SCUOLA
NORMALE
SUPERIORE

Corso di Perfezionamento in Matematica
Triennio 2004–2006

Variational problems in transport theory with mass concentration

Candidato
Filippo Santambrogio

Relatore
Prof. Giuseppe Buttazzo

Acknowledgments

As this thesis includes a large part of the mathematical work I carried out in the last years, I would like here to express my thanks to all the people who accompanied me throughout my mathematical education and researches. I'll stick to professional acknowledgment because for personal ones there's no point in writing them here: those who have been close to me are aware (I hope) of my gratefulness. Besides thanking Scuola Normale Superiore in Pisa, where I have been studying for the last seven years (and, for the first four years, living as well), and those institutions who hosted me for shorter periods, which I'll mention below, I would like to thank real persons.

My fist thank-you goes with no doubt to my advisor, Giuseppe Buttazzo, who has been the main guide for my studies even if I had the chance to work with several other persons. I think the relationship that we built during these years is a very positive one and I'm very glad of the choice I performed from the very beginning of my research career, when I was looking for a Laurea advisor. Prof. Buttazzo has been a huge source of problems (I mean, mathematical problems. . .), and not all of them are included in this thesis: I am sure that those which, up to now, are only at an investigatory stage will give the opportunity of future and fruitful collaborations.

Here in Pisa another person who deserves my sincere thanks is Luigi Ambrosio: recently the collaborations with him have given birth to a short joint article which is part of this thesis but, before that, I had the opportunity, as a student at SNS in Pisa, to learn a lot from his lectures. In particular, all the optimal transport theory I know comes from his 2001/02 class, and I often get wrong when I think that everything we learned in such a course is commonly known by everybody in the field.

Another very important presence here at SNS I learned a lot from is Paolo Tilli: it's a pleasure for me thanking him for his role as a teacher (during my first years here), as an advisor (for my 3rd year colloquio, where I first got introduced to Calculus of Variations and PDEs directly by means of the work of De Giorgi), as a spiritual guide (not only to me, but to several students at SNS), as a co-worker and as a friend. The period of our maximal collaboration, which allowed us to produce a joint work I am very satisfied of, has been for me very pleasant and it's a pity it stopped due to his moving to Torino.

During my PhD studies I had the opportunity to spend a non-negligible amount of time abroad, in particular during two research visits in France, a country that offered me a lot. I stayed three months at Université Montesquieu - Bordeaux IV at the very beginning of my doctorate in order to work together with Guillaume Carlier and I would like to thank him very much for the collaborations we had (in such an occasion and later), for those we are planning and for the friendship we developed.

My second long-term visit to France, at École Normale Supérieure de Cachan, even if quite different from the previous one, has been another very interesting experience. A sufficient reason for it would have been the huge amount of opportunities Paris offers to mathematicians, but in my case I have to add also the role played by Jean-Michel Morel, who supervised my work at ENS Cachan. I'd like to thank him sincerley for the mathematical work we carried out together, for the contribution he gave to several subjects which are part of this thesis, and for his personal attention to me and my future opportunities.

The people I have thanked so far are those who have been somehow, during these years, charged of the direction of my studies; yet my work has actually relied a lot on the interactions with several other mathematicians, co-authors, colleagues, friends… I would like to stress the importance of all the deeply informal discussions I had with them, when mathematics were often mixed with other topics, maybe in front of a tea-cup or in similar occasions (in the SNS corridor, at the CMLA kitchen, at home…).

Contents

Introduction

In this thesis we analyze several problems from the calculus of variations in the framework of optimal transport theory. We are mainly concerned with optimization problems from this theory and with other transportation problem which are alternative to the classical Monge-Kantorovich formalization. Most of the models we present come from application purposes and the set of possible applications includes urban economics, biology, fluid mechanics and geophysics.

The main topic of the thesis consists of the study optimal transport problem where some kind of concentration phenomena occur: on the one hand concentration of the marginal measures and on the other concentration along the transport itself.

The first subject, concentration of marginal measures, has been developed in the form of *Transport and Concentration Problems*, i.e. minimization problems where the unknowns are probability measures and the quantity to be minimized involves some functional encouraging or discouraging their concentration or dispersion and some transport costs as well. In particular, in these problems, transport appear only through their minimal value and we are in general not interested in finding the variables (transport plans, maps...) that actually realize the minimum, and we are only concerned with the properties of "optimal" marginal measures. These problems have some quite natural interpretation from a modelistic point of view. For instance, if we are concerned with the planning of a geographical region, we may be interested in finding a concentrated distribution of production centers together with a spread distribution of consumers, keeping anyway as small as possible the transportation costs for commuting or bringing the product to the consumers. Similar problems may also have an interpretation in the study of the shape of certain biological objects, such as leaves; these are in fact objects whose goal is to maximize their extension to take advantage of sunlight, but they receive their nutrient from a single concentrated source and the transport cost for this nutrient is to be taken into account.

The attention that we give to these problems dates back to the Laurea Thesis [66], prepared in 2003 under the direction of Prof. Buttazzo, and has evolved in the years, considering several models and discussing the corresponding results. A detailed report on these models is exactly the aim of the first part of this thesis. Our attention will be devoted to some modeling aspects and to the analysis of optimality conditions. In particular necessary optimality conditions, typically of the first order, are a key feature of this thesis. They are exploited as much as possible to obtain regularity, qualitative properties and, when possible, explicit expressions for the minimizers. In this case, as the unknowns are measures and hence belong to a nice vector space, differentiation is often feasible and this program gives unexpectedly strong results.

The second of the subjects we mentioned, concentration along transport, has a more classical structure: we are given the starting and arrival measures and we look for the optimal structure which transports the first onto the second. What is important is that in some applications we want to take into account how much the transportation appears to be concentrated: for instance if too many people pass through a same road in a city there could be a congestion effect; on the other hand if we need to create a road system, we would like to concentrate most of the path that different drivers follow on a same road, so that we only need to build few larger roads instead of several smaller ones. This suggests that there should be some quantities measuring how much the transportation is concentrated and, according to the different applications, we would like to consider minimization problem which encourage or discourage this concentration. It seems reasonable to consider Monge's problem as the concentration-neutral one and that we may create variants departing from this one.

In the thesis the main model discouraging concentration of transport, i.e. taking into account congestion effects, is hidden in Chapter 2, in the middle of Transport and Concentration Problems. In such a chapter we only briefly present it and we are much more interested in its minimal value, in the sense that we will use it as a functional of the marginal measures. A more refined and self-contained study of congestion models is in progress in [36] but it has not been included in this thesis due to its preliminary state.

On the other hand a lot of attention is given to transport problems encouraging joint transportation. There is a wide literature on them, as they are very natural in applications, and they give raise to some optimal one-dimensional branching structures. these structures are dealt with in the thesis in three chapters (6,7 and 9), together with a brief review of the already existing results.

Let us see now how the themes above have been developed in the chapters of the thesis. Each chapter corresponds roughly to a paper that has been prepared during this three-years doctoral period: some of these articles have been published and some are accepted. Only the last chapter contains some new computations not yet presented in preprint form.

Chapters 1, 2 and 3 are the first part of the thesis and present transport and concentration problems. In these chapters we are concerned with optimization problems of the following kind:

$$\min_{\mu,\nu\in\mathcal{P}(\Omega)} \mathfrak{F}(\mu,\nu) := T(\mu,\nu) + F(\mu) + G(\nu),$$

where the functional T represents transport costs between the two probability measures μ and ν and F and G are functionals over the space $\mathcal{P}(\Omega)$ of probability measures on Ω with opposite behavior: the first favors spread measures and penalizes concentration while the latter, on the other hand, favors concentrated measures. Chapter 1 is devoted to a problem fitting into this framework that has been proposed in the Laurea Thesis [66] and then in [28]. A particular choice of the functionals T, F and G is performed: we set

$$T(\mu,\nu) = W_p^p(\mu,\nu),$$
$$F(\mu) = \begin{cases} \int_\Omega f(u)\,d\mathcal{L}^d & \text{if } \mu = u\cdot\mathcal{L}^d \\ +\infty & \text{otherwise,} \end{cases}$$
$$G(\nu) = \begin{cases} \sum_{k\in\mathbb{N}} g(a_k) & \text{if} \nu = \sum_{k\in\mathbb{N}} a_k\delta_{x_k} \\ +\infty & \text{otherwise.} \end{cases}$$

The functions f and g must obviously satisfy some conditions, and in particular f must be convex and g subadditive. In this way T is the minimum of a Kantorovich optimal transport problem and F and G are local semicontinuous functionals over measures (see [16], [17] and [18]). A short introduction of this useful class of functionals over measures has been inserted into the chapter: in this way we can see that both concentration preferring functionals (as G is) and functionals favoring spread measures (as F does) fall into this class. Some emphasis is given throughout the Chapter to the interpretation of such a variational model in terms of urban planning: here the measure μ represents residents' distribution in the urban area Ω and ν stands for the distributions of services. The first measure has to be as spread as possible to maximize the average use of land of the citizens, while the second has to be concentrated in order to increase the efficiency f the production (i.e. we have positive externalities for nearby services). However, the commuting transportation cost for people needing to move from home to services must be considered as well. From the interpretation of the model the convexity and subadditivity assumptions on f and g, which are useful for technical reasons, turn out to be quite natural.

Besides the modelistic side, the chapter is devoted to some mathematical aspects of the problem. In this minimization existence results are straightforward, at least when Ω is compact. This is due to the semicontinuity of the functionals with respect to weak convergence of measures. Our attention is consequently mainly devoted to optimality conditions. Notice that any pair (μ, ν) giving finite value to $\mathfrak{F}$ must be necessarily composed by an absolutely continuous measure μ with density $u \in L^1(\Omega)$ and by a purely atomic measure ν. Hence it is interesting to deduce properties on the density u and on the location of the atoms of ν. The main results are obtained by perturbing an optimal μ into a new measure $\mu+\varepsilon(\mu_1-\mu)$ and keeping frozen ν. The duality formula in mass transportation plays a crucial role. The idea is very simple and the computations are simple as well, up to overcoming some technical difficulties about Kantorovich potentials. The result we get is the following: if u is the density of μ and ψ a suitably chosen Kantorovich potential between μ and ν we have

$$f'(u) + \psi = const \quad a.e.$$

For this result two proofs are provided. The second, mainly based on convex analysis, has been suggested by an anonymous referee while reviewing the article [28]. In the original paper, anyway, such a proof was only sketched and it actually requires some preliminary work before being performed. Then, after understanding optimality conditions for fixed ν the attention comes back to the whole problem when both μ and ν vary and the results are applied in order to characterize the global optima. The same results are also useful to gain some compactness when the problem is posed in an unbounded domain, such as $\Omega = \mathbb{R}^d$, where the existence is no longer trivial. The qualitative shape of the optimal configurations is in the end the following: ν is composed by finitely many atoms x_i and $\mu = u \cdot \mathcal{L}^d$ is concentrated on some balls B_i centered at these atoms, with radially decreasing densities given by an explicit formula:

$$u = (f')^{-1}(c_i - |x - x_i|^p) \quad \text{for } x \in B_i.$$

These balls may be interpreted as subcities (or cities if Ω is thought of as a larger region) and the atoms are their centers, where services are located.

In Chapter 2 we introduce in the subject of urban planning the concept of traffic congestion. The source of inspiration is a set of works by Beckmann ([10] and [11]) and the idea is the following: it is well known that the Monge-Kantorovich problem for a distance cost $|x-y|$ is equivalent to the minimal flow problem

$$\inf\left\{\int_\Omega |Y(x)|dx \ : \ \nabla\cdot Y=\mu-\nu \ \text{ in } \Omega, \quad Y\cdot n=0 \ \text{ on } \partial\Omega\right\};$$

if one instead looks at the problem

$$\inf\left\{\int_\Omega |Y(x)|^2dx \ : \ \nabla\cdot Y=\mu-\nu \ \text{ in } \Omega, \quad Y\cdot n=0 \ \text{ on } \partial\Omega\right\}$$

we are not only minimizing the total movement that is necessary to pass from μ to ν, but we are also penalizing an excessive concentration of this movement. At the beginning of the chapter this congestion model for optimal transportation is explained in more details. In minimizing the L^2 norm of the vector field Y under divergence constraints an elliptic equation with Neumann boundary conditions appears as an Euler equation. the optimal flow Y is in fact characterized by

$$Y=\nabla\phi; \qquad \begin{cases} -\Delta\phi=\mu-\nu & \text{in } \Omega, \\ \frac{\partial\phi}{\partial n}=0 & \text{on } \partial\Omega, \end{cases}$$

where the equation as to be taken in the weak sense.

Anyway, as we said above, in the chapter we are not much interested in the minimization problem itself and in understanding how the optimal flow looks like, but we are more interested in using the minimal L^2 norm of the flow as a quantity which represents the congestioned transport cost between the two measures. Then we insert this quantity in a Transport and Concentration Problem as in Chapter 1. It is quite easy to convince oneself, due to the link with elliptic theory, that the infimum of the L^2- flow is infinite if $\mu-\nu$ does not belong to a suitable functional space. This requires a sort of H^{-1} regularity and in particular prevents ν from having atoms. Hence, in the functional $\mathfrak{F}$ we will not only replace T by a congestion term, but G has to be replaced too, as the atomic choice of Chapter 1 is no longer possible. A very natural choice for the functional G is the following one, that we call *interaction energy*, and it is well known in the framework of optimal transport from the work of McCann ([58]):

$$G(\nu)=\int_{\Omega\times\Omega} V(x,y)(\nu\otimes\nu)(dx,dy),$$

where $V(x, y)$ is, for instance, an increasing function of the distance between x and y. The idea is to take the interaction cost for a service located at x and another service located at y and average it with respect to the distribution of services. For the functional F we keep the same choice as in Chapter 1 but we particularize it to the case $f(s) = s^2$, and in the end we get the following functional

$$\mathfrak{F}(\mu, \nu) = ||\mu - \nu||^2_{X'} + ||\mu||^2_{L^2} + G(\nu),$$

which is a quadratic functional. Here the space X' is the dual space of $X = \{\psi \in H^1(\Omega) : \int_\Omega \psi = 0\}$ with norm $||\psi||_X = ||\nabla\psi||_{L^2}$ (this representation of the congestion cost of transport comes from the representation formula for the optimal flow Y).

After setting the problem and getting some existence results, as in Chapter 1 we look for optimality conditions. Here too the first step is freezing ν and getting a convex quadratic problem in μ. Then the attention goes to the problem in ν only. This is more involved than what we had in Chapter 1 since now ν is no longer discrete and it can a priori be any probability measure since $G(\nu) < \infty$ for any $\nu \in \mathcal{P}(\Omega)$. We prove consequently a regularity result for ν which guarantees that ν is actually absolutely continuous with bounded density under certain assumptions (in particular, we need Ω to be convex: in the non-convex case ν may have a singular part concentrated on the non-convex part of $\partial\Omega$). The result is obtained by approximation and it is interesting to see that a very powerful regularization technique comes actually from the use of Monge-Kantorovich theory. In fact the original idea was to perturb the problem by adding a small term $\varepsilon||\nu||^2_{L^2}$ in order to force the optimal ν_ε to have a density and then to get uniform estimates on the L^∞ norm of ν_ε. Unfortunately, in this way we could retrieve at the limit some information only on a particular minimizer of $\mathfrak{F}$, i.e. the one which is approximated by minimizers of the perturbed problems. As we are not facing a convex problem we have no guarantee that there is a unique minimizer and we would like to have a regularity result which is valid for arbitrary minimizers. To do this we decide to add a small perturbation of the kind $\varepsilon W_2^2(\nu, \overline{\nu})$, where $\overline{\nu}$ is a minimizer that we can fix. In this way we are somehow forcing the minimizers ν_ε to converge to $\overline{\nu}$. What is interesting is that, in computing optimality conditions for the minimizers of the perturbed problems, the Kantorovich potential, induced by the presence of the Wasserstein distance W_2, appears. Then, well-known estimates on Kantorovich potentials help in getting uniform bounds on the densities of ν_ε. After a long part devoted to regularity the chapter contains some explicit examples. The one-dimensional case is treated in detail and in this case it turns out that the functionals involved have some displacement convexity properties (and the proof of this fact is interesting in itself). Then, the two-dimensional radial case is treated under some assumptions that ensure the uniqueness (and hence the radiality) of the solution. The case of a ball, of the whole space and of a crown is treated with explicit solutions. The whole work in Chapter 2 comes from a joint paper with Guillaume Carlier ([37]) that has been developed during a three months visit at the University of Bordeaux IV in 2004.

Chapter 3 contains a subsequent short work ([67]) that has been written to complete the subject of Transport and Concentration Problems. The goal was twofold: first, presenting this class of problems as a whole subject, with possible applications in urban planning (but not only); then, completing the framework of the problems we studied. In fact with G. Buttazzo we studied the problem where the transport cost was given by a Wasserstein distance and the concentration one by a local functional on atomic measures and with G. Carlier the case of a congestion transport cost and an interaction concentration cost. We already noticed that the case of congestion + local atomic functionals is not meaningful as it would have lead to a constantly infinite functional. Consequently, to complete the framework it is interesting to consider the Wasserstein + interaction case. The problem is hence the minimization of the functional

$$\mathfrak{F}(\mu,\nu) = W_2^2(\mu,\nu) + F(\mu) + \int_{\Omega\times\Omega} V(x,y)(\nu\otimes\nu)(dx,dy),$$

where F is the same as in Chapter 1. In Chapter 3 the goals are: determining some optimality conditions (mainly on ν for fixed μ, because those on μ for fixed ν come directly from what we saw in Chapter 1), using them to get L^∞ regularity, analyzing an explicit, quadratic, example. The regularity is obtained by the same scheme as in Chapter 2: almost the same kind of perturbations are performed, but the results on elliptic PDEs are here replaced by some results on Monge-Ampère equation. This comes from the fact that the Kantorovich potential plays here the role that in the previous case was played by the solution of the elliptic PDE. A more refined analysis of the behavior of the Kantorovich potential on the boundary of Ω allows to give a result under milder assumptions than what we did in Chapter 2. Anyway, at least a third of the Chapter is devoted to the general topic of transport and concentration problems, and a general definition of *concentration preferring* functional is given: a functional $G : \mathcal{P}(\Omega) \to \mathbb{R}\cup\{+\infty\}$ is said to be concentration preferring if we have $G(t_\sharp\nu) \leq G(\nu)$ for any $\nu \in \mathcal{P}(\Omega)$ and any $t : \Omega \to \Omega$ which is $1-$Lipschitz. Moreover it is shown that several other well-known variational problems fall into this class of problems. This is the case for instance of optimal location problems, which are quite studied in urban economics, but also of other average distance problems that will be discussed in Chapter 8 as well.

From Chapter 4 on our attention begins to be more directed towards branched transport problems. These problems are quite natural when we look at situations where we want to encourage joint transportation, for instance because building a network system with few large roads is cheaper than building several small roads. The structures that arise are all characterized by a first gathering of the masses, followed by joint transportation paths and finally a branching distribution towards the individual destinations. These structures and these problems are likely to appear also in natural phenomena, for instance in leaves, trees, river basins and blood vessels, and not only in human-built systems. The first mathematical precise formulation of the problem is due to Gilbert, who looked at it, in its discrete version, from the point of view of the applications in communication networks (see [48]). Once given some sources x_i with masses a_i and some sinks y_j wth masses b_j, his model consists in solving the following minimization problem

$$\min \; E(G) := \sum_h w_h^\alpha \mathcal{H}^1(e_h),$$

where the infimum is among all weighted oriented graphs $G = (e_h, \hat{e}_h, w_h)_h$ which fulfill the Kirchoff law at any vertex (at any x_i we have a_i+ incoming mass = outcoming mass, at any y_j we have incoming mass = outcoming mass $+b_j$ and at all the other vertices incoming and outcoming mass are equal). The exponent α is a fixed parameter $0 < \alpha < 1$ so that the function $t \mapsto t^\alpha$ is concave and subadditive.

The main recent mathematical interest on this subject has been trying to generalize this problem to the case of non-discrete measures μ and ν. There are very interesting models by Qinglan Xia ([72]), Maddalena-Solimini-Morel ([57]) and Bernot-Caselles-Morel ([13]). Chapter 4 contains a different approach that we tried to give to the problem in a joint work with Alessio Brancolini and Giuseppe Buttazzo. The main idea is looking at an interpolation between μ and ν, i.e a curve γ in the space of probability measures, which minimizes a certain length functional

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|(t)\; dt,$$

where $J : \mathcal{P}(\Omega) \to [0, +\infty]$ is a functional encouraging the curve to pass through concentrated measures and $|\gamma'|$ is the metric derivative according to a suitable distance in $\mathcal{P}(\Omega)$ (for instance the Wasserstein one W_p). In particular, the interest is choosing $J = G_\alpha$ where we have

$$G_\alpha(\nu) = \begin{cases} \sum_{k\in\mathbb{N}} a_k^\alpha & \text{if } \nu = \sum_{k\in\mathbb{N}} a_k \delta_{x_k} \\ +\infty & \text{otherwise,} \end{cases}$$

that is a particular case of the functional G used in Chapter 1. It turns out that this model is not equivalent to those by Gilbert, Xia et al. In fact, as it considers all the atoms of the measure $\gamma(t)$ at time t and it computes its speed in a global way and not for each atom separately, it follows that the functional takes into account also the mass of those atoms that have already reached their destination and stay eventually still. Anyway the model has a certain mathematical simplicity, due to the fact that it is in fact a geodesic problem in the space of probability measures endowed with a conformal perturbation of the Wasserstein distance. Moreover, the same model may work under minor changes to obtain very different functionals and optimal curves. For instance one can replace the functional G_α by F_q, given by

$$F_q(\mu) = \begin{cases} \int_\Omega |u|^q \, d\mathcal{L} & \text{if } \mu = u \cdot \mathcal{L} \\ +\infty & \text{otherwise.} \end{cases}$$

In this case too we have a particular choice of a local semicontinuous functional from those that we used in Chapter 1, but here we are discouraging concentration, favoring on the contrary spread measures.

Throughout the chapter we give some general theoretical existence results for the minimization of this length energies in the framework of metric spaces. Then we analyze separately the two cases of G_α and F_q. In order to have a well-posed problem it is also necessary to answer the question whether the minimal value is finite or not. It is in fact not obvious that a diffuse measure may be reached by a curve of atomic measures keeping the energy finite, as well as reaching an atomic measure with L^q densities could be sometimes impossible. What we get is that it may depend on the exponents α and q: for the case of the functional G_α we have finite energy for any pairs of measures (μ, ν) if $\alpha > 1 - 1/d$ (d being the dimension of the ambient space) and for F_q the same is true if $q < 1 + 1/d$. It is interesting to see how the two cases are somehow specular. At the end of the chapter we also partly approach the case of $\Omega = \mathbb{R}^d$, where existence is less trivial due to a certain lack of compactness. This is anyway solved by a more general theoretical result for metric spaces which are not locally compact. These two geodesic problems (the one with G_α and the one with F_q) could be somehow considered as transport problem where we look at concentration criteria along the transport. We are in fact applying to the interpolating measures the same concentration functional that we used in Chapter 1. Anyway, it looks rather different from what we did to define congestion in Chapter 2 and to what we will see later on for branching problems. Here the approach is less Eulerian and more time-dependent.

In Chapter 5 we develop a little more the case of the geodesic functional based on F_q. In fact this had been presented in Chapter 4 only as a natural counterpart of the concentration case, which was the main object from a branching point of view, but has some interesting feature in itself. First it may model the expansion of a gas whose initial and final configuration are known and which is subject to a negative pressure which leads it to diffuse as much as possible. Second, as we are facing L^q measures, we have in fact densities, i.e. functions of time and space, and we can write optimality conditions on them. The interest is towards the fact that the optimality conditions for those densities are expressed in the form of a system of PDEs which are very similar to the Euler equation for compressible gases. The chapter follows a joint work with Luigi Ambrosio where we rigorously derive this system of PDEs by means of perturbations of the measures through a transport-like variation (i.e. we replace μ_t by $(id + \varepsilon T)_\sharp \mu_t$). The system involves the densities and the velocity fields of the particles composing the densities, for a total of $d+1$ equations. It consists of d equations of kinetic type and the $d+1-$th equation is the continuity equation of conservation of the mass: if we denote by u the density and by v the velocity field we have

$$\begin{cases} H(t)\nabla u^q + K(t)\nabla \cdot \left(u|v|^{p-2}v \otimes v\right) + \frac{d}{dt}\left(K(t)u|v|^{p-2}v\right) = 0 & \text{in } \Omega, \\ \frac{d}{dt}u + \nabla \cdot (vu) = 0 & \text{in } \Omega \\ uv \cdot n = 0 & \text{on } \partial\Omega \\ \lim_{t\downarrow 0} u(t,\cdot)\mathcal{L}^d = \mu_0; \quad \lim_{t\uparrow 1} u(t,\cdot)\mathcal{L}^d = \mu_1, & \end{cases}$$

for suitable time-depending functions K and H.

At the end of the chapter we look for some particular solutions of the system, i.e. self-similar solutions. These are densities which have a certain shape which remains the same during time, up to scaling and translations. It turns out that there exist solutions of this kind (which, obviously, may only link self-similar boundary data μ_0 and μ_1 or at the limit Dirac masses), but they are characterized by a certain special shape: the allowed densities are in fact of the same form of the reversed parabolas that we found in Chapter 1. In the easiest case, i.e when the exponents p and q (for the Wasserstein space W_p and the diffusion functional F_q, respectively) are equal to 2, they have the form

$$u(t,x) = \left(A_t - B_t|x - x_t|^2\right)^+ .$$

The link with the optimal densities of Chapter 1 is evident, as we are minimizing a certain combination of Wasserstein distances and F_q functionals. Moreover, the reference measure is in both cases a Dirac mass (in the first problem ν is a finite sum of Dirac masses, and hence the situation is locally as if it were composed by a single atom; in this case, since if we have self-similar densities, at the limit we also have a single Dirac mass). It is however very interesting to see how this kind of densities appears in several problems involving mass transport.

After presenting the alternative (but different) models viewing branched transport structures as arising from geodesic problems in Wasserstein spaces, we come back in Chapter 6 to the formulations that have been equivalently given by Xia and Maddalena et al. as a generalization of Gilbert's problem. We first present Xia's relaxed problem: the Kirchoff constraint in Gilbert's problem is expressed in [72] as a divergence constraint

$$\nabla \cdot \lambda_G = \mu - \nu, \qquad \text{where } \lambda_G = \sum_h w_h[[e_h]].$$

Here $[[e]]$ is the integration measure measure on the segment e, given by $[[e]] = \hat{e} I_e \cdot \mathcal{H}^1$, and hence λ_G is a vector measure representing the flow which goes from μ to ν through the graph G. After this consideration Xia extended Gilbert's problem by relaxation to generic probabilities μ and ν. The problems becomes

$$\min \;\; \overline{E}(\lambda) : \;\; \nabla \cdot \lambda = \mu - \nu$$

where $\overline{E}(\lambda) := \inf \liminf_n E(\lambda_{G_n})$ and the infimum is over all possible sequences of finite graphs $(G_n)_n$ such that the corresponding vector measures λ_{G_n} converge to λ. It is also possible to prove a representation formula for the relaxed energy $\overline{E}(\lambda)$:

$$\overline{E}(\lambda) = \begin{cases} \int_M \theta^\alpha \, d\mathcal{H}^1, & \text{if } \lambda = (M, \theta, \xi), \\ +\infty & \text{otherwise.} \end{cases} \tag{0.0.1}$$

The equality $\lambda = (M, \theta, \xi)$ means that λ is a vector measure concentrated on the $1-$rectifiable set M and absolutely continuous w.r.t. $\mathcal{H}^1$ with density given by $\theta\xi$ (θ being a real multiplicity and ξ a measurable tangent unit vector field on M).

It is interesting to notice that this problem, as in the congestion problem of Chapter 2 and in the bidual version of Monge-Kantorovich, requires to minimize a quantity on λ under the constraint $\nabla \cdot \lambda = \mu - \nu$. In Monge's case this quantity is just the mass of λ, i.e. the L^1 norm when λ is absolutely continuous, in the congestion case it is the L^2 norm or more generally a convex superlinear functional, and here it is a concave functional also known as $M^\alpha-$mass. This means that, under the same constraints, not only we want to minimize the total movement quantity, but we may encourage or discourage this movement to be concentrated or dispersed. Here we want it to be concentrated (concentrated on one dimensional structures and with a subadditive cost which prefers few larger flows than several small ones), in Chapter 2 we want it to be as spread as possible. Monge's case is somehow in the middle, as a concentration-neutral case. This shows how we have a common Eulerian formulation of some different transport problems, with different features and applications, all starting from Monge (i.e. $c(x, y) = |x - y|$ and we cannot hope to retrieve them by means of other costs, such as $|x - y|^p$).

After presenting the Eulerian approach by Xia the same problem is presented under the Lagrangian approach of two works, the first one by Maddalena, Solimini and Morel, [57], where only the case of a single source (i.e. $\mu = \delta_0$) is dealt with, and the second one by Bernot, Caselles and Morel, [13], where the results are generalized to the case of arbitrary measures. The main idea is to look at measures η on the space Γ of $1-$Lipschitz paths which eventually stop (at a time denoted by $\sigma(\gamma)$) and to define the multiplicity that this system of paths has at a point x: we set $[x]_\eta = \eta(\{\gamma : x \in \gamma\})$. Then we define $Z_\eta(\gamma) = \int_0^{\sigma(\gamma)} [\gamma(t)]_\eta^{\alpha-1} dt$ and we minimize the functional

$$J(\eta) = \int_\Gamma Z_\eta(\gamma)\eta(d\gamma).$$

The constraint in this case is that the initial and terminal measures of η are μ and ν, respectively, i.e. $(\pi_0)_\sharp \eta = \mu$ and $(\pi_\infty)_\sharp \eta = \nu$, where $\pi_0(\gamma) = \gamma(0)$ and $\pi_\infty(\gamma) = \gamma(\sigma(\gamma))$. In a recent paper by the same authors, [14], the equivalences between all these model (i.e. the one by Xia, the one by Maddalena, Solimini and Morel and the one by Bernot, Caselles and Morel) are proven.

What we do in Chapter 6 is mainly looking at the infimum values of these problems and at their dependence on μ and ν. First we recall some finiteness result, and the main one is that the minimum is always finite for any pair of compactly supported measures μ and ν if $\alpha > 1 - 1/d$. It is well-known that this bound is sharp (see [43]): here we provide only a short proof of the fact that, if α is strictly below the threshold, then it is not possible to arrive at the rescaled Lebesgue measure on Ω with finite energy. Notice that this threshold exponent is the same that we had in Chapter 4. By the way, we also take advantage of the Lagrangian formulations that we present, where the time variable is present, and we develop a little more a comparison between the two models. From the formalism of Chapter 4, it turns out that the way to get a problem which is as similar as possible to these branched problems is to take $p = \infty$ in the choice of the distance W_p. After looking at the finiteness of the value, for $\alpha > 1 - 1/d$, we denote the minimum by $d_\alpha(\mu, \nu)$ as in [72]. This quantity turns out to be a distance over the space of probability measures and it was known from Xia that it metrized the weak convergence. In a joint work with Jean Michel Morel, which is the main original part of the chapter, we prove some sharp inequalities between these distances and the Wasserstein distance W_1. Namely, what we prove is

$$W_1(\mu, \nu) \le d_\alpha(\mu, \nu) \le cW_1(\mu, \nu)^{d(\alpha-(1-1/d))},$$

where c is a constant depending only on the dimension d and on the exponent α. We also prove that the exponents of W_1 in the above inequalities are sharp. This result gives an answer to a question posed by Cedric Villani about the comparison of standard Kantorovich transport and branched transport.

If Chapter 6 has also played the role of a general introduction to branched transport problems, in Chapter 7 we develop a very peculiar feature of them whose motivations lie, as far as interdisciplinary applications are concerned, mainly in geophysics. In fact, geophysicists, while studying the shape and evolution of river basins, have two main objects to deal with: the structure of the river network and the elevation of the landscape in the region. In many physical models the landscape elevation is obtained at a point x by integrating along the only stream arriving at x from the outlet of the whole basin the quantity $\theta^{\alpha-1}$, where at any point of the river network θ stands for the multiplicity of the network itself. This topic has been considered in a series of paper (see for instance [9] or the book [64]) mainly under some strong discretization. Anyway, the formula we gave for Z_η and its use in the definition of the functional J suggest that it should be possible to define a similar landscape function also in the continuous case. Roughly speaking, the idea is to take an optimal measure $\eta \in \mathcal{P}(\Gamma)$ minimizing J with initial measure δ_0 and terminal measure $\mu \in \mathcal{P}(\Omega)$ and defining the landscape function z by

$$z(x) = Z_\eta(\gamma) \quad \text{for } \gamma \text{ such that } \pi_\infty(\gamma) = x$$

(this obviously requires to check that it is well-defined, i.e. that different curves give the same result). The chapter follows a recent paper (see [68]) that has been widely discussed with Jean-Michel Morel during the same six-months visit to Cachan in which the results of [60] and of Chapter 9 have been established.. As a first thing we argue in a detailed way the interest of defining a landscape function z associated to branched transport problems and we point out the features it should have: the main one is a certain link with the geometry of the network and in particular we want that at every point of the network the maximal slope direction of z must agree with the direction of the network itself at x. Then it should be interesting to have some regularity property of z, even if one cannot expect it to be Lipschitz continuous, since it must have arbitrarily large derivatives $\theta^{\alpha-1}$ in the direction of the network. This in particular forces us to give a weak concept of maximal slope direction in the above requirement. Anyway, at the end of the chapter, the function z is proven to be Hölder continuous under some extra assumptions (through an interesting use of Campanato spaces), and in general lower semicontinuous.

Another very interesting feature of this study of the landscape function, which is developed in Chapter 7, is the fact that z also acts as a derivative of the functional $\mu \mapsto X_\alpha(\mu) := d_\alpha(\delta_0, \mu)$. In fact we can prove that, if we set $\mu_\varepsilon = \mu + \varepsilon(\mu_1 - \mu)$, it holds

$$\limsup_{\varepsilon \to 0^+} \frac{X_\alpha(\mu_\varepsilon) - X_\alpha(\mu)}{\varepsilon} \leq \alpha \int_\Omega z \, d(\mu_1 - \mu),$$

where z is the landscape function with respect to the fixed measure μ. This is pointed out in the discrete case and then generalized to arbitrary measures μ. This formula may be useful while studying minimization problems for functionals like $F(\mu) + X_\alpha(\mu)$, which was in fact proposed in [57]. Recently similar problems, where F is a functional which encourages the dispersion of μ, have been proposed to model the shape of leaves or flowers. The interpretation comes from the fact that we let μ stand for such a shape and δ_0 represent the source of nutrient for the leave which arrives at a single point. Then, the shape tries to optimize the cost for being irrigated starting from such a single point and the positive effect of being as widespread as possible to take advantage of sunlight. This problem falls easily in the wide framework of Transport and Concentration Problems proposed in Chapter 3 and in the chapter an example of this kind is developed to show how this derivative formula could be useful in getting necessary optimality conditions. This derivative result involving the landscape function may be compared to what happens in the case of usual optimal transportation, where we have

$$\limsup_{\varepsilon\to 0^+} \frac{W_p^p(\mu_\varepsilon,\nu)-W_p^p(\mu,\nu)}{\varepsilon} \le \int_\Omega \psi\, d(\mu_1-\mu),$$

ψ being the Kantorovich potential in the transportation from μ to ν with cost $c(x,y) = |x-y|^p$ (provided it is unique up to constants, otherwise the situation is a little more tricky). This derivative result on Wasserstein distances was in fact the starting point for the results in Chapters 1 and 3. In fact we may realize that the landscape function plays somehow the role of Kantorovich potential in branched transportation and this comes not only from this derivative result, but also from the representation formula

$$X_\alpha(\mu) = \int_\Omega z\, d\mu = \int_\Omega z\, d(\mu-\delta_0),$$

which is proven in the chapter, and from the Hölder continuity result. The Landscape function is proven in fact to be $d(\alpha-(1-1/d))$−Hölder continuous under some conditions on μ, and this result, as the Hölder exponent varies from 0 to 1 as α goes from $1-1/d$ to 1, perfectly fits with the fact that Kantorovich potentials are Lipschitz continuous. Unfortunately, due to the lack of convexity in the minimization problem for branched transport, it seems that there is no interpretation of z as the optimum of a dual problem.

In Chapter 8, we leave the framework of branched transport and we present another optimization problem on one-dimensional structures. This problem, introduced in [27], consists in finding a subset $\Sigma \subset \Omega$ which minimizes the cost function

$$D(\Sigma) = \int_\Omega d(x,\Sigma)\, \mu(dx)$$

among all compact connected sets whose length does not exceed a given value l, i.e. under the constraint $\mathcal{H}^1(\Sigma) \le l$. This means looking for a set which must be as spread as possible (so that the values $d(x,\Sigma)$ are as small as possible), without breaking the connectedness and length constraints. This problem has some interesting interpretations both in terms of applications (in image reconstruction it corresponds to recovering a line Σ from a pixel cloud μ in a picture, recalling the well-known concept of *skeleton* of the image; in urban planning Σ may be interpreted as a subway network in a city Ω with population density μ) and in optimal transport theory. The link with optimal transport theory comes from the equality

$$D(\Sigma) = \inf\left\{W_1(\mu,\nu) \mid \operatorname{spt}(\nu)\subset\Sigma\right\}.$$

In this way we can also see that this problem too falls into the framework of the Transport and Concentration Problems introduced in their generality in Chapter 3: we are just minimizing $\nu \mapsto W_1(\mu, \nu)$ under a constraint $G(\nu) \leq l$, the functional G standing for the minimal length of a compact connected subset containing the support (this functional is explicitly listed in Chapter 3 among those who satisfy the definition on being *concentration preferring*).

This average distance problem with length constraints has been studied as far as existence and qualitative properties of the minimizers are concerned in [27] and the main tool for the existence is Golab's theorem (which justifies the connectedness assumption, which makes anyway sense for several applications). In the joint work ([69]) with Paolo Tilli on which the chapter is based we look at some regularity properties of the minimizers. The main question is the existence of blow-up limits of an optimal Σ around its points. Precisely, we say that Σ has a blow-up limit K at $x_0 \in \Sigma$ if the localized and rescaled sets $(\Sigma \cap \overline{B(x_0, r)} - x_0)/r$ converge, in the Hausdorff distance as $r \to 0$, to some set $K \subset \overline{B(0,1)}$. Due to compactness results on the Hausdorff convergence it is not difficult to have the existence of these limits up to subsequences. It is not even difficult to characterize their shapes: they can be only composed by up to three unit rays, which may form a diameter, a corner or triple 120° configuration when they are not a single ray (this up-to-subsequence result is proven in the chapter). What is not trivial at all is that these limits do not change according to the subsequence and this is proven with different techniques in different cases (endpoints, triple junctions...): these techniques involve stationarity, small perturbations and $\Gamma-$convergence as well. Under an L^∞ assumption on the measure μ it is proven that at any point $x_0 \in \Sigma$ the full limit of the blow-up procedure exists. Moreover, in some cases it is possible to estimate the rate of change of the direction of the rays which form this limit, thus getting a $C^{1,1}$ regularity result. This is proven in the last section of the chapter in a neighborhood of any point x_0 such that the diameter of the set

$$\mathcal{T}(x_0) = \{x \in \Omega \,:\, d(x, \Sigma) = |x - x_0|\}$$

is sufficiently small. In particular this happens if x_0 is a triple point, since in this case we are able to prove that this set reduces to x_0 only. In this way we have a satisfactory description of the behavior of Σ near its triple junctions: it is composed by three $C^{1,1}$ curves whose tangent vectors at x_0 from three angles of 120°. This gives a complete answer to one of the main questions posed in [27] about this problem, the other ones being about regularity (partially answered by this blow-up result), asymptotics as $l \to \infty$ or $l \to 0$, boundary behavior and no-loop properties.

Some of the techniques introduced in Chapter 8 are then used in Chapter 9 on a different problem. We come back to the branched transport framework and we want to study the blow-up. This has been first done by Xia in [74], and we know from it how the blow-up limits up to subsequences look like. In a work in progress with Jean-Michel Morel (see [61]) we try to use a curvature approach to deduce the existence of the limits: we fix a curve in the optimal network, we perturb it and we get optimality conditions. These conditions ensure that the derivative of the curve is a BV function on the interval of parametrization and allow to say that the curve has a side tangent vector at any point. This result requires some strong assumptions on the marginal measures, and in particular a lower bound on the densities. Here in this chapter we propose an alternative approach, which works under different conditions, which are less restrictive on the measures. We suppose that μ belongs to $L^p(\Omega)$ for a certain $p > 1$ and that the couple (μ, ν) satisfies the *regularity assumption*, i.e. either ν is atomic or $\mathrm{spt}(\mu) \cap \mathrm{spt}(\nu) = \emptyset$. On the other hand, the result is only valid in two spatial dimensions and if the point x_0 where to center the blow-up is a branching point (which is anyway the most interesting case, since then we could apply some angle conditions). Under these conditions we are able to perform a procedure exactly as the one used in Chapter 8 for triple points. We prove that the oscillation of the angle $\theta(r)$ which represents the intersection direction of a branch of the optimal network $N = \{x \in \Omega : [x]_\eta > 0\}$ with $\partial B(x_0, r)$ may be estimated by a quantity linked to the mass which is transported onto $N \cap B(x_0, r)$. Then it is sufficient to estimate this mass to get a convergence result and it is what we do, via some geometric and asymptotic estimates.

At the very beginning of the thesis there is a preliminary chapter on optimal transportation where all the results which will be useful later are introduced. There is no proof but only some bibliographical reference to the books by Villani, Ambrosio-Gigli-Savaré and the lecture notes by Ambrosio ([71], [4] and [3], respectively). We deal with the primal and dual Kantorovich problems, with the existence of optimal maps, i.e. solutions to Monge's problem, with the regularity of transports and potentials in the quadratic case by Monge-Ampère equation and with Wasserstein distances, curves in Wasserstein spaces, geodesics and geodesically convex functionals.

As a whole, this thesis presents, in a quite unified setting of transport problems involving concentration criteria, almost all the researches that we carried out during these PhD studies. Only some subjects, mainly related to shape optimization, where the transport component was completely absent has been neglected. Probably the most interesting feature of the thesis are the techniques to get necessary optimality conditions in the set of problems that have been approached. Most of them are not new; we simply use them in a particular way, obtaining sometimes unexpectedly strong results. This is the case of the derivation of some functionals on $\mathcal{P}(\Omega)$ with respect of perturbations such as $\mu + \varepsilon(\mu_1 - \mu)$ or $(id + \varepsilon\xi)_\sharp\mu$. On the other hand some regularization issues such as the L^∞ one in Chapter 2 or the blow-up one in Chapter 8 have required some more technical tools which seem to be quite original. Moreover, also some very classical results, for instance from linear or nonlinear elliptic PDEs, from the theory of Campanato Spaces or from convex analysis, are used throughout the thesis and this completes the picture of the different techniques to get necessary conditions or regularity. Anyway, the thesis does not develop only this aspect of the variational problems that approaches, but devotes also a certain space to the interpretation of the models (as in Chapters 1 and 2 for urban planning and in Chapter 7 for river basins) and to existence results (mainly in Chapter 4).

Notations

We summarize here most of the peculiar notations and expressions which are used throughout the thesis and not always explicitely recalled.

First, let us precise that we will call, for simplicity, *domains* those sets which are the closure of a non-empty connected and bounded open subset of $\mathbb{R}^d$ with negligible boundary. These domains will be often denoted by Ω, so that the reader must not be astonished if Ω denotes a compact set instead of an open one. Moreover, we will silently confuse a domain Ω and its interior $\overset{\circ}{\Omega}$ when some functional spaces involving higher regularity are concerned: for instance when we write $H^1(\Omega)$ we usually mean $H^1(\overset{\circ}{\Omega})$. This is performed in order not to avoid heavy notations, should we distinguish between the closed and the open set.

Given a set C endowed with a natural topology (usually a domain) we will denote by $\mathcal{P}(C)$ the set of all Borel probability measures on C. The set of finite vector on C measures valued in $\mathbb{R}^k$ will be denoted by $\mathcal{M}^k(C)$.

The $d-$dimensional Lebesgue measure will be denoted by $\mathcal{L}^d$, but sometimes we will write $|\Omega|$ for $\mathcal{L}^d(\Omega)$. When we say "the Lebesgue measure on Ω"and we are speaking of a probability measure, we actually mean the rescaled measure $1/|\Omega| \cdot \mathcal{L}^d$ restricted to Ω. The symbol $\mathcal{H}^1$ will denote instead the $1-$dimensional Hausdorff measure.

For a sequence of probability or vector measures on Ω we will use the term *weak convergence* to mean the convergence in the duality with the space $C_b^0(\Omega)$ of bounded continuous functions on Ω. This convergence will be denoted by the symbol $\rightharpoonup$ (with no stars), so that $\mu_n \rightharpoonup \mu$ means $\int_\Omega \phi \, d\mu_n \to \int_\Omega \phi \, d\mu$ for any $\phi \in C_b^0(\Omega)$.

Crucial will also be the concept of image measure: given a measure μ on Ω_1 and a measurable map $T : \Omega_1 \to \Omega_2$ we denote by $T_\sharp\mu$ the image of μ through T, which is a measure on Ω_2 defined by $T_\sharp\mu(A) = \mu(T^{-1}(A))$ for any measurable subset $A \subset \Omega_2$. If $T_\sharp\mu = \nu$ we will also say that T transports (or pushes) μ onto ν.

As far as more transport-related concepts are concerned, we denote by $\Pi(\mu,\nu)$ the set of transport plans with marginal measures μ, $\nu \in \mathcal{P}(\Omega)$ (see Section 0.1) and by $TP(\mu,\nu)$ the set of traffic plans from μ to ν. This latter concept is typical of the theory of branched transport: it consists of the set of all probability measures on the space of Lipschitz curves on $[0,+\infty[$ which eventually stop such that the images under the evaluation at the starting time and at the stopping time are μ and ν, respectively. These two evaluations are denoted by π_0 and π_∞, respectively, as well as the evaluation at a generic time t which is denoted by π_t. See Section 6.2 for details.

The symbol id denotes the identity mapping from a set to itsels $id(x) = x$. The identity matrix is denoted by the symobl $\mathbb{I}$. The symbol I the indicator function: if we write I_A we mean the function whose value is 1 on A and 0 outside. We may also write $I_{condition}$, which means a function of possibly several variables whose value is 1 if the condition is verified and 0 otherwise. For instance, writing $I_{x\in\gamma}$ is a function of two variables (x and γ) which has the same values as $I_\gamma(x)$. When a measure μ (usually the Lebesgue or the Haudorff measures $^\lceil$ or $\mathcal{H}^1$) and a set A are given, we will write $I_A \cdot \mu$ or $\mu \llcorner A$ with the same meaning.

The indicator function in the sense of convex analysis is on the contrary denoted by a δ symbol: $\delta(\cdot|A)$ is the function whose value is 0 on A and $+\infty$ elsewhere. Its Legendre-Flenchel transform is the support function of A and it is given by $\delta^*(y|A) = \sup_{x\in A} y \cdot x$.

Preliminaries on Optimal Transportation

This chapter does not want to be an exhaustive presentation of the topic, but only a short list of useful results with no proofs that will be used later on in thesis. The main reference is [71]. Anyway, the approach is the same used in the lectures given by Prof. L. Ambrosio at SNS Pisa in 2001/02 and hence another possible reference is [3].

The motivation for the whole subject is the following problem proposed by Monge in 1781 ([59]): given two densities of mass f, $g \geq 0$ on $\mathbb{R}^d$, with $\int f = \int g = 1$, find a map $T : \mathbb{R}^d \to \mathbb{R}^d$ pushing the first one onto the other, i.e. such that

$$\int_A g(x)dx = \int_{T^{-1}(A)} f(y)dy \quad \text{for any Borel subset } A \subset \mathbb{R}^d$$

and minimizing the quantity

$$\int_{\mathbb{R}^d} |T(x) - x| f(x) dx$$

among all the maps satisfying this condition.

This problem has stayed with no solution (does a minimizer exist? how to characterize it?...) for centuries. Only with the work by Kantorovich it has been inserted into a suitable framework which gave the possibility to approach it and, later, to find that actually solutions exist and to study them. The problem has been widely generalized, with very general cost functions $c(x, y)$ instead of the euclidean distance $|x - y|$ and more general measures and spaces. For simplicity, here we will not try to present a very wide theory on generic metric spaces, but we will deal only with the euclidean case.

0.1 Primal and dual problems

In what follows we will suppose Ω to be a domain of $\mathbb{R}^d$ and the cost function $c : \Omega \times \Omega \to [0, +\infty[$ will be supposed continuous and symmetric (i.e. $c(x, y) = c(y, x)$).

The generalization that appears as natural from the work of Kantorovich ([53]) of the problem raised by Monge is the following:

Definition 0.1.1. Given two probability measures μ and ν on Ω and a cost function $c : \Omega \times \Omega \to [0, +\infty]$ we consider the problem

$$(K) \quad \min \left\{ \int_{\Omega\times\Omega} c\, d\pi \,|\, \pi \in \Pi(\mu, \nu) \right\}, \tag{0.1.1}$$

where $\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(\Omega \times \Omega) : (p^+)_\sharp \pi = \mu, (p^-)_\sharp \pi = \nu, \}$ and p^+ and p^- are the two projections of $\Omega \times \Omega$ onto Ω. The minimizers for this problem are called *optimal transport plans* between μ and ν. Should π be of the form $(id \times T)_\sharp \mu$ for a measurable map $t : \Omega \to \Omega$, the map T would be called *optimal transport map* from μ to ν.

Remark 0.1.2. It can be easily checked that if $(id \times T)_\sharp \mu$ belongs to $\Pi(\mu, \nu)$ then T pushes μ onto ν (i.e. $\nu(A) = \mu(T^{-1}(A))$ for any Borel set A) and the functional takes the form $\int c(x, T(x))\mu(dx)$, thus generalizing Monge's problem.

Remark 0.1.3. This generalized problem by Kantorovich is much easier to handle than the original one by Monge, for instance because in the Monge case we would need existence of at least a map T satisfying the constraints. This is not the case in the case $\mu = \delta_0$ if ν is not a single Dirac mass. On the contrary, there always exist transport plan in $\Pi(\mu, \nu)$ (for instance $\mu \otimes \nu \in \Pi(\mu, \nu)$). Moreover, one can state that (K) is the relaxation of the original problem by Monge: if one considers the problem in the same setting, where the competitors are transport plans, but sets the functional at $+\infty$ on all the plans that are not of the form $(id \times T)_\sharp \mu$, then one has a functional on $\Pi(\mu, \nu)$ whose relaxation is the functional in (K) (see [5]).

An important tool will be duality theory and to introduce it we need in particular the notion of $c-$transform (a kind of generalization of the well-known Legendre transform).

Definition 0.1.4. Given a function $\chi : \Omega \to \overline{\mathbb{R}}$ we define its $c-$*transform* (or $c-$conjugate function) by

$$\chi^c(y) = \inf_{x\in\Omega} c(x, y) - \chi(x).$$

Moreover, we say that a function ψ is $c-$*concave* if there exists χ such that $\psi = \chi^c$ and we denote by $\Psi_c(\Omega)$ the set of $c-$concave functions.

It is well-known a duality result stating the following equality (see Theorem 1 together with the following Remark on $c-$concavity in [71]):

Proposition 0.1.5. *We have*

$$\min(K) = \sup_{\psi \in \Psi_c(\Omega)} \int_\Omega \psi \, d\mu + \int_\Omega \psi^c \, d\nu. \tag{0.1.2}$$

In particular the minimum value of (K) *is a convex function of* (μ, ν) *as it is a supremum of linear functionals.*

Definition 0.1.6. The functions ψ realizing the maximum in (0.1.2) are called *Kantorovich potentials* for the transport from μ to ν. This is in fact a small abuse, because usually this term is used only in the case $c(x, y) = |x-y|$ only.

Notice that any $c-$concave function shares the same modulus of continuity of the cost c. In particular, in the case $c(x, y) = |x - y|^p$, if Ω is bounded with diameter D, any $\psi \in \Psi_c(\Omega)$ is $pD^{p-1}-$Lipschitz continuous. The case where c is a power of the distance is in fact of particular interest and two values of the exponent p are remarkable: the cases $p = 1$ and $p = 2$. In these two cases we provide characterizations for the set of $c-$concave functions when $\Omega = \mathbb{R}^d$. Let us denote by $\Psi_{(p)}(\Omega)$ the set of $c-$concave functions with respect the cost $c(x, y) = |x - y|^p/p$. It is not difficult to check that

$$\psi \in \Psi_{(1)}(\mathbb{R}^d) \Leftrightarrow \psi \text{ is a 1-Lipschitz function;}$$

$$\psi \in \Psi_{(2)}(\mathbb{R}^d) \Leftrightarrow \; x \mapsto \frac{x^2}{2} - \psi(x) \text{ is a convex function .}$$

The first characterization is true also when Ω does not coincide with the whole space, while the second in fact becomes just an implication (if $\psi \in \Psi_{(2)}$, then $\frac{x^2}{2} - \psi(x)$ is convex, but not any convex function comes from a $c-$concave function, due to the restriction on the Lipschitz constant).

The case $c(x, y) = |x - y|$ shows a lot of interesting features, even if from the point of the existence of an optimal map T it is one of the most difficult. A first interesting property is the following:

Proposition 0.1.7. *For any* $1-$*Lipschitz function* ψ *we have* $\psi^c = -\psi$. *In particular, Formula 0.1.2 may be re-written as*

$$\min(K) = \sup_{\psi \in \mathrm{Lip}_1} \int_\Omega \psi \, d(\mu - \nu).$$

Another peculiar feature of this case is the following:

Proposition 0.1.8. *Consider the problem*

$$(B) \quad \min\left\{M(\lambda)\ :\lambda\in\mathcal{M}^d(\Omega);\nabla\cdot\lambda=\mu-\nu\right\}, \tag{0.1.3}$$

where $M(\lambda)$ denotes the mass of the vector measure λ and the divergence condition is to be read in the weak sense, with Neumann boundary conditions, i.e. $-\int\nabla\phi\cdot d\lambda=\int\phi d(\mu-\nu)$ for any $\phi\in C^1(\Omega)$. If Ω is convex then it holds

$$\min(K)=\min(B).$$

This proposition links the Monge-Kantorovich problem to a minimal flow problem which has been first proposed by Beckmann in [10], under the name of *continuous transportation model*, without knowing this link, as Kantorovich's theory was being developed independently almost in the same years. In Section 2.1 we will see some details more on this model and on the possibility of generalizing it to the case of distances $c(x,y)$ coming from Riemannian metrics. In particular, in the case of a nonconvex Ω, (B) would be equivalent to a Monge-Kantorovich problem where c is the geodesic distance on Ω.

We now come back to the case of a generic cost $c(x,y)$. Another useful result about $c-$transform is the following:

Proposition 0.1.9. *For any cost c and any function $\psi:\Omega\to\overline{R}$ we have $\psi^{cc}\geq\psi$ and the equality holds if and only if ψ is $c-$concave.*

We summarize here some useful results for the case where the cost c is of the form $c(x,y)=h(x-y)$, for a strictly convex function h.

Theorem 0.1.10. *Given μ and ν probability measures on a domain $\Omega\subset\mathbb{R}^d$ there exists unique an optimal transport plan π. It is of the form $(id\times T)_\sharp\mu$, provided μ is absolutely continuous. Moreover there exists also at least a Kantorovich potential ψ, and the gradient $\nabla\psi$ is uniquely determined $\mu-$a.e. (in particular ψ is unique up to additive constants, provided the density of μ is positive a.e. on Ω). The optimal transport map T and the potential ψ are linked by $T(x)=x-(\partial h)^{-1}(\nabla\psi(x))$. Moreover it holds $\psi(x)+\psi^c(T(x))=c(x,T(x))$ for $\mu-$a.e. x. Conversely, every map T which is of the form $T(x)=x-(\partial h)^{-1}(\nabla\psi(x))$ for a function $\psi\in\Psi_c(\Omega)$ is an optimal transport plan from μ to $T_\sharp\mu$.*

Remark 0.1.11. Actually, the existence of an optimal transport map is true under weaker assumptions: we can replace the condition of being absolutely continuous by the condition $\mu(A) = 0$ for any $A \subset \mathbb{R}^d$ such that $\mathcal{H}^{d-1}(A) < +\infty$. Anyway, in this thesis only the absolutely continuous case will be used.

Remark 0.1.12. In Theorem 0.1.10 only the part concerning the optimal map t is not symmetric in μ and ν: hence the uniqueness of the Kantorovich potential is true even if it ν (and not μ) has positive density a.e.

Remark 0.1.13. Theorem 0.1.10 may be particularized to the quadratic case $c(x, y) = |x - y|^2/2$, thus getting the existence of an optimal transport map $t = \nabla\phi$ for a convex ϕ. By using the converse implication (sufficient optimality conditions), this also proves the existence and uniqueness of a gradient of a convex function transporting μ onto ν. This well known fact has been investigated first by Brenier in [21].

All the costs $c(x, y) = |x-y|^p$ with $p > 1$ fall under Theorem 0.1.10. For the case $c(x, y) = |x - y|$ the results are a bit weaker and are summarized below (this is the classical Monge case and we refer to [5] and [45]).

Theorem 0.1.14. *Given μ and ν probability measures on a domain $\Omega \subset \mathbb{R}^d$ there exists at least an optimal transport plan π. Moreover, one of such plans is of the form $(id \times T)_\sharp\mu$ provided μ is absolutely continuous. There exists also at least a Kantorovich potential ψ, and we have $\psi(x) - \psi(T(x)) = |x - T(x)|$ for $\mu-$a.e. x, for any choice of optimal T and ψ.*

Here the absolute continuity assumption is essential to have existence of an optimal transport map, in the sense that in general it cannot be replaced by weaker assumptions as in the strictly convex case. This can be seen from the following example.

Example 0.1.15. Set

$$\mu = \mathcal{H}^1 \llcorner A \quad \text{and} \quad \nu = \frac{\mathcal{H}^1 \llcorner B + \mathcal{H}^1 \llcorner C}{2}$$

where A, B and C are three vertical parallel segments in $\mathbb{R}^2$ whose vertexes lie on the two line $y = 0$ and $y = 1$ and the abscissas are 0, -1 and 1, respectively. In this case one can get a sequence of maps $T_n : A \to B \cup C$ by dividing A into $2n$ equal segments $(A_i)_{i=1,\dots,2n}$ and B and C into n segments each, $(B_i)_{i=1,\dots,n}$ and $(C_i)_{i=1,\dots,n}$ (all ordered upwards). Then define T_n as a piecewise affine map which sends A_{2i-1} onto B_i and A_{21} onto C_i. In this way the cost of the map T_n is less than $1/2 + 1/n$, but no map T may obtain a cost $1/2$, as this would imply that any point is sent horizontally and but this cannot respect the push-forward constraint. On the other hand, the transport plan associated to T_n weakly converge to the transport plan $1/2T^+_\sharp\mu + 1/2T^-_\sharp\mu$, where $T^\pm(x) = x \pm e$ and $e = (1, 0)$. This transport plan turns out to be the only optimal transport plan and has a Kantorovich cost of 1/2.

The same construction provides also an example of the relaxation procedure leading from Monge to Kantorovich.

0.2 Wasserstein distances and spaces

Starting from the values of the problem (K) in (0.1.1) we can define a set of distances over $\mathcal{P}(\Omega)$. For any $p \geq 1$ we can define

$$W_p(\mu, \nu) = \big(\min(K) \text{ with } c(x, y) = |x - y|^p\big)^{1/p}.$$

We recall that it holds, by Duality Formula,

$$\frac{1}{p} W_p^p(\mu, \nu) = \sup_{\psi \in \Psi_p(\Omega)} \int_\Omega \psi \, d\nu + \int_\Omega \psi^c \, d\mu. \tag{0.2.1}$$

Theorem 0.2.1. *If Ω is compact, for any $p \geq 1$ the function W_p is in fact a distance over $\mathcal{P}(\Omega)$ and the convergence with respect to this distance is equivalent to the weak convergence of probability measures. In particular any functional $\mu \mapsto W_p(\mu, \nu)$ is continuous with respect to weak topology.*

The case of a noncompact Ω is a little more difficult. First, the distance must be defined only on a subset of the whole space of probability measures, to avoid infinite values. We will use the space of probabilities with finite $p-$th momentum:

$$\mathcal{W}_p(\Omega) = \{\mu \in \mathcal{P}(\Omega) \, : \, M_p(\mu) := \int_\Omega |x|^p \mu(dx) < +\infty\}.$$

Theorem 0.2.2. *For any $p \geq 1$ the function W_p is a distance over $\mathcal{W}_p(\Omega)$ and, given a measure μ and a sequence $(\mu_n)_n$ in $\mathcal{W}_p(\Omega)$, the following are equivalent:*

- *$\mu_n \to \mu$ according to W_p;*
- *$\mu_n \rightharpoonup \mu$ and $M_p(\mu_n) \to M_p(\mu)$;*
- *$\int_\Omega \phi \, d\mu_n \to \int_\Omega \phi \, d\mu$ for any $\phi \in C^0(\Omega)$ whose growth is at most of order p (i.e. there exist constants A and B depending on ϕ such that $\phi(x) \leq A + B|x|^p$ for any x).*

Notice that, as a consequence of Hölder inequalities, the Wasserstein distances are always ordered, i.e. $W_{p_1} \leq W_{p_2}$ if $p_1 \leq p_2$. Reversed inequalities are possible only if Ω is bounded, and in this case we have, if set $D = \operatorname{diam}(\Omega)$, for $p_1 \leq p_2$,

$$W_{p_1} \leq W_{p_2} \leq D^{1-p_1/p_2} W_{p_1}^{p_1/p_2}.$$

From the monotone behavior of Wasserstein distances with respect to p it is natural to introduce the following distance W_∞: set $\mathcal{W}_\infty(\Omega) = \{\mu \in \mathcal{P}(\Omega) \,:\, \operatorname{spt}(\mu) \text{ is bounded }\}$ (obviously if Ω itself is bounded one has $\mathcal{W}_\infty(\Omega) = \mathcal{P}(\Omega)$) and then

$$W_\infty(\mu, \nu) = \inf \left\{ \pi - \operatorname{ess} \sup_{x,y \in \Omega \times \Omega} |x - y| \,:\, \pi \in \Pi(\mu, \nu) \right\}.$$

It is easy to check that $W_p \nearrow W_\infty$ and it is interesting to study the metric space $\mathcal{W}_\infty(\Omega)$. Curiously enough, this supremal problem in optimal transport theory, even if quite natural, has not deserved much attention up to now.

The W_∞ convergence is stronger than any W_p convergence and hence also than the weak convergence of probability measures. The converse is not true and W_∞ converging turns out to be actually rare: consequently there is a great lack of compactness in $\mathcal{W}_\infty$. For instance it is not difficult to check that, if we set $\mu_t = t\delta_{x_0} + (1-t)\delta_{x_1}$, where $x_0 \neq x_1 \in \Omega$, we have $W_\infty(\mu_t, \mu_s) = |x_0 - x_1|$ if $t \neq s$. This implies that the balls $B(\mu_t, |x_0 - x_1|/2)$ are infinitely many disjoint balls in $\mathcal{W}_\infty$ and prevents compactness.

The following statement summarizes the compactness properties of the spaces W_p for $1 \leq p \leq \infty$ and its proof is a direct application of the considerations above and of Theorem 0.2.2.

Proposition 0.2.3. *For $1 \leq p < \infty$ the space $W_p(\Omega)$ is compact if and only if Ω itself is compact. Moreover, for an unbounded Ω the space $W_p(\Omega)$ is not even locally compact. The space $W_\infty(\Omega)$ is neither compact nor locally compact for any choice of Ω with $\sharp\Omega > 1$.*

0.3 Geodesics, continuity equation and displacement convexity

We are concerned in this sections with several properties linked to the curves in the Wasserstein space W_p. For this subject the main reference is [4].

Before giving the main result we are interested in, we recall the definition of metric derivative, which is a concept that may be useful when studying curves which are valued in generic metric spaces.

Definition 0.3.1. Given a metric space (X, d) and a curve $\gamma : [0, 1] \to X$ we define *metric derivative* of the curve γ at time t the quantity

$$|\gamma'|(t) = \lim_{s \to t} \frac{d(\gamma(s), \gamma(t))}{|s - t|}, \tag{0.3.1}$$

provided the limit exists.

As a consequence of Rademacher Theorem it can be seen (see [7]) that for any Lipschitz curve the metric derivative exists at almost every point $t \in [0, 1]$. We will be concerned quite often with metric derivatives of curves which are valued in the space $\mathcal{W}_p(\Omega)$.

Definition 0.3.2. If we are given a Lipschitz curve $\mu : [0, 1] \to W_p(\Omega)$, we define velocity field of the curve any vector field $v : [0, 1] \times \Omega \to \mathbb{R}^d$ such that for a.e. $t \in [0, 1]$ the vector field $v_t = v(t, \cdot)$ belongs to $[L^p(\mu_t)]^d$ and the *continuity equation*

$$\frac{d}{dt}\mu_t + \nabla \cdot (v \cdot \mu_t) = 0$$

is satisfied in the sense of distributions: this means that for all $\phi \in C_c^1(\Omega)$ and any $t_1 < t_2 \in [0, 1]$ it holds

$$\int_\Omega \phi \, d\mu_{t_2} - \int_\Omega \phi \, d\mu_{t_1} = \int_{t_1}^{t_2} ds \int_\Omega \nabla\phi \cdot v_s \, d\mu_s,$$

or, equivalently, in differential form:

$$\frac{\partial}{\partial t} \int_\Omega \phi \, d\mu_t = \int_\Omega \nabla\phi \cdot v_t \, d\mu_t \qquad \text{for a.e. } t \in [0, 1].$$

We say that v is the *tangent* field to the curve μ_t if, for a.e. t, v_t has minimal $[L^p(\mu_t)]^d$ norm for any t among all the velocity fields.

The following proposition is concerned with the existence of tangent fields and comes from Theorem 8.3.1 and Proposition 8.4.5 in [4].

Theorem 0.3.3. *If $p > 1$ and $\mu = (\mu_t)_t$ is a curve in* $\mathrm{Lip}([0,1]; W_p(\Omega))$ *then there exist unique a tangent vector field v, and it is characterized by*

$$\frac{\partial}{\partial t}\mu + \nabla \cdot (v \cdot \mu) = 0, \tag{0.3.2}$$

$$||v_t||_{L^p(\mu_t)} \leq |\mu'|(t) \ \textit{for a.e. } t, \tag{0.3.3}$$

where the continuity equation is satisfied in the sense of distributions as previously explained. Moreover, if (0.3.2) *holds for a family of vector fields $(v_t)_t$ with $||v_t||_{L^p(\mu_t)} \leq C$ then $\mu \in \mathrm{Lip}([0,1]; \mathcal{W}_p(\Omega))$ and $|\mu'|(t) \leq ||v_t||_{L^p(\mu_t)}$ for a.e. t.*

This characterization of Lipschitz (or, up to reparameterization, absolutely continuous) curves in W_p will be very useful in Chapter 5. Moreover, in the general theory of Wasserstein spaces, it is a key instrument for studying geodesics and other properties linked to them, which will be used in Chapters 2, 5, 6 and 7.

The following result is a characterization of geodesics in $W_p(\Omega)$ when Ω is a convex domain in $\mathbb{R}^d$ (see Proposition 7.2.2 in [4], but some extension to the case of length spaces, for instance non convex domains, may be found in [54]). This procedure is also known as *McCann's linear interpolation.*

Theorem 0.3.4. *All the spaces $\mathcal{W}_p(\Omega)$ are length spaces and if μ and ν belong to $\mathcal{W}_p(\Omega)$, any geodesic curve linking them, when parametrized by arc-length, is of the form*

$$\gamma^\pi(s) = (p_s)_\sharp \pi$$

where $p_s : \Omega \times \Omega \to \Omega$ is given by $p_s(x,y) = x + s(y-x)$ and π is an optimal transport plan from μ to ν for the cost $c_p(x,y) = |x-y|^p$. In the case $p > 1$ and μ, ν absolutely continuous, if T is the corresponding optimal transport map such that $\pi = (id \times T)_\sharp \mu$,then the curve has the form

$$\gamma^\pi(s) = [(1-s)Id + sT]_\# \mu.$$

Conversely, any curve of this form, for a transport plan π or a transport map T, is an arc-length geodesic.

By means of this characterization of geodesics we can also define the useful concept of displacement convexity introduced by McCann in [58].

Definition 0.3.5. Given a functional $F : \mathcal{W}_p(\Omega) \cap L^1 \to [0, +\infty]$, we say that it is *displacement convex* if all the maps $t \mapsto F(\gamma^\pi(t))$ are convex on $[0,1]$ for every choice of μ and ν in $\mathcal{W}_p(\Omega)$ and π optimal transport plan from μ to ν with respect to $c(x,y) = |x-y|^p$.

The following well-known result provides a wide set of displacement convex functionals. In the case $p = 2$ this result is due to McCann ([58]), while the generalization to any p can be found in [4].

Theorem 0.3.6. *Consider the following functionals on the space $\mathcal{W}_p(\Omega)$, where Ω is any convex subset of $\mathbb{R}^N$:*

$$\begin{aligned}
J^1(\mu) &= \begin{cases} \int_\Omega f(u(x))\,dx & \text{if } \mu = u \cdot \mathcal{L}^d \\ +\infty & \text{if } \mu \text{ is not absolutely continuous}; \end{cases} \\
J^2(\mu) &= \int_\Omega V(x)\,\mu(dx); \\
J^3(\mu) &= \int_\Omega \int_\Omega w(x-y)\mu(dx)\mu(dy).
\end{aligned}$$

Suppose $f : [0, +\infty] \to [0, +\infty]$ is a convex and superlinear lower semicontinuous function with $f(0) = 0$, $V : \Omega \to [0, +\infty]$ and $w : \mathbb{R}^d \to [0, +\infty]$ are convex functions. Then the functionals J^2 and J^3 are displacement convex in $\mathcal{W}_p(\Omega)$ and the functional J^1 is displacement convex provided the following additional condition holds: the map

$$r \to r^d f(r^{-d})$$

is required to be convex and non-increasing on $]0, +\infty[$.

0.4 Monge-Ampère equation and regularity

The next step of our analysis is concerned with some regularity properties of t and ψ (the optimal transport map and the Kantorovich potential, respectively) and their relations with the densities of μ and ν. We will consider only the quadratic case $c(x, y) = |x-y|^2/2$, because it is the one where more results have been proven. Very recent results for generic costs may be found in [70].

It is easy, just by a change-of-variables formula, to transform, in the case of regular maps and densities, the equality $\nu = T_\sharp\mu$ into the PDE $v(t(x)) = u(x)/|Jt|(x)$, where u and v are the densities of μ and ν and J denotes the determinant of the Jacobian matrix. Recalling that we may write $t = \nabla\phi$ with ϕ convex (Remark 0.1.13), we get the Monge-Ampère equation

$$M\phi = \frac{u}{v(\nabla\phi)}, \tag{0.4.1}$$

where M denotes the determinant of the Hessian

$$M\phi = \det H\phi = \det \left[\frac{\partial^2 \phi}{\partial x_i \, \partial x_j} \right]_{i,j}.$$

This equation up to now is satisfied by $\phi = id - \psi$ in a formal way only. We define various notions of solutions for (0.4.1):

- we say that ϕ satisfies (0.4.1) in the Brenier sense if $(\nabla\phi)_\sharp u \cdot \mathcal{L}^d = v \cdot \mathcal{L}^d$ (and this is actually the sense to be given to this equation);
- we say that ϕ satisfies (0.4.1) in the Alexandroff sense if $H\phi$, which is always a positive measure for ϕ convex, is absolutely continuous and its density satisfies (0.4.1) a.e.;
- we say that ϕ satisfies (0.4.1) in the viscosity sense if it satisfies the usual comparison properties required by viscosity theory but restricting the comparisons to regular convex test functions (since M is in fact monotone just when restricted to positively definite matrices);
- we say that ϕ satisfies (0.4.1) in the classical sense if it is of class C^2 and the equation holds pointwise.

Notice that any notion except the first may be also applied to the equation $M\phi = f$, while the first one just applies to this specific transportation case. The results we want to use are well summarized in Theorem 50 of [71]:

Theorem 0.4.1. *If u and v are $C^{0,\alpha}(\Omega)$ and are both bounded from above and from below on the whole Ω by positive constants and Ω is a convex open set, then for the unique Brenier solution ϕ of* (0.4.1) *it holds $\phi \in C^{2,\alpha}(\Omega) \cap C^{1,\alpha}(\overline{\Omega})$ and ϕ satisfies the equation in the classical sense (hence also in the Alexandroff and viscosity senses).*

Even if this precise statement is taken from [71], we just detail a possible bibliographical path to arrive at this result. It is not easy to deal with Brenier solutions, so the idea is to consider viscosity solutions, for which it is in general easy to prove existence by Perron's method. Then prove some regularity result on viscosity solutions, up to getting a classical solution. Then, once we have a classical convex solution to Monge-Ampère equation, this will be a Brenier solution too. Since this is unique (up to additive constants) we have got a regularity statement for Brenier solutions. We can find results on viscosity solutions in [31], [33] and [32]. In [31] some conditions to ensure strict convexity of the solution of $M\phi = f$ when f is bounded from above and below are given. In [33] for the same equation it is proved $C^{1,\alpha}$ regularity provided we have strict convexity. In this way the term $u/v(\nabla\phi)$ becomes a $C^{0,\alpha}$ function and in [32] it is proved $C^{2,\alpha}$ regularity for solutions of $M\phi = f$ with $f \in C^{0,\alpha}$.

Chapter 1

An urban planning model by local functionals

This chapter mainly contains results from [28], a joint work with G. Buttazzo which was written in 2003, right after the Laurea Thesis [66]. As it is the oldest of the papers concerned by this thesis, we provide here a slightly different version which has undergone some changes with respect to the published one. In particular the alternative proof for the main theorem has been well detailed and some differences in the presentation of the whole subject may be found.

1.1 Overall optimization of residence and working areas

The efficient planning of a city is a quite complicated problem, possibly involving a huge number of parameters (population density, price of the land, kind and location of the industries working in the area, quality of the life, prices and time for transportations, geographical obstacles, ...). In this study we want to present a simplified model involving only the distribution of inhabitants and of services in the urban area under consideration.

The geographical area will be considered as given and represented by a subset Ω of $\mathbb{R}^d$ ($d = 2$ in the applications to concrete urban planning problems). We want to study the optimal location in Ω of a mass of inhabitants, that we denote by μ, as well as of a mass of services (working places, stores, offices, ...), that we denote by ν. We assume that μ and ν are probability measures on Ω. This corresponds to say that the total amounts of population and production are fixed as problem data: they are exogenous, in economical language. The measures μ and ν represent the unknowns of the problem and have to be found in such a way that they satisfy some criteria:

i) there is a transportation cost for commuting from the residential areas to the services areas;

ii) people desire to live in areas with low population density;

iii) services need to be concentrated as much as possible, in order to increase efficiency and decrease management costs.

Fact i) will be described through a Monge-Kantorovich mass transportation model; the transportation cost will be indeed given by using a p-Wasserstein distance ($p \geq 1$, see Section 0.2): we set $T_p(\mu, \nu) = W_p^p(\mu, \nu)$.

Fact ii) will be described by a penalization functional F, a kind of total unhappiness of citizens due to high density of population, obtained by integrating with respect to the citizens' density their personal unhappiness.

Fact iii) is modeled by a third term G representing the costs for managing services once they are located according to the distribution ν, taking into account that efficiency depends strongly on how much ν is concentrated.

An interesting mathematical model for the description of the equilibrium structure of a city is presented by Carlier and Ekeland in [35]. The same criteria (concentration of services and dispersion of inhabitants) appear and transportation costs are considered as well. Moreover, Monge-Kantorovich optimal transport theory plays an important role. Anyway, the goal is very different since in [35] there is no total performance to be optimized. On the other hand, in this chapter we are precisely looking for a configuration of inhabitants and services which optimizes an overall utility criterion. This will be made by minimizing a suitable total cost functional $\mathfrak{F}(\mu, \nu)$.

The cost functional we will consider is

$$\mathfrak{F}(\mu, \nu) = T_p(\mu, \nu) + F(\mu) + G(\nu) \tag{1.1.1}$$

(notice that we will also refer to this functional as $\mathfrak{F}^p$, when we will need to let p vary) and so the optimal location of μ and ν will be determined by the minimization problem

$$\min \left\{ \mathfrak{F}(\mu, \nu) \ : \ \mu, \nu \text{ probabilities on } \Omega \right\}. \tag{1.1.2}$$

In this way this optimization problem falls into the wider subject of transport and concentration problems, which will be presented in its generality in Chapter 3. In this particular case both F and G will be chosen among local semicontinuous functionals over measures. These functionals have been widely studied by Bouchitté and Buttazzo in [16], [17] and [18] and are briefly presented in the next section.

1.2 Local semicontinuous functionals on measures

The general theory of local functionals over measures has been developed in the framework of vector measures on a metric space Ω. Consequently, we provide here very general definitions and concepts which are valid in $\mathcal{M}^k(\Omega)$, even if later we will particularize our analysis to the case of positive scalar measures. Moreover, in the minimization problem for $\mathfrak{F}$, only probability measures will be actually concerned.

Definition 1.2.1. A functional $J : \mathcal{M}^k(\Omega) \to [0, +\infty]$ is said to be local if it is additive on mutually singular measures, i.e. $J(\mu_1 + \mu_2) = J(\mu_1) + J(\mu_2)$ whenever $\mu_i \in \mathcal{M}^k(\Omega)$ and $\mu_1 \perp \mu_2$.

In [16] and [17] the set of local functionals which are l.s.c. with respect to the weak convergence of measures is characterized as the set of those functionals having this general form:

$$J(\mu) = \int_\Omega f\left(\frac{d\mu}{d\mathrm{m}}\right)\, d\mathrm{m} + \int_{\Omega \setminus A_\mu} f^\infty\left(\frac{d\mu_s}{d|\mu_s|}\right)\, d|\mu_s| + \int_{A_\mu} g(\mu(\{x\}))\#(dx)$$

where

- m is a nonatomic positive measure on Ω;
- $d\mu/d\mathrm{m}$ is the Radon-Nicodym derivative of μ with respect to m;
- $f : \mathbb{R} \to]-\infty, +\infty]$ is convex, lower semicontinuous and proper (i.e. not identically $+\infty$);
- f^∞ is the *recession function* given by

  $$f^\infty(s) := \lim_{t \to +\infty} \frac{f(s_0 + ts)}{t} = \sup_{t>0} \frac{f(s_0 + ts) - f(s_0)}{t}$$

  (the limit is independent on the choice of s_0 in the domain of f, i.e. the set of points such that $f < +\infty$);
- A_μ is the set of atoms of μ, i.e. the points such that $\mu(\{x\}) > 0$;

- $g : \mathbb{R} \to [0, +\infty]$ is a lower semicontinuous subadditive function such that $g(0) = 0$;
- $\#$ is the counting measure;
- f and g satisfy a compatibility condition, namely for any s we have

$$g_0(s) := \lim_{t>0} \frac{g(st)}{t} = f^\infty(s).$$

Notice that these functional can be written in a simpler form since in the case of positive measures $d\mu_s/d|\mu_s| = 1$ for $|\mu_s|-$a.e. x:

$$J(\mu) = \int_\Omega f\left(\frac{d\mu}{d\mathrm{m}}\right)\, d\mathrm{m} + f^\infty(1)|\mu_s|(\Omega \setminus A_\mu) + \int_{A_\mu} g(\mu(x))\, d\#(x).$$

The key point is that, by the results that can be found in [16], these functionals are lower semicontinuous for the weak convergence. Notice that both f^∞ and g^0 are positive $1-$homogeneous functions. In particular, in the positive scalar case, the compatibility condition, which is crucial for semicontinuity, may be checked for $s = 1$ and written as

$$\lim_{t\to 0^+} \frac{g(t)}{t} = \lim_{t\to\infty} \frac{f(st)}{t}.$$

The main advantage of this class of functionals is that it contains both convex and nonconvex functionals (as opposed to what happens in the case of local semicontinuous functionals over L^p functions). In particular the two extreme cases are the ones we get if we let f or g be infinite. In fact, by choosing $g = \delta(\cdot|\{0\})$ (i.e. $g = +\infty$ on $]0, +\infty[$ and $g(0) = 0$), together with a function f such that $\lim_{t\to+\infty} f(t)/t = +\infty$ we get the following functional:

$$F(\mu) = \begin{cases} \int_\Omega f(u) d\mathrm{m} & \text{if } \mu = u \cdot \mathrm{m}; \\ +\infty & \text{if } \mu \text{ is not absolutely continuous w.r.t. m.} \end{cases}$$

Analogously, by setting $f = \delta(\cdot|\{0\})$, together with a function g such that $\lim_{s\to 0^+} g(s)/s = +\infty$ we get

$$G(\mu) = \begin{cases} \sum_i g(a_i) & \text{if } \mu = \sum_i a_i \delta_{x_i}; \\ +\infty & \text{if } \mu \text{ is not atomic.} \end{cases}$$

Typical cases are $f(s) = s^q$ and $g(s) = s^\alpha$, for exponents $q > 1$ and $\alpha < 1$. In general, g is often chosen to be concave, even if subadditivity would be sufficient to apply the general existence theory.

1.3 Interpretation of the model

To define the three terms appearing in our functional $\mathfrak{F}$, we must precise the choices for F and G, since the first term will be a Monge-Kantorovich transport cost, as explained in Section 1.2. For the functional F we take

$$F(\mu) = \begin{cases} \int_\Omega f(u(x))\,dx & \text{if } \mu = u \cdot \mathcal{L}^d,\ u \in L^1(\Omega) \\ +\infty & \text{otherwise,} \end{cases} \tag{1.3.1}$$

where the integrand $f : [0, +\infty] \to [0, +\infty]$ is assumed to be lower semicontinuous and convex, with $f(0) = 0$ and superlinear at infinity, that is

$$\lim_{t \to +\infty} \frac{f(t)}{t} = +\infty. \tag{1.3.2}$$

In this form we have a particular local semicontinuous functional on measures (with m $= \mathcal{L}^d$). Without loss of generality, by subtracting constants to the functional F, we can suppose $f'(0) = 0$. Due to the assumption $f(0) = 0$, the ratio $f(t)/t$ is an incremental ratio of the convex function f and so it is increasing in t. Then, if we write the functional F as

$$\int_\Omega \frac{f(u(x))}{u(x)} u(x)\,dx,$$

we can see the quantity $f(u)/u$, which is increasing in u and tends to ∞ as $u`to\infty$, as the unhappiness of a single citizen when he lives in a place where the population density is u. Integrating it with respect to $\mu = u \cdot \mathcal{L}^d$ gives the total unhappiness of the population.

As far as the concentration term $G(\nu)$ is concerned, we set

$$G(\nu) = \begin{cases} \sum_{i=0}^\infty g(a_i) & \text{if } \nu = \sum_{i=0}^\infty a_i \delta_{x_i} \\ +\infty & \text{if } \nu \text{ is not atomic.} \end{cases} \tag{1.3.3}$$

To the function g we require to be subadditive, lower semicontinuous and such that $g(0) = 0$ and

$$\lim_{t \to 0} \frac{g(t)}{t} = +\infty. \tag{1.3.4}$$

Every single term $g(a_i)$ in the sum in (1.3.3) represents the cost for building and managing a service pole of size a_i, located at the point $x_i \in \Omega$.

In this model, as already pointed out, we fix as a datum the total production of services; moreover, in each service pole the production is required as a quantity proportionally depending on its size (or on the number of inhabitants using such a pole). We may define the productivity P of a pole of mass (size) a as the ratio between the production and the cost to get such a production. Then we have $P(a) = a/g(a)$ and

$$\sum_{i=0}^{\infty} g(a_i) = \sum_{i=0}^{\infty} \frac{a_i}{P(a_i)}.$$

As a consequence of the assumption (1.3.4), the productivity in very small service poles is near 0. When g is also concave, for instance in the case of powers $g(s) = s^r$, the productivity is also an increasing function of the size of the pole.

Notice that in the functional G we do not take into account distances between service poles. It would be interesting to consider also non local functionals involving such distances, taking into account possible interactions and the consequent gain in efficiency. A study of the corresponding problem for an interaction functional G can be found in Chapter 3. The results of next section (since they do not depend on the choice of G) will be used there as well as in the present setting.

For the problem introduced in (1.1.2) existence results are straightforward, especially when we use as an environment a compact set Ω.

Theorem 1.3.1. *Suppose Ω is a domain, $p \geq 1$ and f and g satisfy the conditions listed above. Then the minimization problem (1.1.2) has at least one solution.*

Proof. By the direct method of Calculus of Variations, this result is an easy consequence of the weak compactness of the space $\mathcal{P}(\Omega)$, the space of probability measures on Ω, when Ω itself is compact, and of the weak semicontinuity of the functional $\mathfrak{F}$. The second and third term in the expression (1.1.1) are in fact local semicontinuous functionals (due to results in [16]), while the first term is nothing but a Wasserstein distance raised to a certain power. Since it is known that in compact spaces this distance metrizes the weak topology, T_p is actually continuous. $\square$

In [66], where we first presented the model, other existence results were shown, for instance in the case of a non compact bounded convex set $\Omega \subset \mathbb{R}^d$. Here we will not go through this proof, and will discuss only one existence result in a non-compact setting, obtained as a consequence of a proper use of the optimality conditions presented in next section.

1.4 Necessary optimality conditions on $\mathfrak{F}_\nu$

In this section we find optimality conditions for probability measures on Ω minimizing the functional

$$\mathfrak{F}_\nu(\mu) = T_p(\mu, \nu) + F(\mu).$$

It is clear that, if (μ, ν) is an optimal pair for the whole functional $\mathfrak{F}$, it happens that μ is a minimizer for $\mathfrak{F}_\nu$. We will come back later, in the next section, to the problem of minimizing $\mathfrak{F}$, and we will refer to it as the *whole minimization* problem.

The goal of this section is to derive optimality conditions for $\mathfrak{F}_\nu$, for any ν, without any link to the minimization of $\mathfrak{F}$. There are several different proof for this result, all based on a derivation of the functional $\mathfrak{F}_\nu$. The idea is not difficult but some technical problems, mainly linked to the lack of uniqueness of Kantorovich potentials, arise. We provide here to two different proofs, and another one was present in [66]. The former we give here, as well as the one in [66], rely on a regularizing approach: we start by the easier case $p > 1$ and ν "regular" in some sense, and then recover the general case as a limit. The reason to do so are the conditions ensuring uniqueness properties of the Kantorovich potential presented in Section 0.1. The same idea can be found in [66], where purely atomic probability measures (i.e. finite sums of Dirac masses) were first considered and then, by approximation, the result was extended to any measure ν. The second proof, suggested by an anonimous referee while he/she was reviewing [28], is based on some convex analysis tools and strongly uses the convex structure of the problem.

In the sequel the function f in (1.3.1) will be assumed to be strictly convex, C^1 and with polynomial growth, and we will denote by k the continuous, strictly increasing function $(f')^{-1}$. Strict convexity of f will ensure uniqueness for the minimizer of $\mathfrak{F}_\nu$. Typical choices are $f(t) = t^q$, $q > 1$.

1.4.1 An approximation proof

Lemma 1.4.1. *If μ is optimal for $\mathfrak{F}_\nu$ then, for any other probability measure μ_1 with density u_1, such that $\mathfrak{F}_\nu(\mu_1) < +\infty$, the following inequality holds:*

$$T_p(\mu_1, \nu) - T_p(\mu, \nu) + \int_\Omega f'(u(x))[u_1(x) - u(x)]dx \geq 0.$$

Proof. For any $\varepsilon > 0$, due to the convexity of the transport term and the minimality of μ, it holds

$$
\begin{aligned}
T_p(\mu,\nu) + F(\mu) &\leq T_p(\mu + \varepsilon(\mu_1 - \mu)) + F(\mu + \varepsilon(\mu_1 - \mu), \nu) \\
&\leq T_p(\mu,\nu) + \varepsilon(T_p(\mu_1,\nu) - T_p(\mu,\nu)) + F(\mu + \varepsilon(\mu_1 - \mu)).
\end{aligned}
$$

We deduce that the quantity

$$
T_p(\mu_1,\nu) - T_p(\mu,\nu) + \varepsilon^{-1}\left[F(\mu + \varepsilon(\mu_1 - \mu)) - F(\mu)\right]
$$

is nonnegative. If we let $\varepsilon \to 0$ we obtain the thesis if we prove

$$
\lim_{\varepsilon \to 0} \int \frac{f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon}\, d\mathcal{L}^d = \int f'(u)(u_1 - u)\, d\mathcal{L}^d.
$$

To prove this, notice that by convexity the inequality

$$
\int \frac{f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon}\, d\mathcal{L}^d \geq \int f'(u)(u_1 - u)\, d\mathcal{L}^d
$$

is straightforward. For the opposite inequality, we will use Fatous's Lemma. the pointwise convergence of the integrand is trivial and we can get an upper bound by means ofthe following inequality, which is a consequence, for $\varepsilon < 1$, of the monotonicity of the incremental ratios of convex functions:

$$
\frac{f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon} \leq f(u_1) - f(u).
$$

Since we have $f(u),\ f(u_1) \in L^1(\Omega)$, this is sufficient to apply Fatou's Lemma and get

$$
\limsup_{\varepsilon \to 0} \int \frac{f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon}\, d\mathcal{L}^d \leq \int f'(u)(u_1 - u)\, d\mathcal{L}^d. \quad \square
$$

Lemma 1.4.2. *Let us suppose $\nu = \nu^s + v \cdot \mathcal{L}^d$, with $v \in L^\infty(\Omega)$, $\nu^s \perp \mathcal{L}^d$, $v > 0$ a.e. in Ω. If μ is optimal for $\mathfrak{F}_\nu$, then $u > 0$ a.e. in Ω.*

Proof. The Lemma will be proved by contradiction. If the set $A = \{u = 0\}$ is not negligible, we will find a measure μ_1 for which Lemma 1.4.1 is not verified. Let N be a Lebesgue-negligible set where ν^s is concentrated and T an optimal transport map between μ and ν. Such an optimal transport exists, since $\mu \ll \mathcal{L}^d$, see Section 0.1.

Let $B = T^{-1}(A)$. Up to modifying t on the $\mu-$negligible set A, we may suppose $B \cap A = \emptyset$. Set $\mu_1 = 1_{B^c} \cdot \mu + 1_{A \setminus N} \cdot \nu$: it is a probability measure with density u_1 given by $1_{B^c}u + 1_A v = 1_{B^c\setminus A}u + 1_A v$ (this equality comes from $u = 0$ on A). We have

$$F(\mu_1) = \int_{B^c\setminus A} f(u)\, d\mathcal{L}^d + \int_A f(v)\, d\mathcal{L}^d \leq F(\mu) + ||f(v)||_\infty |\Omega| < +\infty.$$

Setting

$$T^*(x) = \begin{cases} T(x) & \text{if } x \in (A\cup B)^c \\ x & \text{if } x \in (A\cup B) \end{cases},$$

we can see that T^* is a transport map between μ_1 and ν. In fact, for any Borel set $E \subset \Omega$, we may express $(T^*)^{-1}(E)$ as the disjoint union of $E\cap A$, $E\cap B$ and $T^{-1}(E)\cap B^c \cap A^c$, and so

$$\begin{aligned} \mu_1((T^*)^{-1}(E)) &= \nu(E\cap A) + \nu(E\cap B\cap A) + \mu(T^{-1}(E)\cap B^c\cap A^c) \\ &= \nu(E\cap A) + \mu(T^{-1}(E\cap A^c)) = \nu(E), \end{aligned}$$

were we used the fact that $A\cap B = \emptyset$ and that A^c is a set of full measure for μ. Consequently,

$$T_p(\mu_1,\nu) \leq \int_{(A\cup B)^c} |x - T(x)|^p u(x)dx < \int_\Omega |x - T(x)|^p u(x)dx = T_p(\mu,\nu). \tag{1.4.1}$$

From this it follows that for μ_1 Lemma 1.4.1 is not satisfied, as the integral term $\int_\Omega f'(u)(u_1 - u)d\mathcal{L}^d$ is non-positive, because $u_1 > u$ only on A, where $f'(u)$ vanishes. The strict inequality in (1.4.1) follows from the fact that, if $\int_{A\cup B} |x - T(x)|^p u(x)dx = 0$ then for a.e. $x \in B$ we have $u(x) = 0$ or $x = T(x)$, which, by definition of B, implies $x \in A$: in both cases we are led to $u(x) = 0$. This would give $\nu(A) = \mu(B) = 0$, contradicting the assumptions $|A| > 0$ and $v > 0$ a.e. in Ω. □

from now on we will need some of the results from duality theory in mass transportation that we presented in Section 0.1.

Theorem 1.4.3. *Under the same hypotheses of Lemma 1.4.2, assuming also $p > 1$, if μ is optimal for $\mathfrak{F}_\nu$ and we denote by ψ the unique, up to additive constants, Kantorovich potential for the transport between μ and ν, there exists a constant l such that the following relation holds:*

$$u = k(l - \psi) \text{ a.e. in } \Omega. \tag{1.4.2}$$

Proof. Let us choose an arbitrary measure μ_1 with bounded density u_1 (so that $F(\mu_1) < +\infty$) and define $\mu_\varepsilon = \mu + \varepsilon(\mu_1 - \mu)$. Let us denote by ψ_ε a Kantorovich potential between μ_ε and ν, chosen so that all the functions ψ_ε vanish at a same point. We can use the optimality of μ to write

$$T_p(\mu_\varepsilon, \nu) + F(\mu_\varepsilon) - T_p(\mu, \nu) - F(\mu) \geq 0.$$

By means of the duality formula, as $T_p(\mu_\varepsilon, \nu) = \int \psi_\varepsilon d\mu_\varepsilon + \int \psi_\varepsilon^c d\nu$ and $T_p(\mu, \nu) \geq \int \psi_\varepsilon d\mu + \int \psi_\varepsilon^c d\nu$, we can write

$$\int \psi_\varepsilon \, d(\mu_\varepsilon - \mu) + F(\mu_\varepsilon) - F(\mu) \geq 0.$$

Recalling that $\mu_\varepsilon - \mu = \varepsilon(\mu_1 - \mu)$ and that

$$F(\mu_\varepsilon) - F(\mu) = \int (f(u + \varepsilon(u_1 - u)) - f(u)) \; d\mathcal{L}^d,$$

we can divide by ε and pass to the limit. We know from Lemma 1.4.4 that ψ_ε converge towards the unique Kantorovich potential ψ for the transport between μ and ν. For the limit of the F part we use Fatou's Lemma, as in Lemma 1.4.1. We then obtain at the limit

$$\int_\Omega (\psi(x) + f'(u(x)))(u_1(x) - u(x))dx \geq 0.$$

This means that for every probability μ_1 with bounded density u_1 we have

$$\int (\psi(x) + f'(u(x)))u_1(x) \, dx \geq \int (\psi(x) + f'(u(x)))u(x) \, dx.$$

Define first $l = \operatorname{ess\,inf}_{x \in \Omega} \psi(x) + f'(u(x))$. The left hand side, by choosing properly u_1, can be made as close as we want to l. Then we get that the function $\psi + f'(u)$, which is $\mathcal{L}^d-$a.e. (and consequently also μ-a.e.) greater than l, integrated with respect to the probability μ gives a result less or equal than l. It follows

$$\psi(x) + f'(u(x)) = l \quad \mu - a.e. \, x \in \Omega.$$

since by Lemma 1.4.2 we know $u > 0$ a.e., we get an equality valid $\mathcal{L}^d-$a.e.:

$$f'(u) = l - \psi. \tag{1.4.3}$$

We can then compose with k and get the thesis. □

To establish Lemma 1.4.4, that we used in the proof of Theorem 1.4.3, we will use uniqueness properties for Kantorovich potentials when the absolutely continuous part of one of the measures has strictly positive density a.e. in the domain Ω. Notice that proving Lemma 1.4.1 was in fact not essential to get this uniqueness, as in fact we had already supposed that the density of ν was positive (and in fact one of the two densities is sufficient to ensure uniqueness). On the other hand, having $u > 0$ guarantees that (1.4.3) is valid a.e. and not only $\mu-$a.e.

Lemma 1.4.4. *Let ψ_ε be Kantorovich potentials for the transport between $\mu_\varepsilon = \mu + \varepsilon(\mu_1 - \mu)$ and ν, all vanishing at a same point $x_0 \in \Omega$. Suppose that $\mu = u \cdot \mathcal{L}^d$ and $u > 0$ a.e. in Ω and let ψ be the unique Kantorovich potential between μ and ν vanishing at the same point: then ψ_ε converge uniformly to ψ.*

Proof. First, notice that the family $(\psi_\varepsilon)_\varepsilon$ is equicontinuous since any function which is $c-$concave with respect to the cost $c(x, y) = |x - y|^p$ is $pD^{p-1}-$Lipschitz continuous. Moreover, thanks to $\psi_\varepsilon(x_0) = 0$, we get also equiboundedness, and so, by Ascoli-Arzelà Theorem, the existence of uniform limits up to subsequences. Let $\overline{\psi}$ be one of these limits, arising from a certain subsequence. From the optimality of ψ_ε in the duality formula for μ_ε and ν we have, for any $c-$concave function φ,

$$\int \psi_\varepsilon \, d\mu_\varepsilon + \int \psi_\varepsilon^c \, d\nu \geq \int \varphi \, d\mu_\varepsilon + \int \varphi^c \, d\nu.$$

We want to pass to the limit as $\varepsilon \to 0$: we have uniform convergence of ψ_ε but we need uniform convergence of ψ_ε^c as well. To get it, just notice

$$\psi_\varepsilon^c(x) = \inf_y |x - y|^p - \psi_\varepsilon(y), \quad \overline{\psi}^c(x) = \inf_y |x - y|^p - \overline{\psi}(y),$$
$$|\psi_\varepsilon^c(x) - \overline{\psi}^c(x)| \leq ||\psi_\varepsilon - \overline{\psi}||_\infty.$$

Passing to the limit as $\varepsilon \to 0$ along the considered subsequence we get, for any φ

$$\int \overline{\psi} \, d\mu + \int \overline{\psi}^c \, d\nu \geq \int \varphi \, d\mu + \int \varphi^c \, d\nu.$$

This means that $\overline{\psi}$ is a Kantorovich potential for the transport between μ and ν. Then, taking into account that $\overline{\psi}(x_0) = 0$, we get the equality $\overline{\psi} = \psi$. We can also derive that the whole sequence converges to ψ. □

We now highlight that the relation we have proved in Theorem 1.4.3 enables us to choose a density u which is continuous. Moreover, it is also continuous in a quantified way, since it coincides with k composed with a Lipschitz function with a fixed Lipschitz constant. As a next step we will try to extend these results to the case of general ν and to the case $p = 1$. The uniform continuity property we proved will be essential for an approximation process.

In order to go through our approximation approach, we need the following lemma, requiring the well-known theory of $\Gamma-$convergence. For all details about this theory, we refer to [39].

Lemma 1.4.5. *Given a sequence $(\nu_h)_h$ of probability measures on Ω, supposing $\nu_h \rightharpoonup \nu$ and $p > 1$, it follows that the sequence of functionals $(\mathfrak{F}^p_{\nu_h})_h$, $\Gamma-$converges to the functional $\mathfrak{F}^p_\nu$ with respect to weak$-*$ topology on $\mathcal{P}(\Omega)$. Moreover if ν is fixed and we let p vary, we have $\Gamma-$convergence, according to the same topology, of the functionals $(\mathfrak{F}^p_\nu)_p$ to the functional $\mathfrak{F}^1_\nu$ as $p \to 1$.*

Proof. For the first part of the statement, just notice that the Wasserstein distance is a metrization of weak$-*$ topology: consequently, being $T_p(\mu,\nu) = W_p^p(\mu,\nu)$, as $\nu_h \rightharpoonup \nu$ we have uniform convergence of the continuous functionals $T_p(\cdot,\nu_h)$. This implies $\Gamma-$convergence and pointwise convergence. In view of Proposition 6.25 in [39], concerning $\Gamma-$convergence of sums, we achieve the proof. The second assertion follows the same scheme, once we notice that, for each $p > 1$ and every pair (μ,ν) of probability measures, it holds

$$W_1(\mu,\nu) \leq W_p(\mu,\nu) \leq D^{1-1/p} W_1^{1/p}(\mu,\nu).$$

This gives uniform convergence of the transport term, as

$$\begin{aligned} T_p(\mu,\nu) - T_1(\mu,\nu) &\leq (D^{p-1} - 1)T_1(\mu,\nu) \\ &\leq D(D^{p-1} - 1) \to 0. \\ T_p(\mu,\nu) - T_1(\mu,\nu) &\geq T_1^p(\mu,\nu) - T_1(\mu,\nu) \\ &\geq (p-1)c(T_1(\mu,\nu)) \geq \overline{c}\,(p-1) \to 0, \end{aligned}$$

where $c(t) = t\log t$, $\overline{c} = \inf c$ and we used the fact $T_1(\mu,\nu) \leq D$. □

We now state in the form of lemmas two extensions of Theorem 1.4.3

Lemma 1.4.6. *Suppose $p > 1$ and fix an arbitrary $\nu \in \mathcal{P}(\Omega)$: if μ is optimal for $\mathfrak{F}_\nu$ then there exists a Kantorovich potential ψ for the transport between μ and ν such that Formula* (1.4.2) *holds.*

Proof. We choose a sequence $(\nu_h)_h$ approximating ν in such a way that each ν_h satisfies the assumptions of Theorem 1.4.3. By Lemma 1.4.5 and the properties of $\Gamma-$ convergence, the space $\mathcal{P}(\Omega)$ being compact and the functional $\mathfrak{F}_\nu$ having a unique minimizer (see, for instance, Chapter 7 in [39]), we get that $\mu_h \rightharpoonup \mu$, where each μ_h is the unique minimizer of $\mathfrak{F}_{\nu_h}$. Each measure μ_h is absolutely continuous with density u_h. We use (1.4.2) to express u_h in terms of Kantorovich potentials ψ_h and get uniform continuity estimates on u_h. We would like to extract converging subsequences by Ascoli-Arzelà Theorem, but we need also equiboundedness. We may obtain this by using together the integral bound $\int u_h d\mathcal{L}^d = \int k(-\psi_h) d\mathcal{L}^d = 1$ and the equicontinuity. So, up to subsequences, we have this situation:

$$\begin{gathered}\mu_h = u_h \cdot \mathcal{L}^d, \qquad u_h = k(-\psi_h),\\ u_h \to u, \qquad \psi_h \to \psi \text{ uniformly},\\ \mu_h \rightharpoonup \mu, \quad \mu = u \cdot \mathcal{L}^d, \qquad \nu_h \rightharpoonup \nu,\end{gathered}$$

where we have absorbed the constants l into the Kantorovich potentials. Clearly it is sufficient to prove that ψ is a Kantorovich potential between μ and ν to get our goal.

To see this, we consider that, for any $c-$concave function φ, it holds

$$\int \psi_h \, d\mu_h + \int \psi_h^c \, d\nu_h \geq \int \varphi \, d\mu_h + \int \varphi^c \, d\nu_h.$$

The thesis follows passing to the limit with respect to h, as in Lemma 1.4.4. □

Next step will be proving the same relation when ν is generic and $p = 1$. We are in the same situation as before, and we simply need approximation results on Kantorovich potentials, in the more difficult situation when the cost functions $c_p(x, y) = |x - y|^p$ vary with p.

Lemma 1.4.7. *Suppose $p = 1$ and fix an arbitrary $\nu \in \mathcal{P}(\Omega)$: if μ is optimal for $\mathfrak{F}_\nu^1$ then there exists a Kantorovich potential ψ for the transport between μ and ν with cost $c(x, y) = |x - y|$ such that Formula* (1.4.2) *holds.*

Proof. For any $p > 1$ we consider the functional $\mathfrak{F}_\nu$ and its unique minimizer μ_p. Thanks to Lemma 1.4.6 we get the existence of densities u_p and Kantorovich potential ψ_p between μ_p and ν with respect to the cost c_p, such that

$$\mu_p = u_p \cdot \mathcal{L}^d, \quad u_p = k(-\psi_p).$$

By Ascoli-Arzelà compacteness result, as usual, we may suppose, up to subsequences,

$$u_p \to u, \quad \psi_p \to \psi \text{ uniformly,}$$

and, due the $\Gamma-$convergence result in Lemma 1.4.5, since $\mathfrak{F}^1_\nu$ has a unique minimizer denoted by μ, we get also

$$\mu_p \rightharpoonup \mu, \qquad \mu = u \cdot \mathcal{L}^d.$$

As in Lemma 1.4.6, we simply need to prove that ψ is a Kantorovich potential between μ and ν for the cost c_1. The limit function ψ is Lipschitz continuous with Lipschitz constant less or equal than $\liminf_{p\to 1} pD^{p-1} = 1$, since it is approximated by ψ_p. Consequently ψ is $c-$concave for $c = c_1$. We need to show that it is optimal in the duality formula.

Let us recall that, for any real function φ and any cost function c, it holds $\varphi^{cc} \geq \varphi$ and φ^{cc} is a $c-$concave function whose $c-$transform is $\varphi^{ccc} = \varphi^c$. Consequently, by the optimality of ψ_p, we get

$$\int \psi_p d\mu_p + \int \psi_p^{c_p} d\nu \geq \int \varphi^{c_p c_p} d\mu_p + \int \varphi^{c_p} d\nu \geq \int \varphi d\mu_p + \int \varphi^{c_p} d\nu. \quad (1.4.4)$$

We want to pass to the limit in the inequality between the first and the last term. We start by proving that, for an arbitrary sequence $(\varphi_p)_p$, if $\varphi_p \to \varphi_1$, we have the uniform convergence $\varphi_p^{c_p} \to \varphi_1^{c_1}$. Let us take into account that we have uniform convergence on bounded sets of $c_p(x,y) = |x-y|^p$ to $c_1(x,y) = |x-y|$. Then we have

$$\varphi_p^{c_p}(x) = \inf_y |x-y|^p - \varphi_p(y), \quad \varphi_1^{c_1}(x) = \inf_y |x-y| - \varphi_1(y),$$
$$|\varphi_p^{c,p}(x) - \varphi_1^{c,1}(x)| \leq ||c_p - c_1||_\infty + ||\varphi_p - \varphi_1||_\infty,$$

which gives us the convergence we needed. If we apply it to the sequences $\phi_p = \psi_p 4and4\phi_p\phi$, we obtain, passing to the limit as $p \to 1$ in (1.4.4),

$$\int \psi \, d\mu + \int \psi^{c_1} \, d\nu \geq \int \varphi \, d\mu + \int \varphi^{c_1} \, d\nu.$$

By restricting this inequality to the set of c_1-concave functions we get that ψ is a Kantorovich potential for the transport between μ and ν and the cost c_1. □

We can now state the main Theorem of this section, whose proof consists only in putting together all the results we have obtained above.

Theorem 1.4.8. *Let Ω be a bounded domain in $\mathbb{R}^d$, f a C^1 strictly convex function, $p \geq 1$ and ν a probability measure on Ω: then there exists a unique measure $\mu \in \mathcal{P}(\Omega)$ minimizing $\mathfrak{F}_\nu$ and it is absolutely continuous with density u. Moreover, there exists a Kantorovich potential ψ for the transport between μ and ν and the cost $c(x, y) = |x - y|^p$ such that $u = k(-\psi)$, where $k = (f')^{-1}$.*

Consequences on the regularity of u come from this expression, which gives Lipschitz-type continuity, and from the relationship between Kantorovich potentials and optimal transport, which can be expressed through some PDEs. It is not difficult, for instance, in the case $p = 2$, to obtain a Monge-Ampère equation for the density u.

1.4.2 A convex analysis proof

The idea of this proof consists in looking at the subdifferential of the functional $\mathfrak{F}_\nu$, in order to get optimality conditions on the unique minimizer measure μ and its density u (here we will identify any absolutely continuous probability measure with its density). We provide first some lemmas.

Lemma 1.4.9. *If $F : X \to \mathbb{R}$ and $G : X \to \mathbb{R} \cup \{+\infty\}$ are convex functionals with $\partial F(u_0) = \{\xi_0\}$ and such that, for any $u_1 \in X$ we have*

$$(F(u_0 + \varepsilon(u_1 - u_0)) - F(u_0))/\varepsilon \to < u_1 - u_0, \xi_0 >,$$

then $\partial(F + G)(u_0) = \xi_0 + \partial G(u_0)$.

Proof. We only need to prove in fact that if $\xi \in \partial(F + G)(u_0)$, then $\xi - \xi_0 \in \partial G(u_0)$. To do this we take $u_1 \in \mathrm{dom}(F + G) = \mathrm{dom}(G)$, set $u_\varepsilon = u_0 + \varepsilon(u_1 - u_0)$, and we write

$$\frac{F(u_\varepsilon) - F(u_0)}{\varepsilon} + \frac{G(u_\varepsilon) - G(u_0)}{\varepsilon} \geq < u_1 - u_0, \xi_0 > + < u_1 - u_0, \xi - \xi_0 > .$$

Passing to the limit as $\varepsilon \to 0$ gives

$$< u_1 - u_0, \xi - \xi_0 > \leq \lim_{\varepsilon \to 0} \frac{G(u_\varepsilon) - G(u_0)}{\varepsilon} \leq G(u_1) - G(u_0),$$

where the last inequality follows from convexity and gives $\xi - \xi_0 \in \partial G(u_0)$. □

In this subsection, the $c-$transform of an L^1 function will be defined replacing the inf by an ess inf, i.e.

$$\phi^c(x) = \operatorname*{ess\,inf}_y c(x, y) - \phi(y).$$

Lemma 1.4.10. *For any L^1 function ϕ the inequality $\phi^{cc} \geq \phi$ is true almost everywhere.*

Proof. For any x we have

$$\phi^{cc}(x) = \operatorname*{ess\,inf}_y \left[c(x,y) - \operatorname*{ess\,inf}_z \left(c(y,z) - \phi(z) \right) \right].$$

Notice that, in general $\operatorname{ess\,inf}_z \xi(z) = \inf_{z \in A} \xi(z)$ if A is the set of Lebesgue point of ξ. Since c is continuous, the Lebesgue points of $c(y,\cdot) - \phi(\cdot)$ coincide with the Lebesgue points of ϕ. Hence, if x is a Lebesgue point for ϕ, we get $\operatorname{ess\,inf}_z \left(c(y,z) - \phi(z) \right) \leq c(y,x) - \phi(x)$ and, consequently,

$$\phi^{cc}(x) \geq \operatorname*{ess\,inf}_y \left[c(x,y) - c(y,x) + \phi(x) \right] = \phi(x). \quad \square$$

Lemma 1.4.11. *Define $K(\phi) = \int_\Omega \phi^c d\nu$: then K is concave and upper semicontinuous in ϕ with respect to the $\sigma(L^1, L^\infty)-$convergence.*

Proof. To prove the concavity of K, just notice that

$$(t\phi_1 + (1-t)\phi_0)^c(x) = \operatorname*{ess\,inf}_y tc(x,y) + (1-t)c(x,y) - (t\phi_1 + (1-t)\phi_0)(y)$$
$$\geq t \cdot \operatorname*{ess\,inf}_y c(x,y) - \phi_1(y) + (1-t) \cdot \operatorname*{ess\,inf}_y c(x,y) - \phi_0(y) = t\phi_1^c(x) + (1-t)\phi_0^c(x).$$

For semicontinuity, first notice that, once we have concavity, it is sufficient to prove semicontinuity with respect to the strong convergence. Then we prove that, if $\phi_n \to \phi$ a.e., then for any x it holds $\phi^c(x) \geq \limsup_n \phi_n^c(x)$. In fact, suppose by contradiction

$$\phi^c(x) = \operatorname*{ess\,inf}_y c(x,y) - \phi(y) < \lambda < \limsup_n \phi_n^c(x).$$

Then, let A be the set of points which are Lebesgue points for all the ϕ_n and for ϕ and where pointwise convergence happens. For at least a point $y_0 \in A$ we have $c(x,y_0) - \phi(y_0) < \lambda$ and this implies eventually $c(x,y_0) - \phi_n(y_0) < \lambda$. From this we infer $\phi_n^c(x) \leq \lambda$, which is a contradiction to the assumption. Then take a sequence $\phi_n \to \phi$ in L^1 with $K(\phi) < \limsup_n K(\phi_n)$. Up to subsequences we may suppose $\phi_n \to \phi$ pointwisely a.e. and $\phi_n^c \to \psi$ uniformly (by Ascoli-Arzelà Theorem, as all functions ϕ_n^c have the same modulus of continuity as c). By the pointwise semicontinuity we proved we have $\psi \leq \phi^c$ and in the end we get $\lim_n K(\phi_n) = \int_\Omega \psi d\nu \leq K(\phi)$. $\square$

Lemma 1.4.12. *If $F : X \to \mathbb{R}$ and $G : X' \to \mathbb{R}$ are convex l.s.c. (F is $\sigma(X, X')-$lsc and G is $\sigma(X', X)-$lsc) functionals with*

$$F(u) = \sup_v < u, v > -G(v),$$

then for any $u_0 \in X$ we have

$$\partial F(u_0) = \mathrm{argmax}_v < u_0, v > -G(v).$$

Proof. From our assumption we deduce that $F = G^*$ and $G = F^*$. Then, we can use the well-known relation $v \in \partial F(u_0) \Leftrightarrow u_0 \in \partial F^*(v)$ (see [44], Prop. 5.1, for instance). This means that $v \in \partial F(u_0)$ is equivalent to

$$G(w) \geq G(v)+ < w - v, u_0 > \qquad \text{for any } w \in X',$$

which means that v actually maximizes $< u_0, \cdot > -G$. □

We are now ready to give the alternative proof of Theorem 1.4.8.

Proof. Consider the minimizing probability μ with density $u \in L^1(\Omega)$ and define the vector space $X = span\,(L^\infty(\Omega), \{u\})$, and the space X'

$$X' = \left\{ \xi \in L^1(\Omega) : \int_\Omega |\xi| u \, d\mathcal{L}^d < +\infty \right\},$$

which is in duality with X by means of the product $< v, \xi >= \int_\Omega v\xi \, d\mathcal{L}^d$. Then, we consider the minimization problem for the functional H defined on X by

$$H(v) = \begin{cases} \mathfrak{F}_\nu(v) & \text{if } v \in \mathcal{P}(\Omega); \\ +\infty & \text{otherwise.} \end{cases}$$

It is clear that u minimizes H on X. We will prove

$$\partial H(u) = \left\{ f'(u) + \psi : \psi \text{ maximizes } \int_\Omega \phi d\mu + \int_\Omega \phi^c d\nu \text{ for } \phi \in X' \right\}, \tag{1.4.5}$$

and then consider as an optimality condition $0 \in \partial H(u)$. The subdifferential ∂H of the convex functional H is to be considered in the sense of the duality between X and X'.

To prove (1.4.5) we will use the fact that $H = F + T_p$, where both F and T_p are convex functionals. Here F is defined as the usual functional $u \mapsto \int f(u) d\mathcal{L}^d$: notice that, from the growth assumption on f, we have dom$(F) = X$. Moreover we have, we have $\partial F(u) = \{f'(u)\} \subset X'$ and for any $u_1 \in X$ we have

$$(F(u + \varepsilon(u_1 - u)) - F(u))/\varepsilon \to < u_1 - u, f'(u) > .$$

This can be proven by using the same computations as in Lemma 1.4.1. The functional T_p, on the other hand, is defined as usual on $X \cap \mathcal{P}(\Omega)$ and $+\infty$ elsewhere.

Then we may apply Lemma 1.4.9 to get (1.4.5), provided we prove

$$\partial T_p(u) = \left\{ \psi : \psi \text{ maximizes } \int_\Omega \phi d\mu + \int_\Omega \phi^c d\nu \text{ for } \phi \in X' \right\}. \qquad (1.4.6)$$

By Lemma 1.4.12, applied to the spaces X and X', (1.4.6) is a consequence of the equality $T_p(v) = \sup_\phi < v, \phi > + K(\phi)$ and Lemma 1.4.11.

So far we have proven Formula (1.4.5), and, by minimality of u, we get $0 \in \partial H(u)$, which means $0 = f'(u) + \psi$ for a certain ψ attaining the maximum in the duality formula among all functions of X'. It is necessary to prove that ψ is (or agrees a.e.) actually a Kantorovich potential, so that we get the thesis of Theorem 1.4.8. First consider the double transform ψ^{cc} and remember that it holds (Lemma 1.4.10) $\psi^{cc} \geq \psi$ a.e. (see below). Then, by optimality, necessarily we have $\psi^{cc} = \psi$ a.e. on $\{u > 0\}$, since ψ^{cc} belongs to X'(it is a bounded function) and it would improve the value of the integrals in the duality formula. By $\psi^{cc} \geq \psi$ together with $0 = f'(u) + \psi$ we may infer $\psi = \psi^{cc} \wedge 0$ a.e. which shows that ψ agrees a.e. with an infimum of two $c-$concave function (which is itself a $c-$concave function) and concludes the proof. □

1.5 Whole minimization on bounded and unbounded domains

In this section we want to go through the consequences that Theorem 1.4.8 has in the problem of minimizing the whole $\mathfrak{F}$, when this functional is built by using a term G as in (1.3.3), which forces the measure ν, representing services, to be purely atomic. Two are our goals: trying to have an explicit expression for u in the case of a bounded domain Ω and proving an existence result in the case $\Omega = \mathbb{R}^d$.

Theorem 1.5.1. *Suppose (μ, ν) is optimal for problem (1.1.2). Suppose also that the function g is locally Lipschitz in $]0,1]$: then ν has finitely many atoms and is of the form $\nu = \sum_{i=1}^m a_i \delta_{x_i}$.*

Proof. It is clear that ν is purely atomic, i.e. a countable sum of Dirac masses. We want to show their finiteness. Consider $a = \max a_i$ (such a maximum exists since $\lim_i a_i = 0$ and $a_i > 0$) and let L be the Lipschitz constant of g on $[a, 1]$. Now consider an atom with mass a_i and modify ν by moving its mass onto the atom x_j whose mass a_j equals a, obtaining a new measure ν'. The $G-$part of the functional decreases, while it may happen that the transport part increases. Since we do not change μ the $F-$part remains the same. By optimality of ν we get $T_p(\mu, \nu) + G(\nu) \leq T_p(\mu, \nu') + G(\nu')$ and so

$$g(a_i) - La_i \leq g(a_i) + g(a) - g(a + a_i) \leq T_p(\mu, \nu') - T_p(\mu, \nu) \leq a_i D.$$

This implies

$$\frac{g(a_i)}{a_i} \leq D + L,$$

and, by the assumption on the behavior of g at 0, this gives a lower bound δ on a_i. Since we have proved that every atom of ν has a mass greater than δ, we may conclude that ν has finitely many atoms. □

Now we can use the results from last section.

Theorem 1.5.2. *For any $\nu \in \mathcal{P}(\Omega)$ such that ν is purely atomic and composed by finitely many atoms at the points $x_1, \ldots, x_m$, if μ minimizes $\mathfrak{F}_\nu$ there exist constants c_i such that*

$$u(x) = k\left((c_1 - |x - x_1|^p) \vee \ldots (c_m - |x - x_m|^p) \vee 0\right). \tag{1.5.1}$$

In particular the support of u is the intersection with Ω of a finite union of balls centered around the atoms of ν.

Proof. concerning the Kantorovich potential ψ appearing in Theorem 1.4.8 we know that

$$\begin{aligned} \psi(x) + \psi^c(y) &= |x - y|^p \quad \forall (x, y) \in spt(\pi), \\ \psi(x) + \psi^c(y) &\leq |x - y|^p \quad \forall (x, y) \in \Omega \times \Omega, \end{aligned}$$

where π is an optimal transport plan between μ and ν. Taking into account that ν is purely atomic we obtain, defining $c_i = \psi^c(x_i)$,

$$\begin{aligned} -\psi(x) &= c_i - |x - x_i|^p \quad \mu - a.e.\, x \in \Omega_i, \\ -\psi(x) &\geq c_i - |x - x_i|^p \quad \forall x \in \Omega, \forall i, \end{aligned}$$

where $\Omega_i = T^{-1}(x_i)$ and T is an optimal transport map between μ and ν. Since μ–a.e. point in Ω is transported to a point x_i, we know that $u = 0$ a.e. in the complement of $\bigcup_i \Omega_i$. Since, by $f'(u) = -\psi$, it holds $-\psi(x) \geq 0$, one gets that everywhere in Ω the function $-\psi$ is greater than each of the terms $c_i - |x - x_i|^p$ and 0, while a.e. it holds equality with at least one of them. By changing u on a negligible set, one obtains (1.5.1). The support of μ, consequently, turns out to be composed by the union of the balls $B_i = B(x_i, c_i^{1/p})$ intersected with Ω. □

Theorem 1.5.2 allows us to have an almost explicit formula for the density of μ. Formula (1.5.1) becomes more explicit when the balls B_i are disjoint. We give now a sufficient condition on ν under which this fact occurs.

Lemma 1.5.3. *There exists a positive number $\overline{R}$, depending on the function k, such that all the balls B_i have a radius not exceeding $\overline{R}$. In particular, for any atomic probability ν such that the distance between any two of its atoms is larger than $2\overline{R}$, the balls B_i are disjoint.*

Proof. Set $R_i = c_i^{1/p}$ and notice that

$$1 = \int_\Omega u \geq \int_{B_i} k(c_i - |x - x_i|^p)\, dx = \int_0^{R_i} k(R_i^p - r^p) d\omega_d r^{d-1}\, dr,$$

where the number ω_d stands for the volume of the unit ball in $\mathbb{R}^d$. This inequality gives the required upper bound on R_i, since

$$\int_0^{R_i} k(R_i^p - r^p) d\omega_d r^{d-1}\, dr \geq C \int_0^{R_i - 1} n r^{d-1}\, dr = C(R_i - 1)^d.$$

□

When the balls B_i are disjoint we have $B_i = \Omega_i$ for every i and we get a simple relation between radii and masses corresponding to each atom. The constants c_i can then be found by using $R_i = c_i^{1/p}$. In fact, by imposing the equality of the mass of μ in the ball and of ν in the atom, the radius $R(m)$ corresponding to a mass m satisfies

$$m = \int_0^{R(m)} k(R(m)^p - r^p) d\omega_d r^{d-1} dr. \tag{1.5.2}$$

For instance, if $f(s) = s^2/2$, we have

$$R(m) = \left(\frac{m(d+p)}{\omega_d p}\right)^{1/(d+p)}.$$

The second aim of this section is to obtain an existence result for the problem 1.1.2 when $\Omega = \mathbb{R}^d$. A difference from the bounded case is the fact that we must look for minimization among all pairs of measures in $\mathcal{W}_p(\mathbb{R}^d)$, the $p-$th Wasserstein metric space (see Section 0.2), rather than in $\mathcal{P}(\mathbb{R}^d)$.

We start by some simple results about the minimization problem for $\mathfrak{F}_\nu$.

Lemma 1.5.4. *For every fixed $\nu \in \mathcal{P}(\mathbb{R}^d)$ there exist a unique minimizer μ for $\mathfrak{F}_\nu$: it belongs to $\mathcal{W}_p(\mathbb{R}^d)$ if and only if $\nu \in \mathcal{W}_p(\mathbb{R}^d)$, and if ν does not belong to this space the functional $\mathfrak{F}_\nu$ is infinite on the whole $\mathcal{W}_p(\mathbb{R}^d)$. Moreover, if ν is compactly supported, the same happens for μ.*

Proof. The existence of μ comes from the direct method of the calculus of variations and the fact that if $(T_p(\mu_h,\nu))_h$ is bounded, then $(\mu_h)_h$ is tight. Uniqueness follows from the strict convexity of f. The behavior of the functional with respect to the space $\mathcal{W}_p(\mathbb{R}^d)$ is trivial. Finally, the last assertion can be proved by contradiction, supposing $\mu(B(0,R)^c) > 0$ for every $R < +\infty$ and replacing μ by

$$\mu_R = 1_{B_R} \cdot \mu + \frac{\mu(B_R^c)}{|B_r|} I_{B_r} \cdot \mathcal{L}^d,$$

where $B(0,r)$ is a ball containing the support of ν. By optimality, we should have

$$T_p(\mu_R,\nu) + F(\mu_R) \geq T_p(\mu,\nu) + F(\mu), \tag{1.5.3}$$

but we have

$$T_p(\mu_R,\nu) - T_p(\mu,\nu) \leq -((R-r)^p - (2r)^p)\mu(B_R^c), \tag{1.5.4}$$

$$F(\mu_R) - F(\mu) \leq \int_{B_r} \left[f\left(u + \frac{\mu(B_R^c)}{|B_r|}\right) - f(u) \right] d\mathcal{L}^d. \tag{1.5.5}$$

By summing up (1.5.4) and (1.5.5), dividing by $\mu(B_R^c)$ and taking into account (1.5.3), we get

$$-((R-r)^p-(2r)^p)+\frac{1}{\mu(B_R^c)}\int_{B_r} \left[f\left(u + \frac{\mu(B_R^c)}{|B_r|}\right) - f(u) \right] d\mathcal{L}^d \geq 0. \tag{1.5.6}$$

Yet, by passing to the limit as $R \to +\infty$ and $\mu(B_R^c) \to 0$, the first term in (1.5.6) tends to $-\infty$, while the second is decreasing as $R \to +\infty$. This last one tends to $\int_{B_r} f'(u) d\mathcal{L}^d$, provided it is finite for at least a value of R (which ensures the finiteness of the limit as well). To conclude it is sufficient to prove that

$$\int_{B_r} \left[f\left(u + \frac{\mu(B_R^c)}{|B_r|}\right) - f(u) \right] d\mathcal{L}^d < +\infty.$$

This is quite easy in the case $f(z) = Az^q$ with $q > 1$, while for general f the assertion comes from the fact that u is continuous on $\overline{B_r}$, hence bounded. The continuity of u may be obtained by localizing the result of Theorem 1.4.8: just consider $\mu' = 1_{B_r}/\mu(B_r)\cdot\mu$ and $\nu' = T_\sharp\mu'$ for an optimal transport map T between μ and ν and correspondingly rescale the function f. It is clear that μ' minimizes a new functional
$tildeF_{\nu'}$ in the new domain $\Omega' = \overline{B_r}$. Then we may apply Theorem 1.4.8 and get the continuity of its density, which ensures the continuity of u on $\overline{B_r}$. □

To go through our proof we need to manage minimizing sequences, in the sense of Lemma below.

Lemma 1.5.5. *It is possible to choose a minimizing sequence* $((\mu_h, \nu_h))_h$ *in* $\mathcal{W}_p(\mathbb{R}^d) \times \mathcal{W}_p(\mathbb{R}^d)$ *such that for every* h *the measure* ν_h *is finitely supported, and the density of* μ_h *is given by (1.5.1), with disjoint balls centerd at the atoms of* ν_h*.*

Proof. First we start from an arbitrary minimizing sequence $((\mu'_h, \nu'_h))_h$. Then we approximate each ν'_h in $\mathcal{W}_p$ by a finite support measure ν''_h. To do this we truncate the sequence of its atoms and move the mass in excess to the origin. In this way, we have $G(\nu''_h) \leq G(\nu'_h)$, by the subadditivity of g, while the value of the transport term increases of an arbitrary small quantity. Consequently, $((\mu'_h, \nu''_h))_h$ is still a minimizing sequence. Then, we replace μ'_h by μ''_h, chosen in such a way that it minimizes $\mathfrak{F}_{\nu''_h}$. By Lemma 1.5.4, each μ''_h has a compact support. Then, we translate every atom of each ν''_h together with its own set Ω_i, to some disjoint sets Ω^*_i. In this way we get new measures μ'''_h and ν'''_h. The value of the functional in this step has not changed. We may choose to place the atoms of each ν'''_h so far from each other so that each distance between atoms is at least $2\overline{R}$. Then we minimize again in μ, getting a new sequence of pairs $((\mu''''_h, \nu'''_h))_h$ and we set $\nu_h = \nu'''_h$ and $\mu_h = \mu''''_h$. Thanks to Theorem 1.5.2 and Lemma 1.5.3 the requirements of the thesis are fulfilled. □

It is clear now that, if one can obtain a uniform estimate on the number of atoms of the measures ν_h, the existence problem is easily solved: in fact we already know that each ball belonging to the support of μ_h is centered at an atom of ν_h and has a radius not larger than $\overline{R}$. Provided we are able to prove an estimate like $\sharp\,\{\text{atoms of } \nu_h\} \leq N$, it would be sufficient to act by translation on the atoms and their corresponding balls, obtaining a new minimizing sequence (the value of $\mathfrak{F}$ does not change) with supports all contained in a same bounded set (for instance, the ball $B_{N\overline{R}}$).

We now try to give sufficient conditions in order to find minimizing sequences where the number of atoms stays bounded. Notice that, on sequences of the form given by Lemma 1.5.5, the functional $\mathfrak{F}$ has the expression

$$\mathfrak{F}(\mu_h, \nu_h) = \sum_{i=1}^{k(h)} E(m_{i,h}), \quad \text{if } \nu_h = \sum_{i=1}^{k(h)} m_{i,h}\delta_{x_{i,h}}, \tag{1.5.7}$$

where the quantity $E(m)$ is the total contribute given by an atom with mass m to the functional. We may compute:

$$E(m) = g(m) + \int_0^{R(m)} [f(k(R(m)^p - r^p)) + k(R(m)^p - r^p)r^p] \, d\omega_d r^{d-1} dr, \tag{1.5.8}$$

taking into account the particular form of the density in the ball.

Theorem 1.5.6. *Let us suppose* $f \in C^2((0, +\infty))$, $g \in C^2((0, 1]) \cap C^0([0, 1])$, *in addition to all previous assumptions. Then the minimization problem for* $\mathfrak{F}$ *in* $\mathcal{W}_p(\mathbb{R}^d) \times \mathcal{W}_p(\mathbb{R}^d)$ *has a solution, provided*

$$\limsup_{R \to 0^+} g''\left(\int_0^R k(R^p - r^p) d\omega_d r^{d-1} dr\right) \int_0^R k'(R^p - r^p) d\omega_d r^{d-1} dr < -1.$$

Proof. According to what previously proven, it is sufficient to produce a minimizing sequence of the form of Lemma 1.5.5, with a bounded number of atoms. We claim that it is enough to prove that the function E is subadditive on an interval $[0, m_0]$. In fact, once proven it, we start from a sequence $((\mu_h, \nu_h))_h$ built as in Lemma 1.5.5 and use the characterization of $\mathfrak{F}$ given in (1.5.7). Then we modify our sequence by replacing in each ν_h any pair of atoms of mass less than $m_0/2$ by a single atom with the sum of the masses. We keep atoms far away from each other, in order to use (1.5.7). We may perform such a replacement as far as we find more than one atom whose mass is less or equal than $m_0/2$. At the end we get a new pair $((\mu'_h, \nu'_h))_h$ where the number of atoms of ν'_h is less than $N = 1 + \lfloor 2/m_0 \rfloor$. The value of the functional $\mathfrak{F}$ has not increased, thanks to the subadditivity of E on $[0, m_0]$.

Taking into account that $E(0) = 0$ and that concave functions vanishing at 0 are subadditive, we look at concavity properties of the function E in an interval $[0, m_0]$. It is sufficient to compute the second derivative of E and find it negative in a neighborhood of the origin.

By means of the explicit formula (1.5.8), and taking into account also (1.5.2), setting $E(m) = g(m) + K(R(m))$, we start by computing dK/dr.

Using the facts that $f' \circ k = id$ and that $k(0) = 0$, we can obtain the formula

$$\frac{dK(R(m))}{dm}(m) = R(m)^p.$$

From another derivation and some standard computation we finally obtain

$$E''(m) = g''(m) + \frac{1}{\int_0^{R(m)} k'(R(m)^p - r^p) d\omega_d r^{d-1} dr}.$$

The assumption of this Theorem ensures that such a quantity is negative for small m, and so the proof is achieved. $\square$

Remark 1.5.7. Notice that, when the functions f and g are of the form $f(t) = at^q$, $q > 1$, $g(t) = bt^\alpha$, $\alpha < 1$, with a and b positive constants, we have

$$g''\left(\int_0^R k(R^p - r^p) d\omega_d r^{d-1} dr\right) \leq -CR^{(d+\frac{p}{q-1})(\alpha-2)};$$
$$\int_0^R k'(R^p - r^p) d\omega_d r^{d-1} dr \geq CR^{d+p\frac{2-q}{q-1}},$$

and so the lim sup in Theorem 1.5.6 may be estimated from above by

$$\lim_{R \to 0^+} -CR^{\frac{p}{q-1}(\alpha-q)+d(\alpha-1)} = -\infty.$$

Consequently the assumption in Theorem 1.5.6 is always verified when f and g are power functions.

Remark 1.5.8. From the proof of the existence Theorem it is clear that there exists a minimizing pair $(\mu, \nu) \in \mathcal{W}_p(\mathbb{R}^d) \times \mathcal{W}_p(\mathbb{R}^d)$ where ν has finitely many atoms and μ is supported in a finite, disjoint union of balls centered at the atoms of ν and contained in a bounded domain Ω_0, with a density given by Theorem 1.5.2. The same happens if we look for the minimizers in a bounded domain Ω, provided Ω is large enough to contain Ω_0, and hence a solution to the problem in $\mathbb{R}^d$. For instance all the open sets containing N balls of radius $\overline{R}$ admit a minimizing solution supported in disjoint balls.

We conclude by stressing the fact that, in order to solve the problem in $\mathbb{R}^d$, we have only to look at the function E and find out the number of atoms and their respective masses $(m_i)_{i=1...k}$ The problem to solve is then

$$\min\left\{\sum_{i=1}^k E(m_i) \,:\, k \in \mathbb{N}, \sum_{i=1}^k m_i = 1\right\}. \tag{1.5.9}$$

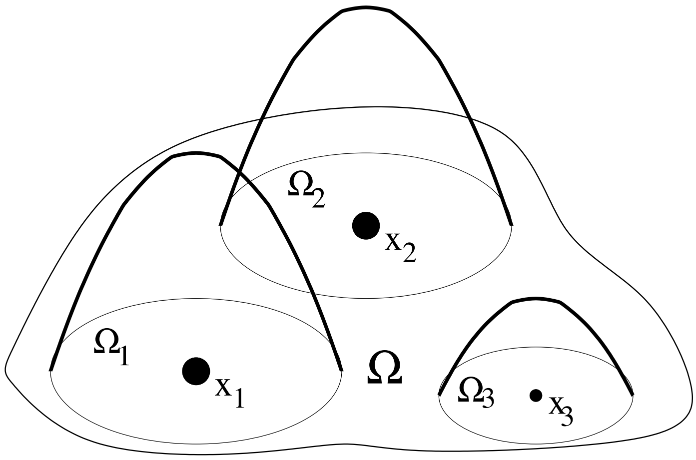

Figure 1.1: Sketch of the solution in a large domain Ω or in the whole space

Typically, for instance when f and g are power functions, the function E involved in (1.5.9) is a concave-convex function, as sketched in picture 2.1. Due to such a concave-convex behavior, it is not in general clear whether the values of the numbers m_i solving (1.5.9) and representing sub-cities' sizes are all equal or may be different.

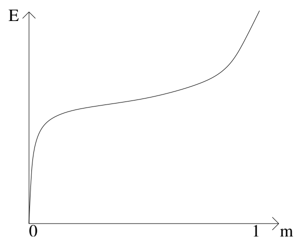

Figure 1.2: Typical behavior of E

1.6 Comments on the model and its results

The model we presented takes into consideration only the optimization of a total welfare parameter of the city, disregarding the effects on each single citizen. In particular no equilibrium condition is considered. This may appear as a fault in the model, since the personal welfare of the citizens (depending on the population density of their zone and on the cost for moving from home to services) could be non-constant. As a consequence, non-stable optimal solutions may occur, where some citizens would prefer to move elsewhere in the city in order to get better conditions. However, this is not necessarily the case, since our model also disregards prices of land and houses in the city, since they do not affect the total wealth of the area. It may turn out that, by a proper, market-determined, choice of prices, welfare differences could be compensated and equilibrium recovered. This fact turns out to be a great difference from the model in [35], both for the importance which is given there to the variable represented by the price of land and for the fact that Carlier and Ekeland exactly look for an equilibrium solution instead of an optimal one.

As we saw, both in the case $\Omega = \mathbb{R}^d$ and Ω bounded, optimal choices for μ and ν are given by the formation of a certain number of sub-cities, which are circular areas with a pole of services in the center (an atom for the measure ν) around which the population is distributed with a decreasing radial density.

Since we have only considered a very simplified model, our goal is neither to suggest a realistic way to design the ideal city, nor to describe in a variational way the formation of existing cities. Anyway, from the analysis of our optimality results (and in particular from the sub-cities phenomena we referred to), we can infer some conclusions.

- This model is not a proper choice to describe the shape of a single existing city, since the delocalization of services we find in an optimal solution does not reflect what reality suggests (in fact we find finitely many disjoint, independent, sub-cities with services only in the center).

- This model is likely to be more realistic on a larger scale, when Ω represents a large urban area composed by several cities: in this case every atom of the optimal ν stands for the center of one of them and includes a complex system of services, located downtown, whose complexity cannot be seen in this scale.

- In this model the concentrated measure ν gives a good representation of the areas where services are offered to citizens and not of areas where commodities are produced (factories), due to the assumption that no land is actually occupied by the service poles (since ν is atomic).

- We conclude by stressing that the same model may be applied as a first simplified approach to other kinds of problems, where we have to choose in some efficient way the distributions of two different parameters, being the first spread and the second concentrated, keeping them as close as possible to each other in some mass transportation sense. This is what will be proposed in a general framework in Chapter 3.

Chapter 2

An urban planning model with traffic congestion

In this chapter we want to introduce another model for the urban planning of residence and working areas, where the transportation costs take into account the effects of traffic congestion. Formalizing traffic congestion is a very interesting matter which is suggested in [11]. The model in [11] has as a starting point the work of Beckmann [10] who, in the early 50's, introduced the so-called *continuous transportation model* in urban economics, leading to a *minimal flow* like problem (see Proposition 0.1.8). Here we will develop a very simplified model whose goal is only to define a transportation cost, sort of a distance between probability measures, which takes care of the idea that "if several people are supposed to pass through a common point, then passing through it will be more costly". Then this cost will be used in a variational problem over probability measures which shares the same structure of the one studied in Chapter 1.

As in [28] and in Chapter 1, given a urban area Ω (a subset of $\mathbb{R}^2$ in the applications), we look for the distribution of residents (or consumption), denoted by μ, and the distribution of services (or production), denoted by ν, so as to minimize a cost involving three terms: an overall transportation term for moving customers to services, a term penalizing dispersion of services and a term penalizing concentration of residents. To take into account (in a special case) congestion effects, we are lead to consider as a transportation cost the squared norm of $\mu-\nu$ in the dual space of some subspace of $H^1(\Omega)$ (see Section 2.1 for details). In dimension 2, this in particular prevents the presence of atoms of $\mu - \nu$. Hence, contrary to [28] and Chapter 1, where a term forcing the distribution of services ν to be concentrated in at most countably many locations was considered, we rather consider an interaction term of the form:

$$H(\nu) := \int_{\Omega \times \Omega} V \, d(\nu \otimes \nu),$$

where $V(x, y)$ is, for instance, an increasing function of $|x-y|$. Such a term, studied in [58] as well, has already been proposed in [66] as a concentration term useful in similar urban planning problems.

2.1 Traffic congestion

In this section, we formally describe how we model congestion effects in the transportation cost functional. Our analysis builds upon the *continuous transportation model* of Beckmann (see [10], [11]).

We are given an urban area Ω, which is an open bounded connected subset of $\mathbb{R}^2$ satisfying some smoothness assumptions that will be made precise later, and we denote by μ and ν the respective distributions of residents and services in the city. As a normalization, we may assume that μ and ν are probability measures on Ω and that μ (respectively ν) also gives the distribution of consumption (respectively of production) so that the signed measure $\mu - \nu$ represents the local measure of excess demand. Following [10], we assume that the consumers' traffic is given by a traffic flow field, i.e. a vector field $Y : \Omega \to \mathbb{R}^2$ whose direction indicates the consumers' travel direction and whose modulus $|Y|$ is the intensity of traffic.

The relationship between the excess demand and the traffic flow is obtained from an equilibrium condition as follows. There is equilibrium in a subregion $K \subset \Omega$ if the outflow of consumers equals the excess demand of K:

$$\int_{\partial K} Y \cdot n \, d\mathcal{H}^{n-1} = (\mu - \nu)(K).$$

Since the previous has to hold for arbitrary K, this formally yields:

$$\nabla \cdot Y = \mu - \nu. \tag{2.1.1}$$

It is also assumed that the urban area is isolated, i.e. no traffic flow should cross the boundary of the city, hence:

$$Y \cdot n = 0 \text{ on } \partial\Omega. \tag{2.1.2}$$

If the transportation cost per consumer is assumed to be uniform, then one may define the transportation cost between μ and ν as the value of the *minimal flow* problem:

$$\inf\left\{\int_\Omega |Y(x)|dx \,:\, Y \text{ satisfies (2.1.1)-(2.1.2)}\right\}. \tag{2.1.3}$$

Of course, one generally has to look for generalized (i.e. vector-valued measures) solutions of the previous problem. As we pointed out in Proposition 0.1.8, this minimal value coincides with the 1-Wasserstein distance between μ and ν. Let us also mention that this problem (or its extension to measures) is tightly connected to the notion of transport density in the Monge-Kantorovich optimal transportation problem (where cost = euclidean distance): we refer to De Pascale and Pratelli [40], for details and very interesting regularity results for transport density in the Monge-Kantorovich problem. These results may give the possibility, under L^p assumption on μ and ν, to extend the problem to the setting of vector measures and then, by regularity, to state that the same problem has also solution among L^p vector fields.

The minimization problem in (2.1.3) may obviously be modified if we want to take into account possible geographical conditions into the following

$$\inf\left\{\int_\Omega k(x)|Y(x)|dx \,:\, Y \text{ satisfies (2.1.1)-(2.1.2)}\right\}. \tag{2.1.4}$$

Here $k(x)$ is a term whose meaning is the transportation cost per consumer at a point x. In this case Problem (2.1.4) is still linked to a transport problem, but for a cost c given by

$$c(x,y) = \inf\left\{\int_0^1 k(\gamma(t))|\gamma'(t)|\,dt \,:\, \gamma \text{ is Lipschitz and } \gamma(0)=x,\, \gamma(1)=y\right\}.$$

Now, in order to take into account congestion effects, it is more realistic to assume that the transportation cost per consumer at a point x depends on the intensity of traffic at x itself. Let $g : [0,+\infty] \to [0,+\infty]$ be a given nondecreasing function, and assume that if the traffic flow is Y then we would like to set $k(x) = g(|Y(x)|)$. It is natural, at this point, to define the transportation cost between μ and ν as:

$$C_g(\mu,\nu) := \inf\left\{\int_\Omega g(|Y(x)|)|Y(x)|dx \,:\, Y \text{ satisfies (2.1.1)-(2.1.2)}\right\}.$$

For the sake of simplicity, we will assume, from now on, that $g(t) = t$ for all $t \geq 0$, and define the cost:

$$C(\mu,\nu) := \inf\left\{\int_\Omega |Y(x)|^2dx \,:\, Y \text{ satisfies (2.1.1)-(2.1.2)}\right\}. \tag{2.1.5}$$

where (2.1.1)-(2.1.2) are understood in the weak sense, hence read as:

$$\int_\Omega Y \cdot \nabla\phi = \int_\Omega \phi \, d(\mu - \nu), \text{ for all } \phi \in C^1(\Omega).$$

Let us define:

$$X := \left\{ \phi \in H^1(\Omega) : \int_\Omega \phi = 0 \right\}.$$

X is a Hilbert space, when equipped with the following inner product and norm:

$$\langle \phi, \psi \rangle_X := \int_\Omega \nabla\phi \cdot \nabla\psi \quad , \quad \|\phi\|_X^2 := \langle \phi, \phi \rangle_X .$$

As usual, we shall identify X and its dual X' by Riesz's isomorphism: for every $f \in X'$, there exists, unique, $\phi \in X$ such that:

$$\langle \phi, \psi \rangle_X = f(\psi) \text{ for all } \psi \in X. \tag{2.1.6}$$

This implies:

$$\|f\|_{X'} = \|\phi\|_X .$$

We shall also write (2.1.6) in the form:

$$\left\{ \begin{array}{ll} -\Delta\phi = f & \text{in } \mathring{\Omega}, \\ \frac{\partial\phi}{\partial n} = 0 & \text{on } \partial\Omega, \phi \in X. \end{array} \right. \tag{2.1.7}$$

With those definitions in mind, it is easy to check that our cost functional given by (2.1.5) may also be written as:

$$C(\mu, \nu) = \left\{ \begin{array}{ll} \|\mu - \nu\|_{X'}^2 & \text{if } \mu - \nu \in X', \\ +\infty & \text{otherwise.} \end{array} \right. \tag{2.1.8}$$

Equivalently, we have:

$$C(\mu, \nu)^{\frac{1}{2}} = \sup \left\{ \int_\Omega \phi \, d(\mu - \nu) : \phi \in C^1(\Omega), \int_\Omega \phi = 0, \|\phi\|_X \leq 1 \right\}. \tag{2.1.9}$$

2.2 The minimization problem

In what follows, Ω will be a domain of $\mathbb{R}^2$ (even if most of the results are actually valid in $\mathbb{R}^d$), V a nonnegative l.s.c. function on $\mathbb{R}^2$ and $\mathcal{L}^2$ will denote the 2-dimensional Lebesgue measure on Ω. We consider the variational problem:

$$\inf\{\mathfrak{F}(\mu,\nu) = C(\mu,\nu) + G(\mu) + H(\nu) \colon \mu,\nu \text{ probabilities on } \Omega\} \tag{2.2.1}$$

where:

$$C(\mu,\nu) = \begin{cases} \|\mu-\nu\|^2_{X'} & \text{if } \mu-\nu \in X', \\ +\infty & \text{otherwise;} \end{cases}$$

$$G(\mu) = \begin{cases} \int_\Omega u^2 & \text{if } \mu = u\cdot\mathcal{L}^2,\ u \in L^2(\Omega), \\ +\infty & \text{otherwise;} \end{cases}$$

and

$$H(\nu) := \int_{\Omega\times\Omega} V(x,y)(\nu\otimes\nu)(dx,dy).$$

Theorem 2.2.1. *Assume that V is bounded from below, l.s.c. and there exist probability measures μ_0 and ν_0 on Ω such that $\mathfrak{F}(\mu_0,\nu_0) < +\infty$. Then the minimization problem (2.2.1) has at least one solution.*

Proof. First it is clear that the infimum of (2.2.1) is finite. Due to the weak compactness of the space of probability measures on Ω, the existence will directly follow from the weak lower semicontinuity of $\mathfrak{F}$. The weak lower semicontinuity of G is clear, that of the interaction functional H is easy to establish and that of C follows from formula (2.1.9). □

2.3 Minimization with respect to μ

In this paragraph, we consider for a fixed probability ν (with $\nu \in X'$) the minimization of $\mathfrak{F}$ with respect to μ:

$$\inf\{C(\mu,\nu) + G(\mu) : \mu \text{ probability measure on } \Omega\ \} \tag{2.3.1}$$

Proposition 2.3.1. *Given $\nu \in \mathcal{P}(\Omega)\cap X'$, then (2.3.1) admits a unique solution μ which is characterized by $\mu = \phi\cdot\mathcal{L}^2$, where $\phi \in H^1(\Omega)$ is the solution of:*

$$\begin{cases} -\Delta\phi + \phi = \nu & \text{in } \mathring{\Omega}, \\ \frac{\partial\phi}{\partial n} = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.3.2}$$

Proof. It is obvious that (2.3.1) admits a unique solution $\mu = u \cdot \mathcal{L}^2$ with $u \in L^2(\Omega)$. Let p be a probability measure on Ω with $p = v \cdot \mathcal{L}^2$ and $v \in L^2(\Omega)$ (which implies at once $p \in X'$). For $\varepsilon \in (0,1)$, one has:

$$0 \le C(\mu + \varepsilon(p-\mu), \nu) + G(\mu + \varepsilon(p-\mu)) - C(\mu,\nu) - G(\mu). \tag{2.3.3}$$

Let $\psi \in X$ be the solution of:

$$\begin{cases} \Delta\psi = \mu - \nu & \text{in } \overset{\circ}{\Omega}, \\ \frac{\partial \psi}{\partial n} = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.3.4}$$

Similarly, let $\chi \in X$ be the solution of:

$$\begin{cases} \Delta\chi = p - \mu & \text{in } \overset{\circ}{\Omega}, \\ \frac{\partial \chi}{\partial n} = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.3.5}$$

We then have:

$$\begin{aligned} C(\mu + \varepsilon(p-\mu), \nu) =& \|\mu - \nu + \varepsilon(p-\mu)\|_{X'}^2 = \|\psi + \varepsilon\chi\|_X^2 \\ =& C(\mu,\nu) + 2\varepsilon \int_\Omega \nabla\psi \cdot \nabla\chi + \varepsilon^2 \int_\Omega |\nabla\chi|^2 \\ =& C(\mu,\nu) - 2\varepsilon \int_\Omega \psi(v-u) + \varepsilon^2 \|p-\mu\|_{X'}^2 \end{aligned} \tag{2.3.6}$$

Similarly

$$G(\mu + \varepsilon(p-\mu)) = G(\mu) + 2\varepsilon \int_\Omega u(v-u) + \varepsilon^2 \int_\Omega (v-u)^2. \tag{2.3.7}$$

Replacing (2.3.6) and (2.3.7) in (2.3.3), dividing by ε and letting $\varepsilon \to 0^+$, yields:

$$\int_\Omega (v-u)(u-\psi) \ge 0. \tag{2.3.8}$$

Since $p = v \cdot \mathcal{L}^2$ is an arbitrary probability measure (with $v \in L^2(\Omega)$), (2.3.8) can also be written as:

$$\text{there exists } m \in \mathbb{R} \text{ such that: } u - \psi \ge m,\ u - \psi = m\ \mu\text{-a.e.}. \tag{2.3.9}$$

Since $u \ge 0$, this also implies $u = (\psi + m) \vee 0$.

Define then $\phi := (\psi + m)$. Let us prove that $\phi \ge 0$ so that we get $u = \phi$. First notice that $\phi \ge 0$ μ-a.e. (using (2.3.9)); then set $\phi_- := -(\phi \wedge 0)$ and get

$$\int_\Omega \nabla\phi \cdot \nabla\phi_- = -\int_{\{\phi<0\}} |\nabla\phi|^2 = \int_\Omega \phi_- d(\nu - \mu) = \int_\Omega \phi_- d\nu \ge 0.$$

This proves $\phi_- = 0$ and hence $u = \phi$; replacing in (2.3.4), we get that ϕ is the solution of (2.3.2). □

2.4 Optimality conditions

Thanks to proposition 2.3.1, we can reformulate the problem (2.2.1) in terms of ν only. More precisely, define for every probability measure ν on Ω:

$$J(\nu) := \inf \{ F(\mu, \nu) : \mu \text{ probability measure on } \Omega \} .$$

By proposition 2.3.1, we have:

$$J(\nu) = \begin{cases} \int_\Omega (|\nabla \phi|^2 + \phi^2) + H(\nu) & \text{with } \phi \text{ the solution of (2.3.2) if } \nu \in X', \\ +\infty & \text{otherwise.} \end{cases}$$

Identifying $H^1(\Omega)$ and its dual $H^1(\Omega)'$ via Riesz's isomorphism for its usual Hilbertian structure:

$$\langle \phi, \psi \rangle_{H^1(\Omega)} := \int_\Omega (\nabla \phi \cdot \nabla \psi + \phi \psi) ,$$

$$\|\phi\|^2_{H^1(\Omega)} := \langle \phi, \phi \rangle_{H^1(\Omega)} ,$$

we may also rewrite J as:

$$J(\nu) = \begin{cases} \|\nu\|^2_{H^1(\Omega)'} + H(\nu) & \text{if } \nu \in H^1(\Omega)', \\ +\infty & \text{otherwise.} \end{cases}$$

Finally, the reformulation of (2.2.1) reads as:

$$\inf \{ J(\nu) : \nu \text{ probability measure on } \Omega \} . \tag{2.4.1}$$

In what follows, for every $\nu \in H^1(\Omega)'$, we will say that $\phi \in H^1(\Omega)$ is the *potential* of ν if:

$$\langle \phi, \psi \rangle_{H^1(\Omega)} = \nu(\psi), \text{ for all } \psi \in H^1(\Omega). \tag{2.4.2}$$

Put differently, the potential of ν is the weak solution of:

$$\begin{cases} -\Delta \phi + \phi = \nu & \text{in } \overset{\circ}{\Omega}, \\ \frac{\partial \phi}{\partial n} = 0 & \text{on } \partial \Omega. \end{cases}$$

Let us also remark that if, in addition, ν is a probability measure on Ω and ϕ its potential, then $\phi \cdot \mathcal{L}^2$ is a probability measure on Ω as well.

Let us denote by $\mathcal{C}$ the set of probability measures belonging to the domain of J:

$$\mathcal{C} = \mathcal{P}(\Omega) \cap H^1(\Omega)' = \left\{ \nu \in H^1(\Omega)' : \nu \geq 0 \text{ in } H^1(\Omega)', \ < \nu \, , 1 >= 1 \right\} . \tag{2.4.3}$$

In general, the interaction functional H is not convex. However, in the *small* case, i.e. when either V or Ω is small (in a sense quantified below) then, due to the term $\|.\|^2_{H^1(\Omega)'}$, the quadratic functional J is in fact strictly convex.

Assume that $V \in C^2(\Omega \times \Omega, \mathbb{R})$ and define:

$$c_{\Omega,V} := \left(\int_\Omega \left(\|V(x,.)\|^2_{H^1(\Omega)} + \|\partial_{x_1} V(x,.)\|^2_{H^1(\Omega)} + \|\partial_{x_2} V(x,.)\|^2_{H^1(\Omega)} \right) dx \right)^{\frac{1}{2}} \tag{2.4.4}$$

Proposition 2.4.1. *Assume that $V \in C^2(\Omega \times \Omega, \mathbb{R})$ and let $c_{\Omega,V}$ be defined by (2.4.4). If $c_{\Omega,V} < 1$, then J is a strictly convex functional on $\mathcal{C}$; (2.4.1) then admits a unique solution.*

Proof. Given $\nu \in \mathcal{C}$, let us define:

$$T_\nu(x) :=< \nu\, ,\, V(x,.) >= \int_\Omega V(x,y)\nu(dy).$$

Since $T_\nu \in H^1(\Omega)$, we have, on the one hand:

$$|H(\nu)| = |\nu(T_\nu)| \leq \|\nu\|_{H^1(\Omega)'} \|T_\nu\|_{H^1(\Omega)}. \tag{2.4.5}$$

On the other hand:

$$T_\nu(x)^2 + |\nabla T_\nu(x)|^2 = \left(\int_\Omega V(x,y)\nu(dy) \right)^2 + \left| \int_\Omega \nabla_x V(x,y)\nu(dy) \right|^2$$
$$\leq \|\nu\|^2_{H^1(\Omega)'} \left(\|V(x,.)\|^2_{H^1(\Omega)} + \|\partial_{x_1} V(x,.)\|^2_{H^1(\Omega)} + \|\partial_{x_2} V(x,.)\|^2_{H^1(\Omega)} \right)$$

Integrating the previous inequality and using (2.4.5), we then get:

$$|H(\nu)| \leq c_{\Omega,V} \|\nu\|^2_{H^1(\Omega)'},$$

so that:

$$J(\nu) \geq (1 - c_{\Omega,V})\, \|\nu\|^2_{H^1(\Omega)'}$$

and the claim of the proposition easily follows using the fact that J is quadratic. □

Let V^s denote the symmetric part of V:

$$V^s(x,y) := \frac{1}{2}(V(x,y) + V(y,x)). \tag{2.4.6}$$

The first-order optimality conditions for (2.4.1) are given by the following result:

Proposition 2.4.2. *Assume that $V \in C^2(\Omega \times \Omega, \mathbb{R})$. Given $\nu \in \mathcal{C}$, let ϕ be the potential of ν and let T^s_ν be defined, for all $x \in \Omega$, by:*

$$T^s_\nu(x) := < \nu \, , \, V^s(x,.) > = \int_\Omega V^s(x,y)\nu(dy).$$

If ν is a solution of (2.4.1), then there exists a constant m such that:

$$\phi + T^s_\nu \geq m, \; \phi + T^s_\nu = m \;\; \nu\text{-a.e.}. \tag{2.4.7}$$

Proof. Let $p \in \mathcal{C}$, and $\chi \in H^1(\Omega)$ be the potential of $p - \nu$. Let $\varepsilon \in (0,1)$; since ν solves (2.4.1), we have:

$$0 \leq J(\nu + \varepsilon(p - \nu)) - J(\nu). \tag{2.4.8}$$

We also have:

$$\|\nu + \varepsilon(p-\nu))\|^2_{H^1(\Omega)'} = \|\phi + \varepsilon\chi\|^2_{H^1(\Omega)} = J(\nu) + 2\varepsilon \int_\Omega \phi d(p-\nu) + \varepsilon^2 \|\chi\|^2_{H^1(\Omega)'}$$

Similarly:

$$H(\nu + \varepsilon(p - \nu)) = H(\nu) + 2\varepsilon \int_\Omega T^s_\nu d(p - \nu) + \varepsilon^2 \int_{\Omega\times\Omega} V d((p - \nu) \otimes (p - \nu)).$$

Replacing in (2.4.8), dividing by ε and letting $\varepsilon \to 0^+$ yields:

$$\int_\Omega (\phi + T^s_\nu)\, d(p - \nu) \geq 0.$$

Since $p \in \mathcal{C}$ is arbitrary in the previous inequality, setting:

$$m := \int_\Omega (\phi + T^s_\nu) d\nu$$

and using the fact that $\phi + T^s_\nu \in H^1(\Omega)$ we then have (in the $H^1(\Omega)$ sense):

$$\phi + T^s_\nu \geq m, \; \phi + T^s_\nu = m \;\; \nu\text{-a.e.}.$$

□

Remark 2.4.3. Let us remark that, thanks to proposition 2.4.1, if $c_{\Omega,V} < 1$, Problem (2.4.1) being strictly convex, condition (2.4.7) is in fact sufficient and fully characterizes the minimizer. This fact will be used several times in the examples of section 8. Secondly, it should be noticed that in (2.4.7) ν appears only indirectly through its potential and T^s_ν.

2.5 Regularity via approximation

The aim of this section is to get some regularity results on the optimal measure ν by approximating the minimization problem and then looking for some properties of minimizers passing to the limit. Our main tools will be the Wasserstein distance from optimal transport theory and elliptic regularity.

The results we need on the distance W_2 are in Sections 0.1 and 0.2. Concerning elliptic regularity, we will deal with the case of Neumann conditions. Precisely, we will use the following.

Proposition 2.5.1. *Consider the elliptic equation* (2.3.2)*, which is always endowed with a unique solution for every $\nu \in X'$. Then it holds:*

- *if Ω is an open set with C^2 boundary and $\nu \in L^p(\Omega)$ then $\phi \in W^{2,p}(\Omega)$;*
- *if Ω is an open set with $C^{2,\alpha}$ boundary and $\nu \in C^{0,\alpha}(\Omega)$ then $\phi \in C^{2,\alpha}(\Omega)$.*

We refer to [1] for both implications. For the Hölder theory we can refer also to [47], whose results in chapter 6, section 7, have to be adapted, while for the L^p theory in the case $p = 2$ the ninth chapter in [26] can be seen as well. From now on, we will call *regular* those open sets whose boundary is $C^{2,\alpha}$ for at least a positive value of α.

In our approximation, we want to retrieve information on all minimizers of our problem (in general, when J is not convex they could be not unique), and so we define some functionals J_ε for every choice of $\overline{\nu} \in \operatorname{argmin} J$.

We set, for small $\varepsilon > 0$,

$$J_\varepsilon(\nu) = J(\nu) + \varepsilon W_2^2(\nu, \nu_\varepsilon) + \delta_\varepsilon \|\nu\|^2_{L^2(\Omega)},$$

where $(\nu_\varepsilon)_\varepsilon$ is a sequence of measures which are absolutely continuous with a strictly positive density, approximating $\overline{\nu}$ in the W_2 distance, and δ_ε is a small parameter depending on ε to be properly chosen.

Since the semicontinuity of the terms we have added with respect to weak* topology is straightforward, we get the existence of at least a minimizer $\overline{\nu}_\varepsilon$ for each functional J_ε. We have the following result, which is nothing but an ad hoc modification of general $\Gamma-$convergence concepts.

Proposition 2.5.2. *It is possible to choose the parameters δ_ε and the sequence $(\nu_\varepsilon)_\varepsilon$ in such a way that the sequence of minimizers $(\overline{\nu}_\varepsilon)_\varepsilon$ of J_ε tends to $\overline{\nu}$ in the weak* topology (or, equivalently, with respect to the W_2 distance).*

Proof. We choose $\nu_\varepsilon \in L^2(\Omega)$ such that $J(\nu_\varepsilon) \leq J(\overline{\nu}) + \varepsilon^2/2$, and $\nu_\varepsilon \rightharpoonup \overline{\nu}$. This is possible thanks to lemma 2.5.3: just choose an L^2 sequence $(\nu_\varepsilon)_\varepsilon$ which approximates $\overline{\nu}$ in the strong topology of $H^1(\Omega)'$ and noticing that also the interaction term is in fact continuous with respect to this convergence. It is not difficult to choose the densities of the measures ν_ε to be positive as required. Then we set $\delta_\varepsilon = \varepsilon^2(\|\nu_\varepsilon\|_{L^2(\Omega)})^{-2}/2$.

So we have

$$J(\overline{\nu}_\varepsilon) + \varepsilon W_2^2(\overline{\nu}_\varepsilon, \nu_\varepsilon) + \delta_\varepsilon \|\overline{\nu}_\varepsilon\|^2_{L^2(\Omega)} \leq J(\nu_\varepsilon) + \delta_\varepsilon \|\nu_\varepsilon\|^2_{L^2(\Omega)} \leq J(\overline{\nu}) + \varepsilon^2.$$

Since $\overline{\nu}$ is a minimizer for J we have $J(\overline{\nu}_\varepsilon) \geq J(\overline{\nu})$, and so we get

$$J(\overline{\nu}) + \varepsilon W_2^2(\overline{\nu}_\varepsilon, \nu_\varepsilon) \leq J(\overline{\nu}) + \varepsilon^2,$$

where we have neglected the positive term $\delta_\varepsilon \|\overline{\nu}_\varepsilon\|^2_{L^2(\Omega)}$. By simplifying and dividing by ε we get

$$W_2^2(\overline{\nu}_\varepsilon, \nu_\varepsilon) \leq \varepsilon,$$

and so

$$W_2(\overline{\nu}_\varepsilon, \overline{\nu}) \leq \sqrt{\varepsilon} + W_2(\overline{\nu}, \nu_\varepsilon) \to 0,$$

which is the thesis.

□

Lemma 2.5.3. *The subspace $C_c^\infty(\Omega) \subset L^2(\Omega)$ is dense in the Hilbert space $H^1(\Omega)'$.*

Proof. It is sufficient to show the following implication:

$$\xi \in H^1(\Omega)',\ \langle \xi, f \rangle_{H^1(\Omega)'} = 0 \text{ for all } f \in C_c^\infty(\Omega) \Rightarrow \xi = 0.$$

After calling ψ_ξ and ψ_f the potentials of ξ and f, respectively, we have

$$\langle \xi, f \rangle_{H^1(\Omega)'} = \langle \psi_\xi, \psi_f \rangle_{H^1(\Omega)} = \int_\Omega \psi_\xi \psi_f + \int_\Omega \nabla\psi_\xi \cdot \nabla\psi_f = \int_\Omega \psi_\xi f.$$

Consequently, the condition of being ξ orthogonal to every $f \in C_c^\infty(\Omega)$ in $H^1(\Omega)'$ implies that the potential of ξ must be orthogonal in $L^2(\Omega)$ to all C_c^∞ functions. So ψ_ξ must be identically 0 and then $\xi = 0$. □

Having established the convergence of the minimizers $\overline{\nu}_\varepsilon$ to $\overline{\nu}$, we look for uniform estimates of such minimizers. From now on, we will make use of the following assumption on the function V:

Vdiod *(V depends increasingly on distances)*: V is a function of the form $V(x,y) = v(|x-y|^2)$ for a C^2 strictly increasing function v with $v'(s) > 0$ for $s > 0$.

Obviously, under this hypothesis, V is a symmetric function and so $V = V^s$ and $T_\nu = T_\nu^s$ for every probability measure ν.

2.5.1 L^∞ estimates in the convex case

Theorem 2.5.4. *Suppose that Ω is a bounded, regular and strictly convex open subset of $\mathbb{R}^2$ and that* **Vdiod** *holds. Then, every minimizer $\overline{\nu}_\varepsilon$ of J_ε is an absolutely continuous measure with L^∞ density, bounded by a uniform constant depending on Ω and on $\|V\|_{C^2(\Omega)}$. Consequently, $\overline{\nu}$ has a density bounded by the same constant as well.*

Proof. We write down a necessary optimality condition on $\overline{\nu}_\varepsilon$. To obtain it, we act as in the proof of proposition 2.4.2. We have only to consider two additional terms. The L^2 term is easy to deal with: if we set $\nu_{\varepsilon,t} = \nu_\varepsilon + t(p - \nu_\varepsilon)$ for small ε, $t \in [0,1]$ and an arbitrary probability $p \in L^2(\Omega)$, we have

$$\lim_{t\to 0} \frac{\|\nu_{\varepsilon,t}\|^2_{L^2(\Omega)} - \|\overline{\nu}_\varepsilon\|^2_{L^2(\Omega)}}{t} = 2\int_\Omega (p - \overline{\nu}_\varepsilon)\overline{\nu}_\varepsilon.$$

For the Wasserstein term, we behave as in last chapter and in [28]. Let us choose for each t a Kantorovich potential $\psi_{\varepsilon,t}$ for the transport between $\nu_{\varepsilon,t}$ and ν_ε, and let ψ_ε be the only Kantorovich potential (up to additive constants) between $\overline{\nu}_\varepsilon$ and ν_ε (uniqueness comes from the fact that the density of ν_ε is positive everywhere on the connected open set Ω). We can choose all these Kantorovich potentials to vanish on a same point. Remember that they are all $L-$Lipschitz functions. We then have

$$W_2^2(\nu_{\varepsilon,t},\nu_\varepsilon) - W_2^2(\overline{\nu}_\varepsilon,\nu_\varepsilon) \le t\int_\Omega \psi_{\varepsilon,t}\, d(p - \overline{\nu}_\varepsilon),$$

and so

$$\limsup_{t\to 0} \frac{W_2^2(\nu_{\varepsilon,t},\nu_\varepsilon) - W_2^2(\overline{\nu}_\varepsilon,\nu_\varepsilon)}{t} \le \int_\Omega \psi_\varepsilon\, d(p - \overline{\nu}_\varepsilon).$$

We have used the fact that, up to subsequences, the sequence $(\psi_{\varepsilon,t})_t$ has a limit and such a limit must be a Kantorovich potential between $\overline{\nu}_\varepsilon$ and ν_ε, and so it must be ψ_ε.

By this considerations and the same technique as in proposition 2.4.2, we get

$$\begin{aligned}
\delta_\varepsilon \overline{\nu}_\varepsilon + \frac{\varepsilon}{2}\psi_\varepsilon + \phi_\varepsilon + T_{\overline{\nu}_\varepsilon} &\geq c_\varepsilon \text{ in } \Omega; &(2.5.1)\\
\delta_\varepsilon \overline{\nu}_\varepsilon + \frac{\varepsilon}{2}\psi_\varepsilon + \phi_\varepsilon + T_{\overline{\nu}_\varepsilon} &= c_\varepsilon \text{ for } \overline{\nu}_\varepsilon - a.e. x \in \Omega. &(2.5.2)
\end{aligned}$$

Here ϕ_ε is the potential of $\overline{\nu}_\varepsilon$, and we have identified $\overline{\nu}_\varepsilon$ with its density (obviously $\overline{\nu}_\varepsilon \in L^2(\Omega)$). We may write

$$\overline{\nu}_\varepsilon = \frac{1}{\delta_\varepsilon}\left(c_\varepsilon - \phi_\varepsilon - T_{\overline{\nu}_\varepsilon} - \frac{\varepsilon}{2}\psi_\varepsilon\right)_+ . \tag{2.5.3}$$

Since $\overline{\nu}_\varepsilon$ is L^2 we have $\phi_\varepsilon \in H^2(\Omega) \subset C^{0,\alpha}(\Omega)$, and this shows that $\overline{\nu}_\varepsilon$ is Hölder continuous, since all the functions appearing in the positive part are at least Hölder continuous. Consequently, since $\overline{\nu}_\varepsilon \in C^{0,\alpha}(\Omega)$, the function ϕ_ε turns out to be a $C^{2,\alpha}$ function.

We look for a maximum point x_ε of $\overline{\nu}_\varepsilon$: in it we have a local minimum of the sum $\frac{\varepsilon}{2}\psi_\varepsilon + \phi_\varepsilon + T_{\overline{\nu}_\varepsilon}$. Thanks to $x \mapsto \psi_\varepsilon(x) - x^2$ being concave, we may write $\psi_\varepsilon \leq l + Q$, with equality in x_ε, where l is an affine function and $Q(x) = x^2$. Consequently x_ε is a local minimum for the sum $\frac{\varepsilon}{2}(l+Q) + \phi_\varepsilon + T_{\overline{\nu}_\varepsilon}$. Lemma 2.5.5 shows that x_ε does not belong to the boundary of Ω, at least for ε sufficiently small. So, since x_ε is an interior point and all the functions involved are at least twice differentiable, we may write, taking the Laplacians,

$$0 \leq 2\varepsilon + \Delta\phi_\varepsilon(x_\varepsilon) + \Delta T_{\overline{\nu}_\varepsilon}(x_\varepsilon). \tag{2.5.4}$$

In this case we can use $\overline{\nu}_\varepsilon = \phi_\varepsilon - \Delta\phi_\varepsilon$ to estimate $\overline{\nu}_\varepsilon(x_\varepsilon)$. In fact in x_ε we have $\overline{\nu}_\varepsilon(x_\varepsilon) > 0$ and so

$$\phi_\varepsilon(x_\varepsilon) - M \leq c_\varepsilon, \tag{2.5.5}$$

where $M = \sup|\psi_\varepsilon| + |T_{\overline{\nu}_\varepsilon}|$ can be uniformly estimated, ψ_ε being $L-$Lipschitz functions vanishing at a given point in Ω and $|T_\nu| \leq \sup|V|$ for every probability ν. So it is sufficient to estimate c_ε. To do this we can integrate (2.5.3), obtaining

$$\delta_\varepsilon \geq c_\varepsilon|\Omega| - M|\Omega| - 1, \tag{2.5.6}$$

where we used the fact that both $\overline{\nu}_\varepsilon$ and ϕ_ε are probability measures. Putting together (2.5.5) and (2.5.6) we get $\phi_\varepsilon \leq C$, being C a constant depending on Ω and $\sup V$, and so, by recalling the equality $\overline{\nu}_\varepsilon = \phi_\varepsilon - \Delta\phi_\varepsilon$ and the inequality (2.5.4) we get

$$\overline{\nu}_\varepsilon(x_\varepsilon) \leq C + 2\varepsilon + \|V\|_{C^2(\Omega)} .$$

Since x_ε is a maximum point we have got an L^∞ estimate on $\overline{\nu}_\varepsilon$. □

Lemma 2.5.5. *Suppose that* Ω *is a bounded, regular and strictly convex open subset of* $\mathbb{R}^2$ *and that* **Vdiod** *holds. Then, at least for small* ε, *we have* $x_\varepsilon \in \overset{\circ}{\Omega}$.

Proof. Suppose, on the contrary, to have a sequence $(x_\varepsilon)_\varepsilon$ contained in the boundary $\partial\Omega$. Such x_ε is a local minimum point for $\frac{\varepsilon}{2}\psi_\varepsilon + \phi_\varepsilon + T_{\overline{\nu}_\varepsilon}$. In a local minimum point on the boundary the normal exterior derivative should be non positive. The derivative of ψ_ε may also not exist, but we may use the fact that ψ_ε is an $L-$Lipschitz function. Being multiplied by ε, and vanishing by definition the normal derivative of ϕ_ε, it is not difficult to check that we should have

$$\limsup_{\varepsilon\to 0} \frac{\partial T_{\overline{\nu}_\varepsilon}}{\partial n}(x_\varepsilon) \leq 0. \tag{2.5.7}$$

On the other hand, we have

$$\frac{\partial T_{\overline{\nu}_\varepsilon}}{\partial n}(x_\varepsilon) = 2\int_\Omega v'(|x_\varepsilon - y|^2)(x_\varepsilon - y)\cdot n(x_\varepsilon)\,\overline{\nu}_\varepsilon(dy) \geq a_\delta \delta \overline{\nu}_\varepsilon(\Omega \setminus S_\delta(x_\varepsilon)),$$

where a_δ is the minimum value of v' on $[\delta^2, \operatorname{diam}\Omega^2]$ and, for every point $x \in \partial\Omega$, we define $S_\delta(x) = \{y \in \Omega \,|\, (x-y)\cdot n(x) \leq \delta\}$. Condition (2.5.7) implies that, for every $\delta > 0$, it holds $\overline{\nu}_\varepsilon(\Omega \backslash S_\delta(x_\varepsilon)) \to 0$. Taking a limit point x_0 of the sequence $(x_\varepsilon)_\varepsilon$, we will show that this implies that the measure which is the limit of the sequence $(\overline{\nu}_\varepsilon)_\varepsilon$ is concentrated on x_0. This is impossible, since this limit measure is $\overline{\nu}$, which is optimal for J, and so it belongs to $H^1(\Omega)'$. Yet in two dimensions a measure concentrated on a single point does not belong to such a space. To conclude, it is then sufficient to show that, for each ball $\overline{B(x_0, r)}$, it holds $S_\delta(x_\varepsilon) \subset \overline{B(x_0,r)}$ for sufficiently small δ and ε, thus getting $\lim_\varepsilon \overline{\nu}_\varepsilon(\overline{B(x_0,r)}) = 1$ for every $r > 0$. If not, we would have a sequence $(y_{\delta,\varepsilon})_{\delta,\varepsilon}$ such that $(x_\varepsilon - y_{\delta,\varepsilon})\cdot n(x_\varepsilon) < \delta$ and $|x - y_{\delta,\varepsilon}| > r$. At the limit we get a point $y \in \partial\Omega$ such that $(x_0 - y)\cdot n(x_0) \leq 0$ (which implies, in a strictly convex Ω, $x_0 = y$), but $|x_0 - y| \geq r$, and this is absurd. □

Remark 2.5.6. If we want to consider the one-dimensional case (with Ω an interval) the proof of lemma 2.5.5 has to be modified: it is sufficient to say that a measure concentrated at a single point, which is a terminal point of the interval, cannot be optimal. The potential of such a measure can be explicitly computed, being an exponential function, and it can be proven that the optimality condition of proposition 2.4.2 cannot hold, at least under the additional assumption $v'(0) = 0$ in **Vdiod**.

Remark 2.5.7. In chapter 3 and in [67] a different proof for the fact that the maximum point does not lie on the boundary is provided. Such a proof has the advantage of getting rid of the strict convexity assumption. Unluckily, it is easy to be shown under some extra regularity assumption on the Kantorovich potential that we cannot ensure here. It seems that it could be extended to the case of Lipschitz potentials but through a quite heavy procedure. This is the reason why we did not develop an alternative proof for Lemma 2.5.5. Anyway, we refer to the proof of Theorem 3.2.2 in next chapter.

We conclude this part of the section by a consideration on the consequences of this result on the regularity of the potential ϕ.

Corollary 2.5.8. *The potential ϕ of an optimal measure is a $W^{2,p}$ function for any $1 \leq p < +\infty$ and then a $C^{1,\alpha}$ function too.*

Proof. Just apply proposition 2.5.1 and consider that $\overline{\nu} \in L^\infty(\Omega) \subset L^p(\Omega)$ for any finite p. The second part of the statement is just a consequence of well-known embedding theorems. □

2.5.2 Interior L^2 estimates in the general case

In this section, we look for weaker estimates which are valid in the case of a non convex domain Ω. Let us write $\partial\Omega = \Gamma_1 \cup \Gamma_2$, where $\Gamma_1 = \partial\Omega \cap \partial(\mathrm{co}\,\Omega)$ and $\Gamma_2 = \partial\Omega \setminus \partial(\mathrm{co}\,\Omega)$.

Theorem 2.5.9. *Suppose that Γ_1 is a strictly convex regular boundary and that* **Vdiod** *holds. Then, given a Lipschitz function θ such that $d(\mathrm{spt}\,\theta, \Gamma_2) > 0$, the sequence of functions $(\theta\overline{\nu}_\varepsilon)_\varepsilon$ is bounded in L^2.*

Proof. We start by testing equation (2.3.2) for $\overline{\nu}_\varepsilon$ against the function $\theta^2\overline{\nu}_\varepsilon$:

$$\int_\Omega \overline{\nu}_\varepsilon^2\theta^2 = \int_\Omega \phi_\varepsilon\theta^2\overline{\nu}_\varepsilon + \int_\Omega \nabla\phi_\varepsilon \cdot \nabla(\theta^2\overline{\nu}_\varepsilon). \tag{2.5.8}$$

Since on $\mathrm{spt}\,\overline{\nu}_\varepsilon$ we have $\nabla\phi_\varepsilon = -\delta_\varepsilon\nabla\overline{\nu}_\varepsilon - \nabla(\frac{\varepsilon}{2}\psi_\varepsilon + T_{\overline{\nu}_\varepsilon})$, we get

$$\begin{aligned}\int_\Omega \overline{\nu}_\varepsilon^2\theta^2 &= \int_\Omega \phi_\varepsilon\theta^2\overline{\nu}_\varepsilon - \delta_\varepsilon\int_\Omega \nabla\overline{\nu}_\varepsilon\cdot\nabla(\theta^2\overline{\nu}_\varepsilon) - \int_\Omega \nabla(\frac{\varepsilon}{2}\psi_\varepsilon + T_{\overline{\nu}_\varepsilon})\cdot\nabla(\theta^2\overline{\nu}_\varepsilon)\\ &\leq \int_\Omega \phi_\varepsilon\theta^2\overline{\nu}_\varepsilon - \delta_\varepsilon\int_\Omega \theta^2|\nabla\overline{\nu}_\varepsilon|^2 - 2\delta_\varepsilon\int_\Omega \overline{\nu}_\varepsilon\theta\nabla\overline{\nu}_\varepsilon\cdot\nabla\theta\\ &\quad + \int_\Omega \Delta(\frac{\varepsilon}{2}\psi_\varepsilon + T_{\overline{\nu}_\varepsilon})\theta^2\overline{\nu}_\varepsilon - \int_{\partial\Omega}\theta^2\overline{\nu}_\varepsilon\left(\frac{\partial T_{\overline{\nu}_\varepsilon}}{\partial n} - \frac{\varepsilon}{2}L\right).\end{aligned} \tag{2.5.9}$$

Using once more $\delta_\varepsilon \nabla \overline{\nu}_\varepsilon = -\nabla\phi_\varepsilon - \nabla(\frac{\varepsilon}{2}\psi_\varepsilon + T_{\overline{\nu}_\varepsilon})$ on $\operatorname{spt}\overline{\nu}_\varepsilon$ in (2.5.9), we get:

$$\int_\Omega \overline{\nu}_\varepsilon^2\theta^2 \le \int_\Omega \phi_\varepsilon\theta^2\overline{\nu}_\varepsilon - \delta_\varepsilon\int_\Omega \theta^2|\nabla\overline{\nu}_\varepsilon|^2 + 2\int_\Omega \overline{\nu}_\varepsilon\theta\nabla\phi_\varepsilon\cdot\nabla\theta$$
$$+2\int_\Omega \overline{\nu}_\varepsilon\theta\nabla(\frac{\varepsilon}{2}\psi_\varepsilon + T_{\overline{\nu}_\varepsilon})\cdot\nabla\theta + \int_\Omega \Delta(\frac{\varepsilon}{2}\psi_\varepsilon + T_{\overline{\nu}_\varepsilon})\theta^2\overline{\nu}_\varepsilon - \int_{\partial\Omega}\theta^2\overline{\nu}_\varepsilon\left(\frac{\partial T_{\overline{\nu}_\varepsilon}}{\partial n} - \frac{\varepsilon}{2}L\right). \tag{2.5.10}$$

Notice that the Laplacian appearing in the fifth term is composed by two parts: the Laplacian of a C^2 function and the Laplacian of a concave one, which is a negative measure. We have six terms that must be estimated:

- the first one is bounded by $\|\theta\overline{\nu}_\varepsilon\|_{L^2(\Omega)}\|\theta\phi_\varepsilon\|_{L^2(\Omega)}$;
- the second is negative;
- the third is bounded by $\|\nabla\phi_\varepsilon\|_{L^2(\Omega)}\|\theta\overline{\nu}_\varepsilon\|_{L^2(\Omega)}\operatorname{lip}\theta$;
- the fourth by $\|\theta\overline{\nu}_\varepsilon\|_{L^2(\Omega)}(\frac{\varepsilon}{2}L + \operatorname{lip}V)\operatorname{lip}\theta$;
- the fifth by $(2\varepsilon + \|V\|_{C^2(\Omega)})\left\|\theta^2\right\|_{L^\infty(\Omega)}$;
- the last one is negative for small ε and it can be proven exactly as in the proof of lemma 2.5.5.

The proof is then achieved, since the sequence $(\phi_\varepsilon)_\varepsilon$ is bounded in $H^1(\Omega)$, thanks to $\|\phi_\varepsilon\|^2_{L^2(\Omega)} + \|\nabla\phi_\varepsilon\|^2_{L^2(\Omega)} \le J_\varepsilon(\overline{\nu}_\varepsilon)$. Moreover, the left hand side in (2.5.10) is quadratic in $\|\theta\overline{\nu}_\varepsilon\|_{L^2(\Omega)}$ and the right hand side at most linear, which gives the estimate we were looking for. □

Theorem 2.5.9 gives a local L^2 bound on the densities $\overline{\nu}_\varepsilon$: this enables us, together with the optimality conditions of section 5, to state a stronger regularity result.

Theorem 2.5.10. *Suppose that Γ_1 is a strictly convex regular boundary and that* **Vdiod** *holds. Then any optimal measure $\overline{\nu}$ for J can be expressed as $\overline{\nu} = \overline{\nu}^a + \overline{\nu}^s$, with $\overline{\nu}^a \in L^\infty(\Omega)$ and $\overline{\nu}^s$ a singular measure supported on $\overline{\Gamma}_2$.*

Proof. By theorem 2.5.9 we get that $\overline{\nu}$ is locally L^2 in $\Omega\setminus\overline{\Gamma}_2$. This means, by interior elliptic regularity theory, that its potential ϕ is locally H^2 in the same set, and thus continuous. Hence the equality $\phi = c - T_{\overline{\nu}}$ given by optimality conditions holds true on the whole $\operatorname{spt}\overline{\nu}$. So the following holds

$$\phi I_{\operatorname{spt}\overline{\nu}} = (c - T_{\overline{\nu}})I_{\operatorname{spt}\overline{\nu}} \in L^\infty(\Omega), \tag{2.5.11}$$

Moreover, $\mathcal{L}^2-$almost everywhere on $\operatorname{spt}\overline{\nu}$, we also have $\Delta\phi = -\Delta T_{\overline{\nu}}$ and so

$$\Delta\phi I_{\operatorname{spt}\overline{\nu}} = -\Delta T_{\overline{\nu}} I_{\operatorname{spt}\overline{\nu}} \in L^\infty(\Omega). \tag{2.5.12}$$

(2.5.11) and (2.5.12) together, imply

$$\overline{\nu} I_A = (\phi - \Delta\phi) I_{\operatorname{spt}\overline{\nu}} \in L^\infty(\Omega),$$

where $A = \overline{\Omega} \setminus \Gamma_2$. Finally, set $\nu^a = \overline{\nu} I_A$ and $\nu^s = \overline{\nu} I_{\Gamma_2}$. □

Remark 2.5.11. In section 8, an example will be given to show that it is in fact possible that an optimal ν gives a positive mass to Γ_2

2.6 Qualitative properties of the minimizers

In this section, we give some qualitative properties regarding the support of an optimal measure ν. This turns out to be very important, thanks to the following result. In all the section Ω will be strictly convex, regular, and condition **Vdiod** will hold.

Proposition 2.6.1. *The L^∞ density of any optimal measure ν coincides almost everywhere in* $\operatorname{spt}\nu$ *with a continuous function.*

Proof. Thanks to the regularity of the potential ϕ we may say that the equality $\phi = c - T_\nu$ holds everywhere in the support and that, for the Laplacian of ϕ, which is an L^p function, it holds $\Delta\phi = -\Delta T_\nu$ a.e. From $\nu = \phi - \Delta\phi$ and $V \in C^2(\Omega)$, which implies $T_\nu \in C^2(\Omega)$, we get the thesis. □

As a consequence of the previous result, we may say that the reason for possible irregular behavior of ν must be traced back to the shape of its support. As far as this shape is concerned, we can only give two general results, whose statements are quite weak.

Proposition 2.6.2. *Suppose, other than the general assumptions of the section, that V is strictly convex. Then the support of ν has non-empty interior.*

Proof. We will show that $\operatorname{spt}\nu$ contains a small ball around the point x_0 defined by $x_0 = \operatorname{argmin} T_\nu$. The function T_ν inherits strict convexity from V, and so there exist just one minimizer and just one critical point for T_ν. We start by saying that, under the assumption of theorem 2.5.4, we must have $\operatorname{spt}\nu \cap \partial\Omega = \emptyset$. Indeed, ϕ being a $C^{1,\alpha}$ function, it holds $\nabla\phi = -\nabla T_\nu$

on the whole support and this, by calculating ∇T_ν as in lemma 2.5.5, would otherwise prevent the normal derivative of ϕ from vanishing on $\partial\Omega$.

We now want to show that $x_0 \in \operatorname{spt}\nu$: to do this consider a maximum point $\overline{x}$ for ϕ. Such a point must be placed in $\operatorname{spt}\nu$, since outside it holds $\Delta\phi = \phi$ and on an interior maximum point we should have a strictly positive value for ϕ. The same consideration can be performed on the boundary, since we already know that the normal derivative vanishes and no maximum point on the boundary with vanishing normal derivative and positive Laplacian is allowed. Notice that outside $\operatorname{spt}\nu$ the function ϕ is an analytic function because of standard elliptic regularity theory and so it makes sense to consider its Laplacian on $\partial\Omega$ too.

Now, it must hold

$$0 = \nabla\phi(\overline{x}) = -\nabla T_\nu(\overline{x}),$$

and so $\overline{x} = x_0$. Consequently x_0 is a point in $\operatorname{spt}\nu$ and then in Ω.

Let us now consider for a fixed small value of $\varepsilon > 0$ and for δ in a ball near $0 \in \mathbb{R}^2$ the functions

$$f_{\varepsilon,\delta}(x) = \phi(x) - \frac{\varepsilon}{2}|x - (x_0 + \delta)|^2.$$

The parameter ε has to be chosen in such a way that in any maximum point of $f_{\varepsilon,\delta}$ it holds $\phi > \varepsilon$ (it is in fact sufficient to satisfy the inequality $|\Omega|^{-1} > \varepsilon(1 + (\operatorname{diam}\Omega)^2/2))$. After choosing $\varepsilon > 0$ in such a way, we will think of it as a fixed parameter.

Now consider $x_\delta \in \operatorname{argmin} f_{\varepsilon,\delta}$. Such a point cannot lie on the boundary because of the sign of the normal derivative, and it cannot be outside $\operatorname{spt}\nu$, by computing the Laplacian. So $x_\delta \in \operatorname{spt}\nu$. The point x_δ is characterized by

$$[\varepsilon id + \nabla T_\nu](x_\delta) = \varepsilon(x_0 + \delta),$$

the application on the left hand side being injective since it is monotone (in the usual sense for vector-valued maps). If $\delta = 0$ the solution is given by $x_\delta = x_0$, and so, by standard local inversion theorems, the set of points x_δ covers a small ball around x_0. Such a ball must consequently be contained in $\operatorname{spt}\nu$. □

Our next result deals with the topology of the support

Proposition 2.6.3. *Suppose, other than the general assumptions of the section, that V is strictly subharmonic, i.e. $\Delta V > 0$. Then the support of ν is simply connected, in the sense that, if $\omega \subset \Omega$ is an open set such that $\partial\omega \subset \operatorname{spt}\nu$, then $\omega \subset \operatorname{spt}\nu$.*

Proof. We consider a maximum point x_0 for $\phi + T_\nu$ in $\overline{\omega}$. Let us recall that $\phi = c - T_\nu$ in $\operatorname{spt}\nu$ and $\phi \geq c - T_\nu$ everywhere. So, if the maximum point belongs to $\operatorname{spt}\nu$, we have $\phi = c - T_\nu$ on ω. On the other hand, it is impossible to have $x_0 \in \omega \setminus \operatorname{spt}\nu$ because there we have $\Delta(\phi + T_\nu) = \phi + \Delta T_\nu > \phi \geq 0$, since T_ν inherits strict subharmonicity from V. Consequently, x_0 must belong to $\overline{\omega} \cap \operatorname{spt}\nu$. Then we have $\phi = c - T_\nu$ and so $\Delta\phi = -\Delta T_\nu$ in the whole ω and so

$$\nu = \phi + \Delta T_\nu > \phi \geq 0 \text{ in } \omega,$$

which obviously implies $\omega \subset \operatorname{spt}\nu$. □

2.7 Geodesic convexity in dimension one

It is worthwhile to consider the case where $\Omega = [-R, R]$ is a bounded interval in $\mathbb{R}$, instead of a two-dimensional open set. Obviously from the point of view of applications it sounds less interesting, even if sometimes in urban economics unidimensional models have been used to deal with the case of very long and narrow cities (and in fact some towns on the sea shore are not far from being one-dimensional). From a mathematical point of view, the main interest lies in the fact that we can show the functional J to be displacement convex (or strictly displacement convex), under convexity assumption on V. See Section 0.3 for details about this notion. This gives uniqueness of the minimizer, but it is also important since displacement convexity has never been studied for functionals of the form of the squared $(H^1)'$ norm. Anyway, the techniques here used to get this term geodesically convex are very specific to the one-dimensional case .

Before presenting the displacement convexity result, we need to recall the concept of Green function and its link to the squared $(H^1)'$ norm. The following result can be adapted to any dimension.

Proposition 2.7.1. *For every measure $\nu \in H^1(\Omega)'$ it holds*

$$\|\nu\|^2_{H^1(\Omega)'} = \int_{\Omega\times\Omega} G(x,y)\, \nu(dx)\nu(dy),$$

the function $G_x = G(x,\cdot)$ being for every $x \in \Omega$ the solution to

$$\begin{cases} -\Delta_y G_x + G_x = \delta_x & \text{in } \overset{\circ}{\Omega}, \\ \frac{\partial G_x}{\partial n} = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.7.1}$$

i.e. G is the Green function for the operator $-\Delta + id$ with Neumann boundary conditions.

Proof. First, we notice that it holds

$$\|\nu\|^2_{H^1(\Omega)'} = \int_\Omega \phi^2 + \int_\Omega |\nabla\phi|^2 = \int_\Omega \phi\, d\nu.$$

Then the general theory on Green functions allows us to say that it holds $\phi(x) = \int_\Omega G(x,y)\nu(dy)$. Integrating once more with respect to ν gives the thesis. □

Now we will take $\Omega = [-R,R] \subset \mathbb{R}$ and we will divide the square $[-R,R]\times[-R,R]$ into two triangles:

$$T^+ = \{(x,y) \in [-R,R]\times[-R,R] \,|\, x < y\}$$
$$T^- = \{(x,y) \in [-R,R]\times[-R,R] \,|\, x > y\}$$

Theorem 2.7.2. *If $\Omega = [-R,R] \subset \mathbb{R}$ and if V is a convex function of the pair (x,y) then the functional J is strictly displacement convex. Consequently, it admits a unique minimizer.*

Proof. An easy computation shows that the Green function in (2.7.1) is given, in the case of the interval $[-R,R]$, by

$$G(x,y) = \begin{cases} \frac{\cosh(x+R)\cosh(y-R)}{\sinh(2R)} & \text{if } (x,y) \in T^+, \\ \frac{\cosh(x-R)\cosh(y+R)}{\sinh(2R)} & \text{if } (x,y) \in T^- \end{cases}$$

denoting by cosh and sinh the hyperbolic cosin and sin, respectively. It is also easy to check that both expressions, the one valid in T^+ and the one in T^-, are strictly convex functions.

Let us now consider a displacement interpolation $\nu_t = [(1-t)id + tS]_\#\nu$ and take

$$J(\nu_t) = \int_{-R}^{R}\int_{-R}^{R} (G+V)(x + t(S(x)-x), y + t(S(y)-y))\,\nu(dx)\nu(dy).$$

Since S must be an optimal transport with respect to the cost $|x-y|^2$ it is well-known that it is an increasing map: consequently $(x,y) \mapsto ((1-t)x + tS(x), (1-t)y + S(y))$ sends each of the triangles T^+, T^- into itself. Then, in order to get $t \mapsto J(\nu_t)$ strictly convex, it is sufficient to have strict convexity of $G+V$ in each triangle. Our hypothesis ensures it and we get the thesis. □

Remark 2.7.3. In the assumptions of theorem 2.7.2 the convexity in each triangle T^+, T^- of $G+V$ is sufficient: in particular, also some concave functions V are allowed.

Remark 2.7.4. In [71], Problem 5.17 addresses at the following question: find new and meaningful functionals which have some displacement convexity properties. The case of the functional $\nu \mapsto \|\nu\|^2_{H^1(\Omega)'}$ in one dimension is quite trivial and specific but maybe it is worth to be linked to this question.

2.8 The quadratic case in two dimensions

We now develop the particular case where $V(x,y) = |x-y|^2$. For such a choice for V and particular Ω we are able to give an almost explicit solution.

First, we make some general considerations on the quadratic kernel. Notice that, for every probability measure ν, we have:

$$T_\nu(x) = |x - \operatorname{bar}(\nu)|^2 + \operatorname{Var}(\nu), \ H(\nu) = 2\operatorname{Var}(\nu),$$
$$J(\nu) = \|\nu\|^2_{H^1(\Omega)'} + 2\int_\Omega |x|^2 \nu(dx) - 2|\operatorname{bar}(\nu)|^2, \tag{2.8.1}$$

denoting by bar and Var the barycenter and the variance of a probability measure, respectively.

We also compute the variation of our functional J when we pass from ν to $\nu + h$, being h an admissible perturbation, i.e. $h = p - \nu$ with $p \in \mathcal{P}(\Omega)$:

$$J(\nu + h) = J(\nu) + 2\int_\Omega (\phi + T_\nu)\, dh + \|h\|^2_{H^1(\Omega)'} + \int_{\Omega\times\Omega} |x-y|^2\, h(dx)h(dy).$$

By using that h is a zero-mean signed measure, we may re-write the last term and get

$$J(\nu + h) = J(\nu) + 2\int_\Omega (\phi + T_\nu)\, dh + \|h\|^2_{H^1(\Omega)'} - \left|\int_\Omega x\, h(dx)\right|^2. \tag{2.8.2}$$

We are now going to analyze the case of Ω being the whole plane, a ball or a crown.

In the case $\Omega = \mathbb{R}^2$ it is clear that we face a lot of symmetries, with respect both to rotations and to translations. This second kind of symmetries enables us to consider just the problem where the barycenter of ν is fixed at 0. Given the set of minimizers for this sub-problem, we will get all the minimizers for the original problem by translating them of an arbitrary vector in $\mathbb{R}^2$.

The problem

$$\inf\{J(\nu) : \nu \text{ probability measure on } \mathbb{R}^2,\ \operatorname{bar}(\nu) = 0\ \}, \tag{2.8.3}$$

thanks to (2.8) or (2.8.2), turns out to be a strictly convex minimization problem. We will then find its unique minimizer by finding a measure ν satisfying the optimality condition, i.e. such that $x \mapsto \phi(x) + |x|^2$ is minimal $\nu-$almost everywhere. Equation (2.8.2) can be used to convince oneself that such a condition is in fact sufficient to have a minimum.

We will build a solution to the minimization problem by looking for a radial measure with radial potential satisfying proper conditions. The following useful result is given without proof because it is only a (nontrivial, we must admit) second-year exercise.

Lemma 2.8.1. *Consider the Cauchy problem*

$$\begin{cases} tu_a''(t) + u_a'(t) = \frac{u_a}{4} & \text{for } t \in (a, +\infty) \\ u_a(a) = C_a - a \\ u_a'(a) = -1, \end{cases} \tag{2.8.4}$$

depending on a parameter $a \in (a^-, a^+)$, where C_a is a decreasing function of a in the interval (a^-, a^+) and the following conditions hold:

$$\lim_{a \to a^+} C_a - a < 0 \quad \text{and} \quad \lim_{a \to a^-} C_a = +\infty.$$

Then there exists a number $\overline{a} \in (a^-, a^+)$ such that:

- *for $a < \overline{a}$ the solution u_a is convex and decreasing up to a point $T(a)$ where $u_a(T(a)) > 0$ and $u_a'(T(a)) = 0$ and then increasing with non-zero derivative;*
- *for $a = \overline{a}$ the solution u_a is convex, decreasing and positive on the whole $(a, +\infty)$ and it is infinitesimal together with its derivative as $t \to +\infty$;*
- *for $a > \overline{a}$ the solution becomes negative.*

Moreover, the map $a \mapsto T(a)$ is increasing and continuous and it holds

$$\lim_{a \to \overline{a}} T(a) = +\infty.$$

Theorem 2.8.2. *The unique solution to problem (2.8.3), and hence to the whole minimization problem in $\mathbb{R}^2$ (up to translations), can be obtained by using lemma 2.8.1, with $a^- = 0$, arbitrary large a^+ and*

$$C_a = \frac{1 + \frac{\pi}{2} a^2}{\pi a} - 4.$$

Then the solution is the measure ν whose density is given in the following, together with its potential ϕ:

$$\nu(x) = \begin{cases} C_{\overline{a}} + 4 - |x|^2 & \textit{for } |x|^2 \le \overline{a} \\ 0 & \textit{for } |x|^2 > \overline{a} \end{cases}; \ \phi(x) = \begin{cases} C_{\overline{a}} - |x|^2 & \textit{for } |x|^2 \le \overline{a} \\ u_{\overline{a}}(|x|^2) & \textit{for } |x|^2 > \overline{a} \end{cases}.$$

Proof. Thanks to the considerations made before, it is sufficient to check that ϕ is the potential of ν (by computing the Laplacian) and that ν is a probability, i.e. $\int \nu \, d\mathcal{L}^2 = 1$ (but C has been properly chosen); the optimality condition being immediately satisfied by construction ($\phi(x) + |x|^2$ is constant for $x \in \operatorname{spt} \nu$ and outside it is greater than this constant as a consequence of the convexity of $u_{\overline{a}}$). Similar computations are detailed in the proof of Theorem 2.8.4 □

The case of a bounded ball in $\mathbb{R}^2$ may be interesting as well. In this case, however, we may suffer of a loss of convexity, because we cannot reduce the problem to the simpler one with fixed barycenter. To avoid this difficulty, we will consider a small ball, such that $c_{\Omega,V} < 1$. Under this assumption, any critical point of the functional will be actually the unique minimizer. We will build the minimizer exactly as in the case of $\mathbb{R}^2$, by using Lemma 2.8.1. We keep the same choice of C, a^- and a^+. By inverting the map T in Lemma 2.8.1 we define a map $R \to a(R)$ given by $T(a(R)) = R^2$: this map is continuously increasing as well, and it obviously holds $a(R) < R^2$.

Theorem 2.8.3. *The unique solution to Problem (2.4.1), when $V(x, y) = |x-y|^2$, $\Omega = B(0, R)$ and R is small enough so that we have $c_{\Omega,V} < 1$, is the measure ν whose density is given in the following, together with its potential ϕ:*

$$\begin{aligned} \nu(x) &= \begin{cases} C_{a(R)} + 4 - |x|^2 & \textit{for } |x|^2 \le a(R), \\ 0 & \textit{for } a(R) < |x|^2 < R^2; \end{cases} \\ \phi(x) &= \begin{cases} C_{a(R)} - |x|^2 & \textit{for } |x|^2 \le a(R), \\ u_{a(R)}(|x|^2) & \textit{for } a(R) < |x|^2 < R^2. \end{cases} \end{aligned}$$

Proof. Simply act as in the proof of theorem 2.8.2 or have a look at computations in theorem 2.8.4. □

Let us now consider a circular crown with radii $R_1 < R_2$, i.e. the domain $\Omega = \overline{B(0, R_2)} \setminus B(0, R_1)$. To give a solution to the problem we will use once more Lemma 2.8.1, but this time we will slightly change the function C. Moreover, exactly as in the case of the ball, we will only deal with a small crown.

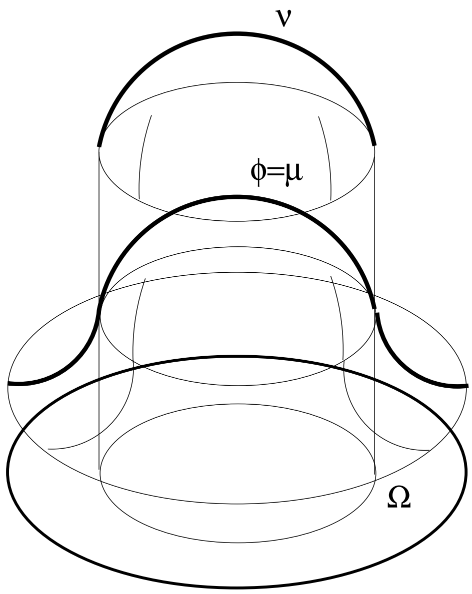

Figure 2.1: The solution in a small ball

Theorem 2.8.4. *The measure ν described in the following, together with its potential ϕ, is the unique solution to Problem (2.4.1), when $V(x,y) = |x-y|^2$, $\Omega = \overline{B(0,R_2)} \setminus B(0,R_1)$ and R_1 and R_2 are small enough so that we have $c_{\Omega,V} < 1$ and $4\pi R_1^2 < 1$:*

$$\begin{aligned}
\nu &= \nu^a + \nu^s; \\
\nu^a(x) &= \begin{cases} C_{a(R_2)} + 4 - |x|^2 & \text{for } R_1^2 < |x|^2 \le a(R_2) \\ 0 & \text{for } a(R_2) < |x|^2 < R_2^2 \end{cases}, \\
\nu^s &= 2R_1 \mathcal{H}^1 \llcorner \partial B(0,R_1), \\
\phi(x) &= \begin{cases} C_{a(R_2)} - |x|^2 & \text{for } R_1^2 < |x|^2 \le a(R_2) \\ u_{a(R_2)}(|x|^2) & \text{for } a(R_2) < |x|^2 < R_2^2 \end{cases},
\end{aligned}$$

where we have chosen, in lemma 2.8.1

$$C_a = \frac{1 - 4\pi R_1^2 + \frac{\pi}{2}(a^2 - R_1^4)}{\pi(a - R_1^2)} - 4,$$

putting $a^- = R_1^2$, and choosing $a(R_2)$ so that it satisfies $T(a(R_2)) = R_2^2$.

Proof. This time we give a quite detailed proof. We start by computing the mass of ν to show that it is a probability on $\overline{\Omega}$.

$$\begin{aligned}\nu(\overline{\Omega}) &= \nu^s(\overline{\Omega}) + \nu^a(\overline{\Omega}) = 4\pi R_1^2 + \int_{R_1}^{\sqrt{a(R_2)}} \left(C_{a(R_2)} + 4 - \rho^2\right) 2\pi\rho\, d\rho \\ &= 4\pi R_1^2 + \left(C_{a(R_2)} + 4\right)\pi\left(a(R_1) - R_1^2\right) - \frac{\pi}{2}\left(a(R_1)^2 - R_1^4\right) = 1.\end{aligned}$$

To show that ϕ is the potential of ν we divide Ω into two crowns: $\Omega_1 = \left\{x \in \Omega \,|\, R_1^2 \leq |x|^2 \leq a(R_2)\right\}$ and $\Omega_2 = \left\{x \in \Omega \,|\, a(R_2) < |x|^2 \leq R_2^2\right\}$. Then we have, for any $\psi \in C^1(\overline{\Omega})$,

$$\begin{aligned}\int_{\Omega_1\cup\Omega_2} \psi\phi + \nabla\psi\cdot\nabla\phi = \int_{\Omega_1} \psi\,(-\Delta\phi+\phi) + \int_{\partial\Omega_1\cap\partial\Omega} \psi\frac{\partial\phi}{\partial n} + \int_{\partial\Omega_1\cap\partial\Omega_2} \psi\frac{\partial\phi}{\partial n} + \\ \int_{\Omega_2} \psi\,(-\Delta\phi+\phi) + \int_{\partial\Omega_2\cap\partial\Omega} \psi\frac{\partial\phi}{\partial n} + \int_{\partial\Omega_2\cap\partial\Omega_1} \psi\frac{\partial\phi}{\partial n}.\end{aligned}$$

Let us have a look at the six terms in the right hand side:

- the first one equals $\int_{\overline{\Omega}} \psi d\nu^a$, because in Ω_1 we have $\nu^a = \phi + 4$ and $\Delta\phi = -4$;
- the second term is zero because ϕ has, by construction, vanishing normal derivative at $|x| = R_2$;
- the third and the sixth one are opposite and so they give a vanishing sum, because ϕ is C^1 by construction;
- the fourth term vanishes because outside $B(0, \sqrt{a(R_2)})$ we have $\Delta\phi = \phi$ as a consequence of (2.8.4);
- the fifth one equals $\int_{\overline{\Omega}} \psi d\nu^s$ by construction of ν^s.

After checking that ϕ is the potential of ν it is immediate to notice that, by construction, the optimality conditions are satisfied. □

Remark 2.8.5. Theorem 2.8.4 gives an example of an optimal ν composed by an L^∞ part and a singular part on $\Gamma_2 = \partial B(0, R_1)$, while giving no mass to $\partial B(0, R_2)$, which is the convex part of the boundary. See also Figure 2.2, where the thick part on the top of the picture stands for the part of measure concentrated on the interior boundary.

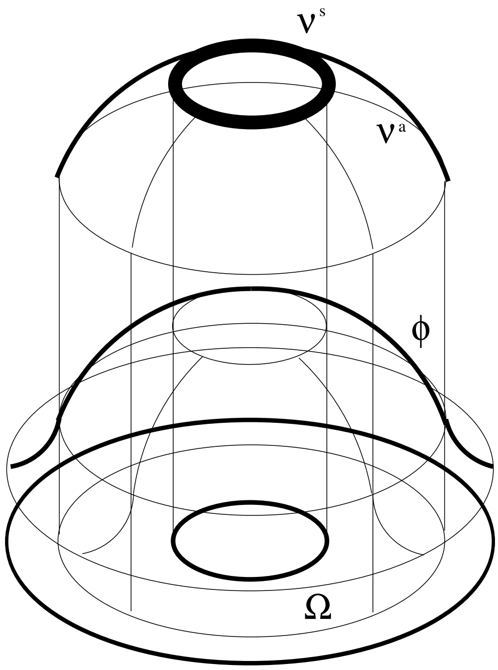

Figure 2.2: The solution in a small crown

Chapter 3

Transport and concentration problems with interaction

One of the aim of this chapter is to underline the possibility of studying variational problems of the kind of those studied in Chapters 1 and 2 as a new general class of optimization problems. Hence we start by a wide framework that contains them all, and then we present a brief study on a problem naturally arising from the comparison of the first two chapters. In this way we consider in an almost exhaustive way the possible problems involving traffic congestion or Wasserstein distances and atomic or interaction concentration functionals. We also refer to some problems which involve as transport costs other models (namely, branched transport models) and which will be more linked to the next parts of the thesis.

3.1 Variational problems for transport and concentration

The general problem we are interested in is

$$\min_{\mu,\nu\in\mathcal{P}(\Omega)} \mathfrak{F}(\mu,\nu) := T(\mu,\nu) + F(\mu) + G(\nu),$$

where the functional T quantifies in some way the distance between the two probability measures μ and ν according to a transport cost criterion, and F and G are functionals over the space $\mathcal{P}(\Omega)$ (the space of probability measures over a domain Ω) with opposite behavior: the first favors spread measures and penalizes concentration while the latter, on the other hand, favors concentrated measures. Obviously there are several other sub-problems that may be of interest, for instance the minimizations of the two separate functionals

$$\mathfrak{F}_\nu(\mu) := T(\mu,\nu) + F(\mu) \quad \text{and} \quad \mathfrak{F}^\mu(\nu) = T(\mu,\nu) + G(\nu),$$

where each time one of the variables is frozen. Also imposing constraints like $F(\mu) \leq H$, $G(\nu) \leq L \dots$ instead of adding penalizations in the functionals may sometimes be considered (and this is in fact the same as adding penalizations through some $0/+\infty$ functions).

These minimization problems are likely to appear in several phenomena both in nature and in decision science. For instance in [28], [37] and [66], as we have seen in the first two chapters of the thesis, these variational problems have been proposed for urban planning models. In this case the spread measure μ stands for inhabitants, the concentrated measure ν for services and they have to be close in a transportation distance sense. On the other hand, a possible choice of the functional $\mathfrak{F}_\nu$ has been proposed recently as a model for the formation of a certain kind of leaves or in general for ramified biological structures: if $\nu = \delta_0$ represents the point where the nutrient arrives to the leaf, then the shape of the leaf is modeled to optimize the quantity of light it receives from the sun taking also into account the effective transport cost for bringing the nutrient all over its shape (see Chapter 7).

We present here some choices for the functionals T and G. The choice of F is in fact easier since a very good class of concentration-penalizing functionals is given by local convex functionals over measures, for instance

$$F(\mu) = \begin{cases} \int_\Omega f(u)\, d\mathcal{L} & \text{if } \mu = u \cdot \mathcal{L} \\ +\infty & \text{otherwise,} \end{cases} \tag{3.1.1}$$

for any convex function f with $f(0) = 0$ which is superlinear at infinity. For these functionals we refer to [16]. Here $\mathcal{L}$ is a reference measure that may be chosen as the Lebesgue measure $\mathcal{L}^d$ if $\Omega \subset \mathbb{R}^d$. By Jensen inequality, spread measures with constant density are optimal for these functionals.

Possible choices of T are the following:

- terms involving Monge-Kantorovich optimal transport cost, through Wasserstein distances, for instance: $T(\mu,\nu) = W_p^p(\mu,\nu)$ (*Wasserstein*);
- terms taking into account traffic congestion effects, as in [37] and in Chapter 2 (*congestion*): here the idea is to encourage separate transportation, because in some applications if too many paths concur in a same point the configuration turns out to be less efficient;

- terms reflecting the natural ramified structure of a transportation network (*branching*) as in [72], where a new distance on probability measures is introduced according to this criterion: on the contrary, here the idea is to encourage joint transportation, in view of the applications where common paths are more efficient.

This last possibility is the most suitable for model involving natural branching structures like leaves (see Section 7.3.3), while the first two seem to be quite natural in urban planning. As we pointed out in the introduction, these three functionals correspond to three different models, where concentration of the transport is indifferent (according to Monge), discouraged (to avoid congestion effects) or encouraged (to reduce the building cost of the network, thus getting ramified structures).

For the functional G, before presenting a list of examples, we give a possible definition of the concept of concentration-preferring.

Definition 3.1.1. We say that $G : \mathcal{P}(\Omega) \to \mathbb{R}$ is a *concentration-preferring* functional if $G(t_\sharp \nu) \leq G(\nu)$ for any measure $\nu \in \mathcal{P}(\Omega)$ and any $1-$Lipschitz continuous map $t : \Omega \to \Omega$.

It is easy to show that any G with this property is minimized by any measure δ_{x_0}, with $x_0 \in \Omega$. We list here some functionals satisfying this definition:

- $G(\nu) = \sharp\, \mathrm{spt}(\nu)$ (*atomic*), as in location problems, where the corresponding T is usually of Wasserstein type;
- (*subadditive*, see [16])

$$G(\nu) = \begin{cases} \sum_{k\in\mathbb{N}} g(a_k) & \text{if } \nu = \sum_{k\in\mathbb{N}} a_k \delta_{x_k} \\ +\infty & \text{otherwise} \end{cases}$$

  for a subadditive function g with $g(0) = 0$ and $g'(0) = +\infty$ (if $g = 1$ on $(0, +\infty)$ and $g(0) = 0$ we recover the previous case);
- $G(\nu) = \inf \left\{ \mathcal{H}^1(\Sigma) \,|\mathrm{spt}(\nu) \subset \Sigma,\, \Sigma \text{ compact and connected } \right\}$ as in the irrigation problem (see [27] and Chapter 8), where $T = W_1$ and a constraint on G is considered (*length*);
- $G(\nu) = \int_{\Omega\times\Omega} w(|x - y|)\nu(dx)\nu(dy)$ for an increasing function w (*interaction*).

Actually the two first cases are functionals which decrease under the effect of any map t and not only under $1-$Lipschitz ones. The first one has been presented separately and not as a particular case of the second because of its large presence in literature. The last choice for G is a well-known functional on probability measures which represents the interaction energy (or cost) between the particles composing ν. It has been first studied in a transportation framework by McCann in[58], where displacement convexity results are given, with the aim of showing uniqueness results for variational problems.

In the two previous chapters two combinations of these functional have been studied for urban planning purposes: the *Wasserstein* + *subadditive* and the *congestion* + *interaction* cases, respectively. The *congestion* + *subadditive* case has been excluded in Chapter 2 since it leads to a trivially infinite functional, and so in this chapter we analyze the remaining one, i.e. the *Wasserstein* + *interaction* case. Many ideas are in common with Chapter 2, up to the fact that elliptic regularity is replaced by considerations on Monge-Ampère equation. Moreover some extra devices are performed and a careful use of Monge-Kantorovich theory is needed.

3.2 Optimality conditions for the interaction case

We are here concerned with the minimization problem for the functional $\mathfrak{F}^\mu$, when the transport term is given by $T(\mu,\nu) = \frac{1}{2}W_2^2(\mu,\nu)$ and the concentration one is an interaction term of the form

$$G(\nu) = \int_{\Omega\times\Omega} V(|x-y|^2)\nu(dx)\nu(dy), \qquad (3.2.1)$$

with $V : [0,+\infty[\to [0,+\infty[$ a regular increasing function. From now on Ω will be a convex domain in $\mathbb{R}^d$.

A priori, a minimizer for this functional may be an arbitrary probability measure on the set Ω, even a singular one. Our goal is to prove, under suitable assumptions and by means of optimality conditions and of an approximation process, that it is in fact an absolutely continuous measure with bounded density.

We provide here an easy optimality condition for the minimization of $\mathfrak{F}^\mu$. We do not go into details in the computation because it follows the same scheme as in [28]. The approximation by measures with positive densities that we are going to use works in this case too, while the alternative proof by convex analysis of Subsection 1.4.2 does not, simply because there is no convexity in the term G.

Theorem 3.2.1. *If a probability measure $\nu \in \mathcal{P}(\Omega)$ is a minimizer for $\mathfrak{F}^\mu$, then there exists a constant m such that*

$$\psi + T_\nu \geq m; \quad \psi + T_\nu = m \quad \nu\text{-a.e.},$$

where ψ is a Kantorovich potential for the transport from ν to μ and we define

$$T_\nu(x) = \int_\Omega 2V(|x-y|^2)\,\nu(dy).$$

Proof. Let us start from the case when μ is absolutely continuous with positive density. In this case we perform convex variations on an optimal measure ν of the form $\nu_t = \nu + t(\nu_1 - \nu)$ for an arbitrary $\nu_1 \in \mathcal{P}(\Omega)$: if we call ψ_t the unique Kantorovich potential from ν_t to μ which vanishes at a certain fixed point $x_0 \in \Omega$, we get (by means of Duality Formula)

$$\begin{aligned}\int_\Omega \psi_t\,d\nu_t + \int_\Omega \psi_t^c\,d\mu + \int_{\Omega\times\Omega} & V(|x-y|^2)\nu_t(dx)\nu_t(dy)\\ &\geq \int_\Omega \psi_t\,d\nu + \int_\Omega \psi_t^c\,d\mu + \int_{\Omega\times\Omega} V(|x-y|^2)\nu(dx)\nu(dy).\end{aligned}$$

After erasing the term $\int_\Omega \psi_t^c\,d\mu$ and dividing by t we pass to the limit, and we know that ψ_t converges uniformly (by Ascoli-Arzelà) to the unique Kantorovich potential ψ from ν to μ vanishing at x_0. This provides, at the limit,

$$\int_\Omega (\psi + T_\nu)\,d\nu_1 \geq \int_\Omega (\psi + T_\nu)\,d\nu.$$

Being ν_1 arbitrary we get that $\nu-$a.e. the function $\psi + T_\nu$ must be equal to its infimum, and this is the thesis.

To generalize the result to an arbitrary measure μ, just proceed by approximation. This can be performed as in [28] and provides the same formula where ψ becomes one of the possibly many Kantorovich potentials instead of the only one. The main difference between this case and the case of a measure μ with positive density is in fact the lack of uniqueness (even up to additive constants) of the Kantorovich potential. □

The problem in the condition of Theorem 3.2.1 lies in the fact that the measure ν appears only in a very implicit way (both in ψ and in T_ν), and this does not allow to derive any estimate on it. We will consequently need to pass through an approximation process, exactly as in Chapter 2. Fixed a minimizer $\overline{\nu}$ for $\mathfrak{F}^{\mu}$, we will consider a sequence of problems $(P_\varepsilon)_\varepsilon$ given by the minimization of

$$\mathcal{P}(\Omega) \ni \nu \mapsto T(\mu_\varepsilon, \nu) + G(\nu) + \delta_\varepsilon A(\nu) + \varepsilon W_2^2(\nu, \overline{\nu}_\varepsilon),$$

where

- $(\mu_\varepsilon)_\varepsilon$ is a sequence of probability measures approximating μ with Lipschitz continuous strictly positive densities u_ε;
- the functional A is given by

$$A(\nu) = \begin{cases} \int_\Omega a(v)\, d\mathcal{L}^d & \text{if } \nu = v \cdot \mathcal{L}^d, \\ +\infty & \text{otherwise,} \end{cases}$$

  for a convex function $a : [0, +\infty[\to [0, +\infty]$ which is superlinear at infinity and blowing up at 0, i.e. $\lim_{t\to 0^+} a(t) = +\infty$, but finite and C^2 with $a'' \geq c > 0$ on $]0, +\infty[$ (for instance $a(t) = t^2 + 1/t$);
- $(\delta_\varepsilon)_\varepsilon$ is a suitably chosen sequence with $\delta_\varepsilon > 0$ and $\delta_\varepsilon \to 0$.
- $(\overline{\nu}_\varepsilon)_\varepsilon$ is a suitably chosen sequence of measures with $\overline{\nu}_\varepsilon \rightharpoonup \overline{\nu}$.

We will prove, exactly as in Chapter 2, that this sequence of problems admits an uniform L^∞ bound for their solutions and that we can choose the parameters so that these solutions converge to $\overline{\nu}$, thus obtaining an L^∞ estimate for $\overline{\nu}$. The existence of solutions for P_ε is trivial by the semicontinuity of each term in the sum with respect to the weak convergence of probability measures on the compact set Ω (and moreover any term but A is actually continuous, while A is semicontinuous by convexity).

Lemma 3.2.2. *Suppose that $\mu = u \cdot \mathcal{L}^d$ with $||u||_{L^\infty} \leq M$ and that V is a C^2 function with $V' > 0$. Then any solution ν_ε to the problem P_ε is absolutely continuous and its density is bounded by a constant C depending only on M, d and V.*

Proof. First we notice that, thanks to the presence of the term $A(\nu)$ in the minimization problem, ν_ε must be absolutely continuous with strictly positive density almost everywhere. Then we write the optimality conditions for ν_ε with respect to variations of the form $\nu_t = \nu_\varepsilon + t(\tilde{\nu} - \nu_\varepsilon)$. From easy computations we get

$$\psi_\varepsilon + T_{\nu_\varepsilon} + \delta_\varepsilon a'(\nu_\varepsilon) + \varepsilon \chi_\varepsilon = m_\varepsilon \text{ a.e.,}$$

where ψ_ε is the Kantorovich potential for the transport from ν_ε to μ_ε and χ_ε from ν_ε to $\overline{\nu}_\varepsilon$ (they are unique up to additive constants) and m_ε is a suitable constant. We get equality almost everywhere due to the fact that we already know $\nu_\varepsilon > 0$ (we identify measures and their densities in this context). Since Kantorovich potentials are Lipschitz functions and T_{ν_ε} shares the same regularity of the integrand $(x,y) \mapsto V(|x-y|^2)$, which is C^2 and then Lipschitz, we get that even $a'(\nu_\varepsilon)$ is Lipschitz continuous, and in particular it is bounded. This prevents ν_ε to be close to 0 since it holds $\lim_{t\to 0^+} a'(t) = -\infty$. Thus we get $\nu_\varepsilon \geq c_\varepsilon > 0$. Moreover, $a'(\nu_\varepsilon)$ is Lipschitz continuous and, being a'' bounded from below by a positive constant, also the inverse of a' is Lipschitz. This proves that ν_ε is a Lipschitz continuous function. We can now use regularity theory on Monge-Ampère equation to get $\psi \in C^{2,\alpha}(\Omega) \cap C^{1,\alpha}(\overline{\Omega})$, since both ν_ε and μ_ε are bounded both from above and from below by positive constants (depending on ε, anyway) and are Lipschitz continuous. The same is true for the Kantorovich potential χ_ε by replacing μ_ε by $\overline{\nu}_\varepsilon$. What we can do now is looking for a maximum point x_0 of ν_ε. Notice that such a point will be a minimum point for $\psi_\varepsilon + T_{\nu_\varepsilon} + \varepsilon\chi_\varepsilon$. First we prove that $x_0 \notin \partial\Omega$. To prove this it is sufficient to prove that the gradient of $\psi_\varepsilon + T_{\nu_\varepsilon} + \varepsilon\chi_\varepsilon$ is directed outwards at any point of $\partial\Omega$, i.e. $\nabla(\psi_\varepsilon + T_{\nu_\varepsilon} + \varepsilon\chi_\varepsilon)(x_0) \cdot n(x_0) > 0$ for any $x_0 \in \partial\Omega$, where n is the outward normal vector. From the fact that the optimal transport map t from ν_ε to μ_ε is given by $t(x) = x - \nabla\psi(x)$ we know that $x - \nabla\psi(x) \in \Omega$ for almost any $x \in \Omega$ (see Figure 3.1). In this case, due to continuity up to the boundary of $\nabla\psi$, this holds for any x and also for $x_0 \in \partial\Omega$ and implies $\nabla\psi(x_0) \cdot n(x_0) \geq 0$. Analogously we get $\nabla\chi(x_0) \cdot n(x_0) \geq 0$. For the gradient of T_{ν_ε} it holds

$$\nabla T_{\nu_\varepsilon}(x_0) = \int_\Omega 4V'(|x_0 - y|^2)(x_0 - y)\,\nu_\varepsilon(dy),$$

and so $\nabla T_{\nu_\varepsilon}(x_0) \cdot n(x_0) > 0$ since $V' > 0$ and for almost any $y \in \Omega$ it holds $(x_0 - y) \cdot n(x_0) > 0$. This proves that x_0 lies in the interior of Ω and this allows us to look at the second derivatives. Taking Hessians we have

$$H\psi_\varepsilon(x_0) + HT_{\nu_\varepsilon}(x_0) + \varepsilon H\chi_\varepsilon(x_0) \geq 0,$$

where the letter H denotes Hessians and the inequality is in the sense of positive definite symmetric matrices. Thus we get

$$H\psi_\varepsilon(x_0) \geq -I\left(2||V||_{C^2(\Omega)} + \varepsilon\right),$$

since the second derivatives of T_{ν_ε} may be estimated by those of V and from the fact that $x^2/2 - \chi(x)$ is convex we deduce $H\chi \leq I$. This is a uniform estimate from below for $H\psi_\varepsilon(x_0)$ and for the convex function ϕ given by $\phi(x) = x^2/2 - \psi_\varepsilon(x)$ we have $H\phi(x_0) \leq I\left(1+\varepsilon+2||V||_{C^2(\Omega)}\right)$. This implies $M\phi(x_0) \leq (1+\varepsilon+2||V||_{C^2(\Omega)})^d$, and, from $\nu_\varepsilon = \mu_\varepsilon(\nabla\phi)M\phi$, we get, for $\varepsilon \leq 1$,

$$\max \nu_\varepsilon = \nu_\varepsilon(x_0) \leq 2^d M \left(1+||V||_{C^2(\Omega)}\right)^d,$$

which is the desired estimate. □

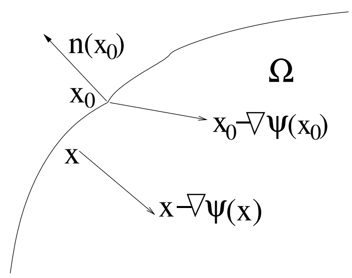

Figure 3.1: Behavior of $\nabla(\psi)$ near $\partial\Omega$

Remark 3.2.3. The proof above of the fact that the gradient is directed outwards (illustrated in figure 3.1 as well) and no maximum point is allowed on the boundary could be used similarly in Chapter 2, thus getting rid of the strict convexity assumption in Theorem 6.5 and of the heavy proof of Lemma 6.6. Notice that it could be possible to get the same result even without C^1 regularity for the potentials, just making the proof a bit trickier. It would be sufficient to evaluate the increments of the potential in small balls around x_0 where the gradient is almost everywhere defined and such that $x - \nabla\psi(x),\ x - \nabla\chi(x) \in \Omega$ a.e. Also the estimate of the last term in Theorem 2.5.9 could be obtained, but it is not immediate because of the non pointwise differentiability of the Kantorovich potential.

Lemma 3.2.4. *It is possible to choose the parameters for the problem P_ε, i.e. the numbers δ_ε and the measures $\overline{\nu}_\varepsilon$ and μ_ε so that any sequence of minimizers $(\nu_\varepsilon)_\varepsilon$ converges to $\overline{\nu}$.*

Proof. It is sufficient to choose for $\overline{\nu}_\varepsilon$ a sequence of absolutely continuous measures with Lipschitz continuous strictly positive densities such that $\mathfrak{F}^\mu(\overline{\nu}_\varepsilon) \leq \mathfrak{F}^\mu(\overline{\nu}) + \varepsilon^2$. Then we have $A(\overline{\nu}_\varepsilon) < +\infty$ and we may choose

$\delta_\varepsilon = \varepsilon^2 A(\overline{\nu}_\varepsilon)^{-1}$. For $(\mu_\varepsilon)_\varepsilon$ we can chose any sequence of absolutely continuous measures with Lipschitz continuous strictly positive densities approximating μ in such a way that $W_2(\mu_\varepsilon, \mu) \leq \varepsilon^2$. Then we have

$$T(\mu_\varepsilon, \nu_\varepsilon) + G(\nu_\varepsilon) + \delta_\varepsilon A(\nu_\varepsilon) + \varepsilon W_2^2(\nu_\varepsilon, \overline{\nu}_\varepsilon) \leq T(\mu_\varepsilon, \overline{\nu}_\varepsilon) + G(\overline{\nu}_\varepsilon) + \delta_\varepsilon A(\overline{\nu}_\varepsilon),$$

which implies

$$\begin{aligned} \mathfrak{F}^\mu(\nu_\varepsilon) + \delta_\varepsilon A(\nu_\varepsilon) + \varepsilon W_2^2(\nu_\varepsilon, \overline{\nu}_\varepsilon) &\leq \mathfrak{F}^\mu(\overline{\nu}_\varepsilon) + 4DW_2(\mu_\varepsilon, \mu) + \delta_\varepsilon A(\overline{\nu}_\varepsilon) \\ &\leq \mathfrak{F}^\mu(\overline{\nu}) + \varepsilon^2 + 4D\varepsilon^2 + \varepsilon^2 \\ &\leq \mathfrak{F}^\mu(\nu_\varepsilon) + \varepsilon^2(2 + 4D). \end{aligned}$$

Finally, this implies $W_2(\nu_\varepsilon, \overline{\nu}_\varepsilon) \leq C\sqrt{\varepsilon}$ and, since $\overline{\nu}_\varepsilon \rightharpoonup \overline{\nu}$, we get $\nu_\varepsilon \rightharpoonup \overline{\nu}$. □

Remark 3.2.5. This is the point where global optimality of $\overline{\nu}$ plays a crucial role. In fact, should $\overline{\nu}$ be only locally minimizing, we could not use the inequality $\mathfrak{F}^\mu(\overline{\nu}) \leq \mathfrak{F}^\mu(\nu_\varepsilon)$, unless we already know that ν_ε is in the domain of minimality of $\overline{\nu}$, i.e. sufficiently close to it.

We can now state our main result and its consequence in the minimization of the whole functional $\mathfrak{F}$.

Theorem 3.2.6. *Given a compact convex set $\Omega \subset \mathbb{R}^d$ with nonempty interior and a probability measure $\mu \in L^\infty(\Omega)$, if the function V appearing in the definition of the functional G is of class C^2 and $V' > 0$, then the minimization problem for the functional $\mathfrak{F}^\mu$ over the space $\mathcal{P}(\Omega)$ admits at least a solution and any minimizer belongs in fact to the space $L^\infty(\Omega)$.*

Proof. As usual the existence is trivial due to continuity and compactness of $\mathcal{P}(\Omega)$ while, for the L^∞ regularity, just apply Lemma 3.2.2 and Lemma 3.2.4 □

Corollary 3.2.7. *Given a compact convex set $\Omega \subset \mathbb{R}^d$ with nonempty interior, a C^1 strictly convex and superlinear function f with polynomial growth and a C^2 function V with $V' > 0$, then the minimization problem over the space $\mathcal{P}(\Omega)^2$ for the functional $\mathfrak{F}(\mu, \nu) = \frac{1}{2}W_2^2(\mu, \nu) + F(\mu) + G(\nu)$, where F is defined by* (3.1.1) *and G by* (3.2.1)*, admits a solution and, in any minimizing pair (μ, ν), both μ and ν are in fact absolutely continuous measures $\mu = u \cdot \mathcal{L}^d$, $\nu = v \cdot \mathcal{L}^d$, with $u \in C^0(\Omega)$ and $v \in L^\infty(\Omega)$.*

Proof. After the usual proof of existence by the direct method in Calculus of Variations, we refer to chapter 1 for the regularity results on μ. Since μ turns out to be absolutely continuous with continuous density (hence bounded), we may apply Theorem 3.2.6 to get the regularity on ν. □

3.3 An explicit example

In this section we come back to the whole problem of minimizing the functional $\mathfrak{F}$ in a very particular case, where we can provide almost explicit densities for the solutions. We consider the case

- $T(\mu,\nu) = \frac{1}{2}W_2^2(\mu,\nu)$ and $G(\nu) = \int_{\Omega\times\Omega} V(|x-y|^2)\nu(dx)\nu(dy)$, as in the previous Section;
- $V(|x-y|^2) = \frac{\lambda}{2}|x-y|^2$ and so, setting $\mathrm{bar}(\nu) = \int_\Omega y\,\nu(dy)$, we have $T_\nu(x) = \lambda|x|^2 - 2\lambda x\cdot\mathrm{bar}(\nu) + \lambda\int_\Omega |y|^2\nu(dy)$;
- $F(\mu) = \frac{1}{2}||\mu||^2_{L^2(\Omega)}$, a particular case of what considered in Chapter 1.

The framework we obtain is very similar to the one in Chapter 2.

Theorem 3.3.1. *In the specific case described above, any pair of minimizers (μ,ν) is shaped as follows:*

- *μ is concentrated on a ball $B(x_0, r_\lambda)$ (intersected with Ω) and has a density u given by*

$$u(x) = \frac{\lambda}{2\lambda+1}(r_\lambda^2 - |x-x_0|^2);$$

- *ν is concentrated on the ball $B(x_0, r_\lambda/(2\lambda+1))$ and it is the image of μ under the homothety of ratio $(2\lambda+1)^{-1}$ and center x_0, hence it has density v given by*

$$v(x) = \lambda(2\lambda+1)^{d-1}(r_\lambda^2 - (2\lambda+1)^2|x-x_0|^2);$$

- *x_0 is the barycenter of both μ and ν.*

Proof. First we write down the optimality conditions given by Theorem 3.2.1 for the minimization in ν with fixed μ and by Chapter 1 for the minimization in μ for fixed ν. We denote by u and v the densities of μ and ν, respectively. We may suppose that the barycenter of ν is the origin, thus obtaining $T_\nu(x) = \lambda|x|^2 + c$. We have

$$\begin{cases} u(x) + \varphi(x) = c_1 & \text{a.e. on } u > 0; \\ \psi(x) + \lambda x^2 = c_2 & \text{a.e. on } v > 0. \end{cases}$$

Here φ and ψ are Kantorovich potentials for the transport from μ to ν and from ν to μ, respectively. From the second condition we can infer

$\nabla\psi(x) = -2\lambda x$ a.e. on $v > 0$. Being ν absolutely continuous, this equality is valid $\nu-$a.e.. This means that the optimal transport map T from ν to μ is given by $T(x) = x - \nabla\psi(x) = (2\lambda + 1)x$. By uniqueness of the optimal transport plan, which is in this case expressed both as a transport map from ν to μ and vice versa, we know also the optimal transport map S from μ to ν which will be $S(x) = x/(2\lambda + 1)$. From duality theory in optimal transportation we know the following equality

$$\varphi(x) + \varphi^c(S(x)) = c(x, S(x)) = \frac{1}{2}|x - S(x)|^2,$$

and thus we get $u(x) = c_1 - \frac{1}{2}|x - S(x)|^2 + \varphi^c(S(x))$. We do already know that u is Lipschitz continuous (by [28]), and this implies that the set $\{u > 0\}$ is an open set. Consequently the same is true for $\{v > 0\}$, which is just an homothety of it. Since φ^c is a Kantorovich potential from ν to μ, we know that it must agree (up to constants) with ψ on any connected component of the open set $\{v > 0\}$. So, let $\omega \subset \Omega$ be a connected component of $\{u > 0\}$. On $(2\lambda + 1)^{-1}\omega$ we have $\varphi^c = \psi + c_3$ and so we get

$$u(x) = c_4 - \frac{1}{2}|x - S(x)|^2 + \psi(S(x)) = c_5 - |x|^2\frac{\lambda}{2\lambda + 1}.$$

From this expression it is clear that $\partial\omega \setminus \partial\Omega$ (where u must vanish) is contained in a sphere around 0. This implies that 0 belongs in fact to ω, since no boundary of ω is allowed in the interior of a certain ball around 0. So there is in fact just one connected component for $\{u > 0\}$ and so we get

$$u(x) = \left[c - |x|^2\frac{\lambda}{2\lambda + 1}\right]^+. \tag{3.3.1}$$

From this it is easy to recover the density v of ν since $\nu = S_\sharp\mu$ and we get the thesis. The point x_0 which turns out to be the center of the balls which are supports for μ and ν is in this notation 0, the barycenter of ν, as in the thesis. It is clear that in this case μ and ν share the same barycenter since they are homothetic. □

Remark 3.3.2. In the example of Theorem 3.3.1 the density v shares the same regularity of u except at the points corresponding to boundary points of Ω where u is positive, i.e. if at $x_0 \in \partial\Omega$ it happens $u(x_0) > 0$ then at $S(x_0)$ we have a jump for v. It is clear from the fact that u is $2\lambda/(2\lambda+1)-$Lipschitz continuous (it follows from the explicit formula) that we have, recalling also $\int_\Omega u\, d\mathcal{L}^d = 1$,

$$1 \leq \left(\inf u + \frac{2\lambda}{2\lambda + 1}D\right)|\Omega|,$$

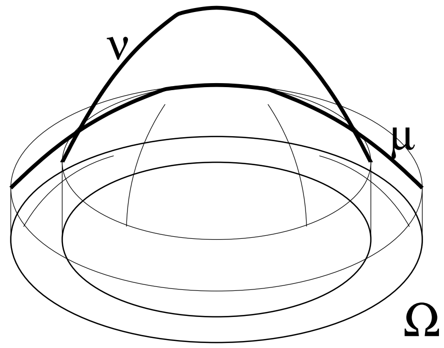

Figure 3.2: The solution for a small ball Ω

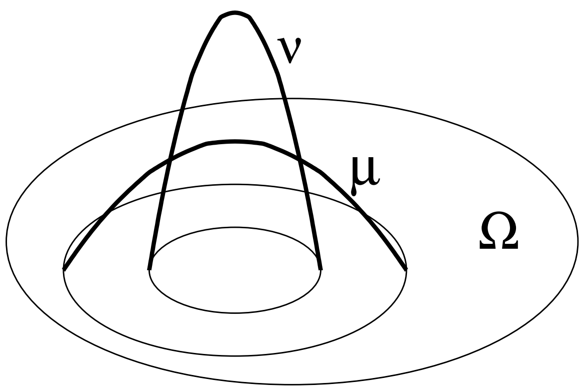

Figure 3.3: A solution for a larger ball

where D is the diameter of Ω. This implies, for small Ω, $\inf u > 0$. In this case u would be positive at any point of $\partial\Omega$ and v discontinuous at any point of $S(\partial\Omega)$. This gives examples when the L^∞ regularity for v cannot be improved up to $v \in C^0(\Omega)$.

Remark 3.3.3. In the explicit example above there remain to be determined both the constant c (or the radius r_λ) and the position of the barycenter x_0 in the formula for u. In some simple cases this is possible too. Notice that, once fixed x_0, the constant c may always be recovered by imposing the condition of being probability measures. For instance if Ω is a ball, we may see that x_0 may not be the barycenter of a density u shaped as in (3.3.1) unless the set $B(x_0, 2\lambda^{-1}(2\lambda+1)) \cap \Omega$ is a ball around x_0. This happens for large Ω whenever the ball $B(x_0, 2\lambda^{-1}(2\lambda+1))$ does not touch the boundary $\partial\Omega$ or, in general, when x_0 is the center of the ball Ω. In the first case (Ω a large ball, as in Figure 3.3, where the case of a generic Ω is represented) we have several solutions for the problem (non-uniqueness), obtained from each other under translations, and u and v are continuous; in the second (Ω a small ball, Figure 3.2) we have uniqueness of the solution, with u a radial continuous function around the center and v a rescaled version of u on a smaller ball.

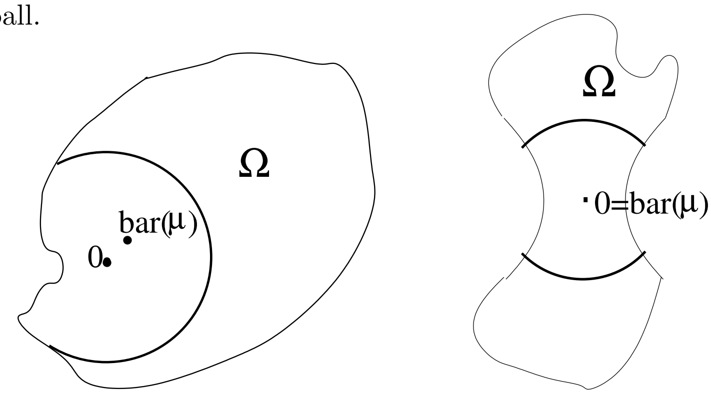

Figure 3.4: The position of the center and of the barycenter

Remark 3.3.4. In general, if Ω is not a ball, the fact that 0 is the barycenter of a distribution of mass which is radial around 0 itself imposes some constraints on the position of 0 with respect to $\partial\Omega$. If the domain Ω cuts a part of the supporting ball from one side, then the center of the ball could be no longer the barycenter. Figure 3.4 shows this effect, as well as a situation where the support touches the boundary on two sides and the center of the ball is actually the barycenter.

Chapter 4

Path functionals in Wasserstein spaces

In this chapter we consider the problem of finding optimal interpolations between two given distributions of masses in order to satisfy some concentration criteria. This is formalized by means of curves in the space of probability measures, endowed with a Wasserstein metric. This distance is then perturbed into a kind of conformal Riemannian metric by means of some penalization encouraging or discouraging concentration. The functional we use to deal with these penalization are the same as in Chapter 1. The starting goal was to give a geodesic approach (curves optimizing a certain perturbed length in a suitable metric space) to some branching transport problems arising in several applications by means of a functional encouraging the curve to pass through concentrated atomic measures, and the case of functionals discouraging concentrations is presented only as a natural counterpart. The results are in fact somehow specular. This chapter follows essentially what presented in a published joint paper with Alessio Brancolini and Giuseppe Buttazzo, [20], up to some preliminary notions which are presented in Sections 0.2 and 0.3. As in [20], we are mainly concerned with existence results and conditions ensuring that the optimal solution has a finite energy. In particular, we will not give necessary optimality conditions and the spirit is hence different with respect to what we did in the previous chapters. Anyway, in the diffusion case (when the energy discourages concentration), as the optimal curve passes through absolutely continuous measures, it is interesting to write down optimality conditions on their densities. These conditions have the form of a system of PDEs and are studied in Chapter 5.

4.1 The metric framework

In this section a generic metric space X with distance d is considered. Under the assumption that closed bounded subsets of X are compact, we will prove an existence result (Theorem 4.1.1) for variational problems with functionals of the type

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|(t)\ dt$$

where $\gamma : [0,1] \to X$ ranges among all Lipschitz curves such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$. We will refer to the value of $\mathcal{J}$ in γ as the *energy* of γ. By $|\gamma'|(t)$ we denote the *metric derivative* of γ at the point $t \in (0,1)$, that we introduced in Section 0.3 and which exists a.e. for Lipschitz curves. Another useful result is that the variation (length) of γ can be written in terms of the metric derivative in integral form:

$$\text{Length}(\gamma) = \int_0^1 |\gamma'|(t)\ dt.$$

By this formula it follows easily that $|\gamma'| \leq M$ a.e. if and only if γ is M-Lipschitz, since when $s < t$

$$d(\gamma(t), \gamma(s)) \leq \text{Length}(\gamma, [s,t]) = \int_s^t |\gamma'|(\tau)\ d\tau \leq M|t-s|,$$

the converse implication being immediate.

Theorem 4.1.1. *Let X be a metric space such that any closed bounded subset of X is compact, $J : X \to [0, +\infty]$ be a lower semicontinuous function and x_0, x_1 arbitrary points in X. Then the functional*

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|(t)\ dt$$

achieves a minimum value among all Lipschitz curves $\gamma : [0,1] \to X$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$, provided the following two assumptions are satisfied:

(H1): there exists a curve γ_0 such that $\mathcal{J}(\gamma_0) < +\infty$;

(H2): $\int_0^\infty \left(\inf_{B(x_0,r)} J\right)\ dr = +\infty$.

The proof of Theorem 4.1.1 relies on the following reparameterization lemma whose proof can be found for example in [7].

Lemma 4.1.2. *Let $\gamma \in \mathrm{Lip}([0,1], X)$ and $L = \mathrm{Var}(\gamma)$ be its total variation. Then there exists a Lipschitz curve $\tilde{\gamma} \in \mathrm{Lip}([0,L], X)$ such that $|\tilde{\gamma}'| = 1$ almost everywhere in $[0,L]$ and $\tilde{\gamma}$ is a parametrization of γ.*

Proof of Theorem 4.1.1. Let $(\gamma_n)_n$ be a minimizing sequence and set $L_n = \mathrm{Var}(\gamma_n)$. The sequence $(\mathcal{J}(\gamma_n))_n$ is bounded by a finite number M. By Lemma 4.1.2 there exists a sequence of curves $\tilde{\gamma}_n : [0, L_n] \to X$ parametrized with unit velocity, reparametrizing the given curves. We have:

$$M \geq \mathcal{J}(\gamma_n) = \int_0^{L_n} J(\tilde{\gamma}_n(t))\, dt \geq \int_0^{L_n} \left(\inf_{B_t(x_0)} J \right)\, dt.$$

Then $(L_n)_n$ is bounded otherwise, by assumption H2, the right hand side would be unbounded. We can reparameterize each curve γ_n at constant speed L_n, thus obtaining a new sequence $(\hat{\gamma_n})_n$ in $\mathrm{Lip}([0,1], X)$, which is still a minimizing sequence, thanks to the equality $\mathcal{J}(\gamma_n) = \mathcal{J}(\hat{\gamma_n})$. Being $(L_n)_n$ bounded, we get that this new minimizing sequence is uniformly bounded and uniformly Lipschitz. By Ascoli-Arzelà Theorem we can suppose, up to a subsequence, that we have uniform convergence $\hat{\gamma}_n \to \hat{\gamma}$ for some L-Lipschitz curve $\hat{\gamma}$ where we have taken $L = \liminf_n L_n$. By recalling the link between Lipschitz conditions and metric derivative we have

$$|\hat{\gamma}'|(t) \leq L \quad \text{for a.e. } t \in [0,1].$$

Now by using the lower semicontinuity of the functional J, we obtain

$$\begin{aligned} \mathcal{J}(\hat{\gamma}) = \int_0^1 J(\hat{\gamma}(t))|\hat{\gamma}'|(t)\, dt &\leq L \int_0^1 \liminf_{n\to+\infty} J(\hat{\gamma}_n(t))\, dt \\ &\leq \liminf_{n\to+\infty} L_n \int_0^1 J(\hat{\gamma}_n(t))\, dt = \liminf_{n\to+\infty} \mathcal{J}(\hat{\gamma}_n), \end{aligned}$$

that is the lower semicontinuity of $\mathcal{J}$ on the considered sequence, which achieves the proof. □

Remark 4.1.3. Notice that the integral assumption H2 is always true if $J \geq c$ for a suitable strictly positive constant. Moreover Theorem 4.1.1 still holds if condition H2 is replaced by the weaker assumption that there exists a curve γ_0 such that

$$\mathcal{J}(\gamma_0) < \int_0^{+\infty} \inf_{B(x_0,r)} J\, dr.$$

We give here a slightly refined version of Theorem 4.1.1, which will be useful in the last section. The goal here is to weaken the compactness assumption on bounded subsets of X.

Theorem 4.1.4. *Let (X, d, d') be a metric space endowed with two distances, such that:*

(K1): $d' \leq d$;

(K2): all $d-$bounded sets in X are relatively compact with respect to d';

(K3): the mapping $d : X \times X \to \mathbb{R}^+$ is a lower semicontinuous function with respect to the distance $d' \times d'$.

Let $J : X \to [0, +\infty]$ be lower semicontinuous with respect to d'. Consider the functional, defined on the set of $d-$Lipschitz curves $\gamma : [0, 1] \to X$, given by

$$\mathcal{J}(\gamma) = \int_0^1 J(\gamma(t))|\gamma'|_d(t)dt,$$

where $|\gamma'|_d(t)$ stands for the metric derivative of γ with respect to d. Then, with the same hypotheses H1 and H2 (where the balls $B(x_0, r)$ are balls in the d-sense) of Theorem 4.1.1, there exists a minimum for $\mathcal{J}$.

Proof. We can take a minimizing sequence $(\gamma_n)_n$ and, as in Theorem 4.1.1, reparameterize it to obtain a sequence $(\hat{\gamma}_n)_n$ in which every curve has constant speed L_n. Hypothesis H2 gives us the boundedness of L_n. Hence the sequence $(\hat{\gamma}_n)_n$ is composed by $d-$equicontinuous functions from $[0, 1]$ to a $d-$bounded subset of X. If we endow X with the distance d' we have an equicontinuous (thanks to assumption K1) sequence of functions whose images are contained in a compact set. We can consequently use Ascoli-Arzelà Theorem to choose a subsequence (not relabeled), such that $\hat{\gamma}_n \to \gamma$, for a suitable curve γ (uniformly in the $d'-$sense). The lower semicontinuity of J with respect to d' allows us to use Fatou Lemma and shows that γ minimizes $\mathcal{J}$, as far as we can show that γ is $d-$Lipschitz with a Lipschitz constant not exceeding $\liminf_n L_n$. To do this we use assumption K3. Taken two points s, t we have in fact:

$$d(\gamma(s), \gamma(t)) \leq \liminf_n d(\hat{\gamma}_n(s), \hat{\gamma}_n(t)) \leq \liminf_n L_n|s - t|,$$

which shows the required Lipschitz property. □

4.2 The case of the space of probability measures

In this section we will choose X to be a space of probability measures on a domain $\Omega \subset \mathbb{R}^d$. To endow it with a distance we will consider actually the Wasserstein space $\mathcal{W}_p(\Omega)$. See Section 0.2 for the basic notions on this space. Once fixed that the elements of the space X will be measures, we will choose the functional J to be a local semicontinuous functional, according to the general theory briefly presented in Section 1.2. Hence J will be of the form

$$J(\mu) = \int_\Omega f\left(\frac{d\mu}{d\mathrm{m}}\right)\, d\mathrm{m} + f^\infty(1)|\mu^s|(\Omega \setminus A_\mu) + \int_{A_\mu} g(\mu(\{x\}))\, d\#(x).$$

To satisfy the assumptions of Theorem 4.1.1, according to Remark 4.1.3, we will prove the following, easy, general estimate, in the case $\mathrm{m}(\Omega) < +\infty$.

Theorem 4.2.1. *Suppose that $f(s) > 0$ for $s > 0$ and $g(1) > 0$. Then we have $J \geq c > 0$. In particular, the assumption H2 is verified.*

Proof. Let us fix some notation. By μ^a we mean the absolutely continuous part of μ with respect to the measure m, and by $\mu^s, \mu^\#, \mu^c$ respectively the singular part, the atomic part and the singular diffused part of μ. Then we have $\mu = \mu^a + \mu^s = \mu^a + \mu^c + \mu^\#$. Since f is convex, by Jensen inequality we have

$$\int_\Omega f\left(\frac{d\mu}{d\mathrm{m}}\right)\, d\mathrm{m} \;\geq\; \mathrm{m}(\Omega) f\left(\frac{1}{\mathrm{m}(\Omega)}\int_\Omega \frac{d\mu}{d\mathrm{m}}\, d\mathrm{m}\right) \;=\; \mathrm{m}(\Omega) f\left(\frac{\mu^a(\Omega)}{\mathrm{m}(\Omega)}\right).$$

Since μ is a positive measure and f^∞ is $1-$homogeneous

$$\int_{\Omega\setminus A_\mu} f^\infty\left(\frac{d\mu^s}{d|\mu^s|}\right)\, d|\mu^s| \;=\; |\mu^s|(\Omega \setminus A_\mu) f^\infty(1) \;=\; \mathrm{m}(\Omega) f^\infty\left(\frac{\mu^c(\Omega)}{\mathrm{m}(\Omega)}\right).$$

Since g is a subadditive function

$$\int_{A_\mu} g(\mu(\{x\}))\, d\#(x) = \sum_{x\in A_\mu} g(\mu(\{x\})) \geq g\left(\sum_{x\in A_\mu} \mu(\{x\})\right) = g(\mu^\#(\Omega)).$$

For the recession function f^∞ we have (thanks to its equivalent definition by means of a sup)

$$f^\infty(x) \geq f(x+y) - f(y) \text{ for all } x, y \in \mathbb{R},$$

and so the sum of the first two terms can be estimated from below by

$$\mathrm{m}(\Omega) f\left(\frac{\mu^a(\Omega) + \mu^c(\Omega)}{\mathrm{m}(\Omega)}\right).$$

Therefore, summing up we obtain

$$J(\mu) \geq \mathrm{m}(\Omega) f\left(\frac{\mu^a(\Omega)+\mu^c(\Omega)}{\mathrm{m}(\Omega)}\right) + g(\mu^{\#}(\Omega)).$$

We set $a = \mu^{\#}(\Omega)$ and $1-a = \mu^a(\Omega)+\mu^c(\Omega)$. Since the function $a \mapsto \mathrm{m}(\Omega)f((1-a)/\mathrm{m}(\Omega)) + g(a)$ is lower semicontinuous, it attains a minimum in the interval $[0,1]$. Thanks to our hypothesis this sum is always positive, and so we have

$$\min_{0\leq a\leq 1} \mathrm{m}(\Omega) f\left(\frac{1-a}{\mathrm{m}(\Omega)}\right) + g(a) = c > 0,$$

that is, we have $J(\mu) \geq c > 0$. □

We now study some special cases of the functional we defined above. In the rest of this section Ω will be a convex domain in $\mathbb{R}^d$ and the measure m will be the Lebesgue measure $\mathcal{L}^d$ on it.

4.2.1 Concentration

Roughly speaking, this is the case we get by setting $f = +\infty$ and $g(z) = |z|^\alpha$ for $0 \leq \alpha < 1$. More precisely we set $f = +\infty$ on $]0,+\infty[$ and $f(0) = 0$, and $g(z) = z^\alpha$. We are in a particular case of what shown as an example in Section 1.2. The functional J will be denoted by G_α and has the following form

$$G_\alpha(\mu) = \begin{cases} \sum_i a_i^\alpha & \text{if } \mu = \sum_i a_i \delta_{x_i}; \\ +\infty & \text{if } \mu \text{ is not atomic.} \end{cases}$$

The corresponding functional $\mathcal{J}$ on Lipschitz paths will be called $\mathcal{G}_\alpha$. This is the case when G_α is finite only on purely atomic measures and minimizing $\mathcal{G}_\alpha$ will look for curves which interpolate μ_0 and μ_1 by passing through concentrated atomic measures.

We are now going to consider the question whether there exists a curve connecting two given measures keeping finite our functional. First we prove that if both the initial and the final measure are atomic the answer is positive. Then we prove that for α in a suitable subinterval of $[0,1]$ every measure can be connected to a Dirac mass, hence every measure can be connected to every other measure by a path of finite energy. Finally we show that this is not possible in general for any $\alpha \in [0,1]$.

Theorem 4.2.2. *Let μ_0 and μ_1 be atomic measures, i.e.,*

$$\mu_0 = \sum_{k=1}^{\infty} a_k \delta_{x_k}, \quad \mu_1 = \sum_{l=1}^{\infty} b_l \delta_{y_l}$$

with $a_k, b_l > 0$, $\sum_k a_k = \sum_l b_l = 1$. Suppose also $G_\alpha(\mu_0) = \sum_k a_k^\alpha < +\infty$, $G_\alpha(\mu_1) = \sum_l b_l^\alpha < +\infty$. Then there exists a Lipschitz curve $\gamma : [0,1] \to \mathcal{W}_p(\Omega)$ such that $\gamma(0) = \mu_0, \gamma(1) = \mu_1$ and

$$\mathcal{G}_\alpha(\gamma) = \int_0^1 G_\alpha(\gamma(t))|\gamma'|(t)\ dt < +\infty.$$

Proof. Suppose $0 \in \Omega$: it is sufficient to prove the theorem when $\mu_0 = \delta_0$ since in the general case one connects the first measure μ_0 to δ_0, then δ_0 to the final measure μ_1. If one can keep finite the functional in both steps, then the result is proved in the general case.

In the transportation from δ_0 to μ_1 there is only one admissible transport plan $\pi \in \Pi(\delta_0, \mu_1)$. Take the curve γ^π which is the geodesic associated to this transport plan. This curve (see Theorem 0.3.4) is given by

$$\gamma^\pi(t) = \sum_{l=1}^{\infty} b_l \delta_{t y_l}.$$

Since it is the geodesic from δ_0 to μ_1 we have $|(\gamma^\pi)'|(t) = W_p(\delta_0, \mu_1)$ for a.e. t. Moreover the atoms of the measure $\gamma^\pi(t)$, for $t > 0$, have the same masses of the atoms of μ_1, and hence we have $G_\alpha(\gamma^\pi(t)) = G_\alpha(\mu_1)$. It follows

$$\mathcal{G}_\alpha(\gamma^\pi) = G_\alpha(\mu_1) W_p(\delta_0, \mu_1) < +\infty$$

and the thesis is proven. □

The proof of the next theorem is related to the one of Proposition 3.1 of [72].

Theorem 4.2.3. *Let $1 - 1/d < \alpha \leq 1$. Then given two arbitrary μ_0 and μ_1 in $\mathcal{W}_p(\Omega)$, there exists a curve joining them such that the functional $\mathcal{G}_\alpha$ is finite.*

Proof. It is sufficient to prove that every measure can be joined to a Dirac mass in an arbitrary point. We prove first the statement for $\Omega = [0,1]^d$. The *dyadic subdivision of order k* of $Q = [0,1]^d$ is given by the family of closed d-dimensional cubes $(Q_h^k)_{h \in I_k}$ where $I_k = \{1, 2, 3, \ldots, 2^k\}^d$ obtained

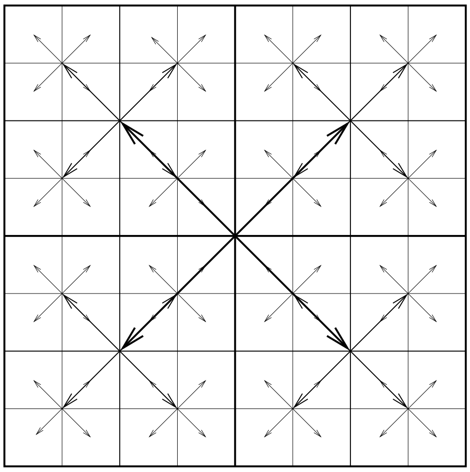

Figure 4.1: Approximation at step $k = 3$.

by Q dividing each edge into 2^k pieces of equal length. We will refer to the elements of $(Q_h^k)_{h\in I_k}$ as k-cubes. To every Borel regular finite measure μ we associate the following sequence of measures:

$$\mu_k = \sum_{h\in I_k} b_h^k \delta_{y_h}$$

where $b_h = \mu(Q_h^k)$ and y_h is the center of Q_h^k. It is straightforward to see that $\mu_k \rightharpoonup^* \mu$ as $k \to +\infty$.

The idea is now simple (see Figure 4.1): first join μ_k to μ_{k+1} with an arc length parametrization γ_k, then put together all these curves to obtain a path from a Dirac mass to the measure μ. At every step a k-cube is divided in 2^d parts which are $(k+1)$-cubes. To bring the Dirac mass in the center of the k-cube to the 2^d centers of the $(k+1)$-cubes with the right weights at each center one splits the center of the k-cube into 2^d parts moving towards the centers of the adjacent $(k+1)$-cubes in such a way that each point moves with unitary speed. At each step (see Figure 4.1 where the first three steps are represented) we obtain a curve γ_k defined on an interval of length $(1/2)^k\sqrt{d}/2$ ($\sqrt{d}$ is the diagonal of Q) such that $|\gamma_k'|(t) = 1$ for all t.

Let us now compute the value of the functional on the curve γ made by joining all curves γ_k above. Since the function $f(x_1, \dots, x_n) = \sum_i^n x_i^\alpha$ with the constraint $\sum_i^n x_i = 1$ reaches its maximum at point $(1/n, \dots, 1/n)$ we have:

$$\mathcal{G}_\alpha(\gamma) = \sum_{k=1}^{+\infty} \left(\frac{1}{2} \frac{1}{2^k} d \sum_{h \in I_k} (b_h^k)^\alpha \right) \leq \sum_{k=1}^{+\infty} \left(\frac{1}{2} \frac{1}{2^k} d 2^{dk} \left(\frac{1}{2^{dk}} \right)^\alpha \right).$$

Since $1 - 1/d < \alpha \leq 1$ the sum considered above is convergent.

In the case of a general bounded Ω it is sufficient to consider a large cube containing the support of the measure μ such that the center is contained in Ω. □

The bound given by $\alpha > 1 - 1/d$ is sharp. We have in fact the following result.

Theorem 4.2.4. *Suppose $\alpha \leq 1 - 1/d$. Then there exists a probability measure μ on Ω such that every non-constant W_p-Lipschitz path γ with $\gamma(0) = \mu$ has $\mathcal{G}_\alpha(\gamma) = +\infty$.*

Proof. Let Ω be the cube $[0,1]^d$ and μ the Lebesgue measure on it. We want to estimate from below

$$\inf \{ G_\alpha(\nu) \,|\, W_p(\mu, \nu) \leq t \}$$

and we will show it to be larger than $ct^{-d(1-\alpha)}$. Therefore, if γ is a W_p-Lipschitz path with constant speed which originates from μ, the integral defining $\mathcal{G}_\alpha$ diverges. We can simply consider $t = 2^{-k}$. To estimate $G_\alpha(\nu)$, when ν is such that $W_p(\mu, \nu) \leq t$, consider a partition of Ω by small cubes of side ε. Let k be the number of those cubes Q_i such that $\nu(Q_i) \leq \mu(Q_i)/2 = \varepsilon^d/2$. In all these cubes we have a zone in which the optimal transport map S between μ and ν must take values outside the cube; this zone, given by $Q_i \setminus S^{-1}(Q_i)$, has a measure of at least $\varepsilon^d/2$. We want to estimate from below the contribute of this zone to the total transport cost between μ and ν. For this contribute we may write

$$\begin{aligned}
\int_{Q_i \setminus S^{-1}(Q_i)} d(x, \partial Q_i)^p dx &= \int_0^{(\varepsilon/2)^p} \left| \left(Q_i \setminus S^{-1}(Q_i) \right) \cap \{ d(x, \partial Q_i)^p > \tau \} \right| d\tau \\
&\geq \int_0^{(\varepsilon/2)^p} \left(\frac{\varepsilon^d}{2} - \left| \{ d(x, \partial Q_i)^p \leq \tau \} \right| \right) d\tau \\
&\geq \int_0^{B^p \varepsilon^p} \left(\frac{\varepsilon^d}{2} - \left| \{ d(x, \partial Q_i) \leq B\varepsilon \} \right| \right) d\tau \\
&\geq c_1 \varepsilon^p \varepsilon^d,
\end{aligned}$$

where B is sufficiently small and c_1 is a positive constant. By recalling that the total transport cost (i.e. the $p-$th power of the distance W_p) is less than t^p, we have

$$kc_1\varepsilon^{d+p} \leq t^p. \tag{4.2.1}$$

On the other hand, the value of G_α can be estimated from below by means of the other cubes and we have

$$G_\alpha(\nu) \geq (\varepsilon^{-d} - k)c_2\varepsilon^{d\alpha}.$$

Let us now choose $\varepsilon = mt$ with m an integer such that $c_1 m^p > 1$ and, by using (4.2.1), we have

$$G_\alpha(\nu) \geq t^{-d}(m^{-d} - m^{-d-p}/c_1)c_2 m^{d\alpha}t^{d\alpha} = c_3 t^{-d(1-\alpha)},$$

where the constant c_3 is positive.

For a general Ω we can simply use a cube contained in Ω and show that the Lebesgue measure on it, rescaled to a probability measure, cannot be reached keeping finite the value of the integral. □

Example 4.2.5 *(Y-shaped paths versus V-shaped paths).* Consider the example in Figure 4.2, where we suppose that l and h are fixed. We define for $0 \leq t \leq l_0$

$$x(t) = (t, 0)$$

and for $l_0 \leq t \leq l_0 + \sqrt{l_1^2 + h^2}$

$$\begin{aligned} x_1(t) =& (l_0 + l_1\frac{t-l_0}{\sqrt{l_1^2+h^2}}, h\frac{t-l_0}{\sqrt{l_1^2+h^2}}) \\ x_2(t) =& (l_0 + l_1\frac{t-l_0}{\sqrt{l_1^2+h^2}}, -h\frac{t-l_0}{\sqrt{l_1^2+h^2}}). \end{aligned}$$

Let us consider the curve $\gamma : [0, l_0 + \sqrt{l_1^2 + h^2}] \to \mathcal{W}_p(\Omega)$ defined by

$$\gamma(t) = \begin{cases} \delta_{x(t)} & \text{if } 0 \leq t < l_0 \\ \frac{1}{2}\delta_{x_1(t)} + \frac{1}{2}\delta_{x_2(t)} & \text{if } l_0 \leq t \leq l_0 + \sqrt{l_1^2 + h^2}. \end{cases}$$

It easy to see that $|\gamma'|(t) = 1$ and that

$$\mathcal{G}_\alpha(\gamma) = l_0 + 2^{1-\alpha}\sqrt{(1-l_0)^2 + h^2}.$$

Then the minimum is achieved for

$$l_0 = l - \frac{h}{\sqrt{4^{1-r} - 1}}.$$

In particular, when $\alpha = 1/2$ we have a Y-shaped path (similar to the one in Figure 4.2) when $l > h$, while the path is V-shaped when $l \leq h$.

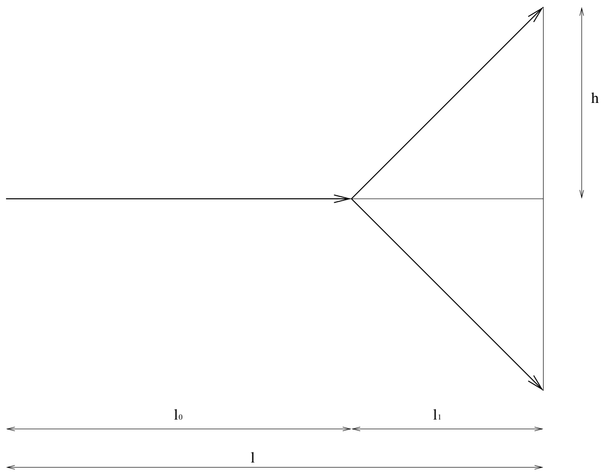

Figure 4.2: A Y-shaped path for $\alpha = 1/2$.

Remark 4.2.6. The result given by Theorem 4.2.3 can clearly be improved for particular choices of μ_0 and μ_1. For instance, we can connect a Dirac mass to the k-dimensional Hausdorff measure on a smooth k-surface for all $\alpha \in]1 - 1/k, 1]$ (see also [42]).

4.2.2 Diffusion

Here we will roughly set $f(z) = |z|^q$ $(q > 1)$ and $g = +\infty$. In fact we are considering the functional

$$F_q(\mu) = \begin{cases} \int_\Omega u^q d\mathcal{L}^d & \text{if } \mu = u \cdot \mathcal{L}^d; \\ +\infty & \text{if } \mu \text{ is not absolutely continuous} \end{cases}$$

This is a particular case of the other case presented in section 1.2 and moreover we can notice that we have $F_q(\mu) = ||\mu||^q_{L^q}$, with the convention that this norm is infinite if the measure does not belong to L^q.

We follow the same structure of the previous section. In this case we will denote the functional J by F_q and $\mathcal{J}$ by $\mathfrak{F}_q$.

We start by proving that when $F_q(\mu_0)$ and $F_q(\mu_1)$ are finite, that is μ_0 and μ_1 are measures with $L^q(\Omega)$ densities, the optimal path problem admits a solution with finite energy.

Theorem 4.2.7. *Assume that $\mu_0 = u_0 \cdot \mathcal{L}^d, \mu_1 = u_1 \cdot \mathcal{L}^d$ with u_0, $u_1 \in L^q(\Omega)$. Then μ_0 and μ_1 can be joined by a finite energy path.*

Proof. The proof of this result relies on the notion of *displacement convexity* (see Section 0.3 for details and useful results). Take T an optimal transport map from μ_0 to μ_1 and let $\pi = (id \times T)_\sharp \mu$ be the associated transport plan. Let $\gamma = \gamma^\pi$ the geodesic in $\mathcal{W}_p$ that it induces according to Theorem 0.3.4. We will show that the energy of the path γ is finite. By Theorem 0.3.6 the functional F_q is displacement convex, so that

$$F_q(\gamma(t)) \le (1-t)F_q(\mu_0) + tF_q(\mu_1).$$

Then

$$\int_0^1 F_q(\gamma(t))|(\gamma)'|(t)\ dt \le \\ W_p(\mu_0,\mu_1)\int_0^1 \left[(1-t)F_q(\mu_0) + tF_q(\mu_1)\right]\ dt = \\ \frac{1}{2}(F_q(\mu_0) + F_q(\mu_1))W_p(\mu_0,\mu_1).$$

Since $F_q(\mu_0)$ and $F_q(u_1)$ are finite, we have that the path $t \mapsto \gamma(t)$ provides a finite value for the energy functional $\mathcal{F}_q$. □

Next step will be the existence of an admissible path for arbitrary extremal measures, if q satisfies some additional constraints.

Recall that if μ_0 and μ_1 are probability measures given by L^1 densities (u_0 and u_1 respectively) and T is a transport map between them with sufficient regularity we have:

$$u_1(y) = u_0(T^{-1}(y))|\text{det } \mathrm{D}T^{-1}(y)|.$$

Lemma 4.2.8. *Let $q < 1 + 1/d$. Let also $\mu = u \cdot \mathcal{L}^d$ with $u \in L^q(\Omega)$ and $\nu = \sum_{j=1}^k b_j\delta_{y_j}$ with $\sum_{j=1}^k b_j = 1$. Then there exists a path between μ and ν with finite energy.*

Proof. Let T be an optimal transport map between μ and ν and π the associated transport plan. Let $B_j := T^{-1}(y_j)$. We now show that the path $\gamma = \gamma^\pi$ has a finite energy. Let us set $T_t = (1-t)id + tT$. If $x \in B_j$, then $T_t(x) = (1-t)x + ty_j$ and $\det \mathrm{D}T_t(x) = (1-t)^d$. Let u_t be the density of the measure $(T_t)_\# \mu$, that is to say:

$$u_t(y) = u(T_t^{-1}(y))|\det \mathrm{D}T_t^{-1}(y)|.$$

We then have:

$$\begin{aligned}\int |u_t(y)|^q \, dy &= \sum_{j=1}^{k} \int \left| u\left(\frac{y - ty_j}{1-t}\right)\right|^q \frac{1}{(1-t)^{dq}} \, dy \\ &= \sum_{j=1}^{k} \int |u(z)|^q (1-t)^{d(1-q)} \, dz \\ &= (1-t)^{d(1-q)} \int |u(z)|^q \, dz.\end{aligned}$$

Moreover, thanks to Theorem 0.3.4, γ is a constant speed geodesic and thus the metric derivative $|\gamma'|(t)$ is constantly equal to the Wasserstein distance $W_p(\mu,\nu)$. Then,

$$\mathcal{F}_q(\gamma) = W_p(\mu,\nu) \int_0^1 \int |u_t(y)|^q \, dy \, dt = \frac{W_p(\mu,\nu)}{d+1-dq} \int_\Omega |u|^q \, dx$$

which is finite since $q < 1 + 1/d$. □

Theorem 4.2.9. *Let $q < 1 + 1/d$. Then every couple of measures can be joined by a path with finite energy.*

Proof. It is enough to link any measure ν to a fixed L^q measure μ (for instance, the normalized Lebesgue measure) with a finite energy path. Let $(\nu_k)_{k\in\mathbb{N}}$ be a sequence of atomic measures approximating ν in the Wasserstein distance W_p. By Lemma 4.2.8, for every k there is a path γ_k with energy

$$\mathcal{F}_q(\gamma_k) = CW_p(\mu,\nu_k)$$

where C is a constant which only depends on d, q, Ω (and of course μ). Extracting a convergent subsequence from $(\gamma_k)_{k\in\mathbb{N}}$ provides a path γ such that, by repeating the lower semicontinuity argument of Theorem 4.1.1, gives

$$\mathcal{F}_q(\gamma) \le \liminf_{k\to+\infty} \mathcal{F}_q(\gamma_k) = \lim_{k\to+\infty} CW_p(\mu,\nu_k) = CW_p(\mu,\nu).$$

Since γ_k connects μ to ν_k, then γ connects μ to ν and the result is established. □

As in the previous section, we show that this result is sharp, as it can be seen from the following statement which is valid in a more general setting. In fact, we prove an estimate which holds for every W_p-Lipschitz curve not only valued in $\mathcal{P}(\Omega)$, but also in $\mathcal{P}(\mathbb{R}^d)$.

Theorem 4.2.10. *Suppose $q \geq 1 + 1/d$. Then there exists $\mu \in \mathcal{W}_p(\Omega)$ such that every non-constant W_p-Lipschitz path γ with $\gamma(0) = \mu$ gives $\mathfrak{F}_q(\gamma) = +\infty$.*

Proof. Let us choose $\mu = \delta_0$. It is sufficient to prove that

$$\inf \{F_q(\nu) \,|\, \nu \in \mathcal{P}(\Omega),\, W_p(\mu, \nu) \leq t\} \geq C t^{-d(q-1)}, \tag{4.2.2}$$

with $C > 0$. In fact, by reparameterization, it is sufficient to prove that the functional is infinite on constant speed paths. Taken such a path γ, with constant speed $L > 0$, we then have

$$\mathcal{F}_q(\gamma) = L \int_0^1 F_q(\gamma(t))dt \geq L \int_0^1 C(Lt)^{-d(q-1)}dt = +\infty,$$

where the integral diverges thanks to the assumption on q. To prove (4.2.2) we can suppose that $\Omega = \mathbb{R}^d$, which is the worst case. This shows that the result does depend neither on the compactness nor on the convexity of Ω. By considering the map that associates to every probability measure ρ the measure $\nu = (m_t)_{\#}\rho$, where $m_t(x) = tx$, one has a one-to-one correspondence between the probabilities whose Wasserstein distance from δ_0 is less than 1 and those whose distance is less than t. It is easy to see that ν is L^q if and only if the same happens for ρ and that the density of ν is the function $x \mapsto t^{-d}u(x/t)$, where u is the density of ρ. Therefore

$$F_q(\nu) = \int \frac{u^q(x/t)}{t^{dq}}dx = \int u^q(y)t^{-dq}t^d dy = F_q(\rho)t^{-d(q-1)}.$$

Consequently, it is now sufficient to evaluate the infimum in (4.2.2) when $t = 1$, and this number will be the constant C we are looking for. We will show that this infimum is in fact a minimum, thus obtaining that it is strictly positive. This problem is quite similar to what we have seen in Chapter 1. To get the existence of a minimum we recall that the functional F_q is sequentially lower semicontinuous with respect to weak topology on probability measures, while the set $\{\nu \in \mathcal{P}(\mathbb{R}^n) \,|\, W_p(\delta_0, \nu) \leq 1\}$ is sequentially compact with respect to the same topology (actually every sequence in such a set turns out to be tight). □

Remark 4.2.11. As in the previous case, it is possible that two measures could be connected by a finite energy path even when q is greater than $1 + 1/d$. For instance, with $N = 2$, the path given by

$$\gamma(t) = \frac{1}{4t} I_{[-1,1]\times[-t,t]} \cdot \mathcal{L}^2$$

is a Lipschitz path in $\mathcal{W}_p([-1,1]\times[-1,1])$ joining $\gamma_0 = 1/2\mathcal{H}^1 \llcorner [-1,1]$ to $\gamma_1 = 1/4\mathcal{L}^2$ (it is in fact a Wasserstein geodesic between them). The energy is finite as far as

$$\int_0^1 \frac{4t}{(4t)^q} dt < +\infty.$$

This condition is fulfilled when $1-q > -1$, i.e. when $q < 2$, instead of the condition $q < 1 + 1/2$ found in Theorem 4.2.9.

4.3 The non-compact case

The existence results of the previous section were based on two important facts: the compactness of Wasserstein spaces $\mathcal{W}_p(\Omega)$ when Ω itself is compact and $1 \le p < +\infty$, and the estimate $F_q \ge c > 0$, proven in Theorem 4.2.1, under the assumption $|\Omega| < +\infty$. Both the facts do not hold when $\Omega = \mathbb{R}^d$, for instance. This is the reason why we developed in Section 4.1 some tools giving the existence of optimal paths under weaker assumption, even in the abstract metric setting. To replace the compactness of Ω we need to use Theorem 4.1.4, while to deal with the fact that we do not have $F_q(\nu) \ge c > 0$ in the case where ν runs over all $\mathcal{W}_p(\mathbb{R}^d)$ we can use the weaker assumption given by hypothesis H2. In this Section we only deal with the case of $\mathfrak{F}_q$-like functionals studied in Section 4.2.2; the case of atomic measures and $\mathcal{G}_\alpha$-like functionals of Section 4.2.1 still present some extra difficulties when Ω is unbounded. We stress the fact that most of the techniques we use can be adapted to deal with several different cases, i.e. Ω unbounded but not necessarily the whole space, or the space $\mathcal{W}_\infty(\Omega)$ (see Section 0.2). Notice that the use of Theorem 4.1.4 is necessary because in general, if Ω is not compact, the corresponding Wasserstein spaces are not even locally compact (see Proposition 0.2.3).

First, we show some lemmas in order to use Theorem 4.1.4.

Lemma 4.3.1. *The weak topology on the space $\mathcal{W}_p(\Omega)$ can be metrized by a distance d' such that $d' \le W_1 \le W_p$.*

Proof. The usual distance metrizing the weak topology is given by

$$d(\mu,\nu) = \sum_{k=1}^{\infty} 2^{-k} \left| \int \phi_k \, d(\mu-\nu) \right|,$$

where $(\phi_k)_k$ is a dense sequence in the unit ball of $C_b^0(\Omega)$. We can choose these functions to be Lipschitz continuous and let, for every index k, c_k be the Lipschitz constant of ϕ_k. Then

$$d'(\mu, \nu) = \sum_{k=1}^{\infty} \frac{2^{-k}}{1 + c_k} \left| \int \phi_k d\,(\mu - \nu) \right|$$

is a distance which metrizes the same topology. Being $\phi_k/(1 + c_k)$ a $1-$Lipschitz function, thanks to the dual formulation of Monge's problem, we have

$$\left| \int \frac{\phi_k}{1 + c_k} d\,(\mu - \nu) \right| \leq W_1(\mu, \nu),$$

and so, by summing up on k, we get $d' \leq W_1$ as required. □

The following two lemmas are well known.

Lemma 4.3.2. *The distance W_p is lower semicontinuous on $\mathcal{W}_p(\Omega) \times \mathcal{W}_p(\Omega)$ endowed with the weak $\times$ weak convergence.*

Proof. Take $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Let γ_n be an optimal transport plan for the cost $|x - y|^p$ between μ_n and ν_n: the sequence of this plans turns out to be tight thanks to tightness of the sequence of the marginal measures, and so we may suppose $\gamma_n \rightharpoonup \gamma$. We can now see that γ is a transport plan between μ and ν and so it holds

$$\begin{aligned} W_p(\mu, \nu) &\leq \left(\int |x - y|^p d\,\gamma \right)^{1/p} \\ &\leq \liminf_{n \to +\infty} \left(\int |x - y|^p d\,\gamma_n \right)^{1/p} = \liminf_{n \to +\infty} W_p(\mu_n, \nu_n). \quad \square \end{aligned}$$

Lemma 4.3.3. *All bounded sets in $\mathcal{W}_p(\mathbb{R}^d)$ are relatively compact with respect to weak topology.*

Proof. Just notice that, in a bounded set, every sequence of probability measures turns out to be tight. The limits up to subsequences (that exist in the weak sense) still belong to the space $\mathcal{W}_p(\mathbb{R}^d)$ as a consequence of the lower semicontinuity of the functional $\mu \mapsto W_p(\mu, \delta_0)$ (which is nothing but the $p-$th momentum of the measure). □

We can give now our result.

Theorem 4.3.4. *Let F_q and $\mathfrak{F}_q$ be defined as in Section 4.2.2 respectively on $\mathcal{W}_p(\mathbb{R}^d)$ and on the set of Lipschitz path in $\mathcal{W}_p(\mathbb{R}^d)$ joining two measures μ_0 and μ_1. Then*

- *if $q < 1 + 1/d$ for every μ_0 and μ_1 there exists a path giving finite and minimal value to $\mathcal{F}_q$;*
- *if $q \geq 1 + 1/d$ there exists a measure μ_0 such that $\mathcal{F}_q = +\infty$ on every non-constant Lipschitz path starting from μ_0.*

Proof. Let us start by the case $q < 1+1/d$: thanks to Lemma 4.3.2 and 4.3.3 we can use Theorem 4.1.4 and so we just need to verify the two assumptions H1 and H2. The existence of a finite-energy path can be achieved in the same way as in Theorem 4.2.9, by passing through a fixed L^q probability measure. Notice that, in order to have the convergence of a subsequence and the lower semicontinuity in the approximation by atomic measures, we will argue as in the proof of Theorem 4.1.4 instead of Theorem 4.1.1. In order to estimate the integral in H2 we will use the same estimate given in Theorem 4.2.10, to achieve

$$\inf \{F_q(\nu) \,|\, \nu \in \mathcal{P}(\Omega),\, W_p(\mu, \nu) \leq t\} \geq Ct^{-d(q-1)},$$

so that the integral diverges as far as $q < 1 + 1/d$.

By repeating the arguments of Theorem 4.2.10, we can then prove also the second part of our result, because $\mu = \delta_0$ cannot be joined to any other probability measure by a finite energy path. □

Remark 4.3.5. In the previous theorem we did not mention the possibility to link, for arbitrary $q > 1$, two measures $\mu_0, \mu_1 \in L^q(\mathbb{R}^d)$. It is easy to check that the same construction used in Theorem 4.2.7 can be used in this setting too. We get in such a way the existence of a path providing a finite value to $\mathfrak{F}_q$, but some problems arise when we look for a minimal one. In fact, for arbitrary q, condition H2 is no longer fulfilled and this prevents us from applying the general existence results.

To conclude this section, we highlight the difference between the case we dealt with (the $\mathfrak{F}_q$ case) and the other important case, represented by the functional $\mathcal{G}_\alpha$. In this latter case it is not necessary to pass through the divergence of the integral in assumption H2, because we actually have $G_\alpha \geq 1$, as already shown. On the other hand, some difficulties arise in verifying assumption H1. In fact the construction we made to build a finite energy path linking δ_0 to a probability measure μ strongly uses the compactness of the support of μ. In order to get a similar construction for the case $\Omega = \mathbb{R}^d$ we would need an estimate such as

$$\inf\{W_p(\mu, \nu) \,|\, \#\,\mathrm{spt}(\nu) \leq k\} \leq C(\mu)k^{-1/d}, \tag{4.3.1}$$

where $C(\mu)$ is a finite constant depending on the measure μ. It is easy to get a similar estimate when μ has compact support, but the constant may depend on the diameter of its support. The existence of a similar estimate for arbitrary measures μ is linked to the asymptotics of the rescaled location problem in $\mathbb{R}^d$. A theory on this asymptotic problem has been explicitly developed (for instance in [19]) only in the case of compact support. However, it leads to a condition like $\mu^{d/(d+p)} \in L^1$, which is always fulfilled for μ compactly supported, while it may fail for general probability measures in $\mathcal{W}_p(\mathbb{R}^d)$. Other interesting estimates that could replace (4.3.1) may be found in [52], where the asymptotics of the same problem under constraints on $G_\alpha(\nu)$ instead of constraints on the cardinality of the support is considered. Unluckily, also these estimates cannot be directly used here.

On the contrary, the case of the functional $\mathcal{G}_\alpha$ for $p = \infty$ and Ω bounded would be feasible. In this case the estimate (4.3.1) is in fact valid and we can prove the existence of a curve with bounded energy reaching any probability measure, provided $\alpha > 1 - 1/d$. We do not develop explicitly the proof, but the reader may convince himself that putting together all the elements that we have proved here one can extend all te results to the case $p = \infty$. We want to stress this fact since the case $p = \infty$ will be important for an informal comparison of this model to the models in [72], [57] and [13] that will be considered in chapters 6, 7 and 9 (see Section 6.2).

Chapter 5

A system of PDEs from a geodesic problem in W_p

In this chapter, mainly based on a joint work with Luigi Ambrosio (see [6]), we consider the diffusion case of Chapter 4 and we look for a system of PDEs characterizing the optimal curves. This can be performed only in the diffusion case, since it in this case that we are concerned with measures which have a density. In fact in the concentration case we are facing a curve of measures which are too singular (usually atomic) to write PDEs on them. This Chapter follows the accepted version of the article [6], up to some difference in the presentation of the topic and some preliminaries which have been moved to Section 0.3.

5.1 Compressible Euler equations from geodesic problems

As in Chapter 4 we will consider a convex domain $\Omega \subset \mathbb{R}^d$ and the corresponding Wasserstein spaces $W_p(\Omega)$. As we mentioned in Section 0.3, it is well known that $W_p(\Omega)$ is a length space, and that (constant speed) geodesics of $W_p(\Omega)$ are in one to one correspondence with optimal transport plans, via McCann's linear interpolation procedure (see for instance Proposition 7.2.2 of [4] and Section 0.3). Here we consider, as in Chapter 4, the case when the Wasserstein metric is perturbed by a conformal factor $J(\mu)$: by minimizing

$$\int_0^1 J(\mu_t)|\mu'|(t)\,dt \tag{5.1.1}$$

among all curves μ connecting $\mu_0 = \mu$ to $\mu_1 = \nu$, one obtains a new distance depending on p and J, and we are interested in computing the geodesics relative to this distance. In (5.1.1), $|\mu'|(t)$ is the metric derivative, see Section 0.3.

In chapter 4 we saw this problem in two opposite cases, namely the concentration and the diffusion one. As we will see later in Chapter 6, the motivation of the study mainly relied on the concentration case, in correlation to some branched transport problems. The case of a functional J which is a local functional preferring spread measures is considered in Chapter 4 only as a natural counterpart and the two problems sound somehow specular. The aim of the present Chapter is to consider this second problem and to find out optimality conditions in the form of PDEs.

Hence, we study in detail the case when $J(\mu)$ is the γ-th power of the L^q norm of the density of μ with respect to Lebesgue measure $\mathcal{L}^d$, with $q > 1$ and $\gamma > 0$ given (and $J(\mu) = +\infty$ if μ is a singular measure). Thus, geodesics with respect to the new metric tend to spread the density as much as possible. Denoting by u_t the density of μ_t, we find that a necessary optimality condition for geodesics is (for $p = 2$, see (5.2.4) for general p)

$$\frac{d}{dt}\left(K(t)vu\right) + K(t)\nabla \cdot (v \otimes vu) + H(t)\nabla u^q = 0, \tag{5.1.2}$$

where v_t is the tangent velocity field of μ_t, linked to u_t via the continuity equation $\frac{d}{dt}u_t + \nabla \cdot (v_t u_t) = 0$. Here $H(t) < 0$ and $K(t) > 0$ are suitable functions depending only on the metric derivative of μ_t and on $J(\mu_t)$. As Brenier pointed to us, this equation is very similar to the compressible Euler equation, but with a negative pressure field $p = H(t)u^q$; a similar equation, with H constant and $q = 3$, recently appeared also in [50], in the one-dimensional case. In fact the main difference appears in the relationship between the L part and the speed part: here it is multiplicative, while in [50] it is additive, as we will explain in a while.

The appearance of the Euler equation as an optimality condition is not surprising, taking into account the approach developed, in the incompressible case, by Brenier (first in a purely Lagrangian framework in [22], [23], and then in a mixed Eulerian-Lagrangian one in [24], [25]). In this connection, we mention that our derivation of the optimality condition differs from [23], [25], where duality is used to perform first variations, and uses instead a perturbation argument directly at the level of the primal problem.

Due to the non-convex nature of this problem, we don't know of any sufficient minimality condition for the geodesics. In this connection, one may notice that, in the case $\gamma = q/2$ and $p = 2$, we have

$$\inf_{\delta>0} \delta \int_\Omega u^q \, dx + \frac{1}{\delta} \int_\Omega |v|^2 u \, dx = 2J(u\mathcal{L}^d) \left(\int_\Omega |v|^2 u \, dx \right)^{1/2}$$

and the minimal $L^2(\mu)$ norm of v is strictly linked to the metric derivative. This suggests a connection between the *multiplicative* model studied here and in Chapter 4, and the *additive* model

$$\min \left\{ \int_0^1 \int_\Omega u^q + |v|^2 u \, dxdt : \ \frac{d}{dt} u + \nabla \cdot (vu) = 0 \right\}$$

subject to Dirichlet conditions at $t = 0$ and $t = 1$. This additive model, in the case $q = 3$, is exactly the one studied in [50] (in this connection, see also [55]). Notice that this problem is *convex* in the pair (u, vu). It turns out, indeed, that the (necessary and sufficient, by the convex nature of the problem) optimality conditions for the additive model are very similar to (5.1.2), the only difference being that H and K do not depend on time.

5.2 Optimality conditions for weighted Wasserstein geodesics

5.2.1 A new velocity vector field

We refer to the notion of tangent and velocity vector fields of Section 0.3. We want here to investigate how velocity fields change if we modify the curve μ_t.

Theorem 5.2.1. *Let a Lipschitz curve $\mu : [0,1] \to W_p(\Omega)$ and a smooth function $T : [0,1] \times \Omega \to \Omega$ be given, such that for any t the function $T_t := T(t, \cdot)$ is a diffeomorphism. Let us consider the new curve μ' given by $\mu'_t = (T_t)_\sharp \mu_t$. If v_t is a velocity field for μ_t, then the vector field v' defined by*

$$v'_t \cdot \mu'_t = (T_t)_\sharp \left[\left(\nabla T_t \cdot v_t + \frac{\partial T}{\partial t} \right) \mu_t \right]$$

is a velocity field for μ'_t.

Proof. We have

$$\int_\Omega \phi\, d\mu'_{t+h} - \int_\Omega \phi\, d\mu'_t = \int_\Omega \phi \circ T_{t+h} d\mu_{t+h} - \int_\Omega \phi \circ T_t\, d\mu_t$$
$$= \int_\Omega (\phi \circ T_{t+h} - \phi \circ T_t)\, d\mu_{t+h} + \int_\Omega \phi \circ T_t\, d(\mu_{t+h} - \mu_t)$$
$$= \int_\Omega \left(\int_t^{t+h} (\nabla \phi) \circ T_s \cdot \frac{\partial T}{\partial t}|_s\, ds \right) d\mu_{t+h} + \int_t^{t+h} ds \int_\Omega (\nabla\phi) \circ T_t \cdot \nabla T_t \cdot v_s\, d\mu_s,$$

where in the last equality we have used the fact that v_t is a velocity field for μ, with test function $\phi \circ T_t$. It is now convenient to divide by h, rewrite and pass to the limit as $h \to 0$:

$$\frac{\int_\Omega \phi d\mu'_{t+h} - \int_\Omega \phi d\mu'_t}{h} = \int_\Omega d\mu_{t+h} \frac{1}{h} \int_t^{t+h} (\nabla\phi) \circ T_s \cdot \frac{\partial T}{\partial t}|_s\, ds$$
$$+ \int_\Omega (\nabla \phi) \circ T_t \cdot \nabla T_t \cdot v_t\, d\mu_t + \frac{1}{h} \int_t^{t+h} ds \int_\Omega \nabla \psi_t \cdot (v_s d\mu_s - v_t d\mu_t)\,, \quad (5.2.1)$$

where $\psi_t = \phi \circ T_t$. In the first term on the right hand side the measures μ_{t+h} weakly converge to μ_t, since $t \mapsto \mu_t$ is Lipschitz continuous, while the integrand uniformly converges as a function of the space variable x to $(\nabla\phi) \circ T_t \cdot \frac{\partial T}{\partial t}$ as $h \to 0$. Hence we get convergence of the integral. If we prove that the last term tends to zero at least for a.e. $t \in [0,1]$ we get the thesis, since then we would have

$$\lim_{h \to 0} \frac{\int_\Omega \phi d\mu'_{t+h} - \int_\Omega \phi d\mu'_t}{h}$$
$$= \int_\Omega \left((\nabla \phi) \circ T_t \cdot \frac{\partial T}{\partial t} + (\nabla\phi) \circ T_t \cdot \nabla T_t \cdot v_t \right) d\mu_t = \int_\Omega \nabla \phi \cdot v'_t d\mu'_t,$$

and this is nothing but the differential version of the continuity equation for v' and μ' (it remains to prove $v'_t \in L^p(\mu'_t)$ but this is straightforward since T_t is a diffeomorphism and this allows to write down the densities and estimate them). To prove that the last term vanishes at the limit we see that, for fixed $\psi \in \mathrm{Lip}(\Omega)$ the function

$$s \mapsto g_\psi(s) := \int_\Omega \nabla \psi \cdot v_s d\mu_s = \frac{d}{ds} \int_\Omega \psi\, d\mu_s$$

is L^∞ since μ_t is a Lipschitz curve in $W_p(\Omega)$ and hence almost any $s \in [0,1]$ is a Lebesgue point. This allows to fix a negligible set $N \subset [0,1]$ such that any point $t \in [0,1] \setminus N$ is a Lebesgue point for all the functions $g_{\psi_{t_1}}$ for $t_1 \in \mathbb{Q}$. We fix now $t \in [0,1] \setminus N$ and try to prove that the last integral in (5.2.1) tends to zero. For $t_1 \in \mathbb{Q}$ it holds

$$\begin{aligned}
&\left| \frac{1}{h} \int_t^{t+h} ds \int_\Omega \nabla\psi_t \cdot (v_s d\mu_s - v_t d\mu_t) \right| \\
&\quad \le \frac{1}{h} \int_t^{t+h} ds \left(\left| \int_\Omega \nabla(\psi_t - \psi_{t_1}) \cdot v_s d\mu_s \right| + \left| \int_\Omega \nabla(\psi_t - \psi_{t_1}) \cdot v_t d\mu_t \right| \right) \\
&\qquad\qquad + \left| \frac{1}{h} \int_t^{t+h} ds \int_\Omega \nabla\psi_{t_1} \cdot (v_s d\mu_s - v_t d\mu_t) \right| \\
&\quad \le \operatorname{Lip}(\psi_t - \psi_{t_1}) \operatorname{Lip}_{W_p}(\mu) + \left| \frac{1}{h} \int_t^{t+h} ds \int_\Omega \nabla\psi_{t_1} \cdot (v_s d\mu_s - v_t d\mu_t) \right|.
\end{aligned}$$

In the last sum the second term tends to zero by the fact that t is a Lebesgue point for $g_{\psi_{t_1}}$ and the first term may be made as small as we want by choosing t_1 close to t, since $\psi_s = \phi \circ T_s$ and both ϕ and T are regular. □

5.2.2 Derivation of the optimality conditions

We consider the diffusion case of the minimization problem presented in Chapter 4, i.e. finding a curve of measures in $W_p(\Omega)$ of minimal length according to a metric which, roughly speaking is the Wasserstein (infinitesimal) metric multiplied by a conformal factor. Precisely, we consider for $q > 1$ the usual functional (see Sections 1.2 and 4.2)

$$F_q(\nu) = \begin{cases} \int_\Omega u^q d\mathcal{L}^d & \text{if } \nu = u \cdot \mathcal{L}^d \\ +\infty & \text{otherwise,} \end{cases}$$

we want to minimize

$$\int_0^1 F_q(\mu_t) |\mu'|(t)\, dt,$$

where $|\mu'|(t)$ is the metric derivative of the curve μ and the minimization occurs among all the W_p-Lipschitz curves $t \mapsto \mu_t$ with given initial and final points, i.e. μ_0 and μ_1 are given probability measures in $W_p(\Omega)$. We will always consider only the non trivial case $\mu_0 \neq \mu_1$. If we define $V(\mu, t) = \int_\Omega |v_t|^p d\mu_t$, where v is the tangent field to the curve μ_t, we know that $|\mu'|(t) = V(\mu, t)^{1/p}$. We may generalize the functional we want to minimize by considering

$$\mathcal{F}(\mu) := \int_0^1 F_q(\mu_t)^\alpha\, V(\mu, t)^\beta dt$$

which reduces to the case studied in [20] and Chapter 4 if $\alpha = 1$ and $\beta = 1/p$. Notice that in this case the functional does not change under reparameterization of curves, while if $\beta > 1/p$ the minimization selects a particular parametrization. For $\beta < 1/p$ the existence of a minimum will in general fail. Anyway we do not deal here with existence results (see [20]), but we only look for necessary optimality conditions. We will consider variations of μ of the form

$$\mu_t^\varepsilon = (T_t^\varepsilon)_\sharp \mu_t \quad \text{with} \quad T^\varepsilon(t,x) = x + \varepsilon\xi(t,x), \;\; T_t^\varepsilon = id + \varepsilon\xi(t,\cdot),$$

for arbitrary regular functions $\xi \in C_c^\infty([0,1]\times\Omega;\mathbb{R}^d)$. In the end optimality conditions will be expressed through a system of PDEs: we will obtain the result after collecting some lemmas. What we want to do now is exploiting the fact that for a minimizing curve μ the following quantity must be minimal for $\varepsilon = 0$:

$$\mathcal{F}(\mu_t^\varepsilon) = \left(\int_0^1 F_\varepsilon(t)^\alpha V_\varepsilon(t)^\beta dt\right),$$

provided we define $F_\varepsilon(t) = F_q(\mu_t^\varepsilon)$ and $V_\varepsilon(t) = V(\mu^\varepsilon, t)$. Since it is not completely easy to deal with the term $V_\varepsilon(t)$, we will replace it by $\tilde{V}_\varepsilon(t)$, with $\tilde{V}_\varepsilon(t)$ given by

$$\tilde{V}_\varepsilon(t) = \int_\Omega |(v^\varepsilon)_t|^p \, d\mu_t^\varepsilon.$$

Here the vector field v^ε is the one we get by Theorem 5.2.1 when the map T is given by T^ε and the initial field v_t is the tangent field to μ_t. In this way we have $\tilde{V}_\varepsilon(t) \geq V_\varepsilon(t)$ (since v_t^ε is a velocity field which is not necessarily of minimal L^p norm) but $\tilde{V}_\varepsilon(0) = V_\varepsilon(0)$. Thus we may switch to considering $\tilde{V}_\varepsilon(t)$ instead of $V_\varepsilon(t)$, getting

$$\tilde{\mathcal{F}}(\mu_t^\varepsilon) = \left(\int_0^1 F_\varepsilon(t)^\alpha \tilde{V}_\varepsilon(t)^\beta dt\right).$$

We will compute the derivative of $\tilde{\mathcal{F}}(\mu_t^\varepsilon)$ with respect to ε and get the conditions we are looking for.

Lemma 5.2.2. *If μ is a curve given by $\mu_t = u_t\mathcal{L}^d$ and such that $\mathcal{F}(\mu) < +\infty$, then for almost any $t \in [0,1]$ it holds*

$$\frac{d}{d\varepsilon}F_\varepsilon(t) = (1-q)\int_\Omega (JT_t^\varepsilon)' \left(\frac{u_t}{JT_t^\varepsilon}\right)^q d\mathcal{L}^d.$$

In particular, if we compute the derivative at $\varepsilon = 0$*, we have*

$$\frac{d}{d\varepsilon} F_\varepsilon(t)|_{\varepsilon=0} = (1-q)\int_\Omega (\nabla \cdot \xi) u_t^q d\mathcal{L}^d.$$

Moreover, for ε *sufficiently small (depending on* T*, but not on* t*) the following inequality holds:*

$$\frac{d}{d\varepsilon} F_\varepsilon(t) \leq C F_q(\mu_t).$$

Proof. We look at the integrand function in the definition of F_ε: to do this it is necessary to look at the density of the measure μ_t^ε. Thanks to the change of variables formula, this density can be easily seen to be given by

$$u_t^\varepsilon = \frac{u_t}{JT_t^\varepsilon} \circ (T_t^\varepsilon)^{-1},$$

where J stands for the Jacobian (this formula is a consequence of T_t^ε being a diffeomorphism at least for small ε). Thus, after changing variables, we have

$$F_\varepsilon(t) = F_q(\mu_t^\varepsilon) = \int_\Omega \left(\frac{u_t}{JT_t^\varepsilon}\right)^q JT_t^\varepsilon \, d\mathcal{L}^d.$$

The derivative of the integral is given by

$$(1-q)(JT_t^\varepsilon)' \left(\frac{u_t}{JT_t^\varepsilon}\right)^q,$$

where $(JT_t^\varepsilon)'$ stands for the derivative w.r.t. ε of JT_t^ε. This quantity may be easily estimated by Cu_t^q, since $1 - a \leq JT_t^\varepsilon \leq 1 + a$ and $(JT_t^\varepsilon)' \leq B$ for suitable constants a and B. Since for almost any t the function u_t belongs to L^q (because the functional we are minimizing is finite) we can apply the dominated convergence theorem and get the thesis. To obtain the derivative at $\varepsilon = 0$ it is sufficient to notice that $(JT_t^\varepsilon)'|_{\varepsilon=0} = \nabla \cdot \xi$, which is well-known. The same estimate we used to get dominated convergence may be used to get the last inequality. □

In the next lemma we consider the term $\tilde{V}_\varepsilon$.

Lemma 5.2.3. *If* μ *is a curve such that* $\mathcal{F}(\mu) < +\infty$*, then for almost any* $t \in [0,1]$ *it holds*

$$\frac{d}{d\varepsilon}\tilde{V}_\varepsilon(t) = p\int_\Omega \left|\nabla T_t^\varepsilon \cdot v_t + \frac{\partial T^\varepsilon}{\partial t}\right|^{p-2} \left(\nabla T_t^\varepsilon \cdot v_t + \frac{\partial T^\varepsilon}{\partial t}\right) \cdot \left(\nabla \xi \cdot v_t + \frac{\partial \xi}{\partial t}\right) d\mu_t. \tag{5.2.2}$$

In particular, if we compute the derivative at $\varepsilon = 0$, we have

$$\frac{d}{d\varepsilon}\tilde{V}_\varepsilon(t)|_{\varepsilon=0} = p\int_\Omega |v_t|^{p-2} v_t \cdot \left(\nabla\xi \cdot v_t + \frac{\partial \xi}{\partial t}\right) d\mu_t.$$

Moreover, for ε sufficiently small (depending on T, but not on t) the following inequality holds:

$$\frac{d}{d\varepsilon}\tilde{V}_\varepsilon(t) \le C(V(\mu,t)+1).$$

Proof. If we compute the densities of μ_t^ε and the expression of the new velocity field and we change variable in the integral by $y = T_t^\varepsilon(x)$, as we did in the previous lemma, we get

$$\tilde{V}_\varepsilon(t) = \int_\Omega \left|\nabla T_t^\varepsilon \cdot v_t + \frac{\partial T^\varepsilon}{\partial t}\right|^p d\mu_t. \tag{5.2.3}$$

When we differentiate the integrand we get exactly the integrand in (5.2.2), and we need only to show that this expression is uniformly dominated, at least for small ε and almost every t to get the result. By boundedness of the derivatives of T^ε it is not difficult to see that the norm of the first vector in the scalar product in the integrand may be estimated by

$$\left|\nabla T_t^\varepsilon \cdot v_t + \frac{\partial T^\varepsilon}{\partial t}\right|^{p-1} \le (C|v_t| + C)^{p-1},$$

while for the second it holds

$$\left|\nabla\xi \cdot v_t + \frac{\partial \xi}{\partial t}\right| \le C|v_t| + C$$

for a suitable constant C. Hence, since $v_t \in [L^p(\mu_t)]^d$ for almost every t the integrability is proved and the differentiation under the integral sign can be performed. □

To conclude, we must put together the two previous results in order to compute the derivative of the integral in t.

Theorem 5.2.4. *If μ is a curve with $\mathcal{F}(\mu) < +\infty$ and $V(\mu,t) \ge V_0 > 0$ for almost every t, then it holds*

$$\frac{d}{d\varepsilon}\tilde{\mathcal{F}}(\mu^\varepsilon)|_{\varepsilon=0} = \alpha(1-q)\int_0^1 F^{\alpha-1}V^\beta \int_\Omega (\nabla\cdot\xi) u_t^q \, d\mathcal{L}^d \, dt + p\beta\int_0^1 F^\alpha V^{\beta-1}\int_\Omega |v_t|^{p-2}\left(\nabla\xi\cdot v_t + \frac{\partial\xi}{\partial t}\right)\cdot v_t \, d\mu_t \, dt,$$

where $F(t) = F_q(\mu_t)$ and $V(t)$ has the usual meaning.

Proof. By the definition of $\tilde{\mathcal{F}}(\mu^\varepsilon)$ we see that the pointwise derivative of the integrand is given by $\alpha F_\varepsilon(t)^{\alpha-1}\frac{dF}{d\varepsilon}\tilde{V}_\varepsilon(t)^\beta + \beta F_\varepsilon(t)^\alpha \tilde{V}_\varepsilon(t)^{\beta-1}\frac{d\tilde{V}}{d\varepsilon}$. By the regularity of T^ε the term $F_\varepsilon(t)$ may be estimated both from above and below by $F(t)$, up to multiplicative constants. As far as $\tilde{V}^\varepsilon(t)$ is concerned, the argument is a little bit more tricky. Indeed we must write $\tilde{V}^\varepsilon(t)$ according to (5.2.3), then estimate

$$A^-|v_t| - B \leq \left|\nabla T_t^\varepsilon \cdot v_t + \frac{\partial T^\varepsilon}{\partial t}\right| \leq A^+|v_t| + B,$$

for ε small enough, where the constants $A^\pm$ are as close to 1 as we want and the constant B is as small as we want (this comes from $\nabla T_t^\varepsilon = id + O(\varepsilon)$ and $\partial T^\varepsilon/\partial t = O(\varepsilon)$), and get

$$A^-\tilde{V}^0 - B \leq \tilde{V}^\varepsilon \leq A^+\tilde{V}^0 + B.$$

The assumption $V \geq V_0 > 0$ allows us to infer from these inequalities that also $\tilde{V}^\varepsilon$ may be estimated both from above and below by V up to multiplicative constants. Finally, by the estimates in Lemmas 5.2.2 and 5.2.3, we bound the whole pointwise derivative by $CF^\alpha V^\beta$ since we have

$$\frac{dF}{d\varepsilon} \leq CF; \qquad \frac{d\tilde{V}}{d\varepsilon} \leq C(V+1) \leq C(1+\frac{1}{V_0})V,$$

)the last inequality too comes from $V \geq V_0$). Since $F_q^\alpha V^\beta$ is integrable on $[0,1]$, we may differentiate under the integral sign and get

$$\frac{d}{d\varepsilon}\mathcal{F}(\mu^\varepsilon)|_{\varepsilon=0} = \int_0^1 \left(\alpha F(t)^{\alpha-1}\frac{dF}{d\varepsilon}|_{\varepsilon=0}\tilde{V}(t)^\beta + \beta F(t)^\alpha \tilde{V}(t)^{\beta-1}\frac{d\tilde{V}}{d\varepsilon}|_{\varepsilon=0}\right) dt.$$

The result follows when we replace the derivatives in ε by the explicit expressions we computed in Lemmas 5.2.2 and 5.2.3. ☐

Remark 5.2.5. If $\beta = 1/p$ and μ is a minimizer, it is always possible to get the lower bound $V \geq V_0$ by reparameterizing in time, for instance by choosing the constant speed parametrization.

Corollary 5.2.6. *If μ minimizes $\mathcal{F}$ with given boundary conditions μ_0 and μ_1, then its density u and its tangent field v satisfy*

$$\alpha(1-q)\int_0^1 F(t)^{\alpha-1}V(t)^\beta \int_\Omega (\nabla\cdot\xi)u_t^q \, d\mathcal{L}^d \, dt$$
$$+ p\beta\int_0^1 F(t)^\alpha V(t)^{\beta-1}\int_\Omega u_t|v_t|^{p-2}(\nabla\xi\cdot v_t + \frac{\partial\xi}{\partial t})\cdot v_t \, d\mathcal{L}^d \, dt = 0,$$

for any vector field $\xi \in C_c^\infty(]0,1[\times\Omega;\mathbb{R}^d)$.

Proof. It is sufficient to notice that when we create the modified curve μ^ε starting form the vector field ξ we do not change the initial and final points of the curve, so that the minimality implies that the derivative of $\tilde{\mathcal{F}}(\mu^\varepsilon)$ at $\varepsilon = 0$ must vanish. $\square$

5.2.3 The resulting system of PDEs

The following theorem follows directly from the previous section.

Theorem 5.2.7. *Let μ_0, $\mu_1 \in \mathcal{W}_p(\Omega)$ and let μ be a curve with $\mathcal{F}(\mu) < +\infty$ which minimizes $\mathcal{F}$ over all the Lipschitz curves with prescribed starting and arrival measures μ_0 and μ_1. Then, denoting by $u(t,\cdot)$ the density of μ_t and by $v(t,\cdot)$ the tangent field to the curve μ, the pair (u,v) provides a weak (distributional) solution of the system*

$$\begin{cases} H(t)\nabla u^q + K(t)\nabla\cdot\left(u|v|^{p-2}v\otimes v\right) + \frac{d}{dt}\left(K(t)u|v|^{p-2}v\right) = 0 & \text{in } \Omega, \\ \frac{d}{dt}u + \nabla\cdot(vu) = 0 & \text{in } \Omega \\ uv\cdot n = 0 & \text{on } \partial\Omega \\ \lim_{t\downarrow 0} u(t,\cdot)\mathcal{L}^d = \mu_0; \quad \lim_{t\uparrow 1} u(t,\cdot)\mathcal{L}^d = \mu_1, & \end{cases} \tag{5.2.4}$$

where $H(t) = \alpha(1-q)F(t)^{\alpha-1}V(t)^\beta$ and $K(t) = p\beta F(t)^\alpha V(t)^{\beta-1}$.
Given (μ_0,μ_1), if $\beta \geq 1/p$, the existence of minimizers is ensured whenever $q < 1 + 1/d$ or, for general q, under the assumption $\mu_0 = u_0\mathcal{L}^d$, $\mu_1 = u_1\mathcal{L}^d$ with u_0, $u_1 \in L^q(\Omega)$ (see Chapter 4). In particular, under these conditions, the existence of solutions to this system is ensured.

It is interesting to rewrite the equations, make some formal simplification and look at some particular cases.

First we expand all the terms in the first equation of System (5.2.4), obtaining

$$\begin{aligned} H(t)\nabla u^q + K(t)\left(u|v|^{p-2}v\cdot\nabla v + v|v|^{p-2}\nabla\cdot(uv) + u\left(v\cdot\nabla|v|^{p-2}\right)v\right) \\ + K(t)\left(v|v|^{p-2}\frac{d}{dt}u + u\frac{d}{dt}\left(v|v|^{p-2}\right)\right) + \frac{d}{dt}K(t)u|v|^{p-2}v = 0. \end{aligned} \tag{5.2.5}$$

Notice that this is always a vector equation, i.e. a system itself, consisting of d equations with $d+1$ unknown functions (the components of v and the density u). This system is then completed by the continuity equation. As usual, by $v\cdot\nabla v$ we mean the vector whose $i-$th component is $\sum_j (v_j\partial v_i/\partial x_j)$.

A formal simplification in (5.2.5) may be done: in fact there is a term $(K(t)v|v|^{p-2})(du/dt+\nabla\cdot(uv))$ that might be removed by using the continuity equation. This is actually possible only under extra regularity assumptions on K and v (it consists of testing the continuity equation against the product $K(t)v|v|^{p-2}$ which is not in general C^1 or regular enough). Anyway, after this formal simplification, (5.2.5) becomes

$$H(t)\nabla u^q + K(t)\left(u|v|^{p-2}v\cdot\nabla v + u\left(v\cdot\nabla|v|^{p-2}\right)v\right) \\ + K(t)u\frac{d}{dt}\left(v|v|^{p-2}\right) + \frac{d}{dt}K(t)u|v|^{p-2}v = 0. \quad (5.2.6)$$

Notice that in the case $\beta = 1/p$ we can reparameterize in time the solution and there are several possible parametrization choices that present some advantages. For instance, we could choose a parametrization so that $K(t)$ is constant, to get rid of the final derivative in time. This choice implies

$$V(t) = \left(\frac{F^\alpha}{K}\right)^{p/(p-1)},$$

and this, in the case of a bounded $|\Omega| < +\infty$, is sufficient to have the lower bound $V \geq V_0$, since F would be bounded from below by a positive constant.

Another important fact to be noticed is that in (5.2.6) there is a common u factor. It is still formal, but in this way we should get, on $\{u > 0\}$,

$$H(t)u^{q-2}\nabla u + K(t)\left(|v|^{p-2}v\cdot\nabla v + \left(v\cdot\nabla|v|^{p-2}\right)v\right) \\ +K(t)\frac{d}{dt}\left(v|v|^{p-2}\right) + \frac{d}{dt}K(t)|v|^{p-2}v = 0.$$

Remark 5.2.8. One might wonder whether the solutions u are automatically positive a.e. in Ω for $t \in]0,1[$. This could be suggested by the fact that in the minimization problem the diffusion of the density is favored and this could entitle us to simplify u in the system. In the next session we will see with explicit examples that this is not necessarily the case.

We finish this overview of simplifications of the system by looking at the simplest case, i.e. $p = q = 2$, $\alpha = 1$, $\beta = 1/2$, in the parametrization regime where K is constant. In this case we get

$$\begin{cases} -2V(t)^{1/2}\nabla u + K\left(v\cdot\nabla v + \frac{d}{dt}v\right) = 0 & \text{in } \{u > 0\}, \\ \frac{d}{dt}u + \nabla\cdot(vu) = 0 & \text{in } \Omega \\ uv\cdot n = 0 & \text{on } \partial\Omega \\ \lim_{t\downarrow 0} u(t,\cdot)\mathcal{L}^d = \mu_0; \quad \lim_{t\uparrow 1} u(t,\cdot)\mathcal{L}^d = \mu_1. \end{cases} \quad (5.2.7)$$

Under no constraint on the parametrization we would have, instead,

$$\begin{cases} -2V(t)^{1/2}\nabla u + K(t)\left(v\cdot\nabla v + \frac{d}{dt}v\right) + v\frac{dK}{dt} = 0 & \text{in } \{u>0\},\\ \frac{d}{dt}u + \nabla\cdot(vu) = 0 & \text{in } \Omega\\ uv\cdot n = 0 & \text{on } \partial\Omega\\ \lim_{t\downarrow 0} u(t,\cdot)\mathcal{L}^d = \mu_0; \quad \lim_{t\uparrow 1} u(t,\cdot)\mathcal{L}^d = \mu_1. \end{cases} \tag{5.2.8}$$

5.3 Self-similar solutions

5.3.1 Homothetic solutions with fixed center

In this section we look for particular solutions of the System (5.2.4) which are self-similar in the sense that, for any t, the measure μ_t is the image under an homothety of a fixed measure. For simplicity we will consider only the case of System (5.2.8), i.e. with $p = q = 2$, and we assume that $0 \in \Omega$. The regularity of the candidate solutions we will propose will be enough to ensure that we can use this simplified system, instead of System (5.2.4). To start this analysis it is necessary to establish the following Lemma.

Lemma 5.3.1. *If μ is a curve in $\mathcal{W}_2(\Omega)$ of the form $\mu_t = (T_{R(t)})_\sharp\overline{\mu}$ for a certain regular function $R : [0,1] \to]0,1]$ (where $T_R(x) = Rx$ is the multiplication by a factor R, hence an homothety), then its tangent field is given by $v_t(x) = xR'(t)/R(t)$.*

Proof. It is not difficult to prove that the field we defined solves the continuity equation and hence is a velocity field. Indeed, if $\phi \in C^1_c(\Omega)$, it holds

$$\begin{aligned}\frac{d}{dt}\int_\Omega \phi d\mu_t = \frac{d}{dt}\int_\Omega \phi(R(t)x)\,d\mu(x) = \int_\Omega \nabla\phi(R(t)x)\cdot R'(t)x\,d\mu(x)\\ = \int_\Omega \nabla\phi(R(t)x)\cdot\frac{R'(t)}{R(t)}R(t)x\,d\mu(x) = \int_\Omega \nabla\phi\cdot v_t\,d\mu_t.\end{aligned}$$

It remains to prove that v is actually the tangent velocity field, i.e. that its L^2 norm is minimal for a.e. t. This is achieved if we are able to prove that $\|v_t\|_{L^2(\mu_t)} = |\mu|'(t)$ for a.e. $t \in [0,1]$. To do this, let us fix two times $t < t+h$ and see that the map $T(x) = xR(t+h)/R(t)$ is a transport between μ_t and μ_{t+h}. Since it is the gradient of the convex function $x \mapsto x^2R(t+h)/2R(t)$, it is actually the optimal transport according to the quadratic cost. Hence

$$\frac{W_2^2(\mu_t,\mu_{t+h})}{h^2} = \frac{1}{h^2}\int_\Omega\left(\frac{R(t+h)}{R(t)} - 1\right)^2 x^2\,d\mu_t(x) \to \int_\Omega\left(\frac{R'(t)}{R(t)}\right)^2 x^2\,d\mu_t(x).$$

Since this last quantity is exactly the norm of v_t in $L^2(\mu_t)$, this proves that v is the tangent field to the curve μ. □

Remark 5.3.2. In the case $p \neq 2$ the same result is true, but one has to use the characterization of tangent velocity fields in terms of closure of gradients of smooth maps, see Proposition 8.4.5 of [4].

A first result we prove is the following:

Theorem 5.3.3. *If* (u, v) *is a self-similar solution of the system* (5.2.4) *with* u *Lipschitz continuous, then necessarily* u *is of the form*

$$u(t, x) = (A_t - B_t|x|^2) \vee 0 \quad \textit{for suitable coefficients } A_t,\ B_t > 0.$$

Proof. We look at the equation (5.2.7) with $p = q = 2$, which is valid on $\{u > 0\}$, and we freeze time, i.e. we look at the resulting space equation for fixed t. We use the fact that v is of the form $v_t(x) = c_t x$, which implies that all the terms v, $v \cdot \nabla v$ and dv/dt are of the same form. This easily implies that also ∇u is of the same form. Hence, at time t, on $\{u > 0\}$, it holds $u(x) = A_t - B_t x^2$, where a priori B_t could also be negative. Anyway we can prove that B_t cannot be negative. In this case in fact, if Ω were a convex unbounded domain, then u could not be the density of a probability measure. On the other hand one can easily see that on bounded convex domains Ω self-similar solutions must vanish on $\partial\Omega$, otherwise we should get a jump of the density at the boundary of $\{u > 0\}$ when rescaling, but u was supposed to be Lipschitz (except in the case that the solution is constant in time). This implies that also in the case of a bounded Ω the coefficient B_t must be positive. For the same continuity reason we get that the region $\{u > 0\}$ must agree with the region $\Omega \cap \{A_t - B_t x^2 > 0\}$ in order to have continuity of u, and this proves the formula. □

Remark 5.3.4. A similar result could be obtained for generic Wasserstein spaces with exponent $p > 1$, getting that any self-similar solution should be of the form $u(t, x) = (A_t - B_t|x|^p) \vee 0$.

Theorem 5.3.5. *If* $\overline{\mu}$ *is a probability measure on* Ω *with density*

$$u(x) = A[(R^2 - |x|^2) \vee 0],$$

then for any regular and monotone function $R : [0, 1] \to [0, 1]$ *the curve* $\mu_t = (T_{R(t)})_\sharp \overline{\mu}$ *is a solution to System* (5.2.4) *together with its tangent field* v.

Proof. It is sufficient to check the first vector equation in the system (5.2.8). First we compute the correct constant A: we must have

$$1 = A \int_0^R (R^2 - r^2) d\omega_d r^{d-1} dr = AR^{d+2}\omega_d \frac{2}{d+2},$$

and hence $A = R^{-d-2}(d+2)/(2\omega_d)$. This allows us to compute the term $F(t)$:

$$F = A^2 \int_0^R (R^2 - r^2)^2 d\omega_d r^{d-1} dr = R^{-d} \frac{2(d+2)}{(d+4)\omega_d}.$$

Then we compute V by recalling that $v_t(x) = xR'(t)/R(t)$. It holds

$$V = \left(\frac{R'}{R}\right)^2 A \int_0^R r^2 (R^2 - r^2) d\omega_d r^{d-1} dr = \frac{d}{d+4}(R')^2.$$

We must also compute dv/dt and $v \cdot \nabla v$:

$$\frac{\partial v}{\partial t} = x \frac{R''R - (R')^2}{R^2}; \quad \nabla v = \left(\frac{R'}{R}\right) I; \quad v \cdot \nabla v = \left(\frac{R'}{R}\right)^2 x.$$

We compute now

$$K(t) = F(t)V(t)^{-1/2} = R^{-d}|R'|^{-1} \frac{2(d+2)}{\sqrt{d(d+4)}\omega_d},$$

$$K'(t) = \operatorname{sign}(R')(-dR^{-d-1} - R^{-d}(R')^{-2}R'') \frac{2(d+2)}{\sqrt{d(d+4)}\omega_d}.$$

If we set $c = \operatorname{sign}(R') \frac{2(d+2)}{\sqrt{d(d+4)}\omega_d}$ we have $K = cR^{-d}(R')^{-1}$ and $K' = c(-dR^{-d-1} - R^{-d}(R')^{-2}R'')$, but also $-2V^{1/2}\nabla u(x) = cdR'R^{-d-2}x$. Inserting everything in the equation we must check that

$$dR'xR^{-d-2} + R^{-d}(R')^{-1}x\frac{R''}{R} - (dR^{-d-1} + R^{-d}(R')^{-2}R'')x\frac{R'}{R} = 0.$$

The proof is achieved as this last equation is (miraculously enough) always satisfied. □

Remark 5.3.6. By a similar proof we can show that, for $p \neq 2$, if $\overline{\mu}$ has a density of the form $u(x) = A[(R^p - |x|^p) \vee 0]$, then μ gives raise to a self-similar solution.

Remark 5.3.7. This kind of self-similar solutions can join two different probability measures which are homothetic, and in particular arrive up to the Dirac mass δ_0. Anyway it is not in general possible to link a measure to δ_0 by a curve with finite energy: in [20], conditions to ensure this possibility are provided, but in general they are not satisfied in the case $q = 2$.

Remark 5.3.8. It is interesting to notice that this self-similar solutions are density of the same kind of the optimal solutions of the problem presented in Chapter 1. This comes from the fact that the measures μ_t are chosen by minimizing a combination of an F_q functional and some Wasserstein distances from the measures μ_s, where μ_s are on the Wasserstein geodesic linking μ_t to δ_0 (and in Chapter one we exactly minimized functionals of the kind $F_q + W_p^p(\delta_{x_i}, \cdot)$).

5.3.2 Moving self-similar solutions

We have characterized all the self-similar solutions which link two homothetic probability measures. It is however interesting to look also at the moving self-similar solutions, i.e. at solutions obtained by homotheties and translations together.

In this case we consider a reference measure $\overline{\mu}$ and we look for solutions of the form $(T^t)_\sharp\overline{\mu}$, where $T^t(x) = R(t)x + \overline{x}(t)$. It is not difficult to replace Lemma 5.3.1 with the following:

Lemma 5.3.9. *If μ is a curve of the form $\mu_t = (T^t)_\sharp\overline{\mu}$, then its tangent field is given by*

$$v_t(x) = \frac{R'(t)}{R(t)}(x - \overline{x}(t)) + \overline{x}'(t).$$

Proof. The result may be proved very similarly to Lemma 5.3.1: it is sufficient to check the continuity equation

$$\frac{d}{dt}\int_\Omega \phi(R(t)x + \overline{x}(t))\,d\mu(x) = \int_\Omega \nabla\phi(R(t)x + \overline{x}(t))\cdot(R'(t)x + \overline{x}'(t))\,d\mu(x)$$
$$= \int_\Omega \nabla\phi(R(t)x + \overline{x}(t))\cdot\frac{R'(t)}{R(t)}(R(t)x + \overline{x}'(t))\,d\mu(x) = \int_\Omega \nabla\phi\cdot v_t\,d\mu_t,$$

and then to check the optimality of the norm by the fact that the map

$$x \mapsto \frac{R(t+h)}{R(t)}(x - \overline{x}(t)) + \overline{x}(t+h)$$

transports μ_t on μ_{t+h} and is optimal, and that

$$\frac{1}{h^2}\int_\Omega \left(\frac{R(t+h)}{R(t)}(x-\overline{x}(t)) + \overline{x}(t+h) - x\right)^2 d\mu_t(x)$$

converges to

$$\int_\Omega \left(\frac{R'(t)}{R(t)}(x-\overline{x}(t)) + \overline{x}'(t)\right)^2 d\mu_t(x) = \|v_t\|_{L^2(\mu_t)}. \quad \square$$

For computational simplicity we consider moving self-similar solutions only under a special reparameterization.

Theorem 5.3.10. *If $\overline{\mu}$ is a probability measure on Ω with density*

$$u(x) = A[(R^2 - |x|^2) \vee 0]$$

and $\overline{x}(0)$, $\overline{x}(1) \in \Omega$ are assigned, a curve $\mu_t = (T^t)_\sharp\overline{\mu}$, parameterized so that $K = FV^{-1/2}$ is constant, is a moving self-similar solution (solving System (5.2.7) *together with its own tangent field) if and only if the vector x moves on the straight line segment from $\overline{x}(0)$ to $\overline{x}(1)$ with constant speed and R is a strictly concave function of t. This means*

$$\overline{x}'' = 0; \quad R^{2d}(d(R')^2 + (d+4)(\overline{x}')^2) \text{ is constant and } R \text{ strictly concave.}$$

Proof. We only need to find conditions for the first equation to be satisfied. We re-write in this case the quantity considered in Theorem 5.3.5: first we compute

$$u(x) = A[(R^2 - |x-\overline{x}|^2) \vee 0]; \;\; A = \frac{(d+2)}{2R^{d+2}\omega_d}; \;\; \nabla u(x) = -\frac{(d+2)}{R^{d+2}\omega_d}(x-\overline{x})$$
$$F = R^{-d}\frac{2(d+2)}{(d+4)\omega_d}; \quad V = \frac{d}{d+4}(R')^2 + (\overline{x}')^2.$$

We have used the fact that u_t is symmetric around $\overline{x}(t)$ and hence there is no mixed term $(x-\overline{x}(t))\cdot\overline{x}'(t)$ in computing $V(t)$. Then we go on with dv/dt and $v\cdot\nabla v$:

$$\frac{\partial v}{\partial t} = (x-\overline{x})\frac{R''R - (R')^2}{R^2} - \overline{x}'\frac{R'}{R} + \overline{x}''; \quad \nabla v = \left(\frac{R'}{R}\right)\mathbb{I};$$
$$v\cdot\nabla v = \left(\frac{R'}{R}\right)^2 (x-\overline{x}) + \frac{R'}{R}\overline{x}' \quad \frac{\partial v}{\partial t} + v\cdot\nabla v = (x-\overline{x})\frac{R''}{R} + \overline{x}''.$$

Then we look at the the condition to have $K'(t) = 0$, which is equivalent to $F^{-2}V$ being constant, and thus $R^{2d}(d(R')^2 + (d+4)(\overline{x}')^2)$ being constant. Assuming K to be constant we try to satisfy the equation, and we write it in the following form that we can reach after multiplying by $V^{1/2}$:

$$-2V\nabla u + F\left(\frac{\partial v}{\partial t} + \frac{1}{2}v \cdot \nabla v\right) = 0.$$

This equation becomes

$$2\left(\frac{d}{d+4}(R')^2 + (\overline{x}')^2\right)\frac{(d+2)}{\omega_d R^{d+2}}(x-\overline{x}(t)) + R^{-d}\frac{2(d+2)}{(d+4)\omega_d}((x-\overline{x})\frac{R''}{R} + \overline{x}'') = 0.$$

To satisfy this equation it is necessary and sufficient that the two parts, the one involving $x - \overline{x}$ and the other with $\overline{x}''$ both vanish. After simplifying we get

$$R^{-2}(d(R')^2 + (d+4)(\overline{x}')^2) + \frac{R''}{R} = 0; \qquad \overline{x}'' = 0.$$

Hence we must have $\overline{x}(t) = (1-t)\overline{x}(0) + t\overline{x}(1)$ and $\overline{x}'(t) = e = \overline{x}(1) - \overline{x}(0)$. Now we recall that $R^{2d}(d(R')^2 + (d+4)(\overline{x}')^2)$ was assumed to be constant and so $d(R')^2 + (d+4)(\overline{x}')^2 = CR^{-2d}$. Hence we get $R'' = -CR^{-2d-1}$. Thus, u is a moving self-similar solutions if and only if the following conditions simultaneously hold:

$$\begin{cases} d(R')^2 + (d+4)e^2 = CR^{-2d} & \text{for a certain } C, \\ R'' = -CR^{-2d-1} & \text{for the same } C, \\ \overline{x}(t) = \overline{x}(0) + te. \end{cases}$$

By differentiating the first equation we get $2dR'R'' = -2dCR^{-2d-1}R'$ and hence the second is automatically satisfied, provided we can ensure that $R' \neq 0$ a.e. This means that R being strict concave is sufficient (it is not possible to have more than a time where R' vanishes), but it is also necessary from the second equation. The result is then proved. □

Chapter 6

Branching transport problems and distances

In this chapter we introduce the subject of branching transport problems, which was the motivation for the geodesic approach of Chapter 4 (concentration case). We start by the Eulerian formulation which was given in the '60s by Gilbert in the discrete case and then generalized to the continuum by Xia. In this chapter we are interested in the minimum value of the variational problems that arise and not in the features of the corresponding optimal structures. We will address the question of the finiteness of the minimum and of the distance d_α induced by it. We will also present the Lagrangian models introduced by Maddalena, Solimini, Morel, Bernot and Caselles. We will take advantage of this to propose an informal comparison between these models and the concentrated case of Chapter 4. The final part of the chapter contains a proof of an inequality between d_α and the classical Wasserstein distance, from a recent work in collaboration with Jean-Michel Morel ([60]).

6.1 Eulerian models by Gilbert and Xia

Lots of branching structures transporting different kind of fluids, such as road systems, communication networks, river basins, blood vessels, leaves and trees and so on, may be easily thought of as coming from a variational principle. They appear when transport costs encourage joint transportation. Recently these problems received a lot of attention by mathematicians, but in fact a mathematical formalization for them is very classical and has been performed first for atomic measures and then generalized. We briefly present here the problem introduced by Gilbert in [48] and [49], where it is presented as an extension of Steiner's minimal length problems. The main applications that Gilbert referred to were in the field of communication networks and the energy to be minimized represents the costs for building the network.

Given two finitely atomic probability measures $\mu = \sum_{i=1}^{m} a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^{n} b_j \delta_{y_j}$, consider

$$(P_G) \qquad \min\ E(G) := \sum_h w_h^\alpha \mathcal{H}^1(e_h), \tag{6.1.1}$$

where the infimum is among all weighted oriented graphs $G = (e_h,\, \hat{e}_h,\, w_h)_h$ (where e_h are the edges, $\hat{e}_h$ represent their orientations and w_h the weights) satisfying Kirchhoff's Law: in each segment vertex which is not one of the x_i's or y_j's the total incoming mass equals the outcoming, while in each x_i we have

$$a_i + \text{ incoming mass } = \text{ outcoming mass}$$

and, conversely, in each y_j we have

$$\text{incoming mass } = \text{ outcoming mass} + b_j.$$

These conditions correspond exactly to the well known Kirchhoff Law for electric circuits. The orientations $\hat{e}_h$ do not appear in the energy E but appear in fact in Kirchhoff constraints. The exponent α is a fixed parameter $0 < \alpha < 1$ so that the function $t \mapsto t^\alpha$ is concave and subadditive. In this way larger links bringing the mass from μ to ν are preferred to several smaller links transporting the same total mass. It is not difficult to check that the energy of any finite graph may be improved if we remove cycles from the graph. In this way we can minimize among finite graphs which are actually trees. This implies a bound on the number of edges and hence ensures a suitable compactness which is enough to prove existence of a minimizer.

More recently Xia, in [72], has proposed a new formalization leading to generalizations of this problem to arbitrary probability measures μ and ν. In this case the interest of the author of [72] is to view this problem as an extension of Monge-Kantorovich optimal transport theory. Actually Steiner and Monge's problems represent the limit cases $\alpha = 0$ and $\alpha = 1$, respectively.

Let us briefly see how Xia extended to the continuous case the discrete irrigation model proposed by Gilbert. The key point is formalizing the problem by using measures (or currents), since the constraint on the incoming and outcoming masses in each vertex (Kirchoff Law) may be easily written as $\nabla \cdot \lambda_G = \mu - \nu$, where $\lambda_G = \sum_h w_h[[e_h]]$ is a vector measure ($[[e]]$ being the integration measure measure on the segment e: $[[e]] = \hat{e} \cdot \mathcal{H}^1 \llcorner e$). This consideration lead Xia in [72] to extend the problem by relaxation to generic probabilities μ and ν. The problems becomes

$$(P_X) \qquad\qquad \min\ \overline{E}(\lambda) : \ \nabla \cdot \lambda = \mu - \nu$$

where

$$\overline{E}(\lambda) := \inf \left\{ \liminf_n E(\lambda_{G_n}) \ : \ G_n \text{ are finite graphs and } \lambda_{G_n} \rightharpoonup \lambda \right\}.$$

It is also possible to prove a representation formula for the relaxed energy $\overline{E}$: we have

$$\overline{E}(\lambda) = \begin{cases} \int_M \theta^\alpha \, d\mathcal{H}^1, & \text{if } \lambda = (M, \theta, \xi), \\ +\infty & \text{otherwise,} \end{cases} \tag{6.1.2}$$

where the equality $\lambda = (M, \theta, \xi)$ means that M is a 1−rectifiable set, θ a real multiplicity, ξ a measurable unit vector field on M tangent to M itself and λ is the vector measure $\theta\xi \cdot \mathcal{H}^1 \llcorner M$.

Notice that (P_X) means minimizing an energy $\overline{E}$ under a divergence constraint, exactly as in the minimal flow problem (Proposition 0.1.8). The difference is that, instead of minimizing the total mass of the vector measure whose divergence is prescribed, we minimize what is sometimes called its $\alpha-$mass M^α (see [73] and [62]).

It should be proven that, when μ and ν are both actually atomic measures, we retrieve the problem by Gilbert. This is not trivial, as we admitted lots of new competitors. Moreover, as our relaxation process did not keep fixed the marginal measures μ and ν, it is not even a priori clear that the infimum value has not changed. To deal with this problem we need some necessary optimality conditions: we would like to state that, once we minimize over vector measures Xia's functional, if μ and ν are themselves finitely atomic, then any minimizer must actually be a finite graph. The problem of regularity is addressed to in [73] and [14], but here we will not be concerned with it.

Another non trivial issue is understanding when the minimum value, which is always finite in the discrete case, is finite in the general case. This leads to some conditions on α and the measures μ and ν. We will resume them in next section.

6.2 Lagrangian models: traffic plans and patterns

This section is an informal summary of the models in [57] and [13] and their properties. Languages and approaches have been sometimes simplified to present them in a more concise way.

Let Ω be a fixed domain in $\mathbb{R}^d$. Let us denote by Γ the set of $1-$Lipschitz curves $\gamma : [0,+\infty[\to \Omega$ that are eventually constant. It means that, if we define the stopping time of a curve γ by

$$\sigma(\gamma) = \inf\{s \,:\, \gamma \text{ is constant on } [s,+\infty[\}\,,$$

these are curves with $\sigma(\gamma) < +\infty$. Let us also denote by Γ_{arc} the set of those curves in Γ which are parametrized by arc length and by Γ_{inj} the set of curves in Γ which are injective on $[0,\sigma(\gamma)[$. In the sequel we will often identify a curve with its image, in the sense that sometimes we will write γ instead of $\gamma([0,\sigma(\gamma)]) = \gamma([0,+\infty[)$.

Given a probability measure η on the space Γ, for any point $x \in \mathbb{R}^d$ the $\eta-$multiplicity of x is defined by

$$[x]_\eta := \eta\{\gamma \in \Gamma \,:\, x \in \gamma([0,\sigma(\gamma)])\}\,. \tag{6.2.1}$$

Then we can define

$$Z_\eta(\gamma) = \int_0^{\sigma(\gamma)} [\gamma(t)]_\eta^{\alpha-1} dt \quad\text{and}\quad J(\eta) = \int_\Gamma Z_\eta\, d\eta. \tag{6.2.2}$$

Notice that, for simplicity, here Z_η is defined without the term $|\gamma'|(t)$ which appears in the original definition in [13]. As a consequence, it will be deduced later that minimizers are actually parametrized by arc length.

Finally, we consider the maps π_0, $\pi_\infty : \Gamma \to \Omega$, given by $\pi_0(\gamma) = \gamma(0)$, and $\pi_\infty(\gamma) = \gamma(\sigma(\gamma))$. The two image measures $(\pi_0)_\sharp\eta$ and $(\pi_\infty)_\sharp\eta$, which belong to $\mathcal{P}(\Omega)$, will be called the starting and the terminal measure of η, respectively. Following the notation of [13] we may define a *traffic plan* as a measure $\eta \in \mathcal{P}(\Gamma)$ such that $\int_\Gamma \sigma(\gamma)\eta(d\gamma) < +\infty$. We will also call *pattern* a traffic plan η such that $(\pi_0)_\sharp\eta = \delta_0$. In the case of a pattern the terminal measure will also be called the measure irrigated by η

The minimization problem proposed in [13] is

$$(P) \quad \min \quad J(\eta) \,:\, \eta \text{ is a traffic plan}, (\pi_\infty)_\sharp\eta = \mu,\, (\pi_0)_\sharp\eta = \nu,$$

where μ and ν are given measures in $\mathcal{P}(\Omega)$. We also denote the set of admissible traffic plans by $TP(\nu,\mu)$. As $[\gamma(t)]_\eta \le 1$, we have $Z_\eta(\gamma) \ge \sigma(\gamma)$. Hence it is straightforward that any η such that $J(\eta) < +\infty$ is actually a traffic plan.

Definition 6.2.1. A traffic plan η which minimizes J among all the traffic plans with the same starting and terminal measures, with $J(\eta) < +\infty$, will be called an *optimal traffic plan.* In the case $\nu = \delta_0$ it will be called *optimal pattern.*

A useful tool developed in [13] (see also [12]) is the following: if η is concentrated on $\Gamma_{arc} \cap \Gamma_{inj}$ then the following remarkable formula holds:

$$J(\eta) = \int_{\mathbb{R}^d} [x]_\eta^\alpha \, \mathcal{H}^1(dx). \tag{6.2.3}$$

This formula gives an evident link with Gilbert and Xia's models.

In the next chapter we will mainly deal with the problem of optimal patterns, i.e. with the case $\nu = \delta_0$. This problem requires some extra tools and concepts that we will present in a while. Before that, let us introduce another concept which is very typical of the general traffic plan case.

Definition 6.2.2. A curve $\gamma_0 : [s_0, t_0] \to \Omega$ is said to be an *arc of* η if

$$\eta \left(\{ \gamma \in \Gamma \, : \, \gamma_0([s_0, t_0]) \subset \gamma \} \right) > 0.$$

We move now to the concepts we need to specifically deal with the case $\nu = \delta_0$.

For any $t \geq 0$ consider an equivalence relation on Γ given by "the two curves γ_1 and γ_2 are in relation at time t if they agree on the interval $[0, t]$", and denote the equivalence classes by $[\cdot]_t$, so that

$$[\gamma]_t = \{ \tilde{\gamma} \, : \, \tilde{\gamma}(s) = \gamma(s) \text{ for any } s \leq t \} \, .$$

For notational simplicity, let us set $|\gamma|_{t,\eta} := \eta([\gamma]_t)$.

Definition 6.2.3. Given $\eta \in \mathcal{P}(\Gamma)$, a curve $\gamma \in \Gamma$ is said to be $\eta-$good if

$$Z_\eta^0(\gamma) := \int_0^{\sigma(\gamma)} |\gamma|_{t,\eta}^{\alpha-1} dt < +\infty.$$

Remark 6.2.4. When $\nu = \delta_0$, the problem of minimizing the functional J^0 given by $J^0(\eta) = \int_\Gamma Z_\eta^0 d\eta$, is exactly the problem adressed in [57]. Its equivalence with the traffic plan model we are presenting here, proposed in [13], is proven in [14] and in [56] and relies on optimality conditions.

Remark 6.2.5. Other intermediate models may be introduced, all differing in the definition of the multiplicity of the curve γ at time t. See for instance [15] or [56].

Here are now the most important optimality results that can be found in [57], [13], [12], [14] and [56] or easily deduced from them.

1. Problem (P) admits a solution, provided the infimum is finite (i.e. there is at least a solution with finite energy).

2. If η is an optimal traffic plan, then η is concentrated on $\Gamma_{arc} \cap \Gamma_{inj}$. In particular, we may apply formula (6.2.3) for J.

3. Suppose that η is an optimal traffic plan, that two curves $\gamma_0, \gamma_1 \in \Gamma_{arc} \cap \Gamma_{inj}$ meet twice (i.e. $\gamma_0(s_0) = \gamma_1(s_1)$, $\gamma_0(t_0) = \gamma_1(t_1)$ and $s_i \neq t_i$) and that γ_0 on the interval $[s_0, t_0]$ ia an arc of η. Then either both curves coincide in the trajectory between the two common points or we have $\int_{s_0}^{t_0} [\gamma_0(t)]_\eta^{\alpha-1} dt < \int_{s_1}^{t_1} [\gamma_1(t)]_\eta^{\alpha-1} dt$. In particular two different arcs of η cannot part and then meet again.

4. If η is an optimal pattern (in particular $\nu = \delta_0$), then for $\eta-$a.e. curve γ and a.e. $t < \sigma(\gamma)$ we have $[\gamma(t)]_\eta = |\gamma|_{t,\eta}$. Roughly speaking this means that if all the mass starts from a common point then there is no parting-and-meeting-again-later (this is the single path property described in [14].

5. As a consequence, any optimal pattern η is concentrated on the set of $\eta-$good curves, and any $\eta-$good curve γ belongs to $\Gamma_{arc} \cap \Gamma_{inj}$ and satisfies $[\gamma(t)]_\eta = \eta([\gamma]_t)$ for any $t < \sigma(\gamma)$.

6. Last but not least $\min(P) = \min(P_X)$, which means that the minima of the Lagrangian and of the Eulerian model coincide.

For the whole set of equivalences between the different models, see [14].

Remark 6.2.6. Notice that an optimal traffic plan η is concentrated on the set of $\eta-$good curves, but this does not mean that this set is linked to the support of η. In fact any restriction of an $\eta-$good curve is itself an $\eta-$good curve and hence, for instance, in the discrete case, we have plenty of $\eta-$good curves but the support of η is finite. In particular the set of $\eta-$good curves may be very different from the set of fibers of a traffic plan that we find in [13] or [14] and does not depend on any parametrization χ, but it is more intrinsic.

Remark 6.2.7. These Lagrangian models may be useful to understand differences and similarities with the concentration case of Chapter 4. In fact it is easy to realize that, even in simple cases such as discrete ones, the way the two models combine length and masses are different. In fact, in the case where some masses $(m_i)_i$ are transported each one on a segment whose length is l_i, in the Xia (or traffic plan or pattern) model the cost is $\sum_i m_i^\alpha l_i$ while in the concentration case of Chapter 4 is $\left(\sum_i m_i^\alpha\right)\left(\sum_i m_i l_i^p\right)^{1/p}$. But the situation changes if we take $p=\infty$ and this is the reason why we insisted on the case of the space $\mathcal{W}_\infty$ in Section 0.2 and in Chapter 4. In fact, if we take a Lipschitz curve μ in $\mathcal{W}_\infty$ (and we can think a $1-$Lipschitz curve up to reparameterization on a different interval), in analogy to Theorem 0.3.3, we may think that there is a velocity field v with $||v|| \leq 1$ and that there is a measure η on Γ (concentrated on solutions of the ODE associated to the vector field v, i.e. on $1-$Lipschitz curves) such that $\mu_t = (\pi_t)_\sharp \eta$ (this is suggested by some results in [54], but it has to be proven). For simplicty let us have a look at the pattern case, i.e. $\mu_0 = \delta_0$. In terms of η the two models give a cost at time t which is $\int_\Gamma |\gamma|_{t,\eta}^{\alpha-1} I_{t<\sigma(\gamma)} \eta(d\gamma) = \sum_{i \in I(t)} m_i^\alpha$ for one and $G_\alpha(\mu_t) = \sum_i m_i^\alpha$ for the other. Here $m_i = \eta([\gamma_i]_t)$ and the curves γ_i are representatives of the equivalence classes of time t, the set $I(t)$ denoting those indexes such that the corresponding classes have not yet stopped. Due to the optimality condition 3 these masses correspond to the masses of the atoms of μ_t (in the sense that two $\eta-$good curves arrive at time t at the same point if and only if they have stayed together from time 0). This shows that the only difference between the two models is the fact that in the model concerning curves in $\mathcal{W}_\infty$ we take into account in the cost also the masses that have stopped. This is in fact the main difference, which is due to the fact that the cost at time t is chosen to depend only on the configuration of masses at time t. It is the price to be paid, having a less accurate and less realistic model, in order to have it mathematically simpler (as a particular case of an abstract geodesic problem).

6.3 Irrigation costs and their finiteness

The minimum value of (P_X) (or of (P)), which obviously depends on μ and ν, will be denoted by $d_\alpha(\mu,\nu)$. About its finiteness, there are results on α ensuring $d_\alpha(\mu,\nu) < +\infty$ for any pair of probabilities (μ,ν) and results concerning the two measures as well, and in particular sort of their dimension.

We know that in the case $\alpha = 1$ any pair of compactly supported measures may be linked with finite energy, because we are actually facing the Monge-Kantorovich problem. It is proven in [72] that, when α is sufficiently close to 1, namely $\alpha > 1 - 1/d$, the minimum stays finite for any pair (μ,ν). This is obtained by means of a dyadic construction which is very similar to the one we did in Chapter 4 (actually, its our construction which is very similar to the one performed by Xia). Moreover the following uniform estimate (see [72]) holds

$$d_\alpha(\mu,\nu) \le C_{\alpha,d}\,\mathrm{diam}(\Omega). \tag{6.3.1}$$

It is not difficult to extend the whole model to the case of finite measures instead of probabilities, thus getting, when μ and ν are two measures with the same mass m,

$$d_\alpha(\mu,\nu) \le C_{\alpha,d} m^\alpha\,\mathrm{diam}(\Omega). \tag{6.3.2}$$

From (6.3.2) and the fact that the distance d_α depends only on $\mu - \nu$ we can deduce a sharper estimate which refines (6.3.1), namely

$$d_\alpha(\mu,\nu) \le C_{\alpha,d}\,\delta^\alpha\,\mathrm{diam}(\omega), \tag{6.3.3}$$

whenever $\mu-\nu = \delta(\mu'-\nu')$ and μ' and ν' are probability measures on $\omega \subset \Omega$ (i.e. we have taken into account the possibility that the two measures differ only on a small set and the mass of the difference is small).

In dimension one this means that for $\alpha > 0$ there is finiteness of the minimum. For $\alpha = 0$ the problem reduces to a length minimization and in the particular case of $d = 1$ this has always a finite solution.

In larger dimensions, however, when α is below this threshold there are pairs of measures which are not linkable by a finite energy configuration. Since in order to link μ to ν and estimate $d_\alpha(\mu,\nu)$, we can always decide to link μ to δ_0 and then δ_0 to ν, we will give the following definition.

Definition 6.3.1. A measure μ is called $\alpha-$*irrigable* if $d_\alpha(\mu,\delta_0) < +\infty$. The quantity $d_\alpha(\mu,\delta_0)$ will also be denoted by $X_\alpha(\mu)$.

In the case $d > 1$ and $\alpha < 1 - 1/d$, for a measure μ being $\alpha-$irrigable is a fact somehow linked to its "dimension". The proofs are in [41] and [43] and give both irrigability and non-irrigability results. In view of the fact that, for lots of applications, it is very interesting to deal with the case of the Lebesgue measure on Ω, we will here presents only the results which are relevant for such a case.

Proposition 6.3.2. *If μ is $\alpha-$irrigable, then μ is concentrated on a set which is $\mathcal{H}^{d(\alpha)}-$negligible, where $d(\alpha) = 1/(1-\alpha)$. In particular the Lebesgue measure is not $\alpha-$irrigable for $\alpha \le 1 - 1/d$.*

We do not provide here the complete proof of this fact, but we want to give a proof of the fact that a measure whose density with respect to the Lebesgue measure is bounded away from zero may not be irrigated for

$\alpha < 1 - 1/d$. It is consequently a very weak result, as it requires the strict inequality on α and very strong assumptions on the measure, but it has the advantage of using only the formulation of the problem given by Xia. This proof comes from some conversations with P. Tilli.

Theorem 6.3.3. *Suppose $\alpha < 1 - 1/d$ and that $\mu \in \mathcal{P}(\Omega)$ is such that $\mu(Q) \geq c|Q|$ for a certain $c > 0$ and any cube $Q \subset \Omega$. Then μ is not $\alpha-$irrigable.*

Proof. Let us divide Ω into small cubes Q_i of side ε, thus having approximatively $C\varepsilon^{-d}$ cubes. Inside any cube we place a subcube Q_i', with side $c\varepsilon$ ($c < 1$). We fix now two sequence of discrete probability measure μ_n and ν_n, converging to μ and δ_0 respectively, such that $d_\alpha(\mu, \nu) = \liminf_n d_\alpha(\mu_n, \nu_n)$. Once we fix the sequence and the cubes, we will eventually have $\mu_n(Q_i') \geq C_1\varepsilon^d$ and $\nu_n(Q_i) \leq C_2\varepsilon^d$, for $C_1 > C_2$ and any index i up to the one for which we have $0 \in Q_i$. Hence we may deduce that, in the optimal discrete graph linking μ_n to ν_n, for all the indexes i but one, there should be at least a mass $(C_2 - C_1)\varepsilon^d$ passing through the region $Q_i \setminus Q_i'$. Since the distance to be covered is at least $(1 - c)\varepsilon$, the energy of the part of the graph contained in $Q_i \setminus Q_i'$ must be at least $C\varepsilon^{1+d\alpha}$. The total energy is hence at least $C\varepsilon^{1+d(\alpha-1)}$. We can deduce $d_\alpha(\mu, \nu) \geq C\varepsilon^{1+d(\alpha-1)}$ and, being ε arbitrary and $1 + d(\alpha - 1) < 0$, we get $d_\alpha(\mu, \nu) = +\infty$. □

Remark 6.3.4. In the previous proof, in the case $\alpha = 1 - 1/d$ we could not get the result. Anyway, notice that the energy has been hugely underestimated, as a consequence of the fact that in any cube Q_i only the contribution of the mass coming from Q_i' has been considered, while for most of the cubes this could be negligible with respect to the mass arriving from other cubes.

Remark 6.3.5. Notice that the threshold $1 - 1/d$ is the same which appears in Chapter 4 for the concentration case.

6.4 The d_α distance and its comparison with W_1

In [72] it is proven that, for $\alpha > 1 - 1/d$, the quantity d_α defines a new distance over the space of probability measures $\mathcal{P}(\Omega)$, which induces the weak topology and endows $\mathcal{P}(\Omega)$ with a structure of length space.

It is natural, as the branching transport problem (P_X) comes from a variant of Monge's problem, to compare the distance arising here (d_α) and the one coming from Monge-Kantorovich theory (W_1). As far as now we know that the two distances induce the same topology on $\mathcal{P}(\Omega)$, which is the same induced by the weak convergence, and it is easily checked ([72]) that $W_1 \leq d_\alpha$. The purpose of this Section is to give a sharp quantitative estimate of the kind $d_\alpha \leq C(W_1)^\beta$. This question was raised as a conjecture by Cedric Villani while reviewing the PhD Thesis [12]. Such an inequality would give an a priori estimate on d^α which is, by the way, numerically relevant. Indeed W_1 is much easier to compute by linear programming than d_α, which involves a non-convex optimization problem.

This estimate, as we avoid using previous results on the topology induced by these distances (i.e. no density argument) gives a direct and quantitative proof of the equivalence between the weak convergence topology and the topology defined by d_α. In fact the only properties on d_α we will use are (6.3.1), (6.3.2) and (6.3.3).

To fix the ideas, we consider two probability measures μ and ν with support in a d-dimensional cube C with edge 1, say $C = [0,1]^d$. It is not difficult to scale the result to any bounded domain in $\mathbb{R}^d$.

Proposition 6.4.1. *The following inequality holds for* $1 > \alpha > 1 - \frac{1}{d}$*:*

$$d_\alpha(\mu, \nu) \leq c W_1(\mu, \nu)^{d(\alpha-(1-1/d))},$$

where c denotes a suitable constant depending only on d and α.

We shall see in Example 6.4.2 that this inequality is sharp.

Proof. Let $\pi_0 \in \mathcal{P}(C \times C)$ be an optimal transport plan between μ and ν. We denote by p^+ and p^- the two projections from $C \times C$ onto C, so that $p^+(x,y) = x$, $p^-(x,y) = y$ and $(p^+)_\sharp \pi_0 = \mu$ and $(p^-)_\sharp \pi_0 = \nu$. In what follows we set $\delta = W_1(\mu,\nu)$ and

$$E_i = \left\{ (x,y) \in C \times C = \Omega,\ (2^i - 1)\frac{\delta}{2} \leq |x-y| < (2^{i+1}-1)\frac{\delta}{2} \right\}.$$

We can limit ourselves to consider those indexes i which are not too large, i.e. up to $(2^{i+1}-1)\frac{\delta}{2} \leq \sqrt{d}$ ($\sqrt{d}$ being the diameter of C). Let I be the maximal index i so that this inequality is satisfied. $C \times C = \cup_{i=0}^I E_i$ is a disjoint union and

$$\sum_{i=0}^{I} (2^i - 1)\frac{\delta}{2}\pi_0(E_i) \leq W_1(\mu,\nu) = \delta \leq \sum_{i=0}^{I}(2^{i+1}-1)\frac{\delta}{2}\pi_0(E_i) \qquad (6.4.1)$$

We call cube with edge e any translate of $[0, e[^d$. For each $i = 0, \cdots, I$, using a regular grid in $\mathbb{R}^d$, one can cover C with disjoint cubes $C_{i,k}$ with edge $(2^{i+1}-1)\delta$. The number of the cubes in the $i-$th covering may be easily be estimated by

$$\left(\frac{1}{(2^{i+1}-1)\delta}+1\right)^d \leq \left(\frac{c}{(2^{i+1}-1)\delta}\right)^d = K(i). \tag{6.4.2}$$

For each index i, it holds $C \subset \cup_{k=1}^{K(i)} C_{i,k}$ and the cubes are disjoint. Set

$$E_{i,k} = (C_{i,k} \times C) \cap E_i, \quad \mu_{i,k} = (p^+)_\#(I_{E_{i,k}} \cdot \pi_0) \text{ and } \nu_{i,k} = (p^-)_\#(I_{E_{i,k}} \cdot \pi_0).$$

In informal terms we have just cut μ and ν into pieces: the μ_i are the pieces of μ for which the Wasserstein distance to the corresponding part ν_i of ν is of order $2^i \frac{\delta}{2}$. The measure $\mu_{i,k}$ is the part of μ_i whose support is in the cube $C_{i,k}$. What we have now gained is that each $\mu_{i,k}$ has a specified diameter of order $2^i\delta$ and is at a distance to its corresponding $\nu_{i,k}$ which is of the same order $2^i\delta$ (see picture 6.1). Let us be a bit more precise. The support of $\mu_{i,k}$ is a cube with edge $(2^i-1)\delta$. By definition of E_i, the maximum distance of a point in the support of $\nu_{i,k}$ to a point in the support of $\mu_{i,k}$ is less than $(2^{i+1}-1)\frac{\delta}{2}$. Thus the supports of $\nu_{i,k}$ and $\mu_{i,k}$ are both contained in a same cube with edge $c2^i\delta$.

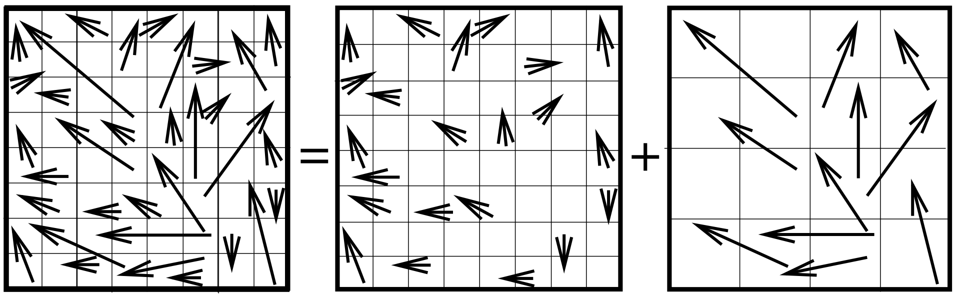

Figure 6.1: Decomposition of Monge's transportation into the sets $E_{i,k}$

By the scaling properties of the d_α distance, see (6.3.1), (6.3.2) and (6.3.3), we deduce:

$$d_\alpha(\mu_{i,k}, \nu_{i,k}) \leq c2^i\delta\pi_0(E_{i,k})^\alpha.$$

From this last relation, the sub-additivity of d_α, Hölder inequality, (6.4.1) and the bound on $K(i)$ given in (6.4.2), one obtains in turn

$$
\begin{aligned}
d_\alpha(\mu,\nu) &\le \sum_{i,k} d_\alpha(\mu_{i,k},\nu_{i,k}) \\
&\le \sum_{i,k} c2^i\delta\pi_0(E_{i,k})^\alpha = c\sum_{i,k}(2^i\delta\pi_0(E_{i,k}))^\alpha(2^i\delta)^{1-\alpha} \\
&\le c(\sum_{i,k}(2^i\delta\pi_0(E_{i,k})))^\alpha(\sum_{i,k}2^i\delta)^{1-\alpha} \\
&\le c(\sum_{i}(2^i\delta\pi_0(E_{i})))^\alpha(\sum_{i=0}^{I}K(i)2^i\delta)^{1-\alpha} \\
&\le c(\delta)^\alpha\left(\sum_{i=0}^{I}\left(\frac{c}{(2^{i+1}-1)\delta}\right)^d 2^i\delta\right)^{1-\alpha} \\
&\le c\delta^{\alpha+(1-d)(1-\alpha)}\left(\sum_{i=0}^{I}2^{i(1-d)}\right)^{1-\alpha} \\
&\le c\delta^{\alpha d-(d-1)} = cW_1(\mu,\nu)^{\alpha d-(d-1)},
\end{aligned}
$$

where c denotes various constants depending only on d and α and where the last two inequalities are valid if $d \ge 2$ so that the series $\sum_{i=0}^{\infty} 2^{i(1-d)}$ is convergent.

In the case $d = 1$ a different proof is needed. In this case we know how does an optimal transportation for $d_\alpha(\mu,\nu)$ look like. We refer to the formulation in (6.1.2), which in the one-dimensional setting gives

$$
d_\alpha(\mu,\nu) = \int_0^1 |\theta(x)|^\alpha dx.
$$

The function θ plays the role of the multiplicity and it is given by

$$
\theta(x) = \rho([0,x]), \quad \rho := \mu - \nu,
$$

as a consequence of its constraint on the derivative. Hence we have

$$
d_\alpha(\mu,\nu) = \int_0^1 |\rho([0,x])|^\alpha dx \le \left[\int_0^1 |\rho([0,x])|dx\right]^\alpha,
$$

where the inequality comes from Jensen inequality. Then we define the set $A = \{x \in [0,1] : \rho([0,x]) > 0\}$ and $h(x) = I_A(x) - I_{[0,1]\setminus A}(x)$ and we have

$$
\begin{aligned}
\int_0^1 |\rho([0,x])|dx &= \int_0^1 \rho([0,x])h(x)dx = \int_0^1 h(x)dx\int_0^1 I_{\{t\le x\}}\rho(dt) \\
&= \int_0^1 \rho(dt)\int_t^1 h(x)dx = \int_0^1 u(t)\rho(dt) \le W_1(\mu,\nu),
\end{aligned}
$$

where $u(t) = \int_t^1 h(x)dx$ is a Lipschitz continuous function whose Lipschitz constant does not exceed 1 as a consequence of $|h(x)| \leq 1$. Thus the last inequality is justified by the duality formula (see Section 0.1, (0.1.2)). Hence it follows easily $d_\alpha(\mu,\nu) \leq W_1(\mu,\nu)^\alpha$, which is the thesis for the one dimensional case. □

As we announced, the result in Proposition 6.4.1 is sharp as far as estimates of d_α in terms of W_1 are concerned. The assumption $\alpha > 1 - 1/d$ cannot be removed: for $d \geq 2$, if we remove it, the quantity d_α could be infinite while W_1 is always finite; in dimension one the only uncovered case is $\alpha = 0$. In this case d_α is in fact always finite but, for instance if $\mu = \delta_0$ and $\nu = (1-\varepsilon)\delta_0 + \varepsilon\delta_1$ we have $d_\alpha(\mu,\nu) = 1$ while $W_1(\mu,\nu) = \varepsilon$. As ε is as small as we want, this excludes any desired inequality. Moreover, the exponent $d(\alpha - (1-1/d))$ cannot be improved as can be seen from the following example.

Example 6.4.2. There exists a sequence of pairs of probability measures (μ_n, ν_n) on the cube C such that

$$d_\alpha(\mu_n,\nu_n) = cn^{-d(\alpha-(1-1/d))} \text{ and } W_1(\mu_n,\nu_n) = c/n.$$

Proof. It is sufficient to divide the cube C into n^d small cubes of edge $1/n$ and to set $\mu_n = \sum_{i=1}^{n^d} \frac{1}{n^d}\delta_{x_i}$ and $\nu_n = \sum_{i=1}^{n^d} \frac{1}{n^d}\delta_{y_i}$, where each x_i is a vertex of one of the n^d cubes (let us say the one with minimal sum of the $d-$coordinates) and the corresponding y_i is the center of the same cube. In this way y_i is one of the points y_j which are the closest to x_i. Thus the optimal configuration both for d_α and W_1 is given by linking any x_i directly to the corresponding y_i. In this way we have

$$\begin{aligned} d_\alpha(\mu_n,\nu_n) &= n^d\left(\frac{1}{n^d}\right)^\alpha \frac{c}{n} = cn^{-d(\alpha-(1-1/d))} \\ W_1(\mu_n,\nu_n) &= n^d\frac{1}{n^d}\frac{c}{n} = \frac{c}{n}. \end{aligned}$$ □

Chapter 7

Landscape function

In this chapter we propose an interesting feature of branching transport, which is a function that is associated to a branching transport problem with one source, and corresponds somehow to the Kantorovich potential of Monge's transport. This chapter follows an accepted paper ([68]) whose motivations lie in different applications. Some of the preliminaries which are needed (mainly we have to look at the Lagrangian formulations in [57] and [13]) have been moved to Chapter 6, so that this chapter is much landscape-focused.

7.1 Motivations

In this chapter we discuss some features of optimal branching structures which are crucial in river basins applications, but we address also applications to other fields.

7.1.1 Landscape equilibrium and OCNs in geophysics

It is interesting to see how people working in geophysics arrive in the study of river basins to some problems which are very similar to the models presented in Chapter 6. There is a wide literature on this geophysical point of view and a quite comprehensive reference is [64]. The specific subject dealt with here is developed both in [64] and in [9] (this last paper being our main reference, but a short previous summary of these ideas can be found in [8] as well).

When studying the configuration of a river basin, the main objects are two: the landscape elevation, which is a function z giving the altitude of any point of the region, and a river network N, which is the datum of all the streams that concur to bring water (which falls on the region as rain) to a single point (where a lake is supposed to be present). A first link between the two objects is the fact that at any point the direction followed by water is the direction of steepest descent of z. Hence, once we know z we are able to deduce N and to compute the multiplicity $\theta(x)$ at any point xç this is the quantity of water passing through x while following the steepest descent lines of z. At first the interest is towards an evolution model, which allows z and N (and hence θ) to depend on time as well. The evolution of z is ruled by an erosion equation of the form

$$\frac{\partial z}{\partial t} = -\theta|\nabla z|^2 + c, \tag{7.1.1}$$

where ∇z is the spatial gradient of z and c is a positive constant. The idea is that the erosion effect increases both with the quantity of water and with the slope. The constant c is called *uplift* and takes care of the fact that all the material brought down by erosion is in the end uniformly redistributed from below in the whole region as a geomorphological effect. Equation (7.1.1) is in fact a simplified version of other more general evolution equations involving higher order terms. The following phenomenon concerning solutions of (7.1.1) can be empirically observed: approximately, up to a certain time scale both z and θ (i.e. N) move, in a very strong erosional evolution; then, up to a larger time scale the network is almost constant, letting $\theta(x,t) = \theta(x)$ depend on the position only, and the landscape function evolves without changing its lines of maximal slope; finally there is a much larger time scale such that z approximatively agrees with a landscape equilibrium, i.e. a stationary solution of (7.1.1). We are interested in studying landscape equilibria. In this case the steepest descent condition, that we can read as "∇z follows the direction of the network", is completed by a second one which we get by imposing $\partial z/\partial t = 0$ in (7.1.1). This leads to $|\nabla z| = c^{1/2}\theta^{-1/2}$ and this last condition is called *slope-discharge relation*. It is explicitly suggested in [9] that in (7.1.1) one could change the exponents of θ and $|\nabla z|$ (preserving anyway the increasing behavior with respect to both variables), thus obtaining different slope-discharge relationships. In general we get $|\nabla z| = c\,\theta^{\alpha-1}$ and the physically interesting case is when the exponent α is very close to $1/2$.

To find landscape equilibria a discretization is performed in [9] and a regular square grid is used. Functions defined on the pixels of the grid and vanishing at a given point x_0, representing the outlet, are considered, as well as networks composed by edges of the grid, directed from every point to one of the neighbors.

- As we already mentioned, the conditions on the direction of the water allow to reconstruct a network from a function. Given a function z with no local minima other than x_0, one can always follow the maximal slope paths of z.

- These paths are obtained by linking any point x of the grid to a point which realizes the minimum of z among the neighbors of z. Notice in particular that these paths are only composed by edges following the two main directions of the grid.

- In this way a network $N = N(z)$ can be deduced from z.

- On the other hand, the slope-discharge condition allows to reconstruct a function from a network N, provided it is tree-shaped.

- In order to make this reconstruction, first compute the multiplicities of the points of the network: at a point x its multiplicity $\theta(x)$ is the number of points which find x on their way to the outlet (this assumes that the quantity of rain falling down at any pixel is the same, i.e. rain falls uniformly on the grid). See also Figure 7.1.1, where the multiplicity of a point x_i is computed as the number of points in the area A_i.

- Then set $z(x_0) = 0$ and for any other point x consider the only path on N linking x_0 to x. Set $z(x) = \sum_i \theta(x_i)^{\alpha-1}$, where the x_i's are the points on the path. In Figure 7.1.1 the path linking x_0 to x is shown.

- In this way we get a function $z = z(N)$.

In general it will not be true that a function $z(N)$ has maximal slope in the direction of the network N. Finding a landscape equilibrium means exactly satisfying both conditions at a time, through a fixed point problem. The algorithm starts from a tree-shaped network N, builds the function $z(N)$, and then the new network $N' = N(z(N))$. If $N' = N$, then the landscape function $z = z(N)$ is a landscape equilibrium.

The important idea presented in [9] is the relation between landscape equilibria and Optimal Channel Networks (OCNs in literature, see for instance [65], [63] and [51]). An OCN is a network N minimizing a certain dissipated energy. The dissipated energy in a system satisfying the slope-discharge relation is the total potential energy that water loses on the network. For each pixel we have a quantity of water θ which falls down to

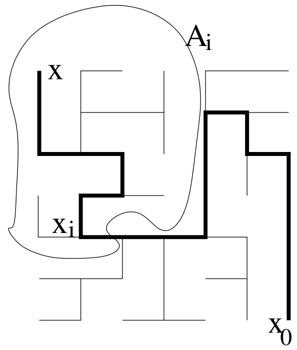

Figure 7.1: The path from x to x_0 and the multiplicity of x_i

the next pixel and its elevation decreases by a quantity which is proportional to $|\nabla z|$ and hence to $\theta^{\alpha-1}$. Hence, the total energy loss is given by $\sum_i \theta(x_i)\theta(x_i)^{\alpha-1} = \sum_i \theta(x_i)^{\alpha}$. It is clear that this energy is the same as in (6.1.1) (no length of segments is involved because in a regular grid they all have the same, given, length). What is proven in [9] is that, if N is an OCN minimizing this energy, then the landscape function $z = z(N)$ reconstructed from N is in fact an equilibrium. This actually means that not only the slope of z in the direction of the network is given by $\theta^{\alpha-1}$, which is true by construction, but also that this direction is the direction of maximal slope.

Notice that the problems studied in [9] and in the other papers on the subject have undergone a very strong discretization. In fact they correspond to solve (P_G) (see Section 6.1) where μ is a discretization on a regular grid of the Lebesgue measure and $\nu = \delta_{x_0}$, but with the extra constraint that only edges e_h which are edges of the grid are allowed. Compared to continuous models there is a loss of rotational invariance, a fixed scale effect due to the mesh, and several questions concerning the river basin may lose their meaning (for instance questions about the interfaces between two separated parts of the basin and points where the water takes two different directions, or most regularity issues). On the other hand, a continuous counterpart for the landscape function could not be simply a regular solution of (7.1.1) or of its statical version: for C^1 functions steepest descent curves are well-defined, but they never merge and therefore do not give raise to a positive multiplicity θ (except for the case $d = 1$, see [9]).

7.1.2 A landscape function appearing for derivative purposes

We will briefly see here another aspect of branching transport problems such as (P_G) where a function similar to the landscape function appears. We recall that the irrigation cost of a finite atomic measure $\mu \in \mathcal{P}(\Omega)$ is the minimum of problem (P_G) for $\nu = \delta_0$. This quantity, as in the generalization to the continuum by Xia, is denoted by $X_\alpha(\mu)$. A variational analysis of the functional X_α yields the following.

Theorem 7.1.1. *Suppose $\mu = \sum_{i=1}^m a_i \delta_{x_i}$ with $a_i > 0$ (so that the finite set $K = \{x_i : i = 1, \dots, n\}$ is actually the support of μ) and that μ_1 is another probability measure concentrated on K with $\mu_1 = \sum_{i=1}^m b_i \delta_{x_i}$. Then we have*

$$X_\alpha(\mu_1) \leq X_\alpha(\mu) + \alpha \sum_{i=1}^m z(x_i)(b_i - a_i),$$

where the function z is defined in this way: take an optimal graph G for the problem (P_G) for the measures μ and δ_0; this graph is a tree; for any x_i define

$$z(x_i) = \sum_{h \in H(i)} w_h^{\alpha-1} \mathcal{H}^1(e_h),$$

where $H(i)$ denotes the sets of indexes of the edges of the unique path from 0 to x_i.

Proof. We will build a new oriented graph which is acceptable for Problem (P_G) when irrigating μ_1. This graph will be built by using the same edges $(e_h)_h$ as in G but changing the weights w_h's. We define the new weights w'_h by

$$w'_h = w_h + \sum_{i:\, h \in H(i)} (b_i - a_i).$$

It is easy to check that this new graph satisfies the constraints, and so we get

$$X_\alpha(\mu_1) \leq \sum_h (w'_h)^\alpha \mathcal{H}^1(e_h) \leq X_\alpha(\mu) + \alpha \sum_h w_h^{\alpha-1} \sum_{i:\, h \in H(i)} (b_i - a_i),$$

where the last inequality is obtained by concavity of $t \mapsto t^\alpha$. By changing the order in performing the sums we easily get the thesis. □

Remark 7.1.2. The link between this function z and the one used in geophysics is straightforward: to compute a value $z(x)$ in fact what we do is integrating the multiplicity of the graph along the river from 0 up to x. See Figure 7.1.2 and compare with Figure 7.1.1: in this case there are in general many more degrees of freedom. The multiplicity of the represented point x_i is the total mass of the region A_i and the geometry of points and edges is not prescribed.

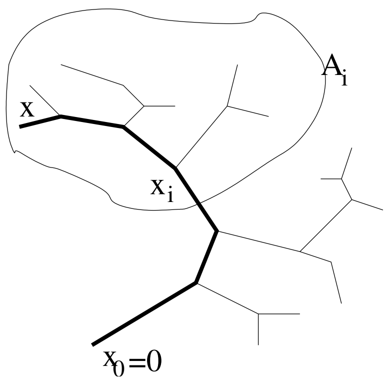

Figure 7.2: The path from x to x_0 and the multiplicity of x_i

Remark 7.1.3. As a consequence of Theorem 7.1.1, if we set $\mu_\varepsilon = \mu + \varepsilon(\mu_1 - \mu)$, we get the following inequality:

$$\limsup_{\varepsilon \to 0^+} \frac{X_\alpha(\mu_\varepsilon) - X_\alpha(\mu)}{\varepsilon} \leq \int z \, d(\mu_1 - \mu).$$

This inequality gives information on the derivative of the functional X_α and this fact is very useful in variational problems of the following kind:

$$(P_+) \qquad \min \quad X_\alpha(\mu) + F(\mu),$$

where F may be any functional whose derivative is known. We will show later an example and briefly explain the interest of these problems.

Remark 7.1.4. The result of Theorem 7.1.1 has been established under no constraints on the direction of the edges, i.e. in the setting of problem (P_G). It is easy to reproduce them in the case of grid-constrained OCNs, as in the proof there is no need to change the edges of the graph. Hence it is a result which is valid also in the setting of [9].

Our main goal is hence to define a landscape function in the continuous case and analyze its properties. We will use the models about these irrigation problems concerning arbitrary probability measures (and not only atomic ones) that we mentioned before. In particular the Lagrangian models will be very useful. We will consider the irrigation of an arbitrary measure on a domain Ω starting from a single source δ_0. Here the main problem is that the optimal structures which arise are not necessarily trees in the sense that there may be points which are reached by several curves. We will anyway propose a landscape function z and check that it is well-defined. Then we will prove that it shares all the properties that we had in the discrete case, in particular the fact that on a point x_0 of the irrigation network it has maximal slope in the direction of the network itself and that this slope is given by $\theta^{\alpha-1}$, where θ is the multiplicity of the network at x_0. Moreover we will prove that even in the continuous case an inequality on the derivative of the energy X_α involving the landscape function is available and finally we will give some continuity and semicontinuity results.

The interest of generalizing the concept of landscape function to the irrigation of arbitrary probability measures does not have only variational applications. In river basins applications, in fact, it is natural to consider directly a configuration where the starting measure is the Lebesgue measure instead of considering a grid discretization. Moreover, getting rid of the discretization will also add isotropy and other features to the models in [9] and [64].

7.2 A general development formula

As we announced we will use the framework that has been introduced in [57] and in [13] and presented in this thesis in Section 6.2. In this section we will develop in a useful way the variation of the functional J (see (6.2.2)) when passing from a traffic plan η to a traffic plan η'. Formula 6.2.3 will be crucial.

Theorem 7.2.1. *Let η and η' be probability measures on Γ and $\Delta\eta = \eta' - \eta$. Let us suppose that both η and $\Delta\eta$ are concentrated on $\Gamma_{arc} \cap \Gamma_{inj}$ and $\int_\Gamma Z_\eta \, d|\Delta\eta| < +\infty$. Then*

$$J(\eta') \leq J(\eta) + \alpha \int_\Gamma Z_\eta \, d\Delta\eta - \alpha(1-\alpha) \int_{\mathbb{R}^d} [x]^2_{\Delta\eta} \, \mathcal{H}^1(dx). \tag{7.2.1}$$

Proof. For any traffic plan η, let us set $S_\eta = \{x \in \mathbb{R}^d : [x]_\eta > 0\}$. First we prove that, under the assumptions of this theorem, we have $\mathcal{H}^1(S_{\eta'} \setminus S_\eta) = 0$. In fact, for any point $x \in S_{\eta'} \setminus S_\eta$ we have necessarily $[x]_\eta = 0$ and $[x]_{\Delta\eta} > 0$. Hence it is sufficient to prove that the integral of $[x]_{\Delta\eta}$ on this set w.r.t. $\mathcal{H}^1$

vanishes to get the desired result. We have

$$\int_{S_{\eta'}\setminus S_\eta}[x]_{\Delta\eta}\mathcal{H}^1(dx)=\int_{S_{\eta'}\setminus S_\eta}\mathcal{H}^1(dx)\int_\Gamma\Delta\eta(d\gamma)I_{x\in\gamma}=\int_\Gamma\Delta\eta(d\gamma)\mathcal{H}^1\big(\gamma\cap(S_{\eta'}\setminus S_\eta)\big).$$

The second assumption of the theorem implies that for $\Delta\eta-$a.e. curve γ the quantity $Z_\eta(\gamma)$ is finite and hence, for a.e. t, we have $\gamma(t)\in S_\eta$. Since γ is 1−Lipschitz continuous, this yields $\mathcal{H}^1(\gamma\setminus S_\eta)=0$. Hence we have $\int_{S_{\eta'}\setminus S_\eta}[x]_{\Delta\eta}\mathcal{H}^1(dx)=0$, which proves $\mathcal{H}^1(S_{\eta'}\setminus S_\eta)=0$.

Now, as both η and η' are concentrated on $\Gamma_{arc}\cap\Gamma_{inj}$, to evaluate J we can use the expression in (6.2.3) and get

$$\begin{aligned}J(\eta')&=\int_{S_\eta}\left([x]_\eta+[x]_{\Delta\eta}\right)^\alpha\mathcal{H}^1(dx)\\&\le J(\eta)+\alpha\int_{S_\eta}[x]_\eta^{\alpha-1}[x]_{\Delta\eta}\mathcal{H}^1(dx)-\alpha(1-\alpha)\int_{S_\eta}[x]_{\Delta\eta}^2\mathcal{H}^1(dx),\quad(7.2.2)\end{aligned}$$

where we have used the fact that $S_{\eta'}\subset S_\eta$ up to $\mathcal{H}^1-$negligible sets and the concavity inequalities

$$(t+s)^\alpha\le t^\alpha+\alpha t^{\alpha-1}s-\alpha(1-\alpha)(\max\{t,t+s\})^{\alpha-2}s^2\le t^\alpha+\alpha t^{\alpha-1}s-\alpha(1-\alpha)s^2$$

(this last inequality being valid when both t and $t+s$ belong to $]0,1]$).

Let us now work on the second term of the last sum we obtained. We have

$$\int_{S_\eta}[x]_\eta^{\alpha-1}[x]_{\Delta\eta}\mathcal{H}^1(dx)=\int_{S_\eta}\mathcal{H}^1(dx)\int_\Gamma\Delta\eta(d\gamma)\,[x]_\eta^{\alpha-1}I_{x\in\gamma}.$$

Here we want to change the order of integration and to do this we check what happens in absolute value:

$$\begin{aligned}\int_{S_\eta}\mathcal{H}^1(dx)\int_\Gamma|\Delta\eta|(d\gamma)[x]_\eta^{\alpha-1}I_{x\in\gamma}&=\int_\Gamma|\Delta\eta|(d\gamma)\int_{S_\eta}\mathcal{H}^1(dx)\,[x]_\eta^{\alpha-1}I_{x\in\gamma}\\&=\int_\Gamma|\Delta\eta|(d\gamma)\int_0^{\sigma(\gamma)}[\gamma(t)]_\eta^{\alpha-1}dt=\int_\Gamma Z_\eta d|\Delta\eta|<+\infty.\quad(7.2.3)\end{aligned}$$

In this series of equality, the first one is just changing the integration order, while the second relies on the fact that $|\Delta\eta|-$a.e. we have $\mathcal{H}^1(\gamma\setminus S_\eta)=0$ and, γ being parametrized by arc length, the $\mathcal{H}^1-$integral on its image may become an integral in dt on $[0,\sigma(\gamma)]$. The finiteness of the last integral in

(7.2.3) allows us to change the order of integration between $\Delta\eta$ and $\mathcal{H}^1$ and by analogous computations we get

$$\int_{S_\eta} \mathcal{H}^1(dx) \int_\Gamma \Delta\eta(d\gamma)[x]_\eta^{\alpha-1} I_{x\in\gamma} = \int_\Gamma Z_\eta \, d\Delta\eta.$$

Inserting this last equality in (7.2.2) gives the thesis. □

7.3 Landscape function: existence and applications

In this section we come specifically back to Problem (P) for $\nu = \delta_0$. Even when not specifically stated, from now on η will be an optimal pattern irrigating an $\alpha-$irrigable measure μ.

7.3.1 Well-definedness of the landscape function

First a very elementary truncation lemma is needed. As it is just the formalization of a well-known principle (that a part of an optimal structure is itself optimal), it will not be proven here. It is in fact proven in [14] when stating the optimality of the connected components of a traffic plan in $\mathbb{R}^d \setminus \{x_0\}$.

Lemma 7.3.1. *If γ_0 is a curve such that $|\gamma_0|_{t_0,\eta} > 0$, set $x_0 = \gamma_0(t_0)$, $A = [\gamma_0]_{t_0}$, $\mu_A = (\pi_\infty)_\sharp(I_A\cdot\eta)$, $\mu' = \mu - \mu_A + \eta(A)\delta_{x_0}$, $\eta' = \eta - I_A\cdot\eta + \eta(A)\delta_{\overline{\gamma_0}}$, where the curve $\overline{\gamma_0}$ is the curve γ_0 stopped at time t_0. Then η' is an optimal pattern irrigating the measure μ'.*

Theorem 7.3.2. *If γ_0 and γ_1 are two $\eta-$good curves sharing the same end-point $\overline{x}$, then $Z_\eta(\gamma_0) = Z_\eta(\gamma_1)$.*

Proof. If the two curves are identical the thesis is easily obtained. If they are not identical, then they must split at a certain time $\overline{t}$. It is possible that one of them stops at time $\overline{t}$, but not both (in this case they would be identical). Then we can choose two times t_0 and t_1 with $|\gamma_i|_{t_i,\eta} > 0$ and $\overline{t} \leq t_i \leq \sigma(\gamma_i)$ for $i = 0, 1$ (if one of the two curves stops at time $\overline{t}$, say for instance $\sigma(\gamma_0) = \overline{t}$, then we are forced to choose $t_0 = \sigma(\gamma_0) = \overline{t}$ and we have $|\gamma_0|_{t_0,\eta} = |\gamma_1|_{\overline{t},\eta} > 0$, where the inequality is a consequence of $\overline{t} < \sigma(\gamma_1)$). Figure 7.3.1 shows the two possible situations.

Let us set $x_i = \gamma_i(t_i)$ and $l = |x_1 - x_0|$. Then we use the notations of the previous Lemma and we write

$$d_\alpha(\delta_0, \mu') \leq d_\alpha(\delta_0, \mu'') + d_\alpha(\mu', \mu''), \tag{7.3.1}$$

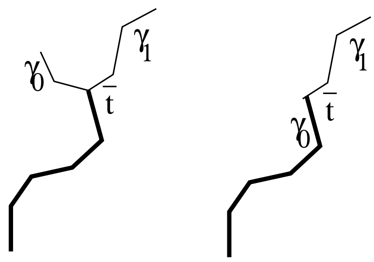

Figure 7.3: Both curves may go on after $\bar t$ or one may stop

where $\mu'' = \mu - \mu_A + \eta(A)\delta_{x_1}$. Define $\eta'' = \eta - I_A \cdot \eta + \eta(A)\delta_{\overline{\gamma_1}}$, where $\overline{\gamma_1}$ is the curve γ_1 stopped at time t_1. It is easy to check $(\pi_\infty)_\sharp \eta'' = \mu''$ and then

$$\begin{aligned} d_\alpha(\delta_0, \mu'') \le J(\eta'') &\le J(\eta') + \alpha \int_\Gamma Z_\eta \, d(\eta'' - \eta') \\ &= J(\eta') + \alpha\eta(A) \left(\int_0^{t_1} |\gamma_1|_{t,\eta}^{\alpha-1} \, dt - \int_0^{t_0} |\gamma_0|_{t,\eta}^{\alpha-1} \right). \end{aligned}$$

Here we have used Theorem 7.2.1 to estimate $J(\eta'')$. Actually by this theorem we should have had $Z_{\eta'}$ instead of Z_η. Yet we can interchange $Z_{\eta'}$ and Z_η because we have only replaced the measure η on A by a same amount of mass concentrated on $\overline{\gamma_0}$, and on $\gamma_0 \cup \gamma_1$ this does not affect multiplicities. As far as the second term of the sum in (7.3.1)is concerned it is easy to see that we have

$$d_\alpha(\mu', \mu'') \le l\eta(A)^\alpha.$$

By inserting these estimates in (7.3.1) we get

$$\int_0^{t_0} |\gamma_0|^{\alpha-1} \, dt - \int_0^{t_1} |\gamma_1|^{\alpha-1} \, dt \le \alpha^{-1} l \eta(A)^{\alpha-1}.$$

Now estimate the length l by

$$\begin{aligned} l &= |x_0 - x_1| \le |x_0 - \overline{x}| + |\overline{x} - x_1| \le (\sigma(\gamma_0) - t_0) + (\sigma(\gamma_1) - t_1) \\ &\le \eta(A)^{1-\alpha} \int_{t_0}^{\sigma(\gamma_0)} |\gamma_0|_{t,\eta}^{\alpha-1} dt + \eta(B)^{1-\alpha} \int_{t_1}^{\sigma(\gamma_1)} |\gamma_1|_{t,\eta}^{\alpha-1} dt. \end{aligned}$$

Hence

$$\int_0^{t_0} |\gamma_0|^{\alpha-1} dt - \int_0^{t_1} |\gamma_1|^{\alpha-1} dt \le \alpha^{-1} \left(\int_{t_0}^{\sigma(\gamma_0)} |\gamma_0|_{t,\eta}^{\alpha-1} dt + \frac{\eta(B)^{1-\alpha}}{\eta(A)^{1-\alpha}} \int_{t_1}^{\sigma(\gamma_1)} |\gamma_1|_{t,\eta}^{\alpha-1} dt \right).$$

Notice that we cannot have $|\gamma_i|_{\sigma(\gamma_i),\eta} > 0$ for both $i = 0, 1$, thanks to the no-loop property (property 3). So, if it is $|\gamma_1|_{\sigma(\gamma_1),\eta} = 0$, once we fix t_0 such that $\eta(A) > 0$, we can choose t_1 so that $\eta(B) \leq \eta(A)$ since $\eta(B) \to 0$ as $t_1 \to \sigma(\gamma_1)$. Otherwise, if $|\gamma_1|_{\sigma(\gamma_1),\eta} > 0$, we can choose directly $t_1 = \sigma(t_1)$. In both cases we have

$$\int_0^{t_0} |\gamma_0|^{\alpha-1}\, dt - \int_0^{t_1} |\gamma_1|^{\alpha-1}\, dt \leq \alpha^{-1} \left(\int_{t_0}^{\sigma(\gamma_0)} |\gamma_0|_{t,\eta}^{\alpha-1} dt + \int_{t_1}^{\sigma(\gamma_1)} |\gamma_1|_{t,\eta}^{\alpha-1} dt \right). \tag{7.3.2}$$

Then we let t_0 and t_1 tend to $\sigma(\gamma_0)$ and $\sigma(\gamma_1)$, according to the criteria for the choice of t_1 we have used so far, and we get at the limit

$$Z_\eta(\gamma_0) - Z_\eta(\gamma_1) \leq 0,$$

because the integral terms on the right hand side of (7.3.2) tend to zero as a consequence of the fact that γ_0 and γ_1 are both $\eta-$good curves. By interchanging the role of γ_0 and γ_1 the thesis is proven. □

Corollary 7.3.3. *If two different $\eta-$good curves γ_0 and γ_1 meet at a certain point $x = \gamma_0(t_0) = \gamma_1(t_1)$, then $|\gamma_0|_{t_0,\eta} = |\gamma_1|_{t_1,\eta} = 0$.*

Proof. If one of the two multiplicities $|\gamma_i|_{t_i,\eta}$ were positive a strict inequality between $Z_\eta(\gamma_0)$ and $Z_\eta(\gamma_1)$ should hold. Yet equality has just been proven and this is a contradiction. □

Corollary 7.3.4. *Any $\eta-$good curve γ is in fact injective on $[0, \sigma(\gamma)]$.*

Proof. The injectivity on $[0, \sigma(\gamma)[$ is already known. Hence, consider the case $\gamma(\sigma(\gamma)) = \gamma(t)$ for $t < \sigma(\gamma)$. This would imply $|\gamma|_{t,\eta} > 0$ but it is contradiction with Corollary 7.3.3, applied to the curve γ and to the curve $\overline{\gamma}$, which is γ stopped at time t. □

Remark 7.3.5. The injectivity on $[0, \sigma(\gamma)]$ was already known for $\eta-$a.e. curve γ (see [13]). Yet, it was not possible to identify an explicit class of curves sharing this property. For our purposes it is important to switch from a generic "a.e." to the fact that this is true for $\eta-$good curves.

The result of Theorem 7.3.2 allows us to define a function on Ω by the values of Z_η.

Definition 7.3.6. We define the landscape function associated to the traffic plan η as the function z_η given by

$$z_\eta(x) = \begin{cases} Z_\eta(\gamma) & \text{if } \gamma \text{ is } \eta-\text{good and } x = \gamma(\sigma(\gamma)); \\ +\infty & \text{if no } \eta-\text{good curve ends at } x. \end{cases}$$

Remark 7.3.7. It was in fact possible to prove more easily that $\mu-$a.e. the value of z was well defined (in the sense that if on a non negligible set of points x we had two different values for Z_η we would have had the possibility to strictly improve the value of J). Yet, we do not want a function z which is defined a.e. but a pointwisely defined value, to deal later with pointwise properties, being also concerned with negligible sets such as S_η.

Remark 7.3.8. Notice, as in Remark 6.2.6, that restrictions of $\eta-$good curves are still $\eta-$good and that this implies that if the landscape function is finite on a point x then it is finite also on the whole $\eta-$good curve arriving up to x.

7.3.2 Variational applications: the functional X_α

Some consequences of the existence of the landscape function are presented here.

Corollary 7.3.9. *For the functional X_α we have the following representation formula $X_\alpha(\mu) = \int_\Omega z\,d\mu$, where $z = z_\eta$ is the landscape function associated to any optimal pattern η irrigating the measure μ.*

Proof. It is sufficient to take the formula $X_\alpha(\mu) = J(\eta) = \int_\Gamma Z_\eta d\eta$ and use the fact that $Z_\eta(\gamma)$ depends only on $\pi_\infty(\gamma)$ through $Z_\eta(\gamma) = z(\pi_\infty(\gamma))$ and get

$$X_\alpha(\mu) = \int_\Gamma Z_\eta d\eta = \int_\Omega z\,d(\pi_\infty)\sharp\eta = \int_\Omega z\,d\mu.$$

□

Corollary 7.3.10. *If μ is $\alpha-$irrigable, then any landscape function z is finite $\mu-$a.e.*

Proof. Corollary 7.3.9 yields $\int z\,d\mu = X_\alpha(\mu) < +\infty$ and from this the result is straightforward. □

Remark 7.3.11. As the word "any" in the previous statement suggests, there is no uniqueness for the landscape function, and there is a landscape function for any optimal pattern.

Moreover, using Theorem 7.2.1 together with the existence of the landscape function, a derivative result extending the discrete case can be obtained. Notice that the following theorem will be useful also for other purposes, for instance when looking for continuity properties of the landscape function (see Section 6).

Theorem 7.3.12. *For a given function g on Ω, such that $||g||_{L^\infty(\mu)} \leq 1$ and $\int_\Omega g\, d\mu = 0$, set $\mu_1 = \mu(1+g)$. Then*

$$X_\alpha(\mu_1) \leq X_\alpha(\mu) + \alpha \int_\Omega z(x)g(x)\mu(dx),$$

where the function $z = z_\eta$ is the landscape function according to an arbitrary optimal pattern η irrigating the measure μ.

Proof. We will consider a variation of η given by $\eta_1 = (1 + (g \circ \pi_\infty)) \cdot \eta$. Since $(\pi_\infty)_\sharp \eta_1 = (1+g) \cdot \mu$, we have

$$X_\alpha(\mu_1) - X_\alpha(\mu) \leq J(\eta_1) - J(\eta).$$

We want to apply Theorem 7.2.1 to this situation, with $\Delta\eta = (g \circ \pi_\infty) \cdot \eta$. Since $\Delta\eta$ is absolutely continuous with respect to η with bounded density, it is straightforward that both the conditions required by the theorem ($\Delta\eta$ being concentrated on $\Gamma_{arc} \cap \Gamma_{inj}$ and Z_η being $|\Delta\eta|-$integrable) are satisfied, so that one gets

$$J(\eta') \leq J(\eta) + \alpha \int_\Gamma Z_\eta d\Delta\eta.$$

Now use the fact that Z_η depends only on its terminal point and get

$$\int_\Gamma Z_\eta d\Delta\eta = \int_\Omega z\, d((\pi_\infty)\sharp\Delta\eta) = \int_\Omega zgd\mu.$$

Putting together all the results yields the thesis. □

A simple consequence of this theorem may be expressed in terms of derivatives.

Corollary 7.3.13. *Set $\mu_\varepsilon = \mu + \varepsilon g \cdot \mu$. Then the following derivative inequality holds:*

$$\limsup_{\varepsilon \to 0^+} \frac{X_\alpha(\mu + \varepsilon g \cdot \mu) - X_\alpha(\mu)}{\varepsilon} \leq \alpha \int_\Omega z(x)g(x)\mu(dx).$$

7.3.3 A transport and concentration problem

As we said the last derivative inequality may be useful in variational problems involving X_α. For the sake of clearness we provide a short example, of the kind we referred to as (P_+) in Section 7.1.

Example 7.3.14. Let us consider the functional $F : \mathcal{P}(\Omega) \to [0, +\infty]$ given by

$$F(\mu) = \begin{cases} \int_\Omega u^2 \, d\mathcal{L}^d & \text{if } \mu = u \cdot \mathcal{L}^d, \\ +\infty & \text{if } \mu \text{ is not absolutely continuous.} \end{cases}$$

If we want to minimize (for $\alpha > 1 - 1/d$) the sum $X_\alpha(\mu) + F(\mu)$ over all probabilities μ on Ω we get as an optimality condition, by deriving and using Corollary 7.3.13,

$$\alpha z + 2u = \mathit{const} \text{ a.e. on } \{u > 0\}. \tag{7.3.3}$$

This implies several interesting properties. First of all we can notice that, both z and u being positive, they are also bounded. It was not *a priori* evident that $u \in L^\infty(\Omega)$, since the natural condition was $u \in L^2(\Omega)$. Since $z(x) \geq |x|$, this gives also an estimate on the support of μ. As the constant appearing in (7.3.3) could be uniformly estimated (it is sufficient to multiply (7.3.3) by u and integrate, thus obtaining $\mathit{const} = \alpha X_\alpha(\mu) + 2F(\mu) \leq 2\min(X_\alpha + F)$), this could also be used to prove an existence result for $\Omega = \mathbb{R}^d$ç we may also prove that actually minimizers of $X_\alpha + F$ are supported in a given bounded ball. Moreover, Formula 7.3.3 yields some regularity result on u according to the results we will prove later on z.

Variational problems such as (P_+) have been first proposed in [57], where the authors suggested to consider problems involving both the irrigation pattern χ and the irrigated measure μ_χ. Moreover they are very similar to what considered in Chapter 1 and they fall into the framework of transport and concentration problems proposed in Chapter 3. It is in fact the minimization of a functional $\mathfrak{F}_\nu$ with $\nu = \delta_0$ (a very much concentrated choice for ν). Exactly as the models in Chapter 1 were proposed to study urban planning problems, with μ standing for the population density in a region, (P_+) may be used in studying the shape of a leaf or a flower, represented by μ. In fact the minimization of a sum of an X_α term and a convex functional on μ could be an easy model taking into account that leaves want to be as spread as possible to catch sunlight but have to be irrigated starting from a single source. This model has been informally proposed by G. Buttazzo. Another model for the formation of tree leaves taking into account irrigation costs may be found in [76], but it has a different nature, as it takes care of the evolution and growth of the leaf.

In the framework of Chapter 1 the key condition coming from optimality was $\psi + f'(u) = \mathit{const}$ and the landscape function z dealt with in this paper plays somehow the role of the Kantorovich potential ψ. Also Corollary

7.3.9 can be seen as a similarity between the landscape function and the Kantorovich potential. Moreover, the Hölder continuity result at the end of this paper perfectly agrees with the fact that Kantorovich potentials (which correspond to $\alpha = 1$) are Lipschitz continuous.

7.4 Properties of the landscape function

7.4.1 Semicontinuity

Lemma 7.4.1. *Given any $\eta \in \mathcal{P}(\Gamma)$, the function $Z_\eta : \Gamma \to \mathbb{R}$ is lower semi-continuous with respect to pointwise convergence.*

Proof. This result is almost implicitly proven both in [57] and in [13], but never explicitly stated. It is anyway proven that $x \mapsto [x]_\eta$ is upper semi-continuous, and hence $x \mapsto [x]_\eta^{\alpha-1}$ is l.s.c. Then, to prove $\liminf_n Z_\eta(\gamma_n) \geq Z_\eta(\gamma)$, fix a time $t_1 < \sigma(\gamma)$ and use $\liminf \sigma(\gamma_n) \geq \sigma(\gamma)$. Eventually we have $\sigma(\gamma_n) > t_1$ and, by Fatou's Lemma, we get

$$\liminf_n Z_\eta(\gamma_n) \geq \liminf_n \int_0^{t_1} [\gamma_n]_\eta^{\alpha-1} dt \geq \int_0^{t_1} [\gamma]_\eta^{\alpha-1} dt.$$

Passing to the limit as $t_1 \to \sigma(\gamma)$ gives the thesis. □

Theorem 7.4.2. *The landscape function z is lower semi-continuous.*

Proof. Consider a sequence $x_n \to x$ and, correspondingly, some $\eta-$good curves γ_n such that $\pi_\infty(\gamma_n) = x_n$ and $z(x_n) = Z_\eta(\gamma_n)$. We may assume $\sup_n z(x_n) < +\infty$. Since it holds $\sigma(\gamma_n) \leq Z_\eta(\gamma_n) = z(x_n)$, we also have $\sup_n \sigma(\gamma_n) < +\infty$ and we can extract a subsequence (not relabeled) such that $\gamma_n \to \gamma$ uniformly. It is not difficult to prove that $\pi_\infty(\gamma) = x$. Thus, it is sufficient to use Lemma 7.4.1 to get $Z_\eta(\gamma) \leq \liminf_n Z_\eta(\gamma_n) = \liminf_n z(x_n)$. This implies that γ is an $\eta-$good curve and that $z(x) = Z_\eta(\gamma)$, which yields the thesis. □

7.4.2 Maximal slope in the network direction

The next property that can be proven in general (i.e., under no extra assumption on α, Ω, $\mu \dots$) on the landscape function is the most important in view of its meaning in river basins applications. Our interest is a continuous counterpart of the landscape function of [9]. What we actually need is a result concerning the fact that, on the points of the irrigation network S_η, the direction of maximal slope of z is exactly the direction of the network. If an $\eta-$good curve γ_0 is fixed, from the definition of z for a.e. t_0 the derivative of z along the curve γ at the point $x_0 = \gamma_0(t_0)$ is exactly $|\gamma_0|_{t_0,\eta}^{\alpha-1}$. This is the reason why we prove the following result. Notice that, as we said, in this continuous case the function z cannot be expected to be very regular, and in fact the maximal slope result we are going to prove involves differentiability in a very pointwise way but very weak as well.

Theorem 7.4.3. *Let $x_0 = \gamma_0(t_0)$, where γ_0 is an $\eta-$good curve, t_0 a time such that $t_0 \leq \sigma(\gamma_0)$ and $\theta_0 := |\gamma_0|_{t_0,\eta} > 0$. Then, for any $x \notin \gamma_0([0,t_0])$, we have*

$$z(x) \geq z(x_0) - \theta_0^{\alpha-1}|x - x_0| - o(|x - x_0|).$$

This corresponds to saying that the slope at x_0 in the direction of the network is actually the maximal slope at x_0.

Proof. Let us fix $x \notin \gamma_0([0,t_0])$ such that $z(x) < z(x_0)$. We may assume that $x = \gamma_x(t_x)$ for an $\eta-$good curve γ_x (otherwise $z(x) = +\infty$) and that the two curves γ_0 and γ_x get apart at a certain time $t_1(x) < t_0$ (the case $t_1(x) \geq t_0$ implies in fact $z(x) \geq z(x_0)$). By Lemma 7.4.4 (see below) we know that $t_1(x) \to t_0$ as $|x - x_0| \to 0$. Let us set $\theta(t) = |\gamma_0|_{t,\eta}$: for $t \in [t_1(x), t_0]$ we may write $\theta(t) \leq \theta_0(1 + \varepsilon_x)$, where ε_x is infinitesimal as $|x - x_0| \to 0$ as a consequence of $t_1(x) \to t_0$. We use again Lemma 7.3.1 and its notations.

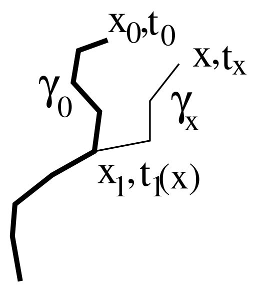

Figure 7.4: Curves and points in the proof

In particular $A = [\gamma_0]_{t_0}$ and $\theta_0 = \theta(t_0) = \eta(A)$. Let us also define, as in Theorem 7.3.2, $\mu'' = \mu - \mu_A + \eta(A)\delta_x$ and $\eta'' = \eta - I_A \cdot \eta + \eta(A)\delta_{\overline{\gamma_x}}$, where the curve $\overline{\gamma_x}$ is the curve γ_x stopped at time t_x, and it is easy to check that $(\pi_\infty)_\sharp\eta'' = \mu''$. Then, by the optimality of η' (we recall, according to the notations of Lemma 7.3.1 $\mu' = \mu - \mu_A + \eta(A)\delta_{x_0}$ and $\eta' = \eta - I_A \cdot \eta + \eta(A)\delta_{\overline{\gamma_0}}$, where $\overline{\gamma_0}$ is the curve γ_0 stopped at time t_0, see also Figure 7.4.2), we have

$$J(\eta') = X_\alpha(\mu') \leq X_\alpha(\mu'') + d_\alpha(\mu'', \mu') \leq J(\eta'') + |x - x_0|\theta_0^\alpha. \tag{7.4.1}$$

We want to compare $J(\eta')$ and $J(\eta'')$ and to do this here we need a more refined estimate than what we could find by using Theorem 7.2.1. As $\eta' - \eta'' = \theta_0(\delta_{\overline{\gamma_0}} - \delta_{\overline{\gamma_x}})$, we have in particular $[y]_{\eta''} = [y]_{\eta'} + \theta_0(I_{y\in\overline{\gamma_x}} - I_{y\in\overline{\gamma_0}})$. By using Formula (6.2.3) we get the following:

$$J(\eta'')-J(\eta') = \int_{\overline{\gamma_x}\setminus\overline{\gamma_0}} \big(([y]_{\eta'} + \theta_0)^\alpha - [y]_{\eta'}^\alpha\big) d\mathcal{H}^1 - \int_{\overline{\gamma_0}\setminus\overline{\gamma_x}} \big([y]_{\eta'}^\alpha - ([y]_{\eta'} - \theta_0)^\alpha\big) d\mathcal{H}^1.$$

It is not difficult to check that, for $y \in \overline{\gamma_x} \cup \overline{\gamma_0}$, it holds $[y]_{\eta'} = [y]_\eta$, as we have replaced the part of η concentrated on A by an equal amount of mass on $\overline{\gamma_0}$. Hence we may estimate (rewriting the integrals w.r.t. $\mathcal{H}^1$ as integrals in dt)

$$J(\eta'') - J(\eta') \leq \alpha \int_{t_1(x)}^{t_x} |\gamma_x|_{t,\eta}^{\alpha-1} \theta_0 \, dt - \int_{t_1(x)}^{t_0} (\theta(t)^\alpha - (\theta(t) - \theta_0)^\alpha) \, dt.$$

Since the function $s \mapsto s^\alpha - (s - \theta_0)^\alpha$ is decreasing and $\theta(t) \leq (1 + \varepsilon_x)\theta_0$, we get $\theta(t)^\alpha - (\theta(t) - \theta_0)^\alpha \geq \theta_0^\alpha \left((1 + \varepsilon_x)^\alpha - \varepsilon_x^\alpha\right)$. Hence we have

$$J(\eta'') - J(\eta') \leq \alpha(z(x) - z(x_1)) - |t_0 - t_1(x)|\theta_0^\alpha \left((1 + \varepsilon_x)^\alpha - \varepsilon_x^\alpha\right),$$

where $x_1 = \gamma_0(t_1(x)) = \gamma_x(t_1(x))$. Write $(1 + \varepsilon_x)^\alpha - \varepsilon_x^\alpha = (1 + \varepsilon'_x)^{-1}$ and $\varepsilon'_x > 0$ is infinitesimal as $x \to x_0$. From $\theta_0^{\alpha-1} \geq (\theta(t))^{\alpha-1}$ we get $|t_0 - t_1(x)|\theta_0^\alpha \geq \theta_0(z(x_0) - z(x_1))$. Now notice that, for $|x - x_0|$ sufficiently small, the inequality $\alpha < (1 + \varepsilon'_x)^{-1}$ is satisfied, and hence

$$J(\eta'') \leq J(\eta') + (1 + \varepsilon'_x)^{-1}\theta_0(z(x) - z(x_0)).$$

If we finally insert it into (7.4.1) we finally get

$$z(x) - z(x_0) \geq -\theta_0^{\alpha-1}|x - x_0|(1 + \varepsilon'_x), \quad \square.$$

Lemma 7.4.4. *According to the notations of Theorem 7.4.3, when $x \to x_0$ and $z(x) \leq z(x_0)$, the parting time $t_1(x)$ tends to t_0.*

Proof. Suppose, by contradiction, that there exists a sequence $x_k \to x_0$ such that $\lim_k t_1(x_k) = \bar{t} < t_0$ and $z(x_k) \leq z(x_0)$. Since γ_0 is injective (Corollary 7.3.4), we may infer the existence of a positive quantity δ such that $|\gamma_0(t_1(x_k)) - x_0| \geq \delta$ (otherwise there would be a time $t \leq \bar{t} < t_0$ with $\gamma_0(t) = x_0$). For any k consider an η-good curve γ_k such that $x_k = \gamma_k(t_k)$. First notice that, at least for k large enough, thanks to $|\gamma_k(t_1(x_k)) - x_k| = |\gamma_0(t_1(x_k)) - x_k| \to |\gamma_0(\bar{t}) - x_0| \geq \delta$, we have $t_k > \bar{t} + \delta/2$. Then let us consider the points $\gamma_k(\bar{t} + \delta/2)$: this collection of points must in fact be finite, otherwise we would have $|\gamma_k|_{\bar{t}+\delta/2,\eta} \to 0$ and hence

$$z(x_k) \geq |\gamma_k|^{\alpha-1}_{\bar{t}+\delta/2,\eta} |t_k - (\bar{t} + \delta/2)| \to +\infty$$

because $|t_k - (\bar{t} + \delta/2)| \geq |x_k - \gamma_0(\bar{t})| - \delta/2 \geq \delta/2$. This is in contradiction with $z(x_k) \leq z(x_0)$ and then we may suppose, up to subsequences, that $\gamma_k(\bar{t} + \delta/2) = \overline{x}$ (for a point $\overline{x}$ which does not belong to the image of γ_0, otherwise we would contradict Property 3) and that γ_k uniformly converges to a curve γ. At the limit we should get a curve γ passing through $\gamma_0(\bar{t})$, $\overline{x}$ and x_0, i.e. we have created a loop because γ_0 does not pass through $\overline{x}$ (see Figure 7.4.2 as well).

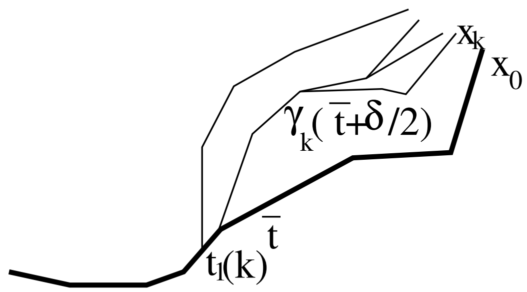

Figure 7.5: A sequence of curves creating a loop at the limit

From $z(x_k) \leq z(x_0)$ we can infer by semicontinuity (Lemma 7.4.1) that γ is an $\eta-$good curve and hence this loop is against Corollary 7.3.3. □

7.5 Hölder continuity under extra assumptions

Here we will be able to prove some extra regularity properties on z, but we have to add some assumptions. The most important ones are on α ($\alpha > 1 - 1/d$ is required) and on the irrigated measure μ (a lower bound on its density).

7.5.1 Campanato spaces by medians

We will here give a simple variant of a well-known result by Campanato (see [34]) about an integral characterization of Hölder continuous functions.

Definition 7.5.1. Given a measurable function u on a domain U we call median of u in U any number m which satisfies the following equivalent conditions:

- $|\{x \in U : u(x) > m\}| \le \frac{1}{2}|U|$ and $|\{x \in U : u(x) < m\}| \le \frac{1}{2}|U|$;
- there exists a measurable subset $A \subset \{x \in U : u(x) = m\}$ such that $|\{x \in U : u(x) > m\} \cup A| = \frac{1}{2}|U|$;
- the function $t \mapsto \int_U |u(x) - t|dx$ achieves its minimum at $t = m$.

The sets of medians of u in U is an interval of $\mathbb{R}$; the middle point of this interval is called central median of u in U

Definition 7.5.2. If A is a given positive number, a domain $\Omega \subset \mathbb{R}^d$ is said to be of type A if it holds $|\Omega_{x_0,r}| \ge Ar^d$ for any $r \in [0, \operatorname{diam}\Omega]$, where $\Omega_{x_0,r} = \Omega \cap B(x_0, r)$.

Lemma 7.5.3. *If Ω is a domain of type A and u is a function in $L^1(\Omega)$ such that*

$$\int_{\Omega_{x_0,r}} |u - \tilde{u}_{x_0,r}|dx \le Cr^{d+\beta},$$

for a finite constant C and any $r \in [0, \operatorname{diam}\Omega]$, where $\tilde{u}_{x_0,r}$ is the central median of u on $\Omega_{x_0,r}$, then u admits a representative which is Hölder continuous of exponent β.

Proof. This is nothing but the fact that Campanato spaces may be built by using medians instead of average values. See the proof of Theorem 1.2 at page 70 in [46] and adapt it. In fact it is easy to see that for each point x_0 the value $\tilde{u}_{x_0,r}$ converges as $r \to 0$ to a value $\tilde{u}(x_0)$ and that

$$|\tilde{u}(x) - \tilde{u}(y)| \le C|x - y|^\beta,$$

exactly as in the proof we mentioned. What we need to prove is that $\tilde{u}(x) = u(x)$ a.e.. This can be obtained in this way: let us denote the average value of u on $\Omega_{x_0,r}$ by $\overline{u}_{x_0,r}$. Then

$$|\overline{u}_{x_0,r} - \tilde{u}_{x_0,r}| \le |\Omega_{x_0,r}|^{-1} \int_{\Omega_{x_0,r}} |u(x) - \tilde{u}_{x_0,r}|dx \le |\Omega_{x_0,r}|^{-1} \int_{\Omega_{x_0,r}} |u(x) - \overline{u}_{x_0,r}|dx,$$

where the second inequality has been established as a consequence of the minimality property of the median. As at Lebesgue points the last expression tends to zero, this implies that the average $\overline{u}_{x_0,r}$ and the median $\tilde{u}_{x_0,r}$ share the same limit a.e. On the same points we also have $\overline{u}_{x_0,r} \to u(x_0)$, and this proves $\tilde{u}(x_0) = u(x_0)$ a.e. $\square$

7.5.2 Hölder continuity of the landscape function

Theorem 7.5.4. *Suppose that Ω is a domain of type A for $A > 0$, that $\alpha > 1 - 1/d$ and that $\mu \in \mathcal{P}(\Omega)$ is a probability measure such that the density of its absolutely continuous part is bounded from below by a positive constant. Then any landscape function z has a representative $\tilde{z}$ which is Hölder continuous of exponent $\beta = d(\alpha - (1 - 1/d))$.*

Proof. Let us fix a measure μ_1 and apply Theorem 7.3.12 to it and μ. By using the triangle inequality for d_α, we get

$$-d_\alpha(\mu, \mu_1) \leq X_\alpha(\mu_1) - X_\alpha(\mu) \leq \alpha \int_\Omega z \, d(\mu_1 - \mu), \tag{7.5.1}$$

provided μ_1 is a measure of the form allowed in Theorem 7.3.12, i.e. $\mu_1 << \mu$ with bounded density. From (7.5.1) we get

$$\alpha \int_\Omega z \, d(\mu - \mu_1) \leq d_\alpha(\mu, \mu_1). \tag{7.5.2}$$

Suppose that μ has an absolutely continuous part with density everywhere larger than $\lambda_0 > 0$ and choose

$$\mu_1 = \mu - \lambda_0 I_A \cdot \mathcal{L}^d + \lambda_0 I_B \cdot \mathcal{L}^d,$$

where A and B are two measurable subsets of $\Omega_{x_0,\varepsilon}$ with $|A| = |B|$, $A \cup B = \Omega_{x_0,\varepsilon}$ and $A \subset \{z \geq m\}$ and $B \subset \{z \leq m\}$ and m is the central median value for z in $\Omega_{x_0,\varepsilon}$. By construction μ_1 is a probability measure to which the estimate of Theorem 7.3.12 may be applied. With this choice of μ and μ_1 we get

$$\begin{aligned}\int_\Omega z \, d(\mu - \mu_1) &= \int_A z(x) \, \lambda_0 \, dx - \int_B z(x) \, \lambda_0 \, dx \\ &= \lambda_0 \left(\int_A (z(x) - m) dx - \int_B (z(x) - m) dx \right) = \lambda_0 \int_{\Omega_{x_0,\varepsilon}} |z(x) - m| dx.\end{aligned}$$

Putting into (7.5.2)

$$\int_{\Omega_{x_0,\varepsilon}} |z(x) - m| dx \leq (\alpha\lambda_0)^{-1} d_\alpha(\mu, \mu_1).$$

To estimate $d_\alpha(\mu, \mu_1)$ use (6.3.3) and get

$$\int_{\Omega_{x_0,\varepsilon}} |z(x) - m| dx \leq \frac{C_{\alpha,d}}{\lambda_0^{1-\alpha}} \varepsilon^{1+\alpha d}.$$

Since $1 + \alpha d = d + \beta$, Lemma 7.5.3 may be applied. $\square$

An important consequence of this fact is the following:

Corollary 7.5.5. *Under the same assumptions on Ω, α and μ of Theorem 7.5.4, the inequality*

$$X_\alpha(\mu_1) \le X_\alpha(\mu) + \int_\Omega \tilde{z} d(\mu_1 - \mu)$$

holds for any measure $\mu_1 \in \mathcal{P}(\Omega)$.

Proof. The inequality holds for μ_1 of the form $\mu_1 = (1+g)\cdot\mu$ with $g \in L^\infty$, but any measure $\mu_1 \in \mathcal{P}(\Omega)$ may be approximated by these kind of measures. Since $\tilde{z}$ is continuous, at both the sides of the inequalities we have quantities which are continuous with respect to weak convergence in the variable μ_1. This allows to conclude that the same inequality is valid for any μ_1. □

Even if we have proven that the landscape function z equals a.e. a function which is Hölder continuous, this is not enough. In fact, this result does not provide information on the behavior of z on negligible sets. Yet, the pointwise values of z on S_η are of particular interest (as in last Section), and S_η is one-dimensional and thus negligible. This is why the next step will be proving that z and $\tilde{z}$ actually agree everywhere.

Theorem 7.5.6. *Let m_ε denote the central median of z in the ball $B(x_0, \varepsilon)$. Under the same assumptions of Theorem 7.5.4 one has $m_\varepsilon \to z(x_0)$ as $\varepsilon \to 0$. Consequently, we have $\tilde{z}(x_0) = z(x_0)$.*

Proof. By the semicontinuity of z it is easy to get $\liminf_{\varepsilon\to 0} m_\varepsilon \ge z(x_0)$, hence only an estimate from above for m_ε is needed. Let us now consider a ball $B(x_0,\varepsilon)$ and a set $A_\varepsilon \subset B(x_0,\varepsilon) \cap \{z \ge m_\varepsilon\}$ such that $|A_\varepsilon| = |B(x_0,\varepsilon)|/2$. Then set $\Gamma_\varepsilon = \{\gamma \in \Gamma : \pi_\infty(\gamma) \in A_\varepsilon\}$, $\mu_\varepsilon = \mu + \mu(A_\varepsilon)\delta_{x_0} - I_{A_\varepsilon}\cdot\mu$, and $\eta_\varepsilon = \eta + \eta(\Gamma_\varepsilon)\delta_{\gamma_0} - I_{\Gamma_\varepsilon}\cdot\eta$, where γ_0 is an $\eta-$good curve stopping at x_0. Theorem 7.2.1 can be applied to η and η_ε and hence we have

$$\begin{aligned} J(\eta_\varepsilon) &\le J(\eta) + \alpha\left(\eta(\Gamma_\varepsilon)Z_\eta(\gamma_0) - \int_{\Gamma_\varepsilon} Z_\eta d\eta\right) \\ &= J(\eta) + \alpha\mu(A_\varepsilon)z(x_0) - \alpha\int_{A_\varepsilon} z(x)\mu(dx) \le J(\eta) + \alpha\mu(A_\varepsilon)(z(x_0) - m_\varepsilon). \end{aligned}$$

Hence we have

$$X_\alpha(\mu) \le X_\alpha(\mu_\varepsilon) + C\varepsilon\mu(A_\varepsilon)^\alpha \le X_\alpha(\mu) + \alpha\mu(A_\varepsilon)(z(x_0) - m_\varepsilon) + C\varepsilon\mu(A_\varepsilon)^\alpha.$$

This implies

$$m_\varepsilon - z(x_0) \le C\varepsilon\mu(A_\varepsilon)^{\alpha-1} \le C\varepsilon^{1+d(\alpha-1)}.$$

Since the exponent $1+d(\alpha-1)$ is larger than 0 we get $\limsup_{\varepsilon\to 0} m_\varepsilon \le z(x_0)$. To get the second part of the thesis, just use $\tilde{z}(x_0) = \lim_{\varepsilon\to 0} m_\varepsilon$. □

Remark 7.5.7. The landscape function z is in general never Lipschitz continuous (not even locally), as on the set S_η it has slopes given by $\theta^{\alpha-1}$. This means that, if we have arbitrarily small values of θ, we cannot have a Lipschitz constant for z. Yet estimates of the kind $\theta \ge c > 0$ would imply $\mathcal{H}^1(S_\eta) < +\infty$ and no measure whose support is not one-dimensional may be irrigated by a set of finite length (or locally finite length).

Remark 7.5.8. Notice that the Hölder exponent β is the same of the inequality $d_\alpha \le cW_1^\beta$ of Section 6.4.

Chapter 8

Blow-up for optimal one-dimensional sets

In this chapter we consider a minimization problem for an average distance functional over one-dimensional sets under length constraints and investigate some regularity properties of the minimizers. This problem, first proposed in [27], has also a transport interpretation. Even if the structure of the problem is quite different, it is interesting to compare it and its solutions to what happens in the branching transport problems we saw in Chapter 6. In fact the results of this chapter, which come from a joint published work with Paolo Tilli ([69]), will be the inspiration source for analogous results in the branching transport framework, as we will see in Chapter 9. This is the main reason to insert this chapter in this thesis.

8.1 Average distance problems and free Dirichlet regions

We present here, in a sketched and simplified way, the framework of the problems that are described in [27] and in [30].

An optimal transport problem with a Dirichlet region Σ is a Kantorovich problem where the cost function $c(x,y)$ takes into account that "transportation is free of charge along Σ". This means that a point x may be linked to a point y by arriving up to $x' \in \Sigma$, moving from $y' \in \Sigma$ up to y and paying only for the transportation between x and x' and between y' and y. If Σ is arcwise connected this means that we moved from x' to y' at no charge, and otherwise this means that free teleport between points of Σ is allowed. If the original cost of transportation was the Monge one, i.e. $c(x,y) = |x-y|$, then this induces a new cost $c_\Sigma(x,y) = |x-y| \wedge (d(x,\Sigma) + d(y,\Sigma))$.

In this kind of problems it turns easily out that if one of the two measures, say ν, has a part which is supported on Σ, then it does not matter where this part is concentrated. In other words, one can always replace ν by another measure $\tilde{\nu}$ such that $\nu - \tilde{\nu}$ is concentrated on Σ and the infimum value of $\int c_\Sigma \, d\pi$ for $\pi \in \Pi(\mu,\nu)$ does not change. In the particular case when ν is completely concentrated on Σ it means that it could be replaced by any other probability measure concentrated on the same set. As a consequence the minimum value of the Kantorovich problem for the cost c_Σ and fixed marginals $\mu \in \mathcal{P}(\Omega)$ and $\nu \in \mathcal{P}(\Sigma)$ is given by

$$D(\mu,\Sigma) = \int_\Omega d(x,\Sigma)\, \mu(dx) = \inf \{W_1(\mu,\nu) \,|\, spt(\nu) \subset \Sigma\}, \tag{8.1.1}$$

where W_1 is the $1-$Wasserstein distance as in Section 0.2. In the end this problem reduces to an average distance from the set Σ, where the average is computed according to the measure μ (and only this measure actually appears in the result).

As a further problem, it is interesting to let Σ itself vary. In this way we get the general version of the average distance problem, i.e.

$$\min\{D(\mu,\Sigma) \,|\, \Sigma \in \mathcal{E}\},$$

where $\mathcal{E}$ is a suitable class of subsets of Ω. As the problem would be trivial if Ω itself were an admissible set, the point is choosing a class Σ which contains only "small" sets in some sense. Typical choices are the set of all the finite Σ such that $\sharp\Sigma \leq k$ or the set of all the one-dimensional connected and compact Σ such that $\mathcal{H}^1(\Sigma) \leq l$. From (8.1.1) it is easy to see that these problems are equivalent to the minimization of $W_1(\mu,\nu)$ under constraints on the support of ν and hence turn out to be transport and concentration problem for a functional $\mathfrak{F}_\mu$ as in Section 3.1.

The first two cases, namely

$$\min\{D(\mu,\Sigma) \,|\, \sharp\Sigma \leq k\} = \min\{W_1(\mu,\nu) \,|\, \sharp\,\mathrm{spt}(\nu) \leq k\},$$

and

$$\min\{D(\mu,\Sigma) \,|\, \Sigma \text{ is compact and connected and } \mathcal{H}^1\Sigma \leq k\} \\ = \min\{W_1(\mu,\nu) \,|\, G(\nu) \leq l\},$$

(where $G : \mathcal{P}(\Omega) \to [0,+\infty]$ is the *length* concentration preferring functional introduced in Section 3.1) are known as *location problem* and *irrigation* *problem*, respectively. As we want to avoid confusion between this kind of irrigation problems, coming from an average distance functional under length constraints, and the branching transport problems of Chapter 6, we will try not to use the name of irrigation problem in this setting.

In this chapter we will be specifically concerned with the following variational problem:

$$\min D(\Sigma) = \int_\Omega d(x, \Sigma)\, \mu(dx) : \Sigma \subset \overline{\Omega} \text{ compact and connected, } \mathcal{H}^1(\Sigma) \leq l. \tag{8.1.2}$$

As a possible interpretation, we may regard Σ as a resource, whose amount is limited by the constraint $\mathcal{H}^1(\Sigma) \leq l$, to be distributed over a region Ω. Since the functional $F(\Sigma)$ is the average distance of a point $x \in \Omega$ to Σ, minimizing $F(\Sigma)$ means letting the resource be as widespread as possible throughout the region Ω. Of course, the measure μ reflects the fact that some subregions of Ω might have a higher or lower priority in being close to the resource Σ. Finally, the imposed connectedness of Σ prevents the infimum of F from being zero, and is a natural constraint in several applications, such as image reconstruction (trying to recover a line Σ from a pixel cloud μ in a picture) or urban planning (when Σ is a subway network in a city Ω with population density μ). The interpretation in image analysis corresponds to what is in general known as the *skeleton* of the image.

Existence results for the minimum problem (8.1.2) can be achieved by Gołab's theorem (see, for instance, [7]) and are the starting point of [27].

In the first paper on the subject [27], after proving the existence of solutions, some questions concerning the minimizers are posed and some partial answers are provided. Stronger results on the structure of a minimizer Σ are obtained in [29]. In the next section we will recall the results from [27, 29] which are relevant to the present paper, after introducing some notation which will be used throughout, and then we will establish the main technical tools that we will use.

We will use several different variational techniques to get necessary optimality conditions; most of them are simply based on first-order perturbations of the minimizer, i.e. the key ingredient is stationarity instead of minimality. We cite in connection to this feature of our work the classical paper [2] on regularity of stationary 1-dimensional structures. However, some of the cases we study here have been solved by more variational techniques, such as $\Gamma-$convergence, and are not easy to translate into stationarity conditions.

8.2 Preliminary and auxiliary results

Let Σ be a minimizer of (8.1.2). We denote by $T : \Omega \to \Sigma$ a fixed measurable selection of the projection multimap, i.e. T is a measurable map such that $|x - T(x)| = d(x, \Sigma)$ for every $x \in \Omega$. See [38] for the existence of such a selection. The measure μ and the map T induce a measure ψ on Σ defined by $\psi = T_\sharp \mu$. Of course, ψ is a probability measure concentrated on Σ. Estimates on the measure ψ will be crucial in the sequel.

The result from [29] which is most relevant to our purpose states that, topologically, Σ is equivalent to a finite graph without cycles, whose vertices have order at most three. More precisely (see [29] for more details):

- Σ is the union of finitely many injective Lipschitz curves with endpoints (called "branches" of Σ).
- Any two branches have at most one endpoint in common;
- Σ has no loops (i.e. $\mathbb{R}^2 \setminus \Sigma$ is connected);
- For every $x \in \Sigma$, exactly one of the following three possibilities occurs:
    1. x is in the relative interior of one branch and belongs to no other branch of Σ (in this case we say that x is an *ordinary point* of Σ);
    2. x is an endpoint of exactly one branch of Σ (we say that x is an *endpoint* of Σ);
    3. x is an endpoint of exactly three branches of Σ (we say that x is a *triple junction*);
- Every endpoint of Σ is an atom for the measure ψ.
- There are at least two endpoints in Σ.

Here we focus on the existence and characterization of blow-up limits of Σ. More precisely, we say that Σ has a blow-up limit K at $x_0 \in \Sigma$ if the localized and rescaled sets $(\Sigma \cap \overline{B(x_0, r)} - x_0)/r$ converge, in the Hausdorff distance as $r \to 0$, to some set $K \subset \overline{B_1}$ (here and throughout, $B(x_0, r)$ denotes the ball of radius r centered at x_0, whereas B_r denotes the ball centered at the origin).

We stress the fact that the existence of the blow-up limits in the Hausdorff distance at a point x_0 is linked to differentiability of Σ at x_0. In fact, parameterizing every branch of Σ by arc length curves, it is not difficult to check that our results on blow-ups will imply the existence of the derivative as a unit vector in the classical sense, or at least the existence of the derivatives from each side in the unlucky case of the limit being a corner.

Let us now add some assumptions and set the last useful notations. From now on Ω will be a convex domain of $\mathbb{R}^2$. Since Ω is supposed to be convex we know that $\Sigma \subset \Omega$ is a solution also to Problem 8.2.1:

$$\min D(\Sigma) = \int_\Omega d(x,\Sigma)\,\mu(dx) : \Sigma \subset \mathbb{R}^2 \text{ compact and connected, } \mathcal{H}^1(\Sigma) \leq l. \tag{8.2.1}$$

This is a consequence of what proven in [30], i.e. that Σ is always contained in the convex hull of the support of μ, and so there is no matter if we enlarge the domain. In the sequel, to get rid of possible boundary difficulties, we will silently use the fact that Σ minimizes also in Problem 8.2.1.

The measure μ will be supposed to be of the form $\mu = f \cdot \mathcal{L}^d$ with $f \in L^\infty$.

We will call C any positive, finite constant depending only on Ω, μ and Σ that may be enlarged at will. Every time a new, larger constant C is needed, the value of the former will be considered to be enlarged as necessary, without changing the notation C.

Finally, given two sets $A, B \subset \mathbb{R}^2$, we denote by $\mathrm{d_H}(A,B)$ their Hausdorff distance, i.e. the infimum of those positive numbers h such that $d(x,B) < h$ for any $x \in A$ and, conversely, $d(y,A) < h$ for any $y \in B$.

Remark 8.2.1. Let S be a Lipschitz curve with endpoints x, y and let $\overline{xy}$ be the segment from x to y. Then the Hausdorff distance $\mathrm{d_H}(S, \overline{xy})$ coincides with the smallest $h \geq 0$ such that S is contained in a h_0-neighborhood of $\overline{xy}$. Indeed, letting h_0 denote such smallest h, it suffices to observe that $\overline{xy}$ is contained in a h-neighborhood of S: in fact, for every $z \in \overline{xy}$, there is at least a point of S on the line through z perpendicular to $\overline{xy}$.

Lemma 8.2.2. *In the same case as before we have*

$$\mathcal{H}^1(S) \geq \sqrt{4\,\mathrm{d_H}(S, \overline{xy})^2 + |x-y|^2}. \tag{8.2.2}$$

Proof. Take a point $p \in S$ a distance exactly $\mathrm{d_H}(S, \overline{xy})$ to the segment $\overline{xy}$ (such a point exists due to the previous remark): then clearly

$$\mathcal{H}^1(S) \geq |x-p| + |p-y|,$$

and minimizing the last expression over all possible p gives the desired estimate (the minimum is achieved at a point p belonging to the axis of $\overline{xy}$). $\square$

We use here the fact that endpoints are atoms for ψ to establish a basic estimate which will be useful in the sequel.

Lemma 8.2.3. *There exists a constant $C > 0$ with the following properties. Let U be any open subset of Σ and V a compact set in $\mathbb{R}^2$ such that $(\Sigma \setminus U) \cup V$ is connected. If $\Sigma \setminus U$ contains at least one endpoint of ψ, then*

$$\mathcal{H}^1(U) \leq \mathcal{H}^1(V) + C\psi(U \setminus V) \max_{z \in U} d(z, V) \tag{8.2.3}$$

and

$$\mathcal{H}^1(U) \leq \mathcal{H}^1(V) + C\psi(U \setminus V)\, \mathrm{d_H}(U, V). \tag{8.2.4}$$

The proof uses some ideas from [29]. For the sake of completeness, we provide all the details.

Proof. We can assume that $\mathcal{H}^1(V) \leq \mathcal{H}^1(U)$.

Let $a \in \Sigma \setminus U$ be an endpoint of Σ (hence an atom for ψ), and set $A = T^{-1}(a)$. Since A is the intersection of a convex set and Ω, and $\mu(A) = \psi(\{a\}) > 0$, then A has nonempty interior. Let $B(y, \rho)$ be a small closed ball of radius $\rho > 0$, centered at some point $y \in A$ and contained in A, such that $a \notin \overline{B(y, \rho)}$ and $\mu(B(y, \rho)) > 0$. Letting $l = \mathcal{H}^1(U) - \mathcal{H}^1(V)$, we construct a new competitor Σ' as follows:

$$\Sigma' = (\Sigma \setminus U) \cup V \cup s_l,$$

where s_l is the closed segment of length l, that lies on the half-line from a to y and has a as one endpoint. Clearly, $\mathcal{H}^1(\Sigma') = \mathcal{H}^1(\Sigma)$. Moreover, Σ' is compact and connected (the segment s_l touches $\Sigma \setminus U$ at a, and $(\Sigma \setminus U) \cup V$ is connected by assumption), hence the minimality of Σ implies that

$$\int_\Omega d(x, \Sigma)\, \mu(dx) \leq \int_\Omega d(x, \Sigma')\, \mu(dx). \tag{8.2.5}$$

Note that, by construction, $d(x, \Sigma') \leq d(x, \Sigma)$ for all $x \in \Omega$ such that $T(x) \notin U \setminus V$, hence in particular, from (8.2.5) we find

$$\int_{B(y,\rho)} \big(d(x, \Sigma) - d(x, \Sigma'))\big)\, \mu(dx) \leq \int_{T^{-1}(U \setminus V)} \big(d(x, \Sigma') - d(x, \Sigma)\big)\, \mu(dx). \tag{8.2.6}$$

For every $x \in B(y, \rho)$, we have $d(x, \Sigma) = |x - a|$, and $d(x, \Sigma') \leq d(x, s_l)$. Hence we find

$$\int_{B(y,\rho)} \big(d(x, \Sigma) - d(x, \Sigma'))\big)\, \mu(dx) \geq \int_{B(y,\rho)} (|x - a| - d(x, s_l))\, \mu(dx) =: I(l). \tag{8.2.7}$$

Considering for a while l as a parameter, we want to estimate from below the last integral $I(l)$ as a function of l (i.e. the length of the segment s_l), on the interval $[0, \mathcal{H}^1(\Sigma)]$ (having defined $l = \mathcal{H}^1(U) - \mathcal{H}^1(V)$, we are not interested in $I(l)$ when $l > \mathcal{H}^1(\Sigma)$): it is clear that $I(l)$ is non decreasing in l, and one can easily check that when l is small enough (for instance, such that $2l < |y-a| - \rho$), there holds $I(l) \geq \varepsilon l$ for some $\varepsilon > 0$ (which depends only on ρ, μ and on the distance $|a-y|$). Since $I(l)$ is nondecreasing, reducing if necessary the value of ε we obtain that an estimate of the kind $I(l) \geq \varepsilon l$ holds for all $l \in [0, \mathcal{H}^1(\Sigma)]$. Therefore, plugging this estimate in (8.2.7) and using (8.2.6), we obtain

$$\varepsilon \left(\mathcal{H}^1(U) - \mathcal{H}^1(V)\right) = \varepsilon l \leq I(l) \leq \int_{T^{-1}(U \setminus V)} \left(d(x, \Sigma') - d(x, \Sigma)\right) \mu(dx). \tag{8.2.8}$$

Note that ε can also be made independent of the particular endpoint a that we have chosen in $\Sigma \setminus U$, since Σ has only finitely many endpoints: therefore, we may work with some ε which depends only on Σ, and not on U. Then, observing that

$$\sup_{x \in T^{-1}(U \setminus V)} \left(d(x, \Sigma') - d(x, \Sigma)\right) \leq \sup_{x \in T^{-1}(U)} \left(d(x, V) - d(x, U)\right) \leq \sup_{z \in U} d(z, V),$$

from (8.2.8) one obtains (8.2.3). Finally, (8.2.4) is an easy consequence of (8.2.3). □

We face an important particular case when the set V is a segment.

Lemma 8.2.4. *There exists C with the following property. Let $S \subset \Sigma$ be a closed injective arc, with endpoints x, y, such that $S \setminus \{x, y\}$ contains no triple junctions of Σ and $C\psi(S \setminus \{x, y\}) < 1/2$. Then*

$$\mathcal{H}^1(S) \leq |x-y| + C\psi(S \setminus \{x, y\})\, \mathrm{d_H}(S, \overline{xy}), \tag{8.2.9}$$

$$\mathrm{d_H}(S, \overline{xy}) \leq C\psi(S \setminus \{x, y\})|x-y|, \tag{8.2.10}$$

$$\mathcal{H}^1(S) \leq |x-y|\left(1 + C\psi(S \setminus \{x, y\})^2\right), \tag{8.2.11}$$

$$\mathcal{H}^1(S) \leq 2|x-y|. \tag{8.2.12}$$

Proof. We apply the previous lemma with $U = S \setminus \{x, y\}$ and V the segment from x to y (note that $(\Sigma \setminus S) \cup V$ is connected since we have replaced a simple curve by another curve with the same endpoints). Then (8.2.9) follows from (8.2.4). Moreover, on squaring (8.2.2) one obtains

$$4\, \mathrm{d_H}(S, \overline{xy})^2 \leq \left(\mathcal{H}^1(S) + |x-y|\right)\left(\mathcal{H}^1(S) - |x-y|\right) \leq 2\mathcal{H}^1(S)\left(\mathcal{H}^1(S) - |x-y|\right). \tag{8.2.13}$$

If we temporarily set $\Delta = \mathcal{H}^1(S) - |x-y|$ and we square (8.2.9), we get

$$\Delta^2 \le C\psi(S \setminus \{x,y\})^2 \, \mathrm{d_H}(S, \overline{xy})^2 \le C\psi(S \setminus \{x,y\})^2 \mathcal{H}^1(S)\Delta,$$

where we used also (8.2.13). Then we get

$$\Delta \le C\psi(S \setminus \{x,y\})^2 \mathcal{H}^1(S). \tag{8.2.14}$$

To estimate $\mathcal{H}^1(S)$ we use the estimate on Δ:

$$\mathcal{H}^1(S) - |x-y| = \Delta \le C\psi(S \setminus \{x,y\})^2 \mathcal{H}^1(S),$$

which, under the assumption $C\psi(S \setminus \{x,y\}) < 1/2$, provides

$$\frac{1}{2}\mathcal{H}^1(S) \le \left(1 - C\psi(S \setminus \{x,y\})^2\right) \mathcal{H}^1(S) \le |x-y|, \tag{8.2.15}$$

which is (8.2.12). Then, from (8.2.14) and (8.2.15), we get also the estimate

$$\Delta \le C\psi(S \setminus \{x,y\})|x-y|$$

and, by recalling (8.2.13) and (8.2.15), (8.2.10) is valid as well. Finally, (8.2.11) follows from (8.2.9) and (8.2.10). □

8.2.1 The function θ and its variation

Take a point $x_0 \in \Sigma$ and consider a branch of Σ starting at x_0: we may regard it as an injective Lipschitz curve $\gamma : [0,\delta] \to \Sigma$, parameterized by arclength, such that $\gamma(0) = x_0$ and $\gamma(\delta)$ is either an endpoint or a triple point of Σ. Of course, we suppose that $\delta > 0$ and that γ contains neither endpoints nor triple junctions in its relative interior.

We will prove that, if $r > 0$ is small enough, then γ touches $\partial B(x_0, r)$ at exactly one point: in this way, we can define for small $r > 0$ the function $\theta(r)$, i.e. (choosing polar coordinates centered at x_0) the angular coordinate θ of the (unique) point where γ touches $\partial B(x_0, r)$. We will also prove some regularity properties of the function $\theta(r)$.

Choose a certain radius $r_0 > 0$, such that the ball $B(x_0, r_0)$ contains no endpoint and no triple junction of Σ, with the only possible exception of x_0. In particular, γ meets $\partial B(x_0, r)$ at least once, for every $r \le r_0$.

We have the following

Theorem 8.2.5. *Consider $x_0 \in \Sigma$ and $r_0 > 0$ such that $B(x_0, r_0)$ contains no endpoint and triple junction other than, possibly, x_0 itself. For any $r < r_0$, set*

$$t_1 = \min\{t \ge 0 \,|\, \gamma(t) \in \partial B(x_0, r)\}, \quad t_2 = \max\{t \le \delta \,|\, \gamma(t) \in \partial B(x_0, r)\}.$$

If $C\psi(\gamma((0,t_2]) < 1$, then $t_1 = t_2$, i.e. γ touches $\partial B(x_0, r)$ exactly once.

Proof. Let us set for brevity $\gamma_{0,1} = \gamma([0,t_1])$, $\gamma_{0,2} = \gamma([0,t_2])$ and $\gamma_{1,2} = \gamma([t_1,t_2])$. We apply Lemma 8.2.3, with $U = \gamma_{0,2} \setminus \{\gamma(0), \gamma(t_2)\}$ and V equal to a suitable rotation of $\gamma_{0,1}$ around x_0, of an angle $\Delta\theta$, in such a way that the point $\gamma(t_1)$, after the rotation, overlaps with $\gamma(t_2)$. Observing that $\mathcal{H}^1(U) - \mathcal{H}^1(V) = \mathcal{H}^1(\gamma_{1,2})$, (8.2.3) implies that

$$\mathcal{H}^1(\gamma_{1,2}) \leq C\psi\left(\gamma_{0,2} \setminus \{x_0\}\right) \max_{x\in\gamma_{0,2}} d(x,V). \tag{8.2.16}$$

To estimate the max in the right hand side, let us split $\gamma_{0,2} = \gamma_{0,1} \cup \gamma_{1,2}$. Since V is a rotation of $\gamma_{0,1}$ which sends the boundary point $\gamma(t_1)$ to $\gamma(t_2)$, there holds

$$\max_{x\in\gamma_{0,1}} d(x,V) \leq |\gamma(t_1) - \gamma(t_2)| \leq \mathcal{H}^1(\gamma_{1,2}).$$

Moreover, since $\gamma(t_2) \in V$ we find

$$d(x,V) \leq |x - \gamma(t_2)| \leq \mathrm{diam}(\gamma_{1,2}) \leq \mathcal{H}^1(\gamma_{1,2})$$

since $\gamma_{1,2}$ is connected. Plugging these estimates into (8.2.16), we find

$$\mathcal{H}^1(\gamma_{1,2}) \leq C\mathcal{H}^1(\gamma_{1,2})\psi(\gamma((0,t_2])). \tag{8.2.17}$$

Under the assumption $C\psi(\gamma((0,t_2])) < 1$ it is clear that we get $\mathcal{H}^1(\gamma_{1,2}) = 0$ and also $t_1 = t_2$, since γ is injective. □

If we want the last result to be useful, it is necessary to establish the following.

Lemma 8.2.6. *For any $x_0 \in \Sigma$ there exists $r_0 = r_0(x_0) > 0$ sufficiently small such that, for any $r < r_0$, the ball $B(x_0,r)$ contains no triple junction nor endpoint other than, possibly, x_0 itself, and $C\psi(\gamma((0,t_2])) < 1$.*

Proof. It is immediate to satisfy the first constraint (no triple junction nor endpoint in the ball) since these points are finite. To satisfy the second requirement it is sufficient to show that the diameter of $\gamma((0,t_2])$ tends to 0 when $r \to 0$. In fact, proven this, we would have $\psi(\gamma((0,t_2])) \leq \psi(B(x_0,\varepsilon(r) \setminus \{x_0\})$ with $\varepsilon(r) \to 0$. Since the measure of the ball without the center tends to vanish with the radius, the thesis would be achieved.

To prove that $\mathrm{diam}(\gamma((0,t_2]))$ tend to 0 suppose, on the contrary, that there exists $\varepsilon_0 > 0$ and a sequence of radii $r_j \to 0$ with $\mathrm{diam}(\gamma((0,t_2^{r_j}]) \geq \varepsilon_0$ and $|x_0 - \gamma(t_2^{r_j})| = r_j$. In the limit we would get a loop in this branch of Σ, and this is a contradiction. □

To strengthen the result, we can make it quite uniform (and in the sequel we will use such an uniformization).

Theorem 8.2.7. *For any $\Sigma_1 \subset \Sigma$ compactly contained in the complement of the atoms of mass at least $(2C)^{-1}$ and of triple junctions and endpoints (which is the complement of a finite set) there exists $r_0 = r_0[\Sigma_1]$ such that, if $r < r_0$ and $x_0 \in \Sigma_1$, then $C\psi(B(x_0, r)) < 1/2$, no triple junction nor endpoint belongs to $B(x_0, r)$, and $C\psi(\gamma((0, t_2]) < 1$ (as in Theorem 8.2.5).*

Proof. We can consider separately the three requirements, and then choose the smallest radius.

It is easy to deal with the statement on $\psi(B(x_0, r))$: otherwise there would exist a sequence of centers x_0^n and of radii $r_n \to 0$ with the mass of the corresponding balls greater than $(2C)^{-1}$. By passing to a converging subsequence it would be straightforward to get the existence of a point $\overline{x_0}$ which would be an atom of mass at least $(2C)^{-1}$, obtained as a limit of the considered sequence of centers, which is a contradiction.

The requirement on triple junctions and endpoints is easily satisfied thanks to the assumption on Σ_1.

As far as the curves $\gamma((0, t_2])$ are concerned it is a little more difficult. Suppose on the contrary that there exists a sequence of arcs $\gamma_n([0, t_2^n])$ for which the distance $|\gamma_n(0) - \gamma_n(t_2^n)| = r_n$ is arbitrarily small and the measure $\psi(\gamma_n([0, t_2^n]))$ larger than C^{-1}. Up to subsequences we may assume that all this arcs are contained in one of the finitely many parts Σ_i consisting in the support of an injective simple curve γ and that they converge in the Hausdorff distance to a closed subset of Σ_i. The map γ provides an homeomorphism between Σ_i and the interval $[0, \delta]$. Because it is well known that Hausdorff convergence on compact sets depends only on topology and not on metric (see, for instance, [38]) we can deduce that we have convergence also of the images of the arcs in $[0, 1]$. The images are clearly segments and so the same holds for the limit. The condition that the distance between the extremal points x_n, y_n of the arcs goes to 0 says that, for a certain $x \in \Sigma_1$, we have $x_n \to x$, $y_n \to x$ and this fact is conserved by the homeomorphism. As a consequence also the extremal points of the segments in $[0, 1]$ collapse to the same point in the limit and so the limit must be a single point. Regarding this fact in Σ_i it is easy to deduce that we have a limit consisting in a single point which must be an atom of at least mass C^{-1}, which is a contradiction. □

As a consequence of what we have just proved, there exists well defined and continuous a function $\theta : (0, r_0] \to S^1$ such that $\theta(r)$ is the angle of the unique point of the curve γ which lies on $\partial B(x_0, r)$. The value r_0 has to be small enough and can be chosen quite uniformly according to Theorem 8.2.7 or depending on x_0 if x_0 is one of the dangerous points (triple junctions, endpoints, atoms of mass at least $(2C)^{-1}$). From now on, r_0 will always denote such a radius chosen in this way.

Theorem 8.2.8. *The function θ is locally Lipschitz on $(0, r_0]$. Moreover, for almost every $r \in (0, r_0)$ we have*

$$|\theta'(r)| \leq C \frac{\psi\left(\overline{B(x_0, r)} \setminus \{x_0\}\right)}{r}.$$

Proof. Consider two rays $0 < r < R \leq r_0$ and the variation $\Delta\theta$ of the angle θ between the two values of the ray.

Let $\Delta r = R - r$, $y = \Sigma \cap \partial B_r$ and $x = \Sigma \cap \partial B_R$. We apply Lemma 8.2.3 with $U = \gamma \cap B_R \setminus \{x_0\}$, and V given by two parts: a rotation of $\gamma \cap \overline{B_r}$ of an angle $\Delta\theta$ around x_0 (in such a way that the image y' of y under the rotation is collinear with x_0 and x), and the segment $\overline{y'x}$.

Setting $S := \gamma \cap \overline{B_R} \setminus B_r$, the Hausdorff distance from S to the segment $\overline{yx}$ can be bounded by $C|y - x|\psi(S \setminus \{x\})$ due to (8.2.10), whereas the distance from $\overline{yx}$ to $\overline{y'x}$ equals $|y - y'|$, hence

$$\max_{z \in S} d(z, V) \leq C|y - x|\psi(S \setminus \{x\}) + |y - y'| \leq C(r\Delta\theta + \Delta r)\psi(B_R \setminus \{x_0\}) + r\Delta\theta. \tag{8.2.18}$$

Moreover, for every point in $\gamma \cap \overline{B_r}$ there is a point in V at a distance less than $r\Delta\theta$ (just follow the point along the arc as it rotates), hence combining this with (8.2.18) we find

$$\max_{z \in U} d(z, V) \leq r\Delta\theta + C(r\Delta\theta + \Delta r)\psi(B_R \setminus \{x_0\}) \leq Cr\Delta\theta + C\Delta r\psi(B_R \setminus \{x_0\}). \tag{8.2.19}$$

Then from (8.2.4) we find

$$\mathcal{H}^1(U) - \mathcal{H}^1(V) \leq C\psi(U \setminus V)\big(r\Delta\theta + \Delta r\psi(B_R \setminus \{x_0\})\big).$$

By our construction, the left hand side equals $\mathcal{H}^1(S) - \Delta r \geq |y - x| - \Delta r$, hence we find

$$|y - x| - \Delta r \leq C\psi(B_R \setminus \{x_0\})\big(r\Delta\theta + \Delta r\psi(B_R \setminus \{x_0\})\big), \tag{8.2.20}$$

where we used $U \setminus V \subset B_R \setminus \{x_0\}$. Now for the left hand side a simple computation yields

$$
\begin{aligned}
|y-x| - \Delta r &= \sqrt{(\Delta r)^2 + 2r(r+\Delta r)(1-\cos\Delta\theta)} - \Delta r \geq \\
= \Delta r &\left(\sqrt{1 + \frac{r^2}{(\Delta r)^2}(1-\cos\Delta\theta)} - 1 \right) \geq C\Delta r \left(\frac{r^2(\Delta\theta)^2}{(\Delta r)^2} \wedge \frac{r\Delta\theta}{\Delta r} \right),
\end{aligned}
$$

where we used some elementary estimates such as $\sqrt{1+x^2} - 1 \geq C(x^2 \wedge x)$ and $(1-\cos t) \geq Ct^2$. Now, getting back to (8.2.20) and writing ψ_R for $\psi(B_R \setminus \{x_0\})$, we see that either

$$
\Delta r \frac{r^2(\Delta\theta)^2}{(\Delta r)^2} \leq C\psi_R r\Delta\theta + C\psi_R^2 \Delta r
$$

or

$$
\Delta r \frac{r\Delta\theta}{\Delta r} \leq C\psi_R r\Delta\theta + C\psi_R^2 \Delta r
$$

is satisfied.

The first case provides a quadratic estimate like $A^2 \leq CAB + CB^2$, which implies $A \leq (1 \vee 2C)B$, where $A = \Delta\theta/\Delta r$ and $B = \psi_R/r$. The second one, under the assumption that $2C\psi_R < 1$, gives $r\Delta\theta \leq 2C\psi_R^2\Delta r \leq \psi_R \Delta r$ and hence the same linear estimate.

So far we have obtained

$$
\frac{\Delta\theta}{\Delta r} \leq C\frac{\psi_R}{r},
$$

which gives local Lipschitz continuity of θ, as far as ψ_R/r remains bounded, i.e. as far as r stays bounded away from 0. By passing to the limit as $R \to r$ we get also the bound on the derivative required by the statement of the theorem. □

Remark 8.2.9. What we have just proved is useful when one wants to show uniqueness of the possible limits of subsequences of $(\Sigma \cap B_r)/r$: it is often enough to find a limit to $\theta(r)$ as $r \to 0$. To achieve it, it would be enough to have $\theta \in BV(0, r_0)$, since any function with bounded variation near 0 satisfies a Cauchy condition near the same point, and so the desired condition is

$$
\int_0^{r_0} \frac{\psi(B_r)}{r} dr < +\infty.
$$

Remark 8.2.10. Actually a Cauchy condition near 0 is sufficient, which is implied by a BV condition but is weaker. A similar idea will be used in Chapter 9.

8.2.2 Blow-up limits, up to subsequences

Lemma 8.2.11. *Choose a point $x_0 \in \Sigma$ and, for every $r > 0$, let $\Sigma_r = \Sigma \cap B(x_0, r)$. The family of rescaled sets $r^{-1}(\Sigma_r - x_0)$ is compact, in the metric space of all non-empty compact subsets of $\overline{B_1}$ endowed with the Hausdorff distance. If $r_j^{-1}(\Sigma_{r_j} - x_0)$ converge to some set $K \subseteq B_1$ for a suitable subsequence $r_j \to 0$, then:*

1. *If x_0 is an endpoint, then K is a radius of $\overline{B_1}$.*
2. *If x_0 is a simple point, then K is the union of two radii of $\overline{B_1}$, which form an angle of at least 120 degrees.*
3. *If x_0 is a triple junction, then K is the union of three radii of $\overline{B_1}$, forming angles of 120 degrees.*

Proof. We can assume that x_0 is the origin of the coordinates. According to the results in [29], there are p branches of Σ going out of x_0, with $p \in \{1, 2, 3\}$ according to the nature of x_0. We can regard these p branches as p injective curves $\gamma_i : [0, \delta] \to \Sigma$, with $\gamma_i(0) = x_0$. Moreover, these curves meet only at the starting point x_0 (otherwise Σ would have a loop), and Theorem 8.2.5 implies that, when r is small enough, each γ_i has a unique intersection with the circle ∂B_r. As a consequence, we can reparameterize each γ_i in such a way that for $r \in [0, r_0]$ we have $\gamma_i \cap \partial B_r = \{\gamma_i(r)\}$, and hence also $\Sigma_r = \bigcup_i \gamma_i([0, r])$.

Thanks to the choice of r_0 (small enough), we can suppose that Σ_r contains no triple junctions, other than (possibly) x_0. Then applying Lemma 8.2.4 with $S = \gamma_i([0, r])$ for some i, (8.2.10) and (8.2.11) yields

$$\mathrm{d_H}\left(\gamma_i([0, r]), \overline{x_0\gamma_i(r)}\right) \le Cr\psi\left(B_r \setminus \{x_0\}\right), \tag{8.2.21}$$

Moreover, letting $K_r = \bigcup_i \overline{x_0\gamma_i(r)}$ denote the union of the p radii of B_r from the center to the points $\gamma_i(r)$, using (8.2.21) for $i = 1, \ldots, p$ yields

$$\mathrm{d_H}(\Sigma_r, K_r) \le Cr\psi\left(B_r \setminus \{x_0\}\right),$$

Now observe that $\psi\left(B_r \setminus \{x_0\}\right) \to 0$ as $r \to 0$. Such an estimate is the key tool. Indeed, given a sequence of radii $r_j \to 0$, passing to subsequences (not relabeled) we may suppose that the sets $r_j^{-1}K_{r_j}$ converge, in the Hausdorff distance, to some set K, which clearly is the union of q radii of B_1, with $1 \le q \le p$. When $p = 1$ this proves the lemma.

Now suppose that $p = 2$. To complete the proof, it suffices to prove that $q = 2$ (i.e. that no two radii of K_r/r overlap in the limit).

To do this, we suppose that there exists some $\varepsilon > 0$ and a subsequence of radii (still denoted by r_j) such that the angle formed by, say, the radii $\overline{x_0\gamma_1(r_j)}$ and $\overline{x_0\gamma_2(r_j)}$ is less than $120° - \varepsilon$ for every j, and we seek a contradiction.

Let us apply Lemma 8.2.3 to the sets $U = (\gamma_1([0, r_j]) \cup \gamma_2([0, r_j])) \setminus \{x_0, \gamma_1(r_j), \gamma_2(r_j)\}$, and V equal to the Steiner connection of the three points x_0, $\gamma_1(r_j)$ and $\gamma_2(r_j)$. Since clearly $\mathcal{H}^1(U) \geq 2r_j$ and $\mathrm{d_H}(U, V) \leq 2r_j$, from (8.2.4) we find

$$2r_j \leq \mathcal{H}^1(V) + C\psi\left(B_{r_j} \setminus \{x_0\}\right) 2r_j.$$

Now, due to our assumption on the angle, one can check that there exists δ_0, depending only on $\varepsilon > 0$, such that $\mathcal{H}^1(V) \leq (2 - \delta_0)r_j$. Then we obtain

$$2r_j \leq (2 - \delta_0)r_j + C\psi(B_{r_j} \setminus \{x_0\})2r_j \quad \forall j,$$

and we get a contradiction for small enough r_j since $\psi(B_{r_j} \setminus \{x_0\}) \to 0$.

When $p = 2$, this shows that the two radii forming K_r tend to form an angle, in the limit, which is at least $120°$. Hence the limit set K is, in this case, the union of exactly two radii of B_1, whose angle is at least $120°$.

Finally, when $p = 3$ is suffices to repeat the same argument to every pair of radii in K_r. □

For the case of ordinary points, we give also a stronger result.

Lemma 8.2.12. *Suppose x_0 is a simple point with $\psi(\{x_0\}) = 0$ and $r_j \to 0$ a sequence of radii such that $r_j^{-1}(\Sigma_{r_j} - x_0)$ converge to a set $K \subset \overline{B_1}$. Then K is a diameter (i.e. the angle between the two radii given by Lemma 8.2.11 is in fact $180°$).*

Proof. As usual, we will suppose $x_0 = 0$. By Lemma 8.2.11 we know that K is the union of two radii, forming an angle $\alpha > 120°$. If we set $\Sigma \cap \partial B_{r_j} = \{x_j^1, x_j^2\}$, we may say

$$2 = \mathcal{H}^1(K) \leq \liminf_{j\to+\infty} \mathcal{H}^1\left(\frac{\Sigma_{r_j}}{r_j}\right) \leq \liminf_{j\to+\infty} \frac{|x_j^1 - x_j^2|}{r_j}\left(1 + C\psi(\Sigma_{r_j})^2\right) = 2\sin\left(\frac{\alpha}{2}\right).$$

Here the first inequality is a consequence of Gołab's Theorem, while the second comes from (8.2.11). This easily implies $\alpha = 180°$ and the thesis. □

8.2.3 $\Gamma-$Convergence

We want here to give a useful $\Gamma-$convergence result finding a $\Gamma-$limit to a sequence of functionals minimized by sets of the form $\Sigma_r := \Sigma \cap B(x_0, r)$. Here we state our theorem by considering only the case when x_0 is an endpoint, but the same result is true, with small modifications, also for any point of Σ which is an atom for ψ. A slightly more sophisticated $\Gamma-$convergence result concerning atomic ordinary point will be developed in section 8.3.

Let us consider an endpoint of Σ which we will call 0 and a small B_r around it. Let x_r be the only point of intersection of Σ and the boundary of the ball B_r. It is clear that Σ_r minimizes, among all compact connected sets S such that $x_r \in S$ and $\mathcal{H}^1(S) \leq \mathcal{H}^1(\Sigma_r)$, the quantity

$$\int_{A_r} d(x, S)\, \mu(dx), \tag{8.2.22}$$

where $A_r = T^{-1}(\Sigma_r)$ is the set of points projected to Σ_r. What we want to investigate now is whether we can find a limit of a proper rescaling of the functional appearing in (8.2.22), in order to get information on the limits of Σ_r/r. Let us consider the functionals defined on the set X of compact connected sets contained in the closed ball of radius 2 and of length less than 2 (which is a compact metric space if endowed with the Hausdorff topology, as a consequence of Gołab's Theorem), given by:

$$F_r(S) = \begin{cases} \int_{A_r/r} (d(x,S) - d(x,0))\, m_{r\sharp}\mu(dx) & \text{if } \frac{x_r}{r} \in S,\ \mathcal{H}^1(S) \leq \frac{\mathcal{H}^1(\Sigma_r)}{r}; \\ +\infty & \text{otherwise.} \end{cases} \tag{8.2.23}$$

Here we denote by m_r the division by r, i.e.: $m_r(x) = x/r$.

Each functional F_r is then minimized by Σ_r/r, as far as such sets have length smaller than 2. This happens for small r, since it holds

$$\lim_{r \to 0} \frac{\mathcal{H}^1(\Sigma_r)}{r} = 1,$$

as a consequence of (8.2.11).

Let us denote by Q the application $Q(x) = x/|x|$ that gives the direction of a non-zero vector and by ν the measure on S^1 given by $\nu = Q_\sharp(\mu I_{A_0})$, where A_0 is the set of point projected to the endpoint 0.

Lemma 8.2.13. *Suppose that on a certain subsequence (not relabeled) we have $x_r/r \to \overline{x}$: then it follows $F_r \stackrel{S}{\to} F$, where the $\Gamma-$convergence is in-*

tended with respect to the Hausdorff convergence on X and F is given by

$$F(S) = \begin{cases} \int_{S^1} -\delta^*(v|S)\nu(dv) & \text{if } x \in S \text{ and } \mathcal{H}^1(S) \leq 1; \\ +\infty & \text{otherwise.} \end{cases} \tag{8.2.24}$$

As usual, $\delta^*(v|S) = \sup_{y\in S} v\cdot y$. As a particular consequence Σ_r/r converges in the Hausdorff distance to a minimizer of F.

Proof. Let us start by the $\Gamma-\liminf$ inequality. We have to prove that, for any $S \in X$ and $S_r \to S$ we have $\liminf F_r(S_r) \geq F(S)$. We may suppose that $F_r(S_r)$ is finite for a subsequence, otherwise the lim inf is $+\infty$, and so we have a sequence of sets S_r approaching S in such a way that $x_r \in S_r$ and $\mathcal{H}^1(S_r) \leq \mathcal{H}^1(\Sigma_r)/r \to 1$. It is clear so far that S satisfies the constraints $\overline{x} \in S$ and $\mathcal{H}^1(S) \leq 1$, as a consequence of Hausdorff convergence's properties.

For every x we denote by $y_x(S_r)$ (one of) the nearest point(s) to x belonging to S_r and by $z_x(S_r)$ (one of) the point(s) realizing the max in $\sup_{z\in S_r} x\cdot z$. Notice that $z_x(S_r)$ depends actually only on $Q(x)$. For every point x we have

$$d(x, S_r) - d(x, 0) = d(x, y_x(S_r)) - d(x, 0) \geq -\frac{x}{|x|} \cdot y_x(S_r) \geq -\frac{x}{|x|} \cdot z_x(S_r).$$

So we may estimate

$$\begin{aligned} F_r(S_r) &\geq \int_{A_r/r} -\frac{x}{|x|} \cdot z_x(S_r) m_{r\sharp}\mu(dx) \\ &= \int_{A_r} -Q(x) \cdot z_{Q(x)}(S_r)\, \mu(dx) \\ &= \int_{A_0} -Q(x) \cdot z_{Q(x)}(S_r)\, \mu(dx) + \int_{A_r \setminus A_0} -Q(x) \cdot z_{Q(x)}(S_r)\, \mu(dx). \end{aligned}$$

The latter term in the last line tends to 0 with r because the integrand is bounded by 2 and the set on which we integrate converges to the empty set. The former indeed is equal to

$$\int_{S^1} -\delta^*(v|S_r)\nu(dv) \to F(S)$$

where the convergence relies on the fact that Hausdorff convergence implies pointwise convergence of the support functions δ^* (see, for instance, [38]). The $\Gamma-$liminf inequality is then proved.

Let us pass to the $\Gamma-$limsup inequality. For every fixed S such that $F(S) < +\infty$, we have to find a sequence $(S_r)_r \to S$ such that

$\limsup F_r(S_r) \le F(S)$. For each r it is sufficient to rotate S so that its intersection with the boundary of the unit ball becomes x_r instead of x and to perform an homothety around x_r in order to satisfy the length constraint. We have hence a sequence of sets S_r such that $F_r(S_r)$ is finite and given by the integral expression in (8.2.23), for which it holds $S_r \to S$ in the Hausdorff distance. This convergence is true thanks to $x_r \to \overline{x}$ and to the fact that convergence holds also for the ratios of the homotheties, which are prescribed by the length constraints. It remains to estimate $F_r(S_r)$. For each couple of point x, z we have

$$|x - z| - |x| = |x| \left(\sqrt{1 - \frac{2x \cdot z}{|x|^2} + \frac{|z|^2}{|x|^2}} - 1 \right) \le -Q(x) \cdot z + \frac{|z|^2}{2|x|}.$$

So we may write

$$\begin{aligned} F_r(S_r) &\le \int_{A_r/r} (|x - z_x(S_r)| - |x|) \; m_{r\sharp}\mu(dx) \\ &\le \int_{A_r/r} \left(-Q(x) \cdot z_{Q(x)}(S_r) + \frac{|z_{Q(x)}(S_r)|^2}{2|x|} \right) m_{r\sharp}\mu(dx) \\ &= \int_{A_0} -Q(x) \cdot z_{Q(x)}(S_r) \, \mu(dx) + \int_{A_r \setminus A_0} -Q(x) \cdot z_{Q(x)}(S_r) \, \mu(dx) \\ &\quad + r \int_{A_r} \frac{|z_{Q(x)}(S_r)|^2}{2|x|} \mu(dx). \end{aligned}$$

In the last sum, the first term yields $F(S)$ in the limit, while the second and the third tend to zero. □

8.2.4 Iterated estimates for small diameters

We show here that, if the diameter of the transported set to a certain point of Σ is sufficiently small, the measure $\psi(B_r)$, for balls centered around that point, can be estimated by r itself.

Lemma 8.2.14. *There exists a constant C such that, given $x_0 \in \Sigma$, and r_0 chosen as usual, if we set $k = diam(T^{-1}(B_{r_0}))$, for all $r \le r_0/2$ it holds*

$$\psi(B_r) \le Ck \left(r + \psi(B_{2r}) \right). \tag{8.2.25}$$

Proof. Consider a point of Σ, which we will call 0, together with two balls around it, of radii r and $2r \le r_0$ respectively. Let x_1, x_2 be the intersection points of Σ and the boundary of the largest ball. We can for each segment $\overline{0x_i}$ consider the Hausdorff distance $\mathrm{d_H}(\Sigma^i_{2r}, \overline{0x_i})$ between it and the corresponding branch of Σ and the distance $\mathrm{d_H}(\Sigma_{2r}, \overline{0x_1} \cup \overline{0x_2})$. We have a set K in which Σ_{2r} is contained, i.e. the set obtained by fattening the two segments by a quantity equal to the latter distance. We may estimate, using each branch of Σ_{2r} as S in (8.2.10),

$$\mathrm{d_H}(\Sigma_{2r}, \overline{0x_1} \cup \overline{0x_2}) \leq \max_{i=1,2} \mathrm{d_H}(\Sigma^i_{2r}, \overline{0x_i}) \leq Cr\psi(B_{2r}).$$

mentioned can be estimated by $Cr\psi(B_{2r})$, thanks to (8.2.10) and the choice of r and r_0, choosing a branch of Σ_{2r} to play the role of S and $x = 0$, $y = x_i$. Consider now the set $K' = K \cap B_r$ and its convex hull K''. Since we want to estimate the area of the set transported to B_r it is sufficient to estimate the area of the set of points which are closer to K'' than to the points x_i. Moreover, being k greater than the diameter of $T^{-1}(B_r)$, we can replace this set by its intersection with B_k. We include this set in the union of

- two stripes S_1, S_2 which are $2r$ wide and each S_i is orthogonal to $\overline{0x_i}$ and has the points 0 and x_i on its boundary,
- a sector E of amplitude $180° - \widehat{x_1 0 x_2}$ starting from 0, delimited by the boundaries of the stripes S_i passing through the origin,
- four small sectors $C_{i,j}$, each of them delimited by the boundary of the stripe S_i passing through x_i and the axis of the segment $\overline{x_i y_{i,j}}$, where the points $y_{i,j}$ are the corner points of the boundary of K'' near x_i.

The amplitude of these last sectors is the same of the angle $0\hat{x_i}y_{i,j}$, which can be estimated by $C\psi(B_{2r})$ thanks to the estimate on the Hausdorff distance. We know that also the amplitude α of the sector E can be estimated the same. In fact, we can consider in (8.2.11) $S = \Sigma_{2r}$, $x = x_1$ and $y = x_2$. We obtain

$$2r \leq \mathcal{H}^1(\Sigma_{2r}) \leq 2r \cos\left(\frac{\alpha}{2}\right)\left(1 + C\psi(\Sigma_{2r})^2\right),$$

and, dividing by $2r$ and using $(\cos\beta)^{-1} \geq 1 + c\beta^2$, which is true for $\beta \leq \pi$, we get

$$\alpha^2 \leq C\psi(\Sigma_{2r})^2.$$

Being μ a measure with an L^∞ density, it is enough to estimate the areas of T_i, E and $C_{i,j}$ intersected with B_k, and we obtain

$$\psi(B_r) \leq C\left(kr + k^2\psi(B_{2r})\right).$$

For simplicity we will estimate k^2 by Ck. □

The interest in the estimate (8.2.25) is that we can iterate it, especially when we have small diameters of the transported sets.

Theorem 8.2.15. *Suppose that there exists $r_1 < r_0/2$ such that*

$$k = \operatorname{diam}(T^{-1}(B(x_0, r_1))) < 1/(2C).$$

Then, for all $r < r_1$ we have the estimate

$$\psi(B_r) \leq \frac{Ckr}{1 - 2Ck} + \left(\frac{2r}{r_1}\right)^{\log_2 1/Ck}. \tag{8.2.26}$$

Proof. Fixed $r < r_1$ we can find an integer h such that $r_1/2 < r2^h \leq r_1$. Iterating (8.2.25) we obtain

$$\psi(B_r) \leq Ckr \sum_{i=0}^{h-1} (2Ck)^i + (Ck)^h \psi(B_{r_1}) \leq \frac{Ckr}{1 - 2Ck} + \left(\frac{2r}{r_1}\right)^{\log_2 1/Ck}.$$

□

Notice that, due to the semicontinuous behavior of the diameter of the transported set, saying that there exists a small r_1 such that $\operatorname{diam}(T^{-1}(B_{r_1})) < 1/(2C)$ is the same as saying that $\operatorname{diam}(T^{-1}(\{x_0\})) < 1/(2C)$.

Notice also the following useful consequence:

Corollary 8.2.16. *If $x_0 \in \Sigma$ and $\operatorname{diam}(T^{-1}(\{x_0\})) < 1/(2C)$, then $\psi(\{x_0\}) = 0$, i.e. all atoms of ψ have transported sets with large diameter.*

Proof. Just use (8.2.26) and $\psi(\{x_0\}) = \lim_r \psi(B_r)$. □

8.3 Blow-up limits

8.3.1 Triple junctions

We make here use of the previous section's tools to establish the expected result regarding singular points of Σ. For simplicity we will always center our analysis in a point x_0 supposed to be the origin.

Theorem 8.3.1. *Suppose $0 \in \Sigma$ is a triple junction: then there exists the limit as $r \to 0$ of Σ_r/r in the Hausdorff distance and it is composed by the union of three rays with $120°$ angles.*

Proof. Thanks to Lemma 8.2.11 (which states that the limits up to subsequences are shaped like the union of three rays angled 120°) we just need to show the uniqueness of those limits. By means of Remark 8.2.9, it is enough to achieve

$$\int_0^{r_0} \frac{\psi(B(0,r))}{r} dr < +\infty. \tag{8.3.1}$$

We will show that, for small r, it holds $\psi(B(0,r)) \le Cr^2$, thus achieving the goal.

Let us consider small values of r, such that the angles between the points of intersection of Σ with the boundary of $B(0,3r)$ are all smaller than 130° (we know that for small r this happens, otherwise we could produce a subsequence having a limit different from the admissible ones). Consider now a point x at a distance $|x| = cr$ from the origin: we want to show that, if c is great enough, it is not possible to have $x \in T^{-1}(B(0,r))$. Supposing on the contrary that x is transported to $B(0,r)$, we get that no point of Σ is contained in the ball centered at x and of radius $r(c-1)$. In particular no point of Σ may lay on the part of $\partial B(0,3r)$ contained in such a ball. Yet, the amplitude of such arc depends only on c and, as c increases to infinity, tends to $2\arccos(1/3) > 130°$. This would mean that, for big c, we would have an arc of 130° on $\partial B(0,3r)$ without any of the three points of intersection with Σ, which is a contradiction. So there exists a constant c_0 such that $T^{-1}(B(0,r)) \subset B(0,c_0 r)$ and this, since $\mu \in L^\infty$, completes the proof. □

Remark 8.3.2. Notice that (8.3.1) remains true also in the case where the measure μ, instead of being L^∞, is simply L^p with $p > 1$, because in this case we can use Holder inequality to get

$$\psi(B(0,r)) \le Cr^{2-2/p},$$

and this is sufficient for the convergence of the desired integral. This idea will be used in chapter 9 as well.

8.3.2 Endpoints

Here we will state an analogous theorem concerning existence of the limit near an endpoint of Σ, giving also a characterization of the direction of the ray we find as a limit. We use the same notation as in the $\Gamma-$convergence subsection, from which this theorem arises.

Theorem 8.3.3. *If* 0 *is an endpoint of* Σ *the limit of* Σ_r/r *in the Hausdorff distance as* $r \to 0$ *exists and is given by a single ray from the origin in the* *direction of $-\overline{v}$, where $\overline{v}$ is given by*

$$\overline{v} = \int_{S^1} v\,\nu(dv)$$

Proof. Thanks to Lemma 8.2.11 it is enough to determine the direction of the rays that can be possible limits of subsequences. To do this we use the $\Gamma-$convergence result provided in Lemma 8.2.13. In fact every set K limit of a subsequence of Σ_r/r intersecting ∂B_1 in a point $\overline{x}$ must be the set maximizing $\int_{S^1} \delta^*(v|S)\nu(dv)$ among all the sets S which are compact and connected, pass through $\overline{x}$ and satisfy $\mathcal{H}^1(S) \leq 1$. This maximizer is always a segment of unit length directed from $\overline{x}$ according to the vector $\overline{v}$. But we also know that $0 \in K$ and the only possible position for $\overline{x}$ so that the maximing set passes through the origin is $\overline{x} = -\lambda\overline{v}$. Then $\overline{x}$ is uniquely determined and the limit of Σ_r/r exists. □

8.3.3 Ordinary points

Our next step is establishing existence of the same limit in the four cases in which we will divide the general case of an ordinary point. In fact we will classify these points x_0 according to the shape of the transported set $\mathcal{T}(x_0) = \{x \in \Omega : d(x,\Sigma) = |x - x_0|\}$. This set coincides up to negligible sets with $T^{-1}(x_0)$. Moreover, $\mathcal{T}(x_0)$ is always a convex set, thus endowed with its own entire dimension: it may be $0, 1$ or 2. The four cases will be given by

1. $\mathcal{T}(x_0) = \{x_0\}$, i.e. dimension 0;
2. $\mathcal{T}(x_0)$ is a segment starting in x_0 (a subcase of dimension 1);
3. $\mathcal{T}(x_0)$ is a segment having x_0 in its relative interior (the other subcase of dimension 1);
4. $\mathcal{T}(x_0)$ is two-dimensional (i.e. with non empty interior).

Let us start from the easiest of the four cases:

Theorem 8.3.4. *Suppose* 0 *is an ordinary point of type* 3*: then there exists the limit of* Σ_r/r *as* $r \to 0$ *in the Hausdorff distance and it is the diameter composed by the two unit rays orthogonal to the segment* $\mathcal{T}(0)$*.*

Proof. No optimality of Σ is here required: just notice that Σ is contained in the complement of two suitable balls tangent in 0 to the segment orthogonal to $\mathcal{T}(0)$. □

Now we move to a case just a little more complicated:

Theorem 8.3.5. *Suppose* 0 *is an ordinary point of type* 2*: then there exists the limit of* Σ_r/r *as* $r \to 0$ *in the Hausdorff distance and it is the diameter composed by the two unit rays orthogonal to the segment* $\mathcal{T}(0)$*.*

Proof. Now we can only ensure that Σ stays outside a single ball tangent in 0 to the segment orthogonal to $\mathcal{T}(0)$: this is enough to say that, provided a limit of a subsequence is a diameter, it must be the diameter orthogonal to $\mathcal{T}(0)$. But every limit of subsequences here is a diameter, thanks to Lemma 8.2.12, since $\psi(\{0\}) = \mu(\mathcal{T}(0)) = 0$. So the limits are uniquely determined and this makes the limit exists. □

Our next case uses something more, because here $\mathcal{T}(0)$ gives no information on the possible limit:

Theorem 8.3.6. *Suppose* 0 *is an ordinary point of type* 1*: then there exists the limit of* Σ_r/r *as* $r \to 0$ *in the Hausdorff distance and it is a diameter.*

Proof. By using Lemma 8.2.12 on subsequences we know that any limit point in the Hausdorff distance has to be a diameter and, to identify it, it is enough to show that the function θ (with respect to any of the two branches of Σ going out from 0) has a limit. As usual, we will look for the inequality $\int_0^{r_0} \frac{\psi(B(0,r))}{r} dr < +\infty$, required by Remark 8.2.9. Here we can use the result valid when $\mathrm{diam}(T^{-1}(0))$ is small , given by Theorem 8.2.15, (we have actually a vanishing diameter) to establish an estimate like $\psi(B_r) \leq Cr$ for small r. This gives the convergence of the integrand and the proof is achieved. □

The last case requires something more, that we will state as another $\Gamma-$convergence lemma. This time we will use the fact that $\Sigma \cap B_r$ minimizes, among all sets S sharing with it the same two intersections with ∂B_r, the functional

$$\int_{A_r} d(x,S)\,\mu(dx) + P\left(\mathcal{H}^1(\Sigma_r) - \mathcal{H}^1(S)\right), \tag{8.3.2}$$

where the quantity $P(\varepsilon)$ is defined, for $\varepsilon < 0$, as the increase in the functional if we cut away a curve of length $-\varepsilon > 0$ starting from a given endpoint in Σ (it is in fact a penalization if S is too long), while for $\varepsilon > 0$ it is the diminution (a negative quantity) of the functional if we add a straight line segment of length ε starting from the same endpoint, in the direction of the tangent vector in it (which exists and coincides with the direction of $\overline{v}(x_0)$, thanks to Theorem 8.3.3). We give now a precise estimate of the term P, in term of the saved/lost length ε (if we save length we have $\varepsilon > 0$ and $P < 0$, and vice versa). Let v_0 be the unit vector in the direction of $\overline{v}(x_0)$, which is the tangent vector to Σ in the endpoint that we call x_0. If $\varepsilon > 0$ we can estimate

$$\begin{aligned} P(\varepsilon) &\leq \int_{A_0(x_0)-x_0} (|x - \varepsilon v_0| - |x|)\ \mu(dx) \\ &\leq \int_{A_0(x_0)-x_0} \left(-\varepsilon v_0 \cdot \frac{x}{|x|} + \frac{\varepsilon^2}{2|x|} \right) \mu(dx) = -\varepsilon v_0 \cdot \overline{v}(x_0) + o(\varepsilon); \end{aligned}$$

if, on the other hand $\varepsilon < 0$ we have

$$P(\varepsilon) \leq \int_{A_{r_\varepsilon}(x_0)} (|x - w_\varepsilon| - d(x, \Sigma))\ \mu(dx),$$

where w_ε is the point of Σ situated after an arc S_ε of length $|\varepsilon|$ starting form the extremal point 0 and $r_\varepsilon = \operatorname{diam}(S_\varepsilon)$ (for small ε it holds $r_\varepsilon = |w_\varepsilon - x_0|$). We may go on with the estimation with

$$\begin{aligned} P(\varepsilon) &\leq \int_{A_0(x_0)} (|x - w_\varepsilon| - |x - x_0|)\, \mu(dx) + \int_{A_{r_\varepsilon(x_0)} \setminus A_0(x_0)} (|x - w_\varepsilon| - d(x, \Sigma))\, \mu(dx) \\ &\leq -(w_\varepsilon - x_0) \cdot \overline{v}(x_0) + o(\varepsilon) + r_\varepsilon \mu(A_{r_\varepsilon}(x_0) \setminus A_0(x_0)). \end{aligned}$$

For small ε it is clear that $r_\varepsilon = |w_\varepsilon - x_0| \leq |\varepsilon|$ and moreover we have $(w_\varepsilon - x_0)/r_\varepsilon = -v_0 - \delta_\varepsilon$ with $\delta_\varepsilon \to 0$ as $\varepsilon \to 0$. We can then estimate again

$$P(\varepsilon) \leq r_\varepsilon v_0 \cdot \overline{v}(x_0) + r_\varepsilon \delta_\varepsilon \cdot \overline{v}(x_0) + o(\varepsilon) + o(r_\varepsilon) \leq -\varepsilon v_0 \cdot \overline{v}(x_0) + o(\varepsilon).$$

Notice that such estimates can be used in fact to get a precise quantitative version of part of the proof Lemma 8.2.4.

So the in the minimization problem given by (8.3.2) it is still true that Σ_r minimizes if we replace the real penalization by a function given by $P(\varepsilon) = -c\varepsilon + o(\varepsilon)$, where $c = v_0 \cdot \overline{v}(x_0) = |\overline{v}(x_0)|$. We may also require

$$P(\varepsilon) \geq -c\varepsilon, \tag{8.3.3}$$

because the minimization is preserved if we make bigger the value of the functional on sets different from the minimizer: in this case no matter if the value of P is made bigger outside 0 (it is the same reason for which we have only given estimate from above of the real penalization). We now rescale the functionals as before, obtaining that Σ_r/r minimizes

$$F_r(S) = \begin{cases} \int_{A_r/r} (d(x, S) - |x|)\ m_{r\sharp}\mu(dx) + \frac{1}{r} P(r(l_r - \mathcal{H}^1(S))) & \text{if } x_r^1, x_r^2 \in S, \\ +\infty & \text{otherwise,} \end{cases}$$

where $l_r = \mathcal{H}^1(\Sigma_r)/r \to 2$ (as usual, it is a consequence of (8.2.12)) and x_r^1 and x_r^2 are the points in which Σ_r/r intersects the boundary of the unit ball.

Lemma 8.3.7. *Let F denote the functional given by*

$$F(S) = \begin{cases} \int_{S^1} -\delta^*(v|S)\nu(dv) - c(2 - \mathcal{H}^1(S)) & \text{if } x^1, x^2 \in S \\ +\infty & \text{otherwise.} \end{cases}$$

Then $F_r \xrightarrow{\Gamma} F$ with respect to the Hausdorff convergence on the space X of compact connected sets contained in a fixed large closed ball, provided $x_r^i \to x^i$ for $i = 1, 2$.

Proof. The proof is close to that of Lemma (8.2.13): for the $\Gamma-$liminf inequality fix a S and an approaching sequence $(S_r)_r$ and use the same estimate to deal with the integral term of the functionals F_r and F. For the penalization term, thanks to (8.3.3), we have

$$-c(2 - \mathcal{H}^1(S)) \leq \liminf_r \frac{1}{r} P(r(l_r - \mathcal{H}^1(S_r))).$$

For the proof of $\Gamma-$limsup inequality it is sufficient to build a sequence $(S_r)_r$ such that it converges to S, the points x_r^i belong to S_r and we have $\mathcal{H}^1(S_r) \to \mathcal{H}^1(S)$: the convergence of the last term follows then from the asymptotic behavior near 0 of the function P and the first can be estimated the same as in the proof of Lemma (8.2.13). To obtain such a sequence it is sufficient to apply to S an affine transformation sending x^i to x_r^i. The convergence $x_r^i \to x^i$ implies the convergences we need. □

We can now state the last theorem regarding existence of the limit.

Theorem 8.3.8. *Suppose* 0 *is an ordinary point of type* 4*: then there exists the limit of Σ_r/r as $r \to 0$ in the Hausdorff distance and it is a corner composed by two unit rays.*

Proof. Being X a compact metric space (see, for instance, [7]), a a consequence of our previous $\Gamma-$convergence result (Lemma 8.3.7), all limits of Σ_r/r must minimize the functional appearing in Lemma 8.3.7. Moreover, we know that they must be the union of two segments (Lemma 8.2.11). We will now try to identify those pair of radii that may be minimizers, exactly as in the proof of Theorem 8.3.3, in order to have uniqueness of the limits and then the existence of the limit.

Let us consider a ball in which the vertical ray directed upwards is given by the vector $\overline{v}$. We want to show the existence of the limit of Σ_r/r, so we must identify the possible limits of subsequences as a unique one. We stress that this is strongly different from saying that the functional F has a unique minimizer: for every converging subsequence we have a different functional F, depending on the limit points x^i. What we want to do is to show that there exists only one possible choice of x^i, $i = 1, 2$ so that the corner composed by the rays arriving in these two points minimizes the corresponding functional. We may identify the points x^i by means of the angles α, β between the corresponding rays and the horizontal line. Notice that if we have some two rays as a limit of subsequence, the set A_0 has to be contained in the sector having 0 has a vertex and the normal vector to the rays as boundary directions. This implies in particular that $\alpha, \beta \geq 0$.

Consider now the ellipse having x^i as focuses and 2 as the length of the greater axis. The center 0 lies on it. The tangent direction to the ellipse is not horizontal unless $\alpha = \beta$. Any S consisting by two segments joining, in order, x^1, y and x^2, where y lies on such an ellipse can be used as a variation to K (where K is the corner we are taking into consideration as a limit of Σ_r/r) and provides the same value as K to the length-penalizing term. Yet, if y has a positive component in the direction of $\overline{v}$, the integral term turns out to be strictly lesser. This shows that only $\alpha = \beta$ is possible.

We are now going to perform variations in which we move the vertex of the corner up or down to a certain value y of the $\overline{v}-$component. The value of F on the set S obtained in such a way can be estimated by

$$\int_{S^1} v \cdot y v_0 \, \nu(dv) - c(2 - \mathcal{H}^1(S)) = -y|\overline{v}| + c\mathcal{H}^1(S) - 2c,$$

where v_0 is the unit vector in the direction of $\overline{v}$. By $\overline{v}$ we mean the vector calculated at 0, while we denote by $\overline{v}(x_0)$ the one obtained at the endpoint x_0. Notice that $c = |\overline{v}(x_0)|$. We have $\mathcal{H}^1(S) = 2\sqrt{\cos^2\alpha + (y - \sin\alpha)^2}$ and we may write at the first order in y:

$$F(S) = -y|\overline{v}| + cy\sin\alpha + o(y).$$

Optimality of K (i.e. $y = 0$) gives so necessarily $|\overline{v}| = c\sin\alpha$, and this completes the determination of α. □

8.4 Something more on regularity

We present in this section a regularity result, as a by subproduct of our previous analysis.

Theorem 8.4.1. *Let γ be an arc length parameterization of a subset $\Sigma_1 \subset \Sigma$ consisting of a simple curve with no triple junction nor endpoint in its relative interior, such that $k = \sup_{x \in \Sigma_1} diam(T^{-1}\{x\}) < 1/(2C)$. Then $\gamma \in C^{1,1}$ and*

$$|\gamma''| \leq \frac{Ck}{1-2Ck}.$$

Proof. Notice that the condition on the diameters of the transported sets prevents Σ_1 to contain atoms, thanks to Corollary 8.2.16. So, writing Σ_1, if necessary, as a countable union of subsets, we can suppose that it is compactly contained in the complement of triple junctions, endpoints and atoms with mass larger than $(2C)^{-1}$.

Because of semicontinuity, for every point $x \in \Sigma_1$ it will exist a ball $B(x, r_1)$ such that $\mathrm{diam}(T^{-1}(B(x, r_1))) < 1/(2C)$. We can also suppose $r_1 < r_0[\Sigma_1]$ (the radius defined in Theorem 8.2.7). Then for every $y \in B(x, r_1/2)$ we have $\mathrm{diam}(T^{-1}(B(y, r_1/2))) < 1/(2C)$. This means that we can use estimate (8.2.26) in all these points. By using also theorem 8.2.8 we can then say that, whenever $y_1 = \gamma(t_1)$ and $y_2 = \gamma(t_2)$ are points in such a neighborhood at distance r, we can estimate

$$\Delta\theta(y_1, y_2) \leq \frac{Ckr}{1-2Ck} + r^\alpha C(r_1),$$

where $\Delta\theta(y_1, y_2)$ is the angle between the tangent vector to Σ in y_1 and the segment $\overline{y_1 y_2}$ (such an angle can be estimated by the variation of the function θ) and α is an exponent greater than 1. By writing the same inequality interchanging the role of y_1 and y_2, summing up, and taking into account that γ is an arc length parameterization, so that all derivatives are unit vector determined only by the direction of the tangent vector, we get

$$|\gamma'(t_1) - \gamma'(t_2)| \leq \frac{Ckr}{1-2Ck} + r^\alpha C(r_1). \tag{8.4.1}$$

Taking into account that $r = |y_1 - y_2| \leq |t_1 - t_2|$, this implies that γ is locally $C^{1,1}$, and so it has almost everywhere a second derivative. Passing to the limit in (8.4.1) we get

$$|\gamma''(t)| \leq \frac{Ck}{1-2Ck}$$

for almost every t. □

Let us have a look to some consequences. First of all we see that the situation analyzed in theorem 8.3.6 is in fact impossible to be found.

Corollary 8.4.2. *No ordinary point x_0 in Σ is such that $\mathcal{T}(x_0) = \{x_0\}$.*

Proof. Just notice, that, thanks to Theorem 8.4.1, in a neighborhood of such a point we should have a $C^{1,1}$ curve. But for $\gamma \in C^{1,1}$ in every point of the curve we have a positive radius ball to which γ is tangent from outside. This ensures the existence of some more points, different from x_0, which are transported to x_0. □

Next consequence deals with triple junctions and can be considered a quite complete answer to the question about them posed in [27]. We will state it in the form of an all-inclusive theorem.

Theorem 8.4.3. *Suppose that $x_0 \in \Sigma$ is a triple junction: then the three branches of Σ starting from x_0 are parameterized by arc length by $C^{1,1}$ curves at least in a neighborhood of x_0 and have tangent vectors in x_0 which form three $120°$ angles.*

Proof. Just use previously proved results (Theorem 8.3.1) and notice that, due to $T^{-1}(B_r) \subset B_{c_0 r}$, we have $\text{diam}(T^{-1}(x_0)) = 0$, which is enough for Theorem 8.4.1 and local $C^{1,1}$ regularity. □

Chapter 9

Blow-up for optimal branching structures

In this chapter we want to apply some techniques from Chapter 8 to the branched transport problems of Chapter 6. The problem of the existence of the blow-up limits for the optimal structures that arise from them has been pointed out by Xia in [74], who mainly considered the limits up to subsequences. As in chapter 8, a delicate point is proving the existence of full limits. the result we present in this chapter are based on an analysis of the variation of the function θ as in Section 8.2.1. They appear here for the first time, but they have been widely discussed with Jean-Michel Morel in last months.

9.1 Techical tools

For a traffic plan η and a function $c : \mathbb{R}^d \times [0,1] \to [0,+\infty]$, set $E_c(\eta) = \int_{\mathbb{R}^d} c(x,[x]_\eta)\mathcal{H}^1(dx)$. We consider only the case of a subadditive function c, i.e. $c(x,s+t) \leq c(x,s) + c(x,t)$. For $c(x,s) = s^\alpha$ we get back to the usual energy J (see Section 6.2). We will also use the definitions of arcs of η (Defintion 6.2.2) introduced in Chapter 6 and coming from [12]. Here and in the sequel, as in Chapters 6 and 7, we identify sometimes a curve with its image.

9.1.1 Geometric estimates

Lemma 9.1.1. *Let η be a traffic plan and γ an arc of η parametrized on the interval $[t_0, t_1]$ with $\gamma(t_0) = x_0$ and $\gamma(t_1) = x_1$. Set $A = \gamma([t_0, t_1])$,*

$\Gamma_A = \{\tilde\gamma \in \Gamma : A \subset \tilde\gamma\}$ and $\theta_0 = \eta(\Gamma_A)$. Suppose that the inequalities $0 < \theta_0 \le [x]_\eta \le \theta_0(1+\varepsilon)$ hold for any $x \in A$. Fix an arc B from x_0 to x_1 and consider the new traffic plan $\eta' = T_\sharp \eta$ where $T : \Gamma \to \Gamma$ is the identity on $\Gamma \setminus \Gamma_A$ and, for $\tilde\gamma \in \Gamma_A$, the curve $T(\tilde\gamma)$ is the curve which agrees with $\tilde\gamma$ up to the part between x_0 and x_1 which is replaced by B. Then we have

$$E_c(\eta') \le E_c(\eta) - \theta_0^\alpha \mathcal{H}^1(A)\left((1+\varepsilon)^\alpha - \varepsilon^\alpha\right) + \int_B c(x,\theta_0)\mathcal{H}^1(dx).$$

Proof. We only have to evaluate $\int_A (\theta(x)^\alpha - (\theta(x)-\theta_0)^\alpha)\,\mathcal{H}^1(dx)$, because the integral on B will be estimated by subadditivity. To evaluate this other term we act as in Theorem 7.4.3. Since the function $s \mapsto s^\alpha - (s-\theta_0)^\alpha$ is decreasing and we know $\theta(x) \le (1+\varepsilon)\theta_0$, we get $\theta(x)^\alpha - (\theta(x)-\theta_0)^\alpha \ge \theta_0^\alpha\left((1+\varepsilon)^\alpha - \varepsilon^\alpha\right)$. Hence we may estimate from below the gain by means of

$$\mathcal{H}^1(A)\theta_0^\alpha\left((1+\varepsilon)^\alpha - \varepsilon^\alpha\right),$$

and then we get the thesis. □

Lemma 9.1.2. *Let η be an optimal traffic plan for J. According to the notations of the previous Lemma, let B be the straight line segment from x_0 to x_1. Suppose again $0 < \theta_0 \le [x]_\eta \le \theta_0(1+\varepsilon)$ for any $x \in A$. Then it holds $\mathcal{H}^1(A)\left((1+\varepsilon)^\alpha - \varepsilon^\alpha\right) \le \mathcal{H}^1(B)$. It also follows*

$$\mathcal{H}^1(A) \le \mathcal{H}^1(B)(1+c\varepsilon^\alpha),$$

for a suitable constant c.

Proof. We consider the new traffic plan η' built as in Lemma 9.1.1. By the optimality of η, we should have $J(\eta') \ge J(\eta)$. Yet, by using Lemma 9.1.1 we get

$$J(\eta') \le J(\eta) - \mathcal{H}^1(A)\theta_0^\alpha\left((1+\varepsilon)^\alpha - \varepsilon^\alpha\right) + \mathcal{H}^1(B)\theta_0^\alpha,$$

which gives the thesis. The final estimate, on the other hand, comes just from a development of $\left((1+\varepsilon)^\alpha - \varepsilon^\alpha\right)^{-1}$ near 0. □

Remark 9.1.3. Notice that in particular, as a consequence of the previous Lemma, the length of a curve of η is always bounded by a constant times the length of the straight line segment between its endpoints, provided ε is sufficiently small.

Lemma 9.1.4. *Under the same notations of Lemma 9.1.2, it holds $\mathcal{H}^1(A) \ge \sqrt{\mathcal{H}^1(B)^2 + 4d_H(A,B)^2}$, where d_H denotes the Hausdorff distance.*

The proof of Lemma 9.1.4 is in fact contained in Lemma 8.2.2 in the previous chapter.

Lemma 9.1.5. *Under the same notations of Lemma 9.1.2, suppose that* η *is an optimal traffic plan and that* γ *is an arc of* η *joining* x_0 *to a point* $x_1 \in \partial B(x_0, R)$*. Moreover, let* $x_2 \in A \cap \partial B(x_0, r)$ *be a point such that the angle* $x_2 \hat{x_0} x_1 = \Delta\theta$*. Suppose* $r \geq R/2$*. Then we have* $\Delta\theta \leq C\varepsilon^{\alpha/2}$*, where* C *is a constant depending only on* α*.*

Proof. We need to notice that $d_H(A, B) \geq cr\Delta\theta$, and, combining the inequalities in Lemmas 9.1.2 and 9.1.4, and inserting this last estimate, we get

$$\sqrt{R^2 + cr^2(\Delta\theta)^2} \leq R(1 + c\varepsilon^\alpha).$$

Then we use $r \geq R/2$ and $\sqrt{1 + x^2} \geq 1 + cx^2$ (which is true as far as x stays bounded, and this is satisfied by $\Delta\theta$) to get

$$1 + c(\Delta\theta)^2 \leq 1 + c\varepsilon^\alpha,$$

and then the thesis. □

9.1.2 Concatenation of traffic plans

We devote this subsection to an useful tool that we need in this Lagrangian setting, i.e. the quite natural concept of concatenation of traffic plans. The idea here is that if we have two traffic plans η_1 and η_2 with the terminal measure of η_1 equal to the starting measure of η_2, then it is possible to merge them in order to form a new traffic plan. In the language of traffic plan (the Lagrangian approach to branching transport) this is not completely trivial and requires an ad-hoc procedure, while in the Eulerian approach this is very easy, as it simply corresponds to the sum of two vector measures.

We need first to introduce the operation of gluing curves: if we are given two curves γ_1, $\gamma_2 \in \Gamma$ such that $\pi_\infty(\gamma_1) = \pi_0(\gamma_2)$ we denote by $g(\gamma_1, \gamma_2)$ the curve given by

$$g(\gamma_1, \gamma_2)(t) = \begin{cases} \gamma_1(t) & \text{if } t \leq \sigma(\gamma_1), \\ \gamma_2(t - \sigma(\gamma_1)) & \text{if } t \geq \sigma(\gamma_1). \end{cases}$$

We also denote by $GP(\Gamma) \subset \Gamma \times \Gamma$ the set of gluable pairs of curves, i.e. $GP(\Gamma) = \{(\gamma_1, \gamma_2) \in \Gamma \times \Gamma \,:\, \pi_\infty(\gamma_1) = \pi_0(\gamma_2)\}$. The application g is a well-defined map from $GP(\Gamma)$ to Γ.

Lemma 9.1.6. *If we are given two traffic plans η_1 and η_2 with $(\pi_0)_\sharp \eta_i = \mu_i$ and $(\pi_\infty)_\sharp \eta_i = \nu_i$, with $\nu_1 = \mu_2$, there exists a traffic plan η with the following properties:*

- $(\pi_0)_\sharp \eta = \mu_1$ *and* $(\pi_\infty)_\sharp \eta = \nu_2$*;*
- $[x]_\eta \leq [x]_{\eta_1} + [x]_{\eta_2}$ *for any $x \in \Omega$ and in particular $E_c(\eta) \leq E_c(\eta_1) + E_c(\eta_2)$ for any subadditive c;*
- *if Γ_1 and Γ_2 are subsets of Γ such that η_1 is concentrated on Γ_1 and η_2 on Γ_2, then η is concentrated on $g(GP(\Gamma) \cap (\Gamma_1 \times \Gamma_2))$.*

Proof. The proof follows a the idea of the gluing lemma in [71]. Let us disintegrate η_1 and η_2 with respect to $\nu_1 = \mu_2$. We get two families of measures $(\eta_1^x)_x$ and $(\eta_2^x)_x$ which give the two disintegrations:

$$\eta_1 = \int_\Omega \eta_1^x \, \nu_1(dx) \quad \text{and} \quad \eta_2 = \int_\Omega \eta_2^x \, \mu_2(dx),$$

and the measures η_1^x and η_2^x are concentrated on the curves stopping at x and leaving x, respectively. Take, for any x, the measure $\eta_1^x \otimes \eta_2^x$: all these measures are probabilities on $\Gamma \times \Gamma$ concentrated on $GP(\Gamma)$. Then consider

$$\eta = g_\sharp \left(\int_\Omega \eta_1^x \otimes \eta_2^x \nu_1(dx) \right).$$

This is the desired measure and it is easy to check the desired properties. To check the first, just notice that $\pi_0(g(\gamma_1, \gamma_2)) = \pi_0(\gamma_1)$ and analogously $\pi_\infty(g(\gamma_1, \gamma_2)) = \pi_\infty(\gamma_2)$. Hence the marginals of η come from the marginals of η_1 and η_2. The last property is a consequence of the definition of η: set $\rho = \int_\Omega \eta_1^x \otimes \eta_2^x \nu_1(dx)$ and see that

$$\begin{aligned} \eta\left(g\left(GP(\Gamma) \cap (\Gamma_1 \times \Gamma_2)\right)\right) &\geq \rho(GP(\Gamma) \cap (\Gamma_1 \times \Gamma_2) = \\ &\int_\Omega \nu_1(dx) \, \eta_1^x \otimes \eta_2^x \left(GP(\Gamma) \cap (\Gamma_1 \times \Gamma_2)\right) = \\ &\int_\Omega \nu_1(dx) \, \eta_1^x \otimes \eta_2^x (\Gamma_1 \times \Gamma_2) = \int_\Omega \nu_1(dx) \, \eta_1^x(\Gamma_1) \, \eta_2^x(\Gamma_2) = 1, \end{aligned}$$

where the first inequality comes from the definition of image measure and then we have used the fact that any measure $\eta_1^x \otimes \eta_2^x$ is concentrated on $GP(\Gamma)$ and at the end that $\eta_i^x(\Gamma_i) = 1$ for ν_1-a.e. x and $i = 1, 2$. To verify the second requirement, just notice that

$$\begin{aligned} [x]_\eta = \eta\left(\{\gamma \, : \, x \in \gamma\}\right) &= \rho\left(\{(\gamma_1, \gamma_2) \, : \, x \in \gamma_1 \cup \gamma_2\}\right) \\ &\leq \rho\left(\{(\gamma_1, \gamma_2) \, : \, x \in \gamma_1\}\right) + \rho\left(\{(\gamma_1, \gamma_2) \, : \, x \in \gamma_2\}\right). \end{aligned}$$

Then we use the fact that the two marginals of ρ, which is a measure on $\Gamma \times \Gamma$, are η_1 and η_2 and we get in the end

$$[x]_\eta \leq \eta_1 (\{\gamma_1 : x \in \gamma_1\}) + \eta_2 (\{\gamma_2 : x \in \gamma_2\}) = [x]_{\eta_1} + [x]_{\eta_2}. \quad \square$$

9.2 Blow-up at branching points

In this section we will use the estimates and the preliminaries that we presented above to get a blow-up result. This result is very weak as it requires several assumptions on the two measures μ and ν (on their summability and support). Moreover, it will be valid only at branching points of optimal traffic plan in $\mathbb{R}^2$. To precise the assumptions on the measures we need to introduce the following concept:

Definition 9.2.1. We say that the pair (μ, ν) satisfies the *regularity assumption* if either one of the two measures μ or ν is finitely atomic or $\operatorname{spt}(\mu) \cap \operatorname{spt}(\nu) = \emptyset$.

Actually most of the results in [14] have been proven under the assumption that (μ, ν) satisfies the regularity assumption and it is not surprising that this hypothesis appears in this framework too.

Notice that, if $\operatorname{spt}(\mu) \cap \operatorname{spt}(\nu) = \emptyset$, by applying the interior regularity results in [14], we know that far away from $\operatorname{spt}(\mu) \cup \operatorname{spt}(\nu)$ the traffic plan is in fact a finite graph. Suppose now $x_0 \notin \operatorname{spt} \mu$ and let ω be an open set containing $\operatorname{spt} \mu$ but such that $d(\omega, \operatorname{spt}(\nu)) > 0$. This means that, locally around x_0, the situation is the same as if we replaced μ by $\mu' = (T_\omega)\sharp\eta$, where $T_\omega : \Gamma \to \partial\omega$ is the the map given by

$$T_\omega(\gamma) = \gamma(t_0) \quad \text{ if } t_0 = \inf\{t : \gamma(t) \notin \omega\}.$$

This measure μ', by the interior regularity result, is a finite atomic measure. So we may always think that the starting measure is finitely atomic (at least if we are looking at a neighborhood of a point far from the support of the starting measure: if it is not so, this means that it is far from the support of the arrival measure and we may reverse the situation by switching the roles of the two measures).

Remark 9.2.2. This procedure replacing μ by μ' corresponds to taking the spatial restriction of the traffic plan η to $\mathbb{R}^2 \setminus \omega$ and then taking the initial measure of the new traffic plan.

We now recall what is known about the behavior of an optimal traffic plan η near a point x_0 and its limits of blow-up procedures. Most of the notions come from [74] or [14]. In [14] the definition of *connected component* of a traffic plan in open set is given. By applying it to the open set $\mathbb{R}^d \setminus \{x_0\}$ we get the definition of the cuts of η around x_0: any cut of η is a branch around x_0. In [14] it is proven that there are finitely many cuts for any point x_0 and an uniform upper bound is given for this number. Any cut gives a new traffic plan with its starting and arrival measures. We say that a cut is *entering in* x_0 if it irrigates a measure of the form $\tilde{\nu} + a_0\delta_{x_0}$ with $\tilde{\nu} \le \nu$ or *leaving* x_0 if its starting measure is of the form $\tilde{\mu} + a_0\delta_{x_0}$ with $\tilde{\mu} \le \mu$. Any cut is an optimal traffic plan between its starting and its irrigated measure; in any cut any two arcs leaving x_0 (or arriving at) agree on a common initial (ending) path; consequently if γ is an arc of the traffic plan leaving x_0 at time t_0, we have $\lim_{t\to t_0}[\gamma(t)]_\eta = a_0$.

We now want to precise what we mean by blow-up in this setting:

Definition 9.2.3. Fix an optimal traffic plan η and a point x_0 and take for each cut of η at x_0 a curve γ_i $(i = 1, \dots, I)$ arriving at or starting from x_0. We say that η admits a blow-up limit at x_0 on the subsequence $(r_j)_j$ with $r_j \to 0$ if there are sets $K_i \subset \overline{B(0,1)}$ such that

$$\frac{\gamma_i \cap \overline{B(0,r_j)} \,-\, x_0}{r_j} \to K_i$$

in the Hausdorff sense for any $i = 1, \dots, I$. This definition does not depend on the choice of the curves γ_i since two possible curves in the same cut coincide on an initial arc near x_0. If the limit exists as a full limit for $r \to 0$ we say that η admits a blow-up limit at x_0 and we will call limit the set $K = \bigcup_{i=1}^{I} K_i$.

In [74] the interest was towards the blow-up limits in the sense of currents: here we will on the contrary look at the Hausdorff limits of the blow-up of the arcs of the cuts. The situation thus corresponds to what we have seen in Chapter 8. Thanks to the correspondence between traffic plans and currents (or vector measures) that has been pointed out in [14] and to the equivalences that have been proven in the same paper, we know that the two notions of blow-up agree. We also know that, as far as the the blow-up limits up to subsequences are concerned, they are a union of a finite number of segments, one for each cut, with directions $\hat{n}_i$ such that $\sum_i^N \theta_i^\alpha \hat{n}_i = 0$. The number N of cuts is uniformly estimated by a constant $N(d, \alpha)$. The important fact is that we do not know whether these are limits up to subsequences or full limits. To prove that the full limit exists we will prove that, on any arc γ arriving at x_0, the angle direction $\theta(r)$ of the point $\gamma \cap \partial B(x_0, r)$ satisfies a Cauchy condition as $r \to 0$ (as we pointed out in Remark 8.2.10). We want to estimate $|\theta(r_0) - \theta(r_1)|$: set $2^{-n} > r_0 \geq 2^{-(n+1)}$ and $2^{-m} > r_1 \geq 2^{-(m+1)}$ with $m \geq n$. Then we write

$$|\theta(r_0) - \theta(r_1)| \leq |\theta(r_0) - \theta(2^{-(n+1)})| + |\theta(2^{-m}) - \theta(r_1)| + \sum_{j=n+1}^{m-1} \Delta\theta_j,$$

where $\Delta\theta_j$ is the angle variation between radii 2^{-j} and $2^{-(j+1)}$. Let us define $\varepsilon(r)$ as the incremental excess of the multiplicity on $\gamma \cap B(x_0, r)$:

$$\varepsilon(r) = \frac{\max_{x \in \gamma \cap B(x_0,r)} [x]_\eta}{\min_{x \in \gamma \cap B(x_0,r)} [x]_\eta} - 1.$$

The quantity $\varepsilon(r)$ is increasing in r and so we may estimate, thanks to Lemma 9.1.5

$$|\theta(r_0) - \theta(r_1)| \leq C \sum_{j=n}^{m} \varepsilon(2^{-j})^{\alpha/2}.$$

To get $|\theta(r_0) - \theta(r_1)|$ as small as we want it is hence sufficient to have

$$\sum_{j=0}^{\infty} \varepsilon(2^{-j})^{\alpha/2} < +\infty, \tag{9.2.1}$$

so that the tail of the sum is infinitesimal. Notice that such a condition is always verified if we have $\varepsilon(r) \leq Cr^\beta$ for a suitable $\beta > 0$.

Now take a point x_0 and an arc γ in one of its cuts. Since μ is finitely atomic, by using the no-loop property, we get that only at a finite number of points of γ there could be some mass which arrives (no more than once for every Dirac mass of μ, even less if they merge before arriving at γ). We are concerned with the arrival points which are not x_0 itself, so that we may simply choose a sufficiently small radius r in order not to have any arrival point in $\gamma \cap B(x_0, r)$. This means that the multiplicity on the curve γ is decreasing, at least if we stay close enough to x_0 and its variation inside $B(x_0, r)$ is due only to the departing mass. Let us now consider the mass irrigated starting from $\gamma \cap B(x_0, r)$. This in fact means taking all the connected components of η in $\mathbb{R}^2 \setminus (\gamma \cap B(x_0, r))$ which touch the relative interior of $\gamma \cap B(x_0, r)$: the traffic plan corresponding to the union of all these last components is an optimal traffic plan which irrigates a measure ν_r starting from a measure μ_r concentrated on $\gamma \cap B(x_0, r)$, with the same mass that we denote by $m(r)$. Call θ_0 the maximal value of the multiplicity on $\gamma \cap B(x_0, r)$. This is the value of the mass at x_0 in the irrigating measure of the cut we are considering. The minimal value, on the contrary, is at x_1, and it is given by $\theta_0 - m(r)$. We have

$$\varepsilon(r) = \frac{\theta_0}{\theta_0 - m(r)} - 1 = \frac{m(r)}{\theta_0 - m(r)}.$$

Since $\theta_0 > 0$ and $m(r)$ is infinitesimal, at least for small r it holds

$$\varepsilon(r) \leq 2\theta_0^{-1} m(r) \tag{9.2.2}$$

and hence estimating $\varepsilon(r)$ is the same as estimating $m(r)$. To do this, we will give a geometric estimate on the area of the set where the measure ν_r is concentrated. Let us call $R(r)$ the highest radius R such that $\nu_r(B(x_0, R)) \leq m(r)/2$.

We will prove the following result.

Lemma 9.2.4. *Suppose x_0 is a branching point of an optimal traffic plan $\overline{\eta}$ in $\mathbb{R}^2$ and that (μ, ν) satisfies the regularity assumption: then there exists a constant k, depending on α, η, x_0, such that for any small r we have $R(r) \leq kr$.*

Proof. We will prove the result by contradiction. Supposing the result to be false, we could build a sequence of measures μ_n whose mass is $m_n \to 0$, each concentrated on $\gamma \cap B(x_0, r_n)$ with $r_n \to 0$, irrigating through a sequence of traffic plans η_n some corresponding measures ν_n. The measures η_n are traffic plans in a quite enlarged sense, as they are not probability measures but they have total mass m_n. Anyway the whole theory extends easily to finite-mass measures on Γ. We are assuming that the radius $R(r_n)$ (the maximal radius to have $\nu_n(B(x_0, R_n)) \leq m_n/2$) is much larger than r_n, in the sense that it holds $R(r_n)/r_n \to +\infty$. Set $N = \cup_i \gamma_i$, where the curves γ_i's are chosen an arc for each cut. We notice that the traffic plans η_n are optimal according to the usual costs J, but also with respect to the energy E_{c_n} given by

$$c_n(x, \theta)) = \begin{cases} \theta^\alpha & \text{if } x \notin N \\ \alpha [x]_{\overline{\eta}}^{\alpha - 1} \theta \wedge \theta^\alpha & \text{if } x \in N. \end{cases}$$

This fact is true since, if we consider a new current η' replacing η_n, we may concatenate it into the optimal traffic plan $\overline{\eta}$ and estimate the difference between the total energy of the new traffic plan and the energy of $\overline{\eta}$ by subadditivity outside N, while on N there is already a positive multiplicity:

adding some multiplicity on N is hence cheaper than doing it elsewhere, and its cost may be estimated by concavity.

Afterwards, we can replace any measure ν_n by the measure we get by stopping the irrigation at $\partial B(x_0, R_n)$. In this way at least half of the mass of the new measure is on the boundary. We can easily renormalize in space what we have so far by a translation (bringing x_0 to 0) and an homothety of ratio R_n^{-1}. In this way we have new sequences μ'_n with $d_H(\mathrm{spt}(\mu'_n), 0) \to 0$ and ν'_n with $\nu'_n(\partial B(0,1)) \geq 1/2m_n$, where m_n is the common mass of μ'_n and ν'_n. After this we renormalize also in mass, multiplying by m_n^{-1}. In this way we get two sequences of probability measures μ''_n, ν''_n and a new sequence of traffic plans η''_n (with mass 1, after the rescaling). This traffic plans are optimal both for the usual energy and for the energy $E''_n := E_{c''_n}$, derived by renormalization by E_{c_n} and given by:

$$c''_n(x,\theta) = \begin{cases} \theta^\alpha & \text{if } x \notin A_n \\ m_n^{1-\alpha}\theta_0(x)^{\alpha-1}\theta \wedge \theta^\alpha & \text{if } x \in A_n, \end{cases}$$

where $A_n = R_n^{-1}(N - x_0) \cap \overline{B(0,1)}$ and $\theta_0(x)$, for $x \in A_n$ is the multiplicity of $\overline{\eta}$ at the point $R_n x + x_0 \in N$. To get this expression it is sufficient to consider the expression of E_{c_n}, renormalize in space, and then in mass. Then we divide the result by m_n^α. Hence it holds

$$d_\alpha(\mu''_n, \nu''_n) = E''_n(\eta''_n) \leq E''_n(\eta),$$

where the inequality holds for any traffic plan $\eta \in TP(\mu''_n, \nu''_n)$. Now we want to let n tend to infinity. Up to subsequences, we may suppose $\mu''_n \rightharpoonup \mu_\infty$, $\nu''_\infty \rightharpoonup \nu_\infty$, $A_n \to A_\infty$ (in the Hausdorff sense). It is straightforward that we have $\mu_\infty = \delta_0$. We also want a limit energy for E''_n. Let us set $E_\infty = E_{c_\infty}$, where c_∞ is given by

$$c_\infty(\theta, x) = \begin{cases} \theta^\alpha & \text{if } x \notin A_\infty \\ 0 & \text{if } x \in A_\infty, \end{cases}.$$

We want now to prove

$$d_\alpha(\mu_\infty, \nu_\infty) \leq E_\infty(\eta),$$

for any $\eta \in TP(\mu_\infty, \nu_\infty)$. It is actually sufficient to prove this for traffic plan whose support is contained in $\overline{B(0,1)}$ (by the convex-hull property, see [14]). In order to do this we can use the approximation in Lemma 9.2.5 and then pass to the limit in the inequality. Since the energy E_∞ allows to move on A_∞ for free, we get

$$\min\{E_\infty(\eta) \,:\, \eta \in TP(\delta_0,\nu_\infty)\} = \min\{d_\alpha(\mu,\nu_\infty \,:\, \mathrm{spt}(\mu)\subset A_\infty\}.$$

Notice that, by the results on limits of the blow-up procedure up to subsequences, the set A_∞ is the union of finitely many rays which form convex angles between them. Hence we may apply Lemma 9.2.6, since $\nu_\infty \neq \delta_0$, and get a contradiction. □

Lemma 9.2.5. *Any traffic plan $\eta \in TP(\delta_0,\nu_\infty)$ whose support is contained in $\overline{B(0,1)}$ may be approximated by a sequence of traffic plan $(\eta_n)_n$ with $\eta_n \in TP(\mu_n'',\nu_n'')$, such that* $\limsup_n E_n''(T_n) \le E_\infty(T)$.

Proof. Let us consider the traffic plan η' which is the restriction of η outside A_∞ and its starting measure σ, concentrated on A_∞. Let $\pi_n : A_\infty \to A_n$ be a measurable map such that $|\pi_n(x)-x| = d(x,A_n) \le d_H(A_\infty,A_n)$ (its existence comes from a measurable selection criterion, see [38]). We consider a traffic plan η_n which is obtained by concatenating the following four traffic plans in order: η_n^1, η_n^2, η' and η_n^4. Here $\eta_n^1 \in TP(\mu_n'',(\pi_n)_\sharp\sigma)$ is a traffic plan which is optimal for the energy E_n''; $\eta_n^2 \in TP((\pi_n)_\sharp\sigma,\sigma)$ and $\eta_n^4 \in TP(\nu_\infty,\nu_n'')$ are optimal traffic plans. We want to estimate $E_n''(\eta_n)$. We have

$$E_n''(\eta_n) \le E_n''(\eta') + E_n''(\eta_n^1) + E_n''(\eta_n^2) + E_n''(\eta_n^4).$$

Notice that, being $c_n''(\cdot,x)$ linear for $x \in A_n$, the cost $E_n''(\eta_n^1)$ coincides, up to the multiplicative constant $m_n^{1-\alpha}$, with the Wasserstein distance from μ_n'' to $(\pi_n)_\sharp\sigma$ according to the geodesic distance on A_n with a weight given by $\theta_0(x)^{\alpha-1}$. This weight being bounded, we deduce (thanks to Remark 9.1.3 too, which implies $\mathcal{H}^1(A_n) \le C$)

$$E_n''(\eta_n^1) \le C m_n^{1-\alpha}\mathcal{H}^1(A_n) \to 0.$$

To estimate $E_n''(\eta_n^2)$we use $E_n''(\eta_n^2) \le E(\eta_n^2) = d_\alpha((\pi_n)_\sharp\sigma,\sigma)$. Then, thanks to $d_H(A_n,A_\infty)\to 0$, we get $(\pi_n)_\sharp\sigma \rightharpoonup \sigma$, which implies $d_\alpha((\pi_n)_\sharp\sigma,\sigma)\to 0$. Similarly one can get $E_n''(\eta_n^4)\to 0$ as a consequence of $\nu_n'' \rightharpoonup \nu_\infty$. In the end we get

$$\limsup_n E_n''(\eta_n) \le E(\eta') + 0 + 0 + 0 = E_\infty(\eta). \quad \square$$

Lemma 9.2.6. *Suppose $d_\alpha(\nu,\mu_0) = \min\{d_\alpha(\nu,\mu) \,|\, \mathrm{spt}(\mu)\subset A_0\}$, where $A_0 \subset \overline{B(0,1)} \subset \mathbb{R}^2$ is a set composed by finitely many rays meeting at 0 with convex angles. Then, if $\nu \neq \delta_0$, we also have $\mu_0 \neq \delta_0$.*

Proof. To prove this we just take an optimal traffic plan $\eta \in TP(\delta_0, \nu)$ and an arc going out of 0. Let x_1 be a point in this arc different from 0 and let A be the part of the arc up to x_1. Let ε be such that $\theta_0 \leq [x]_\eta \leq \theta_0(1+\varepsilon)$ for any $x \in A$. We know that $\varepsilon \to 0$ as $x_1 \to 0$. Let S be the sector bounded by the rays of A_0 containing x_1. Since the amplitude of S is $180° - 2\delta < 180°$, there is one of the two rays composing the boundary of S such that the angle between it and the segment $x_1 0$ is less than $90° - \delta$. This implies $d(x_1, A_0) < c|x_1|$ for a constant c depending on δ and strictly less than 1. Let B be the straight line segment from x_1 to $x_0 \in A_0$ realizing $d(x_1, A_0)$ as a length. Then we consider the new traffic plan $\eta' = T_\sharp \eta$, where $T : \Gamma \to \Gamma$ is the identity on the curves which do not contain A and replaces A by B on the curves containing it. By optimality of η and μ_0 we should have $J(\eta) \leq J(\eta')$. Yet we have, by Lemma 9.1.1,

$$J(\eta') \leq J(\eta) - \theta_0^\alpha \mathcal{H}^1(A) \left((1+\varepsilon)^\alpha - \varepsilon^\alpha\right) + \theta_0^\alpha \mathcal{H}^1(B).$$

By using $\mathcal{H}^1(A) \geq |x_1|$ and $\mathcal{H}^1(B) < c|x_1|$ we get

$$J(\eta') \leq J(\eta) - \theta_0^\alpha |x_1| \left((1+\varepsilon)^\alpha - \varepsilon^\alpha - c\right),$$

which is a contradiction as $c < 1$ and $\varepsilon \to 0$. □

The consequence of what we have proven is the following final theorem.

Theorem 9.2.7. *Suppose x_0 is a branching point of an optimal traffic plan η in $\Omega \subset \mathbb{R}^2$ and that (μ, ν) satisfies the regularity assumption. To fix the ideas, let us think that either $x_0 \notin \mathrm{spt}(\mu)$ or μ is finitely atomic. Suppose moreover that $\nu = f \cdot \mathcal{L}^2$ and that $f \in L^p(\Omega)$ for $p > 1$. Then there exists the full limit of the blow up of the traffic plan η at x_0.*

Proof. We know that there exist limits of blow-up up to subsequences and that these limits are composed by finitely many rays. To show that there is actually a full limit, it is sufficient to prove that the angle direction of the points $\gamma_i \cap \partial B(x_0, r)$ have a limit for each i as $r \to 0$. This is the same idea as in Chapter 8. In fact the limit direction should be, by Hausdorff convergence, the direction of the limit ray and this would fix such a direction, giving uniqueness of the subsequences limits. To do this we prove that these angles satisfy a Cauchy condition and we want hence to prove that (9.2.1) holds. We use (9.2.2) and Lemma 9.2.4 to say that

$$\varepsilon(r) \leq 2\theta_0^{-1} m(r) \leq 4\theta_0^{-1} \nu(B(x_0, R^j(r)) \leq 4\theta_0^{-1} \nu(B(x_0, kr)).$$

Then, by the L^p assumption on ν, we get $\nu(E) \leq ||f||_p |E|^{1-1/p}$, and hence

$$\varepsilon(r) \leq C r^{2(1-1/p)}.$$

We have already noticed that this is sufficient to make the estimate (9.2.1) hold. □

Bibliography

[1] S. Agmon, A. Douglis and L. Nirenberg, Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I. *Comm. Pure Appl. Math.* (12) 623–727, 1959.

[2] W. K. Allard and F. J. Almgren, The structure of Stationary One Dimesional Varifolds with Positive Density. *Invent. Math.* (34), no. 2, 83–97, 1976.

[3] L. Ambrosio, Lecture Notes on Optimal Transport Problems, *Mathematical Aspects of Evolving Interfaces,* Springer Verlag, Berlin, Lecture Notes in Mathematics (1812), 1–52, 2003.

[4] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the spaces of probability measures.* Lectures in Mathematics, ETH Zurich, Birkhäuser, 2005.

[5] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation, in *Optimal transportation and applications*, Lecture Notes in Mathematics (CIME Series, Martina Franca, 2001) 1813, L.A. Caffarelli and S. Salsa Eds., 123-160, 2003.

[6] L. Ambrosio and F. Santambrogio. Necessary optimality conditions for geodesics in weighted Wasserstein spaces. *Rend. Lincei Mat. Appl.*, to appear, available at `cvgmt.sns.it` .

[7] L. Ambrosio and P. Tilli, *Topics on analysis in metric spaces.* Oxford Lecture Series in Mathematics and its Applications (25). Oxford University Press, Oxford, 2004.

[8] J. R. Banavar, F. Colaiori, A. Flammini, A. Giacometti, A. Maritan and A. Rinaldo. Sculpting of a Fractal River Basin *Phys. Rev. Lett.* (78), no. 23, 4522–4525, 1997.

[9] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan and A. Rinaldo. Scaling, Optimality, and Landscape Evolution. *J. Stat. Physics* (104), nos. 1/2, 1–48, 2001.

[10] M. Beckmann, A continuous model of transportation, *Econometrica* (20), 643–660, 1952.

[11] M. Beckmann and T. Puu, *Spatial Economics: Density, Potential and Flow*, North-Holland, Amsterdam, 1985.

[12] M. Bernot, Optimal transport and irrigation, *PhD Thesis*, ENS Cachan, 2005, available at `http://perso.crans.org/~bernot`.

[13] M. Bernot, V. Caselles and J.-M. Morel, Traffic plans, *Publicacions Matemàtiques* (49), no. 2, 417–451, 2005.

[14] M. Bernot, V. Caselles and J.-M. Morel, The structure of branched transportation networks, *Calc. Var. and Par. Diff. Eq.*, to appear. Prépublication du CMLA Cachan, available at `http://www.cmla.ens-cachan.fr/Cmla`, 2006.

[15] M. Bernot and A. Figalli, Stability of traffic plans, preprint in preparation.

[16] G. Bouchitté and G. Buttazzo. New lower semicontinuity results for nonconvex functionals defined on measures. *Nonlinear Analysis* (15), no. 7, 679–692, 1990.

[17] G. Bouchitté and G. Buttazzo. Integral representation of nonconvex functionals defined on measures. *Annales de l'Institut Henry Poincaré-Analyse non Linéaire* (9), no. 1, 101–117, 1992.

[18] G. Bouchitté and G. Buttazzo. Relaxation for a class of nonconvex functionals defined on measures *Annales de l'Institut Henry Poincaré-Analyse non Linéaire* (10), no. 3, 345–361, 1993.

[19] G. Bouchitté, C. Jimenez, M. Rajesh, Asymptotique d'un problème de positionnement optimal (French). *C. R. Math. Acad. Sci. Paris Sér. I Math.* (335), no. 10, 853–858, 2002.

[20] A. Brancolini, G. Buttazzo and F. Santambrogio, Path Functionals over Wasserstein spaces. *J. Eur. Math. Soc.* (8), no.3, 415–434, 2006.

[21] Y. Brenier, Décomposition polaire et réarrangement monotone des champs de vecteurs. (French) *C. R. Acad. Sci. Paris Sér. I Math.* (305), no. 19, 805–808, 1987.

[22] Y. Brenier, The Least Action Principle and the Related Concept of Generalized Flows for Incompressible Perfect Fluids. *J. Amer. Math. Soc.*, (2), 225–255, 1989.

[23] Y. Brenier, The dual least action principle for an ideal incompressible fluid. *Arch. Rational Mech. Anal.*, (123), 323–351, 1993.

[24] Y. Brenier, A homogenized model for vortex sheets. *Arch. Rational Mech. Anal.*, (138), 319–353, 1997.

[25] Y. Brenier, Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equation. *Comm. Pure and Appl. Math.*, (52), 411–452, 1999.

[26] H. Brezis, *Analyse Fonctionnelle*, Masson Editeur, Paris, 1983.

[27] G. Buttazzo, E. Oudet and E. Stepanov, Optimal Transportation Problems with Free Dirichlet Regions. *Progress in Nonlinear Diff. Equations and their Applications* (51), 41–65, Birkhauser, 2002.

[28] G. Buttazzo and F. Santambrogio, A model for the optimal planning of an urban area. *SIAM J. Math. Anal.* (37), no. 2, 514–530, 2005.

[29] G. Buttazzo and E. Stepanov, Optimal transportation networks as free Dirichlet regions for the Monge-Kantorovich problem. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (2), no. 4, 631–678, 2003.

[30] G. Buttazzo and E. Stepanov, Minimization Problems for Average Distance Functionals. *Calculus of Variations: Topics from the Mathematical Heritage of Ennio De Giorgi* D. Pallara (ed.), Quaderni di Matematica, series edited by Dipartimento di Matematica, Seconda Università di Napoli, (14), 48–83, 2004.

[31] L. Caffarelli, A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity. *Ann. of Math.* (131), no. 1, 129–134, 1990.

[32] L. Caffarelli, Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation. *Ann. of Math.* (131), no. 1, 135–150, 1990.

[33] L. Caffarelli, Some regularity properties of solutions of Monge Ampère equation. *Comm. Pure Appl. Math.* (44), no. 8-9, 965–969, 1991.

[34] S. Campanato, Proprietà di Hölderianità di alcune classi di funzioni. *Ann. Scuola Norm. Sup. Pisa* (17), 175–188, 1963.

[35] G. Carlier and I. Ekeland, The Structure of Cities, *Journal of Global Optimization*, (29), no. 4, 2004, 371–376.

[36] G. Carlier, C. Jimenez and F. Santambrogio, Traffic congestion and Wardrope equilibrium, in preparation.

[37] G. Carlier and F. Santambrogio, A Variational Model for Urban Planning with Traffic Congestion, *ESAIM Control Optimization and Calculus of Variations*, (11), no. 4, 2005, 595–613.

[38] C. Castaing and M. Valadier *Convex Analysis and Measurable Multifunction*, Springer Verlag, Berlin, 1977.

[39] G. Dal Maso. *An Introduction to* $\Gamma-$*convergence*. Birkhäuser, Basel, 1992.

[40] L. De Pascale and A. Pratelli, Regularity properties for Monge transport density and for solutions of some shape optimization problem. *Calc. Var. Partial Differential Equations* (14), no. 3, 249–274, 2002.

[41] G. Devillanova, Singular structures in some variational problems, *PhD Thesis*, ENS Cachan, 2005.

[42] G. Devillanova and S. Solimini, Elementary properties of optimal irrigation patterns. *Calc. Var. Par. Diff. Eq.*, to appear.

[43] G. Devillanova and S. Solimini, On the dimension of an irrigable measure. Preprint, 2005.

[44] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems.* North-Holland, Amsterdam, 1976.

[45] L. C. Evans W. Gangbo. *Differential Methods for the the Monge-Kantorovich Mass Transfer Problems.* Memoirs of AMS, (137), no. 653, 1999.

[46] M. Giaquinta, *Multiple integrals in calculus of variations and nonlinear elliptic systems*, Princeton University Press, 1983.

[47] D. Gilbarg and N. Trudinger *Elliptic Partial Differential Equations of Second Order*, Springer Verlag, Berlin, 1977.

[48] E. N. Gilbert, Minimum cost communication networks, *Bell System Tech. J.* (46), 2209–2227, 1967.

[49] E. N. Gilbert and H. O. Pollak, Steiner minimal trees. *SIAM J. Appl. Math.* (16), 1–29, 1968.

[50] A. Guionnet, First order asymptotics of matrix integrals; a rigorous approach towards the understanding of matrix models. *Comm. Math. Phys.*, (224), 527–569, 2004.

[51] E. Ijjasz-Vasquez, R.L. Bras, I. Rodriguez-Iturbe, R. Rigon and A. Rinaldo, Are river basins optimal channel networks? *Advances in Water Resources* (16), 69–79, 1993.

[52] C. Jimenez, Optimisation de problèmes de transport (French), *PhD Thesis*, Univ. du Sud-Toulon Var, 2005, available at `http://www.ceremade.dauphine.fr/∼jimenez`.

[53] L. Kantorovich, On the transfer of masses. *Dokl. Acad. Nauk. USSR*, (37), 7–8, 1942.

[54] S. Lisini, Characterization of absolutely continuous curves in Wasserstein spaces. *Calc. Var. Par. Diff. Eq.*, to appear.

[55] G. Loeper, The reconstruction problem for the Euler-Poisson system in cosmology. *Arch. Rational Mech. Anal.*, to appear.

[56] F. Maddalena and S. Solimini, Transport distances and irrigation models, in preparation.

[57] F. Maddalena, S. Solimini and J.-M. Morel, A variational model of irrigation patterns. *Interfaces and Free Boundaries* (5), 391–415, 2003.

[58] R. J. McCann, A convexity principle for interacting gases. *Adv. Math.* (128), no. 1, 153–159, 1997.

[59] G. Monge, Mémoire sur la théorie des déblais et de remblais, *Histoire de l'Académie Royale des Sciences de Paris*, 666–704, 1781.

[60] J.-M. Morel and F. Santambrogio, Comparison of distances between measures, *Appl. Math. Lett.*, (20), No. 4, 427–432, 2007.

[61] J.-M. Morel and F. Santambrogio, Curvature of paths in optimal traffic plans, in preparation.

[62] E. Paolini and E. Stepanov, Optimal transportation networks as flat chains, *Interfaces and Free Boundaries*, to appear. Available at cvgmt.sns.it.

[63] R. Rigon, A. Rinaldo, I. Rodriguez-Iturbe, E. Ijjasz-Vasquez and R. L. Bras. Optimal channel networks: a framework for the study of river basin morphology, *Water Resources Research*, (29) no. 6, 1635–1646, 1993.

[64] I. Rodriguez-Iturbe and A. Rinaldo *Fractal River Basins*. Cambridge University Press, 1997.

[65] I. Rodriguez-Iturbe, A. Rinaldo, R. Rigon, R.L. Bras, E. Ijjasz-Vasquez and A. Marani. Fractal structures as least energy dissipation patterns: the case of river networks, *Geophysical Research Letters* (5), 2854–2860, 1992.

[66] F. Santambrogio, Misure ottime per costi di trasporto e funzionali locali (Italian), *Laurea Thesis, Università di Pisa*, advisor: G. Buttazzo, *available at* www.unipi.it/etd *and* cvgmt.sns.it, 2003.

[67] F. Santambrogio, Transport and Concentration Problems with Interaction Effects, *J. Glob. Opt.*, in press, availbale at arxiv.org .

[68] F. Santambrogio, Optimal Channel Networks, Landscape Function and Branched Transport, *Interfaces and Free Boundaries*, to appear. Available at cvgmt.sns.it.

[69] F. Santambrogio, P. Tilli, Blow-up of optimal sets in the irrigation problem, *J. Geom. Anal.* (15), no. 2, 343–362, 2005.

[70] N. Trudinger and X.-J. Wang, On the second boundary value problem for Monge-Ampere type equations and optimal transportation, 2006. Preprint available at arxiv.org .

[71] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics, AMS, 2003.

[72] Q. Xia, Optimal Paths related to Transport Problems. *Comm. Cont. Math.* (5), no. 2, 251–279, 2003.

[73] Q. Xia, Interior regularity of optimal transport paths. *Calc. Var. Par. Diff. Eq.* (20), no. 3, 283–299, 2004.

[74] Q. Xia, Boundary Regularity of Optimal Transport Paths, 2003. Preprint available at `http://www.math.ucdavis.edu/ qlxia/`

[75] Q. Xia, An Application of Optimal Transport Paths to Urban Transport Networks. *Discrete and Continuous Dynmical Systems*, Supplement Volume, 904–910, 2005.

[76] Q. Xia, The Formation of Tree Leaves. *ESAIM Control Optimization and Calculus of Variations,* to appear. Available at `http://www.math.ucdavis.edu/ qlxia/` .