Carnegie Mellon University

MELLON COLLEGE OF SCIENCE

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy in Mathematical Sciences

TITLE Nonlocal Interaction Equations in Heterogeneous and Non-convex Environments

PRESENTED BY Lijiang Wu

ACCEPTED BY THE DEPARTMENT OF Mathematical Sciences

David Kinderlehrer
Dejan Slepcev — MAJOR PROFESSOR — March 17, 2015 — DATE

Thomas Bohman — DEPARTMENT HEAD — March 17, 2015 — DATE

APPROVED BY THE COLLEGE COUNCIL

Frederick Gilman — DEAN — March 17, 2015 — DATE

Nonlocal interaction equations in heterogeneous and non-convex environments

by

Lijiang Wu

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
Department of Mathematical Sciences
Pittsburgh, Pennsylvania

Advised by Professors David Kinderlehrer and Dejan Slepčev

Acknowledgments

I wish to take the opportunity to express my deep gratitude to my advisors Professor David Kinderlehrer and Professor Dejan Slepčev for their guidance and encouragements. Professor David Kinderlehrer and Professor Dejan Slepčev have always been most supportive during my Ph.D. study and research along the way, this thesis would not be possible without their constant support and guidance.

I am grateful to my other committee members Professor Gautam Iyer and Professor Adrian Tudorascu for their encouragements and insightful comments.

I also want to thank my colleagues and friends, including but not limited to Laura Bufford, Elena Cristina Canepa, Yuan Chen, Zhe Cheng, Jeff Eisenbeis, Yuanyuan Feng, Justin Hausauer, Tetsuya Ishikawa, Zilin Jiang, Yong Li, Guoqing Liu, Jing Liu, Lifeng Liu, Pan Liu, Xin Yang Lu, Yuhui Ouyang, Chris Potter, Yue Pu, Matteo Rinaldi, Daniel Rodriguez Vanegas, Tanawit SaeSue, Robert Simione, Daniel Spector, Nicolas Garcia Trillos, Daren Wang, Jing Wang, Kimberly Weston, Yan Xu, Xiang Xu, Xiaofeng Yu for their help during my stay here at Pittsburgh.

Finally, I would like to thank the Department of Mathematical Sciences at Carnegie Mellon University for creating a comfortable environment of study. In particular, I want to thank Professor William J. Hrusa for being such an understanding and generous graduate advisor.

Special thanks go to my family for their unconditional support and care.

Contents

List of Figures

Chapter 1

Introduction

In this Thesis, we present two results on nonlocal interaction equations in heterogeneous environments with boundaries and on general non-convex, non-smooth domains.

Nonlocal interaction equations arise in modeling systems with long range interactions. Such systems serve as basic models to a wide range of phenomena ranging from crystallization [7, 42, 106], statistical mechanics [100, 104], chemotaxis [68, 93], granular media [12, 32, 111], coordinated control [10, 71] to biological aggregation [24, 84, 86].

Biological aggregation describes self-organized behavior of large system of biological agents such as swarming insects and flocking birds. The basic assumptions in modeling biological aggregation are long range attraction, short range repulsion and intermediate range alignment. To be more precise, the pairwise interaction between two agents is attractive when they are far from each other to stay in a social group; they are repulsive to each other when they are too close to avoid collision; in the intermediate distance range, they adjust their velocity to align with other agents. Different models on how long range interaction applies have been proposed: by assuming that velocity of an agent is proportional to the force it is subject to, Bertozzi, Topaz and collaborators studied first-order models [107, 108, 109]; when acceleration of the agent is proportional to the force, second-order (self-propelled and alignment) models such as Vicsek model [113], Cucker-Smale model [37, 38] and Motsch-Tadmor model [88] were introduced. These models have been extensively investigated to predict and study flocking [60, 61, 89, 105] and pattern formation in biological aggregation [8, 30, 69, 116, 117].

The dynamics of such systems is governed by long range interactions among agents, therefore such systems are intrinsically nonlocal and lead to the study nonlocal interaction equations. In this Thesis, we study nonlocal interaction equations in heterogeneous environments with boundaries and on general non-convex, non-smooth domains in the gradient flow framework in the space of probability measures.

Optimal transport and gradient flows in the space of probability measures have provided a novel way to establish well-posedness for a class of dissipative equations. Optimal transport [114, 115] defines a metric on the space of probability measures, namely the Wasserstein metric. In [91], Otto explored the differential structure of the metric and used it to study porous media equations. Gradient flow theory in the space of probability measures was further developed to other types of energies, see the book [4, 5] by Ambrosio, Gigli and Savaré and references therein.

Recently, optimal transport and gradient flow theory in Wasserstein metric space have been extensively used to understand Ricci curvature of metric space [77, 102, 103] and discrete graphs (Markov chains) [46, 47, 58, 78, 83], and to solve many different types of PDEs including Keller-Segel system [19, 20, 21, 22, 23], reaction-diffusion equations [59, 74], thin film and quantum drift equations [55, 76, 79]. In biological aggregation, gradient flow theory unifies discrete (particle) and continuum models, and allows mass concentration (blow-up). Motivated by biological aggregation in heterogeneous environments and on general non-convex, non-smooth domains, we develop gradient flow theory of interaction energy in the space of probabilities on Riemannian manifolds with boundaries and on non-convex, non-smooth domains, which applies to several interesting phenomena.

In this Thesis, we are interested in the first order biological aggregation models in heterogeneous environments and non-convex, non-smooth domains, and the well-posedness of the resulting nonlocal interaction equations. In traditional first order models (where velocity of an agent is proportional to the force it is subject to) which we introduce in more detail in Section 2.3, the settings are: let $x^i \in \mathbb{R}^d, m_i \geq 0$ be the position and mass of the i-th agent with $\sum_{i=1}^N m_i = 1$ (after normalization), W, V be the interaction and external potential functions, the dynamics of the configuration follows

$$\dot{x}^i(t) = -\sum_{j\neq i}^N m_j \nabla W(x^i(t) - x^j(t)) - \nabla V(x^i(t)) \qquad \forall i \in \{1, \dots, N\}. \tag{1.0.1}$$

Agents are autonomous in the system and there is no leader in the group. We are interested in large scale collective behavior (well-defined clusters, sharp boundaries) of the system with large number of agents N. Determining the behavior of the system by tracking each individual would involve solving a large system of mutually dependent ODEs, which is computationally expensive. Thus we consider dynamics of the distribution of agents. Denote the empirical distribution by

$$\mu(t) = \sum_{i=1}^N m_i \delta_{x^i(t)},$$

direct calculations show it satisfies the following nonlocal interaction equation

$$\frac{\partial}{\partial t}\mu(t,x) - \operatorname{div}\left(\mu(t,x)\left(\nabla W * \mu(t)(x) + \nabla V(x)\right)\right) = 0 \tag{1.0.2}$$

in the sense of distributions. Note that (1.0.2) is more general than the system of ODEs (1.0.1) since it includes both discrete and continuum distributions, and is independent of the number of agents N.

Nonlocal interaction equations in heterogeneous environments with boundaries.

In nature, space heterogeneity and boundary should be considered to better model biological aggregations in environments, including factors such as variations in mobility of biological agents depending on the environment and physical boundaries. Interesting phenomena (rolling swarm, boundary concentration) emerge from biological aggregations in heterogeneous environments with boundaries which are not seen in traditional homogenous setting. It suggests that space heterogeneity can be used to explain some phenomena such as rolling swarms in locust swarm observed in nature.

The space heterogeneity makes mobility (factor indicating how agents move in response to the force they are subject to) of an individual agent depend on its location in space, while the interaction with other agents is not affected. That is, the agents can still see (interact) with each other directly, only their ability to move depends on their physical location in the environment.

We also assume, naturally, that agents on the boundary cannot go through the boundary, but they can move along it or reenter interior of the domain. In other words, the set of admissible velocities for the agents on the boundary is the tangent cone (inward going directions) at that boundary point (refer to (3.1.3)). In this case the total mass of the agents is preserved (i.e. the agents are not allowed to leave the space), and this induces a projection of the velocities of the agents on the boundary onto the set of admissible velocities. When considered on a subset $\mathcal{M} \subset \mathbb{R}^d$ (environment), the resulting equation we propose has the form

$$\frac{\partial}{\partial t}\mu(t,x) + \operatorname{div}\left(\mu(t,x)P_x A(x)\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right)\right) = 0, \quad \mu(0) = \mu_0, \tag{1.0.3}$$

where $A(x)$ is the mobility of agents, $\nabla, \operatorname{div}$ are Euclidean gradient and divergence, and $\nabla W * \mu$ is defined as

$$\nabla W * \mu(t)(x) = \int_{\mathcal{M}} \nabla W(x-y)d\mu(t,y).$$

We are interested in existence and stability of the nonlocal interaction equations (1.0.3). In the usual Euclidean setting (i.e. $A(x) \equiv I_d$), well-poseness of weak measure solutions to (1.0.2) was obtained by considering solutions as gradient flows to some energy functional in the space of probability measures endowed with (Euclidean) Wasserstein distance, see [5, 28]. Here we follow the ideas and establish well-posedness of weak measure solutions to (1.0.3) by gradient flows to interaction and potential energy functionals in the space of probability measures endowed with Riemannian Wasserstein distance.

To be more specific, assume that $A(x) > 0$ (i.e. symmetric, positive definite) for all $x \in \mathcal{M}$ and let $G(x) = A^{-1}(x)$ be the inverse matrix of A. We assume that $(\mathcal{M}, g)$ is a complete and geodesically convex Riemannian manifold with C^2 boundary under the metric induced by g, where $g_x(v, v) = v^T G(x)v$ (refer to Subsection 3.1 for exact conditions on $(\mathcal{M}, g)$). We denote the space of probability measures with finite second moments (with respect to the Riemannian distance) by $\mathcal{P}_2(\mathcal{M})$ and the Riemannian Wasserstein distance by d_W (refer to (2.1.1), (2.1.3) where d is the Riemannian distance in this case), that is, for any $\mu, \nu \in \mathcal{P}_2(\mathcal{M})$

$$d_W^2(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \left\{ \int_{\mathcal{M}\times\mathcal{M}} \mathrm{dist}^2(x, y) d\gamma(x, y) \right\}, \tag{1.0.4}$$

where $\mathrm{dist}(x, y)$ is the Riemannain distance induced by g between $x, y \in \mathcal{M}$, and $\Gamma(\mu, \nu)$ is the set of transport plans between μ, ν (i.e. the set of joint probability measures on $\mathcal{M} \times \mathcal{M}$ with first marginal μ and second marginal ν, see (2.1.2)).

For interaction and external potentials W and V, and $\mu \in \mathcal{P}_2(\mathcal{M})$, we define the interaction energy

$$\mathcal{W}(\mu) = \frac{1}{2} \int_{\mathcal{M}\times\mathcal{M}} W(x - y) d\mu(x) d\mu(y), \tag{1.0.5}$$

and potential energy

$$\mathcal{V}(\mu) = \int_{\mathcal{M}} V(x) d\mu(x). \tag{1.0.6}$$

We denote the total energy by

$$\mathcal{E}(\mu) = \mathcal{W}(\mu) + \mathcal{V}(\mu). \tag{1.0.7}$$

We develop suitable concepts of subdifferential and gradient flow in our Riemannian setting (refer to Chapter 3) which generalize the notions in the usual Euclidean framework. We then show well-posedness of weak measure solutions to (1.0.3) by establishing existence and stability of gradient flows to the energy functional $\mathcal{E}$ in $(\mathcal{P}_2(\mathcal{M}), d_W)$ given that W, V are geodesically (semi-)convex (with respect to the Riemannian metrc g). To be precise, we use JKO scheme and lower semicontinuity arguments to show the existence of curve of maximal slope. We carry out subdifferential calculus and approximation schemes to get a chain rule, which combined with curve of maximal slope yields gradient flows. We also show quantitative stability estimates of solutions in d_W by generic differentiability of Riemannian Wasserstein distance and (semi-)convexity of the potentials.

There are several challenges in proving the existence and stability of gradient flows (for the interaction energy in particular) in our Riemannian setting. First, existence of the subdifferential requires (semi-)geodesic convexity of the interaction potential on the product manifold $\mathcal{M} \times \mathcal{M}$ (as opposed to (semi-)convexity of W on $\mathbb{R}^d$ suffices in the Euclidean setting) due to the fact that we cannot directly identify tangent spaces at different points as in the Euclidean case. Second, since projection P breaks the continuity of the subdifferential, we have to find a new proof for the lower semicontinuity of the local slope with respect to narrow convergence of probability measures. Last, the breakdown of continuity of the velocity field (due to projection P) raises problems in showing existence of the flow map associated to the velocity field when we try to prove generic differentiability of d_W. Thus we need to use approximations instead to show stability estimates.

It turns that for mildly heterogeneous environments, even some natural interaction potentials (for example $W(x) = \frac{1}{2}|x|^2$) are not globally geodesically (semi-)convex. However, in many applications from biology, the initial distributions μ_0 of biological agents have compact support. In this case, we develop gradient flow theory that applies to a much wider class of interaction potentials (with only weaker, local conditions imposed). The key observation is that we can control the support of discrete solutions from JKO scheme and then show that the limit curve also has compact support for any fixed time, thus local, weaker conditions on W still imply well-posedness.

Nonlocal interaction equations on non-convex, non-smooth domains.

In reality environments have obstacles (such as rivers or mountains) and irregular boundaries. Thus neither convexity nor smoothness of the domain is guaranteed. Therefore, we consider biological aggregations on non-convex domains $\Omega \subset \mathbb{R}^d$ with low regularity (in the Euclidean setting i.e., $A(x) \equiv I_d$ the identity matrix).

The space $\mathcal{P}(\Omega)$ of probability measures on Ω endowed with the Euclidean Wasserstein distance is not geodesically convex, thus general existence of gradient flow theory [5] fails to apply. We instead obtain gradient flows via particle approximations. That is, we approximate the initial data μ_0 by a sequence of sums of Dirac measures (particle measures) $\mu_0^n = \sum_{i=1}^{k(n)} m_i^n \delta_{x_i^n}$ in Wasserstein distance d_W and solve the resulting systems of ODEs for solutions $\mu^n(\,\cdot\,)$ with initial data μ_0^n. We then show that the solutions $\mu^n(\,\cdot\,)$ satisfy quantitative stability estimates (with respect to d_W) and thus the sequence of solutions $\mu^n(\,\cdot\,)$ converges to a limit curve $\mu(\,\cdot\,)$. The goal is to show that $\mu(\,\cdot\,)$ is a weak measure solution to (1.0.3) with $A \equiv I_d$. However, we can not directly take limit in the weak formulation of (1.0.3) again because the projection P breaks the continuity of the velocity fields, and weak convergence of $\mu^n(\,\cdot\,)$ to $\mu(\,\cdot\,)$ is not sufficient for us to take limit in the weak formulation, see Remark 5.3.2. We instead prove $\mu(\,\cdot\,)$ is a solution to (1.0.3) (with $A(x) \equiv I_d$) by establishing that it satisfies the steepest descent property with respect to $\mathcal{E}$. It turns out the notion of domain prox-regularity from the theory of non-convex sweeping process [43, 44, 112] plays a key role in showing both the well-posedness of the system of ODEs (with discontinuous velocity field due to projection) from the particle approximation, and quantitative stability estimates of solutions $\mu^n(\cdot)$ in d_W.

A closed set $\Omega \subset \mathbb{R}^d$ is η-prox-regular if any point in the η-neighborhood of Ω has unique projection onto it. Prox-regularity is an important concept in non-convex analysis; refer to [35, 43, 44, 95, 112] and references therein for details. The first advantage of Ω being prox-regular is that even though Ω is not smooth, we can still define the tangent cone $T(\Omega, x)$ (inward directions) and normal cone $N(\Omega, x)$ (outward directions) which then enables us to define the projection maps P_x in (1.0.3) as projection onto the tangent cone $T(\Omega, x)$, see Figure 5.1 in Chapter 5. Moreover, η-prox-regularity ensures well-posedness of sweeping process on Ω (5.1.17), which we show gives solutions to the ODE systems from particle approximation; the defining property (5.1.4) of prox-regularity yields quantitative stability property of solutions to the ODE systems (with stability constant depending explicitly on η) with respect to Wasserstein distance d_W, see (5.1.8).

Outline

This Thesis is organized as follows:

In Chapter 2, we introduce the background knowledge about optimal transport, gradient flow theory and biological aggregations. We present the Monge's problem and Kantorovich's relaxed formulation, which lead to Wasserstein distance on the space of probability measures. We then show the general gradient flow theory on the space of probability measures via (formal) Otto Calculus and rigorous Subdifferential Calculus. In the last Section, we give the basic models in biological aggregations and some known results in the resulting nonlocal interaction equations.

In Chapter 3, we study nonlocal interaction equations in heterogeneous environments with boundaries. We show that by modeling the heterogeneous environments as Riemannian manifolds $\mathcal{M}$ with boundaries, we are solving nonlocal interaction equations (1.0.3) on Riemannian manifolds with boundaries. We give suitable generalization of Subdifferential Calculus (and thus gradient flows) in our Riemannian setting such that solutions to the desired nonlocal interaction equations are gradient flows to the total energy $\mathcal{E}$ (1.0.7) with respect to the Riemannian Wasserstein distance d_W (1.0.4). We then show the existence and stability of gradient flows (thus also well-posedness of the nonlocal interaction equations) given geodesic (semi-)convexity of interaction and external potentials W, V. We also present some numerical simulations showing that rolling swarms emerge naturally in biological aggregations in heterogeneous environments with boudaries.

In Chapter 4, we show the well-posedness of nonlocal interaction equations (1.0.3) in heterogeneous environments with boundaries given that initial data μ_0 has compactly support. In particular, we relax the strong, global conditions of geodesic (semi-)convexity of interaction and external potentials to weak, local condition on the potentials. We control the support of discrete (approximating) solutions from JKO scheme and show that they have at most exponential growth. Thus a concept of local geodesic convexity of potentials (which can be implied by the weak, local conditions) suffices to ensure the well-posedness of the nonlocal interaction equations (1.0.3).

In Chapter 5, we investigate nonlocal interaction equations on general non-convex, non-smooth domains Ω. Due to the non-convexity of the domain (thus non-geodesic-convexity of the space of probability measures on the domain), the general existence of gradient flow arguments fail to apply. We instead use particle approximations, that is, we approximate the initial data by sum of delta measures (particles) and show the well-posedness of the resulting system of ODEs given that the domain is prox-regular. We then establish the quantitative stability estimates of solutions in Wasserstein distance (explicitly involving convexity constants of potentials and prox-regularity constant of the domain), thus the sequence of solutions converges to a limit curve. We obtain that the limit curve is a gradient flow by showing that it satisfies the steepest descent property, and thus it is a solution to the nonlocal interaction equation we start with (i.e. it satisfies the nonlocal interaction equation (1.0.3) with the desired initial data μ_0). We show well-posedness of nonlocal interaction equations on Ω in three different settings: Ω bounded and η-prox-regular, Ω unbounded and convex (i.e. ∞-prox-regular), and Ω unbounded, η-prox-regular with initial data μ_0 having compact support.

Chapter 2

Background

In this Chapter, we give the background knowledge needed for later chapters. In particular, we first introduce the Wasserstein distance on the space of probability measures via optimal transport. We then recall the results of gradient flow theory on the space of probability measures developed in [5] by Ambrosio, Gigli and Savaré, and the Jordan-Kinderlehrer-Otto (minimizing movement) scheme introduced in [65] by Jordan, Kinderlehrer and Otto to obtain gradient flows. We finish the Chapter by reviewing different models and results in biological aggregation.

2.1 Optimal transport and Wasserstein distance

In this Section, we give the definition of Wasserstein distance between Borel probability measures on a metric space. For general introduction of optimal transport theory we refer to [5, 114].

The optimal transport theory was introduced in 1781 by Monge in [85], with its mathematical formulation, referred as *Monge formulation*, given in the following way on Polish spaces (i.e. complete, separable metric spaces):

Problem 2.1.1. *Given Polish spaces (X, μ), (Y, ν), with μ, ν probability measures, and a cost function $c : X \times Y \longrightarrow [0, \infty]$ define*

$$\mathcal{T}(\mu, \nu) := \{f : X \longrightarrow Y \ \textit{Borel} : f_\sharp \mu = \nu\}$$

and consider the minimization problem

$$\min_{T \in \mathcal{T}(\mu,\nu)} \int_X c(x, T(x)) d\mu(x).$$

Here for a function $f : X \longrightarrow Y$, $f_\sharp \mu$ is the push forward of μ by f, which is a Borel probability measure on Y defined as $f_\sharp \mu(A) = \mu(f^{-1}(A))$ for all Borel measure sets $A \subset Y$.

For the original formulation proposed by Monge in [85], data were $X = Y = \mathbb{R}^d$, $c(x,y) := |x-y|$. Elements of $\mathcal{T}$ are often referred as *transport maps*, between μ and ν. When a transport map realizes the minimization problem, we call it an optimal map between μ and ν, and denote it by t_μ^ν. This formulation presents several undesirable problems:

- $\mathcal{T}(\mu,\nu) \neq \emptyset$ it is not guaranteed: a very easy example is $X = Y := \mathbb{R}$, $c(x,y) := |x-y|$, $\mu := \delta_0$, $\nu := \dfrac{\delta_{-1}+\delta_1}{2}$;
- minimizer may not exist, i.e. (2.1.1) can admit no minima: an easy counterexample is $X = Y := B((0,0),1) \backslash \{(0,0)\} \subset \mathbb{R}^2$, $\mu = \delta_{(1/2,0)}$, $\nu := \delta_{(-1/2,0)}$;
- condition $f_\sharp \mu = \nu$ is not weakly sequentially closed: a counterexample is $T_n : \mathbb{R} \longrightarrow \mathbb{R}$, $T_n(x) := T(nx)$ with $T : \mathbb{R} \longrightarrow \mathbb{R}$ a 1-periodic function equal to 1 on $[0,1/2)$ and -1 on $[1/2,1)$, $\mu := \mathcal{L}_{|[0,1]}$, $\nu := \frac{1}{2}(\delta_{-1}+\delta_1)$. For every n equality $T_{n\sharp}\mu = \nu$ is true, but passing to the limit this becomes $\mathbb{O}_\sharp \mu = \nu$ ($\mathbb{O}$ denoting the null function on $\mathbb{R}$), clearly false.

A way to overcome these difficulties is provided by the Kantorovich formulation, proposed in [66, 67]:

Problem 2.1.2. *Given Polish spaces (X,μ), (Y,ν), with μ,ν probability measures, and a cost function $c : X \times Y \longrightarrow [0,\infty]$, define $\Gamma(\mu,\nu) := \{\gamma \in \mathcal{M}(X \times Y) : \pi_{X\sharp}\gamma = \mu, \pi_{Y\sharp}\gamma = \nu\}$ where $\mathcal{M}(X \times Y)$ denotes the set of probability measures on $X \times Y$, $\pi_X : X \times Y \longrightarrow X$ and $\pi_Y : X \times Y \longrightarrow Y$ the natural projections, and consider the minimization problem*

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x,y) d\gamma(x,y).$$

Elements of $\Gamma(\mu,\nu)$ are often referred as *"transport plans"*, they have first marginal μ and second marginal ν. This formulation provides several advantages over formulation 2.1.1:

- $\Gamma(\mu,\nu) \ni \mu \times \nu$, while $\mathcal{T}(\mu,\nu)$ can be empty,
- there exists a natural injection $i : \mathcal{T}(\mu,\nu) \longrightarrow \Gamma(\mu,\nu)$ defined as
$$i(T) = (id \times T)_\sharp \mu,$$
- $\Gamma(\mu,\nu)$ is convex and compact with respect to the narrow convergence, and
$$\xi \mapsto \int_{X \times Y} c(x,y) d\xi(x,y)$$
is linear,

- as proven in [3, 54, 97], under some additional conditions the infimum of Monge problem is equal to the minimum of Kantorovich problem.

In this Thesis, we only need the special case $X = Y$ a Polish space and $c(x, y) = d^2(x, y)$, where d is the metric on X. Let $\mathcal{P}(X)$ be the space of Borel probability measures on X. Denote by $\mathcal{P}_2(X)$ the space of Borel probability measures with finite 2-moment, i.e.

$$\mathcal{P}_2(X) = \left\{ \mu \in \mathcal{P}(X) : \int_X d^2(x, x_0) d\mu(x) < \infty \right\}, \tag{2.1.1}$$

where $x_0 \in X$ is an arbitrary point on X.

Given $\mu, \nu \in \mathcal{P}(X)$ we define $\Gamma(\mu, \nu)$ as the set of joint distributions on $X \times X$ with first marginal μ, second marginal ν, i.e.

$$\Gamma(\mu, \nu) = \left\{ \gamma \in \mathcal{P}(X \times X) : (\pi_1)_\sharp \gamma = \mu, (\pi_2)_\sharp \gamma = \nu \right\}, \tag{2.1.2}$$

where $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$ are projection operators onto the first and second coordinates, and $(\pi_1)_\sharp \gamma, (\pi_2)_\sharp \gamma$ are push forward of γ by π_1, π_2.

The 2-Wasserstein distance d_W between $\mu, \nu \in \mathcal{P}_2(X)$ is defined as the minimum from Problem 2.1.2 as

$$d_W^2(\mu, \nu) = \min \left\{ \int_{X \times X} d^2(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \right\}. \tag{2.1.3}$$

The existence of the minimum is a direct consequence of direct method in calculus of variations. We denote

$$\Gamma_o(\mu, \nu) = \left\{ \gamma \in \Gamma(\mu, \nu) : \int_{X \times X} d^2(x, y) d\gamma(x, y) = d_W^2(\mu, \nu) \right\} \tag{2.1.4}$$

the set of optimal plans between μ and ν.

Here we recall that, give a sequence $\mu_n \in \mathcal{P}_2(X)$ and $\mu \in \mathcal{P}_2(X)$,

$$\lim_{n \to \infty} d_W(\mu_n, \mu) = 0 \Longleftrightarrow \begin{cases} \mu_n \text{ converges narrowly to } \mu, \\ \lim_{n \to \infty} \int_X d^2(x, x_0) d\mu_n(x) = \int_X d^2(x, x_0) d\mu(x). \end{cases} \tag{2.1.5}$$

Here μ_n converges narrowly to μ if for any bounded, continuous real function f on X,

$$\lim_{n \to \infty} \int_X f(x) d\mu_n(x) = \int_X f(x) d\mu(x).$$

Furthermore, the space $(\mathcal{P}_2(X), d_W)$ is complete and separable. Finally, $\mathcal{K} \subset \mathcal{P}_2(X)$ is relatively compact with respect to the topology induced by d_W if and only if it is tight and 2-uniformly integrable. Refer to Theorem 2.7 from [4] for the detailed proof.

Another interpretation of Wasserstein distance d_W on $\mathcal{P}_2(\mathbb{R}^d)$ is the Benamou-Brenier formulation introduced by Benamou and Brenier in [11]. It is given as the following dynamic Riemannian-like formula: for $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$

$$d_W^2(\mu_0, \mu_1) = \inf_{(\rho,v)\in\mathcal{A}(\rho_0,\rho_1)} \left\{ \int_0^1 \int_{\mathbb{R}^d} |v(t,x)|^2 d\mu(t,x) dt \right\}, \tag{2.1.6}$$

where $(\mu, v) \in \mathcal{A}(\rho_0, \rho_1)$ if $\mu : [0,1] \to \mathcal{P}(\mathbb{R}^d)$ with $\mu(0) = \mu_0, \mu(1) = \mu_1$, and Borel vector field $v : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfy

$$\frac{\partial}{\partial t}\mu(t,x) + \mathrm{div}(\mu(t,x)v(t,x)) = 0,$$

in the sense of distributions. The Benamou-Brenier formulation actually suggests that $(\mathcal{P}_2(\mathbb{R}^d), d_W)$ has a formal infinite dimensional Riemannian manifold structure with Riemannian metric (inner product) at a fixed $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and tangent vector s given by

$$g_\mu(s,s) = \inf_v \left\{ \int_{\mathbb{R}^d} |v(x)|^2 d\mu(x) \right\},$$

with infimum taken over all v such that $s(x) + \mathrm{div}(\mu(x)v(x)) = 0$. It is direct computation (by calculating the Euler-Lagrange equation for v) to show that the infimum is attained at $v = \nabla\phi$ for some ϕ. Thus

$$g_\mu(s,s) = \int_{\mathbb{R}^d} |\nabla\phi(x)|^2 d\mu(x), \tag{2.1.7}$$

for $s + \mathrm{div}(\mu(x)\nabla\phi(x)) = 0$ and formally

$$\begin{cases} d_W^2(\mu_0,\mu_1) = \inf\left\{ \int_0^1 g_{\mu(t)}\left(\frac{\partial\mu}{\partial t}(t), \frac{\partial\mu}{\partial t}(t)\right) dt;\ \mu(0) = \mu_0, \mu(1) = \mu_1 \right\}, \\ g_{\mu(t)}\left(\frac{\partial\mu}{\partial t}(t), \frac{\partial\mu}{\partial t}(t)\right) = \int_{\mathbb{R}^d} |\nabla\phi(x)|^2 d\mu(t,x),\ \frac{\partial\mu}{\partial t}(t) + \mathrm{div}(\mu(t,x)\nabla\phi(x)) = 0. \end{cases} \tag{2.1.8}$$

2.2 Gradient flow in the space of probability measures

We introduce the notions of curves of maximal slope, and subdifferential of energy functional on the space of probability measures endowed with Wasserstein metric, which lead to the notion of gradient flows. For gradient flow theory in Hilbertian setting we refer to [26] by Brézis; in purely metric setting, we refer to [5] by Ambrosio, Gigli and Savaré. Here we mainly focus on gradient flow theory in the space of probability measures and follow the presentation in [5].

Recall that given a Riemannian manifold M, a point $x_0 \in M$ and a smooth function $F : M \longrightarrow \mathbb{R}$, the gradient flow starting from x_0 is a differentiable curve $x : \mathbb{R}_+ \longrightarrow M$ verifying

$$\begin{cases} x'(t) = -\nabla_M F(x(t)), \\ x(0) = x_0, \end{cases} \tag{2.2.1}$$

where ∇_M is the Riemannian gradient on manifold M. An interpretation of this formulation is that the curve $x(\cdot)$ descends along the steepest descent direction of the function F, i.e. along the opposite direction of the gradient of F. An observation is that an equivalent way of describing $x(\cdot)$ is that: $x(\cdot)$ is a differentiable curve on M starting at x_0, satisfying

$$\frac{d}{dt}F(x(t)) \leq -\frac{1}{2}|\nabla_M F(x(t))|^2 - \frac{1}{2}|x'(t)|^2, \tag{2.2.2}$$

with $|\cdot|$ the norm under the Riemannian metric (inner product) on the tangent space of M. Indeed, (2.2.1) implies (2.2.2) by chain rule; a direct Cauchy-Schwarz argument shows that (2.2.2) gives (2.2.1).

We can view $(\mathcal{P}_2(X), d_W)$ as a metric space and use (2.2.2) to generalize the notion of gradient flows in the pure metic setting (Definition 2.2.2). However, since we know formally $(\mathcal{P}_2(\mathbb{R}^d), d_W)$ has an infinite dimensional Riemannian manifold structure (2.1.7), this can be used to perform the so-called (formal) Otto Calculus [91, 114] and define gradient flows, which we describe briefly now.

By (2.1.7) one can define the scalar product of two tangent vectors $\frac{\partial\mu}{\partial t_1}$ and $\frac{\partial\mu}{\partial t_2}$ at μ:

$$\left\langle \frac{\partial\mu}{\partial t_1}, \frac{\partial\mu}{\partial t_2} \right\rangle_\mu = g_\mu\left(\frac{\partial\mu}{\partial t_1}, \frac{\partial\mu}{\partial t_2}\right) = \int_{\mathbb{R}^d} \langle \nabla\phi_1, \nabla\phi_2\rangle d\mu, \tag{2.2.3}$$

where ϕ_1, ϕ_2 solve

$$\frac{\partial\mu}{\partial t_1} + \operatorname{div}(\mu\nabla\phi_1) = 0, \qquad \frac{\partial\mu}{\partial t_2} + \operatorname{div}(\mu\nabla\phi_2) = 0.$$

For F an energy functional on $\mathcal{P}_2(\mathbb{R}^d)$ and $\frac{\partial\mu}{\partial t}$ a tangent vector, we can define the gradient of F which we denote by $\operatorname{grad}_W F$ via

$$\left\langle \operatorname{grad}_W F(\mu), \frac{\partial\mu}{\partial t} \right\rangle_\mu = DF(\mu)\cdot\frac{\partial\mu}{\partial t}. \tag{2.2.4}$$

Let $\mu(\cdot)$ be a curve in $\mathcal{P}_2(\mathbb{R}^d)$ with $\mu(t) = \mu$ and tangent $\frac{\partial\mu}{\partial t}$ at time t. Assume that $\operatorname{grad}_W F + \operatorname{div}(\mu\nabla\phi_1) = 0$ and $\frac{\partial\mu}{\partial t} + \operatorname{div}(\mu\nabla\phi_2) = 0$. If all smoothness issues are left aside,

$$\begin{aligned} DF(\mu)\cdot\frac{\partial\mu}{\partial t} &= \frac{d}{dt}F(\mu(t)) \\ &= \int_{\mathbb{R}^d} \frac{\delta F}{\delta\mu}\cdot\frac{\partial\mu}{\partial t}dx \\ &= -\int_{\mathbb{R}^d} \frac{\delta F}{\delta\mu}\cdot\operatorname{div}(\mu\nabla\phi_2)dx \\ &= \int_{\mathbb{R}^d} \left\langle \nabla\frac{\delta F}{\delta\mu}, \nabla\phi_2 \right\rangle d\mu(x), \end{aligned}$$

where $\frac{\delta F}{\delta \mu}$ is the gradient of the functional F with respect to the standard L^2 Euclidean structure. By (2.2.3), we then know $\nabla \phi_1 = \nabla \frac{\delta F}{\delta \mu}$ and

$$\operatorname{grad}_W F(\mu) = -\operatorname{div}\left(\mu \nabla \frac{\delta F}{\delta \mu}\right). \tag{2.2.5}$$

Thus $\mu(\,\cdot\,)$ is a gradient flow with respect to F on $(\mathcal{P}_2(\mathbb{R}^d), d_W)$ if

$$\frac{\partial}{\partial t}\mu(t) = -\operatorname{grad}_W F(\mu(t)) = \operatorname{div}\left(\mu(t) \nabla \frac{\delta F}{\delta \mu}(t)\right). \tag{2.2.6}$$

Some important examples of such gradient flows are:

$$\begin{aligned}
\mathcal{E}(\mu) &= \int_{\mathbb{R}^d} \frac{d\mu}{d\mathcal{L}^d} \log \frac{d\mu}{d\mathcal{L}^d} dx, & \frac{\partial}{\partial t}\mu &= \Delta \mu;\\
\mathcal{E}(\mu) &= \int_{\mathbb{R}^d} \frac{d\mu}{d\mathcal{L}^d} \log \frac{d\mu}{d\mathcal{L}^d} dx + \int_{\mathbb{R}^d} V d\mu, & \frac{\partial}{\partial t}\mu &= \Delta \mu + \operatorname{div}(\mu \nabla V);\\
\mathcal{E}(\mu) &= \frac{1}{m-1} \int_{\mathbb{R}^d} \left(\frac{d\mu}{d\mathcal{L}^d}\right)^m dx, & \frac{\partial}{\partial t}\mu &= \Delta \mu^m;\\
\mathcal{E}(\mu) &= \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y) d\mu(x) d\mu(y), & \frac{\partial}{\partial t}\mu &= \operatorname{div}\left(\mu \nabla W * \mu\right).
\end{aligned}$$

These equations are known as heat equation, linear Fokker-Planck equation, porous medium equation and nonlocal interaction equation.

To make the formal computations rigorous, we need to introduce gradients in metric space and Subdifferential Calculus in $(\mathcal{P}_2(X), d_W)$, see Chapters 1 and 10 in [5].

For $\mathcal{E} : \mathcal{P}_2(X) \mapsto (-\infty, +\infty]$ an energy functional, we define the the *local slope* of $\mathcal{E}$ with respect to d_W at $\mu \in \mathcal{P}_2(X)$ as

$$|\partial \mathcal{E}|(\mu) = \limsup_{\nu \to \mu} \frac{(\mathcal{E}(\mu) - \mathcal{E}(\nu))^+}{d_W(\mu, \nu)}, \tag{2.2.7}$$

where $f^+ = \max\{f, 0\}$ is positive part of f.

For a locally absolutely continuous curve $[0, +\infty) \ni t \mapsto \mu(t) \in \mathcal{P}_2(X)$ with respect to the Wasserstein distance d_W, we denote its *metric derivative* by

$$|\mu'|(t) = \limsup_{s \to t} \frac{d_W(\mu(t), \mu(s))}{|s-t|}. \tag{2.2.8}$$

We now define the *upper gradient* of $\mathcal{E}$ as a kind of modulus of the gradient for the energy functional $\mathcal{E}$.

Definition 2.2.1 (Upper gradient). *A Borel function $g : \mathcal{P}_2(X) \mapsto [0, +\infty]$ is called a* strong upper gradient *for the functional $\mathcal{E}$ if for every $\mu(\,\cdot\,) \in AC(a, b; \mathcal{P}_2(X))$, the function $g \circ \mu(\,\cdot\,)$ is Borel and*

$$|\mathcal{E}(\mu(t)) - \mathcal{E}(\mu(s))| \leq \int_s^t g \circ \mu(r) |\mu'|(r) dr \qquad \forall a < s \leq t < b. \tag{2.2.9}$$

There is also a notion of weak upper gradient, see for example Definition 1.2.2 from [5]. Since we will not use it, we omit its definition here, and all upper gradient means strong upper gradient. We can now introduce the notion of *curve of maximal slope* with respect to an upper gradient.

Definition 2.2.2. *A locally absolutely continuous curve* $[0, \infty) \ni t \mapsto \mu(t) \in \mathcal{P}_2(X)$ *is a* curve of maximal slope *for the functional* $\mathcal{E}$ *with respect to upper gradient* g*, if* $\mathcal{E} \circ \mu(\,\cdot\,)$ *is* $\mathcal{L}^1$*-a.e. equal to a non-increasing function* $\varphi(\,\cdot\,)$ *and*

$$\varphi'(t) \leq -\frac{1}{2}|\mu'|^2(t) - \frac{1}{2}g^2\left(\mu(t)\right) \tag{2.2.10}$$

for a.e $t \in (0, \infty)$*. Here* $|\mu'|(t)$ *is the metric derivative defined in* (2.2.8)*.*

The general strategy of constructing curves of maximal slope in the space of probability measures is to use the Jordan-Kinderlehrer-Otto (minimizing movement) scheme [65], which we describe in Subsection 2.2.1.

To build connections between gradient flows in the space of probability measures and solutions to continuity equations, recall from [5] that locally absolutely continuous curves $\mu(\,\cdot\,) \subset \mathcal{P}_2(X)$ are solutions to continuity equations among which there exists a velocity field such that the metric slope is realized by it. In particular, we cite the following Theorem 8.3.1 from [5].

Theorem 2.2.3. *Let* I *be an open interval in* $\mathbb{R}_+$*, let* $\mu : I \mapsto \mathcal{P}_2(X)$ *be an absolutely continuous curve and let* $|\mu'| \in L^1(I)$ *be its metric derivative defined in (2.2.8). Then there exists a unique Borel vector field* $v : (t, x) \mapsto v(t, x)$ *such that*

$$\frac{\partial}{\partial t}\mu(t, x) + \operatorname{div}(\mu(t, x)v(t, x)) = 0 \tag{2.2.11}$$

holds in the sense of distributions, with $|\mu'|^2(t) = \int_X |v(t, x)|^2 d\mu(t, x) = \|v(t)\|^2_{L^2(\mu(t),X)}$ *for a.e.* $t \in I$*.*

We call the unique Borel vector field $v(\,\cdot\,)$ the *tangent velocity* to $\mu(\,\cdot\,)$. To define the concept of gradient flow in the space of probability measures, we still need to introduce the subdifferential calculus in $\mathcal{P}_2(X)$ following Chapter 10 from [5]. For fixed $\mu \in \mathcal{P}_2(X)$, when μ is regular (refer to Definition 6.2.2 for definition of regular measures in general Polish space, which corresponds to absolutely continuity with respect to $\mathcal{L}^d$ when $X = \mathbb{R}^d$), we know the optimal map between μ and an arbitrary $\nu \in \mathcal{P}_2(X)$ exists which we denote by t_μ^ν. In that case,

Definition 2.2.4. *We say that* $\xi \in L^2(\mu; X)$ *belongs to the* Fréchet subdifferential $\partial\mathcal{E}(\mu)$ *if*

$$\mathcal{E}(\nu) \geq \mathcal{E}(\mu) + \int_X \langle \xi(x), t_\mu^\nu(x) - x \rangle d\mu(x) + o(d_W(\mu, \nu)). \tag{2.2.12}$$

When the reference probability measure μ is not regular, we instead use the following generalized definition.

Definition 2.2.5. *Fix $\mu \in \mathcal{P}_2(X)$, a vector field $\xi \in L^2(\mu, X)$ is said to be an element of the* subdifferential *of $\mathcal{E}$ at μ, and we denote by $\xi \in \partial\mathcal{E}(\mu)$, if*

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \geq \inf_{\gamma \in \Gamma_o(\mu,\nu)} \int_{X\times X} \langle \xi(x), y - x\rangle d\gamma(x,y) + o\left(d_W(\mu,\nu)\right), \tag{2.2.13}$$

where $\Gamma_o(\mu,\nu)$ is the set of optimal plans between μ and ν as defined in (2.1.4).

We call a locally absolutely continuous curve $\mu(\,\cdot\,) \subset \mathcal{P}_2(X)$ a *gradient flow* with respect to the energy functional $\mathcal{E}$ if for a.e. $t > 0$,

$$v(t) \in -\partial\mathcal{E}(\mu(t)), \tag{2.2.14}$$

where $v(\,\cdot\,)$ is the tangent velocity of $\mu(\,\cdot\,)$ introduced in Theorem 2.2.3.

In gradient flow theory, (semi-)convexity of energy functional is important in showing the existence and quantitative stability estimates of gradient flows. In the space of probability measures, we define the geodesic (semi-)convexity notions as follows.

Definition 2.2.6. *Given $\lambda \in \mathbb{R}$, we say that $\mathcal{E} : \mathcal{P}_2(X) \to (-\infty, +\infty]$ is λ-geosdesically convex if for every couple $\mu^0, \mu^1 \in \mathcal{P}_2(X)$ and any constant speed minimal geodesic μ^t connecting μ^0 and μ^1*

$$\mathcal{E}(\mu^t) \leq (1-t)\mathcal{E}(\mu^0) + t\mathcal{E}(\mu^1) - \frac{\lambda}{2} t(1-t) d_W^2(\mu^0, \mu^1) \qquad \forall t \in [0,1]. \tag{2.2.15}$$

Here μ^t is a constant speed minimal geodesic if $d_W(\mu^t, \mu^s) = |t-s| d_W(\mu^0, \mu^1)$ for all $0 \leq s \leq t \leq 1$.

For λ-geodesically convex energy functional $\mathcal{E}$, we can try to find solutions to a system of variational inequalities (EVI) similar to Hilbertian settings.

Definition 2.2.7. *Given a parameter $\lambda \in \mathbb{R}$, the curve $\mu : [0,\infty) \longrightarrow \mathcal{P}_2(X)$ is gradient flow with parameter λ in the Evolution Variational Inequality (EVI) sense if $\mu(\,\cdot\,)$ is locally absolutely continuous and*

$$\mathcal{E}(\mu(t)) + \frac{1}{2}\frac{d}{dt} d_W^2(\mu(t), \nu) + \frac{\lambda}{2} d_W^2(\mu(t),\nu) \leq \mathcal{E}(\nu), \ \forall \nu \in \mathcal{P}_2(X), \ a.e. \ t > 0. \tag{2.2.16}$$

2.2.1 JKO scheme

In this Subsection, we introduce the JKO scheme introduced by Jordan, Kinderlehrer and Otto in [65] (also refer to [2, 40] for minimizing movement scheme) to construct curves of maximal slope in $(\mathcal{P}_2(X), d_W)$. Here we follow the presentation of Chapter 2 from [5].

Fix a time step $\tau > 0$ and define $\mu_\tau^0 = \mu_0$ where μ_0 are the initial data. Then define iteratively

$$\mu_\tau^{k+1} \in \operatorname{argmin}_{\mu \in \mathcal{P}_2(X)} \left[\frac{d_W^2(\mu, \mu_\tau^k)}{2\tau} + \mathcal{E}(\mu) \right]. \tag{2.2.17}$$

We denote the piecewise constant interpolation by μ_τ. To be more precise, $\mu_\tau(0) = \mu_0$ and

$$\mu_\tau(t) = \mu_\tau^{k+1}, \tag{2.2.18}$$

if $k\tau < t \leq (k+1)\tau$ for $k \geq 0$. The strategy is to show that there exists a subsequence $\tau_n \to 0$, such that $\tilde{\mu}^n(\,\cdot\,) = \mu_{\tau_n}(\,\cdot\,)$ converges narrowly to a curve of maximal slope $\mu(\,\cdot\,)$. Here in order to show the well-posedness of discrete scheme (2.2.17) and the convergence of the piecewise-constant interpolation, we present the general theory developed in [5]. The topological conditions we need to check are given as follows.

- Lower semicontinuity. $\mathcal{E}$ is sequentially lower semicontinuous with respect to narrow convergence of probability measures on d_W bounded sets

$$\sup_{m,n} d_W(\mu_m, \mu_n) < \infty,\ \mu_n \text{ converges narrowly to } \mu \Rightarrow \liminf_{n\to\infty} \mathcal{E}(\mu_n) \geq \mathcal{E}(\mu).$$

- Coercivity. There exists $\tau_* > 0$ and $\mu_* \in \mathcal{P}_2(X)$ such that

$$\inf_{\mu \in \mathcal{P}_2(\mathcal{M})} \left\{ \mathcal{E}(\mu) + \frac{1}{2\tau_*} d_W^2(\mu, \mu_*) \right\} > -\infty.$$

- Compactness. Every d_W bounded set contained in a sublevel of $\mathcal{E}$ is relatively compact with respect to the narrow convergence of probability measures

$$\text{for } (\mu_n) \subset \mathcal{P}_2(X) \text{ with } \sup_n \mathcal{E}(\mu_n) < \infty \text{ and } \sup_{m,n} d_W(\mu_m, \mu_n) < \infty,$$
$$\text{there exists a narrowly convergent subsequence of } (\mu_n).$$

Given that the above three conditions hold, Corollary 2.2.2 from [5] gives the existence of minimizers to (2.2.17).

Lemma 2.2.8 (Existence of the discrete solutions)**.** *If the lower semicontinuity, coercivity and compactness conditions are verified, then for any $\tau < \tau_*$ and $\nu \in \mathcal{P}_2(X)$ there exists $\mu_\infty \in \mathcal{P}_2(X)$ such that*

$$\mathcal{E}(\mu_\infty) + \frac{1}{2\tau} d_W^2(\nu, \mu_\infty) = \inf_{\mu \in \mathcal{P}_2(X)} \left\{ \mathcal{E}(\mu) + \frac{1}{2\tau} d_W^2(\nu, \mu) \right\}. \tag{2.2.19}$$

Proposition 2.2.3 from [5] provides the compactness result for convergence of interpolation curves from the JKO scheme.

Proposition 2.2.9 (Compactness). *If the lower semicontinuity, coercivity and compactness conditions are verified, then there exist a limit curve $\mu \in AC^2_{loc}([0,\infty);\mathcal{P}_2(X))$ and a sequence $\tau_n \to 0^+$ such that the piecewise constant interpolate $\tilde{\mu}^n(\cdot) = \mu_{\tau_n}(\cdot)$ defined as in (2.2.18) satisfies that $\tilde{\mu}^n(t)$ converges narrowly to $\mu(t)$ for any $t \in [0,\infty)$.*

Note that by Lemma 3.2.2 from [5], we actually have a uniform bound on the second moments of $\tilde{\mu}^n$:

$$\sup_{n,\tau} \int_{\mathcal{M}} \mathrm{dist}^2(x, x_0) d\mu_n^\tau(x) < \infty.$$

By Theorem 2.3.3 in [5], the limit curve $\mu(\cdot)$ is a curve of maximal slope with respect to the relaxed local slope $|\partial^- \mathcal{E}|$, defined as

$$|\partial^- \mathcal{E}|(\mu) = \inf \left\{ \liminf_{n\to\infty} |\partial \mathcal{E}|(\mu_n) : \mu_n \rightharpoonup \mu, \sup_n \{d_W(\mu_n, \mu), \mathcal{E}(\mu_n)\} < \infty \right\}, \tag{2.2.20}$$

where $\mu_n \rightharpoonup \mu$ means that μ_n converges narrowly to μ, provided $|\partial^- \mathcal{E}|$ is a strong upper gradient of $\mathcal{E}$.

Theorem 2.2.10. *Assume that $\mathcal{E}$ is lower semicontinuous and coercive. If*

$$\mathcal{P}_2(X) \ni \mu \mapsto |\partial^- \mathcal{E}|(\mu) \text{ is a strong upper gradient for } \mathcal{E}, \tag{2.2.21}$$

then the limit curve $\mu(\cdot)$ is a curve of maximal slope for $\mathcal{E}$ with respect to upper gradient $|\partial^- \mathcal{E}|$ and in particular $\mu(\cdot)$ satisfies the energy identity

$$\frac{1}{2}\int_0^T |\mu'|^2(t)dt + \frac{1}{2}\int_0^T |\partial^- \mathcal{E}|^2(\mu(t))dt + \mathcal{E}(\mu(T)) = \mathcal{E}(\mu_0). \tag{2.2.22}$$

Note that when local slope $|\partial \mathcal{E}|$ is lower semicontinuous with respect to narrow convergence of probability measures, $|\partial^- \mathcal{E}| = |\partial \mathcal{E}|$. In general, we still need to prove the lower semicontinuity of local slope to show that $\mu(\cdot)$ is a curve of maximal slope with respect to upper gradient $|\partial \mathcal{E}|$ instead of $|\partial^- \mathcal{E}|$.

2.3 Biological aggregation

Biological aggregation describes self-organized behavior of large system of biological agents such as swarming insects and flocking birds. The basic assumptions in modeling biological aggregation are long range attraction, short range repulsion and intermediate range alignment. To be more precise, there is pairwise interaction between any two agents in the group, and the pairwise interaction between the two agents is attractive when they are far from each other to stay in a social group; they are repulsive to each other when they are too close to avoid collision; in the intermediate distance range, they adjust their velocity to align with other agents. The dynamics of such systems is governed by long range interactions among agents, therefore such systems are intrinsically nonlocal and lead to the study nonlocal interaction equations.

Different models on how long range interaction applies have been proposed: by assuming that velocity of an agent is proportional to the force it is subject to, Bertozzi, Topaz and collaborators studied first-order models [107, 108, 109]; when acceleration of the agent is proportional to the force, second-order (self-propelled and alignment) models such as Vicsek model [113], Cucker-Smale model [37, 38] and Motsch-Tadmor model [88] were introduced. These models have been extensively investigated to predict and study flocking [60, 61, 89, 105] and pattern formation in biological aggregation [8, 30, 69, 116, 117].

First order models.

We work on the fundamental first order repulsive-attractive models , which can be used and combined with other effects to build more complex models, refer to [107, 108, 109] and the references therein. Let $x^i \in \mathbb{R}^d$ be the location of the i-th agent with mass $m_i > 0$ for $i = 1, \dots, N$. After normalization, we assume $\sum_{i=1}^N m_i = 1$. We also assume that the interaction force between agents i and j are through the gradient of an interaction potential function $W : \mathbb{R}^d \longrightarrow \mathbb{R}$ and only depends on their relative location, i.e., the force on agent i resulting from interaction with j is given by $-\nabla W(x^i - x^j)$. In general, we assume that W is symmetric in the sense $W(x) = W(-x)$. We also allow the existence of an external force such as gravity or wind, we denote the external potential by V. Then the dynamics of the agents follows the system of ODEs,

$$\dot{x}^i(t) = -\sum_{j\neq i}^{N} m_j \nabla W(x^i(t) - x^j(t)) - \nabla V(x^i), \qquad \forall i \in \{1, \dots, N\}. \tag{2.3.1}$$

In many biological relevant applications, we assume that the interaction between two agents only depends on their distance to each other, i.e. $W(x) = w(|x|)$. Then, for the system to be short distance repulsive and long distance attractive, we only need to require that $w'(r) < 0$ for $r < R_r$ and $w'(r) > 0$ for $r > R_a$ for some $0 < R_r < R_a$.

If we denote $x = (x^1, \dots, x^N) \in \mathbb{R}^{dN}$, define the interaction energy

$$\mathcal{W}(x) = \frac{1}{2} \sum_{i\neq j} m_i m_j W(x^i - x^j), \tag{2.3.2}$$

and potential energy

$$\mathcal{V}(x) = \sum_{i=1}^{N} m_i V(x^i). \tag{2.3.3}$$

Then the total energy $\mathcal{E}(x) = \mathcal{W}(x) + \mathcal{V}(x)$ is a dissipative quantity along the solution $x(t) = (x^1(t), \dots, x^N(t))$ of (2.3.1) since

$$
\begin{aligned}
\frac{d}{dt}\mathcal{E}(x(t)) &= \frac{1}{2}\sum_{i\neq j} m_i m_j \left\langle \nabla W(x^i(t) - x^j(t)), \dot{x}^i(t) - \dot{x}^j(t) \right\rangle + \sum_i m_i \langle \nabla V(x^i(t)), \dot{x}^i(t)\rangle \\
&= \sum_i m_i \left\langle \sum_{j\neq i} m_j \nabla W(x^i(t) - x^j(t)) + \nabla V(x^i(t)), \dot{x}^i(t) \right\rangle \\
&= -\sum_i m_i \left| \sum_{j\neq i} m_j \nabla W(x^i(t) - x^j(t)) + \nabla V(x^i(t)) \right|^2 \\
&\leq 0,
\end{aligned}
$$

where we used the fact that $\nabla W(-x) = -\nabla W(x)$ by the symmetry of W. Note that the empirical probability distribution defined by

$$
\mu(t) = \sum_{i=1}^{N} m_i \delta_{x^i(t)} \tag{2.3.4}
$$

satisfies the continuity equation

$$
\frac{\partial}{\partial t}\mu(t,x) - \operatorname{div}\left(\mu(t,x)\left(\nabla W * \mu(t)(x) + \nabla V(x)\right)\right) = 0 \tag{2.3.5}
$$

in the sense of distributions. Indeed for any $\phi \in C_c^\infty(\mathbb{R}^d)$,

$$
\begin{aligned}
\frac{d}{dt}\int_{\mathbb{R}^d} \phi(x) d\mu(t,x) &= \frac{d}{dt}\sum_{i=1} m_i \phi(x^i(t)) \\
&= \sum_{i=1} m_i \left\langle \nabla\phi(x^i(t)), \dot{x}^i(t) \right\rangle \\
&= \sum_{i=1} m_i \left\langle \nabla\phi(x^i(t)), -\sum_{j\neq i} m_j \nabla W(x^i(t) - x^j(t)) - \nabla V(x^i(t)) \right\rangle \\
&= -\int_{\mathbb{R}^d} \langle \nabla\phi(x), \nabla W * \mu(t)(x) + \nabla V(x)\rangle \, d\mu(t,x),
\end{aligned}
$$

which verifies that $\mu(\,\cdot\,)$ is a solution to (2.3.5) as claimed.

We remark here that the passage from discrete to continuum does not involve taking hydrodynamic limit, as the solutions to (2.3.1) are solutions to (2.3.5) via (2.3.4). Thus we can consider the nonlocal interaction equation (2.3.5) to include both discrete and continuum distributions at the same time.

If we define similarly as in discrete setting the interaction energy as

$$
\mathcal{W}(\mu) = \frac{1}{2}\int_{\mathbb{R}^d} W(x-y) d\mu(x) d\mu(y), \tag{2.3.6}
$$

potential energy as

$$\mathcal{V}(\mu) = \int_{\mathbb{R}^d} V(x) d\mu(x), \tag{2.3.7}$$

and total energy

$$\mathcal{E}(\mu) = \mathcal{W}(\mu) + \mathcal{V}(\mu). \tag{2.3.8}$$

Formal computations (ignoring all smoothness issues) show that for $\mu(\cdot)$ a solution to (2.3.5), we have

$$\frac{d}{dt}\mathcal{E}(\mu(t)) = -\int_{\mathbb{R}^d} |\nabla W * \mu(t)(x) + \nabla V(x)|^2 d\mu(t,x) \leq 0. \tag{2.3.9}$$

Again $\mathcal{E}$ is dissipative along the solution to (2.3.5) and actually, solutions to (2.3.5) can be viewed as gradient flows of $\mathcal{E}$ in the space of probability measures under some regularity assumptions on W, V, as shown in [5, 28].

In [18], Bertozzi, Laurent and Rosado studied L^p well-posedness of the aggregation equation (2.3.5). They considered radially symmetric interaction potential function W where the singularity at the origin is of order $|x|^\alpha$ for some $\alpha > 2 - d$, and proved the local well-posedness of (2.3.5) in $\mathcal{P}_2(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ for any $p > p_s$, where $p_s = \frac{d}{d+\alpha-2}$. In [28], Carrillo, Di Francesco, Figalli, Laurent and Slepčev showed the global well-posedness of weak measure solutions to (2.3.5) in $\mathcal{P}_2(\mathbb{R}^d)$ given that W is (semi-)convex, Lipschitz continuous and with at most quadratic growth at infinity. With $V \equiv 0$, they considered solutions to (2.3.5) as gradient flows of the interaction energy $\mathcal{W}$ in the space of probability measures endowed with Wasserstein metric.

Other important properties of nonlocal interaction equations such as blowup (concentration) [14, 15, 16, 17, 62, 64], confinement [9, 29], stability and properties of stationary states [8, 41, 49, 50, 51, 70], asymptotic behavior [27, 63, 72, 98] and related models that incorporate further effects [13, 110] have also been extensively studied.

Second order models.

There are different second order models depending how the velocity of the agents change according to the force they are subject to. Here we briefly introduce Vicsek model, self-propelled interacting particle model, Cucker-Smale model and Motsch-Tadmor model.

In Vicsek model introduced by Vicsek, Czirók, Ben-Jacob, Cohen and Shochet in [113], the i-th agent is described by its position x^i and its velocity $v^i = v_0 e^{i\theta_i}$. Note that $|v^i(t)| = v_0$ for all i and all $t > 0$ in this framework. Fixing a time step $\Delta t > 0$, a radius $r_0 > 0$, a constant $\eta > 0$ and denote $U_i(t) = \{\xi : |\xi - x^i(t)| < r_0\}$. The dynamics of the agents are given by

$$\begin{aligned}
\theta_i(t+\Delta t) &= \langle \theta(t)\rangle_i + \eta\xi_i(t),\\
v^i(t+\Delta t) &= v_0 e^{i\theta_i(t+\Delta t)},\\
x^i(t+\Delta t) &= x^i(t) + v^i(t+\Delta t)\cdot \Delta t,
\end{aligned}$$

where $\langle v(t)\rangle_i = \frac{\sum_{j:x^j\in U_i(t)} v^j(t)}{\sum_{j:x^j\in U_i(t)} 1}$, $\langle\theta(t)\rangle_i$ is the angle of $\langle v(t)\rangle_i$, and $\xi_i(t)$ is a random variable uniformly distributed on $[-\pi,\pi]$. The idea is that the i-th agent interacts with all agents within r_0 radius of it and the mechanism it reacts is to calculate the average velocity of its neighbors $\langle v\rangle_i$ and adjust its direction of velocity θ_i to the average direction$\langle\theta\rangle_i$ plus some noise with strength η. The two main parameters of the Vicsek model are ρ, the density of particles, and η, the noise strength. There is a phase transition from ordered to unordered phase as η increases. Numerical results have shown that there exists an ordered phase for $0<\eta<\eta_c$ and the transition line in the (ρ,η) plane, follows the scaling law (expected from a simple mean-field argument): $\eta_c \sim \rho^{\frac{1}{d}}$ for small ρ.

We now turn to the self-propelled interacting particle model introduced in [73] by Levine, Rappel and Cohen, and extensively studied in [34, 41] by Chuang, D'Orsogna, Marthaler, Chayes and Bertozzi. Given an interaction potential function W and self-propulsion function S, the dynamics of agents are given by

$$\dot{x}^i(t) = v^i(t), \qquad \dot{v}^i(t) = S(|v^i|)v^i + \frac{1}{N}\sum_{j\neq i}\nabla W(x^j - x^i) \qquad \forall i\in\{1,\dots,N\}.$$

An example for the self-propulsion term is given by $S(|v^i|) = \alpha - \beta|v^i|^2, \alpha,\beta>0$ as used in [34, 41]. When taking $W(x) = w(|x|)$ to be the radially symmetric Morse potential $w(r) = C_A e^{-r/l_A} - C_R e^{-r/l_R}$, they find several patterns for the asymptotic behavior such as flocking, mill on a ring, and clustering when particles are milling.

The Cucker-Smale model proposed by Cucker and Smale in [37, 38] describes how agents interact in order to align with their neighbors. The rule is that the closer two individuals are, the more they tend to align with each other (long range cohesion and short range repulsion are ignored). The evolution of each agent is then governed by the following dynamical system,

$$\dot{x}^i(t) = v^i(t), \qquad \dot{v}^i(t) = \frac{\alpha}{N}\sum_{j=1}^{N}\phi_{ij}(v^j(t) - v^i(t)). \tag{2.3.10}$$

Here, α is a positive constant and ϕ_{ij} quantifies the pairwise influence of agent j on the alignment of agent i, as a function of their distance,

$$\phi_{ij} = \phi(|x^j - x^i|).$$

The so-called influence function, $\phi(\,\cdot\,)$, is a strictly positive decreasing function which, by rescaling α if necessary, is normalized so that $\phi(0) = 1$. An example for such an influence function is given by $\phi(t) = (1+r)^{-s}$, $s > 0$. One important feature in the Cuker-Smale model is that it is symmetric in the sense that the coefficients matrix ϕ_{ij} is symmetric $\phi_{ij} = \phi_{ji}$. As a direct consequence, we know that the average velocity $\bar{v}(t) = \frac{1}{N}\sum_{j=1}^{N} v^j(t) \equiv \bar{v}(0)$ remains unchanged.

The Cuker-Smale model with a slowly decaying influence function $\phi(\,\cdot\,)$ such that

$$\int^{\infty} \phi(r)dr = \infty$$

has an unconditional convergence to a so-called flocking dynamics, in the sense that the diameter, $\max_{i,j} |x^i(t) - x^j(t)|$, remains uniformly bounded and all agents of this flock will approach the same velocity $\bar{v}(0)$.

The description of self-organized dynamics by the Cuker-Smale model suffers from several drawbacks, such as after the normalization of the model by the total number of agents N, it is inadequate for far-from-equilibrium scenarios. In [88], Motsch and Tadmor suggest the following modified Motsch-Tadmor model,

$$\dot{x}^i(t) = v^i(t), \qquad \dot{v}^i(t) = \frac{\alpha}{\sum_{k=1}^{N}\phi_{ik}} \sum_{j=1}^{N} \phi_{ij}(v^j(t) - v^i(t)), \qquad \phi_{ij} = \phi(|x^i - x^j|). \quad (2.3.11)$$

In this model the symmetry is lost. However, Motsch and Tadmor show dynamics of the model would experience unconditional flocking provided the influence function ϕ decays sufficiently slowly such that

$$\int^{\infty} \phi^2(r)dr = \infty.$$

Another difference between the flocking behavior of Mostch-Tadmor and Cucker-Smale is that: unlike the Cucker-Smale flocking to the initial bulk velocity $v(0)$, the asymptotic flocking velocity of this Mostch-Tadmor is not necessarily encoded in the initial configuration as an invariant of the dynamics, but it is emerging through the flocking dynamics of the model.

In these second order models, we need to take hydrodynamic limit to pass from particle to kinetic and continuum descriptions. For example in the Mostch-Tadmor models, Motsch and Tadmor showed in [88] the hydrodynamic limit of (2.3.11) is given by

$$\begin{aligned} &\partial_t\rho + \nabla_x \cdot (\rho u) = 0, \\ &\partial_t(\rho u) + \nabla_x(\rho u \otimes u) = \alpha\rho\left(\frac{\langle u\rangle}{\langle 1\rangle} - u\right), \ \langle w\rangle(x) = \int_y \phi(|x-y|)w(y)\rho(y)dy. \end{aligned}$$

For other results on hydrodynamic limits of different flocking models and properties of resulting equations, we refer to [61, 105] and references therein.

Chapter 3

Nonlocal interaction equations in heterogeneous environments with boundaries

In this Chapter, we present results on nonlocal interaction equations modeling biological aggregations in heterogeneous environments with boundaries. We study well-posedness of a class of nonlocal interaction equations with spatially dependent mobility. This leads to the study of the nonlocal interaction equations on subsets $\mathcal{M} \subset \mathbb{R}^d$ endowed with a Riemannian metric g induced by the variable mobility. We obtain conditions, relating the interaction potential and the geometry, which imply existence, uniqueness and stability of solutions. We study the equations in the setting of gradient flows in the space of probability measures on $\mathcal{M}$ endowed with Riemannian 2-Wasserstein metric. The results presented here are based on our paper [119].

Nonlocal interaction equations serve as basic models of biological aggregation, that is collective motion of agents under influence of long-range interactions (via sight, sound, etc.). In this Chapter we investigate the nonlocal interaction equation in heterogeneous environments and also allow for the presence of domain boundaries. On the whole space (when no boundaries are present) the equations are of the form

$$\frac{\partial}{\partial t}\mu(t,x) - \operatorname{div}\left(\mu(t,x)A(x)\nabla\left(W * \mu(t)(x) + V(x)\right)\right) = 0, \tag{3.0.1}$$

where μ describes the agent density, A is the mobility matrix (symmetric and positive definite), W is the interaction potential and V is the external potential.

The mobility endows the subsets of $\mathbb{R}^d$ with Riemannian structure, which leads us to study nonlocal interaction equations on manifolds. We study the well-posedness of the equations in the setting of gradient flows in spaces of probability measures [5, 28]. To extend this setting to manifolds with boundary we need to overcome several challenges. Namely "mass" can accumulate at the boundary and the velocities associated to the gradient flow are not continuous at the boundary. This also causes the problem that in general, we do not have the existence of optimal maps and thus we have to work with optimal plans instead. Furthermore the velocities (of the gradient flows) lack the stability properties used to prove the lower semicontinuity of the slope (see for example Lemma 2.7 in [28]). Studying the equation on a manifold raises issues too. The curvature of the space can cause even the quadratic potential not to be geodesically semi-convex. Thus a particular care and extra conditions are needed when discussing properties like geodesic semi-convexity of energies. Furthermore many standard tools used to study nonlocal equations rely on the linearity of the underlying space and ability to directly identify tangent spaces at different points. Thus these tools do not readily transfer to the manifold setting. For example the standard proof of the characterization of the subdifferential of the interaction energy does not apply in the manifold setting. We develop alternative proofs to handle these challenges.

3.1 Motivation and setup

The studies of the nonlocal equations on heterogeneous environments are in part motivated by the desire to understand mechanisms which give rise to rolling swarms. Such swarms are observed in a number biological swarms, notably the locust swarms (see [107] and references therein). In [107], Topaz, Bernoff, Logan and Toolson propose a model which has a gradient flow structure of an energy that combines the interaction energy and potential energy terms (to model gravity and wind). The mobility in their system is as follows: consider the upper half plan $\mathbb{R}^2_+ = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_2 \geq 0\}$, above the ground the mobility is constant ($A(x) = I_{2\times 2}$), while on the ground the mobility in the horizontal direction if zero ($A(x) = \operatorname{diag}(0, 1)$). They conduct numerical experiments and observe rolling swarms. Here we introduce a model where the change in mobility is more gradual, and thus amenable to rigorous study. The solutions still exhibit the rolling swarms when a smoothed out version of the mobility in [107] is considered. Moreover, rolling forms are present if the horizontal mobility is stratified (increases with height), even if gravity is not present. Figure 3.1 illustrates such a rolling swarm. The interaction potential used is among ones considered in [70], and is given by $W(z) = w(|z|)$ with $w'(r) = \tanh(3(1-r)) + 0.3$. On the right, we also show the corresponding traveling "swarm" in the homogeneous environment. The velocities of all particles are the same. Moreover the configuration seen in the moving coordinate frame is a steady state of the energy $\mathcal{E}_2(\mu) = \iint W(x-y)d\mu(x)d\mu(y)$.

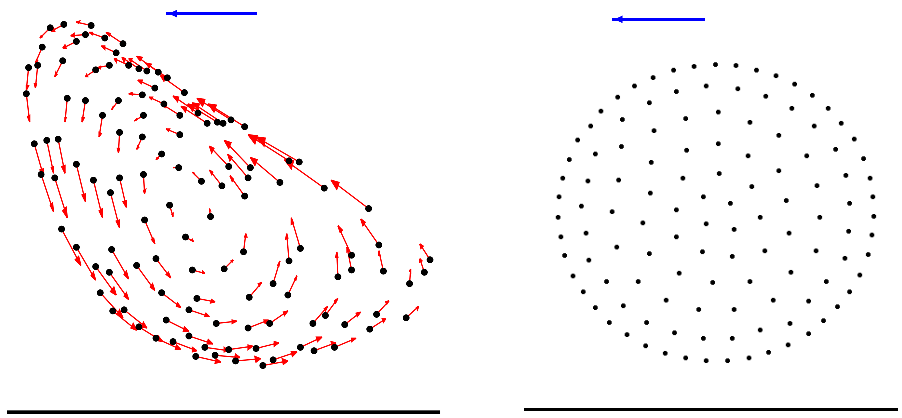


Figure 3.1: Consider gradient flows of $\mathcal{E}(\mu) = \iint W(x-y)d\mu(x)d\mu(y) + \int x_1\, d\mu(x)$, with respect to a stratified metric $G(x) = G((x_1, x_2)) = \text{diag}(\frac{1}{x_2^2}, 1)$ and the usual Euclidean metric. The gradient flow with respect to the stratified metric admits a rolling-wave solution made of a finite number of particles (left). The solution is given in the reference frame of the center of mass. The overall direction of motion is indicated by the large arrow on top (blue). The smaller arrows indicate the velocity of particles in the moving coordinates. The gradient flow with respect to Euclidean metric admits a traveling wave (right). All particle velocities are the same; hence in the moving coordinates the solution appears stationary.

Gradient flows in spaces of probability measures on manifolds: Background.

Let us first recall that the existence of optimal transportation maps on manifolds was first considered by McCann [82], and subsequently generalized in [36, 48]. The regularity of these maps has been the subject of a lot of recent activity and progress (see [52] and references therein). For our purposes however, the results on optimal transportation plans presented in Villani's book [115] are sufficient.

Regarding the gradient flows there has been a significant progress in investigating the gradient flow of entropy (i.e. the heat equation) and other internal energies on manifolds, as well spaces with weaker geometric structure. In particular Lisini [75] considered $\mathbb{R}^d$ endowed with a bounded Riemannian metric G, satisfying $\Lambda_1 I_d \le G \le \Lambda_2 I_d$, and showed the existence of solutions to the equation

$$\frac{\partial}{\partial t}u(t,x) - \text{div}\left(A(x)\left[\nabla(f(u(t,x)) + u(t,x)\nabla V(x)\right]\right) = 0, \tag{3.1.1}$$

on the whole space $\mathbb{R}^d$ with $A(x) = G^{-1}(x)$. In [92] Otto and Westdickenberg used an Eulerian calculus method to give sufficient conditions for the internal and potential energy to be geodesically convex in the space of probability measures endowed with Riemannian Wasserstein metric. In [39], Daneri and Savaré refined the approach of [92] to include the case of geodesic semi-convexity. In [101], Sturm gave the necessary and sufficient conditions for internal and potential energies to be λ-geodesically convex in the space of probability measures endowed with Riemannian Wasserstein metric. Erbar [45] used these conditions to establish well-posedness of heat equation on manifolds in the framework of gradient flows in spaces of probability measures. Gradient flows of the internal energy on manifolds were also discussed in in [115]. Connections with geometry and extensions to weaker spaces have received significant attention, see [6, 56, 57, 90] and references therein. However, to the best of our knowledge the gradient flow of nonlocal interaction energies on manifolds has not been considered.

Description of the problem.

Let $\mathcal{M}$ be a, possibly unbounded, d-dimensional subset of $\mathbb{R}^d$ with C^2 boundary. We consider $\mathcal{M}$ with a Riemannian metric g. Throughout the paper we assume that $(\mathcal{M}, g)$ is complete under the metric induced by g and geodesically convex, that is, for any two points in $\mathcal{M}$ there exists a length minimizing geodesic in $\mathcal{M}$ connecting them. The Riemannian structure encodes the mobility of the agents which depends on the environment. The strength of the interaction is not affected by the environment. To give an example, we study situations where the properties of the terrain affect the mobility of the agents, but not their ability to see each other. Also the density of agents at a given location is with respect to the standard Euclidean volume/area; it is not affected by the metric g. This leads us to study equations in a mixed formulation, where the volume and interaction are with respect to Euclidean structure, while the mobility is with respect to the manifold structure, g.

To study the equations we use their gradient-flow structure, which enables us to write the equations in the form that at the same time applies both to discrete systems with finitely many agents and continuum descriptions. This follows from the theory developed for studies of the nonlocal interaction equations in homogeneous environments [5, 28]. More precisely a configuration (distribution of agents) is described by a measure μ supported on $\mathcal{M}$. The system is assumed to be conservative in the sense that no agents are created or leave the system during the evolution. In other words $\mu(\mathcal{M})$ does not change in time. This allows us to, by renormalizing the problem if needed, assume that configurations μ are probability measures.

The interaction is described by a symmetric interaction potential W. The corresponding interaction energy is defined as in (2.3.6). In addition to interaction we model the environmental influences such as gravity or food distribution by a potential V, which defines the potential energy (2.3.7). Again the total energy is sum of interaction and potential energy as defined in (2.3.8).

Gradient flow structure.

We introduce the geometry on the space of configurations first on a formal level. In nonlocal interaction equations (with no regularizing terms) mass can accumulate at the boundary and furthermore the velocity that describes the gradient flow can be discontinuous at the boundary. For this reason we use a more general way to introduce the gradient flow than is typically the case in heuristic arguments. We use a Lagrangian description of tangent vectors at a configuration. That is tangent vectors to the space of configurations are vector fields on $\mathcal{M}$. As is standard in differential geometry of manifolds with boundary, even at $x \in \partial\mathcal{M}$ we define the tangent space $T_x\mathcal{M}$ to be a vector space, in other words we do allow vectors that point outside the manifold. However since a path in the configuration space cannot take mass outside of $\mathcal{M}$, not all of the vectors in $T_x\mathcal{M}$ are admissible as values of the tangent vector field to the path in the configuration space. To define the set of admissible vectors for $x \in \partial\mathcal{M}$, let $T_x^{\text{in}}\mathcal{M}$ be the *inward sector*, namely the closed half-space of tangent vectors that do not point outside $\mathcal{M}$. That is let $T_x^{\text{in}}\mathcal{M}$ be the set of vectors $\xi \in T_x\mathcal{M}$ for which there exists a differentiable curve $\gamma : [0, \delta) \to \mathcal{M}$ such that $\gamma(0) = x$ and $\gamma'(0) = \xi$. We note that the tangent space to $\partial\mathcal{M}$, considered as a manifold, is a subset of the inward sector: $T\partial\mathcal{M} \subset T^{\text{in}}\mathcal{M}$.

The effort to infinitesimally move configuration μ in by a vector field $v \in T\mathcal{M}$ is

$$\int_{\mathcal{M}} g_x(v(x), v(x))d\mu(x) = \int_{\mathcal{M}} v^T(x)G(x)v(x)d\mu(x)$$

where G is the symmetric matrix which provides the metric g. However not all vector fields in $T\mathcal{M}$ are admissible as tangent vectors to a path in the configuration space. Namely the tangent vector fields must belong to the inward sector $T^{\text{in}}\mathcal{M}$. On the formal level, we consider admissible tangent vectors to the space of configurations to be vector fields in $T^{\text{in}}\mathcal{M}$ which are projections via P of a continuous vector field in $T\mathcal{M}$ as defined in (3.1.3). This is motivated by the fact, which we later establish, that gradient vector of energy $\mathcal{E}$ is given by $v = Pw$ where w is a continuous vector field ($w = (-G^{-1}\nabla(W * \mu + V))$).

The differential of $\mathcal{E}$ in the direction v is given as the directional derivative

$$\begin{aligned}
\operatorname{diff} \mathcal{E}[v] &= \frac{d}{dt}\bigg|_{t=0} \mathcal{E}(\mu_t) \\
&= \frac{d}{dt}\bigg|_{t=0} \left(\frac{1}{2} \int_{\mathcal{M}} \int_{\mathcal{M}} W(\Phi_v[t](x) - \Phi_v[t](y))d\mu(x)d\mu(y) + \int_{\mathcal{M}} V(\Phi_v[t](x))d\mu(x) \right) \\
&= \int_{\mathcal{M}} (\nabla W * \mu + \nabla V)v d\mu.
\end{aligned}$$

Above we used that $\mu_t = \Phi_v[t]_\sharp\mu$ and the symmetry of W, where Φ_v is the flow map associated with the velocity field v.

One can define the gradient descent of $\mathcal{E}$ with respect to metric given by g by defining $-\operatorname{grad}\mathcal{E}$ to be the admissible vector field v which minimizes

$$\int_{\mathcal{M}} g(v,v)d\mu + \operatorname{diff}\mathcal{E}[v]$$

that is

$$\int_{\mathcal{M}} \frac{1}{2}v^T Gv + \nabla(W*\mu+V)v d\mu. \tag{3.1.2}$$

To give this an interpretation of a true gradient flow we need to describe the tangent space to the space of configurations and endow it with an inner product. The issue is that more than one vector field can produce the same curve in the configuration space. Thus tangent vectors to the configuration space are defined as equivalence classes of admissible velocities which, for at least a short time, have the same flow map. The inner product of tangent vector fields at μ is defined as

$$\overline{g}(v,v) = \inf_{\tilde{v}}\left\{\int_{\mathcal{M}} \tilde{v}^T G\tilde{v}d\mu \,:\, (\exists\tilde{\delta}>0)(\forall t\in[0,\tilde{\delta}))\ \Phi_{\tilde{v}}[t]_\sharp\mu = \Phi_v[t]_\sharp\mu\right\}.$$

The tangent vector field v is considered as a representative of the class of velocities which produce the same curve. Since $\operatorname{diff} E[v]$ does not depend on the representative tangent vector field chosen, we note that $-\operatorname{grad}\mathcal{E}$ we defined is also a minimizer of

$$\frac{1}{2}\overline{g}\,(v,v) + \operatorname{diff}\mathcal{E}[v]$$

over all tangent vectors v at μ; which agrees with the standard definition of a gradient flow on a manifold.

To determine the gradient vector we minimize the expression in (3.1.2). We obtain $v(x) = -G^{-1}\nabla(W*\mu+V)(x)$ if x is in the interior of $\mathcal{M}$ and also when $x\in\partial\mathcal{M}$ and $-G^{-1}\nabla(W*\mu+V)(x)$ is in the interior of $T_x^{\text{in}}\mathcal{M}$. Otherwise $v = \Pi_{\partial\mathcal{M}}(-G^{-1}\nabla(W*\mu+V))$, where $\Pi_{\partial\mathcal{M}}$ is the orthogonal projection of $T_x\mathcal{M}$ to $T_x\partial\mathcal{M}$ with respect to g. Setting $A = G^{-1}$ and defining

$$P_x\xi = \begin{cases}\xi, & \text{if } x\notin\partial\mathcal{M} \text{ or } \xi\in T_x^{\text{in}}\mathcal{M}\\ \Pi_{\partial\mathcal{M}}(\xi), & \text{otherwise}\end{cases} \tag{3.1.3}$$

gives that $-\operatorname{grad}\mathcal{E}$ is given by the vector field

$$v = P(-A\nabla(W*\mu+V)). \tag{3.1.4}$$

The gradient flow of $\mathcal{E}$ is thus given by $\frac{\partial}{\partial t}\mu + \operatorname{div}(\mu v) = 0$, that is

$$\frac{\partial}{\partial t}\mu(t,x) + \operatorname{div}\left(\mu(t,x)P_x\left(-A(x)\left(\nabla W*\mu(t)(x) + \nabla V(x)\right)\right)\right) = 0. \tag{3.1.5}$$

Main results.

We denote the usual Euclidean inner product by $\langle , \rangle$. On manifold $(\mathcal{M}, g)$, for $\xi \in T_x\mathcal{M}$ we denote the norm associated to the metric g as $|\xi|_g = \sqrt{g_x(\xi,\xi)}$. We denote the Euclidean gradient and Hessian by ∇ and Hess and Riemannian gradient and Hessian by $\nabla_{\mathcal{M}}$ and $\mathrm{Hess}_{\mathcal{M}}$. For a function $f \in C^0(\mathcal{M})$, we say that f is λ-geodesically convex on $(\mathcal{M}, g)$ if for any $x, y \in \mathcal{M}$ and any constant speed minimal geodesic $\gamma(t)$ connecting x, y with $\gamma(0) = x, \gamma(1) = y$, we have

$$f(\gamma(t)) \leq (1-t)f(x) + tf(y) - \frac{\lambda}{2}t(1-t)\,\mathrm{dist}^2(x,y).$$

Notice that if $f \in C^2(\mathcal{M})$ with $\mathrm{Hess}_{\mathcal{M}} f(x) \geq \lambda G(x)$ for all $x \in \mathcal{M}$, then f is λ-geodesically convex on $(\mathcal{M}, g)$. For $\mathcal{M}$ a d-dimensional subset in $\mathbb{R}^d$ with C^2 boundary, we make the following assumptions on manifold $(\mathcal{M}, g)$:

(M1) The Riemannian metric g is C^2 and satisfies $|\xi|_g^2 \geq \Lambda|\xi|^2$ for some constant $\Lambda > 0$ and all $\xi \in T\mathcal{M}$.

(M2) $(\mathcal{M}, g)$ is geodesically convex in that for all $x, y \in \mathcal{M}$ there exists a length minimizing geodesic contained in $\mathcal{M}$.

We also make the following assumptions on interaction potential W and external potential V:

(NL1) $W(x) = W(-x)$ and $W(0) = 0$.

(NL2) $w(x,y) := W(x-y)$ is λ-geodesically convex on $(\mathcal{M} \times \mathcal{M}, g \times g)$ for some constant λ.

(NL3) $W \in C^1(\mathbb{R}^d)$ and $W(x-y) \leq C\left(1 + \mathrm{dist}^2(x,x_0) + \mathrm{dist}^2(y,x_0)\right)$ for some $C > 0$ and all $x, y \in \mathcal{M}$.

(NL4) $$\liminf_{\mathrm{dist}((x,y),(x_0,x_0))\to\infty} \frac{W(x-y)}{\mathrm{dist}^2(x,x_0) + \mathrm{dist}^2(y,x_0)} \geq 0.$$

(NL5) V is λ-geodesically convex of $(\mathcal{M}, g)$.

(NL6) $V \in C^1(\mathcal{M})$ and $V(x) \leq C\left(1 + \mathrm{dist}^2(x,x_0)\right)$ for all $x \in \mathcal{M}$.

(NL7) $$\liminf_{\mathrm{dist}(x,x_0)\to\infty} \frac{V(x)}{\mathrm{dist}^2(x,x_0)} \geq 0.$$

We list some remarks and direct consequences of the conditions. One can replace the condition (NL3) by the condition that $W \in C^1(\mathbb{R}^d\backslash\{0\})$, W has local minimum at $x = 0$ and satisfies the quadratic growth condition as in [28]. In this case the minimal subdifferential, defined in (3.3.2), has the form $\partial^o\mathcal{E}(\mu) = -P_x\left(-A(x)\left(\partial^o W * \mu + \nabla V\right)\right)$ where $\partial^o W * \mu(x) = \int_{y\neq x} \nabla W(x-y)d\mu(y)$ as defined in [28].

In (NL2) and (NL5), w and V may have different constants for convexity, but we assume that the constants are the same since we can take the minimum of the two constants if necessary.

Conditions (NL2) and (NL3) imply the following linear growth condition on ∇W, $\forall (x_1, y_1) \in \mathcal{M}\times\mathcal{M}$

$$\begin{aligned}\langle A(x_1)\nabla W(x_1-y_1), \nabla W(x_1-y_1)\rangle + \langle A(y_1)\nabla W(x_1-y_1), \nabla W(x_1-y_1)\rangle \\ \leq C\left(1+\operatorname{dist}^2(x_1,x_0)+\operatorname{dist}^2(y_1,x_0)\right).\end{aligned} \tag{3.1.6}$$

Similarly, (NL5) and (NL6) imply the linear growth condition on ∇V, $\forall x \in \mathcal{M}$

$$\langle A(x)\nabla V(x), \nabla V(x)\rangle \leq C(1+\operatorname{dist}^2(x,x_0)). \tag{3.1.7}$$

To see that, for ∇V we notice that

$$\begin{aligned}C\left(1+\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)\right) &\geq V(y)-V(x) \\ &\geq \langle \nabla V(x), T(x,y)\rangle + \frac{\lambda}{2}\operatorname{dist}^2(x,y),\end{aligned}$$

where $T(x,y)$ is the tangent vector at x such that $\exp_x(T(x,y)) = y$ and $|T(x,y)|_g = \operatorname{dist}(x,y)$, which we define in (3.3.1) in Section 3.3. So

$$C\left(1+\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(x,y)\right) \geq \langle \sqrt{A(x)}\nabla V(x), \sqrt{G(x)}T(x,y)\rangle + \frac{\lambda}{2}\operatorname{dist}^2(x,y).$$

$\left\langle \sqrt{G(x)}T(x,y), \sqrt{G(x)}T(x,y)\right\rangle = \operatorname{dist}^2(x,y)$, by taking $\operatorname{dist}(x,y) = \max\{1, \operatorname{dist}(x,x_0)\}$, we get

$$\langle A(x)\nabla V(x), \nabla V(x)\rangle \leq C\left(1+\operatorname{dist}^2(x,x_0)\right).$$

Similar calculations give the growth conditions (3.1.6) on ∇W.

From calculations of Section 3.6, we know that if W and V satisfy (NL2) and (NL5), then $\mathcal{E}$ is geodesically (semi-)convex on $\mathcal{P}_2(\mathcal{M})$ with convexity constant 2λ according to Definition 2.2.6.

Remark 3.1.1. *(Simple conditions for (NL2) and (NL5)) In Section 3.6, we give detailed calculations and precise conditions on W, V and g which guarantee λ-geodesic convexity of V and W. Here we summarize some conclusions.*

- *If there exist constants $c_1 > 0, c_2 > 0$ such that the Riemannian metric $g \in C^1(\mathcal{M})$ with $c_1 I_d \leq G(x) \leq \frac{1}{c_1} I_d$, $|\frac{\partial}{\partial x_k} G_{ij}(x)| \leq \frac{1}{c_1}$ for all $x \in \mathcal{M}$ and W is twice differentiable with $|\nabla W(y)| \leq c_2$, $\operatorname{Hess} W(y) \geq -c_2 I_d$ for all $y \in \mathcal{M}-\mathcal{M} := \{x^1 - x^2 : x^1, x^2 \in \mathcal{M}\}$, then $w(x,y) = W(x-y)$ is λ-geodesically convex on $(\mathcal{M}\times\mathcal{M}, g\times g)$.*

- *For V, if there exists a constant $c_1 > 0$ such that the Riemannian metric $g \in C^1(\mathcal{M})$ satisfies $c_1 I_d \le G(x) \le \frac{1}{c_1} I_d$, $\left|\frac{\partial}{\partial x_k} G_{ij}(x)\right| \le \frac{1}{c_1}$ and $V \in C^2(\mathcal{M})$ satisfies that $|\nabla V(y)| \le c_2$, $\operatorname{Hess} V(y) \ge -c_2 I_d$ for all $y \in \mathcal{M}$, then V is λ-geodesically convex on $(\mathcal{M}, g)$.*

- *In general, the conditions on g, W, V to guarantee λ-geodesic convexity of w, V are more stringent than in the Euclidean space. For example: assuming $g \in C^1(\mathcal{M})$ such that $c_1 I_d \le G(x) \le \frac{1}{c_1} I_d$, $\left|\frac{\partial}{\partial x_k} G_{ij}\right| \le \frac{1}{c_1}$ and $\operatorname{Hess} V(y) \ge -c_2 I_d$, $\operatorname{Hess} W(y) \ge -c_2 I_d$ for some constants $c_1 > 0, c_2 > 0$ does not imply λ-geodesic convexity of the energy. We present an explicit example in Section 3.6 (Example 3.6.3) to show that.*

For manifolds satisfying (M1) and (M2) and potentials W, V satisfying (NL1)-(NL7) we consider (3.1.5) as a gradient flow of $\mathcal{E}$ in space of probability measures endowed with the Riemannian Wasserstein metric. Following the notations introduced in Chapter 2, in this Chapter and the next Chapter 4 we define the Riemannian 2-Wasserstein metric

$$d_W^2(\nu, \mu) = \min\left\{\int_{\mathcal{M}\times\mathcal{M}} \operatorname{dist}^2(x,y)\,d\gamma(x,y) : \gamma \in \Gamma(\mu,\nu)\right\} \tag{3.1.8}$$

and the usual Euclidean 2-Wasserstein metric

$$d_{W,Euc}^2(\nu, \mu) = \min\Big\{\int_{\mathcal{M}\times\mathcal{M}} |x-y|^2 d\gamma(x,y) : \gamma \in \Gamma(\mu,\nu)\Big\},$$

where $\Gamma(\mu,\nu)$ is the set of joint probability distributions on $\mathcal{M}\times\mathcal{M}$ with first marginal μ and second marginal ν.

Denote the set of optimal transport plans between μ and ν with respect to the Riemannian 2-Wasserstein metric d_W by $\Gamma_o(\mu,\nu)$, that is

$$\Gamma_o(\mu,\nu) = \left\{\gamma \in \Gamma(\mu,\nu) : \int_{\mathcal{M}\times\mathcal{M}} \operatorname{dist}^2(x,y)\,d\gamma(x,y) = d_W^2(\mu,\nu)\right\}. \tag{3.1.9}$$

We recall some notions introduced in Chapter 2 in our Riemannian setting. In particular, local slope of $\mathcal{E}$ with respect to the Riemannian 2-Wasserstein metric is defined as is defined as

$$|\partial\mathcal{E}|(\mu) = \limsup_{\nu\to\mu} \frac{(\mathcal{E}(\mu) - \mathcal{E}(\nu))^+}{d_W(\mu,\nu)}. \tag{3.1.10}$$

For a locally absolutely continuous curve $[0,+\infty) \ni t \mapsto \mu(t) \in \mathcal{P}_2(\mathcal{M})$ with respect to Riemannian 2-Wasserstein metric d_W, we denote its metric derivative by

$$|\mu'|(t) = \limsup_{s\to t} \frac{d_W(\mu(t),\mu(s))}{|s-t|}. \tag{3.1.11}$$

We call a locally absolutely continuous curve $[0,+\infty) \ni t \mapsto \mu(t) \in \mathcal{P}_2(\mathcal{M})$ a gradient flow with respect to the energy functional $\mathcal{E}$ if for a.e. $t > 0$,

$$v(t) \in -\partial\mathcal{E}\left(\mu(t)\right)$$

where $\partial\mathcal{E}(\mu(t))$ is the set of subdifferential of $\mathcal{E}$ at $\mu(t)$ and $v(t)$ is the tangent velocity of the curve at $\mu(t)$, which we define in Section 3.3 and Section 3.5.

Define the weak measure solutions to the continuity equation by

Definition 3.1.2. *A locally absolutely continuous curve $\mu : [0,+\infty) \mapsto (\mathcal{P}_2(\mathcal{M}), d_W)$ is a* weak measure solution *to (3.1.5) with initial value μ_0 if*

$$P\left(-A(x)\left(\int_{\mathcal{M}} \nabla W(x-y)d\mu(t,y) + \nabla V(x)\right)\right) \in L^1_{loc}([0,+\infty); L^2(g,\mu(t)))$$

and

$$\begin{aligned}&\int_0^\infty \int_{\mathcal{M}} \frac{\partial \phi}{\partial t}(t,x)d\mu(t,x)dt + \int_{\mathcal{M}} \phi(0,x)d\mu_0(x)\\ &= -\int_0^\infty \int_{\mathcal{M}} \left\langle \nabla\phi(t,x), P\left(-A(x)\left(\int_{\mathcal{M}} \nabla W(x-y)d\mu(t,y) + \nabla V(x)\right)\right)\right\rangle d\mu(t,x)dt\end{aligned} \tag{3.1.12}$$

for all $\phi \in C_c^\infty([0,\infty)\times\mathcal{M})$.

Above we consider $C_c^\infty([0,\infty)\times\mathcal{M})$ to be the set of restrictions of functions in $C_c^\infty([0,\infty)\times\mathbb{R}^d)$ to $\mathcal{M}$. In particular we note that the values of test functions on the boundary of $\mathcal{M}$ may be different from zero. In this way the no-flux boundary conditions are imposed.

The main results of this Chapter are the following theorems regarding existence and stability of gradient flows with arbitrary initial data $\mu_0 \in \mathcal{P}_2(\mathcal{M})$, which we prove in Section 4.2.

Theorem 3.1.3. *Assume (M1)-(M2) and (NL1)-(NL7), then for any $\mu_0 \in \mathcal{P}_2(\mathcal{M})$ there exists a locally absolutely continuous curve $[0,+\infty) \ni t \mapsto \mu(t) \in \mathcal{P}_2(\mathcal{M})$ such that $\mu(0) = \mu_0$ and $\mu(\cdot)$ is a gradient flow of $\mathcal{E}$ with respect to the Riemannian 2-Wasserstein metric d_W. $\mu(\cdot)$ satisfies that for a.e. $t \in (0,+\infty)$*

$$|\partial\mathcal{E}|^2(\mu(t)) = |\mu'|^2(t) = \int_{\mathcal{M}} g_x(\kappa(t,x),\kappa(t,x))\, d\mu(t,x) \tag{3.1.13}$$

and the energy dissipation equality, for $0 \le s \le t < \infty$

$$\mathcal{E}(\mu(s)) - \mathcal{E}(\mu(t)) = \int_s^t \int_{\mathcal{M}} g_x(\kappa(r,x),\kappa(r,x))\, d\mu(r,x)dr, \tag{3.1.14}$$

where we denote $\kappa(r,x) = -P(-A(x)(\nabla W * \mu(r)(x) + \nabla V(x)))$. Moreover, $\mu(\cdot)$ is a weak measure solution to (3.1.5) with initial data μ_0.

Theorem 3.1.4. *Suppose (M1)-(M2) and (NL1)-(NL7) hold true. Let $\mu^1(\cdot), \mu^2(\cdot)$ be two gradient flows of the energy functional $\mathcal{E}$ with initial data μ_0^1, μ_0^2 respectively, then*

$$d_W\left(\mu^1(t), \mu^2(t)\right) \le e^{-2\lambda t} d_W\left(\mu_0^1, \mu_0^2\right) \tag{3.1.15}$$

for any $t \geq 0$. Moreover, the gradient flow solution is characterized by the system of Evolution Variational Inequalities:

$$\frac{1}{2}\frac{d}{dt}d_W^2\left(\mu(t),\nu\right)+\lambda d_W^2\left(\mu(t),\nu\right)\leq\mathcal{E}(\nu)-\mathcal{E}\left(\mu(t)\right), \tag{3.1.16}$$

for a.e. $t > 0$ and for all $\nu \in \mathcal{P}_2(\mathcal{M})$.

Remarks and connections.

Remark 3.1.5. *Recall that for interaction and potential energy on $\mathcal{P}_2\left(\mathbb{R}^d\right)$, the gradient flow of $\mathcal{E}$ with respect to the usual Euclidean 2-Wasserstein metric would be*

$$\frac{\partial}{\partial t}\mu(t,x)-\operatorname{div}\left(\mu(t,x)\left(\int_{\mathbb{R}^d}\nabla W(x-y)d\mu(t,y)+\nabla V(x)\right)\right)=0. \tag{3.1.17}$$

Comparing with (3.1.5), we see that the projection P_x is due to the boundary of $\mathcal{M}$ and the mobility A comes from geometry of $\mathcal{M}$.

We define the set of admissible vector fields $\mathfrak{V}$ at μ to be the set of $L^2(\mu)$ sections of $T^{\text{in}}\mathcal{M}$. That is

$$\mathfrak{V}=\left\{v:\mathcal{M}\to T\mathcal{M}\mid\ (\forall x\in\mathcal{M})\ v(x)\in T_x^{\text{in}}\mathcal{M}\text{ and }\int_{\mathcal{M}}g_x(v(x),v(x))d\mu(x)<\infty\right\}. \tag{3.1.18}$$

Remark 3.1.6. *If we assume that $\mathcal{M}$ has no boundary and we use the Riemannian volume form in defining the probability measures on $\mathcal{M}$, then the gradient flow of $\mathcal{E}$ with respect to the Riemannian 2-Wasserstein metric is*

$$\frac{\partial}{\partial t}\mu(t,x)-\operatorname{div}_{\mathcal{M}}\left(\mu(t,x)\left(\int_{\mathcal{M}}\nabla_{\mathcal{M}}W(x-y)d\mu(t,y)+\nabla_{\mathcal{M}}V(x)\right)\right)=0, \tag{3.1.19}$$

where the divergence and gradient should be understood as the Riemannian divergence and gradient on $\mathcal{M}$, and when test against test functions, it should be integrated against the Riemannian volume form $d\omega(x)$. Writing (3.1.19) in local coordinates, we have

$$\frac{\partial}{\partial t}\mu(t,x)-\frac{1}{\sqrt{\det G(x)}}\operatorname{div}\left(\mu(t,x)\sqrt{\det G(x)}A(x)\left(\nabla W*\mu(t)(x)+\nabla V(x)\right)\right)=0,$$

where the divergence is the Euclidean divergence now. We note that the equation above can be reduced to the form (3.1.5)*. Namely the measure $\tilde{\mu}$ defined by $d\tilde{\mu}(t,x)=\sqrt{\det G(x)}d\mu(t,x)$ solves*

$$\frac{\partial}{\partial t}\tilde{\mu}(t,x)-\operatorname{div}\left(\tilde{\mu}(t,x)A(x)\left(\int_{\mathcal{M}}\nabla W(x-y)d\tilde{\mu}(t,y)+\nabla V(x)\right)\right)=0,$$

which is exactly (3.1.5) without the projection P. So it is similar to consider the gradient flow of $\mathcal{E}$ under the Riemannian and Euclidean volume form. Consequently, (NL1)-(NL7) also imply the existence of the gradient flow of $\mathcal{E}$ with respect to the Riemannian volume form.

Outline.

In Section 3.2, we establish some important properties of the functional $\mathcal{E}$ and the manifold $\mathcal{M}$, in particular the lower semicontinuity of $\mathcal{E}$.

In Section 3.3, we give the definition of subdifferential in the manifold context, which is a natural generalization of the subdifferential in the Euclidean setting. We then identify the minimal subdifferential of $\mathcal{E}$ at μ as

$$\partial^o \mathcal{E}(\mu) = -P_x \left(-A(x) \left(\int_{\mathcal{M}} \nabla W(x-y) d\mu(y) + \nabla V(x) \right) \right).$$

Section 3.4 is devoted to the JKO scheme. We show that the discrete scheme is well-posed and converges to a locally absolutely continuous curve $\mu(t) \in \mathcal{P}_2(\mathcal{M})$. Together with the fact the local slope $|\partial \mathcal{E}|$ is lower semicontinuous, we show that the limit curve $\mu(t)$ is a curve of maximal slope.

In Section 3.5, we establish that the limit curve $\mu(t)$ we get from JKO scheme is actually a gradient flow, thus a weak measure solution to the continuity equation (3.1.5). We then show that geodesic (semi-)convexity of the functional $\mathcal{E}$ implies uniqueness and stability of gradient flow solutions. We remark that the lack of existence of an appropriate flow map due to discontinuity of the velocity fields, makes the proof of differentiability of Wasserstein metric more involved (Lemma 3.5.3).

In Section 3.6, we give some examples of manifolds $(\mathcal{M}, g)$, external and interaction potentials V, W for which V, w are λ-geodesically convex on $(\mathcal{M}, g)$ and $(\mathcal{M} \times \mathcal{M}, g \times g)$ respectively. These imply that functional $\mathcal{E}$ is gedesically (semi-)convex on $(\mathcal{P}_2(\mathcal{M}), d_W)$.

In the last Section 3.7, we present some numerical simulations showing that rolling swarms emerge naturally in biological aggregations in heterogeneous environments.

3.2 Some properties of $\mathcal{E}$ and $\mathcal{M}$

In this Section, we show some basic properties of the functionals $\mathcal{V}, \mathcal{W}$ and the manifold $\mathcal{M}$, which we need in the subsequent sections. First, we show the following simple relation between the distances of two points with respect to the Euclidean and Riemannian metric:

Lemma 3.2.1. *For any $x, y \in \mathcal{M}$,*

$$\operatorname{dist}^2(x, y) \geq \Lambda |x-y|^2. \tag{3.2.1}$$

Proof. Assume that $\gamma(t)$ is a curve which realizes the Riemannian distance between x and y and $\gamma(0) = x, \gamma(1) = y$, then we have

$$\operatorname{dist}^2(x, y) = \int_0^1 g_{\gamma(t)} \left(\gamma'(t), \gamma'(t) \right) dt \geq \int_0^1 \Lambda \langle \gamma'(t), \gamma'(t) \rangle dt \geq \Lambda |x-y|^2.$$

□

We can now compare the Wasserstein distance under Euclidean and Riemannian metric,

Lemma 3.2.2. *For two Borel probability measures μ and ν, we have*

$$d_W^2(\mu,\nu) \geq \Lambda d_{W,Euc}^2(\mu,\nu).$$

Proof. Assume that γ is the optimal transportation plan between μ and ν, that is $\gamma \in \Gamma_o(\mu,\nu)$. Then

$$d_W^2(\mu,\nu) = \int_{\mathcal{M}\times\mathcal{M}} \operatorname{dist}(x,y)^2 d\gamma(x,y) \geq \Lambda \int_{\mathcal{M}\times\mathcal{M}} |x-y|^2 d\gamma(x,y) \geq \Lambda d_{W,Euc}^2(\mu,\nu).$$

□

Now we turn to the properties of $\mathcal{W}$ and $\mathcal{V}$.

Proposition 3.2.3 (Lower semicontinuity of $\mathcal{W}$)**.** *Assume (NL1)-(NL4), then*

$$\liminf_{n\to\infty} \mathcal{W}(\mu_n) \geq \mathcal{W}(\mu), \tag{3.2.2}$$

given that μ_n narrowly converge to μ and μ_n have uniformly bounded second moments.

Proof. By (NL4), $\liminf_{\operatorname{dist}(x,x_0)+\operatorname{dist}(y,x_0)\to\infty} \frac{W(x-y)}{\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)} \geq 0$, for any $\varepsilon > 0$, there exists $R > 0$ such that

$$\frac{W(x-y)}{\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)} > -\varepsilon$$

for all $(x,y) \in \mathcal{M}\times\mathcal{M}$ such that $\operatorname{dist}(x,x_0)+\operatorname{dist}(y,x_0) \geq R$. Thus $W(x-y) + \varepsilon\left(\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)\right)$ is continuous and bounded from below. By Lemma 5.1.7 from [5], we know

$$\begin{aligned}
&\liminf_{n\to\infty} \int_{\mathcal{M}\times\mathcal{M}} \left(W(x-y) + \varepsilon\left(\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)\right)\right) d\mu_n(x)d\mu_n(y) \\
&\quad \geq \int_{\mathcal{M}\times\mathcal{M}} \left(W(x-y) + \varepsilon\left(\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)\right)\right) d\mu(x)d\mu(y),
\end{aligned}$$

which implies

$$\begin{aligned}
\int_{\mathcal{M}\times\mathcal{M}} W(x-y)d\mu(x)d\mu(y) \leq{} & \liminf_{n\to\infty} \int_{\mathcal{M}\times\mathcal{M}} W(x-y)d\mu_n(x)d\mu_n(y) \\
&+ \limsup_{n\to\infty} \int_{\mathcal{M}\times\mathcal{M}} \varepsilon\left(\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)\right) d\mu_n(x)d\mu_n(y)
\end{aligned}$$

On the other hand,

$$\int_{\mathcal{M}\times\mathcal{M}} \varepsilon\left(\operatorname{dist}^2(x,x_0)+\operatorname{dist}^2(y,x_0)\right) d\mu_n(x)d\mu_n(y) \leq 2\varepsilon C$$

where $C = \sup_n \int_{\mathcal{M}} \text{dist}^2(x, x_0) d\mu_n(x) < \infty$. Taking $\varepsilon \to 0^+$ yields

$$\int_{\mathcal{M}\times\mathcal{M}} W(x-y)d\mu(x)d\mu(y) \leq \liminf_{n\to\infty} \int_{\mathcal{M}\times\mathcal{M}} W(x-y)d\mu_n(x)d\mu_n(y).$$

$\square$

For $\mathcal{V}$ the following lower semicontinuity result holds:

Proposition 3.2.4 (Lower semicontinuity of $\mathcal{V}$)**.** *Assume (NL5)-(NL7), then*

$$\liminf_{n\to\infty} \mathcal{V}(\mu_n) \geq \mathcal{V}(\mu), \tag{3.2.3}$$

given that μ_n narrowly converge to μ and μ_n have uniform bounded second moments.

The proof is analogous to the proof of lower semicontinuity of $\mathcal{W}$ and we omit it here.

We list some properties and observations about the projection P:

- For any tangent vector field v in $L^2(\mu)$, $Pv \in \mathfrak{V}$.
- In general for $v, w \in T_x\mathcal{M}$, $P(c_1 v + c_2 w) \neq c_1 Pv + c_2 Pw$ and $g(Pv, w) \neq g(v, Pw)$.
- For any $v, w \in T_x\mathcal{M}$, $|Pv - Pw|_g \leq |v - w|_g$.
- P can break the continuity of the velocity field. In particular if μ_n and μ are absolutely continuous curves in $\mathcal{P}_2(\mathcal{M})$ and v_n and v are corresponding velocities such that $\mu_n(t)$ converges narrowly to $\mu(t)$ then in the Euclidean setting (with no boundary) $v^n d\mu^n$ converges weakly to $v d\mu$, as was shown in Lemma 2.7 from [28]. However this statement does not hold when boundary is present. Thus we need to use a different method to show the lower semicontinuity of the local slope $|\partial\mathcal{E}|$, which we do that in Theorem 3.4.3.
- Even though P is non-linear and breaks continuity, we still have that: The function $\mathcal{M} \times \mathbb{R}^d \ni (x, \xi) \mapsto g_x(P_x\xi, P_x\xi)$ is lower semicontinuous and for all $x \in \mathcal{M}$, the function $\mathbb{R}^d \ni \xi \mapsto g_x(P_x\xi, P_x\xi)$ is convex. Refer to Proposition 3.4.5 for the proof.

3.3 Minimal subdifferential of $\mathcal{E}$

In this Section, we give the definition of subdifferential in the Riemannian geometric setting, which is the natural generalization of the notion in the Euclidean setting. We then identify the minimal subdifferential of $\mathcal{E}$ as $\partial^o\mathcal{E}(\mu) = -P(-A(\nabla W * \mu + \nabla V))$ and show that it realizes the local slope in the sense that $|\partial\mathcal{E}|(\mu) = \|\partial^o\mathcal{E}(\mu)\|_{L^2(g,\mu)}$.

In order to define the subdifferential of a functional in the Riemannian setting, we introduce the exponential map on the space of configurations first. Let $\exp_x : T_x\mathcal{M} \to \mathcal{M}$

be the exponential map on $\mathcal{M}$. It is understood that the domain of $\exp_x$ is actually a subset of $T_x\mathcal{M}$ for which the geodesics of appropriate length and direction exists. We note that if x is in the interior of $\mathcal{M}$ then the domain of $\exp_x$ is an open neighborhood of 0, while if $x \in \partial\mathcal{M}$ then the domain of $\exp_x$ a subset of $T_x^{\text{in}}\mathcal{M}$ and may not be an open neighborhood of 0 even in $T_x^{\text{in}}\mathcal{M}$. For example if $\mathcal{M} = B(0,1)$, g is the Euclidean metric, $x = (1,0)$, and $\xi = (0,1)$, then $\xi \in T_x^{\text{in}}\mathcal{M}$, but $\exp(t\xi)$ is not defined for any $t \neq 0$. This required us to modify a number of standard arguments so that we do not use the exponential map to generate geodesics. We only use the exponential map to parameterize the geodesics which we know to exist.

By our assumptions on $(\mathcal{M}, g)$ we know that there exists a length minimizing geodesics connecting any two points. The problem is that such geodesics may not be unique. However, by Aumann measurable selection theorem, see [53], geodesics can be selected in a measurable way. More precisely there exists a Borel measurable function $T : \mathcal{M} \times \mathcal{M} \to T^{\text{in}}\mathcal{M}$ such that for all $x, y \in \mathcal{M}$

$$\exp_x(T(x,y)) = y \tag{3.3.1}$$

and such that $\gamma(t) = \exp_x(tT(x,y))$, $t \in [0,1]$ gives a minimal geodesic connecting x and y. Note that $g_x(T(x,y), T(x,y)) = \text{dist}^2(x,y)$. Unless otherwise specified, in the remainder of the paper, by T we denote an arbitrary Borel measurable function satisfying the above conditions.

Definition 3.3.1 (Subdifferential)**.** *Fix $\mu \in \mathcal{P}_2(\mathcal{M})$, a vector field $\xi \in L^2(g,\mu)$ is said to be an element of the subdifferential of $\mathcal{E}$ at μ, and we denote as $\xi \in \partial\mathcal{E}(\mu)$, if there exists $T : \mathcal{M} \times \mathcal{M} \to T\mathcal{M}$ as in* (3.3.1) *such that*

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \geq \inf_{\gamma \in \Gamma_o(\mu,\nu)} \int_{\mathcal{M}\times\mathcal{M}} g_x\left(\xi(x), T(x,y)\right) d\gamma(x,y) + o\left(d_W(\mu,\nu)\right), \tag{3.3.2}$$

where $\Gamma_o(\mu,\nu)$ is the set of optimal plans between μ and ν as defined in (3.1.9).

Remark 3.3.2. *The definition of the subdifferential applies to general energy functionals $\mathcal{E}$, not only geodesically (semi-)convex ones. For a λ-geodesically convex energy functional $\mathcal{E}$, if a vector field $\xi \in L^2(g,\mu)$ is an element of the subdifferential of $\mathcal{E}$ at μ then in fact there exists $T : \mathcal{M} \times \mathcal{M} \to T\mathcal{M}$ as in* (3.3.1) *such that*

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \geq \inf_{\gamma \in \Gamma_o(\mu,\nu)} \int_{\mathcal{M}\times\mathcal{M}} g_x\left(\xi(x), T(x,y)\right) d\gamma(x,y) + \frac{\lambda}{2} d_W^2(\mu,\nu).$$

This is the case in our problem since the energy functional $\mathcal{E}$ (defined in (2.3.8)) we are interested in is 2λ-geodesically convex.

We denote the element in $\partial\mathcal{E}(\mu)$ with minimal $L^2(g,\mu)$ norm by $\partial^o\mathcal{E}(\mu)$.

Remark 3.3.3. *Notice that Definition 3.3.1 reduces to the usual definition of subdifferential when g is the Euclidean metric. It is straightforward calculation to show that if $\xi \in \partial\mathcal{E}(\mu)$ then*

$$|\partial\mathcal{E}|(\mu) \leq \|\xi\|_{L^2(g,\mu)}, \tag{3.3.3}$$

where $\|\xi\|^2_{L^2(g,\mu)} = \int_{\mathcal{M}} g_x\left(\xi(x), \xi(x)\right) d\mu(x)$.

We now give the following main theorem of this section regarding the existence of subdifferential and the minimal $L^2(g,\mu)$ element of the subdifferential.

Theorem 3.3.4. *Assume (M1)-(M2), (NL1)-(NL7) hold, then $\partial\mathcal{E}(\mu) \neq \emptyset$ for any $\mu \in \mathcal{P}_2(\mathcal{M})$. Moreover the vector field*

$$\kappa(x) = -P_x\left(-A(x)\left(\int_{\mathcal{M}} \nabla W(x-y)d\mu(y) + \nabla V(x)\right)\right) \tag{3.3.4}$$

is the unique element of minimal $L^2(g,\mu)$-norm in $\partial\mathcal{E}(\mu)$ with

$$|\partial\mathcal{E}|(\mu) = \|\kappa\|_{L^2(g,\mu)}. \tag{3.3.5}$$

Remark 3.3.5. *To consider interaction potentials $W \in C^1(\mathbb{R}^d \setminus \{0\})$, one needs to notice that $0 \in \partial W(0)$. The proof of the above theorem can be used to show that the minimal subdifferential is*

$$\partial^o\mathcal{E}(\mu) = -P\left(-A\left(\partial^o W * \mu + \nabla V\right)\right), \tag{3.3.6}$$

where $\partial^o W(x) = \nabla W(x)$ if $x \neq 0$ and $\partial^o W(0) = 0$.

We also remark that while in the definition of subdifferential Definition 3.3.1, we only require (3.3.2) to hold for some measurable choice of $T(x,y)$ and infimum over $\gamma \in \Gamma_o(\mu,\nu)$, in the proof we actually show that for any $\gamma \in \Gamma_o(\mu,\nu)$ and any measurable selection $T(x,y)$, (3.3.2) holds true with that particular choice of $T(x,y)$ and γ.

To simplify notations, we denote $\nabla W * \mu(x) = \int_{\mathcal{M}} \nabla W(x-y)d\mu(y)$.

Before proving the theorem, we need the following

Lemma 3.3.6. *Let ξ be a vector field in $\mathfrak{V}$ such that there exists $t_0 > 0$ for which $\exp_x(t\xi(x)) \in \mathcal{M}$ for all $0 \leq t \leq t_0$ and μ-a.e. $x \in \mathcal{M}$. Then*

$$\limsup_{t\to 0^+} \frac{d_W\left(\left(\exp(t\xi)\right)_\sharp \mu, \mu\right)}{t} \leq \|\xi\|_{L^2(g,\mu)}, \tag{3.3.7}$$

where we denote $\exp(t\xi)(x) = \exp_x(t\xi(x))$.

Proof of Lemma. For $0 \le t < t_0$, notice that $(id, \exp(t\xi))_\sharp \mu \in \Gamma\left(\mu, (\exp(t\xi))_\sharp \mu\right)$, so

$$d_W^2\left(\mu, (\exp(t\xi))_\sharp \mu\right) \le \int_{\mathcal{M}} \mathrm{dist}^2\left(x, \exp_x\left(t\xi(x)\right)\right) d\mu(x) \\ \le \int_{\mathcal{M}} t^2 g_x\left(\xi(x), \xi(x)\right) d\mu(x).$$

Thus

$$\limsup_{t\to 0^+} \frac{d_W^2\left(\mu, (\exp(t\xi))_\sharp \mu\right)}{t^2} \le \int_{\mathcal{M}} g_x\left(\xi(x), \xi(x)\right) d\mu(x).$$

□

We now prove the theorem.

Proof of Theorem. We divide the proof into two steps.
Step 1. $\kappa \in \partial\mathcal{E}(\mu)$. We need to prove that

$$\int_{\mathcal{M}} g_x\left(\kappa(x), \kappa(x)\right) d\mu(x) < \infty$$

and

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \ge \inf_{\gamma\in\Gamma_o(\mu,\nu)} \int_{\mathcal{M}\times\mathcal{M}} g_x\left(\kappa(x), T(x,y)\right) d\gamma(x,y) + o\big(d_W(\mu,\nu)\big).$$

To prove the first claim, note that

$$\begin{aligned}
&\int_{\mathcal{M}} g_x\left(\kappa(x), \kappa(x)\right) d\mu(x)\\
&\le \int_{\mathcal{M}} g\left(A(x)\left(\nabla W * \mu(x) + \nabla V(x)\right), A(x)\left(\nabla W * \mu(x) + \nabla V(x)\right)\right) d\mu(x)\\
&= \int_{\mathcal{M}} \left\langle A(x)\left(\nabla W * \mu(x) + \nabla V(x)\right), \nabla W * \mu(x) + \nabla V(x)\right\rangle d\mu(x)\\
&\le \int_{\mathcal{M}\times\mathcal{M}} \left\langle A(x)\left(\nabla W(x-y) + \nabla V(x)\right), \nabla W(x-y) + \nabla V(x)\right\rangle d\mu(y)d\mu(x)\\
&\le \int_{\mathcal{M}\times\mathcal{M}} C\left(1 + \mathrm{dist}^2(x,x_0) + \mathrm{dist}^2(y,x_0)\right) d\mu(x)d\mu(y)\\
&< \infty.
\end{aligned}$$

The first inequality above comes from the fact that projection does not increase the length of a vector, while the third inequality holds because ∇W and ∇V have linear growth, as shown in (3.1.6) and (3.1.7). The last inequality holds since μ has finite second moment.

To prove the second claim let $\mu, \nu \in \mathcal{P}_2(\mathcal{M}), \gamma \in \Gamma_o(\mu,\nu)$ be any optimal plan and $T(x,y)$ be as in (3.3.1). Due to λ-convexity of W and V, for fixed $x_1, x_2, y_1, y_2 \in \mathcal{M}$ the function

$$f(t) = \frac{W\left(\exp_{x_1}\left(tT(x_1,y_1)\right) - \exp_{x_2}\left(tT(x_2,y_2)\right)\right) - W(x_1-x_2)}{2t} \\ + \frac{2V\left(\exp_{x_2}\left(tT(x_2,y_2)\right)\right) - 2V(x_2)}{2t} - \frac{\lambda}{2} t\,\mathrm{dist}^2(x_2,y_2) - \frac{\lambda}{4} t\,\mathrm{dist}^2\left((x_1,x_2),(y_1,y_2)\right) \tag{3.3.8}$$

is non-decreasing on $[0,1]$, so $f(1) \geq \liminf_{t\to 0^+} f(t)$. We remark here that the fact that the curve $t \mapsto \exp_{x_1}(tT(x_1,y_1)) - \exp_{x_2}(tT(x_2,y_2))$ is no longer a geodesic on $(\mathcal{M},g)$ is the reason why we need to assume (NL2) of W, i.e. the λ-geodesic convexity of $(x,y) \in \mathcal{M}\times\mathcal{M} \mapsto w(x,y) = W(x-y)$ instead of λ-geodesic convexity of $x \in \mathcal{M} \mapsto W(x)$ as in the Euclidean setting. Note that

$$\lim_{t\to 0^+}\left[\frac{W\left(\exp_{x_1}\left(tT(x_1,y_1)\right) - \exp_{x_2}\left(tT(x_2,y_2)\right)\right) - W(x_1-x_2)}{2t}\right. \\ \left. - \frac{\lambda}{2} t\,\mathrm{dist}^2\left((x_1,x_2),(y_1,y_2)\right)\right] \\ = \frac{1}{2}\left\langle \nabla W(x_1-x_2), T(x_1,y_1) - T(x_2,y_2)\right\rangle,$$

and

$$\lim_{t\to 0^+}\left[\frac{V\left(\exp_{x_2}\left(tT(x_2,y_2)\right)\right) - V(x_2)}{t} - \frac{\lambda}{2} t\,\mathrm{dist}^2(x_2,y_2)\right] = \left\langle \nabla V(x_2), T(x_2,y_2)\right\rangle.$$

Then integrating over $d\gamma(x_1,y_1)d\gamma(x_2,y_2)$ gives

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \\ = \int_{\mathcal{M}\times\mathcal{M}}\int_{\mathcal{M}\times\mathcal{M}} \frac{W(y_1-y_2) + 2V(y_2) - W(x_1-x_2) - 2V(x_2)}{2} d\gamma(x_1,y_1)d\gamma(x_2,y_2) \\ \geq \int_{\mathcal{M}\times\mathcal{M}}\int_{\mathcal{M}\times\mathcal{M}} \left[\frac{1}{2}\left\langle \nabla W(x_1-x_2), T(x_1,y_1) - T(x_2,y_2)\right\rangle + \left\langle \nabla V(x_2), T(x_2,y_2)\right\rangle\right] \\ d\gamma(x_1,y_1)d\gamma(x_2,y_2) + \lambda d_W^2(\mu,\nu) \\ = \int_{\mathcal{M}\times\mathcal{M}}\int_{\mathcal{M}\times\mathcal{M}} \left\langle \nabla W(x_2-x_1) + \nabla V(x_2), T(x_2,y_2)\right\rangle d\gamma(x_1,y_1)d\gamma(x_2,y_2) + \lambda d_W^2(\mu,\nu) \\ = \int_{\mathcal{M}\times\mathcal{M}} \left\langle \nabla W * \mu(x_2) + \nabla V(x_2), T(x_2,y_2)\right\rangle d\gamma(x_2,y_2) + \lambda d_W^2(\mu,\nu) \\ = \int_{\mathcal{M}\times\mathcal{M}} g_{x_2}\left(A(x_2)\left(\nabla W * \mu(x_2) + \nabla V(x_2)\right), T(x_2,y_2)\right) d\gamma(x_2,y_2) + \lambda d_W^2(\mu,\nu) \\ \geq -\int_{\mathcal{M}\times\mathcal{M}} g_{x_2}\left(P_{x_2}\left(-A(x_2)\left(\nabla W * \mu(x_2) + \nabla V(x_2)\right)\right), T(x_2,y_2)\right) d\gamma(x_2,y_2) + \lambda d_W^2(\mu,\nu) \\ = \int_{\mathcal{M}\times\mathcal{M}} g_{x_2}\left(\kappa(x_2), T(x_2,y_2)\right) d\gamma(x_2,y_2) + \lambda d_W^2(\mu,\nu)$$

where the second inequality comes from the fact that: If $x_2 \notin \partial\mathcal{M}$, by definition of P_{x_2} the inequality becomes an equality while if $x_2 \in \partial\mathcal{M}$, then by definition of P_{x_2}

$$
\begin{aligned}
&g_{x_2}\left(A(x_2)\left(\int_{\mathcal{M}} \nabla W(x_2-x_1)d\mu(x_1)+\nabla V(x_2)\right),\xi\right)\\
\geq &g_{x_2}\left(-P_{x_2}\left(-A(x_2)\left(\int_{\mathcal{M}} \nabla W(x_2-x_1)d\mu(x_1)+\nabla V(x_2)\right)\right),\xi\right)
\end{aligned}
$$

for any $\xi \in T^{\mathrm{in}}_{x_2}M$, and we notice that $T(x_2,y_2) \in T^{\mathrm{in}}_{x_2}\mathcal{M}$.

Step 2. κ is the element of minimal $L^2(g,\mu)$-norm in $\partial\mathcal{E}(\mu)$. By Remark 3.3.3, we only need to show $\|\kappa\|_{L^2(g,\mu)} \leq |\partial\mathcal{E}|(\mu)$. Consider first a vector field ξ as in Lemma 3.3.6, i.e. $\xi \in L^2(g,\mu)$ and $\exp_x(t\xi(x)) \in \mathcal{M}$ for all $x \in \mathcal{M}$ and $0 \leq t \leq t_0$,

$$
\begin{aligned}
&\lim_{t\to 0^+} \frac{\mathcal{E}\left(\exp(t\xi)_\sharp \mu\right)-\mathcal{E}(\mu)}{t}\\
&=\lim_{t\to 0^+}\frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} \frac{W\left(\exp_x(t\xi(x))-\exp_z(t\xi(z))\right)+2V\left(\exp_x(t\xi(x))\right)-W(x-z)-2V(x)}{t}\\
&d\mu(x)d\mu(z)\\
&=\frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} \langle \nabla W(x-z), \xi(x)-\xi(z)\rangle + 2\langle \nabla V(x),\xi(x)\rangle\, d\mu(x)d\mu(z)\\
&=\int_{\mathcal{M}}\left\langle \int_{\mathcal{M}} \nabla W(x-z)d\mu(z)+\nabla V(x),\xi(x)\right\rangle d\mu(x)\\
&=\int_{\mathcal{M}}\left\langle A(x)\left(\int_{\mathcal{M}} \nabla W(x-z)d\mu(z)+\nabla V(x)\right), G(x)\xi(x)\right\rangle d\mu(x)\\
&=\int_{\mathcal{M}} g_x\left(A(x)\left(\int_{\mathcal{M}} \nabla W(x-y)d\mu(y)+\nabla V(x)\right),\xi(x)\right) d\mu(x)
\end{aligned}
$$

given that we can prove the second equality. Denote $\gamma_x(t) = \exp_x(t\xi(x))$. By the linear growth condition (3.1.7) on ∇V

$$
\begin{aligned}
|V(\gamma_x(t))-V(\gamma_x(0))| &= \left|\int_0^t \langle \nabla V(\gamma_x(s)), \dot{\gamma}_x(s)\rangle\, ds\right|\\
&\leq \int_0^t \langle A(\gamma_x(s))\nabla V(\gamma_x(s)), \nabla V(\gamma_x(s))\rangle^{\frac{1}{2}} \langle G(\gamma_x(s))\dot{\gamma}_x(s), \dot{\gamma}_x(s)\rangle^{\frac{1}{2}}\, ds\\
&\leq \int_0^t C(1+\mathrm{dist}(\gamma_x(s),x))|\xi(x)|_g ds\\
&\leq \int_0^t C(1+s|\xi(x)|_g)|\xi(x)|_g ds\\
&\leq C(1+t_0|\xi(x)|_g^2)t.
\end{aligned}
$$

Thus

$$
\left|\frac{V(\exp_x(t\xi(x)))-V(x)}{t}\right| = \left|\frac{V(\gamma_x(t))-V(\gamma_x(0))}{t}\right| \leq C(1+t_0|\xi(x)|_g^2) \in L^1(g,\mu).
$$

Similarly for W, by the linear growth condition (3.1.6) on ∇W,

$$
\begin{aligned}
&|W(\gamma_x(t)-\gamma_z(t))-W(x-z)| \\
&=\left|\int_0^t \langle \nabla W\left(\gamma_x(s)-\gamma_z(s)\right), \dot{\gamma}_x(s)-\dot{\gamma}_z(s)\rangle\, ds\right| \\
&\le \int_0^t \Bigg(\langle A(\gamma_x(s))\nabla W(\gamma_x(s)-\gamma_z(s)), \nabla W(\gamma_x(s)-\gamma_z(s))\rangle^{\frac{1}{2}} \langle G(\gamma_x(s))\dot{\gamma}_x(s), \dot{\gamma}_x(s)\rangle^{\frac{1}{2}} \\
&+ \langle A(\gamma_z(s))\nabla W(\gamma_x(s)-\gamma_z(s)), \nabla W(\gamma_x(s)-\gamma_z(s))\rangle^{\frac{1}{2}} \langle G(\gamma_z(s))\dot{\gamma}_z(s), \dot{\gamma}_z(s)\rangle^{\frac{1}{2}} \Bigg) ds \\
&\le \int_0^t C\left(1+\mathrm{dist}(\gamma_x(s),x)+\mathrm{dist}(\gamma_z(s),z)\right)\left(|\xi(x)|_g+|\xi(z)|_g\right) ds \\
&\le C(1+t_0|\xi(x)|_g^2+t_0|\xi(z)|_g^2)t.
\end{aligned}
$$

Thus the second equality follows by Lebesgue's dominated convergence theorem.

By the definition of local slope (3.1.10) and Lemma 3.3.6

$$
\begin{aligned}
|\partial\mathcal{E}|(\mu)\|\xi\|_{L^2(g,\mu)} &\ge |\partial\mathcal{E}|(\mu)\liminf_{t\to 0^+}\frac{d_W\left(\exp(t\xi)\right)_\sharp\mu,\mu)}{t} \\
&\ge -\int_{\mathcal{M}}\left\langle A(x)\left(\int_{\mathcal{M}}\nabla W(x-z)d\mu(z)+\nabla V(x)\right), G(x)\xi(x)\right\rangle d\mu(x) \\
&= \int_{\mathcal{M}} g_x\left(-A(x)\left(\int_{\mathcal{M}}\nabla W(x-z)d\mu(z)+\nabla V(x)\right), \xi(x)\right) d\mu(x).
\end{aligned}
\tag{3.3.9}
$$

We need to plug $\xi=-\kappa$ into (3.3.9), however it is possible that there exists $x\in\partial\mathcal{M}$, such that there exists no $t_0>0$ with $\exp_x(-t\kappa(x))\in\mathcal{M}$ for all $0\le t\le t_0$. Thus we perform the following approximation scheme. For $n\in\mathbb{N}$, denote $\mathcal{M}_n=\{x\in\mathcal{M}:\mathrm{dist}(x,\partial\mathcal{M})>\frac{1}{n}\}$, $B_n=\{x\in\mathcal{M}:\mathrm{dist}(x,x_0)<n\}$ and $n(x)$ the outward normal direction with respect to the Rimmannian metric at $x\in\partial\mathcal{M}$. Define

$$
\xi_n(x)=\begin{cases}
-\kappa(x) & \text{if } x\in B_n\cap\mathcal{M}_n,\\
-\kappa(x)-\frac{1}{n}n(x) & \text{if } x\in B_n\cap\partial\mathcal{M},\\
0 & \text{Otherwise.}
\end{cases}
$$

We claim that ξ_n satisfies the conditions in Lemma 3.3.6 and ξ_n converges to $-\kappa$ in $L^2(g,\mu)$. Indeed, it is straightforward to see that $\xi_n\in L^2(g,\mu)$ and ξ_n converges to $-\kappa$ in $L^2(g,\mu)$. Since κ is continuous in $\mathcal{M}_n$ and $B_n\Subset\mathcal{M}$, we have $\|\kappa\|_{L^\infty(g,\mu)}\le C(n)$ on $B_n\cap\mathcal{M}_n$ and thus for $0\le t\le\frac{1}{nC(n)}$, $\exp_x\left(t\xi_n(x)\right)\in\mathcal{M}$ for $x\in B_n\cap\mathcal{M}_n$. For $x\in\overline{B_n\cap\partial\mathcal{M}}$, we know $g_x\left(\xi_n(x),n(x)\right)\le-\frac{1}{n}$ and $\overline{B_n\cap\partial\mathcal{M}}$ is compact, so there exists $\tilde{t}(n)$ such that $\exp_x\left(t\xi_n(x)\right)\in\mathcal{M}$ for all $0\le t\le\tilde{t}(n)$ and $x\in\overline{B_n\cap\partial\mathcal{M}}$. We can take $t_0=\min\{\frac{1}{nC(n)},\tilde{t}(n)\}$ and $\exp_x\left(t\xi_n(x)\right)\in\mathcal{M}$ for $0\le t\le t_0$ as claimed. Using ξ_n in (3.3.9) yields

$$|\partial\mathcal{E}|(\mu)\|\xi_n\|_{L^2(g,\mu)} \geq \int_{\mathcal{M}} g_x\left(-A(x)\left(\int_{\mathcal{M}}\nabla W(x-z)d\mu(z)+\nabla V(x)\right),\xi_n(x)\right)d\mu(x). \tag{3.3.10}$$

Since $g_x(\xi(x), P\xi(x)) = g_x(P\xi(x), P\xi(x))$, taking $n \to \infty$ then gives

$$\begin{aligned}|\partial\mathcal{E}|(\mu)\|\kappa\|_{L^2(g,\mu)} &\geq \int_{\mathcal{M}} g_x\left(-A(x)\left(\int_{\mathcal{M}}\nabla W(x-z)d\mu(z)+\nabla V(x)\right),-\kappa(x)\right)d\mu(x)\\ &= \int_{\mathcal{M}} g_x\left(\kappa(x),\kappa(x)\right)d\mu(x).\end{aligned}$$

Hence

$$\|\kappa\|_{L^2(g,\mu)} \leq |\partial\mathcal{E}|(\mu), \tag{3.3.11}$$

which completes the proof. □

3.4 JKO scheme: existence of minimizers and convergence

In this Section, we show the existence of curves of maximal slope with respect to $\mathcal{E}$. The general framework, developed in [5], uses the JKO scheme, which we describe in Subsection 2.2.1. We verify the conditions on the functional $\mathcal{E}$ needed to apply the general existence theorem of [5] to get a curve of maximal slope with respect to the relaxed local slope $|\partial^-\mathcal{E}|$. In order to show that the limit curve is a curve of maximal slope with respect to $|\partial\mathcal{E}|$, we proceed to prove that $\mathcal{P}_2(\mathcal{M}) \ni \mu \mapsto \int_s^t |\partial\mathcal{E}|^2(\mu(r))\,dr$ is lower semicontinuous with respect to narrow convergence of probability measures.

Recall the definitions of upper gradient in Definition 2.2.1 and curve of maximal slope in Definition 2.2.2. The general strategy of constructing curves of maximal slope is to use the JKO scheme as described in Subsection 2.2.1 from which we now recall some important notions.

Fix a time step $\tau > 0$ then μ_τ^k are define iteratively

$$\mu_\tau^{k+1} \in \operatorname{argmin}_{\mu\in\mathcal{P}_2(\mathcal{M})}\left[\frac{d_W^2(\mu,\mu_\tau^k)}{2\tau}+\mathcal{E}(\mu)\right], \tag{3.4.1}$$

with $\mu_\tau^0 = \mu_0$. We denote the piecewise constant interpolation by $\mu_\tau(\,\cdot\,)$. To be more precise, $\mu_\tau(0) = \mu_0$ and

$$\mu_\tau(t) = \mu_\tau^{k+1}, \tag{3.4.2}$$

if $k\tau < t \leq (k+1)\tau$ for $k \geq 0$. In order to show that the existence of minimizer in the minimization problem (3.4.1) and to show the convergence of the interplants $\mu_\tau(\,\cdot\,)$ (as $\tau \to 0^+$), we need to check the three topological conditions introduced in Subsection 2.2.1. We now recall and check that the conditions hold true for our energy functional $\mathcal{E}$.

- Lower semicontinuity. $\mathcal{E}$ is sequentially lower semicontinuous with respect to narrow convergence of probability measures on d_W bounded sets

$$\sup_{m,n} d_W(\mu_m,\mu_n) < \infty,\ \mu_n \text{ converges narrowly to } \mu \Rightarrow \liminf_{n\to\infty} \mathcal{E}(\mu_n) \geq \mathcal{E}(\mu).$$

  In Section 3.2 we already show that $\mathcal{E}$ is lower semicontinuous with respect to narrow convergence of probability measures with uniformly bounded second moments.

- Coercivity. There exists $\tau_* > 0$ and $\mu_* \in \mathcal{P}_2(\mathcal{M})$ such that

$$\inf_{\mu\in\mathcal{P}_2(\mathcal{M})} \left\{ \mathcal{E}(\mu) + \frac{1}{2\tau_*} d_W^2(\mu,\mu_*) \right\} > -\infty.$$

  To prove coercivity, let T be as in (3.3.1) and consider $x_0 \in \mathcal{M}$ arbitrary. Then

$$\begin{aligned}
&\mathcal{E}(\mu) + \frac{1}{2\tau} d_W^2(\mu,\delta_{x_0}) \\
&= \int_{\mathcal{M}} V(x)d\mu(x) + \frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} W(x-y)d\mu(x)d\mu(y) + \frac{1}{2\tau}\int_{\mathcal{M}} \operatorname{dist}^2(x,x_0)d\mu(x) \\
&\geq \int_{\mathcal{M}} \left(V(x_0) + \langle \nabla V, T(x_0,x)\rangle + \frac{\lambda}{2}\operatorname{dist}^2(x,x_0) \right) d\mu(x) \\
&+ \int_{\mathcal{M}\times\mathcal{M}} \frac{\lambda}{4}\operatorname{dist}^2\left((x,y),(x_0,x_0)\right) d\mu(x)d\mu(y) + \frac{1}{2\tau}\int_{\mathcal{M}} \operatorname{dist}^2(x,x_0)d\mu(x) \\
&= \int_{\mathcal{M}} \left(\left(\lambda + \frac{1}{2\tau}\right)\operatorname{dist}^2(x_0,x) + \langle \nabla V(x_0), T(x_0,x)\rangle + V(x_0) \right) d\mu(x).
\end{aligned}$$

  Notice that $\langle \nabla V(x_0), T(x_0,x)\rangle = g_{x_0}\left(A(x_0)\nabla V(x_0), T(x_0,x)\right)$ and

$$g_{x_0}\left(T(x_0,x), T(x_0,x)\right) = \operatorname{dist}^2(x_0,x),$$

  so for any $\tau > 0$ such that $\lambda + \frac{1}{2\tau} > 0$, i.e. for $2\lambda^-\tau < 1$, we have

$$\inf_{\mu\in\mathcal{P}_2(\mathcal{M})} \left\{ \mathcal{E}(\mu) + \frac{1}{2\tau} d_W^2(\mu,\delta_{x_0}) \right\} > -\infty,$$

  which implies coercivity for $\mathcal{E}$.

- Compactness. Every d_W bounded set contained in a sublevel of $\mathcal{E}$ is relatively compact with respect to the narrow convergence of probability measures

$$\text{for } (\mu_n) \subset \mathcal{P}_2(\mathcal{M}) \text{ with } \sup_n \mathcal{E}(\mu_n) < \infty \text{ and } \sup_{m,n} d_W(\mu_m,\mu_n) < \infty,$$
$$\text{there exists a narrowly convergent subsequence of } (\mu_n).$$

  To check Compactness condition, note that by Prokhorov's theorem, any sequence $(\mu_n) \subset \mathcal{P}_2(\mathcal{M})$ such that $\sup_{m,n} d_W(\mu_m,\mu_n) < \infty$, μ_n has a narrowly convergent subsequence.

Thus by Lemma 2.2.8 and Proposition 2.2.9, we can show the existence of minimizers of (3.4.1) and convergence of $\mu_\tau(\,\cdot\,)$.

Lemma 3.4.1 (Existence of the discrete solutions)**.** *Suppose $(\mathcal{M}, g)$ satisfies assumptions (M1)-(M2) and W, V satisfy (NL1)-(NL7). Then there exists $\tau_0 > 0$ depending only on V, W such that for all $0 < \tau < \tau_0$ and given $\nu \in \mathcal{P}_2(\mathcal{M})$, there exists $\mu_\infty \in \mathcal{P}_2(\mathcal{M})$ such that*

$$\mathcal{E}(\mu_\infty) + \frac{1}{2\tau} d_W^2(\nu, \mu_\infty) = \inf_{\mu \in \mathcal{P}_2(\mathcal{M})} \Big\{ \mathcal{E}(\mu) + \frac{1}{2\tau} d_W^2(\nu, \mu) \Big\}. \tag{3.4.3}$$

Proposition 3.4.2 (Compactness)**.** *There exist a limit curve $\mu \in AC^2_{loc}([0,\infty); \mathcal{P}_2(\mathcal{M}))$ and a sequence $\tau_n \to 0^+$ such that the piecewise constant interpolate $\tilde{\mu}^n(\,\cdot\,) = \mu_{\tau_n}(\,\cdot\,)$ defined as in (3.4.2) satisfies that $\tilde{\mu}^n(t)$ converges narrowly to $\mu(t)$ for any $t \in [0, \infty)$.*

Note that by Lemma 3.2.2 from [5], we actually have a uniform bound on the second moments of $\tilde{\mu}^n$:

$$\sup_{n,\tau} \int_{\mathcal{M}} \mathrm{dist}^2(x, x_0) d\mu_n^\tau(x) < \infty.$$

By the general theory developed in [5], the limit curve $\mu(\,\cdot\,)$ is a curve of maximal slope with respect to upper gradient $|\partial^- \mathcal{E}|$ defined in (2.2.20). We still need to prove the lower semicontinuity of the slope to show that $\mu(\,\cdot\,)$ is a curve of maximal slope with respect to $|\partial \mathcal{E}|$ instead of $|\partial^- \mathcal{E}|$. We denote by $\kappa^n(t)$ the minimal subdifferential of $\mathcal{E}$ at $\tilde{\mu}^n(t)$. Section 3.3 gives

$$\kappa^n(t,x) = -P_x \left(-A(x) \left(\int_{\mathcal{M}} \nabla W(x-y) d\tilde{\mu}^n(t,y) + \nabla V(x) \right) \right). \tag{3.4.4}$$

Theorem 3.4.3 (Lower semicontinuity of local slope)**.** *Assume that (M1)-(M2) and (NL1)-(NL7) hold true, then the metric slope of the piecewise constant interpolate $\tilde{\mu}^n$ satisfies that for a.e. $t > 0$,*

$$\liminf_{n\to\infty} |\partial \mathcal{E}|^2(\tilde{\mu}^n(t)) \geq |\partial \mathcal{E}|^2(\mu(t)).$$

Remark 3.4.4. *By Fatou's lemma, for any $T > 0$*

$$\begin{aligned} \liminf_{n\to\infty} \int_0^T |\partial \mathcal{E}|^2 \left(\tilde{\mu}^n(t)\right) dt &\geq \int_0^T \liminf_{n\to\infty} |\partial \mathcal{E}|^2(\tilde{\mu}^n(t)) dt \\ &\geq \int_0^T |\partial \mathcal{E}|^2(\mu(t)) dt \end{aligned}$$

In the case $\partial\mathcal{M} = \emptyset$ and $W \in C^1(\mathbb{R}^d \setminus \{0\})$, the lower semicontinuity of local slope can be proved as in Lemma 2.7 from [28]. In the proof below, we allow that $\partial\mathcal{M} \neq \emptyset$.

In the case $\partial\mathcal{M} \neq \emptyset$, the argument in [28] does not work because the projection P breaks the continuity and thus $\kappa^n d\mu^n$ does not necessarily converge narrowly to $\kappa d\mu$. However, the following useful observation holds:

Proposition 3.4.5. *The function* $\mathcal{M}\times\mathbb{R}^d \ni (x,\xi) \mapsto g_x\left(P_x\xi, P_x\xi\right)$ *is lower semicontinuous. For all* $x \in \mathcal{M}$*, the function* $\mathbb{R}^d \ni \xi \mapsto g_x\left(P_x\xi, P_x\xi\right)$ *is convex.*

Proof of Proposition. We first prove the lower semincontinuity property. Assume $\lim_{k\to\infty} x^k = x$ and $\lim_{k\to\infty} \xi^k = \xi$. If $\left\{x^k\right\}_{k=1}^{\infty} \subset \mathring{\mathcal{M}}$ then

$$\begin{aligned} g_x\left(P\xi, P\xi\right) &\le g_x(\xi,\xi) \\ &= \lim_{k\to\infty} g_{x^k}\left(\xi^k, \xi^k\right) \\ &= \lim_{k\to\infty} g_{x^k}\left(P\xi^k, P\xi^k\right). \end{aligned}$$

So lower semicontinuity is verified for $x \in \mathring{\mathcal{M}}$, since for such x and any $\lim_{k\to\infty} x^k = x$, $x^k \in \mathring{\mathcal{M}}$ for all k large enough. For $x \in \partial\mathcal{M}$, due to the fact above, it is enough to consider the case that $x^k \in \partial\mathcal{M}$ for all k. Let $\{e_1, ..., e_d\}$ be a continuous orthonormal basis of $T\mathcal{M}$ near x, such that on $\partial\mathcal{M}, e_d = \vec{n}$ where $\vec{n}$ is the unit outer normal vector with respect to the inner product g. We expand ξ^k in this basis: $\xi^k = \sum_i^d \xi_i^k e_i(x^k)$. Then $P\xi^k = \sum_{i=1}^{d-1} \xi_i^k e_i(x^k) + \left(\xi_d^k\right)^- e_d(x^k)$ for $x^k \in \partial\mathcal{M}$. By the continuity of g and smoothness of $\mathcal{M}$, we have $\lim_{k\to\infty} \xi_i^k = \xi_i$ for all $1 \le i \le d$, thus

$$\begin{aligned} \lim_{k\to\infty} g_{x^k}\left(P\xi^k, P\xi^k\right) &= \lim_{k\to\infty}\left[\sum_{i=1}^{d-1}\left(\xi_i^k\right)^2 + \left(\left(\xi_d^k\right)^-\right)^2\right] \\ &= \sum_{i=1}^{d-1} \xi_i^2 + \left(\xi_d^-\right)^2 \\ &= g_x\left(P\xi, P\xi\right). \end{aligned}$$

We now turn to the convexity property. Similarly for $x \in \mathring{\mathcal{M}}$, since $P_x\xi = \xi$ for all $\xi \in \mathbb{R}^d$, it is straightforward to check that $\xi \mapsto g_x\left(\xi,\xi\right)$ is convex. So we assume $x \in \partial\mathcal{M}$. For any $\xi^1, \xi^2 \in \mathbb{R}^d$, and $0 \le \theta \le 1$ we need to show that

$$g_x\left(P_x\left(\left(1-\theta\right)\xi^1 + \theta\xi^2\right), P_x\left(\left(1-\theta\right)\xi^1 + \theta\xi^2\right)\right) \le (1-\theta)g_x(P_x\xi^1, P_x\xi^1) + \theta g_x(P_x\xi^2, P_x\xi^2).$$

Note that we only need to check that for the last coordinate, that is we only need to prove that

$$\left(\left(\left(1-\theta\right)\xi_d^1 + \theta\xi_d^2\right)^-\right)^2 \le \left(1-\theta\right)\left(\left(\xi_d^1\right)^-\right)^2 + \theta\left(\left(\xi_d^2\right)^-\right)^2,$$

which is a direct consequence of the fact that $f(x) = (x^-)^2$ is a convex function. The proposition is proved. $\square$

We now start to prove the lower semicontinuity of local slope

Proof of Theorem. Since κ is the minimal subdifferential, be Remark 3.3.3 and Theorem 3.3.4, we only need to prove

$$\liminf_{n\to\infty}\int_{\mathcal{M}} g\left(\kappa^n(t,x),\kappa^n(t,x)\right)d\tilde{\mu}^n(t,x)\geq\int_{\mathcal{M}} g\left(\kappa(t,x),\kappa(t,x)\right)d\mu(t,x).$$

Note that the non-negative function $\mathcal{M}\times\mathbb{R}^d\ni(x,\xi)\mapsto g_x\left(P_x\xi,P_x\xi\right)$ satisfies the lower semicontinuity and convexity property. By Proposition 6.42 from [53], we know that for all $(x,\xi)\in\mathcal{M}\times\mathbb{R}^d$,

$$g_x\left(P_x\xi,P_x\xi\right)=\sup_{i\in\mathbb{N}}\{a_i(x)+b_i(x)\xi\}$$

for some bounded continuous functions a_i,b_i. A similar argument to one in Lemma 2.7 of [28] gives that $\nabla W*\mu^n$ converges narrowly to $\nabla W*\mu$. Thus for any $i\in\mathbb{N}$,

$$\begin{aligned}
&\liminf_{n\to\infty}\int_{\mathcal{M}} g_x\left(\kappa^n(t,x),\kappa^n(t,x)\right)d\tilde{\mu}^n(t,x)\\
&=\liminf_{n\to\infty}\int_{\mathcal{M}} g_x\left(P\left(-\nabla W*\tilde{\mu}^n(t,x)-\nabla V(x)\right),P\left(-\nabla W*\tilde{\mu}^n(t,x)-\nabla V(x)\right)\right)d\tilde{\mu}^n(t,x)\\
&\geq\liminf_{n\to\infty}\int_{\mathcal{M}}\left(a_i(x)-b_i(x)\left(\nabla W*\tilde{\mu}^n(t)(x)+\nabla V(x)\right)\right)d\tilde{\mu}^n(t,x)\\
&=\int_{\mathcal{M}}\left(a_i(x)-b_i(x)\left(\nabla W*\mu(t)(x)+\nabla V(x)\right)\right)d\mu(t,x).
\end{aligned}$$

Taking supremum over $i\in\mathbb{N}$ and using Lebesgue's monotone convergence theorem then gives

$$\begin{aligned}
&\liminf_{n\to\infty}\int_{\mathcal{M}} g\left(\kappa^n(t,x),\kappa^n(t,x)\right)d\tilde{\mu}^n(t,x)\\
&\geq\sup_{i\in\mathbb{N}}\int_{\mathcal{M}}\left(a_i(x)-b_i(x)\left(\nabla W*\mu(t)(x)+\nabla V(x)\right)\right)d\mu(t,x)\\
&=\int_{\mathcal{M}} g\left(\kappa(t,x),\kappa(t,x)\right)d\mu(t,x).
\end{aligned}$$

□

Lemma 3.4.6. *Let $\mu\in AC_{loc}(0,+\infty;\mathcal{P}_2(\mathcal{M}))$ and $v(t)$ be its tangent velocity field defined in Lemma 3.5.1. Then for almost every $t\geq 0$*

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right)=\int_{\mathcal{M}} g_x\left(\kappa(t,x),v(t,x)\right)d\mu(t,x),\tag{3.4.5}$$

Proof. Since $\kappa(t)\in\partial\mathcal{E}(\mu(t))$ (by Theorem 3.3.4), we know that

$$\mathcal{E}\left(\mu(t+h)\right)\geq\mathcal{E}\left(\mu(t)\right)+\int_{M\times M} g_x\left(\kappa(t,x),T(x,y)\right)d\gamma_t^h(x,y)+o\left(d_W\left(\mu(t),\mu(t+h)\right)\right).$$

For $h > 0$

$$\begin{aligned}\lim_{h\to 0^+} \frac{\mathcal{E}\left(\mu(t+h)\right) - \mathcal{E}\left(\mu(t)\right)}{h} &\geq \lim_{h\to 0^+} \int_{\mathcal{M}\times\mathcal{M}} g_x\left(\kappa(t,x), \frac{T(x,y)}{h}\right) d\gamma_t^h(x,y)\\ &= \lim_{h\to 0^+} \int_{\mathcal{M}} g_x\left(\kappa(t,x), \int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y)\right) d\mu(t,x)\\ &= \int_{\mathcal{M}} g_x\left(\kappa(t,x), v(t,x)\right) d\mu(t,x).\end{aligned}$$

Similarly, for $h < 0$, we have

$$\lim_{h\to 0^-} \frac{\mathcal{E}\left(\mu(t+h)\right) - \mathcal{E}\left(\mu(t)\right)}{h} \leq \int_{\mathcal{M}} g_x\left(\kappa(t,x), v(t,x)\right) d\mu(t,x).$$

Note that the function $t \to \mathcal{E}\left(\mu(t)\right)$ is non-increasing and thus differentiable for a.e. $t > 0$. Hence

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right) = \int_{\mathcal{M}} g_x\left(\kappa(t,x), v(t,x)\right) d\mu(t,x),$$

for a.e. $t > 0$ as desired. □

Note that by (4.2.3), we get

$$\begin{aligned}\frac{d}{dt}\mathcal{E}(\mu(t)) &= \int_{\mathcal{M}} g_x\left(\kappa(t,x), v(t,x)\right) d\mu(t,x)\\ &\leq \|\kappa(t)\|_{L^2(g,\mu(t))}\|v(t)\|_{L^2(g,\mu(t))}\\ &= |\partial\mathcal{E}|(\mu(t))|\mu'|(t).\end{aligned} \tag{3.4.6}$$

Theorem 3.4.7 (Existence of curves of maximal slope)**.** *Suppose $(\mathcal{M}, g)$ satisfies (M1)-(M2) and W, V satisfy (NL1)-(NL7). Then there exists at least one curve of maximal slope for the functional $\mathcal{E}$, i.e., there exists $\mu \in AC_{loc}([0,\infty); \mathcal{P}_2(\mathcal{M}))$ such that for all $T \geq 0$*

$$\mathcal{E}(\mu_0) \geq \mathcal{E}(\mu(T)) + \frac{1}{2}\int_0^T \left|\mu'\right|^2(t)dt + \frac{1}{2}\int_0^T |\partial\mathcal{E}|^2\left(\mu(t)\right) dt. \tag{3.4.7}$$

Proof. We know that $\mu \mapsto \mathcal{E}(\mu)$ and $\mu \mapsto |\partial\mathcal{E}|(\mu)$ are lower semicontinuous with respect to the narrow convergence. Since $|\partial\mathcal{E}| = |\partial^-\mathcal{E}|$, in order to apply Theorem 2.2.10 (i.e. Theorem 2.3.3 from [5]) to get existence of curves of maximal slope satisfying (3.4.7), we only need to show that $\mathcal{P}_2(\mathcal{M}) \ni \mu \mapsto |\partial\mathcal{E}|(\mu)$ is a strong upper gradient of $\mathcal{E}$, according to Definition 2.2.1.

To show that, consider $\mu \in AC_{loc}(0,+\infty; \mathcal{P}_2(\mathcal{M}))$. We first show that $\mathcal{E}(\mu(\,\cdot\,))$ is also locally absolutely continuous. We note that by the linear growth conditions on ∇V (3.1.7) and ∇W (3.1.6), $|V(x) - V(y)| \leq C(1 + \operatorname{dist}(x,y))\operatorname{dist}(x,y)$ and $|W(x-z) - W(y-w)| \leq C(1 + \operatorname{dist}(x,y) + \operatorname{dist}(z,w))(\operatorname{dist}(x,y) + \operatorname{dist}(z,w))$. Then for $0 \leq s < t < \infty$ and

$\gamma \in \Gamma_o(\mu(t), \mu(s))$ an optimal plan,

$$\begin{aligned} |\mathcal{E}(\mu(t)) - \mathcal{E}(\mu(s))| &\leq \int_{\mathcal{M}\times\mathcal{M}} C(1 + \operatorname{dist}(x,y)) \operatorname{dist}(x,y) d\gamma(x,y) \\ &\leq C(1 + d_W(\mu(t), \mu(s)) d_W(\mu(t), \mu(s)). \end{aligned} \tag{3.4.8}$$

Thus $\mathcal{E}(\mu(t))$ is locally absolutely continuous since $\mu(t)$ is locally absolutely continuous in $(\mathcal{P}_2(\mathcal{M}), d_W)$. Then by consequence of the chain rule (3.4.6), we have

$$\frac{d}{dt}\mathcal{E}(\mu(t)) \leq |\partial\mathcal{E}|(\mu(t))|\mu'|(t).$$

Thus for any $0 < s \leq t < +\infty$, we can integrate to get (2.2.9) and $|\partial\mathcal{E}|(\,\cdot\,)$ is a strong upper gradient for $\mathcal{E}$.

□

3.5 Existence of the gradient flow

In this Section, we first show that locally absolutely continuous curves in $\mathcal{P}_2(\mathcal{M})$ with respect to d_W are solutions to continuity equations in the sense of distributions. Furthermore velocities are in $L^2(g, \mu)$ and belonging to the tangent space to the set of configurations. We then prove the existence of gradient flow and the stability property of the gradient flow.

Lemma 3.5.1. *Let $\mu(\,\cdot\,)$ be an absolutely continuous curve in $\mathcal{P}_2(\mathcal{M})$ and γ_t^h be an optimal plan between $\mu(t)$ and $\mu(t+h)$, i.e. $\gamma_t^h \in \Gamma_o(\mu(t), \mu(t+h))$. Denote the disintegration of γ_t^h with respect to $\mu(t)$ by ν_x^h, then $\int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y)$ converges weakly in $L^2(g, \mu(t))$ to a vector field $v(t,x)$ for a.e. $t > 0$ such that $(\mu(\,\cdot\,), v)$ satisfy the continuity equation*

$$\frac{\partial}{\partial t}\mu(t,x) + \operatorname{div}(\mu(t,x)v(t,x)) = 0 \tag{3.5.1}$$

in the sense of distributions, i.e., test against $\phi \in C_c^\infty([0,\infty) \times \mathcal{M})$, and among all vector fields such that (4.2.1) holds, and v has minimal $L^2(g,\mu)$-norm with

$$\int_{\mathcal{M}} g(v(t,x), v(t,x))\, d\mu(t,x) = |\mu'|^2(t) \tag{3.5.2}$$

for a.e. $t > 0$.

Proof. For the existence of a unique minimal $L^2(g, \mu(t))$-norm vector field $v(t)$ such that $\mu(t)$ satisfies (3.5.1) and (3.5.2), we refer to Theorem 2.29 from [4]. For a fixed $t > 0$ such that the metric derivative $|\mu'|(t)$ exists, we now show that such $v(t)$ is given by the limit of

$\int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y)$. Note that

$$\begin{aligned}
&\int_{\mathcal{M}} g_x \left(\int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y), \int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y) \right) d\mu(t,x) \\
&\leq \int_{\mathcal{M}\times\mathcal{M}} g_x \left(\frac{T(x,y)}{h}, \frac{T(x,y)}{h} \right) d\nu_x^h(y) d\mu(t,x) \\
&= \int_{\mathcal{M}\times\mathcal{M}} g_x \left(\frac{T(x,y)}{h}, \frac{T(x,y)}{h} \right) d\gamma_t^h(x,y) \\
&= \frac{1}{h^2} d_W^2 \left(\mu(t), \mu(t+h)\right).
\end{aligned}$$

Since $\mu(t)$ is absolutely continuous, we know that $\frac{1}{h^2} d_W^2 \left(\mu(t), \mu(t+h)\right) \leq C(t)$ uniformly in h for some constant C. Thus, uniformly in h, we have

$$\int_{\mathcal{M}} g_x \left(\int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y), \int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y) \right) d\mu(t,x) \leq C(t).$$

So there exist a vector field $\tilde{v}(t,x)$ and a sequence $\{h_n\}$ such that $\int_{\mathcal{M}} \frac{T(x,y)}{h_n} d\nu_x^{h_n}(y)$ converges weakly in $L^2(g, \mu(t,x))$ to $\tilde{v}$ as $h_n \to 0$. We claim that

$$\lim_{n\to\infty} \frac{\int_{\mathcal{M}} \phi(t,x) d\mu(t+h_n,x) - \int_{\mathcal{M}} \phi(t,x) d\mu(t,x)}{h_n} = \int_{\mathcal{M}} g_x \left(\nabla\phi(t,x), \tilde{v}(t,x)\right) d\mu(t,x),$$

for a.e. $t > 0$ and for any $\phi \in C_c^\infty\left([0,\infty)\times\mathbb{R}^n\right)$.
Indeed, for the left-hand side we know,

$$\begin{aligned}
&\frac{\int_{\mathcal{M}} \phi(t,x) d\mu(t+h_n,x) - \int_{\mathcal{M}} \phi(t,x) d\mu(t,x)}{h_n} \\
&= \frac{1}{h_n} \int_{\mathcal{M}\times\mathcal{M}} \left(\phi(t,y) - \phi(t,x)\right) d\gamma_t^{h_n}(x,y) \\
&= \frac{1}{h_n} \int_{\mathcal{M}\times\mathcal{M}} \langle \nabla\phi(t,x), T(x,y) \rangle d\gamma_t^{h_n}(x,y) + o(h_n) \\
&= \int_{\mathcal{M}} \left\langle \nabla\phi(t,x), \int_{\mathcal{M}} \frac{T(x,y)}{h_n} d\nu_x^{h_n}(y) \right\rangle d\mu(t,x) + o(h_n) \\
&= \int_{\mathcal{M}} g_x \left(A(x)\nabla\phi(t,x), \int_{\mathcal{M}} \frac{T(x,y)}{h_n} d\nu_x^{h_n}(y) \right) d\mu(t,x) + o(h_n)
\end{aligned}$$

Since $A(x)\nabla\phi(t,x) \in L^2\left(g, \mu(t)\right)$ and $\int_{\mathcal{M}} \frac{T(x,y)}{h_n} d\nu_x^{h_n}(y)$ converges weakly in $L^2\left(g, \mu(t)\right)$ to $\tilde{v}(t,x)$, we get

$$\begin{aligned}
\lim_{n\to\infty} \frac{\int_{\mathcal{M}} \phi(t,x) d\mu(t+h_n,x) - \int_{\mathcal{M}} \phi(t,x) d\mu(t,x)}{h_n} &= \int_{\mathcal{M}} g_x \left(A(x)\nabla\phi(t,x), \tilde{v}(t,x)\right) d\mu(t,x) \\
&= \int_{\mathcal{M}} \langle \nabla\phi(t,x), \tilde{v}(t,x) \rangle d\mu(t,x).
\end{aligned}$$

Since $\mu(t)$ satisfies (4.2.1) with respect to the vector field v, we know that

$$\lim_{h\to 0}\frac{\int_{\mathcal{M}}\phi(t,x)d\mu(t+h,x)-\int_{\mathcal{M}}\phi(t,x)d\mu(t,x)}{h}=\int_{\mathcal{M}}g_x\left(\nabla\phi(t,x),v(t,x)\right)d\mu(t,x),$$

for a.e. $t>0$. Thus $\int_{\mathcal{M}}g_x\left(\nabla\phi(t,x),v(t,x)-\tilde{v}(t,x)\right)d\mu(t,x)=0$ for a.e. $t>0$ and $\mu(t)$ satisfies (4.2.1) with respect to $\tilde{v}(t)$. Now notice that

$$\begin{aligned}
&\int_{\mathcal{M}}g_x\left(\tilde{v}(t,x),\tilde{v}(t,x)\right)d\mu(t,x)\\
&\leq\lim_{n\to\infty}\int_{\mathcal{M}}g_x\left(\int_{\mathcal{M}}\frac{T(x,y)}{h_n}d\nu_x^{h_n}(y),\int_{\mathcal{M}}\frac{T(x,y)}{h_n}d\nu_x^{h_n}(y)\right)d\mu(t,x)\\
&\leq\lim_{n\to\infty}\frac{1}{h_n^2}d_W^2\left(\mu(t),\mu(t+h_n)\right)\\
&=\left|\mu'\right|^2(t)=\int_{\mathcal{M}}g_x\left(v(t,x),v(t,x)\right)d\mu(t,x).
\end{aligned}$$

Together with the minimal $L^2(g,\mu(t))$-norm property of v, we have $\tilde{v}(t)=v(t)$. Since for any $h_n\to 0$ such that $\lim_{n\to\infty}\int_{\mathcal{M}}\frac{T(x,y)}{h_n}d\nu_x^{h_n}(y)$ converges weakly in $L^2(g,\mu(t))$, the weak limit is the same $v(t)$, we have $\int_{\mathcal{M}}\frac{T(x,y)}{h}d\nu_x^h(y)$ converges weakly in $L^2(g,\mu(t))$ to $v(t,x)$. The lemma is proved. $\square$

Recall from Chapter 2 we call $v(t)$ is the tangent velocity field of $\mu(\cdot)$, now we can define gradient flow by

Definition 3.5.2 (Gradient flows)**.** *A locally absolutely continuous curve* $[0,\infty)\ni\ t\mapsto\mu(t)\in\mathcal{P}_2(\mathcal{M})$ *is a* gradient flow *with respect to* $\mathcal{E}$ *if for a.e.* $t>0$

$$v(t)\in-\partial\mathcal{E}\left(\mu(t)\right), \tag{3.5.3}$$

where $v(t)$ *is the tangent velocity field for* $\mu(t)$.

We now show the proof of Theorem 3.1.3

Proof of Theorem 3.1.3. By the chain rule (3.4.5)

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right)=\int_{\mathcal{M}}g_x\left(\kappa(t,x),v(t,x)\right)d\mu(t,x),$$

for a.e. $t>0$, where $v(t)$ is the tangent velocity field for the curve $\mu(t)$. The fact that $\mu(\cdot)$ is a curve of maximal slope implies

$$\frac{d}{dt}\mathcal{E}(\mu(t))\leq-\frac{1}{2}\int_{\mathcal{M}}g_x\left(v(t,x),v(t,x)\right)d\mu(t,x)-\frac{1}{2}\int_{\mathcal{M}}g_x\left(\kappa(t,x),\kappa(t,x)\right)d\mu(t,x). \tag{3.5.4}$$

Combining with (3.4.5) implies $v(t,x)=-\kappa(t,x)$ for a.e. $t>0$. Also since $\partial^o\mathcal{E}(\mu(t))=-\kappa(t,x)$, (3.1.13) is true. Togather with the fact that $\mathcal{E}(\mu)(\cdot)$ is locally absolutely continuous by (3.4.8), we integrate to get (3.1.14). The characterization of the subdifferential of

$\mathcal{E}$ of Theorem 3.3.4 implies that $\mu(t)$ is a gradient flow of $\mathcal{E}$. Then by Lemma 3.5.1, $\mu(\cdot)$ is a weak measure solution to (3.1.5) with initial data μ_0.

□

We now focus on proving Theorem 3.1.4, that is that λ-convexity of $\mathcal{E}$ implies the stability of the gradient flow. We first establish the following lemma:

Lemma 3.5.3. *Let $\mu(t)$ be a locally absolutely continuous curve in $\mathcal{P}_2(\mathcal{M})$ with tangent velocity v, then for a.e. $t > 0$,*

$$\frac{1}{2}\frac{d}{dt}d_W^2\left(\mu(t),\nu\right) = -\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(t,x), T(x,y)\right)d\gamma_t(x,y), \tag{3.5.5}$$

for any fixed $\nu \in \mathcal{P}_2(\mathcal{M})$ and $\gamma_t \in \Gamma_o\left(\mu(t),\nu\right)$ an optimal plan.

Proof of Lemma. We first notice that the function $t \mapsto d_W^2\left(\mu(t),\nu\right)$ is differentiable for a.e. $t > 0$ since $t \mapsto \mu(t)$ is locally absolutely continuous in $(\mathcal{P}_2(\mathcal{M}), d_W)$. In the rest of the proof, we assume that we are working on $t > 0$ such that the function $s \mapsto \frac{1}{2}d_W^2\left(\mu(s),\nu\right)$ is differentiable at t. In the case v is locally Lipschitz in space and $\mathcal{M}$ has no boundary then using the flow map with velocity field v, similar arguments as in [115, 45] imply (3.5.5). However, in our case, we need to deal with the fact that since v is not continuous the flow map is not readily available and furthermore that a geodesic in direction v may not exist at the boundary. We divide the proof into two steps.

Step 1. Consider the case that $\mu(t),\nu$ have compact support for all $t > 0$. To show (3.5.5) we modify the arguments of Theorem 8.4.7 from [5]. An issue is that, as in the proof of Theorem 3.3.4, there may exist $x \in \partial\mathcal{M}$ such that there exists no $t > 0$ for which $\exp_x(tv(x)) \in \mathcal{M}$ exists. To deal with this problem we use the following approximations. For $h \in \mathbb{R}$ with $|h|$ small, define

$$v_h(t,x) = \begin{cases} v(t,x), & \text{if } x \in B(\frac{1}{|h|}) \cap \mathcal{M}_{|h|} \\ v(t,x) - hn(x), & \text{if } x \in B(\frac{1}{|h|}) \cap \partial\mathcal{M} \\ 0, & \text{otherwise.} \end{cases}$$

It is direct to check that v_h converges to v in $L^2(g,\mu(t))$. For fixed $h \in \mathbb{R}$, same argument as in the proof of Theorem 3.3.4 shows that there exists $C(h) > 0$ such that $\exp_x(thv_h(t,x))$ exists for all $0 \le t \le C(h)$ and $x \in \mathcal{M}$. Thus there exists a function f such that $\lim_{h\to 0} f(h) = 0$ and $\exp_x(hv_{f(h)}(t,x)) \in \mathcal{M}$ for all $x \in \mathcal{M}$. We claim that for a.e. $t > 0$

$$\lim_{h\to 0}\frac{d_W^2\left(\left(\exp\left(hv_{f(h)}\right)\right)_\sharp \mu(t), \mu(t+h)\right)}{h^2} = 0. \tag{3.5.6}$$

Indeed, if the claim is true, then for a.e. $t > 0$, we know that $d_W^2(\mu(t), \nu)$ is differentiable and

$$\begin{aligned} \frac{d}{dt} d_W^2(\mu(t), \nu) &= \lim_{h \to 0} \frac{d_W^2\left(\mu(t+h), \nu\right) - d_W^2\left(\mu(t), \nu\right)}{h} \\ &= \lim_{h \to 0} \frac{d_W^2\left(\left(\exp(hv_{f(h)})\right)_\sharp \mu(t), \nu\right) - d_W^2\left(\mu(t), \nu\right)}{h}. \end{aligned}$$

Since $(\exp(hv_{f(h)}), id)_\sharp \gamma_t \in \Gamma\left(\exp(hv_{f(h)})_\sharp \mu(t), \nu\right)$, we get

$$d_W^2\left(\left(\exp(hv_{f(h)})\right)_\sharp \mu(t), \nu\right) \le \int_{\mathcal{M} \times \mathcal{M}} \operatorname{dist}^2\left(\exp_x(hv_{f(h)}(t, x)), y\right) d\gamma_t(x, y).$$

Recall that by the first variation formula, for any $x, y \in \mathcal{M}$, denote

$$D(x, y) = \left\{v \in T_x\mathcal{M} : \exp_x(tv) \in \mathcal{M} \ \forall t \in [0, 1], \exp_x(v) = y, g_x(v, v) = \operatorname{dist}^2(x, y)\right\}, \tag{3.5.7}$$

then

$$\lim_{h \to 0^+} \frac{\operatorname{dist}^2(\exp_x(h\xi), y) - \operatorname{dist}^2(x, y)}{h} = \min\left\{-2g_x\left(\xi, v\right) : v \in D(x, y)\right\}.$$

So taking $h \to 0^+$ and using the Lebesgue's dominated convergence theorem yields

$$\begin{aligned} \frac{d^+}{dt} d_W^2(\mu(t), \nu) &\le \lim_{h \to 0^+} \frac{1}{h} \int_{\mathcal{M} \times \mathcal{M}} \left(\operatorname{dist}^2\left(\exp_x(hv_{f(h)}(t, x)), y\right) - \operatorname{dist}^2(x, y)\right) d\gamma_t(x, y) \\ &\le -2 \int_{\mathcal{M} \times \mathcal{M}} g_x\left(v(t, x), T(x, y)\right) d\gamma_t(x, y). \end{aligned}$$

Similarly, taking $h \to 0^-$ gives

$$\frac{d^-}{dt} d_W^2(\mu(t), \nu) \ge -2 \int_{\mathcal{M} \times \mathcal{M}} g_x\left(v(t, x), T(x, y)\right) d\gamma_t(x, y).$$

Thus we have

$$\frac{1}{2} \frac{d}{dt} d_W^2(\mu(t), \nu) = -\int_{\mathcal{M} \times \mathcal{M}} g_x\left(v(t, x), T(x, y)\right) d\gamma_t(x, y),$$

for a.e. $t > 0$.

We now prove the claim. It is enough to show that

$$\lim_{h \to 0} \int_{\mathcal{M} \times \mathcal{M}} \frac{1}{h^2} \operatorname{dist}^2\left(\exp_x(hv_{f(h)}(t, x)), y\right) d\gamma_t^h(x, y) = 0 \tag{3.5.8}$$

where $\gamma_t^h \in \Gamma_o\left(\mu(t), \mu(t+h)\right)$. Since $\mu(t)$ has compact support for all $t > 0$, we only need to show (3.5.8) for compact subsets of $\mathcal{M}$, i.e., to show

$$\lim_{h \to 0} \int_{K \times K} \frac{1}{h^2} \operatorname{dist}^2\left(\exp_x(hv_{f(h)}(t, x)), y\right) d\gamma_t^h(x, y) = 0$$

for any compact subset $K \Subset \mathcal{M}$. On K, we have that the sectional curvature is bounded from below say by $-k$, then by rescaling, we may assume the constant for the lower bounded of sectional curvature is -1. By the comparison theorem, refer to [94] Theorem 79, we have

$$\begin{aligned}&\cosh\left[\mathrm{dist}(\exp_x(hv_{f(h)}(t,x)),y)\right]\\&\le \cosh\left[\mathrm{dist}(x,y)\right]\cosh\left[h|v_{f(h)}(t,x)|_g\right]-\sinh\left[\mathrm{dist}(x,y)\right]\sinh\left[h|v_{f(h)}(t,x)|_g\right]\cos\alpha,\end{aligned}$$

where α is angle between $v_{f(h)}(t,x)$ and $T(x,y)$, i.e., $\cos\alpha = \frac{g_x\left(v_{f(h)}(t,x),T(x,y)\right)}{\mathrm{dist}(x,y)|v_{f(h)}(t,x)|_g}$. Note that

$$\cosh\left[z\right] = 1+\frac{1}{2}z^2+O(z^4)$$

and

$$\sinh\left[z\right] = z+O(z^3).$$

Expanding cosh, sinh in the comparison formula, we have

$$\begin{aligned}&1+\frac{1}{2}\,\mathrm{dist}^2\left(\exp_x(hv_{f(h)}(t,x)),y\right)\\&\le \cosh\left[\mathrm{dist}(\exp_x(hv_{f(h)}(t,x)),y)\right]\\&\le 1+\frac{1}{2}\,\mathrm{dist}^2(x,y)+\frac{1}{2}h^2|v_{f(h)}(t,x)|_g^2-h\,\mathrm{dist}(x,y)|v_{f(h)}(t,x)|_g\cos\alpha\\&+O(h^3)+O(\mathrm{dist}^3(x,y)).\end{aligned}$$

Thus

$$\begin{aligned}&\lim_{h\to0}\int_{K\times K}\frac{1}{h^2}\,\mathrm{dist}^2\left(\exp_x(hv_{f(h)}(t,x)),y\right)d\gamma_t^h(x,y)\\&\le \lim_{h\to0}\int_{K\times K}\left(\frac{1}{h^2}\,\mathrm{dist}^2(x,y)-\frac{2}{h}\,\mathrm{dist}(x,y)|v_{f(h)}(t,x)|_g\cos\alpha+|v_{f(h)}(t,x)|_g^2\right.\\&\left.+o(h)\right)d\gamma_t^h(x,y)\\&=\lim_{h\to0}\int_{K\times K}\left(\frac{1}{h^2}\,\mathrm{dist}^2(x,y)-2g_x\left(\frac{T(x,y)}{h},v_{f(h)}(t,x)\right)+|v_{f(h)}(t,x)|_g^2\right)d\gamma_t^h(x,y)\\&=\lim_{h\to0}\int_{K\times K}g_x\left(\frac{T(x,y)}{h}-v_{f(h)}(t,x),\frac{T(x,y)}{h}-v_{f(h)}(t,x)\right)d\gamma_t^h(x,y)=0.\end{aligned}$$

Step 2. (3.5.5) holds for general $\mu(t),\nu\in\mathcal{P}_2(\mathcal{M})$. To show that, we need to perform the same approximation as in the proof of Theorem 23.9 from [115], which requires that notion of dynamical coupling, refer to [115]. Here we sketch the approximation and argument, let $A_k=\{\gamma : \sup_t \mathrm{dist}(z,\gamma(t))\le k\}$, where γ is a random curve $\gamma:[0,1]\to\mathcal{M}$ and e_t is the evaluation map $e_t(\gamma)=\gamma(t)$. Define $\mu^k(t)=(e_t)_\sharp\,\Pi_k$ where $\Pi_k(d\gamma)=\frac{\chi_{\gamma\in A_k}\Pi(d\gamma)}{\Pi(A_k)}$

and Π is a probability measure on the action minimizing curves. Denote $Z_k = \Pi(A_k)$ then $Z_k \uparrow 1, Z_k\mu^k(t) \uparrow \mu(t)$ as $k \to \infty$. For each k μ^k solves

$$\frac{\partial \mu^k(t)}{\partial t} + \operatorname{div}(\mu^k(t)v(t)) = 0, \tag{3.5.9}$$

and $\mu^n(t)$ has compact support in $B(z,k)$. So by Step 1,

$$\frac{1}{2}\frac{d}{dt}d_W^2\left(\mu^k(t),\nu^k\right) = -\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(t,x),T(x,y)\right)d\gamma_t^k(x,y).$$

Since $d_W^2(\mu^k(t),\nu^k)$ is locally absolutely continuous, integrating gives

$$\frac{d_W^2\left(\mu^k(t),\nu^k\right)}{2} = \frac{d_W^2\left(\mu^k(0),\nu^k\right)}{2} - \int_0^t\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(s,x),T(x,y)\right)d\gamma_s^k(x,y)ds. \tag{3.5.10}$$

We only need to take $k \to \infty$. By the proof of Theorem 23.9 from [115], $d_W\left(\mu^k(t),\mu(t)\right) = 0$ and we only need to check

$$\lim_{k\to\infty}\int_0^t\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(s,x),T(x,y)\right)d\gamma_s^k(x,y)ds = \int_0^t\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(s,x),T(x,y)\right)d\gamma_s(x,y)ds.$$

Notice that

$$\begin{aligned}
&\left|\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(s,x),T(x,y)\right)d\gamma_s^k(x,y)\right| \\
&\le \left(\int_{\mathcal{M}\times\mathcal{M}} \operatorname{dist}^2(x,y)d\gamma_s^k(x,y)\right)^{\frac{1}{2}}\left(\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(s,x),v(s,x)\right)d\gamma_s^k(x,y)\right)^{\frac{1}{2}} \\
&\le d_W\left(\mu^k(s),\nu^k\right)\left(\frac{1}{Z_k}\int_{\mathcal{M}} g_x\left(v(s,x),v(s,x)\right)d\mu^k(s,x)\right)^{\frac{1}{2}} \\
&\le C\left(\int_{\mathcal{M}} g_x\left(v(s,x),v(s,x)\right)d\mu(s,x)\right)^{\frac{1}{2}},
\end{aligned}$$

and $\int_{\mathcal{M}} g_x\left(v(s,x),v(s,x)\right)d\mu(s,x) \in L^1([0,t])$. It is then sufficient to prove that for a.e. $s \in (0,t)$

$$\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(s,x),T(x,y)\right)d\gamma_s^k(x,y) \to \int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(x,s),T(x,y)\right)d\gamma_s(x,y). \tag{3.5.11}$$

Since

$$\begin{aligned}
&\int_{\mathcal{M}\times\mathcal{M}} \left|g_x\left(v(s,x),T(x,y)\right)\right|d\left|\gamma_s^k-\gamma_s\right|(x,y) \\
&\le \left(\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(s,x),v(s,x)\right)d\left|\gamma_s^k-\gamma_s\right|(x,y)\right)^{\frac{1}{2}}\left(\int_{\mathcal{M}\times\mathcal{M}} \operatorname{dist}^2(x,y)d\left|\gamma_s^k-\gamma_s\right|(x,y)\right)^{\frac{1}{2}} \\
&\le C d_W\left(\mu(s),\nu\right)\left(\int_{\mathcal{M}} g_x\left(v(s,x),v(s,x)\right)d\left|\mu^k(s)-\mu(s)\right|(x)\right)^{\frac{1}{2}},
\end{aligned}$$

and

$$
\begin{aligned}
&\int_{\mathcal{M}} g_x\left(v(s,x), v(s,x)\right) d\left|\mu^k(s)-\mu(s)\right|(x) \\
&\leq (Z_k^{-1}-1)\int_{\mathcal{M}} g_x\left(v(s,x), v(s,x)\right) d\mu(s,x) \\
&+ Z_k^{-1}\int_{\mathcal{M}} g_x\left(v(s,x), v(s,x)\right) d\left|Z_k\mu^k(s)-\mu(s)\right|(x) \\
&\leq (Z_k^{-1}-1)\int_{\mathcal{M}} g_x\left(v(s,x), v(s,x)\right) d\mu(s,x) + Z_k^{-1}\int_{e_s(S)\setminus e_s(A_k)} g_x\left(v(s,x), v(s,x)\right) d\mu(s,x) \\
&\leq (Z_k^{-1}-1)\int_{\mathcal{M}} g_x\left(v(s,x), v(s,x)\right) d\mu(s,x) + Z_k^{-1}\int_{S\setminus A_k} g_{\gamma(s)}\left(v(s,\gamma(s)), v(s,\gamma(s)\right) d\Pi(\gamma),
\end{aligned}
$$

we know

$$
\lim_{k\to\infty}\int_{\mathcal{M}\times\mathcal{M}} \left|g_x\left(v(s,x), T(x,y)\right)\right| d\left|\gamma_s^k-\gamma_s\right|(x,y) = 0.
$$

Thus (3.5.11) holds true. Take $k\to\infty$ in (3.5.10) then gives (3.5.5). $\square$

We now prove Theorem 3.1.4.

Proof of Theorem 3.1.4. Let $\kappa^1 \in \partial^o\mathcal{E}(\mu^1(t))$, $\kappa^2 \in \partial^o\mathcal{E}(\mu^2(t))$ be the minimal subdifferentials and v^1, v^2 be the tangent velocities of the absolutely continuous curves $\mu^1(\,\cdot\,), \mu^2(\,\cdot\,)$ respectively. Also denote by $\gamma_t, \tilde{\gamma}_t \in \Gamma_o\left(\mu^1(t), \mu^2(t)\right)$ optimal transportation plans between $\mu^1(t)$ and $\mu^2(t)$ such that the subdifferential property hold for $\mu^1(t)$ and $\mu^2(t)$ with respect to γ_t and $\tilde{\gamma}_t$ respectively, i.e.

$$
\mathcal{E}\left(\mu^2(t)\right) \geq \mathcal{E}\left(\mu^1(t)\right) + \int_{\mathcal{M}\times\mathcal{M}} g\left(\kappa^1(t,x), T(x,y)\right) d\gamma_t(x,y) + \lambda d_W^2\left(\mu^1(t), \mu^2(t)\right), \quad (3.5.12)
$$

and

$$
\mathcal{E}\left(\mu^1(t)\right) \geq \mathcal{E}\left(\mu^2(t)\right) + \int_{\mathcal{M}\times\mathcal{M}} g\left(\kappa^2(t,y), T(y,x)\right) d\tilde{\gamma}_t(x,y) + \lambda d_W^2\left(\mu^1(t), \mu^2(t)\right). \quad (3.5.13)
$$

By Lemma 4.3.4 from [5] and Lemma 3.5.3 we have

$$
\begin{aligned}
\frac{d}{dt} d_W^2\left(\mu^1(t), \mu^2(t)\right) &\leq -2\int_{\mathcal{M}\times\mathcal{M}} g\left(v^1(t,x), T(x,y)\right) d\gamma_t(x,y) + g\left(v^2(t,y), T(y,x)\right) d\tilde{\gamma}_t(x,y) \\
&= 2\int_{\mathcal{M}\times\mathcal{M}} g\left(\kappa^1(t,x), T(x,y)\right) d\gamma(x,y) + g\left(\kappa^2(t,y), T(y,x)\right) d\tilde{\gamma}_t(x,y) \\
&\leq -4\lambda d_W^2\left(\mu^1(t), \mu^2(t)\right).
\end{aligned}
$$

We can use Gronwall's inequality to get,

$$
d_W\left(\mu^1(t), \mu^2(t)\right) \leq e^{-2\lambda t} d_W\left(\mu_0^1, \mu_0^2\right).
$$

(3.1.15) is proved.

Now we turn to the relationship between gradient flow and system of evolution variational inequalities. If $\mu(\,\cdot\,)$ is a gradient flow with respect to $\mathcal{E}$, by Lemma 3.5.3

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}d_W^2\left(\mu(t),\nu\right) &= -\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(t,x),T(x,y)\right)d\gamma_t(x,y)\\ &= \int_{\mathcal{M}\times\mathcal{M}} g_x\left(\kappa(t,x),T(x,y)\right)d\gamma_t(x,y)\\ &\leq \mathcal{E}(\nu)-\mathcal{E}\left(\mu(t)\right)-\lambda d_W^2\left(\mu(t),\nu\right),\end{aligned}$$

for a.e. $t>0$ and $\gamma_t\in\Gamma_o(\mu(t),\nu)$, which implies the system of evolution variational inequalities.

If $\mu(\,\cdot\,)$ satisfies the system of evolution variational inequalities (3.1.16), then

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}d_W^2(\mu(t),\nu) &= -\int_{\mathcal{M}\times\mathcal{M}} g_x\left(v(t,x),T(x,y)\right)d\gamma_t(x,y)\\ &\leq \mathcal{E}(\nu)-\mathcal{E}(\mu(t))-\lambda d_W^2(\mu(t),\nu).\end{aligned}$$

By the definition of subdifferential of $\mathcal{E}$, we know that $v(t)\in-\partial\mathcal{E}\left(\mu(t)\right)$ for a.e. $t>0$, and thus $\mu(\,\cdot\,)$ is a gradient flow with respect to $\mathcal{E}$.

Thus gradient flow is characterized by the system of evolution variational inequalities. $\square$

3.6 λ-geodesic convexity of $\mathcal{E}$

In this Section, we present the details on obtaining conditions on g, W, V to guarantee λ-geodesic convexity of $\mathcal{W},\mathcal{V}$ and thus $\mathcal{E}$. We also give some examples of Riemannian manifolds $(\mathcal{M},g)$, on which we derive explicit conditions on W,V for $\mathcal{E}$ to be λ-geodesically convex. In particular we consider examples which explore how far can the conditions for λ-convexity be extended. Let us also mention that the general conditions when only the external potential, V, is present follow from the work of Sturm [101], who studied them together with internal energy.

We first show that conditions (NL2) and (NL5) imply the geodesic (semi-)convexity of $\mathcal{W}$ and $\mathcal{V}$ respectively. For any $\mu,\nu\in\mathcal{P}_2(\mathcal{M})$ and $\mu(t)$ a constant speed geodesic connecting them,

$$\mu(t)=(\gamma_t)_\sharp\pi \text{ for some } \pi\in\Gamma_o(\mu,\nu),$$

where for any $x,y\in\mathcal{M}$, γ_t is some constant speed minimal geodesic on $\mathcal{M}$ connecting them, see [39] and reference therein for details. Then by (NL2)

$$\begin{aligned}
\mathcal{W}(\mu(t)) &= \frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} W(x-y)d\mu(t,x)d\mu(t,y) \\
&= \frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}}\int_{\mathcal{M}\times\mathcal{M}} W(\gamma_t(x,z)-\gamma_t(y,w))d\pi(x,z)d\pi(y,w) \\
&\leq \frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}}\int_{\mathcal{M}\times\mathcal{M}} \Big((1-t)W(x-y)+tW(z-w) \\
&- \frac{\lambda}{2}t(1-t)\operatorname{dist}^2((x,y),(z,w))\Big)d\pi(x,z)d\pi(y,w) \\
&= (1-t)\mathcal{W}(\mu)+t\mathcal{W}(\nu)-\frac{\lambda}{2}t(1-t)d_W^2(\mu,\nu).
\end{aligned}$$

Similarly for $\mathcal{V}$, by (NL5)

$$\mathcal{V}(\mu(t)) \leq (1-t)\mathcal{V}(\mu)+t\mathcal{V}(\nu)-\frac{\lambda}{2}t(1-t)d_W^2(\mu,\nu).$$

Thus $\mathcal{W}$ is 2λ-geodesically convex, $\mathcal{V}$ is λ-geodesically convex and $\mathcal{E}$ is 2λ-geodesically convex (i.e. $\mathcal{E}$ is geodesically (semi-)convex with convexity constant 2λ).

We now turn to more detailed investigation of conditions for geodesic convexity. By Proposition 9.1.3 of [5] it is sufficient to verify the convexity along geodesics starting at absolutely continuous measure, μ. We derive the general formula of $\frac{d^2}{dt^2}\mathcal{E}\left(\mu(t)\right)$ for $\mu(t)$ geodesics in $\mathcal{P}_2(\mathcal{M})$ with $\mu(0)=\mu$. By [82, 39, 115], we know that geodesics starting from μ in $\mathcal{P}_2(\mathcal{M})$ are of the form

$$\mu(t) = (F_t)_\sharp \mu$$

where $F_t(x) = \exp_x(t\nabla\phi)$ is the geodesic on $\mathcal{M}$. We write $x_t = F_t(x)$, for simplicity. By definition of push forward of measures and recalling $w(x,y)=W(x-y)$,

$$\mathcal{E}\left(\mu(t)\right) = \mathcal{W}\left(\mu(t)\right)+\mathcal{V}\left(\mu(t)\right) = \frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} w(x_t,y_t)d\mu(x)d\mu(y)+\int_{\mathcal{M}} V(x_t)d\mu(x). \tag{3.6.1}$$

Since x_t and y_t are geodesics on $\mathcal{M}$, (x_t,y_t) is a geodesic on the product manifold $\mathcal{M}\times\mathcal{M}$. When W,V are twice differentiable direct computation shows:

$$\begin{aligned}
\frac{d^2}{dt^2}\mathcal{E}\left(\mu(t)\right) = &\int_{\mathcal{M}} \operatorname{Hess}_{\mathcal{M}} V(x_t)(\dot{x}_t,\dot{x}_t)d\mu(x) \\
&+\frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} \operatorname{Hess}_{\mathcal{M}\times\mathcal{M}} w(x_t,y_t)(\dot{x}_t,\dot{y}_t)(\dot{x}_t,\dot{y}_t)d\mu(x)d\mu(y)
\end{aligned} \tag{3.6.2}$$

where $\operatorname{Hess}_{\mathcal{M}}, \operatorname{Hess}_{\mathcal{M}\times\mathcal{M}}$ are Hessian on $(\mathcal{M},g)$ and $(\mathcal{M}\times\mathcal{M}, g\times g)$. So to verify convexity it suffices to show that there exists $\lambda\in\mathbb{R}$ such that for all vector fields $\dot{x}_t$ as above that

$$\frac{d^2}{dt^2}\mathcal{E}\left(\mu(t)\right) \geq \lambda\int_{\mathcal{M}} g_{x(t)}(\dot{x}_t,\dot{x}_t)d\mu(x).$$

So in general, λ-geodesic convexity of V on $(\mathcal{M}, g)$ and w on $(\mathcal{M}\times\mathcal{M}, g\times g)$ implies λ-geodesic convexity of $\mathcal{E}$. Actually, by [101], the potential energy $\mathcal{V}$ is λ-geodesic convex if and only if $\operatorname{Hess}_{\mathcal{M}} V \geq \lambda g$. Since $\mathcal{M}$ is a subset of $\mathbb{R}^d$ and $w(x,y) = W(x-y)$, we can expand $\frac{d^2}{dt^2}\mathcal{E}(\mu(t))$ in local coordinates,

$$\begin{aligned}\frac{d^2}{dt^2}\mathcal{E}(\mu(t)) =& \frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}}\Big(\operatorname{Hess} W(x_t,y_t)\,(\dot{x}_t,\dot{y}_t)\,(\dot{x}_t,\dot{y}_t)\\ &+\sum_{k,i,j}\frac{\partial W}{\partial z_k}(x_t-y_t)\left(-\Gamma^k_{ij}(x_t)(\dot{x}_t)_i(\dot{x}_t)_j+\Gamma^k_{ij}(y_t)(\dot{y}_t)_i(\dot{y}_t)_j\right)\Big)d\mu(x)d\mu(y)\\ &+\int_{\mathcal{M}}\Big(\operatorname{Hess} V(x_t)\,(\dot{x}_t,\dot{y}_t)+\sum_{k,i,j}\frac{\partial V}{\partial z_k}(x_t)(-1)\Gamma^k_{ij}(x_t)(\dot{x}_t)_i(\dot{x}_t)_j\Big)d\mu(x),\end{aligned} \tag{3.6.3}$$

where Γ^k_{ij} are the Christoffel symbols on $(\mathcal{M}, g)$. This verifies the simple conditions we give in Section 3.1. Indeed, $\operatorname{Hess}_{\mathcal{M}} V_{ij} = \operatorname{Hess} V_{ij} - \frac{\partial V}{\partial z_k}\Gamma^k_{ij}$ and $\Gamma^k_{ij} = \frac{1}{2}A_{km}\left(\frac{\partial G_{mi}}{\partial x_j}+\frac{\partial G_{mj}}{\partial x_i}-\frac{\partial G_{ij}}{\partial x_m}\right)$, the formula (3.6.3) allows us to conclude:

- If $(\mathcal{M}, g)$ is geodesically convex and compact with $G \in C^1(\mathcal{M})$, then any $V \in C^2(\mathcal{M})$ is λ-geodesically convex and $W \in C^2(\mathbb{R}^d)$ is λ-geodesically convex. Indeed, $\operatorname{Hess} V_{ij}, \nabla V$ and Γ^k_{ij} are bounded on $\mathcal{M}$, so $\operatorname{Hess}_{\mathcal{M}} V \geq C I_d \geq \tilde{C} G$ for all $x \in \mathcal{M}$.

- If g is C^1 bounded from below with bounded first derivative, and $V \in C^2(\mathcal{M})$ with bounded first and second derivative, then V is λ-geodesically convex on $(\mathcal{M}, g)$.

- If g is C^1 bounded from below and $V \in C^2(\mathcal{M})$ with $\operatorname{Hess} V \geq c I_d$ such that $\Gamma^k_{ij}\frac{\partial V}{\partial z_k}$ is bounded from above on $\mathcal{M}$, then V is λ-geodesically convex.

One obtains similar conditions on W:

- If g is C^1 bounded from below with bounded first derivative, and $W \in C^2(\mathcal{M})$ with bounded first and second derivative, then $w(x,y) = W(x-y)$ is λ-geodesically convex on $(\mathcal{M}\times\mathcal{M}, g\times g)$.

- If $(\mathcal{M}, g)$ is geodesically convex and compact with $g \in C^1(\mathcal{M})$, then for any W twice differentiable with $\operatorname{Hess} W(y) \geq -cI_d$ for all $y \in \mathcal{M}-\mathcal{M} = \{x^1 - x^2 : x^1 \in \mathcal{M}, x^2 \in \mathcal{M}\}$ and some constant $c > 0$. Note that since $\mathcal{M}$ is compact, $g \in C^1(\mathcal{M})$ and W twice differentiable imply there exist constants $c_1 > 0, c_2 > 0$ such that $c_1 I_d \leq G(x) \leq \frac{1}{c_1} I_d$, $|\frac{\partial}{\partial x_k} G_{ij}| \leq \frac{1}{c_1}$ and $|\nabla W(y)| \leq c_2$ for all $x \in \mathcal{M}$ and $y \in \mathcal{M}-\mathcal{M}$, w is λ-geodesically convex on $(\mathcal{M}\times\mathcal{M}, g\times g)$. In particular, for any $W \in C^2(\mathbb{R}^d)$, $w(x,y) = W(x-y)$ is λ-geodesically convex on $(\mathcal{M}\times\mathcal{M}, g\times g)$ for $(\mathcal{M}, g)$ geodesically convex and compact with $g \in C^1(\mathcal{M})$.

Note that the coupling between ∇W and Γ_{ij}^k is of the form $\frac{\partial W}{\partial z_k}(x-y)\Gamma_{ij}^k(x)$, so we do not have the same conditions as the second item for the λ-geodesic convexity of V. This coupling prevents us from getting some simple conditions of W, g to ensure λ-geodesic convexity of W, even in the 1-D case. It is more transparent in the 1-D examples of W, Example 3.6.3.

We now investigate conditions on V, W. Let us first focus on potential V:

Example 3.6.1. *Consider $d=1$ and $(\mathcal{M}, g) = (\mathbb{R}_+^1, g(x))$, then conditions for λ-geodesic convexity of V is*

$$V''(x) - \frac{g'(x)}{2g(x)}V'(x) \geq \lambda g(x). \tag{3.6.4}$$

- *$g(x) = x^p$ for some $p<0$, then $V(x) = V_0 + \int_1^x y^{\frac{p}{2}}U(y)dy$ is λ-geodesically convex if $U \in C^1(\mathbb{R}_+^1)$ with $x^{-\frac{p}{2}}U'(x) \geq C$ for all $x>0$ and some constant C. Moreover, V is geodesically convex if $U'(x) \geq 0$ for all $x>0$. In particular, it is straightforward to check $V(x) = x^q$ for $q \geq \max\{0, \frac{p}{2}+1\}$ or $q \leq \min\{0, \frac{p}{2}+1\}$ is geodesically convex. Indeed, (3.6.4) becomes*
$$V''(x) - \frac{p}{2x}V'(x) \geq \lambda x^p,$$
which is
$$\left(x^{-\frac{p}{2}}V'(x)\right)' \geq \lambda x^{\frac{p}{2}}$$
for $x>0$. Since $U(x) = x^{-\frac{p}{2}}V'(x)$, the last condition becomes $U'(x) \geq \lambda x^{\frac{p}{2}}$. So for any $U \in C^1\left(\mathbb{R}_+^1\right)$ with $x^{-\frac{p}{2}}U'(x) \geq C$ for some constant C, $V(x) = V_0 + \int_1^x y^{\frac{p}{2}}U(y)dy$ is λ-geodesically convex on $(\mathcal{M}, g)$. If $U'(x) \geq 0$, then V is geodesically convex on $(\mathcal{M}, g)$.

- *$g(x) = e^{\frac{p}{x}}$ for some $p>0$, then $V = V_0 + \int_1^x e^{\frac{p}{2y}}U(y)dy$ is λ-geodesically convex on $(\mathcal{M}, g)$, if $U \in C^1\left(\mathbb{R}_+^1\right)$ with $e^{-\frac{p}{2x}}U'(x) \geq C$ for all $x>0$ and some constant C. If $U'(x) \geq 0$ for all $x>0$, then V is geodesically convex on $(\mathcal{M}, g)$. In particular, $V(x) = x^q$ is geodesically convex for $q \geq 1$ and λ-geodesically convex for $q<1$. Similarly to the above case , the differential inequality (3.6.4) becomes*
$$V''(x) + \frac{p}{2x^2}V'(x) \geq \lambda e^{\frac{p}{x}},$$
which implies
$$\left(e^{-\frac{p}{2x}}V'(x)\right)' \geq \lambda e^{\frac{p}{2x}},$$
for all $x>0$. Take $U(x) = e^{-\frac{p}{2x}}V'(x)$, we have $U'(x) \geq \lambda e^{\frac{p}{2x}}$ and $V(x) = V_0 + \int_1^x e^{\frac{p}{2x}}U(y)dy$. Notice that for any $U \in C^1(\mathbb{R}_+)$ with $U'(x) \geq C$ for some constant C, we have there exists $\lambda \in \mathbb{R}$ such that $U'(x) \geq \lambda e^{\frac{p}{2x}}$ since $e^{\frac{p}{2x}}$ is bounded from below. And if $U'(x) \geq 0$ we can take $\lambda = 0$. So for any $U \in C^1(\overline{\mathbb{R}_+})$, such that U' is bounded from below, then $V(x) = V(0) + \int_0^x e^{\frac{p}{2y}}U(y)dy$ is λ-geodesically convex on $(\mathcal{M}, g)$.

Example 3.6.2. *Consider the upper half space,* $\mathbb{R}^{d-1} \times [0, \infty)$ *endowed with a Riemannian metric given by*

$$G(x) = \begin{bmatrix} g(x_d) I_{d-1} & 0 \\ 0 & 1 \end{bmatrix}.$$

Then

$$\Gamma_{ij}^k = \begin{cases} \frac{1}{2} g^{-1}(x_d) g'(x_d) & \text{if } \{i,j\} = \{k,d\}, k < d, \\ -\frac{1}{2} g'(x_d) & \text{if } i = j < d, k = d, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6.5}$$

Let $\mathcal{M}$ *be a compact, geodesically convex subset of* $\mathbb{R}_+^d$ *with* C^1 *boundary. For any* $V \in C^2(\mathbb{R}_+^d), W \in C^2(\mathbb{R}^d)$*,* V, w *are* λ*-geodesically convex on* $(\mathcal{M}, g)$ *and* $(\mathcal{M} \times \mathcal{M}, g \times g)$*.*

Consider now $d = 2$ *and* $g(x_2) = x_2^p$ *with* $p < 0$*. For simplicity, we assume that* $\mathcal{M}$ *contains portion of* $x_2 = 0$*. We note that the metric is degenerate. Nevertheless investigate if* $V(x) = |x|^2$ *should be* λ*-convex in some generalized sense. Direct computation shows*

$$\operatorname{Hess}_{\mathcal{M}} V(x) = \begin{bmatrix} 2 + p x_2^p & -p x_1 x_2^{-1} \\ -p x_1 x_2^{-1} & 2 \end{bmatrix}.$$

For V *to be* λ*-convex it is necessary that*

$$\begin{aligned} 2 &\geq \lambda, \\ (2 - \lambda)\left(2 + (p - \lambda) x_2^p\right) - p^2 x_1^2 x_2^{-2} &\geq 0. \end{aligned}$$

By taking $x_2 \to 0^+$ *shows that no* $\lambda \in \mathbb{R}$ *can satisfy these conditions.*

In general the conditions for the λ-geodesic convexity of V and w are rather restrictive, as claimed in Remark 3.1.1. The next example illustrates why.

Example 3.6.3. *Take* $(\mathcal{M}, g)$ *to be* $(\mathbb{R}, g)$*. Then the* λ*-geodesic convexity condition for* $w(x, y) = W(x - y)$ *is*

$$\begin{bmatrix} W''(x-y) - \frac{1}{2} W'(x-y) g^{-1}(x) g'(x) & -W''(x-y) \\ -W''(x-y) & W''(x-y) + \frac{1}{2} W'(x-y) g^{-1}(y) g'(y) \end{bmatrix} \geq \lambda \begin{bmatrix} g(x) & 0 \\ 0 & g(y) \end{bmatrix}.$$

In particular it is necessary that for all $x, y \in \mathbb{R}$

$$W''(x-y) - \frac{1}{2} W'(x-y) g^{-1}(x) g'(x) \geq \lambda g(x).$$

One should contrast this condition with condition (3.6.4) *for potential* V*. In particular the condition above shows the presence of long-range effects which make it hard for the condition to be satisfied. For example, if* $W(z) = z^2$*, and* $g(z) = 2 + \frac{\sin(z)}{1+z^2}$ *then the condition above becomes*

$$2 - (x - y) \frac{(1 + x^2)\cos(x) - 2x \sin(x)}{2(1 + x^2) + \sin x} \geq \lambda g(x)$$

taking x such that the term next to $(x-y)$ is negative and then taking $y \to \infty$ shows that there is no λ for which the condition is satisfied.

Nevertheless a usable sufficient condition for $w(x, y) = W(x-y)$ to be λ-convexity can be found. For example $W \in C^2(\mathbb{R})$, with W even, W', W'' bounded, $g \in C^1(\overline{\mathbb{R}})$ with $g \geq C > 0$ and g' bounded suffices.

3.7 Numerical simulations

In this Section, we study rolling swarms from (3.1.5) and show some numerical simulations giving rolling swarms in heterogeneous environments with boundaries.

The rolling motion is an interesting phenomenon observed in real locust swarms, where locusts at the front of the swarm fly downward and those at the back fly upward while all of them are moving forward in pursuit of food. The mathematical models and numerical simulations of rolling locust swarms have been investigated in [13, 107, 110]. In particular, in [13] Bernoff and Topaz introduced a model and performed numerical simulations on the upper half plane $\mathbb{R}^2_+$ showing the existence of rolling swarms by imposing that on the boundary, the horizontal velocity is zero.

We observe that, unlike in homogeneous environments, rolling swarms often emerge in heterogeneous environments. Here we consider the simple heterogeneous environments with stratified mobility

$$A(x) = \begin{bmatrix} x_2^2 & 0 \\ 0 & 1 \end{bmatrix}$$

for $x = (x_1, x_2)$ on $\mathbb{R}^2_+$. We use radially symmetric, repulsive-attractive interaction forces as used in [70],

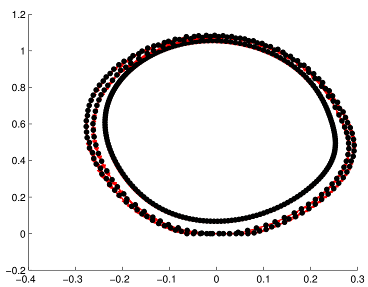
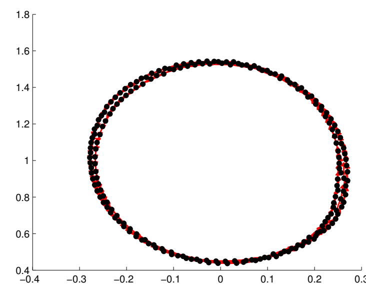
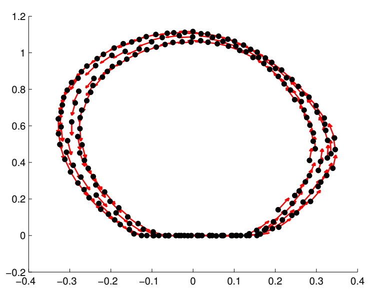
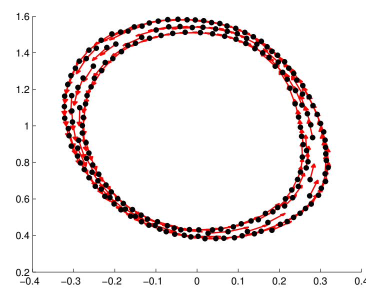
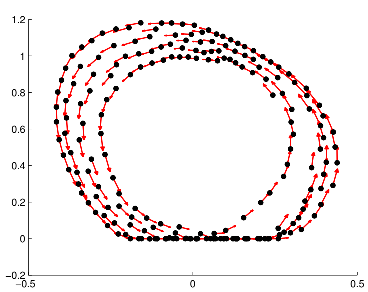
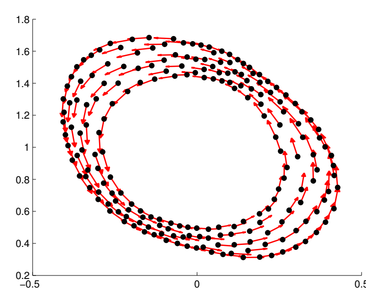
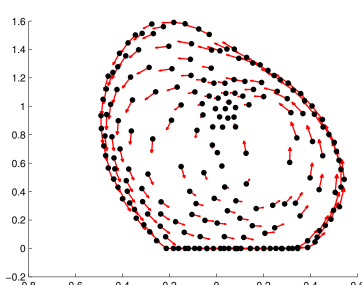
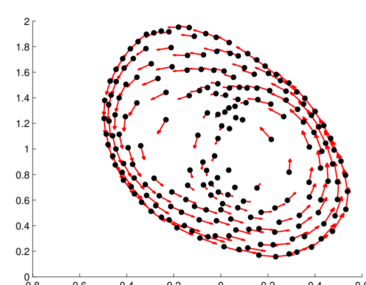
$$F(r) = \tanh[(1-r)a] + b; \quad 0 < a; \quad -\tanh(a) < b < 1.$$

With external potential V and equal mass $\frac{1}{N}$ for every agent, the dynamics become

$$\dot{x}^i(t) = \frac{1}{N} \sum_{j \neq i}^{N} F(|x^i(t) - x^j(t)|) \frac{x^i(t) - x^j(t)}{|x^i(t) - x^j(t)|} - \nabla V(x^i(t)), \qquad \forall i \in \{1, \dots, N\}.$$

In our simulations, we take the number of agents N to be 200; we put the external force to move agents to the left; we either use $\nabla V = (0.1, 0)$ (without gravity) or $\nabla V = (0.1, 0.005)$ (with gravity); we take random initial data from the upper half plane. We perform numerical simulations with different parameters a, b in the interaction force F, it turns out that for a large range of a, b, rolling swarms emerge naturally (with or without gravity). Depending on the strength of interaction potential, external force and the mobility, numerical simulations show various rolling patterns such as rolling with 1-dimensional support and rolling swarms with 2-dimensional support. With gravity, and appropriate scaling between interaction and external force, the rolling swarm has a bubble shape with some portion of agents on the ground (boundary of the domain) and another portion in the air, as desired in locust swarm, see [13] and references therein.

Here we present some results of the numerical simulations of the dynamics for $a = 5$, $b = -0.3, -0.1, 0.1, 0.3$, and with or without gravity at $t = 100$.

	$\nabla V = (0.1, 0.005)$	$\nabla V = (0.1, 0)$
$a = 5, b = -0.3$		
$a = 5, b = -0.1$		
$a = 5, b = 0.1$		
$a = 5, b = 0.3$		

In the figures, the black points and red arrows are positions and velocities of the agents. We can observe from the simulations that, rolling swarms emerge natural from simulations; with the presence of gravity, the swarms have a portion of agents on the boundary (ground) forming a bubble shape; as b increases, the shape of the swarms change from rolling rings (with 1-dimensional support) to rolling swarms with 2-dimensional support.

Chapter 4

Nonlocal interaction equations in heterogeneous environments with boundaries: compactly supported initial data case

We devote this Chapter to the well-posedness of nonlocal interaction equations in heterogeneous environments with boundaries given that the initial data μ_0 has compact support, i.e., $\mathrm{supp}(\mu_0) \Subset \mathcal{M}$. The motivation is that, it turns out that for mildly heterogeneous environments, even some natural interaction potentials (for example $W(x) = \frac{1}{2}|x|^2$) are not globally geodesically (semi-)convex, thus we cannot use our previous general well-posedness results established in Chapter 3. However, in many applications from biology, the initial distributions μ_0 of biological agents have compact support. In this case, we develop gradient flow theory that applies to a much wider class of interaction potentials (with only weaker, local conditions imposed). The strategy is to control the growth of support of the discrete solutions from JKO scheme, thus weaker, local conditions on interaction potential still imply the well-posedness of the nonlocal interaction equations. We are using the notations developed in the previous Chapter 3 and following our paper [118] in this Chapter.

Main assumptions and results.

In the rest of this Chapter, we assume that μ_0 has compact support, i.e., $\mathrm{supp}(\mu_0) \Subset \mathcal{M}$. For $\mathcal{M}$ a d-dimensional subset in $\mathbb{R}^d$ with C^2 boundary, we still assume:

(M1) The Riemannian metric g is C^2 and satisfies $|\xi|_g^2 \geq \Lambda|\xi|^2$ for some constant $\Lambda > 0$ and all $\xi \in T\mathcal{M}$.

(M2) $(\mathcal{M}, g)$ is geodesically convex in that for all $x, y \in \mathcal{M}$ there exists a length minimizing geodesic contained in $\mathcal{M}$.

On the interaction and external potentials W, V, we make the following assumptions:

(LNL1) $W(0) = 0$ and W is symmetric: $W(x) = W(-x)$.

(LNL2) $W \in C^2(\mathbb{R}^d)$ is semi-convex on $\mathbb{R}^d$ with respect to the Euclidean metric (with semi-convexity constant $\Theta \in \mathbb{R}$) and ∇W has the linear growth condition $|\nabla W(x)| \leq C(1 + \mathrm{dist}(x, x_0))$ for some constant $C \geq 0$.

(LNL3) $$\liminf_{\mathrm{dist}((x,y),(x_0,x_0))\to\infty} \frac{W(x-y)}{\mathrm{dist}^2(x, x_0) + \mathrm{dist}^2(y, x_0)} \geq 0.$$

(LNL4) $V \in C^2(\mathcal{M})$ is semi-convex on $\mathcal{M}$ with respect to the Euclidean metric (with semi-convexity constant $\Theta \in \mathbb{R}$) and ∇V has the linear growth condition $|\nabla V(x)| \leq C(1 + \mathrm{dist}(x, x_0))$ for all $x \in \mathcal{M}$.

(LNL5) $$\liminf_{\mathrm{dist}(x,x_0)\to\infty} \frac{V(x)}{\mathrm{dist}^2(x, x_0)} \geq 0.$$

Remark 4.0.1. *About (LNL2) and (LNL3), both Lemma 4.1.3, Theorem 4.1.4 and Theorem 4.1.8 still work for the case when (LNL2) and (LNL3) are replaced by the following repulsive-attractive condition:* $W(x) = w(|x|)$ *with* $w \in C^2((0, \infty))$ *satisfying that there exists constants* $R_a > 0, C_W > 0$*, such that* $w'(r) \geq 0$ *for* $r > R_a$ *and* $w'(r) \geq -C_W$ *for* $0 < r < R_a$.
Thus we still have existence of weak measure solutions for such W*, i.e., Theorem 4.0.3 holds for such* W.

For such potentials W, V and Riemannian manifold $(\mathcal{M}, g)$, we still want to show (3.1.5) as a gradient flow of $\mathcal{E}$ in space of probability measures endowed with the Riemannian Wasserstein metric, refer to Remark 4.0.2.

Remark 4.0.2. *In Chapter 3, when* W, V *are geodesically convex on* $(\mathcal{M} \times \mathcal{M}, g \times g)$ *and* $(\mathcal{M}, g)$*, for any initial data* $\mu_0 \in \mathcal{P}_2(\mathcal{M})$*, we obtain the existence and stability of gradient flows of* $\mathcal{E}$ *in* $(\mathcal{P}_2(\mathcal{M}), d_W)$*. Then we also prove that the gradient flows are weak measure solutions to (3.1.5) and satisfies the system of evolution variational inequalities. In this Chapter we show that under weaker conditions on* W, V*, namely only locally* λ*-geodesic convexity of* W, V *implies the existence of weak measure solutions to (3.1.5) given that the initial data* μ_0 *has compact support, i.e.* $\mathrm{supp}(\mu_0) \Subset \mathcal{M}$*. We also get the stability of weak measure solutions with specific support growth conditions. Note that in general, the* λ*-geodesic convexity of* W, V *can be difficult to check, refer to Section 3.6. However, the local* λ*-geodesic convexity is implied by the smoothness conditions of* W, V *see Proposition*

4.1.7. The weak measure solution we get is not necessarily a gradient flow, in the sense that the tangent velocity $v(t)$ of the absolutely continuous curve $\mu(\cdot)$ only satisfies the local slope definition

$$\mathcal{E}(\nu) - \mathcal{E}(\mu(t)) \geq \int_{\mathcal{M}\times\mathcal{M}} g_x\left(-v(t,x), T(x,y)\right) d\gamma_t(x,y) + o\left(d_W(\mu(t),\nu)\right) \tag{4.0.1}$$

for $\nu \in \mathcal{P}_2(\mathcal{M})$ with compact support and $\gamma_t \in \Gamma_o(\mu(t),\nu)$ an optimal plan. Refer to Section 3.3 for the definitions of $T(x,y)$ and local slope, and Section 3.5 for the definitions of tangent velocity and gradient flow. The advantage is the that we still get the desired energy dissipation (4.0.3) for the weak measure solution even though it is possibly not a gradient flow.

The main result in this Chapter is the following theorem about existence and stability for weak measure solutions for initial data with compact support.

Theorem 4.0.3. *Given that (M1)-(M2), (LNL1)-(LNL5) hold and* $\mathrm{supp}(\mu_0)$ *is compact, i.e.,* $\mathrm{supp}(\mu_0) \Subset \mathcal{M}$*, then there exists a weak measure solution $\mu(\cdot)$ to (3.1.5) satisfying for a.e. $t > 0$*

$$|\mu'|^2(t) = \int_{\mathcal{M}} g_x\left(\kappa(t,x), \kappa(t,x)\right) d\mu(t,x), \tag{4.0.2}$$

and the following energy dissipation equality, for any $0 \leq s < t < \infty$

$$\mathcal{E}\left(\mu(s)\right) = \mathcal{E}\left(\mu(t)\right) + \int_s^t \int_{\mathcal{M}} g_x\left(\kappa(r,x), \kappa(r,x)\right) d\mu(r,x)dr, \tag{4.0.3}$$

where $\kappa(t,x) = -P_x\left(-A(x)\left(\nabla W * \mu(t)(x) + \nabla V(x)\right)\right)$. Moreover, if we have two such solutions $\mu^i(\cdot)$ with initial data μ_0^i satisfying for $i = 1, 2$, $\mathrm{supp}(\mu_0^i) \subset B(r_0)$ *and* $\mathrm{supp}(\mu^i(t)) \subset B\left(r(t)\right)$ *for all $t > 0$, then*

$$d_W\left(\mu^1(t), \mu^2(t)\right) \leq \exp(-\lambda_k t) d_W\left(\mu_0^1, \mu_0^2\right), \tag{4.0.4}$$

where λ_k is the geodesic convexity constant of W, V in $K_k \supset B\left(2r(t)\right)$.

Outline.

Section 4.1 is devoted to the JKO scheme. We show that the discrete scheme is well-posed and converges to a locally absolutely continuous curve $\mu(\cdot)$ in $\mathcal{P}_2(\mathcal{M})$. We then show that the support of the limit curve has exponential growth and lower semi-continuity of $\|\kappa\|_{L^2(g,\mu)}$, which implies the limit curve $\mu(\cdot)$ is a curve of maximal slope with respect to $\|\kappa\|_{L^2(g,\mu)}$.

In Section 4.2, we establish that the limit curve $\mu(\cdot)$ we get from JKO scheme is actually a weak measure solutio to (5.0.1) by showing that $\mathcal{E}(\mu(\cdot))$ satisfies the desired chain rule. We then show that local λ-convexity of the functional $\mathcal{E}$ implies the stability of the weak measure solutions with growth conditions on supports.

4.1 Existence and Convergence of JKO scheme

In this Section, we check the topological conditions about the functional $\mathcal{E}$ to apply the general existing theorem in Subsection 2.2.1 to get a curve of maximal slope with respect to the relaxed local slope $|\partial^-\mathcal{E}|$. We then show the exponential growth of support of the limit curve $\mu(\,\cdot\,)$ and the lower semi-continuity of $\|\kappa\|_{L^2(g,\mu)}$.

Recall from Subsection 2.2.1 and Section 3.4 the notations we use in JKO scheme (3.4.1). The strategy is to show that there exists a subsequence $\tau_n \to 0$, such that $\tilde{\mu}^n(\,\cdot\,) = \mu_{\tau_n}(\,\cdot\,)$ converges narrowly to a curve of maximal slope $\mu(\,\cdot\,)$. We now again check that the topological conditions (Lower semicontinuity, Coercivity and Compactness) introduced in Subsection 2.2.1 and used in Section 3.4 for the general theory to apply still hold true for our energy functional $\mathcal{E}$ in this compactly supported setting.

Notice that by the same arguments as in Proposition 3.2.3 and Proposition 3.2.4 from Chapter 3 give that $\mathcal{E}$ is lower semicontinuous with respect to narrow convergence of probability measures, thus Lower semicontinuity condition is checked. To check Compactness condition, note that by Prokhorov's theorem, any sequence $(\mu_n) \subset \mathcal{P}_2(\mathcal{M})$ such that $\sup_{m,n} d_W(\mu_m, \mu_n) < \infty$, μ_n has a narrowly convergent subsequence. Thus we only need to check Coercivity, that is, there exists $\tau_* > 0$ and $\mu_* \in \mathcal{P}_2(\mathcal{M})$ such that

$$\inf_{\mu \in \mathcal{P}_2(\mathcal{M})} \left\{ \mathcal{E}(\mu) + \frac{1}{2\tau_*} d_W^2(\mu, \mu_*) \right\} > -\infty.$$

To show it is true for $\mathcal{E}$, by (LNL3) and (LNL5), for any $\epsilon > 0$ we know $W(x-y) + \epsilon\left(\mathrm{dist}^2(x,x_0) + \mathrm{dist}^2(y,x_0)\right) > C$ and $V(x) + \epsilon\,\mathrm{dist}^2(x,x_0) > C$ for some constant $C = C(\epsilon)$. So for any fixed $\epsilon > 0$ and $x_0 \in \mathcal{M}$,

$$\begin{aligned}
&\mathcal{E}(\mu) + \frac{1}{2\tau} d_W^2(\mu, \delta_{x_0}) \\
=&\frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} W(x-y)\,d\mu(x)d\mu(y) + \int_{\mathcal{M}} V(x)\,d\mu(x) + \frac{1}{2\tau}\int_{\mathcal{M}} \mathrm{dist}^2(x,x_0)\,d\mu(x) \\
\geq&\frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} \left(-\epsilon\left(\mathrm{dist}^2(x,x_0) + \mathrm{dist}^2(y,x_0)\right) + C\right) d\mu(x)d\mu(y) \\
&+ \int_{\mathcal{M}} \left(-\epsilon\,\mathrm{dist}^2(x,x_0) + C\right) d\mu(x) + \frac{1}{2\tau}\int_{\mathcal{M}} \mathrm{dist}^2(x,x_0)\,d\mu(x) \\
=&\frac{3}{2}C + \int_{\mathcal{M}} \left(\frac{1}{2\tau} - 2\epsilon\right) \mathrm{dist}^2(x,x_0)\,d\mu(x).
\end{aligned}$$

So for any τ such that $\frac{1}{2\tau} \geq 2\epsilon$ i.e. $\tau \leq \frac{1}{4\epsilon}$, we have

$$\inf_{\mu \in \mathcal{P}_2(\mathcal{M})} \left\{ \mathcal{E}(\mu) + \frac{1}{2\tau} d_W^2(\mu, \delta_{x_0}) \right\} > -\infty,$$

which implies Coercivity condition for $\mathcal{E}$. Thus we again get the following existence and compactness results by Lemma 2.2.8 and Proposition 2.2.9.

Lemma 4.1.1 (Existence of the discrete solutions)**.** *Suppose* $(\mathcal{M}, g)$ *satisfies assumptions (M1)-(M2) and* W, V *satisfy (LNL1)-(LNL5). Then there exists* $\tau_0 > 0$ *depending only on* W *such that for all* $0 < \tau < \tau_0$ *and given* $\nu \in \mathcal{P}_2(\mathcal{M})$, *there exists* $\mu_\infty \in \mathcal{P}_2(\mathcal{M})$ *such that*

$$\mathcal{E}(\mu_\infty) + \frac{1}{2\tau} d_W^2(\nu, \mu_\infty) = \inf_{\mu \in \mathcal{P}_2(\mathcal{M})} \Big\{ \mathcal{E}(\mu) + \frac{1}{2\tau} d_W^2(\nu, \mu) \Big\}. \tag{4.1.1}$$

Proposition 4.1.2 (Compactness)**.** *There exist a limit curve* $\mu \in AC^2_{loc}([0,\infty); \mathcal{P}_2(\mathcal{M}))$ *and a sequence* $\tau_n \to 0^+$ *such that the piecewise constant interpolate* $\tilde{\mu}^n = \mu_{\tau_n}$ *defined as in (3.4.2) satisfies that* $\tilde{\mu}^n(t)$ *converges narrowly to* $\mu(t)$ *for any* $t \in [0, \infty)$.

Again the limit curve $\mu(\,\cdot\,)$ is a curve of maximal slope with respect to $|\partial^- \mathcal{E}|$ defined in (2.2.20). We also recall the following definition of

$$G_\tau(t) = \frac{d_W\left(\mu_\tau^{n-1}, \mu_\delta^{n-1}\right)}{\delta} \tag{4.1.2}$$

for $t = t_\tau^{n-1} + \delta \in (t_\tau^{n-1}, t_\tau^n]$, where $\mu_\delta^{n-1} \in \operatorname{argmin}_{\mu \in \mathcal{P}_2(\mathcal{M})} \left[\frac{d_W^2(\mu, \mu_\tau^{n-1})}{2\delta} + \mathcal{E}(\mu) \right]$. Refer to (3.2.2) from [5] for the details.

We now get some properties of the minimizer from the JKO scheme (3.4.1), in particular, the control of the support. Since the support of the initial data μ_0 satisfies $\operatorname{supp}(\mu_0) \subset B_{r_0}(x_0)$ for some $x_0 \in \mathcal{M}$ and $B_{r_0}(x_0) = \{x \in \mathcal{M} : \operatorname{dist}(x, x_0) \le r_0\}$, we now estimate the support of $\mu(t)$ in terms of r_0 and t. Without loss of generality, we assume $x_0 = 0 \in \mathcal{M}$ and denote $B_0(r_0)$ by $B(r_0)$ for short.

Lemma 4.1.3. *For* $0 < \tau \le \frac{\Lambda}{8\Theta^-}$, *we have* $\operatorname{supp}(\mu_1^\tau) \subset B(r_1)$ *where*

$$r_1 \le \frac{r_0 + C\tau}{1 - C\tau}, \tag{4.1.3}$$

for some constant C *depending only on* W, V *and the Riemannian metric* g.

Proof. Let $r_1 > r_0$, and assume that $\mu_1^\tau\left(B(r_1)^\complement\right) > 0$. We consider the variation of μ_1^τ defined by

$$\tilde{\mu}_1^\tau = \mu_1^\tau \lfloor_{B(r_1)} + (\pi_1)_\sharp \left(\gamma \lfloor_{\mathcal{M} \times B(r_1)^\complement}\right),$$

where $\gamma \in \Gamma_o(\mu_0, \mu_1^\tau)$ is an optimal plan between μ_0 and μ_1^τ. Since μ_1^τ is a minimizer of the

JKO scheme,

$$\begin{aligned}\mathcal{E}\left(\mu_1^\tau\right)+\frac{1}{2\tau}d_W^2\left(\mu_1^\tau,\mu_0\right) &\leq \mathcal{E}\left(\tilde{\mu}_1^\tau\right)+\frac{1}{2\tau}d_W^2\left(\tilde{\mu}_1^\tau,\mu_0\right) \\ &\leq \mathcal{E}\left(\mu_1^\tau\right)+\frac{1}{2\tau}d_W^2\left(\mu_1^\tau,\mu_0\right)-\frac{1}{2\tau}\int_{\mathcal{M}\times B(r_1)^{\complement}}\mathrm{dist}^2(x,y)d\gamma(x,y) \\ &+\int_{(y,\tilde{y})\in B(r_1)^{\complement}\times B(r_1)^{\complement}}\left(W(x-\tilde{x})-W(y-\tilde{y})\right)d\gamma(x,y)d\gamma(\tilde{x},\tilde{y}) \\ &+2\int_{(y,\tilde{y})\in B(r_1)^{\complement}\times B(r_1)}\left(W(x-\tilde{y})-W(y-\tilde{y})\right)d\gamma(x,y)d\mu_1^\tau(\tilde{y}) \\ &+\int_{\mathcal{M}\times B(r_1)^{\complement}}\left(V(x)-V(y)\right)d\gamma(x,y).\end{aligned} \tag{4.1.4}$$

For $(x,y)\in\left(\mathcal{M}\times B(r_1)^{\complement}\right)\cap\mathrm{supp}(\gamma)$,

$$V(x)-V(y)\leq C(1+r_0)|x-y|-\frac{\Theta}{2}|x-y|^2. \tag{4.1.5}$$

To see that, since V is Θ convex, we know

$$V(x)-V(y)\leq\langle\nabla V(x),x-y\rangle-\frac{\Theta}{2}|x-y|^2$$

and

$$|\nabla V(x)|\leq C(1+\mathrm{dist}(x,x_0))\leq C(1+r_0)$$

for $x\in\mathrm{supp}(\mu_0)$. So

$$V(x)-V(y)\leq C(1+r_0)|x-y|-\frac{\Theta}{2}|x-y|^2$$

as claimed. Similarly, for W we know:
For $(y,\tilde{y})\in B(r_1)^{\complement}\times B(r_1)$ and $x\in\mathrm{supp}(\mu_0)$,

$$\begin{aligned}W(x-\tilde{y})-W(y-\tilde{y}) &\leq\langle\nabla W(x-\tilde{y}),x-y\rangle-\frac{\Theta}{2}|x-y|^2 \\ &\leq C(1+r_0+r_1)|x-y|-\frac{\Theta}{2}|x-y|^2.\end{aligned} \tag{4.1.6}$$

For $(y,\tilde{y})\in B(r_1)^{\complement}\times B(r_1)^{\complement}$ and $x,\tilde{x}\in\mathrm{supp}(\mu_0)$,

$$\begin{aligned}W(x-\tilde{x})-W(y-\tilde{y}) &\leq\langle\nabla W(x-\tilde{x}),x-y-(\tilde{x}-\tilde{y})\rangle-\frac{\Theta}{2}|x-y-(\tilde{x}-\tilde{y})\,|^2 \\ &\leq C(1+r_0)\left(|x-y|+|\tilde{x}-\tilde{y}|\right)-\Theta\left(|x-y|^2+|\tilde{x}-\tilde{y}|^2\right).\end{aligned} \tag{4.1.7}$$

Plugging back into (4.1.4) and noticing that $r_1 > r_0$, we know for $\tau > 0$ such that $\tau \leq \frac{\Lambda}{8\Theta^-}$

$$
\begin{aligned}
&\left(\frac{1}{2\tau} + \frac{4\Theta}{\Lambda}\right) \int_{\mathcal{M}\times B(r_1)^\complement} \operatorname{dist}^2(x,y) d\gamma(x,y) \\
&\leq \int_{\mathcal{M}\times B(r_1)^\complement} C(1+r_0+r_1)|x-y| d\gamma(x,y) \\
&\leq C(1+r_1) \int_{\mathcal{M}\times B(r_1)^\complement} \operatorname{dist}(x,y) d\gamma(x,y) \\
&\leq C(1+r_1) \left(\int_{\mathcal{M}\times B(r_1)^\complement} \operatorname{dist}^2(x,y) d\gamma(x,y)\right)^{\frac{1}{2}} \left(\mu_1^\tau\left(B(r_1)^\complement\right)\right)^{\frac{1}{2}}.
\end{aligned}
$$

Since for any $(x,y) \in \operatorname{supp}(\gamma) \cap \left(\mathcal{M} \times B(r_1)^\complement\right) \subset B(r_0) \times B(r_1)^\complement$, $\operatorname{dist}(x,y) \geq r_1 - r_0$,

$$
\begin{aligned}
\left(\frac{1}{2\tau} + \frac{4\Theta}{\Lambda}\right)^2 (r_1 - r_0)^2 \mu_1^\tau\left(B(r_1)^\complement\right) &\leq \left(\frac{1}{2\tau} + \frac{4\Theta}{\Lambda}\right)^2 \int_{\mathcal{M}\times B(r_1)^\complement} \operatorname{dist}^2(x,y) d\gamma(x,y) \\
&\leq C(1+r_1)^2 \mu_1^\tau\left(B(r_1)^\complement\right),
\end{aligned}
$$

yielding

$$
r_1 \leq \frac{r_0 + C\tau}{1 - C\tau},
$$

for C a constant depending on W, V and g. $\square$

So after k iterations we have $\operatorname{supp}(\mu_k^\tau) \subset B(r_k)$ with

$$
r_k \leq \frac{r_0 + 1 - (1-C\tau)^k}{(1-C\tau)^k}. \tag{4.1.8}
$$

Fix $t > 0$ such that $(k-1)\tau < t \leq k\tau$, denote $k = \lceil \frac{t}{\tau} \rceil$, we have

$$
\operatorname{supp}(\mu^\tau(t)) = \operatorname{supp}(\mu_k^\tau) \subset B(r_k)
$$

with $r_k \leq \frac{r_0+1-(1-C\tau)^k}{(1-C\tau)^k}$. By taking $\tau_n \to 0$ we have for the limit curve $\mu(t)$, $\operatorname{supp}(\mu(t)) \subset B(r(t))$ with

$$
r(t) \leq \lim_{\tau_n \to 0} \frac{r_0 + 1 - (1-C\tau_n)^{\lceil \frac{t}{\tau_n} \rceil}}{(1-C\tau_n)^{\lceil \frac{t}{\tau_n} \rceil}} = \frac{r_0 + 1 - \exp(-Ct)}{\exp(-Ct)}. \tag{4.1.9}
$$

That is, $\mu(\cdot)$ has at most exponential growth of support.

We now show that for any $\mu \in \mathcal{P}_2(\mathcal{M})$ with $\operatorname{supp}(\mu) \subset B(r) \Subset \mathcal{M}$ for some $r > 0$, denote

$$
\kappa(x) = -P_x\left(-A(x)\left(\int_{\mathcal{M}} \nabla W(x-y) d\mu(y) + \nabla V(x)\right)\right),
$$

we have

Theorem 4.1.4. *Assume (M1)-(M2), (LNL1)-(LNL5) hold, then for any $\mu \in \mathcal{P}_2(\mathcal{M})$ with* $\operatorname{supp}(\mu) \subset B(r) \Subset \mathcal{M}$,

$$\int_{\mathcal{M}} g_x\left(\kappa(x), \kappa(x)\right) d\mu(x) \leq |\partial \mathcal{E}|^2(\mu). \tag{4.1.10}$$

Remark 4.1.5. *Since we already know that* $\operatorname{supp}(\mu) \subset B(r)$*, together with the regularity assumptions on W and V, we know that $\kappa \in L^2(g, \mu)$.*

Before proving the theorem, we need the following definition of local λ-geodesic convexity and proposition about the local λ-geodesic convexity property of W, V.

Definition 4.1.6. *A function $f \in C^0(\mathcal{M})$ is called locally λ-geodesically convex if there exist a sequence of compact subsets $\{K_n\}_{n\in\mathbb{N}}$ and a sequence of real numbers $\{\lambda_n\}_{n\in\mathbb{N}}$ such that $K_n \Subset \mathcal{M}$, $K_n \subset K_{n+1}$, $\bigcup_n K_n = \mathcal{M}$ and f is λ_n-geodesically convex on K_n.*

We now show that W, V are locally λ-geodesically convex.

Proposition 4.1.7. *Given that (M1)-(M2) hold true for $\mathcal{M}$ and W, V satisfy (LNL1)-(LNL5), then W is locally λ-geodesically convex on $\mathcal{M} \times \mathcal{M}$ and V is locally λ-geodesically convex on $\mathcal{M}$.*

Proof. For V, fix $x_0 \in \mathcal{M}$ consider

$$B(n) = \{x \in M : \operatorname{dist}(x, x_0) \leq n\} \Subset \mathcal{M}.$$

We know that $(\operatorname{Hess}_{\mathcal{M}} V)_{ij}(x) = (\operatorname{Hess} V)_{ij}(x) + \Gamma^k_{ij} \frac{\partial V}{\partial x_k}$ in local coordinates. Since $V \in C^2(\mathcal{M})$ we have that $(\operatorname{Hess}_{\mathcal{M}} V)_{ij}$ is uniformly bounded on $B(2n)$, thus $\operatorname{Hess}_{\mathcal{M}} V(x) \geq \lambda_n G(x)$ for all $x \in B(n)$ some λ_n depending on W, V, G. So V is λ_n-geodesically convex on $B(n)$, and V is locally λ-geodesically convex on $\mathcal{M}$. The proof for W is similar and we omit it here. $\square$

We can now prove the theorem.

Proof of Theorem. We claim that for any $\xi \in L^2(g, \mu)$ such that there exists $t_0 > 0$ with $\exp_x(t\xi(x)) \in \mathcal{M}$ for all $0 \leq t \leq t_0$ and $x \in \mathcal{M}$ and $g_x(\xi(x), \xi(x)) \leq n$,

$$\lim_{t\to 0^+} \frac{\mathcal{E}\left(\exp(t\xi)_\sharp \mu\right) - \mathcal{E}(\mu)}{t} = \int_{\mathcal{M}} g_x\left(A(x)\left(\int_M \nabla W(x-y) d\mu(y) + \nabla V(x)\right), \xi(x)\right) d\mu(x). \tag{4.1.11}$$

Indeed, note that if $x \in \operatorname{supp}(\mu) \subset B(r)$ and $g_x(\xi(x), \xi(x)) \leq n^2$, then $\operatorname{dist}^2(x, \exp_x(t\xi(x))) \leq t^2 g_x(\xi(x), \xi(x)) \leq n^2$ for any $0 \leq t \leq t_0 \leq 1$, which implies $\exp_x(t\xi(x)) \in B(r+n)$ for any $0 \leq t \leq t_0$ and $x \in \operatorname{supp}(\mu)$. Since W, V are locally λ-geodesically convex, let $\lambda = \lambda_k$ be such that $B(r+n) \subset K_k$ where W, V are λ-geodesically convex on $K_k \times K_k$ and K_k

respectively. Then by the same argument as in Step 2 in the proof of Theorem 3.3.4 from Chapter 3, we know

$$
\begin{aligned}
&\lim_{t\to 0^+} \frac{\mathcal{E}\left(\exp\left(t\xi\right)_\sharp \mu\right) - \mathcal{E}(\mu)}{t} \\
=&\lim_{t\to 0^+} \left(\int_{\mathcal{M}\times\mathcal{M}} \frac{W\left(\exp_x\left(t\xi(x)\right) - \exp_z\left(t\xi(z)\right)\right) - W(x-z)}{2t} d\mu(x)d\mu(z) \right.\\
&\left. + \int_{\mathcal{M}} \frac{V\left(\exp_x\left(t\xi(x)\right)\right) - V(x)}{t} d\mu(x) \right) \\
=&\frac{1}{2}\int_{\mathcal{M}\times\mathcal{M}} \left\langle \nabla W(x-z), \xi(x) - \xi(z) \right\rangle + 2\left\langle \nabla V(x), \xi(x) \right\rangle d\mu(x)d\mu(z) \\
=&\int_{\mathcal{M}} \left\langle \int_{\mathcal{M}} \nabla W(x-z)d\mu(z) + \nabla V(x), \xi(x) \right\rangle d\mu(x) \\
=&\int_{\mathcal{M}} \left\langle A(x)\left(\int_{\mathcal{M}} \nabla W(x-z)d\mu(z) + \nabla V(x)\right), G(x)\xi(x) \right\rangle d\mu(x) \\
=&\int_{\mathcal{M}} g_x\left(A(x)\left(\int_{\mathcal{M}} \nabla W(x-y)d\mu(y) + \nabla V(x)\right), \xi(x)\right) d\mu(x).
\end{aligned}
$$

For $n \in \mathbb{N}$, fix $x_0 \in \mathcal{M}$ define $B(n) = \{x \in \mathcal{M} : \mathrm{dist}(x, x_0) < n\}$ and $\mathcal{M}_{\frac{1}{n}} = \{x \in \mathcal{M} : \mathrm{dist}(x, \partial\mathcal{M}) \geq \frac{1}{n}\}$. For $x \in \partial\mathcal{M}$, denote $n(x)$ the unit outward normal with respect to the Riemannian metric g. Then define

$$
\xi_n(x) = \begin{cases} -\kappa(x) & \text{if } x \in \{x : g_x\left(\kappa(x), \kappa(x)\right) \leq n^2\} \bigcap \overline{B(n) \bigcap \mathcal{M}_{\frac{1}{n}}}, \\ -\kappa(x) - \frac{1}{n}n(x) & \text{if } x \in \{x : g_x\left(\kappa(x), \kappa(x)\right) \leq n^2\} \bigcap \overline{B(n) \bigcap \partial\mathcal{M}}, \\ 0 & \text{Otherwise.} \end{cases}
$$

Note that κ is continuous on $\{x : g_x\left(\kappa(x), \kappa(x)\right) \leq n^2\} \bigcap \overline{B(n) \bigcap \mathcal{M}_{\frac{1}{n}}}$, thus there exists $t_1 > 0$ such that $\exp_x\left(t\xi_n(x)\right) \in \mathcal{M}$ for all $x \in \{x : g_x\left(\kappa(x), \kappa(x)\right) \leq n^2\} \bigcap \overline{B(n) \bigcap \mathcal{M}_{\frac{1}{n}}}$ and $0 \leq t \leq t_1$. Since $\{x : g_x\left(\kappa(x), \kappa(x)\right) \leq n^2\} \bigcap \overline{B(n) \bigcap \partial\mathcal{M}}$ is compact and $g_x\left(\xi_n(x), n(x)\right) \leq -\frac{1}{n}$, so there exists $t_2 > 0$ such that $\exp_x\left(t\xi_n(x)\right) \in \mathcal{M}$ for all $x \in \{x : g_x\left(\kappa(x), \kappa(x)\right) \leq n^2\} \bigcap \overline{B(n) \bigcap \partial\mathcal{M}}$. Take $t_0 = \min\{t_1, t_2\}$, then for ξ_n, we have $g_x\left(\xi_n(x), \xi_n(x)\right) \leq n^2$ and $\exp_x\left(t\xi_n(x)\right) \in \mathcal{M}$ for all $x \in \mathcal{M}$ and $0 \leq t \leq t_0$. It is direct to check that $\xi_n \in L^2(g, \mu)$ and $\xi_n \to -\kappa$ in $L^2(g, \mu)$ as $n \to \infty$. Recall that from Lemma 3.3 of [119],

$$
\limsup_{t\to 0^+} \frac{d_W\left(\left(\exp(t\xi_n)\right)_\sharp \mu, \mu\right)}{t} \leq \|\xi_n\|_{L^2(g,\mu)}, \tag{4.1.12}
$$

where $\exp(t\xi_n)(x) = \exp_x(\xi_n(x))$. We get

$$\begin{aligned}
|\partial\mathcal{E}|(\mu)\|\xi_n\|_{L^2(g,\mu)} &\geq |\partial\mathcal{E}|(\mu)\liminf_{t\to 0^+}\frac{d_W\left(\exp(t\xi_n)_\sharp\mu,\mu\right)}{t}\\
&\geq -\int_{\mathcal{M}}\left\langle A(x)\left(\int_{\mathcal{M}}\nabla W(x-z)d\mu(z)+\nabla V(x)\right), G(x)\xi_n(x)\right\rangle d\mu(x)\\
&= -\int_{\mathcal{M}} g_x\left(A(x)\left(\int_{\mathcal{M}}\nabla W(x-z)d\mu(z)+\nabla V(x)\right), \xi_n(x)\right) d\mu(x).
\end{aligned}$$

Taking $n\to\infty$ and noting that $g_x(\xi, P\xi) = g_x(P\xi, P\xi)$ then yields

$$|\partial\mathcal{E}|(\mu)\|\kappa\|_{L^2(g,\mu)} \geq \int_{\mathcal{M}} g_x(\kappa(x),\kappa(x))\,d\mu(x).$$

Thus

$$\|\kappa\|_{L^2(g,\mu)} \leq |\partial\mathcal{E}|(\mu).$$

□

We now show the lower semi-continuity of κ with respect to the narrow convergence. Denote

$$\kappa^n(x) = -P_x\left(-A(x)\left(\int_{\mathcal{M}}\nabla W(x-y)d\tilde{\mu}^n(y)+\nabla V(x)\right)\right), \tag{4.1.13}$$

Theorem 4.1.8. *Given that (LNL1)-(LNL5) hold, for a.e. $t>0$,*

$$\liminf_{n\to\infty}\int_{\mathcal{M}}|\kappa^n(t,x)|^2\,d\tilde{\mu}^n(t,x) \geq \int_{\mathcal{M}}|\kappa(t,x)|^2\,d\mu(t,x).$$

For the proof of the theorem, refer to Theorem 3.4.3 and Proposition 3.4.5 in Chapter 3. We now give the main result of this Section regarding the existence of curves of maximal slope with respect to $\|\kappa\|_{L^2(g,\mu)}$.

Theorem 4.1.9. *Suppose $(\mathcal{M},g)$ satisfies (M1)-(M2) and W, V satisfy (LNL1)-(LNL5). Then the limit curve $\mu(\,\cdot\,)\in AC_{loc}([0,\infty);\mathcal{P}_2(\mathcal{M}))$ from JKO scheme satisfies for all $T\geq 0$*

$$\mathcal{E}(\mu_0) \geq \frac{1}{2}\int_0^T|\mu'|^2(t)dt + \frac{1}{2}\int_0^T\|\kappa(t)\|^2_{L^2(g,\mu(t))}dt + \mathcal{E}(\mu(T)). \tag{4.1.14}$$

Proof. Since $\operatorname{supp}(\tilde{\mu}^n(t))\subset B(r(t))\Subset\mathcal{M}$ for n big enough, we have

$$\|\kappa^n(t)\|_{L^2(g,\tilde{\mu}^n(t))} \leq |\partial\mathcal{E}|(\tilde{\mu}^n(t)). \tag{4.1.15}$$

Also straightforward calculation gives

$$\int_0^T\left|\mu'\right|^2(t)dt \leq \liminf_{n\to\infty}\int_0^T\left|(\tilde{\mu}^n)'\right|^2dt. \tag{4.1.16}$$

Recall that from the proof of Theorem 2.3.3 in [5] we have

$$\begin{aligned}
&\frac{1}{2}\int_0^T |\mu'|^2(t)dt + \frac{1}{2}\int_0^T \|\kappa(t)\|^2_{L^2(g,\mu(t))}dt + \mathcal{E}\left(\mu(T)\right) \\
&\leq \frac{1}{2}\int_0^T |\mu'|^2(t)dt + \frac{1}{2}\int_0^T \liminf_{n\to\infty} \|\kappa^n(t)\|^2_{L^2(g,\tilde{\mu}(t))}dt + \mathcal{E}\left(\mu(T)\right) \\
&\leq \liminf_{n\to\infty} \frac{1}{2}\int_0^T |\left(\tilde{\mu}^n\right)'|^2(t)dt + \frac{1}{2}\int_0^T \liminf_{n\to\infty} |\partial\mathcal{E}|^2\left(\tilde{\mu}(t)\right)dt + \liminf_{n\to\infty}\mathcal{E}\left(\tilde{\mu}^n(T)\right) \\
&\leq \liminf_{n\to\infty}\left(\frac{1}{2}\int_0^T \left(\tilde{\mu}^n\right)'|^2(t)dt + \frac{1}{2}\int_0^T G^2_{\tau_n}(t)dt + \mathcal{E}\left(\tilde{\mu}^n(T)\right)\right) \\
&\leq \mathcal{E}\left(\mu_0\right),
\end{aligned}$$

where we use the lower semi-continuity of κ and G_{τ_n} is defined as in (4.1.2). □

4.2 Existence and stability of weak measure solutions

In this Section, we show that the limit curve $\mu(\,\cdot\,)$ is a solution to continuity equation (3.1.5) in the sense of distributions (i.e. a weak measure solution). We then prove the stability properties of weak measure solutions to (3.1.5).

Recall from (3.3.1) of Chapter 3 that, for $x, y \in \mathcal{M}$, $T(x,y) \in T_x\mathcal{M}$ is defined such that the inverse exponential map of x evaluated at y. We also recall from Lemma 3.5.1 that for every $\mu(\,\cdot\,) \in AC_{loc}(\mathcal{P}_2(\mathcal{M}), d_W)$, $\int_{\mathcal{M}} \frac{T(x,y)}{h} d\nu_x^h(y)$ converges weakly in $L^2\left(g, \mu(t)\right)$ to the tangent velocity field $v(t,x)$ for a.e. $t > 0$ such that $\mu(\,\cdot\,)$ satisfies the continuity equation

$$\frac{\partial}{\partial t}\mu(t,x) + \operatorname{div}\left(\mu(t,x)v(t,x)\right) = 0 \tag{4.2.1}$$

in the sense of distributions and

$$\int_{\mathcal{M}} g\left(v(t,x), v(t,x)\right) d\mu(t,x) = \left|\mu'\right|^2(t) \tag{4.2.2}$$

for a.e. $t > 0$.

Now we can show the proof of Theorem 4.0.3.

Proof of Theorem 4.0.3. We only need to prove the following chain rule

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right) = \int_{\mathcal{M}} g_x\left(\kappa(t,x), v(t,x)\right) d\mu(t,x), \tag{4.2.3}$$

for a.e. $t > 0$, where v is the tangent velocity field for the absolutely continuous curve $\mu(\,\cdot\,)$. Indeed, the fact that $\mu(\,\cdot\,)$ satisfies (4.1.14) implies

$$\frac{d}{dt}\mathcal{E}(\mu(t)) \leq -\frac{1}{2}\int_{\mathcal{M}} g_x\left(v(t,x), v(t,x)\right) d\mu(t,x) - \frac{1}{2}\int_{\mathcal{M}} g_x\left(\kappa(t,x), \kappa(t,x)\right) d\mu(t,x). \tag{4.2.4}$$

If (4.2.3) holds, then together with (4.2.4), we have $v(t,x) = -\kappa(t,x)$ for a.e. $t > 0$ and $\mu(\cdot)$ is a weak measure solution to (3.1.5) with $|\mu'|^2(t) = \int_{\mathcal{M}} g_x(\kappa(t,x), \kappa(t,x))\, d\mu(t,x)$.

We now prove the chain rule (4.2.3). We first claim that

$$\begin{aligned}\mathcal{E}(\mu(t+h)) \geq \mathcal{E}(\mu(t)) + &\inf_{\gamma \in \Gamma_o(\mu(t),\mu(t+h))} \int_{\mathcal{M}\times\mathcal{M}} g_x(\kappa(t,x), T(x,y))\, d\gamma_t^h(x,y) \\ &+ o\left(d_W(\mu(t), \mu(t+h))\right).\end{aligned} \tag{4.2.5}$$

For general $\mu, \nu \in \mathcal{P}_2(\mathcal{M})$ with $\mathrm{supp}(\mu) \cup \mathrm{supp}(\nu) \subset B(r(T))$, let $k \in \mathbb{N}$ be such that $B(2r(T)) \subset K_k$ with W is λ_k-geodesically convex on $B(r(T)) \times B(r(T))$ and V λ_k-geodesically convex on $B(r(T))$. Denote $\lambda = \lambda_k$ and $\gamma \in \Gamma_o(\mu,\nu)$ an optimal plan between μ and ν, we then know that the function

$$\begin{aligned}f(t) =& \frac{W\left(\exp_{x_1}(tT(x_1,y_1)) - \exp_{x_2}(tT(x_2,y_2))\right) - W(x_1 - x_2)}{2t} \\ &+ \frac{2V\left(\exp_{x_2}(tT(x_2,y_2))\right) - 2V(x_2)}{2t} - \frac{\lambda}{2} t\,\mathrm{dist}^2(x_2,y_2) - \frac{\lambda}{2} t\,\mathrm{dist}^2\left((x_1,x_2),(y_1,y_2)\right)\end{aligned} \tag{4.2.6}$$

is non-decreasing on $[0,1]$ for all $(x_1,y_1), (x_2,y_2) \in \mathrm{supp}(\gamma)$. So $f(1) \geq \liminf_{t\to 0^+} f(t)$. Integrating over $d\gamma(x_1,y_1)d\gamma(x_2,y_2)$ gives

$$\begin{aligned}&\mathcal{E}(\nu) - \mathcal{E}(\mu) \\ =& \int_{\mathcal{M}\times\mathcal{M}} \int_{\mathcal{M}\times\mathcal{M}} \frac{W(y_1-y_2) + 2V(y_2) - W(x_1-x_2) - 2V(x_2)}{2} d\gamma(x_1,y_1) d\gamma(x_2,y_2) \\ \geq& \int_{\mathcal{M}\times\mathcal{M}} \int_{\mathcal{M}\times\mathcal{M}} \langle \nabla W(x_2 - x_1) + \nabla V(x_2), T(x_2,y_2)\rangle\, d\gamma(x_1,y_1) d\gamma(x_2,y_2) + o\left(d_W(\mu,\nu)\right) \\ =& \int_{\mathcal{M}\times\mathcal{M}} \left\langle \int_{\mathcal{M}} \nabla W(x_2 - x_1) d\mu(x_1) + \nabla V(x_2), T(x_2,y_2) \right\rangle d\gamma(x_2,y_2) + o\left(d_W(\mu,\nu)\right) \\ =& \int_{\mathcal{M}\times\mathcal{M}} g_{x_2}\left(A(x_2)\left(\int_{\mathcal{M}} \nabla W(x_2-x_1) d\mu(x_1) + \nabla V(x_2)\right), T(x_2,y_2)\right) d\gamma(x_2,y_2) \\ &+ o\left(d_W(\mu,\nu)\right) \\ \geq& -\int_{\mathcal{M}\times\mathcal{M}} g_{x_2}\left(P_{x_2}\left(-A(x_2)\left(\nabla W * \mu(x_2) + \nabla V(x_2)\right)\right), T(x_2,y_2)\right) d\gamma(x_2,y_2) + o\left(d_W(\mu,\nu)\right) \\ =& \int_{\mathcal{M}\times\mathcal{M}} g_{x_2}\left(\kappa(x_2), T(x_2,y_2)\right) d\gamma(x_2,y_2) + o\left(d_W(\mu,\nu)\right)\end{aligned}$$

where the second inequality comes from the fact that: If $x_2 \notin \partial\mathcal{M}$, then by definition of P_{x_2} the inequality becomes an equality while if $x_2 \in \partial\mathcal{M}$, then by definition of P_{x_2}. Now note that for $T > 0$ such that $0 \leq t \leq T$ and $0 \leq t+h \leq T$, we have $\mathrm{supp}(\mu(t)) \cup \mathrm{supp}(\mu(t+h)) \subset B(r(T))$, thus by taking $\mu = \mu(t), \nu = \mu(t+h)$ we get (4.2.5). The claim is proved.

By (4.2.5) and Lemma 3.5.1, we have

$$\begin{aligned}\lim_{h\to 0^+}\frac{\mathcal{E}\left(\mu(t+h)\right)-\mathcal{E}\left(\mu(t)\right)}{h} &\geq \int_{\mathcal{M}\times\mathcal{M}} g_x\left(\kappa(t,x),\frac{T(x,y)}{h}\right)d\gamma_t^h(x,y)\\ &=\int_{\mathcal{M}} g_x\left(\kappa(t,x),\int_{\mathcal{M}}\frac{T(x,y)}{h}d\nu_x^h(y)\right)d\mu(t,x)\\ &=\int_{\mathcal{M}} g_x\left(\kappa(t,x),v(t,x)\right)d\mu(t,x).\end{aligned}$$

Similarly

$$\lim_{h\to 0^-}\frac{\mathcal{E}(\mu(t+h))-\mathcal{E}(\mu(t))}{h}\leq\int_{\mathcal{M}} g_x\left(\kappa(t,x),v(t,x)\right)d\mu(t,x).$$

Since the function $t\mapsto\mathcal{E}(\mu(t))$ is non-increasing, it is differentiable for a.e. $t>0$, so

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right)=\int_{\mathcal{M}} g_x\left(\kappa(t,x),v(t,x)\right)d\mu(t,x)$$

for a.e. $t>0$ as desired.

To prove the energy dissipation equality (4.0.3), we only need to show that $\mathcal{E}(\mu(t))$ is locally absolutely continuous. By Θ-convexity and linear growth condition on gradient of W,V, we know $|V(x)-V(y)|\leq C(1+\operatorname{dist}(x,y))|x-y|\leq C(1+\operatorname{dist}(x,y))\operatorname{dist}(x,y)$ and $|W(x-z)-W(y-w)|\leq C(1+\operatorname{dist}(x,y)+\operatorname{dist}(z,w))(\operatorname{dist}(x,y)+\operatorname{dist}(z,w))$, then for $0\leq s<t<\infty$ and $\gamma\in\Gamma_o(\mu(t),\mu(s))$ an optimal plan

$$\begin{aligned}|\mathcal{E}(\mu(t))-\mathcal{E}(\mu(s))| &\leq\int_{\mathcal{M}\times\mathcal{M}} C(1+\operatorname{dist}(x,y))\operatorname{dist}(x,y)d\gamma(x,y)\\ &\leq C\left(1+d_W\left(\mu(t),\mu(s)\right)\right)d_W\left(\mu(t),\mu(s)\right).\end{aligned}$$

Thus $\mathcal{E}(\mu(\,\cdot\,))$ is locally absolutely continuous since $\mu(\,\cdot\,)$ is locally absolutely continuous in $(\mathcal{P}_2(\mathcal{M}),d_W)$.

We now turn to the contraction property, denote $\gamma(t)\in\Gamma_o\left(\mu^1(t),\mu^2(t)\right)$ an optimal plan. Since $\operatorname{supp}\left(\mu^1(t)\right)\cup\operatorname{supp}\left(\mu^2(t)\right)\subset B(r(t))\Subset\mathcal{M}$, by (4.2.6) we have

$$\mathcal{E}\left(\mu^2(t)\right)\geq\mathcal{E}\left(\mu^1(t)\right)+\int_{\mathcal{M}\times\mathcal{M}} g_x\left(\kappa^1(t,x),T(x,y)\right)d\gamma_t(x,y)+\frac{\lambda_k}{2}d_W^2\left(\mu^1(t),\mu^2(t)\right),\tag{4.2.7}$$

and

$$\mathcal{E}\left(\mu^1(t)\right)\geq\mathcal{E}\left(\mu^2(t)\right)+\int_{\mathcal{M}\times\mathcal{M}} g_y\left(\kappa^2(t,y),T(y,x)\right)d\gamma_t(y,x)+\frac{\lambda_k}{2}d_W^2\left(\mu^1(t),\mu^2(t)\right),\tag{4.2.8}$$

for all k such that $B(2r(t))\subset K_k$ and λ_k the geodesic convexity constant of W,V on $K_k\times K_k$ and K_k. Adding them together gives

$$-\lambda_k d_W^2\left(\mu^1(t),\mu^2(t)\right)\geq\int_{\mathcal{M}\times\mathcal{M}}\left(g_x\left(\kappa^1(t,x),T(x,y)\right)+g_y\left(\kappa^2(t,y),T(y,x)\right)\right)d\gamma_t(x,y).\tag{4.2.9}$$

Now since $\mu^1(\,\cdot\,), \mu^2(\,\cdot\,)$ are solutions for (3.1.5), we have $\kappa^i(t) = -v^i(t)$ for $i = 1, 2$. By Lemma 5.3 from [119] and Lemma 4.34 from [5] we then have

$$\begin{aligned}
\frac{d}{dt} d_W^2 \left(\mu^1(t), \mu^2(t)\right) &\leq -2 \int_{\mathcal{M}\times\mathcal{M}} \left(g_x\left(v^1(t,x), T(x,y)\right) + g_y\left(v^2(t,y), T(y,x)\right)\right) d\gamma_t(x,y) \\
&= 2 \int_{\mathcal{M}\times\mathcal{M}} \left(g_x\left(\kappa^1(t,x), T(x,y)\right) + g_y\left(\kappa^2(t,y), T(y,x)\right)\right) d\gamma_t(x,y) \\
&\leq -2\lambda_k d_W^2 \left(\mu^1(t), \mu^2(t)\right).
\end{aligned}$$

Thus

$$d_W\left(\mu^1(t), \mu^2(t)\right) \leq \exp(-\lambda_k t) d_W\left(\mu_0^1, \mu_0^2\right). \tag{4.2.10}$$

□

Chapter 5

Nonlocal interaction equations on non-convex, non-smooth domains

In this Chapter, we discuss well-posedness of a class of nonlocal interaction equations on general non-convex, non-smooth domains. We want to model macroscopic behavior of biological agents interacting in geometrically confined domains Ω with irregular boundaries. The domain boundary may be an environmental obstacle, like a river, or the ground itself, as in the models of locust patterns discussed in [13, 107, 110]. In Chapter 3 systems of interacting agents on domains with boundary are considered in a setting which allows for heterogeneous environments via gradient flow theory developed in [5], but requires the domain (manifold $\mathcal{M}$) to be (geodesically-)convex with C^1 boundary. Here we consider the problem in homogeneous environments (Euclidean setting) but on general domains which are not required to be convex and whose boundary may not be differentiable. Again, the geometrical confinement introduces a constraint on the possible velocity fields of the agents at the boundary. The measure $\mu(\,\cdot\,)$ describing the agent configuration becomes a distributional solution of the equation

$$\begin{cases} \dfrac{\partial}{\partial t}\mu(t,x) + \operatorname{div}\left(\mu(t,x)P_x\left(-\displaystyle\int_\Omega \nabla W(x-y)d\mu(t,y) - \nabla V(x)\right)\right) = 0, \\ \mu(0) = \mu_0, \end{cases} \tag{5.0.1}$$

where P_x is the projection of the velocities to inward pointing ones. Since the domain Ω is non-convex, the space $\mathcal{P}_2(\Omega)$ of probability measures with finite second moments on Ω is not geodesically convex, thus the energy functional $\mathcal{E}$ is not geodesically (semi-)convex and general existence of gradient flow theory [5] (which we introduce in Chapter 2, and use in Chpater 3 and Chapter 4) fails to apply. We instead obtain gradient flows and weak measure solutions via particle approximations. The results presented in this Chapter are obtained in our paper [33].

5.1 Domain regularity and constrained gradient flow structure

When considering domains which are not C^1 the first question is what should the velocity of agents be at a boundary point where the domain is not differentiable. Similar questions have been encountered in studies of differential inclusions on moving domains (general sweeping processes), see [43, 44, 112] and references therein. We rely on notions developed there to properly define the cone of admissible directions at a boundary points and the proper way to project the velocity to the allowable cone. In particular we consider the equation (5.0.1) with projection P_x defined in (5.1.5) and when there is no confusion we denote $P_x = P_{T(\Omega,x)}$.

While one would like to consider very general domains there are limits to possible domains on which well-posedness of weak measure solutions can be developed. Namely, if the domains have an inside corner, then it is not possible for the measure solutions of (5.0.1) to be stable, as we discuss in Remark 5.1.11. It turns out that a class of domains which is rather general and allows for a well-posedness theory are the (uniformly) prox-regular domains (see Definition 5.1.3). Prox-regular domains are the sets which have an outside neighborhood such that for each of its points there exists a unique closest point on the boundary (unique projection). In particular prox-regular domains can have outside corners and outside cusps, but not inside corners.

Our main result is the well-posedness of weak measure solutions of the nonlocal-interaction equation (5.0.1) on uniformly prox-regular domains. We recall from (2.3.6), (2.3.7) and (2.3.8) in Chapter 2 the total energy $\mathcal{E}$ is defined as

$$\mathcal{E}(\mu) = \mathcal{W}(\mu) + \mathcal{V}(\mu) = \frac{1}{2}\int_\Omega \int_\Omega W(x-y)d\mu(x)d\mu(y) + \int_\Omega V(x)d\mu(x). \tag{5.1.1}$$

The energy $\mathcal{E}$ is a dissipated quantity of the evolution (5.0.1), and furthermore the equation can be interpreted as the gradient flow of the energy with respect to the Euclidean Wasserstein metric with the constraint that mass cannot leave the domain Ω. The gradient flow approach was used to study systems in which there are state constraints that determine the set of possible velocities, in particular in crowd motion models [1, 80, 81] where the constraint on the L^∞-norm of the density of agents leading to an L^2-projection of velocity field.

To show the existence of gradient flows, we use particle approximations, that is we use a sequence of delta masses $\mu_0^n = \sum_{j=1}^{k(n)} m_j \delta_{x_j^n}$ to approximate the initial data μ_0 and solve (5.0.1) with initial data μ_0^n. Here the notion of gradient flow solutions (and weak measure solutions) provides the advantage that we can work with delta measures, which makes the particle approximation meaningful. With discrete initial data μ_0^n, (5.0.1) becomes a system of ordinary differential equations. We solve the ODE system and prove that the solutions

$\mu^n(\cdot)$ converges to some $\mu(\cdot)$ by establishing the stability property of solutions to (5.0.1) with different initial data. We then show that the limit curve $\mu(\cdot)$ is a gradient flow with respect to $\mathcal{E}$ swith initial data μ_0 by proving that $\mu(\cdot)$ achieves the maximal dissipation of the associated energy, and is thus the steepest descent of the energy.

The novelty here is that even though the domain Ω is only prox-regular (not necessarily convex or C^1) and the velocity field is discontinuous (due to the projection P), the ODE systems are still well-posed (refer to Theorem 5.2.6) and the stability of solutions $\mu^n(\cdot)$ in Wasserstein metric d_W is valid with explicit dependence on the prox-regularity constant (refer to Proposition 5.3.1). Under semi-convexity assumptions on the potential functions W and V, this enables us to show the well-posedness, that is existence and stability of weak measure solutions to (5.0.1) in three different cases: Ω bounded and prox-regular (Theorem 5.1.5 and Thorem 5.1.6), Ω unbounded and convex (Theorem 5.1.9), and Ω unbounded and prox-regular with compactly supported initial data μ_0 (Theorem 5.1.10). We can also generalize the well-posedness results to time-dependent interaction and external potentials $W = W(t,x), V = V(t,x)$ (Remark 5.5.3). We also give sufficient conditions on the shape of Ω to ensure the existence of an interaction potentials W such that solutions $\mu(\cdot)$ to (5.0.1) aggregate to a single delta mass as time goes to infinity (Theorem 5.1.11 and Remark 5.6.1).

Description of weak measure solutions.

Since we are working in the Euclidean setting, we denote in this Chapter

$$d_W^2(\mu,\nu) = \min\left\{\int_{\Omega\times\Omega} |x-y|^2 d\gamma(x,y) : \gamma \in \Gamma(\mu,\nu)\right\} \tag{5.1.2}$$

the Euclidean Wasserstein distance and $\Gamma_o(\mu,\nu)$ the set of optimal plans with respect to this Euclidean d_W, i.e.

$$\Gamma_o(\mu,\nu) = \left\{\gamma \in \Gamma(\mu,\nu) : \int_{\Omega\times\Omega} |x-y|^2 d\gamma(x,y) = d_W^2(\mu,\nu)\right\}. \tag{5.1.3}$$

We now recall definition of weak measure solutions to the continuity equation to (5.0.1). To be precise, in our Euclidean setting,

Definition 5.1.1. *A locally absolutely continuous curve $\mu(\cdot) \in \mathcal{P}_2(\Omega)$ is a* weak measure solution *to (5.0.1) with initial value μ_0 if*

$$v(t,x) = -P_x\left(\int_\Omega \nabla W(x-y)d\mu(y) + \nabla V(x)\right) \in L^1_{loc}([0,+\infty); L^2(\mu(t)))$$

and

$$\int_0^\infty\int_\Omega \frac{\partial\phi}{\partial t}(t,x)d\mu(t,x)dt + \int_\Omega \phi(0,x)d\mu_0(x) + \int_0^\infty\int_\Omega \langle\nabla\phi(t,x), v(t,x)\rangle\, d\mu(x) = 0\,,$$

for all $\phi \in C_c^\infty([0,\infty)\times\Omega)$*. The projection* P_x *is described below and formally defined in* (5.1.5) *with* $P_x = P_{T(\Omega,x)}$.

Note that the test function ϕ does not have to be zero on the boundary of Ω, and thus the no-flux boundary condition is imposed in a weak form.

We now define the projection P_x. When $\partial\Omega \in C^1$ is smooth and oriented, the definition of P_x is given in Chapter 3 by $P_x(v) = v - \langle v, \nu(x)\rangle\nu(x)$ if $\langle v, \nu(x)\rangle > 0$ and $P_x(v) = v$ otherwise, where $\nu(x)$ is the unit outward normal vector to the boundary at $x \in \partial\Omega$. When Ω is only prox-regular, to define P_x, we need to recall some notations from non-smooth analysis, see [25, 35], in order to replace the normal vector field, and the inward and outward directions.

Definition 5.1.2. *Let* S *be a closed subset of* $\mathbb{R}^d$*. We define the* proximal normal cone *to* S *at* x *by,*

$$N^P(S,x) = \left\{ v \in \mathbb{R}^d \,:\, \exists \alpha > 0,\ x \in P_S\left(x + \alpha v\right) \right\},$$

where

$$P_S(y) = \left\{ z \in S \,:\, \inf_{w \in S} |w - y| = |z - y| \right\}$$

is the projection of y *onto* S.

Note that for $x \in S \setminus \partial S, N^P(S,x) = \{0\}$ and by convention for $x \notin S, N^P(S,x) = \emptyset$. The notion of normal cone extends the concept of outer normal of a smooth set in the sense that if S is a closed subset of $\mathbb{R}^d$ with boundary ∂S an oriented C^2 hypersurface, then for each $x \in \partial S$, $N^P(S,x) = \mathbb{R}^+\nu(x)$ where $\nu(x)$ is the unit outward normal to S at x. We now recall the notion of uniform prox-regular sets.

Definition 5.1.3. *Let* S *be a closed subset of* $\mathbb{R}^d$*.* S *is said to be* η-prox-regular *if for all* $x \in \partial S$ *and* $v \in N^P(S,x), |v| = 1$ *we have*

$$B_\eta(x + \eta v) \cap S = \emptyset,$$

where $B_\eta(y)$ *denotes the open ball centered at* y *with radius* $\eta > 0$.

Note that an equivalent characterization, see [35, 96], is given by: S is η-prox-regular if for any $y \in S, x \in \partial S$ and $v \in N^P(S,x)$,

$$\langle v, y - x\rangle \le \frac{|v|}{2\eta}|y - x|^2. \tag{5.1.4}$$

Observe that if S is closed and convex, then S is ∞-prox-regular, thus η-prox-regularity is a relaxed condition on convexity. We now turn to the tangent cones.

Definition 5.1.4. *Let S be a closed subset of $\mathbb{R}^d$ and $x \in S$, define the* Clarke tangent cone *by*

$$T^C(S,x) = \Big\{ v \in \mathbb{R}^d \,:\, \forall t_n \searrow 0, \forall x_n \in S,\ s.t.\ x_n \to x,\ \exists v_n \to v\ s.t.\ (\forall n)\, x_n + t_n v_n \in S \Big\}\ ,$$

and denote the Clarke normal cone *by*

$$N^C(S,x) = \left\{ \xi \in \mathbb{R}^n : \langle \xi, v \rangle \leq 0 \ \ \forall v \in T^C(S,x) \right\}.$$

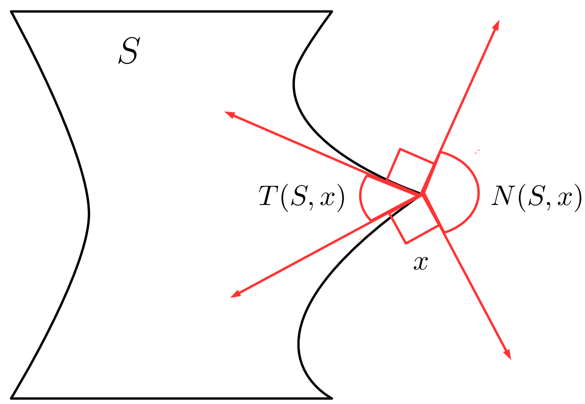


Figure 5.1: The set S is prox-regular but not convex. At the corner point $x \in \partial S$, the tangent and normal cones are denoted by $T(S,x)$ and $N(S,x)$.

Note that $T^C(S,x), N^C(S,x)$ are closed convex cones, also by convention $N^C(S,x) = \emptyset$ for all $x \notin S$. In general, we only have $N^P(S,x) \subset N^C(S,x)$ and the inclusion can be strict. However, for η-prox-regular set S, we have $N^P(S,x) = N^C(S,x)$, see [35, 96]. In that case, we denote the normal cone and tangent cone as $N(S,x) = N^P(S,x) = N^C(S,x)$ and $T(S,x) = T^C(S,x) = T^P(S,x)$ respectively, and for any vector $w \in \mathbb{R}^d$, we define the projection onto the tangent cone by $P_{T(S,x)}(w)$, i.e.,

$$P_{T(S,x)}(w) = \left\{ v \in T(S,x) : |v - w| = \inf_{\xi \in T(S,x)} |\xi - w| \right\}. \tag{5.1.5}$$

Since $T(S,x)$ is a closed convex cone, the infimum is always attained, and $P_{T(S,x)}$ is well-defined. For notation simplicity, since the set we are considering Ω is not changing, we write P_x instead of $P_{T(\Omega,x)}$ and when the context is clear, we put P for P_x. With these preliminaries, we can now state the main results of this work.

Main results.

For any set $A \subset \mathbb{R}^d$, we denote by $A-A = \{x-y : x, y \in A\}$, and the convex hull of A by $\mathrm{Conv}\,(A) = \{\theta x + (1-\theta)y : x, y \in A, 0 \le \theta \le 1\}$. For a function $f \in C^1(\mathbb{R}^d)$, we say that f is λ-geodesically convex on a convex set S if for any $x, y \in S$ we have

$$f(y) \ge f(x) + \langle \nabla f(x), y-x \rangle + \frac{\lambda}{2}|y-x|^2.$$

We call f *locally λ-geodesically convex* if there exist a sequence of compact convex sets $K_n \subset \mathbb{R}^d$ and a sequence of constants λ_n such that $K_n \subset K_{n+1}$, $\bigcup_n K_n = \mathbb{R}^d$ and f is λ_n-geodesically convex on K_n. Note that f is λ-geodesically convex on a convex set S implies for any $x, y \in S$

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \ge \lambda |x-y|^2.$$

The main assumptions depend on the domain Ω and the support of initial data. In fact, we separate our results in three cases: Ω bounded, Ω unbounded and convex, and Ω unbounded with compactly supported initial data. The assumptions are very similar in nature based on the convexity of the potentials V and W and on their growth behavior at ∞ in the unbounded cases. We assume that both potentials V and W are λ_V- and λ_W-convex respectively, possibly locally convex. Finally, in case V and W are λ-locally convex, we can assume, without loss of generality, that V and W share the same sequence of compact convex sets, K_k in the definition of locally λ-geodesic convexity, i.e., $K_k \subset K_{k+1}$, $\bigcup_{k\in\mathbb{N}} K_k = \mathbb{R}^d$ with V and W being $\lambda_{V,k}$ and $\lambda_{W,k}$-geodesically convex on K_k.

Recall the elements of the theory of gradient flows in the space of probability measures in the Euclidean setting (introduced in Chapter 2). In particular the Definition of metric derivative (2.2.8), subdifferenial (Definition 2.2.5), and gradient flow (2.2.14). By Theorem 2.2.3, given a locally absolutely continuous curve $[0,\infty) \ni t \mapsto \mu(t) \in \mathcal{P}_2(\Omega)$, there exists a unique tangent velocity field such that $\mu(\,\cdot\,)$ satisfies the continuity equation in the sense of distributions.

In case Ω is bounded, we assume that

(M1) $\Omega \subset \mathbb{R}^d$ is η-prox-regular with $\eta > 0$.

(A1) $W \in C^1(\mathbb{R}^d)$ is λ_W-geodesically convex on $\mathrm{Conv}\,(\Omega-\Omega)$ for some $\lambda_W \in \mathbb{R}$.

(A2) $V \in C^1(\mathbb{R}^d)$ is λ_V-geodesically convex on $\mathrm{Conv}\,(\Omega)$ for some $\lambda_V \in \mathbb{R}$.

The main results of this paper is the well-posedness of weak measure solutions: existence and stability, with arbitrary initial data. We establish it using an approximation scheme and the theory of gradient flows in spaces of probability measures.

Theorem 5.1.5. *Assume Ω is bounded and satisfies (M1) and W, V satisfy (A1), (A2). Then for any initial data $\mu_0 \in \mathcal{P}_2(\Omega)$, there exists a locally absolutely continuous curve $\mu(\cdot) \in \mathcal{P}_2(\Omega)$ such that $\mu(\cdot)$ is a gradient flow with respect to $\mathcal{E}$ and a weak measure solution to* (5.0.1)*.*

Furthermore for a.e. $t > 0$

$$|\mu'|^2(t) = \int_\Omega |P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)|^2 \, d\mu(t,x), \tag{5.1.6}$$

and for any $0 \le s \le t < \infty$

$$\mathcal{E}(\mu(s)) = \mathcal{E}(\mu(t)) + \int_s^t \int_\Omega |P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)|^2 \, d\mu(r,x)dr. \tag{5.1.7}$$

Theorem 5.1.6. *Assume Ω is bounded and satisfies (M1) and W, V satisfy (A1), (A2). Let $\mu^1(\cdot), \mu^2(\cdot)$ be two weak measure solutions to* (5.0.1) *with initial data μ_0^1, μ_0^2 respectively. Then*

$$d_W\left(\mu^1(t), \mu^2(t)\right) \le \exp\left(\left(-\lambda_W^- - \lambda_V + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{\eta}\right) t\right) d_W\left(\mu_0^1, \mu_0^2\right). \tag{5.1.8}$$

for any $t \ge 0$ where $\lambda_W^- = \min\{\lambda_W, 0\}$. Moreover, the weak measure solution is characterized by the system of Evolution Variational Inequalities:

$$\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu(t), \nu\right) + \left(\frac{\lambda_W^-}{2} + \frac{\lambda_V}{2} - \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{2\eta}\right) d_W^2\left(\mu(t), \nu\right) \le \mathcal{E}(\nu) - \mathcal{E}(\mu(t)), \tag{5.1.9}$$

for a.e. $t > 0$ and for all $\nu \in \mathcal{P}_2(\Omega)$.

Observe that in the stability estimate for solutions (5.1.8), we find two contributions due to the λ-convexity of the potentials and the η-prox-regular property of the domain Ω respectively. We also make a remark here that when Ω bounded, η-prox-reguar (Theorem 5.1.5, Theorem 5.1.6), or unbounded, convex (Theorem 5.1.9), weak measure solutions to (5.0.1) and gradient flows with respect to $\mathcal{E}$ are equivalent, see Remark 5.3.6.

On $\mathbb{R}^n$ when $\mu^1(0)$ and $\mu^2(0)$ have the same center of mass λ_W^- can be replaced by λ_W in (5.1.8). Thus when the potential W is uniformly geodesically convex, $\lambda_W > 0$ and thus there is exponential contraction of solutions. On bounded domains this is not the case since interaction with boundary can change the center of mass of a solution. Nevertheless part of the claim can be recovered. We consider the case that $V \equiv 0$. Denote the set of singletons by $\Xi = \{\delta_x : x \in \mathbb{R}^d\}$. Note that we included the singletons which are not in the set Ω, since the center of mass for measures on a non-convex Ω may lie outside the domain.

Proposition 5.1.7. *Assume Ω is bounded and satisfies (M1) and W satisfies (A1). Let $\mu(\,\cdot\,)$ be a weak measure solutions to* (5.0.1) *with $V \equiv 0$. Then*

$$d_W\left(\mu(t), \Xi\right) \leq \exp\left(\left(-\lambda_W + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)}}{\eta}\right) t\right) d_W\left(\mu_0, \Xi\right). \tag{5.1.10}$$

for any $t \geq 0$.

The proposition implies that solution can aggregate to a point (in perhaps infinite time) even on a nonconvex domain. We ask on what domains there exists a potential for which for any initial datum this aggregation property holds. We provide a sufficient condition on the shape of Ω for aggregation to hold: Let $\operatorname{diam}(\Omega) = \sup_{x,y\in\Omega} |x-y|$.

Theorem 5.1.8. *Assume that Ω is bounded and satisfies (M1). If $\eta > \frac{1}{2}\operatorname{diam}(\Omega)$, then for external potential $V \equiv 0$, there exists an interaction potential W satisfying (A1) for some $\lambda_W > 0$, and constant $C(\Omega) < 0$ such that*

$$d_W(\mu(t), \Xi) \leq d_W(\mu_0, \Xi) \exp\left(C\left(\Omega\right) t\right), \tag{5.1.11}$$

for all $t \geq 0$. In particular, the solution aggregates to a singleton:

$$\lim_{t\to\infty} d_W\left(\mu(t), \Xi\right) = \lim_{t\to\infty} d_W\left(\mu(t), \delta_{\bar{x}(t)}\right) = 0, \tag{5.1.12}$$

where $\bar{x}(t) = \int_\Omega x d\mu(t)$ is the center of mass for $\mu(t)$.

Note that the constant in $\eta > \frac{1}{2}\operatorname{diam}(\Omega)$ cannot be improved, as the example in Remark 5.6.1 shows.

Next we generalize the two existence and stability results to the unbounded domain Ω case in two different settings. In case Ω is unbounded, and for general initial data μ_0, possibly with noncompact support, we give the global assumptions: for some constants $\lambda_W, \lambda_V \in \mathbb{R}$ and $C > 0$,

(GM1) $\Omega \subset \mathbb{R}^d$ is convex, i.e., Ω is ∞-prox-regular.

(GA1) $W \in C^1(\mathbb{R}^d)$ is λ_W-geodesically convex on $\operatorname{Conv}(\Omega-\Omega) = \Omega-\Omega$.

(GA2) ∇W has linear growth, i.e., $|\nabla W(x)| \leq C(1+|x|)$ for all $x \in \mathbb{R}^d$.

(GA3) $V \in C^1(\mathbb{R}^d)$ is λ_V-geodesically convex on $\operatorname{Conv}(\Omega) = \Omega$.

(GA4) ∇V has linear growth, $|\nabla V(x)| \leq C(1+|x|)$ for all $x \in \mathbb{R}^d$.

The main result in this setting reads as:

Theorem 5.1.9. *Assume Ω is unbounded and satisfies (GM1) and W, V satisfy (GA1)-(GA4), then for any $\mu_0 \in \mathcal{P}_2(\Omega)$, there exists a gradient flow $\mu(\,\cdot\,)$ with respect to $\mathcal{E}$ such that $\mu(\,\cdot\,)$ is a weak measure solution to* (5.0.1)*. Moreover, for a.e. $t > 0$*

$$|\mu'|^2(t) = \int_\Omega |P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)|^2 \, d\mu(t,x),$$

and for any $0 \le s \le t < \infty$

$$\mathcal{E}(\mu(s)) = \mathcal{E}(\mu(t)) + \int_s^t \int_\Omega |P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)|^2 \, d\mu(r,x)dr.$$

Similarly, if $\mu^1(\,\cdot\,), \mu^2(\,\cdot\,)$ are two weak measure solutions to (5.0.1) with initial data μ_0^1, μ_0^2 respectively, then

$$d_W\left(\mu^1(t), \mu^2(t)\right) \le \exp\left(-\left(\lambda_W^- + \lambda_V\right)t\right) d_W\left(\mu_0^1, \mu_0^2\right). \tag{5.1.13}$$

for any $t \ge 0$. Also the weak measure solution is characterized by the system of Evolution Variational Inequalities:

$$\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu(t), \nu\right) + \left(\frac{\lambda_W^-}{2} + \frac{\lambda_V}{2}\right) d_W^2\left(\mu(t), \nu\right) \le \mathcal{E}(\nu) - \mathcal{E}(\mu(t)), \tag{5.1.14}$$

for a.e. $t > 0$ and for all $\nu \in \mathcal{P}_2(\Omega)$.

Since Ω is convex means Ω is ∞-prox-regular, the stability estimate (5.1.13) and EVI (5.1.14) in the convex setting are consistent with the estimates in the η-prox-regular setting by taking $\eta = \infty$ in (5.1.8) and (5.1.9).

The convexity assumption is needed since on nonconvex unbounded domains we do not know how to control the error due to lack of convexity (as measured by the prox-regularity (5.1.4)) in the stability of solutions. However, we can show that control assuming compactly supported initial data. Therefore, when Ω is unbounded and the initial data μ_0 has compact support, we assume there exist some constants $\eta > 0, \lambda_W, \lambda_V \in \mathbb{R}, C > 0$ such that the following local assumptions hold

(M1) $\Omega \subset \mathbb{R}^d$ is η-prox-regular.

(LA1) $W \in C^1(\mathbb{R}^d)$ is locally λ-geodesically convex on $\mathbb{R}^d$.

(LA2) ∇W has linear growth, i.e., $|\nabla W(x)| \le C(1 + |x|)$ for all $x \in \mathbb{R}^d$.

(LA3) $V \in C^1(\mathbb{R}^d)$ is locally λ-geodesically convex on $\mathbb{R}^d$.

(LA4) ∇V has linear growth, $|\nabla V(x)| \le C(1 + |x|)$ for all $x \in \mathbb{R}^d$.

Note that the conditions (LA1) and (LA3) are satisfied whenever V and W are C^2 functions on $\mathbb{R}^d$, which is the case in many practical applications.

We show in this setting the following theorem about existence and stability for weak measure solutions for initial data with compact support.

Theorem 5.1.10. *Given that Ω is unbounded and satisfies (M1), and W, V satisfy (LA1)-(LA4). If $\operatorname{supp}(\mu_0) \subset \Omega$ is compact, say $\operatorname{supp}(\mu_0) \subset B(r_0) \cap \Omega$, then there exists a weak measure solution $\mu(\,\cdot\,)$ to (5.0.1) such that $\operatorname{supp}(\mu(t)) \subset B(r(t))$ for $r(t) = (r_0+1)\exp(Ct)$, where $C = C(W,V)$ and $\mu(\,\cdot\,)$ satisfies for a.e. $t > 0$*

$$|\mu'|^2(t) = \int_\Omega |P_x(-\nabla W * \mu(r)(x) - \nabla V(x))|^2 \, d\mu(t,x),$$

and for any $0 \le s \le t < \infty$

$$\mathcal{E}(\mu(s)) = \mathcal{E}(\mu(t)) + \int_s^t \int_\Omega |P_x(-\nabla W * \mu(r)(x) - \nabla V(x))|^2 \, d\mu(r,x)dr.$$

Moreover if we have two such solutions $\mu^i(\,\cdot\,)$ with initial data μ_0^i satisfying for $i = 1,2$, $\operatorname{supp}(\mu_0^i)$ are compact and $\operatorname{supp}(\mu^i(t)) \subset B(r(t))$ for all $t > 0$, then for all $k \in \mathbb{N}$ such that $B(r(t)) \subset K_k$ we have

$$d_W\left(\mu^1(t), \mu^2(t)\right) \le \exp\left(\left(-\lambda_{W,k}^- - \lambda_{V,k} + \frac{\|\nabla W\|_{L^\infty(\Omega_k - \Omega_k)} + \|\nabla V\|_{L^\infty(\Omega_k)}}{\eta}\right) t\right) d_W(\mu_0^1, \mu_0^2). \tag{5.1.15}$$

where $\lambda_{W,k}, \lambda_{V,k}$ are the geodesic convexity constants of W and V in K_k and $\Omega_k = \Omega \cap K_k$.

Let us point out that we are not able to get the system of Evolution Variational Inequalities in its whole generality although they hold for compactly supported reference measures.

Remark 5.1.11. *Here we illustrate on an example that well-posedness of weak measure solutions cannot hold on domains which have an inside corner. Let $\Omega = \{(r\cos(\theta), r\sin(\theta)) \in \mathbb{R}^2 : 0 \le r \le 1, \frac{\pi}{4} \le \theta \le \frac{7\pi}{4}\}$ be as in Figure 5.2. Let $V(x) = -2x_1$ be the external potential and W be any C^2 convex interaction potential with $\nabla W(0) = 0$. Define $\gamma_1(s) = (1,-1)\,s$ and $\gamma_2(s) = (1,1)\,s$ for $0 \le s \le 1$. Then for initial datum $\mu_0 = \delta_0$ both $\mu_1(t) = \delta_{\gamma_1(t)}$ and $\mu_2(t) = \delta_{\gamma_2(t)}$ are weak measure solutions. Thus uniqueness and hence stability of solutions cannot hold.*

Strategy of the proof.

The strategy to construct weak measure solutions to (5.0.1) is to show the existence of gradient flow with respect to $\mathcal{E}$. We approximate the initial data μ_0 in Wasserstein metric

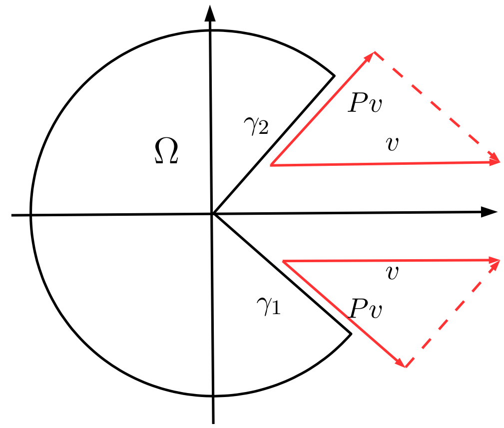


Figure 5.2: The red arrows show the projected velocity field Pv on γ_1 and γ_2, which are driving the particles apart from each other.

by $\mu_0^n = \sum_{i=1}^{k(n)} m_i^n \delta_{x_i^n}$ for $x_i^n \in \Omega \bigcap B(n)$, and solve (5.0.1) with $\mu^n(0) = \mu_0^n$. Then (5.0.1) becomes a discrete projected system, for $1 \le i \le k(n)$

$$\begin{cases} \dot{x}_i^n(t) = P_{x_i^n(t)} \left(-\sum_j m_j^n \nabla W\left(x_i^n - x_j^n\right) - \nabla V(x_i^n) \right) \text{ a.e. } t \ge 0, \\ x_i^n(0) = x_i^n \in \Omega, \end{cases} \tag{5.1.16}$$

which we show its well-posedness based on the well-posedness theory from non-convex sweeping process differential inclusions with perturbations. For the general theory of sweeping processes we refer to [43, 44, 112] and references therein. To be precise, based on [44] there exists a locally absolutely continuous curve $[0, \infty) \ni t \mapsto x(t) = (x_1^n(t), \cdots, x_{k(n)}^n(t)) \in \Omega^{k(n)} = \Omega \times ... \times \Omega$, such that for a.e. $t > 0$,

$$-\dot{x}(t) \in N(\Omega^{k(n)}, x(t)) - v(t, x(t)), \tag{5.1.17}$$

where $v(t, x(t)) = -\sum_j m_j^n \nabla W\left(x_i^n - x_j^n\right) - \nabla V(x_i^n)$ in our case. We then show that the solution to (5.1.17) is actually a solution to (5.1.16). We denote $\mu^n(t) = \sum_{i=1}^{k(n)} m_i^n \delta_{x_i^n(t)}$.

Next we explore the properties of the sequence of solutions $\{\mu^n(\cdot)\}_n$. In particular,

- When Ω is bounded or Ω is unbounded but convex, we first prove the stability of $\mu^n(t)$

$$d_W(\mu^n(t), \mu^m(t)) \le \exp(Ct) d_W(\mu_0^n, \mu_0^m), \tag{5.1.18}$$

where $C = C(W, V)$ is a constant depending only on W, V. Thus $\mu^n(\,\cdot\,)$ converges to some $\mu(\,\cdot\,)$ in $\mathcal{P}_2(\Omega)$ as $n \to \infty$. Since $\mu^n(\,\cdot\,)$ satisfies the energy dissipation inequality,

$$\begin{aligned}\mathcal{E}(\mu^n(s)) \geq \mathcal{E}(\mu^n(t)) &+ \frac{1}{2}\int_s^t |\,(\mu^n)'\,|^2(r)dr\\ &+ \frac{1}{2}\int_s^t \int_\Omega |P_x\left(-\nabla W * \mu^n(r)(x) - \nabla V(x)\right)|^2 d\mu^n(r,x)dr,\end{aligned}$$

by the lower semicontinuity property, we are able to show that $\mu(\,\cdot\,)$ also satisfies the desired energy dissipation inequality

$$\begin{aligned}\mathcal{E}(\mu(s)) \geq \mathcal{E}(\mu(t)) &+ \frac{1}{2}\int_s^t |\mu'|^2(r)dr\\ &+ \frac{1}{2}\int_s^t \int_\Omega |P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)|^2 d\mu(r,x)dr.\end{aligned} \tag{5.1.19}$$

We then show the chain rule, for $\tilde{v}(t)$ is the tangent velocity of $\mu(\,\cdot\,)$ at time t

$$\frac{d}{dt}\mathcal{E}(\mu(t)) = \int_\Omega \left\langle -P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right), \tilde{v}(t,x)\right\rangle d\mu(t,x), \tag{5.1.20}$$

which together with the energy dissipation inequality yields that $\mu(\,\cdot\,)$ is a gradient flow with respect to $\mathcal{E}$ and a weak measure solution to (5.0.1).

- When Ω is unbounded and only η-prox-regular, we first show that the support of the solutions $\mu^n(\,\cdot\,)$ grows at most exponentially, i.e.

$$\operatorname{supp}\left(\mu^n(t)\right) \subset B(r(t)), \tag{5.1.21}$$

for $r(t) = (r_0 + 1)\exp(Ct)$ given that $\operatorname{supp}(\mu_0) \subset B(r_0)$. We then show that, given $\operatorname{supp}\left(\mu^n(t)\right)$ has the same growth condition for all $n \in \mathbb{N}$, $\mu^n(\,\cdot\,)$ still converges to a locally absolutely continuous curve $\mu(\,\cdot\,)$ satisfying (5.1.19) and (5.1.20). Thus $\mu(\,\cdot\,)$ is a weak measure solution to (5.0.1).

Outline

This rest of this Chapter is organized as follows.

In Section 5.2, we show the properties of the projection P and then give the existence results for the discrete projected systems (5.1.16).

In Section 5.3, under the assumption that Ω is bounded, we prove the stability of solutions to the discrete projected systems $\mu^n(\,\cdot\,)$, i.e. (5.1.18). Thus $\mu^n(\,\cdot\,)$ converge to an absolutely continuous curve $\mu(\,\cdot\,)$. We show that $\mu(\,\cdot\,)$ is curve of maximum slope for the energy $\mathcal{E}$ and moreover a gradient flow with respect to $\mathcal{E}$. We then show that $\mu(\,\cdot\,)$ is also a weak measure solution and that weak measure solutions satisfy the stability property

(5.1.8). At the end of the section, we show that solutions are characterized by the system of Evolution Variational Inequalities (5.1.9).

Section 5.4 addresses the case of unbounded, convex Ω and general initial data $\mu_0 \in \mathcal{P}_2(\Omega)$, that is Theorem 5.1.9. The proof of Theorem 5.1.9 is similar to Theorem 5.1.5 and Theorem 5.1.6, we only concentrate on the key differences.

Section 5.5 is devoted to the case when Ω is unbounded and only η-prox-regular with $\mathrm{supp}(\mu_0)$ compact. We show that the support of the solutions to the discrete projected systems (5.1.16) satisfy exponential growth condition (5.1.21). By similar stability results as in Section 5.3, $\mu^n(\,\cdot\,)$ still converges to a locally absolutely continuous curve $\mu(\,\cdot\,)$ and $\mu(\,\cdot\,)$ is a solution to (5.0.1) with the desired energy dissipation (5.1.19). We then give the proof of the stability result (5.1.15) for solutions with control on growth of supports. We end the section by making a remark about well-posedness of (5.0.1) with time-dependent potentials W, V.

In the last Section 5.6, we prove Proposition 5.1.7 and discuss the conditions on the shape of the domain Ω such that there exist interaction potentials W for which solutions $\mu(\,\cdot\,)$ of (5.0.1) aggregate to a singleton (a single delta mass).

5.2 Existence of solutions to discrete systems

In this Section, we first show properties of the projection P, in particular the lower semicontinuity and convexity property of P. Then we give the existence result of solutions to the discrete projected systems (5.1.16).

Recall that the tangent and normal cones $T(\Omega, x)$ and $N(\Omega, x)$ are closed convex cones by Definition 5.1.4.

Proposition 5.2.1. *Suppose Ω satisfies (M1) and $x \in \partial\Omega$. Then for any $v \in \mathbb{R}^d$, there exist a unique orthogonal decomposition $(v_T, v_N) \in T(\Omega, x) \times N(\Omega, x)$ of v such that*

$$\langle v_T, v_N \rangle = 0 \quad \textit{and} \quad v = v_T + v_N.$$

Moreover, $v_T = \mathrm{proj}_{T(\Omega,x)}(v) = P_x(v), v_N = \mathrm{proj}_{N(\Omega,x)}(v)$.

Proposition 5.2.1 is a direct consequence of Moreau's decomposition theorem, see [87, 99] for the proof.

Proposition 5.2.2. *Assume Ω satisfies (M1), then the map $\Omega \times \mathbb{R}^d \ni (x, v) \mapsto |P_x(v)|^2$ is lower semicontinuous and for any fixed $x \in \Omega$, $\mathbb{R}^d \ni v \mapsto |P_x(v)|^2$ is convex.*

Proof. We first show the lower semicontinuity property. Let $\{x_n\}_n \subset \Omega, \{v^n\}_n \subset \mathbb{R}^d$ be such that $\lim_{n\to\infty} x_n = x \in \Omega, \lim_{n\to\infty} v^n = v$. If $x_n \in \mathring{\Omega}$ for all n sufficiently large, then

$P_{x_n}(v^n) = v^n$ and we have $|P_x(v)|^2 \leq |v|^2 = \lim_{n\to\infty} |v^n|^2$. And for any $x \in \mathring{\Omega}$, we have $x_n \in \mathring{\Omega}$ for n sufficiently large, thus

$$\liminf_{n\to\infty} |P_{x_n}(v^n)|^2 \geq |P_x(v)|^2.$$

So we only need to check for $x \in \partial\Omega$ and $\{x_n\}_n \subset \partial\Omega$ such that $\lim_{n\to\infty} x_n = x$. Denote the decomposition of v^n as in Proposition 5.2.1 by

$$v^n = v_T^n + v_N^n$$

where $v_T^n \in T(\Omega, x_n), v_N^n \in N(\Omega, x_n)$ and $\langle v_T^n, v_N^n \rangle = 0$. For any subsequence, which we do not relabel, such that there exists $w_N \in \mathbb{R}^d$ and $\lim_{n\to\infty} v_N^n = w_N$, we claim that $w_N \in N(\Omega, x)$ and $\langle v - w_N, w_N \rangle = 0$. Indeed, since Ω is η-prox-regular,

$$B_\eta\left(x_n + \eta \frac{v_N^n}{|v_N^n|}\right) \cap \Omega = \emptyset.$$

Taking $n \to \infty$ implies

$$B_\eta\left(x + \eta \frac{w_N}{|w_N|}\right) \cap \Omega = \emptyset,$$

which then implies $w_N \in N(\Omega, x)$. Also by taking $n \to \infty$ in $\langle v^n - v_N^n, v_N^n \rangle = 0$ we get $\langle v - w_N, w_N \rangle = 0$. We then know

$$\begin{aligned}
|P_x(v)|^2 &= |v_T|^2 \\
&= |v - v_N|^2 \\
&\leq |v - w_N|^2 \\
&= \lim_{n\to\infty} |v^n - v_N^n|^2 \\
&= \lim_{n\to\infty} |P_{x_n}(v^n)|^2
\end{aligned}$$

So

$$\liminf_{n\to\infty} |P_{x_n}(v^n)|^2 \geq |P_x(v)|^2.$$

We turn to the convexity property. For any fixed $x \in \Omega$, if $x \in \mathring{\Omega}$ then $P_x(v) = v$ for all $v \in \mathbb{R}^d$ and $v \mapsto |v|^2$ is convex. Now for fixed $x \in \partial\Omega$, and any $v^1, v^2 \in \mathbb{R}^d, 0 \leq \theta \leq 1$, denote the unique projection of v^1, v^2 defined in Proposition 5.2.1 by

$$v^i = v_N^i + v_T^i$$

for $i = 1, 2$. Then

$$(1-\theta)v^1 + \theta v^2 = \left((1-\theta)v_T^1 + \theta v_T^2\right) + \left((1-\theta)v_N^1 + \theta v_N^2\right).$$

Note that $(1-\theta)v_T^1 + \theta v_T^2 \in T(\Omega, x)$ and $(1-\theta)v_N^1 + \theta v_N^2 \in N(\Omega, x)$, by Proposition 5.2.1 we have

$$\begin{aligned} |P_x\left((1-\theta)v^1 + \theta v^2\right)|^2 &\leq |(1-\theta)v_T^1 + \theta v_T^2|^2 \\ &\leq (1-\theta)|v_T^1|^2 + \theta|v_T^2|^2 \\ &= (1-\theta)|P_x\left(v^1\right)|^2 + \theta|P_x\left(v^2\right)|^2. \end{aligned}$$

Convexity is verified. □

We cite the following result from [43, 44] about the existence of differential inclusions

Theorem 5.2.3. *Assume that S is η-prox-regular as defined in Definition 5.1.3 and $F : \mathbb{R}^d \ni x \mapsto F(x) \in \mathbb{R}^d$ is a continuous function with at most linear growth, i.e., there exists some constant $C > 0$ such that*

$$|F(x)| \leq C(1 + |x|).$$

Then the differential inclusion

$$\begin{cases} -\dot{x}(t) \in N(S, x(t)) + F(x(t)) \ \ a.e.\ t \geq 0, \\ \ x(0) = x_0 \in S. \end{cases} \tag{5.2.1}$$

has at least one locally absolutely continuous solution.

Note that the theorems, for example Theorem 5.1 from [44], are more general than Theorem 5.2.3. However, we only need the simplified version for our purpose. We also notice that (5.2.1) implies that $x(t) \in S$ for all $t \geq 0$. Indeed, since $N(S, x) = \emptyset$ for all $x \notin S$ we know $x(t) \in S$ for a.e. $t \geq 0$. Then the continuity of $x(t)$ and the fact that S is closed imply that $x(t) \in S$ for all $t \geq 0$. For completeness, we give a sketch of proof here.

Proof. For $T < \frac{1}{2C}$ where C is constant in the growth condition of F. For $n \in \mathbb{N}$, take the partition $0 = t_0^n < t_1^n < ... < t_n^n = T$ and define $\delta_i^n = t_{i+1}^n - t_i^n, x_0^n = x_0, Z_0^n = F(x_0^n)$. Then define iteratively for $0 \leq i \leq n-1$

$$x_{i+1}^n = \text{proj}_S\left(x_i^n - \delta_i^n Z_i^n\right)$$

and

$$Z_{i+1}^n = F(x_{i+1}^n).$$

Note that we have then

$$\|x_{i+1}^n\| \leq \|x_i^n\| + 2\delta_i^n \|Z_i^n\|$$

and

$$\|Z_i^n\| \leq C\left(1 + \|x_i^n\|\right).$$

Thus

$$\begin{aligned}\|x_{i+1}^n\| &\leq \|x_0\| + \sum_{j=0}^{i} 2\delta_j^n C(1+\|x_j^n\|)\\ &\leq \|x_0\| + 2CT(1+\max_{0\leq j\leq i}\|x_j^n\|),\end{aligned}$$

which implies

$$\max_{0\leq i\leq n}\|x_i^n\| \leq \|x_0\| + 2CT(1+\max_{0\leq i\leq n}\|x_i^n\|).$$

Since $2CT < 1$ we have uniformly in n

$$\max_{0\leq i\leq n}\|x_i^n\| \leq \frac{\|x_0\|+2CT}{1-2CT} < \infty,$$

and

$$\max_{0\leq i\leq n}\|Z_i^n\| \leq C(1+\max_{0\leq i\leq n}\|x_i^n\|) < \infty.$$

We now define the approximation solution by

$$x_n(t) = u_i^n + \frac{x_{i+1}^n - x_i^n + \delta_i^n Z_i^n}{\delta_i^n} - (t-t_i^n)Z_i^n,$$

for $t_i^n \leq t < t_{i+1}^n$. Notice that x_n can also be written as

$$x_n(t) = x_0 + \int_0^t [\Pi_n(s) - Z_n(s)]ds \tag{5.2.2}$$

where

$$\Pi_n(t) = \sum_{i=0}^{n} \frac{x_{i+1}^n - x_i^n + \delta_i^n Z_i^n}{\delta_i^n}\chi_{(t_i^n, t_{i+1}^n]}(t)$$

and $Z_n(t) = Z_i^n$ for $t_i^n \leq t < t_{i+1}^n$. We have for a.e. $t \in [t_i^n, t_{i+1}^n)$

$$\dot{x}_n(t) + Z_n(t) = \Pi_n(t) \in N\left(S, x_n(t_i^n)\right).$$

Since $\|\Pi_n(t)\| \leq \|Z_i^n\|$ for $t \in (t_i^n, t_{i+1}^n]$, we know there exists a subsequence of n, which we do not relabel, such that

$$\Pi_n \rightharpoonup \Pi, \quad Z_n \rightharpoonup Z \qquad \text{as } n \to \infty$$

weakly in $L^2[0,T]$. We then have by (5.2.2) that x_n converges locally uniformly to x with

$$x(t) = x_0 + \int_0^t [\Pi(s) - Z(s)]ds.$$

We now claim that $x(t)$ is a solution to the differential inclusion on $[0,T]$. First we check that $x(t) \in S$ for all $t \in [0,T]$. Since

$$\|x_n(t_i^n) - x(t)\| \leq \|x_n(t) - x(t)\| + c|t_i^n - t|,$$

$x(t) = \lim_{n\to\infty} x_n(t_i^n) \in S$. We then verify that $\dot{x}(t) + Z(t) \in -N(S, x(t))$ for a.e. $t \in [0, T]$. Since $\dot{x}_n + Z_n = \Pi_n \rightharpoonup \Pi$ weakly in $L^2([0,T])$ and $\Pi_n(t) \in N(S, x_n(t_i^n))$ for $t_i^n < t \le t_{i+1}^n$, by Mazur's lemma, for a.e. $t \in [0,T]$

$$\dot{x}(t) + Z(t) \in \bigcap_n \{\dot{x}_k(t) + Z_k(t) : k \ge n\}.$$

Then by Proposition 2.1 from [44], we know for a.e. $t \in [0,T]$,

$$\dot{x}(t) + Z(t) \in N(S, x(t)).$$

Now we only need to check that $Z(t) = F(x(t))$. We know that $Z_n(t) = F(x^n(t_i^n)$ for $t_i^n \le t < t_{i+1}^n$. Define $\tilde{u}_n$ by $\tilde{x}_n(t) = x^n(t_i^n)$ for $t_i^n \le t < t_{i+1}^n$ and note $Z_n(t) = F(x^n(t_i^n) = F(\tilde{x}_n(t))$. Then $\tilde{x}_n$ converges locally uniformly to x. Together with the fact that F is continuous, $F(\tilde{x}_n)$ converges to $F(x)$ in $L^2([0,T])$. Since it is direct to check Z_n converges weakly to Z in $L^2([0,T])$, we get $Z(t) = F(x(t))$ for a.e. $t \in [0,T]$. The claim is proved. $\square$

We now show that the solutions for the differential inclusions are actually solutions for the projected systems.

Lemma 5.2.4. *Assume that S is η-prox-regular by Definition 5.1.3 and $x(t)$ is a locally absolutely continuous solution to the differential inclusion (5.2.1). Then*

$$\dot{x}(t) = P_{x(t)}\left(-F\left(x(t)\right)\right) \quad a.e.\ t \ge 0. \tag{5.2.3}$$

Proof. Since S is η-prox-regular, it is tangentially regular, that is

$$T(S,x) = K(S,x)$$

where $T(S,x)$ is defined in Definition 5.1.4 and $K(S,x)$ is the contingent cone defined as

$$K(S,x) = \{v \in \mathbb{R}^d : \exists t_n \searrow 0\ \exists v_n \to v \text{ s.t. } (\forall n)\ x + t_n v_n \in S\}.$$

We refer to [25] for the details. Now note that for a.e. t

$$\dot{x}(t) = \lim_{h\to 0^+} \frac{x(t+h) - x(t)}{h} \in K(S, x(t))$$

and

$$\dot{x}(t) = \lim_{h\to 0^-} \frac{x(t+h) - x(t)}{h} \in -K(S, x(t)).$$

Thus $\langle \dot{x}(t), n(x(t)) \rangle = 0$ for any $n(x(t)) \in N(S, x(t))$. From the differential inclusion (5.2.1),we know that $-F(x(t)) = \dot{x}(t) + n(x(t))$ for some $n(x(t)) \in N(S, x(t))$. Together with fact that $\dot{x}(t) \in T(S, x(t))$ and $\langle \dot{x}(t), n(x(t)) \rangle = 0$, by Proposition 5.2.1

$$\dot{x}(t) = P_{x(t)}\left(-F\left(x(t)\right)\right),$$

as claimed. $\square$

We turn to the existence of solutions to the discrete projected system (5.1.16), which we write as

$$\begin{cases} \dot{x}_i(t) = P_{x_i(t)}\left(v(x^n(t))\right), \\ x_i(0) = x_i \in S. \end{cases} \tag{5.2.4}$$

for $i = 1, \cdots, n$. For that purpose we apply Theorem 5.2.3 and Lemma 5.2.4 for $S = \Omega^n$ and $x_0 = (x_1, \cdots, x_n)$ with $F(x) = (-v_1(x(t)), \cdots, -v_n(x(t)))$, where $v_i(x(t)) = -\nabla W * \mu(t)(x_i(t)) - \nabla V(x_i(t)) = -\sum_{j=1}^{n} m_j \nabla W\left(x_i(t) - x_j(t)\right) - \nabla V(x_i(t))$. To do that, we first check that Ω^n is η-prox-regular.

Proposition 5.2.5. *If $\Omega \subset \mathbb{R}^d$ is η-prox-regular by Definition 5.1.3, then*

$$\Omega^n = \{(x_1, \cdots, x_n) \, : \, x_i \in \Omega, \;\; i = 1, \ldots, n\}$$

is η-prox-regular; Also for any $x = (x_1, \cdots, x_n) \in \Omega^n$ we have

$$N(\Omega^n, x) = N(\Omega, x_1) \times \cdots \times N(\Omega, x_n).$$

Proof. To see Ω^n is also η-prox-regular, first it is direct that Ω^n is a closed set. Now for any $x = (x_1, \cdots, x_n) \in \partial\Omega^n$ and $v = (v^1, \cdots, v^n) \in N(\Omega^n, x)$, by Definition 5.1.4 there exists $\alpha > 0$ such that

$$x \in P_{\Omega^n}\left(x + \alpha v\right),$$

which implies

$$x_i \in P_{\Omega}\left(x_i + \alpha v^i\right)$$

for $1 \le i \le n$. By the equivalent definition of η-prox-regularity of Ω (5.1.4), we then have

$$\langle v^i, y_i - x_i\rangle \le \frac{|v^i|}{2\eta}|y_i - x_i|^2$$

for any $y_i \in \Omega$. Thus

$$\begin{aligned} \langle v, y - x\rangle &= \sum_{i=1}^{n} \langle v^i, y_i - x_i\rangle \\ &\le \sum_{i=1}^{n} \frac{|v^i|}{2\eta}|y_i - x_i|^2 \\ &\le \frac{|v|}{2\eta}|y - x|^2, \end{aligned}$$

for any $y = (y_1, \cdots, y_n) \in \Omega^n$. Thus Ω^n is η-prox-regular by (5.1.4). We now turn to the relations between the normal cones. For $x = (x_1, \cdots, x_n) \in \Omega^n$ and $v = (v^1, \cdots, v^n)$

$$\begin{aligned} v \in N(\Omega^n, x) &\Leftrightarrow \exists \alpha > 0 \text{ s.t. } x \in P_{\Omega^n}\left(x + \alpha v\right) \\ &\Leftrightarrow x_i \in P_{\Omega}\left(x_i + \alpha v^i\right), \;\; i = 1, \ldots, n \\ &\Leftrightarrow v_i \in N(\Omega, x_i), \;\; i = 1, \ldots, n. \end{aligned}$$

Thus $N(\Omega^n, x) = N(\Omega, x_1) \times \cdots \times N(\Omega, x_n)$. $\square$

Now we give the main result regarding the existence of solutions to projected discrete systems.

Theorem 5.2.6. *Assume that Ω is η-prox-regular by Definition 5.1.3. If either Ω is bounded and W, V satisfy (A1)-(A2) or Ω is unbounded and W, V satisfy (GA2) and (GA4) i.e. (LA2) and (LA4), then for any $n \in \mathbb{N}$ and any $(x_1, \cdots, x_n) \in \Omega^n, (m_1, \cdots, m_n) \in \mathbb{R}^n$ with $m_i \geq 0, \sum_{i=1}^n m_i = 1$, the projected discrete system*

$$\begin{cases} \dot{x}_i(t) = P_{x_i(t)}\left(v_i(x(t))\right), \\ x_i(0) = x_i \in \Omega, \end{cases} \tag{5.2.5}$$

for $i = 1, \cdots, n$, where $v_i(x(t)) = -\nabla W * \mu(t)(x_i(t)) - \nabla V(x_i(t)) = -\sum_{j=1}^n m_j \nabla W(x_i(t) - x_j(t)) - \nabla V(x_i(t))$, has a locally absolutely continuous solution.

Proof. We just need to check the conditions for Theorem 5.2.3 to apply. We already know that Ω^n is η-prox-regular. If Ω is bounded and W, V satisfy (A1)-(A2), then the mapping

$$\Omega^n \ni y = (y_1, \cdots, y_n) \mapsto F(y) = (\nabla W * \mu(y_1) + \nabla V(y_1), \cdots, \nabla W * \mu(y_n) + \nabla V(y_n))$$

where $\mu = \sum_{i=1}^n m_i \delta_{y_i}$, is continuous and bounded. Extend F to $\mathbb{R}^{dn}$ so that F is still continuous and bounded. Then by Theorem 5.2.3 there exists an absolutely continuous solution to the differential inclusion

$$\begin{cases} -\dot{x}(t) \in N(\Omega^n, x(t)) + F(x(t)), \\ x(0) = (x_1, \cdots, x_n) \in \Omega^n. \end{cases} \tag{5.2.6}$$

Similarly, if Ω is unbounded and $\nabla W, \nabla V$ satisfy liner growth conditions (GA2) and (GA4), then the mapping

$$\mathbb{R}^{dn} \ni y = (y_1, \cdots, y_n) \mapsto F(y) = (\nabla W * \mu(y_1) + \nabla(y_1), \cdots, \nabla W * \mu(y_n) + \nabla V(y_n))$$

where $\mu = \sum_{i=1}^n m_i \delta_{y_i}$, is continuous and has linear growth on $\mathbb{R}^{dn}$. By Theorem 5.2.3, we still have an absolutely continuous solution to (5.2.6).
Now consider (5.2.6) in components yields for $1 \leq i \leq n$ and $v_i(x) = -\sum_{j=1}^n \nabla W(x_i - x_j)m_j - \nabla V(x_i)$,

$$\begin{cases} -\dot{x}_i(t) \in N(\Omega, x_i(t)) - v_i(x(t)), \\ x_i(0) = x_i \in \Omega. \end{cases}$$

Then similar argument as in Lemma 5.2.4 gives

$$\begin{cases} \dot{x}_i(t) = P_{x_i(t)}\left(v_i(x(t))\right), \\ x_i(0) = x_i \in \Omega. \end{cases}$$

$\square$

Remark 5.2.7. *Here we make a remark that for any continuous vector filed ξ on Ω, if we define*

$$X(t,x) = P_\Omega(x + t\xi(x))$$

for all $t \geq 0$ and $x \in \Omega$, then $X(t,\cdot): \Omega \to \Omega$ is continuous in t for t small and it satisfies

$$\frac{d^+}{dt}X(t,x)|_{t=0} = \lim_{t\to 0^+}\frac{X(t,x)-x}{t} = \lim_{t\to 0^+}\frac{P_\Omega(x+t\xi(x))-x}{t} = P_x(\xi(x)), \tag{5.2.7}$$

for all $x \in \Omega$. We need this in Remark 5.3.6. Local continuity of X with respect to t follows from the local Lipschitz property of P_Ω, see [96].

To see (5.2.7), for any fixed $x \in \Omega$, we make the following two claims.

Claim 1: For any fixed $\xi \in T(\Omega, x)$,

$$\lim_{t\to 0^+}\frac{P_\Omega(x+t\xi)-x}{t} = \xi.$$

Claim 2: For any fixed ξ (not necessarily in $T(\Omega,x)$),

$$\lim_{t\to 0^+}\frac{P_\Omega\left(x+t\xi\right)-P_\Omega\left(x+tP_x\left(\xi\right)\right)}{t} = 0.$$

If Claim 1 and 2 are true, then

$$\begin{aligned}
\lim_{t\to 0^+}\frac{P_\Omega(x+t\xi(x))-x}{t} &= \lim_{t\to 0^+}\frac{P_\Omega(x+t\xi(x))-P_\Omega\left(x+tP_x\left(\xi(x)\right)\right)}{t} \\
&+ \lim_{t\to 0^+}\frac{P_\Omega\left(x+tP_x\left(\xi(x)\right)\right)-x}{t} \\
&= 0 + P_x\left(\xi(x)\right) = P_x\left(\xi(x)\right).
\end{aligned}$$

For Claim1, recall from [25] that for any fixed $\xi \in T(\Omega,x)$ and any sequence t_n decreasing to 0, there exists a sequence ξ_n such that $x + t_n\xi_n \in \Omega$ for all n and $\lim_{n\to\infty}\xi_n = \xi$. Now fix $x \in \Omega$, for any t_n positive, decreasing to 0, take $\xi_n \to \xi$ such that $x + t_n\xi_n \in \Omega$ for all n. Then

$$\lim_{n\to\infty}\frac{P_\Omega(x+t_n\xi)-x}{t_n} = \lim_{n\to\infty}\frac{P_\Omega(x+t_n\xi)-P_\Omega(x+t_n\xi_n)+P_\Omega(x+t_n\xi_n)-x}{t_n}.$$

We note that

$$\lim_{n\to\infty}\frac{P_\Omega(x+t_n\xi_n)-x}{t_n} = \lim_{n\to\infty}\frac{x+t_n\xi_n-x}{t_n} = \lim_{n\to\infty}\xi_n = \xi,$$

and

$$\begin{aligned}
\lim_{n\to\infty}\frac{P_\Omega(x+t_n\xi)-P_\Omega\left(x+t_n\xi_n\right)}{t_n} &\leq \lim_{n\to\infty}\frac{|x+t_n\xi-x-t_n\xi_n|}{t_n} \\
&= \lim_{n\to\infty}|\xi-\xi_n| = 0,
\end{aligned}$$

where we used the local Lipschitz property of the projection P_Ω onto Ω, refer to [96]. Thus Claim 1 holds true.

We now turn to Claim 2. By the proof of Proposition 3.1 of [96], we take $\lambda > 0$ small enough in the proof, then there exists a constant $C = C(\lambda) > 0$ such that for t small enough

$$\begin{aligned}
&\lim_{t\to 0^+} \frac{C}{t^2} \left|P_\Omega(x+t\xi) - P_\Omega(x+tP_x(\xi))\right|^2 \\
\le &\lim_{t\to 0^+} \frac{1}{t^2} \left\langle P_\Omega(x+t\xi) - P_\Omega(x+tP_x(\xi)), x+t\xi - x - tP_x(\xi)\right\rangle \\
= &\lim_{t\to 0^+} \frac{1}{t} \left\langle P_\Omega(x+t\xi) - P_\Omega(x+tP_x(\xi)), \xi - P_x(\xi)\right\rangle \\
= &\lim_{t\to 0^+} \frac{1}{t} \left\langle P_\Omega(x+t\xi) - x + x - P_\Omega(x+tP_x(\xi)), \xi - P_x(\xi)\right\rangle
\end{aligned}$$

Since $\xi - P_x(\xi) \in N(\Omega, x)$, by prox-regular property of Ω and local Lipischitz property of P_Ω,

$$\begin{aligned}
\lim_{t\to 0^+} \frac{1}{t} \left\langle P_\Omega(x+t\xi) - x, \xi - P_x(\xi)\right\rangle &\le \lim_{t\to 0^+} \frac{1}{t} \frac{|\xi - P_x(\xi)|}{2\eta} \left|P_\Omega(x+t\xi) - x\right|^2 \\
&\le \lim_{t\to 0^+} \frac{1}{t} \frac{|\xi - P_x(\xi)|}{2\eta} t^2 |\xi|^2 \\
&= \lim_{t\to 0^+} t \frac{|\xi - P_x(\xi)|}{2\eta} |\xi|^2 = 0.
\end{aligned}$$

Since $P_x(\xi) \in T(\Omega, x)$ and $\langle P_x(\xi), \xi - P_x(\xi)\rangle = 0$, by Claim 1

$$\lim_{t\to 0^+} \frac{1}{t} \left\langle x - P_\Omega(x+tP_x(\xi)), \xi - P_x(\xi)\right\rangle = -\left\langle P_x(\xi), \xi - P_x(\xi)\right\rangle = 0.$$

So

$$\lim_{t\to 0^+} \frac{1}{t^2} \left|P_\Omega(x+t\xi) - P_\Omega(x+tP_x(\xi))\right|^2 = 0.$$

5.3 Existence and stability of solutions with Ω bounded

In this Section, we show the existence and stability of solutions to (5.0.1) for the case when Ω is bounded, prox-regular and W, V satisfy (A1)-(A2).

We approximate $\mu_0 \in \mathcal{P}_2(\Omega)$ by $\mu_0^n = \sum_{i=1}^{k(n)} m_i^n \delta_{x_i^n}$ such that $\lim_{n\to\infty} d_W(\mu_0, \mu_0^n) = 0$ with $x_i^n \in \Omega$. By Theorem 5.2.6, for each $n \in \mathbb{N}$ there exists a a locally absolutely continuous solution to

$$\begin{cases} \dot{x}_i^n(t) = P_{x_i^n(t)}\left(v_i^n(x(t))\right), \quad 1 \le i \le k(n) \\ x_i^n(0) = x_i^n \in \Omega, \end{cases} \tag{5.3.1}$$

for $t \geq 0$, where

$$v_i^n\left(x(t)\right) = -\nabla W * \mu^n(t)(x_i^n(t)) - \nabla V(x_i^n(t)) = -\sum_{j=1}^{k(n)} m_j^n \nabla W\left(x_i^n(t) - x_j^n(t)\right) - \nabla V\left(x_i^n(t)\right)$$

and $\mu^n(t) = \sum_{j=1}^{k(n)} m_j^n \delta_{x_j^n(t)}$. It is a straightforward calculation to see that for any $\phi \in C_c^\infty(\mathbb{R}^d)$

$$\frac{d}{dt}\int_{\mathbb{R}^d} \phi(x) d\mu^n(t,x) = \int_{\mathbb{R}^d} \langle \nabla\phi(x), P_x\left(v^n(t,x)\right)\rangle \, d\mu^n(t,x).$$

Thus $\mu^n(t)$ satisfies

$$\frac{\partial}{\partial t}\mu^n(t,x) + \operatorname{div}\left(\mu^n(t,x) P_x\left(v^n(t,x)\right)\right) = 0,$$

in the sense of distributions for $v^n(t,x) = -\nabla W * \mu^n(t)(x) - \nabla V(x)$.

The following proposition contains the key estimate on the stability of solutions in the discrete case. In particular it shows how the stability in Wasserstein metric d_W defined in (5.1.2) is affected by the lack of convexity of the domain.

Proposition 5.3.1. *Assume that Ω is bounded and satisfies (M1), W, V satisfy (A1) and (A2). Then for two solutions $\mu^n(\,\cdot\,)$ and $\mu^m(\,\cdot\,)$ to the discrete system with different initial data μ_0^n, μ_0^m, we have for all $t \geq 0$*

$$d_W\left(\mu^n(t), \mu^m(t)\right) \leq \exp\left(\left(-\lambda_W^- - \lambda_V + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{\eta}\right) t\right) d_W\left(\mu_0^n, \mu_0^m\right). \tag{5.3.2}$$

Proof. Note that $\mu^n(\,\cdot\,)$ is solution to the continuity equation

$$\partial_t \mu^n(t,x) + \operatorname{div}\left(\mu^n(t,x) P_x\left(v^n(t,x)\right)\right) = 0, \tag{5.3.3}$$

for $v^n(t,x) = -\nabla W * \mu^n(t)(x) - \nabla V(x)$. Since the discrete solutions may have different numbers of particles we use a transportation plan to relate them. Let $\gamma_t \in \Gamma_o\left(\mu^n(t), \mu^m(t)\right)$ be the optimal plan between μ^m and μ^n defined in (5.1.3). By Theorem 8.4.7 and Lemma 4.3.4 from [5]

$$\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu^n(t), \mu^m(t)\right) \leq \int_\Omega \langle P_x\left(v^n(t,x)\right) - P_y\left(v^m(t,y)\right), x - y\rangle \, d\gamma_t(x,y). \tag{5.3.4}$$

We first establish the contractivity the solutions would have if the boundary conditions were not present and then account for the change due to velocity projection at the boundary.

For v^n, v^m, by (A1) and (A2), that is the convexity of W and V,

$$
\begin{aligned}
&\int_{\Omega\times\Omega} \langle v^n(t,x) - v^m(t,y), x-y\rangle d\gamma_t(x,y) \\
=&\int_{\Omega\times\Omega} \langle -\nabla W * \mu^n(t)(x) - \nabla V(x) + \nabla W * \mu^m(t)(y) - \nabla V(y), x-y\rangle d\gamma_t(x,y) \\
=&\frac{1}{2}\int_{\Omega\times\Omega}\int_{\Omega\times\Omega} \langle -\nabla W(x-z) + \nabla W(y-w), x-y-z+w\rangle d\gamma_t(z,w) d\gamma_t(x,y) \\
&+ \int_{\Omega\times\Omega} \langle -\nabla V(x) + \nabla V(y), x-y\rangle d\gamma_t(x,y) \\
\leq& -\frac{1}{2}\lambda_W \int_{\Omega\times\Omega}\int_{\Omega\times\Omega} |x-z-y+w|^2 d\gamma_t(z,w) d\gamma_t(x,y) - \lambda_V \int_{\Omega\times\Omega} |x-y|^2 d\gamma_t(x,y) \qquad (5.3.5) \\
\leq& \left(-\lambda_W^- - \lambda_V\right) \int_{\Omega\times\Omega} |x-y|^2 d\gamma_t(x,y) \\
=& \left(-\lambda_W^- - \lambda_V\right) d_W^2\left(\mu^n(t), \mu^m(t)\right).
\end{aligned}
$$

For the boundary effect, by the fact that Ω is η-prox-regular we have (5.1.4), thus

$$
\begin{aligned}
&\int_{\Omega\times\Omega} \langle P_x\left(v^n(t,x)\right) - v^n(t,x) - P_y\left(v^m(t,y)\right) + v^m(t,y), x-y\rangle d\gamma_t(x,y) \\
\leq& \int_{\Omega\times\Omega} \frac{\|v^n(t)\|_{L^\infty(\Omega)} + \|v^m(t)\|_{L^\infty(\Omega)}}{2\eta} |y-x|^2 d\gamma_t(x,y) \qquad (5.3.6) \\
=& \frac{\|v^n(t)\|_{L^\infty(\Omega)} + \|v^m(t)\|_{L^\infty(\Omega)}}{2\eta} d_W^2\left(\mu^n(t), \mu^m(t)\right).
\end{aligned}
$$

Notice that $v^i(x) = -\nabla W(x) * \mu^i(t)(x) - \nabla V(x)$ implies that for $i = n, m$

$$
\|v^i\|_{L^\infty(\Omega)} \leq \|\nabla W\|_{L^\infty(\Omega - \Omega)} + \|\nabla V\|_{L^\infty(\Omega)} < \infty.
$$

Plugging back into (5.3.4) we have

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu^n(t), \mu^m(t)\right) \\
\leq& \int_{\Omega} \left\langle P_x\left(v^n(t,x)\right) - P_y\left(v^m(t,y)\right), x-y\right\rangle d\gamma_t(x,y) \\
=& \int_{\Omega\times\Omega} \left\langle v^n(t,x) - v^m(t,y), x-y\right\rangle d\gamma_t(x,y) \\
&+ \int_{\Omega\times\Omega} \left\langle P_x\left(v^n(t,x)\right) - v^n(t,x) - P_y\left(v^m(t,y)\right) + v^m(t,y), x-y\right\rangle d\gamma_t(x,y) \\
\leq& \left(-\lambda_W^- - \lambda_V + \frac{\|\nabla W\|_{L^\infty(\Omega - \Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{\eta}\right) d_W^2\left(\mu^n(t), \mu^m(t)\right).
\end{aligned}
$$

By Gronwall's inequality, we know (5.3.2) for all $t \geq 0$. $\square$

Since $n \to \infty, d_W(\mu_0^n, \mu_0) \to 0$, by Proposition 5.3.1 the solutions $\mu^n(\,\cdot\,)$ of (5.3.1) form a Cauchy sequence in n, with respect to Wasserstein metric. Thus

$$\mu^n(t) \xrightarrow{d_W} \mu(t) \quad \text{as } n \to \infty, \tag{5.3.7}$$

for all $t \geq 0$ and some $\mu(t) \in \mathcal{P}_2(\Omega)$.

Remark 5.3.2. *Our goal is to show that $\mu(\,\cdot\,)$ is a weak measure solution of* (5.0.1)*. The most immediate idea would be to try to pass to limit directly in Definition 5.1.1. However note that since P_x is not continuous in x and thus the velocity field governing the dynamics is not continuous (at the boundary of Ω). Given that $\mu^n(\,\cdot\,)$ converge to $\mu(\,\cdot\,)$ only in the weak topology of measures, the lack of continuity of velocities prevents us to directly pass to limit in the integral formulation given in Definition 5.1.1. To show that $\mu(\,\cdot\,)$ is a weak measure solution of* (5.0.1) *we use the theory of gradient flows in the spaces of probability measures $\mathcal{P}_2(\Omega)$. Namely, we establish that $\mu(\,\cdot\,)$ satisfies the steepest descent property with respect to the total energy $\mathcal{E}$ defined in (5.1.1) by showing $\mu^n(\,\cdot\,)$ satisfies such property and the property is stable under the weak topology of measures (convergence in the Wasserstein metric d_W).*

We show that the limit curve $\mu(\,\cdot\,)$ we get from particle approximation is a curve of maximal slope with respect to $\mathcal{E}$.

Theorem 5.3.3. *$\mu(\,\cdot\,)$ satisfies for any $0 \leq s < t < \infty$*

$$\mathcal{E}\left(\mu(s)\right) \geq \mathcal{E}\left(\mu(t)\right) + \frac{1}{2}\int_s^t |\mu'|^2(r)dr + \frac{1}{2}\int_s^t \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x)dr, \tag{5.3.8}$$

where $v(r,x) = -\int_\Omega \nabla W(x-y)d\mu(r,y) - \nabla V(x)$.

Before proving the theorem, we need the following lower semi-continuity result.

Lemma 5.3.4. *Assume (M1) holds for Ω and $\nu^n \in \mathcal{P}_2(\Omega)$ converges narrowly to $\nu \in \mathcal{P}_2(\Omega)$ with $\sup_n \int_\Omega |x|^2 d\nu^n(x) < \infty$, then*

$$\int_\Omega |P_x\left(v(x)\right)|^2 d\nu(x) \leq \liminf_{n\to\infty} \int_\Omega |P_x\left(v^n(x)\right)|^2 d\nu^n(x), \tag{5.3.9}$$

where $v^n(x) = -\int_\Omega \nabla W(x-y)d\nu^n(y) - \nabla V(x)$ and $v(x) = -\int_\Omega \nabla W(x-y)d\nu(y) - \nabla V(x)$.

Proof. Similar argument as in Lemma 2.7 from [28] yields that $\nabla W * \nu^n$ converges weakly to $\nabla W * \nu$, i.e., for any $\phi \in C_b^0\left(\mathbb{R}^d\right)$

$$\lim_{n\to\infty}\int_\Omega \nabla W * \nu^n(x)\cdot\phi(x)d\nu^n(x) = \int_\Omega \nabla W * \nu(x)\cdot\phi(x)d\nu(x).$$

Then by Proposition 5.2.2 we proved in Section 5.2 and Proposition 6.42 from [53], we know that there exist two sequences of bounded continuous functions a_i, b_i such that for all $x \in \Omega, v \in \mathbb{R}^d$

$$|P_x(v)|^2 = \sup_{i\in\mathbb{N}} \{a_i(x) + b_i(x) \cdot v\}.$$

Thus

$$\begin{aligned}
\liminf_{n\to\infty} \int_\Omega |P_x(v^n(x))|^2 d\nu^n(x) &= \liminf_{n\to\infty} \int_\Omega \sup_i \{a_i(x) + b_i(x) \cdot v^n(x)\} d\nu^n(x) \\
&\geq \liminf_{n\to\infty} \int_\Omega (a_i(x) + b_i(x)(-\nabla W * \nu^n(x) - \nabla V(x))) d\nu^n(x) \\
&= \int_\Omega (a_i(x) + b_i(x)(-\nabla W * \nu(x) - \nabla V(x))) d\nu(x).
\end{aligned}$$

Taking supremum over $i \in \mathbb{N}$ and using Lebesgue's monotone convergence theorem then gives

$$\begin{aligned}
\liminf_{n\to\infty} \int_\Omega |P_x(v^n(x))|^2 d\nu^n(x) &\geq \sup_{i\in\mathbb{N}} \int_\Omega (a_i(x) + b_i(x)(\nabla W * \nu(x) + \nabla V(x))) d\nu(x) \\
&= \int_\Omega |P_x(v(x))|^2 d\nu(x).
\end{aligned}$$

□

We now start to prove the theorem.

Proof of Theorem 5.3.3. We first show that the map $t \mapsto \mathcal{E}(\mu^n(t))$ is locally absolutely continuous. Indeed, for $0 \leq s < t < \infty$

$$\begin{aligned}
&|\mathcal{E}(\mu^n(t)) - \mathcal{E}(\mu^n(s))| \qquad (5.3.10)\\
&= \left| \sum_{i=1}^{k(n)} m_i^n (V(x_i^n(t)) - V(x_i^n(s))) + \frac{1}{2} \sum_{i,j=1}^{k(n)} m_i^n m_j^n \left(W(x_i^n(t) - x_j^n(t)) - W(x_i^n(s) - x_j^n(s)) \right) \right| \\
&\leq \sum_{i=1}^{k(n)} m_i^n \left| V(x_i^n) - V(x_j^n) \right| + \frac{1}{2} \sum_{i,j=1}^{k(n)} m_i^n m_j^n \left| W(x_i^n(t) - x_j^n(t)) - W(x_i^n(s) - x_j^n(s)) \right| \\
&\leq \sum_{i=1}^{k(n)} m_i^n \|\nabla V\|_{L^\infty(\mathrm{Conv}(\Omega))} |x_i^n(t) - x_i^n(s)| + \sum_{i=1}^{k(n)} m_i^n \|\nabla W\|_{L^\infty(\mathrm{Conv}(\Omega - \Omega))} |x_i^n(t) - x_i^n(s)| \\
&\leq \left(\|\nabla V\|_{L^\infty(\mathrm{Conv}(\Omega))} + \|\nabla W\|_{L^\infty(\mathrm{Conv}(\Omega - \Omega))} \right) \sum_{i=1}^{k(n)} m_i^n |x_i^n(t) - x_i^n(s)|.
\end{aligned}$$

Thus $t \mapsto \mathcal{E}(\mu(t))$ is locally absolutely continuous since $t \mapsto x_i^n(t)$ is locally absolutely continuous.

Since $\mu^n(\,\cdot\,)$ are solutions to the discrete systems, it is direct to calculate that

$$\frac{d}{dt}\mathcal{E}(\mu^n(t)) = -\int_\Omega |P_x(v^n(t,x))|^2 d\mu^n(t,x),$$

and $|\,(\mu^n)'\,|^2(t) \leq \int_\Omega |P_x(v^n(t,x))|^2 d\mu^n(t,x)$ for a.e. $t > 0$. Combining with the fact that $t \mapsto \mathcal{E}(\mu^n(t))$ is locally absolutely continuous then gives,

$$\mathcal{E}\left(\mu^n(s)\right) \geq \mathcal{E}\left(\mu^n(t)\right) + \frac{1}{2}\int_s^t |\,(\mu^n)'\,|^2(r)dr + \frac{1}{2}\int_s^t \int_\Omega |P_x\left(v^n(r,x)\right)|^2 d\mu^n(r,x)dr. \quad (5.3.11)$$

Note that Ω is bounded, $W, V \in C^1(\mathbb{R}^d)$ and $\lim_{n\to\infty} d_W\left(\mu^n(r), \mu(r)\right) = 0$ for any $0 \leq r < \infty$, we get

$$\lim_{n\to\infty} \mathcal{E}(\mu^n(r)) = \mathcal{E}(\mu(r)).$$

Also by Lemma 5.3.4, for any $0 \leq r < \infty$

$$\liminf_{n\to\infty} \int_\Omega |P_x\left(v^n(r,x)\right)|^2 d\mu^n(r,x) \geq \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x).$$

By Fatou's lemma, we then have

$$\liminf_{n\to\infty} \int_s^t \int_\Omega |P_x\left(v^n(r,x)\right)|^2 d\mu^n(r,x)dr \geq \int_s^t \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x)dr. \quad (5.3.12)$$

We now claim that

$$\liminf_{n\to\infty} \int_s^t |\,(\mu^n)'\,|^2(r)dr \geq \int_s^t |\mu'|^2(r)dr. \quad (5.3.13)$$

To see that, first notice that $\sup_n \int_s^t |\,(\mu^n)'\,|^2(r)dr < \infty$, so $|\,(\mu^n)'\,| \in L^2([s,t])$ and converges weakly in $L^2([s,t])$ to some function A as $n \to \infty$. We then have for any $0 \leq s \leq S \leq T \leq t < \infty$

$$\begin{aligned} d_W\left(\mu(S), \mu(T)\right) &= \lim_{n\to\infty} d_W\left(\mu^n(S), \mu^n(T)\right) \\ &\leq \liminf_{n\to\infty} \int_S^T |\,(\mu^n)'\,|(r)dr \\ &= \int_S^T A(r)dr. \end{aligned}$$

Thus we have

$$|\mu'|(r) \leq A(r)$$

for $s \leq r \leq t$, which then implies

$$\begin{aligned} \int_s^t |\mu'|^2(r)dr &\leq \int_s^t A^2(r)dr \\ &\leq \liminf_{n\to\infty} \int_s^t |\,(\mu^n)'\,|^2(r)dr. \end{aligned}$$

The claim is proved. Now take $n \to \infty$ in (5.3.11)gives

$$\mathcal{E}\left(\mu(s)\right) \geq \mathcal{E}\left(\mu(t)\right) + \frac{1}{2}\int_s^t |\mu'|^2(r)dr + \frac{1}{2}\int_s^t \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x)dr,$$

as desired. $\square$

Note that as a byproduct of the proof, we obtain that $\mu(\cdot)$ is a locally absolutely continuous curve in $\mathcal{P}_2(\Omega)$. We now show the proof of the main Theorem 5.1.5

Proof of Theorem 5.1.5. Since $\mu(\cdot) \in \mathcal{P}_2(\Omega)$ is locally absolutely continuous, by Theorem 2.2.3, there exists a unique Borel vector field $\tilde{v}$ such that the continuity equation

$$\partial_t \mu(t) + \operatorname{div}\left(\mu(t)\tilde{v}(t)\right) = 0, \tag{5.3.14}$$

holds in the sense of distributions, i.e., tested against all $\phi \in C_c^\infty\left([0,\infty)\times\mathbb{R}^d\right)$, and

$$\int_\Omega |\tilde{v}(t,x)|^2 d\mu(t,x) = |\mu'|^2(t),$$

for a.e. $t \geq 0$. Then by Proposition 8.4.6 from [5], for a.e. $t > 0$

$$\lim_{h\to 0}\left(\pi^1, \frac{1}{h}\left(\pi^2 - \pi^1\right)\right)_\sharp \gamma_t^h = \left(\mathrm{Id}\times\tilde{v}(t)\right)_\sharp \mu(t), \tag{5.3.15}$$

in $(\mathcal{P}_2(\Omega), d_W)$ for any $\gamma_t^h \in \Gamma_o(\mu(t),\mu(t+h))$. Here we also need the following stronger convergence: Denote the disintegration of γ_t^h with respect to $\mu(t)$ by ν_x^h, then as $h \to 0$, $\int_\Omega \frac{y-\cdot}{h} d\nu_\cdot^h(y)$ converges to the vector field $\tilde{v}(t,\cdot)$ weakly in $L^2(\mu(t))$. The observation is that

$$\begin{aligned}
\lim_{h\to 0}\left\|\int_\Omega \frac{y-\cdot}{h} d\nu_\cdot^h(y)\right\|_{L^2(\mu(t))}^2 &= \lim_{h\to 0}\int_\Omega \left|\int_\Omega \frac{y-x}{h} d\nu_x^h(y)\right|^2 d\mu(t,x) \\
&\leq \lim_{h\to 0}\int_{\Omega\times\Omega} \frac{|y-x|^2}{h^2} d\gamma_t^h(x,y) \\
&= \lim_{h\to 0}\frac{d_W^2(\mu(t),\mu(t+h))}{h^2} \\
&< \infty.
\end{aligned}$$

Thus $\int_\Omega \frac{y-\cdot}{h} d\nu_\cdot^h(y)$ converges weakly in $L^2(\mu(t))$ to some vector field $\hat{v}(t,\cdot)$. This together with (5.3.15) implies $\hat{v} = \tilde{v}$ and we have the weak $L^2(\mu(t))$ convergence of $\int_\Omega \frac{y-\cdot}{h} d\nu_\cdot^h(y)$ to $\tilde{v}(t)$ as stated.

We now claim the following chain rule: for a.e. $t > 0$

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right) = \int_\Omega \left\langle \nabla W * \mu(t)(x) + \nabla V(x), \tilde{v}(t,x)\right\rangle d\mu(t,x). \tag{5.3.16}$$

Indeed, we first notice that since $\mu(\,\cdot\,)$ is locally absolutely continuous, $\mathcal{E}(\mu(\,\cdot\,))$ is also locally absolutely continuous. To see that we have

$$
\begin{aligned}
|\mathcal{E}(\mu(t)) - \mathcal{E}(\mu(s))| \leq & \frac{1}{2}\left|\int_{\Omega\times\Omega} W(x-y)d\mu(t,x)d\mu(t,y) - \int_{\Omega\times\Omega} W(z-w)d\mu(s,z)d\mu(s,w)\right| \\
& + \left|\int_{\Omega} V(x)d\mu(t,x) - \int_{\Omega} V(z)d\mu(s,z)\right| \\
& \leq \int_{\Omega\times\Omega} \left(\|\nabla V\|_{L^\infty(\mathrm{Conv}(\Omega))} + \|\nabla W\|_{L^\infty(\mathrm{Conv}(\Omega-\Omega))}\right)|x-z|d\gamma(x,z) \\
& \leq \left(\|\nabla V\|_{L^\infty(\mathrm{Conv}(\Omega))} + \|\nabla W\|_{L^\infty(\mathrm{Conv}(\Omega-\Omega))}\right) d_W\left(\mu(t),\mu(s)\right).
\end{aligned}
$$

Thus by the locally absolute continuity of $\mu(\,\cdot\,)$, $\mathcal{E}(\mu(\,\cdot\,))$ is also locally absolutely continuous. Now for any fixed $\mu,\nu \in \mathcal{P}_2(\Omega)$ and $\gamma \in \Gamma_o(\mu,\nu)$, consider the function

$$
\begin{aligned}
f(t) = & \frac{W\left(t\left(x_1-x_2\right)-(1-t)\left(y_1-y_2\right)\right)-W(x_1-x_2)}{2t} \\
& + \frac{2V\left(tx_2+(1-t)y_2\right)-2V(x_2)}{2t} - \frac{\lambda_V}{2}t|x_2-y_2|^2 - \frac{\lambda_W}{2}t\left(|x_1-y_1|^2+|x_2-y_2|^2\right).
\end{aligned}
\tag{5.3.17}
$$

Due to (A1) and (A2), the λ-geodesic convexity of W, V, we know f is non-decreasing on $[0,1]$. So $f(1) \geq \liminf_{t\to 0^+} f(t)$. Integrating over $d\gamma(x_1,y_1)d\gamma(x_2,y_2)$ gives

$$
\begin{aligned}
\mathcal{E}(\nu)-\mathcal{E}(\mu) = & \int_{\Omega\times\Omega}\int_{\Omega\times\Omega} \frac{W(y_1-y_2)+2V(y_2)-W(x_1-x_2)-2V(x_2)}{2} d\gamma(x_1,y_1)d\gamma(x_2,y_2) \\
\geq & \int_{\Omega\times\Omega}\int_{\Omega\times\Omega} \langle \nabla W(x_2-x_1)+\nabla V(x_2), y_2-x_2\rangle\, d\gamma(x_1,y_1)d\gamma(x_2,y_2) + o\left(d_W\left(\mu,\nu\right)\right) \\
= & \int_{\Omega\times\Omega} \left\langle \int_{\Omega} \nabla W(x_2-x_1)d\mu(x_1)+\nabla V(x_2), y_2-x_2\right\rangle d\gamma(x_2,y_2) + o\left(d_W\left(\mu,\nu\right)\right) \\
= & \int_{\Omega\times\Omega} \langle \nabla W*\mu(x_2)+\nabla V(x_2), y_2-x_2\rangle\, d\gamma(x_2,y_2) + o\left(d_W\left(\mu,\nu\right)\right).
\end{aligned}
$$

Denote $v(t,x) = -\nabla W*\mu(t,x)-\nabla V(x)$, we notice that

$$
\begin{aligned}
\langle -v(t,x_2), y_2-x_2\rangle = & \langle -P_{x_2}\left(v(t,x_2)\right), y_2-x_2\rangle + \langle -v(t,x_2)+P_{x_2}\left(v(t,x_2)\right), y_2-x_2\rangle \\
\geq & \langle -P_{x_2}\left(v(t,x_2)\right), y_2-x_2\rangle - \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)}+\|\nabla V\|_{L^\infty(\Omega)}}{2\eta}|y_2-x_2|^2,
\end{aligned}
$$

and

$$
\int_{\Omega\times\Omega} |x_2-y_2|^2 d\gamma(x_2,y_2) = d_W^2(\mu,\nu).
$$

Thus

$$
\mathcal{E}(\nu)-\mathcal{E}(\mu) \geq \int_{\Omega\times\Omega} \langle -P_{x_2}\left(-\nabla W*\mu(x_2)-\nabla V(x_2)\right), y_2-x_2\rangle\, d\gamma(x_2,y_2) + o\left(d_W\left(\mu,\nu\right)\right),
$$

and by Definition 2.2.5

$$
-P\left(v(t)\right) = -P\left(-\nabla W*\mu(t)-\nabla V\right) \in \partial\mathcal{E}(\mu(t)). \tag{5.3.18}
$$

Take $\mu = \mu(t), \nu = \mu(t+h)$ and $\gamma_t^h \in \Gamma_o(\mu(t), \mu(t+h))$ then gives

$$\begin{aligned}
&\lim_{h\to 0^+} \frac{\mathcal{E}(\mu(t+h)) - \mathcal{E}(\mu(t))}{h} \\
&\geq \limsup_{h\to 0^+} \left(\int_{\Omega\times\Omega} \left\langle \nabla W * \mu(t, x_2) + \nabla V(x_2), \frac{y_2 - x_2}{h} \right\rangle d\gamma_t^h(x_2, y_2) + \frac{1}{h} o\left(d_W\left(\mu(t), \mu(t+h)\right)\right) \right) \\
&= \int_\Omega \left\langle \nabla W * \mu(t)(x_2) + \nabla V(x_2), \tilde{v}(t, x_2) \right\rangle d\mu(t, x_2),
\end{aligned}$$

where the last equality comes from (5.3.15). Similarly, by taking $\mu = \mu(t), \nu = \mu(t-h)$, we have

$$\lim_{h\to 0^+} \frac{\mathcal{E}(\mu(t)) - \mathcal{E}(\mu(t-h))}{h} \leq \int_\Omega \left\langle \nabla W * \mu(t)(x_2) + \nabla V(x_2), \tilde{v}(t, x_2) \right\rangle d\mu(t, x_2).$$

Together with the fact that $\mathcal{E}(\mu(\,\cdot\,)$ is locally absolutely continuous, we have for a.e. $t > 0$

$$\frac{d}{dt}\mathcal{E}(\mu(t)) = \int_\Omega \left\langle \nabla W * \mu(t)(x) + \nabla V(x), \tilde{v}(t, x) \right\rangle d\mu(t, x). \tag{5.3.19}$$

The claim is proved. Now for $v_N(t,x) = v(t,x) - P_x(v(t,x))$, we have $v_N(t,x) \in N(\Omega, x)$ and $\|v_N(t)\|_{L^\infty(\Omega)} \leq \|v(t)\|_{L^\infty(\Omega)} < \infty$. Thus

$$\begin{aligned}
\lim_{h\to 0^+} \int_{\Omega\times\Omega} \left\langle v_N(t,x), \frac{y-x}{h} \right\rangle d\gamma_t^h(x,y) &\leq \lim_{h\to 0^+} \int_{\Omega\times\Omega} \frac{\|v_N(t)\|_{L^\infty(\Omega)}}{2\eta} \frac{1}{h} |x-y|^2 d\gamma_t^h(x,y) \\
&\leq \lim_{h\to 0^+} \frac{\|v(t)\|_{L^\infty(\Omega)}}{2\eta} \frac{d_W^2\left(\mu(t), \mu(t+h)\right)}{h} \\
&= 0,
\end{aligned}$$

which together with the weak $L^2(\mu(t))$-convergence of $\int_\Omega \frac{y-\cdot}{h} d\nu.(y)$ implies

$$\int_\Omega \left\langle v_N(t,x), \tilde{v}(t,x) \right\rangle d\mu(t,x) \leq 0.$$

We then know that

$$\frac{d}{dt}\mathcal{E}(\mu(t)) \geq -\int_\Omega \left\langle P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right), \tilde{v}(t,x) \right\rangle d\mu(t,x). \tag{5.3.20}$$

Together with (5.3.8) and (5.3.18) we get for a.e $t > 0$

$$\tilde{v}(t,x) = P_x(v(t,x)) = P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right) \in -\partial\mathcal{E}(\mu(t)),$$

$$|\mu'|^2(t) = \int_\Omega \left|P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right)\right|^2 d\mu(t,x)$$

and for any $0 \leq s \leq t < \infty$

$$\mathcal{E}(\mu(s)) = \mathcal{E}(\mu(t)) + \int_s^t \int_\Omega \left|P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)\right|^2 d\mu(r,x) dr.$$

Thus $\mu(\,\cdot\,)$ is a gradient flow with respect to $\mathcal{E}$ and by (5.3.14), a weak measure solution to (5.0.1). $\square$

Remark 5.3.5. *In [31], Carrillo, Lisini and Mainini showed weak $L^2(\mu(t))$ convergence of $\int_\Omega \frac{y-\cdot}{h} d\nu_\cdot^h(y)$ to $\tilde{v}(t,\cdot)$ in a more general setting than ours.*

Remark 5.3.6. *We can actually show a stronger statement than (5.3.18), that for any $\mu \in \mathcal{P}_2(\Omega)$*

$$-P(v) = -P(-\nabla W * \mu - \nabla V) \in \partial^o \mathcal{E}(\mu), \tag{5.3.21}$$

where $\partial^o\mathcal{E}(\mu)$ denotes the unique minimal $L^2(\mu)$ element in $\partial\mathcal{E}(\mu)$. To show that, we recall the notion of local slope $|\partial\mathcal{E}|$ at $\mu \in \mathcal{P}_2(\Omega)$,

$$|\partial\mathcal{E}|(\mu) = \limsup_{\nu\to\mu} \frac{(\mathcal{E}(\mu) - \mathcal{E}(\nu))^+}{d_W(\mu,\nu)},$$

where $\nu \to \mu$ means $\nu \in \mathcal{P}_2(\Omega)$ approaches μ in d_W. It is straightforward computation to see that for any $\kappa \in \partial\mathcal{E}(\mu)$, $|\partial\mathcal{E}|(\mu) \le \|\kappa\|_{L^2(\mu)}$. Then to get (5.3.21), it is enough to show

$$\|P(-\nabla W * \mu - \nabla V)\|_{L^2(\mu)} \le |\partial\mathcal{E}|(\mu), \tag{5.3.22}$$

which we follow the argument given in [28, 119] to show. For any continuous, $L^2(\mu)$ integrable vector field ξ on Ω, let $X(t)$ be as given in Remark 5.2.7 and define $\mu(t) = (X(t,\cdot))_\sharp \mu$, then

$$\lim_{t\to0^+} \frac{\mathcal{E}(\mu(t)) - \mathcal{E}(\mu)}{t} = \int_\Omega \langle \nabla W * \mu(x) + \nabla V(x), P_x(\xi(x))\rangle \, d\mu(x), \tag{5.3.23}$$

and

$$\lim_{t\to0^+} \frac{d_W(\mu(t),\mu)}{t} \le \|P(\xi)\|_{L^2(\mu)}. \tag{5.3.24}$$

(5.3.24) is a direct consequence of the fact that $(\mathrm{Id}, X(t,\cdot))_\sharp\mu \in \Gamma(\mu,\mu(t))$, see Lemma 3.3.6 from Chapter 3. (5.3.23) comes from a similar argument given in step 2 of the proof of Theorem 3.3.4 from Chapter 3. Thus

$$\begin{aligned} -\int_\Omega \langle \nabla W * \mu(x) + \nabla V(x), P_x(\xi(x))\rangle \, d\mu(x) &= \lim_{t\to0^+} \frac{\mathcal{E}(\mu) - \mathcal{E}(\mu(t))}{t} \\ &= \lim_{t\to0^+} \frac{\mathcal{E}(\mu) - \mathcal{E}(\mu(t))}{d_W(\mu(t),\mu)} \frac{d_W(\mu(t),\mu)}{t} \\ &\le |\partial\mathcal{E}|(\mu)\|P(\xi)\|_{L^2(\mu)}. \end{aligned}$$

Take $\xi(x) = -\nabla W * \mu(x) - \nabla V(x)$, we get (5.3.22) as claimed. Notice that all the above arguments also works when Ω is bounded and η-prox-regular as well as when Ω is unbounded and convex.

A direct consequence of (5.3.21) is that, $\mu(\cdot)$ is a gradient flow with respect to $\mathcal{E}$ if and only if it is a weak measure solution to (5.0.1), given Ω is bounded, η-prox-regular or unbounded, convex. The observation is, if $\mu(\cdot)$ is a gradient flow with tangent velocity *field $\tilde{v}$, then $\tilde{v}(t) = -\partial^o \mathcal{E}(\mu(t))$ for a.e. $t > 0$, see Theorem 11.1.3 from [5]. Thus $\tilde{v}(t) = P(-\nabla W * \mu(t) - \nabla V)$ and $\mu(\,\cdot\,)$ is a weak measure solution to (5.0.1). If $\mu(\,\cdot\,)$ is a weak measure solution to (5.0.1) with tangent velocity $\tilde{v}$, then $\operatorname{div}(\mu(t)\,(P(v(t)) - \tilde{v}(t))) = 0$ for $v(t,x) = -\nabla W * \mu(t)(x) - \nabla V(x)$. By (5.3.19)*

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}(\mu(t)) &= \int_\Omega \langle \nabla W * \mu(t)(x) + \nabla V(x), \tilde{v}(t,x)\rangle \, d\mu(t,x) \\
&= \int_\Omega \langle \nabla W * \mu(t)(x) + \nabla V(x), P_x\left(v(t,x)\right)\rangle \, d\mu(t,x) \\
&= -\|P\left(v(t)\right)\|^2_{L^2(\mu(t))}.
\end{aligned}$$

This together with the fact $\frac{d}{dt}\mathcal{E}(\mu(t)) \geq -\frac{1}{2}|\partial\mathcal{E}|^2(\mu(t)) - \frac{1}{2}|\mu'|^2(t)$ imply $\|P(v(t))\|_{L^2(\mu)} = |\mu'|(t)$ for a.e. $t > 0$ and $P(v(t))$ is the tangent velocity field of $\mu(\,\cdot\,)$ with $P(v(t)) = -\partial^o\mathcal{E}(\mu(t))$ for a.e. $t > 0$. Thus $\mu(\,\cdot\,)$ is a gradient flow with respect to $\mathcal{E}$.

We turn to the proof of Theorem 5.1.6.

Proof of Theorem 5.1.6. We show (5.1.8) first. Let $\mu^1(\,\cdot\,), \mu^2(\,\cdot\,)$ be two solutions to (5.0.1), by Theorem 8.4.7 and Lemma 4.3.4 from [5], we have

$$\frac{d}{dt} d_W^2\left(\mu^1(t), \mu^2(t)\right) \leq 2\int_{\Omega\times\Omega} \langle P_x\left(v^1(t,x)\right) - P_y\left(v^2(t,y)\right), x - y\rangle d\gamma_t(x,y), \tag{5.3.25}$$

where $\gamma_t \in \Gamma_o\left(\mu^1(t), \mu^2(t)\right)$ and $v^i(t,x) = -\nabla W * \mu^i(t)(x) - \nabla V(x)$ for $i = 1, 2$. For v^i, by (A1) and (A2), similar argument as in the proof of Proposition 5.3.1 gives

$$\int_{\Omega\times\Omega} \langle v^1(t,x) - v^2(t,y), x - y\rangle d\gamma_t(x,y) \leq \left(-\lambda_W^- - \lambda_V\right) d_W^2\left(\mu^1(t), \mu^2(t)\right).$$

By the fact that Ω is η-prox-regular we have

$$\begin{aligned}
&\int_{\Omega\times\Omega} \langle P_x\left(v^1(t,x)\right) - v^1(t,x) - P_y\left(v^2(t,y)\right) + v^2(t,y), x - y\rangle d\gamma_t(x,y) \\
&\leq \frac{\|v^1(t)\|_{L^\infty(\Omega)} + \|v^2(t)\|_{L^\infty(\Omega)}}{2\eta} d_W^2\left(\mu^1(t), \mu^2(t)\right),
\end{aligned}$$

where v^i satisfies

$$\|v^i\|_{L^\infty(\Omega)} \leq \|\nabla W\|_{L^\infty(\Omega - \Omega)} + \|\nabla V\|_{L^\infty(\Omega)} < \infty.$$

Plugging back into (5.3.25) yields

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu^1(t), \mu^2(t)\right) &= \int_{\Omega\times\Omega} \langle P_x\left(v^1(t,x)\right) - P_y\left(v^2(t,y)\right), x - y\rangle d\gamma_t(x,y) \\
&= \int_{\Omega\times\Omega} \langle v^1(t,x) - v^2(t,y), x - y\rangle d\gamma_t(x,y) \\
&+ \int_{\Omega\times\Omega} \langle P_x\left(v^1(t,x)\right) - v^1(t,x) - P_y\left(v^2(t,y)\right) + v^2(t,y), x - y\rangle d\gamma_t(x,y) \\
&\leq \left(-\lambda_W^- - \lambda_V + \frac{\|\nabla W\|_{L^\infty(\Omega - \Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{\eta}\right) d_W^2\left(\mu^1(t), \mu^2(t)\right).
\end{aligned}$$

Then by Gronwall's inequality we have for all $t \geq 0$

$$d_W\left(\mu^1(t), \mu^2(t)\right) \leq \exp\left(\left(-\lambda_W^- - \lambda_V + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{\eta}\right) t\right) d_W\left(\mu_0^1, \mu_0^2\right). \tag{5.3.26}$$

(5.1.8) is proved. For (5.1.9), we have if $\mu(\,\cdot\,)$ is a weak measure solution to (5.0.1), then for any $\nu \in \mathcal{P}_2(\Omega)$ and $\gamma_t \in \Gamma_o(\mu(t), \nu)$

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu(t), \nu\right)\\
=&\int_{\Omega\times\Omega} \left\langle P_x\left(v(t,x)\right), x-y\right\rangle d\gamma_t(x,y)\\
=&\int_{\Omega\times\Omega} \left(\langle v(t,x), x-y\rangle + \langle P_x\left(v(t,x)\right) - v(t,x), x-y\rangle\right) d\gamma_t(x,y)\\
\leq&\mathcal{E}(\nu) - \mathcal{E}(\mu(t)) + \int_{\Omega\times\Omega}\left(-\frac{\lambda_W^-}{2} - \frac{\lambda_V}{2} + \frac{\|v(t)\|_{L^\infty(\Omega)}}{2\eta}\right)|x-y|^2 d\gamma_t(x,y)\\
\leq&\mathcal{E}(\nu) - \mathcal{E}(\mu(t)) + \left(-\frac{\lambda_W^-}{2} - \frac{\lambda_V}{2} + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{2\eta}\right) d_W^2\left(\mu(t), \nu\right).
\end{aligned}$$

On the other hand, if a locally absolutely continuous $\mu(\,\cdot\,)$ satisfies (5.1.9), then for any fixed $\nu \in \mathcal{P}_2(\Omega)$,

$$\begin{aligned}
&\int_{\Omega\times\Omega} \langle \tilde{v}(t,x), x-y\rangle d\gamma(x,y)\\
=&\frac{1}{2}\frac{d}{dt} d_W^2(\mu(t), \nu)\\
\leq& \mathcal{E}(\nu) - \mathcal{E}(\mu(t)) + \left(-\frac{\lambda_W^-}{2} - \frac{\lambda_V}{2} + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)} + \|\nabla V\|_{L^\infty(\Omega)}}{2\eta}\right) d_W^2\left(\mu(t), \nu\right),
\end{aligned}$$

where $\tilde{v}$ is the tangent velocity filed of $\mu(\,\cdot\,)$ at t and $\gamma \in \Gamma_o(\mu(t), \nu)$ is an optimal plan. By Definition 2.2.5,

$$\tilde{v}(t) \in -\partial\mathcal{E}(\mu(t))$$

for a.e. $t \geq 0$ and $\mu(\,\cdot\,)$ is a gradient flow to $\mathcal{E}$. Thus we know $\tilde{v}(t) = -\partial^o\mathcal{E}(\mu(t)) = P_x(v(t))$ and $\mu(\,\cdot\,)$ is a weak measure solution to (5.0.1). Thus the weak measure solution is characterized by the system of evolution variational inequalities (5.1.9). $\square$

5.4 Existence and stability of solutions with Ω unbounded: Global case

In this Section we prove the existence and stability of (5.0.1) with Ω unbounded, convex and W, V satisfying (GA1)-(GA4).

For any initial data $\mu_0 \in \mathcal{P}_2(\Omega)$ and fixed $x_0 \in \Omega$, denote $B_n(x_0) = \{x \in \mathbb{R}^d : |x - x_0| < n\}$, we can take $\mu_0^n = \sum_{i=1}^{k(n)} m_j^n \delta_{x_i^n}$ for $x_i^n \in \Omega \cap \overline{B_n(x_0)}$ and $\lim_{n\to\infty} d_W(\mu_0^n, \mu_0) = 0$. To see that, note $\int_\Omega |x - x_0|^2 d\mu_0(x) < \infty$, thus $\lim_{n\to\infty} \int_{\Omega \setminus B_n(x_0)} |x - x_0|^2 d\mu_0(x) = 0$. For $\mu_0 \lfloor_{\Omega \cap \overline{B_n(x_0)}}$, we can find $\tilde{\mu}_0^n$ composed of delta measures with the same total mass as $\mu_0 \lfloor_{\Omega \cap \overline{B_n(x_0)}}$, such that $\mathrm{supp}(\mu_0^n) \subset \Omega \cap \overline{B_n(x_0)}$ and $\lim_{n\to\infty} d_W\left(\mu_0 \lfloor_{\Omega \cap \overline{B_n(x_0)}}, \tilde{\mu}_0^n\right) = 0$. Then $\mu_0^n = \tilde{\mu}_0^n + \left(1 - \mu_0\left(\Omega \cap \overline{B_n(x_0)}\right)\right) \delta_{x_0}$ satisfies the required conditions. Without loss of generality, we assume that $x_0 = 0 \in \Omega$ and denote $B(n) = B_n(0)$.

As in Section 5.3, we first show the convergence of $\mu^n(\,\cdot\,)$ as $n \to \infty$.

Proposition 5.4.1. *Assume that Ω is unbounded and convex, W, V satisfy (GA1)-(GA4). Then for two solutions $\mu^m(\,\cdot\,), \mu^n(\,\cdot\,)$ to the discrete system with different initial data μ_0^m, μ_0^n, we have for all $t \geq 0$*

$$d_W\left(\mu^n(t), \mu^m(t)\right) \leq \exp\left(-\left(\lambda_W^- + \lambda_V\right) t\right) d_W\left(\mu_0^n, \mu_0^m\right). \tag{5.4.1}$$

Proof. The proof is similar to the proof of Proposition 5.3.1 once we notice that since Ω is ∞-prox-regular, by (5.1.4) for any $x, y \in \Omega$

$$\langle P_x\left(v^n(t,x)\right) - v^n(t,x), x - y\rangle \leq 0.$$

$\square$

So as $n \to \infty$ we again know that $\mu^n(t)$ converges to some $\mu(t) \in \left(\mathcal{P}_2\left(\Omega\right), d_W\right)$ for all $t \geq 0$. Before proving that the limit curve $\mu(\,\cdot\,)$ is a curve of maximal slope, we need the following proposition.

Proposition 5.4.2. *Let $\mu_n, \mu \in \mathcal{P}_2(\Omega)$ be such that $\lim_{n\to\infty} d_W\left(\mu_n, \mu\right) = 0$ then*

$$\lim_{n\to\infty} \mathcal{V}(\mu_n) = \mathcal{V}(\mu),$$

and

$$\lim_{n\to\infty} \mathcal{W}(\mu_n) = \mathcal{W}(\mu).$$

Proof. Since the arguments are similar, it is enough for us to show the property for $\mathcal{V}$. By (GA4), there exists a constant $C > 0$ such that $|V(x)| \leq C(1 + |x|^2)$. By Lemma 5.1.7 from [5], since $V(x) + C|x|^2$ is lower semicontinuous and bounded from below,

$$\liminf_{n\to\infty} \int_\Omega \left(V(x) + C|x|^2\right) d\mu_n(x) \geq \int_\Omega \left(V(x) + C|x|^2\right) d\mu(x).$$

$\lim_{n\to\infty} d_W\left(\mu_n, \mu\right) = 0$, we know

$$\lim_{n\to\infty} \int_\Omega |x|^2 d\mu_n(x) = \int_\Omega |x|^2 d\mu(x).$$

Thus

$$\liminf_{n\to\infty}\int_\Omega V(x)d\mu_n(x)\geq\int_\Omega V(x)d\mu(x).$$

Similarly, the condition $C|x|^2 - V(x)$ is lower semicontinuous and bounded from below implies

$$\limsup_{n\to\infty}\int_\Omega V(x)d\mu_n(x)\leq\int_\Omega V(x)d\mu(x).$$

Thus

$$\lim_{n\to\infty}\mathcal{V}(\mu_n)=\mathcal{V}(\mu),$$

as claimed. □

We estimate the growth of support of the solutions $\mu^n(\,\cdot\,)$ to (5.1.16).

Lemma 5.4.3. *Let μ_0^n be the approximation of μ_0 such that* $\operatorname{supp}(\mu_0^n)\subset\Omega\cap B(n)$. *Then* $\operatorname{supp}(\mu^n(t))\subset\Omega\cap B(r(t))$ *for $r(t)\leq(n+1)\exp(Ct)$ for some $C=C(W,V)$ independent of n.*

Proof. Define $r(t)=\sup_i|x_i^n(t)|$. For fixed $t>0$, assume that $x_i^n(t)$ realizes $R(t)$ i.e., $r(t)=|x_i^n(t)|$, then

$$\begin{aligned}
\left|\frac{d}{dt}|x_i^n|^2\right| &= 2\left|\left\langle x_i^n(t), P_{x_i^n}\left(-\sum_{j=1}^{k(n)}m_j\nabla W(x_i^n-x_j^n)-\nabla V(x_i^n)\right)\right\rangle\right|\\
&\leq 2|x_i^n(t)|\left(\sum_{j=1}^{k(n)}m_j\left|\nabla W(x_i^n(t)-x_j^n(t))\right|+|\nabla V(x_i^n(t))|\right)\\
&\leq 2|x_i^n(t)|\left(\sum_{j=1}^{k(n)}m_jC\left(1+|x_i^n(t)+|x_j^n(t)|\right)+C\left(1+|x_i^n(t)|\right)\right)\\
&\leq C\left(1+|x_i^n(t)|^2\right).
\end{aligned}$$

Thus

$$r(t)\leq r(0)\exp(Ct)+\exp(Ct)-1$$

for $r(0)\leq n$ and C depending only on W,V, in particular independent of the number of particles $k(n)$. □

We can now show

Theorem 5.4.4. *Assume Ω is unbounded and convex, W,V satisfy (GA1)-(GA4), then $\mu(\,\cdot\,)$ satisfies for any $0\leq s<t<\infty$*

$$\mathcal{E}(\mu(s))\geq\mathcal{E}(\mu(t))+\frac{1}{2}\int_s^t|\mu'|^2(r)dr+\frac{1}{2}\int_s^t\int_\Omega|P_x(v(r,x))|^2d\mu(r,x)dr, \tag{5.4.2}$$

where $v(r,x)=-\int_\Omega\nabla W(x-y)d\mu(r,y)-\nabla V(x)$.

Proof. We first check that for fixed $n \in \mathbb{N}$, the function $t \mapsto \mathcal{E}(\mu^n(t))$ is locally absolutely continuous. For fixed $0 \leq s < t < \infty$, by Lemma 5.4.3, $\|\nabla V(x)\|_{L^\infty(\Omega \cap B(r(t)))} < \infty$ and $\|\nabla W\|_{L^\infty(\Omega\cap B(r(t)) - \Omega \cap B(r(t))} < \infty$. Then by the same argument as in (5.3.10), $t \mapsto \mathcal{E}(\mu(t))$ is locally absolutely continuous. Together with Proposition 5.4.2, the proof is now identical to the proof of Theorem 5.3.11. We omit it here. $\square$

We proceed to the proof of Theorem 5.1.9

Proof of Theorem 5.1.9. Let $\tilde{v}$ be the tangential velocity field for $\mu(\cdot)$, i.e. $\mu(\cdot)$ satisfies (5.3.14) and $\|\tilde{v}(t)\|_{L^2(\mu(t))} = |\mu'|(t)$. Similar arguments as in the proof of Theorem 5.1.5 still gives that for any $\mu, \nu \in \mathcal{P}_2(\Omega)$

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \geq \int_{\Omega\times\Omega} \langle \nabla W * \mu(x_2) + \nabla V(x_2), y_2 - x_2 \rangle \, d\gamma(x_2, y_2) + o\left(d_W(\mu, \nu)\right,$$

and for a.e. $t > 0$

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right) = \int_\Omega \langle \nabla W * \mu(t)(x) + \nabla V(x), \tilde{v}(t,x) \rangle \, d\mu(t,x).$$

Now since Ω is convex, we have $\langle v_N(t,x), y - x \rangle \leq 0$, thus

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \geq \int_{\Omega\times\Omega} \langle -P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right), y - x \rangle \, d\gamma(x,y),$$

and

$$\lim_{h\to 0^+} \int_{\Omega\times\Omega} \left\langle v_N(t,x), \frac{y-x}{h} \right\rangle d\gamma_t^h(x,y) \leq 0.$$

So we have $-P\left(v(t)\right) = -P\left(-\nabla W * \mu(t) - \nabla V\right) \in \partial\mathcal{E}\left(\mu(t)\right)$ and

$$\int_\Omega \langle v_N(t,x), \tilde{v}(t,x) \rangle d\mu(t,x) \leq 0.$$

Thus

$$\frac{d}{dt}\mathcal{E}\left(\mu(t)\right) \geq -\int_\Omega \langle P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right), \tilde{v}(t,x) \rangle \, d\mu(t,x),$$

which together with Theorem 5.5.1 implies for a.e. $t > 0$

$$\tilde{v}(t,x) = P_x\left(v(t,x)\right) = P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right) \in -\partial\mathcal{E}\left(\mu(t)\right), \tag{5.4.3}$$

$$|\mu'|^2(t) = \int_\Omega \left|P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right)\right|^2 d\mu(t,x) \tag{5.4.4}$$

and for any $0 \leq s \leq t < \infty$

$$\mathcal{E}(\mu(s)) = \mathcal{E}(\mu(t)) + \int_s^t \int_\Omega \left|P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)\right|^2 d\mu(r,x)dr. \tag{5.4.5}$$

Thus $\mu(\,\cdot\,)$ is a gradient flow with respect to $\mathcal{E}$ and by (5.3.14), a weak measure solution to (5.0.1).

For the stability result (5.1.13), we only need to notice that for any two solutions $\mu^1(\,\cdot\,), \mu^2(\,\cdot\,)$ to (5.0.1), since Ω is convex, $\left\langle v^i(t,x) - P_x\left(v^i(t,x)\right), y-x\right\rangle \leq 0$ for $i = 1,2$. Thus

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu^1(t), \mu^2(t)\right) &\leq \int_{\Omega\times\Omega} \langle P_x\left(v^1(t,x)\right) - P_y\left(v^2(t,y)\right), x-y\rangle d\gamma_t(x,y) \\
&= \int_{\Omega\times\Omega} \langle v^1(t,x) - v^2(t,y), x-y\rangle d\gamma_t(x,y) \\
&+ \int_{\Omega\times\Omega} \langle P_x\left(v^1(t,x)\right) - v^1(t,x) - P_y\left(v^2(t,y)\right) + v^2(t,y), x-y\rangle d\gamma_t(x,y) \\
&\leq -\left(\lambda_W^- + \lambda_V\right) d_W^2\left(\mu^1(t), \mu^2(t)\right).
\end{aligned}$$

Then by Gronwall's inequality we get (5.1.13).

For evolution variational inequalities (5.1.14), if $\mu(\,\cdot\,)$ is a solution to (5.0.1) then for any $\nu \in \mathcal{P}_2(\Omega)$ and $\gamma \in \Gamma_o\left(\mu(t), \nu\right)$ an optimal plan

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu(t), \nu\right) &= \int_{\Omega\times\Omega} \langle P_x\left(v(t,x)\right), x-y\rangle\, d\gamma_t(x,y) \\
&= \int_{\Omega\times\Omega} \left(\langle v(t,x), x-y\rangle + \langle P_x\left(v(t,x)\right) - v(t,x), x-y\rangle\right) d\gamma_t(x,y) \\
&\leq \mathcal{E}(\nu) - \mathcal{E}(\mu(t)) - \int_{\Omega\times\Omega} \left(\frac{\lambda_W^-}{2} + \frac{\lambda_V}{2}\right) |x-y|^2 d\gamma_t(x,y) \\
&\leq \mathcal{E}(\nu) - \mathcal{E}(\mu(t)) - \left(\frac{\lambda_W^-}{2} + \frac{\lambda_V}{2}\right) d_W^2\left(\mu(t), \nu\right).
\end{aligned}$$

Similar argument as in the proof of Theorem 5.1.6 shows that if $\mu(\,\cdot\,)$ satisfies (5.1.14), then it is a gradient flow to $\mathcal{E}$ and a weak measure solution to (5.0.1). Thus the weak measure solution to (5.0.1) is characterized by the system of evolution variational inequalities (5.1.14). □

5.5 Existence and stability of solutions with Ω unbounded: Compactly supported initial data case

In this Section, we show the existence and stability results in the case when Ω is unbounded and W, V satisfy (LA1)-(LA4). The novelty is that λ-geodesic convexity of energy is only assumed locally (which is automatically satisfied if V and W are C^2 functions).

We start by giving the control the support of the solutions $\mu^n(t)$ to (5.1.16). Notice that when approximating μ_0 by $\mu_0^n = \sum_{i=1}^{k(n)} m_i^n \delta_{x_i^n}$, since $\mathrm{supp}(\mu_0) \subset \Omega \cap B(r_0)$, we can take $x_i^n \in \Omega \cap B(r_0+1)$ for all $n \in \mathbb{N}$ and $1 \leq i \leq k(n)$ such that we have

$$\lim_{n\to\infty} d_W\left(\mu_0^n, \mu_0\right) = 0.$$

So without loss of generality, we assume $\mathrm{supp}\,(\mu_0^n) \subset B(r_0)$ for all $n \in \mathbb{N}$. Then by Lemma 5.4.3, $\mathrm{supp}\,(\mu^n(t)) \subset \Omega \cap B(r(t))$ for $r(t) \leq (r_0 + 1)\exp(Ct)$ for some $C = C(W, V)$ independent of n.

Proposition 5.5.1. *There exists a locally absolutely continuous curve $\mu(\cdot)$ in $\mathcal{P}_2(\Omega)$ such that $\mu^n(t)$ converges to $\mu(t)$ in $\mathcal{P}_2(\Omega)$ for any $0 \leq t < \infty$.*

Proof. For any fixed $0 < T < \infty$ and any $0 \leq t \leq T$, we know that $\mathrm{supp}(\mu^n(t)) \subset B(r(T))$ for all $0 \leq t \leq T$ uniformly in n. Let K_k and $\lambda_{W,k}, \lambda_{V,k}$ be the sequences of compact convex sets and constants such that W, V are $\lambda_{W,k}$ and $\lambda_{V,k}$-geodesically convex on K_k. Take k_0 be such that $B(2r(T)) \subset K_k$ for all $k \geq k_0$. Still denote $\gamma_t \in \Gamma_o\left(\mu^n(t), \mu^m(t)\right)$ an optimal plan. Now notice that $\mathrm{supp}(\mu^n(t)), \mathrm{supp}(\mu^m(t)) \subset B(r(t)) \cap \Omega \subset K_k \cap \Omega = \Omega_k$, thus

$$\int_{\Omega\times\Omega} \langle v^n(t,x) - v^m(t,y), x - y\rangle \, d\gamma_t(x,y) \leq \int_{\Omega\times\Omega} \left(\lambda_{W,k}^- + \lambda_{V,k}\right) |x-y|^2,$$

and

$$\begin{aligned}
&\int_{\Omega\times\Omega} \langle P_x\left(v^n(t,x)\right) - v^n(t,x) - P_y\left(v^m(t,y)\right) + v^m(t,y), x-y\rangle \, d\gamma_t(x,y) \\
&\leq \int_{\Omega\times\Omega} \left(\frac{\|v^n(t)\|_{L^\infty(\Omega_t)} + \|v^m(t)\|_{L^\infty(\Omega_t)}}{2\eta}\right) |x-y|^2 d\gamma_t(x,y),
\end{aligned}$$

where $\Omega_t = \Omega \cap B(r(t))$. Since $v^n(t,x) = -\int_\Omega \nabla W(x-y) d\mu^n(t,y) - \nabla V(x)$ we know

$$\|v^n(t)\|_{L^\infty(\Omega_t)} \leq \|\nabla W\|_{L^\infty(\Omega_T - \Omega_T)} + \|\nabla V\|_{L^\infty(\Omega_T)} < \infty.$$

Thus as in the proof of Proposition 5.3.1, we have for $0 \leq t \leq T$

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu^n(t), \mu^m(t)\right) \\
&\leq \int_\Omega \langle P_x\left(v^n(t,x)\right) - P_y\left(v^m(t,y)\right), x-y\rangle \, d\gamma_t(x,y) \\
&= \int_{\Omega\times\Omega} \langle v^n(t,x) - v^m(t,y), x-y\rangle \, d\gamma_t(x,y) \\
&\quad + \int_{\Omega\times\Omega} \langle P_x\left(v^n(t,x)\right) - v^n(t,x) - P_y\left(v^m(t,y)\right) + v^m(t,y), x-y\rangle \, d\gamma_t(x,y) \\
&\leq \left(-\lambda_{W,k}^- - \lambda_{V,k} + \frac{\|\nabla W\|_{L^\infty(\Omega_T - \Omega_T)} + \|\nabla V\|_{L^\infty(\Omega_T)}}{\eta}\right) d_W^2\left(\mu^n(t), \mu^m(t)\right).
\end{aligned}$$

By Gronwall's inequality, we have for all $0 \leq t \leq T$

$$d_W\left(\mu^n(t), \mu^m(t)\right) \leq e^{\left(-\lambda_{W,k}^- - \lambda_{V,k} + \frac{\|\nabla W\|_{L^\infty(\Omega_T - \Omega_T)} + \|\nabla V\|_{L^\infty(\Omega_T)}}{\eta}\right)t} d_W(\mu_0^n, \mu_0^m).$$

Thus as $n \to \infty$, $\mu^n(t)$ converges in $\mathcal{P}_2(\Omega)$ to some $\mu(t)$. $\square$

Theorem 5.5.2. *$\mu(\cdot)$ is a curve of maximal slope, for any $0 \le s < t < \infty$*

$$\mathcal{E}\left(\mu(s)\right) \ge \mathcal{E}\left(\mu(t)\right) + \frac{1}{2}\int_s^t |\mu'|^2(r)dr + \frac{1}{2}\int_s^t \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x)dr, \tag{5.5.1}$$

where $v(r,x) = -\int_\Omega \nabla W(x-y)d\mu(r,y) - \nabla V(x)$.

Proof. We use similar argument as in Theorem 5.3.3 and Theorem 5.4.4. For any fixed $n \in \mathbb{N}$, since $\mathrm{supp}(\mu^n(t)) \subset \Omega \cap B(r(t))$, we can still control the L^∞-norm of ∇V and ∇W. Then the same argument as in the proof of Theorem 5.4.4 shows that $t \mapsto \mathcal{E}\left(\mu(t)\right)$ is locally absolutely continuous. Thus the fact that μ^n are solutions to the discrete systems implies,

$$\mathcal{E}\left(\mu^n(s)\right) \ge \mathcal{E}\left(\mu^n(t)\right) + \frac{1}{2}\int_s^t |\left(\mu^n\right)'|^2(r)dr + \frac{1}{2}\int_s^t \int_\Omega |P_x\left(v^n(r,x)\right)|^2 d\mu^n(r,x)dr. \tag{5.5.2}$$

Note that $W, V \in C^1(\mathbb{R}^d)$ and $\lim_{n\to\infty} d_W\left(\mu^n(r), \mu(r)\right) = 0$ with $\mathrm{supp}(\mu^n(r)) \subset \Omega \cap B(r(T)$ for any $0 \le r < t \le T$, we get

$$\lim_{n\to\infty} \mathcal{E}(\mu^n(r)) = \mathcal{E}(\mu(r)).$$

By Lemma 5.3.4 and notice that $\nabla W * \mu^n(r) + \nabla V$ still converges weakly to $\nabla W * \mu(r) + \nabla V$ for any $0 \le r \le T$, then

$$\liminf_{n\to\infty} \int_\Omega |P_x\left(v^n(r,x)\right)|^2 d\mu^n(r,x) \ge \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x).$$

By Fatou's lemma,

$$\liminf_{n\to\infty} \int_s^t \int_\Omega |P_x\left(v^n(r,x)\right)|^2 d\mu^n(r,x)dr \ge \int_s^t \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x)dr.$$

Now by the same argument as in the proof of (5.3.13), we again obtain

$$\liminf_{n\to\infty} \int_s^t |\left(\mu^n\right)'|^2(r)dr \ge \int_s^t |\mu'|^2(r)dr.$$

Take $n \to \infty$ in (5.5.2) gives

$$\mathcal{E}\left(\mu(s)\right) \ge \mathcal{E}\left(\mu(t)\right) + \frac{1}{2}\int_s^t |\mu'|^2(r)dr + \frac{1}{2}\int_s^t \int_\Omega |P_x\left(v(r,x)\right)|^2 d\mu(r,x)dr.$$

□

We now start to prove Theorem 5.1.10

Proof of Theorem 5.1.10. Since $\mu(\cdot)$ is locally absolutely continuous, we know that there exists a unique Borel vector field $\tilde{v}$ such that

$$\partial_t \mu(t) + \mathrm{div}\left(\mu(t)\tilde{v}(t)\right) = 0$$

holds in the sense of distributions. For a fixed $T > 0$ and any $\mu, \nu \in \mathcal{P}_2(\Omega)$ with $\mathrm{supp}(\mu) \subset B(r(T))$, $\mathrm{supp}(\nu) \subset B(r(T))$, let $\gamma \in \Gamma_o(\mu, \nu)$. Since W, V are $\lambda_{W,k}$ and $\lambda_{V,k}$-geodesically convex on $K_k \supset B(r(t)) \cap \Omega$, we have that the function f we defined in (5.3.17) by taking $\lambda = \lambda_k$ is non-decreasing in t for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}\,\gamma$. Thus we still have

$$\mathcal{E}(\nu) - \mathcal{E}(\mu) \geq \int_{\Omega\times\Omega} \langle \nabla W * \mu(x_2) + \nabla V(x_2), y_2 - x_2 \rangle d\gamma(x_2, y_2).$$

For any $0 < t < T$, and $h > 0$ such that $t - h \geq 0, t + h \leq T$, we take $\mu = \mu(t), \nu = \mu(t+h)$ to get

$$\lim_{h\to 0^+} \frac{\mathcal{E}(\mu(t+h)) - \mathcal{E}(\mu(t))}{h} \geq \int_\Omega \langle \nabla W * \mu(t)(x) + \nabla V(x), \tilde{v}(t,x) \rangle d\mu(t,x)$$

Again take $\mu = \mu(t), \nu = \mu(t-h)$ gives

$$\lim_{h\to 0^+} \frac{\mathcal{E}(\mu(t)) - \mathcal{E}(\mu(t-h))}{h} \leq \int_\Omega \langle \nabla W * \mu(t)(x) + \nabla V(x), \tilde{v}(t,x) d\mu(t,x).$$

Also $\mathcal{E}(\mu(t))$ is locally absolutely continuous, so for a.e. $t > 0$

$$\frac{d}{dt}\mathcal{E}(\mu(t)) = \int_\Omega \langle \nabla W * \mu(t)(x) + \nabla V(x), \tilde{v}(t,x) \rangle d\mu(t,x),$$

which again implies

$$\frac{d}{dt}\mathcal{E}(\mu(t)) \geq -\int_\Omega \langle P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right), \tilde{v}(t,x) \rangle d\mu(t,x).$$

Combine with (5.5.1) yields

$$\tilde{v}(t,x) = P_x\left(-\nabla W * \mu(t)(x) - \nabla V(x)\right),$$

and for any $0 \leq s \leq t < \infty$

$$\mathcal{E}(\mu(s)) = \mathcal{E}(\mu(t)) + \int_s^t \int_\Omega \left| P_x\left(-\nabla W * \mu(r)(x) - \nabla V(x)\right)\right|^2 d\mu(r,x) dr.$$

For the contraction property (5.1.15), we notice that for any $0 \leq t \leq T < \infty$ and $k \in \mathbb{N}$ such that $B(r(T)) \subset K_k$

$$\frac{1}{2}\frac{d}{dt} d_W^2\left(\mu^1(t), \mu^2(t)\right) \leq \left(-\lambda_{W,k}^- - \lambda_{V,k} + \frac{\|\nabla W\|_{L^\infty(\Omega_k - \Omega_k)} + \|\nabla V\|_{L^\infty(\Omega_k)}}{\eta}\right) d_W^2\left(\mu^1(t), \mu^2(t)\right).$$

where $\Omega_k = \Omega \cap K_k$. Thus by Gronwall's inequality, we have for all $0 \leq t \leq T$

$$d_W\left(\mu^1(t), \mu^2(t)\right) \leq \exp\left(\left(-\lambda_{W,k}^- - \lambda_{V,k} + \frac{\|\nabla W\|_{L^\infty(\Omega_k - \Omega_k)} + \|\nabla V\|_{L^\infty(\Omega_k)}}{\eta}\right) t\right) d_W(\mu_0^1, \mu_0^2).$$

□

Remark 5.5.3. *When the external and interaction potentials are time-dependent* $V = V(t,x), W = W(t,x)$*, then with some modifications of the arguments we have before, we can still show the existence and stability results of the solutions to (5.0.1) in all the three different cases as in the time-independent settings before. For example, we assume that there are constants* $\lambda \in \mathbb{R}, \eta > 0$ *and a positive function* $\beta \in L^1([0,\infty))$ *such that*

(M1) Ω *is bounded and* η*-prox-regular.*

(TA1) $W \in C^1([0,\infty) \times \mathbb{R}^d)$ *is* λ*-geodesically convex on* $\operatorname{Conv}(\Omega - \Omega)$ *uniformly in* t*.*

(TA2) $V \in C^1([0,\infty) \times \mathbb{R}^d)$ *is* λ*-geodesically convex on* $\operatorname{Conv}(\Omega)$ *uniformly in* t*.*

(TA3) $|\nabla V(t,x)| \leq \beta(t)(1+|x|)$ *and* $|\nabla W(t,x)| \leq \beta(t)(1+|x|)$ *for all* $x \in \mathbb{R}^d$*.*

(TA4) $|\frac{\partial V}{\partial t}(t,x)| \leq \beta(t)(1+|x|^2)$ *and* $|\frac{\partial W}{\partial t}(t,x)| \leq \beta(t)(1+|x|^2)$ *for all* $x \in \mathbb{R}^d$*.*

Then we can show the existence of a weak measure solution $\mu(\,\cdot\,)$ *to (5.0.1) satisfying (5.1.6), (5.1.7) and stability estimate*

$$d_W(\mu^1(t), \mu^2(t)) \leq \exp\left(-2\lambda t + \frac{C(\Omega)}{\eta}\int_0^t \beta(s)ds\right) d_W\left(\mu_0^1, \mu_0^2\right), \tag{5.5.3}$$

where $C(\Omega) = \sup_{x\in\Omega} \operatorname{dist}(x,0)$*. We sketch the proof and concentrate on the differences. Approximate the initial data* μ_0 *by a sequence of particle measures* μ_0^n *as before. Note that we can still show the existence of solutions to the projected ODE system by citing Theorem 5.1 from [44]. Thus for total energy defined as* $\mathcal{E}(t,\mu) = \frac{1}{2}\int_{\Omega\times\Omega} W(t,x-y)d\mu(x)d\mu(y) + \int_\Omega V(t,x)d\mu(x)$*, we have the following energy dissipation along the solutions* $\mu^n(\,\cdot\,)$*,*

$$\begin{aligned}
&\mathcal{E}(s,\mu^n(s)) \\
\geq &\mathcal{E}(t,\mu^n(t)) - \frac{1}{2}\int_s^t \int_{\Omega\times\Omega} \frac{\partial W}{\partial r}(r,x-y)d\mu(r,x)d\mu(r,y)dr \\
&- \int_s^t \int_\Omega \frac{\partial V}{\partial r}(r,x)d\mu(r,x)dr + \frac{1}{2}\int_s^t \left(\left|(\mu^n)'\right|^2(r) + \int_\Omega |P_x(v^n(r,x))|^2 d\mu^n(r,x)\right)dr.
\end{aligned} \tag{5.5.4}$$

Similar stability argument as before shows that the sequence $\{\mu^n(\,\cdot\,)\}_n$ *satisfies the stability estimate (5.5.3). Thus we know* $\mu^n(\,\cdot\,)$ *converges in* d_W *to a locally absolutely curve* $\mu(\,\cdot\,)$ *and* $\mu(\,\cdot\,)$ *satisfies the same energy dissipation (5.5.4) by similar lower semicontinuity arguments. By the* λ*-geodesic convexity and* C^1 *regularity of* W *and* V*, we can then show the following chain rule for* $\mu(\,\cdot\,)$*,*

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}(t,\mu(t)) \geq &\frac{1}{2}\int_{\Omega\times\Omega} \frac{\partial W}{\partial t}(t,x-y)d\mu(t,x)d\mu(t,y) + \int_\Omega \frac{\partial V}{\partial t}(t,x)d\mu(t,x) \\
&- \int_\Omega \langle P_x(v(t,x)), \tilde{v}(t,x)\rangle d\mu(t,x).
\end{aligned} \tag{5.5.5}$$

Combining (5.5.4) with (5.5.5), we show that $\mu(\cdot)$ is a weak measure solution to (5.0.1) satisfying (5.1.6) and (5.1.7). Then (5.5.3) comes from the stability argument of the time-independent setting.

5.6 Aggregation on nonconvex domains

In this Section, we consider the following question: what are the conditions on Ω to ensure the existence of an interaction potential W such that the solution $\mu(\cdot)$ to (5.0.1) aggregates to a singleton (a single delta mass with mass 1) as time goes to infinity?

Let Ω be bounded and η-prox-regular, $V \equiv 0$ and W satisfy (A1) for some $\lambda_W > 0$, such that Theorem 5.1.5 holds and we have a weak measure solution $\mu(\cdot)$ to (5.0.1). We denote $\Xi = \{\delta_x : x \in \mathbb{R}^d\}$ the set of singletons, and start to estimate the evolution of $d_W(\mu(\cdot), \Xi)$, the distance of $\mu(\cdot)$ to Ξ, i.e. we prove Proposition 5.1.7.

Proof of Proposition 5.1.7. It suffices to show that for all $t \geq 0$

$$\frac{1}{2}\frac{d^+}{dt} d_W^2\left(\mu(t), \Xi\right) \leq \left(-\lambda_W + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)}}{2\eta}\right) d_W^2\left(\mu(t), \Xi\right)$$

since then by Gronwall's inequality the result follows.

By shifting time we can assume that $t = 0$. Denote the center of mass for μ_0 by $\bar{x}$, that is $\bar{x} = \int_\Omega x d\mu(0, x)$. It is direct computation to show that $d_W(\mu(0), \Xi) = d_W\left(\mu(0), \delta_{\bar{x}}\right)$, and for any $t > 0$, $d_W(\mu(t), \Xi) \leq d_W\left(\mu(t), \delta_{\bar{x}}\right)$. Thus

$$\begin{aligned}
\frac{1}{2}\frac{d^+}{dt}\bigg|_{t=0} d_W^2\left(\mu(t), \Xi\right) &\leq \frac{1}{2}\frac{d^+}{dt}\bigg|_{t=0} d_W^2\left(\mu(t), \delta_{\bar{x}}\right) \\
&= \int_\Omega \left\langle P_x\left(v(0,x)\right), x - \bar{x}\right\rangle d\mu(0,x) \\
&= \int_\Omega \left(\left\langle v(0,x), x - \bar{x}\right\rangle + \left\langle P_x\left(v(0,x)\right) - v(0,x), x - \bar{x}\right\rangle\right) d\mu(0,x).
\end{aligned}$$

Now we follow similar argument as in the proof of Proposition 5.3.1. To be precise, by (5.3.5) with $\mu^n(t) = \mu(t), \mu^m(t) \equiv \delta_{\bar{x}}$, we have

$$\begin{aligned}
\int_\Omega \left\langle v(0,x), x - \bar{x}\right\rangle d\mu(0,x) &\leq -\frac{\lambda_W}{2}\int_{\Omega\times\Omega} |x-y|^2 d\mu(0,x) d\mu(0,y) \\
&= -\lambda_W \int_\Omega |x - \bar{x}|^2 d\mu(0,x) \\
&= -\lambda_W d_W^2\left(\mu(0), \delta_{\bar{x}}\right),
\end{aligned}$$

where we used the fact that $\int_\Omega (x - \bar{x}) d\mu(0,x) = 0$ for the definition of center of mass.

Also by (5.3.6) with $\mu^n(t) = \mu(t), \mu^m(t) \equiv \delta_{\bar{x}}$,

$$\int_{\Omega\times\Omega} \langle P_x\left(v(0,x)\right) - v(0,x), x - \bar{x}\rangle d\mu(0,x) \leq \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)}}{2\eta} d_W^2\left(\mu(0), \delta_{\bar{x}}\right).$$

Combine the estimates yields

$$\left.\frac{1}{2}\frac{d^+}{dt}\right|_{t=0} d_W^2\left(\mu(t),\Xi\right) \le \left(-\lambda_W + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)}}{2\eta}\right) d_W^2\left(\mu(0),\delta_{\bar{x}}\right)$$
$$= \left(-\lambda_W + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)}}{2\eta}\right) d_W^2\left(\mu(0),\Xi\right).$$

□

We can now prove Theorem 5.1.8.

Proof. It turns out that the quadratic interaction potential leads to the sharpe bound for general domains. Furthermore, since multiplying a potential by a positive constant only leads to a constant rescaling in time of the dynamics, we consider $W(x) = \frac{1}{2}|x|^2$. To verify the inequality (5.1.11) note that $\nabla W(x) = x, \operatorname{Hess} W(x) = \mathrm{Id}$ and $\lambda_W = 1$. Thus $\sup_{\Omega-\Omega}|\nabla W| \le \sup_{x,y\in\Omega}|x-y| = \operatorname{diam}(\Omega)$ and

$$-\lambda_W + \frac{\|\nabla W\|_{L^\infty(\Omega-\Omega)}}{2\eta} \le -1 + \frac{1}{2\eta}\operatorname{diam}(\Omega) =: C(\Omega) < 0$$

which via inequality (5.1.11) implies the desired result. □

Remark 5.6.1. *We notice that (5.1.11) implies that $\lim_{t\to\infty} d_W\left(\mu(t),\delta_{\bar{x}(t)}\right) = 0$ where $\bar{x}(t) = \int_\Omega x d\mu(t,x)$ is the center of mass for $\mu(t)$. Hence as $t\to\infty$, $\mu(t)$ converges in d_W to a singleton, i.e., all mass aggregates to one point to form a delta mass of size 1. Thus Theorem 5.1.8 gives a sufficient condition on the shape of the domain Ω on which there exists a radially symmetric interaction potential W so that solutions aggregate to a point. We note that the simple condition given in the theorem is also sharp in the following sense: for any $\varepsilon > 0$ there exists Ω bounded and η-prox-regular with $0 < \eta \le (\frac{1}{2}-\epsilon)\operatorname{diam}(\Omega)$, and an initial condition μ_0 such that the solution starting from μ_0 does not aggregate to a point.*

Let $\Omega = \{(r\cos\theta, r\sin\theta) \in \mathbb{R}^2 : 1-\epsilon \le r \le 1, -\epsilon \le \theta \le \pi+\epsilon\}$ for $0<\epsilon<\frac{1}{2}$ be as shown in Figure 5.3. Let $x^1 = \left(-(1-\epsilon)\cos\epsilon, -(1-\epsilon)\sin\epsilon\right), x^2 = \left((1-\epsilon)\cos\epsilon, -(1-\epsilon)\sin\epsilon\right)$ and set $\mu_0 = \frac{1}{2}\delta_{x^1} + \frac{1}{2}\delta_{x^2}$. Then Ω is η-prox-regular with $\eta = |x^1-x^2|/2 > 1-2\epsilon$. Since $\operatorname{diam}(\Omega) = 2$, *thus* $(\frac{1}{2}-2\varepsilon)\operatorname{diam}(\Omega) < \eta < \frac{1}{2}\operatorname{diam}(\Omega)$. *For any radially symmetric W which satisfies (A1) for some $\lambda_W > 0$, a direct calculation yields that $v(0,x^1) = -\frac{1}{2}\nabla W(x^1-x^2) \in N(\Omega,x^1)$. Thus $P_{x^1}\left(v(0,x^1)\right) = 0$ and similarly $P_{x^2}\left(v(0,x^2)\right) = 0$. We then see that $\mu(t) \equiv \mu_0$ is the solution to (5.0.1), and hence there is no aggregation to a singleton, (5.1.12) does not hold.*

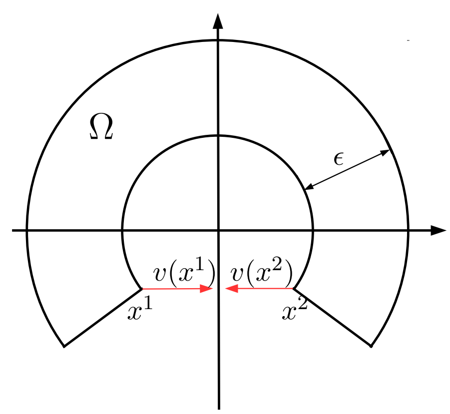


Figure 5.3: The velocity v at x^1 and x^2 are shown as the red arrows, which lie in the normal cones of the points respectively.

Bibliography

[1] D. Alexander, I. Kim, and Y. Yao, *Quasi-static evolution and congested crowd transport*, Nonlinearity, 27 (2014), pp. 823–858.

[2] L. Ambrosio, *Minimizing movements*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5), 19 (1995), pp. 191–246.

[3] ——, *Lecture notes on optimal transport problems*, in Mathematical aspects of evolving interfaces (Funchal, 2000), vol. 1812 of Lecture Notes in Math., Springer, Berlin, 2003, pp. 1–52.

[4] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, in Modelling and Optimisation of Flows on Networks, M. R. Benedetto Piccoli, ed., Lecture Notes in Mathematics, Springer Berlin Heidelberg, Heidelberg, 2013.

[5] 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 ed., 2008.

[6] ——, *Heat flow and calculus on metric measure spaces with Ricci curvature bounded below—the compact case*, Boll. Unione Mat. Ital. (9), 5 (2012), pp. 575–629.

[7] Y. Au Yeung, G. Friesecke, and B. Schmidt, *Minimizing atomic configurations of short range pair potentials in two dimensions: crystallization in the Wulff shape*, Calc. Var. Partial Differential Equations, 44 (2012), pp. 81–100.

[8] D. Balagué, J. A. Carrillo, T. Laurent, and G. Raoul, *Dimensionality of local minimizers of the interaction energy*, Arch. Ration. Mech. Anal., 209 (2013), pp. 1055–1088.

[9] D. Balagué, J. A. Carrillo, and Y. Yao, *Confinement for repulsive-attractive kernels*, Discrete Contin. Dyn. Syst. Ser. B, 19 (2014), pp. 1227–1248.

[10] R. W. Beard, J. Lawton, and F. Y. Hadaegh, *A coordination architecture for space- craft formation control*, IEEE Transactions on Control Systems Technology,, 9 (2001), pp. 777–790.

[11] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[12] D. Benedetto, E. Caglioti, and M. Pulvirenti, *A kinetic equation for granular media*, RAIRO Modél. Math. Anal. Numér., 31 (1997), pp. 615–641.

[13] A. J. Bernoff and C. M. Topaz, *A primer of swarm equilibria*, SIAM J. Appl. Dyn. Syst., 10 (2011), pp. 212–250.

[14] A. L. Bertozzi and J. Brandman, *Finite-time blow-up of L^∞-weak solutions of an aggregation equation*, Commun. Math. Sci., 8 (2010), pp. 45–65.

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

[16] A. L. Bertozzi, J. B. Garnett, and T. Laurent, *Characterization of radially symmetric finite time blowup in multidimensional aggregation equations*, SIAM J. Math. Anal., 44 (2012), pp. 651–681.

[17] A. L. Bertozzi and T. Laurent, *Finite-time blow-up of solutions of an aggregation equation in $\mathbf{R}^n$*, Comm. Math. Phys., 274 (2007), pp. 717–735.

[18] A. L. Bertozzi, T. Laurent, and J. Rosado, *L^p theory for the multidimensional aggregation equation*, Comm. Pure Appl. Math., 64 (2011), pp. 45–83.

[19] 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 (2008), pp. 691–721.

[20] A. Blanchet, E. A. Carlen, and J. A. Carrillo, *Functional inequalities, thick tails and asymptotics for the critical mass Patlak-Keller-Segel model*, J. Funct. Anal., 262 (2012), pp. 2142–2230.

[21] A. Blanchet, J. Carrillo, D. Kinderlehrer, M. Kowalczyk, P. Laurençot, and S. Lisini, *A hybrid variational principle for the keller-segel system in $\mathbb{R}^2$*. Preprint, 2014.

[22] 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 (2009), pp. 133–168.

[23] A. Blanchet and P. Laurençot, *The parabolic-parabolic Keller-Segel system with critical diffusion as a gradient flow in* $\mathbb{R}^d$, $d \geq 3$, Comm. Partial Differential Equations, 38 (2013), pp. 658–686.

[24] S. Boi, V. Capasso, and D. Morale, *Modeling the aggregative behavior of ants of the species polyergus rufescens*, Nonlinear Anal. Real World Appl., 1 (2000), pp. 163–176. Spatial heterogeneity in ecological models (Alcalá de Henares, 1998).

[25] M. Bounkhel, *Regularity concepts in nonsmooth analysis*, vol. 59 of Springer Optimization and Its Applications, Springer, New York, 2012. Theory and applications.

[26] H. Brézis, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, in Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971), Academic Press, New York, 1971, pp. 101–156.

[27] M. Burger and M. Di Francesco, *Large time behavior of nonlocal aggregation models with nonlinear diffusion*, Netw. Heterog. Media, 3 (2008), pp. 749–785.

[28] J. A. Carrillo, M. Di Francesco, 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 (2011), pp. 229–271.

[29] ——, *Confinement in nonlocal interaction equations*, Nonlinear Anal., 75 (2012), pp. 550–558.

[30] J. A. Carrillo, Y. Huang, and S. Martin, *Nonlinear stability of flock solutions in second-order swarming models.* Preprint, 2013.

[31] J. A. Carrillo, S. Lisini, and E. Mainini, *Gradient flows for non-smooth interaction potentials.* To appear in Nonlinear Analysis TMA, 2013.

[32] 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 (2006), pp. 217–263.

[33] J. A. Carrillo, D. Slepčev, and L. Wu, *Nonlocal interaction equations on uniform prox-regular sets.* Submitted, 2014.

[34] Y. L. Chuang, M. R. D'Orsogna, D. Marthaler, L. S. Chayes, and A. L. Bertozzi, *State transitions and the continuum limit for interacting, self-propelled particles*, Phys. D, 232 (2007), pp. 33–47.

[35] F. H. Clarke, R. J. Stern, and P. R. Wolenski, *Proximal smoothness and the lower-C^2 property*, J. Convex Anal., 2 (1995), pp. 117–144.

[36] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math., 146 (2001), pp. 219–257.

[37] F. Cucker and S. Smale, *Emergent behavior in flocks*, IEEE Trans. Automat. Control, 52 (2007), pp. 852–862.

[38] ——, *On the mathematics of emergence*, Jpn. J. Math., 2 (2007), pp. 197–227.

[39] S. Daneri and G. Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal., 40 (2008), pp. 1104–1122.

[40] E. De Giorgi, *New problems on minimizing movements*, in Boundary value problems for partial differential equations and applications, vol. 29 of RMA Res. Notes Appl. Math., Masson, Paris, 1993, pp. 81–98.

[41] M. R. D'Orsogna, Y. L. Chuang, A. L. Bertozzi, and L. S. Chayes, *Self-propelled particles with soft-core interactions: Patterns, stability, and collapse*, Phys. Rev. Lett., 96 (2006), p. 104302.

[42] W. E and D. Li, *On the crystallization of 2D hexagonal lattices*, Comm. Math. Phys., 286 (2009), pp. 1099–1140.

[43] J. F. Edmond and L. Thibault, *Relaxation of an optimal control problem involving a perturbed sweeping process*, Math. Program., 104 (2005), pp. 347–373.

[44] ——, *BV solutions of nonconvex sweeping process differential inclusion with perturbation*, J. Differential Equations, 226 (2006), pp. 135–179.

[45] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Ann. Inst. Henri Poincaré Probab. Stat., 46 (2010), pp. 1–23.

[46] M. Erbar and J. Maas, *Ricci curvature of finite Markov chains via convexity of the entropy*, Arch. Ration. Mech. Anal., 206 (2012), pp. 997–1038.

[47] ——, *Gradient flow structures for discrete porous medium equations*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 1355–1374.

[48] A. Fathi and A. Figalli, *Optimal transportation on non-compact manifolds*, Israel J. Math., 175 (2010), pp. 1–59.

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

[50] ——, *Stability of stationary states of non-local equations with singular interaction potentials*, Math. Comput. Modelling, 53 (2011), pp. 1436–1450.

[51] R. C. Fetecau, Y. Huang, and T. Kolokolnikov, *Swarm dynamics and equilibria for a nonlocal aggregation model*, Nonlinearity, 24 (2011), pp. 2681–2716.

[52] A. Figalli, Y.-H. Kim, and R. McCann, *Regularity of optimal transport maps on multiple products of spheres*, J. Eur. Math. Soc. (JEMS), 15 (2013), pp. 1131–1166.

[53] I. Fonseca and G. Leoni, *Modern methods in the calculus of variations: L^p spaces*, Springer Monographs in Mathematics, Springer, New York, 2007.

[54] W. Gangbo and R. J. McCann, *Optimal maps in Monge's mass transport problem*, C. R. Acad. Sci. Paris Sér. I Math., 321 (1995), pp. 1653–1658.

[55] U. Gianazza, G. Savaré, and G. Toscani, *The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation*, Arch. Ration. Mech. Anal., 194 (2009), pp. 133–220.

[56] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations, 39 (2010), pp. 101–120.

[57] N. Gigli, K. Kuwada, and S.-I. Ohta, *Heat flow on Alexandrov spaces*, Comm. Pure Appl. Math., 66 (2013), pp. 307–331.

[58] N. Gigli and J. Maas, *Gromov-Hausdorff convergence of discrete transportation metrics*, SIAM J. Math. Anal., 45 (2013), pp. 879–899.

[59] A. Glitzky and A. Mielke, *A gradient structure for systems coupling reaction-diffusion effects in bulk and interfaces*, Z. Angew. Math. Phys., 64 (2013), pp. 29–52.

[60] S.-Y. Ha and J.-G. Liu, *A simple proof of the Cucker-Smale flocking dynamics and mean-field limit*, Commun. Math. Sci., 7 (2009), pp. 297–325.

[61] S.-Y. Ha and E. Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinet. Relat. Models, 1 (2008), pp. 415–435.

[62] Y. Huang and A. L. Bertozzi, *Self-similar blowup solutions to an aggregation equation in $\mathbf{R}^n$*, SIAM J. Appl. Math., 70 (2010), pp. 2582–2603.

[63] ——, *Asymptotics of blowup solutions for the aggregation equation*, Discrete Contin. Dyn. Syst. Ser. B, 17 (2012), pp. 1309–1331.

[64] Y. Huang, T. P. Witelski, and A. L. Bertozzi, *Anomalous exponents of self-similar blow-up solutions to an aggregation equation in odd dimensions*, Appl. Math. Lett., 25 (2012), pp. 2317–2321.

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

[66] L. V. Kantorovich, *On a problem of Monge*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 312 (2004), pp. 15–16.

[67] ——, *On mass transportation*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 312 (2004), pp. 11–14.

[68] E. Keller and L. Segel, *Initiation of slime mold aggregation viewed as an instability*, J. Theor. Biol., 26 (1970), pp. 399–415.

[69] T. Kolokolnikov, J. A. Carrillo, A. Bertozzi, R. Fetecau, and M. Lewis, *Emergent behaviour in multi-particle systems with non-local interactions [Editorial]*, Phys. D, 260 (2013), pp. 1–4.

[70] T. Kolokolnikov, H. Sun, D. Uminsky, and A. L. Bertozzi, *Stability of ring patterns arising from two-dimensional particle interactions*, Phys. Rev. E, 84 (2011), p. 015203.

[71] N. E. Leonard and E. Fiorelli, *Virtual leaders, artificial potentials and coordinated control of groups*, in Proceedings of the 40th IEEE Conference on Decision and Control, vol. 5, IEEE, Dec. 2001, pp. 2969–2973.

[72] A. J. Leverentz, C. M. Topaz, and A. J. Bernoff, *Asymptotic dynamics of attractive-repulsive swarms*, SIAM J. Appl. Dyn. Syst., 8 (2009), pp. 880–908.

[73] H. Levine, W.-J. Rappel, and I. Cohen, *Self-organization in systems of self-propelled particles*, Phys. Rev. E, 63 (2000), p. 017101.

[74] M. Liero and A. Mielke, *Gradient structures and geodesic convexity for reaction-diffusion systems*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 371 (2013), pp. 20120346, 28.

[75] S. Lisini, *Nonlinear diffusion equations with variable coefficients as gradient flows in Wasserstein spaces*, ESAIM Control Optim. Calc. Var., 15 (2009), pp. 712–740.

[76] S. Lisini, D. Matthes, and G. Savaré, *Cahn-Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics*, J. Differential Equations, 253 (2012), pp. 814–850.

[77] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2), 169 (2009), pp. 903–991.

[78] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal., 261 (2011), pp. 2250–2292.

[79] D. Matthes, R. J. McCann, and G. Savaré, *A family of nonlinear fourth order equations of gradient flow type*, Comm. Partial Differential Equations, 34 (2009), pp. 1352–1397.

[80] B. Maury, A. Roudneff-Chupin, and F. Santambrogio, *A macroscopic crowd motion model of gradient flow type*, Math. Models Methods Appl. Sci., 20 (2010), pp. 1787–1821.

[81] B. Maury, A. Roudneff-Chupin, F. Santambrogio, and J. Venel, *Handling congestion in crowd motion modeling*, Netw. Heterog. Media, 6 (2011), pp. 485–519.

[82] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal., 11 (2001), pp. 589–608.

[83] A. Mielke, *Geodesic convexity of the relative entropy in reversible Markov chains*, Calc. Var. Partial Differential Equations, 48 (2013), pp. 1–31.

[84] A. Mogilner and L. Edelstein-Keshet, *A non-local model for a swarm*, J. Math. Biol., 38 (1999), pp. 534–570.

[85] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, in Histoire de l'Académie Royale des Sciences de Paris, 1781, pp. 666–704.

[86] D. Morale, V. Capasso, and K. Oelschläger, *An interacting particle system modelling aggregation behavior: from individuals to populations*, J. Math. Biol., 50 (2005), pp. 49–66.

[87] J.-J. Moreau, *Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires*, C. R. Acad. Sci. Paris, 255 (1962), pp. 238–240.

[88] S. Motsch and E. Tadmor, *A new model for self-organized dynamics and its flocking behavior*, J. Stat. Phys., 144 (2011), pp. 923–947.

[89] ——, *Heterophilious dynamics enhances consensus*. Preprint, 2013.

[90] S.-i. Ohta and K.-T. Sturm, *Non-contraction of heat flow on Minkowski spaces*, Arch. Ration. Mech. Anal., 204 (2012), pp. 917–944.

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

[92] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal., 37 (2005), pp. 1227–1255 (electronic).

[93] C. S. Patlak, *Random walk with persistence and external bias*, Bull. Math. Biophys., 15 (1953), pp. 311–338.

[94] P. Petersen, *Riemannian geometry*, vol. 171 of Graduate Texts in Mathematics, Springer, New York, second ed., 2006.

[95] R. A. Poliquin, R. T. Rockafellar, and L. Thibault, *Local differentiability of distance functions*, Trans. Amer. Math. Soc., 352 (2000), pp. 5231–5249.

[96] ——, *Local differentiability of distance functions*, Trans. Amer. Math. Soc., 352 (2000), pp. 5231–5249.

[97] A. Pratelli, *On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation*, Ann. Inst. H. Poincaré Probab. Statist., 43 (2007), pp. 1–13.

[98] G. Raoul, *Non-local interaction equations: Stationary states and stability analysis.* Preprint, 2011.

[99] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970.

[100] D. Ruelle, *Statistical mechanics: Rigorous results*, W. A. Benjamin, Inc., New York-Amsterdam, 1969.

[101] K.-T. Sturm, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, J. Math. Pures Appl. (9), 84 (2005), pp. 149–168.

[102] ——, *On the geometry of metric measure spaces. I*, Acta Math., 196 (2006), pp. 65–131.

[103] ——, *On the geometry of metric measure spaces. II*, Acta Math., 196 (2006), pp. 133–177.

[104] A. Sütő, *Ground state at high density*, Comm. Math. Phys., 305 (2011), pp. 657–710.

[105] E. Tadmor and C. Tan, *Critical thresholds in flocking hydrodynamics with non-local alignment*, Phil. Trans. R. Soc. A, 372 (2014).

[106] F. Theil, *A proof of crystallization in two dimensions*, Comm. Math. Phys., 262 (2006), pp. 209–236.

[107] C. M. Topaz, A. J. Bernoff, S. Logan, and W. Toolson, *A model for rolling swarms of locusts*, Eur. Phys. J. Special Topics, 157 (2008), pp. 93–109.

[108] C. M. Topaz and A. L. Bertozzi, *Swarming patterns in a two-dimensional kinematic model for biological groups*, SIAM J. Appl. Math., 65 (2004), pp. 152–174.

[109] C. M. Topaz, A. L. Bertozzi, and M. A. Lewis, *A nonlocal continuum model for biological aggregation*, Bull. Math. Biol., 68 (2006), pp. 1601–1623.

[110] C. M. Topaz, M. R. D'Orsogna, L. Edelstein-Keshet, and A. J. Bernoff, *Locust dynamics: behavioral phase change and swarming*, PLoS Comput. Biol., 8 (2012), pp. e1002642, 11.

[111] G. Toscani, *One-dimensional kinetic models of granular flows*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 1277–1291.

[112] J. Venel, *A numerical scheme for a class of sweeping processes*, Numer. Math., 118 (2011), pp. 367–400.

[113] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, *Novel type of phase transition in a system of self-driven particles*, Phys. Rev. Lett., 75 (1995), pp. 1226–1229.

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

[115] ——, *Optimal transport*, vol. 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009. Old and new.

[116] J. von Brecht, T. Kolokolnikov, A. L. Bertozzi, and H. Sun, *Swarming on random graphs*, J. Stat. Phys., 151 (2013), pp. 150–173.

[117] J. von Brecht, D. Uminsky, T. Kolokolnikov, and A. L. Bertozzi, *Predicting pattern formation in particle interactions*, Math. Models Methods Appl. Sci., 22 (2012), pp. 1140002, 31.

[118] L. Wu and D. Slepčev, *Nonlocal interaction equations in environments with heterogeneities and boundaries: compactly supported initial data case*. In preparation, 2014.

[119] ——, *Nonlocal interaction equations in environments with heterogeneities and boundaries*, Comm. Partial Differential Equations, (2015). To appear.