Geodesics and PDE methods in transport models

Lorenzo Brasco

Università di Pisa
Dottorato di ricerca in Matematica
XXII ciclo
SSD MAT/05

Université Paris-Dauphine
École doctorale EDDIMO
DFR Mathématiques de la Décision

Geodesics and PDE methods in transport models

Ph.D. Thesis

Lorenzo Brasco

October 11th, 2010

Members of the jury:

Luigi Ambrosio
Scuola Normale Superiore
Guy Bouchitté
Université du Sud-Toulon-Var
Yann Brenier
Université de Nice Sophia Antipolis
Giuseppe Buttazzo (advisor)
Università di Pisa
Guillaume Carlier (advisor)
Université Paris-Dauphine
Luigi De Pascale
Università di Pisa
Aldo Pratelli
Università di Pavia
Filippo Santambrogio (co-advisor)
Université Paris-Sud

“Non senza fatiga si giunge al fine”
G. Frescobaldi, IX Toccata, Secondo Libro, 1627

A Fiammetta, Marco e i miei genitori

Acknowledgements

I would like to express my sincere gratitude to my advisors, Giuseppe Buttazzo and Guillaume Carlier, for their constant support, for their suggestions (the mathematical and the non-mathematical ones) and for having proposed me such an interesting subject for my Ph.D. thesis. Both of them helped me a lot, sharing with me their deep mathematical knowledge and introducing me into the realm of the Calculus of Variations: I am sure that without their encouragements and optimism, I would have not been able to arrive at the end of this work, overcoming many obstacles and hard times.

Besides my advisors, I want to warmly thank Filippo Santambrogio. His criticisms, his skills and his great intuition have given an incisive contribution to this thesis. I must admit that in these years of Ph.D. studies, he has been much more than a simple co-advisor, but rather an enlightening guide and a *model* (a word that occurs many times in this work...). And a friend as well! Now that he has become professor in Orsay, I am very happy for him and I wish to express my pride for having been a sort of first student for him: I am sure that many others will come.

I am also grateful to Rose-Anne Dana and Fabrizio Broglia, for their invaluable work as directors of the Ph.D. programs in Paris-Dauphine and Pisa, respectively.

Rolando Magnanini deserves a special mention: as the advisor of my degree thesis, he has been the first to introduce me to mathematical research. Moreover, when I left Florence after my graduation, he was the one that gave me the precious suggestion of asking Giuseppe Buttazzo to be my Ph.D. advisor. Then this thesis straightforwardly follows from that suggestion (well, maybe not so straightforwardly...).

Finally, I thank Nayam Al-Hassem and Matteo Scienza for these years shared together as Ph.D. students in Pisa.

Sunto: Questa tesi è dedicata allo studio di problemi di trasporto ottimo, alternativi al cosiddetto problema di Monge-Kantorovich: essi appaiono in modo naturale in alcune applicazioni concrete, come nel disegno di reti ottimali di trasporto o nella modellizzazione di problemi di traffico urbano. In particolare, si considerano problemi in cui il costo di trasporto ha una dipendenza non lineare dalla massa: tipicamente in questo tipo di problemi, muovere una massa m per un tratto di lunghezza ℓ costa $\varphi(m)\,\ell$, dove φ è una funzione assegnata, dando perciò luogo ad un costo totale del tipo $\sum \varphi(m)\ell$.

Due casi significativi vengono ampiamente trattati in questo lavoro: il caso in cui la funzione φ è subadditiva (trasporto ramificato), ragion per cui le masse hanno maggiore interesse a viaggiare insieme, in modo da diminuire il costo totale; il caso in cui φ è superadditiva (trasporto congestionato), dove al contrario la massa tende a diffondersi quanto più possibile.

Nel caso del trasporto ramificato, si introducono due nuovi modelli dinamici: nel primo il trasporto è descritto da curve di misure di probabilità che minimizzano un funzionale di tipo geodetico (con un coefficiente penalizzante le misure non atomiche). Il secondo invece è maggiormente nello spirito della formulazione data da Benamou e Brenier per le distanze di Wasserstein: in particolare, il trasporto è descritto per mezzo di coppie "curva di misure–campo di velocità", legate dall'equazione di continuità, che minimizzano un'opportuna energia (non convessa). Per entrambi i modelli, si mostra l'esistenza di configurazioni minimali e si prova l'equivalenza con altre formulazioni esistenti in letteratura.

Per quanto riguarda il caso del trasporto congestionato, si rivedono in dettaglio due modelli già esistenti, provandone l'equivalenza: mentre il primo di questi modelli può essere visto come un approccio Lagrangiano al problema ed ha interessanti legami con questioni di equilibrio per il traffico urbano, il secondo è un problema di ottimizzazione convessa con vincolo di divergenza.

La dimostrazione dell'equivalenza tra i due modelli costituisce il corpo centrale della seconda parte del lavoro e contiene vari elementi di interesse, tra questi: la teoria dei flussi di campi vettoriali poco regolari di DiPerna e Lions, la costruzione di Dacorogna e Moser per mappe di trasporto e soprattutto dei risultati di regolarità (che quivi ricaviamo) per un'equazione ellittica molto degenere, che non sembra essere stata molto studiata.

Parole chiave: Trasporto ottimo, analisi in spazi metrici, trasporto ramificato, equazioni ellittiche degeneri, problemi variazionali non convessi, Di Perna-Lions, congestione di traffico, equazione di continuità, equilibrio di Wardrop

Résumé : Cette thèse est dédiée à l'étude des problèmes de transport optimal, alternatifs au problème de Monge-Kantorovich : ils apparaissent naturellement dans des applications pratiques, telles que la conception des réseaux de transport optimal ou la modélisation des problèmes de circulation urbaine. En particulier, nous considérons des problèmes où le coût du transport a une dèpendance non linèaire de la masse : typiquement dans ce type de problèmes, le côut pour déplacer une masse m pour une longueur ℓ est $\varphi(m)\,\ell$, où φ est une fonction assignée, obtenant ainsi un coût total de type $\sum \varphi(m)\ell$.

Deux cas importants sont abordés en détail dans ce travail : le cas où la fonction φ est subadditive (transport branché), de sorte que la masse a intérêt à voyager ensemble, de manière à réduire le coût total; le cas où φ est superadditive (transport congestionné), où au contraire, la masse tend à diffuser autant que possible.

Dans le cas du transport branché, nous introduisons deux nouveaux modèles: dans le premièr, le transport est décrit par des courbes de mesures de probabilité que minimisent une fonctionnelle de type géodésique (avec un coefficient que pénalise le mesures qui ne sont pas atomiques). Le second est plus dans l'esprit de la formulation de Benamou et Brenier pour les distances de Wasserstein : en particulier, le transport est décrit par paires de "courbe de mesures–champ de vitesse", liées par l'équation de continuité, qui minimisent une énergie adéquate (non convexe). Pour les deux modèles, on démontre l'existence de configurations minimales et l'équivalence avec d'autres formulations existantes dans la littèrature.

En ce qui concerne le cas du transport congestionné, nous passons en revue deux modèles déjà existants, afin de prouver leur équivalence: alors que le premier de ces modèles peut être considéré comme une approche Lagrangienne du problème et il a des liens intéressants avec des questions d'équilibre pour la circulation urbaine, le second est un problème d'optimisation convexe avec contraintes de divergence.

La preuve de l'équivalence entre les deux modèles constitue le corps principal de la deuxième partie de cette thèse et contient différents éléments d'intérêt, y compris: la théorie des flots des champs de vecteurs peu réguliers (DiPerna-Lions), la construction de Dacorogna et Moser pour les applications de transport et en particulier les résultats de régularité (que nous prouvons ici) pour une équation elliptique très dégénérés, qui ne semble pas avoir été beaucoup étudiée.

Mots clés : Transport optimal, analyse dans les espaces métriques, transport branché, équations elliptiques dégénérées, problèmes non convexes du calcul des variations, Di Perna-Lions, congestion du trafic, équation de continuité, équilibre de Wardrop

Abstract: This thesis is devoted to to the study of optimal transport problems, alternative to the so called Monge-Kantorovich one: they naturally arise in some real world applications, like in the design of optimal transportation networks or in urban traffic modeling. More precisely, we consider problems where the transport cost has a nonlinear dependence on the mass: typically in this type of problems, to move a mass m for a distance ℓ costs $\varphi(m)\,\ell$, where φ is a given function, thus giving rise to a total cost of the type $\sum \varphi(m)\,\ell$.

Two interesting cases are widely addressed in this work: the case where φ is subadditive (branched transport), so that masses have the interest to travel together in order to lower the total cost; the case of φ being superadditive (congested transport), where on the contrary the mass tends to be as widespread as possible.

In the case of branched transport, we introduce two new dynamical models: in the first one, the transport is described through the employ of curves of probability measures minimizing a wighted-length functional (with a weight function penalizing non atomic measures). On the other hand, the second model is much more in the spirit of the celebrated Benamou-Brenier formulation for the Wasserstein distances: in particular, the transport is described by means of pairs "curve of measures–velocity vector field", satisfying the continuity equation and minimizing a suitable dynamical energy (which is a non convex one, actually). For both models we prove existence of minimal configurations and equivalence with other modelizations existing in literature.

Concerning the case of congested transport, we review in great details two already existing models, proving their equivalence: while the first one can be viewed as a Lagrangian approach to the problem and it has some interesting links with traffic equilibrium issues, the second one is a divergence-constrained convex optimization problem.

The proof of this equivalence represents the central core of the second part of the work and contains various points of interest: among them, the DiPerna-Lions theory of flows of weakly differentiable vector fields, the Dacorogna-Moser construction for transport maps and, above all, some regularity estimates (that we derive here) for a very degenerate elliptic equation, that seems to be quite unexplored.

Keywords: Optimal transport, analysis in metric spaces, branched transport, degenerate elliptic equations, nonconvex variational problems, Di Perna-Lions, traffic congestion, continuity equation, Wardrop equilibrium

Contents

Introduction

This thesis is devoted to the study of some models of optimal transport, alternative to the so-called Monge-Kantorovich one: as we will try to explain, they naturally arise in some real world applications, like in the design of optimal transportation networks or in traffic congestion modeling. To give the possibility of having a better understanding of the studies we are going to perform in this work, we should start briefly recalling the usual Monge-Kantorovich mass transfer problem, a topic which has received a lot of attention in the last years, contributing to the growth of new techniques which can be applied to various fields of mathematics (to have a clear picture, the reader should consult Villani's monumental book [**85**]). In the usual Monge-Kantorovich problem, one has to find the optimal way to move a given amount of mass from an initial configuration to a prescribed final one, in order to minimize a total cost of transportation. This total cost is obtained by summing the infinitesimal costs relative to each particle: assuming that the initial mass distribution occupies a region V and that its density is described by the function f_0, so that $\int_V f_0(x)\,dx =$ total mass, in each point x we have a quantity of mass given by $f_0(x)\,dx$ which have to be moved to a point $T(x)$. In the case that the infinitesimal cost is of the type *mass*$\times$*distance* (this is the work involved, up to a constant), we then have the total cost

$$\int_V |x - T(x)|\, f_0(x)\,dx,$$

to be minimized among all applications T that transport the initial configuration to the desired one: clearly, the latter requirement has to be rigorously expressed in a mathematical framework. For example, if the density of the final configuration is described by the function f_1, then this requirement can be expressed as

$$\int_A f_1(x)\,dx = \int_{T^{-1}(A)} f_0(x)\,dx, \quad \text{for every set } A \subset V,$$

and T is said to be a *transport map* between f_0 and f_1: obviously, there must result $\int_V f_1(x)\,dx = \int_V f_0(x)\,dx$. This is basically the problem as introduced by Monge in 1781: as we will see, from a mathematical point of view this is a very subtle problem (and thus very interesting!). It can be extended in various way: we can consider more in general a total cost of the type

$$\int_V |x - T(x)|^p\, f_0(x)\,dx,$$

and we can replace $f_0\,dx$ and $f_1\,dx$ with general positive measures ρ_0 and ρ_1, having the same total mass (conventionally, they will be probability measures). In this case, a transport map T is an application satisfying $\rho_1 = (T)_\sharp \rho_0$, that is ρ_1 is the image measure (or push-forward) of ρ_0 through the map T, and the total cost is

$$\int_V |x - T(x)|^p \, d\rho_0(x).$$

Anyway, in this latter case, if one keeps trying to realize the transportation with a map T, there could be some troubles (there are some evident difficulties in sending a single Dirac mass into the sum of two, with a map T...): in other words, the requirement that all the mass located at x must go to the same destination $T(x)$ could be too strong. To allow splitting of mass, new objects have to be introduced: the so-called *transport plans*. They are probability measures γ on the product space $V \times V$, such that $d\gamma(x, y)$ is the amount of mass located at x which has to be sent to y: in this way, the total cost now becomes

$$\int_{V \times V} |x - y|^p \, d\gamma(x, y).$$

The requirement of transporting ρ_0 to ρ_1 is expressed by requiring that $\gamma(A \times V) = \rho_0(A)$ and $\gamma(V \times B) = \rho_1(B)$, i.e. ρ_0 and ρ_1 are the *marginals* of γ: this whole procedure can be seen as a relaxation (in a suitable sense) of the original Monge's problem and it is due to Kantorovich (more than 150 years after Monge!). The reader should not be worried by this rather sloppy and imprecise presentation of the Monge-Kantorovich problem (we will provide more details in Chapter 1), what is important to stress here is that in this type of problems:

(i) the cost linearly depends on the mass;
(ii) the path followed by each mass particle only depend on the initial and final position and it is *not* an unknown: in other words, once an optimal transport plan γ tells you that some particle located at x has to go to y, then this mass moves along the geodesic segment $\overline{xy}$ (as far as the cost is linked to the Euclidean distance).

As we will see, these two facts are tightly related. In a very rough way, we could summarize the scope of this work as that of considering models of transportation where the cost for moving a mass m on a distance ℓ is no more given by $m\,\ell$, but a sublinear or superlinear dependence on the mass is imposed, for example considering $m^\alpha\,\ell$ $(0 < \alpha < 1)$ or $m^p\,\ell$ $(p > 1)$. Let us focus on the first case for a while: observe that with this choice of the parameter α, we get that $(m_1 + m_2)^\alpha < m_1^\alpha + m_2^\alpha$, so that the cost now keeps trace of the fact that "*it is better to keep mass together, in order to save cost*". In this way, the model considered is encoding the cost of the transportation structure in its formulation: in particular, the paths followed by particles do not depend anymore on the coupling *initial position—final position* only, but they also depend on the transportation itself and they are an unknown of the problem. Just to clarify the situation, suppose to have a small town, in which a power supply station has to furnish electricity to the houses of the town: you could see this as a transport problem between a big Dirac mass $M\delta_{x_0}$ (the station) and the sum of a certain number of small Dirac masses $\sum_{i=i}^n m_i\,\delta_{x_i}$, each corresponding to a house (see Figure 1). In this example, M stands for the quantity of energy that the station can produce, while the coefficients m_i can be seen as the amount of energy that every house needs, so that we assume that they satisfy $\sum_{i=1}^k m_i = M$ (clearly, we are disregarding a lot of physical effects, but this is not the point here). Monge-Kantorovich formulation, which looks only at the couplings (in this case, there is only one coupling), gives as optimizer a widespread system of wires, built as follows: the station should provide energy to every house directly, by using a wire for each of them. This would be very costly for the electric company; on the contrary, in a real situation the electric company should try to use the least amount of wires possible (the maintenance of the network should have a cost, probably proportional to its length), building an electric power supply network which minimizes a total cost of the form

$$\sum m^\alpha \ell,$$

under the constraints of Kirchhoff's law for circuits (i.e. electric charge does not disappear) and $\sum_{i=1}^k m_i = M$. Observe that the lower the parameter α, the stronger the branching effects: in particular, for $\alpha = 0$ this is just the so-called Steiner's problem.

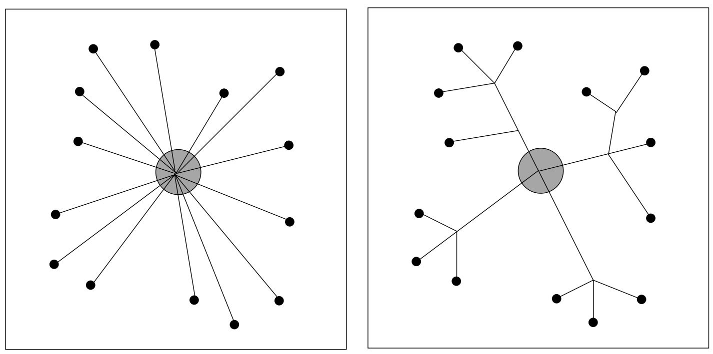

FIGURE 1. It is better to construct an optimal network of wires (right) to save cost: anyway, this is not possible by looking at the Monge-Kantorovich formulation, which rather describes the situation on the left.

In the second case, that is when the infinitesimal cost is of the form $m^p \ell$, with $p > 1$, we see that exactly the opposite phenomenon happens: $m_1^p + m_2^p < (m_1 + m_2)^p$, so it is better to spread the mass during the transport, in order to save cost. In other words, the cost takes into account that letting pass too much mass from a point, could cost a lot in terms of congestion. Let us try to give another simple example: suppose to consider a city with a football stadium located at its center x_0 and having a maximum capacity of M, so that we can think of it as the measure $M\delta_{x_0}$. On sunday, the whole population of the city wants to go to the stadium for the football match: supposing that the city is made of a certain number k of houses, each of them located at x_i and containing m_i inhabitants, we can think to represent the whole of them as $\sum_{i=1}^k m_i \delta_{x_i}$ (and certainly $M = \sum_{i=1}^k m_i$). Clearly, each citizen could decide to take the shortest road to the stadium (this is the Monge-Kantorovich solution): anyway, if everybody behaves like this, they risk to create a lot of traffic congestion, dramatically elevating the trip time. So maybe they would decide to distribute themselves on the whole transportation network, not only on shortest roads, taking into account that traffic congestion could compensate the difference of length.

In both cases, that is branched and congested transport, we will present and discuss various models (some recent, some older and others completely new), then one of the *leitmotiv* of the thesis will be that of *equivalences* between these models: indeed, the importance of having various

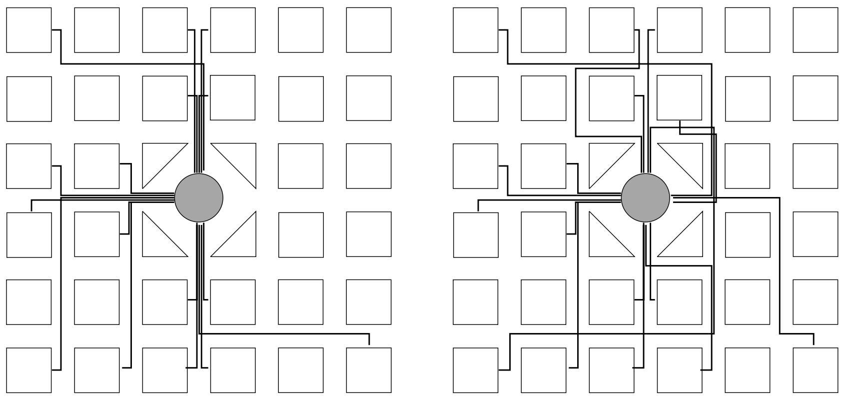

FIGURE 2. A lot of citizens go to stadium, traveling along the shortest paths (left)...but sometimes it is better to take a longer path, avoiding congestion effects (right)

equivalent formulations of the same problem is invaluable, as far as this could provide different points of view on the problem, allowing for a great flexibility. In this way, the more the points of view, the more the machinery you could use to attack the problem (and the more the chances to succeed!): something that seems hard to prove in a formulation, maybe is simpler in another. Moreover the proofs of these equivalences have their own mathematical interest, as far as a lot of interesting mathematical tools have to be exploited in order to achieve the goal: for example, as we will see, the equivalence between the models for congested transport here presented will require some non-trivial regularity results for the very degenerate elliptic equation[1]

$$-\text{div}\left((|\nabla u|-1)_+^{q-1}\frac{\nabla u}{|\nabla u|}\right)=f,$$

which appears to be quite unexplored.

We now leave the reader to the *Plan of the Work*, where a detailed summary of each chapter of the thesis is provided, as well as the principal results obtained and the main techniques used throughout the work.

[1] $(\cdot)_+$ stands for the positive part.

Plan of the work

Chapter 1. This chapter contains standard materials on Optimal Transportation, with particular emphasis on the basic properties of Wasserstein spaces and on the dynamical formulations for the Wasserstein distances. To make this introductory summary more readable, we briefly recall the definition of Wasserstein space and some alternative definitions for the related Wasserstein distance: for simplicity, let us take $\Omega \subset \mathbb{R}^N$ compact convex set, then the space of probability measures $\mathscr{P}(\Omega)$ over Ω can be turned into a metric space, introducing the $p-$Wasserstein distance ($p \geq 1$)

$$w_p(\rho_0, \rho_1) = \inf\left\{\int_{\Omega\times\Omega} |x-y|^p \, d\gamma(x,y) \, : \, \gamma \in \mathscr{P}(\Omega\times\Omega), \, (\pi_x)_\sharp\gamma = \rho_0, \, (\pi_y)_\sharp\gamma = \rho_1\right\}.$$

Indicating with $\mathcal{W}_p(\Omega)$ the space of probability measures equipped with this metric, it turns out that $\mathcal{W}_p(\Omega)$ is a length space and for every $\rho_0, \rho_1 \in \mathcal{W}_p(\Omega)$, their $p-$Wasserstein distance can also be written as

$$w_p(\rho_0, \rho_1) = \inf\left\{\int_0^1 |\mu_t'|_{w_p} \, dt \, : \, \mu_0 = \rho_0, \, \mu_1 = \rho_1\right\}, \tag{1}$$

where the infimum is taken over the set of $\mathcal{W}_p(\Omega)-$valued absolutely continuous curves and $|\mu_t'|_{w_p}$ is the *metric derivative* (see the Appendix A for the definition) of the curve of measures $t \mapsto \mu_t$, with respect to the metric w_p.

This is the first dynamical formulation of the Wasserstein distance: another interesting dynamical formulation, based on the continuity equation (which expresses the conservation of mass), is due to Benamou and Brenier ([**13**]) and reads as

$$w_p(\rho_0, \rho_1) = \inf\left\{\int_0^1 \int_\Omega |v_t(x)|^p \, d\mu_t(x) \, dt \, : \, \begin{array}{c} \partial_t\mu_t + \mathrm{div}_x(v_t\,\mu_t) = 0 \\ \mu_0 = \rho_0, \, \mu_1 = \rho_1 \end{array}\right\}.$$

The link between these two formulations is given by the fact that for every absolutely continuous curve μ over $\mathcal{W}_p(\Omega)$, we can always find an admissible vector field v such that $\|v_t\|_{L^p(\mu_t)} = |\mu_t'|_{w_p}$, for every t (we will give more details on this relation in Section 5), so that this v plays the role of a tangent vector to the curve μ.

Clearly this chapter is far from being an exhaustive presentation of the theme, but it is intended only as an account of the tools that we will need in the course of the work: for all the proofs, further results, details and bibliographical notes, the reader is strongly suggested to consult the two books of Villani [**86, 85**], the lecture notes [**3**] by Ambrosio and the book [**6**] as well, from which much of the materials here presented are taken.

Chapter 2. This chapter has a general character, as far as it deals with lower semicontinuity results for functionals of the type

$$\int_I f(t, \mu(t), |\mu'|(t))\, dt,$$

defined on spaces of absolutely continuous curves over a metric space X. Here $\mu : I \to X$ is an absolutely continuous curve on X, with $I \subset \mathbb{R}$ compact interval, and $|\mu'|(t)$ is again the metric derivative of μ at the point t. We will see that under appropriate growth conditions and (mild) regularity assumptions on the integrand f (and on the metric space X), most of the classical results valid in a Euclidean setting still hold in this general framework (Theorem 2.3.4). Existence of curves of minimal action for the associated variational problems (i.e. minimizing the energy, with given endpoints) are also provided (Theorem 2.4.3), in a setting where X is a metric space not necessarily locally compact: this latter fact prevents a straightforward application of Ascoli-Arzelá Theorem (Theorem 2.4.5).

The results of this chapter have been derived in view of applications to spaces of measures (as the Wasserstein ones, for example), which are particular and significant metric spaces. In particular, this context will be the occasion for introducing the concept of *evolution pairing*, which is, given a curve of measures μ, a way to distinguish between its moving part and its still one. This description is given in an elementary way, avoiding the use of the continuity equation (with some advantages and some drawbacks):

in particular, a couple of curves (ν, μ), with μ_t probability measure and ν_t positive measure with total variation less than or equal to 1, is said an evolution pairing if $\nu_t \leq \mu_t$ for every t (which expresses the fact that the moving mass is just a part of the whole) and $\vartheta_t := \mu_t - \nu_t$ is non-decreasing, that is $\vartheta_s \leq \vartheta_t$ as measures, for $s \leq t$. The latter requirement simply expresses the fact that the mass which is already arrived (that is $\mu_t - \nu_t$) has to grow in time. The main application of this concept will be in Chapter 4, in the context of branched transportation problems (this was the original motivation for its definition in [**B2**]): anyway, in this chapter we will consider quite general action functionals defined on evolution pairings, providing semicontinuity results for them and existence of minimal configurations connecting two given measures ρ_0 and ρ_1.

The whole chapter is taken from the published work [**B2**], with minor changes (in particular, a treatment of supremal functionals has been added).

Chapter 3. Here we start to focus on one of the main subjects of this work, namely *branched transportation problems*, which will occupy this and the next two Chapters. The typical problem one has to face in this context is the following: one has some mass ρ_0 that has to be transported to a destination ρ_1 and wants to find the optimal way to perfom this transportation. The main difference with the classical Monge-Kantorovich mass transportation problem is that optimality should regard the type of structure used to move the mass: in particular, this transportation should be optimal with respect to some energy which takes into account the fundamental principle that "*the more you transport mass together, the more efficient the transport is*". This is, for example, exactly what happens in many natural systems: root systems in a tree, bronchial systems and blood vessels in a human body and so on. Each of them solves the problem of transporting some "mass" (water, oxygen, blood or generic fluids) from a source to a destination, avoiding separation of masses as much as possible. This fundamental principle is translated into the energy by considering, for a mass m moving on a distance ℓ, a cost of the form $m^\alpha \ell$, with the parameter $\alpha \in [0, 1]$ modeling the branching effects: indeed, thanks to the subadditivity of the function $x \mapsto x^\alpha$, we see that it is less expensive to put masses together during the transportation, then giving rise to optimal tree-shaped configurations, i.e. the typical resulting structures are trees made of bifurcating vessels. In the sequel, we will refer to an energy of this type, that is (in an informal way)

$$\sum \text{mass}^\alpha \times \text{length}, \tag{2}$$

as a *Gilbert-Steiner energy.*

First of all, in Chapter 3 we review some of the main models for branched transportation existing in literature, in particular:

- the *transport paths model* by Xia (see [**90**]), in which the transportation structures are modeled as vector measures with prescribed divergence and supported on $1-$dimensional sets;
- what we have called *Lagrangian models*, due to different authors (Maddalena-Morel-Solimini, Bernot-Caselles-Morel, Bernot-Figalli), which slightly differ from one to another, but whose common root is a Lagrangian description of the transportation, through the employing of probability measures over the space of admissible paths;
- the *path functional model* introduced by Brancolini, Buttazzo and Santambrogio in [**24**].

A distinguished feature of the latter model, on which this and the next chapter are mainly focused, is its simplicity and its purely dynamic approach: indeed, it is obtained by perturbing the geodesic formulation for Wasserstein distances (1). This means that the energy to be minimized is defined on Lipschitz curves of probability measures and it is given by a weighted-length functional in the Wasserstein space $\mathcal{W}_p$, with the weight function encouraging aggregation of masses, so that optimal tree-shaped configurations are expected.

More precisely, for the case to study, the choice of the weight function is given by the local and lower semicontinuous functional (here $0 < \alpha < 1$) defined on measures (see [**21**])

$$g_\alpha(\mu) = \begin{cases} \displaystyle\sum_{k\in\mathbb{N}} m_k^\alpha, & \text{if } \mu = \displaystyle\sum_{k\in\mathbb{N}} m_k \delta_{y_k}, \\ +\infty, & \text{otherwise,} \end{cases}$$

and the energy under consideration in [**24**] is

$$\mathcal{P}_{\alpha,p}(\mu) = \int_0^1 g_\alpha(\mu_t)\, |\mu_t'|_{w_p}\, dt,$$

for every Lipschitz curve μ with values in $\mathcal{W}_p(\Omega)$. Then, for every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, a curve μ_t connecting them and minimizing $\mathcal{P}_{\alpha,p}$ provides the optimal structure of branched transportation, as well as the dynamical evolution in time of this transportation.

This model has several drawbacks and in particular it does not correspond, generally speaking, to a Gilbert-Steiner energy: for example, one immediately sees that in the energy $\mathcal{P}_{\alpha,p}$ there is no way to distinguish between moving and still masses (the coefficient g_α is a function of the whole mass configuration at time t, i.e. μ_t), while the metric derivative term does not scale in the appropriate way, that is like a length, but rather like a mass-weighted sum of lengths, so that $\mathcal{P}_{\alpha,p}$ is not exactly of the form (2).

To sum up, the path functional model gives rise to a different energy and to a different model for branched transport, not comparable with the other ones: we will provide some studies on this energy (Section 4), explain the difference with a Gilbert-Steiner energy and suggest how to modify it in order to achieve an equivalent path functional formulation of branched transportation (Section 5), the latter being the content of the next chapter.

Chapter 4. This is based on a joint work with Filippo Santambrogio ([**25**]) and it answers a question raised in Santambrogio's PhD Thesis (see [**76**, Remark 6.2.7]), about the connection between the path functional model and the Lagrangian models. In particular, it is shown that modifying the former

- tuning the Wasserstein exponent p in the energy $\mathcal{P}_{\alpha,p}$ to be ∞ (this settles the unnatural scaling of the energy, as far as the $\infty-$Wasserstein distance corresponds to a supremal mass transportation problem which does not depend on the mass, but only on the maximal displacement);
- introducing a suitable way to distinguish between moving masses and still masses, in such a way that the path functional energy only keeps track of the moving part of the curve μ_t;

makes possible to recover a Gilbert-Steiner energy also via path functional model, thus giving equivalence with the Lagrangian models (Theorem 4.5.1). In doing this, the fundamental tool is the concept of *evolution pairing* introduced in Chapter 2 and the energy to be considered will be

$$\mathfrak{L}_\alpha = \int_0^1 g_\alpha(\nu_t)\,|\mu_t'|_{w_\infty}\,dt,$$

where now the coefficient g_α takes into account only the moving masses, given by the curve ν_t. However, we will see that some troubles still occur, because on the one hand the modified energy has not enough coercivity properties and the time interval $[0,1]$ has to be replaced with $[0,\infty)$, in connection with a $1-$Lipschitz requirement on the curve μ_t, allowing for possibly infinite paths of transportation; on the other hand, the class of evolution pairings enlarges too much the space of admissible configurations and a narrower class is actually required (the so-called *special evolution pairings*, see Definition 4.3.4). As a byproduct, this leads to some non trivial studies about Lipschitz curves in the Wasserstein space $\mathcal{W}_\infty$, a topic which has received little attention up to now and thus it can be of its own interest.

Chapter 5. In this chapter a further model for branched transportation is introduced, again based on an energy functional defined on curves of measures, like the path functional model was. The main novelty of this model is that it can be seen as the natural extension of the Benamou-Brenier approach to the branched setting, thus providing the genuine Eulerian counterpart to the Lagrangian models presented in Chapter 3: more precisely, in this model admissible configurations are couples (μ,ϕ) solving the continuity equation (in a distributional sense)

$$\partial_t\mu_t + \mathrm{div}_x(\phi_t) = 0,$$

and such that μ connects two given measures ρ_0 and ρ_1. Then an energy of the kind (again $0<\alpha<1$, with the same functional g_α as before)

$$\mathcal{G}_\alpha(\mu,\phi) = \int_0^1 g_\alpha\left(\left|\frac{d\phi_t}{d\mu_t}\right|^{1/\alpha}\mu_t\right)\,dt$$

is considered. Observe that the finiteness of the energy implies that $\phi_t \ll \mu_t$ for $\mathscr{L}^1-$a.e. $t \in [0,1]$ and that ϕ_t (not necessarily μ_t) is purely atomic, so that in this case (v stands for the Radon-Nikodym derivative of ϕ with respect to μ) the previous energy takes the form

$$\mathcal{G}_\alpha(\mu,\phi) = \int_0^1 \left[\sum_{i\in\mathbb{N}} |v_{t,i}|\, \mu_{t,i}^\alpha \right] dt,$$

with $|v_{t,i}|\, \mu_{t,i}$ standing for the masses of the atoms of $|\phi_t|$. Again, the link with a Gilbert-Steiner energy is quite evident, interpreting v has the velocity vector fields of the particles. For this model, we will prove (Theorem 5.2.6) the existence of a minimal configuration (μ,ϕ): the method of proof goes along the way of the Direct Methods of the Calculus of Variations, so that the main efforts will be devoted to prove that our variational problem has enough coercivity and semicontinuity properties, with respect to a suitably chosen topology. Moreover equivalence with Lagrangian models (and then with Xia's model and the modified path functional one discussed in Chapter 4) is shown in Theorem 5.4.2, as well as some interesting (though already known) estimates between *branched distances* and the Wasserstein ones, first proven by Devillanova and Solimini ([**43**]) and Morel and Santambrogio ([**70**]): thanks to the tools developed in this chapter, we can provide simpler proofs of these estimates.

The materials of this chapter are taken from the recent paper [**B3**] in collaboration with Giuseppe Buttazzo and Filippo Santambrogio.

Chapter 6. With this chapter (which is a review one) we completely change point of view and we start looking at transportation problems which are, roughly speaking, complementary to the branched ones: we deal with optimal transportation in presence of congestion effects, so that now our total cost for transporting ρ_0 to ρ_1 will take into account the fact that "*the more the mass travels together, the higher the cost is*", giving rise to very spread and as diffused as possible optimal configurations. Here the role of the function $x \mapsto x^\alpha$ will be played by the function $x \mapsto x^p$, with $p > 1$, which is a superadditive one, so that now $m_1^p + m_2^p < (m_1 + m_2)^p$ and concentrations of mass are thus penalized.

This is for example what happens in an urban traffic situation: in this context, we introduce two models for congested dynamics, both having interesting connections with equilibrium issues. The first is the model introduced by Wardrop in a discrete setting ([**88**]) and then recently generalized to the continuous case by Carlier, Jimenez and Santambrogio in [**33**]:

two probability measures ρ_0 and ρ_1 over Ω are given (standing, for example, for centers of production and consumption of a given commodity or for residential and working/commercial areas); the goal is to describe the possible way in which commuters between ρ_0 and ρ_1 choose to distribute on the admissible roads. Probability measures Q over the set of Lipschitz paths (parametrized over $[0,1]$, for example) are introduced, verifying $(e_i)_\sharp Q = \rho_i$, for $i = 0,1$, the function e_t being the application that to every curve σ assigns its position at time t, i.e. $e_t(\sigma) = \sigma(t)$.

In this model, each traffic assignment Q gives rise to a resulting *traffic intensity* i_Q, which is a positive measure over the city Ω (actually, a refinement of the concept of transport density for Monge's problem), representing how much each subregion is congestioned. Then, given a density cost function H, such that $H(i_Q(x))$ stands for the total cost (per unit of volume) of passing from a point x where there is an amount of traffic given by $i_Q(x)$, we obtain an overall transportation cost simply integrating in space, i.e.

$$W(Q) = \int_\Omega H(i_Q(x))\,dx.$$

In order to modelize congestion effects, H typically behaves like a strictly convex power function, for example $H(t) \simeq t^p$, with $p > 1$. Then one looks at existence of configurations Q minimizing W: this is the overall optimization point of view, in which the individual welfare of the commuters between ρ_0 and ρ_1 is disregarded. The main interest of the resulting optimality conditions for W (which are both necessary and sufficient, as far as the problem is convex) is that they are connected with a concept of equilibrium, the so-called *Wardrop equilibrium*: indeed, one can prove that an optimal Q determines a new metric on Ω, which depends on the traffic (i.e. on Q) itself, given by

$$d_Q(x,y) = \inf\left\{\int_0^1 H'(i_Q(\sigma(t)))\,|\sigma'(t)|\,dt\,:\,\sigma(0)=x,\,\sigma(1)=y\right\},$$

and such that Q gives full mass to the set of curves of minimal length for this congested metric. This is exactly the mathematical translation of Wardrop's postulate "*at equilibrium, every actually used road must be of minimal length, taking into account congestion effects*", so that no single commuter has the interest to change his road, provided the others keep their strategy.

It is important to observe that the previous expression d_Q suggests that the cost function H should satisfy $H'(0) > 0$, which rules out a cost function of the type $H(t) = c\,t^p$: roughly speaking, passing from a desert road must have a cost (just for fuel or tires consumption, for example) from an individual point of view. This is the reason why, from Chapter 7 on, we will mainly concentrate on the model case of a cost function of the form

$$H(t) = \frac{1}{p}t^p + t,\ \ t \geq 0.$$

We also observe that in some applications, it is interesting to consider the optimization problem for W in which the coupling $(e_0, e_1)_\sharp Q$ (i.e. a transport plan between ρ_0 and ρ_1) is fixed, that is we can consider what is sometimes called the *who goes where problem*: in other words, for every couple (x, y) we are prescribing the quantity of commuters between x and y, rather than optimizing over the whole set of admissible couplings. Again, optimality conditions for this optimization problem are equivalent to the existence of a Wardrop equilibrium.

The second model we will take into account is the so-called *continuous model of transportation* introduced by Beckmann in [**11**], a particular case of which is the divergence-constrained optimization of the total variation of vector measures, which is another dual formulation of Monge's problem (see Chapter 1, Proposition 1.1.10). This can be seen as the natural counterpart of Xia's model, in particular it gives a static description of the transportation: in Beckmann's own words "*the analysis pertains to a static economy, a single transportation system and fixed production programs for a given commodity [...]* ". In this model, transportation activities are described by vector fields ϕ such that $\mathrm{div}\phi = \rho_0 - \rho_1$ and satisfying a homogeneous Neumann boundary condition (these are just balance conditions), so that $|\phi(x)|$ is the amount of mass passing from x and $\phi(x)/|\phi(x)|$ is the direction of transportation. Then a total cost is considered

$$B(\phi) = \int_\Omega \mathcal{H}(\phi(x))\,dx,$$

with $\mathcal{H}$ typically being a smooth, strictly convex and superlinear function, this last two requirements encoding the congestion effects. One can write down optimality conditions for the problem of minimizing B under the constraint on the divergence and a potential naturally appears, as a Lagrange multiplier for the constraint $\text{div}\phi = \rho_0 - \rho_1$: indeed, for an optimal ϕ we have

$$\phi = \nabla\mathcal{H}^*(\nabla u), \tag{3}$$

where $\mathcal{H}^*$ is the Legendre-Fenchel conjugate of $\mathcal{H}$. We call this function u a *Beckmann potential*: it is easily seen that u acts like a Kantorovich potential for Monge's problem, in the sense that integral curves of its gradient gives the directions of optimal transportation. Also it is a $1-$Lipschitz function, but now with respect to the traffic-dependent metric given by

$$d_u(x,y) = \inf\left\{\int_\Omega |\nabla u(\sigma(t))|\,|\sigma'(t)|\,dt\,:\,\sigma(0)=x,\,\sigma(1)=y\right\}.$$

Moreover as a consequence of the constraints on ϕ, a Beckmann potential u solves the Neumann boundary value problem

$$\begin{cases} -\text{div}\nabla\mathcal{H}^*(\nabla u) &= \rho_0-\rho_1, & \text{in } \Omega, \\ \langle\nabla\mathcal{H}^*(\nabla u),\nu\rangle &= 0, & \text{on } \partial\Omega. \end{cases}$$

Chapter 7. In the case of an isotropic cost function $\mathcal{H}$, i.e. when $\mathcal{H}(z) = H(|z|)$, it is possible to relate Beckmann's problem to Wardrop's one: this is the content of the present chapter. We will see that in general there holds

$$\min B(\phi) \le \min W(Q),$$

and that for every Q, it is possible to construct an admissible ϕ such that $|\phi| \le i_Q$. Then we will obtain the equivalence showing that, on the other hand, for an optimal ϕ we can construct a traffic assignment Q such that

$$|\phi(x)| = i_Q(x).$$

This is the difficult point in the equivalence proof: the main tool to achieve such a construction is a deformation argument which dates back to Moser ([**71**]). Suppose that we are given two measures (let us say absolutely continuous w.r.t. $\mathscr{L}^N$ and with smooth positive densities bounded from below) $\rho_0 = f_0\cdot\mathscr{L}^N$ and $\rho_1 = f_1\cdot\mathscr{L}^N$ and we want to find a change of variable T transforming the first measure into the second, that is such that

$$f_1(T(x))\det\nabla T(x) = f_0(x).$$

One could think to obtain such a construction starting from ρ_0 and continuously deforming it into ρ_1, according to the flow map $X(t,x)$ of a suitable velocity vector field v_t, thus obtaining a one-parameter family of intermediate interpolating measures $\mu_t = X(t,\cdot)_\sharp\rho_0$ between ρ_0 and ρ_1 and such that

$$\partial_t\mu_t + \text{div}_x(v_t\mu_t) = 0.$$

The flow map at time 1 has the required property, i.e. $X(1,\cdot)_\sharp\rho_0 = \rho_1$. In our case, if ϕ is optimal in Beckmann's problem, we see that the velocity field defined as

$$v_t(x) = \frac{\phi}{(1-t)f_0+tf_1},$$

permits to achieve such a construction: then, taking Q of the form $Q = \int \delta_{X(\cdot,x)}\, d\rho_0$, we see that

$$i_Q(x) = |\phi(x)|,$$

as required. In order to give a proper meaning to this construction, regularity results on the optimal ϕ are needed, as well as a suitable notion of flow, which could differ from the usual (let us say Cauchy-Lipschitz) one, when v_t is not regular enough.

In this sense, one of the main results we prove (Theorem 7.5.5) is the equivalence in the sense of DiPerna-Lions of the two problems, that is we consider the flow of v_t in the DiPerna-Lions sense ([**46**]): this requires Sobolev and L^∞ regularity results on the optimal ϕ, that will be derived in Chapters 8 and 9.

Observe that the fact that the two problems are equivalent permits for example to show existence of Wardrop equilibria, just by looking at Beckmann's problem, which is a convex optimization problem, much easier to handle than Wardrop's one: in addition, once the equivalence is established, a Wardrop equilibrium can be explicitly constructed as a probability measure supported on integral curves of the vector field v_t above, i.e. curves parallel to ϕ.

Moreover the equilibrium metric d_Q can be computed in terms of a Beckmann's potential u, as far as using optimality conditions (3) we have

$$H'(i_Q(x)) = H'(|\phi(x)|) = |\nabla u(x)|,$$

so that

$$d_Q(x,y) = \inf\left\{\int_\Omega |\nabla u(\sigma(t))|\,|\sigma'(t)|\,dt\ :\ \sigma(0)=x,\ \sigma(1)=y\right\}.$$

This chapter is based on the joint work [**B4**] with Guillaume Carlier and Filippo Santambrogio.

Chapter 8. A Sobolev regularity result for the optimizer of Beckmann's problem is provided in this chapter, in order to complete the proof of the equivalence given in Chapter 7. In particular it is shown that, when the density cost function is

$$\mathcal{H}(z) = \frac{1}{p}|z|^p + |z|,\ \ z \in \mathbb{R}^N,$$

then the optimal vector field ϕ is in a certain Sobolev space $W^{1,r}_{\mathrm{loc}}(\Omega)$, provided that $\rho_0, \rho_1 \in W^{1,p}(\Omega)$ (the precise relation between r and p is given in Corollary 8.2.4). Moreover under suitable regularity assumptions on the boundary of Ω (a, probably not optimal, $C^{3,1}$ assumption is needed) and slightly enforcing the summability of the data, i.e. $\rho_0, \rho_1 \in W^{1,p}(\Omega) \cap L^{N+\alpha}(\Omega)$, with $\alpha > 0$, then the previous Sobolev property of ϕ_0 becomes global (Theorem 8.3.1 and Corollary 8.3.2). These results are achieved using the optimality conditions, that is

$$\phi = \nabla\mathcal{H}^*(\nabla u) = (|\nabla u| - 1)_+^{q-1}\frac{\nabla u}{|\nabla u|},$$

so that the Beckmann potential u is a weak solution of the very degenerate elliptic equation

$$\operatorname{div}\left((|\nabla u| - 1)_+^{q-1}\frac{\nabla u}{|\nabla u|}\right) = \rho_0 - \rho_1, \tag{4}$$

where $q = p/(p-1)$. The proof is based on difference quotients: this is quite classical (for linear elliptic equations, this is the so-called *Nirenberg method*, see also [**58**]), however some non trivial monotonicity properties peculiar of the operator $\nabla\mathcal{H}^*$ have to be exploited in order to let the proof work. Moreover we observe that this Sobolev regularity result on ϕ_0 is achieved directly, without passing through a higher differentiability result for u (this latter property is probably false): in other words, we are not claiming that $u \in W^{1,q}(\Omega)$ solving (4) possesses second order derivatives, but just that some non-linear function of the gradient ∇u is in a suitable Sobolev space.

In order to prove global regularity, some care at the boundary is needed: here we follow an idea contained in [**35**], up to some significant modifications (in particular, we derive the global Sobolev regularity using the boundedness of the gradient of the solution, whose proof is in Chapter 9). The starting point is classical: one takes a neighborhood in Ω of a boundary point and deforms it (through a diffeomorphism) into a half-ball B^+, correspondingly obtaining a degenerate elliptic equation with variable coefficients in B^+. The idea is then to extend this equation to the whole ball $B = B^+ \cup B^-$, just by reflection: in this way, the required boundary Sobolev estimates immediately translate into interior Sobolev ones, for a variable coefficients equation. Clearly, one has to guarantee that the variable coefficients operator obtained with this construction enjoys some nice regularity properties with respect to the x variable: in order to do this, the initial diffeomorphism has to be suitably chosen and here the somehow heavy request of $\partial\Omega$ being $C^{3,1}$ comes into play.

Observe that the key point, for which it is necessary to reflect the equation extending it to the whole ball, is that we are proving Sobolev regularity for $\nabla\mathcal{H}^*(\nabla u)$ and not for ∇u: this to say that now it is no more useful to work in the half-ball B^+ (this is the classical strategy in non degenerate cases), proving that in the $N-1$ directions which are tangential to the flat part of ∂B^+ the vector field $\nabla\mathcal{H}^*$ is weakly differentiable and then using the equation to recover regularity in the missing direction.

We remark that the regularity results contained in this chapter and in the next one are not trivial and have their own interest, apart from Wardrop equilibria and traffic congestion modeling, as regularity results for a class of very degenerate elliptic equations, whose model example is given by (4) and which deserves more attention.

Also the contents of this chapter are based on the paper [**B4**]: we point out that a first version of this work actually contained some crucial errors in the proof of the Sobolev estimate near the boundary, as well as in the proof of the L^∞ gradient estimate (see next chapter).

Chapter 9. Finally, we show in this chapter that Beckmann's potentials, again in the case of

$$\mathcal{H}(z) = \frac{1}{p}|z|^p + |z|,\ z \in \mathbb{R}^N,$$

are Lipschitz functions: as a byproduct, we obtain that Beckmann's optimizer is an L^∞ vector field, which permits to give a rigorous justification to the proof of the equivalence in the DiPerna-Lions sense. More precisely, we show (Theorem 9.2.1) that taking Ω with a $C^{2,1}$ boundary and $\rho_0, \rho_1 \in L^{N+\alpha}(\Omega)$, with $\alpha > 0$, then every $W^{1,q}$ weak solution of (4) with homogeneous Neumann boundary conditions has a bounded gradient. The proof is based on an approximation procedure which aims to derive a priori L^∞ estimates on the gradient, not depending on the approximation. To be more precise, we introduce the approximating equations

$$\mathrm{div}\nabla\mathcal{H}^*_\varepsilon(\nabla u) = \rho_0^\varepsilon - \rho_1^\varepsilon,$$

which are uniformly elliptic, with ellipticity constants degenerating when ε goes to 0 and we approach the original problem: regularity results are well-known for these equations, in particular they can be differentiated and a linearized equation for the gradient can be derived.

It is then sufficient to choose this approximation in such a way that outside a large (but fixed and independent of ε) ball B, these equations have ellipticity constants independent of ε. The main trick will then be selecting a particular test function and insert it in the weak formulations of the differentiated approximating equations: this will cut away the ball B on which stability of the estimates is not guaranteed. Observe that this procedure formalizes the intuitive idea that, roughly speaking, for large values of the gradient, the equation is uniformly elliptic and so everything goes well, while for small values of $|\nabla u|$ ellipticity breaks down, but in any case the size of the gradient is controlled.

In deriving our L^∞ estimate on the gradient, we will use three different techniques: reverse Hölder inequalities (the so-called Gehring Lemma, [**54**]); Moser's iteration technique ([**72**]) and De Giorgi's truncation argument ([**57**, Chapter 7]); these tools permit to recursively achieve a gain of integrability on the gradient, with estimates independent of ε. The corner-stones of the result are, as always in Elliptic Regularity Theory, the possibility to derive Caccioppoli-type inequalities, as well as the choice of the right test function we mentioned before. The whole proof has benefited of a careful reading of the papers [**45**] and [**63**] by Di Benedetto and Lewis, respectively, dealing with regularity results for the $p-$Laplace operator, i.e.

$$-\Delta_p u := \mathrm{div}(|\nabla u|^{p-2}\nabla u),$$

which is degenerate at critical points of the solution.

The $C^{2,1}$ regularity requirement on $\partial\Omega$ is probably not optimal, but it permits to simplify a bit the proof: indeed, in order to give estimates near the boundary, we perform the same trick as in Chapter 8, reducing ourselves to consider local L^∞ gradient estimates for a variable coefficients equation, using a suitable diffeomorphism and a reflection. If one wants to get rid of this assumption on $\partial\Omega$, the techniques of Lieberman in [**64**] seem to be adaptable also in this case: here we have decided to take a shortcut, not to overburden the reader with unnecessary (at least for the aims of this work) technicalities.

We finally observe that Lipschitz regularity is the best one can hope for solutions of (4), since every $1-$Lipschitz function is a solution of the associated homogeneous equation. Clearly, the fact that $\nabla\mathcal{H}^* \equiv 0$ on $\{z\,:\,|z|<1\}$ does not permit to measure the oscillations of the gradient ∇u of the solution, but just its size: this is the big difference with the case of $p-$Laplacian type operators, for which Hölder continuity of ∇u can be proven.

The proof presented in this chapter (taken from [**B1**]) differs from both that given in the first version of [**B4**] (which was not completely correct, as said) and that contained in the actual version of [**B4**]: in the latter, we followed a shorter strategy based on some regularity results by Fonseca, Fusco and Marcellini contained in [**52**]. Anyway, this shorter proof asked for an unnecessary Hölder regularity on $\rho_0 - \rho_1$: as we will see, this is not restrictive for our purposes in congested transport (indeed, the equivalence statement of Theorem 7.5.5 requires ρ_0 and ρ_1 to be Lipschitz functions), but it is quite clear that a complete proof of the L^∞ gradient estimate for solutions of (4), under sharp assumptions on the data, can be of interest.

Warnings to the reader and main notations

In the course of this work, we will frequently use the *Disintegration Theorem*, a fundamental result in Measure Theory for which the reader is referred to [**38**, Chapter III]. From time to time we will also use *Markov's inequality* (sometimes also called Chebyshev's inequality) which we recall here for convenience: if μ is a measure on the space X and $f : X \to \mathbb{R}$ is $\mu-$measurable, then

$$\mu(\{x \in X \,:\, |f(x)| > M\}) \leq \frac{1}{M} \int_X |f(x)|\, d\mu(x).$$

When speaking of a sequence of probability measures $\{\mu^n\}_{n\in\mathbb{N}} \subset \mathscr{P}(X)$, we will use the term *equi-tightness*[2] to mean the following: for every $\varepsilon > 0$, there exists a compact subset $K_\varepsilon \subset X$ such that

$$\sup_{n\in\mathbb{N}} \mu^n(X \setminus K_\varepsilon) < \varepsilon.$$

When X is a Polish space (i.e. complete and separable metric space), a sequence $\{\mu^n\}_{n\in\mathbb{N}} \subset \mathscr{P}(X)$ is relatively compact if and only if it is equi-tight: this is *Prokhorov's Theorem* (see [**38**, Chapter III]). Here compactness refers to the topology induced by *narrow convergence*, that we indicate with the symbol $\rightharpoonup$, so that $\mu^n \rightharpoonup \mu$ means

$$\lim_{n\to\infty} \int_X \varphi(x)\, d\mu^n(x) = \int_X \varphi(x)\, d\mu(x), \text{ for every } \varphi \in C_b(X).$$

In considering curves $\sigma : I \to X$, we will use two distinct notations: when X is a generic metric space or a subset of $\mathbb{R}^N$, the notation $\sigma(t)$ is adopted, while in the case X is a space of measures, it will be more convenient to write σ_t.

In Chapters 8 and 9, devoted to regularity results for an elliptic equation, we will systematically use the convention of indicating with C or c a generic constant, which may differ from line to line, without keeping track of its precise value, but only of its dependence from the data of the problem. Moreover we will frequently commit the small abuse of using Young's inequality in the (slightly incorrect) form

$$a \cdot b \leq \varepsilon\, a^p + \frac{1}{\varepsilon}\, b^{\frac{p}{p-1}},$$

for every $\varepsilon > 0$, just for ease of computations and we will call this $\varepsilon-$Young's inequality.

Listed below, there are some basic notations used throughout this work:

[2] The terminology does not seem to be standard, in many textbooks usually this is simply called *tightness*.

$\mathscr{L}^k$	$k-$dimensional Lebesgue measure
$\mathscr{H}^k$	$k-$dimensional Hausdorff measure
$(f)_\sharp\mu$	push-forward of the measure μ through the map f, i.e. $(f)_\sharp\mu(A) := \mu(f^{-1}(A))$
$\dfrac{d\mu}{dm}$	Radon-Nikodym derivative of μ w.r.t. m
$\mu \ll m$	the measure μ is absolutely continuous w.r.t. m
$\mu\perp m$	mutually singular measures
$\mu = f\cdot m$	measure absolutely continuous w.r.t. m, with $\dfrac{d\mu}{dm} = f$
$\mathscr{P}(X)$	space of Borel probability measures over X
$C(X)$	continuous functions on X
$C_b(X)$	continuous and bounded functions on X
$C_c(X)$	compactly supported continuous functions
$C_o(X)$	continuous functions vanishing at infinity, i.e. for every $\varepsilon > 0$, $\exists K_\varepsilon \subset X$ compact set s.t. $\sup_{X\setminus K_\varepsilon}\lvert\varphi\rvert < \varepsilon$
$A \Subset B$	A has compact closure in B
1_A	characteristic function of A, i.e. $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ otherwise
$\mathfrak{I}_A$	indicator function of A, i.e. $\mathfrak{I}_A(x) = 0$ if $x \in A$ and $\mathfrak{I}_A(x) = +\infty$ otherwise
$\mathbb{S}^{N-1}$	unit sphere in $\mathbb{R}^N$
$\delta_{h,\omega} f$	incremental ratio in the direction $\omega \in \mathbb{S}^{N-1}$, i.e. $\frac{f(x+h\omega)-f(x)}{h}$
$L^p(X)$	$p-$summable Lebesgue measurable functions
$W^{1,p}(X)$	standard Sobolev space of L^p functions, having L^p distributional gradient
$W^{1,p}_\diamond(X)$	elements of $W^{1,p}(X)$ having zero mean
$\langle v, w\rangle$	standard Euclidean scalar product between v and w
$\langle\cdot,\cdot\rangle_X$	duality product between the Banach space X and its dual X^*
F^*	Legendre-Fenchel conjugate function, i.e. $F^*(x^*) = \sup_{x\in X}\langle x, x^*\rangle_X - F(x)$, $x^* \in X^*$
div	divergence operator, if $\phi : \mathbb{R}^N \to \mathbb{R}^N$ then $\mathrm{div}\phi = \displaystyle\sum_{i=1}^N \frac{\partial \phi_i}{\partial x_i}$
∇	gradient operator, if $u : \mathbb{R}^N \to \mathbb{R}$ then $\nabla u = \left(\dfrac{\partial u}{\partial x_1}, \dots, \dfrac{\partial u}{\partial x_N}\right)$
Δ	Laplace operator, if $u : \mathbb{R}^N \to \mathbb{R}$ then $\Delta u = \displaystyle\sum_{i=1}^N \frac{\partial^2 u}{\partial x_i^2}$
Δ_p	$p-$Laplace operator, if $u : \mathbb{R}^N \to \mathbb{R}$ then $\Delta_p u = \displaystyle\sum_{i=1}^N \frac{\partial}{\partial x_i}\left(\lvert\nabla u\rvert^{p-2}\frac{\partial u}{\partial x_i}\right)$

CHAPTER 1

Basic tools on Optimal Transportation

1. A brief survey on optimal transport

1.1. An old problem by Gaspard Monge. In its simplest form, Monge's transportation problem reads as follows: over an open subset $\Omega \subset \mathbb{R}^N$ (let us furtherly suppose that it is bounded, for the moment) there are given two mass distributions, represented by the two measures ρ_0 and ρ_1, such that

$$\int_\Omega d\rho_0(x) = \int_\Omega d\rho_1(x).$$

For example, with no loss of generality we can think of these as being probability measures: then we want to transport the amount of mass ρ_0 to the location prescribed by ρ_1. This is accomplished through a transport map T, that is an application $T : \Omega \to \Omega$ such that

$$\int_A d\rho_0(x) = \int_{T^{-1}(A)} d\rho_1(y),$$

for every subset $A \subset \Omega$, a condition which can also be rephrased, in a measure-theoretic framework, by saying that ρ_1 has to be the image measure of ρ_0 through the map T, i.e. $\rho_1 = (T)_\# \rho_0$. Supposing that the cost of moving each particle located at x to its destination $T(x)$ is simply given by the the distance $|x - T(x)|$, then integrating this infinitesimal cost with respect to ρ_0 gives the total cost, that is

$$\int_X |x - T(x)|\, d\rho_0(x).$$

Monge's problem can now be formulated as follows

$$(\mathcal{M}) \qquad \min \left\{ \int_\Omega |x - T(x)|\, d\rho_0(x) \, : \, (T)_\# \rho_0 = \rho_1 \right\},$$

that is *find the least expensive way to transport* ρ_0 *to* ρ_1. Clearly, this problem can be generalized in various way, for example considering more general costs other than the Euclidean distance, more general spaces other than bounded subsets of $\mathbb{R}^N$ and so on, but before doing this, one has to observe that Monge's problem is severely ill-posed: this is the reason why, despite having been formulated in 1781, it has received a first complete solution, under appropriate assumptions on the measures ρ_0 and ρ_1, only in 1999 with the impressive PDE-based proof of Evans and Gangbo, contained in [**48**].

To see the ill-posedness of Monge's problem, it is enough to observe that when $\rho_0 = \delta_{x_0}$ and ρ_1 is any measure which is not a single Dirac mass, then the set of transport maps is empty: on the other hand, working only with measures having smooth densities with respect to $\mathscr{L}^N$ does not change too much the situation. Indeed, if $\rho_i = f_i \cdot \mathscr{L}^N$, for $i = 0, 1$, with f_0 and f_1 smooth densities, then by means of the area formula the condition of being a transport map between ρ_0 and ρ_1 is equivalent to the requirement

$$f_1(T(x))|\det \nabla T(x)| = f_0(x),$$

a condition which can not be easily handled, also due to its highly non-linear character. Moreover the class of transport maps between two given measures does not enjoy closedness properties with respect to any reasonable topology, so that it is a hard challenge to look directly at problem $(\mathcal{M})$.

1.2. Kantorovich relaxed formulation. On the other hand, one could observe that to every transport map T it is possible to associate a probability measure γ_T on the product space $\Omega \times \Omega$, that is $\gamma_T = (\mathrm{Id} \times T)_\sharp \rho_0$ which is concentrated on the set graph$(T) = \{(x, y) : y = T(x)\}$ and such that

$$\int_{\Omega\times\Omega} \varphi(x, y)\, d\gamma_T(x, y) = \int_\Omega \varphi(x, T(x))\, d\rho_0(x), \quad \text{for every } \varphi.$$

Observe that such a γ_T has *marginal measures* given by ρ_0 and ρ_1, i.e. $(\pi_x)_\sharp \gamma_T = \rho_0$ and $(\pi_y)_\sharp \gamma_T = \rho_1$, where $\pi_x, \pi_y : \Omega \times \Omega \to \Omega$ are the projections on the first and second component, respectively. In terms of γ_T, the total cost of the map T can be rewritten as

$$\int_{\Omega\times\Omega} |x - y|\, d\gamma_T(x, y),$$

consequently one can think to relax $(\mathcal{M})$, by looking at the problem

$$(\mathcal{K}) \qquad \inf \left\{ \int_{\Omega\times\Omega} |x - y|\, d\gamma(x, y) \,:\, (\pi_x)_\sharp \gamma = \rho_0,\, (\pi_y)_\sharp \gamma = \rho_1 \right\}.$$

Clearly we have

$$(\mathcal{M}) \geq (\mathcal{K}),$$

and we have equivalence between the two problems each time one can prove that there exists an optimal $\gamma \in \Pi(\rho_0, \rho_1)$ concentrated on the graph of a map.

The relaxed formulation $(\mathcal{K})$, introduced by Kantorovich in [**61**], has several advantages: for example, the set

$$\Pi(\rho_0, \rho_1) = \{\gamma \in \mathscr{P}(\Omega \times \Omega) \,:\, (\pi_x)_\sharp \gamma = \rho_0,\, (\pi_y)_\sharp \gamma = \rho_1\},$$

of admissible measures is always not empty and this is the first big difference with Monge's formulation, indeed we have $\rho_0 \otimes \rho_1 \in \Pi(\rho_0, \rho_1)$. More important, $(\mathcal{K})$ is just a linear programming problem, as far as the functional

$$\gamma \mapsto \int_{\Omega\times\Omega} |x - y|\, d\gamma(x, y),$$

is linear, as well as the constraint $\gamma \in \Pi(\rho_0, \rho_1)$. In addition, the set $\Pi(\rho_0, \rho_1)$ is relatively compact (by means of Prokhorov's Theorem) and closed with respect to weak convergence, so that existence of a minimizer for $(\mathcal{K})$ can be easily shown by means of the Direct Methods.

The same considerations apply to the case of a generic cost function $c(x, y)$ replacing $|x - y|$, provided some mild regularity assumptions are made on c. We can also allow for more general spaces in place of $\Omega \subset \mathbb{R}^N$ and we can even take different source and target spaces. Then a general result is the following.

Theorem 1.1.1. *Let X, Y be Polish spaces with $c : X \times Y \to [0, +\infty]$ lower semicontinuous. Then for every $\rho_0 \in \mathscr{P}(X), \rho_1 \in \mathscr{P}(Y)$ the Monge-Kantorovich problem*

$$\inf_{\gamma \in \Pi(\rho_0,\rho_1)} \int_{X\times Y} c(x, y)\, d\gamma(x, y),$$

admits a solution, provided there exists a $\widetilde{\gamma} \in \Pi(\rho_0, \rho_1)$ with finite cost.

1.3. Duality. A very important tool is the so-called *Kantorovich duality* which gives an equivalent formulation of the mass transportation problem in terms of a dual one, the latter involving a maximization instead a minimization. In the general case it reads as follows: we will look in a while at some interesting particular cases.

Theorem 1.1.2 (Kantorovich duality). *The minimum of Kantorovich problem is equal to*

$$\sup \left\{ \int_X \varphi(x)\, d\rho_0(x) + \int_Y \psi(y)\, d\rho_1(y) \right\}, \tag{1.1.1}$$

where the supremum is taken over the set of $\varphi \in L^1(X, \rho_0)$ and $\psi \in L^1(Y, \rho_1)$ satisfying

$$\varphi(x) + \psi(y) \le c(x, y), \quad \textit{for } \rho_0\textit{−a.e. } x \in X,\ \ \rho_1\textit{−a.e. } y \in Y.$$

Observe that given an admissible couple (φ, ψ), it is straightforward to see that substituting φ with ψ^c given by

$$\psi^c(x) := \inf_{y\in Y} c(x, y) - \psi(y),$$

we get that $\varphi \le \psi^c$ and (ψ^c, ψ) is still admissible in (1.1.1). Moreover we get

$$\int_X \varphi(x)\, d\rho_0(x) + \int_Y \psi(y)\, d\rho_1(y) \le \int_X \psi^c(x)\, d\rho_0(x) + \int_Y \psi(y)\, d\rho_1(y).$$

The function ψ^c so defined is called $c-$*transform* of ψ. On the other hand, we see that there still holds $\psi^c(x) + \psi(y) \le c(x, y)$, so that if we further substitute ψ with ψ^{cc}, which as before is given by

$$\psi^{cc}(y) = \inf_{x\in X} c(x, y) - \psi^c(x),$$

we get that $\psi \le \psi^{cc}$ and

$$\int_X \varphi(x)\, d\rho_0(x) + \int_Y \psi(y)\, d\rho_1(y) \le \int_X \psi^c(x)\, d\rho_0(x) + \int_Y \psi^{cc}(y)\, d\rho_1(y),$$

with again $\psi^c + \psi^{cc} \le c$, by construction. All in all, we have shown that (1.1.1) is equivalent to

$$\sup \left\{ \int_X \varphi(x)\, d\rho_0(x) + \int_Y \varphi^c(y)\, d\rho_1(y) \right\}, \tag{1.1.2}$$

where the supremum now is taken over the set of $c-$concave functions, i.e. functions φ such that $\varphi = \psi^c$ for some ψ.

Remark 1.1.3. The concept of $c-$transform is clearly a generalization of the concept of Legendre-Fenchel conjugate function, the latter corresponding to $X = Y$ Hilbert space and $c(x, y) = \langle x, y\rangle_X$. In this case, $c-$concave functions are concave functions in the usual sense.

From now on, in view of our applications, we will mainly confine ourselves to the case $X = Y$, with X being the whole $\mathbb{R}^N$ or a proper subset of it. We now come to illustrate a particular case of the previous construction, characterizing the dual formulation for Monge's problem.

PROPOSITION 1.1.4. *Let us choose* $c(x, y) = |x-y|$*, then the set of* $c-$*concave functions coincides with the set of* $1-$*Lipschitz functions. Moreover for every* $\varphi \in \mathrm{Lip}_1(X)$ *we have* $\varphi^c = -\varphi$ *so that there holds*

$$\inf_{\gamma\in\Pi(\rho_0,\rho_1)} \int_{X\times X} |x-y|\, d\gamma(x,y) = \sup_{\varphi\in \mathrm{Lip}_1(X)} \int_{X\times X} \varphi(x)\, d(\rho_0(x)-\rho_1(x)).$$

Two optimizers γ *and* φ *of the left-hand and right-hand side, respectively, are related through*

$$\varphi(x) - \varphi(y) = |x-y|, \quad \text{for } \gamma-a.e.\ (x,y) \in X \times X.$$

A function maximizing the right-hand side in the previous equality is called a *Kantorovich potential*: observe that if φ is a Kantorovich potential, then the same is true for $\varphi + k$ for any $k \in \mathbb{R}$. The previous result also tells that transport rays correspond to directions of maximal slope for a Kantorovich potential.

REMARK 1.1.5. The previous result contains a complete characterization of $c-$concave functions, in the case $c(x, y) = |x - y|$. There is another interesting choice of c for which a complete characterization can be derived, that is when $c(x, y) = 1/2\,|x - y|^2$ (the factor $1/2$ is just for convenience) and $X = \mathbb{R}^N$. Indeed, in this case φ is $c-$concave if and only φ is such that $1/2\,|x|^2 - \varphi$ is convex[1]. On the other hand, if X is a proper subset of $\mathbb{R}^N$, then one can only say that every $c-$concave function φ is such that $1/2\,|x|^2 - \varphi$ is convex, but the converse implication does not hold: this is a consequence of the fact that now it is no more true that every convex function on X can be pointwisely written as the supremum of its affine minorants.

1.4. Existence of classical solutions. Up to now, we have seen that thanks to Kantorovich formulation it is almost straightforward to obtain a *weak* solution, in the form of a transport plan. Anyway, we could not be content of this weak solution and ask whether there exists at least an optimal transport plan given by a transport map, thus giving a solution to the original problem. Under appropriate assumptions on the first marginal, the answer is positive: we recall a result proven by Ambrosio in [**3**], achieved fixing a crucial gap in the celebrated (yet not completely correct) proof of Sudakov ([**82**]).

THEOREM 1.1.6 (Sudakov-Ambrosio). *Let* $\Omega \subset \mathbb{R}^N$ *be a compact convex set and* $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ *with* $\rho_0 \ll \mathscr{L}^N$*. Then Monge's problem* $(\mathcal{M})$ *admits a solution, that is there exists an optimal transport map* T*.*

Similar results (but with different proofs) are due to Caffarelli, Feldman and McCann ([**31**]), Evans and Gangbo ([**48**]), Trudinger and Wang ([**84**]), among others, while some interesting generalizations, relative to a cost given by an arbitrary convex norm $c(x, y) = \|x - y\|$ other than the Euclidean one, can be found in [**7**] and in recent works of Champion and De Pascale [**36**] and Caravenna [**32**].

[1]A function u such that $C/2\,|x|^2 - u(x)$ is convex, for some C, is called *semiconcave*.

REMARK 1.1.7. As simple examples show, uniqueness of the solution for Monge's problem can not be expected: in other words, there could exist more than one optimal map. This is due to the lack of strict convexity of the cost $|x-y|$. A basic example is the following: $X = \mathbb{R}$, $\rho_0 = \mathscr{L}^1 \llcorner [0,1]$ and $\rho_1 = \mathscr{L}^1 \llcorner [1,2]$ for which two optimal maps are given by

$$T_1(t) = t + 1 \ (\textit{translation}) \ \text{ and } \ T_2(t) = 2 - t \ (\textit{reflection}).$$

Another important case in which existence of an optimal transport map can be proven is when the cost $c(x,y)$ is a strictly convex function of $x-y$, typical examples of this being $c(x,y) = |x-y|^p$, with $p > 1$. Moreover now also uniqueness of the solution can be guaranteed. In this case, we have the following result (see [**86**]).

THEOREM 1.1.8 (Gangbo-McCann). *Let $\rho_0, \rho_1 \in \mathscr{P}(\mathbb{R}^N)$ such that $\rho_0 \ll \mathscr{L}^N$ and let $h : \mathbb{R}^N \to \mathbb{R}$ be a strictly convex and superlinear function. Suppose that there exists $\widetilde{\gamma} \in \Pi(\rho_0, \rho_1)$ such that*

$$\int_{\mathbb{R}^N \times \mathbb{R}^N} h(x-y)\, d\widetilde{\gamma}(x,y) < +\infty. \tag{1.1.3}$$

Then there exists a unique optimizer γ_0 for the Monge-Kantorovich problem

$$\min \left\{ \int_{\mathbb{R}^\times \mathbb{R}^N} h(x-y)\, d\gamma(x,y) \,:\, \gamma \in \Pi(\rho_0, \rho_1) \right\},$$

and moreover γ_0 is induced by a transport, that is

$$\gamma_0 = (\mathrm{Id} \times T)_\sharp \rho_0,$$

where the optimal map T is uniquely determined ρ_0-a.e. by the condition $(T)_\sharp \rho_0 = \rho_1$ and it has the following form

$$T(x) = x - \nabla h^*(\nabla \varphi(x)),$$

for some $c-$concave function φ.

REMARK 1.1.9. As briefly said, the previous result applies to the case $c(x,y) = 1/p\,|x-y|^p$, with $p > 1$, for which the optimal transport map takes the form

$$T(x) = x - |\nabla \varphi|^{q-2} \nabla \varphi,$$

with $q = p/(p-1)$ and φ $c-$concave function, i.e.

$$\varphi(x) = \inf_{y \in \mathbb{R}^N} \frac{1}{p} |x-y|^p - \psi(y),$$

for some function ψ. Moreover in this case in order to have (1.1.3) it is sufficient to assume that ρ_0 and ρ_1 have finite $p-$*moment* (see Section 3), that is

$$\int_{\mathbb{R}^N} |x|^p\, d\rho_0(x) < +\infty \ \text{ and } \ \int_{\mathbb{R}^N} |y|^p\, d\rho_1(y) < +\infty.$$

The choice $p = 2$, corresponding to $c(x,y) = 1/2\,|x-y|^2$, is particularly significant, indeed in this case T can be written as $T(x) = x - \nabla \varphi$, with φ semiconcave (see Remark 1.1.5) which implies that the unique optimal map T is given by the gradient of the convex function $\Psi(x) = 1/2\,|x|^2 - \varphi(x)$.

1.5. A divergence-constrained problem and the role of transport density. Back to Monge's original problem, i.e. when the cost is equal to the distance, we can take advantage of one more equivalent formulation (we could say *dual of the dual*) of $(\mathcal{K})$, which is particularly interesting, as far as it is expressed as a divergence-constrained optimization problem. This is a particular case of the so-called *continuous model of transportation* first proposed by Beckmann in [**11**]: we will come to illustrate this model in its generality in Chapter 6.

PROPOSITION 1.1.10. *For every $\rho_0, \rho_1 \in \mathscr{P}(\mathbb{R}^N)$ we have*

$$\inf_{\gamma\in\Pi(\rho_0,\rho_1)} \int_{\mathbb{R}^N\times\mathbb{R}^N} |x-y|\, d\gamma(x,y) = \inf_{\phi\in\mathcal{M}(\mathbb{R}^N;\mathbb{R}^N)} \left\{|\phi|(\mathbb{R}^N):\ \operatorname{div}\phi = \rho_0-\rho_1\right\}, \tag{1.1.4}$$

where $\mathcal{M}(\mathbb{R}^N;\mathbb{R}^N)$ is the set of vector-valued Radon measures on $\mathbb{R}^N$ and $|\phi|$ is the total variation measure. The condition on the divergence has to be intended in distributional sense, that is

$$\int_{\mathbb{R}^N} \nabla\varphi(x)\cdot d\phi(x) = \int_{\mathbb{R}^N} \varphi(x)\, d(\rho_1(x)-\rho_0(x)), \quad \textit{for every } \varphi\in C^1_c(\mathbb{R}^N).$$

PROOF. The proof is based on a minimax argument: before giving a rigorous justification of all the computations needed, it could be interesting to provide a formal proof (see also [**79**]). One first observes that the indicator function of the set of admissible ϕ can be written as

$$\sup_{\varphi\in C^1_c(\mathbb{R}^N)} \int_\Omega \nabla\varphi(x)\cdot d\phi(x) + \int_\Omega \varphi(x)\, d(\rho_0(x)-\rho_1(x)),$$

so that the original problem can also be stated in an unconstrained form as

$$\inf_{\phi\in\mathcal{M}(\mathbb{R}^N;\mathbb{R}^N)} \sup_{\varphi\in C^1_c} \left(|\phi|(\mathbb{R}^N) + \int_{\mathbb{R}^N} \nabla\varphi(x)\cdot d\phi(x) + \int_{\mathbb{R}^N} \varphi(x)\, d(\rho_0(x)-\rho_1(x))\right).$$

We now (formally, at this level) exchange the infimum and the supremum, noticing that

$$\inf_\phi \left(|\phi|(\mathbb{R}^N) - \int_{\mathbb{R}^N} \nabla\varphi(x)\cdot d\phi(x)\right) = \begin{cases} 0, & \text{if } \|\nabla\varphi\|_\infty\le 1\\ -\infty, & \text{otherwise}\end{cases}$$

so that in the end

$$\inf_{\phi\in\mathcal{M}(\mathbb{R}^N;\mathbb{R}^N)} \left\{|\phi|(\mathbb{R}^N):\ \operatorname{div}\phi=\rho_0-\rho_1\right\} = \sup_{\|\nabla\varphi\|_\infty\le 1} \int_{\mathbb{R}^N} \varphi(x)\, d(\rho_0(x)-\rho_1(x)),$$

the latter being exactly the dual formulation of Monge's problem (Proposition 1.1.4).

The rigorous justification of the previous proof uses a minimax Theorem that can be found for example in [**86**, Theorem 1.9], stating that each time we have a pair of convex functions $\Theta, \Upsilon : X \to \mathbb{R}\cup\{+\infty\}$ on a normed vector space X, such that for a certain $x_0\in X$ we have

$$\Theta(x_0) < +\infty,\ \Upsilon(x_0) < +\infty \ \text{ and } \ \Theta \text{ is continuous at } x_0,$$

then

$$\inf_{x\in X} [\Theta+\Upsilon] = \sup_{x^*\in X^*} \left[-\Theta^*(-x^*) - \Upsilon(x^*)\right]. \tag{1.1.5}$$

We are going to apply this fact with

$$\Theta(\varphi,\psi) = \begin{cases} \int_{\mathbb{R}^N} \varphi(x)\, d(\rho_1(x)-\rho_0(x)), & \text{if } \|\psi\|_\infty \le 1, \\ +\infty, & \text{otherwise,} \end{cases}$$

and

$$\Upsilon(\varphi,\psi) = \begin{cases} 0, & \text{if } \varphi \in C^1 \text{ and } \nabla\varphi = \psi, \\ +\infty, & \text{otherwise,} \end{cases}$$

both being convex functions on the Banach space $X = C_o(\mathbb{R}^N) \times C_o(\mathbb{R}^N;\mathbb{R}^N)$, whose dual space is given by $X^* = \mathcal{M}(\mathbb{R}^N) \times \mathcal{M}(\mathbb{R}^N;\mathbb{R}^N)$. Observe that we have

$$\inf_{(\varphi,\psi)} [\Theta(\varphi,\psi) + \Upsilon(\varphi,\psi)] = -\sup_{\|\nabla\varphi\|_\infty \le 1} \int_{\mathbb{R}^N} \varphi(x)\, d(\rho_0(x)-\rho_1(x)),$$

while

$$\begin{aligned} \sup_{(\mu,\phi)} [-\Theta^*(-\mu,-\phi) - \Upsilon^*(\mu,\phi)] &= \sup_{(\mu,\phi)} \Bigg[-\sup_{\varphi} \sup_{\|\psi\|_\infty \le 1} \left(-\int_{\mathbb{R}^N} \varphi\, d(\mu-\rho_0+\rho_1) - \int_{\mathbb{R}^N} \psi \cdot d\phi \right) \\ &\quad - \sup_{\varphi} \left(\int_{\mathbb{R}^N} \varphi\, d\mu + \int_{\mathbb{R}^N} \nabla\varphi \cdot d\phi \right) \Bigg] \\ &= \sup_{(\mu,\phi)} \Bigg[-|\phi|(\mathbb{R}^N) - \delta_{\{0\}}(\mu-\rho_0+\rho_1) \\ &\quad - \sup_{\varphi} \left(\int_{\mathbb{R}^N} \varphi\, d\mu + \int_{\mathbb{R}^N} \nabla\varphi \cdot d\phi \right) \Bigg] \\ &= \sup_{\phi} \left[-|\phi|(\mathbb{R}^N) - \sup_{\varphi} \left(\int_{\mathbb{R}^N} \varphi\, d(\rho_0-\rho_1) + \int_{\mathbb{R}^N} \nabla\varphi \cdot d\phi \right) \right] \\ &= -\inf_{\phi} \sup_{\varphi} \left[|\phi|(\mathbb{R}^N) + \int_{\mathbb{R}^N} \varphi\, d(\rho_0-\rho_1) + \int_{\mathbb{R}^N} \nabla\varphi \cdot d\phi \right]. \end{aligned}$$

It is then sufficient to use (1.1.5) to conclude. □

REMARK 1.1.11. From the formal proof, it is clear that the optimizers ϕ_0 and φ_0 of the two problems have to satisfy the primal-dual optimality conditions

$$\nabla\varphi_0 \text{ parallel to } \phi_0 \quad \text{and} \quad |\nabla\varphi_0(x)| = 1, \quad \text{for } \phi_0\text{−a.e. } x,$$

that is, in a very informal way, $\nabla\varphi_0 = \phi_0/|\phi_0|$, which is another way of writing the condition that transport takes place on curves of maximal slope for the Kantorovich potential φ_0. Note that in this context φ_0 can be seen as a Lagrange multiplier for the constraint on the divergence.

REMARK 1.1.12. In the case of a proper subset $\Omega \subset \mathbb{R}^N$, formula (1.1.4) still holds, provided a boundary condition is imposed on the admissible vector measures and the Euclidean distance in Monge-Kantorovich problem is substituted with the geodesic one in Ω (for example, in the case of a convex subset, this still coincides with the Euclidean one). The right boundary condition to impose is a homogeneous Neumann one, that is $\langle \phi, \nu \rangle = 0$ on $\partial\Omega$, which has to be intended in the following distributional sense

$$\int_\Omega \nabla\varphi(x) \cdot d\phi(x) = \int_\Omega \varphi(x)\, d(\rho_1(x)-\rho_0(x)), \quad \text{for every } \varphi \in C^1(\overline{\Omega}).$$

Strictly related to this divergence-constrained problem

$$\inf_{\phi\in\mathcal{M}(\Omega;\mathbb{R}^N)} \left\{|\phi|(\Omega):\ \operatorname{div}\phi=\rho_0-\rho_1,\ \langle\phi,\nu\rangle=0\right\}, \tag{1.1.6}$$

is the concept of *transport density*: given an optimal transport plan $\gamma\in\Pi(\rho_0,\rho_1)$ for Monge-Kantorovich problem with cost $c(x,y)=|x-y|$, the relative transport density i_γ is the positive measure defined through

$$\langle i_\gamma,\varphi\rangle=\int_{\Omega\times\Omega}\left(\int_{\overline{xy}}\varphi(z)\,d\mathscr{H}^1(z)\right)d\gamma(x,y),\quad\text{for every }\varphi\in C(\Omega),$$

where $\overline{xy}$ stands for the segment joining x to y. Observe that, by its very definition, for every Borel set $A\subset\Omega$ we have that $i_\gamma(A)$ represents how much the region A is used by the transport relative to γ. The importance of the transport density lies in the fact that

$$i_\gamma(\Omega)=\int_{\Omega\times\Omega}|x-y|\,d\gamma(x,y),$$

that is its mass equals the value of the Monge-Kantorovich problem. Moreover defining the vector version ϕ_γ of i_γ through

$$\langle\phi_\gamma,\varphi\rangle=\int_{\Omega\times\Omega}\left(\int_{\overline{xy}}\langle\varphi(z),\tau\rangle\,d\mathscr{H}^1(z)\right)d\gamma(x,y),\quad\text{for every }\varphi\in C(\Omega;\mathbb{R}^N),$$

τ standing for the tangent versor to the segment $\overline{xy}$, we have that ϕ_γ is admissible for (1.1.6) and

$$|\phi_\gamma|\le i_\gamma,$$

as measures on Ω. On the other hand, as far as transport rays do not cross, the two measures coincides, and ϕ_γ must be an optimizer of the divergence-constrained problem. This implies that regularity results for the transport density would immediately translates into regularity statements for optimizers of (1.1.6).

We summarize here some regularity results on the transport density: the following is a (non exhaustive) collection of results due to De Pascale, Evans, Feldmann, McCann, Pratelli, Santambrogio (see [**40, 41, 42, 50, 75**]).

Theorem 1.1.13. *Let $\rho_0,\rho_1\in\mathscr{P}(\Omega)$ with ρ_0 or ρ_1 absolutely continuous w.r.t. $\mathscr{L}^N$, then i_γ does not depend on the choice of the optimal transport plan γ and $i_\gamma\ll\mathscr{L}^N$, too. If $\rho_0,\rho_1\in\mathscr{P}(\Omega)$ have an L^p density w.r.t. to $\mathscr{L}^N$, for $p\in[1,\infty]$, then the same stays true for i_γ. Let ρ_0 have an L^p density w.r.t. $\mathscr{L}^N$, for $p<N/(N-1)$, then $i_\gamma=f\cdot\mathscr{L}^N$, with $f\in L^p$ (nothing is required on ρ_1 for this latter result to hold).*

Remark 1.1.14. For example, when $\mu_0=f_0\cdot\mathscr{L}^N$ and $\mu_1=f_1\cdot\mathscr{L}^N$ with $f_0,f_1\in L^p$, as a consequence of these results and of the discussion above, problem (1.1.6) can be equivalently settled among L^p vector fields, substituting their total variation with the L^1 norm, that is

$$(1.1.6)=\inf_{\phi\in L^p(\Omega;\mathbb{R}^N)}\left\{\|\phi\|_{L^1(\Omega)}:\ \operatorname{div}\phi=\rho_0-\rho_1,\ \langle\phi,\nu\rangle=0\right\}.$$

2. The discrete case

Let X be a discrete space and take two discrete probability measures, both with a finite number of atoms (not necessarily the same)

$$\rho_0 = \sum_{i=1}^{n} a_i\, \delta_{x_i} \quad \text{and} \quad \rho_1 = \sum_{k=1}^{s} b_k\, \delta_{y_k}, \tag{1.2.1}$$

with $x_i, y_k \in X$, for $i = 1, \ldots, n$ and $k = 1, \ldots, s$. Then any transport plan $\gamma \in \Pi(\rho_0, \rho_1)$ can be written as

$$\gamma = \sum_{i=1}^{n} \sum_{k=1}^{s} \gamma_{ik}\, \delta_{x_i} \otimes \delta_{y_k},$$

with the $n \times m$ matrix $(\gamma_{ik})_{i,k}$ belonging to the *convex set* $\pi(\rho_0, \rho_1)$ given by

$$\pi(\rho_0, \rho_1) = \left\{ M = (m_{ik}) \,:\, m_{ik} \geq 0,\ \sum_{i=1}^{n} m_{ik} = b_k,\ \sum_{k=1}^{s} m_{ik} = a_i \right\}.$$

This means that in this particular context, the Kantorovich problem

$$\inf \left\{ \int_{X \times X} c(x, y)\, d\gamma(x, y) \,:\, \gamma \in \Pi(\rho_0, \rho_1) \right\},$$

can be reformulated as

$$\inf_{M \in \pi(\rho_0, \rho_1)} \sum_{i=1}^{n} \sum_{k=1}^{s} c(x_i, y_k)\, m_{ik}, \tag{1.2.2}$$

that is a linear optimization problem, over a bounded closed convex set[2]. Then by Choquet's Theorem we know that minimizers for this problem do exist in the set $\mathrm{Ext}\,(\pi(\rho_0, \rho_1))$ of extremal points of $\pi(\rho_0, \rho_1)$. In order to characterize this set of extremal points, we need the following definition.

Definition 1.2.1. A matrix $M = (m_{ik})_{i,k=1}^{n}$ is called *acyclic* if the following property holds:

$$\prod_{r=1}^{j} m_{i_r k_r} m_{i_r k_{r+1}} = 0,$$

for every $2 \leq j \leq \min\{n, s\}$ and every set of indexes $i_1 < \cdots < i_j \in \{1, \ldots, n\}$, $k_1 < \cdots < k_j \in \{1, \ldots, s\}$ (the convention $i_{n+1} = i_1$ and $k_{s+1} = k_1$ is used).

Then the set $\mathrm{Ext}\,\big(\pi(\rho_0, \rho_1)\big)$ can be completely characterized as the set of acyclic matrices, that is we have the following (for a proof see [**39**], for example).

Theorem 1.2.2. *A matrix $M = (m_{ik})_{i,k=1}^{n}$ belongs to $\mathrm{Ext}(\pi(\rho_0, \rho_1))$ if and only if $M \in \pi(\rho_0, \rho_1)$ and it is acyclic.*

---

[2] In this form the problem was sometimes referred to as *Hitchcock's distribution problem*, after Frank Lauren Hitchcock who introduced it in [**59**], exactly in the same years of Kantorovich's work. Nowadays, the name of Hitchcock seems to have been completely neglected by the Optimal Transportation community.

Thanks to the previous characterization, going back to Kantorovich problem (1.2.2), we get the following.

PROPOSITION 1.2.3. *Let* $\rho_0, \rho_1 \in \mathscr{P}(X)$ *be as in* (1.2.1). *Then there exists an optimal transport plan* $\gamma \in \Pi(\rho_0, \rho_1)$ *such that*

$$\#\{(x, y) \in \Omega \times \Omega \, : \, \gamma(\{(x, y)\}) \neq 0\} \leq n + s - 1,$$

that is γ *does not move more that* $n + s - 1$ *atoms.*

REMARK 1.2.4. Observe that problem (1.2.2) has the dual formulation

$$\max \left(\sum_{i=1}^{n} u_i \, a_i - \sum_{k=1}^{s} v_k \, b_k \right), \tag{1.2.3}$$

with the unknown $u_1, \dots, u_n$ and $v_1, \dots, v_s$ subject to the constraints

$$u_i - v_k \leq c(x_i, y_k), \text{ for every } i \in \{1, \dots, n\}, \, k \in \{1, \dots, s\}.$$

Then we see that $M = (m_{ik})$ and $\{u_1, \dots, u_n, v_1, \dots, v_s\}$ are solutions of (1.2.2) and (1.2.3), respectively, if and only if they are admissible and they satisfy

$$m_{ik}(c(x_i, y_k) + v_k - u_i) = 0, \quad \text{for every } i, k.$$

The previous clearly implies that if $m_{ik} > 0$, then there must result $u_i - v_k = c(x_i, y_k)$ and conversely, when $u_i - v_k < c(x_i, y_k)$, then $m_{ik} = 0$ (i.e. no mass is moving from x_i to y_k). Obviously, an optimal $\{u_1, \dots, u_n, v_1, \dots, v_s\}$ is nothing but a discrete version of a Kantorovich potential (see also [**51**] for more details and a computational algorithm).

3. Wasserstein spaces and their geometry

3.1. Definitions and topological properties. Given a Polish space (X, d), that is a complete and separable metric space, and an exponent $p \in [1, \infty]$, we define the $p-$Wasserstein space $\mathcal{W}_p(X)$ as the space of all Borel probability measures μ over X, having finite p-momentum, i.e.

$$\|d(\cdot, x_0)\|_{L^p(X,\mu)} < +\infty.$$

Clearly, by means of the triangular inequality, property above does not depend on the choice of $x_0 \in X$.

REMARK 1.3.1. Observe that in the case $p = \infty$, the space $\mathcal{W}_\infty(X)$ is built up of Borel probability measures having bounded support. For example, when $X = \mathbb{R}^N$ with the usual Euclidean topology, the probability measure (c_N is a suitable renormalization constant)

$$\mu = c_N \, e^{-|x|^2} \cdot \mathscr{L}^N,$$

is an element of $\mathcal{W}_p(\mathbb{R}^N)$ for every $p \in [1, \infty)$, but $\mu \notin \mathcal{W}_\infty(\mathbb{R}^N)$.

The set $\mathcal{W}_p(X)$ can be turned into a metric space itself, equipped with the $p-$Wasserstein distance w_p which is defined as

$$w_p(\mu_1, \mu_2) = \min_{\gamma \in \Pi(\mu_1, \mu_2)} \|d(\cdot, \cdot)\|_{L^p(X \times X, \gamma)}, \;\; \mu_1, \mu_2 \in \mathcal{W}_p(X),$$

where $\Pi(\mu_1, \mu_2)$ is the set of *transport plans* between μ_1 and μ_2, that is

$$\Pi(\mu_1, \mu_2) = \{\gamma \in \mathscr{P}(\Omega \times \Omega) \ : \ (\pi_x)_\sharp \gamma = \mu_1, \ (\pi_y)_\sharp \gamma = \mu_2\},$$

with $\pi_x(x, y) = x$ and $\pi_y(x, y) = y$. The fact that w_p is positive and that

$$w_p(\mu_1, \mu_2) = 0 \Longleftrightarrow \mu_1 = \mu_2,$$

are almost straightforward consequences of the definition, while the triangular inequality is a little bit more delicate and it is a consequence of the following result, whose proof can be achieved by means of the Disintegration Theorem.

Lemma 1.3.2 (Gluing lemma). *Let μ_1, μ_2 and μ_3 be probability measures over Polish spaces X_1, X_2 and X_3, respectively. Let $\gamma_{1,2} \in \Pi(\mu_1, \mu_2)$ and $\gamma_{2,3} \in \Pi(\mu_2, \mu_3)$ be two transport plans, then there exists $\gamma \in \mathscr{P}(X_1 \times X_2 \times X_3)$ such that*

$$(\pi_{1,2})_\sharp \gamma = \gamma_{1,2} \ \text{and} \ (\pi_{2,3})_\sharp \gamma = \gamma_{2,3},$$

where $\pi_{1,2}$ and $\pi_{2,3}$ are the projections on $X_1 \times X_2$ and $X_2 \times X_3$, respectively.

The first topological properties of these metric spaces are summarized in the following statement.

Theorem 1.3.3. *Let $p \in [1, \infty]$, then if X is a Polish space, the same is true for $\mathcal{W}_p(X)$.*

Another important fact is the *ordering* relations between Wasserstein distances, which is a straightforward consequence of Hölder's inequality: indeed, we have

$$q \geq p \Longrightarrow w_q \geq w_p. \tag{1.3.1}$$

In the case that the distance d is bounded, it is also possible to give reverse inequalities of the type $w_q \leq w_p^\beta$, but with the power β degenerating as the greater exponent approaches ∞. Namely, we have

$$q \geq p \Longrightarrow w_q \leq w_p^{p/q} \operatorname{diam}(X)^{\frac{q-p}{q}}. \tag{1.3.2}$$

This means that, in the case of d bounded, all distances w_p with $p \in [1, \infty)$ define the same topology on $\mathscr{P}(X)$. This is clearly the case, for example, when $X = \Omega \subset \mathbb{R}^N$ bounded subset, equipped with the Euclidean topology.

Remark 1.3.4. The case $p = \infty$ deserves a little attention more, as far as the topology defined by w_∞ is always stronger than that defined by the other Wasserstein distances. Indeed, it is sufficient to take $x_0, x_1 \in \mathbb{R}^N$ distinct points and to consider $\mu = \delta_{x_0}$ and the sequence $\{\mu_n\}_{n\in\mathbb{N}}$ given by

$$\mu_n = \left(1 - \frac{1}{n}\right)\delta_{x_0} + \frac{1}{n}\delta_{x_1}.$$

Then for every $p \in [1, \infty)$ we get $w_p(\mu, \mu_n) = |x_0 - x_1| n^{-\frac{1}{p}}$, so that μ^n converges to μ in $\mathcal{W}_p$, while $w_\infty(\mu, \mu_n) = |x_0 - x_1|$, which shows that μ_n is not converging in $\mathcal{W}_\infty$. Anyway, it is interesting to remark that, in some very special cases, it is possible to give reverse inequalities of the type (1.3.2) involving w_∞: this is due to Bouchitté, Jimenez and Rajesh. In **[22]** they have shown that when

$\Omega \subset \mathbb{R}^N$ is a bounded open set with Lipschitz boundary and $\mu_0, \mu_1 \in \mathscr{P}(\overline{\Omega})$ with $\mu_0 = f_0 \cdot \mathscr{L}^N$, then for every $p > 1$ we have

$$w_\infty(\mu_0, \mu_1)^{p+N} \le C \left\| \frac{1}{f_0} \right\|_{L^\infty} w_p(\mu_0, \mu_1)^p,$$

with the constant C depending only on N, p and Ω, such that C blows up as p approaches 1.

After these introductory remarks, then a question naturally arise: are the topologies induced by Wasserstein distances comparable to the narrow topology, which is the one induced by convergence in duality with the space $C_b(X)$? The answer is given by the following fundamental result.

THEOREM 1.3.5. *Let $p \in [1, \infty)$ and $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{W}_p(X)$. Let $\mu \in \mathscr{P}(X)$, then the following statements are equivalent:*

(i) $\lim_{n\to\infty} w_p(\mu_n, \mu) = 0$;

(ii) $\mu_n \rightharpoonup \mu$ *and we have convergence of the moment of order p, i.e. for some* $x_0 \in X$

$$\lim_{n\to\infty} \int_X d(x_0, x)^p \, d\mu_n(x) = \int_X d(x_0, x)^p \, d\mu(x);$$

(iii) $\mu_n \overset{p}{\rightharpoonup} \mu$, *that is*

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

for every continuous function φ such that $|\varphi(x)| \le A \, d(x, x_0)^p + B$, for suitable constants A, B and $x_0 \in X$.

REMARK 1.3.6. In particular, when d is bounded and $p \in [1, \infty)$, convergence in the w_p metric is equivalent to the narrow convergence.

When X is in addition locally compact, then Riesz' Theorem identifies the space $\mathcal{M}(X)$ of finite Radon measures with the dual of the Banach space $C_o(X)$, which is given by continuous functions vanishing at infinity, i.e. $\varphi \in C_o(X)$ if φ is a continuous function and for every $\varepsilon > 0$, there exists $K_\varepsilon \subset X$ compact such that

$$\sup_{x \in X \setminus K_\varepsilon} |\varphi(x)| < \varepsilon.$$

Then $\mathcal{M}(X)$ can be equipped with the $*-$weak topology, in which convergence is defined testing against functions of $C_o(X)$: the same topology can be clearly considered on the space of probability measures $\mathscr{P}(X)$. What is important to stress is that in this case, i.e. when X is furtherly assumed to be locally compact, narrow convergence and $*-$weak convergence are equivalent, at least at the level of $\mathscr{P}(X)$.

LEMMA 1.3.7. *Given $\{\mu_n\}_{n\in\mathbb{N}}, \mu \in \mathscr{P}(X)$, we have*

$$\mu_n \rightharpoonup \mu \Longleftrightarrow \mu_n \overset{*}{\rightharpoonup} \mu.$$

The next result summarizes the compactness properties of Wasserstein spaces.

PROPOSITION 1.3.8. *Let $p \in [1, \infty)$, then $\mathcal{W}_p(X)$ is compact if and only if X is compact. The space $\mathcal{W}_\infty(X)$ is neither compact nor locally compact, for any X with $\#X > 1$.*

REMARK 1.3.9. When X is unbounded and $p \in [1, \infty)$, one may think that $\mathcal{W}_p(X)$ should be locally compact, at least if so is X: this is clearly false. Taking $X = \mathbb{R}^N$, let us fix $\varepsilon > 0$ and take a sequence of points $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^N$ such that $|x_n| = n^{1/p}$, defining $\mu = \delta_0$ and the sequence of probability measures

$$\mu_n = \left(1 - \frac{1}{n}\right)\delta_0 + \frac{1}{n}\delta_{x_n},$$

we get that $\mu_n \rightharpoonup \mu$, while $w_p(\mu_n, \mu) = n^{-1/p}\, n^{1/p} = 1$, so that $\{\mu_n\}_{n\in\mathbb{N}}$ can not have any subsequence converging to $\mu = \delta_0$ in $\mathcal{W}_p(\mathbb{R}^N)$.

3.2. Geodesics in $\mathcal{W}_p$. Let us consider the time interval $I = [0, 1]$. Given $\sigma \in AC(I; X)$, we recall that its *length* is given by

$$\ell(\sigma) = \int_I |\sigma'|(t)\, dt,$$

where $|\sigma'|(t)$ is the metric derivative at point t of the curve σ (see the Appendix A for both the definition of the space $AC(I; X)$ and of the metric derivative). Then X is said to be a *length space* if for every $x, y \in X$ there holds

$$d(x, y) = \inf\{\ell(\sigma) \,:\, \sigma \in AC(I; X),\, \sigma(0) = x,\, \sigma(1) = y\}.$$

In the case the infimum is attained for every $x, y \in X$, the space is said to be *geodesic*. Moreover $\sigma \in AC(I; X)$ connecting two distinct points x and y is said to be a *constant speed geodesic* if it is a geodesic, that is $d(x, y) = \ell(\sigma)$ and

$$d(\sigma(t), \sigma(s)) = |t - s|\, d(x, y), \;\; s, t \in I.$$

Observe that the length functional ℓ is invariant under reparametrization, so that restricting the minimization to the class of Lipschitz curves does not affect the minimal value of ℓ, but just the choice of minimizers. Moreover minimizing the $p-$energy with $p > 1$

$$\left(\int_I \left(|\sigma'|(t)\right)^p\, dt\right)^{\frac{1}{p}},$$

this again does not affect the minimal value of ℓ and it selects particular minimizers, the latter being precisely the constant speed geodesics (which are not unique, generally speaking).

PROPOSITION 1.3.10. *If X is a length (respectively, geodesic) space, then the same holds true for $\mathcal{W}_p(X)$.*

This in particular tells us that when X is a length space, the $p-$Wasserstein distance between two elements $\rho_0, \rho_1 \in \mathcal{W}_p(X)$ has the following *dynamical* formulation

$$w_p(\rho_0, \rho_1) = \inf \int_I |\mu_t'|_{w_p}\, dt, \tag{1.3.3}$$

where the infimum is taken among all absolutely continuous curves in $\mathcal{W}_p(X)$ connecting ρ_0 and ρ_1. At this point, we would like to have a characterization of the constant speed geodesics in the Wasserstein spaces: we start with a basic example.

EXAMPLE 1.3.11. Let us fix two distinct points $x_0, x_1 \in \mathbb{R}^N$ and let us consider the constant speed geodesic connecting these two points, that is the segment $\overline{x_0x_1}$ parametrized as $\overline{x_0x_1}(t) = (1-t)\,x_0 + t\,x_1$. Then it is straightforward to verify that the curve of probability measures given by

$$\mu_t = \delta_{\overline{x_0x_1}(t)},\ \ t \in [0,1],$$

is such that for every $p \in [1,\infty]$ we have

$$w_p(\mu_t,\mu_s) = |t-s|\,|x_0 - x_1| = |t-s|\,w_p(\mu_0,\mu_1),$$

that is $\mu(t)$ is a constant speed geodesic in $W_p(\mathbb{R}^N)$, connecting the two Dirac masses $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$. Observe that μ can be rewritten as $\mu_t = (e_t)_\sharp Q$, where e_t is *the evaluation at time* t, given by

$$\begin{array}{rccc} e_t: & C(I;\mathbb{R}^N) & \to & \mathbb{R}^N \\ & \sigma & \mapsto & \sigma(t), \end{array} \tag{1.3.4}$$

and Q is the probability measure over $C(I;\mathbb{R}^N)$ defined by

$$Q = \delta_{\overline{x_0x_1}},$$

that is a Dirac delta supported on the constant speed geodesic between x_0 and x_1.

The following result gives a complete characterization of the geodesics in $\mathcal{W}_p(X)$, in terms of the geodesics of the base space X: the underlying idea is exactly that of the previous example. For the proof, we refer the reader to [**6**, Theorem 7.2.2] when X is a separable Hilbert space and to [**65**, Theorem 6] for the general case.

THEOREM 1.3.12 (Constant speed geodesics in $\mathcal{W}_p$). *Let $p \in (1,\infty)$ and let X be a separable and complete geodesic space. We set*

$$\mathrm{CSG}(I;X) = \{\sigma \in AC(I;X)\ :\ d(\sigma(t),\sigma(s)) = |t-s|\,d(\sigma(0),\sigma(1)),\ \ \textit{for every } t,s \in I\},$$

that is $\mathrm{CSG}(I;X)$ *is the set of constant speed geodesics on X, parametrized over I.*

Then $\mu \in AC(I;\mathcal{W}_p(X))$ is a constant speed geodesic if and only if there exists $Q \in \mathscr{P}(C(I;X))$ such that:

(i) $\mu_t = (e_t)_\sharp Q$, *for every* $t \in I$;
(ii) Q *is concentrated on the set of constant speed geodesics* $\mathrm{CSG}(I;X)$;
(iii) $w_p(\mu_0,\mu_1)^p = \int_{C(I;X)} d(\sigma(0),\sigma(1))^p\,dQ(\sigma)$.

The case of a convex subset $\Omega \subset \mathbb{R}^N$ (which clearly comprises the case $\Omega = \mathbb{R}^N$) will be of particular interest in the sequel, so we want to spend some words more on Theorem 1.3.12 in this setting: indeed, in this situation every element of $\mathrm{CSG}([0,1];\Omega)$ is a segment $\overline{xy}$ with $x,y \in \Omega$, parametrized as $\overline{xy}(t) = (1-t)\,x + t\,y$. Then the probability measure Q given by Theorem 1.3.12 has the following form

$$Q = \int_{\Omega\times\Omega} \delta_{\overline{xy}}\,d\gamma(x,y),$$

for an optimal transport plan $\gamma \in \Pi(\mu_0,\mu_1)$: in other words, Q is the probability whose disintegration with respect to γ is given by the Borel family of Dirac masses $\{\delta_{\overline{xy}}\}_{(x,y)\in\Omega\times\Omega}$ supported on the segments. Then by means of Theorem 1.3.12, μ is a constant speed geodesic in $\mathcal{W}_p(\Omega)$ if and only if it has the form

$$\mu_t = (e_t)_\sharp Q = ((1-t)\,\pi_x + t\,\pi_y)_\sharp \gamma,\ t \in [0,1], \tag{1.3.5}$$

that is

$$\int_\Omega \varphi(x)\,d\mu_t(x) = \int_{\Omega\times\Omega} \varphi((1-t)\,x + t\,y)\,d\gamma(x,y), \quad \text{for every } \varphi \in C_b(\Omega).$$

Formula (1.3.5) is usually referred to as *displacement interpolation*, see [**66**] where the terminology has been introduced for the first time.

Finally, observe that if the optimal γ is given by a transport map T, then formula (1.3.5) can be rephrased into

$$\mu_t = ((1-t)\,\mathrm{Id} + t\,T)_\sharp \mu_0,\ t \in [0,1],$$

that is Q is given by $Q = \int \delta_{\overline{x\,T(x)}}\,d\mu_0(x)$ and

$$\int_\Omega \varphi(x)\,d\mu_t(x) = \int_\Omega \varphi((1-t)\,x + t\,T(x))\,d\mu_0(x), \quad \text{for every } \varphi \in C_b(\Omega).$$

Remark 1.3.13. In the degenerate case $p = 1$, it is no more true that every constant speed geodesics μ can be written as a displacement interpolation (while the conversely is still true). Indeed, going back to the previous example of $\rho_0 = \delta_{x_0}$ and $\rho_1 = \delta_{x_1}$, we see that the linear interpolation $\widetilde{\mu}_t = (1-t)\delta_{x_0} + t\delta_{x_1}$ is such that

$$w_1(\widetilde{\mu}_t, \widetilde{\mu}_s) = |t-s|\,|x_0 - x_1| = |t-s|\,w_1(\rho_1, \rho_0), \quad \text{for every } t, s \in I,$$

and $\widetilde{\mu}(t)$ is not of the form (1.3.5). Observe that the curve $\widetilde{\mu}$ physically corresponds to a *teleport* phenomenon: at every time, you see mass disappearing in x_0 and instantaneously appearing in x_1.

Remark 1.3.14. The same can be said for the case $p = \infty$, that is a constant speed geodesic in $\mathcal{W}_\infty(X)$ is *not* necessarily of the form $\mu_t = (e_t)_\sharp Q$, with Q concentrated on constant speed geodesics. Indeed, let us consider three pairwise distinct points $x_0, x_1, x_2 \in \mathbb{R}^N$ and the curves

$$\sigma_1(t) = (1-t)\,x_0 + t\,x_1 \quad \text{and} \quad \sigma_2(t) = (1-t^2)\,x_0 + t^2\,x_2,\ t \in [0,1],$$

then define $\mu_t = 1/2\,\delta_{\sigma_1(t)} + 1/2\,\delta_{\sigma_2(t)}$, which is a Lipschitz curve in $\mathcal{W}_\infty(\mathbb{R}^N)$ connecting $\rho_0 = \delta_{x_0}$ and $\rho_1 = 1/2\,\delta_{x_1} + 1/2\,\delta_{x_2}$. Observe that μ_t is of the form $\mu_t = (e_t)_\sharp Q$, with Q supported on geodesics, but with σ_2 not having constant speed. On the other hand, supposing that

$$|x_0 - x_1| \ge 2\,|x_0 - x_2|,$$

we have $w_\infty(\rho_0, \rho_1) = |x_0 - x_1|$, and moreover we get

$$\begin{aligned} w_\infty(\mu_t, \mu_s) = \max\{|t-s|\,|x_0 - x_1|,\, |t^2 - s^2|\,|x_0 - x_2|\} &= |t-s|\,|x_0 - x_1| \\ &= |t-s|\,w_\infty(\rho_0, \rho_1), \end{aligned}$$

so that μ is a constant speed geodesic in $\mathcal{W}_\infty(\mathbb{R}^N)$.

4. The Benamou-Brenier formula and its variants

In the paper [**13**], Benamou and Brenier introduced a very interesting alternative formulation for the 2−Wasserstein distance when $X = \mathbb{R}^N$, based on the continuity equation: they introduced it for numerical purposes, but its theoretical impact has been huge since then. We recall here their result and we will see in a while some of its variants and generalizations.

THEOREM 1.4.1 (Benamou-Brenier). *Given $\rho_0, \rho_1 \in \mathscr{P}(\mathbb{R}^N)$ having smooth densities w.r.t. to $\mathscr{L}^N$ and bounded supports, let us set*

$$A(\rho_0, \rho_1) = \{(\mu, v) : \partial_t \mu + \operatorname{div}_x(v\mu) = 0 \text{ in } I \times \mathbb{R}^N,\ \mu_0 = \rho_0,\ \mu_1 = \rho_1\}.$$

Then the 2− Wasserstein distance between ρ_0 and ρ_1 can be characterized as follows:

$$w_2(\rho_0, \rho_1)^2 = \min_{(\mu,v)\in A(\rho_0,\rho_1)} \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^2 \, d\mu_t(x)\, dt. \tag{1.4.1}$$

It is remarkable to point out that the objective functional in the Benamou-Brenier formula is nothing but the integral in time of the *kinetic energy*: in this sense, formula (1.4.1) can be rephrased by saying that the curves minimizing the integral of the kinetic energy are of minimal length in $\mathcal{W}_2(\mathbb{R}^N)$, i.e. geodesics. This immediately leads to think at v as the tangent vector to the curve μ and it suggests that somehow the space $\mathcal{W}_2(\mathbb{R}^N)$ has a kind of (infinite-dimensional) Riemannian manifold structure (the interested reader should consult [**6**]).

Under these comfortable assumptions on ρ_0 and ρ_1 (which can be considerably weakened, see next section), it is quite easy to give a proof of the Benamou-Brenier formula: indeed, to see that the minimization of the kinetic energy

$$\int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^2 \, d\mu_t(x)\, dt,$$

leads to a value which is greater than or equal to the squared 2−Wasserstein distance, it is enough to consider an admissible pair (μ, v) and observe that taking the flow map X of the vector field v, i.e.

$$\begin{cases} X'(t,x) &= v_t(X(t,x)) \\ X(0,x) &= x \end{cases}$$

the curve μ can be represented as $\mu_t = (X(t,\cdot))_\# \rho_0$ (see Appendix B, Theorem B.0.1). Using this, the fact that X is the flow map of v and Jensen's inequality, we can then estimate

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^2 \, d\mu_t(x) dt = \int_0^1 \int_{\mathbb{R}^N} |v_t(X(t,x))|^2 \, d\rho_0(x) dt &= \int_0^1 \int_{\mathbb{R}^N} |X'(t,x)|^2 \, d\rho_0(x) dt \\ &= \int_{\mathbb{R}^N} \int_0^1 |X'(t,x)|^2 \, dt \, d\rho_0(x) \\ &\geq \int_{\mathbb{R}^N} |X(1,x) - x|^2 \, d\rho_0(x) \geq w_2(\rho_0, \rho_1)^2. \end{aligned}$$

Moreover if T is the optimal transport map for $c(x, y) = |x-y|^2$, which exists by means of Theorem 1.1.8, it is straightforward to see that setting $X_t(x) = (1-t)x + tT(x)$ and taking v of the form

$$v_t(x) = T(X_t^{-1}(x)) - X_t^{-1}(x),$$

then X_t is the flow map of v and $\mu_t = (X_t)_{\#}\rho_0$ is a solution of the continuity equation, with velocity field v and ρ_0, ρ_1 as initial and final data. With this choice, we can estimate the kinetic energy as before and then observe

$$\int_{\mathbb{R}^N} \int_0^1 |X_t'(x)|^2 \, dt \, d\rho_0(x) = \int_{\mathbb{R}^N} |T(x) - x|^2 \, d\rho_0(x) = w_2(\mu_0, \mu_1)^2,$$

thus giving (1.4.1).

It is straightforward to extend the previous computations to the case of the $p-$Wasserstein distance, with $p > 1$, thus getting

$$w_p(\rho_0, \rho_1)^p = \min_{(\mu,v)\in A(\rho_0,\rho_1)} \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^p \, d\mu_t(x) \, dt.$$

We can also substitute $\mathbb{R}^N$ with a convex bounded set $\Omega \subset \mathbb{R}^N$, provided that the admissible velocity vector fields v satisfy a Neumann condition $\langle v, \nu \rangle = 0$ at the boundary $\partial\Omega$, with ν standing for the outer normal versor: this has to be intended in a weak sense, that is the formulation of the continuity equation in the sense of distributions will be now given by

$$\begin{aligned}\int_\Omega \varphi(1, x) \, d\rho_1(x) - \int_\Omega \varphi(0, x) \, d\rho_0(x) = &\int_0^1 \int_\Omega \partial_t \varphi(t, x) \, d\mu_t(x) \, dt \\ &+ \int_0^1 \int_\Omega \langle \nabla \varphi(t, x), v_t(x) \rangle \, d\mu_t(x) \, dt,\end{aligned}$$

for every $\varphi \in C^1([0, 1] \times \overline{\Omega})$. Note that from a physical point of view, the homogeneous Neumann boundary condition prevents the necessity of using boundary conditions for μ and let the flow of v stay inside Ω.

Remark 1.4.2. Observe that in the case of a non convex set $\Omega \subset \mathbb{R}^N$, we still obtain a dynamical characterization of the $p-$Wasserstein distance, but with the the Euclidean distance replaced by the geodesic distance of Ω.

One of the distinguished feature of the Benamou-Brenier formula is that it can be simply seen as a convex optimization problem under a linear constraint: to arrive at this point, one has to properly restate it in an equivalent form. First of all, one introduces the new variable $\phi_t = v_t \cdot \mu_t$, so that the continuity equation now simply rewrites as a linear equation in the variables (μ, ϕ), that is

$$\partial_t \mu_t + \mathrm{div}_x \phi_t = 0.$$

Moreover thanks to the Disintegration Theorem, we can identify the curves of measures $t \mapsto \mu_t$ and $t \mapsto \phi_t$ with the measures on $[0, 1] \times \mathbb{R}^N$ given by

$$\mu = \int \mu_t \, dt \quad \text{and} \quad \phi = \int \phi_t \, dt,$$

thus enlarging the class of admissible pairs to couples (μ, ϕ) of measures on $[0, 1] \times \mathbb{R}^N$, setting the value of the energy functional to be $+\infty$ if the requirement $\phi \ll \mu$ is not satisfied. We then observe that introducing

$$f_p(x,y) = \begin{cases} |y|^p\, x^{1-p}, & \text{if } x>0,\ y\in\mathbb{R}^N, \\ 0, & \text{if } x=0,\ y=0, \\ +\infty, & \text{otherwise }, \end{cases}$$

this is a jointly convex and positively 1−homogeneous function, so that the corresponding functional defined on measures (*Benamou-Brenier functional*)

$$\mathcal{F}_p(\mu,\phi) = \int_{[0,1]\times\mathbb{R}^N} f_p\left(\frac{d\mu}{dm},\frac{d\phi}{dm}\right) dm, \tag{1.4.2}$$

is local, lower semicontinuous and, thanks to the 1−homogeneity of f_p, does not depend on the choice of the reference measure m. Using this fact and the definition of f_p, the previous can be rephrased as

$$\mathcal{F}_p(\mu,\phi) = \begin{cases} \displaystyle\int_{[0,1]\times\mathbb{R}^N} \left|\frac{d\phi}{d\mu}(t,x)\right|^p d\mu(t,x), & \text{if } \phi \ll \mu, \\ +\infty, & \text{otherwise,} \end{cases}$$

and finally, if $\mu = \int \mu_t\,dt$ and $\phi \ll \mu$, then the same disintegration holds true for ϕ, that is $\phi = \int \phi_t\,dt = \int v_t\cdot\mu_t\,dt$, which gives

$$\mathcal{F}_p(\mu,\phi) = \int_0^1\int_{\mathbb{R}^N} \left|\frac{d\phi_t}{d\mu_t}(x)\right|^p d\mu_t(x)\,dt = \int_0^1\int_{\mathbb{R}^N} |v_t(x)|^p\, d\mu_t(x)\,dt,$$

thus recovering the objective functional in the Benamou-Brenier formula.

Other variants of mass transportation problems have been studied and can be expressed in this way by considering in (1.4.2) other *convex* functions of the pair (μ,ϕ), in particular breaking the 1−homogeneity of the functional, so that now the corresponding integral functionals also depends on the choice of the reference measure m. For example, Dolbeault, Nazaret and Savaré have introduced in [**44**] new classes of distances over $\mathscr{P}(\mathbb{R}^N)$ based on the minimization of the functional

$$\begin{aligned}\widetilde{\mathcal{F}}_{p,\beta}(\mu,\phi;m) = &\int_0^1\int_{\mathbb{R}^N} \widetilde{f}_{p,\beta}\left(\frac{d\mu_t}{dm}(x),\frac{d\phi_t}{dm}(x)\right) dm(x)\,dt \\ &+ \int_0^1\int_{\mathbb{R}^N} \widetilde{f}^{\infty}_{p,\beta}\left(\frac{d\mu_t^{\perp}}{d\eta}(x),\frac{d\phi_t^{\perp}}{d\eta}(x)\right) d\eta(x)\,dt,\end{aligned}$$

where:

- $\widetilde{f}_{p,\beta}(x,y) = |y|^p\, x^{\beta(1-p)}$ or more precisely

$$\widetilde{f}_{p,\beta}(x,y) = \begin{cases} |y|^p\, x^{\beta(1-p)}, & \text{if } x>0,\ y\in\mathbb{R}^N, \\ 0, & \text{if } x=0,\ y=0, \\ +\infty, & \text{otherwise }, \end{cases}$$

- $\widetilde{f}^{\infty}_{p,\beta}$ is its *recession function*, i.e.

$$\widetilde{f}^{\infty}_{p,\beta}(x,y) = \lim_{t\to+\infty}\frac{\widetilde{f}_{p,\beta}(tx,ty)}{t},\ (x,y)\in\mathbb{R}\times\mathbb{R}^N;$$

- $p\geq 1$ and $0\leq\beta\leq 1$;

- η is a reference measure[3], such that $\mu^\perp \ll \eta$ and $\phi^\perp \ll \eta$.

These distances are connected to the non-linear mobility continuity equation

$$\partial_t \mu_t + \mathrm{div}_x(\mu_t^\beta\, v_t) = 0,$$

and if $\beta \in [0,1]$, i.e. the function $x \mapsto x^\beta$ is concave, the problem of minimizing $\widetilde{\mathcal{F}}_{p,\beta}$ under the constraint of $\partial_t \mu_t + \mathrm{div}_x \phi_t = 0$, where $\phi_t = \mu_t^\beta\, v_t$, turns out to be convex as well: in this way, the authors of [**44**] obtain a wide family of new transport distances on the space of probability measures, interpolating between the usual Wasserstein ones, corresponding to the choice $\beta = 1$ (so that $\widetilde{\mathcal{F}}_{p,1} = \mathcal{F}_p$), and the dual Sobolev ones, corresponding to $\beta = 0$, for which $\widetilde{\mathcal{F}}_{p,0}$ takes the form (here we choose $m = \mathscr{L}^N$)

$$\int_0^1 \int_{\mathbb{R}^N} \left|\frac{d\phi_t}{d\mathscr{L}^N}(x)\right|^p dx\,dt, \tag{1.4.3}$$

with the convention that the value of the energy if $+\infty$ if ϕ_t is not absolutely continuous w.r.t. $\mathscr{L}^N$ (this problem has been addressed by Brenier in [**27**]). Actually, the functionals considered in [**44**] are slightly more general, as far as they treat the case of general increasing and concave functions $x \mapsto h(x)$, not just powers, i.e. $\widetilde{f}_{p,h}(x,y) = |y|^p\, h(x)^{1-p}$.

We point out that the main interest of these distances is in the study of diffusion equations of the type

$$\partial_t \mu_t + \mathrm{div}_x(h(\mu_t)\,|\xi|^{q-2}\xi), \quad \xi = -\nabla\left(\frac{\partial \mathscr{F}}{\partial \mu}\right),$$

where $\partial\mathscr{F}/\partial\mu$ is the first variation of a given functional $\mathscr{F}$ and $q = p/(p-1)$: indeed, at least formally, these equations can be interpreted as gradient flows of $\mathscr{F}$ with respect to these new family of dynamical distances.

Remark 1.4.3. It is straightforward to see that in the variational problem for (1.4.3), the time variable can be eliminated. Namely, we have

$$\min\left\{\int_0^1 \int_{\mathbb{R}^N} \left|\frac{d\phi_t}{d\mathscr{L}^N}(x)\right|^p dx\,dt : \begin{array}{c}\partial_t\mu_t + \mathrm{div}_x \phi_t = 0,\\ \mu_0 = \rho_0,\ \mu_1 = \rho_1\end{array}\right\} = \min\left\{\int_{\mathbb{R}^N} |\Phi(x)|^p\,dx \,:\, \mathrm{div}\,\Phi = \rho_0 - \rho_1\right\},$$

where the problem in the right-hand side will be addressed in great details in Chapter 6 (observe that this is nothing but the dual formulation of the $q-$Laplace equation, with $q = p/(p-1)$). Indeed, taking a minimizer (μ_t, ϕ_t) of the left-hand side and considering

$$\Phi := \int_0^1 \frac{d\phi_t}{d\mathscr{L}^N}\,dt,$$

this is an admissible vector field for the problem in the right-hand side and exchanging the order of integration, we get by means of Jensen inequality

$$\int_0^1\int_{\mathbb{R}^N}\left|\frac{d\phi_t}{d\mathscr{L}^N}(x)\right|^p dx\,dt = \int_{\mathbb{R}^N}\int_0^1\left|\frac{d\phi_t}{d\mathscr{L}^N}(x)\right|^p dt\,dx \ge \int_{\mathbb{R}^N}\left|\int_0^1 \frac{d\phi_t}{d\mathscr{L}^N}(x)\,dt\right|^p dx = \int_{\mathbb{R}^N}|\Phi(x)|^p\,dx.$$

[3] As a consequence of the 1−homogeneity of the recession function, the term of $\widetilde{\mathcal{F}}_{p,\beta}$ containing $\widetilde{f}^{\infty}_{p,\alpha}$ does not depend on the choice of η.

On the other hand, if Φ minimizes the right-hand side, then we can build an admissible pair (μ_t, ϕ_t) just by setting

$$\phi_t \equiv \Phi \cdot \mathscr{L}^N \quad \text{and} \quad \mu_t = (1-t)\rho_0 + t\rho_1,$$

and observe that thanks to the condition on the divergence of ϕ, this pair verifies the continuity equation and has exactly the same energy as ϕ, thus giving the equivalence between the two problems.

REMARK 1.4.4. In connection with *congestion* effects and crowd motion, other variants of the Benamou-Brenier functional include penalizations on high densities: in [**30**] the case

$$\widehat{\mathcal{F}}(\mu, \phi) = \mathcal{F}_p(\mu, \phi) + k \int_{[0,1]\times\Omega} \left| \frac{d\mu}{dm}(t,x) \right|^2 dm(t,x),$$

with $p \geq 1$ and $k > 0$, has been considered as a model for crowd motion in a congested situation (for instance in case of panic). Here the reference measure is given by $m = \mathscr{L}^1 \otimes \mathscr{L}^N$, so that it is understood that if μ is not absolutely continuous w.r.t. m, then the value of the energy functional is infinite. This problem as well is convex: observe that in this model, curves of diffused measure are favoured not only because the functional has a finite value only on them, but also thanks to the fact that the function $x \mapsto x^2$ is super-additive, that is

$$m_1^2 + m_2^2 < (m_1 + m_2)^2,$$

so that the masses have the interest to travel separately, in order to lower the value of the energy functional (see [**30**] for more details and interesting numerical simulations).

5. More on Wasserstein distances and the continuity equation

5.1. The superposition principle. The next result will be useful in Chapters 4 and 7: it can be seen as the probabilistic counterpart to the classical method of characteristics for the continuity equation (a brief account of this is in the Appendix B, while for an excellent and self-contained exposition the reader is referred to [**6**, Chapter 8]), providing the right duality between Eulerian and Lagrangian viewpoints. Roughly speaking, it assures that a solution of the continuity equation with velocity field v, is just a superposition of integral curves of v: this is Theorem 8.2.1 of [**6**].

THEOREM 1.5.1 (Superposition principle). *Let $\mu : I \to \mathscr{P}(\mathbb{R}^N)$ be a narrowly continuous curve, solving the continuity equation*

$$\partial_t \mu_t + \operatorname{div}_x(v_t\, \mu_t) = 0, \quad \text{in } I \times \mathbb{R}^N,$$

for some Borel velocity vector field $v : (x,t) \to v_t(x)$, satisfying the integrability condition

$$\int_I \int_{\mathbb{R}^N} |v_t(x)|^p \, d\mu_t(x)\, dt < +\infty,$$

for some $p > 1$. Then there exists a probability measure $Q \in \mathscr{P}(C(I;\mathbb{R}^N))$ such that

(i) Q is concentrated on curves of $AC^p(I;\mathbb{R}^N)$ which are integral solutions of the ODE $\sigma'(t) = v_t(\sigma(t))$, in the sense that:

$$\int_{C(I;\mathbb{R}^N)} \left| \sigma(t) - \sigma(0) - \int_0^t v_s(\sigma(s))\, ds \right| dQ(\sigma) = 0, \quad \text{for every } t \in I.$$

(ii) $\mu_t = (e_t)_\sharp Q$, *for every* $t \in I$.

Conversely, any Q *satisfying (i) and the integrability condition*

$$\int_I \int_{C(I;\mathbb{R}^N)} |v_t(\sigma(t))| \, dQ(\sigma) \, dt < +\infty,$$

is such that $\mu_t = (e_t)_\sharp Q$ *is a distributional solution of the continuity equation, with velocity vector field* v.

DEFINITION 1.5.2. Let $Q \in \mathscr{P}(C([0,1];\mathbb{R}^N)$ be concentrated on the absolutely continuous solutions of $\sigma'(t) = v_t(\sigma(t))$, in the sense precised before. Then the curve of measures $\mu_t = (e_t)_\sharp Q$ given by the previous Theorem is called *superposition solution* of the continuity equation.

REMARK 1.5.3. It is not hard to see that when v is smooth, formula $\mu_t = (e_t)_\sharp Q$ is exactly equivalent to the method of characteristics. Indeed in this case, for every $x \in \Omega$, there exists a unique curve $X(\cdot, x)$ solving

$$\begin{cases} X'(t,x) &= & v_t(X(t,x)) \\ X(0,x) &= & x \end{cases}$$

so that Q admits the disintegration $Q = \int_\Omega Q^x \; d\mu_0(x)$, where for μ_0-a.e. x, Q^x is a Dirac mass concentrated on this curve, that is

$$Q^x = \delta_{X(\cdot,x)}.$$

This clearly implies the following representation formula for the superposition solution

$$\mu_t = (e_t)_\sharp Q = (X(t,\cdot))_\sharp \mu_0,$$

thus giving the interpretation of the superposition principle as a *probabilistic* version of the method of characteristics.

5.2. Characterization of absolutely continuous curves in $\mathcal{W}_p$. Next, we recall the following important result giving a complete characterization of the space $AC^p(I;\mathcal{W}_p(\mathbb{R}^N))$ in the case $p > 1$. It can be seen as a generalization of the Benamou-Brenier formula for the Wasserstein distance: this is Theorem 8.3.1 of [**6**].

THEOREM 1.5.4. *Let* $\mu \in AC^p(I;\mathcal{W}_p(\mathbb{R}^N))$, *then there exists a Borel vector field* $v : I \times \mathbb{R}^N \to \mathbb{R}^N$ *such that*

$$v_t \in L^p(\mathbb{R}^N, \mu_t), \quad \|v_t\|_{L^p(\mathbb{R}^N,\mu_t)} \le |\mu_t'|_{w_p} \text{ for } \mathscr{L}^1-a.e. \; t \in I,$$

and the continuity equation

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0, \quad \text{in } \mathbb{R}^N \times I,$$

holds in the sense of distributions.

On the other hand, if $\mu : I \to \mathcal{W}_p(\mathbb{R}^N)$ *is a narrowly continuous curve satisfying the continuity equation for some Borel vector field* $v : I \times \mathbb{R}^N \to \mathbb{R}^N$, *with*

$$\int_I \|v_t\|_{L^p(\mathbb{R}^N,\mu_t)} \, dt < +\infty,$$

then $\mu_t \in AC(I;\mathcal{W}_p(\mathbb{R}^N))$ *and* $|\mu_t'|_{w_p} \le \|v_t\|_{L^p(\mathbb{R}^N,\mu_t)}$ *for* $\mathscr{L}^1-a.e.$ $t \in I$.

REMARK 1.5.5. As we said, the previous statement can be seen as a generalization of the Benamou-Brenier formula and in fact one can easily see that it implies this statement, in its general form: indeed for every $\rho_0, \rho_1 \in \mathcal{W}_p(\mathbb{R}^N)$, let us take a constant speed geodesic μ_t between them, then

$$w_p(\rho_0, \rho_1)^p = \int_0^1 |\mu_t'|_{w_p}^p \, dt.$$

On the other hand, by means of the previous Theorem we get that for every velocity vector field v_t such that (ρ_t, v_t) verifies the continuity equation, there holds

$$\int_0^1 |\mu_t'|_{w_p}^p \, dt \leq \int_0^1 \int_\Omega |v_t(x)|^p \, d\mu_t(x) \, dt,$$

and we have equality at least for one of these vector fields, thus recovering the Benamou-Brenier formula in the general setting (no smoothness assumptions are needed).

When X is only a Polish space with no linear or differentiable structures, the continuity equation does not make sense anymore in this setting, so the previous characterization does not apply. Anyway, it is still possible to give a suitable adaptation of Theorem 1.5.4, thus giving a complete characterization of the space $AC^p(I; \mathcal{W}_p(X))$ in terms the elements of $AC^p(I; X)$: this extension is due to Lisini (see [**65**, Theorems 4 and 5]) and it has its own interest.

THEOREM 1.5.6. *Let X be a Polish space and $p \in (1, +\infty)$. If $\mu \in AC^p(I; \mathcal{W}_p(X))$, then there exists $Q \in \mathscr{P}(C(I; X))$ such that*

(i) *$Q(C(I; X) \setminus AC^p(I; X)) = 0$, i.e. Q is concentrated on $p-$absolutely continuous curves;*
(ii) *$(e_t)_\sharp Q = \mu_t$, for every $t \in I$;*
(iii) *$|\mu_t'|_{w_p}^p = \int_C |\sigma'|^p(t) \, dQ(\sigma)$, for $\mathscr{L}^1-$a.e. $t \in I$.*

On the other hand, for every $Q \in \mathscr{P}(C(I; X))$ which is concentrated on $AC^p(I; X)$ and satisfies the integrability condition

$$\int_{C(I;X)} \int_I |\sigma'|^p(t) \, dt \, dQ(\sigma) < +\infty,$$

we get that $(e_t)_\# Q := \mu_t$ is an element of $AC^p(I; \mathcal{W}_p(X))$ and

$$|\mu_t'|_{w_p}^p \leq \int_{C(I;X)} |\sigma'|^p(t) \, dQ(\sigma), \quad \textit{for } \mathscr{L}^1-\textit{a.e. } t \in I.$$

REMARK 1.5.7. Observe that as a Corollary of the previous Theorem, we recover a kind of Benamou-Brenier formula also in this general case, provided the metric space X is furtherly assumed to be geodesic. Indeed, this implies that $\mathcal{W}_p(X)$ is geodesic too (this is Proposition 1.3.10), so that for every $\rho_0, \rho_1 \in \mathcal{W}_p(X)$ there exists a constant speed geodesic μ between them such that

$$w_p(\rho_0, \rho_1)^p = \int_0^1 |\mu_t'|_{w_p}^p \, dt,$$

and thanks to point (iii) of Theorem 1.5.6, we get that there exists Q concentrated on the space $AC^p(I; \mathcal{W}_p(X))$ such that

$$\int_0^1 |\mu_t'|_{w_p}^p \, dt = \int_0^1 \int_{C(I;X)} |\sigma'(t)|^p \, dQ(\sigma) \, dt,$$

and $\mu_t = (e_t)_\sharp Q$. Conversely, to every Q concentrated on $AC^p(I;X)$ and such that $(e_i)_\sharp Q = \rho_i$, for $i = 0, 1$, we associate the curve of measures $\widetilde{\mu}_t = (e_t)_\sharp Q$ and we have

$$w_p(\rho_0, \rho_1) \leq \int_0^1 |\widetilde{\mu}'_t|^p_{w_p}\, dt \leq \int_0^1 \int_{C(I;X)} |\sigma'(t)|^p\, dQ(\sigma)\, dt.$$

All in all, we have proven the following

$$w_p(\rho_0, \rho_1)^p = \min \left\{ \int_0^1 \int_{C(I;X)} |\sigma'(t)|^p\, dQ(\sigma)\, dt \;:\; \begin{array}{c} Q \text{ concentrated on } AC^p(I;X) \\ (e_i)_\sharp Q = \rho_i,\ i = 0,1 \end{array} \right\}. \tag{1.5.1}$$

Remark 1.5.8. Theorem 1.5.6 is still interesting also in the case $X = \mathbb{R}^N$, as far as it provides a nice and explicit representation of the velocity vector field given by the first part of Theorem 1.5.4. Indeed, given $\mu \in AC^p(I; \mathcal{W}_p(\mathbb{R}^N))$, let us consider the corresponding probability measure Q constructed in Theorem 1.5.6. Using the fact that $\mu_t = (e_t)_\sharp Q$ and disintegrating Q as follows

$$Q = \int Q^t_x\, d\mu_t(x),$$

where, for every $t \in I$ and μ_t−a.e. x, Q^t_x is a probability measure concentrated on the set $\{\sigma \in AC^p(I;\mathbb{R}^N) \,:\, \sigma(t) = x\}$, we can define the vector field

$$v_t(x) = \int_C \sigma'(t)\, dQ^t_x(\sigma), \tag{1.5.2}$$

that is $v_t(x)$ is the average of the velocities of the curves passing from x at time t, according to Q. Then using Jensen's inequality and the fact that $Q = \int Q^t_x\, d\mu_t(x)$, we easily get

$$\int_{\mathbb{R}^N} |v_t(x)|^p\, d\mu_t(x) \leq \int_C |\sigma'(t)|^p\, dQ(\sigma) \leq |\mu'_t|^p_{w_p} \text{ for } \mathscr{L}^1\text{−a.e. } t \in I,$$

and moreover it is easily seen that the pair (μ, v) solves the continuity equation, in the sense of distributions.

5.3. The case of $\mathcal{W}_\infty$. Back to the Euclidean case, we have the following analogue of Theorem 1.5.4 in the extremal case $p = \infty$: this result will be useful in Chapter 4. For simplicity, we take the interval $I = [0, 1]$.

Proposition 1.5.9. *Let $\mu \in AC^\infty(I; \mathcal{W}_\infty(\mathbb{R}^N))$, then there exists a Borel vector field $v : I \times \mathbb{R}^N \to \mathbb{R}^N$ such that*

$$v_t \in L^\infty(\mathbb{R}^N, \mu_t), \quad \|v_t\|_{L^\infty(\mathbb{R}^N,\mu_t)} \leq |\mu'_t|_{w_\infty} \text{ for } \mathscr{L}^1\text{−a.e. } t \in I, \tag{1.5.3}$$

and the continuity equation

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0, \quad \text{in } I \times \mathbb{R}^N,$$

holds in distributional sense.

On the other hand, if $\mu : I \to \mathcal{W}_\infty(\mathbb{R}^N)$ is a narrowly continuous curve satisfying the continuity equation for some Borel vector field $v : I \times \mathbb{R}^N \to \mathbb{R}^N$, with

$$\int_I \|v_t\|_{L^\infty(\mathbb{R}^N,\mu_t)}\, dt < +\infty,$$

then $\mu_t \in AC(I; \mathcal{W}_\infty(\mathbb{R}^N))$ and $|\mu'_t|_{w_\infty} \leq \|v_t\|_{L^\infty(\mathbb{R}^N,\mu_t)}$ for $\mathscr{L}^1$−a.e. $t \in I$.

Proof. Let us fix an exponent $p > 1$ and set $p_n = p + n$, with $n \in \mathbb{N}$. Then for every $n \in \mathbb{N}$, by means of Theorem 1.5.4 there exists a vector Borel field v_{p_n} such that

$$\|v_{p_n,t}\|_{L^{p_n}(\mu_t)} = |\mu_t'|_{w_{p_n}} \leq |\mu_t'|_{w_\infty}, \quad \text{for } \mathscr{L}^1- \text{ a.e.} t \in [0,1],$$

and by Jensen's inequality, we also have

$$\|v_{p_n,t}\|_{L^p(\mu_t)} \leq \|v_{p_n,t}\|_{L^{p_n}(\mu_t)}, \quad \text{for every } n \in \mathbb{N},$$

thus getting a uniform bound on the L^p norm of the sequence $\{v_{p_n}\}_{n\in\mathbb{N}}$.

Moreover setting $\mu = \int_0^1 \mu_t\, dt$, i.e. the probability measure on $[0,1] \times \mathbb{R}^N$ whose disintegration with respect to the time variable is given by μ_t for $\mathscr{L}^1-$a.e. $t \in [0,1]$, we introduce the sequence of vector measures $\phi_{p_n} = v_{p_n} \cdot \mu$, which are absolutely continuous w.r.t. to μ. From the previous considerations, we see that

$$|\phi_{p_n}|([0,1] \times \mathbb{R}^N) \leq \int_0^1 \left(\int_{\mathbb{R}^N} \left| \frac{d\phi_{p_n,t}}{d\mu_t}(x) \right|^{p_n} d\mu_t(x) \right)^{\frac{1}{p_n}} dt \leq \operatorname*{ess\,sup}_{t\in[0,1]} |\mu_t'|_{w_\infty},$$

that is $\{\phi_{p_n}\}_{n\in\mathbb{N}}$ have equi-bounded total variations. In particular we have $\phi_{p_n} \rightharpoonup \phi$ as measures on $[0,1] \times \mathbb{R}^N$: at first, we would like to obtain that this limit measure ϕ can still be disintegrated as

$$\phi = \int \phi_t\, dt, \quad \text{with } \phi_t = v_t \cdot \mu_t, \text{ for } \mathscr{L}^1-\text{a.e. } t \in [0,1]. \tag{1.5.4}$$

Anyway, this is a consequence of the lower semicontinuity of the Benamou-Brenier functional (1.4.2): indeed, we get

$$\begin{aligned} \mathcal{F}_p(\mu,\phi) &\leq \liminf_{n\to\infty} \int_0^1 \int_{\mathbb{R}^N} \left| \frac{d\phi_{p_n,t}}{d\mu_t} \right|^p d\mu_t(x)\, dt \\ &= \lim_{n\to\infty} \int_0^1 |\mu_t'|^p_{w_{p_n}}\, dt \leq \int_0^1 |\mu_t'|^p_{w_\infty}\, dt, \end{aligned}$$

and the finiteness of the Benamou-Brenier functional on (μ,ϕ) implies that $\phi \ll \mu$ and then the required disintegration (1.5.4) on ϕ. We can now choose $t_1 < t_2 \in [0,1]$ and use again the lower semicontinuity of the Benamou-Brenier functional, now localized in time on the interval $[t_1,t_2]$, thus getting

$$\int_{t_1}^{t_2} \int_{\mathbb{R}^N} \left| \frac{d\phi_t}{d\mu_t} \right|^p d\mu_t(x)\, dt \leq \int_{t_1}^{t_2} |\mu_t'|^p_{w_\infty}\, dt,$$

and dividing both members by $(t_2 - t_1)$ and taking the limit as $t_2 \to t_1$, Lebesgue Differentiation Theorem then yields

$$\int_{\mathbb{R}^N} \left| \frac{d\phi_t}{d\mu_t} \right|^p d\mu_t(x) \leq |\mu_t'|^p_{w_\infty}, \quad \text{for } \mathscr{L}^1-\text{a.e. } t \in [0,1],$$

that is $\|v_t\|_{L^p(\mathbb{R}^N,\mu_t)} \leq |\mu_t'|_{w_\infty}$, for $\mathscr{L}^1-$a.e. $t \in [0,1]$. Finally, using the fact that

$$\|v_t\|_{L^\infty(\mathbb{R}^N,\mu_t)} = \lim_{p\nearrow\infty} \|v_t\|_{L^p(\mathbb{R}^N,\mu_t)},$$

we get the desired result (1.5.3).

Let us now prove the second assertion: thanks to the assumptions on v_t and μ_t, we have in particular that

$$\int_I \|v_t\|_{L^p(\mathbb{R}^N,\mu_t)}\,dt < +\infty, \quad \text{for every } p > 1,$$

so that for every $p > 1$ we have $\mu_t \in AC(I; \mathcal{W}_p(\mathbb{R}^N))$ and $|\mu_t'|_{w_p} \leq \|v_t\|_{L^p(\mathbb{R}^N,\mu_t)} \leq \|v_t\|_{L^\infty(\mathbb{R}^N,\mu_t)}$ for $\mathscr{L}^1-$a.e. $t \in I$. We then observe that for every $t \in I$ and $h > 0$, we have

$$\begin{aligned} w_\infty(\mu_{t+h}, \mu_t) = \sup_{p>1} w_p(\mu_{t+h}, \mu_t) &\leq \sup_{p>1} \int_t^{t+h} \|v_s\|_{L^p(\mathbb{R}^N,\mu_s)}\,ds \\ &\leq \int_t^{t+h} \|v_s\|_{L^\infty(\mathbb{R}^N,\mu_s)}\,dt, \end{aligned}$$

which implies that $\mu_t \in AC(I; \mathcal{W}_\infty(\mathbb{R}^N))$ and $|\mu_t'|_{w_\infty} \leq \|v_t\|_{L^\infty(\mathbb{R}^N,\mu_t)}$ for $\mathscr{L}^1-$a.e. $t \in I$ (see Appendix A, Theorem A.2.2). □

Example 1.5.10. On the contrary, in the degenerate case $p = 1$, the first implication of Theorem 1.5.4 is no more true, due to the teleport phenomenon we have already encountered. For example, for the absolutely continuous curve $\mu_t = (1-t)\delta_{x_0} + t\delta_{x_1}$ in $\mathcal{W}_1(\mathbb{R}^N)$ there can not exist any velocity field v_t such that

$$v_t \in L^1(\mathbb{R}^N, \mu_t), \quad \int_{\mathbb{R}^N} |v_t(x)|\,d\mu_t(x) \leq |\mu_t'|_{w_1}, \quad \text{for } \mathscr{L}^1-\text{a.e. } t \in I,$$

and $\partial_t \mu_t + \operatorname{div}_x(v_t \mu_t) = 0$ or, in other words, there can not exist any $\mathbb{R}^N-$valued measure ϕ on $[0,1] \times \mathbb{R}^N$ such that $\phi = \int \phi_t\,dt$ with $\phi_t \ll \mu_t$, $|\phi_t|(\mathbb{R}^N) \leq |\mu_t'|_{w_1}$ and $\partial_t \mu + \operatorname{div}_x \phi = 0$: indeed, if on the contrary this were true, then ϕ_t would be a $\mathbb{R}^N-$valued atomic measure whose divergence is still an atomic measure, i.e.

$$\operatorname{div}_x(\phi_t) = \delta_{x_0} - \delta_{x_1},$$

and this is clearly not possible.

Concerning the second statement of Theorem 1.5.4, in the case $p = 1$ it reads as follows: if $\mu : I \to \mathcal{W}_1(\mathbb{R}^N)$ is a narrowly continuous curve such that $\partial_t \mu + \operatorname{div}_x(\phi) = 0$ for some $\mathbb{R}^N-$valued measure $\phi = \int \phi_t\,dt$ on $I \times \mathbb{R}^N$ with $t \mapsto |\phi_t|(\mathbb{R}^N)$ integrable, then $\mu \in AC(I; \mathcal{W}_1(\mathbb{R}^N))$ and $|\mu_t'|_{w_1} \leq |\phi_t|(\mathbb{R}^N)$.

CHAPTER 2

Curves of minimal action over metric spaces

1. Introduction

This chapter is mostly based on the paper [**B2**], where semicontinuity results and minimization problems were presented for action functionals of the form

$$\mathcal{A}(\mu) = \int_I f(t, \mu(t), |\mu'|(t))\ dt \quad \text{and} \quad \overline{\mathcal{A}}(\nu, \mu) = \int_I f(t, \nu(t), |\mu'|(t))\ dt,$$

where $I \subset \mathbb{R}$ is an interval, $\mu \in AC^p(I; X)$ and $\nu \in L^1(I; Y)$, with X and Y Polish metric spaces. The approach here is rather general (general metric spaces, general action functionals etc.), although for the scope of this work, it is particularly tempting to substitute the phrase *metric space* with the more appealing one *Wasserstein space* in every statement that will follow. By the way, the last two sections are exactly focused on the case in which the metric spaces considered are spaces of measures (Wasserstein spaces and spaces of finite Radon measures, equipped with the $*-$weak topology), with an action functional depending on two variables, i.e. of the form $\overline{\mathcal{A}}$: this was one of the main original motivations of [**B2**], in particular we were interested in the possibility to split the curve $\mu : I \to \mathcal{W}_p(\Omega)$ into, roughly speaking, its *moving part* ν and the part that has already reached its final destination. Then one considers an action functionals of the form $\overline{\mathcal{A}}$, which takes into account only the contribution of the moving part, given by the curve ν.

The main motivation for such a study is the following: as we said, the scope of this work is to consider some models of dynamical transportation, in which the transport cost encourages or discourages the aggregation of masses; in these cases, it could happen that in an optimal configuration, some masses arrive at destination and then stop, while others still keep on moving. Then we want our cost to avoid keeping into account the contribution of these stopped masses: this is precisely the case of branched transportation (see next chapter). Indeed, for the aims of this thesis, the main applications of the results contained in Section 5 and Section 6 will be in Chapter 4.

2. Some preliminary semicontinuity results

In this chapter we will always assume that (X, d) is a Polish space, with a given Borel measure $\mathfrak{m}$. Moreover $I = [0, T] \subset \mathbb{R}$ is a compact interval, while by $\mathscr{L}^1$ we mean the 1-dimensional Lebesgue measure. We will use several basic facts about particular spaces of curves (i.e. summable, continuous, absolutely continuous, with bounded variation) in a metric space, together with the definition of *metric derivative* of a curve: the reader is referred to the Appendix A for the main definitions and properties.

We start with the following basic result:

LEMMA 2.2.1. *Let $p \in [1, +\infty]$, for every measurable subset $B \subset I$ such that $\mathscr{L}^1(B) > 0$, the functional*

$$\mu \mapsto \int_B |\mu'|(t)\ dt,\ \ \mu \in AC^p(I; X), \tag{2.2.1}$$

is sequentially l.s.c. on $AC^p(I; X)$, with respect to the weak topology *(see Appendix A, Definition A.2.6).*

PROOF. Let $B \subset I$ be any measurable subset such that $\mathscr{L}^1(B) > 0$ and take $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I; X)$ a sequence weakly converging to $\mu \in AC^p(I; X)$. We can assume that the sequence $\{|\mu_n'|\}_{n\in\mathbb{N}} \subset L^p(I; \mathbb{R})$ weakly ($*$-weakly if $p = +\infty$) converges to a function $v \in L^p(I; \mathbb{R})$.

Then we have

$$\begin{aligned} d(\mu(s), \mu(t)) &= \lim_{n\to+\infty} d(\mu_n(s), \mu_n(t)) \le \lim_{n\to\infty} \int_s^t |\mu_n'|(r)\ dr \\ &= \int_s^t v(r)\ dr, \quad \text{for every } s, t \in I \ \text{ such that } s \le t, \end{aligned}$$

which clearly shows by Lebesgue Differentiation Theorem that

$$|\mu'|(t) \le v(t), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{2.2.2}$$

This in turn implies that

$$\int_B |\mu'|(t)\ dt \le \int_B v(t)\ dt = \liminf_{n\to\infty} \int_B |\mu_n'|(t)\ dt,$$

which gives the lower semicontinuity of (2.2.1). □

With a little extra work, Lemma 2.2.1 can be improved as follows:

LEMMA 2.2.2. *Let $p \in [1, +\infty]$, for every measurable subset $B \subset I$ such that $\mathscr{L}^1(B) > 0$ and every measurable function $\varphi : B \to \mathbb{R}^+$, the functional*

$$\mu \mapsto \int_B \varphi(t)|\mu'|(t)\ dt,\ \ \mu \in AC^p(I; X), \tag{2.2.3}$$

is sequentially l.s.c. on $AC^p(I; X)$, with respect to the weak topology.

PROOF. Take $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I; X)$ a sequence weakly converging to $\mu \in AC^p(I; X)$ and call $v \in L^p(I; \mathbb{R})$ the weak ($*$-weak if $p = +\infty$) limit of $\{|\mu_n'|\}_{n\in\mathbb{N}}$. If we assume for the moment that $\varphi \in L^\infty(B; \mathbb{R}^+)$, using (2.2.2) we get

$$\int_B \varphi(t)|\mu'|(t)\ dt \le \int_B \varphi(t)v(t)\ dt = \lim_{n\to\infty} \int_B \varphi(t)|\mu_n'|(t)\ dt. \tag{2.2.4}$$

In the general case of φ measurable and positive, it is enough to define the sequence

$$\varphi_k(t) = \min\ \{\varphi(t), k\},\ \ t \in B,$$

so that $\varphi_k \in L^\infty(B;\mathbb{R}^+)$ and applying (2.2.4), we get

$$\begin{aligned}\liminf_{n\to\infty}\int_B \varphi(t)|\mu_n'|(t)\ dt &\geq \liminf_{n\to\infty}\int_B \varphi_k(t)|\mu_n'|(t)\ dt\\ &\geq \int_B \varphi_k(t)|\mu'|(t)dt,\ k\in\mathbb{N}.\end{aligned}$$

If we now let k goes to ∞, we can conclude the proof by a simple application of the monotone convergence theorem. □

Finally, we get a semicontinuity result for general *affine* functionals. Before this, we need the following definition.

Definition 2.2.3. A function $h : I\times X \to \mathbb{R}\cup\{+\infty\}$ is said to be a *Carathéodory integrand* if the following are satisfied:

(i) h is $\mathscr{L}^1\otimes\mathfrak{m}$-measurable;
(ii) $h(t,\cdot)$ is finite and continuous on X, for $\mathscr{L}^1$-a.e. $t\in I$.

Lemma 2.2.4. *Let $a : I\times X\to\mathbb{R}$ and $b : I\times X \to \mathbb{R}^+$ be two Carathéodory integrands. If $p\in[1,+\infty]$, then for every measurable subset $B\subset I$ such that $\mathscr{L}^1(B) > 0$, the functional*

$$\mu \mapsto \int_B \left[a(t,\mu(t)) + b(t,\mu(t))|\mu'|(t)\right]\ dt,\ \mu\in AC^p(I;X), \tag{2.2.5}$$

is sequentially l.s.c on $AC^p(I;X)$, with respect to the weak topology.

Proof. The sequential semicontinuity of the term

$$\mu\mapsto\int_B a(t,\mu(t))\ dt,\ \mu\in AC^p(I;X),$$

is straightforward: indeed, it is just a consequence of Fatou Lemma. For the term

$$\mu\mapsto\int_B b(t,\mu(t))|\mu'|(t)\ dt,\ \mu\in AC^p(I;X),$$

we observe that, taking a weakly convergent sequence $\mu_n \rightharpoonup \mu$, if we set

$$g_n^k(t) = \min\{k, b(t,\mu_n(t))\},\ t\in B,$$

and

$$g^k(t) = \min\{k, b(t,\mu(t))\},\ t\in B,$$

by means of the assumptions on b, we have that $g_n^k \to g^k$ $\mathscr{L}^1$-a.e. on B. Moreover $\{g_n^k\}_{n\in\mathbb{N}}$ is equibounded in $L^\infty(B;\mathbb{R})$: Lebesgue Dominated Convergence Theorem implies that $g_n^k\to g^k$ strongly, let's say in $L^{\frac{p}{p-1}}(B;\mathbb{R})$, while $|\mu_n'|$ weakly ($*$-weakly if $p=+\infty$) converges in $L^p(B;\mathbb{R})$, so that

$$\liminf_{n\to\infty}\int_B g_n^k(t)|\mu_n'|(t)\ dt = \liminf_{n\to\infty}\int_B g^k(t)|\mu_n'|(t)\ dt.$$

This, Lemma 2.2.2 and the positivity of b imply

$$\int_B g^k(t)|\mu'|(t)\ dt \leq \liminf_{n\to\infty}\int_B g^k(t)|\mu_n'|(t)\ dt \leq \liminf_{n\to\infty}\int_B b(t,\mu_n(t))|\mu_n'|\ dt,$$

which gives the thesis, passing to the limit as $k \to \infty$. □

We recall that a metric space is said to be *proper* if its closed balls are compact: in particular, a proper metric space is locally compact (the converse is not true) . As we will see when dealing with absolutely continuous curves over a space which is not proper (Section 4 and 6), it is of interest also the case of a metric space with different topologies defined on it. First of all, we introduce some definitions.

DEFINITION 2.2.5. Let (X, τ) be a topological space and $d : X \times X \to [0, +\infty)$ a metric. We say that d is lower semicontinuous on (X, τ) if the following holds: whenever $x_n \stackrel{\tau}{\to} x$ and $y_n \stackrel{\tau}{\to} y$, then

$$d(x, y) \leq \liminf_{n\to\infty} d(x_n, y_n).$$

DEFINITION 2.2.6. Given a space X with two different metrics d_1 and d_2, we set $X_1 = (X, d_1)$ and $X_2 = (X, d_2)$. We indicate with $|\mu'|_{d_1}$ and $|\mu'|_{d_2}$ the metric derivative with respect to d_1 and d_2, respectively. Then a sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I; X_1)$ is said to be d_2-weakly convergent to μ if:

(i) $\max_{t\in I} d_2(\mu_n(t), \mu(t)) \to 0$;
(ii) the sequence $\{|\mu_n'|_{d_1}\}_{n\in\mathbb{N}}$ is equi-bounded in $L^p(I; \mathbb{R})$ and equi-integrable.

We indicate this convergence by $\mu_n \stackrel{d_2}{\rightharpoonup} \mu$.

REMARK 2.2.7. Observer that $AC^p(I; X_1)$ is closed with respect to the d_2-weak convergence. Indeed if $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I; X_1)$ is such that $\mu_n \stackrel{d_2}{\rightharpoonup} \mu$, then by the semicontinuity of d_1 we get

$$d_1(\mu(s), \mu(t)) \leq \liminf_{n\to\infty} d_1(\mu_n(s), \mu_n(t)) \leq \liminf_{n\to\infty} \int_s^t |\mu_n'|_{d_1}(r)\, dr,$$

so that we can still prove property (2.2.2), that is

$$|\mu'|_{d_1}(t) \leq v(t), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I, \tag{2.2.6}$$

where as above $v \in L^p(I; \mathbb{R})$ is the weak ($*$-weak if $p = +\infty$) limit of $\{|\mu_n'|_{d_1}\}_{n\in\mathbb{N}}$: this precisely means that $\mu \in AC^p(I; X_1)$.

Then we can prove the following slight modification of Lemma 2.2.4.

LEMMA 2.2.8. *Let $X_1 = (X, d_1)$ and $X_2 = (X, d_2)$ be two Polish spaces such that d_1 is lower semicontinuous on X_2.*

Fix $p \in [1, +\infty]$. For every pair of Carathéodory integrands $a : I \times X_2 \to \mathbb{R}$, $b : I \times X_2 \to \mathbb{R}^+$ and every measurable subset $B \subset I$ such that $\mathscr{L}^1(B) > 0$, the functional defined on $AC^p(I; X_1)$ by

$$\mu \mapsto \int_B [a(t, \mu(t)) + b(t, \mu(t))|\mu'|_{d_1}(t)]\, dt,$$

is sequentially l.s.c. with respect to the d_2-weak convergence.

PROOF. Again, the key fact is to show that the functional defined on $AC^p(I;X_1)$ by

$$\mu \mapsto \int_B |\mu'|_{d_1}(t)\ dt,$$

is sequentially l.s.c. with respect to the d_2-weak convergence. At this end, it is sufficient to use (2.2.6): then we can repeat the proof of Lemma 2.2.2 and Lemma 2.2.4 and get the thesis. □

3. Semicontinuous action functionals over $AC^p(I;X)$

We now want to consider a generic action functional defined on $AC^p(I;X)$ of the form

$$\mathcal{A}(\mu) = \int_I f(t,\mu(t),|\mu'|(t))\ dt,\ \ \mu \in AC^p(I;X), \tag{2.3.1}$$

for some function $f : I \times X \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, satisfying the following:

$$f \text{ is } \mathscr{L}^1 \otimes \mathfrak{m} \otimes \mathscr{L}^1\text{-measurable;} \tag{2.3.2}$$

$$f(t,\cdot,\cdot) \text{ is l.s.c. on } X \times \mathbb{R} \text{ for every } t \in I; \tag{2.3.3}$$

$$f(t,x,\cdot) \text{ is convex and increasing on } \mathbb{R} \text{ for every } t \in I, x \in X. \tag{2.3.4}$$

We provide some semicontinuity results for such functionals, with respect to the weak convergence in $AC^p(I;X)$.

REMARK 2.3.1. Let us briefly discuss the monotonicity assumption for the function f: at a first glance, assuming (2.3.4) could seem restrictive. Anyway, if you think to the Euclidean case $X = \mathbb{R}^N$, then $g(z) = f(t,x,z)$ would be a function of the modulus $|z|$, that has to be (if we want to ensure the l.s.c. of functional (2.3.1)) convex in z. Clearly, this is possible if and only if g is convex and increasing (see Figure 1).

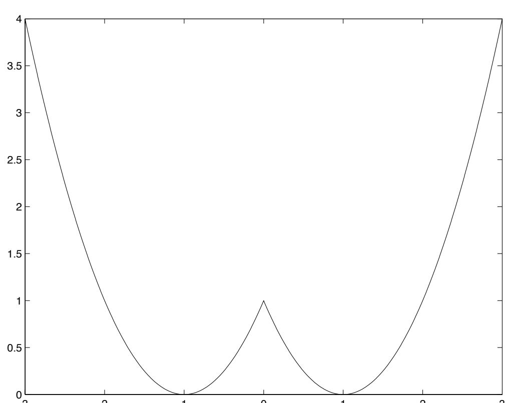

FIGURE 1. A convex function of $|z|$, that is not convex in z.

As usual, the idea is to seek affine approximations of the function f, satisfying (2.3.2), (2.3.3) and (2.3.4): if this can be done, then semicontinuity of $\mathcal{A}$ will result from the application of Lemma 2.2.4.

The following is a crucial result: it is just an adaptation of a classical result, valid in an Euclidean setting (see Lemma 2.2.4 and Remark 2.2.5 of [**29**]).

LEMMA 2.3.2. *Let* $f : I \times X \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *be a function satisfying assumptions* (2.3.2), (2.3.3) *and* (2.3.4). *Assume further that for every* $t \in I$ *the function* $f(t, \cdot, \cdot)$ *satisfies the following condition:*

there exists a function $\theta : \mathbb{R} \to \mathbb{R}$ *such that*

$$\lim_{z\to+\infty} \frac{\theta(z)}{z} = +\infty \quad \textit{and} \quad f(t,x,z) \geq \theta(|z|), \quad \textit{for every } x \in X,\ z \in \mathbb{R}. \tag{2.3.5}$$

Then, there exist two sequences of bounded Carathéodory integrands $a_n : I \times X \to \mathbb{R}$ *and* $b_n : I \times X \to [0, +\infty)$*, such that*

$$f(t,x,z) = \sup_{n\in\mathbb{N}} \{a_n(t,x) + b_n(t,x)z\}, \quad \textit{for every } t \in I,\ x \in X,\ z \in \mathbb{R}.$$

The next general Lemma will be useful in proving our semicontinuity result: the proof can be found in [**29**] (Lemma 2.3.2).

LEMMA 2.3.3. *Let* $\Omega \subset \mathbb{R}^N$ *be any measurable subset and* $g, \{g_n\}_{n\in\mathbb{N}}$ *be measurable functions from* Ω *to* $\mathbb{R} \cup \{+\infty\}$*, such that* $g = \sup \{g_n \ : \ n \in \mathbb{N}\}$ *and* $g_n \geq \varphi$*, for a suitable* $\varphi \in L^1(\Omega; \mathbb{R})$*. Then*

$$\int_\Omega g(x)\, dx = \sup \left\{ \sum_{i\in I} \int_{B_i} g_i(x)\, dx \right\},$$

where the supremum is taken over all finite partitions of Ω*, by pairwise disjoint measurable subsets* B_i*.*

The semicontinuity result now reads as follows:

THEOREM 2.3.4. *Let* $p \in [1, +\infty]$ *and let* $f : I \times X \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *be a function satisfying* (2.3.2), (2.3.3) *and* (2.3.4). *Assume further that there exist two positive constants* α, β*, a point* $\overline{x} \in X$ *and a function* $\varphi \in L^1(I; \mathbb{R})$ *such that* f *satisfies the following estimate*

$$f(t,x,z) \geq -\alpha|z| - \beta d(x, \overline{x})^r - \varphi(t), \tag{2.3.6}$$

for some $r > 0$*. Then the functional*

$$\mathcal{A}(\mu) = \int_I f(t, \mu(t), |\mu'|(t))\, dt, \ \ \mu \in AC^p(I; X), \tag{2.3.7}$$

is well-defined, takes its values in $\mathbb{R} \cup \{+\infty\}$ *and is sequentially l.s.c. on* $AC^p(I;X)$*, with respect to the weak topology.*

PROOF. The fact that the functional $\mathcal{A}$ is well-defined and takes its values in $\mathbb{R}\cup\{+\infty\}$, follows from (2.3.6).

We now proceed to the proof of the sequential lower semicontinuity: let us first assume that f verifies hypothesis (2.3.5) of Lemma 2.3.2, so that we have

$$f(t,x,z) = \sup \{a_n(t,x) + b_n(t,x)z \ : \ n \in \mathbb{N}\},$$

for suitable sequences of bounded Carathéodory integrands $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$, with $b_n \geq 0$. By Lemma 2.3.3, to conclude the proof we can restrict ourselves to prove that for every $n \in \mathbb{N}$ and $B \subset I$ measurable, the functional

$$\mu \mapsto \int_B [a_n(t,\mu(t)) + b_n(t,\mu(t))|\mu'|(t)]\ dt,\ \ \mu \in AC^p(I;X),$$

is sequentially l.s.c. on $AC^p(I;X)$, with respect to the weak topology: this is just a straightforward consequence of Lemma 2.2.4.

We now remove assumption (2.3.5) on f and assume for the moment that $f \geq 0$. Let $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I;X)$ be a weakly convergent sequence: $\{|\mu_n'|\}_{n\in\mathbb{N}}$ is equi-integrable, so there exists a function $\theta : \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{t\to\infty} \frac{\theta(t)}{t} = +\infty \quad \text{and} \quad \sup_{n\in\mathbb{N}} \int_I \theta(|\mu_n'|(t))\ dt \leq 1.$$

For every $\varepsilon > 0$, we set

$$f_\varepsilon(t,x,z) = f(t,x,z) + \varepsilon\theta(|z|),$$

so that f_ε verifies hypothesis (2.3.5) of Lemma 2.3.2 and we can thus obtain

$$\begin{aligned}
\int_I f(t,\mu(t),|\mu'|(t))\ dt &\leq \int f_\varepsilon(t,\mu(t),|\mu'|(t))\ dt \\
&\leq \liminf_{n\to\infty} \int_I f_\varepsilon(t,\mu_n(t),|\mu_n'|(t))\ dt \\
&= \varepsilon \liminf_{n\to\infty} \int_I \theta(|\mu_n'|)(t)\ dt + \liminf_{n\to\infty} \int_I f(t,\mu_n(t),|\mu_n'|(t))\ dt \\
&\leq \varepsilon + \liminf_{n\to\infty} \int_I f(t,\mu_n(t),|\mu_n'|(t))\ dt,
\end{aligned}$$

proving the semicontinuity of $\mathcal{A}$, by the arbitrariness of ε.

Finally, in the general case of a function f satisfying (2.3.6), we proceed as follows: for every $k \in \mathbb{N}$, we define

$$f_k(t,x,z) = \max\{f(t,x,z), -k\},$$

and, taken a weakly convergent sequence $\{\mu_n\} \subset AC^p(I;X)$ such that $\mu_n \rightharpoonup \mu$, we set

$$g_n(t) = \alpha|\mu_n'|(t) + \beta d(\mu_n(t),\overline{x})^r + \varphi(t),\ \ t \in I,$$

$$A_{k,n} = \{t \in I \ : \ f(t,\mu_n(t),|\mu_n'|(t)) < -k\}.$$

We first observe that

$$A_{k,n} \subset \{t \in I \ : \ g_n(t) > k\},$$

and that g_n is equi-integrable, so we obtain

$$\lim_{k\to\infty} \int_{A_{k,n}} g_n(t)\ dt = 0,\ \text{ for every } n \in \mathbb{N}. \tag{2.3.8}$$

Being f_k bounded from below and using (2.3.6), we get

$$\begin{aligned}\int_I f(t,\mu(t),|\mu'|(t))\;dt &\le \int_I f_k(t,\mu(t),|\mu'|(t))\;dt\\ &\le \liminf_{n\to\infty}\int_I f_k(t,\mu_n(t),|\mu_n'|(t))\;dt\\ &= \liminf_{n\to\infty}\left[\int_I f(t,\mu_n(t),|\mu_n'|(t))\;dt - \int_{A_{k,n}} f(t,\mu_n(t),|\mu_n'|(t))\;dt\right]\\ &\le \liminf_{n\to\infty}\int_I f(t,\mu_n(t),|\mu_n'|(t))\;dt + \limsup_{n\to\infty}\int_{A_{k,n}} g_n(t)\;dt,\end{aligned}$$

and this, taking the limit as $k\to\infty$ and taking into account (2.3.8), implies the semicontinuity of $\mathcal{A}$. □

REMARK 2.3.5. In the case $p>1$, we can weaken assumption (2.3.6) of Theorem 2.3.4, by requiring that there exist two positive constants α,β, a point $\overline{x}\in X$ and a function $\varphi\in L^1(I;\mathbb{R})$ such that

$$f(t,x,z)\ge -\alpha|z|^m-\beta d(x,\overline{x})^r-\varphi(t),$$

for $m<p$ and for $r>0$. As in the Euclidean case, we cannot expect any semicontinuity result, if the previous is verified with $m=p>1$ (see [**60**]).

Note that Theorem 2.3.4 can be used to prove lower semicontinuity of *geodesic* functionals, that is functionals of the type

$$\mu\mapsto\int_I g(\mu(t))|\mu'|(t)\;dt,\;\;\mu\in \mathrm{Lip}(I;X)=AC^\infty(I;X),$$

with $g:X\to[0,+\infty]$ lower semicontinuous, which have been studied in detail in the papers [**8**] and [**24**].

Thanks to Theorem 2.3.4, we can also prove the lower semicontinuity of supremal functionals, defined over the space of Lipschitz curves: the proof is just an adaptation of that in [**10**, Theorem 3.4].

THEOREM 2.3.6. *Let $f: I\times X\times\mathbb{R}\to\mathbb{R}\cup+\infty$ be a function satisfying hypothesis* (2.3.2), (2.3.3) *and* (2.3.6). *Suppose moreover that f is level convex with respect to the z variable, that is the sublevel sets*

$$E_\lambda(t,x)=\{z\in\mathbb{R}\;:\;f(t,x,z)\le\lambda\}$$

are convex. Then the functional

$$\mathcal{A}(\mu)=\operatorname*{ess\,sup}_{t\in I} f(t,\mu(t),|\mu'|(t)),\;\;\mu\in AC^\infty(I;X),$$

is sequentially l.s.c. on $AC^\infty(I;X)$, with respect to the weak topology.

PROOF. Let us take a sequence $\{\mu_n\}_{n\in\mathbb{N}}\subset AC^\infty(I;X)$ weakly converging to a curve μ. Then let us set

$$\lambda=\liminf_{n\to\infty}\operatorname*{ess\,sup}_{t\in I} f(t,\mu_n(t),|\mu_n'|(t)),$$

and we choose a subsequence $\{\mu_{n_i}\}_{i\in\mathbb{N}}$ such that

$$\lambda = \lim_{i\to\infty} \operatorname*{ess\,sup}_{t\in I} f(t, \mu_{n_i}(t), |\mu'_{n_i}|(t)),$$

so that for every $\varepsilon > 0$, we can find $i_0 \in \mathbb{N}$ such that for every $i \geq i_0$ we get

$$f(t, \mu_{n_i}(t), |\mu'_{n_i}|(t)) \leq \lambda + \varepsilon, \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{2.3.9}$$

We now consider the integral functional

$$\mathcal{E}(\mu) = \int_I \mathfrak{I}_{E_{\lambda+\varepsilon}}(t, \mu(t), |\mu'|(t))\, dt,\ \mu \in AC^\infty(I;X),$$

where $\mathfrak{I}_{E_{\lambda+\varepsilon}}$ is the indicator function of the set

$$E_{\lambda+\varepsilon} = \{(t,x,z) \in I \times X \times \mathbb{R}\ :\ f(t,x,z) \leq \lambda + \varepsilon\}.$$

Thanks to the assumptions on f, it is easily seen that the function $\mathfrak{I}_{E_{\lambda+\varepsilon}}$ verifies the hypothesis of Theorem 2.3.4, so that $\mathcal{E}$ is l.s.c. on $AC^\infty(I;X)$, with respect to the weak topology:

$$\begin{aligned}\int_I \mathfrak{I}_{E_{\lambda+\varepsilon}}(t, \mu(t), |\mu'|(t))\, dt &\leq \liminf_{n\to\infty} \int_I \mathfrak{I}_{E_{\lambda+\varepsilon}}(t, \mu_n(t), |\mu'_n|(t))\, dt \\ &\leq \lim_{i\to\infty} \int_I \mathfrak{I}_{E_{\lambda+\varepsilon}}(t, \mu_{n_i}(t), |\mu'_{n_i}|(t))\, dt = 0,\end{aligned}$$

where in the last equality we have used (2.3.9). The latter yields

$$f(t, \mu(t), |\mu'|(t)) \leq \lambda + \varepsilon, \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I,$$

which implies, by means of the arbitrariness of ε, that

$$f(t, \mu(t), |\mu'|(t)) \leq \lambda = \liminf_{n\to\infty} \operatorname*{ess\,sup}_{t\in I} f(t, \mu_n(t), |\mu'_n|(t)), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I,$$

thus concluding the proof. □

We conclude this section, giving some refinements of Theorem 2.3.4 that we will need in the sequel.

The first is the following: (X, d_X) and (Y, d_Y) are two Polish spaces and we have an integral functional of the type

$$\mathcal{A}(\nu, \mu) = \int_I f(t, \nu(t), |\mu'|_X(t))\ dt,\ (\nu, \mu) \in L^1(I;Y) \times AC^p(I;X), \tag{2.3.10}$$

where $|\mu'|_X$ stands for the metric derivative of μ, with respect to the metric d_X. It is straightforward to extend the semicontinuity result of Theorem 2.3.4 to this case.

Theorem 2.3.7. *Fix $p \in [1, +\infty]$ and let $f : I \times Y \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a function satisfying hypotheses* (2.3.2), (2.3.3) *and* (2.3.4). *Suppose moreover that there exist two positive constants α, β, a point $\overline{y} \in Y$ and $\varphi \in L^1(I;\mathbb{R})$ such that f satisfies the following estimate*

$$f(t,y,z) \geq -\alpha|z| - \beta d_Y(y, \overline{y}) - \varphi(t). \tag{2.3.11}$$

Then the functional $\mathcal{A} : L^1(I;Y) \times AC^p(I;X) \to \mathbb{R} \cup \{+\infty\}$ defined by (2.3.10) *is well defined and sequentially lower semicontinuous on $L^1(I;Y) \times AC^p(I;X)$, with respect to the strong topology on $L^1(I;Y)$ and the weak topology on $AC^p(I;X)$.*

In the case of a metric space equipped with two different metrics, the following result will be useful: the proof is the same of Theorem 2.3.4, with Lemma 2.2.8 in place of Lemma 2.2.4.

THEOREM 2.3.8. *Let $X_1 = (X, d_1)$ and $X_2 = (X, d_2)$ be two Polish spaces such that d_1 is lower semicontinuous on X_2.*

Fix $p \in [1, +\infty]$. Let $f : I \times X_2 \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a function satisfying (2.3.2), (2.3.3) *and* (2.3.4)*. Assume further that there exist two positive constants α, β, a point $\overline{x} \in X$ and a function $\varphi \in L^1(I; \mathbb{R})$ such that f satisfies the following estimate*

$$f(t, x, z) \geq -\alpha|z| - \beta d_2(x, \overline{x})^r - \varphi(t), \tag{2.3.12}$$

for some $r > 0$. Then the functional

$$\mathcal{A}(\mu) = \int_I f(t, \mu(t), |\mu'|_{d_1}(t))\ dt,\ \ \mu \in AC^p(I; X_1), \tag{2.3.13}$$

is well-defined, takes its values in $\mathbb{R} \cup \{+\infty\}$ and is sequentially l.s.c. on $AC^p(I; X_1)$, with respect to the d_2-weak convergence.

Finally, we can easily obtain a variant of Theorem 2.3.7 for spaces endowed with two metrics: this is motivated by applications to metric spaces which are not proper, the model case being given by the Wasserstein space $\mathcal{W}_\infty$.

THEOREM 2.3.9. *Let $X_1 = (X, d_1)$ and $X_2 = (X, d_2)$ be two Polish spaces such that d_1 is lower semicontinuous on X_2.*

Fix $p \in [1, +\infty]$. Let $f : I \times Y \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a function satisfying hypotheses (2.3.2), (2.3.3) *and* (2.3.4)*. Suppose moreover that there exist two positive constants α, β, a point $\overline{y} \in Y$ and $\varphi \in L^1(I; \mathbb{R})$ such that f satisfies the following estimate*

$$f(t, y, z) \geq -\alpha|z| - \beta d_Y(y, \overline{y}) - \varphi(t). \tag{2.3.14}$$

Then the functional

$$\mathcal{A}(\nu, \mu) = \int_I f(t, \nu(t), |\mu'|_{d_1}(t))\ dt,\ \ (\nu, \mu) \in L^1(I; Y) \times AC^p(I; X_1), \tag{2.3.15}$$

is well-defined, takes its values in $\mathbb{R} \cup \{+\infty\}$ and is sequentially l.s.c. on $L^1(I; Y) \times AC^p(I; X_1)$, with respect to the strong topology on $L^1(I; Y)$ and the d_2-weak topology on $AC^p(I; X_1)$.

4. Minimizing curves

We now turn to the problem of finding a curve minimizing the general cost functional

$$\mathcal{A}(\mu) = \int_I f(t, \mu(t), |\mu'|(t))\ dt, \tag{2.4.1}$$

among all curves $\mu \in AC^p(I; X)$ with fixed endpoints. For every $p \in [1, +\infty]$ and $x_0, x_1 \in X$, we define

$$\mathcal{C}_p(x_0, x_1) = \{\mu \in AC^p(I; X)\ :\ \mu(0) = x_0,\ \mu(T) = x_1\}. \tag{2.4.2}$$

REMARK 2.4.1. In the particular case of $f(t, \mu, |\mu'|) = |\mu'|$, as already observed the problem of minimizing the length functional

$$\ell(\mu) = \int_0^T |\mu'| \, dt,$$

in $\mathcal{C}_p(x_0, x_1)$ admits a solution, which is given by every geodesic in X joining x_0 and x_1, provided that $\mathcal{C}_p(x_0, x_1) \neq \emptyset$ and that X is proper (see [**9**]).

In [**24**] the authors consider the case with $f(t, \mu, |\mu'|) = g(\mu)|\mu'|$: this can now be seen as the problem of finding the geodesics in X, with the respect to some sort of Riemannian distance, whose coefficient is given by g. They prove the following:

THEOREM 2.4.2. *Let X be a proper metric space. If $g : X \to [0, +\infty]$ is a lower semicontinuous function, bounded from below by a constant $c > 0$, we define*

$$\ell_g(\mu) = \int_I g(\mu(t))|\mu'|(t) \, dt, \;\; \mu \in AC^\infty(I; X).$$

Then for every pair of points $x_0, x_1 \in X$, the problem of minimizing ℓ_g in $\mathcal{C}_\infty(x_0, x_1)$ admits a solution, provided that there exists $\overline{\mu} \in \mathcal{C}_\infty(x_0, x_1)$ such that $\ell_g(\overline{\mu})$ is finite.

PROOF. The proof is based on the Reparametrization Lemma A.2.4, the functional considered being invariant under reparametrization. Observe that this clearly also implies that

$$\inf_{\mathcal{C}_p(x_0,x_1)} \int_I g(\mu(t))|\mu'|(t) \, dt = \inf_{\mathcal{C}_\infty(x_0,x_1)} \int_I g(\mu(t))|\mu'|(t) \, dt.$$

Let us take a minimizing sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{C}_\infty(x_0, x_1)$: we can clearly suppose that

$$\sup_{n\in\mathbb{N}} \ell_g(\mu_n) \leq \ell_g(\overline{\mu}) + 1 := C,$$

and moreover up to a reparametrization, we can suppose that each curve is parametrized in such a way that

$$|\mu_n'|(t) \equiv \ell(\mu_n) = \int_I |\mu_n'|(t) \, dt, \;\; t \in I,$$

so that, using the fact that $g \geq c > 0$, we get

$$c \, \ell(\mu_n) \leq \int_I g(\mu_n(t)) \, |\mu_n'|(t) \, dt \leq C,$$

which implies that $\{\mu_n\}_{n\in\mathbb{N}}$ is equi-Lipschitz. Moreover

$$d(\mu_n(t), \mu_0) \leq \int_0^t |\mu_n'|(t) \, dt \leq \ell(\mu_n) \, \mathscr{L}^1(I) \leq C, \quad \text{for every } n \in \mathbb{N}, \; t \in I,$$

so that using the fact that X is proper, we get that $\{\mu_n\}_{n\in\mathbb{N}}$ converges to μ in $AC^\infty(I; X)$, up to a subsequence, by means of Ascoli-Arzelà Theorem A.2.7. Clearly $\mu \in \mathcal{C}_\infty(x_0, x_1)$ and moreover using the semicontinuity of ℓ_g (Theorem 2.3.4), we can conclude that

$$\ell_g(\mu) \leq \liminf_{n\to\infty} \ell_g(\mu_n) = \inf_{\mathcal{C}_\infty(x_0,x_1)} \ell_g,$$

which gives the desired assertion. □

For the case of absolutely continuous curve with finite p-energy, with the general cost functional $\mathcal{A}$ given by (2.4.1), our existence result reads as follows.

THEOREM 2.4.3. *Fix $p \in (1,+\infty)$. Let X be a proper metric space and let $f : I \times X \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a function satisfying* (2.3.2), (2.3.3) *and* (2.3.4). *Assume further that there exist a point $\overline{x} \in X$ and a function $\varphi \in L^1(I;\mathbb{R})$ such that*

$$f(t,x,z) \geq |z|^p - \beta(t)d(x,\overline{x})^r - \varphi(t), \tag{2.4.3}$$

where $0 < r < p$ and $\beta \in L^{\frac{p}{p-r}}(I;\mathbb{R}^+)$. Then for every pair of points $x_0, x_1 \in X$, the problem of minimizing $\mathcal{A}$ in $\mathcal{C}_p(x_0,x_1)$ admits a solution, provided that there exists $\overline{\mu} \in \mathcal{C}_p(x_0,x_1)$ with finite $\mathcal{A}$.

PROOF. Let $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I;X)$ be some minimizing sequence, we can suppose that, up to a subsequence, there exists M such that

$$\mathcal{A}(\mu_n) \leq M, \quad \text{for every } n \in \mathbb{N}.$$

Thanks to the assumptions on f, we immediately obtain that the sequence $\{|\mu_n'|\}_{n\in\mathbb{N}}$ is equi-bounded in $L^p(I;\mathbb{R})$. Indeed, it is enough to use Poincaré-Wirtinger inequality (A.2.3)

$$\int_I d(\mu_n(t),\overline{x})^p\, dt \leq C \int_I |\mu_n'|(t)^p\, dt + A,$$

with A depending only on $\overline{x}$ and the endpoints of μ_n, which are fixed. Then we observe that for every $\varepsilon > 0$, applying Young's inequality, we get

$$\int_I \beta(t)d(\mu_n(t),\overline{x})^r\, dt \leq \left(1-\frac{r}{p}\right)\varepsilon^{\frac{r}{r-p}} \int_I \beta(t)^{\frac{p}{p-r}}\, dt + \frac{r}{p}\varepsilon \int_I d(\mu_n(t),\overline{x})^p\, dt,$$

so that, if we now set

$$\tilde{C}(\varepsilon) = \left(1-\frac{r}{p}\right)\varepsilon^{\frac{r}{r-p}} \int_I \beta(t)^{\frac{p}{p-r}}\, dt + \int_I \varphi(t)\, dt + \frac{r}{p}A\,\varepsilon,$$

then condition (2.4.3) implies

$$M \geq \mathcal{A}(\mu_n) \geq \left(1-\frac{r}{p}C\varepsilon\right) \int_I |\mu_n'|^p(t)\, dt - \widetilde{C}(\varepsilon).$$

With a suitable choice of ε, we obtain the boundedness of $\{|\mu_n'|\}_{n\in\mathbb{N}}$.

This in turn implies that the minimizing sequence is equi-Hölder continuous: in fact by the very definition of absolutely continuous curve and Hölder's inequality, we get

$$\begin{aligned} d(\mu_n(t),\mu_n(s)) &\leq \int_s^t |\mu_n'|(r)\, dr \leq |t-s|^{\frac{p-1}{p}} \left(\int_I |\mu_n'|^p(t)\, dt\right)^{\frac{1}{p}} \\ &\leq C|t-s|^{\frac{p-1}{p}}, \text{ for every } n \in \mathbb{N},\ t,s \in I. \end{aligned}$$

Moreover this sequence is also pointwise relatively compact, because X is proper and there holds

$$d(\mu_n(t),x_0) = d(\mu_n(t),\mu_n(0)) \leq \int_0^t |\mu_n'|(t)\, dt \leq C, \quad \text{for every } n \in \mathbb{N},\ t \in I.$$

We can thus apply Ascoli-Arzelà Theorem (see Appendix A, Theorem A.2.7) to obtain that $\mu_n \rightharpoonup \overline{\mu}$, up to subsequences, where $\overline{\mu} \in \mathcal{C}_p(x_0, x_1)$.

Finally, observe that condition (2.4.3) implies (2.3.6), so that by means of Theorem 2.3.4 the functional $\mathcal{A}$ is l.s.c. on $AC^p(I;X)$, leading us to

$$\mathcal{A}(\overline{\mu}) \leq \liminf_{n\to\infty} \mathcal{A}(\mu_n) = \inf_{\mu\in\mathcal{C}_p(x_0,x_1)} \mathcal{A}(\mu),$$

which concludes the proof. □

REMARK 2.4.4. If condition (2.4.3) is verified with $r = p$, then Theorem 2.4.3 is still valid, provided that the time interval $[0, T]$ is *small enough*, that is we have to guarantee that T is such that

$$C(p, T) < 1,$$

where $C(p, T)$ is the constant given by (A.2.4) in Poincaré-Wirtinger inequality.

The hypothesis that (X, d) is proper can be a very severe one and it could be relaxed somehow, by substituting it with the request that on X there exists another topology τ, such that:

(τ1) there exists a metric $d_\tau \leq d$ which metrizes the topology τ on τ-compact sets;
(τ2) closed balls of (X, d) are τ-compact;
(τ3) d is l.s.c. with respect to τ.

This is a quite standard procedure, which can be also found in [**9**], for example. A typical case in which this occurs is when X is the dual of a separable Banach space, equipped with the norm topology: in this case, τ is just the $*$-weak topology.

The previous considerations lead us to the following result.

THEOREM 2.4.5. *Let $p \in (1, +\infty)$ and $X_1 = (X, d)$ be a Polish space. Suppose that X can be equipped with another topology τ such that $X_2 = (X, \tau)$ satisfies properties (τ1)-(τ3). Let $f : I \times X_2 \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a function satisfying* (2.3.2), (2.3.3) *and* (2.3.4). *Assume further that there exist a point $\overline{x} \in X$ and a function $\varphi \in L^1(I;\mathbb{R})$ such that*

$$f(t, x, z) \geq |z|^p - \beta(t) d_\tau(x, \overline{x})^r - \varphi(t), \tag{2.4.4}$$

where $0 < r < p$ and $\beta \in L^{\frac{p}{p-r}}(I;\mathbb{R}^+)$. Then for every pair of points $x_0, x_1 \in X$, the problem of minimizing

$$\mathcal{A} = \int_I f(t, \mu(t), |\mu'|_d(t))\ dt,$$

in $\mathcal{C}_p(x_0, x_1)$ admits a solution, provided that there exists $\overline{\mu} \in \mathcal{C}_p(x_0, x_1)$ with finite $\mathcal{A}$.

PROOF. Taking a minimizing sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I;X_1)$ and arguing as in the proof of Theorem 2.4.3 (one has to use (2.4.4) in combination with $-d_\tau \geq -d$), we can obtain that $\{|\mu_n'|_d\}_{n\in\mathbb{N}}$ is equi-bounded in $L^p(I;\mathbb{R})$, which in turn implies that for every $n \in \mathbb{N}$ and every $t \in I$ we have

$$\mu_n(t) \in \{x \in X\ :\ d(x, x_0) \leq R\} = B,$$

for a suitable $R > 0$. We now use the fact that (B, d_τ) is a compact metric space and that, due to the fact that $d_\tau \leq d$, we have $\{\mu_n\}_{n\in\mathbb{N}} \in AC^p(I;B) \cap AC^p(I;X_1)$.

Then we apply Ascoli-Arzelà Theorem again: this implies that $\{\mu_n\}_{n\in\mathbb{N}}$ d_τ-weakly converges.

It remains to observe that by Theorem 2.3.8 the functional $\mathcal{A}$ is lower semicontinuous with respect to the d_τ-weak convergence, thus concluding the proof. □

REMARK 2.4.6. We remark that, despite being more general than the case with X proper, Theorem 2.4.5 does not cover some interesting and changeling cases: for example, it does not apply to the case of a functional of the type

$$\mathcal{A}(\mu) = \int_I [|\mu'|^p(t) - \beta(t)d(\mu(t), x_0)^r] \ dt, \tag{2.4.5}$$

because of the fact that, when equipped with the weaker topology τ, the term

$$\mu \mapsto - \int_I \beta(t)d(\mu(t), x_0)^r \ dt,$$

is not τ-l.s.c., due to the fact that d is *only* l.s.c. with respect to this topology and to the presence of the $-$ sign.

A remarkable particular case of (2.4.5) is the following: we choose $(X, d) = (\mathcal{W}_2(\mathbb{R}^N), w_2)$, the 2-Wasserstein metric space and we take the action

$$\mathcal{A}(\mu) = \frac{1}{2} \int_I \left[|\mu_t'|^2_{w_2} - w_2(\mu_t, \nu_0)^2 \right] \ dt, \ \mu \in AC^2(I; \mathcal{W}_2(\mathbb{R}^N)), \tag{2.4.6}$$

for a given reference probability measure $\nu_0 \in \mathcal{W}_2(\mathbb{R}^N)$, for example $\nu_0 = \mathscr{L}^N \llcorner [-1/2, 1/2]^N$

An action like this is considered in the recent paper [**53**] by Gangbo, Nguyen and Tudorascu: the main interest of such a study is that one can write down explicitly an Euler-Lagrange equation for the action $\mathcal{A}$ and discover that this coincides with the so-called Euler-Poisson system (see [**53**, Theorem 3.3]), that is every minimizer $\mu \in \mathcal{C}_2(\mu_0, \mu_1)$ of (2.4.6) satisfies

$$\left\{ \begin{array}{rcll} \partial_t(\mu\, v) + \mathrm{div}\,(\mu\, v \otimes v) & = & \mu\,(\mathrm{barproj}(\gamma) - \mathrm{Id})\,, & \text{in } I \times \mathbb{R}^N, \\ \partial_t \mu + \mathrm{div}\,(\mu\, v) & = & 0, & \text{in } I \times \mathbb{R}^N, \end{array} \right. \tag{2.4.7}$$

in the sense of distributions, where for every $t \in I$ the probability measure $\gamma_t \in \Pi(\nu_0, \mu_t)$ is an optimal plan and $\mathrm{barproj}(\gamma)$ stands for the *barycentric projection* of γ onto its second marginal μ_t, that is disintegrating γ_t as

$$\gamma_t = \int \xi_y \, d\mu_t(y),$$

then $\mathrm{barproj}(\gamma)$ is uniquely defined μ_t−a.e. by

$$\mathrm{barproj}(\gamma_t)(y) = \int_{\mathbb{R}^N} x \, d\xi_y(x), \quad \text{for } \mu_t\text{−a.e. } y \in \mathbb{R}^N.$$

Observe that one can interpret $\mathrm{barproj}(\gamma)$ as a conditional expectation and the rightmost member of the first equation of (2.4.7) as the corresponding momentum. Roughly speaking, for every point y the barycentric projection gives the barycenter of the set where mass located at y is transported: in particular, if the optimal transport is given by a map, then the barycentric projection coincides with this optimal map, that is if $\gamma_t = (\mathrm{Id} \times T_t)_\# \mu_t$, then

$$\mathrm{barproj}(\gamma(t))(y) = T_t(y), \quad \text{for } \mu_t\text{−a.e. } y \in \mathbb{R}^N.$$

This is the case for example when $\mu_t \ll \mathscr{L}^N$ and moreover, as far as the quadratic cost is concerned, we know that $T_t(y) = y - \nabla\psi_t(y)$ with ψ_t semiconcave function (see Chapter 1, Theorem 1.1.8), so that at least formally the first equation in (2.4.7) can be rewritten as

$$\partial_t(\mu\, v) + \mathrm{div}\,(\mu\, v \otimes v) = -\mu\nabla\psi_t.$$

Observe that if one has the right to use the continuity equation to simplify the expression above and to cancel out the common term μ, then one is left with the compressible Euler equation, with pressure term given by the Kantorovich potential ψ_t associated to the transportation from μ_t to ν_0, that is

$$\partial_t v + \langle v, \nabla v\rangle = -\nabla\psi_t.$$

We point out that in the present case, in order to minimize $\mathcal{A}$ given by (2.4.6), neither Theorem 2.4.3 nor Theorem 2.4.5 can be applied, in fact as already observed $\mathcal{W}_2(\mathbb{R}^N)$ is not locally compact, which implies that it is not proper (on the other hand, Theorem 2.4.3 applies if we consider $\mathcal{W}_2(\Omega)$ with $\Omega \subset \mathbb{R}^N$ compact and we add the usual condition $\langle v, \nu\rangle = 0$ at the boundary). Then one can think to equip $\mathcal{W}_2(\mathbb{R}^N)$ with the narrow topology given by the duality with $C_b(\mathbb{R}^N)$: the fact that w_2 is only l.s.c. with respect to this topology, as already observed, implies that the objective functional is no more l.s.c. with respect to this weaker topology, so that this strategy does not work. Indeed, one of the main issues solved by [**53**] is exactly the existence of minimizers of the action (2.4.6), connecting two prescribed measures μ_0 and μ_1, at least in the 1−dimensional case (i.e. $N = 1$) and provided that $T < \pi$: this last condition is clearly connected to the applicability of Poincaré-Wirtinger inequality (see the Appendix, Theorem A.2.3).

5. The case of measures: evolution pairings

We now leave the general setting of metric spaces, particularizing the results of the previous sections to the case of action functionals over the space of probability measures. As in Chapter 1, we will use the notation

$$t \mapsto \mu_t,\ t \in I,$$

to indicate a curve of measures defined over I, rather than the notation $\mu(t)$, used in the previous sections. In particular, we will consider action functionals of the following type (see Theorems 2.3.7 and 2.3.9):

$$\mathcal{A}(\nu,\mu) = \int_I f(t, \nu_t, |\mu_t'|_X)\ dt,\ \ (\nu,\mu) \in L^1(I;Y) \times AC^p(I;X),$$

where X and Y will be suitable spaces of measures which will be made precise in a while.

The main application we have in mind is to provide a dynamical formulation of mass transportation problems, specifically in the context of branched transport: Chapters 3 (specifically, Section 5) and 4 will clarify some of the results contained in these last two sections. Due to this fact, we warn the reader that this section contains some technicalities which can be avoided at a first reading: the main fact is the definition of *evolution pairing* (see Definition 2.5.3).

Let (Ω, d) be a generic locally compact, complete and separable metric space, not necessarily a subset of $\mathbb{R}^N$. From now on, we make the following choice for the two Polish spaces X and Y:

- given $q \in [1,\infty]$, X is the q-Wasserstein metric space $\mathcal{W}_q(\Omega)$;

- Y is the space $\mathcal{M}_1^+(\Omega)$ of positive finite Radon measures over Ω, having total variation less than or equal to 1 and equipped with the distance

$$\mathbf{d}(\nu_1,\nu_2)=\sum_{k=1}^{\infty}\frac{1}{2^k}\left|\int_\Omega \varphi_k(x)\,d(\nu_1(x)-\nu_2(x))\right|,\ \nu_1,\nu_2\in\mathcal{M}_1^+(\Omega), \tag{2.5.1}$$

where $\{\varphi_k\}_{k\in\mathbb{N}}$ is a dense subset of $\{\varphi\in C_o(\Omega)\ :\ \varphi\geq 0,\ \|\varphi\|_\infty\leq 1\}$, and as usual $C_o(\Omega)$ is the completion of the space of compactly supported continuous functions over Ω, with respect to the sup-norm $\|\cdot\|_\infty$.

It is well known that $\mathbf{d}$ metrizes the $*$-weak convergence on the space $\mathcal{M}_1^+(\Omega)$. Moreover $\mathcal{M}_1^+(\Omega)$ is a compact metric space, so that $\mathbf{d}$ is bounded, which means that

$$L^0(I;\mathcal{M}_1^+(\Omega)):=\{\nu:I\to\mathcal{M}_1^+(\Omega)\ :\ \nu \text{ is Borel measurable}\}=L^\infty(I;\mathcal{M}_1^+(\Omega)). \tag{2.5.2}$$

It is clear that by means of Stone-Weierstrass Theorem, we can take the functions φ_k to be Lipschitz in the definition (2.5.1). So, for our purposes it is better to work with the following modified distance

$$\mathfrak{d}(\nu_1,\nu_2)=\sum_{k=1}^{\infty}\frac{1}{2^k\alpha_k}\left|\int_\Omega \varphi_k(x)\,d(\nu_1(x)-\nu_2(x))\right|, \tag{2.5.3}$$

where $\alpha_k=1+\mathrm{Lip}(\varphi_k)$. This distance still metrizes the $*$-weak convergence on $\mathcal{M}_1^+(\Omega)$ and it can be compared with w_q. In fact, we have the following:

LEMMA 2.5.1. *For every $\mu_1,\mu_2\in\mathcal{W}_q(\Omega)$, there holds $\mathfrak{d}(\mu_1,\mu_2)\leq w_q(\mu_1,\mu_2)$.*

PROOF. It is clearly sufficient to prove the thesis in the case $q=1$. We recall the duality formula of Monge's problem, which reads as (see Chapter 1, Proposition 1.1.4)

$$\min_{\gamma\in\Pi(\mu_1,\mu_2)}\int_{\Omega\times\Omega} d(x,y)\ d\gamma(x,y)=\sup_{\varphi\in\mathrm{Lip}_1(\Omega)}\int_\Omega \varphi(x)\ d(\mu_1(x)-\mu_2(x)),$$

where $\mathrm{Lip}_1(\Omega)$ is the space of 1-Lipschitz functions over Ω. Then, for every $\mu_1,\mu_2\in\mathcal{W}_q(\Omega)$ we have

$$\left|\int_\Omega \varphi_k(x)\ d(\mu_1(x)-\mu_2(x))\right|\leq\alpha_k\sup_{\varphi\in\mathrm{Lip}_1(\Omega)}\int_\Omega\varphi(x)\ d(\mu_1(x)-\mu_2(x))=\alpha_k\ w_1(\mu_1,\mu_2),$$

being $\mathrm{Lip}(\varphi_k/\alpha_k)\leq 1$, so that multiplying by $2^{-k}\alpha_k^{-1}$ and summing up, we get

$$\mathfrak{d}(\mu_1,\mu_2)\leq\sum_{k=1}^{\infty}\frac{1}{2^k}w_1(\mu_1,\mu_2)=w_1(\mu_1,\mu_2),$$

proving the assertion. □

REMARK 2.5.2. As a simple consequence of Lemma 2.5.1, we obtain for every $p\in[1,+\infty]$ the inclusion $AC^p(I;\mathcal{W}_q(\Omega))\subset AC^p(I;\mathcal{M}_1^+(\Omega))$, with

$$|\mu_t'|_{\mathfrak{d}}\leq\ |\mu_t'|_{w_q},\ \ \mathscr{L}^1\text{–a.e. } t\in I,\ \mu\in AC^p(I;\mathcal{W}_q(\Omega)). \tag{2.5.4}$$

We now introduce the key concept of *evolution pairing*, which formalizes the idea of associating to every curve of probability measures, a curve which describes the mass that is effectively moving.

DEFINITION 2.5.3. Let $(\nu, \mu) \in L^0(I; \mathcal{M}_1^+(\Omega)) \times AC^p(I; \mathcal{W}_q(\Omega))$ be two curves of measures, such that the following are satisfied:

(E1) $\nu_t \le \mu_t$ in the sense of measures, for $\mathscr{L}^1$-a.e. $t \in I$;

(E2) $\vartheta_t := \mu_t - \nu_t$ is monotone nondecreasing, that is: there exists an $\mathscr{L}^1$-negligible subset $M \subset I$ such that

$$\vartheta_s = \mu_s - \nu_s \le \mu_t - \nu_t = \vartheta_t, \quad \text{for every } s, t \in I \setminus M, \text{ with } s < t;$$

Then we say that (ν, μ) is an *evolution pairing* and we write $\nu \preceq \mu$.

REMARK 2.5.4. We can think of ν as the *moving mass*, while μ is the *total mass*: in this sense, condition (E2) means that the mass that has reached its final destination must increase in time, while (E1) simply states that the moving mass is always less or equal than the total mass. Actually, this really makes sense when the starting measure μ_0 is a Dirac mass[1], so that at time 0 mass starts to move as a whole: on the contrary, when the starting measure is a generic probability, following the same line of reasoning one would have to take into account also the possibility that masses could start to move at different times. In this case, one possibility could be that of defining an evolution pairing as a couple (ν, μ) with the property that $\vartheta = \mu - \nu$ is the sum of an increasing part (the arrived mass) and a decreasing one (the mass which is still not moving), that is a genuinely BV curve. Anyway, in order not to complicate too much the study without a clear scientific gain for the reader, we will not pursue this direction in what follows.

REMARK 2.5.5. The increasing monotonicity of the *arrived mass* ϑ implies the monotonicity of the quantity

$$t \mapsto |\nu_t|(\Omega),$$

while it does not imply that ν has a monotone decreasing (in the sense of measures) behaviour. As an easy counterexample, let us take

$$\sigma_1(t) = (1-t)x_0 + t x_1, \ t \in [0,1],$$

and

$$\sigma_2(t) = \begin{cases} (1-2t)x_0 + 2t x_2, & t \in [0, 1/2], \\ x_2, & t \in [1/2, 1], \end{cases}$$

and consider the curve of probability measures

$$\mu_t = m\delta_{\sigma_1(t)} + (1-m)\delta_{\sigma_2(t)}, \ t \in [0,1],$$

joining $\rho_0 = \delta_{x_0}$ and $\rho_1 = m\delta_{x_1} + (1-m)\delta_{x_2}$, with $x_0, x_1, x_2 \in \mathbb{R}^N$ pairwise distinct and $m \in (0,1)$. If we take

$$\nu_t = \begin{cases} \mu_t, & t \in [0, 1/2], \\ m\delta_{\sigma_1(t)}, & t \in [1/2, 1] \end{cases}$$

then it is easy to verify that $\nu \preceq \mu$, but

$$\nu_{t+h} \perp \nu_t, \ t \in I, \ h > 0.$$

This example should also clarify that in general the curve ν is not continuous.

We exploit the more relevant consequence of evolution pairings in the next Lemma.

[1] In Chapter 4 we will use the results of this and the next section with $\mu_0 = \delta_{x_0}$.

LEMMA 2.5.6. *Let* $(\nu, \mu) \in L^0(I; \mathcal{M}_1^+(\Omega)) \times AC^p(I; \mathcal{W}_q(\Omega))$ *be an evolution pairing. Then* $\nu \in BV(I; \mathcal{M}_1^+(\Omega))$ *and in particular we get*

$$|D\nu|_{\mathfrak{d}}(I) \leq |D\Phi|(I) + \int_I |\mu_t'|_{w_q}\, dt, \tag{2.5.5}$$

where $\Phi : I \to \mathbb{R}^+$ *is the monotone nondecreasing function defined by*

$$\Phi(t) = \mathfrak{d}(\nu_t, \mu_t).$$

PROOF. Before proving the main assertion, we first collect some easy consequences of the definition of evolution pairing:

if the pair (ν, μ) satisfies (E1), this means that

$$\int_\Omega \varphi_k(x)\, d\nu_t(x) \leq \int_\Omega \varphi_k(x)\, d\mu_t(x), \quad \text{for } \mathscr{L}^1\text{-a.e.} t \in I, \ \text{for every } k \in \mathbb{N}.$$

If we now take $\widetilde{\nu} \in \mathcal{M}_1^+(\Omega)$ such that $\widetilde{\nu} \leq \nu_t$ for $\mathscr{L}^1$-a.e. $t \in I$, we obtain

$$0 \leq \int_\Omega \varphi_k(x)\, d(\nu_t(x) - \widetilde{\nu}(x)) \leq \int_\Omega \varphi_k(x)\, d(\mu_t(x) - \widetilde{\nu}(x)),$$

and so, multiplying by $2^{-k}\alpha_k^{-1}$ and summing up, we get

$$\mathfrak{d}(\nu_t, \widetilde{\nu}) \leq \mathfrak{d}(\mu_t, \widetilde{\nu}), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{2.5.6}$$

By hypothesis (E2), we also get

$$0 \leq \int_\Omega \varphi_k(x)\, d(\mu_s(x) - \nu_s(x)) \leq \int_\Omega \varphi_k(x)\, d(\mu_t(x) - \nu_t(x)), \ \text{for every } s < t \in I \setminus M,$$

that is $\Phi(t) = \mathfrak{d}(\nu_t, \mu_t)$ is a real monotone nondecreasing function of a real variable.

To obtain that $\nu \in BV(I; \mathcal{M}_1^+(\Omega))$, it is sufficient to write

$$\nu_t - \nu_s = (\nu_t - \mu_t) - (\nu_s - \mu_s) + (\mu_t - \mu_s), \ s, t \in I \setminus M,$$

and then we use again (E2), so that for every φ_k

$$\left|\int_\Omega \varphi_k(x)\, d(\nu_t(x) - \nu_s(x))\right| \leq \Phi(t) - \Phi(s) + \left|\int_\Omega \varphi_k(x)\, d(\mu_t(x) - \mu_s(x))\right|, \ \text{for every } s < t \in I \setminus M.$$

Again, multiplying by $2^{-k}\alpha_k^{-1}$ and summing up, we get

$$\mathfrak{d}(\nu_t, \nu_s) \leq \Phi(t) - \Phi(s) + \mathfrak{d}(\mu_t, \mu_s), \ s < t \in I \setminus M. \tag{2.5.7}$$

Finally, we observe that

$$|D\Phi|(I) = \lim_{t \to T-} \Phi(t) - \lim_{t \to 0^+} \Phi(t) = \Phi^-(T) - \Phi^+(0),$$

then it follows from (2.5.7) and the definition of essential total variation that

$$\sum_{i=0}^{k} \mathfrak{d}(\nu_{t_i}, \nu_{t_{i+1}}) \le \sum_{i=0}^{k} [\Phi(t_{i+1}) - \Phi(t_i)] + \sum_{i=0}^{k} \mathfrak{d}(\mu_{t_i}, \mu_{t_{i+1}})$$
$$\le \Phi^-(T) - \Phi^+(0) + \int_I |\mu'_t|_{w_q}\, dt,$$

for every finite partitions $0 < t_0 < \cdots < t_{k+1} < 1$ of $I \setminus (M \cup S_\nu)$, proving (2.5.5). □

REMARK 2.5.7. We observe that ν_0^+ and ν_T^- are well defined, thanks to Lemma A.3.3 of the Appendix A. Moreover by the very definition of evolution pairings, we have that if $\nu \preceq \mu$, then $\nu_0^+ \le \mu_0$ and $\nu_T^- \le \mu_T$, in the sense of measures. Indeed, let us prove the first: suppose that there exist $\varphi \in C_o(\Omega; \mathbb{R}^+)$ and $\varepsilon > 0$ such that

$$\int_\Omega \varphi(x)\, d\nu_0^+(x) = \int_\Omega \varphi(x)\, d\mu_0(x) + 4\varepsilon.$$

We can clearly assume that $\|\varphi\|_\infty \le 1$ and we observe that $t \mapsto \int_\Omega \varphi(x)\, d\mu_t(x)$ is a uniformly continuous real function of one variable: then there exists $r_0 < T$ such that

$$\int_\Omega \varphi(x)\, d\mu_t(x) < \int_\Omega \varphi(x)\ d\mu_0(x) + \varepsilon,\ t \in (0, r_0),$$

which implies

$$\int_\Omega \varphi(x)\, d\nu_t(x) \le \int_\Omega \varphi(x)\, d\mu_t(x) < \int_\Omega \varphi(x)\, d\nu_0^+(x) - 3\varepsilon, \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, r_0).$$

Then

$$3\varepsilon < \int_\Omega \varphi(x)\, d\nu_0^+(x) - \int_\Omega \varphi(x)\, d\nu_t(x) = \int_\Omega \varphi(x)\, d(\nu_0^+(x) - \nu_t(x)), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, r_0),$$

and if φ_m is such that $\|\varphi - \varphi_m\|_\infty < \varepsilon$, the previous yields

$$\varepsilon(3 - |\nu_0^+ - \nu_t|(\Omega)) \le \int_\Omega \varphi_m(x)\, d(\nu_0^+(x) - \nu_t(x)), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, r_0).$$

It is enough to observe that $\nu_0^+ - \nu_t$ is a signed Radon measure, with total variation less than or equal to 2, so that we simply obtain

$$\varepsilon \le \int_\Omega \varphi_m(x)\, d(\nu_0^+(x) - \nu_t(x)), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, r_0),$$

and multiplying the terms on both sides by $c = 2^{-m}\alpha_m^{-1}$, we have

$$c\,\varepsilon \le \mathfrak{d}(\nu_t, \nu_0^+), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, r_0).$$

So, if we denote

$$X_\varepsilon^+(0) = \{t > 0\ :\ \mathfrak{d}(\nu_t, \nu_0^+) > c\,\varepsilon\},$$

we have proven (up to an $\mathscr{L}^1$-negligible set) the inclusion $(0, r_0) \subset X_\varepsilon^+(0)$: this in turn contradicts the fact that, by definition of ν_0^+, the set $X_\varepsilon^+(0)$ must have 0-density.

The fact that $\nu_T^- \le \mu_T$ can be proven in the same way.

As already observed (see Proposition 1.3.8), the space $\mathcal{W}_q(\Omega)$ is not locally compact, which in particular means that it is not proper. However, this is not a great trouble, as far as we can endow it with the weaker topology given by $\mathfrak{d}$ and conditions (τ_1)-(τ_3) of Section 4 are satisfied. This is the content of the next Lemma.

LEMMA 2.5.8. *The distance w_q is $\mathfrak{d}$-lower semicontinuous. Moreover all bounded sets in $\mathcal{W}_q(\Omega)$ are $\mathfrak{d}$-relatively compact.*

PROOF. The proof is the same as in [**24**, Lemmas 4.2 and 4.3], the only difference being the fact that $\mathfrak{d}$ metrizes the $*$-weak convergence, instead of the narrow convergence. Anyway, having assumed that Ω is locally compact, we have that, at the level of probability measures, $*$-weak and narrow convergence are actually equivalent (see Chapter 1, Lemma 1.3.7): let us take $\{\mu_n^1\}_{n\in\mathbb{N}}, \{\mu_n^2\}_{n\in\mathbb{N}} \subset \mathcal{W}_q(\Omega)$ such that

$$\mu_n^i \overset{*}{\rightharpoonup} \mu^i \in \mathcal{W}_q(\Omega),\ i = 1, 2.$$

The two sequences are equi-tight (here we use the equivalence between $*$-weak and narrow convergence), so that if for every $n \in \mathbb{N}$ we take $\gamma_n \in \Pi(\mu_n^1, \mu_n^2)$ to be an optimal transport plan, that is

$$w_q(\mu_n^1, \mu_n^2) = \|d(\cdot,\cdot)\|_{(L^q(\Omega\times\Omega);\gamma_n)},$$

then the equi-tightness of the marginals, implies that of $\{\gamma_n\}_{n\in\mathbb{N}} \subset \mathscr{P}(\Omega\times\Omega)$. Thus by Prokhorov's Theorem we have that, up to a subsequence, γ_n narrowly converges to γ and clearly $\gamma \in \Pi(\mu_1, \mu_2)$. This yields

$$\begin{aligned} w_q(\mu^1, \mu^2) \leq \|d(\cdot,\cdot)\|_{(L^q(\Omega\times\Omega);\gamma)} &\leq \liminf_{n\to\infty} \|d(\cdot,\cdot)\|_{(L^q(\Omega\times\Omega);\gamma_n)} \\ &= \liminf_{n\to\infty} w_q(\mu_n^1, \mu_n^2), \end{aligned}$$

proving the first statement.

For the second statement, let us take $x_0 \in \Omega$: we observe that setting

$$B_R(\delta_{x_0}) = \{\mu \in \mathcal{W}_q(\Omega)\ :\ w_q(\mu, \delta_{x_0}) < R\},$$

then every $\{\mu_n\}_{n\in\mathbb{N}} \subset B_R(\delta_{x_0})$ is equi-tight, thanks to Markov's Inequality: using again Prokhorov's Theorem, we get that $\mu_n \overset{*}{\rightharpoonup} \mu$ (up to subsequences). It remains to observe that μ has finite q-momentum: this is just a consequence of the lower semicontinuity of the functional

$$\mu \mapsto w_q(\mu, \delta_{x_0}),$$

thus concluding the proof. □

6. Minimizing evolution pairings

Let $p \in [1, +\infty]$, for every pair $\rho_0, \rho_1 \in \mathcal{W}_q(\Omega)$, we define the following subset of $L^0(I; \mathcal{M}_1^+(\Omega)) \times AC^p(I; \mathcal{W}_q(\Omega))$:

$$EP_{p,q}(\rho_0, \rho_1) = \{\nu \preceq \mu\ :\ \mu_0 = \rho_0,\ \mu_T = \rho_1\}.$$

We are interested in the existence of an evolution pairing (ν, μ) minimizing

$$\overline{\mathcal{A}}(\nu, \mu) = \int_I f(t, \nu_t, |\mu_t'|_{w_q})\ dt, \tag{2.6.1}$$

over the set $EP_{p,q}(\mu_0, \mu_1)$: to this aim, we have to prove that the latter set is closed, with respect to some reasonable topology.

LEMMA 2.6.1. *Let* $\{(\nu^n, \mu^n)\} \subset EP_{p,q}(\rho_0, \rho_1)$ *be such that* $\nu^n \to \nu$ $\mathscr{L}^1$*-a.e. and* $\mu^n \xrightarrow{\mathfrak{d}} \mu$*, then*

$$(\nu, \mu) \in EP_{p,q}(\rho_0, \rho_1).$$

PROOF. We first show that (ν, μ) is an evolution pairing: for every $k \in \mathbb{N}$ we have

$$\begin{aligned}\int_\Omega \varphi_k(x)\, d\nu_t(x) &= \lim_{n\to\infty} \int_\Omega \varphi_k(x)\, d\nu_t^n(x) \leq \lim_{n\to\infty} \int_\Omega \varphi_k(x)\, d\mu_t^n \\ &= \int_\Omega \varphi_k(x)\, d\mu_t(x), \text{ for } \mathscr{L}^1\text{-a.e. } t \in I,\end{aligned}$$

so (ν, μ) verifies (E1).

Then let $M_n \subset I$ be the $\mathscr{L}^1$-negligible set corresponding to ν_n in (E2) and define $M = \bigcup_{n\in\mathbb{N}} M_n$: this is still an $\mathscr{L}^1$-negligible subset of I, on which we have

$$\begin{aligned}\int_\Omega \varphi_k(x)\, d(\mu_s(x) - \nu_s(x)) &= \lim_{n\to\infty} \int_\Omega \varphi_k(x)\, d(\mu_s^n(x) - \nu_s^n(x)) \\ &\leq \lim_{n\to\infty} \int_\Omega \varphi_k(x)\, d(\mu_t^n(x) - \nu_t^n(x)) \\ &= \int_\Omega \varphi_k(x)\, d(\mu_t(x) - \nu_t(x)), \text{ for every } s, t \in I \setminus M, \text{ such that } s < t,\end{aligned}$$

proving property (E2).

It remains to show that $\mu \in AC^p(I; \mathcal{W}_q(\Omega))$ and that it still verifies the conditions on the endpoints: the first is just a consequence of the fact that w_q is $\mathfrak{d}$-l.s.c., while the second straightforwardly follows from the uniform convergence, together with the fact that $\mu_0^n = \rho_0$ and $\mu_T^n = \rho_1$, for every $n \in \mathbb{N}$. □

We are in a position to obtain the existence of a minimal evolution pairing, under the usual appropriate growth conditions on the integrand f.

THEOREM 2.6.2. *Fix* $p \in (1, +\infty)$*. Let* $f : I \times \mathcal{M}_+^1(\Omega) \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *be a function satisfying hypotheses* (2.3.2), (2.3.3) *and* (2.3.4)*. Assume further that there exist a measure* $\overline{\nu} \in \mathcal{M}_1^+(\Omega)$ *and a summable function* h *such that*

$$f(t, \nu, z) \geq |z|^p - \beta(t)\mathfrak{d}(\nu, \overline{\nu})^r - h(t), \tag{2.6.2}$$

where $0 < r < p$ *and* $\beta \in L^{\frac{p}{p-r}}(I; \mathbb{R}^+)$*. Then for every pair* $\rho_0, \rho_1 \in \mathcal{W}_q(\Omega)$*, the minimization problem*

$$\inf_{(\nu,\mu)\in EP_{p,q}(\mu_0,\mu_1)} \overline{\mathcal{A}}(\nu, \mu),$$

admits a solution, provided there exists $(\overline{\nu}, \overline{\mu}) \in EP_{p,q}(\rho_0, \rho_1)$ *with finite* $\overline{\mathcal{A}}$*, where* $\overline{\mathcal{A}}$ *is defined by* (2.6.1).

PROOF. Take a minimizing sequence $\{(\nu^n,\mu^n)\}_{n\in\mathbb{N}}\subset EP_{p,q}(\rho_0,\rho_1)$ and suppose that

$$\overline{\mathcal{A}}(\nu^n,\mu^n)\le L,\quad \text{for every } n\in\mathbb{N}.$$

We consider on $\mathcal{W}_q(\Omega)$ the weaker topology given by $\mathfrak{d}$: then we can repeat the same arguments of Theorem 2.4.5, in combination with Lemma 2.5.8, to get the $\mathfrak{d}$-weak convergence in $AC^p(I;\mathcal{W}_q(\Omega))$ (up to a subsequence) of $\{\mu^n\}_{n\in\mathbb{N}}$ to $\widehat{\mu}\in AC^p(I;\mathcal{W}_q(\Omega))$.

In order to get the convergence of $\{\nu^n\}_{n\in\mathbb{N}}$, we want to use Theorem A.3.5: indeed, it is trivially true that

$$\sup_{n\in\mathbb{N}}\int_I \mathfrak{d}(\nu^n_t,0)\,dt<+\infty.$$

If we want to obtain a bound on the total variations, we can simply use the fact that every (ν^n,μ^n) is an evolution pairing: if we indicate

$$\Phi_n(t)=\mathfrak{d}(\nu^n_t,\mu^n_t),\ t\in I,\ n\in\mathbb{N},$$

we have already seen that these are monotone increasing functions. Moreover they are equi-bounded, because of the boundedness of $\mathfrak{d}$.

This, together with Lemma 2.5.6, implies that $\{\nu^n\}_{n\in\mathbb{N}}\in BV(I;\mathcal{M}^+_1(\Omega))$, with a uniform bound on the total variations. Indeed we have

$$\sup_{n\in\mathbb{N}}|D\nu^n|_{\mathfrak{d}}(I)\le\sup_{n\in\mathbb{N}}(\Phi^-_n(T)-\Phi^+_n(0))+\sup_{n\in\mathbb{N}}\int_I|(\mu^n_t)'|_{w_q}\,dt<+\infty,$$

where we have used that

$$\sup_{n\in\mathbb{N}}\int_I|(\mu^n_t)'|_{w_q}\ dt<+\infty,$$

by the first part of the proof.

So that we can apply Theorem A.3.5, obtaining the convergence of $\{\nu^n\}_{n\in\mathbb{N}}$ in $L^1(I;\mathcal{M}^+_1(\Omega))$ (up to a subsequence) to a curve $\widehat{\nu}\in BV(I;\mathcal{M}^+_1(\Omega))$.

It only remains to observe that, by Lemma 2.6.1 we have $(\widehat{\nu},\widehat{\mu})\in EP_{p,q}(\rho_0,\rho_1)$, while by Theorem 2.3.9 the functional $\overline{\mathcal{A}}$ is lower semicontinuous, so that

$$\overline{\mathcal{A}}(\widehat{\nu},\widehat{\mu})\le\liminf_{n\to+\infty}\overline{\mathcal{A}}(\nu^n,\mu^n)=\min_{(\nu,\mu)\in EP_{p,q}(\rho_0,\rho_1)}\overline{\mathcal{A}}(\nu,\mu),$$

concluding the proof. □

Regarding the Lipschitz case (that is $p=+\infty$), we can prove the analogous of Theorem 2.4.2: namely, we have the existence of an evolution pairing minimizing a *geodesic* functional

$$\tilde{\ell}_g(\nu,\mu)=\int_I g(\nu_t)\,|\mu'_t|_{w_q}\,dt. \tag{2.6.3}$$

THEOREM 2.6.3. *Suppose that $g:\mathcal{M}^+_1(\Omega)\to[0,+\infty]$ is lower semicontinuous and bounded from below by a positive constant $c>0$. Then for every $\rho_0,\rho_1\in\mathcal{W}_q(\Omega)$, the problem*

$$\inf_{(\nu,\mu)\in EP_{\infty,q}(\rho_0,\rho_1)}\widetilde{\ell}_g(\nu,\mu), \tag{2.6.4}$$

admits a solution, provided that there exists an evolution pairing in $EP_{\infty,q}(\rho_0,\rho_1)$, with finite energy.

PROOF. Indeed, taking a minimizing sequence $\{(\nu^n,\mu^n)\}_{n\in\mathbb{N}}$, it should be clear that it is sufficient to obtain the convergence of $\{\mu^n\}_{n\in\mathbb{N}}$: then one can argue as in Theorem 2.6.2.

We suppose that the value (2.6.4) is finite, otherwise the result is trivial, so that it is not restrictive to to suppose that

$$\tilde{\ell}_g(\nu^n,\mu^n)\le C.$$

The functional under consideration is invariant by reparametrization and moreover we observe that if $\nu\preceq\mu$ and $\widetilde{\mu}=\mu\circ\mathfrak{t}$ is a reparametrization of μ, then $\nu\circ\mathfrak{t}=\widetilde{\nu}\preceq\widetilde{\mu}$. So up to reparametrization, we can suppose that

$$|(\mu_t^n)'|_{w_q}\equiv L_n,$$

then

$$cL_n=c\int_I|(\mu_t^n)'|_{w_q}\,dt\le\tilde{\ell}_g(\nu^n,\mu^n)\le C,$$

giving that $\{\mu^n\}_{n\in\mathbb{N}}$ is equi-Lipschitz (with respect to w_q) and

$$\mu_t^n\in\{\mu\in\mathcal{W}_q(\Omega)\,:\,w_q(\mu,\rho_0)\le R\},\ n\in\mathbb{N},\ t\in I,$$

for a suitable $R>0$. This implies the $\mathfrak{d}$-weak convergence of the sequence, with the same line of reasoning of Theorem 2.4.5. Then one can conclude by applying the semicontinuity result of Theorem 2.3.9. $\square$

CHAPTER 3

Branched transportation problems

1. Introduction

With this chapter, we begin the study of *branched transportation problems*, that is transportation problems in which the total cost is lowered by grouping the mass during the movement. As already mentioned in the Plan of the Work, this *energy saving requirement* is usually encoded taking into account an infinitesimal cost of the type

$$\varphi(m)\,\ell,$$

for a mass m moving on a distance ℓ, with φ being a given increasing subadditive function: indeed, thanks to the subadditivity property we have $\varphi(m_1 + m_2) \leq \varphi(m_1) + \varphi(m_2)$, so that it could be less expensive to put different masses together. In this way branching effects arise and the typical optimal configurations are tree-shaped structures. The usual choice for φ is given by $\varphi(m) = m^\alpha$, with $\alpha \in [0, 1]$: observe that in the case $\alpha = 1$ we have a linear dependence on the mass (so it is *concentration indifferent*), corresponding to the infinitesimal cost of the usual Monge-Kantorovich problem. On the other hand, when $\alpha = 0$, branching effects are so strong that one only looks at the minimization of the total length of the transportation structure: this corresponds to a minimal connection (or Steiner) problem.

Nature offers a wide variety of systems which can be seen as solving a branched transportation problem, just think to root systems in a tree or to blood vessels in a human body. Human beings as well have learnt that usually it is better to create optimal transportation networks, in order to distribute some goods in an optimal way from a source to a destination: drainage network systems, telephone cables, electric wires and so on are simple examples of this fact.

From this, the interest in studying variational models giving rise to branched structures of transportation as optimizers, a topic which has been the object of an intensive investigation in recent times. In this chapter we try to offer a presentation of some of these models:

- the *transport paths* one (Section 2) by Xia ([**90**]);
- the Lagrangian models (Section 3), which comprise the *irrigation patterns model* by Maddalena, Morel and Solimini ([**67**]), the *traffic plans model* by Bernot, Caselles and Morel ([**15**]) and the *synchronized traffic plans model* by Bernot and Figalli ([**18**]);
- the *path functional* one (Section 4), which is the one introduced by Brancolini, Buttazzo and Santambrogio (see [**24**]).

Particular emphasis will be put on the latter, the path functional model, which is the less studied: as we will see in Section 5, this is not equivalent to the others, as far as it provides optimal structures which are qualitatively different, indeed it describes a slightly different kind of energy. We will then try to understand the main reasons for this failed equivalence and see how it is possible to modify the path functional model, in order to let it describe an energy equivalent to a Gilbert-Steiner one: then this issue will be addressed in the next chapter, where a fundamental role will be played by the concept of evolution pairing (see Section 5 of the previous chapter) and by action functionals of the form

$$\int_I f(t, \nu_t, |\mu_t'|)\, dt,$$

which has been treated in great details in the previous chapter. In what follows, with Ω we will always indicate a compact convex subset of $\mathbb{R}^N$, having non empty interior.

2. Xia's transport paths model

The interesting feature of this model is that it falls into the class of divergence-constrained optimization problems. To start with, let us consider the discrete case, that is the case of finitely atomic sources and destinations: let ρ_0 and ρ_1 be two probability measures given by

$$\rho_0 = \sum_{k=1}^{m} a_k\, \delta_{x_k} \quad \text{and} \quad \rho_1 = \sum_{j=1}^{n} b_j\, \delta_{x_j},$$

and consider a *weighted oriented graph* $\mathfrak{g}$, consisting of:

- a set $\{v_s\}_{s\in V}$ of vertices of the graph, which comprise the sources $\{x_1, \dots, x_m\}$ and sinks $\{y_1, \dots, y_n\}$;
- a set $\{e_h\}_{h\in J}$ of edges of the graph;
- a set $\{\overrightarrow{\tau_h}\}_{h\in J}$ of orientations for the edges of the graph, that is each $\overrightarrow{\tau_h}$ is the orientation of the corresponding edge e_h;
- a set $\{m_h\}$ of weights defined on the edges of the graph, each m_h standing for the mass transiting on the edge e_h (or the capacity of e_h).

Observe that such a graph $\mathfrak{g}$ encodes all the informations on the transportation structure and it is said to be a *transport path* between ρ_0 and ρ_1 if the following balance conditions are satisfied:

(i) for each source vertex x_k, with $k = 1, \dots, m$, we have

$$a_k = \sum_{e_h^- = x_k} m_h,$$

where e_h^- stands for the starting point of the corresponding edge e_h (remember that the edges are oriented);

(ii) for each sinks vertex y_j, with $j = 1, \dots, n$, we have

$$b_j = \sum_{e_h^+ = y_j} m_h,$$

where e_h^+ stands for the ending point of the edge e_h;

(iii) for any interior vertex v_s, we have (*Kirchhoff's Law*)

$$\sum_{e_h^- = v} m_h = \sum_{e_h^+ = v} m_h,$$

that is at every interior bifurcation point, the total incoming mass must equal the total outcoming one (which expresses the conservation of mass).

Then given ρ_0 and ρ_1 as before, we define the set

$$\mathcal{T}(\rho_0, \rho_1) = \{\mathfrak{g} \text{ transport path between } \rho_0 \text{ and } \rho_1\},$$

and associate to every $\mathfrak{g} \in \mathcal{T}(\rho_0, \rho_1)$ the total cost given by

$$M_\alpha(\mathfrak{g}) = \sum_{h\in J} m_h^\alpha \mathscr{H}^1(e_h), \tag{3.2.1}$$

where $\mathscr{H}^1$ is the one-dimensional Hausdorff measure and $\alpha \in [0,1]$: observe that this is an energy of the form

$$\sum (\text{mass})^\alpha \times \text{length},$$

as discussed in the Introduction, with the subadditive power α favouring the joining of masses. The related optimization problem then reads as

$$\min_{\mathfrak{g}\in T(\rho_0,\rho_1)} M_\alpha(\mathfrak{g}),$$

and it can be viewed as a generalization of Steiner's problem of finding the network of minimal length connecting a set of given points, the latter clearly corresponding to the choice $\alpha = 0$. On the other hand, in the other extremal case, namely when $\alpha = 1$, the cost is linear with respect to the mass and there is no gain in joining masses: in this case, we are simply facing the Monge-Kantorovich problem with cost given by the distance and so

$$\min_{\mathfrak{g}\in T(\rho_0,\rho_1)} M_1(\mathfrak{g}) = w_1(\rho_0, \rho_1).$$

This discrete model, firstly introduced by Gilbert in the '60s ([**56**]) in order to provide a mathematical model for the transportation of signals along telephone cables, has been suitably extended to the case of ρ_0 and ρ_1 generic probability measures by Xia in [**90**], thanks to a relaxation procedure: the idea is to reformulate Gilbert's model in the language of vector measures (or currents, as in [**73**]), so that every weighted oriented graph $\mathfrak{g} \in \mathcal{T}(\rho_0, \rho_1)$ can be viewed as the measure

$$\phi_{\mathfrak{g}} := \sum_h m_h \overrightarrow{\tau_h} \mathscr{H}^1 \llcorner e_h,$$

that is

$$\int_\Omega \varphi(x) \cdot d\phi_{\mathfrak{g}}(x) = \sum_{h\in J} m_h \int_{e_h} \langle \varphi(x), \overrightarrow{\tau_h}(x)\rangle \, d\mathscr{H}^1(x), \quad \text{for every } \varphi \in C^0(\Omega; \mathbb{R}^N),$$

and the balance conditions (i)-(iii) simply rewrite as a constraint on the divergence of this measure, that is

$$\operatorname{div} \phi_{\mathfrak{g}} = \rho_0 - \rho_1, \quad \text{in } \Omega, \ \langle \phi_{\mathfrak{g}}, \nu\rangle = 0, \quad \text{on } \partial\Omega,$$

in the usual sense of distributions, that is

$$\int_\Omega \nabla\varphi(x) \cdot d\phi_{\mathfrak{g}}(x) = \int_\Omega \varphi(x) \, d(\rho_1(x) - \rho_0(x)), \quad \text{for every } \varphi \in C^1(\Omega).$$

In this general case, given $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, a vector measure ϕ is said to be a *transport path* between ρ_0 and ρ_1 if there exists a sequence of triples $\{(\rho_0^n, \rho_1^n, \mathfrak{g}^n)\}_{n\in\mathbb{N}}$ with ρ_0^n, ρ_1^n finitely atomic probability measures and $\mathfrak{g}^n \in \mathcal{T}(\rho_0^n, \rho_1^n)$, such that

$$\rho_i^n \rightharpoonup \rho_i,\ i = 0, 1 \quad \text{and} \quad \phi_{\mathfrak{g}^n} \rightharpoonup \phi.$$

A sequence of triples $\{(\rho_0^n, \rho_1^n, \mathfrak{g}^n)\}_{n\in\mathbb{N}}$ like this is called an *approximating graph sequence* (a. g. s., for short) for ϕ: observe in particular that such a vector field ϕ verifies $\mathrm{div}\phi = \rho_0 - \rho_1$, still in the sense of distributions (with $\langle \phi, \nu \rangle = 0$ on $\partial\Omega$). We still denote with $\mathcal{T}(\rho_0, \rho_1)$ the set of transport paths between ρ_0 and ρ_1 and we define the energy of $\phi \in \mathcal{T}(\rho_0, \rho_1)$ as

$$M_\alpha^*(\phi) = \inf\left\{\liminf_{n\to\infty} M_\alpha(\mathfrak{g}^n)\ :\ \{\rho_0^n, \rho_1^n, \mathfrak{g}^n\} \text{ is an a. g. s. for } \phi\right\}.$$

Moreover this relaxed energy admits the following integral representation (see [**89**, Proposition 4.4])

$$M_\alpha^*(\phi) = \begin{cases} \int_\Sigma m(x)^\alpha \, d\mathscr{H}^1(x), & \text{if } \phi = m \overrightarrow{\tau} \mathscr{H}^1 \llcorner \Sigma, \\ +\infty, & \text{otherwise,} \end{cases} \tag{3.2.2}$$

which is finite only on those vector measures concentrated on a $1-$rectifiable set Σ with a vector density $m\overrightarrow{\tau}$ with respect to $\mathscr{H}^1$, where $\overrightarrow{\tau}$ is an orientation of Σ, which means that the measurable vector field $\overrightarrow{\tau} : \Sigma \to S^{N-1}$ belongs to the approximate tangent space to Σ, for $\mathscr{H}^1-$a.e. point. Obviously, this energy is closely related to the original Gilbert-Steiner one (3.2.1).

In the following, for every $\alpha \in [0, 1]$ and $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ we will set

$$d_\alpha(\rho_0, \rho_1) = \min\{M_\alpha^*(\phi)\ :\ \phi \in \mathcal{T}(\rho_0, \rho_1)\}, \tag{3.2.3}$$

and observe that this minimum value (provided it exists) is unchanged, if we enlarge the class of admissible vector measures to comprise directly those with prescribed divergence $\rho_0 - \rho_1$.

We collect some important results on this model in the next Theorem (see [**90**, Theorems 3.1, 4.2 and 5.1]).

Theorem 3.2.1 (Xia). *Let $\alpha \in (1 - 1/N, 1]$ and $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, then the minimization problem defining $d_\alpha(\rho_0, \rho_1)$ does admit a solution with finite energy. Moreover $d_\alpha : \mathscr{P}(\Omega) \times \mathscr{P}(\Omega) \to [0, \infty)$ defines a distance on $\mathscr{P}(\Omega)$ which metrizes the $*-$weak convergence and such that $(\mathscr{P}(\Omega), d_\alpha)$ is a geodesic space.*

3. Lagrangian models

With this name, we refer to various models due to different authors (Bernot, Caselles, Figalli, Maddalena, Morel, Solimini), whose common root is a Lagrangian description of branched transportation, achieved through the employing of a functional defined on probability measures over the space of Lipschitz paths. This is not surprising, as far as probability measures over the set of paths have shown up to be a very powerful tool in Fluid Mechanics, like in the Lagrangian formulation of incompressible Euler equations (see [**26**]). Usually, in these models the energy associated to the transportation is almost the same, apart for the definition of *multiplicity* (see below) which, in some sense, is characteristic of each model.

We start recalling that the Monge-Kantorovich problem has the following equivalent Lagrangian formulation

$$\min_{Q\in TP(\rho_0,\rho_1)} \int_{C([0,1];\Omega)} \ell(\sigma)\, dQ(\sigma),$$

where

$$TP(\rho_0,\rho_1) = \{Q \in \mathscr{P}((C[0,1];\Omega)) \,:\, Q \text{ is concentrated on } \mathrm{Lip}([0,1];\Omega),\ (e_i)_\sharp Q = \rho_i,\ i = 0,1\},$$

any optimizer being given by $Q = \int_{\Omega\times\Omega} \delta_{\overline{xy}}\, d\gamma(x,y)$, with $\gamma \in \Pi(\rho_0,\rho_1)$ optimal transport plan for the cost $c(x,y) = |x-y|$ and $\overline{xy}$ a parametrization on $[0,1]$ of the segment joining x to y. More generally, one can consider conformal perturbations of the Euclidean metric: defining the distance

$$d_g(x,y) = \inf_{\sigma\in C^{x,y}} \int_0^1 g(\sigma(t))\, |\sigma'(t)|\, dt,$$

with $C^{x,y} = \{\sigma \,:\, \sigma(0) = x,\ \sigma(1) = y\}$ and $g : \Omega \to [0,+\infty]$ lower semicontinuous and bounded from below by $c > 0$, we see that the Monge-Kantorovich problem for this cost d_g has the equivalent Lagrangian description

$$\min_{TP(\rho_0,\rho_1)} \int_{C([0,1];\Omega)} \left(\int_0^1 g(\sigma(t))\, |\sigma'(t)|\, dt\right) dQ(\sigma),$$

any optimal Q now being of the form $Q = \int_{\Omega\times\Omega} Q^{x,y}\, d\gamma(x,y)$, with $Q^{x,y}$ probability measure concentrated on the set of geodesics (with respect to d_g) joining x to y.

Anyway, in this description interaction effects between particles are not taken care of, as far as the quantity to be minimized only involves the lengths of the paths followed by particles. In order to keep track of these effects, the idea is to modify the latter case, the one corresponding to d_g, taking the weight function g to be a function of the mass transiting from $\sigma(t)$ and not only of the position occupied by the particle corresponding to σ at time t. This implies that now the energy mixes masses and lengths, so that it could happen that particles act cooperatively (joining or separating) in order to optimize the total energy. We then arrive to the concept of *multiplicity* which formalizes this idea of *transiting mass*: for example, the multiplicity of the model introduced in [**15**] by Bernot, Caselles and Morel is given by

$$|x|_Q = Q(\{\sigma \,:\, x \in \mathrm{Im}(\sigma)\}). \tag{3.3.1}$$

Then one looks at the minimization of the functional

$$E_g(Q) = \int_C \int_0^1 g(|\sigma(t)|_Q)\, |\sigma'(t)|\, dt\, dQ(\sigma),$$

over the set $TP(\rho_0,\rho_1)$. A straightforward scaling argument shows that the right choice for g, in order to obtain an energy which reproduces the Gilbert-Steiner one (3.2.1) (i.e. sums of mass$^\alpha$ $\times$ length), is

$$g(t) = t^{\alpha-1},$$

with the convention $0^{\alpha-1} = \infty$, so that the energy one wants to consider is given by

$$E_\alpha(Q) = \int_C \int_0^1 |\sigma(t)|_Q^{\alpha-1}\, |\sigma'(t)|\, dt\, dQ(\sigma),\ \ Q \in TP(\rho_0,\rho_1).$$

Observe in particular that if $E_\alpha(Q) < +\infty$, then Q can not give positive mass to a set of paths with zero multiplicity.

REMARK 3.3.1. Using the one-dimensional area formula, it is not difficult to see that for every Q concentrated on injective curves and having finite E_α, we have

$$E_\alpha(Q) = \int_\Omega |x|_Q^\alpha \, d\mathscr{H}^1(x),$$

where it is understood that in the right-hand side we have that the set $\{x \in \Omega \,:\, |x|_Q \neq 0\}$ has $\sigma-$finite $\mathscr{H}^1$ measure (see [**17**, Proposition 4.8]).

So far, this is a heuristic presentation of the Lagrangian models: as briefly mentioned, the choice (3.3.1) for the multiplicity is not the unique that can be made, another possible choice being

$$|(x,t)|_Q = Q(\{\sigma \,:\, x = \sigma(t)\}), \tag{3.3.2}$$

which describes the mass transiting from a point x at a certain time instant t. This is the multiplicity used in the Lagrangian model of Bernot and Figalli (the *synchronized traffic plans model*, see [**18**]): compared with (3.3.1), it has the advantage of being local both in time and space, so that the model presented in [**18**] gives a description of branched transportation problems, which is more dynamical in spirit. Moreover if $\mu_t = (e_t)_\sharp Q$, then

$$|(x,t)|_Q = \mu_t(\{x\}),$$

so that this is the multiplicity which can be naturally compared with the path functional one (we will give the details of this comparison in the next chapter). Another possible definition of multiplicity, and actually the first one which has been introduced, is the multiplicity of the irrigation pattern model of Maddalena, Morel and Solimini, which is only suitable for the case where ρ_0 is a Dirac mass, i.e. $\rho_0 = \delta_{x_0}$. Using the language of probability measures over $C([0,1];\Omega)$, it is defined by

$$[\sigma]_{t,Q} = Q(\{\psi \,:\, \psi(s) = \sigma(s), \text{ for every } s \in [0,t]\}), \tag{3.3.3}$$

which gives, for every curve σ, the quantity of mass which has traveled together with it up to time t.

We now proceed to give a more systematic presentation of these models. We introduce the time interval $I = [0,\infty)$ and instead of the space $C([0,1];\Omega)$, it is considered the space $\mathrm{Lip}_1(I;\Omega)$ of all $1-$Lipschitz curves over Ω, equipped with the topology of the uniform convergence on compact sets, the latter being metrizable.

This space has the advantage of being a compact metric space, so that for example every sequence of probability measures on $\mathrm{Lip}_1(I;\Omega)$ is automatically equi-tight (and thus weakly converging). By the term *traffic plan* we mean every element of $\mathscr{P}(\mathrm{Lip}_1(I;\Omega))$. The function $T : \mathrm{Lip}_1(I;\Omega) \to [0,\infty]$ is the *stopping time* of a curve σ, defined as

$$T(\sigma) = \inf\{t \in [0,\infty) \,:\, \sigma \text{ is constant on } [t,\infty)\}, \tag{3.3.4}$$

and we recall that T is a lower semicontinuous function (see [**15**, Lemma 4.2]). We then define the set of traffic plans with prescribed initial and final measures

$$TP(\rho_0,\rho_1) = \{Q \in \mathscr{P}(\mathrm{Lip}_1(I;\Omega)) \,:\, Q(\{T = +\infty\}) = 0,\, (e_0)_\sharp Q = \rho_0,\, (e_\infty)_\sharp Q = \rho_1\},$$

where for every $t \in I$, the function e_t is the usual evaluation at time t map and the application e_∞ is defined on the set $\{\sigma : T(\sigma) < +\infty\}$ through $e_\infty(\sigma) = \sigma(T(\sigma))$.

Given a traffic plan $Q \in \mathscr{P}(\mathrm{Lip}_1(I;\Omega))$, for every $\alpha \in (0,1)$ we then define its energy as

$$E_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^\infty |\sigma(t)|_Q^{\alpha-1} \, |\sigma'(t)| \, dt \, dQ(\sigma),$$

then we have the following.

THEOREM 3.3.2. *The minimization problem*

$$\min_{TP(\rho_0,\rho_1)} E_\alpha(Q),$$

admits a solution, provided there exists an admissible traffic plan with finite energy.

Alternatively, one can replace the multiplicity defined in (3.3.1) with that given by (3.3.2), thus considering for every traffic plan Q its *synchronized energy* ([**18**])

$$S_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^\infty |(\sigma(t),t)|_Q^{\alpha-1} \, |\sigma'(t)| \, dt \, dQ(\sigma),$$

or, in the case that $\rho_0 = \delta_{x_0}$, its *irrigation energy* ([**67**])

$$I_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^\infty [\sigma]_{t,Q}^{\alpha-1} \, |\sigma'(t)| \, dt \, dQ(\sigma),$$

and consider the relative minimization problems. In the following Theorem, we summarize the main results about comparison between Xia's and Lagrangian models: for the proofs one can see [**17**, Chapter 9] and [**18**, Section 6].

THEOREM 3.3.3. *Let $\rho_0 = \delta_{x_0}$, then the three Lagrangian models are all equivalent. More generally, if ρ_0 is finitely atomic, i.e.*

$$\rho_0 = \sum_{i=1}^{s} a_i \, \delta_{x_i},$$

then the synchronized traffic plan model, corresponding to S_α, is equivalent to the model of Bernot, Caselles and Morel. In any case, the Lagrangian model corresponding to E_α is equivalent to Xia's model. In particular, for every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ we have

$$\min_{Q \in TP(\rho_0,\rho_1)} E_\alpha(Q) = d_\alpha(\rho_0,\rho_1), \tag{3.3.5}$$

where d_α is defined in (3.2.3).

REMARK 3.3.4. It is worth stressing that by *equivalent* we mean not only equality of the minima (which is the less interesting part), but more important that the energies of the models are the same (actually, a Gilbert-Steiner one) and that the optimal structures described do coincide. Moreover a natural way to pass from the minimizers of a problem to another is given.

A nice consequence of the equivalence (3.3.5) is exploited in the next result: it will be useful in the next two Chapters.

PROPOSITION 3.3.5. *For every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, there exist two sequences $\{\rho_0^n\}_{n\in\mathbb{N}}$ and $\{\rho_1^n\}_{n\in\mathbb{N}}$ of finitely atomic probability measures, weakly converging to ρ_0 and ρ_1 respectively, and such that*

$$\lim_{n\to\infty} \min_{Q\in TP(\rho_0^n,\rho_1^n)} E_\alpha(Q) = \min_{Q\in TP(\rho_0,\rho_1)} E_\alpha(Q). \tag{3.3.6}$$

PROOF. Clearly, the proof is just a straightforward consequence of the relaxed formulation of the energy M_α^* in Xia's model and of equivalence (3.3.5). □

REMARK 3.3.6. Observe that the energy E_α has the great advantage of being invariant under time reparametrization (the definition of the multiplicity plays a crucial role in this). This in particular implies that defining the map $\mathfrak{r}(\sigma) = \widetilde{\sigma}$, with $\widetilde{\sigma}$ arc-length reparametrization of σ, and setting $\widetilde{Q} = (\mathfrak{r})_\sharp Q$, we have

$$E_\alpha(Q) = E_\alpha(\widetilde{Q}).$$

In particular, we get that if one withdraws the derivative term and still works with curves which are $1-$Lipschitz, considering the modified energy

$$\widetilde{E}_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^{T(\sigma)} |\sigma(t)|_Q^{\alpha-1}\, dt\, dQ(\sigma),\ \ Q \in TP(\rho_0,\rho_1),$$

then the minimization will give the same result, that is

$$\min_{Q\in TP(\rho_0,\rho_1)} E_\alpha(Q) = \min_{Q\in TP(\rho_0,\rho_1)} \widetilde{E}_\alpha(Q),$$

with the difference that now we are selecting a class of precise minimizers Q, the ones concentrated on curves which move at maximal speed.

Notice that on the contrary the energy I_α is not reparametrization invariant, but anyway we still have that

$$\min_{Q\in TP(\delta_{x_0},\rho_1)} I_\alpha(Q) = \min_{Q\in TP(\delta_{x_0},\rho_1)} \widetilde{I}_\alpha(Q),$$

where

$$\widetilde{I}_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^{T(\sigma)} [\sigma(t)]_{t,Q}^{\alpha-1}\, dt\, dQ(\sigma),$$

once we observe that there holds

$$I_\alpha(Q) \geq I_\alpha(\widetilde{Q}),$$

each time we take $\widetilde{Q} = (\mathfrak{r})_\sharp Q$, that is I_α decrease under arc-length reparametrization.

The previous observations about time reparametrizations do not straightforwardly apply to the case of S_α: anyway, we need to stress the following clarifying result (see for instance [**77**, Section 2.3]) and see in a while some of its consequences.

THEOREM 3.3.7. *If Q_0 is an optimal traffic plan minimizing $\widetilde{E}_\alpha$ on $TP(\delta_{x_0},\rho_1)$, then Q_0 is concentrated on curves σ which are parametrized by arc length (i.e. $|\sigma'(t)| = 1$ a.e. on $[0,T(\sigma)]$) and such that, for all times $t < T(\sigma)$, the equality*

$$|\sigma(t)|_{Q_0} = [\sigma]_{t,Q_0} = |(\sigma(t),t)|_{Q_0},$$

holds.

Then if one considers the problem of minimizing

$$\widetilde{S}_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^{T(\sigma)} |(\sigma(t), t)|_Q^{\alpha-1} \, dt \, dQ(\sigma),$$

thanks to the fact that $\widetilde{E}_\alpha(Q) \leq \widetilde{S}_\alpha(Q)$ for every traffic plan Q and that, by means of Theorem 3.3.7, we have equality on Q_0 optimizer of $\widetilde{E}_\alpha$ (and thus of E_α), we obtain

$$\min_{TP(\rho_0,\rho_1)} S_\alpha = \min_{TP(\rho_0,\rho_1)} E_\alpha = \min_{TP(\rho_0,\rho_1)} \widetilde{E}_\alpha = \min_{TP(\rho_0,\rho_1)} \widetilde{S}_\alpha,$$

in the case that $\rho_0 = \delta_{x_0}$, having used Theorem 3.3.3. This in particular tells that minimizing the synchronized energy with or without the derivative term does not affect the problem and it selects a traffic plan concentrated on curves parametrized by arc-length: this will be particularly useful in the next chapter.

For traffic plans minimizing E_α, Bernot, Caselles and Morel have also shown some regularity properties, asserting that, under suitable conditions, they have the structure of a finite graph (this is not at all obvious when ρ_0 and ρ_1 are general probability measures): just to give a flavour of the kind of results one can expect for, we give the following, which is an *interior* (that is, away from the support of the irrigating and irrigated measure) regularity statement (see [**16**, Theorem 4.7]).

THEOREM 3.3.8. *Let $\alpha \in (1 - 1/N, 1)$ and let $Q \in TP(\rho_0, \rho_1)$ be optimal for E_α. Assume that the supports of ρ_0 and ρ_1 are at positive distance. In any closed ball $B_r(x_0)$ not meeting the supports of ρ_0 and ρ_1, the set*

$$S_Q = \{x \in \Omega \, : \, |x|_Q > 0\},$$

has the structure of a finite graph.

4. The path functional model

As one can figure it out, the concave power α will play a prominent role in this model, too. The point of view introduced in [**24**] is that of studying weighted-length functionals of the type

$$\int_0^1 g(\mu_t) \, |\mu_t'|_{w_p} \, dt, \tag{3.4.1}$$

defined on the space of $\mathcal{W}_p(\Omega)-$valued Lipschitz curves.

Minimizers of functionals (3.4.1), under the constraints $\mu_0 = \rho_0$ and $\mu_1 = \rho_1$, can be seen as geodesics, with respect to a metric which is a conformal perturbation of the Wasserstein one, in the space of probability measures, joining ρ_0 and ρ_1. Indeed, recall that by Chapter 1, formula (1.3.3), we know that $\mathcal{W}_p(\Omega)$ is a geodesic space and moreover the $p-$Wasserstein distance between ρ_0 and ρ_1 can be characterized as

$$w_p(\rho_0, \rho_1) = \min_{\mathcal{C}_{\infty,p}(\rho_0,\rho_1)} \int_0^1 |\mu_t'|_{w_p} \, dt,$$

where, with the notation of the previous chapter, $\mathcal{C}_{\infty,p}(\rho_0, \rho_1)$ is the space of w_p-Lipschitz curve, connecting ρ_0 to ρ_1.

With suitable choices of the conformal factor g in (3.4.1), one can either treat the case of diffusion, where masses spread all over Ω, or that of concentration, where on the contrary masses travel together as much as possible, giving rise to three-shaped optimal structures. For the purposes of this chapter, we will confine ourselves only to this second case.

In order to modelize branched phenomena through (3.4.1), in [**24**] it is considered the lower-semicontinuous weight-function defined over $\mathscr{P}(\Omega)$ by

$$g_\alpha(\mu) = \begin{cases} \sum_{k\in\mathbb{N}} m_k^\alpha, & \text{if } \mu = \sum_{k\in\mathbb{N}} m_k \delta_{y_k}, \\ +\infty, & \text{otherwise,} \end{cases} \tag{3.4.2}$$

with $\alpha \in (0, 1)$, which is a local functional defined on measures, of the kind studied by Bouchitté and Buttazzo in [**21**]. We define the $(\alpha, p)-$*path functional energy* by

$$\mathcal{P}_{\alpha,p}(\mu) = \int_0^1 g_\alpha(\mu_t)\, |\mu_t'|_{w_p}\, dt,\ \ \mu \in \mathrm{Lip}([0,1]; \mathcal{W}_p(\Omega)), \tag{3.4.3}$$

then observing that, due to the sub-additivity of $x \mapsto x^\alpha$, we have

$$g_\alpha(\mu) \geq 1,\ \ \mu \in \mathscr{P}(\Omega),$$

and applying Theorem 2.4.2 of Chapter 2 with $I = [0, 1]$, $X = \mathcal{W}_p(\Omega)$ and $g = g_\alpha$, it is straightforward to obtain the following existence result.

THEOREM 3.4.1. *For every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, the minimization problem*

$$\inf_{\mu\in\mathcal{C}_{\infty,p}(\rho_0,\rho_1)} \mathcal{P}_{\alpha,p}(\mu),$$

admits a solution, provided there exists an element of $\mathcal{C}_{\infty,p}(\rho_0, \rho_1)$ having finite energy.

REMARK 3.4.2. Observe that the term g_α is finite only on atomic measures and reproduces the energy with the masses to the power of α, which is typical of all the other models. Moreover this is a purely dynamical model, as far as any optimal curve provides the evolution of the branched transportation and not just the branched structure underlying the movement: also notice that the term $g_\alpha(\mu_t)$ is local both in space and time.

In the sequel, we will denote by $D_{\alpha,p} : \mathscr{P}(\Omega) \times \mathscr{P}(\Omega) \to [0, +\infty]$ the function defined by the minimization problem of Theorem 3.4.1, that is

$$D_{\alpha,p}(\rho_0, \rho_1) = \min_{\mu\in\mathcal{C}_{\infty,p}(\rho_0,\rho_1)} \mathcal{P}_{\alpha,p}(\mu).$$

We aim to show that for suitable choices of α, these functions are actually always finite and moreover they define a family of distances on the space of probability measure over Ω: then it will be interesting to compare these distances with the usual Wasserstein ones.

In order to do this, we first need to fix some notations: we set $K = [0, 1]^N$ and $K_L = [0, L]^N$, then for every $j \in \mathbb{N}$ we consider the set of multi-indexes

$$B_j = \{z \in \mathbb{N}^N\, :\, \|z\|_\infty \leq 2^j - 1\},$$

and observe that $\sharp(B_j) = 2^{jN}$. We partition the cube K_L with dyadic cubes having edge length $L/2^j$, i.e.

$$K_L = \bigcup_{i=1}^{2^{jN}} K_j^i := \bigcup_{z\in B_j} \frac{L\,K + L\,z}{2^j},$$

and we call $\{K_j^i\}_{i=1}^{2^{jN}}$ the $j-th$ *generation* of cubes. For every $\mu \in \mathscr{P}(\Omega)$ such that $\Omega \subset K_L$, its *dyadic approximation* is given by

$$a_j(\mu) = \sum_{i=1}^{2^{jN}} m_j^i\,\delta_{x_j^i},$$

where $m_j^i = \mu(K_j^i)$ and x_j^i is the center of K_j^i. From now on, we shall always assume that $\Omega \subset K_L$ for a suitable L.

We start giving an estimate for the minimal value of (3.4.1) when ρ_1 is a generic probability measure and ρ_0 is one of its dyadic approximations: this is the main tool is order to guarantee that the function $D_{\alpha,p}$ is always finite, at least for well chosen values of α.

Proposition 3.4.3. *Let $\alpha \in (1-1/N, 1]$ and $p \in [1,\infty]$, then for every $\mu \in \mathscr{P}(\Omega)$ we have*

$$D_{\alpha,p}(a_j(\mu),\mu) \le \frac{2^{(N(1-\alpha)-1)j}}{2^{1-N(1-\alpha)}-1}\,\frac{L\,\sqrt{N}}{2}. \tag{3.4.4}$$

Proof. The idea is simple and nowadays very standard in this context: we have to transport $a_j(\mu)$ on $a_{j+1}(\mu)$, then $a_{j+1}(\mu)$ on $a_{j+2}(\mu)$ and so on; then the $*-$weak convergence of $\{a_k(\mu)\}_{k\ge j}$ to μ will lead to the desired conclusion. In order to do this, it suffices to use a constant speed Wasserstein geodesic at every step, which in this particular case has a very simple structure, due to the fact that the measures are atomic and their supports satisfy a rigid geometrical condition (i.e. they are the centers of dyadic cubes). At every step, we start from the centers $\{x_k^i\}$ of the $k-$generation and we split each mass m_k^i in 2^N parts, sending each of these pieces to the centers of the cubes of the $(k+1)-$generation: the distance covered by each particle is given by $2^{-(k+2)}L\,\sqrt{N}$, so that the Wasserstein distance between two successive dyadic approximations is given by

$$w_p(a_k(\mu), a_{k+1}(\mu)) = \frac{L\,\sqrt{N}}{4\cdot 2^k}.$$

We now define recursively the time steps

$$\begin{cases} t_j &= \quad 0,\\ t_{k+1} &= \quad t_k + \dfrac{L\,\sqrt{N}}{4\cdot 2^{j+k}}, \quad \text{for every } k \ge j,\end{cases}$$

and we consider the geodesic curves $\mu^k : [t_k, t_{k+1}] \to \mathscr{W}_p(\Omega)$ connecting $a_k(\mu)$ and $a_{k+1}(\mu)$, parametrized in such a way that

$$|(\mu^k_t)'|_{w_p} \equiv 1,\ t \in [t_k, t_{k+1}],$$

then interpolating all these curves, we can estimate the energy as follows

$$\begin{aligned} D_{\alpha,p}(a_j(\mu),\mu) &\le \sum_{k=j}^{\infty}\left(\int_{t_k}^{t_{k+1}}\sum_{i=1}^{2^{(k+1)N}}(m_{k+1}^i)^\alpha\,dt\right) \\ &= \frac{L\sqrt{N}}{2}\sum_{k=j}^{\infty}\left(\frac{1}{2^{k+1}}\sum_{i=1}^{2^{(k+1)N}}(m_{k+1}^i)^\alpha\right) \\ &\le \frac{L\sqrt{N}}{2}\sum_{k=j}^{\infty}\left(\frac{1}{2^{k+1}}2^{(k+1)N}\frac{1}{2^{(k+1)N\alpha}}\right) = \frac{L\sqrt{N}}{2}\sum_{s=j+1}^{\infty}\left(2^{N(1-\alpha)-1}\right)^s, \end{aligned}$$

where we have used the fact that the function $f(x_1,\dots,x_n)=\sum_{i=1}^n x_i^\alpha$ achieves its maximum for $x_i=1/n$, $i=1,\dots,n$, under the constraint $\sum_{i=1}^n x_i=1$.

We now observe that thanks to the assumptions on α, the exponent $N(1-\alpha)-1$ is negative, so that simple calculations yields

$$\begin{aligned} \sum_{s=j+1}^{\infty}\left(2^{N(1-\alpha)-1}\right)^s &= \frac{1}{1-2^{N(1-\alpha)-1}} - \frac{1-2^{(N(1-\alpha)-1)(j+1)}}{1-2^{N(1-\alpha)-1}} \\ &= \frac{2^{1-N(1-\alpha)}}{2^{1-N(1-\alpha)}-1} - \frac{2^{1-N(1-\alpha)}-2^{(N(1-\alpha)-1)j}}{2^{1-N(1-\alpha)}-1} \\ &= \frac{2^{(N(1-\alpha)-1)j}}{2^{1-N(1-\alpha)}-1}, \end{aligned}$$

thus giving the desired estimate. □

THEOREM 3.4.4. *Let $\alpha\in(1-1/N,1)$ and $p\in[1,\infty]$, then for every $\rho_0,\rho_1\in\mathscr{P}(\Omega)$ the quantity $D_{\alpha,p}(\rho_0,\rho_1)$ is finite. Moreover we have the following estimate*

$$D_{\alpha,p}(\rho_0,\rho_1)\le L\sqrt{N}\,\frac{1}{2^{1-N(1-\alpha)}-1}. \tag{3.4.5}$$

PROOF. This now follows straightforwardly from Proposition 3.4.3. Indeed, let x_0 denote the center of Q_L, it is sufficient to observe that

$$D_{\alpha,p}(\rho_0,\rho_1)\le D_{\alpha,p}(\rho_0,\delta_{x_0})+D_{\alpha,p}(\delta_{x_0},\rho_1)$$

and then use estimate (3.4.4) with $i=0$. □

REMARK 3.4.5. Observe in particular that we obtain finiteness of the minimum for the same values of α as in Xia's model. In the case $\alpha\in[0,1-1/N]$ on the contrary, counter-examples to the finiteness of $D_{\alpha,p}$ can be constructed (see [**24**, Theorem 3.5]). The interesting problem of characterizing, given a probability measure ρ_0, the values of (α,p) for which there results

$$D_{\alpha,p}(\rho_0,\delta_{x_0})<+\infty,$$

has been addressed in the recent paper [**19**] by Bianchini and Brancolini.

We are now ready to prove that through the path functional model, we can naturally define a new metric on the space of probability measures.

PROPOSITION 3.4.6. *Let $\alpha \in (1-1/N, 1)$ and $p \in [1,\infty]$, then $D_{\alpha,p}(\cdot,\cdot)$ defines a distance on $\mathscr{P}(\Omega)$.*

PROOF. We first observe that thanks to Theorem 3.4.4, with this choice of α we get that $D_{\alpha,p}$ is always finite. Moreover

$$D_{\alpha,p}(\rho_0,\rho_1) = 0 \Longleftrightarrow \rho_0 = \rho_1,$$

one implication being obvious, while the other is just a consequence of the fact that

$$D_{\alpha,p}(\rho_0,\rho_1) \geq \min_{\mu\in\mathcal{C}_\infty(\rho_0,\rho_1)} \int_0^1 |\mu_t'|_{w_p}\, dt = w_p(\rho_0,\rho_1). \tag{3.4.6}$$

Finally, it is straightforward to see that $D_{\alpha,p}$ verifies the triangular inequality. □

REMARK 3.4.7. Thanks to the monotonicity property of Wasserstein distances, the path functionals metrics are ordered, that is

$$p \leq q \Longrightarrow D_{\alpha,p}(\rho_0,\rho_1) \leq D_{\alpha,q}(\rho_0,\rho_1).$$

As one can imagine by means of its very definition, the metric $D_{\alpha,p}$ turns $\mathscr{P}(\Omega)$ into a geodesic space: this is exactly the content of the next simple result.

PROPOSITION 3.4.8. *Let $\alpha \in (1-1/N, 1)$ and $p \in [1,\infty]$. Then $(\mathscr{P}(\Omega), D_{\alpha,p})$ is a geodesic space.*

PROOF. The proof is almost straightforward, using the definition of $D_{\alpha,p}$: it is sufficient to take $\rho_0,\rho_1 \in \mathscr{P}(\Omega)$ and a curve $\mu \in \mathcal{C}_\infty(\rho_0,\rho_1)$ such that

$$D_{\alpha,p}(\rho_0,\rho_1) = \int_0^1 g_\alpha(\mu_t)\, |\mu_t'|_{w_p}\, dt,$$

and show that there results

$$|\mu_t'|_{D_{\alpha,p}} \leq g_\alpha(\mu_t)\, |\mu_t'|_{w_p}, \quad \text{for } \mathscr{L}^1\text{−a.e. } t\in[0,1]. \tag{3.4.7}$$

Indeed, if this is true, then we get

$$\int_0^1 |\mu_t'|_{D_{\alpha,p}}\, dt \leq D_{\alpha,p}(\rho_0,\rho_1),$$

so that μ is a geodesic in $(\mathscr{P}(\Omega), D_{\alpha,p})$ between μ_0 and μ_1. To show (3.4.7), we consider $\widetilde{\mu} \in \mathcal{C}_\infty(\mu_t,\mu_{t+h})$ given by

$$\widetilde{\mu}_s = \mu_{t+sh}, \;\; s\in[0,1],$$

then using the definition of $D_{\alpha,p}$, we get

$$D_{\alpha,p}(\mu_t,\mu_{t+h}) \leq \int_0^1 g_\alpha(\widetilde{\mu}_s)\, |\widetilde{\mu}_s'|_{w_p}\, ds = \int_t^{t+h} g_\alpha(\mu_\tau)\, |\mu_\tau'|_{w_p}\, d\tau,$$

and dividing by $h>0$ and taking the limit as h goes to 0, we obtain (3.4.7). □

We have seen that, thanks to (3.4.6), the topology induced by the distance $D_{\alpha,p}$ is comparable to the $p-$Wasserstein one: it is not difficult to see that actually the two topologies are equivalent, as the next result, which comes from some conversations with Filippo Santambrogio, shows in a quantitative form.

THEOREM 3.4.9. *Let $\alpha \in (1 - 1/N, 1)$ and $p \in [1, \infty]$, then*

$$D_{\alpha,p}(\rho_0, \rho_1) \le C\, w_p(\rho_0, \rho_1)^{1-N(1-\alpha)}, \tag{3.4.8}$$

for a constant C depending on N, α and the diameter of Ω.

PROOF. Using the triangular inequality we get

$$D_{\alpha,p}(\rho_0, \rho_1) \le D_{\alpha,p}(\mu_0, a_j(\rho_0)) + D_{\alpha,p}(a_j(\rho_0), a_j(\rho_1)) + D_{\alpha,p}(a_j(\rho_1), \rho_1),$$

where as before $a_j(\cdot)$ stands for the dyadic approximation of a measure. We then observe that

$$D_{\alpha,p}(\rho_i, a_j(\rho_i)) < C\, 2^{(N(1-\alpha)-1)j},\ i = 0, 1,$$

as a consequence of Proposition 3.4.3, where $N(1 - \alpha) - 1 < 0$ thanks to the assumptions on α. Let us then estimate the term $D_{\alpha,p}(a_j(\rho_0), a_j(\rho_1))$ as follows: first of all, we observe that to give an estimate from above of this term, it is sufficient to exhibit a Lipschitz curve in $\mathcal{W}_p(\Omega)$ connecting the two atomic measures $a_j(\rho_0)$ and $a_j(\rho_1)$ and then to compute its path functional energy. The most simple choice is clearly that of taking $\mu_t^j = ((1 - t)\, \pi_x + t\, \pi_y)_\# \gamma$, with $\gamma \in \Pi(a_j(\mu_0), a_j(\mu_1))$ optimal transport plan (see Chapter 1, Theorem 1.3.12), that is a $p-$Wasserstein constant speed geodesic curve μ_t^j between these two atomic measures (which is a finitely atomic measure, for every t), parametrized by arc-length on the time interval $[0, 1]$, so that

$$D_{\alpha,p}(a_j(\rho_0), a_j(\rho_1)) \le \int_0^1 g_\alpha(\mu_t^j)\, |(\mu_t^j)'|_{w_p}\, dt = w_p(a_j(\rho_0), a_j(\rho_1)) \int_0^1 g_\alpha(\mu_t^j)\, dt,$$

then it is left to estimate the integral term on the right-hand side. For this, we notice that, indicating with $n_j(t) = \#\mathrm{spt}(\mu_t^j)$, i.e. $n_j(t)$ is the number of atoms of μ_t^j, we get that

$$g_\alpha(\mu_t^j) \le n_j(t)^{1-\alpha}.$$

Moreover thanks to Proposition 1.2.3 of Chapter 1, we can suppose that this optimal transport plan does not move more than $2 \cdot 2^{jN}$ atoms (actually we can do slightly better, but this is not the point here). This implies that we can can give a nice estimation on the number $n_j(t)$, that is

$$n_j(t) \le \#\mathrm{spt}(a_j(\rho_0)) + \#\mathrm{spt}(a_j(\rho_1)) = 2 \cdot 2^{jN},$$

so that in the end we get

$$D_{\alpha,p}(a_j(\rho_0), a_j(\rho_1)) \le w_p(a_j(\rho_0), a_j(\rho_1)) \int_0^1 n_j(t)^{1-\alpha}\, dt = C\, w_p(a_j(\rho_0), a_j(\rho_1))\, 2^{jN(1-\alpha)}.$$

Summarizing, up to now we have obtained

$$D_{\alpha,p}(\rho_0, \rho_1) \le C\, 2^{(N(1-\alpha)-1)j} + w_p(a_j(\rho_0), a_j(\rho_1))\, 2^{jN(1-\alpha)}. \tag{3.4.9}$$

To let appear on the right-hand side the $p-$Wasserstein distance between the original measures, we simply use again triangular inequality, that is

$$w_p(a_j(\rho_0), a_j(\rho_1)) \le w_p(a_j(\rho_0), \rho_0) + w_p(\rho_0, \rho_1) + w_p(\rho_1, a_j(\rho_1)),$$

and observe that

$$w_p(a_j(\rho_i), \rho_i) \le C\, 2^{-j},\ i = 0, 1,$$

with C depending only on the diameter of Ω and N, so that plugging these into (3.4.9), we get

$$D_{\alpha,p}(\rho_0,\rho_1) \leq C\, 2^{(N(1-\alpha)-1)j} + w_p(\rho_0,\rho_1)\, 2^{jN(1-\alpha)}.$$

To conclude, it is now sufficient to choose the index j in such a way that $2^{-j} \simeq w_p(\rho_0,\rho_1)$: more precisely, taking $j \in \mathbb{N}$ such that

$$\frac{\mathrm{diam}\,(\Omega)}{2^j} \leq w_p(\rho_0,\rho_1) \leq \frac{\mathrm{diam}\,(\Omega)}{2^{j-1}},$$

a choice which is always possible, as far as $w_p(\rho_0,\rho_1) \leq \mathrm{diam}\,(\Omega)$, we get

$$D_{\alpha,p}(\rho_0,\rho_1) \leq C\, w_p(\rho_0,\rho_1)^{1-N(1-\alpha)},$$

with the constant C depending only on N, α and the diameter of Ω, which is exactly what we wanted to prove. □

REMARK 3.4.10. We remark that the results of Theorems 3.4.4 and 3.4.9, Proposition 3.4.8 are the natural counter parts of the results recalled in Section 2 for the transport path model (Theorem 3.2.1).

5. Some remarks on the path functional model: towards an equivalent formulation

We start this final section with a very simple but instructive example, in which we try to compare the path functional model to the other ones: we will see that equivalence is not guaranteed, generally speaking.

EXAMPLE 3.5.1. Given three pairwise distinct points $x_0, x_1, x_2 \in \mathbb{R}^N$ such that $|x_i - x_0| >> |x_1 - x_2|$, with $i = 1, 2$, we fix two probability measures

$$\rho_0 = \delta_{x_0},\ \ \rho_1 = \frac{1}{2}\delta_{x_1} + \frac{1}{2}\delta_{x_2},$$

and we try to compare the quantities $D_{\alpha,p}(\rho_0,\rho_1)$, corresponding to the path functional model, and $d_\alpha(\rho_0,\rho_1)$ given by Xia's model. We easily see that we have

$$d_\alpha(\rho_0,\rho_1) = |x_0 - x| + \frac{1}{2^\alpha}|x - x_1| + \frac{1}{2^\alpha}|x - x_2|,$$

where the bifurcation point x is determined by the relation

$$-\frac{x - x_0}{|x - x_0|} = \frac{1}{2^\alpha}\frac{x - x_1}{|x - x_1|} + \frac{1}{2^\alpha}\frac{x - x_2}{|x - x_2|}, \tag{3.5.1}$$

which is precisely the balance formula for the optimal path in Xia's model (see [**90**, Example 2.1]). Observe that from (3.5.1) we can easily derive the optimal bifurcation angles θ_1 and θ_2 (see Figure 1), which are given by

$$\cos\theta_1 = \cos\theta_2 = 2^{\alpha-1},$$

and

$$\cos(\theta_1 + \theta_2) = \frac{1 - 2^{1-2\alpha}}{2^{1-2\alpha}}.$$

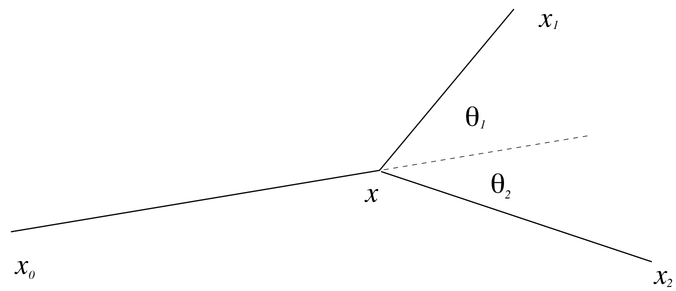

FIGURE 1. Optimal path going from ρ_0 to ρ_1

For the path functional model we see that the optimal curve is given by (observe that this is just the gluing of two Wasserstein geodesics)

$$\mu_t = \begin{cases} \delta_{(1-2t)x_0+2tx}, & t \in [0,1/2], \\ \frac{1}{2}\,\delta_{2(1-t)x+(2t-1)x_1} + \frac{1}{2}\,\delta_{2(1-t)x+(2t-1)x_2}, & t \in [1/2,1], \end{cases}$$

so that

$$D_{\alpha,p}(\rho_0,\rho_1) = \int_0^1 g_\alpha(\mu_t)\,|\mu_t'|_{w_p}\,dt = |x_0 - x| + 2^{1-\alpha-\frac{1}{p}}\left(|x_1 - x|^p + |x_2 - x|^p\right)^{\frac{1}{p}},$$

where the bifurcation point should now satisfy the rather involved relation

$$-\frac{x - x_0}{|x - x_0|} = 2^{1-\alpha-\frac{1}{p}} \frac{|x - x_1|^{p-2}(x - x_1) + |x - x_2|^{p-2}(x - x_2)}{\left(|x - x_1|^p + |x - x_2|^p\right)^{\frac{p-1}{p}}}, \tag{3.5.2}$$

which differs from (3.5.1), also for very special choices of α and p, for example $p = 1/\alpha = 2$. In particular we get that *the path functional model is not equivalent to Xia's one*, as far as they provide different optimal structures. Anyway, observe that in the special case of x_0 belonging to the axis bisecting the segment $\overline{x_1x_2}$, due to the symmetries of the configuration we would have $|x - x_1| = |x - x_2|$, so that (3.5.2) simplifies into

$$-\frac{x - x_0}{|x - x_0|} = 2^{-\alpha}\frac{x - x_1}{|x - x_1|} + 2^{-\alpha}\frac{x - x_2}{|x - x_2|},$$

which are precisely the same optimality conditions as in (3.5.1), so that in this particular case we have

$$D_{\alpha,p}(\rho_0,\rho_1) = d_\alpha(\rho_0,\rho_1),$$

and the two models give rise to the same optimal structures at least in this special symmetric case.

Then, the fact that the path functional model, despite its simple description, has not received much attention, can be seen as a consequence of the fact that it turns out not to be equivalent with the others, in the sense that the optimal structures they describe are not the same, as we have seen. Moreover it shows some unnatural behaviours from a modelization point of view. These are mainly two and we try to explain them in some details, in order to provide a better understanding of the scopes of the next chapter:

(i) *energetic behaviour*: the term g_α is a function of the whole μ, which means that if some masses arrive at their destination and then stop, we continue to pay a cost for them until all the process is over.

Just to clarify, we write down a basic example: suppose you want to transport $\rho_0 = \delta_{x_0}$ to $\rho_1 = m\delta_{x_1} + (1-m)\delta_{x_2}$, where $|x_0 - x_1| = 2|x_0 - x_2|$. A possible connecting curve could be

$$\mu_t = \begin{cases} m\delta_{(1-t)\,x_0 + tx_1} + (1-m)\delta_{(1-2t)\,x_0 + 2t\,x_2}, & t \in [0,1/2], \\ \\ m\delta_{(1-t)\,x_0 + tx_1} + (1-m)\delta_{x_2}, & t \in [1/2,1], \end{cases}$$

but it is easily seen that for a path like this, the path functional energy $\mathcal{P}_{\alpha,p}$ will let you pay a cost for the mass $(1-m)$ also after it is stopped.

On the contrary, it would be desirable to have an energy which takes into account only the moving mass, which in this case is simply given by

$$\nu_t = \begin{cases} \mu_t, & t \in [0,1/2], \\ m\delta_{(1-t)x_0 + tx_1}, & t \in [1/2,1], \end{cases}$$

the latter being no more a curve of probability measures. This is the reason why, at a first stage, the energy (3.4.3) should be modified as follows

$$\widetilde{\mathcal{P}}_{\alpha,p}(\nu,\mu) = \int_0^1 g_\alpha(\nu_t)\,|\mu_t'|_{w_p}\,dt,$$

where now ν is a curve of sub-probability measures, which should represent the moving mass. The curves ν and μ are linked by the condition of being an *evolution pairing*, which is precisely the concept already encountered in Section 5 of Chapter 2: this means that the moving part ν is always less than the total mass μ and that the mass reaching its final destination, given by the difference $\mu - \nu$, has to grow in time (see Section 2). As already discussed (see Chapter 2, Remark 2.5.4), this makes sense when the starting measure $\rho_0 = \delta_{x_0}$ (which is anyway a relevant case, and it was the one studied by Maddalena, Morel and Solimini in [**67**], as we said), so that at time 0 mass starts to move as a whole: for the sake of brevity, our investigation will be strictly confined to these choices of ρ_0, as in [**67**];

(ii) *scaling behaviour*: another problem is the choice of the exponent p, which influences the energy $\mathcal{P}_{\alpha,p}(\mu)$ through the term $|\mu'|_{w_p}$. It seems that the right choice should be $p = +\infty$, for two reasons mainly: the first is that when rescaling a curve μ to be a curve of measures with mass m, we get

$$\mathcal{P}_{\alpha,p}(m\,\mu) = m^{\alpha + \frac{1}{p}}\mathcal{P}_{\alpha,p}(\mu),$$

so that the energy rescales as the power $\alpha + 1/p$, with respect to the mass. Taking $p = +\infty$ clearly settles this behaviour, giving the same scaling as a Gilbert-Steiner energy. The second reason is that the term $|\mu'|_{w_p}$ should play the role of the velocity of the particles, so that it is expected to be *mass-independent*: on the contrary, in the case $p < +\infty$ in general you would have

$$|\mu'|_{w_p} \simeq \left(\sum m\,\ell^p\right)^{\frac{1}{p}},$$

which roughly speaking means that metric velocity is a mass-weighted sum of the velocities of the particles, which strengthen the feeling that $p = +\infty$ should be the right exponent, in order to be able to compare the path functional energy with a Gilbert-Steiner one.

It is the scope of the next chapter to show that, once we correct these unnatural behaviours, it is possible to define a path functional formulation of branched transportation problems, equivalent with the other models. All in all, after this preliminary discussion one is lead to the study of the modified energy given by

$$\widetilde{\mathcal{P}}_{\alpha,\infty}(\nu,\mu) = \int_0^1 g_\alpha(\nu_t)\,|\mu_t'|_{w_\infty}\,dt,$$

but then we have to pay attention to another detail: observe that thanks to the subadditivy of g_α, in the standard path functional model we have

$$g_\alpha(\mu_t) \geq 1,$$

because g_α is evaluated on probability measures: then the existence of a Lipschitz curve minimizing (3.4.3) under a constraint on the endpoints, is almost straightforward as we have seen, thanks to the fact that every minimizing sequence with bounded energy has equi-bounded lengths.

On the contrary, in the modified path functional energy $\widetilde{\mathcal{P}}_{\alpha,\infty}$ one only has

$$g_\alpha(\nu_t) \geq |\nu_t|(\Omega)^\alpha,$$

and the last quantity can go to zero (the moving mass could decrease until it disappears). This fact completely destroys the coercivity of the energy on the space of Lipschitz curves: this means that it could be the case that the transportation process requires an infinite speed (then breaking the Lipschitz constraint), in order to bring all the mass from x_0 to ρ_1 in a finite time, or equivalently, that if you want your curves to stay Lipschitz (i.e. you have an upper bound on the velocities), then you could need an infinite amount of time to complete the transportation. In other words, it may happen that you do not have an upper bound on the length of the paths that particles have to run, because of branching. Curiously enough, this fact is not a drawback, as it is in perfect accordance with the other models, where the existence of an upper bound on the lengths covered by the particles (also for optimal structures) is not known! Indeed, this is still an open problem up to some special cases (see in particular [**16**, Problem 15.13]). We stress the fact that the only case where the answer is known - and it is *yes* - is when the irrigated measure ρ_1 satisfies an *Ahlfors regularity* property, i.e. when its density w.r.t. $\mathscr{H}^s$ is bounded from below for a certain $s \in [0, N]$: in this case, this result is just a consequence of the Hölder continuity of the so-called *landscape function* proven in [**77**] for $s = N$ and then considerably extended in a recent paper by Brancolini and Solimini (see [**25**, Theorem 6.2]).

So in the end, one has to relax the requirement on the finiteness of the time interval and to take advantage of the reparametrization invariance of these weighted length functionals: to keep some compactness one can introduce a bound on the velocities (which does not affect the functional, due to reparametrization), as in the Lagrangian models. It turns out that the kind of energy we are really interested in, as a good candidate to be equivalent to a Gilbert-Steiner energy, is of the form

$$\mathfrak{L}_\alpha(\nu,\mu) := \int_0^\infty g_\alpha(\nu_t)\,|\mu_t'|_{w_\infty}\,dt, \tag{3.5.3}$$

defined for all curves μ which are $\mathcal{W}_\infty(\Omega)-$valued and Lipschitz, with a given Lipschitz constant (let us say 1, for example). It is also clear that keeping the velocity term $|\mu'|_{w_\infty}$ will not be crucial, since if one withdraws it, but keeps the bound $|\mu'|_{w_\infty} \leq 1$, the only effect will be that of selecting those minimizers which move at maximal speed.

CHAPTER 4

An equivalent path functional formulation of branched transportation problems

1. Introduction

This chapter is taken from the joint work [**B5**] with Filippo Santambrogio, where we show equivalence between the Lagrangian models and the path functional model, modified according to the remarks of Section 5 of the previous chapter: for the sake of simplicity, we confine ourselves to the irrigation case, i.e. the case in which the starting measure is a Dirac mass $\rho_0 = \delta_{x_0}$, which is anyway a relevant case and it is the one treated in the model of Maddalena, Morel and Solimini ([**67**]).

Before starting, just some words on the plan of this chapter: in Section 2 we briefly recall the concept of evolution pairing (already encountered in Chapter 2), its main features and we give an existence result for the minimization of functional (3.5.3) over the set of evolution pairings with prescribed endpoints. Section 3 is devoted to a deeper insight into evolution pairings, providing properties and examples that lead us to isolate a good subset (built up of what we call *special evolution pairings*) for which a complete characterization (Section 4) can be given, in terms of the Lipschitz curves of the base space. This characterization is one of the corner-stones of the chapter, which finally permits us to compare, in Section 5, our energy with a Gilbert-Steiner or Bernot-Caselles-Morel one and to show equivalence between our modified path functional model and the other models in the irrigation case.

2. A modified path functional model

Let $\Omega \subset \mathbb{R}^N$ be a compact convex set and let us indicate $I = [0, \infty)$. Moreover since we are interested in studying the branched transport problem with a single Dirac mass as starting measure, in the sequel we will always refer to this configuration and in particular with ρ_0 we will indicate a Dirac mass located at some point of Ω, that is we set $\rho_0 = \delta_{x_0}$, for some $x_0 \in \Omega$.

In this chapter, we will work with the spaces $\mathcal{W}_\infty(\Omega)$ and $\mathcal{M}_1^+(\Omega)$, the latter (see Chapter 2, Section 5) being the space of all positive Radon measures over Ω, with mass smaller than or equal to 1, metrized according to a distance inducing the $*-$weak topology, for instance

$$\mathfrak{d}(\nu_1, \nu_2) = \sum_{k\in\mathbb{N}} \frac{1}{2^k \alpha_k} \left| \int_\Omega \varphi_k(x)\ d(\nu_1(x) - \nu_2(x)) \right|,\ \nu_1, \nu_2 \in \mathcal{M}_1^+(\Omega),$$

where every function φ_k is α_k-Lipschitz and the sequence $\{\varphi_k\}_{k\in\mathbb{N}}$ is dense in

$$\{\varphi \in C(\Omega)\ :\ \varphi \geq 0,\ ||\varphi||_{L^\infty(\Omega)} \leq 1\}.$$

Let us then define the space $\mathrm{Lip}_{1,\mathfrak{d}}(I;\mathcal{W}_\infty(\Omega))$ of all 1-Lipschitz curves in the ∞-Wasserstein space $\mathcal{W}_\infty(\Omega)$, equipped with the $\mathfrak{d}-$*weak convergence* on compact subsets, i.e., indicating with the symbol $\overset{\mathfrak{d}}{\rightharpoonup}$ this convergence, we have

$$\mu^n \overset{\mathfrak{d}}{\rightharpoonup} \mu \Longleftrightarrow \max_{t\in[0,k]} \mathfrak{d}(\mu^n_t,\mu_t) \to 0, \quad \text{for every } k\in\mathbb{N}.$$

Remark 4.2.1. We remark that the use of this convergence is due to the lack of any kind of compactness of the space $\mathcal{W}_\infty(\Omega)$. Moreover we recall that the topology induced by w_∞ is strictly stronger than the $*-$weak topology and we have $\mathfrak{d}\le w_\infty$. What is worthwhile to point out here and crucial for our discussion is that w_∞ is lower semicontinuous with respect to $\mathfrak{d}$ (see Chapter 2, Lemma 2.5.8).

We also recall the definition

$$L^0(I;\mathcal{M}^+_1(\Omega)) := \{\nu : I \to \mathcal{M}^+_1(\Omega) \ :\ \nu \text{ is Borel measurable}\},$$

and on this space we will always consider the pointwise $\mathscr{L}^1-$a.e. convergence. Then in the sequel, when referring to the convergence on the product space $L^0(I;\mathcal{M}^+_1(\Omega))\times \mathrm{Lip}_{1,\mathfrak{d}}(I;\mathcal{W}_\infty(\Omega))$, we will always mean pointwise $\mathscr{L}^1-$a.e. convergence in the first variable and $\mathfrak{d}-$weak in the second.

We want to consider the following energy defined on $L^0(I;\mathcal{M}^+_1(\Omega))\times \mathrm{Lip}_{1,\mathfrak{d}}(I;\mathcal{W}_\infty(\Omega))$

$$\mathfrak{L}_\alpha(\nu,\mu) = \int_0^\infty g_\alpha(\nu_t)|\mu'_t|_{w_\infty}\,dt, \tag{4.2.1}$$

where $g_\alpha : \mathcal{M}^+_1(\Omega)\to\mathbb{R}\cup\{+\infty\}$ is the lower semicontinuous function defined in (3.4.2).

Lemma 4.2.2. *The functional $\mathfrak{L}_\alpha$ defined by* (4.2.1) *is lower semicontinuous on the product space $L^0(I;\mathcal{M}^+_1(\Omega))\times \mathrm{Lip}_{1,\mathfrak{d}}(I;\mathcal{W}_\infty(\Omega))$.*

Proof. The functional under consideration can be written as

$$\mathfrak{L}_\alpha(\nu,\mu) = \sup_{k\in\mathbb{N}} \mathfrak{L}^k_\alpha(\nu,\mu) := \sup_{k\in\mathbb{N}} \int_0^k g_\alpha(\nu_t)|\mu'_t|_{w_\infty}\,dt,$$

and, thanks to the semicontinuity of g_α and of w_∞ with respect to $\mathfrak{d}$, we get that each $\mathfrak{L}^k_\alpha$ is lower semicontinuous with respect to the desired convergence, by means of Theorem 2.3.9 in Chapter 2. It is only left to observe that the supremum of a sequence of lower semicontinuous functions is still a lower semicontinuous function. □

In order to formalize the idea that the curves ν that we aim to consider should represent the moving mass, we give the following slight modification of the concept of *evolution pairing*, already encountered in Chapter 2: the definition is exactly the same, apart for the fact that for technical reasons, here we prefer to work with left-continuous representative (remember that ν in Definition 2.5.3 of Chapter 2 was a BV curve, hence possessing left- and right-continuous representatives).

Definition 4.2.3. Let $(\nu,\mu)\in L^0(I;\mathcal{M}^+_1(\Omega))\times \mathrm{Lip}_{1,\mathfrak{d}}(I;\mathcal{W}_\infty(\Omega))$ be two curves of measures, such that the following are satisfied:

(E1) $\nu_t\le\mu_t$, for every $t\in I$;

(E2) $\vartheta_t := \mu_t - \nu_t$ is monotone non-decreasing and $\mathfrak{d}-$left continuous, that is:

$$\vartheta_s \leq \vartheta_t, \quad \text{for every } s,t \in I, \text{ with } s < t \text{ and } \lim_{s \nearrow t} \mathfrak{d}(\vartheta_s, \vartheta_t) = 0;$$

Then we say that (ν, μ) is an *evolution pairing* and we write $\nu \preceq \mu$.

Notice however that, ϑ being non-decreasing, the condition of left continuity is non-crucial, since one can always modify ϑ_t for t in a $\mathscr{L}^1-$negligible set of times and get a left-continuous curve. It is mainly imposed to give a precise and unambiguous pointwise meaning to ϑ_t for every t, and also to get more easily some of our proofs. Moreover as far as μ is Lipschitz and ϑ is left-continuous, we also get that ν is left-continuous.

Remark 4.2.4. As already observed in Chapter 2, property (E2) implies that the quantity $t \mapsto |\nu_t|(\Omega)$ is non-increasing.

Given a Borel probability measure ρ_1 over Ω, we define the set of admissible evolution pairings

$$EP(\rho_0, \rho_1) = \{\nu \preceq \mu \ : \ \mu_0 = \rho_0, \ \mu_\infty = \rho_1\},$$

where the condition $\mu_\infty = \rho_1$ has to be intended in the sense $\lim_{t \to +\infty} \mathfrak{d}(\mu_t, \rho_1) = 0$, or equivalently, $\mu_t \rightharpoonup \rho_1$ as t goes to $+\infty$.

Recalling the concept of stopping time (3.3.4) for a curve, we introduce the following definition.

Definition 4.2.5. An evolution pairing $(\nu, \mu) \in EP(\rho_0, \rho_1)$ is said to be *normal* if the following conditions hold:

(i) $|\mu'_t|_{w_\infty} = 1$, for a.e. $t \in [0, T(\mu)]$;
(ii) $\nu_t = 0$, for $t \in (T(\mu), +\infty)$, where if $T(\mu) = +\infty$ this condition must intended in the strong sense that $\lim\limits_{t \to +\infty} |\nu_t|(\Omega) = 0$.

In the sequel, with the term *cutting at time T*, we will simply mean the operation that to every ν assigns the product $\nu \cdot 1_{[0,T]}$ of ν for the characteristic function of some time interval $[0, T]$.

We have the following basic result:

Lemma 4.2.6. *Every $(\nu, \mu) \in EP(\rho_0, \rho_1)$ with $\mathfrak{L}_\alpha(\nu, \mu) < +\infty$ is normal, up to a reparametrization of μ and a cutting of ν at the stopping time of μ.*

Proof. Let us take an evolution pairing $(\nu, \mu) \in EP(\rho_0, \rho_1)$ and reparametrize the 1$-$Lipschitz curve μ by arc-length, that is we take $\varphi : [0, +\infty) \to [0, +\infty)$ given by

$$\mathfrak{t}(s) = \inf \left\{ \tau \in I \ : \ s = \int_0^\tau |\mu'_\varrho|_{w_\infty} \, d\varrho \right\}, \ s \in I, \tag{4.2.2}$$

and we set $\widetilde{\mu} = \mu \circ \mathfrak{t}$, then this is a reparametrization of μ (see [**9**, Theorem 4.2.1]) and

$$|\widetilde{\mu}'_t|_{w_\infty} = 1, \ t \in I.$$

Moreover setting $\widetilde{\nu} = \nu \circ \mathfrak{t}$, we clearly get that $(\widetilde{\nu}, \widetilde{\mu})$ is still an evolution pairing contained in $EP(\rho_0, \rho_1)$, for which

$$\mathfrak{L}_\alpha(\widetilde{\nu}, \widetilde{\mu}) = \int_0^\infty g_\alpha(\widetilde{\nu}_t) \, dt = \int_0^\infty g_\alpha(\nu_t) \, |\mu'_t|_{w_\infty} \, dt < +\infty.$$

Using the subadditivity of g_α, the previous in turn implies that the integral

$$\int_0^\infty (|\widetilde{\nu}_t|(\Omega))^\alpha \, dt,$$

must be finite: as far as we are integrating a positive non-increasing function over $[0,\infty)$, we obtain that the integrand must tend to 0, as t tends to ∞.

If $T(\tilde{\mu}) = +\infty$ we have already obtained a normal evolution pairing, otherwise it is sufficient to cut $\tilde{\nu}$ at the time $t = T(\tilde{\mu})$. □

The following Lemma is useful for proving the closedness of the set $EP(\rho_0, \rho_1)$ of evolution pairings joining two given measures, but we will state it in the case where the second measure is not fixed, so as to use it later on in its generality.

LEMMA 4.2.7. *Let $\{(\nu^n, \mu^n)\} \subset EP(\rho_0, \rho_1^n)$ be a sequence of normal evolution pairings such that $(\nu^n, \mu^n) \to (\nu, \mu)$ in $L^0(I; \mathcal{M}_1^+(\Omega)) \times \mathrm{Lip}_{1,\mathfrak{d}}(I; \mathcal{W}_\infty(\Omega))$. Suppose moreover that $\rho_1^n \rightharpoonup \rho_1$ and that*

$$\sup_{n\in\mathbb{N}} \mathfrak{L}_\alpha(\nu^n, \mu^n) < +\infty.$$

Then, up to changing the representative of ν on a negligible set of times $t \in I$, $(\nu, \mu) \in EP(\rho_0, \rho_1)$.

PROOF. We first show that (ν, μ) is an evolution pairing and that $\mu \in \mathrm{Lip}_{1,\mathfrak{d}}(I; \mathcal{W}_\infty(\Omega))$: this can be done as in Lemma 2.6.1 of Chapter 2, since (E1) and (E2) easily pass to limit. Moreover observe that if $\{\nu^n\}_{n\in\mathbb{N}}$ converges to ν $\mathscr{L}^1$–a.e., the same is true for ϑ^n to $\vartheta := \mu - \nu$. In particular, the nondecreasing behaviour of ϑ^n easily passes to the limit, up to the negligible set of non-convergence. Up to replacing ϑ with its left-continuous representative (which means that we only change ϑ_t on a negligible set of times), we get a function which is both monotone and left-continuous.

It remains to show that (ν, μ) still verifies the conditions on the endpoints: the fact that $\mu_0 = \rho_0$ is trivial, so that the only thing to verify is the condition on the final point, that is $\mu_\infty = \rho_1$ in the sense precised before.

In the case that

$$\sup_{n\in\mathbb{N}} T(\mu^n) = T < +\infty, \tag{4.2.3}$$

then we have also $T(\mu) \le T$, using the lower semicontinuity of T. It is now sufficient to use the uniform converge of $\{\mu^n\}_{n\in\mathbb{N}}$ on the interval $[0, T]$ to obtain that

$$\mu_T = \rho_1,$$

which proves the thesis, under the additional hypothesis (4.2.3), by means of the fact that $T(\mu) \le T$.

We now remove assumption (4.2.3), exploiting the concept of evolution pairing. First observe that using property (E2) we have that

$$\vartheta_t^n \le \vartheta_s^n \le \mu_s^n, \quad \text{for every } t, s \in I, \text{ with } t < s,$$

and using the fact that $\mu_\infty^n = \rho_1^n$ we obtain

$$\vartheta_t^n \le \rho_1^n, \tag{4.2.4}$$

and, at the limit as $n \to \infty$, we easily deduce from (4.2.4) that we have $\vartheta_t \le \rho_1$. Moreover the curve ϑ is non-decreasing and

$$|\vartheta_t|(\Omega) = 1 - |\nu_t|(\Omega), \quad \text{for } t \in I,$$

so that, if we are able to prove that $|\nu_t|(\Omega) \to 0$ as $t \to \infty$, we can conclude

$$\lim_{t\to\infty} |\vartheta_t - \rho_1|(\Omega) = 0 \ \text{ and hence } \ \mu_\infty = \vartheta_\infty + \nu_\infty = \rho_1,$$

giving the thesis. At this end we observe that

$$\int_0^\infty g_\alpha(\nu_t)\ dt \le \liminf_{n\to\infty} \int_0^\infty g_\alpha(\nu_t^n)\ dt = \liminf_{n\to\infty} \mathfrak{L}_\alpha(\nu^n,\mu^n) < +\infty,$$

where the first inequality is just a consequence of Fatou Lemma, while the equality right after is a consequence of the normality of each (ν^n,μ^n), so that $\int_0^\infty g_\alpha(\nu_t^n)\,dt = \int_0^\infty g_\alpha(\nu_t^n)\,|\mu_t'|_{w_\infty}\,dt = \mathfrak{L}_\alpha(\nu^n,\mu^n)$. Using again $g_\alpha(\nu_t) \ge |\nu_t|(\Omega)^\alpha$ and the monotone behaviour of $|\nu_t|(\Omega)$ as in Lemma 4.2.6, the latter implies that

$$\lim_{t\to\infty} |\nu_t|(\Omega) = 0,$$

which concludes the proof. □

We are now ready to state and prove a result, about the existence of a minimal evolution pairing connecting two given measures.

Proposition 4.2.8. *The minimization problem*

$$\inf_{(\nu,\mu)\in EP(\rho_0,\rho_1)} \mathfrak{L}_\alpha(\nu,\mu), \tag{4.2.5}$$

admits a solution, provided that there exists an admissible evolution pairing $(\overline{\nu},\overline{\mu})$ having finite $\mathfrak{L}_\alpha$.

Proof. Let $\mathfrak{L}_\alpha(\overline{\nu},\overline{\mu}) = L$ and let us take a minimizing sequence $\{(\nu^n,\mu^n)\}_{n\in\mathbb{N}} \subset EP(\rho_0,\rho_1)$, we can assume that

$$\sup_{n\in\mathbb{N}} \mathfrak{L}_\alpha(\nu^n,\mu^n) \le L + 1.$$

Observe that thanks to Lemma 4.2.6, we can think of every (ν^n,μ^n) as being normal. It is straightforward to see that (up to a subsequence) this minimizing sequence converges in $L^0(I;\mathcal{M}_1^+(\Omega)) \times \mathrm{Lip}_{1,\mathfrak{d}}(I;\mathcal{W}_\infty(\Omega))$ to an evolution pairing (ν,μ): the convergence of $\{\mu^n\}_{n\in\mathbb{N}}$ is just a consequence of the compactness of the space $\mathrm{Lip}_{1,\mathfrak{d}}(I;\mathcal{W}_\infty(\Omega))$, while the convergence of $\{\nu^n\}_{n\in\mathbb{N}}$ follows with a slight modification of the argument in Chapter 2, Theorem 2.6.2.

Moreover (ν,μ) is still admissible, thanks to Lemma 4.2.7, and the thesis follows straightforwardly using the semicontinuity of $\mathfrak{L}_\alpha$ (Lemma 4.2.2). □

3. Further properties of evolution pairings

We start this section with a counter-example, which shows that the class $EP(\rho_0,\rho_1)$ is not the right one in which problem (4.2.5) has to be posed, in order to obtain equivalence with Xia, Bernot-Caselles-Morel and Maddalena-Morel-Solimini models.

EXAMPLE 4.3.1. Let $\rho_0 = \delta_0$ and $\rho_1 = \mathscr{L}^1 \llcorner [-1/2, 1/2]$, we define an evolution pairing (ν, μ) as follows:

$$\mu_t = \begin{cases} \mathscr{L}^1 \llcorner [-t, t] + (1-2t)\delta_t, & t \in [0, 1/2], \\ \mathscr{L}^1 \llcorner \left[-\frac{1}{2}, \frac{1}{2}\right], & t \in (1/2, +\infty) \end{cases}$$

$$\nu_t = \begin{cases} (1-2t)\delta_t, & t \in [0, 1/2], \\ 0, & t \in (1/2, +\infty). \end{cases}$$

Observe that (ν, μ) is normal and it connects ρ_0 to ρ_1. Computing its energy, we have that

$$\mathfrak{L}_\alpha(\nu, \mu) = \int_0^{\frac{1}{2}} g_\alpha(\nu_t) |\mu_t'|_{w_\infty} \, dt = \int_0^{\frac{1}{2}} (1-2t)^\alpha \, dt = \frac{1}{2(\alpha+1)},$$

while the minimal E_α energy is given by

$$2 \int_0^{\frac{1}{2}} \left(\frac{1}{2} - t \right)^\alpha dt = \frac{1}{2^\alpha (\alpha+1)},$$

which is strictly greater than the previous one. We observe that the latter is realized by the traffic plan given by the image measure

$$Q = (\Psi)_\sharp \rho_1,$$

of ρ_1 through the application Ψ that sends every $x \in [-1/2, 1/2]$ to the 1−Lipschitz curve Ψ_x defined by (if $x \geq 0$)

$$\Psi_x(t) = \begin{cases} t, & t \in [0, x], \\ x, & t \in (x, \infty), \end{cases}$$

and by (if $x < 0$)

$$\Psi_x(t) = \begin{cases} -t, & t \in [0, -x], \\ x, & t \in (-x, \infty). \end{cases}$$

Observe that the movement induced by Q is the following: the mass starts to move from the center of the segment, instantaneously splitting in two branches, one going on the right, the other going on the left and continuously disseminating particles on the segment, in an uniform way. This is better visualized by looking at the corresponding evolution pairing, given by

$$\widetilde{\mu}_t = (e_t)_\sharp Q = \begin{cases} \mathscr{L}^1 \llcorner [-t, t] + \frac{1-2t}{2} \delta_{-t} + \frac{1-2t}{2} \delta_t, & t \in [0, 1/2], \\ \mathscr{L}^1 \llcorner \left[-\frac{1}{2}, \frac{1}{2}\right], & t \in (1/2, +\infty), \end{cases}$$

$$\widetilde{\nu}_t = \begin{cases} \frac{1-2t}{2} \delta_{-t} + \frac{1-2t}{2} \delta_t, & t \in [0, 1/2], \\ 0, & t \in (1/2, +\infty), \end{cases}$$

for which $E_\alpha(Q) = \mathfrak{L}_\alpha(\widetilde{\nu}, \widetilde{\mu}) > \mathfrak{L}_\alpha(\nu, \mu)$.

The previous example tells us that in general the elements of $EP(\rho_0, \rho_1)$ (even the minimizers of $\mathfrak{L}_\alpha$, actually) can have strange properties, which has little to do with real physical phenomena of transportation: in fact, in Example 4.3.1 what seems to go wrong is the fact that ν, which is supposed to represent the moving mass, operates a sort of *teleport* from and endpoint of the segment $[-t, t]$ to the opposite one.

Then we have to restrict the class of admissible evolution pairings, isolating those with some good *traveling properties*. In order to do this, we start investigating a property which holds true for a curve having a fixed atomic part. This is a sort of Lipschitz-invariance under mass subtraction, which tells us that once some mass is stopped, then this is no more involved in the transportation process.

LEMMA 4.3.2. *Let $\mu \in \mathrm{Lip}(I;\mathcal{W}_\infty(\Omega))$ be given and suppose that there exists an atomic measure $\mathbf{m} = \sum_{i=1}^{\infty} m_i\delta_{x_i}$ and $t_0 \in I$ such that*

$$\mathbf{m} \le \mu_t, \quad \text{for every } t \in [t_0,\infty).$$

Then the curve $[t_0,\infty) \ni t \mapsto \mu_t - \mathbf{m}$ has the same metric derivative of the curve μ and hence satisfies the following Lipschitz estimate

$$w_\infty(\mu_t - \mathbf{m}, \mu_{t+h} - \mathbf{m}) \le \int_t^{t+h} |\mu_s'|_{w_\infty}\, ds, \quad \text{for every } h \ge 0. \tag{4.3.1}$$

PROOF. The proof may be achieved if one thinks at the characterization of absolutely continuous curves in Wasserstein spaces given in Chapter 1, Section 5. Indeed, as we have already seen in Theorem 1.5.4, if μ is a Lipschitz curve defined on a time interval $[0,T]$ and valued in the space $\mathcal{W}_p(\Omega)$, then there exists a Borel vector field $v : (x,t) \mapsto v_t(x)$ such that

$$v_t \in L^p(\mu_t), \quad \|v_t\|_{L^p(\mu_t)} = |\mu_t'|_{w_p}, \quad \text{for } \mathscr{L}^1\text{–a.e. } t \in I, \tag{4.3.2}$$

and such that the continuity equation holds

$$\partial_t\mu_t + \mathrm{div}_x(v_t\mu_t) = 0. \tag{4.3.3}$$

Moreover this result stays true also for $p = \infty$, as we have proven (see Chapter 1, Proposition 1.5.9).

Another important point is the superposition principle (see Chapter 1, Theorem 1.5.1), which assures that any absolutely continuous curve $t \mapsto \mu_t$ solving (4.3.3) may be obtained as $(e_t)_\sharp Q$, for a probability measure Q on the space of absolutely continuous curves which is concentrated on the solutions of the equation $\sigma'(t) = v_t(\sigma(t))$. As we have seen, for this representation to hold, some integrability conditions on v are needed, but (4.3.2) is widely sufficient.

In our case, since μ is Lipschitz in the w_∞–distance, one knows the existence of a vector field v such that for almost any t the inequality $|v_t(x)| \le 1$ is satisfied (actually, it would be satisfied μ_t–a.e. but one can choose a representative which is everywhere smaller than 1). This implies that the solutions σ of the ODE are Lipschitz continuous curves: as a consequence, they are regular enough to say that, thanks to the one-dimensional area formula, if S is a countable set, then σ' vanishes almost everywhere on the set $\sigma^{-1}(S)$. This implies

$$Q \otimes \mathscr{L}^1\big(\{(\sigma,t)\ :\ \sigma(t) \in S,\, \sigma'(t) \text{ exists and } \sigma'(t) \neq 0\}\big) = 0,$$

and since the curves are solutions of $\sigma'(t) = v_t(\sigma(t))$, using the fact that $\mu_t = (e_t)_\# Q$, this means

$$\int_0^T \mu_t(S \cap \{v_t \neq 0\})dt = Q \otimes \mathscr{L}^1\big(\{(\sigma,t)\ :\ \sigma(t) \in S,\, v_t(\sigma(t)) \neq 0\}\big) = 0.$$

If one chooses as S the set of atoms of $\mathbf{m}$, the previous implies that $\mathbf{m}-$a.e. we have $v_t = 0$, at least for almost any time. Now observe that the continuity equation may obviously be rewritten as

$$\partial_t(\mu_t - \mathbf{m}) + \mathrm{div}_x(v_t\,(\mu_t - \mathbf{m})) + \mathrm{div}_x(v_t\,\mathbf{m}) = 0.$$

and the last term vanishes as a consequence of $v_t\mathbf{m} = 0$: hence one gets that $\mu - \mathbf{m}$ is a solution of the continuity equation with the same velocity field v_t. In particular, since $|v_t| \leq |\mu_t'|_{w_\infty}$, one gets that $\mu - \mathbf{m}$ is Lipschitz according to the $w_\infty-$distance (the latter being easily adapted to the framework of measures with the same mass, instead of probability measures) and its metric derivative with respect to the distance w_∞ does not exceed that of μ. Since it is a straightforward fact to see that there holds

$$w_\infty(\mu_t - \mathbf{m}, \mu_{t+h} - \mathbf{m}) \geq w_\infty(\mu_t, \mu_{t+h}),$$

one can also see the opposite inequality and conclude

$$|(\mu_t - \mathbf{m})'|_{w_\infty} = |\mu_t'|_{w_\infty}, \quad \text{for } \mathscr{L}^1\text{−a. e. } t \in I,$$

which gives the thesis. □

REMARK 4.3.3. It is not difficult to see that the same conclusions of the previous Lemma hold, if we take $\mathbf{m}$ to be a positive Borel measure concentrated on some $\mathscr{H}^1-$negligible Borel set S.

The properties proven in the previous Lemma roughly says that, in certain cases, the speed of a curve μ coincides with the speed of its moving part. This seems a general fact, but we will check that the evolution pairing in Example 4.3.1 is far from satisfying this property. We want hence to introduce a new class of evolution pairings. Thanks to the reparametrization invariance of the functional $\mathfrak{L}_\alpha$, we are more interested in the case where the evolution pairings are normal, i.e. when $|\mu'|_{w_\infty} = 1$ and we give the following definition.

DEFINITION 4.3.4. Let $(\nu, \mu) \in EP(\rho_0, \rho_1)$ be an evolution pairing. If there holds

$$w_\infty(\mu_t - \vartheta_t, \mu_{t+h} - \vartheta_t) \leq h, \text{ for } t \in I,\ h > 0, \tag{4.3.4}$$

then we say that (ν, μ) is a *special evolution pairing* and we denote by $SEP(\rho_0, \rho_1)$ the set of all special evolution pairings contained in $EP(\rho_0, \rho_1)$.

The intuitive idea behind evolution pairings is that, in going from μ_t to μ_{t+h}, the mass which is moving is (or should be) essentially that given by ν_t, which distributes over the difference between the total mass at the time $t + h$ and the mass which was already arrived at time t: property (4.3.4) expresses exactly the requirement that it is this mass that must move at most with unitary speed. Roughly speaking, this means that we can think of the quantity

$$\lim_{h\to 0^+} \frac{w_\infty(\mu_t - \vartheta_t, \mu_{t+h} - \vartheta_t)}{h},\ t \in I,$$

as a kind of velocity of the moving mass ν: the elements of $SEP(\rho_0, \rho_1)$ are exactly those for which ν is $1-$Lipschitz, in this sense.

What is remarkable is that, thanks to Lemma 4.3.2, we can assure that the class $SEP(\rho_0, \rho_1)$ is large enough: for example, it contains all the finitely atomic curves. Even more, it is sufficient that ρ_1 is atomic, then every evolution pairing connecting ρ_0 and ρ_1 is actually a special one. Indeed, we have the following:

PROPOSITION 4.3.5. *Let us take $(\nu,\mu) \in EP(\rho_0,\rho_1)$, with ρ_1 being the sum of (countably many) Dirac masses. Then $(\nu,\mu) \in SEP(\rho_0,\rho_1)$.*

PROOF. It is enough to apply, for any given t, the estimate (4.3.1) with $\mathbf{m} = \vartheta_t$. Since $\vartheta_t \le \rho_1$ and ρ_1 is atomic, the same is true for ϑ_t and hence we get

$$w_\infty(\mu_t - \vartheta_t, \mu_{t+h} - \vartheta_t) \le \int_t^{t+h} |\mu'_s|_{w_\infty}\, ds \le h, \text{ for every } h \ge 0. \qquad \square$$

EXAMPLE 4.3.6. Let us go back to the evolution pairing (ν,μ) given by Example 4.3.1. It is easily seen that this is not an element of $SEP(\rho_0,\rho_1)$: indeed, taking $t < 1/2$ we have for every $0 < h < 1/2 - t$

$$\mu_{t+h} - \vartheta_t = \mathscr{L}^1 \llcorner [-t+h,-t] + \mathscr{L}^1 \llcorner [t,t+h] + (1-2t-2h)\delta_{t+h},$$

so it is not difficult to see that $\Pi(\mu_t - \vartheta_t, \mu_{t+h} - \vartheta_t)$ contains only one element and

$$\frac{w_\infty(\mu_t - \vartheta_t, \mu_{t+h} - \vartheta_t)}{h} = \frac{2t+h}{h},$$

which goes to $+\infty$ as h approaches to 0, while $|\mu'|_{w_\infty} \equiv 1$.

If we see Example 4.3.1 from the point of view of the continuity equation, we may observe that the problem is that the velocity field v_t associated to this curve does not vanish on the part of μ_t which is supposed to be at rest, i.e. on ϑ_t. This is what allows for the teleport phenomenon and this is why the moving measure ν does not satisfy the same Lipschitz estimate as μ. We can also provide another example, that we will not develop in details, where the vector field v_t actually vanishes outside the support of ν_t, but its L^∞ norm is not the same if we consider μ or $\mu - \vartheta$ in the continuity equation, so that the Lipschitz constant increases (without blowing-up) while passing from μ to ν.

EXAMPLE 4.3.7. Consider the measures $\vartheta_0 = \frac{9}{10}\mathscr{L}^1 \llcorner [0,1]$ and $\nu_t = \frac{9}{10}\mathscr{L}^1 \llcorner [t, t+1/9]$ for $t \in [0, 8/9]$. Set $\mu_t = \vartheta_0 + \nu_t$ and then consider the vector field

$$v_t(x) = \overrightarrow{e_1} \cdot 1_{[t,t+\frac{9}{10}]}(x),$$

which at every time t moves rightwards the particles of the interval $[t, t+9/10]$. We have

$$\partial_t \nu_t + \mathrm{div}_x(v_t \nu_t) = 0,$$

and it is quite easy to see that $\|v_t\|_{L^\infty(\nu_t)} = 1 = |\nu'_t|_{w_\infty}$. On the other hand one can see that μ is a solution of the continuity equation with velocity field given by $1/2\, v_t(x)$, that is we have

$$\partial_t \mu_t + \mathrm{div}_x\left(\frac{1}{2} v_t \mu_t\right) = 0,$$

as a consequence of $v_t\, \nu_t = 1/2\, v_t\, \mu_t$. This shows that $|\mu'_t|_{w_\infty} \le \|1/2\, v_t\|_{L^\infty(\mu_t)} = 1/2$, i.e. the speed of the two curves μ and ν is finite in both cases but different.

We then turn our attention to the functional

$$\overline{\mathfrak{L}}_\alpha(\nu,\mu) = \int_0^\infty g_\alpha(\nu_t)\, dt, \ (\nu,\mu) \in SEP(\rho_0,\rho_1),$$

for which the following existence result is almost straightforward. As we said, to give a cleaner definition of the class SEP and of the functional, we decided to stick to the case where the velocity $|\mu'|_{w_\infty}$ (nor, in any sense, $|\nu'|$) does not appear explicitly in the criterion to be minimized, but only in the constraints.

THEOREM 4.3.8. *The minimization problem*

$$\inf_{(\nu,\mu)\in SEP(\rho_0,\rho_1)} \overline{\mathfrak{L}}_\alpha(\nu,\mu), \tag{4.3.5}$$

admits a solution, provided that there exists an admissible special evolution pairing $(\overline{\nu},\overline{\mu})$ *having finite* $\overline{\mathfrak{L}}_\alpha$.

PROOF. It should be clear that it is enough to show that $SEP(\rho_0,\rho_1)$ is closed: then one has to simply reproduce step by step the proof of Proposition 4.2.8, taking into account the fact that every special evolution pairing (ν,μ) having finite $\overline{\mathfrak{L}}_\alpha$, has to satisfy

$$\lim_{t\to\infty} |\nu_t|(\Omega) = 0.$$

Concerning the closedness of $SEP(\rho_0,\rho_1)$, it is enough to use the fact that the distance w_∞ is lower semicontinuous with respect to the $*-$weak convergence of measures, as already pointed out, so that property (4.3.4) easily pass to the limit. □

REMARK 4.3.9. If one wants Theorem 4.3.8 to be interesting, one has to provide conditions for the existence of special evolution pairings with finite energy. The idea is the following: suppose that ρ_1 is a probability measure which is irrigable in the sense of Xia, Solimini et al. This means

$$d_\alpha(\rho_0,\rho_1) = \min\{M^*_\alpha(\Phi)\,:\,\operatorname{div}\Phi = \rho_0-\rho_1\} < +\infty,$$

and thanks to the relaxed definition by Xia, there exists a sequence of finite graphs $\mathfrak{g}_n$, corresponding to traffic plans Q_n, such that $\sup_n E_\alpha(Q_n) < +\infty$ and $(e_\infty)_\sharp Q_n = \rho_1^n \rightharpoonup \rho_1$, with the measures ρ_1^n atomic (see Chapter 3, Proposition 3.3.5). Then one uses the results of Section 5 to see that these traffic plans give rise to some evolution pairings (ν^n,μ^n) which are actually special evolution pairings in $SEP(\rho_0,\rho_1^n)$ and whose energy is the same as $E_\alpha(Q_n)$. Up to subsequences, thanks to the semicontinuity of $\overline{\mathfrak{L}}_\alpha$ and to the closedness result of Lemma 2.6.1, one can get $(\nu^n,\mu^n)\to(\nu,\mu)$ with $(\nu,\mu)\in SEP(\rho_0,\rho_1)$ and $\overline{\mathfrak{L}}_\alpha(\nu,\mu) < +\infty$.

4. Characterization of $SEP(\rho_0,\rho_1)$

The main tool in order to compare $\overline{\mathfrak{L}}_\alpha$ with a Gilbert-Steiner energy, will be a complete characterization of the special evolution pairings, in terms of the Lipschitz curves of the base space. So our aim now is to give a refinement to the case of SEP of the result by Lisini given by Theorem 1.5.6 of Chapter 1, characterizing $p-$absolutely continuous curves in the Wasserstein space $\mathcal{W}_p(\Omega)$ in terms of the $p-$absolutely continuous curves of the ambient space Ω: the main difference (apart from the fact that we explicitly refer to the case $p=+\infty$) is the characterization of the moving part ν in terms of the $1-$Lipschitz curves in Ω which at every fixed time t are still moving.

In order to achieve our scope, we have to start with a couple of technical Lemmas: they are nothing but *ad hoc* adaptations of the gluing Lemma (see Chapter 1, Lemma 1.3.2). First of all we prove the existence of the composition of two transport plans, that takes into account the fact that the mass which arrives at destination must stay in place: at this level, this sentence could sound mysterious, but in the proof of Theorem 4.4.4 it should become clearer. We point out that in the following, given two positive Borel measures $\nu_1, \nu_2 \in \mathcal{M}^+(\Omega)$ with the same mass, by $\Pi(\nu_1, \nu_2)$ we will denote the set of all positive Borel measures over the product space $\Omega \times \Omega$, having fixed marginals ν_1 and ν_2.

Lemma 4.4.1 (Modified gluing lemma). *Let $(\mu_1, \mu_2, \mu_3) \in \mathscr{P}(\Omega)$ and $(\nu_1, \nu_2, \nu_3) \in \mathcal{M}_1^+(\Omega)$ such that*

$$\nu_i \leq \mu_i, \ i = 1, 2, 3,$$

and suppose that, setting $\vartheta_i = \mu_i - \nu_i$, we have $\vartheta_1 \leq \vartheta_2 \leq \vartheta_3$. For every $\gamma_{1,2} \in \Pi(\mu_1 - \vartheta_1, \mu_2 - \vartheta_1)$ and $\gamma_{2,3} \in \Pi(\mu_2 - \vartheta_2, \mu_3 - \vartheta_2)$, there exists $\gamma \in \mathscr{P}(\Omega \times \Omega \times \Omega)$ with the following properties:

(i) $(\pi_{i,i+1})_\sharp \gamma = \gamma_{i,i+1} + (\mathrm{Id} \times \mathrm{Id})_\sharp(\vartheta_i)$, *for $i = 1, 2$;*

(ii) $(\pi_i)_\sharp(\gamma 1_{S_i}) \geq \vartheta_i$, *for $i = 1, 2$, where the set S_i is given by*

$$S_i = \{(x_1, x_2, x_3) \in \Omega \times \Omega \times \Omega \ : \ x_j = x_i, \quad \textit{for } j \geq i\}.$$

Proof. We will make use of the Disintegration Theorem (see [**38**, Chapter III]). First of all, we define

$$\tilde{\gamma}_{1,2} = \gamma_{1,2} + \gamma_{1,2}^0 = \gamma_{1,2} + (\mathrm{Id} \times \mathrm{Id})_\sharp(\vartheta_1),$$

and

$$\tilde{\gamma}_{2,3} = \gamma_{2,3} + \gamma_{2,3}^0 = \gamma_{2,3} + (\mathrm{Id} \times \mathrm{Id})_\sharp(\vartheta_2),$$

which are actually elements of $\Pi(\mu_1, \mu_2)$ and $\Pi(\mu_2, \mu_3)$, respectively. Then we disintegrate $\gamma_{1,2}$ with respect to the x_2 variable, that is

$$\gamma_{1,2} = \int \xi_{x_2}^1 \, d(\mu_2 - \vartheta_1)(x_2) = \int \xi_{x_2}^1 \, d(\mu_2 - \vartheta_2)(x_2) + \int \xi_{x_2}^1 \, d(\vartheta_2 - \vartheta_1)(x_2),$$

where for $(\mu_2 - \vartheta_1)-$a.e. $x_2 \in \Omega$, $\xi_{x_2}^1$ is a Borel probability measure on Ω and equally for $\gamma_{1,2}^0$, thus obtaining

$$\gamma_{1,2}^0 = \int \eta_{x_2}^1 \, d\vartheta_1(x_2).$$

On the other hand, we disintegrate $\gamma_{2,3}$ and $\gamma_{2,3}^0$ with respect to the x_1 variable, that is

$$\gamma_{2,3} = \int \xi_{x_2}^3 \, d(\mu_2 - \vartheta_2)(x_2),$$

$$\gamma_{2,3}^0 = \int \eta_{x_2}^3 \, d\vartheta_2(x_2) = \int \eta_{x_2}^3 \, d(\vartheta_2 - \vartheta_1)(x_2) + \int \eta_{x_2}^3 \vartheta_1(x_2).$$

Observe that actually, by construction we have

$$\eta_{x_2}^1 = \delta_{x_2}, \quad \text{for } \vartheta_1\text{−a.e. } x_2 \in \Omega, \tag{4.4.1}$$

and

$$\eta_{x_2}^3 = \delta_{x_2}, \quad \text{for } \vartheta_2\text{−a.e. } x_2 \in \Omega, \tag{4.4.2}$$

We can rewrite everything as follows

$$\tilde{\gamma}_{1,2} = \int \xi_{x_2}^1 \, d(\mu_2 - \vartheta_2)(x_2) + \int \xi_{x_2}^1 \, d(\vartheta_2 - \vartheta_1)(x_2) + \int \eta_{x_2}^1 \, d\vartheta_1(x_2),$$

$$\tilde{\gamma}_{2,3} = \int \xi^3_{x_2}\, d(\mu_2 - \vartheta_2)(x_2) + \int \eta^3_{x_2}\, d(\vartheta_2 - \vartheta_1)(x_2) + \int \eta^3_{x_2}\, d\vartheta_1(x_2),$$

that is we have "piecewise" disintegrated with respect to their common marginals our transport plans. Then it is natural to glue this two decompositions as follows

$$\gamma = \int \xi^1_{x_2} \otimes \xi^3_{x_2}\, d(\mu_2 - \vartheta_2)(x_2) + \int \xi^1_{x_2} \otimes \eta^3_{x_2}\, (\vartheta_2 - \vartheta_1)(x_2) + \int \eta^1_{x_2} \otimes \eta^3_{x_2}\, d\vartheta_1(x_2), \tag{4.4.3}$$

and it is straightforward to verify that γ has the desired properties: (i) is trivially satisfied, while concerning (ii) let us observe that for every Borel set $A \subset \Omega$, we have

$$\begin{aligned}(\pi_1)_\#(\gamma 1_{S_1})(A) = \gamma\left(\{(a,a,a)\ :\ a \in A\}\right) &\geq \int_A \eta^1_{x_2}(\{x_2\})\eta^3_{x_2}(\{x_2\})\ d\vartheta_1(x_2)\\ &= \int_A d\vartheta_1(x_2) = \vartheta_1(A),\end{aligned}$$

where we have used (4.4.1) and (4.4.2) and the fact that $\vartheta_1 \leq \vartheta_2$. In the end, we have proven property (ii) for $i = 1$, while for $i = 2$ the proof is straightforward. □

REMARK 4.4.2. Observe that the probability measure γ given by (4.4.3) can also be written (with the convention $\vartheta_0 = 0$) as

$$\gamma = \int \xi_{x_3}\, d(\mu_3 - \vartheta_3)(x_3) + \sum_{i=1}^{3} \int \eta^i_{x_3}\, d(\vartheta_i - \vartheta_{i-1})(x_3), \tag{4.4.4}$$

for suitable Borel families of probability measures $\{\xi_{x_3}\}_{x_3\in\Omega}$ and $\{\eta^i_{x_3}\}_{x_3\in\Omega}$ on $\Omega\times\Omega$ such that $\eta^3_{x_3} = \xi_{x_3}$ for ϑ_3-a.e. $x \in \Omega$ and

$$\eta^1_{x_3} = \delta_{(x_3,x_3)}, \quad \text{for } \vartheta_1\text{−a.e. } x_3 \in \Omega,$$

and

$$(\pi_2)_\sharp \eta^2_{x_3} = \delta_{x_3}, \quad \text{for } \vartheta_2\text{−a.e. } x_3 \in \Omega.$$

Indeed, it is sufficient to observe that by construction

$$(\pi_3)_\sharp\left(\int \xi^1_{x_2} \otimes \xi^3_{x_2}\, d(\mu_2 - \vartheta_2)(x_2)\right) = (\mu_3 - \vartheta_3) + (\vartheta_3 - \vartheta_2),$$

then we can disintegrate this measure with respect to the x_3 variable, thus obtaining the existence of a Borel family of probability measures $\{\xi_{x_3}\}_{x_3\in\Omega}$ on the product space $\Omega\times\Omega$ such that

$$\int_\Omega \xi^1_{x_2} \otimes \xi^3_{x_2}\, d(\mu_2 - \vartheta_2)(x_2) = \int_\Omega \xi_{x_3}\, d(\mu_3 - \vartheta_3)(x_3) + \int_\Omega \xi_{x_3}\, d(\vartheta_3 - \vartheta_2)(x_3).$$

Equally, taking into account

$$(\pi_3)_\sharp\left(\int \xi^1_{x_2} \otimes \eta^3_{x_2}\, d(\vartheta_2 - \vartheta_1)(x_2)\right) = \vartheta_2 - \vartheta_1,$$

$$(\pi_3)_\sharp\left(\int \eta^1_{x_2} \otimes \eta^3_{x_2}\, d\vartheta_1(x_2)\right) = \vartheta_1,$$

and disintegrating with respect to the x_3 variable, we obtain the desired representation (4.4.4), keeping in mind (4.4.1) and (4.4.2).

The previous result can be easily generalized to every n-uple of probability measures. More precisely, we have the following:

LEMMA 4.4.3. *For $n \geq 3$, let $\{\mu_i\}_{i=1}^n \subset \mathscr{P}(\Omega)$ and $\{\nu_i\}_{i=1}^n \subset \mathcal{M}_1^+(\Omega)$ be such that*

$$\nu_i \leq \mu_i, \quad \text{for every } i = 1, \dots, n,$$

and suppose that, setting $\vartheta_i = \mu_i - \nu_i$, we have $\vartheta_i \leq \vartheta_{i+1}$. For every $\gamma_{i,i+1} \in \Pi(\nu_i, \mu_{i+1} - \vartheta_i)$, with $i = 1, \dots, n-1$, there exists $\gamma \in \mathscr{P}(\Omega^n)$ with the following properties:

(i) $(\pi_{i,i+1})_\sharp \gamma = \gamma_{i,i+1} + (\mathrm{Id} \times \mathrm{Id})_\sharp(\vartheta_i)$, *for $i = 1, \dots, n-1$;*
(ii) $(\pi_i)_\sharp(\gamma 1_{S_i}) \geq \vartheta_i$, *for $i = 1, \dots, n-1$, where the set S_i is given by*

$$S_i = \{(x_1, \dots, x_n) \in \Omega^n \ : \ x_j = x_i, \quad \text{for } j \geq i\}.$$

Moreover γ can be written as

$$\gamma = \int \xi_{x_n} \, d(\mu_n - \vartheta_n)(x_n) + \sum_{i=1}^n \int \eta^i_{x_n} \, d(\vartheta_i - \vartheta_{i-1})(x_n), \tag{4.4.5}$$

where $\xi_{x_n}, \eta^i_{x_n} \in \mathscr{P}(\Omega^{n-1})$ and every $\eta^i_{x_n}$ is such that

$$(\pi_{i,\dots,n-1})_\sharp \eta^i_{x_n} = \delta_{(x_n,\dots,x_n)}, \quad \text{for } \vartheta_i\text{-a.e. } x_n \in \Omega, \tag{4.4.6}$$

the function $\pi_{i,\dots,n-1}$ being the projection on the $(x_i, \dots, x_{n-1})$ coordinates.

PROOF. We proceed by induction on n, the thesis being true for $n = 3$ thanks to Lemma 4.4.1 and Remark 4.4.2.

Suppose now that the assertion is true for n, that is there exists a probability measure $\gamma \in \mathscr{P}(\Omega^n)$ with the required properties and consider the case $n+1$. As in the proof of Lemma 4.4.1, we can define

$$\tilde{\gamma}_{n,n+1} = \gamma_{n,n+1} + \gamma^0_{n,n+1} = \gamma_{n,n+1} + (\mathrm{Id} \times \mathrm{Id})_\sharp \vartheta_n,$$

and then we disintegrate $\gamma_{n,n+1}$ and $\gamma^0_{n,n+1}$ with respect to x_n, thus getting

$$\begin{aligned}
\tilde{\gamma}_{n,n+1} &= \int \xi^{n+1}_{x_n} \, d(\mu_n - \vartheta_n)(x_n) + \int \eta^{n+1}_{x_n} \, d\vartheta_n(x_n) \\
&= \int \xi^{n+1}_{x_n} \, d(\mu_n - \vartheta_n)(x_n) + \sum_{i=1}^n \int \eta^{n+1}_{x_n} \, d(\vartheta_i - \vartheta_{i-1})(x_n)
\end{aligned}$$

where $\eta^{n+1}_{x_n} = \delta_{x_n}$ for ϑ_n–a.e. $x_n \in \Omega$. Then using the decomposition (4.4.5) for γ, we can define

$$\widehat{\gamma} = \int \xi_{x_n} \otimes \xi^{n+1}_{x_n} \, d(\mu_n - \vartheta_n)(x_n) + \sum_{i=1}^n \int \eta^i_{x_n} \otimes \eta^{n+1}_{x_n} \, d(\vartheta_i - \vartheta_{i-1})(x_n),$$

which is an element of $\mathscr{P}(\Omega^{n+1})$. It is straightforward to see that $\widehat{\gamma}$ satisfies property (i), so let us show that also (ii) holds true: for every Borel subset $A \subset \Omega$ we get

$$\begin{aligned}(\pi_j)_\sharp(\widehat{\gamma}1_{S_j})(A) &= \widehat{\gamma}(\{(x_1,\dots,x_{j-1},a\dots,a)\ :\ x_1,\dots,x_{j-1}\in\Omega,\ a\in A\}) \\ &\geq \sum_{i=1}^{j}\int_A (\pi_{j,\dots,n-1})_\# \eta^i_{x_n}(\{(x_n,\dots,x_n)\})\,\eta^{n+1}_{x_n}(\{x_n\})\,d(\vartheta_i-\vartheta_{i-1})(x_n) \\ &= \sum_{i=1}^{j}(\vartheta_i-\vartheta_{i-1})(A) = \vartheta_j(A), \quad \text{for every } j=1,\dots,n,\end{aligned}$$

where we have used property (4.4.6). To conclude the proof, it remains to show that $\widehat{\gamma}$ can be decomposed as in (4.4.5): observe that by construction we have

$$(\pi_{n+1})_\sharp\left(\int \xi_{x_n}\otimes\xi^{n+1}_{x_n}\,d(\mu_n-\vartheta_n)(x_n)\right) = \mu_{n+1}-\vartheta_n = (\mu_{n+1}-\vartheta_{n+1})+(\vartheta_{n+1}-\vartheta_n),$$

so that as in Remark 4.4.2 we can disintegrate $\int \xi_{x_N}\otimes\xi^{N+1}_{x_n}\,d(\mu_n-\vartheta_n)$ with respect to the x_{n+1} variable, thus obtaining

$$\int \xi_{x_n}\otimes\xi^{n+1}_{x_n}\,d(\mu_n-\vartheta_n)(x_n) = \int \xi_{x_{n+1}}\,d(\mu_{n+1}-\vartheta_{n+1})(x_{n+1}) + \int \xi_{x_{n+1}}\,d(\vartheta_{n+1}-\vartheta_n)(x_{n+1}),$$

where for $(\mu_{n+1}-\vartheta_n)-$a.e. $x_{n+1}\in\Omega$, $\xi_{x_{n+1}}$ is a Borel probability measure over the space Ω^n. The same can be done for each term

$$\int \eta^i_{x_n}\otimes\eta^{n+1}_{x_n}\,d(\vartheta_i-\vartheta_{i-1})(x_n),$$

then taking into account that $\eta^{n+1}_{x_n}=\delta_{x_n}$ for ϑ_n-a.e. $x\in\Omega$ and that $\eta^i_{x_n}$ satisfies (4.4.6) by hypothesis, we can conclude. □

We now have all the elements in order to prove the first main result of this section.

THEOREM 4.4.4. *Let $(\nu,\mu)\in SEP(\rho_0,\rho_1)$. Then there exists $Q\in\mathscr{P}(\mathrm{Lip}_1(I;\Omega))$ such that*

$$(e_t)_\sharp Q = \mu_t \quad \text{and} \quad (e_t)_\sharp Q^t \leq \nu_t,\ t\in I,$$

where $Q^t = Q\llcorner\{\sigma\in\mathrm{Lip}_1(I;\Omega)\ :\ T(\sigma)\geq t\}$.

PROOF. We fix $M\in\mathbb{N}$ and then for every $n\in\mathbb{N}$ we take a dyadic partition

$$t_{i,n} = \frac{M}{2^n}\,i,\ i=0,1,\dots,2^n,$$

of the interval $[0,M]$. Indicating as always

$$\vartheta_t := \mu_t-\nu_t,$$

we take $\tilde{\gamma}_{i,i+1}\in\Pi(\nu_{t_{i,n}},\mu_{t_{i+1,n}}-\vartheta_{t_{i,n}})$ to be optimal for w_∞, that is

$$w_\infty(\nu_{t_{i,n}},\mu_{t_{i+1,n}}-\vartheta_{t_{i,n}}) = \sup\{|x-y|\ :\ (x,y)\in\mathrm{spt}(\tilde{\gamma}_{i,i+1})\} \leq \frac{M}{2^n},$$

where we used that $(\nu,\mu)\in SEP(\rho_0,\rho_1)$. Then we define $\gamma_{i,i+1}\in\Pi(\mu_{t_{i,n}},\mu_{t_{i+1,n}})$ by

$$\gamma_{i,i+1} = \tilde{\gamma}_{i,i+1} + (\mathrm{Id}\times\mathrm{Id})_\sharp\vartheta_{t_{i,n}}.$$

Let $\gamma_M^n \in \mathscr{P}(\Omega^{2^n+1})$ be the multi-transport plan given by Lemma 4.4.3 such that:

(i) $(\pi_{i,i+1})_\sharp \gamma_M^n = \gamma_{i,i+1}$;

(ii) $(\pi_i)_\sharp(\gamma_M^n 1_{S_i}) \geq \vartheta_{t_{i,n}}$, where $S_i = \{\mathbf{x} = (x_0,\dots,x_{2^n}) \in \Omega^n \ : \ x_j = x_i, \text{ for } j \geq i\}$.

We now define the application

$$\begin{array}{rccc} \Theta^n: & \Omega^{2^n+1} & \to & \mathrm{Lip}([0,M];\Omega) \\ & \mathbf{x} & \mapsto & \Theta^n_{\mathbf{x}}, \end{array}$$

where for every $\mathbf{x} = (x_0,\dots,x_{2^n}) \in \Omega^{2^n+1}$, the curve $\Theta^n_{\mathbf{x}}$ is piecewise affine and given by

$$\Theta^n_{\mathbf{x}}(t) = \frac{t_{i+1,n}-t}{t_{i+1,n}-t_{i,n}} x_i + \frac{t-t_{i,n}}{t_{i+1,n}-t_{i,n}} x_{i+1}, \ t \in [t_{i,n},t_{i+1,n}], \ i \in \{0,\dots,2^n-1\},$$

and we further set

$$Q_M^n = (\Theta^n)_\sharp \gamma_M^n \in \mathscr{P}(C([0,M];\Omega)).$$

By construction, it is straightforward to see that every Q_M^n is concentrated on $\mathrm{Lip}_1([0,M];\Omega)$, the latter being a compact space. This in turn implies that the sequence $\{Q_M^n\}_{n\in\mathbb{N}}$ narrowly converges (up to subsequences) to an element Q_M of $\mathscr{P}(\mathrm{Lip}_1([0,M];\Omega))$, by means of Prokhorov's Theorem.

We now show that $(e_t)_\# Q_M = \mu_t$ for every $t \in [0,M]$: first observe that by its very definition, the sequence $\{Q_M^n\}_{n\in\mathbb{N}}$ satisfies

$$\mu_{t_{i,n}} = (e_{t_{i,n}})_\# Q_M^n, \ i = 0,1,\dots,2^n.$$

On the other hand, thanks to the fact that we are considering dyadic partitions of $[0,M]$, we have for every $k < n$

$$\{t_{j,k}\}_{j=0}^{2^k} \subset \{t_{i,n}\}_{i=0}^{2^n},$$

so that for every $k < n$

$$\mu_{t_{i,k}} = (e_{t_{i,k}})_\sharp Q_M^n, \ i = 0,1,\dots,2^k.$$

Letting n go to ∞, we then obtain, for every k, the following equalities

$$\mu_{t_{i,k}} = (e_{t_{i,k}})_\sharp Q_M, \ i = 0,1,\dots,2^k.$$

We have proven that the two uniformly continuous functions μ and $(e_{(\cdot)})_\# Q_M$ coincide on the points $\{t_{j,k}\}_{j=0}^{2^k}$, for every $k \in \mathbb{N}$, thus giving the equality on $[0,M]$ of these functions.

Before going on, we define the following subset of $[0,M]$

$$\mathcal{N} := \left\{ t \in [0,M] \, : \, \text{either } Q_M(T^{-1}(\{t\})) > 0 \ \text{ or there exists } n \text{ such that } Q_M^n(T^{-1}(\{t\})) > 0 \right\},$$

that is $\mathcal{N}$ is the set of times such that $\{\sigma \in \mathrm{Lip}_1([0,M];\Omega) \, : \, T(\sigma) = t\}$ is charged by at least one of the measures Q_M or Q_M^n. Due to the fact that as t varies in $[0,M]$ these sets $T^{-1}(\{t\})$ constitute a partition of the whole space, we observe that $\mathcal{N}$ must be at most countable. We now set

$$Q_M^t = Q_M \llcorner \{\sigma \, : \, T(\sigma) \geq t\}, \ Q_M^{n,t} = Q_M^n \llcorner \{\sigma \, : \, T(\sigma) \geq t\}, \ n \in \mathbb{N},$$

and we first notice that if $t \notin \mathcal{N}$, we have $\liminf_{n\to\infty} Q_M^{n,t} \geq Q_M^t$, in the sense that for every continuous and positive test function φ there holds

$$\liminf_{n\to\infty} \int \varphi \, dQ_M^{n,t} \geq \int \varphi \, dQ_M^t.$$

This is the same as saying that any possible limit measure $\widetilde{Q}$ of a subsequence of $Q_M^{n,t}$ must be larger than Q_M^t. To prove such a property, it is sufficient to notice that this is true if $Q_M^{n,t}$ and Q_M^t are replaced with $1_{\{T>t\}} \cdot Q_M^n$ and $1_{\{T>t\}} \cdot Q_M$, respectively, since the function $1_{\{T>t\}}$ is l.s.c. and this modification may be performed for free if the set $T^{-1}(t)$ is negligible for all these measures.

In order to prove that $(e_t)_\sharp Q_M^t \le \nu_t$, we first observe that using property (ii) of $\{\gamma_M^n\}$ we get that

$$(e_{t_{i,n}})_\sharp Q_M^{n,t_{i,n}} \le \nu(t_{i,n}). \tag{4.4.7}$$

Let us give a brief justification of (4.4.7): indeed, we have

$$\begin{aligned}\int_{C([0,M];\Omega)} \varphi(\sigma(t_{i,n}))\, dQ_M^{n,t_{i,n}}(\sigma) \ge & \int_{C([0,M];\Omega)} \varphi(\sigma(t_{i,n}))\, dQ_M^n(\sigma) \\ & - \int_{\{\sigma\,:\,T(\sigma)\le t_{i,n}\}} \varphi(\sigma(t_{i,n}))\, dQ_M^n(\sigma),\end{aligned}$$

and the first integral in the right-hand side is just the integral of φ with respect to the measure $\mu_{t_{i,n}}$, while for the second we observe that

$$\begin{aligned}\int_{\{\sigma\,:\,T(\sigma)\le t_{i,n}\}} \varphi(\sigma(t_{i,n}))\, dQ_M^n(\sigma) = & \int_{\{\mathbf{x}\,:\,x_j=x_i, \ \text{ for } j\ge i\}} \varphi(\Theta_{\mathbf{x}}^n(t_{i,n}))\, d\gamma_M^n(\mathbf{x}) \\ \ge & \int_\Omega \varphi(x)\, d\vartheta_{t_{i,n}}(x),\end{aligned}$$

having used the definition of Θ^n and property (ii) in the last inequality.

In conclusion, using $\nu_{t_{i,n}} = \mu_{t_{i,n}} - \vartheta_{t_{i,n}}$ we have shown the validity of (4.4.7). Observe moreover that we have $Q_M^{n,t} \le Q_M^{n,t_{i,n}}$ for every $t \ge t_{i,n}$, and using again the fact that the partition under consideration is dyadic, in the end we get

$$(e_{t_{i,k}})_\sharp Q_M^{n,t} \le \nu_{t_{i,k}}, \quad \text{for every } t \ge t_{i,k},$$

for every $k < n$. Taking the limit as n goes to ∞, and using the "semicontinuity" we addressed before, i.e. the fact $\liminf_{n\to\infty} Q_M^{n,t} \ge Q_M^t$, which is true for $t \notin \mathcal{N}$, we get for every i and k

$$(e_{t_{i,k}})_\sharp Q_M^t \le \nu_{t_{i,k}}, \quad \text{for every } t \notin \mathcal{N},\ t \ge t_{i,k}.$$

The condition $t \notin \mathcal{N}$ may be withdrawn, if for $t > t_{i,k}$ one takes $s \in (t_{i,k}, t) \setminus \mathcal{N}$ and uses the inequality $Q_M^t \le Q_M^s$, which gives

$$(e_{t_{i,k}})_\sharp Q_M^t \le (e_{t_{i,k}})_\# Q_M^s \le \nu_{t_{i,k}}, \quad \text{for every } t > t_{i,k}.$$

It is then sufficient to consider a sequence of dyadic numbers $t_{i,k}$ converging to t from the left: we then have $\nu_{t_{i,k}} \to \nu_t$ because of the assumption of left continuity of ν and $(e_{t_{i,k}})_\sharp Q_M^t \to (e_t)_\sharp Q_M^t$ because Q_M^t is a fixed measure on $\mathrm{Lip}_1([0,M];\Omega)$ and the maps $e_{t_{i,k}}$ converge uniformly to e_t on this set. Actually, we can also say

$$w_\infty((e_{t_{i,k}})_\sharp Q_M^t, (e_t)_\sharp Q_M^t) \le |t_{i,k} - t|,$$

thanks to the Lipschitz property of the curves in $\mathrm{Lip}_1([0,M];\Omega)$. This gives

$$(e_t)_\sharp Q_M^t \le \nu_t, \quad \text{for every } t \in [0,M]. \tag{4.4.8}$$

Finally, we have to take the limit as $M \to +\infty$: defining the continuous mapping

$$\Phi_M : \mathrm{Lip}_1([0,M];\Omega) \to \mathrm{Lip}_1(I;\Omega),$$

such that for every $\sigma \in \mathrm{Lip}_1([0,M];\Omega)$, the curve $\Phi_M(\sigma)$ is given by

$$\Phi_M(\sigma)(t) = \begin{cases} \sigma(t), & \text{if } t \leq M, \\ \sigma(M), & \text{if } t > M, \end{cases}$$

we set $\widetilde{Q}_M = (\Phi_M)_\sharp Q_M \in \mathrm{Lip}_1(I;\Omega)$; then the sequence $\{Q_M\}_{M\in\mathbb{N}}$ is narrowly converging (up to subsequences), again thanks to the compactness of the space $\mathrm{Lip}_1(I;\Omega)$. If we call Q its limit, it is not difficult to see that we have $(e_t)_\sharp \widetilde{Q}_M = \mu_t$ on $[0,M]$ and passing to the limit, we obtain that the same holds true for Q on I. Moreover if $\widetilde{Q}^t_M = \tilde{Q}_M \llcorner \{\sigma : T(\sigma) \geq t\}$, then using the fact

$$(e_t)_\sharp \widetilde{Q}^t_M \leq \nu_t, \quad \text{for } t \in [0,M],$$

which is actually equivalent to (4.4.8), and taking again the limit as M goes to $+\infty$, we can show that

$$(e_t)_\sharp Q^t \leq \nu_t, \quad \text{for } t \in I \setminus \widetilde{\mathcal{N}},$$

where the negligible set $\widetilde{\mathcal{N}}$ where the inequality could not hold is, as before, the countable set of times such that $T^{-1}(\{t\})$ is charged by at least one of the measures $\widetilde{Q}_M$ or Q. After that, we consider a general t and $s < t$ with $s \notin \widetilde{\mathcal{N}}$: we have

$$(e_s)_\sharp Q^t \leq (e_s)_\sharp Q^s \leq \nu_s.$$

Taking the limit $s \nearrow t$ we get, as before, $(e_t)_\sharp Q^t \leq \nu_s$, which concludes the proof. □

The next result of this section states that the previous Theorem can be reverted, thus giving a nice correspondence between $SEP(\rho_0,\rho_1)$ and the $1-$Lipschitz curves of Ω.

THEOREM 4.4.5. *Let $Q \in TP(\rho_0,\rho_1)$ be a traffic plan. For every $t \in I$, we set $Q^t = Q \llcorner \{T(\sigma) \geq t\}$ and we define*

$$\mu_t = (e_t)_\sharp Q, \quad \nu_t = (e_t)_\sharp Q^t,$$

then $(\nu,\mu) \in SEP(\rho_0,\rho_1)$.

PROOF. The fact that $\mu \in \mathrm{Lip}_1(I;\mathcal{W}_\infty(\Omega))$ is straightforward, since for every (t,s) the measure $(e_t,e_s)_\sharp Q$ is a transport plan between μ_t and μ_s, providing a cost smaller than $|t-s|$.

We then observe that the set

$$\{\sigma \in \mathrm{Lip}_1(I;\Omega) : T(\sigma) \geq t\},$$

is Borel measurable for every $t \in I$, thanks to the lower semicontinuity of T, so that ν is well-defined. We have to show that $\vartheta_t = \mu_t - \nu_t = (e_t)_\sharp (Q \llcorner \{\sigma \in \mathrm{Lip}_1(I;\Omega) : T(\sigma) < t\})$ is nondecreasing and left continuous. To see the monotonicity property, consider a positive test function $\varphi \in C(\Omega)$ and

$s \leq t$:

$$\begin{aligned}\int_\Omega \varphi(x)\, d\vartheta_s(x) &= \int_{\{\sigma\,:\,T(\sigma)<s\}} \varphi(\sigma(s))\, dQ(\sigma)\\ &= \int_{\{\sigma\,:\,T(\sigma)<s\}} \varphi(\sigma(t))\, dQ(\sigma)\\ &\leq \int_{\{\sigma\,:\,T(\sigma)<t\}} \varphi(\sigma(t))\, dQ(\sigma) = \int_\Omega \varphi(x)\, d\vartheta_t(x).\end{aligned}$$

Once one has the monotonicity, weak continuity is the same as strong continuity and we can turn to prove that $\lim_{s\nearrow t} |\vartheta_s - \vartheta_s|(\Omega) = \lim_{s\nearrow t} (|\vartheta_t|(\Omega) - |\vartheta_s|(\Omega)) = 0$. It is hence sufficient to prove that the mass of ϑ is left continuous, which is the same as looking at the mass of $Q \llcorner \{\sigma \in \mathrm{Lip}_1(I;\Omega) \,:\, T(\sigma) < t\}$. This corresponds to saying that

$$\{\sigma \in \mathrm{Lip}_1(I;\Omega) \,:\, T(\sigma) < t\} = \bigcup_{s<t} \{\sigma \in \mathrm{Lip}_1(I;\Omega) \,:\, T(\sigma) < s\},$$

which is evident.

In order to check that $\nu \preceq \mu$ we notice that property (E1) is evidently verified, since $Q^t \leq Q$, while property (E2) has already been verified when we proved that ϑ is increasing. Hence (ν, μ) is an evolution pairing, which clearly connects ρ_0 and ρ_1. We have to verify that actually it is a special evolution pairing: fixed $h > 0$, let us call

$$\gamma = (e_t, e_{t+h})_\sharp Q^t.$$

It is easy to check that this is a transport plan between $\mu_t - \vartheta_t$ and $\mu_{t+h} - \vartheta_t$ (just check that $(\pi_2)_\sharp \gamma = \mu_{t+h} - \mu_t + \nu_t$). Using the definition of w_∞ and the fact that Q^t is a measure over $\mathrm{Lip}_1(I;\Omega)$, we get

$$w_\infty(\nu_t, \mu_{t+h} - \mu_t + \nu_t) \leq \gamma\text{-}\operatorname*{ess\,sup}_{(x,y)\in\Omega\times\Omega} |x-y| = Q^t\text{-}\operatorname*{ess\,sup}_{\sigma} |\sigma(t) - \sigma(t+h)| \leq h,$$

which finally gives $(\nu, \mu) \in SEP(\rho_0, \rho_1)$. □

5. Equivalence between the models

Up to now, we have collected enough elements to compare our energy $\overline{\mathfrak{L}}_\alpha$ with a Gilbert-Steiner one. We point out that, in view of the results recalled in the previous chapter (Section 3), we are allowed to work with the energy

$$\widetilde{S}_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^{T(\sigma)} |(\sigma(t), t)|_Q^{\alpha-1}\, dt\, dQ(\sigma),$$

for which the minimum problem on $TP(\rho_0, \rho_1)$, with $\rho_0 = \delta_{x_0}$, is equivalent to all the other Lagrangian models and to Xia's model. Then the main result of the chapter is the following:

THEOREM 4.5.1. *Given $\rho_0 = \delta_{x_0}$ and $\rho_1 \in \mathscr{P}(\Omega)$, we get*

$$\min_{SEP(\rho_0,\rho_1)} \overline{\mathfrak{L}}_\alpha = \min_{TP(\rho_0,\rho_1)} \widetilde{S}_\alpha.$$

Moreover, given any optimal traffic plan $Q \in TP(\rho_0, \rho_1)$, the special evolution pairing provided by Theorem 4.4.5 is optimal, and conversely, given an optimal special evolution pairing $(\nu, \mu) \in SEP(\rho_0, \rho_1)$, the construction of Theorem 4.4.4 provides an optimal traffic plan.

PROOF. Let us take $Q \in TP(\rho_0, \rho_1)$ optimal for $\widetilde{S}_\alpha$ and suppose that it has finite energy. We will use the following fact, as a consequence of Theorem 3.3.7 of Chapter 3: for every t, the following equality is satisfied Q^t−a.e.

$$|(\sigma(t), t)|_Q = Q^t\left(\{\eta \,:\, \eta(t) = \sigma(t)\}\right). \tag{4.5.1}$$

This is true since we know $|(\sigma(t), t)|_Q = [\sigma]_{t,Q}$, which means that we can restrict our attention to those curves η who stayed together with σ for all the times between 0 and t. Moreover we can assume that σ is parametrized by arc length on $[0, T(\sigma)]$: this implies that, if σ is still moving, i.e. $T(\sigma) \geq t$, this is the case for all the curves η such that $\eta = \sigma$ on $[0, t]$ and proves that we can furtherly restrict our attention to those curves η with $T(\eta) \geq t$, i.e. switching from Q to Q^t, thus proving assertion (4.5.1).

Exchanging the order of integration, we can write

$$\widetilde{S}_\alpha(Q) = \int_{\mathrm{Lip}_1(I;\Omega)} \int_0^{T(\sigma)} |(\sigma(t), t)|_Q^{\alpha-1}\, dt\, dQ(\sigma) = \int_0^\infty \int_{\mathrm{Lip}_1(I;\Omega)} |(\sigma(t), t)|_Q^{\alpha-1} dQ^t(\sigma)\; dt.$$

Then define the equivalence classes $\Sigma_{t,x} = \{\sigma \in \mathrm{Lip}_1(I;\Omega) \,:\, \sigma(t) = x\}$ and notice that, by finiteness of the energy, for $\mathscr{L}^1$−a.e. t the measure Q^t must be concentrated on those classes $\Sigma_{t,x}$ such that $Q(\Sigma_{t,x}) > 0$. Since they have all positive mass, these classes are no more than a countable number and one can restrict the integral over them:

$$\int_{\mathrm{Lip}_1(I;\Omega)} |(\sigma(t), t)|_Q^{\alpha-1} dQ^t(\sigma) = \sum_i \int_{\Sigma_{t,x_i}} |(\sigma(t), t)|_Q^{\alpha-1} dQ^t(\sigma).$$

Yet, for all the curves $\sigma \in \Sigma_{t,x_i}$ we have $|(\sigma(t), t)|_Q = Q(\Sigma_{t,x_i}) = Q^t(\Sigma_{t,x_i})$, thanks to the fact that the class is non negligible and to the condition (4.5.1). Hence we may go on with

$$\int_{\mathrm{Lip}_1(I;\Omega)} |(\sigma(t), t)|_Q^{\alpha-1} dQ^t(\sigma) = \sum_i \int_{\Sigma_{t,x_i}} Q^t(\Sigma_{t,x_i})^{\alpha-1} dQ^t(\sigma) = \sum_i Q^t(\Sigma_{t,x_i})^\alpha.$$

Moreover if one constructs the measures μ and ν associated to Q thanks to Theorem 4.4.5, ν_t must be atomic and equal to $\sum_i Q^t(\Sigma_{t,x_i})\delta_{x_i}$, which gives in the end

$$\int_{\mathrm{Lip}_1(I;\Omega)} |(\sigma(t), t)|_Q^{\alpha-1} dQ^t(\sigma) = g_\alpha(\nu_t).$$

Hence, if we compare the energy of the special evolution pairing given by Theorem 4.4.5 with the energy of Q, we get

$$\overline{\mathfrak{L}}_\alpha(\nu, \mu) = \int_0^\infty g_\alpha(\nu_t)\, dt = \widetilde{S}_\alpha(Q),$$

which shows that

$$\min_{SEP(\rho_0,\rho_1)} \overline{\mathfrak{L}}_\alpha \leq \min_{TP(\rho_0,\rho_1)} \widetilde{S}_\alpha,$$

using the minimality of Q.

Conversely, let us take $(\nu, \mu) \in SEP(\rho_0, \rho_1)$ optimal and construct the traffic plan $Q \in TP(\rho_0, \rho_1)$ given by Theorem 4.4.4. As before, we revert the order of the integration in the definition of $\widetilde{S}_\alpha$. Let us set $\widehat{\nu}_t = (e_t)_\# Q^t$, and consider

$$\int_{\mathrm{Lip}_1(I;\Omega)} |(\sigma(t),t)|_Q^{\alpha-1}\, dQ^t(\sigma) = \int_\Omega |(x,t)|_Q^{\alpha-1}\, d\widehat{\nu}_t(x).$$

Moreover

$$\begin{aligned}|(x,t)|_Q = Q(\{\sigma \in \mathrm{Lip}_1(I;\Omega)\,:\,\sigma(t)=x\}) &\geq Q^t(\{\sigma \in \mathrm{Lip}_1(I;\Omega)\,:\,\sigma(t)=x\})\\ &= Q^t(\{e_t^{-1}(x)\}) = \widehat{\nu}_t(\{x\}),\end{aligned}$$

which in turn implies

$$\int_{\mathrm{Lip}_1(I;\Omega)} |(\sigma(t),t)|_Q^{\alpha-1}\, dQ^t(\sigma) \leq \int_\Omega \widehat{\nu}_t(\{x\})^{\alpha-1}\, d\widehat{\nu}_t.$$

It is only left to observe that

$$\int_\Omega \widehat{\nu}_t(\{x\})^{\alpha-1}\, d\widehat{\nu}_t = g_\alpha(\widehat{\nu}_t), \quad \text{for } \mathscr{L}^1\text{−a.e. } t \in I,$$

thanks to the fact that $\widehat{\nu}$ is atomic: indeed, ν_t is atomic for $\mathscr{L}^1$−a.e. t and we have $\widehat{\nu}_t \leq \nu_t$. Therefore collecting all these estimates, we end up with

$$\widetilde{S}_\alpha(Q) \leq \int_0^\infty \left(\int_\Omega \widehat{\nu}_t(\{x\})^{\alpha-1}\, d\widehat{\nu}_t(x)\right) dt = \int_0^\infty g_\alpha(\widehat{\nu}_t)\, dt \leq \overline{\mathfrak{L}}_\alpha(\nu,\mu),$$

thus concluding the proof. □

The case of ρ_1 atomic is interesting and deserves some words more: indeed, in this case thanks to Proposition 4.3.5 we obtain $SEP(\rho_0, \rho_1) = EP(\rho_0, \rho_1)$, so that

$$\min_{SEP(\rho_0,\rho_1)} \overline{\mathfrak{L}}_\alpha = \min_{EP(\rho_0,\rho_1)} \mathfrak{L}_\alpha.$$

It is then sufficient to note that by means of Theorem 4.5.1, the left-hand side is equal to the minimum of $\widetilde{S}_\alpha$ over the set $TP(\rho_0, \rho_1)$: summarizing, we have shown the following important fact.

Corollary 4.5.2. *Suppose that ρ_1 is a purely atomic probability measure and $\rho_0 = \delta_{x_0}$. Then*

$$\min_{EP(\rho_0,\rho_1)} \mathfrak{L}_\alpha = \min_{TP(\rho_0,\rho_1)} \widetilde{S}_\alpha.$$

This last connection with atomic measures suggests the following question, somehow in the spirit of Xia's relaxation procedure (see Chapter 3, Section 2): if one considers the functional defined as $\mathfrak{L}_\alpha$ on those evolution pairings (ν, μ) where ρ_1 is finitely atomic and $+\infty$ on the other evolution pairings, what is its relaxation $\mathfrak{L}_\alpha^*$? Is the relaxed functional related to $\mathfrak{L}_\alpha$ on $SEP(\rho_0, \rho_1)$?

CHAPTER 5

A Benamou-Brenier approach to branched transport

1. Introduction

In the previous two Chapters, we have seen various points of view on branched transportation, each of them having its advantages. Anyway, in all of them the branched transportation is studied avoiding the Benamou-Brenier approach consisting in the minimization of a suitable cost $\mathcal{G}(\mu, \phi)$ under the constraint of the continuity equation

$$\partial_t \mu + \operatorname{div}_x \phi = 0,$$

that we believe *is the most natural for this kind of problems*. The only approach to dynamical branched transportation using the continuity equation is, as far as we know, the one of [**23**]. Yet, to prove semicontinuity and hence existence, even in this model, the problem is reduced to the minimization of a functional of the form

$$\int \theta^\alpha \, d\mathscr{H}^1(x,t),$$

which is the energy of Xia, and the dynamical features are not completely exploited.

In the present chapter, based on a paper with Giuseppe Buttazzo and Filippo Santambrogio ([**B3**]), we present a more direct approach: for all pairs (μ, ϕ) verifying the continuity equation, with $\mu_0 = \rho_0$ and $\mu_1 = \rho_1$, we define a functional $\mathcal{G}(\mu, \phi)$ and we show this to be both lower semicontinuous and coercive with respect to a suitable convergence on (μ, ϕ): this will provide directly the existence of an optimal dynamical path.

The chapter is organized as follows: in Section 2 we give the precise setting and state the main results, leaving all the proofs (giving the existence of an optimal path μ_t) for Section 3. As always, we would like to know if our model is in accordance with the other descriptions for branched transportation phenomena: this is done in Section 4, where equivalence with the traffic plan model is shown. Finally, in the last section we deal with some inequalities involving Wasserstein distances and branched ones, i.e. distances over the space of probabilities given by the minima of some branched transportation problems. These inequalities have already been studied in [**70**] and [**43**], but some very precise issues concerning d_α and $w_{1/\alpha}$ are very close to the topics of this chapter and deserve being treated here: new and simpler proofs are provided.

2. Problem setting and main results

In this section we fix the notation and state the main results of the chapter. In the following Ω will denote a given subset of $\mathbb{R}^N$, where all the mass dynamics will take place; for the sake of simplicity we assume that Ω is convex and compact. In the following, we will also use the notation

$\mathcal{M}(\Omega;\mathbb{R}^N)$ to indicate the space of $\mathbb{R}^N-$valued Radon measures over Ω. The main objects to be considered will be pairs (μ,ϕ) with

$$\mu \in C\left([0,1];\mathscr{P}(\Omega)\right),\ \phi \in L^1\big([0,1];\mathcal{M}(\Omega;\mathbb{R}^N)\big) \tag{5.2.1}$$

satisfying the *continuity equation* formally written as (here ν stands for the outer normal versor to $\partial\Omega$)

$$\begin{cases} \partial_t\mu + \operatorname{div}_x\phi &= 0, \quad \text{in } [0,1]\times\Omega, \\ \langle\phi,\nu\rangle &= 0, \quad \text{on } [0,1]\times\partial\Omega, \end{cases} \tag{5.2.2}$$

whose precise meaning, as always, is given in the sense of distributions, that is

$$\int_0^1 \Big[\int_\Omega \partial_t\varphi(t,x)\,d\mu_t(x) + \int_\Omega \nabla_x\varphi(x,t)\cdot d\phi_t(x)\Big]\,dt = 0, \tag{5.2.3}$$

for every smooth function φ with $\varphi(0,x)=\varphi(1,x)=0$.

Definition 5.2.1. We denote with $\mathfrak{D}$ the set of these pairs (μ,ϕ) satisfying (5.2.1) and (5.2.3). Moreover given $\mu_0,\mu_1\in\mathscr{P}(\Omega)$, we define the set $\mathfrak{D}(\rho_0,\rho_1)$ of admissible configurations connecting ρ_0 to ρ_1 in the following way

$$\mathfrak{D}(\rho_0,\rho_1) = \{(\mu,\phi)\in\mathfrak{D}\ :\ \mu_0=\rho_0,\ \mu_1=\rho_1\}.$$

The velocity vector v can be defined as the Radon-Nikodym derivative of the vector measure ϕ with respect to μ:

$$v = \frac{d\phi}{d\mu}.$$

Among all pairs (μ,ϕ) satisfying the continuity equation above, we consider a cost function of the form

$$\mathcal{G}_\alpha(\mu,\phi) = \int_0^1 G_\alpha(\mu_t,\phi_t)\,dt,\ (\mu,\phi)\in\mathfrak{D}, \tag{5.2.4}$$

where G_α is defined through

$$G_\alpha(\mu,\phi) := \begin{cases} g_\alpha(|v|^{1/\alpha}\mu), & \text{if } \phi = v\cdot\mu, \\ +\infty, & \text{if } \phi \not\ll \mu, \end{cases}$$

and g_α $(0<\alpha<1)$ is the same lower semicontinuous functional as in Chapters 3 and 4: $g_\alpha(\lambda)=+\infty$ if λ is not purely atomic, while ($\#$ stands for the counting measure)

$$g_\alpha(\lambda) = \int_\Omega |\lambda(\{x\})|^\alpha\,d\#(x) = \sum_{i\in\mathbb{N}}\lambda_i^\alpha, \ \text{ if } \lambda = \sum_{i\in\mathbb{N}}\lambda_i\delta_{x_i}.$$

In this way our functional $\mathcal{G}_\alpha$ becomes

$$\mathcal{G}_\alpha(\mu,\phi) = \int_0^1 \left[\int_\Omega |v_t(x)|\,\mu_t(\{x\})^\alpha\,d\#(x)\right]dt = \int_0^1\left[\sum_{i\in\mathbb{N}}|v_{t,i}|\,\mu_{t,i}^\alpha\right]dt,$$

and the dynamical model for branched transportation we consider is

$$\mathfrak{B}_\alpha(\rho_0,\rho_1) := \min_{(\mu,\phi)\in\mathfrak{D}(\rho_0,\rho_1)}\mathcal{G}_\alpha(\mu,\phi). \tag{5.2.5}$$

Our main goal is to show that the minimization problem (5.2.5) above admits a solution. This will be obtained through the Direct Methods of the calculus of variations, consisting in proving lower semicontinuity and coercivity of the problem under consideration, with respect to a suitable convergence.

REMARK 5.2.2. We point out that the $*-$weak convergence of the pairs (μ,ϕ) as measures on $[0,1]\times\Omega$ does not directly imply the lower semicontinuity in (5.2.5), since the functional is not jointly convex. On the other hand, if (μ^n,ϕ^n) satisfy the continuity equation (5.2.2) and we assume

$$(\mu^n_t,\phi^n_t)\rightharpoonup(\mu_t,\phi_t),\quad \text{for } \mathscr{L}^1\text{−a.e. } t\in[0,1],$$

then a simple application of Fatou's Lemma would lead to the desired semicontinuity property of $\mathcal{G}_\alpha$, as far as the integrand G_α is a lower semicontinuous functional on measures. To see this, it is enough to observe that considering (as in Chapter 1)

$$f_{1/\alpha}(x,y)=\begin{cases} |y|^{1/\alpha}\,x^{(\alpha-1)/\alpha}, & \text{if } x>0,\ y\in\mathbb{R}^N,\\ 0, & \text{if } x=0,\ y=0,\\ +\infty, & \text{otherwise}\end{cases}$$

this is a convex and $1-$homogeneous, and hence also subadditive, function on $\mathbb{R}\times\mathbb{R}^N$. Then G_α can be equivalently rewritten as

$$G_\alpha(\mu,\phi)=\begin{cases} \displaystyle\int_\Omega f_{1/\alpha}(\mu(\{x\}),\phi(\{x\}))^\alpha\,d\#(x), & \text{if } \phi \text{ is atomic},\\ +\infty, & \text{otherwise}\end{cases}$$

which falls into the class of local and lower semicontinuous functionals defined on measures studied by Bouchitté and Buttazzo (see [**21**, Theorem 3.3]), thanks to the fact that

$$(x,y)\mapsto f_{1/\alpha}(x,y)^\alpha,$$

is lower semicontinuous, subadditive, $f(0,0)=0$ and moreover

$$\lim_{t\to0^+}\frac{f_{1/\alpha}(tx,ty)^\alpha}{t}=+\infty,\quad \text{for } x\neq0.$$

In order to prove in the easiest possible way a semicontinuity result, we will introduce a convergence which is stronger than the weak convergence of measures on $[0,1]\times\Omega$, but weaker than weak convergence for every fixed time t. This convergence will be compatible with the compactness we can infer from our variational problem.

DEFINITION 5.2.3. We say that a sequence $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}$ τ-converges to (μ,ϕ) if $(\mu^n,\phi^n)\rightharpoonup(\mu,\phi)$ in the sense of measures and

$$\sup_{n\in\mathbb{N},\,t\in[0,1]} G_\alpha(\mu^n_t,\phi^n_t)<+\infty.$$

PROPOSITION 5.2.4. *Let $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}\subset\mathfrak{D}$ be a sequence such that $\mathcal{G}_\alpha(\mu^n,\phi^n)\le C$, then up to a time reparametrization, $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}$ is τ-compact.*

PROPOSITION 5.2.5. *Let $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}\subset\mathfrak{D}$ be a sequence τ-converging to (μ,ϕ). Then*

$$\mathcal{G}_\alpha(\mu,\phi)\le\liminf_{n\to\infty}\mathcal{G}_\alpha(\mu^n,\phi^n).$$

As a consequence we obtain the following existence result, which represents the main result of this chapter.

THEOREM 5.2.6. *For every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, the minimization problem* (5.2.5) *admits a solution.*

REMARK 5.2.7. We remark that, for some choices of the data ρ_0, ρ_1 and of the exponent α, the statement of Theorem 5.2.6 could be empty, because the functional $\mathcal{G}_\alpha$ could be constantly $+\infty$ on every admissible path (μ, ϕ) joining ρ_0 to ρ_1. This issue will be solved in Section 4, where the equivalence with Lagrangian models (see Chapter 3) will be proven. Since for these models finiteness of the minima has been widely investigated, we can infer for instance that if $\alpha > 1 - 1/N$ then every pair ρ_0 and ρ_1 can be joined by a path of finite energy. On the other hand, if $\alpha \le 1 - 1/N$, $\rho_0 = \delta_{x_0}$ and ρ_1 is absolutely continuous w.r.t. $\mathscr{L}^N$, then we have $\mathcal{G}_\alpha(\mu, \phi) = +\infty$ for every $(\mu, \phi) \in \mathfrak{D}(\rho_0, \rho_1)$.

3. Proofs

A preliminary inequality to all the proofs is the following: if $\phi \ll \mu$, then $\phi_t = v_t \cdot \mu_t$ and

$$
\begin{aligned}
G_\alpha(\mu_t, \phi_t) &= \sum_i \mu_t(\{x_i\})^\alpha |v_t(x_i)| = \sum_i \Big(\mu_t(\{x_i\}) |v_t(x_i)|^{1/\alpha}\Big)^\alpha \\
&\ge \left(\sum_i \mu_t(\{x_i\}) |v_t(x_i)|^{1/\alpha}\right)^\alpha = \|v_t\|_{L^{1/\alpha}(\mu_t)},
\end{aligned} \tag{5.3.1}
$$

where as always we have used the sub-additivity of the function $x \mapsto x^\alpha$. In particular it also follows

$$
G_\alpha(\mu_t, \phi_t) \ge \|v_t\|_{L^1(\mu_t)} = |\phi_t|(\Omega). \tag{5.3.2}
$$

The next simple result will be quite useful.

LEMMA 5.3.1. *Let $\{(\mu^n, \phi^n)\}_{n\in\mathbb{N}} \subset \mathfrak{D}$ such that $(\mu^n, \phi^n) \xrightarrow{\tau} (\mu, \phi)$. Then $(\mu, \phi) \in \mathfrak{D}$ and $\phi \ll \mu$.*

PROOF. Observe that with the set $\mathfrak{D}$ we a priori restricted our attention to those pairs $(\overline{\mu}, \overline{\phi})$, where $\overline{\mu} \in C\big([0,1]; \mathscr{P}(\Omega)\big)$ and $\overline{\phi} \in L^1\big([0,1]; \mathcal{M}(\Omega; \mathbb{R}^N)\big)$, so that first of all we need to prove that μ is continuous and that ϕ admits a disintegration of the form

$$
\phi = \int \phi_t \, dt,
$$

with respect to the time variable. Yet, using the fact that

$$
\sup_{n\in\mathbb{N},\, t\in[0,1]} G_\alpha(\mu_t^n, \phi_t^n) < +\infty,
$$

the inequality (5.3.1) applied to the pairs (μ^n, ϕ^n) yields a uniform bound on the $L^{1/\alpha}$ norm of the velocities, which implies that the curves $\tilde{\mu}^n$ are uniformly Lipschitz continuous according to the distance $w_{1/\alpha}$ (thanks to Theorem 1.5.4), and this property is inherited by the limit measure μ.

For the decomposition of ϕ, just use again the inequality (5.3.1), thus obtaining a uniform bound on $\|v_t^n\|_{L^{1/\alpha}(\mu_t^n)}$, which a fortiori gives a uniform bound on the Benamou-Brenier functional

$$
\mathcal{F}_{1/\alpha}(\mu^n, \phi^n) = \int_0^1 \|v_t^n\|_{L^{1/\alpha}(\mu_t^n)}^{1/\alpha} \, dt \le C.
$$

This functional being lower semicontinuous w.r.t. to the weak convergence, we can deduce the same bound at the limit: this in particular implies that ϕ is absolutely continuous w.r.t. μ, with an $L^{1/\alpha}$ density (we have already used this argument in the proof of Proposition 1.5.9 of Chapter 1). Since μ is a measure on $[0,1]\times\Omega$ which is of the form $\int \mu_t\,dt$, the same disintegration will be true for ϕ. Finally, it is immediate to check that (μ,ϕ) still solves the continuity equation (5.2.3). □

3.1. Proof of Proposition 5.2.4. Due to the fact the the functional $\mathcal{G}_\alpha$ is $1-$homogeneous in the velocity, it is clear that reparametrizations in time do not change the values of $\mathcal{G}_\alpha$. By reparametrization, we mean replacing a pair (μ,ϕ) with a new pair $(\tilde\mu,\tilde\phi)$ of the form $\tilde\mu_t=\mu_{\mathfrak{s}(t)}$, $\tilde\phi_t=\mathfrak{s}'(t)\phi_{\mathfrak{s}(t)}$ (which equivalently means that $\tilde\phi$ is the image measure of ϕ through the inverse of the map $(t,x)\mapsto(\mathfrak{s}(t),x)$), where $\mathfrak{s}:[0,1]\to[0,1]$ is absolutely continuous and increasing. Observe that we still have $(\tilde\mu,\tilde\phi)\subset\mathfrak{D}$, i.e. this new pair still solves the continuity equation (see [**6**, Lemma 8.1.3]).

Thanks to this invariance, if $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}\subset\mathfrak{D}$ is such that $\mathcal{G}_\alpha(\mu^n,\phi^n)\le C$, then one can define a new sequence of pairs $\{(\tilde\mu^n,\tilde\phi^n)\}_{n\in\mathbb{N}}\subset\mathfrak{D}$, with

$$G_\alpha(\tilde\mu^n_t,\tilde\phi^n_t)\equiv\mathcal{G}_\alpha(\tilde\mu^n,\tilde\phi^n)=\mathcal{G}_\alpha(\mu^n,\phi^n)\le C.$$

After that, we only need to prove compactness for the weak convergence of measures on $[0,1]\times\Omega$, a fact which only requires bounds on the mass of $\tilde\mu^n$ and $\tilde\phi^n$. The bound on $\tilde\mu^n$ is straightforward, since for every t the measure $\tilde\mu^n_t$ is a probability, while for $\tilde\phi^n$, which is absolutely continuous w.r.t. $\tilde\mu^n$, it is enough to use (5.3.2) in order to bound the mass of ϕ by C.

This allows to extract a subsequence $\{(\tilde\mu^{n_k},\tilde\phi^{n_k})\}_{k\in\mathbb{N}}$ which converges weakly to a pair (μ,ϕ). This and the uniform bound on $G_\alpha(\tilde\mu^{n_k}_t,\tilde\phi^{n_k}_t)$ conclude the proof.

3.2. Proof of Proposition 5.2.5. We consider a sequence $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}\subset\mathfrak{D}$ which is $\tau-$converging to (μ,ϕ): observe in particular that, thanks to the uniform bound

$$\sup_{n\in\mathbb{N},\,t\in[0,1]} G_\alpha(\mu^n_t,\phi^n_t)\le C, \tag{5.3.3}$$

we have $\mathcal{G}_\alpha(\mu^n,\phi^n)\le C$ and $\phi^n=v^n\cdot\mu^n$. First of all we define a sequence of measures $\{\mathfrak{m}^n\}_{n\in\mathbb{N}}$ on $[0,1]\times\Omega$ through

$$\mathfrak{m}^n=\int\left(\sum_i\mu^n_t(\{x_{i,t}\})^\alpha|v^n_t(x_{i,t})|\delta_{x_{i,t}}\right)dt,$$

where the points $x_{i,t}$ are the atoms of ϕ^n_t (i.e. the atoms of μ^n_t where the velocity v^n_t does not vanish). We notice that $\mathcal{G}_\alpha(\mu^n,\phi^n)=\mathfrak{m}^n([0,1]\times\Omega)$, then the bound on $\mathcal{G}_\alpha(\mu^n,\phi^n)$ implies the convergence $\mathfrak{m}^n\rightharpoonup\mathfrak{m}$, up to the extraction of a subsequence (not relabeled). It is clear that, on this subsequence, we have

$$\lim_{n\to\infty}\mathcal{G}_\alpha(\mu^n,\phi^n)=\lim_{n\to\infty}\mathfrak{m}^n([0,1]\times\Omega)=\mathfrak{m}([0,1]\times\Omega),$$

then in order to prove the desired semicontinuity property, it is enough to get some proper lower bounds on $\mathfrak{m}$. Observe that, since we have

$$\int_{[0,1]\times\Omega}\xi(t)\,d\mathfrak{m}^n(t,x)=\int_0^1\xi(t)\,G_\alpha(\mu^n_t,\phi^n_t)\,dt,\quad\text{for every }\xi\in C([0,1]),$$

we obtain that the marginal of $\mathfrak{m}^n$ on the time variable is the measure $G_\alpha(\mu_t^n, \phi_t^n)\mathscr{L}^1 \llcorner [0,1]$, which has an L^∞ density uniformly bounded by the constant C on $[0,1]$, again thanks to (5.3.3). This bound is uniform and will be preserved by the limit measure $\mathfrak{m}$: in particular, this implies that we can write $\mathfrak{m} = \int \mathfrak{m}_t \, dt$. Then, in order to prove our semicontinuity result, it will be enough to show the following

$$\mathfrak{m}_t(\Omega) \geq G_\alpha(\mu_t, \phi_t), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{5.3.4}$$

Before going on, observe that thanks to Lemma 5.3.1, we have that $(\mu, \phi) \in \mathfrak{D}$ (so that in particular $\phi = \int \phi_t \, dt$) and $\phi \ll \mu$.

Let us fix a closed set E, as well as a time interval $[a,b]$, and take the function

$$\chi_M(x) := (1 - M \operatorname{dist}(x, E))_+, \; x \in \Omega,$$

where $(\,\cdot\,)_+$ stands for the positive part: observe that χ_M is positive, takes the value 1 on E, is $M-$Lipschitz and identically vanishes outside a $1/M-$neighborhood of E. We then take $\varphi(t,x) = \chi_M(x)^\alpha 1_{[a,b]}(t)$, which is upper semicontinuous on $[0,1] \times \Omega$, then we have

$$\begin{aligned}\int \varphi(t,x) \, d\mathfrak{m}(t,x) &\geq \limsup_{n\to\infty} \int \varphi(t,x)\, d\mathfrak{m}^n(t,x) \\ &= \limsup_{n\to\infty} \int_a^b \left(\sum_i \mu_t^n(\{x_i\})^\alpha |v_t^n(x_i)| \chi_M(x_i)^\alpha \right) dt,\end{aligned}$$

where the points x_i are, as before, the atoms of ϕ^n (and we omitted the dependence on n and t). We then write

$$\mu_t^n(\{x_i\})^\alpha \chi_M(x_i)^\alpha = \left(\mu_t^n(\{x_i\})\chi_M(x_i)\right)^{\alpha-1} \left(\mu_t^n(\{x_i\})\chi_M(x_i)\right),$$

(where $\mu_t^n \, \chi_M > 0$) and we notice that $\mu_t^n(\{x_i\})\chi_M(x_i) \leq \int \chi_M \, d\mu_t^n$. Then we can estimate the right-hand side in the previous inequality as

$$\begin{aligned}\int_a^b \left(\sum_i \mu_t^n(\{x_i\})^\alpha |v_t^n(x_i)| \chi_M(x_i)^\alpha \right) dt &\geq \int_a^b \left(\int \chi_M d\mu_t^n \right)^{\alpha-1} \left(\sum_i \mu_t^n(\{x_i\}) |v_t^n(x_i)| \chi_M(x_i) \right) dt \\ &= \int_a^b \left(\int \chi_M \, d\mu_t^n \right)^{\alpha-1} \left(\int \chi_M \, d|\phi_t^n| \right) dt.\end{aligned}$$

We go on by estimating from above $\int \chi_M \, d\mu_t^n$: we have

$$\int_\Omega \chi_M(x)\, d\mu_t^n(x) \leq \int_\Omega \chi_M(x)\, d\mu_s^n(x) + M \, w_1(\mu_t^n, \mu_s^n),$$

which is a consequence of the definition of w_1 by duality with $1-$Lipschitz functions (see Chapter 1, Proposition 1.1.4). To estimate the w_1 distance we use $w_1 \leq w_{1/\alpha}$ (see Chapter 1, equation (1.3.1)) and the following fact

$$w_{1/\alpha}(\mu_t^n, \mu_s^n) \leq \int_s^t |(\mu_z^n)'|_{w_{1/\alpha}} \, dz \leq \int_s^t \|v_z^n\|_{L^{1/\alpha}(\mu_z)} \, dz,$$

then applying inequality (5.3.1) we have in the end

$$\int_\Omega \chi_M(x)\, d\mu_t^n(x) \le \int_\Omega \chi_M(x)\, d\mu_a^n(x) + CM(b-a), \quad \text{for every } t \in [a,b].$$

In this way we have

$$\int_a^b \left(\int \chi_M\, d\mu_t^n\right)^{\alpha-1}\left(\int \chi_M\, d|\phi_t^n|\, dt\right)\, dt \ge \left(\int \chi_M\, d\mu_a^n + CM(b-a)\right)^{\alpha-1}\int_a^b \left(\int \chi_M\, d|\phi_t^n|\right)\, dt$$
$$= \left(\int \chi_M\, d\mu_a^n + CM(b-a)\right)^{\alpha-1}\int \varphi^{1/\alpha} d|\phi^n|.$$

Hence, we may go on with

$$\int \varphi\, d\mathfrak{m} \ge \limsup_{n\to\infty}\left[\left(\int \chi_M\, d\mu_a^n + CM(b-a)\right)^{\alpha-1}\int \varphi^{1/\alpha} d|\phi^n|\right]$$
$$\ge \left(\int \chi_M d\mu_a + CM(b-a)\right)^{\alpha-1}\int \varphi^{1/\alpha} d|\phi|.$$

In the last inequality the second factor has been dealt with in the following way: suppose $|\phi^n| \rightharpoonup \eta$, then we have $\eta \ge |\phi|$; moreover $\varphi \ge \tilde\varphi$ where $\tilde\varphi(t,x) := \chi_M(x)1_{(a,b)}(t)$, and this last function is l.s.c. and positive, so that

$$\liminf_{n\to\infty}\int \varphi^{1/\alpha} d|\phi^n| \ge \liminf_{n\to\infty}\int \tilde\varphi^{1/\alpha} d|\phi^n| \ge \int \tilde\varphi^{1/\alpha} d\eta \ge \int \tilde\varphi^{1/\alpha} d|\phi| = \int \varphi^{1/\alpha} d|\phi|,$$

where in the last equality we have used that the boundaries $t = a$ and $t = b$ are negligible for $|\phi|$, thanks to the fact that $\phi \in L^1([0,1];\mathcal{M}(\Omega;\mathbb{R}^N))$.

After that, we can divide by $(b-a)$ (keeping for a while M fixed) and pass to the limit as $b \to a$. We have, for $\mathscr{L}^1-$a.e. $a \in [0,1]$,

$$\int \chi_M(x)^\alpha\, d\mathfrak{m}_a(x) \ge \left(\int \chi_M(x)\, d\mu_a(x)\right)^{\alpha-1}\int \chi_M(x)\, d|\phi_a|(x).$$

We let now $M \to \infty$, so that χ_M monotonically converges to the characteristic function of the set E, and we have, by dominated convergence w.r.t. $\mathfrak{m}_a, \mu_a$ and $|\phi_a|$,

$$\mathfrak{m}_a(E) \ge \mu_a(E)^{\alpha-1}|\phi_a|(E). \tag{5.3.5}$$

In the last term the convention $0\cdot\infty = 0$ is used (if $|\phi_a|(E) = 0$). This inequality is proven for closed sets, but by regularity of the measures it is not difficult to prove it for arbitrary sets. Actually, if $S \subset \Omega$ is an arbitrary Borel set, we can write

$$\mathfrak{m}_a(S) \ge \mathfrak{m}_a(E) \ge \mu_a(E)^{\alpha-1}|\phi_a|(E) \ge \mu_a(S)^{\alpha-1}|\phi_a|(E),$$

for every $E \subset S$ closed and take a sequence of closed sets E_k such that $|\phi_a|(E_k) \to |\phi_a|(S)$, since $|\phi_a|$ is, for $\mathscr{L}^1-$a.e. $a \in [0,1]$, a finite (and hence regular) measure on the compact set Ω. We want now to prove that:

- $\phi_a = v_a \cdot \mu_a$ is atomic for $\mathscr{L}^1-$a.e. $a \in [0,1]$ (i.e. μ_a is atomic on $\{v_a \neq 0\}$)
- $\mathfrak{m}_a(\Omega) \ge G_\alpha(\mu_a, \phi_a)$, for $\mathscr{L}^1-$a.e. $a \in [0,1]$.

This would conclude the proof.

For the first statement, take the inequality $\mathfrak{m}_a(S) \geq \mu_a(S)^{\alpha-1}|\phi_a|(S)$ which is valid for any set, and apply it to sets which are contained in the Borel set $V_\varepsilon := \{x \in \Omega : |v_a(x)| > \varepsilon\}$. For those sets, we have easily $\mathfrak{m}_a(S) \geq \varepsilon\mu_a(S)^\alpha$. This means that the measure $\lambda := \varepsilon^{1/\alpha}\mu_a \llcorner V_\varepsilon$ satisfies the inequality $\lambda(S)^\alpha \leq \mathfrak{m}_a(S)$ for every Borel set $S \subset \Omega$, where $\mathfrak{m}_a$ is a finite measure: this implies that λ is atomic (see Lemma 5.3.2 below). If the same is performed for every $\varepsilon = 1/k$, this proves that μ_a is purely atomic on the set $\{x : |v_a(x)| \neq 0\}$, that is $\phi_a = v_a \cdot \mu_a$ is purely atomic.

Once we know that ϕ_a is atomic we infer that $G_\alpha(\mu_a, \phi_a) = \sum_i \mu_a(\{x_i\})^{\alpha-1}|\phi_a|(\{x_i\})$ and we only need to consider $E = \{x_i\}$ in (5.3.5) and add up:

$$\mathfrak{m}_a(\Omega) \geq \sum_i \mathfrak{m}_a(\{x_i\}) \geq G_\alpha(\mu_a, \phi_a),$$

which finally gives (5.3.4). As we said, this concludes the proof.

LEMMA 5.3.2. *Take two finite positive measures λ and μ on a domain Ω, and $\alpha \in (0,1)$. Suppose that the inequality $\lambda(S)^\alpha \leq \mu(S)$ is satisfied for every Borel set $S \subset \Omega$. Then λ is purely atomic.*

PROOF. Consider a regular grid on Ω of size $1/k$, for $k \in \mathbb{N}$, and build a measure λ_k by putting, in every cell of the grid, all the mass of λ in a single point of the cell. This measure λ_k is atomic and we have

$$g_\alpha(\lambda_k) = \sum_i \lambda(S_i)^\alpha \leq \sum_i \mu(S_i) = \mu(\Omega) < +\infty,$$

where the S_i are the cells of the grid. If we let k goes to ∞, the step of the grid goes to zero and we obviously have $\lambda_k \rightharpoonup \lambda$. On the other hand, the functional g_α is lower semicontinuous and this implies $g_\alpha(\lambda) \leq \liminf_{k\to\infty} g_\alpha(\lambda_k) \leq \mu(\Omega) < +\infty$. In particular, λ is atomic, thus proving the assertion. □

3.3. Proof of Theorem 5.2.6. In order to prove existence, one only needs to take a minimizing sequence in $\mathfrak{D}(\rho_0, \rho_1)$ and apply Proposition 5.2.4 to get a new minimizing sequence which is $\tau-$converging: this new sequence is obtained through reparametrization (which does not change the value of $\mathcal{G}_\alpha$) and by extracting a subsequence. Since the constraints in the problem are linear, i.e. $\mu_i = \rho_i$ for $i = 0, 1$ and the continuity equation, the limit (μ, ϕ) still belongs to $\mathfrak{D}(\rho_0, \rho_1)$. Finally, the semicontinuity proven in Proposition 5.2.5 allows to conclude.

4. Equivalence with Lagrangian models

In this section we prove the equivalence of problem (5.2.5) with the other previous formulations of branched transport models. In particular, as a reference model we will take the one by Bernot, Caselles and Morel, in which the energy is defined as

$$E_\alpha(Q) = \int_{C([0,1];\Omega)} \int_0^1 |\sigma(t)|_Q^{\alpha-1} |\sigma'(t)| \, dt \, dQ(\sigma),$$

where Q is a probability measure over $C([0,1];\Omega)$ which gives full mass to the set $\mathrm{Lip}([0,1];\Omega)$ (a *traffic plan*) and for every $x \in \Omega$, the quantity $|x|_Q$ is the *multiplicity* of x with respect to Q, defined by

$$|x|_Q = Q\left(\{\sigma \,:\, x \in \sigma([0,1])\}\right).$$

Given $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, the corresponding minimum problem is then given by

$$\min_{Q \in TP(\rho_0,\rho_1)} E_\alpha(Q),$$

where $TP(\rho_0,\rho_1)$ is the set of traffic plans with prescribed time marginals at $t = 0, 1$

$$TP(\rho_0,\rho_1) = \{Q \,:\, (e_i)_\sharp Q = \mu_i,\ i = 0, 1\},$$

and $e_t : C([0,1];\Omega) \to \Omega$ is as always the evaluation at time t map.

REMARK 5.4.1. Observe that we are making a small abuse in the definition of the energy E_α, which apparently differs from that introduced in Chapter 3: indeed, we are considering the space of Lipschitz curves on the compact time interval $[0,1]$, instead of $1-$Lipschitz curves parametrized over $[0,\infty)$. Clearly this further definition is equivalent to the usual one (as already suggested by the informal presentation in Chapter 3), thanks to the invariance, w.r.t. to reparametrizations in time, of E_α and to the fact that the finiteness of the energy gives the finiteness of the average-length functional. Indeed, if Q satisfies

$$\int_{C([0,1];\Omega)} \int_0^1 |\sigma(t)|_Q^{\alpha-1}\, |\sigma'(t)|\, dt\, dQ(\sigma) \le C,$$

thanks to the fact that $|x|_Q^{\alpha-1} \ge 1$, we obtain that

$$\int_{C([0,1];\Omega)} \ell(\sigma)\, dQ(\sigma) \le C,$$

so that, using Markov inequality and the reparametrization invariance of the functional, it is easily seen that the minimization of this modified energy in $TP(\rho_0,\rho_1)$ admits a solution, provided a Q having finite E_α exists. Moreover to every Q concentrated on $\mathrm{Lip}([0,1];\Omega)$ it is possible to associate a $\widetilde{Q} \in \mathscr{P}(\mathrm{Lip}_1([0,\infty);\Omega))$ (just by arc-length reparametrization of the curves) and the respective energies coincide; on the other hand, as we saw above every $\widetilde{Q} \in \mathscr{P}(\mathrm{Lip}_1([0,\infty);\Omega))$ with finite energy is concentrated on curves with finite length and thus we can revert the previous reasoning, obtaining a Q concentrated in Lipschitz curves on $[0,1]$, without altering the value of the energy.

In this way, we see that the minimization problem considered in this section is equivalent to that for E_α encountered in Section 3 of Chapter 3 and the same results and remarks there stated, apply to this case as well.

As in Chapter 4, we also need to consider the slight modification of E_α above, given by the synchronized energy (the notation is the same as in Chapter 3)

$$S_\alpha(Q) = \int_{C([0,1];\Omega)} \int_0^1 |(\sigma(t),t)|_Q^{\alpha-1}\, |\sigma'(t)|\, dt\, dQ(\sigma),$$

where, we recall that the *synchronized multiplicity* $|(x,t)|_Q$ is given by

$$|(x,t)|_Q = Q(\{\sigma \,:\, \sigma(t) = x\}).$$

As a straightforward consequence of the definition of these two multiplicities, we have already observed that

$$|\sigma(t)|_Q \geq |(\sigma(t), t)|_Q, \tag{5.4.1}$$

so that $E_\alpha(Q) \leq S_\alpha(Q)$. Concerning the comparison between the minimization of E_α and S_α, we recall that they coincide when ρ_0 is finitely atomic (see Chapter 3, Theorem 3.3.3). We are now in a position to state and prove a result giving the equivalence between our model and the one relative to the energy E_α: we will use the notation

$$d_\alpha(\rho_0, \rho_1) = \min_{TP(\rho_0,\rho_1)} E_\alpha,$$

thanks to the equivalence between Xia's model and Lagrangian ones (Chapter 3, Theorem 3.3.3).

THEOREM 5.4.2. *For every $\alpha \in (0,1)$ and $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ we get*

$$\mathfrak{B}_\alpha(\rho_0, \rho_1) = d_\alpha(\rho_0, \rho_1). \tag{5.4.2}$$

PROOF. We first want to prove $\mathfrak{B}_\alpha(\rho_0, \rho_1) \geq d_\alpha(\rho_0, \rho_1)$. If $\mathfrak{B}_\alpha(\rho_0, \rho_1) = +\infty$ there is nothing to prove. Otherwise, take (μ, ϕ) optimal, which implies, by the way, that $\phi = v \cdot \mu$ and that ϕ is atomic. Thanks to the superposition principle (see Chapter 1, Theorem 1.5.1) we can construct a probability measure $Q \in \mathcal{C}$ such that $\mu_t = (e_t)_\sharp Q$ and Q is concentrated on absolutely continuous integral curves of v. Using this information, together with the fact that $E_\alpha \leq S_\alpha$ and exchanging the order of integration, we get

$$\begin{aligned} E_\alpha(Q) \leq S_\alpha(Q) &= \int_{C([0,1];\Omega)} \int_0^1 |(\sigma(t),t)|_Q^{\alpha-1} \, |\sigma'(t)| \, dt \, dQ(\sigma) \\ &= \int_0^1 \int_{C([0,1];\Omega)} |(\sigma(t),t)|_Q^{\alpha-1} \, |\sigma'(t)| \, dQ(\sigma) \, dt \\ &= \int_0^1 \int_{C([0,1];\Omega)} |(\sigma(t),t)|_Q^{\alpha-1} \, |v_t(\sigma(t))| \, dQ(\sigma) \, dt \\ &= \int_0^1 \int_\Omega |(x,t)|_Q^{\alpha-1} \, |v_t(x)| \, d\mu_t(x) \, dt, \end{aligned}$$

then we observe that, by virtue of the fact that $\mu_t = (e_t)_\sharp Q$, there holds (we have already used this observation in the proof of Theorem 4.5.1 of Chapter 4)

$$|(x,t)|_Q = Q(\{\sigma \in \mathcal{C} \, : \, \sigma(t) = x\}) = \mu_t(\{x\}),$$

so that we can rephrase the last integral as follows

$$\int_0^1 \int_\Omega \mu_t(\{x\})^{\alpha-1} \, |v_t(x)| \, d\mu_t(x) \, dt = \int_0^1 \int_\Omega \mu_t(\{x\})^{\alpha-1} \, d|\phi_t|(x) \, dt = \int_0^1 \sum_{i\in\mathbb{N}} |v_{t,i}| \mu_{t,i}^\alpha \, dt,$$

which then gives

$$d_\alpha(\rho_0, \rho_1) \leq \mathcal{G}_\alpha(\mu, \phi) = \mathfrak{B}_\alpha(\rho_0, \rho_1).$$

In order to prove the reverse inequality, we proceed as follows: first of all we prove that

$$\mathfrak{B}_\alpha(\rho_0, \rho_1) \leq \min_{Q\in TP(\rho_0,\rho_1)} S_\alpha(Q). \tag{5.4.3}$$

Let us take $Q \in TP(\rho_0, \rho_1)$ optimal for S_α, then we know that there exists a pair (μ, ϕ) which is a solution of the continuity equation, with $\mu_t = (e_t)_\sharp Q$ and $\phi_t = v_t \cdot \mu_t$. Recalling Remark 1.5.8 of Chapter 1, the velocity v may be chosen as

$$v_t(x) = \int \sigma'(t)\, dQ_x^t(\sigma),$$

where Q_x^t is the disintegration of Q with respect to the evaluation function e_t. This means that for μ_t−a.e. $x \in \Omega$, each Q_x^t is a probability measure concentrated on the set $\{\sigma : \sigma(t) = x\}$ and $Q = \int Q_x^t \, d\mu_t(x)$. Therefore, arguing as before

$$\begin{aligned} S_\alpha(Q) &= \int_0^1 \int_{\mathcal{C}} |(\sigma(t), t)|_Q^{\alpha-1} \, |\sigma'(t)| \, dQ(\sigma)\, dt \\ &= \int_0^1 \int_\Omega |(x,t)|_Q^{\alpha-1} \left(\int |\sigma'(t)| \, dQ_x^t(\sigma) \right) d\mu_t(x)\, dt \\ &\geq \int_0^1 \int_\Omega |(x,t)|_Q^{\alpha-1} \, |v_t(x)| \, d\mu_t(x)\, dt \\ &= \int_0^1 \int_\Omega \mu_t(\{x\})^{\alpha-1} \, |v_t(x)| \, d\mu_t(x)\, dt, \end{aligned}$$

that gives the desired inequality (5.4.3) since, even if we do not know that ϕ_t or μ_t are atomic we can restrict the last integral to the set of atoms of μ.

Summarizing, up to now we have shown

$$d_\alpha(\rho_0, \rho_1) \leq \mathfrak{B}_\alpha(\rho_0, \rho_1) \leq \min_{Q \in TP(\rho_0, \rho_1)} S_\alpha(Q),$$

and equality holds whenever ρ_0 is a finite sum of Dirac masses. In order to conclude, it is enough to notice that thanks to Proposition 3.3.5 of Chapter 3, we may take two sequences $\{\rho_0^n\}_{n\in\mathbb{N}}$ and $\{\rho_1^n\}_{n\in\mathbb{N}}$ of finitely atomic probability measures such that $\rho_0^n \rightharpoonup \rho_0$, $\rho_1^n \rightharpoonup \rho_1$ and

$$\lim_{n\to\infty} d_\alpha(\rho_0^n, \rho_1^n) = d_\alpha(\rho_0, \rho_1),$$

thus getting

$$\begin{aligned} d_\alpha(\rho_0, \rho_1) \leq \mathfrak{B}_\alpha(\rho_0, \rho_1) \leq \liminf_{n\to\infty} \mathfrak{B}_\alpha(\rho_0^n, \rho_1^n) &\leq \lim_{n\to\infty} d_\alpha(\rho_0^n, \rho_1^n) \\ &= d_\alpha(\rho_0, \rho_1), \end{aligned}$$

and hence concluding the proof. □

Remark 5.4.3. Observe that in the previous Theorem, we did not only prove the equality of the minima, but we also provided a natural way to pass from a minimizer of our formulation *à la* Benamou-Brenier to a minimizer of the traffic plans model and back. The two problems are thus equivalent in the sense that they describe the same kind of energy and the same optimal structures of branched transportation: the simple equality of the minima (5.4.2) is just a part of this more important fact.

5. A simple comparison between the distances d_α and $w_{1/\alpha}$

This last section is devoted to estimates between the distance d_α induced by the branched transport and the Wasserstein distances w_p. In particular, in [**70**] the following estimates are proven[1] for $\alpha > 1 - 1/N$:

$$d_\alpha \le C\, w_p^{N(\alpha-1)+1}.$$

As far as lower bounds on d_α are concerned, the most trivial one is $d_\alpha \ge w_1$, but in [**43**, Theorem 8.1] this is improved up to prove $d_\alpha \ge w_{1/\alpha}$, which is slightly better. Moreover for scaling reasons (with respect to the mass) it is not possible to go beyond $p = 1/\alpha$ in this last inequality.

Observe that in this chapter we already needed to estimate some branched transport cost in terms of Wasserstein distances and metric derivatives: this is why this section will be devoted to prove the inequalities

$$w_{1/\alpha} \le d_\alpha \le C\, w_{1/\alpha}^{N(\alpha-1)+1}.$$

Despite being essentially already known, the proof we will provide will be different.

In particular, the formulation of branched transport we gave in this chapter provides an almost straightforward proof of the lower bound: anyway, the main tool (i.e. Inequality (5.3.1)) is essentially in common with [**43**] and [**68**]. What is different is the way to extend this idea to generic measures, i.e. non-atomic ones.

THEOREM 5.5.1. *For every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ we get*

$$w_{1/\alpha}(\rho_0, \rho_1) \le d_\alpha(\rho_0, \rho_1). \tag{5.5.1}$$

PROOF. We first observe that thanks to the results of the previous section, for every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ we get

$$d_\alpha(\rho_0, \rho_1) = \mathfrak{B}_\alpha(\rho_0, \rho_1) = \int_0^1 \left[\int_\Omega |v_t(x)|\, \mu_t(\{x\})^\alpha\, d\#(x)\right] dt.$$

for a suitable (μ, ϕ) admissible in the formulation (5.2.5), with $\phi = v \cdot \mu$. Moreover using once more the inequality (5.3.1) the right-hand side in the previous expression can be estimated as follows

$$\int_0^1 \left[\int_\Omega |v_t(x)|\, \mu_t(\{x\})^\alpha\, d\#(x)\right] dt \ge \int_0^1 \|v_t\|_{L^{1/\alpha}(\mu_t)}\, dt$$

and finally, using the fact that (μ, ϕ) is solution of the continuity equation, we can infer (see Chapter 1, Theorem 1.5.4)

$$|\mu_t'|_{w_{1/\alpha}} \le \|v_t\|_{L^{1/\alpha}(\mu_t)}, \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in [0,1],$$

so that

$$d_\alpha(\rho_0, \rho_1) \ge \int_0^1 |\mu_t'|_{w_{1/\alpha}}\, dt \ge w_{1/\alpha}(\rho_0, \rho_1),$$

where in the last inequality we just estimated the length of a curve with the distance between its endpoints. Thus we have obtained (5.5.1), concluding the proof. □

---

[1] As already observed, for $\alpha \in [0, 1 - 1/N]$ one can always find probability measures ρ_0 and ρ_1 such that $d_\alpha(\rho_0, \rho_1) = +\infty$. Then an inequality of the type $d_\alpha \le C\, w_p^\beta$ is not possible for α belonging to $[0, 1 - 1/N]$.

In order to prove the other inequality, we will use the same notations about dyadic cubes and approximations as in Chapter 3, Section 4. We shall always assume that $\Omega \subset K_L$ for a suitable L, then the following estimate is well-known ([**17**, Proposition 6.6]) and it is exactly the same as in Proposition 3.4.3 (properly speaking, it is exactly the contrary).

PROPOSITION 5.5.2. *Let $\alpha \in (1-1/N, 1]$, then for every $\mu \in \mathscr{P}(\Omega)$ we have*

$$d_\alpha(a_j(\mu), \mu) \leq \frac{2^{(N(1-\alpha)-1)j}}{2^{1-N(1-\alpha)}-1}\, \frac{L\sqrt{N}}{2}. \tag{5.5.2}$$

The fundamental tool in proving the estimate from above

$$d_\alpha(\rho_0, \rho_1) \leq C_N\, w_{1/\alpha}^{N(\alpha-1)+1}(\rho_0, \rho_1),$$

is to show that the distance d_α between two dyadic approximations can be estimated in terms of their $1/\alpha-$Wasserstein distance: this is the content of the next result.

LEMMA 5.5.3. *Let $\alpha \in (1-1/N, 1]$, then for every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ we get*

$$d_\alpha(a_j(\rho_0), a_j(\rho_1)) \leq C\, w_{1/\alpha}(a_j(\rho_0), a_j(\rho_1))\, 2^{jN(1-\alpha)}, \tag{5.5.3}$$

with C depending only on N and α.

PROOF. Let us consider a $\gamma_j \in \Pi(a_j(\rho_0), a_j(\rho_1))$ optimal for the cost $c(x,y) = |x-y|^{1/\alpha}$, that is $\gamma_j \in \mathscr{P}(\Omega \times \Omega)$ is of the form

$$\gamma_j = \sum_{i=1}^n \sum_{k=1}^s m_j(i,k) \delta_{x_j^i} \otimes \delta_{x_j^k}.$$

with $\sum_{i=1}^n m_j(i,k) = \rho_1(K_j^k)$ and $\sum_{k=1}^s m_j(i,k) = \rho_0(K_j^i)$, for every k, i.

Observe that by means of Proposition 1.2.3 of Chapter 1, γ_j can be taken in such a way that

$$\#\{(i,k)\,:\, m_j(i,k) \neq 0\} \leq 2 \cdot 2^{jN}, \tag{5.5.4}$$

that is γ_j does not move more than $2 \cdot 2^{jN}$ atoms. Setting $|x_j^i - x_j^k| = \ell_{i,k}$, we then get

$$w_{1/\alpha}(a_j(\rho_0), a_j(\rho_1)) = \left(\sum_{i=1}^n \sum_{k=1}^s m_j(i,k)\, \ell_{i,k}^{1/\alpha}\right)^\alpha,$$

and using (5.5.4) and Jensen's inequality

$$\begin{aligned} d_\alpha(a_j(\rho_0), a_j(\rho_1)) &\leq \sum_{i=1}^n \sum_{k=1}^s m_j(i,k)^\alpha \ell_{i,k} = \sum_{i=1}^n \sum_{k=1}^s \left(m_j(i,k)\, \ell_{i,k}^{1/\alpha}\right)^\alpha \\ &\leq \left(\sum_{i=1}^n \sum_{k=1}^s m_j(i,k)\, \ell_{i,k}^{1/\alpha}\right)^\alpha (\#\{(i,k)\,:\, m_j(i,k) \neq 0\})^{1-\alpha} \\ &\leq C\, w_{1/\alpha}(a_j(\rho_0), a_j(\rho_1))\, 2^{jN(1-\alpha)}, \end{aligned}$$

concluding the proof. □

THEOREM 5.5.4. *Let $\alpha \in (1-1/N, 1]$, then for every $p \geq 1/\alpha$, we get*

$$d_\alpha(\rho_0,\rho_1) \leq C\, w_p(\rho_0,\rho_1)^{N(\alpha-1)+1}, \tag{5.5.5}$$

with a constant C depending only on N, α and the diameter of Ω.

PROOF. It is enough to show the validity of (5.5.5) for $p = 1/\alpha$, then the general case will be just a consequence of the monotonicity property of the Wasserstein distances, i.e.

$$w_{1/\alpha} \leq w_p, \quad \text{for every } p \geq 1/\alpha.$$

Using triangular inequality, (5.5.2) and (5.5.3), we get for every $j \in \mathbb{N}$

$$\begin{aligned} d_\alpha(\rho_0,\rho_1) &\leq d_\alpha(\rho_0, a_j(\rho_0)) + d_\alpha(a_j(\rho_0), a_j(\rho_1)) + d_\alpha(a_j(\rho_1),\rho_1) \\ &\leq C\, 2^{(N(1-\alpha)-1)j} + d_\alpha(a_j(\rho_0), a_j(\rho_1)) \\ &\leq C\, 2^{(N(1-\alpha)-1)j} + C\, w_{1/\alpha}(a_j(\rho_0), a_j(\rho_1))\, 2^{jN(1-\alpha)}, \end{aligned}$$

and

$$\begin{aligned} w_{1/\alpha}(a_j(\rho_0), a_j(\rho_1)) &\leq w_{1/\alpha}(a_j(\rho_0),\rho_0) + w_{1/\alpha}(\rho_0,\rho_1) + w_{1/\alpha}(a_j(\rho_0),\rho_1) \\ &\leq C\, 2^{-j} + w_{1/\alpha}(\rho_0,\rho_1), \end{aligned}$$

which finally gives

$$\begin{aligned} d_\alpha(\rho_0,\rho_1) &\leq C\, 2^{(N(1-\alpha)-1)j} + C\, w_{1/\alpha}(\rho_0,\rho_1)\, 2^{jN(1-\alpha)} \\ &= C\, 2^{(N(1-\alpha)-1)j} \left(1 + w_{1/\alpha}(\rho_0,\rho_1)\, 2^j\right) \end{aligned}$$

It is now sufficient to choose the index j in such a way that

$$\frac{\operatorname{diam}(\Omega)}{2^j} \leq w_{1/\alpha}(\rho_0,\rho_1) \leq \frac{\operatorname{diam}(\Omega)}{2^{j-1}},$$

which in turn yields

$$2^{(N(1-\alpha)-1)j}(1 + w_{1/\alpha}(\rho_0,\rho_1)2^j) \leq C\, w_{1/\alpha}(\rho_0,\rho_1)^{N(\alpha-1)+1},$$

thus giving the thesis. □

REMARK 5.5.5. As we briefly mentioned, the comparison between d_α and $w_{1/\alpha}$ is the most natural one, as far as

$$d_\alpha(\rho_0,\rho_1) \simeq \sum m^\alpha\, \ell \quad \text{and} \quad w_{1/\alpha}(\rho_0,\rho_1) \simeq \left(\sum m\, \ell^{\frac{1}{\alpha}}\right)^\alpha,$$

and they have the same kind of homogeneity.

REMARK 5.5.6. One may wonder if the exponent $N(\alpha-1)+1$ in inequality (5.5.5) can be improved: actually, the answer is *no*. To see this, it is enough to adapt Example 0.1 of [**70**].

CHAPTER 6

Transportation models for congested dynamics

1. Introduction

With this chapter, we leave the setting of branched transportation, turning out attention to problems in which the cost has to satisfy exactly the opposite requirement: during the transportation, masses have to stay separate from each others, in order to avoid as much as possible *congestion effects*. From a mathematical point of view, this is usually translated into a cost with an increasing, convex and superlinear dependence on the mass, so that moving an amount of mass m on a distance ℓ costs $\psi(m)\,\ell$, with ψ convex and superlinear: typical choices are power-like functions, for example $\psi(t) = t^p$ with $p > 1$, or combinations of them. Observe that in economical terms ψ' corresponds to the *marginal cost*, so that requiring $\psi' > 0$ and ψ' strictly increasing, that is ψ strictly convex, is very natural from a modelization point of view. In this way, provided $\psi(0) = 0$, one also gets that $\psi(m_1) + \psi(m_2) \leq \psi(m_1 + m_2)$, so that it is more convenient to separate the mass in order to save cost.

A daily life situation in which this type of problems occur is certainly well-known to the reader: urban traffic. Anyway, the models that we will introduce in this chapter does not really deal with dynamical situations of traffic (a topic which goes beyond the scopes of this work), but they are more pertinent to model steady-state situations, averaged over a day or a period.

The chapter is organized as follows: in Section 2 we describe a discrete traffic problem (introduced by Wardrop in [**88**]) which has some interesting issues on equilibria and some interesting connections with optimal transport. The main points of this section are the definition of Wardrop equilibrium and the fact that existence for such equilibria can be proven through a variational principle, that is minimizing a total cost for traffic congestion. The next section contains an attempt to extend this model to a continuous setting, by means of employing traffic plans, i.e. probability measures over the set of admissible paths: we briefly review the work [**33**] by Carlier, Jimenez and Santambrogio, where this extension has been proposed. As we will see, it is possible to trace a nice parallel with the discrete case, in particular it is still possible to give a suitable definition of Wardrop equilibrium and to show existence of such equilibria by means of the optimization of an overall congested transportation cost. Finally, in Section 4 the Beckmann's continuous model of transportation is discussed (we have already encountered a particular case of it in Chapter 1, Section 1): this is expressed as a divergence-constrained convex optimization problem, in which the transportation activities are described through vector fields satisfying the balance conditions

$$\mathrm{div}\phi = \rho_0 - \rho_1,\ \langle \phi, \nu \rangle = 0,$$

we have a total cost of integral type $\int_\Omega \mathcal{H}(\phi)\,dx$ for the whole transportation process and we look for a minimizer. This model is much more in the spirit of a global optimization problem, as far as the individual welfare of the commuters between ρ_0 and ρ_1 is disregarded: we are minimizing the total cost of transportation, in the presence of congestion, for a social planner. Anyway, we will see in the next chapter that there is a tight connection between these two models and that Beckmann's problem is directly linked to Wardrop equilibrium issues.

2. Wardrop's model in a discrete setting

We start with the discrete case: this means that sources and destinations are given by a finite number of points, while possible routes connecting them are to be chosen among the edges of a fixed network. We will see in the next section that it is possible to give a suitable extension of this model to a continuous setting.

Specifically, in a discrete setting one considers:

- a finite graph $\mathscr{G}$ consisting of a set of edges E and a set of vertices V, the latter containing $S = \{x_1, \dots, x_n\}$ sources and $D = \{y_1, \dots, y_m\}$ destinations;
- the set $C^{x_i,y_j} = \{\sigma : [0,1] \to \mathscr{G} \,:\, \sigma(0) = x_i,\, \sigma(1) = y_j\}$ of admissible paths (on the graph $\mathscr{G}$) from x_i to y_j;
- a demand input $\gamma = (\gamma(x_i, y_j))_{i,j}$ denoting the quantity of mass from each $x_i \in S$ that has to be sent to each $y_j \in D$: it is also possible to consider a set Π of admissible γ's (in a modern language, this is nothing but an admissible set of transport plans);
- an unknown *repartition strategy* $Q = (Q_\sigma)_\sigma$, such that
$$\sum_{\sigma \in C^{x_i,y_j}} Q_\sigma = \gamma(x_i, y_j),$$
which represents the possible way in which mass can split on the graph in going from x_i to y_j, in order to satisfy the demand input relative to these points;
- a resulting *traffic intensity* $i_Q = (i_Q(e))_{e\in E}$, which clearly depends on the strategy Q, given by
$$i_Q(e) = \sum_{\{\sigma \,:\, \sigma([0,1]) \supset e\}} Q_\sigma,$$
that is for every edge $e \in E$ of the graph, $i_Q(e)$ is the total mass transiting from there, according to the repartition strategy;
- finally, an increasing function $h : [0,\infty) \to [0,\infty)$, such that $h(i_Q(e))$ represents the congestioned cost of the edge e.

Then one defines the total cost of each admissible path $\sigma \in C^{x_i,y_j}$, given by
$$w(\sigma) = \sum_{\{e \in \mathscr{G} \,:\, e \subset \sigma([0,1])\}} h(i_Q(e))\, \mathscr{H}^1(e),$$
and observe that it increasingly depends on the traffic intensity, that is on how much this σ is used by the transportation. The global strategy Q represents the overall traffic distribution on the network Σ, that is it represents the choices of the commuters from S to D. Imposing a *Nash equilibrium* (see [**87**]), which roughly speaking means that no single commuter wants to change his choice, provided all the other players keep the same strategy, gives the following condition:

if on an admissible $\sigma \in C^{x_i,y_j}$ we have $Q_\sigma > 0$ (which means that somebody is using this path σ), then

$$w(\sigma) = \min\{w(\widetilde{\sigma}) \,:\, \widetilde{\sigma} \in C^{x_i,y_j}\},$$

that is *at equilibrium every used path must be minimal, according to the congestioned cost induced by the traffic itself.* This is the well-known concept of *Wardrop equilibrium*, which can also be rephrased as "every actually used (i.e. where the flow is positive) road connecting a source to a destination should be a shortest path (taking into account the congestion effects)".

For our purposes, it is useful to recall the fundamental fact that existence of such an equilibrium can be proven by means of a variational principle: the convexity assumptions in the next statement guarantee that necessary optimality conditions are also sufficient.

Proposition 6.2.1. *Let $H : [0,\infty) \to [0,\infty)$ be a convex function such that $H'(t) = h(t)$, for every $t \geq 0$. Also suppose that the set Π of admissible demand inputs is convex. Then $\widetilde{Q}$ minimizes the overall congestion cost*

$$\mathcal{J}(Q) = \sum_{e\in\mathscr{G}} H(i_Q(e))\,\mathscr{H}^1(e),$$

among all possible strategies if and only if $\widetilde{Q}$ is a Wardrop equilibrium for the congestion function h.

Proof. Let us give a brief justification of this important fact. Using the convexity of H, it is easy to see that $\widetilde{Q}$ minimizes $\mathcal{J}$ if and only if

$$\sum_{e\in\mathscr{G}} H'(i_{\widetilde{Q}(e)})\,(i_Q(e) - i_{\widetilde{Q}}(e))\,\mathscr{H}^1(e) \geq 0, \quad \text{for every admissible } Q.$$

Set $\xi(e) := H'(i_{\widetilde{Q}}(e))$ and use the definition of traffic intensity, so that the left-hand side can be rewritten as

$$\begin{aligned}\sum_{e\in\mathscr{G}} H'(i_{\widetilde{Q}}(e))\,(i_Q(e) - i_{\widetilde{Q}}(e))\,\mathscr{H}^1(e) &= \sum_{e\in\mathscr{G}}\ \sum_{\{\sigma\,:\,\sigma([0,1])\supset e\}} \xi(e)\,\mathscr{H}^1(e)\,\left(Q_\sigma - \widetilde{Q}_\sigma\right)\\ &= \sum_{\sigma}\left(\sum_{\{e\in\mathscr{G}\,:\,e\subset\sigma([0,1])\}} \xi(e)\,\mathscr{H}^1(e)\right)\left(Q_\sigma - \widetilde{Q}_\sigma\right),\end{aligned}$$

and setting $\ell_\xi(\sigma) = \sum_{\{e\subset\sigma([0,1])\}} \xi(e)\,\mathscr{H}^1(e)$ (observe that this can be seen as a weighted length of the curve σ), the previous implies

$$\sum_\sigma \ell_\xi(\sigma)\,Q_\sigma \geq \sum_\sigma \ell_\xi(\sigma)\,\widetilde{Q}_\sigma,$$

so that $\widetilde{Q}$ minimizes $\mathcal{J}$ if and only if

$$\begin{aligned}\sum_\sigma \ell_\xi(\sigma)\,\widetilde{Q}_\sigma &= \inf_Q \sum_\sigma \ell_\xi(\sigma)\,Q_\sigma\\ &= \inf_Q \sum_{x_i\in S,\, y_j\in D}\ \sum_{\sigma\in C^{x_i,y_j}} \ell_\xi(\sigma)\,Q_\sigma.\end{aligned}$$

We then observe that the total quantity of commuters between x_i and y_j is fixed, being given by $\sum_{\sigma\in C^{x_i,y_j}} Q_\sigma = \gamma(x_i,y_j)$, then setting

$$d_\xi(x_i,y_j) = \min\{\ell_\xi(\sigma) \,:\, \sigma\in C^{x_i,y_j}\}, \quad \text{for every } x_i\in S,\ y_j\in D. \tag{6.2.1}$$

we easily see that the previous infimum is given by

$$\sum_\sigma \ell_\xi(\sigma)\,\widetilde{Q}_\sigma = \inf_{\gamma\in\Pi} \sum_{x_i\in S,\, y_j\in D} d_\xi(x_i,y_j)\,\gamma(x_i,y_j).$$

This means that $\widetilde{Q}$ minimizes $\mathcal{J}$ if and only if for every $\sigma\in C^{x_i,y_j}$ such that $\widetilde{Q}_\sigma>0$, then σ must be of minimal length according to ℓ_ξ: thus recalling the definition of ℓ_ξ and w above and taking into account that $H'=h$, we conclude the proof. □

Remark 6.2.2. Observe that in the case Π consists of the whole set of transport plans (i.e. we prescribe only the total amount of traffic exiting $x_i\in S$ and the total amount of traffic entering $y_j\in D$ and we look at all possible couplings), the previous computations show that a minimizer $\widetilde{Q}=(\widetilde{Q}_\sigma)_\sigma$ of the functional $\mathcal{J}$ solves the accessory Monge-Kantorovich problem for the congestioned metric d_ξ, more precisely the element $\widetilde{\gamma}$ defined by $\widetilde{\gamma}(x_i,y_j) = (\sum_{x_i\in S,\, y_j\in D}\widetilde{Q}_\sigma)_{i,j}$ minimizes the quantity

$$\sum_{x_i\in S,\, y_j\in D} d_\xi(x_i,y_j)\,\gamma(x_i,y_j),$$

with $\xi = H'(i_{\widetilde{Q}})$ and d_ξ given by (6.2.1).

Remark 6.2.3. It is worth observing that in general the overall congestion cost

$$\mathcal{J} = \sum_{e\in\mathscr{G}} H(i_Q(e))\,\mathscr{H}^1(e),$$

is different from the total cost paid by vehicles, the latter being given by

$$\sum_{e\in\mathscr{G}} h(i_Q(e))\mathscr{H}^1(e)\,i_Q(e).$$

Clearly if one takes $H'(t)=h(t)$, the two functions $H(t)$ and $h(t)t$ are the same up to a multiplicative constant in the case of power functions; but otherwise they give rise to different optimization problems.

3. Wardrop's model in a continuous setting

Recently, in [**33**] Carlier, Jimenez and Santambrogio have extended Wardrop's model to a continuous setting, i.e. when S and D are replaced by generic probability measures ρ_0 and ρ_1, while the set of admissible paths is no more constrained to live on a given graph, but it is the whole space of continuous curves (see also [**14**] for some numerical simulations based on their model).

For the sake of completeness and to motivate the studies of the next chapter, where we will try to relate Wardrop's and Beckmann's models (the latter will be presented in the next section), we will briefly explain the model and some of the results of [**33**].

In the framework of [**33**], let $\Omega \subset \mathbb{R}^N$ be a bounded open set with smooth boundary and let $\rho_0, \rho_1 \in \mathscr{P}(\overline{\Omega})$ be two given probability measures over its closure, representing the spatial distribution of centers of production and consumption of a given commodity (or residential and working/commercial areas), over the geographical area Ω. The network $\mathscr{G}$ is replaced by the whole space $C([0,1];\overline{\Omega})$, thought as a separable Banach space equipped with the topology of uniform convergence. In order to describe how commuters between ρ_0 and ρ_1 choose their paths, one introduces probability measures Q on $C([0,1];\overline{\Omega})$, concentrated on the set $\mathrm{Lip}([0,1];\Omega)$ and satisfying the compatibility conditions of connecting ρ_0 to ρ_1, in the sense

$$(e_i)_\sharp Q = \rho_i,\ i = 0, 1,$$

where e_t is the evaluation-at-time-t map: this Q is the natural counterpart of the repartition strategy in the discrete model. We set $\mathcal{Q}(\rho_0, \rho_1)$ to denote these admissible probability measures and we can still keep the terminology *traffic plans* for them, as in the previous Chapters. Then the *traffic intensity*, i.e. the transiting mass associated to Q, is represented by a positive Borel measure i_Q on Ω, defined through

$$\int_\Omega \varphi(x)\, di_Q(x) := \int_{C([0,1];\overline{\Omega})} \left(\int_0^1 \varphi(\sigma(t))\, |\sigma'(t))| dt \right) dQ(\sigma), \quad \text{for every } \varphi \in C(\overline{\Omega}).$$

REMARK 6.3.1. It is straightforward to see that i_Q is a generalized transport density: indeed, it could be seen as representing a path-dependent version of the usual transport density for Monge's problem. Taking $\gamma \in \Pi(\rho_0, \rho_1)$ an optimal transport plan for

$$\min \left\{ \int_{\Omega\times\Omega} |x-y|\, d\gamma(x,y) \,:\, \gamma \in \Pi(\rho_0, \rho_1) \right\},$$

and choosing $Q = \int \delta_{\overline{xy}}\, d\gamma(x,y)$, where $\overline{xy}$ is a parametrization of the segment joining x to y, we easily get

$$\langle \varphi, i_Q \rangle = \int_{\Omega\times\Omega} \int_{\overline{xy}} \varphi(z)\, d\mathcal{H}^1(z)\, d\gamma(x,y), \quad \text{for every } \varphi \in C(\Omega),$$

that is i_Q coincides with the usual notion of transport density (see Chapter 1, Section 1).

Then the concept of Wardrop equilibrium can be extended in a natural way to this setting as follows: a $Q \in \mathcal{Q}(\rho_0, \rho_1)$ is said to be *a Wardrop equilibrium for the congestion function h* if it gives full mass to the set of geodesics for the metric d_Q (formally) defined as

$$d_Q(x,y) = \inf_{\sigma \in C^{x,y}} \int_0^1 h(i_Q(\sigma(t)))\, |\sigma'(t)|\, dt. \tag{6.3.1}$$

Observe that this is in fact a metric which depends on the traffic itself, as one should expect, and this is the exact translation in a continuous setting of the principle "every actually used (i.e. where the flow is positive) road connecting a source and a destination must be a shortest path (taking into account the congestion effects)", characterizing Wardrop equilibria in a discrete case.

REMARK 6.3.2. The quantity $h(i_Q(x))$ can be seen as the cost to pay for passing from x, where there is an amount of traffic given by $i_Q(x)$. If $i_Q(x)$ is interpreted as the number (per unit of volume) of vehicles transiting from x, then (formally)

$$\int_\Omega h(i_Q(x))\, i_Q(x)\, dx, \tag{6.3.2}$$

represents the total cost paid by vehicles in commuting between ρ_0 and ρ_1, corresponding to the traffic assignment Q.

Anyway, this is just an informal presentation of what a Wardrop equilibrium should be in the continuous setting: indeed, from a mathematical point of view the definition of the metric d_Q does make little sense, at this level. Observe that for those Q such that $i_Q \ll \mathscr{L}^N$, the previous definition could be interpreted as

$$d_Q(x,y) = \inf_{\sigma\in C^{x,y}} \int_0^1 h\left(\frac{di_Q}{d\mathscr{L}^N}(\sigma(t))\right) |\sigma'(t)|\, dt.$$

In this case one can prove that when the term

$$h\left(\frac{di_Q}{d\mathscr{L}^N}\right),$$

is an L^q function, with $q > N$, this metric d_Q can be rigorously defined (see [**33**, Section 3.2] for more details), through an approximation procedure and the definition of Wardrop equilibrium consequently does make sense.

Regarding the existence of such equilibria, one may wonder if this can be achieved again through a variational principle, exactly as in the discrete case: the answer is *yes*, so let us now turn the attention to the optimization problem. One introduces a total cost functional

$$W(Q) = \begin{cases} \int_\Omega H(i_Q(x))\, dx, & \text{if } i_Q \ll \mathscr{L}^N, \\ +\infty, & \text{otherwise,} \end{cases}$$

where we are confusing the measure i_Q with its density w.r.t. $\mathscr{L}^N$: in particular, by the definition of W we see that the congestion effects in this model will be quite strong. Indeed, with a cost like this only very diffused (i.e. absolutely continuous w.r.t $\mathscr{L}^N$) traffic intensity are allowed and every low-dimensional concentration of the traffic gives rise to an infinite total cost.

The density-cost function $H : [0,\infty) \to [0,\infty)$ is assumed to be increasing and strictly convex, with a $p-$growth ($p > 1$), that is

$$a\, t^p \le H(t) \le b\,(t^p + 1), \quad \text{for every } t \ge 0,$$

and $H(0) = 0$: then the following optimization problem is taken into account

$$\inf_{Q\in\mathcal{Q}^p(\rho_0,\rho_1)} \int_\Omega H(i_Q(x))\, dx, \tag{6.3.3}$$

where the set of admissible traffic plans is given by

$$\mathcal{Q}^p(\rho_0,\rho_1) = \{Q \in \mathcal{Q}(\rho_0,\rho_1) \,:\, i_Q \in L^p(\Omega)\}.$$

REMARK 6.3.3. It is important to observe that $W(Q) = \int_\Omega H(i_Q(x))\,dx$ can be seen as the total cost of congestion for a social planner and in general it differs from the total cost payed by vehicles, the latter being given by (6.3.2). From a modelistic viewpoint, it is also clear that minimizing W with the constraint $(e_i)_\sharp Q = \rho_i$, $i = 0, 1$, we are mimicking a situation in which the social planner has the right to impose *who goes where*, in order to optimize the total cost: in mathematical terms, this corresponds to say that the set of admissible couplings $(e_0, e_1)_\sharp Q$ coincides with the whole set of transport plans $\Pi(\rho_0, \rho_1)$. In a real urban traffic situation, on the other hand, it is more realistic to optimize the total cost under the constraint of $(e_0, e_1)_\sharp Q \in \Pi$, with $\Pi \subset \Pi(\rho_0, \rho_1)$ being a set of admissible transport plans: as a particular case, we have $\Pi = \{\gamma\}$, that is the coupling is given and, roughly speaking, every commuter starting from ρ_0 have decided its destination (this is the situation usually considered by transport analysts, see [**12**]). The definition of Wardrop equilibrium still applies to this situation.

REMARK 6.3.4. At a first sight, it could seem difficult to check that given $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, then there exists Q such that i_Q is admissible, that is $i_Q \in L^p(\Omega)$. Anyway, by means of Theorem 1.1.13 in Chapter 1 about summability results on the transport density, we have for example that ρ_0 and ρ_1 having L^p densities surely implies that $\mathcal{Q}^p(\rho_0, \rho_1) \neq \emptyset$.

Anyway, we can have $\mathcal{Q}^p(\rho_0, \rho_1) \neq \emptyset$ also in the case that neither ρ_0 nor ρ_1 are absolutely continuous w.r.t. $\mathscr{L}^N$, as the next Example shows.

EXAMPLE 6.3.5. Let us take $\rho_0 = (\omega_{N-1})^{-1} \mathscr{H}^{N-1} \llcorner \partial B_1(0)$ and $\rho_1 = \delta_0$, with $\partial B_1(0) = \{x \in \mathbb{R}^N : |x| = 1\}$ and $\omega_{N-1} = \mathscr{H}^{N-1}(\partial B_1(0))$. Let us consider the transport density relative to ρ_0 and ρ_1 for the cost $|x - y|$, that is

$$\begin{aligned}
\langle i_\gamma, \varphi \rangle &= \frac{1}{\omega_{N-1}} \int_{\partial B_1(0)} \int_0^1 \varphi((1-t)x)\,|x|\,dt\,d\mathscr{H}^{N-1}(x) \\
&= \frac{1}{\omega_{N-1}} \int_0^1 \int_{\partial B_s(0)} \varphi(y) \frac{1}{s^{N-1}}\,d\mathscr{H}^{N-1}(y)\,ds \\
&= \frac{1}{\omega_{N-1}} \int_{B_1(0)} \varphi(x)\,\frac{1}{|x|^{N-1}}\,dx, \quad \text{for every } \varphi \in C,
\end{aligned}$$

that is $i_\gamma = \omega_{N-1} |x|^{1-N} \cdot \mathscr{L}^N$, whose density belongs to every L^p with $1 \leq p < N/(N-1)$.

EXAMPLE 6.3.6. For some choices of ρ_0 and ρ_1 singular w.r.t. $\mathscr{L}^N$, we could also obtain a traffic intensity with an L^∞ density w.r.t. $\mathscr{L}^N$. For example, with $N = 2$ it is enough to consider

$$\rho_0 = \mathscr{H}^1 \llcorner (\{0\} \times [0,1]) \quad \text{and} \quad \rho_1 = \mathscr{H}^1 \llcorner (\{1\} \times [0,1]),$$

then taking as i_Q the corresponding transport density, this is the uniform 2−dimensional Lebsegue measure on the square $[0,1] \times [0,1]$.

About existence of solutions for problem (6.3.3), it is proven in [**33**] the following result.

THEOREM 6.3.7. *The minimization problem* (6.3.3) *admits a solution, provided* $\mathcal{Q}^p(\rho_0, \rho_1)$ *is not empty.*

PROOF. We just give a sketch of the proof, highlighting the main ideas and referring the interested reader to [**33**] for further details. The proof goes along the lines of the Direct Methods:

indeed, taking a minimizing sequence $\{Q_n\}_{n\in\mathbb{N}}$, the main issue is to show that this is equi-tight (which implies $*-$weak convergence to a probability Q, by means of Prokhorov's Theorem) and that, if every Q_n is concentrated on Lipschitz curves, then the same holds true for the limit measure Q. To obtain the equi-tightness property, one observes that, assuming

$$\int_\Omega H(i_{Q_n}(x))\,dx \le C \quad \text{for every } n \in \mathbb{N},$$

then $\{i_{Q_n}\}_{n\in\mathbb{N}}$ is equi-bounded in $L^p(\Omega)$, which in turn implies equi-boundedness in $L^1(\Omega)$, so that

$$\int_C \int_0^1 |\sigma'(t)|\,dt\,dQ_n(\sigma) = \int_\Omega i_{Q_n}(x)\,dx \le C,$$

and the required property on the sequence $\{Q_n\}_{n\in\mathbb{N}}$ is now an easy consequence of Markov inequality: to be more precise, we obtain the equi-tightness of the sequence $\{\widetilde{Q}_n\}_{n\in\mathbb{N}}$, where $\widetilde{Q}_n = (\mathfrak{r})_\sharp Q_n$ and $\mathfrak{r}$ is the map that associates to every $\sigma \in \mathrm{Lip}([0,1];\Omega)$ its constant speed reparametrization $\widetilde{\sigma} := \mathfrak{r}(\sigma)$, that is $\widetilde{\sigma} : [0,1] \to \Omega$ is such that $|\widetilde{\sigma}'(t)| \equiv \int_0^1 |\sigma'(t)|\,dt$ and $\widetilde{\sigma}([0,1]) = \sigma([0,1])$. The important fact is that $i_{Q_n} = i_{\widetilde{Q}_n}$.

We also observe that the fact that $\{i_{Q_n}\}_{n\in\mathbb{N}}$ is equi-bounded in $L^p(\Omega)$ implies weak convergence (up to subsequences, as always) to an L^p function i.

Clearly, this would not be enough, as far as we do not know that this limit function i is itself a traffic intensity: anyway, one can show that $\widetilde{Q}_n \overset{*}{\rightharpoonup} Q$ and $i_{Q_n} \rightharpoonup i$, implies that $i_Q \le i$ (observe that this is a sort of lower semicontinuity result). Observing that the set $\mathcal{Q}(\rho_0,\rho_1)$ is closed with respect to the $*-$weak convergence and using the monotonicity and convexity properties of H, then one can conclude

$$\int_\Omega H(i_Q(x))\,dx \le \int_\Omega H(i(x))\,dx \le \liminf_{n\to\infty} \int_\Omega H(i_{Q_n}(x))\,dx,$$

giving the optimality of Q. □

We then write optimality conditions in the continuous setting, which are strictly related to the concept of Wardrop equilibrium and read as follows: as it will be clear in a while, this is the natural counterpart of Proposition 6.2.1 of the previous section.

PROPOSITION 6.3.8. *$Q_0 \in \mathcal{Q}^p(\rho_0,\rho_1)$ is a solution of (6.3.3) if and only if*

$$\int_\Omega H'(i_{Q_0}(x))\,i_{Q_0}(x)\,dx = \inf_{Q\in\mathcal{Q}^p(\rho_0,\rho_1)} \left\{ \int_\Omega H'(i_{Q_0}(x))\,i_Q(x)\,dx \right\} \tag{6.3.4}$$

It is interesting to manipulate a bit (at least at a formal level) the previous optimality condition: as we see, this is nothing but the fact that the optimal transportation strategies are concentrated on geodesics for the congested metric.

Indeed, using the definition of i_Q, setting $\xi = H'(i_{Q_0})$, $\gamma_0 = (e_0,e_1)_\# Q_0 \in \Pi(\rho_0,\rho_1)$ and disintegrating Q_0 as

$$Q_0 = \int_{\Omega\times\Omega} Q^{x,y}\,d\gamma_0(x,y),$$

where for γ_0-a.e. $(x,y)\in\Omega\times\Omega$, the probability measure $Q^{x,y}$ is concentrated on the set $C^{x,y}=\{\sigma : \sigma(0)=x,\ \sigma(1)=y\}$, then the optimality condition (6.3.4) can be formally rewritten as

$$\begin{aligned}\int_\Omega \xi(x)\, i_{Q_0}(x)\, dx &= \inf_Q \int_\Omega \xi(x)\, i_Q(x)\, dx\\ &= \inf_Q \int_{C([0,1];\Omega)} \int_0^1 \xi((\sigma(t)))\, |\sigma'(t)|\, dt\, dQ(\sigma)\\ &= \inf_Q \int_{\Omega\times\Omega} \int_{C^{x,y}} \left(\int_0^1 \xi(\sigma(t))\, |\sigma'(t)|\, dt\right) dQ^{x,y}(\sigma)\, d\gamma(x,y)\\ &= \inf_{\gamma\in\Pi(\rho_0,\rho_1)}\ \inf_{\eta\in\mathscr{P}(C^{x,y})} \int_{C^{x,y}} \left(\int_0^1 \xi(\sigma(t))\, |\sigma'(t)|\, dt\right) d\eta(\sigma)\, d\gamma(x,y)\\ &= \inf_{\gamma\in\Pi(\rho_0,\rho_1)} \int_{\Omega\times\Omega} d_\xi(x,y)\, d\gamma(x,y),\end{aligned}$$

where we have set

$$d_\xi(x,y) = \inf_{\sigma\in C^{x,y}} \int_0^1 \xi(\sigma(t))\, |\sigma'(t)|\, dt.$$

So, at least formally, exactly as in the discrete case we have that if Q_0 is a solution of (6.3.3), then it must be concentrated on the set of geodesics for the metric d_ξ, so that if $H'=h$ then Q_0 is a Wardrop equilibrium for the given congestion function h. Moreover in this case, where only the marginals ρ_0 and ρ_1 are fixed, but not the coupling (i.e. the transport plan), we further obtain that the corresponding $\gamma_0=(e_0,e_1)_\sharp Q_0\in\Pi(\rho_0,\rho_1)$ has to solve the Monge-Kantorovich problem associated with the metric d_ξ. Conversely, both conditions are also sufficient for optimality in (6.3.3).

Under appropriate assumptions, the whole previous discussion can be made rigorous: we summarize the main result about existence and characterization of Wardrop equilibria via optimization of a total cost in the following statement (see [**33**, Theorem 3.10]).

Theorem 6.3.9. *Let us assume that $1<p<N/(N-1)$ and that H is strictly convex. A transportation strategy $Q_0\in\mathcal{Q}^p(\rho_0,\rho_1)$ is optimal for (6.3.3) if and only if one has:*

(i) $\gamma_0=(e_0,e_1)_\# Q_0\in\Pi(\rho_0,\rho_1)$ solves the Monge-Kantorovich problem

$$\inf_{\gamma\in\Pi(\rho_0,\rho_1)} \int_{\Omega\times\Omega} d_{Q_0}(x,y)\, d\gamma(x,y),$$

where the cost d_{Q_0} is defined by

$$d_{Q_0}(x,y) = \inf_{\sigma\in C^{x,y}} \int_0^1 H'(i_{Q_0}(\sigma(t)))\, |\sigma'(t)|\, dt$$

(ii) Q_0 gives full mass to the set of curves such that

$$\int_0^1 H'(i_{Q_0}(\sigma(t)))\, |\sigma'(t)|\, dt = d_{Q_0}(\sigma(0),\sigma(1)),$$

that is Q_0 defines a metric through the coefficient $H'(i_{Q_0})$ and it is indeed concentrated on geodesics for this metric.

It is part of the assertion that, with the previous hypothesis on the exponent p, the equilibrium metric d_{Q_0} is well-defined (see [**33**, Section 3.2]): indeed, this implies that for the conjugate exponent $q = p/(p-1)$ we have $q > N$ and $H'(i_{Q_0}) \in L^q$, which is exactly the condition for the metric d_{Q_0} to be defined, as briefly said.

Thus summarizing, when a congestion function h is given, applying the previous Theorem with H such that $H' = h$, we obtain the existence of a Wardrop equilibrium relative to h. Moreover exactly as in the discrete case, the existence of such an equilibrium is obtained by means of the minimization of a total transportation cost.

REMARK 6.3.10. The conclusions of Theorem 6.3.7 and Theorem 6.3.9 (in the sense that Q is optimal if and only if is a Wardrop equilibrium), still hold in the case of the minimization problem for W under the constraint $Q \in \Pi$, with Π convex subset of $\Pi(\rho_0, \rho_1)$, in particular for the case $\Pi = \{\gamma\}$. On the other hand, condition (i) of Theorem 6.3.9 is peculiar of the optimization problem for W over $\mathcal{Q}^p(\rho_0, \rho_1)$, with no constraints on the coupling $(e_0, e_1)_\sharp Q$.

4. Beckmann's continuous model of transportation

We now turn our attention to the model introduced by Beckmann in [**11**], in which the starting point is the optimization point of view: we have already encountered a particular case of this model in Chapter 1, Section 1.

We can think of modelizing the transportation activities in the city Ω through a vector field $\phi : \Omega \to \mathbb{R}^N$, which in every point x describes the direction of transportation, given by $\phi(x)/|\phi(x)|$, and the total amount of transiting mass, given by $|\phi(x)|$: observe that for every ϕ, the quantity $|\phi|$ somehow plays the role of the traffic intensity i_Q in Wardrop's model. At every point, ϕ should satisfy some balance conditions, assuring that locally the total outcoming/incoming flow of commodity depends on the difference between production and consumption, i.e. $\rho_0 - \rho_1$: that is to say, if in small region $\mathcal{U}$ around x we have more production than consumption, we expect a compensating flow exiting from $\mathcal{U}$ and vice versa.

Mathematically this can be expressed by saying

$$\int_{\partial\mathcal{U}} \langle \phi(x), \nu(x) \rangle \, d\mathscr{H}^{N-1}(x) = \rho_0(\mathcal{U}) - \rho_1(\mathcal{U}),$$

which can also be written as a constraint on the divergence of the vector field ϕ

$$\operatorname{div} \phi = \rho_0 - \rho_1. \tag{6.4.1}$$

Having assumed that ρ_0 and ρ_1 have the same mass, we also have a homogeneous Neumann boundary condition

$$\langle \phi, \nu \rangle = 0, \quad \text{on } \partial\Omega, \tag{6.4.2}$$

stating that the geographical area Ω is self-sufficient with respect to the commodity considered (i.e. we have no import/export of the given commodity). One can also consider situations in which ρ_0 and ρ_1 are positive measures with different masses: in this case, the previous homogeneous boundary condition should be replaced by a non-homogeneous one, meaning that the city Ω needs an outcoming/incoming flow of commodity, in order to be in equilibrium.

Clearly, one can also think of ρ_0 and ρ_1 as the distribution of residents and services (shops, offices etc.) in the city Ω: again, if they have the same mass, then roughly speaking every citizen can satisfy his needs without going outside Ω.

We then introduce some density cost function $\mathcal{H} : \Omega \times \mathbb{R}^N \to \mathbb{R}^+$, depending on the space variable x and on the transiting mass: more precisely, the quantity $\mathcal{H}(x, \phi(x))$ gives the total cost (per unit volume) to let an amount of mass $|\phi(x)|$ transit with direction $\phi(x)/|\phi(x)|$ in the point x, so that in the end the transportation process associated to a certain vector field ϕ satisfying (6.4.1) and (6.4.2) will have a total cost given by

$$\int_\Omega \mathcal{H}(x, \phi(x))\, dx. \tag{6.4.3}$$

Then one is is interested in looking for a ϕ minimizing (6.4.3), under the constraints (6.4.1) and (6.4.2). So far, this problem is clearly too general and not well-posed, unless some specific and reasonable assumptions are made on the cost function $\mathcal{H}$ and on the space of admissible vector fields. In the sequel, for simplicity we will only consider $x-$independent cost $\mathcal{H}$, satisfying the following assumptions:

(i) $\mathcal{H}(z) = H(|z|)$, with $H : [0, \infty) \to [0, \infty)$ strictly convex superlinear function and with $\mathcal{H}(0) = 0$ (we will think of the function H as the same as in Wardrop's model);

(ii) $\mathcal{H}$ has $p-$growth, i.e.

$$a|z|^p \leq \mathcal{H}(z) \leq b(|z|^p + 1),\ z \in \mathbb{R}^N,$$

for some $p \in (1, \infty)$ and a, b positive constants;

(iii) $\mathcal{H}$ is differentiable in $\mathbb{R}^N \setminus \{0\}$ and there exists a positive constant c such that

$$|\nabla \mathcal{H}(z)| \leq c(|z|^{p-1} + 1),\ z \in \mathbb{R}^N \setminus \{0\},$$

then we turn to consider the following optimization problem (*Beckmann's problem*)

$$\inf_{\phi \in L^p(\Omega;\mathbb{R}^N)} \left\{ \int_\Omega \mathcal{H}(\phi(x))\, dx\ :\ \operatorname{div} \phi = \rho_0 - \rho_1,\ \langle \phi, \nu \rangle = 0 \text{ on } \partial\Omega \right\}. \tag{6.4.4}$$

Example 6.4.1. An interesting case is given by the choice $H(z) = |z|^2$, for which the minimal value (6.4.4) is given by (see [**34**] for the details)

$$\mathfrak{A}(\rho_0, \rho_1) = \begin{cases} \|\rho_0 - \rho_1\|^2_{X^*}, & \text{if } \rho_0 - \rho_1 \in X^* \\ +\infty, & \text{otherwise.} \end{cases}$$

where X^* indicates the dual of the Hilbert space $X = W^{1,2}_\diamond(\Omega) = \{\varphi \in W^{1,2}(\Omega)\ :\ \int_\Omega \varphi = 0\}$, equipped with the scalar product

$$\langle \varphi, \psi \rangle_X = \int_\Omega \langle \nabla \varphi(x), \nabla \psi(x) \rangle\, dx,$$

so that in this case the minimal value of Beckmann's problem is just a dual Sobolev norm. We will come back later to this problem in the next chapter.

Optimization problem (6.4.4) enjoys nice optimality conditions that we will exploit extensively in the next chapters: they read as follows.

THEOREM 6.4.2. *Suppose that the infimum in* (6.4.4) *is finite and let* ϕ_0 *be its unique optimizer, then there exists* $\varphi_0 \in W^{1,q}(\Omega)$ *such that*

$$\phi_0 = \nabla\mathcal{H}^*(\nabla\varphi_0), \tag{6.4.5}$$

and φ_0 *is a weak solution of*

$$\begin{cases} \operatorname{div}\nabla\mathcal{H}^*(\nabla u) &= \rho_0 - \rho_1, & \text{in } \Omega, \\ \langle\nabla\mathcal{H}^*(\nabla u), \nu\rangle &= 0, & \text{on } \partial\Omega, \end{cases} \tag{6.4.6}$$

where $\mathcal{H}^*$ *is the Legendre transform of* $\mathcal{H}$ *and* $q = p/(p-1)$.

PROOF. We first observe that problem (6.4.4) consists in minimizing a strictly convex and coercive functional on L^p subject to a convex and closed constraint: then an optimizer ϕ_0 exists and must be unique.

It is well known that problem (6.4.4) has a dual formulation, given by the convex analysis formula (see for instance [**47**])

$$\sup_{\varphi\in W^{1,q}_\diamond(\Omega)} \left\{\langle\rho_1-\rho_0,\varphi\rangle - \int_\Omega \mathcal{H}^*(\nabla\varphi)\right\} = \inf_{\phi\in L^p(\Omega)} \left\{\int_\Omega \mathcal{H}(\phi)\ :\ \operatorname{div}\phi = \rho_0-\rho_1,\ \langle\phi,\nu\rangle = 0\right\},$$

where $W^{1,p}_\diamond(\Omega)$ is the subspace of $W^{1,p}(\Omega)$ consisting of functions with zero mean on Ω and the brackets $\langle\cdot,\cdot\rangle$ denote the duality product between $W^{1,q}_\diamond(\Omega)$ and its dual$\Big(W^{1,q}_\diamond(\Omega)\Big)^*$. Due to the superlinear growth and the strict convexity of $\mathcal{H}$, we get that $\mathcal{H}^* \in C^1$ and it verifies the following growth conditions

$$B(|z|^q - 1) \le \mathcal{H}^*(z) \le A|z|^q,$$

where $q = p/(p-1)$, then using the Direct Methods of the Calculus of Variations it is not difficult to show that the dual problem admits at least a solution φ_0 on $W^{1,q}_\diamond(\Omega)$. Indeed the functional

$$\mathfrak{F}(\varphi) = \int_\Omega \mathcal{H}^*(\nabla\varphi(x))\,dx - \langle\rho_1-\rho_0,\varphi\rangle,$$

is lower semicontinuous with respect to the weak topology of $W^{1,q}_\diamond$ (due to the convexity of $\mathcal{H}^*$) and is coercive, thanks to the growth conditions of $\mathcal{H}^*$ and to Poincarè and Young inequalities

$$\begin{aligned}\mathfrak{F}(\varphi) &\ge B\int_\Omega |\nabla\varphi(x)|^q\,dx - B|\Omega| - \|\rho_0-\rho_1\|\left(\int_\Omega |\varphi(x)|^q\,dx\right)^{\frac{1}{q}} \\ &\ge B\int_\Omega |\nabla\varphi(x)|^q\,dx - B|\Omega| - C\,\|\rho_0-\rho_1\|\left(\int_\Omega |\nabla\varphi(x)|^q\,dx\right)^{\frac{1}{q}} \\ &\ge (B - C\,\varepsilon)\int_\Omega |\nabla\varphi(x)|^q\,dx - B|\Omega| - \frac{C}{\varepsilon}\|\rho_0-\rho_1\|^p.\end{aligned}$$

It is clear that, from a variational point of view, minimizing $\mathfrak{F}$ or maximizing $-\mathfrak{F}$ is exactly the same. We observe further that the Euler-Lagrange equation of $\mathfrak{F}$ is given by (6.4.6), so that φ_0

solves it in distributional sense. Moreover φ_0 and ϕ_0 verify

$$\begin{aligned}\int_\Omega \mathcal{H}(\phi_0(x))\,dx &= \langle \rho_1-\rho_0,\varphi\rangle - \int_\Omega \mathcal{H}^*(\nabla\varphi_0(x))\,dx \\ &= \int_\Omega \langle \nabla\varphi_0(x),\phi_0(x)\rangle\,dx - \int_\Omega \mathcal{H}^*(\nabla\varphi_0(x))\,dx,\end{aligned}$$

where we have used the fact that $\operatorname{div}\phi_0=\rho_0-\rho_1$ and $\langle\phi_0,\nu\rangle=0$. The previous can be recast into

$$\int_\Omega \mathcal{H}(\phi_0(x))\,dx + \int_\Omega \mathcal{H}^*(\nabla\varphi_0(x))\,dx = \int_\Omega \langle \nabla\varphi_0(x),\phi_0(x)\rangle\,dx,$$

which, by means of the so called Legendre reciprocity formula, implies that

$$\phi_0(x)\in\partial\mathcal{H}^*(\nabla\varphi_0(x)),\quad \text{for } \mathscr{L}^N\text{-a.e. } x\in\Omega.$$

Using the fact that $\mathcal{H}^*\in C^1$, we obtain that actually the subgradient set $\partial\mathcal{H}^*$ is made of just an element, namely the gradient $\nabla\mathcal{H}^*$, concluding the proof. □

Definition 6.4.3. In the sequel, we will refer to a function φ_0 verifying (6.4.5) as a *Beckmann potential*: observe that this can be simply seen as a Lagrange multiplier for the divergence constraint on ϕ_0.

Remark 6.4.4. We can (formally, at this stage) define a cost function given by the Riemannian distance associated to $|\nabla\varphi_0|$, that is

$$d_{\varphi_0}(x,y)=\inf_{\sigma\in C^{x,y}}\int_0^1 |\nabla\varphi_0(\sigma(t))||\sigma'(t)|\,dt,$$

where $C^{x,y}=\{\sigma\,:\,\sigma(0)=x,\ \sigma(1)=y\}$, then we see

$$\int_0^1 |\nabla\varphi_0(\sigma(t))||\sigma'(t)|\,dt \geq \int_0^1 \langle\nabla\varphi_0(\sigma(t)),\sigma'(t)\rangle\,dt = \varphi_0(\sigma(1))-\varphi_0(\sigma(0))$$

that is

$$d_{\varphi_0}(x,y)\geq \inf_{\sigma\in C^{x,y}} \varphi_0(\sigma(1))-\varphi_0(\sigma(0)) = \varphi_0(y)-\varphi_0(x),$$

and we have equality if and only if σ is an integral curve of $\nabla\varphi_0$ connecting x to y: in this case, there holds

$$d_{\varphi_0}(x,y)=\varphi_0(y)-\varphi_0(x).$$

This means that a Beckmann potential φ_0 acts like a Kantorovich potential for the metric given by d_{φ_0}, which is induced by φ_0 itself. More precisely (with a little abuse of notation, as far as $\mathcal{H}(z)=H(|z|)$ is not differentiable at 0) notice that d_{φ_0} is induced by $|\nabla\varphi_0|=H'(|\phi_0|)$, in the same way in which the equilibrium metric in Wardrop's model was induced by $H'(i_{Q_0})$: this gives an evident link between the two models, which in particular could suggest that existence of Wardrop equilbria for a congestion function $h=H'$ can be proven by looking at problem (6.4.4). The investigation of such a connection between the two problems is exactly the content of the next chapter.

CHAPTER 7

Equivalence between Wardrop's and Beckmann's models

1. Introduction

In this chapter we discuss how to connect the Wardrop's problem on measures on paths (Chapter 6, Section 3) to the Beckmann's problem on vector fields with prescribed divergence (Chapter 6, Section 4): in which sense and when are they equivalent and, more important, how to pass from a minimizer of the first problem to a minimizer of the second one and back. Throughout this chapter, Wardrop's problem will always be considered in the case that no constraints on the couplings $(e_0, e_1)_\sharp Q$ are imposed: on the other hand, if you fix an admissible proper subset of transport plans $\Pi \subset \Pi(\rho_0, \rho_1)$ (the typical choice is $(e_0, e_1)_\sharp Q = \gamma \in \Pi(\rho_0, \rho_1)$), the two problems would rather describe different situations (see also Remark 7.2.2 below). Then the main result is that of Theorem 7.5.5 where equivalence is proven using the concept of flow *à la* DiPerna-Lions and using some regularity results for the solution of Beckmann problem, whose proof are postponed to Chapters 8 and 9, in order not to bury the central ideas of this result in technical (though interesting) regularity details.

We point out that one of the main features of this equivalence is that, once established, one can prove the existence of Wardrop equilibria for a given congestion function, by solving the more familiar (and more tractable, also from a numerical point of view) convex optimization problem presented in Section 4 of Chapter 6. Moreover one can, roughly speaking, investigate the regularity properties of these Wardrop equilibria, in terms of the regularity of the data ρ_0 and ρ_1, through a detailed analysis of the corresponding Beckmann potentials. In what follows, for convenience we will use the notation

$$(\mathcal{W}) = \min_{Q \in \mathcal{Q}^p(\rho_0, \rho_1)} \int_\Omega H(i_Q(x))\, dx,$$

and

$$(\mathcal{B}) = \min_{\phi \in L^p(\Omega; \mathbb{R}^N)} \left\{ \int_\Omega \mathcal{H}(\phi)\ dx\ :\ \operatorname{div} \phi = \rho_0 - \rho_1,\ \langle \phi, \nu \rangle = 0 \text{ on } \partial\Omega \right\},$$

to denote Wardrop and Beckmann problems, respectively, where we recall that we are considering the isotropic case $\mathcal{H}(z) = H(|z|)$ for every $z \in \mathbb{R}^N$. The model case that we will have in mind for the cost function H is the following

$$\mathcal{H}(z) = \frac{1}{p}|z|^p + |z|,\ z \in \mathbb{R}^N, \tag{7.1.1}$$

with $p > 1$, that is $(\mathcal{B})$ amounts in minimizing a combination of the L^p and L^1 norms, under a divergence constraint. The reason for such a choice is readily said: let us recall that $H' = h$, where h is the congestion function relating the metric to the traffic intensity. It is therefore natural to have $h(0) > 0$: *the metric is positive even if there is no traffic*, which implies that the cost function $\mathcal{H}$

should not be differentiable at 0 and then its subdifferential at 0 contains a ball. In more economic terms, you could rephrase this by saying that we are considering a model where the marginal cost for congestion (represented by H') is strictly increasing and always greater than a threshold $c > 0$.

2. Moser's deformation argument

First of all, we start with the following simple result, stating that actually $(\mathcal{W})$ and $(\mathcal{B})$ can be compared: here the particular choice (7.1.1) does not play any role and the result is valid under the hypothesis that $H(z)$ has $p-$growth, with $p > 1$.

PROPOSITION 7.2.1. *Let* $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ *such that* $\mathcal{Q}^p(\rho_0, \rho_1) \neq \emptyset$. *Then*

$$(\mathcal{B}) \leq (\mathcal{W}). \tag{7.2.1}$$

PROOF. It is sufficient to consider an admissible $Q \in \mathcal{Q}^p(\rho_0, \rho_1)$ and to construct a vector-valued version of the traffic intensity i_Q, namely we define a vector measure ϕ_Q on Ω through

$$\langle \phi_Q, \varphi \rangle := \int_{C([0,1];\overline{\Omega})} \left(\int_0^1 \langle \varphi(\sigma(t)), \sigma'(t) \rangle \, dt \right) dQ(\gamma), \text{ for every } \varphi \in C(\overline{\Omega}, \mathbb{R}^N),$$

then it is not difficult to see that ϕ_Q satisfies the divergence constraint of Beckmann's problem (just taking test functions of the form $\varphi = \nabla \psi$). Moreover by means of Cauchy-Schwarz inequality we easily get $|\phi_Q| \leq i_Q$, so that $\phi_Q \in L^p$ and it is indeed admissible for $(\mathcal{B})$. Using the monotonicity properties of the cost function H, one can say that

$$\int_\Omega \mathcal{H}(\phi_Q(x))\, dx = \int_\Omega H(|\phi_Q(x)|)\, dx \leq \int_\Omega H(i_Q(x))\, dx,$$

thus giving (7.2.1). □

Thanks to Proposition 7.2.1, we see that to prove the full equivalence $(\mathcal{B}) = (\mathcal{W})$, we have to show that given ϕ solving $(\mathcal{B})$, it is possible to construct $Q \in \mathcal{Q}^p(\rho_0, \rho_1)$ such that $i_Q = |\phi|$: this will automatically imply that Q is a solution to $(\mathcal{W})$, with

$$\int_\Omega H(i_Q(x))\, dx = \int_\Omega \mathcal{H}(\phi(x))\, dx,$$

thus giving the desired equivalence. Anyway, before to explain how to achieve such a construction, something has to be precised about the difference between the two models.

REMARK 7.2.2. The construction in the proof of Proposition 7.2.1 should clarify that the main difference between the two formulations is that Beckmann's one, roughly speaking, does not take into account the fact that a huge mass transiting in opposite directions on the same path creates a lot of congestion effects: on the contrary, in Beckmann's construction, this two opposite flows would be added in a vectorial way, thus giving rise to a lot of cancellations. To give a precise flavour of this phenomenon, we give a basic but important example: suppose to have $\rho_0 = \rho_1 = 1/2\, \delta_{x_0} + 1/2\, \delta_{x_1}$, with $x_0 \neq x_1$, then we consider the traffic plan

$$Q = \frac{1}{2}\delta_{\sigma_1} + \frac{1}{2}\delta_{\sigma_2},$$

with $\sigma_1(t) = (1-t)x_0 + tx_1$ and $\sigma_2(t) = (1-t)x_1 + tx_0$, that is the initial and final destinations are the same (two Dirac masses), but we want to exchange the mass in x_0 with that in x_1 and vice versa. Computing the traffic intensity, gives

$$i_Q = \mathscr{H}^1 \llcorner \overline{x_0 x_1},$$

which takes into account the intuitive fact that on the segment $\overline{x_0 x_1}$, globally there is a non negligible amount of transiting mass. On the other hand, the construction of Proposition 7.2.1 gives the vector measure

$$\phi_Q = 0,$$

because of the fact that the definition of ϕ_Q takes into account the orientation of the curves. In any case, it is evident that in Beckmann model, there is no possibility to modelize a situation like this, which is anyway very natural. *A posteriori* we could say that, the equivalence $(\mathcal{B}) = (\mathcal{W})$ should tell us that at the optimum no such cancellations occur and that *transport rays do not intersect*, whatever this means in the present context. Observe that this is clearly strongly linked to the fact that in $(\mathcal{W})$, we are not prescribing the coupling $(e_0, e_1)_\sharp Q$ (see Remark 6.3.3 of the previous chapter), but we are optimizing among all possible transport plans.

We now come to illustrate how to construct Wardrop equilibria, starting from an optimal vector field ϕ: disregarding for a moment regularity issues, we will see that a natural candidate Q_ϕ is given by

$$Q_\phi := \int \delta_{X(.,x)}\, d\rho_0(x),$$

that is

$$\int_{C([0,1];\overline{\Omega})} F(\sigma)\, dQ_\phi(\sigma) = \int_\Omega F(X(\cdot,x))\, d\rho_0(x), \quad \text{for every } F \in C(C([0,1];\overline{\Omega});\mathbb{R}),$$

where $X(.,x)$ is the flow of the non-autonomous ODE

$$\left\{\begin{array}{rcl} \partial_t X(t,x) & = & \widehat{\phi}_t(X(t,x)), \\ X(0,x) & = & x, \end{array}\right. \qquad \widehat{\phi}_t(x) := \frac{\phi(x)}{(1-t)\rho_0(x) + t\rho_1(x)}. \tag{7.2.2}$$

Indeed, suppose that $\rho_0 = f_0 \cdot \mathscr{L}^N$ and $\rho_1 = f_1 \cdot \mathscr{L}^n$, then by its very construction the flow map X satisfies

$$\frac{d}{dt}\Big[\det \nabla_x X(t,x)\, ((1-t) f_0(X(t,x)) + t f_1(X(t,x)))\Big] = 0,$$

so that the quantity inside the square brackets is actually constant in time: then using the fact that the flow at the initial time is the identity map, one obtains

$$f_0(x) = f_1(X(1,x))\, \det \nabla_x X(1,x),$$

which by means of the area formula implies $\rho_1 = (X(1,\cdot))_\sharp \rho_0$. This guarantees that $(e_i)_\sharp Q = \rho_i$, for $i = 0, 1$, and moreover the natural concept of transport density associated to this system, which is precisely the traffic intensity i_Q, is such that (we will provide the details below)

$$i_{Q_\phi} = |\phi|,$$

thus giving the desired equivalence. This is clearly a Lagrangian point of view: on the other hand, one observes that considering the linear interpolating curve $\rho_t(x) = (1-t)\rho_0(x) + t\rho_1(x)$, then

$$\begin{aligned}\partial_t \rho_t + \mathrm{div}_x(\widehat{\phi}_t\, \rho_t) &= \rho_1 - \rho_0 + \mathrm{div}_x \left[\frac{\phi}{(1-t)\rho_0 + t\rho_1} \left((1-t)\rho_0 + t\rho_1\right)\right] \\ &= \rho_1 - \rho_0 + \mathrm{div}_x \phi = 0,\end{aligned}$$

by means of the divergence constraint on ϕ, that is $\widehat{\phi}$ is a velocity field for the measure-valued curve ρ. Moreover provided we have well-posedness of the Cauchy problem for the continuity equation, with initial datum ρ_0, then we have that ρ_t must coincide with the curve $\mu_t = (X(t,\cdot))_\sharp \rho_0$, so that in particular

$$\rho_1 = X(1,\cdot)_\# \rho_0,$$

and the same calculations as before apply, thus giving again $i_Q = |\phi|$: this is the Eulerian point of view.

This deformation argument has been essentially introduced by Moser in [**71**], in order to solve the problem of constructing smooth maps with prescribed Jacobian: further refinements and improvements of this method have then been given by Dacorogna and Moser himself ([**37**]) and by Rivière and Ye ([**74**]), among others. The application of this argument to Optimal Transport problems is not completely new: indeed, we recall that in this context the first striking application has been given by Evans and Gangbo ([**48**]), in their celebrated paper about the existence of a solution for Monge's original problem.

So far, this is the heuristic argument in order to obtain the desired equivalence: we now have to take into account regularity issues and to see if, and eventually in what sense, the previous construction does make sense. This said, the rest of the chapter is devoted to give an answer to the following question:

when and in what sense does the flow of (7.2.2) *exist?*

3. Cauchy-Lipschitz flow

If the vector field $\widehat{\phi}$ defined in (7.2.2) is Lipschitz with respect to the space variable, this flow can be defined in a classical sense (see Appendix B) and the situation is relatively easy to understand: in particular, the computations performed in the previous section are admissible and we obtain the desired equality. Anyway, the regularity properties of $\widehat{\phi}$ strongly depend on two facts:

- the regularity of the data, that is ρ_0, ρ_1 and $\partial\Omega$;
- the regularity of ϕ, which in turns depends not only on the data, but also on the kind of cost function $\mathcal{H}$ we are considering.

Indeed, recall that by Chapter 6, Theorem 6.4.2, we already know that the solution of $(\mathcal{B})$ satisfies

$$\phi = \nabla \mathcal{H}^*(\nabla u),$$

where $\mathcal{H}^*$ is the Legendre-Fenchel conjugate of $\mathcal{H}$ and u is a Beckmann potential, solving the following Neumann value boundary problem

$$\left\{\begin{array}{rcll} -\mathrm{div}\nabla\mathcal{H}^*(\nabla u) &=& \rho_1 - \rho_0, & \text{in } \Omega, \\ \langle \nabla\mathcal{H}^*(\nabla u), \nu\rangle &=& 0, & \text{on } \partial\Omega. \end{array}\right. \tag{7.3.1}$$

So it quite clear that, though we could impose all the needed assumptions on the data (let's say C^∞), the gain of regularity on ϕ depends on the regularity of the solutions of (7.3.1) and thus has some limits, intrinsic in the elliptic operator, beyond which is not possible to go. Let us make some examples.

EXAMPLE 7.3.1. In the case $\mathcal{H}(z) = 1/2\,|z|^2$ (which is the one considered in [**34**]), then the primal-dual optimality condition reads as

$$\phi = \nabla u,$$

and (7.3.1) is a Neumann problem for the Poisson equation $-\Delta u = \rho_1 - \rho_0$. Then to obtain $\phi \in C^{0,1}$ it is widely sufficient to assume that ρ_0 and ρ_1 have densities given by $C^{0,\beta}$ functions: indeed, under these hypotheses, the classical Schauder estimates (see [**55**]) give $\phi = \nabla u \in C^{1,\beta}$, up to the boundary if $\partial\Omega$ is smooth enough. In this case, furtherly assuming that ρ_0 and ρ_1 have Lipschitz densities bounded from below by $c > 0$, we see that $\widehat{\phi}$ defined by (7.2.2) is Lipschitz in the space variable and Moser's argument does apply to this situation.

EXAMPLE 7.3.2. More generally, with the choice $\mathcal{H}(z) = 1/p\,|z|^p$ for $p \neq 2$, condition (6.4.5) leads to (q is the conjugate exponent of p)

$$\phi = |\nabla u|^{q-2}\nabla u,$$

and a Beckmann potential solves the homogeneous Neumann problem for the Poisson-like equation

$$-\Delta_q u = \rho_1 - \rho_0,$$

where Δ_q is the $q-$Laplacian operator. Already in this case, it is not clear if Lipschitz regularity on ϕ (and consequently on $\widehat{\phi}$) can be achieved: indeed, the by now classical regularity results for the $q-$Laplace equation (see for example [**45**] and [**64**]) assures that in general solutions are no more that $C^{1,\beta}$, no matter how regular the data are, which means Hölder regularity for ϕ. A possible higher regularity on the solution should involve a discussion on the critical points of the solution itself, which is quite a delicate matter.

Yet, the situations which are motivated by traffic congestion is even worse: recall our choice (7.1.1) for the cost $\mathcal{H}$. Its Fenchel transform is given by

$$\mathcal{H}^*(\xi) = \frac{1}{q}(|\xi| - 1)_+^q,\ \xi \in \mathbb{R}^N,$$

which turns (7.3.1) into a very degenerate elliptic problem: for its solutions no more than Lipschitz regularity (on the potential u!) can be achieved (see the discussion at the end of the chapter), thus giving just an L^∞ result on ϕ. In the end, for the case to study this Cauchy-Lipschitz interpretation of Moser's argument does not apply.

4. Superposition of flows

For a general vector field v under very mild regularity assumptions, the most general meaning that we can give to the flow of v is in terms of the superposition principle, that we have introduced in Chapter 1, Section 5. As we have seen, this provides a very weak concept of flow (actually, it is a sort of probabilistic one), which anyway is strong enough to still give sense to Moser's deformation argument.

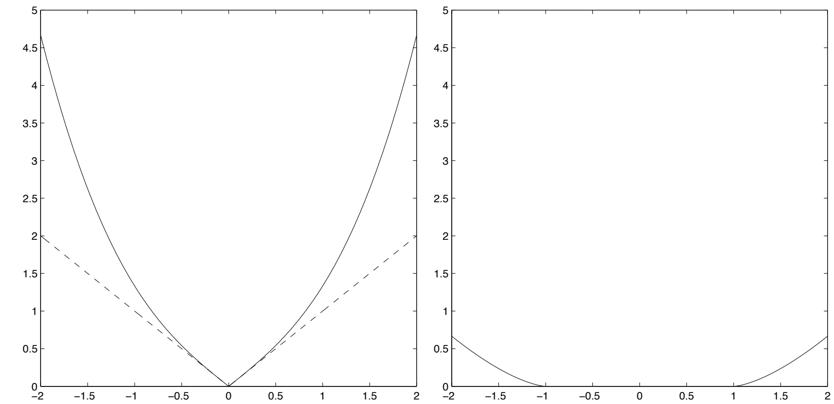

FIGURE 1. The function $\mathcal{H}$ (left) given by (7.1.1) and its Legendre-Fenchel conjugate $\mathcal{H}^*$ (right).

First of all, in this case we need the following slight variation of the superposition principle, dealing with the degenerate case $p = 1$. For the proof, the reader can consult [**5**, Theorem 12], where the integrability condition (7.4.1) is slightly different: this is due to the fact that we are working on a bounded set Ω, which allows to simplify a bit the assumptions.

THEOREM 7.4.1 (Superposition principle, case $p = 1$). *Let μ_t be a positive measure-valued solution of the continuity equation*

$$\partial_t \mu_t + \operatorname{div}(v_t\, \mu_t) = 0,$$

with the vector field v satisfying the following integrability condition

$$\int_0^1 \int_\Omega |v_t(x)|\, d\mu_t(x)\; dt < +\infty, \tag{7.4.1}$$

then μ_t is a superposition solution, that is $\mu_t = (e_t)_\# Q$, with Q concentrated on absolutely continuous integral curves of v (see Chapter 1, Definition 1.5.2).

Using the concept of superposition solution, it is now a straightforward fact to provide a rigorous proof of the equivalence between the two problems $(\mathcal{B})$ and $(\mathcal{W})$.

THEOREM 7.4.2. *Let $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ having L^p density w.r.t. to $\mathscr{L}^N$, given by f_0 and f_1, respectively. Then the equality*

$$(\mathcal{B}) = (\mathcal{W}),$$

holds true.

PROOF. As already illustrated, we take the minimizer ϕ of $(\mathcal{B})$ and we consider the non-autonomous vector field defined by (7.2.2). We point out that the L^p assumption on the densities has been chosen in order to guarantee finiteness of the infima of both problems. With this choice of $\widehat{\phi}$, the linear interpolating curve $\mu_t = (1-t)\rho_0 + t\rho_1$ is a positive measure-valued distributional solution of the continuity equation

$$\partial_t \mu_t + \operatorname{div}(\widehat{\phi}_t\, \mu_t) = 0,$$

with initial datum ρ_0. Moreover $\widehat{\phi}$ satisfies hypothesis (7.4.1), as far as

$$\int_0^1 \int_\Omega |\widehat{\phi}_t(x)| \, d\mu_t(x) \, dt = \int_0^1 \int_\Omega |\widehat{\phi}_t(x)| [(1-t) f_0(x) + t f_1(x)] \, dx \, dt = \int_\Omega |\phi(x)| \, dx < +\infty,$$

so that μ_t is a superposition solution by means of Theorem 7.4.1: this means that there exists a probability measure $Q \in \mathscr{P}(C([0,1]; \overline{\Omega}))$ concentrated on integral curves of $\widehat{\phi}$ and such that $\mu_t = (e_t)_\sharp Q$: observe that in the Cauchy-Lipschitz case, this amounted to say that

$$(1-t)\rho_0 + t\rho_1 = (X(t,\cdot))_\# \rho_0,$$

thanks to the well-posedness of Cauchy problem. This Q is admissible, that is $Q \in \mathcal{Q}(\rho_0, \rho_1)$ and moreover using Fubini Theorem and the disintegration $Q = \int Q_x \, d\rho_0(x)$, we get

$$\begin{aligned}
\int_{\overline{\Omega}} \varphi(x) \, di_Q(x) &= \int_C \int_0^1 \varphi(\sigma(t)) |\sigma'(t)| \, dt \, dQ(\sigma) \\
&= \int_0^1 \int_C \varphi(\sigma(t)) |\sigma'(t)| \, dQ(\sigma) \, dt \\
&= \int_0^1 \int_\Omega \int_C \varphi(\sigma(t)) |\widehat{\phi}_t(\sigma(t))| \, dQ_x(\sigma) \, d\rho_0(x) \, dt \\
&= \int_0^1 \int_\Omega \varphi(x) |\widehat{\phi}(t,x)| \, d\mu_t(x) \, dt = \int_0^1 \int_\Omega \varphi(x) |\phi(x)| \, dx \, dt,
\end{aligned}$$

so that

$$\int_\Omega \varphi(x) \, di_Q(x) = \int_\Omega \varphi(x) |\phi(x)| \, dx, \quad \text{for every } \varphi \in C(\overline{\Omega}).$$

This clearly implies that $i_Q = |\phi|$ and thus $Q \in \mathcal{Q}^p(\rho_0, \rho_1)$ and it solves Wardrop's problem, thus concluding the proof. □

Notice that the regularity of the curves which are charged by the measure Q corresponding to a superposition solution is very poor. On the contrary, if $\widehat{\phi}$ is known to be continuous, these curves are C^1 and they solve their ODE in a classical sense. In a recent paper ([**78**]), Santambrogio and Vespri have proven a (local) C^0 result for the vector field we are interested in, that is

$$\phi = \nabla \mathcal{H}^*(\nabla u),$$

with $\mathcal{H}$ given by (7.1.1), when $N = 2$. Obviously, continuity without Lipschitz continuity or similar conditions (log-Lipschitz or more generally an Osgood condition, for example) is not sufficient for ensuring any kind of uniqueness result for the flow. We will see in a while that some kind of uniqueness may be recovered by an intermediate concept of solution.

5. Wardrop equilibria in the Di Perna-Lions sense

Up to now, we have seen that everything goes well if we face a Lipschitz vector field v and that we can at least prove equality of the minima if, instead, v is merely integrable. In the latter case, it is not evident to add anything else to this equality and in particular one has no real clue to construct a minimizer for Wardrop's problem from a minimizer for Beckmann's one. The problem is mainly linked to the lack of uniqueness of solutions of the ODE. We will see in this section an intermediate concept, for vector fields which are not Lipschitz but much better than merely integrable: this is the flow of a weakly differentiable vector field, in the sense of DiPerna and Lions ([**46**]). In this case we can enforce the conclusion of Theorem 7.4.2 and guarantee that *the optimal* Q_ϕ *associated to the optimizer* ϕ *is actually concentrated on a uniquely defined flow* X *(possibly in a.e. sense), transporting* ρ_0 *to* ρ_1.

First of all, we need to recall some important facts from the theory of Di Perna and Lions: we start with the concept of renormalized solution.

DEFINITION 7.5.1. We say that $\mu \in L^\infty([0,1]; L^\infty(\Omega))$ is a *renormalized solution* of the continuity equation with vector field v if there holds

$$\partial_t \beta(\mu) + \langle v, \nabla_x \beta(\mu)\rangle + (\mathrm{div}_x v)\mu\, \beta'(\mu) = 0, \quad \text{in } (0,1)\times\Omega, \tag{7.5.1}$$

in the sense of distributions, for every $\beta \in C^1(\mathbb{R})$.

Observe that clearly every renormalized solution is a distributional solution (just take $\beta \equiv 1$ in (7.5.1)), while in general the converse does not hold true. It is a remarkable fact of the DiPerna-Lions theory that when v has a Sobolev regularity in x, then v has the *renormalization property*, that is every distributional solution is actually a renormalized one. Moreover renormalized solutions are the right class in which well-posedness for the continuity equation can be proven: *this is crucial for our construction*. We summarize these fundamental results in the following statement (see [**46**, Theorem II.3]): the hypotheses are slightly enforced to be easily adapted to the cases which are of interest for us.

THEOREM 7.5.2. *Let* $\Omega \subset \mathbb{R}^N$ *be a bounded open set having a smooth boundary. Let us assume that* v *satisfies*

$$v \in L^1([0,1]; W^{1,1}(\Omega)) \cap L^1([0,1]; L^\infty(\Omega)), \tag{7.5.2}$$

and

$$\mathrm{div}_x v \in L^1([0,1]; L^\infty(\Omega)), \tag{7.5.3}$$

with $\langle v, \nu\rangle = 0$ *at the boundary* $\partial\Omega$. *Then* $\mu \in L^\infty([0,1]; L^\infty(\mathbb{R}^N))$ *is a renormalized solution of the continuity equation if and only if it is a distributional solution. Moreover given* $\rho_0 \in L^\infty(\mathbb{R}^N)$, *then there exists a unique renormalized solution* μ *of the continuity equation in* $L^\infty([0,1]; L^\infty(\mathbb{R}^N))$ *corresponding to the initial datum* ρ_0.

REMARK 7.5.3. We point out that the renormalization property can be proven also for vector fields with BV regularity (with respect to the space variable), as shown by Ambrosio in [**2**].

THEOREM 7.5.4. *Let* Ω *be a bounded open set with smooth boundary. Let us assume that* v *satisfies the hypothesis of Theorem 7.5.2, with* $\langle v, \nu\rangle = 0$ *at the boundary* $\partial\Omega$. *Then there exists a unique flow map* $X \in C^0([0,1]\times[0,1]; L^1(\Omega; \mathbb{R}^N))$ *which leaves* $\overline{\Omega}$ *invariant and such that:*

(i) if we set $A(t) = \int_0^t \|\mathrm{div}_x v_r(\cdot)\|_\infty\, dr$, *then*

$$e^{-|A(t)-A(s)|}\mathscr{L}^N \le (X(t,s,\cdot))_\sharp \mathscr{L}^N \le e^{|A(t)-A(s)|}\mathscr{L}^N, \quad \text{for every } t,s \in [0,1];$$

(ii) X *satisfies the group property*

$$X(t_3,t_1,x) = X(t_3,t_2,X(t_2,t_1,x)), \quad \text{for } \mathscr{L}^N\text{-a.e. } x\in\Omega, \quad \text{for every } t_1<t_2<t_3 \in [0,1];$$

(iii) for every $s \geq 0$ and for $\mathscr{L}^N$-a.e. $x \in \Omega$, X is an absolutely continuous integral solution of $\sigma'(t) = v_t(\sigma(t))$ with $\sigma(0) = x$, that is

$$X(t, s, x) = x + \int_s^t v_r(X(r, s, x))\, dr, \quad \text{for } \mathscr{L}^N\text{-a.e. } x \in \Omega,\ t \geq s.$$

Moreover if $\rho_0 = f_0 \cdot \mathscr{L}^N$ with $f_0 \in L^\infty(\Omega)$, then for every $s \in [0, 1)$

$$\mu_t(x) = (X(t, s, \cdot))_\sharp \rho_0(x),\ s \leq t \in [0, 1],$$

is a (actually the unique, thanks to Theorem 7.5.2) renormalized solution in $L^\infty([s, 1]; L^\infty(\Omega))$ of the continuity equation, with initial datum $\mu_s(x) = \rho_0(x)$.

We are now ready to state the main result obtained in [**B4**] about the equivalence between Wardrop and Beckmann models and the consequent characterization of Wardrop equilibria, as measures supported on DiPerna-Lions flows. The proof will require quite involved regularity results that we have decided, for the sake of clearness, to postpone and to leave for the last two Chapters.

Theorem 7.5.5. *Let us consider the density cost function*

$$\mathcal{H}(z) = H(|z|),\ z \in \mathbb{R}^N,$$

given by (7.1.1) and assume that $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ are absolutely continuous with respect to the $N-$dimensional Lebesgue measure $\mathscr{L}^N$. Suppose moreover the following:

(i) $\rho_i = f_i \cdot \mathscr{L}^N$, with $f_i \in \mathrm{Lip}\,(\Omega)$ and $f_i \geq c > 0$, for $i = 0, 1$;
(ii) Ω open connected bounded subset of $\mathbb{R}^N$ having smooth boundary[1].

Then the vector field $\widehat{\phi}$ given by (7.2.2) is well-defined and satisfies the hypotheses of DiPerna-Lions Theorem 7.5.4, thus we obtain

$$(\mathcal{W}) = (\mathcal{B}).$$

In particular, a Wardrop equilibrium Q for the congestion function $h = H'$ does exist and it is supported on the flow (in the DiPerna-Lions sense) of $\widehat{\phi}$, i.e. given ϕ optimal for $(\mathcal{B})$, then

$$Q = \int \delta_{X(x,\cdot)}\, d\rho_0(x), \tag{7.5.4}$$

is optimal for $(\mathcal{W})$, where X is the flow map of $\widehat{\phi}$ given by Theorem 7.5.4.

Proof. With these assumptions at hand, assuming for the moment that they imply the following regularity for ϕ optimizing $(\mathcal{B})$

$$\phi \in W^{1,r}(\Omega) \cap L^\infty(\Omega), \tag{7.5.5}$$

for a suitable exponent $r \geq 1$, then it is a straightforward fact to see that $\widehat{\phi}(\cdot, t) \in W^{1,r}(\Omega)$, i.e. the spatial Sobolev regularity of $\widehat{\phi}$ is equivalent to that of ϕ, once f_0 and f_1 are Lipschitz. In the

[1] We will precise the smoothness assumptions on $\partial\Omega$ in the next chapters.

same fashion, for the condition on the divergence, one may see that we have

$$\begin{aligned}\operatorname{div}_x\widehat{\phi} &= \frac{\operatorname{div}\phi}{(1-t)f_0+tf_1} - \frac{(1-t)\langle\phi,\nabla f_0\rangle + t\langle\phi,\nabla f_1\rangle}{((1-t)f_0+tf_1)^2}\\ &= \frac{f_0-f_1}{(1-t)f_0+tf_1} - \frac{(1-t)\langle\phi,\nabla f_0\rangle + t\langle\phi,\nabla f_1\rangle}{((1-t)f_0+tf_1)^2},\end{aligned}$$

so that Lipschitz regularity of f_0 and f_1, lower bounds on f_0 and f_1 and L^∞ on ϕ (again, we use (7.5.5)) easily implies that $\widehat{\phi}$ has a L^∞ spatial divergence, with L^∞ norm integrable in time. Moreover observe that condition (i) guarantees that $\mu_t = (1-t)\rho_0 + t\rho_1$ is a distributional solution in $L^\infty([0,1];L^\infty(\Omega))$ of the continuity equation

$$\partial_t\mu_t + \operatorname{div}_x(\widehat{\phi}_t\,\mu_t) = 0,$$

and thus it is also a renormalized solution, thanks to Theorem 7.5.2. But then μ_t must coincide with $(X(\cdot,t))_\sharp\rho_0$, again thanks to the well-posedness result given by Theorem 7.5.2: this in particular tells us that

$$\mu_1 = \rho_1 = (X(\cdot,1))_\#\rho_0,$$

so that Q defined by (7.5.4) is admissible for $(\mathcal{W})$, that is $(e_i)_\# Q = \rho_i$ for $i=0,1$, and with the same computations performed in Theorem 7.4.2, we obtain

$$i_Q(x) = |\phi(x)|, \quad \text{for } \mathscr{L}^N\text{-a.e. } x\in\Omega,$$

which concludes the proof. □

As one can see, the proof of the previous crucial result is not really complete: what is still missing is the regularity result (7.5.5) for the optimizer of $(\mathcal{B})$. This mission will be accomplished in Chapter 8 (Sobolev regularity) and Chapter 9 (L^∞ estimate). As already said, we will achieve these results thanks to the primal-dual optimality condition given by Theorem 6.4.2, which ensures that

$$\phi = (|\nabla u|-1)_+^{q-1}\frac{\nabla u}{|\nabla u|},$$

where the Beckmann potential $u\in W^{1,q}(\Omega)$ solves the degenerate elliptic equation

$$\operatorname{div}\left((|\nabla u|-1)_+^{q-1}\frac{\nabla u}{|\nabla u|}\right) = f_0 - f_1, \tag{7.5.6}$$

under homogeneous Neumann boundary conditions or, which is the same, is a minimum point of the functional

$$\mathfrak{F}(\varphi) = \frac{1}{q}\int_\Omega (|\nabla\varphi(x)|-1)_+^q\,dx - \int_\Omega (f_0(x)-f_1(x))\,\varphi(x)\,dx,\ \ \varphi\in W^{1,q}_\diamond(\Omega).$$

Remark 7.5.6. Observe that in general no more than $C^{0,1}$ regularity should be expected for solutions of equation (7.5.6), as far as every 1−Lipschitz function is a solution of the homogeneous equation. Moreover regularity results for equation (7.5.6) (or equations exhibiting a similar type of degeneracy) are not at all trivial and they will be of their own interest, apart from its connection with Wardrop equilibria.

CHAPTER 8

Sobolev regularity for the solution of Beckmann's problem

1. Introduction

This chapter is devoted to prove that the solution ϕ of

$$(\mathcal{B}) = \min_{\phi \in L^p(\Omega;\mathbb{R}^N)} \left\{ \int_\Omega \mathcal{H}(\phi(x))\,dx \,:\, \operatorname{div}\phi = \rho_0 - \rho_1,\ \langle \phi, \nu \rangle = 0 \text{ on } \partial\Omega \right\},$$

is a Sobolev vector field, where again

$$\mathcal{H}(z) = \frac{1}{p}|z|^p + |z|,\ z \in \mathbb{R}^N.$$

The hypotheses on ρ_0 and ρ_1 (or more correctly, on their densities) will be slightly weaker than that assumed in Theorem 7.5.5: anyway, we recall that in the proof of that result the Lipschitz requirement on ρ_0 and ρ_1 can not be dropped, because we need the divergence of $\widehat{\phi}$ defined by (7.2.2) to be an L^∞ function. The result will be achieved using the fact that $\phi = \nabla\mathcal{H}^*(\nabla u)$, where u solves

$$\operatorname{div}\nabla\mathcal{H}^*(\nabla u) = \rho_0 - \rho_1,$$

and then proving that a certain non-linear function of the gradient ∇u (the function $\mathcal{V}$, see below) is in $W^{1,2}(\Omega)$. Observe that, due to our assumptions on $\mathcal{H}$, it is not possible to prove Sobolev regularity of ϕ just passing from that of ∇u (as in the case of Example 7.3.1): indeed, when dealing with degenerate elliptic equations, as in this case, it is quite a delicate matter the question of whether or not ∇u is in a Sobolev space, even for very regular data ρ_0 and ρ_1.

2. Interior Sobolev estimates

Looking at equation (7.5.6), we first observe

$$\frac{(|\nabla u|-1)_+^{q-1}}{|\nabla u|}|\xi|^2 \le \langle D^2\mathcal{H}^*(\nabla u)\xi, \xi\rangle \le (q-1)(|\nabla u|-1)_+^{q-2}|\xi|^2,\ \xi \in \mathbb{R}^N,$$

that is the ellipticity constants degenerate in the region $\{|\nabla u| \le 1\}$. We will confine our analysis to the non-singular case $q \ge 2$, which is anyway relevant for the applications of the previous Chapter: indeed, recall that in the continuous Wardrop model, it is assumed $p < N/(N-1)$ in order to give proper sense to the optimality conditions and to the related concept of Wardrop equilibrium (see Chapter 6, Theorem 6.3.9).

First of all, we need the following pointwise inequalities. This is the main point where the precise structure of $\mathcal{H}^*$ plays a role.

LEMMA 8.2.1. *For every $q \geq 2$, let us define the following vector field*

$$V(z) := |\nabla \mathcal{H}^*(z)|^{\frac{p}{2}} \frac{z}{|z|} = (|z|-1)_+^{\frac{q}{2}} \frac{z}{|z|}, \ z \in \mathbb{R}^N. \tag{8.2.1}$$

Then for every $z, w \in \mathbb{R}^N$ we get

$$\langle \nabla \mathcal{H}^*(z) - \nabla \mathcal{H}^*(w), z - w \rangle \geq \frac{4}{q^2} \left| V(z) - V(w) \right|^2, \tag{8.2.2}$$

$$|\nabla \mathcal{H}^*(z) - \nabla \mathcal{H}^*(w)| \leq (q-1) \left(|V(z)|^{\frac{q-2}{q}} + |V(w)|^{\frac{q-2}{q}} \right) |V(z) - V(w)|. \tag{8.2.3}$$

PROOF. We first observe that if

$$\max\{|z|, |w|\} \leq 1,$$

then (8.2.2) and (8.2.3) are trivially true. Secondly, in the case

$$\min\{|z|, |w|\} \leq 1,$$

supposing for example that $|w| \leq 1$ and $|z| > 1$, using Cauchy-Schwarz inequality we get

$$\begin{aligned} \langle \nabla \mathcal{H}^*(z), z - w \rangle &= \frac{(|z|-1)_+^{q-1}}{|z|} \langle z, z - w \rangle \\ &\geq (|z|-1)_+^{q-1} |z| - (|z|-1)_+^{q-1} = (|z|-1)_+^q, \end{aligned}$$

which proves (8.2.2), while (8.2.3) is easily seen to be true in this case, too.

Let us now suppose that $|z| > 1$ and $|w| > 1$. Now, we recall the inequality (see Appendix C, Lemma C.1.3)

$$\langle |s|^{q-2}s - |t|^{q-2}t, s - t \rangle \geq \frac{4}{q^2} \left| |s|^{\frac{q-2}{2}} s - |t|^{\frac{q-2}{2}} t \right|^2, \ s, t \in \mathbb{R}^N, \tag{8.2.4}$$

and we see that if we are able to prove the following

$$\left\langle |s|^{q-2}s - |t|^{q-2}, (|s|+1)\frac{s}{|s|} - (|t|+1)\frac{t}{|t|} \right\rangle \geq \langle |s|^{q-2}s - |t|^{q-2}t, s - t \rangle, \tag{8.2.5}$$

then choosing

$$s = (|z|-1)_+ \frac{z}{|z|}, \ t = (|w|-1)_+ \frac{w}{|w|},$$

and using (8.2.5) in combination with (8.2.4), we obtain (8.2.2). So, it is left to prove inequality (8.2.5): one sees that this is equivalent to

$$|s|^{q-1} + |t|^{q-1} - \langle s, t \rangle \left[\frac{|s|^{q-2}}{|t|} + \frac{|t|^{q-2}}{|s|} \right] \geq 0,$$

which is just a simple consequence of Cauchy-Schwarz inequality $\langle s, t \rangle \leq |s||t|$.

In order to prove (8.2.3), it is enough to start from the inequality (see Appendix C, Lemma C.1.4)

$$||s|^{q-2}s - |t|^{q-2}t| \leq (q-1)(|s|^{\frac{q-2}{2}} + |t|^{\frac{q-2}{2}}) \left| |s|^{\frac{q-2}{2}} s - |t|^{\frac{q-2}{2}} t \right|,$$

which is valid for every $t, s \in \mathbb{R}^N$ and then take s and t as before. □

We are ready to prove the main result of this section: the proof is an adaption of an argument used by Bojarski and Iwaniec (see [**20**]) for the $p-$Laplace operator.

THEOREM 8.2.2. *Let $\Omega \subset \mathbb{R}^N$ be an open set and take $f \in W^{1,p}(\Omega)$, with $p = q/(q-1)$. If $u \in W^{1,q}(\Omega)$ is a local weak solution of*

$$-\mathrm{div}\nabla\mathcal{H}^*(\nabla u) = f, \ \text{in } \Omega, \tag{8.2.6}$$

then we get $\mathcal{V} \in W^{1,2}_{\mathrm{loc}}(\Omega)$, where the function $\mathcal{V}$ is defined by

$$\mathcal{V}(x) := V(\nabla u(x)) = (|\nabla u(x)| - 1)_+^{\frac{q}{2}} \frac{\nabla u(x)}{|\nabla u(x)|}, \ x \in \Omega. \tag{8.2.7}$$

More precisely, for every $\Sigma \Subset \Omega$ there exists a constant $C = C(N, q)$ such that

$$\|\nabla\mathcal{V}(x)\|^2_{L^2(\Sigma)} \leq \frac{C}{\mathrm{dist}\,(\Sigma, \partial\Omega)^2} \|\nabla u\|^q_{L^q(\Omega)} + C\,\|\nabla f\|^p_{L^p(\Omega)}.$$

PROOF. We fix two subsets compactly contained in Ω, that is $\Sigma \Subset \Sigma_0 \Subset \Omega$ and such that $0 < h_0 = \mathrm{dist}\,(\Sigma, \partial\Omega) = 2\,\mathrm{dist}\,(\Sigma_0, \partial\Omega)$: we aim to prove that $\mathcal{V} \in W^{1,2}(\Sigma)$, using integrated difference quotients. First of all, we observe that u local weak solution of (8.2.6) means

$$\int_\Omega \langle\nabla\mathcal{H}^*(\nabla u(x)), \nabla\varphi(x)\rangle\,dx = \int_\Omega f(x)\,\varphi(x)\,dx, \quad \text{for every } \varphi \in W^{1,q}_0(\Omega).$$

In particular, for every h such that $|h| < h_0/2$, taking a $\varphi \in W^{1,q}_0(\Sigma_0)$, we get that

$$\int_\Omega \langle\nabla\mathcal{H}^*(\nabla u(x + h\omega)), \nabla\varphi(x)\rangle\,dx = \int_\Omega f(x + h\omega)\varphi(x)\,dx,$$

for any direction $\omega \in \mathbb{S}^{N-1}$. Hence subtracting and dividing by h, we obtain

$$\int_\Omega \langle\delta_{h,\omega}\nabla\mathcal{H}^*(\nabla u), \nabla\varphi\rangle\,dx = \int_\Omega \delta_{h,\omega} f\,\varphi\,dx, \tag{8.2.8}$$

for every $\varphi \in W^{1,q}_0(\Sigma_0)$, where we have used the notation

$$\delta_{h,\omega} g(x) := \frac{g(x + h\omega) - g(x)}{h}.$$

We now want to exploit (8.2.8) for a suitable choice of the test function φ, in order to obtain $W^{1,2}$ estimates on $\mathcal{V}$. At this end, let us take a smooth cut-off function $\zeta \in C^\infty_c(\Sigma_0)$, such that:

(i) $0 \leq \zeta \leq 1$;
(ii) $\zeta \equiv 1$ on Σ;
(iii) $\|\nabla\zeta\|_\infty \leq C(\mathrm{dist}\,(\Sigma, \partial\Omega))^{-1}$.

Then we make the following choice for the test function φ

$$\varphi(x) = \zeta^2(x)\,\delta_{h,\omega} u(x), \ x \in \Omega,$$

for every pair $(h,\omega)\in\mathbb{R}\times\mathbb{S}^{N-1}$ such that $|h|<h_0/2$: observe that with this choice, this is an admissible test function in (8.2.8). We now develop φ and use Cauchy-Schwarz inequality, getting

$$\int_\Omega \langle \delta_{h,\omega}\nabla\mathcal{H}^*(\nabla u),\delta_{h,\omega}\nabla u\rangle\,\zeta^2\,dx \le 2\int_\Omega |\delta_{h,\omega}\nabla\mathcal{H}^*(\nabla u)|\;\zeta\,|\nabla\zeta|\,|\delta_{h,\omega}u|\,dx + \int_\Omega \zeta^2\,|\delta_{h,\omega}f|\,|\delta_{h,\omega}u|\,dx.$$

An application of the pointwise inequalities (8.2.2) and (8.2.3) yields

$$\int_\Omega |\delta_{h,\omega}\mathcal{V}|^2\zeta^2\;dx \le C\int_\Omega \left(|\mathcal{V}_{h,\omega}|^{\frac{q-2}{q}}+|\mathcal{V}|^{\frac{q-2}{q}}\right)\,|\delta_{h,\omega}\mathcal{V}|\,\zeta\,|\nabla\zeta||\delta_{h,\omega}u|\,dx + \left(\int_\Omega \zeta^p|\delta_{h,\omega}f|^p\,dx\right)^{\frac{1}{p}}\left(\int_\Omega \zeta^q|\delta_{h,\omega}u|^q\,dx\right)^{\frac{1}{q}}$$

where the constant C depends on q only. By means of Young's inequality, we get for every $\varepsilon>0$

$$\left(|\mathcal{V}_{h,\omega}|^{\frac{q-2}{q}}+|\mathcal{V}|^{\frac{q-2}{q}}\right)\,|\delta_{h,\omega}\mathcal{V}|\,\zeta\,|\nabla\zeta||\delta_{h,\omega}u| \le \varepsilon\,|\delta_{h,\omega}\mathcal{V}|^2\,\zeta^2 + \frac{1}{\varepsilon}\left(|\mathcal{V}_{h,\omega}|^{\frac{q-2}{q}}+|\mathcal{V}|^{\frac{q-2}{q}}\right)^2\,|\nabla\zeta|^2|\delta_{h,\omega}u|^2,$$

so that choosing ε small enough, the term on the right-hand side containing $\delta_{h,\omega}\mathcal{V}$ can be absorbed by the term on the left hand-side. Up to now, we have shown

$$\int_\Omega |\delta_{h,\omega}\mathcal{V}|^2\zeta^2\,dx \le C\int_\Omega \left(|\mathcal{V}_{h,\omega}|^{\frac{q-2}{q}}+|\mathcal{V}|^{\frac{q-2}{q}}\right)^2\,|\nabla\zeta|^2|\delta_{h,\omega}u|^2\,dx + \frac{1}{p}\int_\Omega \zeta^p|\delta_{h,\omega}f|^p\,dx+\frac{1}{q}\int_\Omega \zeta^q|\delta_{h,\omega}u|^q\,dx.$$

A simple application of Hölder's inequality to the first term on the right side, yields

$$\int_\Omega |\delta_{h,\omega}\mathcal{V}|^2\,\zeta^2\,dx \le C\left(\int_{\Sigma_0}\left(|\mathcal{V}_{h,\omega}|^{\frac{q-2}{q}}+|\mathcal{V}|^{\frac{q-2}{q}}\right)^{\frac{2q}{q-2}}\;dx\right)^{\frac{q-2}{q}}\left(\int_\Omega |\nabla\zeta|^q|\delta_{h,\omega}u|^q\,dx\right)^{\frac{2}{q}} + \frac{1}{p}\int_\Omega \zeta^p|\delta_{h,\omega}f|^p\,dx+\frac{1}{q}\int_\Omega \zeta^q|\delta_{h,\omega}u|^q\,dx. \tag{8.2.9}$$

It is now sufficient to observe that

$$\left(\int_{\Sigma_0}\left(|\mathcal{V}|^{\frac{q-2}{q}}+|\mathcal{V}_{h,\omega}|^{\frac{q-2}{q}}\right)^{\frac{2q}{q-2}}\;dx\right)^{\frac{q-2}{q}} \le 2\left(\int_\Omega |\mathcal{V}|^2\,dx\right)^{\frac{q-2}{q}},$$

so that inserting the latter in (8.2.9), we easily get

$$\int_\Omega |\mathcal{V}_h-\mathcal{V}|^2\zeta^2\,dx \le \frac{C}{\operatorname{dist}(\Sigma,\partial\Omega)^2}\left(\int_\Omega |\mathcal{V}|^2\,dx\right)^{\frac{q-2}{q}}\left(\int_{\Sigma_0}|\delta_{h,\omega}u|^q\,dx\right)^{\frac{2}{q}} + \frac{1}{p}\int_\Omega \zeta^p|\delta_{h,\omega}f|^p\,dx+\frac{1}{q}\int_\Omega \zeta^q|\delta_{h,\omega}u|^q\,dx.$$

Finally, we just observe that, by means of the characterization of Sobolev spaces in terms of integrated difference quotients (see [**28**]), we have

$$\int_{\Sigma_0} |\delta_{h,\omega} u|^q \, dx \leq C_N \int_\Omega |\nabla u|^q \, dx,$$

and

$$\int_{\Sigma_0} |\delta_{h,\omega} f|^p \, dx \leq C_N \int_\Omega |\nabla f|^p \, dx,$$

and moreover by the very definition of $\mathcal{V}$ we have

$$\int_\Omega |\mathcal{V}|^2 \, dx \leq \int_\Omega |\nabla u|^q \, dx,$$

so that in the end we get

$$\int_\Sigma |\delta_{h,\omega} \mathcal{V}|^2 \, dx \leq \frac{C}{\operatorname{dist}(\Sigma, \Omega)^2} \int_\Omega |\nabla u|^q \, dx + C \int_\Omega |\nabla f|^p \, dx,$$

that is $\mathcal{V}$ has a square-integrable weak derivative along the direction given by $\omega \in \mathbb{S}^{N-1}$. □

REMARK 8.2.3. We observe that as an easy consequence of Theorem 8.2.2 and Sobolev Imbedding Theorems, we get a gain of integrability for ∇u: indeed, in the case $N > 2$, we get $\mathcal{V} \in L^{2^*}_{\text{loc}}(\Omega)$ and then

$$\int_\Omega (|\nabla u(x)| - 1)_+^{\frac{qN}{N-2}} \, dx = \int_\Omega |\mathcal{V}(x)|^{\frac{2N}{N-2}} \, dx < +\infty,$$

which ensures that

$$\nabla u \in L^{q \frac{N}{N-2}}_{\text{loc}}(\Omega), \tag{8.2.10}$$

while if $N = 2$ we get that $\mathcal{V}$ (and so $|\nabla u|$) is in every L^s_{loc}, with $s < \infty$. Moreover in the case $q > N - 2$, then we can assure that $u \in C^{0,\alpha}_{\text{loc}}(\Omega)$, with $\alpha = 1 - (N-2)/q$.

Going back to our vector field $\phi = \mathcal{H}^*(\nabla u)$, Theorem 8.2.2 easily implies the following.

COROLLARY 8.2.4. *Under the assumptions of Theorem 8.2.2, we get*

$$\phi = \nabla \mathcal{H}^*(\nabla u) = |\mathcal{V}|^{\frac{q-2}{q}} \mathcal{V} \in W^{1,r}_{\text{loc}}(\Omega), \tag{8.2.11}$$

for suitable exponents $r = r(N, q)$ given by

$$r(N,q) = \begin{cases} 2, & \textit{if } N = q = 2, \\ \textit{any value } < 2, & \textit{if } N = 2, \ q > 2, \\ \frac{Nq}{(N-1)q + 2 - N}, & \textit{if } N > 2. \end{cases}$$

PROOF. The case $q = 2$ is clearly trivial, in fact in this case $\phi = \mathcal{V} \in W^{1,2}_{\text{loc}}(\Omega)$.

Let us begin with the case $N > 2$: using inequality (8.2.3) with $z = \nabla u(x + h\,\omega)$ and $w = \nabla u(x)$, we get

$$|\delta_{h,\omega} \phi| \leq (q-1) \left(|\mathcal{V}_h(x)|^{\frac{q-2}{q}} + |\mathcal{V}(x)|^{\frac{q-2}{q}} \right) |\delta_{h,\omega} \mathcal{V}|.$$

and we already observed that $\mathcal{V} \in L^{2^*}_{\rm loc}(\Omega)$, so that $|\mathcal{V}|^{\frac{q-2}{q}} \in L^{\frac{2^* q}{(q-2)}}_{\rm loc}(\Omega)$ and the right hand side in the previous inequality belongs to $L^r_{\rm loc}(\Omega)$, with r being given by the relation

$$\frac{1}{r} = \frac{(q-2)}{2^* q} + \frac{1}{2}.$$

This clearly implies that we can control the integrated difference quotients

$$\int |\delta_{h,\omega}\phi(x)|^r \, dx,$$

thus proving the assertion.

Finally, when $N = 2$ and $q > 2$, we can proceed as before, taking into account the fact that $\mathcal{V} \in L^s_{\rm loc}(\Omega)$ for every $s < \infty$.

Notice also that, should $\mathcal{V}$ be bounded, one would automatically get $\phi \in W^{1,2}_{\rm loc}(\Omega)$. □

Remark 8.2.5. The same arguments in the proof of Theorem 8.2.2 may obviously be applied to the case of uniformly elliptic equations, such as $-\mathrm{div}(\nabla\mathcal{K}(\nabla u)) = f$ with

$$\lambda|\xi|^2 \le \langle D^2\mathcal{K}(z)\xi, \xi\rangle \le \Lambda|\xi|^2.$$

In this case they provide local $W^{2,2}$ regularity estimates for the solution u, under the assumption that $f \in W^{1,2}$. It is worth remarking that, as is well-known, these estimates are still true under the sole assumption that $f \in L^2$ (see [**55**, Theorem 8.8]): this somehow could suggest that actually we asked for a regularity assumption on f, stronger than what is really needed. As one can easily guess, this is intimately linked to the degeneracy of our operator $\nabla\mathcal{H}^*$. Actually, in non-degenerate equations, when we arrive to the term

$$\int \delta_{h,\omega} f \, \delta_{h,\omega} u \, dx,$$

we can pass all the increments on the function u, that is we can use the trick

$$\int \delta_{h,\omega} f \, \delta_{h,\omega} u \, dx = -\int \left(\int_0^1 f(x + th\omega)\, dt\right) \delta_{h,\omega} \nabla u \, dx,$$

thus getting something that may be estimated again by the L^2 norm of $\delta_{h,\omega}\nabla u$ (but to the power of one, while at the left hand side it is to the power of two). Yet, here this is no more useful, since $\mathcal{V}$ is not invertible as a function of ∇u: this is why we asked for a higher regularity on f.

It is worthwhile noticing that even in the case of the $q-$Laplacian, where the corresponding quantity $\mathcal{V}$ is given by

$$|\nabla u|^{\frac{q-2}{2}} \nabla u,$$

the difference quotients technique is known to work only with a Sobolev assumption on f and a result of the type

$$f \in L^s \Longrightarrow |\nabla u|^{\frac{q-2}{2}} \nabla u \in W^{1,2}(\Omega),$$

for a suitable exponent $s \ge p = q'$ should not be expected (see the discussion at the end of Appendix C). We end up recalling that even with a very regular datum f, in the case of the $q-$Laplacian with $q > 2$, the weak differentiability of ∇u can be guaranteed only in a fractional sense (see [**69**]):

a simple proof of this fact uses $\mathcal{V} \in W^{1,2}$ (which is true when $f \in W^{1,p}$) and the elementary inequality (see the Appendix C, Lemma C.1.1)

$$|\nabla u(x) - \nabla u(y)| \le C \left| |\nabla u(x)|^{\frac{q-2}{2}} \nabla u(x) - |\nabla u(y)|^{\frac{q-2}{2}} \nabla u(y) \right|^{\frac{2}{q}},$$

so that $\nabla u \in W^{\frac{2}{q}-\varepsilon,q}$, for every $\varepsilon > 0$. Again, such a kind of result fails in the case of our operator $\nabla\mathcal{H}^*$.

3. Sobolev estimates up to the boundary

Under suitable assumptions on the domain Ω and on the boundary datum, Theorem 8.2.2 can be enforced, thus obtaining a global $W^{1,2}$ estimate. This is exactly the content of the next result: observe that we also ask for a higher summability of f. The reason for this further requirement will appear from the proof.

THEOREM 8.3.1. *Let us suppose that Ω is a $C^{3,1}$ domain and take $f \in W^{1,p}_\diamond(\Omega) \cap L^{N+\alpha}(\Omega)$, with $p = q/(q-1)$ and $\alpha > 0$. If $u \in W^{1,q}_\diamond(\Omega)$ is a weak solution of the following Neumann boundary problem*

$$\left\{ \begin{array}{rcll} -\mathrm{div}\,(\nabla\mathcal{H}^*(\nabla u)) & = & f, & \text{in } \Omega, \\ \langle \nabla\mathcal{H}^*(\nabla u), \nu\rangle & = & 0, & \text{on } \partial\Omega, \end{array} \right. \tag{8.3.1}$$

then we get $\mathcal{V} \in W^{1,2}(\Omega)$, where

$$\mathcal{V}(x) = (|\nabla u(x)| - 1)_+^{\frac{q}{2}} \frac{\nabla u(x)}{|\nabla u(x)|}, \ x \in \Omega.$$

PROOF. Let us fix a boundary point x_0 and a neighborhood of it $\mathcal{U}$. We set

$$B^+ = \{x = (x', x_N) \,:\, |x| < 1, \ x_N \ge 0\},$$

and observe that thanks to our assumptions on Ω, we can guarantee the existence of a diffeomorphism ψ, sending B^+ on $\mathcal{U}^+ = \mathcal{U} \cap \Omega$ and the flat part of ∂B^+ on $\mathcal{U} \cap \partial\Omega$.

Let us set $B^- = \mathcal{R}B^+$, $\mathcal{R}$ being the reflection with respect to the hyperplane $\{x_N = 0\}$, then we define

$$\widehat{u}(y) = \left\{ \begin{array}{ll} u(\psi(y)), & \text{if } y \in B^+, \\ u(\psi(\mathcal{R}y)), & \text{if } y \in B^-, \end{array} \right.$$

$$\widehat{f}(x) = \left\{ \begin{array}{ll} f(\psi(y))|\det D\psi(y)|, & \text{if } y \in B^+, \\ f(\psi(\mathcal{R}y))|\det D\psi(\mathcal{R}y)|, & \text{if } y \in B^-, \end{array} \right.$$

and observe that $\widehat{u} \in W^{1,q}(B)$ and $\widehat{f} \in W^{1,p}(B) \cap L^{N+\alpha}(B)$, with $B := B^+ \cup B^-$. Moreover we will use the fact that[1] $u \in W^{1,\infty}(\Omega)$ (see next section, Theorem 9.2.1), so that the same is true for $\widehat{u}$, that is

$$\widehat{u} \in W^{1,\infty}(B). \tag{8.3.2}$$

[1]Here we use the higher summability requirement on f, as far as Theorem 9.2.1 is proven under the assumption that $f \in L^{N+\alpha}$. It is clear that $f \in W^{1,p}(\Omega)$ does not imply this kind of summability in general, and especially in our case, where $p \le 2$.

We now want to find the equation satisfied by $\widehat{u}$ in the unit ball: using the change of variables $x=\psi(y)$ and the fact that u satisfies

$$\int_\Omega \langle \nabla\mathcal{H}^*(\nabla u(x)), \nabla\varphi(x)\rangle\, dx = \int_\Omega f(x)\varphi(x)\, dx, \quad \text{for every } \varphi\in W^{1,p}(\Omega),$$

it is easy to see that $\widehat{u}$ satisfies

$$\int_{B^+} \langle a(y,\nabla\widehat{u}(y)), \nabla\varphi(y)\rangle\, dy = \int_{B^+} \widehat{f}(y)\varphi(y)\, dy, \quad \text{for every } \varphi\in W^{1,p}_0(B),$$

where the function a is defined by

$$a(y,p) = |\det D\psi(y)|\, \nabla\mathcal{H}^*(p\,[D\psi(y)]^{-1})\left([D\psi(y)]^{-1}\right)^t,\ (y,p)\in B^+\times\mathbb{R}^N.$$

On the other hand, using the change of variables $x=\overline{\psi}(y):=\psi(\mathcal{R}y)$ we obtain

$$\int_{B^-} \langle \overline{a}(\nabla\widehat{u}(y)), \nabla\varphi(y)\rangle\, dy = \int_{B^-} \widehat{f}(y)\varphi(y)\, dy, \quad \text{for every } \varphi\in W^{1,p}_0(B),$$

with $\overline{a}$ given by

$$\overline{a}(y,p) = |\det D\overline{\psi}(y)|\, \nabla\mathcal{H}^*(p\,[D\overline{\psi}(y)]^{-1})\left([D\overline{\psi}(y)]^{-1}\right)^t,\ (y,p)\in B^-\times\mathbb{R}^N.$$

Observe that we have

$$|\det D\overline{\psi}(y)| = |\det D\psi(\mathcal{R}y)| \quad\text{and}\quad [D\overline{\psi}(y)]^{-1} = \mathcal{R}[D\psi(\mathcal{R}y)]^{-1},$$

then let us assume for a moment the existence of $\mathcal{O}\in C^{1,1}(B^+;\mathbb{R}^{N\times N})$ such that for every $y\in B^+$, $\mathcal{O}(y)$ is an orthogonal matrix verifying

$$[D\psi(y',0)]^{-1} = \mathcal{R}[D\psi(y',0)]^{-1}\mathcal{O}(y',0),\ y'\in\partial B^+\cap\partial B^-. \tag{8.3.3}$$

Setting for simplicity

$$\widehat{M}(y) = \begin{cases} [D\psi(y)]^{-1}, & \text{if } y\in B^+\\ \mathcal{R}[D\psi(\mathcal{R}y)]^{-1}\mathcal{O}(\mathcal{R}y), & \text{if } y\in B^-\end{cases}$$

the previous discussion, together with assumption (8.3.3) and the fact that

$$\nabla\mathcal{H}^*(z\,\mathcal{O}) = \nabla\mathcal{H}^*(z)\,\mathcal{O},$$

tells us that $\widehat{u}$ is a weak solution in B of the equation

$$-\mathrm{div}\, A(y,\nabla\widehat{u}) = \widehat{f}, \tag{8.3.4}$$

where the operator A is defined by

$$A(y,p) = |\det \widehat{M}(y)|^{-1}\,\nabla\mathcal{H}^*(p\,\widehat{M}(y))\widehat{M}(y)^t,\ (y,p)\in B\times\mathbb{R}^N.$$

Observe that having assumed (8.3.3), is crucial to obtain that $A(\cdot,p)$ is continuous across the hyperplane $\{x_N=0\}$, which in turn implies that $A(\cdot,p)$ is Lipschitz.

So let us verify the existence of such a matrix field $\mathcal{O}$: by polar decomposition, we know that $OU=D\psi$, with O orthogonal and U symmetric and positive definite. This implies that $O=D\psi\,U^{-1}$ and $[D\psi]^tD\psi=U^2$, that is

$$\mathcal{O} = D\psi([D\psi]^tD\psi)^{-\frac{1}{2}}, \tag{8.3.5}$$

which is our candidate for the $C^{1,1}$ matrix field. We now have to make an explicit choice for the diffeomorphism ψ, in order to obtain that this $\mathcal{O}$ further verifies (8.3.3): we can suppose, up to a translation, that $x_0 = 0$ and moreover that, up to a rotation, the set $\mathcal{U} \cap \partial\Omega$ can be represented as the graph of a $g \in C^{3,1}$ defined on $\{|y| \le 1 \,:\, y_N = 0\}$, that is $\mathcal{U} \cap \partial\Omega = \{(y', g(y')) \,:\, |y'| \le 1\}$. Then we see that taking ψ of the following form

$$\psi(y', y_N) = (y', g(y')) - y_N(\nabla g(y'), -1),$$

we get that ψ is diffeomorphism between B^+ and $\mathcal{U}^+$ (up to redefine the neighborhood $\mathcal{U}$, without changing its intersection with the boundary $\mathcal{U} \cap \partial\Omega$). Moreover we have the following expression for the Jacobian matrix

$$D\psi(y) = \begin{bmatrix} \mathrm{Id}_{N-1} - y_N\, D^2 g(y') & \nabla g(y') \\ & \\ -(\nabla g(y'))^t & 1 \end{bmatrix}$$

where Id_{N-1} stands for the $(N-1)\times(N-1)$ identity matrix. Then it is easily seen that with the choice (8.3.5), property (8.3.3) is equivalent to require that

$$[D\psi(y', 0)]^t D\psi(y', 0) = (D\psi(y', 0)\mathcal{R})^2,$$

and this is a straightforward consequence of the structure of $D\psi$. We point out that despite the hypothesis on $\partial\Omega$ of being $C^{3,1}$, the diffeomorphism we have provided is only of class $C^{2,1}$ (we use the gradient of g in the definition of ψ): this is due to the fact that we need a diffeomorphism having a Jacobian matrix with a special structure, because of condition (8.3.3). Moreover we stress that the matrix field $\widehat{M}$ is piecewise $C^{1,1}$ and globally no more than $C^{0,1}$.

We now aim to show that

$$\widehat{\mathcal{V}}(y) = V\left(\nabla \widehat{u}(y)\, \widehat{M}(y)\right) \in W^{1,2}_{\mathrm{loc}}(B),$$

where as before $V(z) = (|z|-1)_+^{q/2} z/|z|$. If this is true, then we will clearly have that $\mathcal{V} \in W^{1,2}$ in a neighborhood of x_0, thus concluding the proof. Let us begin with some manipulations: in order to simplify the notations, we set $\mathfrak{b}(y) = |\det \widehat{M}(y)|^{-1}$, then we begin applying (8.2.2), so that as in Theorem 8.2.2 we obtain

$$\left|\delta_{h,\omega}\widehat{\mathcal{V}}\right|^2 \le \langle \delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M}), \delta_{h,\omega}(\nabla\widehat{u}\,\widehat{M})\rangle,$$

where as always $\delta_{h,\omega}$ denotes the incremental ratio in the direction $\omega \in \mathbb{S}^{N-1}$. Then with some algebraic manipulations, the right-hand side can be re-written as

$$\begin{aligned}\langle \delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M}), \delta_{h,\omega}(\nabla\widehat{u}\,\widehat{M})\rangle &= \langle \delta_{h,\omega}\left(\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})\widehat{M}^t\right), \delta_{h,\omega}\nabla\widehat{u}\rangle \\ &\quad - \langle \nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})(\delta_{h,\omega}\widehat{M}^t), \delta_{h,\omega}\nabla\widehat{u}\rangle \\ &\quad - \langle \delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M}), \nabla\widehat{u}\,\delta_{h,\omega}\widehat{M}\rangle,\end{aligned}$$

and multiplying by $\mathfrak{b}_h$ we obtain

$$\begin{aligned}\mathfrak{b}_h|\delta_{h,\omega}\widehat{\mathcal{V}}|^2 &\le \langle \delta_{h,\omega}A, \delta_{h,\omega}\nabla\widehat{u}\rangle - \langle \nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})(\delta_{h,\omega}\mathfrak{b}\,\widehat{M}^t), \delta_{h,\omega}\nabla\widehat{u}\rangle \\ &\quad - \mathfrak{b}_h\langle \delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M}), \nabla\widehat{u}\,\delta_{h,\omega}\widehat{M}\rangle.\end{aligned}$$

We can now as always take a smooth cut-off function ζ supported in some smaller ball $B' \subset B$, multiply the previous inequality by ζ^2 and then integrate over $\operatorname{supp}\zeta$, so that we get

$$\begin{aligned}\int |\delta_{h,\omega}\widehat{\mathcal{V}}|^2\,dy \le& \int |\delta_{h,\omega}A|\,|\nabla\zeta|\,\zeta\,|\delta_{h,\omega}\widehat{u}|\,dy\\ &- \int \langle \nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})\delta_{h,\omega}(\mathfrak{b}\,\widehat{M}), \delta_{h,\omega}\nabla\widehat{u}\rangle\,\zeta^2\,dy\\ &+ \int |\mathfrak{b}_h||\delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})|\,|\nabla\widehat{u}|\,|\delta_{h,\omega}\widehat{M}|\,\zeta^2\,dy + \int |\delta_{h,\omega}\widehat{f}|\,|\delta_{h,\omega}\widehat{u}|\zeta^2\,dy := \sum_{i=1}^4 \mathcal{I}_i,\end{aligned}$$

where we have used the fact that $\widehat{u}$ is a solution of (8.3.4) and $\mathfrak{b} \in L^\infty$ with $\mathfrak{b} \ge c > 0$. For simplicity, we now discuss separately the estimates of each integral:

Estimate for $\mathcal{I}_1$ We would like to use the basic inequality (8.2.3): we first observe that

$$\begin{aligned}\delta_{h,\omega}A(y,\nabla\widehat{u}) =& \Big[\nabla\mathcal{H}^*(\nabla\widehat{u}(y+h\omega)\,\widehat{M}(y+h\omega))\Big]\;\delta_{h,\omega}(\mathfrak{b}\,\widehat{M}^t)\\ &+ \Big[\mathfrak{b}(y)\widehat{M}^t(y)\Big]\;\delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M}),\end{aligned}$$

so that

$$\begin{aligned}\mathcal{I}_1 \le& \int \Big|\nabla\mathcal{H}^*(\nabla\widehat{u}(y+h\omega)\,\widehat{M}(y+h\omega))\Big|\;|\delta_{h,\omega}(\mathfrak{b}\,\widehat{M}^t)|\,|\nabla\zeta|\,\zeta\,|\delta_{h,\omega}\widehat{u}|\,dy\\ &+ \int \Big|\mathfrak{b}\,\widehat{M}^t\Big|\;|\delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})|\,|\nabla\zeta|\,\zeta\,|\delta_{h,\omega}\widehat{u}|\,dy,\end{aligned}$$

and it is easily seen that the first term does not present any problem, thanks to the fact that $\widehat{u} \in W^{1,q}$, $\mathfrak{b}\,\widehat{M}^t \in W^{1,\infty}$ and $\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M}) \in L^p$. On the contrary, the second integral is a kind of term that has already been estimated in the proof of Theorem 8.2.2, once one takes care of the fact that $\mathfrak{b}\,\widehat{M}^t \in W^{1,\infty}$: indeed, in this case it is only left to estimate

$$\int |\delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})|\,|\nabla\zeta|\,\zeta\,|\delta_{h,\omega}\widehat{u}|\,dy,$$

and it is now sufficient to apply (8.2.3), thus the previous integral can be estimated from above by

$$\int_\Omega \left(|\widehat{\mathcal{V}}_{h,\omega}|^{\frac{q-2}{q}} + |\widehat{\mathcal{V}}|^{\frac{q-2}{q}}\right)\,|\delta_{h,\omega}\widehat{\mathcal{V}}|\,\zeta\,|\nabla\zeta|\,|\delta_{h,\omega}\widehat{u}|\,dx.$$

Then one can use the $\varepsilon-$Young's inequality in order to absorb a term of the kind

$$\int_\Omega |\delta_{h,\omega}\widehat{\mathcal{V}}|^2\zeta^2\,dx,$$

and finally one is left with an integral

$$\int_\Omega \left(|\widehat{\mathcal{V}}_{h,\omega}|^{\frac{q-2}{q}} + |\widehat{\mathcal{V}}|^{\frac{q-2}{q}}\right)^2\,|\nabla\zeta|^2|\delta_{h,\omega}u|^2\,dx,$$

which can be easily estimated just as in the proof of Theorem 8.2.2.

Estimate for $\mathcal{I}_2$ This is the most delicate integral: indeed, we have to integrate by parts in order to avoid the difference quotients of ∇u. As a drawback, this will let appear second-order difference quotients of $\mathfrak{b}\,\widehat{M}$, which in principle can not be easily managed, due to the fact that this is only a Lipschitz function: anyway, thanks to our construction, we know that $\mathfrak{b}\,\widehat{M}$ is a function which is $C^{1,1}$ out of $\{x_N = 0\}$ and globally $C^{0,1}$. This implies that its second-order difference quotients $\delta^2_{h,\omega}(\mathfrak{b}\,\widehat{M})$ (see below for their definition) are uniformly bounded if one stays out of a strip of size h around the hyperplane $\{x_N = 0\}$ and are bounded by $C\,h^{-1}$ in that strip (whose measure is of the order of h), which implies $\int|\delta^2_{h,\omega}(\mathfrak{b}\,\widehat{M})| \le C$.

First of all, we separate the two terms $\nabla\widehat{u}(y+h\omega)$ and $\nabla\widehat{u}(y)$ and we perform the change of variable $z = y + h\omega$ in the first integral, thus obtaining

$$\mathcal{I}_2 = \frac{1}{h}\int \langle \nabla\mathcal{H}^*(\nabla\widehat{u}(y-h\omega)\,\widehat{M}(y-h\omega))\delta_{-h,\omega}(\mathfrak{b}\,\widehat{M}), \nabla\widehat{u}(y)\rangle\,\zeta^2(y-h\omega)\,dy$$
$$+\frac{1}{h}\int\langle\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})\delta_{h,\omega}(\mathfrak{b}\,\widehat{M}), \nabla\widehat{u}(y)\rangle\,\zeta^2\,dy,$$

which can be recast into

$$\mathcal{I}_2 = \int\langle\mathcal{H}^*(\nabla\widehat{u}(y-h\omega)\,\widehat{M}(y-h\omega))\,\delta^2_{h,\omega}(\mathfrak{b}\,\widehat{M}), \nabla\widehat{u}\rangle\,\zeta^2(y-h\omega)\,dy$$
$$+\int\langle\mathcal{H}^*(\nabla\widehat{u}(y-h\omega)\,\widehat{M}(y-h\omega))\,\delta_{h,\omega}(\mathfrak{b}\,\widehat{M}), \nabla\widehat{u}\rangle\,\delta_{-h,\omega}\zeta^2\,dy$$
$$+\int\langle\delta_{-h,\omega}\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M}), \nabla\widehat{u}\rangle\,\zeta^2(y)\,dy,$$

where

$$\delta^2_{h,\omega}(\mathfrak{b}\,\widehat{M}) = \frac{\mathfrak{b}(y+h\omega)\,M(y+h\omega) + \mathfrak{b}(y-h\omega)\,M(y-h\omega) - 2\mathfrak{b}(y)\,M(y)}{h^2}.$$

We now observe that the last two terms can be easily estimated as already seen (see the discussion for $\mathcal{I}_1$ and $\mathcal{I}_3$), while the first contains second-order difference quotients of $\mathfrak{b}\,\widehat{M}$: since all the other factors in the integral are bounded (because $\nabla\widehat{u}$ is bounded), this integral may be estimated with $C\int|\delta^2_{h,\omega}(\mathfrak{d}\,\widehat{M})|\,dx$, and this integral is indeed bounded.

Estimate for $\mathcal{I}_3$ Using the fact that $\widehat{M}$ is Lipschitz, together with estimate (8.2.3), yields

$$I_4 \le C\,\mathrm{lip}(\widehat{M})\int\left|\delta_{h,\omega}\nabla\mathcal{H}^*(\nabla\widehat{u}\,\widehat{M})\right||\nabla\widehat{u}|\,\zeta^2\,dy \le \int\left|\delta_{h,\omega}\widehat{\mathcal{V}}\right|\left(|\widehat{\mathcal{V}}|^{\frac{q-2}{q}} + |\widehat{\mathcal{V}}|^{\frac{q-2}{q}}\right)|\nabla\widehat{u}|\,\zeta^2\,dy,$$

and this can be estimated as before (see the estimation for $\mathcal{I}_1$), absorbing the difference quotients of $\widehat{\mathcal{V}}$ in the left-hand side, thanks to Young's inequality.

Estimate for I_4 This is clearly the easy part: it is sufficient to use Hölder's inequality and the fact that $\widehat{f} \in W^{1,p}$ and $\widehat{u} \in W^{1,q}$. □

Finally, we get the desired global Sobolev estimate on the optimizer ϕ: the proof is a straightforward extension of that of Corollary 8.2.4.

COROLLARY 8.3.2. *Under the assumptions of Theorem 8.3.1, the conclusions of Corollary 8.2.4 are global.*

CHAPTER 9

L^∞ gradient estimates for Beckmann potentials

1. Introduction

To complete our program, we will show in this chapter that the optimizer ϕ of

$$(\mathcal{B}) = \min_{\phi \in L^p(\Omega;\mathbb{R}^N)} \left\{ \int_\Omega \mathcal{H}(\phi(x))\, dx \,:\, \operatorname{div}\phi = \rho_0 - \rho_1,\ \langle \phi, \nu \rangle = 0 \text{ on } \partial\Omega \right\},$$

is actually an element of $L^\infty(\Omega;\mathbb{R}^N)$ ($\mathcal{H}$ is still the same as in (7.1.1)). We recall that, for the scopes of this work, the main application of this result is in the proof of Theorem 7.5.5, where a Lipschitz assumption on ρ_0 and ρ_1 is needed: anyway, in this chapter we will derive this L^∞ estimate under fairly more general assumptions on ρ_0 and ρ_1. Moreover, the result will be true without any restriction on the exponent p.

Again, this result will be achieved by looking at the equation solved by a corresponding Beckmann potential u, that is

$$\operatorname{div} \nabla\mathcal{H}^*(\nabla u) = \rho_0 - \rho_1,$$

with homogeneous Neumann boundary conditions and proving that $\nabla u \in L^\infty$: then, the optimality condition $\phi = \nabla\mathcal{H}^*(\nabla u)$ will do the job.

Before going into the details of this further regularity result, let us spend some words on the method of proof: the desired L^∞ estimate on ∇u is achieved by means of considering more regular (we would say *more elliptic* or *more convex*, depending on the point of view) approximating problems, for which one can provide robust a priori estimates, which in the end depend only on the behaviour at infinity of the operator $\nabla\mathcal{H}^*$. This strategy is nowadays classical: in order to apply it, we have benefited from a careful reading of the seminal papers [**45**] and [**63**] by Di Benedetto and Lewis, respectively. In particular, we underline that the variational nature of the problem seems to play a crucial role in our proof, like in [**63**], while this was unnecessary in [**45**]: this is due to the fact that the convergence we obtain on the solutions u_ε of the approximating problems is not strong enough to permit to work only with the weak formulations of the equations, while in [**45**] this is possible thanks to uniform $C^{1,\alpha}$ estimates, from which one can infer uniform convergence (on compact sets) of u_ε and of their gradients. This is clearly linked to the type of degeneracy of our problem, which allows only for $C^{0,1}$ estimate.

2. Main result, strategy and basic tools

As always, with $\mathcal{H}^*$ we indicate

$$\mathcal{H}^*(z) = 1/q\,(|z|-1)_+^q,\ z \in \mathbb{R}^N,$$

where $1 < q < \infty$ and $(\,\cdot\,)_+$ stands for the positive part.

In this chapter we will finally prove the following crucial result.

THEOREM 9.2.1. *Let $\Omega \subset \mathbb{R}^N$ be a $C^{2,1}$ domain. Given $f \in L^{N+\alpha}$ with zero-mean, every solution $u \in W^{1,q}_\diamond(\Omega)$ of the following Neumann boundary value problem*

$$\begin{cases} -\mathrm{div}\,(\nabla \mathcal{H}^*(\nabla u)) &=& f, & \text{in } \Omega, \\ \langle \nabla \mathcal{H}^*(\nabla u), \nu \rangle &=& 0, & \text{on } \partial\Omega, \end{cases} \tag{9.2.1}$$

is a Lipschitz function.

We regard (9.2.1) as the Euler-Lagrange equation of the original optimization problem

$$\min_{\varphi \in W^{1,q}_\diamond(\Omega)} \int_\Omega \mathcal{H}^*(\nabla \varphi)\, dx - \int_\Omega f\, \varphi\, dx,$$

and then we approximate the latter with a more regular one, depending on a small parameter $\varepsilon \in (0, \varepsilon_0]$, i.e.

$$\min_{\varphi \in W^{1,q}_\diamond(\Omega)} \int_\Omega \mathcal{H}^*_\varepsilon(\nabla \varphi)\, dx - \int_\Omega f_\varepsilon\, \varphi\, dx,$$

possessing a unique solution u_ε such that $u_\varepsilon \rightharpoonup u$ in $W^{1,q}(\Omega)$, with u solution of the original problem. In particular we will suppose that

$$\|u_\varepsilon\|_{W^{1,q}(\Omega)} \le C, \quad \text{for every } 0 < \varepsilon \le \varepsilon_0, \tag{9.2.2}$$

then we aim to prove that the solutions u_ε satisfy uniform L^∞ gradient estimates, independent of ε, which consequently will pass to the limit, showing the required regularity on the original solution u.

REMARK 9.2.2. As far as the original problem is convex but not strictly convex, in general we can not expect any kind of uniqueness for the minimizers of

$$\int_\Omega \mathcal{H}^*(\nabla \varphi)\, dx - \int_\Omega f\, \varphi\, dx.$$

Anyway, for our scope it is important to stress the fact that, given two distinct minimizers u_1, u_2 over $W^{1,q}_\diamond(\Omega)$, we have

$$\nabla \mathcal{H}^*(\nabla u_1(x)) = \nabla \mathcal{H}^*(\nabla u_2(x)), \quad \text{for } \mathscr{L}^N\text{–a.e. } x \in \Omega,$$

which is a consequence of the primal-dual optimality condition $\phi = \nabla \mathcal{H}^*(\nabla u)$ and of the uniqueness for the solution of Beckmann's problem. This in particular implies that, once for a particular minimizer u it is proven that $\nabla u \in L^\infty$, the same must be true for any other minimizer.

2.1. Approximation. For every $\varepsilon \in (0, \varepsilon_0]$ let us consider a smooth function $\mathcal{H}^*_\varepsilon : \mathbb{R}^N \to \mathbb{R}$ with the following basic properties:

(C1) $\mathcal{H}^*_\varepsilon$ is strictly convex and depends only on the modulus, that is $\mathcal{H}^*_\varepsilon(z) = h_\varepsilon(|z|)$;
(C2) $1/q\,(t-1)_+^q \le h_\varepsilon(t) \le A\, t^q + 1$, for some A independent of ε;
(C3) for every $\varepsilon_1 > \varepsilon_2$ we have $h_{\varepsilon_1} \ge h_{\varepsilon_2}$ and h_ε converges to $1/q\,(t-1)_+^q$ as ε goes to 0.

Moreover we require that the functions $\mathcal{H}^*_\varepsilon$ further satisfy the following ellipticity and growth conditions:

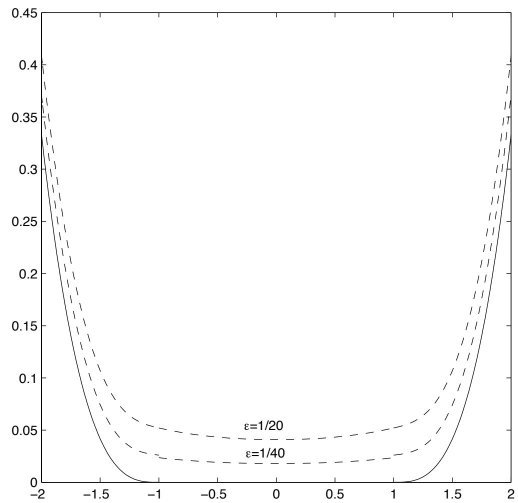

FIGURE 1. The approximating functions $\mathcal{H}^*_\varepsilon$.

(G1) there exist two positive constants $\mu_i = \mu_i(\varepsilon)$ such that

$$\mu_1(1+|z|^2)^{\frac{q-2}{2}}|\xi|^2 \leq \langle D^2\mathcal{H}^*_\varepsilon(z)\,\xi,\xi\rangle \leq \mu_2(1+|z|^2)^{\frac{q-2}{2}}|\xi|^2;$$

(G2) there exist two constants $\lambda > 0$ and $M >> 1$ *independent of* ε such that

$$\frac{1}{\lambda}(1+|z|^2)^{\frac{q-2}{2}}|\xi|^2 \leq \langle D^2\mathcal{H}^*_\varepsilon(z)\,\xi,\xi\rangle \leq \lambda(1+|z|^2)^{\frac{q-2}{2}}|\xi|^2, \quad \text{for every } |z| \geq M; \tag{9.2.3}$$

(G3) there exists a constant κ, independent of ε, such that

$$|\nabla\mathcal{H}^*_\varepsilon(z)| \leq \kappa(1+|z|^2)^{\frac{q-1}{2}}, \quad \text{for every } z \in \mathbb{R}^N.$$

REMARK 9.2.3. For example, in the case $q > 2$, we could simply take $\mathcal{H}^*_\varepsilon(z) = \mathcal{H}^*(z) + \varepsilon(1 + |z|^2)^{\frac{q}{2}}$, while for $q \in (1,2]$ the same choice would be feasible, modulo a smoothing of $\mathcal{H}^*$ around $|z| = 1$.

With the previous assumptions, we have that the equation

$$-\text{div}\,\nabla\mathcal{H}^*_\varepsilon(\nabla u) = f_\varepsilon,$$

f_ε being a smooth approximation of f, is uniformly elliptic outside a ball of radius M, with ellipticity constants independent of ε. Moreover for every fixed $\varepsilon > 0$, this is also a uniformly elliptic equation (but now with ellipticity constants becoming degenerate with ε), so that every solution u_ε is regular enough for computing first derivatives of $\nabla\mathcal{H}^*_\varepsilon(\nabla u_\varepsilon)$: indeed, roughly speaking, our scope will be that of deriving the equation, thus obtaining a linear equation for the gradient which can be used to establish estimates for it.

With the aid of hypotheses (C1)–(C3), it is quite easy to prove the following basic result, granting the convergence of the minimizers of the approximating problem to a minimizer of the original problem.

PROPOSITION 9.2.4. *Let* $f \in L^p(\Omega)$, *with* $p = q/(q-1)$ *and* $\int_\Omega f\,dx = 0$. *We then consider* $\{f_\varepsilon\}_{\varepsilon>0} \subset L^p(\Omega)$ *such that* $f_\varepsilon \rightharpoonup f$ *in* $L^p(\Omega)$ *and having zero-mean. Then every functional*

$$\mathfrak{F}_\varepsilon(u) = \int_\Omega \mathcal{H}^*_\varepsilon(\nabla v(x))\,dx - \int_\Omega f_\varepsilon(x)\,v(x)\,dx,\ \ v \in W^{1,q}_\diamond(\Omega),$$

admits a unique minimizer $u_\varepsilon \in W^{1,q}_\diamond(\Omega)$. *Moreover we get that* $\{u_\varepsilon\}_{\varepsilon>0}$ *weakly converges in* $W^{1,q}$ *to a minimizer of*

$$\mathfrak{F}(u) = \int_\Omega \mathcal{H}^*(\nabla u(x))\,dx - \int_\Omega f(x)\,u(x)\,dx,\ \ u \in W^{1,q}_\diamond(\Omega).$$

PROOF. We are not concerned here with existence and uniqueness of the minimizers, which follow in a standard way, thanks to the assumptions on $\mathcal{H}^*_\varepsilon$. We directly go to the second part of the statement, so, first of all, we show that the sequence of minimizers $\{u_\varepsilon\}_{\varepsilon>0}$ satisfies estimate (9.2.2): we can clearly suppose that

$$\|f_\varepsilon\|_{L^p(\Omega)} \le C, \quad \text{for every } 0 < \varepsilon \le \varepsilon_0,$$

and then by means of the minimality of u_ε we get

$$\int_\Omega \langle \nabla \mathcal{H}^*_\varepsilon(\nabla u_\varepsilon), \nabla u_\varepsilon\rangle\,dx = \int_\Omega f_\varepsilon u_\varepsilon\,dx,$$

and so

$$\begin{aligned}
c_1 \int_\Omega |\nabla u_\varepsilon|^q\,dx - c_2 \le \int_\Omega \langle \nabla \mathcal{H}^*_\varepsilon(\nabla u_\varepsilon), \nabla u_\varepsilon\rangle\,dx &= \int_\Omega f_\varepsilon u_\varepsilon\,dx \\
&\le \|f_\varepsilon\|_{L^p(\Omega)} \|u_\varepsilon\|_{L^q(\Omega)} \\
&\le C\,\|f_\varepsilon\|_{L^p(\Omega)} \|\nabla u_\varepsilon\|_{L^q(\Omega)} \\
&\le \frac{C}{\tau}\|f_\varepsilon\|^p_{L^p(\Omega)} + C\tau\|\nabla u_\varepsilon\|^q_{L^q(\Omega)},
\end{aligned}$$

where we used Poincaré's inequality and Young's inequality, where the constants c_1, c_2 do not depend on ε. From the latter, taking τ small enough, we can easily infer the desired uniform estimate (9.2.2), thus giving the weak compactness of $\{u_\varepsilon\}_{\varepsilon>0}$ in $W^{1,q}(\Omega)$.

We now call u the weak limit (up to a subsequence) of $\{u_\varepsilon\}$ and we show that this is indeed a minimizer of $\mathfrak{F}$. By minimality of every u_ε we know that

$$\int_\Omega \mathcal{H}^*_\varepsilon(\nabla u_\varepsilon)\,dx - \int_\Omega f_\varepsilon\,u_\varepsilon\,dx \le \int_\Omega \mathcal{H}^*_\varepsilon(\nabla v)\,dx - \int_\Omega f_\varepsilon\,v\,dx, \quad \text{for every } v \in W^{1,q}_\diamond(\Omega),$$

then we observe that

$$\int_\Omega \mathcal{H}^*(\nabla u)\,dx \le \liminf_{\varepsilon\to 0} \int_\Omega \mathcal{H}^*(\nabla u_\varepsilon)\,dx \le \liminf_{\varepsilon\to 0} \int_\Omega \mathcal{H}^*_\varepsilon(\nabla u_\varepsilon)\,dx$$

where we used the fact that $\mathcal{H}^*_\varepsilon \ge \mathcal{H}^*$ and the semicontinuity of the term

$$W^{1,q}(\Omega) \ni v \mapsto \int_\Omega \mathcal{H}^*(\nabla v)\,dx.$$

Moreover thanks to the fact that $u_\varepsilon \to u$ in L^q and $f_\varepsilon \rightharpoonup f$ in L^p, we get

$$\int_\Omega f\, u\, dx = \lim_{\varepsilon\to 0} \int_\Omega f_\varepsilon\, u_\varepsilon\, dx.$$

So in the end we have obtained

$$\begin{aligned}\int_\Omega \mathcal{H}^*(\nabla u)\, dx - \int_\Omega f\, u\, dx &\le \liminf_{\varepsilon\to 0} \left(\int_\Omega \mathcal{H}^*_\varepsilon(\nabla u_\varepsilon)\, dx - \int_\Omega f_\varepsilon\, u_\varepsilon\, dx\right)\\ &\le \lim_{\varepsilon\to 0} \left(\int_\Omega \mathcal{H}^*_\varepsilon(\nabla v)\, dx - \int_\Omega f_\varepsilon\, v\, dx\right)\\ &= \int_\Omega \mathcal{H}^*(\nabla v)\, dx - \int_\Omega f\, v\, dx, \text{ for every } v \in W^{1,q}_\diamond(\Omega).\end{aligned}$$

where we used (in last inequality) the monotone convergence of $\mathcal{H}^*_\varepsilon$ to $\mathcal{H}^*$. This clearly proves the minimality of u, thus concluding the proof. □

In what follows, we will set $w_\varepsilon = (1 + |\nabla u_\varepsilon|^2)^{\frac{q}{2}}$ and this function will play a fundamental role in the whole discussion: in particular, we will prove that $w_\varepsilon \in L^\infty$, which in turn will imply that ∇u_ε itself is in L^∞. We will also set

$$k_0 = (M^2 + 1)^{\frac{q}{2}}, \tag{9.2.4}$$

and the set $\{w_\varepsilon > k_0\} = \{|\nabla u_\varepsilon| > M\}$ will be called the *good region* (M is the same as in (G2)).

2.2. Reduction to the boundary. For the solutions u_ε of the approximating problems, we will confine ourselves to prove a uniform L^∞ gradient estimate near the boundary, the interior estimates being simpler and easily deducible from the former. In order to do this, we proceed as in the proof of Theorem 8.3.1 of the previous chapter: this means that, up to apply a diffeomorphism, we can furtherly reduce to prove the required estimate for local solutions in the ball $B = \{x \in \mathbb{R}^N : |x| < 1\}$ of an equation of the type

$$-\text{div}\, A_\varepsilon(x, \nabla u) = f_\varepsilon, \tag{9.2.5}$$

with $z \mapsto A_\varepsilon(x, z)$ satisfying the ellipticity conditions (G1)–(G3) uniformly in x and with $x \mapsto A_\varepsilon(x, z)$ satisfying

$$|A_\varepsilon(x, z) - A_\varepsilon(y, z)| \le L\, (1 + |z|^2)^{\frac{q-1}{2}}\, |x - y|. \tag{9.2.6}$$

To be more precise, the function H_ε so constructed has the form (here we use the same notation as in the proof of Theorem 8.3.1 of Chapter 8)

$$A_\varepsilon(x, z) = |\det \widehat{M}(x)|^{-1}\, \nabla\mathcal{H}^*_\varepsilon(z\, \widehat{M}(x))\, \widehat{M}(x)^t,\ (x, z) \in B \times \mathbb{R}^N,$$

and as already observed, the $C^{2,1}$ assumption on the boundary is needed in order to guarantee that H_ε is Lipschitz with respect to the x variable.

Then, in order to prove Theorem 9.2.1, it will be enough to prove local uniform L^∞ gradient estimates for solutions of (9.2.5) in the ball $B = \{x \in \mathbb{R}^N : |x| < 1\}$.

2.3. Basic tools. Finally, before starting the proof of our main result, we need to recall two fundamental tools in Elliptic Regularity: the first is a higher integrability result for the gradient, which can be found in [**80**, Theorem 3.3.6] (which anyway deals with rather more general situations). This is based on an amended version (proven by Stredulinsky in [**81**]) of the well-known Gehring Lemma ([**54**]). Here we give a slightly simplified version of the statement of Theorem 3.3.6 of [**80**], adapted to our needing.

THEOREM A. *Let $1 < q < N$ and $F \in L^{(q^*)'+\alpha}(\Omega)$ for some $\alpha > 0$, where $q^* = Nq/(N-q)$. Let $\mathcal{A} : \Omega \times \mathbb{R}^N \to \mathbb{R}^N$ be a Borel function satisfying the conditions*

(i) $|\mathcal{A}(x, \nabla u)| \le c_0 |\nabla u|^{q-1} + a$;
(ii) $\langle \mathcal{A}(x, \nabla u), \nabla u\rangle \ge c_1 |\nabla u|^q - \gamma$;

where c_0, c_1 are positive constants and the nonnegative Borel measurable functions a, γ are such that $a \in L^{p+\alpha}(\Omega)$ and $\gamma \in L^{1+\alpha}(\Omega)$, where $p = q/(q-1)$. Then if $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ is a local weak solution of

$$-\operatorname{div} \mathcal{A}(x, \nabla u) = F(x),$$

there exists $\theta > 0$, depending only on N, q, c_0 and α, such that $\nabla u \in L^{q+\theta}_{\mathrm{loc}}(\Omega)$. Moreover for every pair of concentric cubes $Q_\varrho(x_0) \subset Q_{3\varrho}(x_0) \Subset \Omega$ the following estimate holds

$$\left(\fint_{Q_\varrho(x_0)} |\nabla u(x)|^{q+\theta}\, dx \right)^{\frac{q}{q+\theta}} \le C \left[\fint_{Q_{3\varrho}(x_0)} |\nabla u(x)|^q \, dx + \mathcal{Z}(Q_{3\varrho(x_0)}) \right], \tag{9.2.7}$$

with the constant C depending only on N, q, θ and the distance of x_0 from $\partial\Omega$, while $\mathcal{Z}$ is given by

$$\mathcal{Z}(Q_{3\varrho(x_0)}) = \left(\fint_{Q_{3\varrho}(x_0)} |\gamma|^{1+\alpha}\, dx \right)^{\frac{1}{(1+\alpha)}} + \left(\fint_{Q_{3\varrho}(x_0)} |a|^{p+\alpha}\, dx \right)^{\frac{1}{(p+\alpha)}}$$
$$+ \left(\fint_{Q_{3\varrho}(x_0)} |u - \overline{u}_{Q_{3\varrho}(x_0)}|^{q^*}\, dx \right)^{\frac{t}{q^*}} + \left(\fint_{Q_{3\varrho}(x_0)} |F|^{(q^*)'+\alpha}\, dx \right)^{\frac{t'}{(q^*)'+\alpha}},$$

and $\overline{v}_E$ standing for the average of v over a generic set E. Here t is a suitable exponent such that $t < q^*$ and $t' < (q^*)' + \alpha$.

REMARK 9.2.5. The content of Theorem 3.3.6 of [**80**] in indeed more general and it also includes the case $q \ge N$. In this situation, the thesis of Theorem A above is still true taking $F \in L^{1+\alpha}(\Omega)$ and replacing the L^{p^*} norm of $u - \overline{u}_{Q_{3\varrho}}$ in estimate (9.2.7) with any L^r norm (critical case $q = N$) or its L^∞ norm (super critical case $q > N$).

The second result is the following, giving an L^∞ bound for a class of functions which sometimes are called *De Giorgi classes* (but the terminology could not be standard, see also [**62**, Theorem 5.2] for a similar definition and related results). The proof of the following fact can be found in [**57**, Theorem 7.2].

THEOREM B. *Let $v \in W^{1,q}_{\mathrm{loc}}(\Omega)$ be a positive function and suppose that there exist constants $C, \chi > 0$, an exponent $\vartheta > 0$ and a radius $R_0 > 0$, such that for every couple of concentric balls*

$B_\varrho(x_0)$ and $B_R(x_0)$ with $\varrho < R \le R_0$, we get

$$\int_{B\varrho(x_0)} |\nabla(v-k)_+|^q\,dx \le \frac{C}{(R-\varrho)^q}\int_{B_R(x_0)}(v-k)_+^q\,dx + C(\chi^q + k^q R^{-N\vartheta})|\{v>k\}\cap B_R(x_0)|^{1-\frac{p}{N}+\vartheta}, \tag{9.2.8}$$

for every $k \ge k_0$. Then $v \in L^\infty_{\text{loc}}(\Omega)$ and for every $x_0 \in \Omega$ and $R \le \min\{R_0, \text{dist}\,(x_0,\partial\Omega)\}$ we get the estimate

$$\sup_{B_{R/2}(x_0)} v \le C\left[\left(\fint_{B_R(x_0)} v^q\,dx\right)^{\frac{1}{q}} + k_0 + \chi R^{\frac{N\vartheta}{q}}\right]. \tag{9.2.9}$$

Having declared our strategy and introduced all the required tools, we will dedicate the next three sections to the proof of Theorem 9.2.1.

3. Step 1 – Integrability gain

We now want to work with equation (9.2.5), which is defined in the ball B. The first step is to show that $w \in L^2_{\text{loc}}(B)$ uniformly in ε, with an estimate on $\|w\|_2$ depending only on the data and $\|\nabla u\|_q$: in order to guarantee this gain of integrability on w, we wish to use a Moser-type argument ([**72**]), applied to the equation (9.3.1). In the sequel we will drop the subscript ε for the solutions u_ε and for the approximated data f_ε, just for notational convenience: the only important fact is the uniform assumption

$$\|f_\varepsilon\|_{L^{N+\alpha}} \le C, \quad \text{for every } 0<\varepsilon\le\varepsilon_0.$$

The weak formulation of (9.2.5) is given by

$$\int \langle A_\varepsilon(x,\nabla u),\nabla\varphi\rangle\,dx = \int f\,\varphi\,dx, \quad \text{for every } \varphi \in W^{1,q}_0(B),$$

and deriving this equation with respect to x_i, we arrive at

$$\int \langle \nabla_z A_\varepsilon(x,\nabla u)D_i^2 u,\nabla\varphi\rangle\,dx + \int\langle \partial_{x_i}A_\varepsilon(x,\nabla u),\nabla\varphi\rangle\,dx = -\int f\,\varphi_{x_i}\,dx, \tag{9.3.1}$$

where $D_i^2 u$ is the $i-$th column of the Hessian matrix $D^2 u$.

First of all, observe that by means of Theorem A, we already have obtained the gain of integrability $w \in L^{1+\theta/q}$, for some $\theta > 0$ (the A_ε can be easily constructed so to satisfy the mild hypothesis of Theorem A uniformly in ε). Then, let us choose

$$\varphi = u_{x_i}(w^s - k_0^s)_+\zeta^2,$$

with $s \ge \theta/q$, where ζ is a C^∞_c cut-off function supported on some ball $B_R(x_0) \Subset B$, equal to 1 on a smaller concentric ball $B\varrho(x_0)$ and such that its maximal slope is of order $(R-\varrho)^{-1}$.

Inserting φ into (9.3.1) and summing over $i = 1, \dots, N$, we obtain

$$\sum_{i=1}^{N} \int_{\{w>k_0\}} \langle \nabla_z A_\varepsilon(x, \nabla u) D_i^2 u\, u_{x_i}, \nabla w\rangle\, w^{s-1}\, \zeta^2\, dx \tag{9.3.2}$$

$$+\sum_{i=1}^{N} \int \langle \nabla_z A_\varepsilon(x, \nabla u) D_i^2 u, D_i^2 u\rangle\, (w^s - k_0^s)_+\, \zeta^2 dx \tag{9.3.3}$$

$$+2\sum_{i=1}^{N} \int \langle \nabla_z A_\varepsilon(x, \nabla u) D_i^2 u\, u_{x_i}, \nabla \zeta\rangle\, \zeta\, (w^s - k_0^s)_+ u_{x_i}\, dx \tag{9.3.4}$$

$$+\sum_{i=1}^{N} \int_{\{w>k_0\}} \langle \partial_{x_i} A_\varepsilon(x, \nabla u), \nabla w\rangle\, w^{s-1}\, u_{x_i}\, \zeta^2 dx \tag{9.3.5}$$

$$+\sum_{i=1}^{N} \int \langle \partial_{x_i} A_\varepsilon(x, \nabla u), D_i^2 u\rangle\, (w^s - k_0^s)_+\, \zeta^2 dx \tag{9.3.6}$$

$$+2\sum_{i=1}^{N} \int \langle \partial_{x_i} A_\varepsilon(x, \nabla u), \nabla \zeta\rangle\, u_{x_i}\, (w^s - k_0^s)_+\, \zeta dx \tag{9.3.7}$$

$$= -\sum_{i=1}^{N} \int f\, u_{x_i x_i}\, (w^s - k_0^s)_+ \zeta^2 \tag{9.3.8}$$

$$-\sum_{i=1}^{N} \int f\, u_{x_i}\, ((w^s - k_0^s)_+)_{x_i}\, \zeta^2 \tag{9.3.9}$$

$$-\sum_{i=1}^{N} 2 \int f\, \zeta_{x_i} (w^s - k_0^s)_+ u_{x_i} \zeta. \tag{9.3.10}$$

We start by saying that the two main terms are given by (9.3.2) and (9.3.4), which are the cornerstones that in the end will give a Caccioppoli-type inequality. In the sequel, in order to provide a cleaner and easier to follow description of the estimates, we divide the integrals in the previous equation in three groups: the *main terms* (9.3.2), (9.3.3) and (9.3.4), *terms containing* f, i.e. (9.3.8), (9.3.9) and (9.3.10) and *terms containing* $\partial_{x_i} A_\varepsilon$, which are (9.3.5), (9.3.6) and (9.3.7). We aim to analyze each group separately, starting from the basic ones.

3.1. The main terms. So first of all, we proceed to estimate (9.3.2) and (9.3.4): using the fact

$$\nabla w = q\, w^{\frac{q-2}{q}}\, D^2 u \nabla u,$$

the first term can be written as

$$s \int_{\{w>k_0\}} \langle \nabla_z A_\varepsilon(x, \nabla u) D^2 u \nabla u, \nabla w\rangle\, w^{s-1}\, \zeta^2\, dx = \frac{s}{q} \int_{\{w>k_0\}} \langle \nabla_z A_\varepsilon(x, \nabla u) \nabla w, \nabla w\rangle\, \zeta^2\, w^{s-2+\frac{2}{q}} dx$$

and recalling the fact that $\{w > k_0\} = \{|\nabla u| > M\}$ and taking into account (9.2.3), we can estimate this integral from below with

$$\frac{s}{\lambda q}\int_{\{w>k_0\}} |\nabla w|^2\, w^{s-1}\,\zeta^2\,dx = \frac{4s}{\lambda q(s+1)^2}\int_{\{w>k_0\}} \left|\nabla(w^{\frac{s+1}{2}})\right|^2 \zeta^2\,dx$$
$$= \frac{4s}{\lambda q(s+1)^2}\int \left|\nabla\left(w^{\frac{s+1}{2}} - k_0^{\frac{s+1}{2}}\right)_+\right|^2 \zeta^2\,dx := \mathcal{J}.$$

Concerning the other term, we observe

$$2\int \langle \nabla_z A_\varepsilon(x,\nabla u) D^2u\nabla u, \nabla\zeta\rangle\,(w^s - k_0^s)_+\zeta\,dx \le \frac{2}{q}\int |\nabla_z A_\varepsilon(x,\nabla u)||\nabla w|\,|\nabla\zeta|\,\zeta(w^s-k_0^s)_+ w^{\frac{2-q}{q}}\,dx$$
$$\le \frac{2\lambda}{q}\int |\nabla w|\,|\nabla\zeta|\,\zeta\,(w^s-k_0^s)_+\,dx,$$
$$\le \frac{2\lambda}{q}\int_{\{w>k\}} |\nabla w|\,|\nabla\zeta|\,\zeta\,w^s\,dx.$$

Finally, using Young's inequality the last integral can be treated as

$$\int_{\{w>k\}} |\nabla w|\,|\nabla\zeta|\,\zeta\,w^s\,dx \le \tau\int_{\{w>k\}} |\nabla w|^2\,w^{s-1}\,\zeta^2\,dx + \frac{1}{\tau}\int_{\{w>k\}} w^{s+1}\,|\nabla\zeta|^2,\ dx,$$

and the first term can be absorbed in $\mathcal{J}$, taking $\tau > 0$ small enough. Before going on, we observe that the first group contains also the term

$$\sum_{i=1}^N \int \langle \nabla_z A_\varepsilon(\nabla u) D_i^2 u, D_i^2 u\rangle\,(w^s-k_0^s)_+\,\zeta^2 dx,$$

which has positive sign and we could consequently be tempted to drop it: on the contrary, it will be crucial to keep it, in order to absorb similar terms appearing on the right-hand side (for this reason, we will call it *sponge term*, see below). So, it is important to give an estimation from below for it: indeed, we get

$$\text{(9.3.11)}\quad \int \langle \nabla_z A_\varepsilon(\nabla u)D^2u, D^2u\rangle\,(w^s-k_0^s)_+\,\zeta^2 dx \ge \frac{1}{\lambda}\int w^{\frac{q-2}{q}}\,|D^2u|^2\,(w^s-k_0^s)_+\,\zeta^2\,dx := \frac{1}{\lambda}\,\mathcal{S}(B_R),$$

and sometimes, for simplicity, we will call this the *sponge term*. Up to now, we have obtained

$$\int \left|\nabla\left(w^{\frac{s+1}{2}} - k_0^{\frac{s+1}{2}}\right)_+\right|^2 \zeta^2\,dx + \mathcal{S}(B_R) \le C\int_{\{w>k_0\}} \left|w^{\frac{s+1}{2}}\right|^2 |\nabla\zeta^2|\,dx$$
$$+\ \text{Estimates for}\ \left((9.3.5)-(9.3.10)\right),$$

with the constant C depending on q, s and λ, then using the following simple observation

$$\text{(9.3.12)}\qquad w^{s+1}1_{\{w>k_0\}} \le 2\left(w^{\frac{s+1}{2}} - k_0^{\frac{s+1}{2}}\right)_+^2 + 2\,k_0^{s+1},$$

the previous can be recast into

$$
\begin{aligned}
\int_{B_\varrho(x_0)} \left|\nabla\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+\right|^2 \zeta^2\,dx + \mathcal{S}(B_R) \le C \int \left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+^2 |\nabla\zeta|^2\,dx \\
+ \frac{k_0^{2\beta} C}{(R-\varrho)^2}|B_R| \\
+ \text{ Estimates for } ((9.3.5)-(9.3.10))\,,
\end{aligned}
\tag{9.3.13}
$$

which is the kind of estimate which will provide the desired gain of integrability.

3.2. Terms containing the datum f. Using Young's inequality and the fact that $w \ge 1$, we get

$$
\begin{aligned}
-\sum_i \int f\,u_{x_ix_i}\,(w^s-k_0^s)_+\zeta^2\,dx &\le \int |f|\,|D^2u|\,(w^s-k_0^s)_+\,\zeta^2\,dx \\
&\le \tau\int |D^2u|^2\,w^{\frac{q-2}{q}}\,(w^s-k_0^s)_+\,\zeta^2\,dx + \frac{1}{\tau}\int |f|^2\,(w^s-k_0^s)_+\,\zeta^2\,dx,
\end{aligned}
$$

and the first integral can be absorbed in the left-hand side, taking $\tau > 0$ small enough (for example $\tau = 1/(2\lambda)$) by means of the sponge term $\mathcal{S}(B_R)$. Concerning the other term, we can proceed as follows, noticing that

$$
(w^s-k_0^s)_+ \le w^{s+1}1_{\{w>k_0\}} \le 2\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+^2 + 2\,k_0^{s+1}, \tag{9.3.14}
$$

so that using Hölder's inequality and Sobolev inequality we get

$$
\begin{aligned}
\int |f|^2\,(w^s-k_0^s)_+\,\zeta^2\,dx &\le 2\int |f|^2\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+^2\zeta^2\,dx + 2\,k_0^{s+1}\int |f|^2\,\zeta^2\,dx \\
&\le \left(\int_{B_R(x_0)} |f|^N\,dx\right)^{\frac{2}{N}} \left(\int\left(\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+\zeta\right)^{2^*} dx\right)^{\frac{2}{2^*}} \\
&\quad + 2\,k_0^{s+1}\int_{B_R(x_0)} |f|^2\,dx \\
&\le c\,\|f\|^2_{L^{N+\alpha}}\,R^{\frac{2\alpha}{N+\alpha}}\int\left|\nabla\left(\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+\zeta\right)\right|^2 dx \\
&\quad + c\,k_0^{s+1}\|f\|^2_{L^{N+\alpha}}|B_R|^{1-\frac{2}{N+\alpha}},
\end{aligned}
$$

and then we simply observe that

$$
\begin{aligned}
\int\left|\nabla\left(\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+\zeta\right)\right|^2 dx &\le 2\int\left|\nabla\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+\right|^2\zeta^2\,dx \\
&\quad + 2\int|\nabla\zeta|^2\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+^2 dx,
\end{aligned}
$$

so that the term involving the gradient of $(w^{(s+1)/2} - k_0^{(s+1)/2})_+$ will be absorbed in the left-hand side of (9.3.13), choosing the radius R small enough (so that all the estimates will be true for every R smaller than a suitable R_0), while the other term can be kept in the right-hand side of (9.3.13).

Let us go on, with the second term involving f in the derived equation: we use the fact that $|\nabla u| \le w^{\frac{1}{q}}$ and that $w^{\frac{1}{q}+s-1} \le w^s$, so to obtain

$$
\begin{aligned}
-\sum_i \int f\, u_{x_i}((w^s - k_0^s)_+)_{x_i}\, \zeta^2\, dx &\le s \int_{\{w>k_0\}} |f|\, w^{\frac{1}{q}+s-1}\, |\nabla w|\, \zeta^2\, dx \\
&\le \frac{s}{\tau} \int_{\{w>k_0\}} |f|^2\, w^{s+1}\, \zeta^2\, dx + s\tau \int_{\{w>k_0\}} |\nabla w|^2\, w^{s-1}\, \zeta^2\, dx,
\end{aligned}
$$

and the second term can be absorbed by the left-hand side of (9.3.13), while the first can be estimated as before, using (9.3.12). Finally, the last integral: we use the simple fact that $w^{1/q+s} \le w^{s+1}$ and Young's inequality, so that

$$
\begin{aligned}
-2\sum_i \int f\, u_{x_i}\, \zeta_{x_i}\, (w^s - k_0^s)_+\, \zeta\, dx &\le 2 \int_{\{w>k_0\}} |f|\, w^{s+1}\, |\nabla \zeta|\, \zeta\, dx \\
&\le \int_{\{w>k_0\}} |f|^2\, w^{s+1}\, \zeta^2\, dx + \int_{\{w>k_0\}} |\nabla \zeta|^2 \left| w^{\frac{s+1}{2}} \right|^2 dx,
\end{aligned}
$$

and again the first term as already been estimated using (9.3.12), while the second can be recast into

$$
\int_{\{w>k_0\}} |\nabla \zeta|^2 \left| w^{\frac{s+1}{2}} \right|^2 dx \le 2 \int |\nabla \zeta|^2 \left(w^{\frac{s+1}{2}} - k_0^{\frac{s+1}{2}} \right)_+^2 dx + \frac{C\, k_0^{s+1}}{(R-\varrho)^2}\, |B_R|,
$$

again using (9.3.12), the latter being exactly the same term of the right-hand side in (9.3.13).

Putting all together, after the estimation of the first two groups of terms we have obtained

$$
\begin{aligned}
\int \left| \nabla \left(w^{\frac{s+1}{2}} - k_0^{\frac{s+1}{2}} \right)_+ \right|^2 \zeta^2\, dx + \mathcal{S}(B_R) &\le C_1 \int_{\{w>k_0\}} \left| \left(w^{\frac{s+1}{2}} - k_0^{\frac{s+1}{2}} \right)_+ \right|^2 |\nabla \zeta|^2\, dx \\
&\quad + \frac{C\, k_0^{s+1}}{(R-\varrho)^2}\, |B_R| + C\, k_0^{s+1} \|f\|^2_{L^{N+\alpha}}\, |B_R|^{1-\frac{2}{N+\alpha}} \\
&\quad + \text{ Estimates for } ((9.3.5) - (9.3.7)),
\end{aligned}
\tag{9.3.15}
$$

with C_1 depending on q, λ, s, $\|f\|_{L^{N+\alpha}}$.

3.3. Terms involving derivatives of A_ε. We are left with the handling of the terms

$$
\text{Estimates for } ((9.3.5) - (9.3.7)),
$$

and clearly we aim to obtain an inequality of the type (9.3.15). Let us start with the (9.3.5): we have

$$\begin{aligned}\sum_{i=1}^{N}\int_{\{w>k_0\}}\langle \partial_{x_i}A_\varepsilon(x,\nabla u),\nabla w\rangle\, w^{s-1}\,u_{x_i}\,\zeta^2 dx &\le C\int_{\{w>k_0\}}|\nabla w|\,w^{\frac{q-1}{q}}\,w^{s-1}\,w^{\frac{1}{q}}\zeta^2\,dx\\ &\le C\,\tau\int_{\{w>k_0\}}|\nabla w|^2\,w^{s-1}\,\zeta^2\,dx\\ &+\frac{C}{\tau}\int_{\{w>k_0\}}w^{s+1}\,\zeta^2\,dx,\end{aligned}$$

and the first term can be absorbed by the left-hand side of (9.3.15), while for the other we can simply proceed as before, using (9.3.12) in combination with Hölder and Sobolev inequality to get

$$\begin{aligned}\int_{\{w>k_0\}}w^{s+1}\,\zeta^2\,dx \le C\,|B_R|^{\frac{2}{N}}&\int\left|\nabla\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+\right|^2\zeta^2\,dx\\ &+C\,|B_R|^{\frac{2}{N}}\int|\nabla\zeta|^2\left(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}}\right)_+^2 dx\\ &+C\,k_0^{s+1}|B_R|,\end{aligned}$$

the term containing the gradient of $(w^{\frac{s+1}{2}}-k_0^{\frac{s+1}{2}})_+$ being absorbed, taking R small enough.

We go on with (9.3.6), that is

$$\begin{aligned}\sum_{i=1}^{N}\int\langle \partial_{x_i}A_\varepsilon(x,\nabla u),D_i^2 u\rangle\,(w^s-k_0^s)_+\,\zeta^2 dx &\le C\int|D^2u|\,w^{\frac{q-1}{q}}\,(w^s-k_0^s)_+\,\zeta^2\,dx\\ &\le C\tau\int|D^2u|^2\,w^{\frac{q-2}{q}}\,(w^s-k_0^s)_+\,\zeta^2\,dx\\ &+\frac{C}{\tau}\int w\,(w^s-k_0^s)_+\,\zeta^2\,dx,\end{aligned}$$

and the first integral will be absorbed by $\mathcal{S}(B_R)$ in the left-hand side, while the other can be estimated as before.

We are only left with the last integral (9.3.7), for which we can easily derive the following estimate

$$\begin{aligned}2\sum_{i=1}^{N}\int\langle \partial_{x_i}A_\varepsilon(x,\nabla u),\nabla\zeta\rangle\,u_{x_i}\,(w^s-k_0^s)_+\,\zeta dx &\le C\int w\,(w^s-k_0^s)_+\,|\nabla\zeta|\,\zeta\,dx\\ &\le C\int_{\{w>k_0\}}|\nabla\zeta|^2\,w^{s+1}\,dx+C\int_{\{w>k_0\}}\zeta^2\,w^{s+1}\,dx,\end{aligned}$$

and these latter two terms have already been treated.

3.4. Caccioppoli-type inequality. Inserting the latter informations into (9.3.15) and setting $\beta = (s+1)/2$, we can drop the sponge term $\mathcal{S}(B_R)$ and add the term $\int (w^\beta - k_0^\beta)_+^2 \, |\nabla\zeta|^2 \, dx$ on both sides. This yields the Caccioppoli-type inequality

$$\int \left|\nabla\left((w^\beta - k_0^\beta)_+\zeta\right)\right|^2 dx \le C \int \left(w^\beta - k_0^\beta\right)_+^2 |\nabla\zeta|^2 \, dx + \frac{k_0^{2\beta} C}{(R-\varrho)^2}|B_R| + C\, k_0^{2\beta} \|f\|_{L^{N+\alpha}}^2 \, |B_R|^{1-\frac{2}{N+\alpha}}.$$

Finally, we apply the Sobolev inequality and use the properties of the test function ζ, in order to get

$$\left(\int_{B_\varrho(x_0)} \left(w^\beta - k_0^\beta\right)_+^{2^*} dx\right)^{\frac{2}{2^*}} \le \frac{C}{(R-\varrho)^2}\int_{B_R(x_0)} \left(w^\beta - k_0^\beta\right)_+^2 dx + \frac{k_0^{2\beta} C}{(R-\varrho)^2}|B_R| + C\, k_0^{2\beta} \|f\|_{L^{N+\alpha}}^2 \, |B_R|^{1-\frac{2}{N+\alpha}}, \tag{9.3.16}$$

for a constant C that can be chosen so as to depend only on $q, \lambda, \|f\|_{L^{N+\alpha}}$ and β, and not on ε. Starting from $\beta = \frac{1+\theta/q}{2}$, choosing a sequence of concentric balls and iterating a suitable number of times, we finally obtain $w \in L^2_{\text{loc}}$, uniformly in ε. Moreover, for every $B_R \Subset B$ we get an estimate of the type

$$\|w\|_{L^2(B_\varrho)} \le C = C(\lambda, N, \delta, L, \|f\|_{L^{N+\alpha}}, R-\varrho),$$

for every $0 < \varrho < R$.

4. Step 2 – Boundedness of the gradient

The next step is to show that w is in a suitable De Giorgi class, i.e. it satisfies an estimate of the type (9.2.8) (with $p = 2$, as we will see): then using Theorem B we will obtain that $w \in L^\infty_{\text{loc}}$, with an estimate on $\|w\|_\infty$ depending on the L^2 norm of w. Using the fact that $w \in L^2_{\text{loc}}$ uniformly in ε (as shown in the previous section), we will finally obtain that $\nabla u_\varepsilon \in L^\infty_{\text{loc}}$ uniformly in ε.

So, in order to conclude, we proceed exactly as in the last section, but with a different choice for the test function, as far as we now want to show that $(w-k)_+$ satisfies a suitable Caccioppoli inequality, for every $k \ge k_0$.

Keeping this in mind, the right choice for the test function is given by $\varphi = u_{x_i}(w-k)_+\zeta^2$. Inserting this into (9.3.1) and summing over $i = 1,\dots,N$, we obtain

$$\sum_{i=1}^N \int \langle \nabla_z A_\varepsilon(x,\nabla u) D_i^2 u\, u_{x_i}, \nabla(w-k)_+\rangle\, \zeta^2\, dx \tag{9.4.17}$$

$$+\sum_{i=1}^N \int \langle \nabla_z A_\varepsilon(x,\nabla u) D_i^2 u, D_i^2 u\rangle\, (w-k)_+\, \zeta^2 dx \tag{9.4.18}$$

$$+2\sum_{i=1}^N \int \langle \nabla_z A_\varepsilon(x,\nabla u) D_i^2 u\, u_{x_i}, \nabla\zeta\rangle\, \zeta\, (w-k)_+\, u_{x_i}\, dx \tag{9.4.19}$$

$$+\sum_{i=1}^N \int \langle \partial_{x_i} A_\varepsilon(x,\nabla u), \nabla(w-k)_+\rangle\, u_{x_i}\, \zeta^2 dx \tag{9.4.20}$$

$$+\sum_{i=1}^N \int \langle \partial_{x_i} A_\varepsilon(x,\nabla u), D_i^2 u\rangle\, (w-k)_+\, \zeta^2 dx \tag{9.4.21}$$

$$+2\sum_{i=1}^N \int \langle \partial_{x_i} A_\varepsilon(x,\nabla u), \nabla\zeta\rangle\, u_{x_i}\, (w-k)_+\, \zeta dx \tag{9.4.22}$$

$$= -\sum_{i=1}^N \int f\, u_{x_i x_i}\, (w^s - k_0^s)_+ \zeta^2\, dx \tag{9.4.23}$$

$$-\sum_{i=1}^N \int f\, u_{x_i}\, ((w^s - k_0^s)_+)_{x_i}\, \zeta^2\, dx \tag{9.4.24}$$

$$-\sum_{i=1}^N 2\int f\, \zeta_{x_i} (w^s - k_0^s)_+ u_{x_i} \zeta\, dx. \tag{9.4.25}$$

As before, we divide the estimations in three groups.

4.1. The main terms. We observe that, using the fact that

$$\nabla(w-k)_+ = q\, w^{\frac{q-2}{q}}\, D^2 u \nabla u\, 1_{\{w>k\}},$$

the integral (9.4.17) can be written as

$$\int \langle \nabla_z A_\varepsilon(x,\nabla u) D^2 u \nabla u, \nabla(w-k)_+\rangle\, \zeta^2\, dx = \frac{1}{q}\int \langle \nabla_z A_\varepsilon(x,\nabla u)\nabla(w-k)_+, \nabla(w-k)_+\rangle\, \zeta^2\, w^{\frac{2-q}{q}}\, dx$$

and the previous integral is restricted to the set $\{w > k\}$, so that taking $k \geq k_0$, with k_0 given by (9.2.4), and taking into account (9.2.3) we get

$$\int \langle \nabla_z A_\varepsilon(x,\nabla u)\nabla(w-k)_+, \nabla(w-k)_+\rangle\, w^{\frac{2-q}{2}} \zeta^2\, dx \geq \frac{1}{q\lambda}\int |\nabla(w-k)_+|^2 \zeta^2.$$

Then we observe that the term (9.4.18) can be treated as follows (this is the sponge term, as before)

$$\sum_{i=1}^{N}\int \langle \nabla_z A_\varepsilon(x,\nabla u) D_i^2 u, D_i^2 u\rangle\,(w-k)_+\zeta^2\,dx \geq \frac{1}{\lambda}\int |D^2 u|^2\, w^{\frac{q-2}{q}}\,(w-k)_+\,\zeta^2\,dx := \frac{1}{\lambda}\mathcal{S}'(B_R),$$

while concerning the third integral (9.4.19), we get

$$\begin{aligned}
2\int \langle \nabla_z A_\varepsilon(x,\nabla u) D^2 u\nabla u, \nabla\zeta\rangle\,(w-k)_+\zeta\,dx &= \frac{2}{q}\int \langle \nabla_z A_\varepsilon(x,\nabla u)\nabla(w-k)_+, \nabla\zeta\rangle \\
&\quad\times (w-k)_+\, w^{\frac{2-q}{q}}\,\zeta\,dx \\
&\leq \frac{2}{q}\int |\nabla_z A_\varepsilon(x,\nabla u)||\nabla(w-k)_+|\,|\nabla\zeta| \\
&\quad\times \zeta\,(w-k)_+\, w^{\frac{2-q}{q}}\,dx \\
&\leq \frac{2\lambda}{q}\int |\nabla(w-k)_+|\,|\nabla\zeta|\,\zeta(w-k)_+\,dx,
\end{aligned}$$

where we have used again (9.2.3) and the fact that we are integrating over a region where $|\nabla u| \geq M$. Summarizing, we have obtained

$$\begin{aligned}
\int |\nabla(w-k)_+|^2\,\zeta^2\,dx + \int |D^2 u|^2\, w^{\frac{q-2}{q}}\,(w-k)_+\,\zeta^2\,dx &\leq C\int |\nabla(w-k)_+|\,|\nabla\zeta|\,\zeta(w-k)_+\,dx \\
&\quad + \text{Estimates for (9.4.20) – (9.4.25)},
\end{aligned}$$

and with standard calculations we can then obtain the inequality

(9.4.26)

$$\begin{aligned}
\int |\nabla(w-k)_+|^2\,\zeta^2\,dx + \int |D^2 u|^2\, w^{\frac{q-2}{q}}\,(w-k)_+\,\zeta^2\,dx &\leq C\int |\nabla\zeta|^2(w-k)_+^2\,dx, \\
&\quad + \text{Estimates for (9.4.20) – (9.4.25)},
\end{aligned}$$

for every $k \geq k_0$.

4.2. Terms containing f_ε. Using the fact that $w \geq 1$ and Young's inequality, we get

$$\begin{aligned}
\sum_i \int f\, u_{x_i x_i}(w-k)_+\zeta^2\,dx &\leq \int |f|\left(\sum_i |D_i^2 u|^2\right)^{\frac{1}{2}}(w-k)_+\,\zeta^2\,dx \\
&\leq \tau\int |D^2 u|^2 w^{\frac{q-2}{q}}(w-k)_+\zeta^2\,dx + \frac{1}{\tau}\int |f|^2(w-k)_+\zeta^2\,dx
\end{aligned}$$

and observe that the first integral can be absorbed, choosing τ small enough, by the left-hand side. Concerning the second integral, we observe that we can get

$$\begin{aligned}\int |f|^2\,(w-k)_+\,\zeta^2\,dx &\le \frac{1}{2}\int |f|^2\,(w-k)_+^2\,\zeta^2 + \frac{1}{2}\int_{\{w\ge k\}\cap B_R} |f|^2\\ &\le \frac{1}{2}\left(\int |f|^N\,dx\right)^{\frac{2}{N}}\left(\int (w-k)_+^{2^*}\zeta^{2^*}\,dx\right)^{\frac{2}{2^*}}\\ &\quad+\frac{1}{2}\|f\|^2_{L^{N+\alpha}}\,|\{w\ge k\}\cap B_R|^{1-\frac{2}{N+\alpha}}\\ &\le \frac{c}{2}\|f\|^2_{L^{N+\alpha}}\,R^{\frac{2\alpha}{N+\alpha}}\int |\nabla((w-k)_+\zeta)|^2\,dx\\ &\quad+\frac{1}{2}\|f\|^2_{L^{N+\alpha}}\,|\{w\ge k\}\cap B_R|^{1-\frac{2}{N+\alpha}},\end{aligned}$$

just as in the previous section.

Let us then consider (9.4.24): we first observe that $|\nabla u|\le w^{\frac{1}{q}}$, so that

$$\begin{aligned}\sum_i \int f\,u_{x_i}((w-k)_+)_{x_i}\,\zeta^2\,dx &\le \int |f|\,w^{\frac{1}{q}}\,|\nabla(w-k)_+|\,\zeta^2\,dx\\ &\le \frac{1}{\tau}\int |f|^2\,w^{\frac{2}{q}}\,\zeta^2\,dx + \tau\int |\nabla(w-k)_+|^2\,\zeta^2\,dx,\end{aligned}$$

and observe that the second term will be absorbed in the left-hand side.

As before, we write $w^{\frac{1}{q}}$ in place of ∇u and then we use Young's inequality, so that (9.4.25) can be estimated as follows

$$2\sum_i \int f\,u_{x_i}\,\zeta_{x_i}\,(w-k)_+\,\zeta\,dx \le \int |f|^2\,w^{\frac{2}{q}}\zeta^2\,dx + \int |\nabla\zeta|^2\,(w-k)_+^2\,dx.$$

Before putting all the estimates together, we observe that the last two integral have a common term, which can be treated as follows

$$\begin{aligned}\int |f|^2\,w^{\frac{2}{q}}\,\zeta^2\,dx \le \int |f|^2\,w\,\zeta^2\,dx &\le \int |f|^2\,(w-k)_+\,\zeta^2\,dx + k\int_{\{w>k\}\cap B_R} |f|^2\,dx,\\ &\le \int |f|^2\,(w-k)_+\,\zeta^2\,dx + k\,\|f\|^2_{L^{N+\alpha}}\,|\{w\ge k\}\cap B_R|^{1-\frac{2}{N+\alpha}},\end{aligned}$$

and the first integral above has already been estimated.

All in all, we have obtained the following:

$$\begin{aligned}\int |\nabla(w-k)_+|^2\,\zeta^2\,dx + \mathcal{S}'(B_R)dx &\le C_1\int (w-k)_+^2\,|\nabla\zeta|^2\,dx\\ &\quad+C\,(1+k)\|f\|^2_{L^{N+\alpha}}\,|\{w\ge k\}\cap B_R|^{1-\frac{2}{N+\alpha}}\\ &\quad+\text{Estimates for (9.4.20) – (9.4.22)},\end{aligned}\tag{9.4.27}$$

with C_1 depending on $\|f\|_{L^{N+\alpha}}$.

4.3. Terms involving derivatives of A_ε. Using the growth conditions on $\partial_{x_i}\nabla\mathcal{H}^*_\varepsilon(x,\nabla u)$ and the fact that all the integrals are restricted to a region where $\{|\nabla u|>M\}$, we get

$$(9.4.20) \le c\int w^{\frac{q-1}{q}}\,|D^2u|\,(w-k)_+\,\zeta^2\,dx \le c\tau\int w^{\frac{q-2}{q}}\,|D^2u|^2\,(w-k)_+\,\zeta^2\,dx + \frac{c}{\tau}\int w\,(w-k)_+\,\zeta^2\,dx,$$

and we observe that the first term can be absorbed by the sponge term $\mathcal{S}'(B_R)$ as before, while using the simple inequality

$$w(w-k)_+ \le (w-k)_+^2 + k(w-k)_+ \le \frac{3}{2}(w-k)_+^2 + \frac{k^2}{2},$$

and the fact that

$$|\{w>k\}\cap B_R(x_0)| \le |B_R(x_0)|^{\frac{2}{N+\alpha}}|\{w>k\}\cap B_R(x_0)|^{1-\frac{2}{N+\alpha}} \le c\,|\{w>k\}\cap B_R(x_0)|^{1-\frac{2}{N+\alpha}},$$

the second term can be written as

$$\begin{aligned}\int w\,(w-k)_+\,\zeta^2\,dx &\le c\int (w-k)_+^2\,\zeta^2\,dx + ck^2|\{w>k\}\cap B_R(x_0)|\\ &\le cR^2\int|\nabla(w-k)_+|^2\,\zeta^2 dx + cR^2\int|\nabla\zeta|^2\,(w-k)_+^2\,dx\\ &\quad+\bar{c}\,k^2|\{w>k\}\cap B_R(x_0)|^{1-\frac{2}{N+\alpha}},\end{aligned}$$

which is a kind of term that we have already treated (the first can be absorbed, taking R small enough, while the second and the third are good).

Concerning (9.4.21), we easily get

$$\begin{aligned}\sum_{i=1}^N\int\langle\partial_{x_i}A_\varepsilon(x,\nabla u),D_i^2u\rangle\,(w-k)_+\,\zeta^2 dx &\le c\int w^{\frac{q-1}{q}}\,|\nabla(w-k)_+|\,|\nabla u|\,\zeta^2\,dx\\ &\le \int w\,|\nabla(w-k)_+|\,\zeta^2\,dx,\\ &\le c\tau\int|\nabla(w-k)_+|^2\,\zeta^2\,dx + \frac{c}{\tau}\int_{\{w>k\}} w^2\,\zeta^2\,dx,\end{aligned}$$

and the first can be absorbed, while the second has been already estimated in the case of (9.4.20): we can simply use $w^2\le 2(w-k)_+^2+2k^2$ and proceed as before.

Finally, the last term:

$$(9.4.22) \le c\int w\,(w-k)_+\,|\nabla\zeta|\,\zeta\,dx \le \frac{c}{2}\int_{\{w>k\}} w^2\,\zeta^2\,dx + \frac{c}{2}\int(w-k)_+^2\,|\nabla\zeta|^2\,dx,$$

both being terms already estimated.

4.4. Caccioppoli-type inequality. Putting all these estimates together into (9.4.27), we can finally obtain the inequality

$$\int_{B_\varrho(x_0)} |\nabla(w-k)_+|^2\,dx \le \frac{C}{(R-\varrho)^2}\int_{B_R(x_0)} (w-k)_+^2\,dx + C\,k^2\,|\{w\ge k\}\cap B_R(x_0)|^{1-\frac{2}{N+\alpha}}, \tag{9.4.28}$$

which is valid for every $k\ge k_0$, with C depending on $\lambda, N, \|f\|_{L^{N+\alpha}}$, but not on ε. In particular, inequality (9.4.28) implies that w is in a De Giorgi class, so that thanks to Theorem B we get for every $B_R \Subset B$

$$\sup_{B_{R/2}(x_0)} w \le C = C(R, \|w\|_{L^2(B_R)}, k_0),$$

which gives the desired conclusion, thanks to the fact that $w\in L^2_{\mathrm{loc}}$, uniformly in ε as already proven in the previous section.

5. Step 3 – Conclusion

Let us now take $u\in W^{1,q}_\diamond(\Omega)$ weak solution of (9.2.1), under the hypotheses of Theorem 9.2.1. What we have proven so far implies in particular that the minimizers u_ε of $\mathfrak{F}_\varepsilon$, which are equi-bounded in $W^{1,q}$, are also equi-bounded in $W^{1,\infty}$, so that $u_\varepsilon \overset{*}{\rightharpoonup} \widetilde{u}$ in $W^{1,\infty}$, up to a subsequence. It is only left to observe that this limit $\widetilde{u}\in W^{1,\infty}(\Omega)$ is a minimizer of $\mathfrak{F}$ by means of Proposition 9.2.4 and thus another weak solution of (9.2.1): this and the fact that (see Remark 9.2.2)

$$\nabla\mathcal{H}^*(\nabla\widetilde{u}(x)) = \nabla\mathcal{H}^*(\nabla u(x)), \quad \text{for } \mathscr{L}^N\text{–a.e. } x\in\Omega,$$

finally imply $\nabla u\in L^\infty(\Omega)$ as desired.

APPENDIX A

Curves in a metric space

We recall some basic facts about spaces of curves in a metric space. Here (X, d) is a Polish space, equipped with a given Borel measure $\mathfrak{m}$, while $I = [0, T] \subset \mathbb{R}$ is a compact interval.

1. Summable curves

For $p \in [1, +\infty)$, we say that a curve $\mu : I \to X$ belongs to $\mathcal{L}^p(I; X)$ if μ is Borel measurable and

$$\int_I d(\mu(t), x_0)^p \; dt < +\infty,$$

where x_0 is a point of X (clearly the definition does not depend on the choice of x_0, by means of the triangular inequality).

As in the Euclidean case, we call $L^p(I; X)$ the space of equivalence classes (with respect to the relation *equivalence* $\mathscr{L}^1$*-a.e.*) of functions in $\mathcal{L}^p(I; X)$: this is clearly a metric space, endowed with the distance

$$d_p(\mu_1, \mu_2) = \left(\int_I d(\mu_1(t), \mu_2(t))^p \; dt \right)^{1/p}.$$

In the case of $p = +\infty$, we define $\mathcal{L}^\infty(I; X)$ as the space of all curves $\mu : I \to X$ such that

$$\operatorname*{ess\,sup}_{t \in I} d(\mu(t), x_0) < +\infty,$$

for some $x_0 \in X$ and again $L^\infty(I; X)$ is the space of equivalence classes, with the distance

$$d_\infty(\mu_1, \mu_2) = \operatorname*{ess\,sup}_{t \in I} d(\mu_1(t), \mu_2(t)).$$

Remark A.1.1. It is straightforward to see that if X is separable and complete, then for every $p \in [1, +\infty)$ the metric space $L^p(I; X)$ is complete and separable, too. Moreover as in the Euclidean case, it is possible to show that if $\mu_n \to \mu$ in $L^p(I; X)$, then there exists a subsequence $\{\mu_{n_k}\}_{k \in \mathbb{N}}$ converging to μ $\mathscr{L}^1$-a.e.

2. Continuous curves

Let $C(I; X)$ be the space of all continuous curves in X, endowed with the topology of the uniform convergence, that is

$$\mu_n \to \mu \text{ in } C(I; X) \iff d_\infty(\mu_n, \mu) = \max_{t \in I} d(\mu_n(t), \mu(t)) \to 0.$$

We recall the definition of *metric derivative* of $\mu \in C(I;X)$ at the point $t \in I$, defined as

$$|\mu'|(t) = \lim_{s\to t} \frac{d(\mu(t),\mu(s))}{|s-t|}, \tag{A.2.1}$$

every time this limit exists.

Remark A.2.1. When $X = \mathbb{R}^N$ equipped with the usual Euclidean distance, if $\mu : I \to X$ is differentiable at the point t_0, then

$$|\mu'|(t_0) = \left\| \frac{d\mu}{dt}(t_0) \right\|,$$

that is $|\mu'|(t_0)$ is nothing but the Euclidean norm of the derivative of μ at the point t_0.

For $p \in [1,+\infty]$, we consider the space $AC^p(I;X) \subset C(I;X)$, defined as follows: we say that $\mu \in AC^p(I;X)$ if there exists some $\psi \in L^p(I;\mathbb{R})$ such that

$$d(\mu(t),\mu(s)) \le \int_s^t \psi(r)\, dr, \quad \text{for every } s,t \in I \text{ such that } s \le t. \tag{A.2.2}$$

The elements of $AC^p(I;X)$ are called *absolutely continuous curves with finite p-energy* (or simply *absolutely continuous curves*, in the case $p = 1$) and they have the nice property of being almost everywhere metric differentiable, as the following Theorem states (see [**6**, Theorem 1.1.2]).

Theorem A.2.2. *If $\mu \in AC^p(I;X)$, with $p \ge 1$, then the limit (A.2.1) exists for $\mathscr{L}^1$-a.e. $t \in I$. The function $t \mapsto |\mu'|(t)$ belongs to $L^p(I;\mathbb{R})$ and*

$$d(\mu(t),\mu(s)) \le \int_s^t |\mu'|(r)\, dr, \quad \textit{for every } s,t \in I \textit{ such that } s \le t.$$

Moreover we have

$$|\mu'|(t) \le \psi(t), \ \textit{for } \mathscr{L}^1\textit{-a.e. } t \in I,$$

for every $\psi \in L^p(I;\mathbb{R})$ for which (A.2.2) holds.

Next result is a sort of Poincaré-Wirtinger inequality with a trace term, that holds true for curves in an arbitrary metric space.

Theorem A.2.3 (Poincaré-Wirtinger Inequality). *If $\mu \in AC^p(I;X)$, with $p \in (1,+\infty)$, then for every $x_0 \in X$ we get*

$$\begin{aligned} \left(\int_0^T d(\mu(t),x_0)^p\, dt\right)^{\frac{1}{p}} \le C(p,T) &\left[\left(\int_0^T |\mu'|^p(t)\, dt\right)^{\frac{1}{p}} + \frac{|d(\mu(0),x_0) - d(\mu(T),x_0)|}{T^{\frac{p-1}{p}}}\right] \\ &+ \xi_p(\mu(0),\mu(T);x_0), \end{aligned} \tag{A.2.3}$$

where the constant $C(p,T)$ is given by

$$C(p,T) = \frac{pT}{2\pi(p-1)^{\frac{1}{p}}} \sin\left(\pi \frac{p-1}{p}\right), \tag{A.2.4}$$

while the function $\xi_p : (X \times X) \times X \to \mathbb{R}$ *is defined by*

$$\xi_p(x,y;z) = \begin{cases} \left(\dfrac{T}{p+1} \dfrac{d(x,z)^{p+1} - d(y,z)^{p+1}}{d(x,z) - d(y,z)} \right)^{\frac{1}{p}}, & d(x,z) \neq d(y,z), \\ T^{\frac{1}{p}} d(x,z), & d(x,z) = d(y,z). \end{cases}$$

In particular, if $\mu \in AC^p(I;X)$ *happens to be a loop with base point* $x_0 \in X$*, that is* $\mu(0) = \mu(T) = x_0$*, then*

$$\left(\int_0^T d(\mu(t), x_0)^p \, dt \right)^{\frac{1}{p}} \leq C(p,T) \left(\int_0^T |\mu'|^p(t) \, dt \right)^{\frac{1}{p}}. \tag{A.2.5}$$

Proof. The proof is the same as in [**53**], except for the fact that we allow the exponent p to vary in $(1,+\infty)$: we simply use the Poincaré-Wirtinger inequality for real functions of one variable.

Let us set

$$f(t) = d(\mu(t), x_0) - \left(1 - \frac{t}{T}\right) d(\mu(0), x_0) - \frac{t}{T} d(\mu(T), x_0), \ t \in [0,T],$$

then it is easily seen that $f \in AC^p(I;\mathbb{R})$, with $f(0) = f(T) = 0$, so for it the standard Poincarè-Wirtinger inequality holds true, that is

$$\left(\int_0^T |f(t)|^p \, dt \right)^{\frac{1}{p}} \leq C(p,T) \left(\int_0^T |f'(t)|^p \, dt \right)^{\frac{1}{p}},$$

where the best constant $C(p,T)$ is given by (A.2.4) (see [**83**, equation (7a)], for example, where the best constant is computed, together with the function that realizes it).

We now observe that

$$|f'(t)| \leq |\mu'|(t) + \frac{1}{T} |d(\mu(0), x_0) - d(\mu(T), x_0)|, \ \mathscr{L}^1\text{-a.e. } t \in I,$$

so that Minkowski inequality yields

$$\left(\int_0^T |f(t)|^p \, dt \right)^{\frac{1}{p}} \leq C(p,T) \left[\left(\int_0^T |\mu'|^p(t) \, dt \right)^{\frac{1}{p}} + \frac{|d(\mu(0), x_0) - d(\mu(T), x_0)|}{T^{\frac{p-1}{p}}} \right]. \tag{A.2.6}$$

Moreover by Minkowski inequality again we get

$$\left(\int_0^T d(\mu(t), x_0)^p \, dt \right)^{\frac{1}{p}} \leq \left(\int_0^T |f(t)|^p \, dt \right)^{\frac{1}{p}} + \left(\int_0^T \left(\left(1 - \frac{t}{T}\right) d(\mu(0), x_0) + \frac{t}{T} d(\mu(T), x_0) \right)^p \, dt \right)^{\frac{1}{p}}.$$

Computing the integral in the right-hand side and using (A.2.6), we obtain (A.2.3). □

Another remarkable property of curves in AC^p is that they can be reparametrized by arc length. Precisely, we have the following (see [**9**] for a proof):

LEMMA A.2.4 (Reparametrization Lemma). *For* $p \in [1,+\infty]$*, suppose* $\mu \in AC^p(I;X)$ *and let*

$$\ell(\mu) = \int_I |\mu'|(t)\ dt.$$

Then there exists a strictly increasing left-continuous function

$$\mathfrak{t} : [0,\ell(\mu)] \to [0,T],$$

such that:

(1) $\overline{\mu} = \mu \circ \mathfrak{t} \in AC^\infty([0,\ell(\mu)];X)$*;*
(2) $\overline{\mu}([0,\ell(\mu)]) = \mu([0,T])$*;*
(3) $|\overline{\mu}'|(t) = 1$*, for* $\mathscr{L}^1$*-a.e.* $t \in [0,\ell(\mu)]$*.*

REMARK A.2.5. The time rescaling $\mathfrak{t}$ given by the previous Lemma is defined as

$$\mathfrak{t}(s) = \inf\left\{ t \in [0,T]\ :\ s = \int_0^t |\mu'|(r)\ dr \right\}.$$

We remark that in general this is not a continuous function: the important fact is that at its discontinuity points, the jumps of $\mathfrak{t}$ corresponds to time intervals where μ is constant.

DEFINITION A.2.6. The space $AC^p(I;X)$ is endowed with the following notion of convergence: we say that $\{\mu_n\}_{n\in\mathbb{N}} \subset AC^p(I;X)$ *weakly converges*[1] to some $\mu \in AC^p(I;X)$, and we write $\mu_n \rightharpoonup \mu$, if

(i) $\lim_{n\to\infty} \max_{t\in I} d(\mu_n(t),\mu(t)) = 0$;
(ii) the sequence $\{|\mu_n'|\}_{n\in\mathbb{N}}$ is equi-bounded in $L^p(I;\mathbb{R})$ and equi-integrable;

where we intend that, if $p > 1$, then the equi-integrability condition is redundant.

Finally, we recall a compactness criterion for the space of continuous curves $C(I;X)$.

THEOREM A.2.7 (Ascoli-Arzelà). *Given a sequence* $\{\mu_n\}_{n\in\mathbb{N}} \subset C(I;X)$*, this is relatively compact if and only if the following are satisfied:*

(i) $\{\mu_n\}_{n\in\mathbb{N}}$ *is equi-continuous;*
(ii) for every $t \in I$*, the set* $\{\mu_n(t)\ :\ n \in \mathbb{N}\}$ *is relatively compact in* X*.*

3. Curves of bounded variation

Given a curve $\mu : I \to X$, it is possible to define its *pointwise total variation*

$$\mathrm{Var}(\mu;I) = \sup\left\{ \sum_{i=0}^{k} d(\mu(t_i),\mu(t_{i+1}))\ :\ 0 = t_0 < t_1 < \cdots < t_k < t_{k+1} = T \right\}, \tag{A.3.1}$$

where the supremum is taken over all finite partitions of I and we say that μ is *rectifiable* if $\mathrm{Var}(\mu) < +\infty$. For absolutely continuous curves, we have the following (see [**9**] for a proof):

[1]When $X = \mathbb{R}^N$ equipped with the Euclidean metric, the spaces $AC^p(I;\mathbb{R}^N)$ are basically the same as the standard Sobolev spaces $W^{1,p}(I;\mathbb{R}^N)$. Observe that in this case, when $1 < p < \infty$, we have

$$\mu_n \overset{W^{1,p}}{\rightharpoonup} \mu \Longleftrightarrow \sup_n \|\mu_n\|_{W^{1,p}} < +\infty \Longleftrightarrow \mu_n \to \mu \text{ uniformly on } I \text{ and } \sup_n \|\mu_n'\|_{L^p} < +\infty,$$

where convergence is always intended up to a subsequence. This chain of equivalences justifies why we have decided to call *weak convergence* the convergence introduced in $AC^p(I;X)$, rather than simply calling it *uniform convergence*.

Lemma A.3.1. *Let $p \in [1, +\infty]$ and $\mu \in AC^p(I; X)$; then it holds*

$$\mathrm{Var}(\mu; I) = \int_I |\mu'|(t)\ dt. \tag{A.3.2}$$

In particular, every $\mu \in AC^p(I; X)$ is rectifiable.

We now want to introduce the space of curves of bounded variation: we essentially follow [**49**].

Let $\mu : I \to X$ be a Borel measurable curve, we say that μ is *approximately continuous* at $t \in I$ if there exists $x \in X$ such that all the sets

$$X_\varepsilon = \{s \in I\ :\ d(\mu(s), x) > \varepsilon\},$$

have 0-density at t, that is

$$\lim_{r \to 0^+} \frac{\mathscr{L}^1((t - r, t + r) \cap X_\varepsilon)}{r} = 0, \quad \text{for every } \varepsilon > 0.$$

The point x, if it exists, is unique and is called *approximate limit* of μ in t. We indicate with S_μ the set of points where the approximate limit does not exist: we point out that there holds $\mathscr{L}^1(S_\mu) = 0$ (see [**49**, Theorem 2.9.13]).

Given a Borel measurable curve $\mu : I \to X$, we can also define its left and right approximate limits: for every $t \in I$, we define $x = \mu^+(t)$ if the sets

$$\{s \in I\ :\ t < s,\ d(\mu(s), x) > \varepsilon\}$$

have 0-density at t for every $\varepsilon > 0$. Similarly we set $x = \mu^-(t)$ if

$$\{s \in I\ :\ t > s,\ d(\mu(s), x) > \varepsilon\}$$

have 0-density at t for every $\varepsilon > 0$.

Remark A.3.2. It is easily seen that for every $t \in I \setminus S_\mu$, the limits $\mu^+(t)$ and $\mu^-(t)$ exist and they coincide with the approximate limit of μ in t.

Let $\mu \in L^1(I; X)$ be a summable curve, we define its *essential total variation* as

$$|D\mu|(I) = \sup\left\{\sum_{i=0}^{k} d(\mu(t_i), \mu(t_{i+1}))\ :\ 0 < t_0 < \cdots < t_{k+1} < T\right\}, \tag{A.3.3}$$

where the supremum is taken over all finite partitions of $I \setminus S_\mu$.

We then say that μ has *bounded variation* if $|D\mu|(I) < +\infty$ and we write $BV(I; X)$ to indicate the space of curves of bounded variation, with values in the metric space X. This is clearly a metric space, too, with distance given by

$$d_{BV}(\mu_1, \mu_2) = d_1(\mu_1, \mu_2) + \big||D\mu_1|(I) - |D\mu_2|(I)\big|,\ \mu_1, \mu_2 \in BV(I; X).$$

Curves of bounded variation posses left and right approximate limits at every point: we give a proof of this fact (see also [**49**, 2.5.16]).

Lemma A.3.3. *If $\mu \in BV(I; X)$, then for every $t \in (0, T)$ there exist $\mu^+(t)$ and $\mu^-(t)$. Furthermore, the same conclusion holds for $\mu^+(0)$ and $\mu^-(T)$.*

PROOF. We define the nondecreasing function

$$V(t) = |D\mu|([0,t]),\ t \in I,$$

then for every $t \in I$ we have $V(t^-) \le V(t) \le V(t^+)$, where

$$V(t^-) = \sup\{V(s)\ :\ s < t\} = \lim_{s\to t^-} V(s),$$

$$V(t^+) = \inf\{V(s)\ :\ s > t\} = \lim_{s\to t^+} V(s).$$

We just prove that $\mu^-(t)$ exists for every $t \in (0,T]$: the other part of the statement can be proven in the same way. Indeed, observe that

$$d(\mu(s_1),\mu(s_2)) \le V(t^-) - V(s_1),\ s_1, s_2 \in I \setminus S_\mu \text{ such that } s_1 < s_2 < t,$$

which implies, by means of the completeness of X, the existence of

$$\lim_{s\to t^-} \mu(s) \in X.$$

This has to coincide with the approximate limit $\mu^-(t)$, concluding the proof. □

REMARK A.3.4. For every $p \in [1,+\infty]$, if $\mu \in AC^p(I;X)$ we have

$$|D\mu|(I) = \mathrm{Var}(\mu; I).$$

In particular, from Lemma A.3.1 it follows that $AC^p(I;X) \subset BV(I;X)$ and

$$|D\mu|(I) = \int_I |\mu'|(t)\ dt,\ \mu \in AC^p(I;X).$$

We conclude this section with a metric variation of a classical compactness result on BV functions: the proof can be found in [**4**, Theorem 2.4].

THEOREM A.3.5. *Let (X,d) be a locally compact, complete and separable metric space. Let $\{\mu_n\}_{n\in\mathbb{N}} \subset BV(I;X)$ be a sequence such that*

$$\sup_{n\in\mathbb{N}} d_{BV}(\mu_n, x_0) < +\infty,$$

for some $x_0 \in X$. Then there exists a subsequence $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ converging in $L^1(I;X)$ to $\mu \in BV(I;X)$ and

$$|D\mu|(I) \le \liminf_{k\to+\infty} |D\mu_{n_k}|(I).$$

APPENDIX B

The characteristics method for the continuity equation

We have used several times the fact that, for solutions of the continuity equation with smooth vector fields, we have a representation formula in terms of the flow map of these vector fields and of the initial datum. To keep the exposition as self-contained as possible, we recall the precise result: for a proof, the reader is referred to [**6**, Proposition 8.1.8].

THEOREM B.0.1. *Let $\mu : [0,1] \to \mathscr{P}(\mathbb{R}^N)$ be a narrowly continuous curves solving the continuity equation*

$$\partial\mu_t + \mathrm{div}_x(v_t\,\mu_t) = 0,$$

where v_t is a Borel vector field satisfying

$$\int_0^1 \int_{\mathbb{R}^N} |v_t(x)|\, d\mu_t(x)\, dt < \infty \quad \text{and} \quad \int_0^1 \big(\|v_t\|_{L^\infty(\mathcal{K})} + \|\nabla v_t\|_{L^\infty(\mathcal{K})}\big)\, dt < \infty,$$

for every compact set $\mathcal{K} \subset \mathbb{R}^N$. Then for μ_0-a.e. $x \in \mathbb{R}^N$ the Cauchy problem

$$\begin{cases} \sigma'(t) &= v_t(\sigma(t)), \\ \sigma(0) &= x, \end{cases}$$

admits a globally defined solution $X(t,x)$ in $[0,1]$ and

$$\mu_t = (X(t,\cdot))_\sharp\mu_0, \quad \textit{for every } t \in [0,1].$$

APPENDIX C

Basic estimates for the $q-$Laplacian operator

1. Basic inequalities

We collect here some pointwise inequalities, which are particularly useful when dealing with the q-Laplacian operator or similar ones (for example the operator treated in Chapters 8 and 9): we confine ourselves to the case $q \geq 2$.

LEMMA C.1.1. *Let $q \geq 2$, then for every $z, w \in \mathbb{R}^N$ we get*

$$|z-w|^q \leq 2^{q-1}\langle |z|^{q-2}z - |w|^{q-2}w, z-w\rangle. \tag{C.1.1}$$

PROOF. A direct calculation shows that the right-hand side in (C.1.1) can be written as

$$\langle |z|^{q-2}z - |w|^{q-2}w, z-w\rangle = \frac{|z|^{q-2}+|w|^{q-2}}{2}|z-w|^2 + \frac{\left(|z|^{q-2}-|w|^{q-2}\right)\left(|z|^2-|w|^2\right)}{2}$$

and observing that the second term has positive sign, we get

$$\langle |z|^{q-2}z - |w|^{q-2}w, z-w\rangle \geq \frac{|z|^{q-2}+|w|^{q-2}}{2}|z-w|^2.$$

Then we go on with

$$|z-w|^{q-2} \leq (|z|+|w|)^{q-2} \leq \max\{1, 2^{q-3}\}\left(|z|^{q-2}+|w|^{q-2}\right),$$

which enables us to conclude, using the simple fact that $\max\{1, 2^{q-3}\} \leq 2^{q-2}$. Observe that with this proof, the constant 2^{q-1} that we obtain in (C.1.1) is not optimal (it is enough to test it with $q = 2$). □

REMARK C.1.2. By means of (C.1.1) and Cauchy-Schwarz inequality, we obtain

$$|z-w|^q \leq 2^q \left||z|^{\frac{q-2}{2}}z - |w|^{\frac{q-2}{2}}w\right|^2. \tag{C.1.2}$$

Indeed, (C.1.1) implies

$$|z-w|^q \leq 2^{q-1}|z-w|\left||z|^{q-2}z - |w|^{q-2}w\right|,$$

that is

$$|z-w|^{q-1} \leq 2^{q-1}\left||z|^{q-2}z - |w|^{q-2}w\right|,$$

and replacing q with $(q+2)/2$ and raising at the power 2, we obtain (C.1.2).

LEMMA C.1.3. *Let $q \geq 2$, then for every $z, w \in \mathbb{R}^N$ we get*

$$\left||z|^{\frac{q-2}{2}}z - |w|^{\frac{q-2}{2}}w\right|^2 \leq \frac{q^2}{4}\langle |z|^{q-2}z - |w|^{q-2}w, z-w\rangle \tag{C.1.3}$$

PROOF. We start with the simple observation

$$\begin{aligned}|z|^{q-2}z-|w|^{q-2}w &= \int_0^1 \frac{d}{dt}\left[|(1-t)w+tz|^{q-2}((1-t)w+tz)\right]\,dt\\ &= (z-w)\int_0^1 |(1-t)w+tz|^{q-2}\,dt\\ &+ (q-2)\int_0^1 |(1-t)w+tz|^{q-4}\langle(1-t)w+tz,z-w\rangle\,((1-t)w+tz)\,dt,\end{aligned}$$

so that

$$\begin{aligned}\langle|z|^{q-2}z-|w|^{q-2}w,z-w\rangle &= |z-w|^2\int_0^1 |(1-t)w+tz|^{q-2}\,dt\\ &+ (q-2)\int_0^1 |(1-t)w+tz|^{q-4}\langle(1-t)w+tz,z-w\rangle^2\,dt.\end{aligned} \tag{C.1.4}$$

As it is easily seen, the second term in the right-hand side is positive, so that

$$\langle|z|^{q-2}z-|w|^{q-2}w,z-w\rangle \geq |z-w|^2\int_0^1 |(1-t)w+tz|^{q-2}\,dt \tag{C.1.5}$$

Moreover observe that we have

$$\left||z|^{q-2}z-|w|^{q-2}w\right| \leq (q-1)|z-w|\int_0^1 |(1-t)w+tz|^{q-2}\,dt,$$

then, replacing q with $(q+2)/2$ (which is still an exponent ≥ 2) and raising at the power 2, we get

$$\begin{aligned}\left||z|^{\frac{q-2}{2}}z-|w|^{\frac{q-2}{2}}w\right|^2 &\leq \frac{q^2}{4}|z-w|^2\left(\int_0^1 |(1-t)w+tz|^{\frac{q-2}{2}}\,dt\right)^2\\ &\leq \frac{q^2}{4}|z-w|^2\int_0^1 |(1-t)w+tz|^{q-2}\,dt\end{aligned}$$

and finally using (C.1.5), we get the desired inequality. □

LEMMA C.1.4. *Let $q \geq 2$, then for every $z,w \in \mathbb{R}^N$ we get*

$$\left||z|^{q-2}z-|w|^{q-2}w\right| \leq (q-1)\left(|z|^{\frac{q-2}{2}}+|w|^{\frac{q-2}{2}}\right)\left||z|^{\frac{q-2}{2}}z-|w|^{\frac{q-2}{2}}w\right| \tag{C.1.6}$$

PROOF. With the same argument of the previous proof, we arrive at

$$\left||z|^{q-2}z-|w|^{q-2}w\right| \leq (q-1)|z-w|\int_0^1 |(1-t)w+tz|^{q-2}\,dt,$$

and then we observe that

$$|(1-t)w+tz|^{q-2} = |(1-t)w+tz|^{\frac{q-2}{2}}|(1-t)w+tz|^{\frac{q-2}{2}} \leq |(1-t)w+tz|^{\frac{q-2}{2}}\left(|z|^{\frac{q-2}{2}}+|w|^{\frac{q-2}{2}}\right),$$

which yields

$$\left||z|^{q-2}z-|w|^{q-2}w\right| \leq (q-1)|z-w|\left(|z|^{\frac{q-2}{2}}+|w|^{\frac{q-2}{2}}\right)\int_0^1 |(1-t)w+tz|^{\frac{q-2}{2}}\,dt. \tag{C.1.7}$$

To conclude, it is enough to observe that (C.1.5) with the exponent q replaced by $(q+2)/2$, implies

$$\left||z|^{\frac{q-2}{2}}z-|w|^{\frac{q-2}{2}}w\right|\geq|z-w|\int_0^1|(1-t)w+tz|^{\frac{q-2}{2}}\,dt,$$

so that inserting this into (C.1.7), we finally end up with (C.1.6). □

2. Higher differentiability for solutions of the $q-$Laplacian

In this section, we consider $q>2$ and with the term *local weak solution* of the equation $-\Delta_q u=f$ we mean $u\in W^{1,q}_{\rm loc}(\Omega)$ such that

$$\int_\Omega\langle|\nabla u(x)|^{q-2}\nabla u(x),\nabla\varphi(x)\rangle\,dx=\int_\Omega f(x)\,\varphi(x)\,dx,\quad\text{for every }\varphi\in C^\infty_c(\Omega).$$

Theorem C.2.1. *Let $u\in W^{1,q}_{\rm loc}(\Omega)$ be a local weak solution of*

$$-\Delta_q u=f,$$

with $f\in L^p_{\rm loc}(\Omega)$ and $p=q/(q-1)$. Then

$$\mathcal{V}(x):=|\nabla u(x)|^{\frac{q-2}{2}}\nabla u(x)\in W^{\frac{p}{2}-\tau,2}_{\rm loc}(\Omega),$$

for every $\tau>0$.

Proof. Let us set $\mathcal{H}(x)=|\nabla u(x)|^{q-2}\nabla u(x)$, then we have

$$\int_\Omega\langle\mathcal{H}(x),\nabla\varphi(x)\rangle\,dx=\int_\Omega f(x)\varphi(x)\,dx\tag{C.2.1}$$

and

$$\int_\Omega\langle\mathcal{H}(x_h),\nabla\varphi(x)\rangle\,dx=\int_\Omega f(x_h)\varphi(x)\,dx\tag{C.2.2}$$

where $x_h=x+he_k$ (we omit the dependence on $k\in\{1,\dots,N\}$). We set

$$\delta_h u(x)=\frac{u(x_h)-u(x)}{h},$$

then subtracting (C.2.2) to (C.2.1) and dividing by h, we obtain

$$\int_\Omega\langle\delta_h\mathcal{H}(x),\nabla\varphi(x)\rangle\,dx=\int_\Omega\delta_h f(x)\varphi(x)\,dx.$$

Inserting the test function $\varphi=\xi^2\delta_h u$ in the previous equality, where ξ is the usual cut-off function, supported on $B_R(x_0)\Subset\Omega$ and $\xi\equiv1$ on a smaller concentric ball $B_\varrho(x_0)$, with $|\nabla\xi|\simeq(R-\varrho)^{-1}$, we get

$$\begin{aligned}\int_\Omega\xi^2(x)\langle\delta_h\mathcal{H}(x),\nabla\delta_h u(x)\rangle\,dx&=-2\int_\Omega\xi(x)\langle\delta_h\mathcal{H}(x),\nabla\xi(x)\rangle\delta_h u(x)\,dx\\&\quad+\int_\Omega\xi^2(x)\delta_h f(x)\delta_h u(x)\,dx.\end{aligned}$$

We then observe that $\nabla\delta_h u = \delta_h \nabla u$ and using (C.1.3), we can estimate the left-hand side as follows

$$C\int_\Omega \xi^2(x)|\delta_h\mathcal{V}(x)|^2 \le \int_\Omega \xi^2(x)\langle\delta_h\mathcal{H}(x),\nabla\delta_h u(x)\rangle\,dx$$

while for the first term in the right-hand side we can use Cauchy-Schwarz and (C.1.6), in order to get

$$-2\int_\Omega \xi(x)\langle\delta_h\mathcal{H}(x),\nabla\xi(x)\rangle\delta_h u(x)\,dx \le C\int_\Omega \xi(x)|\delta_h\mathcal{G}(x)|\left(|\mathcal{G}(x_h)|^{\frac{q-2}{q}}+|\mathcal{G}(x)|^{\frac{q-2}{q}}\right)\times|\nabla\xi(x)||\delta_h u(x)|\,dx = I_1.$$

Concerning the second term in the right-hand side, we can use

$$\left|\int_\Omega \xi^2(x)\,\delta_h f(x)\,\delta_h u(x)\,dx\right| \le \int_\Omega |\omega(x)|\left|\frac{\partial}{\partial x_k}\left(\xi^2(x)\delta_h u(x)\right)\right|\,dx = I_2,$$

where ω is defined by

$$\omega(x) = \int_0^1 f(x+the_k)\,dt.$$

Let us begin estimating integral I_1: by means of Young's inequality we get

$$\xi(x)|\delta_h\mathcal{V}(x)|\left(|\mathcal{V}(x_h)|^{\frac{q-2}{q}}+|\mathcal{V}(x)|^{\frac{q-2}{q}}\right)|\nabla\xi(x)||\delta_h u(x)| \le \varepsilon|\delta_h\mathcal{V}(x)|^2\xi(x)^2 +\frac{1}{\varepsilon}\left(|\mathcal{V}(x_h)|^{\frac{q-2}{q}}+|\mathcal{V}(x)|^{\frac{q-2}{q}}\right)^2|\delta_h u(x)|^2|\nabla\xi(x)|^2,$$

and we observe that using Hölder's inequality, we get

$$\int_\Omega\left(|\mathcal{V}(x_h)|^{\frac{q-2}{q}}+|\mathcal{V}(x)|^{\frac{q-2}{q}}\right)^2|\delta_h u(x)|^2|\nabla\xi(x)|^2\,dx \le C\left(\int_\Omega|\delta_h u(x)|^q|\nabla\xi(x)|^q\,dx\right)^{\frac{2}{q}}\times\left(\int_\Omega|\mathcal{V}(x)|^2\,dx\right)^{\frac{q-2}{q}},$$

and in the end, taking ε small enough, we can obtain

$$\int_\Omega \xi^2(x)|\delta_h\mathcal{V}(x)|^2 \le C\left(\int_\Omega|\delta_h u(x)|^q|\nabla\xi(x)|^q\,dx\right)^{\frac{2}{q}}\left(\int_\Omega|\mathcal{V}(x)|^2\,dx\right)^{\frac{q-2}{q}}+I_2.$$

It is only left to estimate integral I_2:

$$\left|\frac{\partial}{\partial x_k}\left(\xi^2(x)\delta_h u(x)\right)\right| \le \xi(x)|\nabla\xi(x)||\delta_h u(x)|+\xi^2(x)|\delta_h\nabla u(x)|,$$

so that

$$I_2 \le \int_\Omega \xi(x)|\omega(x)||\nabla\xi(x)||\delta_h u(x)|\,dx+\int_\Omega\xi^2(x)|\omega(x)||\delta_h\nabla u(x)|\,dx.$$

We now proceed as follows: we use Hölder's inequality and (C.1.2), so to obtain

$$\begin{aligned}\int_\Omega \xi^2(x)|\omega(x)||\delta_h \nabla u(x)|\,dx &\le \left(\int_\Omega \xi^2 |\omega(x)|^p\,dx\right)^{\frac{1}{p}} \left(\int_\Omega \xi^2\,|\delta_h \nabla u(x)|^q\,dx\right)^{\frac{1}{q}}\\ &\le C\,h^{\frac{2-q}{q}} \left(\int_\Omega \xi^2 |\omega(x)|^p\,dx\right)^{\frac{1}{p}} \left(\int_\Omega \xi^2\,|\delta_h \mathcal{V}(x)|^2\,dx\right)^{\frac{1}{q}}\\ &\le C\,\frac{h^{\frac{2-q}{q-1}}}{\varepsilon}\int_\Omega \xi^2|\omega(x)|^p\,dx + \varepsilon\int_\Omega \xi^2\,|\delta_h\mathcal{V}(x)|^2\,dx,\end{aligned}$$

so that, for a suitable choice of ε, the second term can be absorbed in the left-hand side. Putting all together, up to now we have shown the following

$$\begin{aligned}\int_{B_\varrho} |\delta_h \mathcal{V}(x)|^2\,dx \le \frac{C}{(R-\varrho)^2}&\left(\int_{B_R}|\nabla u(x)|^q\ dx\right)^{\frac{2}{q}} \left(\int_\Omega |\mathcal{V}(x)|^2\ dx\right)^{\frac{q-2}{q}}\\ &+\frac{C}{R-\varrho}\left(\int_{B_R}|\nabla u(x)|^q\,dx\right)^{\frac{1}{q}}\left(\int_\Omega |f(x)|^p\,dx\right)^{\frac{1}{p}}\\ &+h^{\frac{2-q}{q-1}}\int_\Omega |f(x)|^p\,dx,\end{aligned}$$

so that multiplying both sides by the term $h^{\frac{q-2}{q-1}}$ and using the fact that $h^{\frac{q-2}{q-1}} \le 1$, thanks to the fact that $q > 2$, then we obtain

$$\int_{B_\varrho}\left|\frac{\mathcal{V}(x+h\,e_k)-\mathcal{V}(x)}{h^{\frac{p}{2}}}\right|^2\,dx \le C,$$

which implies that $\mathcal{V}$ belongs to the *Nikolskii space*[1] $\mathscr{N}^{\frac{p}{2},2}$. Then we conclude by means of the imbedding (see [**1**])

$$\mathscr{N}^{\frac{p}{2},2}(\Omega) \hookrightarrow W^{\frac{p}{2}-\tau,2}(\Omega),$$

which holds true for every $\tau > 0$. □

Remark C.2.2. Using the Sobolev imbedding Theorem for fractional Sobolev spaces (see [**69**, Theorem 2.1]), we obtain as a corollary of the previous result

$$\mathcal{V} \in L^{\frac{2N}{N-p}-\tau}_{\rm loc}(\Omega),\quad \text{for every } \tau > 0,$$

[1]Given $\beta \in (0,1)$ and $s \in [1,\infty)$, for every $\varepsilon > 0$ we define $\Omega_\varepsilon = \{x \in \Omega\,:\,\mathrm{dist}(x,\partial\Omega) \ge \varepsilon\}$. Then $u \in \mathscr{N}^{\beta,s}(\Omega)$ if and only if $u \in L^s(\Omega)$ and

$$\sum_{i=1}^N \sup_{\substack{\varepsilon>0,\\ 0<|h|<\varepsilon}} \left(\int_{\Omega_\varepsilon}\frac{|u(x+h\,e_i)-u(x)|^s}{|h|^{\beta s}}\,dx\right)^{1/s} < +\infty.$$

and using the fact that $\phi = |\nabla u|^{q-2}\nabla u = |\mathcal{V}|^{\frac{q-2}{q}}\mathcal{V}$, we also obtain

$$\phi \in L_{\text{loc}}^{\frac{2Nq}{(N-p)(q-2)}-\tau}(\Omega), \quad \text{for every } \tau > 0,$$

which can be read as a higher integrability result for the solution of Beckmann's problem corresponding to the cost $\mathcal{H}(\phi) = 1/p\,|\phi|^p$, with $p < 2$ and $\rho_0, \rho_1 \in L^p(\Omega)$.

Remark C.2.3. Observe that letting q approach 2, from the result of Theorem C.2.1 we recover the classical higher differentiability result

$$f \in L^2_{\text{loc}} \Longrightarrow \nabla u \in W^{1,2}_{\text{loc}}(\Omega),$$

for local weak solutions of $-\Delta u = f$. In this latter case, remembering the Calderon-Zygmund estimates ([**55**]), a better integrability of the datum f affects the integrability of the derivatives of the solution, i.e.

$$f \in L^s_{\text{loc}} \Longrightarrow \nabla u \in W^{1,s}_{\text{loc}}(\Omega),$$

for $s \in (1, \infty)$, but roughly speaking, it does not give *more derivatives*. This should tell that the gain of a fractional order of derivation on the term $\mathcal{V}(x) = |\nabla u(x)|^{\frac{q-2}{2}}\nabla u(x)$ is the better that one can hope for, as far as f acts on the scale of Lebesgue spaces. A tuning of the summability exponent of f can only give a gain of summability of these fractional derivatives: clearly, the proof of this fact should be much more involved and can not be achieved by means of integrated differential quotients. Anyway, the interested reader is warmly suggested to consult [**69**] and the references therein on these delicate topics.

Index

Bibliography

This thesis is based on the following papers:

[B1] L. Brasco, Global L^∞ gradient estimates for solutions to a certain degenerate elliptic equation, to appear on Nonlinear Anal. (2010), available at `http://cvgmt.sns.it/people/brasco`
[B2] L. Brasco, Curves of minimal action over metric spaces, Ann. Mat. Pura Appl., **189** (2010), 95–125.
[B3] L. Brasco, G. Buttazzo, F. Santambrogio, A Benamou-Brenier approach to branched transport, Submitted (2010), available at `http://cvgmt.sns.it/people/brasco`
[B4] L. Brasco, G. Carlier, F. Santambrogio, Congested traffic dynamics, weak flows and very degenerate elliptic equations, J. Math. Pures Appl., **93** (2010), 652–671.
[B5] L. Brasco, F. Santambrogio, An equivalent path functional formulation of branched transportation problems, to appear on Discrete Contin. Dyn. Syst. (2010), available at `http://cvgmt.sns.it/people/brasco`

Other references:

[1] R. A. Adams, Sobolev spaces, Pure and Applied Mathematics, **65**, Academic Press, New York-London, 1975.
[2] L. Ambrosio, Transport equation and Cauchy problem for BV vector fields, Invent. Math., **158** (2004), 227–260.
[3] L. Ambrosio, Lecture notes on optimal transport problems, in *Mathematical aspects of evolving interfaces (Funchal, 2000)*, Lecture Notes in Math. **1812**, Springer, Berlin, 2003.
[4] L. Ambrosio, Metric space valued functions of bounded variation, Ann. Scuola Norm. Sup. Pisa Cl. Sci., **17** (1990), 439–478.
[5] L. Ambrosio, G. Crippa, Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields, in *Transport Equations and Multi-D Hyperbolic Conservation Laws*, Lecture Notes of the Unione Matematica Italiana, 2008.
[6] L. Ambrosio, N. Gigli, G. Savarè, Gradient flows in metric spaces and in the space of probability measure. Second edition. Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.
[7] L. Ambrosio, A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation, in *Optimal transportation and applications*, 122-160, Lecture Notes in Math., **1813**, Springer, Berlin, 2003.
[8] L. Ambrosio, F. Santambrogio, Necessary optimality conditions for geodesics in weighted Wasserstein spaces, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei – Mat. Appl., **18** (2007), 23–37.
[9] L. Ambrosio, P. Tilli, Selected topics on "Analysis in metric spaces", Appunti dei corsi tenuti da docenti della Scuola, Scuola Normale Superiore, Pisa, 2000.
[10] E. N. Barron, R. R. Jensen, C. Y. Wang, Lower semicontinuity of L^∞ functionals, Ann. Inst. H. Poincaré Anal. Non Lineaire, **18** (2001), 495–517.
[11] M. J. Beckmann, A continuous model of transportation, Econometrica, **20** (1952), 643–660.
[12] M. J. Beckmann, C. McGuire, C. Winsten, Studies in Economics of Transportation. Yale University Press, New Haven, 1956.
[13] J.-D. Benamou, Y. Brenier, A computational Fluid Mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math., **84** (2000), 375–393.
[14] F. Benmansour, G. Carlier, G. Peyré, F. Santambrogio, Numerical approximation of continuous traffic congestion equilibria, Netw. Heterog. Media, **4** (2009), 605–623.
[15] M. Bernot, V. Caselles, J.-M. Morel, Traffic plans, Publ. Mat., **49** (2005), 417–451.

[16] M. Bernot, V. Caselles, J.-M. Morel, The structure of branched transportation networks, Calc. Var. Partial Differential Equations, **32** (2008), 279–317.
[17] M. Bernot, V. Caselles, J.-M. Morel, Optimal transportation networks – Models and theory, Lecture Notes in Math., **1955**, Springer-Verlag, Berlin, 2009.
[18] M. Bernot, A. Figalli, Synchronized traffic plans and stability of optima, ESAIM Control Optim. Calc. Var., **14** (2008), 864–878.
[19] S. Bianchini, A. Brancolini, Estimates on path functionals over Wasserstein spaces, SIAM J. Math. Anal., **42** (2010), 1179–1217.
[20] B. Bojarski, T. Iwaniec, p-harmonic equation and quasiregular mappings, Partial Differential Equations (Warsaw 1984), Banach Center Publications, **19** (1987).
[21] G. Bouchitté, G. Buttazzo, New lower semicontinuity results for nonconvex functionals defined on measures, Nonlinear Anal., **15** (1990), 679–692.
[22] G. Bouchitté, C. Jimenez, M. Rajesh, A new L^∞ estimate in optimal transport, Proc. Amer. Math. Soc., **135** (2007), 3525–3535.
[23] G. Bouchitté, P. Seppecher, paper in preparation.
[24] A. Brancolini, G. Buttazzo, F. Santambrogio, Path functionals over Wasserstein spaces, J. Eur. Math. Soc., **8** (2006), 415–434.
[25] A. Brancolini, S. Solimini, On the Hölder regularity of the landscape function, Preprint (2009), available at `http://cvgmt.sns.it/people/brancolin`.
[26] Y. Brenier, The least action principle and the related concept of generalized flows for incompressible perfect fluids, J. Amer. Math. Soc., **2** (1989), 225–255.
[27] Y. Brenier, Extended Monge-Kantorovich theory, in *Optimal transportation and applications*, 91–121, Lecture Notes in Math., **1813**, Springer, Berlin, 2003.
[28] H. Brezis, Functional analysis. Theory and applications (French), Masson, Paris, 1983.
[29] G. Buttazzo, Semicontinuity, relaxation and integral representation in the calculus of variations, Pitman Research Notes in Mathematics Series, **207**, Longman Scientific & Technical, Harlow, 1989.
[30] G. Buttazzo, C. Jimenez, E. Oudet, An optimization problem for mass transportation with congested dynamics, SIAM J. Control Optim., **48** (2009), 1961-1976.
[31] L. Caffarelli, M. Feldman, R. J. McCann, Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs, J. Amer. Math. Soc., **15** (2002), 1–26.
[32] L. Caravenna, A proof of Sudakov theorem with strictly convex norms, to appear on Math. Z., DOI: `10.1007/s00209-010-0677-6`
[33] G. Carlier, C. Jimenez, F. Santambrogio, Optimal transportation with traffic congestion and Wardrop equilibria, SIAM J. Control Optim., **47** (2008), 1330–1350.
[34] G. Carlier, F. Santambrogio, A variational model for urban planning with traffic congestion, ESAIM Control Optim. Calc. Var., **11** (2007), 595–613.
[35] C. Carstensen, S. Müller, Local stress regularity in scalar nonconvex variational problems, SIAM J. Math. Anal., **34** (2002), 495–509.
[36] T. Champion, L. De Pascale, The Monge problem for strictly convex norms in $\mathbb{R}^d$, J. Eur. Math. Soc., **12** (2010), 1355–1369.
[37] B. Dacorogna, J. Moser, On a partial differential equation involving the Jacobian determinant, Annales del'I.H.P., **7** (1990), 1–26.
[38] C. Dellacherie, P.-A. Meyer, Probabilités et potentiel. (French) Chapitres I à IV. Édition entièrement refondue. Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV. Actualités Scientifiques et Industrielles, **1372**, Hermann, Paris, 1975.
[39] J. L. Denny, The support of discrete extremal measures with given marginals, Michigan Math. J., **27** (1980), 59–64.
[40] L. De Pascale, L. C. Evans, A. Pratelli, Integral estimates for transport densities, Bull. London Math. Soc., **36** (2004), 383–395.
[41] L. De Pascale, A. Pratelli, Regularity properties for Monge transport density and for solutions of some shape optimization problems, Calc. Var. Partial Differential Equations, **14** (2002), 249–274.

[42] L. De Pascale, A. Pratelli, Sharp summability for Monge transport density via interpolation, ESAIM Control Optim. Calc. Var., **10** (2004), 549–552.
[43] G. Devillanova, S. Solimini, Elementary properties of optimal irrigation patterns, Calc. Var. Partial Differential Equations, **28** (2007), 317–349.
[44] J. Dolbeault, B. Nazaret, G. Savaré, A new class of transport distances between measures, Calc. Var. Partial Differential Equations, **34** (2009), 193–231.
[45] E. DiBenedetto, $C^{1+\alpha}$ local regularity of weak solutions of degenerate elliptic equations, Nonlinear Anal., **7** (1983), 827-850.
[46] R. J. DiPerna, P.-L. Lions, Ordinary differential equations, transport theory and Sobolev spaces, Invent. Math., **98** (1989), 511–547.
[47] I. Ekeland, R. Temam, Convex analysis and variational problems, Classics in Applied Mathematics, **28**, SIAM, Philadelphia, PA, 1999.
[48] L. C. Evans, W. Gangbo, Differential equations methods for the Monge-Kantorovich mass transfer problem, Mem. Amer. Math. Soc., **137**, 653 (1999).
[49] H. Federer, Geometric measure theory, Die Grundlehren der Mathematischen Wissenschaften, **153**, Springer-Verlag New York Inc., New York, 1969.
[50] M. Feldman, R. McCann, Uniqueness and transport density in Monges's mass transportation problem, Calc. Var. Partial Differential Equations, **15** (2004), 81–113.
[51] M. M. Flood, On the Hitchcock distribution problem, Pacific J. Math., **3** (1953), 369–386.
[52] I. Fonseca, N. Fusco, P. Marcellini, An existence result for a nonconvex variational problem via regularity, ESAIM Control Optim. Calc. Var., **7** (2002), 69–95.
[53] W. Gangbo, T. Nguyen, A. Tudorascu, Euler-Poisson systems as action-minimizing paths in the Wasserstein space, Arch. Ration. Mech. Anal., **192** (2009), 419–452.
[54] F. W. Gehring, The L^p integrability of the partial derivatives of a quasiconformal mapping, Acta Math., **130** (1973), 265–277.
[55] D. Gilbarg and N. Trudinger, Elliptic Partial Differential Equations of Second Order, Springer Verlag, Berlin, 1977.
[56] E. N. Gilbert, Minimum cost communication networks, Bell System Tech. J., **46** (1967), 2209–2227.
[57] E. Giusti, Direct methods in the calculus of variations (Italian), Unione Matematica Italiana, Bologna, 1994.
[58] P. Grisvard, Elliptic problems in nonsmooth domains. Monographs and Studies in Mathematics, **24**. Pitman (Advanced Publishing Program), Boston, MA, 1985.
[59] F. L. Hitchcock, The distribution of a product from several sources to numerous localities, J. Math. Phys. Mass. Inst. Tech., **20** (1941), 224–230.
[60] A. D. Ioffe, On lower semicontinuity of integral functionals. I. II., SIAM J. Cont. and Opt., **15** (1977), 521-538 and 991–1000.
[61] L. V. Kantorovich, On the translocation of masses, C. R. Dokl. Acad. Sci. USSR, **321** (1942), 199–201; English translation J. Math. Sci., **133** (2006), 1381–1382.
[62] O. A. Ladyzhenskaya, N. N. Ural'tseva, Linear and quasilinear elliptic equations, Academic Press, New York and London, 1968.
[63] J. L. Lewis, Regularity of the derivatives of solutions to certain degenerate elliptic equations, Indiana Univ. Math. Jour., **32** (1983), 849–858.
[64] G. M. Lieberman, Boundary regularity for solutions of degenerate elliptic equations, Nonlinear Anal., **12** (1988), 1203–1219.
[65] S. Lisini, Characterization of absolutely continuous curves in Wasserstein spaces, Calc. Var. Partial Differential Equations, **28** (2007), 85–120.
[66] R. J. McCann, A convexity principle for interacting gases, Adv. Math., **128** (1997), 153–179.
[67] F. Maddalena, J.-M. Morel, S. Solimini, A variational model of irrigation patterns, Interfaces Free Bound., **5** (2003), 391–415.
[68] F. Maddalena, S. Solimini, Transport distances and irrigation models, J. Convex Anal., **16** (2009), 121–152.
[69] G. Mingione, The Calderón-Zygmund theory for elliptic problems with data measure, Ann. Sc. Norm. Sup., **6** (2007), 195–261.

[70] J.-M. Morel, F. Santambrogio, Comparison of distances between measures, Appl. Math. Lett., **20** (2007), 427–432.

[71] J. Moser, On the volume elements on a manifold, Trans. Am. Math. Soc., **120** (1965), 286–294.

[72] J. Moser, A new proof of De Giorgi's Theorem concerning the regularity problem for elliptic differential equations, Comm. Pure Appl. Math., **13** (1960), 457–468.

[73] E. Paolini, E. Stepanov, Optimal transportation networks as flat chains, Interfaces Free Bound., **8** (2006), 393–436.

[74] T. Rivière, D. Ye, Resolutions of the prescribed volume form equation, NoDEA Nonlinear Differential Equations Appl., **3** (1996), 323–369.

[75] F. Santambrogio, Absolute continuity and summability of transport densities: simpler proofs and new estimates, Calc. Var. Partial Differential Equations, **36** (2009), 343–354.

[76] F. Santambrogio, Variational problems in transport theory with mass concentration, Tesi. Scuola Normale Superiore di Pisa (Nuova Serie) **4**, Edizioni della Normale, Pisa, 2007.

[77] F. Santambrogio, Optimal channel networks, landscape function and branched transport, Interfaces Free Bound., **9** (2007), 149–169.

[78] F. Santambrogio, V. Vespri, Continuity for a very degenerate elliptic equation in two dimensions, Nonlinear Anal., **73** (2010), 3832–3841.

[79] G. Strang, L^1 and L^∞ approximation of vector fields in the plane, Lecture Notes in Num. Appl. Analysis, **5** (1982), 273–288.

[80] E. W. Stredulinsky, Weighted inequalities and degenerate elliptic partial differential equations, Lecture Notes in Math., **1074**, Springer-Verlag, Berlin, 1984.

[81] E. W. Stredulinsky, Higher integrability from reverse Hölder inequalities, Indiana Univ. Math. J., **29** (1980), 407–413.

[82] V. N. Sudakov, Geometric problems in the theory of infinite-dimensional probability distributions, Proc. Steklov Inst. Math. 1979, no. 2, i–v, 1–178.

[83] G. Talenti, Best constant in Sobolev inequality, Ann. Mat. Pura Appl., **110** (1976), 353–372.

[84] N. S. Trudinger, X.-J. Wang, On the Monge mass transfer problem, Calc. Var. Partial Differential Equations, **13** (2001), 19–31.

[85] C. Villani, Optimal transport. Old and new, Die Grundlehren der Mathematischen Wissenschaften, **338**. Springer-Verlag, Berlin, 2009.

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

[87] N. N. Vorob'ev, Game Theory. Lectures for Economists and Systems Scientists, Applications of Mathematics, **7**, Springer-Verlag, New York-Berlin, 1977.

[88] J. G. Wardrop, Some theoretical aspects of road traffic research, Proc. Inst. Civ. Eng., **2** (1952), 325–378.

[89] Q. Xia, Interior regularity of optimal transport paths, Calc. Var. Partial Differential Equations, **20** (2004), 283–299.

[90] Q. Xia, Optimal paths related to transport problems, Commun. Contemp. Math., **5** (2003), 251–279.