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

Pascal Bianchi[1], Walid Hachem[2], and Victor Priser[1]

[1]LTCI, Télécom Paris, IP Paris, France
[2]CNRS, Laboratoire d'informatique Gaspard Monge (LIGM / UMR 8049), Université Gustave Eiffel, ESIEE Paris, France

March 29, 2024

Abstract

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

Contents

1 Introduction

Given two integers $n, d > 0$, consider the iterative algorithm defined as follows. Starting with the n–uple $(X_0^{1,n}, \ldots, X_0^{n,n})$ 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 $(X_k^{1,n}, \ldots, X_k^{n,n})$, referred to as the *particles*, according to the dynamics:

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

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

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

we furthermore assume that for each n, the sequence $((\xi_k^{i,n})_{i\in[n]})_k$ is a $(\mathcal{F}_k^n)_k$–martingale increment sequence *i.e.,* $\mathbb{E}(\xi_{k+1}^{i,n}|\mathcal{F}_k^n) = 0$. Finally, we assume that $\mathbb{E}(\xi_{k+1}^{i,n}(\xi_{k+1}^{j,n})^T|\mathcal{F}_k^n) = \sigma^2 \mathbb{1}_{i=j} I_d$ for some $\sigma^2 \ge 0$.

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

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

in the regime where both the time index k and the number of particles n tend to infinity (denoted hereinafter as $(k, n) \to (\infty, \infty)$), 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 $\bar{X}^{i,n} : [0, \infty) \to \mathbb{R}^d, t \mapsto \bar{X}_t^{i,n}$ defined as the piecewise linear interpolation of the particles $(X_k^{i,n})_k$. Specifically, writing

$$\tau_k := \sum_{j=1}^k \gamma_j \tag{4}$$

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

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

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

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

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 analyse the convergence in probability, of the shifted random measures

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

when both n and t converge to infinity with arbitrary relative rates, where for every $m \in \mathcal{P}_p(\mathcal{C})$, $\Phi_t(m) \in \mathcal{P}_p(\mathcal{C})$ is defined by $\Phi_t(m)(f) = \int f(x(t+\cdot))dm(x)$ for every bounded continuous function f on $\mathcal{C}$. Under mild assumptions on the vector field b, and some moment assumptions on the iterates and on the noise sequence $((\zeta^{i,n}_k)_{i\in[n]})_k$, 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 $(t,n) \to (\infty,\infty)$ 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 ρ 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 ϕ, the process

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

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

$$x \mapsto \langle b(x,\rho_t), \nabla\phi(x)\rangle + \sigma^2\Delta\phi(x)\,,$$

where Δ is the Laplacian, and where we use the slightly abusive notation $b(x,\rho_t) := \int b(x,y)d\rho_t(y)$.

A McKean-Vlasov distribution ρ is said recurrent if, for some sequence $(t_k) \to \infty$, $\rho = \lim_{k\to\infty} \Phi_{t_k}(\rho)$. The W_p-closure of the set of recurrent McKean-Vlasov distributions will be refered to as the *Birkhoff center*, and denoted by BC, 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:

$$\frac{1}{t}\int_0^t \mathsf{W}_p(\Phi_s(m^n), \mathrm{BC})ds \xrightarrow[(t,n)\to(\infty,\infty)]{} 0\,, \quad \text{in probability.}$$

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 ρ, $F(\Phi_t(\rho))$ is non-increasing in the variable t. Indeed, in such a situation, the Birkhoff center is included in the subset Λ of McKean-Vlasov distributions which satisfy the property that $t \mapsto F(\Phi_t(\rho))$ is constant whenever $\rho \in \Lambda$.

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

$$W_p\left(\mu^n_k, A_p\right) \xrightarrow[(k,n)\to(\infty,\infty)]{} 0 \quad \text{in probability.}$$

To illustrate our results, we provide two important examples of McKean-Vlasov distribution where these results can be applied.

Granular media. Our first example is in $\mathcal{P}_2(\mathcal{C})$ and corresponds to the scenario where the vector field b takes the form:

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

where V and U denote two real differentiable functions on $\mathbb{R}^d$, whose gradients satisfy some linear growth condition. In this case, a Lyapunov function exists, which can be expressed as a function of the so-called *Helmholtz energy.* As a consequence of our main result, we establish that, when $\sigma > 0$, the empirical measures (μ_k^n) converge ergodically in probability as $(k, n) \to (\infty, \infty)$ to the set $\mathcal{S}$ of critical points of the Helmholtz energy, namely:

$$\frac{\sum_{l=1}^k \gamma_l \mathsf{W}_2(\mu_k^n, \mathcal{S})}{\sum_{l=1}^k \gamma_l} \xrightarrow[(n,k)\to(\infty,\infty)]{} 0\,, \quad \text{in probability.}$$

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

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

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

Stochastic gradient descent (SGD) in two layer neural networks. An other archetypal example where a useful Lyapunov function exists is encountered in the field of Machine Learning, when studying the convergence of the popular SGD algorithm. As an illustration, we consider the problem of optimizing the coefficients of the output layer of a two layer network, assuming that the coefficients of the first/hidden layer are sampled, once for all, from a given iid distribution. This scenario is known as the *random features* setting. It captures the asymptotic regime of networks where the width n of the hidden layer goes to infinity, and the ability of the network to reach near perfect reconstruction of a target function, under some hypotheses. More specifically, the output of the neural network for an arbitrary input x of dimension q is assumed to have the form:

$$h(x; \boldsymbol{a}, \boldsymbol{w}_0^n) := \frac{1}{n} \sum_{i=1}^n a^i \varphi(x, w_0^{i,n})\,,$$

where n is the number of neurons at the output of the hidden layer, $\boldsymbol{a} = (a^1, \dots, a^n)$ are the coefficients of the output layer, $\boldsymbol{w}_0 = (w_0^{1,n}, \dots, w_0^{n,n})$ are (random but fixed) $\mathbb{R}^{d-1}$-vectors of the hidden layer, and φ is a real bounded continuous function. We consider the regularized risk minimization problem:

$$\min_{\boldsymbol{a} \in \mathbb{R}^n} \int (h(x; \boldsymbol{a}, \boldsymbol{w}_0^n) - y)^2 d\nu(x, y) + \frac{\lambda}{n} \|\boldsymbol{a}\|^2\,, \tag{7}$$

where ν is a probability measure on $\mathbb{R}^q \times \mathbb{R}$, and where $\lambda > 0$ is a regularization parameter. We assume that the distribution ν is unknown by the observer, but that iid random samples $(x_{k+1}^n, y_{k+1}^n)_{k \in \mathbb{N}}$ with distribution ν are revealed during the iterations of the algorithm. For a fixed n and a learning rate set to $n\gamma_k$, the SGD iterations generate a sequence $(a_k^{1,n}, \dots, a_k^{n,n})$ of random variables. Defining the particles as $X_k^{i,n} := (a_k^{i,n}, w_0^{i,n})$ for each $i \in [n]$, it turns out that the SGD iterations can casted into the form of Eq. (1), for a well chosen vector field b. In this case, the variables $(\zeta_{k+1}^{1,n}, \dots, \zeta_k^{n,n})$ of Eq. (1) are centered random perturbations whose first components represent the difference between the stochastic gradient derived from the new sample (x_{k+1}^n, y_{k+1}^n) and the true gradient of the objective (7). Potentially, we also include the case where a random additive noise of variance σ^2, scaled by $\sqrt{2\gamma_k}$, is artificially added at each iteration to each of the variables $(a_k^{1,n}, \dots, a_k^{n,n})$, in the flavor of a Langevin algorithm. We establish that a Lyapunov function exists, which is built upon the map $\mathcal{R}_\sigma$ given by:

$$\mathcal{R}_\sigma(\mu) := \frac{1}{2} \int \left(\int a\varphi(x, w) d\mu(a, w) - y \right)^2 d\nu(x, y) + \frac{\lambda}{2} \int a^2 d\mu(a, w) \\ + \sigma^2 \int \log\left(\frac{d\mu(\,\cdot\,|w)}{d\mathscr{L}^1}(a) \right) d\mu(a, w)\,,$$

for every probability μ on $\mathbb{R} \times \mathbb{R}^{d-1}$ which admits second order moments. Here, $\mu(da|w)$ is the conditional distribution obtained from the disintegration of μ w.r.t. its marginal in the variable w, and $\frac{d\mu(\cdot|w)}{d\mathscr{L}^1}$ represents its density w.r.t. the Lebesgue measure on $\mathbb{R}$ (setting $\mathcal{R}_\sigma(\mu) = +\infty$ when no such density exist). Specifically, a Lyapunov function can be expressed as the function which, to every McKean-Vlasov distribution, associates the value $\mathcal{R}_\sigma(\rho_\epsilon)$ for an arbitrary $\epsilon > 0$. By further studying the subset Λ of McKean-Vlasov distributions on which this Lyapunov function is constant, we obtain our main corollary. Given some prescribed distribution ϖ for the fixed parameters $w_0^{i,n}$, the empirical measure μ_k^n of the points $(a_k^{i,n}, w_0^{i,n})$ for $i \in [n]$, converges ergodically in probability to the unique minimizer of the risk $\mathcal{R}_\sigma(\mu)$ among all probability measures μ with the prescribed marginal $\mu(\cdot \times \mathbb{R}^{d-1}) = \varpi$.

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, MonteCarlo simulations or numerical solvers. In particular, stability results are more difficult to establish in this setting. Finally, our result focus on a double limit $(k, n) \to (\infty, \infty)$. 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 growth to infinity, ignoring long time convergence, or assume certain constraints on the relative rate of convergence of the two variables.

About the literature. The first results addressing the limiting behavior of a finite system of particles are provided in the context of the propagation of chaos. These findings are discussed in detail in [CD22]. Such results have broad applicability across a variety of particle systems, where the interacting term b can manifest in various forms [MRC87, Oel84, Szn84, ELL21]. In our case, if we set aside the transition from continuous to discrete time, such results typically establish the convergence to zero of the expectation of the squared Wasserstein distance between the empirical measure of the particles, over some fixed time interval $[0, T]$, and a McKean-Vlasov distribution with the same initial measure. Under classical assumptions, this convergence occurs at a rate of $1/n$, where n is the number of particles, but with a constant that grows exponentially with T. This type of result performs poorly in the long run, making the achievement of the double limit in both time and the number of particles unattainable. By imposing additional assumptions, [Mal01, BGM10, CGM08, BRTV98, DEGZ20] derive a bound that is uniform in time, thereby explicitly addressing the double asymptotic regime. However, these works require strong assumptions on the vector field b. For instance, as highlighted in [DMT19], achieving uniform propagation of chaos over time is only possible when a unique McKean-Vlasov stationary distribution exists, a condition that [Tug13] has demonstrated is not always met. In this regard, our assumptions are weaker, allowing the existence of multiple stationary distributions. It is noteworthy that the study of McKean-Vlasov stationary distributions in cases where the uniqueness of such distributions does not hold remains an open area of research. For instance, [Cor23] explore the stability of stationary distributions.

Few works address discrete-time particle systems. The paper [Mal03] employs an implicit Euler scheme for the granular media case, assuming that the potential function is zero and the interaction is strongly convex. The contribution of the paper [Ver06] is the closest to the present one, as it considers an equation very close to Eq. (1). However, this paper assumes that b is globally bounded. Moreover, it does not address the convergence in probability of the empirical measure of the particles but rather the convergence of its expectation. Lastly, another paper closely related to our work is [BS00]. This paper is not specific to the case of McKean-Vlasov processes. In particular, it does not consider a system of particles and does not address double limits. However, it establishes, in the same spirit as ours, the ergodic convergence of the empirical measure of a so-called weak asymptotic pseudotrajectory to the Birkhoff center of a flow on a metric space.

Finally, let us review some applications of our model. Particle systems have historically been motivated by statistical physics. However, in recent decades, they have found utility in various models including neural networks, Markov Chain Monte Carlo theory, mathematical biology, and mean fields game, among others. A well-known model in statistical physics is granular media

[Vil06]. This model has been extensively studied due to its property of being a gradient system, and the uniform propagation of chaos over time works well within this model. It can also be described by a gradient flow [AGS08]. In Markov Chain Monte Carlo theory, the Stein Variational Gradient Descent estimates a target distribution using a particle system [LW16, SSR22], and the convergence of this algorithm remains an open question. Wide Neural Networks can also be represented by particle systems. A convergence result to the minimizers of the risk is attainable when both time and the number of particles tend to infinity [CB18]. Here, the authors establish convergence to gradient descent in continuous time and in the double asymptotic regime. The paper [MMN18] establishes the convergence of noisy stochastic gradient descent when the number of iterations depends on the number of particles. See also [RVE22, SS20] for related works. The case where the parameters of the hidden layer are random but fixed along the optimization process is also known as the random features model [RR07, CRR18].

2 The setting

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

2.1 Notations

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 $\mathbb{R}^d \to \mathbb{R}$ 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 $\mathbb{R}^d \to \mathbb{R}$ functions which are continuously differentiable up to the order p.

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

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

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

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

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

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

We say that a n–uple of random variables $(X_1, \ldots, X_n)$ 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 $(t, n) \to (\infty, \infty)$ if, for every $\epsilon > 0$, the net $(\mathbb{P}(\mathsf{d}(U_t^n, U) > \epsilon) : t \in \mathbb{T}, n \in \mathbb{N})$ converges to zero as t and n both converge to ∞. We denote this by $U_t^n \xrightarrow[(t,n)\to(\infty,\infty)]{\mathbb{P}} U$. Moreover, assuming that the collection of random variables $(U_t^n : t \in \mathbb{T}, n \in \mathbb{N})$ are real valued, we say that the latter collection is *uniformly integrable* if:

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

Finally, for any $d \in \mathbb{N}^*$, $\mathscr{L}^d$ 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 $\mathsf{d}(x, \mathcal{A}) := \inf\{\mathsf{d}(x, y) : y \in \mathcal{A}\}$, with $\inf \emptyset = \infty$. We say that a net (μ_α) converges to $\mathcal{A}$ if $\mathsf{d}(x_\alpha, \mathcal{A}) \to_\alpha 0$.

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

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

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

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

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

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

$$W_p(\mu,\nu) := \left(\inf_{\varsigma \in \Pi(\mu,\nu)} \int \|x-y\|^p d\varsigma(x,y)\right)^{1/p}, \tag{8}$$

where $\Pi(\mu,\nu)$ is the set of measures $\varsigma \in \mathcal{P}(E \times E)$, such that $\varsigma(\,\cdot\, \times E) = \mu$ and $\varsigma(E \times \,\cdot\,) = \nu$. We denote by $\Pi_p^0(\mu,\nu)$ the set of optimal transport plans *i.e.*, the set of measures $\varsigma \in \Pi(\mu,\nu)$ achieving the infemum in Eq. (8). The set $\mathcal{P}_p(E)$ is endowed with the distance W_p. Define:

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

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

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

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

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

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

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

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

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

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

2.3 Spaces of McKean-Vlasov measures

Consider a non-negative number σ^2 and a vector field $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying the following assumption:

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

$$\|b(x,y)\| \leq C(1 + \|x\| + \|y\|).$$

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

$$L(\mu)(\phi)(x) = \langle b(x,\mu), \nabla\phi(x)\rangle + \sigma^2 \Delta\phi(x)\,. \tag{9}$$

Let $(X_t : t \in [0,\infty))$ be the canonical process on $\mathcal{C}$. Denote by $(\mathcal{F}_t^X)_{t\ge 0}$ the natural filtration (*i.e.*, the filtration generated by $\{X_s : 0 \le s \le t\}$).

Definition 1. *Let $p \ge 1$. We say that a measure $\rho \in \mathcal{P}_p(\mathcal{C})$ belongs to the class 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 $(\mathcal{F}_t^X)_{t\ge 0}$-martingale on the probability space $(\mathcal{C}, \mathcal{B}(\mathcal{C}), \rho)$. We denote by V_p the set of such measures.

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

$$G(\rho) := \int \left(\phi(x_t) - \phi(x_s) - \int_s^t L(\rho_u)(\phi)(x_u)\mathrm{d}u\right) \prod_{j=1}^r h_j(x_{v_j}) d\rho(x)\,, \tag{10}$$

where $r \in \mathbb{N}$, $\phi \in C_c^2(\mathbb{R}^d, \mathbb{R})$, $h_1, \dots, h_r \in C_c(\mathbb{R}^d, \mathbb{R})^r$, $0 \le v_1 \le \cdots \le v_r \le s \le t$, are arbitrary. We denote by $\mathcal{G}_p$ the set of such mappings G. Assumption 1 ensures that these mappings are well defined. By Def. 1, every $\rho \in \mathsf{V}_p$ is a root of all $G \in \mathcal{G}_p$. As a matter of fact, a measure $\rho \in \mathcal{P}_p(\mathcal{C})$ belongs to the set V_p, if and only if $G(\rho) = 0$ for every G of the form (10). In other words, Def. 1 is equivalent to the following identity:

$$\mathsf{V}_p = \bigcap_{G \in \mathcal{G}_p} G^{-1}(\{0\})\,. \tag{11}$$

The following lemma is proved in Appendix A.3.

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

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

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

Proof. For all $\rho_n \in \mathsf{V}_p \to \rho_\infty$ in $\mathcal{P}_p(\mathcal{C})$, it holds by Lemma 3 that $G(\rho_\infty) = 0$ for all $G \in \mathcal{G}_p$, which shows that $\rho_\infty \in \mathsf{V}_p$ by (11). Hence, V_p is closed. A closed subset of a Polish space is also Polish. By Prop. 1, V_p is Polish. ◻

2.4 Dynamical systems

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

$$\mathrm{BC}_p := \overline{\{\rho \in \mathsf{V}_p \,:\, \rho \in \omega(\rho)\}}\,.$$

Consider a non-empty set $\Lambda \subset \mathsf{V}_p$.

Definition 2. *Consider the semi-flow Φ. A lower semi-continuous function $F : \mathsf{V}_p \to \mathbb{R}$ is called a Lyapunov function for the set Λ if, for every $\rho \in \mathsf{V}_p$ and every $t > 0$, $F(\Phi_t(\rho)) \le F(\rho)$, 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 Λ, then $\mathrm{BC}_p \subset \overline{\Lambda}$.*

Proof. The limit $\ell := \lim_{t\to\infty} F(\Phi_t(\rho))$ is well-defined because $F(\Phi_t(\rho))$ is non increasing. Consider a recurrent point $\rho \in \mathsf{V}_p$, say $\rho = \lim_n \Phi_{t_n}(\rho)$. Clearly $F(\rho) \ge F(\Phi_{t_n}(\rho)) \ge \ell$. Moreover, by lower semicontinuity of F, $\ell = \lim_n F(\Phi_{t_n}(\rho)) \ge F(\rho)$. Therefore, ℓ 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. □

3 Main results

3.1 Interpolated process and weak⋆ limits

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

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

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

Assumption 2. *The sequence (γ_k) is a non-negative deterministic sequence satisfying*

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

Assumption 3. *The following hold true.*

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

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

$$\mathbb{E}\Big[\xi_{k+1}^{1,n}\Big|\mathcal{F}_k^n\Big] \text{ and } \mathbb{E}\left[\xi_{k+1}^{i,n}\left(\xi_{k+1}^{j,n}\right)^T \mid \mathcal{F}_k^n\right] = \sigma^2 \mathbb{1}_{i=j} I_d\,,$$

for some number $\sigma^2 \ge 0$.

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

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

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

Assumption 4. *We assume either:*

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

or the stronger condition:

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

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

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

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

$$\mathcal{M} := \underset{(t,n)\to(\infty,\infty)}{\mathrm{acc}} \left(\{(\Phi_t(m^n))_\# \mathbb{P}\}\right) \tag{12}$$

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

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

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

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

3.2 Ergodic convergence

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

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

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

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

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

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

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

If Assumption 4–(ii) *additionally holds, the statements remains valid also when $p = 2$.*

3.3 Pointwise convergence to a global attractor

Depending on the vector field b, it is often the case that each measure $\rho \in \mathsf{V}_p$ 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 $\Psi : [0, \infty) \times \mathcal{P}_p(\mathbb{R}^d) \to \mathcal{P}_p(\mathbb{R}^d), (t, \nu) \mapsto \Psi_t(\nu)$, defined on $[0, \infty) \times \mathcal{P}_p(\mathbb{R}^d)$, and such that

$$\rho \in \mathsf{V}_p \Leftrightarrow \forall t \ge 0, \rho_t = \Psi_t(\rho_0). \tag{13}$$

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

The granular media model detailed in Section 4.1 below is a typical example where such a situation occurs. This will also be the case of the random features model detailed in Section 4.2.

In this section, we are interested in the behavior of the measures μ_k^n as $(k, n) \to (\infty, \infty)$, termed the "pointwise" convergence of these measures, when the semi-flow Ψ 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 Ψ 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 Ψ if

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

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

The following result is proven in Section 5.5.

Theorem 2. *Let $p \in [1,2]$, and let Assumptions 1, 2, and 3 hold true. Let Assumption 4–*(i) *or the stronger Assumption 4–*(ii) *hold true according to whether $p < 2$ or $p = 2$ respectively. Assume in addition that the V_p has a semi-flow structure on $\mathcal{P}_p(\mathbb{R}^d)$ as specified in* (13)*, and that this semi-flow Ψ admits a global attractor A_p. Then,*

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

4 Examples

4.1 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

$$b(x,y) := -\nabla V(x) - \nabla U(x-y)\,, \tag{14}$$

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

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

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) $\|\nabla V(x) - \nabla V(y)\| + \|\nabla U(x) - \nabla U(y)\| \leq C(\|x-y\|^\beta \vee \|x-y\|)$, *for every* $(x,y) \in (\mathbb{R}^d)^2$.

Under Assumptions 5, the vector field b satisfies Assumption 1. We will see later, as a byproduct of Th. 3, that the set V_2 of McKean-Vlasov distributions associated to the field b in Eq. (14), is non empty. We say $\mu \ll \mathscr{L}^d$ if $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ admits continuously differentiable density w.r.t. the Lebesgue measure $\mathscr{L}^d$, which we denote by $d\mu/d\mathscr{L}^d$. Define the functional $\mathscr{H} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, \infty]$ as $\mathscr{H}(\mu) = \mathscr{F}(\mu) + \mathscr{V}(\mu) + \mathscr{U}(\mu)$ with

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

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

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

Proposition 6. *Let Assumption 5 hold true, and let b be defined by (14). Assume $\sigma > 0$. Consider $\rho \in \mathsf{V}_2$. Then, for every $t > 0$, ρ_t admits a density $x \mapsto \varrho(t,x)$ in $C^1(\mathbb{R}^d, \mathbb{R})$ w.r.t. the Lebesgue measure. For every $t > 0$, the functional $t \mapsto \mathscr{H}(\rho_t)$ is finite, and satisfies for every $t_2 > t_1 > 0$,*

$$\mathscr{H}(\rho_{t_2}) - \mathscr{H}(\rho_{t_1}) = -\int_{t_1}^{t_2}\int \|v_t(x)\|^2 \varrho(t,x)dxdt\,, \tag{15}$$

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

$$v_t(x) := -\nabla V(x) - \int \nabla U(x-y)d\rho_t(y) - \sigma^2 \nabla \log \varrho(t,x)\,. \tag{16}$$

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

$$\mathcal{S} := \{\mu \in \mathcal{P}_2^r(\mathbb{R}^d) \,:\, \nabla V + \int \nabla U(\,\cdot\, - y)d\mu(y) + \sigma^2 \nabla \log \frac{d\mu}{d\mathscr{L}^d} = 0\, \mu\text{-a.e.}\}\,. \tag{17}$$

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

$$\Lambda_\epsilon := \{\rho \in \mathsf{V}_2 \,:\, \exists \mu \in \mathcal{S}, \forall t \geq \epsilon, \, \rho_t = \mu\}\,. \tag{18}$$

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

$$\mathrm{BC}_2 \subset \overline{\Lambda_0}\,.$$

We also need to consider a setting where V_2 has a semi-flow structure on $\mathcal{P}_2(\mathbb{R}^d)$ as in (13) in order to set the stage for the pointwise convergence of the measures μ_k^n issued from our discrete algorithm. To that end, we shall appeal to the theory of the gradient flows in the space of probability measures as detailed in the treatise [AGS08] of Ambrosio, Gigli and Savaré. The following additional assumption will be needed:

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

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

Proposition 8. *Let Assumption 5 hold true with $\beta = 1$, and let Assumption 6 hold true. Then, for each $\rho \in \mathsf{V}_2$, the curve $t \mapsto \rho_t$ belongs to the set of absolutely continuous functions $\mathrm{AC}^2_{loc}((0,\infty), \mathcal{P}_2(\mathbb{R}^d))$ as defined in [AGS08, Sec. 8.3], and is completely determined by $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ as being the gradient flow of the functional $\mathscr{H}$ in $\mathcal{P}_2(\mathbb{R}^d)$. Thus, V_2 has a semi-flow structure, and we write $\rho_t = \Psi_t(\rho_0)$.*

For completeness, we recall along [AGS08, Chap. 8 and 11] that $t \mapsto \rho_t$ being the solution of the gradient flow of $\mathscr{H}$ in $\mathcal{P}_2(\mathbb{R}^d)$ stands to the existence of a Borel vector field $w_t : \mathbb{R}^d \to \mathbb{R}^d$ such that w_t belongs to the tangent bundle $\mathrm{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d)$ for $\mathscr{L}^1$–almost all $t > 0$, $\|w_t\|_{L^2(\rho_t)} \in L^p_{\mathrm{loc}}(0,\infty)$, the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t w_t) = 0$ holds in general in the sense of distributions, and finally, $w_t \in -\partial \mathscr{H}(\rho_t)$ for $\mathscr{L}^1$–almost each $t > 0$, where $\partial\mathscr{H}$ is the Fréchet sub-differential as defined in [AGS08, Chap. 10], which always exists under our assumptions. Actually, $w_t = v_t$ as given by Equation (16) for almost all t.

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

$$X_{k+1}^{i,n} = X_k^{i,n} - \frac{\gamma_{k+1}}{n} \sum_{j \in [n]} \nabla U(X_k^{i,n} - X_k^{j,n}) - \gamma_{k+1} \nabla V(X_k^{i,n}) + \sqrt{2\gamma_{k+1}} \xi_k^{i,n}\,, \tag{19}$$

for each $i \in [n]$. This is a special case of Eq. (1) with $b(x,y)$ given by Eq. (14) and $\zeta_k^{i,n} = 0$ for all k. For simplicity, Assumption 3 will be replaced by the following stronger assumption:

Assumption 7. *The n-tuple $(X_0^{1,n}, \ldots, X_0^{n,n})$ is exchangeable and satisfies $\sup_n \mathbb{E}(\|X_0^{1,n}\|^4) < \infty$. 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 the condition ii) in Assumption 4 holds.

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

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

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

Theorem 3. *Let Assumptions 2, 5 and 7 be satisfied. Assume $\sigma > 0$. Then, the set $\mathcal{S}$ given by Eq. (17) is non empty, and furthermore,*

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

Proof. Use Cor. 1 with $p = 2$, together with Prop. 7. □

We now turn to the pointwise convergence of the measures μ_k^n.

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

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

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

4.2 Random features

Consider two integers $q \geq 1$, $d \geq 2$. For any fixed n-uple $\boldsymbol{w} = (w^1, \cdots, w^n)$ of $\mathbb{R}^{d-1}$-valued vectors, we consider the following regularized risk minimization problem:

$$\min_{\boldsymbol{a}\in\mathbb{R}^n} R(\boldsymbol{a}, \boldsymbol{w}) := \frac{1}{2}\int (h(x; \boldsymbol{a}, \boldsymbol{w}) - y)^2 d\nu(x, y) + \frac{\lambda}{2n}\|\boldsymbol{a}\|^2 , \tag{20}$$

where $\nu \in \mathcal{P}_2(\mathbb{R}^q \times \mathbb{R})$, $\lambda > 0$ is a regularization parameter, and where for every $x \in \mathbb{R}^q$, every $\boldsymbol{a} = (a^1, \ldots, a^n)$ in $\mathbb{R}^n$,

$$h(x; \boldsymbol{a}, \boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^n a^i \varphi(x, w^i) ,$$

where $\varphi : \mathbb{R}^q \times \mathbb{R}^{d-1} \to \mathbb{R}$ is a function, refered to as the feature map. In the sequel, we consider the process of searching for a minimizer of (20), when the coefficients $\boldsymbol{w}$ are set to an iid sample $\boldsymbol{w}_0^n := (w_0^{1,n}, \ldots, w_0^{n,n})$ of n random variables on $\mathbb{R}^{d-1}$, following a prescribed distribution. We consider the stochastic gradient descent (SGD) on the $\boldsymbol{a}$-parameter, obtained by randomly selecting, at time k, a sample (x_{k+1}^n, y_{k+1}^n) according to the distribution ν. Denoting by $\boldsymbol{a}_k^n = (a_k^{1,n}, \cdots, a_k^{n,n})$ the updated parameters, we consider the iterations:

$$\boldsymbol{a}_{k+1}^n = \boldsymbol{a}_k^n - \gamma_{k+1} n \nabla_{\boldsymbol{a}} \left[\frac{1}{2}(h(x_{k+1}^n; \boldsymbol{a}_k^n, \boldsymbol{w}_0^n) - y_{k+1}^n)^2 + \frac{\lambda\|\boldsymbol{a}_k^n\|^2}{2n}\right] + \sqrt{2\gamma_{k+1}}\tilde{\boldsymbol{\xi}}_{k+1}^n , \tag{21}$$

where $\tilde{\boldsymbol{\xi}}_{k+1}^n = (\tilde{\xi}_{k+1}^{1,n}, \ldots, \tilde{\xi}_{k+1}^{n,n})$ are centered iid random variables with variance $\sigma^2 \geq 0$, where $\gamma_{k+1} n$ is the learning rate, and $\nabla_{\boldsymbol{a}}$ stands for the gradient w.r.t. variable $\boldsymbol{a}$. In the algorithm given by Eq. (21), only the parameter $\boldsymbol{a}$ is updated, while the parameter $\boldsymbol{w}$ is set once for all, at the initialization step. In this setting, the values $(\varphi(x, w_0^{i,n}))_{i\in[n]}$ are refered to as the random features associated with an input x. Note that the learning rate $\gamma_{k+1} n$ is chosen in order to vanish with k, but also to scale with parameter n. Due to the presence of the term $\sqrt{2\gamma_{k+1}}\tilde{\boldsymbol{\xi}}_{k+1}^n$, the iterations (21) should be considered as a noisy version of the classical SGD, in the flavor of a Langevin algorithm. The standard SGD case is obtained by setting $\sigma = 0$. The following assumption summarizes the stated conditions on the above random variables.

Assumption 8. *The following holds.*

i) The r.v. $(\tilde{\xi}_k^{i,n} : k \in \mathbb{N}^*, i \in [n], n \in \mathbb{N}^*)$ are real iid, centered, random variables with variance $\sigma^2 \geq 0$, and satisfy $\mathbb{E}((\tilde{\xi}_1^{1,1})^4) < \infty$.

ii) The r.v. $((x_k^{i,n}, y_k^{i,n}) : k \in \mathbb{N}^*, i \in [n], n \in \mathbb{N}^*)$ are iid, with distribution $\nu \in \mathcal{P}(\mathbb{R}^q \times \mathbb{R})$.

iii) The r.v. $(a_0^{i,n} : i \in [n], n \in \mathbb{N}^*)$ are iid real r.v., and satisfy $\mathbb{E}((a_0^{1,1})^4) < \infty$.

iv) The r.v. $(w_0^{i,n} : i \in [n], n \in \mathbb{N}^*)$ are iid, and satisfy $\mathbb{E}((w_0^{1,1})^2) < \infty$.

v) The families of r.v. respectively mentioned in the four above points i–iv are independent.

Eq. (21) can be expanded, for every $i \in [n]$, as:

$$a_{k+1}^{i,n} = (1 - \frac{\lambda\gamma_{k+1}}{n})a_k^{i,n} + \frac{\gamma_{k+1}}{n}\sum_{j=1}^{n}(y_{k+1}^n - a_k^{j,n}\varphi(x_{k+1}^n, w_0^{j,n}))\varphi(x_{k+1}^n, w_0^{i,n}) + \sqrt{2\gamma_{k+1}}\tilde{\xi}_{k+1}^{i,n}\,.$$

We introduce the following vector field for every $(a, w), (a', w') \in \mathbb{R} \times \mathbb{R}^{d-1}$:

$$\begin{aligned}\tilde{b}((a, w), (a', w')) &:= \int (y - a'\varphi(x, w'))\varphi(x, w)d\nu(x, y) - \lambda a \\ &= Q(w) - a'K(w, w') - \lambda a\,,\end{aligned}$$

where we set:

$$\begin{aligned}K(w, w') &:= \int \varphi(x, w)\varphi(x, w')d\nu(x, y) \\ Q(w) &:= \int y\varphi(x, w)d\nu(x, y)\,.\end{aligned}$$

Then, the SGD iterations can be written as:

$$a_{k+1}^{i,n} = a_k^{i,n} + \gamma_{k+1}\frac{1}{n}\sum_{j=1}^{n}\tilde{b}((a_k^{i,n}, w_0^{i,n}), (a_k^{j,n}, w_0^{j,n})) + \gamma_{k+1}\tilde{\zeta}_{k+1}^{i,n} + \sqrt{2\gamma_{k+1}}\tilde{\xi}_{k+1}^{i,n}\,,$$

where $\tilde{\zeta}_{k+1}^{i,n}$ is the random perturbation given by:

$$\tilde{\zeta}_{k+1}^{i,n} := \frac{1}{n}\sum_{j=1}^{n}\left[(y_{k+1}^n - a_k^{j,n}\varphi(x_{k+1}^n, w_0^{j,n}))\varphi(x_{k+1}^n, w_0^{i,n}) - \tilde{b}((a_k^{i,n}, w_0^{i,n}), (a_k^{j,n}, w_0^{j,n}))\right]\,.$$

The above iterations can be casted into the general form (1), by setting $X_k^{i,n} := (a_k^{i,n}, w_0^{i,n})$, along with $b((a, w), (a', w')) := (\tilde{b}((a, w), (a', w')), 0)$, $\zeta_{k+1}^{i,n} := (\tilde{\zeta}_{k+1}^{i,n}, 0)$ and $\xi_{k+1}^{i,n} := (\tilde{\xi}_{k+1}^{i,n}, 0)$. Consequently, Th. 1 can be used in order to characterize the long run convergence of the occupation measure of the updated parameters. We define $\mathcal{F}_k^n$ as the σ-field generated by the r.v. $a_0^{i,n}, w_0^{i,n}$ and $(x_l^n, y_l^n, \xi_l^{1,n}, \dots, \xi_l^{n,n})$ for $l \in [k]$. Note that $(\zeta_k^{1,n}, \dots, \zeta_k^{n,n})$ is $\mathcal{F}_k^n$-measurable.

Remark 2. *Although the particles $X_k^{i,n} = (a_k^{i,n}, w_0^{i,n})$ satisfy Eq. (1), the notable difference with the model considered in Section 3 lies in the fact that the variables $\xi_{k+1}^{1,n}$ here satisfy:*

$$\mathbb{E}(\xi_{k+1}^{1,n}(\xi_{k+1}^{1,n})^T|\mathcal{F}_k^n) = \sigma^2\text{diag}(1, 0, \dots, 0)\,,$$

whereas our results have been proven under the assumption that $\mathbb{E}(\xi_{k+1}^{1,n}(\xi_{k+1}^{1,n})^T|\mathcal{F}_k^n) = \sigma^2 I_d$. This difference is minor, and our results can be extended without any difficulty to the former case. The most important modification lies in the definition of the McKean-Vlasov distribution V_p in Def. (1), where one should replace the Laplacian term $\Delta\phi(x)$ in the definition (9) of L by the second order partial derivative w.r.t. the first component. This gives rise to the definition:

$$L^{(\sigma^2,\mathbf{0})}(\mu)(\phi)(a, w) = \langle b(a, w), \mu), \nabla\phi(x)\rangle + \sigma^2\partial_a^2\phi(a, w)\,. \tag{22}$$

To avoid any confusion, we now denote by $\mathsf{V}_p^{(\sigma^2,\mathbf{0})}$ the set of McKean-Vlasov distributions defined as in Def. (1), replacing L by $L^{(\sigma^2,\mathbf{0})}$, and we denote by $\text{BC}_p^{(\sigma^2,\mathbf{0})}$ the corresponding Birkhoff center.

Assumption 9. *The following holds.*

i) The function φ is bounded and continuous.

ii) $\int y^4 d\nu(x, y) < \infty$.

Proposition 10. *Let Assumptions 2, 8 and 9 hold true. Then,* $\sup_{n,k}\mathbb{E}((a_k^{1,n})^4 + (\tilde{\zeta}_k^{i,n})^4) < \infty$.

Proof. The proof is provided in Section 7.1. □

We denote by $\tilde{\pi} : (a, w) \mapsto w$ the projection $\mathbb{R} \times \mathbb{R}^{d-1} \to \mathbb{R}^{d-1}$ on the last $d-1$ components. Due to Prop. 10, the conditions of application of Th. 1 are satisfied. As a consequence of Th. 1, the set $\mathrm{BC}_2^{(\sigma^2, \mathbf{0})}$ is non empty. Moreover, as the marginal distribution of the particles w.r.t. the w-variable is a constant, fixed once for all to the distribution of $w_0^{1,1}$, we obtain the following result. As long as Assumptions 2 and 9 hold true, for every $\varpi \in \mathcal{P}_2(\mathbb{R}^{d-1})$, there exists $\rho \in \mathrm{BC}_2^{(\sigma^2, \mathbf{0})}$ such that $\tilde{\pi}_\# \rho_t = \varpi$ for all $t \geq 0$.

For every $\mu \in \mathcal{P}_2(\mathbb{R} \times \mathbb{R}^{d-1})$, define:

$$\mathcal{R}_0(\mu) := \frac{1}{2} \int \left(\int a\varphi(x, w) d\mu(a, w) - y \right)^2 d\nu(x, y) + \frac{\lambda}{2} \int a^2 d\mu(a, w) .$$

The functional $\mathcal{R}_0$ is related to the initial minimization problem through the identity:

$$R(\boldsymbol{a}, \boldsymbol{w}) = \mathcal{R}_0 \left(\frac{1}{n} \sum_{i=1}^{n} \delta_{(a^i, w^i)} \right) ,$$

for every $\boldsymbol{a} = (a^1, \cdots, a^n)$ and $\boldsymbol{w} = (w^1, \cdots, w^n)$. For any $\mu \in \mathcal{P}(\mathbb{R} \times \mathbb{R}^{d-1})$, we write the disintegration of the measure μ as:

$$\mu(da, dw) = \mu(da|w) \tilde{\pi}_\# \mu(dw) .$$

For every $\sigma > 0$, we define the functional $\mathcal{R}_\sigma$ as follows, for every $\mu \in \mathcal{P}_2(\mathbb{R} \times \mathbb{R}^{d-1})$:

$$\mathcal{R}_\sigma(\mu) := \mathcal{R}_0(\mu) + \sigma^2 \int \log \left(\frac{d\mu(\,\cdot\,|w)}{d\mathscr{L}^1}(a) \right) d\mu(a, w) , \tag{23}$$

whenever $\mu(\,\cdot\,|w)$ is absolutely continuous w.r.t. the Lebesgue measure $\mathscr{L}^1$, and $\mathcal{R}_\sigma(\mu) = +\infty$ otherwise. For every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we define:

$$\tilde{v}_\mu(a, w) := \tilde{b}((a, w), \mu) + \sigma^2 \partial_a \log \left(\frac{d\mu(\,\cdot\,|w)}{d\mathscr{L}^1}(a) \right) ,$$

whenever $\mu(\,\cdot\,|w) \ll \mathscr{L}^1$, or $\sigma = 0$.

If $\mu, \mu^* \in \mathcal{P}_2(\mathbb{R}^d)$, by [AGS08, Lem. 12.4.7], there exists a Borel map on $\mathbb{R}^{d-1} \to \mathcal{P}_2(\mathbb{R} \times \mathbb{R})$ which, to every $w \in \mathbb{R}^{d-1}$, associated a probability measure $\boldsymbol{\gamma}(\cdot|w) \in \Pi_2^0(\mu_*(\cdot|w), \mu(\cdot|w))$, where we recall that $\Pi_2^0(\mu^*(\cdot|w), \mu(\cdot|w))$ is the set of 2-Wasserstein optimal transport plans between $\mu^*(\cdot|w)$ and $\mu(\cdot|w)$, as introduced after Eq. (8).

Lemma 4. *Let Assumptions 2 and 9 hold true. Consider $\varpi \in \mathcal{P}_2(\mathbb{R}^{d-1})$ and $\mu, \mu^* \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\tilde{\pi}_\# \mu = \tilde{\pi}_\# \mu^* = \varpi$. Assume that either $\sigma = 0$, or that the following holds for ϖ-almost all w: $\mu^*(\,\cdot\,|w) \ll \mathscr{L}^1$ and*

$$\int \left(\partial_a \log \frac{d\mu(\,\cdot\,|w)}{d\mathscr{L}^1}(a) \right)^2 d\mu(a|w) < \infty .$$

Let $w \mapsto \boldsymbol{\gamma}(\cdot|w)$ be a measurable selection of the correspondence $w \mapsto \Pi_2^0(\mu_*(\cdot|w), \mu(\cdot|w))$. Then,

$$\mathcal{R}_\sigma(\mu) - \mathcal{R}_\sigma(\mu^*) \geq -\int \int \tilde{v}_{\mu^*}(a_*, w)(a - a_*) d\boldsymbol{\gamma}(a_*, a|w) d\varpi(w) + \frac{\lambda}{2} \int \mathsf{W}_2^2(\mu(\,\cdot\,|w), \mu^*(\,\cdot\,|w)) d\varpi(w) .$$

Proof. The proof is provided in Section 7.2. □

Lemma 5. *Let Assumptions 2 and 9 hold true. Consider $\varpi \in \mathcal{P}_2(\mathbb{R}^d)$, and $\rho \in \mathsf{V}_2^{(\sigma^2, \mathbf{0})}$ such that $\tilde{\pi}_\# \rho_0 = \varpi$. In that case, $\tilde{\pi}_\# \rho_t = \varpi$ for all t. If $\sigma > 0$, then, for every $t > 0$, $\rho_t(\,\cdot\,|w) \ll \mathscr{L}^1$ for ϖ-almost every w. Morever, for every $t > 0$ and ϖ-almost every w,*

$$\int \left(\partial_a \log \frac{d\rho_t(\,\cdot\,|w)}{d\mathscr{L}^1}(a) \right)^2 d\rho_t(a|w) < \infty . \tag{24}$$

Finally, $\mathcal{R}_\sigma(\rho_t)$ is finite for all $t > 0$, and for all $t_2 > t_1 > 0$,

$$\mathcal{R}_\sigma(\rho_{t_2}) - \mathcal{R}_\sigma(\rho_{t_1}) \leq -\int_{t_1}^{t_2} \int \tilde{v}_{\rho_t}(a, w)^2 d\rho_t(a, w) dt . \tag{25}$$

Proof. The proof is provided in Section 7.3. □

For every $\varpi \in \mathcal{P}_2(\mathbb{R}^{d-1})$, define the set

$$S(\varpi) := \{\mu \in \mathcal{P}_2(\mathbb{R}^d) : \tilde{\pi}_{\#}\mu = \varpi \text{ and } \tilde{v}_\mu = 0,\ \mu\text{-a.e.}\}\,.$$

Corollary 2. *Let Assumptions 2 and 9 hold true. For every $\epsilon > 0$, the map $\rho \mapsto \mathcal{R}_\sigma(\rho_\epsilon)$ is well defined on $\mathsf{V}_2^{(\sigma^2,\mathbf{0})} \to \mathbb{R}$, lower semicontinuous, and is a Lyapunov function for the set:*

$$\Lambda_\epsilon := \{\rho \in \mathsf{V}_2^{(\sigma^2,\mathbf{0})} : \exists \varpi \in \mathcal{P}_2(\mathbb{R}^{d-1}), \exists \mu \in S(\varpi), \forall t \geq \epsilon, \rho_t = \mu\}\,.$$

Moreover, $\mathrm{BC}_2^{(\sigma^2,\mathbf{0})} \subset \overline{\Lambda_0}$.

Proof. It is an immediate consequence of Lem. 5. □

Proposition 11. *Let Assumption 9 hold true. Then, for every $\varpi \in \mathcal{P}_2(\mathbb{R}^{d-1})$, there exists a unique minimizer, denoted by $\mu^*(\varpi)$, of $\mathcal{R}_\sigma$ among all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\tilde{\pi}_{\#}\mu = \varpi$. Moreover, $\mu^*(\varpi)$ is the unique measure μ satisfying $v_\mu = 0$ μ-a.e., and $\tilde{\pi}_{\#}\mu = \varpi$.*

Proof. As discussed after the statement of Prop. 10, there exists a recurrent point ρ such that $\tilde{\pi}_{\#}\rho_0 = \varpi$. By Cor. 2, $\rho \in \Lambda_0$, which implies that $\rho_0 \in S(\varpi)$. This shows that, for every $\varpi \in \mathcal{P}_2(\mathbb{R}^{d-1})$, $S(\varpi)$ is non empty.

Consider an arbitrary $\mu^* \in S(\varpi)$. By Lem. 4, μ^* is a minimizer of $\mathcal{R}_\sigma$, among the set of measures with marginal ϖ. This shows existence. Let μ be another such minimizer. By Lem. 4, $W_2(\mu^*(\cdot|w), \mu(\cdot|w)) = 0$ for ϖ-almost every w. Thus, $\mu = \mu^*$. □

We are now able to state the main result of this paragraph. Define:

$$\mu_k^n := \frac{1}{n}\sum_{i=1}^n \delta_{(a_k^{i,n}, w_0^{i,n})}\,.$$

Theorem 5. *Let Assumptions 2, 8 and 9 hold true. Assume that $w_0^{1,1}$ has the distribution $\varpi \in \mathcal{P}_2(\mathbb{R}^{d-1})$. Then,*

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

Proof. Put together Prop. 10, Cor.1, Cor. 2 and Prop. 11. □

Remark 3. *Let $\varpi \in \mathcal{P}_2(\mathbb{R}^{d-1})$. As in the proof of Lemma 5, one is able to apply [AGS08, Th. 11.2.1] and obtain the contraction $W_2(\rho_t, \mu^*(\varpi)) \leq \mathrm{e}^{-\lambda t} W_2(\rho_0, \mu^*(\varpi))$ for every $\rho \in \mathsf{V}_2^{(\sigma^2,\mathbf{0})}$ such that $\tilde{\pi}_{\#}\rho_0 = \varpi$. With this in hand, in the same spirit as Theorem 2, one can establish the pointwise convergence*

$$W_2(\mu_k^n, \mu^*(\varpi)) \xrightarrow[(k,n)\to(\infty,\infty)]{\mathbb{P}} 0\,,$$

under the assumptions of Theorem 5.

5 Proofs of Section 3

5.1 Proof of Proposition 4

In this paragraph, consider $1 \leq p \leq 2$. Note that $(\Phi_t(m^n))$ belongs to $\mathcal{P}_p(\mathcal{C})$. In the light of Lemma 2 and Prop 1, we should establish two points: first, the weak$\star$-relatively compactness of the family of intensities $\{\mathbb{I}(\Phi_t(m^n))\}_{t,n}$; second, a uniform integrability condition of the pth order moments of the measures $\mathbb{I}(\Phi_t(m^n)(x))$. These results are respectively stated in Lemmas 6 and 7 below.

Lemma 6. *We posit the assumptions of Prop. 4. The family of intensities $\{\mathbb{I}(\Phi_t(m^n))\}_{t,n}$ is weak$\star$-relatively compact in $\mathcal{P}(\mathcal{C})$.*

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

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

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

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

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

$$\lim_{\delta\to 0}\limsup_{t,n} \mathbb{P}\left(w^T_{\bar{X}^{1,n}_{t+\cdot}}(\delta) \geq \varepsilon\right) = 0 \tag{26}$$

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

$$\begin{aligned}
&\bar{X}^{1,n}_t - \bar{X}^{1,n}_s = P^n_{s,t} + N^n_{s,t} + U^n_{s,t}, \qquad (27)\\
&P^n_{s,t} := \frac{1}{n}\sum_{j=1}^n \left(\sum_{k=k_s}^{k_t-2} \gamma_{k+1} b(X^{1,n}_k, X^{j,n}_k) + (\tau_{k_s} - s)\, b(X^{1,n}_{k_s-1}, X^{j,n}_{k_s-1}) + (\tau_{k_t} - t)\, b(X^{1,n}_{k_t-1}, X^{j,n}_{k_t-1})\right)\\
&N^n_{s,t} := \sum_{k=k_s}^{k_t-2} \sqrt{\gamma_{k+1}}\xi^{1,n}_{k+1} + \frac{\tau_{k_s} - s}{\gamma_{k_s}}\sqrt{\gamma_{k_s}}\xi^{1,n}_{k_s} + \frac{\tau_{k_t} - t}{\gamma_{k_t}}\sqrt{\gamma_{k_t}}\xi^{1,n}_{k_t}\\
&U^n_{s,t} := \frac{1}{n}\sum_{i=1}^n \left(\sum_{k=k_s}^{k_t-2} \gamma_{k+1}\zeta^{i,n}_{k+1} + (\tau_{k_s} - s)\,\zeta^{i,n}_{k_s} + (\tau_{k_t} - t)\,\zeta^{i,n}_{k_t}\right).
\end{aligned}$$

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

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

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 $N^n_{s,t}$ is expressed as a sum of martingale increments, with respect to the filtration $\mathcal{F}^n_k$. Let $\|\cdot\|_\alpha$ denote the α-norm in $\mathbb{R}^d$. We apply Burkholder's inequality stated in [BDG72, Th. 1.1] to the components of the vector $N^n_{s,t}$ in $\mathbb{R}^d$. As Eq. (28) and (29) hold:

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

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

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

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

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

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

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

Also, by Jensen's inequality, we have

$$\left\|U^n_{s,t}\right\|^2 \le \frac{(t-s)}{n} \sum_{i\in[n]} \sum_{k=k_s-1}^{k_t-1} \tilde\gamma_{k+1} \left\|\zeta^{i,n}_{k+1}\right\|^2 \,.$$

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

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

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

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\left\|P^n_{s,t}\right\|^2 + \left\|N^n_{s,t}\right\|^4 + \left\|U^n_{s,t}\right\|^2\right] \le C(t-s)^2 \,, \tag{33}$$

where $0 \le s < t < \infty$, and C is a positive constant, independent of s, t, n. Using [Leo23, Th. 2.8] and Markov's inequality, Eq. (26) hold. □

Lemma 7. *We posit the assumptions of Prop. 4. For every $T > 0$,*

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

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

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

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

$$\left\|\bar X^{1,n}_{t+u}\right\|^p \le 4^{p-1}\left(\left\|\bar X^{1,n}_{t}\right\|^p + \left\|N^n_{t,t+u}\right\|^p + \left\|P^n_{t,t+u}\right\|^p + \left\|U^n_{t,t+u}\right\|^p\right) .$$

Hence,

$$\begin{aligned}\left\|\bar X^{1,n}_{t+u}\right\|^p \mathbb{1}_{\sup\limits_{u\in[0,T]}\|\bar X^{1,n}_{t+u}\|>a} \le 4^p \Bigg(&\left\|\bar X^{1,n}_{t}\right\|^p \mathbb{1}_{\|\bar X^{1,n}_t\|>a/4} + \left\|N^n_{t,t+u}\right\|^p \mathbb{1}_{\sup\limits_{u\in[0,T]}\|N^n_{t,t+u}\|>a/4} \\ &+ \left\|P^n_{t,t+u}\right\|^p \mathbb{1}_{\sup\limits_{u\in[0,T]}\|P^n_{t,t+u}\|>a/4} + \left\|U^n_{t,t+u}\right\|^p \mathbb{1}_{\sup\limits_{u\in[0,T]}\|U^n_{t,t+u}\|>a/4}\Bigg) .\end{aligned}$$

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

$(\|\bar X^{1,n}_t\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*)$ is uniformly integrable by Assumption 4–(i) when $p < 2$, and by Assumption 4–(ii) when $p = 2$. As obtained in Eq. (30), Burkholder inequality stated in [BDG72, Th 1.1] yields:

$$\mathbb{E}\left[\sup_{u\in[0,T]} \left\|N^n_{t,t+u}\right\|^4\right] \le CT^2 \,,$$

where C is a constant independent of t, n, and T. Hence, since $p < 4$, we obtain the uniform integrability of $\{\sup_{u\in[0,T]} \|N^n_{t,t+u}\|^p : t \in \mathbb{R}_+, n \in \mathbb{N}^*\}$. As obtained in Eq. (31) and Eq. (32), we derive:

$$\sup_{u\in[0,T]} \left\|P^n_{t,t+u}\right\|^p \le \frac{CT^{p-1}}{n} \sum_{j\in[n]} \sum_{k=k_t-1}^{k_{t+T}-1} \tilde\gamma_{k+1} \left(1 + \left\|X^{j,n}_k\right\|^p + \left\|X^{1,n}_k\right\|^p\right) ,$$

and

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

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

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

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

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

Consequently,

$$\sup_{t\in\mathbb{R}_+,n\in\mathbb{N}^*} \mathbb{E}\left[F\left(\sup_{u\in[0,T]} \|P^n_{t,t+u}\|^p\right)\right] \le \sup_{k\in\mathbb{N},n\in\mathbb{N}^*} \mathbb{E}\left[F\left(CT^p\left(1+\left\|X^{1,n}_k\right\|\right)^p\right)\right] < \infty\,.$$

Therefore, de la Vallée Poussin theorem yields the uniform integrability of the collection

$$(\sup_{u\in[0,T]} \|P^n_{t,t+u}\|^p \,:\, t\in\mathbb{R}_+, n\in\mathbb{N}^*)\,.$$

The uniform integrability of the collection $(\sup_{u\in[0,T]} \|U^n_{t,t+u}\|^p \,:\, t\in\mathbb{R}_+, n\in\mathbb{N}^*)$ is obtained, by the same arguments. This completes the proof. □

To conclude the proof of Prop. 4, it is sufficient to remark that the tightness conditions provided in Lemma 2 are satisfied, thanks to Lemmas 6 and 7, with Prop. 1.

5.2 Proof of Proposition 5

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

Proposition 12. *Let Assumptions 1, 2, 3 and 4–*(i) *hold,*

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

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

Proof. We need to show that for each $\mathbb{R}_+\times\mathbb{N}$–valued sequence $(t_n,\varphi_n)\to(\infty,\infty)$ as $n\to\infty$, the convergence $\mathbb{E}\,|G(\Phi_{t_n}(m^{\varphi_n}))| \to 0$ holds true, where $G = G_{r,\phi,h_1,\ldots,h_r,t,s,v_1,\ldots,v_r}$ has the form of Eq. (10), with $0\le v_1\le\cdots\le v_r\le s\le t$. We take $\varphi_n = n$ for notational simplicity, and we write $\mathsf{m}_n := \Phi_{t_n}(m^n) \in \mathcal{P}_p(\mathcal{C})$. We have

$$G(\mathsf{m}_n) = \frac{1}{n}\sum_{i\in[n]} \left(\phi(\bar{X}^{i,n}_{t_n+t}) - \phi(\bar{X}^{i,n}_{t_n+s}) - \int_{t_n+s}^{t_n+t} \frac{1}{n}\sum_{j\in[n]} \psi(\bar{X}^{i,n}_u, \bar{X}^{j,n}_u)du\right) Q^{i,n}\,, \tag{34}$$

where se set $\psi(x,y) := \langle\nabla\phi(x), b(x,y)\rangle + \sigma^2\Delta\phi(x)$, and

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

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

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 $\phi(\bar{X}^i_{t_n+t}) - \phi(\bar{X}^i_{t_n+s})$ as a telescoping sum in the discrete random variables X^i_k:

$$\phi(\bar{X}^i_{t_n+t}) - \phi(\bar{X}^i_{t_n+s}) = \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \left(\phi(X^i_{k+1}) - \phi(X^i_k)\right) \\ + \phi(\bar{X}^i_{t_n+t}) - \phi(X^i_{k_{t_n+t}-1}) + \phi(X^i_{k_{t_n+s}}) - \phi(\bar{X}^i_{t_n+s}).$$

The summands at the right hand side of this expression can be decomposed as follows. Remember the form (1) of our algorithm. Denoting as $H_\phi(x)$ the Hessian matrix of ϕ at x, we know by the Taylor-Lagrange formula that there exists $\theta_{k+1} \in [\tau_k, \tau_{k+1}]$ such that

$$\begin{aligned}
\phi(X^i_{k+1}) - \phi(X^i_k) &= \langle \nabla\phi(X^i_k), X^i_{k+1} - X^i_k \rangle + \frac{1}{2}\left(X^i_{k+1} - X^i_k\right)^T H_\phi(\bar{X}^i_{\theta_{k+1}}) \left(X^i_{k+1} - X^i_k\right) \\
&= \gamma_{k+1} \frac{1}{n} \sum_{j\in[n]} \langle \nabla\phi(X^{i,n}_k), b(X^i_k, X^j_k)\rangle + \gamma_{k+1}\sigma^2 \Delta\phi(X^i_k) \\
&\quad + \sqrt{2\gamma_{k+1}} \langle \nabla\phi(X^i_k), \xi^i_{k+1}\rangle + \gamma_{k+1}\langle \nabla\phi(X^i_k), \zeta^i_{k+1}\rangle \\
&\quad + \frac{1}{2}\left(X^i_{k+1} - X^i_k\right)^T H_\phi(\bar{X}^i_{\theta_{k+1}}) \left(X^i_{k+1} - X^i_k\right) - \gamma_{k+1}\sigma^2\Delta\phi(X^i_k) \\
&= \frac{1}{n}\sum_{j\in[n]} \psi(X^i_k, X^j_k) + \sqrt{2\gamma_{k+1}}\langle \nabla\phi(X^i_k), \xi^i_{k+1}\rangle + \gamma_{k+1}\langle\nabla\phi(X^i_k), \zeta^i_{k+1}\rangle \\
&\quad + \frac{1}{2}\left(X^i_{k+1} - X^i_k\right)^T H_\phi(\bar{X}^i_{\theta_{k+1}}) \left(X^i_{k+1} - X^i_k\right) - \gamma_{k+1}\sigma^2\Delta\phi(X^i_k) \\
&= \frac{1}{n}\sum_{j\in[n]} \psi(X^i_k, X^j_k) + \gamma_{k+1}\langle\nabla\phi(X^i_k), \zeta^i_{k+1}\rangle \\
&\quad + \frac{1}{2}\left(X^i_{k+1} - X^i_k\right)^T H_\phi(\bar{X}^i_{\theta_{k+1}}) \left(X^i_{k+1} - X^i_k\right) - \gamma_{k+1}(\xi^i_{k+1})^T H_\phi(X^i_k)\xi^i_{k+1} \\
&\quad + \sqrt{2\gamma_{k+1}}\langle\nabla\phi(X^i_k), \xi^i_{k+1}\rangle + \gamma_{k+1}(\xi^i_{k+1})^T H_\phi(X^i_k)\xi^i_{k+1} - \gamma_{k+1}\sigma^2\Delta\phi(X^i_k).
\end{aligned}$$

In this last expression, the terms $n^{-1}\sum_{j\in[n]}\psi(X^i_k, X^j_k)$ 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. Notice that since $\mathrm{tr}(H_\phi(\bar{X}^i_k)) = \Delta\phi(X^i_k)$, the term

$$\eta^i_{k+1} := \sqrt{2\gamma_{k+1}}\langle\nabla\phi(X^i_k), \xi^i_{k+1}\rangle + \gamma_{k+1}(\xi^i_{k+1})^T H_\phi(X^i_k)\xi^i_{k+1} - \gamma_{k+1}\sigma^2\Delta\phi(X^i_k)$$

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

To proceed, considering the integral at the right hand side of (34), we can write

$$\int_{t_n+s}^{t_n+t} \psi(\bar{X}^i_u, \bar{X}^j_u)du = \\ \int_{\tau_{k_{t_n+s}}}^{\tau_{k_{t_n+t}-1}} \psi(\bar{X}^i_u, \bar{X}^j_u)du + \int_{t_n+s}^{\tau_{k_{t_n+s}}} \psi(\bar{X}^i_u, \bar{X}^j_u)du + \int_{\tau_{k_{t_n+t}-1}}^{t_n+t} \psi(\bar{X}^i_u, \bar{X}^j_u)du,$$

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

$$\chi_1^n := \frac{1}{n}\sum_{i\in[n]}\Big\{\sum_{k=k_{t_n+s}}^{k_{t_n+t}-2}\frac{1}{n}\sum_{j\in[n]}\gamma_{k+1}\psi(X_k^i,X_k^j) - \int_{\tau_{k_{t_n+s}}}^{\tau_{k_{t_n+t}-1}}\frac{1}{n}\sum_{j\in[n]}\psi(\bar X_u^i,\bar X_u^j)du\Big\}Q^i,$$

$$\chi_2^n := \frac{1}{n}\sum_{i\in[n]}\Big\{\phi(\bar X_{t_n+t}^i) - \phi(X_{k_{t_n+t}-1}^i) + \phi(X_{k_{t_n+s}}^i) - \phi(\bar X_{t_n+s}^i)\Big\}Q^i,$$

$$\chi_3^n := -\frac{1}{n}\sum_{i\in[n]}\Big\{\int_{t_n+s}^{\tau_{k_{t_n+s}}}\frac{1}{n}\sum_{j\in[n]}\psi(\bar X_u^i,\bar X_u^j)du + \int_{\tau_{k_{t_n+t}-1}}^{t_n+t}\frac{1}{n}\sum_{j\in[n]}\psi(\bar X_u^i,\bar X_u^j)du\Big\}Q^i,$$

$$\chi_4^n := \frac{1}{n}\sum_{i\in[n]}Q^i\sum_{k=k_{t_n+s}}^{k_{t_n+t}-2}\gamma_{k+1}\langle\nabla\phi\left(X_k^i\right),\zeta_{k+1}^i\rangle,$$

$$\chi_5^n := \frac{1}{n}\sum_{i\in[n]}\sum_{k=k_{t_n+s}}^{k_{t_n+t}-2}\gamma_{k+1}\left(\xi_{k+1}^i\right)^T\Big(H_\phi(\bar X_{\theta_{k+1}}^i) - H_\phi(X_k^i)\Big)\left(\xi_{k+1}^i\right)Q^i,$$

$$\chi_6^n := \frac{1}{n^2}\sum_{i,j\in[n]}\sum_{k=k_{t_n+s}}^{k_{t_n+t}-2}\Big(\sqrt{2}\gamma_{k+1}^{3/2}b(X_k^i,X_k^j)^T H_\phi(\bar X_{\theta_{k+1}}^i)\xi_{k+1}^i\Big)Q^i$$
$$+\frac{1}{n^3}\sum_{i,p,q\in[n]}\sum_{k=k_{t_n+s}}^{k_{t_n+t}-2}\left(\frac{1}{2}\gamma_{k+1}^2 b\left(X_k^i,X_k^p\right)^T H_\phi(\bar X_{\theta_{k+1}}^i)b\left(X_k^i,X_k^q\right)\right)Q^i,$$

$$\chi_7^n := \frac{1}{n^2}\sum_{i,j\in[n]}\sum_{k=k_{t_n+s}}^{k_{t_n+t}-2}\gamma_{k+1}^{3/2}\left(\left(\sqrt{\gamma_{k+1}}\left(b(X_k^i,X_k^j) + \frac{\zeta_{k+1}^i}{2}\right) + \sqrt{2}\xi_{k+1}^i\right)^T H_\phi(\bar X_{\theta_{k+1}}^i)\zeta_{k+1}^i\right)Q^i, \quad \text{and}$$

$$\chi_8^n := \frac{1}{n}\sum_{i\in[n]}\sum_{k=k_{t_n+s}}^{k_{t_n+t}-2}\eta_{k+1}^i Q^i.$$

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

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

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

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

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

We begin by providing a bound on the second moments of $E_{1,1}^n$ and $E_{1,2}^n$. Recalling the definition of ψ, and using the compactness of the support of ϕ along with Assumption 1, we obtain that

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

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

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

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

by the exchangeability, thus, the sequence of random variables $(\bar{X}^{1,n}_{t_n+\cdot})_n$ is tight. Let us work on the random variables $U^1_n := \pi_{[0,t-s]}\#\bar{X}^1_{t_n+s+\cdot}$ and $U^2_n := \pi_{[0,t-s]}\#\bar{X}^2_{t_n+s+\cdot}$ defined on the set $\mathcal{C}([0,t-s])$ of continuous functions on the interval $[0,t-s]$. Since $(\bar{X}^{1,n}_{t_n+\cdot})_n$ is tight, given an arbitrary $\varepsilon > 0$, there is a compact set $\mathcal{K}_\varepsilon \subset \mathcal{C}([0,t-s])$ such that

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

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

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

We thus can write

$$\begin{aligned}\mathbb{E}\left|E^n_{1,2}\right| &= \mathbb{E}\left|E^n_{1,2}\right| \mathbb{1}_{(U^1_n,U^2_n)\in\mathcal{K}^2_\varepsilon} + \mathbb{E}\left|E^n_{1,2}\right| \mathbb{1}_{(U^1_n,U^2_n)\notin\mathcal{K}^2_\varepsilon} \\ &\leq (t-s) \sup_{f,g\in\mathcal{K}_\varepsilon} \max_{\substack{u,v\in[0,t-s]\\ |u-v|\leq\bar{\gamma}_{k_{t_n+s}}}} |\psi(f(u),g(u)) - \psi(f(v),g(v))| + \sqrt{\mathbb{E}(E^n_{1,2})^2}\sqrt{2\mathbb{P}\left[U^1_n \notin \mathcal{K}_\varepsilon\right]}.\end{aligned} \tag{35}$$

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

$$\mathcal{S} := \{u \mapsto \psi(f(u), g(u)) \ : \ f, g \in \mathcal{K}_\varepsilon\}$$

is a set of uniformly equicontinuous functions. As a consequence, the first term at the right hand side of the inequality in (35) converges to zero as $n \to \infty$, since $\bar{\gamma}_{k_{t_n+s}} \to 0$. The second term is bounded by $C\sqrt{\varepsilon}$ thanks to the bound we obtained on $\mathbb{E}(E^n_{1,2})^2$. Since ε is arbitrary, we obtain that $\mathbb{E}|E^n_{1,2}| \to 0$, thus, $\mathbb{E}|\chi^n_1| \to 0$.

The terms χ^2_n, χ^3_n, and χ^5_n are dealt with similarly to χ^1_n. Considering χ^2_n, we have by the exchangeability that $\mathbb{E}|\chi^2_n| \leq C\mathbb{E}|E^n_1|$, with

$$\begin{aligned}E^n_1 &= \phi(\bar{X}^1_{t_n+t}) - \phi(X^1_{k_{t_n+t}-1}) + \phi(X^1_{k_{t_n+s}}) - \phi(\bar{X}^1_{t_n+s}) \\ &= \phi(\bar{X}^1_{t_n+t}) - \phi(\bar{X}^1_{\tau_{k_{t_n+t}-1}}) + \phi(\bar{X}^1_{\tau_{k_{t_n+s}}}) - \phi(\bar{X}^1_{t_n+s}).\end{aligned}$$

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

$$|E^n_1| \leq 2 \max_{\substack{u,v\in[0,t-s]\\ |u-v|\leq\bar{\gamma}_{k_{t_n+s}}}} \left|\phi(U^1_n(u)) - \phi(U^1_n(v))\right|.$$

Taking $\varepsilon > 0$, selecting the compact $\mathcal{K}_\varepsilon \subset \mathcal{C}([0,t-s])$ as we did for χ^n_1, and recalling that the function ϕ is bounded, we have

$$\mathbb{E}|E^n_1| \leq 2 \sup_{f\in\mathcal{K}_\varepsilon} \max_{\substack{u,v\in[0,t-s]\\ |u-v|\leq\bar{\gamma}_{k_{t_n+s}}}} \|\phi(f(u)) - \phi(f(v))\| + C\mathbb{P}\left[U^1_n \notin \mathcal{K}_\varepsilon\right],$$

and we obtain the $\mathbb{E}|\chi^2_n| \to 0$ by the same argument as for χ^1_n.

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

$$\mathbb{E}|\chi^n_5| \leq \sum_{k=k_{t_n+s}}^{k_{t_n+t}-2} \gamma_{k+1}\mathbb{E}|E^{1,n}_k|,$$

where

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

satisfies

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

Therefore,

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

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

Considering the term χ_n^4, we have by exchangeability

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

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

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

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

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

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

We now turn to χ_6^n. Here we write

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

where

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

satisfies

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

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

The treatment of the term χ_7^n is similar and is omitted.

We finally deal with χ_8^n that involves the martingale increments η_k^i. We decompose this term by writing

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

Since the random vectors $\xi_1^{k+1}, \dots, \xi_n^{k+1}$ are decorrelated conditionally to $\mathcal{F}_k^n$ by Assumption 3-(ii), we obtain that

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

and by the martingale property,

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

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

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

It results that $\mathbb{E}(\chi_8^n)^2 \to 0$. The proof of Prop. 12 is completed. □

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

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

5.3 Proof of Theorem 1

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

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

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

Lemma 8. *The collection of r.v. $(M_t^n, \quad t \ge 0, n \in \mathbb{N}^*)$ is tight in $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$.*

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

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

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

Let us denote by $\mathscr{M}$ the set of weak$\star$ accumulation points of the net $((M_t^n)_\#\mathbb{P} : t \ge 0, n \in \mathbb{N}^*)$, as $(t,n) \to (\infty,\infty)$. By Lemma 8, $\mathscr{M}$ is a non empty subset of $\mathcal{P}(\mathcal{P}(\mathcal{P}_p(\mathcal{C})))$. Define:

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

Lemma 9. *For every $\Upsilon \in \mathscr{M}$, $\Upsilon(\mathcal{V}_p) = 1$.*

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

$$\begin{aligned}\int\int \mathsf{W}_p(m,\mathsf{V}_p)dM(m)d\Upsilon(M) &= \lim_{n\to\infty} \mathbb{E}(\langle \mathsf{W}_p(\,.\,,\mathsf{V}_p), M^n_{t_n}\rangle) \\ &= \lim_{n\to\infty} \frac{1}{t_n}\int_0^{t_n} \mathbb{E}(\mathsf{W}_p(\Phi_s(m^n),\mathsf{V}_p))ds \\ &\leq \limsup_{(t,n)\to(\infty,\infty)} \mathbb{E}(\mathsf{W}_p(\Phi_t(m^n),\mathsf{V}_p)) = 0\,,\end{aligned}$$

where the last equality is due to Prop. 5. As V_p is closed by Prop. 2, this concludes the proof. ☐

For every $t \geq 0$, define $(\Theta_t)_{\#\#} = ((\Theta_t)_\#)_\#$. Define:

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

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

Lemma 10. *For every $\Upsilon \in \mathscr{M}$, $\Upsilon(\mathcal{I}) = 1$.*

Proof. Similarly to the proof of Lemma 9, assume without restriction that $\Upsilon = \lim_{n\to\infty}(M^n_{t_n})_\#\mathbb{P}$ in the weak$\star$ sense. Set $t > 0$. The map $M \mapsto d_L(M,(\Theta_t)_{\#\#}M)$ is continuous on $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$, where we recall that d_L stands for the Lévy-Prokhorov distance. Thus, by Fatou's lemma,

$$\int d_L(M,(\Theta_t)_{\#\#}M)d\Upsilon(M) \leq \limsup_{n\to\infty} \mathbb{E}(d_L(M^n_{t_n},(\Theta_t)_{\#\#}M^n_{t_n}))\,. \tag{37}$$

Note that:

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

In particular, for every Borel set $A \subset \mathcal{P}_p(\mathcal{C})$, $|(\Theta_t)_{\#\#}M^n_{t_n}(A) - M^n_{t_n}(A)| \leq 2t/t_n$. The Lévy-Prokhorov distance being bounded by the total variation distance, $d_L(M^n_{t_n},(\Theta_t)_{\#\#}M^n_{t_n}) \leq 2t/t_n$ 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. ☐

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

Proposition 13. *For every $\Upsilon \in \mathscr{M}$, $\Upsilon(\mathcal{B}_p) = 1$.*

Proof. Consider an arbitrary sequence of the form $((M^n_{t_n})_\#\mathbb{P})$ where $t_n \to \infty$, converging in distribution to some measure $\Upsilon \in \mathscr{M}$ as $n \to \infty$. By Lemma 10, the map $(\Theta_t)_\# : \mathcal{P}_p(\mathcal{C}) \to \mathcal{P}_p(\mathcal{C})$ preserves the measure M, for all M Υ-a.e., and for all t. By Lemma 9, $M(\mathsf{V}_p) = 1$. Thus, the restriction of the map $(\Theta_t)_\#$ to V_p, still denoted by $(\Theta_t)_\# : \mathsf{V}_p \to \mathsf{V}_p$ preserves the measure M as well, for all M Υ-a.e.. By the Poincaré recurrence theorem, stated in Theorem 2.3 of [Mañ87], it follows that $M(\mathrm{BC}_p) = 1$ for all M Υ-a.e. ☐

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

$$\mathbb{E}(\langle \mathsf{W}_p(\,.\,,\mathrm{BC}_p), M^n_{t_n}\rangle) > \varepsilon\,, \tag{38}$$

where $\langle\,.\,,\,.\,\rangle$ is the natural dual pairing on $C_b(\mathcal{P}_p(\mathcal{C}))\times\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$. By Lemma 8, one can extract an other subsequence, which we still denote by $((M^n_{t_n})_\#\mathbb{P})$, converging to $\Upsilon \in \mathscr{M}$. As a consequence,

$$\lim_{n\to\infty} \mathbb{E}((\mathsf{W}_p(\,.\,,\mathrm{BC}_p), M^n_{t_n})) = \int\int \mathsf{W}_p(m,\mathrm{BC}_p)dM(m)d\Upsilon(M) = 0\,,$$

where we used the fact that, due to Prop. 13, $\int \mathsf{W}_p(m,\mathrm{BC}_p)dM(m) = 0$ for Υ-amost all M. This contradicts Eq. (38). ☐

5.4 Proof of Corollary 1

Throughout this paragraph, we assume that $1 \le p \le 2$. Consider the r.v.

$$Y_n(s) := W_p\left(\frac{1}{n}\sum_{i=1}^n \delta_{\bar X_s^{i,n}}, (\pi_0)_\# \operatorname{BC}_p\right).$$

Lemma 11. *The r.v.* $(Y_n(s)^p : s > 0, n \in \mathbb{N})$ *are uniformly integrable.*

Proof. Note that $Y_n(s)^p \le C(1 + \frac{1}{n}\sum_i \|\bar X_s^{i,n}\|^p)$. Hence, for each $a > 0$, $Y_n(s)^p \mathbb{1}_{Y_n(s)>a} \le \frac{1}{n}\sum_{i\in n} C(1 + \|\bar X_s^{i,n}\|^p)\mathbb{1}_{C(1+\|\bar X_s^{i,n}\|^p)>a}$. By the exchangeability stated in Assumption 3, the random variables $(Y_n(s) : s > 0, n \in \mathbb{N})$ are uniformly integrable if the random variables $(\|\bar X_s^{1,n}\|^p : s > 0, n \in \mathbb{N})$ are uniformly integrable. We conclude using Assumption 4–(i) if $p < 2$, or Assumption 4–(ii) if $p = 2$. □

Recall the definition of M_t^n in Eq. (36), and recall that $\mathscr{M}$ is the set of cluster points of $((M_t^n)_\#\mathbb{P} : t \ge 0, n \in \mathbb{N}^*)$ as $(t,n) \to (\infty,\infty)$. Consider an arbitrary sequence $t_n \to \infty$, such that $(M_{t_n}^n)_\#\mathbb{P}$ converges to some measure $\Upsilon \in \mathscr{M}$. Consider $\varepsilon > 0$. By Lemma 11; there exists $a > 0$ such that $\sup_{n,s}\mathbb{E}(Y_n(s)\mathbb{1}_{Y_n(s)>a}) < \varepsilon$. Using the inequality $y \le a \wedge y + y\mathbb{1}_{y>a}$, we obtain:

$$\begin{aligned}\mathbb{E}\left(\frac{1}{t_n}\int_0^{t_n} Y_n(s)ds\right) &\le \mathbb{E}\left(\frac{1}{t_n}\int_0^{t_n} a \wedge Y_n(s)ds\right) + \varepsilon \\ &= \mathbb{E}\left(\int a \wedge W_p((\pi_0)_\# m, (\pi_0)_\# \operatorname{BC}_p)dM_{t_n}^n(m)\right) + \varepsilon \end{aligned} \tag{39}$$

The restriction of π_0 to $\mathcal{P}_p(\mathcal{C})$, which we still denote by π_0, is continuous on $(\mathcal{P}_p(\mathcal{C}), \mathsf{W}_p) \to (\mathcal{P}_p(\mathbb{R}^d), W_p)$, where W_p represents the p-th order Wasserstein distance on $\mathcal{P}(\mathbb{R}^d)$. As a consequence, the pushforward map $(\pi_0)_\# : \mathcal{P}(\mathcal{P}_p(\mathcal{C})) \to \mathcal{P}(\mathcal{P}_p(\mathbb{R}^d))$ is continuous. Therefore, as $(\pi_0)_\# \operatorname{BC}_p$ is non empty by Prop. 3, the function $M \mapsto \int a \wedge W_p((\pi_0)_\# m, (\pi_0)_\# \operatorname{BC}_p)dM(m)$ is bounded and continuous on $\mathcal{P}(\mathcal{P}_p(\mathcal{C}))$. Recall that $M_{t_n}^n$ converges in distribution to Υ, and noting that, by Prop. 3,

$$\int\int W_p((\pi_0)_\# m, (\pi_0)_\# \operatorname{BC}_p)dM(m)d\Upsilon(M) = 0.$$

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

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

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 $k_t := \inf\{k : \sum_{i=1}^k \gamma_i \ge t\}$, and τ_k in Eq. (4). Let (α_n) be a sequence of integers tending to infinity. By the triangular inequality,

$$\begin{aligned}\mathbb{E}\left(\frac{\sum_{l=1}^{\alpha_n}\gamma_l W_p(\mu_l^n, (\pi_0)_\# \operatorname{BC}_p)}{\sum_{l=1}^{\alpha_n}\gamma_l}\right) &= \mathbb{E}\left(\frac{1}{\tau_{\alpha_n}}\int_0^{\tau_{\alpha_n}} W_p(\mu_{k_s}^n, (\pi_0)_\# \operatorname{BC}_p)ds\right) \\ &\le \mathbb{E}\left(\frac{1}{\tau_{\alpha_n}}\int_0^{\tau_{\alpha_n}} W_p(\mu_{k_s}^n, \frac{1}{n}\sum_{i\in[n]}\delta_{\bar X_s^{i,n}})ds\right) \\ &\quad + \mathbb{E}\left(\frac{1}{\tau_{\alpha_n}}\int_0^{\tau_{\alpha_n}} Y_n(s)ds\right).\end{aligned}$$

The second 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 term vanishes. For an arbitrary integer l and $s \in [\tau_l, \tau_{l+1}]$,

$$\mathbb{E}\left[W_p\left(\mu_l^n, \frac{1}{n}\sum_{i\in[n]}\delta_{\bar X_s^{i,n}}\right)\right] \le \mathbb{E}\left(\left(\frac{1}{n}\sum_{i\in[n]}\|X_l^{i,n} - \bar X_s^{i,n}\|^p\right)^{1/p}\right) \le (\mathbb{E}(\|X_l^{1,n} - \bar X_s^{1,n}\|^p))^{1/p}.$$

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

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

where we used Assumptions 1, and 3. Consequently,

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

As Assumption 2 holds, $C(\gamma_{l+1}^{p/2} + \gamma_{l+1}^p) \to_{l\to\infty} 0$, 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 $n \to \infty$. Hence,

$$\lim_{n\to 0}\mathbb{E}\left(\frac{\sum_{l=1}^{\alpha_n}\gamma_l W_p(\mu_l^n, (\pi_0)_\# \mathrm{BC}_p)}{\sum_{l=1}^{\alpha_n}\gamma_l}\right) = 0\,,$$

for an arbitrary sequence (α_n) diverging to ∞. By Markov's inequality, Corollary 1 is proven.

5.5 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*

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

Proof. Assume for the sake of contradiction that

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

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

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

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

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

and we obtain our contradiction. □

We now prove Theorem 2. Recall that the collection $\{\Phi_t(m^n)\}$ is tight in $\mathcal{P}_p(\mathcal{C})$ by Prop. 4. Let (t_n, φ_n) be a sequence such that $(t_n, \varphi_n) \to_n (\infty, \infty)$ and such that $(\Phi_{t_n}(m^{\varphi_n}))_n$ converges in distribution to $M \in \mathcal{M}$ as given by (12). To prove Theorem 2, it will be enough to show that

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

This shows indeed that

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

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

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

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

$$\exists T > 0, \quad \limsup_n \widetilde{\mathbb{P}} \left(W_p \left(\tilde{m}^n_T, A_p \right) \geq \delta \right) \leq \varepsilon. \tag{41}$$

to establish our theorem. Observing that the function $\rho \mapsto (\pi_0)_\# \rho$ is a continuous $\mathcal{P}(\mathcal{C}) \to \mathcal{P}_p(\mathbb{R}^d)$ 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 Ψ and to the compact K, we set $T > 0$ in such a way that

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

By the triangular inequality, we have

$$W_p \left(\tilde{m}^n_T, A_p \right) \leq W_p \left(\tilde{m}^n_T, \tilde{m}^\infty_T \right) + W_p \left(\tilde{m}^\infty_T, A_p \right).$$

The first term at the right hand side converges to zero for each $\tilde{\omega} \in \widetilde{\Omega}$ by the continuity of the function $\rho \mapsto (\pi_T)_\# \rho$, thus, this convergence takes place in probability. We also know that for $\widetilde{\mathbb{P}}$–almost all $\tilde{\omega} \in \widetilde{\Omega}$, it holds that $\tilde{m}^\infty \in \mathsf{V}_p$. Thus, regarding the second term, we have $\tilde{m}^\infty_T = \Psi_T(\tilde{m}^\infty_0)$ for these $\tilde{\omega}$, and we can write

$$\widetilde{\mathbb{P}} \left(W_p \left(\tilde{m}^\infty_T, A_p \right) \geq \delta \right) \leq \widetilde{\mathbb{P}} \left(\tilde{m}^\infty \notin \mathcal{K} \right) + \widetilde{\mathbb{P}} \left(\left(W_p \left(\Psi_T(\tilde{m}^\infty_0), A_p \right) \geq \delta \right) \cap \left(\tilde{m}^\infty_0 \in K \right) \right).$$

When $\tilde{m}^\infty_0 \in K$, it holds that $W_p \left(\Psi_T(\tilde{m}^\infty_0), A_p \right) \leq \delta/2$, thus, the second term at the right hand side of the last inequality is zero. The first term satisfies $\widetilde{\mathbb{P}} \left(\tilde{m}^\infty \notin \mathcal{K} \right) = 1 - M(\mathcal{K}) \leq \varepsilon/2$, and the statement (41) follows. Theorem 2 is proven.

6 Proofs of Section 4.1

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

6.1 Proof of Prop. 6

Lemma 13. *Let $\rho \in \mathsf{V}_2$. For every $t > 0$, ρ_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:*

$$\inf_{t \in [t_1, t_2], \|x\| \leq R} \varrho(t, x) \geq C_{R,t_1,t_2}, \tag{42}$$

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

$$\sup_{x \in \mathbb{R}^d, t \in [t_1, t_2]} \|\nabla \varrho(t, x)\| + \varrho(t, x) \leq C_{t_1,t_2}. \tag{43}$$

Finally,

$$\sup_{t \in [t_1, t_2]} \int (1 + \|x\|^2) \|\nabla \varrho(t, x)\| \, dx < \infty. \tag{44}$$

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

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

Moreover,

$$\sup_{t\in[0,T]} \tilde{b}(t,x) \leq C(1+\|x\| + \int \sup_{t\in[0,T]} \|y_t\|\, d\rho(y)) \leq C(1+\|x\|)\,. \tag{45}$$

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

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

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. (45), 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:

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

and $\|\nabla\varrho(t,x)\| \leq C_{2,t_2}t_1^{-(d+1)/2}$. Consequently, ρ satisfies Eq. (42), Eq. (43) and Eq. (44). □

For every $\rho \in \mathsf{V}_2$ and every $t > 0$, recall the definition of the velocity field v_t in Eq. (16): $v_t(x) := -\nabla V(x) - \int \nabla U(x,y)d\rho_t(y) - \sigma^2\nabla \log \varrho(t,x)$, where $\varrho(t,x)$ is the density of ρ_t defined in Lem. 13.

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

$$\int_{t_1}^{t_2} \int \|v_t(x)\| d\rho_t(x)dt < \infty\,. \tag{46}$$

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

$$\int \psi(t_2,x)d\rho_{t_2}(x) - \int \psi(t_1,x)d\rho_{t_1}(x) = \int_{t_1}^{t_2} \int (\partial_t\psi(t,x) + \langle\nabla_x\psi(t,x), v_t(x)\rangle)\rho_t(dx)dt\,. \tag{47}$$

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

$$\int \psi(t_2,x)d\rho_{t_2}(x) - \int \psi(t_1,x)d\rho_{t_1}(x) =$$
$$\int_{t_1}^{t_2} \int (\partial_t\psi(t,x) + \langle\nabla\psi(s,x), b(x,\rho_t)\rangle + \sigma^2\Delta\psi(t,x))\rho_t(dx)dt\,. \tag{48}$$

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 (48) holds in fact for any smooth compactly supported ψ. Using Lemma 13 and an integration by parts of the Laplacian term, Eq. (47) follows. □

The goal now is to establish that the functional $\mathscr{H}$ is a Lyapunov function. This claim will follow from the application of Eq. (47) to the functional $(t,x) \mapsto \sigma^2\log(\varrho(t,x)) + V(x) + \int U(x-y)\varrho(t,y)dy$. However, this function is not necessarily smooth nor compactly supported. In order to be able to apply Lem. 14, mollification should be used. In the sequel, consider two fixed positive numbers $t_2 > t_1$.

Consider a smooth, compactly supported, even function $\eta : \mathbb{R}^d \to \mathbb{R}_+$ such that $\int \eta(x)dx = 1$, and define $\eta_\varepsilon(x) := \varepsilon^{-d}\eta(x/\varepsilon)$ for every $\varepsilon > 0$. For every $t > 0$, we introduce the density

$\varrho_\epsilon(t,\cdot) := \eta_\varepsilon * \rho_\epsilon(t,\cdot)$, and we denote by $\rho_t^\varepsilon(dx) = \varrho_\epsilon(t,x)dx$ the corresponding probability measure. Finally, we define:

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

With these definitions at hand, it is straightforward to check that the statements of Lem. 14 hold when ρ_t, v_t are replaced by $\rho_t^\varepsilon, v_t^\varepsilon$. More specifically, we shall apply Eq. (47) using a specific smooth function $\psi = \psi_{\varepsilon,\delta,R}$, which we will define hereafter for fixed values of $\delta, R > 0$, yealding our main equation:

$$\int \psi_{\varepsilon,\delta,R}(t_2,x)\varrho_\varepsilon(t_2,x)dx - \int \psi_{\varepsilon,\delta,R}(t_1,x)\varrho_\varepsilon(t_1,x)dx = \\ \int_{t_1}^{t_2}\int (\partial_t\psi_{\varepsilon,\delta,R}(t,x) + \langle \nabla\psi_{\varepsilon,\delta,R}(t,x), v_t^\varepsilon(x)\rangle)\varrho_\varepsilon(t,x)dxdt\,. \quad (49)$$

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

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

We extend $\psi_{\varepsilon,\delta,R}$ to a smooth compactly supported function on $\mathbb{R}_+\times\mathbb{R}^d$, and we apply Eq. (49) to the latter. We now investigate the limit of both sides of the equality (49) as δ,ε,R successively tend to $0,0,\infty$. First consider the lefthand side. Note that for all $t\in[t_1,t_2]$,

$$\lim_{\varepsilon\to0}\lim_{\delta\to0}\psi_{\varepsilon,\delta,R}(t,x)\varrho_\varepsilon(t,x) := \left(\sigma^2\log\varrho(t,x) + V(x) + \int U(x-y)\chi_R(y)\varrho(t,y)dy\right)\varrho(t,x)\chi_R(x)\,.$$

The domination argument that allows to interchange limits and integrals is provided by Lem 13. Indeed, for a fixed $R>0$, there exists a constant C_R such that $\varrho^{\epsilon,\delta}(t,x)\le C_R$ and $\psi_{\varepsilon,\delta,R}(t,x)\le C_R$ for all $\|x\|\le R$ and all $t\in[t_1,t_2]$. As a consequence,

$$\lim_{\varepsilon\to0}\lim_{\delta\to0}\int\psi_{\varepsilon,\delta,R}(t,x)\varrho_\varepsilon(t,x) = \sigma^2\int\chi_R(x)\varrho(t,x)\log\varrho(t,x)dx + \\ \int V(x)\chi_R(x)d\rho_t(x) + \int U(x-y)\chi_R(y)\chi_R(x)\varrho(t,x)\varrho(t,y)dxdy\,.$$

By Eq. (43), the first term in the l.h.d. of the above equation converges to $\sigma^2\int\varrho(t,x)\log\varrho(t,x)dx$ as $R\to\infty$. Similarly, $\int V(x)\chi_R(x)d\rho_t(x)$ tends to $\int Vd\rho_t$ as $R\to\infty$, by use of the linear growth condition on ∇V in Assumption 5, along with the fact that ρ_t admits a second order moment. The same holds for the last term. Finally, we have shown that, for every $t\in[t_1,t_2]$,

$$\lim_{R\to\infty}\lim_{\varepsilon\to0}\lim_{\delta\to0}\int\psi_{\varepsilon,\delta,R}(t,x)\varrho_\varepsilon(t,x)dx = \mathscr{H}(\rho_t) + \frac{1}{2}\iint U(x-y)d\rho_t(y)d\rho_t(x)\,,$$

where we recall the definition $\mathscr{H}(\rho_t) := \sigma^2\int\log\varrho(t,\cdot)d\rho_t + \int Vd\rho_t + \frac{1}{2}\iint U(x-y)d\rho_t(y)d\rho_t(x)$. As δ,ε,R successively tend to $0,0,\infty$, we have shown that the l.h.s. of Eq (49) converges to:

$$\mathscr{H}(\rho_{t_2}) - \mathscr{H}(\rho_{t_1}) + \frac{1}{2}\iint U(x-y)d\rho_{t_2}(y)d\rho_{t_2}(x) - \frac{1}{2}\iint U(x-y)d\rho_{t_1}(y)d\rho_{t_1}(x)\,. \quad (51)$$

We should now identify the above term with the limit of the r.h.s. of Eq. (49) in the same regime. The latter is composed of two terms. First consider the second term:

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

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

$$\|\nabla \psi_{\varepsilon,\delta,R}(t,x)\| \le C_R \,. \tag{52}$$

Using Eq. (52) along with Eq. (46), the dominated convergence applies. Letting $\varepsilon \to 0$ in a second step, the exact same argument applies, and we obtain:

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

where the interchange between ∇ and the limits is again a consequence of Lem. 13. We now write the gradient in the above inner product. Note that:

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

We obtain:

$$\begin{aligned}
\lim_{\varepsilon\to 0}\lim_{\delta\to 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon,\delta,R}(t,x), v_t^\varepsilon(x)\rangle \varrho_\varepsilon(t,x)dxdt &= -\int_{t_1}^{t_2}\int \|v_t(x)\|^2 \chi_R(x)\varrho(t,x)dxdt \\
&- \int_{t_1}^{t_2}\int \langle v_t(x), \int (1-\chi_R(y))\nabla U(x-y)d\rho_t(y)\rangle \chi_R(x)d\rho_t(x) \\
&- \int_{t_1}^{t_2}\int \langle v_t(x), \nabla\chi_R(x)(V(x) + \int U(x-y)\chi_R(y)d\rho_t(y))\rangle d\rho_t(x)\,.
\end{aligned} \tag{53}$$

By the dominated convergence theorem, Assumption 5 and Eq. (44), the last two terms in the r.h.s. of Eq.(53) tend to zero as $R \to \infty$, while the first term is handled by the monotone convergence theorem. We thus obtain:

$$\lim_{R\to\infty}\lim_{\varepsilon\to 0}\lim_{\delta\to 0} \int_{t_1}^{t_2} \int \langle \nabla \psi_{\varepsilon,\delta,R}(t,x), v_t^\varepsilon(x)\rangle \varrho_\varepsilon(t,x)dxdt = -\int_{t_1}^{t_2}\int \|v_t(x)\|^2 \varrho(t,x)dxdt\,. \tag{54}$$

As a last step, we should evaluate the limit of the first term in the r.h.s. of Eq. (49), which writes: $\int_{t_1}^{t_2}\int \partial_t \psi_{\varepsilon,\delta,R}(t,x)\varrho_\varepsilon(t,x)dxdt$. 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 $\delta \in (0,1)$, $t \in [t_1,t_2]$, $x \in \mathbb{R}^d$,*

$$|\partial_t \psi_{\varepsilon,\delta,R}(t,x)| \le C_{R,\varepsilon}\,, \tag{55}$$

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

By Eq. (55) and by the continuity of the map $t \mapsto \partial_t \varrho^\varepsilon$ (see the proof of Lemma 15), we can expand the first term in the r.h.s. of Eq. (49) as:

$$\int_{t_1}^{t_2}\int \partial_t \psi_{\varepsilon,\delta,R}(t,x)d\rho_t^\varepsilon(x)dt = \int_{t_1}^{t_2}\int \partial_t \psi_{\varepsilon,\delta,R}(t,x)\varrho^{\varepsilon,\delta}(t,x)dxdt + o_{\varepsilon,R}(\delta)\,, \tag{56}$$

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

$$\partial_t \psi_{\varepsilon,\delta,R}(t,x) = \sigma^2 \frac{\partial_t \varrho^{\varepsilon,\delta}(t,x)}{\varrho^{\varepsilon,\delta}(t,x)}\chi_R(x) + \int U_\varepsilon(x-y)\chi_R(y)\chi_R(x)\partial_t\varrho^{\varepsilon,\delta}(t,y)dy\,. \tag{57}$$

Plugging this equality into (56) and noting that U_ε is even (because U and η_ε are), we obtain:

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

By the dominated convergence theorem, we finally obtain:

$$
\begin{aligned}
&\lim_{R\to\infty}\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\int_{t_1}^{t_2}\int \partial_t\psi_{\varepsilon,\delta,R}(t,x)d\rho_t^\varepsilon(x)dt=\\
&\qquad\frac{1}{2}\int\int U(x-y)\varrho(t_2,x)\varrho(t_2,y)dxdy-\frac{1}{2}\int\int U(x-y)\varrho(t_1,x)\varrho(t_1,y)dxdy\,. \quad (58)
\end{aligned}
$$

Putting together Eq. (51), (54) and (58), and passing to the limit in the continuity equation (49), the statement of Prop. 6 follows.

Proof of Lem. 15. Using Eq. (49) and integration by parts,

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

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

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

Considering the second term in the r.h.s. of Eq. (57), the presence of the product of the compactly supported functions $\chi_R(x)\chi_R(y)$ implies that the former is bounded in absolute value:

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

On the otherhand, using the lower bound (42), the first term in the r.h.s. of Eq. (57), is also bounded, and finally, Eq. (55) follows.

6.2 Proof of Prop. 7

The map $\overline{\mathscr{H}}:\rho\mapsto\mathscr{H}(\rho_\epsilon)$ is real valued and lower semicontinuous by Prop. 6 and Fatou's lemma. Moreover, for every $\rho\in\mathsf{V}_2$, $\overline{\mathscr{H}}(\Phi_t(\rho))-\overline{\mathscr{H}}(\rho)=\mathscr{H}(\rho_{t+\epsilon})-\mathscr{H}(\rho_\epsilon)=-\int_\epsilon^{t+\epsilon}\int\|v_s\|^2d\rho_sds$. Therefore, $\overline{\mathscr{H}}(\Phi_t(\rho))$ is decreasing w.r.t. t, and, as such, $\overline{\mathscr{H}}$ is a Lyapunov function. In addition, the identity $\overline{\mathscr{H}}(\Phi_t(\rho))=\overline{\mathscr{H}}(\rho)$ for all t, is equivalent to: $v_t=0$ ρ_t-a.e., for every $t\ge\epsilon$. By Lem. 14, this implies that $\rho_t=\rho_\epsilon$ for all $t\ge\epsilon$. Thus, $\overline{\mathscr{H}}(\Phi_t(\rho))=\overline{\mathscr{H}}(\rho)$ for all t, if and only if $v_\epsilon=0$ and $\rho_t=\rho_\epsilon$ for all t. This means that $\overline{\mathscr{H}}$ is a Lyapunov function for the set Λ_ϵ. The first point is proven.

Consider a recurrent point $\rho\in\mathsf{V}_2$, say $\rho=\lim\Phi_{t_n}(\rho)$. By Prop. 3, $\rho\in\Lambda_\epsilon$, for any $\epsilon>0$. This means that there exists $\mu\in\mathcal{S}$ such that $\rho_t=\mu$ for all $t>0$. By continuity of the map $(\pi_0)_\#$, $\rho_0=\lim\rho_{t_n}$. Thus, $\rho_0=\mu$. This means that $\rho_t=\mu$ for all $t\ge0$, which writes $\rho\in\Lambda_0$. The proof is complete.

6.3 Proof of Prop. 8

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

1. $\nu_t \to \nu_0$ as $t \downarrow 0$.
2. $\sup_{t\in[0,T]} \int \|x\|^2 \nu_t(dx) < \infty$ for each $T > 0$.
3. The measure ν_t has a density $\eta_t = d\nu_t/d\mathscr{L}^d$ for each $t > 0$. This density satisfies $\eta_t \in L^1_{\text{loc}}((0,\infty); W^{1,1}_{\text{loc}}(\mathbb{R}^d))$.
4. The continuity equation
$$\partial_t \nu_t + \nabla \cdot (\nu_t w_t) = 0$$
is satisfied in the distributional sense, where
$$w_t(x) = -\frac{\sigma^2 \nabla \eta_t(x)}{\eta_t(x)} - \nabla V(x) - \int \nabla U(x-y)\eta_t(y)dy.$$
5. $\|w_t\|_{L^2(\nu_t)} \in L^2_{\text{loc}}(0,\infty)$.

Furthermore, the function $t \mapsto \nu_t$ is the solution of the gradient flow in $\mathcal{P}_2(\mathbb{R}^d)$ of the functional $\mathscr{H}$ provided in the statement, and $w_t \in -\partial\mathscr{H}(\nu_t)$, where $\partial\mathscr{H}$ is the Fréchet sub-differential of $\mathscr{H}$. From the general properties of the gradient flows detailed in [AGS08, Chap. 11], one can then check that we can write $\nu_t = \Psi_t(\nu_0)$ where Ψ 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 $\rho \in \mathsf{V}_2$, the function $t \mapsto \rho_t$ satisfies the five properties stated above. The first two hold true for each $\zeta \in \mathcal{P}_2(\mathcal{C})$: to check the first one, let $X \sim \zeta$. Observe that $X_t \to_{t\to 0} X_0$ by continuity and that $\|X_t - X_0\|^2 \leq 2\sup_{s\in[0,1]} \|X_s\|^2$ for t small, and use the Dominated Convergence. The second property follows from the very definition of $\mathcal{P}_2(\mathcal{C})$. Property 3 follows from Lemma 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, Equation (15). This completes the proof of Proposition 8.

6.4 Proof of Prop. 9

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

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

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

$$\begin{aligned}\|X^i_{k+1}\|^2 = \|X^i_k\|^2 - 2\gamma\langle \nabla V(X^i_k), X^i_k\rangle - 2\frac{\gamma}{n}\sum_j \langle \nabla U(X^i_k - X^j_k), X^i_k\rangle + 2\gamma\|\xi^i_{k+1}\|^2 \\ + \sqrt{2\gamma}T^i_{k+1} + \gamma^2\|\nabla V(X^i_k) + \frac{\gamma}{n}\sum_j \nabla U(X^i_k - X^j_k)\|^2\,.\end{aligned}$$

where we defined:

$$T^i_{k+1} := \langle \xi^i_{k+1}, X^i_k - \gamma\nabla V(X^i_k) - \frac{\gamma}{n}\sum_j \nabla U(X^i_k - X^j_k)\rangle\,.$$

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

$$\|X_{k+1}^i\|^2 \le (1 - \lambda\gamma + C\gamma^2)\|X_k^i\|^2 - \frac{2\gamma}{n}\sum_j \langle \nabla U(X_k^i - X_k^j), X_k^i\rangle + 2\gamma\|\xi_{k+1}^i\|^2 + \sqrt{2\gamma}T_{k+1}^i + C\gamma^2(1 + n^{-1}\sum_j \|X_k^j\|^2)\,. \quad (59)$$

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

$$\sup_{k,n}\left(\mathbb{E}(\|X_k^{1,n}\|^2\|X_k^{2,n}\|^2) + \frac{1}{n}\mathbb{E}(\|X_k^{1,n}\|^4)\right) < \infty\,. \quad (60)$$

To that end, compute the average w.r.t. $i \in [n]$ of both sides of Eq. (59). Setting $S_k := \frac{1}{n}\sum_i \|X_k^i\|^2$, and

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

Eq. (59) leads to, for every k larger than some fixed constant,

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

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

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

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

$$S_{k+1}^2 \le (1-\lambda\gamma)^2 S_k^2 + (2\gamma\chi_{k+1}^\xi + \sqrt{2\gamma}\chi_{k+1}^T + C\gamma)^2 + CS_k(2\gamma\chi_{k+1}^\xi + \sqrt{2\gamma}\chi_{k+1}^T + C\gamma)\,.$$

Thus, for large k,

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

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

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

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

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

We finally obtain: $\mathbb{E}S_{k+1}^2 \le (1-\lambda\gamma)\mathbb{E}S_k^2 + C\gamma^2 + C\gamma\sqrt{\mathbb{E}(S_k^2)}$. This proves that $\mathbb{E}S_k^2$ is bounded uniformly in k, n. By exchangeability, $\mathbb{E}S_k^2 = n^{-1}\mathbb{E}(\|X_k^1\|^4) + (1-1/n)\mathbb{E}(\|X_k^1\|^2\|X_k^2\|^2)$. This proves Eq. (60).

We now expand $\|X_{k+1}^i\|^2$ starting from (59). We use the notation $\nabla U^i := n^{-1}\sum_j \nabla U(X_k^i - X_k^j)$. For all k large enough,

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

We take expectations. Note that $\mathbb{E}\|\xi_{k+1}^i\|^4 \le C$, $\mathbb{E}S_k^2 \le C$ and $\mathbb{E}(\|\xi_{k+1}^i\|^2\|X_k^i\|^2) \le C\mathbb{E}\|X_k^i\|^2 = C\mathbb{E}(S_k) \le C$ (where as usual, C changes at each inequality). Thus

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

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

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

Note that the last term in the above inequality is bounded uniformly in k, n, by Eq. (60). Changing again the constants C, λ, we obtain that for large k,

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

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

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

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

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

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

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

We have shown that:

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

Putting all pieces together,

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

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

6.5 Proof of Th. 4

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

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

With this at hand, using Equation (15) in conjunction with the identity $\rho_\infty = \Psi_t(\rho_\infty)$ for each $t \geq 0$ shows that $u_{\rho_\infty}(x) = 0$ for ρ_∞–almost all x. This shows that $\rho_\infty \in \mathcal{S}$. On the other hand, for $\nu \neq \rho_\infty$ in $\mathcal{P}_2(\mathbb{R}^d)$, we obtain from Equation (15) that the function $t \mapsto \mathscr{H}(\Psi_t(\nu))$ is strictly decreasing. Thus, $\int \|u_\nu\|^2 d\nu > 0$ which shows that $\nu \notin \mathcal{S}$.

7 Proofs of Section 4.2

7.1 Proof of Prop. 10

We recall the iterations:

$$a_{k+1}^{i,n} = a_k^{i,n}(1-\lambda\gamma_{k+1}) + \frac{\gamma_{k+1}}{n}\sum_{j\in[n]}(K(w_0^{i,n}, w_0^{j,n})a_k^{j,n} - Q(w_0^{i,n}) + \gamma_{k+1}\tilde{\zeta}_{k+1}^{i,n} + \sqrt{2\gamma_{k+1}}\tilde{\xi}_{k+1}^{i,n} .$$

We denote $I_{k,n} := \frac{1}{n}\sum_{i\in[n]}(a_k^{i,n})^2$. The proof will be done in two steps. First, we will obtain a bound on $\mathbb{E}(I_{k,n}^2)$. Then, we can bound $\mathbb{E}(a_{k,n}^{i,n})^4$ and the bound on $\mathbb{E}(\tilde{\zeta}_k^{i,n})^4$ follows easily. Observe that:

$$\begin{aligned}(a_{k+1}^{i,n})^2 = (a_k^{i,n})^2(1-\lambda\gamma_{k+1})^2 - 2\frac{\gamma_{k+1}(1-\lambda\gamma_{k+1})}{n}\sum_{j\in[n]} a_k^{i,n}(K(w_0^{i,n}, w_0^{j,n})a_k^{j,n} - Q(w_0^{i,n}))\\ + \sqrt{2\gamma_{k+1}}(1-\gamma_{k+1}\lambda)\tilde{\xi}_{k+1}^{i,n}a_k^{i,n} + (1-\lambda\gamma_{k+1})a_k^{i,n}\gamma_{k+1}\tilde{\zeta}_{k+1}^{i,n}\\ + \left(\frac{\gamma_{k+1}}{n}\sum_{j\in[n]} K(w_0^{i,n}, w_0^{j,n})a_k^{j,n} + \gamma_{k+1}\tilde{\zeta}_{k+1}^{i,n} + \sqrt{2\gamma_{k+1}}\tilde{\xi}_{k+1}^{i,n}\right)^2 . \quad (61)\end{aligned}$$

Since K is a bounded positive semi-definite kernel, $\sum_{i,j\in[n]} a_k^{i,n}K(w_0^{i,n}, w_0^{j,n})a_k^{j,n} \geq 0$. Hence, we obtain:

$$\begin{aligned}\frac{1}{n}\sum_{i\in[n]}(a_{k+1}^{i,n})^2 \leq \frac{1}{n}\sum_{i\in[n]}(a_k^{i,n})^2(1-\lambda\gamma_{k+1}+C\gamma_{k+1}^2)^2 + \frac{2\gamma_{k+1}}{n}\sum_{i\in[n]} a_k^{i,n}Q(w_0^{i,n})\\ + \frac{(1-\lambda\gamma_{k+1})}{n}\sum_{i\in[n]}(\gamma_{k+1}\tilde{\zeta}_{k+1}^{i,n} + \sqrt{2\gamma_{k+1}}\tilde{\xi}_{k+1}^{i,n})a_k^{i,n} + C\gamma_{k+1}^2\frac{1}{n}\sum_{i\in[n]}(\tilde{\zeta}_{k+1}^{i,n})^2 + C\gamma_{k+1}\frac{1}{n}\sum_{i\in[n]}(\tilde{\xi}_{k+1}^{i,n})^2 .\end{aligned}$$

Since φ is bounded, $(\tilde{\zeta}_{k+1}^{i,n})^2 \leq C(1 + y_{k+1}^2 + I_{k,n})$, and

$$\begin{aligned}I_{k+1,n} \leq I_{k,n}(1-2\lambda\gamma_{k+1}+C\gamma_{k+1}^2) + C\gamma_{k+1}\sqrt{I_{k,n}}\\ + \frac{(1-\lambda\gamma_{k+1})}{n}\sum_{i\in[n]}(\gamma_{k+1}\tilde{\zeta}_{k+1}^{i,n} + \sqrt{2\gamma_{k+1}}\tilde{\xi}_{k+1}^{i,n})a_k^{i,n} + C\gamma_{k+1}^2(1+y_{k+1}^2) + C\gamma_{k+1}\sum_{i\in[n]}(\tilde{\xi}_{k+1}^{i,n})^2 .\end{aligned}$$

We have $\mathbb{E}(\tilde{\xi}_{k+1}^{i,n}|\mathcal{F}_k^n) = \mathbb{E}(\zeta_{k+1}^{i,n}|\mathcal{F}_k^n) = 0$ and $\mathbb{E}(y_{k+1}^2 + (\tilde{\xi}_{k+1}^{i,n})^2|\mathcal{F}_k^n) < C$. Raising to the square, taking the expectation of the above inequality, for k large enough, there exists $\tilde{\lambda}$ such that:

$$\mathbb{E}\left[I_{k+1,n}^2\right] \leq \mathbb{E}\left[I_{k,n}^2\right](1-\tilde{\lambda}\gamma_{k+1}) + C\gamma_{k+1}\mathbb{E}\left[I_{k,n}^{3/2}\right] + C\gamma_{k+1}\mathbb{E}\left[I_{k,n}\right] + C\gamma_{k+1}^2 .$$

The latter guarantees, $\sup_{k,n} \mathbb{E}\left[I_{k,n}^2\right] \leq C$. We will keep this in mind and now, going back to Eq. (61), we obtain for k large enough:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i\in[n]}(a_k^{i,n})^4\right] \leq \mathbb{E}\left[\frac{1}{n}\sum_{i\in[n]}(a_{k+1}^{i,n})^4\right](1-\tilde{\lambda}\gamma_{k+1}) + \frac{C\gamma_{k+1}}{n^2}\mathbb{E}\left[\sum_{i,j\in[n]}|a_k^{i,n}|^3(|a_k^{j,n}|+1)\right]$$
$$+\frac{C\gamma_{k+1}}{n}\mathbb{E}\left[\sum_{i\in[n]}(a_k^{i,n})^2\right] + C\gamma_{k+1}^2\mathbb{E}\left[I_{k,n}^2\right] + C\gamma_{k+1}^2\,. \quad (62)$$

The larger term is controlled by Cauchy-Schwartz inequality

$$\sum_{i,j\in[n]}\mathbb{E}\left[|a_k^{i,n}|^3|a_k^{j,n}|\right] \leq \sum_{i,j\in[n]}\sqrt{\mathbb{E}\left[(a_k^{i,n}a_k^{j,n})^2\right]}\sqrt{\mathbb{E}\left[(a_k^{i,n})^4\right]}\,.$$

Using the exchangeability, $\mathbb{E}(I_{k,n}^2) = \frac{1}{n}\mathbb{E}(a_k^{1,n})^4 + \frac{n-1}{n}\mathbb{E}(a_k^{1,n}a_k^{2,n})^2$ and:

$$\frac{1}{n^2}\sum_{i,j\in[n]}\sqrt{\mathbb{E}\left[(a_k^{i,n}a_k^{j,n})^2\right]}\sqrt{\mathbb{E}\left[(a_k^{i,n})^4\right]} = \frac{1}{n}\mathbb{E}\left[(a_k^{1,n})^4\right] + \frac{n-1}{n}\sqrt{\mathbb{E}\left[(a_k^{1,n}a_k^{2,n})^2\right]}\sqrt{\mathbb{E}\left[(a_k^{1,n})^4\right]}$$
$$\leq \sqrt{\mathbb{E}\left[(a_k^{1,n})^4\right]}\sqrt{\frac{1}{n}\mathbb{E}\left[(a_k^{1,n})^4\right] + \frac{n-1}{n}\mathbb{E}\left[(a_k^{1,n}a_k^{2,n})^2\right]}$$
$$\leq \sqrt{\mathbb{E}\left[(a_k^{1,n})^4\right]}\sqrt{\mathbb{E}\left[I_{k,n}^2\right]}\,.$$

Finally, from Eq. (62), the bound on $\mathbb{E}(I_{k,n}^2)$, and for k large enough, we obtain

$$\mathbb{E}\left[(a_{k+1}^{1,n})^4\right] \leq \mathbb{E}\left[(a_{k+1}^{1,n})^4\right](1-\tilde{\lambda}\gamma_{k+1}) + C\gamma_{k+1}\mathbb{E}\left[(a_{k+1}^{1,n})^4\right]^{1/2} + C\gamma_{k+1}\mathbb{E}\left[(a_{k+1}^{1,n})^4\right]^{3/4} + C\gamma_{k+1}^2\,.$$

Consequently, $\sup_{k,n}\mathbb{E}(a_{k+1}^{1,n})^4 < \infty$. Remarking $\mathbb{E}(\zeta_k^{i,n})^4 \leq C\mathbb{E}(a_k^{i,n})^4$, Prop. 10 is proven.

7.2 Proof of Lem.4

Define $A_\mu : w \mapsto \int a d\mu(a|w)$ and $B_\mu : w \mapsto \int a^2 d\mu(a|w)$. We use the notation $K_\varpi f(w) := \int K(w,w')f(w')d\varpi(w)$. We also denote by $\langle f,g\rangle_\varpi := \int f(w)g(w)d\varpi(w)$ the inner product in $L^2(\varpi)$. Define the constant $c := \int y^2 d\nu(x,y)$. Expanding $\mathcal{R}_0(\mu)$, we obtain after some straightforward algebra:

$$\mathcal{R}_0(\mu) = \langle A_\mu, K_\varpi A_\mu - Q\rangle_\varpi + c + \frac{\lambda}{2}\int B_\mu d\varpi\,.$$

Therefore,

$$\mathcal{R}_0(\mu) - \mathcal{R}_0(\mu^*) = \langle A_\mu, K_\varpi A_\mu\rangle - \langle A_{\mu^*}, K_\varpi A_{\mu^*}\rangle_\varpi - \langle Q, A_\mu - A_{\mu^*}\rangle_\varpi + \lambda\int\frac{B_\mu - B_{\mu^*}}{2}d\varpi$$
$$\geq \langle K_\varpi A_{\mu^*} - Q, A_\mu - A_{\mu^*}\rangle_\varpi + \lambda\int\frac{B_\mu - B_{\mu^*}}{2}d\varpi\,, \quad (63)$$

where we use the fact that $\langle A_\mu - A_{\mu^*}, K_\varpi(A_\mu - A_{\mu^*})\rangle_\varpi \geq 0$ to obtain the last inequality. By [AGS08, Lem. 12.4.7], there exists a Borel map on $\mathbb{R}^{d-1} \to \mathcal{P}_2(\mathbb{R}\times\mathbb{R})$ which, to every $w \in \mathbb{R}^{d-1}$, associated a probability measure $\boldsymbol{\gamma}(\cdot|w) \in \Pi_2^0(\mu_*(\cdot|w), \mu(\cdot|w))$, where we recall that $\Pi_2^0(\mu^*(\cdot|w), \mu(\cdot|w))$ is the set of 2-Wasserstein optimal transport plans between $\mu^*(\cdot|w)$ and $\mu(\cdot|w)$, as introduced after Eq. (8). We obtain:

$$\frac{B_\mu(w) - B_{\mu^*}(w)}{2} = \int\frac{a^2 - a_*^2}{2}d\boldsymbol{\gamma}(a_*,a|w) = \int a_*(a-a_*)d\boldsymbol{\gamma}(a_*,a|w) + \int\frac{(a-a_*)^2}{2}d\boldsymbol{\gamma}(a_*,a|w)\,.$$

Substituting this inequality in Eq. (63), and expanding the first term in the r.h.s. of Eq. (63) as a function of $\boldsymbol{\gamma}(\,\cdot\,|w)$, we obtain:

$$\mathcal{R}_0(\mu) - \mathcal{R}_0(\mu^*) \geq \int\int(K_\varpi A_{\mu^*}(w) - Q(w) + \lambda a_*)(a-a_*)d\boldsymbol{\gamma}(a_*,a|w)d\varpi(w)$$
$$+\int\int\frac{(a-a_*)^2}{2}d\boldsymbol{\gamma}(a_*,a|w)d\varpi(w)\,.$$

Note that $K_\varpi A_{\mu^*}(w) - Q(w) + \lambda a_* = -\tilde{b}((a_*, w), \mu^*)$. We obtain:

$$\mathcal{R}_0(\mu)-\mathcal{R}_0(\mu^*) \geq -\int\int \tilde{b}((a_*, w), \mu^*)(a-a_*)d\boldsymbol{\gamma}(a_*, a|w)d\varpi(w)+\int\int \frac{(a-a_*)^2}{2}d\boldsymbol{\gamma}(a_*, a|w)d\varpi(w)\,. \tag{64}$$

We now study $\mathcal{R}_\sigma(\mu) - \mathcal{R}_\sigma(\mu^*)$ for $\sigma > 0$. We make the assumption that $\mu(\,\cdot\,|w)$ admit a density for ϖ-almost every w, which we denote by $\mu(a|w)$ (in the opposite case, $\mathcal{R}_\sigma(\mu) = +\infty$ and there is nothing prove). Then, $\mathcal{R}_\sigma(\mu) = \mathcal{R}_0(\mu) + \sigma^2 \int C_\mu(w)d\varpi(w)$, where $C_\mu(w) := \int \log \mu(a|w)\mu(da|w)$. Since $\int \partial_a(\log \mu^*(a|w))^2 d\mu(a|w) < \infty$ for μ^*-a.e. w, one is able to apply [AGS08, 10.1.1.B, Prop. 9.3.9, Th. 10.4.6], which yields:

$$C_\mu(w) - C_{\mu^*}(w) \geq \int (a-a_*)\partial_a \log \mu^*(a|w)d\boldsymbol{\gamma}(a_*, a|w)\,.$$

Putting all pieces together, the result is complete.

7.3 Proof of Lem. 5

The proof is provided in the case where $\sigma^2 > 0$ (the arguments are simpler when $\sigma^2 = 0$). By Remark 2, the first point is immediate: $\tilde{\pi}_\# \rho_t = \varpi$ for all t.

In order to establish the result, one has two options. The first alternative is to follow step by step the proof of Prop. 6. In the case $\sigma^2 > 0$ (the case $\sigma^2 = 0$ being easier), we establish using [MPZ21], that $\rho_t(\,.\,|w)$ admits a density w.r.t. $\mathscr{L}^1$, which satisfies regularity conditions. In particular, one can prove $\int_{t_1}^{t_2} \int \|v_{\rho_t}\| d\rho_t dt < \infty$ for all $t_2 > t_1 > 0$. Then, by integration by part, it is easy to establish the following continuity equation:

$$\partial_t \rho_t + \nabla \cdot v_{\rho_t}\rho_t = 0\,,$$

in the sense of distributions on $C_c^\infty([t_1, t_2] \times \mathbb{R}^d)$. Using the continuity equation along with the molification technique used in the proof of Prop. 7, Eq. (25) follows. Now, Eq. (25) proves, as a byproduct, that $\int_{t_1}^{t_2} \int \|v_{\rho_t}\|^2 d\rho_t dt < \infty$, which implies Eq. (24). The proof is concluded.

An alternative proof consists in using the concept of gradient flows on $\mathcal{P}_2(\mathbb{R}^d)$. Consider the functional $\mathcal{F}_\varpi(\mu)$ which coincides with $\mathcal{R}_\sigma(\mu)$ if $\tilde{\pi}_\#\mu = \varpi$, and $\mathcal{F}_\varpi(\mu) = +\infty$ otherwise. By [AGS08, Th.11.2.1], there exists a locally Lipschitz curve, say (μ_t), defined on any interval of the form $[0, T]$, such that $\mu_t \to \rho_0$ as $t \to 0$, and whose velocity field (v_t) satisfies $-v_t \in \partial\mathcal{F}_\varpi(\mu_t)$. It holds that $\int_0^T \|v_t\|^2 d\rho_t dt < \infty$ by definition of the velocity. Also using the same result [AGS08, Th.11.2.1], $\mathcal{R}_\sigma(\mu_{t_2}) - \mathcal{R}_\sigma(\mu_{t_1}) \leq -\int_{t_1}^{t_2} \int \|v_t\|^2 d\mu_t dt$ for all $t_2 > t_1 \geq 0$, since $\mathcal{F}_\varpi = \mathcal{R}_\sigma$ along the curve (μ_t). Using [AGS08, Th. 10.4.6] and the same derivations as in the proof of Prop. 4, we establish that $v_t = (\tilde{v}_{\mu_t}, 0)$. Moreover, by [AGS08, Th. 8.2.1], there exists a measure on $\boldsymbol{\mu} \in \mathcal{P}(C([0, T])$ such that $\boldsymbol{\mu}_t = \mu_t$ for all t, and such that $\dot{x}_t = v_{\mu_t}(x_t)$ for $\boldsymbol{\mu}$-almost every x, and almost every t. Thus, on $[0, T]$, $\boldsymbol{\mu}$ satisfies the martingale problem given in Def. 1. As, by [CD22, Prop. 1], this problem has a unique solution, we obtain that $\boldsymbol{\mu}$ coincides with $(\pi_{[0,T]})_\#\rho$. In particular, $\rho_t = \mu_t$.

A Technical proofs

A.1 Proof of Proposition 1

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

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

for all $k \in \mathbb{N}^*$. Let $\mathcal{K} := \bigcap_{k\geq 1} \pi_{[0,k]}^{-1}(K_k) \subset \mathcal{C}$. The union bound yields $\rho_n(\mathcal{K}) > 1 - \varepsilon$. Referring to [Bou89, Theorem 2, Section X, Chapter 5], $\mathcal{K}$ has a compact closure in $\mathcal{C}$. Hence, there exists a converging subsequence (ρ_{φ_n}) converging to $\rho \in \mathcal{P}(\mathcal{C})$. Following the proof of [Vil09, Theorem

6.18], one can readily check that $\lim_{n\to\infty} W_p((\pi_{[0,k]})_{\#}\rho_n, (\pi_{[0,k]})_{\#}\rho) = 0$, for every k. Consequently, $\lim_{n\to\infty} \mathsf{W}_p(\rho_n, \rho) = 0$, which means the completeness of $\mathcal{P}_p(\mathcal{C})$. It remains to obtain its separability.

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

A.2 Proof of Lemma 2

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

$$\mathbb{I}(\rho_n)(K_k) > 1 - \frac{k}{2^k}\,,$$

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

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

The union bound and Markov's inequality yields:

$$\mathbb{P}\left(\rho_n \in \mathcal{K}_\varepsilon\right) > 1 - \varepsilon \tag{65}$$

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

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

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

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

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

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

$$\mathbb{P}\left(\rho_n \in \mathcal{U}_\varepsilon\right) > 1 - \varepsilon\,. \tag{66}$$

Putting together Eq. (65) and Eq. (66),

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

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

A.3 Proof of Lemma 3

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

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

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

$$g(x,y) := \left(\phi(x_t) - \phi(x_s) - \int_s^t \left(\langle \nabla\phi(x_u), b(x_u,y_u)\rangle + \sigma^2 \Delta\phi(x_u)\right) \mathrm{d}u\right) h(x)\,,$$

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

$$|g(x,y)| \leq C\left(1 + \int_s^t \|b(x_u, y_u)\|\, du\right), \tag{67}$$

where $C = \|h\|_\infty \max\left(2\|\phi\|_\infty + \sigma^2(t-s)\|\Delta\phi\|_\infty, \|\nabla\phi\|_\infty\right)$. Note that $\rho_n \otimes \rho_n \to \rho_\infty \otimes \rho_\infty$ in $\mathcal{P}(\mathcal{C} \times \mathcal{C})$. Furthermore, using the bound (67) for our function g, and observing that t is the maximum of the time snapshots intervening in the definition of g, we have for each $a > 0$

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

therefore,

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

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

Acknowledgment. This work was funded by Chair DSAIDIS of Télecom Paris.

References

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[BDG72] D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 223–240. Univ. California Press Berkeley, Calif., 1972.

[Ben99] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.

[BGG13] F. Bolley, I. Gentil, and A. Guillin. Uniform convergence to equilibrium for granular media. *Archive for Rational Mechanics and Analysis*, 208:429–445, 2013.

[BGM10] F. Bolley, A. Guillin, and F. Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent vlasov-fokker-planck equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(5):867–884, 2010.

[Bil99] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[Bou89] N. Bourbaki. *General topology. Chapters 5–10*. Elements of Mathematics (Berlin). Springer-Verlag, Berlin, 1989. Translated from the French, Reprint of the 1966 edition.

[BRTV98] S. Benachour, B. Roynette, D. Talay, and P. Vallois. Nonlinear self-stabilizing processes–i existence, invariant probability, propagation of chaos. *Stochastic processes and their applications*, 75(2):173–201, 1998.

[BS00] M. Benaïm and S. J. Schreiber. Ergodic properties of weak asymptotic pseudotrajectories for semiflows. *Journal of Dynamics and Differential Equations*, 12:579–598, 2000.

[CB18] L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. *Advances in neural information processing systems*, 31, 2018.

[CD22] L.-P. Chaintron and A. Diez. Propagation of chaos: A review of models, methods and applications. i. models and methods. *Kinetic and Related Models*, 15(6):895, 2022.

[CGM08] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability theory and related fields*, 140:19–40, 2008.

[CMV03] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana*, 19(3):971–1018, 2003.

[CMV06] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, 179:217–263, 2006.

[Cor23] Q. Cormier. On the stability of the invariant probability measures of mckean-vlasov equations. *arXiv preprint arXiv:2201.11612*, 2023.

[CRR18] L. Carratino, A. Rudi, and L. Rosasco. Learning with sgd and random features. *Advances in neural information processing systems*, 31, 2018.

[DEGZ20] A. Durmus, A. Eberle, A. Guillin, and R. Zimmer. An elementary approach to uniform in time propagation of chaos. *Proceedings of the American Mathematical Society*, 148(12):5387–5398, 2020.

[DMT19] P. Del Moral and J. Tugaut. Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions. *Stochastic Analysis and Applications*, 37(6):909–935, 2019.

[DS10] S. Daneri and G. Savaré. Lecture notes on gradient flows and optimal transport. *arXiv preprint arXiv:1009.3737*, 2010.

[ELL21] X. Erny, E. Löcherbach, and D. Loukianova. Conditional propagation of chaos for mean field systems of interacting neurons. *Electronic Journal of Probability*, 26(none):1 – 25, 2021.

[GLWZ22] A. Guillin, W. Liu, L. Wu, and C. Zhang. Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems. *The Annals of Applied Probability*, 32(3):1590 – 1614, 2022.

[Leo23] G. Leoni. *A first course in fractional Sobolev spaces*, volume 229. American Mathematical Society, 2023.

[LW16] Q. Liu and D. Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in Neural Information Processing Systems*, 2016.

[Mal01] F. Malrieu. Logarithmic sobolev inequalities for some nonlinear pde's. *Stochastic processes and their applications*, 95(1):109–132, 2001.

[Mal03] F. Malrieu. Convergence to equilibrium for granular media equations and their euler schemes. *The Annals of Applied Probability*, 13(2):540–560, 2003.

[Mañ87] R. Mañé. *Ergodic Theory and Differentiable Dynamics*. Ergebnisse der Mathematik und ihrer Grenzgebiete : a series of modern surveys in mathematics. Folge 3. Springer-Verlag, 1987.

[MMN18] S. Mei, A. Montanari, and P.-M. Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33):E7665–E7671, 2018.

[MPZ21] S. Menozzi, A. Pesce, and X. Zhang. Density and gradient estimates for non degenerate brownian sdes with unbounded measurable drift. *Journal of Differential Equations*, 272:330–369, 2021.

[MRC87] S. Méléard and S. Roelly-Coppoletta. A propagation of chaos result for a system of particles with moderate interaction. *Stochastic Processes and their Applications*, 26:317–332, 1987.

[Oel84] K. Oelschlager. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Annals of Probability*, 12:458–479, 1984.

[RR07] A. Rahimi and B. Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.

[RVE22] G. Rotskoff and E. Vanden-Eijnden. Trainability and accuracy of artificial neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, 75(9):1889–1935, 2022.

[SS20] J. Sirignano and K. Spiliopoulos. Mean field analysis of neural networks: A central limit theorem. *Stochastic Processes and their Applications*, 130(3):1820–1852, 2020.

[SSR22] A. Salim, L. Sun, and P. Richtarik. A convergence theory for svgd in the population limit under talagrand's inequality t1. In *International Conference on Machine Learning*, pages 19139–19152. PMLR, 2022.

[Szn84] A.-S. Sznitman. Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *Journal of functional analysis*, 56(3):311–336, 1984.

[Tug13] J. Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *The Annals of Probability*, 41(3A):1427 – 1460, 2013.

[Ver06] A. Y. Veretennikov. On ergodic measures for mckean-vlasov stochastic equations. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 471–486. Springer, 2006.

[Vil06] C. Villani. Mathematics of granular materials. *Journal of statistical physics*, 124(2-4):781–822, 2006.

[Vil09] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.