \infty} \sup_{\rho \in \mathcal{A}} \int \mathbb{1}_{\sup_{t \in [0, T]} \|x_t\| > a} \left( \sup_{t \in [0, T]} \|x_t\|^p \right) d\rho(x) = 0. \quad (p\text{-UI})$$

In the same way, a sequence  $(\rho_n)$  has uniformly integrable  $p$ -moments if the condition  $(p\text{-UI})$  holds for the sequence  $(\rho_n)$  in place of  $\mathcal{A}$ . Following the same lines as [Vil09, Th. 6.18] and [AGS08, Prop. 7.1.5], we obtain the following lemma. The proof is provided in Appendix A.1.

**Proposition 1.** *i) The space  $\mathcal{P}_p(\mathcal{C})$  is Polish.*

*ii) A subset  $\mathcal{A} \subset \mathcal{P}_p(\mathcal{C})$  is relatively compact if and only if, it is weak\*-relatively compact in  $\mathcal{P}(\mathcal{C})$ , and if  $\mathcal{A}$  has uniformly integrable  $p$ -moments.*

Finally, we will also consider  $\mathcal{P}_p(\mathcal{C})$ -valued sequences of random variables. Therefore, the following extension of Lemma 1, will be useful. It is established in Appendix A.2.

**Lemma 2.** *Let  $(\rho_n)$  be a sequence of random variables valued in  $\mathcal{P}_p(\mathcal{C})$ . Assume that  $(\mathbb{I}(\rho_n))$  is relatively compact in  $\mathcal{P}_p(\mathcal{C})$ . Then,  $(\rho_n)$  is tight in  $\mathcal{P}_p(\mathcal{C})$ .*

## 2.3 Spaces of McKean-Vlasov measures

Let  $d' \in \mathbb{N}^*$ , consider a matrix-valued function  $\sigma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d'}$  and a vector field  $b : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  satisfying the following assumptions:

**Assumption 1.** *The vector field  $b : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ , and  $\sigma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d'}$  are continuous. Moreover, there exists  $C > 0$  such that for all  $x, y \in \mathbb{R}^d$ ,*

$$\|b(x, y)\| \leq C(1 + \|x\| + \|y\|), \text{ and } \|\sigma(x, y)\| \leq C.$$

For every  $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ , we define  $b(x, \mu) := \int b(x, y) d\mu(y)$ , and  $\sigma(x, \mu) = \int \sigma(x, y) d\mu(y)$ , with a slight abuse of notations. We define  $L(\mu)$  which, to every test function  $\phi \in C_c^2(\mathbb{R}^d, \mathbb{R})$ , associates the function  $L(\mu)(\phi)$  given by

$$L(\mu)(\phi)(x) = \langle b(x, \mu), \nabla \phi(x) \rangle + \text{tr}(\sigma(x, \mu)^T H_\phi(x) \sigma(x, \mu)), \quad (8)$$

where  $H_\phi$  is the hessian matrix of  $\phi$ . Let  $(X_t : t \in [0, \infty))$  be the canonical process on  $\mathcal{C}$ . Denote by  $(\mathcal{F}_t^X)_{t \geq 0}$  the natural filtration (i.e., the filtration generated by  $\{X_s : 0 \leq s \leq t\}$ ).

**Definition 1.** *Let  $p \geq 1$ . We say that a measure  $\rho \in \mathcal{P}_p(\mathcal{C})$  belongs to the class  $\mathbf{V}_p$  if, for every  $\phi \in C_c^2(\mathbb{R}^d, \mathbb{R})$ ,*

$$\phi(X_t) - \int_0^t L(\rho_s)(\phi)(X_s) ds$$

*is a  $(\mathcal{F}_t^X)_{t \geq 0}$ -martingale on the probability space  $(\mathcal{C}, \mathcal{B}(\mathcal{C}), \rho)$ .*

In the sequel, it will be convenient to work with the following equivalent characterization. The martingale property implies that every measure  $\rho \in \mathbf{V}_p$  satisfies  $G(\rho) = 0$ , for every function  $G : \mathcal{P}_p(\mathcal{C}) \rightarrow \mathbb{R}$  of the form:

$$G(\rho) := \int \left( \phi(x_t) - \phi(x_s) - \int_s^t L(\rho_u)(\phi)(x_u) du \right) \prod_{j=1}^r h_j(x_{v_j}) d\rho(x), \quad (9)$$

where  $r \in \mathbb{N}$ ,  $\phi \in C_c^2(\mathbb{R}^d, \mathbb{R})$ ,  $h_1, \dots, h_r \in C_c(\mathbb{R}^d, \mathbb{R})^r$ ,  $0 \leq v_1 \leq \dots \leq v_r \leq s \leq t$ , are arbitrary. We denote by  $\mathcal{G}_p$  the set of such mappings  $G$ . Assumption 1 ensures that these mappings are well defined. By Def. 1, every  $\rho \in \mathbf{V}_p$  is a root of all  $G \in \mathcal{G}_p$ . As a matter of fact, a measure  $\rho \in \mathcal{P}_p(\mathcal{C})$  belongs to the set  $\mathbf{V}_p$ , if and only if  $G(\rho) = 0$  for every  $G$  of the form (9). In other words, Def. 1 is equivalent to the following identity:

$$\mathbf{V}_p = \bigcap_{G \in \mathcal{G}_p} G^{-1}(\{0\}). \quad (10)$$

The following lemma is proved in Appendix A.3.

**Lemma 3.** *Let Assumption 1 hold true. Every  $G \in \mathcal{G}_p$  is a continuous function on  $\mathcal{P}_p(\mathcal{C}) \rightarrow \mathbb{R}$ .*

The following result is a consequence of Lemma 3 and Prop. 1.

**Proposition 2.** *Under Assumption 1,  $\mathbf{V}_p$  is a closed subset of  $\mathcal{P}_p(\mathcal{C})$ . Moreover, equipped with the trace topology of  $\mathcal{P}_p(\mathcal{C})$ ,  $\mathbf{V}_p$  is a Polish space.*

*Proof.* For all  $\rho_n \in \mathbf{V}_p \rightarrow \rho_\infty$  in  $\mathcal{P}_p(\mathcal{C})$ , it holds by Lemma 3 that  $G(\rho_\infty) = 0$  for all  $G \in \mathcal{G}_p$ , which shows that  $\rho_\infty \in \mathbf{V}_p$  by (10). Hence,  $\mathbf{V}_p$  is closed. A closed subset of a Polish space is also Polish. By Prop. 1,  $\mathbf{V}_p$  is Polish.  $\square$

## 2.4 Dynamical systems

Recall the definition of the shift  $\Theta_t(x) = x_{t+}$ , defined on  $\mathcal{C}$ . Let us equip the space  $V_p$  assumed nonempty with the trace topology of  $\mathcal{P}_p(\mathcal{C})$ , making it a Polish space (see Prop. 2). With this at hand, one can readily check that the function  $\Phi : [0, \infty) \times V_p \rightarrow V_p$  defined as  $(t, \rho) \mapsto \Phi_t(\rho) = (\Theta_t)_\# \rho$  is a semi-flow on the space  $(V_p, W_p)$ , in the sense that  $\Phi$  is continuous,  $\Phi_0(\cdot)$  coincides with the identity, and  $\Phi_{t+s} = \Phi_t \circ \Phi_s$  for all  $t, s \geq 0$ , see [Ben99] for a nice exposition of the concepts related to semi-flows. The omega limit set of  $\rho \in V_p$  for this semi-flow is the set  $\omega(\rho)$  defined by:

$$\omega(\rho) := \bigcap_{t>0} \overline{\{\Phi_s(\rho) : s > t\}}.$$

Equivalently,  $\omega(\rho)$  is the set of  $W_p$ -limits of sequences of the form  $(\Phi_{t_n}(\rho))$  where  $t_n \rightarrow \infty$ . A point  $\rho \in V_p$  is called recurrent if  $\rho \in \omega(\rho)$ . The Birkhoff center  $BC_p$  is defined as the closure of the set of recurrent points:

$$BC_p := \overline{\{\rho \in V_p : \rho \in \omega(\rho)\}}.$$

Consider a non-empty set  $\Lambda \subset V_p$ .

**Definition 2.** Consider the semi-flow  $\Phi$ . A lower semi-continuous function  $F : V_p \rightarrow \mathbb{R}$  is called a Lyapunov function for the set  $\Lambda$  if, for every  $\rho \in V_p$  and every  $t > 0$ ,  $F(\Phi_t(\rho)) \leq F(\rho)$ , and  $F(\Phi_t(\rho)) < F(\rho)$  whenever  $\rho \notin \Lambda$ .

The following result is standard.

**Proposition 3.** Let  $p > 0$ . If  $F$  is a Lyapunov function for the set  $\Lambda$ , then  $BC_p \subset \overline{\Lambda}$ .

*Proof.* The limit  $\ell := \lim_{t \rightarrow \infty} F(\Phi_t(\rho))$  is well-defined because  $F(\Phi_t(\rho))$  is non increasing. Consider a recurrent point  $\rho \in V_p$ , say  $\rho = \lim_n \Phi_{t_n}(\rho)$ . Clearly  $F(\rho) \geq F(\Phi_{t_n}(\rho)) \geq \ell$ . Moreover, by lower semicontinuity of  $F$ ,  $\ell = \lim_n F(\Phi_{t_n}(\rho)) \geq F(\rho)$ . Therefore,  $\ell$  is finite, and  $F(\rho) = \ell$ . This implies that  $t \mapsto F(\Phi_t(\rho))$  is constant. By definition, this in turn implies  $\rho \in \Lambda$ , which concludes the proof.  $\square$

## 3 Main results

### 3.1 Interpolated process and weak\* limits

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. Let  $d > 0$  be an integer. For each  $n \in \mathbb{N}^*$ , consider the random sequence (1) starting with the  $n$ -uple  $(X_0^{1,n}, \dots, X_0^{n,n})$  of random variables  $X_0^{i,n} \in \mathbb{R}^d$ , with  $((\xi_k^{i,n})_{i \in [n]})_{k \in \mathbb{N}^*}$  and  $((\zeta_k^{i,n})_{i \in [n]})_{k \in \mathbb{N}^*}$  being  $\mathbb{R}^{d \times n}$ -valued random noise sequences. For each of integer  $n > 0$ , define the filtration  $(\mathcal{F}_k^n)_{k \in \mathbb{N}}$  as in Eq. (2) or, more generally, as any filtration such that the following random variables

$$(X_0^{i,n})_{i \in [n]}, ((\xi_\ell^{i,n})_{i \in [n]})_{\ell \leq k}, ((\zeta_\ell^{i,n})_{i \in [n]})_{\ell \leq k}$$

belong to  $\mathcal{F}_k^n$ . Consider the following assumptions:

**Assumption 2.** The sequence  $(\gamma_k)$  is a non-negative deterministic sequence satisfying

$$\lim_{k \rightarrow \infty} \gamma_k = 0, \text{ and } \sum_k \gamma_k = +\infty.$$

**Assumption 3.** The following holds true.

i) For each  $n$ , the  $n$  triplets  $((X_0^{i,n}, (\zeta_k^{i,n})_{k \in \mathbb{N}}, (\xi_k^{i,n})_{k \in \mathbb{N}}))_{i \in [n]}$  is exchangeable as a  $n$ -uple of  $\mathbb{R}^d \times (\mathbb{R}^d)^\mathbb{N} \times (\mathbb{R}^d)^\mathbb{N}$ -valued random variables.

ii) It holds that  $\sup_{k,n} \mathbb{E} \|\xi_k^{1,n}\|^4 < \infty$ . Furthermore, for each  $n > 0$ , and each  $i, j$ ,

$$\mathbb{E} \left[ \xi_{k+1}^{1,n} \middle| \mathcal{F}_k^n \right] = 0 \text{ and}$$

$$\mathbb{E} \left[ \xi_{k+1}^{i,n} \left( \xi_{k+1}^{j,n} \right)^T \middle| \mathcal{F}_k^n \right] = \frac{1}{n^2} \sum_{q, \ell \in [n]} \sigma(X_k^{i,n}, X_k^{q,n}) \sigma(X_k^{j,n}, X_k^{\ell,n})^T \mathbf{1}_{i=j},$$

for some bounded continuous function  $\sigma : (x, y) \in (\mathbb{R}^d)^2 \mapsto \sigma(x, y) \in \mathbb{R}^{d \times d'}$ , given  $d' \in \mathbb{N}$ .

iii) For each  $k$ , and each  $n$ , it holds that  $\mathbb{E}\|\zeta_k^{1,n}\| < \infty$ , and

$$\lim_{(k,n) \rightarrow (\infty, \infty)} \mathbb{E} \left\| \mathbb{E} \left[ \zeta_{k+1}^{1,n} \mid \mathcal{F}_k^n \right] \right\| = 0.$$

**Remark 1.** Assumption 3-(i) holds under the stronger assumption that the  $n$ -uple  $(X_0^{i,n})_{i \in [n]}$  is exchangeable,  $(\xi_k^{i,n})_{i \in [n], k \in \mathbb{N}}$  is an i.i.d. sequence independent of  $(X_0^{i,n})_{i \in [n]}$ , and  $\zeta_k^{1,n} = 0$  for every  $k$ .

**Assumption 4.** We assume either:

$$i) \sup_{k,n} \mathbb{E}[\|X_k^{1,n}\|^2 + \|\zeta_k^{1,n}\|^2] < \infty,$$

or the stronger condition:

ii) The collections of r.v.  $(\|X_k^{1,n}\|^2 : k \in \mathbb{N}, n \in \mathbb{N}^*)$ , and  $(\|\zeta_k^{1,n}\|^2 : k \in \mathbb{N}, n \in \mathbb{N}^*)$  are uniformly integrable.

Recalling the definitions of the interpolated processes  $\bar{X}^{i,n}$  in (5), and the definition of the occupation measure  $m^n$  in (6), we shall consider the *shifted* occupation measure

$$\Phi_t(m^n) = \frac{1}{n} \sum_{i=1}^n \delta_{\Theta_t(\bar{X}^{i,n})},$$

for each  $n \in \mathbb{N}^*$  and each  $t \in (0, +\infty)$ . Note that  $\Phi_t(m^n)$  is a r.v. on  $\mathcal{P}_p(\mathcal{C})$ . We refer to the set

$$\mathcal{M} := \lim_{(t,n) \rightarrow (\infty, \infty)}^{\text{acc}} (\{(\Phi_t(m^n))_{\#} \mathbb{P}\}) \quad (11)$$

as the set of accumulation points of the probability distributions of  $\Phi_t(m^n)$  as  $(t, n) \rightarrow (\infty, \infty)$ . In other words,  $\mathcal{M}$  is the set of measures  $M \in \mathcal{P}(\mathcal{P}_p(\mathcal{C}))$  for which there is a sequence  $(t_n, \varphi_n)_n$  on  $(0, \infty) \times \mathbb{N}^*$ , such that  $t_n \rightarrow_n \infty$ ,  $\varphi_n \rightarrow_n \infty$ , and  $(\Phi_{t_n}(m^{\varphi_n}))$  converges in distribution to  $M$ .

The following two results show that the collection  $(\Phi_t(m^n))$  of random variables is tight (proven in Section 5.1), and that their limits in distribution are supported by the set of McKean-Vlasov distributions (proven in Section 5.2):

**Proposition 4.** Let  $1 \leq p < 2$ , and let Assumptions 1, 2, 3, and 4-(i) hold true. Then, the collection of shifted occupation measures  $\{\Phi_t(m^n) : t \geq 0, n \in \mathbb{N}^*\}$  is tight in  $\mathcal{P}_p(\mathcal{C})$ . If Assumption 4-(ii) additionally holds, the result remains valid when  $p = 2$ .

**Proposition 5.** Let  $1 \leq p < 2$ , and let Assumptions 1, 2, 3, and 4-(i) hold true. Then,  $\mathbf{V}_p$  is a nonempty closed set, and for every  $M \in \mathcal{M}$ , it holds that  $M(\mathbf{V}_p) = 1$ . If Assumption 4-(ii) additionally holds, the result remains valid when  $p = 2$ .

### 3.2 Ergodic convergence

We provide the proof of the following theorem in Section 5.3.

**Theorem 1.** Let  $1 \leq p < 2$ , and let Assumptions 1, 2, 3, and 4-(i) hold true. Then,

$$\frac{1}{t} \int_0^t \mathbb{W}_p(\Phi_s(m^n), \text{BC}_p) ds \xrightarrow[(t,n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0.$$

If Assumption 4-(ii) additionally holds, the result remains valid when  $p = 2$ .

Recall the definition  $\mu_k^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_k^{i,n}}$ . The proof of the following corollary is provided in Section 5.4.

**Corollary 1.** Let  $1 \leq p < 2$ , and let Assumptions 1, 2, 3, and 4-(i) hold true. Then,

$$\frac{\sum_{l=1}^k \gamma_l \mathbb{W}_p(\mu_l^n, (\pi_0)_{\#}(\text{BC}_p))}{\sum_{l=1}^k \gamma_l} \xrightarrow[(k,n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0,$$

and

$$\frac{\sum_{l=1}^k \gamma_l \mathbb{W}_p(\mu_l^n, (\pi_0)_{\#}(\text{BC}_p))^p}{\sum_{l=1}^k \gamma_l} \xrightarrow[(k,n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0.$$

If Assumption 4-(ii) additionally holds, the statements remain valid also when  $p = 2$ .

### 3.3 Pointwise convergence to a global attractor

Depending on the vector field  $b$ , it is often the case that each measure  $\rho \in \mathbf{V}_p$  is uniquely determined by its value  $\rho_0 = (\pi_0)_\# \rho \in \mathcal{P}_p(\mathbb{R}^d)$  in the sense that there exists a semi-flow  $\Psi : [0, \infty) \times \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathcal{P}_p(\mathbb{R}^d)$ ,  $(t, \nu) \mapsto \Psi_t(\nu)$ , defined on  $[0, \infty) \times \mathcal{P}_p(\mathbb{R}^d)$ , and such that

$$\rho \in \mathbf{V}_p \Leftrightarrow \forall t \geq 0, \rho_t = \Psi_t(\rho_0). \quad (12)$$

We shall say that in this situation, the class  $\mathbf{V}_p$  has a semi-flow structure on  $\mathcal{P}_p(\mathbb{R}^d)$ .

The granular media model detailed in Section 4 below is a typical example where such a situation occurs.

In this section, we are interested in the behavior of the measures  $\mu_k^n$  as  $(k, n) \rightarrow (\infty, \infty)$ , termed the “pointwise” convergence of these measures, when the semi-flow  $\Psi$  has a global attractor. We recall here that a set  $A_p \subset \mathcal{P}_p(\mathbb{R}^d)$  is said invariant for the semi-flow  $\Psi$  if  $\Psi_t(A_p) = A_p$  for all  $t \geq 0$ ; A nonempty compact invariant set  $A_p \subset \mathcal{P}_p(\mathbb{R}^d)$  is a global attractor for the semi-flow  $\Psi$  if

$$\forall \nu \in \mathcal{P}_p(\mathbb{R}^d), \quad \lim_{t \rightarrow \infty} W_p(\Psi_t(\nu), A_p) = 0,$$

and furthermore, if there exists a neighborhood  $\mathcal{N}$  of  $A_p$  in  $\mathcal{P}_p(\mathbb{R}^d)$  such that this convergence is uniform on  $\mathcal{N}$ . Such a neighborhood is called a fundamental neighborhood of  $A_p$ .

The following result is proven in Section 5.5.

**Theorem 2.** *Let  $p \in [1, 2]$ , and let Assumptions 1, 2, and 3 hold true. Let Assumption 4-(i) or the stronger Assumption 4-(ii) hold true according to whether  $p < 2$  or  $p = 2$  respectively. Assume in addition that the  $\mathbf{V}_p$  has a semi-flow structure on  $\mathcal{P}_p(\mathbb{R}^d)$  as specified in (12), and that this semi-flow  $\Psi$  admits a global attractor  $A_p$ . Then,*

$$W_p(\mu_k^n, A_p) \xrightarrow[(k, n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0.$$

## 4 Granular media

The proofs of the results relative to this section are provided in Section 6.

In this paragraph, we review some properties of the set  $\mathbf{V}_2$  of McKean-Vlasov processes, in the case where  $(x, y) \mapsto \sigma(x, y) = \sigma I_d$  for some real constant  $\sigma \geq 0$  and,

$$b(x, y) := -\nabla V(x) - \nabla U(x - y), \quad (13)$$

where  $V, U : \mathbb{R}^d \rightarrow \mathbb{R}$  are two functions satisfying the following assumption.

**Assumption 5** (Granular media). *The functions  $V, U$  belong to  $C^1(\mathbb{R}^d, \mathbb{R})$ . Moreover, there exists  $\lambda, C, \beta > 0$ , such that for every  $x, y \in \mathbb{R}^d$ , the following holds:*

- i)  $\langle x, \nabla V(x) \rangle \geq \lambda \|x\|^2 - C$ ,
- ii)  $U(x) = U(-x)$ , and  $\langle x, \nabla U(x) \rangle \geq -C$ ,
- iii)  $\|\nabla V(x)\| + \|\nabla U(x)\| \leq C(1 + \|x\|)$ ,
- iv)  $\|\nabla V(x) - \nabla V(y)\| + \|\nabla U(x) - \nabla U(y)\| \leq C(\|x - y\|^\beta \vee \|x - y\|)$ .

Under Assumptions 5, the vector field  $b$  satisfies Assumption 1. We will see later, as a byproduct of Th. 3, that the set  $\mathbf{V}_2$  of McKean-Vlasov distributions associated to the field  $b$  in Eq. (13), is non empty. We say  $\mu \ll \mathcal{L}^d$  if  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  admits continuously differentiable density w.r.t. the Lebesgue measure  $\mathcal{L}^d$ , which we denote by  $d\mu/d\mathcal{L}^d$ . Define the functional  $\mathcal{H} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, \infty]$  as  $\mathcal{H}(\mu) = \mathcal{F}(\mu) + \mathcal{V}(\mu) + \mathcal{W}(\mu)$  with

$$\mathcal{F}(\mu) = \begin{cases} \int \sigma^2 \log \left( \frac{d\mu}{d\mathcal{L}^d}(x) \right) d\mu(x) & \text{if } \mu \ll \mathcal{L}^d \\ \infty & \text{otherwise,} \end{cases}$$

$$\mathcal{V}(\mu) = \int V(x) d\mu(x), \quad \text{and} \quad \mathcal{W}(\mu) = \frac{1}{2} \iint U(x - y) d\mu(x) d\mu(y).$$

The following central result provides a central properties of the elements of  $\mathbf{V}_2$ .

**Proposition 6.** *Let Assumption 5 hold true, and let  $b$  be defined by (13). Assume  $\sigma > 0$ . Consider  $\rho \in \mathbf{V}_2$ . Then, for every  $t > 0$ ,  $\rho_t$  admits a density  $x \mapsto \varrho(t, x)$  in  $C^1(\mathbb{R}^d, \mathbb{R})$  w.r.t. the Lebesgue measure. For every  $t > 0$ , the functional  $t \mapsto \mathcal{H}(\rho_t)$  is finite, and satisfies for every  $t_2 > t_1 > 0$ ,*

$$\mathcal{H}(\rho_{t_2}) - \mathcal{H}(\rho_{t_1}) = - \int_{t_1}^{t_2} \int \|v_t(x)\|^2 \varrho(t, x) dx dt, \quad (14)$$

where  $v_t$  is the vector field defined for every  $x \in \mathbb{R}^d$  by:

$$v_t(x) := -\nabla V(x) - \int \nabla U(x - y) d\rho_t(y) - \sigma^2 \nabla \log \varrho(t, x). \quad (15)$$

Define  $\mathcal{P}_2^r(\mathbb{R}^d)$  as the set of measures  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  such that  $\mu \ll \mathcal{L}^d$ . Define:

$$\mathcal{S} := \{\mu \in \mathcal{P}_2^r(\mathbb{R}^d) : \nabla V + \int \nabla U(\cdot - y) d\mu(y) + \sigma^2 \nabla \log \frac{d\mu}{d\mathcal{L}^d} = 0 \mu\text{-a.e.}\}. \quad (16)$$

Finally, for every  $\epsilon \geq 0$ , define:

$$\Lambda_\epsilon := \{\rho \in \mathbf{V}_2 : \exists \mu \in \mathcal{S}, \forall t \geq \epsilon, \rho_t = \mu\}. \quad (17)$$

**Proposition 7.** *We posit the assumptions of Prop. 6. For every  $\epsilon > 0$ , the function  $\rho \mapsto \mathcal{H}(\rho_\epsilon)$  is real valued on  $\mathbf{V}_2$ , lower semicontinuous, and is a Lyapunov function for the set  $\Lambda_\epsilon$ . Moreover,*

$$\text{BC}_2 \subset \overline{\Lambda_0}.$$

We also need to consider a setting where  $\mathbf{V}_2$  has a semi-flow structure on  $\mathcal{P}_2(\mathbb{R}^d)$  as in (12) in order to set the stage for the pointwise convergence of the measures  $\mu_k^n$  issued from our discrete algorithm. To that end, we shall appeal to the theory of the gradient flows in the space of probability measures as detailed in the treatise [AGS08] of Ambrosio, Gigli and Savaré. The following additional assumption will be needed:

**Assumption 6.** *The functions  $U$  and  $V$  satisfy the doubling condition. Namely, there exists constants  $C_U, C_V > 0$  such that*

$$U(x + y) \leq C_U (1 + U(x) + U(y)) \quad \text{and} \quad V(x + y) \leq C_V (1 + V(x) + V(y)).$$

**Proposition 8.** *Let Assumption 5 hold true with  $\beta = 1$ , and let Assumption 6 hold true. For each  $\rho \in \mathbf{V}_2$ , the curve  $t \mapsto \rho_t$  belongs to the set of absolutely continuous functions  $\text{AC}_{loc}^2((0, \infty), \mathcal{P}_2(\mathbb{R}^d))$  as defined in [AGS08, Sec. 8.3], and is completely determined by  $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$  as being the gradient flow of the functional  $\mathcal{H}$  in  $\mathcal{P}_2(\mathbb{R}^d)$ . Thus,  $\mathbf{V}_2$  has a semi-flow structure, and we write  $\rho_t = \Psi_t(\rho_0)$ .*

For completeness, we recall along [AGS08, Chap. 8 and 11] that  $t \mapsto \rho_t$  being the solution of the gradient flow of  $\mathcal{H}$  in  $\mathcal{P}_2(\mathbb{R}^d)$  stands to the existence of a Borel vector field  $w_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $w_t$  belongs to the tangent bundle  $\text{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d)$  for  $\mathcal{L}^1$ -almost all  $t > 0$ ,  $\|w_t\|_{L^2(\rho_t)} \in L_{loc}^p(0, \infty)$ , the continuity equation  $\partial_t \rho_t + \nabla \cdot (\rho_t w_t) = 0$  holds in general in the sense of distributions, and finally,  $w_t \in -\partial \mathcal{H}(\rho_t)$  for  $\mathcal{L}^1$ -almost each  $t > 0$ , where  $\partial \mathcal{H}$  is the Fréchet sub-differential as defined in [AGS08, Chap. 10], which always exists under our assumptions. Actually,  $w_t = v_t$  as given by Equation (15) for almost all  $t$ .

We now turn to our discrete algorithm. Consider the iterations:

$$X_{k+1}^{i,n} = X_k^{i,n} - \frac{\gamma_{k+1}}{n} \sum_{j \in [n]} \nabla U(X_k^{i,n} - X_k^{j,n}) - \gamma_{k+1} \nabla V(X_k^{i,n}) + \sqrt{2\gamma_{k+1}} \xi_k^{i,n}, \quad (18)$$

for each  $i \in [n]$ . This is a special case of Eq. (1) with  $b(x, y)$  given by Eq. (13) and  $\zeta_k^{i,n} = 0$  for all  $k$ . For simplicity, Assumption 3 will be replaced by the following stronger assumption:

**Assumption 7.** *The  $n$ -tuple  $(X_0^{1,n}, \dots, X_0^{n,n})$  is exchangeable and  $\sup_n \mathbb{E}(\|X_0^{1,n}\|^4) < \infty$ . Moreover,  $(\xi_k^{i,n})_{i \in [n], k \in \mathbb{N}}$  are i.i.d. centered random variables, with variance  $\sigma^2 I_d$ , and  $\mathbb{E}(\|\xi_1^{1,1}\|^4) < \infty$ .*

The next proposition implies that the condition ii) in Assumption 4 holds.

**Proposition 9.** *Let Assumptions 2, 5 and 7 be satisfied. Then,*

$$\sup_{n \in \mathbb{N}^*, k \in \mathbb{N}} \mathbb{E} \left[ \|X_k^{1,n}\|^4 \right] < \infty.$$

Putting Assumptions 2, 5 and 7 together, the hypotheses of Th. 1 are satisfied for  $p = 2$ .

**Theorem 3.** *Let Assumptions 2, 5 and 7 be satisfied. Assume  $\sigma > 0$ . Then, the set  $\mathcal{S}$  given by Eq. (16) is non empty, and furthermore,*

$$\frac{\sum_{l=1}^k \gamma_l W_2(\mu_l^n, \mathcal{S})}{\sum_{l=1}^k \gamma_l} \xrightarrow[(k,n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0.$$

*Proof.* Use Cor. 1 with  $p = 2$ , together with Prop. 7.  $\square$

We now turn to the pointwise convergence of the measures  $\mu_k^n$ .

**Theorem 4.** *Let Assumption 5 hold true with  $\beta = 1$ , and let Assumption 6 hold true. Assume that the semi-flow  $\Psi$  which existence is stated by Proposition 8 has a global attractor  $A_2$ . In the case where  $A_2$  is a singleton, it holds that  $\mathcal{S} = A_2$ . In any case,*

$$W_2(\mu_k^n, A_2) \xrightarrow[(k,n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0.$$

The classical case when  $A_2$  is reduced to a singleton is the case where the functions  $U$  and  $V$  are both strongly convex; indeed, there exists here  $\lambda > 0$  such that  $W_2(\Psi_t(\nu), \Psi_t(\nu')) \leq e^{-\lambda t} W_2(\nu, \nu')$  [AGS08, Th. 11.2.1]. A rich literature is devoted to relaxing this strong convexity assumption, see [CMV03, CMV06, CGM08, BGG13, GLWZ22] as a non exhaustive list.

**Remark 2.** *A similar proof can be applied to two-layer neural networks with one fixed layer, known as the random feature model [RR07, CRR18]. By analyzing the stochastic gradient descent algorithm for a given risk in this context, we obtain a particle system in the form of Eq. (1).*

## 5 Proofs of Section 3

### 5.1 Proof of Proposition 4

In this paragraph, consider  $1 \leq p \leq 2$ . Note that  $(\Phi_t(m^n))$  belongs to  $\mathcal{P}_p(\mathcal{C})$ . In the light of Lemma 2 and Prop 1, we should establish two points: first, the weak $\star$ -relative compactness of the family of intensities  $\{\mathbb{I}(\Phi_t(m^n))\}_{t,n}$ ; second, a uniform integrability condition of the  $p$ th order moments of the measures  $\mathbb{I}(\Phi_t(m^n)(x))$ . These results are respectively stated in Lemmas 4 and 5 below.

**Lemma 4.** *We posit the assumptions of Prop. 4. The family of intensities  $\{\mathbb{I}(\Phi_t(m^n))\}_{t,n}$  is weak $\star$ -relatively compact in  $\mathcal{P}(\mathcal{C})$ .*

*Proof.* Let us establish the first point. For every bounded continuous function  $\phi : \mathcal{C} \rightarrow \mathbb{R}$ , we have

$$\mathbb{I}(\Phi_t(m^n))(\phi) := \mathbb{E} \left[ \int \phi(x) d(\Phi_t(m^n)(x)) \right] = \frac{1}{n} \sum_{i \in [n]} \mathbb{E} \left[ \phi(\bar{X}_{t+}^{i,n}) \right] = \mathbb{E} \left[ \phi(\bar{X}_{t+}^{1,n}) \right],$$

where we used the exchangeability stated in Assumption 3-(i). Let us define the measure  $\hat{\mathbb{I}}_t^n \in \mathcal{P}(\mathbb{R}^d)$  as

$$\hat{\mathbb{I}}_t^n(\phi) := \mathbb{E} \left[ \phi(\bar{X}_t^{1,n}) \right],$$

for each measurable function  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ . According to Theorem 7.3 in [Bil99], the weak $\star$ -relative compactness of the sequence  $(\mathbb{I}_t^n)_{t,n}$  in  $\mathcal{P}(\mathcal{C})$  is guaranteed if and only if the weak $\star$ -relative compactness of  $(\hat{\mathbb{I}}_t^n)_{t,n}$  in  $\mathcal{P}(\mathbb{R}^d)$  is ensured, and if the following equicontinuity condition

$$\lim_{\delta \rightarrow 0} \limsup_{t,n} \mathbb{P} \left( w_{\bar{X}_{t+}^{1,n}}^T(\delta) \geq \varepsilon \right) = 0 \quad (19)$$

is met for every  $\varepsilon, T > 0$ , where  $w_x^T(\delta)$  is the modulus of continuity of a function  $x$  on the interval  $[0, T]$ . The weak $\star$ -relative compactness of  $(\hat{\Pi}_t^n)_{t,n}$  in  $\mathcal{P}(\mathbb{R}^d)$ , follows directly from Assumption 4. Using the notation  $k_t := \inf\{k : \sum_{i=1}^k \gamma_i \geq t\}$ , and using the definition in Eq. (1), we obtain the decomposition:

$$\begin{aligned} \bar{X}_t^{1,n} - \bar{X}_s^{1,n} &= P_{s,t}^n + N_{s,t}^n + U_{s,t}^n, \\ P_{s,t}^n &:= \frac{1}{n} \sum_{j=1}^n \left( \sum_{k=k_s}^{k_t-2} \gamma_{k+1} b(X_k^{1,n}, X_k^{j,n}) \right. \\ &\quad \left. + (\tau_{k_s} - s) b(X_{k_s-1}^{1,n}, X_{k_s-1}^{j,n}) + (\tau_{k_t} - t) b(X_{k_t-1}^{1,n}, X_{k_t-1}^{j,n}) \right) \\ N_{s,t}^n &:= \sum_{k=k_s}^{k_t-2} \sqrt{\gamma_{k+1}} \xi_{k+1}^{1,n} + \frac{\tau_{k_s} - s}{\gamma_{k_s}} \sqrt{\gamma_{k_s}} \xi_{k_s}^{1,n} + \frac{\tau_{k_t} - t}{\gamma_{k_t}} \sqrt{\gamma_{k_t}} \xi_{k_t}^{1,n} \\ U_{s,t}^n &:= \frac{1}{n} \sum_{i=1}^n \left( \sum_{k=k_s}^{k_t-2} \gamma_{k+1} \zeta_{k+1}^{i,n} + (\tau_{k_s} - s) \zeta_{k_s}^{i,n} + (\tau_{k_t} - t) \zeta_{k_t}^{i,n} \right). \end{aligned} \quad (20)$$

Let the sequence  $(\tilde{\gamma}_k)$  be defined by:  $\tilde{\gamma}_{k_s} := \tau_{k_s} - s$ ,  $\tilde{\gamma}_{k_t} := \tau_{k_t} - t$  and  $\tilde{\gamma}_k := \gamma_k$  for all  $k \neq k_{t_s}, k_{t_t}$ . Note that:

$$\sum_{k=k_s-1}^{k_t-1} \tilde{\gamma}_{k+1} = t - s. \quad (21)$$

Moreover, we have:

$$\frac{\tau_{k_s} - s}{\gamma_{k_s}} \sqrt{\gamma_{k_s}} \leq \sqrt{\tilde{\gamma}_{k_s}}, \quad \text{and} \quad \frac{\tau_{k_t} - t}{\gamma_{k_t}} \sqrt{\gamma_{k_t}} \leq \sqrt{\tilde{\gamma}_{k_t}}. \quad (22)$$

The term  $N_{s,t}^n$  is expressed as a sum of martingale increments, with respect to the filtration  $\mathcal{F}_k^n$ . Let  $\|\cdot\|_\alpha$  denote the  $\alpha$ -norm in  $\mathbb{R}^d$ . We apply Burkholder's inequality stated in [BDG72, Th. 1.1] to the components of the vector  $N_{s,t}^n$  in  $\mathbb{R}^d$ . As Eq. (21) and (22) hold:

$$\mathbb{E} \left( \|N_{s,t}^n\|_4^4 \right) \leq C(t-s) \mathbb{E} \left[ \sum_{k=k_s-1}^{k_t-1} \tilde{\gamma}_{k+1} \left\| \xi_{k+1}^{1,n} \right\|_4^4 \right],$$

where  $C$  is a constant independent  $s, t$  and  $n$ . As Assumption 3-(ii) holds, there exists a constant  $C > 0$  independent of  $s, t$ , and  $n$ , such that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left( \|N_{s,t}^n\|^4 \right) \leq C(t-s)^2. \quad (23)$$

Furthermore, using Jensen's inequality along with Eq. (21), we obtain

$$\|P_{s,t}^n\|^2 \leq \frac{(t-s)}{n} \sum_{j \in [n]} \sum_{k=k_s-1}^{k_t-1} \tilde{\gamma}_{k+1} \left\| b(X_k^{1,n}, X_k^{j,n}) \right\|^2.$$

Using Assumptions 1 and 3, there exists a constant  $C$ , independent of  $s, t, n$ , such that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left( \|P_{s,t}^n\|^2 \right) \leq C(t-s)^2. \quad (24)$$

Also, by Jensen's inequality, we have

$$\|U_{s,t}^n\|^2 \leq \frac{(t-s)}{n} \sum_{i \in [n]} \sum_{k=k_s-1}^{k_t-1} \tilde{\gamma}_{k+1} \left\| \zeta_{k+1}^{i,n} \right\|^2.$$

Since, by Assumption 3, we have  $\sup_{k,n} \mathbb{E}[\|\zeta_k^{1,n}\|^2] < \infty$ , there exists a constant  $C$  independent of  $n, s$ , and  $t$ , such that:

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left( \|U_{s,t}^n\|^2 \right) \leq C(t-s)^2. \quad (25)$$

Combining Equations (24), (23) and (25), we have shown:

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left[ \|P_{s,t}^n\|^2 + \|N_{s,t}^n\|^4 + \|U_{s,t}^n\|^2 \right] \leq C(t-s)^2, \quad (26)$$

where  $0 \leq s < t < \infty$ , and  $C$  is a positive constant, independent of  $s, t, n$ . Using [Leo23, Th. 2.8] and Markov's inequality, Eq. (19) hold.  $\square$

**Lemma 5.** *We posit the assumptions of Prop. 4. For every  $T > 0$ ,*

$$\lim_{a \rightarrow \infty} \sup_{t \in \mathbb{R}_+, n \in \mathbb{N}^*} \mathbb{E} \left[ \int \sup_{s \in [0, T]} \|x_s\|^p \mathbf{1}_{\sup_{s \in [0, T]} \|x_s\| \geq a} d\Phi_t(m^n)(x) \right] = 0.$$

*Proof.* By the exchangeability stated in Assumption 3-(i), we obtain:

$$\mathbb{E} \left[ \int \sup_{u \in [0, T]} \|x_u\|^p \mathbf{1}_{\sup_{u \in [0, T]} \|x_u\| > a} d\Phi_t(m^n)(x) \right] = \mathbb{E} \left[ \sup_{u \in [0, T]} \left\| \bar{X}_{t+u}^{1,n} \right\|^p \mathbf{1}_{\sup_{u \in [0, T]} \left\| \bar{X}_{t+u}^{1,n} \right\| > a} \right],$$

for every  $k, t, n$ . Recalling the decomposition introduced in Eq. (20), for every  $u \in [0, T]$ :

$$\left\| \bar{X}_{t+u}^{1,n} \right\|^p \leq 4^{p-1} \left( \left\| \bar{X}_t^{1,n} \right\|^p + \|N_{t,t+u}^n\|^p + \|P_{t,t+u}^n\|^p + \|U_{t,t+u}^n\|^p \right).$$

Hence,

$$\begin{aligned} \left\| \bar{X}_{t+u}^{1,n} \right\|^p \mathbf{1}_{\sup_{u \in [0, T]} \left\| \bar{X}_{t+u}^{1,n} \right\| > a} &\leq 4^p \left( \left\| \bar{X}_t^{1,n} \right\|^p \mathbf{1}_{\left\| \bar{X}_t^{1,n} \right\| > \frac{a}{4}} \right. \\ &\quad + \left\| N_{t,t+u}^n \right\|^p \mathbf{1}_{\sup_{u \in [0, T]} \|N_{t,t+u}^n\| > \frac{a}{4}} \\ &\quad \left. + \left\| P_{t,t+u}^n \right\|^p \mathbf{1}_{\sup_{u \in [0, T]} \|P_{t,t+u}^n\| > \frac{a}{4}} + \left\| U_{t,t+u}^n \right\|^p \mathbf{1}_{\sup_{u \in [0, T]} \|U_{t,t+u}^n\| > \frac{a}{4}} \right). \end{aligned}$$

Consequently, for each  $T > 0$ , it suffices to obtain the uniform integrability of the four collections of random variables:  $(\|\bar{X}_t^{1,n}\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ ,  $(\sup_{u \in [0, T]} \|N_{t,t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ ,  $(\sup_{u \in [0, T]} \|P_{t,t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$  and  $(\sup_{u \in [0, T]} \|U_{t,t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ .

$(\|\bar{X}_t^{1,n}\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$  is uniformly integrable by Assumption 4-(i) when  $p < 2$ , and by Assumption 4-(ii) when  $p = 2$ . As obtained in Eq. (23), Burkholder inequality stated in [BDG72, Th 1.1] yields:

$$\mathbb{E} \left[ \sup_{u \in [0, T]} \|N_{t,t+u}^n\|^4 \right] \leq CT^2,$$

where  $C$  is a constant independent of  $t, n$ , and  $T$ . Hence, since  $p < 4$ , we obtain the uniform integrability of  $\{\sup_{u \in [0, T]} \|N_{t,t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*\}$ . As obtained in Eq. (24) and Eq. (25), we derive:

$$\sup_{u \in [0, T]} \|P_{t,t+u}^n\|^p \leq \frac{CT^{p-1}}{n} \sum_{j \in [n]} \sum_{k=k_t-1}^{k_{t+T}-1} \tilde{\gamma}_{k+1} \left( 1 + \|X_k^{j,n}\|^p + \|X_k^{1,n}\|^p \right),$$

and

$$\sup_{u \in [0, T]} \|U_{t,t+u}^n\|^2 \leq \frac{CT}{n} \sum_{j \in [n]} \sum_{k=k_t-1}^{k_{t+T}-1} \tilde{\gamma}_{k+1} \left\| \zeta_k^{j,n} \right\|^2,$$

where  $C$  remains a constant independent of  $n$  and  $t$ . Using Assumption 4-(i) when  $p < 2$ , and Assumption 4-(ii) when  $p = 2$ , by de la Vallée Poussin theorem, there exists a non-decreasing, convex, and non-negative function  $F : \mathbb{R}_+^* \rightarrow \mathbb{R}$  such that

$$\lim_{h \rightarrow \infty} \frac{F(h)}{h} = \infty, \text{ and } \sup_{k \in \mathbb{N}, n \in \mathbb{N}^*} \mathbb{E} \left[ F \left( \|X_k^{1,n}\|^p \right) \right] < \infty.$$

Hence, by Jensen's inequality, and the exchangeability stated in Assumption 3,

$$\mathbb{E} \left[ F \left( \sup_{u \in [0, T]} \|P_{t, t+u}^n\|^p \right) \right] \leq \frac{1}{T} \sum_{k=k_t-1}^{k_t+T-1} \tilde{\gamma}_{k+1} \mathbb{E} \left[ F \left( CT^p \left( 1 + \|X_k^{1,n}\| \right)^p \right) \right].$$

Consequently,

$$\begin{aligned} \sup_{t \in \mathbb{R}_+, n \in \mathbb{N}^*} \mathbb{E} \left[ F \left( \sup_{u \in [0, T]} \|P_{t, t+u}^n\|^p \right) \right] &\leq \sup_{k \in \mathbb{N}, n \in \mathbb{N}^*} \mathbb{E} \left[ F \left( CT^p \left( 1 + \|X_k^{1,n}\| \right)^p \right) \right] \\ &< \infty. \end{aligned}$$

Therefore, de la Vallée Poussin theorem yields the uniform integrability of the collection  $(\sup_{u \in [0, T]} \|P_{t, t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ . In the same way, we obtain the uniform integrability of the collection  $(\sup_{u \in [0, T]} \|U_{t, t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ .  $\square$

To conclude the proof of Prop. 4, it is sufficient to remark that the tightness conditions provided in Lemma 2 are satisfied, thanks to Lemmas 4 and 5, with Prop. 1.

## 5.2 Proof of Proposition 5

The core of the proof is provided by the following proposition.

**Proposition 10.** *Let Assumptions 1, 2, 3 and 4-(i) hold,*

$$\lim_{(t, n) \rightarrow (\infty, \infty)} \mathbb{E} |G(\Phi_t(m^n))| = 0,$$

for each function  $G \in \mathcal{G}_p$ .

*Proof.* We need to show that for each  $\mathbb{R}_+ \times \mathbb{N}$ -valued sequence  $(t_n, \varphi_n) \rightarrow (\infty, \infty)$  as  $n \rightarrow \infty$ , the convergence  $\mathbb{E} |G(\Phi_{t_n}(m^{\varphi_n}))| \rightarrow 0$  holds true, where  $G = G_{r, \phi, h_1, \dots, h_r, t, s, v_1, \dots, v_r}$  has the form of Eq. (9), with  $0 \leq v_1 \leq \dots \leq v_r \leq s \leq t$ . We take  $\varphi_n = n$  for notational simplicity, and we write  $\mathbf{m}_n := \Phi_{t_n}(m^n) \in \mathcal{P}_p(\mathcal{C})$ . We have

$$G(\mathbf{m}_n) = \frac{1}{n} \sum_{i \in [n]} \left( \phi(\bar{X}_{t_n+t}^{i,n}) - \phi(\bar{X}_{t_n+s}^{i,n}) - \int_{t_n+s}^{t_n+t} \frac{1}{n^2} \sum_{j, \ell \in [n]} \psi(\bar{X}_u^{i,n}, \bar{X}_u^{j,n}, \bar{X}_u^{\ell,n}) du \right) Q^{i,n}, \quad (27)$$

where we set  $\psi(x, y, y') := \langle \nabla \phi(x), b(x, y) \rangle + \text{tr}(\sigma(x, y)^T H_\phi(x) \sigma(x, y'))$ , and

$$Q^{i,n} := \prod_{j=1}^r h_j(\bar{X}_{t_n+v_j}^{i,n}).$$

We note right away that  $|Q^{i,n}| \leq C$  where  $C$  depends on the functions  $h_j$  only, and furthermore, the random variables  $\{Q^{i,n}\}_{i \in [n]}$  are  $\mathcal{F}_{k_{t_n}+s}^n$ -measurable, where we recall that the integer  $k_t$  is defined by  $k_t := \inf\{k : \sum_{i=1}^k \gamma_i \geq t\}$ .

In the remainder, we suppress the superscript  $(n)$  from most of our notations for clarity. To deal with the right hand side of (27), we begin by expressing  $\phi(\bar{X}_{t_n+t}^i) - \phi(\bar{X}_{t_n+s}^i)$  as a telescoping sum in the discrete random variables  $X_k^i$ :

$$\begin{aligned} \phi(\bar{X}_{t_n+t}^i) - \phi(\bar{X}_{t_n+s}^i) &= \sum_{k=k_{t_n}+s}^{k_{t_n}+t-2} (\phi(X_{k+1}^i) - \phi(X_k^i)) \\ &\quad + \phi(\bar{X}_{t_n+t}^i) - \phi(X_{k_{t_n}+t-1}^i) + \phi(X_{k_{t_n}+s}^i) - \phi(\bar{X}_{t_n+s}^i). \end{aligned}$$

The summands at the r.h.s. of this expression can be decomposed as follows. Remember the form (1) of our algorithm. Denoting as  $H_\phi$  the Hessian matrix of  $\phi$ , by the Taylor-Lagrange

formula, there exists  $\theta_{k+1} \in [\tau_k, \tau_{k+1}]$  such that

$$\begin{aligned}
& \phi(X_{k+1}^i) - \phi(X_k^i) \\
&= \langle \nabla \phi(X_k^i), X_{k+1}^i - X_k^i \rangle + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\
&= \gamma_{k+1} \frac{1}{n} \sum_{j \in [n]} \langle \nabla \phi(X_k^{i,n}), b(X_k^i, X_k^j) \rangle \\
&+ \frac{\gamma_{k+1}}{n^2} \sum_{j, \ell \in [n]} \text{tr} \left( \sigma(X_k^i, X_k^j)^T H_\phi(X_k^i) \sigma(X_k^i, X_k^\ell) \right) \\
&+ \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\
&- \frac{\gamma_{k+1}}{n^2} \sum_{j, \ell \in [n]} \text{tr} \left( \sigma(X_k^i, X_k^j)^T H_\phi(X_k^i) \sigma(X_k^i, X_k^\ell) \right) + \gamma_{k+1} \langle \nabla \phi(X_k^i), \zeta_{k+1}^i \rangle \\
&= \frac{\gamma_{k+1}}{n^2} \sum_{j, \ell \in [n]} \psi(X_k^i, X_k^j, X_k^\ell) + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\
&+ \gamma_{k+1} \langle \nabla \phi(X_k^i), \zeta_{k+1}^i \rangle - \frac{\gamma_{k+1}}{n^2} \sum_{j, \ell \in [n]} \text{tr} \left( \sigma(X_k^i, X_k^j)^T H_\phi(X_k^i) \sigma(X_k^i, X_k^\ell) \right) \\
&+ \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle \\
&= \frac{\gamma_{k+1}}{n^2} \sum_{j, \ell \in [n]} \psi(X_k^i, X_k^j, X_k^\ell) + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\
&+ \gamma_{k+1} \langle \nabla \phi(X_k^i), \zeta_{k+1}^i \rangle - \frac{\gamma_{k+1}}{n^2} \sum_{j, \ell \in [n]} \text{tr} \left( \sigma(X_k^i, X_k^j)^T H_\phi(X_k^i) \sigma(X_k^i, X_k^\ell) \right) + \\
&\sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle + \gamma_{k+1} (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i - \gamma_{k+1} (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i
\end{aligned}$$

In this last expression, the terms  $n^{-2} \sum_{j, \ell \in [n]} \psi(X_k^i, X_k^j, X_k^\ell)$  will be played against the integral term at the right hand side of (27), and the other terms will be proven to have negligible effects. Since  $\text{tr}(\xi_{k+1}^i (\xi_{k+1}^i)^T H_\phi(\bar{X}_k^i)) = (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i$ , the term

$$\begin{aligned}
\eta_{k+1}^i &:= \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle + \gamma_{k+1} (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i \\
&- \frac{\gamma_{k+1}}{n^2} \sum_{j, \ell \in [n]} \text{tr} \left( \sigma(X_k^i, X_k^j)^T H_\phi(X_k^i) \sigma(X_k^i, X_k^\ell) \right)
\end{aligned}$$

in the expression above is a martingale increment term with respect to the filtration  $(\mathcal{F}_k^n)_k$ , thanks to Assumption 3–(ii).

To proceed, considering the integral at the right hand side of (27), we can write

$$\begin{aligned}
\int_{t_n+s}^{t_n+t} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du &= \int_{\tau_{k_{t_n+s}}}^{\tau_{k_{t_n+t-1}}} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du \\
&+ \int_{t_n+s}^{\tau_{k_{t_n+s}}} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du + \int_{\tau_{k_{t_n+t-1}}}^{t_n+t} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du,
\end{aligned}$$

and with these decompositions, we obtain  $G(\mathbf{m}_n) = \sum_{l=1}^8 \chi_l^n$ , where:

$$\begin{aligned}
\chi_1^n &:= \frac{1}{n^3} \sum_{i,j,\ell \in [n]} \left\{ \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1} \psi(X_k^i, X_k^j, X_k^\ell) - \int_{\tau_{k_{t_n+s}}}^{\tau_{k_{t_n+t}-1}} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du \right\} Q^i, \\
\chi_2^n &:= \frac{1}{n} \sum_{i \in [n]} \left\{ \phi(\bar{X}_{t_n+t}^i) - \phi(X_{k_{t_n+t}-1}^i) + \phi(X_{k_{t_n+s}}^i) - \phi(\bar{X}_{t_n+s}^i) \right\} Q^i, \\
\chi_3^n &:= -\frac{1}{n^3} \sum_{i,j,\ell \in [n]} \left\{ \int_{t_n+s}^{\tau_{k_{t_n+t}}} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du + \int_{\tau_{k_{t_n+t}-1}}^{t_n+t} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du \right\} Q^i \\
\chi_4^n &:= \frac{1}{n} \sum_{i \in [n]} Q^i \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1} \langle \nabla \phi(X_k^i), \zeta_{k+1}^i \rangle, \\
\chi_5^n &:= \frac{1}{n} \sum_{i \in [n]} \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1} (\xi_{k+1}^i)^T \left( H_\phi(\bar{X}_{\theta_{k+1}}^i) - H_\phi(X_k^i) \right) (\xi_{k+1}^i) Q^i, \\
\chi_6^n &:= \frac{1}{n^2} \sum_{i,j \in [n]} \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \left( \sqrt{2} \gamma_{k+1}^{3/2} b(X_k^i, X_k^j)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) \xi_{k+1}^i \right) Q^i \\
&\quad + \frac{1}{n^3} \sum_{i,p,q \in [n]} \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \left( \frac{1}{2} \gamma_{k+1}^2 b(X_k^i, X_k^p)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) b(X_k^i, X_k^q) \right) Q^i, \\
\chi_7^n &:= \frac{1}{n^2} \sum_{i,j \in [n]} \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1}^{3/2} \left( \left( \sqrt{\gamma_{k+1}} \left( b(X_k^i, X_k^j) + \frac{\zeta_{k+1}^i}{2} \right) + \sqrt{2} \xi_{k+1}^i \right)^T \right. \\
&\quad \left. H_\phi(\bar{X}_{\theta_{k+1}}^i) \zeta_{k+1}^i \right) Q^i, \quad \text{and} \\
\chi_8^n &:= \frac{1}{n} \sum_{i \in [n]} \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \eta_{k+1}^i Q^i.
\end{aligned}$$

To prove our proposition, we show that  $\mathbb{E}|\chi_l^n| \rightarrow 0$  for all  $l \in [8]$ . The notation  $E_\times^\times$  will be generically used to refer to error terms.

Let us start with  $\mathbb{E}|\chi_1^n|$ . For  $i, j, \ell \in [n]$ , writing

$$E_{i,j,\ell}^n := \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1} \psi(X_k^i, X_k^j, X_k^\ell) - \int_{\tau_{k_{t_n+s}}}^{\tau_{k_{t_n+t}-1}} \psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell) du$$

and using the boundedness of  $Q^i$  and the exchangeability as stated by Assumption 3-(i), we obtain that

$$\mathbb{E}|\chi_1^n| \leq C (\mathbb{E}|E_{1,2,3}^n| + (\mathbb{E}|E_{1,1,2}^n| + \mathbb{E}|E_{2,1,1}^n| + \mathbb{E}|E_{1,2,1}^n|)/n + \mathbb{E}|E_{1,1,1}^n|/n^2).$$

We begin by providing a bound on the second moments of  $E_{i,j,\ell}^n$ . Recalling the definition of  $\psi$ , and using the compactness of the support of  $\phi$  along with Assumption 1, we obtain that

$$\begin{aligned}
\mathbb{E}(E_{i,j,\ell}^n)^2 &\leq 2(t-s)^2 \max_{u \in [t_n+s, t_n+t]} \mathbb{E} \|\psi(\bar{X}_u^i, \bar{X}_u^j, \bar{X}_u^\ell)\|^2 \\
&\leq C(t-s)^2 \left( 1 + \sup_{u \geq 0} \mathbb{E}(\bar{X}_u^1)^2 \right) \\
&\leq C(t-s)^2
\end{aligned}$$

thanks to Assumption 4-(i). To obtain that  $\mathbb{E}|\chi_1^n| \rightarrow 0$ , we thus need to show that  $\mathbb{E}|E_{1,2,3}^n| \rightarrow 0$ .

By Prop. 4 above, the sequence  $(\mathbf{m}_n)$  of  $\mathcal{P}(\mathcal{C})$ -valued random variables is tight. By Lemma 1, this is equivalent to the weak\*-relative compactness of the sequence of intensities  $(\mathbb{I}(\mathbf{m}_n))$ . For each Borel set  $A \in \mathcal{B}(\mathcal{C})$ , we furthermore have that

$$\mathbb{I}(\mathbf{m}_n)(A) = \frac{1}{n} \sum_{i \in [n]} \mathbb{P} \left[ \bar{X}_{t_n+}^{i,n} \in A \right] = \mathbb{P} \left[ \bar{X}_{t_n+}^{1,n} \in A \right]$$

by the exchangeability, thus, the sequence of random variables  $(\bar{X}_{t_n+}^{1,n})_n$  is tight. Given  $i \in [3]$ , let us work on the r. v.  $U_n^i := \pi_{[0,t-s]} \# \bar{X}_{t_n+s+}^i$  defined on the set  $\mathcal{C}([0, t-s])$  of continuous functions on the interval  $[0, t-s]$ . Since  $(\bar{X}_{t_n+}^{1,n})_n$  is tight, given an arbitrary  $\varepsilon > 0$ , there is a compact set  $\mathcal{K}_\varepsilon \subset \mathcal{C}([0, t-s])$  such that

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}[U_n^1 \notin \mathcal{K}_\varepsilon] \leq \varepsilon.$$

Writing  $\bar{\gamma}_l = \sup_{k \geq l} \gamma_k$ , we now have

$$\begin{aligned} |E_{1,2,3}^n| &\leq \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1} \max_{\delta \in [0, \gamma_{k+1}]} |\psi(\bar{X}_{\tau_k+\delta}^1, \bar{X}_{\tau_k+\delta}^2, \bar{X}_{\tau_k+\delta}^3) - \psi(\bar{X}_{\tau_k}^1, \bar{X}_{\tau_k}^2, \bar{X}_{\tau_k}^3)| \\ &\leq (t-s) \max_{\substack{u, v \in [0, t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n+s}}}} |\psi(U_n^1(u), U_n^2(u), U_n^3(u)) - \psi(U_n^1(v), U_n^2(v), U_n^3(v))|. \end{aligned}$$

We thus can write

$$\begin{aligned} \mathbb{E}|E_{1,2,3}^n| &= \mathbb{E}|E_{1,2,3}^n| \mathbf{1}_{(U_n^1, U_n^2, U_n^3) \in \mathcal{K}_\varepsilon} + \mathbb{E}|E_{1,2,3}^n| \mathbf{1}_{(U_n^1, U_n^2, U_n^3) \notin \mathcal{K}_\varepsilon} \\ &\leq (t-s) \sup_{f_1, f_2, f_3 \in \mathcal{K}_\varepsilon} \max_{\substack{u, v \in [0, t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n+s}}}} |\psi(f_1(u), f_2(u), f_3(u)) - \psi(f_1(v), f_2(v), f_3(v))| \\ &\quad + \sqrt{\mathbb{E}(E_{1,2,3}^n)^2} \sqrt{3\mathbb{P}[U_n^1 \notin \mathcal{K}_\varepsilon]}. \end{aligned} \tag{28}$$

By the Arzelà-Ascoli theorem, the functions in  $\mathcal{K}_\varepsilon$  are uniformly equicontinuous and bounded. Since  $\psi$  is a continuous function, one can easily check that the set of functions  $\mathcal{S}$  on  $[0, t-s]$  defined as

$$\mathcal{S} := \{u \mapsto \psi(f_1(u), f_2(u), f_3(u)) : f_1, f_2, f_3 \in \mathcal{K}_\varepsilon\}$$

is a set of uniformly equicontinuous functions. As a consequence, the first term at the right hand side of the inequality in (28) converges to zero as  $n \rightarrow \infty$ , since  $\bar{\gamma}_{k_{t_n+s}} \rightarrow 0$ . The second term is bounded by  $C\sqrt{\varepsilon}$  thanks to the bound we obtained on  $\mathbb{E}(E_{1,2,3}^n)^2$ . Since  $\varepsilon$  is arbitrary, we obtain that  $\mathbb{E}|E_{1,2,3}^n| \rightarrow 0$ , thus,  $\mathbb{E}|\chi_1^n| \rightarrow 0$ .

The terms  $\chi_n^2$ ,  $\chi_n^3$ , and  $\chi_n^5$  are dealt with similarly to  $\chi_n^1$ . Considering  $\chi_n^2$ , we have by the exchangeability that  $\mathbb{E}|\chi_n^2| \leq C\mathbb{E}|E_1^n|$ , with

$$\begin{aligned} E_1^n &= \phi(\bar{X}_{t_n+t}^1) - \phi(X_{k_{t_n+t}-1}^1) + \phi(X_{k_{t_n+s}}^1) - \phi(\bar{X}_{t_n+s}^1) \\ &= \phi(\bar{X}_{t_n+t}^1) - \phi(\bar{X}_{\tau_{k_{t_n+t}-1}}^1) + \phi(\bar{X}_{\tau_{k_{t_n+s}}}^1) - \phi(\bar{X}_{t_n+s}^1). \end{aligned}$$

Keeping the notations  $U_n^1 := \pi_{[0,t-s]} \# \bar{X}_{t_n+s+}^1$  and  $\bar{\gamma}_l$  introduced above, we have

$$|E_1^n| \leq 2 \max_{\substack{u, v \in [0, t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n+s}}}} |\phi(U_n^1(u)) - \phi(U_n^1(v))|.$$

Taking  $\varepsilon > 0$ , selecting the compact  $\mathcal{K}_\varepsilon \subset \mathcal{C}([0, t-s])$  as we did for  $\chi_1^n$ , and recalling that the function  $\phi$  is bounded, we have

$$\mathbb{E}|E_1^n| \leq 2 \sup_{f \in \mathcal{K}_\varepsilon} \max_{\substack{u, v \in [0, t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n+s}}}} \|\phi(f(u)) - \phi(f(v))\| + C\mathbb{P}[U_n^1 \notin \mathcal{K}_\varepsilon],$$

and we obtain the  $\mathbb{E}|\chi_n^2| \rightarrow 0$  by the same argument as for  $\chi_n^1$ .

The treatment of  $\chi_n^3$  is very similar to  $\chi_n^2$  and is omitted. Let us provide some details for  $\chi_n^5$ . Here we have by exchangeability that

$$\mathbb{E}|\chi_5^n| \leq \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1} \mathbb{E}|E_k^{1,n}|,$$

where

$$E_k^{1,n} := (\xi_{k+1}^1)^T \left( H_\phi(\bar{X}_{\theta_{k+1}}^1) - H_\phi(X_k^1) \right) (\xi_{k+1}^1) Q^1.$$

satisfies

$$|E_k^{1,n}| \leq C \|\xi_{k+1}^1\|^2 \max_{\substack{u,v \in [0,t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n}+s}}} \|H_\phi(U_n^1(u)) - H_\phi(U_n^1(v))\|.$$

Therefore,

$$\begin{aligned} \mathbb{E} |E_k^{1,n}| &= \mathbb{E} |E_k^{1,n}| \mathbf{1}_{U_n^1 \in \mathcal{K}_\varepsilon} + \mathbb{E} |E_k^{1,n}| \mathbf{1}_{U_n^1 \notin \mathcal{K}_\varepsilon} \\ &\leq C \mathbb{E} \|\xi_{k+1}\|^2 \sup_{f \in \mathcal{K}_\varepsilon} \max_{\substack{u,v \in [0,t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n}+s}}} \|H_\phi(f(u)) - H_\phi(f(v))\| \\ &\quad + \sqrt{\mathbb{E}(E_k^{1,n})^2} \sqrt{\mathbb{P}[U_n^1 \notin \mathcal{K}_\varepsilon]}. \end{aligned}$$

Since  $\mathbb{E} \|\xi_{k+1}\|^2$  and  $\mathbb{E}(E_k^{1,n})^2$  are bounded, we obtain that  $\mathbb{E}|\chi_5^n| \rightarrow 0$ .

Considering the term  $\chi_n^4$ , we have by exchangeability

$$\begin{aligned} \mathbb{E}|\chi_4^n| &\leq C \mathbb{E} \left| \sum_{k=k_{t_n}+s}^{k_{t_n}+t-2} \gamma_{k+1} \langle \nabla \phi(X_k^1), \zeta_{k+1}^1 \rangle \right| \\ &\leq C \mathbb{E} \left| \sum_{k=k_{t_n}+s}^{k_{t_n}+t-2} \gamma_{k+1} \langle \nabla \phi(X_k^1), \mathbb{E}[\zeta_{k+1}^1 | \mathcal{F}_k^n] \rangle \right| \\ &\quad + C \mathbb{E} \left| \sum_{k=k_{t_n}+s}^{k_{t_n}+t-2} \gamma_{k+1} \langle \nabla \phi(X_k^1), \zeta_{k+1}^1 \rangle \right| \\ &:= \mathbb{E}|\chi_{4,1}^n| + \mathbb{E}|\chi_{4,2}^n|, \end{aligned}$$

where  $\zeta_k^1 = \zeta_k^1 - \mathbb{E}[\zeta_k^1 | \mathcal{F}_{k-1}^n]$  is a martingale increment with respect to the filtration  $(\mathcal{F}_k^n)_k$ . We have

$$\mathbb{E}|\chi_{4,1}^n| \leq C(t-s) \sup_{l \geq k_{t_n}+s} \mathbb{E} \|\mathbb{E}[\zeta_{l+1}^1 | \mathcal{F}_l^n]\|,$$

which converges to zero by Assumption 3-(iii). By the martingale property, we furthermore have

$$\mathbb{E}(\chi_{4,2}^n)^2 \leq C \sum_{k=k_{t_n}+s}^{k_{t_n}+t-2} \gamma_{k+1}^2 \leq C \bar{\gamma}_{k_{t_n}+s} (t-s),$$

which also converges to zero. Thus,  $\mathbb{E}|\chi_4^n| \rightarrow 0$ .

We now turn to  $\chi_6^n$ . Here we write

$$\chi_6^n = \frac{1}{n} \sum_{i \in [n]} \sum_{k=k_{t_n}+s}^{k_{t_n}+t-2} \gamma_{k+1}^{3/2} E_k^i,$$

where

$$\begin{aligned} E_k^i &:= \frac{1}{n} \sum_{j \in [n]} \sqrt{2} b(X_k^i, X_k^j)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) \xi_{k+1}^i Q^i \\ &\quad + \frac{1}{n^2} \sum_{p,q \in [n]} \frac{1}{2} \sqrt{\gamma_{k+1}} b(X_k^i, X_k^p)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) b(X_k^i, X_k^q) Q^i \end{aligned}$$

satisfies

$$|E_k^i| \leq \frac{C}{n} \sum_{j \in [n]} (1 + \|X_k^j\|) \|\xi_{k+1}^i\| + \frac{C}{n^2} \sqrt{\gamma_{k+1}} \sum_{p,q \in [n]} (1 + \|X_k^p\|)(1 + \|X_k^q\|).$$

We readily obtain from Assumptions 3, and 4-(i) that  $\mathbb{E}|E_k^i| \leq C$ , which leads to  $\mathbb{E}|\chi_6^n| \rightarrow 0$ .

The treatment of the term  $\chi_7^n$  is similar and is omitted.

We finally deal with  $\chi_8^n$  that involves the martingale increments  $\eta_k^i$ . We decompose this term by writing

$$\begin{aligned}\chi_8^n &= \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \frac{1}{n} \sum_{i \in [n]} \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle Q^i + \\ &\quad \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \frac{\gamma_{k+1}}{n^3} \sum_{i,j,\ell \in [n]} \left( (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i - \text{tr} \left( \sigma(X_k^i, X_k^j)^T H_\phi(X_k^i) \sigma(X_k^i, X_k^\ell) \right) \right) Q^i \\ &:= \chi_{8,1}^n + \chi_{8,2}^n.\end{aligned}$$

Since the random vectors  $\xi_{k+1}^1, \dots, \xi_{k+1}^n$  are decorrelated conditionally to  $\mathcal{F}_k^n$  by Assumption 3-(ii), we obtain that

$$\mathbb{E} \left[ \left( \frac{1}{n} \sum_{i \in [n]} \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle Q^i \right)^2 \middle| \mathcal{F}_k^n \right] \leq C \frac{\gamma_{k+1}}{n},$$

and by the martingale property,

$$\mathbb{E}(\chi_{8,1}^n)^2 \leq \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} C \frac{\gamma_{k+1}}{n} \leq \frac{C(t-s)}{n}.$$

Using the martingale property again along with the inequality  $(\sum_1^n a_i)^2 \leq n \sum_1^n a_i^2$ , we also have

$$\begin{aligned}\mathbb{E}(\chi_{8,2}^n)^2 &\leq \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1}^2 \mathbb{E} \left( \frac{1}{n^3} \sum_{i,j,\ell \in [n]} \left( (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i \right. \right. \\ &\quad \left. \left. - \text{tr} \left( \sigma(X_k^i, X_k^j)^T H_\phi(X_k^i) \sigma(X_k^i, X_k^\ell) \right) \right) Q^i \right)^2 \\ &\leq C \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1}^2 \\ &\leq \bar{\gamma}_{k_{t_n+s}} C(t-s).\end{aligned}$$

It results that  $\mathbb{E}(\chi_8^n)^2 \rightarrow 0$ . The proof of Prop. 10 is completed.  $\square$

*Proof of Proposition 5.* Let  $(t_n, \varphi_n)_n$  be a  $\mathbb{R}_+ \times \mathbb{N}^*$ -valued sequence such that the distribution of  $(\Phi_{t_n}(m^{\varphi_n}))_n$  converges to a measure  $M \in \mathcal{M}$ , which exists thanks to the tightness of  $(\Phi_{t_n}(m^{\varphi_n}))_n$  as established by Prop. 4. Let  $G \in \mathcal{G}_p$ . By the continuity of  $G$  as established by Lemma 3,  $G(\Phi_{t_n}(m^{\varphi_n}))$  converges in distribution to  $G_\# M \in \mathcal{P}(\mathbb{R})$ . On the other hand, we know by the previous proposition that  $G(\Phi_{t_n}(m^{\varphi_n}))$  converges in probability to zero. Therefore,  $G_\# M = \delta_0$ .

Let  $\text{supp}(M) \subset \mathcal{P}_p(\mathcal{C})$  be the support of  $M$ , and let  $\rho \in \text{supp}(M)$ . By definition of the support,  $M(\mathcal{N}) > 0$  for each neighborhood  $\mathcal{N}$  of  $\rho$ . Therefore, since  $G_\# M = \delta_0$ , there exists a sequence  $(\rho_l)_{l \in \mathbb{N}}$  such that  $\rho_l \in \text{supp}(M)$ ,  $G(\rho_l) = 0$ , and  $\rho_l \rightarrow_l \rho$  in  $\mathcal{P}_p(\mathcal{C})$ . By the continuity of  $G$ , we obtain that  $G(\rho) = 0$ , which shows that  $\text{supp}(M) \subset G^{-1}(\{0\})$ . Since  $G$  is arbitrary, we obtain that  $\text{supp}(M) \subset \bigcap_{G \in \mathcal{G}_p} G^{-1}(\{0\})$ , and the theorem is proven.  $\square$

### 5.3 Proof of Theorem 1

Throughout this paragraph, we assume that  $1 \leq p \leq 2$ .

We define the following collection  $(M_t^n : t \geq 0, n \in \mathbb{N}^*)$  of r.v. on  $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ :

$$M_t^n := \frac{1}{t} \int_0^t \delta_{\Phi_s(m^n)} ds. \quad (29)$$

**Lemma 6.** *The collection of r.v.  $(M_t^n, t \geq 0, n \in \mathbb{N}^*)$  is tight in  $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ .*

*Proof.* Based on Lemma 1, we just need to establish that the family of measures  $(\mathbb{I}(M_t^n))$  is relatively compact in the space  $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ . Recall that  $\mathbb{I}(M_t^n)$  is the probability measure which, to every Borel subset  $A \subset \mathcal{P}_p(\mathcal{C})$ , associates:

$$\mathbb{I}(M_t^n)(A) = \frac{1}{t} \int_0^t \mathbb{P}(\Phi_s(m^n) \in A) ds$$

Consider  $\varepsilon > 0$ . By Prop. 4, there exists a compact set  $\mathcal{K} \in \mathcal{P}_p(\mathcal{C})$  such that  $\mathbb{P}(\Phi_s(m^n) \in \mathcal{K}) > 1 - \varepsilon$ , for all  $s, n$ . As a consequence,  $\mathbb{I}(M_t^n)(\mathcal{K}) > 1 - \varepsilon$ . The proof is completed.  $\square$

Let us denote by  $\mathcal{M}$  the set of weak $\star$  accumulation points of the net  $((M_t^n)_\# \mathbb{P} : t \geq 0, n \in \mathbb{N}^*)$ , as  $(t, n) \rightarrow (\infty, \infty)$ . By Lemma 6,  $\mathcal{M}$  is a non empty subset of  $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ . Define:

$$\mathcal{V}_p = \{M \in \mathcal{P}(\mathcal{P}_p(\mathcal{C})) : M(\mathcal{V}_p) = 1\}.$$

**Lemma 7.** *For every  $\Upsilon \in \mathcal{M}$ ,  $\Upsilon(\mathcal{V}_p) = 1$ .*

*Proof.* Consider  $\Upsilon \in \mathcal{M}$ . Without restriction, we write  $\Upsilon$  as the weak $\star$  limit of some sequence of the form  $(M_{t_n}^n)_\# \mathbb{P}$ . The distance  $W_p(\cdot, \mathcal{V}_p)$  to the set  $\mathcal{V}_p$  (which is non empty by Prop. 5) is a continuous function on  $\mathcal{P}_p(\mathcal{C})$ . Denoting by  $\langle \cdot, \cdot \rangle$  the natural dual pairing on  $C_b(\mathcal{P}_p(\mathcal{C})) \times \mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ , the function  $\langle W_p(\cdot, \mathcal{V}_p), \cdot \rangle$  is a continuous on  $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ . Thus, the sequence of real r.v.  $\langle W_p(\cdot, \mathcal{V}_p), M_{t_n}^n \rangle$  converges in distribution to  $\langle W_p(\cdot, \mathcal{V}_p), \cdot \rangle_\# \Upsilon$ . These variables being bounded, we obtain by taking the limits in expectation:

$$\begin{aligned} \int \int W_p(m, \mathcal{V}_p) dM(m) d\Upsilon(M) &= \lim_{n \rightarrow \infty} \mathbb{E}(\langle W_p(\cdot, \mathcal{V}_p), M_{t_n}^n \rangle) \\ &= \lim_{n \rightarrow \infty} \frac{1}{t_n} \int_0^{t_n} \mathbb{E}(W_p(\Phi_s(m^n), \mathcal{V}_p)) ds \\ &\leq \limsup_{(t, n) \rightarrow (\infty, \infty)} \mathbb{E}(W_p(\Phi_t(m^n), \mathcal{V}_p)) = 0, \end{aligned}$$

where the last equality is due to Prop. 5. As  $\mathcal{V}_p$  is closed by Prop. 2, this concludes the proof.  $\square$

Recall the definition of the shift  $\Theta_t : x \mapsto x_{t+}$ , defined in  $\mathcal{C}$ . For every  $t \geq 0$ , define  $(\Theta_t)_{\#\#} = ((\Theta_t)_\#)_\#$ . Define:

$$\mathcal{I} := \{M \in \mathcal{P}(\mathcal{P}_p(\mathcal{C})) : \forall t > 0, M = (\Theta_t)_{\#\#} M\}.$$

In other words, for every  $M \in \mathcal{I}$  and for every  $t > 0$ ,  $(\Theta_t)_\#$  preserves  $M$ .

**Lemma 8.** *For every  $\Upsilon \in \mathcal{M}$ ,  $\Upsilon(\mathcal{I}) = 1$ .*

*Proof.* Similarly to the proof of Lemma 7, we assume without restriction that  $\Upsilon = \lim_{n \rightarrow \infty} (M_{t_n}^n)_\# \mathbb{P}$  in the weak $\star$  sense. Set  $t > 0$ . The map  $M \mapsto d_L(M, (\Theta_t)_{\#\#} M)$  is continuous on  $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ , where we recall that  $d_L$  stands for the Lévy-Prokhorov distance. Thus, by Fatou's lemma,

$$\int d_L(M, (\Theta_t)_{\#\#} M) d\Upsilon(M) \leq \limsup_{n \rightarrow \infty} \mathbb{E}(d_L(M_{t_n}^n, (\Theta_t)_{\#\#} M_{t_n}^n)). \quad (30)$$

Note that:

$$(\Theta_t)_{\#\#} M_{t_n}^n = \frac{1}{t_n} \int_t^{t+t_n} \delta_{(\Theta_s)_\# m^n} ds.$$

In particular, for every Borel set  $A \subset \mathcal{P}_p(\mathcal{C})$ ,  $|(\Theta_t)_{\#\#} M_{t_n}^n(A) - M_{t_n}^n(A)| \leq 2t/t_n$ . The Lévy-Prokhorov distance being bounded by the total variation distance,  $d_L(M_{t_n}^n, (\Theta_t)_{\#\#} M_{t_n}^n) \leq 2t/t_n$  which tends to zero. The l.h.s. of Eq. (30) is zero, which proves the statement for a fixed value of  $t$ . The proof of the statement for all  $t$ , is easily concluded by a using dense denumerable subset argument.  $\square$

Define:  $\mathcal{B}_p = \{M \in \mathcal{P}(\mathcal{P}_p(\mathcal{C})) : M(\text{BC}_p) = 1\}$ .

**Proposition 11.** *For every  $\Upsilon \in \mathcal{M}$ ,  $\Upsilon(\mathcal{B}_p) = 1$ .*

*Proof.* Consider an arbitrary sequence of the form  $((M_{t_n}^n)_\# \mathbb{P})$  where  $t_n \rightarrow \infty$ , converging in distribution to some measure  $\Upsilon \in \mathcal{M}$  as  $n \rightarrow \infty$ . By Lemma 8, the map  $(\Theta_t)_\# : \mathcal{P}_p(\mathcal{C}) \rightarrow \mathcal{P}_p(\mathcal{C})$  preserves the measure  $M$ , for all  $M$   $\Upsilon$ -a.e., and for all  $t$ . By Lemma 7,  $M(\mathcal{V}_p) = 1$ . Thus, the restriction of the map  $(\Theta_t)_\#$  to  $\mathcal{V}_p$ , still denoted by  $(\Theta_t)_\# : \mathcal{V}_p \rightarrow \mathcal{V}_p$  preserves the measure  $M$  as well, for all  $M$   $\Upsilon$ -a.e.. By the Poincaré recurrence theorem, stated in Theorem 2.3 of [Mañ87], it follows that  $M(\text{BC}_p) = 1$  for all  $M$   $\Upsilon$ -a.e.  $\square$

*Proof of Theorem 1.* To conclude, assume by contradiction that the conclusion of Theorem 1 does not hold. Then, there exists  $\varepsilon > 0$  and a sequence, which, without restriction, we may assume to have the form  $((M_{t_n}^n)_\# \mathbb{P})$ , such that for all  $n$  large enough,

$$\mathbb{E}(\langle W_p(\cdot, \text{BC}_p), M_{t_n}^n \rangle) > \varepsilon, \quad (31)$$

where  $\langle \cdot, \cdot \rangle$  is the natural dual pairing on  $C_b(\mathcal{P}_p(\mathcal{C})) \times \mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ . By Lemma 6, one can extract an other subsequence, which we still denote by  $((M_{t_n}^n)_\# \mathbb{P})$ , converging to  $\Upsilon \in \mathcal{M}$ . As a consequence,

$$\lim_{n \rightarrow \infty} \mathbb{E}(\langle W_p(\cdot, \text{BC}_p), M_{t_n}^n \rangle) = \int \int W_p(m, \text{BC}_p) dM(m) d\Upsilon(M) = 0,$$

where we used the fact that, due to Prop. 11,  $\int W_p(m, \text{BC}_p) dM(m) = 0$  for  $\Upsilon$ -almost all  $M$ . This contradicts Eq. (31).  $\square$

## 5.4 Proof of Corollary 1

Throughout this paragraph, we assume that  $1 \leq p \leq 2$ . We define the functions, for  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ ,

$$g_1(\mu) := W_p(\mu, (\pi_0)_\# \text{BC}_p),$$

and

$$g_2(\mu) := W_p(\mu, (\pi_0)_\# \text{BC}_p)^p.$$

Consider the r.v.

$$Y_{n,\ell}(s) := g_\ell \left( \frac{1}{n} \sum_{i=1}^n \delta_{\bar{X}_s^{i,n}} \right),$$

for  $\ell \in [2]$ .

**Lemma 9.** *The r.v.  $(Y_{n,\ell}(s) : s > 0, n \in \mathbb{N})$  are uniformly integrable for  $\ell \in [2]$ .*

*Proof.* Let  $\ell \in [2]$ . Note that  $Y_{n,\ell}(s) \leq C(1 + \frac{1}{n} \sum_i \|\bar{X}_s^{i,n}\|^p)$ . Hence for a convex, and increasing function  $F : \mathbb{R}_+^* \rightarrow \mathbb{R}$ , by the exchangeability stated in Assumption 3, we obtain  $\mathbb{E}(F(Y_{n,\ell}(s))) \leq \mathbb{E}(F(C(1 + \frac{1}{n} \sum_i \|\bar{X}_s^{i,n}\|^p))) \leq F(C(1 + \mathbb{E}(\|\bar{X}_s^{1,n}\|^p)))$ . By de la Vallée Poussin theorem, the random variables  $(Y_{n,\ell}(s) : s > 0, n \in \mathbb{N})$  are uniformly integrable if the random variables  $(\|\bar{X}_s^{1,n}\|^p : s > 0, n \in \mathbb{N})$  are uniformly integrable. We conclude using Assumption 4-(i) if  $p < 2$ , or Assumption 4-(ii) if  $p = 2$ .  $\square$

Let  $\ell \in [2]$ , recall the definition of  $M_t^n$  in Eq. (29), and recall that  $\mathcal{M}$  is the set of cluster points of  $((M_t^n)_\# \mathbb{P} : t \geq 0, n \in \mathbb{N}^*)$  as  $(t, n) \rightarrow (\infty, \infty)$ . Consider an arbitrary sequence  $t_n \rightarrow \infty$ , such that  $(M_{t_n}^n)_\# \mathbb{P}$  converges to some measure  $\Upsilon \in \mathcal{M}$ . Consider  $\varepsilon > 0$ . By Lemma 9, there exists  $a > 0$  such that  $\sup_{n,s} \mathbb{E}(Y_{n,\ell}(s) \mathbf{1}_{Y_{n,\ell}(s) > a}) < \varepsilon$ . Using the inequality  $y \leq a \wedge y + y \mathbf{1}_{y > a}$ , we obtain:

$$\begin{aligned} \mathbb{E} \left( \frac{1}{t_n} \int_0^{t_n} Y_{n,\ell}(s) ds \right) &\leq \mathbb{E} \left( \frac{1}{t_n} \int_0^{t_n} a \wedge Y_{n,\ell}(s) ds \right) + \varepsilon \\ &= \mathbb{E} \left( \int a \wedge g_\ell((\pi_0)_\# m) dM_{t_n}^n(m) \right) + \varepsilon \end{aligned} \quad (32)$$

The restriction of  $\pi_0$  to the space  $\mathcal{P}_p(\mathcal{C})$ , which we still denote by  $\pi_0$ , is a continuous function on  $(\mathcal{P}_p(\mathcal{C}), W_p) \rightarrow (\mathcal{P}_p(\mathbb{R}^d), W_p)$ , where  $W_p$  represents the  $p$ -th order Wasserstein distance on  $\mathcal{P}(\mathbb{R}^d)$ . As a consequence, the pushforward map  $(\pi_0)_\# : \mathcal{P}(\mathcal{P}_p(\mathcal{C})) \rightarrow \mathcal{P}(\mathcal{P}_p(\mathbb{R}^d))$  is continuous. Therefore, as  $(\pi_0)_\# \text{BC}_p$  is non empty by Prop. 11, the function  $M \mapsto \int a \wedge g_\ell((\pi_0)_\# m) dM(m)$  is bounded and continuous on  $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$ . Recall that  $M_{t_n}^n$  converges in distribution to  $\Upsilon$ , and noting that, by Prop. 11,

$$\int \int g_\ell((\pi_0)_\# m) dM(m) d\Upsilon(M) = 0.$$

Hence, by letting  $n \rightarrow \infty$  in Eq. (32), we obtain  $\limsup_n \mathbb{E}(\frac{1}{t_n} \int_0^{t_n} Y_n(s) ds) \leq \varepsilon$ . As  $\varepsilon$  is arbitrary,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \frac{1}{t_n} \int_0^{t_n} Y_n(s) ds \right) = 0. \quad (33)$$

In order to establish the statement of Corollary 1, we now should consider replacing the integral in Eq. (33) by a sum. This last part is only technical. Recall the definition of  $k_t := \inf\{k : \sum_{i=1}^k \gamma_i \geq t\}$ , and  $\tau_k$  in Eq. (4). Let  $(\alpha_n)$  be a sequence of integers tending to infinity. By the triangular inequality,

$$\begin{aligned} \mathbb{E} \left( \frac{\sum_{l=1}^{\alpha_n} \gamma_l g_\ell(\mu_l^n)}{\sum_{l=1}^{\alpha_n} \gamma_l} \right) &= \mathbb{E} \left( \frac{1}{\tau_{\alpha_n}} \int_0^{\tau_{\alpha_n}} g_\ell(\mu_{k_s}^n) ds \right) \\ &\leq \mathbb{E} \left( \frac{1}{\tau_{\alpha_n}} \int_0^{\tau_{\alpha_n}} W_p(\mu_{k_s}^n, \frac{1}{n} \sum_{i \in [n]} \delta_{\bar{X}_s^{i,n}}) \mathbb{1}_{\ell=1} ds \right) \\ &\quad + \mathbb{E} \left( \frac{1}{\tau_{\alpha_n}} \int_0^{\tau_{\alpha_n}} W_p(\mu_{k_s}^n, \frac{1}{n} \sum_{i \in [n]} \delta_{\bar{X}_s^{i,n}})^p \mathbb{1}_{\ell=2} ds \right) \\ &\quad + \mathbb{E} \left( \frac{1}{\tau_{\alpha_n}} \int_0^{\tau_{\alpha_n}} Y_n(s) ds \right). \end{aligned}$$

The third term in the righthand side of the above inequality tends to zero by Eq. (33) with  $t_n = \tau_{\alpha_n}$ . We should therefore establish that the first and the second term vanish. For an arbitrary integer  $l$  and  $s \in [\tau_l, \tau_{l+1}]$ ,

$$\begin{aligned} \mathbb{E} \left[ W_p \left( \mu_l^n, \frac{1}{n} \sum_{i \in [n]} \delta_{\bar{X}_s^{i,n}} \right) \right] &\leq \mathbb{E} \left( \left( \frac{1}{n} \sum_{i \in [n]} \|X_l^{i,n} - \bar{X}_s^{i,n}\|^p \right)^{1/p} \right) \\ &\leq (\mathbb{E}(\|X_l^{1,n} - \bar{X}_s^{1,n}\|^p))^{1/p}. \end{aligned}$$

where the last inequality uses Jensen's inequality and the exchangeability assumption. Continuing the estimation,

$$\begin{aligned} \mathbb{E}(\|X_l^{1,n} - \bar{X}_s^{1,n}\|^p) &\leq \mathbb{E}(\|X_{l+1}^{1,n} - X_l^{1,n}\|^p) \\ &\leq \mathbb{E} \left[ 3^{p-1} \gamma_{l+1}^p \frac{1}{n} \sum_{j \in [n]} \|b(X_l^{1,n}, X_l^{j,n})\|^p \right] \\ &\quad + \mathbb{E} \left[ 3^{p-1} \gamma_{l+1}^{p/2} \|\xi_{l+1}^{1,n}\|^p \right] + \mathbb{E} \left[ 3^{p-1} \gamma_{l+1}^p \|\zeta_{l+1}^{1,n}\|^p \right]. \\ &\leq C(\gamma_{l+1}^{p/2} + \gamma_{l+1}^p), \end{aligned}$$

where we used Assumptions 1, and 3. Consequently,

$$\mathbb{E} \left( \frac{1}{\tau_{\alpha_n}} \int_0^{\tau_{\alpha_n}} W_p(\mu_{k_s}^n, \frac{1}{n} \sum_{i \in [n]} \delta_{\bar{X}_s^{i,n}}) ds \right) \leq \frac{\sum_{l=1}^{\alpha_n} \gamma_l (C(\gamma_{l+1}^{p/2} + \gamma_{l+1}^p))^{1/p}}{\sum_{l=1}^{\alpha_n} \gamma_l}.$$

and, by the same computation,

$$\mathbb{E} \left( \frac{1}{\tau_{\alpha_n}} \int_0^{\tau_{\alpha_n}} W_p(\mu_{k_s}^n, \frac{1}{n} \sum_{i \in [n]} \delta_{\bar{X}_s^{i,n}})^p ds \right) \leq \frac{\sum_{l=1}^{\alpha_n} \gamma_l (C(\gamma_{l+1}^{p/2} + \gamma_{l+1}^p))}{\sum_{l=1}^{\alpha_n} \gamma_l}.$$

As Assumption 2 holds,  $C(\gamma_{l+1}^{p/2} + \gamma_{l+1}^p) \rightarrow_{l \rightarrow \infty} 0$ , and  $\sum_{l \geq 1} \gamma_l = \infty$ . Therefore, by Stolz-Cesàro theorem, the r.h.s. of the above inequality converges to 0 when  $n \rightarrow \infty$ . Hence,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \frac{\sum_{l=1}^{\alpha_n} \gamma_l g_\ell(\mu_l^n)}{\sum_{l=1}^{\alpha_n} \gamma_l} \right) = 0,$$

for an arbitrary sequence  $(\alpha_n)$  diverging to  $\infty$ . By Markov's inequality, Corollary 1 is proven.

## 5.5 Proof of Theorem 2

We let the assumptions of the theorem hold.

**Lemma 10.** *For a nonempty compact set  $K \subset \mathcal{P}_p(\mathbb{R}^d)$ , it holds that*

$$\lim_{t \rightarrow \infty} \max_{\nu \in K} W_p(\Psi_t(\nu), A_p) = 0.$$

*Proof.* Assume for the sake of contradiction that

$$\exists \varepsilon > 0, \exists (\nu_n) \subset K, \exists (t_n) \rightarrow \infty \text{ such that } W_p(\Psi_{t_n}(\nu_n), A_p) > \varepsilon.$$

Choose  $\delta > 0$  small enough so that the  $\delta$ -neighborhood  $A_p^\delta$  of  $A_p$  for the distance  $W_p$  is included in the fundamental neighborhood of  $A_p$ . Up to taking a subsequence, we can assume by the compactness of  $K$  that there exists  $\nu_\infty \in K$  such that  $\nu_n \rightarrow_n \nu_\infty$ . Since  $A_p$  is a global attractor, there exists  $T > 0$  such that  $W_p(\Psi_T(\nu_\infty), A_p) \leq \delta/2$ . Furthermore, by the continuity of  $\Psi$ , there exists  $n_0$  such that

$$\forall n \geq n_0, \quad W_p(\Psi_T(\nu_n), \Psi_T(\nu_\infty)) \leq \delta/2.$$

This implies that  $\Psi_T(\nu_n) \in A_p^\delta$  for all  $n \geq n_0$ . Since  $A_p^\delta$  is included in the fundamental neighborhood of  $A_p$ , there exists  $\tilde{T} > 0$  such that

$$\forall n \geq n_0, \forall t \geq \tilde{T}, \quad W_p(\Psi_{\tilde{T}+t}(\nu_n), A_p) \leq \varepsilon,$$

and we obtain our contradiction.  $\square$

We now prove Theorem 2. Recall that the collection  $\{\Phi_t(m^n)\}$  is tight in  $\mathcal{P}_p(\mathcal{C})$  by Prop. 4. Let  $(t_n, \varphi_n)$  be a sequence such that  $(t_n, \varphi_n) \rightarrow_n (\infty, \infty)$  and such that  $(\Phi_{t_n}(m^{\varphi_n}))_n$  converges in distribution to  $M \in \mathcal{M}$  as given by (11). To prove Theorem 2, it will be enough to show that

$$\forall \delta, \varepsilon > 0, \exists T > 0, \quad \limsup_n \mathbb{P}(W_p(m_{t_n+T}^{\varphi_n}, A_p) \geq \delta) \leq \varepsilon.$$

This shows indeed that

$$W_p(m_t^n, A_p) \xrightarrow[(t,n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0,$$

and by taking  $t = \tau_k$  and by recalling that  $m_{\tau_k}^n = \mu_k^n$ , we obtain our theorem.

Fix  $\delta$  and  $\varepsilon$ . By the tightness of the family  $\{\Phi_t(m^n)\}$ , there exists a compact set  $\mathcal{D} \subset \mathcal{P}_p(\mathcal{C})$  such that  $\mathbb{P}(\Phi_t(m^n) \in \mathcal{D}) \geq 1 - \varepsilon/2$  for each couple  $(t, n)$ . This implies that  $M(\mathcal{D}) \geq 1 - \varepsilon/2$  by the Portmanteau theorem. Since  $\mathbf{V}_p$  is closed by Prop. 5, the set  $\mathcal{K} = \mathcal{D} \cap \mathbf{V}_p$  is compact in  $\mathcal{P}_p(\mathcal{C})$ , and by consequence, it is compact in  $\mathbf{V}_p$  for the trace topology. By the same proposition,  $M(\mathbf{V}_p) = 1$ , therefore,  $M(\mathcal{K}) \geq 1 - \varepsilon/2$ .

Since  $\mathcal{P}_p(\mathcal{C})$  is Polish, we can apply Skorokhod's representation theorem [Bil99, Th. 6.7] to the sequence  $(\Phi_{t_n}(m^{\varphi_n}))$ , yielding the existence of a probability space  $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ , a sequence of  $\mathcal{P}_p(\mathcal{C})$ -valued random variables  $(\tilde{m}^n)$  on  $\tilde{\Omega}$  and a  $\mathcal{P}_p(\mathcal{C})$ -valued random variable  $\tilde{m}^\infty$  on  $\tilde{\Omega}$  such that  $(\tilde{m}^n)_\# \tilde{\mathbb{P}} = (\Phi_{t_n}(m^{\varphi_n}))_\# \mathbb{P}$ ,  $(\tilde{m}^\infty)_\# \tilde{\mathbb{P}} = M$ , and  $\tilde{m}^n \rightarrow \tilde{m}^\infty$  pointwise on  $\tilde{\Omega}$ . Noting that  $m_{t_n+T}^{\varphi_n}$  and  $\tilde{m}_T^n$  have the same probability distribution as  $\mathcal{P}_p(\mathbb{R}^d)$ -valued random variables, we show that

$$\exists T > 0, \quad \limsup_n \tilde{\mathbb{P}}(W_p(\tilde{m}_T^n, A_p) \geq \delta) \leq \varepsilon. \quad (34)$$

to establish our theorem. Observing that the function  $\rho \mapsto (\pi_0)_\# \rho$  is a continuous  $\mathcal{P}(\mathcal{C}) \rightarrow \mathcal{P}_p(\mathbb{R}^d)$  function, the set  $K = (\pi_0)_\# \mathcal{K}$  is a nonempty compact set of  $\mathcal{P}_p(\mathbb{R}^d)$ . Applying Lemma 10 to the semi-flow  $\Psi$  and to the compact  $K$ , we set  $T > 0$  in such a way that

$$\max_{\nu \in K} W_p(\Psi_T(\nu), A_p) \leq \delta/2.$$

By the triangular inequality, we have

$$W_p(\tilde{m}_T^n, A_p) \leq W_p(\tilde{m}_T^n, \tilde{m}_T^\infty) + W_p(\tilde{m}_T^\infty, A_p).$$

The first term at the right hand side converges to zero for each  $\tilde{\omega} \in \tilde{\Omega}$  by the continuity of the function  $\rho \mapsto (\pi_T)_\# \rho$ , thus, this convergence takes place in probability. We also know that for  $\tilde{\mathbb{P}}$ -almost all  $\tilde{\omega} \in \tilde{\Omega}$ , it holds that  $\tilde{m}^\infty \in \mathbf{V}_p$ . Thus, regarding the second term, we have  $\tilde{m}_T^\infty = \Psi_T(\tilde{m}_0^\infty)$  for these  $\tilde{\omega}$ , and we can write

$$\begin{aligned} \tilde{\mathbb{P}}(W_p(\tilde{m}_T^\infty, A_p) \geq \delta) \\ \leq \tilde{\mathbb{P}}(\tilde{m}^\infty \notin \mathcal{K}) + \tilde{\mathbb{P}}((W_p(\Psi_T(\tilde{m}_0^\infty), A_p) \geq \delta) \cap (\tilde{m}_0^\infty \in K)). \end{aligned}$$

When  $\tilde{m}_0^\infty \in K$ , it holds that  $W_p(\Psi_T(\tilde{m}_0^\infty), A_p) \leq \delta/2$ , thus, the second term at the right hand side of the last inequality is zero. The first term satisfies  $\tilde{\mathbb{P}}(\tilde{m}^\infty \notin \mathcal{K}) = 1 - M(\mathcal{K}) \leq \varepsilon/2$ , and the statement (34) follows. Theorem 2 is proven.

## 6 Proofs of Section 4

The Assumptions 5 and  $\sigma > 0$  are standing in this section.

### 6.1 Proof of Prop. 6

**Lemma 11.** *Let  $\rho \in \mathbf{V}_2$ . For every  $t > 0$ ,  $\rho_t$  admits a density  $x \mapsto \varrho(t, x) \in C^1(\mathbb{R}^d, \mathbb{R})$ . For every  $R > 0, t_2 > t_1 > 0$ , there exists a constant  $C_{R, t_1, t_2} > 0$  such that:*

$$\inf_{t \in [t_1, t_2], \|x\| \leq R} \varrho(t, x) \geq C_{R, t_1, t_2}, \quad (35)$$

and there exist a constant  $C_{t_1, t_2} > 0$ , such that

$$\sup_{x \in \mathbb{R}^d, t \in [t_1, t_2]} \|\nabla \varrho(t, x)\| + \varrho(t, x) \leq C_{t_1, t_2}. \quad (36)$$

Finally,

$$\sup_{t \in [t_1, t_2]} \int (1 + \|x\|^2) \|\nabla \varrho(t, x)\| dx < \infty. \quad (37)$$

*Proof.* The result is an application of Th.1.2 in [MPZ21] with the non homogeneous vector field  $\tilde{b}(t, x) := \int b(x, y) d\rho_t(y)$ . The proof consists in verifying the conditions of the latter theorem. By Assumption 5, for every  $(x, y, T) \in (\mathbb{R}^d)^2 \times \mathbb{R}_+$ ,

$$\begin{aligned} \sup_{t \in [0, T]} \|\tilde{b}(t, x) - \tilde{b}(t, y)\| &\leq \|\nabla V(x) - \nabla V(y)\| \\ &\quad + \sup_{t \in [0, T]} \int \|\nabla U(x - z) - \nabla U(y - z)\| d\rho_t(z) \\ &\leq C(\|x - y\|^\beta \vee \|x - y\|), \end{aligned}$$

Moreover,

$$\sup_{t \in [0, T]} \tilde{b}(t, x) \leq C(1 + \|x\| + \int \sup_{t \in [0, T]} \|y_t\| d\rho(y)) \leq C(1 + \|x\|). \quad (38)$$

As  $\sigma > 0$ , [MPZ21, Th. 1.2] applies:  $\rho$  admits a density  $x \mapsto \varrho(t, x) \in C^1(\mathbb{R}^d)$ , for  $0 < t \leq T$ , and there exists four constants  $(C_{i, T}, \lambda_{i, T})_{i \in [2]}$ , such that:

$$\begin{aligned} \frac{1}{C_{1, T} t^{d/2}} \int \exp\left(-\frac{\|x - \theta_t(y)\|^2}{\lambda_{1, T} t}\right) d\rho_0(y) &\leq \varrho(t, x) \\ \varrho(t, x) &\leq \frac{C_{1, T}}{t^{d/2}} \int \exp\left(-\frac{\lambda_{1, T}}{t} \|x - \theta_t(y)\|^2\right) d\rho_0(y) \\ \|\nabla \varrho(t, x)\| &\leq \frac{C_{2, T}}{t^{(d+1)/2}} \int \exp\left(-\frac{\lambda_{2, T}}{t} \|x - \theta_t(y)\|^2\right) d\rho_0(y), \end{aligned}$$

where the map  $t \mapsto \theta_t(y)$  is a solution to the ordinary differential equation:  $\frac{d\theta_t(y)}{dt} = \tilde{b}(t, \theta_t(y))$  with initial condition  $\theta_0(y) = y$ . By Grönwall's lemma and Eq. (38), there exists a constant  $C_T$

such that  $\|\theta_t(y)\| \leq C_T \|y\|$ , for every  $n, y$ , and  $t \leq T$ . For every  $t_1 \leq t \leq t_2$ , and every  $x$ , we obtain using a change of variables:

$$\begin{aligned} (C_{1,t_2} t_1^{d/2})^{-1} &\geq \varrho(t, x) \\ &\geq C_{1,t_2} t_2^{-d/2} \exp\left(-\frac{2}{\lambda_{1,t_2} t_1} \|x\|^2\right) \int \exp\left(-\frac{2C_{t_2}}{\lambda_{1,t_2} t_1} \|y\|^2\right) d\rho_0(y) \\ &\int (1 + \|x\|^2) \|\nabla \varrho(t, x)\| dx \\ &\leq C_{2,t_2} t_1^{-(d+1)/2} \int (1 + 2\|x\|^2 + 2C_{t_2}^2 \int \|y\|^2 d\rho_0(y)) \exp\left(-\lambda_{2,t_2} t_2^{-1} \|x\|^2\right) dx, \end{aligned}$$

and  $\|\nabla \varrho(t, x)\| \leq C_{2,t_2} t_1^{-(d+1)/2}$ . Consequently,  $\rho$  satisfies Eq. (35), Eq. (36) and Eq. (37).  $\square$

For every  $\rho \in \mathbf{V}_2$  and every  $t > 0$ , recall the definition of the velocity field  $v_t$  in Eq. (15):  $v_t(x) := -\nabla V(x) - \int \nabla U(x, y) d\rho_t(y) - \sigma^2 \nabla \log \varrho(t, x)$ , where  $\varrho(t, x)$  is the density of  $\rho_t$  defined in Lem. 11.

**Lemma 12.** *For every  $\rho \in \mathbf{V}_2$ , and every  $t_2 > t_1 > 0$ ,*

$$\int_{t_1}^{t_2} \int \|v_t(x)\| d\rho_t(x) dt < \infty. \quad (39)$$

Moreover, for every  $\psi \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R})$ ,

$$\begin{aligned} \int \psi(t_2, x) d\rho_{t_2}(x) - \int \psi(t_1, x) d\rho_{t_1}(x) \\ = \int_{t_1}^{t_2} \int (\partial_t \psi(t, x) + \langle \nabla_x \psi(t, x), v_t(x) \rangle) \rho_t(dx) dt. \end{aligned} \quad (40)$$

*Proof.* The first point is a direct consequence of Lemma 11. Consider  $\phi \in C_c^\infty(\mathbb{R}^d, \mathbb{R})$  and  $\eta \in C_c^\infty(\mathbb{R}_+, \mathbb{R})$ . Using Eq. (9) and (10) with  $h_1 = \dots = h_r = 1$ , we obtain that for each  $\psi \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R})$  of the form  $\psi(t, x) = g(t)\phi(x)$ ,

$$\begin{aligned} \int \psi(t_2, x) d\rho_{t_2}(x) - \int \psi(t_1, x) d\rho_{t_1}(x) = \\ \int_{t_1}^{t_2} \int (\partial_t \psi(t, x) + \langle \nabla \psi(s, x), b(x, \rho_t) \rangle + \sigma^2 \Delta \psi(t, x)) \rho_t(dx) dt. \end{aligned} \quad (41)$$

As the functions of the form  $(t, x) \mapsto g(t)\phi(x)$  are dense in  $C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R})$ , Eq (41) holds in fact for any smooth compactly supported  $\psi$ . Using Lemma 11 and an integration by parts of the Laplacian term, Eq. (40) follows.  $\square$

The goal now is to establish that the functional  $\mathcal{H}$  is a Lyapunov function. This claim will follow from the application of Eq. (40) to the functional  $(t, x) \mapsto \sigma^2 \log(\varrho(t, x)) + V(x) + \int U(x-y) \varrho(t, y) dy$ . However, this function is not necessarily smooth nor compactly supported. In order to be able to apply Lem. 12, mollification should be used. In the sequel, consider two fixed positive numbers  $t_2 > t_1$ .

Define a smooth, compactly supported, even function  $\eta : \mathbb{R}^d \rightarrow \mathbb{R}_+$  such that  $\int \eta(x) dx = 1$ , and define  $\eta_\varepsilon(x) := \varepsilon^{-d} \eta(x/\varepsilon)$  for every  $\varepsilon > 0$ . For every  $t > 0$ , we introduce the density  $\varrho_\varepsilon(t, \cdot) := \eta_\varepsilon * \rho_\varepsilon(t, \cdot)$ , and we denote by  $\rho_t^\varepsilon(dx) = \varrho_\varepsilon(t, x) dx$  the corresponding probability measure. Finally, we define:

$$v_t^\varepsilon := \frac{\eta_\varepsilon * (v_t \varrho(t, \cdot))}{\varrho_\varepsilon(t, \cdot)}.$$

With these definitions at hand, it is straightforward to check that the statements of Lem. 12 hold when  $\rho_t, v_t$  are replaced by  $\rho_t^\varepsilon, v_t^\varepsilon$ . More specifically, we shall apply Eq. (40) using a specific smooth

function  $\psi = \psi_{\varepsilon, \delta, R}$ , which we will define hereafter for fixed values of  $\delta, R > 0$ , yielding our main equation:

$$\int \psi_{\varepsilon, \delta, R}(t_2, x) \varrho_{\varepsilon}(t_2, x) dx - \int \psi_{\varepsilon, \delta, R}(t_1, x) \varrho_{\varepsilon}(t_1, x) dx = \int_{t_1}^{t_2} \int (\partial_t \psi_{\varepsilon, \delta, R}(t, x) + \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^{\varepsilon}(x) \rangle) \varrho_{\varepsilon}(t, x) dx dt. \quad (42)$$

We now provide the definition of the function  $\psi_{\varepsilon, \delta, R} \in C_c^{\infty}(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R})$  used in the above equality. Let  $\theta \in C_c^{\infty}(\mathbb{R}, \mathbb{R})$  be a nonnegative function supported by the interval  $[-t_1, t_1]$  and satisfying  $\int \theta(t) dt = 1$ . For every  $\delta \in (0, 1)$ , define  $\theta_{\delta}(t) = \theta(t/\delta)/\delta$ . We define  $\varrho^{\varepsilon, \delta}(\cdot, x) := \theta_{\delta} * \varrho^{\varepsilon}(\cdot, x)$ . The map  $t \mapsto \varrho^{\varepsilon, \delta}(t, \cdot)$  is well defined on  $[t_1, t_2]$ , non negative, and smooth in both variables  $t, x$ . In addition, we define  $V_{\varepsilon} := \eta_{\varepsilon} * V$ ,  $U_{\varepsilon} := \eta_{\varepsilon} * U$ . Finally, we introduce a smooth function  $\chi$  on  $\mathbb{R}^d$  equal to one on the unit ball and to zero outside the ball of radius 2, and we define  $\chi_R(x) := \chi(x/R)$ . For every  $(t, x) \in [t_1, t_2] \times \mathbb{R}$ , we define:

$$\psi_{\varepsilon, \delta, R}(t, x) := (\sigma^2 \log \varrho^{\varepsilon, \delta}(t, x) + V_{\varepsilon}(x) + \int U_{\varepsilon}(x - y) \chi_R(y) \varrho^{\varepsilon, \delta}(t, y) dy) \chi_R(x). \quad (43)$$

We extend  $\psi_{\varepsilon, \delta, R}$  to a smooth compactly supported function on  $\mathbb{R}_+ \times \mathbb{R}^d$ , and we apply Eq. (42) to the latter. We now investigate the limit of both sides of the equality (42) as  $\delta, \varepsilon, R$  successively tend to  $0, 0, \infty$ . First consider the lefthand side. Note that for all  $t \in [t_1, t_2]$ ,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \psi_{\varepsilon, \delta, R}(t, x) \varrho_{\varepsilon}(t, x) \\ := \left( \sigma^2 \log \varrho(t, x) + V(x) + \int U(x - y) \chi_R(y) \varrho(t, y) dy \right) \varrho(t, x) \chi_R(x). \end{aligned}$$

The domination argument that allows to interchange limits and integrals is provided by Lem 11. Indeed, for a fixed  $R > 0$ , there exists a constant  $C_R$  such that  $\varrho^{\varepsilon, \delta}(t, x) \leq C_R$  and  $\psi_{\varepsilon, \delta, R}(t, x) \leq C_R$  for all  $\|x\| \leq R$  and all  $t \in [t_1, t_2]$ . As a consequence,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int \psi_{\varepsilon, \delta, R}(t, x) \varrho_{\varepsilon}(t, x) dx &= \sigma^2 \int \chi_R(x) \varrho(t, x) \log \varrho(t, x) dx + \\ &\int V(x) \chi_R(x) d\rho_t(x) + \int \int U(x - y) \chi_R(y) \chi_R(x) \varrho(t, x) \varrho(t, y) dx dy. \end{aligned}$$

Since  $\rho_t \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\int \varrho(t, x) |\log \varrho(t, x)| dx < \infty$ , and the first term in the r.h.s. of the above equation converges to  $\sigma^2 \int \varrho(t, x) \log \varrho(t, x) dx$  as  $R \rightarrow \infty$ . Similarly,  $\int V(x) \chi_R(x) d\rho_t(x)$  tends to  $\int V d\rho_t$  as  $R \rightarrow \infty$ , by use of the linear growth condition on  $\nabla V$  in Assumption 5, along with the fact that  $\rho_t$  admits a second order moment. The same holds for the last term. Finally, we have shown that, for every  $t \in [t_1, t_2]$ ,

$$\lim_{R \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int \psi_{\varepsilon, \delta, R}(t, x) \varrho_{\varepsilon}(t, x) dx = \mathcal{H}(\rho_t) + \frac{1}{2} \iint U(x - y) d\rho_t(y) d\rho_t(x),$$

recalling  $\mathcal{H}(\rho_t) := \sigma^2 \int \log \varrho(t, \cdot) d\rho_t + \int V d\rho_t + \frac{1}{2} \iint U(x - y) d\rho_t(y) d\rho_t(x)$ . As  $\delta, \varepsilon, R$  successively tend to  $0, 0, \infty$ , we have shown that the l.h.s. of Eq (42) converges to:

$$\begin{aligned} \mathcal{H}(\rho_{t_2}) - \mathcal{H}(\rho_{t_1}) \\ + \frac{1}{2} \iint U(x - y) d\rho_{t_2}(y) d\rho_{t_2}(x) - \frac{1}{2} \iint U(x - y) d\rho_{t_1}(y) d\rho_{t_1}(x). \end{aligned} \quad (44)$$

We should now identify the above term with the limit of the r.h.s. of Eq. (42) in the same regime. The latter is composed of two terms. First consider the second term:

$$\begin{aligned} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^{\varepsilon}(x) \rangle \rho_t^{\varepsilon}(dx) dt \\ = \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), \eta_{\varepsilon} * (v_t(x) \varrho(t, x)) \rangle dx dt. \end{aligned}$$

We can let  $\delta \rightarrow 0$  in this equation and interchange the limit and the integral. This is justified by Lem. 11, which implies that for every  $R > 0$ , there exists a constant  $C_R$  such that for every  $\varepsilon > 0$ ,  $\delta \in (0, 1)$ ,  $t \in [t_1, t_2]$ ,  $x \in \mathbb{R}^d$ ,

$$\|\nabla \psi_{\varepsilon, \delta, R}(t, x)\| \leq C_R. \quad (45)$$

Using Eq. (45) along with Eq. (39), the dominated convergence applies. Letting  $\varepsilon \rightarrow 0$  in a second step, the exact same argument applies, and we obtain:

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^\varepsilon(x) \rangle \varrho_\varepsilon(t, x) dx dt \\ = \int_{t_1}^{t_2} \int \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), \eta_\varepsilon * (v_t(x) \varrho(t, x)) \rangle dx dt \\ = \int_{t_1}^{t_2} \int \langle \nabla (\lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \psi_{\varepsilon, \delta, R}(t, x)), v_t(x) \rangle \varrho(t, x) dx dt, \end{aligned}$$

where the interchange between  $\nabla$  and the limits is again a consequence of Lem. 11. We now write the gradient in the above inner product. Note that:

$$\lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \psi_{\varepsilon, \delta, R}(t, x) = (\sigma^2 \log \varrho(t, x) + V(x) + \int U(x - y) \chi_R(y) \varrho(t, y) dy) \chi_R(x).$$

We obtain:

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^\varepsilon(x) \rangle \varrho_\varepsilon(t, x) dx dt = \\ - \int_{t_1}^{t_2} \int \|v_t(x)\|^2 \chi_R(x) \varrho(t, x) dx dt \\ - \int_{t_1}^{t_2} \int \langle v_t(x), \int (1 - \chi_R(y)) \nabla U(x - y) d\rho_t(y) \rangle \chi_R(x) d\rho_t(x) \\ - \int_{t_1}^{t_2} \int \langle v_t(x), \nabla \chi_R(x) (V(x) + \int U(x - y) \chi_R(y) d\rho_t(y)) \rangle d\rho_t(x). \quad (46) \end{aligned}$$

By the dominated convergence theorem, Assumption 5 and Eq. (37), the last two terms in the r.h.s. of Eq. (46) tend to zero as  $R \rightarrow \infty$ , while the first term is handled by the monotone convergence theorem. We thus obtain:

$$\begin{aligned} \lim_{R \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^\varepsilon(x) \rangle \varrho_\varepsilon(t, x) dx dt \\ = - \int_{t_1}^{t_2} \int \|v_t(x)\|^2 \varrho(t, x) dx dt. \quad (47) \end{aligned}$$

As a last step, we should evaluate the limit of the first term in the r.h.s. of Eq. (42), which writes:  $\int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon, \delta, R}(t, x) \varrho_\varepsilon(t, x) dx dt$ . Here the domination argument allowing to interchange limits and integrals requires more attention, and is justified by the following lemma, whose proof is provided at the end of the section.

**Lemma 13.** *Let  $t_2 > t_1 > 0$  be fixed. For every  $R, \varepsilon > 0$ , there exists a constant  $C_{R, \varepsilon}$  such that for every  $\delta \in (0, 1)$ ,  $t \in [t_1, t_2]$ ,  $x \in \mathbb{R}^d$ ,*

$$|\partial_t \psi_{\varepsilon, \delta, R}(t, x)| \leq C_{R, \varepsilon}, \quad (48)$$

for every  $t \leq T$ ,  $\delta > 0$ , and every  $x \in \mathbb{R}^d$ .

By Eq. (48) and by the continuity of the map  $t \mapsto \partial_t \varrho^\varepsilon$  (see the proof of Lemma 13), we can expand the first term in the r.h.s. of Eq. (42) as:

$$\int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon, \delta, R}(t, x) d\rho_t^\varepsilon(x) dt = \int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon, \delta, R}(t, x) \varrho^{\varepsilon, \delta}(t, x) dx dt + o_{\varepsilon, R}(\delta), \quad (49)$$

where  $o_{\varepsilon,R}(\delta)$  represents a term which tends to zero as  $\delta \rightarrow 0$ , for fixed values of  $\varepsilon, R$ . Note that:

$$\partial_t \psi_{\varepsilon,\delta,R}(t, x) = \sigma^2 \frac{\partial_t \varrho^{\varepsilon,\delta}(t, x)}{\varrho^{\varepsilon,\delta}(t, x)} \chi_R(x) + \int U_\varepsilon(x - y) \chi_R(y) \chi_R(x) \partial_t \varrho^{\varepsilon,\delta}(t, y) dy. \quad (50)$$

Plugging this equality into (49) and noting that  $U_\varepsilon$  is even (because  $U$  and  $\eta_\varepsilon$  are), we obtain:

$$\begin{aligned} & \int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon,\delta,R}(t, x) \varrho^{\varepsilon,\delta}(t, x) dx dt \\ &= \sigma^2 \int_{t_1}^{t_2} \int \partial_t \varrho^\varepsilon(t, x) \chi_R(x) dx dt \\ &+ \frac{1}{2} \int_{t_1}^{t_2} \iint U_\varepsilon(x - y) \partial_t (\varrho^{\varepsilon,\delta}(t, y) \varrho^{\varepsilon,\delta}(t, x)) \chi_R(x) \chi_R(y) dx dy dt \\ &= \sigma^2 \int \varrho^{\varepsilon,\delta}(t_2, x) \chi_R(x) dx - \sigma^2 \int \varrho^{\varepsilon,\delta}(t_1, x) \chi_R(x) dx \\ &+ \frac{1}{2} \int \int U_\varepsilon(x - y) \chi_R(x) \chi_R(y) \varrho^{\varepsilon,\delta}(t_2, x) \varrho^{\varepsilon,\delta}(t_2, y) dx dy \\ &- \frac{1}{2} \int \int U_\varepsilon(x - y) \chi_R(x) \chi_R(y) \varrho^{\varepsilon,\delta}(t_1, x) \varrho^{\varepsilon,\delta}(t_1, y) dx dy. \end{aligned}$$

By the dominated convergence theorem, we finally obtain:

$$\begin{aligned} \lim_{R \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon,\delta,R}(t, x) d\rho_t^\varepsilon(x) dt = \\ \frac{1}{2} \int \int U(x - y) \varrho(t_2, x) \varrho(t_2, y) dx dy - \frac{1}{2} \int \int U(x - y) \varrho(t_1, x) \varrho(t_1, y) dx dy. \end{aligned} \quad (51)$$

Putting together Eq. (44), (47) and (51), and passing to the limit in the continuity equation (42), the statement of Prop. 6 follows.

**Proof of Lem. 13.** Using Eq. (42) and integration by parts,

$$\begin{aligned} & \varrho^\varepsilon(t_2, x) - \varrho^\varepsilon(t_1, x) \\ &= \int_{t_1}^{t_2} \int \langle \nabla \eta_\varepsilon(x - y), b(y, \rho_s) \rangle d\rho_s(y) ds + \sigma^2 \int_{t_1}^{t_2} \int \Delta \eta_\varepsilon(x - y) d\rho_s(y) ds. \end{aligned}$$

Since  $\rho \in \mathcal{P}_2(\mathcal{C})$ ,  $\sup_{t \in [1, T]} \|b(y, \rho_t)\| \leq C(1 + \|y\|) + C \int \sup_{t \in [1, T]} \|x_t\| d\rho(x)$ . As a consequence,  $\sup_{t \in [1, T]} \|b(y, \rho_t)\| \leq C(1 + \|y\|)$ . Along with the observation that, for any fixed  $\varepsilon$ ,  $\nabla \eta_\varepsilon$  and  $\Delta \eta_\varepsilon$  are bounded, it follows that  $t \mapsto \varrho^\varepsilon(t, x)$  is Lipschitz continuous on  $[t_1, t_2]$ , and that its derivative almost everywhere is given by:  $\partial_t \varrho^\varepsilon(t, x) = \int (\langle \nabla \eta_\varepsilon(x - y), b(y, \rho_t) \rangle + \Delta \eta_\varepsilon(x - y)) d\rho_t(y)$ . Thus, there exists a constant  $C_\varepsilon > 0$ , such that:

$$\sup_{t \in [t_1, t_2], x \in \mathbb{R}^d} \partial_t \varrho^\varepsilon(t, x) \leq C_\varepsilon.$$

Considering the second term in the r.h.s. of Eq. (50), the presence of the product of the compactly supported functions  $\chi_R(x) \chi_R(y)$  implies that the former is bounded in absolute value:

$$\left| \int U_\varepsilon(x - y) \chi_R(y) \chi_R(x) \partial_t \varrho^{\varepsilon,\delta}(t, y) dy \right| \leq C_{R,\varepsilon}.$$

On the otherhand, using the lower bound (35), the first term in the r.h.s. of Eq. (50), is also bounded, and finally, Eq. (48) follows.

## 6.2 Proof of Prop. 7

The map  $\overline{\mathcal{H}} : \rho \mapsto \mathcal{H}(\rho_\epsilon)$  is real valued and lower semicontinuous by Prop. 6 and Fatou's lemma. Moreover, for every  $\rho \in \mathbb{V}_2$ ,  $\overline{\mathcal{H}}(\Phi_t(\rho)) - \overline{\mathcal{H}}(\rho) = \mathcal{H}(\rho_{t+\epsilon}) - \mathcal{H}(\rho_\epsilon) = - \int_\epsilon^{t+\epsilon} \int \|v_s\|^2 d\rho_s ds$ . Therefore,  $\overline{\mathcal{H}}(\Phi_t(\rho))$  is decreasing w.r.t.  $t$ , and, as such,  $\overline{\mathcal{H}}$  is a Lyapunov function. In addition,

the identity  $\overline{\mathcal{H}}(\Phi_t(\rho)) = \overline{\mathcal{H}}(\rho)$  for all  $t$ , is equivalent to:  $v_t = 0$   $\rho_t$ -a.e., for every  $t \geq \epsilon$ . By Lem. 12, this implies that  $\rho_t = \rho_\epsilon$  for all  $t \geq \epsilon$ . Thus,  $\overline{\mathcal{H}}(\Phi_t(\rho)) = \overline{\mathcal{H}}(\rho)$  for all  $t$ , if and only if  $v_\epsilon = 0$  and  $\rho_t = \rho_\epsilon$  for all  $t$ . This means that  $\overline{\mathcal{H}}$  is a Lyapunov function for the set  $\Lambda_\epsilon$ . The first point is proven.

Consider a recurrent point  $\rho \in \mathbb{V}_2$ , say  $\rho = \lim \Phi_{t_n}(\rho)$ . By Prop. 3,  $\rho \in \Lambda_\epsilon$ , for any  $\epsilon > 0$ . This means that there exists  $\mu \in \mathcal{S}$  such that  $\rho_t = \mu$  for all  $t > 0$ . By continuity of the map  $(\pi_0)_\#$ ,  $\rho_0 = \lim \rho_{t_n}$ . Thus,  $\rho_0 = \mu$ . This means that  $\rho_t = \mu$  for all  $t \geq 0$ , which writes  $\rho \in \Lambda_0$ . The proof is complete.

### 6.3 Proof of Prop. 8

Since  $\beta = 1$ , we obtain by Assumption 5 that  $\nabla U$  and  $\nabla V$  are Lipschitz continuous, therefore, the functions  $U$  and  $V$  are weakly convex. Thus, we obtain from our assumptions that the functions  $U$  and  $V$  with  $U$  being even are differentiable, weakly convex, and they satisfy the doubling assumption. In these conditions, the following facts hold true by [AGS08, Th. 11.2.8] (see also, *e.g.*, [DS10]): for each measure  $\nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ , there exists a unique function  $t \mapsto \nu_t \in \mathcal{P}_2(\mathbb{R}^d)$  that satisfies the following properties:

- i)  $\nu_t \rightarrow \nu_0$  as  $t \downarrow 0$ .
- ii)  $\sup_{t \in [0, T]} \int \|x\|^2 \nu_t(dx) < \infty$  for each  $T > 0$ .
- iii) The measure  $\nu_t$  has a density  $\eta_t = d\nu_t/d\mathcal{L}^d$  for each  $t > 0$ . This density satisfies  $\eta_t \in L^1_{\text{loc}}((0, \infty); W^{1,1}_{\text{loc}}(\mathbb{R}^d))$ .
- iv) The continuity equation

$$\partial_t \nu_t + \nabla \cdot (\nu_t w_t) = 0$$

is satisfied in the distributional sense, where

$$w_t(x) = -\frac{\sigma^2 \nabla \eta_t(x)}{\eta_t(x)} - \nabla V(x) - \int \nabla U(x-y) \eta_t(y) dy.$$

- v)  $\|w_t\|_{L^2(\nu_t)} \in L^2_{\text{loc}}(0, \infty)$ .

Furthermore, the function  $t \mapsto \nu_t$  is the solution of the gradient flow in  $\mathcal{P}_2(\mathbb{R}^d)$  of the functional  $\mathcal{H}$  provided in the statement, and  $w_t \in -\partial \mathcal{H}(\nu_t)$ , where  $\partial \mathcal{H}$  is the Fréchet sub-differential of  $\mathcal{H}$ . From the general properties of the gradient flows detailed in [AGS08, Chap. 11], one can then check that we can write  $\nu_t = \Psi_t(\nu_0)$  where  $\Psi$  is a semi-flow on  $\mathcal{P}_2(\mathbb{R}^d)$ .

With this at hand, all we have to do is to check that for each  $\rho \in \mathbb{V}_2$ , the function  $t \mapsto \rho_t$  satisfies the five properties stated above. The first two hold true for each  $\zeta \in \mathcal{P}_2(\mathcal{C})$ : to check the first one, let  $X \sim \zeta$ . Observe that  $X_t \rightarrow_{t \rightarrow 0} X_0$  by continuity and that  $\|X_t - X_0\|^2 \leq 2 \sup_{s \in [0,1]} \|X_s\|^2$  for  $t$  small, and use the Dominated Convergence. The second property follows from the very definition of  $\mathcal{P}_2(\mathcal{C})$ . Property 3 follows from Lemma 11. By Lemma 12, the continuity equation is satisfied by the function  $t \mapsto \rho_t$  with  $v_t = w_t$ , hence Property 4. Finally, Property 5 follows from Proposition 6, Equation (14). This completes the proof of Proposition 8.

### 6.4 Proof of Prop. 9

In this paragraph, Assumptions 2 and 5 hold. Therefore, Assumptions 1 and 3 also hold. Let  $k \in \mathbb{N}, n \in \mathbb{N}^*$ . We recall Eq. (1)

$$X_{k+1}^{i,n} = X_k^{i,n} - \gamma_{k+1} \nabla V(X_k^{i,n}) - \frac{\gamma_{k+1}}{n} \sum_{j \in [n]} \nabla U(X_k^{i,n} - X_k^{j,n}) + \sqrt{2\gamma_{k+1}} \xi_{k+1}^{i,n}.$$

Let us momentarily drop the superscript  $n$  to simplify the notations, and we write  $\gamma$  as a shorthand notation for  $\gamma_{k+1}$ . Note that  $\nabla U(-x) = -\nabla U(x)$ . We expand:

$$\begin{aligned} \|X_{k+1}^i\|^2 &= \|X_k^i\|^2 - 2\gamma \langle \nabla V(X_k^i), X_k^i \rangle - 2\frac{\gamma}{n} \sum_j \langle \nabla U(X_k^i - X_k^j), X_k^i \rangle \\ &\quad + 2\gamma \|\xi_{k+1}^i\|^2 + \sqrt{2\gamma} T_{k+1}^i + \gamma^2 \|\nabla V(X_k^i)\|^2 + \frac{\gamma}{n} \sum_j \nabla U(X_k^i - X_k^j) \|^2. \end{aligned}$$

where we defined:

$$T_{k+1}^i := \langle \xi_{k+1}^i, X_k^i - \gamma \nabla V(X_k^i) - \frac{\gamma}{n} \sum_j \nabla U(X_k^i - X_k^j) \rangle.$$

Using Assumption 5 and Cauchy-Schwartz inequality, there exists constants  $C, \lambda > 0$  such that:

$$\begin{aligned} \|X_{k+1}^i\|^2 &\leq (1 - \lambda\gamma + C\gamma^2) \|X_k^i\|^2 - \frac{2\gamma}{n} \sum_j \langle \nabla U(X_k^i - X_k^j), X_k^i \rangle + 2\gamma \|\xi_{k+1}^i\|^2 \\ &\quad + \sqrt{2\gamma} T_{k+1}^i + C\gamma^2 (1 + n^{-1} \sum_j \|X_k^j\|^2). \end{aligned} \quad (52)$$

Note that  $\mathbb{E}(T_{k+1}^i | \mathcal{F}_k^n) = 0$ . As a preliminar, we first establish the bound:

$$\sup_{k,n} \left( \mathbb{E}(\|X_k^{1,n}\|^2 \|X_k^{2,n}\|^2) + \frac{1}{n} \mathbb{E}(\|X_k^{1,n}\|^4) \right) < \infty. \quad (53)$$

To that end, compute the average w.r.t.  $i \in [n]$  of both sides of Eq. (52). Setting  $S_k := \frac{1}{n} \sum_i \|X_k^i\|^2$ , and

$$\begin{aligned} \chi_k^U &:= \frac{1}{n^2} \sum_i \sum_j \langle \nabla U(X_k^i - X_k^j), X_k^i \rangle \\ \chi_{k+1}^\xi &:= \frac{1}{n} \sum_i \|\xi_{k+1}^i\|^2 \\ \chi_{k+1}^T &:= \frac{1}{n} \sum_i T_{k+1}^i. \end{aligned}$$

Eq. (52) leads to, for every  $k$  larger than some fixed constant,

$$S_{k+1} \leq (1 - \lambda\gamma) S_k - 2\gamma \chi_k^U + 2\gamma \chi_{k+1}^\xi + \sqrt{2\gamma} \chi_{k+1}^T + C\gamma^2.$$

Moreover, using that  $\nabla U(X_k^j - X_k^i) = -\nabla U(X_k^i - X_k^j)$ , we obtain:

$$\chi_k^U = \frac{1}{2n^2} \sum_i \sum_j \langle \nabla U(X_k^j - X_k^i), X_k^i - X_k^j \rangle \geq -C.$$

Therefore,  $S_{k+1} \leq (1 - \lambda\gamma) S_k + 2\gamma \chi_{k+1}^\xi + \sqrt{2\gamma} \chi_{k+1}^T + C\gamma$ , for large  $k$ . Raising to the square,

$$\begin{aligned} S_{k+1}^2 &\leq (1 - \lambda\gamma)^2 S_k^2 + (2\gamma \chi_{k+1}^\xi + \sqrt{2\gamma} \chi_{k+1}^T + C\gamma)^2 \\ &\quad + CS_k(2\gamma \chi_{k+1}^\xi + \sqrt{2\gamma} \chi_{k+1}^T + C\gamma). \end{aligned}$$

Thus, for large  $k$ ,

$$\begin{aligned} \mathbb{E} S_{k+1}^2 &\leq (1 - \lambda\gamma) \mathbb{E} S_k^2 + C\gamma^2 \mathbb{E}(\chi_{k+1}^\xi)^2 \\ &\quad + C\gamma \mathbb{E}(\chi_{k+1}^T)^2 + C\gamma^2 + C\gamma \mathbb{E}(S_k \chi_{k+1}^\xi) + C\gamma \mathbb{E} S_k. \end{aligned}$$

Note that  $\mathbb{E}((\chi_{k+1}^\xi)^2)$  is bounded uniformly in  $k, n$ . Moreover, by Jensen inequality,  $\mathbb{E}(S_k) \leq \sqrt{\mathbb{E}(S_k^2)}$ . Finally, using that  $\mathbb{E}(S_k \chi_{k+1}^\xi) = d\sigma^2 \mathbb{E}(S_k)$ , we obtain:

$$\mathbb{E} S_{k+1}^2 \leq (1 - \lambda\gamma) \mathbb{E} S_k^2 + C\gamma^2 + C\gamma \mathbb{E}((\chi_{k+1}^T)^2) + C\gamma^2 + C\gamma \sqrt{\mathbb{E}(S_k^2)}.$$

We inspect the term  $\mathbb{E}((\chi_{k+1}^T)^2)$ :

$$\begin{aligned} \mathbb{E}((\chi_{k+1}^T)^2) &= \frac{\sigma^2}{n^2} \sum_i \mathbb{E} \|X_k^i - \gamma \nabla V(X_k^i) - \frac{\gamma}{n} \sum_j \nabla U(X_k^i - X_k^j)\|^2 \\ &\leq \frac{C}{n^2} \sum_i \mathbb{E} \|X_k^i\|^2 + \gamma^2 \frac{C}{n^2} \sum_i \left( \frac{1}{n} \sum_j \|\nabla U(X_k^i - X_k^j)\| \right)^2 \\ &\leq \frac{C}{n^2} \sum_i \mathbb{E} \|X_k^i\|^2 + \gamma^2 \frac{C}{n^2} \sum_i \frac{1}{n} \sum_j (1 + \|X_k^i\|^2 + \|X_k^j\|^2) \\ &\leq \frac{C}{n} S_k + \gamma^2 \frac{C}{n^2}. \end{aligned}$$

We finally obtain:  $\mathbb{E}S_{k+1}^2 \leq (1 - \lambda\gamma)\mathbb{E}S_k^2 + C\gamma^2 + C\gamma\sqrt{\mathbb{E}(S_k^2)}$ . This proves that  $\mathbb{E}S_k^2$  is bounded uniformly in  $k, n$ . By the exchangeability stated in Assumption 7,  $\mathbb{E}S_k^2 = n^{-1}\mathbb{E}(\|X_k^1\|^4) + (1 - 1/n)\mathbb{E}(\|X_k^1\|^2\|X_k^2\|^2)$ . This proves Eq. (53).

We now expand  $\|X_{k+1}^i\|^2$  starting from (52). First, we define  $\nabla U^i := n^{-1} \sum_j \nabla U(X_k^i - X_k^j)$ . For all  $k$  large enough,

$$\begin{aligned} \|X_{k+1}^i\|^4 &\leq (1 - \gamma\lambda)\|X_k^i\|^4 + (-\gamma\langle \nabla U^i, X_k^i \rangle + 2\gamma\|\xi_{k+1}^i\|^2 \\ &\quad + \sqrt{2\gamma}T_{k+1}^i + C\gamma^2(1 + S_k))^2 - 2\gamma\langle \nabla U^i, X_k^i \rangle\|X_k^i\|^2 \\ &\quad + 2\gamma\|\xi_{k+1}^i\|^2\|X_k^i\|^2 + \sqrt{2\gamma}T_{k+1}^i\|X_k^i\|^2 + C\gamma^2\|X_k^i\|^2(1 + S_k). \end{aligned}$$

We take expectations in the above equation. Remark the following  $\mathbb{E}\|\xi_{k+1}^i\|^4 \leq C$ ,  $\mathbb{E}S_k^2 \leq C$  and  $\mathbb{E}(\|\xi_{k+1}^i\|^2\|X_k^i\|^2) \leq C\mathbb{E}\|X_k^i\|^2 = C\mathbb{E}(S_k) \leq C$  (where as usual,  $C$  changes at each inequality). Thus

$$\begin{aligned} \mathbb{E}\|X_{k+1}^i\|^4 &\leq (1 - \gamma\lambda)\mathbb{E}\|X_k^i\|^4 + C\gamma^2\mathbb{E}\langle \nabla U^i, X_k^i \rangle^2 + C\gamma\mathbb{E}((T_{k+1}^i)^2) + C\gamma \\ &\quad - 2\gamma\mathbb{E}\langle \nabla U^i, X_k^i \rangle\|X_k^i\|^2 + C\gamma^2\mathbb{E}(\|X_k^i\|^2 S_k). \end{aligned}$$

It is straightforward to show that  $\mathbb{E}((T_{k+1}^i)^2) \leq C(1 + \gamma^2\mathbb{E}(\|\nabla U^i\|^2))$  and that, in turn,  $\mathbb{E}(\|\nabla U^i\|^2) \leq C(1 + \mathbb{E}S_k) \leq C$ . Thus,  $\mathbb{E}((T_{k+1}^i)^2) \leq C$ . Moreover, by Cauchy-Schwartz inequality followed by the triangular inequality,

$$\begin{aligned} \mathbb{E}\langle \nabla U^i, X_k^i \rangle^2 &\leq \mathbb{E}\left(\frac{1}{n} \sum_j \|\nabla U(X_k^i - X_k^j)\|^2\|X_k^i\|^2\right) \\ &\leq C\mathbb{E}\left(\frac{1}{n} \sum_j (1 + \|X_k^i\|^2 + \|X_k^j\|^2)\|X_k^i\|^2\right) \\ &\leq C + C\mathbb{E}\|X_k^1\|^4 + C\mathbb{E}\|X_k^1\|^2\|X_k^2\|^2. \end{aligned}$$

Note that the last term in the above inequality is bounded uniformly in  $k, n$ , by Eq. (53). Changing again the constants  $C, \lambda$ , we obtain that for large  $k$ ,

$$\mathbb{E}\|X_{k+1}^i\|^4 \leq (1 - \gamma\lambda)\mathbb{E}\|X_k^i\|^4 - 2\gamma\mathbb{E}\langle \nabla U^i, X_k^i \rangle\|X_k^i\|^2 + C\gamma.$$

The crux is to estimate the term  $-2\gamma\mathbb{E}\langle \nabla U^i, X_k^i \rangle\|X_k^i\|^2$ . By exchangeability, and using that  $\nabla U(0) = 0$ ,

$$\begin{aligned} \mathbb{E}\langle \nabla U^i, X_k^i \rangle\|X_k^i\|^2 &= \frac{1}{n} \sum_{j \neq i} \mathbb{E}\langle \nabla U(X_k^i - X_k^j), X_k^i \rangle\|X_k^i\|^2 \\ &= \left(\frac{n-1}{n}\right) \mathbb{E}\langle \nabla U(X_k^1 - X_k^2), X_k^1 \rangle\|X_k^1\|^2. \end{aligned}$$

Moreover, using that  $\langle \nabla U(X_k^1 - X_k^2), X_k^1 - X_k^2 \rangle \geq -C$ ,

$$\begin{aligned} &\mathbb{E}\langle \nabla U(X_k^1 - X_k^2), X_k^1 \rangle\|X_k^1\|^2 \\ &\geq -C\mathbb{E}\|X_k^1\|^2 + \mathbb{E}[\langle \nabla U(X_k^1 - X_k^2), X_k^2 \rangle\|X_k^1\|^2] \\ &\geq -C\mathbb{E}\|X_k^1\|^2 - \mathbb{E}[\|\nabla U(X_k^1 - X_k^2)\|X_k^2\|\|X_k^1\|^2] \\ &\geq -C\mathbb{E}\|X_k^1\|^2 - C\mathbb{E}[(1 + \|X_k^1\| + \|X_k^2\|)\|X_k^2\|\|X_k^1\|^2] \\ &\geq -C(\mathbb{E}\|X_k^1\|^2 + \mathbb{E}[\|X_k^2\|\|X_k^1\|^3] + \mathbb{E}[\|X_k^2\|^2\|X_k^1\|^2] + 1) \\ &\geq -C\mathbb{E}[\|X_k^2\|\|X_k^1\|^3] - C, \end{aligned}$$

where we used the fact, proven above, that  $\mathbb{E}\|X_k^1\|^2$  and  $\mathbb{E}[\|X_k^2\|^2\|X_k^1\|^2]$  are bounded, uniformly in  $n, k$ . The term  $\mathbb{E}[\|X_k^2\|\|X_k^1\|^3]$  can be handled by Cauchy-Schwartz inequality:

$$\mathbb{E}[\|X_k^2\|\|X_k^1\|^3] \leq \mathbb{E}(\|X_k^2\|^2\|X_k^1\|^2)^{\frac{1}{2}} \mathbb{E}(\|X_k^1\|^4)^{\frac{1}{2}} \leq C\mathbb{E}(\|X_k^1\|^4)^{\frac{1}{2}}.$$

We have shown that:

$$\mathbb{E}\langle \nabla U^i, X_k^i \rangle\|X_k^i\|^2 \geq -C\mathbb{E}(\|X_k^1\|^4)^{\frac{1}{2}} - C.$$

Putting all pieces together,

$$\mathbb{E}\|X_{k+1}^1\|^4 \leq (1 - \gamma\lambda)\mathbb{E}\|X_k^1\|^4 + C\gamma\sqrt{\mathbb{E}(\|X_k^1\|^4)} + C\gamma.$$

This proves that  $\mathbb{E}\|X_k^1\|^4$  is bounded, uniformly in  $k, n$ . The proof is complete.

## 6.5 Proof of Th. 4

The convergence provided in the statement follows at once from Proposition 8 and Theorem 2. We need to prove that  $\mathcal{S} = A_2$  when  $A_2 = \{\rho_\infty\}$ . For an absolutely continuous probability measure  $d\nu(x) = \eta(x)dx \in \mathcal{P}_2(\mathbb{R}^d)$  with  $\eta \in C^1(\mathbb{R}^d, \mathbb{R})$ , write

$$u_\nu(x) = -\nabla V(x) - \int \nabla U(x-y)\eta(y)dy - \sigma^2 \nabla \log \eta(x).$$

With this at hand, using Equation (14) in conjunction with the identity  $\rho_\infty = \Psi_t(\rho_\infty)$  for each  $t \geq 0$  shows that  $u_{\rho_\infty}(x) = 0$  for  $\rho_\infty$ -almost all  $x$ . This shows that  $\rho_\infty \in \mathcal{S}$ . On the other hand, for  $\nu \neq \rho_\infty$  in  $\mathcal{P}_2(\mathbb{R}^d)$ , we obtain from Equation (14) that the function  $t \mapsto \mathcal{H}(\Psi_t(\nu))$  is strictly decreasing. Thus,  $\int \|u_\nu\|^2 d\nu > 0$  which shows that  $\nu \notin \mathcal{S}$ .

## A Technical proofs

### A.1 Proof of Proposition 1

Let  $I \subset \mathbb{R}$ , we denote by  $C(I, \mathbb{R}^d)$  the set of continuous function from  $I$  to  $\mathbb{R}^d$ . One can show, that  $(\rho_n)$  is a Cauchy sequence in the complete space  $(\mathcal{P}_p(C([0, k], \mathbb{R}^d)), W_p)$ . Thus, there exists a sequence of compact sets  $(K_k)$  in  $C([0, k], \mathbb{R}^d)$  such that:

$$(\pi_{[0, k]})_\# \rho_n(K_k) > 1 - \frac{\varepsilon}{2^k},$$

for all  $k \in \mathbb{N}^*$ . Let  $\mathcal{K} := \bigcap_{k \geq 1} \pi_{[0, k]}^{-1}(K_k) \subset \mathcal{C}$ . The union bound yields  $\rho_n(\mathcal{K}) > 1 - \varepsilon$ . Referring to [Bou89, Theorem 2, Section X, Chapter 5],  $\mathcal{K}$  has a compact closure in  $\mathcal{C}$ . Hence, there exists a converging subsequence  $(\rho_{\varphi_n})$  converging to  $\rho \in \mathcal{P}(\mathcal{C})$ . Following the proof of [Vil09, Theorem 6.18], one can readily check that  $\lim_{n \rightarrow \infty} W_p((\pi_{[0, k]})_\# \rho_n, (\pi_{[0, k]})_\# \rho) = 0$ , for every  $k$ . Consequently,  $\lim_{n \rightarrow \infty} W_p(\rho_n, \rho) = 0$ , which means the completeness of  $\mathcal{P}_p(\mathcal{C})$ . It remains to obtain its separability.

As  $\mathcal{C}$  is Polish, there exists a dense sequence  $(x_n)$  in  $\mathcal{C}$ . Following the proof of [Vil09, Theorem 6.18], one can construct a sequence  $(\rho_n)$  in  $\mathcal{P}_p(\mathcal{C})$  from  $(x_n)$ , such that  $((\pi_{[0, k]})_\# \rho_n)$  is dense in  $C([0, k], \mathbb{R}^d)$  for every  $k$ . With this result, it can be verified that  $(\rho_n)$  is dense in  $\mathcal{P}_p(\mathcal{C})$ .

### A.2 Proof of Lemma 2

Since Prop. 1 holds,  $(\mathbb{I}(\rho_n))$  is a weak $\star$ -relatively compact sequence in  $\mathcal{P}(\mathcal{C})$ , and there exists a sequence of compact sets  $(K_k)$  in  $\mathcal{C}$ , such that

$$\mathbb{I}(\rho_n)(K_k) > 1 - \frac{1}{k2^k},$$

for every  $k \in \mathbb{N}^*$  and every  $n \in \mathbb{N}^*$ . Let  $\varepsilon > 0$ . We define the relatively compact set in  $\mathcal{P}(\mathcal{C})$ :

$$\mathcal{K}_\varepsilon := \left\{ \rho \in \mathcal{P}(\mathcal{C}) : \rho(K_k) > 1 - \frac{1}{k\varepsilon}, \text{ for every } k \in \mathbb{N}^*, \text{ such that } k\varepsilon > 1 \right\}.$$

The union bound and Markov's inequality yields:

$$\mathbb{P}(\rho_n \in \mathcal{K}_\varepsilon) > 1 - \varepsilon \tag{54}$$

for every  $n \in \mathbb{N}^*$ .

To be relatively compact in  $\mathcal{P}_p(\mathcal{C})$ , the set  $\mathcal{K}_\varepsilon$  must satisfy Eq. (p-UI). Since the sequence  $(\mathbb{I}(\rho_n))$  has uniformly integrable  $p$ -moments, there exists a sequence  $(a_{k,l})_{(k,l) \in (\mathbb{N}^*)^2}$ , such that for every  $l \in \mathbb{N}^*$ ,  $\lim_{k \rightarrow \infty} a_{k,l} = \infty$ , and

$$\forall (k, l) \in (\mathbb{N}^*)^2, \sup_{n \in \mathbb{N}^*} \mathbb{E} \left[ \int \sup_{t \in [0, l]} \|x_t\|^p \mathbf{1}_{\sup_{t \in [0, l]} \|x_t\| > a_{k,l}} d\rho_n(x) \right] \leq \frac{1}{kl2^{k+l}}.$$

For  $\varepsilon > 0$ , we define a set that satisfies Eq. (p-UI):

$$\mathcal{U}_\varepsilon := \left\{ \rho \in \mathcal{P}_p(\mathcal{C}) : \int \sup_{t \in [0, l]} \|x_t\|^p \mathbf{1}_{\sup_{t \in [0, l]} \|x_t\| > a_{k,l}} d\rho(x) \leq \frac{1}{\varepsilon kl}, k, l \in \mathbb{N}^* \right\}.$$

Using Markov's inequality and the union bound, we obtain

$$\mathbb{P}(\rho_n \in \mathcal{U}_\varepsilon) > 1 - \varepsilon. \quad (55)$$

Putting together Eq. (54) and Eq. (55),

$$\mathbb{P}(\rho_n \in \mathcal{K}_\varepsilon \cap \mathcal{U}_\varepsilon) > 1 - 2\varepsilon.$$

$\mathcal{K}_\varepsilon \cap \mathcal{U}_\varepsilon$  is a relatively compact set in  $\mathcal{P}_p(\mathcal{C})$ . Thus,  $(\rho_n)$  is tight in  $\mathcal{P}_p(\mathcal{C})$ .

### A.3 Proof of Lemma 3

Given  $G = G_{r,\phi,h_1,\dots,h_r,t,s,v_1,\dots,v_r} \in \mathcal{G}_p$ , we first want to show that  $G(\rho_n) \rightarrow G(\rho_\infty)$  as  $\rho_n \rightarrow \rho_\infty$  in  $\mathcal{P}_p(\mathcal{C})$ . This last convergence is characterized by the fact that  $\rho_n \rightarrow \rho_\infty$  in  $\mathcal{P}(\mathcal{C})$ , and that the sequence  $(\rho_n)$  has uniformly integrable  $p$ -moments as shown by ( $p$ -UI), which is written here as

$$\forall T > 0, \quad \lim_{a \rightarrow \infty} \sup_{n \in \mathbb{N}} \int \mathbf{1}_{\sup_{u \in [0,T]} \|y_u\| > a} \left( \sup_{u \in [0,T]} \|y_u\|^p \right) d\rho_n(y) = 0.$$

We write  $G(\rho_n) = \int g(x, y, z) d(\rho_n \otimes \rho_n \otimes \rho_n)(x, y, z)$ , where for  $x, y, z$  in  $\mathcal{C}$ :

$$g(x, y, z) := (\phi(x_t) - \phi(x_s) - \int_s^t (\langle \nabla \phi(x_u), b(x_u, y_u) \rangle + \sigma(x_u, y_u)^T H_\phi(x_u) \sigma(x_u, z_u)) du) h(x),$$

and  $h(x) := \prod_{j=1}^r h_j(x_{t_j})$ . Using Cauchy-Schwartz inequality, we state a useful inequality:

$$|g(x, y, z)| \leq C \left( 1 + \int_s^t \|b(x_u, y_u)\| du \right), \quad (56)$$

where  $C = \|h\|_\infty \max \left( 2\|\phi\|_\infty + \|\sigma\|_\infty^2 (t-s) \|\Delta\phi\|_\infty, \|\nabla\phi\|_\infty \right)$ . Note that  $\rho_n \otimes \rho_n \otimes \rho_n \rightarrow \rho_\infty \otimes \rho_\infty \otimes \rho_\infty$  in  $\mathcal{P}(\mathcal{C} \times \mathcal{C} \times \mathcal{C})$ . Furthermore, using the bound (56) for our function  $g$ , and observing that  $t$  is the maximum of the time snapshots intervening in the definition of  $g$ , we have for each  $a > 0$

$$\begin{aligned} \sup_{n \in \mathbb{N}} \int \mathbf{1}_{|g(x,y,z)| > a} |g(x, y, z)| d(\rho_n \otimes \rho_n \otimes \rho_n)(x, y, z) \\ \leq \sup_{n \in \mathbb{N}} \int \mathbf{1}_{C \left( 1 + t \sup_{u \in [0,t]} \|y_u\| \right) > a} C^p \left( 1 + t \sup_{u \in [0,t]} \|y_u\| \right)^p d\rho_n(y), \end{aligned}$$

therefore,

$$\lim_{a \rightarrow \infty} \sup_{n \in \mathbb{N}} \int \mathbf{1}_{|g(x,y,z)| > a} |g(x, y, z)| d(\rho_n \otimes \rho_n \otimes \rho_n)(x, y, z) = 0,$$

which shows that  $G(\rho_n) \rightarrow G(\rho)$  by uniform integrability, and the first result of the lemma is established.

## References

- [AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [BCEM24] M. Benko, I. Chlebicka, J. Endal, and B. Miasojedow. Convergence rates of particle approximation of forward-backward splitting algorithm for granular medium equations, 2024.
- [BDFR15] A. Budhiraja, P. Dupuis, M. Fischer, and K. Ramanan. Limits of relative entropies associated with weakly interacting particle systems. *Electronic Journal of Probability*, 20, 2015.

- [BDG72] D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 223–240. Univ. California Press Berkeley, Calif., 1972.
- [Ben99] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.
- [BGG13] F. Bolley, I. Gentil, and A. Guillin. Uniform convergence to equilibrium for granular media. *Archive for Rational Mechanics and Analysis*, 208:429–445, 2013.
- [Bil99] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [Bou89] N. Bourbaki. *General topology. Chapters 5–10*. Elements of Mathematics (Berlin). Springer-Verlag, Berlin, 1989. Translated from the French, Reprint of the 1966 edition.
- [BS00] M. Benaïm and S. J. Schreiber. Ergodic properties of weak asymptotic pseudotrajectories for semiflows. *Journal of Dynamics and Differential Equations*, 12:579–598, 2000.
- [CB18] L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. *Advances in neural information processing systems*, 31, 2018.
- [CD22] L.-P. Chaintron and A. Diez. Propagation of chaos: A review of models, methods and applications. i. models and methods. *Kinetic and Related Models*, 15(6):895, 2022.
- [CGM08] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability theory and related fields*, 140:19–40, 2008.
- [Chi22] L. Chizat. Mean-field langevin dynamics: Exponential convergence and annealing. *Transactions on Machine Learning Research*, 2022.
- [CLRW24] F. Chen, Y. Lin, Z. Ren, and S. Wang. Uniform-in-time propagation of chaos for kinetic mean field langevin dynamics. *Electronic Journal of Probability*, 29:1–43, 2024.
- [CMV03] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matemática Iberoamericana*, 19(3):971–1018, 2003.
- [CMV06] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, 179:217–263, 2006.
- [Cor23] Q. Cormier. On the stability of the invariant probability measures of mckean-vlasov equations. *arXiv preprint arXiv:2201.11612*, 2023.
- [CRR18] L. Carratino, A. Rudi, and L. Rosasco. Learning with sgd and random features. *Advances in neural information processing systems*, 31, 2018.
- [DEGZ20] A. Durmus, A. Eberle, A. Guillin, and R. Zimmer. An elementary approach to uniform in time propagation of chaos. *Proceedings of the American Mathematical Society*, 148(12):5387–5398, 2020.
- [DMT19] P. Del Moral and J. Tugaut. Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions. *Stochastic Analysis and Applications*, 37(6):909–935, 2019.
- [DS10] S. Daneri and G. Savaré. Lecture notes on gradient flows and optimal transport. *arXiv preprint arXiv:1009.3737*, 2010.

- [ELL21] X. Erny, E. Löcherbach, and D. Loukianova. Conditional propagation of chaos for mean field systems of interacting neurons. *Electronic Journal of Probability*, 26(none):1 – 25, 2021.
- [GLWZ22] A. Guillin, W. Liu, L. Wu, and C. Zhang. Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems. *The Annals of Applied Probability*, 32(3):1590 – 1614, 2022.
- [HRŠS21] K. Hu, Z. Ren, D. Šiška, and L. Szpruch. Mean-field langevin dynamics and energy landscape of neural networks. In *Annales de l’Institut Henri Poincaré (B) Probabilités et statistiques*, volume 57, pages 2043–2065. Institut Henri Poincaré, 2021.
- [KJHK24] MR. Karimi Jaghargh, YP. Hsieh, and A. Krause. Stochastic approximation algorithms for systems of interacting particles. *Advances in Neural Information Processing Systems*, 36, 2024.
- [Leo23] G. Leoni. *A first course in fractional Sobolev spaces*, volume 229. American Mathematical Society, 2023.
- [LLF23] D. Lacker and L. Le Flem. Sharp uniform-in-time propagation of chaos. *Probability Theory and Related Fields*, 187(1-2):443–480, 2023.
- [LW16] Q. Liu and D. Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in Neural Information Processing Systems*, 2016.
- [Mal01] F. Malrieu. Logarithmic sobolev inequalities for some nonlinear pde’s. *Stochastic processes and their applications*, 95(1):109–132, 2001.
- [Mal03] F. Malrieu. Convergence to equilibrium for granular media equations and their euler schemes. *The Annals of Applied Probability*, 13(2):540–560, 2003.
- [Mañ87] R. Mañé. *Ergodic Theory and Differentiable Dynamics*. Ergebnisse der Mathematik und ihrer Grenzgebiete : a series of modern surveys in mathematics. Folge 3. Springer-Verlag, 1987.
- [MMN18] S. Mei, A. Montanari, and P.-M. Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33):E7665–E7671, 2018.
- [MPZ21] S. Menozzi, A. Pesce, and X. Zhang. Density and gradient estimates for non degenerate brownian sdes with unbounded measurable drift. *Journal of Differential Equations*, 272:330–369, 2021.
- [MRC87] S. Méléard and S. Roelly-Coppoletta. A propagation of chaos result for a system of particles with moderate interaction. *Stochastic Processes and their Applications*, 26:317–332, 1987.
- [MRW24] P. Monmarché, Z. Ren, and S. Wang. Time-uniform log-sobolev inequalities and applications to propagation of chaos. *arXiv preprint arXiv:2401.07966*, 2024.
- [NWS22] A. Nitanda, D. Wu, and T. Suzuki. Convex analysis of the mean field langevin dynamics. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 9741–9757. PMLR, 28–30 Mar 2022.
- [Oel84] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Annals of Probability*, 12:458–479, 1984.
- [RR07] A. Rahimi and B. Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.
- [RVE22] G. Rotskoff and E. Vanden-Eijnden. Trainability and accuracy of artificial neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, 75(9):1889–1935, 2022.

- [SS20] J. Sirignano and K. Spiliopoulos. Mean field analysis of neural networks: A central limit theorem. *Stochastic Processes and their Applications*, 130(3):1820–1852, 2020.
- [SSR22] A. Salim, L. Sun, and P. Richtarik. A convergence theory for svgd in the population limit under talagrand’s inequality t1. In *International Conference on Machine Learning*, pages 19139–19152. PMLR, 2022.
- [Szn84] A.-S. Sznitman. Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *Journal of functional analysis*, 56(3):311–336, 1984.
- [Tug13] J. Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *The Annals of Probability*, 41(3A):1427 – 1460, 2013.
- [Ver06] A. Y. Veretennikov. On ergodic measures for mckean-vlasov stochastic equations. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 471–486. Springer, 2006.
- [Vil06] C. Villani. Mathematics of granular materials. *Journal of statistical physics*, 124(2-4):781–822, 2006.
- [Vil09] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.