AN INVARIANCE PRINCIPLE FOR GRADIENT FLOWS IN THE SPACE OF PROBABILITY MEASURES

JOSÉ A. CARRILLO†, RISHABH S. GVALANI‡, & JEREMY WU†

ABSTRACT. We seek to establish qualitative convergence results to a general class of evolution PDEs described by gradient flows in optimal transportation distances. These qualitative convergence results come from dynamical systems under the general name of LaSalle Invariance Principle. By combining some of the basic notions of gradient flow theory and dynamical systems, we are able to reproduce this invariance principle in the setting of evolution PDEs under general assumptions. We apply this abstract theory to a non-exhaustive list of examples that recover, simplify, and even extend the results in their respective literatures.

1. INTRODUCTION

Our goal is to show that gradient flows in metric spaces can be cast into the framework of dynamical systems to apply generic qualitative long-time convergence results. In particular, we seek to reformulate the LaSalle invariance principle in the setting of evolution PDEs. Let us provide a formal and instructive review of this concept in finite dimensional ODEs, we refer to [CH98, Chapter 9] for a general semigroup description. For $x_0 \in \mathbb{R}^d$, consider a curve $x : [0, \infty) \to \mathbb{R}^d$ solving

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = F(x(t)), \quad x(0) = x_0,$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a sufficiently smooth driving vector field. The *LaSalle invariance principle* in this case reads

Theorem 1.1 (LaSalle invariance principle: ODE version). *Suppose $\Phi : \mathbb{R}^d \to \mathbb{R}$ is continuous and satisfies the strict Lyapunov condition; $t \mapsto \Phi(x(t))$ is non-increasing and*

$$\forall t \geq 0, \, \Phi(x(t)) = \Phi(x_0) \implies \forall t \geq 0, \, x(t) = x_0.$$

Assume further that the set $\bigcup_{t \geq 0}\{x(t)\}$ is relatively compact. Then, the distance between $x(t)$ and the set of equilibrium points $F^{-1}(\{0\})$ converges to 0 as $t \to \infty$.

If we specialize to *gradient flows*, the situation simplifies significantly. If $E : \mathbb{R}^d \to \mathbb{R}$ is some (smooth) function, let us now consider

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = -\nabla E(x(t)), \quad x(0) = x_0.$$

In this setting, E is immediately a strict Lyapunov function since

$$\frac{\mathrm{d}}{\mathrm{d}t}E(x(t)) = -|\nabla E(x(t))|^2 \leq 0.$$

Here, it suffices to check the compactness of trajectories $x(t)$ to apply the invariance principle. The compactness property can also be pivoted onto E since trajectories are contained in sublevel sets of E; it suffices to check that sublevel sets of E are relatively compact. We seek to extend these advantages in finite dimensional gradient flows to infinite dimensional gradient flows commonly describing evolution PDEs in biology, physics, social networks, and many other applications. The function E describes quite a lot of the dynamics of the ODEs and this is also true for PDEs which are the infinite dimensional analogues.

(†) MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, OXFORD OX2 6GG, UK
(‡) MAX-PLANCK-INSTITUT FÜR MATHEMATIK IN DEN NATURWISSENSCHAFTEN, 04103 LEIPZIG, GERMANY

Let us formally introduce the basic notion of gradient flows in the space of probability measures by describing a typical example of such an equation. We refer the reader to Section 2 or [AGS08] for more details and greater generality. For $\Omega = \mathbb{R}^d$ or $\mathbb{T}^d$, and functions $F : \mathbb{R}_{\geq 0} \to \mathbb{R}$; $V, W : \Omega \to \mathbb{R}$ we consider the functional on probability densities with finite second moment, denoted by $\mathcal{P}_2(\Omega)$,

$$E(\rho) = \int_\Omega F(\rho)\,\mathrm{d}x + \int_\Omega V(x)\,\mathrm{d}\rho(x) + \frac{1}{2}\int_\Omega (W * \rho)(x)\,\mathrm{d}\rho(x).$$

We refer to the following PDE or its solution as Wasserstein gradient flow of E

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla (F' + V + W * \rho) \right).$$

In fact, following Otto's formal Riemannian calculus [Ott01], this PDE can be interpreted closer to the ODE gradient flow by

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = -\nabla_{W_2} E(\rho),$$

where W_2 is the 2-Wasserstein distance on $\mathcal{P}_2(\Omega)$. Here ∇_{W_2} denotes the gradient operator

$$\nabla_{W_2} := \nabla \cdot \left(\rho \nabla \frac{\delta}{\delta\rho} \right),$$

on $(\mathcal{P}_2(\Omega), W_2(\cdot, \cdot))$, which can be thought of as an infinite-dimensional Riemannian manifold (cf., for example, [Lot08, Gig12]). By a formal computation, we see that

$$\frac{\mathrm{d}}{\mathrm{d}t} E(\rho(t)) = - \int_\Omega |\nabla(F' + V + W * \rho)|^2 \, \mathrm{d}\rho(x) \leq 0,$$

which is in direct analogy with the ODE setting, i.e. E satisfies the strict Lyapunov condition. This property has been the basis of a very active field of research in the last couple of decades. Our focus is to exploit this property at the abstract level of dynamical systems described above. Several difficulties arise before we can formulate a a version of the LaSalle invariance principle in this context. Firstly, the regularity of $E : \rho \mapsto E(\rho)$ is usually limited to lower semicontinuity in the most interesting cases. Secondly, the compactness of trajectories is usually established in a *weaker* topology than the W_2 topology. We now list a few prominent examples of Wasserstein gradient flows, some of which we consider in Section 4, are given by

(1) Fokker-Planck equations (cf. [JKO98]): $F(x) = x \log x$, $V \in C^\infty(\Omega)$, $W = 0$. Assuming V satisfies certain growth assumptions, it is known that solutions converge exponentially in the 2-Wasserstein distance towards the unique stationary solution given by

$$\rho_s(x) = \frac{e^{-V(x)}}{\int_\Omega e^{-V(y)}\,\mathrm{d}y}.$$

(2) Porous medium (and fast diffusion) equations (cf. [Ott01, Vó7]):

$$F(x) = \begin{cases} x \log x & m = 1 \\ \dfrac{1}{m-1} x^m & m > 1 \end{cases}$$

and $V = W = 0$. It is known that self-similar solutions exist and give rise to finite or infinite time extinction depending on the value of m.

(3) Aggregation equations (cf. [CDF$^+$11]): $F = V = 0$ with W typically chosen to be radial. These equations can describe a variety of interaction-based phenomena from biology, physics, and social networks. One of the main topics of interest is what conditions guarantee consensus concentration or segregation.

(4) Parabolic-elliptic Keller-Segel models (cf. [BCC08, BCL09b]): $F = x \log x, V = 0$, and W is the negative Green's function in $\mathbb{R}^d$. Understanding how certain infinite-time profiles are generated is a very active area of research. The long-time behaviour of solutions strongly depends on the initial mass.

(5) Machine Learning (c.f. [Woj20, WE20] and the references therein): An important step in the implementation of multi-layer deep neural networks is to optimise the network parameters (with respect to some appropriately chosen cost function) by evolving them using a gradient descent scheme. The mean field equation associated to this system corresponds to a gradient flow in the 2-Wasserstein metric. The long-time asymptotics of the density

of the parameters then describes, in some sense, the optimal choice of parameters in the training.

We postpone further discussion on the literature of some of these particular examples and others to the relevant subsections below (cf. Section 4).

This paper seeks to unify the convergence results of many different PDEs under one general theory. However, one of the limitations of this soft approach is the lack of quantitative rates of convergence. We note that deriving explicit rates for convergence usually relies on á la carte techniques that are problem-dependent. Furthermore, these techniques only provide convergence of solutions towards the set of equilibrium points. In specific cases, one can extract more information about the structure of this set and thus obtain a stronger convergence results e.g. to a particular stationary solution. In the situation where there are multiple equilibrium points (cf. [CGPS20, CG19]), we can only hope to obtain local rates of convergence in the possible basins of attraction for stable stationary solutions. Our result gives the first step in this direction; for instance, in the presence of phase transitions, depending on the energy of the initial data, we can identify the limiting stationary solutions (cf. Section 4.2 and Section 4.3).

The plan of the paper is as follows: In Section 2, we develop the notation and assumptions necessary to state the main results of Theorems 2.12 to 2.14. Section 3 is dedicated to the proofs of these results. Section 4 is dedicated to examples of gradient flows in which our abstract theory applies as well as a small overview of the existing literature in each example. Appendix A contains a brief review of the abstract formalism of dynamical systems in metric spaces with a mild adaptation to allow for non-unique solutions. Appendix B contains a technical lemma allowing us to easily check lower semicontinuity of certain functionals, which we repeatedly use in our examples.

2. Preliminaries and the main abstract result

We denote by $\mathcal{P}(X)$ the space of all Borel probability measures on some Polish metric space (X, d). We equip this with the so-called narrow or weak topology, i.e. the coarsest topology on $\mathcal{P}(X)$ such that all functions of the form $\rho \mapsto \int_X \varphi \, d\rho$, $\varphi \in C_b(X)$ are continuous, where $C_b(X)$ is the space of bounded, continuous functions on X. This topology is metrisable by, for instance, the Lévy–Prokhorov metric, which we denote by $d_{LP}(\cdot, \cdot)$. Furthermore, we define $\mathcal{P}_p(X)$ to be the subset of $\mathcal{P}(X)$ consisting of all probability measures ρ with finite p-moment. We equip this space with the topology generated by the so-called p-Wasserstein or transportation cost distance, $W_p(\cdot, \cdot)$. Note that the space $\mathcal{P}_p(X)$ also inherits a topology from $\mathcal{P}(X)$ as a subspace of $\mathcal{P}(X)$. This topology, i.e. the narrow topology on $\mathcal{P}_p(X)$, is coarser than the W_p-topology. We will denote this topology by σ. Note that this topology is metrisable (for example by the restriction of d_{LP} to $\mathcal{P}_p(X)$) and we denote the associated metric by $d_\sigma(\cdot, \cdot)$. Given a sequence of probability measures $(\rho_n)_{n \in \mathbb{N}}$ and a topology τ, we denote convergence to a point $\rho \in \mathcal{P}(X)$ in this topology by $\rho_n \xrightarrow{\tau} \rho$ as $n \to \infty$. For weak convergence in $\mathcal{P}(X)$, we will often just write $\rho_n \to \rho$ as $n \to \infty$.

We are interested in studying the long time behaviour of W_p gradient flows in $\mathcal{P}_p(X)$ ($1 < p < \infty$). We start by introducing some preliminary notions from the theory of gradient flows that will play a role in our subsequent analysis. We will use a weaker notion of solution referred to as a *curve of maximal slope* rather than the stronger *gradient flow* notion of solution based on the so-called *Evolutionary Variational Inequality (EVI)*. The advantage of using this notion is that it is softer and allows for non-uniqueness of solutions. For the rest of the paper, even though we will use the two terms interchangeably, we remind the reader that we will only be working with the curve of maximal slope notion of solution.

Definition 2.1. (Absolutely continuous curves, [AGS08, Definition 1.1.1]) Fix $T > 0$ and $1 < p < \infty$. We say that that $\rho \in C([0, T]; \mathcal{P}_p(X))$ is *absolutely continuous*, denoted by $\rho \in AC([0, T]; \mathcal{P}_p(X))$, if there exists an $m \in L^1([0, T])$, such that

$$W_p(\rho(t), \rho(s)) \leq \int_s^t m(r) \, dr \,, \tag{2.1}$$

for all $0 \leq s < t \leq T$. We say that $\rho \in AC([0, \infty); \mathcal{P}_p(X))$ if $\rho \in AC([0, T]; \mathcal{P}_p(X))$ for all $T > 0$.

Definition 2.2. (Metric derivative, [AGS08, Definition 1.1.2]) Fix $T > 0$ and $1 < p < \infty$ and consider a curve $\rho \in AC([0,T]; \mathcal{P}_p(X))$. Then, the limit

$$|\rho'|(t) := \lim_{s \to t} \frac{W_p(\rho(t), \rho(s))}{|t - s|},$$

exists t a.e. The function $t \mapsto |\rho'|(t)$ is in $L^1([0,T])$ and is referred to as the *metric derivative* of the curve ρ. Furthermore, $|\rho'|$ is the minimal admissible function for the inequality in (2.1).

Consider a proper function $E : \mathcal{P}_p(X) \to (-\infty, +\infty]$. This function should be thought of as the free energy driving our system. We now present the final ingredient needed to define the notion of solution we will use in this paper, i.e. the weak upper gradient.

Definition 2.3. (Weak upper gradient, [AGS08, Definition 1.2.2]) A function $G : \mathcal{P}_p(X) \to [0, +\infty]$ is a *weak upper gradient* for the energy E if for every curve $\rho \in AC([0,T], \mathcal{P}_p(X))$ such that

  (i) $G \circ \rho |\rho'| \in L^1([0,T])$
  (ii) $E \circ \rho$ is a.e. equal to a function φ with finite pointwise variation in $[0,T]$ and there holds

$$|\varphi'(t)| \leq G(\rho(t))|\rho'|(t), \quad \text{a.e. } t \in [0,T].$$

We introduce the definition of a p-curve of maximal slope:

Definition 2.4. (Curve of maximal slope, [AGS08, Definition 1.3.2]) A curve $\rho \in AC([0,\infty); \mathcal{P}_p(X))$ is said to be a *p-curve of maximal slope* for the functional E with respect to its weak upper gradient G if $E \circ \rho$ is a.e. equal to a non-increasing map $\varphi_t = E(\rho(t))$ and

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t = -\frac{1}{p}|\rho'|^p(t) - \frac{1}{q}G^q(\rho(t)) \quad \text{a.e. } t \in [0,\infty), \tag{MS}$$

where $q = p/(p-1)$ and $|\rho'|$ is the metric derivative of ρ in the sense of Definition 2.2. We will often abbreviate 'p-curve of maximal slope for the functional E with respect to its upper gradient G' by *curve of maximal slope*. Furthermore, we will call $\rho \in AC([0,\infty); \mathcal{P}_p(X))$ a *p-curve of maximal slope with initial measure/datum* $\rho_0 \in \mathcal{P}_p(X), E(\rho_0) < \infty$ if ρ is a p-curve of maximal slope and $\rho(0) = \rho_0$.

Remark 2.5. The fact that curves of maximal slope exist under rather mild assumptions on E and G follows from [AGS08, Theorems 2.3.1 and 2.3.3]. Furthermore, it is known that under the additional assumption of λ-convexity (in the sense of McCann [McC94]) on E and for $p = 2$ for every $\rho_0 \in \mathcal{P}_2(X)$, there exists a unique (up to a.e. equality) curve of maximal slope ρ with $\rho(0) = \rho_0$ (c.f. [AGS08, Theorem 11.1.4]). As mentioned earlier, we will not work with guaranteed uniqueness, instead we choose to fix some curve of maximal slope for which existence is guaranteed; our example in Section 4.8 illustrates how our convergence results work even in the presence of non-uniqueness.

Assumption A. We will impose the following set of assumptions on our energy, E, and its weak upper gradient, G:

  (A1) $E : \mathcal{P}_p(X) \to (-\infty, +\infty]$ is a proper, bounded below, and lower semicontinuous (l.s.c.) functional with respect to the σ-topology.
  (A2) The sublevel set $L_{\leq C}(E) := \{\rho \in \mathcal{P}_p(X) : E(\rho) \leq C\}$ is relatively compact in the σ-topology for all $C \in \mathbb{R}$.
  (A3) $G : \mathcal{P}_p(X) \to [0, \infty]$ is l.s.c. with respect to the σ-topology.

Assumption B. In certain cases, the energy E and its weak upper gradient, G, may have extensions $\tilde{E}$ and $\tilde{G}$ defined on $\mathcal{P}(X)$ such that $\tilde{E}(\rho) = E(\rho)$ and $\tilde{G}(\rho) = G(\rho)$ for all $\rho \in \mathcal{P}_p(X)$. In this setting, we make the following set of assumptions:

  (B1) $\tilde{E} : \mathcal{P}(X) \to (-\infty, +\infty]$ is a proper, bounded below, and l.s.c. functional.
  (B2) The sublevel set $L_{\leq C}(\tilde{E}) := \{\rho \in \mathcal{P}(X) : \tilde{E}(\rho) \leq C\}$ is relatively compact in $\mathcal{P}(X)$.
  (B3) $\tilde{G} : \mathcal{P}(X) \to [0, \infty]$ is l.s.c. in $\mathcal{P}(X)$.

Note that (B1) implies (A1) and (B3) implies (A3). Furthermore, (A2) implies (B2). The set of assumptions (B1)–(B3) handles precisely the case where gradient flows exist in $\mathcal{P}_p(X)$ for every $t > 0$, while allowing for convergence towards steady states in $\mathcal{P}(X)$ (but possibly not in $\mathcal{P}_p(X)$). We refer to the example in Section 4.8 which illustrates this.

Assumption C. We may sometimes replace assumptions (A2) and (B2) by the following assumptions:

(A2') The sublevel set $L_{\leq C}(E) := \{\rho \in \mathcal{P}_p(X) \cap \mathcal{Y} : E(\rho) \leq C\}$ is relatively compact in the σ-topology on $\mathcal{P}_p(X)$.

(B2') The sublevel set $L_{\leq C}(E) := \{\rho \in \mathcal{P}_p(X) \cap \mathcal{Y} : E(\rho) \leq C\}$ is relatively compact in $\mathcal{P}(X)$.

Here, $\mathcal{Y}$ is a subset of $\mathcal{P}_p(X)$ (resp. $\mathcal{Y} \subset \mathcal{P}(X)$) such that $\bigcup_{t \geq 0}\{\rho(t)\} \subseteq \mathcal{Y}$, where $\rho(t)$ is a p-curve of maximal slope of E with respect to G for some initial measure ρ_0. Here, $\mathcal{Y}$ could be a subset which is invariant under the gradient flow (eg. radial measures) or could be the curve of maximal slope itself.

We wish to better understand stationary solutions and convergence towards these states under a dynamical systems viewpoint. One of the main points to clarify is the relation between the set of stationary solutions and the ω-limit set of (MS) when treated as an abstract dynamical system for a given initial condition. To be precise, we define these notions in our present context. We have adapted some classical definitions from dynamical systems (which can be found in the Appendix A) for our present context to allow for non-unique trajectories. We start with the following result:

Proposition 2.6. *Assume that for all $\rho_0 \in \mathcal{P}_p(X), E(\rho_0) < \infty$ there exists a p-curve of maximal slope with initial measure ρ_0. We define the set*

$$Z_{E,p} := \{\rho \in \mathcal{P}_p(X) : E(\rho) < \infty\},$$

which we equip with the σ-topology.

Then, the family of mappings $\{S_t\}_{t \geq 0}$, $S_t : Z_{E,p} \to 2^{Z_{E,p}}$ sending $\rho_0 \mapsto \bigcup_{j \in J} \rho_j(t)$, where J indexes the set of all possible curves of maximal slope with initial measure ρ_0, defines a metric dynamical system in the sense of Definition A.1.

Proof. We need to check that the conditions of Definition A.1 are satisfied for curves of maximal slope. The condition (1) is trivially satisfied because curves of maximal slope decrease the energy E. For condition (2), one can check that S_0 maps any initial datum ρ_0 to the initial point of the curve of maximal slope $\rho(0) = \rho_0$. Thus, it coincides with the identity, I. Condition (3) is satisfied because curves of maximal slope live in $AC([0, \infty); \mathcal{P}_p(X))$. $\qquad\square$

We now define what we mean by a stationary state of a curve of maximal slope.

Definition 2.7. We say that $\mu \in \mathcal{P}_p(X)$ is a *stationary state* of (MS) if $E(\mu) < \infty$ and the curve defined for all $t \geq 0$ by $\rho(t) := \mu$ is a curve of maximal slope according to Definition 2.4. We denote the set of stationary states by $\mathcal{E}$.

Note that Proposition 2.6 tells us that curves of maximal slope generate a metric dynamical system on the space $Z_{E,p}$ equipped with the σ-topology. Thus, the definition of the corresponding ω-limit set should be with respect to this topology. For the convenience of the reader, we recast Definition A.2 into our particular setting.

Definition 2.8. For $\rho_0 \in Z_{E,p}$ and $\rho \in AC([0, \infty); \mathcal{P}_p(X))$ a curve of maximal slope with $\rho(0) = \rho_0$, we denote by $\omega^\rho(\rho_0)$ the *ω-limit set* with initial data ρ_0 subordinate to ρ, defined by

$$\omega^\rho(\rho_0) := \left\{ \bar{\rho} \in \mathcal{P}_p(X) : \exists t_n \to \infty \text{ s.t. } \rho(t_n) \xrightarrow{\sigma} \bar{\rho} \text{ as } n \to \infty \right\}.$$

We now give a characterisation of stationary states in relation to the weak upper gradient G which we will use extensively in the sequel.

Lemma 2.9. *$\rho \in \mathcal{P}_p(X)$ is a stationary state of (MS) if and only if $G(\rho) = 0$ and $E(\rho) < +\infty$.*

Proof. ($\implies$) By Definition 2.7, the curve $\rho(t) = \rho$ solves (MS). Since $\rho(t)$ is constant for all times t, the first two terms of (MS) are a.e. zero which implies $G(\rho) = G(\rho(t)) = 0$.
($\impliedby$) By defining the curve $\rho(t) = \rho$, the first two terms of (MS) are always zero. Since $G(\rho(t)) = G(\rho) = 0$, equation (MS) holds automatically. $\qquad\square$

The above lemma leads to the following natural relaxation of the notion of stationary state.

Definition 2.10. Assume the energy E and the weak upper gradient G have extensions to $\mathcal{P}(X)$, i.e. there exist functions $\tilde{E} : \mathcal{P}(X) \to (-\infty, +\infty]$ and $\tilde{G} : \mathcal{P}(X) \to [0, +\infty]$, such that $\tilde{E}(\rho) = E(\rho)$ and $\tilde{G}(\rho) = G(\rho)$ for all $\rho \in \mathcal{P}(X)$. Then, we say that a measure $\mu \in \mathcal{P}_p(X)$ is a *weak stationary state* of (MS) if $\tilde{E}(\mu) < +\infty$ and $\tilde{G}(\mu) = 0$. We denote the set of weak stationary states by $\mathcal{E}_w$.

The motivation for the above definition follows from the fact that for certain curves of maximal slope the long time limits may lie outside the set $\mathcal{P}_p(X)$. This is especially true for curves of maximal slope associated to drift-diffusion type Fokker–Planck equations when the confining potential V does not grow fast enough at $+\infty$. The above definition then allows us to characterise the stationary states of these PDEs as zeroes of the associated extended weak upper gradient $\tilde{G}$ and obtain useful information about the long time behaviour of these curves of maximal slope even if the stationary states lie outside $\mathcal{P}_p(X)$. We will discuss a specific instance of this in Section 4.1. Accordingly, we give a relaxed definition of ω-limit sets.

Definition 2.11. For $\rho_0 \in Z_{E,p}$ and $\rho \in AC([0,\infty); \mathcal{P}_p(X))$ a curve a maximal slope with $\rho(0) = \rho_0$, we denote by $\omega'^\rho(\rho_0)$ the *relaxed ω-limit set* with initial data ρ_0 subordinate to ρ, defined by

$$\omega'^\rho(\rho_0) := \{\bar{\rho} \in \mathcal{P}(X) : \exists t_n \to \infty \text{ s.t. } \rho(t_n) \to \bar{\rho} \text{ in } \mathcal{P}(X) \text{ as } n \to \infty\}.$$

With these notations, our first main result is the relation between $\omega^\rho(\rho_0)$ (resp. $\omega'^\rho(\rho_0)$) and $\mathcal{E}$ (resp. $\mathcal{E}_w$).

Theorem 2.12. *Fix $\rho_0 \in Z_{E,p}$ and a curve of maximal slope $\rho \in AC([0,\infty); \mathcal{P}_p(X))$ with initial condition $\rho(0) = \rho_0$. Under the set of assumptions* (A1), (A2) *(or* (A2')*),* (A3)*, we have*

$$\omega^\rho(\rho_0) \subset \mathcal{E},$$

so that ω-limits are stationary. Under the set of assumptions (B1), (B2) *(or* (B2')*),* (B3)*, we have $\omega'^\rho(\rho_0) \subset \mathcal{E}_w$, so that relaxed ω-limits are weak stationary states.*

The set inclusions from Theorem 2.12 are essential to the following convergence results.

Theorem 2.13. *Assume the energy E and weak upper gradient G satisfy assumptions* (A1), (A2) *(or* (A2')*), and* (A3)*. Furthermore, assume that for any $\rho_0 \in Z_{E,p}$ there exists a curve of maximal slope, $\rho \in AC([0,\infty); \mathcal{P}_p(X))$ of E with respect to G. Then, it follows that*

$$\lim_{t \to \infty} d_\sigma(\rho(t), \mathcal{E}) = \lim_{t \to \infty} \inf_{\mu \in \mathcal{E}} d_\sigma(\rho(t), \mu) = 0.$$

Furthermore if (A2') *is satisfied and $\mathcal{Y}$ is closed in the σ-topology on $\mathcal{P}_p(X)$, then*

$$\lim_{t \to \infty} d_\sigma(\rho(t), \mathcal{E} \cap \mathcal{Y}) = 0.$$

Theorem 2.14. *Assume the energy E and the weak upper gradient G have extensions to $\mathcal{P}(X)$ such that $\tilde{E}$ and weak upper gradient $\tilde{G}$ satisfy assumptions* (B1), (B2) *(or* (B2')*), and* (B3)*. Furthermore, assume that for any $\rho_0 \in Z_{E,p}$ there exists a curve of maximal slope, $\rho \in AC([0,\infty); \mathcal{P}_p(X))$ of E with respect to G. Then, it follows that*

$$\lim_{t \to \infty} d_{LP}(\rho(t), \mathcal{E}_w) = \lim_{t \to \infty} \inf_{\mu \in \mathcal{E}_w} d_{LP}(\rho(t), \mu) = 0.$$

Furthermore if (B2') *is satisfied and $\mathcal{Y}$ is closed in $\mathcal{P}(X)$, then*

$$\lim_{t \to \infty} d_{LP}(\rho(t), \mathcal{E}_w \cap \mathcal{Y}) = 0.$$

We end this section with the following small lemma that can be used to improve the convergence in Theorems 2.13 and 2.14:

Lemma 2.15. *Let Y be a set equipped with two metrics d_1 and d_2 such that d_2 induces a finer topology on Y than d_1. Furthermore, let $y : [0,\infty) \to Y$ be a curve such that*

$$\lim_{t \to \infty} d_1(y(t), Y^*) = \lim_{t \to \infty} \inf_{y \in Y^*} d_1(y(t), Y^*) = 0, \tag{2.2}$$

for some $Y^* \subseteq Y$. Assume that $\bigcup_{t \geq 0} \{y(t)\}$ is compact with respect to the topology generated by d_2, then

$$\lim_{t \to \infty} d_2(y(t), Y^*) = 0.$$

Proof. Let us assume by contradiction that we can find a sequence $(y(t_n))_{n \in \mathbb{N}}$, $t_n \to \infty$, such that

$$\liminf_{n \to \infty} d_2(y(t_n), Y^*) \geq \varepsilon \,,$$

for some $\varepsilon > 0$. By the compactness of $\bigcup_{t \geq 0} \{y(t)\}$, it follows that there exists a subsequence $(y(t_{n_k}))_{k \in \mathbb{N}}$ and a $y^* \in Y$ such that

$$\lim_{k \to \infty} d_2(y(t_{n_k}), y^*) = 0 \,.$$

It follows that

$$\lim_{k \to \infty} d_1(y(t_{n_k}), y^*) = 0 \,.$$

This implies, from (2.2), that $y^* \in Y^*$. This is a contradiction and so the result of the lemma follows. $\qquad\square$

3. Proof of the main abstract result

We prove Theorem 2.12 after a few lemmata. Using this result, we then prove Theorems 2.13 and 2.14. Throughout this section, we use the same notation as in Theorem 2.12; we fix $\rho_0 \in Z_{E,p}$ and a curve of maximal slope $\rho \in AC([0, \infty); \mathcal{P}_p(X))$ with initial condition $\rho(0) = \rho_0$.

Lemma 3.1. *We take the set of assumptions* (A1), (A2) *(or* (A2')*), and* (A3). *We have that the trajectory $\bigcup_{t \geq 0} \{\rho(t)\}$ is relatively compact in $\mathcal{P}_p(X)$ with respect to the σ-topology.*

Proof. According to equation (MS), since the right-hand side is non-positive, we have that $E(\rho(t))$ is a non-increasing function of t. By $E(\rho_0) < \infty$, we have that $E(\rho(t)) \leq E(\rho_0) < \infty$ for all $t \geq 0$. Thus,

$$\bigcup_{t \geq 0} \{\rho(t)\} \subset \{\rho \in \mathcal{P}_p(X) : E(\rho) \leq E(\rho_0)\} \,,$$

or in the case of (A2'),

$$\bigcup_{t \geq 0} \{\rho(t)\} \subset \{\rho \in \mathcal{P}_p(X) \cap \mathcal{Y} : E(\rho) \leq E(\rho_0)\} \,.$$

It follows by (A2) or (A2'), that $\bigcup_{t \geq 0} \{\rho(t)\}$ is relatively compact in $\mathcal{P}_p(X)$ with respect to the σ-topology. $\qquad\square$

Lemma 3.1 allows us to extract a time-diverging sequence $t_n \to \infty$ and probability measure $\rho^* \in \mathcal{P}_p(X)$ such that $\rho(t_n) \xrightarrow{\sigma} \rho^*$ in $\mathcal{P}_p(X)$. So we have established the

Corollary 3.2. *Under the set of assumptions* (A1), (A2) *(or* (A2')*),* (A3)*; we have $\omega^\rho(\rho_0) \neq \emptyset$.*

Lemma 3.1 also has a generalisation to the relaxed ω-limit sets.

Lemma 3.3. *We take the set of assumptions* (B1), (B2) *(or* (B2')*), and* (B3). *We have that the trajectory $\bigcup_{t \geq 0} \{\rho(t)\}$ is relatively compact in $\mathcal{P}(X)$.*

Proof. As in the proof of Lemma 3.1, we can use equation (MS) to deduce the bound $E(\rho(t)) \leq E(\rho_0) < \infty$ for all $t \geq 0$.

$$\bigcup_{t \geq 0} \{\rho(t)\} \subset \{\rho \in \mathcal{P}_p(X) : E(\rho) \leq E(\rho_0)\} \,,$$

or in the case of (B2'),

$$\bigcup_{t \geq 0} \{\rho(t)\} \subset \{\rho \in \mathcal{P}_p(X) \cap \mathcal{Y} : E(\rho) \leq E(\rho_0)\} \,.$$

It follows by (B2) or (B2'), that $\bigcup_{t \geq 0} \{\rho(t)\}$ is relatively compact in $\mathcal{P}(X)$. $\qquad\square$

Corollary 3.4. *Under the set of assumptions* (B1), (B2) *(or* (B2')*),* (B3)*; we have $\omega'^\rho(\rho_0) \neq \emptyset$.*

With these basic properties of $\omega^\rho(\rho_0)$ and $\omega'^\rho(\rho_0)$, we now turn to the proof of Theorem 2.12.

Proof of Theorem 2.12. We only prove the first containment $\omega^\rho(\rho_0) \subset \mathcal{E}$ under the set of assumptions (A1), (A2) (or (A2')), (A3). Fix $\rho^* \in \omega(\rho_0) := \omega^\rho(\rho_0)$ obtained as a limit from some time-diverging sequence $t_n \to \infty$ where $\rho(t_n) \xrightarrow{\sigma} \rho^*$. It suffices to show $G(\rho^*) = 0$ by Lemma 2.9. The strategy is to construct a limit curve $\rho^\infty : [0,1] \to \mathcal{P}_p(X)$ for (a further subsequence of) $\rho(t_n)$ coming from gradient flow theoretic compactness properties of equation (MS) and then to firstly deduce that $G(\rho^\infty(t)) = 0$ for all $t \in [0,1]$. By showing that $\rho^\infty(0) = \rho^*$, we conclude the proof.

We define $\rho^n \in AC([0,1]; \mathcal{P}_p(X))$ where $\rho^n(t) = \rho(t_n + t)$, $t \in [0,1]$. We will prove compactness of this sequence ρ^n based on a generalised Arzelà-Ascoli result. By a similar argument to Lemma 3.1, for every $t \in [0,1]$, $\overline{\{\rho^n(t)\}_{n\in\mathbb{N}}}^\sigma \subset \{\rho \in \mathcal{P}_p(X) : E(\rho) \le E(\rho_0)\}$ which is compact by assumption (A2) or (A2'). Fix $0 \le s < t \le 1$. By construction and definition of the metric derivative (cf. Definition 2.2), we have for every $n \in \mathbb{N}$

$$W_p^p(\rho^n(t), \rho^n(s)) = W_p^p(\rho(t_n + t), \rho(t_n + s)) \le \left(\int_{t_n+s}^{t_n+t} |\rho'|(u)\frac{\mathrm{d}u}{t-s} \right)^p (t-s)^p .$$

Applying Jensen's inequality, we estimate further

$$W_p^p(\rho^n(t), \rho^n(s)) \le \left(\int_{t_n+s}^{t_n+t} |\rho'|^p(u)\,\mathrm{d}u \right) |t-s|^{p-1} \le C|t-s|^{p-1} ,$$

where C can be taken to be $\int_0^\infty |\rho'|^p(u)\,\mathrm{d}u < \infty$ due to the absolute continuity and maximal slope property of the curve and we stress that this constant is independent of n, s, t. Finally, we take the p-th root of the above inequality to obtain

$$W_p(\rho^n(t), \rho^n(s)) \le C|t-s|^{\frac{p-1}{p}} , \tag{†}$$

where we have abused notation by recycling C for the constant. Incidentally, estimate (†) is precisely the reason we have excluded the values $p = 1, \infty$. Applying [AGS08, Proposition 3.3.1], we can find a further subsequence which we omit and a limit curve $\rho^\infty \in C([0,1]; \mathcal{P}_p(X))$ such that pointwise $t \in [0,1]$

$$\rho(t_n + t) = \rho^n(t) \xrightarrow{\sigma} \rho^\infty(t) \text{ in } \mathcal{P}_p(X).$$

It follows then that $\rho^\infty(0) = \rho^*$. Additionally from (A3), it follows that

$$G^q(\rho^\infty(t)) \le \liminf_{n\to\infty} G^q(\rho^n(t)) .$$

By Fatou's lemma, we then have

$$\int_0^1 G^q(\rho^\infty(t))\,\mathrm{d}t \le \liminf_{n\to\infty} \int_0^1 G^q(\rho^n(t))\,\mathrm{d}t . \tag{G∞}$$

Returning to equation (MS), we integrate in time to obtain, for a.e. $t \in (0, \infty)$,

$$E(\rho(t)) - E(\rho(0)) = -\frac{1}{p} \int_0^t |\rho'|^p(s)\,\mathrm{d}s - \frac{1}{q} \int_0^t G^q(\rho(s))\,\mathrm{d}s.$$

Reversing the sign, applying the lower bound assumption (A1), and dropping the metric derivative term, we arrive at the bound

$$\int_0^t G^q(\rho(s))\,\mathrm{d}s \le C.$$

The universal constant C depends only on the lower bound of E, the initial value $E(\rho(0))$, and q. Therefore, sending $t \to \infty$, we have the integrability of $G^q(\rho(t))$

$$\int_0^\infty G^q(\rho(t))\,\mathrm{d}t < \infty. \tag{Gq}$$

From the above estimate (G∞) and the full integrability of $G^q(\rho(t))$ in (Gq), we have that

$$\int_0^1 G^q(\rho^\infty(t))\,\mathrm{d}t \le \liminf_{n\to\infty} \int_0^1 G^q(\rho^n(t))\,\mathrm{d}t = \liminf_{n\to\infty} \int_{t_n}^{t_n+1} G^q(\rho(t))\,\mathrm{d}t = 0 .$$

Since $G^q(\mu) \ge 0$ for all $\mu \in \mathcal{P}_p(X)$, it follows that $G^q(\rho^\infty(t)) = 0$ a.e. $t \in [0,1]$. Thus, there exists a sequence of times $t_m \to 0$ such that $G^q(\rho^\infty(t_m)) = 0$. Additionally, since $\rho^\infty \in C([0,1]; \mathcal{P}_p(X))$, $\rho^\infty(t_m) \xrightarrow{\sigma} \rho^*$ in $\mathcal{P}_p(X)$ as $t_m \to 0$. Applying (A3) again, we obtain

$$G^q(\rho^*) \le \liminf_{t_m\to 0} G^q(\rho^\infty(t_m)) = 0 .$$

By Lemma 2.9, the result follows. The proof of the other containment $\omega'^{\rho}(\rho_0) \subset \mathcal{E}_w$ under the assumptions (B1), (B2) (or (B2')), (B3) is exactly the same with appropriate modifications. $\qquad\square$

We now have all the ingredients to prove Theorems 2.13 and 2.14.

Proof of Theorem 2.13. From the first part of Theorem 2.12, it suffices to show

$$\lim_{t \to \infty} d_\sigma(\rho(t), \omega^{\rho}(\rho_0)) = \lim_{t \to \infty} \inf_{\mu \in \omega^{\rho}(\rho_0)} d_\sigma(\rho(t), \mu) = 0.$$

This is essentially identical to the proof of Theorem A.3. For the convenience of the reader, we redo the proof in our setting. Assume for a contradiction that this is not true, so there exists a $\delta > 0$ and a time-diverging sequence $t_n \to \infty$ as $n \to \infty$ such that

$$\inf_{\mu \in \omega^{\rho}(\rho_0)} d_\sigma(\rho(t_n), \mu)) \geq \delta, \quad \text{for all } n \in \mathbb{N}.$$

Apply Lemma 3.1 to find $\rho^* \in \mathcal{P}_p(X)$ such that, up to subsequence,

$$d_\sigma(\rho(t_n), \rho^*) \to 0, \quad \text{as } n \to \infty.$$

This is in contradiction with the previous lower bound since $\rho^* \in \omega^{\rho}(\rho_0)$. Finally, if (A2') is satisfied and $\mathcal{Y}$ is closed on $\mathcal{P}_p(X)$, it follows that $\omega^{\rho}(\rho_0) \subset \mathcal{Y}$. $\qquad\square$

Proof of Theorem 2.14. This proof follows the same steps of the previous proof using the second part of Theorem 2.12. Here, one needs to replace $d_\sigma, \omega^{\rho}(\rho_0)$, and Lemma 3.1 with $d_{LP}, \omega'^{\rho}(\rho_0)$, and Lemma 3.3, respectively. $\qquad\square$

We conclude this section with a result on the limiting behaviour of the energy functional E following similar arguments to [CH98, Theorem 9.2.3].

Proposition 3.5. *For $\rho_0 \in Z_{E,p}$ and $\rho \in AC([0, \infty); \mathcal{P}_p(X))$ the corresponding curve of maximal slope. Assume assumptions assumptions (A1), (A2) (or (A2')), and (A3) are satisfied. Then, we have*

(a) *The limit $E_\infty := \lim_{t \to \infty} E(\rho(t)) = \inf_{t>0} E(\rho(t))$ exists.*
(b) *For every $\rho^* \in \omega^{\rho}(\rho_0)$, we have $E(\rho^*) \leq E_\infty$. In particular, if*

$$E_\infty = \inf_{\rho \in \mathcal{P}_p(X)} E(\rho),$$

then every $\rho^* \in \omega^{\rho}(\rho_0)$ attains the infimum of E over $\mathcal{P}_p(X)$.
(c) *If E is continuous on $\mathcal{P}_p(X)$ with respect to the σ-topology, then*

$$E(\rho) = E_\infty,$$

for all $\rho \in \omega^{\rho}(\rho_0)$.
(d) *If $\mathcal{E}$ is a finite set, then $\omega^{\rho}(\rho_0)$ is a singleton $\{\rho^*\}$ and*

$$\lim_{t \to \infty} d_\sigma(\rho(t), \rho^*) = 0.$$

Proof. Based on Eq. (MS), we know that $t \mapsto E(\rho(t))$ is a non-increasing function. Based on assumption (A1), this function is also bounded below so its limit exists which gives (a).

For (b), we take a fixed $\rho^* \in \omega^{\rho}(\rho_0)$, so there exists a time-diverging sequence $t_n \to \infty$ as $n \to \infty$ such that $\rho(t_n) \xrightarrow{\sigma} \rho^*$. By the l.s.c. property in (A1), we have

$$E(\rho^*) \leq \liminf_{n \to \infty} E(\rho(t_n)) = \inf_{t>0} E(\rho(t)).$$

The proof of (c) is a straightforward consequence of the convergence in Theorem 2.13 and the continuity of E. For (d), we note from Theorem 2.12 that $\omega^{\rho}(\rho_0) \subset \mathcal{E}$. Thus, $\omega^{\rho}(\rho_0)$ is finite and by Theorem A.3 it is connected. It follows that it must be a singleton. The convergence follows from Theorem 2.13. $\qquad\square$

Remark 3.6. We note that an analogous result to Proposition 3.5 can be derived for the relaxed ω-limit set $\omega'^{\rho}(\rho_0)$ under assumptions (B1), (B2) (or (B2')), and (B3).

4. Applications to specific PDE

This section is devoted to applying our previous results to a wide range of gradient flows. Many of these examples have been studied extensively in the literature; we demonstrate how our abstract theory recovers and even extends some of the results. We emphasise Section 4.1 for a Fokker-Planck equation with weak confining potential in which the curve belongs to $\mathcal{P}_p(X)$ for all $t > 0$ while its steady state belongs in $\mathcal{P}(X) \setminus \mathcal{P}_p(X)$ and Section 4.8 for a pure aggregation equation in which non-unique curves converge to disjoint subsets of the set of stationary states. In Section 4.2, we examine a one dimensional noisy Kuramoto model with no confining potential although the compactness is compensated by restricting to the torus. Section 4.3 discusses the asymptotics of the McKean-Vlasov equation, that is general interaction potential with linear diffusion to get some asymptotic results even in the presence of phase transitions. We generalise the diffusion term to porous medium type in Section 4.4 with a complete description in the radial symmetric case. In Section 4.5, we turn to q-Wasserstein gradient flows for $q \neq 2$ on the torus extending the convergence results of [Agu03]. In Section 4.6, the porous medium equation is investigated to recover the well-known result of convergence to the Barenblatt profile. Removing diffusion entirely, we study a pure aggregation equation on $\mathbb{R}$ in Section 4.7, providing convergence results using our theory and classical techniques to complement the results of [GPY17] on the torus.

4.1. **Fokker–Planck equation with weak confinement.**
In this section, we study the Fokker–Planck equation on $\mathbb{R}^d$ with gradient drift which can be written as

$$\begin{cases} \partial_t \rho = \Delta \rho + \nabla \cdot (\rho \nabla V) \\ \rho(0) = \rho_0 \in \mathcal{P}_2(\mathbb{R}^d) \end{cases}, \tag{4.1}$$

where we choose V to be slowly growing at infinity as

$$V(x) = \log(1 + |x|^{2d}),$$

see [TV00, AMTU01] for related linear Fokker-Planck equations. Equation (4.1) is the forward Kolmogorov equation associated to the following Itô diffusion process:

$$\begin{cases} \mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t \\ \mathrm{Law}(X_0) = \rho_0 \in \mathcal{P}_2(\mathbb{R}^d) \end{cases}$$

where B_t is a d-dimensional Wiener process. Thus, $\mathrm{Law}(X_t) = \rho(t)$, where $\rho(t)$ is the unique weak solution of (4.1). As mentioned in the introduction, following the seminal work in [JKO98], (4.1) can be viewed as gradient flow of the free energy, $E_{FP} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$,

$$E_{FP}(\rho) = \int_{\mathbb{R}^d} \rho \log \rho \,\mathrm{d}x + \int_{\mathbb{R}^d} \log(1 + |x|^{2d})\,\mathrm{d}\rho(x),$$

with respect to the weak upper gradient $G_{FP} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$,

$$G_{FP}(\rho) = \left(\int_{\mathbb{R}^d} \left| \nabla \log(\rho(1 + |x|^{2d})) \right|^2 \,\mathrm{d}\rho(x) \right)^{1/2}.$$

Indeed, we have the following result

Proposition 4.1. *For all $\rho_0 \in Z_{E_{FP},2}$, there exists a unique 2-curve of maximal slope $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ of the energy E_{FP} with respect to the weak upper gradient G_{FP} such that $\rho(0) = \rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$. Furthermore, any such 2-curve of maximal is equivalent to a weak solution of (4.1).*

Proof. The proof of this result follows from [AGS08, Theorem 11.2.8] and the fact that V is λ-convex. $\qquad\square$

The reason we have made this choice of the confinement V is because it allows us to demonstrate our abstract convergence results in the setting in which the stationary solutions are not necessarily in $\mathcal{P}_2(\mathbb{R}^d)$, i.e. Theorem 2.14. Indeed, we have the following result

Proposition 4.2. *Consider the extensions $\tilde{E}_{FP}$ and $\tilde{G}_{FP}$ of E_{FP} and G_{FP} to $\mathcal{P}(\mathbb{R}^d)$. Then,*

$$\rho_\infty(\mathrm{d}x) = \frac{Z_\infty}{1 + |x|^{2d}}\,\mathrm{d}x\,, \quad Z_\infty = \left(\int_{\mathbb{R}^d} \frac{1}{1 + |x|^{2d}}\,\mathrm{d}x\right)^{-1},$$

is the unique minimiser of $\tilde{E}_{FP}$ and unique zero of $\tilde{G}_{FP}$ over $\mathcal{P}(\mathbb{R}^d)$. It follows that it is the unique weak stationary state of the associated curve of maximal slope in the sense of Definition 2.10. Furthermore, $\rho_\infty \in \mathcal{P}(\mathbb{R}^d) \setminus \mathcal{P}_2(\mathbb{R}^d)$ and is a stationary weak solution of (4.1).

Proof. That ρ_∞ is the unique minimiser follows from Jensen's inequality. Indeed, we have that for any $\rho \in \mathcal{P}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$

$$\begin{aligned}
\tilde{E}_{FP}(\rho) &= \int_{\mathbb{R}^d} \rho \log \rho \,\mathrm{d}x + \int_{\mathbb{R}^d} \log(1 + |x|^{2d})\,\mathrm{d}\rho(x) \\
&= \int_{\mathbb{R}^d} \frac{\rho}{\rho_\infty} \log\left(\frac{\rho}{\rho_\infty}\right)\mathrm{d}\rho_\infty + \log(Z_\infty) \\
&\geq \log(Z_\infty) = \tilde{E}_{FP}(\rho_\infty)\,.
\end{aligned}$$

Note that the final inequality is strict unless $\rho = \rho_\infty$. Thus, the result follows. That ρ_∞ is a zero of $\tilde{G}_{FP}$ is apparent by plugging it into the expression for $\tilde{G}_{FP}$. Furthermore, for uniqueness we note that any zero $\rho \in \mathcal{P}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ of $\tilde{G}_{FP}$ must satisfy

$$\rho(x) = \frac{C_A}{1 + |x|^{2d}}\,,$$

for a.e. $x \in A$, for every connected component A of its support. If ρ is not fully supported then this would imply that $\log(\rho(1 + |x|^{2d}))$ is sum of indicator functions of disjoint sets. Since the indicator function of any set is not in $H^1(\mathbb{R}^d)$ (unless the set is $\mathbb{R}^d$), it follows that

$$\tilde{G}_{FP} = +\infty\,.$$

Thus, ρ must have full support and so is equal to ρ_∞. The fact that ρ_∞ is a weak stationary solution of (4.1) follows by plugging it in. $\qquad\square$

We can now state and prove our convergence result:

Proposition 4.3. *Let $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ be the unique 2-curve of maximal slope of the energy E_{FP} with respect to the weak upper gradient G_{FP}, for some initial datum $\rho_0 \in Z_{E_{FP},2}$. Then,*

$$\lim_{t \to \infty} d_{LP}(\rho(t), \rho_\infty) = 0\,,$$

where $\rho_\infty \in \mathcal{P}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ is as given in Proposition 4.1.

Proof. We check that the conditions of Theorem 2.14 hold true. Note that we have already shown in the proof of Proposition 4.1 that $\tilde{E}_{FP}$ is bounded below and thus it is proper. Furthermore, the fact that it is l.s.c. follows from standard results (cf. [JLJ98, Lemma 4.3.1]) and Fatou's lemma. Thus, (B1) is satisfied. Note that $\tilde{G}_{FP}$ is of the form (LSC) and so (B3) is satisfied. Furthermore, for any element of a sublevel set of the energy $\rho \in L_{\leq C}(\tilde{E}_{FP})$, we have that

$$\frac{1}{2}\int_{\mathbb{R}^d} \log(1 + |x|^{2d})\,\mathrm{d}\rho \leq C - \int_{\mathbb{R}^d} \rho \log \rho \,\mathrm{d}x - \frac{1}{2}\int_{\mathbb{R}^d} \log(1 + |x|^{2d})\,\mathrm{d}\rho(x)\,.$$

Using a similar argument as in the proof of Proposition 4.1, we have that

$$\int_{\mathbb{R}^d} \rho \log \rho \,\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}^d} \log(1 + |x|^{2d})\,\mathrm{d}\rho(x) \geq -C_1\,,$$

for some $C_1 > 0$ independent of ρ. It follows that for all $\rho \in L_{\leq C}(\tilde{E}_{FP})$

$$\frac{1}{2}\int_{\mathbb{R}^d} \log(1 + |x|^{2d})\,\mathrm{d}\rho \leq C_2\,.$$

Thus, for all $\rho \in L_{\leq C}(\tilde{E}_{FP})$

$$\int_{B_R^c(0)} \mathrm{d}\rho \leq \frac{2C_2}{\log(1 + R^{2d})}\,.$$

By Prokhorov's theorem, (B2) is satisfied. Thus, by Theorem 2.14,

$$\lim_{t \to \infty} d_{LP}(\rho(t), \mathcal{E}_w) = 0 \,,$$

where $\mathcal{E}_w$ is the set of weak stationary states as defined in Definition 2.10. But by Proposition 4.1, $\mathcal{E}_w = \{\rho_\infty\}$. Thus, the result follows. □

4.2. **The noisy Kuramoto model.**   In this section, we discuss a special case of the McKean–Vlasov equation posed on the one-dimensional unit torus, $\mathbb{T}$ (or, equivalently, the unit circle $\mathbb{S}$). This corresponds to the choice of the interaction potential $W_\kappa = -\kappa \cos(2\pi x), \kappa > 0$, which leads to the following PDE:

$$\begin{cases} \partial_t \rho = \partial_x^2 \rho + \partial_x(\rho W_\kappa' * \rho) & (x,t) \in \mathbb{T} \times [0, \infty) \\ \rho(0) = \rho_0 \in \mathcal{P}_2(\mathbb{T}) \end{cases}. \qquad (4.2)$$

The above equation describes the so-called noisy Kuramoto model for synchronising oscillators. It can be derived as the mean field limit of the following set of interacting SDEs on the torus:

$$\begin{cases} dX_t^i = -\dfrac{1}{N} \sum\limits_{j=1}^{N} W_\kappa'(X_t^i - X_t^j) \, dt + \sqrt{2} \, dB_t^i \\ \mathrm{Law}(X_t^1, \ldots, X_t^N) = \rho_0^{\otimes N} \in \mathcal{P}_2(\mathbb{T}^N) \end{cases}, \qquad (4.3)$$

where the X_t^i represent the phases of the individual oscillators and the B_t^i are independent $\mathbb{T}$-valued Wiener processes. We refer the reader to [Kur81, SSK88, ABPV$^+$05] for a review of the Kuramoto model and its variants. We note that the Kuramoto model and the dynamics described in (4.3) also correspond to the overdamped Langevin dynamics associated to the so-called Heisenberg XY model from statistical physics for continuous spins on a lattice with mean field interaction (cf. [FV18, Chapter 9]).

As expected, we can think of (4.2) as a curve of maximal slope of the energy $E_K : \mathcal{P}_2(\mathbb{T}) \to (-\infty, +\infty]$,

$$E_K(\rho) := \int_\mathbb{T} \rho \log \rho \, dx + \frac{1}{2} \iint_{\mathbb{T} \times \mathbb{T}} W_\kappa(x - y) \, d\rho(x) \, d\rho(y) \,,$$

with respect to the weak upper gradient $G_K : \mathcal{P}_2(\mathbb{T}) \to (-\infty, +\infty]$,

$$G_K(\rho) := \left(\int_\mathbb{T} \left| \nabla \log \frac{\rho}{e^{-W_k * \rho}} \right|^2 d\rho \right)^{1/2} .$$

Indeed, we have the following result which we state without proof:

Proposition 4.4. *Let $\rho_0 \in Z_{E_K, 2}$. Then, there exists a unique 2-curve of maximal slope $\rho \in AC([0, \infty); \mathcal{P}_2(\mathbb{T}))$ of the energy E_K with respect to the weak upper gradient G_K such that $\rho(0) = \rho_0$. Furthermore, curves of maximal slope are equivalent to weak solutions of (4.2).*

The reason we have made this choice of the interaction potential is because the noisy Kuramoto model is one of the simplest models of this form which exhibits the phenomenon of phase transitions, i.e. a change in structure of stationary solutions and minimisers of the energy as the parameter $\kappa > 0$ is varied. Of course the Kuramoto model is not the only model which exhibits this phenomenon. We refer the reader to [CP10, CGPS20, CG19] where the phenomenon of phase transitions and bifurcations for such equations is discussed in detail in a more general context. We remark that phase transitions also occur in the case of the McKean–Vlasov equation on $\mathbb{R}^d$ discussed in Section 4.3, see [Daw83, Tug14a, BCnCD16, Li19]. We now state without proof the following result which provides a complete characterisation of the set of stationary states and minimisers of the curve of maximal slope associated to the Kuramoto model:

Proposition 4.5. [CGPS20, Proposition 6.1] *Assume $\kappa \leq 2$. Then, there exists a unique minimiser of E_K and stationary state of the associated curve of maximal slope given by $\rho_\infty := \mathcal{L}_{\mathbb{T}^d}$, the normalised Lebesgue measure on $\mathbb{T}^d$. On the other hand, if $\kappa > 2$, there exist two (up to translations) stationary states of the curve of maximal slope given by ρ_∞ and*

$$\rho^*(dx) := Z_\kappa^{-1} e^{\sigma(\kappa) \cos(2\pi x)} \, dx, \qquad Z_\kappa := \int_\mathbb{T} e^{\sigma(\kappa) \cos(2\pi x)} \, dx \,,$$

for some $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{\kappa \to \infty} \sigma(\kappa) = +\infty \,.$$

Furthermore, ρ^* is the unique (up to translations) minimiser of E_K for $\kappa > 2$.

We are finally in a position to state our convergence result:

Proposition 4.6. *Let $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{T}))$ be the unique 2-curve of maximal slope of the energy E_K with respect to the weak upper gradient G_K, for some initial datum $\rho_0 \in \mathcal{P}_2(\mathbb{T})$ with $E_K(\rho_0) < \infty$. Then, if $\kappa \leq 2$,*

$$\lim_{t \to \infty} W_\infty(\rho(t), \rho_\infty) = 0 \,,$$

where $\rho_\infty(\mathrm{d}x) = \mathrm{d}x$. On the other hand if $\kappa > 2$ and $E_K(\rho_0) < E_K(\rho_\infty) = 0$, then

$$\lim_{t \to \infty} W_\infty(\rho(t), T_\rho^*) = 0 \,,$$

where ρ^* is as defined in Proposition 4.5 and T_{ρ^*} is the set of all translates of ρ^*. Furthermore, if ρ_0 is symmetric about some $x^* \in \mathbb{T}$, i.e.

$$\rho_0(x^* + \cdot) = \rho_0(x^* - \cdot) \,,$$

then

$$\lim_{t \to \infty} W_\infty(\rho(t), \rho_{x^*}^*) = 0 \,,$$

where $\rho_{x^*}^* \in T_{\rho^*}$ is such that

$$\int_{\mathbb{T}} (x - x^*) \rho_{x^*}^* \, \mathrm{d}x = 0 \,.$$

Proof. That (A1) is satisfied follows from the fact that the energy is finite at ρ_∞ and we have

$$E_K(\rho) \geq -\frac{1}{2} \|W_\kappa\|_{L^\infty(\mathbb{T})}.$$

Again, the fact that E_K is l.s.c. follows from standard results. Since $\mathcal{P}(\mathbb{T}) = \mathcal{P}_2(\mathbb{T})$ and G_K is of the form (LSC), it follows that (A3) is satisfied. Finally, since $\mathcal{P}(\mathbb{T})$ is compact, (A2) is trivially satisfied. Applying Theorem 2.13, we have that

$$\lim_{t \to \infty} d_\sigma(\rho(t), \mathcal{E}^K) = 0 \,,$$

where $\mathcal{E}^K$ is set of all stationary states of the associated curve of maximal slope. We know from Proposition 4.5 that, for $\kappa \leq 2$, $\mathcal{E}^K = \{\rho_\infty\}$. Thus, for $\kappa \leq 2$,

$$\lim_{t \to \infty} d_\sigma(\rho(t), \rho_\infty) = 0 \,.$$

On the other hand for $\kappa > 2$, we know from Proposition 4.5, that $E_K(\rho^*) < E_K(\rho_\infty)$. Since $t \mapsto E_K(\rho(t))$ is non-increasing, it follows that if $E_K(\rho_0) < E_K(\rho_\infty) = 0$ and $\kappa > 2$, we have that

$$\lim_{t \to \infty} d_\sigma(\rho(t), T_{\rho^*}) = 0 \,,$$

where T_{ρ^*} is the set of all translates of the measure ρ^* defined in Proposition 4.5. We now note that weak solutions of (4.2) with symmetric initial data about some $x^* \in \mathbb{T}$ remain symmetric about x^* for all $t \geq 0$. This can be seen by performing a change of variables $x \mapsto -x$ in the associated curve of maximal slope (MS) and noting, from Proposition 4.4, that curves of maximal slope are unique and equivalent to weak solutions of (4.2). Since elements of T_{ρ^*} are all symmetric about some point and since this symmetry is preserved under weak convergence, the convergence to a distinct $\rho_{x^*}^* \in T_{\rho^*}$ follows such that

$$\int_{\mathbb{T}} (x - x^*) \rho_{x^*}^* \, \mathrm{d}x = 0 \,.$$

We can improve the convergence in $d_\sigma(\cdot, \cdot)$ to any $W_p(\cdot, \cdot)$, $1 \leq p < \infty$, by using the fact that $\mathcal{P}_p(\mathbb{T}^d)$ is compact for all $1 \leq p < \infty$. Since ρ_∞ is bounded below away from zero, we can apply [BJR07, Theorem 1.2] to obtain convergence in $W_\infty(\cdot, \cdot)$, thus completing the proof of the result. $\qquad \square$

Remark 4.7. We note that in the case of general initial data $\rho_0 \in \mathcal{P}_2(\mathbb{T})$ it is non-trivial to identify which distinct element of T_{ρ^*} is seen in the long-time limit. We refer the reader to the discussion in [GPP12, Lemma 2.2, Theorem 4.6] in which it is shown that under certain mild assumptions a distinct limit in T_{ρ^*} is selected.

4.3. **McKean–Vlasov equation on $\mathbb{R}^d$.** We consider the McKean–Vlasov equation on $\mathbb{R}^d$ as considered in [CMV03, Vil03, CMV06, Tug13, Tug14b]. It describes the density of a so-called self-stabilising mean field McKean SDE and is given by

$$\begin{cases} \partial_t \rho = \nabla \cdot (\rho \nabla (\log \rho + V + W * \rho)) \\ \rho(0) = \rho_0 \in \mathcal{P}_2(\mathbb{R}^d) \end{cases}, \tag{4.4}$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a confining potential and $W : \mathbb{R}^d \to \mathbb{R}$ is an interaction potential satisfying the following assumptions:

$V \in C^2(\mathbb{R}^d)$ and there exists $\lambda > 0$ such that $D^2V(x) \geq \lambda$ for all $x \notin K \subset \mathbb{R}^d$, compact, (V1)

$$\lim_{|x| \to \infty} \|D^2V(x)\| = +\infty, \tag{V2}$$

$W \in C^2(\mathbb{R}^d)$ even and positive. (W1)

Under these rather minimal assumptions we can apply [AGS08, Theorem 11.2.8] to assert that weak solutions of (4.4) are equivalent to 2-curves of maximal slope of the energy $E_{MV} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$, given by,

$$E_{MV}(\rho) := \int_{\mathbb{R}^d} \rho \log \rho \, \mathrm{d}x + \int_{\mathbb{R}^d} V(x) \, \mathrm{d}\rho(x) + \frac{1}{2} \int_{\mathbb{R}^d} (W * \rho)(x) \, \mathrm{d}\rho(x) \,.$$

with respect to the weak upper gradient $G : \mathcal{P}_2(\mathbb{R}^d) \to [0, \infty]$ which is given by

$$G_{MV}(\rho) := \int_{\mathbb{R}^2} \left| \nabla \log \frac{\rho}{e^{-V-W*\rho}} \right|^2 \mathrm{d}\rho(x) \,,$$

if the above quantities are finite or as $+\infty$ otherwise. It follows that stationary states of the 2-curve of maximal slope are equivalent to stationary solutions of (4.4). To apply our results, we need to check that (A2) and (A3) hold true:

Lemma 4.8. *Under the assumptions* (V1), (V2), *and* (W1), *the energy* E_{MV} *satisfies* (A1) *and* (A2). *Furthermore, the sublevel set* $L_{\leq C}(E_{MV}) := \{\rho \in \mathcal{P}_2(\mathbb{R}^d) : E_{MV}(\rho) \leq C\}, C \in \mathbb{R}$ *is compact in* $\mathcal{P}_2(\mathbb{R}^d)$ *with respect to the* $W_{2-\varepsilon}(\cdot, \cdot)$ *metric.*

Proof. The fact that E_{MV} is proper and l.s.c, i.e. (A1), follows from standard results (cf. [JLJ98, Lemma 4.3.1]) and Fatou's lemma. In order to prove (A2), we use the fact that assumptions (V1) and (V2) imply that

$$\lim_{|x| \to \infty} \left(\frac{V(x)}{|x|^2} - 1 \right) = +\infty \,.$$

Indeed, for all $C > 0$ we can find $x_0 \in \mathbb{R}^d$ such that $D^2V(x) > C$ for all $|x| > |x_0|$, such that

$$\lim_{|x| \to \infty} \frac{V(x)}{|x|^2} - 1 = +\infty \,.$$

Thus, we can find a ball of size R_0 such that

$$\frac{V(x)}{|x|^2} - 1 > 1, \tag{4.5}$$

for all $x \in B_{R_0}^c$. Now, using (W1) we have that

$$\frac{1}{2} \int_{\mathbb{R}^d} (W * \rho)(x) \, \mathrm{d}\rho(x) \geq 0 \,.$$

Also, we have the following bound for the entropic term

$$\int_{\mathbb{R}^d} \rho \log \rho \, \mathrm{d}x = \int_{\mathbb{R}^d} \rho \log \frac{\rho}{e^{-|x|^2}} \, \mathrm{d}x - \int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\rho$$

$$\geq C_1 - \int_{\mathbb{R}^d} |x|^2 \, d\rho \,,$$

where in the last step we simply apply Jensen's inequality. Putting these together we have that any element ρ of a sublevel set of the energy E_{MV}, $L_{\leq C}(E_{MV}), C \in \mathbb{R}$ must satisfy

$$\int_{\mathbb{R}^d} \left(\frac{V(x)}{|x|^2} - 1 \right) |x|^2 \, d\rho(x) \leq C - C_1 =: C_2 \,.$$

Fix some $\delta > 0$ and consider $R_1 > R_0$. We then have for any $\rho \in L_{\leq C}$ (which is necessarily absolutely continuous) that

$$\int_{B_{R_1}} |x|^2 \, d\rho = \int_{\mathbb{R}^d} |x|^2 \, d\rho - \int_{B_{R_1}^c} |x|^2 \, d\rho \,.$$

Applying (4.5), we have that

$$\frac{V(x)}{|x|^2} - 1 > \log \left(\frac{\inf_{\partial B_{R_1}} V(x)}{R_1^2} - 1 \right), \quad \text{for all } x \in B_{R_1}^c \,.$$

Thus, we obtain

$$\begin{aligned}
\int_{B_{R_1}} |x|^2 \, d\rho &= \int_{\mathbb{R}^d} |x|^2 \, d\rho - \int_{B_{R_1}^c} |x|^2 \, d\rho \\
&> \int_{\mathbb{R}^d} |x|^2 \, d\rho - \frac{1}{\log \left(\frac{\inf_{\partial B_{R_1}} V(x)}{R_1^2} - 1 \right)} \int_{B_{R_1}^c} \left(\frac{V(x)}{|x|^2} - 1 \right) |x|^2 \, d\rho \\
&\geq \int_{\mathbb{R}^d} |x|^2 \, d\rho - \frac{1}{\log \left(\frac{\inf_{\partial B_{R_1}} V(x)}{R_1^2} - 1 \right)} (C - C_1) \,.
\end{aligned}$$

Making R_1 large enough (independent of the choice of $\rho \in L_{\leq C}$), we can obtain

$$\int_{B_{R_1}} |x|^2 \, d\rho \geq \int_{\mathbb{R}^d} |x|^2 \, d\rho - \delta \,.$$

Thus, we have

$$\int_{B_{R_1}^c} |x|^2 \, d\rho \leq \delta \,.$$

Thus, by Prokhorov's theorem, we have compactness of $L_{\leq C}(E_{MV})$ in $\mathcal{P}_{2-\varepsilon}(\mathbb{R}^d)$ with respect to the metric $W_{2-\varepsilon}(\cdot, \cdot)$. for all $0 < \varepsilon \leq 1$. Thus, (A2) is satisfied. $\qquad \square$

Remark 4.9. Note that the compactness we have derived is stronger than that required by (A2). Indeed, we have obtained compactness of the sublevel set $L_{\leq C}$ in $\mathcal{P}_{2-\varepsilon}(\mathbb{R}^d)$ equipped with $W_{2-\varepsilon}$ rather than with respect to the σ-topology on $\mathcal{P}_2(\mathbb{R}^d)$.

Lemma 4.10. *Under the assumptions* (V1), (V2), *and* (W1), *the weak upper gradient* G_{MV} *satisfies* (A3).

Proof. We note that G_{MV} can be trivially extended to a function $\tilde{G}_{MV} : \mathcal{P}(\mathbb{R}^d) \to [0, \infty]$ such that $\tilde{G}_{MV}(\rho) = G_{MV}(\rho)$ for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$. $\tilde{G}_{MV}$ is of the form (LSC) and thus is l.s.c. on $\mathcal{P}(\mathbb{R}^d)$ by Theorem B.1. It follows that G_{MV} is l.s.c with respect to the σ-topology on $\mathcal{P}_2(\mathbb{R}^d)$. $\qquad \square$

We conclude this subsection with the following convergence result which recovers the main results of [Tug13]:

Proposition 4.11. *Let* $\rho \in AC([0, \infty); \mathcal{P}_2(\mathbb{R}^d))$ *be a 2-curve of maximal slope of the energy* E_{MV} *with respect to the weak upper gradient* G_{MV}, *for some initial datum* $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *with* $E_{MV}(\rho_0) < \infty$. *Then,*

$$\lim_{t \to \infty} W_{2-\varepsilon}(\rho(t), \mathcal{E}^{MV}) = 0 \,,$$

for all $0 < \varepsilon \leq 1$ *where* $\mathcal{E}^{MV} \subseteq \mathcal{P}_2(\mathbb{R}^d)$ *is the set of stationary states of the 2-curve of maximal slope associated to* E_{MV} *and* G_{MV}.

Proof. Since by Lemmas 4.8 and 4.10, (A1), (A2), and (A3) are satisfied, we have by application of Theorem 2.13 that
$$\lim_{t\to\infty} d_\sigma(\rho(t), \mathcal{E}^{MV}) = 0 \,.$$
Using Lemma 2.15 and the stronger $W_{2-\varepsilon}$ compactness from Lemma 4.8, the result of the proposition follows. □

Remark 4.12. As shown in [Daw83, Tug14a, BCnCD16, GP18, Li19] in the case of bistable or multivalleys confinement potentials, McKean-Vlasov equations posed on $\mathbb{R}^d$ of the form discussed in (4.4) also exhibit the phenomenon of phase transitions (cf. Section 4.2). A typical example of such a system is provided by the so-called Desai–Zwanzig model. To obtain this model, we set in one dimension
$$V(x) = x^2 - \frac{x^4}{2} \qquad \text{and} \qquad W_\kappa(x) = \kappa \frac{x^2}{2}$$
in (4.4). This system also exhibits a phase transition, i.e. as the value of the parameter κ is increased, the system goes from exhibiting one stationary solution to exactly three. We remark that the results we present in Section 4.2, especially Proposition 4.6, hold true analogously in the setting of the Desai–Zwanzig model or in more complicated models of phase transitions in higher dimensions as discussed in [Tug14a, Li19]. Furthermore, for the Desai–Zwanzig model, we can apply Proposition 3.5(c) to argue that the solution converges to a unique stationary state (due to the finiteness of the set of stationary solutions).

4.4. Aggregation-diffusion equation.
In this section we study the aggregation diffusion equation equation given by
$$\partial_t \rho = \Delta(\rho^m) + \nabla \cdot (\rho \nabla W * \rho) \qquad (t,x) \in (0,\infty) \times \mathbb{R}^d \tag{4.6}$$
in the regime discussed in [CHVY19, DYY19], see [CCY19] for a recent survey. We place the following assumptions on the interaction potential W

$$W \in C^\infty(\mathbb{R}^d) \text{ is radially symmetric, non-negative, and } W'(r) > 0, r > 0 \tag{W2}$$

$$\text{There exists some } C_W > 0 \text{ such that } W'(r) \le C_W \tag{W3}$$

$$\text{For all } a, b \ge 0 \text{ it holds that } W(a+b) \le C_W(1 + W(1+a) + W(1+b)) \tag{W4}$$

$$W \text{ is } \lambda\text{-convex} \tag{W5}$$

$$\lim_{r\to\infty} W(r) = +\infty \tag{W6}$$

$$\left(\lim_{r\to\infty} W(r)\right) - W \in L^p(\mathbb{R}^d) \text{ for some } 1 \le p < \infty \tag{W7}$$

Note that we have modified the assumptions of [CHVY19, DYY19] so as to ensure that curves of maximal slope exist from [AGS08, Theorem 11.2.8]. We have excluded singular interactions so that we do not have to deal with technical difficulties in the gradient flow formulation of (4.6) arising from singularities at the origin. Before proceeding to state and prove our results on convergence for (4.6), we refer the reader to [Shu20] where convergence to the steady state for (4.6) is discussed for $d = 1$ and with less restrictive growth assumptions on W.

Proposition 4.13. *Let $\rho_0 \in Z_{E_{AD},2}$. Then, under assumptions (W2), (W3), and (W4), there exists a unique 2-curve of maximal slope $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ of the energy*
$$E_{AD}(\rho) := \frac{1}{m-1} \int_{\mathbb{R}^d} \rho(x)^m \, dx + \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y) \, d\rho(x) \, d\rho(y) \,,$$
with respect to the weak upper gradient
$$G_{AD}(\rho) := \left(\int_{\mathbb{R}^d} \left| \nabla \left(\frac{m}{m-1} \rho^{m-1} + W * \rho \right) \right|^2 d\rho(x) \right)^{1/2} ,$$
such that $\rho(0) = \rho_0$. Furthermore, curves of maximal slope are equivalent to weak solutions of (4.6).

Proof. The proof of this result is an application of [DS10, Theorem 4.20]. We just need to check that W satisfies the so-called doubling condition and is λ-convex. The doubling condition is exactly (W4). The fact that it is λ-convex is exactly (W5). □

We now characterise the minimisers of the associated free energy:

Proposition 4.14 (Existence and uniqueness of minimisers)**.** *Assume that $m \geq 2$ and that assumptions* (W2)*,* (W3)*,* (W4)*, and either* (W6) *or* (W7)*, are satisfied. Note that E_{AD} and G_{AD} have extensions to $\mathcal{P}(\mathbb{R}^d)$ in the sense of Definition* 2.10*. Then, $\tilde{E}_{AD}$ has a unique (up to translations) minimiser $\rho_* \in \mathcal{P}(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ such that ρ_* is compactly supported, radially symmetric and decreasing. Furthermore, ρ^* is the unique weak stationary state of the associated curve of maximal slope.*

Proof. The proof of this result is a combination of results from [CHVY19, DYY19]. We first apply [CHVY19, Lemma 3.9] to argue that all stationary weak solutions of (4.6) are essentially bounded. It then follows from [CHVY19, Theorem 2.2] that any stationary solution is compactly supported, radially symmetric, and decreasing (up to a translation). Finally, from [DYY19, Theorem 1.1] for $m \geq 2$, we know that stationary solutions are unique up to translations. Since, by the result of [AGS08, Theorem 10.4.13], stationary weak solutions of (4.6) are equivalent to weak stationary states, we have that there exists a unique up to translation weak stationary state, $\rho^* \in \mathcal{P}(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$, of the associated curve of maximal slope which compactly supported, radially symmetric, and decreasing. Finally, we argue that $\tilde{E}_{AD}$ has a minimiser. Since any minimiser is a weak stationary state it follows that ρ^* is also the unique minimiser of the free energy. $\square$

Motivated by the previous result we introduce the set of all radial probability measures centred at $x_0 \in \mathbb{R}^d$, which is defined as follows

$$\mathcal{P}_2^{R,x_0}(\mathbb{R}^d) := \left\{ \rho \in \mathcal{P}_2(\mathbb{R}^d) : \rho(A - x_0) = \rho(\pi(A - x_0)), \forall \pi \in O(d), \forall A \subseteq \mathbb{R}^d, \text{ measurable} \right\},$$

where $O(d)$ is the group of all $d \times d$ orthogonal matrices. We finally are the position to state our convergence result:

Proposition 4.15. *Assume that $m \geq 2$ and that assumptions* (W2)*,* (W3)*,* (W4)*, and either* (W6) *or* (W7)*, are satisfied. Let $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ be the unique 2-curve of maximal slope of the energy E_{AD} with respect to the weak upper gradient G_{AD}, for some initial datum $\rho_0 \in \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$, $x_0 \in \mathbb{R}^d$, with $E_{MV}(\rho_0) < \infty$. Then,*

$$\lim_{t \to \infty} d_\sigma(\rho(t), \rho^*) = 0,$$

where $\rho^* \in L^\infty(\mathbb{R}^d) \cap \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$ is as given in Proposition 4.14*, i.e. it is the unique stationary weak solution of* (4.6) *and minimiser of E_{AD} with mean $x_0 \in \mathbb{R}^d$, i.e.*

$$\int_{\mathbb{R}^d} (x - x_0) \, d\rho^*(x) = 0.$$

Proof. We first argue that if $\rho \in \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$, then

$$\int_{\mathbb{R}^d} (x - x_0) \, d\rho = 0.$$

Indeed, we have that

$$\int_{\mathbb{R}^d} (x - x_0) \, d\rho = \int_{\mathbb{R}^d} x \, d(\tau_{-x_0}\rho),$$

where $\tau_x \rho$ is characterised by

$$(\tau_x \rho)(A) = \rho(A - x),$$

for all measurable $A \subseteq \mathbb{R}^d$. Note that if $\rho \in \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$, then $\tau_{-x_0}\rho \in \mathcal{P}_2^{R,0}(\mathbb{R}^d)$. We thus have that

$$\int_{\mathbb{R}^d} (x - x_0) \, d\rho = \int_{\mathbb{R}^d} x \, d(\tau_{-x_0}\rho)$$
$$= \int_{H_1^+} x \, d(\tau_{-x_0}\rho) + \int_{H_1^-} x \, d(\tau_{-x_0}\rho) + \int_{H_1} x \, d(\tau_{-x_0}\rho),$$

where H_i is the hyperplane given by

$$H_i = \{x \in \mathbb{R}^d : x_i = 0\},$$

and

$$H_i^+ = \{x \in \mathbb{R}^d : x_i > 0\}$$
$$H_i^- = \{x \in \mathbb{R}^d : x_i < 0\}.$$

Choosing $\pi : \mathbb{R}^d \to \mathbb{R}^d, x \mapsto -x \in O(d)$ and using , we obtain

$$\int_{\mathbb{R}^d} (x - x_0) \, \mathrm{d}\rho = \int_{H_1^+} x \, \mathrm{d}(\tau_{-x_0}\rho) + \int_{H_1^-} x \, \mathrm{d}(\tau_{-x_0}\rho)(x) + \int_{H_1} x \, \mathrm{d}(\tau_{-x_0}\rho)$$
$$= \int_{H_1^+} x \, \mathrm{d}(\tau_{-x_0}\rho) - \int_{H_1^+} x \, \mathrm{d}(\tau_{-x_0}\rho) + \int_{H_1} x \, \mathrm{d}(\tau_{-x_0}\rho)$$
$$= \int_{H_1} x \, \mathrm{d}(\tau_{-x_0}\rho).$$

If $\rho \in L^1(\mathbb{R}^d)$, the proof is complete since H_1 has Lebesgue measure zero. If not we can continue decomposing H_1 into $H_1 \cap H_2^+, H_1 \cap H_2^-, H_1 \cap H_2$ and apply the same argument. The argument terminates at some $H_i, i < d$, unless ρ is concentrated at x_0 in which case the proof is trivial. Let $\rho \in AC([0, \infty); \mathcal{P}_2(\mathbb{R}^d))$ be the curve of maximal slope defined in the statement of the proposition. Since ρ is also a weak solution of (4.6), testing against x in the weak formulation of (4.6) we obtain

$$\int_{\mathbb{R}^d} x \, \mathrm{d}\rho(t) = x_0,$$

for all $t \geq 0$. Furthermore, one can check that the weak formulation of (4.6) is invariant under the action of elements of $\pi \in O(d)$. Thus, if we set $\mathcal{Y} = \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$ we have that $\bigcup_{t \geq 0} \{\rho(t)\} \subseteq \mathcal{Y}$. Note that E_{AD} and G_{AD} have extensions to $\mathcal{P}(\mathbb{R}^d)$ in the sense of Definition 2.10. The fact that $\tilde{E}_{AD}$ is proper and l.s.c. follows from standard results (cf. [JLJ98, Lemma 4.3.1]). Thus, (B1) is satisfied. Furthermore, $\tilde{G}_{AD}$ is of the form specified in (LSC) and thus (B3) is satisfied. Now consider the set

$$L_{\leq C}(\tilde{E}_{AD}) = \left\{ \mu \in \mathcal{P}_2(\mathbb{R}^d) \cap \mathcal{Y} : \tilde{E}_{AD}(\mu) \leq C \right\}.$$

We now follow the standard argument to obtain compactness in the radial setting as illustrated in [CCV15, Theorem 2.1] or [McC94]. For any $\mu \in L_{\leq C}(\tilde{E}_{AD})$, we have that

$$C \geq \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} W(x - y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)$$
$$\geq \frac{1}{2} \int_{\mathbb{R}^d} \int_{|x+x_0-y|>1} W(x - y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) + \frac{1}{2} \int_{\mathbb{R}^d} \int_{|x+x_0-y|\leq 1} W(x - y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)$$
$$\geq \frac{1}{2} \int_{\mathbb{R}^d} \int_{|x+x_0-y|>1} W(x - y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) - \frac{1}{2} \|W\|_{L^\infty(B_{1+|x_0|}(0))}$$

We now note that for any $x \in \mathbb{R}^d$ with $|x| \geq 1$, $\{y \in \mathbb{R}^d : x \cdot (y - x_0) \leq 0\} \subseteq \{y \in \mathbb{R}^d : |x + x_0 - y| > 1\}$. It follows that for every $R \geq 1$ it holds that

$$C + \frac{1}{2} \|W\|_{L^\infty(B_{1+|x_0|}(0))} \geq \frac{1}{2} \int_{|x| \geq R} \int_{x \cdot (y - x_0) \leq 0} W(x - y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y).$$

Note that $x \cdot (y - x_0) \leq 0$ implies that $|x + x_0 - y| \geq |x|$. Since W is radial and monotone increasing, we obtain

$$C + \frac{1}{2} \|W\|_{L^\infty(B_{1+|x_0|}(0))} \geq \frac{1}{2} \int_{|x| \geq R} \int_{x \cdot (y - x_0) \leq 0} W(x) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)$$
$$\geq \frac{W(R)}{2} \int_{|x| \geq R} \int_{x \cdot (y - x_0) \leq 0} \mathrm{d}\mu(x) \, \mathrm{d}\mu(y).$$

We now use the fact that $\mu \in \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$, that $\{y \in \mathbb{R}^d : x \cdot (y - x_0) \leq 0\}$ is a half space passing through x_0, and that μ is not singular since $\tilde{E}_{AD}(\mu) \leq C$ to assert that

$$\int_{x \cdot (y - x_0) \leq 0} \mathrm{d}\mu(y) = \frac{1}{2},$$

for all $x \in \mathbb{R}^d$. It follows that

$$C + \frac{1}{2}\|W\|_{L^\infty(B_{1+|x_0|}(0))} \geq \frac{W(R)}{4}\int_{|x|\geq R} \mathrm{d}\mu(x).$$

Since $W(R) \to \infty$ as $R \to \infty$, by Prokhorov's theorem, the set $L_{\leq C}(\tilde{E}_{AD})$ is compact in $\mathcal{P}(\mathbb{R}^d)$. Thus, (B2') is satisfied. Applying the result of Theorem 2.14 we obtain that

$$\lim_{t\to\infty} d_{LP}(\rho(t), \mathcal{E}_w) = 0,$$

where $\mathcal{E}_w$ is the set of weak stationary states of the curve of maximal slope in the sense of Definition 2.10. Note however that by Proposition 4.14 it follows that

$$\mathcal{E}_w \subset \mathcal{P}_2(\mathbb{R}^d),$$

and that $\mathcal{E}_w$ consists of translates of a single measure ρ^*. It follows that

$$\lim_{t\to\infty} d_\sigma(\rho(t), \mathcal{E}_w) = 0,$$

since $d_\sigma(\cdot, \cdot)$ is just the restriction of $d_{LP}(\cdot, \cdot)$ to $\mathcal{P}_2(\mathbb{R}^d)$. Even though we cannot extract enough compactness to pass to the limit in the first moment of the curve of maximal slope, we still have enough rigidity to say something about the limit. Let us assume that $\rho(t)$ does not converge to a single measure in $\mathcal{E}_w$, i.e. there exist subsequences $\rho(t_n)$ and $\rho(t_m)$, $t_m, t_n \to \infty$ and measures $\rho_1, \rho_2 \in \mathcal{E}_w$ such that

$$\lim_{n\to\infty} d_\sigma(\rho(t_n), \rho_1) = 0$$
$$\lim_{m\to\infty} d_\sigma(\rho(t_m), \rho_2) = 0.$$

Thus, for any open ball $B_r(x), x \in \mathbb{R}^d$ and $\pi \in O(d)$, we have that

$$\int_{B_r(x)} \mathrm{d}\rho_1 = \lim_{n\to\infty}\int_{B_r(x)} \mathrm{d}\rho(t_n) = \lim_{n\to\infty}\int_{\pi(B_r(x-x_0))} \mathrm{d}\rho(t_n) = \int_{\pi(B_r(x-x_0))} \mathrm{d}\rho_1,$$

and

$$\int_{B_r(x)} \mathrm{d}\rho_2 = \lim_{m\to\infty}\int_{B_r(x)} \mathrm{d}\rho(t_m) = \lim_{m\to\infty}\int_{\pi(B_r(x-x_0))} \mathrm{d}\rho(t_m) = \int_{\pi(B_r(x-x_0))} \mathrm{d}\rho_2.$$

Here, we have used the fact that $\bigcup_{t\geq 0}\{\rho(t)\} \in \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$, the Portmanteau lemma, and the fact $B_r(x)$ and its image under $\pi \in O(d)$ is a continuity set of $\rho_1, \rho_2 \in \mathcal{E}_w$ (since they are not singular measures). Since open balls generate the Borel σ-algebra, it follows that both $\rho_1, \rho_2 \in \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$. Thus, $\rho_1 = \rho_2$. Indeed, they must be translates of each other and if $\rho_2 = \tau_x\rho_1$ for some $x \neq 0$, then it follows that $\rho_2 \in \mathcal{P}_2^{R,x_0+x}$ which would be a contradiction. Thus, we have that

$$\lim_{t\to\infty} d_\sigma(\rho(t), \rho^*) = 0,$$

where $\rho^* \in \mathcal{P}_2^{R,x_0}(\mathbb{R}^d)$ and so

$$\int_{\mathbb{R}^d} (x - x_0)\,\mathrm{d}\rho^* = 0.$$

$\square$

4.5. **Degenerate diffusion equations on $\mathbb{T}^d$.** Our theory applies to parabolic equations of the form

$$\partial_t \rho = \nabla \cdot (\rho \nabla c^* (\nabla F'(\rho))), \quad \rho(0) = \rho_0 \in \mathcal{P}_q(\mathbb{T}^d), \tag{4.7}$$

where $1 < q < \infty$, c^* denotes the Legendre transform of $c(z) = \frac{1}{q}|z|^q$, the usual cost function in the q-Wasserstein metric W_q, and F is any of

$$F(x) = \frac{1}{p-1}x\log x, \quad F(x) = \frac{1}{m(m-1)}x^m, \quad m \geq 1.$$

Here, $1 < p < \infty$ is the conjugate exponent of q. It is by now standard that these choices for F guarantee that the energy functional

$$E(\rho) = \int_{\mathbb{T}^d} F(\rho)\,\mathrm{d}x$$

satisfies the proper, lower bounded, and lower semicontinuity properties of (A1). Restricting to $\mathbb{T}^d$ guarantees the compactness (A2). Included in equation (4.7) are the so-called generalized heat equation, parabolic p-Laplacian equation [KV88], doubly degenerate diffusion equation (see [SV94] for an L^p theory), and the porous medium equation.

Proposition 4.16. *Gradient flow solutions of* (4.7) *exist provided the initial condition is in the domain of* $E, \rho_0 \in D(E)$ *and coincide with q-curves of maximal slope for the functional E with respect to the weak upper gradient*

$$G(\rho) = \left(\int_{\mathbb{T}^d} |\nabla F'(\rho)|^p \rho \, \mathrm{d}x \right)^{\frac{1}{p}},$$

which is l.s.c. In particular, G satisfies (A3).

Proof. We use [AGS08, Theorems 11.1.3 and 11.3.2] to obtain the equivalence and existence of gradient flow solutions to (4.7) and q-curves of maximal slope with respect to the upper gradient $|\partial E|$. It remains to check that G is an upper gradient of E. However, this is true using [AGS08, Theorem 10.4.6] which shows $G = |\partial E|$ since we are working with absolutely continuous measures.

The convexity of the functionals E is given by [AGS08, Proposition 9.3.9] and finally $|\partial E|$ is l.s.c. by [AGS08, Corollay 2.4.10]. Alternatively, G is of the form (LSC) and hence is l.s.c by Theorem B.1. $\square$

Proposition 4.17. *Let $\rho \in AC([0,\infty); \mathcal{P}_q(\mathbb{T}^d))$ be a q-curve of maximal slope of the energy E with respect to the weak upper gradient G, for some initial datum $\rho_0 \in \mathcal{P}_q(\mathbb{T}^d)$ with $E(\rho_0) < \infty$. Then,*

$$\lim_{t \to \infty} W_\infty(\rho(t), \rho_\infty) = 0,$$

where $\rho_\infty := \mathcal{L}_{\mathbb{T}^d}$ is the normalized Lebesgue measure on $\mathbb{T}^d$.

Proof. Since we are considering the bounded domain, $\mathbb{T}^d$, the only stationary state is ρ_∞. This can be seen by finding the unique zero of G together with the convexity of the energy. By the previous discussion and proposition, we have that (A1), (A2), and (A3) are satisfied. An application of Theorem 2.13 gives

$$\lim_{t \to \infty} d_\sigma(\rho(t), \rho_\infty) = 0.$$

Owing to the boundedness of $\mathbb{T}^d$, we can improve this convergence to

$$\lim_{t \to \infty} W_q(\rho(t), \rho_\infty) = 0.$$

Furthermore, by [BJR07, Theorem 1.2], since ρ_∞ has a density which is uniformly lower bounded away from zero, we have

$$\lim_{t \to \infty} W_\infty(\rho(t), \rho_\infty) = 0.$$

$\square$

For more details in a general setting, we refer to [Agu05, Ott96]. For the asymptotic behaviour in the full space $\mathbb{R}^d$, we lose the compactness; there is no mechanism to guarantee sublevel sets of E are compact in $\mathcal{P}$ in our theory. However, by introducing a change of variables as in [Agu03], equation (4.7) can be re-written with a confining potential term, unfortunately breaking the gradient flow structure. There, convergence to an equilibrium solution can be established with exponential rate. We defer an explicit instance of the change of variables to the next example.

4.6. **Porous Medium Equation on $\mathbb{R}^d$.** In this subsection, we study how our general asymptotic theory can be applied to the porous medium equation on $\mathbb{R}^d$

$$\partial_t \rho = \Delta \rho^m, \quad m > 1.$$

It is known in the vast literature surrounding this equation and its variants (cf. [CT00, DPD02, Vó3, Vó7]) that there is a self-similar structure to this equation and in the large time limit tends to a Barenblatt profile. Mathematically, we shall see that seeking a self-similar solution will introduce

an equation that has nicer gradient flow properties for which we can apply our theory.Following Otto [Ott01], let us introduce the following substitution

$$\rho(t, x) = \frac{1}{t^{d\alpha}} \hat{\rho} \left(\log t, \frac{x}{t^{\alpha}} \right), \quad \alpha = \frac{1}{d(m-1)+2}.$$

Labelling the change of variables as

$$\tau = \log t, \quad y = \frac{x}{t^{\alpha}},$$

we obtain the following evolution equation for $\hat{\rho}$

$$\partial_{\tau} \hat{\rho} = \Delta_y \hat{\rho}^m + \alpha \nabla_y \cdot (y \hat{\rho}).$$

This is a gradient flow with energy functional

$$E(\hat{\rho}) = F_m(\hat{\rho}) + \frac{\alpha}{2} \int_{\mathbb{R}^d} |y|^2 \, \mathrm{d}\hat{\rho},$$

where the internal energy F_m is given by

$$F_m(\hat{\rho}) = \frac{1}{m-1} \int_{\mathbb{R}^d} \hat{\rho}^m \, \mathrm{d}x.$$

Arguing similarly to the previous section, this functional and its associated dissipation

$$G^2(\hat{\rho}) = \int_{\mathbb{R}^d} \hat{\rho} \left| \frac{m}{m-1} \nabla \hat{\rho}^{m-1} + \alpha y \right|^2 \, \mathrm{d}y,$$

fulfil (A1), (A2), and (A3). Here, we note that the compactness on $\mathbb{R}^d$ is owed to the confining potential for the second moment of $\hat{\rho}$ in E. Hence, after a self-similar scaling, the porous medium equation fits into our abstract framework. Again, the lower semicontinuity of G is given by Theorem B.1. Before stating the convergence result, we repeat the well known fact

Lemma 4.18 ([Ott01]). *The unique stationary state in the evolution of $\hat{\rho}$ is*

$$\hat{\rho}_{\infty}(y) = \left(C - \frac{\alpha(m-1)}{2m} |y|^2 \right)_+^{\frac{1}{m-1}},$$

where $C > 0$ is a constant such that $\hat{\rho}_{\infty}$ has unit mass.

Sketch proof. We only show here the formal computation that determines $\hat{\rho}_{\infty}$. For the uniqueness of the stationary state, we refer to [Ott01] for more details. The zeroes of $G^2(\hat{\rho})$ are precisely given by

$$0 = \alpha y + \frac{m}{m-1} \nabla \hat{\rho}^{m-1},$$

on the support of $\hat{\rho}$. Integrating out y gives, for some constant C_1,

$$C_1 = \frac{\alpha}{2} |y|^2 + \frac{m}{m-1} \hat{\rho}^{m-1}.$$

The formula is proven noting the positive part has to be taken since these computations are on the support of $\hat{\rho}_{\infty}$. $\qquad \square$

Proposition 4.19. *Let $\hat{\rho} \in AC([0, \infty); \mathcal{P}_2(\mathbb{R}^d))$ be a 2-curve of maximal slope of the energy E with respect to the weak upper gradient G, for some initial datum $\rho_0 \in Z_{E,2}$. Then, for every $0 < \varepsilon < 1$*

$$\lim_{\tau \to \infty} W_{2-\varepsilon}(\hat{\rho}(\tau), \hat{\rho}_{\infty}) = 0.$$

Proof. In the introductory discussion, (A1), (A2), and (A3) are all satisfied so we can apply Theorem 2.13. Furthermore, the previous Lemma 4.18 shows that the set of stationary states is the singleton $\{\hat{\rho}_{\infty}\}$ so that $d_{\sigma}(\hat{\rho}(\tau), \hat{\rho}_{\infty}) \to 0$ as $\tau \to \infty$. We can upgrade the convergence to $W_{2-\varepsilon}$ since the compactness of sublevel sets of the energy $\{E \leq C\}$ holds up to $W_{2-\varepsilon}$; second moment control is built into E. $\qquad \square$

4.7. **Nonlocal interaction equations: a model for consensus convergence.** In this example, we study pure interaction in one dimension which is still quite complicated. One important example arises in the family of Hegselmann-Krause models for opinion dynamics [HK⁺02]. It can be written as the following aggregation equation on $\mathbb{R}$

$$\begin{cases} \partial_t \rho = \partial_x(\rho \partial_x(\psi * \rho)) \\ \rho(0) = \rho_0 \in \mathcal{P}_2(\mathbb{R}) \end{cases}, \tag{4.8}$$

where the interaction kernel $\psi : \mathbb{R} \to \mathbb{R}$ is given as a primitive

$$\psi(x) = \int_{-\infty}^x y\phi(y)\,\mathrm{d}y,$$

for a compactly supported even function $\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$. We refer to [LT04, BV06, BL09, BCL09a, FR10, FR11, CDF⁺11, BCDFP15] for more details on aggregation equations and finite time blowup of L^1-solutions. For simplicity, we will assume ϕ is qualitatively similar but more regular than the examples considered in [GPY17]. We assume that $\phi \in C_c^\infty(\mathbb{R})$ so that the resultant ψ has the following properties;

 (1) $\psi \in C_c^\infty(\mathbb{R})$,
 (2) $\mathrm{supp}(\psi) \subset \mathrm{supp}(\phi)$,
 (3) and ψ is non-positive and even.

In [GPY17], the examples of ϕ considered are all L^∞ which only guarantees almost everywhere Lipschitz continuity of ψ. As mentioned previously, we require the further regularity assumptions for simplicity. The associated energy functional and candidate weak upper gradient are

$$E_{GPY}(\rho) := \frac{1}{2}\int_{\mathbb{R}} [\psi * \rho](x)\,\mathrm{d}\rho(x), \qquad G_{GPY}^2(\rho) := \frac{1}{2}\int_{\mathbb{R}} |\partial_x(\psi*\rho)|^2\,\mathrm{d}\rho(x).$$

Since ψ is C_c^∞, one can check that the regularity and λ-convexity assumptions of [CDF⁺11, Assumptions NL0-3] all hold and there exists a unique 2-curve of maximal slope for (4.8) with G_{GPY} being the upper gradient of E_{GPY}. Moreover, weak solutions of (4.8) are equivalent to the associated curves of maximal slope.

We explain below that all the assumptions of our abstract theory are satisfied. However, in the sequel we verify some Dobrushin type estimates to obtain rates of convergence. We start with the following soft convergence result:

Proposition 4.20. *Let $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}))$ be a 2-curve of maximal slope of the energy E_{GPY} with respect to the weak upper gradient G_{GPY}, for some initial datum $\rho_0 \in Z_{E_{GPY},2}$. Then,*

$$\lim_{t\to\infty} d_\sigma\left(\rho(t), \mathcal{E}_{GPY}\right) = 0\,,$$

where $\mathcal{E}_{GPY} \subseteq \mathcal{P}_2(\mathbb{R})$ is the set of stationary states of the 2-curve of maximal slope associated to E_{GPY} and G_{GPY}.

Proof. The listed properties of ψ guarantee that E_{GPY} satisfies (A1), (indeed it is continuous on $\mathcal{P}_2(\mathbb{R})$). Furthermore, G_{GPY} has an extension to $\mathcal{P}(\mathbb{R})$ which is exactly of the form (LSC). Thus, by Theorem B.1, (A3) also holds. We verify (A2') in the sequel with Proposition 4.22 and then apply Theorem 2.13. □

It is natural to ask if there is a characterisation for the stationary states for (4.8). The situation is complicated even in one dimension. A priori, we can generate an entire family of linear combination of Diracs belonging to the set of stationary states. Suppose the support of ψ is contained in $[-R, R]$, $\{x_i\}_{i \in I} \subset \mathbb{R}$ is a discrete set of points (finite or infinite), and $\{m_i\}_{i \in I} \subset \mathbb{R}_{>0}$ is indexed the same such that

$$\sum_{i\in I} m_i = 1, \quad \sum_{i\in I} m_i x_i^2 < \infty.$$

If we furthermore assume $\inf_{i\neq j\in I} |x_i - x_j| > 2R$, then the combination $\mu = \sum_{i\in I} m_i \delta_{x_i}$ is a stationary state for (4.8) since each particle x_i has zero interaction with different particles x_j. We are not aware if all stationary states to (4.8) are combinations of Diracs described in this way. If instead we assume ψ is analytic (hence, not compactly supported) and satisfies a growth condition, there is a characterisation of stationary states of (4.8) as a finite sum of Diracs [FR10, Proposition 2.2].

Lemma 4.21 (Second moment estimate). *For $\rho_0 \in Z_{E_{GPY},2}$, let $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}))$ be the associated 2-curve of maximal slope for E_{GPY} with respect to the upper gradient G_{GPY}. Then, we have the estimate*

$$\frac{\mathrm{d}}{\mathrm{d}t} \int |x|^2 \, \mathrm{d}\rho_t(x) \leq 0 \,,$$

for all $t \geq 0$.

Proof. Based on [CDF+11], $\rho : t \mapsto \rho_t$ can be thought of as a solution to (4.8) in the sense of distributions. Formally, we test equation (4.8) against $|x|^2/2$ yielding

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \int \frac{|x|^2}{2} \, \mathrm{d}\rho(x) &= -\int x[(\psi') * \rho](x) \, \mathrm{d}\rho(x) \\
&= -\int x \int (x-y)\phi(x-y) \, \mathrm{d}\rho(y) \, \mathrm{d}\rho(x) \\
&= -\frac{1}{2} \iint |x-y|^2 \phi(x-y) \, \mathrm{d}\rho(y) \, \mathrm{d}\rho(x) \leq 0.
\end{aligned}$$

Here, we have used $\psi'(x) = x\phi(x)$, symmetry in swapping variables $x \leftrightarrow y$, and the positivity of ϕ to arrive at the conclusion. To make this argument rigorous, one can use a standard cut-off approximation of $|x|^2/2$. $\qquad\square$

Proposition 4.22 (Compactness of trajectories). *For $\rho_0 \in Z_{E_{GPY},2}$, let $\rho \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}))$ be the associated 2-curve of maximal slope for E_{GPY} with respect to the upper gradient G_{GPY}. We have that $\{\rho_t\}_{t \in [0,\infty)}$ is relatively compact in $\mathcal{P}(\mathbb{R})$ with limits in $\mathcal{P}_2(\mathbb{R})$. In particular, (A2') holds.*

Proof. Our strategy for this proof will be to first show that the family $\{\rho_t\}_{t \in [0,\infty)}$ is tight in $\mathcal{P}(\mathbb{R})$ using the previous Lemma 4.21. The gain of moments will then also come from Lemma 4.21.
Step 1: $\{\rho_t\}_{t \in [0,\infty)}$ is tight in $\mathcal{P}(\mathbb{R})$

Set $E_0 := \int |x|^2 \, \mathrm{d}\rho_0(x) \geq \int |x|^2 \, \mathrm{d}\rho_t(x)$ where the inequality comes from Lemma 4.21. Fix $\varepsilon > 0$ and choose $R > 0$ large enough such that $E_0/R^2 < \varepsilon$. For every $t \geq 0$, we have the following chain of inequalities;

$$\int_{B_R} \mathrm{d}\rho_t(x) = 1 - \int_{\mathbb{R} \setminus B_R} \mathrm{d}\rho_t(x) \geq 1 - \frac{1}{R^2} \int_{\mathbb{R} \setminus B_R} |x|^2 \, \mathrm{d}\rho_t(x) \geq 1 - \frac{E_0}{R^2} \geq 1 - \varepsilon.$$

This establishes tightness in $\mathcal{P}(\mathbb{R})$.
Step 2: Limits belong to $\mathcal{P}_2(\mathbb{R})$

From the previous step, we can find $\bar\rho \in \mathcal{P}(\mathbb{R})$ and a time-divergent sequence $(t_n)_{n \in \mathbb{N}}$ such that

$$\rho_{t_n} \overset{*}{\rightharpoonup} \bar\rho \quad \text{in } \mathcal{P}(\mathbb{R}).$$

We use Fatou's lemma and lemma 4.21 to give

$$\int |x|^2 \, \mathrm{d}\bar\rho(x) \leq \liminf_{n \to \infty} \int |x|^2 \, \mathrm{d}\rho_{t_n}(x) \leq E_0.$$

$\qquad\square$

At this point, we may conclude that our abstract theory applies and use the previous results to deduce convergence towards stationary states (which exist by Corollary 3.2). In fact, we may improve the convergence of Proposition 4.20 to $W_{2-\varepsilon}$ or, in the case of compactly supported initial data, W_∞. We can provide a more concrete description with classical methods.

Proposition 4.23 (Sum of Diracs). *Consider $\phi \in C_c^\infty(\mathbb{R})$ a non-negative and even function described previously but now with the specific conditions*

$$\operatorname{supp}(\phi) = [-R, R], \quad \phi > 0 \text{ on } (-R, R).$$

Let $N \in \mathbb{N}$ be fixed with weights $\omega^i \in [0,1]$ and initial positions $x_0^i \in \mathbb{R}$ such that

$$\sum_{i=1}^N \omega^i = 1, \quad \sup_{i,j=1,\dots,N} |x_0^i - x_0^j| =: D < R.$$

Then, for $\rho_0 = \sum_{i=1}^{N} \omega^i \delta_{x_0^i}$ as an initial condition to equation (4.8), the corresponding solution ρ converges weakly and in 1-Wasserstein exponentially to $\delta_{\sum_{i=1}^{N} \omega^i x_0^i}$ as $t \to \infty$. The rate depends on ϕ and D.

We will prove this result after a few key lemmata.

Lemma 4.24 (Evolution of particles). *Assume the same notation and setting of Proposition 4.23. Then, $\rho_t = \sum_{i=1}^{N} \omega^i \delta_{x_t^i}$ is the solution to (4.8) where each particle evolves according to*

$$\frac{\mathrm{d}}{\mathrm{d}t} x^i = -\sum_{j=1}^{N} \omega^j \psi'(x^i - x^j), \quad x^i(0) = x_0^i. \tag{4.9}$$

Proof. The right-hand side of (4.9) is smooth and compactly supported in each x^i, hence classical Cauchy-Lipschitz theory guarantees existence and uniqueness of solutions for each x^i for all times. For any test function τ, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{-\infty}^{\infty} \tau(x) \,\mathrm{d}\rho_t(x) = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i=1}^{N} \omega^i \tau(x^i) = \sum_{i=1}^{N} \omega^i \tau'(x^i) \frac{\mathrm{d}}{\mathrm{d}t} x^i = -\sum_{i=1}^{N} \omega^i \tau'(x^i) \sum_{j=1}^{N} \omega^j \psi'(x^i - x^j)$$

$$= -\sum_{i=1}^{N} \omega^i \tau'(x^i)(\psi' * \rho_t)(x^i) = -\int_{-\infty}^{\infty} \tau'(x)\partial_x(\psi * \rho_t)(x) \,\mathrm{d}\rho_t(x).$$

This is the weak formulation of (4.8) hence ρ_t defined in the statement of the lemma is a solution. The interaction ψ' has the necessary regularity to apply standard results that can be found in [Gol16] to deduce uniqueness. $\qquad\square$

We turn to studying properties of the dynamical system given in (4.9) since these translate into crucial statements at the level of the PDE (4.8). Using an argument from [MPPD19], we have

Lemma 4.25 (Particles do not collide in finite time). *Assume the same notation and setting of Proposition 4.23. For $i = 1, \ldots, N$ take the unique solutions x^i to the ODE system (4.9). There exists a constant $K > 0$ depending only on ψ such that*

$$|x^i(t) - x^j(t)| \geq |x_0^i - x_0^j| e^{-Kt},$$

for all $i, j = 1, \ldots, N$ and all $t > 0$.

Proof. We study the evolution of $|x^i - x^j|^2$ which yields

$$\frac{\mathrm{d}}{\mathrm{d}t} |x^i - x^j|^2 = 2(x^i - x^j)\left(\frac{\mathrm{d}}{\mathrm{d}t} x^i - \frac{\mathrm{d}}{\mathrm{d}t} x^j\right) = -2(x^i - x^j)\sum_{k=1}^{N} \omega^k[\psi'(x^i - x^k) - \psi'(x^j - x^k)]$$

$$\geq -2|x^i - x^j|\sum_{k=1}^{N} \omega^k|\psi'(x^i - x^k) - \psi'(x^j - x^k)|$$

$$\geq -2|x^i - x^j|^2 \sum_{k=1}^{N} \omega^k \sup_{\xi \in \mathbb{R}} |\psi''(\xi - x^k)|$$

$$\geq -2K|x^i - x^j|^2.$$

The second line comes from Cauchy-Schwarz inequality. The third line makes use of a Mean-Value type inequality. Finally, the fourth line comes from using the smoothness and compact support of ψ to produce the constant K. We conclude after applying Grönwall's inequality on the differential inequality. $\qquad\square$

One important consequence of this lemma is the following corollary.

Corollary 4.26 (Contracting particle cloud). *Assume the same notation and setting of Proposition 4.23. For $i = 1, \ldots, N$ take the unique solutions x^i to the ODE system (4.9). Assume each initial particle is unique so that $x_0^i \neq x_0^j$, for all $i \neq j$. Let k, l denote the indices of the right-most and left-most particles at time zero, respectively. In other words,*

$$x_0^k = \max_{i=1,\ldots,N} x_0^i, \quad x_0^l = \min_{i=1,\ldots,N} x_0^i.$$

Then, these indices still mark the right-most and left-most particles at all future times, meaning

$$x^k(t) = \max_{i=1,\dots,N} x^i(t), \quad x^l(t) = \min_{i=1,\dots,N} x^i(t), \quad \text{for all } t \geq 0.$$

Furthermore, the particle cloud is contracting in the sense that

$$\frac{\mathrm{d}}{\mathrm{d}t} x^k < 0, \quad \frac{\mathrm{d}}{\mathrm{d}t} x^l > 0, \quad \text{for all } t > 0.$$

Proof. From Lemma 4.25, we immediately establish the first part of the result saying

$$x^k(t) = \max_{i=1,\dots,N} x^i(t), \quad x^l(t) = \min_{i=1,\dots,N} x^i(t), \quad \text{for all } t \geq 0.$$

For the second part, we first claim that the particle cloud remains in $(-R, R)$, that is

$$\sup_{i,j=1,\dots,N} |x^i(t) - x^j(t)| < R, \quad \text{for all } t \geq 0. \tag{4.10}$$

We prove this claim by re-writing the particle cloud diameter in terms of $x^k - x^l$, indeed Lemma 4.25 guarantees

$$\sup_{i,j=1,\dots,N} |x^i(t) - x^j(t)| = x^k(t) - x^l(t).$$

Now, we study the evolution of the right-hand side.

$$\frac{\mathrm{d}}{\mathrm{d}t}(x^k - x^l) = -\sum_{j=1}^{N} \omega^j (\psi'(x^k - x^j) - \psi'(x^l - x^j)) \leq 0.$$

The inequality comes from the following. By definition, $x^k - x^j > 0$ and $x^l - x^j < 0$ for every index $j = 1, \dots, N$. Recalling that $\phi \geq 0$ and $\psi'(x) = x\phi(x)$, we have that

$$\psi'(x^l - x^j) \leq 0 \leq \psi'(x^k - x^j), \quad \text{for all } j = 1, \dots, N.$$

Therefore, the claimed inequality (4.10) holds since $\sup_{i,j=1,\dots,N} |x_0^i - x_0^j| = D < R$.

We return to $\frac{\mathrm{d}}{\mathrm{d}t} x^k$ and remark that the following computations are easily extended to $\frac{\mathrm{d}}{\mathrm{d}t} x^l$. Observe

$$\frac{\mathrm{d}}{\mathrm{d}t} x^k = -\sum_{j=1}^{N} \omega^j \psi'(x^k - x^j) = -\sum_{j=1}^{N} \omega^j (x^k - x^j)\phi(x^k - x^j) < 0, \quad \text{for all } t \geq 0.$$

Here, we once again use the property that $x^k - x^j > 0$ for all times $t \geq 0$. Furthermore, the claimed inequality (4.10) guarantees $\phi(x^k - x^j) > 0$ by our assumption on ϕ in Proposition 4.23. $\square$

We are now in a position to prove Proposition 4.23.

Proof of Proposition 4.23. The proof is broken down into a few steps. To fix ideas, we will take the x^i satisfying (4.9) from Lemma 4.24 and the indices k, l denoting the right-most and left-most particles, respectively. Recall that these indices are well-defined and persist for all time by Corollary 4.26. This strategy is similar to the argument in [CDF+11].

(1) Establish the stationarity of the centre of mass $\sum_{j=1}^{N} \omega^j x_t^j = \sum_{j=1}^{N} \omega^j x_0^j$.
(2) Show that each x^i converges with exponential rate to the centre of mass $\sum_{j=1}^{N} \omega^j x_0^j$.
(3) Convert the convergence at the particle level to convergence at the PDE level.

Step 1: We compute

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\sum_{j=1}^{N} \omega^j x_t^j \right) = -\sum_{j=1}^{N} \omega^j \sum_{i=1}^{N} \omega^i \psi'(x^j - x^i)$$

$$= -\frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{N} \omega^j \omega^i \psi'(x^j - x^i) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega^i \omega^j \psi'(x^i - x^j)$$

$$= -\frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{N} \omega^j \omega^i \psi'(x^j - x^i) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega^i \omega^j \psi'(x^j - x^i) = 0.$$

Here we recall that ψ' is odd since $\psi'(x) = x\phi(x)$ and ϕ is assumed to be even.

Step 2: It suffices to consider the convergence of the right-most particle x^k. The same computation will give the convergence of the left-most particle x^l. We compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\left|x^k - \sum_{j=1}^N \omega^j x_0^j\right|^2\right) = -\left(x^k - \sum_{j=1}^N \omega^j x^j\right)\sum_{m=1}^N \omega^m \psi'(x^k - x^m)$$

$$= -\left(\sum_{j=1}^N \omega^j(x^k - x^j)\right)\sum_{m=1}^N \omega^m(x^k - x^m)\phi(x^k - x^m)$$

$$\leq -\delta\frac{1}{2}\left(\sum_{j=1}^N \omega^j(x^k - x^j)\right)^2 = -\delta\frac{1}{2}\left|x^k - \sum_{j=1}^N \omega^j x_0^j\right|^2.$$

We explain the appearance of $\delta > 0$ in the last line. We assumed that the initial particle cloud has diameter strictly less than R which leads to the estimate, using the contracting property of Corollary 4.26

$$\sup_{m=1,\dots,N}|x^k - x^m| \leq \sup_{m=1,\dots,N}|x_0^k - x_0^m| \leq \sup_{i,j=1,\dots,N}|x_0^i - x_0^j| = D < R.$$

Hence, we have a lower bound of $\phi(x^k - x^m)$ for any $m = 1, \dots, N$ by

$$\inf_{m=1,\dots,N}\phi(x^k - x^m) \geq \inf_{m=1,\dots,N}\phi(x_0^k - x_0^m) \geq \inf_{x\in[-D,D]}\phi(x) =: \delta/2 > 0.$$

We are guaranteed that $\delta > 0$ by assumption that $\phi > 0$ on $(-R, R)$ and the regularity of ϕ. We conclude by Grönwall's inequality

$$\left|x^k - \sum_{j=1}^N \omega^j x_0^j\right|^2 \leq \left|x_0^k - \sum_{j=1}^N \omega^j x_0^j\right|^2 e^{-\delta t}.$$

Step 3: We take the convergence result of Step 2 at the level of the characteristic system (4.9) and interpret this in the context of the original PDE (4.8). We will use the well-known dual form of the 1-Wasserstein distance. Fix a Lipscthitz test function τ with $|\tau'| \leq 1$. By Lemma 4.24, we test the solution $\rho_t = \sum_{i=1}^N \omega^i \delta_{x_t^i}$ against τ

$$\left|\langle\rho_t, \tau\rangle - \left\langle\delta_{\sum_{j=1}^N \omega^j x_0^j}, \tau\right\rangle\right| = \left|\sum_{i=1}^N \omega^i\left[\tau(x^i) - \tau\left(\sum_{j=1}^N \omega^j x_0^j\right)\right]\right|$$

$$\leq \sum_{i=1}^N \omega^i\left|\tau(x^i) - \tau\left(\sum_{j=1}^N \omega^j x_0^j\right)\right| \leq \sum_{i=1}^N \omega^i\left|x^i - \sum_{j=1}^N \omega^j x_0^j\right||\tau'(\xi^i)|$$

$$\leq \sum_{i=1}^N \omega^i\left|x_0^i - \sum_{j=1}^N \omega^j x_0^j\right|e^{-(\delta/2)t} \lesssim e^{-(\delta/2)t}.$$

In the second line, we have used the Mean-Value inequality where ξ^i is some number between x^i and $\sum_{j=1}^N \omega^j x_0^j$ as well as the Lipschitz assumption of τ. This shows the exponential convergence as $t \to \infty$ in 1-Wasserstein. $\qquad\square$

By well-posedness theory with probability measures as initial data [CDF$^+$11] or a standard mean-field limit Dobrushin's argument in [Gol16], the previous results can be extended to general probability measures.

Proposition 4.27. *Consider $\phi \in C_c^\infty(\mathbb{R})$ satisfying the same assumptions as in Proposition 4.23. Take $\rho_0 \in \mathcal{P}(\mathbb{R})$ such that*

$$\mathrm{diam}(\mathrm{supp}(\rho_0)) \leq D < R,$$

with mean $m = \int_{-\infty}^\infty x\,\mathrm{d}\rho_0(x)$. Then, there is a unique weak solution $\rho \in C([0,\infty); \mathcal{P}(\mathbb{R}))$ to equation (4.8) with initial condition ρ_0. Furthermore, there is exponential convergence in 1-Wasserstein

to consensus;

$$\rho \to \delta_m, \quad t \to \infty.$$

4.8. **Application to a case of non-unique curves of maximal slope.** The purpose of this example is to emphasise the dependence of the convergence results and the ω-limit sets to the initial data and the curves of maximal slope emanating from it. Our definition of stationary states is independent of initial data and solutions, however in Theorems 2.13 and 2.14, it is possible that different curves of maximal slope from the same initial data might be converging to disjoint subsets of the set of stationary states. Consider the following interaction potential

$$W(x) = \begin{cases} \operatorname{sign}(|x| - 1)||x| - 1|^{3/2} & |x| > \frac{3}{4} \\ \dfrac{|x|^2}{2} - \dfrac{5}{32} & |x| \le \frac{3}{4} \end{cases},$$

along with the associated aggregation equation given by

$$\partial_t \rho = \partial_x \left(\rho \partial_x W * \rho \right).$$

Note that $W \in C^1(\mathbb{R})$ (see (4.11)) and $W(x) \le C\left(1 + |x|^2\right)$ but it is not λ-convex for any $\lambda \in \mathbb{R}$. Thus, it is not a *pointy potential* or a variation of it in the sense of [CDF$^+$11, CLM14]. We search for solutions $\rho \in AC([0, \infty); \mathcal{P}_2(\mathbb{R}))$ that are 2-curves of maximal slope of the energy $E : \mathcal{P}_2(\mathbb{R}) \to (-\infty, +\infty]$ given by

$$E(\mu) = \frac{1}{2} \iint_{\mathbb{R} \times \mathbb{R}} W(x - y)\mu(\mathrm{d}x)\mu(\mathrm{d}y),$$

with respect to the upper gradient $G : \mathcal{P}_2(\mathbb{R}) \to [0, \infty]$ given by

$$G(\mu) = \left(\int_{\mathbb{R}} |\partial_x W * \mu|^2 \mu(\mathrm{d}x) \right)^{1/2},$$

for some initial data $\rho(0) = \mu_0 \in \mathcal{P}_2(\mathbb{R})$.

Based on the forms of E and G, it is easy to check that (B1) and (B3) are true. For simplicity, we will not address (B2) in full generality. In the setting of Proposition 4.30, the initial datum is symmetric on $\mathbb{R}$. Since the interaction potential, W, is also symmetric, any gradient flow solution is also symmetric. Therefore, we verify (B2') with $\mathcal{Y}$ being the set of all symmetric probability measures.

Lemma 4.28. *The sublevel set $L_{\le C}(E) = \{\mu \in \mathcal{P}_2(\mathbb{R}) \cap \mathcal{Y} : E(\mu) \le C\}$ is relatively compact in $\mathcal{P}(\mathbb{R})$ where $\mathcal{Y}$ is the set of symmetric probability measures on $\mathbb{R}$.*

Proof. The strategy for this proof is similar to Proposition 4.14 borrowing methods from [CCV15, Theorem 2.1] or [McC94]. We repeatedly make use of the fact that W is an increasing function of $|x|$. We define first the reduced energy

$$E_1(\mu) := \frac{1}{2} \int_{-\infty}^{\infty} \int_{|x-y| \ge 1} W(x - y)\mu(\mathrm{d}x)\mu(\mathrm{d}y).$$

If $E(\mu) \le C$ for $C \in \mathbb{R}$, then we can also upper bound $E_1(\mu)$ using $W \ge -\frac{5}{32}$

$$E_1(\mu) = E(\mu) - \frac{1}{2} \int_{-\infty}^{\infty} \int_{|x-y| < 1} W(x - y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \le C + \frac{5}{64} \int_{-\infty}^{\infty} \int_{|x-y| < 1} \mu(\mathrm{d}x)\mu(\mathrm{d}y)$$

$$\le C + \frac{5}{64}.$$

For fixed $R \ge 1$ and $x \in \mathbb{R}$ such that $|x| \ge 1$, we have

$$\{y \in \mathbb{R} : xy \le 0\} \subset \{y \in \mathbb{R} : |x - y| \ge 1\}.$$

This provides a lower bound for $E_1(\mu)$ by

$$E_1(\mu) \ge \int_{|x|>R} \int_{xy \le 0} W(x - y)\mu(\mathrm{d}x)\mu(\mathrm{d}y).$$

In fact, $xy \le 0$ implies the better estimate $|x - y| \ge |x|$ which minorises W

$$E_1(\mu) \ge \int_{|x|>R} \int_{xy \le 0} W(x)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \ge W(R) \int_{|x|>R} \int_{xy \le 0} \mu(\mathrm{d}x)\mu(\mathrm{d}y) = \frac{W(R)}{2} \int_{|x|>R} \mu(\mathrm{d}x).$$

The last equality uses the symmetry of μ; $\{y \in \mathbb{R} : xy \leq 0\}$ is one of $(-\infty, 0]$ or $[0, +\infty)$, both of which have μ-mass $1/2$. The lower and upper bounds of $E_1(\mu)$ just computed give the estimate

$$\sup_{\mu \in L_{\leq C}(E)} \int_{|x| > R} \mu(\mathrm{d}x) \leq \frac{2(C + \frac{5}{64})}{W(R)} \overset{R \to \infty}{\to} 0.$$

This establishes compactness by Prokhorov's theorem. $\qquad\square$

Lemma 4.28 verifies the remaining assumption in our abstract result.

Proposition 4.29. *Let $\mu \in AC([0, \infty); \mathcal{P}_2(\mathbb{R}))$ be a 2-curve of maximal slope of the energy E with respect to the weak upper gradient G, for initial data $\mu_0 \in Z_E \cap \mathcal{Y}$. Then,*

$$\lim_{t \to \infty} d_{LP}(\mu(t), \mathcal{E}_w) = 0,$$

where $\mathcal{E}_w \subset \mathcal{P}(\mathbb{R})$ is the set of weak stationary states of the 2-curves of maximal slope associated to E and G.

Proof. This is an application of Theorem 2.14 since Lemma 4.28 is precisely (B2'). The other assumptions (B1) and (B3) are easily verified based on the forms of E and G. $\qquad\square$

We emphasise here that our abstract convergence result is between a curve and a set; different curves of maximal slope could converge to disjoint subsets of the set of stationary states. We illustrate this in the proof of Proposition 4.30 below.

Proposition 4.30. *Let $\mu_0 = \frac{1}{2}(\delta_{\frac{1}{2}} + \delta_{-\frac{1}{2}}) \in \mathcal{P}_2(\mathbb{R})$. Then, there exist two distinct curves $\rho, \eta \in AC([0, \infty); \mathcal{P}_2(\mathbb{R}))$ with $\rho(0) = \eta(0) = \mu_0$ such that ρ and η are both curves of maximal slope of the energy E with respect to the upper gradient G. Furthermore,*

$$\lim_{t \to \infty} d_{LP}(\rho(t), \mu_0) = 0 \qquad and \qquad \lim_{t \to \infty} d_{LP}(\eta(t), \delta_0) = 0.$$

Proof. We first compute the gradient of W as follows

$$W'(x) = \begin{cases} \mathrm{sign}(x)\frac{3}{2}||x| - 1|^{1/2} & |x| > \frac{3}{4} \\ x & |x| \leq \frac{3}{4} \end{cases}. \tag{4.11}$$

Note now that $\rho \equiv \mu_0$ is a curve of maximal slope of E with respect to G. Indeed, we can check that $G(\rho) = G(\mu_0) = 0$. Thus, by Lemma 2.9, we have that ρ is a (stationary) curve of maximal slope. To construct η we first analyse the following ODEs

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = -\dfrac{1}{2}W'(x_1 - x_2) \\[2mm] \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = -\dfrac{1}{2}W'(x_2 - x_1) \end{cases}, \tag{4.12}$$

with $x_1(0) = -1/2, x_2(0) = 1/2$. We define $x_3 = x_2 - x_1$ and $x_4 = x_1$, which in turn satisfy the following system of ODEs

$$\begin{cases} \dfrac{\mathrm{d}x_3}{\mathrm{d}t} = -W'(x_3) \\[2mm] \dfrac{\mathrm{d}x_4}{\mathrm{d}t} = \dfrac{1}{2}W'(x_3) \end{cases},$$

with $x_3(0) = 1, x_4(0) = -1/2$. Thus,if we solve for x_3, x_4 (and by extension x_1 and x_2) can be obtained easily. Note that

$$x_3(t) = \begin{cases} 1 - \frac{9}{16}t^2 & 0 \leq t \leq 2/3 \\ \frac{3}{4}e^{-(t-2/3)} & 2/3 < t < \infty \end{cases}$$

is a solution to the ODE. Further calculations tell us that

$$x_1(t) = \begin{cases} -1/2 + \frac{9}{32}t^2 & 0 \leq t \leq 2/3 \\ -\frac{3}{8}e^{-(t-2/3)} & 2/3 < t < \infty \end{cases}$$

$$x_2(t) = \begin{cases} 1/2 - \frac{9}{32}t^2 & 0 \le t \le 2/3 \\ \frac{3}{8}e^{-(t-2/3)} & 2/3 < t < \infty \end{cases}$$

are solutions of (4.12). Define the curve

$$\eta := \frac{1}{2}\big(\delta_{x_1(t)} + \delta_{x_2(t)}\big).$$

We will now argue that $\eta \in AC([0,\infty); \mathcal{P}_2(\mathbb{R}))$ is a curve of maximal slope. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\eta) = \frac{1}{4}\frac{\mathrm{d}}{\mathrm{d}t}W(x_2(t) - x_1(t)) = \frac{1}{4}\frac{\mathrm{d}}{\mathrm{d}t}W(x_3(t)) = -\frac{1}{4}|W'|^2(x_3),$$

where for the last equality we have simply used the fact that x_3 is the Euclidean gradient flow of W. Additionally,

$$\frac{1}{2}G^2(\eta) = \frac{1}{2}\left(\frac{1}{8}|W'|^2(x_2(t) - x_1(t)) + \frac{1}{8}|W'|^2(x_1(t) - x_2(t))\right) = \frac{1}{8}|W'|^2(x_3(t)).$$

Finally, we are left to compute the metric derivative of η.

$$d_2^2(\eta(t), \eta(s)) = \frac{1}{2}\big(|x_2(t) - x_2(s)|^2 + |x_1(t) - x_1(s)|^2\big),$$

for $0 \le s \le t < \infty$ and $|t - s| \ll 1$. The above equality follows by explicitly constructing the transport map and can be simplified using symmetry to

$$d_2^2(\eta(t), \eta(s)) = |x_2(t) - x_2(s)|^2.$$

Dividing by $|t - s|$ and passing to the limit as $t \to s$ we obtain

$$\lim_{t \to s}\frac{d_2(\eta(t), \eta(s))}{|t - s|} = \lim_{t \to s}\frac{|x_2(t) - x_2(s)|}{|t - s|} = \frac{1}{2}|W'|(x_3(s)),$$

from which we obtain that

$$d_2(\eta(t), \eta(s)) = \int_s^t \frac{1}{2}|W'|(x_3(u))\,\mathrm{d}u.$$

Thus, $\frac{1}{2}|W'|(x_3(t))$ is necessarily the metric derivative of η and

$$\frac{1}{2}|\eta'|^2(t) = \frac{1}{8}|W'|^2(x_3(t)).$$

Thus, it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\eta(t)) = -\frac{1}{2}|\eta'|^2(t) - \frac{1}{2}G^2(\eta(t)),$$

from which the result follows. $\qquad\square$

Acknowledgements

Appendix A. Dynamical systems

In this section, we discuss some preliminaries from the theory of the continuous-time metric dynamical systems. We follow the discussion in [CH98, Chapter 9]. We make slight modifications to deal with the fact that trajectories of the dynamical systems we consider are not necessarily unique. Let (Z, d) be a complete metric space. We start with the definition of a metric dynamical system:

Definition A.1 (Metric dynamical system)**.** A *metric dynamical system* on Z is a family of mappings $\{S_t\}_{t \ge 0}$ on Z such that

(1) $S_t : Z \to 2^Z$ for all $t \ge 0$;

(2) $S_0 = I$ where $I : Z \to 2^Z, z \mapsto \{z\}$;
(3) For all $z_0 \in Z$ there exists a family of trajectories $(t \mapsto z_j(t)) \in C([0, \infty); Z), j \in J$ for some, possibly uncountable, index set J, such that $S_t z = \bigcup_{j \in J} \{z_j(t)\}$.

A *uniquely-defined metric dynamical system* is a continuous selection of the family of mappings $\{S_t\}_{t \geq 0}$ on Z such that

(1) $S_t \in C(Z; Z)$ for all $t \geq 0$;
(2) $S_0 = I$;
(3) $S_{t+s} = S_t \circ S_s$ for all $s, t \geq 0$;
(4) For all $z_0 \in Z$ there exists a trajectory $(t \mapsto z(t)) \in C([0, \infty); Z)$ such that $S_t z_0 = z(t)$.

We now introduce the concept of the ω-limit set associated to a point $z \in Z$.

Definition A.2 (ω-limit set)**.** Let $\{S_t\}_{t \geq 0}$ be a metric dynamical system and $z_0 \in Z$. Then, the set

$$\omega^j(z_0) := \left\{ z^* \in Z : \exists t_n \to \infty, \lim_{n \to \infty} d(z_j(t_n), z^*) = 0 \right\}$$

is called the ω-*limit set of z_0 subordinate to* $j \in J$. If the dynamical system $\{S_t\}_{t \geq 0}$ is uniquely-defined, then we call

$$\omega(z_0) := \left\{ z^* \in Z : \exists t_n \to \infty, \lim_{n \to \infty} d(z(t_n), z^*) = 0 \right\},$$

the ω-*limit set of z_0.*

We then have the following result:

Theorem A.3. *Let $\{S_t\}_{t \geq 0}$ be a metric dynamical system and $z_0 \in Z$. Assume that $\bigcup_{t \geq 0} \{z_j(t)\}$ is relatively compact in Z for some $j \in J$. Then,*

(a) $\omega^j(z_0)$ *is non-empty, compact, and connected.*
(b) $\lim_{t \to \infty} d(z_j(t), \omega^j(z_0)) = 0$, *where*

$$d(z_j(t), \omega^j(z_0)) := \inf_{z^* \in \omega^j(z_0)} d(z_j(t), z^*) = \min_{z^* \in \omega^j(z_0)} d(z_j(t), z^*).$$

Assume $\{S_t\}_{t \geq 0}$ is uniquely-defined and $\bigcup_{t \geq 0} \{z(t)\} = \bigcup_{t \geq 0} \{S_t z_0\}$ is relatively compact. Then, the same results hold for $\omega(z_0)$.

Proof. We note that the proof of Theorem A.3(a) follows from the fact that we can find subsequence $(z_i(t_n))_{n \in \mathbb{N}}$ and $z^* \in Z$ such that

$$\lim_{n \to \infty} d(z_j(t_n), z^*) = 0 .$$

thus, $z^* \in \omega^j(z_0)$ and $\omega^j(z_0) \neq \emptyset$. Furthermore, by definition, we have that

$$\omega^j(z_0) = \bigcap_{s > 0} \overline{\bigcup_{t \geq s} \{z_j(t)\}} .$$

$\omega^j(z_0)$ is a decreasing intersection of compact and connected sets and thus is compact and connected itself.

For the proof of Theorem A.3(b), we assume by contradiction that we can find a sequence $t_n \to \infty$ such that

$$\liminf_{n \to \infty} d(z_j(t_n), \omega^j(z_0)) \geq \delta .$$

By compactness, there exists a subsequence $t_{n_k} \to \infty$ and some z^* such that

$$\lim_{k \to \infty} d(z_j(t_{n_k}), z^*) = 0 .$$

Clearly, $z^* \in \omega^j(z_0)$ which is absurd. Thus, (b) follows. $\qquad \square$

Appendix B. Lower semicontinuity of weak upper gradients

In this section we include a short justification on a general proof strategy for deducing lower semicontinuity of the weak upper gradients that share certain structural features. The abstract setting is as follows: X is $\mathbb{R}^d$ or $\mathbb{T}^d$ space and $G : \mathcal{P}(X) \to (-\infty, +\infty]$ is a proper function with domain

$$D(G) = \{\rho \in \mathcal{P}(X) : G(\rho) < +\infty\}.$$

We assume that G has the following form:

$$G(\rho) := \left(\int_X |v_\rho(x)|^p \, \mathrm{d}\rho(x) \right)^{\frac{1}{p}}, \tag{LSC}$$

for some $1 < p < \infty$ and some Borel measurable $v_\rho : X \to \mathbb{R}^d$ dependent on ρ. We then have the following result:

Theorem B.1. *The function $G : \mathcal{P}(X) \to (-\infty, +\infty]$ is l.s.c.*

Proof. We need to check that G is l.s.c with respect to weak convergence of probability measures, i.e. given $\rho_n \to \rho$ in $\mathcal{P}(X)$ it holds that

$$G(\rho) \le \liminf_{n \to \infty} G(\rho_n). \tag{B.1}$$

Assume first that ρ_n has no subsequence ρ_{n_k} such that $\sup_{n_k \in \mathbb{N}} G(\rho_{n_k}) < \infty$. Then, we have that

$$\liminf_{n \to \infty} G(\rho_n) = +\infty,$$

and thus (B.1) follows trivially. Indeed, assume this is not the case, i.e.

$$\liminf_{n \to \infty} G(\rho_n) < C < \infty.$$

Then, for some $\varepsilon > 0$ we can find an $N_0 < \infty$, such that for all $N \ge N_0$ it holds that

$$\inf_{N \ge N_0} G(\rho_n) \le C + \varepsilon/2.$$

Thus, there exists some $n_1 \ge N$ such that $G(\rho_{n_1}) \le C + \varepsilon$. Now setting $N_1 = n_1 + 1$, we have that

$$\inf_{N \ge N_1} G(\rho_n) \le C + \varepsilon/2.$$

Thus, there exists some $n_2 > n_1$ such that $G(\rho_{n_2}) \le C + \varepsilon$. Proceeding like this, we can construct a subsequence ρ_{n_k} such that $G(\rho_{n_k}) \le C + \varepsilon$, which provides us with a contradiction. Thus, it is sufficient to check lower semicontinuity of G along sequences with at least one bounded subsequence. In fact, we have that

$$\liminf_{n \to \infty} G(\rho_n) = \inf \left\{ \liminf_{n \to \infty} G(\rho_{n_k}) : \sup_{n_k \in \mathbb{N}} G(\rho_{n_k}) < \infty \right\} \tag{B.2}$$

Indeed, we have the bound

$$\liminf_{n \to \infty} G(\rho_n) = \lim_{N \infty} \inf_{n \ge N} G(\rho_n) \le \lim_{N \to \infty} \inf_{n_k \ge N} G(\rho_{n_k}),$$

for an arbitrary bounded subsequence ρ_{n_k}. Taking the infimum the we have the bound in one direction. For the other direction we note that, as before, for all $\varepsilon > 0$ we can find a bounded subsequence ρ_{n_k} such that

$$\liminf_{n_k \to \infty} G(\rho_{n_k}) \le \liminf_{n \to \infty} G(\rho_n) + \varepsilon.$$

Thus, (B.2) follows. For every G-bounded sequence $(\rho_n)_{n \in \mathbb{N}}$ weakly convergent to some $\rho \in \mathcal{P}(X)$, we set $v_n(x) = v_{\rho_n}(x)$ and note that

$$\sup_{n \in \mathbb{N}} \|v_n\|_{L^p(X, \rho_n)} < \infty.$$

Applying [AGS08, Theorem 5.4.4], we have that

$$G(\rho) \le \liminf_{n_k \to \infty} G(\rho_n).$$

Taking the infimum over all bounded subsequences and using (B.2), the result of the theorem follows. $\square$

References

[ABPV+05] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.*, 77:137–185, Apr 2005.

[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.

[Agu03]   M. Agueh. Asymptotic behavior for doubly degenerate parabolic equations. *C. R. Math. Acad. Sci. Paris*, 337(5):331–336, 2003.

[Agu05]   M. Agueh. Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory. *Adv. Differential Equations*, 10(3):309–360, 2005.

[AMTU01]   A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1-2):43–100, 2001.

[BCC08]   A. Blanchet, V. Calvez, and J. A. Carrillo. Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model. *SIAM J. Numer. Anal.*, 46(2):691–721, 2008.

[BCDFP15] G. A. Bonaschi, J. A. Carrillo, M. Di Francesco, and M. A. Peletier. Equivalence of gradient flows and entropy solutions for singular nonlocal interaction equations in 1D. *ESAIM Control Optim. Calc. Var.*, 21(2):414–441, 2015.

[BCL09a]   A. L. Bertozzi, J. A. Carrillo, and T. Laurent. Blow-up in multidimensional aggregation equations with mildly singular interaction kernels. *Nonlinearity*, 22(3):683–710, 2009.

[BCL09b]   A. Blanchet, J. A. Carrillo, and P. Laurençot. Critical mass for a Patlak-Keller-Segel model with degenerate diffusion in higher dimensions. *Calc. Var. Partial Differential Equations*, 35(2):133–168, 2009.

[BCnCD16] A. B. T. Barbaro, J. A. Cañizo, J. A. Carrillo, and P. Degond. Phase transitions in a kinetic flocking model of Cucker-Smale type. *Multiscale Model. Simul.*, 14(3):1063–1088, 2016.

[BJR07]   G. Bouchitté, C. Jimenez, and M. Rajesh. A new L^∞ estimate in optimal mass transport. *Proc. Amer. Math. Soc.*, 135(11):3525–3535, 2007.

[BL09]   A. L. Bertozzi and T. Laurent. The behavior of solutions of multidimensional aggregation equations with mildly singular interaction kernels. *Chin. Ann. Math. Ser. B*, 30(5):463–482, 2009.

[BV06]   M. Bodnar and J. J. L. Velazquez. An integro-differential equation arising as a limit of individual cell-based models. *J. Differential Equations*, 222(2):341–380, 2006.

[CCV15]   J. A. Carrillo, D. Castorina, and B. Volzone. Ground states for diffusion dominated free energies with logarithmic interaction. *SIAM J. Math. Anal.*, 47(1):1–25, 2015.

[CCY19]   J. A. Carrillo, K. Craig, and Y. Yao. Aggregation-diffusion equations: dynamics, asymptotics, and singular limits. In *Active particles, Vol. 2*, Model. Simul. Sci. Eng. Technol., pages 65–108. Birkhäuser/Springer, Cham, 2019.

[CDF+11]   J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev. Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.*, 156(2):229–271, 2011.

[CG19]   J. A. Carrillo and R. S. Gvalani. Phase transitions for nonlinear nonlocal aggregation-diffusion equations. *arXiv e-prints*, Dec 2019.

[CGPS20]   J. A. Carrillo, R. S. Gvalani, G. A. Pavliotis, and A. Schlichting. Long-Time Behaviour and Phase Transitions for the Mckean–Vlasov Equation on the Torus. *Arch. Ration. Mech. Anal.*, 235(1):635–690, 2020.

[CH98]   T. Cazenave and A. Haraux. *An introduction to semilinear evolution equations*, volume 13 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1998. Translated from the 1990 French original by Yvan Martel and revised by the authors.

[CHVY19]   J. A. Carrillo, S. Hittmeir, B. Volzone, and Y. Yao. Nonlinear aggregation-diffusion equations: radial symmetry and long time asymptotics. *Invent. Math.*, 218(3):889–977, 2019.

[CLM14]   J. A. Carrillo, S. Lisini, and E. Mainini. Gradient flows for non-smooth interaction potentials. *Nonlinear Anal.*, 100:122–147, 2014.

[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. *Rev. Mat. 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. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[CP10]   L. Chayes and V. Panferov. The McKean-Vlasov equation in finite volume. *J. Stat. Phys.*, 138(1-3):351–380, 2010.

[CT00]   J. A. Carrillo and G. Toscani. Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J.*, 49(1):113–142, 2000.

[Daw83]   D. A. Dawson. Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Statist. Phys.*, 31(1):29–85, 1983.

[DPD02]   M. Del Pino and J. Dolbeault. Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions. *J. Math. Pures Appl. (9)*, 81(9):847–875, 2002.

[DS10]   S. Daneri and G. Savaré. Lecture Notes on Gradient Flows and Optimal Transport. *arXiv e-prints*, page arXiv:1009.3737, September 2010.

[DYY19] M. G. Delgadino, X. Yan, and Y. Yao. Uniqueness and non-uniqueness of steady states of aggregation-diffusion equations. *arXiv e-prints*, August 2019.

[FR10] K. Fellner and G. Raoul. Stable stationary states of non-local interaction equations. *Math. Models Methods Appl. Sci.*, 20(12):2267–2291, 2010.

[FR11] K. Fellner and G. Raoul. Stability of stationary states of non-local equations with singular interaction potentials. *Math. Comput. Modelling*, 53(7-8):1436–1450, 2011.

[FV18] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems*. Cambridge University Press, Cambridge, 2018. A concrete mathematical introduction.

[Gig12] N. Gigli. Second order analysis on $(\mathscr{P}_2(M), W_2)$. *Mem. Amer. Math. Soc.*, 216(1018):xii+154, 2012.

[Gol16] F. Golse. On the dynamics of large particle systems in the mean field limit. In *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, volume 3 of *Lect. Notes Appl. Math. Mech.*, pages 1–144. Springer, [Cham], 2016.

[GP18] S. N. Gomes and G. A. Pavliotis. Mean field limits for interacting diffusions in a two-scale potential. *J. Nonlinear Sci.*, 28(3):905–941, 2018.

[GPP12] G. Giacomin, K. Pakdaman, and X. Pellegrin. Global attractor and asymptotic dynamics in the Kuramoto model for coupled noisy phase oscillators. *Nonlinearity*, 25(5):1247–1273, 2012.

[GPY17] J. Garnier, G. Papanicolaou, and T.-W. Yang. Consensus convergence with stochastic effects. *Vietnam J. Math.*, 45(1-2):51–75, 2017.

[HK+02] R. Hegselmann, U. Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation*, 5(3), 2002.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[JLJ98] J. Jost and X. Li-Jost. *Calculus of variations*, volume 64 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1998.

[Kur81] Y. Kuramoto. Rhythms and turbulence in populations of chemical oscillators. *Phys. A*, 106(1-2):128–143, 1981. Statphys 14 (Proc. Fourteenth Internat. Conf. Thermodynamics and Statist. Mech., Univ. Alberta, Edmonton, Alta., 1980).

[KV88] S. Kamin and J. L. Vázquez. Fundamental solutions and asymptotic behaviour for the p-Laplacian equation. *Rev. Mat. Iberoamericana*, 4(2):339–354, 1988.

[Li19] X. Li. Flocking: Phase transition and asymptotic behaviour, 2019.

[Lot08] J. Lott. Some geometric calculations on Wasserstein space. *Comm. Math. Phys.*, 277(2):423–437, 2008.

[LT04] H. Li and G. Toscani. Long-time asymptotics of kinetic models of granular flows. *Arch. Ration. Mech. Anal.*, 172(3):407–428, 2004.

[McC94] R. J. McCann. *A convexity theory for interacting gases and equilibrium crystals*. ProQuest LLC, Ann Arbor, MI, 1994. Thesis (Ph.D.)–Princeton University.

[MPPD19] S. McQuade, B. Piccoli, and N. Pouradier Duteil. Social dynamics models with time-varying influence. *Math. Models Methods Appl. Sci.*, 29(4):681–716, 2019.

[Ott96] F. Otto. Doubly degenerate diffusion equations as steepest descent, 1996.

[Ott01] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[Shu20] R. Shu. Equilibration of aggregation-diffusion equations with weak interaction forces. *arXiv e-prints*, March 2020.

[SSK88] H. Sakaguchi, S. Shinomoto, and Y. Kuramoto. Phase transitions and their bifurcation analysis in a large population of active rotators with mean-field coupling. *Progr. Theoret. Phys.*, 79(3):600–607, 1988.

[SV94] G. Savaré and V. Vespri. The asymptotic profile of solutions of a class of doubly nonlinear equations. *Nonlinear Anal.*, 22(12):1553–1565, 1994.

[Tug13] J. Tugaut. Self-stabilizing processes in multi-wells landscape in $\mathbb{R}^d$-convergence. *Stochastic Process. Appl.*, 123(5):1780–1801, 2013.

[Tug14a] J. Tugaut. Phase transitions of McKean-Vlasov processes in double-wells landscape. *Stochastics*, 86(2):257–284, 2014.

[Tug14b] J. Tugaut. Self-stabilizing processes in multi-wells landscape in $\mathbb{R}^d$-invariant probabilities. *J. Theoret. Probab.*, 27(1):57–79, 2014.

[TV00] G. Toscani and C. Villani. On the trend to equilibrium for some dissipative systems with slowly increasing a priori bounds. *J. Statist. Phys.*, 98(5-6):1279–1309, 2000.

[V03] J. L. Vázquez. Asymptotic beahviour for the porous medium equation posed in the whole space. *J. Evol. Equ.*, 3(1):67–118, 2003. Dedicated to Philippe Bénilan.

[V07] J. L. Vázquez. *The porous medium equation*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford, 2007. Mathematical theory.

[Vil03] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[WE20] S. Wojtowytsch and W. E. Can Shallow Neural Networks Beat the Curse of Dimensionality? A mean field training perspective. *arXiv e-prints*, page arXiv:2005.10815, May 2020.

[Woj20] S. Wojtowytsch. On the Convergence of Gradient Descent Training for Two-layer ReLU-networks in the Mean Field Regime. *arXiv e-prints*, May 2020.