



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

Pascal Bianchi <sup>a</sup>, Walid Hachem <sup>b</sup>, Victor Priser <sup>a,\*</sup>

<sup>a</sup> LTCI, Télécom Paris, France

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

## ARTICLE INFO

### MSC:

primary 35B40  
65C35  
secondary 35K55  
37M25  
60H35  
82C22

### Keywords:

Particle systems  
Ergodic convergence  
McKean–Vlasov equation  
Discrete-time  
Granular media equation

## ABSTRACT

We consider a discrete-time system of  $n$  coupled random vectors, a.k.a. interacting particles. The dynamics involve 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 McKean–Vlasov distributions. We also consider the pointwise convergence of the empirical measures of the particles. We consider the example of the granular media equation, where the particles are shown to converge to a critical point of the Helmholtz energy.

## 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:

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

for each  $i \in [n]$  where  $[n] := \{1, \dots, n\}$ , where

$$\mu_k^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_k^{i,n}}. \quad (2)$$

In this equation,  $b : \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathbb{R}^d$  is a continuous vector field, where, for some  $p \in [1, 2]$ ,  $\mathcal{P}_p(\mathbb{R}^d)$  is the space of probability measures with a finite  $p$ th order, equipped with the Wasserstein distance. Moreover,  $(\gamma_k)_k$  is a vanishing sequence of deterministic positive step sizes,  $((\xi_k^{i,n})_{i \in [n]})_{k \in \mathbb{N}^*}$  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. We also assume the exchangeability of 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:

$$\mathcal{F}_k^n := \sigma((X_0^{i,n})_{i \in [n]}, ((\xi_\ell^{i,n})_{i \in [n]})_{\ell \leq k}, ((\zeta_\ell^{i,n})_{i \in [n]})_{\ell \leq k}), \quad (3)$$

\* Corresponding author.

E-mail address: [victor.priser@telecom-paris.fr](mailto:victor.priser@telecom-paris.fr) (V. Priser).

we assume that for each  $n$ , the sequence  $((\xi_k^{i,n})_{i \in [n]})_k$  is a  $(F_k^n)_k$ -martingale increment sequence i.e.,  $\mathbb{E}(\xi_{k+1}^{i,n} | F_k^n) = 0$ . Finally, we assume that

$$\mathbb{E}(\xi_{k+1}^{i,n} (\xi_{k+1}^{j,n})^T | F_k^n) = \sigma(X_k^{i,n}, \mu_k^n) \sigma(X_k^{j,n}, \mu_k^n)^T \mathbb{1}_{i=j}$$

for some  $\sigma : \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times d'}$ , with  $d' > 0$ .

The aim of the paper is to characterize the asymptotic behavior of the empirical measure of the particles  $\mu_k^n$  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  $((\xi_k^{i,n})_{i \in [n]})_k$ , 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  $\phi$  and  $\text{tr}$  denotes the Trace operator.

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:

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

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  $\mathcal{A}$  of McKean–Vlasov distributions which satisfy the property that  $t \mapsto F(\Phi_t(\rho))$  is constant whenever  $\rho \in \mathcal{A}$ .

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

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

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  $\mathcal{P}_2(C)$  and corresponds to the scenario where  $\sigma(x, \mu) = \sigma I_d$  for some real constant  $\sigma \geq 0$ , and with a slight abuse of notation the vector field  $b$  takes the form  $b(x, \mu) = \int b(x, y) d\mu(y)$ , with:

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

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 is provided by the *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  $S$  of critical points of the Helmholtz energy, namely:

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

where, this time,  $W_2$  represents the classical Wasserstein distance, and where  $S$  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 [1]. Such results have broad applicability across a variety of particle systems, where the interacting term  $b$  can manifest in various forms [2–5]. 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 [6] 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 [6], 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 [7], which requires the strong convexity of the confinement potential and the convexity of the interaction potential. Later, [8] relaxed the strong convexity assumption on the confinement potential. [9] 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, [10–12] 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 [13], achieving uniform propagation of chaos over time is only possible when a unique McKean–Vlasov stationary distribution exists. A condition that [14] 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, [15] explores the stability of stationary distributions. Additionally, [16] 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, [17] employs an implicit Euler scheme for the granular media case, assuming a zero potential function and strongly convex interaction. The work in [18] studies a Jordan–Kinderlehrer–Otto (JKO) scheme for granular media, assuming a strongly convex confinement potential. The contribution of [19] 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, [20] 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 [21]. 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 [22]. In Markov Chain Monte Carlo theory, the Stein Variational Gradient Descent estimates a target distribution using a particle system [23,24], 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 [25]. Here, the authors establish convergence to gradient descent in continuous time and in the double asymptotic regime. The paper [26] establishes the convergence of noisy stochastic gradient descent when the number of iterations depends on the number of particles. See also [27–31] 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  $C$  the set of the  $[0, \infty) \rightarrow \mathbb{R}^d$  continuous functions. It is well-known that the space  $C$  endowed with the topology of the uniform convergence on the compact intervals of  $[0, \infty)$  is a Polish space.

We denote by  $\text{conv}(A)$  the convex hull of a set  $A$ .

#### 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  $\rho \in \mathcal{P}_p(C)$ , 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\*-relatively compact in  $\mathcal{P}(E)$  i.e., has a weak\* 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  $U_t^n \xrightarrow[(t,n) \rightarrow (\infty, \infty)]{\mathbb{P}} U$ . When  $(U_t^n)$  is deterministic, we write  $U_t^n \xrightarrow[(t,n) \rightarrow (\infty, \infty)]{ } U$ . 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:

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

We define  $\limsup_{(t,n) \rightarrow (\infty, \infty)} U_t^n := \inf_{t \in \mathbb{T}, n \in \mathbb{N}, s \geq t, k \geq n} \sup U_s^k$ . 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  $d(x_\alpha, \mathcal{A}) \rightarrow_a 0$ .

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\* 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\* 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** ([2]). 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\*-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

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

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

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

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

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

In the same way, a sequence  $(\rho^n)$  has uniformly integrable  $p$ -moments if the condition (p-UI) holds for the sequence  $(\rho^n)$  in place of  $\mathcal{A}$ . Following the same lines as [32, Th. 6.18] and [22, Prop. 7.1.5], we obtain the following lemma. The proof is provided in Appendix A.1.

**Proposition 1.**

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

Finally, we will also consider  $\mathcal{P}_p(C)$ -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  $\mathcal{P}_p(C)$ . Assume that  $(\mathbb{I}(\rho^n))$  is relatively compact in  $\mathcal{P}_p(C)$ . Then,  $(\rho^n)$  is tight in  $\mathcal{P}_p(C)$ .

### 2.3. Spaces of McKean–Vlasov measures

Let  $d' \in \mathbb{N}^*$ . Consider a matrix-valued function  $\sigma : \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times d'}$  and a vector field  $b : \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathbb{R}^d$  satisfying the following assumptions:

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

$$\|b(x, \mu)\| \leq C(1 + \|x\| + \int \|y\| d\mu(y)),$$

and  $\|\sigma(x, \mu)\| \leq C$ .

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

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

where  $H_\phi$  is the hessian matrix of  $\phi$ . Let  $(X_t : t \in [0, \infty))$  be the canonical process on  $C$ . 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  $\rho \in \mathcal{P}_p(C)$  belongs to the class  $\mathcal{V}_p$  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  $(C, \mathcal{B}(C), \rho)$ .

The elements of  $\mathbb{V}_p$  will be referred to as McKean–Vlasov distributions of order  $p$ . In the sequel, it will be convenient to work with the following equivalent characterization. The martingale property implies that every measure  $\rho \in \mathbb{V}_p$  satisfies  $G(\rho) = 0$ , for every function  $G : \mathcal{P}_p(C) \rightarrow \mathbb{R}$  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  $r \in \mathbb{N}$ ,  $\phi \in C_c^2(\mathbb{R}^d, \mathbb{R})$ ,  $h_1, \dots, h_r \in C_c(\mathbb{R}^d, \mathbb{R})$ ,  $0 \leq v_1 \leq \dots \leq v_r \leq s \leq t$ , are arbitrary. We denote by  $\mathcal{G}_p$  the set of such mappings  $G$ . [Assumption 1](#) ensures that these mappings are well defined. By [Definition 1](#), every  $\rho \in \mathbb{V}_p$  is a root of all  $G \in \mathcal{G}_p$ . As a matter of fact, a measure  $\rho \in \mathcal{P}_p(C)$  belongs to the set  $\mathbb{V}_p$ , if and only if  $G(\rho) = 0$  for every  $G$  of the form (9). In other words, [Definition 1](#) is equivalent to the following identity:

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

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  $\mathcal{P}_p(C) \rightarrow \mathbb{R}$ .*

The following result is a consequence of [Lemma 3](#) and [Proposition 1](#).

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

**Proof.** For all  $\rho^n \in \mathbb{V}_p \rightarrow \rho^\infty$  in  $\mathcal{P}_p(C)$ , it holds by [Lemma 3](#) that  $G(\rho^\infty) = 0$  for all  $G \in \mathcal{G}_p$ , which shows that  $\rho^\infty \in \mathbb{V}_p$  by (10). Hence,  $\mathbb{V}_p$  is closed. A closed subset of a Polish space is also Polish. By [Proposition 1](#),  $\mathbb{V}_p$  is Polish.  $\square$

## 2.4. Dynamical systems

Recall the definition of the shift  $\Theta_t(x) = x_{t+}$ , defined on  $C$ . Let us equip the space  $\mathbb{V}_p$  assumed nonempty with the trace topology of  $\mathcal{P}_p(C)$ , making it a Polish space (see [Proposition 2](#)). With this at hand, one can readily check that the function  $\Phi : [0, \infty) \times \mathbb{V}_p \rightarrow \mathbb{V}_p$  defined as  $(t, \rho) \mapsto \Phi_t(\rho) = (\Theta_t)_\# \rho$  is a semi-flow on the space  $(\mathbb{V}_p, W_p)$ , 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 [\[33\]](#) for an exposition of the concepts related to semi-flows. The omega limit set of  $\rho \in \mathbb{V}_p$  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  $\rho \in \mathbb{V}_p$  is called recurrent if  $\rho \in \omega(\rho)$ . The Birkhoff center  $BC_p$  is defined as the closure of the set of recurrent points:

$$BC_p := \overline{\{\rho \in \mathbb{V}_p : \rho \in \omega(\rho)\}}.$$

By extension, given a measure  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ , we say that  $\mu$  is a *recurrent marginal McKean–Vlasov measure* if there exists a recurrent measure  $\rho \in \mathbb{V}_p$  such that  $\rho_0 = \mu$ . We denote by  $BC_p^0$  the closure of recurrent marginal McKean–Vlasov measures, that is,

$$BC_p^0 = \overline{\{\rho_0 : \rho \in \mathbb{V}_p, \rho \in \omega(\rho)\}}, \quad (11)$$

or in short,  $BC_p^0 = (\pi_0)_\#(BC_p)$ .

**Definition 2.** Consider the semi-flow  $\Phi$  and a non-empty set  $\Lambda \subset \mathbb{V}_p$ . A lower semi-continuous function  $F : \mathbb{V}_p \rightarrow \mathbb{R}$  is called a Lyapunov function for the set  $\Lambda$  if, for every  $\rho \in \mathbb{V}_p$  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  $\rho \in \mathbb{V}_p$ , 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. (3) 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

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

Recall the definition  $\mu_k^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_k^{i,n}}$ .

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

- (i) For each  $n$ ,  $((X_0^{i,n}, (\xi_k^{i,n})_{k \in \mathbb{N}}, (\zeta_k^{i,n})_{k \in \mathbb{N}}))_{i \in [n]}$  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  $\sup_{k,n} \mathbb{E} \|\xi_k^{1,n}\|^4 < \infty$ . Moreover, for each  $n > 0$ , and each  $i, j$ ,

$$\begin{aligned} \mathbb{E} \left[ \xi_{k+1}^{1,n} \middle| \mathcal{F}_k^n \right] &= 0 \\ \mathbb{E} \left[ \xi_{k+1}^{i,n} \left( \xi_{k+1}^{j,n} \right)^T \middle| \mathcal{F}_k^n \right] &= \sigma(X_k^{i,n}, \mu_k^n) \sigma(X_k^{j,n}, \mu_k^n)^T \mathbb{1}_{i=j}, \end{aligned}$$

- (iii) For each  $k$ , and each  $n$ ,  $\mathbb{E} \|\xi_k^{1,n}\| < \infty$ , and

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

**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  $k$ .

Define  $\kappa_k(\mu) = \int \|x\|^k d\mu(x)$ .

**Assumption 4.** The following conditions hold:

- (i)  $\sup_n \mathbb{E} \|X_0^{1,n}\|^2 < \infty$  and  $\sup_{k,n} \mathbb{E} \|\xi_k^{1,n}\|^2 < \infty$ ,
- (ii) There exist  $c, C > 0$ , such that for all  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$\int \langle x, b(x, \mu) \rangle d\mu(x) \leq -c\kappa_2(\mu) + C. \quad (12)$$

**Assumption 4'.** In addition to Assumption 4, the following hold:

- (i)  $\sup_n \mathbb{E} \|X_0^{1,n}\|^4 < \infty$  and  $\sup_{k,n} \mathbb{E} \|\xi_k^{1,n}\|^4 < \infty$ ,
- (ii) There exists constants  $c, C > 0$  such that for all  $\mu \in \mathcal{P}_4(\mathbb{R}^d)$ ,

$$\int \langle x, b(x, \mu) \rangle \|x\|^2 d\mu(x) \leq -c\kappa_4(\mu) + C \left( 1 + \sqrt{\kappa_4(\mu)} \right), \quad (13)$$

Section 4 includes an example for which Assumptions 4 and 4' are satisfied.

**Remark 2.** Assumption 4 can be replaced by the milder condition that  $\sup_{k,n} \mathbb{E} \|X_k^{1,n}\|^2 + \mathbb{E} \|\xi_k^{1,n}\|^2 < \infty$ . Similarly, Assumption 4' can be replaced by the condition that  $(\|X_k^{1,n}\|^2)_{k,n}$  and  $(\|\xi_k^{1,n}\|^2)_{k,n}$  are uniformly integrable. The results of this paper hold under these milder, but less easily verifiable assumptions.

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  $\mathcal{P}_p(C)$ . We refer to the set

$$\mathcal{M} := \underset{(t,n) \rightarrow (\infty, \infty)}{\text{acc}} \left( \left\{ (\Phi_t(m^n))_{\#} \mathbb{P} \right\} \right) \quad (14)$$

as the set of weak $\star$  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  $M \in \mathcal{P}(\mathcal{P}_p(C))$  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  $M$ .

We now state the main results of this paragraph. [Proposition 4](#) shows that the set  $\mathcal{M}$  is non-empty. [Proposition 5](#) shows that any  $M \in \mathcal{M}$  is supported by the set of McKean–Vlasov distributions.

**Proposition 4.** *Let  $1 \leq p < 2$ . Let [Assumptions 1–4](#) hold. Then,*

$$\sup_{k,n} \mathbb{E} \|X_k^{1,n}\|^2 < \infty.$$

Moreover, for any  $j \in \mathbb{N}^*$ , the family of measures  $\{(\Phi_t(m^n))_{\#} \mathbb{P} : t \geq 0, n \in \mathbb{N}^*\}$  is relatively compact in  $\mathcal{P}(\mathcal{P}_p(C))$ .

If [Assumption 4'](#) holds, the conclusion is still valid for  $p = 2$ , and, moreover,  $\sup_{k,n} \mathbb{E} \|X_k^{1,n}\|^4 < \infty$ .

**Proof.** See [Section 5.1](#).  $\square$

**Proposition 5.** *Let  $1 \leq p < 2$ . Under [Assumptions 1–4](#),  $\mathcal{V}_p$  is a non-empty closed set. Moreover,  $M(\mathcal{V}_p) = 1$  for every  $M \in \mathcal{M}$ . If [Assumption 4'](#) holds, the conclusion is still valid for  $p = 2$ .*

**Proof.** See [Section 5.2](#).  $\square$

### 3.2. Ergodic convergence

We provide the proof of the following theorem in [Section 5.3](#).

**Theorem 1.** *Let  $1 \leq p < 2$ . Under [Assumptions 1–4](#),  $\text{BC}_p$  is non-empty.*

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

If [Assumption 4'](#) holds, the statement is still valid for  $p = 2$ .

**Corollary 1.** *Let  $1 \leq p < 2$ . Under [Assumptions 1–4](#),*

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

The same statement holds if  $W_p(\cdot, \cdot)$  is replaced by  $W_p(\cdot, \cdot)^p$ . Finally, if [Assumption 4'](#) holds, the conclusion is still valid for  $p = 2$ .

**Proof.** The proof is provided in [Section 5.4](#).  $\square$

**Remark 3.** The fact that the Birkhoff center  $\text{BC}_p$  is non empty follows from the combination of [Lemmas 7](#) and [9](#). Specifically, [Lemma 7](#) establishes the existence of measures, which, by [Lemma 9](#), can only be supported by  $\text{BC}_p$ .

**Remark 4.** In simple cases,  $\text{BC}_p$  is reduced to a singleton, which corresponds to the unique stationary McKean–Vlasov distribution. For instance, this happens under sufficient but strong assumptions on  $b$  and  $\sigma$ , which ensure a *uniform-in-time propagation of chaos* [[1,7,9](#)]. We refer to [Section 4](#) for a discussion.

Besides this case, the McKean–Vlasov process potentially admits multiple stationary measures. In such a case,  $\text{BC}_p$  contains multiple points. This scenario is common, and interesting regarding practical applications. A first example can be found in [[13,14](#)], in the context of the Granular media equation, see also [Remark 6](#). A second example is encountered in the case of consensus based optimization methods [[34,35](#)], where, under the assumption of a constant noise intensity  $\sigma$ , the limiting McKean–Vlasov process potentially admits several stationary measures. A third example, in the case  $\sigma = 0$ , is given by Stein Variational Gradient Descent (SVGD) algorithm [[36](#)].

Finally, let us review some consequences regarding linear functionals. Denote by  $\text{Lip}_L$  the set of Lipschitz continuous functions on  $\mathbb{R}^d \rightarrow \mathbb{R}$ , whose Lipschitz constant is no larger than  $L \geq 0$ . Define:

$$\text{BC}_p^0(f) := \left\{ \int f d\mu : \mu \in \text{BC}_p^0 \right\}.$$



**Corollary 2.** Let  $1 \leq p < 2$ , and let [Assumptions 1–4](#) hold true. Then, for every  $L \geq 0$

$$\sup_{f \in \text{Lip}_L} d \left( \frac{\sum_{l \in [n], l \in [k]} \gamma_l f(X_l^{i,n})}{n \sum_{l \in [k]} \gamma_l}, \text{conv}(\text{BC}_p^0(f)) \right) \xrightarrow[(k,n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0.$$

The conclusion remains valid for  $p = 2$  under [Assumption 4](#).

**Proof.** See Section 5.5.  $\square$

### 3.3. The case of a unique recurrent point

In this subsection, we will present additional results in the special case where the following assumption holds.

**Assumption 5.** There exists  $\rho^* \in \mathcal{P}_p(C)$  such that  $\text{BC}_p \subset \{\rho^*\}$ .

We observe that, under [Assumptions 1–4](#),  $\text{BC}_p$  is non-empty (see [Theorem 1](#)). Consequently, under [Assumptions 1–5](#), we have  $\text{BC}_p = \{\rho^*\}$ .

Let  $n \in \mathbb{N}^*$  and  $j \leq n$ . One may consider the law of the family of random variables  $(X_l^{1,n}, \dots, X_l^{j,n})$ :

$$I_l^{j,n} := (X_l^{1,n}, \dots, X_l^{j,n})_{\#} \mathbb{P} = \mathbb{P}((X_l^{1,n}, \dots, X_l^{j,n}) \in \cdot).$$

For instance,  $I_l^{1,n}$  is the law of the particle  $X_l^{1,n}$ , which is equal to the law of  $X_l^{i,n}$  for any  $i$ , due to the exchangeability.

**Corollary 3.** Under [Assumptions 1–5](#), we obtain for every  $j \in \mathbb{N}$

$$\frac{\sum_{\ell \in [k]} \gamma_{\ell} W_p(I_{\ell}^{j,n}, (\rho_0^*)^{\otimes j})}{\sum_{\ell \in [k]} \gamma_{\ell}} \xrightarrow[(k,n) \rightarrow (\infty, \infty)]{} 0, \quad (15)$$

where  $W_p$  denotes the Wasserstein distance of order  $p$  on  $\mathcal{P}_p((\mathbb{R}^d)^j)$ , and  $(\rho_0^*)^{\otimes j}$  is the  $j$ -fold tensor product of  $\rho_0^*$ .

**Proof.** See Section 5.6.  $\square$

Eq. (15) can be interpreted as a propagation of chaos result in the long run. This should be compared to standard propagation of chaos results, which are usually stated over a finite time interval [1].

Following [37,38], let us introduce the notion of essential accumulation set. We say that a measure  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$  is an *essential accumulation point* of  $(I_k^{1,n})_{k,n}$ , if for every neighborhood  $U$  of  $\mu$ ,

$$\limsup_{(k,n) \rightarrow (\infty, \infty)} \frac{\sum_{\ell \in [k]} \gamma_{\ell} \mathbb{1}_U(I_{\ell}^{1,n})}{\sum_{\ell \in [k]} \gamma_{\ell}} > 0.$$

This can be interpreted as follows. An essential accumulation point  $\mu$  is an accumulation point, with the property that the particle distribution  $I_k^{1,n} = \mathbb{P}(X_k^{1,n} \in \cdot)$  spends substantial time in the neighborhood of  $\mu$ .

**Corollary 4.** Under [Assumptions 1–5](#),  $\rho_0^*$  is the unique essential accumulation point of  $(I_k^{1,n})_{k,n}$ .

**Proof.** See Section 5.7.  $\square$

In other terms, as  $(k, n)$  tend to infinity, the law  $\mathbb{P}(X_k^{1,n} \in \cdot)$  spends most of its time in the neighborhood of  $\rho_0^*$ .

### 3.4. Pointwise convergence to a global attractor

Depending on the vector field  $b$ , it is often the case that each measure  $\rho \in \mathcal{V}_p$  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

$$\rho \in \mathcal{V}_p \Leftrightarrow \forall t \geq 0, \rho_t = \Psi_t(\rho_0). \quad (16)$$

We shall say that in this situation, the class  $\mathcal{V}_p$  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 5.8.

**Theorem 2.** Let  $p \in [1, 2]$ , and let [Assumption 1](#), [2](#), and [3](#) hold true. Let [Assumption 4](#) or the stronger [Assumption 4'](#) hold true according to whether  $p < 2$  or  $p = 2$  respectively. Assume in addition that the  $V_p$  has a semi-flow structure on  $\mathcal{P}_p(\mathbb{R}^d)$  as specified in [\(16\)](#), and that this semi-flow  $\Psi$  admits a global attractor  $A_p$ . Then,

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

**Remark 5.** A typical scenario where the set  $A_p$  exists and contains a single element is provided in [Remark 6](#).

#### 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  $V_2$  of McKean–Vlasov processes, in the case where  $\sigma(x, \mu) = \sigma I_d$  for some real constant  $\sigma \geq 0$  and with a slight abuse of notation  $b(x, \mu) = \int b(x, y) d\mu(y)$ , with:

$$b(x, y) := -\nabla V(x) - \nabla U(x - y), \quad (17)$$

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

**Assumption 6 (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 [Assumption 6](#), the vector field  $b$  and  $\sigma$  satisfies [Assumption 1](#). We will see later, as a byproduct of [Theorem 3](#), that the set  $V_2$  of McKean–Vlasov distributions associated to the field  $b$  in [Eq. \(17\)](#), 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  $\mathcal{H}(\mu) = \mathcal{F}(\mu) + \mathcal{V}(\mu) + \mathcal{U}(\mu)$  with

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

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

The following central result provides a central properties of the elements of  $V_2$ .

**Proposition 6.** Let [Assumption 6](#) hold true, and let  $b$  be defined by [\(17\)](#). Assume  $\sigma > 0$ . Consider  $\rho \in V_2$ . Then, for every  $t > 0$ ,  $\rho_t$  admits a density  $x \mapsto \rho(t, x)$  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  $t_2 > t_1 > 0$ ,

$$\mathcal{H}(\rho_{t_2}) - \mathcal{H}(\rho_{t_1}) = - \int_{t_1}^{t_2} \int \|v_t(x)\|^2 \rho(t, x) dx dt, \quad (18)$$

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

$$v_t(x) := -\nabla V(x) - \int \nabla U(x - y) d\rho_t(y) - \sigma^2 \nabla \log \rho(t, x). \quad (19)$$

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:

$$S := \{ \mu \in \mathcal{P}_2^r(\mathbb{R}^d) : \nabla V + \int \nabla U(\cdot - y) d\mu(y) + \sigma^2 \nabla \log \frac{d\mu}{d\mathcal{L}^d} = 0 \text{ } \mu\text{-a.e.} \}. \quad (20)$$

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

$$A_\epsilon := \{ \rho \in V_2 : \exists \mu \in S, \forall t \geq \epsilon, \rho_t = \mu \}. \quad (21)$$

**Proposition 7.** We posit the assumptions of [Proposition 6](#). For every  $\epsilon > 0$ , the function  $\rho \mapsto \mathcal{H}(\rho_\epsilon)$  is real valued on  $V_2$ , lower semicontinuous, and is a Lyapunov function for the set  $A_\epsilon$ . Moreover,

$$BC_2 \subset \overline{A_0}.$$

We also need to consider a setting where  $V_2$  has a semi-flow structure on  $\mathcal{P}_2(\mathbb{R}^d)$  as in [\(16\)](#) 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 [\[22\]](#) of Ambrosio, Gigli and Savaré. The following additional assumption will be needed:

**Assumption 7.** 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 6](#) hold true with  $\beta = 1$ , and let [Assumption 7](#) hold true. Then, for each  $\rho \in \mathcal{V}_2$ , the curve  $t \mapsto \rho_t$  belongs to the set of absolutely continuous functions  $\text{AC}_{\text{loc}}^2((0, \infty), \mathcal{P}_2(\mathbb{R}^d))$  as defined in [\[22, Sec. 8.3\]](#), 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,  $\mathcal{V}_2$  has a semi-flow structure, and we write  $\rho_t = \Psi_t(\rho_0)$ .

For completeness, we recall along [\[22, Chap. 8 and 11\]](#) 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  $t > 0$ ,  $\|w_t\|_{L^2(\rho_t)} \in L_{\text{loc}}^p(0, \infty)$ , 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 [\[22, Chap. 10\]](#), which always exists under our assumptions. Actually,  $w_t = v_t$  as given by [Eq. \(19\)](#) for almost all  $t$ .

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

$$X_{k+1}^{i,n} = X_k^{i,n} - \frac{\gamma_{k+1}}{n} \sum_{j \in [n]} \nabla U(X_k^{i,n} - X_k^{j,n}) - \gamma_{k+1} \nabla V(X_k^{i,n}) + \sqrt{2\gamma_{k+1}} \xi_k^{i,n}, \quad (22)$$

for each  $i \in [n]$ . This is a special case of [Eq. \(1\)](#) with  $b(x, y)$  given by [Eq. \(17\)](#) and  $\zeta_k^{i,n} = 0$  for all  $k$ . For simplicity, [Assumption 3](#) will be replaced by the following stronger assumption:

**Assumption 8.** We assume that 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 such that  $\mathbb{E}(\|\xi_1^{1,1}\|^4) < \infty$ .

The next proposition implies that [Assumption 4'](#) holds.

**Proposition 9.** Let [Assumption 2](#), [6](#) and [8](#) be satisfied. Then, [Eq. \(12\)](#) and [\(13\)](#) hold.

Putting [Assumption 2](#), [6](#) and [8](#) together, the hypotheses of [Theorem 1](#) are satisfied for  $p = 2$ .

**Theorem 3.** Let [Assumption 2](#), [6](#) and [8](#) be satisfied. Assume  $\sigma > 0$ . Then, the set  $S$  given by [Eq. \(20\)](#) is non empty, and furthermore,

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

**Proof.** Use [Corollary 1](#) with  $p = 2$ , together with [Proposition 7](#).  $\square$

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

**Theorem 4.** Let [Assumption 6](#) hold true with  $\beta = 1$ , and let [Assumption 7](#) 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  $S = A_2$ . In any case,

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

**Remark 6.** Many authors have been interested in the long-time convergence of granular media equations under hypotheses ensuring the uniqueness of the stationary distribution [\[8,39–42\]](#). The most obvious case where such a situation arises, is the case where the functions  $U$  and  $V$  are both strongly convex.; Then, there exists  $\lambda > 0$  such that  $W_2(\Psi_t(\nu), \Psi_t(\nu')) \leq e^{-\lambda t} W_2(\nu, \nu')$  [\[22, Th. 11.2.1\]](#). Here, [Theorem 4](#) applies, with  $A_2$  being reduced to the unique stationary measure.

On the other hand, the coexistence of multiple stationary measures typically corresponds to the case of metastable behaviors, where the Helmholtz energy admits several critical points. For instance, this situation arises in the case of a multi-well potential with low noise intensity [\[14,43\]](#). Although it can be challenging to characterize such phase transition phenomena, our work supports the assertion that a numerical system with  $n$  particles provides an estimate, in the sense that the  $n$ -system inherits the same asymptotic behavior as its mean-field approximation.

## 5. Proofs of Section 3

### 5.1. Proof of [Proposition 4](#)

In this paragraph, consider  $1 \leq p \leq 2$ . We recall that, when  $p < 2$ , [Assumption 4](#) holds and [Assumption 4'](#) holds when  $p = 2$ . First, we need the following lemma.

**Lemma 4.**

Let [Assumption 4](#) with [Assumption 1](#), [2](#), and [3](#) hold true, it holds that  $\sup_{k,n} \mathbb{E}\|X_k^{1,n}\|^2 < \infty$ . Furthermore, when [Assumption 4'](#) holds, we have  $\sup_{k,n} \mathbb{E}\|X_k^{1,n}\|^4 < \infty$ .

**Proof.** First, we will show the first point of the lemma. We recall the iteration

$$X_{k+1}^{i,n} = X_k^{i,n} + \gamma_{k+1} b(X_k^{i,n}, \mu_k^n) + \gamma_k \zeta_k^{i,n} + \sqrt{2\gamma_{k+1}} \varepsilon_{k+1}^{i,n}.$$

In this proof, we denote by  $C > 0$  a generic constant that is sufficiently large, and by  $c > 0$  a generic constant that is sufficiently small. We take  $k$  large enough such that  $\gamma_k + \gamma_k^2 \leq C\gamma_k$  and  $-\gamma_k + \gamma_k^2 \leq -c\gamma_k$ . For simplicity, we remove the superscript  $n$  from  $X_k^{i,n}$ ,  $\mu_k^n$ ,  $\zeta_k^{i,n}$ , and  $\varepsilon_k^{i,n}$ . Moreover, we remove the subscript  $k+1$  from  $\gamma_{k+1}$ .

By [Assumption 1](#), for  $i \in [n]$ , we obtain

$$\begin{aligned} & \|X_{k+1}^i\|^2 - \|X_k^i\|^2 \\ &= \gamma \langle X_k^i, b(X_k^i, \mu_k) \rangle + \sqrt{2\gamma} \langle X_k^i, \xi_{k+1}^i \rangle + \gamma \langle X_k^i, \zeta_{k+1}^i \rangle \\ &+ \|\gamma b(X_k^i, \mu_k) + \sqrt{2\gamma} \xi_{k+1}^i + \gamma \zeta_{k+1}^i\|^2 \\ &\leq \gamma \langle X_k^i, b(X_k^i, \mu_k) \rangle + \sqrt{2\gamma} \langle X_k^i, \xi_{k+1}^i \rangle + \gamma \langle X_k^i, \zeta_{k+1}^i \rangle + \\ &6\gamma^2 \|X_k^i\|^2 + 6\gamma^2 \int \|x\|^2 d\mu_k(x) + 6\gamma \|\xi_{k+1}^i\|^2 + 3\gamma^2 \|\zeta_{k+1}^i\|^2. \end{aligned} \quad (23)$$

Summing the latter with respect to  $i$ , with [Eq. \(12\)](#), we obtain

$$\begin{aligned} & \frac{1}{C} \left( \int \|x\|^2 d\mu_{k+1}(x) - \int \|x\|^2 d\mu_k(x) \right) \\ &\leq \gamma \int \langle x, b(x, \mu_k) \rangle d\mu_k(x) + \frac{\sqrt{\gamma}}{n} \sum_{i \in [n]} \langle X_k^i, \xi_{k+1}^i \rangle + \frac{\gamma}{n} \sum_{i \in [n]} \langle X_k^i, \zeta_{k+1}^i \rangle \\ &+ \gamma^2 \int \|x\|^2 d\mu_k(x) + \frac{1}{n} \sum_{i \in [n]} (\gamma \|\xi_{k+1}^i\|^2 + \gamma^2 \|\zeta_{k+1}^i\|^2) \\ &\leq -c\gamma \int \|x\|^2 d\mu_k(x) + \frac{\sqrt{\gamma}}{n} \sum_{i \in [n]} \langle X_k^i, \xi_{k+1}^i \rangle + \frac{\gamma}{n} \sum_{i \in [n]} \langle X_k^i, \zeta_{k+1}^i \rangle \\ &+ \gamma^2 \int \|x\|^2 d\mu_k(x) + \frac{1}{n} \sum_{i \in [n]} (\gamma \|\xi_{k+1}^i\|^2 + \gamma^2 \|\zeta_{k+1}^i\|^2) + C\gamma. \end{aligned} \quad (24)$$

Taking the expectation, by the exchangeability given by [Assumption 3](#), the assumption on  $(\zeta_k^i)_{i,k}$ , and [Assumption 3](#), we obtain

$$\mathbb{E} \|X_{k+1}^1\|^2 - \mathbb{E} \|X_k^1\|^2 \leq -c\gamma \mathbb{E} \|X_k^1\|^2 + C\gamma.$$

As a consequence, we obtain the first point of the lemma.

Now, we proceed to demonstrate the second point of the lemma. But first we claim that

$$\sup_{k \in \mathbb{N}} \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \right)^2 < \infty. \quad (25)$$

Indeed, by raising to the square [Eq. \(24\)](#) and taking the expectation, we obtain

$$\begin{aligned} & \mathbb{E} \left( \int \|x\|^2 d\mu_{k+1}(x) \right)^2 - \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \right)^2 \\ &\leq -c\gamma \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \right)^2 + C\gamma \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \right) + C\gamma. \end{aligned}$$

Now, we will obtain the second point of the lemma. By raising to the square [Eq. \(23\)](#), we obtain

$$\begin{aligned} & \frac{1}{C} (\|X_{k+1}^i\|^4 - \|X_k^i\|^4) \\ &\leq \gamma \langle X_k^i, b(X_k^i, \mu_k) \rangle \|X_k^i\|^2 + \sqrt{\gamma} \langle X_k^i, \xi_{k+1}^i \rangle \|X_k^i\|^2 + \gamma \langle X_k^i, \zeta_{k+1}^i \rangle \|X_k^i\|^2 + \\ &\gamma^2 \|X_k^i\|^4 + \gamma^2 \int \|x\|^2 d\mu_k(x) \|X_k^i\|^2 + \gamma \|\xi_{k+1}^i\|^2 \|X_k^i\|^2 + \gamma^2 \|\zeta_{k+1}^i\|^2 \|X_k^i\|^2 \\ &\gamma^4 \|X_k^i\|^4 + \gamma^4 \left( \int \|x\|^2 d\mu_k(x) \right)^2 + \gamma^2 \|\xi_{k+1}^i\|^4 + \gamma^4 \|\zeta_{k+1}^i\|^4. \end{aligned}$$

Summing over  $i \in [n]$ , we obtain

$$\begin{aligned} & \frac{1}{C} \left( \int \|x\|^4 d\mu_{k+1}(x) - \int \|x\|^4 d\mu_k(x) \right) \\ & \leq \gamma \int \langle x, b(x, \mu_k) \rangle \|x\|^2 d\mu_k(x) + \frac{\sqrt{\gamma}}{n} \sum_{i \in [n]} \langle X_k^i, \xi_{k+1}^i \rangle \|X_k^i\|^2 + \\ & \quad \frac{\gamma}{n} \sum_{i \in [n]} \langle X_k^i, \xi_{k+1}^i \rangle \|X_k^i\|^2 + \gamma^2 \int \|x\|^4 d\mu_k(x) + \gamma^2 \left( \int \|x\|^2 d\mu_k(x) \right)^2 + \\ & \quad \frac{\gamma}{n} \sum_{i \in [n]} \|\xi_{k+1}^i\|^2 \|X_k^i\|^2 + \frac{\gamma^2}{n} \sum_{i \in [n]} \|\xi_{k+1}^i\|^2 \|X_k^i\|^2 + \gamma^4 \int \|x\|^4 d\mu_k(x) + \\ & \quad \gamma^4 \left( \int \|x\|^2 d\mu_k(x) \right)^2 + \frac{\gamma^2}{n} \sum_{i \in [n]} \|\xi_{k+1}^i\|^4 + \frac{\gamma^4}{n} \sum_{i \in [n]} \|\xi_{k+1}^i\|^4. \end{aligned}$$

Taking the expectation, by Eq. (13), and by the assumption on  $(\xi_k^i)_{k,i}$ , we obtain

$$\begin{aligned} & \frac{1}{C} (\mathbb{E} \|X_{k+1}^1\|^4 - \|X_k^1\|^4) \\ & \leq -c\gamma \mathbb{E} \|X_k^1\|^4 + \gamma \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \left( \int \|x\|^4 d\mu_k(x) \right)^{1/2} \right) \\ & \quad + \gamma^2 \mathbb{E} \|X_k^1\|^4 + \gamma^2 \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \right)^2 + \gamma \mathbb{E} \|X_n^1\|^2 \\ & \quad + \gamma \mathbb{E} \left( \int \|x\|^4 d\mu_k(x) \right)^{1/2} + \gamma. \end{aligned}$$

Cauchy–Schwarz inequality yields

$$\begin{aligned} & \frac{1}{C} (\mathbb{E} \|X_{k+1}^1\|^4 - \|X_k^1\|^4) \\ & \leq -c\gamma \mathbb{E} \|X_k^1\|^4 + \gamma \left( \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \right)^2 \right)^{1/2} (\mathbb{E} \|X_k^1\|^4)^{1/2} \\ & \quad + \gamma^2 \mathbb{E} \|X_k^1\|^4 + \gamma^2 \mathbb{E} \left( \int \|x\|^2 d\mu_k(x) \right)^2 + \gamma (\mathbb{E} \|X_n^1\|^4)^{1/2} + \gamma. \end{aligned}$$

Finally, by Eq. (25), we obtain

$$\begin{aligned} & \frac{1}{C} (\mathbb{E} \|X_{k+1}^1\|^4 - \|X_k^1\|^4) \\ & \leq -c\gamma \mathbb{E} \|X_k^1\|^4 + \gamma^2 \mathbb{E} \|X_k^1\|^4 + \gamma (\mathbb{E} \|X_n^1\|^4)^{1/2} + \gamma, \end{aligned}$$

which concludes the proof.  $\square$

Note that  $(\Phi_t(m^n))$  belongs to  $\mathcal{P}_p(C)$ .

In the light of Lemma 2 and Proposition 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 5 and 6 below.

**Lemma 5.** *We posit the assumptions of Proposition 4. The family of intensities  $\{\mathbb{I}(\Phi_t(m^n))\}_{t,n}$  is weak $\star$ -relatively compact in  $\mathcal{P}(C)$ .*

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

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

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

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

for each measurable function  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ . According to Th. 7.3 in [44], the weak $\star$ -relative compactness of the sequence  $(\hat{\mathbb{I}}_t^n)_{t,n}$  in  $\mathcal{P}(C)$  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

$$\lim_{\delta \rightarrow 0} \limsup_{t,n} \mathbb{P} \left( w_{\bar{X}_{t+}^{1,n}}^T(\delta) \geq \varepsilon \right) = 0 \quad (26)$$

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\*-relative compactness of  $(\mathbb{I}_t^n)_{t,n}$  in  $\mathcal{P}(\mathbb{R}^d)$ , follows directly from Lemma 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:

$$\begin{aligned}\bar{X}_t^{1,n} - \bar{X}_s^{1,n} &= P_{s,t}^n + N_{s,t}^n + U_{s,t}^n, \\ P_{s,t}^n &:= \sum_{k=k_s}^{k_t-2} \gamma_{k+1} b(X_k^{1,n}, \mu_k^n) \\ &\quad + \left(\tau_{k_s} - s\right) b(X_{k_s-1}^{1,n}, \mu_{k_s-1}^n) + \left(\tau_{k_t} - t\right) b(X_{k_t-1}^{1,n}, \mu_{k_t-1}^n) \\ N_{s,t}^n &:= \sum_{k=k_s}^{k_t-2} \sqrt{\gamma_{k+1}} \xi_{k+1}^{1,n} + \frac{\tau_{k_s} - s}{\gamma_{k_s}} \sqrt{\gamma_{k_s}} \xi_{k_s}^{1,n} + \frac{\tau_{k_t} - t}{\gamma_{k_t}} \sqrt{\gamma_{k_t}} \xi_{k_t}^{1,n} \\ U_{s,t}^n &:= \sum_{k=k_s}^{k_t-2} \gamma_{k+1} \zeta_{k+1}^{1,n} + \left(\tau_{k_s} - s\right) \zeta_{k_s}^{1,n} + \left(\tau_{k_t} - t\right) \zeta_{k_t}^{1,n}.\end{aligned}\tag{27}$$

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  $k \neq k_s, k_t$ . Note that:

$$\sum_{k=k_s-1}^{k_t-1} \tilde{\gamma}_{k+1} = t - s.\tag{28}$$

Moreover, we have:

$$\frac{\tau_{k_s} - s}{\gamma_{k_s}} \sqrt{\gamma_{k_s}} \leq \sqrt{\tilde{\gamma}_{k_s}}, \quad \text{and} \quad \frac{\tau_{k_t} - t}{\gamma_{k_t}} \sqrt{\gamma_{k_t}} \leq \sqrt{\tilde{\gamma}_{k_t}}.\tag{29}$$

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 [45, Th. 1.1] to the components of the vector  $N_{s,t}^n$  in  $\mathbb{R}^d$ . As Eq. (28) and (29) hold:

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

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

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left( \|N_{s,t}^n\|_4^4 \right) \leq C(t-s)^2.\tag{30}$$

Furthermore, using Jensen's inequality along with Eq. (28), we obtain

$$\|P_{s,t}^n\|^2 \leq (t-s) \sum_{k=k_s-1}^{k_t-1} \tilde{\gamma}_{k+1} \|b(X_k^{1,n}, \mu_k^n)\|^2.$$

Using Assumption 1 and Lemma 4, there exists a constant  $C$ , independent of  $s, t, n$ , such that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left( \|P_{s,t}^n\|^2 \right) \leq C(t-s)^2.\tag{31}$$

Also, by Jensen's inequality, we have

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

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

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left( \|U_{s,t}^n\|^2 \right) \leq C(t-s)^2.\tag{32}$$

Combining Eqs. (31), (30) and (32), we have shown:

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left[ \|P_{s,t}^n\|^2 + \|N_{s,t}^n\|^4 + \|U_{s,t}^n\|^2 \right] \leq C(t-s)^2,\tag{33}$$

where  $0 \leq s < t < \infty$ , and  $C$  is a positive constant, independent of  $s, t, n$ . Using [46, Th. 2.8] and Markov's inequality, Eq. (26) hold.  $\square$

**Lemma 6.** We posit the assumptions of Proposition 4. For every  $T > 0$ ,

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

**Proof.** By the exchangeability stated in [Assumption 3–\(i\)](#), we obtain:

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

for every  $k, t, n$ . Recalling the decomposition introduced in [Eq. \(27\)](#), for every  $u \in [0, T]$ :

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

Hence,

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

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  $(\sup_{u \in [0, T]} \|U_{t,t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ .  $(\|\bar{X}_t^{1,n}\|^p : k \in \mathbb{R}_+, n \in \mathbb{N}^*)$  is uniformly integrable by the first point of [Lemma 4](#) when  $p < 2$ , and by the second point of [Lemma 4](#) when  $p = 2$ . As obtained in [Eq. \(30\)](#), Burkholder inequality stated in [\[45, Th 1.1\]](#) 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 [Eqs. \(31\) and \(32\)](#), we derive:

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

and

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

where  $C$  remains a constant independent of  $n$  and  $t$ . Using the first point of [Lemma 4](#) when  $p < 2$ , and the second point of [Lemma 4](#) 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

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

Hence, by Jensen's inequality,

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

Consequently,

$$\sup_{t \in \mathbb{R}_+, n \in \mathbb{N}^*} \mathbb{E} \left[ F \left( \sup_{u \in [0, T]} \|P_{t,t+u}^n\|^p \right) \right] < \infty.$$

Therefore, de la Vallée Poussin theorem yields the uniform integrability of the collection  $(\sup_{u \in [0, T]} \|P_{t,t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ . The uniform integrability of the collection  $(\sup_{u \in [0, T]} \|U_{t,t+u}^n\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$  is obtained, by the same arguments. This completes the proof.  $\square$

To conclude the proof of [Proposition 4](#), it is sufficient to remark that the tightness conditions provided in [Lemma 2](#) are satisfied, thanks to [Lemmas 5 and 6](#), with [Proposition 1](#).

## 5.2. Proof of [Proposition 5](#)

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

**Proposition 10.** Let [Assumption 1](#), [2](#), [3](#) and [4](#) 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  $m_n := \Phi_{t_n}(m^n) \in \mathcal{P}_p(C)$ . We have

$$G(m_n) = \frac{1}{n} \sum_{i \in [n]} \left( \phi(\bar{X}_{t_n+t}^{i,n}) - \phi(\bar{X}_{t_n+s}^{i,n}) - \int_{t_n+s}^{t_n+t} \psi(\bar{X}_u^{i,n}, m_u^n) du \right) \mathcal{Q}^{i,n}, \quad (34)$$

where we set  $\psi(x, \mu) := \langle \nabla \phi(x), b(x, \mu) \rangle + \text{tr}(\sigma(x, \mu)^T H_\phi(x) \sigma(x, \mu))$ , and

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

We note right away that  $|\mathcal{Q}^{i,n}| \leq C$  where  $C$  depends on the functions  $h_j$  only, and furthermore, the random variables  $\{\mathcal{Q}^{i,n}\}_{i \in [n]}$  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 (34), 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

$$\begin{aligned} &\phi(X_{k+1}^i) - \phi(X_k^i) \\ &= \langle \nabla \phi(X_k^i), X_{k+1}^i - X_k^i \rangle + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\ &= \gamma_{k+1} \langle \nabla \phi(X_k^i), b(X_k^i, \mu_k^n) \rangle \\ &\quad + \gamma_{k+1} \text{tr}(\sigma(X_k^i, \mu_k^n)^T H_\phi(X_k^i) \sigma(X_k^i, \mu_k^n)) \\ &\quad + \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\ &\quad - \gamma_{k+1} \text{tr}(\sigma(X_k^i, \mu_k^n)^T H_\phi(X_k^i) \sigma(X_k^i, \mu_k^n)) + \gamma_{k+1} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle \\ &= \gamma_{k+1} \psi(X_k^i, \mu_k^n) + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\ &\quad + \gamma_{k+1} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle - \gamma_{k+1} \text{tr}(\sigma(X_k^i, \mu_k^n)^T H_\phi(X_k^i) \sigma(X_k^i, \mu_k^n)) \\ &\quad + \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle \\ &= \gamma_{k+1} \psi(X_k^i, \mu_k^n) + \frac{1}{2} (X_{k+1}^i - X_k^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i) (X_{k+1}^i - X_k^i) \\ &\quad + \gamma_{k+1} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle - \gamma_{k+1} \text{tr}(\sigma(X_k^i, \mu_k^n)^T H_\phi(X_k^i) \sigma(X_k^i, \mu_k^n)) + \\ &\quad \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle + \gamma_{k+1} (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i - \gamma_{k+1} (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i \end{aligned}$$

In this last expression, the terms  $\psi(X_k^i, \mu_k^n)$  will be played against the integral term at the right hand side of (34), and the other terms will be proven to have negligible effects. Since  $\text{tr}(\xi_{k+1}^i (\xi_{k+1}^i)^T H_\phi(\bar{X}_{\theta_{k+1}}^i)) = (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i$ , the term

$$\begin{aligned} \eta_{k+1}^i &:= \sqrt{2\gamma_{k+1}} \langle \nabla \phi(X_k^i), \xi_{k+1}^i \rangle + \gamma_{k+1} (\xi_{k+1}^i)^T H_\phi(X_k^i) \xi_{k+1}^i \\ &\quad - \gamma_{k+1} \text{tr}(\sigma(X_k^i, \mu_k^n)^T H_\phi(X_k^i) \sigma(X_k^i, \mu_k^n)) \end{aligned}$$

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 (34), we can write

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



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

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

To prove our proposition, we show that  $\mathbb{E}|\chi_l^n| \rightarrow 0$  for all  $l \in [8]$ . The notation  $E_\infty^\times$  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

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

and using the boundedness of  $Q^i$  and the exchangeability as stated by [Assumption 3–\(i\)](#), we obtain that

$$\mathbb{E}|\chi_1^n| \leq C \mathbb{E}|E_1^n|.$$

We begin by providing a bound on the second moments of  $E_i^n$ . Recalling the definition of  $\psi$ , and using the compactness of the support of  $\phi$  along with [Assumption 1](#), we obtain that

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

thanks to [Lemma 4](#). To obtain that  $\mathbb{E}|\chi_1^n| \rightarrow 0$ , we thus need to show that  $\mathbb{E}|E_1^n| \rightarrow 0$ .

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

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

by the exchangeability, thus, the sequence of random variables  $(\bar{X}_{t_n+}^{1,n})_n$  is tight. Let us work on the r. v.  $U_n := \pi_{[0,t-s]} \# \bar{X}_{t_n+s+}^1$  defined on the set  $C([0, t-s])$  of continuous functions on the interval  $[0, t-s]$ . Since  $(\bar{X}_{t_n+}^{1,n})_n$  and  $((\Theta_t)_\# m^n)_{t,n}$  are tight (by prop. 4), given an arbitrary  $\varepsilon > 0$ , there exists two compact sets  $K_\varepsilon \subset C([0, t-s])$  and  $\mathcal{K}_\varepsilon \subset \mathcal{P}_p(C)$  such that

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P} \left[ U_n \notin K_\varepsilon \right] + \mathbb{P} \left[ \Phi_{t_n+s}(m^n) \notin \mathcal{K}_\varepsilon \right] \leq \varepsilon.$$

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

$$\begin{aligned} |E_1^n| &\leq \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1} \max_{\delta \in [0, \bar{\gamma}_{k+1}]} \left| \psi(\bar{X}_{\tau_k+\delta}^1, m_{\tau_k+\delta}^n) - \psi(\bar{X}_{\tau_k}^1, m_{\tau_k}^n) \right| \\ &\leq (t-s) \max_{\substack{u,v \in [0,t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n+s}}}} \left| \psi(U_n(u), m_{t_n+s+u}^n) - \psi(U_n(v), m_{t_n+s+v}^n) \right|. \end{aligned}$$

We thus can write

$$\begin{aligned} \mathbb{E} |E_1^n| &= \mathbb{E} |E_1^n| \mathbb{1}_{(U_n, \Phi_{t_n+s}(m^n)) \in K_\varepsilon \times \mathcal{K}_\varepsilon} + \mathbb{E} |E_1^n| \mathbb{1}_{(U_n, \Phi_{t_n+s}(m^n)) \notin K_\varepsilon \times \mathcal{K}_\varepsilon} \\ &\leq (t-s) \sup_{f, \rho \in K_\varepsilon \times \mathcal{K}_\varepsilon} \max_{\substack{u,v \in [0,t-s] \\ |u-v| \leq \bar{\gamma}_{k_{t_n+s}}}} \left| \psi(f(u), \rho_u) - \psi(f(v), \rho_v) \right| \\ &\quad + \sqrt{\mathbb{E}(E_1^n)^2} \sqrt{\mathbb{P}[U_n \notin K_\varepsilon] + \mathbb{P}[\Phi_{t_n+s}(m^n) \notin \mathcal{K}_\varepsilon]}. \end{aligned} \quad (35)$$

By the Arzelà-Ascoli theorem, the functions in  $K_\varepsilon$  are uniformly equicontinuous and bounded. Moreover, the set  $\{u \in [0, t-s] \mapsto \rho_u, \rho \in \mathcal{K}_\varepsilon\}$  is also uniformly equicontinuous and  $\{\rho_u : u \in [0, t-s], \rho \in \mathcal{K}_\varepsilon\}$  is include in a compact subspace of  $\mathcal{P}_p(\mathbb{R}^d)$ .

Since  $\psi$  is a continuous function, by Heine theorem,  $\psi$  is equicontinuous, when we restrict  $\psi$  to a compact space. Therefore, one can easily check that the set of functions  $S$  on  $[0, t-s]$  defined as

$$S := \{u \mapsto \psi(f(u), \rho_u) : (f, \rho) \in K_\varepsilon \times \mathcal{K}_\varepsilon\}$$

is a set of uniformly equicontinuous functions. As a consequence, the first term at the right hand side of the inequality in (35) 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  $\mathbb{E}(E_1^n)^2$ . Since  $\varepsilon$  is arbitrary, we obtain that  $\mathbb{E}|E_1^n| \rightarrow 0$ , 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  $U_n := \pi_{[0,t-s]}^\# \bar{X}_{t_n+s+}^1$  and  $\bar{\gamma}_l$  introduced above, we have

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

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

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

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

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

satisfies

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

Therefore,

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

Since  $\mathbb{E} \|\xi_{k+1}^1\|^2$  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

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

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

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

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,  $\mathbb{E}|\chi_n^4| \rightarrow 0$ .

We now turn to  $\chi_n^6$ . Here we write

$$\chi_n^6 = \frac{1}{n} \sum_{i \in [n]} \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1}^{3/2} E_k^i,$$

where

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

satisfies

$$|E_k^i| \leq C \|b(X_k^i, \mu_k^n)\| \|\xi_{k+1}^i\| + C \sqrt{\gamma_{k+1}} \|b(X_k^i, \mu_k^n)\|^2.$$

We readily obtain from [Assumption 1, 3](#) and [Lemma 4](#) that  $\mathbb{E}|E_k^i| \leq C$ , which leads to  $\mathbb{E}|\chi_n^6| \rightarrow 0$ .

The treatment of the term  $\chi_n^7$  is similar and is omitted.

We finally deal with  $\chi_n^8$  that involves the martingale increments  $\eta_k^i$ . We decompose this term by writing

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

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

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

$$\begin{aligned} &\leq C \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1}^2 \\ &\leq \bar{\gamma}_{k_{t_n+s}} C(t-s). \end{aligned}$$

It results that  $\mathbb{E}(\chi_8^n)^2 \rightarrow 0$ . The proof of Proposition 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 Proposition 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  $\text{supp}(M) \subset \mathcal{P}_p(C)$  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  $\mathcal{P}_p(C)$ . 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  $\mathcal{P}(\mathcal{P}_p(C))$ :

$$M_t^n := \frac{1}{t} \int_0^t \delta_{\Phi_s(m^n)} ds. \quad (36)$$

**Lemma 7.** The collection of r.v.  $(M_t^n, t \geq 0, n \in \mathbb{N}^*)$  is tight in  $\mathcal{P}(\mathcal{P}_p(C))$ .

**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  $\mathcal{P}(\mathcal{P}_p(C))$ . Recall that  $\mathbb{I}(M_t^n)$  is the probability measure which, to every Borel subset  $A \subset \mathcal{P}_p(C)$ , 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 Proposition 4, there exists a compact set  $\mathcal{K} \in \mathcal{P}_p(C)$  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 7,  $\mathcal{M}$  is a non empty subset of  $\mathcal{P}(\mathcal{P}_p(C))$ . Define:

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

**Lemma 8.** For every  $Y \in \mathcal{M}$ ,  $Y(\mathcal{V}_p) = 1$ .

**Proof.** Consider  $Y \in \mathcal{M}$ . Without restriction, we write  $Y$  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 Proposition 5) is a continuous function on  $\mathcal{P}_p(C)$ . Denoting by  $\langle \cdot, \cdot \rangle$  the natural dual pairing on  $C_b(\mathcal{P}_p(C)) \times \mathcal{P}(\mathcal{P}_p(C))$ , the function  $\langle W_p(\cdot, \mathcal{V}_p), \cdot \rangle$  is a continuous on  $\mathcal{P}(\mathcal{P}_p(C))$ . Thus, the sequence of real r.v.  $\langle W_p(\cdot, \mathcal{V}_p), M_{t_n}^n \rangle$  converges in distribution to  $\langle W_p(\cdot, \mathcal{V}_p), \cdot \rangle_\# Y$ . These variables being bounded, we obtain by taking the limits in expectation:

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

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

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

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

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

**Lemma 9.** For every  $Y \in \mathcal{M}$ ,  $Y(\mathcal{I}) = 1$ .

**Proof.** Similarly to the proof of Lemma 8, we assume without restriction that  $Y = \lim_{n \rightarrow \infty} (M_{t_n}^n)_\# \mathbb{P}$  in the weak $\star$  sense. Set  $t > 0$ . The map  $M \mapsto d_L(M, (\Theta_t)_\# M)$  is continuous on  $\mathcal{P}(\mathcal{P}_p(C))$ , where we recall that  $d_L$  stands for the Lévy–Prokhorov distance. Thus, by Fatou’s lemma,

$$\int d_L(M, (\Theta_t)_\# M) dY(M) \leq \limsup_{n \rightarrow \infty} \mathbb{E}(d_L(M_{t_n}^n, (\Theta_t)_\# M_{t_n}^n)). \quad (37)$$

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  $A \subset \mathcal{P}_p(C)$ ,  $|(\Theta_t)_\# M_{t_n}^n(A) - M_{t_n}^n(A)| \leq 2t/t_n$ . 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. (37) 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:  $B_p = \{M \in \mathcal{P}(\mathcal{P}_p(C)) : M(\text{BC}_p) = 1\}$ .

**Proposition 11.** For every  $Y \in \mathcal{M}$ ,  $Y(B_p) = 1$ .

**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  $Y \in \mathcal{M}$  as  $n \rightarrow \infty$ . By Lemma 9, the map  $(\Theta_t)_\# : \mathcal{P}_p(C) \rightarrow \mathcal{P}_p(C)$  preserves the measure  $M$ , for all  $M$   $Y$ -a.e., and for all  $t$ . By Lemma 8,  $M(\mathbb{V}_p) = 1$ . Thus, the restriction of the map  $(\Theta_t)_\#$  to  $\mathbb{V}_p$ , still denoted by  $(\Theta_t)_\# : \mathbb{V}_p \rightarrow \mathbb{V}_p$  preserves the measure  $M$  as well, for all  $M$   $Y$ -a.e. By the Poincaré recurrence theorem, stated in Th. 2.3 of [47], it follows that  $M(\text{BC}_p) = 1$  for all  $M$   $Y$ -a.e.  $\square$

**Proof of Theorem 1.** By Lemma 7, the set  $\mathcal{M}$  is non-empty. Consequently, by Proposition 11,  $B_p$  is non-empty, which implies that  $\text{BC}_p$  is also non-empty. 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,

$$\mathbb{E}(\langle W_p(\cdot, \text{BC}_p), M_{t_n}^n \rangle) > \varepsilon, \quad (38)$$

where  $\langle \cdot, \cdot \rangle$  is the natural dual pairing on  $C_b(\mathcal{P}_p(C)) \times \mathcal{P}(\mathcal{P}_p(C))$ . Using Lemma 7, one can extract an other subsequence, which we still denote by  $((M_{t_n}^n)_\# \mathbb{P})$ , converging to  $Y \in \mathcal{M}$ . As a consequence,

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

where we used the fact that, due to Proposition 11,  $\int W_p(m, \text{BC}_p) dM(m) = 0$  for  $Y$ -almost all  $M$ . This contradicts Eq. (38).  $\square$

#### 5.4. Proof of Corollary 1

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

$$g_1(\mu) := W_p(\mu, \text{BC}_p^0),$$

and

$$g_2(\mu) := W_p(\mu, \text{BC}_p^0)^p.$$

Consider the r.v.

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

for  $\ell \in [2]$ .

**Lemma 10.** 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  $Y_{n,\ell}(s) \leq C(1 + \frac{1}{n} \sum_i \|\tilde{X}_s^{i,n}\|^p)$ . Hence for a convex, and increasing function  $F : \mathbb{R}_+^* \rightarrow \mathbb{R}$ , by the exchangeability stated in Assumption 3, we obtain  $\mathbb{E}(F(Y_{n,\ell}(s))) \leq \mathbb{E}(F(C(1 + \frac{1}{n} \sum_i \|\tilde{X}_s^{i,n}\|^p))) \leq F(C(1 + \mathbb{E}(\|\tilde{X}_s^{1,n}\|^p)))$ . 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  $(\|\tilde{X}_s^{1,n}\|^p : s > 0, n \in \mathbb{N})$  are uniformly integrable. We conclude using Lemma 4.  $\square$

Let  $\ell \in [2]$ , recall the definition of  $M_t^n$  in Eq. (36), 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  $Y \in \mathcal{M}$ . Consider  $\varepsilon > 0$ . By Lemma 10, there exists  $a > 0$  such that  $\sup_{n,s} \mathbb{E}(Y_{n,\ell}(s) \mathbb{1}_{Y_{n,\ell}(s) > a}) < \varepsilon$ . Using the inequality  $y \leq a \wedge y + y \mathbb{1}_{y > a}$ , we obtain:

$$\mathbb{E} \left( \frac{1}{t_n} \int_0^{t_n} Y_{n,\ell}(s) ds \right) \leq \mathbb{E} \left( \frac{1}{t_n} \int_0^{t_n} a \wedge Y_{n,\ell}(s) ds \right) + \varepsilon$$

$$= \mathbb{E} \left( \int a \wedge g_{\ell}((\pi_0)_{\#} m) dM_{t_n}^n(m) \right) + \varepsilon \quad (39)$$

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

$$\iint g_{\ell}((\pi_0)_{\#} m) dM(m) dY(M) = 0.$$

Hence, by letting  $n \rightarrow \infty$  in Eq. (39), we obtain  $\limsup_n \mathbb{E}(\frac{1}{t_n} \int_0^{t_n} Y_n(s) ds) \leq \varepsilon$ . As  $\varepsilon$  is arbitrary,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \frac{1}{t_n} \int_0^{t_n} Y_n(s) ds \right) = 0. \quad (40)$$

In order to establish the statement of [Corollary 1](#), we now should consider replacing the integral in Eq. (40) 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. (40) 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,

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

where we used [Assumption 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 Corollary 2

Let  $A \subset \mathcal{P}_p(\mathbb{R}^d)$ , we define

$$\text{conv}(A) := \left\{ \sum_{i \in [n]} \lambda_i \mu_i : (\mu_i)_{i \in [n]} \in A^n, \lambda_i > 0, \sum_{i \in [n]} \lambda_i = 1, n \in \mathbb{N}^* \right\}.$$

Let  $(\mu_i, \nu_i)_{i \in [2]}$  be measures in  $\mathcal{P}_p(\mathbb{R}^d)$ , and let  $\lambda \in [0, 1]$ . We claim that

$$W_p(\lambda \mu_1 + (1 - \lambda) \mu_2, \lambda \nu_1 + (1 - \lambda) \nu_2)^p \leq \lambda W_p(\mu_1, \nu_1)^p + (1 - \lambda) W_p(\mu_2, \nu_2)^p. \quad (41)$$

Indeed let  $(\pi_1^\varepsilon, \pi_2^\varepsilon) \in \Pi(\mu_1, \nu_1) \times \Pi(\mu_2, \nu_2)$  satisfying for  $i \in [2]$ :

$$\left| \int \|x - y\|^p d\pi_i^\varepsilon(x, y) - W_p(\mu_i, \nu_i)^p \right| \leq \varepsilon.$$

Since  $\lambda \pi_1^\varepsilon + (1 - \lambda) \pi_2^\varepsilon \in \Pi(\lambda \mu_1 + (1 - \lambda) \mu_2, \lambda \nu_1 + (1 - \lambda) \nu_2)$ , we obtain

$$\begin{aligned} & W_p(\lambda \mu_1 + (1 - \lambda) \mu_2, \lambda \nu_1 + (1 - \lambda) \nu_2)^p \\ & \leq \lambda \int \|x - y\|^p d\pi_1^\varepsilon(x, y) + (1 - \lambda) \int \|x - y\|^p d\pi_2^\varepsilon(x, y) \\ & \leq \lambda W_p(\mu_1, \nu_1)^p + (1 - \lambda) W_p(\mu_2, \nu_2)^p + 2\varepsilon. \end{aligned}$$

Since it is true for every  $\varepsilon > 0$ , this proves our claim.

Now, let  $A \subset \mathcal{P}_p(\mathbb{R}^d)$ , there exists  $\nu_1^\varepsilon, \nu_2^\varepsilon \in A$  satisfying

$$\lambda W_p(\mu_1, A)^p + (1 - \lambda) W_p(\mu_2, A)^p \geq \lambda W_p(\mu_1, \nu_1^\varepsilon)^p + (1 - \lambda) W_p(\mu_2, \nu_2^\varepsilon)^p - 2\varepsilon.$$

Since this is true for every  $\varepsilon > 0$ , by Eq. (41):

$$\lambda W_p(\mu_1, A)^p + (1 - \lambda) W_p(\mu_2, A)^p \geq W_p(\lambda \mu_1 + (1 - \lambda) \mu_2, \text{conv}(A))^p. \quad (42)$$

Applying Eq. (42) to the second claim of Corollary 1, we obtain

$$W_p \left( \frac{\sum_{l \in [k]} \gamma_l \mu_l^n}{n \sum_{l \in [k]} \gamma_l}, \text{conv}((\pi_0)_\#(\text{BC}_p)) \right) \xrightarrow[(k, n) \rightarrow (\infty, \infty)]{\mathbb{P}} 0.$$

Since  $W_p(\cdot, \cdot) \geq W_1(\cdot, \cdot)$ , we can apply the Kantorovich duality theorem and interchange the inf and sup to obtain our result.

### 5.6. Proof of Corollary 3

Let  $i \leq n$ . We define the  $\tilde{I}_t^{i,n}$  as

$$\tilde{I}_t^{i,n} := (\tilde{X}_t^{1,n}, \dots, \tilde{X}_t^{i,n})_\# \mathbb{P}.$$

Define the measure

$$J_t^{i,n} = \frac{1}{t} \int_0^t \delta_{\tilde{I}_s^{i,n}} ds. \quad (43)$$

We define the measure  $\tilde{J}_t^{i,n} \in \mathcal{P}_p((\mathbb{R}^d)^i)$

$$\tilde{J}_t^{i,n}(A) := \int \mu(A) dJ_t^{i,n}(\mu),$$

for every  $A \in \mathcal{B}((\mathbb{R}^d)^i)$ . Recalling the definition of  $M_t^n$  in Eq. (36), we remark that

$$(\pi_0)_\# M_t^n = \frac{1}{t} \int_0^t \delta_{m_s^n} ds.$$

We define the measure  $\tilde{M}_t^{i,n} \in \mathcal{P}_p((\mathbb{R}^d)^i)$  as

$$\tilde{M}_t^{i,n}(A) := \mathbb{E} \left( \int \mu^{\otimes i}(A) d(\pi_0)_\# M_t^n(\mu) \right),$$

for every  $A \in \mathcal{B}((\mathbb{R}^d)^i)$ .

**Lemma 11.** *There exists a constant  $C$ , independent of  $t$  and  $n$ , such that*

$$\sup_{A \in \mathcal{B}((\mathbb{R}^d)^i)} \left| \tilde{M}_t^{i,n}(A) - \tilde{J}_t^{i,n}(A) \right| \leq \frac{C}{n},$$

for every  $t, n$ .

**Proof.** First we assume that  $A = (A_1, \dots, A_i) \in (\mathcal{B}(\mathbb{R}^d))^i$ . Remark that

$$\tilde{J}_t^{i,n}(A) = \frac{1}{t} \int_0^t \mathbb{P}((\bar{X}_s^{1,n}, \dots, \bar{X}_s^{i,n}) \in (A_1, \dots, A_i)) ds,$$

and

$$\tilde{M}_t^{i,n}(A) = \frac{1}{tn^i} \int_0^t \sum_{j_1, \dots, j_i=1}^n \mathbb{P}((\bar{X}_s^{j_1,n}, \dots, \bar{X}_s^{j_i,n}) \in (A_1, \dots, A_i)) ds.$$

By exchangeability,

$$|\tilde{M}_t^{i,n}(A) - \tilde{J}_t^{i,n}(A)| \leq \frac{C}{n},$$

for our specific choice of  $A$  and a constant  $C$  independent of  $A$ . We conclude by a density argument.  $\square$

Since the total variation distance is greater than the Lévy-Prokhorov distance denoted by  $d_L$ , by the triangular inequality and [Lemma 11](#)

$$d_L(\tilde{J}_t^{i,n}, (\rho_0^*)^{\otimes i}) \leq d_L(\tilde{M}_t^{i,n}, (\rho_0^*)^{\otimes i}) + \frac{C}{n}.$$

By [Assumption 5](#) and [Proposition 11](#), we obtain  $\mathcal{M} = \{\delta_{\rho_0^*}\}$ . Consequently, for every  $(t_n, \varphi_n) \rightarrow (\infty, \infty)$ ,  $d_L(\tilde{M}_{t_n}^{i,\varphi_n}, (\rho_0^*)^{\otimes i}) \rightarrow 0$ , which means that  $\tilde{J}_{t_n}^{i,n}$  converges to  $(\rho_0^*)^{\otimes i}$  in  $\mathcal{P}((\mathbb{R}^d)^i)$ . By [1, Lem. 3.14],  $J_{t_n}^{i,\varphi_n}$  converges to  $\delta_{(\rho_0^*)^{\otimes i}}$  in  $\mathcal{P}(\mathcal{P}((\mathbb{R}^d)^i))$ .

By an application of [Proposition 4](#) with [Lemma 2](#),  $\{J_{t_n}^{i,\varphi_n} : t \geq 0, n \in \mathbb{N}\}$  is a compact subspace of  $\mathcal{P}(\mathcal{P}_p((\mathbb{R}^d)^i))$ . Consequently, for every  $(t_n, \varphi_n) \rightarrow (\infty, \infty)$ ,  $J_{t_n}^{i,\varphi_n}$  converges to  $\delta_{(\rho_0^*)^{\otimes i}}$  in  $\mathcal{P}(\mathcal{P}_p((\mathbb{R}^d)^i))$ . The conclusion follows from the same proof as in [Corollary 1](#).

### 5.7. Proof of [Corollary 4](#)

In the proof of [Corollary 3](#), we showed that for every subsequence  $(t_n, \varphi_n) \rightarrow (\infty, \infty)$ ,  $J_{t_n}^{1,\varphi_n} \rightarrow \delta_{\rho_0^*}$ .

Let  $U \subset \mathcal{P}_p(\mathbb{R}^d)$  be an open neighborhood of  $\rho_0^*$ . Then, by the Portmanteau theorem,

$$\limsup_{n \rightarrow \infty} J_{t_n}^{1,\varphi_n}(U) \geq \liminf_{n \rightarrow \infty} J_{t_n}^{1,\varphi_n}(U) \geq \mathbb{1}_{\rho_0^* \in U} = 1,$$

for every  $(t_n, \varphi_n) \rightarrow (\infty, \infty)$ . By similar arguments as in [Corollary 1](#),  $\rho_0^*$  is an essential accumulation point of  $(I_k^{1,n})_{k,n}$ .

Let  $\tilde{\mu}$  be a essential accumulation point of  $(I_k^{1,n})_{k,n}$ . Then, by similar arguments as in [Corollary 1](#), for every open neighborhood  $U \subset \mathcal{P}_p(\mathbb{R}^d)$  of  $\tilde{\mu}$

$$\limsup_{n \rightarrow \infty} J_{t_n}^{1,\varphi_n}(U) > 0,$$

for some  $(t_n, \varphi_n) \rightarrow (\infty, \infty)$ . Assume that  $\tilde{\mu} \neq \rho_0^*$ . Define the closed set  $F_0 := \{\mu \in \mathcal{P}_p(\mathbb{R}^d) : W_p(\mu, \tilde{\mu}) \leq W_p(\rho_0^*, \tilde{\mu})/2\}$ . The open set  $U_0 = \{\mu \in \mathcal{P}_p(\mathbb{R}^d) : W_p(\mu, \tilde{\mu}) < W_p(\rho_0^*, \tilde{\mu})/2\}$  is a neighborhood of  $\tilde{\mu}$  satisfying  $U_0 \subset F_0$ . Then by Portmanteau theorem

$$0 = \mathbb{1}_{\rho_0^* \in F_0} \geq \limsup_{n \rightarrow \infty} J_{t_n}^{1,\varphi_n}(F_0) \geq \limsup_{n \rightarrow \infty} J_{t_n}^{1,\varphi_n}(U_0) > 0,$$

for every  $(t_n, \varphi_n) \rightarrow (\infty, \infty)$ . This contradicts our claim:  $\tilde{\mu} \neq \rho_0^*$ . Consequently,  $\rho_0^*$  is the unique accumulation point of  $(I_k^{1,n})_{k,n}$ .

### 5.8. Proof of [Theorem 2](#)

We let the assumptions of the theorem hold.

**Lemma 12.** For a nonempty compact set  $K \subset \mathcal{P}_p(\mathbb{R}^d)$ , it holds that

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

**Proof.** Assume for the sake of contradiction that

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

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  $v_\infty \in K$  such that  $v_n \rightarrow_n v_\infty$ . Since  $A_p$  is a global attractor, there exists  $T > 0$  such that  $W_p(\Psi_T(v_\infty), A_p) \leq \delta/2$ . Furthermore, by the continuity of  $\Psi$ , there exists  $n_0$  such that

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

This implies that  $\Psi_T(v_n) \in A_p^\delta$  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

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

and we obtain our contradiction.  $\square$



We now prove [Theorem 2](#). Recall that the collection  $\{\Phi_t(m^n)\}$  is tight in  $\mathcal{P}_p(C)$  by [Proposition 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 (14). To prove [Theorem 2](#), it will be enough to show that

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

This shows indeed that

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

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  $D \subset \mathcal{P}_p(C)$  such that  $\mathbb{P}(\Phi_t(m^n) \in D) \geq 1 - \varepsilon/2$  for each couple  $(t, n)$ . This implies that  $M(D) \geq 1 - \varepsilon/2$  by the Portmanteau theorem. Since  $V_p$  is closed by [Proposition 5](#), the set  $\mathcal{K} = D \cap V_p$  is compact in  $\mathcal{P}_p(C)$ , and by consequence, it is compact in  $V_p$  for the trace topology. By the same proposition,  $M(V_p) = 1$ , therefore,  $M(\mathcal{K}) \geq 1 - \varepsilon/2$ .

Since  $\mathcal{P}_p(C)$  is Polish, we can apply Skorokhod's representation theorem [44, 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  $\mathcal{P}_p(C)$ -valued random variables  $(\tilde{m}^n)$  on  $\tilde{\Omega}$  and a  $\mathcal{P}_p(C)$ -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  $\tilde{m}_{t_n+T}^n$  have the same probability distribution as  $\mathcal{P}_p(\mathbb{R}^d)$ -valued random variables, we show that

$$\exists T > 0, \limsup_n \tilde{\mathbb{P}} \left( W_p \left( \tilde{m}_{t_n+T}^n, A_p \right) \geq \delta \right) \leq \varepsilon. \quad (44)$$

to establish our theorem. Observing that the function  $\rho \mapsto (\pi_0)_\# \rho$  is a continuous  $\mathcal{P}(C) \rightarrow \mathcal{P}_p(\mathbb{R}^d)$  function, the set  $K = (\pi_0)_\# \mathcal{K}$  is a nonempty compact set of  $\mathcal{P}_p(\mathbb{R}^d)$ . Applying [Lemma 12](#) to the semi-flow  $\Psi$  and to the compact  $K$ , we set  $T > 0$  in such a way that

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

By the triangular inequality, we have

$$W_p \left( \tilde{m}_{t_n+T}^n, A_p \right) \leq W_p \left( \tilde{m}_{t_n+T}^n, \tilde{m}_T^\infty \right) + W_p \left( \tilde{m}_T^\infty, A_p \right).$$

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  $\tilde{m}^\infty \in V_p$ . 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

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

When  $\tilde{m}_0^\infty \in K$ , it holds that  $W_p \left( \Psi_T(\tilde{m}_0^\infty), A_p \right) \leq \delta/2$ , thus, the second term at the right hand side of the last inequality is zero. The first term satisfies  $\tilde{\mathbb{P}} \left( \tilde{m}^\infty \notin K \right) = 1 - M(\mathcal{K}) \leq \varepsilon/2$ , and the statement (44) follows. [Theorem 2](#) is proven.

## 6. Proofs of Section 4

The [Assumption 6](#) and  $\sigma > 0$  are standing in this section.

### 6.1. Proof of [Proposition 6](#)

**Lemma 13.** Let  $\rho \in V_2$ . 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:

$$\inf_{t \in [t_1, t_2], \|x\| \leq R} \varrho(t, x) \geq C_{R, t_1, t_2}, \quad (45)$$

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

$$\sup_{x \in \mathbb{R}^d, t \in [t_1, t_2]} \|\nabla \varrho(t, x)\| + \varrho(t, x) \leq C_{t_1, t_2}. \quad (46)$$

Finally,

$$\sup_{t \in [t_1, t_2]} \int (1 + \|x\|^2) \|\nabla \varrho(t, x)\| dx < \infty. \quad (47)$$

**Proof.** The result is an application of Th.1.2 in [48] 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 6](#), for every  $(x, y, T) \in (\mathbb{R}^d)^2 \times \mathbb{R}_+$ ,

$$\sup_{t \in [0, T]} \|\tilde{b}(t, x) - \tilde{b}(t, y)\| \leq \|\nabla V(x) - \nabla V(y)\|$$

$$+ \sup_{t \in [0, T]} \int \|\nabla U(x - z) - \nabla U(y - z)\| d\rho_t(z) \\ \leq C(\|x - y\|^\beta \vee \|x - y\|),$$

Moreover,

$$\sup_{t \in [0, T]} \tilde{b}(t, x) \leq C(1 + \|x\| + \int \sup_{t \in [0, T]} \|y_t\| d\rho(y)) \leq C(1 + \|x\|). \quad (48)$$

As  $\sigma > 0$ , [48, Th. 1.2] applies:  $\rho$  admits a density  $x \mapsto \rho(t, x) \in C^1(\mathbb{R}^d)$ , for  $0 < t \leq T$ , and there exists four constants  $(C_{i,T}, \lambda_{i,T})_{i \in [2]}$ , such that:

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

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. (48), 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:

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

and  $\|\nabla \rho(t, x)\| \leq C_{2,t_2} t_1^{-(d+1)/2}$ . Consequently,  $\rho$  satisfies Eq. (45), Eq. (46) and (47).  $\square$

For every  $\rho \in \mathbb{V}_2$  and every  $t > 0$ , recall the definition of the velocity field  $v_t$  in Eq. (19):  $v_t(x) := -\nabla V(x) - \int \nabla U(x, y) d\rho_t(y) - \sigma^2 \nabla \log \rho(t, x)$ , where  $\rho(t, x)$  is the density of  $\rho_t$  defined in Lemma 13.

**Lemma 14.** For every  $\rho \in \mathbb{V}_2$ , and every  $t_2 > t_1 > 0$ ,

$$\int_{t_1}^{t_2} \int \|v_t(x)\| d\rho_t(x) dt < \infty. \quad (49)$$

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

$$\int \psi(t_2, x) d\rho_{t_2}(x) - \int \psi(t_1, x) d\rho_{t_1}(x) \\ = \int_{t_1}^{t_2} \int (\partial_t \psi(t, x) + \langle \nabla_x \psi(t, x), v_t(x) \rangle) \rho_t(dx) dt. \quad (50)$$

**Proof.** The first point is a direct consequence of Lemma 13. 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  $\psi(t, x) = g(t)\phi(x)$ ,

$$\int \psi(t_2, x) d\rho_{t_2}(x) - \int \psi(t_1, x) d\rho_{t_1}(x) = \\ \int_{t_1}^{t_2} \int (\partial_t \psi(t, x) + \langle \nabla \psi(s, x), b(x, \rho_t) \rangle + \sigma^2 \Delta \psi(t, x)) \rho_t(dx) dt. \quad (51)$$

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. (51) holds in fact for any smooth compactly supported  $\psi$ . Using Lemma 13 and an integration by parts of the Laplacian term, Eq. (50) 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. (50) to the functional  $(t, x) \mapsto \sigma^2 \log(\rho(t, x)) + V(x) + \int U(x - y)\rho(t, y)dy$ . However, this function is not necessarily smooth nor compactly supported. In order to be able to apply Lemma 14, 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  $\rho_\varepsilon(t, \cdot) := \eta_\varepsilon * \rho_\varepsilon(t, \cdot)$ , and we denote by  $\rho_t^\varepsilon(dx) = \rho_\varepsilon(t, x)dx$  the corresponding

probability measure. Finally, we define:

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

With these definitions at hand, it is straightforward to check that the statements of Lemma 14 hold when  $\rho_t, v_t$  are replaced by  $\rho_t^\varepsilon, v_t^\varepsilon$ . More specifically, we shall apply Eq. (50) 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:

$$\begin{aligned} & \int \psi_{\varepsilon, \delta, R}(t_2, x) \rho_\varepsilon(t_2, x) dx - \int \psi_{\varepsilon, \delta, R}(t_1, x) \rho_\varepsilon(t_1, x) dx = \\ & \int_{t_1}^{t_2} \int (\partial_t \psi_{\varepsilon, \delta, R}(t, x) + \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^\varepsilon(x) \rangle) \rho_\varepsilon(t, x) dx dt. \end{aligned} \quad (52)$$

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  $\rho^{\varepsilon, \delta}(\cdot, x) := \theta_\delta * \rho^\varepsilon(\cdot, x)$ . The map  $t \mapsto \rho^{\varepsilon, \delta}(t, \cdot)$  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:

$$\psi_{\varepsilon, \delta, R}(t, x) := (\sigma^2 \log \rho^{\varepsilon, \delta}(t, x) + V_\varepsilon(x) + \int U_\varepsilon(x - y) \chi_R(y) \rho^{\varepsilon, \delta}(t, y) dy) \chi_R(x). \quad (53)$$

We extend  $\psi_{\varepsilon, \delta, R}$  to a smooth compactly supported function on  $\mathbb{R}_+ \times \mathbb{R}^d$ , and we apply Eq. (52) to the latter. We now investigate the limit of both sides of the equality (52) as  $\delta, \varepsilon, R$  successively tend to  $0, 0, \infty$ . First consider the lefthand side. Note that for all  $t \in [t_1, t_2]$ ,

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

The domination argument that allows to interchange limits and integrals is provided by Lemma 13. Indeed, for a fixed  $R > 0$ , there exists a constant  $C_R$  such that  $\rho^{\varepsilon, \delta}(t, x) \leq C_R$  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,

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

Since  $\rho_t \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\int \rho(t, x) |\log \rho(t, x)| dx < \infty$ , and the first term in the r.h.s. of the above equation converges to  $\sigma^2 \int \rho(t, x) \log \rho(t, x) dx$  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 6, 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  $t \in [t_1, t_2]$ ,

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

recalling  $\mathcal{H}(\rho_t) := \sigma^2 \int \log \rho(t, \cdot) d\rho_t + \int V d\rho_t + \frac{1}{2} \int \int U(x - y) d\rho_t(y) d\rho_t(x)$ . As  $\delta, \varepsilon, R$  successively tend to  $0, 0, \infty$ , we have shown that the l.h.s. of Eq. (52) converges to:

$$\begin{aligned} & \mathcal{H}(\rho_{t_2}) - \mathcal{H}(\rho_{t_1}) \\ & + \frac{1}{2} \int \int U(x - y) d\rho_{t_2}(y) d\rho_{t_2}(x) - \frac{1}{2} \int \int U(x - y) d\rho_{t_1}(y) d\rho_{t_1}(x). \end{aligned} \quad (54)$$

We should now identify the above term with the limit of the r.h.s. of Eq. (52) in the same regime. The latter is composed of two terms. First consider the second term:

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

We can let  $\delta \rightarrow 0$  in this equation and interchange the limit and the integral. This is justified by Lemma 13, which implies that for every  $R > 0$ , there exists a constant  $C_R$  such that for every  $\varepsilon > 0$ ,  $\delta \in (0, 1)$ ,  $t \in [t_1, t_2]$ ,  $x \in \mathbb{R}^d$ ,

$$\|\nabla \psi_{\varepsilon, \delta, R}(t, x)\| \leq C_R. \quad (55)$$

Using Eq. (55) along with Eq. (49), the dominated convergence applies. Letting  $\varepsilon \rightarrow 0$  in a second step, the exact same argument applies, and we obtain:

$$\lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^\varepsilon(x) \rangle \rho_\varepsilon(t, x) dx dt$$

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

where the interchange between  $\nabla$  and the limits is again a consequence of [Lemma 13](#). We now write the gradient in the above inner product. Note that:

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

We obtain:

$$\begin{aligned} &\lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^\varepsilon(x) \rangle \rho_\varepsilon(t, x) dx dt = \\ &- \int_{t_1}^{t_2} \int \|v_t(x)\|^2 \chi_R(x) \rho(t, x) dx dt \\ &- \int_{t_1}^{t_2} \int \langle v_t(x), \int (1 - \chi_R(y)) \nabla U(x - y) d\rho_t(y) \rangle \chi_R(x) d\rho_t(x) \\ &- \int_{t_1}^{t_2} \int \langle v_t(x), \nabla \chi_R(x) (V(x) + \int U(x - y) \chi_R(y) d\rho_t(y)) \rangle d\rho_t(x). \end{aligned} \quad (56)$$

By the dominated convergence theorem, [Assumption 6](#) and Eq. (47), the last two terms in the r.h.s. of Eq. (56) tend to zero as  $R \rightarrow \infty$ , while the first term is handled by the monotone convergence theorem. We thus obtain:

$$\begin{aligned} &\lim_{R \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon, \delta, R}(t, x), v_t^\varepsilon(x) \rangle \rho_\varepsilon(t, x) dx dt \\ &= - \int_{t_1}^{t_2} \int \|v_t(x)\|^2 \rho(t, x) dx dt. \end{aligned} \quad (57)$$

As a last step, we should evaluate the limit of the first term in the r.h.s. of Eq. (52), which writes:  $\int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon, \delta, R}(t, x) \rho_\varepsilon(t, x) dx dt$ . 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 15.** *Let  $t_2 > t_1 > 0$  be fixed. For every  $R, \varepsilon > 0$ , there exists a constant  $C_{R, \varepsilon}$  such that for every  $\delta \in (0, 1)$ ,  $t \in [t_1, t_2]$ ,  $x \in \mathbb{R}^d$ ,*

$$|\partial_t \psi_{\varepsilon, \delta, R}(t, x)| \leq C_{R, \varepsilon}, \quad (58)$$

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

By Eq. (58) and by the continuity of the map  $t \mapsto \partial_t \rho^\varepsilon$  (see the proof of [Lemma 15](#)), we can expand the first term in the r.h.s. of Eq. (52) as:

$$\int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon, \delta, R}(t, x) d\rho_t^\varepsilon(x) dt = \int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon, \delta, R}(t, x) \rho^{\varepsilon, \delta}(t, x) dx dt + o_{\varepsilon, R}(\delta), \quad (59)$$

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

$$\partial_t \psi_{\varepsilon, \delta, R}(t, x) = \sigma^2 \frac{\partial_t \rho^{\varepsilon, \delta}(t, x)}{\rho^{\varepsilon, \delta}(t, x)} \chi_R(x) + \int U_\varepsilon(x - y) \chi_R(y) \chi_R(x) \partial_t \rho^{\varepsilon, \delta}(t, y) dy. \quad (60)$$

Plugging this equality into (59) and noting that  $U_\varepsilon$  is even (because  $U$  and  $\eta_\varepsilon$  are), we obtain:

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

By the dominated convergence theorem, we finally obtain:

$$\lim_{R \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \int_{t_1}^{t_2} \int \partial_t \psi_{\varepsilon, \delta, R}(t, x) d\rho_t^\varepsilon(x) dt =$$

$$\frac{1}{2} \iint U(x-y) \varphi(t_2, x) \varphi(t_2, y) dx dy - \frac{1}{2} \iint U(x-y) \varphi(t_1, x) \varphi(t_1, y) dx dy. \quad (61)$$

Putting together Eq. (54), (57) and (61), and passing to the limit in the continuity equation (52), the statement of Proposition 6 follows.

**Proof of Lemma 15.** Using Eq. (52) and integration by parts,

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

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

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

Considering the second term in the r.h.s. of Eq. (60), 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:

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

On the otherhand, using the lower bound (45), the first term in the r.h.s. of Eq. (60), is also bounded, and finally, Eq. (58) follows.

## 6.2. Proof of Proposition 7

The map  $\overline{\mathcal{H}} : \rho \mapsto \mathcal{H}(\rho_\varepsilon)$  is real valued and lower semicontinuous by Proposition 6 and Fatou's lemma. Moreover, for every  $\rho \in \mathcal{V}_2$ ,  $\mathcal{H}(\Phi_t(\rho)) - \mathcal{H}(\rho) = \mathcal{H}(\rho_{t+\varepsilon}) - \mathcal{H}(\rho_\varepsilon) = - \int_{\varepsilon}^{t+\varepsilon} \int \|v_s\|^2 d\rho_s ds$ . Therefore,  $\mathcal{H}(\Phi_t(\rho))$  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  $t \geq \varepsilon$ . By Lemma 14, this implies that  $\rho_t = \rho_\varepsilon$  for all  $t \geq \varepsilon$ . Thus,  $\overline{\mathcal{H}}(\Phi_t(\rho)) = \overline{\mathcal{H}}(\rho)$  for all  $t$ , if and only if  $v_\varepsilon = 0$  and  $\rho_t = \rho_\varepsilon$  for all  $t$ . This means that  $\overline{\mathcal{H}}$  is a Lyapunov function for the set  $A_\varepsilon$ . The first point is proven.

Consider a recurrent point  $\rho \in \mathcal{V}_2$ , say  $\rho = \lim \Phi_{t_n}(\rho)$ . By Proposition 3,  $\rho \in A_\varepsilon$ , for any  $\varepsilon > 0$ . 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  $\rho \in A_0$ . The proof is complete.

## 6.3. Proof of Proposition 8

Since  $\beta = 1$ , we obtain by Assumption 6 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 [22, Th. 11.2.8] (see also, e.g., [49]): 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 [22, 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  $\rho \in \mathcal{V}_2$ , the function  $t \mapsto \rho_t$  satisfies the five properties stated above. The first two hold true for each  $\zeta \in \mathcal{P}_2(C)$ : 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  $\mathcal{P}_2(C)$ . Property 3 follows from Lemma 13. By Lemma 14, 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, Eq. (18). This completes the proof of Proposition 8.

#### 6.4. Proof of Proposition 9

First, we will show Eq. (12). Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$\int \langle x, b(x, \mu) \rangle d\mu(x) = - \int \langle x, \nabla V(x) \rangle d\mu(x) - \int \int \langle x, \nabla U(x-y) \rangle d\mu(x) d\mu(y).$$

Since  $U$  is even,  $\nabla U(-x) = -\nabla U(x)$ . Therefore,

$$\int \int \langle x, \nabla U(x-y) \rangle d\mu(x) d\mu(y) = \frac{1}{2} \int \int \langle x-y, \nabla U(x-y) \rangle d\mu(x) d\mu(y).$$

Recalling that  $\langle \nabla U(x), x \rangle \geq -C$  and  $\langle x, \nabla V(x) \rangle \geq \lambda \|x\|^2$ , Eq. (12) holds:

$$\int \langle x, b(x, \mu) \rangle d\mu(x) \leq -\lambda \int \|x\|^2 d\mu(x) + C.$$

Eq. (13) is obtained by the same computation as above, where in addition, we used  $\|\nabla U(x)\| \leq C(1 + \|x\|)$ . Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$\begin{aligned} & \int \langle x, b(x, \mu) \rangle \|x\|^2 d\mu(x) \\ &= - \int \langle x, \nabla V(x) \rangle \|x\|^2 d\mu(x) - \int \int \langle x, \nabla U(x-y) \rangle \|x\|^2 d\mu(x) d\mu(y) \\ &= - \int \langle x, \nabla V(x) \rangle \|x\|^2 d\mu(x) - \int \int \langle x-y, \nabla U(x-y) \rangle \|x\|^2 d\mu(x) d\mu(y) \\ &\quad - \int \int \langle y, \nabla U(x-y) \rangle \|x\|^2 d\mu(x) d\mu(y) \\ &\leq -\lambda \int \|x\|^4 d\mu(x) + C \int \|x\|^2 d\mu(x) + C \int \int \|x\|^2 \|y\| d\mu(x) d\mu(y) \\ &\quad + C \int \int \|x\|^3 \|y\| d\mu(x) d\mu(y) + C \int \int \|x\|^2 \|y\|^2 d\mu(x) d\mu(y). \end{aligned}$$

By Cauchy–Schwarz’s inequality,

$$\begin{aligned} \int \int \|x\|^3 \|y\| d\mu(x) d\mu(y) &\leq \int \|x\|^2 d\mu(x) \left( \int \|x\|^4 d\mu(x) \right)^{1/2}, \\ \int \int \|x\|^2 \|y\| d\mu(x) d\mu(y) &\leq \int \|x\|^2 d\mu(x) \left( \int \|x\|^4 d\mu(x) \right)^{1/4}. \end{aligned}$$

Therefore, we obtain Eq. (13)

$$\begin{aligned} & \int \langle x, b(x, \mu) \rangle \|x\|^2 d\mu(x) \\ &\leq -\lambda \int \|x\|^4 d\mu(x) + C \left( 1 + \int \|x\|^2 d\mu(x) \right) \left( 1 + \left( \int \|x\|^4 d\mu(x) \right)^{1/2} \right). \end{aligned}$$

#### 6.5. Proof of Theorem 4

The convergence provided in the statement follows at once from Proposition 8 and Theorem 2. We need to prove that  $S = A_2$  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 Eq. (18) 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  $\rho^\infty \in S$ . On the other hand, for  $\nu \neq \rho^\infty$  in  $\mathcal{P}_2(\mathbb{R}^d)$ , we obtain from Eq. (18) that the function  $t \mapsto \mathcal{H}(\Psi_t(\nu))$  is strictly decreasing. Thus,  $\int \|u_\nu\|^2 d\nu > 0$  which shows that  $\nu \notin S$ .

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Victor Priser reports financial support was provided by chair DSAIDIS. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgement

The third author thanks the DSAIDIS Chair for funding his work.

## Appendix. 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:

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

for all  $k \in \mathbb{N}^*$ . Let  $\mathcal{K} := \bigcap_{k \geq 1} \pi_{[0, k]}^{-1}(K_k) \subset C$ . The union bound yields  $\rho^n(\mathcal{K}) > 1 - \varepsilon$ . Referring to [50, Th. 2, Sec. X, Chapter 5],  $\mathcal{K}$  has a compact closure in  $C$ . Hence, there exists a converging subsequence  $(\rho_{\varphi_n})$  converging to  $\rho \in \mathcal{P}(C)$ . Following the proof of [32, Th. 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  $\mathcal{P}_p(C)$ . It remains to obtain its separability.

As  $C$  is Polish, there exists a dense sequence  $(x_n)$  in  $C$ . Following the proof of [32, Th. 6.18], one can construct a sequence  $(\rho^n)$  in  $\mathcal{P}_p(C)$  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  $\mathcal{P}_p(C)$ .

### A.2. Proof of Lemma 2

Since Proposition 1 holds,  $(\mathbb{I}(\rho^n))$  is a weak\*-relatively compact sequence in  $\mathcal{P}(C)$ , and there exists a sequence of compact sets  $(K_k)$  in  $C$ , 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  $\mathcal{P}(C)$ :

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

The union bound and Markov's inequality yields:

$$\mathbb{P}(\rho^n \in \mathcal{K}_\varepsilon) > 1 - \varepsilon \tag{A.1}$$

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

To be relatively compact in  $\mathcal{P}_p(C)$ , 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

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

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

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

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

$$\mathbb{P}(\rho^n \in \mathcal{U}_\varepsilon) > 1 - \varepsilon. \tag{A.2}$$

Putting together Eqs. (A.1) and (A.2),

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

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

### 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  $\mathcal{P}_p(C)$ . This last convergence is characterized by the fact that  $\rho^n \rightarrow \rho^\infty$  in  $\mathcal{P}(C)$ , and that the sequence  $(\rho^n)$  has uniformly integrable  $p$ -moments defined by (p-UI). We write  $G(\rho^n) = \int g(x, \rho^n) d\rho^n(x)$ , where for  $x$  in  $C$  and  $\rho \in \mathcal{P}_p(C)$ :

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

and  $h(x) := \prod_{j=1}^r h_j(x_{t_j})$ .

We claim that  $g$  is a continuous bounded function on  $C \times \overline{\{\rho^n : n \in \mathbb{N}\}}$ . The continuity is given by the assumptions on  $b$ . Using Cauchy–Schwarz inequality, we state a useful inequality:

$$|g(x, \rho)| \leq C \left( 1 + \int_s^t \|\nabla \phi(x_u)\| \|b(x_u, \rho_u)\| du \right),$$

where  $C = \|h\|_\infty \max(2\|\phi\|_\infty + \|\sigma\|_\infty^2(t-s)\|H_\phi\|_\infty, 1)$ . Since  $\phi$  is compactly supported, by [Assumption 1](#):

$$b(x, \rho^n) \leq C \left( 1 + t \sup_{n \in \mathbb{N}} \int \sup_{u \in [0, t]} \|y_u\| d\rho^n(y) \right).$$

The sequence  $(\rho^n)$  has uniformly integrable  $p$ -moments in  $\mathcal{P}_p(C)$ , consequently we obtain the bound:

$$\sup_{x \in C, n \in \mathbb{N} \cup \{\infty\}} b(x, \rho^n) < \infty.$$

Let  $\varepsilon > 0$ . Since,  $\rho^n \rightarrow \rho^\infty$  in  $\mathcal{P}_p(C)$ , the set  $\overline{\{\rho^n : n \in \mathbb{N}\}}$  is a compact subspace of  $\mathcal{P}_p(C)$ . Hence, there exists a compact subspace  $\mathcal{K} \subset C$  satisfying

$$\sup_{n \in \mathbb{N} \cup \{\infty\}} \rho^n(\mathcal{K}^c) \leq \varepsilon,$$

where  $\mathcal{K}^c$  denotes set of function  $x \in C$  that doesn't belong to  $\mathcal{K}$ . By Stone–Weierstrass's theorem, there exists  $k_\varepsilon \in \mathbb{N}^*$  and continuous bounded functions  $(f_i, h_i)_{i \in [k_\varepsilon]} \in (C(C, \mathbb{R}) \times C(\overline{\{\rho^n : n \in \mathbb{N}\}}, \mathbb{R}))^{k_\varepsilon}$  satisfying

$$\forall (x, n) \in \mathcal{K} \times \mathbb{N} \cup \{\infty\}, \quad \left| \sum_{i \in [k_\varepsilon]} f_i(x) h_i(\rho^n) - g(x, \rho^n) \right| \leq \varepsilon.$$

Note that for  $n \in \mathbb{N}$ ,

$$\begin{aligned} |G(\rho^n) - G(\rho^\infty)| &\leq \left| G(\rho^n) - \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^n(x) h_i(\rho^n) \right| \\ &\quad + \left| \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^\infty(x) h_i(\rho^\infty) - G(\rho^\infty) \right| \\ &\quad + \left| \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^n(x) h_i(\rho^n) - \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^\infty(x) h_i(\rho^\infty) \right|. \end{aligned} \quad (\text{A.3})$$

For  $n \in \mathbb{N} \cup \{\infty\}$ , we decompose  $G(\rho^n)$  as follows

$$\begin{aligned} G(\rho^n) &= \int \mathbb{1}_{x \in \mathcal{K}} \left( g(x, \rho^n) - \sum_{i \in [k_\varepsilon]} f_i(x) h_i(\rho^n) \right) d\rho^n(x) \\ &\quad + \int \mathbb{1}_{x \in \mathcal{K}^c} \left( g(x, \rho^n) - \sum_{i \in [k_\varepsilon]} f_i(x) h_i(\rho^n) \right) d\rho^n(x) + \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^n(x) h_i(\rho^n). \end{aligned}$$

For every  $\varepsilon > 0$ , since  $g$  is bounded, we obtain

$$\sup_{n \in \mathbb{N} \cup \{\infty\}} \left| G(\rho^n) - \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^n(x) h_i(\rho^n) \right| \leq 2\varepsilon.$$

Consequently, using the latter result in [Eq. \(A.3\)](#), we obtain

$$\begin{aligned} |G(\rho^n) - G(\rho^\infty)| &\leq 4\varepsilon + \left| \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^n(x) h_i(\rho^n) - \sum_{i \in [k_\varepsilon]} \int f_i(x) d\rho^\infty(x) h_i(\rho^\infty) \right|. \end{aligned}$$

Since,  $f_i$  and  $h_i$  are continuous bounded functions, we obtain for every  $\varepsilon > 0$

$$\limsup_{n \rightarrow \infty} |G(\rho^n) - G(\rho^\infty)| \leq 4\varepsilon,$$

which concludes the proof.

## References

- [1] L.-P. Chaintron, A. Diez, Propagation of chaos: A review of models, methods and applications. i. models and methods, *Kinet. Relat. Model.* 15 (6) (2022) 895, <https://doi.org/10.3934/krm.2022017>.
- [2] S. Méléard, S. Roelly-Coppoletta, A propagation of chaos result for a system of particles with moderate interaction, *Stochastic Process. Appl.* 26 (1987) 317–332, URL: <https://api.semanticscholar.org/CorpusID:119750253>.



- [3] K. Oelschläger, A martingale approach to the law of large numbers for weakly interacting stochastic processes, *Ann. Probab.* 12 (1984) 458–479, URL: <https://api.semanticscholar.org/CorpusID:121628874>.
- [4] A.-S. Sznitman, Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated, *J. Funct. Anal.* 56 (3) (1984) 311–336.
- [5] X. Erny, E. Löcherbach, D. Loukianova, Conditional propagation of chaos for mean field systems of interacting neurons, *Electron. J. Probab.* 26 (none) (2021) 1–25, <http://dx.doi.org/10.1214/21-EJP580>.
- [6] M. Karimi Jaghargh, Y. Hsieh, A. Krause, Stochastic approximation algorithms for systems of interacting particles, *Adv. Neural Inf. Process. Syst.* 36 (2024).
- [7] F. Malrieu, Logarithmic Sobolev inequalities for some nonlinear PDE's, *Stochastic Process. Appl.* 95 (1) (2001) 109–132.
- [8] P. Cattiaux, A. Guillin, F. Malrieu, Probabilistic approach for granular media equations in the non-uniformly convex case, *Probab. Theory Related Fields* 140 (2008) 19–40.
- [9] A. Durmus, A. Eberle, A. Guillin, R. Zimmer, An elementary approach to uniform in time propagation of chaos, *Proc. Amer. Math. Soc.* 148 (12) (2020) 5387–5398.
- [10] P. Monmarché, Z. Ren, S. Wang, Time-uniform log-Sobolev inequalities and applications to propagation of chaos, 2024, arXiv preprint [arXiv:2401.07966](https://arxiv.org/abs/2401.07966).
- [11] F. Chen, Y. Lin, Z. Ren, S. Wang, Uniform-in-time propagation of chaos for kinetic mean field Langevin dynamics, *Electron. J. Probab.* 29 (2024) 1–43.
- [12] D. Lacker, L. Le Flem, Sharp uniform-in-time propagation of chaos, *Probab. Theory Related Fields* 187 (1–2) (2023) 443–480.
- [13] P. Del Moral, J. Tugaut, Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions, *Stoch. Anal. Appl.* 37 (6) (2019) 909–935.
- [14] S. Herrmann, J. Tugaut, Non-uniqueness of stationary measures for self-stabilizing processes, *Stochastic Process. Appl.* 120 (7) (2010) 1215–1246.
- [15] Q. Cormier, On the stability of the invariant probability measures of McKean–Vlasov equations, 2023, arXiv preprint [arXiv:2201.11612](https://arxiv.org/abs/2201.11612).
- [16] A. Budhiraja, P. Dupuis, M. Fischer, K. Ramanan, Limits of relative entropies associated with weakly interacting particle systems, *Electron. J. Probab.* 20 (2015).
- [17] F. Malrieu, Convergence to equilibrium for granular media equations and their Euler schemes, *Ann. Appl. Probab.* 13 (2) (2003) 540–560.
- [18] M. Benko, I. Chlebicka, J. Endal, B. Miasojedow, Convergence rates of particle approximation of forward-backward splitting algorithm for granular medium equations, 2024, URL: <https://arxiv.org/abs/2405.18034>, [arXiv:2405.18034](https://arxiv.org/abs/2405.18034).
- [19] A.Y. Veretennikov, On ergodic measures for McKean–Vlasov stochastic equations, in: *Monte Carlo and Quasi-Monte Carlo Methods 2004*, Springer, 2006, pp. 471–486.
- [20] M. Benaïm, S.J. Schreiber, Ergodic properties of weak asymptotic pseudotrajectories for semiflows, *J. Dynam. Differential Equations* 12 (2000) 579–598.
- [21] C. Villani, Mathematics of granular materials, *J. Stat. Phys.* 124 (2–4) (2006) 781–822.
- [22] L. Ambrosio, N. Gigli, G. Savaré, Gradient flows in metric spaces and in the space of probability measures, second ed., in: *Lectures in Mathematics ETH Zürich*, Birkhäuser Verlag, Basel, 2008, p. x+334.
- [23] Q. Liu, D. Wang, Stein variational gradient descent: A general purpose bayesian inference algorithm, *Adv. Neural Inf. Process. Syst.* (2016).
- [24] A. Salim, L. Sun, P. Richtarik, A convergence theory for SVGD in the population limit under Talagrand's inequality T1, in: *International Conference on Machine Learning*, PMLR, 2022, pp. 19139–19152.
- [25] L. Chizat, F. Bach, On the global convergence of gradient descent for over-parameterized models using optimal transport, *Adv. Neural Inf. Process. Syst.* 31 (2018).
- [26] S. Mei, A. Montanari, P.-M. Nguyen, A mean field view of the landscape of two-layer neural networks, *Proc. Natl. Acad. Sci.* 115 (33) (2018) E7665–E7671.
- [27] G. Rotskoff, E. Vanden-Eijnden, Trainability and accuracy of artificial neural networks: An interacting particle system approach, *Comm. Pure Appl. Math.* 75 (9) (2022) 1889–1935.
- [28] J. Sirignano, K. Spiliopoulos, Mean field analysis of neural networks: A central limit theorem, *Stochastic Process. Appl.* 130 (3) (2020) 1820–1852.
- [29] K. Hu, Z. Ren, D. Šiška, Ł. Szpruch, Mean-field langevin dynamics and energy landscape of neural networks, in: *Annales de l'Institut Henri Poincaré (B) Probabilités et statistiques*, vol. 57, Institut Henri Poincaré, 2021, pp. 2043–2065.
- [30] L. Chizat, Mean-field langevin dynamics: Exponential convergence and annealing, *Trans. Mach. Learn. Res.* (2022).
- [31] A. Nitanda, D. Wu, T. Suzuki, Convex analysis of the mean field langevin dynamics, in: G. Camps-Valls, F.J.R. Ruiz, I. Valera (Eds.), *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, in: *Proceedings of Machine Learning Research*, vol. 151, PMLR, 2022, pp. 9741–9757, URL: <https://proceedings.mlr.press/v151/nitanda22a.html>.
- [32] C. Villani, Optimal transport, in: *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 338, Springer-Verlag, Berlin, 2009, p. xxii+973, <http://dx.doi.org/10.1007/978-3-540-71050-9>, Old and new.
- [33] M. Benaïm, Dynamics of stochastic approximation algorithms, in: *Séminaire de Probabilités, XXXIII*, in: *Lecture Notes in Math.*, vol. 1709, Springer, 1999, pp. 1–68, <http://dx.doi.org/10.1007/BFb0096509>.
- [34] J.A. Carrillo, Y.-P. Choi, C. Totzeck, O. Tse, An analytical framework for a consensus-based global optimization method, 2018, URL: <https://arxiv.org/abs/1602.00220>, [arXiv:1602.00220](https://arxiv.org/abs/1602.00220).
- [35] M. Fornasier, T. Klock, K. Riedl, Consensus-based optimization methods converge globally, *SIAM J. Optim.* 34 (3) (2024) 2973–3004, <http://dx.doi.org/10.1137/22m1527805>.
- [36] Q. Liu, Stein variational gradient descent as gradient flow, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [37] J. Bolte, E. Pauwels, R. Rios-Zertuche, Long term dynamics of the subgradient method for Lipschitz path differentiable functions, 2020, URL: <https://arxiv.org/abs/2006.00098>, [arXiv:2006.00098](https://arxiv.org/abs/2006.00098).
- [38] P. Bianchi, R. Rios-Zertuche, A closed-measure approach to stochastic approximation, 2023, URL: <https://arxiv.org/abs/2112.05482>, [arXiv:2112.05482](https://arxiv.org/abs/2112.05482).
- [39] J.A. Carrillo, R.J. McCann, C. Villani, Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates, *Rev. Mat. Iberoam.* 19 (3) (2003) 971–1018.
- [40] J.A. Carrillo, R.J. McCann, C. Villani, Contractions in the 2-Wasserstein length space and thermalization of granular media, *Arch. Ration. Mech. Anal.* 179 (2006) 217–263.
- [41] F. Bolley, I. Gentil, A. Guillin, Uniform convergence to equilibrium for granular media, *Arch. Ration. Mech. Anal.* 208 (2013) 429–445.
- [42] A. Guillin, W. Liu, L. Wu, C. Zhang, Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems, *Ann. Appl. Probab.* 32 (3) (2022) 1590–1614, <http://dx.doi.org/10.1214/21-AAP1707>.
- [43] J.A. Carrillo, R.S. Gvalani, G.A. Pavliotis, A. Schlichting, Long-time behaviour and phase transitions for the McKean–Vlasov equation on the torus, *Arch. Ration. Mech. Anal.* 235 (1) (2020) 635–690.
- [44] P. Billingsley, Convergence of probability measures, in: *Wiley Series in Probability and Statistics: Probability and Statistics*, second ed., John Wiley & Sons, Inc., New York, 1999, p. x+277, <http://dx.doi.org/10.1002/9780470316962>, A Wiley-Interscience Publication.
- [45] D.L. Burkholder, B.J. Davis, R.F. Gundy, Integral inequalities for convex functions of operators on martingales, in: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, Univ. California Press Berkeley, Calif., 1972, pp. 223–240.
- [46] G. Leoni, A First Course in Fractional Sobolev Spaces, vol. 229, American Mathematical Society, 2023.
- [47] R. Mañé, Ergodic theory and differentiable dynamics, in: *Ergebnisse der Mathematik und ihrer Grenzgebiete : a series of modern surveys in mathematics. Folge 3*, Springer-Verlag, 1987, URL: <https://books.google.fr/books?id=0CxrgEACAAJ>.
- [48] S. Menozzi, A. Pesce, X. Zhang, Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift, *J. Differential Equations* 272 (2021) 330–369.
- [49] S. Daneri, G. Savaré, Lecture notes on gradient flows and optimal transport, 2010, arXiv preprint [arXiv:1009.3737](https://arxiv.org/abs/1009.3737).
- [50] N. Bourbaki, General topology. Chapters 5–10, in: *Elements of Mathematics (Berlin)*, Springer-Verlag, Berlin, 1989, p. iv+363, Translated from the French, Reprint of the 1966 edition.