NNT : 2019SACLS112

THÈSE DE DOCTORAT

de

l'UNIVERSITÉ PARIS-SACLAY

École doctorale de mathématiques Hadamard (EDMH, ED 574)

Établissement d'inscription : Université Paris-Sud
Laboratoire d'accueil : Laboratoire de mathématiques d'Orsay, UMR 8628 CNRS

Spécialité de doctorat : Mathématiques fondamentales

Hugo LAVENANT

Courbes et applications optimales à valeurs dans l'espace de Wasserstein

Date de soutenance : 24 mai 2019

Après avis des rapporteurs :

PIERRE CARDALIAGUET (Université Paris-Dauphine)
GIUSEPPE SAVARÉ (Università di Pavia)

Jury de soutenance :

YANN BRENIER (CNRS & ÉNS de Paris)	Président du jury
PIERRE CARDALIAGUET (Université Paris-Dauphine)	Rapporteur
QUENTIN MÉRIGOT (Université Paris-Sud)	Examinateur
FILIPPO SANTAMBROGIO (Université Claude Bernard Lyon 1)	Directeur de thèse
KARL-THEODOR STURM (Universität Bonn)	Examinateur
DANIELA TONON (Université Paris-Dauphine)	Examinatrice

Remerciements

Mes premiers remerciements vont à mon directeur de thèse Filippo Santambrogio. C'est sous sa direction que j'ai découvert le monde de la recherche, et j'en suis très honoré. Merci pour la confiance que tu m'as accordée, ton soutien permanent et tes nombreux conseils. Discuter avec toi a toujours été un plaisir, ta tenacité et ta curiosité mathématique resteront un modèle pour moi.

A significant part of my research was done in interaction with Justin Solomon and his team. I warmly thank him for his enthusiasm. Being overwhelmed by his endless ideas and suggestions was always truly stimulating, and a great pleasure.

Je remercie sincèrement Pierre Cardaliaguet et Giuseppe Savaré pour l'attention qu'ils ont portée à mes travaux en rapportant cette thèse ; ainsi que Yann Brenier, Quentin Mérigot, Karl-Theodor Sturm et Daniela Tonon qui m'ont fait l'honneur d'être membres de mon jury.

J'ai une pensée pour tout·e·s les enseignant·e·s que j'ai eu la chance d'avoir au cours de ma scolarité, je les remercie de m'avoir transmis le goût de la réflexion et de m'avoir amené jusqu'ici.

Ces trois années de thèse furent riches en rencontres et en émotions. Je remercie sincèrement toutes les personnes qui ont rendu cette expérience si agréable : mes camarades doctorant·e·s ; toute l'équipe administrative du LMO ; et plus généralement celles et ceux que j'ai pu croiser, à Orsay ou ailleurs, avec qui j'ai eu la chance de parler de mathématiques et de bien d'autres choses.

Je remercie mes ami·e·s, qui ont toujours été là et qui savent tout ce que je leur dois. Enfin, j'ai une pensée toute particulière pour ma famille, mes parents, mes frères, ma cousine préférée, pour me supporter tous les jours et croire en moi.

Contents

Chapter 1

Introduction

This work is concerned with variational problems where the unknowns are curves or mappings valued in the Wasserstein space. The so-called Wasserstein space[1] is the space of distributions of mass over a fixed domain whose total mass is fixed. Putting by convention a total mass equal to 1, this space can be identified with the one of probability measures. It is endowed with the Wasserstein distance: such a distance measures how much it costs to transport mass from one configuration to another. Hence, it is not surprising that this space arises naturally when one aims at modeling phenomena dealing with the evolution (*via* transport) of a configuration of mass, with the constraint that the total amount of mass is preserved.

A curve valued in the Wasserstein space should be thought as the evolution in time of a distribution of mass: a crowd of people, a herd of sheeps, an assembly of particles (from molecules to stars), etc. We will be interested in boundary value problems, i.e. problems where the configuration of mass at the initial and final temporal horizons are prescribed, or at least penalized. From the physical point of view, it would amount to look at instanciations of the *least action principle*; on the other hand if one looks at a crowd of rational agents (like people), it amounts to assume that the agents try to anticipate the consequences of theirs actions. With the help of the metric structure on the Wasserstein space, one can define the *action* of a given curve, which measures how costly it is for the motion of the curve to occur. Minimizing the action with fixed boundary values leads to the geodesic problem in the Wassertein space, which is now a well understood concept: it reveals the optimal way to transport the mass from its initial to its final configuration. We will add congestion terms that have, in general, the effect of penalizing the configurations where the distribution of mass is too concentrated. In a crowd of people, it comes from the aversion (and the physical impossibility) for people to be too close to each other; in a fluid it is what stands for incompressbility. The effect of this penalization is the appearance of *pressure* forces and, of course, the spreading of the mass. The particles can have a natural dynamic that tends to concentrate them: for instance, people might all want to go to the same place, while physical particles, submitted to the force generated by a potential, have a tendency to move to the minimum of this potential. From a mathematical point of view, our study will focus on the regularity of the solutions of such problems where there is an interplay between optimal evolution, penalization of congestion, but also favor of congestion through natural dynamic of the particles. Broadly speaking, these results can be classified as *elliptic regularity*: we will show that solutions of some (convex) problems in calculus of variations exhibit more regularity than what expected *a priori*.

[1] We are aware that this denomination is controversial. Even though we will stick to the most common usage, as the reader can see in the next section, this distance could also be associated to the names of Monge, Lévy, Fréchet, Kantorovich, or Rubinstein.

A natural extension of curves are mappings depending on more than one parameter (whereas a curve is a mapping depending only on one parameter, namely time). A mapping valued in the Wasserstein space can be thought as a distribution of mass depending on several parameters. Although evoked briefly later in this introduction, the link between this mappings and applications or modeling of actual phenomena remains tiny and would need to be explored. The generalization of the geodesic problem is straightforward and leads to the concept of harmonic mappings valued in the Wasserstein space. The Dirichlet energy of a mapping, which represents heuristically the integral of the square of the magnitude (measured with the Wasserstein distance) of the gradient of a mapping is a natural extension of the *action* of a curve. Minimizing the Dirichlet energy with fixed values on the boundary of the source domain, i.e. the *Dirichlet problem*, leads to minimizers that we call *harmonic mappings valued in the Wasserstein space*. Following the work of Otto, the Wasserstein space can be seen formally as an infinite dimensional Riemannian manifold whose sectional curvature is positive. The theory of harmonic mappings valued in Riemannian manifold and metric spaces of negative curvature is now well understood, but to the best of our knowledge, our work is one of the first one to consider harmonic mappings valued in infinite dimensional Riemannian manifolds of positive curvature, though our technique are very specific to the Wasserstein space and could hardly be generalized to other spaces. Our contribution is a sound a thorough mathematical study of the Dirichlet energy and the Dirichlet problem in the Wasserstein space, the proof of a maximum principle (more specifically a Ishihara-type property) in this setting, a specific study of the case where the mappings are valued in a family of elliptically contoured distributions, and a discretization of the problem leading to an algorithm to compute (approximation of) these harmonic mappings.

In the rest of this introduction, we present a brief overview of the history of the optimal transport theory, focusing on its link with the distances on the space of probability measures (this part is can be skipped without impacting the comprehension of the rest). Then we describe the kind of variational problems in the Wassertein space we are interested with. On a toy model, we exhibit the key estimate which is at the basis of most of the results of this work, and conclude by an overview of the manuscript.

1.1 Optimal transport and Wasserstein distances: a brief historical survey

The birth of optimal transport is usually dated back to the *Mémoire* of Monge [Mon81] published in 1781 where he formulated the problem: if one wants to move a configuration of mass from one place to another, such that the cost of moving mass is proportional to the mass and the distance traveled, what is the most efficient way to do it? Monge was a *geometer* and its main interest was about the geometric characterization of the solution when the dimension of the ambient space is 2 or 3. He gave a partial answer in terms of congruence of lines, developable surfaces, etc. We refer to [Ghy12] for a detailed account of the *mémoire* of Monge and the subsequent work by Dupin, Appel and others with the same geometric focus. However, the metric point of view on the optimal transport problem started only in the 20th century.

Before going further, let us fix some notations. From the modern point of view of calculus of variations, the optimal transport problem can be stated as follows. Take $c(x,y)$ a function describing the cost of moving mass from x to y, and two configurations of mass $\mu(x)\mathrm{d}x$, $\nu(y)\mathrm{d}y$ sharing the same total mass. We want to find the coupling $\gamma(x,y)\mathrm{d}x\mathrm{d}y$, which describes the amount of mass sent from x onto y, such that γ actually transports μ onto ν (meaning that $\int \gamma(x,y)\mathrm{d}y = \mu(x)$ and $\int \gamma(x,y)\mathrm{d}x = \nu(x)$) and minimizes the total cost $\iint c(x,y)\gamma(x,y)\mathrm{d}x\mathrm{d}y$. In short, it can be written

$$\min_{\gamma} \left\{ \iint c(x,y)\gamma(x,y)\mathrm{d}x\mathrm{d}y \ : \ \gamma \geqslant 0, \ \int \gamma(x,y)\mathrm{d}y = \mu(x) \text{ and } \int \gamma(x,y)\mathrm{d}x = \nu(x) \right\}. \tag{1.1}$$

Usually, μ and ν are normalized so that their total mass is 1, i.e. they are seen as probability distributions. From the point of view of probability theory, the goal is to find the joint law of (X,Y) such that the law of X is μ, the law of Y is ν, which minimizes the cost $\mathbb{E}[c(X,Y)]$. This problem has three inputs (the configurations of mass μ, ν and the cost c) and two outputs (the optimal value of the problem and the optimizer γ).

Distances on the space of probability measures: early history If the cost c is fixed and one is only interested in the value of the problem (1.1), then in some specific situations it provides a *distance* on the space of probability measures over a fixed metric space which we will call the Wassertein distance. We mention that the formalization of the concept of distance and of metric spaces (*espaces distanciés*) dates back to the work of Fréchet at the beginning of the 20th century. Around that time the French school of mathematics, under the impulsion of Poincaré and Borel, started to study actively the calculus of probability from a mathematical point of view, and made the link with measure theory and the Lebesgue integral, developed shortly before.

In 1925, in his book *Calcul des probabilités* [Lév25, p. 199-200], Paul Lévy introduced a distance between probability distributions over the real line for technical reasons, in order to handle an approximation process. This distance was not the Wasserstein distance but was rather inspired by geometric considerations about the cumulative distribution functions of the laws. In a note that he wrote in a book for Fréchet (see [Fré50, p. 331-337], first published in 1935), he proposed different notions of distances between probability distributions among which one can read the Wasserstein distance. This distance is seen as way to lift a distance between random variables into a distance between laws of random variable, but he wrote that such a distance lacks from explicit expression. An explicit expression of the Wassertein distance in the case where the ambient space is one-dimensional, which amounts to say that the optimal coupling γ is the increasing one, was given by Fréchet in 1957 [Fré57], though it is possible that the solution (whose proof is not so involved) was found before.

We mention that few years before, in 1948, Fréchet [Fré48] proposed a definition of random variables valued in metric spaces. The main issue was the definition of a mean (*une position typique*) and the topology on the space of such random variables. At least in the optimal transport community, this work is now mainly known for its definition of a (Fréchet) barycenter in metric spaces.

As the reader can see, as soon as the notions of metric spaces and probability distributions were settled, the interplay between them, which is what the Wasserstein distance is about, has naturally been a subject of interest for researchers. However, the Wasserstein distance is not *any* distance on the space of probability measures and features a lot of additional properties explored later in the 20th century.

Economical interpretation The celebrated article *On the translocation of masses* of Kantorovich was published in 1942 (see [Kan58] for the english translation) and introduced what is considered as the modern formulation of the optimal transport problem, namely (1.1). In this article, he introduced the dual problem associated to it, which reads as the maximization of

$$\max_{\varphi,\psi} \left\{ \int \varphi(x)\mu(x)\mathrm{d}x + \int \psi(y)\nu(y)\mathrm{d}y \ : \ \varphi(x) + \psi(y) \leqslant c(x,y) \right\} \tag{1.2}$$

He showed that with this dual problem one can give a necessary and sufficient condition to characterize the optimal γ. It is only a few years later, in [Kan48] that he made the link with the problem phrased by Monge. The problem (1.1) is now usually called the *Monge-Kantorovich* problem, while the equivalence between this problem and its dual (1.2) is the *Kantorovich* duality.

This formulation shows that solving the optimal transport problem is a *linear programming* problem: minimization of a linear function under linear equalities and inequalities constraints. Actually, the formulation of the problem as a linear program coincides with the developpement of the linear programming theory, which happened after the second world war, in connection with military and industrial interests [Dan83]. In some sense, the first instance of a problem really thought as a linear programming one (namely (1.1), though the work of Kantorovich was not known in the West until the end of the 1950s), was set in the infinite dimensional framework. As mentioned by Dantzig, this linear programming structure, including the power of duality, also appeared in the work of Von Neumann and Morgenstern [MVN53] in game theory (published also around the end of the second world war).

In view of the applications, the optimal transport problem is seen as a planning problem, more specifically an assignment problem: for instance μ represents a distribution of workers, ν a distribution of tasks, and one wants to find the optimal way to assign each worker to a specific task. The cost $c(x, y)$ would be the efficiency of worker x when doing task y, and one would rather try to maximize the total efficiency. The economical interpretations of optimal transport are completely out of the scope of this work, for a modern reference we refer to [Gal16]. From the applied point of view, the great achievement of the linear programming approach was the conception of scalable and efficient algorithms as the simplex algorithm. Indeed from a combinatorial point of view the assignment problem is untractable as soon as the number of workers is larger than a few dozens, but if the problem has a linear programming structure then it becomes scalable with e.g. the simplex algorithm.

A flexible tool It is often said, sometimes even written[2] in the optimal transport community (especially the one working on quadratic optimal transport) that (almost) nothing happened between the work of Kantorovich and the one of Brenier [Bre87]. We would like to moderate this assertion.

Making sense of linear programming problems and proving duality results in the most general settings has been a research program conducted after the work of Kantorovich. One can look for instance at the survey by Rachev [Rac85] or the book by Rachev and Rüschendorf [RR98]. Quite quickly, it has been noticed, if one chooses the distance over the underlying space as the cost function, that the optimal value of (1.1) defines a distance on the space of probability measures over a given metric space, distance which in fact comes from a norm: it is now what is called the 1-Wasserstein distance W_1. In the work of Kantorovich and co-authors, the first occurence seems to be [KR58].

This 1-Wasserstein has revealed itself to be a great and flexible tool to study the space of probability distributions. In the article [Was69] by Wasserstein, published in 1969, the Wasserstein distance is used as a technical tool to study Markov processes[3]. In the 70s, the 1-Wasserstein distance is used (sometimes rediscovered) to tackle different problems: identifying the dynamical systems which are isomorphic to Bernoulli shifts by characterizing the rates

[2]We want to avoid to put the blame on anyone, hence the absence of citations to back up this claim.

[3]It seems that it is Dobrushin who introduced the terminology Wasserstein distance (written Vasershtein in [Dob70]). A different name may have been chosen at some point, but the article [JKO98], which showed the relevance of the 2-Wasserstein metric, followed this terminology. All the subsequent works on the topic stick to this denomination, and this manuscript makes no exception.

of decrease of correlations [Orn74]; constructing random fields with prescribed distributions [Dob70] or proving the mean field limit in kinetic theory [Dob79] to cite some examples. Here the 1-Wasserstein distance appeared to be the most relevant and the most easily manipulated (because of its dual formulation) metric on the space of probability distributions.

As pointed out by Vershik [Ver06], being rediscovered and used by many different communities, the Wasserstein distances received many different names and it was not apparent that all the formulations were related to each other. It is only in the beginning of the 21st century, with the publication of reference textbooks [RR98, Vil03] and its increase in popularity that optimal transport metrics became considered as a part of the legacy of the work of Kantorovich and co-authors.

The quadratic case At the end of the 80s, the quadratic Wassertein distance, i.e. considering the distance squared for the cost, began to draw more and more attention. Independent works by Knott and Smith [KS84], Brenier [Bre87] (with the english version [Bre91]), Cuesta and Matrán [CM89] and Rüschendorf and Rachev [RR90] have indeed provided a characterization of the optimal γ in this case. Brenier is usually the one credited for this result, which he formulated as a polar decomposition theorem. He was working on incompressible fluid mechanics: assume that $S : \Omega \to \Omega$ is mapping the initial position of particles of a fluid to their final one, where the fluid is constrained to stay in a bounded domain $\Omega \subset \mathbb{R}^d$. Incompressibility is expressed by the constraint that the push forward of the Lebesgue measure $\mathcal{L}$ by S, denoted $S\#\mathcal{L}$, is equal to $\mathcal{L}$: it means that the distribution particles at the initial and final time is uniform over the domain. A map S such that $S\#\mathcal{L} = \mathcal{L}$ is called a measure-preserving map. One can be interested in computing the projection (and the distance) of a map $S : \Omega \to \Omega$ onto the set of measure preserving maps w.r.t. the Hilbertian metric on $L^2(\Omega, \mathbb{R}^d)$ to quantify how far from being incompressible a map is. Brenier showed that, to compute this projection, one just has to solve an optimal transport problem (with the distance squared as the cost) between $\mathcal{L}$ and $S\#\mathcal{L}$. Moreover, $S = T \circ U$, where U is a measure preserving map, namely the projection of S on the set of the measure-preserving maps, and T is the gradient of a convex function and the optimal transport map between $\mathcal{L}$ and $S\#\mathcal{L}$ (meaning that $(\mathrm{Id}, T)\#\mathcal{L}$ is the optimal γ in (1.1)).

As a byproduct, Brenier showed that in the quadratic case in $\mathbb{R}^d$ (provided that measures have densities w.r.t. the Lebesgue measure), the optimal γ has a very nice structure: it is unique and concentrated on the graph of the gradient of a convex function. In particular, the optimal transport problem does not split mass: each point x is sent onto a unique $y = T(x)$. Later, McCann [McC97] showed that the optimal γ can be used to construct an interpolation between probability measures which is aware of the geometry of the underlying space: this is what is known as McCann's interpolation, and corresponds to geodesic in the Wasserstein space. If a particle of mass is supposed to be sent from x onto y, then it does following the geodesic at constant speed from x to y. Moreover, he showed that there exist some relevant functionals over the space of probability measures which are convex along this interpolation. He used this property to study the uniqueness of solutions of variational problem modeling the behavior of a gas. A few years later, Otto [Ott98, Ott01] together with Jordan and Kinderlehrer [JKO98] understood that this way of interpolating between probability measures reveals an underlying structure of Riemannian manifold which is physically relevant and that some well known parabolic PDEs (in particular the heat equation) could be expressed as gradient flows, w.r.t. this Riemannian structure, of functionals on the space of probability distributions. These functionals were precisely the ones shown by McCann to be convex w.r.t. McCann's interpolation, which is now interpreted as a geodesic interpolation in this Riemannian structure.

In other words, the *Wasserstein space*, which is the space of probability distributions over a given space endowed with the quadratic Wasserstein distance, has at least formally a structure of Riemannian manifold, and gradient flows w.r.t. this structure coincide with actual physical equations. We emphasize that in Otto's original work on the subject [Ott98], the goal was to pass to the limit some non linear PDE, and the gradient flow structure in the Wasserstein space appeared to be the right framework to achieve this end. Later, this point of view appeared to offer many advantages: it enables to get explicit rate of convergence to equilibrium [Vil08, Chapter 24], to make sense of models whose writing in terms of PDEs can only be formal [MRCS10], or to give a way to numerically compute the gradient flows [Pey15]. The publication, in 2005 for the first edition, of the book by Ambrosio, Gigli and Savaré [AGS08], exhibited a sound framework for existence, uniqueness and characterization of these gradient flows.

Gradient flows are first order evolution in time. Second order (in time) equations appear naturally when one considers curves minizing Lagrangians depending on the velocity, measured with the Wasserstein distance. It is the case for the variational model of the incompressible Euler equations of Brenier [Bre89, Bre99], which is itself inspired by the more geometry-oriented works of Moreau [Mor59] and Arnold [Arn66] about the least action principle for incompressible fluid dynamics. Around 2006, the theory of Mean Field Games was introduced by Lasry and Lions in [LL06b, LL07] and, independently, by Caines, Huang and Malamé in [HMC06]. Though apparently disconnected from the theory of optimal transport, it was realized that some instances of these problems share a deep link with it and could be thought as second order in time equations in the Wasserstein space.

Following Brenier's work, some people [McC01] have realized that if the underlying space is not $\mathbb{R}^d$ but has a richer geometric structure, the Wasserstein space was able to reveal it. Indeed, the coupling between probability distributions that it provides has a lot of geometric information in it. It became apparent, with the work of Sturm [Stu06] and Lott and Villani [LV09] that the convexity of the functionals studied by McCann was closely related to the *Ricci* curvature of the underlying space. Leveraging from this observation, a synthetic theory of Ricci curvature was developed with the help of optimal transport, enabling to define and study non smooth spaces with Ricci curvature bounded from below and dimension bounded from above, the so-called $\mathrm{CD}(K,N)$ spaces, later refined in $\mathrm{RCD}(K,N)$ by Gigli [Gig13] by imposing a requirement of being infinitesimally hilbertian. Bounds on the Ricci curvature deal with rate at which the volumes grow or shrink along geodesic interpolation, and optimal transport has provided new proofs of results involving Ricci-curvature related results, namely Brun Minkowski inequalities and functional analysis estimates [McC94, Bar97], isoperimetric inequalities [CM17], etc. These proofs are more robust than the previous ones, hence they can be more easily generalized, and one line of research of this past ten years has been to prove that all these results stay true in $\mathrm{(R)CD}(K,N)$ spaces, i.e. in the non smooth setting.

Numerical optimal transport As mentioned above, in the 40s and the 50s were simultaneously introduced the modern formulation of optimal transport as a linear programming problem and efficient numerical algorithms to tackle linear programming. In short, the first way to solve (1.1) is to consider measures μ, ν with finite support and to use a solver for finite dimensional linear programming. There are some clever refinements that can leverage the precise structure of the cost function, but these kind of methods become untractable when the support of the measures is moderately large.

At the end of the 90s, as the picture of the Wasserstein space as a Riemannian manifold was emerging, Benamou and Brenier proposed in [BB00] the following algorithm: the idea was to compute the whole (McCann) interpolation by solving one single convex problem. In other words, to compute the optimal transport and the Wasserstein distance between two measures μ and ν, one computes a time-dependent curve ρ_t valued in the Wasserstein space which is μ at time $t = 0$ and ν at time $t = 1$, while being a constant-speed geodesic in between. The price to pay is the increase in the number of variables of the problem (namely, by adding time as a variable), but the gain is that the problem stays convex and can be formulated in terms of PDEs. With the help of primal-dual iterative methods to face the optimization problem, one can solve efficiently the problem, especially if the measures have densities. Indeed, in the latter case, PDE-based discretization are more suited than the ones with finitely supported measures. Moreover, this method applies with very few changes for the computation of gradient flows [BCL16] and to solve instances of Mean Field Games [BC15].

We also mention other PDEs based solvers, which leverage on the fact that the optimal γ, in some cases, is concentrated on the graph of a mapping T. Indeed, the computation of the optimal T amounts to solve a Monge-Ampère equation, for which there is now efficient and robust solvers [BFO14, BCM16].

Another class of method are the semi-discrete ones [Mér11, Lév15, KMT16]. One of the measure is supposed to be discrete, i.e. sum of Dirac masses, while the other one has a density. Then, provided that one can compute the integral of the measure with a density over simple cells (typically convex polyhedra), the optimal transport can be computed *exactly* in a reasonable time if one knows how to compute Laguerre diagrams (a generalization of Voronoi diagrams) quickly; and they are indeed efficient solvers for the latter task. This type of computation is well suited for problems where the precise structure of the transport is needed, as the solver returns exactly the solution. It has been applied successively to fluid mechanics computations [GM18a] or design of optical components [MMT18] for instance.

In 2013, Marco Cuturi [Cut13] (see also [SDGP$^+$15]) showed how, by adding an entropic regularization of the transport plan to the linear problem (1.1), one obtains a problem which can be solved very quickly. More precisely, if ε is the scale of the entropic regularization, using Sinkhorn's algorithm, one can solve the problem thanks to a very easy iterative scheme, where each iteration amounts to compute a matrix vector product; however the number of iterations needed increases as $\varepsilon \to 0$ (as well as the quality of approximation). This regularization introduces spreading of mass, i.e. the support of the optimal transport plan γ is no longer supported on the graph of a function: in some cases this is a good thing (from the modeling point of view it corresponds to add noise, see [BCDMN18]), in others it is something undesirable (for instance if one is interested in the transport map and not just the value of the problem).

Applications of the Wasserstein distance On the more applied side, the Wasserstein distance has found numerous applications. We do not at all pretend to be exhaustive and we refer the interested reader to to [San15, PC17, KPT$^+$16] and references therein. As mentioned before, the optimal transport problem, as an instance of linear programming, was naturally suited for economical applications. In the beginning of the 21st century, it was (re)-discovered and introduced in image processing under the name of *Earth Mover's distance* [RTG00]. Around the same time, transport maps have also been used as a way to interpolate colors between images, see for instance [MS03]. With the explosion of numerical methods to solve optimal transport problem after 2010, Wasserstein distances have been used in machine learning as a loss function [FZM$^+$15, ACB17, FSV$^+$18] or for domain adaptation [CFTR17]. As the computation of the Wasserstein distance is one operation among others in the machine learning pipeline, scalability becomes a real issue, and entropic regularization has been the most commonly used tool to bypass it.

As it bears some tiny links with the second part of our manuscript, we mention the line of work developed by Solomon and co-authors [SGB13, SRGB14, SDGP$^+$15]. The idea is to use optimal transport for surface processing. If one has a Riemannian manifold $(\mathcal{N}, \mathfrak{g})$ (specifically: a 2-dimensional Riemannian submanifold of $\mathbb{R}^3$), many tasks in geometry processing imply to deal with functions valued in $\mathcal{N}$. The set of such functions is non convex and defined by non linear constraints. On the other hand, one can use the Wasserstein space $(\mathcal{P}(\mathcal{N}), W_2)$ as a substitute for $(\mathcal{N}, \mathfrak{g})$: the space becomes convex, easy to discretize, and still encodes the geometry of the surface. The (huge) price to pay is an increase in the number of unknowns. In any case, with this heuristic, one can see why considering mappings valued in the Wasserstein space can appear in surface processing.

1.2 Variational problems in the Wasserstein space

Let us specify what variational problems in the Wasserstein space look like and the ones we are interested in. From now on we stick to the setting of this manuscript. We take $\Omega \subset \mathbb{R}^d$ a convex bounded domain which we endow with the quadratic Wasserstein distance W_2, see (2.1) in the next chapter. The space of probability measures over Ω is denoted by $\mathcal{P}(\Omega)$.

1.2.1 A toy model

Let $\nu_1, \nu_2, \dots, \nu_N$ be given probability distributions and $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ a convex functional over the Wasserstein space. As an example, on can think of F as defined by

$$F(\mu) = \begin{cases} \displaystyle\int_\Omega \mu(x) \ln(\mu(x)) \mathrm{d}x & \text{if } \mu \text{ has a density,} \\ +\infty & \text{otherwise,} \end{cases} \tag{1.3}$$

i.e. $F(\mu)$ is the Boltzmann entropy[4] w.r.t. $\mathcal{L}$ the Lebesgue measure restricted to Ω. The functional F favors diffuse densities and is minimal for $\mu = \mathcal{L}/\mathcal{L}(\Omega)$. Then, we fix $\lambda_1, \lambda_2, \dots, \lambda_N$ positive weights and we consider the calculus of variation problem

$$\min_\mu \left\{ F(\mu) + \sum_{i=1}^N \lambda_i \frac{W_2^2(\mu, \nu_i)}{2} \ : \ \mu \in \mathcal{P}(\Omega) \right\}. \tag{1.4}$$

In other words, we are looking for a measure $\mu \in \mathcal{P}(\Omega)$ which is close to the minimum of F (which means, with the example of the entropy, that the measure should be diffuse) and, in the same time, is close to the measures ν_i. Existence of a solution is granted provided F exhibits lower semi-continuity, and uniqueness can be shown under suitable assumptions (either strict convexity of F or absolute continuity of at least one measure ν_i w.r.t. $\mathcal{L}$).

This toy problem is at the same time very simple, because we know how to characterize explicitely the solutions, but on the other hand the informations that can be extracted from the optimality conditions are very useful. Indeed, it appears as a discretization of variational problems involving curves and mappings. The main part of the present manuscript just amounts to show that a complicated problem can be reduced to (a sequence of problems like) (1.4), and to use our understanding of the latter to say something that can be translated at the level of the (complicated) original problem. In some specific cases detailed just below, Problem (1.4) boils down to already studied objects.

[4] By abuse of notation, we call F the Boltzmann entropy rather than *minus* the Boltzmann entropy.

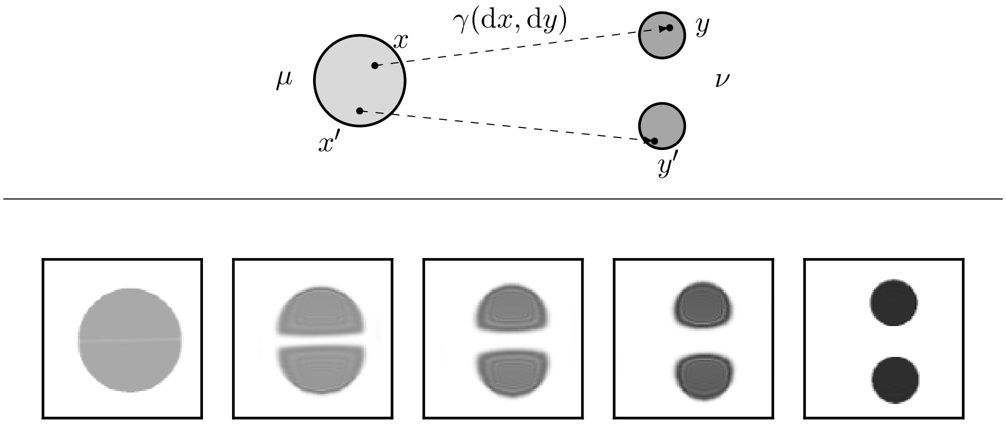

Figure 1.1: Top: schematic view of formulation (2.1) of optimal transport between μ, on the left, and ν, on the right. The quantity $\gamma(x, y)\mathrm{d}x\mathrm{d}y$ represents the amount of mass that is transported from x to y. The coupling γ is chosen in such a way that the total cost is minimal. Bottom: geodesic in the Wasserstein space between the same distributions (computed with an adaptation of the algorithm of Chapter 11). To go from the top to the bottom row, once one has the optimal γ, a proportion $\gamma(x, y)\mathrm{d}x\mathrm{d}y$ of particles follows the straight line between x and y with constant speed. The macroscopic result of all these motions is a time-varying probability distributions, whose snapshots are displayed.

Figure 1.2: Barycenter in the Wasserstein space. The shape in the middle represents the barycenter with equal weights of the probability measures which are (normalized) indicators of the shapes in the corners. Taken from [SDGP$^+$15] with permission of the authors.

If $N = 2$ and $F = 0$, i.e. if we minimize

$$\mu \mapsto \lambda_1 \frac{W_2^2(\mu, \nu_1)}{2} + \lambda_2 \frac{W_2^2(\mu, \nu_2)}{2},$$

then, denoting $\lambda = \lambda_1/(\lambda_1 + \lambda_2)$, the solution of this problem is nothing else than ρ_λ, where $t \mapsto \rho_t$ is the geodesic, in the Wasserstein space, joining ν_1 to ν_2. An example of geodesic in the Wasserstein space is displayed in Figure 1.1. More generally, if $N \geqslant 2$, then this problem amounts to compute the so-called Wasserstein barycenters of the measures ν_i with weights λ_i [AC11], see Figure 1.2 for an illustration. Notice that this definition would be valid for elements of arbitrary metric spaces, and it coincides with the notion of Fréchet barycenter [Fré48]. In short, a problem like (1.4), provided we set $F = 0$, answers the question: in the sense of optimal transport, what is the best way to summarize the data of many measures $\nu_1, \nu_2, \ldots, \nu_N$ in a single one?

If $N = 1$ and $F \neq 0$, i.e. if we solve (setting $\tau = 1/\lambda_1$)

$$\min_{\mu} \left\{ F(\mu) + \frac{W_2^2(\mu, \nu_1)}{2\tau} \; : \; \mu \in \mathcal{P}(\Omega) \right\}, \tag{1.5}$$

then this problem is one step of the minimizing movement scheme (sometimes called the JKO scheme because of the work [JKO98]) used to compute gradient flows in the Wasserstein space. Notice that the problem amounts to find μ which is close to ν_1 but at the same time decreases the energy F. If we define a sequence recursively by taking μ^{k+1} the solution of the problem above with data $\nu_1 = \mu^k$, then, by sending $\tau \to 0$ (in this case τ is interpreted as a time step), μ^k will converge to $\rho_{k\tau}$ where the curve $t \to \rho_t$ is the Wasserstein gradient flow of F. Namely, $t \mapsto \rho_t$, which is a curve valued in the Wasserstein space, is the curve which always follows the direction of steepest descent of F, but where this direction is computed w.r.t. the Wasserstein geometry.

1.2.2 Flow interchange

Problem (1.4) is a convex problem, and one can write the optimality conditions which are (by convexity) necessary and sufficient, hence entirely characterize the solutions of the problem. However, in this work we will concentrate on a single estimate that we extract from (1.4) and that we will use again and again. It corresponds to the perturbation of the optimizer along the gradient flow of a functional which is convex along generalized geodesics. Its use in the case of problems like (1.4) was introduced by Matthes, McCann and Savaré [MMS09] in the context of minimizing movement schemes, under the name *flow interchange*, in order to prove regularity results for gradient flows, see for instance [CGM17] for some recent application of the same technique.

Specifically, let $G : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ be a functional convex along generalized geodesics in the Wasserstein space: it (almost) means that, along the geodesics in the Wasserstein space, the function G is convex. In short, G is convex w.r.t. the Riemannian structure of the Wasserstein space. The typical example is

$$G_m(\mu) = \begin{cases} \dfrac{1}{m-1} \displaystyle\int_\Omega \mu(x)^m \mathrm{d}x & \text{if } \mu \text{ has a density,} \\ +\infty & \text{otherwise,} \end{cases}$$

for $m > 1$, the case $m = 1$ would correspond to the Boltzmann entropy. The gradient flow of G_m is the curve $t \mapsto \rho_t$ satisfying the PDE called the porous medium equation

$$\partial_t \rho = \Delta(\rho^m).$$

Take μ a solution of (1.4) and consider $t \mapsto \rho_t$ the gradient flow of G starting from μ. We use ρ_t for small t as a competitor in the problem defining μ. The key point is that, G being convex along generalized geodesics, we can estimate the derivative of the Wasserstein distance along the flow of G: this is called the *Evolution Variational Inequality* and it reads

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\frac{W_2^2(\rho_t,\nu)}{2}\right|_{t=0} \leqslant G(\nu) - G(\rho_0) = G(\nu) - G(\mu)$$

for any $\nu \in \mathcal{P}(\Omega)$. With the help of this inequality we can write, by optimality of μ,

$$\sum_{k=1}^{m} \lambda_i \left(G(\nu_i) - G(\mu)\right) \geqslant - \left.\frac{\mathrm{d}}{\mathrm{d}t}F(\rho_t)\right|_{t=0}. \tag{1.6}$$

This is the *flow interchange* estimate that we will use over and over.

If $F = 0$, i.e. if μ is the barycenters of the ν_i, we see that, provided the normalization $\sum \lambda_i = 1$ holds,

$$G(\mu) \leqslant \sum_{i=1}^{m} \lambda_i G(\nu_i),$$

which is nothing else than Jensen's inequality: for a convex (w.r.t. the Wasserstein geometry) functional, the value of G at the (Wasserstein) barycenter is smaller than the mean of the values of G.

Moreover, if for instance F is the Boltzmann entropy and $G = G_m$, i.e. ρ satisfies the porous medium equation then

$$- \left.\frac{\mathrm{d}}{\mathrm{d}t}F(\rho_t)\right|_{t=0} = m \int_\Omega |\nabla \mu|^2 \mu^{m-2} = \frac{4}{m} \int_\Omega \left|\nabla \left(\mu^{m/2}\right)\right|^2.$$

As a consequence, (1.6) gives an upper bound on a Sobolev norm of $\mu^{m/2}$. For this to hold, it is enough for the $G_m(\nu_i)$ to be finite. In short: provided the ν_i are in $L^m(\Omega)$ and F is the Boltzmann entropy, if μ is the solution of (1.4), then $\mu^{m/2}$ is in $H^1(\Omega)$. This is an example of elliptic regularity: the minimizer of a variational problem is smoother than the data.

Let us conclude this subsection by explaining where the name *flow interchange* comes from. Assume that we use (1.4) in the framework $N = 1$, i.e. we use the minimizing movement scheme to compute an approximation of the gradient flow of the functional F. In this setting, ν_1 and μ are thought as two samples at times $k\tau$ and $(k+1)\tau$ of a smooth curve valued in the Wasserstein space. Hence, the l.h.s. of (1.6) is nothing else than an approximation of (minus) the dissipation of G along the Wasserstein gradient flow of F, whereas the r.h.s. is the dissipation of F along the Wasserstein gradient flow of G. In short: we can compare the dissipation of G along the flow of F and the dissipation of F along the flow of G, i.e. we can *interchange* the flows.

We emphasize that the flow interchange, in the setting of Hilbert spaces, is immediate. Indeed, let $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ be smooth and a point $z_0 \in \mathbb{R}^d$. Let $t \mapsto x_t$ the curve such that $x_0 = z_0$ and $\dot{x}_t = -\nabla f(x_t)$ (the gradient flow of f) and similarly $t \mapsto y_t$ the curve such that $y_0 = z_0$ and $\dot{y}_t = -\nabla g(y_t)$ (the gradient flow of g). Then

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}g(x_t)\right|_{t=0} = \left.\frac{\mathrm{d}}{\mathrm{d}t}f(y_t)\right|_{t=0}$$

as both quantities are equal to $-\nabla f(z_0) \cdot \nabla g(z_0)$. The dissipation of g along the flow of f is indeed the same as the dissipation of f along the flow of g. In the (not as smooth) setting of the Wasserstein space, one has to assume convexity along geodesics of the functionals and ends up with a one-sided estimate only (because of the time-discretization).

1.2.3 Curves and mappings valued in the Wasserstein space

In the present manuscript, we are interested in problems where the unknowns are curves and mappings valued in the Wasserstein space. Let $\rho : [0,1] \to \mathcal{P}(\Omega)$ a curve valued in the Wasserstein space, i.e. a distribution of mass evolving in time. The main metric quantity associated to this curve is the *action*

$$\int_0^1 \frac{1}{2} |\dot{\rho}_t|^2 \mathrm{d}t,$$

where $\dot{\rho}_t$ is the metric derivative of the curve, i.e. measures the speed of the curve in the Wasserstein space. It is defined by

$$|\dot{\rho}_t| := \lim_{h \to 0} \frac{W_2(\rho_{t+h}, \rho_t)}{|h|}.$$

Curves minimizing the action with fixed endpoint are the geodesics in the Wasserstein space.

On the other hand, let us consider $\boldsymbol{\mu} : D \to \mathcal{P}(\Omega)$ a mapping valued in the Wasserstein space[5]. Here $D \subset \mathbb{R}^p$ is the source space while the target space is $\mathcal{P}(\Omega)$ the Wasserstein space built over $\Omega \subset \mathbb{R}^d$. The *Dirichlet energy* of the mapping $\boldsymbol{\mu}$, which is the multi-dimensional equivalent of the action and heuristically correspond to $\int_D |\nabla \boldsymbol{\mu}|^2/2$ (with the magnitude of the gradient measured in the Wasserstein space), is defined as

$$\mathrm{Dir}(\boldsymbol{\mu}) = \lim_{\varepsilon \to 0} C_p \int_D \left(\int_{B(\xi,\varepsilon)} \frac{W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta))}{2\varepsilon^{p+2}} \mathrm{d}\eta \right) \mathrm{d}\xi, \tag{1.7}$$

with C_p a dimensional constant depending on p the dimension of D. The reader can check that if $\boldsymbol{\mu}$ were a smooth mapping valued in a Hilbert space and W_2 the Hilbertian metric, than $\mathrm{Dir}(\boldsymbol{\mu})$ would really coincide with $\int_D |\nabla \boldsymbol{\mu}|^2/2$. This definition is the one of Korevaar, Schoen [KS93] and Jost [Jos94] for Dirichlet energy of mappings valued in metric spaces, and it coincides with the action of curve if the source space D is a segment of $\mathbb{R}$.

We will be looking at three different classes of problem, which can be roughly stated as follows.

- *Variational problems arising in Mean Field Games.* Find $\rho : [0,1] \to \mathcal{P}(\Omega)$ which minimizes

  $$\int_0^1 \frac{1}{2} |\dot{\rho}_t|^2 \mathrm{d}t + \int_0^1 E(\rho_t) \mathrm{d}t$$

  with fixed or penalized values at $t = 0$ and $t = 1$. Here $E : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ is a functional introducing congestion effect. It can be of the form (1.3), or a constraint on the maximal value of the density, augmented by the integral of ρ against a potential. This problem, which gained interest because of the theory of Mean Field Games [Lio12, Car10, BCS17] (see Section 3.1) features competition between optimal density evolution (minimization of the action), penalization of congestion (through E) and favor of congestion (through boundary conditions and also E if it includes a potential energy). Figure 1.3 illustrates what the solutions look like.

[5] To keep consistent notations in the introduction, the source space is denoted by D and the target space by $\mathcal{P}(\Omega)$. However, in the second part of the manuscript, mainly for contingent reasons, $\Omega \subset \mathbb{R}^p$ will be the source space while $\mathcal{P}(D)$ (with $D \subset \mathbb{R}^q$) will be the target space. As the two parts of this manuscript are independent from one another, we hope that this will not be too confusing for the reader.

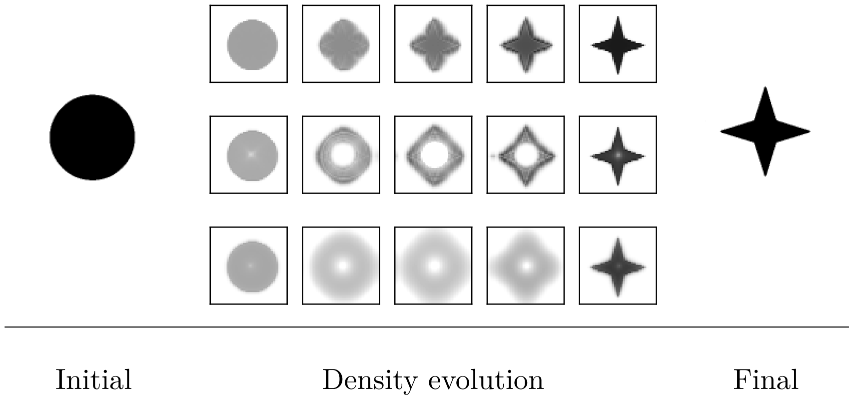

Figure 1.3: Illustration of the optimal density evolution problem in the case where Ω is the 2-dimensional torus. On the left and right are the probability measures corresponding to the initial and final (temporal) value of the curve valued in the Wasserstein space. The first row is the geodesic in the Wasserstein space between the two measures: no congestion effects. In the second row, we have added a potential taking high values in the center of the domain, forcing the optimal curve to avoid this region. On the last row, we still have a potential penalizing presence of mass in the center, but we also penalize congested densities by adding in the running cost the L^2 norm (squared) of the density. As a result, mass has a tendency to spread. These pictures are computed by adapting the algorithm of Chapter 11: to take for the source space a segment is in fact simpler than what is done in this chapter, and following [BCS17], the adaptation to optimal density evolution requires to modify only a few lines of the code.

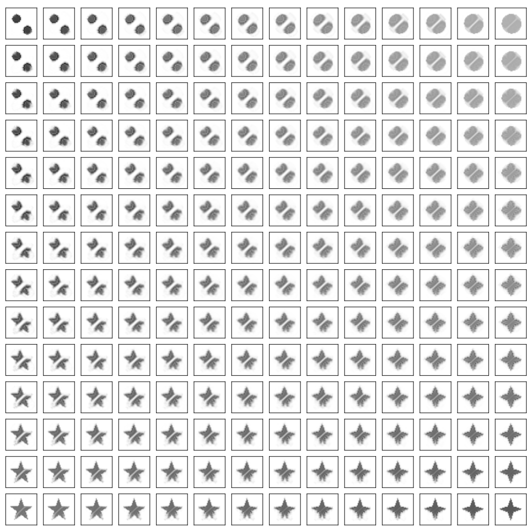

Figure 1.4: Example of a mapping valued in the Wasserstein space. Each little square corresponds to the value of the mapping at one point, which is a probability distribution (represented by its density). This mapping is harmonic, which means that it minimizes the Dirichlet energy among all mappings sharing the same boundary conditions. More on this figure and its computation is said in Chapter 11.

- *Incompressible Euler equations.* Find $Q \in \mathcal{P}(C([0,1],\mathcal{P}(\Omega)))$ a probability measure on the set of curves valued in the Wasserstein space, i.e. the law of a random variable taking values in $C([0,1],\mathcal{P}(\Omega))$ the space of continuous curves valued in the Wasserstein space, which minimizes
$$\int_{C([0,1],\mathcal{P}(\Omega))} \left(\int_0^1 \frac{1}{2} |\dot{\rho}_t|^2 \mathrm{d}t \right) Q(\mathrm{d}\rho),$$
(the expected action of the random curve) with fixed values at $t \in \{0,1\}$ and under the incompressibility constraint that for all $t \in [0,1]$,
$$\int_{C([0,1],\mathcal{P}(\Omega))} \rho_t Q(\mathrm{d}\rho) = \mathcal{L},$$
where $\mathcal{L}$ is the Lebesgue measure restricted to Ω. This constraint states that in expectation the random curve is the Lebesgue measure. As it will be explained in Section 3.2, this problem can be seen as an instance of the least action principle for the incompressible Euler equations [Bre89, Bre99, DF12]. Compared to the previous problem, we face now a continuum of curves valued in the Wasserstein space and the congestion effects are trickier as they are encoded in this incompressibility constraint. Note, however, that this is a (huge) infinite-dimensional linear programming problem in the variable Q.

- *Dirichlet problem in the Wasserstein space.* Find $\boldsymbol{\mu} : D \to \mathcal{P}(\Omega)$ a mapping valued in the Wasserstein space with fixed values $\boldsymbol{\mu}_b : \partial D \to \mathcal{P}(\Omega)$ on the boundary of D which minimizes the Dirichlet energy. The natural terminology, by analogy with the case of mappings valued in Riemannian manifolds, is to call it the *Dirichlet problem* and to consider the minimizers as *harmonic mappings valued in the Wasserstein space*. An example of such a mapping is presented in Figure 1.4. This problem was introduced more than 15 years ago by Brenier [Bre03, Section 3] but the study left more open questions than sound results. Independently, it was reintroduced in the framework of geometry processing by [SGB13] and studied in the PhD thesis of [Lu17], though we argue that a good theoretical framework was still missing, and we hope that this work is a step in this direction.

Notice that the optimality conditions of these problems are, roughly speaking, second order elliptic equations for curves and mappings valued in the Wasserstein space, as they arise when minimizing functionals involving convex functionals of first order derivatives.

As we said, the common point in our work is the use of the flow interchange estimate to tackle these problems and extract interesting features.

For problems involving curves, if one discretize in time the curve ρ with $N+1$ time steps $0, \tau, 2\tau, \dots, 1$ ($\tau := 1/N$ being the distance between two time steps), then the action of the curve is approximated by
$$\sum_{k=1}^{N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_k \tau)}{2\tau}.$$
In particular, at optimality, the k-th component $\rho_{k\tau}$ solves the problem (1.4) with $M = 2$, $\nu_1 = \rho_{(k-1)\tau}$, $\nu_2 = \rho_{(k+1)\tau}$, $\lambda_1 = \lambda_2 = \tau^{-2}$ and a functional F whose precise expression depend on the congestion effect modeled by E. Using the flow interchange estimate (1.6), we deduce that for any G convex w.r.t. the Wasserstein geometry, we have an estimate between the (discrete) second derivative of $k \mapsto G(\rho_{k\tau})$ and the dissipation of E along the flow of G.

On the other hand, for the problem involving the Dirichlet energy of mappings, the formulation by Korevaar, Schoen and Jost directly comes with a natural approximation process: just take at the r.h.s. of (1.7) and look for a mapping minimizing it for a fixed ε. By doing so, we obtain a sequence $\boldsymbol{\mu}_\varepsilon$ converging to a solution of the Dirichlet problem. For a given ε, the mapping $\boldsymbol{\mu}_\varepsilon$ satisfies the following optimality conditions: for any point ξ, the probability distribution $\boldsymbol{\mu}_\varepsilon(\xi)$ is the Wassertein barycenter of the $\boldsymbol{\mu}_\varepsilon(\eta)$ for $\eta \in B(\xi, \varepsilon)$ as already noticed by Jost [Jos94]. Then, applying the flow interchange estimate (more specifically Jensen's inequality), we deduce that $G(\boldsymbol{\mu}_\varepsilon(\xi))$ is smaller than the mean of $G \circ \boldsymbol{\mu}_\varepsilon$ on the ball $B(\xi, \varepsilon)$. Sending $\varepsilon \to 0$, we deduce that the composition $G \circ \boldsymbol{\mu}$ between a functional G convex w.r.t. the Wassertein geometry and a harmonic mapping $\boldsymbol{\mu}$ is a subharmonic function, i.e. satisfies $\Delta(G \circ \boldsymbol{\mu}) \geqslant 0$. We call this property the Ishihara type property [Ish78], which can be thought as a maximum principle for harmonic mappings valued in the Wasserstein space.

Now, considering what is above, let us briefly summarize the new results contained in this manuscript.

- *Variational problems arising in Mean Field Games: soft congestion.* In the case where the penalization E is the integral of a convex function of the density (which can be augmented by a potential), we are able to prove L^∞ regularity results for ρ, locally in time and globally in space. These results do not depend on the temporal boundary conditions, only on a quantification of the convexity of E. These results are proved starting from the flow interchange estimate followed by a iteration process reminiscent of Moser's proof of regularity for elliptic equations [Mos60]. The proof is really different from previous attempts, either based on maximum principle for degenerate elliptic PDEs [Lio12], which gives L^∞ regularity provided the boundary data are regular; or on regularity by duality [San18, CMS16, GM18b], which gives Sobolev regularity for some function of the density. However, previous works deal with quite generic Lagrangians, while our technique applies only to quadratic ones.

- *Variational problems arising in Mean Field Games: hard congestion.* In the case were E enforces the constraint for the density not to exceed a given threshold, we are able to show that the pressure arising from this constraint is not only a measure, but belongs to $L^\infty_t H^1_x$ under loose regularity assumptions on the potential driving the dynamic. We do not rely on a flow interchange estimate, but we still discretize in time and end up with problems like (1.4). This result improves the previous work [CMS16], but at the price that we look only at quadratic Lagrangians.

- *Incompressible Euler equations.* We prove that the averaged entropy (i.e. the expectation of the entropy of a random curve ρ drawn according to the optimal Q) is a convex function of time, a result which was conjectured by Brenier [Bre03, Section 4] but wasn't proved until now. Our proof relies on the use of the flow interchange which directly gives us the convexity of the entropy in problem discretized in time. Posterior to the publication our work [Lav17], Baradat and Monsaigeon [BM18] gave a simpler proof of this conjecture. Contrary to us, they are able to show the convexity for all solutions, whereas we can only show it for a particular one. Indeed, they do not rely on an approximation process and directly work at the level of curves valued in the Wasserstein space.

- *Dirichlet problem in the Wasserstein space.* Our first contribution is to give a proper functional analysis framework for the analysis of the Dirichlet problem. We show equivalences between the definition of the Dirichlet energy of Korevaar, Schoen and Jost and the one based on the extension of the Benamou Brenier formula proposed by Brenier in [Bre03, Section 3]. We show the failure of the so-called superposition principle, answering to [Bre03, Problem 3.5]. Using the flow interchange estimate, we are able to prove the Ishihara-type property. In the case of mappings valued in the space of Gaussian[6] measures, we show well posedness of the Dirichlet problem and write explicit PDEs satisfied by the covariance matrices. Eventually, we propose a numerical discretization based on the Benamou-Brenier formulation that we used to make the illustrations present in this manuscript.

1.3 Organization of the manuscript

Optimal transport toolbox We recall the main definitions and results of optimal transport that we use in the sequel. We mention that we present briefly the so-called *Otto calculus* about the Riemannian structure of the Wasserstein space, which can be thought as hidden behind all of our work.

After the optimal transport toolbox, this manuscript is divided into two main parts, concerned with variational problems about *curves*, and variational problems about *mappings* respectively. For each part, we advise to read the first chapter of it, which we have tried to free from technical details, before going into the more specialized chapters.

Part I: optimal density evolution with congestion

The first part of this manuscript is concerned with problems involving curves valued in the Wasserstein space, namely the variational problems arising in Mean Field Games and the incompressible Euler equations.

Introduction to optimal density evolution We specify the problems about curves valued in the Wasserstein space that we will tackle. We provide a heuristic derivation of the optimality conditions and how, from these optimality conditions, one can guess the results proved in the next chapters. We also make the link between these variational problems and modeling, i.e. what they have to do with Mean Field Games and Incompressible Euler equations.

Regularity of the density in the case of soft congestion The content of this chapter is based on the article *Optimal density evolution with congestion: L^∞ bounds via flow interchange techniques and applications to variational Mean Field Games* written with Filippo Santambrogio [LS18]. We prove the L^∞ regularity of the density in the case of variational problem arising in Mean Field Games, relying on a flow interchange estimate and an iterative process reminiscent of Moser's proof of regularity for elliptic equations. To make the computations rigorous, we discretize the problem in time, and we show that this discretization leads indeed to a good approximation of the original problem.

Regularity of the pressure in the case of hard congestion The content of this chapter is based on the article *New estimates on the regularity of the pressure in density-constrained Mean Field Games* written with Filippo Santambrogio [LS19]. We prove that in density-constrained Mean Field Games, which amounts to problems where the density is forced to stay below a given threshold, then the pressure arising from this constraint, which is *a priori* only a measure, belongs in fact to $L^\infty_t H^1_x$ or even $L^\infty_{t,x}$ provided some regularity assumptions on the potential. The time-discretization used to make the computations rigorous are the same as in the previous chapter, however the estimates at the discrete level are quite different; and the passage to the

[6]In fact, we rather work with elliptically contoured distributions but this subtlety is irrelevant at this point.

limit now deals with dual variables (i.e. the pressure and the value function) and no longer primal variable (i.e. the density, as in the previous chapter).

Time-convexity of the entropy in the multiphasic formulation of the incompressible Euler equations The content of this chapter is based on our article *Time-convexity of the entropy in the multiphasic formulation of the incompressible Euler equations* [Lav17]. We prove the conjecture of Brenier about the convexity of the averaged entropy in the variational formulation of the incompressible Euler equations. The techniques are very similar to the ones of the two previous chapters, however in this case the additional issue is that we deal with a continuum of curves (more specifically a measure on the set of curves valued in the Wasserstein space). At the end of the chapter, we also prove that our formulation of the problem, which looks slightly different than the one of Brenier [Bre99], is in fact equivalent to it.

Part II: Harmonic mappings valued in the Wasserstein space

The second part of this manuscript deals exclusively with the Dirichlet problem for mappings valued in the Wasserstein space. It is mainly based on our article *Harmonic mappings valued in the Wasserstein space* [Lav19].

Introduction to harmonic mappings in the Wasserstein space We specify the link between this problem and the more general one of harmonic mappings valued in Riemannian manifolds and metric spaces. We highlight that the issue is the *positive* curvature of the Wasserstein space which prevents from applying already known theories. We also give an overview of the main arguments and ideas present in the rest of this part.

The Dirichlet energy and the Dirichlet problem In this chapter, we show that one can define the Dirichlet energy in two different ways which turn out to be equivalent: either by relying on the theory of Korevaar, Schoen and Jost which is valid for mappings valued in arbitrary metric spaces, or by an extension of the Benamou-Brenier formulation of the action for curves. Moreover, the space of mappings with finite Dirichlet energy is shown to be identical to $H^1(\Omega, \mathcal{P}(D))$ where the latter is defined in the sense of Reshetnyak [Res97]. We state the Dirichlet problem, prove its well-posedness and derive a dual formulation. Eventually, we show that the superposition principle does not hold, which is one of the main reason why the study of mappings valued in the Wasserstein space turns out to be more involved than the one of curves.

The maximum principle In this chapter, we show the Ishihara-type property: the composition of a functional convex along generalized geodesics with an harmonic mapping valued in the Wasserstein space is a subharmonic function. The proof bears many similarities with the previous part, as it also relies on a approximation process (this time with ε-Dirichlet energies) and the use of the flow interchange estimate, which this time translates as Jensen's inequality for Wasserstein barycenters.

Special cases We first evoke results by other people about the case where the measures on the boundary are Dirac masses: as Ω is flat, the solution of the Dirichlet problem stays valued in the set of Dirac masses. Then we briefly say what happens in the case where the target space is the Wasserstein space over a segment of $\mathbb{R}$: in this very special case the geometry of the Wasserstein space is flat and we do not need to rely on the theory presented in the previous chapters. On the other hand, we also study what happens if all the boundary data belong to $\mathcal{P}_{\text{ec}}(\Omega)$ a family of elliptically contoured distributions (this is a generalization of the gaussians measures), where measures are characterized by their covariance matrix. In this case, we show that a solution of the Dirichlet problem stays valued in $\mathcal{P}_{\mathrm{ec}}(\Omega)$, that we have uniqueness under minor regularity assumptions, and that we can write the PDE satisfied by the covariance matrix. Eventually, we give an example where the solution is (almost) explicitly known, which still features interesting effects of the geometry of the Wasserstein space.

Numerical computations Although not identical, numerical methods very similar to the one of this chapter have been published in the article *Dynamical Optimal Transport on Discrete Surfaces* written with Sebastian Claici, Ed Chien and Justin Solomon [LCCS18]. As we concentrate in the present manuscript on the Dirichlet problem, while the article was mainly aimed at the computation of geodesics (over curved surfaces), the content of this chapter is quite different from the article, though the core ideas are the same. We tackle the problem of the computation of harmonic mappings valued in the Wasserstein space. The only tool at our disposal suited for numerics is the Benamou-Brenier formulation. Inspired by works on geodesics in the Wasserstein space, we propose a finite difference discretization that we mainly use for illustration purposes. The implementation of the algorithm presented in this chapter can be found online at

`https://github.com/HugoLav/PhD`

Perspectives and open questions Being a relatively unexplored topic, we point out some open questions related to harmonic mappings valued in the Wasserstein space that we think are of some interest. We have not included a similar chapter for the first part of this manuscript: of course, the regularity results that we proved in the first part are far from being optimal, but we have no clue about directions for improvement. On the other hand, for harmonic mappings valued in the Wasserstein space, some of our attempts are failed but other gave promising preliminary results, though not conclusive, that we would like to expose.

Notations

We set some useful notations that we will not always recall. Throughout the whole manuscript, we will use the abbreviations w.r.t. (with respect to), l.h.s. (left hand side), r.h.s. (right hand side) and l.s.c. (lower semi-continuous).

If X is any set, the mapping $\mathrm{Id} : X \to X$ denotes the identity mapping.

The symbol $\overline{\mathbb{R}}$ will denote $\mathbb{R} \cup \{+\infty\}$. Though the value $+\infty$ will be allowed, we will *never* consider functionals taking the value $-\infty$.

If X is a polish space (metric, complete and separable), it is endowed with its Borel σ-algebra. We define $\mathcal{P}(X)$ as the space of Borel positive measure with unit mass. It is endowed with the topology of weak convergence, which means convergence in duality with $C(X)$ the space of continuous bounded and real-valued functions defined on X. We also define $\mathcal{M}(X, \mathbb{R}^n)$, for $n \geqslant 1$ as the space of Borel (vectorial) measures valued in $\mathbb{R}^n$ with finite mass, still endowed with the topology of weak convergence. In the case $n = 1$, we use the shortcut $\mathcal{M}(X) := \mathcal{M}(X, \mathbb{R})$. In particular, $\mathcal{P}(X)$ is a convex subset of the linear space $\mathcal{M}(X)$. If $\mu \in \mathcal{P}(X)$ or $\mathcal{M}(X, \mathbb{R}^n)$, integration w.r.t. μ is denoted by $\mathrm{d}\mu$, or by $\mu(\mathrm{d}x)$ if the variable cannot be omitted. If $x \in X$, the Dirac mass at point x is denoted by δ_x. The indicator function of a set X, which is a function taking the value 1 on X and 0 elsewhere, will be denoted by $\mathbb{1}_X$.

The Euclidean spaces $\mathbb{R}^d$ will be endowed with their canonical Euclidean structure with norm denoted by $|\ |$. The notation $B(x, r)$ is used for the closed ball of center x and radius r. The outward normal to a domain X, whenever it exists, is denoted by $\mathbf{n}_X$.

If X is a subset of a Euclidean space $\mathbb{R}^d$, the d-dimensional Lebesgue measure restricted to X will be denoted by $\mathcal{L}_X$ or simply $\mathcal{L}$ if X is clear from the context. If no measure is specified or we write simply $\mathrm{d}x$ for some variable x belonging to a subset of a Euclidean space, then the integration is performed w.r.t. the Lebesgue measure.

If $T : X \to Y$ is a measurable application between two measurable spaces X and Y and μ is a measure on X, then the image measure (or push forward) of μ by T, denoted by $T\#\mu$, is the measure defined on Y by $(T\#\mu)(B) = \mu(T^{-1}(B))$ for any measurable set $B \subset Y$. It can also be defined by

$$\int_Y a(y)(T\#\mu)(\mathrm{d}y) := \int_X a(T(x))\mu(\mathrm{d}x),$$

this identity being valid as soon as $a : Y \to \mathbb{R}$ is an integrable function [AGS08, Section 5.2].

If (X, μ) is a measured space and (Y, d) is any metric separable space, $L^p_\mu(X, Y)$ will denote the space of measurable mappings $f : X \to Y$ for which $d(f, y)^p$ integrable w.r.t. μ for some $y \in Y$. If $Y = \mathbb{R}$, then the letter Y is omitted, and if μ is the Lebesgue measure, then the letter μ is omitted. If Y is an Euclidean space, then we set

$$\|f\|^p_{L^p_\mu(X,Y)} := \int_X |f(x)|^p \mu(\mathrm{d}x).$$

If X and Y are two subsets of Euclidean spaces, the L^∞ norm of a measurable function $f : X \to Y$

is defined as $\|f\|_\infty := \operatorname{ess\,sup}_{x\in X} |f(x)|$, where the essential supremum is taken w.r.t. the Lebesgue measure.

If X and Y are two subsets of Euclidean spaces, $C(X,Y)$ and $C^1(X,Y)$ will denote respectively the continuous and C^1 functions defined on X and valued in Y. If $Y = \mathbb{R}$, then the target space is omitted and we use $C(X)$ or $C^1(X)$. The notation ∇f will stand for the gradient: if X is of dimension d then $\nabla f \in C(X,\mathbb{R}^d)$ for a function $f \in C^1(X)$. On the other hand, $\nabla\cdot$ will stand for the divergence: $\nabla \cdot f \in C(X)$ for a vector field $f \in C^1(X,\mathbb{R}^d)$. If X is of dimension 1, the derivative of $f : X \to Y$ is simply denoted by $\dot{f}$ (if X stands for the time) or f'. Actually, if X is a segment of $\mathbb{R}$ and $f \in C(X,Y)$, the value of f at time $t \in X$ will be denoted by $f_t \in Y$ rather than $f(t)$.

Notations specific to the case of harmonic mappings In the case of harmonic mappings valued in the Wasserstein space, we will consider two domains $\Omega \subset \mathbb{R}^p$ and $D \subset \mathbb{R}^q$. In general, all elements related to Ω will be denoted with Greek letters, and those related to D with Latin ones. For instance, points in Ω (resp. D) will be denoted by ξ, η (resp. x, y), and $(e_\alpha)_{1\leqslant\alpha\leqslant p}$ (resp. $(e_i)_{1\leqslant i\leqslant q}$) is the canonical basis of $\mathbb{R}^p$ (resp. $\mathbb{R}^q$).

On the space $C^1(\Omega \times D, Y)$ the following differential operators can be defined. The derivatives w.r.t. variables in Ω will be denoted by ∇_Ω, or simply $(\partial_\alpha)_{1\leqslant\alpha\leqslant p}$, and those w.r.t. variables in D by ∇_D, or simply $(\partial_i)_{1\leqslant i\leqslant q}$. As an example, if $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$, with components $(\varphi^\alpha)_{1\leqslant\alpha\leqslant p}$, then $\nabla_\Omega \cdot \varphi \in C(\Omega \times D)$ is defined as

$$\nabla_\Omega \cdot \varphi(\xi, x) = \sum_{\alpha=1}^{p} \partial_\alpha \varphi^\alpha(\xi, x),$$

for all $\xi \in \Omega$ and $x \in D$; and $\nabla_D\varphi \in C(\Omega \times D, \mathbb{R}^{pq})$ is defined as, for any $\alpha \in \{1, 2, \dots, p\}$ and $i \in \{1, 2, \dots, q\}$,

$$(\nabla_D\varphi)^{\alpha i}(\xi, x) = \partial_i \varphi^\alpha(\xi, x) \in \mathbb{R}.$$

The notation $C^1_c(\mathring{\Omega} \times D, Y)$ will stand for the smooth functions which are compactly supported in $\mathring{\Omega}$ but not necessarily in D (and valued in Y): if $\varphi \in C^1_c(\mathring{\Omega} \times D, Y)$, it means that there exists a compact set $X \subset \mathring{\Omega}$ such that $\varphi(\xi, x) = 0$ as soon as $\xi \notin X$.

Chapter 2

Optimal Transport toolbox

The goal of this chapter is to present (a tiny part of) the theory of optimal transport with an emphasis on the tools and results that we use in the rest of this manuscript. We do not claim to be able to show all the richness and the level of generality reached by this topic, on the contrary we will focus and deal only with the aspects relevant for our research. This chapter does not provide new contents, and was inspired by the standard textbooks [San15, Vil03, AGS08, Vil08].

In all this chapter, we consider Ω the closure of a bounded convex open set of $\mathbb{R}^d$. By doing so, we restrict ourselves to the case where the underlying space (i.e. Ω) has no curved geometry. The convexity assumption is crucial as it will prevent any congestion effect coming from the presence of a boundary: to move inside Ω following shortest paths (i.e. straight lines), one never meets $\partial\Omega$ the boundary of Ω. Hence the boundary will never be a cause of congestion effects. Eventually, we look only at a bounded Ω. This assumption may be removed at the price of the study of the quadratic moments of the probability measures, but we deliberately prefer to avoid these complications. The generalization to the case where Ω is the d-dimensional torus is straightforward and we do not address it explicitly: actually, this case would be even simpler because there is no boundary term to handle.

The set of probability measures on Ω, denoted by $\mathcal{P}(\Omega)$ is endowed with the topology of weak convergence of measures. It is a convex compact subspace of the set $\mathcal{M}(\Omega)$ of all finite measures on Ω. The space $\mathcal{P}(\Omega)$, endowed with the quadratic Wasserstein distance defined below, is what we call the *Wasserstein space*.

2.1 The Wasserstein distance

2.1.1 The optimal transport problem

If $\mu, \nu \in \mathcal{P}(\Omega)$ are two probability measures, the quadratic Wasserstein distance between them is defined as

$$W_2(\mu,\nu) := \sqrt{\min_\gamma \left\{ \iint_{\Omega\times\Omega} |x-y|^2 \gamma(\mathrm{d}x,\mathrm{d}y) \ : \ \gamma \in \mathcal{P}(\Omega\times\Omega) \text{ and } \pi_0\#\gamma = \mu, \ \pi_1\#\gamma = \nu \right\}}. \tag{2.1}$$

In the formula above, $\pi_0, \pi_1 : \Omega \times \Omega \to \Omega$ stand for the projections on respectively the first and second component of $\Omega \times \Omega$. A $\gamma \in \mathcal{P}(\Omega \times \Omega)$ satisfying the constraints $\pi_0\#\gamma = \mu$ and $\pi_1\#\gamma = \nu$ is called a *transport plan* and an optimal γ is called an *optimal transport plan*. Clearly, $W_2^2 : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}$ is a convex function of its two arguments.

Let us recall briefly the interpretation of this definition. The measures μ and ν are thought as distributions of mass sharing the same total mass (1 by convention). The measure γ on the product space describes a way of moving mass from μ to ν: the quantity $\gamma(x,y)\mathrm{d}x\mathrm{d}y$ is the infinitesimal amount of mass moving from x to y. The constraints $\pi_0\#\gamma = \mu$ and $\pi_1\#\gamma = \nu$ are the ones prescribing that γ actually represents a way to move mass from μ to ν. The cost of moving mass from x to y is $|x-y|^2$, hence the name *quadratic* Wasserstein distance. Then, we take the infimum over γ, namely we look for the cheapest way to move mass from μ onto ν. The distance between μ and ν is the square root of the total cost for this cheapest way.

The optimization problem defining the Wasserstein distance can be seen as a linear programming problem in the variable γ: it consists in the minimization of a linear functional under linear constraints. In particular, it admits a dual problem which reads

$$\frac{W_2^2(\mu,\nu)}{2} = \max_{\varphi,\psi} \left\{ \int_\Omega \varphi(x)\mu(\mathrm{d}x) + \int_\Omega \psi(x)\nu(\mathrm{d}x) \ : \right. \\ \left. \varphi,\psi \in C(\Omega) \text{ and } \varphi(x)+\psi(y) \leqslant \frac{|x-y|^2}{2} \forall x,y\in\Omega \right\}. \quad (2.2)$$

Beware that we have inserted a factor $1/2$ because of the simplifications it leads to in the sequel. The optimal φ and ψ in the problem above are called the *Kantorovich potentials*: the economical interpretation is that $\varphi(x)$ (resp. $\psi(y)$) is the cost of loading (resp. unloading) a unit of mass at x (resp. y). The constraint $\varphi(x)+\psi(y) \leqslant |x-y|^2/2$ states that the total price of loading and unloading cannot excess the cost for moving from x to y, and the total cost is nothing else than the sum of the total cost of loading and unloading.

We will not prove it but these two optimization problems admit solutions $\gamma \in \mathcal{P}(\Omega\times\Omega)$ and $\varphi,\psi \in C(\Omega)$. Moreover, we have the following relations between the optimizers of the primal and dual formulation of the Wasserstein distance.

Proposition 2.1. *Let $\mu,\nu \in \mathcal{P}(\Omega)$ be given. Let us call γ and (φ,ψ) any solutions in the optimization problems* (2.1) *and* (2.2) *respectively.*

1. *There holds*
$$\varphi(x)+\psi(y) = \frac{|x-y|^2}{2}$$
for γ-a.e. $(x,y)\in\Omega\times\Omega$.

2. *Moreover, one can choose φ and ψ in such a way that they are c-transform one from another, namely that*
$$\begin{cases} \varphi(x) &= \inf_{y\in\Omega} \left(\dfrac{|x-y|^2}{2} - \psi(y) \right) \\ \psi(y) &= \inf_{x\in\Omega} \left(\dfrac{|x-y|^2}{2} - \varphi(x) \right). \end{cases}$$

The interpretation of the first point is that, the price of loading at x and unloading at y is equal to the price $|x-y|^2/2$ to move from x to y if some mass is actually moved (by γ) from x to y. Notice that the second point implies that $|\cdot|^2/2-\varphi$ and $|\cdot|^2/2-\psi$ are convex functions. Moreover, it also implies that φ and ψ are Lipschitz functions, with a Lipschitz constant that is bounded by the one of $x\mapsto|x|^2/2$ on Ω, the latter quantity being independent on μ and ν.

Now, and this was understood by Brenier and others, much more can be said when one restricts its attention to measures which are not too singular. The right assumption is for the measures not to give mass to $(d-1)$-dimensional subsets but we will rather consider the stronger assumption that the measures have a density w.r.t. $\mathcal{L}$ the d-dimensional Lebesgue measure. Indeed, in this case, there is a unique γ and it is concentrated on the graph of a function T, meaning that each particle x of the initial measure μ is sent onto a unique point $y = T(x)$. This is expressed in the next proposition, which is usually called Brenier's theorem.

Proposition 2.2. *Assume that μ has a density w.r.t. $\mathcal{L}$. Then there exists a unique γ solution of* (2.1) *and it can be written as $\gamma = (\mathrm{Id}, \mathrm{Id} - \nabla\varphi)\#\mu$ where (φ, ψ) is any solution of the dual problem* (2.2)*. In particular, $\nu = (\mathrm{Id} - \nabla\varphi)\#\mu$.*

Let us underline that $\mathrm{Id} - \nabla\varphi$ is the gradient of the convex function $|\cdot|^2/2 - \varphi$. In fact, Brenier's theorem comes with a reciprocal: under the assumption of the proposition, if there exists a map $T : \Omega \to \Omega$ which is the gradient of a convex function such that $T\#\mu = \nu$ then $\gamma = (\mathrm{Id}, T)\#\mu$ is the (unique) optimal transport plan between μ and ν. In other words, Brenier's theorem states that *there exists a unique way to write $\nu = T\#\mu$ with T being the gradient of a convex function, and such a T can be found by solving the optimal transport problem.*

Notice that Brenier's theorem does not imply the uniqueness of φ, in fact we have only the uniqueness of $\nabla\varphi$ on the support of μ. Let us now introduce an even stronger assumption: that the support of μ is Ω, or more precisely that the density of μ w.r.t. $\mathcal{L}$ is strictly positive a.e. With that in hand, we have uniqueness in the dual problem and we can compute *derivatives* of the Wasserstein distance w.r.t. its inputs.

Proposition 2.3. *Let $\mu, \nu \in \mathcal{P}(\Omega)$ be two absolutely continuous probability measures with strictly positive density w.r.t. $\mathcal{L}$. Then there exists a unique (up to adding a constant to φ and subtracting it from ψ) pair (φ, ψ) of Kantorovich potentials. Moreover the "vertical" derivative of $W_2^2(\cdot, \nu)$ at μ is φ: if $\tilde\rho \in \mathcal{P}(\Omega)$ is any probability measure, then*

$$\lim_{\varepsilon \to 0} \frac{W_2^2((1-\varepsilon)\mu + \varepsilon\tilde\rho, \nu) - W_2^2(\mu, \nu)}{2} = \int_\Omega \varphi(\tilde\rho - \mu).$$

For a proof, we refer to [San15, Propositions 7.18 and 7.17]. We underline that this result is not surprising: with (2.2), one sees that $\mu \mapsto W_2^2(\mu, \nu)/2$ is the supremum of functional linear w.r.t. μ. Hence its derivative is the slope of the linear functional for which the maximum is reached, i.e. φ.

In the proof of Proposition 9.5, we will make a brief use of the 1-Wasserstein distance W_1. It can defined by duality in the following way: for any $\mu, \nu \in \mathcal{P}(\Omega)$,

$$W_1(\mu, \nu) := \sup_\varphi \left\{ \int_D \varphi \mathrm{d}\,(\mu - \nu) \ : \ \varphi \in C(\Omega) \text{ is } 1-\text{Lispchitz} \right\}.$$

The only property that will be of interest to us is that this 1-Wasserstein distance controls the 2-Wasserstein distance in the sense that $W_2 \leqslant C\sqrt{W_1}$, see [San15, Equation (5.1)], where C is related to the diameter of Ω.

2.1.2 The Wasserstein space

Proposition 2.4. *The function $W_2 : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}_+$ defines a distance over $\mathcal{P}(\Omega)$ which metrizes the weak convergence of measures.*

This property is obviously the main reason why W_2 is called a distance. If one were not working on a bounded Ω, the Wasserstein distance would metrize the weak convergence of measures together with the convergence of the second moments, but we avoid this kind of subtleties by assuming compactness.

Definition 2.5. *The* Wasserstein space *is the metric space* $(\mathcal{P}(\Omega), W_2)$.

Based on the property above, it is a compact metric space.

It is important to understand that the Wasserstein space contains the geometry of the underlying space. Indeed, every $x \in \Omega$ can be seen as an element of $\mathcal{P}(\Omega)$ by identifying it with δ_x the Dirac mass located in x. From the very definition of the Wasserstein distance,

$$W_2(\delta_x, \delta_y) = |x - y|.$$

As a consequence, when we will talk in the sequel about differential concepts on the Wasserstein space, the reader can, as a safety check, look at what happens when the curves or mappings are valued in the set of Dirac masses. For instance, if $f : [0, 1] \to \Omega$ is a smooth curve valued in Ω, we can see it as a curve ρ valued in the Wasserstein space by setting $\rho_t = \delta_{f(t)}$, and, for instance the *action* $A(\rho)$ of the curve, defined later in (2.7), can be seen to be the L^2 norm of the derivative of f, up to a factor $1/2$.

We mention that in the Wasserstein space the translations "commute" with the optimal transport plans in the following sense: if $\mu, \nu \in \mathcal{P}(\Omega)$ and if $T_x, T_y : \mathbb{R}^d \to \mathbb{R}^d$ are translations by x and y respectively, then $\gamma \in \mathcal{P}(\Omega \times \Omega)$ is an optimal transport plan between μ and ν if and only if $(T_x, T_y)\#\gamma$ is an optimal transport plan between $T_x\#\mu$ and $T_y\#\nu$. [Vil03, Problem 1]. In particular, if we take for $m(\mu)$ and $m(\nu)$ the centers of mass of μ and ν, and we call $\mu_0 := T_{-m(\mu)}\#\mu$ (and similarly for ν) the centered measure built from μ then we have the decomposition

$$W_2^2(\mu, \nu) = W_2^2(\mu_0, \nu_0) + |m(\mu) - m(\mu)|^2. \tag{2.3}$$

It tells us that we can decouple the effects of the center of mass and the centered part in the Wasserstein distance. It also implies that the mapping $\mu \to m(\mu)$ is a retraction from $(\mathcal{P}(\Omega), W_2)$ onto $(\Omega, |\ |)$: it is a 1-Lipschitz mapping which leaves Ω, identified with the set of Dirac masses, invariant.

2.2 Curves valued in the Wasserstein space and Otto Calculus

2.2.1 Metric derivative

The main interest of the Wasserstein distance –for what we have in mind– is that it endows $\mathcal{P}(\Omega)$ with a differential structure. One can define what a smooth curve valued in $\mathcal{P}(\Omega)$ is and compute its speed in a way which is relevant to modeling.

We will denote by Γ the space of continuous curves from $[0, 1]$ to $\mathcal{P}(\Omega)$. This space will be equipped with the distance d_∞ of the uniform convergence, i.e.

$$d_\infty(\rho^1, \rho^2) := \max_{t \in [0,1]} W_2(\rho^1_t, \rho^2_t).$$

One has to think at an element of Γ as a distribution of mass evolving in time: a pile of sand, the assembly of molecules in a gas, a crowd of people, a herd of sheeps, etc.

Following [AGS08, Definition 1.1.1], we will use the following definition.

Definition 2.6. *We say that a curve* $\rho \in \Gamma$ *is* 2*-absolutely continuous if there exists a function* $\lambda \in L^2([0, 1])$ *such that, for every* $0 \leqslant t \leqslant s \leqslant 1$,

$$W_2(\rho_t, \rho_s) \leqslant \int_t^s \lambda(r) \mathrm{d}r.$$

For real-valued functions, this definition would single out the H^1 functions–those whose speed is square integrable. It follows the same purpose here. The main interest of this notion lies in the following result.

Theorem 2.7. *If $\rho \in \Gamma$ is a 2-absolutely continuous curve, then the quantity*

$$|\dot{\rho}_t| := \lim_{h \to 0} \frac{W_2(\rho_{t+h}, \rho_t)}{|h|}$$

exists and is finite for a.e. t. Moreover,

$$\int_0^1 |\dot{\rho}_t|^2 \mathrm{d}t = \sup_{N \geqslant 2} \sup_{0 \leqslant t_1 < t_2 < \ldots < t_N \leqslant 1} \sum_{k=2}^{N} \frac{W_2^2(\rho_{t_{k-1}}, \rho_{t_k})}{t_k - t_{k-1}}. \tag{2.4}$$

Proof. The first part is just [AGS08, Theorem 1.1.2]. The proof of the representation formula (2.4) can easily be obtained by adapting the proof of [AT03, Theorem 4.1.6]. □

The quantity $|\dot{\rho}_t|$ is called the metric derivative of the curve ρ and heuristically corresponds to the norm of the derivative of ρ at time t in the metric space $(\mathcal{P}(\Omega), W_2)$. Up to now, this definition would make sense and the theorem would be true for curves valued in arbitrary (though separable) metric spaces. However, in the case of the Wasserstein distance, there is this beautiful link between analysis in metric spaces and fluid dynamics which goes as follows, see also [AGS08, Theorem 8.3.1] or [San15, Theorem 5.14].

Theorem 2.8. *Let $\rho \in \Gamma$ be a 2-absolutely continuous curve. Then*

$$\frac{1}{2} \int_0^1 |\dot{\rho}_t|^2 \mathrm{d}t = \min_{\mathbf{v}} \left\{ \int_0^1 \left(\int_\Omega \frac{1}{2} |\mathbf{v}_t|^2 \mathrm{d}\rho_t \right) \mathrm{d}t \right\}, \tag{2.5}$$

where the minimum is taken over all families $(\mathbf{v}_t)_{t \in [0,1]}$ such that $\mathbf{v}_t \in L^2_{\rho_t}(\Omega, \mathbb{R}^d)$ for a.e. t and such that the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t \mathbf{v}_t) = 0$ with no-flux boundary conditions is satisfied in a weak sense. Moreover, there exists a unique optimal $(\mathbf{v}_t)_{t \in [0,1]}$ and it is characterized by the fact that for a.e. $t \in [0,1]$, the field $\mathbf{v}_t$ belongs to the closure of $\{\nabla \phi, \phi \in C^\infty(\Omega)\}$ in the Hilbert space $L^2_{\rho_t}(\Omega, \mathbb{R}^d)$.

The optimal family $(\mathbf{v}_t)_{t \in [0,1]}$ is called the tangent *velocity field to ρ.*

The continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t \mathbf{v}_t) = 0$ describes the evolution of an assembly of particles, whose distribution at time t is ρ_t, and such that the velocity of a particle located at time t in x is $\mathbf{v}_t(x)$. The no-flux boundary conditions expresses the conditions that particles do not leave the domain Ω, mathematically they read $(\rho_t \mathbf{v}_t) \cdot \mathbf{n}_\Omega = 0$ on $\partial\Omega$, where $\mathbf{n}_\Omega$ is the outward normal to Ω.

In other words, if $\rho \in \Gamma$ is a 2-absolutely continuous curve, there exists a time-dependent velocity field $\mathbf{v}_t$ which "represents" the motion of ρ (in the sense that the continuity equation is satisfied) and such that, at for a.e. t,

$$\frac{1}{2} |\dot{\rho}_t|^2 = \int_\Omega \frac{1}{2} |\mathbf{v}_t|^2 \mathrm{d}\rho_t.$$

The latter expression is nothing else than the kinetic energy of the assembly of particles at time t.

2.2.2 A word on Otto calculus

Theorem 2.8 allows us to talk about the so-called *Otto calculus* and the interpretation of the Wasserstein space as a Riemannian manifold. The following discussion will stay at a very formal level, actually one could see the book [AGS08] as clean formalization of it. Let $\mu \in \mathcal{P}(\Omega)$ admitting a smooth density bounded from below and above. To describe the tangent space at μ, one needs to consider the set of all curves $\rho : [-1, 1] \to \mathcal{P}(\Omega)$ with $\rho_0 = \mu$. The question is how one can characterize the speed of ρ at time $t = 0$ and measure its magnitude.

One is tempted to compute $\partial_t \rho|_{t=0}$ to evaluate the speed of ρ at time $t = 0$. But as far as transport is concerned, $\partial_t \rho$ is not the right quantity: it tells you that mass is created at some place (where $\partial_t \rho > 0$) and removed elsewhere (where $\partial_t \rho < 0$), not what is transported where. On the other hand, one has rather to represent the motion with the help of a velocity field $\mathbf{v}_t$: the interpretation of the continuity equation is that a particle located at x at time 0 will move to $x + t\mathbf{v}_0(x)$ at time t at least at first order in t. Then, among all the velocity fields $\mathbf{v}_0$ representing the motion of ρ at time $t = 0$, there exists an optimal one $\bar{\mathbf{v}}_0$, the one such that the square of the metric speed of ρ is nothing else than the kinetic energy of the particles which are moving according to $\bar{\mathbf{v}}_0$. According to Theorem 2.8, the optimal one is characterized by $\bar{\mathbf{v}}_0 = \nabla \phi$ where ϕ satisfies

$$\begin{cases} \nabla \cdot (\mu \nabla \phi) = - \left.\partial_t \rho\right|_{t=0} & \text{in } \mathring{\Omega}, \\ \nabla \phi \cdot \mathbf{n}_\Omega = 0 & \text{on } \partial \Omega, \end{cases} \tag{2.6}$$

which is a rephrasing of the continuity equation. Notice that (2.6) is an elliptic equation, whose r.h.s. $- \partial_t \rho|_{t=0}$ is seen as given, and which is well-posed if $\mu = \rho_0$ is bounded from below and above. Here one difficulty of optimal transport is apparent: it is difficult to handle the situations where there is not mass everywhere (i.e. if μ vanishes), because in this case (2.6) cannot be studied by standard tools. Anyway, provided ϕ is defined (according to Theorem 2.8, $\mathbf{v}_t$ is well defined at least for a.e. t), the square of the speed of ρ at time 0 is nothing else than $\int_\Omega |\nabla \phi|^2 \mathrm{d}\mu$, i.e. the H^1 norm of ϕ weighted by μ.

Given $\mu \in \mathcal{P}(\Omega)$ there are two ways to describe its tangent space $T_\mu \mathcal{P}(\Omega)$, in other words two *bases* for it. The first one, which corresponds to *vertical* motion, is to see an element of $T_\mu \mathcal{P}(\Omega)$ as a function $\partial_t \rho|_{t=0}$ which has 0-mean (because of mass conservation). The second one, which corresponds to *horizontal* motion, is to see an element of $T_\mu \mathcal{P}(\Omega)$ as a function ϕ and to think at $\nabla \phi$ as a velocity field. The change of coordinates formula is nothing else than the elliptic equation (2.6) which is well posed at least if μ is smooth enough. Eventually, the metric tensor has a better expression in the basis for *horizontal* motion, as, if ϕ, ψ are two elements of the tangent space,

$$\langle \phi, \psi \rangle_{T_\mu \mathcal{P}(\Omega)} = \int_\Omega (\nabla \phi \cdot \nabla \psi) \mathrm{d}\mu.$$

Actually, one could guess this metric tensor by doing a formal Taylor expansion of the Wasserstein distance. If ε is small, then at the leading order $W_2^2(\mu, \nu) \simeq \varepsilon^2 \langle \phi, \phi \rangle_{T_\mu \mathcal{P}(\Omega)}$ provided that $\nabla \cdot (\mu \nabla \phi) = -(\nu - \mu)/\varepsilon$.

As an example, if $\mu = \mathcal{L}$, then (2.6) boils down to the Poisson equation with Neumann boundary conditions, and the metric tensor is the H^{-1} scalar product in the basis for vertical motion (while staying the H^1 one in the basis for horizontal motion). It explains why the Wasserstein distance is sometimes considered as a weighted H^{-1} norm, see for instance [San15, Section 5.5.2].

2.2.3 Action of a curve

A crucial quantity is the action of the curve, which is defined as

$$A(\rho) := \begin{cases} \dfrac{1}{2}\displaystyle\int_0^1 |\dot{\rho}_t|^2 \mathrm{d}t & \text{if } \rho \text{ is } 2 - \text{absolutely continuous,} \\ +\infty & \text{else.} \end{cases} \tag{2.7}$$

Thanks to Theorem 2.8, one can interpret $A(\rho)$ as the integral over time of kinetic energy, hence from a physical point of view $A(\rho)$ is an *action* (the integral over time of a Lagrangian). This action looks like a H^1-norm, hence the following results is not surprising.

Proposition 2.9. *The functional $A : \Gamma \to \overline{\mathbb{R}}$ is convex, l.s.c. and its sublevel sets are compact in Γ.*

We recall that Γ is a convex subspace of the set of functions defined on $[0,1]$ and valued in $\mathcal{M}(\Omega)$, hence convexity of A has a well-defined meaning.

Proof. To prove that A is convex and l.s.c., we rely on the representation formula (2.4) which shows that A is the supremum of convex continuous functions. Moreover if $\rho \in \Gamma$ is a curve with finite action and $s < t$, then, again with (2.4), one can see that $W_2(\rho_s, \rho_t) \leqslant \sqrt{2A(\rho)}\sqrt{t-s}$. This shows that the sublevel sets of A are uniformly equicontinuous, therefore they are relatively compact thanks to Ascoli-Arzela's theorem. ☐

Let us underline that the whole goal of the second part of this work is to give a meaning to this action A when one faces no longer curves valued in the Wasserstein space, but mappings, i.e. probability measures depending on more than one parameter.

2.2.4 Geodesics

There is a particular class of curves valued in the Wasserstein space, namely the (constant-speed) geodesic. A curve is a geodesic if it is the shortest path between two points. With the additional requirement that this geodesic is traveled at constant-speed, a curve $\rho \in \Gamma$ is by definition a geodesic if and only if for any $t, s \in [0,1]$,

$$W_2(\rho_t, \rho_s) = |t-s| W_2(\rho_0, \rho_1).$$

In the sequel, *geodesic* will always mean *constant-speed geodesic*.

There are two main features that we want to underline: the first one is the characterization of geodesics as solutions of a problem of calculus of variations, and the second one is the fact that geodesics between two measures μ and ν can be computed automatically if one solves (2.1) the optimal transport problem between them.

The first statement would be in fact true for geodesics valued in arbitrary (though separable) metric spaces. It amounts to say that given $\mu, \nu \in \mathcal{P}(\Omega)$ the solutions of the problem

$$\min_\rho \left\{ A(\rho) = \frac{1}{2}\int_0^1 |\dot{\rho}_t|^2 \mathrm{d}t \ : \ \rho \in \Gamma \text{ such that } \rho_0 = \mu \text{ and } \rho_1 = \nu \right\}$$

are exactly the constant-speed geodesics joining μ to ν, and the value of the problem is $W_2^2(\mu, \nu)$ /2. Notice that Proposition 2.9 can help to show the existence of at least one solution. Switching to the fluid dynamic formulation with Theorem 2.8, we can see that we can also write

$$W_2^2(\mu,\nu) = \min_{\rho,\mathbf{v}} \left\{ \int_0^1 \int_\Omega |\mathbf{v}_t|^2 \mathrm{d}\rho_t \mathrm{d}t \ : \ \rho \in \Gamma \text{ such that } \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0 \text{ and } \rho_0 = \mu, \ \rho_1 = \nu \right\}, \tag{2.8}$$

which is sometimes called the Benamou-Brenier formulation of the Wasserstein distance. This expression actually perfectly fits in the framework of Otto calculus in the sense that it shows that W_2 is the Riemannian distance coming from the metric tensor described above.

If one knows how to solve the optimal transport problem (2.1) between μ and ν, then the geodesics between these two measures can be deduced. In fact, if $\gamma \in \mathcal{P}(\Omega \times \Omega)$ is an optimal transport plan between μ and ν, then the curve $t \mapsto \rho_t := ((1-t)\pi_0 + t\pi_1)\#\gamma$ is a constant speed geodesic between μ and ν for t running between 0 and 1, and reciprocally every geodesic can be written that way [San15, Proposition 5.32]. The interpretation is that, if a particle must be sent from x to y, then it moves on the segment (i.e. the geodesic in Ω) joining x to y at constant-speed. An example of geodesic in the Wasserstein space is displayed in Figure 1.1, page 9. Combining this result with the structure of the optimal transport plans, one can write the following.

Proposition 2.10. *Let $\mu, \nu \in \mathcal{P}(\Omega)$ and assume that μ has a density w.r.t. $\mathcal{L}$. Then there exists a unique geodesic $\rho \in \Gamma$ joining μ to ν and it can be written, for $t \in [0,1]$,*

$$\rho_t = (\mathrm{Id} - t\nabla\varphi)\#\mu,$$

where (φ, ψ) is any pair of Kantorovich potentials between μ and ν.

2.3 Gradient flows and functional over the Wasserstein space

2.3.1 Gradient flows

An other class of remarkable curves are the gradient flows generated by functionals convex along geodesics. Roughly speaking, if $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ is a given functional, a gradient flow is a curve $\rho : [0, +\infty) \to \mathcal{P}(\Omega)$ along which F decreases "the most" w.r.t. the Wasserstein distance, in a formal way it can be written

$$\frac{\mathrm{d}\rho_t}{\mathrm{d}t} = -\nabla F(\rho_t). \tag{2.9}$$

Of course nor the notion of gradient or of time derivative make sense as vectors in the Wassertsein space, but the Otto calculus, by providing a formal Riemannian structure on the space $\mathcal{P}(\Omega)$, indicates that there is some hope to make sense of it. In [AGS08] (see also [San15, Chapter 8]), it is shown how the notion of gradient flow can still be defined through the use of metric quantities only.

A standard assumption to ensure the existence and uniqueness of a gradient flow with a given initial value ρ is that F is convex along generalized geodesic. If μ_0, μ and ν are three probability measures on Ω, one can always build a transport plan $\gamma \in \mathcal{P}(\Omega \times \Omega \times \Omega)$ such that the 1-marginals are respectively μ_0, μ and ν and the 2-marginals are optimal transport plans between μ_0, μ on the one hand and μ_0, ν on the other hand (notice that in general the last 2-marginal is not an optimal plan between μ and ν). Then, the generalized geodesic $\rho : [0,1] \to \mathcal{P}(D)$ between μ and ν with base point μ_0 is defined as $\rho_t := a_t\#\gamma$, with $a_t : (x,y,z) \in \Omega^3 \mapsto (1-t)y + tz \in \Omega$. A functional $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ is said convex along generalized geodesics if for any points μ_0, μ and ν, there exists a generalized geodesic ρ joining μ to ν with base point μ_0 such that $F \circ \rho : [0,1] \to \overline{\mathbb{R}}$

is a convex function. As a particular case, a function convex along generalized geodesics is convex along geodesics in the Wasserstein space, but the reciprocal is not always true [AGS08, Remark 9.2.8]

Gradient flows in the Wasserstein space are a very large topic, we will only need what are called the Energy Dissipation Equality (EDE) and the Evolution Variational Inequality (EVI) formulations of gradient flows, which are ways to make sense of (2.9) in the metric framework. They are summarized in the following theorem, whose proof can be found in [AGS08, Theorem 11.2.1].

Theorem 2.11. *Let $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ a functional l.s.c. and convex along generalized geodesics. Then, for any $\rho \in \mathcal{P}(\Omega)$ such that $F(\rho) < +\infty$, there exists a 2-absolutely continuous curve $t \in [0, +\infty) \mapsto S_t^F \rho \in \mathcal{P}(\Omega)$ such that $S_0^F \rho = \rho$ and for any $t \geqslant 0$ and any ν such that $F(\nu) < +\infty$,*

$$\limsup_{h \to 0,\ h > 0} \frac{W_2^2(S_{t+h}^F \rho, \nu) - W_2^2(S_t^F \rho, \nu)}{2h} \leqslant F(\nu) - F(S_t^F \rho). \tag{2.10}$$

Moreover, the function $t \mapsto F\left(S_t^F \rho\right)$ is decreasing and more precisely for any $t \geqslant 0$,

$$\int_0^t |\dot{S_s^F \rho}|^2 \mathrm{d}s = F(\rho) - F(S_t^F \rho). \tag{2.11}$$

The curve $S^F \rho$ (which can be shown to be unique) is nothing else than the gradient flow of F starting form ρ.

We have a few comments to make. This result is by no way trivial, on the contrary it can be seen as a great achievement of the theory of gradient flows in the Wasserstein space. To get convinced that it might be true, the reader can replace the Wasserstein space by a Hilbert space, take F to be convex, and check that indeed something like (2.10) is actually true if ρ satisfies (2.9). In view of the problems we will tackle later, notice that Theorem 2.11 tells us that, to build an interesting competitor in a problem involving the squared Wasserstein distance, it might be useful to follow the gradient flow of some functional convex along generalized geodesics.

At this point, it might be necessary to make the difference between functional which are convex and the ones convex along (generalized) geodesics. If $\mu, \nu \in \mathcal{P}(\Omega)$, there are at least two ways to compute the "mean" between them:

- take $\dfrac{\mu + \nu}{2}$ the usual mean between measures;
- or take $\rho_{1/2}$ where $\rho \in \Gamma$ is a geodesic joining μ to ν.

These two means do not coincide, see Figure 2.1. The first one will be called the linear one, while the second will be the metric one. By a convex functional F over $\mathcal{P}(\Omega)$, we mean a functional such that the value of the linear mean is smaller than the mean of the values; whereas for a functional convex along geodesics we need the value of the metric mean to be smaller than the mean of the values. The functional $F\mu \mapsto W_2^2(\mu_0, \mu)$ (square distance to a fixed measure) is the example of a convex functional (this can easily be seen from (2.2)) which is not convex along geodesics (a feature which expresses the positive curvature of the Wasserstein space in the sense of Alexandrov, see [AGS08, Section 7.3]). Actually, this function F is convex along *generalized* geodesics if the base point is μ_0.

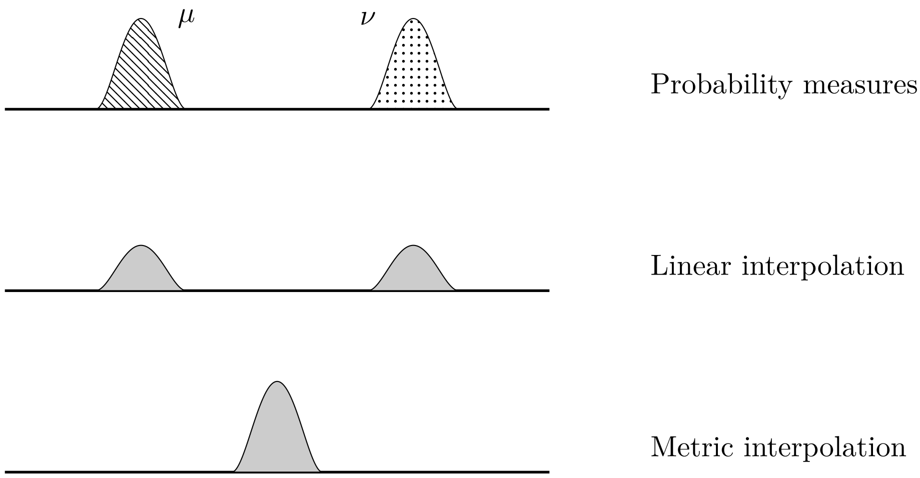

Figure 2.1: On the different ways of interpolating between probability measures. Top row: two probability measures μ and ν on the real line. Middle row: linear interpolation $(\mu+\nu)/2$ of the measures. Bottom row: metric interpolation of the measures, i.e. middle point of the geodesic in the Wasserstein space joining μ to ν.

2.3.2 Examples of functionals defined over the Wasserstein space

Let us introduce a class of functionals defined over the Wasserstein space which will be of great importance in this manuscript (for a complete overview of the topic, we refer the reader to [San15, Chapter 7]). The most intriguing ones are the functionals of the density, which take the form

$$F(\mu) = \int_\Omega f(\mu^{\mathrm{ac}}(x))\mathrm{d}x + f'(+\infty)\mu^{\mathrm{sing}}(\Omega) \tag{2.12}$$

where $f : [0, +\infty) \to \mathbb{R}$ is convex and bounded from below, and $\mu =: \mu^{\mathrm{ac}}\mathcal{L} + \mu^{\mathrm{sing}}$ is the decomposition of μ as an absolutely continuous part $\mu^{\mathrm{ac}}\mathcal{L}$ and a singular part μ^{sing} w.r.t. $\mathcal{L}$. If f is superlinear, which is equivalent to $f'(+\infty) = +\infty$, then this functional is infinite if μ is not absolutely continuous w.r.t. $\mathcal{L}$. In the latter case, by convexity of f, this functional is minimized when μ is constant (i.e. proportional to $\mathcal{L}$): it penalizes congested densities. Some standard and useful properties of F are summarized below.

Proposition 2.12. *Assume that f is convex, bounded from below and that $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ is defined by* (2.12)*. Then the following assertions hold:*

1. *The functional F is convex and l.s.c. on $\mathcal{P}(\Omega)$.*
2. *If $s^d f(s^{-d})$ is convex and decreasing, then the functional F is convex along generalized geodesics in $(\mathcal{P}(\Omega), W_2)$.*

Let us underline that the convexity of Ω is crucial for the latter point to actually hold. If this is the case, then the gradient flow of f, defined in Theorem 2.11, can be shown to satisfy the PDE

$$\begin{cases} \partial_t \rho = \nabla \cdot (\rho \nabla (f'(\rho))) & \text{in } \mathring{\Omega}, \\ \rho \nabla (f'(\rho)) \cdot \mathbf{n}_\Omega = 0 & \text{on } \partial\Omega. \end{cases}$$

The typical function f satisfying the two assumptions above is $f(s) = s^m$ for some $m > 1$. According to the computation above, the gradient flow of the associated F would lead to the porous medium equation $\partial_t \rho = C(m)\Delta(\rho^m)$.

2.4 Heat flow

However, the central functional which will appear everywhere is the (negative) Boltzmann entropy which is defined as

$$H(\mu) = \begin{cases} \displaystyle\int_\Omega \ln(\mu(x))\mu(x)\mathrm{d}x & \text{if } \mu \text{ has a density w.r.t. } \mathcal{L}, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.13}$$

Using Proposition 2.12, one can say that H is l.s.c., convex and convex along generalized geodesics. The central result is that *the gradient flow of H in the Wasserstein space is the heat flow with Neumann boundary conditions.*

The heat flow denotes the flow of the heat equation. This equation will be of great importance as it will be the tool used to regularize probability measures: indeed, following the heat flow with Neumann boundary conditions is the best way, in a convex domain with boundary, to regularize a probability measure without leaving the Wasserstein space. Moreover, in Chapter 6, the link between the heat flow and the Boltzmann entropy will be fully exploited.

We will denote by $\Phi : [0, +\infty) \times \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ the heat flow with Neumann boundary conditions acting on Ω. If $\mu \in \mathcal{P}(\Omega)$ and $t > 0$, then $\Phi_t\mu \in \mathcal{P}(\Omega)$ is defined as the measure $u(t,x)\mathrm{d}x$ with a density $u : (0, +\infty) \times \Omega \to \mathbb{R}$ which is the solution of the Cauchy Problem

$$\begin{cases} \partial_s u(s,x) = \Delta u(s,x) & \text{if } (s,x) \in (0,+\infty) \times \mathring{\Omega}, \\ \nabla u(s,x) \cdot \mathbf{n}_\Omega(x) = 0 & \text{if } (s,x) \in (0,+\infty) \times \partial\Omega, \\ \lim\limits_{s\to 0}[u(s,x)\mathrm{d}x] = \mu & \text{in } \mathcal{P}(\Omega), \end{cases}$$

where $\mathbf{n}_\Omega$ is the outward normal to Ω.

A closely related object is the so-called *heat kernel*. We denote by $K : (0, +\infty) \times \Omega \times \Omega \to \mathbb{R}_+$ the heat kernel associated to the Laplacian on Ω with Neumann boundary conditions [Are02, Section 7]. It is the function such that for any $t > 0$

$$\Phi_t(u_0\mathcal{L}) := \left(\int_\Omega K_t(x,y)u_0(y)\mathrm{d}y \right) \mathrm{d}x,$$

at least if $u_0 \in L^1(\Omega)$. Notice, as a constant function is preserved by the heat flow, that the integral of $K_t(x, \cdot)$ is 1 for a.e. $x \in \Omega$.

As said above, the key point is that the heat flow Φ is the gradient flow of the entropy w.r.t. the Wasserstein geometry, in the sense of Theorem 2.11. As an immediate consequence, (2.10) and (2.11) hold if one replaces F by H and S^F by the heat flow Φ. The useful properties of the heat flow are summarized in the following proposition.

Proposition 2.13. *The heat flow Φ satisfies the following properties:*

(i) For any $\mu \in \mathcal{P}(\Omega)$ and any $t > 0$, the measure $\Phi_t\mu$ has a density w.r.t. $\mathcal{L}$ which is bounded from below by a strictly positive constant and belongs to $C^1(\mathring{\Omega})$.

(ii) For any $t > 0$, the density of $\Phi_t\mu$ w.r.t. $\mathcal{L}$ is bounded in $L^\infty(\Omega)$ by a constant that depends on t, but not on $\mu \in \mathcal{P}(\Omega)$.

(iii) For a fixed $t > 0$ and for any $\mu \in \mathcal{P}(\Omega)$ and $a \in C(\Omega)$, one has

$$\int_\Omega a \,\mathrm{d}\,(\Phi_t\mu) = \int_\Omega (\Phi_t a)\,\mathrm{d}\mu.$$

(iv) For any $\mu, \nu \in \mathcal{P}(\Omega)$ *and any* $t \geqslant 0$,

$$W_2(\Phi_t \mu, \Phi_t \nu) \leqslant W_2(\mu, \nu). \tag{2.14}$$

Proof. Point (i) is standard interior parabolic regularity. Point (ii) comes from $L^\infty - L^1$ estimates for the Neumann Laplacian, see [Are02, Section 7]. Point (iii) just states that the heat flow is self-adjoint. Point (iv) comes from the convexity along generalized geodesics of the entropy H and the fact that the heat flow is the gradient flow of the latter, see [AGS08, Theorem 11.2.1]. □

Except for (iv), all of the statements of Proposition 2.13 remain true if we drop the convexity assumption on Ω, and only assume that Ω is connected and has a Lipschitz boundary.

With the help of the last point, we can prove this uniform estimate about the behavior of the heat flow for small values of t.

Proposition 2.14. *There exists a function* $\omega : [0, +\infty) \to \mathbb{R}$*, continuous and with* $\omega(0) = 0$ *such that, for any* $\mu \in \mathcal{P}(\Omega)$ *and any* $t \geqslant 0$,

$$W_2(\Phi_t \mu, \mu) \leqslant \omega(t).$$

Proof. The only thing to check is that ω is continuous in 0. Assume by contradiction that it is not the case. We can find $(\mu_n)_{n \in \mathbb{N}}$ a sequence in $\mathcal{P}(\Omega)$ and $(t_n)_{n \in \mathbb{N}}$ a sequence that tends to 0 such that, for some $\delta > 0$, there holds $W_2(\Phi_{t_n} \mu_n, \mu_n) \geqslant \delta$. Up to extraction, we can assume that μ_n converges to some limit μ. We can write

$$W_2(\Phi_{t_n} \mu_n, \mu_n) \leqslant W_2(\Phi_{t_n} \mu_n, \Phi_{t_n} \mu) + W_2(\Phi_{t_n} \mu, \mu) + W_2(\mu, \mu_n) \leqslant W_2(\Phi_{t_n} \mu, \mu) + 2 W_2(\mu, \mu_n),$$

where we have used the last point of Proposition 2.13. But then it is clear that the two terms of the r.h.s. tend to 0, which is a contradiction. □

Part I

Optimal density evolution with congestion

Chapter 3

Introduction to optimal density evolution

The goal of this chapter is to give an overview of the present part about optimal density evolution, to present the results that we will prove, and to provide a flavor of the techniques of proof. The discussion in this chapter will stay at a formal level, with non rigorous arguments, all the technical details are provided in the next chapters.

3.1 Variational problem arising in Mean Field Games

The problems we are interested in deal with the temporal evolution of a density subject to congestion effect. Namely, we consider curves $\rho \in \Gamma = C([0,1], \mathcal{P}(\Omega))$ which are continuous and valued in the space of probability measures over a fixed bounded convex domain Ω. The measure ρ_t denotes the density of agents, or particles, at time t. In all the sequel, we always identify a measure with its density w.r.t. $\mathcal{L}$ the Lebesgue measure restricted to Ω. We will look for curves solving a variational problem of the form

$$\min_{\rho} \left\{ \int_0^1 \frac{1}{2} |\dot{\rho}_t|^2 \mathrm{d}t + \int_0^1 E(\rho_t) \mathrm{d}t + \Psi(\rho_1) \ : \ \rho \in \Gamma, \ \rho_0 \text{ given} \right\}. \tag{3.1}$$

Let us describe in details the different terms in this objective functional.

- The first term is the integral over time of the square of the speed of the curve ρ in the Wasserstein space, which can also be seen as the *action* of the curve, namely the integral over time of the kinetic energy, see Section 2.2. If one would only minimize this term, with ρ_0 and ρ_1 fixed, the set of solutions would be the set of geodesics (in the Wasserstein space) between ρ_0 and ρ_1.

- To define the second term, we need to specify a functional $E : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ which will describe congestion effects. In Chapter 4, the functional E (the "running cost") takes the form of an integral functional such as

$$E(\rho) := \int_\Omega f(\rho(x)) \mathrm{d}x + \int_\Omega V(x)\rho(x) \mathrm{d}x$$

  for a convex function f and a fixed time-independent potential V. The function f penalizes concentrated densities while on the contrary the potential V favors them (namely those which are concentrated in the minima of V). This is what we call *soft congestion*, as very peaked densities are penalized (through f), but still allowed. On the other hand, in Chapter 5, we will deal with *hard congestion*, as we will forbid densities whose L^∞ norm is above a fixed threshold, namely 1. Specifically, the functional E will take the form

$$E(\rho) := \begin{cases} \int_\Omega V(x)\rho(x)\mathrm{d}x & \text{if } \rho(x) \leqslant 1 \text{ for a.e. } x \in \Omega, \\ +\infty & \text{else.} \end{cases}$$

- The final penalization Ψ can be either a functional of the same form of E, or a constraint which prescribes ρ_T. According to us, the most interesting results are interior regularity (away from $t = 0$ and 1), thus the precise form of the final penalization is most of the time irrelevant.

These variational problem can be thought as interesting in themselves, as an illustration for the interplay between optimal density evolution (the action of the curve), favor of congestion (through V and Ψ), and penalization of congestion (through f or the hard congestion constraint $\rho \leqslant 1$), see for instance [BJO09] for an early introduction of them. On the other hand, they are closely connected to the Mean Field Game theory as detailed below. We mention that the case where $E(\rho)$ is the H^{-1} norm of $\rho - 1$, which corresponds to the least action principle for a cloud of galaxies with Newtonian interaction has been introduced in [BFH$^+$03] under the name *reconstruction of the early universe*. It has been studied in depth in [Loe06] with an approach based on the dual problem, which has a regularizing effect not present in the case we are interested in.

We will only look at the cases where E and Ψ are convex functional over $\mathcal{P}(\Omega)$, hence the "primal problem" (3.1) is a convex one. To understand the optimality conditions, the main tool is the dual problem which can obtained by a formal inf − sup exchange. Indeed, we use Theorem 2.8 to express the action of the curve as the kinetic energy with a velocity field $\mathbf{v}$ submitted to a continuity equation. Using ϕ as a Lagrange multiplier to enforce the continuity equation, the solution of (3.1) is given by the saddle point

$$\min_{\rho,\mathbf{v}} \sup_{\phi} \left\{ \int_0^1 \int_\Omega \frac{1}{2}|\mathbf{v}|^2 \mathrm{d}\rho \mathrm{d}t + \int_0^1 E(\rho_t)\mathrm{d}t + \Psi(\rho_1) \right. \\ \left. + \int_\Omega \phi_0 \mathrm{d}\rho_0 - \int_\Omega \phi_1 \mathrm{d}\rho_1 + \int_0^1 \int_\Omega (\partial_t \phi + \mathbf{v} \cdot \nabla \phi)\, \mathrm{d}\rho \mathrm{d}t \right\}.$$

and the only constraint is that ρ_0 is fixed. Now we exchange the infimum and the supremum, something which can be justified with the help the Fenchel-Rockafellar theorem like in [Car15]. In the saddle point formulation, let us do the optimization in ρ and $\mathbf{v}$. The one is $\mathbf{v}$ is straightforward as the Lagrangian is quadratic in $\mathbf{v}$, we hasve$\mathbf{v} = -\nabla \phi$ and the remaining part, calling $h = -\partial_t \phi + \frac{1}{2}|\nabla \phi|^2$ can be written

$$\sup_{\phi,h} \min_{\rho} \left\{ -\int_0^1 \int_\Omega h \mathrm{d}\rho \mathrm{d}t + \int_0^1 E(\rho_t)\mathrm{d}t + \Psi(\rho_1) \right. \\ \left. + \int_\Omega \phi_0 \mathrm{d}\rho_0 - \int_\Omega \phi_1 \mathrm{d}\rho_1 \ : \ -\partial_t \phi + \frac{1}{2}|\nabla \phi|^2 = h \right\}.$$

Now the minimization is formally over. Indeed, ρ_0 is fixed and optimizing on the other values of ρ, the Fenchel transform of E and Ψ appear. Namely, the dual problem reads

$$\sup_{\phi,h} \left\{ \int_\Omega \phi_0 \mathrm{d}\rho_0 - \int_0^1 E^\star(h_t)\mathrm{d}t - \Psi^\star(\phi_1) \ : \ \phi, h : [0,1] \times \Omega \to \mathbb{R} \text{ and } -\partial_t \phi + \frac{1}{2}|\nabla \phi|^2 = h \right\}, \tag{3.2}$$

where $E^\star, \Psi^\star : C(\Omega) \to \overline{\mathbb{R}}$ denote the Fenchel transforms of E and Ψ. Here ϕ is the so-called value function and $p = h - V$ can be thought as a pressure or a price as explained below. The existence of a solution to this dual problem is guaranteed if one relaxes the space of function in which ϕ and h lives (the precise choice depends on E and Ψ). In any case, let ρ be a solution of the primal problem, call $\mathbf{v} : [0,1] \times \Omega \to \mathbb{R}^d$ its tangent velocity field obtained thanks to Theorem 2.8, and take ϕ, h a solution of the dual problem. Then the absence of duality gap leads to the system of equations

$$\begin{cases} \nabla \phi & = -\mathbf{v}, \\ h & \in \partial E(\rho), \\ \phi_1 & \in \partial \Psi(\rho_1), \end{cases} \tag{3.3}$$

where $\partial E, \partial \Psi$ denote the subdifferentials in the sense of convex analysis. Recall that to these equations one has to add

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) & = 0, \\ -\partial_t \phi + \frac{1}{2}|\nabla \phi|^2 & = h, \end{cases} \tag{3.4}$$

which are the constraints of the primal and dual variables respectively. The first equation, namely the continuity equation, is supplemented with no-flux boundary conditions $\nabla(\rho \mathbf{v}) \cdot \mathbf{n}_\Omega = 0$, where $\mathbf{n}_\Omega$ is the outward normal to Ω. In short, the optimality conditions (3.3), (3.4) are a coupling between backward a Hamilton-Jacobi equation for ϕ with r.h.s. belonging to $\partial E(\rho)$ (i.e. depending on the density) and terminal cost Ψ, and a forward continuity equation with initial density given and velocity $-\nabla \phi$.

Now, if one chooses the right functional spaces, one can make sense of (3.1) and (3.2) and get an existence result, but with a pretty weak notion of solution. Namely, ρ is only a probability measure at any time t, and the pressure $p = h - V$ is merely a positive measure. The goal of Chapters 4 and 5 is to prove additional regularity for the solutions of the primal and dual problems.

3.1.1 On the link with Mean Field Games

Other than the intrinsic interest of (3.1) as an interplay between optimal density evolution and congestion, the main motivation for the regularity of the study of these variational problems was about Mean Field Games (MFG).

MFG aim at modeling situations where there is a large number of rational agents who are playing a game (i.e. having to take decisions whose payoff depends on what the others do) where the payoff depends only on the average (i.e. mean field) behavior of the other agents. One example, which is where the models in this manuscript come from, is the one of crowd motion. Imagine a crowd of people who want to escape a given place, and they are not in an emergency situation so that they can take time to think and adopt a rational behavior. Each agent wants to escape the room, but on the other hand also wants to avoid congested area. He or she will choose his trajectory to reach the exit while avoiding others, but the latter condition depends on the choice of other agents, hence the game aspect. There is a *mean field* effect because each agent is only interested in the average behavior of the rest of the crowd, i.e. the density of other agents, and not in the specific trajectory of each other agent. The reader might begin to see the link with the problem (3.1) introduced above: the terminal condition (reaching the exit) favors congestion, while the aversion of people for crowded areas plays in the opposite direction.

For the whole theory of Mean Field Games, introduced by Lasry and Lions in [LL06b, LL07] and, independently, by Caines, Huang and Malamé in [HMC06], we refer to the lecture notes by Cardaliaguet [Car10] and to the video-recorded lectures by Lions, [Lio12]. From the mathematical point of view, we study the situation where there is an infinite number of players, which means that the situation is modeled through concepts of fluid mechanics (density, velocity, pressure, etc.) and characterize with the help of PDEs. What there is to characterize are Nash equilibria, i.e. a set of strategy where each player has no interest in deviating from its strategy if other players do not. Most models assume stochastic effects on the trajectory of the agents, and the corresponding PDEs include diffusion terms which make the solution smooth and simplify the analysis, besides being reasonable from the modeling point of view. Analytically, the most difficult case consists in problems where the interaction between players is local (i.e. the cost at point x and time t depends on the value of the density $\rho_t(x)$, without averaging it in a neighborhood) and no diffusion is present. This case is essentially attacked when the game is of variational origin, i.e. it is a potential game, and the equilibrium condition arises as an optimality condition for an optimization problem in the class of density evolutions. For local potential MFG, we refer to [Car15, CG15] and to the survey [BCS17].

More specifically, we assume that we have a continuum of agent, and each agent has a given position $x(0)$ and chooses its trajectory $x : [0,1] \to \Omega$ by solving a control problem, with a finite temporal horizon (taken equal to 1) of the form

$$\min_x \left\{ \int_0^1 \left(\frac{|\dot{x}(t)|^2}{2} + V(x(t)) + p(t,x(t)) \right) \mathrm{d}t + \Psi(x(1)) \right\}, \tag{3.5}$$

The function $p : [0,1] \times \Omega \to \mathbb{R}$ is a pressure, or a price if one thinks in economical terms which depends on the mean field effect, i.e. the density of other agents. In the case of soft congestion, the pressure is just a function of the density ρ of other agents, namely, to keep the same notations as above, $p(t,x) = f'(\rho_t(x))$ where f is a convex function. On the other hand, in the case of hard congestion, namely if the density if forced to stay below 1, we just know that the pressure is a positive function, which does not vanish only on areas where the constraint $\rho \leqslant 1$ is saturated, and whose role is to prevent it from being violated. From the economical point of view, it is a *price* that agents have to pay to pass through congest areas. Here the terminal cost $\Psi : \Omega \to \mathbb{R}$ is the price paid by the players at the final time, to make the link with (3.1) the final penalization of the density would be $\int_\Omega \Psi \mathrm{d}\rho_1$.

The striking result, already understood by Lasry and Lions [LL06b] (see also [BCS17] for a short and self contained introduction) is that to find the evolution of the density of agents, under a monotonicity assumption (which, in our setting, translates in the convexity of the running cost E and the final cost Ψ), it is enough to solve the variational problem (3.1) and its dual (3.2). Indeed, let us take $\rho, \mathbf{v}, \phi, h$ solutions of such problems, recall that they satisfy the optimality conditions (3.3), (3.4). In particular, ϕ solves the Hamilton-Jacobi equation associated to the control problem (3.5) with terminal cost Ψ, i.e. ϕ is the value function for such problem. On the other hand, as $\mathbf{v} = -\nabla\phi$, it means that if an agent located in x at time t moves with velocity $\mathbf{v}(t,x) = -\nabla\phi(t,x)$, then the resulting motion of all the agents is indeed described by ρ. In short: the optimality conditions (3.3), (3.4) exactly describe the mean field game model.

Alternatively, the same equilibrium problem can be formulated in terms of a probability measure Q on the set $C([0,1],\Omega)$ of paths valued in Ω. This measure Q represents the distribution of strategy of the agents: $Q(\gamma)\mathrm{d}\gamma$ describes the proportion of agent choosing the strategy γ, i.e.

moving along γ. With $e_t : C([0,1],\Omega) \to \Omega$ is the evaluation map at time t, defining $\rho_t = (e_t)_\# Q$, the measure ρ_t becomes the spatial distribution of agents at time t. Then, for the measure Q to be an equilibrium, we require $(e_0)_\# Q = \rho_0$ and that Q-a.e. curve is optimal for (3.5) with this definition of ρ_t and a pressure $p(t,\cdot)$ which belongs to $\partial E(\rho_t)$.

Yet, these considerations are essentially formal and not rigorous, so far. The objects $\rho, \mathbf{v}$ and ϕ live in rather big functional spaces and giving a precise meaning to the optimality conditions is not obvious. Moreover, the difficulty with the interpretation about individual agents solving (3.5) is the following: the function $h(t,x) := V(x) + p(t,x)$ is *a priori* defined a.e.: indeed, either p is a function of the density ρ in the case of soft congestion, or is merely a positive measure in the case of hard congestion. Integrating it on a curve, as we do when we consider the action $\int_0^1 h(t,x(t))\mathrm{d}t$ in (3.5) has absolutely no meaning! Of course, it would be different if we could prove some regularity (for instance, continuity) on ρ and p. The question of the regularity in mean field games is a very challenging one and is not entirely understood yet. In [CMS16] a stategy to overcome this difficulty, taken from [AF09], is used: indeed, it is sufficient to choose a suitable representative of h to give a precise meaning to the integral of h on a curve, and the correct choice is

$$\hat{h}(t,x) := \limsup_{r\to 0} h_r(t,x) := \frac{1}{|B(x,r)|}\int_{B(x,r)} h(t,y)\mathrm{d}y.$$

To prove that Q is concentrated on optimal curves for $\hat{h}$ it is then enough to write estimates with h_r and then pass to the limit as $r \to 0$. This requires an upper bound on h_r, and the natural assumption is that the maximal function $Mh := \sup_r h_r$ is L^1 in space and time. Thanks to well-known results in harmonic analysis, $h \in L^1$ is not enough for this but $h \in L^m$ for $m > 1$ is instead enough. Once integrability of Mh is obtained, then one can say that the optimal measure Q is concentrated on curves which minimize in (3.5) in the class of curves $x(\cdot)$ such that $\int_0^1 Mh(t,x(t))\mathrm{d}t < +\infty$. These curves are almost all curves in a suitable sense, thanks to the integrability properties of Mh in space-time, but they are in general *not all* curves. To be able to compare with all curves, what would be needed is Mh bounded, or in other words that h (hence p as the potential V is assumed to be bounded) belongs to L^∞.

Thus, the key point is to get summability estimates on the pressure p, which in the case of soft congestion, translates into summablity estimates on ρ as $p = f'(\rho)$. Moreover, if one proves them, then it is possible to infer regularity for the value function ϕ: indeed, p appears in the r.h.s. of the Hamilton-Jacobi equation and it implies, if p is in L^m with $m > 1 + d/2$, that ϕ exhibits Hölder and Sobolev regularity, as proved by Cardaliaguet and collaborators [Car15, CG15, CPT15]. To summarize,

1. If $p \in L^m$ with $m > 1$ then one can build $Q \in \mathcal{P}(C[0,1],\Omega)$ which represents the strategy of the agents in such a way that Q-a.e. curve is optimal in (3.5), but optimal among the class of curves satisfying some integrability assumption involving the maximal function of p.

2. If $p \in L^m$ with $m > 1 + d/2$, then the value function ϕ is Hölder-continuous and satisfies $\partial_t\phi \in L^{1+\varepsilon}$, $\nabla\phi \in L^{2+\varepsilon}$ for some $\varepsilon > 0$, at least locally in space and time.

3. If $p \in L^\infty$ one can build $Q \in \mathcal{P}(C[0,1],\Omega)$ in such a way that Q-a.e. curve is optimal in (3.5) compared to *all* other curves. Moreover, ϕ is Hölder-continuous and exhibits Sobolev regularity just as above.

What we provide, in the next two chapters, is precisely L^∞ regularity of p in both soft and hard congestion. However, the main restriction of our work is that we consider only a quadratic Lagrangian (i.e. only $L(x,\dot{x}) = |\dot{x}|^2/2$ appear (3.5)) to be able to import optimal transport techniques, while previous results that we will describe just below work with more general Lagrangians. Their Lagrangians $L(x, \dot{x})$ can depend on x and can behave like $|\dot{x}|^r$ when $r \to +\infty$ with $r \neq 2$.

In the case of soft congestion, given the formula $p = f'(\rho)$, if f' is bounded from below and $f(s)$ behaves as s^q as $q \to +\infty$ then automatically $p \in L^{q/(q-1)}$. Thus we can always build the measure Q, but regularity of ϕ is *a priori* true only under the condition $q < 1 + 2/d$. Of course, if $\rho \in L^\infty$ (and this what we will prove!) and f' is bounded from below then automatically $p \in L^\infty$. The question of the L^∞ regularity of ρ was already studied, in the MFG framework, by P.-L. Lions (see the second hour of the video of the lecture of November 27, 2009, in [Lio12]). The analysis by P.-L. Lions was more general than ours in what concerns the Lagrangian. On the other hand, we are able to include a potential $V(x)$ and to obtain local regularity results, which were not present in [Lio12]. Indeed, the results presented by P.-L. Lions only concerned the case where both ρ_0 and ρ_1 are fixed (planning problem) and belong to L^∞, and no potential V is considered. The technique was essentially taken from maximum principles in degenerate elliptic PDEs; it could be adapted to the case where ρ_1 is penalized instead of fixed (which amounts to changing a Dirichlet boundary condition at $t = 1$ into a Neumann one), but adapting it in order to obtain local results seems out of reach. Indeed, local estimates in degenerate elliptic equations usually require quantitative information on the degeneracy and the growth of the different terms, which are in general not available in this setting. Here what we do is different, as detailed in the next subsection.

In the case of hard congestion, we are able to prove that $p \in L^\infty$ as soon as the potential V belongs to $W^{1,q}(\Omega)$ with $q > d$, where d is the dimension of the ambiant space. The only previous study of the regularity of the pressure we are aware of is the one of [CMS16], where the authors obtain $p \in L^2_{t,loc} BV_x$. It allows, thanks to the injection $BV \hookrightarrow L^{d/(d-1)}$ to say that p is in L^m with $m > 1$ and hence recover the interpretation with Q representing the distribution of strategy. Such regularity was obtained by mimicking the proof of the regularity of the pressure in the case of the incompressible Euler equations first investigated by [Bre99] and later refined in [AF08]. The main strategy is what was latter called *regularity by duality* [San18]: one evaluates the dual gap between space-time translations of the primal solution and the (untranslated) dual solution, quantifies precisely the discrepancy, and uses it to deduce Sobolev regularity. We underline that this strategy was used in [AF08] precisely to be able to deduce in [AF09] an interpretation of the model of the incompressible Euler equations in term of a measure Q on the set of curves. Here we get higher regularity for the pressure than in [CMS16], with less assumption on the data (we require $V \in W^{1,q}(\Omega)$ with $q > d$ while they assume $V \in C^{1,1}(\Omega)$), and on more general domain (they work on the torus, we work in a general convex domain); however we handle only quadratic Lagrangians whereas they work in a more general setting. As we describe briefly below, our strategy is different than theirs: we really get an explicit inequality involving the Laplacian of the pressure, from which we apply standard elliptic regularity techniques. Eventually, it seems that the strategy that we use cannot be applied to study the regularity of the pressure for the more complicated setting of the incompressible Euler equations.

Now, let us detail in the two following subsections the strategy to get these regularity estimates.

3.1.2 Soft congestion

We concentrate on the case of soft congestion, namely

$$E(\rho) := \int_\Omega f(\rho(x))\mathrm{d}x + \int_\Omega V(x)\rho(x)\mathrm{d}x.$$

In this case, the optimality conditions (3.3), (3.4) read

$$\begin{cases} \partial_t \rho - \nabla \cdot (\rho \nabla \phi) & = 0, \\ \rho_0 & \text{given}, \\ -\partial_t \phi + \dfrac{1}{2}|\nabla \phi|^2 & \leqslant f'(\rho) + \nabla V, \\ \phi_1 & \in \partial \Psi(\rho_1), \end{cases}$$

and in the third equation there is equality ρ-a.e. We will get rid of the subtilty as far as this informal presentation is concerned and pretend that the third equation is in fact an equality and not an inequality.

Our result is that, *provided that f is convex enough, the solution ρ to* (3.1) *is unique and belongs to* $L^\infty([T_1, T_2] \times \Omega)$ *for any* $0 < T_1 < T_2 < 1$, see Theorems 4.4, 4.5 and 4.6 for the precise statements. In other words, there is L^∞ regularity for ρ, global in space and local in time. Let us underline that this result is surprising: if for instance one takes $f(\rho) = \rho^2$, what we prove is that not only $\rho \in L^2([0,1] \times \Omega)$ (which would be true for any $\rho \in \Gamma$ competitor for which the objective functional is finite) but that ρ is bounded in L^∞. In particular, no conditions are asked on ρ_0 and Ψ (other than the fact that there exists at least on $\rho \in \Gamma$ with finite energy) for this result to hold.

Let us explain how one proves –formally– this estimate. We work in the case $V = 0$, as computations are already quite involved in this simpler setting. We introduce the functionals U_m, where

$$U_m(\rho) := \frac{1}{m(m-1)} \int_\Omega \rho(x)^m \mathrm{d}x,$$

and $m > 1$ ($U_1(\rho)$ can be defined as the Boltzmann entropy of ρ, and the normalization constants are chosen for coherence with this case). The idea is to look at the behavior w.r.t. time of $U_m(\rho)$, for ρ the solution of (3.1). We will control in a fine way the growth of the quantities $U_m(\rho)$ when $m \to +\infty$, relying on an iterative process reminiscent of Moser's proof of regularity for elliptic equations [Mos60].

Specifically, we are interested in the second derivative w.r.t. time of $U_m(\rho)$ where ρ is a solution of (3.1). To guess the result, one can introduce the convective derivative $D_t := \partial_t - \nabla \phi \cdot \nabla$ which is the derivative along the flow of the velocity field $\mathbf{v} = -\nabla \phi$. As the continuity equation is satisfied, for any function $g : [0,1] \times \Omega \to \mathbb{R}$, there holds

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega g(t,x) \mathrm{d}\rho_t = \int_\Omega (D_t g) \mathrm{d}\rho_t$$

The continuity equation can be written $D_t \rho = \rho \Delta \phi$ and, taking the Laplacian of the Hamilton-Jacobi equation, dropping a positive term in the process, we get $D_t(\Delta \phi) \geqslant -\Delta(f'(\rho))$. With these identities in mind,

$$\begin{aligned} \frac{\mathrm{d}^2}{\mathrm{d}t^2} U_m(\rho) = \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{m} \int_\Omega (D_t \rho) \rho^{m-1} &= \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{m} \int_\Omega (\Delta \phi) \rho^m \\ &= \frac{1}{m} \int_\Omega (D_t \Delta \phi) \rho^m + \frac{m-1}{m} \int_\Omega (D_t \rho)(\Delta \phi) \rho^{m-1} \\ &\geqslant -\frac{1}{m} \int_\Omega \Delta(f'(\rho)) \rho^m + \frac{m-1}{m} \int_\Omega (\Delta \phi)^2 \rho^m. \end{aligned}$$

Doing an integration by parts in the first integral, and dropping the second one as it is positive, we are left with the estimate

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} U_m(\rho_t) \geqslant \int_\Omega |\nabla \rho_t|^2 \rho_t^{m-1} f''(\rho_t). \tag{3.6}$$

In particular, as f is convex, we see that the r.h.s. is positive, hence the function $U_m(\rho_t)$ is a convex function of time. In the case $f = 0$, where ρ is simply a geodesic in the Wasserstein space, we recover the fact that $U_m(\rho)$ is a convex function of time: this is the geodesic convexity of U_m, see Proposition 2.12. As a consequence, in the case f arbitrary (though convex), if ρ_0 and ρ_1 are fixed and belong to some L^m for $m > 1$, then so does ρ_t for any $t \in [0,1]$. Similarly, if ρ_0, ρ_1 belong to L^∞, so does ρ_t for any $t \in [0,1]$.

Now we want to go further and drop assumptions on ρ_0, ρ_1. To estimate more precisely the r.h.s. of (3.6), a natural assumption is $f''(s) \geqslant s^\alpha$ (with α which could be negative, of course): if this is the case, one can check that the integrand of the r.h.s. is larger than $|\nabla(\rho_t^{(m+1+\alpha)/2})|^2$ (up to a constant depending polynomially in m). Using the Sobolev injection $H^1 \hookrightarrow L^{2d/(d-2)}$, one can conclude (neglecting the 0-order term of the H^1 norm of $\rho_t^{(m+1+\alpha)/2}$), with $1 < \beta < d/(d-2)$, that

$$C(m)\frac{\mathrm{d}^2}{\mathrm{d}t^2}U_m(\rho_t) \geqslant \left(\int_\Omega \rho_t^{\beta(m+1+\alpha)}\right)^{1/\beta}$$

In the case $\alpha \geqslant -1$, we see that the r.h.s. is larger than $U_{\beta m}(\rho_t)^{1/\beta}$. In other words, we have obtained a control of $U_{\beta m}(\rho)$ in terms of $U_m(\rho)$. Such a control can be iterated. If we take a positive cutoff function χ which is equal to 1 on $[T_1 - \varepsilon, T_2 + \varepsilon]$ and which is null outside $[T_1 - 2\varepsilon, T_2 + 2\varepsilon]$, multiplying (3.6) by χ and integrating the left hand side by parts twice, we can say that

$$\int_{T_1-\varepsilon}^{T_2+\varepsilon} U_{\beta m}(\rho_t)^{1/\beta}\mathrm{d}t \leqslant C(m,\varepsilon)\int_{T_1-2\varepsilon}^{T_2+2\varepsilon} U_m(\rho_t)\mathrm{d}t,$$

where the constant $C(m,\varepsilon)$ grows not faster than a polynomial function of m and ε^{-1}. We have to work a little bit more on the l.h.s. because we want to exchange the power $1/\beta$ and the integral sign, and unfortunately Jensen's inequality gives it the other way around. To this extent, we rely on the following observation: as the function $U_{\beta m}$ is convex (this can be seen in (3.6)) and positive, it is bounded on $[T_1, T_2]$ either by its values on $[T_1, T_1 - \varepsilon]$ or on $[T_2, T_2 + \varepsilon]$, thus we have a "reverse Jensen's inequality"

$$\left(\int_{T_1}^{T_2} U_{\beta m}(\rho_t)\mathrm{d}t\right)^{1/\beta} \leqslant \frac{(T_2 - T_1)^{1/\beta}}{\varepsilon}\left(\int_{T_1-\varepsilon}^{T_1} U_{\beta m}(\rho_t)^{1/\beta}\mathrm{d}t + \int_{T_2}^{T_2+\varepsilon} U_{\beta m}(\rho_t)^{1/\beta}\right).$$

Combining this inequality with the estimation we have on the r.h.s., we deduce that

$$\left(\int_{T_1}^{T_2} U_{\beta m}(\rho_t)\mathrm{d}t\right)^{1/\beta} \leqslant C(m,\varepsilon)\int_{T_1-2\varepsilon}^{T_2+2\varepsilon} U_m(\rho_t)\mathrm{d}t,$$

where the new constant $C(m,\varepsilon)$ has also a polynomial behavior in m and ε^{-1}. This estimation is ready to be iterated. Indeed, setting $m_n := \beta^n m_0$ and $\varepsilon_n = 2^{-n}\varepsilon_0$, given the moderate growth of $C(m,\varepsilon)$, it is not difficult to conclude that

$$\limsup_{n\to+\infty}\left(\int_{T_1-\varepsilon_n}^{T_2+\varepsilon_n} U_{m_n}(\rho_t)\mathrm{d}t\right)^{1/m_n} < +\infty.$$

As the l.h.s. controls the L^∞ norm of ρ on $[T_1, T_2] \times \Omega$, this is enough to conclude that ρ is bounded locally in time and globally in space.

Let us comment on some technical refinements that arise in the actual proof, presented in Chapter 4.

- In practice, we do not have enough temporal regularity to differentiate twice w.r.t. time. To bypass this issue, we introduce a discrete in time version of (3.1) and we prove all the estimates at the discrete level. More is said on that in Section 3.3 at the end of this chapter.

- If we add an interior potential V, the r.h.s. of (3.6) contains lower order terms that are controlled by the term involving f''. However, the sign of the l.h.s. is no longer known and the function U_m is no longer convex but rather satisfies

$$\frac{d^2}{dt^2} U_m(\rho_t) + \omega^2 U_m(\rho_t) \geqslant 0,$$

  where ω grows linearly with m. In particular, the "reverse Jensen inequality" becomes more difficult to prove, but it is still doable.

- With assumptions on the final penalization, the regularity can be extended to the final time. More precisely, if we assume that the final penalization Ψ is given by the sum of a potential term and a congestion term, then formally (and again we use a discrete in time version to make this rigorous),

$$\left. \frac{d}{dt} U_m(\rho_t) \right|_{t=1} \leqslant b(m) U_m(\rho_1), \tag{3.7}$$

  where the constant $b(m)$ depends on the potential and can be taken equal to 0 if there is no potential. This inequality enables to control the value of U_m at the boundary $t = 1$ by its values in the interior. Thus the same kind of iterations can be performed and gives L^∞ regularity up to the boundary.

- If $\alpha < -1$, we only have a control of U_m by $U_{\beta(m+1+\alpha)}$. Thus we must start the iterative procedure with a value m such that $m < \beta(m + 1 + \alpha)$, i.e. we must impose *a priori* some L^m regularity on ρ (with a m which depends on α and β, the latter depending itself only on the dimension of the ambient space). Such a regularity is imposed by assuming that ρ_0 (which is fixed) is in $L^m(\Omega)$ and that the boundary penalization in $t = 1$ is the sum of a potential and a congestion term. Indeed, if this is the case, the boundary condition (3.7) combined with the interior estimate (3.6) shows that the potentials V, W are small enough (compared to something that depends on the function f and m), the L^m norm of ρ on $[0, 1] \times \Omega$ must be bounded.

3.1.3 Hard congestion

Now we tackle *a priori* estimates in the case of hard congestion. Namely we assume that the running cost is

$$E(\rho) := \begin{cases} \displaystyle\int_\Omega V(x)\rho(\mathrm{d}x) & \text{if } \rho(x) \leqslant 1 \text{ a.e. } x \in \Omega, \\ +\infty & \text{else.} \end{cases}$$

In other words, we forbid the density to be above the threshold 1. In particular, we must assume that the Lebesgue measure of Ω is larger than 1 in order for probability measures satisfying the constraint to exist. Also, the terminal density will be penalized by $\int_\Omega \Psi \mathrm{d}\rho_1$, where $\Psi \in C(\Omega)$ is a fixed potential. In this case, the optimality conditions (3.3), (3.4) read

$$\begin{cases} \partial_t \rho - \nabla \cdot (\rho \nabla \phi) & = 0, \\ -\partial_t \phi + \frac{1}{2}|\nabla \phi|^2 & \leqslant P + V \quad \text{(with equality on } \{\rho > 0\}\text{)}, \\ \rho_0 & \text{given}, \\ \phi_1 & \leqslant \Psi \quad \text{(with equality on } \{\rho_1 > 0\}\text{)}, \end{cases}$$

where $P \geqslant 0$ is a measure concentrated on the set $\{\rho = 1\}$. Indeed $h \in \partial E(\rho)$ reads, in this case, $h = P + V$ and $P \geqslant 0$. Compared to what is above we use the letter P instead of p to denote the pressure: we will have to distinguish between the measure P and its density p w.r.t. $\mathcal{L}$.

The density ρ already belongs to L^∞ (by the very definition of the constraint is satisfies), the object for which we will improve regularity is the pressure. Our result is that, *provided that V is smooth enough (namely $V \in W^{1,q}(\Omega)$ for $q \geqslant d$), then the pressure P has a density w.r.t. Lebesgue which belong to $H^1(\Omega)$ for a.e. time*, see Theorem 5.5 for the precise statement.

Let us give a heuristic derivation of this result. For simplicity, we will just consider the conditions which are satisfied on the support of ρ, where the inequalities become equalities. Anyway, this is not restrictive since we are interested in estimates on the pressure P, i.e. on the set $\{\rho = 1\}$. As we will see later, the pressure P is a measure which can be decomposed into two parts: its restriction to $[0, 1) \times \Omega$ is absolutely continuous w.r.t. the Lebesgue measure on $[0, 1) \times \Omega$, and its density is denoted by p; on the other hand, there is also a part on $\{1\} \times \Omega$ which is singular, but absolutely continuous w.r.t. the Lebesgue measure on Ω, and its density is denoted by P_1. This second part represents a jump of the function ϕ at $t = 1$, which allows to re-write the system as follows.

$$\begin{cases} \partial_t \rho - \nabla \cdot (\rho \nabla \phi) & = 0, \\ -\partial_t \phi + \frac{1}{2}|\nabla \phi|^2 & = p + V, \\ \rho_0 & \text{given}, \\ \phi_1 & = \Psi + P_1. \end{cases} \tag{3.8}$$

where the density ρ satisfies $\rho \leqslant 1$ everywhere and $p, P_1 \geqslant 0$ are strictly positive only on the regions where the constraint involving ρ is saturated, i.e. where $\rho = 1$ ($\rho_1 = 1$ in the case of P_1).

Similarly to the case of soft congestion, we denote by $D_t := \partial_t - \nabla\phi \cdot \nabla$ the convective derivative. This time, the idea is to look at the quantity $-D_{tt}(\ln \rho)$. Indeed, the first equation of (3.8) can be rewritten $D_t(\ln \rho) = \Delta\phi$. On the other hand, taking the Laplacian of the second equation in (3.8), it is easy to get, dropping a positive term, $-D_t(\Delta\phi) \leqslant \Delta(p + V)$. Hence,

$$-D_{tt}(\ln \rho) \leqslant \Delta(p + V). \tag{3.9}$$

Notice that if $\rho(t, x) = 1$, then ρ is maximal at (t, x) hence $-D_{tt}(\ln \rho)(t, x) \geqslant 0$. On the other hand, if $\rho(t, x) < 1$ then $p(t, x) = 0$. In other words, p satisfies $\Delta(p + V) \geqslant 0$ on $\{p > 0\}$, which looks like an obstacle problem. Multiplying (3.9) by p, integrating w.r.t. space at a given instant in time and doing an integration by parts, for all t

$$\int_\Omega \nabla p(t, \cdot) \cdot \nabla(p(t, \cdot) + V) \leqslant \int_{\partial\Omega} p(t, \cdot)[\nabla(p(t, \cdot) + V) \cdot \mathbf{n}_\Omega], \tag{3.10}$$

where $\mathbf{n}_\Omega$ is the outward normal to Ω. As $\nabla(p + V)$ is the acceleration of the agents and they are constrained to stay in Ω, under the assumption that the latter is convex, $\nabla(p(t, \cdot) + V) \cdot \mathbf{n}_\Omega \leqslant 0$, hence the l.h.s. of (3.10) is negative. From this we immediately see that $\|\nabla p(t, \cdot)\|_{L^2(\Omega)} \leqslant$

$\|\nabla V\|_{L^2(\Omega)}$, i.e. that $p \in L^\infty((0,1); H^1(\Omega))$. Moreover, taking $m > 1$, mutliplying (3.9) by p^m, and provided that $\nabla V \in L^q(\Omega)$ with $q > d$, using Moser iterations, we are able to prove that $p(t,\cdot) \in L^\infty(\Omega)$ with a norm depending only on V and Ω. For the final pressure P_1, we only look at $D_t(\ln \rho) = \Delta\phi$. Using the equation for the terminal value of ϕ,

$$D_t(\ln \rho)(1,\cdot) = \Delta(P_1 + \Psi). \tag{3.11}$$

The l.h.s. is positive at every point x such that $\rho(1,x) = 1$, hence we get $\Delta(\Psi + P_1) \geqslant 0$ on $\{P_1 > 0\}$. From exactly the same computations, we deduce $\|\nabla P_1\|_{L^2(\Omega)} \leqslant \|\nabla \Psi\|_{L^2(\Omega)}$ and the $L^\infty(\Omega)$ norm of P_1 depends only on Ω and Ψ provided that $\nabla\Psi \in L^q(\Omega)$ with $q > d$.

Let us say that this strategy, namely looking at the convective derivative of quantities such as $\ln\rho$ was in fact already used by Loeper [Loe06] to study a problem similar to ours (related to the reconstruction of the early universe), but in a case without potential and where $\Delta p := \rho - 1$. In his case, (3.9) led to a differential inequality involving only ρ from which a L^∞ bound on ρ was deduced.

From this heuristic computation, one can guess when the same result could applied to more general Lagrangians as the question is reduced to what happens when one takes the Laplacian of the Hamilton-Jacobi equation. For instance, if we replace Ω by a Riemannian manifold, it is clear that the heuristic computation can be performed exactly in the same way provided that the manifold has a positive Ricci curvature, as the inequality involving the Laplacian of the Hamilton-Jacobi equation can be deduced from Bochner's formula.

However, this strategy seems bound to fail when applied to the more involved setting of the incompressible Euler equations (see Section 3.2 for the definition of the model). The first hint is that the regularity of the pressure depends on the potential V, which does not appear in this other setting. Moreover, here we have used that if ρ is maximal, so is $\ln\rho$. Applying the same strategy to the incompressible Euler equations, we would get (see in the next Section for the notations)

$$\int_{\mathfrak{A}} [D_{tt}(\ln \rho^\alpha)]\, \theta(\mathrm{d}\alpha) \leqslant [\text{ something with } \Delta p \text{ }],$$

but the constraint is about $\int_{\mathfrak{A}} \rho^\alpha \, \theta(\mathrm{d}\alpha)$ hence we cannot have any information on the sign of the l.h.s. of the equation above.

As in the case of soft congestion, these computations are only formal because the quantities involved cannot be differentiated twice in time, hence we have to work with a discrete version in time of the problem which is detailed in Section 3.3.

3.2 Incompressible Euler equations

3.2.1 Model and convexity of the entropy

The incompressible Euler equations aim at describing the motion of an inviscid and incompressible fluid. From the physical point of view, this system is conservative, hence one can hope to instantiate the *least action principle* and to write a variational formulation of these equations. This was done by Arnold [Arn66] with a geometric point of view: the incompressible Euler equations are seen as a geodesic equation on the infinite-dimensional manifold of measure-preserving maps. Later, Brenier introduced relaxations leading to generalized geodesics on the group of measure-preserving maps: in [Bre89], he identified the correct calculus of variations framework for this problem to make sense and admit solutions. Translated at a microscopic level, fluid particles are allowed to split and diffuse on the whole space: for a general survey, see for instance [DF12]. We will concentrate in this paper on one of Brenier's model with a flavor of Eulerian point of view introduced in [Bre99] (see also [Bre03, Section 4], [DF12, Section 1.5.3] and [AF09]).

More specifically, the goal is to study the evolution of particles subject to an *incompressibility* constraint, namely that the average distribution of particles is, at any given time, uniform. In particular, if we look just at ρ the distribution of particles, then it is a constant (proportional to the Lebesgue measure) hence we see no evolution. To be able to analyze efficiently the motion of the particles, one refines the description and looks at the individual behavior of the particles. The model goes as follows.

There are (possibly infinitely) many phases indexed by a parameter α which belongs to some probability space $(\mathfrak{A}, \mathcal{A}, \theta)$. At a fixed time t, each phase is described by its density ρ^α_t and its velocity field $\mathbf{v}^\alpha_t$, which are functions of the position x. We assume that all the densities are confined in a fixed bounded domain Ω with Lebesgue measure 1, and up to a normalization constant ρ^α_t can be seen as a probability measure on Ω. The evolution in time of the phase α is done according to the continuity equation

$$\partial_t \rho^\alpha_t + \nabla \cdot (\rho^\alpha_t \mathbf{v}^\alpha_t) = 0. \tag{3.12}$$

We assume no-flux boundary conditions on $\partial\Omega$, thus the total mass of ρ^α is preserved over time. The different phases are coupled through the incompressibility constraint: at a fixed t the density of all the different phases must sum up to the Lebesgue measure $\mathcal{L}$ (restricted to Ω). In other words, for any t we impose that

$$\int_{\mathfrak{A}} \rho^\alpha_t \theta(\mathrm{d}\alpha) = \mathcal{L}. \tag{3.13}$$

Looking at the problem from a variational point of view, we assume that the values of ρ^α_t are fixed for $t = 0$ and $t = 1$ and that the trajectories observed are those solving the following variational problem:

$$\min_{(\rho^\alpha, \mathbf{v}^\alpha)_\alpha} \left\{ \int_{\mathfrak{A}} \int_0^1 \int_\Omega \frac{1}{2} |\mathbf{v}^\alpha_t(x)|^2 \rho^\alpha_t(x) \mathrm{d}x \mathrm{d}t \theta(\mathrm{d}\alpha) \ : \ (\rho^\alpha, \mathbf{v}^\alpha) \text{ satisfies (3.12) and (3.13)} \right\}. \tag{3.14}$$

From a physical point of view, the functional which is minimized corresponds to the average (over all phases) of the integral over time of the kinetic energy, namely the global *action* of all the phases. From the point of view of this manuscript, given Theorem 2.8, the functional which is minimized is the average of the quantities $A(\rho^\alpha)$ (see equation (2.7)) where $t \mapsto \rho^\alpha_t$ is thought as a curve valued in the Wassersteins space. Without the incompressibility constraint, each phase would evolve independently and follow a geodesic in the Wasserstein space joining ρ^α_0 to ρ^α_1 (this is precisely what the Benamou-Brenier formula (2.8) says).

In Brenier's original formulation, the space $(\mathfrak{A}, \mathcal{A}, \theta)$ is the domain Ω endowed with the Lebesgue measure $\mathcal{L}$. In fact, the phase $\alpha \in \Omega$ represents the trajectory of a particle whose initial position is α. If $T : \Omega \to \Omega$ is a measure-preserving map, "classical" boundary conditions are those where ρ^α_0 is the Dirac mass located at α and ρ^α_1 is the Dirac mass located at $T(\alpha)$: it says that the particles located at $t = 0$ in α must be in $T(\alpha)$ at $t = 1$. In a classical solution, each phase α will be of the form $\rho^\alpha_t = \delta_{y^\alpha(t)}$, where $y^\alpha : [0, 1] \to \Omega$ is a curve joining α to $T(\alpha)$. But, even if one starts with "classical" boundary conditions, there are cases where the phase α may split and ρ^α may not be a Dirac mass for any $t \in (0, 1)$, leading to a "non-classical" solution (for examples of such cases, the reader can consult [Bre89, Section 6] or the detailed study [BFS09]).

It happens that all the quantities involved do not really depend on the particular dependence of the ρ^α in α. Indeed, recall that Γ is the space of continuous curves valued in the probability measures on Ω endowed with the Wasserstein distance: in short $\Gamma = C([0,1], \mathcal{P}(\Omega))$. Everything only depends on the image measure of θ through the map $\alpha \mapsto \rho^\alpha$. The natural object we are dealing with is therefore a probability measure on Γ, something that one can call (by analogy with [BCM05]) a W_2-traffic plan. We will use the letter Q to denote those W_2 traffic plans: compared to Section 3.1, where Q denoted a measure on the set of curves valued in Ω, here Q will be a measure on the set of curves valued in $\mathcal{P}(\Omega)$. In a way, the application $\alpha \mapsto \rho^\alpha$ is a parametrization of a W_2-traffic plan: that's why we will call Brenier's formulation the parametric one, while we will work in the non parametric setting, dealing directly with probability measures on Γ. In our setting, most topological properties are easier to handle, and notations are according to us simplified. Even though any probability measure on Γ cannot be *a priori* parametrized, we will show that it is the case for the solutions of the variational formulation of the Euler equations. Therefore, our results can be translated in Brenier's parametric setting.

More precisely, if $Q \in \mathcal{P}(\Gamma)$ is a W_2-traffic plan, given Theorem 2.8, the energy that we seek to minimize is

$$\min_Q \left\{ \int_\Gamma \left(\int_0^1 \frac{1}{2} |\dot\rho_t|^2 \mathrm{d}t \right) Q(\mathrm{d}\rho) \ : \ Q \in \mathcal{P}(\Gamma) \text{ and } \forall t \in [0,1], \ \int_\Gamma \rho_t Q(\mathrm{d}\rho) = \mathcal{L} \right\}. \tag{3.15}$$

and the joint law of Q at time $t \in \{0,1\}$ is fixed. The constraint is nothing else than the translation, in the setting of W_2 traffic plans, of the incompressibility constraint. The continuity equation has disappeared, it is now implicit in the definition of the action for curves valued in the Wasserstein space. Compared to the setting of the previous Section 3.1, the cost functional is just (the expectation of) the action, but one faces a continuum of curves which interact through this global incompressbility constraint. Until the end of this section, we will keep using the notations of the parametric setting, as they are more suited for the exposition (but not for the proofs!).

With formal considerations (see for instance [Bre03, Section 4]) which amount basically to write the dual formulation of this convex problem and explicit the absence of dual gap, one can be convinced that for each phase α, the optimal velocity field is the gradient of a scalar field ϕ^α (i.e. $\mathbf{v}^\alpha_t = -\nabla \phi^\alpha_t$), and that each ϕ^α evolves according to a Hamilton-Jacobi equation

$$-\partial_t \phi^\alpha_t + \frac{|\nabla \phi^\alpha_t|^2}{2} = -p_t,$$

with a pressure field p that does not depend on α and that arises from the incompressibility constraint. As discussed below, the actual regularity of the pressure is a hard and challenging question. If we look at the Boltzmann entropy (see (2.13)) of the phase α, a formal computation, which is almost the same as the one done to obtain (3.6), leads to

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \int_\Omega \rho^\alpha_t(x) \ln \rho^\alpha_t(x) \mathrm{d}x \geqslant \int_\Omega \Delta p_t(x) \rho^\alpha_t(x) \mathrm{d}x, \tag{3.16}$$

Thus, if one defines the averaged entropy $\mathcal{H}$ as a function of time by

$$\mathcal{H}(t) := \int_{\mathfrak{A}} H(\rho^\alpha_t) \mathrm{d}t = \int_{\mathfrak{A}} \left(\int_\Omega \rho^\alpha_t(x) \ln \rho^\alpha_t(x) \mathrm{d}x \right) \theta(\mathrm{d}\alpha),$$

the previous computation leads to

$$\mathcal{H}''(t) \geqslant \int_\Omega \Delta p_t = \int_{\partial\Omega} \nabla p_t \cdot \mathbf{n}_\Omega$$

where $\mathbf{n}_\Omega$ stands for the outward normal to Ω. In the setting of W_2 traffic plans, it is enough to replace $\mathcal{H}(t)$ by the integral of $H(\rho_t)$ against the measure $Q \in \mathcal{P}(\Gamma)$. Let us underline that it is crucial that the r.h.s. of (3.16) depends linearly on ρ^α, so that, integrated w.r.t. α, this dependency disappears. The same computation with a functional different from the entropy would not lead to any relevant result. At this point, notice that the convexity of Ω becomes a natural requirement. Indeed, if this is the case, the acceleration of a fluid particle located on the boundary will be directed toward the interior of Ω because the particle is constrained to stay in Ω. As the acceleration of the fluid particles is – at least heuristically – equal to $-\nabla p$, it is reasonable to expect that $\nabla p \cdot \mathbf{n}_\Omega \geqslant 0$ on $\partial\Omega$. Therefore, at a formal level, assuming the convexity of Ω leads to $\mathcal{H}'' \geqslant 0$, i.e. to the property that the averaged entropy $\mathcal{H}$ is a convex function of time. This was remarked and conjectured by Brenier in [Bre03, section 4].

Our contribution is to show that the conjecture of Brenier is true, namely that *the averaged entropy $\mathcal{H}$ is, at least for one solution of the variational formulation of the incompressible Euler equation, a convex function of time*, see Theorem 6.9 for the precise statement. This result is somewhat disappointing because we can prove convexity only for one solution and not for all: it is because we use an approximation process to prove the result, and uniqueness of the solution is known to be false in general. As for the two previous sections, the main difficulty lies in the fact that the solutions are not regular enough to make the computations rigorous, and we bypass this difficulty by introducing a time-discretization described below. As the reader will be able to see in Chapter 6, once the time discretization is performed we never have to worry about the regularity of the pressure p nor the value functions ϕ^α and the latter objects do not even appear in the proof.

Posterior to the publication of our work [Lav17], Baradat and Monsaingeon [BM18, Proposition 5.2] have provided a simpler proof of this result. The main idea is the same: perturb the solution w.r.t. the heat flow and use the result as a competitor. Because of our time discretization, what we do can be see as fixing an instant t_0 in time and then letting ρ_t^α unchanged except if $t = t_0$ where in this case we change it into $\Phi_s \rho_{t_0}^\alpha$ for a small $s > 0$. On the other hand, in [BM18], they directly work at the continuous level and change ρ_t^α in $\Phi_{st(1-t)}\rho_t^\alpha$ for some small s: their perturbation is not localized in time. Nevertheless, thanks to nice algebraic properties of the heat flow, they are able to write the derivative of the action and conclude to the convexity of the entropy. As they do not have a time-discretization procedure, they are able to retrieve the convexity of the entropy for *all* solutions of (3.14) and not only one. On the other hand, in their framework Ω is the torus and the adaptation to a convex domain seems doable, but not immediate, whereas our proof is identical whether we are in the torus or a general convex domain. Actually, as we just use the linearity of the heat flow and the EVI estimate, we believe that our proof could be copied *mutatis mutandis* in $\mathrm{RCD}(0,\infty)$ spaces. We have still included our proof in this manuscript for its similarity with the techniques in Chapters 4 and 5, but we would advise a reader which just want to have a nice proof of the convexity of the entropy to read [BM18] rather than Chapter 6.

Let us briefly mention here some already known results and directions of study of this problem not related to the convexity the entropy. With the relevant framework, existence of a solution to (3.15) is rather easy, although one has to show that the problem is not empty, which is not immediate [Bre89, Section 4]. In some cases there is no uniqueness in (3.15), we refer the reader to [BFS09] for a comprehensive study of one of such cases. Most of the research has been dedicated to the characterization of optimality conditions: the main issue is to show existence, uniqueness and regularity of a pressure field. It was accomplished by Brenier [Bre99] and regularity of the pressure was later refined by Ambrosio and Figalli [AF08], though it is believed that the current result is not sharp yet. In [AF09], the authors used the improved regularity of the pressure to get a Lagrangian interpretation, i.e. to characterize the trajectories of the different fluid particles. They also show that, from the value of the problem (3.15), one can infer a distance on the set of measure-preserving plans (i.e. elements of $\{\mu \in \mathcal{P}(\Omega \times \Omega) \ : \pi_0 \# \mu = \mathcal{L} \text{ and } \pi_1 \# \mu = \mathcal{L}\}$). We also mention the recent result [Bar19] which proves continuous dependence on the pressure w.r.t. the data, i.e. the initial and final configuration of the phases.

3.2.2 An explicit example

We end this section by explaining how, in basically the only situation where explicit solutions of (3.14) are known, one can check by hand that indeed the entropy is convex. The situation is the following: we take $\Omega = B(0,1)$ the ball of center 0 and radius 1 in dimension d with $d = 1$ or $d = 2$. The parameter space is the domain itself, i.e. $(\mathfrak{A}, \mathcal{A}, \theta)$ is the domain Ω endowed with its Borel σ-algebra and the normalized Lebesgue measure. Instead of the final time taken to be 1, for normalization reasons we rather choose it to be π. Let us put $\rho_0^\alpha = \delta_\alpha$ and $\rho_\pi^\alpha = \delta_{-\alpha}$ for $\alpha \in \Omega$. Hence the phase α must describe a particle which is located at time $t = 0$ in α and at time $t = \pi$ in $-\alpha$.

The idea is that we know what the solutions look like. In dimension $d = 2$, there exists two smooth solutions of this problem, namely the ones where the particles rotate at unit angular speed in the clockwise and counter clockwise directions. For these solutions, as the acceleration of a particle located at x is $-x$, the pressure field can be computed explicitly and is given by $p(x) = |x|^2/2$. Actually, and this was one of the main result of [Bre99], the pressure field is the same for all solutions. In dimension $d = 1$, we refer to [Bre89, BFS09] for the justification that the pressure is also equal to $p(x) = x^2/2$. Hence, whatever the solution we pick (with the boundary conditions described above), we know that particles must have an acceleration equal to $-\nabla p(x) = -x$. We introduce $\Psi : [0, \pi] \times \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ the flow of the equation $\ddot{x} = -x$, given by

$$\Psi(t, x, v) = x \cos(t) + v \sin(t).$$

If the velocity of the particle α at time $t = 0$ is given, then we will have $\rho_t^\alpha = \delta_{\Psi(t,\alpha,v)}$. In dimension 2 we can set $v = \alpha^\perp$ the rotation of α by $\pi/2$. Physically, each particle α moves in the counter clockwise direction along the circle of radius $|\alpha|$ with unit angular speed: we recover the solution described at the begining of this paragraph. Such a flow is incompressible and is a solution of (3.14); and if we make particles flow in the clockwise direction we get also an optimal incompressible flow (which shows in particular that uniqueness does not hold). But what was understood by Brenier [Bre89] and later exhaustively explored in [BFS09] is that we can allow for a phase α to diffuse: one can choose $\eta^\alpha \in \mathcal{P}(\mathbb{R}^d)$ a distribution of velocity over $\mathbb{R}^d$, depending on α, and say that it stands for the initial distribution of velocity of the phase α. The two rotations described above correspond to cases where this distribution is a Dirac mass. More precisely, let us state [BFS09, Lemma 2.3], while being sloppy about measurability issues.

Proposition 3.1. *Let $(\eta^\alpha)_{\alpha \in \Omega}$ a family of probability measures over $\mathbb{R}^d$ indexed by $\alpha \in \Omega$. For each $\alpha \in \Omega$ and each $t \in [0, \pi]$ we define*

$$\rho_t^\alpha := \Psi(t, \alpha, \cdot) \# \eta^\alpha.$$

Then $(\rho^\alpha)_{\alpha \in \Omega}$ is a solution of (3.14) if and only if it is incompressible, i.e. if and only if for all $t \in [0, \pi]$,

$$\int_\Omega \rho_t^\alpha \theta(\mathrm{d}\alpha) = \frac{\mathcal{L}}{\mathcal{L}(\Omega)}$$

and reciprocally every solution of (3.14) is of this form.

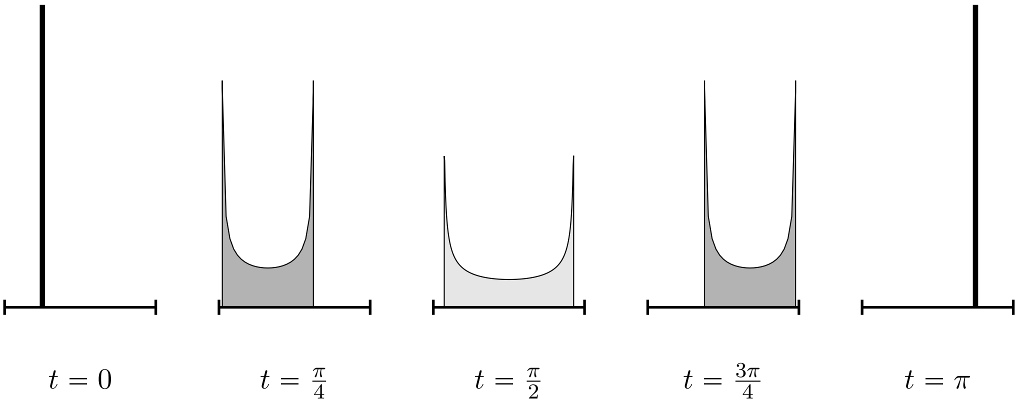

Figure 3.1: Temporal evolution, for the unique solution of the variational formulation of the Euler equations in dimension 1, of the phase $\alpha = -\frac{1}{2}$ which is a Dirac mass in $x = -\frac{1}{2}$ at $t = 0$ and a Dirac at $x = \frac{1}{2}$ at $t = \pi$. One can see that, along the evolution, the mass is spreading. Though not exactly represented, the density of this phase is not *bounded* for a fixed t. Note that the support of the density at $t = \frac{\pi}{2}$ is *not* the whole domain Ω.

We recall that $\theta = \mathcal{L}/\mathcal{L}(\Omega)$, and that, as $\mathcal{L}(\Omega) \neq 1$, we are forced to rescale the Lebesgue measure over Ω when expressing incompressibility.

Working with the family $(\eta^\alpha)_{\alpha \in \Omega}$, let us check the convexity of the entropy. Indeed, for a fixed t and α, the map $\Psi(t, \alpha, \cdot)$ is just an affine transformation with slope $\sin(t)$. Hence it is quite easy to compute

$$H(\rho_t^\alpha) = -d \ln(\sin(t)) + H(\eta^\alpha),$$

where we recall that $d \in \{1, 2\}$ is the dimension of space. Notice that the dependency on t has been decoupled from the one in η. Integrating this equality w.r.t. α,

$$\mathcal{H}(t) = -d \ln(\sin(t)) + \int_\Omega H(\eta^\alpha) \theta(\mathrm{d}\alpha).$$

Now if we evaluate this identity at $t = \pi/2$, we can conclude that

$$\mathcal{H}(t) = -d \ln(\sin(t)) + \mathcal{H}\left(\frac{\pi}{2}\right). \tag{3.17}$$

In this identity, $\mathcal{H}$ can be computed with any solution of (3.14), but of course $\mathcal{H}(\pi/2)$ depends on the solution and may be infinite. This is the case in the clockwise and counter clockwise rotations describe above. Even in these cases, (3.17) tells us that $\mathcal{H}(t)$ is identically $+\infty$, which is true but not really relevant.

On the other hand, if $\mathcal{H}(\pi/2) < +\infty$, then (3.17) shows that $\mathcal{H}$ is indeed convex, and also belongs to $L^1([0, \pi])$ as such properties are true for $t \mapsto -\ln \sin(t)$. In particular, the latter property shows that we fall under Assumption 6.1 which we will make later in Chapter 6.

We confess that we do not know, if $d = 2$, whether there are explicit solutions, i.e. explicit families $(\eta^\alpha)_{\alpha \in \Omega}$ for which $\mathcal{H}(\pi/2) < +\infty$. Indeed, the families proposed in [BFS09] are concentrated on one dimensional sets, hence have an infinite entropy[1]. However, if $d = 1$, i.e. when $\Omega = [-1, 1]$, then there is only one solution [BFS09, Theorem 3.1] and it is given by

$$\eta^\alpha(\mathrm{d}v) = \frac{\mathbb{1}_{\alpha^2 + v^2 \leqslant 1}(v)}{\pi\sqrt{1 - \alpha^2 - v^2}} \mathrm{d}v.$$

[1] One could try to compute a convex combination of different solutions given in [BFS09] to build one with finite entropy, but the computations are quite heavy and we did not have the courage to finish them.

One can look at Figure 3.1 to understand what the evolution of the phase α (with $\alpha = -1/2$) looks like. As $\Psi(\pi/2, \alpha, \cdot)$ is just the identity, $\rho^\alpha_{\pi/2}(\mathrm{d}x) = \eta^\alpha(\mathrm{d}x)$. Hence, the total entropy at time $\pi/2$ can be written

$$\mathcal{H}(\pi/2) = -\frac{1}{\mathcal{L}(\Omega)} \iint_{\mathbb{R}^2} \frac{\ln(f(1-\alpha^2-v^2))}{f(1-\alpha^2-v^2)} \mathrm{d}\alpha \mathrm{d}v,$$

where we just set $f(s) = \pi\sqrt{s}\mathbb{1}_{s>0}$. Doing a polar change of variables and calling $r^2 = \alpha^2 + v^2$, one can check that the integral is finite: the only issue would be close to $r = 1$, but we have the expansion $\ln f(1-r^2)/f(1-r^2) \sim C \ln(1-r)(1-r)^{-1/2}$. As a conclusion, *if we look at the inversion of the segment in dimension $d = 1$, then the unique solution of* (3.14) *is such that* $\mathcal{H} \in L^1([0,\pi])$ *and $\mathcal{H}$ is indeed a convex function of time.*

3.3 Time discretization and flow interchange

In the next three chapters, there is a common feature: the use of a time discretization to make rigorous estimates established formally *via* time differentiation. As the technicalities involved are very similar in all three chapters, we will try, in this subsection, to give a flavor of them.

3.3.1 The discrete problem

Let us explain how one can discretize a problem like (3.1). We choose $N + 1 \geqslant 2$ an integer which will denote the number of time steps. We will write $\tau := 1/N$ for the distance between two time steps. The set T^N will stand for the set of all time steps, namely

$$T^N := \{k\tau;\ k = 0, 1, \dots, N\}.$$

We set $\Gamma_N := \mathcal{P}(\Omega)^{T^N} \simeq \mathcal{P}(\Omega)^{N+1}$: i.e. an element $\rho \in \Gamma_N$ is a $N+1$-uplet $(\rho_0, \rho_\tau, \dots, \rho_1)$ of probability measures indexed by T^N. Given (2.4), a natural discretization of the action is

$$\int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t \simeq \sum_{k=1}^N \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau}.$$

Hence, the continuous problem (3.1) will be replaced by

$$\min_{\rho \in \Gamma_N} \left\{ \sum_{k=1}^N \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} + \sum_{k=1}^N \tau E(\rho_{k\tau}) + \Psi(\rho_1) \ :\ \rho_0 \text{ given} \right\}. \tag{3.18}$$

Existence of solutions to (3.18) is easy to get, but the main interest of this problem is the way optimality conditions are written. Indeed, let $\hat{\rho}$ be a solution of (3.18) and fix $k \in \{1, 2, \dots, N-1\}$. We use the shortcut $\bar{\rho} := \hat{\rho}_{k\tau}$. Let us also denote $\mu := \hat{\rho}_{(k-1)\tau}$ and $\nu := \hat{\rho}_{(k+1)\tau}$ the values of the curve at the previous and next time step respectively. By optimality we know that $\bar{\rho}$ is a minimizer (among all probability measures) of

$$\rho \mapsto \frac{W_2^2(\mu, \rho) + W_2^2(\rho, \nu)}{2\tau} + \tau E(\rho). \tag{3.19}$$

Notice that this is an instance of the toy model (1.4) presented in Chapter 1. The key idea, which was introduced precisely in the context of the JKO scheme under the name *flow interchange* [MMS09], is to take F a function convex along generalized geodesics and to use $S^F_t \bar{\rho}$ (the

Wasserstein gradient flow of F starting at $\bar{\rho}$, see Theorem 2.11) as a competitor. Indeed, using the inequality (2.10) and saying that $S^F_t \bar{\rho}$ cannot do better than $\bar{\rho}$, one ends up with

$$\frac{F(\mu) + F(\nu) - 2F(\bar{\rho})}{\tau^2} \geqslant - \left. \frac{\mathrm{d}}{\mathrm{d}t} E\left(S^F_t \bar{\rho}\right) \right|_{t=0} \tag{3.20}$$

The l.h.s. is nothing else than the discrete second derivative in time of the quantity $k \mapsto F(\hat{\rho}_{k\tau})$. Taking $E = 0$, we recover that the latter quantity is a convex function of time, which is precisely the geodesic convexity of F.

As far as the study of soft congestion is concerned, it is enough to take $F = U_m$ for some $m \geqslant 1$. Indeed, as recall in Section 2.2, one can write a PDE satisfy by the gradient flow $S^{U_m} \bar{\rho}$ and evaluate precisely the rate of dissipation of E along S^{U_m}. Once the computation is done, what we get is exactly a discrete version in time of (3.6). Actually, for the case of soft congestion we write explicitly the optimality conditions of (3.19): calling $(\tilde{\varphi}, \varphi)$ and $(\psi, \tilde{\psi})$ pairs of Kantorovich potentials between $\mu, \bar{\rho}$ and $\bar{\rho}, \nu$ respectively, they read

$$\frac{\varphi + \psi}{2\tau^2} + \frac{\delta E}{\delta \rho}(\bar{\rho}) = \text{[constant]}. \tag{3.21}$$

The presence of the Kantorovich potentials should not be surprising: they appear as the derivative of the Wasserstein distance w.r.t. ρ, see Proposition 2.3. Then we multiply this optimality condition by the relevant quantity, which is nothing than the gradient (in the Wasserstein space) of U_m at the point $\bar{\rho}$, and we integrate w.r.t. Ω to get (3.20).

For the study of hard congestion, we do not rely on a flow interchange estimate, though we use the optimality conditions (3.21) of the discrete problem. Indeed, the idea is to translate all the formal computations at the discrete level, with the Hamilton-Jacobi equation being translated in (3.21).

Eventually, for the incompressible Euler equations, we use the flow interchange estimate, as we perturb each curve by making the component at time $k\tau$ follow the heat flow, which is nothing else than the Wasserstein gradient flow of the entropy. By doing that, we preserve the incompressibility constraint: this is just a consequence of the linearity of the heat flow and the fact that the Lebesgue measure is invariant under the heat flow. Hence, when we write (3.20) there is no r.h.s. Integrating w.r.t. all the phases leads to the discrete time-convexity of the entropy.

There is actually a technical refinement present in all of the three chapters: in the discrete problem, we add a vanishing entropic penalization, namely

$$\lambda \sum_{k=0}^{N} \tau H(\rho_{k\tau}),$$

where we recall that H is the Boltzmann entropy, see (2.13). The parameter λ is then sent to 0, together or after that $N \to +\infty$. The goal of this entropic penalization is twofold:

It will force the minimizers of the discrete problem to be measures with strictly positive density a.e.: this comes from the fact that the derivative of $x \mapsto x \ln x$ at $x = 0$ is $-\infty$. As seen for instance in Proposition 2.3, it is great help to handle derivatives of the Wasserstein distance. We think that it reveals an unavoidable issue: when one studies optimal transport, *it is hard to handle the regions where there is no mass, because in these regions mass can only be added, not removed* (hence all the optimality conditions with an inequality which becomes an equality where there is mass). The role of entropic penalization is precisely to remove this potential issue and greatly helps to write the optimality conditions. Of course, one has to pay a price: passing to the limit $N \to +\infty$ with entropic penalization is more involved than without.

On the other hand, in the case of the incompressible Euler equations, it will guarantee convergence of the discrete entropy to the continuous one. Such a convergence is in fact not true *a priori* (because H is only l.s.c. on $\mathcal{P}(\Omega)$) but necessary if one wants to pass to the limit a feature such as convexity.

3.3.2 Passing to the limit

In each chapter, the strategy is always to prove estimates at the discrete level, i.e. for problem (3.18) and then to pass to the limit $N \to +\infty$ to get estimates that are true at the continuous level, i.e. for problem (3.1). To achieve this end, we basically prove a Γ-limit. We first emphasize that, in case of non-uniquness in the limit problem, we can prove something only for *one* solution of the continuous problem, not for all. It happens for the convexity of the entropy in the variational formulation of the incompressible Euler equations and for the regularity of the pressure in the case of hard congestion (as there is no uniqueness in the dual problem defining the pressure).

For the $\Gamma - \liminf$, the first step is to identify competitors at the discrete level with competitors at the continuous one. This is in fact quite easy: if one has $\rho \in \Gamma_N$ (i.e. one knows the value of ρ_t only for $t \in \{0, \tau, 2\tau, \dots, 1\}$), then by interpolating along constant-speed geodesics on each segment $[k\tau, (k+1)\tau]$ one easily get a competitor in Γ. Along this process, the discrete action is equal to the continuous one. Hence standard lower semi-continuity arguments allow to handle the limit $N \to +\infty$.

On the other hand, the $\Gamma - \limsup$ is done by sampling a continuous curve to get a discrete one. The only issue that might appear is the presence of entropic regularization: indeed, in this case, one must first regularize a curve before sampling it to ensure a control on the discrete entropic penalization term.

We mention that for hard congestion, in Chapter 5, what we do in rather pass to the limit in the dual problem, as the pressure P is a dual variable. As we already know that we have convergence of the values of the problem (because of the convergence of the primal problem and the absence of duality gap), it is enough to show that the limiting pressure does at least as good, when evaluated in the continuous dual problem, as the solution of the continuous primal problem.

Chapter 4

Regularity of the density in the case of soft congestion

In this chapter, we tackle the problem of optimal density evolution with soft congestion which reads

$$\min_{\rho}\left\{\int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t + \int_0^1 E(\rho_t)\mathrm{d}t + \Psi(\rho_1) \ : \ \rho \in \Gamma, \ \rho_0 \text{ given}\right\}.$$

where $\Gamma = C([0,1], \mathcal{P}(\Omega))$ and $|\dot{\rho}_t|$ is the metric derivative of ρ. The functional $E : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ will have the form

$$E(\rho) = \int_\Omega f(\rho) + \int_\Omega V \mathrm{d}\rho$$

The goal is to show that the optimal ρ is in L^∞ globally in space, locally in time provided we can quantify how much convex f is and V has some regularity.

4.1 Statement of the problem and regularity of the density

Assumptions. *The assumptions that will hold throughout this chapter are the following.*

- *Recall that Ω is the closure of an open convex bounded domain with smooth boundary. To simplify the constants, we assume that its Lebesgue measure is* 1.
- *We assume that $f : [0, +\infty) \to \mathbb{R}$ is a strictly convex function, bounded from below and C^2 on $(0, +\infty)$. We define the congestion penalization F by, for any $\rho \in \mathcal{P}(\Omega)$,*

  $$F(\rho) := \int_\Omega f(\rho^{ac}) + f'(+\infty)\rho^{sing}(\Omega),$$

  where $\rho =: \rho^{ac}\mathcal{L} + \rho^{sing}$ is the decomposition of ρ as an absolutely continuous part ρ^{ac} (identified with its density) and a singular part ρ^{sing} w.r.t. $\mathcal{L}$. Thanks to Proposition 2.12, we know that F is a convex l.s.c. functional on $\mathcal{P}(\Omega)$.
- *We assume that $V : \Omega \to \mathbb{R}$ is a Lipschitz function.*
- *We assume that $\Psi : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ is a l.s.c. and convex functional, bounded from below.*

We will consider variational problems with a running cost of the form

$$\rho \mapsto E(\rho) := F(\rho) + \int_\Omega V \mathrm{d}\rho$$

while Ψ will penalize the final density, and the initial one will be prescribed. Namely, we fix $\bar{\rho}_0$ a fixed element of $\mathcal{P}(\Omega)$. We recall that $\Gamma := C([0,1], \mathcal{P}(\Omega))$ where $\mathcal{P}(\Omega)$ is endowed with the Wasserstein metric W_2, and that the metric derivative of a 2-absolutely continuous curve is defined in Theorem 2.7.

Definition 4.1. *We define the the functional $\mathcal{A} : \Gamma \to \mathbb{R}$ by*

$$\mathcal{A}(\rho) := \int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t + \int_0^1 E(\rho_t)\mathrm{d}t + \Psi(\rho_1).$$

We state the continuous problem as

$$\min\{\mathcal{A}(\rho) \;:\; \rho \in \Gamma,\ \rho_0 = \bar{\rho}_0\}. \tag{4.1}$$

A curve $\rho \in \Gamma$ with $\bar{\rho}_0$ that minimizes $\mathcal{A}$ will be called a solution of the continuous problem.

Proposition 4.2. *Let us assume that there exists $\rho \in \Gamma$ with $\rho_0 = \bar{\rho}_0$ such that $\mathcal{A}(\rho) < +\infty$. Then the problem* (4.1) *admits a unique solution.*

Proof. The functional $\mathcal{A}$ is the sum of l.s.c., convex and bounded functionals. Moreover, as $\mathcal{A}(\rho) \geqslant \int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t - C$ (where C depends on the lower bounds of f, V and Ψ), we know, thanks to Proposition 2.9, that the sublevel sets of $\mathcal{A}$ are compact. The existence of a solution to (4.1) follows from the direct method of calculus of variations.

To prove uniqueness, we need to prove that $\mathcal{A}$ is strictly convex. If ρ^1 and ρ^2 are two distinct minimizers of $\mathcal{A}$, we define $\rho := (\rho^1 + \rho^2)/2$. As ρ^1 and ρ^2 are distinct, by continuity there exists $T_1 < T_2$ such that ρ^1_t and ρ^2_t differ for every $t \in [T_1, T_2]$. In particular, for any $t \in [T_1, T_2]$, by strict convexity of F, $F(\rho) < (F(\rho^1) + F(\rho^2))/2$. Thus,

$$\int_0^1 F(\rho_t)\mathrm{d}t < \frac{1}{2}\int_0^1 F(\rho^1_t)\mathrm{d}t + \frac{1}{2}\int_0^1 F(\rho^2_t)\mathrm{d}t.$$

As all the other terms appearing in $\mathcal{A}$ are convex, one concludes that $\mathcal{A}(\rho) < (\mathcal{A}(\rho^1) + \mathcal{A}(\rho^2))/2$, which contradicts the optimality of ρ^1 and ρ^2. □

As we will be interested in the regularity of the solutions of (4.1), we will not discuss the existence of admissible competitors, i.e. the existence of $\rho \in \Gamma$ with $\rho_0 = \bar{\rho}_0$ such that $\mathcal{A}(\rho) < +\infty$. However, let us just say that if $f(s)$ growths at most like s^m as $s \to +\infty$ with $m < 1 + 1/d$ (where d is the dimension of the space), and if $\Psi(\mathcal{L}) < +\infty$, then existence of such a ρ is guaranteed for any $\bar{\rho}_0 \in \mathcal{P}(\Omega)$. Indeed, by convexity of $\mathcal{A}$ it is enough to check that $\mathcal{A}(\rho)$ is finite if ρ is the geodesic joining a Dirac mass at time $t = 0$ to the Lebesgue measure at time $t = 1$.

In order to get the L^∞ bounds, we will consider two different cases (strong and weak congestion), depending on the second derivative of f. Let us start by introducing the typical functions f that we will consider.

Definition 4.3. *For any $m \geqslant 1$, we define $u_m : [0, +\infty) \to \mathbb{R}$ for any $t \geqslant 0$ through*

$$u_m(t) := \begin{cases} t \ln t + 1 & \text{if } m = 1 \\ \dfrac{t^m}{m(m-1)} & \text{if } m > 1 \end{cases}.$$

For any $m \geqslant 1$, the functional $U_m : \mathcal{P}(\Omega) \to \mathbb{R}$ is defined, for $\rho \in \mathcal{P}(\Omega)$, via

$$U_m(\rho) := \begin{cases} \displaystyle\int_\Omega u_m(\rho) & \text{if } \rho \text{ is absolutely continuous w.r.t. } \mathcal{L} \\ +\infty & \text{else} \end{cases}.$$

One can notice that $u_m''(t) = t^{m-2}$ for any $m \geqslant 1$ and any $t > 0$, hence the functions u_m are convex for all m. One can also notice that U_1 is (up to an additive constant) the entropy w.r.t. $\mathcal{L}$ that we already defined in (2.13). Moreover, thanks to Proposition 2.12, we see that U_m is l.s.c., convex and convex along generalized geodesics in $\mathcal{P}(\Omega)$. Let us underline also that a direct application of Jensen's inequality yields $m^2 U_m$ for any $m \geqslant 1$.

Let us now state the different assumptions to quantify how much F penalizes concentrated measures.

Assumption 4.1 (strong congestion). *There exists $\alpha \geqslant -1$ and $C_f > 0$ such that $f''(t) \geqslant C_f t^\alpha$ for any $t > 0$.*

Assumption 4.2 (strong congestion-variant). *There exist $\alpha \geqslant -1$, $t_0 > 0$ and $C_f > 0$ such that $f''(t) \geqslant C_f t^\alpha$ for any $t \geqslant t_0$.*

In particular, integrating twice, we see that under either of the above assumptions, for $\rho \in \mathcal{P}(\Omega)$ we have $U_{\alpha+2}(\rho) \leqslant C_f F(\rho) + C$, where C is a constant that depends on f (but not on ρ). One can also see that $f'(+\infty) = +\infty$. The function u_m is the typical example of a function satisfying Assumption 4.1 with $\alpha = m - 2$. To produce functions satisfying Assumption 4.2 but not Assumption 4.1, think at $f(t) = \sqrt{1 + t^4}$ (if we try to satisfy Assumption 4.1 we need $\alpha \leqslant 0$ for large t, and $\alpha \geqslant 2$ for small t) or at $f(t) = (t - 1)_+^2$ (the difference between these two examples is that in the first case on could choose an aribtrary $t_0 > 0$, while in the second it is necessary to use $t_0 \geqslant 1$).

Assumption 4.3 (weak congestion). *There exist $\alpha < -1$, $t_0 > 0$ and $C_f > 0$ such that $f''(t) \geqslant C_f t^\alpha$ for any $t \geqslant t_0$.*

For example, $f(t) := \sqrt{1 + t^2}$ satisfies $f''(t) \geqslant C_f t^\alpha$ for $t \geqslant 1$ with $\alpha = -3$.

Assumption 4.4 (higher regularity of the potential). *The potential V is of class $C^{1,1}$ (it is C^1 and its gradient is Lipschitz) and $\nabla V \cdot \mathbf{n}_\Omega \geqslant 0$ on $\partial\Omega$, where $\mathbf{n}_\Omega$ is the outward normal to Ω.*

We will see that only Assumption 4.1, where we require a control of f'' everywhere, allows to deal with Lipschitz potentials, while in general we will need the use of Assumption 4.4. The condition $\nabla V \cdot \mathbf{n}_\Omega \geqslant 0$ on $\partial\Omega$ can be interpreted by the fact that the minimum of V is reached in the interior of Ω: it prevents the mass of ρ to concentrate on the boundaries.

Assumption 4.5 (final penalization). *The penalization Ψ is of the following form*

$$\Psi(\rho_1) = \begin{cases} \displaystyle\int_\Omega g(\rho_1) + \int_\Omega W \mathrm{d}\rho_1 & \text{if } \rho_1 \text{ is absolutely continuous w.r.t. } \mathcal{L} \\ +\infty & \text{if } \rho_1 \text{ is singular w.r.t. } \mathcal{L}, \end{cases}$$

where $g : [0, +\infty) \to \mathbb{R}$ is a convex and superlinear (i.e. $g'(+\infty) = +\infty$) function, bounded from below, and $W : \Omega \to \mathbb{R}$ is a potential of class $C^{1,1}$ satisfying $\nabla W \cdot \mathbf{n}_\Omega \geqslant 0$ on $\partial\Omega$.

The mains results of this chapter can be stated as follows.

Theorem 4.4 (strong congestion, interior regularity). *Suppose that either Assumption 4.1 holds or Assumption 4.2 and 4.4 hold, and that $\mathcal{A}(\rho) < +\infty$ for some $\rho \in \Gamma$ with $\rho_0 = \bar{\rho}_0$. Let ρ be the unique solution to* (4.1)*. Then for any $0 < T_1 < T_2 < 1$, the restriction of ρ to $[T_1, T_2]$ belongs to $L^\infty([T_1, T_2] \times \Omega)$.*

Theorem 4.5 (strong congestion, boundary regularity). *Suppose that either Assumption 4.1 holds or Assumption 4.2 and 4.4 hold, and that Assumption 4.5 holds as well, and that* $\mathcal{A}(\rho) < +\infty$ *for some* $\rho \in \Gamma$ *with* $\rho_0 = \bar{\rho}_0$*. Let* ρ *be the unique solution to* (4.1)*. Then, for any* $0 < T_1 < 1$*, the restriction of* ρ *to* $[T_1, 1]$ *belongs to* $L^\infty([T_1, 1] \times \Omega)$*.*

Theorem 4.6 (weak congestion case). *Suppose Assumptions 4.5, 4.3 and 4.4 hold and that* $\mathcal{A}(\rho) < +\infty$ *for some* $\rho \in \Gamma$ *with* $\rho_0 = \bar{\rho}_0$*. We assume that the prescribed initial measure* $\bar{\rho}_0$ *satisfies* $\bar{\rho}_0 \in L^{m_0}$ *with* $m_0 > d|\alpha + 1|/2$ *and* $F(\bar{\rho}_0) < +\infty$*, and that* $\|\Delta V\|_\infty$*,* $\|\Delta W\|_\infty$ *are small enough (smaller than a constant that depends on* m_0*). Let* ρ *be the unique solution to* (4.1)*. Then* $\rho \in L^{m_0}([0, 1] \times \Omega)$ *and for any* $0 < T_1 < 1$*, the restriction of* ρ *to* $[T_1, 1]$ *belongs to* $L^\infty([T_1, 1] \times \Omega)$*.*

Actually, in the last theorem, the constant should also depend on the Lebesgue measure of Ω but we do not see it as we have normalized Ω to have unit Lebesgue measure.

The rest of the chapter is devoted to the proof of these theorems. In particular, we will always assume in the sequel that there exists $\rho \in \Gamma$ with $\rho_0 = \bar{\rho}_0$ such that $\mathcal{A}(\rho) < +\infty$. In order to prove these theorems, we will introduce a discrete (in time) variational problem that will approximate the continuous one. For this problem, we will be able to show the existence of a unique smooth (in space) solution and write down the optimality conditions. From these optimality conditions, we will be able to derive a *flow interchange* estimate whose iteration will give uniform (in the approximation parameters, and in m) L^m estimates.

Let us introduce the discrete problem here. As explained in the previous chapter, we will use two approximations parameters:

- $N + 1 \geqslant 2$ will denote the number of time steps. We will write $\tau := 1/N$ for the distance between two time steps. The set T^N will stand for the set of all time steps, namely
$$T^N := \{k\tau;\ k = 0, 1, \dots, N\}.$$
We set $\Gamma_N := \mathcal{P}(\Omega)^{T^N} \simeq \mathcal{P}(\Omega)^{N+1}$ (i.e. an element $\rho \in \Gamma_N$ is a $N + 1$-uplet $(\rho_0, \rho_\tau, \dots, \rho_1)$ of probability measures indexed by T^N).
- We will also add a (vanishing) entropic penalization (recall that U_1 denotes the entropy w.r.t. $\mathcal{L}$). It will ensure that the solution of the discrete problem is smooth. The penalization will be a discretized version of
$$\lambda \int_0^1 U_1(\rho_t)\mathrm{d}t,$$
where λ is a parameter that will be sent 0.

Let us state formally our problem. We fix $N \geqslant 1$ ($\tau := 1/N$) and $\lambda > 0$, and we set $\lambda_N = \lambda$ if Assumption 4.5 is satisfied, $\lambda_N = 0$ otherwise. We define $\mathcal{A}^{N,\lambda} : \Gamma_N \to \mathbb{R}$ by

$$\mathcal{A}^{N,\lambda}(\rho) := \sum_{k=1}^{N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} + \sum_{k=1}^{N-1} \tau \left(E(\rho_{k\tau}) + \lambda U_1(\rho_{k\tau})\right) + \Psi(\rho_1) + \lambda_N U_1(\rho_1).$$

This means that in the case of Assumption 4.5 we penalize ρ_1 by $\int_\Omega g(\rho_1) + \lambda U_1(\rho_1) + \int_\Omega W \mathrm{d}\rho_1$, while we do not modify the boundary condition otherwise (the reason for not always adding $\lambda U_1(\rho_1)$ lies in the possibility of having a prescribed value for ρ_1 with infinite entropy). In all the cases, we enforce strictly $\rho_0 = \bar{\rho}_0$. The discrete minimization problem reads

$$\min\{\mathcal{A}^{N,\lambda}(\rho)\ :\ \rho \in \Gamma_N,\ \rho_0 = \bar{\rho}_0\}, \tag{4.2}$$

and a $\rho \in \Gamma_N$ which minimizes $\mathcal{A}^{N,\lambda}$ will be called a solution of (4.2).

Theorem 4.7. *For any $N \geqslant 1$ and any $\lambda > 0$, the discrete problem* (4.2) *admits a solution.*

Proof. The functional $\mathcal{A}^{N,\lambda}$ is a sum of convex and l.s.c. functionals, bounded from below, hence it is itself convex, l.s.c. and bounded from below. Moreover, the space $\Gamma_N = \mathcal{P}(\Omega)^{N+1}$ is compact (for the weak convergence). Thus, to use the direct method of calculus of variations, it is enough to show that $A^{N,\lambda}(\rho) < +\infty$ for some $\rho \in \Gamma_N$.

This is easy in this discrete framework: just take $\rho_{k\tau} = \mathcal{L}$ if $k \in \{1, 2, \ldots, N-1\}$, $\rho_0 = \bar{\rho}_0$ and $\rho_{N\tau}$ equal to an arbitrary measure ρ such that $\Psi(\rho) + \lambda_N U_1(\rho) < +\infty$. □

We did not adress the uniqueness of the minimizer in the above problem since we do not really care about it, but indeed it also holds. Indeed, the strict convexity of F (or the term λU_1 that we added) guarantees uniqueness of $\rho_{k\tau}$ for all $k \leqslant N-1$. The uniqueness of the last measure (which cannot be deducted from strict convexity for an arbitrary functional Ψ, as we do not always add a term of the form $\lambda U_1(\rho_1)$) can be obtained from the strict convexity of the last Wasserstein distance term $\rho \mapsto W_2^2(\rho, \rho_{(N-1)\tau})$, as $\rho_{(N-1)\tau}$ is absolutely continuous (see [San15, Proposition 7.19]).

In all the following, for any $N \geqslant 1$ and $\lambda > 0$, we denote by $\bar{\rho}^{N,\lambda} \in \Gamma_N$ the unique solution of (4.2) with parameters N and λ. Moreover, In all the sequel, we fix $1 < \beta < d/(d-2)$. It is well known that the space $H^1(\Omega)$ is continuously embedded into $L^{2\beta}(\Omega)$. Moreover, in the case where the assumptions of Theorem 4.6 are satisfied, we choose β in such a way that

$$\frac{\beta}{\beta - 1} m_0 > |\alpha + 1|. \tag{4.3}$$

4.2 Flow interchange estimate

4.2.1 Interior flow interchange

In this subsection, we study the optimality conditions of (4.2) away from the temporal boundaries. We fix for the rest of the subsection $N \geqslant 1$, $0 < \lambda \leqslant 1$ and $0 < k < N$, and we use the shortcut $\bar{\rho} := \bar{\rho}_{k\tau}^{N,\lambda}$. Let us also denote $\mu := \bar{\rho}_{(k-1)\tau}^{N,\lambda}$ and $\nu := \bar{\rho}_{(k+1)\tau}^{N,\lambda}$. As $\bar{\rho}^{N,\lambda}$ is a solution of the discrete problem, we know that $\bar{\rho}$ is a minimizer (among all probability measures) of

$$\rho \mapsto \frac{W_2^2(\mu, \rho) + W_2^2(\rho, \nu)}{2\tau} + \tau \left(F(\rho) + \lambda U_1(\rho) + \int_\Omega V \mathrm{d}\rho \right).$$

In particular, we know that $U_1(\bar{\rho}) < +\infty$, thus $\bar{\rho}$ is absolutely continuous w.r.t. $\mathcal{L}$.

Lemma 4.8. *The density $\bar{\rho}$ is strictly positive a.e.*

Proof. For $0 < \varepsilon < 1$, we define $\rho_\varepsilon := (1-\varepsilon)\bar{\rho} + \varepsilon \mathcal{L}$. As $\mathcal{L}$ is a probability measure, we know that ρ_ε is a probability measure too. Thus, using ρ_ε as a competitor, we get

$$\lambda(U_1(\bar{\rho}) - U_1(\rho_\varepsilon)) \leqslant \frac{W_2^2(\mu, \rho_\varepsilon) + W_2^2(\rho_\varepsilon, \nu)}{2\tau} + \tau E(\rho_\varepsilon) - \frac{W_2^2(\mu, \bar{\rho}) + W_2^2(\bar{\rho}, \nu)}{2\tau} - \tau E(\bar{\rho}).$$

We estimate the r.h.s. by convexity (as W_2^2 and F are convex) to see that

$$U_1(\bar{\rho}) - U_1(\rho_\varepsilon) \leqslant \frac{\varepsilon}{\lambda} \left(\frac{W_2^2(\mu, \mathcal{L}) + W_2^2(\mathcal{L}, \nu)}{2\tau} + \tau E(\mathcal{L}) - \frac{W_2^2(\mu, \bar{\rho}) + W_2^2(\bar{\rho}, \nu)}{2\tau} - \tau E(\bar{\rho}) \right).$$

Thus, there exists a constant C, independent of ε, such that $U_1(\bar{\rho}) - U_1(\rho_\varepsilon) \leqslant C\varepsilon$. This can be easily seen to imply (see for instance the proof of [San15, Lemma 8.6]) that $\bar{\rho}$ is strictly positive a.e. □

We can then write the first-order optimality conditions.

Proposition 4.9. *The measure $\bar{\rho}$ (or more precisely its density w.r.t. $\mathcal{L}$) is Lipschitz and bounded away from 0 and ∞. Moreover, let us denote by φ_μ and φ_ν the Kantorovich potentials for the transport from $\bar{\rho}$ to respectively μ and ν. Then the following identity holds a.e.:*

$$\frac{\nabla\varphi_\mu + \nabla\varphi_\nu}{\tau^2} + \left(f''(\bar{\rho}) + \frac{\lambda}{\bar{\rho}}\right)\nabla\bar{\rho} + \nabla V = 0. \tag{4.4}$$

Proof. Let $\tilde{\rho} \in \mathcal{P}(\Omega) \cap L^\infty(\Omega)$ and for $0 < \varepsilon < 1$ define $\rho_\varepsilon = (1-\varepsilon)\bar{\rho} + \varepsilon\tilde{\rho}$. We use ρ_ε as a competitor. We use Proposition 2.3 as $\bar{\rho} > 0$ a.e., the Kantorovich potentials φ_μ and φ_ν for the transport from $\bar{\rho}$ to respectively μ and ν are unique and

$$\lim_{\varepsilon\to 0} \frac{W_2^2(\mu,\bar{\rho}) - W_2^2(\mu,\rho_\varepsilon) + W_2^2(\bar{\rho},\nu) - W_2^2(\rho_\varepsilon,\nu)}{2\tau^2} = \int_\Omega \frac{\varphi_\mu + \varphi_\nu}{\tau}(\bar{\rho} - \tilde{\rho}).$$

The term involving V is straightforward to handle as it is linear. Hence, by optimality of $\bar{\rho}$ we get

$$\int_\Omega \left(\frac{\varphi_\mu + \varphi_\nu}{\tau^2} + V\right)(\bar{\rho} - \tilde{\rho}) \leqslant \liminf_{\varepsilon\to 0} \frac{F(\rho_\varepsilon) + \lambda U_1(\rho_\varepsilon) - F(\bar{\rho}) - \lambda U_1(\bar{\rho})}{\varepsilon}. \tag{4.5}$$

By definition of the objects involved,

$$\frac{F(\rho_\varepsilon) + \lambda U_1(\rho_\varepsilon) - F(\bar{\rho}) - \lambda U_1(\bar{\rho})}{\varepsilon} = \int_\Omega \frac{f_\lambda[(1-\varepsilon)\bar{\rho} + \varepsilon\tilde{\rho}] - f_\lambda[\bar{\rho}]}{\varepsilon}.$$

The integrand of the integral of the r.h.s. converges pointewisely, as $\varepsilon \to 0$, to $(f'(\bar{\rho}) + \lambda\ln\bar{\rho})(\tilde{\rho} - \bar{\rho})$. Moreover, as the function f_λ is convex, we see that for $0 < \varepsilon < 1$,

$$\frac{f_\lambda[(1-\varepsilon)\tilde{\rho} + \varepsilon\bar{\rho}] - f_\lambda[\bar{\rho}]}{\varepsilon} \leqslant f_\lambda(\tilde{\rho}) - f_\lambda(\bar{\rho}).$$

As $\tilde{\rho} \in L^\infty(\Omega)$ and $F(\bar{\rho}) + \lambda U_1(\bar{\rho}) < +\infty$, the r.h.s. of the equation is integrable on Ω. Thus, by a reverse Fatou's lemma,

$$\limsup_{\varepsilon\to 0} \int_\Omega \frac{F(\rho_\varepsilon) + \lambda U_1(\rho_\varepsilon) - F(\bar{\rho}) - \lambda U_1(\bar{\rho})}{\varepsilon} \leqslant \int_\Omega \left(f'(\bar{\rho}) + \lambda\ln\bar{\rho}\right)(\tilde{\rho} - \bar{\rho}).$$

Combing this equation with (4.5), we see that $\int_\Omega h \mathrm{d}(\tilde{\rho} - \bar{\rho}) \geqslant 0$ with

$$h := \frac{\varphi_\mu + \varphi_\nu}{\tau^2} + f'(\bar{\rho}) + \lambda\ln\bar{\rho} + V.$$

We know that h is finite a.e., thus its essential infimum cannot be $+\infty$. Moreover, starting from $\bar{\rho}f'(\bar{\rho}) \geqslant f(\bar{\rho}) - f(0)$, we see that $\int_\Omega h\bar{\rho} > -\infty$. Taking probability measures $\tilde{\rho}$ concentrated on sets where h is close to its essential infimum, we see that the essential infimum of h cannot be $+\infty$ and that h coincides with its essential infimum $\bar{\rho}$-a.e. As $\bar{\rho} > 0$ a.e., there exists C such that we have a.e. on Ω

$$f'(\bar{\rho}) + \lambda\ln\bar{\rho} = C - \frac{\varphi_\mu + \varphi_\nu}{\tau^2} - V. \tag{4.6}$$

As f' is C^1 and increasing, it is easy to see that $f' + \lambda\ln$ is an homeomorphism of $(0,+\infty)$ on $(-\infty,+\infty)$ which is bilipschitz on compact sets. As the function $C - (\varphi_\mu + \varphi_\nu)/\tau^2 - V$ takes its values in a compact set and is Lispchitz, we see that $\bar{\rho}$ is bounded away from 0 and ∞ and is Lipschitz. With all this regularity (recall that f is assumed to be C^2 on $(0,+\infty)$), we can take the gradient of (4.6) to obtain (4.4). □

Theorem 4.10 (Flow interchange inequality)**.** *For any $m \geqslant 1$, the following inequality holds:*

$$\int_\Omega |\nabla\bar\rho|^2 f''(\bar\rho)\bar\rho^{m-1} + \int_\Omega (\nabla\bar\rho\cdot\nabla V)\bar\rho^{m-1} \leqslant \frac{U_m(\mu)+U_m(\nu)-2U_m(\bar\rho)}{\tau^2}.$$

The reader can see that this result is a discrete version of (3.6) derived heuristically in the previous chapter.

Proof. We multiply pointewisely (4.4) by $\bar\rho^{m-1}\nabla\bar\rho$ and integrate over Ω. Dropping the entropic term, we easily get

$$\int_\Omega |\nabla\bar\rho|^2 f''(\bar\rho)\bar\rho^{m-1} + \int_\Omega (\nabla\bar\rho\cdot\nabla V)\bar\rho^{m-1} \leqslant -\frac{1}{\tau^2}\int_\Omega \left[\nabla\bar\rho\cdot(\nabla\varphi_\mu+\nabla\varphi_\nu)\right]\bar\rho^{m-1}.$$

To prove the flow interchange inequality, it is enough to show that

$$-\int_\Omega (\nabla\bar\rho\cdot\nabla\varphi_\mu)\bar\rho^{m-1} \leqslant U_m(\mu)-U_m(\bar\rho),$$

as a similar inequality will hold for the term involving φ_ν. To this purpose, we denote by $\rho:[0,1]\to\mathcal{P}(\Omega)$ the constant-speed geodesic joining $\bar\rho$ to μ. By Proposition 2.10, we know that it is given by

$$\rho_t = (\mathrm{Id}-t\nabla\varphi_\mu)\#\bar\rho.$$

By geodesic convexity of U_m, the function $t\mapsto U_m(\rho_t)$ is convex. Hence,

$$\begin{aligned} U_m(\mu)-U_m(\bar\rho) &= U_m(\rho_1)-U_m(\rho_0)\\ &\geqslant \limsup_{t\to0}\frac{U_m(\rho_t)-U_m(\rho_0)}{t}\\ &= \limsup_{t\to0}\int_\Omega \frac{u_m(\rho_t)-u_m(\bar\rho)}{t}\\ &\geqslant \limsup_{t\to0}\int_\Omega \frac{(\rho_t-\bar\rho)u_m'(\bar\rho)}{t}\\ &= \limsup_{t\to0}\int_\Omega \frac{u_m'(\bar\rho[x-t\nabla\varphi_\mu(x)])-u_m'(\bar\rho[x])}{t}\bar\rho(x)\mathrm{d}x, \end{aligned}$$

where we also have used that u_m is convex. It is clear that for a.e. $x\in\Omega$,

$$\lim_{t\to0}\frac{u_m'(\bar\rho[x-t\nabla\varphi_\mu(x)])-u_m'(\bar\rho[x])}{t} = -\left[(\nabla\bar\rho\cdot\nabla\varphi_\mu)u_m''(\bar\rho)\right](x).$$

Moreover, we have the uniform (in t) bound

$$\left|\frac{u_m'(\bar\rho[x-t\nabla\varphi_\mu(x)])-u_m'(\bar\rho[x])}{t}\right| \leqslant \|u_m''(\bar\rho)\|_\infty\|\nabla\bar\rho\|_\infty\|\nabla\varphi_\mu\|_\infty.$$

At this point, one can remember that $u_m''(x)=x^{m-2}$. Moreover, as $\bar\rho$ is bounded away from 0 and ∞ and Lipschitz, the r.h.s. of the equation above is finite. Thus, by dominated convergence,

$$\limsup_{t\to0}\int_\Omega \frac{u_m'(\bar\rho[x-t\nabla\varphi_\mu(x)])-u_m'(\bar\rho[x])}{t}\bar\rho(x)\mathrm{d}x = -\int_\Omega(\nabla\bar\rho\cdot\nabla\varphi_\mu)\bar\rho^{m-1}. \qquad \square$$

From the result of Theorem 4.10 we need to deduce estimates on improved L^m norms. To this aim, we treat in a slightly different way the cases of weak and strong congestion even if the result are similar. The main issue is to control the term involving ∇V.

Corollary 4.11 (Strong congestion case)**.** *Suppose that Assumption 4.1 holds. Then, for any* $m \geqslant \alpha + 2$ *one has*

$$U_{\beta m}(\bar{\rho})^{1/\beta} \leqslant Cm^2 \left[\frac{U_m(\mu) + U_m(\nu) - 2U_m(\bar{\rho})}{\tau^2} + Cm^2 U_m(\bar{\rho}) \right],$$

where $C > 0$ *depends only on* f, V *and* Ω.

Proof. Let us start from the case of Assumption 4.1. In this case, we recall that C_f is the constant such that $f''(t) \geqslant C_f t^\alpha$ for any $t > 0$. We transform the term involving ∇V in the following way:

$$\begin{aligned} \int_\Omega (\nabla \bar{\rho} \cdot \nabla V) \bar{\rho}^{m-1} &= \int_\Omega (\bar{\rho}^{\alpha/2} \nabla \bar{\rho}) \cdot (\bar{\rho}^{-\alpha/2} \nabla V) \bar{\rho}^{m-1} \\ &\geqslant -\frac{C_f}{2} \int_\Omega |\bar{\rho}^{\alpha/2} \nabla \bar{\rho}|^2 \bar{\rho}^{m-1} - \frac{1}{2C_f} \int_\Omega |\bar{\rho}^{-\alpha/2} \nabla V|^2 \bar{\rho}^{m-1} \\ &= -\frac{C_f}{2} \int_\Omega |\nabla \bar{\rho}|^2 \bar{\rho}^{m-1+\alpha} - \frac{1}{2C_f} \int_\Omega |\nabla V|^2 \bar{\rho}^{m-1-\alpha} \\ &\geqslant -\frac{C_f}{2} \int_\Omega |\nabla \bar{\rho}|^2 \bar{\rho}^{m-1+\alpha} - \frac{\|\nabla V\|_\infty^2 m^2}{2C_f} U_m(\bar{\rho}). \end{aligned}$$

For the last inequality, we have used the fact that

$$\int_\Omega \bar{\rho}^{m-1-\alpha} \leqslant \left(\int_\Omega \bar{\rho}^m \right)^{(m-1-\alpha)/m} \leqslant \int_\Omega \bar{\rho}^m \leqslant m^2 U_m(\bar{\rho}),$$

which is valid because $1 \leqslant m - 1 - \alpha \leqslant m$ and $\mathcal{L}(\Omega) = 1$. Thus, using Theorem 4.10, we get

$$\begin{aligned} \frac{C_f}{2} \int_\Omega |\nabla \bar{\rho}|^2 \bar{\rho}^{m-1+\alpha} &\leqslant \int_\Omega |\nabla \bar{\rho}|^2 f''(\bar{\rho}) \bar{\rho}^{m-1} - \frac{C_f}{2} \int_\Omega |\nabla \bar{\rho}|^2 \bar{\rho}^{m-1+\alpha} \\ &\leqslant \left[\frac{U_m(\mu) + U_m(\nu) - 2U_m(\bar{\rho})}{\tau^2} + \frac{\|\nabla V\|_\infty^2}{2C_f} m^2 U_m(\bar{\rho}) \right]. \end{aligned}$$

We are interested only in the large values taken by $\bar{\rho}$. Let us introduce $\hat{\rho} := \max(1, \bar{\rho})$. This function is larger than $\bar{\rho}$ and 1 and its gradient satisfies $|\nabla \hat{\rho}| = |\nabla \bar{\rho}| \mathbb{1}_{\bar{\rho} \geqslant 1}$. Thus,

$$\int_\Omega |\nabla \hat{\rho}^{m/2}|^2 = \frac{m^2}{4} \int_\Omega |\nabla \hat{\rho}|^2 \hat{\rho}^{m-2} \leqslant \frac{m^2}{4} \int_\Omega |\nabla \hat{\rho}|^2 \hat{\rho}^{m-1+\alpha} \leqslant \frac{m^2}{4} \int_\Omega |\nabla \bar{\rho}|^2 \bar{\rho}^{m-1+\alpha}.$$

(the last inequality is true since $\nabla \hat{\rho} = 0$ on the points where $\hat{\rho} > \bar{\rho}$, and the first inequality is exactly the point where we exploit the fact $\hat{\rho} \geqslant 1$, which explains the use of $\hat{\rho}$ instead of $\bar{\rho}$). On the other hand, if we use the injection of $H^1(\Omega)$ into $L^{2\beta}(\Omega)$ for the function $\hat{\rho}^{m/2}$, we get (with C_Ω a constant that depends only on Ω),

$$\left(\int_\Omega \hat{\rho}^{m\beta} \right)^{1/\beta} \leqslant C_\Omega \left(\int_\Omega |\nabla \hat{\rho}^{m/2}|^2 + \int_\Omega \hat{\rho}^m \right).$$

As $\bar\rho^{\beta m} \leqslant \hat\rho^{\beta m}$ and $\hat\rho^m \leqslant 1 + \bar\rho^m$, we see that

$$\begin{aligned}\left(\int_\Omega \bar\rho^{m\beta}\right)^{1/\beta} &\leqslant \left(\int_\Omega \hat\rho^{m\beta}\right)^{1/\beta} \\ &\leqslant C_\Omega\left(\frac{m^2}{4}\int_\Omega |\nabla\bar\rho|^2\bar\rho^{m-1+\alpha} + \int_\Omega \bar\rho^m + 1\right) \\ &\leqslant C_\Omega m^2\left(\frac{1}{4}\int_\Omega |\nabla\bar\rho|^2\bar\rho^{m-1+\alpha} + 2U_m(\bar\rho)\right) \\ &\leqslant Cm^2\left[\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar\rho)}{\tau^2} + Cm^2U_m(\bar\rho)\right].\end{aligned}$$

Notice that to go from the second to the third line, we have used the fact that $1 \leqslant \int_\Omega \bar\rho^m \leqslant m^2U_m(\bar\rho)$. To conclude, it remains to notice that, as $m\beta \geqslant \beta > 1$, that we can control (uniformly in m) $U_{m\beta}(\bar\rho)$ by $\int_\Omega \bar\rho^{m\beta}$. Indeed

$$\left(\int_\Omega \bar\rho^{m\beta}\right)^{1/\beta} \geqslant \frac{1}{(\beta(\beta-1))^{1/\beta}}U_{m\beta}(\bar\rho)^{1/\beta}.$$

Thus, up to a change in the constant C, we get the result we claimed. □

Corollary 4.12 (Weak congestion case)**.** *Suppose Assumption 4.3 and 4.4 both hold. Then, for any $m \geqslant 1$ such that $\beta(m+\alpha+1) \geqslant 1$ one has*

$$\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar\rho)}{\tau^2} + CmU_m(\bar\rho) \geqslant 0$$

and

$$U_{\beta(m+1+\alpha)}(\bar\rho)^{1/\beta} \leqslant Cm^2\left[\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar\rho)}{\tau^2} + CmU_m(\bar\rho)\right] + Ct_0^{m+1+\alpha},$$

where C depends only on f, V and Ω.

Proof. We use an integration by parts to treat the term involving ∇V. Recall that $\mathbf{n}_\Omega$ denotes the exterior normal to Ω.

$$\begin{aligned}\int_\Omega (\nabla\bar\rho\cdot\nabla V)\bar\rho^{m-1} &= \frac{1}{m}\int_\Omega \nabla(\bar\rho^m)\cdot\nabla V \\ &= \frac{1}{m}\int_{\partial\Omega}(\nabla V\cdot\mathbf{n}_\Omega)\bar\rho^m - \frac{1}{m}\int_\Omega \Delta V\bar\rho^m \\ &\geqslant -\|\Delta V\|_\infty mU_m(\bar\rho),\end{aligned}$$

where we have used the assumption $\nabla V\cdot\mathbf{n}_\Omega \geqslant 0$ on $\partial\Omega$. Thus, using Theorem 4.10, we get (recall that $f''(t) \geqslant C_f t^\alpha$ but only for $t \geqslant t_0$)

$$C_f\int_{\{\bar\rho\geqslant t_0\}} |\nabla\bar\rho|^2\bar\rho^{m-1+\alpha} \leqslant C_f\int_\Omega |\nabla\bar\rho|^2\bar\rho^{m-1}f''(\bar\rho) \tag{4.7}$$

$$\leqslant \left[\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar\rho)}{\tau^2} + \|\Delta V\|_\infty mU_m(\bar\rho)\right]. \tag{4.8}$$

This gives us the first inequality of the corollary. In a similar manner to the strong congestion case, we introduce $\hat\rho := \max(t_0, \bar\rho)$. This time we notice that

$$\int_\Omega |\nabla\hat\rho^{(m+1+\alpha)/2}|^2 \leqslant \frac{m^2}{4}\int_{\{\bar\rho\geqslant t_0\}} |\nabla\bar\rho|^2\bar\rho^{m-1+\alpha}.$$

Thus, if we use the injection of $H^1(\Omega)$ into $L^{2\beta}(\Omega)$ with the function $\hat{\rho}^{(m+1+\alpha)/2}$,

$$\left(\int_\Omega \hat{\rho}^{\beta(m+1+\alpha)}\right)^{1/\beta} \leqslant C_\Omega \left(\int_\Omega |\nabla \hat{\rho}^{(m+1+\alpha)/2}|^2 + \int_\Omega \hat{\rho}^{m+1+\alpha}\right).$$

Then, we proceed as in the proof of the strong congestion case, but this time $m+1+\alpha \leqslant m$ and $\hat{\rho}^{m+1+\alpha} \leqslant \bar{\rho}^{m+1+\alpha} + t_0^{m+1+\alpha}$:

$$\begin{aligned}\left(\int_\Omega \bar{\rho}^{\beta(m+1+\alpha)}\right)^{1/\beta} &\leqslant \left(\int_\Omega \hat{\rho}^{\beta(m+1+\alpha)}\right)^{1/\beta} \\ &\leqslant C_\Omega \left(\frac{m^2}{4}\int_{\{\bar{\rho}\geqslant 1\}} |\nabla \bar{\rho}|^2 \bar{\rho}^{m-1+\alpha} + \int_\Omega \bar{\rho}^{m+1+\alpha} + t_0^{m+1+\alpha}\right) \\ &\leqslant C_\Omega \left(\frac{m^2}{4}\int_{\{\bar{\rho}\geqslant 1\}} |\nabla \bar{\rho}|^2 \bar{\rho}^{m-1+\alpha} + \int_\Omega \bar{\rho}^{m} + t_0^{m+1+\alpha}\right) \\ &\leqslant Cm^2\left[\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar{\rho})}{\tau^2} + CmU_m(\bar{\rho})\right] + Ct_0^{m+1+\alpha}.\end{aligned}$$

Notice that if $t_0 \leqslant 1$, we can control t_0^m by $m^2 U_m(\bar{\rho})$ (as we did in the strong congestion case), but in the general case this is not possible and we have to keep an explicit dependence in t_0. To conclude, we notice that, thanks to (4.3), one has $\beta(m+1+\alpha) \geqslant m \geqslant m_0$ and thus

$$\left(\int_\Omega \bar{\rho}^{\beta(m+1+\alpha)}\right)^{1/\beta} \geqslant \frac{1}{(m_0(m_0-1))^{1/\beta}} U_{\beta(m+1+\alpha)}(\bar{\rho})^{1/\beta}.$$ ☐

The last case is a combination of the previous two cases.

Corollary 4.13 (Strong congestion case-variant)**.** *Suppose that Assumption 4.2 and 4.4 both hold. Then, for any $m \geqslant m_0$ one has*

$$\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar{\rho})}{\tau^2} + CmU_m(\bar{\rho}) \geqslant 0 \tag{4.9}$$

and

$$U_{\beta m}(\bar{\rho})^{1/\beta} \leqslant Cm^2\left[\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar{\rho})}{\tau^2} + CmU_m(\bar{\rho})\right] + Ct_0^m, \tag{4.10}$$

where C depends only on f, V and Ω.

Proof. We begin with the same computations as in Corollary 4.12. We can obtain the same result as in (4.7), but on the set $\{\bar{\rho} \geqslant t_0\}$ we can use $\alpha \geqslant -1$ to write

$$\int_{\{\bar{\rho}\geqslant t_0\}} |\nabla \bar{\rho}|^2 \bar{\rho}^m \leqslant C\left[\frac{U_m(\mu)+U_m(\nu)-2U_m(\bar{\rho})}{\tau^2} + \|\Delta V\|_\infty m U_m(\bar{\rho})\right].$$

With $\hat{\rho} := \max(t_0, \bar{\rho})$ we get

$$\int_\Omega |\nabla \hat{\rho}^{m/2}|^2 \leqslant C\frac{m^2}{4}\int_{\{\bar{\rho}\geqslant t_0\}} |\nabla \bar{\rho}|^2 \bar{\rho}^m,$$

and the conclusion comes from the same Sobolev injection, with the function $\hat{\rho}^{m/2}$, and similar computations as in the previous cases. ☐

For simplicity, inequalities (4.9) and (4.10) will be used by replacing the term $mU_m(\bar{\rho})$ with $m^2 U_m(\bar{\rho})$, so as to allow a unified presentation with the inequality obtained in Corollary 4.11. Notice also that Corollary 4.11 is basically giving us the same inequality as (4.10), as long as we set $t_0 = 0$.

4.2.2 Boundary flow interchange

In the case of Assumption 4.5, we can derive some estimate right at the final time $t = 1$ ($k = N$). We will only sketch the proof, at it mimicks the proof of the interior case and these computations are well-known in the case of the applications to the JKO scheme. We know that, with $\bar\rho = \bar\rho_1^{N,\lambda}$ and $\mu := \bar\rho_{1-\tau}^{N,\lambda}$, the measure $\bar\rho$ is a minimizer (among all probability measures) of

$$\rho \mapsto \frac{W_2^2(\mu,\rho)}{2\tau} + G(\rho) + \lambda U_1(\rho) + \int_\Omega W \mathrm{d}\rho.$$

Let us remark that it correspond to one step of the JKO scheme: it is in the context of such variational problems that the flow interchange was firstly used, see [MMS09]. In any case, with these notation, we obtain:

Proposition 4.14. *Suppose Assumption 4.5 holds. Then, for any $m \geqslant 1$,*

$$\frac{U_m(\mu) - U_m(\bar\rho)}{\tau} \geqslant -(m-1)\|\Delta W\|_\infty U_m(\bar\rho).$$

Proof. Following the same strategy than in Lemma 4.8 and Proposition 4.9, we know that $\bar\rho$ is bounded away from 0 and ∞, is a Lipschitz function, and that

$$\frac{\nabla\varphi_\mu}{\tau} + \left(g''(\bar\rho) + \frac{\lambda}{\bar\rho}\right)\nabla\bar\rho + \nabla W = 0$$

a.e. on Ω, where φ_μ is the unique Kantorovich potential for the transport from $\bar\rho$ to μ. Thus, if we multiply by $\bar\rho^{m-1}\nabla\bar\rho$, we get, by the same estimation than in Theorem 4.10 (we drop both the entropic penalization and the congestion term),

$$\frac{U_m(\mu) - U_m(\bar\rho)}{\tau} \geqslant \int_\Omega (\nabla W \cdot \nabla\bar\rho)\bar\rho^{m-1}.$$

It remains to perform an integration by parts, using the sign of $\nabla W \cdot \mathbf{n}_\Omega$ on $\partial\Omega$, to conclude that

$$\frac{U_m(\mu) - U_m(\bar\rho)}{\tau} \geqslant -\frac{1}{m}\int_\Omega \Delta W \bar\rho^m \geqslant -(m-1)\|\Delta W\|_\infty U_m(\bar\rho).$$ □

4.3 Moser-like iterations

Corollaries 4.11, 4.12 and 4.13 allow us to control the $L^{m\beta}$ or $L^{(m+1+\alpha)\beta}$ norm of $\bar\rho$ in terms of its L^m norm. The strategy will consist in integrating w.r.t. to time and iterating such a control in order to get a bound on the $L^m([T_1,T_2]\times\Omega)$ norm of $\bar\rho^{N,\lambda}$ that does not depend on λ and N and to control how this bounds grows in m. For any $N \geqslant 1$ and any $0 < \lambda < 1$, recall that $\bar\rho^{N,\lambda}$ is a solution of the discrete problem (4.2).

Definition 4.15. *For any $m \geqslant 1$ and any $0 \leqslant T_1 \leqslant T_2 \leqslant 1$, we define $L^m_{T_1,T_2}$ as*

$$L^m_{T_1,T_2} := \liminf_{N\to+\infty,\lambda\to 0} \left(\sum_{T_1 \leqslant k\tau \leqslant T_2} \tau U_m(\bar\rho_{k\tau}^{N,\lambda})\right)^{1/m}.$$

The quantity $L^m_{T_1,T_2}$ can be seen as a discrete counter part of (up to a factor $1/(m(m-1))^{1/m}$) the L^m norm of the restriction to $[T_1,T_2]$ of the limit (whose existence will be proven in the next section) of $\bar\rho^{N,\lambda}$ when $N \to +\infty$ and $\lambda \to 0$.

4.3.1 The strong congestion case

First, we integrate w.r.t. time the estimate obtained in Corollary 4.11.

Proposition 4.16. *Suppose that either Assumption 4.1 holds or Assumptions 4.2 and 4.4 both hold. Then there exists two constants C_1 and C_2 (depending on f, V and Ω) such that, for any $0 < \varepsilon \leqslant C_1/m$ and any $0 < T_1 < T_2 < 1$ such that $[T_1 - \varepsilon, T_2 + \varepsilon] \subset (0,1)$, and any $m \geqslant \alpha + 2$,*

$$L^{\beta m}_{T_1,T_2} \leqslant \left[C_2 \frac{m^3}{\varepsilon} \left(m^2 + \frac{1}{\varepsilon^2} \right) \right]^{1/m} \max \left(L^m_{T_1-\varepsilon, T_2+\varepsilon}, t_0 \right).$$

As pointed out earlier after Corollary 4.13, in the case where Assumption 4.1 holds, we set $t_0 = 0$.

Proof. Let us recall that in Corollary 4.11 and Corollary 4.13, we have proved (if we explicit the dependence in N and λ) that for any $N \geqslant 1$, $\lambda > 0$ and any $k \in \{1, 2, \dots, N-1\}$, one has

$$U_{\beta m}(\bar{\rho}^{N,\lambda}_{k\tau})^{1/\beta} \leqslant Cm^2 \left[\frac{U_m(\bar{\rho}^{N,\lambda}_{(k-1)\tau}) + U_m(\bar{\rho}^{N,\lambda}_{(k+1)\tau}) - 2U_m(\bar{\rho}^{N,\lambda}_{k\tau})}{\tau^2} + Cm^2 U_m(\bar{\rho}^{N,\lambda}_{k\tau}) \right] + Ct_0^m. \tag{4.11}$$

Let us take $\chi : [0,1] \to [0,1]$ a positive $C^{1,1}$ cutoff function such that $\chi(t) = 1$ if $t \in [T_1 - \varepsilon/3, T_2 + \varepsilon/3]$ and $\chi(t) = 0$ if $t \notin [T_1 - 2\varepsilon/3, T_2 + 2\varepsilon/3]$. Such a function χ can be chosen with $\|\chi''\|_\infty \leqslant 54/\varepsilon^2$. We multiply (4.11) by $\tau\chi(k\tau)$ and sum over $k \in \{1, 2, \dots, N-1\}$. After performing a discrete integration by parts, we are left with

$$\sum_{k=1}^{N-1} \tau \chi(k\tau) U_{m\beta}(\bar{\rho}^{N,\lambda}_{k\tau})^{1/\beta} \leqslant Ct_0^m + Cm^2 \sum_{k=1}^{N-1} \tau U_m(\bar{\rho}^{N,\lambda}_{k\tau}) \left[Cm^2 + \frac{\chi((k+1)\tau) + \chi((k-1)\tau) - 2\chi(k\tau)}{\tau^2} \right].$$

Given the bound on the second derivative of χ, and if $\tau \leqslant \varepsilon/3$, we get

$$\sum_{T_1 - \varepsilon/3 \leqslant k\tau \leqslant T_2 + \varepsilon/3} \tau U_{m\beta}(\bar{\rho}^{N,\lambda}_{k\tau})^{1/\beta} \leqslant Ct_0^m + Cm^2 \left(m^2 + \frac{1}{\varepsilon^2} \right) \sum_{T_1 - \varepsilon \leqslant k\tau \leqslant T_2 - \varepsilon} \tau U_m(\bar{\rho}^{N,\lambda}_{k\tau}).$$

The l.h.s. is not exactly $\left(L^{m\beta}_{T_1,T_2} \right)^{1/m}$ as we would like to exchange the sum and the power $1/\beta$. Unfortunately, Jensen's inequality gives the inequality the other way around. To overcome this difficulty, we will use the fact that the function $k \mapsto U_{\beta m}(\bar{\rho}^{N,\lambda}_{k\tau})$ is almost a convex function of k. More precisely, we will use the "reverse Jensen inequality", whose proof is postponed at the end of this chapter in Section 4.5.

Lemma 4.17. *Let $(u^\tau_k)_{k \in \mathbb{Z}}$ be a family of real sequences indexed by a parameter τ. We assume that there exists $\omega \geqslant 0$ such that for any $k \in \mathbb{Z}$ and any τ, one has $u^\tau_k > 0$ and*

$$\frac{u^\tau_{k+1} + u^\tau_{k-1} - 2u^\tau_k}{\tau^2} + \omega^2 u^\tau_k \geqslant 0. \tag{4.12}$$

Then, for any $T_1 < T_2$ and any $\eta < \pi/(8\omega)$, there exists τ_0 (which depends on ω), such that, if $\tau \leqslant \tau_0$, then

$$\left(\sum_{T_1 \leqslant k\tau \leqslant T_2} \tau u^\tau_k \right)^{1/\beta} \leqslant C \frac{(\omega+1)(T_2 - T_1 + 1)^{1+1/\beta}}{\eta} \sum_{T_1 - \eta \leqslant k\tau \leqslant T_2 + \eta} \tau (u^\tau_k)^{1/\beta},$$

where C is a universal constant.

To use this lemma, we observe that $u_k^\tau := U_{m\beta}(\bar\rho_{k\tau}^{N,\lambda})$ satisfies (4.12) with $\omega^2 = Cm^2$ (thanks again to Corollary 4.11 and Corollary 4.13). Thus, if we take C_1 small enough, we have $\varepsilon/3 < \pi/(8\omega)$ as soon as $\varepsilon \leqslant C_1/m$. If τ is small enough, we can exchange the sum and the power $1/\beta$ to get

$$\begin{aligned}\left(\sum_{T_1 \leqslant k\tau \leqslant T_2} \tau U_{m\beta}(\bar\rho_{k\tau}^{N,\lambda})\right)^{1/\beta} &\leqslant C\frac{Cm+1}{\varepsilon} \sum_{T_1-\varepsilon/3 \leqslant k\tau \leqslant T_2+\varepsilon/3} \tau U_{m\beta}(\bar\rho_{k\tau}^{N,\lambda})^{1/\beta} \\ &\leqslant C_2 \frac{m^3}{\varepsilon}\left(m^2 + \frac{1}{\varepsilon^2}\right)\left(t_0 + \sum_{T_1-\varepsilon \leqslant k\tau \leqslant T_2+\varepsilon} \tau U_m(\bar\rho_{k\tau}^{N,\lambda})\right).\end{aligned}$$

Notice that we have put the constant $C_2 m^3 \varepsilon^{-1}(m^2+\varepsilon^{-2})$ also in factor of t_0^m, as it is anyway larger than 1 as soon as ε is small enough. Then we take the power $1/m$ on both sides, use the identity $(a+b)^{1/m} \leqslant C\max(a^{1/m}, b^{1/m})$ and send $N \to +\infty$ and $\lambda \to 0$ to get the result. $\square$

In other words, on a slightly larger time interval, the $L^{\beta m}$ norm is control by the L^m norm. We just have to iterate this inequality.

Proposition 4.18. *Suppose that either Assumption 4.1 holds or Assumptions 4.2 and 4.4 both hold. For any* $0 < T_1 < T_2 < 1$*, there exists* C *(that depends on* T_1, T_2, f, V *and* Ω*) such that*

$$\limsup_{m\to+\infty} L^m_{T_1,T_2} \leqslant C\max\left(L^{\alpha+2}_{0,1}, t_0\right).$$

Proof. Let $\varepsilon_0 > 0$ be small enough such that $0 < T_1 - \varepsilon_0\beta/(\beta-1) \leqslant T_2 + \varepsilon_0\beta/(\beta-1) < 1$ and $\varepsilon_0 \leqslant C_1/(\alpha+2)$ (where C_1 is the constant defined in Proposition 4.16). For any $n \in \mathbb{N}$, let us define

$$T_1^n := T_1 - \sum_{k=n}^{+\infty} \frac{\varepsilon_0}{\beta^n} \quad\text{and}\quad T_2^n := T_2 + \sum_{k=n}^{+\infty}\frac{\varepsilon_0}{\beta^n},$$

and set $m_n := (\alpha+2)\beta^n$. Using Proposition 4.16, as we have $|T_i^{n+1} - T_i^n| = \varepsilon_0\beta^{-n} \leqslant C_1/m_n$ for $i \in \{1,2\}$, we can say that, with $l_n := \max\left(L^{m_n}_{T_1^n,T_2^n}, t_0\right)$

$$\begin{aligned} l_{n+1} &\leqslant \left[\max\left\{1, C_2\frac{m_n^3}{\varepsilon_0\beta^{-n}}\left(m_n^2 + \frac{1}{(\varepsilon_0\beta^{-n})^2}\right)\right\}\right]^{1/m_n} l_n \\ &\leqslant \left[C\beta^{6n}\right]^{\beta^{-n}/(\alpha+2)} l_n.\end{aligned}$$

One can easily check, as $\beta > 1$, that

$$\prod_{n=0}^{+\infty}\left[C\beta^{6n}\right]^{\beta^{-n}/(\alpha+2)} < +\infty,$$

thus we get that

$$\sup_{n\in\mathbb{N}} L^{m_n}_{T_1,T_2} \leqslant \sup_{n\in\mathbb{N}} L^{m_n}_{T_1^n,T_2^n} \leqslant \sup_{n\in\mathbb{N}} l_n \leqslant Cl_0 = C\max\left(L^{m_0}_{T_1^0,T_2^0}, t_0\right) \leqslant C\max\left(L^{\alpha+2}_{0,1}, t_0\right).$$

To conclude, we notice that, if $m > 1$ and $m_n \geqslant m$, one has (using Jensen's inequality)

$$L^m_{T_1,T_2} \leqslant \frac{(m_n(m_n-1))^{1/m_n}}{(m(m-1))^{1/m}} L^{m_n}_{T_1,T_2},$$

thus sending $m \to +\infty$ (hence $n \to +\infty$) we conclude that

$$\limsup_{m\to+\infty} L^m_{T_1,T_2} \leqslant \sup_{n\in\mathbb{N}} L^{m_n}_{T_1,T_2}. \tag{4.13}$$

$\square$

As we will see later, the fact that $L^{\alpha+2}_{0,1}$ is finite is a consequence of the fact that the solution $\bar\rho$ of the continuous problem (4.1) satisfies $\int_0^1 F(\bar\rho_t)\mathrm{d}t < +\infty$.

4.3.2 Estimates up to the final time

In this subsection, still supposing that either Assumption 4.1 holds or Assumptions 4.2 and 4.4 both hold, we exploit Assumption 4.5 to extend the L^∞ bound up to the final time $t = 1$. We will prove a result similar to Proposition 4.16, but this time up to the boundary.

Proposition 4.19. *Suppose that either Assumption 4.1 holds or Assumptions 4.2 and 4.4 both hold, and that Assumption 4.5 also holds. Then there exists two constants C_1 and C_2 (depending on f, V, g, W and Ω) such that for any $0 < \varepsilon < C_1/m$ and any $0 < T_1 < 1$ with $0 < T_1 - \varepsilon$, then for any $m \geqslant \alpha + 2$,*

$$L^{\beta m}_{T_1,1} \leqslant \left[C_2 \frac{m^3}{\varepsilon} \left(\frac{m}{\varepsilon} + m^2 + \frac{1}{\varepsilon^2} \right) \right]^{1/m} \max\left(L^m_{T_1-\varepsilon,1}, t_0 \right).$$

Again, we recall that if we are under Assumption 4.1, we take $t_0 = 0$.

Proof. Let us recall that equation (4.11) holds for any $N \geqslant 1$, $\lambda > 0$ and $k \in \{1, 2, \dots, N-1\}$. We take $\chi : [0,1] \to [0,1]$ a positive $C^{1,1}$ cutoff function such that $\chi(t) = 1$ if $t \in [T_1 - \varepsilon/3, 1]$ and $\chi(t) = 0$ if $t \in [0, T_1 - 2\varepsilon/3]$. Such a function χ can be chosen with $\|\chi''\|_\infty \leqslant 54/\varepsilon^2$. We multiply (4.11) by $\tau\chi(k\tau)$ and sum over $k \in \{1, 2, \dots, N-1\}$. After performing a discrete integration by parts, we are left with (now a boundary term is appearing):

$$\sum_{k=1}^{N-1} \tau\chi(k\tau) U_{m\beta}(\bar\rho^{N,\lambda}_{k\tau})^{1/\beta} \leqslant Cm^2 \Bigg(\frac{U_m(\bar\rho^{N,\lambda}_1) - U_m(\bar\rho^{N,\lambda}_{1-\tau})}{\tau} \chi(1) + \sum_{k=1}^{N-1} \tau U_m(\bar\rho^{N,\lambda}_{k\tau}) \left[Cm^2 + \frac{\chi(k\tau+\tau) + \chi(k\tau-\tau) - 2\chi(k\tau)}{\tau^2} \right] \Bigg) + Ct_0^m.$$

With the help of Proposition 4.14 and Corollary 4.11 or Corollary 4.13, and as $\chi(1) = 1$, we are able to write (provided that $\tau \leqslant \varepsilon/3$)

$$\sum_{T_1-\varepsilon/3 \leqslant k\tau \leqslant 1} \tau U_{m\beta}(\bar\rho^{N,\lambda}_{k\tau})^{1/\beta} \leqslant Cm^2 \left(mU_m(\bar\rho^{N,\lambda}_1) + \left[m^2 + \frac{1}{\varepsilon^2} \right] \sum_{T_1-\varepsilon \leqslant k\tau \leqslant 1} \tau U_m(\bar\rho^{N,\lambda}_{k\tau}) \right) + Ct_0^m.$$

To transform the boundary term $U_m(\bar\rho^{N,\lambda}_1)$ into an integral term, we use the following lemma, whose proof is also postponed at the end of this chapter in Section 4.5.

Lemma 4.20. *Let $(u^\tau_k)_{k\in\mathbb{Z}}$ be a family of real sequences indexed by a parameter τ. We assume that there exists $\omega \geqslant 0$ such that for any $k \in \mathbb{Z}$ and any τ, one has $u^\tau_k > 0$ and (4.12). We also assume that there exists $b \geqslant 0$ such that for some $N \in \mathbb{Z}$,*

$$\frac{u^\tau_N - u^\tau_{N-1}}{\tau} \leqslant b u^\tau_N.$$

Then, there exists C_1 and C_2 universal constants and τ_0 (which depends on ω and b), such that for any $\eta \leqslant \min\{\pi/(32\omega), \pi/(32b)\}$ and any $\tau \leqslant \tau_0$, then

$$u^\tau_N \leqslant \frac{C_1}{\eta} \sum_{N\tau-\eta \leqslant k\tau \leqslant N\tau} \tau u^\tau_k, \tag{4.14}$$

and for any $T_1 < N\tau$,

$$\left(\sum_{T_1 \leqslant k\tau \leqslant N\tau} \tau u^\tau_k \right)^{1/\beta} \leqslant C_2 \frac{(\omega+1)(N\tau - T_1 - +1)^{1+1/\beta}}{\eta} \sum_{T_1-\eta \leqslant k\tau \leqslant N\tau} \tau (u^\tau_k)^{1/\beta}. \tag{4.15}$$

We are in the case where this lemma can be applied because of Corollary 4.11 or Corollary 4.13 and Proposition 4.14 with $u_k^\tau = U_m(\bar{\rho}_{k\tau}^{N,\lambda})$, $\omega = Cm$ and $b = Cm$ and $N\tau = 1$. Thus, if $\varepsilon < C/m$, we can guarantee that we can use equation (4.14) of Lemma 4.20 (with $\varepsilon = \eta$), thus

$$\sum_{T_1 - \varepsilon/3 \leqslant k\tau \leqslant 1} \tau U_{m\beta}(\bar{\rho}_{k\tau}^{N,\lambda})^{1/\beta} \leqslant C t_0^m + Cm^2 \left[\frac{m}{\varepsilon} + m^2 + \frac{1}{\varepsilon^2} \right] \sum_{T_1 - \varepsilon \leqslant k\tau \leqslant 1} \tau U_m(\bar{\rho}_{k\tau}^{N,\lambda}).$$

Then we use equation (4.15) of Lemma 4.20 (but this time with $u_k^\tau = U_{\beta m}(\bar{\rho}_{k\tau}^{N,\lambda})$) to exchange the sum and the power $1/\beta$ on the l.h.s. to conclude that

$$\left(\sum_{T_1 \leqslant k\tau \leqslant 1} \tau U_{m\beta}(\bar{\rho}_{k\tau}^{N,\lambda}) \right)^{1/\beta} \leqslant C \frac{m^3}{\varepsilon} \left[\frac{m}{\varepsilon} + m^2 + \frac{1}{\varepsilon^2} \right] \left(t_0^m + \sum_{T_1 - \varepsilon \leqslant k\tau \leqslant 1} \tau U_m(\bar{\rho}_{k\tau}^{N,\lambda}) \right).$$

Again, we have put $m^3 \varepsilon^{-1}(m\varepsilon^{-1} + m^2 + \varepsilon^{-2})$ in factor of t_0^m, which is legit because this factor is larger than 1 for ε small enough. Taking the power $1/m$ on each side, using the identity $(a+b)^{1/m} \leqslant C \max(a^{1/m}, b^{1/m})$, and letting $N \to +\infty$ and $\lambda \to 0$, we get the result. □

It is then very easy to iterate this result, which looks exactly like Proposition 4.16. Thus, the proof of the following proposition, which is exactly the same as Proposition 4.18, is left to the reader.

Proposition 4.21. *Suppose that either Assumption 4.1 holds or Assumptions 4.2 and 4.4 both hold, and that Assumption 4.5 also holds. Then, for any $0 < T_1 < 1$, there exists C (that depends on T_1, f, V and Ω) such that*

$$\limsup_{m \to +\infty} L_{T_1,1}^m \leqslant C \max\left(L_{0,1}^{\alpha+2}, t_0 \right).$$

4.3.3 The weak congestion case

The scheme is very similar in the weak congestion case, even though the iteration is not as direct as in the strong congestion case. Moreover, we will directly prove an L^∞ bound up to $t = 1$, because, as we will see, Assumption 4.5 will be needed anyway to initialize the iterative process. The proofs will be less detailed in this case: the reading on the two previous subsections is advised to understand this one.

Proposition 4.22. *Suppose Assumptions 4.3, 4.4 and 4.5 hold. Then there exist constants C_1 and C_2 (depending on f, V and Ω) such that, for any $0 < \varepsilon \leqslant C_1/m$ and any $\varepsilon < T_1 < 1$, then for any $m \geqslant m_0$,*

$$L_{T_1,1}^{\beta(m+1+\alpha)} \leqslant \left[C_2 \frac{m^{5/2}}{\varepsilon} \left(\frac{m}{\varepsilon} + m + \frac{1}{\varepsilon^2} \right) \right]^{1/(m+1+\alpha)} \max\left[\left(L_{T_1-\varepsilon,1}^m \right)^{m/(m+1+\alpha)}, t_0 \right].$$

Proof. The proof starts the same way: starting from Corollary 4.12, we write

$$U_{\beta(m+1+\alpha)}(\bar{\rho}_{k\tau}^{N,\lambda})^{1/\beta} \leqslant Cm^2 \left[\frac{U_m(\bar{\rho}_{(k-1)\tau}^{N,\lambda}) + U_m(\bar{\rho}_{(k+1)\tau}^{N,\lambda}) - 2U_m(\bar{\rho}_{k\tau}^{N,\lambda})}{\tau^2} + CmU_m(\bar{\rho}_{k\tau}^{N,\lambda}) \right] + Ct_0^{m+1+\alpha}. \quad (4.16)$$

Because of Assumption 4.5, we can also write, tanks to Proposition 4.14, that

$$\frac{U_m(\bar{\rho}^{N,\lambda}_{1-\tau}) - U_m(\bar{\rho}^{N,\lambda}_1)}{\tau} \geqslant -(m-1)\|\Delta W\|_\infty U_m(\bar{\rho}^{N,\lambda}_1).$$

We use the same cutoff function χ that in the proof of Proposition 4.19. We multiply (4.16) by $\tau\chi(k\tau)$, perform a discrete integration by parts and end up with

$$\begin{aligned}
\sum_{T_1-\varepsilon/3 \leqslant k\tau \leqslant 1} & \tau U_{\beta(m+1+\alpha)}(\bar{\rho}^{N,\lambda}_{k\tau})^{1/\beta} \\
& \leqslant Cm^2 \left(\frac{U_m(\bar{\rho}^{N,\lambda}_1) - U_m(\bar{\rho}^{N,\lambda}_{1-\tau})}{\tau} + \left[m + \frac{1}{\varepsilon^2} \right] \sum_{T_1-\varepsilon \leqslant k\tau \leqslant 1} \tau U_m(\bar{\rho}^{N,\lambda}_{k\tau}) \right) + Ct_0^{m+1+\alpha} \\
& \leqslant Cm^2 \left(mU_m(\bar{\rho}^{N,\lambda}_1) + \left[m + \frac{1}{\varepsilon^2} \right] \sum_{T_1-\varepsilon \leqslant k\tau \leqslant 1} \tau U_m(\bar{\rho}^{N,\lambda}_{k\tau}) \right) + Ct_0^{m+1+\alpha}.
\end{aligned}$$

We also use Lemma 4.20 but this time with $\omega^2 = Cm$ (this is Corollary 4.12) and $b = Cm$. The frequency ω^2 is smaller than in the strong congestion case (where it was of the order m^2) because we have made stronger assumptions on the potential V, though this is not important as we only use the fact that ω grows not faster than a polynomial of m. With this lemma we can both transform the boundary term into an integral term and exchange the sum and the power $1/\beta$: there exists C_1 such that if $0 < \varepsilon \leqslant C_1/m$ and if τ is small enough,

$$\begin{aligned}
\left(\sum_{T_1 \leqslant k\tau \leqslant 1} \tau U_{\beta(m+1+\alpha)}(\bar{\rho}^{N,\lambda}_{k\tau}) \right)^{1/\beta} & \leqslant C\frac{\sqrt{m}+1}{\varepsilon} \sum_{T_1-\varepsilon/3 \leqslant k\tau \leqslant 1} \tau U_{\beta(m+1+\alpha)}(\bar{\rho}^{N,\lambda}_{k\tau})^{1/\beta} \\
& \leqslant C\frac{m^{5/2}}{\varepsilon} \left(\frac{m}{\varepsilon} + m + \frac{1}{\varepsilon^2} \right) \left(t_0^{m+1+\alpha} + \sum_{T_1-\varepsilon \leqslant k\tau \leqslant 1} \tau U_m(\bar{\rho}^{N,\lambda}_{k\tau}) \right).
\end{aligned}$$

We take the power $1/(m+1+\alpha)$ on both sides, use the fact that

$$(a+b)^{1/(m+1+\alpha)} \leqslant C \max(a^{1/(m+1+\alpha)}, b^{1/(m+1+\alpha)})$$

and let $N \to +\infty, \lambda \to 0$ to get the result. □

We proceed the same way by iterating the inequality, even though this expressions are slightly more complicated. Let us underline that the condition (4.3) on β is precisely the one that ensures that $\beta(m+1+\alpha) > m$ as soon as $m \geqslant m_0$: it is only thanks to this condition that the iteration of Proposition 4.22 will give useful information.

Proposition 4.23. *Suppose Assumptions 4.3, 4.4 and 4.5 hold. Then, there exists* $\gamma < +\infty$ *such that, for any* $0 < T_1 < 1$*, there exists* C *(that depends on* T_1, f, V *and* Ω*) such that*

$$\limsup_{m \to +\infty} L^m_{T_1,1} \leqslant C \left(\max \left[L^{m_0}_{0,1}, t_0 \right] \right)^\gamma.$$

Proof. As we know, thanks to our normalization choices, that $L^{m_0}_{0,1} \geqslant 1$, it is not restrictive that assume that $t_0 \geqslant 1$ (indeed, if this is not the case, Assumption 4.3 is still valid with $t_0 = 1$ and the content of Proposition 4.23 does not change).

Once we have chosen $\varepsilon_0 < \beta T_1/(\beta - 1)$, we define T_1^n by the same formula as in the proof of Proposition 4.18. We also define m_n by recurrence: for any $n \in \mathbb{N}$, we take $m_{n+1} = \beta(m_n + 1 + \alpha)$. Thus, we have the explicit expression

$$m_n = \left(m_0 + (\alpha + 1)\frac{\beta}{\beta - 1} \right) \beta^n - (\alpha + 1)\frac{\beta}{\beta - 1}.$$

In particular, $(m_n)_{n \in \mathbb{N}}$ diverges exponentially fast to $+\infty$ as $n \to +\infty$. Using Proposition 4.22 and as $t_0 \geqslant 1$, we get

$$\begin{aligned} L^{m_{n+1}}_{T_1^{n+1},1} &\leqslant \left[C_2 \frac{m_n^{5/2}}{\varepsilon_0 \beta^{-n}} \left(\frac{m_n}{\varepsilon_0 \beta^{-n}} + m_n + \frac{1}{(\varepsilon_0 \beta^{-n})^2} \right) \right]^{1/(m_n+1+\alpha)} \max \left[\left(L^{m_n}_{T_1^n,1} \right)^{m_n/(m_n+1+\alpha)}, t_0 \right] \\ &\leqslant \left[C\beta^{11n/2} \right]^{1/(m_n+\alpha+1)} \max \left(\left[L^{m_n}_{T_1^n,1}, t_0 \right] \right)^{m_n/(m_n+\alpha+1)}. \end{aligned}$$

Denoting by $l_n := \ln \left(\max \left[L^{m_n}_{T_1^n,1}, t_0 \right] \right)$, we see that

$$l_{n+1} \leqslant C_3 \frac{11n}{2(m_n + \alpha + 1)} + \frac{C_4}{m_n + \alpha + 1} + \frac{m_n}{m_n + \alpha + 1} l_n.$$

Given the exponential asymptotic growth of $(m_n)_{n \in \mathbb{N}}$, we leave it to the reader to check that is enough to conclude that $\limsup_{n \to +\infty} l_n \leqslant \gamma l_0 + C_5$ for some $\gamma < +\infty$ and $C_4 < +\infty$. Taking the exponential gives

$$\limsup_{n \in \mathbb{N}} L^{m_n}_{T_1^n,1} \leqslant C \left(\max \left[L^{m_0}_{0,1}, t_0 \right] \right)^{\gamma}.$$

To conclude, we use again (4.13), which is valid independently of Assumption 4.1 or Assumption 4.3. □

However, in the weak congestion case, the fact that $L^{m_0}_{0,1} < +\infty$ will require a little bit more of work and relies on the particular form of the boundary conditions.

Proposition 4.24. *Suppose Assumptions 4.3, 4.4 and 4.5 hold. Then there exists C_{max} (which depends on m_0) such that, if $\|\Delta V\|_\infty$ and $\|\Delta V\|_\infty$ are smaller than C_{max} then*

$$L^{m_0}_{0,1} < +\infty.$$

Proof. Again we will use the almost convexity of $U_{m_0}(\bar{\rho}^{N,\lambda})$. Indeed, we rely on the following lemma, which has the same flavor as the “reverse Jensen inequality” and whose proof is postponed at the end of this chapter in Section 4.5.

Lemma 4.25. *Let $a > 0$, $b \geqslant 0$ and $\omega \geqslant 0$ and assume $T_{\max} = \min\{\pi/(32\omega), \pi/(32b)\}$ is bounded by 1. Then there exist some constants $C < +\infty$ and $\tau_0 > 0$ (all depending on a, b and ω) such that for any $N > 1/\tau_0$ ($\tau := 1/N$) and for any sequence $(u_k^\tau)_k \in \mathbb{Z}$ of strictly positive numbers satisfying (4.12) for $k \in \{1, 2, \ldots, N-1\}$, and such that $u_0^\tau = a$ and*

$$\frac{u_{N-1}^\tau - u_N^\tau}{\tau} \geqslant -b u_N^\tau,$$

one has $u_k^\tau \leqslant C$ for any $k \in \{1, 2, \ldots, N\}$.

We use this lemma with $u_k^\tau = U_{m_0}(\bar{\rho}_{k\tau}^{N,\lambda})$. Equation (4.12) is satisfied with $\omega^2 = \|\Delta V\|_\infty m_0$ (Corollary 4.12, one should look closely at the proof to see that the dependence is indeed linear in $\|\Delta V\|_\infty$); one can take

$$a = U_{m_0}(\bar{\rho}_0^{N,\lambda}) = U_{m_0}(\bar{\rho}_0) = \frac{1}{m_0(m_0-1)} \int_\Omega \bar{\rho}_0^{m_0};$$

and we take $b = (m_0-1)\|\Delta W\|_\infty$ (cf. Proposition 4.14). Thus, one can conclude that if $1 \leqslant T_{\max}$, then $U_{m_0}(\bar{\rho}_{k\tau}^{N,\lambda})$ is bounded independently on N. The latter condition can be rephrased as $\max(\|\Delta V\|_\infty, \|\Delta V\|_\infty) \leqslant C_{\max}$ once one plugs the formula for $T_{\max}$. This is enough to conclude that $L_{0,1}^{m_0}$ is finite. □

4.4 Limit of the discrete problems

In this section, we will see that the solutions $\bar{\rho}^{N,\lambda}$ of the discrete problems (4.2) converge to the solution $\bar{\rho}$ of the continuous one (4.1) when $N \to +\infty$ and $\lambda \to 0$. Then, using the results of the previous sections, we will be able to show the L^∞ bound on $\bar{\rho}$.

4.4.1 Building discrete curves from continuous one

In our construction we will need to work with curves with finite entropy. This is easy under Assumption 4.1 of 4.2, but requires an approximation in the case of Assumption 4.3. Hence, we will show that in this case we can approximate curves in Γ by curves in Γ with finite entropy. In order to do this, we will use the heat flow, whose definition and useful properties are recalled in Section 2.4.

Recall that the heat flow with Neumann boundary conditions on Ω is denoted by Φ and the heat kernel by K. We mention that if $h : \mathbb{R} \to \mathbb{R}$ is any convex function bounded from below, $\rho \in \mathcal{P}(\Omega) \cap L^1(\Omega)$, and $s \geqslant 0$ then

$$\int_\Omega h\left[(\Phi_s\rho)(x)\right] \mathrm{d}x \leqslant \int_\Omega h[\rho(x)]\mathrm{d}x;$$

If h is not superlinear, the same stays true for any $\rho \in \mathcal{P}(\Omega)$ by replacing the integral $\int_\Omega h[\rho(x)]\mathrm{d}x$ with the expression in (2.12). Indeed, using in particular Jensen's inequality and the fact that $\int_\Omega K_t(x,y)\mathrm{d}x = 1$ for any y and t,

$$\begin{aligned}\int_\Omega h\left[(\Phi_s\rho)(x)\right] \mathrm{d}x &= \int_\Omega h\left(\int_\Omega K_s(x,y)\rho(y)\mathrm{d}y\right)\mathrm{d}x \\ &\leqslant \iint_{\Omega\times\Omega} K_s(x,y)h(\rho(y))\mathrm{d}y\mathrm{d}x = \int_\Omega h[\rho(y)]\mathrm{d}y.\end{aligned}$$

The proof in the case where h is not superlinear and ρ is not absolutely continuous is obtained by writing $\rho =: \rho^{\mathrm{ac}}\mathcal{L} + \rho^{\mathrm{sing}}$. Observing that $h'(\infty)$ is the Lipschitz constant of h, we have

$$\int_\Omega h\left[(\Phi_s\rho)(x)\right] \mathrm{d}x - \int_\Omega h\left[(\Phi_s\rho^{\mathrm{ac}})(x)\right] \mathrm{d}x \leqslant h'(\infty) \int_\Omega |\Phi_s\rho^{\mathrm{sing}}|(x)\mathrm{d}x = h'(\infty)\rho^{\mathrm{sing}}(\Omega).$$

Proposition 4.26. *Suppose Assumption 4.5 holds and that $\bar{\rho}_0$ is such that $U_1(\bar{\rho}_0), F(\bar{\rho}_0) < +\infty$, and let $\rho \in \Gamma$ with $\rho_0 = \bar{\rho}_0$. Then, for any $\varepsilon > 0$, there exists $\tilde{\rho} \in \Gamma$ with $\tilde{\rho}_0 = \bar{\rho}_0$ and $C < +\infty$ such that $\mathcal{A}(\tilde{\rho}) \leqslant \mathcal{A}(\rho) + \varepsilon$ and $U_1(\tilde{\rho}_t) \leqslant C$ for any $t \in [0,1]$.*

Proof. Without loss of generality, we assume $\mathcal{A}(\rho) < +\infty$. The idea is to use the heat flow to regularize solutions. But we cannot apply the heat flow uniformly, as we would loose the boundary condition $\rho_0 = \bar{\rho}_0$. Consequently, for any $0 < s < 1$, we define $\tilde{\rho}^s \in \mathcal{P}(\Gamma)$ by

$$\tilde{\rho}^s(t) := \begin{cases} \Phi_t(\rho_0) & \text{if } 0 \leqslant t \leqslant s \\ \Phi_s\left(\rho\left[\dfrac{t-s}{1-s}\right]\right) & \text{if } s \leqslant t \leqslant 1 \end{cases}.$$

In other words, we take the curve $\Phi_s\rho$, squeeze it into $[s,1]$, and use the heat flow to join ρ_0 to ρ_s on $[0,s]$. In particular, $\tilde{\rho}^s_0 = \rho_0 = \bar{\rho}_0$ and $\tilde{\rho}^s_1 = \Phi_s\rho_1$. From $U_1(\rho_0) < +\infty$ and the fact that U_1 is decreasing along the heat flow (see Theorem 2.11), $U_1(\tilde{\rho}_t)$ is bounded by $U_1(\rho_0)$ if $t \in [0,s]$ and by a constant C_s (depending only on s and Ω) if $t \in [s,1]$. Hence, for any $s > 0$, there exists $C < +\infty$ such that $U_1(\tilde{\rho}^s_t) \leqslant C$ for any $t \in [0,1]$.

It remains to show that $\mathcal{A}$ does not increase too much because of our regularization process. Because of the remark made above, one can see that

$$\int_0^1 F(\tilde{\rho}^s_t)\mathrm{d}t + G(\tilde{\rho}^s_1) \leqslant sF(\bar{\rho}_0) + (1-s)\int_0^1 F(\rho_t)\mathrm{d}t + G(\rho_1).$$

To handle the action of $\tilde{\rho}^s$, we remark thanks to the fourth point of Proposition 2.13 and the representation formula (2.4) that applying uniformly the heat flow decreases the action. Hence, performing a affine change of variables on $[s,1]$ and using (2.11),

$$\begin{aligned}\int_0^1 |\dot{\tilde{\rho}}^s_t|^2\mathrm{d}t &= \int_0^s |\dot{\Phi_t\rho_0}|^2\mathrm{d}t + \frac{1}{1-s}\int_0^1 |\dot{\Phi_s\rho_t}|^2\mathrm{d}t \\ &\leqslant U_1(\rho_0) - U_1(\Phi_s\rho_0) + \frac{1}{1-s}\int_0^1 |\dot{\rho}_t|^2\mathrm{d}t.\end{aligned}$$

By lower semi-continuity of U_1 and as $U_1(\rho_0) = U_1(\bar{\rho}_0)$ is finite, one concludes that

$$\limsup_{s\to 0}\int_0^1 |\dot{\tilde{\rho}}^s_t|^2\mathrm{d}t \leqslant \int_0^1 |\dot{\rho}_t|^2\mathrm{d}t.$$

Finally to handle the term involving the potentials, one uses, by continuity of the heat flow, that $\tilde{\rho}^s_t$ converges to ρ_t for any $t \in [0,1]$ as s goes to 0. As $\int_0^1 |\dot{\tilde{\rho}}^s_t|^2\mathrm{d}t$ is uniformly bounded, the family $(\tilde{\rho}^s)_{0\leqslant s<1}$ is uniformly equicontinuous, hence $\tilde{\rho}^s$ converges uniformly to ρ as $s \to 0$. This allows us to write

$$\lim_{s\to 0}\left[\int_0^1\int_\Omega V\mathrm{d}\tilde{\rho}^s_t\mathrm{d}t + \int_\Omega W\mathrm{d}\tilde{\rho}^s_1\right] = \int_0^1\int_\Omega V\mathrm{d}\rho_t\mathrm{d}t + \int_\Omega W\mathrm{d}\rho_1.$$

Gluing all the inequalities that we have collected, we see that $\limsup_{s\to 0}\mathcal{A}(\tilde{\rho}^s) \leqslant \mathcal{A}(\rho)$. Hence, it is enough to take $\tilde{\rho} := \tilde{\rho}^s$ for s small enough. □

Now, let us show how one can sample a continuous curve to get a discrete one that approximates it.

Proposition 4.27. *Let $\rho \in \Gamma$ with $\rho_0 = \bar{\rho}_0$ be such that $\int_0^1 U_1(\rho_t)\mathrm{d}t < +\infty$ and $\lambda > 0$ be fixed. For any $N \geqslant 1$ we can build a curve $\rho^N \in \Gamma_N$ with $\rho^N_0 = \bar{\rho}_0$ in such a way that*

$$\limsup_{N\to+\infty}\mathcal{A}^{N,\lambda}(\rho^N) \leqslant \mathcal{A}(\rho) + \lambda\int_0^1 U_1(\rho_t)\mathrm{d}t + \lambda_N U_1(\rho_1).$$

We recall that $\lambda_N = 0$ by default except if Assumption 4.5 holds.

Proof. We can assume $\mathcal{A}(\rho) < +\infty$. The idea is to sample ρ on a grid translated w.r.t. T^N. We start with the following observation.

$$\begin{aligned}\int_0^\tau \sum_{k=1}^{N-1} (F(\rho_{k\tau+s}) + \lambda U_1(\rho_{k\tau+s}))\,\mathrm{d}s &= \int_0^{1-\tau} (F(\rho_t) + \lambda U_1(\rho_t))\,\mathrm{d}t \\ &\leqslant \int_0^1 (F(\rho_t) + \lambda U_1(\rho_t))\,\mathrm{d}t + C\tau,\end{aligned}$$

where C depends only on the lower bounds of F and U_1. Therefore, there exists $s_N \in (0, \tau)$ such that

$$\tau \sum_{k=1}^{N-1} (F(\rho_{k\tau+s_N}) + \lambda U_1(\rho_{k\tau+s_N})) \leqslant \int_0^1 (F(\rho_t) + \lambda U_1(\rho_t)) + C\tau.$$

Let us define $\rho^N \in \Gamma_N$ by sampling ρ on the grid translated by s_N: for any $k \in \{0, 1, \dots, N\}$,

$$\rho^N := \begin{cases} \rho_0 & \text{if } k = 0 \\ \rho_1 & \text{if } k = N \\ \rho_{k\tau+s_N} & \text{if } 1 \leqslant k \leqslant N-1 \end{cases}.$$

As the boundary values are left unchanged and given the choice of s_N, it is clear that

$$\left(\mathcal{A}(\rho) + \lambda \int_0^1 U_1(\rho_t)\mathrm{d}t + \lambda_N U_1(\rho_1)\right) - \mathcal{A}^{N,\lambda}(\rho^N) \geqslant \int_0^1 \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t - \sum_{k=1}^N \frac{W_2^2(\rho^N_{(k-1)\tau}, \rho^N_{k\tau})}{2\tau} - C\tau.$$

The r.h.s. of the above equation is delicate to evaluate because of the non uniformity of the grid near the boundaries. Recall that if $t \leqslant s$ then $W_2^2(\rho_t, \rho_s) \leqslant (s-t)\int_t^s |\dot{\rho}_r|^2\mathrm{d}r$, hence

$$\begin{aligned}&\sum_{k=1}^N \frac{W_2^2(\rho^N_{(k-1)\tau}, \rho^N_{k\tau})}{2\tau} \\ &= \frac{W_2^2(\rho_0, \rho_{\tau+s_N})}{2\tau} + \sum_{k=2}^{N-1} \frac{W_2^2(\rho_{(k-1)\tau+s_N}, \rho_{k\tau+s_N})}{2\tau} + \frac{W_2^2(\rho_{(k-1)\tau+s_N}, \rho_1)}{2\tau} \\ &\leqslant \frac{\tau+s_N}{2\tau}\int_0^{\tau+s_N} \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t + \sum_{k=2}^{N-1}\int_{(k-1)\tau+s_n}^{k\tau+s_N} \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t + \frac{\tau-s_N}{2\tau}\int_{1-\tau+s_N}^1 \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t \\ &\leqslant \int_0^{\tau+s_N} |\dot{\rho}_t|^2\mathrm{d}t + \int_{\tau+s_N}^{1-\tau+s_N} \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t + \int_{1-\tau+s_N}^1 \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t \\ &\leqslant \int_0^1 \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t + \int_0^{2\tau} \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t.\end{aligned}$$

In particular, we have used $\tau + s_N \leqslant 2\tau$ and $\tau - s_N \leqslant \tau$. Letting $N \to +\infty$ (hence $\tau \to 0$), we end up with

$$\limsup_{N\to+\infty} \sum_{k=1}^N \frac{W_2^2(\rho^N_{(k-1)\tau}, \rho^N_{k\tau})}{2\tau} \leqslant \int_0^1 \frac{1}{2}|\dot{\rho}_t|^2\mathrm{d}t,$$

and this is enough to conclude. □

Corollary 4.28. *Under the assumptions of Theorems 4.4, 4.5 or 4.6, there exists $C < +\infty$ such that, uniformly in $N \geqslant 1$ and $\lambda \in (0, 1]$, one has*

$$\mathcal{A}^{N,\lambda}(\bar{\rho}^{N,\lambda}) \leqslant C.$$

Proof. If we are under the assumptions of Theorems 4.4 ot 4.5, we take $\rho \in \Gamma$ such that $\mathcal{A}(\rho) < +\infty$. As $U_1 \leqslant C_f F + C$, we see that $\int_0^1 U_1(\rho_t)\mathrm{d}t < +\infty$. If we are under the assumptions of 4.6, we take $\rho \in \Gamma$ such that $\mathcal{A}(\rho) < +\infty$ and regularize it thanks to Proposition 4.26. For this regularized curve, one has $\int_0^1 U_1(\rho_t)\mathrm{d}t + \lambda U_1(\rho_1) < +\infty$.

In any of these two cases, we construct ρ^N as in Proposition 4.27 and define $C := \sup_{N \geqslant 1} \mathcal{A}^{N,\lambda}(\rho^N)$, then we use the fact that $\mathcal{A}^{N,\lambda}(\bar{\rho}^{N,\lambda}) \leqslant \mathcal{A}^{N,\lambda}(\rho^N) \leqslant C$. □

4.4.2 Solution of the continuous problem as limit of discrete curves

We will build a suitable interpolation of the discrete curves $\bar{\rho}^{N,\lambda}$ that will converge to some continuous curve $\bar{\rho}$ as $N \to +\infty$ and $\lambda \to 0$, and we will show that $\bar{\rho}$ is a solution of (4.1).

As the order in which the limits $N \to +\infty$ and $\lambda \to 0$ are taken does not matter, we will do them in the same time. We take two sequences $(N_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ that go respectively to $+\infty$, and 0 (the second one being strictly positive). We will not relabel the sequences when extracting subsequences. Moreover, to avoid heavy notations, we will drop the index n, and $\lim_{n\to+\infty}$ will be denoted by $\lim_{N\to+\infty,\lambda\to 0}$. We will need to define two kind of interpolations: one filling the gaps with constant-speed geodesics, and the other one by using piecewise constant curves.

Definition 4.29. *If $N \geqslant 1$ and $\lambda > 0$, we define $\hat{\rho}^{N,\lambda} \in \Gamma$ as the curve such that $\hat{\rho}^{N,\lambda}$ coincides with $\bar{\rho}^{N,\lambda}$ on T^N, and such that for any $k \in \{0,1,\dots,N-1\}$, the restriction of $\hat{\rho}^{N,\lambda}$ to $[k\tau,(k+1\tau)]$ is the constant-speed geodesic joining $\bar{\rho}^{N,\lambda}_{k\tau}$ to $\bar{\rho}^{N,\lambda}_{(k+1)\tau}$.*

As $\bar{\rho}^{N,\lambda}_{k\tau}$ is absolutely continuous w.r.t. $\mathcal{L}$ for any $k \in \{1,2,\dots,N-1\}$, the constant-speed geodesic between $\bar{\rho}^{N,\lambda}_{k\tau}$ and $\bar{\rho}^{N,\lambda}_{(k\pm 1)\tau}$ is always unique. From the characterization of constant-speed geodesics, one has, for any $k \in \{0,1,\dots,N-1\}$,

$$\int_{k\tau}^{(k+1)\tau} \frac{1}{2}\left|\dot{\hat{\rho}}^{N,\lambda}_t\right|^2 \mathrm{d}t = \frac{W_2^2(\bar{\rho}^{N,\lambda}_{k\tau}, \bar{\rho}^{N,\lambda}_{(k+1)\tau})}{2\tau}.$$

Summing these identities over k,

$$\int_0^1 \frac{1}{2}\left|\dot{\hat{\rho}}^{N,\lambda}_t\right|^2 \mathrm{d}t = \sum_{k=1}^{N} \frac{W_2^2(\bar{\rho}^{N,\lambda}_{(k-1)\tau}, \bar{\rho}^{N,\lambda}_{k\tau})}{2\tau}. \tag{4.17}$$

In other words, the continuous action of the interpolated curve $\hat{\rho}^{N,\lambda}$ is equal to the discrete action of the discrete curve $\bar{\rho}^{N,\lambda}$.

Definition 4.30. *If $N \geqslant 1$ and $\lambda > 0$, we define $\tilde{\rho}^{N,\lambda} : [0,1] \to \mathcal{P}(\Omega)$ as the function such that $\tilde{\rho}^{N,\lambda}$ coincides with $\bar{\rho}^{N,\lambda}$ on T^N, and such that for any $k \in \{0,1,\dots,N-1\}$, the restriction of $\tilde{\rho}^{N,\lambda}$ to $[k\tau,(k+1\tau))$ is equal to $\bar{\rho}^{N,\lambda}_{k\tau}$.*

The curve $\tilde{\rho}^{N,\lambda}$ is not continuous as it might admit discontinuities at every point in T^N. Let us underline that the following identity trivially holds:

$$\sum_{k=1}^{N-1} \tau\left(F(\bar{\rho}^{N,\lambda}_{k\tau}) + \int_\Omega V \mathrm{d}\bar{\rho}^{N,\lambda}_{k\tau}\right) = \int_0^{1-\tau}\left(F(\tilde{\rho}^{N,\lambda}_t) + \int_\Omega V \mathrm{d}\tilde{\rho}^{N,\lambda}_t\right)\mathrm{d}t \tag{4.18}$$

Proposition 4.31. *Under the assumptions of Theorems 4.4, 4.5 or 4.6, there exists $\bar{\rho} \in \Gamma$ such that $\hat{\rho}^{N,\lambda}$ and $\tilde{\rho}^{N,\lambda}$ converge uniformly to $\bar{\rho}$ as $N \to +\infty$ and $\lambda \to 0$.*

Proof. Let us denote by C the constant given in Corollary 4.28. As all the terms in $\mathcal{A}^{N,\lambda}$ are bounded from below and given identity (4.17), one can see that there exists C_1 such that

$$\int_0^1 \frac{1}{2}\left|\dot{\hat{\rho}}_t^{N,\lambda}\right|^2 \mathrm{d}t \leqslant C_1$$

uniformly in $N \geqslant 1$ and $\lambda \in (0,1]$. Thus, by compactness of the sublevel sets of the action (Proposition 2.9), one concludes of the existence of $\bar{\rho} \in \Gamma$ such that $\hat{\rho}^{N,\lambda}$ converges uniformly (up to extraction) to $\bar{\rho}$ as $N \to +\infty$ and $\lambda \to 0$. Moreover, one can see that for any $t \in [0,1]$ and any $N \geqslant 1$, by setting k to be the largest integer such that $k\tau \leqslant t$, one has

$$W_2\left(\hat{\rho}_t^{N,\lambda}, \tilde{\rho}_t^{N,\lambda}\right) = W_2\left(\hat{\rho}_t^{N,\lambda}, \hat{\rho}_{k\tau}^{N,\lambda}\right) \leqslant \sqrt{\tau}\sqrt{\int_{k\tau}^t \left|\dot{\hat{\rho}}_s^{N,\lambda}\right|^2 \mathrm{d}s} \leqslant \sqrt{2C_1\tau}.$$

This allows to conclude that $\tilde{\rho}^{N,\lambda}$ also converges uniformly to $\bar{\rho}$ as $N \to +\infty$ and $\lambda \to 0$. □

Proposition 4.32. *Under the assumptions of Theorems 4.4, 4.5 or 4.6, the curve $\bar{\rho}$ is the solution to the continuous problem* (4.1)*.*

Proof. Taking the limit $N \to +\infty$ and $\lambda \to 0$ in (4.17), as the action is l.s.c., we end up with

$$\int_0^1 \frac{1}{2}|\dot{\bar{\rho}}_t|^2 \mathrm{d}t \leqslant \liminf_{N\to+\infty,\lambda\to 0} \sum_{k=1}^N \frac{W_2^2(\bar{\rho}_{(k-1)\tau}^{N,\lambda}, \bar{\rho}_{k\tau}^{N,\lambda})}{2\tau}.$$

Then, to handle the terms with the potential and the congestion, one can notice that for any $t \in [0,1]$, by lower semi-continuity of F and the convergence of $\tilde{\rho}_t^{N,\lambda}$ to $\bar{\rho}_t$,

$$F(\bar{\rho}_t) + \int_\Omega V \mathrm{d}\bar{\rho}_t \leqslant \liminf_{N\to+\infty,\lambda\to 0} F(\tilde{\rho}_t^{N,\lambda}) + \int_\Omega V \mathrm{d}\tilde{\rho}_t^{N,\lambda}.$$

Thus, using Fatou's lemma, as F, V and U_1 are bounded from below, one has for any $\tau_0 > 0$,

$$\begin{aligned}
\int_0^{1-\tau_0} \left(F(\bar{\rho}_t) + \int_\Omega V \mathrm{d}\bar{\rho}_t\right) &\leqslant \liminf_{N\to+\infty,\lambda\to 0} \int_0^{1-\tau} \left(F(\tilde{\rho}_t^{N,\lambda}) + \int_\Omega V \mathrm{d}\tilde{\rho}_t^{N,\lambda}\right) \mathrm{d}t \\
&= \liminf_{N\to+\infty,\lambda\to 0} \sum_{k=1}^{N-1} \tau \left(F(\bar{\rho}_{k\tau}^{N,\lambda}) + \int_\Omega V \mathrm{d}\bar{\rho}_{k\tau}^{N,\lambda}\right) \mathrm{d}t \\
&\leqslant \liminf_{N\to+\infty,\lambda\to 0} \sum_{k=1}^{N-1} \tau \left(F(\bar{\rho}_{k\tau}^{N,\lambda}) + \int_\Omega V \mathrm{d}\bar{\rho}_{k\tau}^{N,\lambda} + \lambda U_1(\bar{\rho}_{k\tau}^{N,\lambda})\right).
\end{aligned}$$

In the equation above, τ_0 is arbitrary thus it is still valid for $\tau_0 = 0$. As moreover the boundary penalization Ψ is l.s.c. and the entropic penalization $\lambda_N U_1(\rho_1)$ is positive, one is allowed to write that

$$\mathcal{A}(\bar{\rho}) \leqslant \liminf_{N\to+\infty,\lambda\to 0} \mathcal{A}^{N,\lambda}(\bar{\rho}^{N,\lambda}).$$

Let us assume by contradiction that there exists $\rho \in \Gamma$ such that $\mathcal{A}(\rho) < \mathcal{A}(\bar{\rho})$. Using, if needed, Proposition 4.26, we can assume without loss of generality that $\mathcal{A}(\rho) < \mathcal{A}(\bar{\rho})$ and $\int_0^1 U_1(\rho_t)\mathrm{d}t + \lambda_N U_1(\rho_1) < +\infty$. Using Proposition 4.27, for any $N \geqslant 1$, we can build $\rho^N \in \Gamma_N$ in such a way that

$$\limsup_{N\to+\infty} \mathcal{A}^{N,\lambda}(\rho^N) \leqslant \mathcal{A}(\rho) + \lambda \int_0^1 U_1(\rho_t)\mathrm{d}t + \lambda_N U_1(\rho_1).$$

Taking the limit $\lambda \to 0$, one can see that

$$\limsup_{N\to+\infty,\lambda\to 0} \mathcal{A}^{N,\lambda}(\rho^N) \leqslant \mathcal{A}(\rho) < \mathcal{A}(\bar\rho) \leqslant \liminf_{N\to+\infty,\lambda\to 0} \mathcal{A}^{N,\lambda}(\bar\rho^{N,\lambda}).$$

Taking N large enough and λ small enough, we conclude that $\mathcal{A}^{N,\lambda}(\rho^N) < \mathcal{A}^{N,\lambda}(\bar\rho^{N,\lambda})$, which is a contradiction with the optimality of $\bar\rho^{N,\lambda}$. □

4.4.3 Uniform bounds on $\bar\rho$

To conclude and prove the Theorems 4.4, 4.5 and 4.6, it is enough to show the L^∞ bounds on $\bar\rho$, which of course we will prove using the discrete solutions $\bar\rho^{N,\lambda}$. The key is the following proposition.

Proposition 4.33. *Let $0 < T_1 < T_2 \leqslant 1$. Then for any $0 < T_1' < T_1$ and any $T_2 < T_2' < 1$ (or $T_2' = T_2 = 1$ in the case $T_2 = 1$),*

$$\operatorname*{ess\,sup}_{T_1 \leqslant t \leqslant T_2, x\in\Omega} |\bar\rho_t(x)| \leqslant \limsup_{m\to+\infty} L^m_{T_1',T_2'}.$$

Proof. We rely on the well-known identity

$$\operatorname*{ess\,sup}_{T_1\leqslant t\leqslant T_2,\ x\in\Omega} |\bar\rho_t(x)| = \limsup_{m\to+\infty} \left(\int_{T_1}^{T_2}\int_\Omega \bar\rho_t^m \mathrm{d}t\right)^{1/m} = \limsup_{m\to+\infty}\left(\int_{T_1}^{T_2} U_m(\bar\rho_t)\mathrm{d}t\right)^{1/m}.$$

For a fixed $m > 1$ and for $\tau > 0$ small enough, one has

$$\int_{T_1}^{T_2} U_m(\tilde\rho_t^{N,\lambda})\mathrm{d}t \leqslant \sum_{T_1'\leqslant k\tau\leqslant T_2'} \tau U_m(\bar\rho^{N,\lambda}_{k\tau}).$$

When sending $N \to \infty$ and $\lambda \to 0$, by lower semi-continuity of U_m and by convergence of $\tilde\rho^{N,\lambda}$ to $\bar\rho$, we know that

$$\begin{aligned}\int_{T_1}^{T_2} U_m(\bar\rho_t)\mathrm{d}t &\leqslant \liminf_{N\to+\infty,\lambda\to 0}\int_{T_1}^{T_2} U_m(\tilde\rho_t^{N,\lambda})\mathrm{d}t \\ &\leqslant \liminf_{N\to+\infty,\lambda\to 0} \sum_{T_1'\leqslant k\tau\leqslant T_2'} \tau U_m(\bar\rho^{N,\lambda}_{k\tau}).\end{aligned}$$

Taking the power $1/m$ on each side and by definition of $L^m_{T_1',T_2'}$, one gets

$$\left(\int_{T_1}^{T_2} U_m(\bar\rho_t)\mathrm{d}t\right)^{1/m} \leqslant L^m_{T_1',T_2'}.$$

It is enough to take the limit $m \to +\infty$ to get the announced inequality. □

We can now conclude the desired bounds:

Proof of Theorem 4.4. Combining Proposition 4.33 and Proposition 4.18, it is enough to show that $L^{\alpha+2}_{0,1} < +\infty$. Because of Assumption 4.1 or 4.2, we know that $U_{\alpha+2} \leqslant C_1 F + C_2$ with $C_1 > 0$. Hence, in order to conclude that $L^{\alpha+2}_{0,1} < +\infty$, it is enough to use Corollary 4.28, which provides a constant $C < +\infty$ such that for any $N \geqslant 1$ and any $\lambda \in (0,1]$ we have

$$\sum_{k=1}^{N-1} \tau F(\bar\rho^{N,\lambda}_{k\tau}) \leqslant C.$$ □

Proof of Theorem 4.5. We combine Proposition 4.33 and Proposition 4.21, as we saw that $L_{0,1}^{\alpha+2} < +\infty$ (in the proof of Theorem 4.4). □

Proof of Theorem 4.6. It is enough to combine Proposition 4.33 with Propositions 4.23 and 4.24. □

4.5 Apprendix: reverse Jensen inequality

In this section, we prove Lemma 4.17 (the "reverse Jensen inequality") as well as Lemmas 4.20 and 4.25, whose proofs were postponed in order not to overload the key arguments of this chapter. In all the sequel, we consider a family of sequences $(u_k^\tau)_{k\in\mathbb{Z}}$ indexed by a parameter $\tau > 0$. We assume that there exists $\omega \geqslant 0$ such that for any $k \in \mathbb{Z}$, one has $u_k^\tau > 0$ and

$$\frac{u_{k+1}^\tau + u_{k-1}^\tau - 2u_k^\tau}{\tau^2} + \omega^2 u_k^\tau \geqslant 0. \tag{4.19}$$

This inequation is a discrete counterpart of the differential inequality $u'' + \omega^2 u \geqslant 0$. Let us remark, by the positivity of u_k^τ, that we can assume without loss of generality that $\omega > 0$, even though the proofs are considerably simpler if $\omega = 0$: the constants would be better, and the strategy of the proof would be slightly different. The key point to handle u_k^τ is to compare it with explicit sequences realizing the opposite inequality in (4.19).

Definition 4.34. *For any $\tau > 0$, let $\mathcal{T}^\tau$ be the set of sequences $(v_k)_{k\in\mathbb{Z}}$ of the form $v_k = A\cos(2\omega k\tau + \delta)$.*

Lemma 4.35. *There exists $\tau_0 > 0$ such that for any $\tau \leqslant \tau_0$, if $(v_k)_{k\in\mathbb{Z}} \in \mathcal{T}^\tau$ and k is such that $v_k > 0$ then*

$$\frac{v_{k+1} + v_{k-1} - 2v_k}{\tau^2} + \omega^2 v_k < 0$$

Proof. This is a consequence of the trigonometric identity

$$\frac{v_{k+1} + v_{k-1} - 2v_k}{\tau^2} + \omega^2 v_k = \left(2\frac{\cos(2\omega\tau) - 1}{\tau^2} + \omega^2\right) v_k$$

and the fact that $2\dfrac{\cos(2\omega\tau) - 1}{\tau^2} + \omega^2 \sim -3\omega^2$ as τ goes to 0. □

We also note the following properties on the sequences in $\mathcal{T}^\tau$, that we do not prove and leave to the reader as an exercise.

- if $k_1 < k_2$ are fixed with $|k_2 - k_1|\tau\omega < \pi/8$ and τ is small enough, then for every fixed positive values $a_1, a_2 > 0$ there exists a unique sequence in $\mathcal{T}^\tau$ with $v_{k_1} = a_1$ and $v_{k_2} = a_2$. Moreover, such a sequence $(v_k)_{k\in\mathbb{Z}}$ is such that there exists an open interval I of the form either $(k_0\tau, k_1\tau)$ or $(k_2\tau, k_3\tau)$, with length at least $\pi/(8\omega)$, with $v_k > 0$ for all the indices k such that $k\tau \in I$.

- if $k_1 < N$ and $b \geqslant 0$ are fixed and $|N - k_1|\tau < \min\{\pi/(8\omega), \pi/(8b)\}$ and τ is small enough, then for every $a > 0$ there exists a unique sequence in $\mathcal{T}^\tau$ with $v_{k_1} = a$ and $(v_N - v_{N-1})/\tau = bv_N$. Moreover, such a sequence $(v_k)_{k\in\mathbb{Z}}$ is such that there exists an open interval I of the form $(k_0\tau, k_1\tau)$ with length at least $\min\{\pi/(32\omega), \pi/(32b)\}$, with $v_k > 0$ for all the indices k such that $k\tau \in I$.

Note that, for the purpose of Lemma 4.35 and of the subsequent observations other choices of v_k were possible, such as $v_k = A\cos((1+\varepsilon)\omega k\tau + \delta)$ for some $\varepsilon > 0$, but we chose $\varepsilon = 1$ for simplicity in the next computations (more generally in this appendix we have not been looking for the sharpest constants). Indeed, all these results are not surprising: at the continuous level v solves $v'' + 4\omega^2 v = 0$ and most of the discrete results are just an adaptation of this property. The important point is the following comparision principle between $(u^\tau_k)_{k\in\mathbb{Z}}$ and $(v_k)_{k\in\mathbb{Z}}$.

Lemma 4.36. *Let $k_1 < k_2$ such that $|k_2 - k_1|\tau\omega < \pi/8$ and assume $\tau \leqslant \tau_0$. Let $v \in \mathcal{T}^\tau$ the unique element of $\mathcal{T}^\tau$ such that $v_{k_1} = u^\tau_{k_1}$ and $v_{k_2} = u^\tau_{k_2}$. Let k_0 (resp. k_3) be the largest (resp. the smallest) index smaller that k_1 (resp. larger than k_2) such that $v_{k_0-1} < 0$ (resp. $v_{k_3+1} < 0$).*

Then $u^\tau_k \leqslant v_k$ for any $k_1 \leqslant k \leqslant k_2$ and $u^\tau_k \geqslant v_k$ for any $k_0 \leqslant k \leqslant k_1$ and any $k_2 \leqslant k \leqslant k_3$.

In other words, u^τ is below v between k_1 and k_2 and above outside k_1 and k_2 (as long as $v \geqslant 0$).

Proof. The fact that there exists only one $v \in \mathcal{T}^\tau$ such that $v_{k_1} = u^\tau_{k_1}$ and $v_{k_2} = u^\tau_{k_2}$ has been already observed above. Let us define $w_k = u^\tau_k - v_k$. By (4.19) and Lemma 4.35,

$$\frac{w_{k+1} + w_{k-1} - 2w_k}{\tau^2} + \omega^2 w_k > 0 \tag{4.20}$$

for any $k_0 \leqslant k \leqslant k_3$ and $w_{k_1} = w_{k_2} = 0$. We want to prove $w_k \leqslant 0$ for every $k_1 \leqslant k \leqslant k_2$. We consider the piecewise affine interpolation $\bar{w}$ of the values w_k: a function which is affine on each interval $[k\tau, (k+1)\tau]$ and is equal to w_k at the point $k\tau$. The condition (4.20) translates on $\bar{w}$ as differential inequality in the sense of distributions:

$$\bar{w}'' + \omega^2 \sum_k \tau w_k \delta_{k\tau} \geqslant 0. \tag{4.21}$$

Let us assume by contradiction that there is an open interval $I \subset (k_1\tau, k_2\tau)$ on which $\bar{w} > 0$, with $\bar{w} = 0$ on ∂I. We denote by $|I|$ the length of such an interval, and we have $|I| \leqslant |k_2 - k_1|\tau$. By multiplying the above inequality by $\bar{w}$ and integrating by parts we get

$$\int_I |\bar{w}'|^2 = -\int_I \bar{w}''\bar{w} \leqslant \omega^2 \sum_{k\,:\,k\tau\in I} \tau|w_k|^2.$$

Then, we observe that we have, for each k s.t. $k\tau \in I$,

$$|w_k| \leqslant \frac{1}{2}\int_I |\bar{w}'| \leqslant \frac{1}{2}\sqrt{|I|\int_I |\bar{w}'|^2}.$$

The reason for the factor 1/2 in the above inequality is the possibility to choose to integrate $\bar{w}'$ on an interval at the right or at the left of $k\tau$, and to choose the one where the integral of $|\bar{w}'|$ is smaller. This implies

$$\int_I |\bar{w}'|^2 \leqslant \omega^2\tau\ \#\{k : k\tau \in I\}\frac{1}{4}|k_2 - k_1|\tau \int_I |\bar{w}'|^2.$$

Since $\{k : k\tau \in I\} \subset \{k : k_1 < k < k_2\}$, we have $\#\{k : k\tau \in I\} < |k_2 - k_1|$ and the contradiction comes from the assumption $\omega\tau|k_2 - k_1| < \pi/8 < 2$.

In order to prove $w_k \geqslant 0$ for $k_0 \leqslant k \leqslant k_1$, we first observe that (4.20) for $k = k_1$, now that we know $w_{k_1+1} \leqslant 0$, implies $w_{k_1-1} > 0$. If for some k with $k_0 \leqslant k \leqslant k_1$ we had $w_k < 0$, then we could find an open interval $J \subset (k_0\tau, k_1\tau)$ where $\bar{w} > 0$ with $\bar{w} = 0$ on ∂J. We then apply the same approach as above, thus obtaining

$$\int_J |\bar{w}'|^2 \leqslant \omega^2 \tau \; \#\{k : k\tau \in J\} \frac{1}{4} |J| \int_J |\bar{w}'|^2.$$

It is important to not that J is contained in an interval of positivity of a function of the form $A\cos(2\omega t + \delta)$, whose length is $\pi/(2\omega)$; the number of points of the form $k\tau$ contained in an interval of such a length is at most $\pi/(2\omega\tau) + 1$ but for $k = k_1, k_2$ the point $k\tau$ does not belong to the open interval J. Hence $\#\{k : k\tau \in J\} \leqslant \pi/(2\omega\tau)$, and we have a contradiction since $\pi^2 < 16$. □

We provide now a variant in the case where on the interval $(k_1\tau, k_2\tau)$ we impose a different boundary condition on the right end side.

Lemma 4.37. *Let $k_1 < N$ and $b \geqslant 0$ such that $|N - k_1|\tau < \min\{\pi/(8\omega), \pi \; (8b)\}$ and assume $\tau \leqslant \tau_0$. Suppose $(u_N - u_{N-1})/\tau \leqslant b u_N$. Let $v \in \mathcal{T}^\tau$ the unique element of $\mathcal{T}^\tau$ such that $v_{k_1} = u^\tau_{k_1}$ and $(v_N - v_{N-1})/\tau = b v_N$. Let k_0 be the largest (resp. the smallest) index smaller that $k_1$such that $v_{k_0-1} < 0$.*

Then $u^\tau_k \leqslant v^\tau_k$ for any $k_1 \leqslant k \leqslant N$ and $u^\tau_k \geqslant v^\tau_k$ for any $k_0 \leqslant k \leqslant k_1$.

Proof. The argument is very similar to the one in Lemma 4.36. We first define $w_k = u_k - v_k$, as well as the piecewise affine interpolation $\bar{w}$ of the values w_k, which satisfies again (4.21), but also $w'(1) \leqslant bw(1)$, (recall tha $N\tau = 1$).

Then, we assume by contradiction that there is an open interval $I \subset (k_1\tau, N\tau)$ on which $\bar{w} > 0$. If $\bar{w} = 0$ on ∂I (i.e., on both points on the boundary), the argument is really the same. Otherwise, we can assume $I = (t, 1)$, with $\bar{w}(t) = 0$. By multiplying by $\bar{w}$ and integrating by parts we get

$$\int_I |\bar{w}'|^2 = \bar{w}(1)\bar{w}'(1) - \int \bar{w}''\bar{w} \leqslant b|\bar{w}(1)|^2 + \omega^2 \sum_{k \,:\, k\tau \in I} \tau |w_k|^2.$$

Then, we use that on I we have

$$|\bar{w}| \leqslant \int_I |\bar{w}'| \leqslant \sqrt{|I| \int_I |\bar{w}'|^2}.$$

We do not have anymore the factor 1/2 because $\bar{w}$ only vanishes at one end, now. This implies

$$\int_I |\bar{w}'|^2 \leqslant |I| \left(\omega^2 \tau \; \#\{k : k\tau \in I\} + b\right) \int_I |\bar{w}'|^2.$$

Since $\#\{k : k\tau \in I\} < |N - k_1|$ and $|I| \leqslant |N - k_1|\tau$, using the assumptions on $|N - k_1|$ we have

$$\int_I |\bar{w}'|^2 \leqslant \left(\frac{\pi}{8} + \left(\frac{\pi}{8}\right)^2\right) \int_I |\bar{w}'|^2.$$

This is a contradiction, since

$$\frac{\pi}{8} + \left(\frac{\pi}{8}\right)^2 < \frac{1}{2} + \frac{1}{4} < 1.$$ □

With the two lemma above, we are able to deduce some Harnack-type inequality, which means that we can control the values of a u satisfying (4.19) in the interior of an interval with the values of u outside the interval.

Lemma 4.38. *Let $k_1 < k_2$ such that $|k_2 - k_1|\tau\omega < \pi/8$ and assume $\tau \leqslant \tau_0$. Let k_0 (resp. k_3) be the smallest (resp. largest) integer smaller than k_1 (resp. larger than k_2) such that $(k_1 - k_0)\tau\omega < \pi/8$ (resp. $(k_3 - k_2)\tau\omega < \pi/8$). Then one has*

$$\sup_{k_1 \leqslant k \leqslant k_2} u_k^\tau \leqslant C \max\left(\inf_{k_0 \leqslant k \leqslant k_1} u_k^\tau \ , \ \inf_{k_2 \leqslant k \leqslant k_3} u_k^\tau \right),$$

where the constant C is universal.

Proof. Given the symmetry of the property we want to prove w.r.t. to time reversal, we can assume that $u_{k_1}^\tau \leqslant u_{k_2}^\tau$. Let $v \in \mathcal{T}^\tau$ be the unique element of $\mathcal{T}^\tau$ such that $v_{k_1} = u_{k_1}^\tau$ and $v_{k_2} = u_{k_2}^\tau$. We know that it can be written in the form $v_k = A\cos(k\tau\omega + \delta)$ with $A \geqslant 0$. In particular, $A \geqslant |v_k|$ for any $k \in \mathbb{Z}$. Up to a time translation, we can assume that $\delta = 0$ and $k_1 \leqslant 0 \leqslant k_2$. By the hypothesis $u_{k_1}^\tau \leqslant u_{k_2}^\tau$, and $|k_2 - k_1|\tau\omega < \pi/8$, we can even say that $|k_2| \leqslant |k_1|$; thus, one has $k_2\tau \leqslant \pi/(16\omega)$. In particular, for any $k_2 \leqslant k \leqslant k_3$, we can say more than $v_k > 0$:

$$\begin{aligned} v_k &\geqslant A\cos(2\omega k_3 \tau) \\ &\geqslant A\cos(2k_2\omega\tau + 2(k_3 - k_2)\omega\tau) \\ &\geqslant \cos\left(\frac{\pi}{8} + \frac{\pi}{4}\right) \sup_{k' \in \mathbb{Z}} |v_{k'}| \\ &\geqslant \frac{1}{C} \sup_{k' \in \mathbb{Z}} |v_{k'}|, \end{aligned}$$

with $C = \cos(3\pi/8)^{-1} < +\infty$. Thus, by using the comparison between u^τ and v (Lemma 4.36), one can say that, for any $k_2 \leqslant k \leqslant k_3$,

$$u_k^\tau \geqslant \frac{1}{C} \sup_{k_1 \leqslant k'\tau \leqslant k_2} u_{k'}^\tau,$$

which easily implies the claim. □

We also provide the same type of lemma but where a different condition is imposed on the right end side, namely a Neumann-type boundary condition.

Lemma 4.39. *Let $k_1 < N$ and $b \geqslant 0$ such that $|N - k_1|\tau \leqslant \min\{\pi/(32\omega), \pi/(32b)\}$ and assume $\tau \leqslant \tau_0$. Suppose $(u_N - u_{N-1})/\tau \leqslant bu_N$. Let k_0 be the smallest integer smaller than k_1 such that $(k_1 - k_0)\tau \leqslant \min\{\pi/(32\omega), \pi/(32b)\}$. Then one has*

$$\sup_{k_1 \leqslant k \leqslant N} u_k^\tau \leqslant C \inf_{k_0 \leqslant k \leqslant k_1} u_k^\tau,$$

where the constant C is universal.

Proof. The strategy of the proof is the same than for Lemma 4.38. We take v to be the unique element of $\mathcal{T}^\tau$ such that $v_{k_1} = u_{k_1}^\tau$ and $(v_N - v_{N-1})/\tau = bv_N$. We know that v is of the form $v_k = A\cos(2k\tau\omega + \delta)$. Up to a time translation, we can assume that $N\tau = 0$ and take $\delta \in (-\pi/2, \pi/2)$. Starting from $(v_N - v_{N-1})/\tau = bv_N$ and using well known factorization formulas, one ends up with

$$b = -2\omega\tan(\delta) + O(\omega\tau).$$

Thus, if $\tau \leqslant \tau_0$, one can say that $\arctan(-b/\omega) \leqslant \delta \leqslant \arctan(-b/(4\omega))$. Hence, using the fact that $\arctan(t) + \arctan(1/t) = -\pi/2$ (if $t < 0$) and that $\min\{\pi t/4, \pi/4\} \leqslant \arctan(t) \leqslant t$ (if $t \geqslant 0$), one concludes that

$$\min\left\{-\frac{\pi}{2} + \frac{\pi\omega}{4b}, -\frac{\pi}{4}\right\} \leqslant \delta \leqslant \min\left\{-\frac{\pi}{2} + \frac{4\omega}{b}, 0\right\}.$$

In other words, δ cannot be too close to $-\pi/2$, the point where the cosine vanishes. Given the information that we have on k_1 and k_0, one can check that

$$\begin{aligned}\delta - 2\omega\tau k_0 &= \delta - 2\omega\tau k_1 - 2\omega\tau(k_0 - k_1) \\ &\geqslant \min\left\{-\frac{\pi}{2} + \frac{\pi\omega}{4b}, -\frac{\pi}{4}\right\} - 2\min\left\{\frac{\pi}{16}, \frac{\pi\omega}{16b}\right\} \\ &\geqslant \min\left\{-\frac{\pi}{2} + \frac{\pi\omega}{8b}, -\frac{3\pi}{8}\right\}.\end{aligned}$$

As, for every $k_0 \leqslant k \leqslant N$, one has

$$A\cos(\delta - 2\omega\tau k_0) \leqslant v_k \leqslant A\cos(\delta),$$

it is easy to conclude that

$$\frac{\sup_{k_0 \leqslant k \leqslant N} v_k}{\inf_{k_0 \leqslant k \leqslant N} v_k} \leqslant \frac{\cos(\min\{-\frac{\pi}{2} + \frac{4\omega}{b}, 0\})}{\cos(\min\{-\frac{\pi}{2} + \frac{\pi\omega}{8b}, \frac{3\pi}{8}\})} \leqslant C,$$

where the value of C can be estimated by noting that if $\omega/b \ll 1$ both the numerator and the denominator are of the order of ω/b and if ω/b is not small the denominator is far from 0 and the numerator is bounded by 1. This proves that C is a universal constant. It remains to use Lemma 4.37 to transfer the above inequality into an information on u^τ. □

To conclude, we can prove the Lemmas 4.17, 4.20 and 4.25 that we used throughout the chapter, by using the above results. To prove Lemma 4.17, we cut the interval $[T_1, T_2]$ into several pieces of length of order $1/\omega$, on each piece we use the Harnack inequality to exchange the sum and the power $1/\beta$, and we use rough comparisons to put the pieces together.

Proof of Lemma 4.17. Let M be the smallest integer larger than $8\omega(T_2 - T_1)/\pi + 1$. We cut the interval $[T_1, T_2]$ into M closed intervals $I_1, I_2, \dots, I_M$ of equal length (all equal to $(T_2 - T_1)/M < \pi/(8\omega)$). Let us choose an interval I_i, we can use Lemma 4.38 to write

$$\begin{aligned}\left(\sum_{k\,:\,k\tau\in I_i} \tau u_k^\tau\right)^{1/\beta} &\leqslant (|I_i| + \tau)^{1/\beta} \sup_{k\tau \in I_i} (u_\tau^k)^{1/\beta} \\ &\leqslant C(|I_i| + \tau)^{1/\beta}\left(\inf_{T_1^i - \eta \leqslant k\tau \leqslant T_1^i} (u_k^\tau)^{1/\beta} + \inf_{T_2^i \leqslant k\tau \leqslant T_2^i + \eta} (u_k^\tau)^{1/\beta}\right) \\ &\leqslant C\frac{(|I_i| + \tau)^{1/\beta}}{\eta} \sum_{k\,:\,k\tau \in I_i \pm \eta} (u_k^\tau)^{1/\beta},\end{aligned}$$

where $I_i \pm \eta$ denotes the set of real numbers which are at a distance at most η of I_i. Then we put together the estimate for each I_i:

$$\begin{aligned}
\left(\sum_{T_1 \leqslant k\tau \leqslant T_2} \tau u_k^\tau\right)^{1/\beta} &\leqslant \left(\sum_{i=1}^{M} \sum_{k\,:\,k\tau \in I_i} \tau u_k^\tau\right)^{1/\beta} \\
&\leqslant M^{1/\beta} \sum_{i=1}^{M} \left(\sum_{k\,:\,k\tau\in I_i} \tau u_k^\tau\right)^{1/\beta} \\
&\leqslant C M^{1/\beta} \left(\frac{T_2 - T_1}{M} + \tau\right)^{1/\beta} \frac{1}{\eta} \sum_{i=1}^{M} \sum_{k\,:\,k\tau \in I_i \pm \eta} \tau \left(u_k^\tau\right)^{1/\beta} \\
&\leqslant C \frac{M(T_2 - T_1 + M\tau)^{1/\beta}}{\eta} \sum_{T_1 - \eta \leqslant k\tau \leqslant T_2 + \eta} \tau \left(u_k^\tau\right)^{1/\beta} \\
&\leqslant C \frac{(\omega+1)(T_2 - T_1 + 1)^{1+1/\beta}}{\eta} \sum_{T_1 - \eta \leqslant k\tau \leqslant T_2 + \eta} \tau \left(u_k^\tau\right)^{1/\beta},
\end{aligned}$$

where we have used the fact that $M\tau \leqslant 1$ if $\tau \leqslant \tau_0$ (where τ_0 depends on ω) and also that M can be estimated by a constant times $\omega + 1$. ☐

Proof of Lemma 4.20. For the first part, we apply Lemma 4.39 with $k_1 = N$. With the choice of η, one has $(k_1 - k_0)\tau \leqslant \min\{\pi/(32\omega), \pi/(32b)\}$. Thus, one can write that

$$u_N^\tau \leqslant C \inf_{1-\eta \leqslant k\tau \leqslant 1} u_k^\tau,$$

which is enough to to conclude as the r.h.s. is bounded by the mean of u_k^τ, for $1 - \eta \leqslant k\tau \leqslant 1$.

For the second part (which is a variant of Lemma 4.17, but with Neumann boundary conditions on one side), we can say with the help of Lemma 4.39 that with k_1 the smallest integer smaller than N such that $|N - k_1|\tau \max\{\omega, b\} < \pi/32$,

$$\begin{aligned}
\left(\sum_{k_1\tau \leqslant k\tau \leqslant 1} \tau u_k^\tau\right)^{1/\beta} &\leqslant |1 - k_1\tau + \tau|^{1/\beta} \sup_{k_1\tau \leqslant k\tau \leqslant 1} (u_k^\tau)^{1/\beta} \\
&\leqslant C|1 - k_1\tau + \tau|^{1/\beta} \inf_{k_1\tau - \eta \leqslant k\tau \leqslant k_1\tau} (u_k^\tau)^{1/\beta} \\
&\leqslant \frac{C|1 - k_1\tau + \tau|^{1/\beta}}{\eta} \sum_{k_1\tau - \eta \leqslant k\tau \leqslant k_1\tau} \tau (u_k^\tau)^{1/\beta}.
\end{aligned}$$

Then, we combine this estimate with the interior estimate Lemma 4.17 (with $T_2 = 1 - k_1\tau$) to end up with the announced result. ☐

Proof of Lemma 4.25. We apply Lemma 4.39 with $k_1 = 0$. Thus if $1 = \tau N \leqslant T_{\max} = \min\{\pi/(32\omega), \pi/(32b)\}$, one has

$$\sup_{0 \leqslant k\tau \leqslant 1} u_k^\tau \leqslant C u_0^\tau = Ca.$$

Thus, the l.h.s. is bounded by a constant which does not depend on N. ☐

Chapter 5

Regularity of the pressure in the case of hard congestion

In this chapter, we tackle the problem of optimal density evolution with hard congestion which reads

$$\min_{\rho} \left\{ \int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t + \int_0^1 E(\rho_t)\mathrm{d}t + \int_\Omega \Psi \mathrm{d}\rho_1 \ : \ \rho \in \Gamma, \ \rho_0 \text{ given} \right\}.$$

where $\Gamma = C([0,1], \mathcal{P}(\Omega))$ and $|\dot{\rho}_t|$ is the metric derivative of ρ. The functional E will have the form

$$E(\rho) = \begin{cases} \int_\Omega V \mathrm{d}\rho & \text{if } \rho \leqslant 1 \text{ a.e. on } [0,1] \times \Omega, \\ +\infty & \text{otherwise} \end{cases}$$

The goal is to show that the pressure arising from the incompressibility constraint exhibits some Sobolev regularity if this is the case for the potentials V and Ψ. Compared to the previous chapter, we underline that the final cost is the integration against a given potential (called $\Psi : \Omega \to \mathbb{R}$) and not an arbitrary function $\Psi : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$

In this chapter, we say that a measure $\mu \in \mathcal{P}(\Omega)$ satisfies $\mu \leqslant 1$ if μ has a density w.r.t. $\mathcal{L}$ and this density is bounded a.e. by 1.

5.1 Statement of the problem and regularity of the pressure

5.1.1 Primal and dual problem

Assumptions. *The assumptions that will hold throughout this chapter are the following.*

(A1) The domain Ω is the closure of an open bounded convex subset of $\mathbb{R}^d$ with Lebesgue measure $|\Omega|$ strictly larger than 1.

(A2) We fix $V \in H^1(\Omega)$ (the "running cost") and assume that it is positive.

(A3) We fix $\Psi \in H^1(\Omega)$ (the "final cost") and assume that it is positive.

(A4) We take $\bar{\rho}_0 \in \mathcal{P}(\Omega)$ (the initial probability measure) such that $\bar{\rho}_0 \leqslant 1$.

We denote by $\Gamma_0 \subset \Gamma$ the set of curves such $\rho \in \Gamma$ that $\rho_0 = \bar{\rho}_0$. As we will see below in the definition of the primal problem, it does not change anything to add a constant to V or Ψ, hence (A2) and (A3) are equivalent to ask that V and Ψ are bounded from below.

The primal objective functional reads

$$\mathcal{A}(\rho) := \begin{cases} \int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t + \int_0^1 \left(\int_\Omega V \mathrm{d}\rho_t \right) \mathrm{d}t + \int_\Omega \Psi \mathrm{d}\rho_1 & \text{if } \rho_t \leqslant 1 \text{ for all } t \in [0,1], \\ +\infty & \text{else.} \end{cases}$$

Definition 5.1. *The primal problem is*

$$\min_\rho \{\mathcal{A}(\rho) \ : \ \rho \in \Gamma_0\} \tag{5.1}$$

We will need to consider the dual of this problem. Let $\phi \in C^1([0,1] \times \Omega)$ and $P \in C([0,1] \times \Omega)$ be smooth functions with P positive and in such a way that the Hamilton Jacobi equation is satisfied as an inequality

$$-\partial_t \phi + \frac{1}{2}|\nabla \phi|^2 \leqslant V + P \tag{5.2}$$

and the final value of ϕ is constrained by

$$\phi(1, \cdot) \leqslant \Psi. \tag{5.3}$$

The dual functional, at least for smooth functions, is defined as follows:

$$\mathcal{B}(\phi, P) := \int_\Omega \phi(0, \cdot)\bar{\rho}_0 - \iint_{[0,1]\times\Omega} P.$$

and there is no duality gap between the primal and the dual problem. The reader can refer to Section 3.1 to understand where this expression comes from. However, to get existence of a solution of the dual problem, it is too restrictive to look only at smooth functions. As understood in [CMS16], the right functional space is the following.

Definition 5.2. *Let $\mathcal{K}$ be the set of pairs (ϕ, P) where $\phi \in \mathrm{BV}([0,1] \times \Omega) \cap L^2([0,1], H^1(\Omega))$ and $P \in \mathcal{M}_+([0,1] \times \Omega)$ is a positive measure, and the Hamilton Jacobi equation (5.2) is understood in the distributional sense, provided we set $\phi(1^+, \cdot) = \Psi$ and that we take in account the possible jump from $\phi(1^-, \cdot)$ to $\phi(1^+, \cdot)$ in the temporal distributional derivative.*

For $(\phi, P) \in \mathcal{K}$, the dual functional is understood in the following sense:

$$\mathcal{B}(\phi, P) := \int_\Omega \phi(0^+, \cdot)\bar{\rho}_0 - P([0,1] \times \Omega).$$

Notice, that we can always assume $\phi(0^-, \cdot) = \phi(0^+, \cdot)$ (no jump for $t = 0$, otherwise it would increase $\mathcal{B}(\phi, P)$) but we set $\phi(1^+, \cdot) = \Psi$. The measure P can have a part concentrated on $t = 1$, which may lead to $\phi(1^-, \cdot) > \phi(1^+, \cdot) = \Psi$, provided the jump is compensated by the part of P on $\{1\} \times \Omega$.

Definition 5.3. *The dual problem is*

$$\max_{\phi, P} \{\mathcal{B}(\phi, P) \ : \ (\phi, P) \in \mathcal{K}\}. \tag{5.4}$$

These two problems are in duality in the following sense [CMS16, Propositions 3.3 and 3.8].

Theorem 5.4. *There holds*

$$\min_\rho \{\mathcal{A}(\rho) \ : \ \rho \in \Gamma_0\} = \max_{\phi, P} \{\mathcal{B}(\phi, P) \ : \ (\phi, P) \in \mathcal{K}\}.$$

Notice that the existence of a solution to both the primal and the dual problem are included in this statement. The main result of this chapter is the following:

Theorem 5.5. *There exists a solution* $(\bar{\phi}, \bar{P})$ *of the dual problem such that:*

- *The restriction of* $\bar{P}$ *to* $[0,1) \times \Omega$ *has a density w.r.t. to the* $d+1$*-dimensional Lebesgue measure and this density, denoted by* $\bar{p}$*; satisfies* $\|\nabla \bar{p}(t,\cdot)\|_{L^2(\Omega)} \leqslant \|\nabla V\|_{L^2(\Omega)}$ *for a.e.* $t \in [0,1]$*. Moreover, if* $V \in W^{1,q}(\Omega)$ *with* $q > d$*, then* $\|\bar{p}\|_{L^\infty([0,1)\times\Omega)} \leqslant C < +\infty$ *with* C *depending only on* $\|\nabla V\|_{L^q(\Omega)}$ *and* Ω*.*

- *The restriction of* $\bar{P}$ *to* $\{1\} \times \Omega$ *has a density w.r.t. to the* d*-dimensional Lebesgue measure and this density (denoted by* $\bar{P}_1$*) satisfies* $\|\nabla \bar{P}_1\|_{L^2(\Omega)} \leqslant \|\nabla \Psi\|_{L^2(\Omega)}$*. Moreover, if* $\Psi \in W^{1,q}(\Omega)$ *with* $q > d$*, then* $\|\bar{P}_1\|_{L^\infty(\Omega)} \leqslant C < +\infty$ *with* C *depending only on* $\|\nabla \Psi\|_{L^q(\Omega)}$ *and* Ω*.*

As already understood in [CMS16, Section 5], there are situations where the pressure is concentrated on $\{1\} \times \Omega$: one cannot expect $\bar{P}$ to have a density with respect to the Lebesgue measure on the closed interval $[0,1]$. Nevertheless we prove in our theorem that the part of the pressure concentrated on $\{1\} \times \Omega$ has spatial regularity, namely $H^1(\Omega)$ and even $L^\infty(\Omega)$ if $\Psi \in W^{1,q}(\Omega)$ with $q > d$. The rest of this chapter is devoted to the proof of this theorem, the wrapping of the arguments being located at page 99.

5.1.2 The discrete problem

To tackle this problem and make rigorous the estimate presented in Chapter 3, we will approximate the primal problem in the following way:

- We introduce a time-discretization. The integer $N+1$ denotes the number of time steps, $\tau = 1/N$ will denote the time step.

- We add an infinitesimal entropic penalization. The goal is to make sure that the density of the minimizers of the discrete problem will be bounded from below, which is necessary when we want to write the optimality conditions.

- For technical reasons, we need to regularize V and Ψ. We take $(V_N)_{N\in\mathbb{N}}$ a sequence which converges to V in $H^1(\Omega)$ and such that V_N is Lipschitz for any $N \geqslant 1$. We can assume moreover that $\|\nabla V_N\|_{L^2(\Omega)} \leqslant \|\nabla V\|_{L^2(\Omega)}$ and V_N is positive. Similarly, we take a sequence Ψ_N going to Ψ in $H^1(\Omega)$ satisfying analogous properties.

The entropic penalization will be realized with the help of the Boltzmann entropy H whose definition is recalled in (2.13). As recalled in Proposition 2.12, the functional H is lower semi-continuous on $\mathcal{P}(\Omega)$. Moreover, a simple application of Jensen's inequality yields

$$-\ln(|\Omega|) \leqslant H(\rho) \leqslant 0$$

as soon as $\rho \leqslant 1$.

To define the discrete problem, we take $N \geqslant 1$ and denote by

$$\Gamma_0^N := \{(\rho_{k\tau})_{k\in\{0,1,\dots,N\}} \ : \ \rho_{k\tau} \in \mathcal{P}(\Omega) \text{ and } \rho_0 = \bar{\rho}_0\} \subset (\mathcal{P}(\Omega))^{N+1}$$

the set of discrete curves starting from $\bar{\rho}_0$. We denote by $\tau := 1/N$ the time step. We choose $(\lambda_N)_{N\in\mathbb{N}}$ which goes to 0 while being strictly positive, it will account for the scale of the entropic penalization. The speed at which $\lambda_N \to 0$ is irrelevant for the analysis, hence we do not need to specify it. The discrete functional $\mathcal{A}^N$ is defined on Γ_0^N as

$$\mathcal{A}^N(\rho) := \sum_{k=0}^{N-1} \frac{W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau} + \sum_{k=1}^{N-1} \tau \left(\int_\Omega V_N \rho_{k\tau} + \lambda_N H(\rho_{k\tau}) \right) + \int_\Omega \Psi_N \rho_1 + \lambda_N H(\rho_1)$$

if $\rho_{k\tau} \leqslant 1$ for all $k \in \{0, 1, \dots, N\}$ and $+\infty$ otherwise. The discrete problem reads as

$$\min_\rho \left\{ \mathcal{A}^N(\rho) \ : \ \rho \in \Gamma_0^N \right\}. \tag{5.5}$$

Proposition 5.6. *For any $N \geqslant 1$, there exists a unique solution to the discrete problem.*

Proof. The functional $\mathcal{A}^N$ is l.s.c. on Γ_0^N. Moreover, the curve ρ which is constant and equal to $\bar{\rho}_0$ belongs to Γ_0^N and is such that $\mathcal{A}^N(\rho) < +\infty$. As Γ_0^N is compact (for the topology of the weak convergence of measures), the direct method of calculus of variations ensures the existence of a minimizer.

Uniqueness clearly holds as $\lambda_N > 0$ and the entropy is a strictly convex function on $\mathcal{P}(\Omega)$. □

From now on, for any $N \geqslant 1$, we fix $\bar{\rho}^N$ the unique solution of the discrete problem

5.2 Estimates on the discrete problem

Let us comment on a technical refinement: for some computations to be valid, we will need to assume that $\bar{\rho}_0$ is smooth is strictly positive. If it is not the case, it is easy to approximate (for fixed N) the measure $\bar{\rho}_0$ with a sequence $\bar{\rho}_0^{(n)}$ of smooth densities. For such a $\bar{\rho}_0^{(n)}$, the estimates obtained below for a given N (Corollary 5.12) do not depend on n. Hence it is easy to send n to $+\infty$, using the stability of the Kantorovich potentials [San15, Theorem 1.52] to see that these estimates are still satisfied for the solution of the discrete problem with initial condition $\bar{\rho}_0$. In short: we will do as if our initial condition $\bar{\rho}_0$ were smooth, and as long as the final estimates do not depend on the smoothness of $\bar{\rho}_0$ this will be legitimate.

5.2.1 Interior regularity

We begin with the interior regularity. In this subsection, we fix $N \geqslant 1$ and $k \in \{1, 2, \dots, N-1\}$ a given instant in time. We use the shortcut $\bar{\rho} := \bar{\rho}_{k\tau}^N$ and we also denote $\mu := \bar{\rho}_{(k-1)\tau}^N$ and $\nu := \bar{\rho}_{(k+1)\tau}^N$. As $\bar{\rho}^N$ is a solution of the discrete problem, we know that $\bar{\rho}$ is a minimizer, among all probability measures with density bounded by 1, of

$$\rho \mapsto \frac{W_2^2(\mu, \rho) + W_2^2(\rho, \nu)}{2\tau} + \tau \left(\int_\Omega V_N \rho + \lambda_N H(\rho) \right).$$

Lemma 5.7. *The density $\bar{\rho}$ is strictly positive a.e.*

Proof. This is exactly the same proof as Lemma 4.8, as the construction done in this proof preserves the constraint of having a density smaller than 1. □

Proposition 5.8. *Let us denote by φ_μ and φ_ν the Kantorovich potentials for the transport from $\bar{\rho}$ to μ and ν respectively. There exists $p \in L^1(\Omega)$, positive, such that $\{p > 0\} \subset \{\bar{\rho} = 1\}$ and a constant C such that*

$$\frac{\varphi_\mu + \varphi_\nu}{\tau^2} + V_N + p + \lambda_N \ln(\bar{\rho}) = C \quad a.e. \tag{5.6}$$

Moreover p and $\ln(\bar{\rho})$ are Lipschitz and $\nabla p \cdot \nabla \ln(\bar{\rho}) = 0$ a.e.

Proof. Let $\tilde\rho \in \mathcal{P}(\Omega)$ such that $\tilde\rho \leqslant 1$. We define $\rho_\varepsilon := (1-\varepsilon)\bar\rho + \varepsilon\tilde\rho$ and use it as a competitor. Clearly $\rho_\varepsilon \leqslant 1$, i.e. it is an admissible competitor. Comparing $\mathcal{A}^N(\rho_\varepsilon)$ to $\mathcal{A}^N(\rho)$, we extract the following information. Using Proposition 2.3, as $\bar\rho > 0$, the Kantorovich potentials φ_μ and φ_ν are unique (up to a constant) and

$$\lim_{\varepsilon\to 0} \frac{W_2^2(\mu,\rho_\varepsilon) - W_2^2(\mu,\bar\rho) + W_2^2(\rho_\varepsilon,\nu) - W_2^2(\bar\rho,\nu)}{2\tau^2} = \int_\Omega \frac{\varphi_\mu + \varphi_\nu}{\tau}(\tilde\rho - \bar\rho).$$

The term involving V_N is straightforward to handle as it is linear. The only remaining term is the one involving the entropy. But here, using the same reasoning as in Proposition 4.9, we can say that

$$\limsup_{\varepsilon\to 0} \frac{H(\rho_\varepsilon) - H(\bar\rho)}{\varepsilon} \leqslant \int_\Omega \ln(\bar\rho)(\tilde\rho - \bar\rho).$$

Putting the pieces together, we see that $\int_\Omega h\,(\tilde\rho - \bar\rho) \geqslant 0$ for any $\tilde\rho \in \mathcal{P}(\Omega)$ with $\tilde\rho \leqslant 1$, provided that h is defined by

$$h := \frac{\varphi_\mu + \varphi_\nu}{\tau^2} + V_N + \lambda_N \ln(\bar\rho)$$

It is known, analogously to [MRCS10, Lemma 3.3], that this leads to the existence of a constant C such that

$$\begin{cases} \bar\rho = 1 & \text{on } \{h < C\} \\ \bar\rho \leqslant 1 & \text{on } \{h = C\} \\ \bar\rho = 0 & \text{on } \{h > C\} \end{cases} \tag{5.7}$$

Not that the case $\{h > C\}$ can be excluded by Lemma 5.7. The pressure p is defined as $p = (C-h)_+$, thus (5.6) holds. It satisfies $p \geqslant 0$ and $\bar\rho < 1$ implies $p = 0$.

It remains to answer the question of the integrability properties of p and $\ln(\bar\rho)$. Notice that p is positive, and non zero only on $\{\bar\rho = 1\}$. On the other hand, $\ln(\bar\rho) \leqslant 0$ and it is non zero only on $\{\bar\rho < 1\}$. Hence, one can write

$$p = \left(C - \frac{\varphi_\mu + \varphi_\nu}{\tau^2} + V_N\right)_+ \quad \text{and} \quad \ln(\bar\rho) = -\frac{1}{\lambda_N}\left(C - \frac{\varphi_\mu + \varphi_\nu}{\tau^2} + V_N\right)_- . \tag{5.8}$$

Given that the Kantorovich potentials and V_N are Lipschitz, it implies the Lipschitz regularity for p and $\ln(\bar\rho)$. Moreover, the identity $\nabla p \cdot \nabla \ln(\bar\rho) = 0$ is straightforward using $\nabla f_+ = \nabla f \mathbb{1}_{f>0}$ a.e., which is valid for any $f \in H^1(\Omega)$. □

Let us note that φ_μ and φ_ν have additional regularity properties, even though they depend heavily on N.

Lemma 5.9. *The Kantorovich potentials φ_μ and φ_ν belong to $C^{2,\alpha}(\mathring{\Omega}) \cap C^{1,\alpha}(\Omega)$.*

Proof. If $k \in \{1, \dots, N-1\}$, thanks to Proposition 5.8 (applied in $k-1$ and $k+1$), we know that μ and ν have a Lipschitz density and are bounded from below. Using the regularity theory for the Monge Ampère-equation [Vil03, Theorem 4.14], we can conclude that φ_μ and φ_ν belong to $C^{2,\alpha}(\mathring{\Omega}) \cap C^{1,\alpha}(\Omega)$. □

Theorem 5.10. *For any $m \geqslant 1$, the following inequality holds:*

$$\int_\Omega \nabla(p^m) \cdot \nabla(p + V_N) \leqslant 0. \tag{5.9}$$

Proof. The (optimal) transport map from $\bar{\rho}$ to μ is given by $\mathrm{Id} - \nabla\varphi_\mu$, and similarly for ν. We consider the following quantity, (defined on the whole Ω given the regularity of μ, ν, φ_μ and φ_ν), which is a discrete analogue of the l.h.s. of (3.9):

$$D(x) := -\frac{\ln(\mu(x - \nabla\varphi_\mu(x))) + \ln(\nu(x - \nabla\varphi_\nu(x))) - 2\ln(\bar{\rho}(x))}{\tau^2}.$$

Notice that if $\bar{\rho}(x) = 1$, then by the constraint $\mu(x - \nabla\varphi_\mu(x)) \leqslant 1$ and $\nu(x - \nabla\varphi_\nu(x)) \leqslant 1$ the quantity $D(x)$ is positive. On the other hand, using $(\mathrm{Id} - \nabla\varphi_\mu)\#\bar{\rho} = \mu$ and the Monge-Ampère equation, for all $x \in \mathring{\Omega}$ there holds

$$\mu(x - \nabla\varphi_\mu(x)) = \frac{\bar{\rho}(x)}{\det(\mathrm{Id} - D^2\varphi_\mu(x))},$$

and a similar identity holds for φ_ν. Hence the quantity $D(x)$ is equal, for all $x \in \mathring{\Omega}$, to

$$D(x) = \frac{\ln(\det(\mathrm{Id} - D^2\varphi_\mu(x))) + \ln(\det(\mathrm{Id} - D^2\varphi_\nu(x)))}{\tau^2}.$$

Diagonalizing the matrices $D^2\varphi_\mu, D^2\varphi_\nu$ and using the convexity inequality $\ln(1-y) \leqslant -y$, we end up with

$$D(x) \leqslant -\frac{\Delta(\varphi_\mu(x) + \varphi_\nu(x))}{\tau^2}.$$

We multiply this identity by p^m and integrate. Thanks to the fact that D is positive on $\{\bar{\rho} = 1\}$, as p is positive and does not vanish only on $\{\bar{\rho} = 1\}$, the quantity $p^m D$ is positive on $\mathring{\Omega}$. As the latter coincides, up to a Lebesgue negligible set, with Ω, we get

$$\int_\Omega p^m \frac{\Delta(\varphi_\mu + \varphi_\nu)}{\tau^2} \leqslant 0. \tag{5.10}$$

We do an integration by parts, which reads

$$\int_\Omega p^m \frac{\Delta(\varphi_\mu + \varphi_\nu)}{\tau^2} = \int_{\partial\Omega} p^m \frac{\nabla(\varphi_\mu + \varphi_\nu)}{\tau^2} \cdot \mathbf{n}_\Omega - \int_\Omega \nabla(p^m) \cdot \frac{\nabla(\varphi_\mu + \varphi_\nu)}{\tau^2} \tag{5.11}$$

To handle the boundary term, recall that $\nabla\varphi_\mu$ is continuous up to the boundary and that $x - \nabla\varphi_\mu(x) \in \Omega$ for every $x \in \Omega$ as $(\mathrm{Id} - \nabla\varphi_\mu)\#\bar{\rho} = \mu$. Given the convexity of Ω, it implies $\nabla\varphi_\mu(x) \cdot \mathbf{n}_\Omega(x) \geqslant 0$ for every point $x \in \partial\Omega$ for which the outward normal $\mathbf{n}_\Omega(x)$ is defined. As it is the case for a.e. point of the boundary, as a similar inequality holds for φ_ν, and given that p^m is positive, we can drop the boundary term in (5.11) and get

$$\int_\Omega p^m \frac{\Delta(\varphi_\mu + \varphi_\nu)}{\tau^2} \geqslant -\int_\Omega \nabla(p^m) \cdot \frac{\nabla(\varphi_\mu + \varphi_\nu)}{\tau^2}.$$

Using the optimality conditions (5.6), we see that

$$0 \geqslant \int_\Omega p^m \frac{\Delta(\varphi_\mu + \varphi_\nu)}{\tau^2} \geqslant \int_\Omega \nabla(p^m) \cdot \nabla(p + V_N + \lambda_N \ln(\bar{\rho}))$$

Now remember that in Proposition 5.8 we have proved that $\nabla p \cdot \nabla \ln(\bar{\rho}) = 0$ a.e., which is sufficient to drop the term involving $\nabla \ln(\bar{\rho})$ and get (5.9). □

The inequation (5.9) implies the $H^1(\Omega)$ and $L^\infty(\Omega)$ regularity for the pressure: this can be seen as a consequence of Moser's regularity for elliptic equations [Mos60]. We still give the proof for the sake of completeness, and also because in the inequality (5.9), the boundary terms have already been taken in account, which enables to get regularity up to the spatial boundary in a single set of iterations.

Lemma 5.11. *Let f, W be Lipschitz functions defined on Ω such that f vanishes on a set of measure at least $|\Omega| - 1 > 0$ and such that, for any $m \geqslant 1$,*

$$\int_\Omega \nabla(f^m) \cdot \nabla(f + W) \leqslant 0.$$

Then there holds $\|\nabla f\|_{L^2(\Omega)} \leqslant \|\nabla W\|_{L^2(\Omega)}$. Moreover, if $\nabla W \in L^q(\Omega)$ with $q > d$, then $f \in L^\infty(\Omega)$ and $\|f\|_{L^\infty(\Omega)}$ is bounded by a constant which depends only on Ω and $\|\nabla W\|_{L^q(\Omega)}$.

Proof. With $m = 1$ we immediately get

$$\|\nabla f\|_{L^2(\Omega)} \leqslant \|\nabla W\|_{L^2(\Omega)},$$

In particular, using the Poincaré inequality and the fact that $|\{f = 0\}| \geqslant |\Omega| - 1$, we see that $\|f\|_{L^1(\Omega)}$ is bounded by a constant depending only on Ω and V.

In the rest of the proof, we denote by C a constant which depends only on Ω and $\|\nabla W\|_{L^q(\Omega)}$, and can change from line to line. We write the estimate, for any $m \geqslant 1$, as

$$\int_\Omega |\nabla f|^2 f^{m-1} \leqslant -\int_\Omega (\nabla f \cdot \nabla W) f^{m-1}.$$

Using Young's inequality, it is clear that

$$\frac{2}{(m+1)^2}\int_\Omega \left|\nabla\left(f^{(m+1)/2}\right)\right|^2 = \frac{1}{2}\int_\Omega |\nabla f|^2 f^{m-1} \leqslant \frac{1}{2}\int_\Omega |\nabla W|^2 f^{m-1}.$$

Take $\tilde\beta < \beta < \frac{d}{d-2}$ sufficiently close to $\frac{d}{d-2}$ in such a way that $2\tilde\beta/(\tilde\beta - 1) \leqslant q$. In particular, the $L^{2\tilde\beta/(\tilde\beta-1)}(\Omega)$ norm of ∇W is bounded by $C\|\nabla W\|_{L^q(\Omega)}$. Moreover, we know that $H^1(\Omega) \hookrightarrow L^{2\beta}(\Omega)$. Considering the fact that $f^{(m+1)/2}$ vanishes on a subset of measure at least $|\Omega| - 1$, it enables us to write [Mos60, Lemma 2]

$$\begin{aligned}\left(\int_\Omega f^{(m+1)\beta}\right)^{1/\beta} &\leqslant C\int_\Omega \left|\nabla\left(f^{(m+1)/2}\right)\right|^2 \\ &\leqslant C(m+1)^2\int_\Omega |\nabla W|^2 f^{m-1} \\ &\leqslant C(m+1)^2\left(\int_\Omega |\nabla W|^{2\tilde\beta/(\tilde\beta-1)}\right)^{(\tilde\beta-1)/\tilde\beta}\left(\int_\Omega f^{(m-1)\tilde\beta}\right)^{1/\tilde\beta},\end{aligned}$$

where the last inequality is Hölder's inequality with an exponent $\tilde\beta$. Thanks to this choice, taking the power $1/(m+1)$ on both sides,

$$\|f\|_{L^{(m+1)\beta}(\Omega)} \leqslant \left[C(m+1)^2\right]^{1/(m+1)} \|f\|_{L^{(m-1)\tilde\beta}}^{(m-1)/(m+1)}.$$

It is easy to iterate this inequation. With $r = (m-1)\tilde\beta$, as $(m+1)\beta \geqslant \beta r/\tilde\beta$, one can write that

$$\|f\|_{L^{\beta/\tilde\beta r}(\Omega)} \leqslant [C(r+1)]^{C/r}\max\left(\|f\|_{L^r(\Omega)}, 1\right).$$

An easy induction (recall that we already know that f is bounded in $L^1(\Omega)$ by a constant depending only on Ω and W) with $r_n = \left(\beta/\tilde\beta\right)^n$ shows that $\|f\|_{L^{r_n}(\Omega)}$ is bounded by a constant which depends only on $\|\nabla W\|_{L^q(\Omega)}$ and Ω, which implies the claimed $L^\infty(\Omega)$ bound. ☐

Corollary 5.12. *There holds* $\|\nabla p\|_{L^2(\Omega)} \leqslant \|\nabla V\|_{L^2(\Omega)}$. *Moreover, if* $V \in W^{1,q}(\Omega)$ *with* $q > d$, *then* $p \in L^\infty(\Omega)$ *and* $\|p\|_{L^\infty(\Omega)}$ *is bounded by a constant which depends only on* Ω *and* $\|\nabla V\|_{L^q(\Omega)}$.

Proof. It is enough to combine Lemma 5.11 and (5.9): one has to remember that p vanishes where $\bar{\rho} = 1$, which is of measure at least $|\Omega| - 1$, that $\|\nabla V_N\|_{L^2(\Omega)} \leqslant \|\nabla V\|_{L^2(\Omega)}$, and that $\|\nabla V_N\|_{L^q(\Omega)}$ is bounded independently on N if $V \in W^{1,q}(\Omega)$. □

5.2.2 Boundary regularity

As we said, we will see that the pressure has a part which is concentrated on the temporal boundary $t = 1$. The regularity of this part is proved exactly by the same technique than in the interior, hence we will only sketch it. In this subsection, we fix $N \geqslant 1$. We use the shortcut $\bar{\rho} := \bar{\rho}^N_{N\tau} = \bar{\rho}^N_1$ for the final measure and we also denote $\mu := \bar{\rho}^N_{(N-1)\tau}$. As $\bar{\rho}^N$ is a solution of the discrete problem, we know that $\bar{\rho}$ is a minimizer, among all probability measures with density bounded by 1, of

$$\rho \mapsto \frac{W_2^2(\mu, \rho)}{2\tau} + \left(\int_\Omega \Psi_N \rho + \lambda_N H(\rho) \right).$$

Lemma 5.13. *The density* $\bar{\rho}$ *is strictly positive a.e.*

Proof. This property holds for exactly the same reason as in Lemma 5.7. □

Proposition 5.14. *Let us denote by* φ_μ *the Kantorovich potential for the transport from* $\bar{\rho}$ *to* μ. *There exists* $p \in L^1(\Omega)$, *positive, such that* $\{p > 0\} \subset \{\bar{\rho} = 1\}$ *and a constant* C *such that*

$$\frac{\varphi_\mu}{\tau} + \Psi_N + p + \lambda_N \ln(\bar{\rho}) = C. \tag{5.12}$$

Moreover p *and* $\ln(\bar{\rho})$ *are Lipschitz and* $\nabla p \cdot \nabla \ln(\bar{\rho}) = 0$ *a.e.*

Proof. We use exactly the same competitor as in the proof of Proposition 5.8. It leads to the conclusion that $\int_\Omega h(\tilde{\rho} - \bar{\rho}) \geqslant 0$ for any $\tilde{\rho} \in \mathcal{P}(\Omega)$ with $\tilde{\rho} \leqslant 1$ where h is defined as

$$h := \frac{\varphi_\mu}{\tau} + \Psi_N + \lambda_N \ln(\bar{\rho}).$$

It implies the existence of a constant C such that (5.7) holds, and we define p exactly in the same way, as $p := (C - h)_+$. The integrability properties of p and $\ln(\bar{\rho})$ are derived in the same way as in the proof of Proposition 5.8. □

The additional regularity for φ_μ is exactly the same than for the interior case (this is why we have also used an entropic penalization at the boundary).

Lemma 5.15. *The Kantorovich potential* φ_μ *belongs to* $C^{2,\alpha}(\mathring{\Omega}) \cap C^{1,\alpha}(\Omega)$.

Theorem 5.16. *For any* $m \geqslant 1$, *the following inequality holds:*

$$\int_\Omega \nabla(p^m) \cdot \nabla(p + \Psi_N) \leqslant 0. \tag{5.13}$$

Proof. On the set $\mathring{\Omega}$ we consider the following quantity, which is the analogue of the l.h.s. of (3.11):

$$D(x) := \frac{\ln(\bar{\rho}(x)) - \ln(\mu(x - \nabla\varphi_\mu(x)))}{\tau}.$$

If $\bar{\rho}(x) = 1$, then by the constraint $\mu(x - \nabla\varphi_\mu(x)) \leqslant 1$ the quantity $D(x)$ is positive. On the other hand, exactly by the same estimate than in the proof of Theorem 5.10,

$$D(x) \leqslant -\frac{(\Delta\varphi_\mu)(x)}{\tau}.$$

We multiply this inequality by p^m, do an integration by parts (the boundary term is handled exactly as in Theorem 5.10), and we end up with (5.13). □

Corollary 5.17. *There holds $\|\nabla p\|_{L^2(\Omega)} \leqslant \|\nabla\Psi\|_{L^2(\Omega)}$. Moreover, if $\Psi \in W^{1,q}(\Omega)$ with $q > d$, then $p \in L^\infty(\Omega)$ and $\|p\|_{L^\infty(\Omega)}$ is bounded by a constant which depends only on Ω and $\|\nabla\Psi\|_{L^q(\Omega)}$.*

Proof. Exactly as in the proof of Corollary 5.12, it is enough to combine Lemma 5.11 and the estimate (5.13). □

5.3 Convergence to the continuous problem

Recall that for any $N \geqslant 1$, $\bar{\rho}^N \in \Gamma_0^N$ denotes the solution of the discrete problem.

5.3.1 Convergence of the primal problem

This convergence is very similar to the one performed in Chapter 4 hence we will not really reproduce it. Furthermore, as we are ultimately interested in the dual problem, we need only the convergence of the value of the primal problem, not of the minimizers.

Define $\tilde{\mathcal{A}}^N$ on Γ_0^N exactly as the discrete primal functional $\mathcal{A}^N$, but where the regularized potentials V_N and Ψ_N are replaced by the true potentials V and Ψ. Given the L^∞ bound on ρ (which holds if $\mathcal{A}^N$ or $\tilde{\mathcal{A}}^N$ are finite), one can see that for any $\rho \in \Gamma_0^N$ with density bounded by 1,

$$\left|\mathcal{A}^N(\rho) - \tilde{\mathcal{A}}^N(\rho)\right| \leqslant \|V - V_N\|_{L^1(\Omega)} + \|\Psi - \Psi_N\|_{L^1(\Omega)}, \tag{5.14}$$

and the r.h.s. goes to 0 uniformly in ρ as $N \to +\infty$.

On the other hand, using exactly the same proofs as in Chapter 4, Section 4.4, one can easily check (the only thing to check is that all the constructions are compatible with the constraint of having a density bounded by 1 but it is straightforward) that the value of the discrete problem

$$\min_\rho \left\{\tilde{\mathcal{A}}^N(\rho) \ : \ \rho \in \Gamma_0^N\right\}$$

converges to the minimal value of the primal problem (notice that it is for this result that we need the scale λ_N of the entropic penalization to go to 0). Combined with (5.14), one can conclude the following.

Proposition 5.18. *The value of the discrete problem converges to the continuous one in the sense that*

$$\lim_{N\to+\infty} \mathcal{A}^N(\bar{\rho}^N) = \min_\rho \left\{\mathcal{A}(\rho) \ : \ \rho \in \Gamma_0\right\}.$$

5.3.2 Convergence to the dual problem

In this subsection, we want to build a value function ϕ^N which will go, as $N \to +\infty$, to a solution of the (continuous) dual problem. Notice that the discrete functional $\mathcal{A}^N$ is convex, hence we could consider discrete dual problem but we will not do it explicitly: indeed, the approximate value function ϕ^N will not be a solution of the discrete dual problem and we will not prove a duality result at the discrete level.

On the contrary, we will just guess the expression of ϕ^N (we have to say to the inspiration for this kind of construction was found in the work of Loeper [Loe06]) and use the explicit expression to prove that the value of some quantity which looks like the continuous dual objective, evaluated at ϕ^N, is close to the value of the discrete primal problem. Then, sending N to $+\infty$, we recover an admissible $(\bar{\phi}, \bar{P})$ for the continuous dual problem such that $\mathcal{B}(\bar{\phi}, \bar{P})$ is larger than the optimal value of the continuous primal problem (and this comes from estimates proved at the discrete level). It will allow us to conclude that $(\bar{\phi}, \bar{P})$ is a solution of the dual problem thanks to the absence of duality gap at the continuous level. Eventually, we pass to the limit the discrete estimates in Corollary 5.12 and Corollary 5.17 to get the ones for $\bar{p}$ and $\bar{P}_1$.

Let us recall that $\bar{\rho}^N$ is the solution of the discrete problem. For any $k \in \{0, 1, \dots, N-1\}$, we choose $(\varphi^N_{k\tau}, \psi^N_{k\tau})$ a pair of Kantorovich potential between $\bar{\rho}^N_{k\tau}$ and $\bar{\rho}^N_{(k+1)\tau}$, such choice is unique up to an additive constant. According to Proposition 5.8 and Proposition 5.14, making the dependence on N and k explicit, for any $k \in \{1, 2, \dots, N\}$, there exists a pressure $p^N_{k\tau}$ positive and Lipschitz, and a constant $C^N_{k\tau}$ such that

$$\begin{cases} \dfrac{\psi^N_{(k-1)\tau} + \varphi^N_{k\tau}}{\tau^2} + V_N + p^N_{k\tau} + \lambda_N \ln(\bar{\rho}^N_{k\tau}) = C^N_{k\tau} & k \in \{1, 2, \dots, N-1\}, \\ \dfrac{\psi^N_{(k-1)\tau}}{\tau} + \Psi_N + p^N_1 + \lambda_N \ln(\bar{\rho}^N_{k\tau}) = C^N_1 & k = N. \end{cases} \tag{5.15}$$

We define the following value function, defined on the whole interval $[0, 1]$ which can be thought as a function which looks like a solution of what could be called a discrete dual problem.

Definition 5.19. *Let $\phi^N \in \mathrm{BV}([0,1] \times \Omega) \cap L^2([0,1] \times H^1(\Omega))$ the function defined as follows. The "final" value is given by*

$$\phi^N(1^-, \cdot) := \Psi_N + p^N_1. \tag{5.16}$$

Provided that the value $\phi^N((k\tau)^-, \cdot)$ is defined for some $k \in \{1, 2, \dots, N\}$, the value of ϕ^N on $((k-1)\tau, k\tau) \times \Omega$ is defined by

$$\phi^N(t, x) := \inf_{y \in \Omega} \left(\frac{|x-y|^2}{2(k\tau - t)} + \phi^N((k\tau)^-, y) \right). \tag{5.17}$$

If $k \in \{1, 2, \dots, N-1\}$, the function ϕ^N has a temporal jump at $t = k\tau$ defined by

$$\phi^N((k\tau)^-, x) := \phi^N((k\tau)^+, x) + \tau \left(V_N + p^N_{k\tau} \right)(x) \tag{5.18}$$

Notice that we have not included the entropic term: its only effect would have been to decrease ϕ^N (which in the end decreases the value of the dual functional) and it would have prevented us from getting compactness on the sequence ϕ^N. The link between this value function and the Kantorovich potentials is the following.

Lemma 5.20. *For any $k \in \{1, 2, \dots, N\}$, one has*

$$\phi^N((k\tau)^-, \cdot) \geqslant C^N_1 + \tau \sum_{j=k}^{N-1} C^N_{j\tau} - \frac{\psi^N_{(k-1)\tau}}{\tau}. \tag{5.19}$$

For any $k \in \{0, 1, \dots, N-1\}$, one has

$$\phi^N((k\tau)^+, \cdot) \geqslant C^N_1 + \tau \sum_{j=k+1}^{N-1} C^N_{j\tau} + \frac{\varphi^N_{k\tau}}{\tau}. \tag{5.20}$$

Proof. We will prove it by (decreasing) induction on $k \in \{0, 1, \ldots, N\}$. For $k = N$, by the optimality conditions (5.15) and the fact that $\ln(\bar{\rho}_1^N) \leqslant 0$, it is clear that (5.19) holds.

Now assume that (5.19) holds for some k. Using (5.17), one has

$$\begin{aligned}
\phi^N(((k-1)\tau)^+, x) &= \inf_{y \in \Omega} \left(\frac{|x-y|^2}{2\tau} + \phi^N((k\tau)^-, y) \right) \\
&\geqslant C_1^N + \tau \sum_{j=k}^{N-1} C_{j\tau}^N + \inf_{y \in \Omega} \left(\frac{|x-y|^2}{2\tau} - \frac{\psi_{(k-1)\tau}^N(y)}{\tau} \right) \\
&= C_1^N + \tau \sum_{j=k}^{N-1} C_{j\tau}^N + \frac{1}{\tau} \inf_{y \in \Omega} \left(\frac{|x-y|^2}{2} - \psi_{(k-1)\tau}^N(y) \right) \\
&= C_1^N + \tau \sum_{j=k}^{N-1} C_{j\tau}^N + \frac{\varphi_{(k-1)\tau}^N(x)}{\tau},
\end{aligned}$$

where the last inequality comes from the fact that $\varphi_{(k-1)\tau}^N$ is the c-transform of $\psi_{(k-1)\tau}^N$. This gives us (5.20) for $k-1$. On the other hand, assume that (5.20) holds for some k. Using (5.18) and the optimality conditions (5.15) ,

$$\begin{aligned}
\phi^N((k\tau)^-, x) &= \phi^N((k\tau)^+, x) + \tau \left(V_N + p_{k\tau}^N \right) \\
&\geqslant C_1^N + \tau \sum_{j=k+1}^{N-1} C_{j\tau}^N + \frac{\varphi_{k\tau}^N}{\tau} + \tau \left(V_N + p_{k\tau}^N \right) \\
&= C_1^N + \tau \sum_{j=k+1}^{N-1} C_{j\tau}^N + C_{k\tau}^N - \frac{\psi_{k\tau}^N}{\tau} - \lambda_N \tau \ln(\bar{\rho}_{k\tau}^N) \geqslant C_1^N + \tau \sum_{j=k}^{N-1} C_{j\tau}^N - \frac{\psi_{k\tau}^N}{\tau},
\end{aligned}$$

which means that (5.19) holds for k. □

From this identity, we can express some kind of duality result at the discrete level, which reads as follows.

Proposition 5.21. *For $N \geqslant 1$, the following inequality holds:*

$$\mathcal{A}^N(\bar{\rho}^N) \leqslant \int_\Omega \phi^N(0^+, \cdot)\bar{\rho}_0 - \tau \sum_{k=1}^{N-1} \int_\Omega p_{k\tau}^N - \int_\Omega p_1^N. \tag{5.21}$$

We have an inequality and not an equality because we have not included the entropic terms in the value function.

Proof. The idea is to evaluate $\mathcal{A}^N(\bar{\rho}^N)$ by expressing the Wasserstein distances with the help of the Kantorovich potentials.

$$\begin{aligned}
\mathcal{A}^N(\bar\rho^N) &= \sum_{k=0}^{N-1} \frac{W_2^2(\bar\rho^N_{k\tau}, \bar\rho^N_{(k+1)\tau})}{\tau} + \sum_{k=1}^{N-1} \tau \left(\int_\Omega V_N \bar\rho^N_{k\tau} + \lambda_N H(\bar\rho^N_{k\tau}) \right) + \int_\Omega \Psi_N \bar\rho^N_1 + \lambda_N H(\bar\rho^N_1) \\
&= \frac{1}{\tau} \sum_{k=0}^{N-1} \left(\int_\Omega \varphi^N_{k\tau} \bar\rho^N_{k\tau} + \int_\Omega \psi^N_{k\tau} \bar\rho^N_{(k+1)\tau} \right) + \sum_{k=1}^{N-1} \tau \left(\int_\Omega V_N \bar\rho^N_{k\tau} + \lambda_N H(\bar\rho^N_{k\tau}) \right) + \int_\Omega \Psi_N \bar\rho^N_1 \\
&\qquad + \lambda_N H(\bar\rho^N_1) \\
&= \frac{1}{\tau} \int_\Omega \varphi^N_0 \bar\rho_0 + \sum_{k=1}^{N-1} \int_\Omega \left(\frac{\varphi^N_{k\tau} + \psi^N_{(k-1)\tau}}{2\tau} + \tau (V_N + \lambda_N \ln(\bar\rho^N_{k\tau})) \right) \bar\rho^N_{k\tau} \\
&\qquad + \int_\Omega \left(\frac{\psi^N_{(N-1)\tau}}{2\tau} + \Psi_N + \lambda_N \ln(\bar\rho^N_1) \right) \bar\rho^N_1,
\end{aligned}$$

where the last equality comes from a reindexing of the sums. Now we use the optimality conditions (5.15) to handle the second and third term. Notice that, as $p^N_{k\tau}$ lives only where $\bar\rho^N_{k\tau} = 1$, that we can replace $\bar\rho^N_{k\tau}$ by 1 when it is multiplied by the pressure. Recall also that the probability distributions, when integrated against a constant, are equal to this constant. We are left with

$$\begin{aligned}
\mathcal{A}^N(\bar\rho^N) &= \frac{1}{\tau} \int_\Omega \varphi^N_0 \bar\rho_0 + \sum_{k=1}^{N-1} \left(C^N_{k\tau} - \tau \int_\Omega p^N_{k\tau} \right) + \left(C^N_1 - \int_\Omega p^N_1 \right) \\
&\leqslant \int_\Omega \phi^N(0^+, \cdot) \bar\rho_0 - \tau \sum_{k=1}^{N-1} \int_\Omega p^N_{k\tau} - \int_\Omega p^N_1,
\end{aligned}$$

where the last equality comes from Lemma 5.20 which allows to make the link between the Kantorovich potential φ^N_0 and $\phi^N(0^+, \cdot)$. □

We want to pass to the limit $N \to +\infty$. To this extent, we rely on the fact that ϕ^N satisfies an explicit equation in the sense of distributions. We start to define the distribution which will be the r.h.s. of the Hamilton Jacobi equation.

Definition 5.22. *Let α^N and P^N the positive measures on $[0,1] \times \Omega$ defined as*

$$\begin{cases}
\alpha^N & := \tau \sum_{k=1}^{N-1} \delta_{t=k\tau} (p^N_{k\tau} + V_N) + \delta_{t=1} p^N_1, \\
P^N & := \tau \sum_{k=1}^{N-1} \delta_{t=k\tau} p^N_{k\tau} + \delta_{t=1} p^N_1.
\end{cases}$$

More precisely, for any test function $a \in C([0,1] \times \Omega)$,

$$\iint_{[0,1]\times\Omega} a \,\mathrm{d}\alpha^N := \tau \sum_{k=1}^{N-1} \int_\Omega a(k\tau, \cdot) \left(V_N + p^N_{k\tau} \right) + \int_\Omega a(1, \cdot) p^N_1,$$

and similarly for P^N.

In other words, α^N is, from the temporal point of view, a sum of delta function, each of them corresponding to the jump of the value function ϕ^N.

Proposition 5.23. *Provided that we set $\phi^N(0^-,\cdot) = \phi^N(0^+,\cdot)$ and $\phi^N(1^+,\cdot) = \Psi_N$, the following equation holds in the sense of distributions on $[0,1]\times\Omega$:*

$$- \partial_t \phi^N + \frac{1}{2}|\nabla \phi^N|^2 \leqslant \alpha^N. \tag{5.22}$$

Proof. As the pressures and the potentials V_N, Ψ_N are Lipschitz, for any $t \in [0,1]$, the value function $\phi^N(t^+,\cdot)$ and $\phi^N(t^-,\cdot)$ are Lipschitz (but with a Lipschitz constant which may diverge as $N \to +\infty$).

Notice that on each interval $(((k-1)\tau)^+,(k\tau)^-)$, the function ϕ^N is defined by the Hopf-Lax formula, hence solves the Hamilton-Jacobi equation $-\partial_t \phi^N + \frac{1}{2}|\nabla\phi^N|^2 = 0$ a.e. [Eva10, Section 3.3]. It implies that the inequality $-\partial_t\phi^N + \frac{1}{2}|\nabla \phi^N|^2 \leqslant 0$ is also satisfied in the sense of distributions, as $\nabla\phi^N$ is bounded and $\partial_t \phi$ may have some singular parts, but they are positive.

Provided that we set $\phi^N(0^-,\cdot) = \phi^N(0^+,\cdot)$ and $\phi^N(1^+,\cdot) = \Psi_N$, the measure $\partial_t\phi^N$ has a singular negative part at $\{\tau, 2\tau, \dots, 1\}$ corresponding to the jumps of the function ϕ^N; but, given (5.16) and (5.18), the negative part of $\partial_t \phi^N$ is exactly $-\alpha^N$. □

The next step is to pass to the limit $N \to +\infty$. To this extent, we need uniform bounds on α^N, which derive easily from the bounds that we have on the pressure.

Lemma 5.24. *There exists a constant C, independent of N, such that $\alpha^N([0,1]\times\Omega) \leqslant C$ and $P^N([0,1]\times\Omega)\leqslant C$.*

Recall that both α^N and P^N are positive measures as we have chosen V_N in such a way that it is positive.

Proof. We know that the $p^N_{k\tau}$, for $k \in \{1,2,\dots,N\}$ have a gradient which is bounded uniformly in $L^2(\Omega)$. As moreover they all vanish on a set of measure at least $|\Omega| - 1$, they are bounded uniformly (w.r.t. N) in $L^1(\Omega)$. This is enough, in order to get the uniform bound on P^N. Given the way V_N is built, the one for α^N is a straightforward consequence of the one on P^N. □

Now that we have a bound on α^N, to get compactness on the sequence ϕ^N, we use the same kind of estimates used to prove existence of a solution in the dual at the continuous level, see for instance [CMS16, Section 3]. We recall that $\mathcal{K}$, the set of admissible competitors for the dual problem, was defined in Definition 5.2.

Proposition 5.25. *There exists $(\bar\phi, \bar P) \in \mathcal{K}$ admissible for the dual problem such that*

$$\begin{cases} \lim\limits_{N\to+\infty} \phi^N = \bar\phi & \textit{weakly in } \mathrm{BV}([0,1]\times\Omega)\cap L^2([0,1],H^1(\Omega)), \\ \lim\limits_{N\to+\infty} P^N = \bar P & \textit{in } \mathcal{M}_+([0,1]\times\Omega). \end{cases}$$

Proof. Given Lemma 5.24, we know that P^N is bounded in $\mathcal{M}_+([0,1]\times\Omega)$ independently of N. Up to the extraction of a subsequence, it converges weakly as a measure to some $\bar P$. On the other hand, once we know this convergence, it is easy to see that α^N converges as a measure on $\mathcal{M}_+([0,1]\times\Omega)$ to $\bar P + V$.

We have assumed that V and Ψ are positive, and so are V_N and Ψ_N, independently of N. Using the definition of ϕ^N and the positivity of the pressures, it is not hard to see that ϕ^N is positive $[0,1]\times\Omega$. Integrating the Hamilton Jacobi equation w.r.t. space and time and using the bound on α^N (Lemma 5.24), we see that

$$\int_\Omega \phi^N(0^+,\cdot) - \int_\Omega \Psi_N + \frac{1}{2}\iint\limits_{[0,1]\times\Omega} |\nabla\phi^N|^2 \leqslant C. \tag{5.23}$$

Combined with the positivity of $\phi^N(0^+, \cdot)$ and a $L^1(\Omega)$ bound on Ψ_N, we see that $\nabla\phi^N$ is uniformly bounded in $L^2([0,1] \times \Omega)$.

It remains to get a bound on $\partial_t \phi^N$. Of course, as a measure, it can be decomposed as a positive and a negative part. The negative part is concentrated on the instants $\{\tau, 2\tau, \dots, 1\}$ as $\partial_t\phi^N \geqslant 0$ on the intervals $(((k-1)\tau)^+, (k\tau)^-)$. On the other hand, on $\{\tau, 2\tau, \dots, 1\}$, the temporal derivative $\partial_t\phi^N$ coincides with $-\alpha^N$, hence the negative part is bounded as a measure. On the other hand, given that

$$\iint_{[0,1]\times\Omega} \partial_t \phi^N = \int_\Omega \Psi_N - \int_\Omega \phi^N(0^+, \cdot) \leqslant \int_\Omega \Psi_N$$

is bounded independently of N, we see that $(\partial_t\phi^N)_+ = \partial_t\phi^N + (\partial_t\phi^N)_-$ is also bounded as a measure.

As a consequence, up to the extraction of a subsequence we know that ϕ^N converges weakly in $\mathrm{BV}([0,1] \times \Omega) \cap L^2([0,1], H^1(\Omega))$ to some $\bar{\phi}$. This convergence allows easily to pass to the limit in the Hamilton-Jacobi equation satisfied (in the sense of distributions) by ϕ^N, hence $(\bar{\phi}, \bar{P})$ is admissible in the dual problem. □

The last step, to show the optimality of the limit $(\bar{\phi}, \bar{P})$, is to pass to the limit in (5.21).

Proposition 5.26. *The pair $(\bar{\phi}, \bar{P}) \in \mathcal{K}$ is a solution of the dual problem.*

Proof. We have already proved in Proposition 5.18 that

$$\lim_{N\to+\infty} \mathcal{A}^N(\bar{\rho}^N) = \min_\rho \{\mathcal{A}(\rho) \ : \ \rho \in \Gamma_0\}.$$

Given (5.21) and the duality result which holds for the continuous problem (Theorem 5.4), it is enough to show that

$$\limsup_{N\to+\infty} \left(\int_\Omega \phi^N(0^+, \cdot)\bar{\rho}_0 - \tau \sum_{k=1}^{N-1} \int_\Omega p^N_{k\tau} - \int_\Omega p^N_1 \right) \leqslant \mathcal{B}(\bar{\phi}, \bar{P}) = \int_\Omega \bar{\phi}(0^+, \cdot)\bar{\rho}_0 - \bar{p}([0,1] \times \Omega).$$

The convergence of the term involving the pressure is quite easy to show. Indeed, given the positivity of the pressures,

$$\tau \sum_{k=1}^{N-1} \int_\Omega p^N_{k\tau} + \int_\Omega p^N_1 = P^N([0,1] \times \Omega) \to \bar{p}([0,1] \times \Omega)$$

by weak convergence. On the other hand, using the definition of the trace,

$$\int_\Omega \bar{\phi}(0^+, x)\bar{\rho}_0 = \lim_{t\to 0} \frac{1}{t} \iint_{[0,t]\times\Omega} \bar{\phi}(s,x)\bar{\rho}_0(x) \mathrm{d}s\mathrm{d}x.$$

We fix some $t > 0$. Due to the convergence of ϕ^N to $\bar{\phi}$, it clearly holds

$$\lim_{N\to+\infty} \frac{1}{t} \iint_{[0,t]\times\Omega} \phi^N(s,x)\bar{\rho}_0(x)\mathrm{d}s\mathrm{d}x = \frac{1}{t} \iint_{[0,t]\times\Omega} \bar{\phi}(s,x)\bar{\rho}_0(x)\mathrm{d}s\mathrm{d}x.$$

For the value function ϕ^N, we can use the information that we have on the temporal derivative, namely $\partial_t \phi^N \geqslant -\alpha^N$. It allows us to write, given the positivity of $\bar\rho_0$,

$$\begin{aligned}\frac{1}{t}\iint_{[0,t]\times\Omega} \phi^N(s,x)\bar\rho_0(x)\mathrm{d}s\mathrm{d}x &= \frac{1}{t}\iint_{[0,t]\times\Omega}\left(\phi^N(0^+,x)+\int_0^s \partial_t\phi^N(r,x)\mathrm{d}r\right)\bar\rho_0(x)\mathrm{d}s\mathrm{d}x \\ &\geqslant \int_\Omega \phi^N(0^+,\cdot)\bar\rho_0 - \frac{1}{t}\iint_{[0,t]\times\Omega} s\alpha^N(s,x)\bar\rho_0(x)\mathrm{d}s\mathrm{d}x \\ &\geqslant \int_\Omega \phi^N(0^+,\cdot)\bar\rho_0(x) - \iint_{[0,t]\times\Omega}\alpha^N(s,x)\bar\rho_0(x)\mathrm{d}s\mathrm{d}x.\end{aligned}$$

Now, recall that $\bar\rho_0 \leqslant 1$ and α^N converges as a measure to $\bar p + V$, hence

$$\begin{aligned}\frac{1}{t}\iint_{[0,t]\times\Omega}\bar\phi(s,x)\bar\rho_0(x)\mathrm{d}s\mathrm{d}x &= \lim_{N\to+\infty}\frac{1}{t}\iint_{[0,t]\times\Omega}\phi^N(s,x)\bar\rho_0(x)\mathrm{d}s\mathrm{d}x \\ &\geqslant \limsup_{N\to+\infty}\left(\int_\Omega \phi^N(0^+,\cdot)\bar\rho_0\right) - (\bar p+V)([0,t]\times\Omega).\end{aligned}$$

Now we send t to 0, and use the fact that $(\bar p + V)(\{0\}\times\Omega) = 0$ (this can be seen as a consequence of Corollary 5.12) to conclude that

$$\limsup_{N\to+\infty}\int_\Omega \phi^N(0^+,\cdot)\bar\rho_0 \leqslant \int_\Omega \bar\phi(0^+,x)\bar\rho_0,$$

which gives us the announced result. □

To reach the conclusion of our main theorem, it is enough to show that $\bar P$ has the regularity we announced. But this easily derives from the weak convergence of P^N to $\bar P$ and the estimates of Corollary 5.12 and Corollary 5.17.

Proof of Theorem 5.5. For any smooth test functions a, b (with a being real-valued and b being vector valued), given the convergence of P^N to $\bar p$ it holds

$$\begin{cases}\displaystyle\iint_{[0,1]\times\Omega} a\bar P &= \displaystyle\lim_{N\to+\infty}\left(\sum_{k=1}^{N-1}\tau\int_\Omega a(k\tau,\cdot)p^N_{k\tau} + \int_\Omega a(1,\cdot)p^N_1\right), \\ \displaystyle\iint_{[0,1]\times\Omega}(\nabla\cdot b)\bar P &= \displaystyle -\lim_{N\to+\infty}\left(\sum_{k=1}^{N-1}\tau\int_\Omega g(k\tau,\cdot)\cdot\nabla(p^N_{k\tau}) + \int_\Omega b(1,\cdot)\cdot\nabla p^N_1\right),\end{cases}$$

where on the second line we have done an integration by parts in the r.h.s. Using the estimates given by Corollary 5.12 and Corollary 5.17, it is clear that

$$\iint_{[0,1]\times\Omega}(\nabla\cdot b)\bar P \leqslant \int_0^1\left(\|b(t,\cdot)\|_{L^2(\Omega)}\|\nabla V\|_{L^2(\Omega)}\right)\mathrm{d}t + \|b(1,\cdot)\|_{L^2(\Omega)}\|\nabla\Psi\|_{L^2(\Omega)}.$$

On the other hand, if $V, \Psi \in W^{1,q}$ with $q > d$ then, using the same propositions,

$$\iint_{[0,1]\times\Omega} a\bar P \leqslant C\int_0^1\left(\|a(t,\cdot)\|_{L^1(\Omega)}\right)\mathrm{d}t + \|a(1,\cdot)\|_{L^\infty(\Omega)},$$

where C depends only on $\|\nabla V\|_{L^q(\Omega)}, \|\nabla\Psi\|_{L^q(\Omega)}$ and Ω. Standard functional analysis manipulations provide the conclusions of Theorem 5.5. □

Chapter 6

Time-convexity of the entropy in the multiphasic formulation of the Euler equations

In this chapter we study the minimization problem

$$\min_Q \left\{ \int_\Gamma \left(\int_0^1 \frac{1}{2} |\dot\rho_t|^2 \mathrm{d}t \right) Q(\mathrm{d}\gamma) \ : \ Q \in \mathcal{P}(\Gamma) \right\},$$

where $\Gamma = C([0,1], \mathcal{P}(\Omega))$ is the set of continuous curves valued in the Wasserstein space $(\mathcal{P}(\Omega), W_2)$ and $|\dot\rho_t|$ is the metric derivative of a curve $\rho \in \Gamma$; and where Q is submitted to two constraints:

- The temporal boundary conditions are given, in the sense that if we denote $(e_0, e_1) : \Gamma \to \mathcal{P}(\Omega)^2$ the evaluation operator at time $t = 0$ and 1, the measure $(e_0, e_1)\#Q \in \mathcal{P}(\mathcal{P}(\Omega)^2)$ is fixed.

- The incompressibility constraint which states that for all $t \in [0,1]$,

  $$\int_\Gamma \rho_t Q(\mathrm{d}\rho) = \mathcal{L};$$

  in other words, if $\rho \in \Gamma$ is a random curve drawn according to Q then in expectation $\rho_t = \mathcal{L}$.

Although it may not be clear at first sight, but as it was detailed in Section 3.2, this variational problem is an instance of the *least action principle* for the incompressible Euler equations.

Recall that $H : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ is the Boltzmann entropy, see (2.13). The main goal of this chapter is to show that if ones defines, for an optimal Q,

$$\mathcal{H}(t) := \int_\Gamma H(\rho_t) Q(\mathrm{d}\rho)$$

the averaged entropy then $\mathcal{H}$ is a convex function of time. Moreover, in Section 6.4, we will show that the model used in this chapter is equivalent to the "parametric" model introduced by Brenier [Bre99, AF09].

6.1 Statement of the problem and the main result

Assumptions. *The assumption that will hold throughout this chapter is the following: the domain Ω is the closure of an open bounded convex subset of $\mathbb{R}^d$. Without loss of generality, we assume that the Lebesgue measure of Ω is 1.*

Recall (see Section 2.2) that $\Gamma = C([0,1], \mathcal{P}(\Omega))$ is the set of continuous curves valued in $(\mathcal{P}(\Omega), W_2)$ endowed with the topology of uniform convergence. More generally, if S is a closed subset of $[0,1]$, Γ_S will denote the set of continuous functions on S valued in $\mathcal{P}(\Omega)$ (in practice, we will only consider subsets S that have a finite number of points or that are subintervals of $[0,1]$). In the case where the index S is omitted, it is assumed that $S = [0,1]$. For any closed subset S' of S, the application $e_{S'} : \Gamma_S \to \Gamma_{S'}$ is the restriction operator. In the case where $S' = \{t\}$ is a singleton, we will use the notation $e_t := e_{\{t\}}$ and often use the compact writing ρ_t for $e_t(\rho) = \rho(t)$. One can see that Γ_S is a polish space, and that it is compact if S contains a finite number of points.

The space $\mathcal{P}(\Gamma_S)$, which is the space of Borel probability measures over Γ_S, is endowed with the topology of weak-$\star$ convergence of measures. As explained in Section 3.2, the object on which we will work, a "W_2-traffic plan", is a probability measure on the set of curves valued in $\mathcal{P}(\Omega)$, i.e. an element of $\mathcal{P}(\Gamma)$. If $Q \in \mathcal{P}(\Gamma)$, we need to translate the constraints, namely the fact that the values of the curves at $t = 0$ and $t = 1$ are fixed, and the incompressibility at each time t.

Incompressibility means that at each time t, the measure $e_t \# Q$ (which is an element of $\mathcal{P}(\mathcal{P}(\Omega))$) when averaged (its mean value is an element of $\mathcal{P}(\Omega)$), is equal to $\mathcal{L}$. We therefore need to define what the mean value of $e_t \# Q$ is.

Definition 6.1. *Let S be a closed subset of $[0,1]$ and $t \in S$. If $Q \in \mathcal{P}(\Gamma_S)$, we denote by $m_t(Q)$ the probability measure on Ω defined by*

$$\forall a \in C(\Omega), \quad \int_\Omega a(x)[m_t(Q)](\mathrm{d}x) := \int_{\Gamma_S} \left(\int_\Omega a(x) \rho_t(\mathrm{d}x) \right) Q(\mathrm{d}\rho). \tag{6.1}$$

We can easily see that, for a fixed t, $Q \mapsto m_t(Q)$ is continuous. It is an easy application of Fubini's theorem to show that, if Q-a.e. ρ_t is absolutely continuous w.r.t. to $\mathcal{L}$, then $m_t(Q)$ is also absolutely continuous w.r.t. $\mathcal{L}$, and its density is the mean density of the ρ_t w.r.t. Q. Incompressibility is then expressed by the fact that $m_t(Q) = \mathcal{L}$ for any t.

To encode the boundary conditions, we just consider a coupling $\gamma \in \mathcal{P}(\Gamma_{\{0,1\}}) = \mathcal{P}(\mathcal{P}(\Omega) \times \mathcal{P}(\Omega))$ between the initial and final values, compatible with the incompressibilty constraint (i.e. $m_0(\gamma) = m_1(\gamma) = \mathcal{L}$), and we impose that $(e_0, e_1)\# Q = \gamma$.

Definition 6.2. *Let $\gamma \in \mathcal{P}(\Gamma_{\{0,1\}})$ be a coupling compatible with the incompressibility constraint (i.e. $m_0(\gamma) = m_1(\gamma) = \mathcal{L}$) and S be a closed subset of $[0,1]$ containing 0 and 1. The space of incompressible W_2-traffic plans is*

$$\mathcal{P}_{\mathrm{in}}(\Gamma_S) := \{ Q \in \mathcal{P}(\Gamma_S) \ : \ \forall t \in S, \ m_t(Q) = \mathcal{L} \}.$$

The space of W_2-traffic plans satisfying the boundary conditions is

$$\mathcal{P}_{\mathrm{bc}}(\Gamma_S) := \{ Q \in \mathcal{P}(\Gamma_S) \ : \ (e_0, e_1)\# Q = \gamma \}.$$

The space of admissible W_2-traffic plans is

$$\mathcal{P}_{\mathrm{adm}}(\Gamma_S) := \mathcal{P}_{\mathrm{in}}(\Gamma_S) \cap \mathcal{P}_{\mathrm{bc}}(\Gamma_S).$$

The following proposition derives directly from the definition.

Proposition 6.3. *If S is a closed subset of $[0,1]$ containing 0 and 1, the spaces $\mathcal{P}_{\text{in}}(\Gamma_S)$, $\mathcal{P}_{\text{bc}}(\Gamma_S)$ and $\mathcal{P}_{\text{adm}}(\Gamma_S)$ are closed in $\mathcal{P}(\Gamma_S)$.*

We have now enough vocabulary to state the minimization problem we are interested in, namely to minimize the averaged action over the set of admissible W_2-traffic plans. We denote by $\mathcal{A} : \mathcal{P}(\Gamma) \to [0,+\infty]$ the functional defined by, for any $Q \in \mathcal{P}(\Gamma)$,

$$\mathcal{A}(Q) := \int_\Gamma A(\rho) Q(\mathrm{d}\rho) = \int_\Gamma \left(\int_0^1 \frac{1}{2} |\dot{\rho}_t|^2 \mathrm{d}t \right),$$

where we recall that $A(\rho)$ is the action of the curve ρ, see (2.7).

Definition 6.4. *The continuous problem is defined as*

$$\min_Q \{\mathcal{A}(Q) \ : \ Q \in \mathcal{P}_{\text{adm}}(\Gamma)\}. \tag{6.2}$$

Any $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$ with $\mathcal{A}(Q) < +\infty$ realizing the minimum will be referred as a solution of the continuous problem.

In order to prove the existence of a solution to (6.2), we rely on the classical following lemma which is valid if Γ_S is replaced by any metric space (see for instance [San15, Proposition 7.1] and [AGS08, Remark 5.15]).

Lemma 6.5. *Let S be a closed subset of $[0,1]$ and $F : \Gamma_S \to [0,+\infty]$ a l.s.c. positive function. Then the function $\mathcal{F} : \mathcal{P}(\Gamma_S) \to [0,+\infty]$ defined by*

$$\mathcal{F}(Q) = \int_{\Gamma_S} F(\rho) Q(\mathrm{d}\rho)$$

is convex and l.s.c. Moreover, if the sublevel sets of F are compact, so are those of $\mathcal{F}$.

The existence of a solution to (6.2) is then a straightforward application of the direct method of calculus of variations.

Theorem 6.6. *There exists at least one solution to* (6.2)*.*

Proof. The functional $\mathcal{A}$ is l.s.c. and has compact sublevel sets thanks to Proposition 2.9 and Lemma 6.5. Moreover the set $\mathcal{P}_{\text{adm}}(\Gamma)$ is closed. To use the direct method of calculus of variations, we only need to prove that there exists $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$ such that $\mathcal{A}(Q) < +\infty$.

Notice that as Ω is convex, it is the image of the unit cube of $\mathbb{R}^d$ by a Lipschitz and measure-preserving map (see [FP92, Theorem 5.4][1]). It is known (see [AF09, Theorem 3.3] and Proposition 6.26 to translate the result in our setting) that the fact that Ω is the image of the unit cube by a Lipschitz and measure-preserving map ensures the existence of an admissible W_2-traffic plan with finite action. □

In this chapter we are interested in the temporal behavior of the entropy when averaged over all phases. Recall that for any $\mu \in \mathcal{P}(\Omega)$, the entropy $H(\mu)$ of μ is defined through (2.13).

[1] Strictly speaking, in [FP92], it is required that Ω has a piecewise C^1 boundary, but this assumption is only used to prove that the Minkowski functional of Ω is Lipschitz. If Ω is convex, then its Minkowski functional is convex, hence Lipschitz. Thus, one can drop the assumption of a piecewise C^1 boundary if Ω is convex.

Definition 6.7. *Let S be a closed subset of $[0,1]$. For any $Q \in \mathcal{P}(\Gamma_S)$, we define the* averaged entropy *$\mathcal{H}_Q : S \to [0,+\infty]$ by, for any $t \in S$,*

$$\mathcal{H}_Q(t) := \int_{\Gamma_S} H(\rho_t) Q(\mathrm{d}\rho).$$

If $Q \in \mathcal{P}(\Gamma)$, the quantity $\displaystyle\int_0^1 \mathcal{H}_Q(t)\mathrm{d}t$ will be called the total *entropy of Q.*

By lower semi-continuity of H and Lemma 6.5, we can see that the function (of the variable t) $\mathcal{H}_Q$ is l.s.c. In the sequel, we will concentrate on the cases where the averaged entropy belongs to $L^1([0,1])$, i.e. where the total entropy is finite. By doing so, we exclude classical solutions: indeed, for a classical solution $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$, for any t the measure $e_t \# Q$ is concentrated on Dirac masses, for which the entropy is infinite. We denote by $\mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$ the set of admissible W_2-traffic plans for which the total entropy is finite:

$$\mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma) := \mathcal{P}_{\mathrm{adm}}(\Gamma) \cap \left\{ Q \in \mathcal{P}(\Gamma) \ : \ \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t < +\infty \right\}$$

The main (and restrictive) assumption that we will consider is that there exists a solution of the continuous problem (6.2) in $\mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$:

Assumption 6.1. *There exists $Q \in \mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$ such that $\mathcal{A}(Q) = \min\{\mathcal{A}(Q') \ : \ Q' \in \mathcal{P}_{\mathrm{adm}}(\Gamma)\}$.*

We will also work with a second assumption which will turn out to be more restrictive than Assumption 6.1, but which has the advantage of involving only the boundary terms, namely the fact that the initial and final values have finite averaged entropy.

Assumption 6.2. *The coupling γ is such that $\mathcal{H}_\gamma(0)$ and $\mathcal{H}_\gamma(1)$ are finite.*

In other words, we impose that

$$\int_\gamma \left(\int_\Omega \rho_0 \ln \rho_0 \right) \gamma(\mathrm{d}\rho) < +\infty \quad \text{and} \quad \int_\gamma \left(\int_\Omega \rho_1 \ln \rho_1 \right) \gamma(\mathrm{d}\rho) < +\infty.$$

In particular, Assumption 6.2 implies that $e_0 \# \gamma$ and $e_1 \# \gamma$ are concentrated on measures that are absolutely continuous w.r.t. $\mathcal{L}$: it excludes any classical boundary data.

The two main results of this chapter can be stated as follows. Recall that Ω is assumed to be convex.

Theorem 6.8. *Suppose that Assumption 6.2 holds. Then there exists a solution $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ of the continuous problem (6.2) such that $\mathcal{H}_Q(t) \leqslant \max(\mathcal{H}_\gamma(0), \mathcal{H}_\gamma(1))$ for any $t \in [0,1]$.*

In other words, if the initial and final averaged entropy are finite, then there exists a solution of the continuous problem with a uniformly bounded averaged entropy. In particular, Assumption 6.2 implies Assumption 6.1.

Theorem 6.9. *Suppose that Assumption 6.1 holds. Then, among all the solutions of the continuous problem (6.2), the unique $Q \in \mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$ which minimizes the total entropy $\int_0^1 \mathcal{H}_Q(t)\mathrm{d}t$ is such that $\mathcal{H}_Q$ is convex.*

In other words, we are able to prove the convexity of the averaged entropy for the solution which is "the most mixed", i.e. the one for which the total entropy is minimal. This statement contains the fact that the criterion of minimization of the total entropy selects a unique solution among the – potentially infinitely many – solutions of (6.2). Anyway, as we explained above in Section 3.2, it is now proved in [BM18] that the result in fact holds for all solutions, at least when Ω is the d-dimensional torus.

The next two sections are devoted to the proof of these two theorems. As explained in Chapter 3, we will introduce a discrete (in time) problem (6.3) which approximates the continuous one. Without any assumption, we will be able to prove the convexity of the averaged entropy at the discrete level (Theorem 6.12). Then we will show that, under Assumption 6.1 or Assumption 6.2, the solutions of the discrete problems converge to a solution of the continuous one (Proposition 6.18). Under Assumption 6.2, this solution will happen to have a uniformly bounded entropy (Corollary 6.20). Then we will show that, under Assumption 6.1, this solution will be the one with minimal total entropy (Corollary 6.21) and that its averaged entropy is a convex function of time (Corollary 6.25).

Finally, the uniqueness of such a $Q \in \mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$ with minimal total entropy has nothing to do with the discrete problem, it is a simple consequence of the strict convexity of $\mathcal{H}$. We will therefore prove it here to end this section. Indeed, it is a consequence of the following proposition.

Proposition 6.10. *Let Q^1 and $Q^2 \in \mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$ be two distinct admissible W_2-traffic plans. Then there exists $Q \in \mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(Q)$ with*

$$\mathcal{A}(Q) \leqslant \frac{1}{2}\left(\mathcal{A}(Q^1) + \mathcal{A}(Q^2)\right)$$

and

$$\int_0^1 \mathcal{H}_Q(t)\mathrm{d}t < \frac{1}{2}\left(\int_0^1 \mathcal{H}_{Q^1}(t)\mathrm{d}t + \int_0^1 \mathcal{H}_{Q^2}(t)\mathrm{d}t\right).$$

Proof. As $Q \mapsto \mathcal{H}_Q$ is linear, it is not sufficient to consider the mean of Q^1 and Q^2. Instead, we will need to take means in Γ. In order to do so, we disintegrate Q^1 and Q^2 w.r.t. $e_{\{0,1\}} = (e_0, e_1)$. We obtain two families $Q^1_{\rho_0,\rho_1}$ and $Q^2_{\rho_0,\rho_1}$ of W_2-traffic plans indexed by $(\rho_0, \rho_1) \in \Gamma_{\{0,1\}} = \mathcal{P}(\Omega)^2$. We define Q by its disintegration w.r.t. $e_{\{0,1\}}$: we set $Q := Q_{\rho_0,\rho_1} \otimes \gamma$ where Q_{ρ_0,ρ_1} is taken to be the image measure of $Q^1_{\rho_0,\rho_1} \otimes Q^2_{\rho_0,\rho_1}$ by the map $(\rho^1, \rho^2) \mapsto (\rho^1 + \rho^2)/2$ (where the + refers to the usual affine structure on Γ). In other words, for any $a \in C(\Gamma)$,

$$\int_\Gamma a(\rho)Q(\mathrm{d}\rho) := \int_{\Gamma_{\{0,1\}}} \left(\int_\Gamma a\left[\frac{\rho^1 + \rho^2}{2}\right] Q^1_{\rho_0,\rho_1}(\mathrm{d}\rho^1)Q^2_{\rho_0,\rho_1}(\mathrm{d}\rho^2)\right)\gamma(\mathrm{d}\rho_0, \mathrm{d}\rho_1).$$

As $(e_0, e_1)\#Q^1_{\rho_0,\rho_1}$ and $(e_0, e_1)\#Q^2_{\rho_0,\rho_1}$ are Dirac masses concentrated on (ρ_0, ρ_1), we can easily see that $Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma)$. The incompressibility constraint is straightforward to obtain: for any $a \in C(\Omega)$ and any $t \in [0, 1]$,

$$\begin{aligned}
&\int_\Omega a(x)[m_t(Q)](\mathrm{d}x) \\
&\quad = \int_{\Gamma_{\{0,1\}}} \left(\int_\Gamma \left[\int_\Omega a(x)\frac{\rho^1_t(\mathrm{d}x) + \rho^2_t(\mathrm{d}x)}{2}\right] Q^1_{\rho_0,\rho_1}(\mathrm{d}\rho^1)Q^2_{\rho_0,\rho_1}(\mathrm{d}\rho^2)\right)\gamma(\mathrm{d}\rho_0, \mathrm{d}\rho_1) \\
&\quad = \int_{\Gamma_{\{0,1\}}} \left(\int_\Gamma \left[\int_\Omega a(x)\frac{\rho^1_t(\mathrm{d}x)}{2}\right] Q^1_{\rho_0,\rho_1}(\mathrm{d}\rho^1) + \int_\Gamma \left[\int_\Omega a(x)\frac{\rho^2_t(\mathrm{d}x)}{2}\right] Q^2_{\rho_0,\rho_1}(\mathrm{d}\rho^2)\right)\gamma(\mathrm{d}\rho_0, \mathrm{d}\rho_1) \\
&\quad = \frac{1}{2}\int_\Omega a(x)\mathrm{d}x + \frac{1}{2}\int_\Omega a(x)\mathrm{d}x = \int_\Omega a(x)\mathrm{d}x.
\end{aligned}$$

Thus, we have $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$. To handle the action, let us just remark that for any ρ^1 and ρ^2 in Γ, by convexity of A,

$$A\left(\frac{\rho^1+\rho^2}{2}\right) \leqslant \frac{1}{2}\left(A(\rho^1)+A(\rho^2)\right).$$

Integrating this inequality w.r.t. to $Q^1_{\rho_0,\rho_1} \otimes Q^2_{\rho_0,\rho_1}$ and then w.r.t. γ gives the result. We use the same kind of reasoning for the entropy, but this functional is strictly convex. Hence, for any $t \in [0,1]$,

$$H\left(\frac{\rho^1_t+\rho^2_t}{2}\right) \leqslant \frac{1}{2}\left(H(\rho^1_t)+H(\rho^2_t)\right)$$

with a strict inequality if $\rho^1_t \neq \rho^2_t$ and if the r.h.s. is finite. Integrating w.r.t. t and w.r.t. $Q^1_{\rho_0,\rho_1} \otimes Q^2_{\rho_0,\rho_1}$ we get,

$$\int_\Gamma \left(\int_0^1 H\left[\frac{\rho^1_t+\rho^2_t}{2}\right]\mathrm{d}t\right) Q^1_{\rho_0,\rho_1}(\mathrm{d}\rho^1) Q^2_{\rho_0,\rho_1}(\mathrm{d}\rho^2) \leqslant \\ \frac{1}{2}\left(\int_\Gamma \left(\int_0^1 H[\rho^1_t]\mathrm{d}t\right) Q^1_{\rho_0,\rho_1}(\mathrm{d}\rho^1) + \int_\Gamma \left(\int_0^1 H[\rho^2_t]\mathrm{d}t\right) Q^2_{\rho_0,\rho_1}(\mathrm{d}\rho^2)\right),$$

with a strict inequality if $Q^1_{\rho_0,\rho_1} \neq Q^2_{\rho_0,\rho_1}$ and if the r.h.s. is finite. Then, we integrate w.r.t. γ and notice that, as $Q^1 \neq Q^2$, then $Q^1_{\rho_0,\rho_1} \neq Q^2_{\rho_0,\rho_1}$ for a γ-non negligible sets of (ρ_0,ρ_1), and as Q^1 and $Q^2 \in \mathcal{P}^{\mathcal{H}}_{\text{adm}}(\Gamma)$, the r.h.s. of the equation above is finite for γ-a.e. (ρ_0,ρ_1). Using Fubini's theorem, we are led to the announced conclusion. □

6.2 Analysis of the discrete problem

As we explained before, to tackle the continuous problem (6.2), we will introduce a discretized (in time) variational problem that approximates the continuous one. In this section, we prove its well-posedness, and show that the discrete averaged entropy is convex. In the proof of the latter property, we use the *flow interchange* technique as explained in Chapter 3.

The discrete problem is obtained by performing two different approximations:

- We consider a number of discrete times $N+1 \geqslant 2$. We will use $\tau := 1/N$ as a notation for the time step. The set $T^N \subset [0,1]$ will stand for the set of all discrete times, namely

  $$T^N := \{k\tau \ : \ k = 0,1,\dots,N\}.$$

  We use the compact notation $\Gamma_N := \Gamma_{T^N} = \mathcal{P}(\Omega)^{N+1}$. We will work with W_2-traffic plans on Γ_N, i.e. elements of $\mathcal{P}(\Gamma_N)$.

- We will also add an entropic penalization, i.e. a discretized version of

  $$\lambda \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t,$$

  with λ a small parameter. This term explains why we select, at the limit $\lambda \to 0$, the minimizers whose total entropy is minimal. It is crucial because it enables us to show that the averaged entropy of the discrete problem converges pointwisely to the averaged entropy of the continuous problem. This pointwise convergence is necessary to ensure that the averaged entropy of the continuous problem is convex. In particular, the limit $\lambda \to 0$ must be taken after $N \to +\infty$.

Let us state formally our discrete minimization problem. We fix $N \geqslant 1$ ($\tau := 1/N$) and $\lambda > 0$ and define $T^N = \{k\tau \ : \ k = 0, 1, \dots, N\}$. We denote by $\mathcal{A}^{N,\lambda} : \mathcal{P}(\Gamma_N) \to [0, +\infty]$ the functional defined by, for any $Q \in \mathcal{P}(\Gamma_N)$,

$$\mathcal{A}^{N,\lambda}(Q) := \sum_{k=1}^{N} \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} Q(\mathrm{d}\rho) + \lambda \sum_{k=1}^{N-1} \tau \mathcal{H}_Q(k\tau) .$$

The *Discrete Problem* consists in minimizing this functional under the constraint that the initial and final values are coupled through γ and the incompressibility constraint, the set of such W_2-traffic plans being $\mathcal{P}_{\mathrm{adm}}(\Gamma_N)$ (cf. Definition 6.2):

$$\min_Q \left\{ \mathcal{A}^{N,\lambda}(Q) \ : \ Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma_N) \right\} . \tag{6.3}$$

A solution of the discrete problem is a $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma_N)$ with $\mathcal{A}^{N,\lambda}(Q) < +\infty$ which minimizes $\mathcal{A}^{N,\lambda}$.

Proposition 6.11. *The discrete problem* (6.3) *admits a solution.*

Proof. We can see that $\mathcal{A}^{N,\lambda}$ is a positive l.s.c. functional. Lower semi-continuity of the discretized action and of the entropic penalization are not difficult to see thanks to Lemma 6.5.

As the space $\mathcal{P}(\Gamma_N) = \mathcal{P}(\mathcal{P}(\Omega)^{N+1})$ is compact, $\mathcal{P}_{\mathrm{adm}}(\Gamma_N)$ is also a compact space, thus it is enough to show that there exists one $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma_N)$ such that $\mathcal{A}^{N,\lambda}(Q) < +\infty$. We take Q to be equal to γ on the endpoints, and such that $e_{k\tau}\#Q$ is a Dirac mass concentrated on the Lebesgue measure $\mathcal{L}$ for any $k \in \{1, 2, \dots, N-1\}$. As $H(\mathcal{L}) = 0$ and as the incompressibility constraint $m_{k\tau}(Q) = \mathcal{L}$ is satisfied for every $k \in \{0, 1, \dots, N\}$, we can see that for this Q we have

$$\mathcal{A}^{N,\lambda}(Q) = \int_{\Gamma_{\{0,1\}}} \frac{W_2^2(\rho_0, \mathcal{L}) + W_2^2(\mathcal{L}, \rho_1)}{2\tau} \gamma(\mathrm{d}\rho).$$

As the Wasserstein distance is uniformly bounded by the diameter of Ω, the r.h.s. of the above equation is finite. The conclusion derives from a straightforward application of the direct method of calculus of variations. □

One could show that the discrete problem (6.3) admits a unique solution (it is basically the same proof as Proposition 6.10), but we will not need it. The key result of this section is the following.

Theorem 6.12. *Let $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma_N)$ be a solution of the discrete problem* (6.3)*. Then the function $k \in \{0, 1, \dots, N\} \mapsto \mathcal{H}_Q(k\tau)$ is convex, i.e. for every $k \in \{1, 2, \dots, N-1\}$,*

$$\mathcal{H}_Q(k\tau) \leqslant \frac{1}{2} \mathcal{H}_Q((k-1)\tau) + \frac{1}{2} \mathcal{H}_Q((k+1)\tau) . \tag{6.4}$$

Proof. As $\mathcal{A}^{N,\lambda}(Q)$ is finite we know that for every $k \in \{1, 2, \dots, N-1\}$, $\mathcal{H}_Q(k\tau) < +\infty$. Let us remark that if $\mathcal{H}_Q(0) = +\infty$ then there is nothing to prove in equality (6.4) for $k = 1$ (the r.h.s. being infinite); and, equivalently, if $\mathcal{H}_Q(1) = +\infty$ there is nothing to prove for $k = N-1$. So from now on, we fix $k \in \{1, 2, \dots, N-1\}$ such that $\mathcal{H}_Q((k-1)\tau)$, $\mathcal{H}_Q(k\tau)$ and $\mathcal{H}_Q((k+1)\tau)$ are finite, and it is enough to show (6.4) for such a k.

We recall that $\Phi : [0, +\infty) \times \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ denotes the heat flow, let us call $\Phi^k : [0, +\infty) \times \Gamma_N \to \Gamma_N$ the heat flow acting only on the k-th component: for any $s \geqslant 0$, $\rho \in \Gamma_N$ and $l \in \{0, 1, \dots, N\}$,

$$\Phi^k_s(\rho)(l\tau) := \begin{cases} \Phi_s(\rho_{l\tau}) & \text{if } l = k, \\ \rho_{l\tau} & \text{if } l \neq k. \end{cases}$$

If $s \geqslant 0$, it is clear that Φ_s^k leaves unchanged the boundary values, thus $\Phi_s^k \# Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma_N)$. Concerning the term $m_{l\tau}(Q)$, the linearity of the flow enables us to write

$$m_{l\tau}(\Phi_s^k \# Q) = \begin{cases} \Phi_s\left(m_{l\tau}[Q]\right) & \text{if } l = k, \\ m_{l\tau}(Q) & \text{if } l \neq k. \end{cases}$$

As $m_{l\tau}[Q] = \mathcal{L}$ and $\mathcal{L}$ is preserved by the heat flow, we conclude that $m_{l\tau}(\Phi_s^k \# Q) = \mathcal{L}$ for any $l \in \{0, 1, \dots, N\}$ hence $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma_N)$. Let us underline that the linearity of the heat flow is crucial to handle the incompressibility constraint. Our proof would not have worked if we would have wanted to show the convexity (w.r.t. time) of a functional (different from the entropy) whose gradient flow in the Wasserstein space were not linear. Using $\Phi_s^k \# Q$ as a competitor in (6.3),

$$\mathcal{A}^{N,\lambda}(Q) \leqslant \mathcal{A}^{N,\lambda}(\Phi_s^k \# Q). \tag{6.5}$$

Let us expand this formula. We can see (by definition of $\mathcal{H}_Q$) that

$$\mathcal{H}_{\Phi_s^k \# Q}\left(l\tau\right) = \begin{cases} \displaystyle\int_{\Gamma_N} H(\Phi_s[\rho_{l\tau}]) Q(\mathrm{d}\rho) & \text{if } l = k, \\ \mathcal{H}_Q\left(l\tau\right) & \text{if } l \neq k. \end{cases}$$

We can rewrite (6.5) in the following form (all the terms that do not involve the time $k\tau$ cancel):

$$\begin{aligned} \int_{\Gamma_N} & \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau}) + W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau} Q(\mathrm{d}\rho) + \lambda\tau \int_{\Gamma_N} H(\rho_{k\tau}) Q(\mathrm{d}\rho) \\ & \leqslant \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau}) + W_2^2(\Phi_s\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau} Q(\mathrm{d}\rho) + \lambda\tau \int_{\Gamma_N} H(\Phi_s\rho_{k\tau}) Q(\mathrm{d}\rho). \end{aligned}$$

It is known that the heat flow decreases the entropy (it is for example encoded in (2.11)), thus

$$\int_{\Gamma_N} H(\Phi_s\rho_{k\tau}) Q(\mathrm{d}\rho) \leqslant \int_{\Gamma_N} H(\rho_{k\tau}) Q(\mathrm{d}\rho).$$

Therefore, multiplying by τ and dividing by s, we are left with the following inequality, valid for any $s > 0$:

$$\begin{aligned} \int_{\Gamma_N} & \frac{W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau}) - W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2s} Q(\mathrm{d}\rho) \\ & \qquad + \int_{\Gamma_N} \frac{W_2^2(\Phi_s\rho_{k\tau}, \rho_{(k+1)\tau}) - W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2s} Q(\mathrm{d}\rho) \geqslant 0. \end{aligned}$$

The integrand of the first integral is exactly the rate of increase of the function $s \mapsto W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau})/2$ whose lim sup is bounded, when $s \to 0$, by $H(\rho_{(k-1)\tau}) - H(\rho_{k\tau})$ according to (2.10). Moreover, as the entropy is positive, the same inequality (2.10) shows that this rate of increase is uniformly (in s) bounded from above by $H(\rho_{(k-1)\tau})$, and the latter is integrable w.r.t. to Q. Hence by applying a reverse Fatou's lemma, we see that

$$\int_{\Gamma_N} [H(\rho_{(k-1)\tau}) - H(\rho_{k\tau})] Q(\mathrm{d}\rho) \geqslant \limsup_{s \to 0} \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau}) - W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2s} Q(\mathrm{d}\rho).$$

We have a symmetric minoration for $\int_{\Gamma_N} [H(\rho_{(k+1)\tau}) - H(\rho_{k\tau})] Q(\mathrm{d}\rho)$, hence we end up with

$$\begin{aligned} 0 &\leqslant \int_{\Gamma_N} [H(\rho_{(k-1)\tau}) - H(\rho_{k\tau})] Q(\mathrm{d}\rho) + \int_{\Gamma_N} [H(\rho_{(k+1)\tau}) - H(\rho_{k\tau})] Q(\mathrm{d}\rho) \\ &= \int_{\Gamma_N} [H(\rho_{(k-1)\tau}) + H(\rho_{(k+1)\tau}) - 2H(\rho_{k\tau})] Q(\mathrm{d}\rho) \\ &= \mathcal{H}_Q\left((k-1)\tau\right) + \mathcal{H}_Q\left((k+1)\tau\right) - 2\mathcal{H}_Q\left(k\tau\right). \end{aligned}$$

□

6.3 Limit of the discrete problems to the continuous one

In all this section, let us denote by $\bar{Q}^{N,\lambda}$ a solution (in fact there exists only one but this is not important) of the discrete problem (6.3) with parameters N and λ. We want to pass to the limit in the following way:

- We will interpolate geodesically between discrete instants and show that this builds a sequence of W_2-traffic plans which converges to a limit $\bar{Q}^\lambda \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ when $N \to +\infty$. This $\bar{Q}^\lambda$ is expected to be a solution

$$\min_Q \left\{ \mathcal{A}(Q) + \lambda \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t \ : \ Q \in \mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma) \right\}.$$

- Then, when $\lambda \to 0$, the W_2-traffic plans $\bar{Q}^\lambda$ will converge to the solution $\bar{Q}$ of the original problem with minimal total entropy and $\int_0^1 \mathcal{H}_{\bar{Q}^\lambda}(t)\mathrm{d}t$ will converge to $\int_0^1 \mathcal{H}_{\bar{Q}}(t)\mathrm{d}t$. This is the convergence of the total entropy that enables us to get a pointwise convergence of the averaged entropy.

Basically, we are performing two successive Γ-limits. Let us stress out that the order in which the limits are taken is important, though this importance may be hard to see under the various technical details. Taking the limit $\lambda \to 0$ at the end is needed to show that at the limit the selected minimizer of the continuous problem is the one with minimal total entropy (cf. the proof of Proposition 6.21).

This section is organized as follows. First we show some kind of $\Gamma - \limsup$, i.e. given continuous curves we build discrete ones whose discrete action and total entropy are close to their continuous counterparts. Then, and thanks to these constructions, we show a uniform bound on $\bar{Q}^{N,\lambda}$ that allows us to extract converging subsequences toward a limit $\bar{Q}$, and we show that $\bar{Q}$ is a solution of the continuous problem. Finally, we show that $\bar{Q}$ is the minimizer of $\mathcal{A}$ with minimal total entropy and that its averaged entropy is convex.

6.3.1 Building discrete curves from continuous ones

Let us first show a result that will be crucial to handle Assumption 6.2, namely a procedure to regularize curves in order for the total entropy to be finite.

Proposition 6.13. *Under Assumption 6.2, for any $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ and for any $\varepsilon > 0$, there exists $Q' \in \mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$ such that $\mathcal{A}(Q') \leqslant \mathcal{A}(Q) + \varepsilon$ and $\mathcal{H}_{Q'} \in L^\infty([0,1])$.*

Proof. Let us fix $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. Almost identically to the proof of Proposition 4.26, the idea is to use the heat flow Φ to regularize the curves: indeed, we know thanks to point (ii) of Proposition 2.13 that if $s > 0$ is fixed, then for any $\rho \in \Gamma$, $H(\Phi_s \rho_t)$ is bounded independently on t and ρ.

Moreover, applying uniformly the heat flow decreases the action, as already recalled in the proof of Proposition 4.26. However, by doing this, we lose the boundary values. To recover them, we squeeze the curve $\Phi_s\rho$ into the subinterval $[s, 1-s]$, and then use the heat flow (acting on ρ_0) to join ρ_0 to $\Phi_s(\rho_0)$ on $[0, s]$ and $\Phi_s(\rho_1)$ to ρ_1 on $[1-s, 1]$. Formally, for $0 < s \leqslant 1/2$, let us define the regularizing operator $R_s : \Gamma \to \Gamma$ by

$$\forall \rho \in \Gamma, \forall t \in [0,1],\ R_s(\rho)(t) := \begin{cases} \Phi_t(\rho_0) & \text{if } 0 \leqslant t \leqslant s, \\ \Phi_s\left(\rho\left[\frac{t-s}{1-2s}\right]\right) & \text{if } s \leqslant t \leqslant 1-s, \\ \Phi_{1-t}(\rho_1) & \text{if } 1-s \leqslant t \leqslant 1. \end{cases}$$

The continuity of the heat flow allows us to assert that $R_s(\rho)$ is a continuous curve. As the entropy decreases along the heat flow, and as $H(R_s[\rho])$ is uniformly bounded on $[s, 1-s]$ (independently on ρ), we can see that there exists a constant C_s depending only on s such that

$$\forall \rho \in \Gamma,\ \forall t \in [0,1],\ H[R_s(\rho)(t)] \leqslant \max(H(\rho_0), H(\rho_1), C_s). \tag{6.6}$$

To estimate the action of $R_s(\rho)$, we use the the fourth point of Proposition 2.13 and the representation formula (2.4) on $[s, 1-s]$ and the identity (2.11) to handle the boundary terms:

$$\begin{aligned} A(R_s(\rho)) &\leqslant \int_0^s \frac{1}{2}|\dot{\Phi_t\rho_0}|^2 \mathrm{d}t + \int_s^{1-s} \frac{1}{2}|\dot{\rho}_{(t-s)/(1-2s)}|^2 \mathrm{d}t + \int_{1-s}^1 \frac{1}{2}|\dot{\Phi_{1-t}\rho_1}|^2 \mathrm{d}t \\ &= \frac{H(\rho_0) - H(\Phi_s[\rho_0])}{2} + \frac{1}{1-2s}\int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t + \frac{H(\rho_1) - H(\Phi_s[\rho_1])}{2} \\ &= \frac{1}{1-2s}A(\rho) + \frac{1}{2}\left(H(\rho_0) - H(\Phi_s[\rho_0]) + H(\rho_1) - H(\Phi_s[\rho_1])\right). \end{aligned}$$

In particular, using the lower semi-continuity of the entropy H and the continuity w.r.t. s of the heat flow, we see that if $H(\rho_0)$ and $H(\rho_1)$ are finite,

$$\limsup_{s \to 0} A(R_s(\rho)) \leqslant A(\rho). \tag{6.7}$$

We are now ready to use the regularization operator on the W_2-traffic plan Q. For a fixed $0 < s \leqslant 1/2$, we define $Q_s := R_s \# Q$. As R_s does not change the boundary points, we still have $(e_0, e_1)\# Q_s = \gamma$. Integrating (6.6) w.r.t. Q, we get that

$$\forall t \in [0,1],\ \mathcal{H}_{Q_s}(t) \leqslant \mathcal{H}_{Q_s}(0) + \mathcal{H}_{Q_s}(1) + C_s = \mathcal{H}_\gamma(0) + \mathcal{H}_\gamma(1) + C_s,$$

and we know that the r.h.s. is finite because of Assumption 6.2. Concerning the action, since $\mathcal{H}(\rho_0)$ and $\mathcal{H}(\rho_1)$ are finite for Q-a.e. $\rho \in \Gamma$, we can integrate (6.7) w.r.t. Q by using a reverse Fatou's lemma to get

$$\limsup_{s \to 0} \mathcal{A}(Q_s) \leqslant \mathcal{A}(Q).$$

It remains to check the incompressibility. For a fixed $t \in [0,1]$, we notice that $e_t \# Q_s$ is of the form $(\Phi_r \circ e_{t'})\# Q$ for a some $r \geqslant 0$ and $t' \in [0,1]$ (for example, $r = t$ and $t' = 0$ if $t \in [0, s]$, and $r = s$ and $t' = (t-s)/(1-2s)$ if $t \in [s, 1-s]$). Thus, by linearity of the heat flow, $m_t(Q_s) = \Phi_r(m_{t'}[Q])$. But $m_{t'}(Q) = \mathcal{L}$ for any t' and the Lebesgue measure is preserved by the heat flow, hence $m_t(Q_s) = \mathcal{L}$.

Therefore, the Q' that we take is just Q_s for $s > 0$ small enough. □

It is then possible to show how one can build a discrete curve from a continuous one in such a way that the action and the total entropy do not increase too much. This is a standard procedure which would be valid for probability on curves valued in arbitrary geodesic spaces.

Proposition 6.14. *Let $Q \in \mathcal{P}_{\mathrm{adm}}^{\mathcal{H}}(\Gamma)$ be an admissible W_2-traffic plan with finite total entropy. For any $N \geqslant 1$, we can build a W_2-traffic plan $Q_N \in \mathcal{P}_{\mathrm{adm}}(\Gamma_N)$ in such a way that*

$$\limsup_{N \to +\infty} \mathcal{A}^{N,\lambda}(Q_N) \leqslant \mathcal{A}(Q) + \lambda \int_0^1 \mathcal{H}_Q(t) \mathrm{d}t.$$

Proof. We can assume that $\mathcal{A}(Q) < +\infty$. Similarly to Proposition 4.27, the idea is to sample each curve on a uniform grid, but not necessarily on T^N. Indeed, the key point in this sampling is to ensure that the discrete entropic penalization of the functional $\mathcal{A}^{N,\lambda}$ is bounded by $\lambda \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t$. Let us fix $N \geqslant 1$ and recall that $\tau = 1/N$. We can see that

$$\int_0^\tau \sum_{k=1}^{N-1} \mathcal{H}_Q(k\tau + s) \,\mathrm{d}s = \int_\tau^1 \mathcal{H}_Q(t)\mathrm{d}t \leqslant \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t.$$

Therefore, there exists $s_N \in (0, \tau)$ such that

$$\tau \sum_{k=1}^{N-1} \mathcal{H}_Q(k\tau + s_N) \leqslant \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t.$$

We define the sampling operator $S_N : \Gamma \to \Gamma_N$ (which samples on the grid $\{k\tau + s_N \;:\; k = 1, 2, \dots, N-1\}$) by

$$\forall \rho \in \Gamma, \forall k \in \{0, 1, \dots, N\},\; S_N(\rho)(k\tau) = \begin{cases} \rho_0 & \text{if } k = 0, \\ \rho_1 & \text{if } k = N, \\ \rho_{k\tau + s_N} & \text{if } 1 \leqslant k \leqslant N-1. \end{cases}$$

Then we simply define $Q_N := S_N \# Q$. As the initial and final values are left unchanged, it is clear that $(e_0, e_1)\#Q_N = (e_0, e_1)\#Q = \gamma$, i.e. $Q_N \in \mathcal{P}_{\mathrm{bc}}(\Gamma_N)$. By construction, we have that

$$\lambda \sum_{k=1}^{N-1} \tau \mathcal{H}_{Q_N}(k\tau) = \lambda\tau \sum_{k=1}^{N-1} \mathcal{H}_Q(k\tau + s_N) \leqslant \lambda \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t.$$

Moreover, as $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ is incompressible, it is clear that Q_N is incompressible too. The last term to handle is the action. Indeed, we have to take care of the fact that we use a translated grid which is not uniform close to the boundaries. After a standard computation (which would be valid in any geodesic space) which we already did in the proof of Proposition 4.27, one finds that

$$\sum_{k=1}^N \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} Q_N(\mathrm{d}\rho) \leqslant \mathcal{A}(Q) + \int_\Gamma \left(\int_0^{2\tau} \frac{1}{2} |\dot{\rho}_s|^2 \mathrm{d}s \right) Q(\mathrm{d}\rho).$$

For every 2-absolutely continuous curve, it is clear that the quantity $\int_0^{2\tau} \frac{1}{2}|\dot{\rho}_s|^2 \mathrm{d}s$ goes to 0 as $N \to +\infty$ and it is dominated by $A(\rho)$ which is integrable w.r.t. Q. Therefore, by dominated convergence,

$$\limsup_{N \to +\infty} \left(\sum_{k=1}^N \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} Q_N(\mathrm{d}\rho) \right) \leqslant \mathcal{A}(Q).$$

Gluing all the inequalities we have collected on Q_N, we see that $\mathcal{A}^{N,\lambda}(Q_N)$ satisfies the desired asymptotic bound. ◻

Corollary 6.15. *Under Assumption 6.1 or Assumption 6.2, there exists $C < +\infty$, such that, uniformly in $N \geqslant 1$, $\lambda \in (0,1]$ and $q > 1$, we have*

$$\mathcal{A}^{N,\lambda}(\bar{Q}^{N,\lambda}) \leqslant C.$$

Proof. Indeed, it is enough to take Q any element of $\mathcal{P}^{\mathcal{H}}_{\mathrm{adm}}(\Gamma)$ with finite action (it exists by definition under Assumption 6.1 and we use Proposition 6.13 under Assumption 6.2), to construct Q_N as in Proposition 6.14, to define $C := \sup_{N \geqslant 1} \mathcal{A}^{N,\lambda}(Q_N)$, and to use the fact that $\mathcal{A}^{N,\lambda}(\bar{Q}^{N,\lambda}) \leqslant \mathcal{A}^{N,\lambda}(Q_N) \leqslant C$. □

6.3.2 Solution of the continuous problem as a limit of discrete solutions

To go from W_2-traffic plans on discrete curves to W_2-traffic plans on continuous ones, we will need an extension operator $E_N : \Gamma_N \to \Gamma$ that interpolates a discrete curve along geodesics in $(\mathcal{P}(\Omega), W_2)$. More precisely,

Definition 6.16. *Let $N \geqslant 1$. If $\rho \in \Gamma_N$, the curve $E_N(\rho) \in \Gamma$ is defined as the one that coincides with ρ on T^N and such that for any $k \in \{0, 1, \dots, N-1\}$, the restriction of $E_N(\rho)$ to $[k\tau, (k+1)\tau]$ is a*[2] *constant-speed geodesic joining $\rho_{k\tau}$ to $\rho_{(k+1)\tau}$.*

In particular, for any $k \in \{0, 1, 2, \dots, N-1\}$, $|\dot{E_N}(\rho)|$ is constant on $[k\tau, (k+1)\tau]$ and equal to $W_2(\rho_{k\tau}, \rho_{(k+1)\tau})/\tau$. Thus, we have the identity

$$\int_{k\tau}^{(k+1)\tau} \frac{1}{2} |\dot{E_N}(\rho)_t|^2 \mathrm{d}t = \frac{W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau},$$

summed over $k \in \{0, 1, \dots, N-1\}$, these identities led to

$$A(E_N[\rho]) = \sum_{k=1}^{N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau}. \tag{6.8}$$

In other words, the action of the extended curve $E_N(\rho)$ is equal to the discrete one of ρ.

We are now ready to show the convergence of $\bar{Q}^{N,\lambda}$ to some limit $\bar{Q} \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. We take two sequences $(N_n)_{n \in \mathbb{N}}$ and $(\lambda_m)_{m \in \mathbb{N}}$ that converge respectively to $+\infty$ and 0. We will not relabel the sequences when extracting subsequences. Moreover, to avoid heavy notations, we will drop the indexes n and m, and $\lim_{n \to +\infty}$ and $\lim_{m \to +\infty}$ will be denoted respectively by $\lim_{N \to +\infty}$ and $\lim_{\lambda \to 0}$.

Proposition 6.17. *Under Assumption 6.1 or Assumption 6.2, there exists $\bar{Q} \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$, and a family $(\bar{Q}^\lambda)_\lambda \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ such that (up to extraction)*

$$\lim_{N \to +\infty} (E_N \# \bar{Q}^{N,\lambda}) = \bar{Q}^\lambda \ \textit{in } \mathcal{P}(\Gamma),$$
$$\lim_{\lambda \to 0} \bar{Q}^\lambda = \bar{Q} \quad \textit{in } \mathcal{P}(\Gamma).$$

[2]One may worry about the non uniqueness of the geodesic and hence of the fact that the extension operator E_N is ill-defined. However, it is a classical result of optimal transport that the constant-speed geodesic joining two measures is unique as soon as one of the two measures is absolutely continuous w.r.t. $\mathcal{L}$. Moreover, for a traffic plan $Q \in \mathcal{P}(\Gamma_N)$, if $\mathcal{H}_Q(t) < +\infty$ for $t \in T^N$, then Q-a.e. ρ is absolutely continuous w.r.t. $\mathcal{L}$ at time t. Thus as long as we work with W_2-traffic pans Q such that $\mathcal{H}_Q(k\tau) < +\infty$ for any $k \in \{1, 2, \dots, N-1\}$ (and we leave it to the reader to check that it is the case), the operator E_N is well defined.

Proof. We denote by C the constant given by Corollary 6.15.

We use (6.8), namely the fact that E_N transforms the discrete action into the continuous one:

$$\mathcal{A}(E_N \# \bar{Q}^{N,\lambda}) = \int_{\Gamma_N} A(E_N(\rho))\bar{Q}^{N,\lambda}(\mathrm{d}\rho) = \sum_{k=1}^{N} \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \bar{Q}^{N,\lambda}(\mathrm{d}\rho) \leqslant C,$$

where the last estimate comes from the definition of C and the positivity of the entropy. We know that the functional $\mathcal{A}$ is l.s.c. and that its sublevel sets are compact. Hence, we get the existence of $(\bar{Q}^\lambda)_\lambda$ such that

$$\lim_{N\to+\infty} (E_N \# \bar{Q}^{N,\lambda}) = \bar{Q}^\lambda$$

in $\mathcal{P}(\Gamma)$ and $\mathcal{A}(\bar{Q}^\lambda) \leqslant C$. Applying exactly the same argument, we can conclude at the existence of $\bar{Q} \in \mathcal{P}(\Gamma)$ with

$$\lim_{\lambda\to 0} \bar{Q}^\lambda = \bar{Q}$$

in $\mathcal{P}(\Gamma)$ together with $\mathcal{A}(\bar{Q}) \leqslant C$.

It is easy to show that $(e_0, e_1)\#\bar{Q} = \gamma$ as we have that $(e_0, e_1)\#\bar{Q}^{N,\lambda} = \gamma$: this condition passes to the limit and is preserved by E_N.

The part which is not direct is the incompressibility of $\bar{Q}^\lambda$. We recall that $m_{k\tau}(\bar{Q}^{N,\lambda}) = \mathcal{L}$ for any $k \in \{0, 1, \ldots, N\}$ and any λ. Then to show that the incompressibility constraint is satisfied by $\bar{Q}^\lambda$ for every t, we proceed as follows: let us consider $t \in [0,1]$ and $N \geqslant 1$. Let $k \in \{0, 1, \ldots, N-1\}$ such that $k\tau \leqslant t \leqslant (k+1)\tau$. We denote by $s \in [0,1]$ the real such that $t = (k+s)\tau$. By definition of E_N, if $\rho \in \Gamma_N$, there exists $\bar{\gamma}$ an optimal transport plan between $\rho_{k\tau}$ and $\rho_{(k+1)\tau}$ (i.e. an optimal γ in formula (2.1) with $\mu = \rho_{k\tau}$ and $\nu = \rho_{(k+1)\tau}$) such that $E_N(\rho)(t) = \pi_s \# \bar{\gamma}$ with $\pi_s : (x,y) \mapsto (1-s)x + sy$. For any $a \in C^1(\Omega)$, we can see that

$$\begin{aligned}
\left| \int_\Omega a \mathrm{d}[E_N(\rho)(t)] - \int_\Omega a \mathrm{d}\rho_{k\tau} \right| &= \left| \int_{\Omega\times\Omega} (a[(1-s)x + sy] - a[x]) \bar{\gamma}(\mathrm{d}x, \mathrm{d}y) \right| \\
&\leqslant \int_{\Omega\times\Omega} s |\nabla a(x)| |x-y| \bar{\gamma}(\mathrm{d}x, \mathrm{d}y) \\
&\leqslant \sqrt{\int_{\Omega\times\Omega} |\nabla a(x)|^2 \bar{\gamma}(\mathrm{d}x, \mathrm{d}y)} \sqrt{\int_{\Omega\times\Omega} |x-y|^2 \bar{\gamma}(\mathrm{d}x, \mathrm{d}y)} \\
&\leqslant \|\nabla a\|_{L^\infty} W_2(\rho_{k\tau}, \rho_{(k+1)\tau}).
\end{aligned}$$

Therefore, if we estimate the action of $m_t(E_N \# \bar{Q}^{N,\lambda})$ on a C^1 function a, we find that

$$\begin{aligned}
\left| \int_\Omega a \mathrm{d}[m_t(E_N \# \bar{Q}^{N,\lambda})] - \int_\Omega a(x) \mathrm{d}x \right| &= \left| \int_\Omega a \mathrm{d}[m_t(E_N \# \bar{Q}^{N,\lambda})] - \int_\Omega a \mathrm{d}[m_{k\tau}(\bar{Q}^{N,\lambda})] \right| \\
&\leqslant \int_{\Gamma_N} \left| \int_\Omega a \mathrm{d}[E_N(\rho)(t)] - \int_\Omega a \mathrm{d}\rho_{k\tau} \right| \bar{Q}^{N,\lambda}(\mathrm{d}\rho) \\
&\leqslant \|\nabla a\|_{L^\infty} \int_{\Gamma_N} W_2(\rho_{k\tau}, \rho_{(k+1)\tau}) \bar{Q}^{N,\lambda}(\mathrm{d}\rho) \\
&\leqslant \sqrt{2\tau} \|\nabla a\|_{L^\infty} \sqrt{\int_{\Gamma_N} \frac{W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau} \bar{Q}^{N,\lambda}(\mathrm{d}\rho)} \\
&\leqslant \sqrt{2C\tau} \|\nabla a\|_{L^\infty}.
\end{aligned}$$

Taking the limit $N \to +\infty$ (hence $\tau \to 0$), we know that $m_t(E_N \# \bar{Q}^{N,\lambda})$ converges to $m_t(\bar{Q}^\lambda)$, thus we get

$$\int_\Omega a \mathrm{d}[m_t(\bar{Q}^\lambda)] = \int_\Omega a(x)\mathrm{d}x.$$

As a is an arbitrary C^1 function, we have the equality $m_t(\bar{Q}^\lambda) = \mathcal{L}$ for any t, in other words, $\bar{Q}^\lambda \in \mathcal{P}_{\text{in}}(\Gamma)$. As we already know that $\bar{Q}^\lambda \in \mathcal{P}_{\text{bc}}(\Gamma)$, we conclude that $\bar{Q}^\lambda \in \mathcal{P}_{\text{adm}}(\Gamma)$ for any $\lambda > 0$. But $\mathcal{P}_{\text{adm}}(\Gamma)$ is closed, therefore $\bar{Q} \in \mathcal{P}_{\text{adm}}(\Gamma)$. □

With all the previous work, it is easy to conclude that $\bar{Q}$ is a minimizer of $\mathcal{A}$: we just copy a standard proof of Γ-convergence.

Proposition 6.18. *Under Assumption 6.1 or Assumption 6.2, $\bar{Q}$ is a solution of the continuous problem* (6.2)*.*

Proof. We have already seen that $\mathcal{A}(E_N \# \bar{Q}^{N,\lambda}) \leqslant \mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,\lambda})$. By lower semi-continuity of $\mathcal{A}$, we deduce that

$$\mathcal{A}(\bar{Q}) \leqslant \liminf_{\lambda \to 0} \left(\liminf_{N \to +\infty} \mathcal{A}^{N,\lambda}(\bar{Q}^{N,\lambda}) \right).$$

By contradiction, let us assume that there exists $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$ such that $\mathcal{A}(Q) < \mathcal{A}(\bar{Q})$. If we are under Assumption 6.2, we can regularize it thanks to Proposition 6.13, and under Assumption 6.1 we know that we can assume that $Q' \in \mathcal{P}^{\mathcal{H}}_{\text{adm}}(\Gamma)$ and $\mathcal{A}(Q') \leqslant \mathcal{A}(Q)$. In any of these two cases, we can assume that there exists $Q \in \mathcal{P}^{\mathcal{H}}_{\text{adm}}(\Gamma)$ such that $\mathcal{A}(Q) < \mathcal{A}(\bar{Q})$. Thanks to Proposition 6.14, we know that we can construct a sequence Q_N with

$$\limsup_{N \to +\infty} \mathcal{A}^{N,\lambda}(Q_N) \leqslant \mathcal{A}(Q) + \lambda \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t$$

Taking the limit $\lambda \to 0$ and using $\mathcal{A}(Q) < \mathcal{A}(\bar{Q})$, we get

$$\limsup_{\lambda \to 0} \left(\limsup_{N \to +\infty} \mathcal{A}^{N,\lambda}(Q_N) \right) < \mathcal{A}(\bar{Q}) \leqslant \liminf_{\lambda \to 0} \left(\liminf_{N \to +\infty} \mathcal{A}^{N,\lambda}(\bar{Q}^{N,\lambda}) \right).$$

Taking N large enough and λ small enough, one has $\mathcal{A}^{N,\lambda}(Q_N) < \mathcal{A}^{N,\lambda}(\bar{Q}^{N,\lambda})$, which contradicts the optimality of $\bar{Q}^{N,\lambda}$. □

6.3.3 Behavior of the averaged entropy of $\bar{Q}$

Now, we will show that $\mathcal{H}_{\bar{Q}} \in L^1([0,1])$ and that $\bar{Q}$ is the minimizer of $\mathcal{A}$ with minimal total entropy. If $Q \in \mathcal{P}(\Gamma_N)$, let us denote by $\mathcal{H}^{\text{int}}_Q : [0,1] \to [0,+\infty]$ the piecewise affine interpolation of $\mathcal{H}_Q$. More precisely, if $k \in \{0,1,\dots,N-1\}$ and $s \in [0,1]$, we define

$$\mathcal{H}^{\text{int}}_Q\left((k+s)\tau\right) := (1-s)\mathcal{H}_Q\left(k\tau\right) + s\mathcal{H}_Q\left((k+1)\tau\right).$$

We show the following estimate, which relies on the lower semi-continuity of the entropy:

Proposition 6.19. *For any $t \in [0,1]$, we have the following upper bound for $\mathcal{H}_{\bar{Q}}(t)$:*

$$\mathcal{H}_{\bar{Q}}(t) \leqslant \liminf_{\lambda \to 0} \left(\liminf_{N \to +\infty} \mathcal{H}^{int}_{\bar{Q}^{N,\lambda}}(t) \right).$$

Proof. We will use the fact that the entropy is geodesically convex, i.e. convex along the constant-speed geodesics. Recall that $E_N : \Gamma_N \to \Gamma$ is the extension operator that interpolates along constant-speed geodesics. Let us take $\rho \in \Gamma_N$. By geodesic convexity, we have for any $k \in \{0, 1, \dots, N-1\}$ and $s \in [0,1]$

$$H\left[E_N(\rho)\left((k+s)\tau\right)\right] \leqslant (1-s)H(\rho_{k\tau}) + sH(\rho_{(k+1)\tau}).$$

Integrating this inequality over Γ_N w.r.t. $\bar{Q}^{N,\lambda}$, we get

$$\begin{aligned}\mathcal{H}_{E_N\#\bar{Q}^{N,\lambda}}\left((k+s)\tau\right) &\leqslant (1-s)\mathcal{H}_{\bar{Q}^{N,\lambda}}\left(k\tau\right) + s\mathcal{H}_{\bar{Q}^{N,\lambda}}\left((k+1)\tau\right) \\ &\leqslant \mathcal{H}^{\text{int}}_{\bar{Q}^{N,q,\lambda}}\left((k+s)\tau\right).\end{aligned}$$

We take the limit $N \to +\infty$, followed by $\lambda \to 0$ to get (thanks to the lower semi-continuity of the averaged entropy) the announced inequality. □

We derive a useful consequence, which implies Theorem 6.8.

Corollary 6.20. *Under Assumption 6.2, the function $\mathcal{H}_{\bar{Q}}$ is bounded by* $\max(\mathcal{H}_\gamma(0), \mathcal{H}_\gamma(1))$.

Proof. This is where we use the work of Section 6.2: thanks to Theorem 6.12, we know that $\mathcal{H}_{\bar{Q}^{N,\lambda}}$ is convex and therefore bounded by the values at its endpoints which happen to be finite (independently of N and λ):

$$\forall k \in \{0, 1, 2, \dots, N\},\ \mathcal{H}_{\bar{Q}^{N,\lambda}}\left(k\tau\right) \leqslant \max(\mathcal{H}_\gamma(0), \mathcal{H}_\gamma(1)).$$

Thus the function $\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}$ is also bounded uniformly on $[0,1]$ by $\max(\mathcal{H}_\gamma(0), \mathcal{H}_\gamma(1))$. Proposition 6.19 allows us to conclude that the same bound holds for $\mathcal{H}_{\bar{Q}}$. □

As we have now proved Theorem 6.8, we will work only under Assumption 6.1. It remains to show that the $\bar{Q}$ we constructed is the one with minimal total entropy. This is done thanks to the entropic penalization, and is standard in Γ-convergence theory, the specific structure of the Wasserstein space does not play any role.

Proposition 6.21. *For any $Q \in \mathcal{P}^{\mathcal{H}}_{\text{adm}}(\Gamma)$ solution of the continuous problem* (6.2)*, we have*

$$\int_0^1 \mathcal{H}_{\bar{Q}}(t)\mathrm{d}t \leqslant \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t$$

Proof. Let us start with an exact quadrature formula for $\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}$:

$$\int_\tau^{1-\tau} \mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}(t)\mathrm{d}t = \frac{\tau}{2}\mathcal{H}_{\bar{Q}^{N,\lambda}}\left(\tau\right) + \tau\sum_{k=2}^{N-2}\mathcal{H}_{\bar{Q}^{N,\lambda}}\left(k\tau\right) + \frac{\tau}{2}\mathcal{H}_{\bar{Q}^{N,\lambda}}\left(1-\tau\right) \leqslant \tau\sum_{k=1}^{N-1}\mathcal{H}_{\bar{Q}^{N,\lambda}}\left(k\tau\right)$$

Then we take successively the limits $N \to +\infty$ and $\lambda \to 0$, applying Fatou's lemma and using Proposition 6.19 to get

$$\int_0^1 \mathcal{H}_{\bar{Q}}(t)\mathrm{d}t \leqslant \liminf_{\lambda\to 0}\left(\liminf_{N\to+\infty}\left(\tau\sum_{k=1}^{N-1}\mathcal{H}_{\bar{Q}^{N,\lambda}}\left[k\tau\right]\right)\right). \tag{6.9}$$

On the other hand, let us show that the r.h.s. of (6.9) is smaller than the total entropy of any minimizer of (6.2). Indeed, assume that this is not the case for some $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$ solution of (6.2). In particular, for some $\lambda > 0$ small enough, we have the strict inequality

$$\int_0^1 \mathcal{H}_Q(t)\mathrm{d}t < \liminf_{N\to+\infty}\left(\tau\sum_{k=1}^{N-1}\mathcal{H}_{\bar{Q}^{N,\lambda}}\left[k\tau\right]\right).$$

Using the fact that $\mathcal{A}(Q) \leqslant \mathcal{A}(\bar{Q}^\lambda)$ by optimality of Q, and thanks to the lower semi-continuity of the action,

$$\mathcal{A}(Q) \leqslant \mathcal{A}(\bar{Q}^\lambda) \leqslant \liminf_{N\to+\infty} \left(\sum_{k=1}^{N} \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \bar{Q}^{N,\lambda}(\mathrm{d}\rho) \right).$$

Therefore, gluing these two estimates together, we obtain

$$\begin{aligned}\mathcal{A}(Q) &+ \lambda \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t \\ &< \liminf_{N\to+\infty} \left(\liminf_{q\to+\infty} \left(\sum_{k=1}^{N} \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \bar{Q}^{N,\lambda}(\mathrm{d}\rho) + \lambda \sum_{k=1}^{N-1} \tau \mathcal{H}_{\bar{Q}^{N,\lambda}}[k\tau] \right) \right).\end{aligned}$$

But if we build the Q_N from Q as in Proposition 6.14, we get, for N large enough,

$$\mathcal{A}^{N,\lambda}(Q_N) < \sum_{k=1}^{N} \int_{\Gamma_N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \bar{Q}^{N,\lambda}(\mathrm{d}\rho) + \lambda \sum_{k=1}^{N-1} \tau \mathcal{H}_{\bar{Q}^{N,\lambda}}(k\tau) \leqslant \mathcal{A}^{N,\lambda}(\bar{Q}^{N,\lambda}),$$

which is a contradiction with the optimality of $\bar{Q}^{N,\lambda}$. Hence, we have proved that for any $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ solution of the continuous problem,

$$\int_0^1 \mathcal{H}_{\bar{Q}}(t)\mathrm{d}t \leqslant \liminf_{\lambda\to 0} \left(\liminf_{N\to+\infty} \left(\tau \sum_{k=1}^{N-1} \mathcal{H}_{\bar{Q}^{N,\lambda}}[k\tau] \right) \right) \leqslant \int_0^1 \mathcal{H}_Q(t)\mathrm{d}t. \tag{6.10}$$

□

Now it remains to show that $\mathcal{H}_{\bar{Q}}$ is a convex function of time. This will be done by proving that $\mathcal{H}_{\bar{Q}}$ is the limit of $\mathcal{H}^{\mathrm{int}}_{\bar{Q}^{N,\lambda}}$.

Proposition 6.22. *Under Assumption 6.1, for a.e. $t \in [0,1]$,*

$$\mathcal{H}_{\bar{Q}}(t) = \lim_{\lambda\to 0} \left(\lim_{N\to+\infty} \left(\mathcal{H}^{int}_{\bar{Q}^{N,\lambda}}(t) \right) \right).$$

Proof. Taking $Q = \bar{Q}$ in (6.10), we see that, up to extraction,

$$\int_0^1 \mathcal{H}_{\bar{Q}}(t)\mathrm{d}t = \lim_{\lambda\to 0} \left(\lim_{N\to+\infty} \left(\tau \sum_{k=1}^{N-1} \mathcal{H}_{\bar{Q}^{N,\lambda}}[k\tau] \right) \right).$$

In other words, the integral over time of the discrete averaged entropy converges to the integral of the continuous one. As we know moreover that the discrete averaged entropy is an upper bound for the continuous one (Proposition 6.19), it is not difficult to show that the discrete averaged entropy converges (up to extraction) a.e. to the continuous one. □

6.3.4 From convexity a.e. to true convexity

Proposition 6.22 is slightly weaker than the result we claimed, as we get information about $\mathcal{H}_{\bar{Q}}$ only for a.e. time. The first step toward true convexity is to show that, under Assumption 6.2, the averaged entropy is everywhere below the line joining the endpoints.

Proposition 6.23. *Under Assumption 6.2, for any $t \in [0,1]$, we have*

$$\mathcal{H}_{\bar{Q}}(t) \leqslant (1-t)\mathcal{H}_{\bar{Q}}(0) + t\mathcal{H}_{\bar{Q}}(1).$$

Proof. From Proposition 6.22, we know that $\mathcal{H}_{\bar{Q}}$ is a.e. the limit of the functions $\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}$. Thanks to Theorem 6.12, we can assert that for any $t \in [0,1]$, one has $\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}(t) \leqslant (1-t)\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}(0) + t\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}(1)$. We also know that $\mathcal{H}_{\bar{Q}}$ and $\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}$ coincide for $t = 0$ and $t = 1$. Therefore, for a.e. $t \in [0,1]$,

$$\begin{aligned}
\mathcal{H}_{\bar{Q}}(t) &= \lim_{\lambda \to 0} \left(\lim_{N \to +\infty} \left(\mathcal{H}^{\text{int}}_{\bar{Q}^{N,\lambda}}(t) \right) \right) \\
&\leqslant \lim_{\lambda \to 0} \left(\lim_{N \to +\infty} \left((1-t)\mathcal{H}^{\text{int}}_{\bar{Q}^{N,q,\lambda}}[0] + t\mathcal{H}^{\text{int}}_{\bar{Q}^{N,q,\lambda}}[1] \right) \right) \\
&= (1-t)\mathcal{H}_{\bar{Q}}(0) + t\mathcal{H}_{\bar{Q}}(1).
\end{aligned}$$

As $\mathcal{H}_{\bar{Q}}$ is l.s.c., we see that the above inequality is valid for any $t \in [0,1]$. □

Now, if $\bar{Q}$ is the solution of the continuous problem (6.2) with minimal total entropy, then its restriction to any subinterval of $[0,1]$ is also optimal: for any $0 \leqslant t_1 < t_2 \leqslant 1$, $e_{[t_1,t_2]}\#\bar{Q}$ is also the solution of the continuous problem (on $[t_1,t_2]$) with boundary conditions $e_{\{t_1,t_2\}}\#\bar{Q}$ with minimal total entropy. This is already known [AF09, Remark 3.2 and below] and comes from the fact that we can concatenate traffic plans.

Proposition 6.24. *Let $0 \leqslant t_1 < t_2 \leqslant 1$. Then for any $Q \in \mathcal{P}_{\text{adm}}(\Gamma_{[t_1,t_2]})$ such that $e_{\{t_1,t_2\}}\#Q = e_{\{t_1,t_2\}}\#\bar{Q}$, we have*

$$\int_\Gamma \left(\int_{t_1}^{t_2} \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t \right) \bar{Q}(\mathrm{d}\rho) \leqslant \int_{\Gamma_{[t_1,t_2]}} \left(\int_{t_1}^{t_2} \frac{1}{2}|\dot{\rho}_t|^2 \mathrm{d}t \right) Q(\mathrm{d}\rho).$$

Moreover, if the inequality above is an equality, then

$$\int_{t_1}^{t_2} \mathcal{H}_{\bar{Q}}(t)\mathrm{d}t \leqslant \int_{t_1}^{t_2} \mathcal{H}_Q(t)\mathrm{d}t.$$

Proof. This property relies on the fact that if $Q \in \mathcal{P}_{\text{adm}}(\Gamma_{[t_1,t_2]})$ with $e_{\{t_1,t_2\}}\#Q = e_{\{t_1,t_2\}}\#\bar{Q}$, we can concatenate Q and $\bar{Q}$ together to build a W_2-traffic plan $Q' \in \mathcal{P}(\Gamma)$ such that $e_{[0,1]\setminus[t_1,t_2]}\#Q' = e_{[0,1]\setminus[t_1,t_2]}\#\bar{Q}$ and $e_{[t_1,t_2]}\#Q' = e_{[t_1,t_2]}\#Q$. To do that, it is enough to disintegrate the measures $\bar{Q}$ and Q w.r.t. $e_{\{t_1,t_2\}}$ and then to concatenate elements of $\Gamma_{[0,1]\setminus[t_1,t_2]}$ and $\Gamma_{[t_1,t_2]}$ which coincides on $\{t_1,t_2\}$: we leave the details to the reader. □

Combining the two above propositions, we recover the convexity of $\mathcal{H}_{\bar{Q}}$. Let us remark that we rely on the fact that the minimizer of $\mathcal{A}$ with minimal total entropy is unique.

Corollary 6.25. *Under Assumption 6.1 or Assumption 6.2, for any $0 \leqslant t_1 < t_2 \leqslant 1$ and any $s \in (0,1)$,*

$$\mathcal{H}_{\bar{Q}}((1-s)t_1 + st_2) \leqslant (1-s)\mathcal{H}_{\bar{Q}}(t_1) + s\mathcal{H}_{\bar{Q}}(t_2).$$

Proof. If the r.h.s. is infinite, there is nothing to prove. Therefore, we can assume that $\mathcal{H}_{\bar{Q}}(t_1)$ and $\mathcal{H}_{\bar{Q}}(t_2)$ are finite. By uniqueness of the solution with minimal total entropy (Proposition 6.10), we know that $e_{[t_1,t_2]}\#\bar{Q}$ coincides with the solution of the continuous problem (6.2) with minimal total entropy on $[t_1,t_2]$ with boundary conditions $e_{\{t_1,t_2\}}\#\bar{Q}$ (Proposition 6.24). As $\mathcal{H}_{\bar{Q}}(t_1)$ and $\mathcal{H}_{\bar{Q}}(t_2)$ are finite, Assumption 6.2 is satisfied for the continuous problem on $[t_1,t_2]$ and therefore we can apply Proposition 6.23 to get

$$\mathcal{H}_{\bar{Q}}((1-s)t_1 + st_2) \leqslant (1-s)\mathcal{H}_{\bar{Q}}(t_1) + s\mathcal{H}_{\bar{Q}}(t_2).$$ □

6.4 Equivalence with the parametric formulation of the Euler equation

In this section we will explain why our non-parametric formulation is equivalent to Brenier's parametric one that we presented in Section 3.2. From the way we build it, it is clear that our formulation admits more potential solutions than Brenier's one, so the only technical point will be to show that, if the boundary data are in a parametric form, it is possible to parametrize the *a priori* non-parametric solution of the continuous problem.

Let us take $\mathfrak{A}$ a polish space and consider $\theta \in \mathcal{P}(\mathfrak{A})$ a Borel probability measure on $\mathfrak{A}$. We will assume that we have two families (the initial and the final) $(\rho_i^\alpha)_{\alpha\in\mathfrak{A}}$ and $(\rho_f^\alpha)_{\alpha\in\mathfrak{A}}$ of probabilities measures on Ω indexed by $\mathfrak{A}$. We denote by $P_{\text{bc}} : \mathfrak{A} \to \Gamma_{\{0,1\}} = \mathcal{P}(\Omega)^2$ the parametrization of the boundary conditions, simply defined by $P_{\text{bc}}(\alpha) = (\rho_i^\alpha, \rho_f^\alpha)$ and assume that it is measurable. We assume that the boundary data satisfy the incompressibility condition, i.e.

$$\int_{\mathfrak{A}} \rho_i^\alpha \theta(\mathrm{d}\alpha) = \mathcal{L} \quad \text{and} \quad \int_{\mathfrak{A}} \rho_f^\alpha \theta(\mathrm{d}\alpha) = \mathcal{L}.$$

Translated in our language, if we set $\gamma := P_{\text{bc}}\#\theta$, we simply impose that $m_0(\gamma) = m_1(\gamma) = \mathcal{L}$.

A measurable family $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t)\in\mathfrak{A}\times[0,1]}$ indexed by α and t such that, for θ-a.e. α, $(t \mapsto \rho_t^\alpha) \in \Gamma$ and $\mathbf{v}_t^\alpha \in L^2(\Omega, \mathbb{R}^d, \rho_t^\alpha)$ for a.e. t, is said to be admissible if

$$\begin{cases} \rho_0^\alpha = \rho_i^\alpha \text{ and } \rho_1^\alpha = \rho_f^\alpha & \text{for } \theta\text{-a.e. } \alpha, \\ \partial_t \rho_t^\alpha + \nabla \cdot (\rho_t^\alpha \mathbf{v}_t^\alpha) = 0 & \text{in a weak sense with no-flux boundary conditions for } \theta\text{-a.e. } \alpha, \\ \displaystyle\int_{\mathfrak{A}} \rho_t^\alpha \theta(\mathrm{d}\alpha) = \mathcal{L} & \text{for all } t \in [0,1]. \end{cases}$$

The first equation corresponds to the temporal boundary conditions, the second one is the continuity equation while the last one is the coding of the incompressibility. If $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t)\in\mathfrak{A}\times[0,1]}$ is an admissible family, we define its (parametrized) action $\mathcal{A}_P$ by

$$\mathcal{A}_P(\rho, \mathbf{v}) := \int_{\mathfrak{A}} \int_0^1 \int_\Omega \frac{1}{2} |\mathbf{v}_t^\alpha(x)|^2 \rho_t^\alpha(\mathrm{d}x)\mathrm{d}t\theta(\mathrm{d}\alpha)$$

and its parametrized averaged entropy $\mathcal{H}_P(\rho, \mathbf{v}) : [0,1] \to \mathbb{R}$ by, for any $t \in [0,1]$,

$$\mathcal{H}_P(\rho, \mathbf{v})(t) := \int_{\mathfrak{A}} H(\rho_t^\alpha)\theta(\mathrm{d}\alpha).$$

The first proposition is very simple: it asserts that every parametric family can be seen as an non parametric one. In the sequel, we define the boundary conditions $\gamma \in \mathcal{P}_{\text{in}}(\Gamma_{\{0,1\}})$ for the non-parametric problem by $\gamma := P_{\text{bc}}\#\theta$.

Proposition 6.26. *Let $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t)\in\mathfrak{A}\times[0,1]}$ be an admissible family. Then there exists $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$ such that $\mathcal{A}(Q) \leqslant \mathcal{A}_P(\rho, \mathbf{v})$ and $\mathcal{H}_Q(t) = \mathcal{H}_P(\rho, \mathbf{v})(t)$ for any $t \in [0,1]$.*

Proof. Let $P : \mathfrak{A} \to \Gamma$, defined by $P(\alpha) = (t \mapsto \rho_t^\alpha)$ be the parametrization. We set $Q := P\#\theta$ and leave it to the reader to check that this choice works (Theorem 2.8 might be useful). □

The reverse proposition is slightly more difficult to prove: it asserts that one can always build a parametric family from a non-parametric W_2-traffic plan in such a way that the global action and the total entropy decrease. In particular, it implies together with Proposition 6.26 that (provided that the boundary conditions are in a parametric form) the solution of the continuous problem (6.2) with minimal total entropy can be parametrized.

Proposition 6.27. *Let $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. Then there exists an admissible family $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t)\in\mathfrak{A}\times[0,1]}$ such that $\mathcal{A}_P(\rho,\mathbf{v}) \leqslant \mathcal{A}(Q)$ and $\mathcal{H}_P(\rho,\mathbf{v})(t) \leqslant \mathcal{H}_Q(t)$ for any $t \in [0,1]$.*

Proof. Let us disintegrate Q w.r.t. to $e_{\{0,1\}} = (e_0, e_1)$. We obtain a family $(Q_{\rho_0,\rho_1})_{\rho_0,\rho_1}$ of W_2-traffic plans indexed by $(\rho_0,\rho_1) \in \Gamma_{\{0,1\}} = \mathcal{P}(\Omega)^2$. We define the curve ρ_t^α as the average of all the curves in Γ w.r.t. to $Q_{\rho_i^\alpha,\rho_f^\alpha}$: for any $t \in [0,1]$ and any α for which $Q_{\rho_i^\alpha,\rho_f^\alpha}$ is defined (and this property holds for θ-a.e. α), we set

$$\rho_t^\alpha := m_t\left(Q_{\rho_i^\alpha,\rho_f^\alpha}\right).$$

By definition of disintegration, $e_{\{0,1\}}\#Q_{\rho_i^\alpha,\rho_f^\alpha}$ is a Dirac mass at the point $(\rho_i^\alpha, \rho_f^\alpha)$, thus the boundary conditions are satisfied. The incompressibility condition is just a consequence of the incompressibility of Q: for any $a \in C(\Omega)$,

$$\begin{aligned}
\int_{\mathfrak{A}}\left(\int_\Omega a(x)\rho_t^\alpha(\mathrm{d}x)\right)\theta(\mathrm{d}\alpha) &= \int_{\mathfrak{A}}\left(\int_\Gamma\left[\int_\Omega a(x)\rho_t(\mathrm{d}x)\right]Q_{\rho_i^\alpha,\rho_f^\alpha}(\mathrm{d}\rho)\right)\theta(\mathrm{d}\alpha) \\
&= \int_{\Gamma_{\{0,1\}}}\left(\int_\Gamma\left[\int_\Omega a(x)\rho_t(\mathrm{d}x)\right]Q_{\rho_0,\rho_1}(\mathrm{d}\rho)\right)\gamma(\mathrm{d}\rho_0,\mathrm{d}\rho_1) \\
&= \int_\Gamma\left(\int_\Omega a(x)\rho_t(\mathrm{d}x)\right)Q(\mathrm{d}\rho) \\
&= \int_\Omega a(x)\mathrm{d}x.
\end{aligned}$$

To handle the action, we use the fact that A is convex and l.s.c. Thus, thanks to Jensen's inequality, for θ-a.e. α,

$$A(\rho^\alpha) \leqslant \int_\Gamma A(\rho)Q_{\rho_i^\alpha,\rho_f^\alpha}(\mathrm{d}\rho).$$

Integrating w.r.t. θ, we end up with

$$\int_{\mathfrak{A}} A(\rho^\alpha)\theta(\mathrm{d}\alpha) \leqslant \mathcal{A}(Q).$$

We consider only the case $\mathcal{A}(Q) < +\infty$ (else there is nothing to prove). Thus, for θ-a.e. α the quantity $A(\rho^\alpha)$ is finite. By Theorem 2.8, we can find for each α a family $(\mathbf{v}_t^\alpha)_{t\in[0,1]}$ of functions $\Omega \to \mathbb{R}^d$ such that the continuity equation is satisfied, $\mathbf{v}_t^\alpha \in L^2(\Omega, \mathbb{R}^d, \rho_t^\alpha)$ for a.e. t and such that the following identity holds

$$\int_0^1\int_\Omega \frac{1}{2}|\mathbf{v}_t^\alpha(x)|^2\rho_t^\alpha(\mathrm{d}x)\mathrm{d}t = \int_{\mathfrak{A}}\int_0^1 \frac{1}{2}|\dot\rho_t^\alpha|^2\mathrm{d}t.$$

Therefore, we see that the family $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t)\in\mathfrak{A}\times[0,1]}$ is admissible and, integrating the last equality w.r.t. θ, that $\mathcal{A}_P(\rho,\mathbf{v}) \leqslant \mathcal{A}(Q)$.

To get the inequality involving the entropy, we use the fact that the functional $\mathcal{H}$ is convex and l.s.c. on $\mathcal{P}(\Omega)$, thus by Jensen's inequality,

$$H\left(m_t\left(Q_{\rho_i^\alpha,\rho_f^\alpha}\right)\right) \leqslant \int_\Gamma H(\rho_t)Q_{\rho_i^\alpha,\rho_f^\alpha}(\mathrm{d}\rho).$$

Integrating w.r.t. θ leads to the announced inequality. □

Part II

Harmonic mappings valued in the Wasserstein space

Chapter 7

Introduction to harmonic mappings valued in the Wasserstein space

The goal of this chapter is to introduce and motivate the notion of harmonic mappings valued in the Wassertein space, and to give a brief overview of the rest of this part. Compared to the previous part, Ω will be the source space which is not necessarily assumed to be convex, while D will be the convex domain over which the Wasserstein space $\mathcal{P}(D)$ is defined.

Namely, throughout this whole part we take $p, q \geqslant 1$ some integers and we make the following assumptions.

Assumptions. *We assume that Ω is a connected compact subset of $\mathbb{R}^p$. Moreover, $\partial\Omega$ is assumed to be Lipschitz, which means that around any point of $\partial\Omega$, up to a rotation, Ω is the epigraph of a Lipschitz function. The Lebesgue measure of Ω is assumed to be* 1.

We assume that D is a convex compact subset of $\mathbb{R}^q$.

As a general rule, Greek letters will be associated to objects related to Ω, while Latin ones will be for objects related to D. For instance, generic points in Ω (resp. D) will usually be denoted by ξ, η (resp. x, y); and derivatives w.r.t. variables in Ω (resp. D) will be denoted by $(\partial_\alpha)_{1 \leqslant \alpha \leqslant p}$ (resp. $(\partial_i)_{1 \leqslant i \leqslant q}$). The notation $\mathcal{L}_\Omega$ (resp. $\mathcal{L}_D$) will stand for the Lebesgue measure restricted to Ω (resp. D). Notice that by assumption $\mathcal{L}_\Omega \in \mathcal{P}(\Omega)$.

7.1 Harmonic mappings

If $f : \Omega \to \mathbb{R}$ is a real-valued function defined on a subset Ω of $\mathbb{R}^p$, one says that f is *harmonic* if

$$\Delta f = 0, \tag{7.1}$$

where $\Delta = \sum_{\alpha=1}^p \partial_{\alpha\alpha}$ denotes the Laplacian operator. Although this equation can be traced back to physics (for instance it corresponds to the equation satisfied by the electric potential in the absence of charge, or the one satisfied by the temperature in some homogeneous and isotropic medium when the permanent regime is reached), it has revealed to have its own mathematical interest [HW08]. In particular it is associated to a concept of *equilibrium*, as for an harmonic function f, the value of f at a point $\xi \in \Omega$ is always equal to the mean of the values of f on a ball centered at ξ. A whole line of research has been devoted to define harmonic mappings $f : X \to Y$ where X and Y are spaces without a structure as strong as the Euclidean one. If X and Y are Riemannian manifolds, one can define an analogue of (7.1) which involves the metric tensors of both X and Y (see for instance [ES64] or, for a modern presentation, [Jos08, HW08]).

The standard assumption to get existence results and nice properties of harmonic mappings is that X has a positive curvature and Y has a negative curvature. In the 90s, Korevaar and Schoen [KS93] on one side and Jost [Jos94] on the other side, presented independently a more general setting and showed that one can define harmonic mappings $f : \Omega \to Y$ provided that Ω is a compact Riemannian manifold (in fact a more general object in Jost's work) and Y is a metric space with negative curvature in the sense of Alexandrov [KS93, Section 2.1].

The most robust point of view for the definition of harmonic mappings valued in metric spaces is related to the Dirichlet problem. Indeed, if we go back to the case where $Y = \mathbb{R}$, a function $f : \Omega \to \mathbb{R}$ is harmonic if and only if it is a minimizer of the Dirichlet energy

$$\mathrm{Dir}(g) := \int_\Omega \frac{1}{2} |\nabla g(\xi)|^2 \mathrm{d}\xi$$

among all functions $g : \Omega \to \mathbb{R}$ having the same values as f on $\partial\Omega$ the boundary of Ω. The main advantage of this formulation is that it involves only first order derivatives, and most of the concepts involving first order derivatives can be defined on metric spaces even without any vectorial structure [AT03]. Korevaar, Schoen and Jost proved that for every separable metric space Y, one can define the analogue of the Dirichlet energy of any mapping $f : \Omega \to Y$. Then, under the assumption that Y has a negative curvature in the sense of Alexandrov, they proved existence and uniqueness of a minimizer of the Dirichlet energy (provided that the values at the boundary $\partial\Omega$ are fixed), interior and boundary regularity of the minimizer and lots of other properties similar to harmonic mappings between manifolds. Most of the proofs mimic the ones in the Euclidean case and rely only on the curvature properties of the target space Y. To quote Korevaar and Schoen: "*We find the generality, elegance, and simplicity of the proofs presented here to be an indication that we have found the proper framework for their expression*" [KS93, p. 614].

In this part, our goal is to define and to study harmonic mappings defined over a compact domain Ω of $\mathbb{R}^p$ and valued in the space of probability measures over a convex domain D of $\mathbb{R}^q$ endowed with the quadratic Wasserstein distance W_2. We will define the Dirichlet energy for mappings $\boldsymbol{\mu} : \Omega \to (\mathcal{P}(D), W_2)$ and study its minimizers under the constraint that the values at the boundary $\partial\Omega$ are fixed. It is known that $(\mathcal{P}(D), W_2)$ is a positively curved space in the sense of Alexandrov [AGS08, Section 7.3], hence the whole theory of Korevaar, Schoen and Jost does not apply: we have to leave the world of "*generality, elegance and simplicity*". Though we manage to develop a fairly satisfying theory of Dirichlet energy and harmonic mappings valued in the Wasserstein space, it is *ad hoc*: it intensively relies on specific properties of $(\mathcal{P}(D), W_2)$ and is hardly generalizable to other positively curved spaces. We have already presented in the introduction of this manuscript, in Figure 1.4, an example of what these harmonic mappings look like.

7.2 Related works

This work can be seen as an extension of an article written by Brenier [Bre03] almost 15 years ago. Recently, few articles [SNB$^+$12, SGB13, SRGB14, VL18, Lu17] have been published on related topics even though none of them seems aware of Brenier's work.

In Section 3 of [Bre03], Brenier proposed a definition of what he called *generalized harmonic functions* which is the same thing as our harmonic mappings valued in the Wasserstein space. He defined the Dirichlet energy for such mappings; proved the existence of harmonic mappings in some special cases and gave an explicit solution in the very special case where all measures on $\partial\Omega$ are Dirac masses; indicated the formulation of the dual problem; and formulated some conjectures. In the present work, we will rely on the same definition of Dirichlet energy as in Brenier's article, but we push the analysis much further: we provide a rigorous functional analysis framework; link the Dirichlet energy with already known notions of analysis in metric spaces (in particular with the definition of Korevaar, Schoen and Jost); prove the existence of harmonic mappings in a more general context; and answer Brenier's conjectures.

In [SNB$^+$12, SGB13], the authors studied *soft maps* (which are nothing more than maps $\Omega \to \mathcal{P}(D)$ except that Ω and D are surfaces, i.e. Riemannian manifolds of dimension 2) and define a Dirichlet energy in the same way as Korevaar, Schoen and Jost. These maps are seen as relaxations of "classical" maps $\Omega \to D$, and they focus on numerical computation and visualization of theses soft maps, see also [SRGB14] for applications to supervised learning. On the other hand, they do not analyze in detail the theoretical properties of the Dirichlet energy and harmonic mappings, which in contrast is the main topic of the present work. In [Lu17], the author provides some theoretical analysis of soft maps by focusing on the cases where the boundary measures on $\partial\Omega$ are either Dirac masses or Gaussian measures. He uses only the metric definition of the Dirichlet energy, i.e. the one of Korevaar, Schoen and Jost.

Finally, in [VL18] the authors also study mappings valued in the space of probability measures, but are rather interested in the bounded variation norm (the integral of the norm of the gradient) than in the Dirichlet energy. Their provide applications to the denoising of measure-valued images.

Apart from these articles, let us underline the interest of our work by relating it to other already known concepts:

- It is well known that harmonic mappings defined over an interval of $\mathbb{R}$ and valued in a geodesic space are precisely the constant-speed geodesics, and it is the case with our definition. Thus our work can be seen as extending the definition of geodesics in the Wasserstein space, the latter being an object which is now well understood.

- As we said above, our definition of Dirichlet energy coincides with the one of Korevaar, Schoen and Jost. In particular, our work shows that their definition can be applied to positively curved spaces and still get some non trivial result, even though we rely on the very special structure of the Wasserstein space.

- In connection to soft maps, Justin Solomon and co-authors have introduced the concept of *Wasserstein propagation* [SRGB14]. They take a finite graph (V, E) with positive weights $(\omega_e)_{e \in E}$ on the edges. If $\boldsymbol{\mu} : V \to \mathcal{P}(D)$ is a mapping defined over the vertices the graph and valued in the Wasserstein space, its Dirichlet energy is defined as

$$\mathrm{Dir}(\boldsymbol{\mu}) := \sum_{e=(v,w) \in E} \omega_e \frac{W_2^2(\boldsymbol{\mu}(v), \boldsymbol{\mu}(w))}{2}.$$

  It could be seen, at least formally, as an analogue as the Dirichlet energy defined in this work when the source space is discrete. Then they assume that they have a distinguished subset $V_0 \subset V$ of the set of vertices, thought as the boundary of the graph. The Wasserstein propagation problem amounts to find a mapping minimizing the Dirichlet energy among all mappings having given values on V_0. Indeed, the (Wasserstein-valued) labels on V_0 are *propagated* to the rest of the graph. Already in [SRGB14], or for instance in the recent article [GAHE18], the link between this problem and statistical questions is raised.

- To study the regularity of minimal surfaces, Almgren proposed the notion of Q-valued functions (see [AJ00] or [DLS11] for a clear and self-contained reference), which can be seen (up to renormalization) as mappings defined on $\Omega \subset \mathbb{R}^p$ and valued in the subset $\mathcal{A}_Q(D)$ (where $Q \geqslant 1$ is an integer) of the Wasserstein space $(\mathcal{P}(D), W_2)$ defined as

$$\mathcal{A}_Q(D) := \left\{ \frac{1}{Q} \sum_{i=1}^{Q} \delta_{x_i} \ : \ (x_1, x_2, \dots, x_Q) \in D^Q \right\}.$$

In other words, $\mathcal{A}_Q(D)$ is the set of probability measures which are combinations of at most Q Dirac masses with weights which are multiples of $1/Q$, and is endowed with the Wasserstein distance W_2. To put it shortly, a Q-function is a function which in every point takes Q unordered different values (counted with multiplicity). There exists a beautiful existence and regularity theory for harmonic Q-functions. As $\bigcup_{Q \geqslant 1} \mathcal{A}_Q(D)$ is dense in $\mathcal{P}(D)$, it would be tempting to see the Dirichlet problem for mappings valued in the Wasserstein space $\mathcal{P}(D)$ as the limit as $Q \to +\infty$ of the Dirichlet problem for Q-functions. However, it is not so obvious that this limit really holds, and most of the results in the theory of Q-functions are proved by induction on Q through clever decompositions and combinatorial arguments, hence they depend heavily on Q and not much can be passed to the limit $Q \to +\infty$. Notice that the space $\mathcal{A}_Q(D)$ is also positively curved in the sense of Alexandrov (the example in [AGS08, Section 7.3] lives in $\mathcal{A}_2(D)$), hence the theory of Q-functions is a theory of harmonic mappings valued in a positively curved space. However, it is known that $\mathcal{A}_Q(D)$ is in a bilipschitz bijection with a subset of $\mathbb{R}^N$ for some large N [DLS11, Theorem 2.1]: with Q-functions we stay in the finite-dimensional world. On the contrary, in the present article, the target space $(\mathcal{P}(D), W_2)$ will be both positively curved and genuinely infinite-dimensional.

On the other hand, to avoid confusion, let us mention briefly about some works that are *not* really related to the present one.

There has been a lot of works recently about analysis on non smooth spaces using optimal transport as a central tool (see for instance [Gig15] and references therein); and also some works defining Hamilton-Jacobi equations on the Wasserstein space [GNT08, GŚ15] in link with the so-called *master equation* in Mean Field Games [CDLL15].

In these works, one studies mappings which are defined over a non smooth space (a RCD(K, N) one in the first case, the Wasserstein space in the second) but which are valued in $\mathbb{R}$. All the issues (and interesting questions) come from the lack of smoothness of the source space. In the present work, we study mappings which are defined over a smooth space, but valued in a non smooth one, namely the Wasserstein space. At some point, it might be possible to look at mappings defined over a non smooth space and valued in the Wasserstein space but, as the reader will see in the sequel, there is already some work to do when the source space is smooth.

7.3 Main definitions and results

Let us go into the details and summarize the content of this part as well as the key insights. In this discussion we will stay informal, with sometimes sloppy or non rigorous statements.

Dirichlet energy and Dirichlet problem Chapter 8 is concerned with the definition of the Dirichlet energy and the Dirichlet problem.

More specifically, Section 8.1 is devoted to the different definitions of the Dirichlet energy of a mapping $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$, the equivalence between these definitions and some properties of this Dirichlet energy

The idea is to start from curves valued in the Wasserstein space and the so-called Benamou-Brenier formula [BB00]. If I is a segment of $\mathbb{R}$ and $\boldsymbol{\mu} : I \to \mathcal{P}(D)$ is an absolutely continuous curve, then its Dirichlet energy is nothing else than its action defined by (see Theorem 2.8)

$$\mathrm{Dir}(\boldsymbol{\mu}) = \inf_{\mathbf{v}} \left\{ \int_I \left(\int_D \frac{1}{2} |\mathbf{v}(t,x)|^2 \boldsymbol{\mu}(t, \mathrm{d}x) \right) \mathrm{d}t \ : \ \mathbf{v} : I \times D \to \mathbb{R}^q \ \text{ and } \partial_t \boldsymbol{\mu} + \nabla \cdot (\boldsymbol{\mu}\mathbf{v}) = 0 \right\},$$

which means that one minimizes the integral over time of the kinetic energy among all velocity fields $\mathbf{v}$ such that the continuity equation $\partial_t \boldsymbol{\mu} + \nabla \cdot (\boldsymbol{\mu}\mathbf{v}) = 0$ is satisfied. This continuity equation is supplemented with the non-flux condition $\nabla(\boldsymbol{\mu}\mathbf{v}) \cdot \mathbf{n}_D = 0$ on ∂D to ensure preservation of mass. What Benamou and Brenier understood is that the correct variable is the momentum $\mathbf{E} = \mathbf{v}\boldsymbol{\mu}$[1]. Indeed, the continuity equation $\partial_t \boldsymbol{\mu} + \nabla \cdot \mathbf{E} = 0$ becomes a linear constraint and

$$\int_I \left(\int_D \frac{1}{2} |\mathbf{v}(t,x)|^2 \boldsymbol{\mu}(t, \mathrm{d}x) \right) \mathrm{d}t = \iint_{I \times D} \frac{|\mathbf{E}|^2}{2\boldsymbol{\mu}}$$

is a convex function of the pair $(\boldsymbol{\mu}, \mathbf{E})$. In particular, to find the constant-speed geodesic between μ and $\nu \in \mathcal{P}(D)$, assuming that $I = [0,1]$, one minimizes the convex Dirichlet energy over the pairs $(\boldsymbol{\mu}, \mathbf{E})$ with linear constraints given by the continuity equation, that $\boldsymbol{\mu}(0) = \mu$ and that $\boldsymbol{\mu}(1) = \nu$.

As noticed in [Bre03, Section 3], this formulation can be directly extended to the case where the source space is no longer of dimension 1: if Ω is a subset of $\mathbb{R}^p$, one can define a (generalized) continuity equation for the pair $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ and $\mathbf{E} : \Omega \times D \to \mathbb{R}^{pq}$ by

$$\nabla_\Omega \boldsymbol{\mu} + \nabla_D \cdot \mathbf{E} = 0, \tag{7.2}$$

where ∇_Ω stands for the gradient w.r.t. variables in Ω and $\nabla_D \cdot$ stands for the divergence w.r.t. variables in D. Notice that if $\mathbf{E}$ is thought as a matrix-valued measure, its dimension is the same as the Jacobian of a map defined on Ω and valued in D. More precisely if $(\mathbf{E}^{\alpha i})_{1 \leqslant \alpha \leqslant p, 1 \leqslant i \leqslant q}$ denote the components of $\mathbf{E}$, and if the derivatives w.r.t. variables in Ω (resp. D) are denoted by $(\partial_\alpha)_{1 \leqslant \alpha \leqslant p}$ (resp. $(\partial_i)_{1 \leqslant i \leqslant q}$) then the the continuity equation reads: for any $\alpha \in \{1, 2, \ldots, p\}$,

$$\partial_\alpha \boldsymbol{\mu} + \sum_{i=1}^{q} \partial_i \mathbf{E}^{\alpha i} = 0.$$

The Dirichlet energy of the pair $(\boldsymbol{\mu}, \mathbf{E})$ is defined as

$$\iint_{\Omega \times D} \frac{|\mathbf{E}|^2}{2\boldsymbol{\mu}} = \iint_{\Omega \times D} \sum_{\alpha=1}^{p} \sum_{i=1}^{q} \frac{|\mathbf{E}^{i\alpha}|^2}{2\boldsymbol{\mu}},$$

and $\mathrm{Dir}(\boldsymbol{\mu})$, the Dirichlet energy of $\boldsymbol{\mu}$, is the minimal Dirichlet energy of the pairs $(\boldsymbol{\mu}, \mathbf{E})$ among all $\mathbf{E}$ such that the continuity equation is satisfied (Definition 8.7). It is a straightforward copy of the classical proofs of optimal transport to show that there exists a unique optimal momentum $\mathbf{E}$ (which we call the tangent momentum) which is written $\mathbf{E} = \mathbf{v}\boldsymbol{\mu}$ for some velocity field $\mathbf{v} : \Omega \times D \to \mathbb{R}^{pq}$, and that Dir is convex and lower semi-continuous.

[1]The notation $\mathbf{E}$ can look unusual for a momentum. We have taken it from [San15] while the author of this book found it in the works of Brenier [Bre03]. When asked, Brenier answered that he introduced this notation in [Bre01], where the momentum coming from optimal transport was representing an electric field, while there was a "true" momentum $\mathbf{j}$ representing a courant density. To make sure there was no confusion, Brenier chose to use E for the momentum of optimal transport, and J for the current density. For very contingent reasons, it seems that this notation has been perpetuated ever since.

We will prove that for $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$, one has $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ if and only if for any $u : \mathcal{P}(D) \to \mathbb{R}$ which is 1-Lipschitz, one has that $u \circ \boldsymbol{\mu}$ belongs to $H^1(\Omega)$ with $|\nabla(u \circ \boldsymbol{\mu})| \leqslant g$, where $g \in L^2(\Omega)$ is independent of u. Moreover, the minimal g will be shown to be controlled from above and below by

$$\sqrt{\int_D |\mathbf{v}(\cdot, x)|^2 \boldsymbol{\mu}(\cdot, \mathrm{d}x)} \in L^2(\Omega),$$

where $\mathbf{E} = \mathbf{v}\boldsymbol{\mu}$ is the tangent momentum (Theorem 8.20). This precisley shows that the space $\{\boldsymbol{\mu} : \Omega \to \mathcal{P}(D) \ : \ \mathrm{Dir}(\boldsymbol{\mu}) < +\infty\}$ coincides with the set $H^1(\Omega, \mathcal{P}(D))$, where the latter is defined in the sense of Reshetnyak [Res97], and that the gradient of $\boldsymbol{\mu}$ in the sense of Reshetnyak (the minimal g above) is related to the tangent velocity field $\mathbf{v}$. The Dirichlet energy is not equal to the L^2 norm of g, as it is already the case in the classical framework [Chi07]: if we see $\mathbf{v} : \Omega \times D \to \mathbb{R}^{pq}$ as a matrix-valued field, the Benamou-Brenier definition measures the magnitude of $\mathbf{v}$ with the Hilbert-Schmidt norm, whereas the optimal g from the definition of Reshetnyak is rather related to the operator norm of the matrices. Nevertheless, it implies that Lipschitz mappings $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ (i.e. such that $W_2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta)) \leqslant C|\xi - \eta|$ for any $\xi, \eta \in \Omega$) have a finite Dirichlet energy.

We will also prove that our Dirichlet energy coincides with the one of Korevaar and Schoen, as well as Jost. The definitions of these authors can appear slightly different though they turn out to be equivalent, see [Chi07]. Their idea goes as follows: if $f : \Omega \to \mathbb{R}$ is smooth, then for any $\xi \in \mathbb{R}^p$,

$$|\nabla f(\xi)|^2 = \lim_{\varepsilon \to 0} C_p \int_{B(\xi,\varepsilon)} \frac{|f(\eta) - f(\xi)|^2}{\varepsilon^{p+2}} \mathrm{d}\eta,$$

for some constant C_p which depends on p the dimension of Ω, where $B(\xi, \varepsilon)$ is the ball of center ξ and radius ε. Thus, if $\varepsilon > 0$ is small, a good approximation of the Dirichlet energy of f would be

$$\mathrm{Dir}(f) = \int_\Omega \frac{1}{2} |\nabla f(\xi)|^2 \mathrm{d}\xi \simeq C_p \iint_{\Omega \times \Omega} \frac{|f(\xi) - f(\eta)|^2}{2\varepsilon^{p+2}} \mathbb{1}_{|\xi - \eta| \leqslant \varepsilon} \mathrm{d}\xi \mathrm{d}\eta.$$

Notice that the right hand side involves only metric quantities, thus its definition can be extended if $f : \Omega \to Y$ where (Y, d) is an arbitrary metric space by replacing $|f(\xi) - f(\eta)|^2$ by $d(f(\xi), f(\eta))^2$: this is what is done and extensively studied in [KS93, Section 1] (curvature assumptions on Y are not required for the definition of the Dirichlet energy, but are used to derive existence, uniqueness and properties of the minimizers). The counterpart in our case is to define the ε-Dirichlet energy of a mapping $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ by

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) := C_p \iint_{\Omega \times \Omega} \frac{W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta))}{2\varepsilon^{p+2}} \mathbb{1}_{|\xi - \eta| \leqslant \varepsilon} \mathrm{d}\xi \mathrm{d}\eta.$$

We are able show that Dir_ε converges to Dir as $\varepsilon \to 0$: it holds pointwisely but also in the sense of Γ-convergence (Theorem 8.26). For both the equivalence with the definition of Korevaar, Schoen and Jost, or with the one of Reshetnyak, the difficulty is not to guess them (they are fairly simple at the formal level) but to conduct careful approximation arguments.

We will show how one can define values on $\partial\Omega$ for mappings $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ with finite Dirichlet energy. There already exists a trace theory in [KS93], however in view of the dual formulation for the Dirichlet problem, we prefer to define trace values by extending the continuity equation up to the boundary of Ω. Indeed, multiplying (7.2) by a test function $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$

valued in $\mathbb{R}^p$, we get the following weak formulation:

$$\iint_{\Omega\times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint_{\Omega\times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E} = \int_{\partial\Omega} \left(\int_D \varphi(\xi, x) \cdot \mathbf{n}_\Omega(\xi) \boldsymbol{\mu}(\xi, \mathrm{d}x) \right) \sigma(\mathrm{d}\xi),$$

where $\mathbf{n}_\Omega$ is the outward normal to $\partial\Omega$ and σ the surface measure. We will show that, if $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$, then the r.h.s. can always be defined as a finite vector-valued measure acting on φ called $\mathrm{BT}_{\boldsymbol{\mu}}$ (Theorem 8.27). Two mappings will have the same values on the boundary $\partial\Omega$ if, by definition, they have the same boundary term.

In Section 8.2 we define the Dirichlet problem and establish its dual formulation. This is fairly classic in optimal transport theory, our proofs do not bring any new ideas.

To define the Dirichlet problem, we assume that a mapping $\boldsymbol{\mu}_b : \Omega \to \mathcal{P}(D)$ with finite Dirichlet energy is given and we study

$$\min_{\boldsymbol{\mu}} \{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} = \boldsymbol{\mu}_b \text{ on } \partial\Omega \}.$$

Thanks to the Benamou-Brenier formulation, existence of a solution is a straightforward application of the direct method of calculus of variations (Theorem 8.32). As we discuss it in Chapter 12, we do not know if uniqueness holds. Only in some particular case where the boundary values belong to a family of elliptically contoured distributions, we are able to prove uniqueness.

In the formulation of the Dirichlet problem, we define the boundary conditions through a mapping $\boldsymbol{\mu}_b$ defined on the whole Ω. A natural question arises: if $\boldsymbol{\mu}_b : \partial\Omega \to \mathcal{P}(D)$ is given, is it possible to extend it on Ω in such a way that $\mathrm{Dir}(\boldsymbol{\mu}_b) < +\infty$? We will show that the answer to this question is positive if $\boldsymbol{\mu}_b$ is Lipschitz on $\partial\Omega$, indeed in this case one can extend it as a Lipschitz mapping on Ω. The question of the existence of a Lipschitz extension for mappings $f : Z \to Y$, where $Z \subset X$ and X, Y are metric spaces has been intensively studied, see for instance [LS97, Oht09] and references therein. The general philosophy is that lower bounds on the curvature are required for the source space X, whereas upper bounds on the curvature are required for the target space Y. In our case, there are no upper bounds for the curvature of the target space $\mathcal{P}(D)$, hence we cannot apply classical results. However, we use the fact that we want to extend Lipschitz mappings defined not on an arbitrary closed subset of Ω, but on the boundary $\partial\Omega$ which has some regularity. By some *ad hoc* construction, we are able to treat the case where Ω is a ball, but we cannot control the Lipschitz constant of the extension on Ω by the Lipschitz constant of the mapping on $\partial\Omega$. Nevertheless, we can conclude for smooth domains, as they can be cut in a finite number of pieces, each piece being in a bilipschitz bijection with a ball (Theorem 8.33).

Let us establish here the dual formulation *via* a formal $\inf - \sup$ exchange, it was already done in [Bre03]. Indeed, given the definition of Dir and the weak formulation of the continuity equation,

$$\begin{aligned}
&\min_{\boldsymbol{\mu}} \{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} = \boldsymbol{\mu}_b \text{ on } \partial\Omega \} \\
&\quad = \inf_{\boldsymbol{\mu}, \mathbf{v}} \left[\iint_{\Omega\times D} \frac{1}{2} |\mathbf{v}|^2 \boldsymbol{\mu} + \sup_{\varphi \in C^1(\Omega\times D, \mathbb{R}^p)} \left(\mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) - \iint_{\Omega\times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} - \iint_{\Omega\times D} \nabla_D \varphi \cdot \mathbf{v}\boldsymbol{\mu} \right) \right] \\
&\quad = \sup_{\varphi \in C^1(\Omega\times D, \mathbb{R}^p)} \left[\mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) + \inf_{\boldsymbol{\mu}, \mathbf{v}} \iint_{\Omega\times D} \left(\frac{1}{2} |\mathbf{v}|^2 - \nabla_D \varphi \cdot \mathbf{v} - \nabla_\Omega \cdot \varphi \right) \boldsymbol{\mu} \right].
\end{aligned}$$

Optimizing in $\mathbf{v}$, we have that $\mathbf{v} = \nabla_D \varphi$, and then the infimum in $\boldsymbol{\mu}$ is translated into the constraint $\nabla_\Omega \cdot \varphi + \frac{1}{2}|\nabla_D \varphi|^2 \leqslant 0$. Hence, we have (formally, and it is proved rigorously in the core of the article, see Theorem 8.36) the following identity:

$$\sup_{\varphi} \left\{ \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) \ : \ \varphi \in C^1(\Omega \times D, \mathbb{R}^p) \ \text{ and } \ \nabla_\Omega \cdot \varphi + \frac{|\nabla_D \varphi|^2}{2} \leqslant 0 \right\}$$
$$= \min_{\boldsymbol{\mu}} \{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} = \boldsymbol{\mu}_b \text{ on } \partial\Omega \}.$$

We do not have an existence result for solutions φ of the dual problem. Notice that φ is a vector-valued function, but there is only a scalar constraint on it: the dual problem looks harder than in the case where Ω is a segment of $\mathbb{R}$. We are aware that "multitime Hamilton Jacobi equations" have been studied [LR86] but the setting is different: in the latter case, one has a scalar unknown which is submitted to as many equations as there are of "temporal" dimension.

Formally, as it is done in [Bre03], one can get optimality conditions out of the dual formulation. Indeed, we have that $\mathbf{v} = \nabla_D \varphi$ and, from the optimization in $\boldsymbol{\mu}$, that $\nabla_\Omega \cdot \varphi + \frac{1}{2}|\nabla_D \varphi|^2 = 0$ $\boldsymbol{\mu}$-a.e. If we assume that $\boldsymbol{\mu}$ is strictly positive a.e., we end up with the following system for $\mathbf{v}$ (the first equation is just a rewriting of the fact that $\mathbf{v}$ is a gradient, the second one is obtained by differentiating $\nabla_\Omega \cdot \varphi + \frac{1}{2}|\nabla_D \varphi|^2 = 0$ w.r.t. D):

$$\begin{cases} \partial_i \mathbf{v}^{\alpha j} = \partial_j \mathbf{v}^{\alpha i} & \text{for } \alpha \in \{1, 2, \ldots, p\} \text{ and } i, j \in \{1, 2, \ldots, q\}, \\ \displaystyle\sum_{\alpha=1}^{p} \partial_\alpha \mathbf{v}^{\alpha i} + \sum_{\alpha=1}^{p} \sum_{j=1}^{q} \mathbf{v}^{\alpha j} \partial_j \mathbf{v}^{\alpha i} = 0 & \text{for } i \in \{1, 2, \ldots, q\}. \end{cases} \tag{7.3}$$

However, we will not push the analysis further and try to derive a rigorous version of theses optimality conditions, it might be the topic of an other study. As the reader can see in the sequel, even without them, we can already say a lot.

In Section 8.3, we answer to a problem formulated by Brenier [Bre03, Problem 3.1]. The question is the following: if $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$, does there exists a probability Q over functions $f : \Omega \to D$ such that $\boldsymbol{\mu}$ is represented by Q, i.e.

$$\int_D a(x) \boldsymbol{\mu}(\xi, \mathrm{d}x) = \int a(f(\xi)) Q(\mathrm{d}f)$$

for all $a \in C(D)$ continuous and $\xi \in \Omega$; and such that the Dirichlet energy is the mean of the Dirichlet energy of the f:

$$\mathrm{Dir}(\boldsymbol{\mu}) = \int \left(\int_\Omega \frac{1}{2} |\nabla f(\xi)|^2 \mathrm{d}\xi \right) Q(\mathrm{d}f)?$$

If Ω is a segment of $\mathbb{R}$ the answer is positive as shown in [AGS08, Section 8.2]: it is known as the probabilistic representation or the superposition principle. However, as soon as Ω is two or more dimensional (in fact it already fails if Ω is a circle), the answer becomes negative (Proposition 8.41). We will provide a counterexample and explain the obstruction.

The main consequence is the following: there is no Lagrangian formulation for mappings $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$. There can be no static formulation of the Dirichlet problem analogue to transport plans or multimarginal formulation. One is forced to work only with the Eulerian formulation, namely the Benamou-Brenier formula. It explains why it is substantially more difficult to study mappings $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ as soon as the dimension of Ω is larger than 2, as most of the difficult results of optimal transport are proved thanks to the Lagrangian point of view.

Maximum principle In Chapter 9, we prove a maximum principle (more specifically a Ishihara-type property) for harmonic mappings, meaning roughly speaking that *harmonic mappings reach their maximum on the boundary of the domain* Ω. Of course, there is no canonical order on the Wasserstein space, thus this assertion does not really make sense: only the composition of a (real-valued) geodesically convex function over $\mathcal{P}(D)$ with an harmonic mapping will satisfy the maximum principle.

If $f : \Omega \to \mathbb{R}$ is a real valued harmonic function, then $(F \circ f) : \Omega \to \mathbb{R}$ is a subharmonic function for every $F : D \to \mathbb{R}$ convex, which means that $\Delta(F \circ f) \geqslant 0$. It can be checked by a direct computation using the chain rule. If we take $f : X \to Y$, where X and Y are two Riemannian manifolds, then the result still holds (provided that harmonicity, subharmonicity and convexity are properly defined through the Riemannian structures) and it is even a characterization of harmonic mappings: this was first remark by Ishihara [Ish78] (hence we will denote this assertion as a “Ishihara type property” rather than a maximum principle), one can find a statement and a proof in [Jos08, Corollary 8.2.4]. In short: once composed with a convex real-valued function, an harmonic mapping satisfies the maximum principle. Extensions of this result when the target is a metric space with negative curvature are available, see for instance [Stu05, Section 7].

In the Wassertein space, mappings which are convex w.r.t. the metric structure, which means convex along geodesics, are well understood. Actually, we will need something a little stronger, which is convexity along generalized geodesics (see Section 2.2) as it guarantees existence and uniqueness of its gradient flow. In our case the Ishihara property reads: if $F : \mathcal{P}(D) \to \mathbb{R}$ is convex along generalized geodesics and if $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ is a solution of the Dirichlet problem, then $(F \circ \boldsymbol{\mu}) : \Omega \to \mathbb{R}$ is subharmonic (Theorem 9.3). This can be considered as the main result of this part, and the proof bears many similarities with what is done in the first part of this manuscript.

The proof of geodesic convexity usually relies on the Lagrangian formulation, which, as we said above, is not available in our case. To overcome this difficulty, we use the approximate Dirichlet energies Dir_ε as a substitute for Dir. Indeed, as explained by Jost [Jos94], if $\boldsymbol{\mu}_\varepsilon$ is a minimizer of Dir_ε (with for instance fixed values around the boundary $\partial\Omega$), then for a.e. $\xi \in \Omega$, $\boldsymbol{\mu}_\varepsilon(\xi)$ is a minimizer of

$$\nu \mapsto \int_{B(\xi,\varepsilon)} W_2^2(\nu, \boldsymbol{\mu}_\varepsilon(\eta)) \mathrm{d}\eta,$$

in other words $\boldsymbol{\mu}_\varepsilon(\xi)$ is a barycenter of the $\boldsymbol{\mu}_\varepsilon(\eta)$, for $\eta \in B(\xi, \varepsilon)$. Notice that if $f : \Omega \to \mathbb{R}$ is real-valued and harmonic, then for any $\varepsilon > 0$ $f(\xi)$ is the barycenter of $f(\eta)$ for $\eta \in B(\xi, \varepsilon)$, while in the metric case this property only holds asymptotically as $\varepsilon \to 0$. For barycenters in the Wasserstein space, there exists a generalized Jensen inequality: it was already proved for the barycenter of a finite number of measures by Agueh and Carlier [AC11, Proposition 7.6] under the assumption that F is convex along generalized geodesics, and in a more general case (in particular with an infinite numbers of measures defined on a compact manifold, whereas Agueh and Carlier worked in the Euclidan space) by Kim and Pass [KP17, Section 7], but with rather strong regularity assumptions on the measures. As explained in the introduction of this manuscript, we provide a new proof of this Jensen inequality in a case adapted to our context by letting the barycenter $\boldsymbol{\mu}_\varepsilon(\xi)$ follow the gradient of the functional F and use the result as a competitor: through arguments first advanced in [MMS09] in a very different context under the name of *flow interchange*, one can show (estimating the derivative of the Wasserstein distance along the flow of F with the so-called (EVI) inequality) that for a.e. $\xi \in \Omega$

$$\int_{B(\xi,\varepsilon)} [F(\boldsymbol{\mu}_\varepsilon(\eta)) - F(\boldsymbol{\mu}_\varepsilon(\xi))] \mathrm{d}\eta \geqslant 0. \tag{7.4}$$

Then, as Dir_ε Γ-converges to Dir, one knows that $\boldsymbol{\mu}_\varepsilon$ converges to $\boldsymbol{\mu}$ a solution of the Dirichlet problem. Passing in the limit (7.4), one concludes that $(F \circ \boldsymbol{\mu})$ is subharmonic in the sense of distributions.

Let us make a few comments. The main drawback of the proof, as we proceed by approximation and that uniqueness in the Dirichlet problem is not known, is that we are only able to show subharmonicity of $F \circ \boldsymbol{\mu}$ for one solution of the Dirichlet problem (which moreover depends on F), and not for all. To overcome this limitation, the best thing to do would be to prove uniqueness in the Dirichlet problem. Let us also discuss the regularity that we need on F. Either we require F to be continuous (which is very restrictive: it excludes the internal energies); or, if F is only lower semi-continuous, we need F to be bounded on bounded subsets of $L^\infty(D) \cap \mathcal{P}(D)$ (which is not very restrictive), but we also need the weak lower semi-continuity of

$$\boldsymbol{\mu} \mapsto \int_\Omega F(\boldsymbol{\mu}(\xi))\mathrm{d}\xi.$$

More precisely, a mapping $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ can be seen as an element of $\mathcal{P}(\Omega \times D)$ (by "fubinization") and we require lower semi-continuity of $\boldsymbol{\mu} \mapsto \int_\Omega (F \circ \boldsymbol{\mu})$ w.r.t. the weak convergence on $\mathcal{P}(\Omega \times D)$. This weak lower semi-continuity holds heuristically if F is convex for the usual (and not geodesic) convexity on $\mathcal{P}(D)$. At the end of the day, the Ishihara property works for potential energies (for a convex, L^1 and lower semi-continuous potential), for internal energies (which have a super linear growth and satisfy McCann's conditions) and for the interaction energies (but only for a convex continuous interaction potential). Eventually, notice that we do not have the converse statement: we do not know if the fact that $F \circ \boldsymbol{\mu}$ is subharmonic for any F convex along generalized geodesics is enough to prove that $\boldsymbol{\mu}$ is harmonic. To prove such a result, one would need a better understanding of the optimality conditions of the Dirichlet problem.

Special case In Chapter 10 we provide specific situations where we can say more about harmonic mappings.

In Section 10.1, we briefly present the results of other people, namely Brenier [Bre03] and Lu [Lu17]. More precisely, we say what happens when the boundary data $\boldsymbol{\mu}_b$ is valued in the set of Dirac masses: the solution of the Dirichlet problem stays valued in this set. The shortest argument relies on the existence of a retraction onto the set of Dirac masses. As understood by Lu, this argument would also work if the space D over which the Wasserstein space is defined is replaced by a Riemannian manifold with *negative* curvature, while it fails for some *positively* curved manifolds.

Another simple situation, in Section 10.2, is the case where the set D, on which the target space $\mathcal{P}(D)$ is modeled, is a segment of $\mathbb{R}$. In this case, the Wasserstein space $(\mathcal{P}(D), W_2)$ is in an isometric bijection with a convex subset of the Hilbert space $L^2([0,1])$. Hence, the Dirichlet problem reduces to the study of the Dirichlet problem for mappings valued in a Hilbert space, which is more standard.

In Section 10.3 we provide an example where we can do explicit computations, namely when we restrict our attention to a family of *elliptically contoured distributions*. This terminology comes from [Gel90] and denotes a generalization of the family of Gaussian measures. In statistics this type of family is sometimes called a *location-scatter family*. More precisely, we take $\rho \in L^1(\mathbb{R}^q)$ a positive and compactly supported function such that the measure $\rho(x)\mathrm{d}x$ has a unit mass, zero mean, and the identity matrix as covariance matrix. The family of elliptically contoured distributions built on ρ is nothing else than the sets of measures obtained as image measures from $\rho(x)\mathrm{d}x$ by symmetric positive linear transformations. For instance, if ρ is the indicator function of a ball, the family of elliptically contoured distributions built on ρ consists in probability measures uniformly distributed on centered ellipsoids. In general the level sets of the density are ellipsoids, hence the terminology. The Gaussian case would be obtained by taking for $\rho(x)\mathrm{d}x$ a centered standard Gaussian, but this probability measure is not compactly supported (recall that we work in $\mathcal{P}(D)$ where $D \subset \mathbb{R}^q$ is compact). As in the Gaussian case, the elements of the family of elliptically contoured distributions are parametrized by their covariance matrix. Notice that it is already known that the geodesic between Gaussian measures and more generally the barycenter of Gaussian measures stay in the Gaussian family [AC11, Section 6.3]. If the boundary values $\boldsymbol{\mu}_b : \partial\Omega \to \mathcal{P}(D)$ are valued in a family of elliptically contoured distributions, we show that there exists at least one solution of the Dirichlet problem which takes values in the same family everywhere on Ω (Theorem 10.9): it relies on a simple argument, the existence of a *retraction* on the family of elliptically contoured distributions.

Under the additional assumption that the covariance matrices on the boundary $\partial\Omega$ are non singular we are able to show much more (Theorem 10.10). It implies that there is a solution of the Dirichlet problem with covariance matrices non singular everywhere in Ω: to prove it we use the maximum principle for the Boltzmann entropy, which translates in a minimum principle for the determinant of the covariance matrices. From this we are able to derive the Euler-Lagrange equation satisfied by the covariance matrix.

Moreover we can show the uniqueness of the solution to the Dirichlet problem among all competitors, not necessarily those valued in the family of elliptically contoured distributions. Let us give the structure of the proof as it is almost the only case where we know how to prove uniqueness. The observation is that all solutions of the Dirichlet problem must have the same tangent velocity field. Indeed, if φ is a solution of the dual problem, from optimality the tangent velocity field to any solution must be equal to $\nabla_D \varphi$. Now, if the velocity field $\nabla_D \varphi$ is regular enough (namely Lipschitz w.r.t. variables in D), then the solution of the (1-dimensional) continuity equation with velocity field $\nabla_D \varphi$ is unique. As the (generalized) continuity equation implies the 1-dimensional one, and as all solutions of the Dirichlet problem coincide on $\partial\Omega$ they must be equal everywhere. In the case of a family of elliptically contoured distributions the tangent velocity field is linear w.r.t. variables in D with some uniform bounds which allow us to make this argument rigorous.

Still under this additional assumption, we are also able to show the regularity of the minimizer: as the problem boils down to the study of Dirichlet minimizing mappings valued in a Riemannian manifold, the only thing to show, following the theory of Schoen and Uhlenbeck [SU82, SU83] is the absence of non-constant tangent minimizing mappings. We prove the latter property with the help of the maximum principle: even though the Wasserstein space is positively curved, there is a lot of functionals convex along geodesics defined on it.

In summary, under the assumption that the covariance matrices on the boundary $\partial\Omega$ are non singular we are able to give a full solution to the problem: existence, uniqueness, regularity and Euler-Lagrange equation.

In Section 10.4, we give an example of an harmonic mapping valued in a family of elliptically contoured distributions for which we have enough symmetry to give an almost explicit formula: at this point, it boils down to solve a 1-dimensional problem of calculus of variations for a curve valued in $\mathbb{R}^2$. The interest of this example is that, despite its simplicity, its encodes some characteristic features of the geometry of the Wasserstein space. Indeed, on this example, we know that the superposition principle must fail. Moreover, we are able to show that $\boldsymbol{\mu}(\xi)$ can *not* be written as the (weighted) barycenter of the $\boldsymbol{\mu}(\eta)$ for $\eta \in \partial\Omega$. In other words, for harmonic mappings valued in the Wasserstein space, there is no hope for a Green formula to be true, i.e. to express in a simple way the value at one point as a barycenter of the values at the boundary.

Numerical illustrations In Chapter 11, we describe the method that we use to compute harmonic mappings valued in the Wasserstein space. As there is no Lagrangian point of view nor static formulation, the Benamou-Brenier formulation appears to be the most adapted to tackle numerics. Indeed, this formulation can be read as a convex optimization problem with a linear constraint involving differential operator.

We have started form the dual formulation of the Dirichlet problem, and, inspired by ideas from [PPO14] about the use of staggered grid, we provide a finite difference formulation of the dual problem. We can prove existence of a solution to this discrete (i.e. finite dimensional) dual problem, however this proof is very specific to the finite-dimensional case and cannot be generalized to the continuous case. Then, taking the dual of the dual discrete problem, we obtain a (primal) discrete problem which looks like the continuous one. In short: we obtain two discrete (i.e. finite dimensional) convex optimization problems in duality, which mimic the primal and dual Benamou-Brenier formulations of the Dirichlet problem. However, as discussed more in details in Chapter 12, we do not have a proof of convergence if we refine the discretization.

Then, to solve efficiently these problems, as in the original paper by Benamou and Brenier [BB00], we use the iterative method called *Alternating Direction Method of Multipliers* (ADMM). Our unknowns live on a discretization of the space $\Omega \times D$, which is typically a space of dimension 4. However, the only non local step of each ADMM iteration is the resolution of a Poisson problem on $\Omega \times D$, for which we use leverage our Cartesian discretization and use FFT. However, due to the lack of strict convexity of the Dirichlet energy, the number of ADMM iterations required is important and the method is quite slow.

Although we have no guarantee that our method indeed computes an approximation of harmonic mappings, we show on some examples that it gives plausible mappings. An example of the output of our algorithm has already been presented in the Introduction, with Figure 1.4.

Perspectives and open questions In Chapter 12, we present some problems that are very natural but still left unanswered, and we explain the obstructions to our current attempts of proof.

To end this introduction, let us comment the somehow restrictive framework that we have chosen. The compactness assumption of Ω and D allows to simplify proofs by avoiding tails estimates: we believe that there is enough technical difficulties and non trivial statements even in this case, and that the key features of the Dirichlet problem are captured, which is the reason why we have restricted ourselves to the compact case. Although we have stuck to the Euclidean case, we see no deep reason which would prevent our definitions and results to be applied to the case where Ω and D are compact Riemannian manifolds. In particular, our regularization procedures rely on heat flows which are available in Riemannian manifolds. Finally, we have stick to the quadratic Wasserstein distance. We believe that if $p \in (1, +\infty)$ is given, the machinery that we use can be adapted in a straightforward way to define

$$\int_\Omega \frac{1}{p} |\nabla \boldsymbol{\mu}|^p,$$

where $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ but $\mathcal{P}(D)$ is endowed with the p-Wassertsein distance. However the Ishihiara type property is related to the Riemannian framework; also the explicit computations in the case of a family of elliptically contoured distributions are no longer avalaible. As mentionned above, the case $p = 1$, which corresponds the *total variation* of $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ (where $\mathcal{P}(D)$ is equipped with the 1-Wasserstein distance), has been defined and studied very recently [VL18] in the context of image denoising.

Chapter 8

The Dirichlet energy and the Dirichet problem

In this chapter, we define the Dirichlet energy of a mapping $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ following the idea of [Bre03, Section 3]. We relate the space of $\boldsymbol{\mu}$ with finite Dirichlet energy with $H^1(\Omega, \mathcal{P}(D))$ using the theory of Sobolev spaces valued into metric spaces of Reshetnyak [Res97, Res04], and we also prove that this Dirichlet energy coincides with the limit of ε-Dirichlet energies introduced by Korevaar, Schoen and Jost [KS93, Jos94].

Let us first define the space $L^2(\Omega, \mathcal{P}(D))$. As $\mathcal{P}(D)$ is bounded, it coincides with the measurable mappings valued in $\mathcal{P}(D)$.

Definition 8.1. *We denote by $L^2(\Omega, \mathcal{P}(D))$ the quotient space of measurable mappings $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ by the equivalence relation of being equal $\mathcal{L}_\Omega$-a.e. This space is endowed with the distance d_{L^2} defined by: for any $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in $L^2(\Omega, \mathcal{P}(D))$,*

$$d_{L^2}^2(\boldsymbol{\mu}, \boldsymbol{\nu}) := \int_\Omega W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\nu}(\xi)) \mathrm{d}\xi.$$

If $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$, we can define a probability measure on $\Omega \times D$, that we will call temporary $\bar{\boldsymbol{\mu}}$, in the following way: for any $a \in C(\Omega \times D)$,

$$\iint\limits_{\Omega \times D} a \mathrm{d}\bar{\boldsymbol{\mu}} := \int_\Omega \left(\int_D a(\xi, \cdot) \mathrm{d}\boldsymbol{\mu}(\xi) \right) \mathrm{d}\xi. \tag{8.1}$$

As we have assumed that the Lebesgue measure of Ω is 1, the measure $\bar{\boldsymbol{\mu}}$ is an actual probability measure on $\Omega \times D$. If we take a function $a \in C(\Omega)$ which depends only on variables in Ω, one can see that

$$\iint\limits_{\Omega \times D} a \mathrm{d}\bar{\boldsymbol{\mu}} = \int_\Omega a(\xi) \mathrm{d}\xi. \tag{8.2}$$

In other words, the marginal of $\bar{\boldsymbol{\mu}}$ is the Lebesgue measure (restricted to Ω). We will denote by $\mathcal{P}_0(\Omega \times D)$ the subspace of $\mathcal{P}(\Omega \times D)$ such that (8.2) is satisfied for all $a \in C(\Omega)$. Thanks to the disintegration Theorem [AGS08, Theorem 5.3.1], one can see that, reciprocally, to each $\bar{\boldsymbol{\mu}} \in \mathcal{P}_0(\Omega \times D)$, one can associate a unique element $\boldsymbol{\mu}$ of $L^2(\Omega, \mathcal{P}(D))$ such that (8.1) holds. In all the sequel, we will drop the "bar" on $\bar{\boldsymbol{\mu}}$ and use the same letter $\boldsymbol{\mu}$ to denote an element of $L^2(\Omega, \mathcal{P}(D))$ and its counterpart in $\mathcal{P}_0(\Omega \times D)$ through the bijection that we have just described. Any $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ can be seen in two different ways: either as a mapping $\Omega \to \mathcal{P}(D)$, or as a probability measure on $\Omega \times D$, and we will very often switch between the two points of view. To clarify the notations:

- if $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$, then $\boldsymbol{\mu}(\xi)$ or $\boldsymbol{\mu}(\xi, \mathrm{d}x)$, which is an element of $\mathcal{P}(D)$, will denote the mapping $\boldsymbol{\mu}$ evaluated at ξ;
- $\boldsymbol{\mu}(\mathrm{d}\xi, \mathrm{d}x)$ will indicate that we consider $\boldsymbol{\mu}$ as an element of $\mathcal{P}_0(\Omega \times D)$, integration on $\Omega \times D$ will be denoted by $\mathrm{d}\boldsymbol{\mu}$ or $\boldsymbol{\mu}(\mathrm{d}\xi, \mathrm{d}x)$, notice that we have the following relation: $\boldsymbol{\mu}(\mathrm{d}\xi, \mathrm{d}x) = \boldsymbol{\mu}(\xi, \mathrm{d}x)\mathrm{d}\xi$;
- the mapping $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ is said continuous (resp. Lipschitz) if there is one representative of $\boldsymbol{\mu}$ such that $W_2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta))$ goes to 0 if $\eta \to \xi$ (resp. is bounded by $C|\xi - \eta|$ for some $C < +\infty$).

The topologies on $L^2(\Omega, \mathcal{P}(D))$ are defined as follows.

Definition 8.2. *The strong topology on $L^2(\Omega, \mathcal{P}(D))$ is the one induced by the distance d_{L^2}, and the weak topology is the one induced on $\mathcal{P}_0(\Omega \times D)$ by the weak topology on $\mathcal{P}(\Omega \times D)$.*

Proposition 8.3. *W.r.t. the strong topology, $L^2(\Omega, \mathcal{P}(D))$ is a polish space. W.r.t. the weak topology, $L^2(\Omega, \mathcal{P}(D))$ is a separable compact space. Moreover, the strong topology is finer than the weak topology.*

Proof. The statement concerning the strong topology is a consequence of the fact that $\mathcal{P}(D)$ is itself a polish space, see for instance [KS93, Section 1.1]. As $\mathcal{P}_0(\Omega \times D)$ is closed in $\mathcal{P}(\Omega \times D)$, for the second statement we simply use the fact that $\mathcal{P}(\Omega \times D)$ is itself a separable compact space.

To compare the topologies we take a sequence $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ which converges strongly to some $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$. Up to extraction, we know that we can assume that $\boldsymbol{\mu}_n(\xi)$ converges in $\mathcal{P}(D)$ to $\boldsymbol{\mu}(\xi)$ for a.e. $\xi \in \Omega$. In particular, if $a \in C(\Omega \times D)$, we have that $\int_D a(\xi, \cdot)\mathrm{d}\boldsymbol{\mu}_n(\xi)$ converges to $\int_D a(\xi, \cdot)\mathrm{d}\boldsymbol{\mu}(\xi)$ for a.e. $\xi \in \Omega$. With the help of Lebesgue dominated convergence Theorem, we see that

$$\lim_{n\to+\infty} \iint_{\Omega\times D} a\mathrm{d}\boldsymbol{\mu}_n = \lim_{n\to+\infty} \int_\Omega \left(\int_D a(\xi, \cdot)\mathrm{d}\boldsymbol{\mu}_n(\xi) \right) \mathrm{d}\xi = \int_\Omega \left(\int_D a(\xi, \cdot)\mathrm{d}\boldsymbol{\mu}(\xi) \right) \mathrm{d}\xi = \iint_{\Omega\times D} a\mathrm{d}\boldsymbol{\mu}.$$

As a is arbitrary, this allows us to conclude that $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ converges to $\boldsymbol{\mu}$ for the weak topology. □

8.1 The Dirichlet energy

8.1.1 A Benamou-Brenier type definition

We are now ready to define the Dirichlet energy. The first step is to define the (generalized) continuity equation. Recall that $C^1_c(\mathring{\Omega} \times D, \mathbb{R}^p)$ is the set of C^1 functions defined on $\Omega \times D$ and valued in $\mathbb{R}^p$, whose support is compactly included in $\mathring{\Omega}$, but not necessarily in D, and $\mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ denotes the space of vector-valued measures on $\Omega \times D$ with finite mass.

Definition 8.4. *If $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ and if $\mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$, we say that the pair $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation if, for every $\varphi \in C^1_c(\mathring{\Omega} \times D, \mathbb{R}^p)$, one has*

$$\iint_{\Omega\times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint_{\Omega\times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E} = 0.$$

In other words, the pair $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation if the equation

$$\nabla_\Omega \boldsymbol{\mu} + \nabla_D \cdot \mathbf{E} = 0.$$

with no-flux boundary conditions on ∂D is satisfied in a weak sense. If we develop in coordinates, it means that for every $\alpha \in \{1, 2, \ldots, p\}$, one has $\partial_\alpha \boldsymbol{\mu} + \sum_{i=1}^q \partial_i \mathbf{E}^{i\alpha} = 0$. If the pair $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation, we want to define its Dirichlet energy by $\iint_{\Omega \times D} \frac{|\mathbf{E}|^2}{2\boldsymbol{\mu}}$. It is well known in optimal transport that this definition can be made by duality.

Definition 8.5. *If* $(\boldsymbol{\mu}, \mathbf{E})$ *satisfies the continuity equation, we define its Dirichlet energy* $\mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E})$ *by*

$$\mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) := \sup_{a,b} \left\{ \iint_{\Omega \times D} a \mathrm{d}\boldsymbol{\mu} + \iint_{\Omega \times D} b \cdot \mathrm{d}\mathbf{E} \ : \ (a,b) \in C(\Omega \times D, \mathcal{K}) \right\},$$

where $\mathcal{K} \subset \mathbb{R}^{1+pq}$ *is the set of pair* (x, y) *with* $x \in \mathbb{R}$ *and* $y \in \mathbb{R}^{pq}$ *such that* $x + \frac{1}{2}|y|^2 \leqslant 0$.

Note that $|y|$ is the Euclidean norm of $y \in \mathbb{R}^{pq}$. In other words, if y is seen as $p \times q$ matrix, $|y|$ is the Hilbert-Schmidt norm of the matrix. The following proposition is identical to the case of the Benamou-Brenier formula.

Proposition 8.6. *If* $(\boldsymbol{\mu}, \mathbf{E})$ *satisfies the continuity equation and* $\mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) < +\infty$, *then* $\mathbf{E}$ *is absolutely continuous w.r.t.* $\boldsymbol{\mu}$, *and if* $\mathbf{v} : \Omega \times D \to \mathbb{R}^{pq}$ *is the density of* $\mathbf{E}$ *w.r.t.* $\boldsymbol{\mu}$, *then one has*

$$\mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) = \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{v}\boldsymbol{\mu}) = \iint_{\Omega \times D} \frac{1}{2} |\mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu}.$$

Proof. There is nothing to add to the proof of this when Ω is 1-dimensional, and such a proof can be found for instance in [San15, Proposition 5.18]. □

Definition 8.7. *Let* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$. *Its Dirichlet energy* $\mathrm{Dir}(\boldsymbol{\mu})$ *is defined by*

$$\mathrm{Dir}(\boldsymbol{\mu}) := \inf_{\mathbf{E}} \left\{ \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) \ : \ \mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq}) \text{ and } (\boldsymbol{\mu}, \mathbf{E}) \text{ satisfies the continuity equation} \right\}.$$

Let us underline that if there exists no $\mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ such that $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation, then by convention $\mathrm{Dir}(\boldsymbol{\mu}) = +\infty$. To be sure that it is written somewhere, let us state the following proposition which identifies the Dirichlet energy if Ω is a segment of $\mathbb{R}$ with what we called previously the *action* of a curve. It is a consequence of Theorem 2.8.

Proposition 8.8. *Assume that* I *is a segment of* $\mathbb{R}$ *and let* $\boldsymbol{\mu} \in L^2(I, \mathcal{P}(D))$. *Then* $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ *if and only if* $\boldsymbol{\mu}$ *is* 2*-absolutely continuous, and in this case*

$$\mathrm{Dir}(\boldsymbol{\mu}) = \int_I \frac{1}{2} |\dot{\boldsymbol{\mu}}|^2(t) \mathrm{d}t.$$

Now, let us show easy properties of the functional Dir and the optimal $\mathbf{v}$ and $\mathbf{E}$. The proofs are straightforward adaptation of the case where Ω is a segment of $\mathbb{R}$.

Proposition 8.9. *If* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ *is such that* $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$, *then there exists a unique* $\mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ *such that* $(\boldsymbol{\mu}, \mathbf{E})$ *satisfies the continuity equation and* $\mathrm{Dir}(\boldsymbol{\mu}) = \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E})$.

Definition 8.10. *If* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ *and if* $\mathbf{E} = \mathbf{v}\boldsymbol{\mu}$ *is such that* $(\boldsymbol{\mu}, \mathbf{E})$ *satisfies the continuity equation and* $\mathrm{Dir}(\boldsymbol{\mu}) = \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) < +\infty$, *then* $\mathbf{E}$ *and* $\mathbf{v}$ *are said* tangent *to* $\boldsymbol{\mu}$.

The terminology *tangent* comes from [AGS08]. As in the case of absolutely continuous curves, there is a characterization of the tangent velocity field $\mathbf{v}$ which looks like the one of Theorem 2.8.

Proposition 8.11. *Let* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ *such that* $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ *and* $\mathbf{v} \in L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$ *such that* $(\boldsymbol{\mu}, \mathbf{v}\boldsymbol{\mu})$ *satisfies the continuity equation. Then* $\mathbf{v}$ *is tangent to* $\boldsymbol{\mu}$ *if and only if there exists a sequence* $(\psi_n)_{n\in\mathbb{N}}$ *in* $C^1(\Omega \times D, \mathbb{R}^p)$ *such that* $(\nabla_D \psi_n)_{n\in\mathbb{N}}$ *converges to* $\mathbf{v}$ *in* $L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$.

Proof of Proposition 8.9 and Proposition 8.11. In the Hilbert space $L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$ the set X of $\mathbf{v}$ such that $(\boldsymbol{\mu}, \mathbf{v}\boldsymbol{\mu})$ satisfies the continuity equation is clearly an affine set, and it is not empty as $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$. Denoting by $Y = \{\nabla\psi \ : \ \psi \in C^1(\mathring{\Omega} \times D, \mathbb{R}^p)\}$, it is clear that X is parallel to $Y^\perp$.

Thanks to Proposition 8.6, the problem of calculus of variations in Definition 8.7 corresponds to finding the orthogonal projection of the vector $0 \in L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$ on the set of X, i.e. Proposition 8.9 is proved.

It is well known that the projection $\mathbf{v}$ is characterized by the fact that $\mathbf{v}$ is orthogonal to any vector in the linear space parallel to X. In other words, $\mathbf{v}$ is characterized (beside the fact that it satisfies the continuity equation) by $\mathbf{v} \in X^\perp = (Y^\perp)^\perp$. The latter is nothing else than the closure in $L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$ of Y. An easy argument involving cutoff functions shows that this closure is the same as the closure of the set of $\nabla_D \psi$ for $\psi \in C^1(\Omega \times D, \mathbb{R}^p)$, hence Proposition 8.11 is proved. ☐

As an immediate corollary, Proposition 8.11 implies a localization property: the tangent velocity field $\mathbf{v}$, depends only locally on the values of $\boldsymbol{\mu}$. In the next proposition, $\boldsymbol{\mu}|_{\tilde{\Omega}}$ and $\mathbf{v}|_{\tilde{\Omega}}$ will denote the restrictions of $\boldsymbol{\mu}$ and $\mathbf{v}$ to a subset $\tilde{\Omega}$ of Ω.

Corollary 8.12. *Let* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ *such that* $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ *and let* $\mathbf{v} \in L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$ *be tangent to* $\boldsymbol{\mu}$*. Then, if* $\tilde{\Omega}$ *is any subdomain compactly supported in* $\mathring{\Omega}$*,* $\mathbf{v}|_{\tilde{\Omega}}$ *is tangent to* $\boldsymbol{\mu}|_{\tilde{\Omega}}$.

Still building from Proposition 8.11, we can build some sort of dual representation for the Dirichlet energy. Namely, we can say that

$$\mathrm{Dir}(\boldsymbol{\mu}) = \sup_{\varphi} \left\{ - \iint_{\Omega \times D} \left(\nabla_\Omega \cdot \varphi + \frac{1}{2} |\nabla_D \varphi|^2 \right) \mathrm{d}\boldsymbol{\mu} \ : \ \varphi \in C^1_c(\mathring{\Omega} \times D, \mathbb{R}^p) \right\}. \tag{8.3}$$

Indeed, if $\mathbf{v}$ is the tangent velocity field to $\boldsymbol{\mu}$, given the continuity equation and elementary algebra,

$$\begin{aligned} - \iint_{\Omega \times D} \left(\nabla_\Omega \cdot \varphi + \frac{1}{2} |\nabla_D \varphi|^2 \right) \mathrm{d}\boldsymbol{\mu} &= \iint_{\Omega \times D} \left(\nabla_D \varphi \cdot \mathbf{v} - \frac{1}{2} |\nabla_D \varphi|^2 \right) \mathrm{d}\boldsymbol{\mu} \\ &= \mathrm{Dir}(\boldsymbol{\mu}) - \frac{1}{2} \iint_{\Omega \times D} |\nabla_D \varphi - \mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu}. \end{aligned}$$

Hence the l.h.s. is always smaller than $\mathrm{Dir}(\boldsymbol{\mu})$, and we can make the discrepancy arbitrary small thanks to Proposition 8.11.

Proposition 8.13. *The mapping* $\mathrm{Dir} : L^2(\Omega, \mathcal{P}(D)) \to \overline{\mathbb{R}}$ *is l.s.c. w.r.t. weak convergence. Moreover it is convex: for any* $\boldsymbol{\mu}$ *and* $\boldsymbol{\nu}$ *in* $L^2(\Omega, \mathcal{P}(D))$ *and any* $t \in [0, 1]$,

$$\mathrm{Dir}((1-t)\boldsymbol{\mu} + t\boldsymbol{\nu}) \leqslant (1-t)\mathrm{Dir}(\boldsymbol{\mu}) + t\mathrm{Dir}(\boldsymbol{\nu}).$$

Proof. From (8.3), we see that Dir is the supremum of linear and continuous (w.r.t. weak convergence) functionals on $L^2(\Omega, \mathcal{P}(D))$. Hence it is convex and continuous. ☐

We will conclude this subsection by showing the following approximation result, which will be useful to prove the equivalences with the metric definitions. We will not be able to regularize up to the boundary of Ω, though it will be sufficient for our purpose.

Theorem 8.14. *Fix $\tilde{\Omega} \subset \mathring{\Omega}$ compactly embedded in $\mathring{\Omega}$. Let $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ with $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$. Then there exists a sequence $\boldsymbol{\mu}_n \in L^2(\tilde{\Omega}, \mathcal{P}(D))$ with the following properties:*

(i) For any $n \in \mathbb{N}$, $\boldsymbol{\mu}_n(\mathrm{d}\xi, \mathrm{d}x) = \rho_n(\xi, x)\mathrm{d}\xi\mathrm{d}x$, where the density ρ_n of $\boldsymbol{\mu}_n$ w.r.t. to $\mathcal{L}_{\tilde{\Omega}} \otimes \mathcal{L}_D$ satisfies $\rho_n \in C^\infty(\tilde{\Omega}, L^\infty(D))$ and $\operatorname{ess\,inf}_{\tilde{\Omega}\times D} \rho_n > 0$.

(ii) The sequence $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ converges weakly to $\boldsymbol{\mu}$ in $L^2(\tilde{\Omega}, \mathcal{P}(D))$.

(iii) There holds

$$\lim_{n\to+\infty} \mathrm{Dir}(\boldsymbol{\mu}_n) = \mathrm{Dir}(\boldsymbol{\mu}|_{\tilde{\Omega}}).$$

Notice that $\boldsymbol{\mu}_n$ is defined only on $\tilde{\Omega}$, i.e. not on the full domain Ω.

Proof. On Ω, we will regularize with a convolution kernel χ. Specifically, we fix $\chi : \mathbb{R}^p \to [0,1]$ a smooth function, radial, compactly supported in $B(0,1)$ and of total integral 1, and we set $\chi_n(\xi) = n^p \chi(n\xi)$. On the other hand, on D we will regularize with the heat flow that we denote by Φ^D, see Section 2.4. We set $\tilde{\boldsymbol{\mu}}_n(\xi) := [\Phi^D_{1/n}][\boldsymbol{\mu}(\xi)]$ for any $\xi \in \Omega$. Hence $\tilde{\boldsymbol{\mu}}_n \in L^2(\Omega, \mathcal{P}(D))$ is defined on the whole Ω. For n large enough and $\xi \in \tilde{\Omega}$ we define

$$\boldsymbol{\mu}_n(\xi) := \int_\Omega \chi_n(\xi - \eta)\tilde{\boldsymbol{\mu}}_n(\eta)\mathrm{d}\eta,$$

where here we do the usual (linear) mean of probability measures. In short, $\boldsymbol{\mu}_n = \chi_n \star_\Omega \tilde{\boldsymbol{\mu}}_n$. Here we need n such that the support of χ_n is small compared to the distance between $\tilde{\Omega}$ and $\partial\Omega$.

Assertion (i) holds because of the regularization properties of the convolution and the lower bound on the solution of the heat flow.

Assertion (ii) is standard: if we fix $a \in C(\tilde{\Omega} \times D)$, given the self-adjacency of the heat flow and the symmetry of the heat kernel,

$$\iint_{\tilde{\Omega}\times D} a \mathrm{d}\boldsymbol{\mu}_n = \iint_{\tilde{\Omega}\times D} \Phi^D_{1/n}[\chi_n \star_\Omega a] \; \mathrm{d}\boldsymbol{\mu}$$

and the r.h.s. converges strongly to the integral of a against $\boldsymbol{\mu}$ because of standard functional analysis.

Assertion (iii) is slightly trickier. As we have already seen earlier in this manuscript (see for instance Proposition 4.26), applying the heat flow decreases the Dirichlet energy, at least for curves valued in the Wasserstein space. With mappings, provided we admit the representation given below by Theorem 8.26 and the contraction property of the heat flow, it is straightforward that we should have $\mathrm{Dir}(\tilde{\boldsymbol{\mu}}_n) \leqslant \mathrm{Dir}(\boldsymbol{\mu})$. But the current theorem will be used to prove Theorem 8.26, hence we cannot invoke it. We adopt a different strategy: we start with the "dual" representation for the Dirichlet energy given by (8.3). We want to show that $\mathrm{Dir}(\tilde{\boldsymbol{\mu}}_n) \leqslant \mathrm{Dir}(\boldsymbol{\mu})$. For any fixed $\varphi \in C^1_c(\mathring{\Omega} \times D, \mathbb{R}^p)$, and given that the heat flow is self-adjoint,

$$\iint_{\Omega\times D} \left(\nabla_\Omega \cdot \varphi + \frac{1}{2}|\nabla_D \varphi|^2 \right) \mathrm{d}\tilde{\boldsymbol{\mu}}_n = \iint_{\Omega\times D} \left(\nabla_\Omega \cdot (\Phi^D_{1/n}\varphi) + \Phi^D_{1/n}\left(\frac{1}{2}|\nabla_D \varphi|^2 \right) \right) \mathrm{d}\boldsymbol{\mu}.$$

Notice that we used the property that the heat flow acting on D commutes with $\nabla_\Omega \cdot$. Now, the key point is the so-called Bakery-Émery estimate

$$\frac{1}{2}\left|\nabla\left(\Phi^D_{1/n}\varphi\right)\right|^2 \leqslant \Phi^D_{1/n}\left(\frac{1}{2}|\nabla_D\varphi|^2\right)$$

which is valid because D is a convex domain [GKM18, Equation (2.4)]. Hence

$$\iint_{\Omega\times D}\left(\nabla_\Omega\cdot\varphi+\frac{1}{2}|\nabla_D\varphi|^2\right)\mathrm{d}\tilde{\boldsymbol{\mu}}_n \geqslant \iint_{\Omega\times D}\left(\nabla_\Omega\cdot(\Phi^D_{1/n}\varphi)+\frac{1}{2}\left|\nabla\left(\Phi^D_{1/n}\varphi\right)\right|^2\right)\mathrm{d}\boldsymbol{\mu} \geqslant -\mathrm{Dir}(\boldsymbol{\mu}),$$

where the last inequality comes from (8.3). Taking the supremum in φ and using the representation formula (8.3) we conclude that $\mathrm{Dir}(\tilde{\boldsymbol{\mu}}_n) \leqslant \mathrm{Dir}(\boldsymbol{\mu})$. Now we want to control the Dirichlet energy of $\boldsymbol{\mu}_n$ with the one of $\tilde{\boldsymbol{\mu}}_n$. Recall that Dir is a convex function. But $\boldsymbol{\mu}_n$ is the average, w.r.t. to the weights $\chi_n(\eta)$, of the mappings $\xi \mapsto \tilde{\boldsymbol{\mu}}_n(\xi-\eta)$. Hence, by Jensen's inequality,

$$\mathrm{Dir}(\boldsymbol{\mu}_n) \leqslant \int_{B(0,1/n)} \chi_n(\eta)\mathrm{Dir}\left(\tilde{\boldsymbol{\mu}}_n|_{\tilde{\Omega}}(\cdot-\eta)\right).$$

Hence, calling Ω_n the set of points which are distant at most $1/n$ from $\tilde{\Omega}$, one has $\mathrm{Dir}(\boldsymbol{\mu}_n) \leqslant \mathrm{Dir}(\tilde{\boldsymbol{\mu}}_n|_{\Omega_n})$. Sending n to $+\infty$ and using the lower semi-continuity of Dir and assertion (ii) to get the reverse inequality, we get (iii). □

8.1.2 The smooth case

In this subsection, we will briefly study the smooth case, i.e. the one where $\boldsymbol{\mu}$ has a smooth and strictly positive density w.r.t. $\mathcal{L}_\Omega \otimes \mathcal{L}_D$. It will help us to understand the meaning of the continuity equation and we will use it in the sequel when reasoning by approximation. Basically, if we are in a sufficiently smooth setting, we can give a precise meaning to the arguments evoked in Section 2.2 about Otto calculus.

Definition 8.15. *A mapping $\boldsymbol{\mu} \in L^2(\Omega,\mathcal{P}(D))$ with $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ is said smooth if it admits a density ρ w.r.t. $\mathcal{L}_\Omega \otimes \mathcal{L}_D$ satisfying $\rho \in C^\infty(\Omega, L^\infty(D))$ and uniformly bounded from below.*

In particular, it implies ρ is uniformly bounded (from above) on the closed set Ω. Notice that Theorem 8.14 says that any $\boldsymbol{\mu} \in L^2(\Omega,\mathcal{P}(D))$ with finite Dirichlet energy can be approximated by a sequence of smooth functions (only in the interior of Ω) according to Definition 8.15. Let us start by explaining how, in the smooth case, one can compute the tangent velocity field.

Proposition 8.16. *Let $\boldsymbol{\mu} \in L^2(\Omega,\mathcal{P}(D))$ be smooth. Then, for every $\xi \in \mathring{\Omega}$, there exists a unique $\varphi(\xi,\cdot) \in H^1(D,\mathbb{R}^p)$ with 0-mean solution to the elliptic equation*

$$\begin{cases} \nabla_D\cdot(\rho(\xi,\cdot)\nabla_D\varphi(\xi,\cdot)) = -\nabla_\Omega\rho(\xi,\cdot) & \text{in } \mathring{D} \\ \nabla_D\varphi(\xi,\cdot)\cdot\mathbf{n}_D = 0 & \text{on } \partial D. \end{cases} \tag{8.4}$$

Moreover $\nabla_D\varphi \in L^2_{\boldsymbol{\mu}}(\Omega\times D,\mathbb{R}^{pq})$ is the tangent velocity field to $\boldsymbol{\mu}$ and it is continuous as a mapping from Ω to $L^2(D,\mathbb{R}^{pq})$.

Proof. The existence of a unique solution to the elliptic equation (8.4) derives from standard arguments. Notice that $\nabla_\Omega\rho(\xi,\cdot)$ has always 0-mean on D, hence the equation is well-posed. In particular, as ρ is bounded from below, the equation is uniformly elliptic. We have the usual estimate

$$\|\nabla_D \varphi(\xi,\cdot)\|_{L^2(D,\mathbb{R}^{pq})} \leqslant C \|\nabla_D \varphi(\xi,\cdot)\|_{L^2_{\rho(\xi,\cdot)}(D,\mathbb{R}^{pq})} \leqslant C \|\nabla_\Omega \rho(\xi,\cdot)\|_\infty,$$

which tells us that $\nabla_D \varphi(\xi,\cdot)$ is uniformly bounded (w.r.t. ξ) in $L^2(D,\mathbb{R}^{pq})$. By construction, $\mathbf{v} := \nabla_D \varphi$ is such that $(\boldsymbol{\mu}, \mathbf{v}\boldsymbol{\mu})$ satisfies the continuity equation.

To prove continuity of $\xi \mapsto \nabla_D \varphi(\xi,\cdot)$, let us fix $\xi \in \mathring{\Omega}$ and a sequence ξ_n which converges to ξ. We use momentarily the compact notations $\bar{\varphi} = \varphi(\xi,\cdot) \in H^1(D,\mathbb{R}^{pq})$ and $\varphi_n = \varphi(\xi_n,\cdot) \in H^1(D,\mathbb{R}^{pq})$. Similarly, we set $\bar{\rho} = \rho(\xi,\cdot)$ and $\rho_n = \rho(\xi_n,\cdot)$. The r.h.s. of the elliptic equations will be $\bar{h} = -\nabla_\Omega \rho(\xi,\cdot)$ and $h_n = -\nabla_\Omega \rho(\xi_n,\cdot)$. We want to show that φ_n converges to $\bar{\varphi}$ in $H^1(D,\mathbb{R}^{pq})$, while we know that $\bar{\rho}, \rho_n$ are uniformly bounded from below and above, and that ρ_n (resp. h_n) converges to $\bar{\rho}$ (resp. $\bar{h}$) in $L^\infty(D)$. Clearly, $\varphi_n - \bar{\varphi}$ satisfies the elliptic equation

$$\nabla_D \cdot (\bar{\rho} \nabla_D(\varphi_n - \bar{\varphi})) = h_n - \bar{h} + \nabla_D \cdot ((\rho_n - \bar{\rho}) \nabla_D \varphi_n)$$

with Neumann boundary conditions. Testing this equation against $\varphi_n - \bar{\varphi}$, we deduce that

$$\|\nabla_D(\varphi_n - \bar{\varphi})\|_{L^2(D,\mathbb{R}^{pq})} \leqslant C\left(\|h_n - \bar{h}\|_{L^2(D)} + \|\rho_n - \bar{\rho}\|_\infty \|\nabla_D \varphi_n\|_{L^2(D,\mathbb{R}^{pq})}\right).$$

We can use the convergence of ρ_n to $\bar{\rho}$, h_n to $\bar{h}$ and the fact that $\|\nabla_D \varphi_n\|_{L^2(D,\mathbb{R}^{pq})}$ is uniformly bounded in n to conclude that the l.h.s. goes to 0 as $n \to +\infty$. ☐

Now take $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ smooth and denote by $\mathbf{v} = \nabla_D \varphi$ its tangent velocity field. If $\gamma : I \to \mathring{\Omega}$ is a smooth curve going from an interval of $\mathbb{R}$ to $\mathring{\Omega}$, then, multiplying (8.4) by $\dot{\gamma}$, one can see that $\boldsymbol{\mu}^\gamma = \boldsymbol{\mu} \circ \gamma : I \to \mathcal{P}(D)$ defines a curve valued in the Wasserstein space for which the (classical) continuity equation $\partial_t \boldsymbol{\mu}^\gamma + \nabla \cdot (\mathbf{v}^\gamma \boldsymbol{\mu}^\gamma)$ with Neumann boundary conditions is satisfied (at least in a weak sense), provided that we define $\mathbf{v}^\gamma := \mathbf{v} \cdot \dot{\gamma} : I \times D \to \mathbb{R}^q$. More precisely, if $i \in \{1, 2, \ldots, q\}$, the i-the component of $\mathbf{v}^\gamma$ at time $t \in I$ and at the point $x \in D$ is

$$(\mathbf{v}^\gamma(t,x))^i = \sum_{\alpha=1}^{p} \mathbf{v}(\gamma(t), x)^{i\alpha} \dot{\gamma}^\alpha(t).$$

In other words, the (generalized) continuity equation implies that we get (classical) continuity equation for every curve of Ω. In some sense, the (generalized) continuity equation is much stronger in higher dimensions.

As we recalled previously, the velocity field $\mathbf{v}^\gamma$ is related to the metric derivative of the curve $\boldsymbol{\mu}^\gamma$ in the Wasserstein space. As the tangent velocity field $\mathbf{v} \in L^2(\Omega \times D, \mathbb{R}^{pq})$ is the gradient of a function $\nabla_D \varphi$, by Proposition 8.11 $\mathbf{v}^\gamma$ is the tangent velocity field to the curve $\boldsymbol{\mu}^\gamma$. Using Theorems 2.7 and 2.8, we see that for all $s \leqslant t \in I$,

$$\frac{W_2^2(\boldsymbol{\mu}(\gamma(t), \boldsymbol{\mu}(\gamma(s)))}{t-s} \leqslant \int_s^t \int_D |\mathbf{v}^\gamma(r,x)|^2 \boldsymbol{\mu}(\gamma(r), \mathrm{d}x) \mathrm{d}r. \tag{8.5}$$

But in fact, we can say more and go from a global estimate to a local one, this is the object of the following proposition.

Proposition 8.17. *Let $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ be smooth and let $\mathbf{v} \in C(\Omega, L^2(D, \mathbb{R}^{pq}))$ its tangent velocity field.*

Then the function $\boldsymbol{\mu}$ is Lipschitz. Moreover, if $\xi \in \mathring{\Omega}$ and $\eta \in \mathbb{R}^p$,

$$\lim_{\varepsilon \to 0} \frac{W_2(\boldsymbol{\mu}(\xi + \varepsilon\eta), \boldsymbol{\mu}(\xi))}{|\varepsilon|} = \sqrt{\int_D |\mathbf{v}(\xi,x) \cdot \eta|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x)}. \tag{8.6}$$

The important point of this proposition is that the estimate holds for *all* points of Ω, there is no more "almost everywhere" in the statement.

Proof. We fix $\xi \in \mathring{\Omega}$ and use $\gamma(t) := \xi + t\eta$ which is defined for t sufficiently close to 0. Notice that $\mathbf{v}^\gamma(t,x) = \mathbf{v}(\xi + t\eta, x) \cdot \eta$.

We denote by ρ the density of $\boldsymbol{\mu}$ w.r.t. $\mathcal{L}_\Omega \times \mathcal{L}_D$. To prove that $\boldsymbol{\mu}$ is Lipschitz, we use (8.5) and the fact that $\rho \in L^\infty(\Omega \times D)$ and $\mathbf{v} \in L^\infty(\Omega, L^2(D, \mathbb{R}^{pq}))$.

The fact the l.h.s. of (8.6) (provided the lim is replaced by a lim sup) is bounded by the r.h.s. comes directly from (8.5) and the continuity of $\mathbf{v} : \Omega \to L^2(D, \mathbb{R}^{pq})$.

To prove the reverse inequality, take a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ realizing the lim inf for the l.h.s. of (8.6). Call $\psi_n \in C^0(D)$ the function with 0-mean such that $\varepsilon_n\psi_n$ is the Kantorovich potential from $\boldsymbol{\mu}(\xi)$ to $\boldsymbol{\mu}(\xi + \varepsilon_n\eta)$, it is unique because $\boldsymbol{\mu}(\xi)$ is supported on the whole D, see Proposition 2.3. As $\mathrm{Id} - \varepsilon_n\nabla_D\psi_n$ is the optimal transport map from $\boldsymbol{\mu}(\xi)$ onto $\boldsymbol{\mu}(\xi + \varepsilon_n\eta)$, there holds

$$\varepsilon_n \|\nabla_D\psi_n\|_{L^2_{\boldsymbol{\mu}(\xi)}(D)} = W_2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\xi + \varepsilon_n\eta)) \leqslant C\varepsilon_n,$$

where C is the Lipschitz constant of $\boldsymbol{\mu}$. In particular, using the lower bound on ρ, one sees that, up to a subsequence, $(\psi_n)_{n\in\mathbb{N}}$ converges weakly in $H^1_{\boldsymbol{\mu}(\xi)}(D)$ to some function ψ such that

$$\sqrt{\int_D |\nabla_D\psi(x)|^2\boldsymbol{\mu}(\xi, \mathrm{d}x)} = \|\nabla_D\psi\|_{L^2_{\boldsymbol{\mu}(\xi)}(D)} \leqslant \liminf_{n\to+\infty} \frac{W_2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\xi + \varepsilon_n e))}{\varepsilon_n}.$$

Thus, to conclude, it is enough to show that $\nabla_D\psi = -\mathbf{v}(\xi, \cdot) \cdot \eta$.

As $\mathrm{Id} - \varepsilon_n\nabla_D\psi_n$ transports $\boldsymbol{\mu}(\xi)$ onto $\boldsymbol{\mu}(\xi + \varepsilon_n\eta)$, for any $f \in C^1(D)$, one has

$$\int_D f(x - \varepsilon_n\nabla_D\psi_n(x))\rho(\xi, x)\mathrm{d}x = \int_D f(x)\rho(\xi + \varepsilon_n\eta, x)\mathrm{d}x.$$

Using a Taylor expansion on f and dividing by ε_n,

$$\left| \int_D \nabla_D\psi_n(x) \cdot \nabla_D f(x)\rho(\xi, x)\mathrm{d}x + \int_D \frac{\rho(\xi + \varepsilon_n\eta, x) - \rho(\xi, x)}{\varepsilon_n} f(x)\mathrm{d}x \right| \leqslant C\varepsilon_n \int_D |\nabla_D\psi_n(x)|^2\mathrm{d}x,$$

where the constant C is a bound on the second derivative of f. Using the H^1 bound on ψ_n and the weak convergence to ψ, as well as the fact that ρ is differentiable w.r.t. variables in Ω, we conclude that ψ solves weakly the elliptic equation

$$\nabla_D \cdot (\rho(\xi, \cdot)\nabla_D\psi) = -\nabla_\Omega\rho(\xi, \cdot) \cdot \eta.$$

Using the uniqueness (recall that ψ_n has 0-mean, hence ψ too) result for equation (8.4), this allows to conclude that $\nabla_D\psi = \mathbf{v}(\xi, \cdot) \cdot \eta$ where $\mathbf{v}$ is the tangent velocity field to $\boldsymbol{\mu}$, hence the proposition is proved. □

8.1.3 Equivalence with Sobolev spaces valued in metric spaces

Until now, we have not discussed the existence of solutions to the (generalized) continuity equation: this notion could be too strong or too loose. In this subsection, we will show that the set of $\boldsymbol{\mu}$ with finite Dirichlet energy coincides with an already known definition of Sobolev spaces valued in metric spaces given by Reshetnyak [Res97, Res04]. This definition is restricted to the case where the source space has a smooth structure (which is precisely our framework), but can be seen as particular case of a more general definition given by Hajłasz (a pedagogic and clear introduction to the latter can be found in [AT03, Chapter 5]).

Definition 8.18. *Let* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$*. For any* $\nu \in \mathcal{P}(D)$*, define* $[\boldsymbol{\mu}]_\nu \in L^2(\Omega)$ *by* $[\boldsymbol{\mu}]_\nu(\xi) := W_2(\boldsymbol{\mu}(\xi), \nu)$*. We say that* $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ *if there exists a countable family* $(\nu_n)_{n \in \mathbb{N}}$ *dense in* $\mathcal{P}(D)$ *such that* $[\boldsymbol{\mu}]_{\nu_n} \in H^1(\Omega)$ *for all* $n \in \mathbb{N}$ *and there exists a function* $g \in L^2(\Omega)$ *such that, for every* $n \in \mathbb{N}$*,*

$$|\nabla [\boldsymbol{\mu}]_{\nu_n}| \leqslant g \tag{8.7}$$

a.e. on Ω*. The smallest* g *for which* (8.7) *holds is called the metric gradient of* $\boldsymbol{\mu}$ *and is denoted by* $g_{\boldsymbol{\mu}}$*.*

Notice that $g_{\boldsymbol{\mu}} = \sup_n |\nabla [\boldsymbol{\mu}]_{\nu_n}|$. The definition looks slightly different than in [Res97]. However, it is equivalent because of the following result:

Proposition 8.19. *Let* $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ *and* $g_{\boldsymbol{\mu}} \in L^2(\Omega)$ *be its metric gradient. Then for all mappings* $u : \mathcal{P}(D) \to \mathbb{R}$ *which are* C*-Lipschitz,* $u \circ \boldsymbol{\mu} \in H^1(\Omega)$ *and* $|\nabla(u \circ \boldsymbol{\mu})| \leqslant C g_{\boldsymbol{\mu}}$ *a.e. on* Ω*.*

Proof. Is is enough to copy the proof of [Res97, Theorem 5.1]. Indeed, in this proof, one only uses the functions $[\boldsymbol{\mu}]_\nu$ for measures ν belonging to a dense and countable subset of $\mathcal{P}(D)$. □

In particular, if $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$, then $[\boldsymbol{\mu}]_\nu \in H^1(\Omega)$ with gradient bounded by $g_{\boldsymbol{\mu}}$ for all $\nu \in \mathcal{P}(D)$. Notice that the definition above can be stated for mappings valued in arbitrary metric spaces (separability of the target space is required). The main theorem of this subsection is the following, which states that the framework that we have developed coincides with the one of Reshetnyak.

Theorem 8.20. *Let* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$*. Then* $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ *if and only if* $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$*. Moreover, if* $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ *and if* $\mathbf{v}$ *is tangent to* $\boldsymbol{\mu}$*, then for a.e.* $\xi \in \Omega$*,*

$$g_{\boldsymbol{\mu}}(\xi) \leqslant \sqrt{\int_D |\mathbf{v}(\xi, x)|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x)} \leqslant \sqrt{p}\, g_{\boldsymbol{\mu}}(\xi).$$

Notice that the inequalities are sharp. The function $g_{\boldsymbol{\mu}}$ measures the norm of the gradient of $\boldsymbol{\mu}$ as an operator norm, whereas the norm of the velocity field $\mathbf{v}$ is measured with an Hilbert-Schmidt norm, which explains the discrepancy, see [Chi07] for a more detailed discussion.

We will prove this theorem in three steps. The first one is to prove it if Ω is a segment of $\mathbb{R}$ (Proposition 8.21). It is just a rewriting of the definition of Reshetnyak and does not rely of the special structure of the Wasserstein space. The second step is to say that, roughly speaking, a function is in $H^1(\Omega)$ if it is in H^1 for a.e. lines, with some uniform control on the gradients. It enables us to get the result if Ω is a cube (Proposition 8.23). The third step is simply to write that every domain can be written as a (countable) union of cubes.

Proposition 8.21. *Theorem 8.20 holds if* Ω *is a segment of* $\mathbb{R}$*.*

Proof. Assume $\Omega = I$ is a segment of $\mathbb{R}$. The set of curves with finite Dirichlet energy coincides with the set of absolutely continuous curves, see Proposition 8.8. Given Theorem 2.8, we want to prove the equality $g_{\boldsymbol{\mu}} = |\dot{\boldsymbol{\mu}}|$ a.e. on I.

Assume that $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ and take $\nu \in \mathcal{P}(D)$. Then, as $W_2(\cdot, \nu)$ is 1-Lipschitz, for all $s < t$ elements of I,

$$|[\boldsymbol{\mu}]_\nu(t) - [\boldsymbol{\mu}]_\nu(s)| \leqslant W_2(\boldsymbol{\mu}(t), \boldsymbol{\mu}(s)) \leqslant \int_s^t |\dot{\boldsymbol{\mu}}|(r) \mathrm{d}r.$$

It shows that the function $[\boldsymbol{\mu}]_\nu$ is in $H^1(I)$ and its gradient is smaller than $|\dot{\boldsymbol{\mu}}|$. Hence, as ν is arbitrary, $\boldsymbol{\mu} \in H^1(I, \mathcal{P}(D))$ and $g_{\boldsymbol{\mu}} \leqslant |\dot{\boldsymbol{\mu}}|$.

Reciprocally, assume $\boldsymbol{\mu} \in H^1(I, \mathcal{P}(D))$, take $(\nu_n)_{n\in\mathbb{N}}$ countable and dense in $\mathcal{P}(D)$ such that $[\boldsymbol{\mu}]_{\nu_n} \in H^1(I)$ for every $n \in \mathbb{N}$ with gradient bounded by $g_{\boldsymbol{\mu}}$. In particular, for any $n \in \mathbb{N}$ and any $s < t$ elements of I,

$$|[\boldsymbol{\mu}]_{\nu_n}(t) - [\boldsymbol{\mu}]_{\nu_n}(s)| \leqslant \int_s^t g_{\boldsymbol{\mu}}(r)\mathrm{d}r.$$

Then we choose ν_n arbitrary close to $\boldsymbol{\mu}(t)$: the r.h.s. is unchanged and the l.h.s. is arbitrary close to $W_2(\boldsymbol{\mu}(t), \boldsymbol{\mu}(s))$. Hence we conclude that

$$W_2(\boldsymbol{\mu}(s), \boldsymbol{\mu}(t)) \leqslant \int_t^s g_{\boldsymbol{\mu}}(r)\mathrm{d}r,$$

which is enough to say that $\boldsymbol{\mu}$ is an absolutely continuous curve and $|\dot{\boldsymbol{\mu}}| \leqslant g_{\boldsymbol{\mu}}$ a.e. on I by minimality of $|\dot{\boldsymbol{\mu}}|$. ☐

Now we will prove Theorem 8.20 at least locally, which means in the case where Ω is a cube. Up to an isometry and a dilatation, we can assume that Ω is the unit cube of $\mathbb{R}^p$. Recall that $(e_\alpha)_{1\leqslant\alpha\leqslant p}$ is the canonical basis of $\mathbb{R}^p$. In the sequel, we will denote by $\Omega_\alpha \subset \mathbb{R}^p$ the α-face of the cube, which means the set of $(\xi^1, \dots, \xi^{\alpha-1}, 0, \xi^{\alpha+1}, \dots, \xi^p)$, with $0 \leqslant \xi^\beta \leqslant 1$ for all $\beta \neq \alpha$. The measure on Ω_α will be the $p-1$-dimensional Lebesgue measure. If $f : \Omega \to X$ is a given mapping (where X is any set) and $\xi \in \Omega_\alpha$ is fixed, then $f_\xi : [0,1] \to X$ is defined by $f_\xi(t) = f(\xi + te_\alpha)$: it is the restriction of f to a line directed by e_α and crossing Ω_α at ξ. Recall the following characterization for real-valued mappings:

Proposition 8.22. *Assume Ω is the unit cube of $\mathbb{R}^p$ and let $f \in L^2(\Omega)$ be a given function. The function f belongs to $H^1(\Omega)$ if and only if for any $\alpha \in \{1, 2, \dots, p\}$, for a.e. $\xi \in \Omega_\alpha$, the function f_ξ is in $H^1([0,1])$ and*

$$\int_{\Omega_\alpha} \left(\int_0^1 |\dot{f}_\xi(t)|^2 \mathrm{d}t \right) \mathrm{d}\xi < +\infty.$$

Moreover, for a.e. $\xi \in \Omega_\alpha$ and a.e. $t \in [0,1]$,

$$(\partial_\alpha f)(\xi + te_\alpha) = \dot{f}_\xi(t).$$

Proof. One can look at [EG92, Section 4.9]. ☐

Proposition 8.23. *Theorem 8.20 holds if Ω is the unit cube of $\mathbb{R}^p$.*

Proof. Implication $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty \Rightarrow \boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. Assume first that $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ is such that $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ and take $\mathbf{v} \in L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$ the velocity field tangent to $\boldsymbol{\mu}$. Fix $\alpha \in \{1, 2, \dots, p\}$. Take two compactly supported test functions $\psi \in C^1_c(]0,1[\times D)$ and $a \in C^1_c(\Omega_\alpha)$. As a test function $\varphi \in C^1_c(\Omega \times D, \mathbb{R}^p)$ in the weak formulation of the continuity equation, choose $\varphi(\xi + te_\alpha, x) := (0, 0, \dots, 0, \psi(t,x)a(\xi), 0, \dots, 0)$ for $\xi \in \Omega_\alpha$ and $t \in [0,1]$ (only the α-th component of φ is not 0). If we expand we find that $\nabla_\Omega \cdot \varphi = a\partial_t \psi$ hence

$$0 = \iint\limits_{\Omega\times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint\limits_{\Omega\times D} \nabla_D \varphi \cdot \mathbf{v}\mathrm{d}\boldsymbol{\mu} = \int_{\Omega_\alpha} \left(\iint\limits_{[0,1]\times D} \partial_t \psi(t,x) \mathrm{d}t \boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x) \right) a(\xi)\mathrm{d}\xi$$
$$+ \int_{\Omega_\alpha} \left(\iint\limits_{[0,1]\times D} \nabla_D \psi(t,x) \cdot (\mathbf{v}(\xi + te_\alpha, x) \cdot e_\alpha) \mathrm{d}t \boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x) \right) a(\xi)\mathrm{d}\xi.$$

Using the arbitrariness of a, we deduce that for a.e. $\xi \in \Omega_\alpha$, and for a fixed $\psi \in C^1_c(]0,1[\times D, \mathbb{R}^p)$,

$$\iint_{[0,1]\times D} \partial_t \psi(t,x) \mathrm{d}t \boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x) + \iint_{[0,1]\times D} \nabla_D \psi(t,x) \cdot (\mathbf{v}(\xi + te_\alpha, x) \cdot e_\alpha) \mathrm{d}t \boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x) = 0. \quad (8.8)$$

Now, taking a sequence $(\psi_n)_{n\in\mathbb{N}}$ which is dense in $C^1_c(]0,1[\times D, \mathbb{R}^p)$, we can say that for a.e. $\xi \in \Omega_\alpha$, for all $\psi \in C^1_c(]0,1[\times D, \mathbb{R}^p)$, (8.8) holds. For $\xi \in \Omega_\alpha$ define $\boldsymbol{\mu}_\xi : [0,1] \to \mathcal{P}(D)$ by $\boldsymbol{\mu}_\xi(t) = \boldsymbol{\mu}(\xi + te_\alpha)$ and $\mathbf{v}_\xi : [0,1] \times D \to \mathbb{R}^q$ by $\mathbf{v}_\xi(t,x) = \mathbf{v}(\xi + te_\alpha, x) \cdot e_\alpha$. By Fubini's theorem, for a.e. $\xi \in \Omega_\alpha$, $\mathbf{v}_\xi \in L^2_{\boldsymbol{\mu}_\xi}([0,1] \times D, \mathbb{R}^q)$. Hence (8.8) rewrites as: for a.e. $\xi \in \Omega_\alpha$, the curve $\boldsymbol{\mu}_\xi$ is an absolutely continuous curve in the Wasserstein space with a velocity field given by $\mathbf{v}_\xi$. By Proposition 8.21, if $\nu \in \mathcal{P}(D)$, then the function $[\boldsymbol{\mu}_\xi]_\nu$ is in $H^1([0,1])$ and

$$|\partial_t [\boldsymbol{\mu}_\xi]_\nu(t)| \leqslant \sqrt{\int_D |\mathbf{v}_\xi(t,x)|^2 \boldsymbol{\mu}_\xi(t, \mathrm{d}x)} = \sqrt{\int_D |\mathbf{v}(\xi + te_\alpha, x) \cdot e_\alpha|^2 \boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x)}.$$

As the r.h.s. is integrable over $[0,1] \times \Omega_\alpha$ and α is arbitrary, we can use Proposition 8.22 to see that $[\boldsymbol{\mu}]_\nu \in H^1(\Omega)$. Moreover, taking the square of the previous equation and summing over $\alpha \in \{1, 2, \ldots, p\}$, we see that for a.e. $\xi \in \Omega$

$$|\nabla [\boldsymbol{\mu}]_\nu(\xi)|^2 \leqslant \int_D |\mathbf{v}(\xi, x)|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x).$$

Thus, we conclude that $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ and for a.e. $\xi \in \Omega$,

$$g_{\boldsymbol{\mu}}(\xi) \leqslant \sqrt{\int_D |\mathbf{v}(\xi, x)|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x)}. \quad (8.9)$$

Implication $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D)) \Rightarrow \mathrm{Dir}(\boldsymbol{\mu}) < +\infty$. Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. Take $(\nu_n)_{n\in\mathbb{N}}$ a sequence which is dense in $\mathcal{P}(D)$. For any $n \in \mathbb{N}$, the function $[\boldsymbol{\mu}]_{\nu_n}$ belongs to $H^1(\Omega)$. Fix $\alpha \in \{1, 2, \ldots, p\}$. For any $n \in \mathbb{N}$, for a.e. $\xi \in \Omega_\alpha$, the function $[\boldsymbol{\mu}_\xi]_{\nu_n} : t \mapsto W_2(\boldsymbol{\mu}(\xi + te_\alpha), \nu_n)$ is in $H^1([0,1])$ with a gradient bounded by $g_{\boldsymbol{\mu}}(\xi + te_\alpha)$. As $\mathbb{N}$ is countable, we can exchange the "for a.e. $\xi \in \Omega_\alpha$" and the "for all $n \in \mathbb{N}$". Hence, for a.e. $\xi \in \Omega_\alpha$, the function $\boldsymbol{\mu}_\xi : [0,1] \to \mathcal{P}(D)$ belongs to $H^1([0,1], \mathcal{P}(D))$ with a gradient bounded by $g_{\boldsymbol{\mu}}(\xi + te_\alpha)$. For a given $\xi \in \Omega_\alpha$, we can use Proposition 8.21 and Theorem 2.8 to get the existence of a velocity field $\mathbf{w}^\alpha_\xi \in L^2_{\boldsymbol{\mu}_\xi}([0,1] \times D, \mathbb{R}^q)$ such that $(\boldsymbol{\mu}_\xi, \mathbf{w}^\alpha_\xi \boldsymbol{\mu}_\xi)$ satisfies the (1-dimensional) continuity equation and for a.e. $t \in [0,1]$,

$$\sqrt{\int_D |\mathbf{w}^\alpha_\xi(t,x)|^2 \boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x)} \leqslant |\dot{\boldsymbol{\mu}}_\xi(t)| = |g_{\boldsymbol{\mu}_\xi}(t)| \leqslant g_{\boldsymbol{\mu}}(\xi + te_\alpha). \quad (8.10)$$

Now, do this for a.e. $\xi \in \Omega_\alpha$ and then for any $\alpha \in \{1, 2, \ldots, p\}$. Define the velocity field $\mathbf{v} : \Omega \times D \to \mathbb{R}^{pq}$ component by component, the α-th component at the point $\xi + te_\alpha$ (with $\xi \in \Omega_\alpha$) being defined as $\mathbf{w}^\alpha_\xi(t)$. To justify that $\mathbf{v}$ is measurable, notice that $\mathbf{w}^\alpha_\xi$ is the solution of an optimization problem [AGS08, Equation (8.3.11)] which depends in a measurable way of ξ, thus one can apply Proposition 8.45 (see below at the end of the chapter). By the bound (8.10), it is clear that $\mathbf{v} \in L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$. Moreover, if $\varphi \in C^1_c(\Omega \times D, \mathbb{R}^p)$,

$$\begin{aligned}
\iint\limits_{\Omega\times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} &= \sum_{\alpha=1}^{p} \iint\limits_{\Omega\times D} \partial_\alpha \varphi^\alpha(\xi, x)\boldsymbol{\mu}(\mathrm{d}\xi, \mathrm{d}x) \\
&= \sum_{\alpha=1}^{p} \int_{\Omega_\alpha} \left(\int_0^1 \partial_\alpha \varphi^\alpha(\xi + te_\alpha, x)\boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x)\mathrm{d}t \right) \mathrm{d}\xi \\
&= -\sum_{\alpha=1}^{p} \int_{\Omega_\alpha} \left(\int_0^1 \nabla_D \varphi^\alpha(\xi + te_\alpha, x) \cdot \mathbf{w}^\alpha_\xi(t,x)\boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x)\mathrm{d}t \right) \mathrm{d}\xi \\
&= -\sum_{\alpha=1}^{p} \int_{\Omega_\alpha} \left(\int_0^1 \nabla_D \varphi^\alpha(\xi + te_\alpha, x)(\mathbf{v}(\xi + te_\alpha, x) \cdot e_\alpha)\boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x)\mathrm{d}t \right) \mathrm{d}\xi \\
&= -\iint\limits_{\Omega\times D} \nabla_D \varphi \cdot \mathbf{v}\mathrm{d}\boldsymbol{\mu}.
\end{aligned}$$

(The second and last inequalities are Fubini's theorem and the third one comes from the 1-dimensional continuity equations). Hence, we see that $(\boldsymbol{\mu}, \mathbf{v}\boldsymbol{\mu})$ satisfies the continuity equation.

To conclude, we need to show a control of $\mathbf{v}$ by $g_{\boldsymbol{\mu}}$. If $\alpha \in \{1, 2, \ldots, p\}$, for a.e. $\xi \in \Omega_\alpha$ and a.e. $t \in [0, 1]$, one has, by definition of $g_{\boldsymbol{\mu}_\xi}$ and Proposition 8.21,

$$\sqrt{\int_D |\mathbf{w}^\alpha_\xi(t,x)|^2 \boldsymbol{\mu}(\xi + te_\alpha, \mathrm{d}x)} = g_{\boldsymbol{\mu}_\xi}(t) = \sup_{n\in\mathbb{N}} |\partial_\alpha[\boldsymbol{\mu}]_{\nu_n}(\xi + te_\alpha)|,$$

which can be rewritten as: for a.e. $\xi \in \Omega$, for all $\alpha \in \{1, 2, \ldots, p\}$,

$$\sqrt{\int_D |\mathbf{v}(\xi, x) \cdot e_\alpha|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x)} = \sup_{n\in\mathbb{N}} |\nabla[\boldsymbol{\mu}]_{\nu_n}(\xi)e_\alpha| \leqslant g_{\boldsymbol{\mu}}(\xi). \tag{8.11}$$

Squaring, summing over α and taking the square root, we see that for a.e. $\xi \in \Omega$

$$\sqrt{\int_D |\mathbf{v}(\xi, x)|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x)} \leqslant \sqrt{p} g_{\boldsymbol{\mu}}(\xi).$$

Even though one could prove that $\mathbf{v}$ is the tangent velocity field (using the fact that the $\mathbf{w}^\alpha$ are and the characterization given in Proposition 8.11), it is enough to use Corollary 8.12 to see that the l.h.s. is a.e. larger than the $L^2_{\boldsymbol{\mu}(\xi)}(D, \mathbb{R}^{pq})$-norm of the tangent velocity field. □

To conclude the proof of the theorem, we just have to justify that we can put the pieces together.

Proof of Theorem 8.20. The domain $\mathring{\Omega}$ can be cut in a (countable) number of cubes $(\Omega_m)_{m\in\mathbb{N}}$. The boundary $\partial\Omega$ does not play any role as $\mathcal{L}_\Omega(\partial\Omega) = 0$.

Implication $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty \Rightarrow \boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. Assume first that $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ is such that $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ and take $\mathbf{v} \in L^2_{\boldsymbol{\mu}}(\Omega \times D, \mathbb{R}^{pq})$ the velocity field tangent to $\boldsymbol{\mu}$. Fix $n \in \mathbb{N}$. On each cube Ω_m, we know that the function $[\boldsymbol{\mu}]_{\nu_n}$ is in $H^1(\Omega_m)$ with a gradient which is bounded by a function which does not depend on n and is in $L^2(\Omega)$, which is sufficient to say that $[\boldsymbol{\mu}]_{\nu_n} \in H^1(\Omega)$ with a gradient bounded by a function which does not depend on $n \in \mathbb{N}$.

Implication $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D)) \Rightarrow \mathrm{Dir}(\boldsymbol{\mu}) < +\infty$. Assume that $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. For any $m \in \mathbb{N}$, one can construct a tangent velocity field $\mathbf{v} \in L^2_{\boldsymbol{\mu}}(\Omega_m \times D, \mathbb{R}^{pq})$. Combining Proposition 8.9 giving the uniqueness $\boldsymbol{\mu}$-a.e. of the tangent velocity field and Corollary 8.12 which enables to localize, one sees that if $\Omega_{m_1} \cap \Omega_{m_2} \neq \varnothing$, then the tangent velocity fields $\mathbf{v}_1 \in L^2_{\boldsymbol{\mu}}(\Omega_{m_1} \times D, \mathbb{R}^{pq})$ and $\mathbf{v}_2 \in L^2_{\boldsymbol{\mu}}(\Omega_{m_2} \times D, \mathbb{R}^{pq})$ coincide $\boldsymbol{\mu}$-a.e. on $\Omega_{m_1} \cap \Omega_{m_2}$. Thus, one can define a velocity field $\mathbf{v}$ on the whole Ω, and it is straightforward to check that $\mathbf{v}$ is tangent to $\boldsymbol{\mu}$. □

8.1.4 Equivalence with Dirichlet energy in metric spaces

In this subsection we will show that our definition coincides with the one of Korevaar, Schoen, and Jost [KS93, Jos94]. As explained in the introduction, their formulation goes as follows.

Definition 8.24. *Let $\varepsilon > 0$ and $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$. We define the ε-Dirichlet energy of $\boldsymbol{\mu}$ by*

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) := C_p \iint_{\Omega\times\Omega} \frac{W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta))}{2\varepsilon^{p+2}} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \mathrm{d}\xi \mathrm{d}\eta,$$

where the normalization constant C_p is defined as $C_p := |\eta|^2 \left(\int_{B(0,1)} |\xi \cdot \eta|^2 \mathrm{d}\xi \right)^{-1}$.

One can notice that the ε-Dirichlet energy is always finite as $\mathcal{P}(D)$ has a finite diameter, but it can blow up when $\varepsilon \to 0$. The goal is to prove that Dir_ε is a good approximation of Dir if ε is small enough. Before stating the main result, let us do the following observation, which will be useful in the sequel.

Proposition 8.25. *Let $\varepsilon > 0$ be fixed. Then the functional $\mathrm{Dir}_\varepsilon : L^2(\Omega, \mathcal{P}(D)) \to \mathbb{R}$ is continuous w.r.t. strong convergence and l.s.c. w.r.t. the weak convergence.*

Proof. The continuity w.r.t. strong convergence is simple: recall that $\mathcal{P}(D)$ has a finite diameter, thus Lebesgue dominated convergence theorem is enough. The lower semi-continuity relies on the fact that W_2^2 is a supremum of continuous linear functionals, thus is l.s.c. and convex.

More precisely, fix $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ and a sequence $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ which converges weakly to $\boldsymbol{\mu}$ in $L^2(\Omega, \mathcal{P}(D))$. If ξ and η are points of Ω, take $(\varphi(\xi, \eta, \cdot), \psi(\xi, \eta, \cdot))$ a pair of Kantorovich potential between $\boldsymbol{\mu}(\xi)$ and $\boldsymbol{\mu}(\eta)$. In other words, $\varphi(\xi, \eta, \cdot)$ and $\psi(\xi, \eta, \cdot)$ are continuous functions (in fact uniformly Lipschitz), such that $\varphi(\xi, \eta, x) + \psi(\xi, \eta, y) \leqslant |x - y|^2/2$ for any $x, y \in D$, and such that

$$\frac{1}{2} W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta)) = \int_D \varphi(\xi, \eta, x) \boldsymbol{\mu}(\xi, \mathrm{d}x) + \int_D \psi(\xi, \eta, y) \boldsymbol{\mu}(\eta, \mathrm{d}y). \tag{8.12}$$

One can do that in such a way that $\varphi : \Omega \times \Omega \to C(D)$ and $\psi : \Omega \times \Omega \to C(D)$ are measurable. Indeed, for fixed ξ and η, $(\varphi(\xi, \eta, \cdot), \psi(\xi, \eta, \cdot)) \in C(D) \times C(D)$ is a maximizer a functional which is continuous on $C(D) \times C(D)$ and which depends on ξ and η in a measurable way: hence we can apply Proposition 8.45 (see below at the end of the chapter). Then, using the double convexification trick (see [Vil03, Section 2.1]) which is a measurable operation, we can assume that (φ, ψ) are uniformly (w.r.t. ξ and η) Lipschitz and bounded as elements of $C(D)$. By the Kantorovich duality, for every $n \in \mathbb{N}$,

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_n) \geqslant \frac{C_p}{\varepsilon^{p+2}} \iint_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \varphi(\xi, \eta, x) \boldsymbol{\mu}_n(\xi, \mathrm{d}x) + \int_D \psi(\xi, \eta, y) \boldsymbol{\mu}_n(\eta, \mathrm{d}y) \right) \mathrm{d}\xi \mathrm{d}\eta. \tag{8.13}$$

Now, apply Lusin's theorem to the mapping $\varphi : \Omega \times \Omega \to C(D)$ (for Lusin's theorem to other spaces than $\mathbb{R}$, see for instance [San15, Box 1.6]). For any $\delta > 0$, we can find a compact $X \subset \Omega \times \Omega$ such that $\mathcal{L}_\Omega \otimes \mathcal{L}_\Omega([\Omega \times \Omega] \backslash X) \leqslant \delta$ and $\varphi : X \to C(D)$ is continuous on X. Now notice, as $|\varphi(\xi, \eta, x) - \varphi(\xi, \eta, y)| \leqslant C|x - y|$ uniformly in ξ and η, that $\varphi : X \times D \to \mathbb{R}$ is a continuous function for the product topology on $X \times D \subset \Omega \times \Omega \times D$. This function can be extended in a function $\tilde{\varphi} \in C(\Omega \times \Omega \times D)$. To sum up, there exists a continuous function $\tilde{\varphi}$, which coincides with φ on $X \times D$ (the important point is that there is coincidence on all D). Thus, denoting by

C a uniform bound of φ and $\tilde{\varphi}$, one has that for every $\boldsymbol{\nu} \in L^2(\Omega, \mathcal{P}(D))$,

$$\left| \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \varphi(\xi,\eta,x)\boldsymbol{\nu}(\xi,\mathrm{d}x) \right) \mathrm{d}\xi\mathrm{d}\eta - \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \tilde{\varphi}(\xi,\eta,x)\boldsymbol{\nu}(\xi,\mathrm{d}x) \right) \mathrm{d}\xi\mathrm{d}\eta \right| \leqslant C\delta. \tag{8.14}$$

On the other hand, using Fubini's theorem one sees that

$$\iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \tilde{\varphi}(\xi,\eta,x)\boldsymbol{\mu}_n(\xi,\mathrm{d}x) \right) \mathrm{d}\xi\mathrm{d}\eta = \iint\limits_{\Omega\times D} \left(\int_{B(\xi,\varepsilon)\cap\Omega} \tilde{\varphi}(\xi,\eta,x)\mathrm{d}\eta \right) \boldsymbol{\mu}_n(\mathrm{d}\xi,\mathrm{d}x).$$

As $\tilde{\varphi}$ is continuous and bounded, it is not difficult to see that

$$(\xi,x) \in \Omega\times D \mapsto \int_{B(\xi,\varepsilon)\cap\Omega} \tilde{\varphi}(\xi,\eta,x)\mathrm{d}\eta \in \mathbb{R}$$

is continuous. Hence, using the weak convergence of $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$,

$$\lim_{n\to+\infty} \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \tilde{\varphi}(\xi,\eta,x)\boldsymbol{\mu}_n(\xi,\mathrm{d}x) \right) \mathrm{d}\xi\mathrm{d}\eta = \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \tilde{\varphi}(\xi,\eta,x)\boldsymbol{\mu}(\xi,\mathrm{d}x) \right) \mathrm{d}\xi\mathrm{d}\eta.$$

Using equation (8.14) with both $\boldsymbol{\mu}_n$ and $\boldsymbol{\mu}$ as $\boldsymbol{\nu}$, and using moreover the arbitrariness of δ, we conclude that we can replace $\tilde{\varphi}$ by φ in the equation above:

$$\lim_{n\to+\infty} \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \varphi(\xi,\eta,x)\boldsymbol{\mu}_n(\xi,\mathrm{d}x) \right) \mathrm{d}\xi\mathrm{d}\eta = \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \varphi(\xi,\eta,x)\boldsymbol{\mu}(\xi,\mathrm{d}x) \right) \mathrm{d}\xi\mathrm{d}\eta.$$

Of course there is exactly the same statement with ψ. With the help of this information, combining (8.13) and (8.12), we reach the conclusion that

$$\begin{aligned}
&\liminf_{n\to+\infty} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_n) \\
&\qquad \geqslant \liminf_{n\to+\infty} \frac{C_p}{\varepsilon^{p+2}} \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \varphi(\xi,\eta,x)\boldsymbol{\mu}_n(\xi,\mathrm{d}x) + \int_D \psi(\xi,\eta,y)\boldsymbol{\mu}_n(\eta,\mathrm{d}y) \right) \mathrm{d}\xi\mathrm{d}\eta \\
&\qquad = \frac{C_p}{\varepsilon^{p+2}} \iint\limits_{\Omega\times\Omega} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \left(\int_D \varphi(\xi,\eta,x)\boldsymbol{\mu}(\xi,\mathrm{d}x) + \int_D \psi(\xi,\eta,y)\boldsymbol{\mu}(\eta,\mathrm{d}y) \right) \mathrm{d}\xi\mathrm{d}\eta \\
&\qquad = \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}).
\end{aligned}$$

□

We are now ready to state and prove the main theorem of this subsection.

Theorem 8.26. *Let* $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$*. Then* $\mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$ *converges to* $\mathrm{Dir}(\boldsymbol{\mu})$ *as* $\varepsilon \to 0$*, and the sequence* $(\mathrm{Dir}_{2^{-n}\varepsilon_0}(\boldsymbol{\mu}))_{n\in\mathbb{N}}$ *is increasing for any* $\varepsilon_0 > 0$*.*

In addition for any $\varepsilon_0 > 0$*,* $\mathrm{Dir}_{2^{-n}\varepsilon_0}$ *Γ-converges to* Dir *on the space* $L^2(\Omega, \mathcal{P}(D))$ *endowed with the weak topology as* $n \to +\infty$*.*

In the case of a smooth mapping $\boldsymbol{\mu}$, the equivalence will directly derives from Proposition 8.17. The difficulty of the proof is to study the behavior of Dir_ε w.r.t. approximations.

Proof. Monotonicity of Dir_ε. If $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$, $\varepsilon > 0$ and $\lambda \in (0,1)$ then one has

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) \leqslant \lambda \mathrm{Dir}_{\lambda\varepsilon}(\boldsymbol{\mu}) + (1-\lambda)\mathrm{Dir}_{(1-\lambda)\varepsilon}(\boldsymbol{\mu}).$$

Indeed, this is a consequence of the triangle inequality and is valid for mappings valued in arbitrary metric spaces, see for instance [Jos94, Example 1) (i)] or [Jos08, Equation (8.3.4)] for a proof. In particular, by taking $\lambda = 1/2$, we see that the sequence $(\mathrm{Dir}_{2^{-n}\varepsilon_0}(\boldsymbol{\mu}))_{n\in\mathbb{N}}$ is increasing for any $\varepsilon_0 > 0$. Moreover, with well chosen λ, one sees that for a fixed $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ the function $\varepsilon \mapsto \varepsilon \mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$ is subadditive, which is enough to ensure the convergence of $\mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$ to some limit in $[0, +\infty]$ as $\varepsilon \to 0$.

The smooth case. Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ be smooth in the sense of Definition 8.15. Let $\mathbf{v}$ be its tangent velocity field, by Proposition 8.16, there holds $\mathbf{v} \in C(\Omega, L^2(D, \mathbb{R}^{pq}))$. We will show that the limit of $\mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$ is equal to $\mathrm{Dir}(\boldsymbol{\mu})$. Indeed, one can write

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) = \int_\Omega \mathrm{dir}_\varepsilon(\xi)\mathrm{d}\xi,$$

where

$$\mathrm{dir}_\varepsilon(\xi) := C_p \int_{\Omega\cap B(\xi,\varepsilon)} \frac{W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta))}{2\varepsilon^{p+2}}\mathrm{d}\eta.$$

If $\xi \notin \partial\Omega$ (it happens for a.e. ξ), for ε small enough, $B(\xi, \varepsilon) \subset \Omega$ and we can perform the following change of variables in spherical coordinates: denoting by $\mathbb{S}^{p-1}$ the unit sphere of $\mathbb{R}^p$ and σ its surface measure,

$$\mathrm{dir}_\varepsilon(\xi) = \frac{C_p}{2}\int_{\mathbb{S}^{d-1}}\left(\int_0^1 \frac{W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\xi + r\varepsilon\theta))}{\varepsilon^2} r^{p-1}\mathrm{d}r\right)\sigma(\mathrm{d}\theta).$$

Thanks to Proposition 8.17 we have the pointwise limit of the integrand, and we can pass to the limit as $\varepsilon \to 0$: recall that $\boldsymbol{\mu}$ is Lipschitz, which gives a uniform bound from above of the Wasserstein distances. Hence, for a.e. $\xi \in \Omega$,

$$\begin{aligned}
\lim_{\varepsilon\to 0} \mathrm{dir}_\varepsilon(\xi) &= \frac{C_p}{2}\int_{\mathbb{S}^{d-1}}\left[\int_0^1\left(\int_D |\mathbf{v}(\xi, x)\cdot(r\theta)|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x)\right) r^{p-1}\mathrm{d}r\right]\sigma(\mathrm{d}\theta) \\
&= \frac{C_p}{2}\int_D\left(\int_{B(0,1)} |\mathbf{v}(\xi, x)\cdot\eta|^2\mathrm{d}\eta\right)\boldsymbol{\mu}(\xi, \mathrm{d}x) \\
&= \frac{1}{2}\int_D |\mathbf{v}(\xi, x)|^2\boldsymbol{\mu}(\xi, \mathrm{d}x),
\end{aligned}$$

where the last inequality comes from the definition of C_p. To integrate this equality over Ω, we still use the fact that $\boldsymbol{\mu}$ is Lipschitz to get the appropriate bounds, hence

$$\lim_{\varepsilon\to 0}\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) = \int_\Omega\left(\lim_{\varepsilon\to 0}\mathrm{dir}_\varepsilon(\xi)\right)\mathrm{d}\xi = \int_\Omega\left(\int_D \frac{1}{2}|\mathbf{v}(\xi, x)|^2\boldsymbol{\mu}(\xi, \mathrm{d}x)\right)\mathrm{d}\xi = \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{v}\boldsymbol{\mu}) = \mathrm{Dir}(\boldsymbol{\mu}).$$

General case: $\lim_\varepsilon \mathrm{Dir}_\varepsilon \leqslant \mathrm{Dir}$. Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. As $\boldsymbol{\mu}$ is in H^1 in the sense of Reshetnyak, and using the main result of [Res04], we know that $l := \lim_\varepsilon \mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$ is finite. It implies, thanks to the theory of Korevaar and Schoen [KS93, Theorem 1.10], that the so-called energy density is absolutely continuous w.r.t. $\mathcal{L}_\Omega$ which means $\lim_\varepsilon \mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$ does not decrease too much if we restrict $\boldsymbol{\mu}$ to a domain $\tilde{\Omega}$ slightly smaller than Ω. More precisely, it implies that for any δ there exists $\tilde{\Omega}$ compactly embedded in $\mathring{\Omega}$ such that, for some ε_0 small enough,

$$l - \delta \leqslant \mathrm{Dir}_{\varepsilon_0}(\boldsymbol{\mu}|_{\tilde{\Omega}}) \leqslant l.$$

Let $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ the sequence of elements of $L^2(\tilde{\Omega}, \mathcal{P}(D))$ given by Theorem 8.14. We choose n large enough so that $\mathrm{Dir}(\boldsymbol{\mu}_n) \leqslant \mathrm{Dir}(\boldsymbol{\mu}|_{\tilde{\Omega}}) + \delta$ and $\mathrm{Dir}_{\varepsilon_0}(\boldsymbol{\mu}_n) \geqslant \mathrm{Dir}_{\varepsilon_0}(\boldsymbol{\mu}|_{\tilde{\Omega}}) - \delta$: it is possible because $\mathrm{Dir}_{\varepsilon_0}$ is lower semi-continuous w.r.t. weak convergence on $L^2(\tilde{\Omega}, \mathcal{P}(D))$. Hence,

$$l \leqslant \mathrm{Dir}_{\varepsilon_0}(\boldsymbol{\mu}|_{\tilde{\Omega}}) + \delta \leqslant \mathrm{Dir}_{\varepsilon_0}(\boldsymbol{\mu}_n) + 2\delta \leqslant \mathrm{Dir}(\boldsymbol{\mu}_n) + 2\delta \leqslant \mathrm{Dir}(\boldsymbol{\mu}|_{\tilde{\Omega}}) + 3\delta \leqslant \mathrm{Dir}(\boldsymbol{\mu}) + 3\delta,$$

where the third inequality comes from monotonicity and the smooth case treated above. As δ is arbitrary, we get that $l \leqslant \mathrm{Dir}(\boldsymbol{\mu})$, which means

$$\lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) \leqslant \mathrm{Dir}(\boldsymbol{\mu}).$$

This equation still holds if $\boldsymbol{\mu} \notin H^1(\Omega, \mathcal{P}(D))$ as the r.h.s. is infinite.

General case: $\lim_\varepsilon \mathrm{Dir}_\varepsilon \geqslant \mathrm{Dir}$. For this part, we need to control in a fine way the behavior of Dir_ε w.r.t. the approximation procedure of Theorem 8.14. Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ be given. Fix $\tilde{\Omega} \subset \mathring{\Omega}$ compactly included and let $\tilde{\boldsymbol{\mu}}_n, \boldsymbol{\mu}_n$ the sequences used in the proof of Theorem 8.14. We recall that they are defined by

$$\tilde{\boldsymbol{\mu}}_n(\xi) := [\Phi^D_{1/n}][\boldsymbol{\mu}(\xi)], \quad \boldsymbol{\mu}_n(\xi) := \int_\Omega \chi_n(\xi - \eta)\tilde{\boldsymbol{\mu}}_n(\eta)\mathrm{d}\eta,$$

where $\chi_n : \mathbb{R}^p \to \mathbb{R}$ is a compactly supported convolution kernel and $\boldsymbol{\mu}_n$ is defined only over $\tilde{\Omega}$. Using the result for the smooth case,

$$\lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_n|_{\tilde{\Omega}}) = \mathrm{Dir}(\boldsymbol{\mu}_n|_{\tilde{\Omega}}). \tag{8.15}$$

As the heat flow is a contraction in the Wasserstein space (Proposition 2.13), we know that $\mathrm{Dir}_\varepsilon(\tilde{\boldsymbol{\mu}}_n) \leqslant \mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$. As W_2^2 is jointly convex w.r.t. to its two arguments, the function Dir_ε is convex for the affine structure on $L^2(\tilde{\Omega}, \mathcal{P}(D))$. Hence, exactly by the same argument than in the proof of Theorem 8.14,

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_n) \leqslant \mathrm{Dir}_\varepsilon(\tilde{\boldsymbol{\mu}}_n) \leqslant \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}),$$

and the important point is that the r.h.s. does not depend on n. Taking the limit $\varepsilon \to 0$ and using equation (8.15), we see that

$$\mathrm{Dir}(\boldsymbol{\mu}_n) = \lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_n) \leqslant \lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}).$$

Now we can send $n \to +\infty$ and use Theorem 8.14 to say that the l.h.s. converges to $\mathrm{Dir}(\boldsymbol{\mu}|_{\tilde{\Omega}})$. As $\tilde{\Omega}$ is now arbitrary, it yields the result

$$\mathrm{Dir}(\boldsymbol{\mu}) \leqslant \lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}).$$

In the case $\boldsymbol{\mu} \notin H^1(\Omega, \mathcal{P}(D))$, to justify that $\lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$, we can use for instance [Chi07, Proposition 4] which is valid for mappings valued in arbitrary metric spaces.

The Γ*-convergence.* The statement of Γ-convergence is now easy. To summarize, until now we have proved the monotonicity and that

$$\mathrm{Dir}(\boldsymbol{\mu}) = \lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$$

for every $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$. It is an exercise that we leave to the reader to check that any sequence of functionals which are l.s.c. (which is the case for the Dir_ε, see Proposition 8.25) and which converges in a increasing way in fact Γ-converges. □

8.1.5 Boundary values

It is well known that it is possible to make sense of the values of a H^1 real-valued function on hypersurfaces, in particular to give a meaning to the values of such a function on the boundary of a domain. As we want to define the Dirichlet problem, which consists in minimizing the Dirichlet energy with fixed values on the boundary $\partial\Omega$, we need to give a meaning to the boundary values of elements of $H^1(\Omega, \mathcal{P}(D))$. Korevaar and Schoen have already developed a trace theory in a fairly general context [KS93, Section 1.12]. However, in our specific situation and in view of proving the dual formulation of the Dirichlet problem, we will define the boundary values by showing how one can extend the continuity equation for test functions $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$ which are no longer compactly supported in $\mathring{\Omega}$. Even if we do not prove it in this article, our definition of trace coincides with the one of [KS93]: to be convinced one can look at Proposition 9.6 and compare it to [KS93, Theorem 1.12.3]. Recall that $\mathbf{n}_\Omega$ denotes the outward normal to $\partial\Omega$.

Theorem 8.27. *Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. Then there exists a vector-valued measure $\mathrm{BT}_{\boldsymbol{\mu}} \in \mathcal{M}(\Omega \times D, \mathbb{R}^p)$ supported on $\mathbf{n}_\Omega \partial\Omega \times D$ (which means that $\mathrm{BT}_{\boldsymbol{\mu}}(\varphi) = 0$ if $\varphi \cdot \mathbf{n}_\Omega = 0$ on $\partial\Omega \times D$) such that for any $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$ and for any $\mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ for which $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation and $\mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) < +\infty$,*

$$\iint_{\Omega \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint_{\Omega \times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E} = \mathrm{BT}_{\boldsymbol{\mu}}(\varphi). \tag{8.16}$$

Moreover if $\boldsymbol{\mu}$ is continuous as a mapping valued in $(\mathcal{P}(D), W_2)$ then for any $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$,

$$\mathrm{BT}_{\boldsymbol{\mu}}(\varphi) = \int_{\partial\Omega} \left(\int_D \varphi(\xi, x) \cdot \mathbf{n}_\Omega(\xi) \boldsymbol{\mu}(\xi, \mathrm{d}x) \right) \sigma(\mathrm{d}\xi),$$

where σ is the surface measure on $\partial\Omega$.

$\mathrm{BT}_{\boldsymbol{\mu}}$ stands for "Boundary Term" of $\boldsymbol{\mu}$. It is not surprising that, if $\boldsymbol{\mu}$ is continuous, the value of $\mathrm{BT}_{\boldsymbol{\mu}}$ depends only on the values of $\boldsymbol{\mu}$ on the boundary.

Proof. Take $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ and $\mathbf{E} = \mathbf{v}\boldsymbol{\mu} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ such that $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation and $\mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) < +\infty$. The l.h.s. of (8.16) defines a vector-valued distribution on $\Omega \times D$ acting on φ. We need to show that it is of order 0 and that it does not depend on $\mathbf{E}$.

Take $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$. We define $f : \Omega \to \mathbb{R}^p$ by, for a.e. $\xi \in \Omega$,

$$f(\xi) := \int_D \varphi(\xi, x) \boldsymbol{\mu}(\xi, \mathrm{d}x).$$

Using the continuity equation with test functions of the form $\chi\varphi^\alpha$, for $\chi \in C^1_c(\mathring{\Omega}, \mathbb{R}^p)$ and $\alpha \in \{1, 2, \ldots, p\}$, one can see that $f \in H^1(\mathring{\Omega}, \mathbb{R}^p)$ and

$$\partial_\alpha f^\beta(\xi) = \int_D \partial_\alpha \varphi^\beta(\xi, x) \boldsymbol{\mu}(\xi, \mathrm{d}x) + \int_D \nabla_D \varphi^\beta(\xi, x) \cdot \mathbf{v}^\alpha(\xi, x) \boldsymbol{\mu}(\xi, \mathrm{d}x).$$

for all $\alpha, \beta \in \{1, 2, \ldots, p\}$. In particular f admits on $\partial\Omega$ a trace $\bar{f} : \partial\Omega \to \mathbb{R}^p$. We apply the divergence theorem: one can find in [EG92, Section 4.3] a statement when $\partial\Omega$ is only Lipschitz and f has Sobolev regularity. In our case, given the expression of ∇f, it reads

$$\iint_{\Omega \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint_{\Omega \times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E} = \int_\Omega \nabla \cdot f = \int_{\partial\Omega} \bar{f}(\xi) \cdot \mathbf{n}_\Omega(\xi) \sigma(\mathrm{d}\xi) \tag{8.17}$$

where $\mathbf{n}_\Omega$ is the outward normal to $\partial\Omega$ and σ its the surface measure. In particular we see that the r.h.s. of (8.16) does not depend on $\mathbf{E}$. Moreover, as $\|f\|_\infty \leqslant \|\varphi\|_\infty$, the same L^∞ bounds holds for $\bar{f}$, thus

$$\left| \int_{\partial\Omega} \bar{f}(\xi) \cdot \mathbf{n}_\Omega(\xi)\sigma(\mathrm{d}\xi) \right| \leqslant \sigma(\partial\Omega)\|\varphi\|_\infty.$$

It allows to conclude that the l.h.s. of (8.16) is a distribution of order 0 acting on φ, hence it can be represented by a measure $\mathrm{BT}_{\boldsymbol{\mu}} \in \mathcal{M}(\Omega \times D, \mathbb{R}^p)$. From (8.17) it is clear that $\mathrm{BT}_{\boldsymbol{\mu}}$ is supported on $\mathbf{n}_\Omega\partial\Omega \times D$.

If we assume moreover that $\boldsymbol{\mu}$ is continuous, so is f. Indeed, for any $\xi, \eta \in \Omega$,

$$\begin{aligned}
|f(\xi) - f(\eta)| &= \left| \int_D \varphi(\xi, x)\boldsymbol{\mu}(\xi, \mathrm{d}x) - \int_D \varphi(\eta, x)\boldsymbol{\mu}(\eta, \mathrm{d}x) \right| \\
&\leqslant \int_D |\varphi(\xi, x) - \varphi(\eta, x)|\boldsymbol{\mu}(\xi, \mathrm{d}x) + \left| \int_D \varphi(\eta, x)\boldsymbol{\mu}(\xi, \mathrm{d}x) - \int_D \varphi(\eta, x)\boldsymbol{\mu}(\eta, \mathrm{d}x) \right| \\
&\leqslant \|\nabla_\Omega\varphi\|_\infty|\xi - \eta| + \left| \int_D \varphi(\eta, x)\boldsymbol{\mu}(\xi, \mathrm{d}x) - \int_D \varphi(\eta, x)\boldsymbol{\mu}(\eta, \mathrm{d}x) \right|.
\end{aligned}$$

When $\xi \to \eta$, the first term obviously goes to 0, and the second one too by definition of the weak convergence (by assumption $\boldsymbol{\mu}(\xi) \to \boldsymbol{\mu}(\eta)$ in the weak sense). Thus $\bar{f}$ coincides with f, which gives the announced result. □

If $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$, using the disintegration theorem and testing against well chosen functions, one can show that there exists $\bar{\boldsymbol{\mu}} : \partial\Omega \to \mathcal{P}(D)$ defined σ-a.e. such that $\mathrm{BT}_{\boldsymbol{\mu}} = \mathbf{n}_\Omega\bar{\boldsymbol{\mu}} \otimes \sigma$. The mapping $\bar{\boldsymbol{\mu}}$ can be seen as a definition of the values of $\boldsymbol{\mu}$ on $\partial\Omega$.

Now we can define what it means to share the same boundary values and prove that the set of $\boldsymbol{\mu}$ with fixed boundary values is closed.

Definition 8.28. *Let $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ two elements of $H^1(\Omega, \mathcal{P}(D))$. We say that $\boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\nu}|_{\partial\Omega}$ if $\mathrm{BT}_{\boldsymbol{\mu}} = \mathrm{BT}_{\boldsymbol{\nu}}$.*

Proposition 8.29. *Let $\boldsymbol{\mu}_b \in H^1(\Omega, \mathcal{P}(D))$ and $C \in \mathbb{R}$ be fixed. Then the set*

$$\{\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D)) \ : \ \boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega} \text{ and } \mathrm{Dir}(\boldsymbol{\mu}) \leqslant C\}$$

is closed for the weak topology on $L^2(\Omega, \mathcal{P}(D))$.

Proof. The proof is straightforward. Indeed, take a sequence $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ in $\in L^2(\Omega, \mathcal{P}(D))$ such that $\boldsymbol{\mu}_n|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega}$ and $\mathrm{Dir}(\boldsymbol{\mu}_n) \leqslant C$ for any $n \in \mathbb{N}$, and assume it converges weakly to some $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$. By lower semi-continuity of Dir, we know that $\mathrm{Dir}(\boldsymbol{\mu}) \leqslant C$. For any $n \in \mathbb{N}$ choose $\mathbf{E}_n \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ tangent to $\boldsymbol{\mu}_n$, similarly take $\mathbf{E}_b$ tangent to $\boldsymbol{\mu}_b$. The identity $\boldsymbol{\mu}_n|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega}$ can be written: for every $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$

$$\iint\limits_{\Omega\times D} \nabla \cdot \varphi \mathrm{d}\boldsymbol{\mu}_n + \iint\limits_{\Omega\times D} \nabla_D\varphi \cdot \mathrm{d}\mathbf{E}_n = \iint\limits_{\Omega\times D} \nabla \cdot \varphi \mathrm{d}\boldsymbol{\mu}_b + \iint\limits_{\Omega\times D} \nabla_D\varphi \cdot \mathrm{d}\mathbf{E}_b. \tag{8.18}$$

Notice that this simple estimate holds for the total mass of $\mathbf{E}_n$: provided $\mathbf{v}_n$ is the tangent velocity field to $\boldsymbol{\mu}_n$,

$$|\mathbf{E}_n|(\Omega \times D) = \|\mathbf{v}_n\|_{L^1_{\boldsymbol{\mu}}(\Omega\times D, \mathbb{R}^{pq})} \leqslant C_1\|\mathbf{v}_n\|_{L^2_{\boldsymbol{\mu}}(\Omega\times D, \mathbb{R}^{pq})} \leqslant C_1\sqrt{2\mathrm{Dir}(\boldsymbol{\mu}_n)} \leqslant C_1\sqrt{2C}.$$

Hence one can assume that, up to extraction, $(\mathbf{E}_n)_{n\in\mathbb{N}}$ weakly converges to some $\mathbf{E}$. It is easy to see that $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation and that $\mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) \leqslant C < +\infty$. Thus, we can pass to the limit in (8.18) and see that for any $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$,

$$\iint_{\Omega\times D} \nabla \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint_{\Omega\times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E} = \iint_{\Omega\times D} \nabla \cdot \varphi \mathrm{d}\boldsymbol{\mu}_b + \iint_{\Omega\times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E}_b,$$

which exactly means that $\boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega}$. □

8.2 The Dirichlet problem and its dual

8.2.1 Statement of the problem

With all the tools at our disposal, we are ready to state the Dirichlet problem. It simply consists in minimizing the Dirichlet energy under the constraint that the values at the boundary are fixed.

Definition 8.30. *Let $\boldsymbol{\mu}_b \in H^1(\Omega, \mathcal{P}(D))$. Then the Dirichlet problem with boundary values $\boldsymbol{\mu}_b$ is defined as*

$$\min_{\boldsymbol{\mu}} \left\{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D)) \text{ and } \boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega} \right\}.$$

A mapping $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ which realizes the minimum is called a solution of the Dirichlet problem.

Definition 8.31. *Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. We say that $\boldsymbol{\mu}$ is harmonic if it is a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}$.*

With the work of the previous section, the existence of at least one solution is a straightforward application of the direct method of calculus of variations.

Theorem 8.32. *Let $\boldsymbol{\mu}_b \in H^1(\Omega, \mathcal{P}(D))$. Then there exists at least one solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_b$.*

Proof. There exists at least one $\boldsymbol{\mu}$ with finite Dirichlet energy which satisfies the boundary conditions, namely $\boldsymbol{\mu}_b$. Thus, one can consider a minimizing sequence $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$. By compactness of $L^2(\Omega, \mathcal{P}(D))$, we can assume, up to extraction, that this sequence converges weakly to some $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$. By Proposition 8.29, we know that $\boldsymbol{\mu}$ also satisfies $\boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega}$. The lower semi-continuity of Dir allows to conclude that $\boldsymbol{\mu}$ is a minimizer of Dir. □

Let us spend a few words about the question of uniqueness, more is said in Chapter 12. Of course, the proof above provides no information about it. By convexity of the Dirichlet energy (Proposition 8.13), we know that the set of solutions of the Dirichlet problem is convex. However, Dir is *not* strictly convex. Recall that if $\Omega = [0,1]$ is a segment of $\mathbb{R}$, then the Dirichlet problem reduces to the problem of finding a geodesic between the two endpoints $\boldsymbol{\mu}_b(0)$ and $\boldsymbol{\mu}_b(1)$. It is well known that a sufficient condition for uniqueness is to impose that either $\boldsymbol{\mu}_b(0)$ or $\boldsymbol{\mu}_b(1)$ are absolutely continuous w.r.t. $\mathcal{L}_D$, and there can be non uniqueness when it is not the case. Hence, it would natural, in order to investigate the question of uniqueness, to impose that for every $\xi \in \partial\Omega$, the measure $\boldsymbol{\mu}_b(\xi)$ is absolutely continuous w.r.t. $\mathcal{L}_D$. We do not know if uniqueness holds under this hypothesis: a difference with the case where Ω is a segment is the fact that we do not know a static or Lagrangian formulation. In other words, we do not know the equivalent of transport plans, which in the case of a 1-dimensional Ω, allow to parametrize geodesics and to greatly simplify the problem. However we are able to prove uniqueness in a non trivial case: the one of a family of elliptically contoured distributions treated in Subsection 10.3.

8.2.2 Lipschitz extension

To give ourselves the boundary conditions, we need a mapping $\boldsymbol{\mu}_b$ defined on the whole Ω, even though only its values near $\partial\Omega$ will play a role. Thus a natural question arises: if $\boldsymbol{\mu}_b$ is only defined on $\partial\Omega$, is it possible to extend it on Ω? The next theorem shows that the answer is positive in the case where $\boldsymbol{\mu}_b$ is Lipschitz on $\partial\Omega$. Indeed, in this case we can build an extension which is Lipschitz on Ω, thus in $H^1(\Omega, \mathcal{P}(D))$ thanks to Theorem 8.20.

Theorem 8.33. *Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(D)$ a Lipschitz mapping. Then there exists $\boldsymbol{\mu} : \Omega \to \mathcal{P}(D)$ Lipschitz such that $\boldsymbol{\mu}(\xi) = \boldsymbol{\mu}_l(\xi)$ for every $\xi \in \partial\Omega$.*

For a continuous $\boldsymbol{\mu}$ the boundary term $\mathrm{BT}_{\boldsymbol{\mu}}$ depends only on the values of $\boldsymbol{\mu}$ on $\partial\Omega$ (Theorem 8.27), hence the boundary term of the Lipschitz extension of $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(D)$ does not depend on the extension. In other words, the following problem is well defined:

Definition 8.34. *Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(D)$ a Lipschitz mapping. Then the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ is defined as the Dirichlet problem with boundary values $\boldsymbol{\mu}_b$, where $\boldsymbol{\mu}_b$ is any Lipschitz extension of $\boldsymbol{\mu}_l$ on Ω.*

Now, let us prove the Lipschitz extension theorem. It relies on the following Lemma, which allows to treat the case where Ω is a ball.

Lemma 8.35. *Let $B(0,1)$ be the unit ball of $\mathbb{R}^p$ and $\mathbb{S}^{p-1} := \partial B(0,1)$ its boundary. Let $\boldsymbol{\mu}_l : \mathbb{S}^{d-1} \to \mathcal{P}(D)$ a Lipschitz mapping and take $x_0 \in D$. Define, for any $r \in [0,1]$ the map $T_r : D \to D$ by $T_r(x) = rx + (1-r)x_0$. Then the mapping $\boldsymbol{\mu} : B(0,1) \to \mathcal{P}(D)$ defined by*

$$\boldsymbol{\mu}(r\xi) := T_r\#[\boldsymbol{\mu}(\xi)]$$

for any $r \in [0,1]$ and any $\xi \in \mathbb{S}^{d-1}$ is Lipschitz.

Proof. If $\xi \in \mathbb{S}^{d-1}$ is fixed, then $r \in [0,1] \mapsto \boldsymbol{\mu}(r\xi)$ is the constant speed geodesic joining δ_{x_0} to $\boldsymbol{\mu}_l(\xi)$. Hence, we can write that $W_2(\boldsymbol{\mu}(r\xi), \boldsymbol{\mu}(s\xi)) \leqslant C|r-s|$, where C depends only on the diameter of $\mathcal{P}(D)$. On the other hand, as T_r is r-Lipschitz in D, then $\nu \mapsto T_r\#\nu$ is also r-Lipschitz in $\mathcal{P}(D)$. Hence, for any ξ and η in $\mathbb{S}^{d-1}$, one has $W_2(\boldsymbol{\mu}(r\xi), \boldsymbol{\mu}(r\eta)) \leqslant Cr|\xi - \eta|$, where C is the Lipschitz constant of $\boldsymbol{\mu}_l$. Putting the two estimates together, we deduce that for any $r, s \in [0,1]$ and any $\xi, \eta \in \mathbb{S}^{p-1}$,

$$W_2(\boldsymbol{\mu}(r\xi), \boldsymbol{\mu}(s\eta)) \leqslant C[|r-s| + \min(r,s)|\xi - \eta|],$$

which is enough to conclude that $\boldsymbol{\mu}$ is Lipschitz. □

Notice that the Lipschitz constant of the extension is not controlled by the Lipschitz constant of $\boldsymbol{\mu}_l$: the distance between δ_{x_0} and the range of $\boldsymbol{\mu}_l$ also plays a role as $\boldsymbol{\mu}(0) = \delta_{x_0}$. Hence, we cannot use a decomposition with Withney cubes to extend mappings defined on arbitrary closed subsets Ω, but only on the boundary of smooth sets: basically we need to use Lemma 8.35 only a finite number of times.

Proof of Theorem 8.33. We will use Lemma 8.35 in the following form: if Ω is a domain which is in a bilipschitz bijection with a ball, then Theorem 8.33 holds for this domain.

We reason by induction on $p \geqslant 1$ the dimension of Ω. In dimension 1, $\Omega = I$ is a segment. To extend a mapping defined only on the boundary of the segment I, we take the constant speed geodesic in $\mathcal{P}(D)$ between the values of $\boldsymbol{\mu}_l$ at the two endpoints of I.

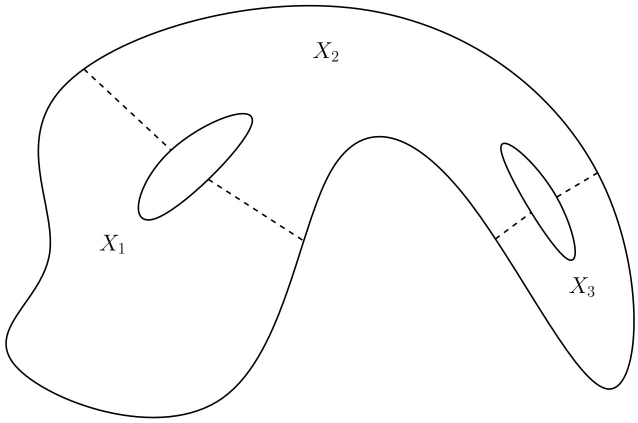

Figure 8.1: Idea of the proof of Theorem 8.33: every domain Ω with Lipschitz boundary (solid line), even with an intricate topology, can be decomposed in a finite number of pieces X_1, X_2, X_3 such that each of them is in a bilipschitz bijection with a ball. The boundaries between the pieces (dashed lines) are in bilipschitz bijection with balls of a smaller dimension (here segments).

Now assume that the result holds for some $p-1 \geqslant 1$ and let Ω be a compact domain with Lipschitz boundary in $\mathbb{R}^p$. The goal is to cut Ω in a finite number of pieces on which Lemma 8.35 apply. For each $\xi \in \Omega$ we choose $r_\xi > 0$ such that $B(\xi, r_\xi) \cap \Omega$ is in a bilipschitz bijection with a ball. It is obvious that we can do that for $\xi \in \mathring{\Omega}$, and for points on $\partial\Omega$ we use the fact that Ω is locally the epigraph of a Lipschitz function. By compactness, we find balls $B_1, B_2, \ldots, B_N$ covering Ω such that $B_n \cap \Omega$ is in a bilipschitz bijection with a ball for any $n \in \{1, 2, \ldots, N\}$. We can of course assume that B_n is not included in B_m for any $n \neq m$. Then we define recursively $X_1 := B_1 \cap \Omega$ and $X_n = (B_n \cap \Omega) \backslash \mathring{X}_{n-1}$ for $n \in \{2, \ldots, N\}$. For any $n \in \{1, 2, \ldots, N\}$, X_n is still in a bilipschitz bijection with a ball (see Figure 8.1 to understand what we are trying to do). On $\bigcup_n \partial X_n$, which is made of $\partial\Omega$ and of pieces of spheres of $\mathbb{R}^p$, thus locally in bilipschitz bijection with Lipschitz domains of $\mathbb{R}^{p-1}$, we can use the induction assumption and extend $\boldsymbol{\mu}_l$. Then, we use Lemma 8.35 to extend $\boldsymbol{\mu}$ on $\mathring{X}_n$ for each $n \in \{1, 2, \ldots, N\}$. We have obtained a function $\boldsymbol{\mu}$ which is continuous and Lipschitz on each $X_n, n \in \{1, 2, \ldots, N\}$: it is globally Lipschitz on Ω. $\square$

8.2.3 The dual problem

We will know show a rigorous proof of the absence of duality gap. The dual problem was already obtained, at least formally, in Chapter 7.

Theorem 8.36. *Let $\boldsymbol{\mu}_b \in H^1(\Omega, \mathcal{P}(D))$. Then one has*

$$\sup_{\varphi} \left\{ \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) \ : \ \varphi \in C^1(\Omega \times D, \mathbb{R}^p) \ \ \text{and} \ \ \nabla_\Omega \cdot \varphi + \frac{|\nabla_D \varphi|^2}{2} \leqslant 0 \ \text{on} \ \Omega \times D \right\}$$
$$= \min_{\boldsymbol{\mu}} \left\{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D)) \ \text{and} \ \boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega} \right\}.$$

Proof. We rely on the Fenchel-Rockafellar duality theorem which can be found in [Vil03, Theorem 1.9]. Let $X := C(\Omega \times D, \mathbb{R}^{1+pq})$ the space of continuous functions defined on the compact space $\Omega \times D$ and valued in $\mathbb{R}^{1+pq}$ endowed with the norm of uniform convergence. An element of X will be written (a, b), where $a \in C(\Omega \times D)$ and $b \in C(\Omega \times D, \mathbb{R}^{pq})$. The dual space $X^\star$ is, by the Riesz theorem, $\mathcal{M}(\Omega \times D, \mathbb{R}^{1+pq})$. Again an element of $X^\star$ will be written $(\boldsymbol{\mu}, \mathbf{E})$ where

$\boldsymbol{\mu} \in \mathcal{M}(\Omega \times D)$ is a signed measure and $\mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ is a vector-valued measure. We introduce the functionals $F : X \to \overline{\mathbb{R}}$ and $G : X \to \overline{\mathbb{R}}$ defined as, for any $(a, b) \in X$,

$$F(a,b) = \begin{cases} 0 & \text{if } a(\xi,x) + \dfrac{|b(\xi,x)|^2}{2} \leqslant 0 \text{ for every } (\xi,x) \in \Omega \times D \\ +\infty & \text{else,} \end{cases}$$
$$G(a,b) = \begin{cases} -\mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) & \text{if } (a,b) = (\nabla_\Omega \cdot \varphi, \nabla_D \varphi) \text{ for some } \varphi \in C^1(\Omega \times D, \mathbb{R}^p) \\ +\infty & \text{else.} \end{cases}$$

Notice that thanks to (8.16), G is well defined and does not depend on the choice of φ such that $(a, b) = (\nabla_\Omega \cdot \varphi, \nabla_D \varphi)$. Notice also that at the point $(-1, 0) \in X$, one has that F is finite and continuous and that G is finite (take $\varphi(\xi, x) := (-\xi^1, 0, 0, \ldots, 0)$, where ξ^1 is the first component of ξ). As moreover F and G are convex, one can apply Fenchel-Rockafellar duality which means

$$\begin{aligned} - \min_{(\boldsymbol{\mu},\mathbf{E}) \in X^\star} [F^\star(\boldsymbol{\mu}, \mathbf{E}) + G^\star(-\boldsymbol{\mu}, -\mathbf{E})] &= \inf_X (F + G) \\ &= -\sup_\varphi \left\{ \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) \ : \ \varphi \in C^1(\Omega \times D, \mathbb{R}^p) \ \text{ and } \ \nabla_\Omega \cdot \varphi + \frac{|\nabla_D \varphi|^2}{2} \leqslant 0 \right\}, \end{aligned}$$

where the last inequality is just a rewriting of the definition of F and G. Let us compute $F^\star(\boldsymbol{\mu}, \mathbf{E})$. By definition,

$$F^\star(\boldsymbol{\mu}, \mathbf{E}) = \sup_{a,b} \left\{ \iint\limits_{\Omega \times D} a \mathrm{d}\boldsymbol{\mu} + \iint\limits_{\Omega \times D} b \cdot \mathrm{d}\mathbf{E} \ : (a,b) \in C(\Omega \times D, \mathcal{K}) \right\},$$

where $\mathcal{K}$ is defined in Definition 8.5. In particular, if $\boldsymbol{\mu}$ is not a positive measure, then choosing suitable negative a, one sees that $F^\star(\boldsymbol{\mu}, \mathbf{E}) = +\infty$. Moreover, if $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$ and $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation, then $F^\star(\boldsymbol{\mu}, \mathbf{E}) = \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E})$: this is precisely Definition 8.5. On the other hand, we can compute $G^\star$: for any $(\boldsymbol{\mu}, \mathbf{E}) \in X^\star$,

$$G^\star(-\boldsymbol{\mu}, -\mathbf{E}) = \sup_{\varphi \in C^1(\Omega \times D, \mathbb{R}^p)} \left(\mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) - \iint\limits_{\Omega \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} - \iint\limits_{\Omega \times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E} \right).$$

By linearity of the expression inside the sup w.r.t. φ, we see that $G^\star(-\boldsymbol{\mu}, -\mathbf{E}) < +\infty$ if and only if $G^\star(-\boldsymbol{\mu}, -\mathbf{E}) = 0$, which translates in

$$\mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) = \iint\limits_{\Omega \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint\limits_{\Omega \times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E}$$

for every $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$. Let $a \in C(\Omega)$ a continuous function. It can always be written $a = \nabla_\Omega \cdot \varphi$, where $\varphi \in C^1(\Omega, \mathbb{R}^p)$ (take $\varphi = \nabla f$ where f solves $\Delta f = a$), thus using the fact that for such a φ,

$$\mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) = \iint\limits_{\Omega \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu}_b = \iint\limits_{\Omega \times D} a \mathrm{d}\boldsymbol{\mu}_b = \int_\Omega a(\xi) \mathrm{d}\xi,$$

one sees that if $G^\star(-\boldsymbol{\mu}, -\mathbf{E}) < +\infty$, then

$$\int_\Omega a(\xi) \mathrm{d}\xi = \iint\limits_{\Omega \times D} a \mathrm{d}\boldsymbol{\mu}.$$

Provided that $\boldsymbol{\mu}$ is a positive measure (recall that it happens if $F^\star(\boldsymbol{\mu}, \mathbf{E}) < +\infty$) and by arbitrariness of a, it implies that the disintegration of $\boldsymbol{\mu}$ w.r.t. $\mathcal{L}_\Omega$ is made of probability measures on D, in other words that $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$. Once we have this information, testing with functions φ which are compactly supported on Ω, we see that if $G^\star(-\boldsymbol{\mu}, -\mathbf{E}) < +\infty$ then $(\boldsymbol{\mu}, \mathbf{E})$ satisfies the continuity equation, and testing with arbitrary φ, we see that $\mathrm{BT}_{\boldsymbol{\mu}} = \mathrm{BT}_{\boldsymbol{\mu}_b}$. In the end, one concludes that

$$\min_{(\boldsymbol{\mu}, \mathbf{E}) \in X^\star} [F^\star(\boldsymbol{\mu}, \mathbf{E}) + G^\star(-\boldsymbol{\mu}, -\mathbf{E})] = \min_{\boldsymbol{\mu}} \left\{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D)) \text{ and } \boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_b|_{\partial\Omega} \right\}. \quad \square$$

A natural question which arises is the existence of an optimal $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$ (or in a space of less regular functions). Actually, as detailed in Chapter 12, we do not know the answer to this question and we believe that it can be substantially more complicated than in the case where Ω is a segment of $\mathbb{R}$.

8.3 Failure of the superposition principle

8.3.1 The superposition principle

In this section, we want to explain why a powerful tool to study curves valued in the Wasserstein space (i.e. the case where Ω is a segment of $\mathbb{R}$), namely the superposition principle, fails in higher dimensions. To say it briefly, there is no Lagrangian point of view for mappings valued into the Wasserstein space, one has to work only with the Eulerian one. Notice that the question of the existence of a superposition principle was already formulated by Brenier [Bre03, Problem 3.1], but left unanswered. As we want to prove a negative result, we will not only provide a counterexample to the superposition principle, but also try to explain the obstruction and why this principle fails for all but few exceptional cases. Let us first recall the superposition principle for absolutely continuous curves.

The set Ω will be replaced by the unit segment $I = [0, 1]$. As stated in Proposition 8.8, the set $H^1(I, \mathcal{P}(D))$ coincides with the set of 2-absolutely continuous curves whose definition is recalled in Section 2.2. We denote by $\mathcal{C} = C(I, D)$ the set of continuous curves valued in D endowed with the norm of uniform convergence, it is a polish space. If $f \in \mathcal{C}$, then $\dot{f}$ denotes the derivative w.r.t. time of f provided that it exists. For any $t \in I$, $e_t : \mathcal{C} \to D$ is the evaluation operator, which means $e_t(f) = f(t)$ for any $f \in \mathcal{C}$. The following result can be found in [AGS08, Section 8.2], see also [Lis07] for a more general framework.

Theorem 8.37. *Let $\boldsymbol{\mu} \in H^1(I, \mathcal{P}(D))$. Then there exists a probability measure $Q \in \mathcal{P}(\mathcal{C})$ such that*

(i) for any $t \in I$, $e_t \# Q = \boldsymbol{\mu}(t)$;

(ii) the following equality holds:

$$\mathrm{Dir}(\boldsymbol{\mu}) = \int_{\mathcal{C}} \left(\int_I \frac{1}{2} |\dot{f}(t)|^2 \mathrm{d}t \right) Q(\mathrm{d}f).$$

The measure Q can be seen as a multimarginal transport plan coupling all the different instants, whose 2-marginals are almost optimal transport plans if they are taken between two very close instants. In other words, for any t and s in I, $(e_s, e_t)\#Q$ is a transport plan between $\boldsymbol{\mu}(s)$ and $\boldsymbol{\mu}(t)$ (by (i)), and it is almost an optimal transport plan if s is very close to t by (ii).

Another way to see it is the following: if $f \in \mathcal{C}$, then we can also see it as an element $\boldsymbol{\mu}_f$ of $H^1(I, \mathcal{P}(D))$. Indeed, just set $\boldsymbol{\mu}(t) = \delta_{f(t)}$ for any $t \in I$, and one can define $\mathbf{E}_f \in \mathcal{M}(I \times D, \mathbb{R}^q)$ by, for any $b \in C(I \times D, \mathbb{R}^q)$,

$$\iint_{I \times D} b \cdot \mathrm{d}\mathbf{E}_f := \int_I b(t, f(t)) \cdot \dot{f}(t)\mathrm{d}t.$$

With this choice, one can check that

$$\mathrm{Dir}(\boldsymbol{\mu}_f) = \mathrm{Dir}(\boldsymbol{\mu}_f, \mathbf{E}_f) = \int_I \frac{1}{2}|\dot{f}(t)|^2\mathrm{d}t.$$

Then, Theorem 8.37 is saying that there exists $Q \in \mathcal{P}(\mathcal{C})$ such that $\boldsymbol{\mu}$ is the mean w.r.t. Q of the $\boldsymbol{\mu}_f$ (this is (i)), and such the $\mathbf{E}$ which is tangent to $\boldsymbol{\mu}$ is the mean w.r.t. Q of the $\mathbf{E}_f$. Indeed, by linearity of the continuity equation the mean of the $\mathbf{E}_f$ is an admissible momentum. Using Jensen's inequality,

$$\mathrm{Dir}(\boldsymbol{\mu}) = \mathrm{Dir}\left(\int_{\mathcal{C}} \boldsymbol{\mu}_f Q(\mathrm{d}f)\right) \leqslant \mathrm{Dir}\left(\int_{\mathcal{C}} \boldsymbol{\mu}_f Q(\mathrm{d}f), \int_{\mathcal{C}} \mathbf{E}_f Q(\mathrm{d}f)\right) \leqslant \int_{\mathcal{C}} \mathrm{Dir}(\boldsymbol{\mu}_f, \mathbf{E}_f) Q(\mathrm{d}f)$$

and the r.h.s. is equal to the l.h.s. by (ii). Hence, all inequalities are equalities, which tells us that $\int_{\mathcal{C}} \mathbf{E}_f Q(\mathrm{d}f)$ is the tangent momentum to $\boldsymbol{\mu}$.

Let us try to see what a superposition principle would look like if the dimension of Ω is larger than 1. We denote by $\mathcal{F}$ the space $L^2(\Omega, D)$ which is a polish space. As it was already done in [Bre03], if $f \in H^1(\Omega, D)$, then we can see it as an element $\boldsymbol{\mu}_f$ of $H^1(\Omega, \mathcal{P}(D))$ by setting $\boldsymbol{\mu}_f(\xi) := \delta_{f(\xi)}$. In other words, a classical function can be seen as a mapping valued in the Wasserstein space by identifying $f(\xi) \in D$ with $\delta_{f(\xi)} \in \mathcal{P}(D)$. More precisely, we define $\boldsymbol{\mu}_f \in L^2(\Omega, \mathcal{P}(D))$ and $\mathbf{E}_f \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ by, for any $a \in C(\Omega \times D)$ and $b \in C(\Omega \times D, \mathbb{R}^{pq})$,

$$\iint_{\Omega \times D} a \mathrm{d}\boldsymbol{\mu}_f := \int_\Omega a(\xi, f(\xi))\mathrm{d}\xi,$$

$$\iint_{\Omega \times D} b \cdot \mathrm{d}\mathbf{E}_f := \int_\Omega b(\xi, f(\xi)) \cdot \nabla f(\xi)\mathrm{d}\xi.$$

Proposition 8.38. *If $f \in H^1(\Omega, D)$, and if $\boldsymbol{\mu}_f$ and $\mathbf{E}_f$ are defined as above, then $\mathbf{E}_f$ is tangent to $\boldsymbol{\mu}_f$ and*

$$\mathrm{Dir}(\boldsymbol{\mu}_f) = \mathrm{Dir}(\boldsymbol{\mu}_f, \mathbf{E}_f) = \int_\Omega \frac{1}{2}|\nabla f(\xi)|^2\mathrm{d}\xi.$$

Proof. To check the first part, take $\varphi \in C^1_c(\mathring{\Omega} \times D, \mathbb{R}^p)$. Defining $\tilde{\varphi} \in H^1(\Omega, \mathbb{R}^p)$ by $\tilde{\varphi}(\xi) = \varphi(\xi, f(\xi))$, we have that $\tilde{\varphi}$ is compactly supported in $\mathring{\Omega}$ and

$$\nabla \cdot \tilde{\varphi} = (\nabla_\Omega \cdot \varphi)(\xi, f(\xi)) + (\nabla_D \varphi)(\xi, f(\xi)) \cdot \nabla f(\xi).$$

Integrating this identity w.r.t. Ω, as the l.h.s. vanishes by compactness of the support of $\tilde{\varphi}$, we see that we can conclude that $(\boldsymbol{\mu}_f, \mathbf{E}_f)$ satisfies the continuity equation.

Notice that $\mathbf{E}_f$ has a density $\mathbf{v}_f \in L^2_{\boldsymbol{\mu}_f}(\Omega \times D, \mathbb{R}^{pq})$ w.r.t. $\boldsymbol{\mu}$ given by $\mathbf{v}_f(\xi, x) = \nabla f(\xi)$. In particular, for a fixed ξ, $\mathbf{v}_f(\xi, \cdot)$ is constant hence the gradient of a function. Using Proposition 8.11, one sees that it is enough to conclude that $\mathbf{E}_f$ is tangent. Moreover, as $\mathbf{v}_f$ does not depend on x,

$$\mathrm{Dir}(\boldsymbol{\mu}_f) = \mathrm{Dir}(\boldsymbol{\mu}_f, \mathbf{E}_f) = \iint_{\Omega \times D} \frac{1}{2}|\mathbf{v}_f(\xi)|^2 \boldsymbol{\mu}(\mathrm{d}\xi, \mathrm{d}x) = \int_\Omega \frac{1}{2}|\mathbf{v}_f(\xi)|^2\mathrm{d}\xi = \int_\Omega \frac{1}{2}|\nabla f(\xi)|^2\mathrm{d}\xi. \quad \square$$

We mention that Brenier proved that if $f : \Omega \to D$ is a (classical) harmonic map, then $\boldsymbol{\mu}_f$ is also an harmonic mapping, see Proposition 10.1 below.

By analogy, the superposition principle would read as follows: If $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ and $\mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ is tangent to $\boldsymbol{\mu}$, does there exist $Q \in \mathcal{P}(\mathcal{F})$ such that $\boldsymbol{\mu}$ is the mean of $\boldsymbol{\mu}_f$ w.r.t. Q and $\mathbf{E}$ is the mean of $\mathbf{E}_f$ w.r.t. Q? Thanks to Jensen's inequality and the uniqueness of the tangent momentum, the second condition can in fact be rewritten as

$$\mathrm{Dir}(\boldsymbol{\mu}) = \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) = \int_{\mathcal{F}} \mathrm{Dir}(\boldsymbol{\mu}_f, \mathbf{E}_f) Q(\mathrm{d}f) = \int_{\mathcal{F}} \left(\int_\Omega \frac{1}{2} |\nabla f(\xi)|^2 \mathrm{d}\xi \right) Q(\mathrm{d}f).$$

These considerations can be summarized by the following definition, which is the same as [Bre03, Problem 3.1]. For $f \in \mathcal{F}$ we define its "classical" Dirichlet energy $\mathrm{Dir}_c(f)$ by

$$\mathrm{Dir}_c(f) = \begin{cases} \int_\Omega \frac{1}{2} |\nabla f(\xi)|^2 \mathrm{d}\xi & \text{if } f \in H^1(\Omega, D), \\ +\infty & \text{else.} \end{cases}$$

Definition 8.39. *Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$. We say that $\boldsymbol{\mu}$ admits a superposition principle if there exists $Q \in \mathcal{P}(\mathcal{F})$ such that*

(i) for any $a \in C(\Omega \times D)$;

$$\iint_{\Omega \times D} a \mathrm{d}\boldsymbol{\mu} = \int_{\mathcal{F}} \left(\int_\Omega a(\xi, f(\xi)) \mathrm{d}\xi \right) Q(\mathrm{d}f),$$

(ii) the following identity holds:

$$\int_{\mathcal{F}} \mathrm{Dir}_c(f) Q(\mathrm{d}f) \leqslant \mathrm{Dir}(\boldsymbol{\mu}).$$

In particular, with our definition, if Q represents $\boldsymbol{\mu} \in H^1(\Omega, D)$, then for Q-a.e. function f one has $\mathrm{Dir}_c(f) < +\infty$ hence f belongs to $H^1(\Omega, D)$. Let us underline that (i) is heuristically the same as (i) of Theorem 8.37, but in a form integrated over Ω because the evaluation operator does not make sense in higher dimensions: the elements of $\mathcal{F}$ are not necessarily continuous. In Definition 8.39, if (i) and (ii) holds, then the inequality in (ii) is in fact an equality because the reverse inequality always holds. Indeed, if $\boldsymbol{\mu}$ satisfies the superposition principle, we can say that $\boldsymbol{\mu} = \int_{\mathcal{F}} \boldsymbol{\mu}_f Q(\mathrm{d}f)$. By convexity of the Dirichlet energy (Proposition 8.13), we can apply Jensen's inequality, thus

$$\mathrm{Dir}(\boldsymbol{\mu}) \leqslant \int_{\mathcal{F}} \mathrm{Dir}(\boldsymbol{\mu}_f) Q(\mathrm{d}f) = \int_{\mathcal{F}} \mathrm{Dir}_c(f) Q(\mathrm{d}f).$$

8.3.2 Counterexample

We will first provide a counterexample which we will try to make as generic as possible. In what follows, we take $\Omega := \mathbb{B}$ to be the unit disk of $\mathbb{R}^2$ and $\mathbb{S}^1 = \partial\mathbb{B}$ its boundary. We also take $D = \mathbb{B}$. We view $\mathbb{B}$ as a subset of the complex plane $\mathbb{C}$: multiplication on $\mathbb{B}$ means complex multiplication.

Let $\boldsymbol{\mu}_s : \mathbb{S}^1 :\to \mathcal{P}(\mathbb{B})$ be the (complex) square root: it is the mapping defined by, for $\xi \in \mathbb{S}^1$,

$$\boldsymbol{\mu}_s(\xi) := \frac{1}{2} \sum_{z^2 = \xi} \delta_z = \frac{1}{2}(\delta_{\sqrt{\xi}} + \delta_{-\sqrt{\xi}}),$$

where $\sqrt{\xi}$ is a (complex) square root of ξ. The function $\boldsymbol{\mu}_s$ is clearly Lipschitz (with Lipschitz constant equals to 2). In fact, if $\xi = e^{it}$ with $t \in \mathbb{R}$, one can write

$$\boldsymbol{\mu}_s(e^{it}) = \frac{1}{2}\left(\delta_{\exp(it/2)} + \delta_{\exp(it/2+i\pi)}\right).$$

The function $t \mapsto \boldsymbol{\mu}_s(e^{it})$ is 2π-periodic, but it cannot be written as a superposition of continuous 2π-periodic functions, only 4π-periodic ones. Hence, the superpositon principle with continuous functions fails for this mapping. This example is well known in the theory of Q-functions [DLS11], we took it from there. To our purpose, we will need the fact that the superposition principle with $H^{1/2}$ functions fails for the mapping $\boldsymbol{\mu}_s$: roughly speaking, it holds because $H^{1/2}$ functions, in dimension 1, cannot have jumps.

Lemma 8.40. *There is no function $f \in H^{1/2}(\mathbb{S}^1, \mathbb{B})$ such that $f(\xi)^2 = \xi$ for a.e. $\xi \in \mathbb{S}^1$.*

As this lemma is not directly related to harmonic mappings, we postpone its proof to the end of this chapter in Section 8.4. With the help of this lemma, we can prove that no mapping $\boldsymbol{\mu} \in H^1(\mathbb{B}, \mathcal{P}(\mathbb{B}))$ such that $\boldsymbol{\mu}|_{\partial\mathbb{B}} = \boldsymbol{\mu}_s$ can have a superposition principle: indeed, if it were the case, then we could restrict the superposition to $\partial\mathbb{B}$, and we would have a superposition principle for $\boldsymbol{\mu}_s$ with functions in $H^{1/2}$ which is a contradiction. To make this argument rigorous is a bit technical given the definition we chose for the boundary values of mappings in $H^1(\mathbb{B}, \mathcal{P}(\mathbb{B}))$: $\boldsymbol{\mu}$ is not necessarily continuous.

Proposition 8.41. *Let $\boldsymbol{\mu} \in H^1(\mathbb{B}, \mathcal{P}(\mathbb{B}))$ such that $\boldsymbol{\mu}|_{\partial\mathbb{B}} = \boldsymbol{\mu}_s$. Then $\boldsymbol{\mu}$ cannot admit a superposition principle.*

Proof. We will of course reason by contradiction. We assume that there exists $Q \in \mathcal{P}(\mathcal{F})$ which satisfies the points (i) and (ii) of Definition 8.39 (in fact only point (i) will be sufficient). Let $\mathbf{E} = \mathbf{v}\boldsymbol{\mu}$ tangent to $\boldsymbol{\mu}$. Take $\delta > 0$ and $\varepsilon > 0$. We choose $\chi_\varepsilon \in C^1([0,1])$ an increasing function supported on $[1-\varepsilon, 1]$, such that $\chi_\varepsilon(1) = 1$. Define $a_\varepsilon \in C^1(\mathbb{B}, \mathbb{R}^2)$ and $b_\delta \in C^1(\mathbb{B} \times \mathbb{B})$ by, for any $\xi, x \in \mathbb{B}$,

$$\begin{aligned} a_\varepsilon(\xi) &= \frac{\xi}{|\xi|}\chi_\varepsilon(|\xi|), \\ b_\delta(\xi, x) &= \frac{|\xi - x^2|^2}{\delta^2}. \end{aligned}$$

In words, a_ε is a vector-valued function, parallel to lines issued from the origin, and whose norm is increasing on the annulus of radii $1-\varepsilon$ and 1 from 0 to 1. Define $A_\varepsilon = \{\xi \in \mathbb{B} \ : \ 1-\varepsilon \leqslant |\xi| \leqslant 1\}$ the annulus outside which a_ε vanishes. A simple computation gives

$$\left|\nabla \cdot a_\varepsilon(\xi) - \chi'_\varepsilon(|\xi|)\right| \leqslant C\mathbb{1}_{A_\varepsilon}(\xi),$$

where C does not depend on ε. On the other hand, b_δ is a smooth scalar function, which vanishes if $x^2 = \xi$, which is larger than 1 if $|x^2 - \xi| \geqslant \delta$ and whose derivative is bounded by $C\delta^{-2}$. As a test function for the continuity equation, we take $\varphi(\xi, x) = a_\varepsilon(\xi) b_\delta(\xi, x)$. With this choice, for every $\xi \in \mathbb{S}^1$, one has

$$\int_{\mathbb{B}} \varphi(\xi, x)\boldsymbol{\mu}_s(\xi, \mathrm{d}x) = \frac{1}{2}\sum_{x^2=\xi} \varphi(\xi, x) = 0.$$

Thus, $\mathrm{BT}_{\boldsymbol{\mu}_s}(\varphi) = 0$ and the continuity equation tested against φ reads

$$\Bigg| \iint_{\mathbb{B}\times\mathbb{B}} \chi_\varepsilon'(|\xi|) b_\delta(\xi, x)\boldsymbol{\mu}(\mathrm{d}\xi, \mathrm{d}x) + \iint_{\mathbb{B}\times\mathbb{B}} [a_\varepsilon(\xi)\cdot\nabla_\Omega b_\delta(\xi,x) + (a_\varepsilon(\xi)\otimes\nabla_D b_\delta(\xi,x))\cdot\mathbf{v}(\xi,x)]\boldsymbol{\mu}(\mathrm{d}\xi,\mathrm{d}x)\Bigg| \leqslant C\varepsilon.$$

Indeed, in the r.h.s, the reminder $\nabla\cdot a_\varepsilon - \chi_\varepsilon'(|\xi|)$ of order 1 has been integrated over A_ε whose area scales like ε. For the first integral, we use the assumption that $\boldsymbol{\mu}$ satisfies the superposition principle. For the second one, we bound ∇b_δ by $C\delta^{-2}$, notice that a_ε vanishes outside A_ε and use Cauchy-Schwarz:

$$\begin{aligned}
\int_{\mathcal{F}}\left(\int_{\mathbb{B}} \chi_\varepsilon'(|\xi|)b_\delta(\xi, f(\xi))\mathrm{d}\xi\right) Q(\mathrm{d}f) &= \iint_{\mathbb{B}\times\mathbb{B}} \chi_\varepsilon'(|\xi|)b_\delta(\xi,x)\boldsymbol{\mu}(\mathrm{d}\xi,\mathrm{d}x)\\
&\leqslant \frac{C}{\delta^2}\iint_{A_\varepsilon\times\mathbb{B}} (1+|\mathbf{v}(\xi,x)|)\boldsymbol{\mu}(\mathrm{d}\xi,\mathrm{d}x) + C\varepsilon\\
&\leqslant \frac{C}{\delta^2}\sqrt{\iint_{\mathbb{B}\times\mathbb{B}}(1+|\mathbf{v}(\xi,x)|^2)\boldsymbol{\mu}(\mathrm{d}\xi,\mathrm{d}x)}\sqrt{\iint_{A_\varepsilon\times\mathbb{B}}\boldsymbol{\mu}(\mathrm{d}\xi,\mathrm{d}x)} + C\varepsilon\\
&\leqslant \frac{C}{\delta^2}\sqrt{1+2\mathrm{Dir}(\boldsymbol{\mu})}\sqrt{\varepsilon} + C\varepsilon \leqslant C\frac{\sqrt{\varepsilon}}{\delta^2},
\end{aligned}$$

where C denotes a generic constant which changes from one line to another and the inequality may hold only for small ε and δ. Let us call $\mathcal{F}_{\delta,\varepsilon}\subset\mathcal{F}$ the set of $f\in\mathcal{F}$ such that

$$\int_{\mathbb{B}} \chi_\varepsilon'(|\xi|)|f(\xi)^2 - \xi|^2\mathrm{d}\xi \geqslant \delta^2.$$

By Markov's inequality, one can say that

$$\begin{aligned}
Q(\mathcal{F}_{\delta,\varepsilon}) = Q\left(\left\{ f\in\mathcal{F} \ :\ \int_{\mathbb{B}} \chi_\varepsilon'(|\xi|)b_\delta(\xi, f(\xi))\mathrm{d}\xi \geqslant 1\right\}\right)&\\
\leqslant \int_{\mathcal{F}}\left(\int_{\mathbb{B}}\chi_\varepsilon'(|\xi|)b_\delta(\xi,f(\xi))\mathrm{d}\xi\right)Q(\mathrm{d}f) \leqslant C\frac{\sqrt{\varepsilon}}{\delta^2}.&
\end{aligned}$$

Now take the sequence $\varepsilon_n := 2^{-n}$. By the previous estimate, one sees that

$$\sum_{n=1}^{+\infty} Q(\mathcal{F}_{\delta,\varepsilon_n}) < +\infty.$$

By the Borel-Cantelli lemma, one has that $Q(\limsup_n \mathcal{F}_{\delta,\varepsilon_n}) = 0$ which means that for Q-a.e. $f\in\mathcal{F}$, there exists n_0 (which may depend on f) such that

$$\int_{\mathbb{B}} \chi_{\varepsilon_n}'(|\xi|)|f(\xi)^2 - \xi|^2\mathrm{d}\xi \leqslant \delta^2$$

for all $n\geqslant n_0$. Recall also that Q-a.e. f belongs to $H^1(\Omega, D)$. For such an f, sending n to $+\infty$ and by definition of the trace of f,

$$\int_{\mathbb{S}^1} |\bar{f}(\xi)^2 - \xi|^2\sigma(\mathrm{d}\xi) \leqslant \delta^2,$$

where in this formula $\bar{f}$ stands for the trace of f on $\mathbb{S}^1$ and σ the surface measure on $\partial\mathbb{B}$. Then using this estimate for smaller and smaller δ along a countable sequence, we conclude that Q-a.e. function f satisfies $\bar{f}(\xi)^2 = \xi$ a.e. on $\mathbb{S}^1$. But on the other hand the trace of Q-a.e. function f belongs to $H^{1/2}(\mathbb{S}^1, \mathbb{B})$, which is a clear contradiction with Lemma 8.40. □

From this Proposition, we deduce that there exists an harmonic and a Lipschitz mapping $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ for which the superposition principle fails: just take respectively a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_s$, or a Lipschitz extension of $\boldsymbol{\mu}_s$.

Though, these examples can seem too particular and rely too much on some singular boundary conditions. To produce stronger examples, we will use the fact that, roughly speaking, the set of $\boldsymbol{\mu}$ admitting a superposition principle is stable by approximation. Thus, by contraposition, any neighborhood of a $\boldsymbol{\mu}$ which does not admit a superposition principle will contain other measures not admitting a superposition principle.

Proposition 8.42. *Let $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ a sequence of elements of $H^1(\Omega, \mathcal{P}(D))$ such that, for every $n \in \mathbb{N}$, $\boldsymbol{\mu}_n$ admits a superposition principle. We assume that $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ converges weakly to $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ and that $\lim_n \mathrm{Dir}(\boldsymbol{\mu}_n) = \mathrm{Dir}(\boldsymbol{\mu})$. Then $\boldsymbol{\mu}$ admits a superposition principle.*

Proof. For any $n \in \mathbb{N}$, let $Q_n \in \mathcal{P}(\mathcal{F})$ such that (i) and (ii) of Definition 8.39 are satisfied. By Rellich's theorem (recall that D is compact), the functional $\mathrm{Dir}_c : \mathcal{F} \to \mathbb{R}$ has compact sublevel sets in the $L^2(\Omega, D)$-topology. As

$$\sup_{n\in\mathbb{N}} \int_{\mathcal{F}} \mathrm{Dir}_c(f) Q_n(\mathrm{d}f) = \sup_{n\in\mathbb{N}} \mathrm{Dir}(\boldsymbol{\mu}_n) < +\infty,$$

we can say [AGS08, Remark 5.1.5] that $(Q_n)_{n\in\mathbb{N}}$ is tight, hence up to extraction it weakly converges in $\mathcal{P}(\mathcal{F})$ to some $Q \in \mathcal{P}(\mathcal{F})$. We will show that Q represents $\boldsymbol{\mu}$.

Let us take $a \in C(\Omega \times D)$ and define $A : \mathcal{F} \to \mathbb{R}$ by, for any $f \in \mathcal{F}$,

$$A(f) := \int_\Omega a(\xi, f(\xi)) \mathrm{d}\xi.$$

The function A is continuous for the L^2 topology. Thus, starting from

$$\int_{\mathcal{F}} A(f) Q_n(\mathrm{d}f) = \iint\limits_{\Omega \times D} a \,\mathrm{d}\boldsymbol{\mu}_n,$$

which is valid by Definition 8.39, we can pass both terms to the limit (recall that $\boldsymbol{\mu}_n$ weakly converges to $\boldsymbol{\mu}$) and see that $(\boldsymbol{\mu}, Q)$ satisfies (i) of Definition 8.39.

Moreover, as Dir_c is l.s.c. (for the $L^2(\Omega, D)$ topology), we can say that

$$\int_{\mathcal{F}} \mathrm{Dir}_c(f) Q(\mathrm{d}f) \leqslant \liminf_{n\to+\infty} \int_{\mathcal{F}} \mathrm{Dir}_c(f) Q_n(\mathrm{d}f) = \liminf_{n\to+\infty} \mathrm{Dir}(\boldsymbol{\mu}_n) = \mathrm{Dir}(\boldsymbol{\mu}),$$

which gives point (ii) of Definition 8.39 and concludes the proof. □

With this proposition, one can use for instance the heat flow to regularize mappings and produce "smoother" counterexamples. For instance, let $\boldsymbol{\mu} \in H^1(\mathbb{B}, \mathcal{P}(\mathbb{B}))$ which does not satisfy the superposition principle. Set $\boldsymbol{\mu}_n(\xi) := \Phi^{\mathbb{B}}_{1/n} \boldsymbol{\mu}(\xi)$: for a fixed $\xi \in \mathbb{B}$, we regularize $\boldsymbol{\mu}(\xi)$ with the help of the heat flow acting on $\mathcal{P}(\mathbb{B})$. One can check easily that $\boldsymbol{\mu}_n$ converges weakly in $L^2(\mathbb{B}, \mathcal{P}(\mathbb{B}))$ to $\boldsymbol{\mu}$. As $\Phi^{\mathbb{B}}_{1/n}$ is a contraction in the Wasserstein space (Proposition 2.13), $\mathrm{Dir}(\boldsymbol{\mu}_n) \leqslant \mathrm{Dir}(\boldsymbol{\mu})$ and by lower semi-continuity of Dir we deduce that $\lim_n \mathrm{Dir}(\boldsymbol{\mu}_n) = \mathrm{Dir}(\boldsymbol{\mu})$. According to Proposition 8.42, we deduce that $\boldsymbol{\mu}_n$ does not satisfy the superposition principle for n large enough. On the other hand, the for any ξ and any n the measure $\boldsymbol{\mu}_n(\xi)$ is smooth: it admits a density bounded from below and from above.

8.3.3 Local obstruction to the superposition principle

The counterexample provided above shows a *global* obstruction. Indeed, the mapping $\boldsymbol{\mu}_s$ can be thought locally in Ω as a superposition of classical functions, but there is a contradiction if we try to make this superposition global. On the other hand, there is also (at least formally) *local obstructions* to the superposition principle. To describe them we will stay sloppy about the regularity issues and concentrate on heuristic explanations.

Indeed, if $\boldsymbol{\mu}$ admits a superposition principle given by $Q \in \mathcal{P}(\mathcal{F})$, and if $\mathbf{v}$ is the velocity field tangent to $\boldsymbol{\mu}$, then for Q-a.e. f, one has $\nabla f(\xi) = \mathbf{v}(\xi, f(\xi))$. To prove this fact, notice that the tangent momentum $\mathbf{E} = \mathbf{v}\boldsymbol{\mu}$ is equal to $\int_{\mathcal{F}} \mathbf{E}_f Q(\mathrm{d}f)$: this is exactly the same proof as the case where Ω is a segment of $\mathbb{R}$ which we did at the beginning of this section. In other words, for any $b \in C(\Omega \times D, \mathbb{R}^{pq})$,

$$\iint_{\Omega \times D} b \cdot \mathrm{d}\mathbf{E} := \int_{\mathcal{F}} \left(\int_{\Omega} b(\xi, f(\xi)) \cdot \nabla f(\xi) \mathrm{d}\xi \right) Q(\mathrm{d}f).$$

Thus, one can say that

$$\begin{aligned}
\mathrm{Dir}(\boldsymbol{\mu}) = \iint_{\Omega \times D} \frac{1}{2}|\mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu} = \iint_{\Omega \times D} \frac{1}{2}\mathbf{v} \cdot \mathrm{d}\mathbf{E} &= \int_{\mathcal{F}} \left(\int_{\Omega} \frac{1}{2}\mathbf{v}(\xi, f(\xi)) \cdot \nabla f(\xi) \mathrm{d}\xi \right) Q(\mathrm{d}f) \\
&\leqslant \int_{\mathcal{F}} \left(\int_{\Omega} \frac{1}{4} \left[|\mathbf{v}(\xi, f(\xi))|^2 + |\nabla f(\xi)|^2 \right] \mathrm{d}\xi \right) Q(\mathrm{d}f) \\
&= \frac{1}{4} \iint_{\Omega \times D} |\mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu} + \frac{1}{2} \int_{\mathcal{F}} \left(\int_{\Omega} \frac{1}{2} |\nabla f(\xi)|^2 \mathrm{d}\xi \right) Q(\mathrm{d}f) \\
&= \mathrm{Dir}(\boldsymbol{\mu}).
\end{aligned}$$

In particular, the inequality is an equality: one sees that for Q-a.e. $f \in \mathcal{F}$, one has $\nabla f(\xi) = \mathbf{v}(\xi, f(\xi))$ for a.e. $\xi \in \Omega$.

The analogue if Ω is a segment is the fact that (using notations from Theorem 8.37) for Q-a.e. there holds f, $\dot{f}(t) = \mathbf{v}(t, f(t))$: the measure Q is supported on the flow of the vector field $\mathbf{v}$ (see [AGS08, Theorem 8.2.1]). In dimension larger than 1, the constraint $\nabla f = \mathbf{v}(\cdot, f)$ is much stronger. In particular, it implies that along every curve $\gamma : I \to \Omega$, the function $f \circ \gamma$ follows the flow of $\mathbf{v} \cdot \dot{\gamma}$. However, there are many different curves going from one point to another: if we want all the results to be coherent, some commutation properties of the flow of $\mathbf{v}$ along different directions are needed, which turns out to be a very strong constraint. Indeed, coordinatewise, the constraint reads for every $\alpha \in \{1, 2, \dots, p\}$ and $i \in \{1, 2, \dots, q\}$,

$$\partial_\alpha f^i(\xi) = \mathbf{v}^{\alpha i}(\xi, f(\xi)).$$

If we differentiate w.r.t. β, we find that

$$\partial_{\beta\alpha} f^i(\xi) = \partial_\beta \mathbf{v}^{\alpha i}(\xi, f(\xi)) + \sum_{j=1}^{q} \partial_\beta f^j(\xi) \partial_j \mathbf{v}^{\alpha i}(\xi, f(\xi)) = \left(\partial_\beta \mathbf{v}^{\alpha i} + \sum_{j=1}^{q} \mathbf{v}^{\beta j} \partial_j \mathbf{v}^{\alpha i} \right) (\xi, f(\xi)).$$

The l.h.s is clearly symmetric if we exchange the role of α and β, so must be the r.h.s. It implies that for all $\alpha, \beta \in \{1, 2, \dots, p\}$,

$$\partial_\alpha \mathbf{v}^{\beta i} + \sum_{j=1}^{q} \mathbf{v}^{\alpha j} \partial_j \mathbf{v}^{\beta i} = \partial_\beta \mathbf{v}^{\alpha i} + \sum_{j=1}^{q} \mathbf{v}^{\beta j} \partial_j \mathbf{v}^{\alpha i}, \tag{8.19}$$

at least on the support of $\boldsymbol{\mu}$ in $\Omega \times D$. In other words, we see that $\mathbf{v}$ must satisfy a differential constraint for the superposition principle to hold, and there is no reason why this constraint would be satisfied for a generic $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$, even for a harmonic mapping. Actually, we provide in Section 10.4 an explicit example where this commutativity relation (8.19) does not hold.

An other way to understand the *local* failure of the superposition principle is the following. We will be sloppy and use the evaluation operators $e_\xi : \mathcal{F} \to D$ defined by $e_\xi(f) := f(\xi)$ (these operators are in principle not defined as elements of $\mathcal{F}$ are not continuous). If $\boldsymbol{\mu}$ admits a superposition principle, it would mean that for ξ and η very close, $(e_\xi, e_\eta)\#Q \in \mathcal{P}(D \times D)$ is a transport plan between $\boldsymbol{\mu}(\xi)$ and $\boldsymbol{\mu}(\eta)$ (because of point (i)) which is almost optimal (between of point (ii)). It also works with three measures: if ξ, η and θ are three points of Ω very close to each other (for instance located at the vertices of an equilateral triangle), then $(e_\xi, e_\eta, e_\theta)\#Q \in \mathcal{P}(D \times D \times D)$ is a coupling between $\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta)$ and $\boldsymbol{\mu}(\theta)$ whose 2-marginals are almost optimal transport plans. However, it is known that, if μ_1, μ_2 and $\mu_3 \in \mathcal{P}(D)$, then in general there exists no coupling between the three whose 2-marginals are optimal transport plans.

8.4 Appendix: $H^{1/2}$ determination of the square root

In this subsection we want to prove Lemma 8.40, which states that, with $\mathbb{S}^1$ the unit circle of the complex plane $\mathcal{C}$ and $\mathbb{B}$ its unit disk, there is no function $f \in H^{1/2}(\mathbb{S}^1, \mathbb{S}^1)$ such that $f(\xi)^2 = \xi$ for a.e. $\xi \in \mathbb{S}^1$ (where the multiplication is understood as a complex multiplication). We take for granted that there is no continuous function $f \in C(\mathbb{S}^1, \mathbb{S}^1)$ such that $f(\xi)^2 = \xi$ for all $\xi \in \mathbb{S}^1$. Hence, it is enough to reason by contradiction and to prove that a function $f \in H^{1/2}(\mathbb{S}^1, \mathbb{S}^1)$ such that $f(\xi)^2 = \xi$ for a.e. $\xi \in \mathbb{S}^1$ admits a continuous representative.

We start with some easy lemma which states that $H^{1/2}(\mathbb{S}^1, \mathbb{B})$ is stable by composition with Lipschitz function.

Lemma 8.43. *Let $u : \mathbb{S}^1 \to \mathbb{R}$ a Lipschitz function and $f \in H^{1/2}(\mathbb{S}^1, \mathbb{S}^1)$. Then $(u \circ f) \in H^{1/2}(\mathbb{S}^1, \mathbb{R})$.*

Proof. It is well known (see [McL00, Chapter 3]) that there exists $\tilde{f} \in H^1(\mathbb{B}, \mathbb{B})$ whose trace on $\mathbb{S}^1$ is f. Clearly, the function $u \circ \tilde{f}$ stays in $H^1(\mathbb{B}, \mathbb{R})$, hence its trace, which is nothing else than $u \circ f$, is in $H^{1/2}(\mathbb{S}^1, \mathbb{R})$. □

Then, let us prove that an $H^{1/2}$ function cannot have a jump.

Proposition 8.44. *Let $f \in H^{1/2}([0,1], \mathbb{R})$ such that $f(\xi) \in \{0, 1\}$ for a.e. $\xi \in [0,1]$. Then there is a representative of f which is constant.*

Proof. We reason by contraposition: we assume that f is not constant, which translates in $0 < \int_0^1 f < 1$ and we want to show that $f \notin H^{1/2}([0,1], \mathbb{R})$. Recall that it is sufficient to prove, given the definition of the $H^{1/2}$ norm [McL00, Chapter 3], that

$$\iint_{[0,1] \times [0,1]} \frac{|f(\eta) - f(\theta)|}{|\theta - \eta|^2} \mathrm{d}\eta \mathrm{d}\theta = +\infty.$$

Take $t > 0$ large enough. The function

$$\xi \mapsto \frac{1}{\sqrt{t}} \int_{\xi - t^{-1/2}/2}^{\xi + t^{-1/2}/2} f(\eta) \mathrm{d}\eta$$

is continuous on $[t^{-1/2}/2, 1-t^{-1/2}/2]$ and has a means which belongs to $[c, 1-c]$, where $0 < c < 1$ is independent of t (provided it is large enough). Hence, there exists ξ_t such that

$$\int_{\xi_t - t^{-1/2}/2}^{\xi_t + t^{-1/2}/2} f(\eta)\mathrm{d}\eta \in \left[\frac{c}{\sqrt{t}}, 1 - \frac{c}{\sqrt{t}}\right].$$

Heuristically, ξ_t is close to a point where f "jumps". On the segment $[\xi_t - t^{-1/2}/2, \xi_t + t^{-1/2}/2]$, there must be points for which $f = 0$ and points for which $f = 1$, and at least $ct^{-1/2}$ of each kind. In particular, it implies that

$$\mathcal{L}_{[0,1]} \otimes \mathcal{L}_{[0,1]} \left(\left\{ (\eta, \theta) \in \left[\xi_t - \frac{1}{2\sqrt{t}}, \xi_t + \frac{1}{2\sqrt{t}}\right]^2 \ : \ f(\eta) = 1 \text{ and } f(\theta) = 0 \right\} \right) \geqslant \frac{c^2}{t}.$$

As a consequence,

$$\mathcal{L}_{[0,1]} \otimes \mathcal{L}_{[0,1]} \left(\left\{ (\eta, \theta) \in [0,1]^2 \ : \ \frac{|f(\eta) - f(\theta)|}{|\theta - \eta|^2} \geqslant t \right\} \right) \geqslant \frac{c^2}{t}.$$

This estimate leads to

$$\iint_{[0,1]\times[0,1]} \frac{|f(\eta) - f(\theta)|}{|\theta - \eta|^2} \mathrm{d}\eta \mathrm{d}\theta$$
$$= \int_0^{+\infty} \left[\mathcal{L}_{[0,1]} \otimes \mathcal{L}_{[0,1]} \left(\left\{ (\eta, \theta) \in [0,1]^2 \ : \ \frac{|f(\eta) - f(\theta)|}{|\theta - \eta|^2} \geqslant t \right\} \right) \right] \mathrm{d}t = +\infty. \quad \square$$

With these two lemmas, we can easily arrive to our conclusion.

Proof of Lemma 8.40. Let $f \in H^{1/2}(\mathbb{S}^1, \mathbb{S}^1)$ such that $f(\xi)^2 = \xi$ for a.e. $\xi \in \mathbb{S}^1$. We want to show that f is continuous. Take X an arc of circle of $\mathbb{S}^1$. If X is small enough, there are two continuous functions f_0 and f_1 (the complex square roots) defined on X such that for all $\xi \in X$, $z^2 = \xi$ if and only if $z \in \{f_0(\xi), f_1(\xi)\}$. Moreover, if X is small, the ranges of f_0 and f_1 are far apart, hence we can find a Lipschitz function $u : \mathbb{B} \to \{0, 1\}$ such that $u \circ f_0 = 0$ and $u \circ f_1 = 1$ on X. Thus, $(u \circ f)(\xi) \in \{0, 1\}$ for $\xi \in X$. The previous lemmas allow us to conclude that the function is in $H^{1/2}(X, \{0, 1\})$, hence constant, which means that f is continuous on X. As X is arbitrary, f is continuous on $\mathbb{S}^1$, which is a contradiction. $\square$

8.5 Appendix: Measurable selection of the arg min

We want to show a result which states that if $F : X \times Y \to \mathbb{R}$ is a function which is measurable w.r.t. X, then one can find a selection $m : X \to Y$ such that $F(x, m(x)) = \min_Y F(x, \cdot)$ for every $x \in X$, i.e. such that $m(x) \in \arg\min_Y F(x, \cdot)$. First we recall the following result which can be found in [AB06, Theorem 18.19].

Proposition 8.45. *Let X be a measured space and Y a polish space. Let $F : X \times Y \to \mathbb{R}$ a function such that $F(x, \cdot) : Y \to \mathbb{R}$ is continuous for every $x \in X$, and $F(\cdot, y) : X \to \mathbb{R}$ is measurable for every $y \in Y$. Assume that for every $x \in X$, the function $F(x, \cdot)$ has a minimizer over Y.*

Then there exists $m : X \to Y$ a measurable function such that for all $x \in X$,

$$F(x, m(x)) = \min_{y \in Y} F(x, y).$$

However, in particular for Proposition 9.8 below, we need a case where $F(x, \cdot)$ is only l.s.c.. Thus, we prove some *ad hoc* result relying on the particular structure of our problem which allows to treat lower semi-continuity.

Lemma 8.46. *Let X be a measured space and Y a compact metrizable space. Let $F : X \times Y \to \mathbb{R}$ a function such that $F(x, \cdot) : Y \to \mathbb{R}$ is continuous for every $x \in X$, and $F(\cdot, y) : X \to \mathbb{R}$ is measurable for every $y \in Y$; and let $G : Y \to \mathbb{R}$ a l.s.c. function.*

Then the function $H : X \to \mathbb{R}$ defined by

$$H(x) := \min_y \{F(x, y) + G(y) \ : \ y \in Y\}$$

is measurable.

Proof. Notice that Y is separable as it is compact and metrizable. For any rational number a, the exists a sequence dense in $\{y \in Y \ : \ G(y) \leqslant a\}$. Hence, we can construct a sequence $(y_n)_{n\in\mathbb{N}}$ such that for any rational number a there is a subsequence of $(y_n)_{n\in\mathbb{N}}$ which is included and dense in $\{y \in Y \ : \ G(y) \leqslant a\}$.

Set $\tilde{H}(x) := \inf_n F(x, y_n) + G(y_n)$ which is measurable and larger than H. Let us prove that it is equal to H. Indeed, if $x \in X$, by standard arguments of calculus of variations, there exists $\bar{y}$ such that $H(x) = F(x, \bar{y}) + G(\bar{y})$. For any $a > G(\bar{y})$ rational, take a subsequence $(y_{n_k})_{k\in\mathbb{N}}$ which belongs to $\{y \in Y \ : \ G(y) \leqslant a\}$ and which converges to $\bar{y}$. By continuity of F, one has

$$\tilde{H}(x) \leqslant \liminf_{k\to+\infty} \left(F(x, y_{n_k}) + G(y_{n_k})\right) \leqslant F(x, \bar{y}) + a.$$

As a can be chosen arbitrary close to $G(\bar{y})$, we have that $\tilde{H}(x) \leqslant F(x, \bar{y}) + G(\bar{y}) = H(x)$. □

Proposition 8.47. *Let X be a measured space and Y a compact metrizable space. Let $F : X \times Y \to \mathbb{R}$ a function such that $F(x, \cdot) : Y \to \mathbb{R}$ is continuous for every $x \in X$, and $F(\cdot, y) : X \to \mathbb{R}$ is measurable for every $y \in Y$; and let $G : Y \to \mathbb{R}$ a l.s.c. function.*

Then there exists $m : X \to Y$ a measurable function such that for any $x \in X$,

$$F(x, m(x)) + G(m(x)) = \min_y \{F(x, y) + G(y) \ : \ y \in Y\}.$$

Proof. As in the previous lemma, define $H(x) := \min\{F(x, y) + G(y) \ : \ y \in Y\}$, it is a measurable function valued in $\mathbb{R}$. Let Γ be the mapping going from X and valued in the compact subsets of Y defined by $\Gamma(x) = \arg\min_x (F(x, \cdot) + G(\cdot))$ which means

$$\Gamma(x) := \{y \in Y \ : \ F(x, y) + G(y) = H(x)\}.$$

Notice that $\Gamma(x)$ is never empty thanks to standard arguments of calculus of variations. To prove the existence of a measurable selection of Γ, we rely on [AB06, Theorem 18.13]: it is sufficient to show that Γ is measurable, which means that $\{x \in X \ : \ \Gamma(x) \cap Z \neq \emptyset\}$ is a measurable set of X for any closed set $Z \subset Y$. But one can be convinced that, for a fixed $Z \subset Y$ closed,

$$\Gamma(x) \cap Z \neq \emptyset \ \Leftrightarrow \ H(x) = H_Z(x),$$

where $H_Z(x) := \min\{F(x, z) + G(z) \ : \ z \in Z\}$. Thanks to Lemma 8.46, both H and H_Z are measurable, thus the set on which they coincide is measurable, which concludes the proof. □

Chapter 9

The maximum principle

As explained in Chapter 7, we want to show in this chapter that $F \circ \boldsymbol{\mu}$ is subharmonic (which means $\Delta(F \circ \boldsymbol{\mu}) \geqslant 0$) as soon as $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ is harmonic and $F : \mathcal{P}(D) \to \mathbb{R}$ is convex along generalized geodesics. As far as the regularity of F is concerned the simplest would be to assume that F is continuous on $\mathcal{P}(D)$. Nevertheless, this assumption is very strong and excludes natural functionals (like the internal energies). In the case where F is only l.s.c., we will need additional assumptions: it is the object of the following definition.

Definition 9.1. *We say that $F : \mathcal{P}(D) \to \overline{\mathbb{R}}$ is regular if it is l.s.c. on $\mathcal{P}(D)$, if*

$$\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D)) \mapsto \int_\Omega F(\boldsymbol{\mu}(\xi)) \mathrm{d}\xi$$

is l.s.c. for the weak convergence on $L^2(\Omega, \mathcal{P}(D))$, and if F is bounded on the bounded sets of $L^\infty(D) \cap \mathcal{P}(D)$.

Lower semi-continuity of F is a reasonable assumption. To impose that F is bounded on bounded sets of $L^\infty(D) \cap \mathcal{P}(D)$ is not a strong constraint as D is compact, we will need it to ensure that, by regularizing probability measures with the heat flow, we get measures for which F is finite.

Lower semi-continuity of $\mathfrak{F} : \boldsymbol{\mu} \mapsto \int_\Omega (F \circ \boldsymbol{\mu})$ is less usual: by a standard argument left to the reader, it implies that F is convex for the affine structure on $\mathcal{P}(D)$. However, we do not know in the general case if the fact that F is convex and l.s.c. on $\mathcal{P}(D)$ is enough to ensure lower semi-continuity of $\mathfrak{F}$. Indeed, to apply abstract functional analysis arguments, we would like to work in the space $\mathcal{M}(\Omega \times D)$ endowed with the total variation norm: it is the dual of the Banach space $(C(\Omega \times D), \| \cdot \|_\infty)$. If F is convex and l.s.c. on $\mathcal{P}(D)$, it can be shown easily that $\mathfrak{F}$ is convex and l.s.c. on $\mathcal{M}(\Omega \times D)$ endowed with the total variation norm. However, it only implies that $\mathfrak{F}$ is l.s.c. for the topology on $\mathcal{M}(\Omega \times D)$ defined by duality w.r.t. the dual of $\mathcal{M}(\Omega \times D)$, the latter being strictly larger than $C(\Omega \times D)$.

However, for the usual functionals on $\mathcal{P}(D)$ we can do an *ad hoc* analysis and we have the following results.

Proposition 9.2. *Let $V \in L^1(D)$ a l.s.c. function. Then the functional*

$$F : \mu \in \mathcal{P}(D) \mapsto \int_D V \mathrm{d}\mu$$

is regular.

Let $f : [0, +\infty) \to \mathbb{R}$ a proper and convex function such that $\lim_{t\to+\infty} f(t)/t = +\infty$. Then the functional defined by

$$F : \mu \in \mathcal{P}(D) \mapsto \begin{cases} \int_D f(\mu(x))\mathrm{d}x & \text{if } \mu \text{ is absolutely continuous w.r.t. } \mathcal{L}_D \\ +\infty & \text{else,} \end{cases}$$

is regular.

Proof. As V is l.s.c. on the compact D, it is bounded from below. As V is in $L^1(\Omega)$, the function F is clearly bounded on bounded sets of $L^\infty(D) \cap \mathcal{P}(D)$. Then, we can use [San15, Proposition 7.1], seeing either V as a l.s.c. function on D, or as a l.s.c. on $\Omega \times D$ (constant w.r.t. its first variable) to get that both F and $\int_\Omega (F \circ \cdot)$ are l.s.c.

For the internal energy, to get lower semi-continuity of F we rely on [San15, Proposition 7.7]. To get the lower semi-continuity of $\int_\Omega (F \circ \cdot)$, we can see that

$$\int_\Omega F(\boldsymbol{\mu}(\xi))\mathrm{d}\xi = \begin{cases} \iint\limits_{\Omega\times D} f(\boldsymbol{\mu}(\xi, x))\mathrm{d}\xi\mathrm{d}x & \text{if } \boldsymbol{\mu} \text{ is absolutely continuous w.r.t. } \mathcal{L}_\Omega \otimes \mathcal{L}_D \\ +\infty & \text{else,} \end{cases}$$

thus [San15, Proposition 7.7] still applies. As f is bounded on bounded sets of $[0, +\infty)$, we see that F is bounded on bounded sets of $L^\infty(D) \cap \mathcal{P}(D)$. □

However, the interaction energy is not regular: it lacks convexity w.r.t. the affine structure on $\mathcal{P}(D)$ [San15, Chapter 7]. For instance, take $\Omega = D = [0, 1]$ and define $F : \mathcal{P}(D) \to \mathbb{R}$ by

$$F(\mu) := \iint\limits_{D\times D} |x - y|^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y).$$

This functional is continuous and bounded on $\mathcal{P}(D)$. However, if we define $\boldsymbol{\mu}_n(\xi) := \delta_{x_n(\xi)}$ with $x_n(\xi) = 1/2 + 1/2\cos(n\xi)$, one can see that $F(\boldsymbol{\mu}_n(\xi)) = 0$ for all $\xi \in \Omega$ and $n \in \mathbb{N}$, but $(\boldsymbol{\mu}_n)_{n\in\mathbb{N}}$ converges weakly on $\mathcal{P}(\Omega \times D)$ to $\boldsymbol{\mu} := \mathcal{L}_\Omega \otimes \mathcal{L}_D$, for which the value $\int_\Omega (F \circ \boldsymbol{\mu})$ is strictly positive. On the other hand, as soon as the interaction potential is continuous, the interaction energy is continuous on $\mathcal{P}(D)$.

Finally, let us recall that a function $f : \Omega \to \mathbb{R}$ is said subharmonic on $\mathring{\Omega}$ in the sense of distributions if $\Delta f \geqslant 0$ as a distribution in $\mathring{\Omega}$.

Theorem 9.3. *Let $F : \mathcal{P}(D) \to \overline{\mathbb{R}}$ a functional which is convex along generalized geodesics. Assume either that F is continuous on $\mathcal{P}(D)$ or that F is regular. Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(D)$ a Lipschitz mapping such that $\sup(F \circ \boldsymbol{\mu}_l) < +\infty$.*

Then there exists at least one solution $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ of the Dirichlet problem with boundary conditions $\boldsymbol{\mu}_l$ such that $(F \circ \boldsymbol{\mu}) : \Omega \to \mathbb{R}$ is subharmonic in $\mathring{\Omega}$ in the sense of distributions and

$$\operatorname*{ess\,sup}_\Omega (F \circ \boldsymbol{\mu}) \leqslant \sup_{\partial\Omega} (F \circ \boldsymbol{\mu}_l). \tag{9.1}$$

Moreover, if F is regular then $\boldsymbol{\mu}$ can be chosen in such a way that

$$\int_\Omega F(\boldsymbol{\mu}(\xi))\mathrm{d}\xi \leqslant \int_\Omega F(\boldsymbol{\nu}(\xi))\mathrm{d}\xi. \tag{9.2}$$

if $\boldsymbol{\nu}$ is any other solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$.

Let us make some comments. The first one is that (9.1) is nothing else than the maximum principle. It is not implied by the subharmonicity of $(F \circ \boldsymbol{\mu})$ as the latter holds only in $\mathring{\Omega}$ and we do not know if $(F \circ \boldsymbol{\mu})$ is continuous. The second one is that (9.2) characterizes $\boldsymbol{\mu}$ if F is strictly convex. More generally, the subharmonicity of $F \circ \boldsymbol{\mu}$ would hold for $\boldsymbol{\mu}$ solution of the Dirichlet problem minimizing

$$\int_\Omega a(\xi) F(\boldsymbol{\mu}(\xi)) \mathrm{d}\xi,$$

where $a \in C(\Omega)$ is a continuous and strictly positive function (it comes from a slight modification of the proof which is left to the reader). The last comment is that this result is somehow disappointing because we cannot guarantee the subharmonicity to hold for all solutions. The main issue is that we reason by approximation, thus the solution $\boldsymbol{\mu}$ is constructed as the limit of some approximate mappings, the existence of the limit is coming from compactness. But as we have no uniqueness result for the Dirichlet problem, we can only identify the limit through (9.2) (which is a byproduct of the approximation process) but we cannot say much more.

The rest of this chapter is devoted to the proof of Theorem 9.3. In Section 9.1 we prove some preliminary results. The most difficult and interesting case is the one where F is not assumed to be continuous but only regular: it is the object of Sections 9.2 and 9.3. To conclude, in Section 9.4, we briefly comment about the simplifications of the proof in the case of a continuous F.

9.1 Preliminary results

We prove first some technical results which would have overburden the previous chapters. The first one deals with Rellich compactness theorem, as we will want some strong convergence of our solutions of the approximate problems.

Proposition 9.4. *Let $(\boldsymbol{\mu}_n)_{n \in \mathbb{N}}$ a sequence in $H^1(\Omega, \mathcal{P}(D))$ such that $\sup_n \mathrm{Dir}(\boldsymbol{\mu}_n) < +\infty$. Then, up to extraction, the sequence $(\boldsymbol{\mu}_n)_{n \in \mathbb{N}}$ converges strongly in $L^2(\Omega, \mathcal{P}(D))$ to some $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$.*

Proof. This is nothing else than the Rellich compactness theorem, but for mappings valued in metric spaces. Remark that $\mathcal{P}(D)$ has a finite diameter, thus in this result we only need a control on the Dirichlet energy of $\boldsymbol{\mu}_n$. We can find this result for instance in [KS93, Theorem 1.13] or in [AT03, Theorem 5.4.3]. In any way, this result is also a consequence of the next proposition. □

In fact, we will need a stronger result, as we want to show compactness if we only have a control of the approximate Dirichlet energies.

Proposition 9.5. *Let $(\boldsymbol{\mu}_\varepsilon)_{\varepsilon > 0}$ a family in $L^2(\Omega, \mathcal{P}(D))$ such that $\liminf_\varepsilon \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_\varepsilon) < +\infty$. Then there exists a sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ which goes to 0 such that $(\boldsymbol{\mu}_{\varepsilon_n})_{n \in \mathbb{N}}$ converges strongly in $L^2(\Omega, \mathcal{P}(D))$ to some $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$.*

There is a well known criterion for compactness in $L^2(\Omega)$: the Riesz-Fréchet-Kolmogorov theorem. It requires a uniform control of the L^2-norm of the difference between a function and its translation. Here, we have only a control of the distance between a function and its translated in average (thanks to Dir_ε), and our mappings take values in $\mathcal{P}(D)$ rather than $\mathbb{R}$. Nevertheless, the strategy of the proof of the Riesz-Fréchet-Kolmogorov theorem is rather straightforward to adapt.

Proof. There exists a sequence $(\varepsilon_m)_{m \in \mathbb{N}}$, converging to 0, such that $\sup_m \mathrm{Dir}_{\varepsilon_m}(\boldsymbol{\mu}_{\varepsilon_m}) < +\infty$.

As in the proof of Theorem 8.14, let χ be a smooth function, radial, compactly supported in $B(0,1)$ and we set $\chi_t(\xi) = t^{-p} \chi(\xi / t)$. We will regularize $\boldsymbol{\mu}_{\varepsilon_m}$ only w.r.t. the source space Ω. More specifically, for any $\tilde{\Omega}$ compactly supported in $\mathring{\Omega}$ and t small enough, we define $\tilde{\boldsymbol{\mu}}_{m,t} \in L^2(\tilde{\Omega}, \mathcal{P}(D))$ by

$$\tilde{\boldsymbol{\mu}}_{m,t}(\xi) := \int_\Omega \chi_t(\xi - \eta)\boldsymbol{\mu}_{\varepsilon_m}(\eta)\mathrm{d}\eta \tag{9.3}$$

for any $\xi \in \tilde{\Omega}$. We first estimate $d_{L^2}(\tilde{\boldsymbol{\mu}}_{m,t}, \boldsymbol{\mu}_{\varepsilon_m}|_{\tilde{\Omega}})$. Using Jensen's inequality and the definition of Dir_t,

$$\begin{aligned} d_{L^2}(\tilde{\boldsymbol{\mu}}_{m,t}, \boldsymbol{\mu}_{\varepsilon_m}|_{\tilde{\Omega}}) &= \int_{\tilde{\Omega}} W_2^2\left(\int_{B(0,t)} \chi_t(\eta)\boldsymbol{\mu}_{\varepsilon_m}(\xi - \eta)\mathrm{d}\eta, \boldsymbol{\mu}(\xi)\right)\mathrm{d}\xi \\ &\leqslant \int_{\tilde{\Omega}}\int_{B(0,t)} \chi_t(\eta) W_2^2\left(\boldsymbol{\mu}_{\varepsilon_m}(\xi-\eta), \boldsymbol{\mu}(\xi)\right)\mathrm{d}\eta\mathrm{d}\xi \\ &\leqslant \frac{2t^{p+2}\|\chi_t\|_\infty}{C_p}\mathrm{Dir}_t(\boldsymbol{\mu}_{\varepsilon_m}) = Ct^2\mathrm{Dir}_t(\boldsymbol{\mu}_{\varepsilon_m}). \end{aligned}$$

Now, because of the monotonicity of Dir_t (Theorem 8.26) remember that $\mathrm{Dir}_t(\boldsymbol{\mu}_{\varepsilon_m}) \leqslant \mathrm{Dir}_{\varepsilon_m}(\boldsymbol{\mu}_{\varepsilon_m})$ if m is large enough (and t should in fact be of the form $2^N\varepsilon_m$ but it does not really matter). In consequence, for any $\delta > 0$, there exists $t > 0$ (small) and $m_0 \in \mathbb{N}$, such that for any $m \geqslant m_0$,

$$d_{L^2}(\tilde{\boldsymbol{\mu}}_{m,t}, \boldsymbol{\mu}_{\varepsilon_m}|_{\tilde{\Omega}}) \leqslant \delta.$$

On the other hand, for a fixed $t > 0$, we want to show compactness of the family $(\tilde{\boldsymbol{\mu}}_{m,t})$ in $L^2(\tilde{\Omega}, \mathcal{P}(D))$. We will show that this family is uniformly equi-Hölder as mappings defined on $\tilde{\Omega}$ and valued in $(\mathcal{P}(D), W_2)$: it implies compactness in $C(\tilde{\Omega}, \mathcal{P}(D))$ from which we easily deduce compactness in $L^2(\tilde{\Omega}, \mathcal{P}(D))$. Here $\tilde{\Omega}$ is a compact subset of Ω lying at a distance larger than t from $\partial\Omega$. We prefer to work on the 1-Wasserstein distance whose definition is recalled in Section 2.1. Take $\varphi \in C(D)$ a 1-Lipschitz function, up to translation by a constant we can assume that $\|\varphi\|_\infty \leqslant C$ with C independent of φ. Then for any $\xi, \eta \in \tilde{\Omega}$,

$$\begin{aligned} \int_D \varphi(x)\tilde{\boldsymbol{\mu}}_{m,t}(\xi, \mathrm{d}x) - \int_D \varphi(x)\tilde{\boldsymbol{\mu}}_{m,t}(\eta, \mathrm{d}x) &= \iint\limits_{\tilde{\Omega}\times D} \varphi(x)\left(\chi_t(\xi-\theta) - \chi_t(\eta-\theta)\right)\boldsymbol{\mu}(\theta, \mathrm{d}x)\mathrm{d}\theta \\ &\leqslant |\xi - \eta|\frac{1}{t^{p+1}}\|\chi'\|_\infty\|\varphi\|_\infty. \end{aligned}$$

As the bound is independent on φ, we deduce that $W_1(\tilde{\boldsymbol{\mu}}_{m,t}(\xi), \tilde{\boldsymbol{\mu}}_{m,t}(\eta)) \leqslant Ct^{-(p+1)}|\xi - \eta|$ for all ξ and η in $\tilde{\Omega}$. Using $W_2 \leqslant C\sqrt{W_1}$ [San15, Equation (5.1)], we see that, for a fixed t, the family $(\tilde{\boldsymbol{\mu}}_{m,t})_{m\in\mathbb{N}}$, defined on $\tilde{\Omega}$, is uniformly equi-continuous (more precisely 1/2-Hölder continuous).

Now we put the pieces together. For each $n \geqslant 1$, take $\tilde{\Omega}_n \subset \mathring{\Omega}$ compactly supported in $\mathring{\Omega}$ such that $\mathcal{L}_\Omega(\Omega\backslash\tilde{\Omega}_n) \leqslant 1/n$. Choose also t_n small enough such that $d_{L^2}(\tilde{\boldsymbol{\mu}}_{m,t_n}, \boldsymbol{\mu}_{\varepsilon_m}|_{\tilde{\Omega}_n}) \leqslant 1/n$ holds for m large enough and the distance between $\tilde{\Omega}_n$ and $\partial\Omega$ is smaller than t_n. Then, using Ascoli-Arzelà theorem, up to a subsequence, we know that $(\tilde{\boldsymbol{\mu}}_{m,t_n})_{m\in\mathbb{N}}$ converges strongly in $L^2(\tilde{\Omega}_n, \mathcal{P}(D))$, in particular it is a Cauchy sequence. Up to a diagonal extraction in $(\varepsilon_m)_{m\in\mathbb{N}}$ (we do not relabel the sequence), we can assume that $(\tilde{\boldsymbol{\mu}}_{m,t_n}|_{\tilde{\Omega}_n})_{m\in\mathbb{N}}$ is a Cauchy sequence for all $n \in \mathbb{N}$. Notice, as $\mathcal{P}(D)$ has a finite diameter, that $|d_{L^2}(\boldsymbol{\mu}, \boldsymbol{\nu}) - d_{L^2}(\boldsymbol{\mu}|_{\tilde{\Omega}_n}, \boldsymbol{\nu}|_{\tilde{\Omega}_n})| \leqslant C/n$ for all $\boldsymbol{\mu}, \boldsymbol{\nu} \in L^2(\Omega, \mathcal{P}(D))$. Hence, for any $n \in \mathbb{N}$, one has for m and m' large enough,

$$\begin{aligned} d_{L^2}(\boldsymbol{\mu}_{\varepsilon_m}, \boldsymbol{\mu}_{\varepsilon_{m'}}) &\leqslant d_{L^2}(\boldsymbol{\mu}_{\varepsilon_m}|_{\tilde{\Omega}_n}, \tilde{\boldsymbol{\mu}}_{m,t_n}) + d_{L^2}(\tilde{\boldsymbol{\mu}}_{m,t_n}, \tilde{\boldsymbol{\mu}}_{m',t_n}) + d_{L^2}(\boldsymbol{\mu}_{\varepsilon_{m'}}|_{\tilde{\Omega}_n}, \tilde{\boldsymbol{\mu}}_{m',t_n}) + \frac{2C}{n} \\ &\leqslant \frac{2+2C}{n} + d_{L^2}(\tilde{\boldsymbol{\mu}}_{m,t_n}, \tilde{\boldsymbol{\mu}}_{m',t_n}), \end{aligned}$$

and $d_{L^2}(\tilde{\boldsymbol{\mu}}_{m,t_n}, \tilde{\boldsymbol{\mu}}_{m',t_n})$ can be made arbitrary small for m and m' large enough. In other words, $(\boldsymbol{\mu}_{\varepsilon_m})_{m\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{P}(D))$, thus it converges strongly. □

We will also need a result about the boundary conditions. Indeed, as the minimizers of Dir_ε will only live in $L^2(\Omega, \mathcal{P}(D))$, we cannot define and impose boundary values. To bypass this difficulty, we extend slightly our domain into a larger domain $\Omega_e \supset \Omega$ and impose the values of the mappings everywhere on $\Omega_e \backslash \mathring{\Omega}$.

Proposition 9.6. *Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(D)$ a Lipschitz mapping. There exists a compact Ω_e such that $\Omega \subset \mathring{\Omega}_e$, and a Lipschitz mapping $\boldsymbol{\mu}_e \in L^2(\Omega_e\backslash\mathring{\Omega}, \mathcal{P}(D))$ such that $\boldsymbol{\mu}_e = \boldsymbol{\mu}_l$ on $\partial\Omega$ and*

$$\{\boldsymbol{\mu}_e(\xi) \ : \ \xi \in \Omega_e\backslash\mathring{\Omega}\} = \{\boldsymbol{\mu}_l(\xi) \ : \ \xi \in \partial\Omega\}. \tag{9.4}$$

Moreover, a mapping $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ satisfies $\boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_l$ if and only if the mapping $\tilde{\boldsymbol{\mu}}$ defined on Ω_e by

$$\tilde{\boldsymbol{\mu}}(\xi) = \begin{cases} \boldsymbol{\mu}(\xi) & \text{if } \xi \in \mathring{\Omega} \\ \boldsymbol{\mu}_e(\xi) & \text{if } \xi \in \Omega_e\backslash\mathring{\Omega}, \end{cases}$$

belongs to $H^1(\Omega_e, \mathcal{P}(D))$.

Proof. As Ω has a Lipschitz boundary, one can say [KS93, Section 1.12] that there exists a compact Ω_e such that $\Omega \subset \mathring{\Omega}_e$, and $\Psi : [0,1] \times \partial\Omega \to \Omega_e\backslash\mathring{\Omega}$ a bilipschitz mapping such that $\Psi(0,\cdot)$ is the identity on $\partial\Omega$. Roughly speaking, $\Psi(t,\xi)$ should be thought as $\xi + t\mathbf{n}_\Omega(\xi)$ where $\mathbf{n}_\Omega$ is the outward normal to $\partial\Omega$. Then, one can define

$$\boldsymbol{\mu}_e(\Psi(t,\xi)) := \boldsymbol{\mu}_l(\xi)$$

for every $t \in [0,1]$ and $\xi \in \partial\Omega$: we extend $\boldsymbol{\mu}_l$ by keeping it constant along the normal to $\partial\Omega$. Because Ψ is bilipschitz and $\boldsymbol{\mu}_l$ is Lipschitz, it is clear that $\boldsymbol{\mu}_e$ is a Lipschitz mapping. Moreover, by construction, (9.4) obviously holds.

Let us prove the second point. Take $\mathbf{E} \in \mathcal{M}(\Omega \times D, \mathbb{R}^{pq})$ and $\mathbf{E}_e \in \mathcal{M}((\Omega_e\backslash\mathring{\Omega}) \times D, \mathbb{R}^{pq})$ the momenta tangent to respectively $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_e$. The tangent momentum of $\tilde{\boldsymbol{\mu}}$, if it were to exist, must coincide with $\mathbf{E}$ on $\Omega \times D$ and with $\mathbf{E}_e$ on $(\Omega_e\backslash\mathring{\Omega}) \times D$ because of Corollary 8.12. Hence, if must be $\tilde{\mathbf{E}} \in \mathcal{M}(\Omega_e \times D, \mathbb{R}^{pq})$ defined by

$$\iint_{\Omega_e \times D} b \cdot \mathrm{d}\tilde{\mathbf{E}} = \iint_{\Omega \times D} b \cdot \mathrm{d}\mathbf{E} + \iint_{(\Omega_e\backslash\mathring{\Omega}) \times D} b \cdot \mathrm{d}\mathbf{E}_e.$$

As we already have $\mathrm{Dir}(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{E}}) < +\infty$, we see that $\tilde{\boldsymbol{\mu}} \in H^1(\Omega_e, \mathcal{P}(D))$ if and only if $(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{E}})$ satisfies the continuity equation. If $\varphi \in C^1_c(\Omega_e, \mathbb{R}^p)$,

$$\begin{aligned} &\iint_{\Omega_e \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\tilde{\boldsymbol{\mu}} + \iint_{\Omega_e \times D} \nabla_D \varphi \cdot \mathrm{d}\tilde{\mathbf{E}} \\ &= \iint_{\Omega \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu} + \iint_{\Omega \times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E} + \iint_{(\Omega_e\backslash\mathring{\Omega}) \times D} \nabla_\Omega \cdot \varphi \mathrm{d}\boldsymbol{\mu}_e + \iint_{(\Omega_e\backslash\mathring{\Omega}) \times D} \nabla_D \varphi \cdot \mathrm{d}\mathbf{E}_e \\ &= \mathrm{BT}_{\boldsymbol{\mu}}(\varphi) + \mathrm{BT}_{\boldsymbol{\mu}_e}(\varphi). \end{aligned}$$

By Whitney's theorem, the restriction of functions in $C^1_c(\mathring{\Omega}_e, \mathbb{R}^p)$ to Ω coincide with $C^1(\Omega, \mathbb{R}^p)$, thus we see that $\tilde{\boldsymbol{\mu}} \in H^1(\Omega_e, \mathcal{P}(D))$ if and only if $\mathrm{BT}_{\boldsymbol{\mu}} = -\mathrm{BT}_{\boldsymbol{\mu}_e}$. Considering the fact that the outward normal to $\Omega_e\backslash\mathring{\Omega}$ is $-\mathbf{n}_\Omega$, and that $\boldsymbol{\mu}_e$ is continuous with values on $\partial\Omega$ given by $\boldsymbol{\mu}_l$, there holds $\mathrm{BT}_{\boldsymbol{\mu}_e} = -\mathrm{BT}_{\boldsymbol{\mu}_l}$ hence the proposition is proved. □

9.2 The approximate problems and their optimality conditions

In all this subsection, we assume that F is regular. As explained before, we use Dir_ε to approximate Dir, as the optimality conditions of Dir_ε imply that for each $\xi \in \Omega$, $\boldsymbol{\mu}(\xi)$ is a barycenter of all $\boldsymbol{\mu}(\eta)$ for η in the ball of center ξ and radius ε.

Let us introduce some notations that we will keep during the rest of the proof. We denote by $\Omega_e \supset \Omega$ and $\boldsymbol{\mu}_e \in H^1(\Omega_e \backslash \mathring{\Omega}, \mathcal{P}(D))$ the objects given by Proposition 9.6. Take $\varepsilon_0 > 0$ such that $B(\xi, \varepsilon_0) \subset \Omega_e$ for all $\xi \in \partial\Omega$. We denote by

$$L^2_e(\Omega_e, \mathcal{P}(D)) := \{\boldsymbol{\mu} \in L^2(\Omega_e, \mathcal{P}(D)) \ : \ \boldsymbol{\mu}|_{\Omega_e \backslash \mathring{\Omega}} = \boldsymbol{\mu}_e\}$$

the set of L^2 mappings which coincide with $\boldsymbol{\mu}_e$ on $\Omega_e \backslash \mathring{\Omega}$. This set $L^2_e(\Omega_e, \mathcal{P}(D))$ is clearly closed for the weak convergence on $L^2(\Omega_e, \mathcal{P}(D))$, in particular it is compact for the weak convergence. We also define $H^1_e(\Omega_e, \mathcal{P}(D)) := H^1(\Omega_e, \mathcal{P}(D)) \cap L^2_e(\Omega_e, \mathcal{P}(D))$. In the rest of the proof, we extend the definitions of Dir_ε and Dir on $L^2_e(\Omega_e, \mathcal{P}(D))$. More precisely, if $\boldsymbol{\mu} \in L^2_e(\Omega_e, \mathcal{P}(D))$,

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) := C_p \iint\limits_{\Omega_e \times \Omega_e} \frac{W_2^2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta))}{2\varepsilon^{p+2}} \mathbb{1}_{|\xi - \eta| \leqslant \varepsilon} \mathrm{d}\xi \mathrm{d}\eta,$$

and

$$\begin{aligned} \mathrm{Dir}(\boldsymbol{\mu}) := \inf_{\mathbf{E}} \{ & \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) \ : \ \mathbf{E} \in \mathcal{M}(\Omega_e \times D, \mathbb{R}^{pq}) \\ & \text{and } (\boldsymbol{\mu}, \mathbf{E}) \text{ satisfies the continuity equation on } \Omega_e \times D\}. \end{aligned}$$

(we integrate over Ω_e and not only on Ω). We also use the notation

$$M := \sup_{\partial\Omega} (F \circ \boldsymbol{\mu}_l),$$

by assumption M is finite. Remark that by construction, if $\boldsymbol{\mu} \in L^2_e(\Omega_e, \mathcal{P}(D))$, then for all $\xi \in \Omega_e \backslash \mathring{\Omega}$ one has $F(\boldsymbol{\mu}(\xi)) \leqslant M$.

As F is l.s.c. on the compact set $\mathcal{P}(D)$, it is bounded from below. Hence, we can translate it by a constant and assume that $F \geqslant 0$ on $\mathcal{P}(D)$.

Let $\varepsilon > 0$ and $\lambda > 0$ be fixed. The approximate problem is defined as

$$\min_{\boldsymbol{\mu}} \left\{ \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) + \lambda \int_{\Omega_e} F(\boldsymbol{\mu}(\xi)) \mathrm{d}\xi \ : \ \boldsymbol{\mu} \in L^2_e(\Omega_e, \mathcal{P}(D)) \right\}. \tag{9.5}$$

To add the term $\lambda \int_{\Omega_e} F \circ \boldsymbol{\mu}$ has two purposes: on the one hand, it ensures that $F \circ \boldsymbol{\mu}$ will be regular enough (namely in $L^1(\Omega_e)$) to extract information from the optimality conditions; on the other hand by taking the limit $\varepsilon \to 0$ and then $\lambda \to 0$, we will be able to say that $F \circ \boldsymbol{\mu}_{\varepsilon,\lambda}$ (where $\boldsymbol{\mu}_{\varepsilon,\lambda}$ is a minimizer of the approximate problem) converges pointewisely, and it is necessary to pass to the limit the (approximate) subharmonicity that we will get from the optimality conditions of the approximate problem.

The following result is easy with all the tools developed above.

Proposition 9.7. *For any $\varepsilon > 0$ and $\lambda > 0$, there exists a solution to the approximate problem* (9.5).

Proof. Let $\nu \in \mathcal{P}(D)$ any measure such that $F(\nu) < +\infty$ (it exists as F is regular). If we define $\boldsymbol{\mu} \in L^2_e(\Omega_e, \mathcal{P}(D))$ by $\boldsymbol{\mu}|_{\mathring{\Omega}} := \nu$ and $\boldsymbol{\mu}|_{\Omega_e \backslash \mathring{\Omega}} := \boldsymbol{\mu}_e$, one can see that $\int_{\Omega_e} F(\boldsymbol{\mu}(\xi))\mathrm{d}\xi < +\infty$, moreover as $\mathcal{P}(D)$ has a finite diameter $\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) < +\infty$. Hence, the minimization problem is non empty. In consequence, we are minimizing over the set $L^2_e(\Omega_e, \mathcal{P}(D))$, which is compact for the weak convergence, a functional which is l.s.c. (see Proposition 8.25 and the regularity assumption on F): we can use the direct method of calculus of variations. □

Starting from now, for any $\varepsilon > 0$ and $\lambda > 0$, we denote by $\boldsymbol{\mu}_{\varepsilon,\lambda}$ a solution of the approximate problem (9.5).

Proposition 9.8. *Let $0 < \varepsilon \leqslant \varepsilon_0$ and $\lambda > 0$ be fixed. Then for a.e. $\xi \in \Omega$, $\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)$ is a minimizer over $\mathcal{P}(D)$ of*

$$\nu \mapsto \frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(\nu, \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\mathrm{d}\eta + \lambda F(\nu).$$

Proof. We reason by contradiction. If the property does not hold, there exists $c > 0$ and a set $X \subset \mathring{\Omega}$ of strictly positive measure such that for all $\xi \in X$,

$$\frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\mathrm{d}\eta + \lambda F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \\ \geqslant c + \min_{\nu \in \mathcal{P}(D)} \left(\frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(\nu, \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\mathrm{d}\eta + \lambda F(\nu) \right). \tag{9.6}$$

Now, consider $\delta > 0$ small and $Y \subset X$ such that $\mathcal{L}_\Omega(Y) = \delta$. On every point of $\xi \in Y$, we want to select a minimizer ν (which depends on ξ) of the r.h.s. of (9.6), and we want to dot it in a measurable way. Notice that

$$\nu \mapsto \frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(\nu, \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\mathrm{d}\eta + \lambda F(\nu)$$

is the sum of a functional continuous w.r.t. ν and measurable w.r.t. ξ, and the functional λF which is l.s.c. w.r.t. ν but which does not depend on ξ. The fact that F is only l.s.c. prevents us from using directly Proposition 8.45, though by some *ad hoc* measurable selection result which is stated and proved in the appendix at the end Chapter 8 (Proposition 8.47), one can still choose $\nu(\xi)$ a minimizer in such a way that it is measurable in ξ. In other words, we construct $\tilde{\boldsymbol{\mu}} \in L^2_e(\Omega_e, \mathcal{P}(D))$ such that $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_{\varepsilon,\lambda}$ on $\Omega_e \backslash Y$ and

$$\frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\mathrm{d}\eta + \lambda F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \\ \geqslant c + \left(\frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(\tilde{\boldsymbol{\mu}}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\mathrm{d}\eta + \lambda F(\tilde{\boldsymbol{\mu}}(\xi)) \right)$$

for all $\xi \in Y$. Now we evaluate:

$$\left(\mathrm{Dir}_\varepsilon(\tilde{\boldsymbol{\mu}}) + \lambda \int_{\Omega_e} F(\tilde{\boldsymbol{\mu}}(\xi))\mathrm{d}\xi \right) - \left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda \int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))\mathrm{d}\xi \right) \\ = \frac{C_p}{2\varepsilon^{p+2}} \iint_{\Omega_e \times \Omega_e} \left[W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\tilde{\boldsymbol{\mu}}(\xi), \tilde{\boldsymbol{\mu}}(\eta)) \right] \mathbb{1}_{|\xi-\eta| \leqslant \varepsilon} \mathrm{d}\xi \mathrm{d}\eta \\ + \lambda \int_Y [F(\tilde{\boldsymbol{\mu}}(\xi)) - F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))]\mathrm{d}\xi$$

The integral over $\Omega_e \times \Omega_e$ can be split over four parts: the one over $(\Omega_e \backslash Y) \times (\Omega_e \backslash Y)$, which vanishes because $\boldsymbol{\mu}_{\varepsilon,\lambda} = \tilde{\boldsymbol{\mu}}$ on this set; the one over $Y \times Y$, which can be bounded by $C\delta^2$, where C depends on the diameter of $\mathcal{P}(D)$ and on ε; and the ones over $(\Omega_e \backslash Y) \times Y$ and $Y \times (\Omega_e \backslash Y)$ which are equal by symmetry. Moreover, one has

$$\begin{aligned}
&\frac{C_p}{2\varepsilon^{p+2}} \iint_{Y \times (\Omega_e \backslash Y)} \left[W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\tilde{\boldsymbol{\mu}}(\xi), \tilde{\boldsymbol{\mu}}(\eta))\right] \mathbb{1}_{|\xi-\eta| \leqslant \varepsilon} \mathrm{d}\xi \mathrm{d}\eta \\
&\quad = \frac{C_p}{2\varepsilon^{p+2}} \iint_{Y \times (\Omega_e \backslash Y)} \left[W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\tilde{\boldsymbol{\mu}}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\right] \mathbb{1}_{|\xi-\eta| \leqslant \varepsilon} \mathrm{d}\xi \mathrm{d}\eta \\
&\quad \leqslant C\delta^2 + \frac{C_p}{2\varepsilon^{p+2}} \iint_{Y \times \Omega_e} \left[W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\tilde{\boldsymbol{\mu}}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\right] \mathbb{1}_{|\xi-\eta| \leqslant \varepsilon} \mathrm{d}\xi \mathrm{d}\eta \\
&\quad = C\delta^2 + \frac{C_p}{2\varepsilon^{p+2}} \int_Y \left(\int_{B(\xi,\varepsilon)} \left[W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\tilde{\boldsymbol{\mu}}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))\right] \mathrm{d}\eta \right) \mathrm{d}\xi,
\end{aligned}$$

where the inequality comes from the fact that we have add the piece $Y \times Y$ which is of size δ^2 and over which we integrate a function which is bounded. Notice that we have used that $B(\xi, \varepsilon) \subset \Omega_e$ for $\xi \in \Omega$ as $\varepsilon < \varepsilon_0$. The part on $(\Omega_e \backslash Y) \times Y$ gives exactly the same amount, thus

$$\begin{aligned}
&\left(\mathrm{Dir}_\varepsilon(\tilde{\boldsymbol{\mu}}) + \lambda \int_{\Omega_e} F(\tilde{\boldsymbol{\mu}}(\xi)) \mathrm{d}\xi \right) - \left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda \int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \mathrm{d}\xi \right) \\
&\quad \leqslant C\delta^2 + \int_Y \left(\frac{C_p}{\varepsilon^{p+2}} \left[\int_{B(\xi,\varepsilon)} [W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\tilde{\boldsymbol{\mu}}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))] \mathrm{d}\eta \right] \right. \\
&\qquad\qquad \left. + \lambda \left[F(\tilde{\boldsymbol{\mu}}(\xi)) - F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \right] \right) \mathrm{d}\xi \leqslant C\delta^2 - c\delta,
\end{aligned}$$

where the last inequality comes precisely form the way we chose $\tilde{\boldsymbol{\mu}}$ on Y and of $\mathcal{L}_\Omega(Y) = \delta$. Hence, taking δ small enough, the r.h.s. is strictly negative, which is a contradiction with the optimality of $\boldsymbol{\mu}_{\varepsilon,\lambda}$. ☐

Remark that if $\lambda = 0$, our proof still works, and it precisely shows that $\boldsymbol{\mu}_{\varepsilon,0}(\xi)$ is a barycenter of the $\boldsymbol{\mu}_{\varepsilon,0}(\eta)$ for η running over the ball of center ξ and radius ε, a fact which was already stated by Jost [Jos94]. The crucial result which allows us to get subharmonicity is the following, namely Jensen's inequality for functionals convex along generalized geodesics. Notice that $F \circ \boldsymbol{\mu}_{\varepsilon,\lambda}$ is integrable on Ω_e.

Proposition 9.9. *Let* $0 < \varepsilon \leqslant \varepsilon_0$ *and* $\lambda > 0$ *be fixed. Then, for a.e.* $\xi \in \Omega$,

$$F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \leqslant \frac{1}{|B(\xi,\varepsilon)|} \int_{B(\xi,\varepsilon)} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) \mathrm{d}\eta.$$

Proof. Let us take a point $\xi \in \Omega$ for which the conclusion of Proposition 9.8 holds and such that $F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) < +\infty$: it is the case for a.e. points of Ω. As a competitor, we use $S_t^F[\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)]$ for small $t > 0$, which means that we let $\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)$ follow the gradient flow of F, see Theorem 2.11. By Proposition 9.8,

$$\begin{aligned}
&\frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) \mathrm{d}\eta + \lambda F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \\
&\quad \leqslant \frac{C_p}{\varepsilon^{p+2}} \int_{B(\xi,\varepsilon)} W_2^2(S_t^F[\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)], \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) \mathrm{d}\eta + \lambda F(S_t^F[\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)]).
\end{aligned}$$

By the very definition of gradient flows, $F(S_t^F[\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)]) \leqslant F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))$. Thus, rearranging the terms and dividing by $2t > 0$,

$$\int_{B(\xi,\varepsilon)} \frac{W_2^2(S_t^F[\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)], \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))}{2t} \mathrm{d}\eta \geqslant 0.$$

For a.e. $\eta \in B(\xi,\varepsilon)$, one has that $F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) < +\infty$. Hence, using Theorem 2.11, we see that for a.e. $\eta \in B(\xi,\varepsilon)$, the quantity

$$\frac{W_2^2(S_t^F[\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)], \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - W_2^2(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi), \boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))}{2t}$$

has a lim sup bounded by $F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))$ and is uniformly bounded in t by $F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\eta))$ (by Theorem 2.11 and positivity of F), the latter being integrable on $B(\xi,\varepsilon)$. Hence, by Fatou's lemma, we can pass to the limit $t \to 0$ and conclude that

$$\int_{B(\xi,\varepsilon)} [F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))] \mathrm{d}\eta \geqslant 0.$$

The result follows by just rearranging the terms. □

Let us conclude this subsection by proving a maximum principle, but for mappings which are ε-subharmonic. Recall that M is the supremum of $F \circ \boldsymbol{\mu}$ on $\Omega_e \backslash \mathring{\Omega}$ for any $\boldsymbol{\mu} \in L_e^2(\Omega_e, \mathcal{P}(D))$.

Proposition 9.10. *Let $0 < \varepsilon \leqslant \varepsilon_0$ and $\lambda > 0$ be fixed. Then, for a.e. $\xi \in \Omega_e$, one has $F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \leqslant M$.*

Proof. Let $\delta > 0$ be fixed and consider $f_\delta : \Omega_e \to \overline{\mathbb{R}}$ defined by $f_\delta(\xi) = F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) + \delta|\xi - \xi_0|^2$, where ξ_0 is any point of Ω. By strict convexity of the square function and thanks to Proposition 9.9, for a.e. $\xi \in \Omega$,

$$\int_{B(\xi,\varepsilon)} [f_\delta(\eta) - f_\delta(\xi)] \mathrm{d}\eta > 0.$$

In particular, the essential supremum of f_δ cannot be reached on $\mathring{\Omega}$, it must be reached on $\Omega_e \backslash \mathring{\Omega}$. On $\Omega_e \backslash \mathring{\Omega}$ we control the values of $F \circ \boldsymbol{\mu}_{\varepsilon,\lambda}$ by M, in consequence $\operatorname{ess\,sup}_{\Omega_e} f_\delta \leqslant M + C\delta$, where C depends on the diameter of Ω. Sending δ to 0 (along a sequence), we get the result. □

9.3 Limit to the Dirichlet problem

In all this section, we still assume that F is regular.

The goal is now to pass to the limit and to show that $\boldsymbol{\mu}_{\varepsilon,\lambda}$ converges to $\boldsymbol{\mu}$ a solution of the Dirichlet problem such that $F \circ \boldsymbol{\mu}$ is subharmonic. Recall that Dir_ε Γ-converges to Dir when $\varepsilon \to 0$, see Theorem 8.26. To get subharmonicity, we will need strong convergence, it implies to take first the limit $\varepsilon \to 0$ and then $\lambda \to 0$. But on the other hand, we need a uniform bound on the minimal values of the approximate problems to pass to the limit. To get them implies that we need to produce at least one mapping $\boldsymbol{\mu}$ in $H_e^1(\Omega_e, \mathcal{P}(D))$ such that $\int_{\Omega_e} (F \circ \boldsymbol{\mu}) < +\infty$. To do this, we cannot rely on the Lipschitz extension: there is no way to guarantee that $\int_\Omega (F \circ \boldsymbol{\mu}) < +\infty$ with the construction used in the proof of Theorem 8.33. To get this uniform bound, we will take first the limit $\lambda \to 0$ and then $\varepsilon \to 0$ (relying only on weak convergence). It will produce a solution $\tilde{\boldsymbol{\mu}} \in H_e^1(\Omega_e, \mathcal{P}(D))$ of the Dirichlet problem with $\int_{\Omega_e} (F \circ \tilde{\boldsymbol{\mu}}) < +\infty$ but we cannot guarantee subharmonicity of $F \circ \tilde{\boldsymbol{\mu}}$. However it brings uniform bounds and enables us to take the limit $\varepsilon \to 0$, $\lambda \to 0$ and get a solution $\bar{\boldsymbol{\mu}}$ of the Dirichlet problem for which $F \circ \bar{\boldsymbol{\mu}}$ is subharmonic.

We take two sequences $(\varepsilon_n)_{n\in\mathbb{N}}$, $(\lambda_m)_{m\in\mathbb{N}}$ that both converge to 0 while being strictly positive. More precisely we take $\varepsilon_n := \varepsilon_0 2^{-n}$ for any $n \in \mathbb{N}$, thus we always have $\varepsilon_n \leqslant \varepsilon_0$ and $\mathrm{Dir}_{\varepsilon_n}$ converges in an increasing way and Γ-converges to Dir. We will not relabel the sequences when extracting subsequences. Moreover, to avoid heavy notations, we will drop the indexes n and m; and $\lim_{n\to+\infty}$, $\lim_{m\to+\infty}$ will be denoted respectively by $\lim_{\varepsilon\to 0}$ and $\lim_{\lambda\to 0}$.

Proposition 9.11. *Up to extraction, there exists $\tilde{\boldsymbol{\mu}} \in H^1_e(\Omega_e, \mathcal{P}(D))$ such that*

$$\tilde{\boldsymbol{\mu}} := \lim_{\varepsilon\to 0}\left(\lim_{\lambda\to 0} \boldsymbol{\mu}_{\varepsilon,\lambda}\right),$$

where the limits are taken weakly in $L^2_e(\Omega_e, \mathcal{P}(D))$. Moreover, $\tilde{\boldsymbol{\mu}}$ is a minimizer of Dir *in the space $H^1_e(\Omega_e, \mathcal{P}(D))$ and*

$$\int_{\Omega_e} F(\tilde{\boldsymbol{\mu}}(\xi))\mathrm{d}\xi < +\infty. \tag{9.7}$$

Proof. The existence of $\tilde{\boldsymbol{\mu}} \in L^2_e(\Omega_e, \mathcal{P}(D))$ is trivial: recall that $L^2_e(\Omega_e, \mathcal{P}(D))$ is compact for the weak convergence. Moreover, using Proposition 9.10, we have that for $\varepsilon \leqslant \varepsilon_0$ and $\lambda > 0$,

$$\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))\mathrm{d}\xi \leqslant M|\Omega_e|.$$

By the regularity assumption on F, we can pass this inequality to the weak limit and get (9.7).

The minimizing property of $\tilde{\boldsymbol{\mu}}$ is more involved. Assume by contradiction that there exists $\boldsymbol{\nu} \in H^1_e(\Omega, \mathcal{P}(D))$ such that $\mathrm{Dir}(\boldsymbol{\nu}) < \mathrm{Dir}(\tilde{\boldsymbol{\mu}})$. By the Γ-convergence of Dir_ε to Dir and the positivity of F, one has

$$\mathrm{Dir}(\boldsymbol{\nu}) < \mathrm{Dir}(\tilde{\boldsymbol{\mu}}) \leqslant \liminf_{\varepsilon\to 0}\left(\liminf_{\lambda\to 0}\left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))\mathrm{d}\xi\right)\right).$$

In particular, we can choose $\varepsilon > 0$ small enough such that (by monotonicity of Dir_ε)

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\nu}) \leqslant \mathrm{Dir}(\boldsymbol{\nu}) < \liminf_{\lambda\to 0}\left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))\mathrm{d}\xi\right).$$

We regularize $\boldsymbol{\nu}$ in the following way: for $t > 0$, we denote by $\boldsymbol{\nu}_t := (\mathbb{1}_{\mathring{\Omega}}\Phi^D_t)\boldsymbol{\nu}$ the element of $L^2_e(\Omega_e, \mathcal{P}(D))$ for which the heat flow on D has been followed only in $\mathring{\Omega}$: in other words, for any $t > 0$,

$$\boldsymbol{\nu}_t(\xi) := \begin{cases} (\Phi^D_t)[\boldsymbol{\nu}(\xi)] & \text{if } \xi \in \mathring{\Omega}, \\ \boldsymbol{\nu}(\xi) = \boldsymbol{\mu}_e(\xi) & \text{if } \xi \in \Omega_e\backslash\mathring{\Omega}. \end{cases}$$

Clearly, $\boldsymbol{\nu}_t \in L^2_e(\Omega_e, \mathcal{P}(D))$. Moreover, as $W_2(\boldsymbol{\nu}_t(\xi), \boldsymbol{\nu}(\xi)) \leqslant \omega(t)$ with $\omega(t) \to 0$ as $t \to 0$ (see Proposition 2.14), we see that $\boldsymbol{\nu}_t$ converges strongly in $L^2_e(\Omega_e, \mathcal{P}(D))$ to $\boldsymbol{\nu}$. In particular, thanks to the continuity of Dir_ε, there exists t small enough such that

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\nu}_t) < \liminf_{\lambda\to 0}\left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))\mathrm{d}\xi\right).$$

Because of the standard $L^\infty - L^1$ estimate for the heat flow (see (ii) of Proposition 2.13), one has that $\{\boldsymbol{\nu}_t(\xi) \ : \ \xi \in \mathring{\Omega}\}$ is included in a bounded set of $L^\infty(D) \cap \mathcal{P}(D)$. In particular, $F \circ \boldsymbol{\nu}_t$ is bounded on $\mathring{\Omega}$. As it is also bounded on $\Omega_e\backslash\mathring{\Omega}$ by M, we see that $\int_{\Omega_e} F \circ \boldsymbol{\nu}_t < +\infty$. Hence, for some λ small enough,

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\nu}_t) + \lambda\int_{\Omega_e} F(\boldsymbol{\nu}_t(\xi))\mathrm{d}\xi < \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))\mathrm{d}\xi,$$

which is a contradiction with the optimality of $\boldsymbol{\mu}_{\varepsilon,\lambda}$. □

Proposition 9.12. *Up to extraction, there exists* $\bar{\boldsymbol{\mu}} \in H^1_e(\Omega_e, \mathcal{P}(D))$ *such that*

$$\bar{\boldsymbol{\mu}} := \lim_{\lambda \to 0} \left(\lim_{\varepsilon \to 0} \boldsymbol{\mu}_{\varepsilon,\lambda} \right),$$

where the limits are taken strongly in $L^2_e(\Omega_e, \mathcal{P}(D))$. *Moreover,* $\bar{\boldsymbol{\mu}}$ *is a minimizer of* Dir *in the space* $H^1_e(\Omega_e, \mathcal{P}(D))$ *and for any other minimizer* $\boldsymbol{\nu}$ *of* Dir *in* $H^1_e(\Omega_e, \mathcal{P}(D))$,

$$\int_{\Omega_e} F(\bar{\boldsymbol{\mu}}(\xi)) \mathrm{d}\xi \leqslant \liminf_{\lambda \to 0} \left(\liminf_{\varepsilon \to 0} \left(\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \mathrm{d}\xi \right) \right) \leqslant \int_{\Omega_e} F(\boldsymbol{\nu}(\xi)) \mathrm{d}\xi. \tag{9.8}$$

Proof. Using $\tilde{\boldsymbol{\mu}}$ as a competitor in the approximate problem, given the monotonicity of Dir_ε, one has that

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda \int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \mathrm{d}\xi \leqslant \mathrm{Dir}(\tilde{\boldsymbol{\mu}}) + \lambda \int_{\Omega_e} F(\tilde{\boldsymbol{\mu}}(\xi)) \mathrm{d}\xi \leqslant C,$$

where the constant C is uniform in $\varepsilon > 0$ and $0 < \lambda \leqslant 1$. In particular, using the Rellich-like theorem (Proposition 9.5), we see that, up to extraction, $\boldsymbol{\mu}_{\varepsilon,\lambda}$ converges strongly in $L^2_e(\Omega_e, \mathcal{P}(D))$ to some $\bar{\boldsymbol{\mu}}_\lambda$ when $\varepsilon \to 0$. Moreover, by Γ-convergence of Dir_ε and the regularity of F,

$$\mathrm{Dir}(\bar{\boldsymbol{\mu}}_\lambda) + \lambda \int_{\Omega_e} F(\bar{\boldsymbol{\mu}}_\lambda(\xi)) \mathrm{d}\xi \leqslant \liminf_{\varepsilon \to 0} \left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda \int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \mathrm{d}\xi \right) \leqslant C. \tag{9.9}$$

Hence, we have a uniform bound on $\mathrm{Dir}(\bar{\boldsymbol{\mu}}_\lambda)$, and we can apply Rellich theorem (Proposition 9.4) to see that $\bar{\boldsymbol{\mu}}_\lambda$ converges strongly in $L^2(\Omega_e, \mathcal{P}(D))$ to some $\bar{\boldsymbol{\mu}} \in H^1_e(\Omega_e, \mathcal{P}(D))$ when $\lambda \to 0$. Moreover, using the lower semi-continuity of Dir and positivity of F,

$$\mathrm{Dir}(\bar{\boldsymbol{\mu}}) \leqslant \liminf_{\lambda \to 0} \left(\mathrm{Dir}(\bar{\boldsymbol{\mu}}_\lambda) + \lambda \int_{\Omega_e} F(\bar{\boldsymbol{\mu}}_\lambda(\xi)) \mathrm{d}\xi \right). \tag{9.10}$$

Let us assume by contradiction that $\bar{\boldsymbol{\mu}}$ is not a minimizer of Dir. Thanks to Proposition 9.11, it boils down to assume that $\mathrm{Dir}(\tilde{\boldsymbol{\mu}}) < \mathrm{Dir}(\bar{\boldsymbol{\mu}})$. In particular, as $F \circ \tilde{\boldsymbol{\mu}}$ is integrable on Ω_e and with the help of (9.10), it means that there exists λ small enough such that

$$\mathrm{Dir}(\tilde{\boldsymbol{\mu}}) + \lambda \int_{\Omega_e} F(\tilde{\boldsymbol{\mu}}(\xi)) \mathrm{d}\xi < \mathrm{Dir}(\bar{\boldsymbol{\mu}}_\lambda) + \lambda \int_{\Omega_e} F(\bar{\boldsymbol{\mu}}_\lambda(\xi)) \mathrm{d}\xi.$$

Using the fact that $\mathrm{Dir}_\varepsilon(\tilde{\boldsymbol{\mu}}) \to \mathrm{Dir}(\tilde{\boldsymbol{\mu}})$ to handle the l.h.s. and (9.9) to deal with the r.h.s., we see that for $\varepsilon > 0$ small enough,

$$\mathrm{Dir}_\varepsilon(\tilde{\boldsymbol{\mu}}) + \lambda \int_{\Omega_e} F(\tilde{\boldsymbol{\mu}}(\xi)) \mathrm{d}\xi < \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda \int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \mathrm{d}\xi,$$

which is a contradiction with the optimality of $\boldsymbol{\mu}_{\varepsilon,\lambda}$. Hence, $\bar{\boldsymbol{\mu}}$ is a minimizer of Dir over $H^1_e(\Omega_e, \mathcal{P}(D))$.

Remark that in (9.8) the first inequality is a consequence of the fact that F is regular. Assume by contradiction that there exists $\boldsymbol{\nu} \in H^1_e(\Omega_e, \mathcal{P}(D))$ a minimizer of Dir such that the second inequality of (9.8) does not hold. In particular as $\mathrm{Dir}(\bar{\boldsymbol{\mu}}) = \mathrm{Dir}(\boldsymbol{\nu})$, and by Γ-convergence of Dir_ε and lower semi-continuity of Dir,

$$\mathrm{Dir}(\boldsymbol{\nu}) = \mathrm{Dir}(\bar{\boldsymbol{\mu}}) \leqslant \liminf_{\lambda \to 0} \left(\liminf_{\varepsilon \to 0} \left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) \right) \right),$$

thus one can write that for some λ small enough,

$$\mathrm{Dir}(\boldsymbol{\nu}) + \lambda \int_{\Omega_e} F(\boldsymbol{\nu}(\xi)) \mathrm{d}\xi < \liminf_{\varepsilon \to 0} \left(\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_{\varepsilon,\lambda}) + \lambda \int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \mathrm{d}\xi \right):$$

it leads to the same contradiction as before by taking $\varepsilon > 0$ small enough. ☐

Now, the key result to get subharmonicity of $F \circ \bar{\boldsymbol{\mu}}$ is that we can pass at the pointwise limit the quantity $F \circ \boldsymbol{\mu}_{\varepsilon,\lambda}$.

Proposition 9.13. *For a.e. $\xi \in \Omega$, there holds*

$$F(\bar{\boldsymbol{\mu}}(\xi)) = \lim_{\lambda \to 0} \left(\lim_{\varepsilon \to 0} \left(F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \right) \right).$$

Proof. As the convergence of $\boldsymbol{\mu}_{\varepsilon,\lambda}$ to $\bar{\boldsymbol{\mu}}$ holds strongly in $L^2_e(\Omega, \mathcal{P}(D))$, we can, up to extraction, assume that it holds a.e. In other words, for a.e. $\xi \in \Omega$,

$$\bar{\boldsymbol{\mu}}(\xi) = \lim_{\lambda \to 0} \left(\lim_{\varepsilon \to 0} \left(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi) \right) \right)$$

in $\mathcal{P}(D)$. By lower semi-continuity of F on $\mathcal{P}(D)$, the inequality

$$F(\bar{\boldsymbol{\mu}}(\xi)) \leqslant \liminf_{\lambda \to 0} \left(\liminf_{\varepsilon \to 0} \left(F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \right) \right)$$

holds for a.e. $\xi \in \Omega$. On the other hand, use (9.8) with $\boldsymbol{\nu} = \bar{\boldsymbol{\mu}}$: up to extraction one has

$$\int_{\Omega_e} F(\bar{\boldsymbol{\mu}}(\xi)) \mathrm{d}\xi = \lim_{\lambda \to 0} \left(\lim_{\varepsilon \to 0} \left(\left(\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \mathrm{d}\xi \right) \right) \right).$$

By combining the two equations above (recall that all the functions $F \circ \boldsymbol{\mu}_{\varepsilon,\lambda}$ and $F \circ \bar{\boldsymbol{\mu}}$ are positive and bounded above by M thanks to Proposition 9.10), we reach the desired conclusion (this is just an adaptation of the proof of Scheffé's lemma). ☐

Proposition 9.14. *The function $F \circ \bar{\boldsymbol{\mu}}$ is subharmonic on $\mathring{\Omega}$. Moreover,*

$$\operatorname*{ess\,sup}_{\Omega} (F \circ \bar{\boldsymbol{\mu}}) \leqslant M.$$

Proof. The fact that the essential supremum of $F \circ \bar{\boldsymbol{\mu}}$ is bounded by M is a simple combination of Propositions 9.10 and 9.13. For the subharmonicity, take $\psi \in C^\infty_c(\mathring{\Omega})$ a smooth and positive function compactly supported in $\mathring{\Omega}$. For $0 < \varepsilon \leqslant \varepsilon_0$ small enough, one has, thanks to Proposition 9.9,

$$\int_{\Omega_e} \psi(\xi) \left(\frac{1}{\varepsilon^{d+2}} \int_{B(\xi,\varepsilon)} [F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\eta)) - F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))] \mathrm{d}\eta \right) \mathrm{d}\xi \geqslant 0.$$

Performing a discrete integration by parts (which is possible if ε is smaller than the distance between $\partial\Omega$ and the support of ψ), one sees that

$$\int_{\Omega_e} F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi)) \left(\frac{1}{\varepsilon^{d+2}} \int_{B(\xi,\varepsilon)} [\psi(\eta) - \psi(\xi)] \mathrm{d}\eta \right) \mathrm{d}\xi \geqslant 0.$$

Now send $\varepsilon \to 0$ and then $\lambda \to 0$. By smoothness of ψ, the quantity $\varepsilon^{-(d+2)} \int_{B(\xi,\varepsilon)} [\psi(\eta) - \psi(\xi)] \mathrm{d}\eta$ converges to $\Delta\psi(\xi)$ (up to a multiplicative constant). On the other hand, $F(\boldsymbol{\mu}_{\varepsilon,\lambda}(\xi))$ converges pointwisely to $F(\bar{\boldsymbol{\mu}})$ (see Proposition 9.13) while being bounded by M. By Lebesgue dominated convergence theorem,

$$\int_{\Omega_e} F(\boldsymbol{\mu}(\xi)) \Delta\psi(\xi) \mathrm{d}\xi \geqslant 0,$$

which exactly means that $F \circ \boldsymbol{\mu}$ is subharmonic in the sense of distributions as ψ is an arbitrary smooth and positive function. ☐

Now we can conclude:

Proof of Theorem 9.3 if F is regular. We take $\boldsymbol{\mu}$ the restriction of $\bar{\boldsymbol{\mu}}$ to Ω. Thanks to Proposition 9.6, the fact that $\bar{\boldsymbol{\mu}}$ is a minimizer of Dir among $H^1_e(\Omega_e, \mathcal{P}(D))$ is translated into the fact that $\boldsymbol{\mu}$ is a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$. The subharmonicity and the upper bound of $F \circ \bar{\boldsymbol{\mu}}$ are preserved by restriction. To get the minimality of $\int_\Omega (F \circ \bar{\boldsymbol{\mu}})$ among all other solutions, we just use (9.8). □

9.4 Simplifications in the continuous case

In this section, we assume that F is continuous. In particular, as $\mathcal{P}(D)$ is compact, it implies that F is bounded. The proof is simpler because we do not need to add the term $\lambda \int F \circ \boldsymbol{\mu}$ in the approximate problem. Indeed, strong convergence in $L^2(\Omega, \mathcal{P}(D))$ of a sequence $\boldsymbol{\mu}_n$ to $\boldsymbol{\mu}$ implies, up to extraction, the convergence a.e. of $(F \circ \boldsymbol{\mu}_n)$ to $(F \circ \boldsymbol{\mu})$.

We define $\Omega_e, \boldsymbol{\mu}_e$ and the functional spaces $L^2_e(\Omega_e, \mathcal{P}(D)), H^1_e(\Omega_e, \mathcal{P}(D))$ as in the beginning of Section 9.2.

Proof of Theorem 9.3 if F is continuous. For any $\varepsilon > 0$, we take $\boldsymbol{\mu}_\varepsilon \in L^2_e(\Omega_e, \mathcal{P}(D))$ a minimizer of Dir_ε over $L^2_e(\Omega_e, \mathcal{P}(D))$.

We can still apply Proposition 9.8 and conclude that for a.e. $\xi \in \Omega$, $\boldsymbol{\mu}_\varepsilon(\xi)$ is a barycenter of the $\boldsymbol{\mu}_\varepsilon(\eta)$ for $\eta \in B(\xi, \varepsilon)$. The proof of Jensen's inequality (Proposition 9.9) works in the same way as F is bounded on $\mathcal{P}(D)$. Hence, the maximum principle given by Proposition 9.10 is still true as it is only implied by Proposition 9.9.

To pass to the limit $\varepsilon \to 0$, we use the fact that (along an appropriate sequence) Dir_ε Γ-converges to Dir. Hence, up to extraction, $\boldsymbol{\mu}_\varepsilon$ converges to $\bar{\boldsymbol{\mu}}$ which is a minimizer of Dir over $L^2_e(\Omega_e, \mathcal{P}(D))$. Thanks to Proposition 9.5, the convergence takes place strongly in $L^2_e(\Omega_e, \mathcal{P}(D))$ and a.e. By continuity of F, we deduce that the conclusion of Proposition 9.13 still holds: $F \circ \boldsymbol{\mu}_\varepsilon$ converges a.e. to $F \circ \bar{\boldsymbol{\mu}}$ as $\varepsilon \to 0$. Thus the proof of Proposition 9.14 works exactly in the same way and it is enough to take for $\boldsymbol{\mu}$ the restriction of $\bar{\boldsymbol{\mu}}$ to Ω. □

Chapter 10

Special cases

In this chapter, we give examples of situations where more can be said about harmonic mappings. The first ones are rather simple: if the boundary conditions are valued in the set of Dirac masses, then so does the solution of the Dirichlet problem; and when D is a segment of $\mathbb{R}$ the space $\mathcal{P}(D)$ is isometric to a convex subset of a Hilbert space, hence the study is considerably simpler and all the machinery developed above is too heavy. The third one is trickier: we restrict ourselves to a family of elliptically contoured distributions, which is a geodesically convex subset of finite dimension. Thus we end up with mappings valued in a finite-dimensional Riemannian manifold, on which we can show existence, uniqueness, regularity and write explicit Euler-Lagrange equation.

10.1 Dirac masses

In this section, we say briefly what happens when the boundary data $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(D)$ is valued in the set of Dirac masses. We underline that all results in this section were proved by other people. We define

$$\mathcal{P}_{\mathrm{dc}}(D) := \{\delta_x \ : \ x \in D\} \subset \mathcal{P}(D)$$

the set of Dirac masses. The proof of the following result can be found in [Bre03, Theorem 3.1]

Proposition 10.1. *Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}_{\mathrm{dc}}(D)$ a Lipschitz mapping valued in the set of Dirac masses. Then there exists a unique solution to the Dirichlet problem with boundary conditions $\boldsymbol{\mu}_l$ and it is valued in $\mathcal{P}_{\mathrm{dc}}(D)$.*

Actually, if $\boldsymbol{\mu}_l(\xi) = \delta_{f_l(\xi)}$ for $\xi \in \partial\Omega$ then the solution of the Dirichlet problem is $\boldsymbol{\mu}(\xi) = \delta_{f(\xi)}$ where $f : \Omega \to D$ is the (classical) harmonic extension of f_l. The proof by Brenier relied on the exhibition of a solution to the dual problem in this particular case. Actually, there are at least two other arguments to reach the conclusion that at least one solution of the Dirichlet problem is valued in $\mathcal{P}_{\mathrm{dc}}(D)$.

- Denoting by $F : \mu \mapsto \iint_{D\times D} |x-y|^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y)$ the variance functional, and recalling that F is convex along generalized geodesic, we can apply Theorem 9.3 to say that (for at least one solution of the Dirichlet problem), the minimum of the variance is reached on $\partial\Omega$, where it is 0. But $\mathcal{P}_{\mathrm{dc}}(D)$ coincides with $F^{-1}(\{0\})$, hence the result.

- The mapping $\mu \mapsto \delta_{m(\mu)}$ with $m(\mu) := \int_D x\mu(\mathrm{d}x)$ is a contraction in the Wasserstein space, see (2.3). Moreover, it leaves $\mathcal{P}_{\mathrm{dc}}(D)$ invariant. Take a solution of the Dirichlet problem, *a priori* valued in $\mathcal{P}(D)$ and compose it with this mapping: thanks to Lemma 10.7 (see below), the boundary conditions are not changed and the Dirichlet energy decreases, hence we get a solution valued in $\mathcal{P}_{\mathrm{dc}}(D)$.

Now, let us do a short digression about curved geometries. As we already mention, we see no obstruction to extend our definition to the case where D is replaced by $(\mathcal{N}, \mathfrak{g})$ a Riemannian manifold. The striking point is that, if $\mathcal{N}$ is negatively curved, then these two arguments should still work. Indeed, if $\mathcal{N}$ has negative curvature and is simply connected, the variance functional is convex along geodesics, see [KP15]. However, to really make the argument working, we would need convexity along generalized geodesics as well as the validity of the Evolution Variational Inequality on Riemannian manifolds. On the other hand, still if $\mathcal{N}$ has negative curvature, the barycenter of a measure is uniquely defined and the mapping sending a measure on its barycenter is a contraction for the Wasserstein distance (see [Stu03, Theorem 6.3] for a proof for the 1-Wasserstein distance which can be easily adapted to W_2). Actually, in the case where $(\mathcal{N}, \mathfrak{g})$ is a simply connected manifold with negative curvature, a result similar to Proposition 10.1 has been proved in [Lu17, Theorem 3.3]. The proof by Lu relies on the second argument, i.e. the existence of a retraction onto $\mathcal{P}_{\mathrm{dc}}(\mathcal{N})$.

On the other hand, as understood by Lu, if $\mathcal{N}$ has positive curvature the result is no longer true. Indeed, he provided an example [Lu17, Example 3.6] of a domain Ω and some boundary conditions valued in $\mathcal{P}_{\mathrm{dc}}(\mathcal{N})$ (where $\mathcal{N}$ is the unit circle) such that any solution of the Dirichlet problem is *not* valued in $\mathcal{P}_{\mathrm{dc}}(\mathcal{N})$.

10.2 One dimensional target

In this section, we assume that $D = I = [0,1]$ is the unit interval. We underline that [Lu17, Theorem 2.2] provides a result similar to what follows in this section, we do not claim novelty here either. The important point is that the space $\mathcal{P}(I)$ has a very simple structure: the right object to characterize an element $\mu \in \mathcal{P}(D)$ is its inverse distribution function $F_\mu^{[-1]} : [0,1] \to [0,1]$ defined by

$$F_\mu^{[-1]}(t) := \inf\{x \in [0,1] \ : \ \mu([0,x]) \geqslant t\}.$$

It is well known that $F_\mu^{[-1]}$ is increasing, right continuous, and that there is a bijection between the set of increasing and right continuous mappings $[0,1] \to [0,1]$ and $\mathcal{P}(I)$. Moreover, for any $\mu, \nu \in \mathcal{P}(I)$, one has (see for instance [San15, Proposition 2.17])

$$W_2^2(\mu,\nu) = \int_0^1 \left| F_\mu^{[-1]}(t) - F_\nu^{[-1]}(t) \right|^2 \mathrm{d}t. \tag{10.1}$$

Introduce the Hilbert space $\mathcal{H} := L^2([0,1])$ with its usual norm (denoted by $|\cdot|_{\mathcal{H}}$) and the subspace $\mathcal{H}_{\mathrm{i}}$ of increasing functions: if $f \in \mathcal{H}$, then we say that $f \in \mathcal{H}_{\mathrm{i}}$ if $f(t) \in [0,1]$ for a.e. $t \in [0,1]$ and if for any $0 \leqslant t_1 < t_2 \leqslant t_3 < t_4 \leqslant 1$, one has

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f(t) \mathrm{d}t \leqslant \frac{1}{t_4 - t_3} \int_{t_3}^{t_4} f(t) \mathrm{d}t$$

Notice that $\mathcal{H}_{\mathrm{i}}$ is clearly a convex and closed subset of $\mathcal{H}$. Any $f \in \mathcal{H}_{\mathrm{i}}$ has a unique increasing and right continuous representative. Indeed, take the representative given by the Lebesgue differentiation theorem: except on a subset N which is negligible, it is increasing. Then, on N and on any point of discontinuity, choose the right limit. Uniqueness is easy as any increasing and right continuous representative is continuous except at a countable number of points. This discussion can be summarized in the following proposition.

Proposition 10.2. *If we define* $\Psi(\mu) := F_\mu^{[-1]}$*, then* Ψ *is a one-to-one isometry between* $\mathcal{P}(I)$ *and* $\mathcal{H}_i$*.*

Now we need to make the bridge between the Dirichlet energy in the space $H^1(\Omega, \mathcal{P}(I))$ and the one in $H^1(\Omega, \mathcal{H})$. In fact, it was already proved by Korevaar and Schoen [KS93] that their definition of Dirichlet energy coincides with the usual one if the target space is $\mathbb{R}$. By Pythagore's theorem, the equivalence still holds if the target space is a separable Hilbert space, as one can work on the coordinates in an orthogonal basis. As our definition of Dirichlet energy coincides with the one of Korevaar and Schoen, see Theorem 8.26, we can conclude that

$$\mathrm{Dir}(\boldsymbol{\mu}) := \int_\Omega |\nabla(\Psi \circ \boldsymbol{\mu})(\xi)|^2_{\mathcal{H}} \mathrm{d}\xi. \tag{10.2}$$

for any $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(I))$. Thus, we can say the following:

Theorem 10.3. *Let* $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(I)$ *a given Lipschitz mapping. Then there exists a unique* $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(I))$ *solution of the Dirichlet problem with boundary values* $\boldsymbol{\mu}_l$*. Moreover,* $\Psi \circ \boldsymbol{\mu}$ *is the solution of the minimization problem*

$$\min_f \left\{ \int_\Omega |\nabla f(\xi)|^2_{\mathcal{H}} \mathrm{d}\xi \ : \ f \in H^1(\Omega, \mathcal{H}) \ \text{ and } \ f|_{\partial\Omega} = \Psi \circ \boldsymbol{\mu}_l \right\}. \tag{10.3}$$

Proof. Everything relies on (10.2). With the help of Proposition 9.6, one can be convinced that imposing $\mathrm{BT}_{\boldsymbol{\mu}} = \mathrm{BT}_{\boldsymbol{\mu}_b}$ is the same as saying that the the trace of $(\Psi \circ \boldsymbol{\mu})$ is $(\Psi \circ \boldsymbol{\mu}_l)$. Then, one takes f to be the unique harmonic extension of $(\Psi \circ \boldsymbol{\mu}_l)$ in $H^1(\Omega, \mathcal{H})$: it is the minimizer of (10.3). By the maximum principle, as $(\Psi \circ \boldsymbol{\mu}_l) \in \mathcal{H}_\mathrm{i}$ on $\partial\Omega$, it is clear that $f \in H^1(\Omega, \mathcal{H}_\mathrm{i})$. Thus, we can simply set $\boldsymbol{\mu} := \Psi^{-1} \circ f$. □

10.3 Family of elliptically contoured distributions

We study the case where the boundary values belong to a family of elliptically contoured distributions: they are parametrized by their covariance matrix. It can be seen as a generalization of the case where the measures are Gaussian. In this section, we would like to show that at least one solution of the Dirichlet problem is valued in the family of elliptically contoured distributions if it is the case for the boundary values, and to give a full solution (existence, uniqueness, regularity and Euler-Lagrange equation) under the additional assumption that the covariance matrices of the boundary values are non singular.

We will deal with centered measures (i.e. measures with zero mean) because the contribution of the mean to the Dirichlet energy can be handled independently. More precisely if $\mu \in \mathcal{P}(D)$ we denote by $m(\mu) := \int_D x\mu(\mathrm{d}x) \in D$ its mean and μ_0 the centered measured defined as the push forward of μ by $(x \mapsto x - m(\mu))$. As recalled in Section 2.1, if $\mu, \nu \in \mathcal{P}(D)$ then

$$W_2^2(\mu, \nu) = W_2^2(\mu_0, \nu_0) + |m(\mu) - m(\nu)|^2.$$

If $\boldsymbol{\mu} \in L^2(\Omega, \mathcal{P}(D))$, we use the formula above on $\mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$:

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) = \mathrm{Dir}_\varepsilon(\boldsymbol{\mu}_0) + C_p \iint_{\Omega\times\Omega} \frac{|m(\boldsymbol{\mu}(\xi)) - m(\boldsymbol{\mu}(\eta))|^2}{2\varepsilon^{p+2}} \mathbb{1}_{|\xi-\eta|\leqslant\varepsilon} \mathrm{d}\xi \mathrm{d}\eta.$$

Then, sending ε to 0 and using [Jos08, Theorem 8.3.1] to handle the part involving the Dirichlet energy of the means, one sees that

$$\mathrm{Dir}(\boldsymbol{\mu}) = \mathrm{Dir}(\boldsymbol{\mu}_0) + \frac{1}{2} \int_\Omega |\nabla[m(\boldsymbol{\mu})](\xi)|^2 \mathrm{d}\xi.$$

The term involving $m(\boldsymbol{\mu})$ is easy to minimize (because $m(\boldsymbol{\mu})$ is a vector-valued function, it boils down to take the harmonic extension) and it can be done independently from the term involving $\mathrm{Dir}(\boldsymbol{\mu}_0)$. In other words, it is not restrictive to work only with centered measures.

Let us go define what is a family of elliptically contoured distributions. As we have assumed that D is compact, we cannot work with non compactly supported measures, in particular with Gaussian measures. For the rest of the section, we fix $\rho \in L^1(\mathbb{R}^q)$ a positive function compactly supported such that $\rho\mathcal{L}_D$ is a probability measure with zero mean and the identity matrix as a covariance matrix. Recall that the covariance matrix $\mathrm{cov}(\mu)$ of a centered measure $\mu \in \mathcal{P}(\mathbb{R}^q)$ with finite second moments is defined as: for any $i, j \in \{1, 2, \dots, q\}$,

$$\mathrm{cov}(\mu)_{ij} := \int_{\mathbb{R}^q} x_i x_j \mu(\mathrm{d}x);$$

and that the covariance matrix of an non-centered measure μ is defined as the covariance matrix of its centered part. For technical reasons, we also assume that ρ is radial and that the Boltzmann entropy of $\rho\mathcal{L}_D$ (see (10.8) below) is finite. Let us denote by $S_q(\mathbb{R})$ the set of symmetric $q \times q$ matrices and $S_q^+(\mathbb{R}) \subset S_q(\mathbb{R})$ the set of symmetric and semi-definite positive $q \times q$ matrices. The space $S_q(\mathbb{R})$ is equiped with its canonical scalar product $\langle \cdot, \cdot \rangle$ defined by $\langle A, B \rangle = \mathrm{Tr}(AB)$. The unique symmetric square root of a matrix $A \in S_q^+(\mathbb{R})$ is denoted by $A^{1/2}$. Instead of parametrizing measures by their covariance matrix we will do it by the square root of their covariance matrix, i.e. by their standard deviation: it is more natural for homogeneity reasons and the formulas are slightly simpler.

Definition 10.4. *For any $A \in S_q^+(\mathbb{R})$ we denote by $\rho_A \in \mathcal{P}(\mathbb{R}^q)$ the push-forward of $\rho\mathcal{L}_D$ by the map $x \in \mathbb{R}^q \mapsto Ax \in \mathbb{R}^q$.*

The set of all ρ_A for $A \in S_q^+(\mathbb{R})$ is denoted by $\mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$ and is called a family of elliptically contoured distributions (with reference measure $\rho\mathcal{L}_D$).

Thanks to the normalization of ρ, the measure ρ_A has zero mean and covariance matrix A^2. Notice that if A is invertible then

$$\rho_A(\mathrm{d}x) := \frac{1}{\det(A)} \rho \left(A^{-1} x \right) \mathrm{d}x.$$

We would recover the Gaussian case by taking $\rho(x) = (2\pi)^{-q/2} \exp(-|x|^2/2)$, but this function is not compactly supported.

The crucial tool to establish that an harmonic extension of a mapping valued in a family of elliptically contoured distributions stays in the same family is the existence of a retraction on the set $\mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$. Let us call $\mathcal{P}_2(\mathbb{R}^q)$ the set of probability measures on $\mathbb{R}^q$ with finite second moment.

Definition 10.5. *Let $R : \mathcal{P}_2(\mathbb{R}^q) \to \mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$ the application defined by $R(\mu) := \rho_A$, where $A := \mathrm{cov}(\mu)^{1/2}$ is the symmetric square root of the covariance matrix of μ.*

Proposition 10.6. *The application $R : \mathcal{P}_2(\mathbb{R}^q) \to \mathcal{P}_{\mathrm{ec}}(D)$ leaves $\mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$ unchanged and is a contraction (i.e. is 1-Lipschitz) provided that $\mathcal{P}_2(\mathbb{R}^q)$ and $\mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$ are endowed with the quadratic Wasserstein distance W_2.*

Proof. The first part is obvious by the way we normalize ρ. The second part is a reformulation of Theorem 2.1 and Theorem 2.4 of [Gel90]. Nevertheless, for the convenience of the reader we provide a simpler argument.

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^q)$, without loss of generality we can assume that they are centered. Let (φ, ψ) be a pair of Kantorovich potential between $R(\mu)$ and $R(\nu)$. It is well known that φ and ψ are quadratic functions (see for instance [PC17, Remark 2.29]), hence

$$\frac{W_2^2(\mu,\nu)}{2} \geqslant \int_{\mathbb{R}^q} \varphi \mathrm{d}\mu + \int_{\mathbb{R}^q} \psi \mathrm{d}\nu = \int_{\mathbb{R}^q} \varphi \mathrm{d}R(\mu) + \int_{\mathbb{R}^q} \psi \mathrm{d}R(\nu) = \frac{W_2^2(R(\mu), R(\nu))}{2}.$$

Indeed, the first inequality is Kantorovich's duality and the first equality comes from the fact that the integral of a quadratic function against a centered probability measure depends only on the covariance matrix of the probability measure. □

Let us prove state and prove here an easy technical lemma which will be crucial in the sequel.

Lemma 10.7. *Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(D)$ a Lipschitz function and $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ such that $\boldsymbol{\mu}|_{\partial\Omega} = \boldsymbol{\mu}_l$. Let $T : \mathcal{P}(D) \to \mathcal{P}(D)$ a 1-Lipschitz mapping. Then $T \circ \boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ with $(T \circ \boldsymbol{\mu})|_{\partial\Omega} = (T \circ \boldsymbol{\mu}_l)$ and*

$$\mathrm{Dir}(T \circ \boldsymbol{\mu}) \leqslant \mathrm{Dir}(\boldsymbol{\mu}).$$

Proof. As T is a contraction and from the definition of Dir_ε it is obvious that

$$\mathrm{Dir}_\varepsilon(T \circ \boldsymbol{\mu}) \leqslant \mathrm{Dir}_\varepsilon(\boldsymbol{\mu})$$

holds for any $\varepsilon > 0$. Then it is sufficient to send ε to 0. To get the assertion involving the boundary conditions, one can use for instance Proposition 9.6. □

As we work in the compactly supported case, we add some assumption that D is large enough in order for the boundary of D to be invisible. More precisely, the following lemma will help us to handle the finiteness of D.

Lemma 10.8. *Let $\tilde{D} \subset D$ be a convex compact subset of D. Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}(\tilde{D})$ be a Lipschitz mapping. If $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(\tilde{D}))$ is a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$, then, seen as an element of $H^1(\Omega, \mathcal{P}(D))$ (extending $\boldsymbol{\mu}$ by 0 on $D \backslash \tilde{D}$), $\boldsymbol{\mu}$ is also a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ (with $\boldsymbol{\mu}_l$ seen as a mapping valued in $\mathcal{P}(D)$).*

Proof. It relies on a simple observation. Let $P_{\tilde{D}} : D \to \tilde{D}$ be the Euclidean projection on $\tilde{D}$. One has that $\nu \mapsto P_{\tilde{D}} \# \nu$ is a 1-Lipschitz function from $(\mathcal{P}(D), W_2)$ to $(\mathcal{P}(\tilde{D}), W_2)$ which leaves the boundary values $\boldsymbol{\mu}_l$ unchanged. Thus we can apply Lemma 10.7 to see that $P_{\tilde{D}}$ maps any competitor from $H^1(\Omega, \mathcal{P}(D))$ into a competitor in $H^1(\Omega, \mathcal{P}(\tilde{D}))$. □

We will say that $\tilde{D} \subset D$ is compatible with ρ if it is a compact convex subset of D and for any $\mu \in \mathcal{P}(\tilde{D})$, one has $R(\mu) \in \mathcal{P}(D)$. It holds if D is large enough compared to $\tilde{D}$ and the diameter of the support of ρ. In the sequel, we will use the notations $\mathcal{P}_{\mathrm{ec}}(\tilde{D}) := \mathcal{P}(\tilde{D}) \cap \mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$ and $\mathcal{P}_{\mathrm{ec}}(D) := \mathcal{P}(D) \cap \mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$. The first main result of this section is the following.

Theorem 10.9. *Take $\tilde{D} \subset D$ compatible with ρ. Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}_{\mathrm{ec}}(\tilde{D})$ a Lipschitz mapping valued in the family of elliptically contoured distributions. Then there exists $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ such that $\boldsymbol{\mu}(\xi) \in \mathcal{P}_{\mathrm{ec}}(D)$ for a.e. $\xi \in \Omega$.*

The assumption that $\tilde{D}$ is compatible with D can be translated in the fact that the supports of the $\boldsymbol{\mu}_l(\xi)$ for $\xi \in \partial\Omega$ are small compared to D.

Proof. Let $\tilde{\boldsymbol{\mu}}$ be a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$, it exists thanks to Theorem 8.33 and Theorem 8.32. According to Lemma 10.8, we can choose $\tilde{\boldsymbol{\mu}}$ such that $\tilde{\boldsymbol{\mu}} \in \mathcal{P}(\tilde{D})$ a.e. As R is a contraction which leaves the boundary values unchanged, it is clear thanks to Lemma 10.7 that $\boldsymbol{\mu} := R \circ \tilde{\boldsymbol{\mu}}$ is a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$. By construction, $\boldsymbol{\mu}$ is valued in $\mathcal{P}_{\text{ec}}(\mathbb{R}^q)$ and also in $\mathcal{P}(D)$ as $\tilde{D}$ is compatible with ρ. □

We believe that, conducting a careful analysis, one can prove that all solutions of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ are valued in $\mathcal{P}_{\text{ec}}(D)$.

Now, we want to go further and give a more explicit description of the solution valued in the family of elliptically contoured distributions. To this extent, we rely on the fact that the manifold $S_q^+(\mathbb{R})$, when endowed with the distance induced by W_2 through the application $A \mapsto \rho_A$, has a structure of Riemannian manifold, at least when restricted to the set of non singular matrix. The computation of Wasserstein distance between gaussians distributions has been discovered independently many times (see for instance [DL82, Gel90]), while the resulting geometry was first investigated by Takatsu [Tak11] and has recently regained some interest [MMP18, BJL18]. The restriction of the Wasserstein distance to the set of gaussian measures is sometimes called the Bures metric.

More precisely, if A and B are in $S_q^+(\mathbb{R})$ it is known (see for instance [Gel90]) that (up to a global multiplicative constant that depends only on ρ)

$$W_2^2(\rho_A, \rho_B) = \operatorname{Tr}\left(A^2 + B^2 - 2(AB^2A)^{1/2}\right).$$

Notice that if A and B commute then $W_2^2(\rho_A, \rho_B) = \operatorname{Tr}((A-B)^2)$ is the squared Euclidean distance between A and B, which justifies that the right choice is to parametrize elements of the family of elliptically contoured distributions by the square root of their covariance matrix. If $A \in S_q^+(\mathbb{R})$, we can define the linear map $L_A : S_q(\mathbb{R}) \to S_q(\mathbb{R})$ by $L_A := A \otimes \operatorname{Id} + \operatorname{Id} \otimes A$. More explicitly for any $B \in S_q(\mathbb{R})$

$$L_A(B) = AB + BA.$$

The map L_A is symmetric, and is moreover positive definite as soon as A is positive definite (in this case in particular it is invertible). If A is diagonal, then L_A is also diagonal in the canonical basis for matrices. In particular, if A and B commute, then L_A and L_B also commute. Denote by $S_q^{++}(\mathbb{R})$ the set of $q \times q$ symmetric definite positive matrices. If $A \in S_q^{++}(\mathbb{R})$ and $B \in S_q(\mathbb{R})$, a lengthy but straightforward computation leads to

$$\lim_{t \to 0} \frac{W_2^2(\rho_A, \rho_{A+tB})}{t^2} = \langle B, \mathfrak{g}_A(B) \rangle \tag{10.4}$$

where $\mathfrak{g}_A : S_q(\mathbb{R}) \to S_q(\mathbb{R})$ is a linear map defined as

$$\mathfrak{g}_A := \frac{1}{2}(L_A)^2 (L_{A^2})^{-1}.$$

More explicitly, if A is a diagonal matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_q$ and $B = (B_{ij})_{1 \leqslant i,j \leqslant q}$ then

$$\langle B, \mathfrak{g}_A(B) \rangle = \frac{1}{2} \sum_{1 \leqslant i,j \leqslant q} \frac{(\lambda_i + \lambda_j)^2}{\lambda_i^2 + \lambda_j^2} B_{ij}^2. \tag{10.5}$$

Notice that $\mathfrak{g}_A$ always defines a scalar product on the space $S_q(\mathbb{R})$. As a consequence, we can define the Riemannian manifold $(S_q^{++}(\mathbb{R}), \mathfrak{g})$: at each point $A \in S_q^{++}(\mathbb{R})$ the tangent space, which is isomorphic to $S_q(\mathbb{R})$, is endowed with the scalar product $\mathfrak{g}_A$. If we do that, as we already know that $\mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$ is a geodesic space and thanks to (10.4), we see that the Riemannian distance $d_{\mathfrak{g}}$ induced by $\mathfrak{g}$ satisfies $d_{\mathfrak{g}}(A,B) = W_2(\rho_A, \rho_B)$ for any $A, B \in S_q^{++}(\mathbb{R})$. From this identity we can derive the following consequence. Take $\mathbf{A} \in H^1(\Omega, (S_q^{++}(\mathbb{R}), \mathfrak{g}))$ a matrix-valued function and define $\rho_{\mathbf{A}} \in L^2(\Omega, \mathcal{P}(D))$ by $\rho_{\mathbf{A}}(\xi) = \rho_{\mathbf{A}(\xi)}$ for a.e. $\xi \in \Omega$. Then $\rho_{\mathbf{A}} \in H^1(\Omega, \mathcal{P}(D))$ and

$$\mathrm{Dir}(\rho_{\mathbf{A}}) = \int_\Omega \frac{1}{2} \sum_{\alpha=1}^p \langle \partial_\alpha \mathbf{A}(\xi), \mathfrak{g}_{\mathbf{A}(\xi)}(\partial_\alpha \mathbf{A}(\xi)) \rangle \mathrm{d}\xi. \tag{10.6}$$

To justify this identity, one can use for instance the formulation with Dir_ε (Theorem 8.26), replace the Wasserstein distance W_2 by the Riemannian distance $d_{\mathfrak{g}}$, and use the already known equivalence between Dir and the limit of Dir_ε when $\varepsilon \to 0$ for mappings valued in a Riemannian manifold [Jos08, Theorem 8.3.1].

Notice that the metric tensor $\mathfrak{g}_A$ diverges as A becomes singular. Thus, it is natural to assume that the boundary values have non singular covariance matrices. With this assumption we are able to provide a full solution of the Dirichlet problem, which is the second main result of this section.

Theorem 10.10. *Take $\tilde{D} \subset D$ compatible with ρ. Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}_{\mathrm{ec}}(\tilde{D})$ a Lipschitz mapping such that* $\det(\mathrm{cov}(\boldsymbol{\mu}_l(\xi))) > 0$ *for all $\xi \in \partial\Omega$ and define $\mathbf{A}_l(\xi) = \mathrm{cov}(\boldsymbol{\mu}_l(\xi))^{1/2}$ for all $\xi \in \partial\Omega$.*

Then there exists a unique solution $\bar{\boldsymbol{\mu}} \in H^1(\Omega, \mathcal{P}(D))$ of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ and $\bar{\boldsymbol{\mu}}(\xi) \in \mathcal{P}_{\mathrm{ec}}(D)$ for a.e. $\xi \in \Omega$. Moreover, if $\bar{\mathbf{A}} \in H^1(\Omega, (S_q^{++}(\mathbb{R}), \mathfrak{g}))$ is defined by $\bar{\mathbf{A}}(\xi) := \mathrm{cov}(\bar{\boldsymbol{\mu}}(\xi))^{1/2}$ for a.e. $\xi \in \Omega$, then the following holds:

(i) $\operatorname*{ess\,inf}_{\xi \in \Omega} \det(\bar{\mathbf{A}}(\xi)) > 0$*;*

(ii) $\bar{\mathbf{A}}$ *is a minimizer of*

$$\int_\Omega \frac{1}{2} \sum_{\alpha=1}^p \langle \partial_\alpha \mathbf{B}(\xi), \mathfrak{g}_{\mathbf{B}(\xi)}(\partial_\alpha \mathbf{B}(\xi)) \rangle \mathrm{d}\xi.$$

among all $\mathbf{B} \in H^1(\Omega, (S_q^{++}(\mathbb{R}), \mathfrak{g}))$ *which have boundary values* $\mathbf{A}_l$*;*

(iii) $\bar{\mathbf{A}}$ *is a weak solution of*

$$\sum_{\alpha=1}^p \partial_\alpha \left(L_{\bar{\mathbf{A}}} L^{-1}_{\bar{\mathbf{A}}^2}(\partial_\alpha \bar{\mathbf{A}}) \right) + \sum_{\alpha=1}^p \left(L_{\bar{\mathbf{A}}} L^{-1}_{\bar{\mathbf{A}}^2}(\partial_\alpha \bar{\mathbf{A}}) \right)^2 = 0. \tag{10.7}$$

(iv) The mapping $\bar{\mathbf{A}}$ is smooth (namely C^∞) in the interior of Ω, and regularity up to the boundary holds provided $\mathbf{A}_l$ and $\partial\Omega$ are smooth enough.

Notice that we are able to prove uniqueness among *all* mappings valued in the Wasserstein space and not only those valued in the family of elliptically contoured distributions: it is one of the only case where we can prove that uniqueness holds for the Dirichlet problem. Remark also that (10.7) is nothing else than the Euler-Lagrange equation associated to the problem of calculus of variations (ii). The last point is the application of the standard theory of elliptic regularity for harmonic mappings valued in Riemannian manifolds, in particular we refer the reader to [SU83] for the precise assumptions required for the regularity to hold up to the boundary. The only thing we will need to show is the absence of non constant *minimizing tangent maps*, which we will prove thanks to an argument based on the maximum principle.

The rest of this section is dedicated to the proof of Theorem 10.10 which is obtained by putting together Propositions 10.11, 10.12,10.15 and 10.17. More precisely, the first step is to show the existence of one solution $\bar{\boldsymbol{\mu}}$ of the Dirichlet problem taking values in the family of elliptically contoured distributions for which the covariance matrices stay non singular inside Ω (Proposition 10.11). Then, using the explicit expression (10.6), it is fairly easy to show that (ii) and (iii) are satisfied (Proposition 10.12). The hardest part is the question of uniqueness. As explained in Chapter 7, we will first show that any solution $\boldsymbol{\mu}$ of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ must have $\bar{\mathbf{v}}$ as tangent velocity field, where $\bar{\mathbf{v}}$ is the tangent velocity field of $\bar{\boldsymbol{\mu}}$. Then, as $\bar{\mathbf{v}}$ will happen to be smooth enough (linear, hence Lipschitz w.r.t. variables in D), we will use the results about uniqueness of the (1-dimensional) continuity equation for smooth velocity field (Proposition 10.15). For the last point of the theorem, as $\bar{\mathbf{A}}$ is a Dirichlet minimizing mapping valued in a compact subset of the Riemannian manifold $(S_q^{++}(\mathbb{R}), \mathfrak{g})$ (thanks to point (i)), we can apply the classical theory: see [SU82, Theorem IV] for the interior regularity and [SU83] for the boundary regularity. The only point to show is the absence of non constant minimizing tangent maps, which a consequence of Proposition 10.17 proved below.

Let us begin by showing that for at least one solution of the Dirichlet problem the covariance matrices stay non singular inside Ω. As a tool to measure regularity of elliptically contoured distributions, we will use the Boltzmann entropy, see (2.13). More precisely, we define $H : \mathcal{P}(D) \to \overline{\mathbb{R}}$ by

$$H(\mu) := \begin{cases} \displaystyle\int_D \mu(x) \ln(\mu(x)) \mathrm{d}x & \text{if } \mu \text{ is absolutely continuous w.r.t. } \mathcal{L}_D, \\ +\infty & \text{else.} \end{cases} \tag{10.8}$$

It is known that H is convex along generalized geodesics [AGS08, Theorem 9.4.10] and it is regular according to Proposition 9.2. Moreover, an explicit computation leads to $H(\rho_A) = -\ln(\det A) + H(\rho\mathcal{L}_D)$ (with the convention $\ln(0) = -\infty$). Also, using the fact that Gaussian measures are the ones which minimize H for a given covariance matrix, we get that for any $\mu \in \mathcal{P}(D)$,

$$H(\mu) \geqslant -\frac{1}{2} \ln\left(\det\left(\mathrm{cov}(\mu)\right)\right) + C, \tag{10.9}$$

where the constant C is the entropy of a standard normal distribution.

Proposition 10.11. *Take $\tilde{D} \subset D$ compatible with ρ. Let $\boldsymbol{\mu}_l : \partial\Omega \to \mathcal{P}_{\mathrm{ec}}(\tilde{D})$ a Lipschitz mapping such that $\det\left(\mathrm{cov}(\boldsymbol{\mu}_l(\xi))\right) > 0$ for all $\xi \in \partial\Omega$. Then there exists $\bar{\boldsymbol{\mu}} \in H^1(\Omega, \mathcal{P}(D))$ a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ such that $\bar{\boldsymbol{\mu}}(\xi) \in \mathcal{P}_{\mathrm{ec}}(D)$ for a.e. $\xi \in \Omega$ and such that*

$$\operatorname*{ess\,inf}_{\xi\in\Omega} \left[\det\left(\mathrm{cov}(\bar{\boldsymbol{\mu}}(\xi))\right)\right] > 0.$$

Proof. Notice, thanks to the explicit formula for H on $\mathcal{P}_{\mathrm{ec}}(\mathbb{R}^q)$ and as $\boldsymbol{\mu}_l$ is continuous, that $\sup_{\partial\Omega}(H \circ \boldsymbol{\mu}_l) < +\infty$. Take $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(\tilde{D}))$ the solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ given by Theorem 9.3 (with $F = H$). Set $\bar{\boldsymbol{\mu}} := R \circ \boldsymbol{\mu}$. By the same argument as in Theorem 10.9, $\bar{\boldsymbol{\mu}} \in H^1(\Omega, \mathcal{P}_{\mathrm{ec}}(D))$ is a solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$. Using first the estimate (10.9) and then the maximum principle (9.1),

$$\begin{aligned} \operatorname*{ess\,sup}_{\xi\in\Omega} \left[-\ln\left(\det\left(\mathrm{cov}(\bar{\boldsymbol{\mu}}(\xi))\right)\right)\right] &= \operatorname*{ess\,sup}_{\xi\in\Omega} \left[-\ln\left(\det\left(\mathrm{cov}(\boldsymbol{\mu}(\xi))\right)\right)\right] \\ &\leqslant -2C + 2 \operatorname*{ess\,sup}_{\xi\in\Omega} H(\boldsymbol{\mu}(\xi)) \\ &\leqslant -2C + 2 \sup_{\xi\in\partial\Omega} H(\boldsymbol{\mu}_l(\xi)) < +\infty. \end{aligned}$$

□

Until the end of the section, $\bar{\boldsymbol{\mu}} \in H^1(\Omega, \mathcal{P}_{\mathrm{ec}}(D))$ will denote the object defined in Proposition 10.11 and for a.e. $\xi \in \Omega$, one defines $\bar{\mathbf{A}}(\xi) = \mathrm{cov}(\bar{\boldsymbol{\mu}}(\xi))^{1/2}$. Notice that point (i) of Theorem 10.10 is proved. Now let us derive the equation satisfied by $\bar{\mathbf{A}}$.

Proposition 10.12. *The mapping $\bar{\mathbf{A}} \in H^1(\Omega, (S_q^{++}(\mathbb{R}), \mathfrak{g}))$ is a weakly harmonic map, more precisely a minimizer of*

$$\mathbf{B} \in H^1(\Omega, (S_q^{++}(\mathbb{R}), \mathfrak{g})) \mapsto \int_\Omega \frac{1}{2} \sum_{\alpha=1}^{p} \langle \partial_\alpha \mathbf{B}(\xi), \mathfrak{g}_{\mathbf{B}(\xi)}(\partial_\alpha \mathbf{B}(\xi)) \rangle \mathrm{d}\xi.$$

among all $\mathbf{B}$ which have boundary values $\mathbf{A}_l$. In particular, $\bar{\mathbf{A}}$ satisfies the Euler-Lagrange equation (10.7)*.*

Proof. We need to prove that, for any $\mathbf{B} \in H^1(\Omega, (S_q^{++}(\mathbb{R}), \mathfrak{g}))$ with boundary values $\mathbf{A}_l$ one has

$$\int_\Omega \frac{1}{2} \sum_{\alpha=1}^{p} \langle \partial_\alpha \mathbf{B}(\xi), \mathfrak{g}_{\mathbf{B}(\xi)}(\partial_\alpha \mathbf{B}(\xi)) \rangle \mathrm{d}\xi \geqslant \int_\Omega \frac{1}{2} \sum_{\alpha=1}^{p} \langle \partial_\alpha \bar{\mathbf{A}}(\xi), \mathfrak{g}_{\bar{\mathbf{A}}(\xi)}(\partial_\alpha \bar{\mathbf{A}}(\xi)) \rangle \mathrm{d}\xi = \mathrm{Dir}(\rho_{\bar{\mathbf{A}}}) = \mathrm{Dir}(\bar{\boldsymbol{\mu}}).$$

To prove it, if we take any $\mathbf{B} \in H^1(\Omega, (S_q^{++}(\mathbb{R}), \mathfrak{g}))$ we can build $\boldsymbol{\mu} := \rho_{\mathbf{B}}$ and we have, thanks to (10.6), the identity

$$\mathrm{Dir}(\boldsymbol{\mu}) = \int_\Omega \frac{1}{2} \sum_{\alpha=1}^{p} \langle \partial_\alpha \mathbf{B}(\xi), \mathfrak{g}_{\mathbf{B}(\xi)}(\partial_\alpha \mathbf{B}(\xi)) \rangle \mathrm{d}\xi.$$

A priori, $\boldsymbol{\mu}$ is valued in $\mathcal{P}(\mathbb{R}^q)$. If we denote by $P_D : \mathbb{R}^q \to D$ the Euclidean projection on D, then

$$\mathrm{Dir}(\bar{\boldsymbol{\mu}}) \leqslant \mathrm{Dir}(P_D \# \boldsymbol{\mu}) \leqslant \mathrm{Dir}(\boldsymbol{\mu}),$$

where the first inequality comes from the optimality of $\bar{\boldsymbol{\mu}}$ (notice that $P_D\#\cdot$ leaves the boundary values unchanged) and the second one from the fact that $P_D\#\cdot$ is a contraction (Lemma 10.7).

To get the Euler-Lagrange equation it is actually easier if we take the covariance matrix and not its square root as the variable. In other words we define $\bar{\mathbf{C}} := \bar{\mathbf{A}}^2$. As $\bar{\mathbf{A}}$ is never singular, this change of variables is smooth. We have $\partial_\alpha \bar{\mathbf{C}} = L_{\bar{\mathbf{A}}}(\partial_\alpha \bar{\mathbf{A}})$ and in particular

$$\langle \partial_\alpha \bar{\mathbf{A}}, \mathfrak{g}_{\bar{\mathbf{A}}}(\partial_\alpha \bar{\mathbf{A}}) \rangle = \langle \partial_\alpha \bar{\mathbf{C}}, L_{\bar{\mathbf{C}}}^{-1}(\partial_\alpha \bar{\mathbf{C}}) \rangle.$$

If we take $\mathbf{D} : \Omega \to S_q(\mathbb{R})$ smooth and compactly supported on Ω and that we consider $\mathbf{B} := \bar{\mathbf{C}} + t\mathbf{D}$ as a competitor for small t, we reach the conclusion that

$$\sum_{\alpha=1}^{p} \langle \partial_\alpha \mathbf{D}, L_{\bar{\mathbf{C}}}^{-1}(\partial_\alpha \bar{\mathbf{C}}) \rangle + \frac{1}{2} \sum_{\alpha=1}^{p} \left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_{t=0} \langle \partial_\alpha \bar{\mathbf{C}}, L_{\bar{\mathbf{C}}+t\mathbf{D}}^{-1}(\partial_\alpha \bar{\mathbf{C}}) \rangle = 0.$$

A simple computation leads to

$$L_{\bar{\mathbf{C}}+t\mathbf{D}}^{-1}(\partial_\alpha \bar{\mathbf{C}}) = L_{\bar{\mathbf{C}}}^{-1}(\partial_\alpha \bar{\mathbf{C}}) - t L_{\bar{\mathbf{C}}}^{-1} \left[\mathbf{D}(L_{\bar{\mathbf{C}}}^{-1}(\partial_\alpha \bar{\mathbf{C}})) + (L_{\bar{\mathbf{C}}}^{-1}(\partial_\alpha \bar{\mathbf{C}}))\mathbf{D} \right] + o(t^2).$$

Using the properties of the Trace and the symmetry of $L_{\bar{\mathbf{C}}}^{-1}$, we conclude that the Euler-Lagrange equation reads

$$\sum_{\alpha=1}^{p} \langle \partial_\alpha \mathbf{D}, L_{\bar{\mathbf{C}}}^{-1}(\partial_\alpha \bar{\mathbf{C}}) \rangle - \sum_{\alpha=1}^{p} \langle \mathbf{D}, (L_{\bar{\mathbf{C}}}^{-1}(\partial_\alpha \bar{\mathbf{C}}))^2 \rangle = 0.$$

Coming back to $\bar{\mathbf{C}} = \bar{\mathbf{A}}^2$ and $\partial_\alpha \bar{\mathbf{C}} = L_{\bar{\mathbf{A}}}(\partial_\alpha \bar{\mathbf{A}})$, as $\mathbf{D}$ is arbitrary we see that we get the weak formulation of (10.7). □

As far as the regularity issues are concerned, notice that $\bar{\mathbf{A}}$ is uniformly bounded from below as a symmetric matrix (this is (i) of Theorem 10.10) and also bounded from above as a symmetric matrix (as $\rho_{\bar{\mathbf{A}}} \in \mathcal{P}(D)$ and D is compact), hence the operators $L_{\bar{\mathbf{A}}(\xi)} : S_q(\mathbb{R}) \to S_q(\mathbb{R})$ are bounded with a bounded inverse uniformly in $\xi \in \Omega$. In other words, the metric tensor $\mathfrak{g}_{\bar{\mathbf{A}}(\xi)}$ is equivalent to the canonical scalar product uniformly in $\xi \in \Omega$. In particular, the regularity $\bar{\boldsymbol{\mu}} \in H^1(\Omega, \mathcal{P}(D))$ translates in $\bar{\mathbf{A}} \in H^1(\Omega, S_q(\mathbb{R}))$ where $S_q(\mathbb{R})$ is endowed with its usual Euclidean norm $|\cdot|$.

We need to prove uniqueness. The first step is to identify the tangent velocity field to $\bar{\boldsymbol{\mu}}$ and a (at least formal) solution of the dual problem.

Proposition 10.13. *For any $\alpha \in \{1, 2, \dots, p\}$ we define $\bar{\mathbf{B}}^\alpha := L_{\bar{\mathbf{A}}} L^{-1}_{\bar{\mathbf{A}}^2}(\partial_\alpha \bar{\mathbf{A}}) \in L^2(\Omega, S_q(\mathbb{R}))$ and we set*

$$\bar{\mathbf{v}}^\alpha(\xi, x) := \bar{\mathbf{B}}^\alpha(\xi) x \in \mathbb{R}^q.$$

for $\xi \in \Omega$ and $x \in D$. Then $\bar{\mathbf{v}} \in L^2_{\bar{\boldsymbol{\mu}}}(\Omega \times D, \mathbb{R}^{pq})$ is the tangent velocity field to $\bar{\boldsymbol{\mu}}$.

Proof. Take $\psi \in C^1_c(\Omega \times D, \mathbb{R}^p)$ a test function. If we define $\tilde{\psi} \in H^1(\Omega, \mathbb{R}^p)$ by

$$\tilde{\psi}(\xi) := \int_D \psi(\xi, x) \bar{\boldsymbol{\mu}}(\xi, \mathrm{d}x) = \int_D \psi(\xi, \bar{\mathbf{A}}(\xi) x) \rho(x) \mathrm{d}x,$$

then we see that $\tilde{\psi}$ is compactly supported in Ω, in particular the integral of $\nabla \cdot \tilde{\psi}$ over Ω vanishes. It reads

$$\iint_{\Omega \times D} (\nabla_\Omega \cdot \psi)(\xi, \bar{\mathbf{A}}(\xi) x) \rho(x) \mathrm{d}x + \iint_{\Omega \times D} \sum_{\alpha=1}^{p} (\partial_\alpha \bar{\mathbf{A}}(\xi) x) \cdot (\nabla_D \psi^\alpha)(\xi, \bar{\mathbf{A}}(\xi) x) \rho(x) \mathrm{d}x = 0.$$

By doing for a fixed $\xi \in \Omega$ the change of variables $y = \bar{\mathbf{A}}(\xi) x$, one can see that $(\bar{\boldsymbol{\mu}}, \mathbf{w}\bar{\boldsymbol{\mu}})$ satisfies the continuity equation where $\mathbf{w} : \Omega \times D \to \mathbb{R}^p$ is given by

$$\mathbf{w}^\alpha(\xi, y) := [\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1}](\xi)\ y.$$

Notice that $\mathbf{w}(\xi, \cdot)$ is not a gradient because $\partial_\alpha \bar{\mathbf{A}}(\xi)$ and $\bar{\mathbf{A}}(\xi)^{-1}$ do not necessarily commute. On the contrary, as the matrices $\bar{\mathbf{B}}^\alpha(\xi)$ for $\alpha \in \{1, 2, \dots, p\}$ are symmetric, $\bar{\mathbf{v}}(\xi, \cdot)$ is a gradient.

Fix $\xi \in \Omega$ and $\alpha \in \{1, 2, \dots, p\}$. We claim that the velocity field $\bar{\mathbf{v}}^\alpha(\xi, \cdot)$ is the orthogonal projection in $L^2_{\bar{\boldsymbol{\mu}}(\xi)}(D, \mathbb{R}^q)$ of $\mathbf{w}^\alpha(\xi, \cdot)$ on the space of gradients (actually, this is exactly how $\bar{\mathbf{v}}^\alpha$ was chosen). Not to overburden the notations, we drop momentarily the dependence on ξ, i.e. $\bar{\mathbf{A}} := \bar{\mathbf{A}}(\xi)$, $\bar{\mathbf{B}}^\alpha := \bar{\mathbf{B}}^\alpha(\xi)$ and $\partial_\alpha \bar{\mathbf{A}} := \partial_\alpha \bar{\mathbf{A}}(\xi)$ are considered as given matrices. Take $f \in C^1(D)$ a test function defined on D and compute:

$$\begin{aligned} \int_D \nabla f(x) \cdot (\mathbf{w}^\alpha(\xi, x) - \bar{\mathbf{v}}^\alpha(\xi, x)) \bar{\boldsymbol{\mu}}(\xi, \mathrm{d}x) &= \int_D (\nabla f)(\bar{\mathbf{A}} x) \cdot \left((\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1} - \bar{\mathbf{B}}^\alpha) \bar{\mathbf{A}} x \right) \rho(x) \mathrm{d}x \\ &= \int_D (\nabla \tilde{f})(x) \cdot \left(\bar{\mathbf{A}}^{-1} (\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1} - \bar{\mathbf{B}}^\alpha) \bar{\mathbf{A}} x \right) \rho(x) \mathrm{d}x, \end{aligned}$$

where $\tilde{f}(x) := f(\bar{\mathbf{A}} x)$. On the other hand, as the reader can check, $\bar{\mathbf{B}}^\alpha$ is the projection on the set of symmetric matrices of $\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1}$ where the scalar product between two matrices C and D is given by $\mathrm{Tr}(\bar{\mathbf{A}} C^\top D \bar{\mathbf{A}})$. In particular, the matrix $(\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1} - \bar{\mathbf{B}}^\alpha) \bar{\mathbf{A}}^2$ is skew-symmetric, thus the matrix $\bar{\mathbf{A}}^{-1} (\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1} - \bar{\mathbf{B}}^\alpha) \bar{\mathbf{A}}$ is also skew-symmetric. As ρ is radial, it implies that the function

$$x \in D \mapsto \left(\bar{\mathbf{A}}^{-1} (\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1} - \bar{\mathbf{B}}^\alpha) \bar{\mathbf{A}} x \right) \rho(x)$$

is divergence-free. It allows us to conclude that

$$\int_D \nabla f(x) \cdot (\mathbf{w}^\alpha(\xi, x) - \bar{\mathbf{v}}^\alpha(\xi, x))\bar{\boldsymbol{\mu}}(\xi, \mathrm{d}x) = \int_D \tilde{f}(x) \nabla \cdot \left[\left(\bar{\mathbf{A}}^{-1}(\partial_\alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{-1} - \bar{\mathbf{B}}^\alpha) \bar{\mathbf{A}} x \right) \rho(x) \right] \mathrm{d}x = 0,$$

hence the claim is proved as f is arbitrary.

The claim implies that $(\bar{\boldsymbol{\mu}}, \bar{\mathbf{v}}\bar{\boldsymbol{\mu}})$ also satisfies the continuity equation: for any $\psi \in C^1_c(\Omega \times D, \mathbb{R}^p)$,

$$\iint_{\Omega \times D} \nabla_\Omega \cdot \psi \mathrm{d}\bar{\boldsymbol{\mu}} + \iint_{\Omega \times D} \nabla_D \psi \cdot \bar{\mathbf{v}} \mathrm{d}\bar{\boldsymbol{\mu}} = \iint_{\Omega \times D} \nabla_\Omega \cdot \psi \mathrm{d}\bar{\boldsymbol{\mu}} + \iint_{\Omega \times D} \nabla_D \psi \cdot \mathbf{w} \mathrm{d}\bar{\boldsymbol{\mu}} + \iint_{\Omega \times D} \nabla_D \psi \cdot (\bar{\mathbf{v}} - \mathbf{w}) \mathrm{d}\bar{\boldsymbol{\mu}} = 0,$$

as the last integral vanishes because of the claim.

As $\bar{\mathbf{v}}(\xi, \cdot)$ is a gradient (because the $\bar{\mathbf{B}}^\alpha$ are symmetric), Proposition 8.11 implies that $\bar{\mathbf{v}}$ is the tangent velocity field to $\bar{\boldsymbol{\mu}}$. □

Notice that if we define $\bar{\varphi} : \Omega \times D \to \mathbb{R}^p$ by, for any $\xi \in \Omega, x \in D$ and $\alpha \in \{1, 2, \ldots, p\}$,

$$\bar{\varphi}^\alpha(\xi, x) := \frac{1}{2} \bar{\mathbf{B}}^\alpha(\xi) x \cdot x;$$

then $\bar{\mathbf{v}} = \nabla_D \varphi$. More precisely, for a.e. $\xi \in \Omega$, $\bar{\varphi}(\xi, \cdot)$ (resp. $\bar{\mathbf{v}}(\xi, \cdot)$) is defined *everywhere* on D as a smooth function belonging to $C^1(D, \mathbb{R}^p)$ (resp. $C^1(D, \mathbb{R}^{pq})$). Moreover the Euler-Lagrange equation (10.7), which can be written

$$\sum_{\alpha=1}^p \partial_\alpha \bar{\mathbf{B}}^\alpha + \sum_{\alpha=1}^p (\bar{\mathbf{B}}^\alpha)^2 = 0, \tag{10.10}$$

translates at the level of $\bar{\varphi}$ in

$$\nabla_\Omega \cdot \bar{\varphi} + \frac{1}{2} |\nabla_D \bar{\varphi}|^2 = 0. \tag{10.11}$$

In fact, at least formally (because of the lack of smoothness of $\bar{\varphi}$), the function $\bar{\varphi}$ is a solution of the dual problem. For $\bar{\varphi}$ to be an actual solution, we would need the $\bar{\mathbf{B}}^\alpha$ to be C^1 up to the boundary: even with the elliptic regularity proved below (i.e. point (iv) of Theorem 10.10), we would not reach such a strong result if we just assume that $\partial\Omega$ and $\mathbf{A}_l$ are Lipschitz. We will use $\bar{\varphi}$ to show that the tangent velocity field of any other solution of the Dirichlet problem with boundary values $\boldsymbol{\mu}_l$ must coincide with $\bar{\mathbf{v}}$. About the smoothness of the objects involved, notice that for any $\alpha \in \{1, 2, \ldots, p\}$ one has $\bar{\mathbf{B}}^\alpha \in L^2(\Omega, S_q(\mathbb{R}))$ and, given (10.10), the function

$$\sum_{\alpha=1}^p \partial_\alpha \bar{\mathbf{B}}^\alpha$$

belongs to $L^1(\Omega, S_q(\mathbb{R}))$.

Proposition 10.14. *Let $\boldsymbol{\mu}$ a solution of the Dirichlet problem with boundary conditions $\boldsymbol{\mu}_l$ and $\mathbf{v}$ its tangent velocity field. Then, for a.e. $\xi \in \Omega$, one has $\mathbf{v}(\xi, x) = \bar{\mathbf{v}}(\xi, x)$ for $\boldsymbol{\mu}(\xi)$-a.e. x.*

Proof. If $\varphi \in C^1(\Omega \times D, \mathbb{R}^p)$ then, as $\boldsymbol{\mu}$ and $\bar{\boldsymbol{\mu}}$ share the same boundary conditions,

$$\iint_{\Omega \times D} (\nabla_\Omega \cdot \varphi + \nabla_D \varphi \cdot \mathbf{v}) \mathrm{d}\boldsymbol{\mu} = \mathrm{BT}_{\boldsymbol{\mu}_l}(\varphi) = \iint_{\Omega \times D} (\nabla_\Omega \cdot \varphi + \nabla_D \varphi \cdot \bar{\mathbf{v}}) \mathrm{d}\bar{\boldsymbol{\mu}}.$$

We claim that we can insert $\varphi = \bar{\varphi}$ even though $\bar{\varphi}$ is *a priori* not regular enough. In other words, given (10.11) and the fact that $\bar{\mathbf{v}} = \nabla_D \bar{\varphi}$, we claim that

$$\iint\limits_{\Omega\times D} \left(-\frac{1}{2}|\bar{\mathbf{v}}|^2 + \bar{\mathbf{v}} \cdot \mathbf{v} \right) \mathrm{d}\boldsymbol{\mu} = \iint\limits_{\Omega\times D} \frac{1}{2}|\bar{\mathbf{v}}|^2 \mathrm{d}\bar{\boldsymbol{\mu}}. \tag{10.12}$$

Notice that the r.h.s. is (formally) equal to both $\mathrm{BT}_{\boldsymbol{\mu}_l}(\bar{\varphi})$ and $\mathrm{Dir}(\bar{\boldsymbol{\mu}})$: it is not surprising as $\bar{\varphi}$ is a solution of the dual problem.

To prove such an equality we regularize $\bar{\varphi}$ in the following way. For each $\alpha \in \{1, 2, \dots, p\}$ we apply to the matrix field $\bar{\mathbf{B}}^\alpha$ the standard truncation and convolution procedure (see [EG92, Theorem 3 of Section 4.2]) to produce a sequence $(\bar{\mathbf{B}}^\alpha_n)_{n\in\mathbb{N}}$ which belongs to $C^1(\Omega, S_q(\mathbb{R}))$ and which converges to $\bar{\mathbf{B}}^\alpha$ in $L^2(\Omega, S_q(\mathbb{R}))$. Moreover, as derivatives commute with convolution, we can say that

$$\lim_{n\to+\infty} \sum_{\alpha=1}^{p} \partial_\alpha \bar{\mathbf{B}}^\alpha_n = \sum_{\alpha=1}^{p} \partial_\alpha \bar{\mathbf{B}}^\alpha = -\sum_{\alpha=1}^{p} (\bar{\mathbf{B}}^\alpha)^2,$$

and the limit takes place in $L^1(\Omega, S_q(\mathbb{R}))$ as we already know that the r.h.s. belongs to such a space. In particular, up to extraction the convergences hold a.e. on Ω. Then we set

$$\varphi^\alpha_n(\xi, x) := \frac{1}{2}\bar{\mathbf{B}}^\alpha_n(\xi) x \cdot x.$$

for $\xi \in \Omega$ and $x \in D$. By construction $\varphi_n \in C^1(\Omega \times D, \mathbb{R})$ so that

$$\iint\limits_{\Omega\times D} (\nabla_\Omega \cdot \varphi_n + \nabla_D \varphi_n \cdot \mathbf{v}) \mathrm{d}\boldsymbol{\mu} = \mathrm{BT}_{\boldsymbol{\mu}_l}(\varphi_n) = \iint\limits_{\Omega\times D} (\nabla_\Omega \cdot \varphi_n + \nabla_D \varphi_n \cdot \bar{\mathbf{v}}) \mathrm{d}\bar{\boldsymbol{\mu}}. \tag{10.13}$$

It remains to show that we can pass to the limit $n \to +\infty$. Given the convergence a.e. of the $\bar{\mathbf{B}}^\alpha_n$ and of $\sum \partial_\alpha \bar{\mathbf{B}}^\alpha_n$, we can assume that for a.e. $\xi \in \Omega$, the functions $\nabla_\Omega \cdot \varphi_n(\xi, \cdot)$ and $\nabla_D \varphi_n(\xi, \cdot)$ converge to respectively $-\frac{1}{2}|\bar{\mathbf{v}}|^2(\xi, \cdot)$ and $\bar{\mathbf{v}}(\xi, \cdot)$ in respectively $C(D)$ and $C(D, \mathbb{R}^{pq})$ respectively (notice that we use the fact that D is bounded). Hence for a.e. $\xi \in \Omega$,

$$\lim_{n\to+\infty} \int_D (\nabla_\Omega \cdot \varphi_n(\xi, x) + \nabla_D \varphi_n(\xi, x) \cdot \mathbf{v}(\xi, x))\, \boldsymbol{\mu}(\xi, \mathrm{d}x)$$
$$= \int_D \left(-\frac{1}{2}|\bar{\mathbf{v}}|^2(\xi, x) + \bar{\mathbf{v}}(\xi, x) \cdot \mathbf{v}(\xi, x) \right) \boldsymbol{\mu}(\xi, \mathrm{d}x). \tag{10.14}$$

It remains to integrate this limit over Ω. The natural upper bound for the l.h.s. of (10.14) is obtained by Cauchy-Schwarz and the boundedness of D: for any $n \in \mathbb{N}$,

$$\left| \int_D (\nabla_\Omega \cdot \varphi_n(\xi, x) + \nabla_D \varphi_n(\xi, x) \cdot \mathbf{v}(\xi, x))\, \boldsymbol{\mu}(\xi, \mathrm{d}x) \right|$$
$$\leqslant C \left(\sum_{\alpha=1}^{p} |\bar{\mathbf{B}}^\alpha_n(\xi)|^2 + \sqrt{\int_D |\mathbf{v}(\xi, x)|^2 \boldsymbol{\mu}(\xi, \mathrm{d}x)} \sqrt{\sum_{\alpha=1}^{p} |\bar{\mathbf{B}}^\alpha_n(\xi)|^2} \right),$$

where C depends only on D. The r.h.s. is not bounded uniformly w.r.t. $n \in \mathbb{N}$ but on the other hand it converges in $L^1(\Omega)$ which is enough to say that the l.h.s. is uniformly integrable. Hence, up to extraction we can integrate (10.14) w.r.t. Ω:

$$\lim_{n\to+\infty} \iint\limits_{\Omega\times D} (\nabla_\Omega \cdot \varphi_n + \nabla_D \varphi_n \cdot \mathbf{v}) \mathrm{d}\boldsymbol{\mu} = \iint\limits_{\Omega\times D} \left(-\frac{1}{2}|\bar{\mathbf{v}}|^2 + \bar{\mathbf{v}} \cdot \mathbf{v} \right) \mathrm{d}\boldsymbol{\mu}.$$

Of course, the result still holds if we take $(\boldsymbol{\mu}, \mathbf{v}) = (\bar{\boldsymbol{\mu}}, \bar{\mathbf{v}})$. Thus, passing in the limit in (10.13) we get (10.12).

Until now we have not used the optimality of $\boldsymbol{\mu}$. We notice that the r.h.s. of (10.12) is nothing else than $\mathrm{Dir}(\bar{\boldsymbol{\mu}})$ which coincides with $\mathrm{Dir}(\boldsymbol{\mu}) = \iint_{\Omega \times D} \frac{1}{2}|\mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu}$ by optimality of $\boldsymbol{\mu}$. From there, an algebraic manipulation leads to

$$\iint_{\Omega \times D} \frac{1}{2}|\mathbf{v} - \bar{\mathbf{v}}|^2 \mathrm{d}\boldsymbol{\mu} = 0,$$

which easily implies the result: recall that for a.e. $\xi \in \Omega$, the velocity field $\bar{\mathbf{v}}$ is continuous on D. □

Proposition 10.15. *Let $\boldsymbol{\mu}$ a solution of the Dirichlet problem with boundary conditions $\boldsymbol{\mu}_l$. Then $\boldsymbol{\mu} = \bar{\boldsymbol{\mu}}$.*

Proof. Take $\boldsymbol{\mu}$ a solution of the Dirichlet problem with boundary conditions $\boldsymbol{\mu}_l$ and define $\boldsymbol{\nu} = \boldsymbol{\mu} - \bar{\boldsymbol{\mu}}$. We extend $\boldsymbol{\nu}$ on $\mathbb{R}^p \backslash \Omega$ by 0: with such a choice $\boldsymbol{\nu} \in L^2(\mathbb{R}^p, \mathcal{M}(D))$ is a (signed) measure-valued mapping defined on the whole space $\mathbb{R}^p$ which vanishes outside a compact set. We also define $\bar{\mathbf{v}}$ as a function $\mathbb{R}^p \times D \to \mathbb{R}^{pq}$ by extending it to 0 outside $\Omega \times D$. If $\varphi \in C^1(\mathbb{R}^p \times D, \mathbb{R}^p)$ is any smooth function then

$$\begin{aligned}
\iint_{\mathbb{R}^p \times D} (\nabla_\Omega \cdot \varphi + \nabla_D \varphi \cdot \bar{\mathbf{v}})\, \mathrm{d}\boldsymbol{\nu} &= \iint_{\Omega \times D} (\nabla_\Omega \cdot \varphi + \nabla_D \varphi \cdot \bar{\mathbf{v}})\, \mathrm{d}\boldsymbol{\nu} \\
&= \iint_{\Omega \times D} (\nabla_\Omega \cdot \varphi + \nabla_D \varphi \cdot \bar{\mathbf{v}})\, \mathrm{d}\boldsymbol{\mu} - \iint_{\Omega \times D} (\nabla_\Omega \cdot \varphi + \nabla_D \varphi \cdot \bar{\mathbf{v}})\, \mathrm{d}\bar{\boldsymbol{\mu}} \\
&= \mathrm{BT}_{\boldsymbol{\mu}_l}(\varphi) - \mathrm{BT}_{\boldsymbol{\mu}_l}(\varphi) = 0,
\end{aligned}$$

where we have used the fact that both $(\boldsymbol{\mu}, \bar{\mathbf{v}}\boldsymbol{\mu})$ and $(\bar{\boldsymbol{\mu}}, \bar{\mathbf{v}}\bar{\boldsymbol{\mu}})$ satisfy the continuity equation. In other words, $(\boldsymbol{\nu}, \bar{\mathbf{v}}\boldsymbol{\nu})$ satisfy the continuity equation on the whole space $\mathbb{R}^p \times D$.

We take an arbitrary direction in $\mathbb{R}^p$: we fix $\alpha = 1$. As we have seen in the proof of Proposition 8.23, the (generalized) continuity equation implies that for a.e. $\xi \in \mathbb{R}^{p-1} = (e_\alpha)^\perp$, the curve $t \in \mathbb{R} \mapsto \boldsymbol{\nu}((t, \xi))$ satisfies the (1-dimensional) continuity equation with a velocity field given by $\mathbf{w}(t, x) = \bar{\mathbf{v}}^\alpha((t, \xi), x)$. Notice that for a fixed t the velocity field $\mathbf{w}(t, \cdot)$ is Lipschitz and bounded with Lipschitz constant and upper bound controlled by $C \mathbb{1}_{(t,\xi) \in \Omega} |\bar{\mathbf{B}}^\alpha((t, \xi))|$ where $C < +\infty$ depends only on D. Given that $\bar{\mathbf{B}}^\alpha \in L^2(\Omega)$, for a.e. $\xi \in \mathbb{R}^{p-1}$ one has that

$$\int_{\mathbb{R}} \mathbb{1}_{(t,\xi) \in \Omega} |\bar{\mathbf{B}}^\alpha((t, \xi))| \mathrm{d}t < +\infty.$$

Hence for a.e. $\xi \in \mathbb{R}^{p-1}$ the assumptions of [AGS08, Proposition 8.1.7] are satisfied: the curve $t \in \mathbb{R} \mapsto \boldsymbol{\nu}((t, \xi))$ is solution of a continuity equation which has at most one solution. As the curve identically equal to 0 is a solution (recall that $\boldsymbol{\nu}((t, \xi)) = 0$ for $|t|$ large enough), so must be $\boldsymbol{\nu}((\cdot, \xi))$. As this result holds for a.e. $\xi \in \mathbb{R}^{p-1}$, it implies that $\boldsymbol{\nu}$ is identically zero, which is the desired result. □

Eventually, to prove regularity, following the theory of Schoen and Uhlenbeck [SU82, SU83], we only need to show that there is no *minimizing tangent maps*, i.e. no Dirichlet minimizing mapping which is 0-homogeneous. We start with the following result.

Proposition 10.16. *Let* $\mathbf{A} \in H^1(\Omega, S_q^{++}(\mathbb{R}))$ *be a weak solution of* (10.7)*, bounded from above and uniformly away from singular matrices, and* $C \in S_q^+(\mathbb{R})$ *a semi-definite positive matrix. Then the (real-valued) mapping*

$$f : \xi \in \Omega \to \langle \mathbf{A}(\xi)^2, C \rangle$$

is subharmonic.

Actually, this is nothing else than the Ishihara property (Theorem 9.3) for the functional $\mu \mapsto \int_D \xi \cdot (C\xi)\mu(\mathrm{d}\xi)$, though in this simpler case we can show that it holds for any solution, as we can check it by a straightforward computation.

Proof. As in Proposition 10.13, for $\alpha \in \{1, 2, \ldots, p\}$, we set $\mathbf{B}^\alpha := L_{\mathbf{A}} L_{\mathbf{A}^2}^{-1}(\partial_\alpha \mathbf{A})$. Thanks to the assumptions on $\mathbf{A}$, we know that $\mathbf{B}^\alpha \in L^2(\Omega, S_q(\mathbb{R}))$: this regularity is enough to justify the following computations. Indeed, with this notation at hand, for any $\alpha \in \{1, 2, \ldots, p\}$

$$\partial_\alpha f = \langle L_{\mathbf{A}}(\partial_\alpha \mathbf{A}), C \rangle = \langle L_{\mathbf{A}^2}(\mathbf{B}^\alpha), C \rangle = \langle \mathbf{B}^\alpha, L_{\mathbf{A}^2}(C) \rangle.$$

Hence, taking the derivative again and summing over α,

$$\begin{aligned}
\Delta f &= \sum_{\alpha=1}^{p} \left(\langle \partial_\alpha \mathbf{B}^\alpha, L_{\mathbf{A}^2}(C) \rangle + \langle \mathbf{B}^\alpha, L_{L_{\mathbf{A}}(\partial_\alpha \mathbf{A})}(C) \rangle \right) \\
&= \sum_{\alpha=1}^{p} \left(\langle \partial_\alpha \mathbf{B}^\alpha, L_{\mathbf{A}^2}(C) \rangle + \langle \mathbf{B}^\alpha, L_{L_{\mathbf{A}^2}(\mathbf{B}^\alpha)}(C) \rangle \right) \\
&= \sum_{\alpha=1}^{p} \left(\langle \partial_\alpha \mathbf{B}^\alpha, L_{\mathbf{A}^2}(C) \rangle + \operatorname{Tr}\left(\left[2\mathbf{B}^\alpha \mathbf{A}^2 \mathbf{B}^\alpha + (\mathbf{B}^\alpha)^2 \mathbf{A}^2 + \mathbf{A}^2 (\mathbf{B}^\alpha)^2 \right] C \right) \right).
\end{aligned}$$

Now, using (10.7) which reads $\sum_\alpha \partial_\alpha \mathbf{B}^\alpha = -\sum_\alpha (\mathbf{B}^\alpha)^2$, one reaches the conclusion that

$$\Delta f = 2 \sum_{\alpha=1}^{p} \operatorname{Tr}\left(\mathbf{B}^\alpha \mathbf{A}^2 \mathbf{B}^\alpha C \right).$$

The matrix $\mathbf{B}^\alpha \mathbf{A}^2 \mathbf{B}^\alpha$ belongs to $S_q^+(\mathbb{R})$ because $\mathbf{A}$ does, and so does C by assumption. As the trace of the product of two elements of $S_q^+(\mathbb{R})$ is non negative, we deduce $\Delta f \geqslant 0$ which was the claim. □

With this result, it is easy to see that there exists no non constant 0-homogeneous tangent maps. Notice, by point (i) of Theorem 10.10, and as D is bounded, that any minimizing tangent map, if it were to exist, would be bounded from above and uniformly away from singular matrices.

Proposition 10.17. *Assume* $\Omega = \mathbb{B}$ *the unit ball of dimension* p *and* $\mathbf{A} \in H^1(\Omega, S_q^{++}(\mathbb{R}))$ *is a weak solution of* (10.7)*, bounded from above and uniformly away from singular matrices, which is* 0*-homogeneous, meaning that* $\mathbf{A}(\lambda\xi) = \mathbf{A}(\xi)$ *for any* $\lambda > 0$*. Then* $\mathbf{A}$ *is constant.*

Proof. According to Proposition 10.16, for any $C \in S_q^+(\mathbb{R})$, the function

$$f : \xi \in \Omega \to \langle \mathbf{A}(\xi)^2, C \rangle$$

is subharmonic and 0-homogeneous, hence it is constant by the maximum principle. But clearly, the scalar product between $\mathbf{A}$ and any given symmetric positive matrix is constant if and only if $\mathbf{A}$ is itself constant. □

10.4 An (almost) explicit example

In this section we want to give a case where the solution to the Dirichlet problem can be (almost) exactly computed and which, in the same time, exhibits interesting effects of the geometry of the Wasserstein space. This example deals with mappings valued in the set of elliptically contoured distributions (see the previous section), hence we will look only at the square root of the covariance matrices.

We choose $\Omega := \mathbb{B}$ the unit disk of $\mathbb{R}^2$. We will work in polar coordinate, i.e. a generic point of Ω will be characterize by r the distance to the origin and θ the angle made with the axis Ox. The domain D is included $\mathbb{R}^2$: as explained in the previous section, we don't really care about the specific form of D as we will work with a family of elliptically contoured distributions.

For any $\theta \in \mathbb{R}$ and any $(\kappa_1, \kappa_2) \in \mathbb{R}^2$ real numbers, we define the following 2×2 matrices:

$$\mathcal{R}(\theta) := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad \mathcal{D}(\kappa_1, \kappa_2) := \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}, \quad \mathcal{S}(\theta) := \mathcal{R}(-\theta)\mathcal{D}(1, -1)\mathcal{R}(\theta)$$

$\mathcal{R}(\theta)$ is the rotation by an angle θ, while $\mathcal{D}(\kappa_1, \kappa_2)$ is just a diagonal matrix and $\mathcal{S}(\theta)$ is the orthogonal symmetry w.r.t. the line making an angle θ with the horizontal axis. Now, we fix numbers $0 < \bar{\kappa}_1 \leqslant \bar{\kappa}_2$ and we define the matrix field

$$\mathbf{A}_l(\theta) := \mathcal{R}(-\theta)\mathcal{D}(\bar{\kappa}_1, \bar{\kappa}_2)\mathcal{R}(\theta)$$

which is defined on $\partial\mathbb{B}$ (parametrized in polar coordinates). We set $\boldsymbol{\mu}_l = \rho_{\mathbf{A}_l}$. The matrices $\mathbf{A}_l$ are uniformly bounded from below, and the mapping $\theta \to \mathbf{A}_l(\theta)$ is Lipschitz hence we can apply Theorem 10.10 and conclude that there exists a unique solution $\boldsymbol{\mu}$ to the Dirichlet problem with boundary values $\rho_{\mathbf{A}_l}$. Moreover, this solution is valued in the set of elliptically contoured distributions. Let us denote $\mathbf{A}(r, \theta) := \mathrm{cov}(\boldsymbol{\mu}(r, \theta))^{1/2}$. Then we can give an almost explicit expression for $\mathbf{A}$.

Theorem 10.18. *Let $\mathbf{A} : \mathbb{B} \to S_2^{++}(\mathbb{R})$ defined by $\mathbf{A} := \mathrm{cov}(\boldsymbol{\mu})^{1/2}$ where $\boldsymbol{\mu}$ is the unique solution of the Dirichlet problem with boundary conditions $\boldsymbol{\mu}_l$ as described above. There exists two functions $\kappa_1, \kappa_2 : [0, 1] \to [\bar{\kappa}_1, \bar{\kappa}_2]$ such that for any $(r, \theta) \in \mathbb{B}$,*

$$\mathbf{A}(r, \theta) = \mathcal{R}(-\theta)\mathcal{D}(\kappa_1(r), \kappa_2(r))\mathcal{R}(\theta). \tag{10.15}$$

Moreover, the functions κ_1, κ_2 satisfy the following properties.

(i) The functions κ_1, κ_2 are smooth with $\kappa_1(0) = \kappa_2(0)$ and $\kappa_1(1) = \bar{\kappa}_1$, $\kappa_2(1) = \bar{\kappa}_2$.

(ii) For any $r \geqslant 0$, there holds $\kappa_1(r) \leqslant \kappa_2(r)$.

(iii) The pair (κ_1, κ_2) minimizes

$$\int_0^1 \left(\frac{r}{2}((\kappa_1')^2 + (\kappa_2')^2) + \frac{1}{r} \frac{(\kappa_1^2 - \kappa_2^2)^2}{\kappa_1^2 + \kappa_2^2} \right) \tag{10.16}$$

among all pairs satisfying (i). In particular, it solves the following system

$$\begin{cases} (r\kappa_1')' &= \dfrac{1}{r} \left(\dfrac{4\kappa_1(\kappa_1^2 - \kappa_2^2)}{\kappa_1^2 + \kappa_2^2} - \dfrac{2\kappa_1(\kappa_1^2 - \kappa_2^2)^2}{(\kappa_1^2 + \kappa_2^2)^2} \right), \\ (r\kappa_2')' &= \dfrac{1}{r} \left(\dfrac{4\kappa_2(\kappa_2^2 - \kappa_1^2)}{\kappa_1^2 + \kappa_2^2} - \dfrac{2\kappa_2(\kappa_1^2 - \kappa_2^2)^2}{(\kappa_1^2 + \kappa_2^2)^2} \right). \end{cases}$$

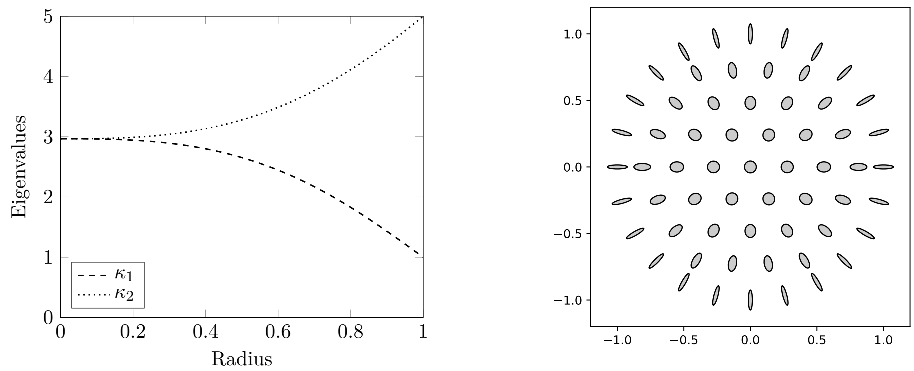

Figure 10.1: Numerical computation of the solution of the problem explored in this section with $\bar{\kappa}_1 = 1$ and $\bar{\kappa}_2 = 5$. On the left, plot of (κ_1, κ_2) minimizing a finite difference version of (10.16) with $N = 100$ discretization points for r. On the right, representation of the resulting harmonic mapping: on points $\xi \in \mathbb{B}$ the source space are displayed ellipses defined by $\mathbf{A}(\xi)$. The matrices $\mathbf{A}(\xi)$ were computed with (10.15) and the optimizer of the finite difference version of (10.16).

Here κ_1', κ_2' denote the derivatives of κ_1, κ_2. A plot of the solution is displayed in Figure 10.1. The interpretation is that, along a given radius of $\mathbb{B}$, all the matrices $\mathbf{A}$ are diagonal in the same basis, but the eigenvalues κ_1, κ_2 depend on the distance to the center. On the other hand, on a given circle around the origin, the matrices $\mathbf{A}$ share the same eigenvalues but the eigenvectors depend on the angle θ.

Proof. The smoothness of $\mathbf{A}$ directly derives from point (iv) of Theorem 10.10. The regularity up to the boundary holds because here both $\mathbf{A}_l$ and $\partial\mathbb{B}$ are C^∞ objects. As $\mathbf{A}$ is uniformly non singular, it easily implies that, valued in $S_2^{++}(\mathbb{R})$ endowed with its Euclidean structure, $\mathbf{A}$ is also a C^∞ mapping.

To prove the specific form that $\mathbf{A}$ takes, we will use symmetry arguments and *uniqueness* of the solution to the Dirichlet problem. For a given $\theta_0 \in \mathbb{R}$, we consider the matrix field $\mathbf{B}$ defined by

$$\mathbf{B}(r, \theta) := \mathcal{R}(-\theta_0)\mathbf{A}(r, \theta - \theta_0)\mathcal{R}(\theta_0).$$

One can see easily that $\mathbf{B}$ shares the same boundary conditions as $\mathbf{A}$. On the other hand, $(r, \theta) \mapsto \rho_{\mathbf{A}(r, \theta-\theta_0)}$ share the same Dirichlet energy as $\mathbf{A}$ and $C \mapsto \mathcal{R}(-\theta_0)C\mathcal{R}(\theta_0)$ is an isometry of $(S_2^{++}(\mathbb{R}), \mathfrak{g})$ hence the Dirichlet energy of $\rho_{\mathbf{B}}$ is the same as the one of $\rho_{\mathbf{A}}$. By uniqueness in the Dirichlet problem, $\mathbf{A} = \mathbf{B}$ which means in particular

$$\mathbf{A}(r, \theta) := \mathcal{R}(-\theta)\mathbf{A}(r, 0)\mathcal{R}(\theta)$$

for any $(r, \theta) \in \mathbb{B}$. We still have to justify that $\mathbf{A}(r, 0)$ is diagonal in the canonical basis of $S_2^{++}(\mathbb{R})$. To this end, we now use the competitor $\mathbf{B}(r, \theta) := \mathcal{S}(\theta)\mathbf{A}(r, \theta)\mathcal{S}(\theta)$. With this $\mathbf{B}$, one can check that

$$\mathbf{B}(\theta, r) := \mathcal{R}(-\theta)\mathbf{B}(r, 0)\mathcal{R}(\theta)$$

still holds and $\mathbf{B}$ shares the same boundary conditions as $\mathbf{A}$. Along a radius of $\mathbb{B}$, as $C \mapsto \mathcal{S}(\theta)C\mathcal{S}(\theta)$ is an isometry of $(S_2^{++}(\mathbb{R}), \mathfrak{g})$, thus the contribution to the Dirichlet energy of the radial derivatives of $\mathbf{B}$ is the same as the one of $\mathbf{A}$. On the other hand, for each r we know that there exists θ_r such that one eigenvector of $\mathbf{A}(0, r)$ makes an angle θ_r with Ox. In particular, there holds $\mathbf{A}(0,r) = \mathcal{R}(-2\theta_r)\mathbf{B}(0,r)\mathcal{R}(2\theta_r) = \mathbf{B}(2\theta r, r)$. As a consequence $\mathbf{A}(\theta,r) = \mathbf{B}(\theta + 2\theta_r, r)$ for all $\theta \in \mathbb{R}$. It shows that the tangential part of the Dirichlet energy on the circle of radius r is the same for $\mathbf{A}$ and $\mathbf{B}$. As it is the case for every r, we deduce that $\mathbf{A}$ and $\mathbf{B}$ share the same Dirichlet energy, hence coincide. It reads $\mathbf{A}(0,r)$ commutes with $\mathcal{S}(0)$, which translates in the fact that $\mathbf{A}(0,r)$ is a diagonal matrix.

Thus, we just define κ_1, κ_2 as the function such that $\mathbf{A}(r,0) = \mathcal{D}(\kappa_1(r), \kappa_2(r))$. Given that $\mathbf{A}$ is smooth, we also know that κ_1, κ_2 are smooth. As $\mathbf{A}$ has boundary conditions $\mathbf{A}_l$, we easily identify $\kappa_1(1) = \bar{\kappa}_1,\ \kappa_2(1) = \bar{\kappa}_2$.

Notice that the Dirichlet energy of $\rho_{\mathbf{A}}$ is given, as we work in polar coordinates, by

$$\mathrm{Dir}(\rho_{\mathbf{A}}) = \int_0^1 \int_0^{2\pi} \frac{1}{2} \left(r \langle \partial_r \mathbf{A}, \mathfrak{g}_{\mathbf{A}}(\partial_r \mathbf{A}) \rangle + \frac{1}{r} \langle \partial_\theta \mathbf{A}, \mathfrak{g}_{\mathbf{A}}(\partial_\theta \mathbf{A}) \rangle \right) \mathrm{d}\theta \mathrm{d}r.$$

We need to develop this expression given the specific form of $\mathbf{A}$. For the radial component it is easy as $\partial_r \mathbf{A}(r,\theta) = \mathcal{R}(-\theta)\mathcal{D}(\kappa_1'(r), \kappa_2'(r))\mathcal{R}(\theta)$: given (10.5), we have $\langle \partial_r \mathbf{A}, \mathfrak{g}_{\mathbf{A}}(\partial_r \mathbf{A}) \rangle = (\kappa_1')^2 + (\kappa_2')^2$. The derivation for the tangential part is more tedious but straightforward: we compute

$$\partial_\theta \mathbf{A}(r,\theta) = (\kappa_2(r) - \kappa_1(r))\mathcal{R}(-\theta) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathcal{R}(\theta),$$

and then we plug in (10.5) to get

$$\langle \partial_\theta \mathbf{A}, \mathfrak{g}_{\mathbf{A}}(\partial_\theta \mathbf{A}) \rangle = \frac{(\kappa_1^2 - \kappa_2^2)^2}{\kappa_1^2 + \kappa_2^2}.$$

Hence, we conclude that the Dirichlet energy of $\mathbf{A}$ is nothing else, up to a multiplicative constant, than (10.16).

From this information, we see that indeed (iii) is satisfied. Moreover, we infer that $\kappa_1(0) = \kappa_2(0)$, otherwise $\mathbf{A}$ would not be a smooth mapping. Of course, the system of ODEs satisfied by (κ_1, κ_2) is nothing else than the Euler Lagrange equations associated to the minimization of (10.16).

Eventually, if $\kappa_2(r_0) < \kappa_1(r_0)$ for some r_0, then there exists $r_1 \in (0,1)$ such that $\kappa_1(r_1) = \kappa_2(r_1)$. Then, setting $\kappa_1(r) = \kappa_2(r) = \kappa_1(r_1)$ for all $r \in [0, r_1]$ would decrease the Dirichlet energy, hence a contradiction. □

Notice that the condition $\kappa_1(0) = \kappa_2(0)$ means that the measure $\boldsymbol{\mu}(0)$ is isotropic, in particular that $\boldsymbol{\mu}$ is continuous in 0, there is no blow up. We call this example *almost* explicit because we do not have an analytical formula for (κ_1, κ_2). On the other hand, we have computed a finite difference approximation of the solution. Indeed, denoting by $N+1$ the number of discretization points, $\tau = 1/N$ the spatial step, a finite difference approximation of (10.16) is

$$\sum_{k=0}^{N-1} \tau \frac{1}{2} \cdot \left(k\tau + \frac{1}{2} \right) \left[\left(\frac{\kappa_1((k+1)\tau) - \kappa_1(\tau)}{\tau} \right)^2 + \left(\frac{\kappa_2((k+1)\tau) - \kappa_2(\tau)}{\tau} \right)^2 \right] + \sum_{k=1}^{N} \tau \left[\frac{1}{k\tau} \frac{(\kappa_1(k\tau)^2 - \kappa_2(k\tau)^2)^2}{\kappa_1(k\tau)^2 + \kappa_2(k\tau)^2} \right].$$

We have minimized this functional thanks to a simple gradient descent algorithm[1], the result is displayed in Figure 10.1. We have reached a critical point of this functional, and though we have

[1] The code is available at `https://github.com/HugoLav/PhD`.

no guarantee that it is indeed a minimizer (this problem is *not* convex in κ_1, κ_2), we have found that random initialization leads to the same output, at that this output is in accordance with Theorem 10.18.

Moreover, with this (almost) explicit expression at hand, we want to derive two consequences.

First, we know that there is some r such that $\boldsymbol{\mu}(0)$ is *not* the barycenter of the $\boldsymbol{\mu}(r,\theta)$ for $\theta \in [0, 2\pi]$. In other words, by calling $\mathbb{B}_r := B(0,r)$, we see that $\boldsymbol{\mu}$ is harmonic on $\mathbb{B}_r$ but $\boldsymbol{\mu}(0)$ is *not* the Wasserstein barycenter of the values of $\boldsymbol{\mu}$ on the boundary on $\partial\mathbb{B}_r$. It shows that there is no hope of writing a Green formula stating that the values of $\boldsymbol{\mu}$ at one point are the (weighted) Wasserstein barycenters of the values of $\boldsymbol{\mu}$ on the boundary (which is true for harmonic mappings valued in $\mathbb{R}$).

To back such a claim, it is enough to notice that the Wasserstein barycenter of the $\boldsymbol{\mu}(r,\theta)$ for $\theta \in [0, 2\pi]$ is $\rho_{\kappa \mathrm{Id}}$ with $\kappa = (\kappa_1(r) + \kappa_2(r))/2$. Indeed, by symmetry this barycenter is of the form $\rho_{\kappa \mathrm{Id}}$ and a very simple optimization problem leads to the explicit expression of κ. Indeed, we recall that if two matrices A, B commute, then the Wasserstein distance between ρ_A and ρ_B coincides with the euclidean distance (the one induced by the Hilbert-Schmidt norm) between A and B. Now, if $\boldsymbol{\mu}(0)$ were the barycenter of the $(\boldsymbol{\mu}(r,\theta))_{\theta \in [0,2\pi]}$ it would mean that the function $\kappa := (\kappa_1 + \kappa_2)/2$ is constant. Using the system of ODE for κ_1 and κ_2, it would lead to

$$0 = \left(r \left(\frac{\kappa_1 + \kappa_2}{2} \right)' \right)' = \frac{\kappa_1 + \kappa_2}{r} \left(\frac{2(\kappa_1 - \kappa_2)^2}{\kappa_1^2 + \kappa_2^2} - \frac{(\kappa_1^2 - \kappa_2^2)^2}{(\kappa_1^2 + \kappa_2^2)^2} \right) = \frac{(\kappa_1 + \kappa_2)}{r(\kappa_1^2 + \kappa_2^2)^2} (\kappa_1 - \kappa_2)^4.$$

Hence, it is straightforward that the r.h.s. does not vanish if $\kappa_1 \neq \kappa_2$, which happens at least close to $r = 1$ if $\bar{\kappa}_1 < \bar{\kappa}_2$. On Figure 10.1, though it is not really visually apparent, we were able to check that $(\kappa_1 + \kappa_2)/2$ is not constant.

The second interesting consequence is that it provides an explicit example where we can show that the commutativity relation (8.19) does not hold, hence we know that the superposition principle should fail for this mapping. Recall that this relation is the following: for every $\alpha, \beta \in \{1, 2\}$,

$$\partial_\alpha \mathbf{v}^{\beta i} + \sum_{j=1}^q \mathbf{v}^{\alpha j} \partial_j \mathbf{v}^{\beta i} = \partial_\beta \mathbf{v}^{\alpha i} + \sum_{j=1}^q \mathbf{v}^{\beta j} \partial_j \mathbf{v}^{\alpha i}. \tag{10.17}$$

We claim that this relation does not hold. We have to write the tangent velocity field for the mapping $\mathbf{A}$. We know that, for a direction α, they are of the form $\mathbf{v}^\alpha(\xi, x) = \mathbf{B}^\alpha(\xi)x$ for $x \in D$, where $\mathbf{B}^\alpha$ is defined in Proposition 10.13. In our specific example we have, using the computations of the proof of Theorem 10.18,

$$\begin{aligned}
\mathbf{B}^r(r,\theta) &= \mathcal{R}(-\theta) \begin{pmatrix} \kappa_1'(r)/\kappa_1(r) & 0 \\ 0 & \kappa_2'(r)/\kappa_2(r) \end{pmatrix} \mathcal{R}(\theta), \\
\mathbf{B}^\theta(r,\theta) &= \frac{\kappa_2^2 - \kappa_1^2}{r(\kappa_1^2 + \kappa_2^2)} \mathcal{R}(-\theta) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \mathcal{R}(\theta).
\end{aligned}$$

Now, we notice that in (10.17), the terms in $\partial_\alpha \mathbf{v}^\beta$ and $\partial_\beta \mathbf{v}^\alpha$ would lead to terms of the form $(\xi, x) \mapsto \mathbf{C}(\xi)x$ where $\mathbf{C}$ are symmetric matrices (because the derivatives of symmetric matrices are symmetric matrices). On the other hand, choosing as directions $(e_\alpha, e_\beta) = (e_r, e_\theta)$ one can write

$$\left(\sum_{j=1}^q \mathbf{v}^{\alpha j} \partial_j \mathbf{v}^{\beta i} - \sum_{j=1}^q \mathbf{v}^{\beta j} \partial_j \mathbf{v}^{\alpha i} \right) ((r,\theta), x) = \Big[(\mathbf{B}^\theta \mathbf{B}^r - \mathbf{B}^r \mathbf{B}^\theta)(r,\theta) x \Big]_i,$$

in
Hence it is enough for this part not to vanish in order to conclude that (10.17) does not hold. But this part vanishes only if $\mathbf{B}^r$ and $\mathbf{B}^\theta$ commute. From the explicit expressions that we have, if $\kappa_1 \neq \kappa_2$ they do not share the same eigenvectors hence do not commute.

Chapter 11

Numerical illustrations

The goal of this chapter is to present the numerical method that we use to compute approximations of the harmonic mappings valued in the Wasserstein space. The actual implementation of this method can be found online at the following address

`https://github.com/HugoLav/PhD`

As we said earlier, there is no Lagrangian point of view nor static formulation for mappings valued in the Wasserstein space. Hence, the main tool to handle numerics appears to be the so-called Benamou Brenier formula. We underline that the content of this chapter is not really satisfactory: we only provide a consistent discretization, but we are unable to prove the convergence of it when one refines the discretization.

We will work with finite difference discretizations. Hence, to simplify the analysis, we restrict ourselves to the following framework in this whole chapter.

Assumptions. *The domain Ω is the unit square of $\mathbb{R}^2$. The domain D is the 2-dimensional torus $(\mathbb{R}/\mathbb{Z})^2$.*

We only work on spaces of dimension 2 because of scalability issues: as our unknowns will be defined on the space $\Omega \times D$, we cannot really afford the dimension of this space to be larger than 4. On the other hand, having Ω of dimension 1 is the already known case of geodesics in the Wasserstein space, for which several algorithms exist; and if D is of dimension 1 the problem becomes too simple as explained in Section 10.2. Eventually, we take D to be the torus because it helps us to avoid handling what happens at the boundaries of D. As mentioned earlier all of the theory developed in the previous chapters can be adapted straightforwardly to this case.

We want to discretize the variational problem stated in Definition 8.30. To this end, we will use the Benamou-Brenier formulation of the Dirichlet energy. Actually, we will rather start with a discretization of the dual problem, see Theorem 8.36. In short, given a boundary condition $\boldsymbol{\mu}_b : \partial\Omega \to \mathcal{P}(D)$, the (continuous) Dirichlet problem consist in these two formulations, that we call respectively primal and dual:

$$\min_{\boldsymbol{\mu},\mathbf{E}} \left\{ \iint_{\Omega\times D} \frac{|\mathbf{E}|^2}{2\boldsymbol{\mu}} \ : \ \nabla_\Omega \mu + \nabla_D \cdot \mathbf{E} = 0 \text{ and } \boldsymbol{\mu} = \boldsymbol{\mu}_b \text{ on } \partial\Omega \right\},$$

$$\sup_{\varphi} \left\{ \int_{\partial\Omega} \left(\int_D \varphi(\xi, x) \cdot \mathbf{n}_\Omega(\xi) \boldsymbol{\mu}_b(\xi, \mathrm{d}x) \right) \sigma(\mathrm{d}\xi) \ : \ \nabla_\Omega \cdot \varphi + \frac{1}{2} |\nabla_D \varphi|^2 \leqslant 0 \right\}.$$

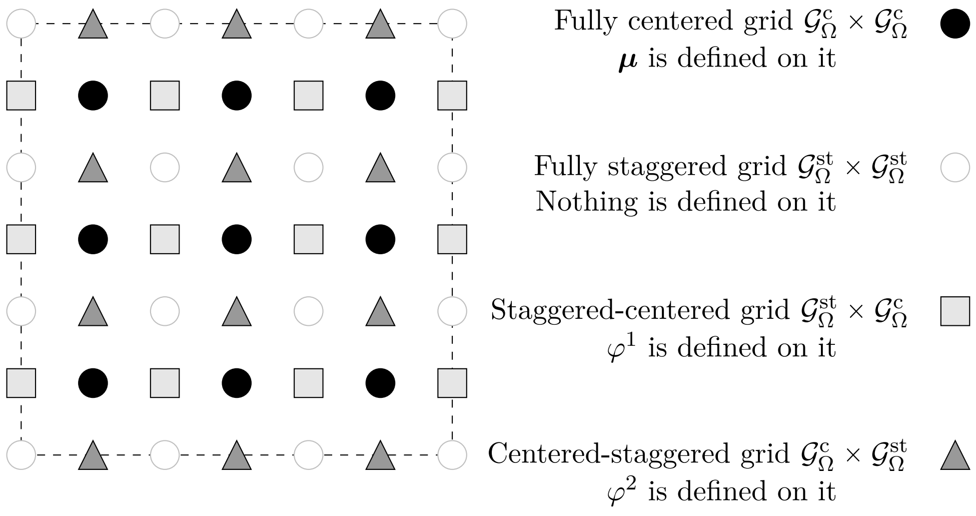

Figure 11.1: Example of the different grids considered on Ω for the case $N = 3$. We mention that the boundary values $\boldsymbol{\mu}$ are defined on the intersection of the staggered-centered and centered-staggered with the boundary of Ω (displayed as a dashed line).

The primal unknowns $\mu, \mathbf{E}$ belong to $\mathcal{P}(\Omega \times D)$ and $\mathcal{M}(\Omega \times D, \mathbb{R}^4)$, while the dual unknown φ is an element of $C^1(\Omega \times D, \mathbb{R}^2)$. The link between the optimizers of these two problems is $\mathbf{E} = (\nabla_D \varphi)\boldsymbol{\mu}$, but we do not know if an optimal φ exists, and even if it does, if it is not smooth enough the knowledge of φ does not determine uniquely $\boldsymbol{\mu}$ and $\mathbf{E}$.

11.1 Discretization

Both the unit square Ω and the torus D will be discretized with uniform grids. As argued in [PPO14], if one uses finite differences it is better to use grids which are staggered with respect to each other.

We denote by N the number of discretization points per dimension in Ω. The grid step is $\tau = 1/N$. We consider two 1-dimensional grids, called respectively the centered and the staggered grid.

$$\mathcal{G}_\Omega^c = \left\{ \left(i + \frac{1}{2} \right) \tau \ : \ \tau = 0, 1, \ldots, N-1 \right\} \subset [0,1],$$
$$\mathcal{G}_\Omega^{st} = \{ i\tau \ : \ \tau = 0, 1, \ldots, N \} \subset [0,1].$$

The staggered grid has $N+1$ points, while the centered one has only N. We will consider the 2-dimensional grids $\mathcal{G}_\Omega^c \times \mathcal{G}_\Omega^c$, $\mathcal{G}_\Omega^{st} \times \mathcal{G}_\Omega^c$ and $\mathcal{G}_\Omega^c \times \mathcal{G}_\Omega^{st}$: if we were to divide Ω in N^2 equal squares, the first grid would be located on the centers of the squares, the second one the centers of the vertical interfaces between squares, and the third one on the centers of the horizontal interfaces between squares (see Figure 11.1).

On the other hand, we denote by M the number of discretization points in D and $\delta = 1/M$ the grid step. We consider two 1-dimensional grids, called respectively the centered and the staggered grid.

$$\mathcal{G}_D^{\mathrm{c}} = \left\{ \left(i + \frac{1}{2} \right) \delta \ : \ i = 0, 1, \ldots, M-1 \right\} \subset \mathbb{R}/\mathbb{Z},$$
$$\mathcal{G}_D^{\mathrm{st}} = \{ i\delta \ : \ i = 0, 1, \ldots, M-1 \} \subset \mathbb{R}/\mathbb{Z}$$

Contrary to Ω, the grids $\mathcal{G}_D^{\mathrm{c}}$ and $\mathcal{G}_D^{\mathrm{st}}$ have the same cardinality: this is because of the absence of boundary on D.

We start by explaining how we discretize the *dual* formulation of the Dirichlet problem. In the continuous world, an unknown of this problem is defined on $\Omega \times D$ and valued in $\mathbb{R}^2$. The function φ has two components φ^1, φ^2. They will be defined over the following girds:

$$\varphi^1 : \mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \to \mathbb{R},$$
$$\varphi^2 : \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \to \mathbb{R}$$

More precisely, we call $X_{N,M} := \mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}} \times \mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}} \simeq \mathbb{R}^{2(N+1)NM^2}$ the finite-dimensional space to which such a φ belong. This space is endowed with the scalar product $\langle \ , \ \rangle_X := \tau^2 \langle \ , \ \rangle$ which is the canonical scalar product on $\mathbb{R}^{2(N+1)NM^2}$ multiplied by the scaling factor τ^2. For a $\varphi \in X_{N,M}$, it is not difficult to find a consistent discretization of the divergence w.r.t. variables in the source space Ω, actually this space was chosen for that. Indeed, this discretization should be, for $(\xi^1, \xi^2, x^1, x^2) \in \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}$,

$$(\nabla_\Omega^{\mathrm{dsc}} \cdot \varphi)_{\xi^1,\xi^2,x^1,x^2} := \left(\frac{\varphi^1_{\xi^1+\tau/2,\xi^2,x^1,x^2} - \varphi^1_{\xi^1-\tau/2,\xi^2,x^1,x^2}}{\tau} \right) + \left(\frac{\varphi^2_{\xi^1,\xi^2+\tau/2,x^1,x^2} - \varphi^2_{\xi^1,\xi^2-\tau/2,x^1,x^2}}{\tau} \right).$$

Notice that the discrete version of $\nabla_\Omega \cdot \varphi$ ends up on a single grid, which is the "fully" centered one. On the other hand, if one uses finite differences to compute the gradients $\nabla_D \varphi$, they are not defined on the same grid. For instance, the discrete derivative of φ^1 w.r.t. the first coordinate of D, denoted by $\partial_{D,1}^{\mathrm{dsc}} \varphi^1$ is naturally defined on $\mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{st}} \times \mathcal{G}_D^{\mathrm{c}}$ by

$$\partial_{D,1}^{\mathrm{dsc}} \varphi^1_{\xi^1,\xi^2,x^1,x^2} := \frac{\varphi^1_{\xi^1,\xi^2,x^1+\delta/2,x^2} - \varphi^1_{\xi^1,\xi^2,x^1-\delta/2,x^2}}{\delta},$$

and $\partial_{D,i}^{\mathrm{dsc}} \varphi^\alpha$ are defined similarly by permutation of the indices $\alpha, i \in \{1, 2\}$.

As they will be important later, and to compactify (a little bit) the notations, we introduce variables A and $B^{\alpha i}$ (with $\alpha, i \in \{1, 2\}$) such that the following constraints hold

$$A = \nabla_\Omega^{\mathrm{dsc}} \cdot \varphi,$$
$$B^{\alpha i} = \partial_{D,i}^{\mathrm{dsc}} \varphi^\alpha \quad \forall \alpha, i \in \{1, 2\}.$$

In particular, A (resp. $B^{\alpha i}$) is defined on the same grid as $\nabla_\Omega^{\mathrm{dsc}} \cdot \varphi$ (resp. $\partial_{D,i}^{\mathrm{dsc}} \varphi^\alpha$). We will call $Y_{N,M}$ the finite-dimensional space to which (A, B) belongs: it is the space

$$\underbrace{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}}}_{\nabla_\Omega^{\mathrm{dsc}} \cdot \varphi} \times \underbrace{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{st}} \times \mathcal{G}_D^{\mathrm{c}}}}_{\partial_{D,1}^{\mathrm{dsc}} \varphi^1} \times \underbrace{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{st}}}}_{\partial_{D,2}^{\mathrm{dsc}} \varphi^1} \times \underbrace{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_D^{\mathrm{st}} \times \mathcal{G}_D^{\mathrm{c}}}}_{\partial_{D,1}^{\mathrm{dsc}} \varphi^2} \times \underbrace{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{st}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{st}}}}_{\partial_{D,2}^{\mathrm{dsc}} \varphi^2}.$$

Similarly to $X_{N,M}$, this space is endowed with the scalar product $\langle \ , \ \rangle_Y := \tau^2 \langle \ , \ \rangle$ which is the canonical scalar product on $\mathbb{R}^{(N^2+4N(N+1))M^2}$ multiplied by the scaling factor τ^2. The discrete differentiation operator will be subsumed under the letter $\mathcal{D}^{\mathrm{dsc}}$, namely

$$\mathcal{D}^{\mathrm{dsc}} := \begin{pmatrix} \nabla_\Omega^{\mathrm{dsc}} \cdot \\ \left(\partial_{D,i}^{\mathrm{dsc}} \varphi^\alpha \right)_{\alpha, i \in \{1,2\}} \end{pmatrix},$$

which is a linear operator going from $X_{N,M}$ to $Y_{N,M}$. The relation between (A, B) and φ is simply written $(A, B) = \mathcal{D}^{\mathrm{dsc}}\varphi$.

We need to define the constraint corresponding to the Hamilton-Jacobi equation, which would read like $A + |B|^2/2$, but A and the $B^{\alpha i}$ do not live on the same grid. To still get a well-defined constraint, we use the following heuristics: compute finite differences approximations of $\nabla_D \varphi$, then square them, and after that average them from the staggered grids onto the centered one $\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}$. More specifically, let us introduce the average operators $\mathrm{Avg}_\Omega : \mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{st}}} \to \mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}}}$ and $\mathrm{Avg}_D : \mathbb{R}^{\mathcal{G}_D^{\mathrm{st}}} \to \mathbb{R}^{\mathcal{G}_D^{\mathrm{c}}}$ defined by

$$(\mathrm{Avg}_\Omega u)_\xi := \frac{u_{\xi+\tau/2} + u_{\xi-\tau/2}}{2}, \ \forall \xi \in \mathcal{G}_\Omega^{\mathrm{c}}$$
$$(\mathrm{Avg}_D u)_x := \frac{u_{x+\delta/2} + u_{\xi-\delta/2}}{2}, \ \forall x \in \mathcal{G}_D^{\mathrm{c}}$$

Following the rule of thumb, we define $F^{\mathrm{dsc}}(A, B)$ the function whose value is 0 if

$$A + \frac{1}{2}\Bigg\{ \left[\mathrm{Avg}_\Omega \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}}}} \otimes \mathrm{Avg}_D \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}_D^{\mathrm{c}}}} \left(\left(B^{1,1}\right)^2\right)\right]$$
$$+ \left[\mathrm{Avg}_\Omega \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}}}} \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}_D^{\mathrm{c}}}} \otimes \mathrm{Avg}_D \left(\left(B^{1,2}\right)^2\right)\right] + \left[\mathrm{Id}_{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}}}} \otimes \mathrm{Avg}_\Omega \otimes \mathrm{Avg}_D \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}_D^{\mathrm{c}}}} \left(\left(B^{2,1}\right)^2\right)\right]$$
$$+ \left[\mathrm{Id}_{\mathbb{R}^{\mathcal{G}_\Omega^{\mathrm{c}}}} \otimes \mathrm{Avg}_\Omega \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}_D^{\mathrm{c}}}} \otimes \mathrm{Avg}_D \left(\left(B^{2,2}\right)^2\right)\right] \Bigg\} \leqslant 0 \quad (11.1)$$

on $\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times Gc_D \times Gc_D$ and $+\infty$ otherwise. Here $(B^{\alpha i})^2$ is understood component by component. The formula can look complicated, but the idea is simple: if $B^{\alpha i}$ is not on the centered grid in one direction, we average it in this direction. The function F^{dsc} is a discrete analogue of the functional F introduced in the proof of Theorem 8.36. The constraint on the discrete dual problem will be written

$$F^{\mathrm{dsc}}\left[\nabla_\Omega^{\mathrm{dsc}} \cdot \varphi, \left(\partial_{D,i}^{\mathrm{dsc}} \varphi^\alpha\right)_{\alpha, i \in \{1,2\}}\right] = 0,$$

it is a discrete analogue of the Hamilton-Jacobi constraint.

We also need to discretize the objective functional. Recalling that $\boldsymbol{\mu}_b : \partial\Omega \to \mathcal{P}(D)$, by a slight abuse of notation we denote by $\boldsymbol{\mu}_b(\xi^1, \xi^2, x^1, x^2)$ the integral of $\boldsymbol{\mu}_b(\xi^1, \xi^2, \mathrm{d}x^1, \mathrm{d}x^2)$ on the square centered in (x^1, x^2) of side length δ^2. In particular the sum of $\boldsymbol{\mu}_b(\xi^1, \xi^2, x^1, x^2)$ for $(x^1, x^2) \in \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}$ is equal to 1. For $\varphi \in X_{N,M}$ we define,

$$\mathrm{BT}_{\boldsymbol{\mu}_b}^{\mathrm{dsc}}(\varphi) := \tau \sum_{\xi \in \mathcal{G}_\Omega^{\mathrm{c}}} \sum_{(x^1, y^1) \in \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}} \Bigg(\varphi^1_{1,\xi,x^1,x^2} \boldsymbol{\mu}_b(1, \xi, x^1, x^2) - \varphi^1_{0,\xi,x^1,x^2} \boldsymbol{\mu}_b(0, \xi, x^1, x^2)$$
$$+ \varphi^2_{\xi,1,x^1,x^2} \boldsymbol{\mu}_b(\xi, 1, x^1, x^2) - \varphi^2_{\xi,0,x^1,x^2} \boldsymbol{\mu}_b(\xi, 0, x^1, x^2) \Bigg),$$

which is just a linear form on $X_{N,M}$ looking like $\mathrm{BT}_{\boldsymbol{\mu}_b}$. By an abuse of notation, we denote by $\mathrm{BT}_{\boldsymbol{\mu}_b}^{\mathrm{dsc}}$ both the linear form on $X_{N,M}$ and the element of $X_{N,M}$ representing this linear form with the help of the scalar product $\langle\ ,\ \rangle_X$.

Definition 11.1. *We define the discrete (i.e. finite-dimensional) dual problem as*

$$\max_{\varphi, A, B} \left\{ \mathrm{BT}_{\boldsymbol{\mu}_b}^{dsc}(\varphi) - F^{dsc}(A, B) \ : \varphi \in X_{N,M}, \ (A, B) \in Y_{N,M} \ \text{and } \mathcal{D}^{dsc}\varphi = (A, B) \right\}$$

The variables (A, B) look superfluous for the moment but they play an important role for the actual resolution of the problem. The discrete dual problem is a *convex* problem. Moreover, it reads as a *quadratically constrained linear program*, more specifically as the quadratic part is semi-definite positive, it could be rewritten as a *second-order cone program.*

Proposition 11.2. *For any $N, M \geqslant 2$, there exists at least one solution to the discrete dual problem.*

Proof. Once (A, B) has been eliminated, the constraint on φ can be written as a intersection of quadratic constraints, but the set of admissible φ is not compact. Using $\varphi = 0$ as a competitor, we know that the value of the dual problem is positive.

Let $(\varphi_n)_{n\in\mathbb{N}}$ a maximizing sequence with $\mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\varphi_n) \geqslant 0$ for every n. For a fixed n, we define $\bar{\varphi_n}$ by

$$(\bar{\varphi_n})_{\xi^1,\xi^2,x^1,x^2} := \frac{1}{M^2} \sum_{(y^1,y^2)\in\mathcal{G}^{\mathrm{c}}_D\times\mathcal{G}^{\mathrm{c}}_D} (\varphi_n)_{\xi^2,\xi^2,y^1,y^2}.$$

The function $\bar{\varphi}_n$ is defined on the same grid as φ_n, but does not depend on variables in D. We write $\tilde{\varphi}_n = \varphi_n - \bar{\varphi}_n$. Thus, we can decompose $\varphi_n = \tilde{\varphi}_n + \bar{\varphi}_n$ and $\tilde{\varphi}_n$ has 0 mean in the sense that

$$\sum_{(x^1,x^2)\in\mathcal{G}^{\mathrm{c}}_D\times\mathcal{G}^{\mathrm{c}}_D} (\tilde{\varphi}_n)_{\xi^1,\xi^2,x^1,x^2} = 0. \tag{11.2}$$

As $\bar{\varphi}_n$ does not depend on the variables in D and given the normalization of $\boldsymbol{\mu}_b$,

$$\tau^2 \sum_{(\xi^1,\xi^2,x^1,x^2)\in\mathcal{G}^{\mathrm{c}}_\Omega\times\mathcal{G}^{\mathrm{c}}_\Omega\times\mathcal{G}^{\mathrm{c}}_D\times\mathcal{G}^{\mathrm{c}}_D} \left(\nabla^{\mathrm{dsc}}_\Omega \cdot \bar{\varphi}_n\right)_{\xi^1,\xi^2,x^1,x^2} = \mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\bar{\varphi}_n) \geqslant -\mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\tilde{\varphi}_n),$$

where the inequality comes from $\mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\varphi_n) \geqslant 0$. Hence, summing over the grid $\mathcal{G}^{\mathrm{c}}_\Omega\times\mathcal{G}^{\mathrm{c}}_\Omega\times\mathcal{G}^{\mathrm{c}}_D\times\mathcal{G}^{\mathrm{c}}_D$ the inequalities coming from $F^{\mathrm{dsc}}(\mathcal{D}^{\mathrm{dsc}}\varphi_n) = 0$, taking in account that $\bar{\varphi}_n$ does not depend on the variables (x^1, x^2),

$$-\mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\tilde{\varphi}_n) + \tau^2 \sum_{(\xi^1,\xi^2,x^1,x^2)\in\mathcal{G}^{\mathrm{c}}_\Omega\times\mathcal{G}^{\mathrm{c}}_\Omega\times\mathcal{G}^{\mathrm{c}}_D\times\mathcal{G}^{\mathrm{c}}_D} \left(\nabla^{\mathrm{dsc}}_\Omega \cdot \tilde{\varphi}_n\right) + Q(\tilde{\varphi}_n) \leqslant 0.$$

where the last term of the sum is the quadratic form Q defined by

$$\begin{aligned} Q(\varphi) := \frac{1}{2}\Bigg\{ &\left[\mathrm{Avg}_\Omega \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_\Omega}} \otimes \mathrm{Avg}_D \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_D}} \left(\left(\partial^{\mathrm{dsc}}_{D,1}\varphi^1\right)^2\right)\right] \\ &+ \left[\mathrm{Avg}_\Omega \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_\Omega}} \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_D}} \otimes \mathrm{Avg}_D \left(\left(\partial^{\mathrm{dsc}}_{D,2}\varphi^1\right)^2\right)\right] \\ &+ \left[\mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_\Omega}} \otimes \mathrm{Avg}_\Omega \otimes \mathrm{Avg}_D \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_D}} \left(\left(\partial^{\mathrm{dsc}}_{D,1}\varphi^2\right)^2\right)\right] \\ &+ \left[\mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_\Omega}} \otimes \mathrm{Avg}_\Omega \otimes \mathrm{Id}_{\mathbb{R}^{\mathcal{G}^{\mathrm{c}}_D}} \otimes \mathrm{Avg}_D \left(\left(\partial^{\mathrm{dsc}}_{D,2}\varphi^2\right)^2\right)\right] \Bigg\} \end{aligned}$$

The quadratic form Q is definite positive over the set of $\tilde{\varphi}_n$ satisfying (11.2). (Note that it explains why it was important to square the derivatives of φ *before* averaging). On the other hand, the first two terms in the inequality above are linear in $\tilde{\varphi}_n$. Hence, we deduce that $\tilde{\varphi}_n$ is bounded, thus converges up to extraction.

Then we want to say something about $\bar{\varphi}_n$. Thanks to the convergence of $\tilde{\varphi}_n$ and the constraint $F^{\mathrm{dsc}}(\mathcal{D}^{\mathrm{dsc}}\varphi_n) = 0$, it is clear that $\nabla_\Omega^{\mathrm{dsc}} \cdot \bar{\varphi}_n$ is bounded from above uniformly on the grid $\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}$. On the other hand, the sum of such $\nabla_\Omega^{\mathrm{dsc}} \cdot \bar{\varphi}_n$ on the same grid is larger than $-\mathrm{BT}_{\boldsymbol{\mu}_b}(\tilde{\varphi}_n)$, hence bounded from below. Thus, up to extraction, $\nabla_\Omega^{\mathrm{dsc}} \cdot \bar{\varphi}_n$ converges for every point of the grid $\mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_\Omega^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}} \times \mathcal{G}_D^{\mathrm{c}}$.

Let $\hat{\varphi}_n$ be the projection of $\bar{\varphi}_n$ on the orthogonal of the kernel of $(\nabla_\Omega^{\mathrm{dsc}} \cdot)$. By the previous observation, $\hat{\varphi}_n$ converges to some limit up to extraction. The sequence $\hat{\varphi}_n + \tilde{\varphi}_n$ satisfies

$$\begin{aligned}\mathrm{BT}_{\boldsymbol{\mu}_b}(\hat{\varphi}_n + \tilde{\varphi}_n) - F^{\mathrm{dsc}}(\mathcal{D}^{\mathrm{dsc}}(\hat{\varphi}_n + \tilde{\varphi}_n)) &= \mathrm{BT}_{\boldsymbol{\mu}_b}(\bar{\varphi}_n + \tilde{\varphi}_n) - F^{\mathrm{dsc}}(\mathcal{D}^{\mathrm{dsc}}(\tilde{\varphi}_n)) \\ &= \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi_n) - F^{\mathrm{dsc}}(\mathcal{D}^{\mathrm{dsc}}(\varphi_n))\end{aligned}$$

because $\nabla_\Omega^{\mathrm{dsc}} \cdot \hat{\varphi}_n = \nabla_\Omega^{\mathrm{dsc}} \cdot \bar{\varphi}_n$, and is convergent up to extraction. Its limit is nothing else, as φ_n is a maximizing sequence, than a solution of the dual problem. □

The next step is to derive the dual of this discrete problem and observe that it looks like the continuous primal problem. Moreover, as we will use a primal-dual algorithm to efficiently solve this convex optimization problem, we will need an expression of the Lagrangian at some point. The derivation of the dual is very similar to what was done with the formal inf − sup exchange in the introductory Chapter 7.

We introduce $(\boldsymbol{\mu}, (\mathbf{E}^{\alpha i})_{\alpha, i \in \{1,2\}}) \in Y_{N,M}$ Lagrange multipliers for the constraint $A = \nabla_\Omega^{\mathrm{dsc}} \cdot \varphi$ and $B^{\alpha i} = \partial_i^{\mathrm{dsc}} \varphi^\alpha$. The Lagrangian of the problem can be written

$$L(\varphi, A, B, \boldsymbol{\mu}, \mathbf{E}) := \mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\varphi) - F^{\mathrm{dsc}}(A, B) + \langle (\boldsymbol{\mu}, \mathbf{E}), (A, B) - \mathcal{D}^{\mathrm{dsc}} \varphi \rangle_Y. \tag{11.3}$$

The objective value of the discrete dual problem is recovered by minimizing the Lagrangian in $\boldsymbol{\mu}$ and $\mathbf{E}$. To get the dual of the dual discrete problem, we first maximize in φ and A, B. Maximization in the (now unconstrained) variable φ is straightforward as the Lagrangian is linear in φ. It can be written abstractly

$$(\mathcal{D}^{\mathrm{dsc}})^\top \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{E} \end{pmatrix} = \mathrm{BT}_{\boldsymbol{\mu}_b}, \tag{11.4}$$

where $(\mathcal{D}^{\mathrm{dsc}})^\top$ is the adjoint of the operator $\mathcal{D}^{\mathrm{dsc}}$. This equation, stating the equality of two vectors in $X_{N,M}$, is nothing else than a discrete version of the (generalized) continuity equation; but we will not try to write it explicitly. Then, for the maximization in (A, B), we know that we end up by definition with $(F^{\mathrm{dsc}})^\star(\boldsymbol{\mu}, \mathbf{E})$ the Fenchel transform of F^{dsc}.

Proposition 11.3. *The value of the discrete dual problem is equal to*

$$\min_{\boldsymbol{\mu}, \mathbf{E}} \left\{ (F^{dsc})^\star(\boldsymbol{\mu}, \mathbf{E}) \ : \ (\boldsymbol{\mu}, \mathbf{E}) \in Y_{N,M} \text{ and } (\mathcal{D}^{dsc})^\top \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{E} \end{pmatrix} = \mathrm{BT}_{\boldsymbol{\mu}_b} \right\}.$$

The latter problem will be called the discrete primal problem

Proof. As the problem is finite-dimensional, standard arguments about convex duality guarantee the existence of a solution to the discrete dual problem and the absence of duality gap, see [BV04, Chapter 5]. □

Actually, Theorem 8.36 is the infinite-dimensional analogue of the proposition above. We will *not* provide an explicit expression for $(F^{\mathrm{dsc}})^\star(\boldsymbol{\mu}, \mathbf{E})$. It leads to some expression of the form $\sum \frac{|\mathbf{E}|^2}{2\tilde{\boldsymbol{\mu}}}$, where $\tilde{\boldsymbol{\mu}}$ is a linear averaging of the $\boldsymbol{\mu}$ defined on grids on which the components of $\mathbf{E}$

are defined, which means that the discrete primal problem really looks like the continuous one. However, the precise expression of $(F^{\text{dsc}})^\star(\boldsymbol{\mu}, \mathbf{E})$ is quite heavy and will not be relevant in the sequel, but we mention that it is reminiscent of the formulas of Maas and Gigli [Maa11, GM13] about optimal transport on graphs. In any case, $\boldsymbol{\mu}_{\xi^1,\xi^2,x^1,x^2}$ can be interpreted as the mass given by $\boldsymbol{\mu}(\xi^1,\xi^2)$ to the square of center (x^1,x^2) and side length δ^2.

We do not know whether there is uniqueness in the discrete primal problem. However, in practice, when we used the iterative algorithm described in the next section to solve these convex problems, we have found that the solution $\boldsymbol{\mu}$ does not depend on the initial guess. On the other hand, for the discrete dual problem, as the constraint on $\mathcal{D}^{\text{dsc}}\varphi$ will be saturated only where $\boldsymbol{\mu}$ is strictly positive, it is highly unlikely for uniqueness to hold.

If we summarize, we have two convex finite dimensional problems in duality, which look like the continuous ones. As discussed later in Chapter 12, we do not know if the discrete problems converge, when $N, M \to +\infty$, to the continuous ones. However, we can prove these easy properties on the solution $\boldsymbol{\mu}$ of the discrete primal problem.

Proposition 11.4. *Let $(\boldsymbol{\mu}, \mathbf{E})$ a solution of the discrete primal problem. Then $\boldsymbol{\mu} \geqslant 0$ and for all $(\xi^1,\xi^2) \in \mathcal{G}^c_\Omega \times \mathcal{G}^c_\Omega$,*

$$\sum_{(x^1,x^2)\in\mathcal{G}^c_D\times\mathcal{G}^c_D} \boldsymbol{\mu}_{\xi^1,\xi^2,x^1,x^2} = 1.$$

In other words, positivity and preservation of the mass hold.

Proof. If $\boldsymbol{\mu}_{\xi^1,\xi^2,x^1,x^2} < 0$ at some $(\xi^1,\xi^2,x^1,x^2) \in \mathcal{G}^c_\Omega \times \mathcal{G}^c_\Omega \times \mathcal{G}^c_D \times \mathcal{G}^c_D$, then it is enough to take A_{ξ^1,ξ^2,x^1,x^2} very negative and large (in absolute value) to conclude that the Lagrangian $L(\varphi, A, B, \boldsymbol{\mu}, \mathbf{E})$ goes to $+\infty$, which would say that the value of the discrete primal problem is $+\infty$, hence a contradiction.

For the preservation of mass, we use the discrete version of the continuity equation. Namely, we fix $(\xi^1,\xi^2) \in \mathcal{G}^c_\Omega \times \mathcal{G}^c_\Omega$ and we take $\varphi^1_{\xi^1+\tau/2,\xi^2,x^1,x^2} = 1$ for all $(x^1,x^2) \in \mathcal{G}^c_D \times \mathcal{G}^c_D$ while all other values of φ^1 are set to 0. All the values of φ^2 are set to 0. Because of the discrete continuity equation,

$$\tau^2 \langle \nabla^{\text{dsc}}_\Omega \cdot \varphi, \boldsymbol{\mu} \rangle = \text{BT}_{\boldsymbol{\mu}_b}(\varphi).$$

Developing the l.h.s. and using the exact expression for the r.h.s. (and the normalization for $\boldsymbol{\mu}_b$),

$$\sum_{(x^1,x^2)\in\mathcal{G}^c_D\times\mathcal{G}^c_D} \left(\boldsymbol{\mu}_{\xi^1+\tau,\xi^2,x^1,x^2} - \boldsymbol{\mu}_{\xi^1,\xi^2,x^1,x^2} \right) = \begin{cases} 1 & \text{if } \xi^1 + \tau = 1, \\ 0 & \text{otherwise.} \end{cases}$$

From this set of equations, it is not difficult to see that mass is preserved along each line parallel to the second component of Ω, which eventually proves the claim. □

Let us conclude this section by insisting that the grid on which $\boldsymbol{\mu}$ is defined is *not* the same than the one on which the boundary conditions are defined. Indeed, $\boldsymbol{\mu}$ is defined on the grid $\mathcal{G}^c_\Omega \times \mathcal{G}^c_\Omega$ which never actually touches the boundary $\partial\Omega$ of Ω, while the boundary conditions $\boldsymbol{\mu}_b$ live on $\partial\Omega$. The meaning given to the boundary conditions is only through (11.4), namely the values of $\boldsymbol{\mu}_b$ on $\partial\Omega$ and $\boldsymbol{\mu}$ close to $\partial\Omega$ are coupled with the momentum $\mathbf{E}$, which is itself defined up to the boundary of Ω.

11.2 Effective resolution

In practice, the algorithm that we have implemented computes a saddle point of the Lagrangian (11.3). The input of the algorithm is the boundary data $\boldsymbol{\mu}_b$, and the outputs are $\varphi, A, B, \boldsymbol{\mu}$ and $\mathbf{E}$. To compute a saddle point, we first *augment* the Lagragian with a quadratic penalization and then use the *Alternative Direction Method of Multipliers* (ADMM), see [BPC+11]. Notice that this is very similar to what happens for computation of geodesics in the Wasserstein space, see [BB00] or [San15, Section 6.1]. But before describing the method, let us first emphasize that we augment the number of unknowns in the Lagrangian, with no impact on the saddle points.

Instead of storing only $\nabla^{\mathrm{dsc}}_{\Omega} \cdot \varphi$, we store the whole four derivatives w.r.t. variables in Ω of φ. More precisely, we denote by $\partial^{\mathrm{dsc}}_{\Omega} : \mathbb{R}^{\mathcal{G}^{\mathrm{st}}_{\Omega}} \to \mathbb{R}^{\mathcal{G}^{\mathrm{c}}_{\Omega}}$ the finite difference operator Ω and $(\partial^{\mathrm{dsc}}_{\Omega})^{\top} : \mathbb{R}^{\mathcal{G}^{\mathrm{c}}_{\Omega}} \to \mathbb{R}^{\mathcal{G}^{\mathrm{st}}_{\Omega}}$ its adjoint, which is almost the opposite of the previous one, except for what happens at the boundary. By definition, in $A = (A^{\alpha\beta})_{\alpha,\beta\in\{1,2\}}$, we store

$$\begin{cases} A^{1,1} &= \partial^{\mathrm{dsc}}_{\Omega,1}\varphi^1, \\ A^{1,2} &= -(\partial^{\mathrm{dsc}}_{\Omega,1})^{\top}\varphi^2, \\ A^{2,1} &= -(\partial^{\mathrm{dsc}}_{\Omega,2})^{\top}\varphi^1, \\ A^{2,2} &= \partial^{\mathrm{dsc}}_{\Omega,2}\varphi^2, \end{cases}$$

Notice that, given the previous definitions, $\nabla^{\mathrm{dsc}}_{\Omega} \cdot \varphi = A^{1,1} + A^{2,2}$. For $A^{1,2}$ and $A^{2,1}$ we use rather the adjoint of $\partial^{\mathrm{dsc}}_{\Omega}$ because of the grid on which φ is defined. As A has four components, the Lagrange mutliplier $\boldsymbol{\mu}$ will be a vector with four different components $(\boldsymbol{\mu}^{\alpha\beta})_{\alpha,\beta\in\{1,2\}}$.

As far as B is concerned, we will split the variables. Indeed, each value of $B^{\alpha i}_{\xi^2,\xi^2,x^1,x^2}$ appears in four different inequalities involved in the definition of F^{dsc} in (11.1). So each value of $B^{\alpha i}$ will be stored four times, in such a way that each component of B is constrained to a unique inequality, to which we of course add the equality constraints $B^{\alpha i} = \partial^{\mathrm{dsc}}_{D,i}\varphi^{\alpha}$. Automatically, the number of Lagrange multipliers, i.e. the dimension of the vector $\mathbf{E}$, is multiplied by 4.

We will define by $\mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}}\varphi$ the "augmented" differentiation operator by

$$\mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}}\varphi = \begin{pmatrix} \partial^{\mathrm{dsc}}_{\Omega,1}\varphi^1 \\ -(\partial^{\mathrm{dsc}}_{\Omega,1})^{\top}\varphi^2 \\ -(\partial^{\mathrm{dsc}}_{\Omega,2})^{\top}\varphi^1 \\ \partial^{\mathrm{dsc}}_{\Omega,2}\varphi^2 \\ \left(\partial^{\mathrm{dsc}}_{D,i}\varphi^{\alpha}\right)_{\alpha,i\in\{1,2\}} \end{pmatrix},$$

where in the last row it is tacitly assumed that each component of $\partial^{\mathrm{dsc}}_{D,i}\varphi^{\alpha}$ is duplicated four times. Notice that $\mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}}\varphi$ just looks like a gradient of φ w.r.t. all the variables.

Then, the Lagrangian (11.3) can be rewritten as

$$L(\varphi, A, B, \boldsymbol{\mu}, \mathbf{E}) = \mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\varphi) - F^{\mathrm{dsc}}(A^{11} + A^{22}, B) + \langle(\boldsymbol{\mu}, \mathbf{E}), (A, B) - \mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}}\varphi\rangle_Y.$$

Although we have increased the number of variables, it is straightforward to see that if $(\varphi, A, B, \boldsymbol{\mu}, \mathbf{E})$ is a saddle point of this Lagrangian then $\boldsymbol{\mu}^{11} = \boldsymbol{\mu}^{22}$ so that we really recover a saddle point of the previous Lagrangian. Following [BPC+11], we augment the Lagrangian by adding a quadratic penalization. Specifically, we set, for $r > 0$,

$$L^{\mathrm{aug}}(\varphi, A, B, \boldsymbol{\mu}, \mathbf{E}) = \mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\varphi) - F^{\mathrm{dsc}}(A^{11} + A^{22}, B) + \langle(\boldsymbol{\mu}, \mathbf{E}), (A, B) - \mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}}\varphi\rangle_Y - \frac{r}{2}\left\|\mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}}\varphi - (A, B)\right\|^2_Y. \quad (11.5)$$

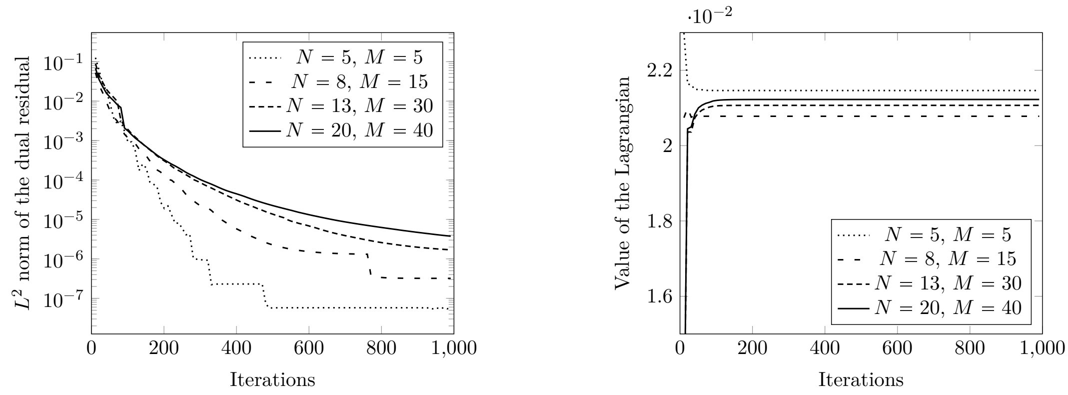

Figure 11.2: Convergence of the ADMM iterations. Left: evolution of the L^2 norm of the dual residual as a function of the number of iterations. Right: evolution of the value of the Lagrangian as a function of the number of iterations. The dual residual, as defined in [BPC$^+$11, Section 3.3], corresponds roughly to $r((\mathcal{D}^{\text{dsc}}_{\text{aug}})^\top(\boldsymbol{\mu}, \mathbf{E}) - \text{BT}^{\text{dsc}}_{\boldsymbol{\mu}_b})$. The jumps in its values are due to an update of the augmenting parameter r. We plot these quantities for different value of the discretization parameters N, M, but we the same boundary conditions: those corresponding to a family of elliptically contoured distributions as in Figure 11.3.

This augmented Lagrangian has the same saddle point than the previous one.

Now, the algorithm consists in the iteration of the following steps. Given $\varphi, A, B, \boldsymbol{\mu}, \mathbf{E}$,

1. Replace φ by the one that maximizes $L^{\text{aug}}(\cdot, A, B, \boldsymbol{\mu}, \mathbf{E})$.
2. Replace (A, B) by the ones that maximize $L^{\text{aug}}(\varphi, \cdot, \cdot, \boldsymbol{\mu}, \mathbf{E})$.
3. Do the dual update $(\boldsymbol{\mu}, \mathbf{E}) \leftarrow (\boldsymbol{\mu}, \mathbf{E}) - r\left[(A, B) - \mathcal{D}^{\text{dsc}}_{\text{aug}}\varphi\right]$.

We emphasize that the step used in the dual update is precisely r the augmentation parameter. In practice, the value of this parameter was tuned dynamically during the iterations according to the heuristic rule of [BPC$^+$11, Section 3.4.1]. As the problem is finite-dimensional, convergence to a saddle point is guaranteed [BPC$^+$11, Section 3.2]. Actually, if the scalar products are well scaled, we noticed experimentally that the number of ADMM iterations to reach convergence was quite independent on the sizes N, M of the grids, see Figure 11.2 (though we would agree that this is not so apparent, but we have trouble to do computation with more than a few dozens discretization points per dimension). At least for the case of geodesics, this feature comes from the fact that, in the limit $N, M \to +\infty$, the ADMM iterations still make sense and converge also holds in the infinite-dimensional setting [Gui03, Hug16]. Let us now detail in practice how the different steps are handled. As the reader can see below, each step of the ADMM has a complexity of $O(N^2M^2\log(NM))$.

Maximization in φ Once the other variables are fixed, the augmented Lagrangian L^{aug} is a quadratic function of φ. Hence its maximization amounts to invert a linear system, whose matrix, namely $(\mathcal{D}^{\text{dsc}}_{\text{aug}})^\top\mathcal{D}^{\text{dsc}}_{\text{aug}}$ is the same for every iteration. Notice that this matrix $(\mathcal{D}^{\text{dsc}}_{\text{aug}})^\top\mathcal{D}^{\text{dsc}}_{\text{aug}}$ has a kernel of dimension 1, which corresponds to functions that are constant. Hence, once we impose that φ has 0-mean (i.e. that φ lives in the space orthogonal to the kernel), inverting the linear system is a matter of linear algebra.

More specifically, $(\mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}})^\top \mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}}$ is a finite difference discretization of the full Laplacian $\Delta_\Omega + \Delta_D$. To invert this matrix, we leverage the fact that we work on Cartesian grids: we use a Discrete Cosine Transform on $\mathcal{G}^{\mathrm{c}}_\Omega, \mathcal{G}^{\mathrm{st}}_\Omega$ and a Fast Fourier Transform on $\mathcal{G}^{\mathrm{c}}_D, \mathcal{G}^{\mathrm{st}}_D$. Notice that for the grids in Ω we work with *cosine* transforms because of the boundary conditions. Provided we use efficient routines for these transforms, the overall complexity of this step is $O(N^2 M^2 \log(NM))$.

Maximization in (A, B) Once the other variables are fixed, maximizing the augmented Lagrangian L^{aug} in (A, B) amounts to project a vector, namely $\mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}} \varphi + \frac{1}{r}(\boldsymbol{\mu}, \mathbf{E})$, onto the set of (A, B) satisfying the constraint $F^{\mathrm{dsc}}(A^{11} + A^{22}, B) < +\infty$. In particular, notice that A^{12} and A^{21} are not submitted to any constraints, hence the projection is straightforward. On the other hand, as each component of $(A^{11} + A^{22}, B)$ is subject to a unique inequality constraint (hence the interest of the splitting of variables), we have to solve for each point of the grid $\mathcal{G}^{\mathrm{c}}_\Omega \times \mathcal{G}^{\mathrm{c}}_\Omega \times \mathcal{G}^{\mathrm{c}}_D \times \mathcal{G}^{\mathrm{c}}_D$ a problem of the type

$$\min_{a,b} \left\{ \frac{1}{2}|a - a_0|^2 + \frac{1}{2}|b - b_0|^2 \ : \ (a_0, b_0) \text{ given and } a + \frac{|b|^2}{2} \leqslant 0 \right\},$$

where a, a_0 are scalar and b, b_0 are vectors. It is well known (see for instance [PPO14, Proposition 1]) that solving this problem amounts to find the root of a third-order polynomial. The latter search was performed using Newton's method. Provided all the problems are solved in parallel, the overall complexity of this step is $O(N^2 M^2)$.

Dual update The dual update $(\boldsymbol{\mu}, \mathbf{E}) \leftarrow (\boldsymbol{\mu}, \mathbf{E}) - r\left[(A, B) - \mathcal{D}^{\mathrm{dsc}}_{\mathrm{aug}} \varphi\right]$ just amounts to subtract some arrays, the overall complexity of this step is $O(N^2 M^2)$.

11.3 Examples

We present in this section actual computations of harmonic mappings by the discretization and algorithm described above. We insist that we have no proof of convergence if one refines the discretization. The only hint in this direction is that the outputs are visually plausible, but we consider our method as a way to provide illustrations rather than a solid and guaranteed numerical discretization. Notice that it would be hard to compare the computed solution with a theoretical one. Indeed, the only case where we have explicit formulas, namely Section 10.4, does not fit the geometry that we can reach with our finite difference discretization: with our current implementation, we can only handle the source space Ω being a square, not a disk. It might be possible to propose finite difference discretizations for more complicated geometries of the source space, but it would for sure require heavy changes in the implementation.

Our method is quite slow. Indeed, the number of ADMM iterations required to reach convergence (setting a primal and dual residual lower than 10^{-4}, where these residuals are defined in [BPC+11, Section 3.3]) is of the order of 10^3 and the time per iteration, for instance for $N = 14$ and $M = 40$ is 4 seconds. Hence, the total time required to reach convergence can be of the order of several hours.

To plot the results, we have displayed N^2 copies of D. Each copy of D corresponds to a point $(\xi^1, \xi^2) \in \mathcal{G}^{\mathrm{c}}_\Omega \times \mathcal{G}^{\mathrm{c}}_\Omega$, on which is represented the measure with density $\boldsymbol{\mu}_{\xi^1, \xi^2}$. As pointed out earlier, the boundary conditions $\boldsymbol{\mu}_b$ are *not* on the same grid as the final solution $\boldsymbol{\mu}$. However, at least visually, there seems to be a good agreement between the boundary values $\boldsymbol{\mu}_b$ and the values of $\boldsymbol{\mu}$ close to the boundary.

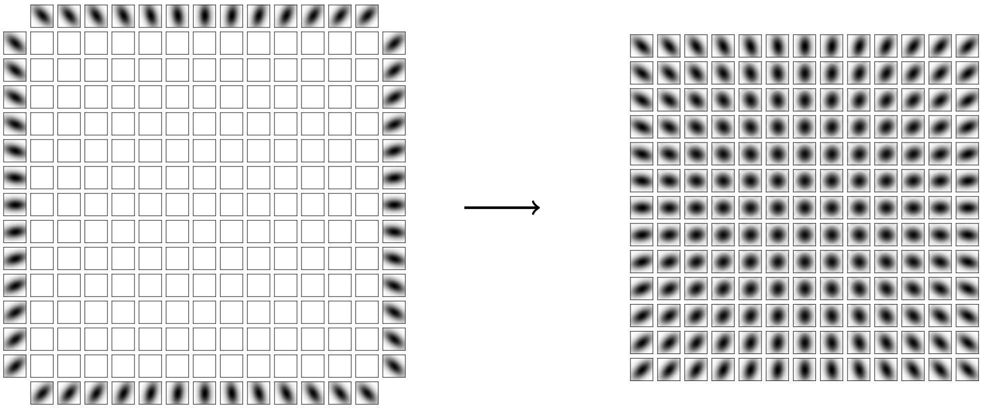

Figure 11.3: Left: boundary conditions, inspired from the example described in Section 10.4. Right: solution of the Dirichlet problem with $N = 13,\ M = 30$.

Constant boundary conditions If $\boldsymbol{\mu}_b(\xi)$ does not depend on $\xi \in \partial\Omega$, then our algorithm indeed converges to a function which is constant over Ω and takes the same value than $\boldsymbol{\mu}_b$. The associated value of the Lagrangian is 0.

Family of elliptically contoured distributions We have tried to mimic the situation described in Section 10.4. As underlined before, we cannot really do it as Ω is the unit square and not the unit disk, and the target space is the Wasserstein space built over the torus. Nevertheless, as the reader can see in Figure 11.3, we still observe the symmetry predicted in Section 10.4 and the fact that, close to the center of Ω, the measures are more isotropic than at the boundary.

Interpolation between shapes Figure 11.4 is built according to the following process. For each corner of Ω, we have selected a probability measure which is just the normalized indicator of a shape. Then we have computed the geodesics in the Wasserstein space between each of these shapes, and used the geodesics as boundary data $\boldsymbol{\mu}_b$ (i.e. on each edge of the square we put a geodesic). Eventually, we have solved the Dirichlet problem.

We have chosen this example because of the similarity with [SDGP$^+$15, Figure 12] which is reproduced in Figure 11.5. In the latter, for each point of Ω, one computes the barycenter with bilinear weights of the shapes in the corners. Hence the edges of Figure 11.5 coincide with the boundary conditions of our figure. However, as explained in Section 10.4, the value of the solution of the Dirichlet problem at one point cannot be expressed as the (weighted) barycenter of the values at the boundary.

Hence, though visually similar, the interpolation in Figure 11.4 and 11.5 are very likely to differ. However, we couldn't prove the discrepancy analytically, because of the more complicated geometry than in Section 10.4 and the absence of a closed formula for the Wasserstein distance when we no longer face measures valued in a family of elliptically contoured distributions. Identifying the discrepancy numerically is likely to be challenging: as seen in Section 10.4, we expect it to be small, hence we would not be able to distinguish between it and the errors due to the discretization. Moreover, Figure 11.4 and Figure 11.5 are computed by different means (the method of the present chapter for the former, entropic regularization for the latter), which would make a precise comparison even more delicate.

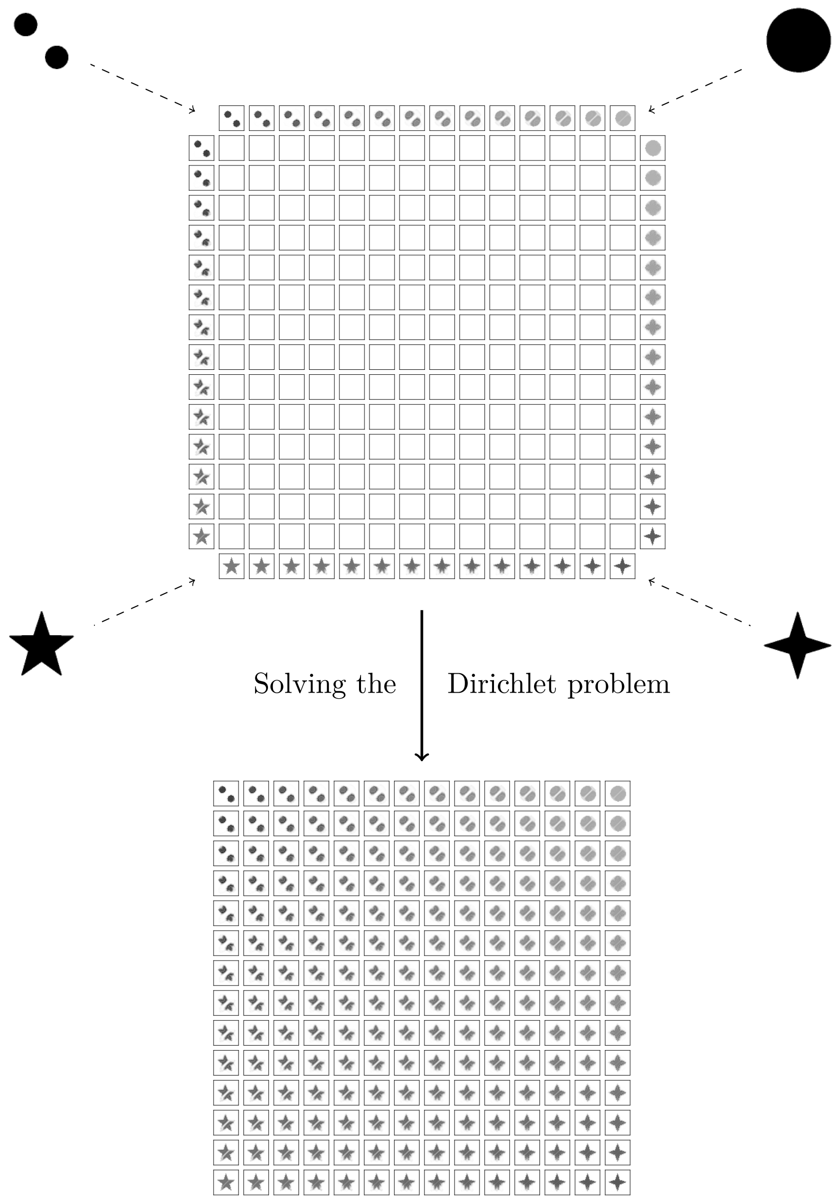

Figure 11.4: Top: the boundary conditions are geodesics in the Wasserstein space between the shapes displayed in the corners. The one dimensional geodesics were computed by adapting the method described in this chapter to the simpler case where the source space is a segment. Bottom: solution of the Dirichlet problem with $N = 14,\ M = 40$.

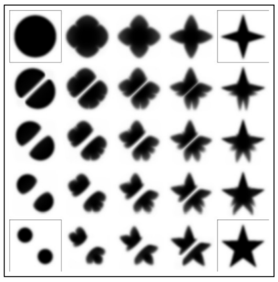

Figure 11.5: Figure taken from [SDGP$^+$15, Figure 12] with permission of the authors. On each corner of the square there is a given shape. Then, on points of the square, the Wasserstein barycenter with bilinear weights between the four probability distributions on the corners is computed (with the help of entropic regularization) and displayed.

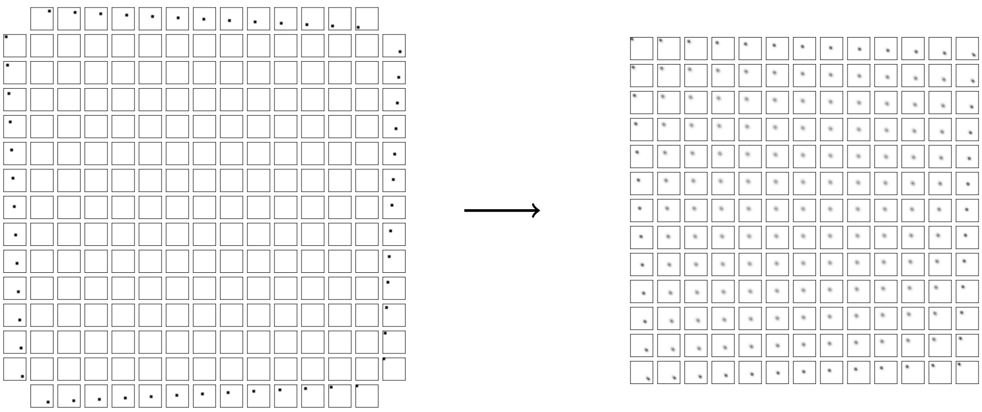

Figure 11.6: Left: boundary conditions which are supposed to look like Dirac masses. In practice, each probability distribution on the boundary is non zero on $4 = 2 \times 2$ points of the grid $\mathcal{G}_D^c \times \mathcal{G}_D^c$. Right: solution of the Dirichlet problem with $N = 13,\ M = 17$. As one can see, the solution of the Dirichlet problem takes values in only very peaked probability distributions.

Dirac masses As a test case, we have put for boundary values Dirac masses, i.e. a data $\boldsymbol{\mu}_b(\xi^1, \xi^2)$ is 0 for all but a few vertices of $\mathcal{G}_D^c \times \mathcal{G}_D^c$. As mentioned before, the solution of the (continuous) Dirichlet problem is a mapping valued in the set of Dirac masses. When we put such boundary conditions in our algorithm, as one can see in Figure 11.6, we observe that $\boldsymbol{\mu}$ stays very peaked in the middle, which reproduces a feature of the continuous case. This test case is somehow an extreme one: we use a PDE formulation of our problem but we test it on very singular measures. Nevertheless, we recover a result which is visually satisfactory. We mention that we have tried other discretizations which gave worse output on this kind of test, and that the present method, which we chose in the end, was the one performing the better on this example.

Chapter 12

Perspectives and open questions

In this chapter, we would like to present questions that we have faced but left unanswered and give some possible directions for future research. For most of these questions, we have tried the standard approaches but they were not conclusive, and their resolution are likely to need new ideas which are not present in this manuscript. In the rest of this chapter, we will be quite sloppy with regularity issues and most of the computations will be purely formal.

12.1 Uniqueness in the primal problem, existence in the dual problem

Even some very natural questions about the primal and dual formulation of the Dirichlet problem are not answered.

Question 12.1. *Under which assumptions can one guarantee the uniqueness of the solution* $\boldsymbol{\mu}$ *to the Dirichlet problem (Definition 8.30)?*

Question 12.2. *In which functional space, and in which sense, can one find a* φ *which realizes the supremum in the dual formulation of the Dirichlet problem, as defined in Theorem 8.36?*

Existence in the dual problem We will start with the second question. If the source space Ω is a segment, the constraint to which φ is submitted is the Hamilton Jacobi equation, namely

$$\partial_t \varphi + \frac{1}{2} |\nabla \varphi|^2 \leqslant 0.$$

Usually, one relaxes the set of admissible φ by admitting continuous functions as competitors, and the Hamilton Jacobi equation is understood in the viscosity sense. The key point is that there exists an explicit expression of $\varphi(t = 1, \cdot)$ as a function of $\varphi(t = 0, \cdot)$, namely the Hopf-Lax formula. For our vector-valued unknown φ, we don't know what would be the meaning of our constraint in the viscosity sense, and we are unaware of any explicit formula related to it. On the other hand, when one works on Mean Field Games, as stated for instance in Definition 5.2, the Hamilton Jacobi equation is rather understood in the distributional sense. Whatever meaning we choose, it is important that, provided φ satisfies the constraint, and $\boldsymbol{\mu}$ has boundary conditions $\boldsymbol{\mu}_b : \partial\Omega \to \mathcal{P}(D)$,

$$\mathrm{Dir}(\boldsymbol{\mu}) \geqslant \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi).$$

Such a computation is usually justified with a regularization procedure, where there is an interplay between the meaning given to the Hamilton Jacobi equation and the regularity of the measures

$\boldsymbol{\mu}_b$ at the boundary. Until now, we have not really studied this interplay, but we think that, provided the $\boldsymbol{\mu}_b$ are regular enough (with a density, maybe bounded from below), one can allow a quite loose meaning (maybe distributional sense) to our Hamilton Jacobi constraint.

Provided we know the meaning of the Hamilton Jacobi constraint, we still need to prove compactness of a maximizing sequence (maybe up to transformation like the double convexification trick in the case of geodesics [Vil03, Section 2.1]). We have proved in Proposition 11.2 such a compactness result for a discretized version of our dual problem. However, this proof really relies on the fact that we work in a finite-dimensional setting and the constants used to get compactness blow up when one refines the discretization. We have no idea of any estimate which would lead us to compactness in the infinite dimensional case.

Uniqueness in the primal problem The Dirichlet energy is not strictly convex, hence uniqueness is not automatically guaranteed. However, when Ω is a segment, we know that, provided the values at the boundary are regular (at least one of them has a density w.r.t. $\mathcal{L}_D$ for instance), uniqueness holds. But the proof of such a result relies on a static formulation which is no longer available if the dimension of Ω is larger than 1.

We want to highlight, as already observed and used in the proof of Theorem 10.10 that all solutions should share the same velocity field. Indeed, let $\boldsymbol{\mu}, \varphi$ be admissible competitors for the primal and dual problems respectively, and $\mathbf{v}$ the tangent velocity field to $\boldsymbol{\mu}$. The dual gap can be written

$$\begin{aligned}
\mathrm{Dir}(\boldsymbol{\mu}) - \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi) &= \iint\limits_{\Omega\times D} \frac{1}{2}|\mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu} - \iint\limits_{\Omega\times D} (\nabla_\Omega \cdot \varphi + \nabla_D \varphi \cdot \mathbf{v})\, \mathrm{d}\boldsymbol{\mu} \\
&= \iint\limits_{\Omega\times D} \frac{1}{2}|\mathbf{v} - \nabla_D\varphi|^2 \mathrm{d}\boldsymbol{\mu} - \iint\limits_{\Omega\times D} \underbrace{\left(\nabla_\Omega \cdot \varphi + \frac{1}{2}|\nabla_D\varphi|^2\right)}_{\leqslant 0} \mathrm{d}\boldsymbol{\mu} \\
&\geqslant \iint\limits_{\Omega\times D} \frac{1}{2}|\mathbf{v} - \nabla_D\varphi|^2 \mathrm{d}\boldsymbol{\mu}.
\end{aligned}$$

Hence, the dual gap controls how much $\nabla_D\varphi$ and $\mathbf{v}$ are close to each other. Now, provided that there exists a solution φ (which is, as we have seen above, not guaranteed), that this solution φ is C^1 (which is likely to be false), then all solutions of the Dirichlet problem must have $\nabla_D\varphi$ as their tangent velocity field. On the other hand, on the set of mappings in $H^1(\Omega, \mathcal{P}(D))$ sharing the same velocity field, the Dirichlet energy is linear. In other words: concerning uniqueness, the convexity of Dir can only tell us that all solutions share the same velocity field.

The question becomes: from the knowledge of the velocity field, can one recover the mapping $\boldsymbol{\mu}$? The usual answer involves Lipschitz continuity assumptions on this velocity field [AGS08, Proposition 8.1.7], which translates in the control of the *second* derivatives (w.r.t. variables in D) of an hypothetical φ solution of the dual problem. In the case of mappings valued in a family of elliptically contoured distribution, we were able to have this control because of available explicit expressions, but in the general case it seems out of reach. We mention that Hug [Hug16, Section IV.2], in the case of geodesics in the Wasserstein space, showed how one can recover uniqueness of the mapping $\boldsymbol{\mu}$ once one can prove uniqueness of the tangent velocity field. However, his proof relies on the explicit expression of the velocity field once the Kantorovich potentials are known, and he only proves uniqueness in the class of $\boldsymbol{\mu}$ such that $\boldsymbol{\mu} \in L^2(\Omega \times D)$: he needs to assume some *a priori* regularity of $\boldsymbol{\mu}$ w.r.t. variables in D. We are aware of the Ambrosio-DiPerna-Lions theory [DL89, Amb04], but this theory still requires higher regularity on the velocity fields than the one we have, and only proves uniqueness among mappings which have some regularity w.r.t.

variables in D (the typical assumption would be that all $\boldsymbol{\mu}(\xi)$ have a density w.r.t. to $\mathcal{L}_D$ and that this density is uniformly bounded from above).

If for instance all boundary measures $\boldsymbol{\mu}_b(\xi)$ belong to some space $L^m(D)$ with a uniform bound on the $L^m(D)$ norm, then by Theorem 9.3 we know that $\boldsymbol{\mu}(\xi)$ also belong to $L^m(D)$ for a.e. $\xi \in \Omega$. (However, we know it that it holds for only one solution of the Dirichlet problem). Provided that we manage to prove that it is the case for *all* solutions of the Dirichlet problem (at least for the entropy one could adapt the proof of Baradat and Monsaigeon [BM18]), then it would be one step in the direction of using the result of Hug or Ambrosio-DiPerna-Lions: we would know that the measures have some regularity w.r.t. variables in D. Even with this wishful thinking, we are not over because we need to provide regularity on the velocity field, and we have no idea how.

12.2 Regularity of harmonic mappings

A general feature of harmonic mappings valued in Riemannian manifolds is that they exhibit a lot of regularity, maybe except on a singular set of small dimension. Notice that it is not possible for harmonic mappings valued in the Wasserstein space to gain regularity w.r.t. variables in D. Indeed, if the boundary conditions are very irregular, for instance if $\boldsymbol{\mu}_b(\xi)$ is a Dirac mass for all $\xi \in \partial\Omega$, then $\boldsymbol{\mu}(\xi)$ is also a Dirac mass for all $\xi \in \Omega$ where $\boldsymbol{\mu}$ is the solution of the Dirichlet problem (Proposition 10.1). Hence, seen as objects living on D, the measures $\boldsymbol{\mu}(\xi)$ are no more regular than the boundary conditions. In the first part of this manuscript, we were able to provide more regularity on our solutions because of congestion effects: the problem of Chapter 4 is of the form

$$\min_{\boldsymbol{\mu}} \left\{ \mathrm{Dir}(\boldsymbol{\mu}) + \int_\Omega E(\boldsymbol{\mu}(\xi)) \mathrm{d}\xi \right\},$$

for some $E : \mathcal{P}(D) \to \overline{\mathbb{R}}$ (and with the source space Ω being a segment), and all regularity was related to E. On the other hand, seen as mappings $\Omega \to (\mathcal{P}(D), W_2)$, one can expect harmonic mappings to be more regular. Actually, we think that the answer to the following question is positive.

Question 12.3. *If $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ is an harmonic mapping, can one show that, at least locally in the interior of Ω, the mapping $\boldsymbol{\mu}$ is Hölder continuous, in the sense that for all ξ, η away from the boundary $\partial\Omega$,*

$$W_2(\boldsymbol{\mu}(\xi), \boldsymbol{\mu}(\eta)) \leqslant C|\xi - \eta|^\gamma$$

for some $C, \gamma > 0$.

Of course, if the source space Ω is a segment, this answer is positive as geodesics are Lipschitz. We underline, as explained below, that we conjecture that there is no singular set, whatever the dimension of the source space is. However, because of the absence of a bootstrapping argument, we don't know if one could reach Lipschitz regularity. Notice, in the case of mappings valued in a family of elliptically contoured distributions (Section 10.3), that this conjecture is actually true as proved in Theorem 10.10. As we will explain below, the proof for this special case indicates why this result should be true in full generality.

Regularity theory for harmonic mappings valued in Riemannian manifolds Let $(\mathcal{N}, \mathfrak{g})$ be a (compact) Riemannian manifold. We want first to give a very quick overview of the proof for the regularity theory of Dirichlet minimizing mappings $f : \Omega \to (\mathcal{N}, \mathfrak{g})$. We refer to the articles by Schoen and Uhlenbeck [SU82, SU83] for the original investigation of the question and [HW08, Section 4] for a survey on the topic. We introduce the notation

$$E_{\xi,r}(f) := \frac{1}{2r^{p-2}} \int_{B(\xi,r)} |\nabla f(\eta)|^2_{\mathfrak{g}} \mathrm{d}\eta$$

which denotes the amounts of energy present in the ball of center ξ and radius r, rescaled in such a way that this quantity becomes invariant under dilatation (recall that p is the dimension of Ω). The general strategy is the following.

1. Show that, if f is harmonic, then for a given ξ the quantity $r \to E_{\xi,r}(f)$ is non increasing. To prove this result, one usually uses "interior" perturbation, i.e. one takes Φ a diffeomorphism of Ω close to identity and compare the Dirichlet energy of $f \circ \Phi$ with the one of f. Once the monotonicity is established, the idea is to look at the limit, when $r \to 0$, of $E_{\xi,r}(f)$ and to distinguish between two cases: either the limit is 0, either it is not.

2. Show that if $E_{\xi,r}(f)$ is smaller than a constant $\varepsilon > 0$ (which does not depend on r), then f is Hölder in the ball of center ξ and radius $r/2$. This result, known as a ε-regularity result, usually relies on an iterative argument whose key estimate is that, if $E_{\xi,r}(f) \leqslant \varepsilon$ than $E_{\xi,\delta r}(f) \leqslant \frac{1}{2} E_{\xi,r}(f)$ for some $\delta \in (0,1)$. This key estimate is proved thanks to a linearization of the manifold, estimates for harmonic mappings valued in a Euclidean space, and a fine control of the error made by linearization. It is valid for any compact manifold $\mathcal{N}$, seen as a submanifold of some Euclidean space thanks to the Nash embedding theorem. Then, with this key estimate at hand, one can control the precise speed at which $E_{\xi,r}(f)$ tends to 0 when $r \to 0$ and prove Hölder continuity thanks to Morrey-Campanato inclusions. Once Hölder continuity is proved, provided $(\mathcal{N}, \mathfrak{g})$ is smooth enough, the usual theory for elliptic equations comes into play and, by bootstrapping, one gets $f \in C^\infty(B(\xi, r/2), \mathcal{N})$.

3. The previous paragraph handled the case where the limit of $E_{\xi,r}(f)$ is 0. On the other hand, if this limit is strictly positive, then one can consider a rescaling of f, namely $f_r = f((\cdot - \xi)/r)$. This is sometimes called a *blow up* argument. The scaling on $E_{\xi,r}$ was chosen in such a way that $E_{\xi,r}(f) = E_{0,1}(f_r)$. By some compactness arguments, one can extract from f_r a subsequence converging to some $\bar{f} : B(0,1) \to \mathcal{N}$. This function $\bar{f}$ is 0-homogeneous (constant along the radii issued from 0) and, by minimality of f, it is harmonic, in the sense that it satisfies the Euler Lagrange equations coming from the minimization of the Dirichlet energy.

   Then one studies the 0-homogenous harmonic mappings valued in $\mathcal{N}$. By the Ishihara property [Ish78, Jos08], if $F : \mathcal{N} \to \mathbb{R}$ is a convex function, $F \circ \bar{f}$ is real-valued, convex and 0-homogeneous, hence constant. If $\mathcal{N}$ has negative curvature, there are enough convex functions (namely, the distance square to a given point of $\mathcal{N}$) to conclude that $\bar{f}$ is constant. But this implies that $E_{0,1}(\bar{f}) = 0$, which contradicts the assumption that $E_{\xi,r}(f)$ does not tend to 0 when $r \to 0$. On the other hand, if $\mathcal{N}$ has positive curvature, there may exist such 0-homogeneous harmonic mappings: the typical example is $\xi \to \xi/|\xi|$, defined on the unit ball of $\mathbb{R}^3$ and valued in the unit sphere of $\mathbb{R}^3$.

   A point ξ for which the limit of $E_{\xi,r}(f)$ is not 0 will correspond to a singular point, because f is not continuous around that point. Indeed, close to that point, f will behave like a 0-homogeneous harmonic mapping valued in $\mathcal{N}$. What was said just above is that there are no singular points for harmonic mappings valued in negatively curved manifolds, but there may be in positively curved ones: although the ε-regularity theory is generic, the study of 0-homogeneous harmonic mappings strongly depends on the geometry of $\mathcal{N}$.

Putting all the pieces together, the regularity result reads as follows. A mapping $f : \Omega \to (\mathcal{N}, \mathfrak{g})$ which is Dirichlet minimizing is smooth, except on a singular set $\Sigma \subset \Omega$. Moreover, the Hausdorff dimension of the singular set can be bounded by $p - \hat{p}$, where $\hat{p}$ is the smallest integer for which there exists non constant 0-homogeneous harmonic mappings defined on $\mathbb{R}^{\hat{p}}$ valued in $\mathcal{N}$.

The set Σ is defined as the set of ξ such that $E_{\xi,r}(f)$ does not tend to 0 when $r \to 0$. The estimation on the Hausdorff dimension of Σ comes from techniques originating from the work of Federer [Fed70]. In particular, if $\mathcal{N}$ is negatively curved then $\hat{p} = +\infty$ and Dirichlet minizing mappings are always smooth. Notice also, by definition of $E_{\xi,r}(f)$, that, if Ω is of dimension 2, then $E_{\xi,r}(f)$ always tends to 0 as $r \to 0$. Hence, the codimension of the singular set is always larger than 2, and all Dirichlet minizing mappings defined over a space of dimension 2 are smooth.

We finish this brief overview by mentioning two regularity results for mappings valued in metric spaces.

For harmonic mappings valued in metric spaces negatively curved in the sense of Alexandrov, as already proved in the original article by Korevaar and Schoen, harmonic mappings are *Lipschitz* in the interior of Ω [KS93, Theorem 2.4.6] and Hölder continuous up to the boundary [Ser94]. The proof of such a result does not rely on the general strategy described above, the authors directly showed that the metric counterpart of $|\nabla f|^2_{\mathfrak{g}}$ (the local density of energy) is a subharmonic function (provided f is Dirichlet minimizing), hence bounded in the interior of Ω. However, the subharmonicity of such a quantity is really a feature of negatively curved space, and it is false for harmonic mappings valued in positively curved (finite-dimensional) Riemannian manifolds.

As far as Q-functions are concerned (see Page 125), which is an example of mappings valued in a metric space of positive curvature, Almgren [AJ00] proved that Dirichlet minimizing Q-fonctions are Hölder continuous. De Lellis and Spadaro [DLS11] later proposed a simpler proof of this result. The latter proof relies on an estimate, for a Dirichlet minimizing mapping f, between the Dirichlet energy on a ball and the Dirichlet energy on the boundary of the ball, i.e. the sphere. This estimate leads to an ODE enabling to control the speed at which $E_{\xi,r}(f)$ tends to 0 when $r \to 0$. The main tool is comparison with clever explicit constructions. However, all the constants depend on Q, and there is no hope to take the limit $Q \to +\infty$.

What about harmonic mappings valued in the Wasserstein space? Let us try to explain what can adapted from the general strategy to prove regularity of harmonic mappings. Let $\boldsymbol{\mu} \in H^1(\Omega, \mathcal{P}(D))$ an harmonic mapping and let $\mathbf{v}$ be its tangent velocity field, we denote by

$$E_{\xi,r}(\boldsymbol{\mu}) = \frac{1}{2r^{p-2}} \int_{B(\xi,r)} \int_D |\mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu}$$

the rescaled energy over $B(\xi, r) \subset \Omega$.

1. The monotonicity formula, i.e. the fact that $r \mapsto E_{\xi,r}(\boldsymbol{\mu})$ is non increasing for $\boldsymbol{\mu}$ harmonic, is very likely to stay true. Indeed the proof of it in the Riemannian case uses "interior" perturbation which are also available here. Actually, a formal computation from the (expected) optimality conditions (7.3) leads to a divergence free stress-energy tensor [HW08, equation (29)] which is known to imply the monotonicity formula [HW08, Section 4.3].

2. We do not know how to prove an ε-regularity result, i.e. to prove that $\boldsymbol{\mu}$ is Hölder on $B(\xi, r/2)$ provided $E_{\xi,r}(\boldsymbol{\mu})$ is small enough. This is at this point that we face the infinite-dimensionality of the Wasserstein space and its positive curvature. Arguments of the Riemannian case completely fail as there is no embedding in a Euclidean space. Moreover, the metric tensor of the Wasserstein space does not depend smoothly on the point at which

it is computed. Eventually, comparisons with well-chosen competitors have not lead to any result yet, the main issue being the difficulty to control the energy of these other competitors, due to the positive curvature of the Wasserstein space. Although we think that this ε-regularity result holds, our attempts to prove it have failed.

3. On the other hand we think we can exclude *blow up* configurations, in the sense that there are no non constant 0-homogeneous harmonic mappings valued in the Wasserstein space. Indeed, we will rely on the Ishihara property, see Theorem 9.3. For a given $V \in C(D)$, we denote by $F_V : \mathcal{P}(D) \to \mathbb{R}$ the functional defined by

$$F_V(\mu) := \int_D V \mathrm{d}\mu. \tag{12.1}$$

As already mentioned before, if V is convex then F_V is convex along generalized geodesics. Thus, if $\boldsymbol{\mu}$ is harmonic and 0-homogeneous, then $F_V \circ \boldsymbol{\mu}$ is subharmonic and 0-homogeneous provided V is convex. (Actually, our result holds for only one minimizer of the Dirichlet energy, but we think that, at least in the case of potential energies, one can prove that it actually holds for all harmonic mappings). As a consequence, $F_V \circ \boldsymbol{\mu}$ is constant. It implies that the integral of $\boldsymbol{\mu}(\xi)$ against any convex potential does not depend on ξ. As the linear span of convex functions include all functions, we deduce that $\boldsymbol{\mu}(\xi)$ does not depend on ξ.

We emphasize that in the case of mappings valued in a family of elliptically contoured distributions, we used exactly this argument in Propositions 10.16 and 10.17. Actually, in this case, as our mappings were valued in a finite-dimensional Riemannian manifold, the ε-regularity derived from the general theory of [SU82], and we are able to give a positive answer to the question of regularity.

As we see the key point is that, even though the Wasserstein space is positively curved, there exists a lot of geodesically convex functions defined over it, and it is enough to exclude blow up. This argument is, for the moment, mainly heuristic but we really think that it could be implemented rigorously.

If we summarize, provided we can come up one day with a proof of an ε-regularity result, and provided that we write a rigorous proof of the absence of non constant 0-homogeneous mappings valued in the Wasserstein space, we would be able to provide a positive answer to the question raised at the beginning of this section.

12.3 Convergence of the numerical method

The numerical method that we proposed has no guarantee of convergence if we refine the discretization. We could have chosen a finite element discretization of the problem. It would have implied to choose finite element spaces in which $\boldsymbol{\mu}$ and $\mathbf{E}$ live, give a variational meaning to the (generalized) continuity equation, and, for the Dirichlet energy, to compute exactly or choose an approximation of the integral

$$\iint_{\Omega \times D} \frac{|\mathbf{E}|^2}{2\boldsymbol{\mu}}$$

for $\boldsymbol{\mu}, \mathbf{E}$ which belong to the finite element spaces. Notice, as soon as $\boldsymbol{\mu}$ is not piecewise constant, that analytical integration of the formula above promises to be very tedious. We have not followed this strategy, mostly for contingent reasons: we have originally worked on finite differences to be able to use FFT to run fast ADMM iterations, and we did not dare to implement another version. Moreover, as we will explain below, the obstructions in the proof of convergence are also present with finite element methods.

Recall that the setting of Chapter 11 is that, from the continuous primal and dual (Dirichlet) problems, which are two convex optimization problems in duality, we derive two discrete (i.e. finite-dimensional) problems, the two of them being in duality, and which are consistent approximations of the primal and dual continuous problems. As we do not work with finite elements, it would be hard to prove convergence of the *solutions* of the discrete problems: we do not know how to see a discrete $\boldsymbol{\mu}$ as a continuous one. However, we will indicate that, even trying to prove the convergence of the (numerical) *values* of the discrete problems to the continuous ones, we run into issues.

Before going into details, we indicate some related work about the computation of geodesics in the Wasserstein space, which would amount to take for Ω a segment of $\mathbb{R}$. Starting from the work of Maas [Maa11], some people have started to be interested in Wasserstein spaces over finite spaces using a formulation mimicing the Benamou-Brenier one. From a numerical point of view (though this was not the aim of such an article), it would be like solving the geodesic problem in the Wasserstein space with a continuous time but a discrete space. In this setting, convergence if one refines the spatial discretization has been obtained, first for a uniform cartesian mesh on the torus [GM13], and then in the more general framework of finite volumes [GKM18]. These proofs rely on careful regularization procedures with the help of heat flows and a fine study of the metric tensor of the discrete Wasserstein space. On the other hand, for a fixed discrete Wasserstein space, a proof of the convergence of a time discretization was obtained in [ERSS17] but all the constants of this proof blow up when one refines the space discretization. We also mention [BC15, Section 3] which gives a proof of convergence for some static problem related to optimal transport, and then asserts that the dynamical case (which corresponds to computing geodesics in the Wasserstein space) is likely to be more involved. To the best of our knowledge, there is no proof of convergence of algorithms computing geodesics in the Wasserstein space, defined from Benamou-Brenier formulations, when one refines both the temporal and the spatial grid.

Now we go to the framework of Chapter 11, where Ω is the unit square discretized with N points and D is the 2-dimensional torus discretized with M points. We fix some boundary conditions $\boldsymbol{\mu}_b : \partial\Omega \to \mathcal{P}(D)$ which we assume Lipschitz w.r.t. variables in $\partial\Omega$. We want to show convergence of the values of the problem. Let us call $V_{N,M}$ the value of the discrete dual problem (see Definition 11.1) and V the value of the continuous problem (see Definition 8.30).

We claim that, quite easily, one should get

$$V \leqslant \liminf_{N,M \to +\infty} V_{N,M}.$$

If we were working with finite elements, this identity would be automatic. Indeed, from solutions $(\boldsymbol{\mu}_{N,M}, \mathbf{E}_{N,M})$ of the discrete primal problem, we would extract a converging subsequence to $(\boldsymbol{\mu}, \mathbf{E})$. The continuity equation is a linear constraint, hence should pass to the limit. On the other hand, the Dirichlet energy is l.s.c. hence

$$V \leqslant \mathrm{Dir}(\boldsymbol{\mu}, \mathbf{E}) \leqslant \liminf_{N,M \to +\infty} \mathrm{Dir}(\boldsymbol{\mu}_{N,M}, \mathbf{E}_{N,M}) = \liminf_{N,M \to +\infty} V_{N,M}$$

In our finite difference setting, what we can do instead is to sample the continuous dual problem. Indeed, take φ a solution of the continuous dual problem. It is not difficult to smooth it a little bit while still respecting the differential constraint to which it is submitted and not changing too much $\mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi)$. Then, it is enough to take for $\varphi_{N,M}$ the sampled values of φ on the relevant grid. By consistency of our discrete formulation, this discrete $\varphi_{N,M}$ will satisfy the discrete constraint. Moreover, $\mathrm{BT}^{\mathrm{dsc}}_{\boldsymbol{\mu}_b}(\varphi_{N,M}) \simeq \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi)$ at least if N, M are large enough. Using $\varphi_{N,M}$ as a competitor in the discrete dual problem, we get a lower bound on $V_{N,M}$ by something close to V, which was our claim. We see that the key point is that the constraint in the dual problem is one-sided, which leaves us some room for regularization techniques.

On the other hand, the reverse inequality, namely

$$V \geqslant \limsup_{N,M \to +\infty} V_{N,M}.$$

is currently out of reach. The natural idea would be to take $\boldsymbol{\mu}, \mathbf{E}$ solution of the continuous primal problem and to sample them on a grid to get $\boldsymbol{\mu}_{N,M}, \mathbf{E}_{N,M}$ discrete competitors. Provided there is a clean discretization, we could expect $(\boldsymbol{\mu}_{N,M}, \mathbf{E}_{N,M})$ to solve exactly the discrete continuity equation. However, the function

$$(\boldsymbol{\mu}, \mathbf{E}) \mapsto \frac{|\mathbf{E}|^2}{2\boldsymbol{\mu}}$$

is not uniformly continuous: its derivative has a singularity in 0. Hence, if the density $\boldsymbol{\mu}$ vanishes at some point, we could make a large error in the Dirichlet energy with this sampling process. To counter this effect, we would like to regularize $\boldsymbol{\mu}$ a little bit, for instance with a convolution or a heat flow. But we have to take care of the boundary conditions on $\partial\Omega$! One option would be to do a regularization which preserves the boundary conditions. This is what is done for instance in [ERSS17], but in a simpler setting: as they work on geodesics (corresponding to $\Omega = [0,1]$), there are only two boundary points $t = 0$ and $t = 1$. It is not clear how to adapt their proof when Ω is no longer a segment. Another option would be to act that we loose the boundary conditions, but it naturally leads to the following question.

Question 12.4. *Is the mapping which sends $\boldsymbol{\mu}_b : \partial\Omega \to \mathcal{P}(D)$ onto the (numerical) value of the Dirichlet problem with boundary conditions $\boldsymbol{\mu}_b$ continuous? If the answer is positive, for which topology on the set of boundary conditions?*

If Ω is a segment (which is the framework of [ERSS17, GKM18]), then the value of the Dirichlet problem is the squared Wasserstein distance, which is of course a continuous functions of its inputs. In the case of harmonic mappings, from the dual formulation (Theorem 8.36) we know that the mapping is l.s.c. as a supremum of continuous mappings $\boldsymbol{\mu}_b \mapsto \mathrm{BT}_{\boldsymbol{\mu}_b}(\varphi)$. On the other hand, showing an upper semi-continuity would amount to prove a stability result for the optimal φ in the dual formulation. Such a feature is not known yet, and probably related to the question of the existence of a solution to the dual problem. Moreover, these continuity properties should rather hold for the discrete problem (uniformly in N, M), which means the proofs should be adaptable to the discrete setting. In any case, a tentative proof for the regularization of the primal problem, followed by a sampling procedure, has failed. In the primal problem the constraint is an equality (namely the generalized continuity equation together with the boundary conditions), hence there is much less room for regularization. We emphasize that this whole discussion could be applied to a finite element discretization.

Another idea would be to interpolate a solution of the dual discrete problem. If $\varphi_{N,M}$ is a solution of the discrete dual problem, one could try to interpolate it to produce a competitor on the continuous dual problem, hence giving a lower bound on V which is close to $V_{N,M}$. With our current discretization, to show that from $\varphi_{N,M}$ one can indeed build an admissible competitor at the continuous level, we would need estimates on the (discrete) second derivatives of $\varphi_{N,M}$, uniformly on N, M. We don't even know yet how to get these estimates at the continuous level.

Appendix A

Résumé des résultats de la thèse

Ce chapitre contient un résumé des problèmes abordés dans cette thèse et des résultats auxquels nous sommes parvenus. Il a été écrit pour se lire indépendamment du reste de ce travail. Pour le garder concis, nous avons fait le choix de n'évoquer que très brièvement les modèles dont les problèmes abordés proviennent : l'unité de notre travail se trouve plus dans la structure mathématique commune aux différents problèmes variationnels étudiés qu'aux phénomènes qu'ils prétendent décrire. Signalons aussi que, dans un souci de clarté, les énoncés des résultats donnés dans ce chapitre sont parfois peu rigoureux et ne décrivent pas le cadre le plus général traité dans le cœur de ce manuscrit.

A.1 Cadre de la thèse

Fixons nous Ω un domaine convexe et borné de l'espace euclidien $\mathbb{R}^d$. L'espace de Wasserstein n'est autre que l'ensemble des mesures de probabilité sur Ω, que l'on munit de la distance (quadratique) de Wasserstein dont la définition est rappelée ci-dessous, cf. (A.1). Un élément de l'espace de Wasserstein est pensé comme une distribution de masse dans le domaine Ω, et tous les éléments de cet espace partagent la même masse totale, à savoir 1. La distance de Wasserstein entre deux distributions μ et ν représente alors le coût minimal nécessaire pour déplacer la masse de la configuration μ vers la configuration ν. Au vu de cette définition, l'espace de Wasserstein est souvent un cadre naturel lorsque l'on cherche à modéliser des phénomènes comme l'évolution d'une configuration de masse lorsque la masse totale est conservée. Dans ce travail, nous nous intéressons à des problèmes variationnels dans lesquels les inconnues sont soit des courbes, soit des applications, prenant leurs valeurs dans l'espace de Wasserstein.

Une courbe à valeurs dans l'espace de Wasserstein est pensée comme l'évolution temporelle d'une configuration de masse : une foule, un troupeau de moutons, un ensemble de particules (depuis les molécules jusqu'aux étoiles), etc. Nous nous intéresserons à des problèmes aux limites, c'est-à-dire lorsque les valeurs de la courbe à l'instant initial et l'instant final sont données (ou du moins pénalisées), et pour lesquels la courbe minimise une certaine énergie faisant intervenir sa vitesse, mesurée dans l'espace de Wasserstein.

Une extension naturelle des courbes consiste en les applications, c'est-à-dire que nous nous intéressons aussi aux situations où l'espace de départ n'est plus seulement uni-dimensionnel (la variable correspondant dans ce cas au temps), mais est un domaine de l'espace euclidien. Nous considérons alors des problèmes variationnels pour des applications à valeurs dans l'espace de Wasserstein prenant des valeurs fixées sur le bord du domaine (c'est-à-dire que nous regardons toujours des problèmes aux limites), et qui minimisent leur énergie de Dirichlet, à savoir l'intégrale du carré de la norme de leur gradient, où la norme du gradient est mesurée à l'aide de la distance de Wasserstein. Les solutions de ces problèmes variationnel sont naturellement appelées les applications harmoniques (à valeurs dans l'espace de Wasserstein).

Plus précisément, commençons par quelques rappels à propos de la théorie du transport optimal [Vil03, Vil08, AGS08, San15]. Si $\mu, \nu \in \mathcal{P}(\Omega)$, sont deux mesures de probabilité (i.e. deux éléments de l'espace de Wasserstein), la distance entre les deux est définie par

$$W_2(\mu, \nu) := \sqrt{\min_{\gamma} \left\{ \iint_{\Omega \times \Omega} |x-y|^2 \gamma(\mathrm{d}x, \mathrm{d}y) \ : \ \gamma \in \mathcal{P}(\Omega \times \Omega) \text{ et } \pi_0 \# \gamma = \mu, \ \pi_1 \# \gamma = \nu \right\}}. \quad \text{(A.1)}$$

Dans cette formule, π_0 et $\pi_1 : \Omega \times \Omega \to \Omega$ sont les projections sur respectivement la première et la deuxième composante de $\Omega \times \Omega$. Une mesure $\gamma \in \mathcal{P}(\Omega \times \Omega)$ qui satisfait les contraintes $\pi_0 \# \gamma = \mu$ et $\pi_1 \# \nu = \gamma$ est appelé un plan de transport entre μ et ν, et il est dit optimal s'il réalise le minimum du membre de droite de (A.1).

Un plan de transport γ décrit une manière de transporter la masse de la configuration $\mu(x)\mathrm{d}x$ vers $\nu(y)\mathrm{d}y$: la quantité de masse qui est transportée de x à y n'est autre que $\gamma(x,y)\mathrm{d}x\mathrm{d}y$. Le coût pour un tel transport est $|x-y|^2$, et le carré de la distance de Wasserstein correspond au coût le minimal parmi tous les transports possibles.

L'application $W_2 : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}_+$ définit une distance sur l'ensemble $\mathcal{P}(\Omega)$, et elle métrise la convergence en mesure. L'espace métrique $(\mathcal{P}(\Omega), W_2)$ est appelé l'espace de Wasserstein. Nous soulignons ici que Ω est supposé compact : dans cette thèse, nous travaillons sous cette hypothèse qui simplifie certains aspects techniques tout en conservant les caractéristiques typiques de l'espace de Wasserstein. Pour travailler dans un cadre non compact, il faudrait aussi s'intéresser aux moments d'ordre 2 des mesures de probabilité considérées.

Nous mentionnons le théorème de Brenier [Bre87], qui montre que les plans de transport optimaux ont une structure bien particulière : sous réserve que μ a une densité par rapport à la mesure de Lebesgue (en fait cette hypothèse peut être affaiblie), il existe un unique plan de transport optimal $\gamma \in \mathcal{P}(\Omega \times \Omega)$ entre μ et ν, et il est concentré sur le graphe du gradient d'une fonction convexe, c'est-à-dire qu'il existe $T : \Omega \to \Omega$, gradient d'une fonction convexe, tel que $\gamma = (\mathrm{Id}, T)\#\mu$. Ce résultat montre que le couplage entre mesures de probabilité donné par la théorie du transport optimal est en fait un objet avec une structure très rigide.

Nous nous intéressons principalement au point de vue différentiel sur l'espace de Wasserstein. Si $\rho : [0,1] \to \mathcal{P}(\Omega)$ est une courbe à valeurs dans l'espace de Wasserstein, par exemple Lipschitz par rapport à la distance de Wasserstein, on peut lui associer une vitesse, qui est la quantité scalaire définie par, à t fixé,

$$|\dot{\rho}_t| = \lim_{h \to 0} \frac{W_2(\rho_{t+h}, \rho_t)}{|h|}.$$

Sous réserve que la courbe soit Lipschitz (en fait on peut affaiblir cette hypothèse), cette quantité est bien définie et finie pour presque tout temps. La quantité centrale est l'*action* de la courbe, définie par

$$A(\rho) := \int_0^1 \frac{1}{2} |\dot{\rho}_t|^2 \mathrm{d}t, \tag{A.2}$$

et qui se comporte comme une norme H^1 (mise au carré). Cette quantité a un lien avec des considérations de mécanique de fluide. Comme cela a été compris par Benamou et Brenier [BB00], l'action $A(\rho)$ d'une courbe coïncide avec

$$\min_{\mathbf{v}} \left\{ \int_0^1 \left(\int_\Omega \frac{1}{2} |\mathbf{v}|^2 \mathrm{d}\rho \right) \mathrm{d}t \ : \ \mathbf{v} : [0,1] \times \Omega \to \mathbb{R}^d \ \text{ et } \ \partial_t \rho + \nabla \cdot (\mathbf{v}\rho) = 0 \right\}. \tag{A.3}$$

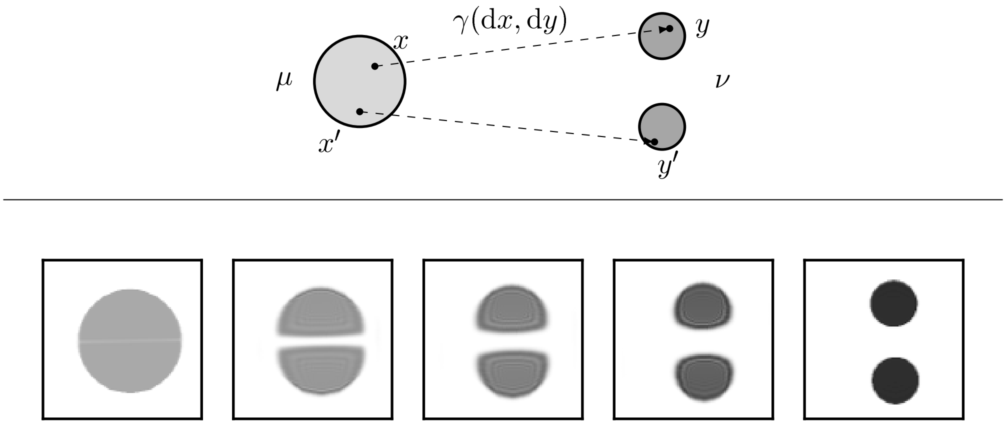

Figure A.1: En haut : vue schématique de la formulation (A.1) du transport optimal entre μ, à gauche, et ν, à droite. La quantité $\gamma(x, y)\mathrm{d}x\mathrm{d}y$ représente la quantité de masse qui est transportée de x à y. Le couplage γ est ensuite choisi de manière à minimiser le coût total de transport. En bas : géodésique dans l'espace de Wasserstein entre les mêmes mesures. Une fois que le γ optimal est choisi, une proportion $\gamma(x, y)\mathrm{d}x\mathrm{d}y$ se déplace en ligne droite à vitesse constante entre x et y. Le résultat macroscopique de tous ces mouvements microscopiques est une mesure de probabilité évoluant dans le temps, dont des instantanés sont affichés.

Plus précisément, l'équation de continuité $\partial_t \rho + \nabla \cdot (\mathbf{v}\rho) = 0$ dit que le champ de vitesse $\mathbf{v}$ (dépendant du temps) représente le mouvement de particules composant ρ, dans le sens où si une assemblée de particules a une vitesse en un point x et un instant t donnée par $\mathbf{v}_t(x)$, alors le mouvement collectif est décrit par une densité ρ évoluant en temps selon l'équation de continuité. Dès lors, parmi tous les champs de vitesse $\mathbf{v}$ qui représentent le mouvement de masse décrit par ρ, on choisit celui qui minimise l'intégrale en temps de l'énergie cinétique (l'intégrale temporelle d'un Lagrangien, c'est-à-dire une action), et la valeur minimale n'est autre que l'action de la courbe $A(\rho)$, qui avait été définie de manière purement métrique.

Il existe une classe particulière de courbes à valeurs dans l'espace de Wasserstein, à savoir les géodésiques parcourues à vitesse constante. On peut en donner une définition purement métrique : une courbe $\rho : [0, 1] \to \mathcal{P}(\Omega)$ est une géodésique (parcourue à vitesse constante) si et seulement si, pour tous les instants s et t,

$$W_2(\rho_t, \rho_s) = |t - s| W_2(\rho_0, \rho_1).$$

Étant données deux mesures de probabilité μ, ν, il existe toujours une géodésique telle que $\rho_0 = \mu$ et $\rho_1 = \nu$. D'ailleurs, cette géodésique est une courbe solution du problème variationnel

$$\min_\rho \{A(\rho) \ : \ \rho_0 = \mu \text{ and } \rho_1 = \nu\}.$$

De plus, la structure des géodésiques a un lien fort avec le problème de transport optimal. En effet, si $\gamma \in \mathcal{P}(\Omega \times \Omega)$ est un plan de transport optimal entre μ et ν, alors la courbe définie par $\rho_t = ((1 - t)\pi_0 + t\pi_1)\#\gamma$ est une géodésique entre μ et ν, et réciproquement toute géodésique est de cette forme. Un exemple de géodésique dans l'espace de Wasserstein est affiché dans la Figure A.1.

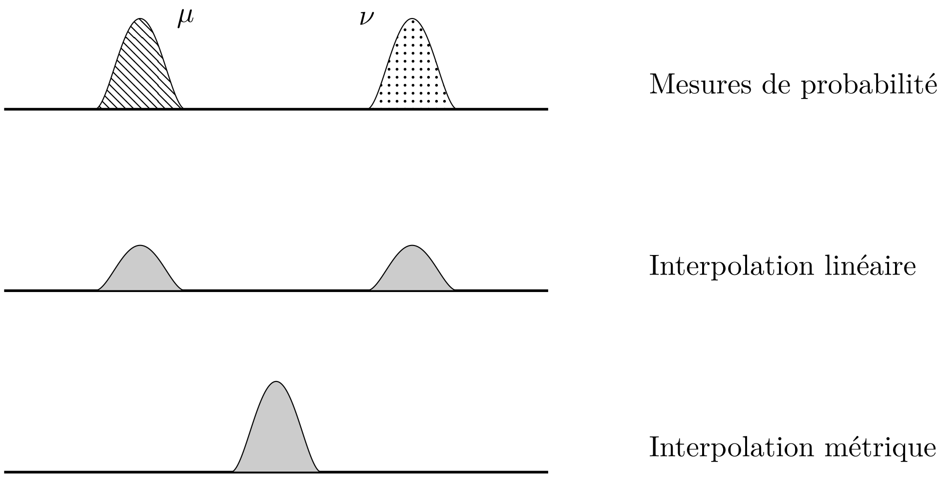

Figure A.2: À propos des différentes manières d'interpoler entre deux mesures de probabilité. En haut : deux mesures de probabilité μ et ν sur l'axe réel. Au milieu : interpolation linéaire $(\mu + \nu)/2$ des deux mesures. En bas : interpolation métrique entre les deux mesures, c'est-à-dire le point milieu de la géodésique dans l'espace de Wasserstein joignant μ à ν.

L'espace de Wasserstein est un sous-ensemble convexe de l'ensemble des mesures sur Ω. En termes moins savants, si μ et ν sont deux mesures de probabilité sur Ω, la mesure $(\mu+\nu)/2$, c'est-à-dire la moyenne (linéaire) des deux, est encore une mesure de probabilité. Une fonctionnelle $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ sera dite convexe si $F((\mu + \nu)/2) \leqslant (F(\mu) + F(\nu))/2$. D'un autre côté, il existe un autre moyen de faire la moyenne de μ et ν : il s'agit de prendre $\rho_{1/2}$, où $t \to \rho_t$ est une géodésique à vitesse constante joignant μ à ν. Pour peu que l'une des mesures ait une densité par rapport à la mesure de Lebesgue, la géodésique joignant les deux est unique, de sorte que $\rho_{1/2}$, que l'on appellera moyenne métrique est bien définie, cf. Figure A.2. Si $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ est une fonctionnelle semi-continue inférieurement, on dira que F est géodésiquement convexe si $t \mapsto F(\rho_t) \in \overline{\mathbb{R}}$ est convexe pour toute géodésique ρ. En fait, on aura besoin de la propriété légèrement différente de convexité le long de géodésiques généralisées, nous renvoyons le lecteur ou la lectrice au corps de ce manuscrit pour une explication de la différence, qui est peu pertinente pour la suite.

L'exemple type de fonction géodésiquement convexe est l'(opposé de l') entropie de Boltzmann, définie par

$$F(\mu) = \begin{cases} \displaystyle\int_\Omega \mu(x) \ln(\mu(x)) \mathrm{d}x & \text{si } \mu \text{ a une densité par rapport à Lebesgue,} \\ +\infty & \text{sinon.} \end{cases} \tag{A.4}$$

Plus généralement, toute fonction définie par $F(\mu) = \int_\Omega f(\mu)$ (dans le cas où μ est absolument continu par rapport à Lebesgue) avec f convexe, superlinéaire, et $s \mapsto s^d f(s^{-d})$ convexe décroissante, est convexe le long des géodésiques (généralisées).

Une fois que l'on se donne une fonctionnelle F convexe le long des géodésiques (généralisées), il est possible de considérer un problème d'évolution dans l'espace de Wasserstein, appelé flot gradient de F qui s'écrit heuristiquement

$$\frac{\mathrm{d}\rho_t}{\mathrm{d}t} = -\nabla F(\rho).$$

Bien sûr, ni la dérivée temporelle, ni le gradient n'ont de sens dans l'espace de Wasserstein, mais il est possible de donner un sens à cette équation à l'aide de quantités ne faisant intervenir que la distance de Wasserstein. Ainsi, une caractérisation métrique des flots gradient, appelée (EVI) pour *Evolution variational inequality*, se lui comme suit : une courbe $t \mapsto \rho_t$ est un flot gradient pour la fonctionnelle F si et seulement si, pour tout $\nu \in \mathcal{P}(\Omega)$ fixé et tout $t \geqslant 0$,

$$\lim_{h \to 0, h>0} \frac{W_2(\rho_{t+h}, \nu) - W_2(\rho_t, \nu)}{h} \leqslant F(\nu) - F(\rho_t). \tag{A.5}$$

Nous soulignons qu'il est crucial de supposer que F soit convexe le long des géodésiques généralisées (en fait on pourrait étendre cette définition au cas où F est λ-convexe). Un des résultats majeurs de [AGS08] est que, pour toute donnée initiale μ telle que $F(\mu) < +\infty$, il existe une unique courbe $\rho : [0, +\infty) \to \mathcal{P}(\Omega)$, qui vérifie (A.5), et telle que $\rho_0 = \mu$. Dans le cas où F est l'entropie de Boltzmann, c.f. (A.4), cette courbe est la solution de l'équation d'évolution

$$\partial_t \rho = \Delta \rho$$

avec condition de Neumann sur le bord, c'est-à-dire que le flot gradient de l'entropie dans l'espace de Wasserstein n'est autre que le flot de la chaleur. Cette remarquable propriété, observée dans [JKO98], est notamment ce qui a motivé l'étude des flots gradients dans l'espace de Wasserstein.

A.2 Courbes optimales à valeurs dans l'espace de Wasserstein

Nous nous sommes intéressés à des problèmes variationnels dans lesquels l'inconnue est une courbe à valeurs dans l'espace de Wasserstein. Plus précisément, nous désignerons dans la suite par $\Gamma = C([0,1], \mathcal{P}(\Omega))$ l'ensemble des courbes continues à valeurs dans l'espace de Wasserstein $(\mathcal{P}(\Omega), W_2)$. Nous rappelons que $A(\rho)$, pour $\rho \in \Gamma$, désigne l'action d'une courbe et est définie dans (A.2), on lui attribue éventuellement la valeur $+\infty$ si ρ n'est pas assez régulière pour que l'intégrale ait un sens.

A.2.1 Géodésiques parcourues à vitesse constante

Le problème le plus simple (et bien compris) pour de telles courbes s'écrit

$$\min_{\rho} \left\{ A(\rho) \ : \ \rho \in \Gamma, \ \rho_0, \rho_1 \text{ données} \right\}. \tag{A.6}$$

Les solutions de ce problème sont les géodésiques joignant ρ_0 à ρ_1, parcourues à vitesse constante, dont on peut trouver une illustration Figure A.1.

Une variante de ce problème consiste, par exemple, à pénaliser la valeur finale au lieu de la fixer. Si l'on considère $\Psi : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ une fonction convexe et semi-continue inférieurement, on peut considérer le problème

$$\min_{\rho} \left\{ A(\rho) + \Psi(\rho_1) \ : \ \rho \in \Gamma, \ \rho_0 \text{ donnée} \right\}.$$

Ce problème apparaît d'ailleurs quand on regarde le *minimizing movement scheme* pour les flots gradients dans l'espace de Wasserstein (parfois appelé schéma *JKO* d'après [JKO98] dans ce contexte). Nous soulignons qu'une fois ρ_1 connue (elle ne dépend que de ρ_0 et Ψ), la solution aux instants intermédiaires reste une géodésique dans l'espace de Wasserstein. Ainsi, dans ce qui nous intéressera par la suite, à savoir de la régularité locale en temps, que la valeur au temps final soit imposée ou juste pénalisée ne changera pas grand chose.

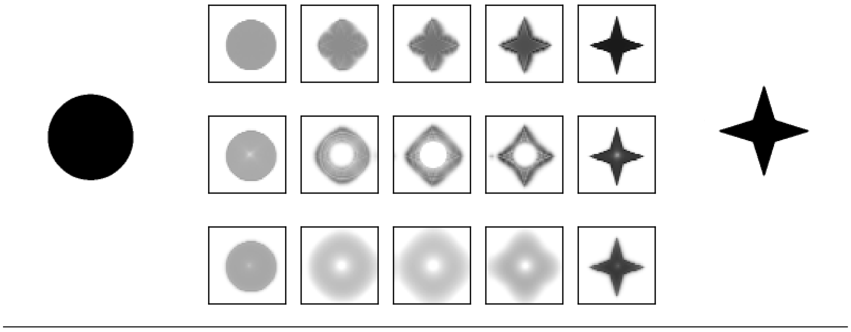

Figure A.3: Illustration du problème de l'évolution optimale de la densité avec de la congestion considéré dans le Théorème A.1 dans le cas où Ω est le tore de dimension 2. À gauche et à droite se trouvent des mesures de probabilité correspondant à la valeur initiale et la valeur finale de la courbe à valeurs dans l'espace de Wasserstein. La première ligne est la géodésique dans l'espace de Wasserstein entre les deux mesures : pas d'effets de congestion. Pour la deuxième ligne, nous avons ajouté un potentiel prenant des valeurs élevées au centre du domaine, forçant ainsi la densité à éviter cette région. Pour la dernière ligne, le potentiel pénalisant la présence de masse au centre est toujours présent, mais nous pénalisons aussi les densités congestionnées par le carré de la norme L^2 de la densité. En conséquence, la masse a tendance à s'étaler.

Terminons par une remarque tautologique, mais qui sera utile dans des cadres plus compliqués. Si $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ est une fonctionnelle définie sur l'espace de Wasserstein et ρ une solution de notre problème variationnel, on peut regarder l'évolution temporelle de $t \mapsto F(\rho_t)$. Dans le cas où F est convexe le long des géodésiques, alors $t \mapsto F(\rho_t)$ est bien entendu une fonction convexe du temps. Dans des problèmes variationnels plus élaborés, la majeure partie de notre analyse consistera justement à choisir une fonctionnelle F est pertinente et à décrire son évolution le long de la solution.

A.2.2 Évolution optimale de densité avec de la congestion

Inspiré par une modélisation issue des *jeux à champ moyen* [LL06a, LL06b, HMC06, Car10], nous nous intéressons à des problèmes où une courbe à valeurs dans l'espace de Wasserstein ne minimise pas seulement son action, mais aussi une énergie E qui pénalise certaines configurations où la densité est trop élevée. Le problème type est de la forme

$$\min \left\{ A(\rho) + \int_0^1 E(\rho_t) \mathrm{d}t \ : \ \rho \in \Gamma \right\}$$

où $E(\rho)$ est par exemple l'intégrale (spatiale) du carré de la densité, ou l'entropie de Boltzmann. La valeur initiale et la valeur finale de la courbe peuvent être imposées, ou pénalisées. Une alternative est de dire que $E(\rho)$ vaut $+\infty$ si la valeur de la densité dépasse un certain seuil critique et (par exemple) l'intégrale de ρ contre un potentiel sinon. Cela revient à mettre la contrainte que la densité de ρ ne dépasse pas un seuil, et mène à l'apparition de forces de pression concentrées sur l'endroit où la contrainte est saturée. Nous signalons que pour la structure linéaire sur l'espace des mesures probabilité, ce problème est *convexe* (pour peu que E le soit), ainsi l'existence d'une solution est assez standard une fois les espaces fonctionnels appropriés choisis.

D'un point de vue de la modélisation, ρ représente une densité d'individus (ou plutôt d'agents) qui cherchent à se rendre à un endroit (d'où par exemple la pénalisation de la valeur finale) mais qui tentent d'anticiper le comportement des autres agents de façon à éviter les zones congestionnées (d'où la pénalisation des configurations où la densité est trop élevée). Nous insistons qu'il est indispensable de prendre en compte l'*anticipation* des agents pour se retrouver avec un problème variationnel avec des conditions aux bords temporels données et pas une équation d'évolution.

La question qui nous intéresse est celle de la régularité des solutions : naturellement la courbe optimale ρ sera telle que $E(\rho_t)$ soit intégrable, mais en réalité la régularité spatiale pourra être plus forte : nous montrons des résultats de *régularité elliptique*, c'est-à-dire que la solution de certains problèmes variationnels convexes est plus régulière que ce que l'on pouvait penser *a priori*.

Congestion douce Nous nous intéressons au cas où l'énergie de pénalisation de la congestion E prend la forme

$$E(\rho) := \int_\Omega f(\rho(x))\mathrm{d}x + \int_\Omega V(x)\rho(x)\mathrm{d}x,$$

où f est une fonction convexe et bornée par en dessous, tandis que $V : \Omega \to \mathbb{R}$ est un potentiel au moins continu. Une approximation numérique du problème résultant est montrée dans la Figure A.3. Nous arrivons alors à montrer de la régularité L^∞ sur ρ.

Théorème A.1. *On considère le problème*

$$\min_\rho \left\{ A(\rho) + \int_0^1 \left(\int_\Omega f(\rho_t(x))\mathrm{d}x + \int_\Omega V(x)\rho_t(x)\mathrm{d}x \right) \mathrm{d}t + \Psi(\rho_1) \ : \ \rho \in \Gamma, \ \rho_0 \text{ donnée} \right\},$$

où l'inconnue ρ est une courbe continue à valeurs dans l'espace de Wasserstein $(\mathcal{P}(\Omega), W_2)$.

On suppose que le domaine Ω est convexe, que le potentiel $V : \Omega \to \mathbb{R}$ est Lipschitz et que $\Psi : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ est convexe et semi-continue inférieurement. Enfin, on suppose que la fonction f est convexe et qu'il existe $C_f > 0$, $\alpha \geqslant -1$ tels que $f''(s) \geqslant C_f s^\alpha$ pour $s > 0$.

Alors, pour peu qu'il existe un compétiteur d'énergie finie[1]*, le problème admet une unique solution ρ et pour tout $0 < T_1 < T_2 < 1$, la restriction de ρ à $[T_1, T_2] \times \Omega$ appartient à $L^\infty([T_1, T_2] \times \Omega)$.*

Ce résultat, obtenu initialement dans l'article [LS18] écrit en collaboration avec F. Santambrogio, se trouve dans le Chapitre 4. Nous soulignons qu'en réalité dans ce chapitre de nombreuses variations autour de ce résultat sont étudiées. Pour peu que l'on spécifie la forme précise de Ψ, la régularité peut être étendue jusqu'à l'instant final (i.e. on peut prendre $T_2 = 1$), mais l'étendre jusqu'à l'instant initial reste une question ouverte. De plus, l'estimation sur f peut être affaiblie, quitte à supposer de la régularité supplémentaire sur le potentiel : par exemple on peut imposer $f''(s) \geqslant C_f s^\alpha$ seulement pour $s \geqslant s_0$ pour un certain s_0 (mais alors il faut supposer $V \in C^{1,1}$), et l'on peut même regarder $\alpha < -1$ sous des hypothèses supplémentaires assez lourdes.

Ce résultat apporte réellement une information supplémentaire : l'hypothèse sur f autorise toutes les fonctions puissances $f(s) = s^q$ avec $q > 1$, et même l'entropie $f(s) = s\log(s)$. Avec une telle pénalisation, automatiquement un ρ d'énergie finie appartient à $L^q([0,1] \times \Omega)$, notre résultat nous apprend qu'en plus le ρ optimal appartient à L^∞ localement en temps et globalement en espace. Et ce sans aucune hypothèse de régularité sur la donnée initiale ρ_0, ni sur la pénalisation finale Ψ. Du moins la seule hypothèse à vérifier est l'existence d'un compétiteur d'énergie finie,

[1]Par abus de notation, l'*énergie* désigne la quantité que l'on cherche à minimiser

et cela est par exemple garanti automatiquement si f croît au plus comme s^q avec $q < 1 + 1/d$, où d est la dimension de l'espace.

Avant de donner une idée de la démonstration, nous mentionnons qu'un résultat de régularité L^∞ avait déjà été obtenu pour un problème similaire de jeux à champ moyen par P.-L. Lions (dans la deuxième heure de la vidéo du cours au Collège de France du 27 Novembre 2009 [Lio12]). Ce résultat repose sur un principe du maximum pour les équation elliptiques dégénérées : pour ce faire, il est nécessaire que les mesures ρ_0 et ρ_1 soit fixées et dans L^∞ (ainsi que bornées par en dessous), et qu'il n'y ait pas de potentiel V, contrairement à notre résultat. En revanche, P.-L. Lions peut traiter des Lagrangiens généraux (c'est-à-dire remplacer l'action A par une fonctionnelle plus générale), tandis que nous devons nous restreindre au cas quadratique.

Donnons très brièvement une idée de la démonstration : elle repose sur une idée similaire à la démonstration d'une borne L^∞ pour les solutions d'équations elliptiques par Moser [Mos60]. Pour simplifier, regardons seulement le cas où $V = 0$, notons $\rho \in \Gamma$ la solution de notre problème variationnel et introduisons les fonctionnelles $U_m : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ définies par

$$U_m(\rho) := \frac{1}{m(m-1)} \int_\Omega \rho(x)^m \mathrm{d}x,$$

(ou $+\infty$ si la mesure ρ n'a pas de densité par rapport à Lebesgue)avec U_1 qui serait l'entropie de Boltzmann. Ces fonctionnelles sont convexes le long des géodésiques, en particulier $t \mapsto U_m(\rho_t)$ est une fonction convexe du temps lorsque $f = 0$. Dans le cas $f \neq 0$, un calcul formel nous mène à

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} U_m(\rho_t) \geqslant \int_\Omega |\nabla \rho_t|^2 \rho_t^{m-1} f''(\rho_t),$$

c'est-à-dire que l'on peut quantifier la convexité de $U_m(\rho)$. À partir de là, en utilisant l'hypothèse $f''(s) \geqslant C_f s^\alpha$ et des injections de Sobolev, on tombe sur

$$C(m, C_f) \frac{\mathrm{d}^2}{\mathrm{d}t^2} U_m(\rho_t) \geqslant \left(\int_\Omega \rho_t^{\beta(m+1+\alpha)} \right)^{1/\beta}$$

où $\beta > 1$ est lié à la dimension de l'espace d. En bref, nous pouvons contrôler une puissance $\beta(m + 1 + \alpha) > m$ de ρ par une puissance m de ρ (tout cela intégré en espace). Certes il y a une dérivée seconde en temps qui apparaît, mais avec un certain travail (très similaire à [Mos60]) l'itération de l'estimée ci-dessus suffit à borner uniformément les normes L^m de ρ, et ainsi conclure à une borne L^∞. En pratique les calculs ci-dessus sont formels, et il faut introduire une discrétisation en temps pour les rendre rigoureux.

Congestion dure Nous nous intéressons maintenant au cas où les configurations pour lesquelles la densité dépasse un certain seuil sont tout simplement interdites. Cela correspond à une contrainte de capacité maximale, l'environnement ne peut pas accueillir plus qu'une certaine densité d'agents. Mathématiquement, l'énergie E prend la forme

$$E(\rho) := \begin{cases} \displaystyle\int_\Omega V(x)\rho(x)\mathrm{d}x & \text{si } \rho(x) \leqslant 1 \text{ pour presque tout } x \in \Omega, \\ +\infty & \text{sinon,} \end{cases}$$

où $V : \Omega \to \mathbb{R}$ est un potentiel fixé. Par convention, la seuil maximal pour la densité a été fixé à 1. Dans ce cas la densité ρ est automatiquement dans L^∞, la problématique se déplace vers la régularité de la pression qui est le multiplicateur de Lagrange associé à la contrainte sur la densité. Plus précisément, la pression P donne lieu à la force forçant la contrainte à être respectée et vérifie $P \geqslant 0$ (c'est une mesure positive) ainsi que $P = 0$ si $\rho < 1$ (elle n'est active que si la contrainte est saturée). L'enjeu est de prouver un résultat de régularité L^∞ sur la pression, cette question étant liée à des problématiques d'interprétation lagrangienne en jeux à champ moyen [CMS16].

Théorème A.2. *On considère le problème*

$$\min_\rho \left\{ A(\rho) + \int_0^1 \left(\int_\Omega V(x)\rho_t(x)\mathrm{d}x \right) \mathrm{d}t + \int_\Omega \Psi(x)\rho_1(x)\mathrm{d}x \ : \ \rho \in \Gamma, \ \rho_0 \ \textit{donnée} \right\},$$

où l'inconnue $\rho \in \Gamma$ satisfait la contrainte $\rho_t \leqslant 1$ pour tout instant $t \in [0,1]$. On suppose que Ω est convexe et que les potentiels $V, \Psi : \Omega \to \mathbb{R}$ appartiennent à $W^{1,q}(\Omega)$ avec $q > d$, où d est la dimension de l'espace ambiant.

Alors il existe $p \in L^\infty([0,1] \times \Omega)$ et $P_1 \in L^\infty(\Omega)$ tels que

$$P = p(x,t)(\mathrm{d}x \otimes \mathrm{d}t) + P_1(x)(\mathrm{d}x \otimes \delta_{t=1})$$

soit le multiplicateur de Lagrange associé à la contrainte $\rho \leqslant 1$.

Ce résultat, obtenu initialement dans l'article [LS19] écrit en collaboration avec F. Santambrogio, se trouve dans le Chapitre 5. Comme déjà remarqué dans [CMS16], même si la pression est régulière (avec une densité L^∞ par rapport à Lebesgue) sur $[0,1) \times \Omega$, il n'est pas possible d'exclure une singularité temporelle de la pression pour l'instant final $t = 1$.

La seule étude précédente concernant la régularité de la pression dans les jeux à champ moyen avec contrainte de densité dont nous soyons au courant est celle de [CMS16], où les auteurs obtiennent $P \in L^2_{t,loc} BV_x$. Grâce à l'injection $BV \hookrightarrow L^{d/(d-1)}$, cela permet de de dire que $P \in L^m$ avec $m > 1$ (localement en temps, globalement en espace). Une telle régularité avait été obtenue en adaptant la preuve de la régularité de la pression dans le cas des équations d'Euler incompressible, d'abord étudiée par [Bre99], et par la suite raffinée dans [AF08]. La stratégie générale, appelée par la suite *régularité par dualité* [San18] permet d'obtenir de la régularité Sobolev pour des équations elliptiques très dégénérées. Dans notre cas nous adoptons une stratégie complètement différente, qui nous conduit à une information sur le laplacien de la pression, mais qui ne marche que pour des lagrangiens quadratiques (alors que [CMS16] permet de traiter des quantités plus générales que l'action A). D'un autre côté, nous avons besoin de moins de régularité sur le potentiel (V doit avoir une régularité Sobolev au lieu de $V \in C^{1,1}$ dans [CMS16]) et notre stratégie marche sur des domaines convexes, pas seulement sur le tore comme pour [CMS16].

Plus précisément, soit ρ une solution du problème qui nous intéresse. On note $\mathbf{v}$ son champ de vitesse tangent, c'est-à-dire le $\mathbf{v}$ optimal dans la formule de Benamou-Brenier (A.3). La dérivée convective associée à $\mathbf{v}$, c'est-à-dire $\partial_t + \mathbf{v} \cdot \nabla$ est notée D_t. À partir des conditions d'optimalité, un calcul formel conduit à

$$-D_{tt} \ln(\rho) \leqslant \Delta(P + V).$$

Or, si $P > 0$, ce qui ne peut arriver que si $\rho = 1$ (la contrainte est saturée), alors $\ln(\rho)$ atteint un maximum de sorte que $-D_{tt}\ln(\rho) \geqslant 0$. En bref, on arrive à la conclusion que

$$\Delta(P + V) \geqslant \quad \text{sur } \{P > 0\}.$$

C'est une sorte de problème de l'obstacle pour P. Des résultats de régularité elliptique plutôt standards, couplés à de bonnes conditions au bord (car Ω est supposé convexe) permettent de conclure, à t fixé, que $P \in H^1(\Omega)$ si $V \in H^1(\Omega)$ et $P \in L^\infty(\Omega)$ si $V \in W^{1,q}(\Omega)$ avec $q > d$. Un argument similaire permet de conclure pour la régularité de la pression lorsque $t = 1$. De manière analogue au paragraphe précédent, les calculs présentés ici sont formels et une approximation par discrétisation temporelle est nécessaire pour les rendre rigoureux.

A.2.3 Formulation variationnelle des équations d'Euler

Une autre variation autour du problème (A.6) apparaît dans la formulation variationnelle des équations d'Euler incompressible [Bre89, Bre99, AF09, DF12]. Nous énonçons ici le problème sous une forme qui le rend similaire à (A.6), même si ce n'est pas sous celle-ci qu'il est originellement apparu dans [Bre89].

Plus précisément, rappelons que $\Gamma = C([0,1], \mathcal{P}(\Omega))$ désigne l'ensemble des courbes continues à valeurs dans l'espace de Wasserstein. Notre inconnue ne sera pas un élément de Γ, mais Q une mesure de probabilité sur Γ. En d'autres termes, notre inconnue est la loi d'une courbe aléatoire. La quantité que nous cherchons à minimiser est l'espérance de l'action $\mathbb{E}_Q[A(\rho)]$ sous deux contraintes : celle que la loi jointe de Q aux instants $t=0$ et $t=1$ soit fixée et surtout celle que pour tout t, $\mathbb{E}_Q[\rho_t] = \mathcal{L}$ la mesure de Lebesgue sur Ω. Cette deuxième contrainte exprime l'incompressibilité et force la masse à se répartir uniformément sur Ω.

D'un point de vue de la modélisation, il faut imaginer Q comme décrivant la cinématique d'un fluide composé d'une infinité de phases : $Q(\rho)\mathrm{d}\rho$ donne la proportion de phases suivant la trajectoire ρ. Les configurations des différentes phases sont fixées à l'instant initial et l'instant final, et comme le fluide est incompressible, au niveau global la somme des toutes les phases se répartit uniformément sur le domaine. Enfin, en suivant le principe de moindre action, la quantité à minimiser n'est autre que l'action totale, c'est-à-dire l'intégrale en temps de l'énergie cinétique.

Ce modèle a été étudié sous l'angle de l'existence des solutions [Bre89], de la description du défaut d'unicité [BFS09] et surtout de l'existence et la régularité du multiplicateur de Lagrange correspondant à la contrainte d'incompressibilité, à savoir la pression [Bre99, AF08, AF09]. Pour notre part, nous nous sommes intéressés à une conjecture laissée ouverte par Brenier [Bre03, Section 4], à savoir le comportement en temps de l'entropie moyenne définie par

$$\mathcal{H}_Q(t) = \mathbb{E}_Q \left[\int_\Omega \rho_t(x) \ln(\rho_t(x)) \mathrm{d}x \right]. \tag{A.7}$$

En effet, un calcul formel indique que la fonction $\mathcal{H}_Q$, pour Q solution du problème, devrait être une fonction convexe du temps, mais ce problème est resté ouvert. Nous y apportons une réponse positive.

Théorème A.3. *Soient Ω un domaine convexe, de mesure de Lebesgue unité, et $\gamma \in \mathcal{P}(\mathcal{P}(\Omega) \times \mathcal{P}(\Omega))$ une mesure de probabilité sur le produit $\mathcal{P}(\Omega) \times \mathcal{P}(\Omega)$ (satisfaisant une condition d'incompressibilité). On considère le problème*

$$\min_Q \left\{ \mathbb{E}_Q[A(\rho)] \ : \ Q \in \mathcal{P}(\Gamma), \ (e_0, e_1)\#Q = \gamma \ \text{ et } \ \forall t, \mathbb{E}_Q[\rho_t] = \mathcal{L} \right\},$$

où $(e_0, e_1) : \Gamma \to \mathcal{P}(\Omega) \times \mathcal{P}(\Omega)$ désigne l'évaluation aux instants $t = 0$ et $t = 1$.

Supposons qu'il existe une solution du problème pour laquelle $\mathcal{H}_Q$ (définie dans (A.7)) soit dans $L^1([0,1])$, et considérons alors la solution Q telle que la norme $L^1([0,1])$ de $\mathcal{H}_Q$ est minimale. Alors $\mathcal{H}_Q$ est une fonction convexe du temps.

Ce résultat, obtenu initialement dans l'article [Lav17], se trouve dans le Chapitre 6. Il peut être vu comme décevant car nous pouvons garantir la convexité de l'entropie moyenne seulement pour une solution des équations d'Euler incompressible. Après la publication de notre article, Baradat et Monsaingeon [BM18] ont prouvé, par une approche différente, au moins quand le domaine Ω est un tore, qu'en fait toute solution des équation d'Euler sous forme variationnelle possède une entropie moyenne temporellement convexe.

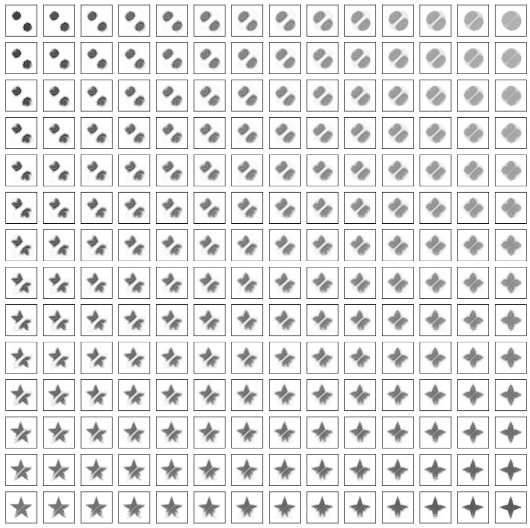

Figure A.4: Exemple d'application harmonique à valeurs dans l'espace de Wasserstein. Chaque petit carré correspond à la valeur de l'application en un point, qui est une mesure de probabilité (représentée par sa densité). Cette application est harmonique, c'est-à-dire qu'elle minimise l'énergie de Dirichlet parmi toutes les applications partageant les mêmes conditions aux bords.

Formellement la convexité de $\mathcal{H}_Q$ est vraie, c'est d'ailleurs au vu d'un calcul formel que Brenier avait formulé sa conjecture. Pour la démontrer rigoureusement, nous utilisons des arguments de discrétisation en temps similaires au deux chapitres précédents. C'est d'ailleurs parce que nous raisonnons par approximation que nous ne pouvons obtenir le résultat pour toutes les solutions du problème considéré, mais seulement pour une.

A.3 Applications harmoniques à valeurs dans l'espace de Wasserstein

Dans la deuxième partie de ce manuscrit, issue en majorité de l'article [Lav19], nous nous intéressons à des applications, et non plus seulement des courbes, à valeurs dans l'espace de Wasserstein. C'est-à-dire que nous considérons des applications $\boldsymbol{\mu} : D \to \mathcal{P}(\Omega)$ définies sur un domaine $D \subset \mathbb{R}^p$ prenant leurs valeurs dans $\mathcal{P}(\Omega)$. Le cas où D est un segment de $\mathbb{R}$ nous ramène aux problématiques précédentes. Dans le cas général, l'enjeu est la définition de l'équivalent de l'action A, que l'on appelle plutôt l'énergie de Dirichlet et qui sera notée Dir. Heuristiquement, $\mathrm{Dir}(\boldsymbol{\mu}) = \frac{1}{2} \int_D |\nabla \boldsymbol{\mu}|^2$, où $|\nabla \boldsymbol{\mu}|$ est mesuré à l'aide de la distance de Wasserstein. Une Dir définie, prenons D un domaine borné avec un bord ∂D Lipschitz. Pour peu que $\boldsymbol{\mu}_b : \partial D \to \mathcal{P}(\Omega)$ soit une application fixée définie sur le bord de D, on peut s'intéresser au *problème de Dirichlet*

$$\min_{\boldsymbol{\mu}} \left\{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} : D \to \mathcal{P}(\Omega) \text{ et } \boldsymbol{\mu} = \boldsymbol{\mu}_b \text{ sur } \partial D \right\}. \tag{A.8}$$

Par analogie avec le cas des applications à valeurs réelles, les solutions du problème de Dirichlet sont naturellement appelées les application *harmoniques* à valeurs dans l'espace de Wasserstein. Cette appellation rentre parfaitement dans le cadre du calcul d'Otto [Ott01], qui montre que formellement l'espace de Wasserstein est une variété Riemannienne de dimension infinie. Notons que si D est un segment, comme Dir coïncide alors avec A, (A.8) est identique à (A.6) et les applications harmoniques coïncident avec les géodésiques à vitesse constante.

Le premier enjeu est la définition de l'énergie de Dirichlet. Une définition, fondée sur une extension de la formule de Benamou-Brenier (A.3) a été proposée par Brenier dans [Bre03, Section 3]. D'un autre côté, l'énergie de Dirichlet et les applications harmoniques à valeurs dans les espaces métriques ont été étudiées dans les années 90 dans les travaux de Korevaar, Schoen et Jost [KS93, Jos94]. En plus de proposer une définition valable dans un cadre très général, ces auteurs ont montré qu'une étude approfondie est possible si l'espace cible possède une courbure négative au sens de Alexandrov. Or l'espace de Wasserstein a une courbure positive au sens de Alexandrov [AGS08, Section 7.3], de sorte que la définition de Korevaar, Schoen et Jost a toujours un sens, mais la plupart de leur théorie est inutilisable.

Notre première contribution est de montrer que ces deux définitions, celle se fondant sur une extension de la formulation de Benamou-Brenier et celle issue de la théorie des applications à valeurs dans les espaces métriques, coïncident : c'est l'objet de la Section 8.1. Plus précisément L'énergie de Dirichlet telle que proposée par Brenier prend la forme suivante.

Définition A.4. *Soit $\boldsymbol{\mu} : D \to \mathcal{P}(\Omega)$. On définit son énergie de Dirichlet par la formule*

$$\mathrm{Dir}(\boldsymbol{\mu}) = \min_{\mathbf{v}} \left\{ \iint_{D\times\Omega} \frac{1}{2}|\mathbf{v}|^2 \mathrm{d}\boldsymbol{\mu} \ : \ \mathbf{v} : D \times \Omega \to \mathbb{R}^{pd} \text{ et } \nabla_D \boldsymbol{\mu} + \nabla_\Omega \cdot (\boldsymbol{\mu}\mathbf{v}) = 0 \right\},$$

avec la convention que le minimum de l'ensemble vide est $+\infty$.

On pourra noter la ressemblance avec (A.3). Pour introduire la formulation métrique, commençons par définir l'énergie de Dirichlet au niveau ε, à savoir

$$\mathrm{Dir}_\varepsilon(\boldsymbol{\mu}) = C_p \iint_{D\times D} \frac{W_2^2(\boldsymbol{\mu}(x), \boldsymbol{\mu}(y))}{2\varepsilon^{p+2}} \mathbb{1}_{|x-y|\leqslant\varepsilon} \mathrm{d}x\mathrm{d}y, \tag{A.9}$$

où C_p est une constante de normalisation qui dépend de p la dimension du domaine D. Il est facile de vérifier que si l'espace de Wasserstein est remplacé par la droite réelle, l'énergie de Dirichlet au niveau ε converge simplement vers l'énergie de Dirichlet au sens usuel, à savoir l'intégrale du carré du gradient. Dans notre situation, nous arrivons au résultat suivant.

Théorème A.5. *On a*

$$\lim_{\varepsilon\to 0} \mathrm{Dir}_\varepsilon = \mathrm{Dir},$$

simplement et au sens de la Γ-convergence (le long de la suite $(\varepsilon_n)_{n\in\mathbb{N}} = (2^{-n})_{n\in\mathbb{N}}$).

Le membre de gauche correspond à l'énergie de Dirichlet au sens de Korevaar, Schoen et Jost, tandis que le membre de droite est celui de la Définition A.4. L'équivalence de ces définitions est assez facile au niveau formel (et donc dans le cas où tous les objets sont réguliers), la preuve de ce résultat repose sur des procédures d'approximation pour se ramener au cas régulier.

Nous signalons, sans le détailler ici, que si $\mathrm{Dir}(\boldsymbol{\mu}) < +\infty$ alors il est possible de donner un sens aux valeurs de $\boldsymbol{\mu}$ sur le bord de D, pour peu que ce dernier soit Lipschitz.

À partir de là, il est facile de montrer que le problème de Dirichlet (A.8) est bien posé si les données aux bords ont une certaine régularité, c'est l'objet de la Section 8.2.

Théorème A.6. *Soit $\boldsymbol{\mu}_b : \partial D \to (\mathcal{P}(\Omega), W_2)$ une application Lipschitz. Alors il existe au moins une solution au problème* (A.8).

L'hypothèse sur $\boldsymbol{\mu}_b$ permet de construire au moins un compétiteur d'énergie de Dirichlet finie grâce à une extension Lipschitz de $\boldsymbol{\mu}_b$ sur tout le domaine D. Nous soulignons que, puisque l'espace de Wasserstein a une courbure positive, l'existence d'une extension Lipschitz n'est pas automatique [Oht09] et repose sur des arguments *ad hoc* adaptés seulement à l'espace de Wasserstein. L'unicité de la solution reste une question ouverte : elle ne peut pas être vraie en toute généralité (c'est déjà faux si D est un segment), mais même avec des hypothèses de régularité supplémentaires sur $\boldsymbol{\mu}_b$ nous ne savons pas si nous pouvons conclure à l'unicité.

Du point de vue de la structure linéaire sur $\mathcal{P}(\Omega)$, le problème de Dirichlet est un problème convexe. En particulier, il lui est associé un problème dual qui s'écrit de la façon suivante. La lettre σ désigne la mesure de surface sur ∂D.

Théorème A.7. *Soit $\boldsymbol{\mu}_b : \partial D \to (\mathcal{P}(\Omega), W_2)$ une application Lipschitz. Alors on peut écrire*

$$\sup_{\varphi \in C^1(D\times\Omega,\mathbb{R}^p)} \left\{ \int_{\partial D} \left(\int_\Omega \varphi(x,\cdot) \mathrm{d}\boldsymbol{\mu}_b(x) \right) \sigma(\mathrm{d}x) \ : \ \nabla_D \cdot \varphi + \frac{|\nabla_\Omega \varphi|^2}{2} \leqslant 0 \ \textit{sur } D \times \Omega \right\}$$
$$= \min_{\boldsymbol{\mu}} \left\{ \mathrm{Dir}(\boldsymbol{\mu}) \ : \ \boldsymbol{\mu} : D \to \mathcal{P}(\Omega) \ \textit{et } \boldsymbol{\mu} = \boldsymbol{\mu}_b \ \textit{sur } \partial D \right\}.$$

Dans ce problème dual apparaît une équation d'Hamilton-Jacobi dans laquelle l'inconnue est φ une fonction à valeurs vectorielles, tandis qu'il ne porte sur elle qu'une contrainte scalaire. Nous ne savons pas si le supremum dans le membre de gauche est atteint, pas même le sens dans lequel il pourrait être atteint : déjà dans le cas où D est un segment, il faut par exemple autoriser des solutions de viscosité.

Étant un problème convexe, le problème de Dirichlet se prête bien à une approximation numérique. Plus précisément, nous proposons dans le Chapitre 11 un problème convexe de dimension finie qui, formellement, est une discrétisation du problème de Dirichlet. La preuve de la convergence des minimiseurs de ce problème de dimension finie vers les applications harmoniques reste néanmoins une question ouverte. Grâce à des méthodes de "*splitting proximal*", plus précisément l'*Alternating Direction Method of Multipliers* (ADMM), nous pouvons résoudre algorithmiquement le problème de dimension finie en un temps long (mais raisonnable) et nous nous en sommes servis pour produire les illustrations qui parsèment ce manuscrit. Nous soulignons qu'une telle méthode n'est en rien nouvelle, elle a déjà utilisé pour approcher numériquement des géodésiques dans l'espace de Wasserstein : c'est même pour cette raison que Benamou et Brenier ont introduit la formulation (A.3) pour l'action d'une courbe [BB00].

Le principe de superposition [AGS08, Section 8.2] est un outil très pratique pour étudier les courbes à valeurs dans l'espace de Wasserstein. Lorsque D cesse d'être un segment (et que Ω n'est pas de dimension 1), alors il est mis en échec, comme décrit dans la Section 8.3, et cela répond d'ailleurs à une question soulevée par Brenier [Bre03, Problem 3.1]. En bref, il n'existe pas de point de vue Lagrangien pour les applications harmoniques à valeurs dans l'espace de Wasserstein, seule la formulation eulérienne est disponible. Cela explique que l'étude des applications soit substantiellement plus difficile que celle des courbes.

Le résultat théorique principal auquel nous arrivons, pour les applications harmoniques, est qu'un principe du maximum reste valide : c'est l'objet du Chapitre 9. Bien sûr il n'y a pas d'ordre canonique sur l'ensemble des mesures de probabilité. Pour donner un sens au principe du maximum, il faut prendre une application harmonique $\boldsymbol{\mu} : D \to \mathcal{P}(\Omega)$ et la composer avec $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ convexe le long des géodésiques (généralisées). Si D est un segment (donc $\boldsymbol{\mu}$ est une géodésique), on obtient une fonction convexe par définition. Dans le cas général, on obtient une fonction sous-harmonique, c'est-à-dire $\Delta(F \circ \boldsymbol{\mu}) \geqslant 0$. Cette propriété n'est pas si surprenante : dans le cadre lisse, à savoir si l'on prend une application harmonique à valeurs dans une variété Riemannienne, et qu'on la compose avec une application convexe à valeurs réelles, on obtient une application sous-harmonique, comme remarqué par Ishihara [Ish78]. Dans le cas où la variété Riemannienne est l'espace de Wasserstein (c'est-à-dire celui que nous étudions), nous arrivons à prouver le résultat suivant.

Théorème A.8. *Soit* $F : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ *une fonctionnelle convexe le long des géodésiques généralisées (ainsi que quelques propriétés de régularité supplémentaires que l'on ne détaille pas). Soit* $\boldsymbol{\mu}_b : \partial D \to (\mathcal{P}(\Omega), W_2)$ *une application Lipschitz telle que* $\sup_{\partial D}(F \circ \boldsymbol{\mu}_b) < +\infty$.

Alors il existe au moins une solution $\boldsymbol{\mu} : D \to \mathcal{P}(\Omega)$ *du problème de Dirichlet avec conditions au bord* $\boldsymbol{\mu}_b$ *telle que* $(F \circ \boldsymbol{\mu}) : D \to \mathbb{R}$ *est sous-harmonique au sens des distributions dans* $\mathring{D}$ *et*

$$\operatorname*{ess\,sup}_{D}(F \circ \boldsymbol{\mu}) \leqslant \sup_{\partial D}(F \circ \boldsymbol{\mu}_b). \tag{A.10}$$

L'équation (A.10) n'est autre que le principe du maximum, il ne découle pas directement de l'équation vérifiée par $F \circ \boldsymbol{\mu}$ à l'intérieur de D à cause de l'éventuelle discontinuité de $F \circ \boldsymbol{\mu}$ au bord de D. Ce résultat est bien sûr décevant car nous pouvons affirmer quelque chose seulement pour une solution du problème de Dirichlet, pas pour toutes : la raison est que nous procédons par approximation et que nous ne savons pas garantir l'unicité dans le problème de Dirichlet.

Comme c'est l'objet principal de cette partie, nous présentons la stratégie de notre démonstration. L'idée consiste à minimiser Dir_ε l'énergie de Dirichlet au niveau ε définie dans (A.9), grâce au Théorème A.5 nous savons que si $\boldsymbol{\mu}_\varepsilon$ est un minimiseur de Dir_ε (avec des conditions aux bords appropriées) alors $\boldsymbol{\mu}_\varepsilon$ va converger vers une solution du problème de Dirichlet lorsque $\varepsilon \to 0$. Par optimalité de $\boldsymbol{\mu}_\varepsilon$, on arrive à la conclusion que pour presque tout x dans D, la mesure $\boldsymbol{\mu}_\varepsilon(x)$ est un barycentre des $\boldsymbol{\mu}_\varepsilon(y)$ pour $y \in B(x, \varepsilon)$: cette mesure minimise la fonctionnelle

$$\mu \mapsto \int_{B(y,\varepsilon)} W_2^2(\mu, \boldsymbol{\mu}_\varepsilon(y))\mathrm{d}y. \tag{A.11}$$

On en tire un lien entre barycentre et applications harmoniques (vrai dans un cadre très général, voir par exemple [Jos94]) : la mesure $\boldsymbol{\mu}(x)$ est le barycentre de ses voisins sur une boule de taille ε, du moins dans la limite $\varepsilon \to 0$. Comme F est convexe le long des géodésiques généralisées, on peut utiliser l'inégalité de Jensen pour dire que

$$F(\boldsymbol{\mu}_\varepsilon(x)) \leqslant \int_{B(y,\varepsilon)} F(\boldsymbol{\mu}_\varepsilon(y))\mathrm{d}y. \tag{A.12}$$

Plus précisément, pour établir cette inégalité de Jensen, on part de la caractérisation du barycentre par (A.11) et on perturbe le barycentre en lui faisant suivre le flot gradient de la fonctionnelle F. L'utilisation de l'inégalité EVI, rappelée dans (A.5), donne exactement l'inégalité de Jensen. Il reste alors à passer à la limite $\varepsilon \to 0$ dans (A.12) pour obtenir $\Delta(F \circ \boldsymbol{\mu}) \geqslant 0$ au sens des distributions ainsi que (A.10). C'est à ce moment que des considérations techniques, que nous ne détaillons pas ici, entrent en jeu.

Terminons ce résumé en expliquant ce que l'on peut dire de plus dans certains cas particuliers, qui sont développés dans le Chapitre 10.

L'espace Ω s'injecte de façon isométrique dans $(\mathcal{P}(\Omega), W_2)$: il suffit d'associer à $x \in \Omega$ la mesure δ_x, à savoir la masse de Dirac concentrée en x. En particulier, si $f : D \to \Omega$, on peut naturellement la voir comme une application $\boldsymbol{\mu}_f : x \in D \mapsto \delta_{f(x)} \in \mathcal{P}(\Omega)$ à valeurs dans l'espace de Wasserstein. En un sens, une application à valeurs dans l'espace de Wasserstein peut être vue comme la relaxation d'une application à valeurs dans Ω [SGB13]. Une question que l'on peut se poser est de savoir si cette relaxation apporte de meilleurs compétiteurs pour le problème de Dirichlet. Brenier [Bre03, Theorem 3.1] a déjà apporté une réponse négative.

Proposition A.9. *Soit* $\boldsymbol{\mu}_b : x \in \partial D \mapsto \delta_{f_b(x)} \in (\mathcal{P}(\Omega), W_2)$ *une application Lipschitz à valeurs dans l'ensemble des masses de Dirac. Alors il existe une unique solution* $\boldsymbol{\mu}$ *au problème de Dirichlet avec conditions au bord* $\boldsymbol{\mu}_b$ *et elle s'écrit* $\boldsymbol{\mu}(x) = \delta_{f(x)}$*, où* f *est l'unique extension harmonique de* f_b*.*

La preuve de Brenier repose sur l'exhibition de la solution du problème dual dans ce cas. Nous ne l'avions pas encore signalé mais si $\boldsymbol{\mu} : x \mapsto \delta_{f(x)}$ pour un certain $f : D \to \Omega$ alors

$$\mathrm{Dir}(\boldsymbol{\mu}) = \int_D \frac{1}{2}|\nabla f|^2.$$

En d'autres termes, l'énergie de Dirichlet pour les applications à valeurs dans l'espace de Wasserstein étend bien la définition usuelle de l'énergie de Dirichlet pour les fonctions.

Un autre exemple révèle plus de la géométrie de l'espace de Wasserstein : il s'agit des applications qui sont à valeurs dans l'ensemble des mesures gaussiennes. Comme nous travaillons avec l'hypothèse que Ω est *borné*, nous ne pouvons pas vraiment considérer des mesures gaussiennes. À la place, prenons $\rho \in \mathcal{P}(\mathbb{R}^d)$ une mesure radiale, à support compact, et dont la matrice de covariance est l'identité. Pour des raisons techniques, nous supposons aussi que son entropie est finie. On note $S_d(\mathbb{R})$ l'ensemble des matrices $d \times d$ symétriques, ainsi que $S_d^+(\mathbb{R})$ l'ensemble des matrices symétrique semi-définies positives.

Définition A.10. *Pour* $A \in S_d^+(\mathbb{R})$*, on note* $\rho_A \in \mathcal{P}(\mathbb{R}^d)$ *la mesure image de* ρ *par l'application* $x \mapsto Ax$*.*

L'ensemble des ρ_A*, pour* $A \in S_d^+(\mathbb{R})$ *est noté par* $\mathcal{P}_{\mathrm{ec}}(\mathbb{R}^d)$ *et appelé une famille de distributions aux contours elliptiques. On utilisera aussi la notation* $\mathcal{P}_{\mathrm{ec}}(\Omega) = \mathcal{P}(\Omega) \cap \mathcal{P}_{\mathrm{ec}}(\mathbb{R}^d)$

Par exemple, si ρ est l'indicatrice (renormalisée) d'un disque en dimension 2, alors les ρ_A sont des indicatrices (renormalisées) d'ellipses. Plus généralement, le paramétrage $A \mapsto \rho_A$ est injectif puisque A peut se retrouver en prenant la racine carré de $\mathrm{cov}(\rho_A)$ la matrice de covariance de ρ_A. Le cas gaussien s'obtiendrait en prenant pour ρ la mesure gaussienne centrée réduite. L'ensemble des matrices symétriques définies positives muni de la distance $(A,B) \mapsto W_2(\rho_A, \rho_B)$ forme une variété Riemannienne de dimension finie. Elle n'est cependant pas complète car les matrices singulières (au voisinage desquelles la métrique dégénère) sont à distance finie des matrices inversibles. Notre résultat principal est le suivant.

Théorème A.11. *Soit* $\boldsymbol{\mu}_b : \partial D \to \mathcal{P}_{\mathrm{ec}}(\Omega)$ *une application Lipschitz (vérifiant certaines conditions de compatibilité entre leur support et* Ω *non spécifiées) telle que* $\det(\mathrm{cov}(\boldsymbol{\mu}_b(x))) > 0$ *pour tout* $x \in \partial D$ *and définissons* $\mathbf{A}_b(x) = \mathrm{cov}(\boldsymbol{\mu}_b(x))^{1/2}$ *pour tout* $x \in \partial D$*.*

Alors il existe une unique solution $\boldsymbol{\mu} : D \to \mathcal{P}(\Omega)$ *au problème de Dirichlet avec valeurs au bord* $\boldsymbol{\mu}_b$ *et* $\boldsymbol{\mu}$ *prend ses valeurs dans* $\mathcal{P}_{\mathrm{ec}}(\Omega)$*. De plus si* $\mathbf{A} : D \to S_d^+(\mathbb{R})$ *est défini par* $\mathbf{A}(x) := \mathrm{cov}(\boldsymbol{\mu}(x))^{1/2}$ *pour* $x \in D$ *alors on a :*

(i) $\operatorname*{ess\,inf}_{x \in D} \det(\mathbf{A}(x)) > 0$*.*

(ii) L'application $\mathbf{A}$ *est régulière (spécifiquement* C^∞*) dans l'intérieur de* D *et la régularité jusqu'au bord est vraie si* $\mathbf{A}_b$ *et* ∂D *sont assez réguliers.*

(iii) L'application $\mathbf{A}$ *vérifie une équation aux dérivées partielles explicite (mais que l'on ne reproduit pas ici).*

La preuve de ce résultat repose sur deux idées simples. La première est l'existence d'une rétraction sur l'ensemble $\mathcal{P}_{\mathrm{ec}}(\Omega)$ (i.e. une application 1-Lipschitz valant l'identité sur $\mathcal{P}_{\mathrm{ec}}(\Omega)$) de sorte qu'en prenant n'importe quel compétiteur et en le composant avec cette rétraction on diminue son énergie de Dirichlet tout en laissant les conditions au bord inchangées : cela garantit l'existence d'au moins une solution au problème de Dirichlet à valeurs dans $\mathcal{P}_{\mathrm{ec}}(\Omega)$. Puis on utilise le principe du maximum (Théorème A.8) en prenant pour F l'entropie de Boltzmann (A.4). En effet, pour un tel choix de F,

$$F(\rho_A) = -\ln(\det A) + C,$$

où la constante C dépend seulement de ρ. Ainsi, le principe du maximum pour l'entropie se transforme en principe du minimum pour le déterminant de $\mathbf{A}$: le point (i) est prouvé. Une fois ces deux arguments utilisés, on sait en fait que $\mathbf{A}$ est une application harmonique à valeurs dans une variété Riemannienne de dimension finie, et des arguments standards permettent d'en déduire l'équation aux dérivées partielles qu'elle satisfait ainsi que d'en inférer sa régularité.

Bibliography

[AB06] C. D. Aliprantis and K. Border. *Infinite Dimensional Analysis: a Hitchhiker's Guide*. Springer, 2006.

[AC11] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[ACB17] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, 2017.

[AF08] L. Ambrosio and A. Figalli. On the regularity of the pressure field of Brenier's weak solutions to incompressible Euler equations. *Calculus of Variations and Partial Differential Equations*, 31(4):497–509, 2008.

[AF09] L. Ambrosio and A. Figalli. Geodesics in the space of measure-preserving maps and plans. *Archive for rational mechanics and analysis*, 194(2):421–462, 2009.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Birkhäuser Verlag, 2008.

[AJ00] F. J. Almgren Jr. *Almgren's big regularity paper: Q-valued functions minimizing Dirichlet's integral and the regularity of area-minimizing rectifiable currents up to codimension 2*. World Scientific, 2000.

[Amb04] L. Ambrosio. Transport equation and Cauchy problem for BV vector fields. *Inventiones mathematicae*, 158(2):227–260, 2004.

[Are02] W. Arendt. Semigroups and evolution equations: functional calculus, regularity and kernel estimates. *Handbook of Differential Equations: Evolutionary Equations*, 1:1–85, 2002.

[Arn66] V. Arnold. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. In *Annales de l'institut Fourier*, volume 16, pages 319–361, 1966.

[AT03] L. Ambrosio and P. Tilli. *Topics on Analysis in Metric Spaces*. Oxford Lecture Series in Mathematics and Its Applications. 2003.

[Bar97] F. Barthe. *Inégalités fonctionnelles et géométriques obtenues par transport des mesures*. PhD thesis, Marne-la-Vallée, 1997.

[Bar19] A. Baradat. Continuous dependence of the pressure field with respect to endpoints for ideal incompressible fluids. *Calculus of Variations and Partial Differential Equations*, 58(1):25, 2019.

[BB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[BC15] J.-D. Benamou and G. Carlier. Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *Journal of Optimization Theory and Applications*, 167(1):1–26, 2015.

[BCDMN18] J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna. An entropy minimization approach to second-order variational mean-field games. *arXiv preprint arXiv:1807.09078*, 2018.

[BCL16] J.-D. Benamou, G. Carlier, and M. Laborde. An augmented Lagrangian approach to Wasserstein gradient flows and applications. *ESAIM: Proceedings and Surveys*, 54:1–17, 2016.

[BCM05] M. Bernot, V. Caselles, and J.-M. Morel. Traffic plans. *Publicacions Matemàtiques*, 49(2):417–451, 2005.

[BCM16] J.-D. Benamou, F. Collino, and J.-M. Mirebeau. Monotone and Consistent discretization of the Monge-Ampère operator. *Mathematics of computation*, 85(302):2743–2775, 2016.

[BCS17] J.-D. Benamou, G. Carlier, and F. Santambrogio. Variational Mean Field Games. In *Active Particles, Volume 1 : Advances in Theory, Models, and Applications*, pages 141–171. Springer, 2017.

[BFH+03] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskiĭ. Reconstruction of the early Universe as a convex optimization problem. *Monthly Notices of the Royal Astronomical Society*, 346(2):501–524, 2003.

[BFO14] J.-D. Benamou, B. D. Froese, and A. M. Oberman. Numerical solution of the optimal transportation problem using the monge–ampère equation. *Journal of Computational Physics*, 260:107–126, 2014.

[BFS09] M. Bernot, A. Figalli, and F. Santambrogio. Generalized solutions for the Euler equations in one and two dimensions. *Journal de mathématiques pures et appliquées*, 91(2):137–155, 2009.

[BJL18] R. Bhatia, T. Jain, and Y. Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 2018.

[BJO09] G. Buttazzo, C. Jimenez, and É. Oudet. An optimization problem for mass transportation with congested dynamics. *SIAM Journal on Control and Optimization*, 48(3):1961–1976, 2009.

[BM18] A. Baradat and L. Monsaingeon. Small noise limit and convexity for generalized incompressible flows, Schrödinger problems, and optimal transport. *arXiv preprint arXiv:1810.12036*, 2018.

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[Bre87] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[Bre89] Y. Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.

[Bre91] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[Bre99] Y. Brenier. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Communications on pure and applied mathematics*, 52(4):411–452, 1999.

[Bre01] Y. Brenier. A Monge-Kantorovich Approach to the Maxwell Equations. In H. Freistühler and G. Warnecke, editors, *Hyperbolic Problems: Theory, Numerics, Applications*, pages 179–186, Basel, 2001. Birkhäuser Basel.

[Bre03] Y. Brenier. Extended Monge-Kantorovich theory. In *Optimal transportation and applications*, pages 91–121. Springer, 2003.

[BV04] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[Car10] P. Cardaliaguet. Notes on mean field games, 2010. `https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf`.

[Car15] P. Cardaliaguet. Weak solutions for first order mean field games with local coupling. In *Analysis and geometry in control theory and its applications*, pages 111–158. Springer, 2015.

[CDLL15] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. The master equation and the convergence problem in mean field games. *arXiv preprint arXiv:1509.02505*, 2015.

[CFTR17] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2017.

[CG15] P. Cardaliaguet and J. P. Graber. Mean field games systems of first order. *ESAIM: Control, Optimisation and Calculus of Variations*, 21(3):690–722, 2015.

[CGM17] C. Cancès, T. O. Gallouët, and L. Monsaingeon. Incompressible immiscible multiphase flows in porous media: a variational approach. *Analysis & PDE*, 10(8):1845–1876, 2017.

[Chi07] D. Chiron. On the definitions of Sobolev and BV spaces into singular spaces and the trace problem. *Communications in Contemporary Mathematics*, 9(04):473–513, 2007.

[CM89] J. A. Cuesta and C. Matrán. Notes on the Wasserstein metric in Hilbert spaces. *The Annals of Probability*, 17(3):1264–1276, 1989.

[CM17] F. Cavalletti and A. Mondino. Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds. *Inventiones mathematicae*, 208(3):803–849, 2017.

[CMS16] P. Cardaliaguet, A. R. Mészáros, and F. Santambrogio. First order mean field games with density constraints: pressure equals price. *SIAM Journal on Control and Optimization*, 54(5):2672–2709, 2016.

[CPT15] P. Cardaliaguet, A. Porretta, and D. Tonon. Sobolev regularity for the first order Hamilton–Jacobi equation. *Calculus of Variations and Partial Differential Equations*, 54(3):3037–3065, 2015.

[Cut13] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.

[Dan83] G. B. Dantzig. Reminiscences about the origins of linear programming. In *Mathematical Programming The State of the Art*, pages 78–86. Springer, 1983.

[DF12] S. Daneri and A. Figalli. Variational models for the incompressible Euler equations. *HCDTE Lecture Notes. Part II. Nonlinear Hyperbolic PDEs, Dispersive and Transport Equations. AIMS Book Series, Applied Mathematics, to appear*, 2012.

[DL82] D.C. Dowson and B. V. Landau. The Fréchet distance between multivariate normal distributions. *Journal of multivariate analysis*, 12(3):450–455, 1982.

[DL89] R. J. DiPerna and P.-L. Lions. Ordinary differential equations, transport theory and Sobolev spaces. *Inventiones mathematicae*, 98(3):511–547, 1989.

[DLS11] C. De Lellis and E. N. Spadaro. *Q-valued functions revisited*. American Mathematical Soc., 2011.

[Dob70] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970.

[Dob79] R. L. Dobrushin. Vlasov equations. *Functional Analysis and its Applications*, 13(2):115–123, 1979.

[EG92] L. C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. Taylor & Francis, 1992.

[ERSS17] M. Erbar, M. Rumpf, B. Schmitzer, and S. Simon. Computation of optimal transport on discrete metric measure spaces. *arXiv preprint arXiv:1707.06859*, 2017.

[ES64] J. Eells and J. H. Sampson. Harmonic mappings of Riemannian manifolds. *American journal of mathematics*, 86(1):109–160, 1964.

[Eva10] L. C. Evans. Partial differential equations. 2010.

[Fed70] H. Federer. The singular sets of area minimizing rectifiable currents with codimension one and of area minimizing flat chains modulo two with arbitrary codimension. *Bulletin of the American Mathematical Society*, 76(4):767–771, 1970.

[FP92] I. Fonseca and G. Parry. Equilibrium configurations of defective crystals. *Archive for rational mechanics and analysis*, 120(3):245–283, 1992.

[Fré48] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10(3):215–310, 1948.

[Fré50] M. Fréchet. *Recherches théoriques modernes sur le calcul des probabilités*, chapter Distance de deux variables aléatoires et distance de deux lois de probabilité. Gauthier-Villars, 2 edition, 1950.

[Fré57] M. Fréchet. Sur la distance de deux lois de probabilité. *Comptes rendus hebdomadaires des séances de l'académie des sciences*, 244(6):689–692, 1957.

[FSV+18] J. Feydy, T. Séjourné, F.-X. Vialard, S.-I. Amari, A. Trouvé, and G. Peyré. Interpolating between optimal transport and MMD using Sinkhorn divergences. *arXiv preprint arXiv:1810.08278*, 2018.

[FZM+15] C. Frogner, C. Zhang, H. Mobahi, M. Araya, and T. A. Poggio. Learning with a Wasserstein loss. In *Advances in Neural Information Processing Systems*, pages 2053–2061, 2015.

[GAHE18] T. Gao, S. Asoodeh, Y. Huang, and J. Evans. Wasserstein Soft Label Propagation on Hypergraphs: Algorithm and Generalization Error Bounds. *arXiv preprint arXiv:1809.01833*, 2018.

[Gal16] A. Galichon. *Optimal transport methods in economics*. Princeton University Press, 2016.

[Gel90] M. Gelbrich. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990.

[Ghy12] É. Ghys. Gaspard Monge – Le mémoire sur les déblais et les remblais. Images des Mathématiques, CNRS, 2012.

[Gig13] N. Gigli. The splitting theorem in non-smooth context. *arXiv preprint arXiv:1302.5555*, 2013.

[Gig15] N. Gigli. On the differential structure of metric measure spaces and applications. *American Mathematical Society*, 236(1113), 2015.

[GKM18] P. Gladbach, E. Kopfer, and J. Maas. Scaling limits of discrete optimal transport. *arXiv preprint arXiv:1809.01092*, 2018.

[GM13] N. Gigli and J. Maas. Gromov–Hausdorff Convergence of Discrete Transportation Metrics. *SIAM Journal on Mathematical Analysis*, 45(2):879–899, 2013.

[GM18a] T. O. Gallouët and Q. Mérigot. A Lagrangian scheme à la Brenier for the incompressible Euler equations. *Foundations of Computational Mathematics*, 18(4):835–865, 2018.

[GM18b] J. P. Graber and A. R. Mészáros. Sobolev regularity for first order mean field games. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*. Elsevier, 2018.

[GNT08] W. Gangbo, T. Nguyen, and A. Tudorascu. Hamilton-Jacobi equations in the Wasserstein space. *Methods and Applications of Analysis*, 15(2):155–184, 2008.

[GŚ15] W. Gangbo and A. Święch. Existence of a solution to an equation arising from the theory of mean field games. *Journal of Differential Equations*, 259(11):6573–6643, 2015.

[Gui03] K. Guittet. On the time-continuous mass transport problem and its approximation by augmented Lagrangian techniques. *SIAM Journal on Numerical Analysis*, 41(1):382–399, 2003.

[HMC06] M. Huang, R. P. Malhamé, and P. E. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.

[Hug16] R. Hug. *Analyse mathématique et convergence d'un algorithme pour le transport optimal dynamique : cas des plans de transports non réguliers, ou soumis à des contraintes*. PhD thesis, 2016.

[HW08] F. Hélein and J. C. Wood. Harmonic maps. *Handbook of global analysis*, 1213:417–491, 2008.

[Ish78] T. Ishihara. A mapping of Riemannian manifolds which preserves harmonic functions. 1978.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[Jos94] J. Jost. Equilibrium maps between metric spaces. *Calculus of Variations and Partial Differential Equations*, 2(2):173–204, 1994.

[Jos08] J. Jost. *Riemannian geometry and geometric analysis*. Springer, 2008.

[Kan48] L. Kantorovich. On a problem of Monge. In *CR (Doklady) Acad. Sci. URSS (NS)*, volume 3, pages 225–226, 1948.

[Kan58] L. Kantorovich. On the translocation of masses. *Management Science*, 5(1):1–4, 1958.

[KMT16] J. Kitagawa, Q. Mérigot, and B. Thibert. Convergence of a Newton algorithm for semi-discrete optimal transport. *arXiv preprint arXiv:1603.05579*, 2016.

[KP15] Y.-H. Kim and B. Pass. Nonpositive curvature, the variance functional, and the Wasserstein barycenter. *arXiv preprint arXiv:1503.06460*, 2015.

[KP17] Y.-H. Kim and B. Pass. Wasserstein barycenters over Riemannian manifolds. *Advances in Mathematics*, 307:640–683, 2017.

[KPT+16] S. Kolouri, S. Park, M. Thorpe, D. Slepčev, and G. K. Rohde. Transport-based analysis, modeling, and learning from signal and data distributions. *arXiv preprint arXiv:1609.04767*, 2016.

[KR58] L. Kantorovich and G. S. Rubinstein. On a space of completely additive functions. *Vestnik Leningrad. Univ*, 13(7):52–59, 1958.

[KS84] M. Knott and C. S. Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43(1):39–49, 1984.

[KS93] N. J. Korevaar and R. Schoen. Sobolev spaces and harmonic maps for metric space targets. *Communications in Analysis and Geometry*, 1(4):561–659, 1993.

[Lav17] H. Lavenant. Time-convexity of the entropy in the multiphasic formulation of the incompressible euler equation. *Calculus of Variations and Partial Differential Equations*, 56(6):170, sep 2017.

[Lav19] H. Lavenant. Harmonic mappings valued in the Wasserstein space. *Journal of Functional Analysis*, 277(3):688 – 785, 2019.

[LCCS18] H. Lavenant, S. Claici, E. Chien, and J. Solomon. Dynamical Optimal Transport on Discrete Surfaces. *ACM Transactions on Graphics*, 37(6), 2018.

[Lév25] P. Lévy. *Calcul des probabilités*. Gauthier-Villars, 1925.

[Lév15] B. Lévy. A numerical algorithm for L^2 semi-discrete optimal transport in 3D. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1693–1715, 2015.

[Lio12] P.-L. Lions. Jeux à champ moyen et applications. Videos available on the website of Coll. de France, 2006-12.

[Lis07] S. Lisini. Characterization of absolutely continuous curves in Wasserstein spaces. *Calculus of Variations and Partial Differential Equations*, 28(1):85–120, 2007.

[LL06a] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. I–le cas stationnaire. *Comptes Rendus Mathématique*, 343(9):619–625, 2006.

[LL06b] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. II–Horizon fini et contrôle optimal. *Comptes Rendus Mathématique*, 343(10):679–684, 2006.

[LL07] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[Loe06] G. Loeper. The reconstruction problem for the Euler-Poisson system in cosmology. *Archive for rational mechanics and analysis*, 179(2):153–216, 2006.

[LR86] P.-L. Lions and J.-C. Rochet. Hopf formula and multitime Hamilton-Jacobi equations. *Proceedings of the American Mathematical Society*, 96(1):79–84, 1986.

[LS97] U. Lang and V. Schroeder. Kirszbraun's theorem and metric spaces of bounded curvature. *Geometric & Functional Analysis GAFA*, 7(3):535–560, 1997.

[LS18] H. Lavenant and F. Santambrogio. Optimal density evolution with congestion: L^∞ bounds via flow interchange techniques and applications to variational Mean Field Games. *Communications in Partial Differential Equations*, 43(12):1761–1802, 2018.

[LS19] H. Lavenant and F. Santambrogio. New estimates on the regularity of the pressure in density-constrained mean field games. *Journal of the London Mathematical Society*, 2019.

[Lu17] Z. Lu. *Properties of Soft Maps on Riemannian Manifolds*. PhD thesis, New York University, 2017.

[LV09] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.

[Maa11] J. Maas. Gradient flows of the entropy for finite Markov chains. *Journal of Functional Analysis*, 261(8):2250 – 2292, 2011.

[McC94] R. J. McCann. *A convexity theory for interacting gases and equilibrium crystals*. PhD thesis, Princeton University, 1994.

[McC97] R. J. McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.

[McC01] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geometric & Functional Analysis GAFA*, 11(3):589–608, 2001.

[McL00] W. McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge University Press, 2000.

[Mér11] Q. Mérigot. A multiscale approach to optimal transport. In *Computer Graphics Forum*, volume 30, pages 1583–1592. Wiley Online Library, 2011.

[MMP18] L. Malago, L. Montrucchio, and G. Pistone. Wasserstein-Riemannian Geometry of Positive-definite Matrices. *arXiv preprint arXiv:1801.09269*, 2018.

[MMS09] D. Matthes, R. J. McCann, and G. Savaré. A family of nonlinear fourth order equations of gradient flow type. *Communications in Partial Differential Equations*, 34(11):1352–1397, 2009.

[MMT18] J. Meyron, Q. Mérigot, and B. Thibert. Light in Power: A General and Parameter-free Algorithm for Caustic Design. *ACM Transactions on Graphics*, 37(6), 2018.

[Mon81] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mémoires de l'Académie Royale des Sciences*, 1781.

[Mor59] J.-J. Moreau. Une méthode de cinématique fonctionnelle en hydrodynamique. *Comptes rendus hebdomadaires des séances de l'académie dess sciences*, 249(21):2156–2158, 1959.

[Mos60] J. Moser. A new proof of De Giorgi's theorem concerning the regularity problem for elliptic differential equations. *Communications on Pure and Applied Mathematics*, 13(3):457–468, 1960.

[MRCS10] B. Maury, A. Roudneff-Chupin, and F. Santambrogio. A macroscopic crowd motion model of gradient flow type. *Mathematical Models and Methods in Applied Sciences*, 20(10):1787–1821, 2010.

[MS03] J. Morovic and P.-L. Sun. Accurate 3d image colour histogram transformation. *Pattern Recognition Letters*, 24(11):1725–1735, 2003.

[MVN53] O. Morgenstern and J. Von Neumann. *Theory of games and economic behavior*. Princeton university press, 1953.

[Oht09] S.-I. Ohta. Extending Lipschitz and Hölder maps between metric spaces. *Positivity*, 13(2):407–425, 2009.

[Orn74] D. S. Ornstein. *Ergodic theory, randomness, and dynamical systems*. Yale UP, 1974.

[Ott98] F. Otto. Dynamics of labyrinthine pattern formation in magnetic fluids: A mean-field theory. *Archive for Rational Mechanics and Analysis*, 141(1):63–103, 1998.

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

[PC17] G. Peyré and M. Cuturi. Computational optimal transport, 2017.

[Pey15] G. Peyré. Entropic approximation of Wasserstein gradient flows. *SIAM Journal on Imaging Sciences*, 8(4):2323–2351, 2015.

[PPO14] N. Papadakis, G. Peyré, and É. Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[Rac85] S. T. Rachev. The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory of Probability & Its Applications*, 29(4):647–676, 1985.

[Res97] Y. G. Reshetnyak. Sobolev-type classes of functions with values in a metric space. *Siberian Mathematical Journal*, 38(3):657–675, 1997.

[Res04] Y. G. Reshetnyak. Sobolev-type classes of functions with values in a metric space. II. *Siberian Mathematical Journal*, 45(4):709–721, 2004.

[RR90] L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L2-distance. *Journal of Multivariate Analysis*, 32(1):48–54, 1990.

[RR98] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory*, volume 1. Springer Science & Business Media, 1998.

[RTG00] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[San15] F. Santambrogio. Optimal transport for applied mathematicians. *Progress in Nonlinear Differential Equations and their applications*, 87, 2015.

[San18] F. Santambrogio. Regularity via duality in calculus of variations and degenerate elliptic PDEs . *Journal of Mathematical Analysis and Applications*, 2018.

[SDGP+15] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics*, 34(4), 2015.

[Ser94] T. Serbinowski. Boundary regularity of harmonic maps to nonpositively curved metric spaces. *Communications in Analysis and Geometry*, 2(1):139–153, 1994.

[SGB13] J. Solomon, L. Guibas, and A. Butscher. Dirichlet energy for analysis and synthesis of soft maps. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing*, pages 197–206. Eurographics Association, 2013.

[SNB+12] J. Solomon, A. Nguyen, A. Butscher, M. Ben-Chen, and L. Guibas. Soft maps between surfaces. In *Computer Graphics Forum*, volume 31, pages 1617–1626. Wiley Online Library, 2012.

[SRGB14] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher. Wasserstein propagation for semi-supervised learning. In *International Conference on Machine Learning*, pages 306–314, 2014.

[Stu03] K.-T. Sturm. Probability measures on metric spaces of nonpositive. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, 338:357, 2003.

[Stu05] K.-T. Sturm. A semigroup approach to harmonic maps. *Potential Analysis*, 23(3):225–277, 2005.

[Stu06] K.-T. Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.

[SU82] R. Schoen and K. Uhlenbeck. A regularity theory for harmonic maps. *Journal of Differential Geometry*, 17(2):307–335, 1982.

[SU83] R. Schoen and K. Uhlenbeck. Boundary regularity and the Dirichlet problem for harmonic maps. *Journal of Differential Geometry*, 18(2):253–268, 1983.

[Tak11] A. Takatsu. Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, 48(4):1005–1026, 2011.

[Ver06] A. M. Vershik. Kantorovich metric: Initial history and little-known applications. *Journal of Mathematical Sciences*, 133(4):1410–1417, 2006.

[Vil03] C. Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[Vil08] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[VL18] T. Vogt and J. Lellmann. Measure-valued variational models with applications to diffusion-weighted imaging. *Journal of Mathematical Imaging and Vision*, 60(9):1482–1502, 2018.

[Was69] L. N. Wasserstein. Markov processes over denumerable products of spaces describing large systems of automata. *Problems of Information Transmission*, 5(3):47–52, 1969.

École doctorale de mathématiques Hadamard (EDMH)

Titre : Courbes et applications optimales à valeurs dans l'espace de Wasserstein

Mots Clefs : Calcul des variations, Transport optimal, Régularité elliptique, Analyse dans les espaces métriques.

Résumé : L'espace de Wasserstein est l'ensemble des mesures de probabilité définies sur un domaine fixé et muni de la distance de Wasserstein quadratique. Dans ce travail, nous étudions des problèmes variationnels dans lesquels les inconnues sont des applications à valeurs dans l'espace de Wasserstein.

Quand l'espace de départ est un segment, c'est-à-dire quand les inconnues sont des courbes à valeurs dans l'espace de Wasserstein, nous nous intéressons à des modèles où, en plus de l'action des courbes, des termes pénalisant les configurations de congestion sont présents. Nous développons des techniques permettant d'extraire de la régularité à partir de l'interaction entre l'évolution optimale de la densité (minimisation de l'action) et la pénalisation de la congestion, et nous les appliquons à l'étude des jeux à champ moyen et de la formulation variationnelle des équations d'Euler.

Quand l'espace de départ n'est plus seulement un segment mais un domaine de l'espace euclidien, nous considérons seulement le problème de Dirichlet, c'est-à-dire la minimisation de l'action (qui peut être appelée l'énergie de Dirichlet) parmi toutes les applications dont les valeurs sur le bord du domaine de départ sont fixées. Les solutions sont appelées les applications harmoniques à valeurs dans l'espace de Wasserstein. Nous montrons que les différentes définitions de l'énergie de Dirichlet présentes dans la littérature sont en fait équivalentes ; que le problème de Dirichlet est bien posé sous des hypothèses assez faibles ; que le principe de superposition est mis en échec lorsque l'espace de départ n'est pas un segment ; que l'on peut formuler une sorte de principe du maximum ; et nous proposons une méthode numérique pour calculer ces applications harmoniques.

Title: Optimal curves and mappings valued in the Wasserstein space

Keys words: Calculus of variations, Optimal Transport, Elliptic regularity, Analysis in metric spaces.

Abstract: The Wasserstein space is the space of probability measures over a given domain endowed with the quadratic Wasserstein distance. In this work, we study variational problems where the unknowns are mappings valued in the Wasserstein space.

When the source space is a segment, i.e. when the unknowns are curves valued in the Wasserstein space, we are interested in models where, in addition to the action of the curves, there are some terms which penalize congested configurations. We develop techniques to extract regularity from the minimizers thanks to the interplay between optimal density evolution (minimization of the action) and penalization of congestion, and we apply them to the study of Mean Field Games and the variational formulation of the Euler equations.

When the source space is no longer a segment but a domain of a Euclidean space, we consider only the Dirichlet problem, i.e. the minimization of the action (which can be called the Dirichlet energy) among mappings sharing a fixed value on the boundary of the source space. The solutions are called harmonic mappings valued in the Wasserstein space. We prove that the different definitions of the Dirichlet energy in the literature turn out to be equivalent; that the Dirichlet problem is well-posed under mild assumptions; that the superposition principle fails if the source space is no longer a segment; that a sort of maximum principle holds; and we provide a numerical method to compute these harmonic mappings.