PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Université Paris Dauphine

Optimization in spaces of measures: Optimal Transport, Geometric Structures and Game Theory

Soutenue par

João Miguel Machado

Le 17 Septembre 2024

École doctorale n°543

École doctorale SDOSE

Spécialité

Mathématiques

Composition du jury :

Annalisa Massaccesi Associate Professeur Università di Padova	*Rapporteure*
Antoine Lemenant Professeur Université de Lorraine	*Rapporteur*
Filippo Santambrogio Professeur Université Claude Bernard - Lyon 1	*Examinateur*
Frédéric De Gournay Maître de Conférences Institut de Mathématiques de Toulouse	*Examinateur*
Yana Teplitskaya Université Paris-Saclay, LMO	*Examinatrice*
Antonin Chambolle Directeur de recherche CNRS & Université Paris Dauphine, PSL	*Directeur de thèse*
Vincent Duval Directeur de recherche INRIA & Université Paris Dauphine, PSL	*Directeur de thèse*

Dauphine UNIVERSITÉ PARIS | PSL

Acknowledgments

I can not express how impactful these last years during my thesis have been in my life in terms of professional, personal and intellectual growth. It was a time of great fulfillment and also of frustration whenever the mathematics just did not work. Whenever I found myself in one of those moments in my office at Dauphine on the second floor, I would go to the hall and stare at OTAN's motto stamped on the ceiling: "*Animus in consulendo liber*". It seems that the best translation from latin is the French one "*L'esprit est libre dans la consultation*". This must have been the most valuable lesson from this period: no matter how smart or how much we master a field, our minds are always dragged by our personal biases, and we are only truly free upon averaging this noise out by consulting with our peers and mentors.

This lesson may seem a bit cliché, but I believe we can only learn it after much practice. I thank my advisors Antonin Chambolle and Vincent Duval for their patience and being catalysts for me to learn this lesson, which I am still doing, through our numerous discussions. I hope to have learned from Antonin how to think of creative solutions, exploiting a geometric intuition, and always try to approach a problem from different perspectives, while Vincent taught me to always be very meticulous about every claim, justifying them with transparent reasoning and basic principles, as a true Stoic sage.

To all the members of the jury: Filippo Santambrogio, Frédéric de Gournay, Yana Teplitskaya and specially Antoine Lemenant and Annalisa Massaccesi, who very kindly accepted to be rapporteurs of this thesis, I thank you for letting me share my work with you.

To all the people that enabled the start of this PhD. To Maria Soledad Aronna, who has been a mentor since the beginning of my path in mathematics. Without her I would never have come to France and had this amazing experience. And to all the amazing professors I had in Orsay during my first year in France, specially Quentin Mérigot, Laurent Pfeiffer, Luca Nenna and Jean-François Babadjian, from whom I learned all the basis that lead to this thesis and also made flourish my appetite for the Calculus of Variations.

Thanks to all the people I have collaborated during this thesis. To Idriss Mazari and Paul Pegon, who taught me a lot and the for the delicious coffee we shared at Paul's office. To Guilherme Mazanti and Laurent Pfeiffer, thanks for always having me at Centrale Supelec. It is always nice to come back to the plateau, I hope our collaboration will still give many fruits.

To all my colleagues at Dauphine and Mokaplan, I thank you for the times we've spent together and everything you have taught me. To the members of the office B227: Charly, Lorenzo, Lucas, Théo, I will never forget all the kitty pictures that Charly would share with us daily. To Adrien, Daniele, Giovanni, Umberto, Tommy many thanks for all the moments we had together, either at Dauphine and the super fast queue from CROUS or outside.

To everyone at Inria and Mokaplan, thank you so much for the enormous privilege to be a part of this team. Thanks to all the permanents: Antonin, Derya, Flavien, Irène, Jean-David, Guillaume, Paul, Thomas (to who I wish the best of luck in the new PARMA team), Vincent and Yann. Thanks to the Mokabébés: Adrian, Gabriele, Guillaume, Romain, Tommy and specially to Hugo and Maxime, it was really a pleasure to co-organize the Mokameetings and share some beers along the way, I hope there will be many more to come.

Most important of all I thank my family. From all my heart I thank for the support of my parents and my brother, who always encourage me to pursue what was first just an interest in mathematics, which eventually became a passion and, by the time I got to France, my biggest goal was to turn it into a career and become the best mathematician that I could. This was my biggest goal until I met the last person I must thank, but in fact the one who deserves the most acknowledgment for the countless hours of support, my fiancé Rafaela. Thank you for your unconditional love, for hearing me when I claimed to have proved something new and twenty minutes later realizing that the proof was not quite there yet and for the amazing journey we have shared so far. Your are my biggest goal.

Contents

Introduction - English version

1. Why optimization in spaces of measures?

One can say that the Calculus of Variations has its birth in the antiquity, where the ancient Greeks had already formulated the first instances of the isoperimetric problem, which given all shapes of the plan with a fixed area, consists in finding the one with the smallest perimeter. There is some debate to whether this is considered the founding problem in the field as its first solutions were purely geometric, as opposed to other problems that appeared in the seventeenth century from the development of classical mechanics, which were formulated in an analytic terminology. The first and maybe most famous of them is the brachystochrone problem, first proposed by Galileo and later reformulated with a precise mathematical model by Bernoulli.[1]

Both of these are optimization problems, but this is not what makes them variational in nature, for instance Combinatorial Optimization share the same nature of searching for a minimal object and however such problems are quite far from our methods. The name Calculus of Variations was attributed by Euler after Lagrange's *method of variations*, see for instance the introduction of Goldstein's historical treatise on the subject [Goldstine, 1980]. At their time, the paradigm in physics was that nature does not waste energy and trajectories of particles are determined by the minimization of a notion of energy. They were then interested in problems of the form

$$\min\left\{\mathcal{L}(x) \stackrel{\text{def.}}{=} \int_0^T L(t, x(t), \dot{x}(t))\mathrm{d}t : x(0) = x_0,\ x(T) = x_T\right\}, \tag{0.1}$$

[1] Goldstein for instance [Goldstine, 1980], places the origins of the Calculus of Variations in the brachystochrone problem as it was the first variational problem to be solved through analytic methods instead of geometric.

the minimization taking place over all curves $x : [0, T] \to \mathbb{R}$ connecting points x_0 and x_T. The idea of Lagrange to compute a minimizer of this energy was to perturb the optimal curve as $x + \varepsilon h$, where h is a curve such that $h(0) = h(T) = 0$ called a variation. The endpoint constraints on h ensures that $x + \varepsilon h$ is still admissible for the minimization of $\mathcal{L}$ so that, as soon as x is a minimizer, $\varepsilon \mapsto \mathcal{L}(x + \varepsilon h)$ attains its minimum at $\varepsilon = 0$ and the derivative w.r.t. ε at $\varepsilon = 0$ is zero, which gives the famous Euler-Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}}(t, x, \dot{x}) = \frac{\partial L}{\partial x}(t, x, \dot{x}). \tag{0.2}$$

Notice that we have not made explicit in problem (0.1) what is the class of curves in which the minimization takes place, and indeed in the early days of the calculus of variations this was not explicitly stated. In practice admissible curves were assumed smooth, which suggest an implicit belief that nature is continuous (that is not necessarily true). This is intimately related with question of existence of minimizers, and indeed the question of existence was not addressed in the early days of the field. It was only in 1915 that Tonelli suggested a result of existence for a wide class of energies among the space of absolutely continuous functions, see for instance the discussion in [Clarke, 2013, Chap. 16].

The main idea of Tonelli's proof is now known as the *Direct Method of the Calculus of Variations* and has proven to be sufficiently flexible to be applied in much more general settings such as minimization problems in metric spaces. To state it in a modern flavor, let (X, d) be a metric space, $F : X \to \mathbb{R} \cup \{+\infty\}$ be a functional such that

- F is lower semi-continuous, *i.e.* for every sequence $x_n \xrightarrow[n\to\infty]{d} x$, it holds that
$$F(x) \leq \liminf_{n\to\infty} F(x_n)$$
- F has compact level sets, *i.e.* the sets $\{x \in X : F(x) \leq \ell\}$ are compact.

Then, supposing that $\inf_{x\in X} F(x) < +\infty$, and letting $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence, that is a sequence such that $F(x_n)$ converges to $\inf_X F$, since F has compact level sets, $(x_n)_{n\in\mathbb{N}}$ admits a subsequence converging to some x. It then follows from the lower semi-continuity of F that

$$F(x) \leq \liminf_{n\to\infty} F(x_n) = \inf_X F,$$

and x is a minimizer.

This method is now standard and possibly the most popular technique to show existence of solutions to variational problems, hence it is most interesting when it does not work. It might fail either due to a lack of compactness or to F not being lower semi-continuous. In second case, a natural approach is to define an l.s.c. functional that is as close to F as possible. This is the *lower semi-continuous relaxation* that is defined as

$$\overline{F}(x) \overset{\text{def.}}{=} \inf\left\{\liminf_{n\to\infty} F(x_n) :\ x_n \xrightarrow[n\to\infty]{d} x\right\} = \sup_{\substack{G\leq F\\ G \text{ is l.s.c.}}} G(x), \tag{0.3}$$

that is the largest lower semi continuous function that is smaller that F. It can be shown that $\overline{F}$ admits minimizers and that

$$\min_X \overline{F} = \inf_X F.$$

Whenever the Direct Method fails due to a lack of compactness, one approach is to embed the functional into another space for which good compactness results are available, compute its relaxation and employ the Direct Method in this setting. It is then desirable to exploit necessary conditions of optimality for the relaxed formulation to show that the solution can actually be represented by an element of the original space. In the sequel, we will discuss a few examples of problems that were studied with this approach.

Equilibrium shapes of liquids and drops

Our first example is a class of *geometric variational problem*, see for instance [Maggi, 2012], that gives the optimal shape that a liquid assumes inside a contained Ω that is assumed to be an open, bounded and connected subset of $\mathbb{R}^d$. From physical principles, a liquid occupies a region E of volume m inside a container Ω that minimizes a free energy given by a surface tension of the liquid and a total potential energy, being written as

$$\inf \left\{ (\operatorname{Per}(E;\Omega) - \beta \operatorname{Per}(E;\partial\Omega)) + \int_E g(x)\mathrm{d}x : {}^{E\subset A}_{|E|=m} \right\} \tag{0.4}$$

where $\beta > 0$ is the adhesion coefficient of the liquid to the material of the container, the liquid is submitted to the potential g, typically a gravitational potential, and $\operatorname{Per}(E;A)$ denotes the perimeter of a set E with smooth boundary inside an open set A and can be written as

$$\operatorname{Per}(E;A) = \int_{\partial E\cap A} 1\mathrm{d}\mathscr{H}^{d-1},$$

where $\mathscr{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. It turns out that the class of subsets of $\mathbb{R}^d$ with smooth boundary is not adapted to apply the Direct Method, since the usual notions of set convergence, as the Hausdorff distance (see Chapter 1), does not preserve smoothness of the boundary. The alternative is to look at the indicator functions of the these sets 1_E, since from the Gauss-Green Theorem, these functions admit a derivative in the sense of distributions given by

$$D1_E = \nu_E \mathscr{H}^{d-1} \llcorner \partial E, \text{ if } E \text{ has smooth boundary}$$

where $\nu_E : \partial E \to \mathbb{S}^{d-1}$ is the outwards normal of ∂E. One can then define a set of finite perimeter as a set E such that 1_E has a weak derivative given by a Radon measure, with bounded total variation norm $\|D1_E\|_{\mathcal{M}(\Omega)} < +\infty$. This definition is much weaker and allows for much less regular sets, for instance polygons, but it is designed in such a way that sets of finite perimeter have similar operation properties to those of smooth sets. For instance, one can show that a set of finite perimeter has a normal ν_E inside a set

$\partial^\star E \subset \operatorname{supp} D1_E$ called its reduced boundary. In addition, the Gauss-Green formula also holds, and we have that if E is of finite perimeter

$$D1_E = \nu_E \mathscr{H}^{d-1} \llcorner \partial^\star E, \text{ and } \operatorname{Per}(E) \stackrel{\text{def.}}{=} \mathscr{H}^{d-1}(\partial^\star E).$$

Since now the topology of Radon measures is much more flexible, for any sequence of sets of finite perimeter $(E_n)_{n\in\mathbb{N}}$ such that $|E_n \Delta E| \xrightarrow[n\to\infty]{} 0,$ we have $D1_{E_n} \xrightharpoonup[n\to\infty]{\star} D1_E$. The energy in (0.4) can be proven to be lower semi-continuous and the Direct Method is applicable.

The question that remains is if minimizers are actually in the original class of sets with smooth boundary. This was later confirmed by the regularity theory of quasi-minimizers of the perimeter, but without the detour on the general theory of sets of finite perimeter, the direct method would not be applicable.

The Optimal Transportation problem

Our next example dates from 1781, when Monge first proposed his formulation of the Optimal Transportation problem [Monge, 1781]: given two prescribed distributions of particles, the question is how to transport one onto the other while minimizing the total work that is proportional to the total traveled distance. It is naturally written in a modern terminology with probability measures. Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be two Polish spaces, *i.e.* complete and separable metrizable topological spaces, given $\mu \in \mathscr{P}(\mathcal{X})$, $\nu \in \mathscr{P}(\mathcal{Y})$ initial and final distributions and let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a continuous and bounded function; Monge's problem is then written as

$$\inf_{T_\sharp \mu = \nu} \int_\mathcal{X} c(x, T(x)) \mathrm{d}\mu, \tag{MP}$$

where the infimum is taken among all Borel measurable maps $T : \mathcal{X} \to \mathcal{Y}$ that pushes μ onto ν, that is such that the measure $T_\sharp \mu(A) \stackrel{\text{def.}}{=} \mu(T^{-1}(A)) = \nu(A)$, for all Borel measurable sets A.

Monge made important contributions about the qualitative properties of minimizers in the particular case that $c(x, y) = |x - y|$ and distributions that are absolutely continuous with respect to the Lebesgue measure, but as the rest of the Calculus of Variations community at the time he did not address the question of existence. The first difficulty is the feasibility of transporting μ onto ν; indeed if μ is given by a Dirac mass and ν is diffuse there is no map that can perform this transportation.

But that is not all, the Direct method is not enough to obtain existence even in the simplest case that the infimum is finite, $c(x, y) = |x-y|^2$ is the squared euclidean distance the measure, and μ has no atoms, so there exist maps achieving the transportation $T_\sharp \mu = \nu$. Let us try to apply the Direct method and let $(T_n)_{n\in\mathbb{N}}$ be a minimizing sequence, so we have that $\|\mathrm{id} - T_n\|_{L^2(\mu)} \le C$, for all $n \in \mathbb{N}$. From weak compactness in $L^2(\mu)$ we can extract a weakly convergent subsequence, with limit T. The problem is that this new limit does not in general respect the constraint $T_\sharp \mu = \nu$.

In 1942 [Kantorovich, 1942], L.Kantorovitch identified two difficulties with the previous formulation. Firstly, the feasibility of the transportation via maps does not allow the splitting of mass into different destinations, the main impediment for the case that μ has a Dirac mass. In addition, the pushforward operation introduces a highly non-linear constraint. For these reasons he introduced the following problem

$$W_c(\mu, \nu) \overset{\text{def.}}{=} \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\gamma(x, y), \tag{KP}$$

where $\Pi(\mu, \nu)$ denotes the class of all transportation plans coupling μ and ν, the measures $\gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ such that $\mu(\cdot) = \gamma(\cdot \times \mathcal{Y})$ and $\nu(\cdot) = \gamma(\mathcal{X} \times \cdot)$. Now instead of saying explicitly where each particle is sent, the quantity $\gamma(A \times B)$ represents the probability that a particle in A is sent to B.

This reformulation solves all the shortcomings of Monge's problem. For starters, for any pair of measures μ, ν, the product measure $\mu \otimes \nu$ is always an admissible transportation plan. In addition, the constraints are now linear, indeed defining $\pi_{\mathcal{X}} : (x, y) \mapsto x$, and similarly $\pi_{\mathcal{Y}} : (x, y) \mapsto y$, we have that

$$\gamma \in \Pi(\mu, \nu) \text{ if and only if } (\pi_{\mathcal{X}})_\sharp \gamma = \mu, \; (\pi_{\mathcal{Y}})_\sharp \gamma = \nu.$$

Since the space of probability measures endowed with the narrow topology, the weak topology in duality with the space of continuous and bounded functions, enjoys very flexible compactness properties such as Prokhorov's Theorem, we see that Kantorovitch's problem reduces to the minimization of a continuous functional over a compact set.

It was shown in [Pratelli, 2007] that Kantorovitch's problem corresponds to the lower semi-continuous relaxation of Monge's and the question is then to characterize when minimal transportation plans are induced by maps, that is when solutions are of the form $\gamma = (\mathrm{id}, T)_\sharp \mu$. The first result in this direction is due to [Brenier, 1987, Brenier, 1991] when $c(x, y) = |x - y|^2$ is the squared euclidean distance in $\mathbb{R}^d$. He proved that whenever μ is absolutely continuous there is an unique optimal transportation plan induced by a map $T = \nabla \phi$ that is the gradient of a convex function.

Game Theory and existence of Nash equilibria

Although not explicitly stated, these ideas of relaxation are present in the literature of Game Theory since the seminal work [Nash, 1951]. The concept of mixed strategies introduced by Nash is a clever way to convexify the original N-player game and be able to apply Browder's fixed point theorem and obtain an equilibrium.

Afterwards, there came an interest to understand games where the choice of an individual cannot affect the global outcome of game, but rather the mean collective choice induces a mean field that guides the choice of all players [Aumann, 1964, Aumann, 1966]. To this end consider two Polish spaces, $(\mathcal{X}, d_{\mathcal{X}})$ representing the space of types of players and $(\mathcal{Y}, d_{\mathcal{Y}})$ representing the space of admissible plays. To model the fact that now we have a continuum of players, we let $\mu \in \mathscr{P}(\mathcal{X})$ represent the distribution of their types and some $\nu \in \mathscr{P}(\mathcal{Y})$ represent the mean field of strategies chosen. In this scenario, a player of type x seeks to

$$\min_{\mathcal{Y}} \Phi(x, \cdot, \nu)$$

where $\Phi : \mathcal{X} \times \mathcal{Y} \times \mathscr{P}(\mathcal{Y})$ is the cost depending on their type and the mean field.

A first idea to define equilibria, as in [Schmeidler, 1973], would be to define as maps $T : \mathcal{X} \to \mathcal{Y}$ such that $T_\sharp \mu = \nu$ and almost every player solving the above problem. In [Mas-Colell, 1984], and as for the optimal transportation literature, the notion of Cournot-Nash equilibria was defined as a coupling $\gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ that satisfies $(\pi_{\mathcal{Y}})_\sharp \gamma = \nu$ and

$$\gamma\left(\left\{(x, y) : \ y \in \operatorname*{argmin}_{\mathcal{Y}} \Phi(x, \cdot, \nu)\right\}\right) = 1.$$

In the particular case that

$$\Phi(x, y, \nu) = c(x, y) + \frac{\delta \mathcal{E}}{\delta \nu}(\nu),$$

it was proven in [Blanchet and Carlier, 2016] that equilibria can be found via a variational approach. The authors show that if

$$\nu \in \operatorname*{argmin}_{\nu' \in \mathscr{P}(\mathcal{Y})} W_c(\mu, \nu') + \mathcal{E}(\nu'),$$

where W_c denotes the value of the optimal transportation problem with cost c, and $\gamma \in \Pi(\mu, \nu)$ is an optimal transportation plan, then γ is a Cournot-Nash equilibria. As a result, they also capitalize on the well established characterization of when optimal transportation is achieved by a map to answer to the original question of existence of Cournot-Nash equilibria in pure strategies. ■

2. Contributions of this thesis

The previous discussion, although certainly not exhaustive, motivates the framework of relaxing variational problems in the space of probability measures, or at least of describing the space of admissible competitors for the minimization with Radon measures as is the case of sets of finite perimeter and their Gauss-Green measure. It also exemplifies how this framework can be applied to various domains that will be studied in this thesis. In the following I list the contributions of this thesis.

2.1. 1D approximation of measures in Wasserstein spaces

Since the work of Kantorovitch, the field of optimal transportation has flourished with many new developments, both theoretical and in diverse applications. A major breakthrough in the theory was the definition of the Wasserstein distances, through the value of the optimal transportation problem with the cost given by a distance, see the bibliographical notes in [Villani, 2009, Chap. 6] for a detailed historical discussion. Indeed, letting $c(x,y) = |x-y|^p$ in $\mathbb{R}^d$, for $1 \leq p < +\infty$ one can define the p-Wasserstein distance between two measures $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$, the space of probability measures with finite p-moments see 4.3, as

$$W_p^p(\mu,\nu) \overset{\text{def.}}{=} \min_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^d \mathrm{d}\gamma. \tag{0.5}$$

The new space $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ can be shown to be a Polish space and the topology induced by W_p is very similar to the narrow convergence of measures, and coincides with it in compact domains. This has motivated the use of this distance as a data fidelity term in many applications. In [Lebrat et al., 2019, Chauffert et al., 2017], the authors propose an optimal transport based method for projecting images onto spaces of measures with sparse supports, for instance curves or point clouds.

Their methods are parametric by nature as their class of admissible minimizers can be described by curves, this motivated us to propose a non-parametric problem for the approximation of measures with 1-dimensional structures. In other words, given $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$, we seek to approximate it with a measure uniformly distributed over a 1-dimensional (1D) set. To this end, we consider the following variational problem

$$\inf_{\Sigma \text{ closed and connected}} W_p^p\left(\rho_0, \frac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma\right) + \Lambda\mathscr{H}^1(\Sigma). \tag{$W\mathscr{H}^1$}$$

The Wasserstein distance acts as a data-fidelity term and the 1-dimensional Hausdorff measure $\mathscr{H}^1(\Sigma)$, penalizes the length and forces competitors with finite energy to be 1-rectifiable, since they are connected. Notice that without the regularization one could make the Wasserstein as small as wanted by taking a space-filling curve. On the other hand, without the connectivity constraint one can also approximate any probability measure with a sum of Dirac masses while keeping the length term equal zero.

To show existence of solution to ($W\mathscr{H}^1$), the natural topology to work with such sets is the Hausdorff distance defined as

$$d_H(A,B) \overset{\text{def.}}{=} \max\left\{\sup_{a\in A} \operatorname{dist}(a,B), \sup_{b\in B} \operatorname{dist}(b,A).\right\} \tag{0.6}$$

However, one cannot easily resort to the direct method as the set of measures of the form $\mathscr{H}^1 \llcorner \Sigma$ is not closed due to concentration effects as illustrated in Figure 1 bellow.

We then propose the following relaxation of ($W\mathscr{H}^1$)

$$\inf_{\nu\in\mathscr{P}(\Omega)} W_p^p(\rho_0,\nu) + \Lambda\mathcal{L}(\nu), \text{ where } \mathcal{L}(\nu) \overset{\text{def.}}{=} \inf_{\alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu} \alpha, \tag{$\overline{W\mathscr{H}^1}$}$$

and proceed to show that the functional $\mathcal{L}$ is the l.s.c. relaxation of the functional

$$\ell(\nu) \overset{\text{def.}}{=} \begin{cases} \mathscr{H}^1(\Sigma), & \text{if } \nu = \dfrac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma \text{ for } \Sigma \text{ closed and connected,} \\ +\infty, & \text{otherwise.} \end{cases}$$

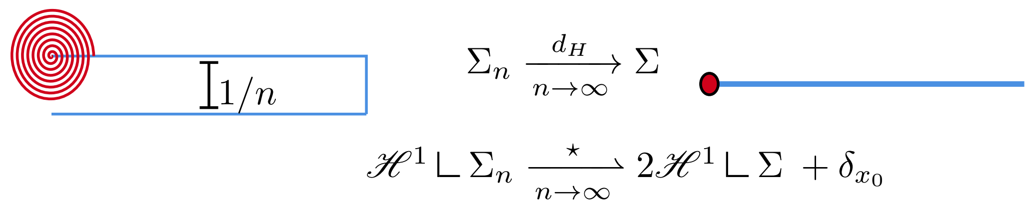

Figure 1: Concentration effects on the weak convergence of measures. In this image, Σ_n consists on two strips becoming closer and closer and a spiral converging very rapidly to a single point. In the Hausdorff convergence this gives only a segment, we lose the information of the total mass.

To compute this relaxation, the key tool we require is a density version of Gołab's Theorem. In [Gołąb, 1928], it is shown that the length is lower semi-continuous for sequences of connected sets converging in the Hausdorff distance, that is if $(\Sigma_n)_{n\in\mathbb{N}}$ is a sequence of compact and connected sets such that $\Sigma_n \xrightarrow[n\to\infty]{d_H} \Sigma$, then

$$\mathscr{H}^1(\Sigma) \leq \liminf_{n\to\infty} \mathscr{H}^1(\Sigma_n). \tag{0.7}$$

In [Ambrosio and Tilli, 2004, Paolini and Stepanov, 2013] a density version of this result was proved. Under the same conditions, if $\mathscr{H}^1 \llcorner \Sigma_n$ converges to a measure μ, then $\mu \geq \mathscr{H}^1 \llcorner \Sigma$. We have further generalized this result by weakening the notion of set convergence to the Kuratowski convergence, see Chapter 1 for a precise definition, and by allowing the sequence of sets to have locally finite length.

Theorem 0.1 (Density version of Golab's Theorem)**.** *Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a sequence of closed and connected subsets of $\mathbb{R}^d$ converging in the sense of Kuratowski to some closed set Σ and having locally uniform finite length,* i.e. *for all $R > 0$*

$$\sup_{n\in\mathbb{N}} \mathscr{H}^1(\Sigma_n \cap B_R(x_0)) < +\infty.$$

Define the measures $\mu_n \overset{\text{def.}}{=} \mathscr{H}^1 \llcorner \Sigma_n$, and let μ be a weak-$\star$ cluster point of this sequence. Then $\operatorname{supp} \mu \subset \Sigma$ and it holds that

$$\mu \geq \mathscr{H}^1 \llcorner \Sigma,$$

in the sense of measures.

With the stronger assumption of Hausdorff convergence of a sequence of compact sets, which was already known for instance from [Paolini and Stepanov, 2013], we can show that $\mathcal{L} = \bar{\ell}$ and argue that ($\overline{W\mathscr{H}^1}$) is the relaxation of the original problem ($W\mathscr{H}^1$). The feature of allowing for possibly unbounded sets with infinite length allows us to use it in the study of blow ups of solutions. We proceed then to showing that solutions to the relaxed problem are absolutely continuous with respect to $\mathscr{H}^1$, as long as the original measure also is, and that, whenever $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$ does not give mass to 1D sets, any solution of the relaxed problem has constant density and therefore must be a minimizer for the original problem.

Theorem 0.2. *Let $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$, $\Lambda > 0$. Then ($\overline{W\mathscr{H}^1}$) admits a solution ν, and there exists $\Lambda_\star \geq 0$ such that if $\Lambda > \Lambda_\star$, ν is a Dirac mass. For $\Lambda < \Lambda_\star$, ν is supported by a set $\Sigma \in \mathcal{A}$ and the following properties hold.*

1. *If ρ_0 is absolutely continuous w.r.t. $\mathscr{H}^1$, or has an L^∞ density w.r.t. $\mathscr{H}^1$, then so does ν.*

2. *If ρ_0 does not give mass to $1D$ sets, then ν is uniformly distributed over a connected set Σ of finite length, which is therefore is a solution to the original problem ($W\mathscr{H}^1$).*

In the sequel, we study qualitative properties of optimal networks Σ. First we prove Ahlfors regularity of minimizers.

Theorem 0.3. *Assume that $\rho_0 \in L^{\frac{d}{d-1}}(\mathbb{R}^d)$, and Λ is sufficiently small, so that minimizers ν of ($\overline{W\mathscr{H}^1}$) are not a Dirac mass. Then $\Sigma = \operatorname{supp} \nu$ is Ahlfors regular, i.e. there is r_0 depending on d, p, ρ_0 and $\mathcal{L}(\nu)$ and C depending only on d, p such that for any $x \in \Sigma$ and $r \leq r_0$ it holds that*

$$r \leq \mathscr{H}^1(\Sigma \cap B_r(x)) \leq Cr.$$

Next we study when is it that solutions are trees. Given a connected set Σ, we say that $\Gamma \subset \Sigma$ is a loop if it is homeomorphic to $\mathbb{S}^1$, we then say that Σ is a tree if it has no loops.

Theorem 0.4. *Whenever $\rho_0 = \sum_{i=1}^{N} a_i \delta_{x_i}$, for $a_i > 0$ for all $i = 1, \ldots, N$ and $\sum_{i=1}^{N} a_i = 1$ optimal networks Σ are trees.*

2.2. Phase field approximation of 1D variational problems

The access to reliable numerical solutions for the problem ($W\mathscr{H}^1$) is particularly important for the understanding of its properties. However, as the class of connected sets with finite length is not parametric, it is challenging to represent such objects numerically. A popular approach in the literature to deal with geometric variational problems is then to approximate such sets with the level sets of a Sobolev function, which can then be efficiently represented numerically. In other words we seek to find an approximation for a 1D set Σ of the form

$$\Sigma \approx \{\varphi_\varepsilon \leq \varepsilon^s\},$$

where φ_ε belongs to an appropriate space.

This approximation must be done in such a way that these Sobolev functions, called *phase fields* since they are designed to approximate two phases when they are close to 0 or to 1, can also be used to approximate a certain geometrical quantity. The approach of [Modica and Mortola, 1977] was to approximate the perimeter of a set with an elliptic functional. Later on, Ambrosio and Tortorelli proposed a similar approach to study the Mumford-Shah problem, see [Ambrosio et al., 2000, Chap. 6]. We then work with a modification of the Ambrosio-Tortorelli functional that is similar to the approach proposed in [Chambolle et al., 2019b, Chambolle et al., 2019a]: for a function φ_ε such that $\Sigma \approx \{\varphi_\varepsilon \le \varepsilon^s\}$, we approximate its length as

$$\mathscr{H}^1(\Sigma) \approx \mathcal{AT}_p(\varphi_\varepsilon) \stackrel{\text{def.}}{=} \frac{1}{\Lambda_{p,d}} \int_\Omega \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathrm{d}x.$$

The highest difficulty is actually to incorporate the connectivity constraints in this diffuse formulation. For this reason we consider the case $p > d \ge 2$, so that phase fields with finite $\mathcal{AT}_p$ belong to the Sobolev space $W^{1,p}$ and hence are Hölder continuous. This allows to control the level sets and have a nice synergy with the *diffuse connectivity functional* $\mathcal{C}_\varepsilon(\varphi_\varepsilon)$ proposed by Dondl and Wojtowytsch, see e.g. [Dondl et al., 2017], which is designed to measure how disconnected the level set $\{\varphi_\varepsilon \le \varepsilon^s\}$ is:

$$\mathcal{C}_\varepsilon(\varphi_\varepsilon) \stackrel{\text{def.}}{=} \int_{\Omega\times\Omega} \beta_\varepsilon(\varphi_\varepsilon(x))\beta_\varepsilon(\varphi_\varepsilon(y)) d^{F_\varepsilon\circ\varphi_\varepsilon}(x,y)\mathrm{d}x\mathrm{d}y.$$

The function β_ε is designed to select only a small level set of φ_ε and the geodesic distance $d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}$ is given by

$$d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}(x,y) \stackrel{\text{def.}}{=} \inf \left\{ \int_K F_\varepsilon \circ \varphi_\varepsilon(x) \mathrm{d}\mathscr{H}^1(x) : \begin{array}{c} K \text{ connected, } x,y \in K \\ \mathscr{H}^1(K) \le \varepsilon^{-1} \end{array} \right\},$$

where $F_\varepsilon(z)$ assumes the value 0 for $z \le \varepsilon^s$, so that $d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}(x,y) = 0$ if x and y are contained in the same connected component of $\{\varphi_\varepsilon \le \varepsilon^s\}$.

The interplay between these two functionals is the key to the diffuse approximation results proposed in this thesis and has proven itself sufficiently flexible to treat a variety of 1-dimensional shape optimization problems. In particular, we managed to exploit a relation of problem ($W\mathscr{H}^1$) with the *average distance minimizers problem* introduced by Butazzo and Stepanov in [Buttazzo and Stepanov, 2003], that can be described as follows: given a demographic density ρ_0 over a region Ω that can be for instance a city, one seeks to construct a metro network Σ in such a way to minimize the average distance of an individual to the network. The problem is then given by

$$\inf_{\Sigma \text{ connected}} \int_\Omega \operatorname{dist}(x,\Sigma)^q \mathrm{d}\rho_0(x) + \mathscr{H}^1(\Sigma). \tag{ADM}$$

By noticing that $\displaystyle\int_\Omega \operatorname{dist}(x,\Sigma)^q \mathrm{d}\rho_0(x) = \inf_{\operatorname{supp}\nu\subset\Sigma} W_q^q(\rho_0,\nu)$, we propose a unified approach to approximate both problems.

Problem (ADM) is approximated with the *diffuse average distance* functional:

$$\mathcal{AD}_\varepsilon(\nu_\varepsilon,\varphi_\varepsilon) \stackrel{\text{def.}}{=} \begin{cases} W_q^q(\rho_0,\nu_\varepsilon) + \Lambda\mathcal{AT}_p(\varphi_\varepsilon) + \dfrac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) & \nu_\varepsilon \in \mathscr{P}(\Omega), \\ \qquad\qquad + \dfrac{1}{\varepsilon^\ell}\displaystyle\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon, & \varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega) \\ +\infty, & \text{otherwise.} \end{cases}$$

For the problem ($\overline{W\mathscr{H}^1}$), all the terms have similar functions, except the Ambrosio-Tortorelli term since we wish to approximate $\mathcal{L}(\nu)$ instead of $\mathscr{H}^1(\Sigma)$. The proposed functional is then

$$\mathcal{WH}^1_\varepsilon(\alpha_\varepsilon,\nu_\varepsilon,\varphi_\varepsilon) \overset{\text{def.}}{=} \begin{cases} W^q_q(\rho_0,\nu_\varepsilon) + \Lambda\alpha_\varepsilon + \frac{1}{\varepsilon}\left\|\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon\right\|^2_{L^2(\Omega)} & \alpha_\varepsilon \geq 0, \\ & \nu_\varepsilon \in \mathscr{P}(\Omega) \\ \qquad + \frac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) + \frac{1}{\varepsilon^\ell}\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon, & \varphi_\varepsilon \in 1 + W^{1,p}_0(\Omega), \\ +\infty, & \text{otherwise,} \end{cases}$$

where the measure $\mu_\varepsilon = \mu_\varepsilon(\varphi_\varepsilon)$ is the diffuse transition measure and is defined as

$$\mu_\varepsilon \overset{\text{def.}}{=} \frac{1}{\Lambda_{p,d}}\left(\frac{\varepsilon^{p-d+1}}{p}|\nabla\varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'}(1-\varphi_\varepsilon)^2\right)\mathcal{L}^d \llcorner \Omega. \tag{0.8}$$

The notion of approximation we use is that of Γ-convergence of functionals. It was introduced by De Giorgi, see [Dal Maso, 1993] and Chapter 1 of this thesis for an introduction, for its pertinent properties for variational problems. Indeed, if $(F_\varepsilon)_{\varepsilon>0}$ converges in the sense of Γ convergence to a functional F, and $x_\varepsilon \in \operatorname{argmin} F_\varepsilon$ is a sequence converging to x, then this limit point is a minimizer of the Γ-limit, $x \in \operatorname{argmin} F$.

The rigorous results proved in this thesis on the phase field approximation of 1D shape optimization problems are the following.

Theorem 0.5. *Assume that $p > d \geq 2$, $\ell > s$ and that $\kappa > (2d+1)\frac{(s+1)p-d+1}{p-d}$, then*

- *the diffuse average distance functional approximates* (ADM)

$$\mathcal{AD}_\varepsilon \xrightarrow[\varepsilon\to 0]{\Gamma} \mathcal{AD}(\nu,\varphi) \overset{\text{def.}}{=} \begin{cases} W^q_q(\rho_0,\nu) + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu), & \nu \in \mathscr{P}(\Omega),\ \varphi \equiv 1, \\ +\infty, & \textit{otherwise,} \end{cases}$$

  where $\mathscr{H}^1_S(\operatorname{supp}\nu)$ is the length of the minimal Steiner tree connecting $\operatorname{supp}\nu$. The Γ-convergence holds in the strong topology of L^2 and weak topology of $\mathscr{P}(\Omega)$.

  In addition, let $(\nu_\varepsilon,\varphi_\varepsilon)_{\varepsilon>0}$ be a family of minimizers of $\mathcal{AD}_\varepsilon$, it admits a cluster point $(\nu,\varphi\equiv 1)$, which then achieves the infimum and

$$\min_\Sigma\,(\text{ADM}) = \min_{(\nu,\varphi)} \mathcal{AD}(\nu,\varphi),$$

  and it holds that

  - *Σ is a minimizer of* (ADM) *if, and only if, it is a minimal Steiner tree of $\operatorname{supp}\nu$, for some ν minimizer of $\mathcal{AD}$;*
  - *ν is a minimizer of $\mathcal{AD}$ if, and only if, it can be written as $\nu = (\pi_\Sigma)_\sharp\rho_0$, where π_Σ is a measurable selection of the projection operator onto some Σ minimizer of* (ADM).

- *for problem* ($\overline{W\mathscr{H}^1}$)*, it holds that*

$$\mathcal{WH}^1_\varepsilon \xrightarrow[\varepsilon\to 0]{\Gamma} \mathcal{WH}^1(\alpha,\nu,\varphi) \overset{\text{def.}}{=} \begin{cases} W_q^q(\rho_0,\nu) + \Lambda\mathcal{L}(\nu), & \nu \in \mathscr{P}(\Omega),\ \alpha \geq \mathcal{L}(\nu), \\ +\infty, & \textit{otherwise,} \end{cases}$$

the Γ*-convergence being held in* $\mathbb{R}$*, the strong topology of* L^2 *and weak topology of* $\mathscr{P}(\Omega)$*.*

If in addition, whenever ρ_0 *does not charge countably* $\mathscr{H}^1$*-rectifiable sets, if* $(\alpha_\varepsilon,\nu_\varepsilon,\varphi_\varepsilon)_{\varepsilon>0}$ *is a sequence of minimizers of* $\mathcal{WH}^1_\varepsilon$*, then it has a cluster point* $(\alpha,\nu,\varphi\equiv 1)$ *of the form*

$$\alpha = \mathscr{H}^1(\Sigma),\ \nu = \frac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma,\ \textit{ where } \Sigma \textit{ is connected } \mathscr{H}^1\textit{-rectifiable,}$$

so that Σ *minimizes* ($W\mathscr{H}^1$)*.*

2.3. The Wasserstein gradient flow of the total variation

In the seminal paper [Jordan et al., 1998] the authors proposed a variational interpretation of the Fokker-Planck equation as the gradient flow of the entropy functional. There have been many developments since then, notably from [Ambrosio et al., 2008], see also [Santambrogio, 2015, Chap. 8]. Given $\Omega \subset \mathbb{R}^d$ convex and compact, and a functional $\mathscr{F} : \mathscr{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$, the *JKO scheme* consists of

$$\rho_{k+1} \in \operatorname*{argmin}_{\rho\in\mathscr{P}(\Omega)} \mathscr{F}(\rho) + \frac{1}{2\tau}W_2^2(\rho_k,\rho),\ \text{for } \rho_0 \text{ given.} \tag{JKO}$$

Once this sequence $(\rho_k)_{k\in\mathbb{N}}$ is obtained, one can define a curve, depending on the parameter τ, in the Wasserstein space as

$$\rho_\tau(t) \overset{\text{def.}}{=} \rho_k, \qquad \text{if } t \in [k\tau,(k+1)\tau).$$

Letting $\tau \to 0$, one then obtains a curve in the Wasserstein space that solves the following PDE

$$\begin{cases} \partial_t\rho(t) = \operatorname{div}\left(\rho\nabla\frac{\delta\mathscr{F}}{\delta\rho}(\rho(t))\right), & \text{in } [0,T]\times\Omega \\ \frac{\partial\rho}{\partial n} = 0, & \text{on } \partial\Omega \\ \rho(0) = \rho_0. \end{cases}$$

As mentioned before, [Jordan et al., 1998] proved the convergence of this scheme when we consider $\mathscr{F}(\rho) = \int_\Omega \rho\log\rho\,\mathrm{d}x$, so that the limit PDE becomes the heat equation, being of great philosophical importance as the heat equation is then formally interpreted as a minimization of entropy.

On the other hand, using gradient flows is standard technique for image denoising and inpainting in the image processing community. In particular, the *total variation functional* defined as

$$\mathrm{TV}(u) \stackrel{\text{def.}}{=} \sup\left\{\int_\Omega \operatorname{div} z(x)u(x)\mathrm{d}x : z \in C_c^1\left(\Omega;\mathbb{R}^N\right), \|z\|_\infty \leq 1\right\},$$

is known for its properties of preserving the edges and promoting parsimonious reconstruction of noisy images, see [Chambolle et al., 2010, Chambolle et al., 2016]. Indeed, since the seminal work [Rudin et al., 1992] the flow of the total variation in the L^2-topology has become a standard benchmark for image denoising, the now known as *Rudin-Osher-Fatemi problem* given by

$$\inf_{u\in L^2(\Omega)} \mathrm{TV}(u) + \frac{1}{2\lambda}\|u-g\|^2_{L^2(\Omega)}. \tag{ROF}$$

Recently, the Wasserstein gradient flow of TV has been studied for applications in image processing [Burger et al., 2012, Benning et al., 2013, Carlier and Poon, 2019].

In this thesis, we revisit the work of Carlier & Poon [Carlier and Poon, 2019] and derive Euler-Lagrange equations for the problem

$$\inf_{\rho\in\mathscr{P}(\Omega)} \mathrm{TV}(\rho) + \frac{1}{2\tau}W_2^2(\rho_0,\rho). \tag{TV-W}$$

Differently from [Carlier and Poon, 2019], that further regularize the problem with an entropy term, our approach to derive optimality conditions for the problem (TV-W), is to relate it to a suitable (ROF) problem, whose optimality conditions are well understood. This way we can derive further regularity for the involved quantities, which is crucial for understanding the limit of the gradient flow scheme (JKO).

Theorem 0.6. *Let $\Omega \subset \mathbb{R}^d$ be a compact and convex domain. For any $\rho_0 \in L^1(\Omega)\cap\mathscr{P}(\Omega)$, let ρ_1 be the unique minimizer of (TV-W). The following hold.*

1. *There is a vector field $z \in H_0^1(\operatorname{div};\Omega)\cap L^\infty(\Omega;\mathbb{R}^d)$ and a Lagrange multiplier $\beta \geq 0$ such that*
$$\begin{cases} \operatorname{div} z + \dfrac{\psi_1}{\tau} = \beta, & \textit{a.e. in } \Omega \\ z\cdot\nu = 0, & \textit{on } \partial\Omega \\ \beta\rho_1 = 0, & \textit{a.e. in } \Omega \\ \operatorname{div} z \in \partial\,\mathrm{TV}(\rho_1), \end{cases} \tag{0.9}$$
where ψ_1 is a Kantorovitch potential associated with ρ_1.

2. *The Lagrange multiplier β is the unique solution to (ROF) with $\lambda = 1$ and $g = \psi_1/\tau$.*

3. *The functions $\operatorname{div} z$, ψ_1 and β are Lipschitz continuous.*

In the sequel, we propose a Douglas-Rachford type algorithm for its numerical optimization. We validate our algorithm with an example for which we can derive an explicit solution and later apply it to a problem consisting in the reconstruction of quantized images.

2.4. From Nash to Cournot-Nash

One of the central questions in the game theory and particularly, mean field games, is the following:

> *Given a sample of players following a continuous distribution, when does a sequence of Nash equilibria for the associated finite game will converge to a notion of equilibrium for a game with infinitely many players?*

In this work we focus on answering this question in the context of *Cournot-Nash equilibria.*

Consider two Polish spaces $\mathcal{X}$ and $\mathcal{Y}$, one can interpret $\mathcal{X}$ as the space of types of players with distribution given by $\mu \in \mathscr{P}(\mathcal{X})$, $\mathcal{Y}$ to be the space admissible strategies for said players, with distribution given by $\nu \in \mathscr{P}(\mathcal{Y})$. In this context, a coupling $\gamma \in \Pi(\mu, \nu)$ represents the joint distribution of players and strategies. We consider also a function $\Phi : \mathcal{X} \times \mathcal{Y} \times \mathscr{P}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$, such that $\Phi(x, y, \nu)$ denotes the cost of a player of type x to choose the strategy y, in a mean field of strategies represented by the distribution ν. We then say that a coupling $\gamma \in \Pi(\mu, \nu)$ is a Cournot-Nash equilibrium if

$$\gamma\left(\left\{(x, y) \in \mathcal{X} \times \mathcal{Y} : y \in \operatorname*{argmin}_{y' \in \mathcal{Y}} \Phi(x, y', \nu)\right\}\right) = 1, \tag{0.10}$$

Results guaranteeing the existence of equilibria have been established with fixed point methods in the above-mentioned works [Schmeidler, 1973, Mas-Colell, 1984]. This approach relies strongly on the continuity of the cost function. Building upon the work of [Blanchet and Carlier, 2016], whenever Φ is composed of an individual cost plus a mean pair-wise interaction term, having the following particular form

$$\Phi(x, y, \nu) = c(x, y) + \int_{\mathcal{Y}} L(y) \mathrm{d}\nu(y) + \int_{\mathcal{Y}} H(y, y') \mathrm{d}\nu(y'), \tag{0.11}$$

we characterize the Cournot-Nash equilibria of the game associated with Φ as the extremal points of the following potential functional

$$\mathcal{J}(\gamma) \stackrel{\text{def.}}{=} \begin{cases} \displaystyle\int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\mathrm{d}\gamma + \int_{\mathcal{Y}} L(y)\mathrm{d}\nu(y) + \int_{\mathcal{Y}\times\mathcal{Y}} H(y, y')\mathrm{d}\nu \otimes \nu, & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{if } \gamma \notin \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y}). \end{cases}$$

In the game theory literature, this sort of game is said to have a potential structure, in the sense that equilibria can be obtained by minimizing this potential functional.

With this observation, we propose a sequence of N-player games also exhibiting a potential structure, that is each one admitting a potential function $\mathcal{J}_N$, whose minimization yields Nash equilibria for their associated games. We then show that this sequence of potential functionals Γ-converges to $\mathcal{J}$. In particular, this means that a sequence of Nash equilibria obtained as minimizers of this sequence of potential functionals converge to minimizer of $\mathcal{J}$ and, therefore to a Cournot-Nash equilibrium giving a positive answer to the question we were interested in for a fairly general class of games.

Publications. This thesis gave rise to the following publications and preprints

- *The Total Variation-Wasserstein Problem*, joint work with Antonin Chambolle and Vincent Duval, published at GSI'23
- *1D approximation of measures in Wasserstein spaces*, joint work with Antonin Chambolle and Vincent Duval, preprint
- *Phase-field approximation for 1-dimensional shape optimization problems*, preprint
- *From Nash to Cournot-Nash via Γ-convergence*, joint work with Guilherme Mazanti and Laurent Pfeiffer, in preparation.

Presentation and Awards The works conducted in this thesis were presented in the following conferences

- The results in Chapter 3 were presented as a poster in the Lantin American Congress of Industrial and Applied Mathematics (LACIAM), held in Rio de Janeiro in January of 2023. **(Best poster award)**
- The results in Chapter 3 and Chapter 4 were presented as a contributed talk in the conference "Calculus of Variations and Applications", held in Paris from 19 to 23 of June of 2023.
- The results in Chapter 6 were published and presented in the conference "Geometric Science of Information" (GSI 2023) in Saint-Malo from 30 of August to September 1st of 2023.
- The results in Chapter 7 were presented in the Journées SMAI-MODE of 2024 in Lyon from 27 to 29 march of 2024. **(Awarded the Prix Dodu)**

3. Organization of this thesis

This theses is organized as follows.

- In Chapter 1 we introduce the mathematical theories and methods from the calculus of variations employed in this thesis. We start our discussion with standard tools to study variational problems: the Direct Method, relaxations and Γ-convergence. We then discuss the topologies used throughout this work, describing notions of convergence of sets and measures. The tools from geometric measure theory that we will require are surveyed and we finish with a more in depth discussion of the optimal transport problem and its properties.
- Chapter 2 is dedicated to the theory of metric continuums, in other words connected sets with finite $\mathscr{H}^1$-measure. We profit the situation to gather classical results from the literature and to emphasize some small advancements obtained in this thesis for our main goals that are described in the following Chapters.

- In Chapter 3 we introduce the Wasserstein-$\mathscr{H}^1$ problem and its regularization, passing though the length functional described in 2.1. We discuss thoroughly the conditions on the regularization parameter Λ for the problem to have non-trivial solutions, that is for solutions not to be reduced to a Dirac delta. Next, we prove that solutions to the relaxed problem are rectifiable measures, as long as the original measure is absolutely continuous with respect to $\mathscr{H}^1$, and we proceed to prove existence with a blow-up argument.

- In Chapter 4 we continue with the study of the Wasserstein-$\mathscr{H}^1$ problem, but now we are mostly concerned with qualitative properties of minimizers. First we show that if the original measure ρ_0 is sufficiently integrable (belongs to $L^{\frac{d}{d-1}}(\mathbb{R}^d)$), then any solution to the relaxed problem is Ahlfors regular. Under the same hypothesis we show that minimizers are trees, do not present any subset that is a homeomorphic image of $\mathbb{S}^{d-1}$. Next we show that the same is true in the opposite case that, where ρ_0 is singular, given by a sum of Dirac masses.

- In Chapter 5 we pass to the phase field approximation of 1D shape optimization problems described in 2.2

- We proceed in Chapter 6 with the study of the gradient flow of the total variation functional in the Wasserstein topology. We derive the Euler-Lagrange equations be means of the optimality conditions of a suitable Rundin-Osher-Fatemi problem and show that all level sets of the solution are solutions to the same prescribed curvature problem. In the sequel we proposed a proximal splitting algorithm to solve it, which is validated with an example whose solution can be explicitly computed.

- In Chapter 7 we show that a sequence of Nash equilibria to a suitable family of N-players converge to a Cournot-Nash equilibria by exploiting a particular potential structure. This potential structure implies the existence of a potential function whose minimization yields equilibria. We show that the sequence of potential functions to the N-players games Γ-converge with full probability to the potential function of the game with a continuum of players.

Introduction - version Française

1. Pour quoi optimiser dans les espaces de mesures?

On peut dire que le calcul des variations trouve ses origines dans l'Antiquité, où les Grecs anciens avaient déjà formulé les premières instances du problème isopérimétrique. Celui-ci consiste, parmi toutes les formes d'un plan ayant une aire fixe, à trouver celle ayant le plus petit périmètre. Il existe un débat pour savoir si ce problème peut être considéré comme fondateur du domaine, car ses premières solutions étaient purement géométriques. En revanche, d'autres problèmes apparus au XVIIe siècle, avec le développement de la mécanique classique, ont été formulés dans une terminologie analytique. Le premier, et peut être le plus célèbre d'entre eux, est le problème de la brachystochrone, initialement proposé par Galilée et reformulé ensuite avec un modèle mathématique précis par Bernoulli.

Bien que ces deux problèmes relèvent de l'optimisation, cela ne les caractérise pas nécessairement comme variationnels. Par exemple, l'optimisation combinatoire partage également cette nature de recherche d'un objet minimal, mais ce genre de problème reste assez éloigné des méthodes employées en calcul des variations.

Le terme « Calcul des Variations » a été attribué par Euler, à partir de la méthode des variations développée par Lagrange. On peut se référer, par exemple, à l'introduction du traité historique de Goldstine sur le sujet [Goldstine, 1980]. À leur époque, le paradigme en physique était que la nature ne gaspille pas d'énergie et que les trajectoires des particules sont déterminées par la minimisation d'une certaine notion d'énergie. Ils s'intéressaient alors aux problèmes de la forme

$$\min\left\{\mathcal{L}(x)\stackrel{\text{def.}}{=}\int_0^T L(t,x(t),\dot{x}(t))\mathrm{d}t : x(0)=x_0,\ x(T)=x_T\right\}, \tag{0.12}$$

la minimisation s'effectuant parmi toutes les courbes $x : [0,T] \to \mathbb{R}$ connectant les points x_0 et x_T. L'idée de Lagrange pour calculer un minimiseur de cette énergie était de perturber la courbe optimale avec des variations sous la forme $x+\varepsilon h$, où h est une courbe telle que $h(0)=h(T)=0$ qu'on appelle une variation. Les contraintes initiale et finale sur h garantissent que $x+\varepsilon h$ est encore admissible pour la minimisation de $\mathcal{L}$ de façon que, dès que x est un minimiseur, $\varepsilon \mapsto \mathcal{L}(x+\varepsilon h)$ atteint son minimum en $\varepsilon=0$ et la dérivée par rapport à ε en $\varepsilon=0$ est zéro, ce qui donne la célèbre équation d'Euler-Lagrange:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}}(t,x,\dot{x}) = \frac{\partial L}{\partial x}(t,x,\dot{x}). \tag{0.13}$$

Remarquons que nous n'avons pas explicitement précisé dans le problème (0.12) quelle est la classe de courbes dans laquelle la minimisation a lieu, et en effet, aux débuts du calcul des variations, cela n'était pas clairement énoncé. En pratique, les courbes admissibles étaient supposées lisses, ce qui suggère une croyance implicite selon laquelle la nature est continue (ce qui n'est pas nécessairement vrai). Cela est intimement lié à la question de l'existence des minimiseurs, et effectivement, cette question n'a pas été abordée aux débuts du domaine. Ce n'est qu'en 1915 que Tonelli a proposé un résultat d'existence pour une large classe d'énergies dans l'espace des fonctions absolument continues; voir, par exemple, la discussion dans [Clarke, 2013, Chap. 16].

L'idée principale de la preuve de Tonelli est aujourd'hui connue sous le nom de *méthode directe du calcul des variations*, et elle s'est révélée suffisamment flexible pour être appliquée dans des cadres beaucoup plus généraux, comme les problèmes de minimisation dans les espaces métriques. Pour l'exprimer dans un cadre moderne, soit (X,d) un espace métrique, et $F : X \to \mathbb{R}\cup\{+\infty\}$ une fonctionnelle telle que :

- F est semi-continue inférieurement, *i.e.* pour toute suite $x_n \xrightarrow[n\to\infty]{d} x$, on a
$$F(x) \le \liminf_{n\to\infty} F(x_n).$$
- F possède des ensembles de niveaux compacts, *i.e.* les ensembles $\{x\in X : F(x)\le \ell\}$ sont compacts.

En supposant que $\inf\limits_{x\in X} F(x) < +\infty$, soit $(x_n)_{n\in\mathbb{N}}$ une suite minimisante, c'est-à-dire une suite telle que $F(x_n)$ converge vers $\inf\limits_X F$, comme les sous-ensembles de niveau de F sont compacts, $(x_n)_{n\in\mathbb{N}}$ admet une sous-suite convergente vers un certain x. Il suit de la semi-continuité inférieure de F que

$$F(x) \le \liminf_{n\to\infty} F(x_n) = \inf_X F,$$

d'où la optimalité de x.

Cette méthode est maintenant largement acceptée et peut être considérée comme la méthode la plus répandue pour prouver l'existence de solutions aux problèmes variationnels. Cependant, on s'intéresse notamment aux situations dont elle ne marche pas. Cela peut échouer soit à cause d'un manque de compacité ou si F n'est pas semi-continue inférieurement. Dans ce deuxième cas, une approche naturelle est de définir une fonctionnelle s.c.i. qui soit aussi proche de F que possible. Il s'agit de la *relaxation semi-continue inférieure* définie comme

$$\overline{F}(x) \overset{\text{def.}}{=} \inf\left\{ \liminf_{n\to\infty} F(x_n) :\ x_n \xrightarrow[n\to\infty]{d} x \right\} = \sup_{\substack{G \le F \\ G \text{ est s.c.i.}}} G(x), \tag{0.14}$$

c'est-à-dire la plus grande fonctionnelle s.c.i. plus petite que F. On peu montrer que $\overline{F}$ admets des minimiseurs et que

$$\min_X \overline{F} = \inf_X F.$$

Quand la Méthode Directe ne marche pas à cause d'un manque de compacité, une approche possible consiste à intégrer la fonctionnelle dans un autre espace plus large avec des bonnes propriétés de compacité, calculer la relaxation s.c.i. et utiliser de bonnes conditions nécessaires d'optimalité pour le problème relaxé pour montrer que les minimiseurs peuvent en fait être représentés par un élément du espace d'origine. Dans la suite, nous discutons quelques exemples de problèmes qui ont été étudiés avec cette approche.

Formes d'équilibre des liquides

Notre premier exemple est une classe de *problèmes variationnels géométriques*, voir par exemple [Maggi, 2012], qui donne la forme optimale qu'un liquide assume dans un récipient Ω, qu'on suppose être un sous-ensemble ouvert, connexe, borné de $\mathbb{R}^d$. Selon des principes physiques, le liquide occupe une région E de volume m dans Ω que minimise une énergie libre donnée par la tension superficielle et l'énergie potentielle, étant écrite comme

$$\inf\left\{ (\operatorname{Per}(E;\Omega) - \beta \operatorname{Per}(E;\partial\Omega)) + \int_E g(x)\mathrm{d}x : \substack{E\subset A \\ |E|=m} \right\} \tag{0.15}$$

où $\beta > 0$ est un coefficient d'adhésion du liquide à la surface du récipient, liquide est soumis à un potentiel g, typiquement gravitationnel et $\operatorname{Per}(E;A)$ représente le périmètre d'un ensemble E avec bord lisse à l'intérieur d'un ensemble ouvert A et peut être écrit comme

$$\operatorname{Per}(E;A) = \int_{\partial E\cap A} 1\mathrm{d}\mathscr{H}^{d-1},$$

où $\mathscr{H}^{d-1}$ est la mesure d'Hausdorff de dimension $(d-1)$. Il se trouve que la classe des sous-ensembles de $\mathbb{R}^d$ dont le bord est lisse n'est pas adapté à la Méthode Directe, une fois que la notion de convergence usuelle des ensembles, comme au sens de la distance d'Hausdorff par exemple (voir le Chapitre 1), ne préserve pas la régularité du bord. L'alternative est d'utiliser les fonctions indicatrices de ces ensembles 1_E, car d'après le Théorème de Gauss-Green, ces fonctions admettent une dérivée au sens des distributions donnée par

$$D1_E = \nu_E \mathscr{H}^{d-1} \llcorner \partial E,$$

où $\nu_E : \partial E \to \mathbb{S}^{d-1}$ est le vecteur normal unitaire extérieur à ∂E.

On peut ensuite définir un ensemble de périmètre fini, comme un ensemble E tel que 1_E a une dérivée au sens des distributions donnée par une mesure de Radon, avec une variation totale finie $\|D1_E\|_{\mathcal{M}(\Omega)} < +\infty$. Cette définition est beaucoup plus faible et nous permet de considérer des ensembles moins réguliers commes des polygones, mais elle est conçue de façon que à avoir des opérations analogues aux ensembles réguliers. Par exemple, on peut montrer que un ensemble de périmètre fini admets un vecteur normal ν_E dans un ensemble $\partial^\star E \subset \operatorname{supp} D1_E$ appelé la frontière réduite. De plus, la formule de Gauss-Green est également valide, et on a que si E est de périmètre fini

$$D1_E = \nu_E \mathscr{H}^{d-1} \llcorner \partial^\star E, \text{ et } \operatorname{Per}(E) \stackrel{\text{def.}}{=} \mathscr{H}^{d-1}(\partial^\star E).$$

Comme la topologie des mesures de Radon est beaucoup plus flexible, pour n'importe quelle suite d'ensembles de périmètre fini $(E_n)_{n\in\mathbb{N}}$ telle que $|E_n \Delta E| \xrightarrow[n\to\infty]{} 0$, nous avons $D1_{E_n} \xrightarrow[n\to\infty]{\star} D1_E$. On peut montrer que l'énergie (0.15) est s.c.i. et la Méthode Directes'applique.

La question qui reste est si les minimiseurs sont en réalité dans la classe originale des ensembles réguliers. Cela a été confirmé par la théorie des quasi-minimiseurs du périmètre, mais sans le détour dans la théorie générale des ensembles à périmètre fini, la Méthode Directe n'apporte pas d'information.

Le problème du Transport Optimal

Notre prochain exemple date de 1781, quand Monge a proposé sa formulation du problème de Transport Optimal [Monge, 1781]: données deux distributions de particules, comment transporter une vers l'autre en minimisant le travail de transport total, qui est proportionnel à la distance totale parcourue. Ce problème peut être décrit naturellement avec la terminologie moderne des mesures de probabilité. Soit $(\mathcal{X}, d_{\mathcal{X}})$ et $(\mathcal{Y}, d_{\mathcal{Y}})$ deux espaces Polonais, *i.e.* des espaces topologiques complètes et métrisables, données $\mu \in \mathscr{P}(\mathcal{X})$, $\nu \in \mathscr{P}(\mathcal{Y})$ les distributions initiale et finale et soit $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ une fonction continue et bornée; le problème de Monge s'écrit comme

$$\inf_{T_\sharp \mu = \nu} \int_{\mathcal{X}} c(x, T(x)) \mathrm{d}\mu, \tag{MP}$$

où l'infimum se passe parmi le maps Borel mesurables $T : \mathcal{X} \to \mathcal{Y}$ tels que $T_\sharp \mu(A) \stackrel{\text{def.}}{=} \mu(T^{-1}(A)) = \nu(A)$, pour tout ensemble mesurable A.

Monge a apporté d'importantes contributions sur les propriétés qualitatives des minimiseurs, dans le cas particulier où $c(x, y) = |x - y|$ et les distributions sont absolument continues par rapport à la mesure de Lebesgue, mais ainsi, comme la communauté de Calcul des Variations à l'époque, il n'a pas étudié la question d'existence. La première difficulté concerne l'admissibilité de transporter μ vers ν ; en effet, si μ est donnée par une masse de Dirac et ν est une mesure diffuse, il n'existe pas un map qui réalise ce transport.

De plus, la Méthode Directe ne suffit pas pour obtenir l'existence dans le cas encore plus simple quand l'infimum est fini, $c(x, y) = |x - y|^2$ est la distance euclidienne au carré et μ n'a pas d'atomes, de formes qu'il existe un map T tel que $T_\sharp\mu = \nu$. Si on essaie d'appliquer la Méthode Directe, prenons une suite minimisante $(T_n)_{n\in\mathbb{N}}$, de formes que $\|\mathrm{id} - T_n\|_{L^2(\mu)} \leq C$, pour tout $n \in \mathbb{N}$. En utilisant la compactité faible dans $L^2(\mu)$, on peut extraire une sous-suite convergente dans la topologie faible, avec limite T. Le problème c'est que cette limite ne respecte pas en général la contrainte $T_\sharp\mu = \nu$.

En 1942 [Kantorovich, 1942], L.Kantorovitch identifie les deux difficultés principales dans la formulation précédante. D'abord, la admissibilité du transport avec les maps n'admets pas la séparation des masses ponctuelles en destinations différentes. Ensuite, l'opération de pushforward introduit une contrainte fortement non-linéaire, alors que la convergence faible (que on peut espérer la compacité en dimension infinie) est définie pour marcher bien avec des opérations linéaires. Pour ces raisons, il a introduit le problème suivant

$$W_c(\mu, \nu) \overset{\text{def.}}{=} \min_{\gamma\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\mathrm{d}\gamma(x, y), \tag{KP}$$

où $\Pi(\mu, \nu)$ dénote la classe de couplages entre μ et ν, les mesures $\gamma \in \mathscr{P}(\mathcal{X}\times\mathcal{Y})$ telles que $\mu(\cdot) = \gamma(\cdot \times \mathcal{Y})$ et $\nu(\cdot) = \gamma(\mathcal{X} \times \cdot)$. Au lieu d'expliciter où chaque particule est envoyée, la quantité $\gamma(A \times B)$ représente la probabilité qu'une particule dans A est envoyée à B.

Cette reformulation résout toutes les lacunes du problème de Monge. D'abord, pour toute paire de mesures μ, ν, la mesure produit $\mu \otimes \nu$ est toujours admissible en tant que plan de transport. De plus, les contraintes sont maintenant linéaires, en effet en définissant $\pi_{\mathcal{X}} : (x, y) \mapsto x$, ainsi comme $\pi_{\mathcal{Y}} : (x, y) \mapsto y$, nous avons que

$$\gamma \in \Pi(\mu, \nu) \text{ si e seulement si } (\pi_{\mathcal{X}})_\sharp\gamma = \mu,\ (\pi_{\mathcal{Y}})_\sharp\gamma = \nu.$$

Comme l'espace de mesures de probabilité muni de la topologie étroite, la topologie faible en dualité avec l'espace des fonctions continues et bornées, présente des propriétés très flexibles de compacité, comme le Théorème de Prokhorov, nous voyons que le problème de Kantorovitch se réduit à minimiser une fonctionnelle linéaire continue dans un ensemble compact.

Il a été démontré dans [Pratelli, 2007] que le problème de Kantorovitch s'agit de la relaxation semi-continue inférieure du problème de Monge et la question qui reste est de characteriser quand un plan de transport optimal est induit par des maps, c'est-à-dire quand les solutions sont de la forme $\gamma = (\mathrm{id}, T)_\sharp\mu$. Le premier résultat dans cette direction a été proposé par Brenier dans [Brenier, 1987, Brenier, 1991] quand $c(x, y) = |x - y|^2$. Il a prouvé que quand μ est absolument continue il y a un unique plan de transport optimal induit par un map $T = \nabla\phi$ donné par le gradient d'une fonction convexe.

Théorie des Jeux et existence des équilibres de Nash

Les mêmes idées de relaxation ont été employées dans la littérature de théorie des jeux, même si pas explicitement, depuis le travail séminal [Nash, 1951]. Le concept des stratégies mixtes introduites par Nash est une façon astucieuse de convexifier le jeu à N-joueurs original, pour pouvoir appliquer le théorème de point fixe de Brouwer et obtenir un équilibre.

Ensuite, il y a eu un intérêt de comprendre le comportement des jeux où un individu n'affecte pas le résultat global du jeu, mais où le choix collectif induit un champ moyen qui guide le comportement des joueurs [Aumann, 1964, Aumann, 1966]. Pour cela, considérons deux espaces Polonais, $(\mathcal{X}, d_{\mathcal{X}})$ représentant l'espace des types des joueurs et $(\mathcal{Y}, d_{\mathcal{Y}})$ qui représente l'espace de stratégies admissibles. Pour modéliser le fait que maintenant nous avons un continuum de joueurs, prenons $\mu \in \mathscr{P}(\mathcal{X})$ une distribution de probabilité des types de ceux-ci et $\nu \in \mathscr{P}(\mathcal{Y})$ représentant le champ moyen des stratégies choisies. Dans cette situation, un joueur de type x cherche à minimiser

$$\min_{\mathcal{Y}} \Phi(x, \cdot, \nu)$$

où $\Phi : \mathcal{X} \times \mathcal{Y} \times \mathscr{P}(\mathcal{Y})$ est le coût qui dépend de leur type et du champ moyen.

Une façon préliminaire de définir un équilibre, comme dans [Schmeidler, 1973], serait de définir des maps $T : \mathcal{X} \to \mathcal{Y}$ telles que $T_\sharp \mu = \nu$ de façon que presque tout joueur résout le problème de minimisation précédant. Dans [Mas-Colell, 1984], comme dans la littérature de transport optimal, la notion de équilibre de Cournot-Nash a été définie comme un couplage $\gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ que satisfait $(\pi_{\mathcal{Y}})_\sharp \gamma = \nu$ et

$$\gamma\left(\left\{(x, y) : \; y \in \operatorname*{argmin}_{\mathcal{Y}} \Phi(x, \cdot, \nu)\right\}\right) = 1.$$

Dans le cas particulier

$$\Phi(x, y, \nu) = c(x, y) + \frac{\delta \mathcal{E}}{\delta \nu}(\nu),$$

il a été démontré dans [Blanchet and Carlier, 2016] qu'un équilibre peut être trouvé à travers un principe variationnel. Les auteurs prouvent que si

$$\nu \in \operatorname*{argmin}_{\nu' \in \mathscr{P}(\mathcal{Y})} W_c(\mu, \nu') + \mathcal{E}(\nu'),$$

où W_c dénote la valeur du problème de transport optimal à coût c, et $\gamma \in \Pi(\mu, \nu)$ est un plan de transport optimal, alors γ est un équilibre du type Cournot-Nash. Par conséquent, ils se servent de la théorie bien établie du transport optimal, notamment la caractérisation de quand le transport optimal est effectué par un map de transport pour répondre à la question d'existence des équilibres du type Cournot-Nash en stratégies pures. ■

2. Contributions du présent travail

La discussion précédente, certainement pas exhaustive, motive le cadre de relaxer des problèmes variationnels dans les espaces de mesures de probabilité, ou moins motivent la description de l'espace de compétiteurs comme les mesures de Radon. Il sert aussi à exemplifier comment ce cadre peut être appliqué en domaines divers qui seront étudiés dans cette thèse. Dans la suite, je listerai les contributions de ce travail.

2.1. Approximation 1-dimensionnelle des mesures dans les espaces de Wasserstein

Depuis les travaux de Kantorovitch, le domaine du transport optimal a fleuri avec plusieurs nouveaux développements, à la fois théoriques et aussi en divers domaines d'applications. Un avancement majeur a été la définition des distances de Wasserstein à travers de la fonction valeur du problème de transport optimal dont le coût est donné par une distance, voir les notes bibliographiques dans [Villani, 2009, Chap. 6] pour une discussion historique détaillée. En effet, si $c(x,y) = |x-y|^p$ dans $\mathbb{R}^d$, pour $1 \leq p \leq +\infty$ on peut définir la p-distance de Wasserstein entre deux mesures $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$, l'espace de mesures de probabilité à p-moment fini, voir 4.3:

$$W_p^p(\mu,\nu) \stackrel{\text{def.}}{=} \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^d \mathrm{d}\gamma. \tag{0.16}$$

On peut montrer que le nouveau espace $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ est aussi un espace Polonais et la topologie induite par la distance W_p est très similaire à la convergence étroite des mesures, et coïncide avec celle-ci quand l'espace ambient est compact. Cela a motivé l'utilisation de cette distance comme un terme d'attache aux données dans plusieurs applications. Par exemple, dans [Lebrat et al., 2019, Chauffert et al., 2017], les auteurs proposent une méthode basée sur le transport optimal pour projeter les images dans un espace de mesures dont le support est de dimension réduite, comme une courbe ou un nuage de points.

Ces méthodes sont paramétriques par nature, une fois que la classe de minimiseurs peut être décrite par des courbes, cela nous a motivés à proposer un problème non-paramétrique pour l'approximation des mesures avec des structures de dimension 1. Autrement dit, soit $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$, nous cherchons à l'approcher avec une mesure uniformément distribuée parmi un ensemble 1-dimensionnel. Pour cela, nous considérons le problème variationnel suivant

$$\inf_{\Sigma \text{ fermé et connexe}} W_p^p\left(\rho_0, \frac{1}{\mathscr{H}^1(\Sigma)} \mathscr{H}^1 \llcorner \Sigma\right) + \Lambda \mathscr{H}^1(\Sigma). \tag{$W\mathscr{H}^1$}$$

La distance de Wasserstein fonctionne comme un terme d'attache aux données et la mesure d'Hausdorff 1-dimensionnelle, $\mathscr{H}^1(\Sigma)$, pénalise la longueur et force les compétiteurs avec énergie finie à être 1-rectifiable, une fois qu'ils sont connexes. Notons que sans cette régularisation, on pourrait rendre la distance de Wasserstein aussi petite qu'on veut en

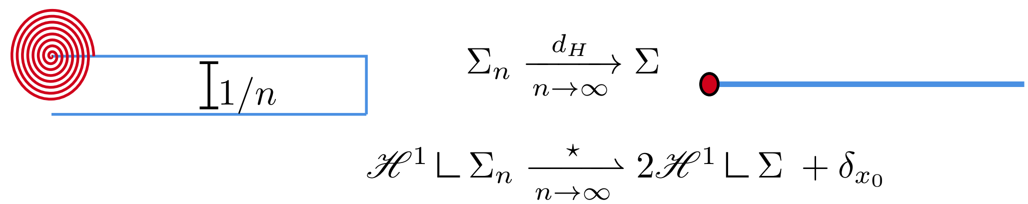

Figure 2: Effets de concentration dans la convergence faible des mesures. Dans cette image, Σ_n est formé par deux lignes horizontales qui deviennent de plus en plus proches, connectées par une verticale de longueur $1/n$, et une spirale qui converge très rapidement vers un point. La convergence au sens de la distance d'Hausdorff des ensembles donne une ligne droite horizontale comme limite, on perd l'information sur la masse totale.

prenant une courbe qui remplisse tout l'espace (en anglais *space filling curve*). Sans la contrainte de connectivité on pourrait approcher n'importe quelle mesure de probabilité arbitrairement avec une somme des masses de Dirac, en gardant la longueur totale nulle.

Pour démontrer l'existence de miniseurs pour ($W\mathscr{H}^1$), la topologie naturelle pour travailler avec ces ensembles est celle d'Hausdorff définie comme

$$d_H(A,B) \overset{\text{def.}}{=} \max\left\{\sup_{a\in A}\operatorname{dist}(a,B), \sup_{b\in B}\operatorname{dist}(b,A).\right\} \tag{0.17}$$

En revanche, nous ne pouvons pas nous en servir de la Méthode Directe car la classe de mesures de la forme $\mathscr{H}^1 \llcorner \Sigma$ n'est pas fermée à cause des effets de concentration illustrés dans la Figure 2.

Pour cela nous proposons la relaxation suivante du problème ($W\mathscr{H}^1$)

$$\inf_{\nu\in\mathscr{P}(\Omega)} W_p^p(\rho_0,\nu) + \Lambda\mathcal{L}(\nu), \text{ où } \mathcal{L}(\nu) \overset{\text{def.}}{=} \inf_{\alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu} \alpha, \tag{$\overline{W\mathscr{H}^1}$}$$

ensuite nous démontrons que cette fonctionnelle $\mathcal{L}$ est bien la relaxation s.c.i. de la fonctionnelle

$$\ell(\nu) \overset{\text{def.}}{=} \begin{cases} \mathscr{H}^1(\Sigma), & \text{si } \nu = \dfrac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma \text{ pour } \Sigma \text{ fermé et connexe,} \\ +\infty, & \text{sinon.} \end{cases}$$

Pour calculer cette relaxation, l'ingrédient principal est une version à densité du Théorème de Gołab. Dans [Gołąb, 1928], il a été démontré que la longueur est s.c.i. pour des suites d'ensembles connexes convergentes pour la distance d'Hausdorff, c'est-à-dire si $(\Sigma_n)_{n\in\mathbb{N}}$ est une suite d'ensembles compacts et connexes telle que $\Sigma_n \xrightarrow[n\to\infty]{d_H} \Sigma$, alors

$$\mathscr{H}^1(\Sigma) \leq \liminf_{n\to\infty} \mathscr{H}^1(\Sigma_n). \tag{0.18}$$

Dans [Ambrosio and Tilli, 2004, Paolini and Stepanov, 2013] une version à densité de ce résultat est démontré. Sous les mêmes conditions, si $\mathscr{H}^1 \llcorner \Sigma_n$ converge vers une mesure

μ, alors $\mu \geq \mathscr{H}^1 \llcorner \Sigma$. Nous avons généralisé davantage ce résultat en relaxant la notion de convergence des ensembles pour prendre en compte la notion plus faible de convergence de Kuratowski, voir Chapitre 1 pour une définition précise, en permettant que la suite d'ensembles soit à longueur localement finie.

Theorem 0.7 (Version à densité du Théorème de Gołab)**.** *Soit une suite $(\Sigma_n)_{n\in\mathbb{N}}$ de sous-ensembles de $\mathbb{R}^d$ connexes et fermés, qui converge dans le sens de Kuratowski vers un ensemble connexe et fermé Σ, tel que Σ est à longueur localement uniformément finie,* i.e. *pour tout $R > 0$*

$$\sup_{n\in\mathbb{N}} \mathscr{H}^1(\Sigma_n \cap B_R(x_0)) < +\infty.$$

Définissons la suite de mesures $\mu_n \stackrel{\text{def.}}{=} \mathscr{H}^1 \llcorner \Sigma_n$, et soit un point d'accumulation μ dans la topologie faible-$\star$. Alors $\operatorname{supp} \mu \subset \Sigma$ et nous avons que

$$\mu \geq \mathscr{H}^1 \llcorner \Sigma,$$

au sens des mesures.

Avec l'hypothèse plus forte de convergence au sens d'Hausdorff, pour laquelle ce résultat était déjà connu par exemple dans [Paolini and Stepanov, 2013], on peut démontrer que $\mathcal{L} = \overline{\ell}$ et conclure que ($\overline{W\mathscr{H}^1}$) est bien la relaxation s.c.i. du problème d'origine ($W\mathscr{H}^1$). L'avantage de permettre des ensembles possiblement non bornés nous permet d'utiliser ce résultat dans l'étude des familles de blowup des solutions du problème relaxé. Nous démontrons ainsi que les solutions du problème relaxé sont absolument continues par rapport à $\mathscr{H}^1$, dès que la mesure originale le soir aussi, et si $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$ ne donne pas la masse aux ensembles 1-rectifiables, n'importe quelle solution du problème relaxé aura une densité par rapport à $\mathscr{H}^1$ constante, étant alors une solution du problème de départ.

Theorem 0.8. *Soit $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$, $\Lambda > 0$. Alors ($\overline{W\mathscr{H}^1}$) admets une solution ν, et il existe une $\Lambda_\star \geq 0$ tel que si $\Lambda > \Lambda_\star$, ν est une masse de Dirac. Pour $\Lambda < \Lambda_\star$, ν est concentrée sur un ensemble $\Sigma \in \mathcal{A}$ et nous avons les propriétés suivantes*

1. *Si ρ_0 est absolument continue par rapport à $\mathscr{H}^1$, ou admet une densité L^∞ par rapport à $\mathscr{H}^1$, alors le même est vérifié pour ν.*

2. *Si ρ_0 ne donne pas la masse aux ensembles 1-rectifiables, alors ν est uniformément distribuée dans un ensemble connexe Σ à longueur finie, qui sera alors une solution du problème d'origine ($W\mathscr{H}^1$).*

Ensuite, nous étudions des propriétés qualitatives des réseaux optimales Σ. D'abord nous démontrons un résultat de régularité Ahlfors des minimiseurs.

Theorem 0.9. *Supposons que $\rho_0 \in L^{\frac{d}{d-1}}(\mathbb{R}^d)$, est Λ est suffisamment petit, de forme que les minimiseurs ν de ($\overline{W\mathscr{H}^1}$) ne sont pas des masses de Dirac. Alors $\Sigma = \operatorname{supp} \nu$ est Ahlfors régulier, i.e. il existe r_0 qui dépend de d, p, ρ_0 et $\mathcal{L}(\nu)$, ainsi que C qui dépend de d, p tels que pour tout $x \in \Sigma$ et $r \leq r_0$ nous avons que*

$$r \leq \mathscr{H}^1(\Sigma \cap B_r(x)) \leq Cr.$$

Ensuite nous passons à étudier quand les solutions sont-elles des arbres. Soit un ensemble connexe Σ on dit que $\Gamma \subset \Sigma$ est une boucle de Σ s'il est une image homomorphe de $\mathbb{S}^1$; on dit que Σ s'il n'a pas de boucles.

Theorem 0.10. *Supposons que* $\rho_0 = \sum_{i=1}^{N} a_i \delta_{x_i}$, *où* $a_i > 0$ *pour tout* $i = 1, \ldots, N$ *et* $\sum_{i=1}^{N} a_i = 1$ *les réseaux optimaux* Σ *sont des arbres.*

2.2. Approximation de champ de phase des problèmes variationnels 1D

Avoir l'accès aux méthodes numériques fiables pour le problème ($W\mathscr{H}^1$) est particulièrement important pour mieux comprendre ces propriétés. En revanche, comme la classe des ensembles connexes avec longueur finie est non paramétrique, il se trouve difficile de représenter ces objets numériquement. Une approche populaire dans la littérature pour traiter les problèmes variationnels géométriques consiste à approcher ces ensembles avec les sous-ensembles de niveau d'une fonction Sobolev, qui peut être représentée numériquement de façon efficace. Autrement dit, on cherche une approximation d'un ensemble 1D Σ de la forme

$$\Sigma \approx \{\varphi_\varepsilon \leq \varepsilon^s\},$$

où φ_ε appartient à un espace approprié.

Cette approximation doit être faite de façon que ces fonctions Sobolev, appelées *champs de phase* une fois qu'elles sont conçues pour approximer deux phases quand elles sont proches de 0 et 1, peuvent aussi être utilisées pour approcher une quantité géométrique d'intérêt. L'approche de [Modica and Mortola, 1977] était d'approcher le périmètre d'un ensemble avec une fonctionnelle elliptique. Plus tard, Ambrosio et Tortorelli ont proposé une méthode similaire pour étudier le problème de Mumford-Shah, voir [Ambrosio et al., 2000, Chap. 6]. Nous travaillons avec une modification de la fonctionnelle de Ambrosio-Tortorelli, pareille à ce qu'a été proposé dans [Chambolle et al., 2019b, Chambolle et al., 2019a]: pour une fonction φ_ε telle que $\Sigma \approx \{\varphi_\varepsilon \leq \varepsilon^s\}$, nous proposons l'approximation de sa longueur comme

$$\mathscr{H}^1(\Sigma) \approx \mathcal{AT}_p(\varphi_\varepsilon) \stackrel{\text{def.}}{=} \frac{1}{\Lambda_{p,d}} \int_\Omega \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathrm{d}x.$$

La difficulté majeure est en fait d'incorporer la constrainte de connectivité dans la formulation diffuse. Pour cela, on considère le cas $p > d \geq 2$, de forme que les champs de phase tels que l'énergie $\mathcal{AT}_p$ est finie appartiennent à l'espace de Sobolev $W^{1,p}$ et donc sont Hölder continues. Cela nous permet de contrôler les sous-ensembles de niveau et à avoir une bonne synergie avec la *fonctionnelle de connectivité diffuse* $\mathcal{C}_\varepsilon(\varphi_\varepsilon)$ proposée par

Dondl et Wojtowytsch, voir e.g. [Dondl et al., 2017]. Cette fonctionnelle est projetée pour mesurer le niveau de dis-connectivité de l'ensemble de niveau $\{\varphi_\varepsilon \leq \varepsilon^s\}$:

$$\mathcal{C}_\varepsilon(\varphi_\varepsilon) \stackrel{\text{def.}}{=} \int_{\Omega\times\Omega} \beta_\varepsilon(\varphi_\varepsilon(x))\beta_\varepsilon(\varphi_\varepsilon(y))d^{F_\varepsilon\circ\varphi_\varepsilon}(x,y)\mathrm{d}x\mathrm{d}y.$$

La fonction β_ε sélectionne seulement les ensembles de niveau petits de φ_ε, alors que la distance géodésique $d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}$ est donnée par

$$d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}(x,y) \stackrel{\text{def.}}{=} \inf\left\{\int_K F_\varepsilon\circ\varphi_\varepsilon(x)\mathrm{d}\mathscr{H}^1(x) : \begin{array}{c} K \text{ connexe, } x,y\in K \\ \mathscr{H}^1(K)\leq\varepsilon^{-1}\end{array}\right\},$$

où $F_\varepsilon(z)$ assume la valeur 0 pour $z \leq \varepsilon^s$, de forme que $d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}(x,y) = 0$ si x et y sont dans la même composante connexe de $\{\varphi_\varepsilon \leq \varepsilon^s\}$.

La relation entre ces deux fonctionnelles est clé pour les résultats d'approximation diffuse proposés dans cette thèse et s'est montrée suffisamment flexible pour traiter une variété de problèmes d'optimisation de forme 1-dimensionnels. En particulier, nous exploitons un lien entre le problème ($W\mathscr{H}^1$) et le problème de *minimisation de la distance moyenne* introduit par dans [Buttazzo and Stepanov, 2003], qui peut être décrit comme : soit une densité démographique donnée par une mesure ρ_0 dans une région Ω, que peut être par exemple une ville, on cherche à construire un réseau de métro Σ de formes à minimiser la distance moyenne de chaque citoyen aux réseaux. Telle problème s'écrit comme

$$\inf_{\Sigma\text{ connexe}} \int_\Omega \mathrm{dist}(x,\Sigma)^q\mathrm{d}\rho_0(x) + \mathscr{H}^1(\Sigma). \tag{ADM}$$

En remarquant que $\int_\Omega \mathrm{dist}(x,\Sigma)^q\mathrm{d}\rho_0(x) = \inf_{\operatorname{supp}\nu\subset\Sigma} W_q^q(\rho_0,\nu)$, nous proposons une approche unifiée pour approcher les deux problèmes.

Le problème (ADM) est approché par la fonctionnelle de *distance moyenne diffuse*:

$$\mathcal{AD}_\varepsilon(\nu_\varepsilon,\varphi_\varepsilon) \stackrel{\text{def.}}{=} \begin{cases} W_q^q(\rho_0,\nu_\varepsilon) + \Lambda\mathcal{AT}_p(\varphi_\varepsilon) + \dfrac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) & \nu_\varepsilon\in\mathscr{P}(\Omega), \\ \qquad\qquad + \dfrac{1}{\varepsilon^\ell}\displaystyle\int_\Omega \varphi_\varepsilon\mathrm{d}\nu_\varepsilon, & \varphi_\varepsilon\in 1+W_0^{1,p}(\Omega) \\ +\infty, & \text{sinon.}\end{cases}$$

Concernant le problème ($\overline{W\mathscr{H}^1}$), chaque terme fonctionne de façon similaire, sauf pour le terme d'Ambrosio-Tortorelli une fois que nous cherchons à approcher $\mathcal{L}(\nu)$ au lieu de $\mathscr{H}^1(\Sigma)$. La fonctionnelle proposée dévient

$$\mathcal{WH}_\varepsilon^1(\alpha_\varepsilon,\nu_\varepsilon,\varphi_\varepsilon) \stackrel{\text{def.}}{=} \begin{cases} W_q^q(\rho_0,\nu_\varepsilon) + \Lambda\alpha_\varepsilon + \dfrac{1}{\varepsilon}\left\|\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon\right\|_{L^2(\Omega)}^2 & \alpha_\varepsilon\geq 0, \\ & \nu_\varepsilon\in\mathscr{P}(\Omega) \\ \qquad + \dfrac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) + \dfrac{1}{\varepsilon^\ell}\displaystyle\int_\Omega\varphi_\varepsilon\mathrm{d}\nu_\varepsilon, & \varphi_\varepsilon\in 1+W_0^{1,p}(\Omega), \\ +\infty, & \text{otherwise,}\end{cases}$$

où la mesure $\mu_\varepsilon = \mu_\varepsilon(\varphi_\varepsilon)$ est la mesure de transition diffuse définie comme

$$\mu_\varepsilon \stackrel{\text{def.}}{=} \frac{1}{\Lambda_{p,d}} \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathcal{L}^d \llcorner \Omega. \tag{0.19}$$

La notion d'approximation choisie est celle de la Γ-convergence. Elle a été introduite par De Giorgi, voire [Dal Maso, 1993] et le Chapitre 1 de cette thèse pour une introduction, pour ses propriétés pertinentes aux problèmes variationnels. En effet, si $(F_\varepsilon)_{\varepsilon>0}$ converge au sens de la Γ-convergence vers une fonctionnelle F, et $x_\varepsilon \in \operatorname{argmin} F_\varepsilon$ est une suite convergente vers x, alors cette limite minimise aussi la Γ-limite, $x \in \operatorname{argmin} F$.

Les résultats précis démontrés dans cette thèse concernant l'approximation de champ de phase sont décrits ensuite.

Theorem 0.11. *Supposons que $p > d \geq 2$, $\ell > s$ et $\kappa > (2d+1)\frac{(s+1)p-d+1}{p-d}$, alors*

- *la fonctionnelle de distance moyenne diffuse approche* (ADM)

$$\mathcal{AD}_\varepsilon \xrightarrow[\varepsilon\to 0]{\Gamma} \mathcal{AD}(\nu,\varphi) \stackrel{\text{def.}}{=} \begin{cases} W_q^q(\rho_0,\nu) + \Lambda \mathscr{H}^1_S(\operatorname{supp}\nu), & \nu \in \mathscr{P}(\Omega),\ \varphi \equiv 1, \\ +\infty, & \textit{sinon}, \end{cases}$$

  où $\mathscr{H}^1_S(\operatorname{supp}\nu)$ correspond à la longueur de l'arbre de Steiner minimale que connecte $\operatorname{supp}\nu$. La Γ-convergence se passe dans la topologie forte de L^2 et la topologie faible de $\mathscr{P}(\Omega)$.

  De plus, soit $(\nu_\varepsilon, \varphi_\varepsilon)_{\varepsilon>0}$ une famille de minimiseurs de $\mathcal{AD}_\varepsilon$, elle admet un point d'accumulation $(\nu, \varphi \equiv 1)$ qui minimise

$$\min_{\Sigma} (\text{ADM}) = \min_{(\nu,\varphi)} \mathcal{AD}(\nu,\varphi),$$

  nous avons aussi les propriétés suivantes

  - *Σ est un minimiseur de* (ADM) *si, et seulement si, il s'agit d'un arbre d'Steiner minimal pour $\operatorname{supp}\nu$, pour une certaine mesure ν qui minimise $\mathcal{AD}$;*
  - *ν minimise $\mathcal{AD}$ si, et seulement si, il peut s'écrire comme $\nu = (\pi_\Sigma)_\sharp \rho_0$, où π_Σ est une sélection mesurable de l'opérateur de projection sur une certain Σ qui minimise* (ADM).

- *Pour le problème ($\overline{W\mathscr{H}^1}$), nous avons que*

$$\mathcal{WH}^1_\varepsilon \xrightarrow[\varepsilon\to 0]{\Gamma} \mathcal{WH}^1(\alpha,\nu,\varphi) \stackrel{\text{def.}}{=} \begin{cases} W_q^q(\rho_0,\nu) + \Lambda \mathcal{L}(\nu), & \nu \in \mathscr{P}(\Omega),\ \alpha \geq \mathcal{L}(\nu), \\ +\infty, & \textit{sinon}, \end{cases}$$

  la Γ-convergence se passe dans la topologie de $\mathbb{R}$, la topologie forte de L^2 et la topologie faible de $\mathscr{P}(\Omega)$.

De plus, en supposant que ρ_0 ne charge pas les ensembles $\mathscr{H}^1$-rectifiables, si $(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon)_{\varepsilon>0}$ est une suite de minimiseurs de $\mathcal{WH}^1_\varepsilon$, alors il admet un point d'accumulation $(\alpha, \nu, \varphi \equiv 1)$ de la forme

$$\alpha = \mathscr{H}^1(\Sigma),\ \nu = \frac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma,\ \textit{où } \Sigma \textit{ est connexe } \mathscr{H}^1\textit{-rectifiable},$$

de façon que Σ minimise ($W\mathscr{H}^1$).

2.3. Le flot de gradient Wasserstein de la variation totale

Dans le travail séminal [Jordan et al., 1998] les auteurs proposent une interprétation variationnelle pour l'équation de Fokker-Planck comme un flot de gradient de l'entropie dans la topologie induite par la distance de Wasserstein-2. Il y a eu plusieurs avancements depuis, notamment dans [Ambrosio et al., 2008], voir aussi [Santambrogio, 2015, Chap. 8]. Décrivons cette structure de flot de gradient: soit $\Omega \subset \mathbb{R}^d$ convexe et compact, et une fonctionnelle $\mathscr{F} : \mathscr{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$, le *schéma JKO* consiste à

$$\rho_{k+1} \in \underset{\rho \in \mathscr{P}(\Omega)}{\operatorname{argmin}}\, \mathscr{F}(\rho) + \frac{1}{2\tau} W_2^2(\rho_k, \rho),\ \text{pour } \rho_0 \text{ donné.} \tag{JKO}$$

Une fois que cette suite $(\rho_k)_{k\in\mathbb{N}}$ est obtenue, on peut définir une courbe, dépendant du paramètre τ, dans l'espace de Wasserstein comme

$$\rho_\tau(t) \overset{\text{def.}}{=} \rho_k, \qquad \text{if } t \in [k\tau, (k+1)\tau).$$

En prenant la limite quand $\tau \to 0$, on obtient une courbe dans l'espace de Wasserstein que résout l'EDP suivante

$$\begin{cases} \partial_t \rho(t) = \operatorname{div}\left(\rho \nabla \dfrac{\delta \mathscr{F}}{\delta \rho}(\rho(t))\right), & \text{dans } [0,T] \times \Omega \\ \dfrac{\partial \rho}{\partial n} = 0, & \text{sur } \partial\Omega \\ \rho(0) = \rho_0. \end{cases}$$

Comme susmentionné, [Jordan et al., 1998] a démontré la convergence de ce schéma quand on considère $\mathscr{F}(\rho) = \displaystyle\int_\Omega \rho \log \rho \mathrm{d}x$, et l'équation limite obtenue dévient l'équation de la chaleur, un résultat de grande importance philosophique une fois que l'équation de la chaleur est formellement interprété comme un flot qui cherche à minimiser l'entropie.

D'un autre côté, l'utilisation des flots de gradient est une technique standard pour le traitement d'image. En particulier, la fonctionnelle de *variation totale* définie comme

$$\mathrm{TV}(u) \overset{\text{def.}}{=} \sup\left\{ \int_\Omega \operatorname{div} z(x) u(x) \mathrm{d}x : z \in C_c^1\left(\Omega; \mathbb{R}^N\right), \|z\|_\infty \leq 1 \right\},$$

est connue par ses propriétés de préservation des contours et permet une reconstruction parcimonieuse des images bruitées, voire [Chambolle et al., 2010, Chambolle et al., 2016]. En effet, depuis le travail séminal [Rudin et al., 1992] le flot de la variation totale dans la topologie L^2 est devenu un benchmark pour le débruitage d'image, et maintenant connu comme le problème de *Rudin-Osher-Fatemi* et peut s'écrire comme

$$\inf_{u \in L^2(\Omega)} \mathrm{TV}(u) + \frac{1}{2\lambda} \|u - g\|^2_{L^2(\Omega)} . \tag{ROF}$$

Récemment, le flot de gradient de TV dans la topologie Wasserstein a été étudié pour le traitement d'image [Burger et al., 2012, Benning et al., 2013, Carlier and Poon, 2019]. Dans cette thèse, nous revenons au travail de Carlier & Poon [Carlier and Poon, 2019] et dérivons les équations de Euler-Lagrange pour ce problèmes

$$\inf_{\rho \in \mathscr{P}(\Omega)} \mathrm{TV}(\rho) + \frac{1}{2\tau} W_2^2(\rho_0, \rho). \tag{TV-W}$$

Différemment de [Carlier and Poon, 2019], qu'utilise la version du même problème avec une régularisation entropique, notre approche pour dériver les conditions d'optimalité pour (TV-W) consiste à faire le lien entre ceci et un problème du typo (ROF) opportun, pour lequel les conditions d'optimalité sont bien comprises. Ainsi nous pouvons déduire plus de régularité pour les quantités impliquées, ce qui est crucial pour comprendre la limite de ce schéma de flot de gradient (JKO).

Theorem 0.12. *Soit $\Omega \subset \mathbb{R}^d$ un domaine compact et convexe. Pour tout $\rho_0 \in L^1(\Omega) \cap \mathscr{P}(\Omega)$, soit ρ_1 l'unique minimiseur de* (TV-W). *Nous avons les propriétés suivantes*

1. *Il existe un champ vectoriel $z \in H^1_0(\mathrm{div}; \Omega) \cap L^\infty(\Omega; \mathbb{R}^d)$ et un multiplicateur de Lagrange $\beta \geq 0$ tels que*

$$\begin{cases} \mathrm{div}\, z + \dfrac{\psi_1}{\tau} = \beta, & \textit{a.e. dans } \Omega \\ z \cdot \nu = 0, & \textit{sur } \partial\Omega \\ \beta \rho_1 = 0, & \textit{a.e. dans } \Omega \\ \mathrm{div}\, z \in \partial\, \mathrm{TV}(\rho_1), & \end{cases} \tag{0.20}$$

   où ψ_1 est le potentiel de Kantorovitch associé à ρ_1.

2. *Le multiplicateur de Lagrange β est l'unique solution du problème* (ROF) *avec $\lambda = 1$ et $g = \psi_1/\tau$.*

3. *Les fonctions* $\mathrm{div}\, z, \psi_1$ *et β sont Lipschitz continues.*

Par la suite, nous proposons un algorithme du type Douglas-Rachford pour l'optimisation numérique du même problème. Nous validons notre algorithme avec un exemple pour lequel nous avons une solution théorique explicite et ensuite nous l'appliquons à un problème de reconstruction d'une image quantifiée.

2.4. De Nash à Cournot-Nash

Une des questions centrales en théorie des jeux et, en particulier, les jeux à champ moyen, est la suivante:

> *Donné un échantillon de joueurs suivant une distribution continue, quand est-ce qu'une suite d'équilibres de Nash pour une famille de jeux finis associée converge vers une notion d'équilibre pour un jeu avec une infinité de joueurs?*

Dans ce travail nous répondons à cette question dans le contexte des *équilibres de Cournot-Nash.*

Soit deux espaces Polonais $\mathcal{X}$ et $\mathcal{Y}$, on peut interpréter $\mathcal{X}$ comme l'espace des types de joueurs distribués selon une mesure de probabilité $\mu \in \mathscr{P}(\mathcal{X})$, et $\mathcal{Y}$ comme l'espace de stratégies admissibles pour ceux-ci, distribués selon $\nu \in \mathscr{P}(\mathcal{Y})$. Dans ce contexte, un couplage $\gamma \in \Pi(\mu, \nu)$ représente la distribution jointe de joueurs et stratégies. On considère aussi une fonction $\Phi : \mathcal{X} \times \mathcal{Y} \times \mathscr{P}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$, telle que $\Phi(x, y, \nu)$ dénote le coût qu'un joueur de type x choisisse l'stratégie y, dans un champ moyen d'stratégies donné par la distribution ν. On dit alors qu'un couplage $\gamma \in \Pi(\mu, \nu)$ est un équilibre du type Cournot-Nash si

$$\gamma\left(\left\{(x,y) \in \mathcal{X} \times \mathcal{Y} : y \in \underset{y' \in \mathcal{Y}}{\operatorname{argmin}}\, \Phi(x, y', \nu)\right\}\right) = 1, \tag{0.21}$$

Des résultats d'existence d'équilibres ont été établis avec des méthodes du type point fixe dans les travaux [Schmeidler, 1973, Mas-Colell, 1984]. Cette approche dépend fortement la continuité de la fonction coût. En se basent sur le travail de [Blanchet and Carlier, 2016], si Φ est composé par une partie de coût individuel plus un terme d'interaction moyenne deux-à-deux, ayant la forme

$$\Phi(x, y, \nu) = c(x, y) + \int_{\mathcal{Y}} L(y) \mathrm{d}\nu(y) + \int_{\mathcal{Y}} H(y, y') \mathrm{d}\nu(y'), \tag{0.22}$$

nous caractérisons les équilibres de Cournot-Nash de ce jeu comme les points extrémaux de la fonctionnelle potentielle suivante

$$\mathcal{J}(\gamma) \overset{\text{def.}}{=} \begin{cases} \displaystyle\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\gamma + \int_{\mathcal{Y}} L(y)\mathrm{d}\nu(y) + \int_{\mathcal{Y}\times\mathcal{Y}} H(y,y')\mathrm{d}\nu\otimes\nu, & \text{if } \gamma \in \Pi(\mu,\nu), \\ +\infty & \text{if } \gamma \notin \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y}). \end{cases}$$

Dans la littérature de la théorie des jeux, on dit que ce type de problème présente une structure potentielle, au sens que des équilibres peuvent être trouvés en minimisant cette fonction potentielle.

Avec cette observation, on propose une suite des jeux à N-joueurs, admettant aussi une fonction de potentiel $\mathcal{J}_N$, dont les minimiseurs sont des équilibres de Nash pour les jeux associés. Nous montrons ensuite que cette suite de fonctions de potentiel Γ-converge vers $\mathcal{J}$. En particulier, cela implique qu'une suite d'équilibres de Nash obtenue comme des minimiseurs de $\mathcal{J}_N$ converge vers un minimiseur de $\mathcal{J}$, et alors vers un équilibre de type Cournot-Nash. Ainsi répondant positivement à la question qu'on s'intéresse pour une classe assez large de jeux.

Publications. Cette thèse apporte les publications et pre-publications suivantes

- *The Total Variation-Wasserstein Problem*, collaboration avec Antonin Chambolle et Vincent Duval, publié dans GSI'23
- *1D approximation of measures in Wasserstein spaces*, collaboration avec Antonin Chambolle et Vincent Duval, preprint
- *Phase-field approximation for 1-dimensional shape optimization problems*, preprint
- *From Nash to Cournot-Nash via* Γ*-convergence*, collaboration avec Guilherme Mazanti and Laurent Pfeiffer, en praparation.

Presentations et Distinctions Les travaux de cette thèse ont été présentés dans le conférences suivantes

- Les résultats du Chapitre 3 ont été présentés sur le format de poster dans la conférence "Lantin American Congress of Industrial and Applied Mathematics" (LACIAM), en Janvier de 2023 à Rio de Janeiro **(Prix de meilleur poster)**
- Les résultats des Chapitres 3 et 4 ont été présentés sur le format de 'contributed talk' dans la conférence "Calculus of Variations and Applications", en Juin de 2023 à Paris.
- Les résultats du Chapitre 6 ont été publiés et présentés dans la conférence "Geometric Science of Information" (GSI 2023) en Août de 2023 à Saint-Malo.
- Les résultats du Chapitre 7 ont été présentés dans les Journées SMAI-MODE en mars de 2024 à Lyon. **(Laureat du Prix Dodu)**

3. Organisation de cette thèse

Ensuite nous détaillons l'organisation de la thèse présente.

- Dans le Chapitre 1 nous introduisons les théories mathématiques venant du Calcul des Variations employées dans cette thèse. En commençant notre discussion par les outils standards pour l'étude des problèmes problèmes variationnels : la Méthode Directe, les relaxations et la Γ-convergence. Ensuite, nous discutons des topologies utilisées le long du travail, décrivant les notions de convergence des ensembles et des mesures. Les outils de la théorie géométrique de la mesure que nous aurons besoin sont détaillés et nous finissons avec une discussion plus approfondie du transport optimal et ses propriétés.

- Le Chapitre 2 est dédié à la théorie des continuums, c'est-à-dire les ensembles connexes avec mesure $\mathscr{H}^1$ finie. On profite de l'occasion pour collecter les résultats classiques de la littérature et mettre à valeur des généralisations obtenues dans cette thèse pour nos objectifs principaux des Chapitres suivants.

- Dans le Chapitre 3 on introduit le problème Wasserstein-$\mathscr{H}^1$ et sa régularisation, en passant par la fonctionnelle de longueur dans la section 2.1. Nous discutions en détails les conditions sur le paramètre de régularisation Λ pour que le problème ait une solution non triviale, c'est-à-dire qu'elle ne soit pas une masse de Dirac. Ensuite, nous démontrons que les solutions du problème relaxé sont des mesures rectifiables, dès que la mesure optimale soit absolument continue par rapport à $\mathscr{H}^1$. Finalement, on passe à la preuve d'existence pour le problème du départ avec un argument du type blow-up.

- Dans le Chapitre 4 on continue avec l'étude du problème Wasserstein-$\mathscr{H}^1$, en s'intéressant notamment aux propriétés qualitatives des minimiseurs. D'abord on montre que si la mesure originale ρ_0 est suffisamment intégrable (appartient à $L^{\frac{d}{d-1}}(\mathbb{R}^d)$), alors toute solution du problème relaxé est Ahlfors régulière. Si ρ_0 est une mesure atomique, on démontre que les minimiseurs sont des arbres, ils ne contiennent pas des parties homeomorphes à $\mathbb{S}^{d-1}$.

- Dans le Chapitre 5 on étudie une approximation les approximations de champ de phase pour les problème d'optimisation de forme 1D décrits dans la section 2.2

- Dans le Chapitre 6 on s'intéresse aux flots de gradient de la fonctionnelle de variation totale dans la topologie de Wasserstein. On dérive les équations d'Euler Lagrange en se servant des conditions d'optimalité bien comprises d'un problème de Rundin-Osher-Fatemi approprié, et on montre que tous les ensembles de niveau sont des solutions du même problème de courbure prescrite. Ensuite, nous proposons un algorithme proximal pour sa résolution numérique, qu'est validé d'abord avec un exemple dont nous connaissons la solution explicite.

- Dans le Chapitre 7 nous démontrons qu'une suite des équilibres de Nash d'une famille de jeux à N-joueurs converge vers les équilibres du type Cournot-Nash en exploitant une structure potentielle particulière. Celle-ci nous permet d'obtenir une fonction potentielle dont la minimisation nous donne des équilibres. On montre que la suite des fonctions potentielles pour les jeux à N-joueurs Γ-converge Avec probabilité de 1 vers la fonctionnalité de potentiel du jeu avec un continuum de joueurs.

Chapter 1

Calculus of Variations, Optimal Transport and Geometric Measure Theory

Contents

1. Abstract variational problems

1.1. The Direct Method and relaxations

The Direct Method discussed in the introduction of this thesis might be by now the most used tool to prove existence of minimizers to variational problems. It is in fact very simple and stems from the well-known Weierstrass' extreme value Theorem, which states that any lower semi-continuous function defined over the reals admits a point that attains its infimum over closed and bounded intervals.

During Weierstrass' time mathematics was going through a transition from calculus to analysis, it was a moment when the mathematical community started questioning the foundations of the field [Gray, 2015]. After the establishment of other areas as fields of study in mathematics, such as Measure Theory and Topology, the Direct method is a dissection of the fundamental topological properties that are used in the proof of Weierstrass' theorem, namely sequential compactness and lower semi-continuity.

In this sense, the Direct Method turned the matter of existence of minimizers into a topological question and the field profited from the developments of functional analysis results regarding compactness of various abstract spaces, for instance Blaschke's compactness for the Hausdorff distance 2.6, Prokhorov's Theorem 1.5, Arsela-Ascoli's compactness for uniform convergence, Rellich-Kondrachev [Evans, 2022, Chap. 5.7], Ambrosio's compactness Theorem for SBV spaces [Ambrosio et al., 2000, Thm. 4.8].

The lower semi-continuity is not trivial either and a lot of effort was done to characterize such functionals, see for instance [Dacorogna, 2008]. Whenever the functional we wish to minimize is not lower semi-continuous, as many of the examples in the introduction, the alternative is to look at the relaxed formulation of the problem. This can be done systematically by considering the *lower semi-continuous relaxation* or *l.s.c. envelope* of a functional $\mathscr{F}$ one wishes to minimize, which is defined as the *biggest lower semi-continuous function* that is smaller than $\mathscr{F}$, see [Attouch et al., 2014, Def. 3.2.2]. Given $\mathscr{F} : \mathcal{X} \mapsto \mathbb{R} \cup \{+\infty\}$, it can be proven that its relaxation is characterized by

$$\overline{\mathscr{F}}(x) \overset{\text{def.}}{=} \inf \left\{ \liminf_{x_n \to x} \mathscr{F}(x_n) : \ x_n \xrightarrow[n\to\infty]{} x \right\}, \tag{1.1}$$

see [Attouch et al., 2014, Prop. 3.2.6].

1.2. The notion of Γ-convergence

The notion of Γ-convergence was introduced by De Giorgi in order to have good properties concerning the limits of minimizers of variational problems, see for instance [Dal Maso, 1993].

Definition 1.1. *Let $(\mathcal{X}, d_{\mathcal{X}})$ be a complete metric space, a sequence of functionals $\mathscr{F}_n : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ Γ-converges to $\mathscr{F}$, if*

- $\underline{\Gamma - \liminf}$: *for every sequence* $x_N \xrightarrow[N\to\infty]{d_\mathcal{X}} x$ *in* $\mathcal{X}$*, it holds that*

$$\mathscr{F}(x) \leq \liminf_{N\to\infty} \mathscr{F}_N(x_N).$$

- $\underline{\Gamma - \limsup}$: *for every* $x \in \mathcal{X}$*, there is a sequence* $x_N \xrightarrow[N\to\infty]{d_\mathcal{X}} x$ *such that*

$$\limsup_{N\to\infty} \mathscr{F}_N(x_N) \leq \mathscr{F}(x),$$

$(x_N)_{N\in\mathbb{N}}$ *is called the recovery sequence of* x.

Equivalently, given a sequence of functionals $\mathscr{F}_N$, we define the lower and upper Γ limits, respectively, as

$$\begin{aligned} \Gamma\text{-}\liminf \mathscr{F}_N(x) &\overset{\text{def.}}{=} \inf\left\{\liminf_{N\to\infty} \mathscr{F}_N(x_N) \,:\, x_N \xrightarrow[N\to\infty]{} x\right\}, \\ \Gamma\text{-}\limsup \mathscr{F}_N(x) &\overset{\text{def.}}{=} \inf\left\{\limsup_{N\to\infty} \mathscr{F}_N(x_N) \,:\, x_N \xrightarrow[N\to\infty]{} x\right\}. \end{aligned} \tag{1.2}$$

From [Braides, 2002, Prop. 1.28], both Γ upper and lower limits are lower semi-continuous and $\mathscr{F}_N \xrightarrow[N\to\infty]{\Gamma} \mathscr{F}$ if and only if $\Gamma\text{-}\liminf \mathscr{F}_N = \Gamma\text{-}\limsup \mathscr{F}_N = \mathscr{F}$.

The fundamental property that makes it interesting is the fact that cluster points of minimizers of a sequence of minimizers of $\mathscr{F}_N$, which Γ-converges to $\mathscr{F}$, are minimizers of $\mathscr{F}$. Indeed, let $(x_N)_{N\in\mathbb{N}}$ be a sequence of minimizers of $(\mathscr{F}_N)_{N\in\mathbb{N}}$ converging to x. For an arbitrary $x' \in \mathcal{X}$, let x'_N be a corresponding recovery sequence, then it follows that

$$\begin{aligned} \mathscr{F}(x) &\leq \liminf_{N\to\infty} \mathscr{F}_N(x_N) && \text{by the } \Gamma - \liminf \text{ inequality} \\ &\leq \liminf_{N\to\infty} \mathscr{F}_N(x'_N) && \text{by the minimality of } x_N \\ &\leq \limsup_{N\to\infty} \mathscr{F}_N(x'_N) \leq \mathscr{F}(x') && \text{since } x'_N \text{ is a recovery sequence.} \end{aligned}$$

As x' was an arbitrary point of $\mathcal{X}$, it follows that x is a minimizer of $\mathscr{F}$.

This gives an extremely versatile tool to approach different problems with different goals. In the sequel, we shall discuss a few examples of well-known Γ-converge results that are particularly relevant to this thesis.

Geometric variational problems and phase-field approximations

One important application of Γ-convergence is in the so-called *phase-field* approximations of geometric variational problems. The notion of perimeter of a set, or in general hyper-surface of its boundary, is central in geometric problems arising from physical phenomena. Indeed, this quantity is used to model the surface tension of a region occupied by a liquid or a drop [Figalli et al., 2010], or the shape of atomic nuclei [Choksi et al., 2017].

However, it is a challenge to represent sets in a computer and solve nontrivial optimization problems over this space. A possible approach is to approximate them with

Sobolev functions and search for families of functionals that approximate the perimeter in the sense of Γ-convergence.

This was done in [Modica and Mortola, 1977], see also [Braides, 1998], where it was first proven that

$$\int_\Omega \left(\varepsilon|\nabla u|^2 + \frac{1}{\varepsilon}W(u)\right) \mathrm{d}x \xrightarrow[\varepsilon\to 0]{L^1(\Omega)-\Gamma} \mathscr{F}(u) = \begin{cases} \sigma \operatorname{Per}(E;\Omega), & \text{if } u = \mathbb{1}_E, \\ +\infty, & \text{otherwise,} \end{cases} \tag{1.3}$$

where W is a continuous function such that $W(t) = 0$ if and only if $t \in \{0,1\}$.

Since the original result, there have been many developments in this direction, see for instance the discussion on the introduction of Chapter 5 of this thesis, where we propose a new phase-field approximation result of the problem introduced in Chapter 3.

Limits of particle systems via Γ-convergence

The notion of Γ-convergence has also been used in [Serfaty, 2015] to study the asymptotic behavior of a Coulomb gas as the number of particles goes to infinity. In this statistical mechanical model, one considers a system of particles whose positions are $(x_1,\ldots,x_N) \in (\mathbb{R}^d)^N$, under the influence of an external potential V and electrostatic pair-wise interactions. The optimal configuration of the particles minimize the (rescaled) Hamiltonian

$$H_N(x_1,\ldots,x_N) \overset{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^N V(x_i) + \frac{1}{N^2}\sum_{i\neq j} g(x_i - x_j),$$

where g is the Green function of the Laplacian, so that the double sum corresponds to the classical electrostatic potential induced by this charged particle system.

Through the map $(x_1,\ldots,x_N) \mapsto \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, the Hamiltonian can be canonically extended to $\mathscr{P}(\mathbb{R}^d)$ by defining

$$\mathscr{F}_N(\mu) \overset{\text{def.}}{=} \begin{cases} H_N(x_1,\ldots,x_N), & \text{if } \mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \\ +\infty, & \text{otherwise.} \end{cases}$$

In was proved in [Serfaty, 2015] that

$$\mathscr{F}_N \xrightarrow[N\to\infty]{\mathscr{P}(\mathbb{R}^d)-\Gamma} \mathscr{F}(\mu) \overset{\text{def.}}{=} \int_{\mathbb{R}^d} V(x)\mathrm{d}\mu(x) + \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x-y)\mathrm{d}\mu\otimes\mu(x,y). \tag{1.4}$$

The limit energy turns out to be the energy minimized in the classical *capacitor problem*, solved by Frostman in [Frostman, 1935].

If the main motivation of the previous example was computational, *i.e.* obtain an approximate problem in a space that is easier to work with computationally, the convergence (1.4) is a hydrodynamic limit in the sense of statistical physics which derives rigorously the properties of a macroscopic system through the interactions of microscopic particles. Chapter 7 builds upon these ideas in order to understand the hydrodynamic limit of a sequence of N-player games as a game with a continuum of players.

2. The space of Radon measures

2.1. Radon measures

Given a Polish space $(\mathcal{X}, d_{\mathcal{X}})$, that is a complete separable space equipped with a metric topology, we define the space of signed, scalar-valued, Radon measures $\mathcal{M}(\mathcal{X})$ as the space of all regular Borel measures over $\mathcal{X}$, assuming real values over compact subsets of $\mathcal{X}$. The total variation measure induced by some $\mu \in \mathcal{M}(\mathcal{X})$ is defined as

$$|\mu|(A) \overset{\text{def.}}{=} \sup \left\{ \sum_{k=1}^{N} |\mu(E_k)| : \ (E_k)_{k=1}^{N} \text{ is a finite partition of } A \right\}. \tag{1.5}$$

It induces the so-called total variation norm $\|\mu\|_{\mathcal{M}(\mathcal{X})} \overset{\text{def.}}{=} |\mu|(\mathcal{X})$ and the space of *bounded Radon measures*

$$\mathcal{M}_b(\mathcal{X}) \overset{\text{def.}}{=} \left\{ \mu \text{ Radon measure } : \|\mu\|_{\mathcal{M}(\mathcal{X})} < +\infty \right\}$$

is a Banach space when equipped with it.

Since convergence in the total variation norm is very strong and hard to verify, it is natural to look for nice dual topologies for this space. A natural candidate is to pair $\mathcal{M}_b(\mathcal{X})$ with a space of continuous functions since the quantity

$$\langle \mu, \phi \rangle \overset{\text{def.}}{=} \int_{\mathcal{X}} \phi \mathrm{d}\mu. \tag{1.6}$$

This quantity will naturally be finite whenever we take $\phi \in C_c(\mathcal{X})$, the space of continuous functions with compact support, or its closure in the norm $\|\cdot\|_\infty$ given by the continuous functions converging to 0 at infinity

$$\overline{C_c(\mathcal{X})} = C_0(\mathcal{X}) \overset{\text{def.}}{=} \left\{ \phi : \mathcal{X} \to \mathbb{R} : \begin{array}{c} \text{for every } \varepsilon > 0, \text{ there is } K \text{ compact s.t.} \\ |f(x)| < \varepsilon \text{ for all } x \in \mathcal{X} \setminus K \end{array} \right\}.$$

This closure can be proven with Urysohn's Lemma, which is quite simple in metric spaces since we can rely on the distance, see [Rudin, 1986]. Riesz's representation Theorem states that the space $\mathcal{M}_b(\mathcal{X})$ is the dual space of $C_0(\mathcal{X})$.

Theorem 1.2. *Let $\mathcal{X}$ be a locally compact Hausdorff space. Then every bounded linear functional $L : C_0(\mathcal{X}) \to \mathbb{R}$ is uniquely represented via the duality product (1.6) by a bounded Radon measure $\mu \in \mathcal{M}_b(\mathcal{X})$. In addition, the operator norm of L is given by $\|L\|_{op} = \|\mu\|_{\mathcal{M}_b(\mathcal{X})}$.*

We are particularly interested in the space of probability measures, defined as

$$\mathscr{P}(\mathcal{X}) \overset{\text{def.}}{=} \left\{0 \leq \mu \in \mathcal{M}_b(\mathcal{X}) : \ \|\mu\|_{\mathcal{M}(\mathcal{X})} = 1\right\}$$

The previous result suggests using Banach-Alaoglu-Bourbaki theorem to exploit compactness of the unit ball of $\mathcal{M}_b(\mathcal{X})$ in the weak-$\star$ topology obtained with the duality paring with $C_0(\mathcal{X})$, however as the total variation norm is not continuous in the weak-$\star$ topology, we cannot guarantee that a cluster point of a sequence of probability measures will be in $\mathscr{P}(\mathcal{X})$. This is the case if $\mathcal{X}$ is compact since then

$$C_0(\mathcal{X}) = C_c(\mathcal{X}) = C(\mathcal{X}),$$

so that the function $\phi \equiv 1$ is admissible. Hence, if $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{X})$ converges to μ in the weak-$\star$ topology, it holds that

$$\mu(\mathcal{X}) = \int_{\mathcal{X}} 1\mathrm{d}\mu = \lim_{n\to\infty} \mu_n(\mathcal{X}) = 1.$$

For a general, non-compact space, the solution is to obtain a duality theorem with the space of bounded continuous functions $C_b(\mathcal{X})$, which contains the constant functions, so that the corresponding notion of weak convergence preserves the total mass, and is sufficiently large so that we can consider general Polish spaces.

Theorem 1.3. *Let $\mathcal{X}$ be a normal Hausdorff space. Then every bounded linear functional $L : C_b(\mathcal{X}) \to \mathbb{R}$ is uniquely represented via the duality product (1.6) by a regular bounded and finitely additive signed measure μ. In addition, the operator norm of L is given by $\|L\|_{op} = \|\mu\|_{\mathcal{M}_b(\mathcal{X})}$.*

Notice that in 1.3, the representation is not in necessarily a Radon measure. We refer the reader to [Fonseca and Leoni, 2007, Chap. 1.3.3] and the references therein for various duality theorems between spaces of measures and different classes of continuous functions. In what follows, we will use the weak topologies induced by the duality with $C_0(\mathcal{X})$ and $C_b(\mathcal{X})$, which of course are the same if $\mathcal{X}$ is compact.

Definition 1.4. *Given a suitable space of functions $\mathcal{K}$, we say a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ if*

$$\int_{\mathcal{X}} \phi \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int_{\mathcal{X}} \phi \mathrm{d}\mu_n \text{ for every } \phi \in \mathcal{K}.$$

If $\mathcal{K} = C_0(\mathcal{X})$, the continuous functions converging to 0 at infinity, we obtain the convergence in the weak-$\star$ sense, and we write $\mu_n \xrightarrow[n\to\infty]{\star} \mu$.

If $\mathcal{K} = C_b(\mathcal{X})$, the continuous and bounded functions, we say μ_n converges to μ in the narrow topology, and we write $\mu_n \xrightarrow[n\to\infty]{} \mu$.

From the previous discussion, since we are interested in working with the space $\mathscr{P}(\mathcal{X})$, the narrow topology is more suitable for most applications, even tough we cannot simply use Banach-Alaoglu-Bourbaki as a compactness criterion. Fortunately, we have Prokhorov's compactness theorem as a practical tool for compactness in the narrow topology.

Theorem 1.5 (Prokhorov)**.** *Let $\mathcal{F} \subset \mathscr{P}(\mathcal{X})$ be a family of probability measures over $\mathcal{X}$. Then $\mathcal{F}$ is compact for the narrow topology, if and only if, it is a tight family, i.e. for all $\varepsilon > 0$ there is a compact set K such that*

$$\mu(\mathcal{X} \setminus K) < \varepsilon, \text{ for all } \mu \in \mathcal{F}.$$

Actually, the set $\mathcal{K} = C_b(\mathcal{X})$ is not the minimal set for which we can define a weak topology that yields the narrow convergence. This is clear since we can always approximate functions in $C_b(\mathcal{X})$ with Lipschitz functions, but we can even construct a countable set of test functions yielding the narrow convergence.

Proposition 1.6 ([Ambrosio et al., 2008, Chapter 5])**.** *There exists a countable set $\mathcal{K} = (f_k)_{k\in\mathbb{N}}$ of Lipschitz functions, such that any sequence $(\mu_n)_{n\in\mathbb{N}}$ converges narrowly to μ, if and only if,*

$$\int_{\mathcal{X}} f_k \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int_{\mathcal{X}} f_k \mathrm{d}\mu, \text{ for all } k \in \mathbb{N}.$$

Now consider a pair of Polish spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ and let $\mathcal{X} \ni x \mapsto \nu^x \in \mathscr{P}(\mathcal{Y})$ be a measure-valued map.

Definition 1.7. *We say $(\nu^x)_{x\in\mathcal{X}}$ is measurable if for any Borel set $B \subset \mathcal{Y}$, the function $x \mapsto \nu^x(B)$ is Borel measurable.*

Now given $\mu \in \mathscr{P}(\mathcal{Y})$ we can define a new probability measure $\gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ in the product space through duality as

$$\int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\mathrm{d}\gamma(x,y) \overset{\text{def.}}{=} \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} f(x,y)\mathrm{d}\nu^x(y) \right) \mu(x),$$

and we use the notation $\gamma = \mu \otimes \nu^x$. It turns out that all measures $\gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ can be written in this way as a consequence of the *disintegration theorem*, see for instance [Stroock and Varadhan, 1997, Thm. 1.1.6] for a proof in Polish spaces.

Theorem 1.8. *Let $\mathcal{X}_0$ and $\mathcal{X}_1$ be Polish spaces, and two probability measures $\mu_0 \in \mathscr{P}(\mathcal{X}_0)$ and $\mu_1 \in \mathscr{P}(\mathcal{X}_1)$. If $\pi : \mathcal{X}_0 \to \mathcal{X}_1$ is a measurable map such that $\pi_\sharp\mu_0 = \mu_1$, then there exists a μ_1-a.e. uniquely determined Borel family $(\mu_0^{x_1})_{x_1\in\mathcal{X}_1} \subset \mathscr{P}(\mathcal{X}_0)$ such that*

$$\mu_0^{x_1}(\mathcal{X}_0 \setminus \pi^{-1}(x_1)) = 0 \text{ for } \mu_1\text{-a.e. } x_1 \in \mathcal{X}_1,$$

and for every measurable function $f : \mathcal{X}_0 \to [0, +\infty]$ if holds that

$$\int_{\mathcal{X}_0} f(x_0)\mathrm{d}\mu_0(x_0) = \int_{\mathcal{X}_1} \left(\int_{\pi^{-1}(x_1)} f(x_0)\mathrm{d}\mu_0^{x_1}(x_0) \right) \mathrm{d}\mu_1(x_1).$$

Any such $(\mu_0^{x_1})_{x_1\in\mathcal{X}_1}$ is called a disintegration family and we write $\mu_0 = \mu_0^{x_1} \otimes \mu_1$.

Whenever $\gamma \in \Pi(\mu, \nu)$, we apply the previous theorem with $\mathcal{X}_0 = \mathcal{X} \times \mathcal{Y}$, $\mathcal{X}_1 = \mathcal{X}$ and $\pi = \pi_{\mathcal{X}}$ to write $\gamma = \mu \otimes \nu^x$. One of the most useful, yet simple, applications of the disintegration theorem is the gluing lemma.

Lemma 1.9. *[Ambrosio et al., 2008, Lemma 5.3.2] Let $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ be Polish spaces, $\gamma_{1,2} \in \mathscr{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ and $\gamma_{1,3} \in \mathscr{P}(\mathcal{X}_1 \times \mathcal{X}_3)$ such that*

$$(\pi_{\mathcal{X}_1})_\sharp \gamma_{1,2} = (\pi_{\mathcal{X}_1})_\sharp \gamma_{1,3} = \mu_1.$$

Then there exists $\gamma_{1,2,3} \in \mathscr{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ such that

$$(\pi_{\mathcal{X}_1,\mathcal{X}_2})_\sharp \gamma_{1,2,3} = \gamma_{1,2} \text{ and } (\pi_{\mathcal{X}_1,\mathcal{X}_3})_\sharp \gamma_{1,2,3} = \gamma_{1,3}.$$

Proof. The proof consists on taking the disintegration families $\gamma_{1,2} = \gamma_{1,2}^{x_1} \otimes \mu_1(x_1)$, $\gamma_{1,3} = \gamma_{1,3}^{x_1} \otimes \mu_1(x_1)$ and defining the new measure as

$$\gamma_{1,2,3} \overset{\text{def.}}{=} \int_{\mathcal{X}_1} \gamma_{1,2}^{x_1} \otimes \gamma_{1,3}^{x_1} \mathrm{d}\mu_1(x_1).$$

□

2.2. Random Radon measures

We will also use in this work, the notion of random probability measure. The simplest example of this kind of object is a sequence of empirical measures, that is, given an *i.i.d.* sample of random variables $(X_i)_{i\in\mathbb{N}}$ we define the measures

$$\boldsymbol{\mu}_N \overset{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}.$$

Clearly, for each realization of the random variables we obtain a different discrete measure. For a random sample of agents $(X_i)_{i\in\mathbb{N}}$, we will describe a profile of strategies with the measures

$$\boldsymbol{\gamma}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i} \otimes \nu_i,$$

where $\nu_i \in \mathscr{P}(\mathcal{Y})$ represents the strategy, possibly in mixed plays, of player i. In general, a random measure is defined as follows.

Definition 1.10. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Polish space $\mathcal{X}$, a random measure $\boldsymbol{\mu}$ is a map from Ω into the space of Radon measures*

$$\boldsymbol{\mu} : \Omega \ni \omega \mapsto \boldsymbol{\mu}(\omega) \in \mathcal{M}_b(\mathcal{X}),$$

which is measurable for the Borel σ-algebra defined with respect to the narrow topology, in duality with $C_b(\mathcal{X})$. We let $\mathcal{M}_\Omega(\mathcal{X})$ denote the space of all random measures, and $\mathscr{P}_\Omega(\mathcal{X})$ is the convex subset of $\mathcal{M}_\Omega(\mathcal{X})$ consisting of all $\mathscr{P}(\mathcal{X})$-valued random probability measures.

Given $\mu_b \in \mathscr{P}_\Omega(\mathcal{X})$, the map

$$C_b(\mathcal{X}) \ni \phi \mapsto \mathbb{E}\left[\int_{\mathcal{X}} \phi \mathrm{d}\boldsymbol{\mu}(\omega)\right],$$

is a bounded linear map over $C_b(\mathcal{X})$, so from Riesz' representation theorem this defines a non-random measure via duality, the *expectation measure* $\mathbb{E}\boldsymbol{\mu} \in \mathscr{P}(\mathcal{X})$ as

$$\int_{\mathcal{X}} \phi \mathrm{d}\mathbb{E}\boldsymbol{\mu} \overset{\text{def.}}{=} \mathbb{E}\left[\int_{\mathcal{X}} \phi \mathrm{d}\boldsymbol{\mu}(\omega)\right]. \tag{1.7}$$

In particular, a random measure can be identified by a non-random measure if, and only if, it coincides with its expectation almost surely.

The Glivenko-Cantelli law of large numbers, also known as the Glivenko-Cantelli theorem [Dudley, 1969], states that the empirical measures $\boldsymbol{\mu}_N$ converge in the narrow topology to μ with probability 1. Hence, in order to give a topology to $\mathscr{P}_\Omega(\mathcal{X})$, the first naive candidate would be to consider $\mathbb{P}$-a.s. convergence of the random measures in the narrow topology. However, this topology would not be metrizable, and it also does not enjoy good compactness properties as Prokhorov's Theorem [Dudley, 2002]. For these reasons, we consider the narrow topology in $\mathscr{P}_\Omega(\mathcal{X})$.

Definition 1.11. *We say that an $f : \Omega \times \mathcal{X} \to \mathbb{R}$ is a random bounded continuous function, and we let $C_\Omega(\mathcal{X})$ denote the class of all such functions, if*[1]

1. $x \mapsto f(\omega, x) \in C_b(\mathcal{X})$ *almost surely;*
2. $\omega \mapsto f(\omega, x)$ *is $\mathcal{F}$-measurable for all $x \in \mathcal{X}$;*
3. $\omega \mapsto \|f(\omega, \cdot)\|_{L^\infty(\mathcal{X})}$ *is integrable with respect to $\mathbb{P}$.*

The narrow topology of random measures is then the weakest topology that makes

$$\mathscr{P}_\Omega(\mathcal{X}) \ni \boldsymbol{\mu} \mapsto \mathbb{E}_{\mathbb{P}}\left[\int_X f(\omega, x) d\boldsymbol{\mu}_\omega(x)\right] \text{ continuous for all } f \in C_\Omega(\mathcal{X}).$$

Since the functions of the form

$$\sum_{i=1}^N 1_{A_i}(\omega) f_i(x), \text{ for } A_i \quad \mathbb{P}\text{-measurable and } f_i \in C_b(\mathcal{X}),$$

are dense in $C_\Omega(\mathcal{X})$, it follows that $\mathbb{P}$-a.s. convergence implies convergence in the narrow topology of $\mathscr{P}_\Omega(\mathcal{X})$. The advantage is that the latter enjoys compactness properties analogous to Prokhorov's Theorem 1.5. This is extremely useful since, even if one can show that a sequence of random probability measures is tight almost surely and apply the classical version of Prokhorov's Theorem, for each event ω will be associated a subsequence where narrow convergence holds, but we cannot in general obtain a single subsequence that converges $\mathbb{P}$-almost surely. Hence the usefulness of the following result, see [Crauel, 2002, Thm. 4.29].

[1] In [Aubin and Frankowska, 2009] the random continuous functions are also called Carathéodory integrands.

Theorem 1.12 (Random Prokhorov's Theorem)**.** *A family of random measures $\mathcal{F} \subset \mathscr{P}_\Omega(\mathcal{X})$ is pre-compact for the narrow topology of random measures, if and only if, it is tight: for any $\varepsilon > 0$ there is a compact set K_ε such that*

$$\mathbb{E}\left[\boldsymbol{\mu}(\mathcal{X} \setminus K_\varepsilon)\right] \leq \varepsilon \textit{ for every } \boldsymbol{\mu} \in \mathcal{F}.$$

3. Geometric Measure Theory

3.1. Hausdorff measures and the Area and Co-area formulas

The Hausdorff measures over $\mathbb{R}^d$ where introduced in [Hausdorff, 1918]; the major advantage of introducing them is to quantify the dimension of parts of $\mathbb{R}^d$ in an intrinsic manner, without resorting to any sub-vector space, as well as to compute the k-dimensional volume of arbitrary sets. First consider the *k-dimensional Hausdorff measure of step δ* defined as

$$\mathscr{H}^k_\delta(E) \stackrel{\text{def.}}{=} \inf\left\{\sum_{n\in\mathbb{N}} \omega_k \left(\frac{\operatorname{diam} E_n}{2}\right)^k : \begin{array}{c}(E_n)_{n\in\mathbb{N}} \text{ covers } E\\ \operatorname{diam}(E_n) \leq \delta \text{ for all } n \in \mathbb{N}\end{array}\right\}, \tag{1.8}$$

where ω_k denotes the volume of the unit k-dimensional euclidean ball. For each $\delta > 0$ it is clear that $\mathscr{H}^k_\delta$ defines an outer-measure, *i.e.* a countably sub-additive set function, and given $E \subset \mathbb{R}^d$ the sequence $\delta \mapsto \mathscr{H}^k_\delta(E)$ is non-increasing so that the *k-dimensional Hausdorff measure* is defined as

$$\mathscr{H}^k(E) \stackrel{\text{def.}}{=} \sup_{\delta>0} \mathscr{H}^k_\delta(E) = \lim_{\delta\to 0} \mathscr{H}^k_\delta(E). \tag{1.9}$$

For the case $k = 0$, $\mathscr{H}^0$ denotes the counting measure.

Whenever $E \subset \mathbb{R}^d$ is a sufficiently nice set and $f : \mathbb{R}^d \to \mathbb{R}^d$ is a C^1 function, we know that

$$\operatorname{vol}(f(E)) = \int_E |\det \nabla f(x)| \mathrm{d}x.$$

The first advantage of introducing the Hausdorff measures is that it allows us to measure less regular sets, for instance Lipschitz images, instead of C^1 graphs. Given $f : \mathbb{R}^k \to \mathbb{R}^d$ and $E \subset \mathbb{R}^k$, it holds that

$$\mathscr{H}^k(f(E)) = \int_E J_k f(x) \mathrm{d}x,$$

where $J_k f$ is a generalization of the jacobian. This is a particular case of the area formula that is stated below.

Theorem 1.13 (Area Formula). *Let $f : \mathbb{R}^k \to \mathbb{R}^d$, for $1 \le k \le d$, be a Lipschitz function. Then for any measurable subset E of $\mathbb{R}^k$ it follows that*

$$\int_E J_k f(x) \mathrm{d}\mathcal{L}^k(x) = \int_{\mathbb{R}^d} \mathscr{H}^0 \left(E \cap \{f = y\}\right) \mathrm{d}\mathscr{H}^k(y),$$

where $J_k f \overset{\text{def.}}{=} \left(|\det(\nabla f^\top \nabla f)|\right)^{1/2}$ is the k-dimensional jacobian of f.

In the sequel we consider a few particular cases and direct consequences:

- if in addition, we assume that f is injective, the term $\mathscr{H}^0 \left(E \cap \{f = y\}\right)$, called the multiplicity, equals 1 if $y \in f(E)$ and 0 otherwise, so that we obtain

  $$\mathscr{H}^k(f(E)) = \int_E J_k f(x) \mathrm{d}x,$$

  which is a generalization of the change of variables formula, the case when $k = d$ so that $J_k f = |\det \nabla f|$.

- if $g : \mathbb{R}^k \to [0, +\infty]$ is a Borel function we have that

  $$\int_{f(E)} \left(\int_{\mathbb{R}^k \cap \{f = y\}} g \mathrm{d}\mathscr{H}^0 \right) \mathrm{d}\mathscr{H}^k(y) = \int_E g(x) J_k f(x) \mathrm{d}\mathcal{L}^k(x).$$

  Which can also be written with a disintegration notation from 1.8 as

  $$J_k f \mathcal{L}^k \llcorner E = \mathscr{H}^0 \llcorner \{f = y\} \otimes \mathscr{H}^k_{\mathbb{R}^d} \llcorner f(E),$$

  where the subindex $\mathbb{R}^d$ is to emphasize that we are referring to k-dimensional Hausdorff measure over $\mathbb{R}^d$.

- Whenever f is injective, taking $g = \phi \circ f$ in the above case, for an arbitrary $\phi \in C_b(\mathbb{R}^d)$, implies that

  $$f_\sharp \left(J_k f \mathcal{L}^k \llcorner E \right) = \mathscr{H}^k_{\mathbb{R}^d} \llcorner f(E).$$

The area formula can then be seen as a way to embed arbitrary lower dimensional structures into a bigger dimensional space. The dual approach operation is furnished by the co-area formula that consists in foliating a set into smaller dimensional slices.

Theorem 1.14 (Co-Area Formula). *Let $f : \mathbb{R}^d \to \mathbb{R}^k$, for $1 \le k \le d$, be a Lipschitz function. Then for any measurable subset E of $\mathbb{R}^d$ it follows that*

$$\int_E C_k f(x) \mathrm{d}\mathcal{L}^d(x) = \int_{\mathbb{R}^k} \mathscr{H}^{d-k} \left(E \cap \{f = y\}\right) \mathrm{d}\mathcal{L}^k(y),$$

where $C_k f \overset{\text{def.}}{=} \left(|\det(\nabla f \nabla f^\top)|\right)^{1/2}$ is the k-dimensional co-area factor of f.

As for the area formula, we present some interesting cases of use.

- Let $\psi : \mathbb{R}^d \to \mathbb{R}$ be a summable Borel function, then

$$\int_E \psi(x) C_k f(x) \mathrm{d}\mathcal{L}^d(x) = \int_{f(E)} \left(\int_{E\cap\{f=y\}} \psi \mathrm{d}\mathscr{H}^{d-k} \right) \mathrm{d}\mathcal{L}^k(y),$$

  which can also be written in disintegration form as

$$C_k f \mathcal{L}^d \llcorner E = \mathscr{H}^{d-k} \llcorner \{f = y\} \otimes \mathcal{L}^k \llcorner f(E).$$

- The particular case $k = 1$, we have $C_k f = |\nabla f|$, which gives the integration over level sets formula

$$\int_{E\cap\{f>t\}} \psi(x) |\nabla f(x)| \mathrm{d}x = \int_t^\infty \left(\int_{E\cap\{f=s\}} \psi \mathrm{d}\mathscr{H}^{d-1} \right) \mathrm{d}s,$$

  so that

$$\Psi : t \mapsto \int_{E\cap\{f>t\}} \psi(x) |\nabla f(x)| \mathrm{d}x \in AC(\mathbb{R}_+), \text{ and } \Psi'(t) = -\int_{E\cap\{f=t\}} \psi \mathrm{d}\mathscr{H}^{d-1}.$$

3.2. Rectifiable sets

We have introduced the Hausdorff measures $\mathscr{H}^k$ with the promise of being able to compute the area of arbitrary hyper-surfaces of dimension k without resorting to a parametrization. A natural example are smooth manifolds of $\mathbb{R}^d$ of dimension k. We shall see that the Hausdorff measures will allow us to study a much wider class of sets. More specifically, we introduce the notion of *rectifiable set*, see [Ambrosio et al., 2000, Definition 2.57] or [Maggi, 2012, Chapter 10], and more generally the notion of *rectifiable measure*.

Definition 1.15. *Let $M \subset \mathbb{R}^d$ be a Borel set and $k \in \mathbb{N}$, we say that M is countably $\mathscr{H}^k$-rectifiable, or shortly k rectifiable, if there are countably many Lipschitz functions $f_i : \mathbb{R}^k \to \mathbb{R}^d$ such that*

$$\mathscr{H}^k \left(M \setminus \bigcup_{i\in\mathbb{N}} f_i\left(\mathbb{R}^k\right) \right) = 0.$$

A Radon measure μ is said to be k-rectifiable if it is supported over a k-rectifiable set and $\mu \ll \mathscr{H}^k$.

In the simple case $M = f(E)$, for $E \subset \mathbb{R}^k$, one can define the tangent space at a point of differentiability of f as

$$\nabla f(z) \left(\mathbb{R}^k\right), \text{ for } x = f(z).$$

This is a parametric definition that can be extended to k-rectifiable sets. It turns out the parametric notion of tangentiability can be expressed in terms of measure theory. Given a Borel set M, we set the measure $\mu = \mathscr{H}^k \llcorner M$, and we consider the family of blow-up measures

$$\mu_r \overset{\text{def.}}{=} r^{-k}\Phi^{x,r}_{\sharp}\mu = \mathscr{H}^k \llcorner \left(\frac{M-x}{r}\right), \text{ for } \Phi^{x,r} \overset{\text{def.}}{=} \frac{\mathrm{id}-x}{r}. \tag{1.10}$$

The blow-up Theorem, see [Maggi, 2012, Theorem 10.2], states that for $\mathscr{H}^k$-a.e. $x \in M$ this family of measures converges in the weak-$\star$ topology to a measure of the form $\mathscr{H}^k \llcorner \pi_x$, for a unique k-plane $\pi_x \in \mathrm{G}(k,d)$, the Grassmannian of k-planes of $\mathbb{R}^d$.

More generally define the k-density, whenever the limit exists, of a Radon measure μ as

$$\theta_k(\mu,x) \overset{\text{def.}}{=} \lim_{r\to 0^+} \frac{\mu(B_r(x))}{\omega_k r^k} \text{ and } \theta_k(M,x) \overset{\text{def.}}{=} \theta_k\left(\mathscr{H}^k \llcorner M, x\right), \tag{1.11}$$

see [Ambrosio et al., 2000, Maggi, 2012]. A direct consequence of the blow-up Theorem is that $\mathscr{H}^k$-a.e. point of a k-rectifiable set has k-density 1. Analogously for a k-rectifiable measure μ it holds that $\mu = \theta_k(\mu,x)\mathscr{H}^k \llcorner M$.

The equivalence between all notions was completed with the work of Preiss and the notion of a tangent space to a measure, see for instance the monograph [De Lellis, 2006] and [Mattila, 1995]. If a measure (*resp.* a set) has a finite k-density, *i.e.* the limit in (1.11) exists and is finite $\mathscr{H}^k$-a.e., then this measure (*resp.* set) is k-rectifiable. The previous discussion is summarized in the following theorem.

Theorem 1.16. *Let μ be a Radon measure over $\mathbb{R}^d$, the following are equivalent.*

(i) *μ is k-rectifiable*

(ii) *For $\mathscr{H}^k$-a.e. $x \in \operatorname{supp}\mu$, it holds that*

$$r^{-k}\Phi^{x,r}_{\sharp}\mu \xrightarrow[r\to 0]{\star} \theta_k(\mu,x)\mathscr{H}^k \llcorner \pi_x,$$

for a unique k-plane $\pi_x \in G(k,d)$. In this case, we say that Σ is flat at x.

(iii) *For $\mathscr{H}^k$-a.e. $x \in \operatorname{supp}\mu$, the k-density of μ in* (1.11) *exists and is finite. In the particular case $\mu = \mathscr{H}^k \llcorner \Sigma$, it follows that Σ is k-rectifiable if and only if $\theta_k(\Sigma,x) = 1$ for $\mathscr{H}^k$-a.e. $x \in \Sigma$.*

4. The Optimal Transportation problem

The optimal transportation problem was originally proposed by Monge in [Monge, 1781] and by now it has been extensively documented in a vast selection of recent monographs, see [Ambrosio et al., 2008, Villani, 2009, Santambrogio, 2015, Ambrosio et al., 2021]. It can be stated in modern terminology as follows: Given two Polish spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$, one seeks to transport a distribution given by a probability measure $\mu \in \mathscr{P}(\mathcal{X})$ onto a target distribution $\nu \in \mathscr{P}(\mathcal{Y})$, while minimizing the total transportation work, where the cost of transporting a unity of mass from the point x to y is given by $c(x, y)$, for a lower semi-continuous function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$. This can be expressed as the following variational problem

$$\inf_{T_\sharp \mu = \nu} \int_{\mathcal{X}} c(x, T(x)) \mathrm{d}\mu, \tag{1.12}$$

where $T_\sharp \mu$ denotes the push-forward measure and is defined as

$$T_\sharp \mu(A) \overset{\text{def.}}{=} \mu(T^{-1}(A)), \text{ for any Borel set } A \subset \mathcal{Y}, \tag{1.13}$$

or equivalently via duality as

$$\int_{\mathcal{Y}} \phi \mathrm{d} T_\sharp \mu \overset{\text{def.}}{=} \int_{\mathcal{X}} \phi(T(x)) \mathrm{d}\mu(x), \text{ for any } \phi \in C_b(\mathcal{Y}). \tag{1.14}$$

This problem is highly non-linear, and the existence and the properties of an optimal map remained not understood for two centuries until Kantorovitch proposed a reformulation that reinvigorated the field [Kantorovich, 1942]. Kantorovitch's reformulation is then written as a linear program in the space of probability measures

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\gamma(x, y), \tag{1.15}$$

where

$$\Pi(\mu, \nu) \overset{\text{def.}}{=} \left\{ \gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{Y}) : \quad (\pi_{\mathcal{X}})_\sharp \gamma = \mu, \ (\pi_{\mathcal{Y}})_\sharp \gamma = \nu \right\} \tag{1.16}$$

is the space of transport couplings, $\pi_{\mathcal{X}}$ denotes the projection onto $\mathcal{X}$, *i.e.* $\pi_{\mathcal{X}}(x, y) = x$ and similarly for $\pi_{\mathcal{Y}}$. Therefore, (1.15) corresponds to the minimization of a linear functional under linear constraints, so that existence follows from the direct method. Indeed, if c is a continuous and bounded function, the map $\gamma \mapsto \langle c, \gamma \rangle$ is continuous from the definition of the narrow topology. If c is l.s.c., it can be approximated by bounded Lipschitz functions and one can show that is this case $\gamma \mapsto \langle c, \gamma \rangle$ is l.s.c. in the narrow topology, we refer the reader to [Ambrosio et al., 2021, Thm. 2.6 and Thm. 2.10] for a proof of these claims.

4.1. Kantorovitch's problem as a relaxation of Monge's

Problems (1.15) and (1.12) can be respectively stated as the minimization of the following functionals

$$\mathscr{K}(\gamma) \overset{\text{def.}}{=} \begin{cases} \displaystyle\int_{\mathcal{X} \times \mathcal{Y}} c \mathrm{d}\gamma, & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty, & \text{otherwise,} \end{cases} \quad \mathscr{M}(\gamma) \overset{\text{def.}}{=} \begin{cases} \mathscr{K}(\gamma), & \text{if } \gamma = (\mathrm{id}, T)_\sharp \mu \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.17}$$

The natural question is when does it hold that $\mathscr{K} = \overline{\mathscr{M}}$, that is, Kantorovitch's formulation is given by the lower semi-continuous envelope of Monge's in the sense of (1.1).

As we have discussed in the introduction, Kantorovitch identified the two things that could go wrong to prevent optimal maps to exist. First was the fact that Monge's problem was highly non-linear and hence the topology to apply the Direct method was not clear. This is not an issue for (1.15) as it is a linear program. The second issue was the possibility of not having any maps at all, for instance if μ is a single Dirac delta and ν is the convex combination of two. This is not an issue for (1.15) either since $\mu \otimes \nu$ is always an admissible coupling.

While fixing these two issues, Kantorovitch obtained formally a relaxation of Monge's problem, under the minimal hypothesis that μ is atomless. This assumption guarantees that there always exists a map transporting μ to ν, and in the case that $\mathcal{X} = \mathcal{Y} = \Omega$ is a compact subset of $\mathbb{R}^d$, in [Ambrosio, 2003] Ambrosio rigorously proved this relaxation by showing that the set of transportation couplings induced by maps is dense.

Proposition 1.17 (Ambrosio). *If $\mathcal{X} = \mathcal{Y} = \Omega$, a compact subset of $\mathbb{R}^d$, c is a continuous and bounded function and μ is atomless, then for any $\gamma \in \Pi(\mu, \nu)$ there is a sequence of maps such that*

$$(T_n)_\sharp \mu = \nu, \text{ and } \gamma_{T_n} \overset{\text{def.}}{=} (\mathrm{id}, T_n)_\sharp \mu \xrightarrow[n\to\infty]{} \gamma.$$

In other words, the set of couplings induced by transportation maps is dense.

In particular, this implies the equality between the infimum of Monge and the minimization of Kantorovitch

$$\inf \mathscr{M} = \min \mathscr{K}. \tag{1.18}$$

Ambrosio's density result 1.17 is much stronger than what is needed to show (1.18). In [Pratelli, 2007] Pratelli showed that this equality holds under much more general assumptions, which are sharp as he provides many counterexamples that contradict (1.18) whenever they do not hold.

Proposition 1.18 (Pratelli). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and let c be an l.s.c. function that is continuous in the interior of its domain. Then if μ is atomless, the equality (1.18) holds.*

4.2. The dual problem and existence of an optimal map

To gain some insight on the problem, let us consider some optimal coupling γ and a finite set of pairs $(x_i, y_i)_{i=1}^N \subset \operatorname{supp} \gamma$. Then, an optimality condition for the plan γ is that for any choice of such pairs it holds that

$$\sum_{i=1}^N c(x_i, y_i) \le \sum_{i=1}^N c(x_i, y_{i+1}),$$

with the convention that $y_{N+1} = y_1$. If it was not the case, we can use the lower semicontinuity of c to build a strictly better competitor. Any set $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ with this property is called $c-$*cyclically-monotone*, they encode the geometric information that no permutation can yield a better transportation cost.

This phenomenon relies on the interpretation of the quantity $c(x, y)$ as the cost of transporting one unit of goods from x to y, under the global transportation of resources distributed by μ onto the target distribution ν. Instead, we can imagine that this global distribution will be done by an external consortium. This consortium will then charge the price $\varphi(x)$ to pick up one unit of this resource at the position x, and charge $\psi(y)$ to deliver it at y. But then, this consortium will face the constraint in this pricing given by

$$\varphi(x) + \psi(y) \leq c(x, y),$$

otherwise another consortium could still profit with strictly better prices.

It then follows that any admissible ψ will be such that

$$\psi(y) \leq \inf_{x \in \mathcal{X}} c(x, y) - \varphi(x) \stackrel{\text{def.}}{=} \varphi^c(y), \text{ for all } y \in \mathcal{Y}.$$

So that the best possible ψ, for φ fixed, is given by φ^c, the so-called c-transform of φ. So for any φ it follows from the definition that $\varphi(x) + \varphi^c(y) \leq c(x, y)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, and we define the *contact set*, or the c *sub-differential* of φ be defined as

$$\partial^c \varphi \stackrel{\text{def.}}{=} \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \ \varphi(x) + \varphi^c(y) = c(x, y)\}.$$

Since we can decouple the contributions of x and y with the prices φ, ψ, a direct inspection shows that $\partial^c \varphi$ is a c-cyclically monotone subset of $\mathcal{X} \times \mathcal{Y}$.

One can then expect that the transportation consortium will have its maximum profit when it chooses prices $(\varphi, \psi) \in L^1(\mu) \times L^1(\nu)$ attaining

$$\max_{\substack{(\varphi,\psi) \in L^1(\mu) \times L^1(\nu) \\ \varphi(x)+\psi(y) \leq c(x,y)}} \int_{\mathcal{X}} \varphi \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu = \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c \mathrm{d}\gamma. \tag{1.19}$$

The maximization on the left-hand side of (1.19) is called *Kantorovitch's dual problem* and its maximizers are called *Kantorovitch potentials*. We summarize the previous discussion as follows, see [Villani, 2009, Thm. 5.10].

Theorem 1.19. *Let* $(\mathcal{X}, d_{\mathcal{X}})$ *and* $(\mathcal{Y}, d_{\mathcal{Y}})$ *be Polish spaces and* $c : \mathcal{X} \times \mathcal{Y} \to [0, +\infty]$ *be an l.s.c. function, then the duality holds*

$$\begin{aligned} \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c \mathrm{d}\gamma &= \sup_{\substack{(\varphi,\psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y}) \\ \varphi(x)+\psi(y) \leq c(x,y)}} \int_{\mathcal{X}} \varphi \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu \\ &= \max_{\substack{(\varphi,\psi) \in L^1(\mu) \times L^1(\nu) \\ \varphi(x)+\psi(y) \leq c(x,y)}} \int_{\mathcal{X}} \varphi \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu \end{aligned} \tag{1.20}$$

In addition, if the minimization above is finite, the following are equivalent

- $\gamma \in \Pi(\mu, \nu)$ *is optimal;*

- *there is* φ *such that* $\operatorname{supp}\gamma \subset \partial^c\varphi$ *and hence it is* c*-cyclically monotone;*

- $\operatorname{supp}\gamma$ *is concentrated in a* c*-cyclically monotone set* $\Gamma \subset \mathcal{X} \times \mathcal{Y}$*.*

With Kantorovitch's duality theory it was finally possible to answer to the question of existence of optimal maps for the problem of Monge. Consider $(x, y) \in \operatorname{supp}\gamma \subset \partial^c\varphi$, for some optimal Kantorovitch potential as in 1.19, it follows that

$$\varphi(x) + \varphi^c(y) = c(x, y).$$

Hence, assuming that $\mathcal{X} = \mathcal{Y} = \Omega$ is a compact subset of $\mathbb{R}^d$, cost $c \in C^1$ and satisfies what is known as the *twist condition*

$$y \mapsto \nabla_x c(x, y) \text{ is injective,} \tag{1.21}$$

we can take gradients w.r.t. the variable x, assuming sufficient regularity of the potentials, and obtain that

$$\nabla\varphi(x) = \nabla_x c(x, y).$$

As a result, if μ is absolutely continuous with respect the Lebesgue measure, this equality combined with the twist condition defines a measurable map $T(x) \overset{\text{def.}}{=} (\nabla_x c(x, \cdot))^{-1}(\nabla\varphi(x))$, which is the unique optimal map for Monge's problem.

4.3. The Wasserstein distances

The Wasserstein distances W_p are defined through the value function of an optimal transport problem, see [Ambrosio et al., 2008, Santambrogio, 2015, Villani, 2009] for further details. Given two probability measures $\mu, \nu \in \mathscr{P}(\mathcal{X})$, we set

$$W_p^p(\mu, \nu) \overset{\text{def.}}{=} \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d_{\mathcal{X}}(x, y)^p \mathrm{d}\gamma = \sup_{\substack{(\varphi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y}) \\ \varphi(x)+\psi(y)\le d_{\mathcal{X}}(x,y)^p}} \int_{\mathcal{X}} \varphi \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu. \tag{1.22}$$

In the sequel we list some properties of the Wasserstein distances

- W_p is l.s.c. with respect to the narrow convergence, since from the Kantorovitch-duality formula, it can be written as the supremum of continuous functions. Moreover it is continuous if $\mathcal{X}$ is compact, [Villani, 2009, Lemma 4.3]

- $W_p^p(\mu, \nu) < +\infty$ if and only if $\mu, \nu \in \mathscr{P}_p(\mathcal{X})$ which is the space of probability measures with finite p-moments, *i.e.*

$$\mathscr{P}_p(\mathcal{X}) \overset{\text{def.}}{=} \left\{ \mu \in \mathscr{P}(\mathcal{X}) : \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)^p \mathrm{d}\mu(x) < +\infty \right\};$$

- $(\mathscr{P}_p(\mathcal{X}), W_p)$ is a Polish space, whenever $(\mathcal{X}, d_{\mathcal{X}})$ is;

- convergence for the distance W_p is equivalent to narrow convergence plus convergence of the p-moments

$$W_p(\mu_n, \mu) \xrightarrow[n\to\infty]{} 0 \Longleftrightarrow \begin{cases} \mu_n \xrightarrow[n\to\infty]{} \mu, \\ \displaystyle\int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)^p \mathrm{d}\mu_n(x) \xrightarrow[n\to\infty]{} \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)^p \mathrm{d}\mu(x). \end{cases}$$

Chapter 2

Topological and measure theoretic properties of metric continuums

Contents

In this Chapter we are interested in studying the properties of the class of subsets Σ of $\mathbb{R}^d$ such that Σ is connected and has finite length *i.e.* $\mathscr{H}^1(\Sigma) < +\infty$, which are called *metric continuums*. First we see that this class itself can be seen as a metric space, for which we shall introduce the notions of *Hausdorff* and *Kuratowski* convergence and study their properties. We shall see that these topologies enjoy all the properties we usually require to define variational problems in this class, and also to prove existence of solutions, for instance compactness (Blaschke's Theorem) and lower semi-continuity (Gołab's Theorem).

We can also study the fine properties of one fixed element of this class. Given one such Σ, we can use the tools from geometric measure theory developed in the previous chapter to endow it with an intrinsic metric structure, one that is independent of the underlying space $\mathbb{R}^d$, or even another ambient metric space since our arguments generalize to this setting. Then, we revisit the notions of approximate tangentiability studied previously for general countably $\mathscr{H}^k$-rectifiable sets. Not only the elements of this class are automatically 1-rectifiable sets, but we can show that their notion of blow-ups is stronger than the usual measure-theoretic sense, since the convergence of blow-ups also holds in the topology induced by the Hausdorff metric. Finally, we study some geometric qualitative properties that such sets might present, such as what it means to be a tree.

We should remark that most of the results of this chapter are true when the ambient space is a general metric space, instead of $\mathbb{R}^d$. We have decided to keep the discussion in the euclidean setting to simplify the presentation since the structure of blow-ups is by nature euclidean and most of the applications throughout this thesis are for variational problems among subsets of euclidean spaces.

1. Preliminary properties: Hausdorff and Kuratowski convergence

We start by clarifying what we mean by a connected set with finite length and its basic working properties. By connected, we use the standard definition, see [Munkres, 2017].

Definition 2.1. *Consider a non-empty set Ω, which can be a subset of any topological space. We say that two non-empty and disjoint open subsets A, B are a separation of Ω if $A \cup B = \Omega$. The set Ω is connected if it does not admit a separation. It is said to be path or arc-wise connected if for any two points $x, y \in \Omega$ there is a continuous function $f : [0, 1] \to \Omega$ such that $f(0) = x$ and $f(1) = y$.*

Some standard properties of connected sets can be worked out from the definition, for instance that the continuous images of connected sets are connected and that path connected implies connected. The converse is not true in general, the typical example being the *topologist's sine curve*

$$\begin{aligned} S &= \overline{\{(x, \sin(1/x)) \in \mathbb{R}^2 : \ x \in (0, 1)\}} \\ &= (\{0\} \times [-1, 1]) \cup \left\{(x, \sin(1/x)) \in \mathbb{R}^2 : \ x \in (0, 1]\right\}. \end{aligned}$$

It is tempting to say that the issue is that $\mathscr{H}^1(S) = +\infty$, and this is indeed the feature that prevents this set from being path connected, as we see in the following Proposition, whose proof can be found in [Alberti and Ottolini, 2017].

Proposition 2.2. *Let Σ be a closed connected subset of $\mathbb{R}^d$ with finite length $\mathscr{H}^1(\Sigma) < +\infty$. Then there is a surjective Lipschitz curve f from $[0,1]$ to Σ. In particular Σ is countably $\mathscr{H}^1$-rectifiable and path-connected.*

This proposition is a great example of the interplay between the topological and measure theoretical properties of the sets we are interested in. In what follows, whenever we are dealing with connected sets of finite length, we shall immediately assume that they are path connected and rectifiable. Besides, in general it is not easy to verify neither that a set is rectifiable, not that it is path connected, so the previous proposition gives us a very practical criterion to verify these working properties.

In order for the space of continua to be a viable class to define variational problems, we need to endow it with a topology that preserves the properties defining it, *i.e.* being connected and having finite length. For this we introduce the notions of *Hausdorff* and *Kuratowski* convergence of sets, see for instance [Rockafellar and Wets, 2009]. After discussing their basic properties and some reformulations, we show that they have the desired properties when restricted to the class connected sets with bounded length.

Definition 2.3. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of closed sets of $\mathbb{R}^d$. If $A \subset \mathbb{R}^d$ is closed, we say that*

- *A_n converges in the Hausdorff sense to A if $d_H(A_n, A) \xrightarrow[n\to\infty]{} 0$, where d_H is called the Hausdorff distance and is defined as*
$$d_H(A,B) \stackrel{\text{def.}}{=} \max\left\{\sup_{a\in A}\operatorname{dist}(a,B), \sup_{b\in B}\operatorname{dist}(b,A)\right\}, \text{ we write } A_n \xrightarrow[n\to\infty]{d_H} A. \quad (2.1)$$
- *A sequence of closed sets C_n converges in the sense of Kuratowski to C, and we write $C_n \xrightarrow[n\to\infty]{K} C$, when*
  1. *for all sequences $x_n \in C_n$, all its cluster points are contained in C.*
  2. *For all points $x \in C$ there exists a sequence $x_n \in C_n$, converging to x.*

It can be verified that the quantity $d_H(\cdot,\cdot)$ defines indeed a distance over the compact subsets of $\mathbb{R}^d$, or any metric space in general. Manipulating its definition we obtain the following equivalent characterizations.

Proposition 2.4. *For any two closed sets $X, Y \subset \mathbb{R}^d$, the following characterizations of the Hausdorff distance hold*

$$\begin{aligned} d_H(X,Y) &= \sup_{z\in\mathbb{R}^d} |\operatorname{dist}(z,X) - \operatorname{dist}(x,Y)| & (2.2)\\ &= \inf\left\{\varepsilon > 0 : \; X \subseteq B_\varepsilon(Y) \text{ and } Y \subseteq B_\varepsilon(X)\right\}. & (2.3)\end{aligned}$$

Proof. For the first equality, let us first show that

$$\sup_{z\in\mathbb{R}^d} |\mathrm{dist}(z,X) - \mathrm{dist}(z,Y)| \le d_H(X,Y).$$

For a given z, suppose without loss of generality that $\mathrm{dist}(z,X) \ge \mathrm{dist}(z,Y)$ and take $\bar{x} \in X$ attaining the $\mathrm{dist}(z,X)$ and take $\bar{y} \in X$ attaining $\mathrm{dist}(\bar{x},Y)$. Then it follows that

$$\begin{aligned} |\mathrm{dist}(z,X) - \mathrm{dist}(z,Y)| &= \mathrm{dist}(z,X) - \mathrm{dist}(z,Y) \\ &\le |z-\bar{x}| - |z-\bar{y}| \le |\bar{x}-\bar{y}| \\ &= \mathrm{dist}(\bar{x},Y) \le \sup_{x\in X} \mathrm{dist}(x,Y) \le d_H(X,Y). \end{aligned}$$

Taking the sup over z gives the desired inequality. For the converse one, assume now that $\sup\limits_{y\in Y} \mathrm{dist}(y,X) \ge \sup\limits_{x\in X} \mathrm{dist}(x,Y)$ and for each $n \in \mathbb{N}$ there is $y_n \in Y$ such that

$$\begin{aligned} d_H(X,Y) = \sup_{y\in Y} \mathrm{dist}(y,X) \le \mathrm{dist}(y_n,X) + \frac{1}{n} &= |\mathrm{dist}(y_n,X) - \mathrm{dist}(y_n,Y)| + \frac{1}{n} \\ &\le \sup_{z\in\mathbb{R}^d} |\mathrm{dist}(z,X) - \mathrm{dist}(z,Y)| + \frac{1}{n}. \end{aligned}$$

Letting $n \to \infty$, we obtain the first characterization.

For the second characterization, consider $\varepsilon > d_H(X,Y)$, then from the definition (2.1) for any $x \in X$, there is $y \in Y$ such that $d(x,y) < \varepsilon$. As a result we obtain that X is contained in the neighborhood $B(Y,\varepsilon)$. The corresponding inclusion for Y also holds by symmetry of the definition. This implies that

$$d_H(X,Y) \le \inf\left\{\varepsilon > 0 : \ X \subseteq B_\varepsilon(Y) \text{ and } Y \subseteq B_\varepsilon(X)\right\}.$$

Now consider some $\varepsilon < d_H(X,Y)$, then there is either $x \in X$ such that $\mathrm{dist}(x,Y) > \varepsilon$, in which case $X \subsetneq B(Y,\varepsilon)$, or there is $y \in Y$ such that $\mathrm{dist}(y,X) > \varepsilon$, so that $Y \subsetneq B(X,\varepsilon)$. Either way, we get that the infimum we are interested in is at least $d_H(X,Y)$ and the result follows. □

A direct consequence of the equivalent characterization (2.3) is that Hausdorff convergence preserves connectedness.

Proposition 2.5. *Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a sequence of closed and connected sets converging to Σ for the Hausdorff distance. Then Σ is connected.*

Proof. Suppose that Σ is not connected then there are two disjoint and closed sets C_1 and C_2 such that $\Sigma \subset C_1 \cup C_2$. As closed and disjoint sets they must be at a positive distance from each other,

$$0 < \delta \overset{\text{def.}}{=} \mathrm{dist}(C_1, C_2),$$

Then, taking n large enough so that $d_H(\Sigma,\Sigma_n) < \delta/4$, we obtain that $\Sigma_n \subset B(\Sigma,\delta/4) \subset B(C_1 \cup C_2, \delta/4)$. But this is a contradiction with the connectedness of Σ_n since $B(C_1,\delta/4)$ and $B(C_2,\delta/4)$ are open sets. □

Another direct consequence, this time from characterization (2.2), is that $A_n \xrightarrow[n\to\infty]{d_H} A$, if and only if $\mathrm{dist}(\cdot, A_n) \xrightarrow[n\to\infty]{} \mathrm{dist}(\cdot, A)$ uniformly. The Kuratowski convergence, exhibits a similar behavior, which can be more easily seen once we define the notions of Kuratowski inner and outer limits, see [Rockafellar and Wets, 2009, Definition 4.1], as follows

$$\liminf_{n\to\infty} C_n \overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^d : \limsup_{n\to\infty} \mathrm{dist}(x, C_n) = 0 \right\},$$
$$\limsup_{n\to\infty} C_n \overset{\text{def.}}{=} \left\{ x \in \mathbb{R}^d : \liminf_{n\to\infty} \mathrm{dist}(x, C_n) = 0 \right\},$$

and are always well-defined. The outer and inner limits correspond, respectively, to the points satisfying properties (1) and (2) in the definition of Kuratowski convergence. Therefore, the convergence is equivalent to the liminf coinciding with the limsup, which translates into pointwise convergence of $\mathrm{dist}(\cdot, C_n)$ towards $\mathrm{dist}(;C)$. Recalling that they are all 1-Lipschitz, by Ascoli-Arzela's Theorem,

$$C_n \xrightarrow[n\to\infty]{K} C \text{ if and only if } \mathrm{dist}(\cdot, C_n) \xrightarrow[n\to\infty]{} \mathrm{dist}(\cdot, C) \text{ locally uniformly.}$$

As a consequence, Hausdorff implies Kuratowski convergence, and whenever we consider subsets of a compact set, the two notions coincide. This characterization via the uniform convergence of the distance functions has an important consequence, which is that the topology induced by the Hausdorff distance is compact.[1] This result is known as Blaschke's Theorem, see [Ambrosio et al., 2000, Chap. 6].

Theorem 2.6 (Blaschke)**.** *Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of closed subsets of a compact set Ω, then it admits a subsequence converging in the Hausdorff distance.*

Proof. Consider the family of 1-Lipschitz functions $d_n(\cdot) \overset{\text{def.}}{=} \mathrm{dist}(\cdot, C_n)$. Since the sequence C_n is contained in a compact set, the corresponding sequence of distance functions is uniformly bounded and uniformly continuous, satisfying the hypothesis of the Ascoli-Arzela Theorem. So up to a subsequence we assume that $d_n \xrightarrow[n\to\infty]{\|\cdot\|_\infty} d$, converges uniformly to some 1-Lipschitz function d.

Define the set

$$C \overset{\text{def.}}{=} \{d = 0\},$$

and our goal is to prove that C_n converges to C in the Kuratowski sense, since it is equivalent to Hausdorff convergence as we are in a compact set.

The first point in the definition of Kuratowski convergence is trivially satisfied since if $C_n \ni x_n \to x$, then $d(x) = \lim d_n(x_n) = 0$ and hence $x \in C$. For the second, for each n let $x_n \in C_n$ be a point attaining $\mathrm{dist}(x, C_n)$. Then

$$d(x, x_n) = \mathrm{dist}(x, C_n) \xrightarrow[n\to\infty]{} d(x) = 0.$$

[1] In what follows, we shall refer to this topology as the Hausdorff topology, which is not to be mistaken with general Hausdorff topologies, that is topologies such that every two points can be separated by disjoint open sets.

The result follows. □

A less clear behavior however is what happens upon restriction to a closed set. Given a compact set L, it is not true in general that $C_n \xrightarrow[n\to\infty]{K} C$ implies the convergence of $C_n \cap L$ to $C \cap L$. Indeed, it is hard to verify property (2) since given $x \in \partial(L \cap C)$ and a sequence $x_n \to x$, this sequence might be coming from the complement of L. This indicates that the Kuratowski limit loses information at the boundary, unless this information is coming from the interior of the set. In general, we have the following Lemma.

Lemma 2.7. *Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of closed sets in $\mathbb{R}^d$, converging to C in the sense of Kuratowski. Then, for any $x_0 \in \mathbb{R}^d$, there exists a countable set $I \subset [0, +\infty)$ such that*

$$C_n \cap \overline{B_R(x_0)} \xrightarrow[n\to\infty]{d_H} C \cap \overline{B_R(x_0)}, \text{ for all } R \in [0, +\infty) \setminus I.$$

In the next proof, use the following notation $B_R = \{|x| < R\}$ and $\overline{B_R} = \{|x| \le R\}$.

Proof. Up to a translation, it suffices to prove the result for $x_0 = 0$. We can also assume that $C \neq \emptyset$, otherwise for any $R > 0$, $C_n \cap B_R = \emptyset$ for n large enough and the result holds. Defining $R_0 = \inf\{R > 0 : C \cap \overline{B_R} \neq \emptyset\}$, we have that if $R < R_0$, one has $C_n \cap \overline{B_R} = \emptyset$ for n large enough and the Hausdorff limit is empty, as expected.

Now we take $R \ge R_0$ and consider a subsequence $(C_{n_k})_{k\in\mathbb{N}}$ and a closed set C^R such that

$$C_{n_k} \cap \overline{B_R} \xrightarrow[n\to\infty]{d_H} C^R.$$

Since $C_{n_k} \cap \overline{B_R} \subset C_{n_k}$, it holds that $C^R \subset C$. On the other hand, given $x \in C \cap B_R$, if there exists $x_n \in C_n \cap \overline{B_R}$ with $x_n \to x$, then $x \in C^R$. Therefore

$$\overline{C \cap B_R} \subset C^R \subset C \cap \overline{B_R}$$

and to finish the proof it suffices to show that there is a countable set $I \subset [R_0, +\infty)$ such that if $R \notin I$, $R > R_0$, then $C \cap \overline{B_R} = \overline{C \cap B_R}$.

Let $\xi \in \partial B_1$ and consider the function $R \mapsto \operatorname{dist}(R\xi, C \cap \overline{B_R})$. If $R > R' \ge R_0$ it holds that

$$\operatorname{dist}(R\xi, C \cap \overline{B_R}) \le \operatorname{dist}(R'\xi, C \cap \overline{B_{R'}}) + R - R'.$$

Indeed, let $x_{R'}$ be the point minimizing the distance from $R'\xi$ to $C \cap \overline{B_{R'}}$, then

$$\begin{aligned}\operatorname{dist}(R\xi, C \cap \overline{B_R}) &\le d(R\xi, x_{R'}) \le d(R\xi, R'\xi) + d(R'\xi, x_{R'}) \\ &= \operatorname{dist}(R'\xi, C \cap \overline{B_{R'}}) + R - R'.\end{aligned}$$

Hence the function $\varphi_\xi : R \mapsto \operatorname{dist}(R\xi, C \cap \overline{B_R}) - R$, is nonincreasing in $[R_0, +\infty)$ and in particular it has at most a countable number of discontinuity points. In addition, given $\xi, \xi' \in \partial B_1$, it holds that

$$\begin{aligned}|\varphi_\xi(R) - \varphi_{\xi'}(R)| &= \left| \inf_{x\in\overline{B_R}} d(x, R\xi) - \inf_{x\in\overline{B_R}} d(x, R\xi') \right| \\ &\le \sup_{x\in\overline{B_R}} |d(x, R\xi) - d(x, R\xi')| \le R|\xi - \xi'|.\end{aligned}$$

Therefore if R is a point of discontinuity for φ_ξ, then for all ξ' in a neighborhood of ξ, R is a point of discontinuity for $\varphi_{\xi'}$.

Let $(\xi_n)_{n\in\mathbb{N}}$ be a dense sequence in ∂B_1. For each n we can find a countable subset $I_n \subset [R_0, +\infty)$, such that φ_{ξ_n} is continuous at any $R \in (R_0, +\infty) \setminus I_n$. Finally, we define the countable set I as $I = \bigcup_{n\in\mathbb{N}} I_n$.

If $R \notin I$, then either $R < R_0$ and $C \cap \overline{B_R} = \overline{C \cap B_R} = \emptyset$, or $R \geq R_0$. In that case, for any $\xi \in \partial B_1$, φ_ξ is continuous. Otherwise, there would be some ξ_n, close enough to ξ, such that φ_{ξ_n} is discontinuous, a contradiction. Let $x = R\xi \in C$, the continuity of φ_ξ implies that

$$\lim_{R'\uparrow R} \operatorname{dist}(R'\xi, C \cap \overline{B_{R'}}) = 0.$$

Hence take $R_n \uparrow R$, set $\varepsilon_n \stackrel{\text{def.}}{=} \operatorname{dist}(R_n\xi, C \cap \overline{B_{R_n}})$ and let $x_n \in C \cap \overline{B_{R_n}}$ be a vector attaining this distance. As $x_n \in C \cap B_R$ and $|x - x_n| \leq \varepsilon_n + R - R_n$, x_n converges to x, and $x \in \overline{C \cap B_R}$. It follows that $(C \cap \overline{B_R}) \setminus \overline{C \cap B_R} = \emptyset$, completing the proof. □

2. Gołab's Theorem

Besides compactness, the second property we need to attack a variational problem is lower semi-continuity. The most basic quantity we might wish to minimize in the context of this thesis is the length of a set, that is, its $\mathscr{H}^1$-measure. Although $\Sigma \mapsto \mathscr{H}^1(\Sigma)$ is not lower semi-continuous in general, as one can find examples of sequences converging in the Hausdorff sense contradicting the lower semi-continuity see [Morel and Solimini, 2012, Chapter 10], it is lower semi-continuous among sequences $(\Sigma_n)_{n\in\mathbb{N}}$ of connected sets. This is the thesis of Gołab's Theorem.

Lemma 2.8 (Gołab's Theorem)**.** *Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a sequence of closed and connected subsets of a complete metric space converging in the Hausdorff metric to Σ. Then Σ is a closed and connected set and*

$$\mathscr{H}^1(\Sigma) \leq \liminf_{n\to\infty} \mathscr{H}^1(\Sigma_n). \tag{2.4}$$

Corollary 2.9. *If $(\Sigma_n)_{n\in\mathbb{N}}$ has a uniformly bounded number of connected components, then (2.4) also holds.*

Proof. Suppose that the sequence $(\Sigma_n)_{n\in\mathbb{N}}$ can be decomposed in at most k families of connected sets

$$\Sigma_n = \bigcup_{i=1}^k \Sigma_n^i, \quad \text{for each } n \in \mathbb{N}.$$

So up to subsequences we can assume that for each $i = 1, \ldots, k$ we have the convergence $\Sigma_n^i \xrightarrow[n\to\infty]{d_H} \Sigma^i$, so that applying Gołab's Theorem k times yields

$$\mathscr{H}^1(\Sigma) \leq \sum_{i=1}^k \mathscr{H}^1(\Sigma^i) \leq \sum_{i=1}^k \liminf_{n\to\infty} \mathscr{H}^1(\Sigma_n^i) \leq \liminf_{n\to\infty} \sum_{i=1}^k \mathscr{H}^1(\Sigma_n^i) = \liminf_{n\to\infty} \mathscr{H}^1(\Sigma_n).$$

□

Besides the original proof of the original result from [Gołąb, 1928] there has been many alternative proofs of this result. In the sequel we give a (non-exhaustive) list of different approaches used in the literature to prove/generalize this result.

- In [Morel and Solimini, 2012], a general lower semi-continuity result in metric spaces is derived for $\mathscr{H}^\alpha$ measures, under the assumption that Σ_n satisfies a *uniform concentration property*. It is then shown in $\mathbb{R}^d$ that sequences of connected sets satisfy this property and Lemma 2.8 follows.

- Lemma 2.8 is also proved in [Ambrosio and Tilli, 2004] and [Paolini and Stepanov, 2013] using abstract density properties in metric spaces.

- The approach of [Alberti and Ottolini, 2017] is to find a curve that passes through each point of a general continuum exactly two times, reducing the problem to the lower semi-continuity of the length of a curve, which is a direct consequence of the definition.

In the sequel, we consider a sequence of continua $(\Sigma_n)_{n\in\mathbb{N}}$ converging to Σ in the sense of Kuratowski. We are mostly interested in the sequence of measures $\mathscr{H}^1 \llcorner \Sigma_n$, up to subsequences, we can always assume it to converge weakly to a measure μ. In the spirit of Gołab's Theorem we expect that $\mu \geq \mathscr{H}^1 \llcorner \Sigma$. This is proven in the sequel, and we shall call this result as the *density version of Gołab's Theorem*. This result is hidden in the proof found in [Ambrosio and Tilli, 2004] of the usual thesis of Gołab's Theorem in metric spaces, see also [Paolini and Stepanov, 2013]. But while these works assume Hausdorff convergence, we assume the weaker Kuratowski convergence and we do not restrict the sequence to be bounded, in fact it can have infinite length, as long as it is locally finite.

Theorem 2.10 (Density version of Gołab's Theorem)**.** *Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a sequence of closed and connected subsets of $\mathbb{R}^d$ converging in the sense of Kuratowski to some closed set Σ and having locally uniform finite length,* i.e. *for all $R > 0$*

$$\sup_{n\in\mathbb{N}} \mathscr{H}^1(\Sigma_n \cap B_R(x_0)) < +\infty.$$

Define the measures $\mu_n \overset{\text{def.}}{=} \mathscr{H}^1 \llcorner \Sigma_n$, and let μ be a weak-$\star$ cluster point of this sequence. Then supp$\mu \subset \Sigma$ and it holds that

$$\mu \geq \mathscr{H}^1 \llcorner \Sigma,$$

in the sense of measures.

Proof of Theorem 2.10. We will show that $\mu(\Sigma \cap B_r(y_0)) \geq \mathscr{H}^1(\Sigma \cap B_r(y_0))$ for $\mathscr{H}^1$-a.e. $y_0 \in \Sigma$ and for all $r > 0$ small enough. This implies that $\theta_1(\mu, y_0) \geq 1$, and the result follows by integrating. Assume that Σ is not a singleton, otherwise there is nothing to prove, so that taking any $y_0 \in \Sigma$, for $r > 0$ small enough $\Sigma \cap B_r^c(y_0) \neq \emptyset$. From the Kuratowski convergence, for n large enough, each set Σ_n has a point inside and another outside the ball $\overline{B_r(y_0)}$.

We start by fixing some $0 < \delta < r$ and looking at the smaller ball $B_{r-\delta}(y_0)$. Consider the following class

$$\mathcal{A}_n \stackrel{\text{def.}}{=} \left\{ \gamma \text{ connected component of } \Sigma_n \cap \overline{B_r(y_0)} \text{ which intersects } B_{r-\delta}(y_0) \right\}.$$

Each $\gamma \in \mathcal{A}_n$ must be such that $\mathscr{H}^1(\gamma) \geq \delta$. Indeed, as for each $n \in \mathbb{N}$ there is a point in $\Sigma_n \cap B_r(y_0)^c$ and another in $\gamma \cap \partial B_{r-\delta}(y_0)$, the connectedness implies γ is contained in an arc joining these two points, but then it must have length at least δ, as it is the smallest distance between the two balls. So define

$$\tilde{\Sigma}_n \stackrel{\text{def.}}{=} \bigcup_{\gamma \in \mathcal{A}_n} \gamma,$$

which is a bounded sequence of closed sets, but not necessarily connected. However this sequence has a uniformly bounded number of connected components since

$$\delta \sharp \mathcal{A}_n \leq \sum_{\gamma \in \mathcal{A}_n} \mathscr{H}^1(\gamma) \leq \mathscr{H}^1(\Sigma_n \cap B_R(x_0)), \text{ hence } \sharp \mathcal{A}_n \leq \sup_{n \in \mathbb{N}} \frac{\mathscr{H}^1(\Sigma_n \cap B_R(y_0))}{\delta} < +\infty,$$

for $R > 0$ large enough.

As $\tilde{\Sigma}_n$ is a bounded sequence, by Blaschke's Theorem we can assume up to an extraction that $\tilde{\Sigma}_n \xrightarrow[n \to \infty]{d_H} \tilde{\Sigma}$. In fact, for a.e. $0 < \delta < r$, using Lemma 2.7, it holds that

$$\tilde{\Sigma} \cap \overline{B_{r-\delta}(y_0)} = \overline{\Sigma \cap B_{r-\delta}(y_0)}, \tag{2.5}$$

since by the construction, $\tilde{\Sigma}_n \cap \overline{B_{r-\delta}(y_0)} = \Sigma_n \cap \overline{B_{r-\delta}(y_0)}$ and choosing δ such that $\Sigma_n \cap \overline{B_{r-\delta}(y_0)} \xrightarrow[n \to \infty]{K} \overline{\Sigma \cap B_{r-\delta}(y_0)}$.

This way, we can apply the global version of Gołab's Theorem with a uniformly bounded number of connected components to the sequence $\tilde{\Sigma}_n \cap B_{r-\delta}(y_0)$ so that we write

$$\begin{aligned}
\mu\left(\overline{B_r(y_0)}\right) &\geq \limsup_{n \to \infty} \mathscr{H}^1\left(\Sigma_n \cap B_r(y_0)\right) \geq \limsup_{n \to \infty} \mathscr{H}^1\left(\tilde{\Sigma}_n\right) \\
&\geq \liminf_{n \to \infty} \mathscr{H}^1\left(\tilde{\Sigma}_n \cap B_{r-\delta}\right) \\
&\geq \mathscr{H}^1\left(\tilde{\Sigma} \cap B_{r-\delta}(y_0)\right) = \mathscr{H}^1\left(\overline{\Sigma \cap B_{r-\delta}(y_0)}\right) \\
&\geq \mathscr{H}^1\left(\Sigma \cap B_{r-\delta}(y_0)\right),
\end{aligned}$$

where the first inequality is due to the weak-$\star$ convergence of the measures and the forth is given by Gołab's Theorem. But as this estimate is true for any $\delta > 0$, it must hold that

$\mu\left(\overline{B_r(y_0)}\right) \geq \mathscr{H}^1\left(\Sigma \cap B_r(y_0)\right)$ for any $y_0 \in \Sigma$ and $r > 0$. To extend this to open balls as well we use the following estimates

$$\mu(B_r) = \lim_{n\to\infty} \mu\left(\overline{B_{r-1/n}}\right) \geq \lim_{n\to\infty} \mathcal{H}^1\left(\Sigma \cap B_{r-1/n}\right) = \mathcal{H}^1\left(\Sigma \cap B_r\right).$$

□

Remark 2.11. *As we have not used any properties from the vector space structure of* $\mathbb{R}^d$*, this proof works in the case a locally compact metric space, as in [Ambrosio and Tilli, 2004].*

The technology developed so far allows us to define the most basic, and still extremely relevant, 1D-shape optimization problem, the Steiner tree problem [Brazil et al., 2014], [Paolini and Stepanov, 2013]. It can be formulated as follows: given some Borel set K, we seek a network that will connect while minimizing the quantity

$$\inf\left\{\mathscr{H}^1(\Sigma): \quad K \subset \Sigma \text{ and } \Sigma \text{ is connected}\right\}, \tag{2.6}$$

with possibly infinite length. With an application of the Direct method of the calculus of variations, one can prove the existence of a minimal network by combining Blaschke's and Gołab's Theorems, whenever the infimum is finite.

3. Blowup of 1-dimensional sets

From Proposition 2.2, we know that connected sets Σ with finite length are actually countably $\mathscr{H}^1$-rectifiable, and as such they enjoy the properties described in the previous chapter, such as the existence of an approximate tangent space and of blow-ups, in the measure theoretical sense. In other words, we know that for *a.e.* $x \in \Sigma$, it holds that

$$\mathscr{H}^1 \llcorner \left(\frac{\Sigma - x}{r}\right) \xrightarrow[r\to 0]{\star} \mathscr{H}^1 \llcorner T_x\Sigma. \tag{2.7}$$

Differently from the k-dimensional case, we can use the tools developed so far to prove the convergence of blow-ups in the Hausdorff and Kuratowski topologies.

Lemma 2.12. *Let* $\Sigma \subset \mathbb{R}^d$ *be closed and connected with* $\mathscr{H}^1(\Sigma) < +\infty$*, then for every* $x \in \Sigma$ *admitting an approximate tangent space* $T_x\Sigma$ *as in Thm. 1.16, and for all* $R > 0$ *it holds that*

$$\frac{\Sigma - x}{r} \cap \overline{B_R(0)} \xrightarrow[r\to 0^+]{d_H} T_x\Sigma \cap \overline{B_R(0)}, \tag{2.8}$$

and for every $r >$ *it holds that*

$$d_H\left(\Sigma \cap B_r(x) - x, T_x\Sigma \cap B_r(0)\right) = r d_H\left(\frac{\Sigma - x}{r} \cap B_1, T_x\Sigma \cap B_1\right) = o(r). \tag{2.9}$$

In particular, the global convergence holds in the Kuratowski sense

$$\frac{\Sigma - x}{r} \xrightarrow[r\to 0^+]{K} T_x\Sigma.$$

Proof. First we take a rectifiability point $x \in \Sigma$ with tangent space $T_x\Sigma$, which we know to be $\mathscr{H}^1$ a.a. of Σ, so that (2.7) holds. Let T be the (Kuratowski) limit of a subsequence $\dfrac{\Sigma - x}{r_k}$. From (2.7) we have that $T_x\Sigma \subset T$. Thanks to Lemma 2.7 and Theorem 2.10, for almost all $R > 0$ if holds that

$$\mathscr{H}^1(T \cap B_R(0)) \leq \liminf_{k\to\infty} \mathscr{H}^1\left(\frac{\Sigma - y}{r_k} \cap B_R(0)\right) = \mathscr{H}^1(T_y\Sigma \cap B_R(0)), \tag{2.10}$$

which shows $T \Delta T_x\Sigma$ is $\mathscr{H}^1$-negligible.

Notice that, if there is some $z \in T \setminus T_x\Sigma$, we may consider some ball $B_s(z)$ which does not intersect $T_x\Sigma$. Since T is the limit of connected sets, z must be path-connected in T to some point in $(B_s(z))^c$, so that $\mathscr{H}^1(T \cap B_s(z)) \geq s$. This contradicts (2.10). Hence, $T = T_x\Sigma$, and is independent of the subsequence, and we deduce the localized Hausdorff and the Kuratowski convergences.

To check (2.9), notice that from homogeneity of the distance in $\mathbb{R}^d$ it holds that

$$\frac{d_H\left((\Sigma - x) \cap B_r, T_x\Sigma \cap B_r\right)}{r} = d_H\left(\frac{\Sigma - x}{r} \cap B_1, T_x\Sigma \cap B_1\right)$$

and the RHS converges to zero as $r \to 0$ from the previous reasoning. $\square$

In the sequel we present a slight stronger consequence of the blow-up in the Hausdorff topology, instead of the weak measure theoretical blow-up. We exploit the characterization of the Hausdorff distance by means of inclusions in neighborhoods around the limit (2.3) to show that if Σ is flat at a point x_0, admits an approximate tangent space, any slice perpendicular to the tangent space $T_{x_0}\Sigma$ will contain a point of Σ near x_0. In the sequel we show that this property can be partially transferred to any sequence of connected sets $(\Sigma_\varepsilon)_{\varepsilon>0}$ that converge to Σ.

Theorem 2.13. *Let $\Sigma \subset \mathbb{R}^d$ be closed and connected with $\mathscr{H}^1(\Sigma) < +\infty$, then the following hold. Assume that Σ is flat at x_0 with approximate tangent space $T_{x_0}\Sigma = \mathbb{R}\tau$, for $\tau \in \mathbb{S}^{d-1}$ and let π_τ denote the projection onto it.*

(1) For $1/2 < \delta < 1$, there is some r_0 such that

$$[-\delta r, \delta r]\tau \subset \pi_\tau(\Sigma \cap B_r(x_0)), \text{ for all } r < r_0.$$

That is, for any $t \in [-\delta r, \delta r]$ there is $x \in \Sigma \cap B_r(x_0)$ such that $\langle \tau, x\rangle = t$. In addition, x belongs to the connected component of $\Sigma \cap B_r(x_0)$ that contains x_0.

(2) Let $(\Sigma_\varepsilon)_{\varepsilon>0}$ be a family of connected sets such that $\Sigma_\varepsilon \xrightarrow[\varepsilon\to 0]{d_H} \Sigma$. Then for $1/2 < \delta < 1$, there are r_0 and ε_0 such that, if $r < r_0$ and $\varepsilon < \varepsilon_0$, for each $t \in (-\delta r, \delta r)$, there exists

$$x \in \Sigma_\varepsilon \cap B_r(x_0), \text{ such that } \pi_\tau(x) = x_0 + t\tau, \tag{2.11}$$

except in a set that is either a singleton, or a connected interval $(a_\varepsilon, b_\varepsilon)$ such that $b_\varepsilon - a_\varepsilon \leq 2 d_H(\Sigma_\varepsilon, \Sigma)$.

Proof. Item (1) is proven in [Bonnivard et al., 2015] in the case $d = 2$, for completeness we prove it here in $\mathbb{R}^d$. Using (2.9), proved in Lemma 2.12, take r_0 small enough such that

$$d_H\left(\frac{\Sigma \cap B_r(x_0) - x_0}{r}, [-\tau, \tau]\right) \leq (1 - \delta),$$
$$\Sigma \cap B_{\delta r}(x_0) \subset [-\delta r, \delta r]\tau + B_{(1-\delta)r}(x_0). \tag{2.12}$$

Therefore, there must be points $z_+, z_- \in (\Sigma \cap B_r(x_0) - x_0)$ such that $|z_\pm - (\pm\tau)| \leq (1-\delta)r$ and paths $\gamma_\pm \subset [-\delta r, \delta r]\tau + B_{(1-\delta)r}(x_0)$ connecting x_0 and $z_\pm$. Therefore, we must have that $\pi_\tau(z_+) \geq \delta r$ and $\pi_\tau(z_-) \geq -\delta r$ so that $\pi_\tau(\gamma_+)$ (resp. $\pi_\tau(\gamma_-)$) must be a connected set containing $x_0 + [0, \delta r]$ (resp. $x_0 + [-\delta r, 0]$).

Item (2) can be interpreted as a partial transfer of property (2) to any sequence of connected sets Σ_ε converging to Σ, up to a small set that can be quantified. From the Hausdorff convergence in item (1), we can choose r_0 such that for $r < r_0$ we have

$$\Sigma \cap B_{\delta r}(x_0) \subset x_0 + B_{\frac{r\delta'}{2}}([-\tau, \tau]), \text{ with } \delta' = \sqrt{1 - \delta^2}.$$

In addition, the cylinder

$$C_{\delta,r}(x_0) \overset{\text{def.}}{=} \left\{ x : \begin{array}{c} |\pi_\tau(x - x_0)| < \delta r \\ |\pi_{\tau^\perp}(x - x_0)| < \delta' r \end{array} \right\} \text{ is contained in the ball } B_{\delta r}(x_0).$$

Suppose by contradiction that the set of points that do not satisfy (2.11) is disconnected and take tow points t and t' in two distinct connected components. From property (2), between these sections there is a point of Σ, *i.e.* there exists $y \in \Sigma$ inside the smaller cylinder $\pi_\tau^{-1}((t'r, tr)\tau) \cap C_{\delta,r}(x_0)$.

From the Hausdorff convergence of Σ_ε to Σ, for ε small enough, there exists $y_\varepsilon \in \Sigma_\varepsilon$ that can be made arbitrarily close to y taking ε small enough, so that

$$y_\varepsilon \in \pi_\tau^{-1}((t'r, tr)\tau) \cap C_{\delta,r}(x_0), \text{ and } \Sigma_\varepsilon \cap B_r(x_0) \subset x_0 + B_{\delta' r}([-\tau, \tau]).$$

But then as Σ_ε is connected, there is a path connecting y_ε to some point of Σ_ε outside $B_r(x_0)$. This path must then intersect $(\pi_\tau)^{-1}(\{t'\tau, t\tau\})$, which contradicts the fact that t and t' do not satisfy (2.11).

This proves that if the set of values $t \in (-\delta r, \delta r)$ not satisfying (2.11) is either a singleton or a connected interval. In the latter case, assume it to be given by $(a_\varepsilon, b_\varepsilon) \subset (-\delta r, \delta r)$; suppose that $b_\varepsilon - a_\varepsilon > 2d_H(\Sigma_\varepsilon, \Sigma)$. In this case take

$$y \in \Sigma \cap B_r(x_0), \text{ such that } \langle \tau, y_0 \rangle = \frac{b_\varepsilon + a_\varepsilon}{2} > d_H(\Sigma_\varepsilon, \Sigma),$$

such point exists from item (2). This means that the closest point of Σ_ε to y is at distance bigger than $d_H(\Sigma_\varepsilon, \Sigma)$, which contradicts the definition of the Hausdorff distance between Σ_ε and Σ. $\square$

4. Loops and tree structure

In the original formulation of the Steiner problem, defined in (2.6), the set K is a discrete set of points in $\mathbb{R}^2$. It can then be proven that any optimal network is a tree (in the classical sense of graph theory [Diestel, 2017, Chap. 1.5]) made of finitely many segments connected by triple junctions forming 120 degrees. The property of being a tree, or not having cycles, is very economical and is independent of the underlying geometry, so it is natural to expect that it will also be the case for any solution of the Steiner problem, even in a metric setting. Hence, we give a definition of loop that is topological rather than geometric.

Definition 2.14. *We say that a set Γ is a loop whenever it is homeomorphic to $\mathbb{S}^1$. Any connected set Σ which contains no loops it is said to be a tree.*

A point $x \in \Sigma$ is a non-cut point of Σ if $\Sigma \setminus \{x\}$ remains connected. Otherwise, x is called a cut point.

It turns out that $\mathscr{H}^1$ almost every point in a loop is a non-cut point. This is proved for instance in [Paolini and Stepanov, 2013, Lemma 5.6] when the ambient space is a general metric space. In the following Lemma, we exploit the geometric structure of $\mathbb{R}^d$ to prove this result, while more information in the process.

Lemma 2.15. *Let $\Sigma \subset \mathbb{R}^d$ be a closed connected set with $\mathscr{H}^1(\Sigma) < +\infty$, consisting of more than one point and containing a loop Γ. Then $\mathscr{H}^1$-a.e. point $x \in \Gamma$ is such that for any $r > 0$ small enough, there exists $\bar{r} \in \left(\frac{r}{2}, r\right)$, such that $\Sigma \setminus B_{\bar{r}}(x)$ and $\Sigma \cap B_{\bar{r}}(x)$ are connected and*

$$\mathscr{H}^0(\Sigma \cap \partial B_{\bar{r}}(x)) = \mathscr{H}^0(\Gamma \cap \partial B_{\bar{r}}(x)) = 2.$$

In addition, it holds that $\mathscr{H}^1$-a.e. point of Γ is a non-cut point.

Proof. Let Γ be a loop of Σ, from Thm. 1.16 and [Maggi, 2012, Prop. 10.5], we know that $\mathscr{H}^1$-a.e. point of $\Sigma \cap \Gamma$ admits an approximate tangent plane such that

$$T_x\Sigma = T_x\Gamma.$$

Fix one such point x where the approximate tangents w.r.t. Σ and Γ coincide and let $\mathbb{R}\tau$ be the common tangent space. Given $r > 0$, it holds from the area formula and point (iii) of Thm. 1.16 that

$$\int_0^r \mathscr{H}^0(\partial B_s(x) \cap \Gamma)\mathrm{d}s \leq \int_0^r \mathscr{H}^0(\partial B_s(x) \cap \Sigma)\mathrm{d}s \leq \mathscr{H}^1(B_r(x) \cap \Sigma) = 2r + o(r). \quad (2.13)$$

In addition, from the Hausdorff convergence of the blow-ups from $\Sigma \cap B_r(x)$, Lemma 2.12, we can assume for n large enough that

$$\Sigma \cap B_r(x) \subset \left\{ z : \begin{array}{c} |\langle z - x, \tau \rangle| < r \\ |\langle z - x, \tau^\perp \rangle| < \frac{r}{100} \end{array} \right\}.$$

Since $\dfrac{\Gamma - x}{r}$ is a curve converging to the segment $\mathbb{R}\tau$, it must cross all the surfaces

$$\partial \left(B_s(0) \cap \{\pm \langle z, \tau \rangle > 0\} \right) \quad 0 < s < r,$$

so that $2 \le \mathscr{H}^0(\Gamma \cap \partial B_s(x)) \le \mathscr{H}^0(\Sigma \cap \partial B_s(x))$. As a result, from (2.13) we have that

$$0 \le \frac{1}{r} \int_0^r \underbrace{\left(\mathscr{H}^0(\partial B_s(x) \cap \Gamma) - 2 \right)}_{\ge 0} \mathrm{d}s \le \frac{o(r)}{r}.$$

Hence, for r small enough, we can find

$$\bar{r} \in \left(\frac{r}{2}, r \right) \text{ such that } \mathscr{H}^0(\Sigma \cap \partial B_{\bar{r}}(x)) = \mathscr{H}^0(\Gamma \cap \partial B_{\bar{r}}(x)) = 2.$$

For such radius we have that $\partial B_{\bar{r}}(x) \cap \Sigma = \partial B_{\bar{r}}(x) \cap \Gamma = \{y_{1,n}, y_{2,n}\}$ and $\Gamma \setminus B_{\bar{r}}(x)$ is a path between $y_{1,n}$ and $y_{2,n}$.

It follows that both $\Sigma \cap B_{\bar{r}}(x)$ and $\Sigma \setminus B_{\bar{r}}(x)$ remain connected. Indeed, for the former, it suffices to notice that since $\mathscr{H}^0(\Gamma \cap B_{\bar{r}}(x)) = 2$, $\Gamma \cap B_{\bar{r}}(x)$ is homeomorphic to an arc of $\mathbb{S}^1$ and so it is connected, as continuous images of connected sets are connected. As a result, it must also hold that $\Sigma \cap B_{\bar{r}}(x)$ is connected since if it was not, there would a connected component Γ' that is disjoint fom $\Gamma \cap B_{\bar{r}}(x)$. But since $\Sigma \cap \partial B_{\bar{r}}(x) = \Gamma \cap \partial B_{\bar{r}}(x)$, Γ' would also be disjoint from $\Sigma \setminus B_{\bar{r}}(x)$, contradicting the connectedness of Σ.

To prove the connectedness of $\Sigma \setminus B_{\bar{r}}(x)$, consider $z_1, z_2 \in \Sigma \setminus B_{\bar{r}}(x)$ and let $\gamma \subset \Sigma$ be a path between them. If $\gamma \subset \Sigma \setminus B_{\bar{r}}(x)$, there is nothing to prove, otherwise γ must contain either $y_{1,n}$ $y_{2,n}$, or both. If it contains only one of them, $\gamma \setminus B_{\bar{r}}(x)$ remains connected. In the case that it contains both, we can create a new path $\gamma \cup \Gamma \setminus B_{\bar{r}}(x)$ that must be connected, contained in $\Sigma \setminus B_{\bar{r}}(x)$ and has the points z_1, z_2. It follows that $\Sigma \setminus B_{\bar{r}}(x)$ is connected.

Let us show that x is a non-cut point. Indeed, for any $y_1, y_2 \in \Sigma \setminus \{x\}$, use the previous construction to obtain a radius such that $\Sigma \setminus B_r(x)$ is connected and contains y_1, y_2. Therefore, we can find a path in $\Sigma \setminus \{x\}$ connecting them proving that $\{x\}$ is a non-cut point. $\square$

The previous Lemma will be used in Chapter 4 to show that minimizers of a variational problem are trees. If x is a general non-cut point of Σ, a more involved construction proposed in [Buttazzo and Stepanov, 2003], gives a sequence of connected sets D_n, containing x, whose diameters converge to 0 and such that $\Sigma \setminus D_n$ remains connected. This construction is useful in the proof of absence of loops for many geometric variational problems: see for instance [Paolini and Stepanov, 2013] for the Steiner problem, [Buttazzo and Stepanov, 2003] and [Santambrogio and Tilli, 2005] for the average distance minimizers problem, [Chambolle et al., 2017] for the optimal compliance problem, and it is also used in Chapter 4 to prove that solutions to the new problem proposed in Chapter 3 of this thesis are trees.

Chapter 3

The Wasserstein-$\mathscr{H}^1$ problem

Contents

1. Introduction

Consider the following 1D-shape optimization problem: given a reference probability measure $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$ (the set of probability measures ρ with $\int_{\mathbb{R}^d} |x|^p \mathrm{d}\rho < +\infty$, $p \geq 1$), we seek to approximate ρ_0 with measures supported over a connected 1-dimensional subset of $\mathbb{R}^d$. This approximation is done by means of the following variational problem

$$\inf_{\Sigma \in \mathcal{A}} W_p^p(\rho_0, \nu_\Sigma) + \Lambda \mathscr{H}^1(\Sigma), \tag{P_Λ}$$

where the measure ν_Σ is defined as

$$\nu_\Sigma \overset{\text{def.}}{=} \frac{1}{\mathscr{H}^1(\Sigma)} \mathscr{H}^1 \llcorner \Sigma, \text{ for } \Sigma \in \mathcal{A} \overset{\text{def.}}{=} \left\{ \Sigma \subset \mathbb{R}^d : \begin{array}{l} 0 < \mathscr{H}^1(\Sigma) < +\infty \\ \text{compact, connected.} \end{array} \right\}, \tag{3.1}$$

and $\mathscr{H}^1$ denotes the 1-dimensional Hausdorff measure in $\mathbb{R}^d$. The term W_p denotes the usual Wasserstein distance on the space of probability measures (see [Santambrogio, 2015, Villani, 2009]).

One can trace the idea of approximating a probability measure by a 1D set back to the concept of principal curves from the seminal paper [Hastie and Stuetzle, 1989], which extends linear regression to regression using general curves, and introduces a variational problem to define such curves. In this variational sense, a principal curve minimizes the expectation of the distance to the curve, w.r.t. a probability measure describing a data set (with some regularization to ensure existence). This problem was introduced with the goal of performing regression in a way that intrinsically describes the data, instead of relying on a previous belief that it can be accurately described by some class of curves, such as splines. As proposed in [Kégl et al., 2000], a length constraint is a simple and intrinsic way to ensure existence. The properties of such minimizers have been studied in detail in e.g. [Lu and Slepčev, 2016, Delattre and Fischer, 2020].

A further generalization consists in replacing the curve with a more general 1-dimensional compact and connected set, yielding the *average distance minimizer problem* introduced in [Buttazzo and Stepanov, 2003], and its dual counterpart *maximum distance minimizer problem* [Paolini and Stepanov, 2004, Lemenant, 2010]. Such problems were conceived for applications in urban planning, where one seeks to minimize the average distance to a transportation network, giving rise to the need for a larger class of 1D sets that might present bifurcations.

While the above-mentioned problems only focus on some geometric approximation of the support of the measure, approximating a measure in the sense of weak convergence is sometimes more desirable. In [Lebrat et al., 2019, Chauffert et al., 2017], the authors have proposed optimal transport based methods for the projection of probability measures onto classes of measures supported on simple curves, using the Wasserstein distance as a data term. Potential applications range from 3D printing to image compression and reconstruction. In [Ehler et al., 2021], the data fidelity term is chosen to be a discrepancy, see also [Neumayer and Steidl, 2021]. The advantage of using discrepancies is that approximation rates can be given independently from the dimension, being therefore a good alternative to overcome the curse of dimensionality. The problem we study is an attempt to generalize this class of problems to the approximation with one-dimensional connected sets.

One difficulty when studying (P_Λ) is that the class of measures ν_Σ is not closed in the usual weak topologies considered for the space of probability measures. While a sequence of sets $(\Sigma_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with uniformly bounded length will have subsequences converging (in the Hausdorff sense) either to a point or a set in $\mathcal{A}$, the corresponding measures ν_{Σ_n} might converge to a measure which is not necessarily uniform on that set: longer parts of Σ_n might concentrate in the limit on shorter parts of Σ, as illustrated in Figure 1.

Hence, minimizing sequences converge in general to a measure which is not of the form ν_Σ, and we need to determine a relaxation of our energy in a topology for which the Wasserstein distance is lower semi-continuous, such as the narrow convergence. The relaxed energy takes the form

$$\inf_{\nu\in\mathcal{P}_p(\mathbb{R}^d)} W_p^p(\rho_0,\nu)+\Lambda\mathcal{L}(\nu), \tag{$\overline{P}_\Lambda$}$$

where the length functional $\mathcal{L}$, defined in Section 2.1, generalizes the notion of length of the support of a measure, see for instance Example 2.2. We will show later on, in Proposition 3.6, that $\mathcal{L}$ is the lower semi-continuous relaxation, for the narrow topology, of the functional ℓ given by $\mathscr{H}^1(\Sigma)$ for measures of the form ν_Σ, and $+\infty$ else, see (3.2). We also find that $\mathcal{L}(\nu)<\infty$ if and only if $\operatorname{supp}\nu\in\mathcal{A}$ or ν is a Dirac mass

The following theorem gathers the various results proved throughout this chapter.

Theorem 3.1. *Let $\rho_0\in\mathscr{P}_p(\mathbb{R}^d)$, $\Lambda>0$. Then ($\overline{P}_\Lambda$) admits a solution ν, and there exists $\Lambda_\star\geq 0$ such that if $\Lambda>\Lambda_\star$, ν is a Dirac mass. For $\Lambda<\Lambda_\star$, ν is supported by a set $\Sigma\in\mathcal{A}$ and the following properties hold.*

1. *If ρ_0 is absolutely continuous w.r.t. $\mathscr{H}^1$, or has an L^∞ density w.r.t. $\mathscr{H}^1$, then so does ν.*

2. *If ρ_0 does not give mass to $1D$ sets, then $\nu=\nu_\Sigma$ and therefore is a solution to the original problem (P_Λ).*

This chapter is organized as follows: in Section 2 we go through the definition of the length functional and its properties as well as the relaxed problem and existence of solution for it. In Section 3 we discuss the existence of Λ_*. In Section 4 we prove point (1) from Theorem 3.1, while the existence is proved in Section 5 (Theorem 3.24).

After introducing the problem and showing the results described above, that make it well-posed, in the present Chapter we shall continue its study in Chapters 4 and 5, where we prove further qualitative properties of minimizers such as Ahlfors regularity and absence of loops, as well as provide a phase-field approximation result in the form of Γ-convergence.

2. The length functional and the relaxed problem

If a minimizing sequence Σ_n converges to some set Σ, we cannot expect weak cluster points of the measures ν_{Σ_n} to have the form ν_Σ, see Figure 1. Hence the objective of (P_Λ) is not lower semi-continuous for the narrow convergence, and, in this section, we introduce a relaxation for (P_Λ). First, we define a functional which extends the length of the support and we discuss some of its properties, then we use it to define the relaxed problem.

2.1. Definition and elementary properties

Recalling that $\mathcal{A}$ is the collection of the compact connected sets $\Sigma \subset \mathbb{R}^d$ with $0 < \mathscr{H}^1(\Sigma) < +\infty$, we consider

$$\ell : \mathcal{P}(\mathbb{R}^d) \ni \nu \mapsto \begin{cases} \mathscr{H}^1(\Sigma), & \text{if } \nu = \dfrac{1}{\mathscr{H}^1(\Sigma)} \mathscr{H}^1 \llcorner \Sigma \text{ for some } \Sigma \in \mathcal{A}, \\ +\infty, & \text{otherwise,} \end{cases} \tag{3.2}$$

so that (P_Λ) becomes $\inf_\nu W_p^p(\rho_0, \nu) + \Lambda \ell(\nu)$. As discussed above, ℓ is not l.s.c., hence we introduce the following relaxation, which we call the *length functional*. For any $\nu \in \mathcal{P}(\mathbb{R}^d)$, we define

$$\mathcal{L}(\nu) \overset{\text{def.}}{=} \begin{cases} \inf \left\{ \alpha \geq 0 \mid \alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp} \nu \right\}, & \text{if } \operatorname{supp} \nu \text{ is connected,} \\ +\infty, & \text{otherwise,} \end{cases} \tag{3.3}$$

with the convention that $\inf \emptyset \overset{\text{def.}}{=} +\infty$. Notice that, since ν is a probability measure, $\mathcal{L}(\nu) \geq \mathscr{H}^1(\operatorname{supp} \nu)$, and that $\mathcal{L}(\nu) = 0$ if and only if $\nu = \delta_x$ for some $x \in \mathbb{R}^d$. As a result, $0 < \mathcal{L}(\nu) < \infty$ if and only if $\operatorname{supp} \nu \in \mathcal{A}$. Moreover, for any $\Sigma \in \mathcal{A}$ and $\nu_\Sigma \overset{\text{def.}}{=} \dfrac{1}{\mathscr{H}^1(\Sigma)} \mathscr{H}^1 \llcorner \Sigma$, we have $\mathcal{L}(\nu_\Sigma) = \mathscr{H}^1(\Sigma) = \ell(\nu_\Sigma)$.

Remark 3.2. *Definition (3.3) also makes sense for any positive measure $\mu \in \mathcal{M}_+(\mathbb{R}^d)$. In that case, thanks to Theorem 2.10, it may be easily shown to be lower semi-continuous with respect to the weak convergence, defining $\mathcal{L}(0) = 0$ (see also Section 2.2). Yet then, of course, even for uniformly distributed measures such as $\nu = \theta \mathscr{H}^1 \llcorner \Sigma$ for some $\theta > 0$, its value does not coincide with the length of the support anymore (it rather is $\mathscr{H}^1(\Sigma)/\nu(\mathbb{R}^d)$).*

In Section 2.2 below, we prove that $\mathcal{L}$ is the lower semi-continuous enveloppe of ℓ for the narrow topology of probability measures. Before that, let us discuss some alternative formulations for $\mathcal{L}$. Following [?, Sec. 2.4], we consider the upper derivative,

$$\forall x \in \operatorname{supp} \nu, \quad D_\nu^+(\mathscr{H}^1 \llcorner \operatorname{supp} \nu)(x) \overset{\text{def.}}{=} \limsup_{r \to 0^+} \frac{\mathscr{H}^1(B_r(x) \cap \operatorname{supp} \nu)}{\nu(B_r(x))}. \tag{3.4}$$

Proposition 3.3 (Alternative definitions of $\mathcal{L}$)**.** *Let $\nu \in \mathcal{P}(\mathbb{R}^d)$ be such that* $\operatorname{supp}\nu$ *is connected. Then*

$$\mathcal{L}(\nu) = \sup\left\{ \frac{\mathscr{H}^1(U \cap \operatorname{supp}\nu)}{\nu(U)} \mid U \textit{ open}, U \cap \operatorname{supp}\nu \neq \emptyset \right\} \tag{3.5}$$

$$= \sup\left\{ \frac{\mathscr{H}^1(B_r(x) \cap \operatorname{supp}\nu)}{\nu(B_r(x))} \mid r > 0, x \in \operatorname{supp}\nu \right\} \tag{3.6}$$

$$= \left\| D_\nu^+(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)\right\|_\infty, \tag{3.7}$$

where $\|\cdot\|_\infty$ denotes the supremum norm over $\operatorname{supp}\nu$*.*

Proof. It is immediate that

$$(\text{R.H.S. of (3.3)}) \geq (\text{R.H.S. of (3.5)}) \geq (\text{R.H.S. of (3.6)}) \geq (\text{R.H.S. of (3.7)})\,.$$

Now, assume that $\|D_\nu^+(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)\|_\infty < +\infty$ and let $\alpha > \|D_\nu^+(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)\|_\infty$. For every compact set $K \subset \mathbb{R}^d$ and every $x \in K \cap (\operatorname{supp}\nu)$, there is some $r(x) > 0$ such that $\mathscr{H}^1(B_r(x) \cap (\operatorname{supp}\nu)) \leq \alpha\nu(B_r(x))$. We may extract from the covering $(B_{r(x)}(x))_{x \in K\cap(\operatorname{supp}\nu)}$ with open sets a finite covering $(B_{r_i}(x_i))_{i=1}^N$ of $K \cap (\operatorname{supp}\nu)$. As a result

$$\mathscr{H}^1(K \cap (\operatorname{supp}\nu)) \leq \sum_{i=1}^N \alpha\nu(B_{r_i}(x_i)) \leq N\alpha < +\infty,$$

so that $\mathscr{H}^1 \llcorner (\operatorname{supp}\nu)$ is a Radon measure. We may thus apply [**?**, Prop. 2.21] to deduce

$$(\text{R.H.S. of (3.7)}) \geq (\text{R.H.S. of (3.3)})\,.$$

If $\|D_\nu^+(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)\|_\infty = +\infty$, the inequality holds trivially, which completes the proof. □

The length functional inherits some of the properties of the $\mathscr{H}^1$ measure.

Proposition 3.4. *Let $f : \mathbb{R}^d \to \mathbb{R}^d$, be a k-Lipschitz function, with $k > 0$. Then*

$$\forall \nu \in \mathcal{P}(\mathbb{R}^d), \quad \mathcal{L}(f_\sharp\nu) \leq k\mathcal{L}(\nu). \tag{3.8}$$

Proof. If $\mathcal{L}(\nu) = +\infty$, there is nothing to prove. Otherwise, $\operatorname{supp}\nu$ is compact, and $\operatorname{supp}(f_\sharp\nu) = f(\operatorname{supp}\nu)$. Moreover, for any open set $U \subset \mathbb{R}^d$, since $f^{-1}(U)$ is open,

$$U \cap (\operatorname{supp} f_\sharp\nu) \neq \emptyset \iff \nu(f^{-1}(U)) > 0 \iff f^{-1}(U) \cap (\operatorname{supp}\nu) \neq \emptyset.$$

Now, let U be an open set which intersects $\operatorname{supp}(f_\sharp\nu)$. Using that

$$U \cap f(\operatorname{supp}\nu) \subset f\left(f^{-1}(U) \cap \operatorname{supp}\nu\right)$$

we get

$$\frac{\mathscr{H}^1\left(U\cap\operatorname{supp}(f_\sharp\nu)\right)}{f_\sharp\nu(U)} = \frac{\mathscr{H}^1\left(U\cap f(\operatorname{supp}\nu)\right)}{\nu(f^{-1}(U))} \leq \frac{\mathscr{H}^1\left(f\left(f^{-1}(U)\cap\operatorname{supp}\nu\right)\right)}{\nu(f^{-1}(U))}$$
$$\leq k\frac{\mathscr{H}^1\left(f^{-1}(U)\cap\operatorname{supp}\nu\right)}{\nu(f^{-1}(U))}$$
$$\leq k\mathcal{L}(\nu)$$

since $f^{-1}(U)$ is an open set which intersects $\operatorname{supp}\nu$. Taking the supremum over all U yields the claimed inequality. □

It is also possible to express the length-functional using the Besicovitch differentiation theorem [?, Thm. 2.22]. Assume that $\mathscr{H}^1(\operatorname{supp}\nu) < +\infty$ (otherwise $\mathcal{L}(\nu) = +\infty$). Then, the measure $\mathscr{H}^1 \llcorner \operatorname{supp}\nu$ is Radon, and the limit

$$D_\nu(\mathscr{H}^1\llcorner\operatorname{supp}\nu)(x) \overset{\text{def.}}{=} \lim_{r\to0^+}\frac{\mathscr{H}^1(B_r(x)\cap\operatorname{supp}\nu)}{\nu(B_r(x))} \tag{3.9}$$
$$\left(\text{resp. } D_{\mathscr{H}^1\llcorner\operatorname{supp}\nu}(\nu)(x) \overset{\text{def.}}{=} \lim_{r\to0^+}\frac{\nu(B_r(x))}{\mathscr{H}^1(B_r(x)\cap\operatorname{supp}\nu)}\right) \tag{3.10}$$

exists for ν-a.e. x (resp. $\mathscr{H}^1\llcorner\operatorname{supp}\nu$-a.e. x).

Proposition 3.5 (Alternative definitions, II). *Let* $\nu\in\mathcal{P}(\mathbb{R}^d)$ *such that* $\operatorname{supp}\nu$ *is connected and* $\mathscr{H}^1(\operatorname{supp}\nu)<+\infty$*. Then*

$$\mathcal{L}(\nu) = \begin{cases}\left\|\frac{\mathrm{d}\left(\mathscr{H}^1\llcorner\operatorname{supp}\nu\right)}{\mathrm{d}\nu}\right\|_{L^\infty_\nu} & \textit{if } (\mathscr{H}^1\llcorner\operatorname{supp}\nu)\ll\nu,\\ +\infty & \textit{otherwise.}\end{cases} \tag{3.11}$$
$$= \begin{cases}0 & \textit{if } \operatorname{supp}\nu \textit{ is a singleton,}\\ \left\|\left(\frac{\mathrm{d}\nu}{\mathrm{d}\left(\mathscr{H}^1\llcorner\operatorname{supp}\nu\right)}\right)^{-1}\right\|_{L^\infty_{\mathscr{H}^1\llcorner\operatorname{supp}\nu}} & \textit{otherwise.}\end{cases} \tag{3.12}$$

Notice that in Proposition 3.5, both “norms” may take the value $+\infty$, and in (3.12), we adopt the convention that $1/0=+\infty$.

Proof of Proposition 3.5. First, we prove (3.11). If $(\mathscr{H}^1\llcorner\operatorname{supp}\nu)\ll\nu$ then the Lebesgue-Besicovitch differentiation theorem ensures that

$$\mathscr{H}^1\llcorner\operatorname{supp}\nu = \left(\frac{\mathrm{d}\left(\mathscr{H}^1\llcorner\operatorname{supp}\nu\right)}{\mathrm{d}\nu}\right)\nu \leq \left\|\frac{\mathrm{d}\left(\mathscr{H}^1\llcorner\operatorname{supp}\nu\right)}{\mathrm{d}\nu}\right\|_{L^\infty_\nu}\nu.$$

Therefore,

$$\mathcal{L}(\nu) \leq \left\|\frac{\mathrm{d}\left(\mathscr{H}^1\llcorner\operatorname{supp}\nu\right)}{\mathrm{d}\nu}\right\|_{L^\infty_\nu} \leq \left\|D^+_\nu(\mathscr{H}^1\llcorner\operatorname{supp}\nu)\right\|_\infty = \mathcal{L}(\nu).$$

If $(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)$ is not absolutely continuous w.r.t. ν, there is no $\alpha > 0$ such that $\alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu$, and $\mathcal{L}(\nu) = +\infty$.

Now, we prove (3.12). The case where $\operatorname{supp}\nu$ is a singleton is already known. We assume now that $\mathscr{H}^1(\operatorname{supp}\nu) > 0$, and using the Besicovitch differentiation theorem [?, Thm. 2.22], we decompose

$$\nu = \theta\mathscr{H}^1 \llcorner \operatorname{supp}\nu + \nu^s, \tag{3.13}$$

where

$$\theta(x) \stackrel{\text{def.}}{=} \frac{\mathrm{d}\nu}{\mathrm{d}\left(\mathscr{H}^1 \llcorner \operatorname{supp}\nu\right)}(x) = \lim_{r\to 0^+} \frac{\nu(B_r(x))}{\mathscr{H}^1(B_r(x)\cap \operatorname{supp}\nu)} = \left(D^+_\nu(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)(x)\right)^{-1}$$

for $(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)$-a.e. x. From the last equality, we get

$$\left\|\theta^{-1}\right\|_{L^\infty_{\mathscr{H}^1 \llcorner \operatorname{supp}\nu}} \leq \left\|D^+_\nu(\mathscr{H}^1 \llcorner \operatorname{supp}\nu)(x)\right\|_\infty = \mathcal{L}(\nu).$$

To prove the converse inequality, we assume $\|\theta^{-1}\|_{L^\infty_{\mathscr{H}^1 \llcorner \operatorname{supp}\nu}} < +\infty$ (otherwise there is nothing to prove). Using (3.13), we note that

$$\left(\left\|\theta^{-1}\right\|_{L^\infty_{\mathscr{H}^1 \llcorner \operatorname{supp}\nu}}\right)\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu,$$

so that $\mathcal{L}(\nu) \leq \|\theta^{-1}\|_{L^\infty_{\mathscr{H}^1 \llcorner \operatorname{supp}\nu}}$. □

We may now examine a few examples.

Example 2.1. *Let $\nu = \sum_{n=1}^\infty 2^{-n}\delta_{q_n}$, where $(q_n)_{n\geq 1}$ is a dense sequence in $[0,1]$. The support being the set of points x such that $\nu(B_r(x)) > 0$ for all $r > 0$, one has $\operatorname{supp}\nu = [0,1]$ which is connected. However, using (3.3), we see that $\mathcal{L}(\nu) = +\infty$.*

Example 2.2 (Densities on a $(\mathscr{H}^1, 1)$-rectifiable set)**.** *Let $\Sigma \subseteq \mathbb{R}^d$ be a closed connected set with $0 < \mathscr{H}^1(\Sigma) < +\infty$, $\theta\colon \Sigma \to \mathbb{R}_+$ a Borel function such that $\displaystyle\int_\Sigma \theta \mathrm{d}\mathscr{H}^1 < 1$, and let $\nu = \theta\mathscr{H}^1 \llcorner \Sigma + \nu^s$ be a probability measure, where $\operatorname{supp}\nu^s \subset \Sigma$ and the measures ν^s and $\mathscr{H}^1 \llcorner \Sigma$ are mutually singular. Then $\mathcal{L}(\nu) = \|1/\theta\|_{L^\infty_{\mathscr{H}^1 \llcorner \Sigma}}$: the length functional ignores the singular part.*

Example 2.3 (Parametrized Lipschitz curves)**.** *Let $\gamma\colon\ [0,1] \to \mathbb{R}^d$ be a non-constant Lipschitz curve, and let ν such that for all $f \in C_b(\mathbb{R}^d)$,*

$$\langle f, \nu\rangle \stackrel{\text{def.}}{=} \frac{1}{\operatorname{len}(\gamma)}\left(\int_0^1 f(\gamma(t))\,|\dot\gamma(t)|\,\mathrm{d}t\right), \quad \textit{where } \operatorname{len}(\gamma) \stackrel{\text{def.}}{=} \int_0^1 |\dot\gamma(t)|\,\mathrm{d}t$$

is the length of the curve. By the area formula [Federer, 2014, Thm. 3.2.5],

$$\mathrm{d}\nu(y) = \frac{1}{\mathrm{len}(\gamma)}\,\mathrm{card}(\gamma^{(-1)}(y))\mathrm{d}\left(\mathscr{H}^1 \llcorner \Sigma\right)(y)$$

where $\Sigma = \gamma([0,1])$. As a result,

$$\mathcal{L}(\nu) = \frac{\mathrm{len}(\gamma)}{\text{ess-}\min_{y\in\Sigma}\left(\mathrm{card}(\gamma^{(-1)}(y))\right)}, \tag{3.14}$$

where the minimum is an *essential minimum* with respect to $\mathscr{H}^1 \llcorner \Sigma$.

2.2. Lower semi-continuity of the length functional

Now, we prove that $\mathcal{L}$ is the lower semi-continuous envelope of ℓ for the narrow convergence.

Proposition 3.6. *The functional $\mathcal{L}$ is the lower semi-continuous envelope of ℓ for the narrow topology. Moreover, for every ν such that $\mathcal{L}(\nu) < +\infty$,*

$$\mathscr{H}^1(\mathrm{supp}\,\nu) \leq \mathcal{L}(\nu) \tag{3.15}$$

with equality if and only if $\nu = \delta_x$ for some $x \in \mathbb{R}^d$, or $\mathscr{H}^1(\mathrm{supp}\,\nu) > 0$ and $\nu = \dfrac{1}{\mathscr{H}^1(\mathrm{supp}\,\nu)}\mathscr{H}^1 \llcorner \mathrm{supp}\,\nu$, i.e. $\nu = \nu_\Sigma$ for some $\Sigma \in \mathcal{A}$, as defined in (3.1).

Proof of Proposition 3.6: The inequality (3.15) is clear from the definition of (3.3), so we study the equality case.

If $\nu = \delta_x$ or $\nu = \dfrac{1}{\mathscr{H}^1(\mathrm{supp}\,\nu)}\mathscr{H}^1 \llcorner \mathrm{supp}\,\nu$ with $\mathscr{H}^1(\mathrm{supp}\,\nu) > 0$, one readily checks that $\mathcal{L}(\nu) = \mathscr{H}^1(\mathrm{supp}\,\nu)$. Conversely, if (3.15) is an equality, for every Borel set B,

$$\begin{aligned} 0 &= \mathcal{L}(\nu) - \mathscr{H}^1(\mathrm{supp}\,\nu) \\ &= \underbrace{\left(\mathcal{L}(\nu)\nu(B) - \mathscr{H}^1(B \cap \mathrm{supp}\,\nu)\right)}_{\geq 0} + \underbrace{\left(\mathcal{L}(\nu)\nu(B^\complement) - \mathscr{H}^1(B^\complement \cap \mathrm{supp}\,\nu)\right)}_{\geq 0} \end{aligned}$$

so that both terms must be zero. If $\mathcal{L}(\nu) > 0$, we deduce

$$\forall B \subset \mathbb{R}^d \text{ Borel}, \quad \nu(B) = \frac{\mathscr{H}^1(B \cap \mathrm{supp}\,\nu)}{\mathcal{L}(\nu)} = \frac{\mathscr{H}^1(B \cap \mathrm{supp}\,\nu)}{\mathscr{H}^1(\mathrm{supp}\,\nu)}.$$

If $\mathcal{L}(\nu) = 0$, $\mathscr{H}^1(\mathrm{supp}\,\nu) = 0$ and since $\mathrm{supp}\,\nu$ is connected, ν is a Dirac mass.

Next we prove that $\mathcal{L}$ is sequentially lower semi-continuous. We consider $(\nu_n)_{n\in\mathbb{N}}$ such that $\nu_n \xrightarrow[n\to\infty]{} \nu \in \mathcal{P}(\mathbb{R}^d)$ and we show that $\alpha \overset{\text{def.}}{=} \liminf_{n\to\infty} \mathcal{L}(\nu_n) \geq \mathcal{L}(\nu)$. If

$\alpha = +\infty$, we have nothing to prove. Otherwise, up to the extraction of a subsequence, we may assume that $\lim_{n\to\infty} \mathcal{L}(\nu_n) = \alpha$ and that $\mathcal{L}(\nu_n) < +\infty$ for all $n \in \mathbb{N}$.

Defining the sequence of compact and connected sets $\Sigma_n \overset{\text{def.}}{=} \operatorname{supp} \nu_n$, it holds that $\mathscr{H}^1(\Sigma_n) \leq \mathcal{L}(\nu_n)$, so that

$$\sup_{n\geq N} \mathscr{H}^1(\Sigma_n) \leq \alpha + 1 < +\infty$$

for N large enough. Hence, for all $n \geq N$, $\operatorname{diam}(\Sigma_n) \leq \alpha + 1$. In addition, let $x \in \operatorname{supp} \nu$. Since $0 < \nu(B_1(x)) \leq \liminf_{n\to\infty} \nu_n(B_1(x))$, for all n large enough $(\operatorname{supp} \nu_n) \cap B_1(x) \neq \emptyset$, thus $\operatorname{supp} \nu_n \subset \overline{B_{\alpha+2}(x)}$.

Therefore, we may apply Blaschke's Theorem and assume, up to extracting a subsequence, that $\Sigma_n \xrightarrow[n\to\infty]{d_H} \Sigma$. From the weak convergence of measures one has $\operatorname{supp} \nu \subset \Sigma$. Let us show that $\operatorname{supp} \nu = \Sigma$. If Σ is a singleton $\{x_0\}$, we have $\nu = \delta_{x_0}$. Otherwise, Theorem 2.10 implies that $\Sigma \in \mathcal{A}$ and furthermore, as $\mathcal{L}(\nu_n)\nu_n \geq \mathscr{H}^1 \llcorner \Sigma_n$, that

$$\alpha\nu \geq \mathscr{H}^1 \llcorner \Sigma. \tag{3.16}$$

Hence, as Σ is connected, for all $z \in \Sigma$ it holds $\nu(B_r(z)) > 0$, confirming that $\operatorname{supp} \nu = \Sigma$. Finally from (3.16) we get that

$$\liminf_{n\to\infty} \mathcal{L}(\nu_n) = \alpha \geq \mathcal{L}(\nu),$$

proving that $\mathcal{L}$ is l.s.c.

As a result, we have proved that $\mathcal{L}$ is l.s.c. and that $\mathcal{L} \equiv \ell$ on the effective domain of ℓ. To show that $\mathcal{L}$ is the l.s.c. enveloppe of ℓ, we prove that it is above any l.s.c. functional $\mathcal{G} \leq \ell$. Let $\nu \in \mathscr{P}(\mathbb{R}^d)$. If $\mathcal{L}(\nu) = +\infty$, we have $\mathcal{G}(\nu) \leq \mathcal{L}(\nu)$. If $\mathcal{L}(\nu) < +\infty$, using Lemma 3.7 below, we can find a sequence $\nu_{\Sigma_n} \xrightarrow[n\to\infty]{} \nu$ such that $\mathscr{H}^1(\Sigma_n) \to \mathcal{L}(\nu)$. The lower semi-continuity of $\mathcal{G}$ yields

$$\mathcal{G}(\nu) \leq \liminf_{n\to\infty} \mathcal{G}(\nu_{\Sigma_n}) \leq \liminf_{n\to\infty} \ell(\nu_{\Sigma_n}) = \liminf_{n\to\infty} \mathscr{H}^1(\Sigma_n) = \mathcal{L}(\nu).$$

□

The proof of Proposition 3.6 relies on the following approximation Lemma.

Lemma 3.7. *Let $\nu \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathcal{L}(\nu) < \infty$. Then, there exists a sequence $(\Sigma_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ such that*

- $\Sigma_n \xrightarrow[n\to\infty]{d_H} \operatorname{supp} \nu$,
- $\nu_{\Sigma_n} \xrightarrow[n\to\infty]{} \nu$ *and* $W_p(\nu_{\Sigma_n}, \nu) \xrightarrow[n\to\infty]{} 0$ *for any* $p \geq 1$, *where* ν_{Σ_n} *is defined as in* (3.1).

We also have $\mathscr{H}^1(\Sigma_n) \xrightarrow[n\to\infty]{} \mathcal{L}(\nu)$ *and if, in addition* $\mathcal{L}(\nu) > 0$, *we can take* $\mathscr{H}^1(\Sigma_n) = \mathcal{L}(\nu)$ *for all* $n \in \mathbb{N}$.

Proof. To simplify the notation, we set $\alpha = \mathcal{L}(\nu)$ and $\Sigma = \operatorname{supp}\nu$. For $\alpha = 0$ (that is, $\nu = \delta_{x_0}$ for some x_0), we consider

$$\Sigma_n = x_0 + [0, 1/n] \times \{0\}^{d-1}$$

which provides the desired approximation, with $\mathscr{H}^1(\Sigma_n) = 1/n \to 0 = \mathcal{L}(\delta_{x_0})$.

For $\alpha > 0$, we start by covering the space with cubes of the form

$$Q_{z,n} \overset{\text{def.}}{=} \frac{1}{n}\left(z + [0,1)^d\right), \quad \text{for } z \in \mathbb{Z}^d.$$

For some fixed n, let $(Q_{i,n})_{i\in I_n}$ be the collection of the cubes such that $\nu\left(Q_{z,n}\right) > 0$, since the set Σ is compact, I_n is finite for a given n. We define the quantities

$$m_{i,n} \overset{\text{def.}}{=} \alpha\nu(Q_{i,n}) - \mathscr{H}^1(\Sigma \cap Q_{i,n}) \leq \alpha,$$

as the excess mass of ν in the cube $Q_{i,n}$ (note that $m_{i,n} \geq 0$ in view of (3.3)). Our strategy is to modify $\nu \llcorner Q_{i,n}$ by adding segments with uniform measure inside the cube and having a total length equal to the excess mass $m_{i,n}$.

If $\Sigma \cap \operatorname{int} Q_{i,n} \neq \emptyset$, take x_i in this intersection, so that $B_{\delta_i}(x_i) \subset Q_{i,n}$ for some $\delta_i > 0$. Then, set $N_{i,n} \overset{\text{def.}}{=} \left\lceil \frac{m_{i,n}}{\delta_i} \right\rceil$, and choose $\delta_{i,j} \geq 0$ for $j = 1, \dots, N_{i,n}$ such that

$$\sum_{j=1}^{N_{i,n}} \delta_{i,j} = m_{i,n}, \text{ and } 0 \leq \delta_{i,j} < \delta_i.$$

Since $\mathscr{H}^1(\Sigma \cap Q_{i,n}) < +\infty$, it is possible to choose $N_{i,n}$ vectors $v_{i,j} \in \mathbb{S}^{d-1}$ such that the segments $S_{i,j} \overset{\text{def.}}{=} [x_i, x_i + \delta_{i,j} v_{i,j}]$ are contained in $\operatorname{int} Q_{i,n}$ and satisfy $\mathscr{H}^1(\Sigma \cap S_{i,j}) = 0$, for $j = 1, \dots, N_{i,n}$.

If $\Sigma \cap \operatorname{int} Q_{i,n} = \emptyset$, as the cubes have positive mass, it means that ν is concentrated on the boundary of the cube, in which case we take $x_i \in \Sigma \cap \partial Q_i$ and any family of segments entering the cube will suffice.

Next, we define the measures

$$\nu_{\Sigma_n} \overset{\text{def.}}{=} \frac{1}{\mathscr{H}^1(\Sigma_n)} \mathscr{H}^1 \llcorner \Sigma_n \text{ for } \Sigma_n \overset{\text{def.}}{=} \Sigma \cup \bigcup_{i\in I_n} \bigcup_{j=1}^{N_{i,n}} S_{i,j}.$$

From the construction, the Hausdorff distance between Σ and Σ_n is at most the diagonal of the cube $[0, 1/n)^d$, so that

$$d_H(\Sigma, \Sigma_n) \leq \frac{\sqrt{d}}{n} \xrightarrow[n\to\infty]{} 0,$$

and the total length of Σ_n is given by

$$\begin{aligned}\mathscr{H}^1(\Sigma_n) &= \sum_{i\in I_n} \mathscr{H}^1(\Sigma\cap Q_{i,n}) + \sum_{i\in I_n}\sum_{j=1}^{N_{i,n}} \mathscr{H}^1(S_{i,j}) \\ &= \sum_{i\in I_n} \mathscr{H}^1(\Sigma\cap Q_{i,n}) + m_{i,n} = \alpha \sum_{i\in I_n} \nu(Q_{i,n}) = \alpha.\end{aligned}$$

Each $\Sigma_n \in \mathcal{A}$ since it is connected and compact (as a finite union of compact sets).

To finish the proof, it remains to show that $\nu_{\Sigma_n} \xrightharpoonup[n\to\infty]{} \nu$. By construction, there exists a compact set $K \subset \mathbb{R}^d$ such that $(\operatorname{supp}\nu) \cup \bigcup_{n\geq 1} (\operatorname{supp}\nu_{\Sigma_n}) \subset K$. Then any function $\phi \in C_b(\mathbb{R}^d)$ is uniformly continuous on K, and we denote by ω its modulus of continuity. Observing that $\nu_{\Sigma_n}(Q_{i,n}) = \nu(Q_{i,n})$, we note that

$$\begin{aligned}\left|\int_{\mathbb{R}^d} \phi \mathrm{d}\nu_{\Sigma_n} - \int_{\mathbb{R}^d} \phi \mathrm{d}\nu\right| &\leq \sum_{i\in I_n} \left|\int_{Q_{i,n}} \phi \mathrm{d}\nu_{\Sigma_n} - \int_{Q_{i,n}} \phi \mathrm{d}\nu\right| \\ &\leq \sum_{i\in I_n} \omega(\mathrm{diam} Q_{i,n})\nu(Q_{i,n}) \leq \omega\left(\sqrt{d}/n\right) \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

Hence $\nu_{\Sigma_n} \xrightharpoonup[n\to\infty]{} \nu$. But as the support of all such measures is contained in the compact K and the Wasserstein distance metrizes the weak convergence in $\mathcal{P}_p(K)$, see [Santambrogio, 2015, Thm. 5.10], it holds that $W_p(\nu_{\Sigma_n}, \nu) \xrightarrow[n\to\infty]{} 0$. $\square$

Remark 3.8. *The conclusions of Proposition 3.6 and Lemma 3.7 still hold when replacing the narrow topology with the local weak-$\star$ topology.*

2.3. A relaxed problem with existence of solutions

The relaxed problem $(\overline{P}_\Lambda)$ introduced on page 81 is defined by replacing ℓ in the orginal problem with its l.s.c. envelope $\mathcal{L}$. We define the energy $\mathcal{E}(\nu) \overset{\text{def.}}{=} W_p^p(\rho_0, \nu) + \Lambda\mathcal{L}(\nu)$, and with a slight abuse of notation, we sometimes write $\mathcal{E}(\Sigma) = \mathcal{E}(\nu_\Sigma)$ for $\Sigma \in \mathcal{A}$. The main point of considering this relaxed problem is that the existence of solutions for $(\overline{P}_\Lambda)$ follows from the direct method of the calculus of variations.

Theorem 3.9. *The relaxed problem $(\overline{P}_\Lambda)$ admits a solution. In addition, $\mathcal{E}$ is the l.s.c. enveloppe of $W_p^p(\rho_0, \cdot) + \Lambda\ell$, and:*

$$\inf (P_\Lambda) = \min (\overline{P}_\Lambda).$$

Proof. Let $(\nu_n)_{n\in\mathbb{N}}$ be a minimizing sequence for $\mathcal{E}$. Since $\left(\sup_n W_p^p(\rho_0, \nu_n)\right) < +\infty$, the moments of order p of ν_n are uniformly bounded (see for instance [Santambrogio, 2015, Thm. 5.11]), and we may then extract a (not relabeled) subsequence converging to some

$\nu \in \mathscr{P}(\mathbb{R}^d)$ in the narrow topology (by Prokhorov's theorem). From Proposition 3.6 and the fact that the Wasserstein distance is lower semi-continuous, the functional $\mathcal{E}$ is l.s.c. and we have that

$$\mathcal{E}(\nu) \leq \liminf_{n\to\infty} \mathcal{E}(\nu_n) = \inf(\overline{P}_\Lambda).$$

The measure ν is a minimizer of $(\overline{P}_\Lambda)$.

To show that $\mathcal{E}$ is the l.s.c. enveloppe of the original energy one may argue as in the proof of Proposition 3.6. Consider any l.s.c. functional $\mathcal{G}$ such that

$$\forall \nu \in \mathscr{P}(\mathbb{R}^d), \quad \mathcal{G}(\nu) \leq W_p^p(\rho_0, \nu) + \Lambda \ell(\nu).$$

For every ν with $\mathcal{L}(\nu) < +\infty$, we use Lemma 3.7 to build a sequence $(\nu_n)_{n\in\mathbb{N}}$ such that $W_p^p(\rho_0, \nu_{\Sigma_n}) \to W_p^p(\rho_0, \nu)$. Indeed, as ν_{Σ_n} converges to ν for the Wasserstein distance, the triangle inequality gives

$$|W_p(\rho_0, \nu_{\Sigma_n}) - W_p(\rho_0, \nu)| \leq W_p(\nu_{\Sigma_n}, \nu) \xrightarrow[n\to\infty]{} 0.$$

Hence for any $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ it holds that

$$\mathcal{G}(\nu) \leq \liminf_{n\to\infty} \left(W_p^p(\rho_0, \nu_{\Sigma_n}) + \Lambda \ell(\nu_{\Sigma_n}) \right) = W_p^p(\rho_0, \nu) + \Lambda \mathcal{L}(\nu) = \mathcal{E}(\nu),$$

and we conclude that $\mathcal{E}$ is the l.s.c. envelope. $\square$

3. On the support of optimal measures

Our goal for this section is to answer the question of "how small" Λ must be in Theorem 3.1. For this, in Theorem 3.10 we study when solutions of the relaxed problem $(\overline{P}_\Lambda)$ are Dirac masses. Keeping this in mind the rest of this section can be skipped and the reader can move on to the major results of this chapter.

The following notation will be useful: a point x_0 is said to be a *p-mean* of ρ_0 if

$$x_0 \in \operatorname*{argmin}_{y\in\mathbb{R}^d} \int_{\mathbb{R}^d} |x-y|^p \mathrm{d}\rho_0(x) = \operatorname*{argmin}_{y\in\mathbb{R}^d} W_p^p(\rho_0, \delta_y).$$

A 2-mean is just the mean of ρ_0, that is, $m_{\rho_0} \overset{\text{def.}}{=} \displaystyle\int_{\mathbb{R}^d} x \mathrm{d}\rho_0(x)$. For $p > 1$, the p-mean is uniquely defined, but for $p = 1$ the collection of 1-means is a closed convex set which is not reduced to a singleton in general.

Theorem 3.10. *For a fixed measure $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$ there exists a critical parameter $\Lambda_\star \in [0, \infty)$ such that*

- *for $\Lambda < \Lambda_\star$ no solution is a Dirac measure;*
- *for $\Lambda > \Lambda_\star$ it holds that $\operatorname{argmin}(\overline{P}_\Lambda)$ is the set of p-means of ρ_0.*

Moreover, $\Lambda_\star = 0$ *if and only if* ρ_0 *is a Dirac mass.*

We start by studying the support of the optimal measure, showing that it is contained in the convex hull of the support of ρ_0. In the sequel the proof of Theorem 3.10 will be divided in several steps. We end the section with an exemple of ρ_0 composed of 2 Dirac masses.

3.1. Elementary properties of the support

Given a set $A \subset \mathbb{R}^d$ we denote by $\overline{\mathrm{conv}}\, A$ its closed convex hull.

Lemma 3.11. *Let* $\nu \in \mathscr{P}(\mathbb{R}^d)$ *be a solution to* ($\overline{P}_\Lambda$). *Then the following properties hold*

(1) $\mathscr{H}^1(\operatorname{supp}\nu) \leq \frac{1}{\Lambda} W_p^p(\rho_0, \delta_{m_{\rho_0}})$, *where* m_{ρ_0} *is any* p*-mean of* ρ_0*. In particular,* Σ *is contained in some ball of diameter* $d_0 \overset{\text{def.}}{=} \frac{1}{\Lambda} W_p^p(\rho_0, \delta_{m_{\rho_0}})$.

(2) $\operatorname{supp}\nu \subset \overline{\mathrm{conv}}\,(\operatorname{supp}\rho_0)$

Proof. For the first point, let Σ denote the support of ν. Since ν has finite energy we have that $\mathcal{L}(\nu) \geq \mathscr{H}^1(\Sigma)$. Thus, since it is also optimal

$$\Lambda \mathscr{H}^1(\Sigma) \leq W_p^p(\rho_0, \nu) + \Lambda \mathcal{L}(\nu) \leq W_p^p(\rho_0, \delta_{m_{\rho_0}}) + \Lambda \mathcal{L}(\delta_{m_{\rho_0}}) = W_p^p(\rho_0, \delta_{m_{\rho_0}}).$$

For the second point, let $C \overset{\text{def.}}{=} \overline{\mathrm{conv}}\,(\operatorname{supp}\rho_0)$. It is a nonempty closed convex set, therefore the projection onto C is well-defined and 1-Lipschitz. We denote it by f. By Proposition 3.4, it holds that $\mathcal{L}(\nu) \geq \mathcal{L}(f_\sharp \nu)$. Moreover, for every $(x, y) \in C \times \mathbb{R}^d$,

$$|x-y|^2 = |x-f(y)|^2 + |f(y)-y|^2 + 2\underbrace{\langle x-f(y),\, f(y)-y\rangle}_{\geq 0} \geq |x-f(y)|^2$$

with equality if and only if $y \in C$. As a result, if γ is an optimal transport plan for (ρ_0, ν),

$$\begin{aligned} W_p^p(\rho_0, \nu) = \int |x-y|^p \mathrm{d}\gamma(x,y) &\geq \int |x-f(y)|^p \mathrm{d}\gamma(x,y) \\ &= \int |x-y|^p \mathrm{d}\left((\mathrm{id}, f)_\sharp \gamma\right)(x,y) \geq W_p^p(\rho_0, f_\sharp \nu) \end{aligned}$$

with strict inequality unless $y \in C$ for γ-a.e. (x, y) (hence ν-a.e. y).

But ν is a solution to ($\overline{P}_\Lambda$), therefore the inequality

$$W_p^p(\rho_0, \nu) + \Lambda \mathcal{L}(\nu) \geq W_p^p(\rho_0, f_\sharp \nu) + \Lambda \mathcal{L}(f_\sharp \nu)$$

cannot be strict. We deduce that $y \in C$ for ν-a.e. y, and C being closed, that $\operatorname{supp}\nu \subset C$. □

Example 3.1. *Let* $\rho_0 = \delta_{x_0}$ *for some* $x_0 \in \mathbb{R}^d$*. From Lemma 3.11 above, we deduce that for all* $\Lambda > 0$, $\operatorname{argmin}\ (\overline{P}_\Lambda) = \{\delta_{x_0}\}$.

3.2. When solutions are Dirac masses

Now, we discuss whether or not Dirac masses may appear in the case where ρ_0 is not a Dirac measure. We start with the following Lemma.

Lemma 3.12. *Let $\Lambda > 0$ such that $\delta_{x_0} \in \operatorname{argmin}\left(\overline{P}_\Lambda\right)$, for $\Lambda' > \Lambda$ it holds*

- *for $p > 1$ that δ_{x_0} is the unique solution of $\left(\overline{P}_{\Lambda'}\right)$,*
- *for $p = 1$ that $\operatorname{argmin}\left(\overline{P}_{\Lambda'}\right)$ consists of only Dirac masses.*

Proof. If $\delta_{x_0} \in \operatorname{argmin}\left(\overline{P}_\Lambda\right)$, for any $p \geq 1$, and for any measure (ν) with $\mathcal{L}(\nu) > 0$ it holds that

$$W_p^p(\rho_0, \delta_{x_0}) \leq W_p^p(\rho_0, \nu) + \Lambda \mathcal{L}(\nu) < W_p^p(\rho_0, \nu) + \Lambda' \mathcal{L}(\nu),$$

and hence ν cannot be a minimizer of $\left(\overline{P}_{\Lambda'}\right)$. Then for any $p \geq 1$ it holds that $\operatorname{argmin}\left(\overline{P}_{\Lambda'}\right)$ consists of Dirac measures. Whenever $p > 1$, the function $y \mapsto W_p^p(\rho_0, \delta_y)$ is strictly convex and hence $\operatorname{argmin}\left(\overline{P}_{\Lambda'}\right)$ is a singleton. $\square$

This simple Lemma allows for the definition of the critical value $\Lambda_\star$ as follows

$$\Lambda_\star \stackrel{\text{def.}}{=} \inf\left\{\Lambda \geq 0 : \operatorname{argmin}\left(\overline{P}_\Lambda\right) \subset (\delta_x)_{x\in\mathbb{R}^d}\right\}. \tag{3.17}$$

As stated in Theorem 3.10, $\Lambda_\star > 0$ whenever ρ_0 is not a single Dirac mass, which is a direct consequence of the convergence of solutions to ρ_0 when Λ goes to 0.

Lemma 3.13. *For every $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$, and $\Lambda > 0$, let ν_Λ be any solution to $(\overline{P}_\Lambda)$. Then*

$$\nu_\Lambda \xrightharpoonup[\Lambda\to 0^+]{} \rho_0. \tag{3.18}$$

Proof. If $\mathcal{L}(\rho_0) < +\infty$, it suffices to notice that

$$W_p^p(\rho_0, \nu_\Lambda) \leq W_p^p(\rho_0, \nu_\Lambda) + \Lambda\mathcal{L}(\nu_\Lambda) \leq W_p^p(\rho_0, \rho_0) + \Lambda\mathcal{L}(\rho_0) = \Lambda\mathcal{L}(\rho_0) \xrightarrow[\Lambda\to 0^+]{} 0.$$

However, we need to handle the case where $\mathcal{L}(\rho_0) = +\infty$.

Let $\varepsilon > 0$. By the density of discrete measures in the Wasserstein space, there exists a probability measure of the form $\mu = \sum_{i=1}^N a_i \delta_{x_i}$ such that $W_p^p(\rho_0, \mu) \leq \varepsilon$. We may assume that $N \geq 2$. By connecting all the points $\{x_i\}_{1\leq i\leq N}$, we obtain a compact connected set Σ with $0 < \mathscr{H}^1(\Sigma) < +\infty$. For every $\theta \in (0,1)$, we then define

$$\tilde{\rho}_0 \stackrel{\text{def.}}{=} \frac{\theta}{\mathscr{H}^1(\Sigma)} \mathscr{H}^1 \llcorner \Sigma + (1-\theta)\mu$$

and we note that $\mathcal{L}(\tilde{\rho}_0) \leq \frac{\mathscr{H}^1(\Sigma)}{\theta} < +\infty$.

Moreover, by the optimality of ν_Λ,

$$W_p^p(\rho_0,\nu_\Lambda) \leq \Lambda\mathcal{L}(\nu_\Lambda) + W_p^p(\rho_0,\nu_\Lambda) \leq \Lambda\mathcal{L}(\tilde{\rho}_0) + W_p^p(\rho_0,\tilde{\rho}_0).$$

Taking the upper limit as $\Lambda \to 0^+$, and using the convexity of the Wasserstein distance yields

$$\limsup_{\Lambda\to 0^+} \left(W_p^p(\rho_0,\nu_\Lambda)\right) \leq W_p^p(\rho_0,\tilde{\rho}_0) \leq \theta W_p^p\left(\rho_0, \frac{\mathscr{H}^1 \llcorner \Sigma}{\mathscr{H}^1(\Sigma)}\right) + (1-\theta)W_p^p(\rho_0,\mu).$$

Letting $\theta \to 0^+$ we obtain $\limsup_{\Lambda\to 0^+}\left(W_p^p(\rho_0,\nu_\Lambda)\right) \leq \varepsilon$ for every $\varepsilon > 0$, which yields $\lim_{\Lambda\to 0^+} W_p^p(\rho_0,\nu_\Lambda) = 0$, hence the claimed result. $\square$

As a consequence of Lemma 3.13, we note that $\liminf_{\Lambda\to 0^+}(\operatorname{supp}\nu_\Lambda) \supset \operatorname{supp}\rho_0$, so that if ρ_0 is not a Dirac mass, neither is ν_Λ for $\Lambda > 0$ small enough.

Next, we show that for Λ large enough, the solution becomes a Dirac measure.

Proposition 3.14. *For every $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$, $\Lambda_\star < +\infty$.*

Proof. Up to a change of the origin, we may assume that $\int_{\mathbb{R}^d} x \mathrm{d}\rho_0(x) = 0$.

We let $\nu \in \operatorname{argmin}\ (\overline{P}_\Lambda)$, $\Sigma \stackrel{\text{def.}}{=} \operatorname{supp}\nu$, and we define $y_0 \in \operatorname{argmin}_{y\in\Sigma}|y|$.

Setting $r \stackrel{\text{def.}}{=} \min\{\, r' \geq 0 \mid \operatorname{supp}\nu \subset B(y_0,r)\,\}$, we note from the connectedness of Σ that $r \leq \mathscr{H}^1(\Sigma) < +\infty$. Moreover, the convexity of the p-norm yields

$$\forall x,y \in \mathbb{R}^d, \quad |x-y|^p \geq |x-y_0|^p - p\,|x-y_0|^{p-1}\,|y-y_0|\,.$$

As a result, if γ is an optimal transport plan for (ρ_0,ν),

$$\begin{aligned}
\mathcal{E}(\nu) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p\,\mathrm{d}\gamma(x,y) + \Lambda\mathcal{L}(\nu)\\
&\geq \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y_0|^p\,\mathrm{d}\gamma(x,y) - p\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y_0|^{p-1}\,|y-y_0|\,\mathrm{d}\gamma(x,y) + \Lambda\mathscr{H}^1(\Sigma)\\
&\geq \mathcal{E}(\delta_{y_0}) + r\left(\Lambda - p\int_{\mathbb{R}^d} |x-y_0|^{p-1}\,\mathrm{d}\rho_0(x)\right).
\end{aligned}$$

By optimality of ν, we have $\mathcal{E}(\nu) \leq \mathcal{E}(\delta_{y_0})$, so that $r = 0$ and ν is a Dirac mass provided that $\left(\Lambda - p\int_{\mathbb{R}^d}|x-y_0|^{p-1}\,\mathrm{d}\rho_0(x)\right) > 0$.

Now, we show that $\int_{\mathbb{R}^d}|x-y_0|^{p-1}\,\mathrm{d}\rho_0(x)$ can be bounded independently from ν. For any optimal ν, since $\mathcal{E}(\nu) \leq \mathcal{E}(\delta_0)$, we note that $W_p^p(\rho_0,\nu) \leq W_p^p(\rho_0,\delta_0)$. Hence

$$|y_0| \leq W_p(\delta_0,\nu) \leq W_p(\delta_0,\rho_0) + W_p(\rho_0,\nu) \leq 2W_p(\delta_0,\rho_0).$$

Setting $R \stackrel{\text{def.}}{=} 2W_p(\delta_0,\rho_0)$, we see that it is sufficient to take

$$\Lambda > \max_{y_0\in B(0,R)}\left(p\int_{\mathbb{R}^d}|x-y_0|^{p-1}\,\mathrm{d}\rho_0(x)\right),$$

to ensure that ν is a Dirac mass. $\square$

Remark 3.15. *In some cases, it is possible to provide sharper bounds on $\Lambda_\star$:*

- *If $p = 1$, we see that $\Lambda_\star \le 1$.*
- *If $p = 2$, it can be shown by a simple translation argument that ν and ρ_0 have the same barycenter. Then, one may adapt the above argument to get $\Lambda_\star \le 2 \displaystyle\int |x - x_0| \,\mathrm{d}\rho_0(x)$, where $x_0 = \displaystyle\int x \mathrm{d}\rho_0(x) = 0$.*
- *If $\operatorname{supp} \rho_0$ is bounded, it is possible to obtain $\Lambda_\star \le p(\operatorname{diam}(\operatorname{supp} \rho_0))^{p-1}$ for any $p \ge 1$, by exploiting the Lipschitzianity of the dual potentials: there exists (ϕ, ψ), solution to the dual Kantorovitch problem (see [Santambrogio, 2015, Sec. 1.2])*

  $$W_p^p(\mu, \nu) = \max \left\{ \int \phi \mathrm{d}\mu + \int \psi \mathrm{d}\nu : \begin{array}{c} \phi \in L^1(\mu), \psi \in L^1(\nu), \\ \phi(x) + \psi(y) \le |x-y|^p \end{array} \right\},$$

  such that $\operatorname{Lip}(\psi) \le p(\operatorname{diam}(\operatorname{supp} \rho_0))^{p-1}$. Then,

  $$\begin{aligned} W_p^p(\rho_0, \delta_{y_0}) - W_p^p(\rho_0, \nu) &\le \psi(y_0) - \int_\Sigma \psi \mathrm{d}\nu \le \int_\Sigma |\psi(y_0) - \psi(x)| \,\mathrm{d}\nu(x) \\ &\le \operatorname{Lip}(\psi) \cdot \mathscr{H}^1(\Sigma) \le \Lambda \mathcal{L}(\nu) \end{aligned}$$

  and for $\Lambda > \operatorname{Lip}(\psi)$, the last inequality is strict, yielding the contradiction $\mathcal{E}(\delta_{y_0}) < \mathcal{E}(\nu)$, unless $\mathscr{H}^1(\Sigma) = 0$.

The example of an input with two Dirac masses

In this subsection we consider the case $p = 2$. Let $x_{-1} = (-1, 0, \ldots, 0)$, $x_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$, and let $\rho_0 = \frac{1}{2}\left(\delta_{x_{-1}} + \delta_{x_1}\right)$. By Lemma 3.11, we know that the solutions to $(\overline{P}_\Lambda)$ are supported on line segments which are contained in $[x_{-1}, x_1]$. We may thus reduce the problem to the one-dimensional setting, with $x_{-1} = -1$, $x_1 = 1$. The solution to that problem is given by the following proposition.

Proposition 3.16. *For $p = 2$ and $\rho_0 = \frac{1}{2}(\delta_{-1} + \delta_1)$, the unique solution to $(\overline{P}_\Lambda)$ is given by*

$$\nu_\Lambda = \begin{cases} \sqrt{\dfrac{3\Lambda}{2}} \mathscr{H}^1 \llcorner [-1, 1] + \left(\dfrac{1}{2} - \sqrt{\dfrac{3\Lambda}{2}} \right) (\delta_{-1} + \delta_1) & \text{if } 0 < \Lambda < \frac{1}{6}, \\ \dfrac{1}{3(1-2\Lambda)} \mathscr{H}^1 \llcorner \left[-\dfrac{3}{2}(1-2\Lambda), \dfrac{3}{2}(1-2\Lambda) \right] & \text{if } \frac{1}{6} \le \Lambda < \frac{1}{2} \\ \delta_0 & \text{if } \Lambda \ge \frac{1}{2}. \end{cases} \tag{3.19}$$

Proof. We fix $\Lambda > 0$ and denote ν a solution. Let $\alpha = \mathcal{L}(\nu)$. If $\alpha = 0$, ν is a Dirac mass. If $\alpha > 0$, we know that the support of ν is a connected subset of $\overline{\mathrm{conv}}\{-1,1\} = [-1,1]$, so that $\operatorname{supp}\nu = [a,b]$ for $-1 \leq a < b \leq 1$. In addition, letting $c \in [a,b]$ such that $\nu([a,c[) \leq 1/2$ and $\nu([a,c]) \geq 1/2$, one can check that if some mass is sent from $\{-1\}$ to $]c,b]$, then exchanging it with the same amount of mass sent from $\{+1\}$ to $[a,c[$ we reduce the Wasserstein distance. Hence one may assume that the mass coming from $\{-1\}$ is sent to a measure ν^- supported on $[a,c]$ while the mass from $\{+1\}$ is sent to a measure ν^+ supported on $[c,b]$, with $\nu^- + \nu^+ = \nu$. Observing that $\nu \geq \frac{1}{\alpha}\mathscr{H}^1 \llcorner [a,b]$ (we are in the case $\alpha > 0$), we introduce the non-negative excess measures:

$$\nu^-_{\text{exc}} = \nu^- - \frac{1}{\alpha}\mathscr{H}^1 \llcorner [a,c], \quad \nu^+_{\text{exc}} = \nu^+ - \frac{1}{\alpha}\mathscr{H}^1 \llcorner [c,b],$$

and $\nu_{\text{exc}} = \nu^-_{\text{exc}} + \nu^+_{\text{exc}}$. Once more, we see that the Wasserstein distance is reduced if all the mass sent from $\{-1\}$ to ν^-_{exc} is sent to the point $\{a\}$, closest to $\{-1\}$. Hence, we may assume that $\nu^-_{\text{exc}} = x\delta_a$, for $x \geq 0$, and similarly, $\nu^+_{\text{exc}} = y\delta_b$, for $y \geq 0$. Eventually, we easily see that if $a > -1$ and $x > 0$, then we can extend the segment $[a,b]$ towards $\{-1\}$, adding a small piece $[a-\delta,\delta]$ for $\delta \leq \min\{\alpha x, a+1\}$, send a fraction δ/α of the measure $x\delta_a$ rather to $\frac{1}{\alpha}\mathscr{H}^1 \llcorner [a-\delta,a]$, and reduce again the Wasserstein distance without changing $\mathcal{L}(\nu)$. We deduce that $x = 0$ if $a > -1$, similarly $y = 0$ if $b < 1$.

Since the solutions are supported on a line segment in $[-1,1]$, they are of the form $\nu = \delta_a$ or $\nu = \frac{1}{\alpha}\mathscr{H}^1 \llcorner [a,b] + \nu_{\text{exc}}$, with $\alpha = \mathcal{L}(\nu)$ and $\operatorname{supp}\nu_{\text{exc}} \subset [a,b] \subset [-1,1]$. Recalling that for $p = 2$, ν must have the same center of mass as ρ_0, we deduce that ν must be equal to

$$\begin{aligned}
&\nu_{0,0} \overset{\text{def.}}{=} \delta_0, \\
\text{or}\quad &\nu_{b,2b} \overset{\text{def.}}{=} \frac{1}{2b}\mathscr{H}^1 \llcorner [-b,b] \quad \text{for some } b \in (0,1) \\
\text{or}\quad &\nu_{1,\alpha} = \frac{1}{\alpha}\mathscr{H}^1 \llcorner [-1,1] + \left(\frac{1}{2} - \frac{1}{\alpha}\right)(\delta_{-1} + \delta_1) \quad \text{for some } \alpha \geq 2.
\end{aligned}$$

Let $\mathcal{E}(\nu) = \Lambda\mathcal{L}(\nu) + W_2^2(\rho_0,\nu)$ denote the energy to minimize. We have $\mathcal{E}(\nu_{0,0}) = 1 = \lim_{b\to 0^+}\mathcal{E}(\nu_{b,2b})$, and

$$\begin{aligned}
\mathcal{E}(\nu_{b,2b}) &= 2\Lambda b + 2\int_0^b (1-x)^2 \frac{\mathrm{d}x}{2b} = \frac{b^2}{3} + (2\Lambda - 1)b + 1 \\
\text{with}\quad \frac{\mathrm{d}}{\mathrm{d}b}\mathcal{E}(\nu_{b,2b}) &= \frac{2b}{3} + 2\Lambda - 1, \\
\mathcal{E}(\nu_{1,\alpha}) &= \Lambda\alpha + 2\int_0^1 (1-x)^2 \frac{\mathrm{d}x}{\alpha} + 0 = \Lambda\alpha + \frac{2}{3\alpha}, \\
\text{with}\quad \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathcal{E}(\nu_{1,\alpha}) &= \Lambda - \frac{2}{3\alpha^2}.
\end{aligned}$$

For $0 < \Lambda < \frac{1}{6}$, we check that ν_{1,α^*}, for $\alpha^* \overset{\text{def.}}{=} \sqrt{\frac{2}{3\Lambda}}$, is the unique solution.

For $\frac{1}{6} \leq \Lambda < \frac{1}{2}$, we get that $\nu_{b^*,2b^*}$ is the unique solution, with $b^* \overset{\text{def.}}{=} \frac{3}{2}(1-2\Lambda)$.

For $\Lambda \geq \frac{1}{2}$, the functions $\alpha \mapsto \mathcal{E}(\nu_{1,\alpha})$ and $b \mapsto \mathcal{E}(\nu_{b,2b})$ are strictly decreasing on $[2,+\infty[$ and $]0,1]$ respectively. Therefore $\nu_{0,0}$ is the unique solution to ($\overline{P}_\Lambda$). □

4. Solutions are rectifiable measures

Our goal here is to show that whenever $\rho_0 \ll \mathscr{H}^1$, any solution ν is a rectifiable measure of the form

$$\nu = \theta \mathscr{H}^1 \llcorner \Sigma, \quad \text{for } \theta \in L^1(\Sigma; \mathscr{H}^1)$$

To this end we introduce the excess measure ν_{exc} as the positive measure given by the mass of ν that exceeds the density constraints. We first show that this measure solves a family of localized problems. This is used to prove the absolute continuity w.r.t. $\mathscr{H}^1 \llcorner \Sigma$, that is, point (1) of Theorem 3.1.

4.1. The excess measure

Let ν be a minimizer of ($\overline{P}_\Lambda$) with support Σ not reduced to a singleton. From the definition of the length functional we have:

$$\mathcal{L}(\nu) < \infty \text{ if and only if there is } \alpha \geq 0 \text{ such that } \alpha\nu \geq \mathscr{H}^1 \llcorner \Sigma.$$

Setting $\alpha \overset{\text{def.}}{=} \mathcal{L}(\nu) > 0$, we define the following decomposition

$$\nu = \nu_{\mathscr{H}^1} + \nu_{\text{exc}}, \text{ where } \nu_{\mathscr{H}^1} \overset{\text{def.}}{=} \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma \text{ and } \nu_{\text{exc}} \overset{\text{def.}}{=} \nu - \nu_{\mathscr{H}^1}. \tag{3.20}$$

The part $\nu_{\mathscr{H}^1}$ is the measure which saturates the density constraint, and the support of the *excess measure* ν_{exc} is where the constraint is inactive.

We define an analogous (nonunique) decomposition of γ and ρ_0 by disintegrating γ w.r.t. the second marginal. From the disintegration theorem [Ambrosio et al., 2000, Theorem 2.28], there exists a ν-measurable family $\{\gamma_y\}_{y\in\mathbb{R}^d} \subset \mathscr{P}(\mathbb{R}^d)$, such that $\gamma = \gamma_y \otimes \nu$, that is

$$\int_{\mathbb{R}^d\times\Sigma} \psi(x,y)\mathrm{d}\gamma(x,y) = \int_\Sigma \left(\int_{\mathbb{R}^d} \psi(x,y)\mathrm{d}\gamma_y(x)\right)\mathrm{d}\nu(y), \text{ for all } \phi \in L^1(\gamma). \tag{3.21}$$

We define a decomposition $\gamma = \gamma_{\mathscr{H}^1} + \gamma_{\text{exc}}$ as

$$\gamma_{\mathscr{H}^1}(A\times B) \overset{\text{def.}}{=} \int_{\Sigma\cap B} \gamma_y(A)\mathrm{d}\nu_{\mathscr{H}^1}(y), \quad \gamma_{\text{exc}}(A\times B) \overset{\text{def.}}{=} \int_{\Sigma\cap B} \gamma_y(A)\mathrm{d}\nu_{\text{exc}}(y). \tag{3.22}$$

The decomposition $\rho_0 = \rho_{\mathscr{H}^1} + \rho_{\text{exc}}$ can be defined as the marginals of $\gamma_{\mathscr{H}^1}$ and γ_{exc}

$$\rho_{\mathscr{H}^1} \overset{\text{def.}}{=} \pi_{0\sharp}\gamma_{\mathscr{H}^1}, \quad \rho_{\text{exc}} \overset{\text{def.}}{=} \pi_{0\sharp}\gamma_{\text{exc}}. \tag{3.23}$$

This way $\gamma_{\mathscr{H}^1} \in \Pi(\rho_{\mathscr{H}^1}, \nu_{\mathscr{H}^1})$, $\gamma_{\mathscr{H}^1} \in \Pi(\rho_{\text{exc}}, \nu_{\text{exc}})$ and they are optimal transportation plans between their respective marginals. Indeed if we find a better transportation plan for either problem we can construct a better plan for the original problem, contradicting the minimality of γ. We therefore also have a decomposition between the Wasserstein distances

$$W_p^p(\rho_0, \nu) = W_p^p(\rho_{\mathscr{H}^1}, \nu_{\mathscr{H}^1}) + W_p^p(\rho_{\text{exc}}, \nu_{\text{exc}}). \tag{3.24}$$

It is important to point out that, although the decomposition of ν is natural, there are many ways to decompose γ and ρ_0. In the sequel we show that for any such decomposition the excess must be concentrated on the graph of the operator given by the (multivalued) projection onto Σ

$$\Pi_\Sigma(x) \stackrel{\text{def.}}{=} \operatorname*{argmin}_{y\in\Sigma} |x-y|^2. \tag{3.25}$$

Note that Π_Σ is a multivalued operator which is included in the subgradient of the convex conjugate of the function: $y \mapsto |y|^2/2$ if $y \in \Sigma$ and $+\infty$ else.

Lemma 3.17. *Let ν be a minimizer of ($\overline{P}_\Lambda$) and γ an optimal transport plan from ρ_0 to ν. Then, for any decomposition $\gamma = \gamma_{\mathscr{H}^1} + \gamma_{exc}$, s.t. $\pi_{1\sharp}\gamma_1 = \nu_{\mathscr{H}^1}$, it holds that*

$$\operatorname{supp}\gamma_{exc} \subset \operatorname{graph}(\Pi_\Sigma). \tag{3.26}$$

In addition, for any π_Σ measurable selection of $x \mapsto \Pi_\Sigma(x)$, the measure

$$\nu_{\mathscr{H}^1} + \pi_{\Sigma\sharp}\rho_{exc}$$

is optimal for ($\overline{P}_\Lambda$).

Proof. Consider the problem

$$\inf_{\substack{\gamma\in\mathscr{P}_p(\mathbb{R}^d\times\mathbb{R}^d)\\ \pi_{0\sharp}\gamma=\rho_0,}} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \,\mathrm{d}\gamma(x,y) + \Lambda\mathcal{L}(\pi_{0\sharp}\gamma), \tag{$\overline{Q}_\Lambda$}$$

which is a reformulation of ($\overline{P}_\Lambda$) in terms of the transport plan γ from ρ_0 to ν.

Now, let $(\gamma_{\mathscr{H}^1}, \gamma_{\text{exc}})$ be any suitable decomposition of γ and let π_Σ be a measurable selection of Π_Σ. We set $\rho_{\text{exc}} \stackrel{\text{def.}}{=} \pi_{0\sharp}\gamma_{\text{exc}}$ and define $\tilde\gamma = \gamma_{\mathscr{H}^1} + (\mathrm{id}, \pi_\Sigma)_\sharp\rho_{\text{exc}}$. Then it holds that $\mathcal{L}(\pi_{1\sharp}\tilde\gamma) \le \mathcal{L}(\nu)$ and

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \,\mathrm{d}\tilde\gamma &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \,\mathrm{d}\gamma_{\mathscr{H}^1} + \int_{\mathbb{R}^d} |x-\pi_\Sigma(x)|^p \,\mathrm{d}\rho_{\text{exc}}\\ &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \,\mathrm{d}\gamma_{\mathscr{H}^1} + \int_{\mathbb{R}^d\times\Sigma} |x-y|^p \,\mathrm{d}\gamma_{\text{exc}} = \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \,\mathrm{d}\gamma\end{aligned}$$

Since γ is a minimizer of ($\overline{Q}_\Lambda$), both inequalities must be equalities, in particular we must have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \left(|x-y|^p - |x-\pi_\Sigma(x)|^p\right) \mathrm{d}\gamma_{\text{exc}} = 0.$$

Since γ-a.e. (x,y) is in $\mathbb{R}^d \times \Sigma$, the integrand is non-negative and must vanish γ_{exc}-a.e. Hence $(x,y) \in \operatorname{Graph}(\Pi_\Sigma)$ for γ_{exc}-a.e. (x,y) and (3.26) follows since $\operatorname{Graph}(\Pi_\Sigma)$ is closed. As a consequence, the measure $\nu_{\mathscr{H}^1} + \pi_{\Sigma\sharp}\rho_{\text{exc}}$ reaches the minimimum for $(\overline{P}_\Lambda)$ and is optimal. ☐

4.2. Solutions are absolutely continuous

Now we prove that the solutions to the relaxed problem $(\overline{P}_\Lambda)$ are absolutely continuous w.r.t. $\mathscr{H}^1 \llcorner \Sigma$. The proof is based on the construction of a localized variational problem.

Lemma 3.18. *Let ν be an optimal solution for the relaxed problem $(\overline{P}_\Lambda)$ and set $\alpha = \mathcal{L}(\nu)$. Let $\mathcal{S} = \mathcal{S}_0 \times \mathcal{S}_1 \subset \mathbb{R}^d \times \mathbb{R}^d$ be a Borel set and define the transportation plan*

$$\gamma_{\mathcal{S}} \overset{\text{def.}}{=} \gamma_{exc} \llcorner \mathcal{S}_0 \times \mathcal{S}_1$$

along with its marginals

$$\rho_{\mathcal{S}} \overset{\text{def.}}{=} \pi_{0\sharp}\gamma_{\mathcal{S}} = \rho_{exc} \llcorner \mathcal{S}_0, \quad \nu_{\mathcal{S}} \overset{\text{def.}}{=} \pi_{1\sharp}\gamma_{\mathcal{S}}.$$

Then the measure $\nu_{\mathcal{S}}$ solves the following variational problem

$$\inf \left\{ W_p^p(\rho_{\mathcal{S}}, \nu') : \begin{array}{c} \text{there is } \Gamma \text{ such that} \\ \nu' \in \mathcal{M}_+(\Sigma \cup \Gamma), \\ \nu' \geq \alpha^{-1}\mathscr{H}^1 \llcorner \Gamma \setminus \Sigma, \\ \Sigma \cup \Gamma \in \mathcal{A},\ \nu'(\mathbb{R}^d) = \nu_{\mathcal{S}}(\mathbb{R}^d) \end{array} \right\} \tag{3.27}$$

More generally, let $(\sigma_{\mathcal{S},t})_{t\in[0,1]}$ be the constant speed geodesic between $\rho_{\mathcal{S}}$ and $\nu_{\mathcal{S}}$ defined through $\sigma_{\mathcal{S},t} \overset{\text{def.}}{=} \pi_{(1-t)\sharp}\gamma_{\mathcal{S}}$, where $\pi_t(x,y) \overset{\text{def.}}{=} (1-t)x + ty$. Then for any $t \in [0,1]$, the measure $\nu_{\mathcal{S}}$ minimizes the variational problem

$$\inf \left\{ W_p^p(\sigma_{\mathcal{S},t}, \nu') : \begin{array}{c} \text{there is } \Gamma \text{ such that} \\ \nu' \in \mathcal{M}_+(\Sigma \cup \Gamma), \\ \nu' \geq \alpha^{-1}\mathscr{H}^1 \llcorner \Gamma \setminus \Sigma, \\ \Sigma \cup \Gamma \in \mathcal{A},\ \nu'(\mathbb{R}^d) = \nu_{\mathcal{S}}(\mathbb{R}^d) \end{array} \right\}. \tag{3.28}$$

Proof. First, we fix some arbitrary Γ such that $\Sigma \cup \Gamma \in \mathcal{A}$. We consider measures $\nu' \in \mathcal{M}_+(\Sigma \cup \Gamma)$ such that $\nu'(\mathbb{R}^d) = \nu_{\mathcal{S}}(\mathbb{R}^d)$ and $\nu' \geq \alpha^{-1}\mathscr{H}^1 \llcorner \Gamma$, and we build competitors to ν of the form $\nu - \nu_{\mathcal{S}} + \nu'$. Such measures are supported over $\Sigma \cup \Gamma \in \mathcal{A}$ and

$$\begin{aligned} \nu - \nu_{\mathcal{S}} + \nu' &= \nu_{\mathscr{H}^1} + (\nu_{\text{exc}} - \nu_{\mathcal{S}}) + \nu' \\ &\geq \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma + \alpha^{-1}\mathscr{H}^1 \llcorner \Gamma \geq \alpha^{-1}\mathscr{H}^1 \llcorner (\Sigma \cup \Gamma), \end{aligned}$$

so that $\mathcal{L}(\nu - \nu_{\mathcal{S}} + \nu') \leq \alpha = \mathcal{L}(\nu)$. By optimality of ν, we deduce that

$$W_p^p(\rho_0, \nu) \leq W_p^p(\rho_0, \nu - \nu_{\mathcal{S}} + \nu').$$

Now, as the support of γ is c-cyclically monotone (see [Ambrosio et al., 2021, Def. 3.10 and Thm. 3.17]), so is the support of $\gamma_{\mathcal{S}}$, making it an optimal transportation plan between its marginals (see [Ambrosio et al., 2021, Thm. 4.2]). Since the same argument applies to $\gamma - \gamma_{\mathcal{S}}$, we get

$$W_p^p(\rho_0, \nu) = W_p^p\left(\rho_0 - \rho_{\mathcal{S}}, \nu - \nu_{\mathcal{S}}\right) + W_p^p\left(\rho_{\mathcal{S}}, \nu_{\mathcal{S}}\right). \tag{3.29}$$

Besides, let γ' be an optimal transportation plan from $\rho_{\mathcal{S}}$ to ν'. Then $(\gamma - \gamma_{\text{exc}}) + \gamma'$ is a transportation plan from ρ_0 to $(\nu - \nu_{\mathcal{S}} + \nu')$, hence

$$W_p^p(\rho_0, \nu) \leq \int |x-y|^p \, \mathrm{d}\gamma_{\mathcal{S}} + \int |x-y|^p \, \mathrm{d}\gamma' = W_p^p\left(\rho_0 - \rho_{\mathcal{S}}, \nu - \nu_{\mathcal{S}}\right) + W_p^p\left(\rho_{\mathcal{S}}, \nu'\right).$$

Substracting (3.29), we deduce that $W_p^p\left(\rho_{\mathcal{S}}, \nu_{\mathcal{S}}\right) \leq W_p^p\left(\rho_{\mathcal{S}}, \nu'\right)$ for all the admissible variations ν' of the excess measure.

As $\gamma_{\mathcal{S}}$ is an optimal transportation plan between $\rho_{\mathcal{S}}$ and $\nu_{\mathcal{S}}$, from [Santambrogio, 2015, Theorem 5.27] one can define a constant speed geodesic between such measures as

$$\sigma_{\mathcal{S},t} \stackrel{\text{def.}}{=} \pi_{(1-t)_\sharp}\gamma_{\mathcal{S}}, \text{ where } \pi_t(x,y) \stackrel{\text{def.}}{=} (1-t)x + ty.$$

Hence for any variation ν', admissible in the sense of the previous problem, and for any $t \in [0,1]$, it holds that

$$\begin{aligned} W_p\left(\rho_{\mathcal{S}}, \sigma_{\mathcal{S},t}\right) + W_p\left(\sigma_{\mathcal{S},t}, \nu_{\mathcal{S}}\right) = W_p\left(\rho_{\mathcal{S}}, \nu_{\mathcal{S}}\right) &\leq W_p\left(\rho_{\mathcal{S}}, \nu'\right) \\ &\leq W_p\left(\rho_{\mathcal{S}}, \sigma_{\mathcal{S},t}\right) + W_p\left(\sigma_{\mathcal{S},t}, \nu'\right). \end{aligned}$$

Where the equality comes from general properties of constant speed geodesics in metric spaces, while the inequalities come from the minimality of $\nu_{\mathcal{S}}$ and the triangle inequality, respectively. We conclude that in fact, the measures $\nu_{\mathcal{S}}$ minimize the Wasserstein distance to the family of geodesic interpolations $\sigma_{\mathcal{S},t}$. □

We now craft a specific set $\mathcal{S}$ to apply the lemma. Given $\delta > 0$, we define the set

$$D_\delta \stackrel{\text{def.}}{=} \left\{ x \in \operatorname{supp} \rho_{\text{exc}} : \ \delta \leq \operatorname{dist}(x, \Sigma) \leq \delta^{-1} \ \right\}, \tag{3.30}$$

And for a fixed point $y_0 \in \Sigma$, and $\delta, r > 0$ consider the new transportation plan

$$\gamma_{\delta,r} \stackrel{\text{def.}}{=} \gamma_{\text{exc}} \llcorner D_\delta \times B_r(y_0) \tag{3.31}$$

along with its marginals

$$\rho_{\delta,r} \stackrel{\text{def.}}{=} \pi_{0\sharp}\gamma_{\delta,r} \leq \rho_{\text{exc}} \llcorner D_\delta, \quad \nu_{\delta,r} \stackrel{\text{def.}}{=} \pi_{1\sharp}\gamma_{\delta,r}. \tag{3.32}$$

From Lemma 3.18 it holds that

$$\nu_{\delta,r} \in \operatorname{argmin}\left\{ W_p^p\left(\rho_{\delta,r},\nu'\right) : \begin{array}{c} \text{there is } \Gamma \text{ such that,} \\ \nu' \in \mathcal{M}_+(\Sigma\cup\Gamma), \\ \nu' \geq \alpha^{-1}\mathscr{H}^1 \llcorner \Gamma\setminus\Sigma, \\ \Sigma\cup\Gamma\in\mathcal{A},\ \nu'(\mathbb{R}^d)=\nu_{\delta,r}(\mathbb{R}^d) \end{array} \right\}. \tag{3.33}$$

We also introduce

$$\gamma_\delta \overset{\text{def.}}{=} \gamma_{\text{exc}} \llcorner D_\delta \times \Sigma \text{ and } \nu_\delta \overset{\text{def.}}{=} \pi_{1\sharp}\gamma_\delta, \tag{3.34}$$

so that by definition, $\nu_{\delta,r} = \nu_\delta \llcorner B_r(y_0)$ and ν_{exc} can be further decomposed as $\nu_{\text{exc}} = \nu_\delta + \pi_{1\sharp}\left(\gamma_{\text{exc}} \llcorner D_\delta^c \times \mathbb{R}^d\right)$. As D_δ is a nested sequence of sets, $(\nu_\delta)_{\delta>0}$ is a monotone sequence and taking the limit as $\delta\to 0$ we have

$$\nu_{\text{exc}} = \sup_{\delta>0} \nu_\delta + \rho_{\text{exc}} \llcorner \Sigma, \tag{3.35}$$

the second limit being $\rho_{\text{exc}} \llcorner \Sigma$ because of Lemma 3.17 and since the only projection of a point in Σ is itself.

In the next Theorem 3.20 we show that the measures ν_δ have a uniform L^∞ bounded density w.r.t. $\mathscr{H}^1$. So when $\rho_0 \ll \mathscr{H}^1$, (3.35) shows that any optimal $\nu \ll \mathscr{H}^1$. The argument consists in crafting a competitor for the localized problem (3.33), built as a measure supported on a curve with controlled length, defined over small sphere, centered at an arbitrary point of the support of ν_δ. Letting the radius of this sphere go to zero, and comparing the energy of this competitor and the optimal measure, gives a uniform bound on the density. This strategy is illustrated in Figure 3.

Lemma 3.19. *Let B_2 be the ball on $\mathbb{R}^d$ centered at the origin. There exists a connected set $\Gamma_d \subset \partial B_2$ with $\mathscr{H}^1(\Gamma_d) < +\infty$ and such that*

$$\operatorname{dist}(x,\Gamma_d) \leq |x-y| - \frac{1}{2}$$

for any $x \notin B_2$ and for all $y \in B_1$.

Proof. We start by covering the sphere ∂B_2 with finitely many balls $\left(B_{1/2}(x_i)\right)_{i=1}^{N_d}$, each having radius $1/2$. The number of balls N_d being dependent on the dimension. In the sequel we define Γ_d with geodesics on ∂B_2 connecting the centers $(x_i)_{i=1}^{N_d}$.

As we have finitely many points, we will also have finitely many curves and hence $\mathscr{H}^1(\Gamma_d)$ must also be finite. We can even choose the connected set Γ_d with minimal length, which is a solution to Steiner's problem on the spheres and has a tree structure, so that we can bound $\mathscr{H}^1(\Gamma_d) \leq (N_d-1)D_d$, where D_d is the diameter of ∂B_2 in its Riemannian metric.

To prove the desired property, take $x \notin B_2$ and $y \in B_1$. Let $\{\hat{y}\} = [x,y]\cap\partial B_2$. Then $\hat{y} \in B_{1/2}(x_i)$ for some x_i while $|x-\hat{y}| \leq |x-y| - 1$, and it follows that

$$\operatorname{dist}(x,\Gamma_d) \leq |x-x_i| \leq |x-\hat{y}| + |\hat{y}-x_i| \leq |x-y| - \frac{1}{2}.$$

□

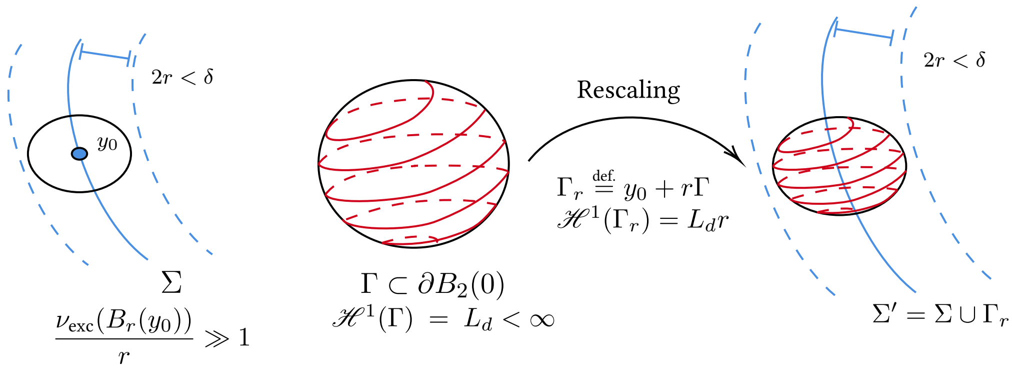

Figure 3: Scheme of the proof of Thm. 3.20. For the new competitor, created with the curve Γ from Lemma 3.19, we pay a little more in the transportation cost to generate $\alpha^{-1}\mathscr{H}^1 \llcorner \Gamma_r$, but pay much less by projecting the remaining mass onto it.

Theorem 3.20. *Given $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$, let ν be a solution to $(\overline{P}_\Lambda)$. Then it holds that the measures $(\nu_\delta)_{\delta>0}$ are of the form*

$$\nu_\delta = \theta_\delta \mathscr{H}^1 \llcorner \Sigma, \ \text{with } \|\theta_\delta\|_{L^\infty(\Sigma,\mathscr{H}^1)} \leq \frac{7}{2}\frac{C_d}{\mathcal{L}(\nu)},$$

for $C_d = 2 + \mathscr{H}^1(\Gamma_d)$, Γ_d being the set from Lemma 3.19.

Therefore, if $\rho_0 \ll \mathscr{H}^1$ or has a L^∞ density w.r.t. $\mathscr{H}^1$, so does ν, in particular it is a rectifiable measure.

Proof. For $y_0 \in \Sigma$, let us define the one-dimensional upper density [Ambrosio et al., 2000, Def. 2.55]

$$\theta_\delta(y_0) \overset{\text{def.}}{=} \limsup_{r\to 0} \frac{\nu_\delta(B_r)}{2r}.$$

We will show that $\theta_\delta(y_0) \leq \frac{9}{2}\frac{C_d}{\mathcal{L}(\nu)}$, so that thanks to [Ambrosio et al., 2000, Thm. 2.56], $\nu_\delta \ll \mathscr{H}^1 \llcorner \Sigma$. Since Σ is 1-rectifiable, it follows that for $\mathscr{H}^1$-a.e. $y_0 \in \Gamma$, $\theta_\delta(y_0)$ is the Radon-Nikodým derivative of ν_δ w.r.t. $\mathscr{H}^1 \llcorner \Sigma$, and the claim of the theorem follows.

From the optimality of ν, the measure $\nu_{\delta,r}$ solves problem (3.33). In order to build a competitor we consider the set Γ_d from Lemma 3.19, choose some point $\bar{y} \in \Gamma_d$ and define

$$\Gamma_r \overset{\text{def.}}{=} [y_0, y_0 + r\bar{y}] \cup (y_0 + r\Gamma_d)\,,$$

which is contained in $\overline{B}_{2r}(y_0)$. Notice that $\Sigma \cup \Gamma_r$ is always a compact, connected and 1-rectifiable set and one has

$$\mathscr{H}^1(\Gamma_r) = C_d r,$$

where $C_d = 2 + \mathscr{H}^1(\Gamma_d)$ is a constant depending only on the dimension.

In the sequel, setting $\alpha = \mathcal{L}(\nu)$ we define the following parameter

$$m_r \overset{\text{def.}}{=} \frac{\mathscr{H}^1(\Gamma_r)}{\alpha\nu_\delta(B_r)}.$$

Suppose that $C_d/\alpha < 2\theta_\delta(y_0)$. Then,

$$1 > m_0 \overset{\text{def.}}{=} \frac{C_d}{2\alpha\theta_\delta(y_0)} = \liminf_{r\to 0} m_r.$$

Now, we consider a subsequence $(r_k)_{k\in\mathbb{N}} \searrow 0$ such that $\lim_{k\to\infty} m_{r_k} = \liminf_{r\to 0} m_r$. In particular, $m_{r_k} \in (0,1)$ for r_k sufficiently small. For simplicity, in the sequel, we drop the subscript k, yet we consider only $r \in \{r_k\}_{k\in\mathbb{N}}$.

Let γ_{Γ_r} be an optimal transportation plan between $m_r\rho_{\delta,r}$ and $\alpha^{-1}\mathscr{H}^1 \llcorner \Gamma_r$ for the Wasserstein-p distance and define the new plan

$$\tilde{\gamma}_{\delta,r} \overset{\text{def.}}{=} \gamma_{\Gamma_r} + (1-m_r)(\mathrm{id}, \pi_{\Gamma_r})_\sharp \rho_{\delta,r}, \text{ and } \tilde{\nu}_{\delta,r} \overset{\text{def.}}{=} \pi_{1\sharp}\tilde{\gamma}_{\delta,r},$$

where π_{Γ_r} is a measurable selection of the projection operator onto Γ_r, this construction is illustrated in Figure 3. Therefore, $\tilde{\nu}_{\delta,r}$ is admissible for (3.33) and we have the following estimate

$$W_p^p(\rho_{\delta,r}, \tilde{\nu}_{\delta,r}) \le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \mathrm{d}\gamma_{\Gamma_r} + (1-m_r)\int_{\mathbb{R}^d} \mathrm{dist}(x,\Gamma_r)^p \mathrm{d}\rho_{\delta,r}.$$

We will estimate each term of the previous inequality separately. For the first one, notice that as $\mathrm{supp}\,\gamma_{\Gamma_r} \subset \Pi_\Sigma^{-1}(B_r(y_0)) \times \overline{B}_{2r}(y_0)$, it holds that

$$|x-y| \le \mathrm{dist}(x,\Sigma) + 3r, \quad \text{for } \gamma_{\Gamma_r}\text{-a.e. } (x,y).$$

For the second term, as the projection of x onto Σ is inside $B_r(y_0)$, if follows from Lemma 3.19 that

$$\mathrm{dist}(x,\Gamma_r) \le \mathrm{dist}(x,\Sigma) - \frac{r}{2}, \text{ for } \mathrm{dist}(x,\Sigma) > 2r.$$

Therefore, for a fixed δ and taking $2r < \delta$, the Wasserstein distance is bounded by

$$W_p^p(\rho_{\delta,r}, \tilde{\nu}_{\delta,r}) \le m_r \int_{\mathbb{R}^d} (\mathrm{dist}(x,\Sigma) + 3r)^p \mathrm{d}\rho_{\delta,r} \\ + (1-m_r)\int_{\mathbb{R}^d} (\mathrm{dist}(x,\Sigma) - r/2)^p \mathrm{d}\rho_{\delta,r}$$

Notice that $W_p^p(\rho_{\delta,r}, \nu_{\delta,r}) = \int_{\mathbb{R}^d} \mathrm{dist}(x,\Sigma)^p \mathrm{d}\rho_{\delta,r}$, so in order to compare the Wassertein distances we use the following inequalities

$$(\mathrm{dist}(x,\Sigma) + 3r)^p \le \mathrm{dist}(x,\Sigma)^p + 3rp(\mathrm{dist}(x,\Sigma) + 3r)^{p-1}$$
$$\Big(\mathrm{dist}(x,\Sigma) - \frac{r}{2}\Big)^p \le \mathrm{dist}(x,\Sigma)^p - \frac{r}{2}p\Big(\mathrm{dist}(x,\Sigma) - \frac{r}{2}\Big)^{p-1}$$

which follow from the convexity of $t \mapsto |t|^p$. Then, given $\varepsilon > 0$, if $r \leq \delta\varepsilon$ one deduces, for $\operatorname{dist}(x,\Sigma) \geq \delta$, that:

$$(\operatorname{dist}(x,\Sigma) + 3r)^p \leq \operatorname{dist}(x,\Sigma)^p + 3rp(1+3\varepsilon)^{p-1}\operatorname{dist}(x,\Sigma)^{p-1}$$
$$\Big(\operatorname{dist}(x,\Sigma) - \frac{r}{2}\Big)^p \leq \operatorname{dist}(x,\Sigma)^p - \frac{r}{2}p\Big(1-\frac{\varepsilon}{2}\Big)^{p-1}\operatorname{dist}(x,\Sigma)^{p-1}.$$

Therefore it holds that

$$W_p^p(\rho_{\delta,r}, \tilde{\nu}_{\delta,r}) \leq W_p^p(\rho_{\delta,r}, \nu_{\delta,r}) + pr\Delta_{r,\varepsilon}\int_{\mathbb{R}^d} \operatorname{dist}(x,\Sigma)^{p-1} \mathrm{d}\rho_{\delta,r}$$
$$\text{for } \Delta_{r,\varepsilon} = 3m_r(1+3\varepsilon)^{p-1} - \frac{1-m_r}{2}\Big(1-\frac{\varepsilon}{2}\Big)^{p-1}$$

Hence from the optimality of $\nu_{\delta,r}$ we have $\Delta_{r,\varepsilon} \geq 0$, so that letting $r \to 0$ and then $\varepsilon \to 0$, it must hold that $4m_0 \geq (1-m_0)/2$, that is:

$$\theta_\delta(y_0) \leq \frac{7}{2}\frac{C_d}{\alpha}.$$

As a result, the family $(\nu_\delta)_{\delta>0}$ has a uniform L^∞ density bounds, and so does the limit measure $\sup_{\delta>0}\nu_\delta = (\sup_{\delta>0}\theta_\delta)\,\mathscr{H}^1 \llcorner \Sigma$. But as the exceeding measure can be decomposed as (3.35) we deduce that whenever the initial measure $\rho_0 \ll \mathscr{H}^1$ or has a L^∞ density w.r.t. $\mathscr{H}^1$, so does the solution ν. □

5. Existence of solutions to (P_Λ)

This section is dedicated to the proof of Theorem 3.1, item (2). Knowing that the excess measure is absolutely continuous (Theorem 3.20), we use a blow up argument near a rectifiability point y_0 of Σ. From Lemma 3.18, the blow-ups of ν_{exc} minimize a family of functionals $(F_r)_{r>0}$, which in turn Γ-converge to some functional F. Since these blow-ups also converge (for $\mathscr{H}^1$-a.e. y_0) to a uniform density on $T_{y_0}\Sigma$, this limit measure must also minimize the Γ-limit F. Yet if it is not zero, we can build a better competitor (Lemma 3.23 below), giving a contradiction to the minimality of the uniform measure. We deduce that ν_{exc} vanishes.

5.1. Blow-up and Γ-convergence

In the sequel, we assume that $\rho_0 \ll \mathcal{H}^1$, so that from Theorem 3.20 any minimizer ν, as well as $(\nu_\delta)_{\delta>0}$ (defined in (3.34)), are rectifiable measures and we can write

$$\nu_\delta = \theta_\delta \mathcal{H}^1 \llcorner \Sigma, \text{ for } \theta_\delta \in L^1(\mathscr{H}^1 \llcorner \Sigma).$$

Observe that ν_δ-a.e. $y \in \Sigma$ is a rectifiability point, and we choose $y_0 \in \Sigma$ such that:

$$T_{y_0}\Sigma \text{ exists} \quad \text{ and } \quad y_0 \text{ is a Lebesgue point of } \theta_\delta. \tag{3.36}$$

We then use Lemma 3.18 with the choice $\mathcal{S}_0 \times \mathcal{S}_1 = D_\delta \times B_r(y_0)$, and we focus on the variational problem (3.28): we obtain the families of measures $(\nu_{\delta,r})_{r>0}$ and $(\sigma_{\delta,r})_{r>0}$ as $\nu_{\delta,r} \overset{\text{def.}}{=} \nu_\delta \llcorner B_r(y_0)$ and $\sigma_{\delta,r} \overset{\text{def.}}{=} \pi_{(1-r)_\sharp}\gamma_{\delta,r}$, where $(\sigma_{\delta,t})_{t\in[0,1]}$ is a family of geodesic interpolations, as in Lemma 3.18, so that

$$\nu_{\delta,r} \in \operatorname{argmin} \left\{ W_p^p(\sigma_{\delta,r},\nu') : \begin{array}{c} \text{there is } \Gamma \text{ such that} \\ \nu' \in \mathcal{M}_+(\Sigma\cup\Gamma), \\ \nu' \geq \alpha^{-1}\mathcal{H}^1 \llcorner (\Gamma\setminus\Sigma), \\ \Sigma\cup\Gamma \in \mathcal{A},\ \nu'(\mathbb{R}^d) = \nu_{\delta,r}(\mathbb{R}^d) \end{array} \right\}. \tag{3.37}$$

From Lemma 3.17 the optimal transport plan between $\nu_{\delta,r}$ and $\sigma_{\delta,r}$ is supported on $\operatorname{graph}(\Pi_\Sigma)$.

The sequence of measures $\nu_{\delta,r}$ are essentially a localization of ν_δ around y_0 so, by the blow-up Theorem 1.16 (see also [?, Theo. 2.83]), it holds that

$$r^{-1}\Phi_\sharp^{y_0,r}\nu_{\delta,r} \xrightarrow[r\to 0]{\star} \theta_\delta(y_0)\mathcal{H}^1 \llcorner [-\tau,\tau], \text{ where } \mathbb{R}\tau = T_{y_0}\Sigma. \tag{3.38}$$

Up to a subsequence (not labelled) we also have:

$$r^{-1}\Phi_\sharp^{y_0,r}\sigma_{\delta,r} \xrightarrow[r\to 0]{\star} \bar{\sigma}_\delta \tag{3.39}$$

for some measure $\bar{\sigma}_\delta$. By construction $\sigma_{\delta,r}$ is supported on $\{r\delta^{-1} \geq \operatorname{dist}(\cdot,\Sigma) \geq r\delta\}$, so that $\operatorname{supp}\bar{\sigma}_\delta \subset \{x : \delta^{-1} \geq \operatorname{dist}(x,\mathbb{R}\tau) \geq \delta\}$.

In view of (3.38) and (3.39), we introduce the blow-ups of the measures $\nu_{\delta,r}$ and $\sigma_{\delta,r}$,

$$\bar{\nu}_{\delta,r} \overset{\text{def.}}{=} \frac{1}{r}\Phi_\sharp^{y_0,r}\nu_{\delta,r}, \quad \bar{\sigma}_{\delta,r} \overset{\text{def.}}{=} \frac{1}{r}\Phi_\sharp^{y_0,r}\sigma_{\delta,r}, \text{ and the set } \Sigma_r \overset{\text{def.}}{=} \frac{\Sigma - y_0}{r} \cap \overline{B_1(0)}. \tag{3.40}$$

In addition, we define a family of functionals $(F_r)_{r>0}$ as

$$F_r(\nu') \overset{\text{def.}}{=} \begin{cases} W_p^p(\bar{\sigma}_{\delta,r},\nu'), & \begin{array}{c} \text{there is } \Gamma \subset \overline{B_1(0)} \text{ such that} \\ \nu' \in \mathcal{M}_+(\Sigma_r\cup\Gamma),\ \nu' \geq \alpha^{-1}\mathscr{H}^1 \llcorner (\Gamma\setminus\Sigma_r), \\ \left(\frac{\Sigma-y_0}{r}\right)\cup\Gamma \text{ closed and connected}, \\ \nu'(\overline{B_1(0)}) = \dfrac{\nu_\delta(B_r(y_0))}{r}, \end{array} \\ +\infty, & \text{otherwise,} \end{cases} \tag{3.41}$$

where $\alpha = \mathcal{L}(\nu)$. Observing that for any given measures μ', ν' we have

$$W_p^p\left(\frac{1}{r}\Phi_\sharp^{y_0,r}\mu', \frac{1}{r}\Phi_\sharp^{y_0,r}\nu'\right) = \frac{1}{r^{p+1}}W_p^p(\mu',\nu'). \tag{3.42}$$

and recalling (3.37), we see that $\bar{\nu}_{\delta,r} \in \operatorname{argmin} F_r$ for any $r > 0$.

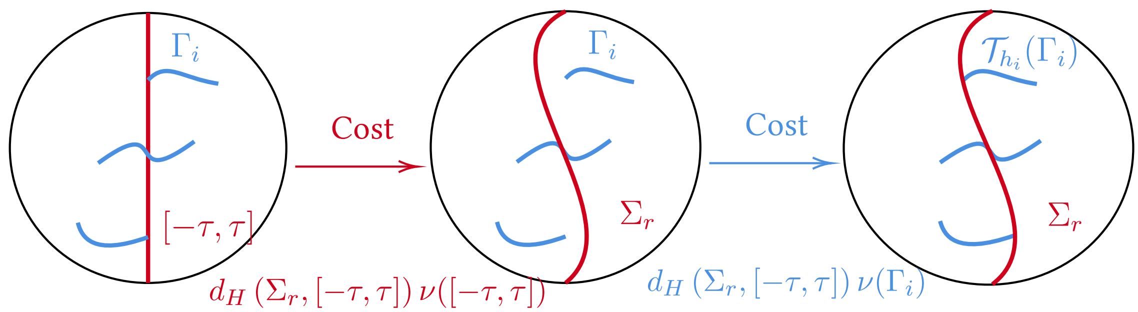

Figure 4: Transportation argument for the construction of a recovery sequence in the Γ convergence of $(F_r)_{r>0}$. Both operations have a transportation cost of the order $d_H\left(\Sigma_r,[-\tau,\tau]\right)$, and hence converge to 0.

The natural candidate for the limit of this family is the following:

$$F(\nu') \overset{\text{def.}}{=} \begin{cases} W_p^p\left(\bar\sigma_\delta,\nu'\right), & \begin{array}{c}\text{there is } \Gamma\subset\overline{B_1(0)} \text{ such that}\\ \nu'\in\mathcal{M}_+\left([-\tau,\tau]\cup\Gamma\right),\ \nu'\geq\alpha^{-1}\mathscr{H}^1\llcorner\left(\Gamma\setminus[-\tau,\tau]\right),\\ \mathbb{R}\tau\cup\Gamma \text{ closed and connected,}\\ \nu'(\overline{B_1(0)})=2\theta_\delta(y_0),\end{array}\\ +\infty, & \text{otherwise.}\end{cases} \tag{3.43}$$

We prove in Theorem 3.21 below that F_r Γ-converges to F as $r\to0^+$. We refer to [?, Braides, 2002] and in particular to [Braides, 2002, Def. 1.24]) for the definition of the (lower and upper) Γ-limit. From the properties of the Γ-convergence, see [?, Cor. 7.20], it follows that $\theta_\delta(y_0)\mathscr{H}^1\llcorner[-\tau,\tau]$ must be a minimizer of F (as the limit of minimizers of F_r). The estimate from below of the Γ-liminf is obtained with the tools developed so far, while estimating the Γ-limsup will require an appropriate construction illustrated in Figure 4.

Theorem 3.21. *The family of functionals $(F_r)_{r>0}$ Γ-converges to F as $r\to0^+$, in the narrow topology.*

Proof. Γ-liminf: we consider an infinitesimal sequence $(r_n)_{n\in\mathbb{N}}$ such that $(\nu'_n)_{n\in\mathbb{N}}$ converges to ν' in the narrow sense in $\overline{B_1(0)}$, and that $\liminf_{n\to\infty}F_{r_n}(\nu'_n)<\infty$ for all $n\in\mathbb{N}$, otherwise there is nothing to prove.

First we look at the first marginals in the definition of F_{r_n}. From (3.39) we know that $\bar\sigma_{\delta,r_n}\xrightarrow[n\to\infty]{\star}\bar\sigma_\delta$. By the lower semi-continuity of the Wasserstein distance w.r.t. the narrow convergence, if we prove that $F(\nu')<\infty$, that is, if the limit satisfies the constraints in the definition of F, we will have that

$$F(\nu')\leq\liminf_{n\to\infty}F_{r_n}(\nu'_n).$$

As $\alpha\nu'_n \geq \mathscr{H}^1 \llcorner (\Gamma_n \setminus \Sigma_{r_n})$ for some $\Gamma_n \subset \overline{B_1(0)}$ such that $\left(\frac{\Sigma - y_0}{r_n}\right) \cup \Gamma_n \in \mathcal{A}$, Blaschke's Theorem [?, Thm. 6.1] and Lemma 2.12 imply that, up to a subsequence, $\Gamma_n \xrightarrow[n\to\infty]{d_H} \Gamma$ for some closed set $\Gamma \subset \overline{B_1(0)}$ and $\frac{\Sigma - y_0}{r_n} \xrightarrow[n\to\infty]{K} \mathbb{R}\tau$. Hence,

$$\Xi_n \overset{\text{def.}}{=} \left(\frac{\Sigma - y_0}{r_n}\right) \cup \Gamma_n \xrightarrow[n\to\infty]{K} \Xi \overset{\text{def.}}{=} \mathbb{R}\tau \cup \Gamma.$$

Let us check that Ξ is connected (which is not immediate since the Kuratowski limit of connected sets is not necessarily connected). Assume by contradiction that there are two disjoint open sets $U, V \subset \mathbb{R}^d$ such that $U \cap \Xi$ and $V \cap \Xi$ form a partition of Ξ. Since $\mathbb{R}\tau \subset \Xi$ is connected, it is contained in either U or V (say, U). As a result, $V \cap \Xi \subset \Gamma \subset \overline{B_1(0)}$ is bounded, and possibly replacing V with $V \cap B_2(0)$, we may assume that V is bounded too, so that ∂V is compact. Since $\Xi \subset V \cap (\mathbb{R}^d \setminus \overline{V})$, we note that $\partial V \cap \Xi = \emptyset$, and we deduce that $\min_{x\in\partial V} \operatorname{dist}(x, \Xi) > 0$.

Now, the Kuratowski convergence of Ξ_n towards Ξ implies that, for all n large enough, Ξ_n intersects both V and $U \subset \mathbb{R}^d \setminus \overline{V}$, hence, by the connectedness of Ξ_n, there exists $x_n \in \Xi_n \cap \partial V$. But the Kuratowski convergence also implies that $\operatorname{dist}(\cdot, \Xi_n) \to \operatorname{dist}(\cdot, \Xi)$ locally uniformly (hence uniformly on ∂V), which contradicts that $\min_{x\in\partial V} \operatorname{dist}(x, \Xi) > 0$. As a result, Ξ is connected.

The fact that $\operatorname{supp} \nu' \subset [-\tau, \tau] \cup \Gamma$ comes from the weak convergence of ν'_n to ν'. As this convergence takes place in a compact set it also holds that $\nu'(\overline{B_1(0)}) = \lim_{n\to\infty} \nu'_n(\overline{B_1(0)}) = 2\theta_\delta(y_0)$ since $\theta_\delta(y_0)$ is the density of ν_δ at y_0.

It only remains to verify the density constraints, $\alpha\nu' \geq \mathcal{H}^1 \llcorner (\Gamma \setminus [-\tau, \tau])$. We cannot apply Gołab's Theorem to ν'_n since, although $\alpha\nu'_n \geq \mathscr{H}^1 \llcorner (\Gamma_n \setminus \Sigma_{r_n})$, we do not have an upper bound on the number of connected components of $\Gamma_n \setminus \Sigma_{r_n}$. What we do know is that the sequence $\Xi_n = r_n^{-1}(\Sigma - y_0) \cup \Gamma_n$ satisfies the assumptions of Theorem 2.10, so we apply it to the measures $\mathscr{H}^1 \llcorner \Xi_n$ instead, remembering that

$$\mathcal{H}^1 \llcorner \left(\frac{\Sigma - y_0}{r_n}\right) + \alpha\nu'_n \geq \mathcal{H}^1 \llcorner \left(\frac{\Sigma - y_0}{r_n} \cup \Gamma_n\right).$$

The left-hand side converges in the local weak-$\star$ sense to $\mathscr{H}^1 \llcorner \mathbb{R}\tau + \alpha\nu'$. The right-hand side (which is bounded by the left-hand side) converges in the same sense, up to a subsequence. We let λ denote a limit and Theorem 2.10 implies that $\lambda \geq \mathscr{H}^1 \llcorner (\mathbb{R}\tau \cup \Gamma)$, which gives $\mathscr{H}^1 \llcorner \mathbb{R}\tau + \alpha\nu' \geq \mathscr{H}^1 \llcorner (\mathbb{R}\tau \cup \Gamma)$, and thus

$$\alpha\nu' \geq \mathscr{H}^1 \llcorner (\Gamma \setminus [-\tau, \tau]).$$

Γ-limsup: Let $(r_n)_{n\in\mathbb{N}}$ be an infinitesimal sequence. By Lemma 2.12, we know that $(\Sigma - y_0)/r_n$ converges in the Kuratowski sense towards $\mathbb{R}\tau$, and $\Sigma_{r_n} \overset{\text{def.}}{=} (\Sigma - y_0)/r_n \cap \overline{B_1(0)}$ converges towards $[-\tau, \tau]$ for the Hausdorff distance.

The strategy to prove the limsup is illustrated in Figure 4, and roughly explained as follows. We concatenate three steps. First we renormalize ν' to satisfy the mass constraint in F_{r_n}. But this normalization may break the condition $\alpha\nu'_n \geq \mathscr{H}^1 \llcorner (\Gamma \setminus [-\tau,\tau])$, so we slightly shrink the support to satisfy this constraint again. We also need the measure ν'_n to be supported on some connected set $\Sigma_{r_n} \cup \Gamma_n$, hence we move the mass of ν' from $[-\tau,\tau]$ to Σ_{r_n} by projection, and we translate the mass of each connected component of the (shrinked) $\Gamma \setminus [-\tau,\tau]$ so that it is connected to Σ_{r_n}. Eventually, by doing so, some parts of the support may get out of $\overline{B_1(0)}$, so we project the residual mass onto $\overline{B_1(0)}$.

To be more precise, we first address the case $\theta_\delta(y_0) = 0$. As $F(\nu') < +\infty$ if and only if $\nu' = 0$, we need only prove the result for $\nu' = 0$. Let P_n be any measurable selection of the projection onto Σ_{r_n}, and define $\nu'_n \overset{\text{def.}}{=} P_{n\sharp}\bar{\sigma}_{\delta,r_n}$. With $\Gamma = \emptyset$, and since $|x - P_n(x)| \leq \delta^{-1}$ for all $x \in \operatorname{supp} \bar{\sigma}_{\delta,r_n}$, we observe that

$$F_{r_n}(\nu'_n) \leq W_p^p(\bar{\sigma}_{\delta,r_n}, \nu'_n) \leq \delta^{-p} \frac{\nu_\delta(B_{r_n})}{r_n} \xrightarrow[n\to+\infty]{} 0 = F(\nu').$$

Moreover, as $\nu'_n \xrightarrow[n\to+\infty]{} \nu'$ in the narrow topology, we have built a recovery sequence for ν'.

Now, we deal with the case $\theta_\delta(y_0) > 0$. Let ν' such that $F(\nu') < +\infty$, and let Γ be a set as in (3.43). Observe that $[-\tau,\tau] \cup \Gamma$ is connected, being the projection of $\mathbb{R}\tau \cup \Gamma$ onto $\overline{B_1(0)}$, and since it has finite $\mathscr{H}^1$ measure, it is arcwise connected, by [David, 2006, Prop. 30.1, Cor. 30.2]. As a result, $\mathbb{R}\tau \cup \Gamma$ is arcwise connected too.

Let $(C_i)_{i\in I}$ denote the arcwise connected components of $\Gamma \setminus (\mathbb{R}\tau)$. For each $i \in I$, as the set $\mathbb{R}\tau \cup \Gamma$ is arcwise connected, one may check that there exists some $z_i \in [-\tau,\tau]$ such that $\{z_i\} \cup C_i$ is arcwise connected. As a result, the set $C_i \subset \mathbb{R}^d \setminus (\mathbb{R}\tau)$ cannot consist of one single point, and $\mathscr{H}^1(C_i) > 0$. Therefore, the index set I is at most countable.

Let us construct a recovery sequence $(\nu'_n)_{n\in\mathbb{N}}$. By the Kuratowski (even Hausdorff) convergence of Σ_{r_n} towards $[-\tau,\tau]$, for each $i \in I$, there exists a sequence $(z_{n,i})_{n\in\mathbb{N}}$ such that $z_{n,i} \in \Sigma_{r_n}$ for each $n \in \mathbb{N}$, and $z_{n,i} \to z_i$. We then define

$$a_n \overset{\text{def.}}{=} \frac{\nu_\delta(B_{r_n})}{2r_n\theta_\delta(y_0)}, \quad \text{and} \quad s_n \overset{\text{def.}}{=} \max(1, a_n^{-1}),$$

noting that $a_n \to 1$ and $s_n \to 1$, and we introduce the map T_n,

$$T_n(y) \overset{\text{def.}}{=} \begin{cases} P_n(y/s_n), & \text{if } y \in [-\tau,\tau], \\ (y - z_i)/s_n + z_{n,i}, & \text{if } y \in C_i, \end{cases}$$

where, as before, P_n is some measurable selection of the projection onto Σ_{r_n}. The map T_n shrinks each connected component C_i and translates it to the corresponding $z_{n,i} \in \Sigma_{r_n}$ so as to ensure connectedness (see below). Letting P_B denote the projection onto the unit ball $\overline{B_1(0)}$, we eventually define

$$\nu'_n \overset{\text{def.}}{=} (P_B \circ T_n)_\sharp(a_n\nu').$$

Let us check that ν'_n converges to ν' in the narrow topology. We note that for $y \in [-\tau,\tau]$,

$$|y/s_n - P_n(y/s_n)| = \operatorname{dist}(y/s_n, \Sigma_{r_n}) \leq d_H([-\tau,\tau], \Sigma_{r_n}) \xrightarrow[n\to+\infty]{} 0,$$

so that $T_n(y) \to y$, and for $y \in C_i$,

$$|y - T_n(y)| \leq |y|(1 - 1/s_n) + |z_i/s_n - z_{n,i}| \xrightarrow[n\to+\infty]{} 0.$$

As a result, for $y \in [-\tau, \tau] \cup \Gamma$, $T_n(y) \to y$, and eventually $P_B \circ T_n(y) \to y$. By the dominated convergence theorem, we get that for any $\phi \in C_b(\mathbb{R}^d)$,

$$\int \phi \mathrm{d}\nu'_n = a_n \int_{[-\tau,\tau]\cup\Gamma} \phi\left(P_B(T_n(y))\right) \mathrm{d}\nu'(y) \xrightarrow[n\to+\infty]{} \int_{[-\tau,\tau]\cup\Gamma} \phi\left(y\right) \mathrm{d}\nu'(y)$$

so that $\nu'_n \xrightarrow[n\to+\infty]{} \nu'$ in the narrow topology.

Let us now check the constraints in F_{r_n}. From the properties of image measures, we see that $\operatorname{supp} \nu'_n \subset \overline{B_1(0)}$, and that $\nu'_n(\overline{B_1(0)}) = \nu'_n(\mathbb{R}^d) = a_n \nu'(\mathbb{R}^d) = \nu_\delta(B_{r_n})/r_n$, so that ν'_n has the mass prescribed by F_{r_n}. Consider the set

$$\Gamma_n \overset{\text{def.}}{=} \bigcup_{i\in I} \Gamma_{n,i} \quad \text{where} \quad \Gamma_{n,i} \overset{\text{def.}}{=} \overline{(P_B \circ T_n)(C_i)}. \tag{3.44}$$

In addition, the mass of ν'_n is concentrated in $\Sigma_{r_n} \cup \Gamma_n$, and we prove below that satisfies all the constraints in F_{r_n}.

First let us show that $\dfrac{\Sigma - y_0}{r_n} \cup \Gamma_n$ is connected. For each $i \in I$, as the set $\{z_i\} \cup C_i$ is arcwise connected, so is its image by the map $y \mapsto (y - z_i)/s_n + z_{n,i}$, which is equal to $\{z_{n,i}\} \cup T_n(C_i)$. As a result $\{z_{n,i}\} \cup \overline{P_B \circ T_n(C_i)} = \{z_{n,i}\} \cup \Gamma_{n,i}$ is connected, as well as $\dfrac{\Sigma - y_0}{r_n} \cup \Gamma_n$.

Now, let us show that $\dfrac{\Sigma - y_0}{r_n} \cup \Gamma_n$ is closed. If I is finite, then, by (3.44), $r_n^{-1}(\Sigma - y_0) \cup \Gamma_n$ is closed as the finite union of closed sets. Otherwise, I is countable, and from [Paolini and Stepanov, 2013, Lemma 2.6], we have

$$\mathscr{H}^1(\Gamma_{n,i}) = \mathscr{H}^1(P_B \circ T_n(C_i)) \leq \mathscr{H}^1(T_n(C_i)) = s_n^{-1} \mathscr{H}^1(C_i) \xrightarrow[i\to\infty]{} 0.$$

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence contained in $r_n^{-1}(\Sigma - y_0) \cup \Gamma_n$, such that $x_k \to x$. If there is an infinite amount of terms of this sequence in either $\dfrac{\Sigma - y_0}{r_n}$ or any of the $\Gamma_{n,i}$, since these sets are closed, then $x \in \dfrac{\Sigma - y_0}{r_n} \cup \Gamma_n$. Otherwise, we can find a sub-sequence $x_{k'} \in \Gamma_{n,i_{k'}}$, so that

$$\operatorname{dist}\left(x, \frac{\Sigma - y_0}{r_n}\right) = \lim_{k'\to\infty} \operatorname{dist}\left(x_{k'}, \frac{\Sigma - y_0}{r_n}\right) \leq \lim_{k'\to\infty} \mathscr{H}^1(\Gamma_{n,i_{k'}}) = 0,$$

and we conclude that $\dfrac{\Sigma - y_0}{r_n} \cup \Gamma_n$ is closed.

To show it satisfies the density constraints, take any non-negative $\phi \in C_b(\mathbb{R}^d)$,

$$\begin{aligned}\alpha \int \phi \mathrm{d}\nu'_n &= \alpha a_n \int_{[-\tau,\tau]\cup\Gamma} \phi\left(P_B(T_n(y))\right) \mathrm{d}\nu'(y) \\ &\geq \alpha a_n \sum_{i\in I} \int_{C_i} \phi\left(P_B(T_n(y))\right) \mathrm{d}\nu'(y) \\ &\geq a_n \sum_{i\in I} \int_{C_i} \phi\left(P_B((y-z_i)/s_n + z_{n,i})\right) \mathrm{d}\mathscr{H}^1(y) \\ &= a_n s_n \sum_{i\in I} \int_{\Gamma_{n,i}} \phi\left(P_B(y')\right) \mathrm{d}\mathscr{H}^1(y') \\ &\geq \int_{\Gamma_n} \phi \mathrm{d}\mathscr{H}^1.\end{aligned}$$

It follows that $\alpha\nu'_n \geq \mathscr{H}^1 \llcorner \Gamma_n$ and we conclude that $F_{r_n}(\nu_n) < \infty$, for all $n \in \mathbb{N}$.

By the continuity of the Wasserstein distance with respect to the narrow convergence (provided the measures are supported in some common compact set), we have that:

$$F_{r_n}(\nu'_n) \xrightarrow[n\to\infty]{} F(\nu').$$

The Γ-convergence follows. $\square$

Now that we have characterized the limit problem, we show that the optimal transportation is given by projections as the blow-up family.

Lemma 3.22. *If $\theta_\delta(y_0) > 0$, the optimal transport plan between the measure σ_δ, defined in* (3.39)*, and $\bar\nu = \theta_\delta(y_0)\mathscr{H}^1 \llcorner [-\tau, \tau]$, defined in* (3.38)*, is unique and given by the projection map $\Pi_{[-\tau,\tau]}$.*

Proof. Consider a family $\bar\gamma_r$ of optimal transport plans from $\bar\sigma_{\delta,r}$ to $\bar\nu_{\delta,r}$. Up to a subsequence it converges to some $\bar\gamma$, which, by the stability of optimal transport plans, also transports σ_δ to $\bar\nu$ optimally. Since $\bar\sigma_{\delta,r}, \bar\nu_{\delta,r}$ are generated by the pushforward of $\nu_{\mathrm{exc}} \llcorner B_r(y_0)$ by $\Phi^{y_0,r}$, from Lemma 3.17 we know that

$$\operatorname{supp} \bar\gamma_r \subset \operatorname{graph}\left(\Pi_{\Sigma_r}\right).$$

Let us show that $\operatorname{supp}\bar\gamma \subset \operatorname{graph}\left(\Pi_{[-\tau,\tau]}\right)$. Indeed if $(x,p) \in \operatorname{supp}\bar\gamma$, there is an open ball B centered at (x,p) such that

$$0 < \bar\gamma(B) \leq \liminf_{r\to 0} \bar\gamma_r(B).$$

In particular, we can find $\operatorname{supp}\bar\gamma_r \ni (x_r, p_r) \xrightarrow[r\to 0]{} (x,p)$. So it holds that

$$|x-p| = \lim_{r\to 0} |x_r - p_r| = \lim_{r\to 0} \operatorname{dist}\left(x_r, \Sigma_r\right) = \operatorname{dist}(x, [-\tau,\tau]),$$

where the last equality comes from the uniform convergence of the distance functions, recalling from Lemma 2.12 that $\Sigma_r \xrightarrow[r\to 0]{d_H} [-\tau,\tau]$.

Now we show that this property is true for any other optimal plan. Consider γ transporting σ_δ to $\bar\nu$ optimaly, then by the optimality of $\bar\gamma$ it holds that

$$\begin{aligned}\int_{\mathbb{R}^d} (\mathrm{dist}(x,[-\tau,\tau]))^p \mathrm{d}\sigma_\delta &= \int |x-y|^p \mathrm{d}\bar\gamma = \int |x-y|^p \mathrm{d}\gamma \\ &\geq \int (\mathrm{dist}(x,[-\tau,\tau]))^p \mathrm{d}\gamma = \int_{\mathbb{R}^d} \mathrm{dist}(x,[-\tau,\tau])^p \mathrm{d}\sigma_\delta.\end{aligned}$$

Since $|x-y| - \mathrm{dist}(x,[-\tau,\tau]) \geq 0$ for γ-a.e. (x,y) and the inequality above must be an equality, we must have $\operatorname{supp}\gamma \subset \operatorname{graph}\left(\Pi_{[-\tau,\tau]}\right)$ for any optimal γ. In particular, as $\Pi_{[-\tau,\tau]}$ is uni-valued, it means that the optimal transport plan is unique and given by the projection map. □

5.2. Competitor for the limit problem and existence for (P_Λ)

We now show that if $\theta_\delta > 0$ on a set of positive measure, we reach a contradiction, by building a better competitor for the Γ-limit problem. It follows from Theorem 3.21 that:

$$\lambda^{y_0} \stackrel{\text{def.}}{=} \theta_\delta(y_0)\mathscr{H}^1 \llcorner [-\tau,\tau] \in \operatorname{argmin} F,$$

where F is defined in (3.43). In addition, Lemma 3.22 shows that that the optimal transportation of σ^{y_0} to λ^{y_0} is given by the orthogonal projection. We show that we can lower the energy by projecting part of the mass to a (closer) horizontal line as in Figure 5. This contradicts the existence of rectifiability points of Σ such that $\theta_\delta(y_0) > 0$ so that $\nu_\delta \equiv 0$, and shows the following Lemma:

Lemma 3.23. *For any $\delta > 0$, the measures ν_δ defined in* (3.34) *vanish.*

Proof. Up to a rotation, we may assume that $\tau = e_d$, where $(e_i)_{i=1}^d$ is a basis of $\mathbb{R}^d$. Since σ^{y_0} is supported on $\left\{x = (x', x_d) \in \mathbb{R}^d : |x'| > \delta,\ |x_d| \leq 1\right\}$, we can cover its support with finitely many sets $(E_i)_{i=1}^N$ defined as:

$$E_i \stackrel{\text{def.}}{=} \left\{x = (x', x_d) \in \mathbb{R}^d : \langle \xi_i, x\rangle > \delta/2,\ |x_d| \leq 1\right\}$$

where $\xi_i \in \mathbb{S}^{d-1} \cap [e_d]^\perp$ are unit vectors and N depends only on the dimension. We then define a disjoint family

$$F_1 = E_1, \quad F_{i+1} = E_{i+1} \setminus \bigcup_{j=1}^{i} F_i \text{ for } i \geq 1$$

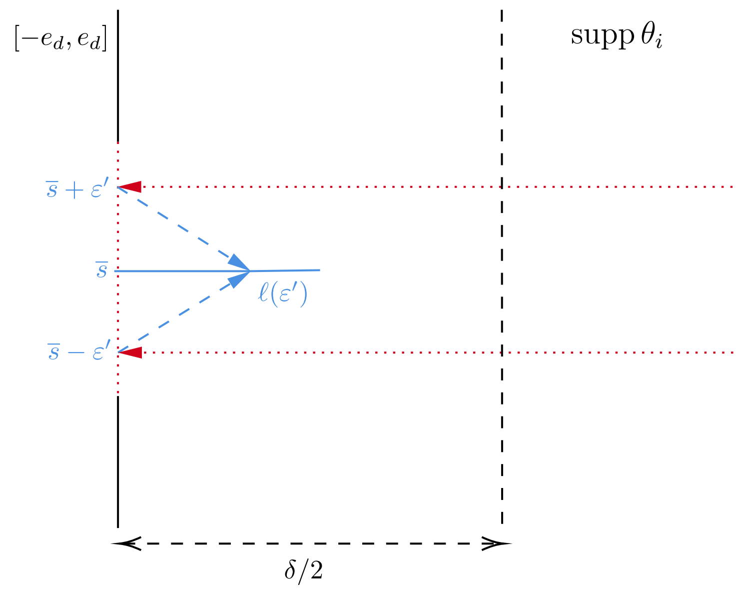

Figure 5: Construction of a competitor for the minimization of F.

and decompose our measures σ^{y_0} and λ^{y_0} as

$$\sigma^{y_0} = \sum_{i=1}^N \sigma_i,\ \lambda^{y_0} = \sum_{i=1}^N \lambda_i \text{ where } \sigma_i \overset{\text{def.}}{=} \sigma^{y_0} \llcorner F_i \text{ and } \lambda_i \overset{\text{def.}}{=} \operatorname{proj}_{d\sharp}\sigma_i,$$

with $\operatorname{proj}_d : x \mapsto x_d e_d$ the projection onto the vertical axis. By Radon-Besicovitch's differentiation theorem, $\lambda_i = \theta_i \mathscr{H}^1 \llcorner [-e_d, e_d]$, where $\theta_i(s) = \theta_i(se_d) \geq 0$ are such that

$$\sum_{i=1}^N \theta_i = \theta_\delta(y_0).$$

Consider $\bar{s} \in (-1,1)$ a common Lebesgue point of all θ_i, $i = 1, \ldots, N$. Let i be the index for which $\theta_i(\bar{s})$ is maximal: then $\theta_i(\bar{s}) \geq \theta_\delta(y_0)/N$. Up to a change of coordinates, we assume that $\xi_i = e_1$, and we introduce the notation: $\mathbb{R}^d \ni x = (x_1, x'', x_d)$ for $x'' \in \mathbb{R}^{d-2}$. Let now:

$$C_\varepsilon \overset{\text{def.}}{=} F_i \cap \{x \in \mathbb{R}^d : |x_d - \bar{s}| < \varepsilon\} \subset \{x = (x_1, x'', x_d) : x_1 > \delta/2,\ |x_d - \bar{s}| < \varepsilon\}.$$

We obtain, from the fact that $(\operatorname{proj}_d)_\sharp \sigma_i = \theta_i \mathscr{H}^1 \llcorner [-e_d, e_d]$, that

$$\frac{\sigma_i(C_\varepsilon)}{2\varepsilon} = \frac{1}{2\varepsilon}\int_{\bar{s}-\varepsilon}^{\bar{s}+\varepsilon} \theta_i(t)\mathrm{d}t \xrightarrow[\varepsilon\to 0]{} \theta \overset{\text{def.}}{=} \theta_i(\bar{s}) \geq \frac{\theta_\delta(y_0)}{N}.$$

Now, assume by contradiction that $\theta > 0$. If ε is small enough, we have:

$$\theta \leq \frac{\sigma_i(C_{\varepsilon'})}{\varepsilon'} \leq 3\theta. \tag{3.45}$$

for all $\varepsilon' \leq \varepsilon$. Now let us exploit the fact that, from Lemma 3.22, the optimal transport is given by projections to propose a new transportation map, sending the mass in C_ε to a segment pointing towards e_1:

$$T(x) \stackrel{\text{def.}}{=} \begin{cases} \ell(|x_d - \bar{s}|)e_1 + \bar{s}e_d, & \text{if } x \in C_\varepsilon \\ \operatorname{proj}_d(x), & \text{otherwise,} \end{cases}$$

where $\ell : [0, \varepsilon] \to \mathbb{R}_+$ is defined via the conservation of mass relation

$$\ell(\varepsilon') = \alpha\sigma_i(C_{\varepsilon'}). \tag{3.46}$$

In other words, the mass that was sent to the vertical segment $[\bar{s} - \varepsilon', \bar{s} + \varepsilon']e_d$ is now sent to the horizontal segment $\bar{s}e_d + [0, \ell(\varepsilon')]e_1$, for each $\varepsilon' \in [0, \varepsilon]$. This construction is illustrated in Figure 5.

Thanks to (3.46), the map T sends $\sigma_i \llcorner C_\varepsilon$ to the measure $\alpha^{-1}\mathscr{H}^1 \llcorner L$ where $L \stackrel{\text{def.}}{=} \bar{s}e_d + [0, \ell(\varepsilon)]e_1$, hence, the transported measure $T_\sharp\sigma^{y_0}$ satisfies the constraints in the definition (3.43) of the limiting functional F and one has $F(T_\sharp\sigma^{y_0}) < +\infty$.

We shall now see that for each point $x \in C_\varepsilon$ with $x_d \neq \bar{s}$, it holds that

$$|x - \operatorname{proj}_d(x)|^p > |x - T(x)|^p. \tag{3.47}$$

To show (3.47), recalling the notation $x = (x_1, x'', x_d)$, it suffices that

$$\begin{aligned} |x - \operatorname{proj}_d(x)|^2 > |x - &T(x)|^2 \\ &\iff x_1^2 + |x''|^2 > (x_1 - \ell(|x_d - \bar{s}|))^2 + |x''|^2 + (x_d - \bar{s})^2 \\ &\iff 2x_1\ell(|x_d - \bar{s}|) > \ell(|x_d - \bar{s}|)^2 + (x_d - \bar{s})^2. \end{aligned}$$

In addition to (3.45), we choose ε in such a way that for any $x \in C_\varepsilon$ we have

$$\alpha\theta|x_d - \bar{s}| \leq \ell(|x_d - \bar{s}|) = \alpha\sigma_i(C_{|x_d - \bar{s}|}) \leq 3\alpha\theta\varepsilon < \left(1 + \frac{1}{(\alpha\theta)^2}\right)^{-1}\delta$$

and hence

$$\begin{aligned} \ell(|x_d - \bar{s}|)^2 + (x_d - \bar{s})^2 \leq \left(1 + \frac{1}{(\alpha\theta)^2}\right)\ell(|x_d - \bar{s}|)^2 < \delta\ell(|x_d - \bar{s}|) \\ \leq 2x_1\ell(|x_d - \bar{s}|), \end{aligned}$$

for all $x \in C_\varepsilon$, with $x_d \neq \bar{s}$, so that (3.47) holds. Since $\theta = \theta_i(\bar{s}) > 0$, it follows that

$$F(T_\sharp\sigma^{y_0}) = W_p^p(\sigma^{y_0}, T_\sharp\sigma^{y_0}) < W_p^p(\sigma^{y_0}, \lambda^{y_0}) = F(\lambda^{y_0}).$$

This contradicts the fact that $\theta_\delta(y_0)\mathscr{H}^1 \llcorner [-e_d, e_d]$ is a minimizer of F, showing that we must have $\theta_i(\bar{s}) = 0$ and, in turn, $\theta_\delta(y_0) = 0$. As this holds for $\mathscr{H}^1$-a.e. point $y_0 \in \Sigma$, we deduce that $\nu_\delta \equiv 0$. $\square$

The previous lemma, combined with the characterization of solutions, as in (3.35),

$$\nu = \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma + \sup_{\delta>0} \nu_\delta + \rho_{\text{exc}} \llcorner \Sigma$$

proves point (2) of Theorem 3.1, giving existence of solutions of our original problem (P_Λ) whenever the initial measure ρ_0 does not give mass to 1-dimensional sets. In fact, we have proven the following, slightly stronger, result.

Theorem 3.24. *Let $\rho_0 \in \mathscr{P}_p(\mathbb{R}^d)$ and suppose that the parameter $\Lambda < \Lambda_\star$. Then the solution to the relaxed problem ($\overline{P}_\Lambda$) is of the form*

$$\nu = \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma + \rho_{exc} \llcorner \Sigma,$$

where $\alpha = \mathcal{L}(\nu)$ and ρ_{exc} was defined in (3.23).

In addition, if ρ_0 does not give mass to 1-rectifiable sets, any solution of the relaxed problem ($\overline{P}_\Lambda$) corresponds to a solution of the original shape optimization problem (P_Λ).

Remark 3.25. *In the characterization of solutions given by*

$$\nu = \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma + \rho_{exc} \llcorner \Sigma,$$

the last term is reminiscent of Lemma 3.17, that says that the excess measure ν_{exc} is formed through projections. Indeed, rewriting it as

$$\nu_{exc} = \sup_{\delta>0} \nu_\delta + \rho_{exc} \llcorner \Sigma,$$

as in equation (3.35), we have shown that in Lemma 3.23 that the components ν_δ coming from a distance δ to Σ are in fact null.

6. Discussion

In this Chapter we have introduced the Wasserstein-$\mathscr{H}^1$ problem and constructed its basic theory. As a 1-dimensional shape optimization problem, we have profited from the various techniques developed in the literature for this class of problems, but the unavoidable use of the narrow topology of Radon measures makes the proof of existence for (P_Λ) deviate considerably from other problems of this class.

This was our motivation to obtain the slight generalization of Gołab's Theorem in 2.10, where we allow for Kuratowski convergence and for sequences of sets that might have infinite length. However, it is important to point out that we have only required this generalization in the proof of Γ-convergence of the blow-ups in Theorem 4.7. This technique is very similar to the one employed by Santambrogio and Tilli [Santambrogio and Tilli, 2005] to study the topology of blow-ups of solutions to the average distance minimizers problem [Lemenant, 2010]. Although convoluted, the strategy of proof by blow-up presented in Section 5 seems to be very useful, as we shall also employ a variant in Chapter 4.

Of course, the proof of existence of minimizers to a variational problem is only the beginning of its study and we hope that the present chapter has instigated the appetite of the reader for more. Two natural questions about problem (P_Λ) arise:

- What kind of qualitative properties does minimizers of such problems enjoy?
- How can we compute them?

These questions shall be the focus of Chapters 4 and 5.

Chapter 4

Qualitative properties for the Wasserstein-$\mathscr{H}^1$ problem

Contents

1. Introduction

In this Chapter we are concerned with qualitative properties of minimizers of the variational problem (P_Λ). More specifically, we show that under certain conditions on ρ_0, solutions are Ahlfors regular (*cf.* Definition 4.1 below) and do not have loops, that is they do not contain a subset that is homeomorphic to $\mathbb{S}^1$, see Definition 2.14 from Chapter 2.

The Ahlfors regularity for instance, will be proved in a high integrability regime, *i.e.* for $\rho_0 \in L^{\frac{d}{d-1}}(\mathbb{R}^d)$. In this case, we know from Theorem 3.24 that solutions of the relaxed problem ($\overline{W\mathscr{H}^1}$) are necessarily uniformly distributed over their support, being therefore a solution to the original problem (P_Λ). However, even in this regime, the relaxed problem is still very valuable since we can create many more variations for it than for the original formulation, so one can expect that more optimality conditions can be derived from the relaxed formulation. Indeed, in the proof of Ahlfors regularity we shall craft competitors that are suitable for the relaxed problem, but not for the original one, see Figure 6 and its description in the beginning of Section 2.

In the sequel we study when is it that solutions are absent of loops, or *tree property*. The first key result we show in Proposition 4.5 that loops are formed through projections. We then use this result to prove the tree property in two cases; when ρ_0 is a point cloud measure, a convex combination of Dirac masses; and $\rho_0 \in L^{\frac{d}{d-1}}(\Omega)$, where Ω is a compact subset of $\mathbb{R}^d$, exploiting the Ahlfors regularity presented by minimizers in this regime.

This is somehow a surprising behavior since an intermediate level of regularity, for instance $\rho_0 = \mathscr{H}^1 \llcorner S$, will necessarily have $\mathscr{H}^1 \llcorner S$ as a solution for Λ sufficiently small. Therefore, if S has a loop, there is at least one minimizer that also have.

2. Ahlfors regularity

In this section we prove that whenever the initial measure $\rho_0 \in L^{\frac{d}{d-1}}(\mathbb{R}^d)$, the optimal solutions to the relaxed problem ($\overline{P}_\Lambda$) have an Ahlfors regular support.

Definition 4.1. *We say that a set $\Sigma \subset \mathbb{R}^d$ is Ahlfors regular whenever there exist $r_0 > 0$ and $c, C > 0$ such that for $r \leq r_0$ it holds that*

$$cr \leq \mathscr{H}^1(\Sigma \cap B_r(x)) \leq Cr, \text{ for all } x \in \Sigma.$$

We prove in this section the following result.

Theorem 4.2. *If $\rho_0 \in L^{\frac{d}{d-1}}(\mathbb{R}^d)$, let ν be a solution of the relaxed problem ($\overline{P}_\Lambda$) and Σ its support. Then Σ is Ahlfors-regular, there exist $r_0 > 0$ depending on d, p, ρ_0 and α and $\bar{C} > 0$ depending only on d and p such that, for all $\bar{x} \in \Sigma$ and $r \leq r_0$,*

$$r \leq \mathscr{H}^1(\Sigma \cap B_r(\bar{x})) \leq \bar{C}r.$$

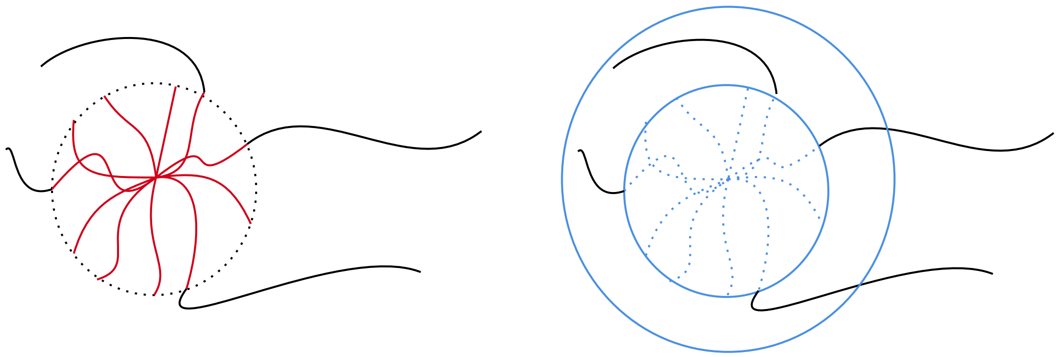

Figure 6: Construction for a proof of Ahlfors regularity in 2D. Adding the inner circumference we preserve connectedness, and adding the second allows for a smaller transportation cost for a large portion of the mass.

The lower bound (with $c = 1$ and $r_0 = \operatorname{diam} \Sigma$) follows directly from the connectedness of Σ, hence we skip the proof. The upper bound will follow as a corollary of Lemma 4.3 below. Let us describe the strategy for proving this estimate.

The idea is similar to proving the L^∞ bound on the excess measure: if in a small ball $B_r(\bar{x})$ the measure ν has too much mass, we build another "closer" 1D structure onto which the mass is transported with a smaller cost. Yet there is an additional difficulty: when replacing $\Sigma \cap B_r(\bar{x})$ with another set we should preserve connectedness. In Theorem 3.20, we were rearranging only the excess mass, and it was not an issue.

In dimension $d = 2$, we can simply remove all the mass from $\Sigma \cap B_r(x_0)$ and add the structures $\partial B_r(x_0) \cup \partial B_R(x_0)$, for some $R > 0$ large enough. If Σ does not satisfy the definition of Ahlfors regularity at x_0 for any constant $C > 0$, we are sure to have enough mass in $B_R(x_0)$ that is sent to $\Sigma \cap B_r(x_0)$ to form these circumferences, while still having plenty of mass left outside $B_R(x_0)$. The inner circumference preserve the connectedness of this new competitor, while we can project the remaining mass, which is a much greater fraction than the rest, onto the outer circumference, gaining a lot in terms of transportation cost. This argument is illustrated in Figure 6.

For a general dimension d, we need to control the number of connected components of $\Sigma \setminus B_r(\bar{x})$ and find a way to connect them back without adding too much length. This number of connected components is controlled by the quantity $\mathscr{H}^0(\Sigma \cap \partial B_r(\bar{x}))$, which we can control on average by means of the *generalized area formula* [Ambrosio et al., 2000, Theorem 2.91]: If $f : \mathbb{R}^M \to \mathbb{R}^N$ is a Lipschitz function and $E \subset \mathbb{R}^M$ is a k-rectifiable set then it holds that

$$\int_{\mathbb{R}^N} \mathscr{H}^0(E \cap f^{-1}(y)) \mathrm{d}\mathscr{H}^k(y) = \int_E J_k \mathrm{d}^E f_x \mathrm{d}\mathscr{H}^k(x), \tag{4.1}$$

where $\mathrm{d}^E f_x$ is the restriction of $\nabla f(x)$ (when f is smooth) to the approximate tangent space of E. Hence, choosing $E = \Sigma \cap (B_{r_1}(\bar{x}) \setminus B_{r_2}(\bar{x}))$ and $f : x \mapsto |x - \bar{x}|$, we deduce from (4.1) that

$$\int_{r_2}^{r_1} \mathscr{H}^0(\Sigma \cap \partial B_s(\bar{x}))\mathrm{d}s \leq \mathscr{H}^1(\Sigma \cap B_{r_1}(\bar{x})) - \mathscr{H}^1(\Sigma \cap B_{r_2}(\bar{x})) \tag{4.2}$$

Using this we first prove the following lemma:

Lemma 4.3. *Assume $\rho_0 \in L^{\frac{d}{d-1}}(\mathbb{R}^d)$. There exist $\bar{C}(d,p) > 0$ and r_0 depending on ρ_0, α, d, p, such that for any $C \geq \bar{C}$, if $r \leq r_0$ and $x \in \Sigma$, then either $\mathscr{H}^1(\Sigma \cap B_r(x)) \leq Cr$ or $\mathscr{H}^1(\Sigma \cap B_{2r}(x)) \geq 10Cr$.*

Proof. Let $r > 0$ and $C \geq 1$, and let $\bar{x} \in \Sigma$ such that both $\mathscr{H}^1(\Sigma \cap B_r(\bar{x})) > Cr$ and $\mathscr{H}^1(\Sigma \cap B_{2r}(\bar{x})) < 10Cr$. We show that if $r \leq r_0$ and $C \geq \bar{C}$, which will both be chosen later, then we can contruct a better competitor to the minimizer ν.

The function $f : s \mapsto \mathscr{H}^1(\Sigma \cap B_s(\bar{x}))$ is nondecreasing, hence in $\mathrm{BV}(\mathbb{R}_+)$ and satisfies, thanks to (4.2), that $\mathscr{H}^0(\Sigma \cap \partial B_s(\bar{x}))\mathrm{d}s \leq Df$ in the sense of measures (equivalently, $\mathscr{H}^0(\Sigma \cap \partial B_s(\bar{x}))$ is less than, or equal to $f'(s)\mathrm{d}s$, the absolutely continuous part of Df).

We note that

$$\begin{aligned}
\inf_{s \in (3r/2, 2r)} \left(\frac{s\mathscr{H}^0\left(\Sigma \cap \partial B_s(\bar{x})\right)}{\mathscr{H}^1(\Sigma \cap B_s(\bar{x}))} \right) &\leq \frac{2}{r} \int_{3r/2}^{2r} \frac{s\mathscr{H}^0\left(\Sigma \cap \partial B_s(\bar{x})\right)}{\mathscr{H}^1(\Sigma \cap B_s(\bar{x}))} ds \\
&\leq 4 \int_{3r/2}^{2r} \frac{1}{f(s)} f'(s)\mathrm{d}s \\
&\leq 4 \ln \left(\frac{f(2r)}{f(3r/2)} \right),
\end{aligned}$$

where we have used the classical chain rule at almost every point and [Ambrosio et al., 2000, Cor. 3.29]. Since $f(2r)/f(3r/2) < (10Cr)/(Cr) = 10$, we deduce that there exists $\bar{s} \in (3r/2, 2r)$ such that

$$\bar{\delta}\bar{s}\mathscr{H}^0(\Sigma \cap \partial B_{\bar{s}}(\bar{x})) \leq \mathscr{H}^1(\Sigma \cap B_{\bar{s}}(\bar{x})) \quad \text{where } \bar{\delta} \overset{\text{def.}}{=} \tfrac{1}{4 \ln 10}. \tag{4.3}$$

Now, we let

$$M = 2\left(1 + 10 \cdot \left(\frac{40}{17}\right)^{p-1}\right) \tag{4.4}$$

(this choice will be made clear at the end of this proof) and we consider

$$\delta \overset{\text{def.}}{=} \frac{\bar{\delta}}{10M} < \bar{\delta} < \frac{1}{2}. \tag{4.5}$$

We define a set Γ as follows: we choose a finite covering of $\partial B_1(0)$ with balls $B(x_i, \delta/2)$ centered at points $(x_i)_{i=1}^N$ (the minimal number N depends only on d and p, through δ). Then, we find a minimal tree connecting the points $(x_i)_{i=1}^N$ through geodesics on the sphere. We add to this minimal tree the segments $[x_i, (1+\delta)x_i]$, $i = 1, \ldots, N$. We call Γ

the resulting (connected) set, whose total length $L \stackrel{\text{def.}}{=} \mathscr{H}^1(\Gamma)$ is of order at most $2N\delta$ and depends only on d and p. Notice that each point of ∂B_1 is at distance at most δ, along the geodesic curve on the sphere, to a point of Γ, and that thanks to the "spikes" $[x_i, (1+\delta)x_i]$, any point with, say, $|x| \geq 10$ is closer to a point of Γ than from any point in $B_1(0)$.

Now, we define

$$\Gamma_{\bar{s}} \stackrel{\text{def.}}{=} (\bar{x} + \bar{s}\Gamma) \cup \bigcup_{x \in \Sigma \cap \partial B_{\bar{s}}} S_x,$$

where S_x denotes a geodesic connecting x to $\bar{x} + \bar{s}\Gamma$, of length at most $\mathscr{H}^1(S_x) \leq \bar{s}\delta$. Since $\bar{s} < 2r$ and $\delta < 1/2$, it follows that $\Gamma_{\bar{s}} \subset B_{3r}(\bar{x})$. We define the competitor set as

$$\Sigma' \stackrel{\text{def.}}{=} \Sigma \setminus B_{\bar{s}}(\bar{x}) \cup \Gamma_{\bar{s}}.$$

The addition of the geodesics S_x ensures that Σ' remains connected, and using (4.3), we estimate the length of $\Gamma_{\bar{s}}$ as

$$\begin{aligned} \mathscr{H}^1(\Gamma_{\bar{s}}) &\leq L\bar{s} + \delta\bar{s}\mathscr{H}^0(\Sigma \cap \partial B_{\bar{s}}(\bar{x})) \leq 2Lr + \tfrac{1}{10M}\mathscr{H}^1(\Sigma \cap B_{\bar{s}}(\bar{x})) \\ &< (2L + \tfrac{C}{M})r, \end{aligned} \tag{4.6}$$

Now we define a new competitor ν' whose support is Σ'. If γ denotes the optimal transportation plan from ρ_0 to ν, given $s > 0$ let

$$\rho_s \stackrel{\text{def.}}{=} \pi_{0\sharp}\left(\gamma \llcorner \left(\mathbb{R}^d \times B_s\right)\right)$$

denote the portion of the measure ρ_0 which is transported to the ball B_s. In particular, the above length estimates imply that

$$Lr \leq \mathscr{H}^1(\Gamma_{\bar{s}}) < (2L + \tfrac{C}{M})r \leq (2\tfrac{L}{C} + \tfrac{1}{M})\alpha\nu(B_r) \leq \alpha\rho_r(\mathbb{R}^d) \leq \alpha\rho_{\bar{s}}(\mathbb{R}^d), \tag{4.7}$$

where $\alpha \stackrel{\text{def.}}{=} \mathcal{L}(\nu)$, and using that $M \geq 2$ (see (4.4)) and assuming $\bar{C} \geq 4L$ (which we recall depends only on d and p). But, if r is small enough (not depending on $\bar{x}$, by uniform equi-integrability of $\rho_0^{d/(d-1)}$) Holder's inequality implies that

$$\alpha\rho_{\bar{s}}(B_{10r}(\bar{x})) \leq \alpha\|\rho_0\|_{L^{\frac{d}{d-1}}(B_{10r}(\bar{x}))}|B_{10r}(\bar{x})|^{\frac{1}{d}} \leq Lr. \tag{4.8}$$

We fix $r_0 > 0$, which depends only on the dimension (through L), the integrability of ρ_0, and α, such that the above inequality holds for $r \leq r_0$.

Equations (4.7)-(4.8) show that for r small enough, part of the mass transported to $\nu \llcorner B_{\bar{s}}$ must come from outside of the ball B_{10r}. In particular, since $t \mapsto \rho_{\bar{s}}(B_t(\bar{x}))$ is continuous, there is $R > 10r$ such that

$$\rho_{\bar{s}}(B_R(\bar{x})) = \alpha^{-1}\mathscr{H}^1(\Gamma_{\bar{s}}). \tag{4.9}$$

To form the new competitor we use the following strategy: the mass sent to $\Sigma \setminus B_{\bar{s}}$ remains untouched, the mass $\rho_{\bar{s}} \llcorner B_R$ previously used to form $\nu \llcorner B_{\bar{s}}$ is transported to $\alpha^{-1}\mathscr{H}^1 \llcorner \Gamma_{\bar{s}}$ and the remaining mass is projected onto $\Gamma_{\bar{s}}$.

So, letting $\tilde\gamma$ denote the optimal transportation plan between $\rho_{\bar s}\llcorner B_R$ and $\alpha^{-1}\mathscr{H}^1\llcorner\Gamma_{\bar s}$, define the new plan

$$\gamma' = \gamma\llcorner\mathbb{R}^d\times B_{\bar s}(\bar x)^c + \tilde\gamma\llcorner B_R\times\mathbb{R}^d + (\mathrm{id},\pi_{\Gamma_{\bar s}})_\sharp\left(\rho_{\bar s}\llcorner B_R^c\right),$$

and the new competitor ν' as its second marginal. By construction, $\alpha\nu'\geq\mathscr{H}^1\llcorner\Sigma'$ so that $\mathcal{L}(\nu')\leq\mathcal{L}(\nu)$. We now estimate the gain in terms of transportation cost.

- For $(x,y)\in B_R\times B_{\bar s}$ and for any $y'\in\Gamma_{\bar s}\subset B_{3r}$, as $\bar s\leq 2r$ and $10r<R$, the convexity of $t\mapsto t^p$ yields

$$\begin{aligned}|x-y'|^p &\leq (|x-y|+5r)^p \leq |x-y|^p + 5rp\left(|x-y|+5r\right)^{p-1}\\ &\leq |x-y|^p+5rp(2R)^{p-1}.\end{aligned}$$

  Hence integrating w.r.t. the transportation plans we get

$$\int_{B_R\times\Gamma_{\bar s}}|x-y'|^p\mathrm{d}\tilde\gamma\leq\int_{B_R\times B_{\bar s}}|x-y|^p\mathrm{d}\gamma+5rp\left(2R\right)^{p-1}\rho_{\bar s}\left(B_R\right),$$

  (this can be checked by disintegration w.r.t. their common first marginal, which is the measure $\rho_{\bar s}\llcorner B_R$).

- Similarly, for $x\in B_R^c$ and $y\in B_{\bar s}\setminus B_r$ the addition of the spikes ensures that

$$|x-\pi_{\Gamma_{\bar s}}(x)|\leq|x-y|.$$

  However if $x\in B_R^c$ and $y\in B_r$ it holds that

$$|x-\pi_{\Gamma_{\bar s}}(x)|\leq|x-y|-\frac{r}{2}\text{ and }|x-y|\geq R-r,$$

  so that once again using the convexity of $t\mapsto t^p$ we have

$$\begin{aligned}|x-\pi_{\Gamma_{\bar s}}(x)|^p &\leq \left(|x-y|-\frac{r}{2}\right)^p\leq|x-y|^p-p\frac{r}{2}\left(|x-y|-\frac{r}{2}\right)^{p-1}\\ &\leq|x-y|^p-p\frac{r}{2}\left(\frac{17}{20}R\right)^{p-1}.\end{aligned}$$

  So, decomposing the integration for the points going to B_r and to $B_{\bar s}\setminus B_r$, this time the transportation cost can be bound by:

$$\begin{aligned}\int_{B_R^c}|x-\pi_{\Gamma_{\bar s}}(x)|^p\mathrm{d}\rho_{\bar s} &= \int_{B_R^c}|x-\pi_{\Gamma_{\bar s}}(x)|^p\mathrm{d}(\rho_{\bar s}-\rho_r)+\int_{B_R^c}|x-\pi_{\Gamma_{\bar s}}(x)|^p\mathrm{d}\rho_r\\ &\leq\int_{B_R^c\times B_{\bar r}}|x-y|^p\mathrm{d}\gamma-p\frac{r}{2}\left(\frac{17}{20}R\right)^{p-1}\rho_r\left(B_R^c\right).\end{aligned}$$

We get:

$$W_p^p(\rho_0, \nu') \leq W_p^p(\rho_0, \nu) + 5rp\,(2R)^{p-1}\,\rho_{\bar{s}}\,(B_R) - p\frac{r}{2}\left(\frac{17}{20}R\right)^{p-1}\rho_r\,(B_R^c)\,.$$

As $\mathcal{L}(\nu') \leq \mathcal{L}(\nu)$, the optimality of ν gives that $W_p^p(\rho_0, \nu) \leq W_p^p(\rho_0, \nu')$, which, along with the previous estimates, implies

$$0 \leq 5 \cdot 2^{p-1}\rho_{\bar{s}}\,(B_R) - \frac{1}{2}\left(\frac{17}{20}\right)^{p-1}\rho_r\,(B_R^c) \;\Leftrightarrow\; \rho_r\,(B_R^c) \leq 10 \cdot \left(\frac{40}{17}\right)^{p-1}\rho_{\bar{s}}\,(B_R)\,.$$

On the other hand, since

$$\rho_r\,(B_R(\bar{x})^c) = \nu(B_r) - \rho_r(B_R(\bar{x})) \geq \alpha^{-1}Cr - \rho_r(B_R(\bar{x})) \geq \alpha^{-1}Cr - \rho_{\bar{s}}(B_R(\bar{x})),$$

and recalling (4.6) and (4.9), we deduce:

$$C \leq \left(1 + 10 \cdot \left(\frac{40}{17}\right)^{p-1}\right)(2L + \tfrac{C}{M})$$

We conclude that with the choice (4.4) of M, one has $C \leq 2ML$, which depends only on p and d and a contradiction follows if we choose $\bar{C} = 1 + 2ML$. □

Proof of Theorem 4.2. Consider $\bar{C}$, r_0 from Lemma 4.3. Fix $x \in \Sigma$ and assume there is $r \in (0, r_0)$ such that $\mathscr{H}^1(\Sigma \cap B_r(x)) \geq \bar{C}r$. Then the thesis of the lemma applies and it must hold that $\mathscr{H}^1(\Sigma \cap B_{2r}(x)) \geq 10\bar{C}r$. By induction, we find that for $k \geq 1$, one of the following holds:

- either $2^k r > r_0$;
- or we apply the lemma again, using that $\mathscr{H}^1(\Sigma \cap B_{2^k r}(x)) \geq 5^{k-1}\bar{C}(2^k r)$, and we get
$$\mathscr{H}^1(\Sigma \cap B_{2^{k+1}r}(x)) \geq 5^k\bar{C}(2^{k+1}r).$$

Let $k \geq 1$ be the first integer such that $2^k r > r_0$, so that $2^{k-1}r \leq r_0$ and

$$5^{k-1}\bar{C}(2^k r) \leq \mathscr{H}^1(\Sigma \cap B_{2^k r}(x)).$$

Hence, $r_0 \leq 2^k r \leq 5^{-k+1}\bar{C}^{-1}\mathscr{H}^1(\Sigma)$ and it holds that $k \leq k_0 \stackrel{\text{def.}}{=} \log_5(5\mathscr{H}^1(\Sigma)/\bar{C}r_0)$, and

$$r \geq r_0 2^{-k} \geq \bar{r}_0 \stackrel{\text{def.}}{=} r_0 \cdot 2^{-k_0}.$$

We find that if $r \leq \bar{r}_0$ then for $x \in \Sigma$, $\mathscr{H}^1(\Sigma \cap B_r(x)) \leq \bar{C}r$. □

Remark 4.4. *It is interesting to observe here that the regularity constant $\bar{C}$ depends only on d and p, while the scale $\bar{r}_0$ at which the Ahlfors-regularity holds gets smaller as ρ_0 gets more singular or when α (or $\mathscr{H}^1(\Sigma)$) increases.*

3. Absence of loops

Recall from Definition 2.14 of Chapter 2 that given a connected set Σ, $\Gamma \subset \Sigma$ is a loop if it is homeomorphic to $\mathbb{S}^1$, Σ is a tree if it has no loops. In this section we give a general strategy of proof for the absence of loops to solutions of problem ($\overline{P}_\Lambda$). We obtain this conclusion in the case that

$$\rho_0 = \mu_N \overset{\text{def.}}{=} \sum_{i=1}^{N} a_i \delta_{x_i} \text{ where } \sum_{i=1}^{N} a_i = 1,\ a_i > 0 \text{ for all } i = 1, \ldots, N, \tag{4.10}$$

but we believe it to be also true for $\rho_0 \in L^{\frac{d}{d-1}}$.

Throughout this section, we fix an optimal measure $\nu_\star$ for ($\overline{P}_\Lambda$) and set $\Sigma = \operatorname{supp} \nu_\star$. First we assume by contradiction that Σ contains a loop Γ. We will follow closely the argument from the existence proof carried throughout Section 5 of Chapter 3. The crucial result that enables this general proof strategy is given in Prop. 4.5 which states that loops are almost formed through projections onto Σ, and plays a similar result to Lemma 3.17 from Chapter 3.

3.1. Loops are formed through projections

Next we show that, under certain conditions on the measure ρ_0, loops must be formed through projections onto an optimal network Σ. In the course of the proof we will need to transport part of the measure ρ_0 with an arbitrary measurable selection of the projection operator

$$\Pi_\Sigma(x) = \operatorname*{argmin}_{y \in \Sigma} \frac{1}{2}|x-y|^2. \tag{4.11}$$

Therefore, we assume that

$$\text{there is a measurable selection } \pi_\Sigma \text{ of (4.11) that is } \rho_0\text{-a.e. uniquely defined.} \tag{4.12}$$

This is the case for instance if

- $\rho_0 \ll \mathcal{L}^d$, since the projection map is Lebesgue-a.e. single valued;
- ρ_0 is as in (4.10), since we can choose $y_i \in \operatorname*{argmin}_{\Sigma} |x_i - y|^2$ and set $\pi_\Sigma(x_i) \overset{\text{def.}}{=} y_i$.

Proposition 4.5. *Suppose that ρ_0 has a compact support and that (4.12) holds. Let $\nu_\star$ be a minimizer of ($\overline{P}_\Lambda$). If γ is an optimal transportation plan between ρ_0 and $\nu_\star$ and $\Gamma \subset \Sigma$ is a loop, then*

$$|x-y| = \operatorname{dist}(x,\Sigma) \text{ for } \gamma\text{-a.e. } (x,y) \in \mathbb{R}^d \times \Gamma.$$

Proof. Given $\eta > 0$, define the set

$$E_\eta \overset{\text{def.}}{=} \left\{(x,y) \in \mathbb{R}^d \times \Gamma :\ |x-y|^p > \operatorname{dist}(x,\Sigma)^p + \eta\right\}$$

and consider the measure ν_η defined for a Borel set A as

$$\nu_\eta(A) \stackrel{\text{def.}}{=} \gamma(E_\eta \cap (\mathbb{R}^d \times A)).$$

From its construction, it follows that $\nu_\eta \leq \nu_\star$. Therefore, to conclude it suffices to show that for any $\bar{y} \in \Sigma$, admitting an approximate tangent space $T_{\bar{y}}\Sigma = T_{\bar{y}}\Gamma$, it holds that

$$\theta_1(\nu_\eta, \bar{y}) = 0 \text{ for } \mathscr{H}^1\text{-a.e. } y \in \Gamma.$$

Let $(r_n)_{n\in\mathbb{N}}$ be an infinitesimal sequence obtained from Lemma 2.15 such that $\Sigma_n \stackrel{\text{def.}}{=} \Sigma \setminus B_{r_n}(\bar{y})$ remains connected. For n large enough, let us show that if

$$(x, y) \in E_\eta \cap (\mathbb{R}^d \times B_{r_n}(\bar{y})) \text{ then } \pi_\Sigma(x) \in \Sigma \setminus B_{r_n}(x).$$

Indeed, for such a pair (x, y) we have that

$$\begin{aligned}
\operatorname{dist}(x, \Sigma)^p + \eta \leq |x - y|^p &\leq (\operatorname{dist}(x, \Sigma) + |\pi_\Sigma(x) - y|)^p \\
&\leq \operatorname{dist}(x, \Sigma)^p + p(\operatorname{dist}(x, \Sigma) + |y - \pi_\Sigma(x)|)^{p-1}|y - \pi_\Sigma(x)| \\
&\leq \operatorname{dist}(x, \Sigma)^p + p(2\operatorname{diam}(\operatorname{supp}\rho_0))^{p-1}|y - \pi_\Sigma(x)|,
\end{aligned}$$

where the third inequality follows from the convexity of $t \mapsto |t|^p$. As a result, for n sufficiently large, we obtain that

$$2r_n < \frac{\eta}{p(2\operatorname{diam}(\operatorname{supp}\rho_0))^{p-1}} \leq |y - \pi_\Sigma(x)|.$$

Since $y \in B_{r_n}(\bar{y})$, it must follow that $\pi_\Sigma(x) \in \Sigma \setminus B_{r_n}(\bar{y})$, for n large enough.

In the sequel, we write $B_{r_n} = B_{r_n}(\bar{y})$ to simplify notation, and we define an alternative transportation plan as follows

$$\gamma' \stackrel{\text{def.}}{=} \gamma \llcorner \mathbb{R}^d \times \Sigma_n + (\pi_0, \pi_\Sigma \circ \pi_0)_\sharp \gamma \llcorner E_\eta \cap \mathbb{R}^d \times B_{r_n} + (\pi_0, y_n)_\sharp \gamma \llcorner \mathbb{R}^d \times B_{r_n} \setminus E_\eta, \quad (4.13)$$

where π_0, π_1 denote the projections onto the first and second marginal, *i.e* $\pi_0(x, y) = x$, and $y_n \in \Sigma_n \cap \partial B_{r_n}(\bar{y})$. Its second marginal then defines a new competitor as

$$\nu' \stackrel{\text{def.}}{=} \nu_\star \llcorner \Sigma_n + \nu_\eta \llcorner B_{r_n} + \gamma\left(\mathbb{R}^d \times B_{r_n} \setminus E_\eta\right)\delta_{y_n}. \quad (4.14)$$

The first term preserves the transportation plan that does not concern $\Sigma \cap B_{r_n}$, the second projects onto Σ all the mass that is sent to $\Sigma \cap B_{r_n}$, and the last term sends all the mass whose projection is close to $\Sigma \cap B_{r_n}$ to the point y_n, creating a Dirac mass at y_n.

Since the mass on the second term of the transportation plan γ' in (4.13) is sent to Σ_n, it follows that $\operatorname{supp}\nu' = \Sigma_n$. But since this operation can only increase the density of $\nu_\star$ over Σ_n, we have that $\nu' \llcorner \Sigma_n \geq \nu_\star \llcorner \Sigma_n$ and it follows that

$$\mathcal{L}(\nu_\star) \geq \mathcal{L}(\nu'). \quad (4.15)$$

This construction yields

$$\begin{aligned} W_p^p(\rho_0, \nu_\star) &= \int_{\mathbb{R}^d \times \Sigma_n} |x-y|^p \mathrm{d}\gamma + \int_{\mathbb{R}^d \times B_{r_n} \cap E_\eta} |x-y|^p \mathrm{d}\gamma + \int_{\mathbb{R}^d \times B_{r_n} \setminus E_\eta} |x-y|^p \mathrm{d}\gamma \\ &\geq \int_{\mathbb{R}^d \times \Sigma_n} |x-y|^p \mathrm{d}\gamma + \int_{\mathbb{R}^d \times B_{r_n} \cap E_\eta} (\mathrm{dist}(x,\Sigma)^p + \eta)\, \mathrm{d}\gamma \\ &+ \int_{\mathbb{R}^d \times B_{r_n} \setminus E_\eta} |x-y_n|^p \mathrm{d}\gamma - p \int_{\mathbb{R}^d \times B_{r_n} \setminus E_\eta} \underbrace{\big||x-y_n| - |x-y|\big|}_{\leq |y-y_n| \leq 2r_n} |x-y_n|^{p-1} \mathrm{d}\gamma \\ &\geq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \mathrm{d}\gamma' + \eta \nu_\eta(B_{r_n}) - 2pr_n \int_{\mathbb{R}^d \times B_{r_n} \setminus E_\eta} |x-y_n|^{p-1} \mathrm{d}\gamma, \end{aligned}$$

so that from the minimality of $\nu_\star$ and (4.15), the previous estimate gives

$$\frac{\nu_\eta(B_{r_n}(\bar{y}))}{2r_n} \leq \frac{p}{\eta} \int_{\mathbb{R}^d \times B_{r_n} \setminus E_\eta} |x-y_n|^{p-1} \mathrm{d}\gamma \xrightarrow[n\to\infty]{} 0.$$

We conclude that for all $\bar{y}$ that is a rectifiability point of Γ, it holds that $\theta_1(\nu_\eta, \bar{y}) = 0$, and the result follows. □

3.2. Localizations and blow-up

To begin our proof of absence of loops when ρ_0 is of the form (4.10), we suppose by contradiction that Σ contains a loop Γ. We first notice that, as a direct consequence of the characterization of solutions given in Thm 3.24 we have that in the case (4.10)

$$\nu_\star = \alpha^{-1} \mathscr{H}^1 \llcorner \Sigma + \sum_{i=1}^N b_i \delta_{x_i},$$

where $0 \leq b_i \leq a_i$ for all $i = 1, \dots, N$.

In order to adapt the scheme of proof from Section 5 from Chap. 3, our first step is to choose a suitable point of Γ to make localizations. Consider

$$y_0 \in \Gamma \setminus (x_i)_{i=1}^N, \text{ such that } T_{y_0}\Sigma = T_{y_0}\Gamma \text{ exists}, \tag{4.16}$$

and we set

$$0 < L \overset{\text{def.}}{=} \min_{i=1,\dots,N} |y_0 - x_i|.$$

In the sequel, let $(r_n)_{n\in\mathbb{N}}$ be a sequence of radii, obtained from Lemma 2.15, such that

$$\Sigma_n \overset{\text{def.}}{=} \Sigma \setminus B_{r_n}(y_0) \text{ is connected and } r_n \to 0, \tag{4.17}$$

and we assume without loss of generality that $(x_i)_{i=1}^N \subset B^c_{r_n}(y_0)$ for all $n \in \mathbb{N}$. We proceed by defining a sequence of localized measures

$$\nu_n \overset{\text{def.}}{=} \nu_\star \llcorner B_{r_n}(y_0),$$

and our goal is to define a suitable sequence of localized problems solved by ν_n. Letting γ be an optimal transportation plan from ρ_0 to the fixed solution $\nu_\star$, a natural candidate is to consider the problem of transporting the measure

$$\rho_n \overset{\text{def.}}{=} (\pi_0)_\sharp \left(\gamma \llcorner \mathbb{R}^d \times B_{r_n}(y_0)\right)$$

to all measures of the form $\nu' = \theta \mathscr{H}^1 \llcorner \Sigma'$ such that $(\Sigma \setminus B_{r_n}(y_0)) \cup \Sigma'$ remains connected.

Afterwards, we define a blow-up of this sequence of problems, as (3.41) in Chapter 3, and extract a limit. This is almost the strategy of proof from Section 5 of Chapter 3. As before, to prevent the measures from losing mass at infinity in the blow-up step, we need to let ρ_n follow a constant speed geodesic almost until it reaches ν_n.

Defining the maps $\pi_t \overset{\text{def.}}{=} t\pi_1 + (1-t)\pi_0$, where π_0 and π_1 corresponds to the projections onto the first and second marginals respectively, the curve

$$t \mapsto \lambda_t \overset{\text{def.}}{=} (\pi_t)_\sharp \gamma \text{ is a constant speed geodesic between } \rho_0 \text{ and } \nu_\star, \tag{4.18}$$

see [Santambrogio, 2015, Chap. 5]. Hence, we set

$$\sigma_n = (\pi_{1-r_n})_\sharp \gamma. \tag{4.19}$$

With these elements, we prove the next result, that is completely analogous to Lemma. 3.18 of Chap. 3, the difference is that before since we only used the excess mass, the connectedness constraints were not an issue; this time the choice of the radius r_n needed to be more careful in order to preserve connectedness.

Lemma 4.6. *The localized measure ν_n solves the following minimization problem*

$$\min \left\{ W_p^p(\rho_n, \nu') : \begin{array}{c} \Sigma' = \operatorname{supp} \nu' \subset \overline{B_{r_n}(y_0)} \\ \nu' \geq \alpha^{-1} \mathscr{H}^1 \llcorner \Sigma', \\ \Sigma' \textit{ and } \Sigma_n \cup \Sigma' \textit{ are connected}, \\ \nu'(\mathbb{R}^d) = \nu_\star(B_{r_n}(y_0)), \end{array} \right\} \tag{4.20}$$

It is also a minimizer for this problem with ρ_n replaced by σ_n. That is, for all ν' admissible for (4.20), it holds that

$$W_p^p(\sigma_n, \nu_n) \leq W_p^p(\sigma_n, \nu').$$

Proof. Let γ be the optimal transportation plan between ρ_0 and $\nu_\star$ and define the new transportation plan

$$\tilde{\gamma} \overset{\text{def.}}{=} \gamma \llcorner \mathbb{R}^d \times \Sigma_n + \gamma',$$

where γ' is optimal between ρ_n and ν'. Then the new competitor $\tilde{\nu} \overset{\text{def.}}{=} (\pi_1)_\sharp \tilde{\gamma}$ is such that $\mathcal{L}(\tilde{\nu}) \leq \mathcal{L}(\nu_\star)$, and the optimality of $\nu_\star$ gives that

$$\int_{\mathbb{R}^d \times \Sigma_n} |x-y|^p \mathrm{d}\gamma + \int_{\mathbb{R}^d \times B_{r_n}(y_0)} |x-y|^p \mathrm{d}\gamma \leq \int_{\mathbb{R}^d \times \Sigma_n} |x-y|^p \mathrm{d}\gamma + \int |x-y|^p \mathrm{d}\gamma'$$

Giving that $W_p^p(\rho_n, \nu_n) \leq W_p^p(\rho_n, \nu')$ for all ν' admissible.

To study the problem with ρ_n replaced by σ_n, we first recall that since σ_n follows the constant speed geodesic we have that

$$\begin{aligned} W_p(\rho_n, \sigma_n) + W_p(\sigma_n, \nu_n) = W_p(\rho_n, \nu_n) &\leq W_p(\rho_n, \nu') \\ &\leq W_p(\rho_n, \sigma_n) + W_p(\sigma_n, \nu'). \end{aligned}$$

Cancelling the term $W_p(\rho_n, \sigma_n)$, the result follows. □

In the sequel, recalling the definition of the blow-up operator $\Phi^{y_0,r} = \frac{\mathrm{id}-y_0}{r}$, notice that for any given measures μ, ν we have

$$W_p^p\left(\frac{1}{r}(\Phi^{y_0,r})_\sharp\mu, \frac{1}{r}(\Phi^{y_0,r})_\sharp\nu\right) = \frac{1}{r^{p+1}} W_p^p\left(\mu,\nu\right). \tag{4.21}$$

Define the sequences of blow-up probability measures as

$$\bar{\sigma}_n \stackrel{\text{def.}}{=} \frac{1}{\nu_\star(B_{r_n})}(\Phi^{y_0,r_n})_\sharp\sigma_n, \quad \bar{\nu}_n \stackrel{\text{def.}}{=} \frac{1}{\nu_\star(B_{r_n})}(\Phi^{y_0,r_n})_\sharp\nu_n. \tag{4.22}$$

From Lemma 4.6 and (4.21), each element from the sequence $(\bar{\nu}_n)_{n\in\mathbb{N}}$ is also a minimizer of a sequence of functionals $(F_n)_{n\in\mathbb{N}}$ defined as

$$F_n(\varrho) \stackrel{\text{def.}}{=} \begin{cases} W_p^p\left(\bar{\sigma}_n, \varrho\right), & \begin{array}{c} \varrho \in \mathscr{P}\left(\Sigma'\right), \Sigma' \subset \overline{B_1(0)} \\ \varrho \geq \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma', \\ \Sigma' \text{ and } \dfrac{\Sigma_n - y_0}{r_n} \cup \Sigma' \text{ are connected,} \end{array} \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.23}$$

Now, recall that from the blow-up properties of Σ, it follows that

$$\frac{\Sigma \cap B_{r_n}(y_0) - y_0}{r_n} \xrightarrow[n\to\infty]{d_H} [-\tau, \tau].$$

We can also extract a subsequence for the convergence of the measures, so that it holds that

$$\bar{\sigma}_n \xrightharpoonup[n\to\infty]{\star} \bar{\sigma}, \quad \bar{\nu}_n \xrightharpoonup[n\to\infty]{\star} \bar{\nu}. \tag{4.24}$$

This motivates the following limit problem, which is minimized by $\bar{\nu}$ as we shall prove later,

$$F(\varrho) \stackrel{\text{def.}}{=} \begin{cases} W_p^p\left(\bar{\sigma}, \varrho\right), & \begin{array}{c} \varrho \in \mathscr{P}\left(\Sigma'\right), \Sigma' \subset \overline{B_1(0)} \\ \varrho \geq \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma', \\ \Sigma' \text{ is connected,} \\ \Sigma' \cap \left(T_{y_0}\Sigma \setminus (-\tau,\tau)\right) \neq \emptyset \end{array} \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.25}$$

Theorem 4.7. *The sequence $(F_n)_{n\in\mathbb{N}}$, defined in (4.23), Γ-converges to F defined in (4.25).*

Proof. Let us start with the $\Gamma - \liminf$, so consider a sequence $(\varrho_n)_{n\in\mathbb{N}}$ converging in the narrow topology to ϱ, and such that $\liminf_{n\to\infty} F_n(\varrho_n) < +\infty$, so we can assume that for each $n \in \mathbb{N}$ there is a connected set Σ_n' such that

$$\Sigma_n' = \operatorname{supp} \varrho_n, \quad \Sigma_n' \subset \overline{B_1(0)}, \quad \alpha\varrho_n \geq \mathscr{H}^1 \llcorner \Sigma_n' \text{ and } \frac{\Sigma_n - y_0}{r_n} \cup \Sigma_n' \text{ is connected.}$$

From Blaschke's Theorem, we can assume that $\Sigma_n' \xrightarrow[n\to\infty]{d_H} \Sigma' \subset \overline{B_1(0)}$, since this sequence is contained in a compact subset of $\mathbb{R}^d$. As the Hausdorff limit of a sequence of connected sets, Σ' is itself connected and from Gołab's Theorem, ϱ is such that

$$\alpha\varrho \geq \mathscr{H}^1 \llcorner \Sigma' \text{ and } \varrho \in \mathscr{P}(\Sigma').$$

Finally, to prove that $(T_{y_0}\Sigma \cap (-\tau,\tau)) \cap \Sigma' \neq \emptyset$, consider a sequence

$$y_n \in \frac{\Sigma_n - y_0}{r_n} \cap \Sigma_n'.$$

One again, up to subsequences, it follows that $y_n \xrightarrow[n\to\infty]{} y$, but since

$$\frac{\Sigma_n - y_0}{r_n} \xrightarrow[n\to\infty]{K} T_{y_0}\Sigma \setminus (-\tau,\tau) \text{ and } \Sigma_n' \xrightarrow[n\to\infty]{d_H} \Sigma',$$

it follows that $y \in (T_{y_0}\Sigma \setminus (-\tau,\tau)) \cap \Sigma'$.

As a result, ϱ is in the domain of F and from the lower semi-continuity of the Wasserstein distance we get that

$$F(\varrho) = W_p^p(\bar\sigma, \varrho) \leq \liminf_{n\to\infty} W_p^p(\bar\sigma_n, \varrho_n) = \liminf_{n\to\infty} F_n(\varrho_n).$$

To prove the $\Gamma - \limsup$, let ϱ be such that $F(\varrho) < +\infty$, otherwise there is nothing to prove. To define a recovery sequence for one such measure, set $\Sigma' \overset{\text{def.}}{=} \operatorname{supp} \varrho$ and if Σ' touches $\dfrac{\Sigma_n - y_0}{r_n}$, we set $\varrho_n \overset{\text{def.}}{=} \varrho$.

Otherwise, let $y \in (T_{y_0}\Sigma \setminus (-\tau,\tau)) \cap \Sigma'$ and since $\dfrac{\Sigma_n - y_0}{r_n} \xrightarrow[n\to\infty]{} T_{y_0}\Sigma \setminus (-\tau,\tau)$, there is a sequence

$$y_n \in \frac{\Sigma_n - y_0}{r_n} \text{ such that } |y_n - y| \xrightarrow[n\to\infty]{} 0.$$

So we consider the translation operator $\mathcal{T}_n(x) \overset{\text{def.}}{=} x - (y_n - y)$ and define the measures

$$\varrho_n \overset{\text{def.}}{=} (\mathcal{T}_n)_\sharp \varrho.$$

This way, the new measures are such that

$$\operatorname{supp} \varrho_n = \Sigma_n' \overset{\text{def.}}{=} \mathcal{T}_n(\Sigma') \text{ is connected and } \alpha\varrho_n \geq \mathscr{H}^1 \llcorner \Sigma_n'.$$

In addition, $\dfrac{\Sigma_n - y_0}{r_n} \cup \Sigma_n'$ is also connected by construction so that $F_n(\varrho_n) < \infty$ for all $n \in \mathbb{N}$. To conclude, we must show that $\varrho_n \xrightarrow[n\to\infty]{} \varrho$. Indeed, if follows that

$$W_p^p(\varrho, \varrho_n) = \int_{\Sigma'} |x - \mathcal{T}_n(x)|^p \mathrm{d}\varrho = |y_n - y|^p \xrightarrow[n\to\infty]{} 0.$$

Since ϱ and ϱ_n are supported in $\overline{B_1(0)}$, the Wasserstein distance metrizes the narrow convergence, and we obtain that

$$\limsup_{n\to\infty} F_n(\varrho_n) = \limsup_{n\to\infty} W_p^p(\bar{\sigma}_n, \varrho_n) = W_p^p(\bar{\sigma}, \varrho) = F(\varrho),$$

and we conclude that F_n Γ-converges to F. □

We can now transfer a lot of information about the minimization of F_n to the minimization of F, by means of the Γ-convergence result and the fact that the optimal transportation in the definition of F_n is almost achieved via projections. In fact, only the transportation onto $\Gamma \cap B_{r_n}(y_0)$ is given by projections, and there might be some mass in the set $(\Sigma \setminus \Gamma) \cap B_{r_n}(y_0)$, but since Σ and Γ have the same approximate tangent space at y_0, this contribution vanishes as $n \to \infty$, and the limit inherits the projection properties from Γ.

Lemma 4.8. *For ρ_0 given by a sum of Dirac measures as in (4.10), the following assertions are true:*

(i) $\operatorname{supp} \bar{\sigma} \subset \{x : \operatorname{dist}(x, T_{y_0}\Sigma) \geq L\}$;

(ii) The measure $\bar{\nu} = \frac{1}{2}\mathscr{H}^1 \llcorner T_{y_0}\Sigma \cap B_1(0)$ and it is a minimizer of F;

(iii) the optimal transportation from $\bar{\sigma}$ to $\bar{\nu}$ is attained by the projection map onto $T_{y_0}\Sigma$.

Proof. Item (i) follows from the fact that, by construction, $|y_0 - x_i| \geq L$ for all $i = 1, \ldots, N$. To prove item (ii), it follows from the fact that $\nu_n = \alpha^{-1}\mathscr{H}^1 \llcorner \Sigma \cap B_{r_n}$ and the blow-up theorem for rectifiable measures that

$$\begin{aligned}
\bar{\nu} &\overset{\text{def.}}{=} \lim_{n\to\infty} \frac{1}{\nu_\star(B_{r_n}(y_0))} (\Phi^{y_0, r_n})_\sharp \nu_\Sigma \llcorner B_{r_n} \\
&= \lim_{n\to\infty} \underbrace{\frac{r_n}{\nu_\star(\Sigma \cap B_{r_n}(y_0))}}_{\to 1/2\theta(y_0)} \underbrace{\frac{1}{r_n} (\Phi^{y_0, r_n})_\sharp \nu_\star \llcorner \Sigma \cap B_{r_n}(y_0)}_{\to \theta(y_0)\mathscr{H}^1 \llcorner \Sigma \cap B_1(0)} \\
&= \frac{1}{2}\mathscr{H}^1 \llcorner \Sigma \cap B_1(0).
\end{aligned}$$

In addition, $\bar{\nu} \in \operatorname{argmin} F$ as the limit of minimizers of F_n, from the fundamental property of Γ-convergence.

Recall the sequences σ_n and $\nu_\star \llcorner B_{r_n}$, and let γ_n be the optimal transportation plan between them. From Prop. 4.5, it follows that

$$\operatorname{supp} \gamma_n \llcorner \mathbb{R}^d \times \Gamma \subset \operatorname{graph}(\Pi_\Sigma).$$

Since $\bar\sigma_n, \bar\nu_n$ are generated by the push-forward of σ_n and $\nu_\star \llcorner B_{r_n}$ by Φ^{y_0,r_n}, the optimal transportation between them in given by the plan

$$\bar\gamma_n \overset{\text{def.}}{=} \frac{1}{\nu_\star(B_{r_n})}\left(\Phi^{(y_0,y_0),r_n}\right)_\sharp \gamma_n, \text{ so that } \operatorname{supp}\left(\bar\gamma_n \llcorner \left(\mathbb{R}^d \times \frac{\Gamma - y_0}{r_n}\right)\right) \subset \operatorname{graph}\left(\Pi_{\frac{\Sigma - y_0}{r_n}}\right).$$

Up to a subsequence it converges to some $\bar\gamma$, which, by the stability of optimal transportation plans, also transports $\bar\sigma$ to $\bar\nu$ optimally, let us show that $\operatorname{supp}\bar\gamma \subset \operatorname{graph}\left(\Pi_{T_{y_0}\Sigma}\right)$. Notice that for any $A \subset \mathbb{R}^d$, we have that

$$\bar\gamma_n\left(A \times \left(\frac{(\Sigma \setminus \Gamma) \cap B_{r_n} - y_0}{r_n}\right)\right) \leq \frac{1}{\nu_\star(B_{r_n})}\nu_\star\left((\Sigma \setminus \Gamma) \cap B_{r_n}\right) = \frac{o(r_n)}{r_n} \xrightarrow[n\to\infty]{} 0,$$

since the tangent space of Σ and Γ coincide at y_0, from (4.16), and $\theta_1(\nu_\star, y_0) < +\infty$.

As a result, given $(x,p) \in \operatorname{supp}\bar\gamma$, there is an open ball B centered at (x,p) such that

$$0 < \bar\gamma(B) \leq \liminf_{n\to\infty} \bar\gamma_n(B) = \liminf_{n\to\infty} \bar\gamma_n\left(B \cap \left(\mathbb{R}^d \times \frac{\Gamma - y_0}{r_n}\right)\right).$$

In particular, we can find $\operatorname{supp}\bar\gamma_n \llcorner \left(\mathbb{R}^d \times \frac{\Gamma - y_0}{r_n}\right) \ni (x_n, p_n) \xrightarrow[n\to\infty]{} (x,p)$. So it holds that

$$|x - p| = \lim_{n\to\infty} |x_n - p_n| = \lim_{n\to\infty} \operatorname{dist}\left(x_n, \frac{\Sigma - y_0}{r_n}\right) = \operatorname{dist}(x, T_{y_0}\Sigma),$$

where the last equality comes from the point-wise convergence of the distance functions from Kuratowski convergence of blow-ups from Lemma 2.12. □

3.3. Better competitor and absence of loops

We now obtain a contradiction to the fact that the optimal set Σ contains a loop. Let us recall the construction done so far; if Σ is optimal for (P_Λ) such that it contains a loop Γ, we let y_0 be a flat non-cut point of Σ inside Γ. Then we can perform the construction done, via localizations around y_0, in the previous subsection and obtain the measures $\bar\sigma$ and $\bar\nu$, as in (4.24). From Lemma 4.8, the latter is a minimizer of the functional F defined in (4.25) and

$$\bar\nu = \frac{1}{2}\mathscr{H}^1 \llcorner T_{y_0}\Sigma \in \operatorname{argmin} F.$$

As the optimal transportation from $\bar\sigma$ to $\bar\nu$ is attained by the projection map onto $T_{y_0}\Sigma$, we use a refined version of the argument done in Section 5 to construct a strictly better

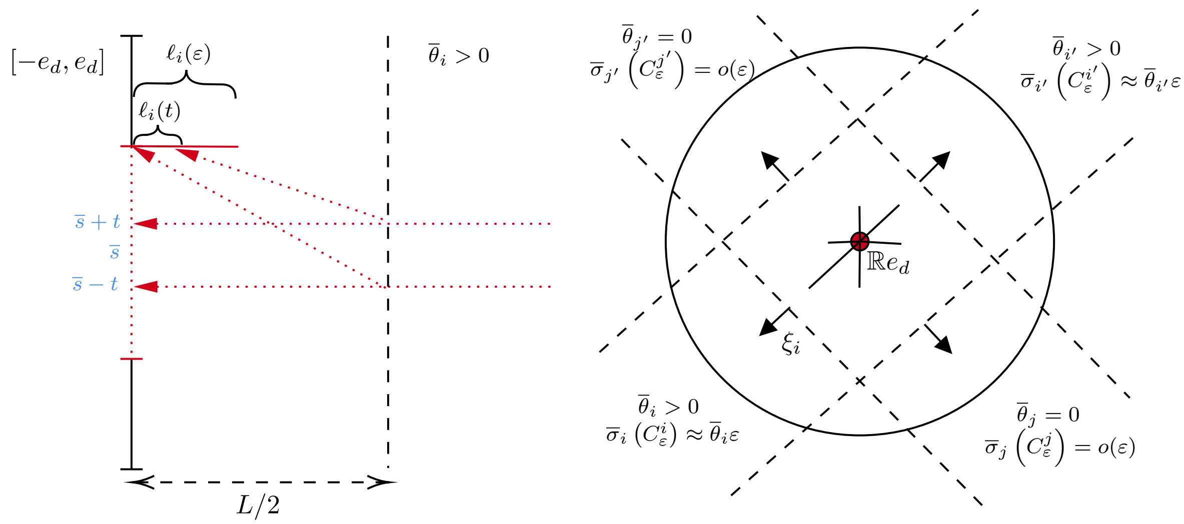

Figure 7: Construction of a better competitor in Lemma. 4.9. On the right, the partition of the space into sections. For sections i, i' such that $\bar\theta_i, \bar\theta_{i'} > 0$ we add a segment in their direction. For $\bar\theta_j, \bar\theta_{j'} = 0$ we construct a Dirac mass. On the cases of positive density we have a gain of order ε^2 in transportation cost, for zero density we lose $o(\varepsilon^2)$. On the left the transportation strategy of each section of the partitioned space.

competitor to F. The further complexity of this case stems from the fact that we must remove all the mass of a small segment and create an advantageous structure, see Figure 7. This construction will then contradict the existence of loops, so that any optimal Σ must be a tree.

Lemma 4.9. *If $\bar\sigma$ satisfies the thesis of Lemma 4.8, there exists a measure ν' such that $F(\nu') < F(\bar\nu)$.*

Proof. Suppose by contradiction that $\Sigma = \operatorname{supp}\nu_\star$ contains a loop, and let y_0 be a flat non-cut point inside this loop. Up to a rotation, we may assume that $T_{y_0}\Sigma = \mathbb{R}^d e_d$, where $(e_i)_{i=1}^d$ is a basis of $\mathbb{R}^d$. From item (i) of Lemma 4.8, $\bar\sigma$ is supported on $\{x = (x', x_d) \in \mathbb{R}^d : |x'| > L,\ |x_d| \le 1\}$, so we can cover its support with finitely many sets $(E_i)_{i=1}^N$ defined as:

$$E_i \overset{\text{def.}}{=} \{x = (x', x_d) \in \mathbb{R}^d : \langle \xi_i, x\rangle > L/2,\ |x_d| \le 1\}$$

where $\xi_i \in \mathbb{S}^{d-1} \cap [e_d]^\perp$ are unit vectors and N depends only on the dimension. We then define a disjoint family

$$F_1 = E_1, \quad F_{i+1} = E_{i+1} \setminus \bigcup_{j=1}^{i} F_i \text{ for } i \ge 1$$

and decompose our measures $\bar{\sigma}$ and $\bar{\nu}$ as

$$\bar{\sigma} = \sum_{i=1}^{N} \bar{\sigma}_i,\ \bar{\nu} = \sum_{i=1}^{N} \bar{\nu}_i \text{ where } \bar{\sigma}_i \overset{\text{def.}}{=} \bar{\sigma} \llcorner F_i \text{ and } \bar{\nu}_i \overset{\text{def.}}{=} (\mathrm{proj}_d)_\sharp \bar{\sigma}_i,$$

where $\mathrm{proj}_d : x \mapsto x_d e_d$ is the projection onto the vertical axis. By the Radon-Besicovitch's differentiation theorem, $\bar{\nu}_i = \theta_i \mathscr{H}^1 \llcorner [-e_d, e_d]$, where $\theta_i(s) = \theta_i(se_d) \geq 0$ are such that

$$\sum_{i=1}^{N} \theta_i(s) = \frac{1}{2}.$$

In the sequel, introduce the notation: $\mathbb{R}^d \ni x = (x_i, x_i'', x_d)$ where $x_i = \langle \xi_i, x \rangle$ is the component of x parallel to ξ_i and $x_i'' \in [\xi_i, e_d]^\perp$. Defining the sets

$$C_t^i \overset{\text{def.}}{=} F_i \cap \{x \in \mathbb{R}^d : |x_d - \bar{s}| \leq t\} \subset \left\{ x = (x_i, x_i'', x_d) : \begin{matrix} x_i > L/2, \\ |x_d - \bar{s}| \leq t \end{matrix} \right\},$$

and letting $\bar{s} \in (-1, 1)$ be a common Lebesgue point of all θ_i, $i = 1, \dots, N$, it follows from the fact that $(\mathrm{proj}_d)_\sharp \bar{\sigma}_i = \theta_i \mathscr{H}^1 \llcorner [-e_d, e_d]$ that, for every $i = 1, \dots, N$

$$\frac{\bar{\sigma}_i(C_\varepsilon^i)}{2\varepsilon} = \frac{1}{2\varepsilon} \int_{\bar{s}-\varepsilon}^{\bar{s}+\varepsilon} \theta_i(t) \mathrm{d}t \xrightarrow[\varepsilon \to 0]{} \theta_i(\bar{s}). \tag{4.26}$$

Consider now the two subfamilies of indexes

$$I_1 \overset{\text{def.}}{=} \{i :\ \theta_i(\bar{s}) > 0\}, \quad I_2 \overset{\text{def.}}{=} \{i :\ \theta_i(\bar{s}) = 0\}.$$

In particular, for each $i \in I_1$, there is a constant $\bar{\theta}_i > 0$ and $\varepsilon > 0$ such that for $t < \varepsilon$ we have

$$\frac{1}{\bar{\theta}_i} \leq \frac{\bar{\sigma}_i(C_t^i)}{t} \leq \bar{\theta}_i. \tag{4.27}$$

Now let us exploit the fact that, from Lemma 4.8 the optimal transport is given by projections to propose a new transportation map, sending the mass in C_ε^i to a segment pointing towards ξ_i:

$$\bar{T}(x) \overset{\text{def.}}{=} \begin{cases} \ell_i(|x_d - \bar{s}|)\xi_i + (\bar{s} + \varepsilon)e_d, & \text{if } x \in C_\varepsilon^i \text{ and } i \in I_1, \\ (\bar{s} + \varepsilon)e_d, & \text{if } x \in C_\varepsilon^i \text{ and } i \in I_2, \\ \mathrm{proj}_d(x), & \text{otherwise,} \end{cases}$$

where $\ell_i : [0, \varepsilon] \to \mathbb{R}_+$ is defined via the conservation of mass relation

$$\ell_i(t) = \alpha \bar{\sigma}_i(C_t^i). \tag{4.28}$$

In other words, the mass that was sent to the vertical segment $[\bar{s} - \varepsilon', \bar{s} + \varepsilon']e_d$ is now used to form the horizontal segments

$$L_i \overset{\text{def.}}{=} (\bar{s} + \varepsilon)e_d + [0, \ell_i(\varepsilon)]\xi_i,$$

for each $i \in I_1$. The mass corresponding to the remaining indexes form a Dirac measure concentrated in $(\bar{s}+\varepsilon)e_d$, but with a mass of order $o(\varepsilon)$.

Thanks to (3.46), the map $\bar{T}$ sends $\bar{\sigma}_i \llcorner C^i_\varepsilon$ to the measure $\alpha^{-1}\mathscr{H}^1 \llcorner L_i$, hence the transported measure $\bar{T}_\sharp\bar{\sigma}$ satisfies the constraints in the definition (4.25) of the limiting functional F, since the newly added structure, given by

$$\Sigma' = \bigcup_{i \in I_1} L_i,$$

is a connected set. As a result, one has that $F(\bar{T}_\sharp\bar{\sigma}) < +\infty$.

So for $i \in I_\star$ and $x \in C^i_\varepsilon$, recalling the notation $x = (x_i, x''_i, x_d)$, we have that

$$\begin{aligned}
|x - \mathrm{proj}_d(x)|^2 - |x - \bar{T}(x)|^2 &= x_i^2 + |x''_i|^2 - (x_i - \ell_i(|x_d - \bar{s}|))^2 - |x''_i|^2 - (x_d - \bar{s} - \varepsilon)^2 \\
&= 2x_i\ell_i(|x_d - \bar{s}|) - \ell_i(|x_d - \bar{s}|)^2 - (x_d - \bar{s})^2 + 2\varepsilon|x_d - \bar{s}| - \varepsilon^2 \\
&\geq 2\left(\frac{L}{\alpha\bar{\theta}_i} + \varepsilon\right)|x_d - \bar{s}| - \left(1 + (\alpha\bar{\theta}_i)^2\right)|x_d - \bar{s}|^2 - \varepsilon^2 \\
&\geq \frac{2L}{\alpha\bar{\theta}_i}|x_d - \bar{s}| - \left(2 + (\alpha\bar{\theta}_i)^2\right)\varepsilon^2,
\end{aligned}$$

This is a qualitative estimate on the difference of the squared distance, to extend it to the p-power, we use that for any $a, b \in \mathbb{R}$

$$a^{p/2} - b^{p/2} = \frac{p}{2}b^{\frac{p}{2}-1}(a-b) + o(a-b), \tag{4.29}$$

so that since $|x_d - \bar{s}| < \varepsilon$ for $x \in C^i_\varepsilon$ and $|x - \bar{T}(x)| > \frac{1}{2}$, taking $a = |x - \mathrm{proj}_d(x)|^2$ and $b = |x - \bar{T}(x)|^2$, we obtain for some constant C_p that

$$\begin{aligned}
|x - \mathrm{proj}_d(x)|^p - |x - \bar{T}(x)|^p &\geq C_p\left(|x - \mathrm{proj}_d(x)|^2 - |x - \bar{T}(x)|^2\right) + o(\varepsilon) \\
&\geq C_p(x_d - \bar{s}) + o(\varepsilon).
\end{aligned}$$

Notice that given $n_i \in \mathbb{N}$, to be fixed later, for any $x \in C^i_\varepsilon \setminus C_{\frac{\varepsilon}{n_i}}$ we have that $|x_d - \bar{s}| \geq \frac{\varepsilon}{n_i}$. Hence, integrating with respect to $\bar{\sigma}_i$ over C^i_ε yields

$$\begin{aligned}
\int_{C^i_\varepsilon}\left(|x - \mathrm{proj}_d(x)|^p - |x - \bar{T}(x)|^p\right)\mathrm{d}\bar{\sigma}_i &\geq C_p\int_{C^i_\varepsilon\setminus C^i_{\frac{\varepsilon}{n_i}}}|x_d - \bar{s}|\mathrm{d}\bar{\sigma}_i + o(\varepsilon^2) \\
&\geq C_p\frac{\varepsilon}{n_i}\bar{\sigma}_i\left(C^i_\varepsilon \setminus C^i_{\frac{\varepsilon}{n_i}}\right) + o(\varepsilon^2) = C_p\frac{\varepsilon}{n_i}\left(\bar{\sigma}_i\left(C^i_\varepsilon\right) - \bar{\sigma}_i\left(C^i_{\frac{\varepsilon}{n_i}}\right)\right) + o(\varepsilon^2) \\
&\geq \frac{C_p}{n_i}\left(\frac{1}{\bar{\theta}_i} - \frac{\bar{\theta}_i}{n_i}\right)\varepsilon^2 + o(\varepsilon^2) \geq \frac{C_p}{2\bar{\theta}_i n_i}\varepsilon^2 + o(\varepsilon^2),
\end{aligned}$$

where in the last inequality we choose $n_i \geq 2\bar{\theta}_i^2$.

For the indexes $i \notin I_2$, we observe that the error committed by using the map $\bar{T}$ is given by $|x - \mathrm{proj}_d(x)|^2 - |x - \bar{s}e_d|^2 = -(x_d - \bar{s})^2 \geq -\varepsilon^2$. So using once again (4.29) we get that

$$|x - \mathrm{proj}_d(x)|^p - |x - \bar{s}e_d|^p \geq -C_p\varepsilon^2 + o(\varepsilon^2).$$

Now setting $\nu' \stackrel{\text{def.}}{=} \bar{T}_\sharp \bar{\sigma}$, we obtain that

$$\begin{aligned}
W_p^p(\bar{\sigma}, \bar{\nu}) - W_p^p(\bar{\sigma}, \nu') &\geq \int \left(|x - \operatorname{proj}_d(x)|^p - |x - \bar{T}(x)|^p \right) \mathrm{d}\bar{\sigma} \\
&= \sum_{i=1}^{N} \int_{C_\varepsilon^i} \left(|x - \operatorname{proj}_d(x)|^p - |x - \bar{T}(x)|^p \right) \mathrm{d}\bar{\sigma}_i \\
&\geq C_p \left(\sum_{i \in I_1} \left(\frac{1}{2\bar{\theta}_i n_i} \varepsilon^2 + o(\varepsilon^2) \right) - \sum_{i \in I_2} (\varepsilon^2 + o(\varepsilon^2)) \bar{\sigma}_i(C_\varepsilon^i) \right) \\
&= C_p \varepsilon^2 \left(\sum_{i \in I_1} \left(\frac{1}{2\bar{\theta}_i n_i} + \frac{o(\varepsilon^2)}{\varepsilon^2} \right) - \sum_{i \in I_2} \left(1 + \frac{o(\varepsilon^2)}{\varepsilon^2} \right) \bar{\sigma}_i(C_\varepsilon^i) \right) > 0
\end{aligned}$$

for ε small enough since $\bar{\sigma}_i(C_\varepsilon^i) = o(\varepsilon)$, for each $i \in I_2$. But as the new competitor ν' is admissible for the minimization of F, we obtain a contradiction with the fact that $\bar{\nu}$ is a minimizer from Lemma 4.8. □

Using this Lemma and the properties of solutions of for ρ_0 given by a sum of Dirac masses as in (4.10), we have proved the following result.

Theorem 4.10. *If ρ_0 given in (4.10), then the support of any minimizer to ($\overline{P}_\Lambda$) is a tree.*

4. Discussion

We have already seen in the last Chapter that it can be really hard to produce variations for the Wasserstein-$\mathscr{H}^1$ problem (P_Λ). This is essentially because of the sharp density constraints of the measures $\nu_\Sigma = \frac{1}{\mathscr{H}^1(\Sigma)} \mathscr{H}^1 \llcorner \Sigma$ must satisfy, *i.e.* have a constant density. The introduction of the length functional serves as a way of relaxing these constraints from equality to inequality, making it much more manageable. However, from the characterization **??** the term $\mathcal{L}$ is still an L^∞ energy, and hence notoriously difficult to systematically craft variations. This is the reason we have not succeeded to derive a satisfactory *Euler-Lagrange equation* for problems (P_Λ) or ($\overline{P}_\Lambda$).

Nevertheless, it has been shown that we can extract a lot of information for these problems by hand-crafting variations of ($\overline{P}_\Lambda$) that are well suited for the property one wishes to show. The density constraint, although easier to deal with in the relaxed formulation, requires a very fine control of "who goes where" in the global transportation of mass. In this sense, we can group the strategies of proof from the two last Chapters by how we can exploit a prior knowledge of the distribution of mass:

1. In the proofs of L^∞ bounds and Ahlfors regularity, Thm 3.20 of Chap. 3 and Thm. 4 of the present one, we are able to identify a region that contains a lot of mass that travels a very long distance to the optimal network Σ. In these cases, the strategy of projecting the mass coming from far away into a newly added network seems to be a promising route. The design of this new structure is then very dependent of the

geometry of the disposition of mass, whereas in Thm 3.20 we knew it came from outside of a tubular region of the optimal Σ, in Thm 4.2, we used the integrability of ρ_0 to infer that the mass came from the complement of a ball with very high radius.

2. In the proofs of existence of solutions and absence of loops, Thm. 3.24 and Thm. 4.10, we identify a part of the mass of an optimal measure ν that is formed through projections onto its support, 3.17 and 4.5. We then manage to pass these properties to the blow-up of a limit problem and exploit this to create better competitors. One could conjecture that this strategy is behind a general principle, that non-optimal structures of minimizers, such as loops, should not happen, and if they do, they must be formed though projections. The contradiction can then be obtained by means of the assumptions made on ρ_0.

There are many qualitative properties of minimizers that are left as open questions.

- The presentation we have chosen for the proof of absence of loops in the point cloud case is very flexible and should also be applicable in other cases. Notice that since the projection principle onto loops from Prop. 4.5 is would still works for $\rho_0 \ll \mathcal{L}^d$, the proof still works for this case, except the property $\operatorname{supp} \bar{\sigma} \subset \{\operatorname{dist}(\cdot, T_{y_0}\Sigma) \geq L\}$ from Lemma 4.8. This raises the question if the proof can be adapted to this case, or even assuming more integrability for ρ_0, for instance $L^{\frac{d}{d-1}}$ in order to exploit the Ahlfors regularity proved for this case.

- What can we say about the topology of an optimal network? We can give some sufficient conditions for it to be a tree, but can we say something about the topology of bifurcations? The Ahlfors regularity indicates that the number of bifurcations at any given point is globally bounded, but could we expect the problem (P_Λ) to present only triple bifurcations like some of its counterparts, as the Steiner problem and average distance minimizers problem [Lemenant, 2010]?

- Once the topology of bifurcations is identified the natural question is if one can also explicitly characterize the blow-ups of any point in the network, as done by Bonnet for the Mumford-Shah problem [Ambrosio et al., 2000, Thm. 6.10] and by Santambrogio-Tilli in [Santambrogio and Tilli, 2005] for the average distance minimizers problem.

- A much harder problem would be to prove the $C^{1,\alpha}$ regularity of an optimal network Σ. A first naive approach would be trying to verify Morgan's criterion [Morgan, 1994] by producing competitors of original problem (P_Λ). However, if we can expect this problem to behave similarly to the average distance minimizer problem, as we argue in the following Chapter, its regularity theory should be much more nuanced than the thesis of Morgan's criterion.

Chapter 5

Phase field approximation for 1D variational problems

Contents

1. Introduction

In this work we propose a unified phase field approximation for the Wasserstein-$\mathscr{H}^1$ problem, introduced in Chapter 3, and the average distance minimizers problem, for a review see [Lemenant, 2010], in any dimension. For facility of reference throughout the chapter, we recall these problems in the sequel and give the precise framework we will be interested in the present chapter.

Henceforth, we let Ω be a compact and connected subset of $\mathbb{R}^d$ with nonempty interior and Lipschitz boundary. Given $\rho_0 \in \mathscr{P}(\Omega)$, we recall that the Wasserstein-$\mathscr{H}^1$ problem from Chapter 3 consists of finding the best approximation of ρ_0 among measures that are uniformly distributed over a 1-dimensional set Σ. It is given by the following variational problem

$$\inf_{\Sigma \text{ connected}} W_q^q\left(\rho_0, \frac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma\right) + \Lambda \mathscr{H}^1(\Sigma). \tag{$W\mathscr{H}^1$}$$

Here W_q denotes the q-Wasserstein distance between two probability measures, and $\mathscr{H}^1$ denotes the 1-Hausdorff measure. [1]

Without the regularization term, this problem would be trivial as one could find a space-filling curve that makes the Wasserstein term converge to zero. On the other hand, the problem would also be trivial without the connectedness constraints, as one could approximate the measure ρ_0 with an empirical measure, paying nothing for the length term and making the Wasserstein distance arbitrarily small. It is proven in Chapter 3, see Theorem 3.24, that this problem has a solution once Λ is sufficiently small and ρ_0 is smooth enough (does not give mass to 1D sets).

The average distance minimizers problem, introduced in [Buttazzo and Stepanov, 2003], is defined as

$$\inf_{\Sigma \text{ connected}} \int_\Omega \operatorname{dist}(x,\Sigma)^q \mathrm{d}\rho_0(x) + \Lambda \mathscr{H}^1(\Sigma). \tag{ADM}$$

Now the measure ρ_0 represents a distribution of population over the domain Ω and the integral term models the average distance of this population to a transportation network Σ. The minimizers of (ADM) can then be interpreted as the best possible transportation network, in the sense that the average individual, distributed with law ρ_0, is closest to the transportation network.

As was pointed out in [Bonnivard et al., 2015], the average distance functional can be expressed as

$$\int_\Omega \operatorname{dist}(x,\Sigma)^q \mathrm{d}\rho_0(x) = \inf_{\operatorname{supp}\nu\subset\Sigma} W_q^q(\rho_0,\nu),$$

so the average distance minimizers problem can be rewritten as

$$(\text{ADM}) \equiv \inf_{\Sigma \text{ connected}} \inf_{\operatorname{supp}\nu\subset\Sigma} W_q^q(\rho_0,\nu) + \Lambda\mathscr{H}^1(\Sigma).$$

[1] The change from p to q in the Wasserstein distance is motivated by the introduction in the present Chapter of the p-Ambrosio Tortorelli functional. Due to its importance in this section of the work, we have decided to slightly change the notation.

This is the key observation that will allow us to propose a unified approach to study both problems.

The relaxation of ($W\mathscr{H}^1$) is given by

$$\inf_{\nu\in\mathscr{P}(\Omega)} W_q^q(\rho_0,\nu)+\Lambda\mathcal{L}(\nu), \tag{$\overline{W\mathscr{H}^1}$}$$

where $\nu\mapsto\mathcal{L}(\nu)$ is the *length functional* defined as

$$\mathcal{L}(\nu)\overset{\text{def.}}{=}\inf\left\{\alpha\geq 0:\alpha\nu\geq\mathscr{H}^1\llcorner\operatorname{supp}\nu\right\}, \tag{5.1}$$

which, from Prop. 3.6, is the l.s.c. relaxation of the functional $\nu\mapsto\mathscr{H}^1(\Sigma)$ if ν is the probability measure uniformly distributed over a connected set Σ, *i.e.* $\nu=\dfrac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1\llcorner\Sigma$, and $+\infty$ otherwise. As a consequence, one can rewrite the relaxation as a minimization in three variables: the measure ν, its support Σ and a new scalar variable α that measures the saturation of the density constraints $\alpha\nu\geq\mathscr{H}^1\llcorner\Sigma$:

$$(\overline{W\mathscr{H}^1})\equiv\inf_{\Sigma\text{ connected}}\ \inf_{\substack{\alpha\nu\geq\mathscr{H}^1\llcorner\Sigma\\ \operatorname{supp}\nu=\Sigma}} W_q^q(\rho_0,\nu)+\Lambda\alpha. \tag{5.2}$$

Under the assumptions for existence to ($W\mathscr{H}^1$), the optimal α is given by $\mathscr{H}^1(\Sigma)$ and problem ($\overline{W\mathscr{H}^1}$) can formally be seen as (ADM) with additional density constraints.

Problems ($W\mathscr{H}^1$) and (ADM) fall in the category of 1D shape optimization and are notoriously difficult to solve numerically in general. Perhaps the most famous of them is the Steiner tree problem [Brazil et al., 2014]. It has many modern reformulations [Paolini and Stepanov, 2013], one of which can be stated in the language of geometric measure theory as follows: given some Borel set K, we let $\mathscr{H}^1_S(K)$ denote the length of the minimal Steiner tree that connects K, therefore being defined as

$$\mathscr{H}^1_S(K)\overset{\text{def.}}{=}\inf\left\{\mathscr{H}^1(\Sigma):\quad K\subset\Sigma\text{ and }\Sigma\text{ is connected}\right\}, \tag{5.3}$$

and we let $\mathcal{S}(K)$ denote some tree that attains this value. From [Paolini and Stepanov, 2013], the above minimization has a solution with possibly infinite length. In its original formulation, where K is a discrete set of points in $\mathbb{R}^2$, it can be proven that any optimal network is made of finitely many segments connected by triple junctions forming 120 degrees. It can therefore be seen as a combinatorial problem and is one of Karp's original NP-hard problems [Karp, 1972]. In the computer science and combinatorial optimization communities the natural approach to solving this type of problems is to resort to heuristic methods [Voß, 1992], and even in the calculations of variations this approach has been exploited in [Alberti et al., 2019].

Another popular approach, which is variational in nature and indeed the one we shall adopt, is to resort to *phase field approximations*, that is a family of Sobolev functions whose level sets are good approximations of the set of small dimension we wish to approximate. The idea, originally from Modica and Mortola in [Modica and Mortola, 1977] to study the Cahn-Hillard equations, and later of Ambrosio and Tortorelli in [Ambrosio, 1992, Ambrosio et al., 2000] for the Mumford-Shah problem [Lemenant, 2016, Fusco, 2003], is to find a family of elliptic functionals Γ-converging ([Braides, 2002]) to the functional one wishes to minimize.

The Ambrosio-Tortorelli functional, see for instance the monograph [Braides, 1998],

$$\mathcal{AT}_p(\varphi_\varepsilon) = \frac{1}{\Lambda_{p,d}} \int_\Omega \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathrm{d}x$$

approximates formally the quantity $\mathscr{H}^1(\{\varphi = 0\})$, where $\Lambda_{p,d}$ is a renormalization constant, which will be obtained from an auxiliary variational problem see (5.10) later on, and p' is the conjugate exponent of p *i.e.* $\dfrac{1}{p} + \dfrac{1}{p'} = 1$. However, it does not penalize the connectedness of the set $\{\varphi = 0\}$. For more information on this functional see the discussion at the end of this Section and subsection 2.1.

In the original works from Ambrosio and Tortorelli about the Mumford-Shah functional, see [Ambrosio et al., 2000, Chap. 6], the elliptic integrand was of the form

$$\varepsilon |\nabla \varphi_\varepsilon|^2 + \frac{(1-\varphi_\varepsilon)^2}{\varepsilon},$$

which coincides with ours if $p = d = 2$. We must emphasize that in the d-dimensional case, the Ambrosio-Tortorelli functional actually is meant to approximate $\mathcal{H}^{d-1}(J_u)$, where J_u is the jump set of an SBV function u. This way, the exponent $-(d-1)$ is meant to be the codimension of the type of structures we wish to approximate. This idea was proposed and further exploited for more general codimensions in [Chambolle et al., 2019a].

Actually, the connectedness constraint in ($W\mathscr{H}^1$)(ADM) is the hardest to deal with phase field approximations, having only recently being treated in [Bonnivard et al., 2015, Chambolle et al., 2019b, Stuhmer et al., 2013], where the authors' strategy to impose connectedness was to explore properties of the solutions, for instance a priori knowledge that certain points belong to the optimal networks. As we don't have such a priori knowledge for ($W\mathscr{H}^1$) and (ADM), our strategy to impose such constraints will be to employ the connectedness functional proposed in [Dondl and Wojtowytsch, 2021] and later used to study several problems. For instance in [Dondl et al., 2017] it was employed to minimize a variation of the Willmore energy with connectedness constraints, in [Dondl et al., 2018] it is used to propose a phase-field approximation of the connected perimeter and in [Dondl et al., 2021] it is used in the study of a liquid drop model with Coulomb type interaction. Their approach was to define the so-called *diffuse connectedness functional* as

$$\mathcal{C}_\varepsilon(\varphi_\varepsilon) \overset{\text{def.}}{=} \int_{\Omega\times\Omega} \beta_\varepsilon(\varphi_\varepsilon(x))\beta_\varepsilon(\varphi_\varepsilon(y)) d^{F_\varepsilon \circ \varphi_\varepsilon}(x,y) \mathrm{d}x\mathrm{d}y, \tag{5.4}$$

where $d^{F_\varepsilon\circ\varphi_\varepsilon}$ is a geodesic distance that penalizes the part of the path between its endpoints outside the level set $\{\varphi_\varepsilon \leq \varepsilon^s\}$, where s is a parameter that allows us to control the thickness of the transition regions from 0 to 1 of the optimal phase fields. The function β_ε is then designed to select only this level set on the integration over $\Omega \times \Omega$, for further details see Section 2.

The diffuse approximation results, in the sense of Γ-convergence, that we prove in this work are formulated with respect to the following functionals:

$$\mathcal{AD}_\varepsilon(\nu_\varepsilon, \varphi_\varepsilon) \overset{\text{def.}}{=} \begin{cases} W_q^q(\rho_0, \nu_\varepsilon) + \Lambda \mathcal{AT}_p(\varphi_\varepsilon) + \dfrac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) & \nu_\varepsilon \in \mathscr{P}(\Omega), \\ \qquad\qquad + \dfrac{1}{\varepsilon^\ell}\displaystyle\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon, & \varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega) \\ +\infty, & \text{otherwise,} \end{cases} \tag{5.5}$$

the *diffuse average distance* functional, and

$$\mathcal{WH}^1_\varepsilon(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon) \overset{\text{def.}}{=} \begin{cases} W_q^q(\rho_0, \nu_\varepsilon) + \Lambda\alpha_\varepsilon + \dfrac{1}{\varepsilon}\left\|\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon\right\|^2_{L^2(\Omega)} & \alpha_\varepsilon \geq 0, \\ & \nu_\varepsilon \in \mathscr{P}(\Omega) \\ \qquad + \dfrac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) + \dfrac{1}{\varepsilon^\ell}\displaystyle\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon, & \varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega), \\ +\infty, & \text{otherwise.} \end{cases} \tag{5.6}$$

where the measure $\mu_\varepsilon = \mu_\varepsilon(\varphi_\varepsilon)$ is the diffuse transition measure and is defined as

$$\mu_\varepsilon \overset{\text{def.}}{=} \frac{1}{\Lambda_{p,d}}\left(\frac{\varepsilon^{p-d+1}}{p}|\nabla\varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'}(1-\varphi_\varepsilon)^2\right)\mathcal{L}^d \llcorner \Omega, \tag{5.7}$$

where $\Lambda_{p,d}$ depends only on p and d, see Section 2 for further details and properties about these measures and, in particular the variational interpretation of this constant.

Before stating our results, we make the following hypothesis that will be assumed without statement throughout this work.

(H1) $\Omega \subset \mathbb{R}^d$ is a compact, connect set with Lipschitz boundary and such that $\overline{\mathrm{int}\Omega} = \Omega$, and it is *star-shaped* that is there exists $x_\star \in \mathrm{int}\,\Omega$ such that $\lambda x_\star + (1-\lambda)x \in \mathrm{int}\,\Omega$ for any $x \in \Omega$ and $\lambda \in (0,1)$.

Hypothesis (H1) is to prevent the loss of mass, due to concentration on the boundary, while passing to the limit in the weak-$\star$ topology of probability measures.

The first result concerns the approximation for the average distance minimizers problem, in the spirit of the results found in [Bonnivard et al., 2015] for instance. The difference of our approach is that, due to the diffuse connectedness functional, we do not need the a priori knowledge that the optimal network contains any specific point.

Theorem 5.1. *Assume that $\ell > s > 1$, $\kappa > (2d+1)\frac{(s+1)p-d+1}{p-d}$ and that $p > d \geq 2$. Then it holds that*

$$\mathcal{AD}_\varepsilon \xrightarrow[\varepsilon\to 0]{\Gamma} \mathcal{AD}(\nu, \varphi) \overset{\text{def.}}{=} \begin{cases} W_q^q(\rho_0, \nu) + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu), & \nu \in \mathscr{P}(\Omega),\ \varphi \equiv 1, \\ +\infty, & \textit{otherwise,} \end{cases}$$

where $\mathscr{H}^1_S(\operatorname{supp}\nu)$ *is the length of the minimal Steiner tree connecting* $\operatorname{supp}\nu$*, defined in* (5.3)*. The* Γ*-convergence holds in the strong topology of* L^2 *and weak-*$\star$ *topology of* $\mathscr{P}(\Omega)$*.*

In addition, let $(\nu_\varepsilon, \varphi_\varepsilon)_{\varepsilon>0}$ *be a family of minimizers of* $\mathcal{AD}_\varepsilon$*, it admits a cluster point* $(\nu, \varphi \equiv 1)$*, which then achieves the infimum and*

$$\min_{\Sigma}\,(\text{ADM}) = \min_{(\nu,\varphi)} \mathcal{AD}(\nu, \varphi),$$

and it holds that

- Σ *is a minimizer of* (ADM) *if, and only if, it is a minimal Steiner tree of* $\operatorname{supp}\nu$*, for some* ν *minimizer of* $\mathcal{AD}$*;*
- ν *is a minimizer of* $\mathcal{AD}$ *if, and only if, it can be written as* $\nu = (\pi_\Sigma)_\sharp\rho_0$*, where* π_Σ *is a measurable selection of the projection operator onto some* Σ *minimizer of* (ADM)*.*

It is important to point out that, given an optimal network Σ for (ADM), the measure $\nu = (\pi_\Sigma)_\sharp\rho_0$, that is the minimizers of $\mathcal{AD}$, carries important information about the topology of Σ. It was shown in [Buttazzo and Stepanov, 2003, Santambrogio and Tilli, 2005], that points $y \in \Sigma$ such that $\nu(\{y\}) > 0$ are either end-points or corner points of Σ, see also the survey [Lemenant, 2012]. Therefore, the approximation we propose carries the information of the optimal network though the level sets of phase fields, and of the expected topology, though the approximations of the measure ν.

Our second Γ convergence result concerns the relaxed problem ($\overline{W\mathscr{H}^1}$). Since the energy in ($W\mathscr{H}^1$) is not l.s.c., as seen as a functional in $\mathscr{P}(\Omega)$, we cannot hope to prove a Γ-convergence result for it, since Γ limits always are l.s.c. in the topology inducing the Γ convergence. What we strive instead, is to approximate the relaxed problem, so that under the assumptions on ρ_0 that guaranties existence for ($W\mathscr{H}^1$), cf. Chapter 3, any cluster points of minimizers of $\mathcal{WH}^1_\varepsilon$ will also minimize the original problem ($W\mathscr{H}^1$).

Theorem 5.2. *Assume that* $\ell > s > 1$*,* $\kappa > (2d+1)\frac{(s+1)p-d+1}{p-d}$ *and* $p > d \geq 2$*. Then it holds that*

$$\mathcal{WH}^1_\varepsilon \xrightarrow[\varepsilon\to 0]{\Gamma} \overline{\mathcal{WH}^1}(\alpha,\nu,\varphi) \overset{\text{def.}}{=} \begin{cases} W_q^q(\rho_0,\nu) + \Lambda\mathcal{L}(\nu), & \nu \in \mathscr{P}(\Omega),\ \alpha \geq \mathcal{L}(\nu), \\ +\infty, & \textit{otherwise}, \end{cases}$$

the Γ*-convergence being held in* $\mathbb{R}$*, the strong topology of* L^2 *and weak-*$\star$ *topology of* $\mathscr{P}(\Omega)$*.*

In addition, whenever ρ_0 *does not charge countably* $\mathscr{H}^1$*-rectifiable sets, if* $(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon)_{\varepsilon>0}$ *is a sequence of minimizers of* $\mathcal{WH}^1_\varepsilon$*, then it has a cluster point* $(\alpha, \nu, \varphi \equiv 1)$ *of the form*

$$\alpha = \mathscr{H}^1(\Sigma),\ \nu = \frac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma,\ \textit{where } \Sigma \textit{ is connected } \mathscr{H}^1\textit{-rectifiable},$$

so that Σ *minimizes* ($W\mathscr{H}^1$)*.*

A few remarks about Theorems 5.1 and 5.2 and their proofs are in order. First of all, the formal relation between ($\overline{W\mathscr{H}^1}$) and (ADM) becomes more evident from the proposed phase field approximations. Indeed, the proofs of our Γ-convergence approximations only differ on how we deal with the support of the measures ν. In the average distance minimizers problem we do not need to control it as much as in the Wasserstein-$\mathscr{H}^1$ problem, since in the former we only need the support to be contained in a 1-dimensional set (the support doesn't even need to be rectifiable), in the latter we must distribute a minimal amount of mass everywhere.

The lower bound for κ is not very encouraging for numerics since the quantity $\varepsilon^{-\kappa}$ can very quickly exceed machine precision with not so small values for ε. In other models, for instance the connected Willmore energy, Γ convergence results have been achieved with $\kappa = 2$, see [Dondl et al., 2017]. Since in our problems the phase fields approximate 1-dimensional structures, instead of sets of finite perimeter as in [Dondl et al., 2017, Dondl et al., 2018], it is expected the value for κ in our problems to be larger. That said, the argument in Theorem 5.10 is probably not optimal and there might be another argument that gives a smaller bound for κ. However, in practice numerical experiments have shown to work with different constants penalizing $\mathcal{C}_\varepsilon$, [Dondl et al., 2018].

Finally, from a numerical point of view, the case $p = 2$ in the Ambrosio-Tortorelli term is much more convenient. However, in this work we need to assume

$$p > d \text{ in the functional } \mathcal{AT}_p. \tag{5.8}$$

This will imply that phase-fields with finite energy belong in the space $W^{1,p}(\Omega)$, and from standard Sobolev injections they must be Hölder continuous. Not only the enhanced regularity is of paramount importance in controlling the small level sets of the phase fields, synergizing well with the connectedness functional in the general $\Gamma - \liminf$ inequality, see Thm. 5.10, but it also helps in the matter of existence of solutions for the sequence of approximated problems.

In [Bonnivard et al., 2015], the matter of existence was solved by adding a penalization of $\|\nabla\varphi_\varepsilon\|_{L^p(\Omega)}$, having the same effect, but possibly affecting the numerics, as it was only a parasite term in the minimization and not contributing to the approximation of the length. In [Bonnivard et al., 2018, Bonnivard et al., 2020], to approximate the Steiner tree problem, the question of existence was dealt with by means of a regularization with a term reminiscent of the Willmore energy [Willmore, 1992, Willmore, 1993], forcing phase fields to be in $W^{2,2}(\Omega)$. A disadvantage of our approach is that the first variation of our energy will have a p-Laplacian term. On the other hand, computing variations for the Willmore energy require the solution of a fourth order PDE. It is not clear then which approach would be more computationally demanding and extensive numerical experiments and testing are in order.

In addition to the aforementioned reasons, for the case $p > d = 2$ case, we shall also see in Proposition 5.3 and Theorem 5.6 that it provides a better optimal profile, in the sense that its transition width is of the order $\frac{p}{p-2}\varepsilon$, as opposed to $\varepsilon \log \varepsilon$ in the case $p = 2$, see [Bonnivard et al., 2015]. As a result, one can expect sharper transitions with $p > 2$, and therefore, better qualitative results. This intuition is corroborated in Section 2.1, see in particular Figures 8 and 9. We also characterize the optimal profiles for $d \geq 3$, but in this case its computation is no longer explicit.

This chapter is organized as follows: in Section 2 we start by proving general results about the Ambrosio-Tortorelli and the connectedness functional to then discuss the central results in this Chapter, Theorems 5.10 and 5.11. In these results we give the major arguments that are then adapted in Section 3 to the Γ-convergence for both the average distance minimizers and the $W - \mathscr{H}^1$ problems. In Section 4 we give our concluding remarks.

2. The Γ-convergence: the general theory

In this section we write general results about the interplay between the Ambrosio-Tortorelli term and connectedness functional in Γ-convergence. We start by studying $\mathcal{AT}_p$ and $\mathcal{C}_\varepsilon$ separately in paragraphs 2.1 and 2.2, and finally we give flexible Theorems in Subsection 2.3 that give conditions under which these functionals behave well together, allowing to prove both Γ-convergence Theorems 5.1 and 5.2. Hopefully this modular presentation will be helpful in the analysis of phase-field approximations for other problems in the future.

2.1. Properties of $\mathcal{AT}_p$

In this section we discuss the individual properties of the Ambrosio-Tortorelli type functional defined as

$$\mathcal{AT}_p(\varphi_\varepsilon) \overset{\text{def.}}{=} \frac{1}{\Lambda_{p,d}} \int_\Omega \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathrm{d}x. \tag{5.9}$$

As we have chosen $p > d$, see the discussion surrounding this choice in the introduction, if $\mathcal{AT}_p(\varphi_\varepsilon) < \infty$ the family of phase fields φ_ε belongs to the Sobolev space $W^{1,p}(\Omega)$, and since we assume from the start that Ω has a Lipschitz boundary, from the classical Sobolev injection $W^{1,p}(\Omega) \hookrightarrow C^{0,\beta}(\overline{\Omega})$, φ_ε is β-Hölder continuous for $\beta = 1 - \frac{d}{p}$.

We recall the definition of the diffuse transition measure defined in (5.7) as

$$\mu_\varepsilon \overset{\text{def.}}{=} \frac{1}{\Lambda_{p,d}} \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathcal{L}^d \llcorner \Omega.$$

We shall see that not only $\mathcal{AT}_p(\varphi_\varepsilon)$ approximates the quantity $\mathscr{H}^1(\Sigma)$, but one can also find a good family of phase fields φ_ε such that $\mu_\varepsilon \xrightarrow[\varepsilon \to \infty]{\star} \mathscr{H}^1 \llcorner \Sigma$, whenever Σ is a connected set. With this goal, let us start with a simple example, of approximating a segment $L = [0,1]e_d$ in $\mathbb{R}^d$. By symmetry, it is natural to expect that a radially symmetric profile around the d-axis would suffice. This motivates the following $(d-1)$-variational problem

$$\Lambda_{p,d} \overset{\text{def.}}{=} \min \left\{ \mathcal{C}_{p,d}(u) \overset{\text{def.}}{=} \int_{\mathbb{R}^{d-1}} \left(\frac{1}{p} |\nabla u|^p + \frac{1}{p'} (1-u)^2 \right) \mathrm{d}x : \begin{array}{c} u(0) = 0 \\ \nabla u \in L^p(\mathbb{R}^{d-1}) \\ 1-u \in L^2(\mathbb{R}^{d-1}) \end{array} \right\}, \tag{5.10}$$

which is inspired on the analysis of [Chambolle et al., 2019a] and will be used in the proofs of the Γ-liminf and Γ-limsup. Clearly, functions u with finite energy above are equal to 1 at infinity. Moreover, we can relate solutions of (5.10) with the following problem in $\mathbb{R}$

$$\lambda_{p,d} \overset{\text{def.}}{=} \min \left\{ c_{p,d}(f) \overset{\text{def.}}{=} \int_0^{+\infty} t^{d-2} \left(\frac{1}{p} |f'|^p + \frac{1}{p'} (1-f)^2 \right) \mathrm{d}t : \begin{array}{c} f(0) = 0, \\ f \in AC^p(\mathbb{R}_+) \end{array} \right\}, \tag{5.11}$$

where $AC^p(\mathbb{R}_+)$ denotes the space of p-absolutely continuous curves.

It is harder to apply the Direct method to this second problem since the term t^{d-2} gets in the way of bounding the L^p-norm of the velocities, but we manage to derive existence and uniqueness for (5.11) from (5.10). In dimension $d = 2$, it can be solved explicitly without resorting to the Euler-Lagrange equations.

Proposition 5.3. *For any $p \geq 2$, the variational problem (5.10) admits a unique minimizer, which is radially symmetric of the form $u(x) = f_p(|x|)$, where $f_p : \mathbb{R}_+ \to [0,1]$ is the unique non-decreasing Hölder continuous minimizer of (5.11), and we have that*

$$\Lambda_{p,d} = \sigma_{d-2} \lambda_{p,d}, \text{ where } \sigma_{d-2} = \mathscr{H}^{d-2}(\mathbb{S}^{d-2})$$

is the area of the $d-2$ unit sphere in $\mathbb{R}^{d-1}$.

In addition, for the case $p > d = 2$ the optimal profile f_p is given by

$$f_p(t) \overset{\text{def.}}{=} \begin{cases} 1 - \left(1 - \frac{p-2}{p} t\right)^{\frac{p}{p-2}}, & 0 \leq t \leq \frac{p}{p-2}, \\ 1, & t \geq \frac{p}{p-2}, \end{cases}$$

so that the value can be computed explicitly as

$$\lambda_{p,2} = \int_0^1 (1-u)^{2/p'} \mathrm{d}u = \frac{p}{3p-2}.$$

Proof. Existence and uniqueness of a minimizer of (5.10), follows from a classical argument using the direct method and the fact that the energy $\mathcal{C}_{p,d}$ is strictly convex. In addition, as this energy is invariant with respect to rotations around the origin, the solution must be radially symmetric and given by

$$u(x) = f_p(|x|).$$

From Morrey's inequality

$$[u]_{C^{0,\beta}(\mathbb{R}^{d-1})} \leq C_d \left\| \nabla u \right\|_{L^p(\mathbb{R}^{d-1})}, \text{ for } \beta = 1 - \frac{d}{p},$$

we conclude that f_p must be β-Hölder continuous.

Let us show that f_p is the unique minimizer of (5.11). Given any u admissible for (5.10), from the coarea formula [Evans and Gariepy, 2015, Thm. 3.13] and a change of variables, it holds that

$$\begin{aligned}\mathcal{C}_{p,d}(u) &= \int_0^\infty \left(\int_{t\mathbb{S}^{d-2}} \frac{1}{p}|\nabla u|^p + \frac{1}{p'}(1-u)^2 \mathrm{d}\mathscr{H}^{d-2}\right)\mathrm{d}t \\ &= \int_{\mathbb{S}^{d-2}} \int_0^\infty t^{d-2}\left(\frac{1}{p}|\nabla u(t\xi)|^p + \frac{1}{p'}(1-u(t\xi))^2 \mathrm{d}t\right)\mathrm{d}\mathscr{H}^{d-2}(\xi) \\ &\geq \int_{\mathbb{S}^{d-2}} \int_0^\infty t^{d-2}\left(\frac{1}{p}|(u(t\xi))'|^p + \frac{1}{p'}(1-u(t\xi))^2 \mathrm{d}t\right)\mathrm{d}\mathscr{H}^{d-2}(\xi) \geq \sigma_{d-2}\lambda_{p,d},\end{aligned}$$

where the last inequality comes from the fact that, for all $\xi \in \mathbb{S}^{d-2}$, the function $t \mapsto u(t\xi)$ is admissible for the problem (5.11), as $1-u \in L^2(\mathbb{R}^{d-1})$. As a result, if f_p is not the unique solution to (5.11), then there is another function $\bar{f}$ such that

$$\lambda_{p,d} < c_{p,d}(\bar{f}) < c_{p,d}(f_p),$$

so that $\bar{u}(x) = \bar{f}(|x|)$ has a strictly smaller energy that $u(x) = f_p(|x|)$. Hence, f_p must be a minimizer of (5.11), which must be unique from the uniqueness of solutions to (5.10). As a consequence, we conclude that $\Lambda_{p,d} = \sigma_{d-2}\lambda_{p,d}$.

The fact that f_p has image in $[0,1]$ comes from the fact that, if it was not the case, we could replace it with $\max\{0, \min\{f_p, 1\}\}$ and obtain a strictly smaller energy. Similarly, the strict monotonicity of f_p is achieved by replacing f_p with

$$\bar{f}_p(t) \overset{\text{def.}}{=} \max\{f_p(s) : \ 0 \leq s \leq t\},$$

yielding a strictly better energy if f_p and $\bar{f}_p$ do not coincide.

The variational problem (5.11) can become quite intractable for general p and d, but for the special case $d = 2$, we can refine the previous argument since the Lagrangian for (5.11) is now autonomous. In this case, notice that for any f admissible, we obtain a lower bound for $\lambda_{p,2}$ as

$$\begin{aligned}\int_0^{+\infty}\left(\frac{1}{p}|f'(t)|^p + \frac{1}{p'}(1-f(t))^2\right)\mathrm{d}t &\geq \int_0^{+\infty} (1-f(t))^{2/p'}|f'(t)|\mathrm{d}t \\ &= \lim_{t\to\infty}\int_0^{f(t)} (1-u)^{2/p'}\mathrm{d}u = \int_0^1 (1-u)^{2/p'}\mathrm{d}u\end{aligned}$$

where the inequality comes from Young's inequality, $ab \leq \dfrac{1}{p}a^p + \dfrac{1}{p'}b^{p'}$ which is an equality if and only if $a = b^{p'-1}$, or equivalently $b = a^{p-1}$. An optimal solution f_p to the 1D variational problem must then satisfy Young's inequality with equality for a.e. t, and hence it must solve the ODE

$$f_p'(t) = (1-f_p(t))^{2/p}, \quad f_p(0) = 0. \tag{5.12}$$

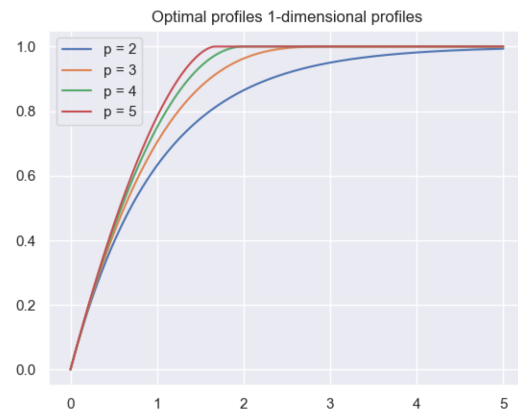

Figure 8: Optimal profiles induced by the Ambrosio-Tortorelli functional in $\mathbb{R}^2$ with different values of p. From the behavior of the optimal solution as p grows, one can expect that phase-field approximations with $p > 2$ yields sharper results.

It is then straightforward to verify that f_p is given by

$$f_p(t) = 1 - \left(1 - \frac{p-2}{p}t\right)^{\frac{p}{p-2}}, \text{ for } t \in \left[0, \frac{p}{p-2}\right],$$

and for $t > p/(p-2)$ we extend f_p with 1 and the integral remains unchanged. Using f_p in the energy from (5.11), we attain the lower bound for $\lambda_{p,2}$ above, and it follows that f_p must be the unique minimizer. □

An analogous analysis can be performed for the case $p = d = 2$. In this case, the same argument with Young's inequality gives the ODE $f'(t) = 1 - f(t)$ with the same boundary conditions, whose solution is given by $f_2(t) = 1 - e^{-t}$, for $t \geq 0$. Now the optimal profile never attains the value 1 and is more diffuse, see Figure 8. Compare it also with the proof of Lemma 2.8 in [Bonnivard et al., 2015].

Remark 5.4 (On the regularity and the support of f_p)**.** *In the case $d > 2$, it is not clear if f_p attains the value 1 in finite time. This is the case for $d = 2$, and should also hold for $d > 2$, as the extra term t^{d-2} penalizes even more f_p being away from 1.*

In addition, besides being globally Hölder continuous, from the Euler-Lagrange equations, it follows that f_p is C^1 with a Hölder continuous derivative.

In the sequel, given a connected and countably $\mathscr{H}^1$-rectifiable set Σ, we use this optimal profile to construct a family of phase-fields $(\varphi_\varepsilon)_{\varepsilon>0}$ such that the associated diffuse transition measures μ_ε approximate $\mathscr{H}^1 \llcorner \Sigma$. Our strategy will be to combine the optimal profile obtained in Prop. 5.3 with the fact that the Minkowski content coincides with the Hausdorff measure, see [Ambrosio et al., 2000, Thm. 2.104]. More precisely, if Σ is closed and countably $\mathscr{H}^1$-rectifiable, defining $\Sigma_t \stackrel{\text{def.}}{=} \{x : \operatorname{dist}(x, \Sigma) \leq t\}$ it holds that

$$\lim_{t \to 0} \frac{\mathcal{L}^d(\Sigma_t)}{\omega_{d-1} t^{d-1}} = \mathscr{H}^1(\Sigma), \tag{5.13}$$

where ω_{d-1} denotes the volume of the unitary $d-1$-dimensional ball. This property is not always true; if Σ is a rectifiable curve it is known to be true, see [Federer, 2014, Thm. 3.2.39]. Alternatively, the conclusion (5.13) also holds if there is a Radon measure μ over Σ that is Ahlfors regular from below, [Ambrosio et al., 2000, Thm. 2.104]; that is, there exists a constant $c > 0$ and some $r_0 > 0$ such that for all $x \in \Sigma$ if holds that

$$\mu(B_r(x)) \geq cr, \text{ for all } x \in \Sigma \text{ and } r < r_0.$$

Of course this is true for any path-connected set Σ by taking $\mu = \mathscr{H}^1 \llcorner \Sigma$ since for any $r < \operatorname{diam}(\Sigma)/2$ and $x \in \Sigma$ we have $\mathscr{H}^1(\Sigma \cap B_r(x)) \geq r$, so for any set we might be interested in this work, its Hausdorff measure coincides with the Minkowski content. We shall use in fact that this equality implies a weak convergence in the space of measures.

Lemma 5.5. *Let Σ be a compact, connected and countably $\mathscr{H}^1$-rectifiable subset of $\mathbb{R}^d$ with finite length $\mathscr{H}^1(\Sigma) < \infty$, then it holds that*

$$\frac{1}{\omega_{d-1}t^{d-1}}\mathcal{L}^d \llcorner \Sigma_t \xrightarrow[t\to 0]{\star} \mathscr{H}^1 \llcorner \Sigma.$$

Proof. Set $\nu_t \overset{\text{def.}}{=} \frac{1}{\omega_{d-1}t^{d-1}}\mathcal{L}^d \llcorner \Sigma_t$ for $t > 0$, notice that property (5.13) implies that $\nu_t(\mathbb{R}^d) \xrightarrow[t\to 0]{} \mathscr{H}^1(\Sigma)$. Let ν be a weak cluster point of ν_t, if we show that $\nu \geq \mathscr{H}^1 \llcorner \Sigma$, the convergence of the total mass implies that $\nu = \mathscr{H}^1 \llcorner \Sigma$.

From [Ambrosio et al., 2000, Prop 2.101], which shows that the "lower Minkowski content" of a rectifiable set is larger than its Hausdorff measure, as the set $\Sigma \cap \overline{B_r(x)}$ is closed and countably $\mathscr{H}^1$-rectifiable, for any $x \in \mathbb{R}^d$ and $0 < r' < r$ it holds that

$$\begin{aligned}\nu(B_r(x)) &\geq \limsup_{t\to 0} \nu_t\left(\overline{B_{r'}(x)}\right) \geq \liminf_{t\to 0} \frac{\mathcal{L}^d\left(\left\{\operatorname{dist}(\cdot, \Sigma \cap \overline{B_{r'}(x)}) \leq t\right\}\right)}{\omega_{d-1}t^{d-1}} \\ &\geq \mathscr{H}^1\left(\Sigma \cap \overline{B_{r'}(x)}\right), \text{ for } r' < r.\end{aligned}$$

Letting $r' \to r$, we conclude that $\nu \geq \mathscr{H}^1 \llcorner \Sigma$, and equality follows since both measures have the same total mass. Since all cluster points of ν_t are $\mathscr{H}^1 \llcorner \Sigma$, it must be the weak limit of the entire family. □

Now we prove the promised approximation result in $\mathbb{R}^d$, which is a strengthened version of [Bonnivard et al., 2015, Lemma 2.8], from where the main idea of the proof is borrowed. In [Bonnivard et al., 2015] the corresponding result is not stated as a weak convergence of the diffuse transition measure and is only proved in $\mathbb{R}^2$. Although the weak convergence is a small improvement to [Bonnivard et al., 2015], it is crucial to the proof of the Γ-convergence result for ($\overline{P}_\Lambda$). We expect that the diffuse transition measures μ_ε associated with the family

$$\varphi_\varepsilon(x) \overset{\text{def.}}{=} \begin{cases} f_p\left(\dfrac{d_\Sigma(x) - b_\varepsilon}{\varepsilon}\right), & \text{if } d_\Sigma(x) \geq b_\varepsilon \\ 0, & \text{otherwise,} \end{cases} \tag{5.14}$$

will converge to $\mathscr{H}^1 \llcorner \Sigma$, where $d_\Sigma(\cdot) \overset{\text{def.}}{=} \operatorname{dist}(\cdot, \Sigma)$ and $b_\varepsilon = o(\varepsilon)$. We only need to be careful with the boundary condition $\varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega)$, this is the case if f_p reaches 1 in finite time, for instance if $p > d = 2$.

Theorem 5.6 (Approximation with diffuse measures). *Given $\Sigma \subset \Omega$ closed, connected with $\mathscr{H}^1(\Sigma) < \infty$. Then, there is a family $(\varphi_\varepsilon)_{\varepsilon>0} \subset 1 + W_0^{1,p}(\Omega)$ whose corresponding diffuse approximation measures defined in (5.7) are such that*

$$\mu_\varepsilon \xrightarrow[\varepsilon\to 0]{} \mathscr{H}^1 \llcorner \Sigma,$$

in both the narrow topology and the weak-$\star$ topologies of $\mathcal{M}_b(\Omega)$.

If $\Sigma \subset \operatorname{int}\Omega$, this family $(\varphi_\varepsilon)_{\varepsilon>0}$ can be constructed such that $\varphi_\varepsilon \equiv 0$ over the set $\{\operatorname{dist}(\cdot, \Sigma) \leq b_\varepsilon\}$, for $b_\varepsilon = o(\varepsilon)$.

Proof. The proof will be done in multiple cases of increasing generality, depending if Σ has a part contained in the boundary of Ω and if the optimal profile f_p reaches 1 in finite time. As a preliminary result, we prove an approximation result in the entire space $\mathbb{R}^d$. Define $d_\Sigma(x) \overset{\text{def.}}{=} \operatorname{dist}(x, \Sigma)$, set the notation

$$\Sigma_r \overset{\text{def.}}{=} \{x \in \Omega : d_\Sigma(x) \leq r\}$$

and let $(\varphi_\varepsilon)_{\varepsilon>0}$ be the family defined in (5.14). Consider now the measures over $\mathcal{M}_b(\mathbb{R}^d)$

$$\varrho_\varepsilon \overset{\text{def.}}{=} \frac{1}{\Lambda_{p,d}} \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1 - \varphi_\varepsilon)^2 \right) \mathcal{L}^d. \tag{5.15}$$

To show the convergence of $(\varrho_\varepsilon)_{\varepsilon>0}$ to $\mathscr{H}^1 \llcorner \Sigma$, our strategy will be to use the Minkowski content of Σ, and more specifically Lemma 5.5, to verify

$$\int_{\mathbb{R}^d} \psi \mathrm{d}\varrho_\varepsilon \xrightarrow[\varepsilon\to 0]{} \int_\Sigma \psi \mathrm{d}\mathscr{H}^1 \text{ for any } \psi \in C_b(\mathbb{R}^d). \tag{5.16}$$

Fixing $\psi \in C_b(\mathbb{R}^d)$, from the coarea formula [Evans and Gariepy, 2015, Thm. 3.13], we can define

$$\Psi : t \mapsto \int_{\Sigma_t} \psi \mathrm{d}x, \text{ such that } \Psi'(t) = \int_{\partial\Sigma_t} \psi \mathrm{d}\mathscr{H}^{d-1} \text{ for a.e. } t > 0.$$

It follows that Ψ is bounded, and Lemma 5.5 implies that

$$\Psi(t) = \omega_{d-1} t^{d-1} \int_\Sigma \psi \mathrm{d}\mathscr{H}^1 + o(t^{d-1}), \tag{5.17}$$

where the $o(t^{d-1})$ depends only on Σ and the function ψ.

Since the functions φ_ε are defined as the composition of a 1 dimensional profile with the distance function d_Σ, we can disintegrate then with the sets $\partial\Sigma_t$, over which φ_ε is constant. For this reason, we define the quantity

$$\begin{aligned} h_\varepsilon(t) &\overset{\text{def.}}{=} \frac{\varepsilon^{p-d+1}}{p}\left(\frac{\mathrm{d}}{\mathrm{d}t} f_p\left(\frac{t}{\varepsilon}\right)\right)^p + \frac{\varepsilon^{-d+1}}{p'}\left(1 - f_p\left(\frac{t}{\varepsilon}\right)\right)^2 \\ &= \varepsilon^{-d+1}\left(\frac{1}{p}\left|f_p'\left(\frac{t}{\varepsilon}\right)\right|^p + \frac{1}{p'}\left(1 - f_p\left(\frac{t}{\varepsilon}\right)\right)^2\right). \end{aligned}$$

Notice that since f_p is the optimal 1-dimensional profile from the problem (5.11), it follows that $h_\varepsilon(t) \xrightarrow[t\to\infty]{} 0$, for all $\varepsilon > 0$.

In the sequel we decompose the integral in (5.16) as

$$\int_{\mathbb{R}^d} \psi \mathrm{d}\varrho_\varepsilon = \int_{\Sigma_{b_\varepsilon}} \psi \mathrm{d}\varrho_\varepsilon + \int_{\mathbb{R}^d\setminus\Sigma_{b_\varepsilon}} \psi \mathrm{d}\varrho_\varepsilon$$

Since $\varphi_\varepsilon \equiv 0$ over Σ_{b_ε}, the first integral on the right-hand side above becomes

$$\int_{\Sigma_{b_\varepsilon}} \psi \mathrm{d}\varrho_\varepsilon = \frac{\varepsilon^{-d+1}}{p'\Lambda_{p,d}}\Psi(b_\varepsilon) = \left(\frac{\omega_{d-1}}{p'\Lambda_{p,d}}\int_\Sigma \psi \mathrm{d}\mathscr{H}^1\right)\left(\frac{b_\varepsilon}{\varepsilon}\right)^{d-1} + \frac{o({b_\varepsilon}^{d-1})}{\varepsilon^{d-1}} \xrightarrow[\varepsilon\to 0]{} 0,$$

which converges to 0 since $b_\varepsilon = o(\varepsilon)$.

Hence, in order to study the convergence (5.16), it suffices to consider the second term, which can be rewritten with the coarea formula as

$$\begin{aligned} \int_{\mathbb{R}^d\setminus\Sigma_{b_\varepsilon}} \psi \mathrm{d}\varrho_\varepsilon &= \frac{1}{\Lambda_{p,d}}\int_0^{+\infty} h_\varepsilon(t)\Psi'(t+b_\varepsilon)\mathrm{d}t \\ &= \frac{1}{\Lambda_{p,d}}\left(h_\varepsilon(t)\Psi(t+b_\varepsilon)\big|_0^{+\infty} - \int_0^\infty h_\varepsilon'(t)\Psi(t+b_\varepsilon)\mathrm{d}t\right) \end{aligned}$$

Recalling that $h_\varepsilon(t) \xrightarrow[t\to\infty]{} 0$ and that Ψ is a bounded function such that $\Psi(b_\varepsilon) = o(\varepsilon^{d-1})$, from the Minkowski content, the boundary terms vanish at the limit and we retain once again just the integral part, which we develop further as

$$\begin{aligned} \left(-\int_0^\infty h_\varepsilon'(t)\Psi(t+b_\varepsilon)\mathrm{d}t\right) &= \omega_{d-1}\left(\int_\Sigma \psi \mathrm{d}\mathscr{H}^1 + o(1)\right)\int_0^\infty -(t+b_\varepsilon)^{d-1}h_\varepsilon'(t)\mathrm{d}t \\ &= \omega_{d-1}(d-1)\left(\int_\Sigma \psi \mathrm{d}\mathscr{H}^1 + o(1)\right)\int_0^\infty (t+b_\varepsilon)^{d-2}h_\varepsilon(t)\mathrm{d}t, \end{aligned}$$

where we have used (5.17) in the first equality. Using the fact that h_ε is obtained with the optimal profile defining the constant $\lambda_{p,d}$ we obtain

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d} \psi \mathrm{d}\varrho_\varepsilon = \frac{\omega_{d-1}(d-1)}{\Lambda_{p,d}}\left(\int_\Sigma \psi \mathrm{d}\mathscr{H}^1\right)\lim_{\varepsilon\to 0}\underbrace{\int_0^\infty \left(t + \frac{b_\varepsilon}{\varepsilon}\right)^{d-2}\left(\frac{1}{p}\left|f_p'\right|^p + \frac{1}{p'}(1-f_p)^2\right)\mathrm{d}t}_{=:\lambda_{p,d,\varepsilon}}$$

From Lebesgue's dominated convergence theorem, $\lambda_{p,d,\varepsilon} \xrightarrow[\varepsilon\to 0]{} \lambda_{p,d}$, since $b_\varepsilon = o(\varepsilon)$. Hence, recalling that $\omega_{d-1}(d-1) = \sigma_{d-2}$ and that, $\Lambda_{p,d} = \sigma_{d-2}\lambda_{p,d}$ from Proposition 5.3, we obtain the desired convergence

$$\int_{\mathbb{R}^d} \psi \mathrm{d}\varrho_\varepsilon \xrightarrow[\varepsilon\to 0]{} \int_\Sigma \psi \mathrm{d}\mathscr{H}^1$$

for all $\psi \in C_b(\mathbb{R}^d)$, so that $\varrho_\varepsilon \xrightarrow[\varepsilon\to 0]{} \mathscr{H}^1 \llcorner \Sigma$ in both the narrow and the weak-$\star$ topologies.

Now let us make the construction of approximating phase-fields with the additional constraint that $\varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega)$. As mentioned in the beginning, we shall divide our construction in different cases.

Case 1 ($\Sigma \subset \operatorname{int}\Omega$ and $\operatorname{supp}(1-f_p)$ is compact):

Given $\Sigma \subset \operatorname{int}\Omega$, since $1-f_p$ has compact support, the family $(\varphi_\varepsilon)_{\varepsilon>0}$ defined in (5.14) is contained in $1 + W_0^{1,p}(\Omega)$. Indeed, setting $t_p \overset{\text{def.}}{=} \inf\{t \geq 0 : f_p(t) = 1\} < \infty$, we get that $\varphi_\varepsilon < 1$ only when

$$\frac{d_\Sigma(\cdot) - b_\varepsilon}{\varepsilon} \leq t_p, \text{ so that } \operatorname{supp}(1-\varphi_\varepsilon) \subset \Sigma_{\varepsilon t_p + b_\varepsilon} \subset \operatorname{int}\Omega,$$

whenever $\varepsilon t_p + b_\varepsilon < \operatorname{dist}(\Sigma, \partial\Omega)$. Therefore, for ε small enough, we have $\mu_\varepsilon = \varrho_\varepsilon$ and the result follows.

Case 2 ($\Sigma \subset \operatorname{int}\Omega$ and $\operatorname{supp}(1-f_p)$ not compact):

When we can no longer assume that the support of $1-f_p$ is compact, we approximate it with another profile with compact support. In this case, set

$$t_\varepsilon \overset{\text{def.}}{=} \frac{\operatorname{dist}(\Sigma, \partial\Omega)}{2\varepsilon}, \quad \lambda_\varepsilon \overset{\text{def.}}{=} f_p(t_\varepsilon) \xrightarrow[\varepsilon\to 0]{} 1,$$

and define the new profiles

$$f_{p,\varepsilon}(t) \overset{\text{def.}}{=} \begin{cases} \dfrac{1}{\lambda_\varepsilon} f_p(t), & \text{if } t \leq t_\varepsilon, \\ 1, & \text{otherwise,} \end{cases} \qquad \bar{\varphi}_\varepsilon(x) \overset{\text{def.}}{=} \begin{cases} f_{p,\varepsilon}\left(\dfrac{d_\Sigma(x) - b_\varepsilon}{\varepsilon}\right), & \text{if } d_\Sigma(x) \geq b_\varepsilon \\ 0, & \text{otherwise.} \end{cases}$$

Similarly to the previous case, $\bar{\varphi}_\varepsilon \in 1 + W_0^{1,p}(\Omega)$ for ε small enough.

Now let μ_ε denote the diffuse approximation measures referent to $\bar{\varphi}_\varepsilon$. Since we know that ϱ_ε from (5.15) converge weakly to $\mathscr{H}^1 \llcorner \Sigma$, to obtain the same limit for μ_ε it suffices to show that

$$\|\mu_\varepsilon - \varrho_\varepsilon\|_{L^1(\mathbb{R}^d)} \xrightarrow[\varepsilon\to 0]{} 0.$$

Indeed, a similar computation to the start of the proof using the coarea formula gives

$$\|\mu_\varepsilon - \varrho_\varepsilon\|_{L^1(\mathbb{R}^d)} \leq \int_{\mathbb{R}^d} \left(\frac{\varepsilon^{p-d+1}}{p} \left(\frac{1}{\lambda_\varepsilon^p} - 1 \right) |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p} \left((1-\varphi_\varepsilon)^2 - \left(1 - \frac{1}{\lambda_\varepsilon}\varphi_\varepsilon\right)^2 \right) \right) \mathrm{d}x$$

$$= \varepsilon^{-d+1} \int_{b_\varepsilon}^{\infty} \mathscr{H}^1(\Sigma_t) \left[\left(\frac{1}{\lambda_\varepsilon^p} - 1 \right) \frac{1}{p} \left| f_p'\left(\frac{t}{\varepsilon}\right) \right|^p \right.$$

$$\left. + \frac{1}{p'} \left(\left(1 - f_p\left(\frac{t}{\varepsilon}\right)\right)^2 - \left(1 - \frac{1}{\lambda_\varepsilon} f_p\left(\frac{t}{\varepsilon}\right)\right)^2 \right) \right] \mathrm{d}t$$

$$\leq C \int_{b_\varepsilon/\varepsilon}^{\infty} s^{d-2} \left[\left(\frac{1}{\lambda_\varepsilon^p} - 1 \right) \frac{1}{p} |f_p'|^p + \frac{1}{p'} \left((1-f_p)^2 - \left(1 - \frac{1}{\lambda_\varepsilon} f_p\right)^2 \right) \right] \mathrm{d}s.$$

From Lebesgue's dominated convergence theorem, we conclude that $\|\mu_\varepsilon - \varrho_\varepsilon\|_{L^1(\mathbb{R}^d)} \xrightarrow[\varepsilon\to 0]{} 0$ so that $\mu_\varepsilon \xrightarrow[\varepsilon\to 0]{} \mathscr{H}^1 \llcorner \Sigma$.

Case 3 ($\Sigma \subset \Omega$ and $\mathrm{supp}(1-f_p)$ not compact): For this case we exploit the assumption that Ω is star-shaped to define a sequence of sets $\Sigma_n \subset \mathrm{int}\,\Omega$ and such that $\mathscr{H}^1 \llcorner \Sigma_n \xrightarrow[N\to\infty]{} \mathscr{H}^1 \llcorner \Sigma$. Notice that we make this assumption to slightly simplify the proof, but a similar construction can be made by assuming that Ω has a continuous boundary any using a partition of the unity over the boundary.

We consider $x_\star \in \mathrm{int}\,\Omega$ such that $tx_\star + (1-t)x \in \mathrm{int}\,\Omega$ for any $x \in \Omega$ and any $t \in (0,1)$.

So considering the sequence

$$\Sigma_n \overset{\text{def.}}{=} \frac{1}{n}x_\star + \left(1 - \frac{1}{n}\Sigma\right), \text{ if holds } \mathscr{H}^1 \llcorner \Sigma_n \xrightarrow[\varepsilon\to 0]{} \mathscr{H}^1 \llcorner \Sigma.$$

We let $(\varphi_{n,\varepsilon})_{\varepsilon>0}$ be the family in $1 + W_0^{1,p}(\Omega)$ obtained in the previous case whose diffuse approximation measures $(\mu_{n,\varepsilon})_{\varepsilon>0}$ are such that

$$\mu_{n,\varepsilon} \xrightarrow[\varepsilon\to 0]{} \mathscr{H}^1 \llcorner \Sigma_n \xrightarrow[n\to\infty]{} \mathscr{H}^1 \llcorner \Sigma.$$

Hence, a diagonal extraction argument yields the desired sequence. □

Notice that this proof also works for the case $p = 2$, using the corresponding optimal 1-dimensional profile, see Figure 8, as done in [Bonnivard et al., 2015]. As discussed after Prop. 5.3, since the 1-dimensional profile for $p > 2$ promotes a sharper transition, the optimal sequence of phase fields constructed in the previous Theorem should have a better perceptual reconstruction. This is corroborated in Figure 9.

Remark 5.7. *In Theorem 5.6, we have actually shown that the sequence of diffuse transition measures corresponding to the family defined in (5.14) is such that*

$$\mu_\varepsilon \xrightarrow[\varepsilon\to 0]{} \mathscr{H}^1 \llcorner \Sigma, \text{ in } \mathcal{M}_b(\mathbb{R}^d).$$

If we had simply restricted φ_ε from (5.14) to Ω it would follow that

$$\mu_\varepsilon \xrightarrow[\varepsilon\to 0]{} \mathscr{H}^1 \llcorner (\Sigma \cap \mathrm{int}\,\Omega) + \frac{1}{2}\mathscr{H}^1 \llcorner (\Sigma \cap \partial\Omega), \text{ in } \mathcal{M}_b(\Omega).$$

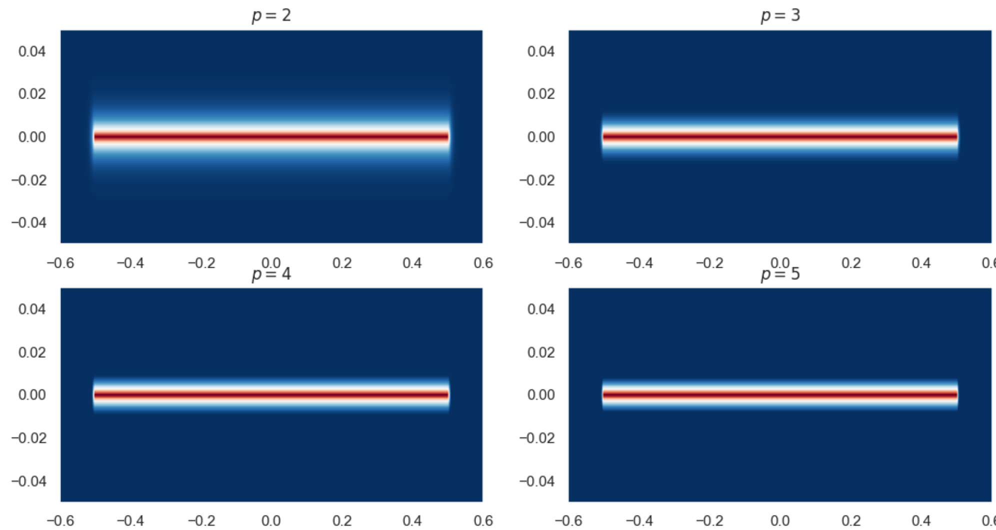

Figure 9: Recovery sequences (for $b_\varepsilon = 0$) obtained with the optimal profile from Prop. 5.3 for different values of p and $\varepsilon = 0.01$.

2.2. Properties of $\mathcal{C}_\varepsilon$

In the sequel, we survey some properties of the connectedness functional. For a fixed parameter $s > 0$, the functional $\mathcal{C}_\varepsilon$ is designed to penalize the non-connectedness of the set $\{\varphi \leq \varepsilon^s\}$. Given a function $\Phi : \Omega \to [0,1]$ we define the weighted distance d_ε^Φ as

$$d_\varepsilon^\Phi(x,y) \overset{\text{def.}}{=} \inf \left\{ \int_K \Phi(x) \mathrm{d}\mathscr{H}^1(x) : \begin{array}{c} K \text{ connected, } x, y \in K \subset \Omega \\ \mathscr{H}^1(K) \leq \omega(\varepsilon) \end{array} \right\}, \tag{5.18}$$

where $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous, monotone increasing function such that $\omega(\varepsilon) \to \infty$ as $\varepsilon \to 0$. This quantity can only be a distance if $\Phi > 0$, except for a set of Hausdorff dimension strictly smaller than 1, but we shall commit the abuse of calling it a geodesic distance even if it is not necessarily the case.

However, there is no guarantee of being able to find a such K connecting x and y with a length smaller than $\omega(\varepsilon)$. For this reason, let $\mathrm{diam}_{\mathrm{geo}}(\Omega)$ denote the diameter of Ω w.r.t. the geodesic distance inside Ω, which can be defined as

$$\mathrm{dist}_{\mathrm{geo}}(x,y) \overset{\text{def.}}{=} \min \left\{ \mathscr{H}^1(\gamma) : x, y \in \gamma \text{ and } \gamma \subset \Omega \text{ is connected} \right\}.$$

As Ω is bounded and connected with Lipschitz boundary $\mathrm{diam}_{\mathrm{geo}}(\Omega) < \infty$ and for ε small enough so that $\omega(\varepsilon) > \mathrm{diam}_{\mathrm{geo}}(\Omega)$, there must be an admissible curve connecting x, y, so that the infimum (5.18) is bounded by $\|\Phi\|_\infty \, \mathrm{dist}_{\mathrm{geo}}(x,y)$.

Either way, assuming ε small enough, if we compose φ_ε with a function $F_\varepsilon(z)$ that is zero if $z \leq \varepsilon^s$, the quantity $d_\varepsilon^{F_\varepsilon \circ \varphi_\varepsilon}(x,y)$ gives a quantitative notion of how disconnected the two points x, y are in the set $\{\varphi_\varepsilon \leq \varepsilon^s\}$. To get a global notion, we must integrate among all pairs of points in this level set. This way, the *diffuse connectedness functional* $\mathcal{C}_\varepsilon$ from [Dondl and Wojtowytsch, 2021, Dondl et al., 2018] is then defined as

$$\mathcal{C}_\varepsilon(\varphi_\varepsilon) \stackrel{\text{def.}}{=} \int_{\Omega\times\Omega} \beta_\varepsilon(\varphi_\varepsilon(x))\beta_\varepsilon(\varphi_\varepsilon(y))d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}(x,y)\mathrm{d}x\mathrm{d}y, \tag{5.19}$$

where β_ε, F_ε are continuous functions such that

$$\beta_\varepsilon(z) = \begin{cases} 1 & \text{if } z \leq \varepsilon^s, \\ 0 & \text{if } z \geq 2\varepsilon^s, \end{cases} \quad F_\varepsilon(z) = \begin{cases} 0 & \text{if } z \leq \varepsilon^s/2, \\ 1 & \text{if } z \geq \varepsilon^s. \end{cases} \tag{5.20}$$

In addition, make the following hypothesis on these functions that, as (H1), will be assumed without statement

(H2) $\beta_\varepsilon, F_\varepsilon$ are strictly monotone in the intervals $(\varepsilon^s, 2\varepsilon^s)$ and $(\frac{1}{2}\varepsilon^s, \varepsilon^s)$ respectively and

$$F_\varepsilon\left(\frac{3}{4}\varepsilon^s\right) \geq \frac{1}{2}.$$

Next, in order to prove existence of solutions to the approximate functionals used in the Γ convergence result, we show that $\mathcal{C}_\varepsilon$ is continuous for uniform convergence.

Lemma 5.8. *Let Ω be a compact, connect set with Lipschitz boundary, and ε small enough so that $\omega(\varepsilon) > \operatorname{diam}_{geo}(\Omega)$. For $\varepsilon > 0$ fixed, the following facts hold*

1. *If $\Phi \in C\left(\overline{\Omega}\right)$, for every pair $x, y \in \Omega$ there is an optimal set K attaining the geodesic distance $d_\varepsilon^\Phi(x,y)$ defined in (5.18). In addition, this set can be taken a curve.*

2. *The geodesic distance $d_\varepsilon^\Phi(\cdot,\cdot)$ is Lipschitz continuous w.r.t. Φ for the uniform convergence, with Lipschitz constant given by $\omega(\varepsilon)$.*

3. *The connectedness functional $\mathcal{C}_\varepsilon$ is continuous w.r.t. uniform convergence of continuous functions.*

Proof. To prove (1), fix $x, y \in \Omega$ and Φ continuous. Consider a minimizing sequence of compact and connected sets K_n, with uniformly bounded length $\mathscr{H}^1(K_n) \leq \omega(\varepsilon)$ approximating the infimum in (5.18). From Blaschke's Theorem [Ambrosio et al., 2000, Thm. 6.1], we can extract a subsequence (not relabelled) converging in the Hausdorff metric to a connected set K, which must also contain the points x, y.

Consider now the measures $\nu_n \stackrel{\text{def.}}{=} \mathscr{H}^1 \llcorner K_n$, from the uniform bound on the lengths of K_n, we obtain that $\nu_n(\overline{\Omega}) \leq \omega(\varepsilon)$. Hence, as we are in a compact set, Prokhorov's compactness theorem implies that ν_n has a weak cluster point ν. In addition, from Gołab's theorem 2.10, from Chapter 2, we know that $\nu \geq \mathscr{H}^1 \llcorner K$ and the lower semi-continuity of the total variation norm w.r.t. weak convergence of measures gives that

$$\mathscr{H}^1(K) \leq \nu(\overline{\Omega}) \leq \liminf_{n\to\infty} \nu_n(\overline{\Omega}) = \liminf_{n\to\infty} \mathscr{H}^1(K_n) \leq \omega(\varepsilon),$$

so that K remains admissible for (5.18).

Finally, since K_n is a minimizing sequence and from the continuity of Φ we get

$$d_\varepsilon^\Phi(x,y) \le \int_K \Phi \mathrm{d}\mathscr{H}^1 \le \int_\Omega \Phi \mathrm{d}\nu = \lim_{n\to\infty} \int_{K_n} \Phi \mathrm{d}\mathscr{H}^1 = d_\varepsilon^\Phi(x,y),$$

so K attains the distance $d^\Phi(x,y)$. But as $\mathscr{H}^1(K) < \infty$ and it is connected, it must be pathwise connected and countably $\mathscr{H}^1$-rectifiable, so that it can be covered by countably many Lipschitz curves. But as $x, y \in K$, we can find a curve $\gamma \subset K$ whose end points are x, y. From the rectifiability of K, γ must be composed of countably many Lipschitz arcs.

To prove (2), consider two continuous functions Φ_1 and Φ_2 and let K_2 be optimal for the definition of $d_\varepsilon^{\Phi_2}(x,y)$, so that in particular $x, y \in K_2$ and $\mathscr{H}^1(K_2) \le \omega(\varepsilon)$. It then holds that

$$\begin{aligned} d_\varepsilon^{\Phi_1}(x,y) &\le \int_{K_2} \Phi_1 \mathrm{d}\mathscr{H}^1 = d_\varepsilon^{\Phi_2}(x,y) + \int_{K_2} (\Phi_1 - \Phi_2)\mathrm{d}\mathscr{H}^1 \\ &\le d_\varepsilon^{\Phi_2}(x,y) + \omega(\varepsilon)\left\|\Phi_1 - \Phi_2\right\|_\infty . \end{aligned}$$

Changing the roles of Φ_1 and Φ_2 the result follows.

Item (3) then becomes a consequence of the dominated convergence theorem: let $\varphi_n \xrightarrow[n\to\infty]{} \varphi$ uniformly, so that $\beta_\varepsilon \circ \varphi_n \xrightarrow[n\to\infty]{} \beta_\varepsilon \circ \varphi$ uniformly. The sequence $\beta_\varepsilon \circ \varphi_n$ remains uniformly bounded in n and so does $d_\varepsilon^{F_\varepsilon\circ\varphi_n}(x,y)$. As Ω is a bounded set, the dominated convergence theorem yields

$$\begin{aligned} \mathcal{C}_\varepsilon(\varphi) &= \int_{\Omega\times\Omega} \beta_\varepsilon(\varphi(x))\beta_\varepsilon(\varphi(y)) d^{F_\varepsilon\circ\varphi}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{\Omega\times\Omega} \lim_{n\to\infty} \beta_\varepsilon(\varphi_n(x))\beta_\varepsilon(\varphi_n(y)) d^{F_\varepsilon\circ\varphi_n}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \lim_{n\to\infty} \int_{\Omega\times\Omega} \beta_\varepsilon(\varphi_n(x))\beta_\varepsilon(\varphi_n(y)) d^{F_\varepsilon\circ\varphi_n}(x,y)\mathrm{d}x\mathrm{d}y = \lim_{n\to\infty} \mathcal{C}_\varepsilon(\varphi_n), \end{aligned}$$

proving the continuity of $\varphi \mapsto \mathcal{C}_\varepsilon(\varphi)$ w.r.t. uniform convergence. □

Remark 5.9. *Without the constraint $\mathscr{H}^1(K) \le \omega(\varepsilon)$ in the definition of $d_\varepsilon^\Phi(\cdot,\cdot)$ in (5.18), we believe it is possible to construct counter examples to the continuity property (2), for instance if we take $\Phi = \mathrm{dist}(\cdot, F)$, the distance function to a fractal set F as Koch's snowflake [Falconer, 1986].*

2.3. The fundamental liminf and limsup inequalities

In order to establish Γ-convergence results for problems ($W\mathscr{H}^1$) and (ADM) we will study the behavior of families of functions such that the following functional is uniformly bounded

$$F_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon) \overset{\text{def.}}{=} \mathcal{AT}_p(\varphi_\varepsilon) + \frac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) + \frac{1}{\varepsilon^\ell}\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon. \tag{5.21}$$

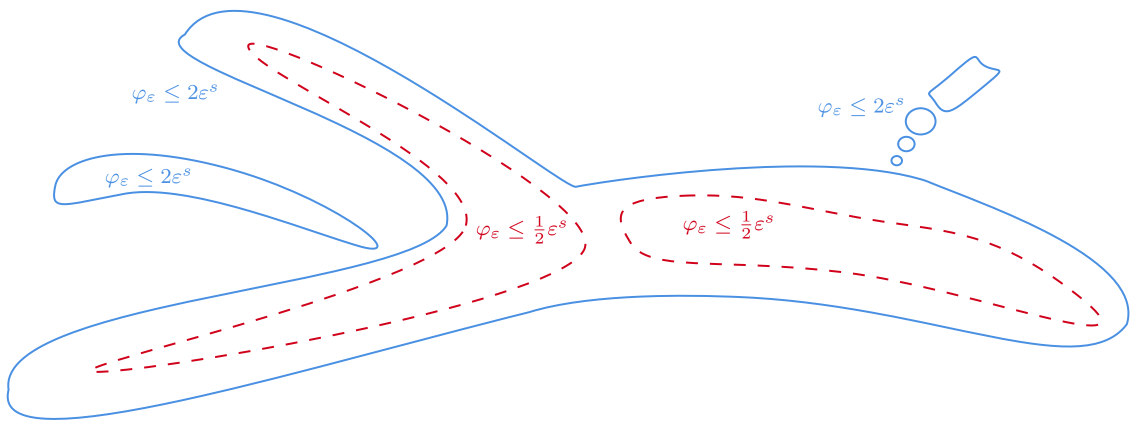

Figure 10: A single connected component of $\{\varphi_\varepsilon \le 2\varepsilon^s\}$, Σ_ε, contains all of $\{\varphi_\varepsilon \le \frac{1}{2}\varepsilon^s\}$ in red. Neither level sets are necessarily connected, but Σ_ε contains almost all the mass.

Our first result in this direction is a characterization of cluster points from families of functions such that $F_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon) \le C$ remains bounded for all $\varepsilon > 0$. We show this limit is supported in a connected, countably $\mathscr{H}^1$-rectifiable set whose length is bounded by the liminf of $F_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon)$.

Theorem 5.10. *Let Ω be a compact, connected subset of $\mathbb{R}^d$ with Lipschitz boundary, and suppose that $\ell > s$ and $\kappa > (2d+1)\frac{(s+1)p-d+1}{p-d}$. For any family $(\varphi_\varepsilon, \nu_\varepsilon)_{\varepsilon>0}$ such that for every $\varepsilon > 0$ it holds that $\varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega)$, $\nu_\varepsilon \in \mathscr{P}_{ac}(\Omega)$ and*

$$F_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon) \le C,$$

then it follows that:

(i) For ε small enough, there exists a connected component of $\{\varphi_\varepsilon \le 2\varepsilon^s\}$, denoted by Σ_ε, that contains $\{\varphi_\varepsilon \le \varepsilon^s/2\}$.

(ii) Up to subsequences, Σ_ε converges in the Hausdorff distance to a connected countably $\mathscr{H}^1$-rectifiable set Σ and $\varphi_\varepsilon \xrightarrow[\varepsilon\to 0]{} 1$, strongly in $L^2(\Omega)$. The families of measures $(\nu_\varepsilon, \mu_\varepsilon)_{\varepsilon>0}$, for μ_ε defined in (5.7), also converge $\nu_\varepsilon \xrightarrow[\varepsilon\to 0]{\star} \nu$, $\mu_\varepsilon \xrightarrow[\varepsilon\to 0]{\star} \mu$, and the limits satisfy

$$\mu \ge \mathscr{H}^1 \llcorner \Sigma, \text{ and } \operatorname{supp}\nu \subset \Sigma \subset \operatorname{supp}\mu.$$

In particular, it holds that

$$\mathscr{H}^1(\Sigma) \le \liminf_{\varepsilon\to 0} \frac{1}{\Lambda_{p,d}} \int_\Omega \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla\varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathrm{d}x. \tag{5.22}$$

Proof. **Step 1:** Our first step is to construct the connected and countably $\mathscr{H}^1$-rectifiable set Σ, where the cluster points of the diffuse measures μ_ε are concentrated. This will be done by studying the small level sets of the family $(\varphi_\varepsilon)_{\varepsilon>0}$. First we show that, for ε small enough, a single connected component of $\{\varphi_\varepsilon \leq 2\varepsilon^s\}$ contains all of $\{\varphi_\varepsilon \leq \varepsilon^s/2\}$, see Figure 10. Write

$$\{\varphi_\varepsilon < \varepsilon^s\} = \bigcup_{i\in I} \Sigma_{\varepsilon,i},$$

where $(\Sigma_{\varepsilon,i})_{i\in I}$ denote the set of connected components of $\{\varphi_\varepsilon < \varepsilon^s\}$. We distinguish the components that intersect the set we are interested in, $\{\varphi_\varepsilon < \varepsilon^s/2\}$, by defining the subset of indices

$$I_\star \stackrel{\text{def.}}{=} \{i \in I : \quad \Sigma_{\varepsilon,i} \cap \{\varphi_\varepsilon < \varepsilon^s/2\} \neq \emptyset\}\,.$$

Since $\varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega)$, any sublevel set $\{\varphi_\varepsilon \leq l\}$ is compactly contained in Ω for $0 \leq l < 1$. Hence, we can manipulate these sets without worrying about border effect.

First we check that $I_\star$ is not empty; if it were, we would have $\varphi_\varepsilon \geq \frac{1}{2}\varepsilon^s$ everywhere in Ω so that the term

$$\frac{\varepsilon^{s-\ell}}{2}|\nu_\varepsilon|(\Omega) \leq \frac{1}{\varepsilon^\ell}\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon \leq C$$

would yield a contradiction letting $\varepsilon \to 0$ since we have assumed $\ell > s$.

We claim that there is a radius r_ε such that for all $i \in I_\star$ and any $x \in \Sigma_{\varepsilon,i}\cap\{\varphi_\varepsilon \leq \varepsilon^s/2\}$ or $x' \in \{\varphi_\varepsilon = \varepsilon^s\}$ one has that

$$B(x; r_\varepsilon) \subset \Sigma_{\varepsilon,i} \text{ and } B(x'; r_\varepsilon) \subset \left\{\frac{3}{4}\varepsilon^s \leq \varphi_\varepsilon \leq 2\varepsilon^s\right\}. \tag{5.23}$$

This is a consequence of the fact that φ_ε is Hölder continuous. Indeed, since $p > d$, from Morrey's inequality (see [Evans, 2022, Thm. 5.4]) it holds that $\varphi_\varepsilon \in C^{0,\beta}$ with $\beta = 1 - \frac{d}{p}$ and a Hölder constant bounded by

$$[\varphi_\varepsilon]_{C^{0,\beta}(B_r)} \leq c\,\|\nabla\varphi_\varepsilon\|_{L^p(B_r)} \leq c\varepsilon^{-\frac{p-d+1}{p}},$$

where the two constants above differ, the first depends only on the dimension of Ω and the second inequality follows from the bound on $F_\varepsilon(\varphi_\varepsilon)$. Hence, it follows directly from the definition of Hölder continuity that (5.23) holds with

$$r_\varepsilon = c_0\varepsilon^{\beta'} \text{ with } \beta' > \frac{(s+1)p - d + 1}{p - d}. \tag{5.24}$$

In particular, we conclude that $I_\star$ is finite since taking exactly one ball for each of these connected components we obtain the bound $|I_\star|\omega_d r_\varepsilon^d \leq |\{\varphi_\varepsilon \leq 2\varepsilon^s\}| < \infty$.

In the sequel we define the quantities

$$\delta_{ij} \stackrel{\text{def.}}{=} d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}(\Sigma_{\varepsilon,i}, \Sigma_{\varepsilon,j}) \stackrel{\text{def.}}{=} \min_{x\in\Sigma_{\varepsilon,i}, y\in\Sigma_{\varepsilon,j}} d_\varepsilon^{F_\varepsilon\circ\varphi_\varepsilon}(x,y), \quad \text{for } i \neq j \in I_\star.$$

In the sequel, using the balls of radius r_ε defined in (5.23), we can bound δ_{ij} from above and from below. Starting with the upper bound, by definition, it must hold that

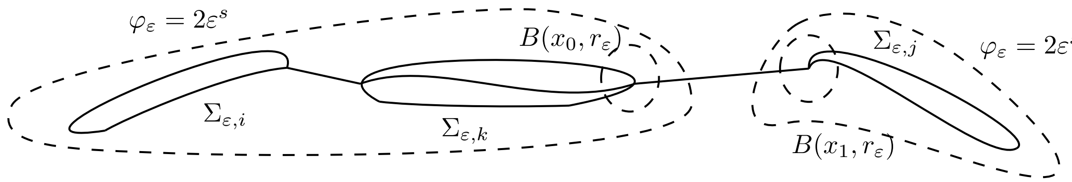

Figure 11: The optimal path between $\Sigma_{\varepsilon,i}$ and $\Sigma_{\varepsilon,j}$ has at least two segments of length r_ε.

$\delta_{ij} \leq d^{F_\varepsilon \circ \varphi_\varepsilon}(x, y)$ for any pair $x \in \Sigma_{\varepsilon,i}$ and $y \in \Sigma_{\varepsilon,j}$. Therefore, taking B_i, B_j as in (5.23) with centers in $\{\varphi_\varepsilon \leq \varepsilon^s/2\}$, so that they are contained in $\{\varphi_\varepsilon \leq \varepsilon^s\}$, we can bound the connectedness functional from below as

$$\left(\omega_d r_\varepsilon^d\right)^2 \delta_{ij} \leq \int_{B_i \times B_j} \beta_\varepsilon(\varphi_\varepsilon(x))\beta_\varepsilon(\varphi_\varepsilon(y)) d^{F_\varepsilon \circ \varphi_\varepsilon}(x, y) \mathrm{d}x \mathrm{d}y \leq \mathcal{C}_\varepsilon(\varphi_\varepsilon) \leq C\varepsilon^\kappa,$$

where we have used the fact that $\beta_\varepsilon \circ \varphi_\varepsilon \equiv 1$ inside the set $\{\varphi_\varepsilon \leq \varepsilon^s\}$, so that

$$\delta_{ij} \leq C' \varepsilon^{\kappa - 2d\beta'}. \tag{5.25}$$

We conclude that, for ε small enough, $\delta_{ij} < r_\varepsilon$ for all $i \neq j \in I_\star$, since if it was not the case, we obtain a contradiction in

$$C'' \varepsilon^{\beta'} \leq \delta_{ij} \leq C' \varepsilon^{\kappa - 2d\beta'}, \text{ by taking } (2d+1)\beta' < \kappa,$$

so we chose $\frac{(2d+1)((s+1)-d+1)}{p-d} < \kappa$ and $\frac{(s+1)-d+1}{p-d} < \beta' < \frac{\kappa}{2d+1}$.

Now given any two $i, j \in I_\star$, letting γ be a curve attaining δ_{ij}. If $\Sigma_{\varepsilon,i}$ and $\Sigma_{\varepsilon,j}$ are not in the same connected component of $\{\varphi_\varepsilon \leq 2\varepsilon^s\}$, then there are at least two points x_0, x_1 in this curve γ such that $\varphi_\varepsilon(x_0) = \varphi_\varepsilon(x_1) = \varepsilon^s$. From (5.23), there are balls $B(x_0, r_\varepsilon)$ and $B(x_1, r_\varepsilon)$, since each of then remains in the connected components' of $\{\varphi_\varepsilon \leq 2\varepsilon^s\}$ that contain $\Sigma_{\varepsilon,i}$ and $\Sigma_{\varepsilon,j}$, respectively, so that $B(x_0, r_\varepsilon)$ and $B(x_1, r_\varepsilon)$ must be disjoint. This construction is illustrated in Figure 11.

Also from (5.23), both of these balls must be contained in $\{\frac{3}{4}\varepsilon^s \leq \varphi_\varepsilon \leq 2\varepsilon^s\}$ and, it holds from (H2) that $F_\varepsilon \circ \varphi_\varepsilon \geq \frac{1}{2}$ over $B(x_0, r_\varepsilon)$ and $B(x_1, r_\varepsilon)$. We then have that

$$\delta_{ij} \geq \int_{B(x_0, r_\varepsilon) \cup B(x_1, r_\varepsilon)} F_\varepsilon \circ \varphi_\varepsilon \mathrm{d}\mathscr{H}^1 \llcorner \gamma \geq \frac{1}{2} \mathscr{H}^1 \left(\gamma \cap \left(B(x_0, r_\varepsilon) \cup B(x_1, r_\varepsilon)\right)\right) \geq r_\varepsilon.$$

As this is not true for ε sufficiently small, all $\Sigma_{\varepsilon,i}$ must be contained in the same connected component of $\{\varphi_\varepsilon \leq 2\varepsilon^s\}$.

Let Σ_ε denote this connected component; up to subsequences, we can assume that $\Sigma_\varepsilon \xrightarrow[\varepsilon \to 0]{d_H} \Sigma$. As the Hausdorff limit of connected sets, Σ is itself connected. We can now show that if $\nu_\varepsilon \xrightarrow[\varepsilon \to 0]{\star} \nu$ then ν is concentrated in Σ. Since $\{\varphi_\varepsilon \leq \frac{1}{2}\varepsilon^s\} \subset \Sigma_\varepsilon$, the energy bound $F_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon) \leq C$ implies

$$\nu_\varepsilon\left(\Omega \setminus \Sigma_\varepsilon\right) \leq \frac{2}{\varepsilon^s} \int_{\Omega\setminus\Sigma_\varepsilon} \varphi_\varepsilon \mathrm{d}\nu_\varepsilon \leq 2C\varepsilon^{\ell-s} \xrightarrow[\varepsilon\to 0]{} 0,$$

as we have assumed that $\ell > s$. Therefore, if $x \notin \Sigma$, there is a radius r such that $B_r(x) \cap \Sigma_\varepsilon = \emptyset$ for all $\varepsilon > 0$ small enough. From the previous estimate and the properties of weak convergence, we obtain

$$\nu(B_r(x)) \leq \liminf_{\varepsilon\to 0} \nu_\varepsilon(B_r(x)) \leq \liminf_{\varepsilon\to 0} \nu_\varepsilon(\Omega \setminus \Sigma_\varepsilon) = 0.$$

This show that $\operatorname{supp} \nu \subset \Sigma$.

The rest of the proof is dedicated to show that any cluster point μ of μ_ε is such that $\mu \geq \mathscr{H}^1 \llcorner \Sigma$. We first show in Step 2 that $\mathscr{H}^1(\Sigma) < +\infty$, which will imply that Σ is countably $\mathscr{H}^1$-rectifiable. In Step 3 we use this fact to refine the estimates from Step 2 and conclude. Both of these arguments will be based on the fact that, see [Ambrosio et al., 2000, Thm. 2.56],

$$\theta_1^\star(\mu, x) \overset{\text{def.}}{=} \limsup_{r\to 0} \frac{\mu\left(B_r(x)\right)}{2r} \geq \theta \text{ for } \mathscr{H}^1\text{-a.e. } x \in \Sigma \Longrightarrow \mu \geq \theta \mathscr{H}^1 \llcorner \Sigma. \tag{5.26}$$

Hence, in each of these Steps we prove an estimate of the form: for all $x \in \Sigma$

$$\liminf_{\varepsilon\to 0} \mu_\varepsilon(B_r(x)) \geq \theta 2r, \tag{5.27}$$

for different values of θ. As a consequence, this implies (5.26) by means of classical properties of weak convergence of measures.

Step 2: Given $x_0 \in \Sigma$, fix $r < \min\{\operatorname{dist}(x_0, \partial\Omega), \operatorname{diam}(\Sigma)/2\}$ so that $B_r(x_0) \subset \Omega$. Defining $v_\varepsilon = \varphi_\varepsilon(\varepsilon\cdot)$, we can rewrite $\mu_\varepsilon(B_r(x_0))$ as

$$\begin{aligned} \mu_\varepsilon(B_r(x_0)) &= \frac{\varepsilon^{-d+1}}{\Lambda_{p,d}} \int_0^r \left(\int_{\partial B_\rho(x_0)} \left[\frac{\varepsilon^p}{p} |\nabla \varphi_\varepsilon|^p + \frac{1}{p'}(1-\varphi_\varepsilon)^2 \right] \mathrm{d}\mathscr{H}^{d-1} \right) \mathrm{d}\rho \\ &= \frac{1}{\Lambda_{p,d}} \int_0^r \left(\int_{\partial B\left(x_0, \frac{\rho}{\varepsilon}\right)} \left[\frac{1}{p} |\nabla v_\varepsilon|^p + \frac{1}{p'}(1-v_\varepsilon)^2 \right] \mathrm{d}\mathscr{H}^{d-1} \right) \mathrm{d}\rho. \end{aligned} \tag{5.28}$$

From Fatou's Lemma we obtain that

$$\liminf_{\varepsilon\to 0} \mu_\varepsilon(B_r(x_0)) \geq \frac{1}{\Lambda_{p,d}} \int_0^r \left(\liminf_{\varepsilon\to 0} \int_{\partial B\left(x_0, \frac{\rho}{\varepsilon}\right)} \left[\frac{1}{p} |\nabla v_\varepsilon|^p + \frac{1}{p'}(1-v_\varepsilon)^2 \right] \mathrm{d}\mathscr{H}^{d-1} \right) \mathrm{d}\rho. \tag{5.29}$$

Since the total mass of μ_ε is given by $\mathcal{AT}_p(\varphi_\varepsilon)$, the LHS above remains bounded and hence the liminf on the RHS of (5.29) is finite for a.e. $\rho \in (0, r)$. Hence, it suffices to bound this liminf from below with a constant that holds for almost every $\rho \in (0, r)$, in particular every ρ such that this liminf is finite suffices. To this end, our strategy will be to compare the inner integral in (5.29) with the auxiliary variational problem (5.10) that defines the constant $\Lambda_{p,d}$.

Our first step is to find some $x_\varepsilon \in \partial B_\rho(x_0)$ such that $\varphi_\varepsilon(x_\varepsilon) \leq 2\varepsilon^s$, for a fixed $\rho \in (0, r)$. We can assume that $\Sigma \setminus B_r(x_0)$ is not empty, so from the Hausdorff convergence of Σ_ε to Σ, for ε small enough, there is $z_\varepsilon \in B_\rho(x_0)$ and another point of Σ_ε outside B_r. But since Σ_ε is connected, there is some $x_\varepsilon \in \partial B_\rho(x_0) \cap \Sigma_\varepsilon$, with the desired property.

Up to a translation and a rotation, we may assume that $x_0 = -\rho e_d$ and $x_\varepsilon = 0$; and to define our new function over $\mathbb{R}^{d-1}$, first introduce the notation $\mathbb{R}^d \ni x = (x', x_d) \in \mathbb{R}^{d-1} \times \mathbb{R}$ and define a map Φ_ε from the ball of $\mathbb{R}^{d-1}$, $B_{\mathbb{R}^{d-1}}(0, \frac{\rho}{\varepsilon})$, to the sphere $\partial B_{\mathbb{R}^d}(x_0, \frac{\rho}{\varepsilon})$ as

$$\Phi_\varepsilon(x') \overset{\text{def.}}{=} (x', \phi_\varepsilon(x')) \text{ where } \phi_\varepsilon(x') \overset{\text{def.}}{=} \sqrt{\left(\frac{\rho}{\varepsilon}\right)^2 - |x'|^2} - \frac{\rho}{\varepsilon}.$$

In the sequel, we obtain the new function $\tilde{v}_\varepsilon \in W^{1,p}(B_{\mathbb{R}^{d-1}}(0, \rho/2\varepsilon))$ as

$$\tilde{v}_\varepsilon(x') = v_\varepsilon(\Phi_\varepsilon(x')), \text{ for } x' \in B_{\mathbb{R}^{d-1}}(0, \rho/2\varepsilon).$$

Notice that $\nabla_{x'} \tilde{v}_\varepsilon = \nabla_{x'} \Phi_\varepsilon^\top \nabla_x v_\varepsilon \circ \Phi_\varepsilon$ so that

$$|\nabla_{x'} \tilde{v}_\varepsilon| = |\nabla_{x'} v_\varepsilon + \nabla_{x'} \phi_\varepsilon \partial_d v_\varepsilon| \leq C |\nabla_x v_\varepsilon|,$$

and using the area formula, one obtains that

$$\int_{B_{\mathbb{R}^{d-1}}(0, \frac{\rho}{2\varepsilon})} \left(\frac{1}{p} |\nabla \tilde{v}_\varepsilon|^p + \frac{1}{p'} (1 - \tilde{v}_\varepsilon)^2 \right) \mathrm{d}x' \leq C \int_{\partial B(x_0, \frac{\rho}{\varepsilon})} \left(\frac{1}{p} |\nabla v_\varepsilon|^p + \frac{1}{p'} (1 - v_\varepsilon)^2 \right) \mathrm{d}\mathscr{H}^{d-1}, \tag{5.30}$$

for some constant $C > 0$ depending on the $(d-1)$-Jacobian of Φ_ε, more specifically the quantity

$$\det \left| \nabla_{x'} \Phi_\varepsilon^\top \nabla_{x'} \Phi_\varepsilon \right| = \left(1 + |\nabla_{x'} \phi_\varepsilon(x')|^2\right)^{1/2} = \left(1 + \frac{|\varepsilon x'|^2}{\rho^2 - |\varepsilon x'|^2}\right)^{1/2} = \begin{cases} \geq 1, \\ \leq \sqrt{3}, \end{cases}$$

which can be bounded from above and from below independently of ε for $x' \in B_{\mathbb{R}^{d-1}}(0, \frac{\rho}{2\varepsilon})$. Up to a subsequence, the left hand side of (5.30) is uniformly bounded since we have assumed the liminf in the right hand sind of (5.29) to be finite.

These estimates motivate the definition of a family of variational problems, indexed by ε, that approximate (5.10), the problem whose value defines the constant $\Lambda_{p,d}$, as follows

$$\Lambda_{p,d,\varepsilon} \overset{\text{def.}}{=} \min \left\{ \mathcal{C}_{p,d,\varepsilon}(u) \overset{\text{def.}}{=} \int_{B_{\mathbb{R}^{d-1}}(0, \frac{\rho}{2\varepsilon})} \left(\frac{1}{p} |\nabla w|^p + \frac{1}{p'} (1 - w)^2 \right) \mathrm{d}x : \begin{array}{c} 1 - w \in L^2(\mathbb{R}^{d-1}), \\ \nabla w \in L^p(\mathbb{R}^{d-1}) \\ w(0) \leq 2\varepsilon^s \end{array} \right\}, \tag{5.31}$$

and our goal is to show that $\Lambda_{p,d,\varepsilon} \xrightarrow[\varepsilon \to 0]{} \Lambda_{p,d}$. So let u be optimal for (5.10), then its restriction to $B_{\mathbb{R}^{d-1}}(0, \frac{\rho}{2\varepsilon})$ is admissible and we have

$$\Lambda_{p,d} \geq \int_{B_{\mathbb{R}^{d-1}}(0, \frac{\rho}{2\varepsilon})} \left(\frac{1}{p} |\nabla u|^p + \frac{1}{p'} (1 - u)^2 \right) \mathrm{d}x \geq \Lambda_{p,d,\varepsilon}.$$

As a result, $\Lambda_{p,d,\varepsilon}$ is uniformly bounded and letting w_ε be a solution to (5.31), we will show that it converges locally uniformly to a function w. Indeed, for any $R > 0$ fixed, for ε large enough we have that $B_{\mathbb{R}^{d-1}(0,R)} \subset B_{\mathbb{R}^{d-1}(0,\frac{\rho}{2\varepsilon})}$, so the energy bound on $(w_\varepsilon)_{\varepsilon>0}$ implies that this sequence is Hölder continuous, with the same constant, hence equicontinuous and equibounded from the fact that it converges to 1 at infinity. So, from Ascoli-Arzelà, this sequence converges to some w, uniformly in $B_{\mathbb{R}^{d-1}(0,R)}$. As this also implies the existence of a subsequence whose gradients converge weakly in L^p_{loc}, we get that

$$\liminf_{\varepsilon\to 0} \Lambda_{p,d,\varepsilon} \geq \int_{B_{\mathbb{R}^{d-1}(0,R)}} \left(\frac{1}{p}|\nabla w|^p + \frac{1}{p}(1-w)^2\right) \mathrm{d}x.$$

Since w remains admissible for the problem defining $\Lambda_{p,d}$, taking the supremum on $R > 0$ we get that

$$\liminf_{\varepsilon\to 0} \Lambda_{p,d,\varepsilon} \geq \mathcal{C}_{p,d}(w) \geq \Lambda_{p,d}.$$

The desired convergence $\Lambda_{p,d,\varepsilon} \to \Lambda_{p,d}$ follows and we have that

$$C \liminf_{\varepsilon\to 0} \int_{\partial B\left(x_0,\frac{\rho}{2\varepsilon}\right)} \left[\frac{1}{p}|\nabla v_\varepsilon|^p + \frac{1}{p'}(1-v_\varepsilon)^2\right] \mathrm{d}\mathscr{H}^{d-1} \geq \liminf_{\varepsilon\to 0} \Lambda_{p,d,\varepsilon} = \Lambda_{p,d}.$$

Combining these estimates with (5.29), from the weak convergence of μ_ε to μ, we obtain

$$\begin{aligned}\mu\left(\overline{B_r(x_0)}\right) &\geq \liminf_{\varepsilon\to 0} \mu_\varepsilon(B_r(x_0)) \\ &\geq \frac{1}{C\Lambda_{p,d}} \int_0^r \left(\liminf_{\varepsilon\to 0} \int_{\partial B\left(x_0,\frac{\rho}{\varepsilon}\right)} \left[\frac{1}{p}|\nabla v_\varepsilon|^p + \frac{1}{p'}(1-v_\varepsilon)^2\right] \mathrm{d}\mathscr{H}^{d-1}\right) \mathrm{d}\rho \geq \theta 2r\end{aligned}$$

for $\theta = 1/(2C)$. So $\mu \geq \theta \mathscr{H}^1 \llcorner \Sigma$, from (5.26), and in particular $\mathscr{H}^1(\Sigma) < +\infty$.

Step 3: Now that we know that Σ has finite length, we deduce that it is rectifiable and we can use the rectifiability of Σ to refine the previous estimate, showing that $\theta^\star_1(\mu, x) \geq 1$ for $\mathscr{H}^1$-a.e. $x \in \Sigma$ and from (5.26) conclude that $\mu \geq \mathscr{H}^1 \llcorner \Sigma$ and (5.22) will follow from the properties of weak convergence of measures.

From the rectifiability of Σ it holds that $\mathscr{H}^1$-a.e. $x_0 \in \Sigma$ admits an approximate tangent space. Let x_0 be one of such points and assume, without loss of generality, that $T_{x_0}\Sigma = \mathbb{R}e_d$. So given a small radius r and $\delta \in (0,1)$ close to 1, we consider the cylinder

$$C_{\delta,r}(x_0) \overset{\text{def.}}{=} x_0 + \left\{ x = (x', x_d) : \begin{array}{c} |x'| < \delta r \\ |x_d| < \delta' r \end{array} \right\} \text{ and } \delta' = \sqrt{1-\delta^2}.$$

Our goal is to refine the estimations from the previous step by taking a foliation given by planes orthogonal to $\mathbb{R}e_d$, instead of the spheres. In the sequel, for each t, we define a disc by slicing the cylinder $C_{\delta,r}(x_0)$ with the hyperplane $\{x_d = t\}$.

$$D_t \overset{\text{def.}}{=} C_{\delta,r}(x_0) \cap (\pi_d)^{-1}(x_0 + te_d), \text{ for } t \in (-\delta r, \delta r).$$

We can now obtain a more precise estimate than in Step 2. However, to obtain the point x_ε, such that $\varphi_\varepsilon(x_\varepsilon) \leq 2\varepsilon^s$, the connectedness of Σ_ε was sufficient since we could count on the spherical symmetry of ∂B_ρ. Now, we need a refined argument that will give a point $x_{\varepsilon,t} \in D_t$ such that $\varphi_\varepsilon(x_{\varepsilon,t}) \leq 2\varepsilon^s$, for almost every t. From item (3) of Theorem (2.13), we can find such point for all $t \in [-\delta r, \delta r] \setminus (a_\varepsilon, b_\varepsilon)$ with $b_\varepsilon - a_\varepsilon < 2d_H(\Sigma_\varepsilon, \Sigma)$. Now we can perform a computation analogous to the one presented in Step 2:

$$\begin{aligned}\mu_\varepsilon(B_r(x_0)) &\geq \mu_\varepsilon(C_{r,\delta}(x_0)) \\ &\geq \frac{1}{\Lambda_{p,d}} \int_{[-\delta r,\delta r]\setminus(a_\varepsilon,b_\varepsilon)} \left(\int_{D_t} \left(\frac{\varepsilon^{p-d+1}}{p} |\nabla \varphi_\varepsilon|^p + \frac{\varepsilon^{-d+1}}{p'} (1-\varphi_\varepsilon)^2 \right) \mathrm{d}\mathscr{H}^{d-1} \right) \mathrm{d}t.\end{aligned} \tag{5.32}$$

Let us focus on the L^2 term for the moment. From the energy bound on $\mathcal{AT}_p(\varphi_\varepsilon)$, we know that

$$\int_{-\delta r}^{\delta r} \int_{D_t} (1-\varphi_\varepsilon)^2 \mathrm{d}\mathscr{H}^{d-1} \mathrm{d}t \leq C\varepsilon^{d-1}.$$

Hence, from the converse of the dominated convergence Theorem, up to a subsequence we can assume that, for a.e. $t \in (-\delta r, \delta r)$,

$$\int_{D_t} (1-\varphi_\varepsilon)^2 \mathrm{d}\mathscr{H}^{d-1} \xrightarrow[\varepsilon\to 0]{} 0.$$

Disintegrating once again, we can write the previous term as

$$\int_{D_t} (1-\varphi_\varepsilon)^2 \mathrm{d}\mathscr{H}^{d-1} = \int_{\mathbb{S}^{d-2}} \left(\int_0^{\delta' r} l^{d-2} (1-\varphi_\varepsilon(l\xi + te_d))^2 \mathrm{d}l \right) \mathrm{d}\mathscr{H}^{d-2}(\xi).$$

The same argument gives that, for a.e. $t \in (-\delta r, \delta r)$, for $\mathscr{H}^{d-2}$-a.e. $\xi \in \mathbb{S}^{d-2}$,

$$\varphi_\varepsilon(l\xi + te_d) \xrightarrow[\varepsilon\to 0]{} 1, \text{ for a.e. } l \in [0, \delta' r].$$

Now, fix $t \in (-\delta r, \delta r)$ and $\xi \in \mathbb{S}^{d-2}$ such that the previous limit holds and consider the point $x_{\varepsilon,t} \in D_t$ such that $\varphi_\varepsilon(x_{\varepsilon,t}) \leq 2\varepsilon^s$. Up to a translation, we can assume that $x_{\varepsilon,t} = te_d$ to simplify our notation. We can then define a family of 1-dimensional functions $(f_\varepsilon^{t,\xi})_{\varepsilon>0}$ which we can compare with the optimal 1D profile from Prop. 5.3 such that

$$1 - f_\varepsilon^{t,\xi} \in W_0^{1,p}(\mathbb{R}_+), \text{ and } f_\varepsilon^{t,\xi}(l) = \varphi_\varepsilon\left(te_d + \varepsilon l\xi\right), \text{ for } l \in \left[0, \frac{\bar{l}}{\varepsilon}\right],$$

where $\bar{l}$ is some point close to $\delta' r$ such that $\varphi_\varepsilon(\bar{l}\xi + te_d) = f_{t,\xi,\varepsilon}(\bar{l}) \xrightarrow[\varepsilon\to 0]{} 1$.

As in Step 2, the family $(1 - f_{t,\xi,\varepsilon})_{\varepsilon>0}$ is equibounded in $W^{1,p}(\mathbb{R}_+)$ so that up to a subsequence, it converges weakly to some $f_{t,\xi}$. It also holds that

$$f_{t,\xi}(0) = \lim_{\varepsilon\to 0} \varphi_\varepsilon(x_{\varepsilon,t}) = 0 \text{ and } \lim_{l\to\infty} f_{t,\xi}(l) = \lim_{\varepsilon\to 0} \varphi_\varepsilon(te_d + \delta' r\xi) = 1,$$

so that the limit $f_{t,\xi}$ is admissible for the 1D optimization problem (5.11), and we have that

$$\liminf_{\varepsilon\to 0}\int_0^{\frac{\bar{l}}{\varepsilon}} l^{d-2}\left(\frac{1}{p}|f'_{t,\xi,\varepsilon}|^p+\frac{1}{p'}(1-f_{t,\xi,\varepsilon})^2\right)\mathrm{d}l \geq \liminf_{\varepsilon\to 0}\int_0^{+\infty} l^{d-2}\left(\frac{1}{p}|f'_{t,\xi}|^p+\frac{1}{p'}(1-f_{t,\xi})^2\right)\mathrm{d}l \geq \lambda_{p,d}$$

Let us now gather these ingredients to estimate $\mu(\overline{B_r(x_0)})$. From our previous considerations, it follows that

$$\begin{aligned}
\mu(\overline{B_r(x_0)}) &\geq \liminf_{\varepsilon\to 0}\frac{1}{\Lambda_{p,d}}\int_{[-\delta r,\delta r]\setminus(a_\varepsilon,b_\varepsilon)}\int_{D_t}\left(\frac{\varepsilon^{p-(d-1)}}{p}|\nabla\varphi_\varepsilon|^p+\frac{\varepsilon^{-(d-1)}}{p'}(1-\varphi_\varepsilon)^2\right)\mathrm{d}\mathscr{H}^{d-1}\mathrm{d}t\\
&\geq \liminf_{\varepsilon\to 0}\frac{1}{\Lambda_{p,d}}\int_{[-\delta r,\delta r]\setminus(a_\varepsilon,b_\varepsilon)}\int_{\mathbb{S}^{d-2}}\int_0^{\frac{\bar{l}}{\varepsilon}} l^{d-2}\left(\frac{1}{p}|f'_{t,\xi,\varepsilon}|^p+\frac{1}{p'}(1-f_{t,\xi,\varepsilon})^2\right)\mathrm{d}l\mathrm{d}\mathscr{H}^{d-2}\mathrm{d}t\\
&\geq \frac{1}{\Lambda_{p,d}}\int_{[-\delta r,\delta r]}\int_{\mathbb{S}^{d-2}}\underbrace{\liminf_{\varepsilon\to 0}\left(1_{(a_\varepsilon,b_\varepsilon)^c}(t)\int_0^{\frac{\bar{l}}{\varepsilon}} l^{d-2}\left(\frac{1}{p}|f'_{t,\xi,\varepsilon}|^p+\frac{1}{p'}(1-f_{t,\xi,\varepsilon})^2\right)\mathrm{d}l\right)}_{\geq\lambda_{p,d}}\mathrm{d}\mathscr{H}^{d-2}\mathrm{d}t\\
&\geq \frac{\sigma_{d-2}\lambda_{p,d}}{\Lambda_{p,d}}\delta 2r=\delta 2r.
\end{aligned}$$

Where we have used the fact that $b_\varepsilon - a_\varepsilon < d_H(\Sigma_\varepsilon,\Sigma)\to 0$ and the definition of $\Lambda_{p,d}$.

We conclude that $\theta_1^\star(\mu,x)\geq\delta$, where $\delta\in(0,1)$ is arbitrary. Letting $\delta\to 1$ it follows that $\mu\geq\mathscr{H}^1\llcorner\Sigma$, as well as (5.22). □

For the $\Gamma-\limsup$ inequality, we will use precisely the approximating sequence φ_ε proposed in Theorem 5.6 for a given Σ. On the other hand, if ν is a probability measure concentrated in Σ, it is not hard to construct a sequence of absolutely continuous measures approximating it, it suffices to take a mollification with a smooth kernel. With this construction, we have already proven that $\mathcal{AT}_p(\varphi_\varepsilon)$ converges to $\mathscr{H}^1(\Sigma)$, the only work that is left is to check that the terms $\mathcal{C}_\varepsilon(\varphi_\varepsilon)$ and $\varepsilon^{-\ell}\int\varphi_\varepsilon\mathrm{d}\nu_\varepsilon$ converge to 0.

Theorem 5.11 (A recovery sequence). *Suppose that Ω satisfies Hypothesis (H1) and that $s>1$. Then, for any closed $\Sigma\subset\Omega$ such that $\mathscr{H}^1(\Sigma)<\infty$ and $\nu\in\mathscr{P}(\Sigma)$, there exists a family $(\varphi_\varepsilon,\nu_\varepsilon)_{\varepsilon>0}\subset(1+W_0^{1,p}(\Omega))\times\mathscr{P}_{ac}(\Omega)$ such that*

$$\varphi_\varepsilon\xrightarrow[\varepsilon\to 0]{L^2(\Omega)}1,\quad \nu_\varepsilon\xrightarrow[\varepsilon\to 0]{\star}\nu,\quad \mu_\varepsilon\xrightarrow[\varepsilon\to 0]{\star}\mathscr{H}^1\llcorner\Sigma,\quad \lim_{\varepsilon\to 0}\mathcal{AT}_p(\varphi_\varepsilon)=\mathscr{H}^1(\Sigma)\tag{5.33}$$

and

$$\mathcal{C}_\varepsilon(\varphi_\varepsilon)=0,\ \textit{for all}\ \varepsilon>0\ \textit{and}\ \frac{1}{\varepsilon^\ell}\int_\Omega\varphi_\varepsilon\mathrm{d}\nu_\varepsilon\xrightarrow[\varepsilon\to 0]{}0.\tag{5.34}$$

Proof. Following the construction Thm. 5.6, it suffices to consider $\Sigma \subset \operatorname{int}\Omega$, otherwise we exploit the star-shape property of Ω to find a sequence $\Sigma_n \subset \operatorname{int}\Omega$ such that $\mathscr{H}^1 \llcorner \Sigma_n \xrightarrow[n\to\infty]{} \mathscr{H}^1 \llcorner \Sigma$ and perform a diagonal extraction argument with the familes of phase fields approximating Σ_n to obtain the desired result for Σ.

Assuming $\Sigma \subset \operatorname{int}\Omega$, we recall the following notation from the proof of Thm. 5.6: $d_\Sigma(x) \stackrel{\text{def.}}{=} \operatorname{dist}(x, \Sigma)$ so that $\Sigma_r \stackrel{\text{def.}}{=} \{x \in \Omega : d_\Sigma(x) \leq r\}$.

Let f_p be the optimal profile from Prop. 5.3. If $1 - f_p$ has compact support, the recovery sequence $(\varphi_\varepsilon)_{\varepsilon>0}$ is then defined as in Theorem 5.6 as

$$\varphi_\varepsilon(x) \stackrel{\text{def.}}{=} f_p\left(\frac{d_\Sigma(x) - b_\varepsilon}{\varepsilon}\right), \tag{5.35}$$

where b_ε will be chosen shortly. If f_p reaches 1 only asymptotically, we can replace f_p with a suitable $f_{p,\varepsilon}$ that attains 1 in finite time. Either way, we have from Thm. 5.6 $\varphi_\varepsilon \in 1 + W_0^{1,p}(\Omega)$. Since f_p is increasing and continuous, we have that

$$\{\varphi_\varepsilon \leq 2\varepsilon^s\} = \left\{d_\Sigma(\cdot) \leq b_\varepsilon + \varepsilon f_p^{-1}(2\varepsilon^s)\right\} = \{d_\Sigma(\cdot) \leq 2\varepsilon^s\} \tag{5.36}$$

for $b_\varepsilon \stackrel{\text{def.}}{=} 2\varepsilon^s - \varepsilon f_p^{-1}(2\varepsilon^s)$. From the Hölder continuity of f_p, $b_\varepsilon \geq 0$ for ε small enough and $b_\varepsilon = o(\varepsilon)$.

It follows from Thm. 5.6 that $\varphi_\varepsilon \xrightarrow[\varepsilon\to 0]{L^2(\Omega)} 1$, its corresponding family of diffuse transition measures is such that $\mu_\varepsilon \xrightarrow[\varepsilon\to 0]{\star} \mathscr{H}^1 \llcorner \Sigma$ and $\lim\limits_{\varepsilon\to 0} \mathcal{AT}_p(\varphi_\varepsilon, \nu_\varepsilon) = \mathscr{H}^1(\Sigma)$. For the family $(\nu_\varepsilon)_{\varepsilon>0}$ let $(\eta_t)_{t>0}$ be a sequence of mollifiers $\eta_t = t^{-d}\eta\left(\frac{\cdot}{t}\right)$, with η supported at the unitary ball and set $\nu_\varepsilon \stackrel{\text{def.}}{=} \eta_{c_\varepsilon} \star \nu$, for c_ε small enough so that

$$f_p\left(\frac{c_\varepsilon - b_\varepsilon}{\varepsilon}\right) \leq \varepsilon^{2\ell} \text{ and } 0 \leq c_\varepsilon - b_\varepsilon \xrightarrow[\varepsilon\to 0]{} 0.$$

It then holds that $\nu_\varepsilon \xrightarrow[\varepsilon\to 0]{\star} \nu$, see [Ambrosio et al., 2000, Thm. 2.2], and $\operatorname{supp}\nu_\varepsilon \subset \Sigma_{c_\varepsilon}$.

To finish the proof, it only remains to show (5.34). First notice that as ν_ε is concentrated in Σ_{c_ε}, it holds that

$$\frac{1}{\varepsilon^\ell}\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon = \frac{1}{\varepsilon^\ell}\int_\Omega f_p\left(\frac{d_\Sigma(x) - b_\varepsilon}{\varepsilon}\right)\mathrm{d}\nu_\varepsilon \leq \frac{1}{\varepsilon^\ell} f_p\left(\frac{c_\varepsilon - b_\varepsilon}{\varepsilon}\right) \leq \varepsilon^\ell \xrightarrow[\varepsilon\to 0]{} 0.$$

To compute the term $\mathcal{C}_\varepsilon(\varphi_\varepsilon)$, observe that, from the connectedness of Σ, the set $\{\varphi_\varepsilon \leq 2\varepsilon^s\}$ is connected. Given any two points in $\{\varphi_\varepsilon \leq 2\varepsilon^s\}$, we project each one onto Σ, since Σ is itself connected, we can find a path in Σ connecting the projections. From (5.36), the union of these three arcs forms a path inside $\{\varphi_\varepsilon \leq 2\varepsilon^s\}$ connecting the two original points. Since inside this level set $F_\varepsilon \circ \varphi_\varepsilon \equiv 0$ by construction, for any two points x, y in this level set, we conclude that $d^{F_\varepsilon\circ\varphi_\varepsilon}(x,y) = 0$. Since the connectedness functional can be written as

$$\mathcal{C}_\varepsilon(\varphi_\varepsilon) = \int_{\{\varphi_\varepsilon \leq 2\varepsilon^s\}\times\{\varphi_\varepsilon\leq 2\varepsilon^s\}} \beta_\varepsilon(\varphi_\varepsilon(x))\beta_\varepsilon(\varphi_\varepsilon(y)) d^{F_\varepsilon\circ\varphi_\varepsilon}(x,y)\mathrm{d}x\mathrm{d}y,$$

one has $\mathcal{C}_\varepsilon(\varphi_\varepsilon) = 0$ for all $\varepsilon > 0$. $\square$

3. Approximation of (ADM) and ($\overline{W\mathscr{H}^1}$)

In this section we finally profit from the general analysis done previously to study the problems we were interested in the first place. We first use the properties of the connectedness functional proved in Lemma 5.8 to show existence of minimizers to the phase-field approximations $\mathcal{AD}_\varepsilon$ and $\overline{\mathcal{WH}^1}_\varepsilon$, defined in the introduction in (5.5) and (5.6), and proceed to the proof of our main Theorems 5.1 and 5.2.

Theorem 5.12. *For $\varepsilon > 0$ fixed, both functional $\mathcal{AD}_\varepsilon$ and $\mathcal{WH}^1_\varepsilon$ admit minimizers.*

Proof. First notice that both

$$\inf_{(\nu,\varphi)} \mathcal{AD}_\varepsilon(\nu,\varphi) \text{ and } \inf_{(\alpha,\nu,\varphi)} \mathcal{WH}^1_\varepsilon(\alpha,\nu,\varphi)$$

are finite. This can be seen by considering for instance the recovery sequence from Thm. 5.11 of a segment contained in Ω. Now we can apply the direct method of the calculus of variations to both functionals.

Starting with $\mathcal{AD}_\varepsilon$, let $(\nu_n,\varphi_n)_{n\in\mathbb{N}}$ be a minimizing sequence. Since the infimum of $\mathcal{AD}_\varepsilon$ is finite, it follows that

$$\sup_{n\in\mathbb{N}} \mathcal{AT}_p(\varphi_n) < +\infty,$$

and hence φ_n is bounded in $W^{1,p}(\Omega)$. From Morrey's inequality this sequence is equicontinuous, and it can be taken to be uniformly bounded since the energy can be reduced by thresholding them with the constant 1. From Ascoli-Arzela, it converges uniformly and in $W^{1,p}(\Omega)$, up to a subsequence, to some φ. Similarly, using Banach-Alaoglu we can extract a subsequence such that ν_n converges weakly to some measure ν. We than have that

$$\begin{aligned}
&W_q^p(\rho_0,\nu) = \lim_{n\to\infty} W_q^p(\rho_0,\nu), && \text{from the weak continuity of } W_q\\
&\mathcal{AT}_p(\varphi) \le \liminf_{n\to\infty} \mathcal{AT}_p(\varphi_n), && \text{from the weak convergence in } W^{1,p}(\Omega)\\
&\mathcal{C}_\varepsilon(\nu) = \lim_{n\to\infty} \mathcal{C}_\varepsilon(\nu), && \text{since } \mathcal{C}_\varepsilon \text{ is } C^0 \text{ for uniform convergence, Lemma 5.8}\\
&\int_\Omega \varphi \mathrm{d}\nu = \lim_{n\to\infty} \int_\Omega \varphi_n \mathrm{d}\nu_n, && \text{since } \nu_n \rightharpoonup \text{ in } \mathscr{P}(\Omega) \text{ and } \varphi_n \to \varphi \text{ uniformly.}
\end{aligned}$$

From the fact that (ν_n,φ_n) is a minimizing sequence, it follows that (ν,φ) attains the infimum of $\mathcal{AD}_\varepsilon$.

For $\mathcal{WH}^1_\varepsilon$, a for a minimizing $(\alpha_n,\nu_n,\varphi_n)_{n\in\mathbb{N}}$, a similar argument from the previous case, we can assume up to a subsequence that

$$\begin{aligned}
&\alpha_n \xrightarrow[n\to\infty]{} \alpha, &&\\
&\nu_n \xrightarrow[n\to\infty]{} \nu, && \text{weakly in } L^2(\Omega) \text{ and } \mathscr{P}(\Omega)\\
&\varphi_n \xrightarrow[n\to\infty]{} \varphi, && \text{weakly in } W^{1,p}(\Omega) \text{ and uniformly.}
\end{aligned}$$

And the same continuity and lower semi-continuity property let us conclude that (α,ν,φ) is optimal. $\square$

3.1. Proof of Γ-convergence for average distance minimizers

Now we are in position to prove the Γ-convergence result for the average distance minimizers problem as direct consequence of Theorems 5.10 and 5.11.

Proof of Theorem 5.1: Starting with the $\Gamma-\liminf$, let $(\varphi_\varepsilon, \nu_\varepsilon)_{\varepsilon>0}$ such that $\varphi_\varepsilon \xrightarrow[\varepsilon\to 0]{L^2(\Omega)} \varphi$ and $\nu_\varepsilon \xrightarrow[\varepsilon\to 0]{\star} \nu$. Suppose w.l.o.g. that $\liminf_{\varepsilon\to 0} \mathcal{AD}_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon) < \infty$. Up to taking a subsequence attaining the $\liminf$, it holds that

$$F_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon) \le C \text{ for all } \varepsilon > 0.$$

Once again up to subsequences, it follows from Theorem (5.10) that $\varphi \equiv 1$ and there exists a countably $\mathscr{H}^1$-rectifiable set Σ such that $\operatorname{supp}\nu \subset \Sigma$. This implies that the Steiner tree connecting $\operatorname{supp}\nu$ exists and has a finite length, [Paolini and Stepanov, 2013], since we have that $\mathscr{H}^1_S(\operatorname{supp}\nu) \le \mathscr{H}^1(\Sigma)$. From the lower semi-continuity of the Wasserstein distance w.r.t. weak convergence and the previous properties it holds that

$$\begin{aligned} W_q^q(\rho_0,\nu) + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu) &\le W_q^q(\rho_0,\nu) + \Lambda\mathscr{H}^1(\Sigma) \\ &\le \liminf_{\varepsilon\to 0} \left(W_q^q(\rho_0,\nu_\varepsilon) + \Lambda\mathcal{AT}_p(\varphi_\varepsilon)\right) \\ &\le \liminf_{\varepsilon\to 0} \mathcal{AD}_\varepsilon(\varphi_\varepsilon,\nu_\varepsilon). \end{aligned}$$

For the $\Gamma-\limsup$, for some $\nu \in \mathscr{P}(\Omega)$, suppose that $\mathscr{H}^1_S(\operatorname{supp}\nu) < +\infty$, otherwise there is nothing to prove. This implies that there exists a Steiner tree $\mathcal{S}(\operatorname{supp}\nu)$ attaining the infimum $\mathscr{H}^1_S(\operatorname{supp}\nu)$ with finite length and is hence a countably $\mathscr{H}^1$-rectifiable set. We can then use the recovery sequence proposed in Theorem 5.11 with $\Sigma = \mathcal{S}(\operatorname{supp}\nu)$. As we are in a bounded domain, the Wasserstein distance is continuous for the weak convergence of measures and it holds that

$$\mathcal{AD}_\varepsilon(\varphi_\varepsilon,\nu_\varepsilon) = W_q^q(\rho_0,\nu_\varepsilon) + \Lambda\mathcal{AT}_p(\varphi_\varepsilon) \xrightarrow[\varepsilon\to 0]{} W_q^q(\rho_0,\nu) + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu).$$

To finish the proof we verify that

$$\min_{\Sigma}(\text{ADM}) = \min_{(\nu,\varphi)} \mathcal{AD}.$$

Let Σ and ν be minimizers of (ADM) and $\mathcal{AD}$, respectively, whereas let $\mathcal{S}(\operatorname{supp}\nu)$ and π_Σ denote a Steiner tree of $\operatorname{supp}\nu$ and a measurable selection of the projection operator onto Σ. It holds that

$$\begin{aligned}\min\,(\text{ADM}) &\leq \int_\Omega \operatorname{dist}(x, \mathcal{S}(\operatorname{supp}\nu))^q \mathrm{d}\rho_0(x) + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu)\\ &= \int_\Omega \operatorname{dist}(x, \operatorname{supp}\nu)^q \mathrm{d}\rho_0(x) + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu)\\ &\leq W_q^q(\rho_0,\nu) + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu) = \min\mathcal{AD}\\ &\leq W_q^q(\rho_0, (\pi_\Sigma)_\sharp\rho_0) + \Lambda\mathscr{H}^1(\Sigma)\\ &= \int_\Omega \operatorname{dist}(x,\Sigma)^q \mathrm{d}\rho_0(x) + \Lambda\mathscr{H}^1(\Sigma) = \min\,(\text{ADM}).\end{aligned}$$

It is then clear that if there is ν optimal for $\mathcal{AD}$ then $\mathcal{S}(\operatorname{supp}\nu)$ is optimal for (ADM) and similarly, if is Σ optimal for (ADM) then $(\pi_\Sigma)_\sharp\rho_0$ is optimal for $\mathcal{AD}$. Let us prove the converse of these propositions.

If ν is optimal and cannot be written this way, then $\Sigma = \mathcal{S}(\operatorname{supp}\nu)$ is a minimizer and $W_q^q(\rho_0,\nu) > \int_\Omega \operatorname{dist}(x,\Sigma)^q \mathrm{d}\rho_0$, otherwise it would follow necessarily that $\nu = (\pi_\Sigma)_\sharp\rho_0$, hence

$$\mathcal{AD}(\nu, \varphi \equiv 1) > \int_\Omega \operatorname{dist}(x,\Sigma)^q \mathrm{d}\rho_0 + \Lambda\mathscr{H}^1(\Sigma) \geq \min\mathcal{AD}$$

which contradicts the minimality of ν.

Similarly suppose that Σ is optimal and cannot be written as the Steiner tree of the support of any minimizer ν. We know that $\nu' = (\pi_\Sigma)_\sharp\rho_0$ is a minimizer whose support is contained in Σ, so

$$\min\,(\text{ADM}) = W_q^q(\rho_0,\nu') + \Lambda\mathscr{H}^1(\Sigma) > W_q^q(\rho_0,\nu') + \Lambda\mathscr{H}^1_S(\operatorname{supp}\nu') \geq \min\mathcal{AD},$$

contradicting the minimality of Σ. ■

3.2. Proof of Γ-convergence for ($\overline{W\mathscr{H}^1}$)

Now we move on to the Γ-convergence result for the problem ($\overline{W\mathscr{H}^1}$). We shall use two results from the theory developed for this problem in Section 2 of Chapter 3. Recall that the relaxed problem is stated in terms of the length functional defined as

$$\mathcal{L}(\nu) \overset{\text{def.}}{=} \inf\left\{\alpha \geq 0 : \alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu\right\}, \tag{5.37}$$

which is the l.s.c. relaxation of the functional $\nu_\Sigma \mapsto \mathscr{H}^1(\Sigma)$ if ν_Σ is the probability measure uniformly distributed over a connected set Σ, *i.e.* $\nu_\Sigma = \frac{1}{\mathscr{H}^1(\Sigma)}\mathscr{H}^1 \llcorner \Sigma$ and $+\infty$ otherwise. In the sequel, we recall Lemma 3.7 which is an approximation result for measures ν such that $\mathcal{L}(\nu) < +\infty$.

Lemma 5.13. *Let $\nu \in \mathscr{P}(\Omega)$ such that $\mathcal{L}(\nu) < \infty$, there exists a sequence of connected sets $(\Sigma_n)_{n\in\mathbb{N}}$ such that*

- $\Sigma_n \xrightarrow[n\to\infty]{d_H} \Sigma$ *and* $\mathscr{H}^1(\Sigma_n) \xrightarrow[n\to\infty]{} \mathcal{L}(\nu)$*;*
- $\nu_{\Sigma_n} \overset{\text{def.}}{=} \dfrac{1}{\mathscr{H}^1(\Sigma_n)}\mathscr{H}^1 \llcorner \Sigma_n \xrightarrow[n\to\infty]{} \nu.$

As we have discussed, our Γ-convergence result actually approximates the relaxed problem ($\overline{W\mathscr{H}^1}$) instead of ($W\mathscr{H}^1$), but we cannot expect anything better since Γ-limits are always l.s.c. [Attouch et al., 2014], which is not the case for the energy of the orginal problem.

Proof of Theorem 5.2:

Let us start with the $\Gamma - \liminf$ inequality. Consider $(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon) \xrightarrow[\varepsilon\to 0]{} (\alpha, \nu, \varphi)$ in the product topology of $\mathbb{R}$, weak convergence of measures and strong $L^2(\Omega)$ convergence. Up to extracting a subsequence in ε, we can suppose w.l.o.g. that there is a constant $C > 0$ such that $\mathcal{WH}^1_\varepsilon(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon) \leq C$, otherwise there is nothing to prove. This can be done by first assuming the $\liminf$ is finite and taking a subsequence attaining it. Clearly from the continuity of the Wasserstein distance, it holds that

$$W_q^q(\rho_0, \nu_\varepsilon) + \Lambda\alpha_\varepsilon \xrightarrow[\varepsilon\to 0]{} W_q^q(\rho_0, \nu) + \Lambda\alpha,$$

hence to conclude, we must verify the constraints $\varphi \equiv 1$ and $\alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu$, where $\operatorname{supp}\nu$ is connected.

Recalling that μ_ε is the diffuse transition measure defined at (5.7) with the function φ_ε, notice that

$$\|\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon\|_{\mathcal{M}(\Omega)} = \|\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon\|_{L^1(\Omega)} \leq |\Omega|^{1/2}\,\|\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon\|_{L^2(\Omega)} \leq (C|\Omega|\varepsilon)^{1/2}.$$

Therefore, $\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon$ converge to 0 strongly, and hence $\mu_\varepsilon \xrightarrow[\varepsilon\to 0]{\star} \alpha\nu$. It also holds that

$$\mathcal{AT}_p(\varphi_\varepsilon) = \mu_\varepsilon(\Omega) \leq \|\alpha_\varepsilon\nu_\varepsilon - \mu_\varepsilon\|_{L^1(\Omega)} + \|\alpha_\varepsilon\nu_\varepsilon\|_{L^1(\Omega)} \leq \alpha_\varepsilon + (C|\Omega|\varepsilon)^{1/2},$$

so we can find another constant $C' > 0$ such that

$$F_\varepsilon(\varphi_\varepsilon, \nu_\varepsilon) = \mathcal{AT}_p(\varphi_\varepsilon) + \frac{1}{\varepsilon^\kappa}\mathcal{C}_\varepsilon(\varphi_\varepsilon) + \frac{1}{\varepsilon^\ell}\int_\Omega \varphi_\varepsilon \mathrm{d}\nu_\varepsilon \leq C' \text{ for all } \varepsilon > 0.$$

From Theorem (5.10), the sequence φ_ε converges strongly in $L^2(\Omega)$ to the constant 1, and there is a connected, countably $\mathscr{H}^1$-rectifiable set Σ such that $\operatorname{supp}\nu \subset \Sigma \subset \operatorname{supp}\mu$ and $\mu = \alpha\nu \geq \mathscr{H}^1 \llcorner \Sigma$. Hence, since $\mu = \alpha\nu$ and $\alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu$ we have that $\operatorname{supp}\nu = \Sigma = \operatorname{supp}\mu$ and $\alpha \geq \mathcal{L}(\nu)$, implying

$$\begin{aligned} W_q^q(\rho_0, \nu) + \Lambda\mathcal{L}(\nu) \leq W_q^q(\rho_0, \nu) + \Lambda\alpha &\leq \liminf_{\varepsilon\to 0}\left(W_q^q(\rho_0, \nu_\varepsilon) + \Lambda\alpha_\varepsilon\right) \\ &\leq \liminf_{\varepsilon\to 0} \mathcal{WH}^1(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon). \end{aligned}$$

Moving on to the construction of the recovery sequence, our strategy is to combine Lemma 5.13 with the fact that the diffuse transition measures μ_ε related to the recovery sequence from Theorem (5.11) converge to uniform measures of the form $\mathscr{H}^1 \llcorner \Sigma$.

Given $\alpha\nu \geq \mathscr{H}^1 \llcorner \operatorname{supp}\nu$ such that ν is a probability measure, $\operatorname{supp}\nu$ is connected and $0 < \alpha = \mathcal{L}(\nu) < +\infty$, since if $\alpha = \mathcal{L}(\nu) = 0$ then ν is concentrated in a single point. For clarity of notation set $\Sigma = \operatorname{supp}\nu$ and let Σ_n be the approximating sequence from Lemma (5.13). For each $n \in \mathbb{N}$, construct the recovery sequence $(\varphi_{n,\varepsilon})_{\varepsilon>0}$ from Theorem (5.11), built from the set Σ_n. From the construction, $\mathcal{C}_\varepsilon(\varphi_{n,\varepsilon}) = 0$ and it holds that

$$\mu_{n,\varepsilon} \xrightarrow[\varepsilon\to 0]{\star} \mathscr{H}^1 \llcorner \Sigma_n, \text{ for all } n \in \mathbb{N},$$

where $\mu_{n,\varepsilon}$ is the diffuse transition measure associated with $\varphi_{n,\varepsilon}$. Define

$$\alpha_{n,\varepsilon} \stackrel{\text{def.}}{=} \mu_{n,\varepsilon}(\Omega) \text{ and } \nu_{n,\varepsilon} \stackrel{\text{def.}}{=} \frac{1}{\alpha_{n,\varepsilon}}\mu_{n,\varepsilon} \in \mathscr{P}(\Omega),$$

so that

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\alpha_{n,\varepsilon} = \lim_{n\to\infty}\mathscr{H}^1(\Sigma_n) = \alpha,$$
$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\nu_{n,\varepsilon} = \lim_{n\to\infty}\frac{1}{\mathscr{H}^1(\Sigma_n)}\mathscr{H}^1 \llcorner \Sigma_n = \nu.$$

With a diagonal argument, we select a decreasing sequence $\varepsilon_n \to 0$ such that

$$\mathcal{C}_{\varepsilon_n}(\varphi_{n,\varepsilon_n}) = \|\alpha_{n,\varepsilon_n}\nu_{n,\varepsilon_n} - \mu_{n,\varepsilon_n}\|_{L^2(\Omega)} = 0 \text{ and } \alpha_{n,\varepsilon_n}, \nu_{n,\varepsilon_n} \to \alpha, \nu.$$

Our recovery sequence is then defined as

$$(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon) \stackrel{\text{def.}}{=} (\alpha_{n,\varepsilon_n}, \nu_{n,\varepsilon_n}, \varphi_{n,\varepsilon_n}) \text{ if } \varepsilon_n \leq \varepsilon < \varepsilon_{n-1},$$

so the continuity of the Wasserstein distance yields

$$\mathcal{WH}^1_\varepsilon(\alpha_\varepsilon, \nu_\varepsilon, \varphi_\varepsilon) \xrightarrow[\varepsilon\to 0]{} W_q^q(\rho_0, \nu) + \Lambda\mathcal{L}(\nu),$$

and the result follows.

The fact that, whenever ρ_0 does not charge 1-dimensional sets, cluster points of minimizers of $\mathcal{WH}^1_\varepsilon$ converge to a measure ν_Σ, where Σ minimizes the original problem ($W\mathscr{H}^1$) follows from the fundamental property of Γ-convergence and the fact that under these conditions, from the existence Theorem for the orginal problem 3.24, any solution to the relaxation ($\overline{W\mathscr{H}^1}$) is in fact a solution to ($W\mathscr{H}^1$). ■

4. Conclusion

In this Chapter we have discussed a general approach to define phase-field approximations for the Wasserstein-$\mathscr{H}^1$ problem as well as the average distance minimizers problem, the key ingredients being the interplay between the Ambrosio-Tortorelli functional $\mathcal{AT}_p$ and the connectivity functional $\mathcal{C}_\varepsilon$ in the general Thm. 5.10 and the approximation properties of the diffuse transition measures in the weak topology from Thm. 5.6. These results can be easily applied to other approximation schemes. One example that would not be as simple is the general Monge-Kantorovitch model proposed in [Buttazzo and Stepanov, 2003] since the network Σ appears in the definition of a metric that is used inside an optimization problem, a 1-Wasserstein distance.

Many questions are left unanswered, on the theoretical side, recalling that the original model of Modica and Mortola was motivated by the Cahn-Hillard equations, one could ask if there is a connection between a modified model with a p-Laplacian and a suitable family of p-elliptic functionals as $\mathcal{AT}_p$ employed in the present work. Also inspired on previous phase-field models, one could ask if optimal or almost optimal phase-fields enjoy some sort of equipartition of energy. We forced this to be the case in the recovery sequence constructed in Thm. 5.6, but it might be a more general phenomenon.

Numerical implementations of the approximations will be investigated in future work and might serve as a source of conjectures for theoretical questions and qualitative properties about both the Wasserstein-$\mathscr{H}^1$ and the average distance minimization problems.

Chapter 6

The Wasserstein gradient flow of the total variation

Contents

1. Introduction

The Wasserstein gradient flow of the total variation functional has been studied in a series of recent papers [Burger et al., 2012, Benning et al., 2013, Carlier and Poon, 2019], for applications in image processing. In the present chapter, we revisit the work of Carlier & Poon [Carlier and Poon, 2019] and derive Euler-Lagrange equations for the problem: given $\Omega \subset \mathbb{R}^d$ open, bounded and convex, $\tau > 0$ and an absolutely continuous probability measure $\rho_0 \in \mathscr{P}(\Omega)$

$$\inf_{\rho \in \mathscr{P}(\Omega)} \mathrm{TV}(\rho) + \frac{1}{2\tau} W_2^2(\rho_0, \rho), \tag{TV-W}$$

where τ is interpreted as a time discretization parameter for an implicit Euler scheme, as we shall see below.

The *total variation functional* of a Radon measure $\rho \in \mathcal{M}(\Omega)$ is defined as

$$\mathrm{TV}(\rho) = \sup \left\{ \int_\Omega \operatorname{div} z \mathrm{d}\rho : z \in C_c^1\left(\Omega; \mathbb{R}^d\right), \|z\|_\infty \leq 1 \right\}, \tag{TV}$$

which is not to be mistaken here with the *total variation measure* $|\mu|$ of a Radon measure μ or its *total variation norm* $|\mu|(\Omega)$. We recall that $\mathrm{BV}(\Omega)$ denotes the subspace of functions $u \in L^1(\Omega)$ whose weak derivative Du is a *finite Radon measure*. It can also be characterized as the L^1 functions such that $\mathrm{TV}(u) < \infty$, where $\mathrm{TV}(u)$ should be understood as in (TV) with the measure $u\mathcal{L}^d \llcorner \Omega$, and it holds that $\mathrm{TV}(u) = |Du|(\Omega)$. As $\mathrm{BV}(\mathbb{R}^d) \hookrightarrow L^{\frac{d}{d-1}}(\mathbb{R}^d)$, solutions to (TV-W) are also absolutely continuous w.r.t. the Lebesgue measure. Therefore, w.l.o.g. we can minimize on $L^{\frac{d}{d-1}}(\Omega)$, which is a reflexive Banach space. In addition, a function ρ will have finite energy only if $\rho \in \mathscr{P}(\Omega)$.

The data term is given by the Wasserstein distance, which we recall, is defined through the value of the optimal transportation problem (see [Santambrogio, 2015])

$$W_2^2(\mu, \nu) \stackrel{\text{def.}}{=} \min_{\gamma \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} |x-y|^2 \mathrm{d}\gamma = \sup_{\substack{\varphi, \psi \in C_b(\Omega) \\ \varphi \oplus \psi \leq |x-y|^2}} \int_\Omega \varphi \mathrm{d}\mu + \int_\Omega \psi \mathrm{d}\nu, \tag{6.1}$$

where the minimum is taken over all the probability measures on $\Omega \times \Omega$ whose marginals are μ and ν. An optimal pair (φ, ψ) for the dual problem is referred to as Kantorovitch potentials.

Using total variation as regularization was suggested in [Rudin et al., 1992] with a L^2 data term for the *Rudin-Osher-Fatemi problem*

$$\inf_{u \in L^2(\Omega)} \mathrm{TV}(u) + \frac{1}{2\lambda} \|u - g\|_{L^2(\Omega)}^2, \tag{ROF}$$

see [Chambolle et al., 2010] for an overview. Other data terms were considered to better model the oscillatory behavior of the noise [Meyer, 2001, Lieu and Vese, 2008]. More recently (for instance [Cuturi and Peyré, 2018]) Wasserstein energies have shown success in the imaging community, the model (TV-W) being used for image denoising in [Benning et al., 2013, Burger et al., 2012].

Existence and uniqueness of solutions for (TV-W) follow from the direct method in the calculus of variations, and the strict convexity of $W_2^2(\rho_0, \cdot)$ whenever ρ_0 is absolutely continuous, see [Santambrogio, 2015, Prop. 7.19]. However, it is not easy to compute the subdifferential of the sum, which makes the derivation of the Euler-Lagrange equations not trivial.

In [Carlier and Poon, 2019], the authors studied the gradient flow scheme defined by the successive iterations of (TV-W), and following the seminal work [Jordan et al., 1998] they showed that, in dimension 1 as the parameter $\tau \to 0$, the discrete scheme converges to the solution of a fourth order PDE. They used an entropic regularization approach, followed by a Γ-convergence argument, to derive an Euler-Lagrange equation, which states that there exists a Kantorovitch potential ψ_1 coinciding with some $\operatorname{div} z \in \partial\,\mathrm{TV}(\rho_1)$ in the set $\{\rho_1 > 0\}$. On $\{\rho_1 = 0\}$, these quantities are related through a bounded Lagrange multiplier β associated with the nonnegativity constraint $\rho_1 \geq 0$.

In this work we propose an alternative way to derive the Euler-Lagrange equations which relies on the well established properties of solutions of (ROF) and shows further regularity of the quantities $\operatorname{div} z, \beta$.

Theorem 6.1. *For any $\rho_0 \in L^1(\Omega) \cap \mathscr{P}(\Omega)$, let ρ_1 be the unique minimizer of (TV-W). The following hold.*

1. *There is a vector field $z \in H_0^1(\operatorname{div}; \Omega) \cap L^\infty(\Omega; \mathbb{R}^d)$ and a Lagrange multiplier $\beta \geq 0$ such that*
$$\begin{cases} \operatorname{div} z + \dfrac{\psi_1}{\tau} = \beta, & \text{a.e. in } \Omega \\ z \cdot \nu = 0, & \text{on } \partial\Omega \\ \beta\rho_1 = 0, & \text{a.e. in } \Omega \\ z \cdot D\rho_1 = |D\rho_1|, & \|z\|_\infty \leq 1, \end{cases} \tag{TVW-EL}$$
where ψ_1 is a Kantorovitch potential associated with ρ_1.

2. *The Lagrange multiplier β is the unique solution to (ROF) with $\lambda = 1$ and $g = \psi_1/\tau$.*

3. *The functions $\operatorname{div} z, \psi_1$ and β are Lipschitz continuous.*

2. The Euler-Lagrange equation

In this section we derive optimality conditions for (TV-W) by relying on some properties of the subdifferential of TV in an appropriate space and on the optimality conditions for (ROF). In the following, we let X and $X^\star$ be duality-paired spaces and $f : X \to \mathbb{R} \cup \{\infty\}$ be a convex function, the subdifferential of f on X is given by the set

$$\partial_X f(u) \overset{\text{def.}}{=} \{p \in X^\star : f(v) \geq f(u) + \langle p, v - u \rangle \text{ , for all } v \in X\} . \tag{6.2}$$

2.1. The Rudin-Osher-Fatemi (ROF) problem

In this subsection we prove some well-known properties about problem (ROF) and the total variation functional that are used throughout the text. It is easy to see that (ROF) has a unique solution since the functional $u \mapsto \mathrm{TV}(u) + \|u-g\|^2_{L^2(\Omega)}$ is strongly convex and l.s.c. in the weak topology of $L^2(\Omega)$, a standard application of the direct method of the calculus of variations gives the result. In particular, this means that the Euler-Lagrange equations for this problem are necessary and sufficient conditions for optimality.

Therefore, let u be such minimizer and $v \in L^2(\Omega)$ arbitrary. Then comparing the energies of u and v we have

$$\begin{aligned}\lambda\left(\mathrm{TV}(v) - \mathrm{TV}(u)\right) &\geq \frac{1}{2}\int_\Omega \left((u-g)^2 - (v-g)^2\right)\mathrm{d}x \\ &= \int_\Omega (v-u)(g-u)\mathrm{d}x - \frac{1}{2}\int_\Omega (u-v)^2 \mathrm{d}x.\end{aligned}$$

Now taking $u + t(v-u)$, for $t \in [0,1]$, as test function in the previous inequality we obtain that

$$\lambda\left(\mathrm{TV}(u+t(v-u)) - \mathrm{TV}(u)\right) - t\int_\Omega (v-u)(g-u)\mathrm{d}x \geq -\frac{t^2}{2}\int_\Omega (u-v)^2\mathrm{d}x, \quad (6.3)$$

which implies that

$$\lambda\left(\mathrm{TV}(u+t(v-u)) - \mathrm{TV}(u)\right) - t\int_\Omega (v-u)(g-u)\mathrm{d}x \geq 0. \quad (6.4)$$

Taking $t = 1$, we actually obtain the following Euler-Lagrange equation

$$\frac{g-u}{\lambda} \in \partial\,\mathrm{TV}(u). \quad (6.5)$$

To extract further information, we need to characterize the subdifferential of TV, which we will do in the next section in the more general case of one-homogeneous functionals. We shall start with general properties of 1-homogeneous functionals and use it to derive the characterization of the subdifferential of TV from Proposition 6.5.

Lemma 6.2. *[Aliprantis and Border, 2006, Thm. 7.57] Let X be a reflexive Banach space and $J : X \to \mathbb{R} \cup \{+\infty\}$ be a convex, positively one-homogeneous functional,* i.e. $J(\lambda u) = |\lambda| J(u)$ *for all $\lambda \in \mathbb{R}$ and $u \in X$. Then*

$$\partial J(u) = \left\{p \in \partial J(0) : \langle p, u\rangle = J(u)\right\}. \quad (6.6)$$

In particular, if J is the support function of a convex set $C \subset X^\star$

$$J(u) = \sup_{p\in C} \langle p, u\rangle\,, \quad (6.7)$$

then the subdifferential $\partial J(0) = \overline{\mathrm{conv} C}$.

Proof. Setting $A \overset{\text{def.}}{=} \{p \in \partial J(0) : \langle p, u\rangle = J(u)\}$, we claim this set is contained in $\partial J(u)$. Take $p \in A$, so that since $J(0) = 0$ by one-homogeneity, and as $p \in \partial J(0)$, we have

$$J(v) \geq \langle p, v\rangle\,, \text{ for all } v \in X.$$

Summing and subtracting $J(u) = \langle p, u\rangle$ we obtain the subdifferential inequality.

For the converse inclusion, let $p \in \partial J(u)$, and taking $v = 0$ in the sub-differential inequality, we have

$$J(u) \leq \langle p, u\rangle\,.$$

Now it suffices to prove that $\langle p, v\rangle \leq J(v)$ for all $v \in X$, as this is equivalent to $p \in \partial J(0)$ and implies the equality $J(u) = \langle p, u\rangle$. By contradiction, if there is some v such that $J(v) - \langle p, v\rangle < 0$, take λv instead of v for $\lambda > 0$, then by the one-homogeneity we have

$$\lambda \underbrace{(J(v) - \langle p, v\rangle)}_{<0} \geq J(u) - \langle p, u\rangle\,.$$

Taking λ large enough we arrive at a contradiction.

For the second statement, notice that it suffices to prove that $C \subset \partial J(0) \subset \overline{C}$, since $\partial J(0)$ being a subdifferential implies that it is already weak-$\star$ closed and convex.

The first inclusion follows from the definition; we pass to the second. Suppose there is some $p_0 \in \overline{C} \setminus \partial J(0)$, then $\{p_0\}$ is compact, convex and disjoint from $\partial J(0)$. Using the convex separation Theorem in $X^\star$, see [Bourbaki, 2003, Chap. 2.5 Prop. 4] or [Brezis, 2010, Problem 9, page 447], we can find a continuous linear functional in the weak-$\star$ topology, that is some $v \in X$, that strictly separates $\{p_0\}$ and $\partial J(0)$ giving

$$\langle p_0, v\rangle > \sup_{p \in \partial J(0)} \langle p_0, v\rangle \geq J(v).$$

We then arrive at a contradiction. □

Recalling the definition of the TV functional as

$$\mathrm{TV}(u) = \sup_{p \in \mathcal{K}} \int_\Omega p(x)u(x)\mathrm{d}x, \text{ where } \mathcal{K} \overset{\text{def.}}{=} \left\{ p = \mathrm{div}\phi : \begin{array}{c} \phi \in C_c^1(\Omega; \mathbb{R}^N) \\ \|\phi\|_\infty \leq 1 \end{array} \right\}, \tag{6.8}$$

from the previous Lemma 6.2 it holds that $\partial_{L^2} \mathrm{TV}(0) = \overline{\mathcal{K}}$, with the closure being taken with respect to the weak topology of $L^2(\Omega)$. Hence, for Ω convex and bounded, the subdifferential of TV in L^2 assumes the form

$$\partial\,\mathrm{TV}(u) = \left\{ p = -\,\mathrm{div}\,z : \begin{array}{c} z \in H_0^1(\mathrm{div}, \Omega),\ \|z\|_\infty \leq 1 \\ \mathrm{TV}(u) = \int_\Omega p(x)u(x)\mathrm{d}x \end{array} \right\},$$

where $H_0^1(\mathrm{div}; \Omega)$ denotes the closure of $C_c^\infty(\Omega; \mathbb{R}^d)$ with respect to the norm $\|z\|^2_{H^1(\mathrm{div})} = \|z\|^2_{L^2(\Omega)} + \|\mathrm{div}\,z\|^2_{L^2(\Omega)}$.

Next we prove another property of the subdifferential of TV based on the coarea formula (see [Ambrosio et al., 2000, Thm. 3.40]): for any $u \in \mathrm{BV}(\Omega)$ it holds that

$$\mathrm{TV}(u) = \int_{\mathbb{R}} \mathrm{Per}(\{u \geq t\})\mathrm{d}t. \tag{6.9}$$

Lemma 6.3. *For any $u \in \mathrm{BV}(\Omega)$ and $p \in \partial\,\mathrm{TV}(u)$, it holds that*

$$p \in \partial\,\mathrm{TV}(u^+), \quad -p \in \partial\,\mathrm{TV}(u^-).$$

Proof. We can show that for a.e. $s \in \mathbb{R}$ it holds that $\mathrm{Per}\left(\{u > s\}\right) = \mathrm{Per}\left(\{u \geq s\}\right) = \mathrm{Per}\left(\{u \geq s\}^c\right)$. In addition, by Lemma 6.2, if $p \in \partial\,\mathrm{TV}(u)$, one has

$$\begin{aligned} \mathrm{TV}(u) &= \int_0^{+\infty} \mathrm{Per}(\{u^+ > s\})\mathrm{d}s + \int_0^{+\infty} \mathrm{Per}(\{u^- > s\})\mathrm{d}s \\ = \int_\Omega p(x)u(x)\mathrm{d}x &= \int_0^{+\infty}\int_{\{u^+>s\}} p\mathrm{d}x\mathrm{d}s - \int_0^{+\infty}\int_{\{u^->s\}} p\mathrm{d}x\mathrm{d}s. \end{aligned}$$

In particular, we have

$$\begin{aligned} &\int_0^{+\infty} \underbrace{\left(\mathrm{Per}(\{u^+ > s\}) - \int_{\{u^+>s\}} p(x)\mathrm{d}x\right)}_{\geq 0}\mathrm{d}s \\ &\qquad\qquad = \int_0^{+\infty} \underbrace{\left(-\mathrm{Per}(\{u^- > s\}) - \int_{\{u^->s\}} p(x)\mathrm{d}x\right)}_{\leq 0}\mathrm{d}s. \end{aligned}$$

Since the integrands on each side have constant and opposite signs, for a.e. $s \in \mathbb{R}$

$$\mathrm{Per}(\{u^+ > s\}) = \int_{\{u^+>s\}} p(x)\mathrm{d}x \text{ and } \mathrm{Per}(\{u^- > s\}) = \int_{\{u^->s\}} -p(x)\mathrm{d}x.$$

Integrating over $[0, +\infty]$, the co-area formula gives

$$\mathrm{TV}(u^+) = \int_\Omega p(x)u^+(x)\mathrm{d}x \text{ and } \mathrm{TV}(u^-) = \int_\Omega -p(x)u^-(x)\mathrm{d}x,$$

and the result follows from Lemma 6.2. □

The final property of the solutions of (ROF) we will require is as follows: if u solves (ROF), then its positive part solves the same problem with an additional positivity constraint (see also [Chambolle, 2004, Lemma A.1]).

Theorem 6.4. *A function u is the solution of (ROF) if, and only if,*

$$u - g \in \lambda\partial_{L^2}\,\mathrm{TV}(u).$$

In addition, it holds that

$$u^+ \in \operatorname*{argmin}_{v \geq 0} \mathrm{TV}(v) + \frac{1}{2\lambda}\int_\Omega |v(x) - g(x)|^2\mathrm{d}x. \tag{6.10}$$

Proof. The first property is a direct consequence of Fermat's rule and the fact that the L^2 norm is smooth. For the second, since the constrained problem remains strictly convex, it suffices to show that

$$g - u^+ \in \partial_{L^2} \left(\lambda \operatorname{TV} + \chi_{v\geq 0}\right)(u^+). \tag{6.11}$$

Writing $u = u^+ - u^-$ with both $u^+, u^- \geq 0$, from Lemma (6.3) and the Euler-Lagrange equation of (ROF):

$$g - u \in \lambda \partial_{L^2} \operatorname{TV}(u),$$

we know that

$$g - u^+ = -u^- + \lambda p, \text{ with } p \in \partial_{L^2} \operatorname{TV}(u^+).$$

So it suffices to show that $-u^- + \lambda \partial_{L^2} \operatorname{TV}(u^+) \subset \partial_{L^2} \left(\lambda \operatorname{TV} + \chi_{v\geq 0}\right)(u^+)$. Take some $v \geq 0$ and since $p \in \partial_{L^2} \operatorname{TV}(u^+)$, we have

$$\lambda \left(\operatorname{TV}(v) - \operatorname{TV}(u^+) - \left\langle p, v - u^+ \right\rangle\right) \geq 0 \geq - \left\langle u^-, v \right\rangle.$$

Rearranging the terms we obtain the desired relation $\lambda p - u^- \in \partial_{L^2} \left(\lambda \operatorname{TV} + \chi_{v\geq 0}\right)(u^+)$. □

The results of this section are summarized in the following.

Proposition 6.5. *[Bredies and Holler, 2016, Chambolle et al., 2010, Mercier, 2018] If $u \in \operatorname{BV}(\Omega)$, then the subdifferential of TV at u assumes the form*

$$\partial_{L^2} \operatorname{TV}(u) = \left\{ p \in L^2(\Omega) : \begin{array}{c} p = -\operatorname{div} z, \ z \in H^1_0(\operatorname{div}; \Omega), \\ \|z\|_\infty \leq 1, \ |Du| = z \cdot Du \end{array} \right\}.$$

If $p \in \partial_{L^2} \operatorname{TV}(u)$, then

$$p \in \partial_{L^2} \operatorname{TV}(u^+), \quad -p \in \partial_{L^2} \operatorname{TV}(u^-).$$

If in addition u solves (ROF), then

1. *u^+ solves (ROF) with the constraint $u \geq 0$;*
2. *it holds that*
$$0 \in \frac{u - g}{\lambda} + \partial_{L^2} \operatorname{TV}(u), \tag{6.12}$$
and conversely, if u satisfies (6.12), u minimizes (ROF);
3. *for Ω convex, if g is uniformly continuous with modulus of continuity ω, then u has the same modulus of continuity.*

2.2. Derivation of the Euler-Lagrange equation for (TV-W)

Unless otherwise stated, we consider in the sequel $X = L^{\frac{d}{d-1}}(\Omega)$, $X^\star = L^d(\Omega)$ and we drop the index X in the notation ∂_X. Under certain regularity, one can see the Kantorovitch potentials as the first variation of the Wasserstein distance, [Santambrogio, 2015]. As a consequence, Fermat's rule $0 \in \partial\left(W_2^2(\rho_0, \cdot) + \mathrm{TV}(\cdot)\right)(\rho_1)$ assumes the following form.

Lemma 6.6. *Let ρ_1 be the unique minimizer of (TV-W), then there exists a Kantorovitch potential ψ_1 associated to ρ_1 such that*

$$-\frac{\psi_1}{\tau} \in \partial\left(\mathrm{TV} + \chi_{\mathscr{P}(\Omega)}\right)(\rho_1). \tag{6.13}$$

Proof. For simplicity, we assume $\tau = 1$. Take $\rho \in \mathrm{BV}(\Omega) \cap \mathscr{P}(\Omega)$ and define $\rho_t \overset{\text{def.}}{=} \rho + t(\rho_1 - \rho)$. Since $\overline{\Omega}$ is compact, the sup in (6.1) admits a maximizer [Santambrogio, 2015, Prop. 1.11]. Let φ_t, ψ_t denote a pair of Kantorovitch potentials between ρ_0 and ρ_t. From the optimality of ρ_1 it follows

$$\begin{aligned}
&\frac{1}{2}W_2^2(\rho_0, \rho_1) + \mathrm{TV}(\rho_1) \leq \int_\Omega \varphi_t \mathrm{d}\rho_0 + \int_\Omega \psi_t \mathrm{d}\rho_t + \mathrm{TV}(\rho_t) \\
&\leq \int_\Omega \varphi_t \mathrm{d}\rho_0 + \int_\Omega \psi_t \mathrm{d}\rho_1 + \mathrm{TV}(\rho_1) + (1-t)\left(\int_\Omega \psi_t \mathrm{d}(\rho - \rho_1) + \mathrm{TV}(\rho) - \mathrm{TV}(\rho_1)\right) \\
&\leq \frac{1}{2}W_2^2(\rho_0, \rho_1) + \mathrm{TV}(\rho_1) + (1-t)\left(\int_\Omega \psi_t \mathrm{d}(\rho - \rho_1) + \mathrm{TV}(\rho) - \mathrm{TV}(\rho_1)\right).
\end{aligned}$$

Hence, $-\psi_t \in \partial\left(\mathrm{TV} + \chi_{\mathscr{P}(\Omega)}\right)(\rho_1)$ for all $t \in (0, 1)$. Notice that as the optimal transport map from ρ_0 to ρ_t is given by $T_t = \mathrm{id} - \nabla\psi_t$ and assumes values in the bounded set Ω, the family $(\psi_t)_{t\in[0,1]}$ is uniformly Lipschitz so that by Arzelà-Ascoli's Theorem ψ_t converges uniformly to ψ_1 as t goes to 1 (see also [Santambrogio, 2015, Thm. 1.52]). Therefore, $-\psi_1 \in \partial\left(\mathrm{TV} + \chi_{\mathscr{P}(\Omega)}\right)(\rho_1)$. $\square$

With these results we can prove Theorem 6.1.

Proof of Theorem 6.1. Here, to simplify, we still assume $\tau = 1$. The subdifferential inclusion (6.13) is conceptually the Euler-Lagrange equation for (TV-W), however it can be difficult to verify the conditions for direct sum between subdifferentials and give a full characterization. Therefore, for some arbitrary $\rho \in \mathcal{M}_+(\Omega)$ and $t > 0$, set

$$\rho_t = \frac{\rho_1 + t(\rho - \rho_1)}{1 + t\alpha}, \text{ where } \alpha = \int_\Omega \mathrm{d}(\rho - \rho_1).$$

Now ρ_t is admissible for the subdifferential inequality and using the positive homogeneity of TV we can write

$$\mathrm{TV}(\rho_1) - \int_\Omega \psi_1 \mathrm{d}\left(\rho_t - \rho_1\right) \leq \frac{\mathrm{TV}(\rho_1) + t\left(\mathrm{TV}(\rho) - \mathrm{TV}(\rho_1)\right)}{1 + t\alpha}.$$

After a few computations we arrive at $\mathrm{TV}(\rho) \geq \mathrm{TV}(\rho_1) + \int_\Omega (C - \psi_1)\mathrm{d}(\rho - \rho_1)$, where $C = \mathrm{TV}(\rho_1) + \int_\Omega \psi_1 \mathrm{d}\rho_1$. Notice that $(\phi + C, \psi - C)$ remains an optimal potential. So we can replace ψ_1 by $\psi_1 - C$, and obtain that for all $\rho \geq 0$ the following holds

$$\mathrm{TV}(\rho) \geq \mathrm{TV}(\rho_1) + \int_\Omega -\psi_1 \mathrm{d}(\rho - \rho_1), \text{ with } \mathrm{TV}(\rho_1) = \int_\Omega -\psi_1 \mathrm{d}\rho_1. \tag{6.14}$$

In particular, this means $-\psi_1 \in \partial\left(\mathrm{TV} + \chi_{\mathcal{M}_+(\Omega)}\right)(\rho_1)$ and ρ_1 is optimal for

$$\inf_{\rho \geq 0} \mathcal{E}(\rho) \overset{\text{def.}}{=} \mathrm{TV}(\rho) + \int_\Omega \psi_1(x)\rho(x)\mathrm{d}x. \tag{6.15}$$

This suggests a penalization with an L^2 term *e.g.*

$$\inf_{u \in L^2(\Omega)} \mathcal{E}_t(u) \overset{\text{def.}}{=} \mathrm{TV}(u) + \int_\Omega \psi_1(x)u(x)\mathrm{d}x + \frac{1}{2t}\int_\Omega |u - \rho_1|^2 \mathrm{d}x \tag{6.16}$$

which is a variation of (ROF) with $g = \rho_1 - t\psi_1$. In order for (6.16) to make sense, we need $\rho_1 \in L^2(\Omega)$, which is true if ρ_0 is L^∞ since then [Carlier and Poon, 2019, Thm. 4.2] implies $\rho_1 \in L^\infty$. Suppose for now that ρ_0 is a bounded function.

Let u_t denote the solution of (6.16), from Prop. 6.5 if u_t solves (6.16), then u_t^+ solves the same problem with the additional constraint that $u \geq 0$, see [Chambolle, 2004, Lemma A.1]. As $\rho_1 \geq 0$ we can compare the energies of u_t^+ and ρ_1 and obtain the following inequalities

$$\mathcal{E}(\rho_1) \leq \mathcal{E}(u_t^+) \text{ and } \mathcal{E}_t(u_t^+) \leq \mathcal{E}_t(\rho_1).$$

Summing both inequalities yields

$$\int_\Omega |u_t^+ - \rho_1|^2 \mathrm{d}x \leq 0, \text{ therefore } u_t^+ = \rho_1 \text{ a.e. on } \Omega. \tag{6.17}$$

In particular, we also have that $u_t \leq \rho_1$. But as u_t solves a (ROF) problem, the optimality conditions from Prop. 6.5 give

$$\beta_t - \psi_1 \in \partial_{L^2} \mathrm{TV}(u_t), \text{ where } \beta_t \overset{\text{def.}}{=} \frac{\rho_1 - u_t}{t} \geq 0. \tag{6.18}$$

Notice from the characterization of $\partial_{L^2} \mathrm{TV}(\cdot)$ that $\partial_{L^2} \mathrm{TV}(u) \subset \partial_{L^2} \mathrm{TV}(u^+)$. Since $u_t^+ = \rho_1$, we have that

$$\beta_t - \psi_1 \in \partial_{L^2} \mathrm{TV}(\rho_1), \tag{6.19}$$

which proves (TVW-EL).

Now we move on to study the family $(\beta_t)_{t>0}$. Since $\rho_1 = u_t^+$, by definition $\beta_t = u_t^-/t$ and using the fact that $\partial_{L^2} \mathrm{TV}(u) \subset \partial_{L^2} \mathrm{TV}(u^-)$ in conjunction with equation (6.18), it holds that

$$\psi_1 - \beta_t \in \partial_{L^2} \mathrm{TV}(\beta_t). \tag{6.20}$$

But then, from Prop. 6.5, β_t solves (ROF) with $g = \psi_1$ and $\lambda = 1$. As this problem has a unique solution, the family $\{\beta_t\}_{t>0} = \{\beta\}$ is a singleton.

Since Ω is convex, and we know that the Kantorovitch potentials are Lipschitz continuous, cf. [Santambrogio, 2015], so β, as a solution of (ROF) with Lipschitz data $g = \psi_1$, is also Lipschitz continuous with the same constant, following [Mercier, 2018, Theo. 3.1].

But from (6.19) and the characterization of the subdifferential of TV, there is a vector field z such that $z \cdot D\rho_1 = |D\rho_1|$ such that

$$\beta - \psi_1 = \operatorname{div} z,$$

and as a consequence $\operatorname{div} z$ is also Lipschitz continuous, with constant at most twice the constant of ψ_1.

In the general case of $\rho_0 \in L^1(\Omega)$, define $\rho_{0,N} \overset{\text{def.}}{=} c_N(\rho_0 \wedge N)$ for $N \in \mathbb{N}$, where c_N is a renormalizing constant. Then $\rho_{0,N} \in L^\infty(\Omega)$ and $\rho_{0,N} \xrightarrow[N\to\infty]{L^1} \rho_0$. Let $\rho_{1,N}$ denote the unique minimizer of (TV-W) with data term $\rho_{0,N}$, we can assume that $\rho_{1,N}$ w-$\star$ converges to some $\tilde{\rho}$. Then for any $\rho \in \mathscr{P}(\Omega)$ we have

$$\mathrm{TV}(\rho_{1,N}) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho_{1,N}) \leq \mathrm{TV}(\rho) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho).$$

Passing to the limit on $N \to \infty$ we have that $\tilde{\rho}$ is a minimizer and from uniqueness it must hold that $\tilde{\rho} = \rho_1$.

Hence, consider the functions $z_N, \psi_{1,N}, \beta_N$ that satisfy (TVW-EL) for $\rho_{1,N}$. Up to a subsequence, we may assume that z_N converges weakly-$\star$ to some $z \in L^\infty(\Omega; \mathbb{R}^d)$. Since $\psi_{1,N}$, β_N and $\operatorname{div} z_N$ are Lipschitz continuous with the same Lipschitz constant for all N, by Arzelà-Ascoli, we can assume that $\psi_{1,N}, \beta_N$ and $\operatorname{div} z_N$ converge uniformly to Lipschitz functions $\psi_1, \beta, \operatorname{div} z = \beta - \psi_1$. In addition, passing to the limit in (6.20), we find that β solves (ROF) for $\lambda = 1$ and $g = \psi_1$.

Since β_N converges uniformly and $\rho_{1,N}$ converges w-$\star$ we have

$$0 = \lim_{N\to\infty} \int_\Omega \beta_N \rho_{1,N} \mathrm{d}x = \int_\Omega \beta \rho_1 \mathrm{d}x,$$

and hence $\beta\rho_1 = 0$ a.e. in Ω since both are nonnegative. In addition, ψ_1 is a Kantorovitch potential associated to ρ_1 from the stability of optimal transport (see [Santambrogio, 2015, Thm. 1.52]). From the optimality of $\rho_{1,N}$ it holds that

$$\mathrm{TV}(\rho_{1,N}) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho_{1,N}) \leq \mathrm{TV}(\rho_1) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho),$$

so that $\lim \mathrm{TV}(\rho_{1,N}) \leq \mathrm{TV}(\rho_1)$. Changing the roles of ρ_1 and $\rho_{1,N}$ we get an equality. So it follows that

$$\int_\Omega (\beta - \psi_1)\rho_1 \mathrm{d}x = \lim_{N\to\infty} \int_\Omega (\beta_N - \psi_{1,N})\rho_{1,N} \mathrm{d}x = \lim_{N\to\infty} \mathrm{TV}(\rho_{1,N}) = \mathrm{TV}(\rho_1),$$

Since TV is 1-homogeneous we conclude that $\beta - \psi_1 \in \partial\, \mathrm{TV}(\rho_1)$. □

We say E is a set of finite perimeter if the indicator function $\mathbb{1}_E$ is a BV function, and we set $\operatorname{Per}(E) = \operatorname{TV}(\mathbb{1}_E)$. As a byproduct of the previous proof we conclude that the level sets $\{\rho_1 > s\}$ are all solutions to the same prescribed curvature problem.

Corollary 6.7. *The following properties of the level sets of ρ_1 hold.*

1. *For $s > 0$ and ψ_1 in* (TVW-EL)
$$\{\rho_1 > s\} \in \underset{E\subset\Omega}{\operatorname{argmin}} \operatorname{Per}(E;\Omega) + \frac{1}{\tau}\int_E \psi_1 \mathrm{d}x$$

2. *$\partial\{\rho_1 > s\} \setminus \partial^*\{\rho_1 > s\}$ is a closed set of Hausdorff dimension at most $d-8$, where ∂^* denotes the reduced boundary of a set, see [Ambrosio et al., 2000]. In addition, $\partial^*\{\rho_1 > s\}$ is locally the graph of a function of class $W^{2,q}$ for all $q < +\infty$.*

Proof. For simplicity take $\tau = 1$. Inside the set $\{\rho_1 > s\}$, for $s > 0$, we have $-\psi_1 = \operatorname{div} z$, so from the definition of the perimeter we have
$$\int_{\{\rho_1>s\}} -\psi_1 \mathrm{d}x = \int_{\{\rho_1>s\}} \operatorname{div} z \mathrm{d}x \leq \operatorname{Per}\left(\{\rho_1 > s\}\right).$$
So using the fact that $\operatorname{TV}(\rho_1) = \int_\Omega -\psi_1 \mathrm{d}x$, the coarea formula and Fubini's Theorem give
$$\int_0^{+\infty} \operatorname{Per}(\mathbb{1}_{\{\rho_1>s\}})\mathrm{d}s = \int_\Omega -\psi_1 \int_0^{\rho_1(x)} \mathrm{d}s\mathrm{d}x = \int_0^{+\infty}\int_{\{\rho_1>s\}} -\psi_1 \mathrm{d}x\mathrm{d}s.$$
Hence, $\operatorname{Per}(\{\rho_1 > s\}) = \int_{\{\rho_1>s\}} -\psi_1 \mathrm{d}x$ for *a.e.* $s > 0$. But as $\beta\psi_1 = 0$ a.e., we have $-\psi_1 = \operatorname{div} z$ in $\{\rho_1 > s\}$, so that $-\psi_1 \in \partial \operatorname{TV}(\mathbb{1}_{\{\rho_1>s\}})$ for *a.e.* $s > 0$; and by a continuity argument, for all $s > 0$. The subdifferential inequality with $\mathbb{1}_E$ gives
$$\{\rho_1 > s\} \in \underset{E\subset\Omega}{\operatorname{argmin}} \operatorname{Per}(E) + \int_E \psi_1(x)\mathrm{d}x. \tag{6.21}$$

Item (2) follows directly from the properties of (ROF), see [Chambolle et al., 2010], since $\rho_1 = u^+$, where u solves a problem (ROF). □

3. Numerical Experiments

In this section we have designed two numerical experiments involving the numerical solution of (TV-W) using a Douglas-Rachford algorithm [Combettes and Pesquet, 2011] with Halpern acceleration [Contreras and Cominetti, 2022], see table 1. For this we need subroutines to compute the prox operators defined, for a given $\lambda > 0$, as
$$\operatorname{prox}_{\lambda\operatorname{TV}}(\bar\rho) \stackrel{\text{def.}}{=} \underset{\rho\in L^2(\Omega)}{\operatorname{argmin}} \operatorname{TV}(\rho) + \frac{1}{2\lambda}\left\|\rho - \bar\rho\right\|^2_{L^2(\Omega)}, \tag{6.22}$$
$$\operatorname{prox}_{\lambda W_2^2}(\bar\rho) \stackrel{\text{def.}}{=} \underset{\rho\in L^2(\Omega)}{\operatorname{argmin}} \frac{1}{2\tau}W_2^2(\rho_0,\rho) + \frac{1}{2\lambda}\left\|\rho - \bar\rho\right\|^2_{L^2(\Omega)}. \tag{6.23}$$

We implemented the prox of TV with the algorithm from [Condat, 2017], modified to account for Dirichlet boundary conditions. From [Chambolle and Pock, 2021, Theo. 2.4] it is consistent with the continuous total variation. The prox of W_2^2 is computed by expanding the L^2 data term as

$$\begin{aligned}\operatorname{prox}_{\lambda W_2^2}(\bar\rho) &= \operatorname*{argmin}_{\rho\in L^2(\Omega)} \frac{1}{2\tau}W_2^2(\rho_0,\rho) + \frac{1}{2\lambda}\int_\Omega \rho^2 \mathrm{d}x + \int_\Omega \rho \underbrace{\left(-\frac{\bar\rho}{\lambda}\right)}_{=V} \mathrm{d}x + \underbrace{\frac{1}{2\lambda}\bar\rho^2 \mathrm{d}x}_{cst} \\ &= \operatorname*{argmin}_{\rho\in L^2(\Omega)} \frac{1}{2\tau}W_2^2(\rho_0,\rho) + \frac{1}{2\lambda}\int_\Omega \rho^2 \mathrm{d}x + \int_\Omega \rho V \mathrm{d}x,\end{aligned}$$

which is one step of the Wasserstein gradient flow of the porous medium equation $\partial_t\rho_t = \lambda^{-1}\Delta(\rho_t^2) + \operatorname{div}(\rho_t\nabla V)$, where the potential is $V = -\bar\rho/\lambda$, see [Santambrogio, 2015]. To compute it we have used the back-n-forth algorithm from [Jacobs et al., 2021].

Algorithm 1 Halpern accelerated Douglas-Rachford algorithm

$\beta_0 \leftarrow 0$
$x_0 \leftarrow$ Initial Image
while $n \geq 0$ **do**
  $y_n \leftarrow \operatorname{prox}_{\lambda\,\mathrm{TV}}(x_n)$
  $\lambda_n \in [\varepsilon, 2-\varepsilon]$
  $z_n \leftarrow x_n + \lambda_n\left(\operatorname{prox}_{\lambda W_2^2}(2y_n - x_n) - y_n\right)$
  $\beta_n \leftarrow \frac{1}{2}\left(1+\beta_{n-1}^2\right)$ ▷ Optimal constants for Halpern acceleration from
  $x_{n+1} \leftarrow (1-\beta_n)x_0 + \beta_n z_n$
end while

3.1. Evolution of balls

Following [Carlier and Poon, 2019], in dimension 1, whenever the initial measure is uniformly distributed over a ball, the solutions remain balls. In $\mathbb{R}^d$, one can prove this remains true. If ρ_0 is uniformly distributed over a ball of radius r_0, then the solution to (TV-W) is uniformly distributed in a ball of radius r_1 solving

$$r_1^2(r_1 - r_0) = r_0^2(d+2)\tau.$$

In this appendix we prove the theoretical characterization of the optimal radius used in the first experiment of Section 3.

Lemma 6.8 ([Carlier and Poon, 2019], Lemma 2.1)**.** *Let* $\rho_0 \in \mathscr{P}_2(\mathbb{R}^d)$ *and take* $\Omega = \mathbb{R}^d$. *If* $\rho_1 \in BV\left(\mathbb{R}^d\right) \cap \mathscr{P}_2(\mathbb{R}^d)$ *and there exists* $p \in \partial\,\mathrm{TV}(\rho_1)$ *satisfying*

$$\frac{\psi}{\tau} + p \geq 0, \text{ with equality } \rho_1\text{-a.e.} \tag{6.24}$$

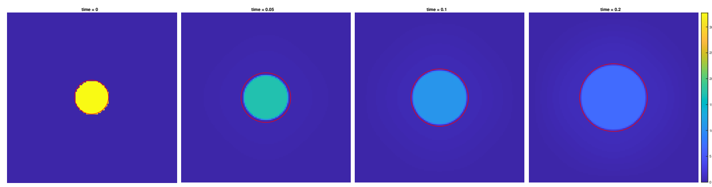

Figure 12: Evolution of circles: from left to right initial condition and solutions for $\tau = 0.05, 0.1, 0.2$. The red circles correspond to the theoretical radius.

then ρ_1 *solves* (TV-W).

Proof. By definition, for any $\rho \in \mathrm{BV}(\Omega) \cap \mathscr{P}_2(\mathbb{R}^d)$, one has

$$\mathrm{TV}(\rho) \geq \int_{\mathbb{R}^d} p \mathrm{d}\rho,$$

So using Kantorovich duality and taking (φ, ψ) the potentials between ρ_0 and ρ_1, for every ρ one obtains

$$\begin{aligned} \frac{1}{2} W_2^2(\rho_0, \rho) &\geq \int \varphi \mathrm{d}\rho_0 + \int \psi \mathrm{d}\rho = \int \varphi \mathrm{d}\rho_0 + \int \psi \mathrm{d}\rho_1 + \int \psi \mathrm{d}(\rho - \rho_1) \\ &= \frac{1}{2} W_2^2(\rho_0, \rho_1) + \int \psi \mathrm{d}(\rho - \rho_1) \\ &\geq \frac{1}{2} W_2^2(\rho_0, \rho_1) + \tau \left(\mathrm{TV}(\rho_1) - \int p \mathrm{d}\rho \right) \end{aligned}$$

Which, by the definition of TV, implies that

$$\mathrm{TV}(\rho_1) + \frac{1}{2\tau} W_2^2(\rho_0, \rho_1) \leq \mathrm{TV}(\rho) + \frac{1}{2\tau} W_2^2(\rho_0, \rho),$$

for all $\rho \in \mathrm{BV}(\Omega) \cap \mathscr{P}_2(\Omega)$, and therefore ρ_1 is a minimizer of (TV-W). $\square$

Given an initial radius r_0, we set

$$\mathcal{E}_{\tau, r_0}(\rho) \stackrel{\text{def.}}{=} \frac{1}{2\tau} W_2^2(\rho_{r_0}, \rho) + \mathrm{TV}(\rho). \tag{6.25}$$

We minimize $\mathcal{E}_{\tau, r_0}(\rho_{r_1})$ as a real valued function of r_1 and check that the minimizer ρ_{r_1} satisfies the sufficient condition 6.8 with a well tailored z.

The first step is to find the optimal transport map (if it exists). From [Santambrogio, 2015, Thm. 1.48] it suffices to find a map $T_\sharp \rho_{r_0} = \rho_{r_1}$ that can be written as the gradient of a convex function. It is easy to check that $T = \frac{r_1}{r_0}\mathrm{id}$ is the gradient of $u(x) \stackrel{\text{def.}}{=} \frac{r_1}{2r_0}|x|^2$ and $\rho_{r_1} = T_\sharp \rho_{r_0}$.

In the sequel, let us compute $\mathcal{E}_{\tau,r_0}(\rho_{r_1})$ for some r_1. The Wasserstein term can be easily computed using the optimal map, namely

$$
\begin{aligned}
\frac{1}{2\tau}W_2^2\left(\rho_{r_0},\rho_{r_1}\right) &= \frac{1}{2\tau}\int_{B(0,r_0)}\left|x-\frac{r_1}{r_0}x\right|^2\frac{\mathrm{d}x}{\omega_d r_0^d} = \frac{1}{2\tau}\left(1-\frac{r_1}{r_0}\right)^2\frac{1}{\omega_d r_0^d}\int_{B(0,r_0)}|x|^2\,\mathrm{d}x \\
&= \frac{1}{2\tau}\left(1-\frac{r_1}{r_0}\right)^2\frac{1}{\omega_d r_0^d}\int_0^{r_0} r^2\mathscr{H}^{d-1}(\partial B(0,r))\mathrm{d}r \\
&= \frac{1}{2\tau}\left(1-\frac{r_1}{r_0}\right)^2\frac{1}{\omega_d r_0^d}\int_0^{r_0} 2\pi\omega_{d-1}r^{d+1}\mathrm{d}r \\
&= \frac{1}{2(d+2)\tau}\frac{2\pi\omega_{d-1}}{\omega_d}(r_1-r_0)^2.
\end{aligned}
$$

To compute the total variation term, we will use the coarea formula.

$$
\begin{aligned}
\mathrm{TV}(\rho_{r_1}) &= \int_{\mathbb{R}}\mathrm{Per}\left(\{\rho_{r_1}>s\}\right)\mathrm{d}s = \int_0^{1/\omega_d r_1^d}\mathrm{Per}\left(\{\rho_{r_1}>s\}\right)\mathrm{d}s \\
&= \frac{1}{\omega_d r_1^d}\mathrm{Per}\left(B(0,r_1)\right) = \frac{1}{\omega_d r_1^d}\mathscr{H}^{d-1}\left(\partial B(0,r_1)\right) \\
&= \frac{2\pi\omega_{d-1}}{\omega_d}\frac{1}{r_1}.
\end{aligned}
$$

Hence, setting $K_d \overset{\text{def.}}{=} \frac{2\pi\omega_{d-1}}{\omega_d}$ we obtain

$$
\mathcal{E}_{\tau,r_0}(r_1) = \frac{K_d}{2(d+2)\tau}(r_1-r_0)^2 + \frac{K_d}{r_1},
$$

which is minimized by the positive root of

$$
r_1^2(r_1-r_0) = r_0^2(d+2)\tau. \tag{6.26}
$$

Now, supposing that some function z satisfying the conditions of Lemma 6.8 exists, let us try to find it explicitly. Starting with the Kantorovitch potential ψ, we know that $T = \frac{r_0}{r_1}\mathrm{id} = \mathrm{id} - \nabla\psi$, which means that ψ is of the form

$$
\psi(x) = \frac{r_1-r_0}{2r_1}|x|^2 + C.
$$

Hence, we look for z such that

$$
\operatorname{div} z(x) = \frac{r_0-r_1}{2\tau r_1}|x|^2 - \frac{C}{\tau},\ \text{for all } x\in B(0,r_1),
$$

and the constant C can be computed explicitly with the relation

$$
\mathrm{TV}(\rho_{r_1}) = \frac{K_d}{r_1} = \int_{B(0,r_1)}\operatorname{div} z\,\mathrm{d}\rho_{r_1}.
$$

In particular, we can take z of the form

$$z(x) = -\frac{r_1 - r_0}{2\tau r_1} x^3 + \frac{2}{r_1} x + z(0), \text{ for } x \in B(0, r_1),$$

and x^3 stands for the vector $(x_i^3)_{i=1}^d$. The condition $\|z\|_{L^\infty(B(0,r_1))} \leq 1$ also holds for a suitable choice of $z(0)$ and to conclude it suffices to extend z outside $B(0, r_1)$ in order to keep this bound and have compact support. As such extension always exists, we conclude.

3.2. Reconstruction of dithered images

In this experiment we use model (TV-W) to reconstruct dithered images. In $\mathscr{P}(\mathbb{R}^2)$ the dithered image is a sum of Dirac masses, so the model (TV-W) outputs a new image which is close in the Wasserstein topology, but with small total variation. In Figure 13 below, we compared the result with the reconstruction given by (ROF), both with a parameter $\tau = 0.2$. Although the classical (ROF) model was able to create complex textures, these remain granulated, whereas the (TV-W) model is able to generate both smooth and complex textures.

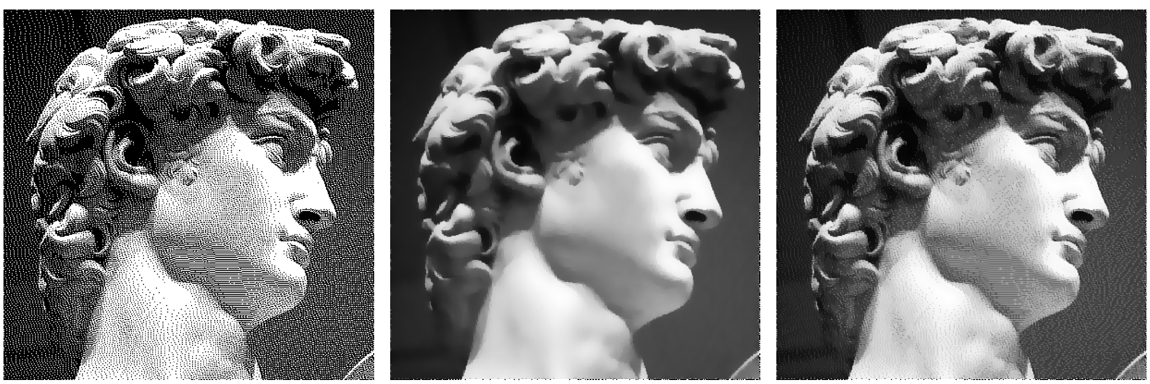

Figure 13: Dithering reconstruction problem. From left to right: Dithered image, TV-Wasserstein and ROF results.

4. Conclusion

In this work we revisited the TV-Wasserstein problem. We showed how it can be related to the classical (ROF) problem and how to exploit this to derive the Euler-Lagrange equations, obtaining further regularity. We proposed a Douglas-Rachford algorithm to solve it and presented two numerical experiments: the first one being coherent with theoretical predictions and the second being an application to the reconstruction of dithered images.

Chapter 7

From Nash to Cournot-Nash via Γ convergence

This work is a collaboration with Guilherme Mazanti and Laurent Pfeiffer.

Contents

1. Introduction

One of the central questions in game theory is how the mean collective choice affects a game as the number of players grows and how the notion of equilibrium for such games is affected. In the category of differential games, *i.e.* when players are constrained by controlled dynamics and hence solve an optimal control problem each, the notion of equilibrium in a setting with a continuum of players has been established by Larsy and Lions as *Mean Field Games* in their seminal papers [Lasry and Lions, 2006a, Lasry and Lions, 2006b, Lasry and Lions, 2007]. In their formulation, an equilibrium in this continuum setting becomes the solution of a pair of coupled PDEs, one describing the evolution of the probability distribution of players and another describing the optimality conditions of the underlying optimal control problem solved by each player. The term Mean Field Game now refers to a wider class of equilibrium problems in economics, statistical physics, biology and social sciences, and for instance we can mention the subclass of *Lagrangian Mean Field Games*, see [Santambrogio and Shim, 2021, Bonnans et al., 2021]. In this case, instead of being constrained by a certain differential equation, players seek to minimize a certain energy by choosing a continuous curve and an optimal dynamics is then selected from the optimality conditions related to this energy.

Although the literature in Mean Field Games is vastly expanding since the works of Larsy and Lions, the notion of games with a continuum of players is much older and has been studied in the economics literature since the 60's by Aumann [Aumann, 1964, Aumann, 1966]. Although Aumann formulated his notion of equilibrium with preference relations, as was noted in [Mas-Colell, 1984], an equivalent way of defining equilibria in the continuous setting is to consider a cost function, indexed by the player and depending on the collective distribution of plays. The game then consists of finding an equilibrium between each player trying to minimize the cost despite the effects of the collective distribution. In [Schmeidler, 1973], Schmeidler was interested in a such a model with a continuum of players, but more specifically he wanted to obtain existence of equilibria in pure strategies. He described a profile of strategies as a measurable function from the space of players to the space of admissible strategies. Later on, this notion of equilibrium was relaxed, for instance in [Hart et al., 1974, Mas-Colell, 1984], defining equilibria as probability measures over the space of admissible strategies, introducing the notion that is now known as *Cournot-Nash* equilibrium, see Definition 7.1 below.

This relaxation also played an important part in the development of the optimal transportation problem, described as follows: given two Polish spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$, a pair of Borel probability measures $\mu \in \mathscr{P}(\mathcal{X})$, $\nu \in \mathscr{P}(\mathcal{Y})$ and a transportation cost $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, one seeks to minimize the following

$$\mathcal{W}_c(\mu, \nu) \stackrel{\text{def.}}{=} \inf_{T_\sharp \mu = \nu} \int_{\mathcal{X}} c(x, T(x)) \mathrm{d}\mu = \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\gamma. \tag{7.1}$$

The infimum on the left is known as Monge's formulation [Monge, 1781] and is taken over all measurable maps T that transport μ to ν, in the sense that for all measurable sets $B \subset \mathcal{Y}$ it holds that $T_\sharp \mu(B) \stackrel{\text{def.}}{=} \mu(T^{-1}(B)) = \nu(B)$, where $T_\sharp \mu$ is called the push-forward measure.

The minimum on the right side is called Kantorovitch formulation [Kantorovich, 1942], which is taken among all couplings of μ and ν,

$$\Pi(\mu,\nu) \stackrel{\text{def.}}{=} \{\gamma \in \mathscr{P}(\mathcal{X}\times\mathcal{Y}) : (\pi_{\mathcal{X}})_\sharp\gamma = \mu,\ (\pi_{\mathcal{Y}})_\sharp\gamma = \nu\}, \tag{7.2}$$

that is the probability measures in the product space whose marginals are μ and ν.

As in the game theory community, the original interest of Monge was to find transportation strategies that are given by maps. However, it is fairly easy to obtain a pair of measures not admitting any map transporting one onto the other, for instance if μ is a Dirac mass and ν is diffuse. In that case one cannot avoid splitting the mass to perform the transportation, in the jargon of game theory the players would be forced to mix their plays in order to attain a fixed distribution of strategies. For this reason Kantorovitch proposed his formulation in order for the problem to be well posed for any pair of Borel measures. In this sense, we lose the initial intuition of Monge for the sake of existence of solutions. Afterwards, Monge's intuition was recovered from Kantorovitch's general formulation by Brenier's seminal papers [Brenier, 1987, Brenier, 1991] in the euclidean case with quadratic cost, and later by McCann [McCann, 1995, McCann, 2001] in the case of a Riemannian manifold, by finding conditions guaranteeing that optimal transportation plans are concentrated in the graph of a measurable map.

One can interpret $\mathcal{X}$ as the space of types of players with distribution given by $\mu \in \mathscr{P}(\mathcal{X})$, $\mathcal{Y}$ to be the space admissible strategies for said players, with distribution given by $\nu \in \mathscr{P}(\mathcal{Y})$, and a coupling $\gamma \in \Pi(\mu,\nu)$ as the joint distribution of players and strategies. In other words, given $A \times B \subset \mathcal{X}\times\mathcal{Y}$, the quantity $\gamma(A\times B)$ represents the probability that a player with type in A chooses a strategy in B. Given a function $\Phi : \mathcal{X}\times\mathcal{Y}\times\mathscr{P}(\mathcal{Y}) \to \mathbb{R}\cup\{+\infty\}$, let $\Phi(x,y,\nu)$ denote the cost of a player of type x to choose the strategy y, in a mean field of strategies represented by the distribution ν, we obtain the notion of *Cournot-Nash* equilibrium.

Definition 7.1. *A probability measure $\gamma \in \Pi(\mu,\nu)$ is a Cournot-Nash equilibrium for the game referent to the cost $\Phi : \mathcal{X}\times\mathcal{Y}\times\mathscr{P}(\mathcal{Y})$ if it satisfies the equilibrium condition*

$$\gamma\left(\left\{(x,y)\in\mathcal{X}\times\mathcal{Y} : y \in \operatorname*{argmin}_{y'\in\mathcal{Y}} \Phi(x,y',\nu)\right\}\right) = 1, \tag{7.3}$$

it is called an equilibrium of finite social cost if

$$\int_{\mathcal{X}\times\mathcal{Y}} \Phi(x,y,\nu)\mathrm{d}\gamma < +\infty. \tag{7.4}$$

An equilibrium γ is called pure if it can be written as $\gamma = (\mathrm{id},T)_\sharp\mu$, where $T : \mathcal{X}\to\mathcal{Y}$ is a measurable map.

Results guaranteeing the existence of equilibria have been established with fixed point methods in the above-mentioned works. This approach relies strongly on the continuity of the cost function. In [Blanchet and Carlier, 2016], whenever Φ can be written as

$$\Phi(x,y,\nu) = c(x,y) + \frac{\delta\mathcal{E}}{\delta\nu}[\nu](y), \tag{7.5}$$

the sum of an individual continuous cost $c(x,y)$ and the first variation of a functional $\mathcal{E}$, Blanchet and Carlier showed that if $\gamma \in \Pi(\mu,\nu)$ is an optimal transportation plan for the cost c and

$$\nu \in \underset{\nu' \in \mathscr{P}(\mathcal{Y})}{\operatorname{argmin}} \mathcal{W}_c(\mu,\nu') + \mathcal{E}(\nu), \tag{7.6}$$

then γ is a Cournot-Nash equilibria in the sense of Definition 7.1. The advantage of their approach is twofold: firstly, as their proof of existence is of variational nature it provides a natural approach to compute equilibria numerically, as done in [Blanchet and Carlier, 2014b, Blanchet and Carlier, 2016], see also [Blanchet et al., 2018] for an approach using entropic regularization of the OT term. The characterization via optimal transport also gives information about the existence of pure Cournot-Nash equilibria, the original motivation of Schmeidler, that is at first glace abandoned when we define equilibria as couplings instead of maps. Since equilibria are described as optimal solutions to an optimal transport problem with first marginal given by μ, one can then use the well established conditions from OT to conclude that equilibria are pure, under suitable assumptions on μ and c, see for instance [Gangbo and McCann, 1996, Carlier, 2003] or the recent monograph [Santambrogio, 2015].

It is worth noting that definition 7.1 deviates from the literature, for instance from the one from [Blanchet and Carlier, 2016], as we also require the finite social cost condition. It is not restrictive to the analysis from [Blanchet and Carlier, 2016],where the cost c is assumed continuous and the underlying spaces compact, so that (7.4) holds trivially. In economic modeling we want to take into account that not all strategies are accessible to all types of individuals. Richer individuals have access to better education, health care and financial products and services. On the other hand, it is not reasonable that they benefit from governmental aids. For such reasons it is pertinent to allow $c(x,y)$ to be $+\infty$ to model the fact that not all strategies are attainable for all players, so that (7.4) is no longer trivially satisfied.

Our model and convergence of Nash to Cournot-Nash equilibria

In the present Chapter we will study a model similar to [Blanchet and Carlier, 2014b, Blanchet and Carlier, 2014a, Blanchet and Carlier, 2016]. Assume that $\mathcal{X}$ and $\mathcal{Y}$ are Polish (separable, complete and metric) spaces denoting the spaces of types of players and of admissible strategies, respectively. Our first result is a full characterization of Cournot-Nash equilibria of games with costs having the potential structure from (7.5), where we assume only lower semi-continuity of c and $\mathcal{E}$. Instead of working with the energy $\nu \mapsto \mathcal{W}_c(\mu,\nu) + \mathcal{E}(\nu)$, used in [Blanchet and Carlier, 2016], we use a lifted energy over the space of transportation plans with fixed marginal μ, which is defined as

$$\mathcal{J}(\gamma) \stackrel{\text{def.}}{=} \begin{cases} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\gamma + \mathcal{E}(\nu), & \text{if } \gamma \in \Pi(\mu,\nu), \\ +\infty & \text{if } \gamma \notin \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y}), \end{cases} \tag{7.7}$$

where

$$\mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y}) \stackrel{\text{def.}}{=} \left\{\gamma \in \mathscr{P}(\mathcal{X}\times\mathcal{Y}) : \quad (\pi_{\mathcal{X}})_\sharp \gamma = \mu\right\}. \tag{7.8}$$

We show in Theorem 7.4 that γ is a Cournot-Nash equilibrium if, and only if, it is an critial point of the energy $\mathcal{J}$, see Definition 7.2.

As the cost c can now be any l.s.c. function, we cannot resort to the same techniques as [Blanchet and Carlier, 2016], as the latter relies strongly on the continuity of the cost to have uniqueness of solutions of the dual problem. On the other hand, since the cost function c no longer needs to be continuous, one can take into account social or feasibility constraints, since a player of type x is now obliged to choose strategies in the set $\mathcal{Y}_x \stackrel{\text{def.}}{=} \{y \in \mathcal{Y} : c(x,y) < +\infty\}$. Such constraints can be used to model stratification phenomena in a society where only certain individuals have access to certain advantages or possibilities, for instance from their financial resources or from governmental incentives to vulnerable groups.

In the sequel we focus our attention to the case where $\mathcal{E}$ can be decomposed into a mean individual and interaction energies as follows.

$$\mathcal{E}(\nu) = \mathcal{L}(\nu) + \mathcal{H}(\nu,\nu), \text{ where } \mathcal{L}(\nu) = \int_{\mathcal{Y}} L\mathrm{d}\nu \text{ and } \mathcal{H}(\nu,\nu) = \int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu\otimes\nu, \tag{7.9}$$

where the first variation of the energy, see Definition (7.15), can be explicitly computed as

$$\frac{\delta\mathcal{E}}{\delta\nu}(\nu) = L + 2\int_{\mathcal{Y}} H(\cdot,y)\mathrm{d}\nu(y), \tag{7.10}$$

whenever H is symmetric. We make the following assumptions on these functionals

(H3) $\mu \in \mathscr{P}(\mathcal{X})$ is atomless.

(H4) $L\colon \mathcal{Y} \to \bar{\mathbb{R}}_+$ is lower semi-continuous and $H\colon \mathcal{Y}\times\mathcal{Y} \to \bar{\mathbb{R}}_+$ is Borel measurable.

(H5) H is symmetric, i.e., $H(y,\widetilde{y}) = H(\widetilde{y},y)$ for every $(y,\widetilde{y}) \in \mathcal{Y}\times\mathcal{Y}$.

(H6) The function $\mathcal{Y}\times\mathcal{Y} \ni (y,\widetilde{y}) \mapsto L(y) + L(\widetilde{y}) + H(y,\widetilde{y}) \in \bar{\mathbb{R}}_+$ is lower semi-continuous.

(H7) L has compact sub-level sets, *i.e.* for every $\kappa > 0$, the set $\{L \leq \kappa\}$ is compact.

Under these conditions the characterization of Cournot-Nash equilibria holds, and we focus our attention to a stability result for the minimization of the energy $\mathcal{J}$ with respect to the marginal μ. In particular, under additional assumption (H8), in Theorem 7.9 we show a Lipschitz dependence of the value function for the 1-Wasserstein distance.

Once we understand the structure of the game with a continuum of players, we wish to answer the following question:

> *Given a sample of players following a continuous distribution, when does a sequence of Nash equilibria for the associated finite game will converge to a Cournot-Nash equilibrium?*

This is a broad question in game theory and our main contribution in this work is how we exploit the variational criterion to obtain Cournot-Nash equilibria in order to show that Nash equilibria of a suitable sequence of N-player games converge to Cournot-Nash equilibria. Understanding this limiting procedure is one of the major motivations for the study of Mean Field Games, or more generally games with a continuum of players. In many economic and social scenarios, the number of agents acting in the game becomes rapidly intractable, hence it is of great theoretical and practical importance to be able to rigorously describe the model with infinitely many players as the limit of a sequence of N-players games. This question has been addressed since the inception of the Mean Field Games theory in the lectures of Pierre-Louis Lions at Collège de France, see the notes of Cardaliaguet [Cardaliaguet, 2010], and also for the convergence to Cournot-Nash equilibria in [Blanchet and Carlier, 2014a]. Both references treat a case where the N players of the game are fixed from the start and seek each to optimize a continuous cost.

As we do not assume any continuity of Φ, we cannot resort to fixed point methods, but we can exploit the structure given by the pairwise interaction between players to define a sequence of N-player games whose equilibria can also be obtained through the minimization of an energy $\mathcal{J}_N$ and show that this sequence of functions converge in the sense of Γ-convergence to $\mathcal{J}$. The latter is a variational notion of convergence proposed by De Giorgi, see [Dal Maso, 1993], with the property that if γ_N is a sequence of minimizers of $\mathcal{J}_N$, which converge to γ, then γ is a minimizer of $\mathcal{J}$.

The situation we wish to model is the following: given a sample of the types of players obtained through an *i.i.d.* sample $(X_i)_{i\in\mathbb{N}}$ with common distribution $\mu \in \mathscr{P}(\mathcal{X})$, we let the first N elements represent the type of the agents in our N-player game. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ induced by the sample $(X_i)_{i\in\mathbb{N}}$, so that Ω represents all the possible realizations of this sampling, $\mathcal{F}$ is the σ-algebra generated by the random variables X_i and, $\mathbb{P} = \mu^{\otimes\mathbb{N}}$.

In Section 3, we then consider two types of information structures, in the *open loop information structure* players only know the distribution of types μ and hence choose an execution profile that assigns a strategy in $\mathcal{Y}$, or more generally $\mathscr{P}(\mathcal{Y})$ in mixed plays, for each realization of their type random variable X_i. This gives rise to a stochastic game with a potential function on the space of *random probability measures*, which Γ-converges to the functional $\mathcal{J}$ over the space of *non-random* transportation plans from (7.7). Making the further assumption that H is a continuous and bounded function, we prove that any cluster point of a sequence of Nash equilibria in the N-players game is a Cournot-Nash equilibria by mens of the characterization of such equilibria via the stationarity of the potential function.

On the other hand, in a *closed loop information structure* for a given realization of the sample $\omega = (X_i = x_i)_{i\in\mathbb{N}}$, we define a sequence of N-player games that also admit a potential function $\mathcal{J}_{\omega,N}$, which Γ-converges to the same functional $\mathcal{J}$ with full probability.

Under the same assumptions of the open loop case, we also prove that the cluster points of any sequence of Nash equilibria is a Cournot-Nash equilibria for the limiting game.

Examples

In this paragraph we discuss multiple examples that are covered by our model and their relevance in the literature.

Potential Cournot-Nash equilibria [Blanchet and Carlier, 2016]

The first clear example is the model proposed in [Blanchet and Carlier, 2016]. As discussed above, they proposed a variational principle to find equilibria as in Definition 7.1 and gave plenty of examples of economic applications for this model as the holiday choice and technology choice models. The distinctions from ours is that, for them c had to be a continuous cost in order to give the variational characterization of equilibria using the OT problem as in equation (7.6). We do not need this assumption since we propose the lifting to the space of transportation plans $\gamma \in \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$. This lift is purely technical and the characterization via the value of the associated OT problem still holds in our case and is useful for numerical purposes, since one can use the dual formulation of the OT problem as a dimensionality reduction technique. On the other hand, taking c to be lower semi-continuous allows us to make a link with the next class of examples.

Abstract Lagrangian Mean Field Games [Santambrogio and Shim, 2021]

Consider a crowd motion, where the starting point of each agent is distributed by a probability measure μ and the final goal of each agent is to reach a target set while minimizing a cost depending on their own trajectory and on the distribution of trajectories of all agents Q. One can think of the target set as the exit of a metro, for instance. In [Santambrogio and Shim, 2021], Santambrogio and Shim propose a model where each agent chooses their trajectory among all possible continuous curves respecting their given initial condition. In this case, $\mathcal{X} = \Omega$ is a compact subset of $\mathbb{R}^d$ and $\mathcal{Y} = C^0([0,T];\Omega)$. Each agent then tries to find a curve σ, such that $\sigma(0) = x_0$, the given initial condition, while minimizing an energy of the form

$$F(\sigma, Q) \overset{\text{def.}}{=} \int_0^T \left(|\sigma'(t)|^2 + \int_{\mathcal{Y}} \eta(\sigma(t) - \tilde{\sigma}(t)) |\sigma'(t) - \tilde{\sigma}'(t)|^2 \mathrm{d}Q(\tilde{\sigma}) \right) \mathrm{d}t. + \Psi(\sigma(T))$$

Here Ψ is an end point cost and η is an interaction kernel of Cucker-Smale type in order to observe a phenomenon of consensus of the velocities as in the seminal paper [Cucker and Smale, 2007]. The measure Q corresponds to the distribution of trajectories of all agents so that the integral term becomes a mean interaction cost and the initial condition is then imposed by the constraint $(e_0)_\sharp Q = \mu$. They defined equilibria as measures $Q \in \mathscr{P}(\mathcal{Y})$ such that $(e_0)_\sharp Q = \mu$ and

$$\int_{\mathcal{Y}} F(\sigma, Q)\mathrm{d}Q < \infty, \text{ and } F(\sigma, Q) = \inf_{\tilde{\sigma}(0)=\sigma(0)} F(\tilde{\sigma}, Q), \text{ for all } \sigma \in \operatorname{supp} Q. \tag{7.11}$$

In [Sadeghi Arjmand, 2022], this model was generalized into an abstract model, where F is given by

$$F(\sigma, Q) \stackrel{\text{def.}}{=} L(\sigma) + \int_{\mathcal{Y}} H(\sigma, \tilde{\sigma})\mathrm{d}Q(\tilde{\sigma}),$$

where $L\colon \mathcal{Y} \to \mathbb{R}$ is an individual cost, $H(\cdot,\cdot)\colon \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is a symmetric cost of pairwise interaction, where $\mathcal{Y}$ is an abstract space of admissible strategies, being a Polish space. The players are now labelled by some $x \in \mathcal{X}$, another Polish space of types of players following a distribution $\mu \in \mathscr{P}(\mathcal{X})$. The initial condition map is now replaced by a continuous map $e : \mathcal{Y} \to \mathcal{X}$, and we consider measures $Q \in \mathscr{P}(\mathcal{Y})$ such that $\pi_\sharp Q = \mu$. Their notion of equilibrium is the same as in (7.11), but they show that equilibria are critical points of the following functional

$$Q \mapsto \int_{\mathcal{Y}} L\mathrm{d}Q + \int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}Q \otimes Q.$$

This suggests a link with the previous model of Cournot-Nash equilibria and indeed, for $c(x, y) = 0$ if $x = e(y)$ and $+\infty$ otherwise, we can rewrite the constraints as

$$\pi_\sharp \nu = \mu \iff \mathcal{W}_c(\mu, \nu) < \infty, \text{ since } \mathcal{W}_c(\mu, \nu) = \begin{cases} 0, & \text{if } e_\sharp \nu = \mu, \\ +\infty, & \text{otherwise.} \end{cases}$$

Conversely, if we propose the lifted energy to the space of transportation plans (7.7), the variational criterion for Cournot-Nash equilibria from Blanchet and Carlier is of the same form as the one for Lagrangian MFGs.

Wasserstein gradient flows (JKO schemes)

For the final example, let $\mathcal{X} = \mathcal{Y} = \Omega$ be a compact subset of $\mathbb{R}^d$. We wish to discuss the case of Wasserstein gradient flows, also known as JKO schemes in reference to the seminal paper of Jordan, Kinderlehrer, and Otto [Jordan et al., 1998], where the authors proposed a variational formulation of the Fokker–Planck equation. Their ideas were later generalized to other evolution equations, for instance in [Ambrosio et al., 2008], see also [Santambrogio, 2015, Chap. 8]. The scheme consists in solving the following variational problem iteratively

$$\rho_{k+1} \in \operatorname*{argmin}_{\rho\in\mathscr{P}(\Omega)} \frac{1}{2\tau} W_2^2(\rho_k, \rho) + \mathcal{F}(\rho) \text{ with } \rho_0 \text{ given}, \tag{7.12}$$

where W_2^2 corresponds to the value of the OT problem with $c(x, y) = |x - y|^2$. By solving this sequence of variational problems, one obtains a sequence $(\rho_k)_{k\in\mathbb{N}}$ and define an interpolation depending on the parameter τ as

$$\rho_\tau(t) \stackrel{\text{def.}}{=} \rho_k \text{ if } t \in [k\tau, (k + 1)\tau).$$

For a variety of choices of $\mathcal{F}$, it can be shown that ρ_τ converges as $\tau \to 0$ to a solution of the evolution equation

$$\partial_t \rho + \operatorname{div}\left(\rho \nabla \frac{\delta \mathcal{F}}{\delta \rho}(\rho)\right) = 0, \quad \rho(0) = \rho_0, \tag{7.13}$$

with no-flux boundary conditions.

The case, see [Carrillo et al., 2003, Carrillo et al., 2006],

$$\mathcal{F}(\rho) = \int_\Omega V(x)\mathrm{d}\rho(x) + \int_{\Omega\times\Omega} W(x-y)\mathrm{d}\rho\otimes\rho(x,y),$$

corresponds to an advection plus aggregation phenomenon, that is covered by our Γ-convergence results. Hence, it would be interesting to investigate if, under suitable conditions, our convergence result could possibly be used to show that given an i.i.d. sample of initial conditions $(X_i)_i$ with law ρ_0, letting $(x_i(\cdot))_i$ denote a family of integral curves, to an appropriate vector field, with initial condition $x_i(0) = X_i$, then the limits of the measures $\frac{1}{N}\sum_{i=1}^{N} \delta_{x_i(t)}$ are almost surely solutions to the corresponding evolution equation (7.13).

2. Potential structure for Cournot-Nash equilibria

In this section, our objective is twofold, first we extend the results of Blanchet and Carlier about the potential structure for Cournot-Nash equilibria, allowing for individual costs c that are l.s.c. instead of continuous. In the sequel we show a stability result of the value function w.r.t. the fixed marginal μ.

2.1. Potential structure for Cournot-Nash equilibria

The goal of this section is to characterize equilibria in the sense of Definition 7.1 as critical points of an energy functional. For now, we assume that the optimization problem a player of type $x \in \mathcal{X}$ tries to solve among a mean field of plays $\nu \in \mathscr{P}(\mathcal{Y})$ is given by

$$\min_{y\in\mathcal{Y}} \Phi(x,y,\nu) \stackrel{\text{def.}}{=} c(x,y) + \frac{\delta\mathcal{E}}{\delta\nu}(\nu)(y), \tag{7.14}$$

where c is l.s.c. and the second term can be written as the first variation of an energy $\mathcal{E} : \mathscr{P}(\nu) \to \mathbb{R}$, which is defined below.

Definition 7.2. *We say that a functional $\mathcal{F}$ defined over the probability measures $\mathscr{P}(\mathcal{X})$ over a Polish space $\mathcal{X}$ admits a first variation at $\mu_0 \in \mathscr{P}(\mathcal{X})$ if there exists a Borel measurable* *function $f : \mathcal{X} \to \mathbb{R}$ such that f is $\mu - \mu_0$ integrable for all μ in the domain of $\mathcal{F}$, that is the integral of f against either μ or μ_0 is finite, and such that*

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0^+} \mathcal{F}(\mu_0 + \varepsilon(\mu - \mu_0)) = \langle f, \mu - \mu_0 \rangle \stackrel{\text{def.}}{=} \int_{\mathcal{X}} f \mathrm{d}(\mu - \mu_0), \tag{7.15}$$

and we write $f = \dfrac{\delta \mathcal{F}}{\delta \mu}(\mu_0)$. In addition, we say that μ_0 is a critical point of $\mathcal{F}$ if

$$\left\langle \frac{\delta \mathcal{F}}{\delta \mu}(\mu_0), \mu - \mu_0 \right\rangle \geq 0 \textit{ for all } \mu \in \operatorname{dom} \mathcal{F}.$$

It is clear that the first variation as in Definition 7.2 cannot be unique, since summing a constant to a function satisfying (7.15) will still satisfy the same relation, as the integration is taken against $\mu - \mu_0$, which integrates to 0. It is, however, unique up to a constant.

For the rest of this paragraph, we let $\mathcal{E}$ be an l.s.c. functional over $\mathscr{P}(\mathcal{Y})$, and we consider the energy

$$\mathcal{J}(\gamma) \stackrel{\text{def.}}{=} \begin{cases} \displaystyle\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\gamma + \mathcal{E}(\nu), & \text{if } \gamma \in \Pi(\mu,\nu), \\ +\infty & \text{if } \gamma \notin \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y}), \end{cases} \tag{7.16}$$

with a general functional $\mathcal{E}$, so that

$$\Phi(x,y,\nu) = \frac{\delta \mathcal{J}}{\delta \gamma}(\gamma) \text{ for } \nu = (\pi_{\mathcal{Y}})_\sharp \gamma.$$

Our goal is to show that critical points of this energy are Cournot-Nash equilibria, notice however that satisfying the equilibrium condition (7.3) is independent of having a finite social cost (7.4), we can have bad equilibria that represents a society with infinite poverty, for instance if a non-negligible part of the population is infinitely poor. For this we make the following definition.

Definition 7.3. *A measure $\varrho \in \mathscr{P}(\mathcal{Y})$ is a distribution of finite social cost for the distribution μ if there is a function $\kappa \in L^1(\mu)$ such that for μ-a.e. $x \in \mathcal{X}$ there is $y_x \in \mathcal{Y}$ satisfying*

$$\Phi(x, y_x, \varrho) \leq \kappa(x).$$

The main result of this section is the following.

Theorem 7.4. *Assume that $0 \leq \mathcal{E}$ admits a first variation given by an l.s.c. function with compact sub-level sets over $\mathcal{Y}$ and let Φ be as (7.14). It follows that*

(i) $\gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$ is a Cournot-Nash equilibrium in the sense of Definition 7.1, if and only if it is a critical point of $\mathcal{J}$ defined in (7.16). If in addition, $\nu = (\pi_{\mathcal{Y}})_\sharp\gamma$ is a distribution of finite social cost, then γ is a Cournot-Nash equilibrium of finite social cost.

(ii) if $\mathcal{J}$ admits a minimizer, then

$$\min_{\gamma\in\mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})} \mathcal{J}(\gamma) = \min_{\nu\in\mathscr{P}(\mathcal{Y})} \mathcal{W}_c(\mu,\nu)+\mathcal{E}(\nu),$$

and it is a Cournot-Nash equilibrium.

Proof. The proof is inspired by the arguments in [Sadeghi Arjmand, 2022, Thm. 4.5.1] for the case of an abstract Lagrangian Mean Field Game and [Liu and Pfeiffer, 2023, Appendix A].

First suppose that $\gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$ is a critical point, and define the function

$$\phi(x) \overset{\text{def.}}{=} \inf_{\mathcal{Y}} \Phi(x,\cdot,\nu).$$

It follows that ϕ is Borel measurable since it is lower semi-continuous as we prove next. Take $x_k \xrightarrow[k\to\infty]{} x$ such that $\liminf \phi(x_k)$ is finite, otherwise there is nothing to prove, and assume up to the extraction of a subsequence that the $\liminf$ is a limit. Consider $y_k \in \operatorname{argmin} \Phi(x_k,\cdot,\nu)$ so that $\Phi(x_k,\cdot,\nu)\le C$ is uniformly bounded. Therefore, as $c\ge 0$ it holds that

$$(y_k)_{k\in\mathbb{N}} \subset \left\{\frac{\delta\mathcal{E}}{\delta\nu}(y)\le C\right\},$$

which is a compact set. Up to another extraction, we may assume that $y_k \to y$, so that the lower semi-continuity of Φ gives

$$\phi(x)\le\Phi(x,y,\nu)\le\liminf_{k\to\infty}\Phi(x_k,y_k,\nu)=\liminf_{k\to\infty}\phi(x_k).$$

To prove item (i), if suffices to show that the set

$$A=\{(x,y)\in\mathcal{X}\times\mathcal{Y} : \phi(x)<\Phi(x,y,\nu)\}$$

is γ-negligible. Suppose this is not the case, and our goal is to construct a Borel measurable selection of the argmin operator, that is a Borel function $T:\mathcal{X}\to\mathcal{Y}$ such that

$$T(x)\in\operatorname*{argmin}_{\mathcal{Y}}\Phi(x,\cdot,\nu) \text{ for all } x\in X.$$

From [Brown and Purves, 1973, Thm. 1] it holds that if $E\subset\mathcal{X}\times\mathcal{Y}$ is a Borel set with the property that $E_x \overset{\text{def.}}{=} \{y\in\mathcal{Y} : (x,y)\in E\}$ is σ-compact for all $x\in\pi_{\mathcal{X}}(E)$, then there is a Borel measurable selection $T:\pi_{\mathcal{X}}(E)\to\pi_{\mathcal{Y}}(E)$. And from [Brown and Purves, 1973, Cor. 1], the measurable selection of the argmin operator can be obtained since A is a Borel set, as ϕ and Φ are Borel measurable, and the sub-level sets of the first variation of $\mathcal{E}$ are compact, so that

$$\mathcal{Y}=\bigcup_{n\in\mathbb{N}}\left\{y : \frac{\delta\mathcal{E}}{\delta\nu}(y)\le n\right\}, \text{ is } \sigma\text{-compact}.$$

In the sequel, we use it to define a transportation plan given by

$$\bar\gamma \overset{\text{def.}}{=} \gamma \llcorner (\mathcal{X}\times\mathcal{Y}\setminus A) + (\pi_{\mathcal{X}}, T\circ\pi_{\mathcal{X}})_\sharp \gamma \llcorner A.$$

Recalling that Φ is precisely the first variation of $\mathcal{J}$ evaluated at γ, we have

$$
\begin{aligned}
0 \leq \left\langle \frac{\delta \mathcal{J}}{\delta \gamma}(\gamma), \bar{\gamma} - \gamma \right\rangle &= \int_{\mathcal{X}\times\mathcal{Y}} \Phi(\bar{x}, \bar{y}, \nu) \mathrm{d}\bar{\gamma} - \int_{\mathcal{X}\times\mathcal{Y}} \Phi(x, y, \nu) \mathrm{d}\gamma \\
&= \int_A \underbrace{(\Phi(x, T(x), \nu) - \Phi(x, y, \nu))}_{<0} \mathrm{d}\gamma \leq 0.
\end{aligned}
$$

This contradicts the fact that $\gamma(A) > 0$, and we conclude that γ is a Cournot-Nash equilibrium.

Conversely, suppose that γ is an equilibrium, from Def. (7.1) and it follows that

$$
\int_{\mathcal{X}\times\mathcal{Y}} \phi(x) \mathrm{d}\gamma = \int_{\mathcal{X}\times\mathcal{Y}} \Phi(x, y, \nu) \mathrm{d}\gamma.
$$

Hence, for any other admissible transportation plan $\bar{\gamma} \in \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$, it holds that

$$
\int_{\mathcal{X}\times\mathcal{Y}} \Phi(x, y, \nu) \mathrm{d}\bar{\gamma} \geq \int_{\mathcal{X}\times\mathcal{Y}} \phi(x) \mathrm{d}\bar{\gamma} = \int_{\mathcal{X}} \phi(x) \mathrm{d}\mu = \int_{\mathcal{X}\times\mathcal{Y}} \Phi(x, y, \nu) \mathrm{d}\gamma.
$$

From the fact that $\Phi(x, y, \nu) = \dfrac{\delta \mathcal{J}}{\delta \gamma}(\gamma)$, we conclude that γ is a critical point of $\mathcal{J}$.

Given $\gamma \in \Pi(\mu, \nu)$ that is a critical point of $\mathcal{J}$, hence is also a Cournot-Nash equilibrium, suppose in addition that ν is a distribution of finite social cost. By definition, μ-a.e. we have that $\phi(x) \leq \kappa(x)$ so that

$$
\int_{\mathcal{X}\times\mathcal{Y}} \Phi(x, y, \nu) \mathrm{d}\gamma = \int_{\mathcal{X}} \phi(x) \mathrm{d}\mu \leq \int_{\mathcal{X}} \kappa(x) \mathrm{d}\mu < +\infty.
$$

As any minimizer is a critical point, item (ii) follows. □

In the previous Theorem, the condition that the infimum is finite is non-trivial. Imposing further conditions on $\mathcal{E}$, such as strict convexity, this can be verified as done in [Blanchet and Carlier, 2016]. For the rest of this work, specially for the proof of convergence of Nash to Cournot-Nash equilibria, we concentrate on a case where $\mathcal{E}$ is given as the sum of a linear and an interaction term, as in [Santambrogio and Shim, 2021, Sadeghi Arjmand, 2022]. That is, when $\mathcal{E}$ can be written as follows

$$
\mathcal{E}(\nu) = \mathcal{L}(\nu) + \mathcal{H}(\nu, \nu), \text{ where } \mathcal{L}(\nu) = \int_{\mathcal{Y}} L \mathrm{d}\nu \text{ and } \mathcal{H}(\nu, \nu) = \int_{\mathcal{Y}\times\mathcal{Y}} H \mathrm{d}\nu \otimes \nu, \quad (7.17)
$$

and satisfy the hypothesis (H3)-(H7). It is then the sum of an individual cost L and an interaction cost. In this case, the lifted energy from (7.16) becomes

$$
\mathcal{J}(\gamma) \stackrel{\text{def.}}{=} \int_{\mathcal{X}\times\mathcal{Y}} c \mathrm{d}\gamma + \mathcal{L}(\nu) + \mathcal{H}(\nu, \nu), \quad (7.18)
$$

where $\nu = (\pi_{\mathcal{Y}})_\sharp \gamma$, and game with a continuum of players that we are interested is described by the cost

$$\Phi\colon \begin{cases} \mathcal{X} \times \mathcal{Y} \times \mathscr{P}(\mathcal{Y}) & \to \mathbb{R}_+ \cup \{+\infty\} \\ (x, y, \nu) & \mapsto c(x,y) + L(y) + 2\int_{\mathcal{Y}} H(y, y')\mathrm{d}\nu(y'). \end{cases} \tag{7.19}$$

In addition, we recall that we assume hypothesis (H3)-(H7) from the introduction. In particular, assumption (H3) that μ does not contain atoms is not restrictive, as discussed in Remark 7.5 below.

Remark 7.5. *If μ has atoms, we can work in the lifted space*

$$\mathcal{X}' = [0,1] \times \mathcal{X} \textit{ and } \mu' \in \Pi(\mathcal{L}^1 \llcorner [0,1], \mu),$$

that is a coupling between the Lebesgue measure on the interval $[0,1]$ and μ. On the other hand, there is a map $T' : \mathcal{X}' \to \mathcal{X}$ such that $T'_\sharp \mu' = \mu$, simply given by the projection $T' = \pi_{\mathcal{X}}$. Then we can formulate a new game with c replaced by $c'(x', y) = c(\pi_{\mathcal{X}}(x'), y)$, which remains l.s.c. in the product space $\mathcal{X}' \times \mathcal{Y}$. This new game will then satisfy all hypothesis (H3)-(H7).

As the integral of l.s.c. functionals, both $\mathcal{L}$ and $\mathcal{H}$ are l.s.c. as functionals over $\mathscr{P}(\mathcal{Y})$, see for instance [Santambrogio, 2015, Prop. 7.1]. Since the sub-level sets of $\mathcal{L}$ are compact, we would be able to prove existence of minimizers for $\mathcal{J}$, were it not for the term $\mathcal{H}$ that can be $+\infty$, for instance if H diverges in the diagonal.

In this case, we can characterize the cases where there the infimum is finite, and hence when we have existence, with a measure defined with the individual transportation cost c and the interaction energy H as follows: For $K \subset \mathcal{Y}$ compact, define

$$\operatorname{cap}_{c,H}(K) \overset{\text{def.}}{=} \left(\inf_{\varrho \in \mathscr{P}_{c,\mu}(K)} \int_{\mathcal{Y}\times\mathcal{Y}} H \mathrm{d}\varrho \otimes \varrho \right)^{-1}, \tag{7.20}$$

where $\mathscr{P}_{c,\mu}(K) \overset{\text{def.}}{=} \{\varrho \in \mathscr{P}(K) :\ \mathcal{W}_c(\mu, \varrho) < +\infty\}$. The capacity of an open set $U \subset \mathcal{Y}$ can then be defined through outer regularity

$$\operatorname{cap}_{c,H}(U) \overset{\text{def.}}{=} \sup\left\{\operatorname{cap}_{c,H}(K) : \quad K \subset U\right\},$$

and for a general set A as the inf of the same quantity among all the open sets U containing A. This defines a monotone set function that can be used to characterize when the infimum of $\mathcal{J}$ is finite.

Lemma 7.6. *Under hypotheses (H4)–(H7), it holds that*

$$\inf \mathcal{J} < +\infty \iff \operatorname{cap}_{c,H}(\{L < +\infty\}) > 0.$$

Proof. Starting with the direct implication, suppose that there exists $\gamma \in \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$ such that $\mathcal{J}(\gamma) < +\infty$. In particular, letting ν denote the second marginal of γ, it follows that $\mathcal{W}_c(\mu, \nu) < +\infty$ and $\operatorname{supp} \nu \subset \{L < +\infty\}$. It then follows that

$$\operatorname{cap}_{c,H}(\{L < +\infty\}) \geq \mathcal{J}(\gamma)^{-1} > 0.$$

Conversely, if $\operatorname{cap}_{c,H}(\{L < +\infty\}) > 0$, there is some $N \in \mathbb{N}$ such that

$$\operatorname{cap}_{c,H}(\{L \leq N\}) > 0.$$

Hence there is a measure ϱ concentrated over the compact set $\{L \leq N\}$ such that $\mathcal{W}_c(\mu, \varrho) < +\infty$ and $\mathcal{H}(\varrho, \varrho) < +\infty$. Taking γ as an optimal transportation plan between μ and ϱ gives that $\mathcal{J}(\gamma) < +\infty$. $\square$

The previous Lemma seems almost tautological, but in some particular cases there are strong results in the literature that characterize exactly which are the sets with positive capacity. In examples 2.1 and 2.1 we treat two models whose particular properties allow to verify the capacity criterion from Lemma 7.6.

Example 2.1. *In the Lagrangian mean field game of Mazanti et.al. the interaction term is shown to be bounded by the individual cost, that is, there is a constant $C > 0$ such that for all $\nu \in \mathscr{P}(\mathcal{Y})$ it holds that $\mathcal{H}(\nu, \nu) \leq C(1 + 2\mathcal{L}(\nu))$, which trivializes the capacity condition since $\mathcal{L}$ is not identically $+\infty$.*

Example 2.2. *Consider now a simpler case where $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$,*

$$c \in C_b(\mathbb{R}^d \times \mathbb{R}^d), \text{ and } H(y, \bar{y}) = |y - \bar{y}|^{-\alpha} \text{ for some } 0 < \alpha < d.$$

The condition that c is bounded implies that the set $\mathscr{P}_{c,\mu}(\mathbb{R}^d) = \mathscr{P}(\mathbb{R}^d)$ since the optimal transportation problem $\mathcal{W}_c(\mu, \varrho)$ is finite for any probability measure ϱ. This way, the capacity condition becomes

$$\operatorname{cap}_\alpha(\{L < +\infty\}) > 0,$$

where cap_s denotes the usual capacity, with $H(y, \bar{y}) = |y - \bar{y}|^{-\alpha}$. In this case, Frostman's Lemma, see [Falconer, 2004, Chap. 4.3] and [Ponce, 2016, Appendix B] or the original thesis of Frostman [Frostman, 1935], gives a charaterization of sets with strictly positive α-capacity in $\mathbb{R}^d$. Indeed, for a general Borel set $A \subset \mathbb{R}^d$ it holds that

$$d_H(A) = \inf\{s \geq 0 : \operatorname{cap}_s(A) = 0\},$$

where $d_H(A)$ denotes the Haussdorff dimension of the set A.

It follows that, in order to satisfy the capacity condition, it suffices to verify that the set $\{L < +\infty\}$ is of dimension bigger than α.

Remark 7.7. *Example 2.2 above motivates the characterization of sets with strictly positive $\operatorname{cap}_{c,H}$ for more general choices of c and H. As mentioned above the first difficulty is to choose a class of pairs (c, H) that do not make the infimum in the capacity $+\infty$. We have trivialized this question by considering c bounded, but it excludes the examples of Lagrangian Mean Field Games where $c(x, \sigma) = +\infty$ if $\sigma(0) \neq x$.*

2.2. Stability of the value function

In this paragraph our primary goal is to show the following estimate

$$\left| \inf_{\gamma \in \mathscr{P}_{\mu_0}(\mathcal{X}\times\mathcal{Y})} \mathcal{J} - \inf_{\gamma \in \mathscr{P}_{\mu_1}(\mathcal{X}\times\mathcal{Y})} \mathcal{J} \right| \leq C W_1(\mu_0, \mu_1) \text{ for some } C > 0, \tag{7.21}$$

where $\mu_0, \mu_1 \in \mathscr{P}(\mathcal{X})$ are two distribution of agents. Intuitively, it says that if we compute an equilibrium with respect to an estimation of the distribution of agents that is close to the real one, then one can expect that this estimated equilibrium is also close to equilibria for the real distribution, as it is quasi optimal to the minimization yield them.

To prove (7.21) we will exploit the *gluing method*, introduced in [Liu and Pfeiffer, 2023]. This method depends on the existence of a gluing operator as described in the following assumption:

(H8) There exists an operator $\mathcal{G} : \mathcal{X} \times \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ such that:

- $\mathcal{G}$ is consistent: for every $y \in \mathcal{Y}$ and $x \in \mathcal{X}$ it holds that $\mathcal{G}(x, x, y) = y$;
- there exists a positive constant $C > 0$ satisfying

$$\begin{aligned} c\left(x_1, \mathcal{G}(x_1, x_0, y)\right) &\leq c(x_0, y) + C d_{\mathcal{X}}(x_1, x_0), \\ L\left(\mathcal{G}(x_1, x_0, y)\right) &\leq L(y) + C d_{\mathcal{X}}(x_1, x_0), \\ H(\mathcal{G}(x_1, x_0, y), \mathcal{G}(\tilde{x}_1, \tilde{x}_0, \tilde{y})) &\leq H(y, \tilde{y}) + C\left(d_{\mathcal{X}}(x_1, x_0) + d_{\mathcal{X}}(\tilde{x}_1, \tilde{x}_0)\right). \end{aligned} \tag{7.22}$$

  for any pairs $x_0, x_1 \in \mathcal{X}$ and $y, \tilde{y} \in \mathcal{Y}$.

Essentially, hypothesis (H8) says that there is an operator that given some player of type x_0 choosing play y, any other player of type x_1 can choose a strategy $\mathcal{G}(x_1, x_0, y)$ paying a perturbation, of order $d_{\mathcal{X}}(x_0, x_1)$, of the cost paid by the first player. With this assumption we can prove that

Lemma 7.8 (Gluing method). *Let μ_0, μ_1 be probability measures in $\mathscr{P}(\mathcal{X})$ and $\gamma_0 \in \mathscr{P}_{\mu_0}(\mathcal{X} \times \mathcal{Y})$. Under the hypothesis (H8), there exists a measure $\gamma_1 \in \mathscr{P}_{\mu_1}(\mathcal{X} \times \mathcal{Y})$ such that*

$$\mathcal{J}(\gamma_1) \leq \mathcal{J}(\gamma_0) + 4C W_1(\mu_0, \mu_1),$$

where W_1 denotes the Kantorovitch-Rubinstein distance.

Proof. Given measures $\mu_0, \mu_1 \in \mathscr{P}(\mathcal{X})$ and let $\pi_{1,0} \in \Pi(\mu_1, \mu_0)$ be an optimal transportation plan between μ_0 and μ_1, *i.e.*

$$\int_{\mathcal{X}_1 \times \mathcal{X}_1} d_{\mathcal{X}}(x_1, x_0) \mathrm{d}\pi_{1,0}(x_1, x_0) = W_1(\mu_1, \mu_0),$$

where $\mathcal{X}_0$ and $\mathcal{X}_1$ are identical copies of the space $\mathcal{X}$.

Let $\Gamma \in \mathscr{P}(\mathcal{X}_1 \times \mathcal{X}_0 \times \mathcal{Y})$ denote the gluing of $\pi_{1,0}$ and γ_0, as in Lemma 1.9 from Chap. 1 or Lemma 5.3.2 of [Ambrosio et al., 2008], so that $(\pi_{\mathcal{X}_0,\mathcal{Y}})_\sharp \Gamma = \gamma_0$ and $(\pi_{\mathcal{X}_1,\mathcal{X}_0})_\sharp \Gamma = \pi_{1,0}$. The measure γ_1 is then defined as $\gamma_1 := (\pi_{\mathcal{X}_1}, \mathcal{G})_\sharp \Gamma$. It follows from these definitions that

$$\mathcal{J}(\gamma_0) = \int_{\mathcal{X}_0 \times \mathcal{Y}} (c + L + H)\, \mathrm{d}\Gamma \otimes \Gamma$$
$$\mathcal{J}(\gamma_1) = \int_{\mathcal{X}_1 \times \mathcal{Y}} (c + L + H)\, \mathrm{d}\left((\pi_{\mathcal{X}_1}, \mathcal{G})_\sharp \Gamma \right) \otimes \left((\pi_{\mathcal{X}_1}, \mathcal{G})_\sharp \Gamma \right).$$

Using the definition of the gluing operator from (H8), we get the following estimates

$$\begin{aligned}
\mathcal{J}(\gamma_1) &= \int_{(\mathcal{X}_0 \times \mathcal{X}_1 \times \mathcal{Y})^2} (c(x_1, \mathcal{G}(x_1, x_0, y)) + L(\mathcal{G}(x_1, x_0, y)) + \\
&\qquad H(\mathcal{G}(x_1, x_0, y), \mathcal{G}(\bar{x}_1, \bar{x}_0, \bar{y})))\mathrm{d}\Gamma \otimes \Gamma \\
&\leq \int_{(\mathcal{X}_0 \times \mathcal{X}_1 \times \mathcal{Y})^2} (c + L + H)\, \mathrm{d}\Gamma \otimes \Gamma + C \int_{(\mathcal{X}_0 \times \mathcal{X}_1)^2} (3 d_{\mathcal{X}}(x_1, x_0) + d_{\mathcal{X}}(\tilde{x}_1, \tilde{x}_0))\mathrm{d}\Gamma \otimes \Gamma \\
&= \mathcal{J}(\gamma_0) + 4C \int_{\mathcal{X}_1 \times \mathcal{X}_0} d_{\mathcal{X}}(x_1, x_0) \mathrm{d}\pi_{1,0} \\
&= \mathcal{J}(\gamma_0) + 4C W_1(\mu_0, \mu_1).
\end{aligned}$$

The result follows. □

The previous Lemma 7.8 will also be useful in the proof of Γ convergence in the open closed loop formulation. For now, we use it to prove the following:

Theorem 7.9. *Under the hypothesis (H8), the stability inequality* (7.21) *for the value function holds.*

Proof. Let $\gamma_0 \in \mathscr{P}_{\mu_0}(\mathcal{X} \times \mathcal{Y})$ optimal, so that

$$\mathcal{J}(\gamma_0) = \min_{\mathscr{P}_{\mu_0}(\mathcal{X} \times \mathcal{Y})} \mathcal{J}.$$

So let γ_1 be the measure obtained from the gluing method in Lemma 7.8. It then holds that

$$\inf_{\mathscr{P}_{\mu_1}(\mathcal{X} \times \mathcal{Y})} \mathcal{J} - \inf_{\mathscr{P}_{\mu_0}(\mathcal{X} \times \mathcal{Y})} \mathcal{J} \leq \mathcal{J}(\gamma_1) - \mathcal{J}(\gamma_0) \leq 4C W_1(\mu_0, \mu_1),$$

where C is the constant from (H8). Changing the roles of μ_0 and μ_1, we conclude. □

Example 2.3 (Back to example 2.1)**.** *Let us now give further details for the Lagrangian MFG discussed in example 2.1. We consider a model where a population of agents tries to read a target set in minimal time under pair-wise interactions.*

For simplicity, let $\Omega \subset \mathbb{R}^d$ be a convex set, and let $\Gamma \subset \Omega$ be the target set of the players. In this case $\mathcal{X} = \Omega$ and $\mathcal{Y} = C(\mathbb{R}_+; \Omega)$, the continuous functions with values in Ω. For $\sigma \in \mathcal{Y}$ we set

$$\tau(\sigma) \stackrel{\text{def.}}{=} \inf\{t \geq 0 : \sigma(t) \in \Gamma\},$$

the minimal time to reach the target and

$$c(x_0, \sigma) = \begin{cases} 0, & \textit{if } \sigma(0) = x_0, \\ +\infty, & \textit{otherwise.} \end{cases}$$

The individual and interaction energies are given by

$$L(\sigma) \overset{\text{def.}}{=} \int_0^{\tau(\sigma)} \ell(t, \sigma(t), \dot{\sigma}(t)) \mathrm{d}t + \Psi(\sigma_\tau),$$
$$H(\sigma, \bar{\sigma}) \overset{\text{def.}}{=} \int_0^{\tau(\sigma) \wedge \tau(\bar{\sigma})} h(t, \sigma(t), \dot{\sigma}(t), \bar{\sigma}(t), \dot{\bar{\sigma}}(t)) \mathrm{d}t.$$

For simplicity, we assume that ℓ and h are bounded non-negative functions, that σ remains constant after reaching Γ for the first time and that if $\dot{\sigma}(t) = 0$, then $\ell(t, \sigma(t), \dot{\sigma}(t)) = h(t, \sigma(t), \dot{\sigma}(t), \bar{\sigma}(t), \dot{\bar{\sigma}}(t)) = 0$.

In order to have a small perturbation of the energies, the easiest way is to preserve the stopping time, hence given σ such that $\sigma(0) = x_0$ we search for a curve of the form

$$\sigma_{x_1}(t) \overset{\text{def.}}{=} \begin{cases} \left(1 - \frac{t}{t_0}\right) x_1 + \frac{t}{t_0} \sigma(t_0), & \textit{if } t \in [0, t_0], \\ \sigma(t), & \textit{otherwise.} \end{cases}$$

Therefore, choosing $t_0 \leq \min\{\tau_\sigma, |x_0 - x_1|\}$ we obtain that

$$L(\sigma_{x_1}(t)) \leq L(\sigma) + \int_0^{t_0} \ell(t, \sigma_{x_1}(t), \dot{\sigma}_{x_1}(t)) \mathrm{d}t \leq L(\sigma) + C|x_0 - x_1|.$$

An analogous reasoning for H gives the required result.

3. The N-player game: open and closed loop decisions

In this section we consider an N-player formulation of the abstract game we have discussed in Section 1.2. Given an i.i.d. sample of agents $(X_i)_{i \in \mathbb{N}}$ with common law $\mu \in \mathscr{P}(\mathcal{X})$, where the first N elements represent the type of the agents in our N-player game. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ induced by the sample $(X_i)_{i \in \mathbb{N}}$, so that $\Omega = \mathcal{X}^{\otimes \mathbb{N}}$ represents all the possible realizations of this sampling, $\mathcal{F}$ is the σ-algebra generated by the random variables X_i and, $\mathbb{P} = \mu^{\otimes \mathbb{N}}$.

There are two possible information structures in the decision-making of the players we can think of. The first and more natural one is a *closed loop* information structure: once player i has the knowledge of its type x_i, they choose a play y_i among the set of feasible plays, that is such that $c(x_i, y_i) < \infty$. Alternatively, one can imagine an *open loop* information structure, where each player chooses beforehand some sort of optimal execution strategy, in the sense that it chooses a map from $\mathcal{X}$ to $\mathcal{Y}$, and given only the knowledge of the realization of its own type, it follows this optimal execution plan. Assuming that the state variables of all players are identically distributed and that those are rational and indistinguishable, this model implicitly takes into account the distribution of the others since each player can expect the others to play as optimally as themselves.

First let us recall the definition of Nash-equilibria and introduce some notation. An N-player game in pure strategies is a tuple $(g_i, S_i)_{i=1}^N$ where S_i denotes the space of admissible plays for player i and g_i is a function

$$g_i : S_i \times S_{-i} \ni (x_i, x_{-i}) \mapsto g_i(x_i, x_{-i}) \in \mathbb{R} \text{ where } S_{-i} \stackrel{\text{def.}}{=} \prod_{j \neq i} S_j.$$

Given an admissible profile of strategies $(x_j)_{j=1}^N$, x_{-i} corresponds to the tuple of strategies deprived of x_i and the quantity $g_i(x_i, x_{-i})$ represents the cost of player i choosing x_i given that the remaining players choose x_{-i}.

A game in mixed strategies, or mixed plays, is a tuple $(g_i, \mathscr{P}(S_i))_{i=1}^N$, such that

$$g_i(\nu_i, \nu_{-i}) \stackrel{\text{def.}}{=} \int_{S_i \times S_{-i}} g_i(x_i, x_{-i}) \mathrm{d}\nu_i \otimes \nu_{-i}(x_i, x_{-i}),$$

where $\nu_{-i} \stackrel{\text{def.}}{=} \nu_1 \otimes \cdots \otimes \nu_{i-1} \otimes \nu_{i+1} \otimes \nu_N$.

In the sequel, we recall the definition of Nash equilibrium, but notice that since a game in mixed plays is just a game in pure strategies with a different set of admissible plays, we only write it explicitly for the pure strategies' formulation.

Definition 7.10. *A Nash equilibrium of an N-players game $(g_i, S_i)_{i=1}^N$ is a profile of strategies $(x_i)_{i=1}^N$ where no player has a unilateral incentive to deviate, that is, all $i = 1, \ldots, N$ if holds that*

$$g_i(x_i, x_{-i}) \leq g_i(x_i', x_{-i}) \textit{ for all } x_i' \in S_i.$$

We proceed with the definition of the closed an open loop games described above.

3.1. Closed and Open loop information structures

In the sequel, we discuss both of these information structures. We show that each of them is associated with a functional, whose minimization yields Nash equilibria, that is these games also have a potential structure. The remarkable thing is that both these functionals Γ-converge, each on its appropriated topology, to the same limit, being the functional $\mathcal{J}$ as proved in Section 4.

Closed loop game

Given an event $\omega = (x_i)_{i\in\mathbb{N}}$, each player seeks to

$$\begin{aligned}
\underset{\nu_i\in\mathscr{P}(\mathcal{Y})}{\text{minimize}}\, J_{\omega,i}(\nu_i,\nu_{-i}) &\stackrel{\text{def.}}{=} \int_{\mathcal{Y}} c(x_i,y)\mathrm{d}\nu_i + \int_{\mathcal{Y}} L\mathrm{d}\nu_i + \frac{2}{N}\sum_{j\neq i}\int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu_i\otimes\nu_j \\
&= \int_{\mathcal{Y}} c(x_i,y)\mathrm{d}\nu_i + \mathcal{L}(\nu_i) + \frac{2}{N}\sum_{j\neq i}\mathcal{H}(\nu_i,\nu_j).
\end{aligned} \tag{7.23}$$

Notice that we have written the relaxed formulation in mixed strategies and a profile in pure strategies is just a tuple $(\nu_i)_{i=1}^N$ such that $\nu_i = \delta_{y_i}$ for all players.

In (7.23) we have excluded the cross terms $\mathcal{H}(\nu_i,\nu_i)$ for two reasons. From a modeling perspective, it makes sense that agents do not interact with themselves, which is exactly what this term represents. In addition, this formulation makes sense in pure strategies even in the case where the pairwise interaction diverges in the diagonal, $H(y,y) = +\infty$, for instance the case of an electrostatic interaction, see *e.g.* example 2.2. If we had kept the self interaction in this case, any pure strategy would yield the player the value $+\infty$. We shall also consider the case that the pairwise interaction vanishes in the diagonal, *i.e.* $H(y,y) = 0$. For the formulation in pure strategies this does not affect the cost functions of each player, but in mixed plays the diagonal terms $\mathcal{H}(\nu_i,\nu_i)$ can be included, see Proposition 7.12.

Open loop game

In the open loop formulation, as each player chooses a strategy before having the knowledge of the realization of the sample, the type of player i is better described by the random variable X_i and an admissible strategy must be given by a measurable family $(\nu^x)_{x\in\mathcal{X}} \subset \mathscr{P}(\mathcal{Y})$, being uniquely described with a random probability measure.

We recall the notion of *random probability measure* introduced in Section 2.2 of Chap. 1, see also [Crauel, 2002, Kallenberg et al., 2017]. Given a probability space $(\Omega,\mathcal{F},\mathbb{P})$, a random measure $\boldsymbol{\mu}$ is a measurable map

$$\omega \mapsto \boldsymbol{\mu}(\omega) \in \mathscr{P}(\mathcal{X}),$$

w.r.t. the Borel σ-algebra induced by the narrow topology of $\mathscr{P}(\mathcal{X})$. We let $\mathscr{P}_\Omega(\mathcal{X})$ denote the space of all random probability measures over $\mathcal{X}$ equipped with the *narrow topology of random measures*, that is the topology induced by the duality with *random bounded continuous functions*, see Definition 1.11 in Chap. 1.

With this terminology, in the context of the open loop game, each player must explicit its play for any realization of the random variable X_i describing them. Therefore, instead of choosing a deterministic strategy, for each state x a player chooses some ν^x, in such a way that the family $(\nu^x)_{x\in\mathcal{X}}$ is a measurable family, as in Definition 1.7 from Chap. 1. The criterion that each player seeks to minimize is then

$$\min_{\boldsymbol{\nu}_i \in \mathscr{P}_\Omega(\mathcal{Y})} \mathcal{J}_{\Omega,i}\left(\boldsymbol{\nu}_i, \boldsymbol{\nu}_{-i}\right) \stackrel{\text{def.}}{=} \mathbb{E}_{\mathbb{P}}\left[\int_{\mathcal{Y}} c(X_i, y)\mathrm{d}\nu_i^{X_i} + \mathcal{L}(\nu_i^{X_i}) + \frac{2}{N}\sum_{i\neq j}\mathcal{H}\left(\nu_i^{X_i}, \nu_j^{X_j}\right)\right], \tag{7.24}$$

where $\boldsymbol{\nu}_i = \nu_i^{X_i}$ for some measurable map $(\nu_i^x)_{x\in\mathcal{X}}$.

A profile $(\boldsymbol{\nu}_i)_{i=1}^N$ is pure if each $\boldsymbol{\nu}_i$ is a Dirac delta with full probability and can then be described with a map as measures of the form $\boldsymbol{\nu}_i = \delta_{T_i(X_i)}$. The formulation in pure strategies can then be expressed as

$$\min_{T_i} \mathcal{J}_{\Omega,i}(T_i, T_{-i}) = \mathbb{E}_{\mathbb{P}}\left[c(X_i, T_i(X_i)) + L(T_i(X_i)) + \frac{2}{N}\sum_{i\neq j}\mathcal{H}(T_i(X_i), T_j(X_j))\right]. \tag{7.25}$$

3.2. Potential structure for N-player games

As in the game with a continuum of players, the N-player games described above also enjoy a potential structure, in both pure and mixed strategies. In this section we argue that each of the previously described games admits a functional whose minimizers generate Nash equilibria for their corresponding game. To describe Nash-equilibria as transportation plans, we will consider plans whose marginals are given by the empirical measure of the random sample $(X_i)_{i\in\mathbb{N}}$

$$\boldsymbol{\mu}_N \stackrel{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{X_i}.$$

Under assumption (H3), that μ has no atoms, with full $\mathbb{P}$-probability, the event $\omega = (x_i)_{i\in\mathbb{N}}$ has distinct realizations, *i.e.* $x_i \neq x_j$ for all $i \neq j$. In these case, the realization of the empirical measure $\boldsymbol{\mu}_N$ is uniquely represented as

$$\boldsymbol{\mu}_N(\omega) = \frac{1}{N}\sum_{i=1}^N \delta_{X_i(\omega)}.$$

For every such event, there is a bijection between the strategy profiles $(\nu_i)_{i=1}^N$ and the measures $\gamma_N \in \mathscr{P}_{\boldsymbol{\mu}_N(\omega)}(\mathcal{X}\times\mathcal{Y})$ by means of the disintegration theorem, which guarantees that each such measure is uniquely written as

$$\gamma_N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \nu^{x_i}. \tag{7.26}$$

This representation can be seen as a lift of a profile of strategies $(\nu_i = \nu^{x_i})_{i=1}^N$ to the space of plans $\mathscr{P}(\mathcal{X} \times \mathcal{Y})$. We can define a potential function in the lifted space as

$$\mathcal{J}_{\omega,N}(\gamma_N) \overset{\text{def.}}{=} \begin{cases} \displaystyle\int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\gamma_N + \frac{1}{N}\sum_{i=1}^N \mathcal{L}\left(\nu^{x_i}\right) \\ \displaystyle\quad + \quad \frac{1}{N^2}\sum_{j\neq i} \mathcal{H}\left(\nu^{x_i}, \nu^{x_j}\right), & \text{if } \gamma_N \in \mathscr{P}_{\mu_N(\omega)}(\mathcal{X}\times\mathcal{Y}), \\ +\infty, & \text{otherwise,} \end{cases} \tag{7.27}$$

where $(\nu^{x_i})_{i=1}^N$ denotes the unique profile obtained though the representation (7.26).

The formulation in pure strategies can then be obtained by considering the following potential functional

$$J_{\omega,N}(y_1,\ldots,y_N) \overset{\text{def.}}{=} \mathcal{J}_{\omega,N}\left(\frac{1}{N}\sum_{i=1}^N \delta_{(x_i,y_i)}\right). \tag{7.28}$$

This is equivalent to restricting $\mathcal{J}_{\omega,N}$ to the set

$$\mathscr{P}^{\text{pure}}_{\mu_N}(\mathcal{X}\times\mathcal{Y}) \overset{\text{def.}}{=} \left\{\gamma_N \in \mathscr{P}_{\mu_N}(\mathcal{X}\times\mathcal{Y}) :\ \gamma_N = \frac{1}{N}\sum_{i=1}^N \delta_{(x_i,y_i)}\right\}.$$

Remark 7.11. *Assumption (H3), that μ is atomless, is particularly relevant here in order to make the disintegration representation uniquely well-defined with full probability. As showed in Remark 7.5, this is not restrictive since we can replace the space $\mathcal{X}$ with $\mathcal{X}' = [0,1]\times\mathcal{X}$. In the context of the sampling, we would obtain an i.i.d. sequence $(X_i')_{i\in\mathbb{N}} = (T_i, X_i)_{i\in\mathbb{N}}$ with common law given by $\mu' = \mathcal{L}^1 \llcorner [0,1]\otimes\mu$, which has no atoms since the Lebesgue measure is non-atomic. Therefore, any event $\omega' = ((T_i,X_i) = (t_i,x_i))_{i\in\mathbb{N}}$ is such that $(t_i,x_i)\neq(t_j,x_j)$ with full probability.*

For the game in open loop, we let $\boldsymbol{\mu}_N \in \mathscr{P}_\Omega(\mathcal{X})$, recall Def. 1.10 from Chap. 1, be the random measure obtained via the sample of random variables $\boldsymbol{\mu}_N \overset{\text{def.}}{=} \dfrac{1}{N}\displaystyle\sum_{i=1}^N \delta_{X_i}$, and we define the space of random transportation plans

$$\mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X}\times\mathcal{Y}) \overset{\text{def.}}{=} \left\{\boldsymbol{\gamma}_N = \boldsymbol{\mu}_N\otimes\nu^x = \frac{1}{N}\sum_{i=1}^N \delta_{X_i}\otimes\nu^{X_i} :\ (\nu^x)_{x\in\mathcal{X}} \text{ is measurable}\right\},$$

where we recall the definition of measurable family of measures from Def. 1.7. The potential function in open loop formulation is defined as

$$\mathcal{J}_{\Omega,N}(\boldsymbol{\gamma}_N) \overset{\text{def.}}{=} \begin{cases} \displaystyle\mathbb{E}_{\mathbb{P}}\left[\int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\boldsymbol{\gamma}_N + \frac{1}{N}\sum_{i=1}^N \mathcal{L}\left(\nu^{X_i}\right)\right. \\ \displaystyle\qquad\left. + \frac{1}{N^2}\sum_{i\neq j}\mathcal{H}\left(\nu^{X_i},\nu^{X_j}\right)\right], & \text{if } \boldsymbol{\gamma}_N \in \mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X}\times\mathcal{Y}), \\ +\infty, & \text{otherwise.} \end{cases} \tag{7.29}$$

As for the closed loop formulation, there is a canonical bijection between the set of random measures $\mathscr{P}_{\Omega,\mu}(\mathcal{X})$ and the set of symmetric strategy profiles, obtained through the disintegration theorem.

First notice that all the above potential functionals admit minimizers since c, L and H are l.s.c. and L has compact sub-level sets. We shall prove that minimizers for each potential functional yield Nash equilibria for the corresponding game and, in the case that H vanishes in the diagonal and is strictly positive elsewhere, we can prove that any minimizer induces a Nash equilibrium in pure strategies.

Proposition 7.12. *The following assertions hold:*

(i) It is equivalent to minimize $J_{\omega,N}$ and $\mathcal{J}_{\omega,N}$, minimizers of the latter are supported on the set of minimizers of the former and it holds that

$$\min_{\mathcal{Y}^N} J_{\omega,N} = \min_{\mathscr{P}^{pure}_{\mu_N}(\mathcal{X}\times\mathcal{Y})} \mathcal{J}_{\omega,N} = \min_{\mathscr{P}_{\mu_N}(\mathcal{X}\times\mathcal{Y})} \mathcal{J}_{\omega,N}. \tag{7.30}$$

(ii) Let

$$(y_i)_{i=1}^N \in \operatorname{argmin} J_{\omega,N}, \quad \gamma_N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \nu_{\omega,i} \in \operatorname{argmin} \mathcal{J}_{\omega,N},$$

$$\boldsymbol{\gamma}_N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i} \otimes \nu^{X_i} \in \operatorname{argmin} \mathcal{J}_{\Omega,N}$$

then $(y_i)_{i=1}^N$, $(\nu_{\omega,i})_{i=1}^N$ induce Nash equilibria for the game (7.23) and $\left(\nu^{X_i}\right)_{i=1}^N$ induces an equilibrium for (7.24).

(iii) Suppose that H vanishes on the diagonal, that it is strictly positive outside it and that we allow for self interaction in our game, i.e. *we replace $\mathcal{L}$ with $\mathcal{L}_H(\gamma) = \mathcal{L}(\gamma) + \mathcal{H}(\gamma,\gamma)$. Then minimizers of $\mathcal{J}_{\omega,N}$ are of the form*

$$\gamma_N = \frac{1}{N}\sum_{i=1}^N \delta_{(x_i,y_i)}, \text{ where } (y_1,\dots,y_N) \in \operatorname{argmin} J_{\omega,N}. \tag{7.31}$$

(iv) If $H = +\infty$ in the diagonal, any minimizer of $\mathcal{J}_{\omega,N}$ is atomless.

Proof. The first equality in (7.30) comes from the bijection between the set of pure equilibrium measures and $\mathcal{Y}^{\otimes N}$. The second is a direct consequence of the fact that the measures γ_N in the domain of $\mathcal{J}_{\omega,N}$ can be written as

$$\gamma_N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \otimes \nu_i,$$

so that we can write

$$\mathcal{J}_{\omega,N}(\gamma_N) = \int_{\mathcal{Y}^{\otimes N}} J_{\omega,N}(y_1,\dots,y_N) \mathrm{d}\nu_1 \otimes \cdots \otimes \nu_N.$$

Then for any admissible γ_N, we have

$$\min_{\mathscr{P}_{\mu_N(\omega)}(\mathcal{X}\times\mathcal{Y})} \mathcal{J}_{\omega,N}(\gamma) \geq \min_{\mathcal{Y}^{\otimes N}} J_{\omega,N}.$$

Taking γ_N with second marginal supported on the set of minimizers of $J_{\omega,N}$ gives the result.

To check assertion (ii), notice that a minimizer of $J_{\omega,N}$ will directly satisfy the definition of Nash equilibrium by considering variations of the minimum index by index. Indeed, let $y = (y_i)_{i=1}^N$ be a minimizer and suppose that player i deviates, choosing $\bar{y}_i$ instead of y_i and yielding a new profile $\bar{y} = (y_1, \ldots, y_{i-1}, \bar{y}_i, y_{i+1}, \ldots, y_N)$. First notice that from the symmetry of H we have

$$\begin{aligned}\sum_{j\neq k} H(y_j, y_k) &= \sum_{j\neq k, j\neq i, k\neq i} H(y_j, y_k) + \sum_{k\neq i} H(y_i, y_k) + \sum_{j\neq i} H(y_j, y_i) \\ &= \sum_{j\neq k, j\neq i, k\neq i} H(y_j, y_k) + 2\sum_{j\neq i} H(y_j, y_i),\end{aligned}$$

so the minimality of y gives

$$\begin{aligned}J_{\omega,N}(y) &= \frac{1}{N}\left(\sum_{j\neq i}[c(x_j, y_j) + L(y_j)] + \frac{1}{N}\sum_{j\neq k, j, k\neq i} H(y_j, y_k) + J_{\omega,i}(y_i, y_{-i})\right) \\ &\leq \frac{1}{N}\left(\sum_{j\neq i}[c(x_j, y_j) + L(y_j)] + \frac{1}{N}\sum_{j\neq k, j, k\neq i} H(y_j, y_k) + J_{\omega,i}(\bar{y}_i, y_{-i})\right) = J_{\omega,N}(\bar{y}),\end{aligned}$$

where we recall that $J_{\omega,i}(\bar{y}_i, y_{-i})$ is given as in (7.23), by considering Dirac measures. Canceling out the repeated terms we obtain that $J_{\omega,i}(y_i, y_{-i}) \leq J_{\omega,i}(\bar{y}_i, y_{-i})$, meaning that the profile $(y_1, \ldots, y_N)$ is a Nash equilibria in pure strategies. Similarly, if $\gamma_N = \dfrac{1}{N}\displaystyle\sum_{i=1}^N \delta_{x_i} \otimes \nu_{\omega,i}$ is a minimizer for $\mathcal{J}_{\omega,N}$, then the profile $(\nu_{\omega,1}, \ldots, \nu_{\omega,N})$ is a Nash equilibria in mixed strategies.

To prove (iii), notice that from item (i) and the fact that $H \geq 0$ it holds that

$$\inf \mathcal{J}_{\omega,N} \geq \inf J_{\omega,N} + \frac{1}{N^2}\sum_{i=1}^N \mathcal{H}(\nu_{\omega,i}, \nu_{\omega,i}) \geq \inf \mathcal{J}_{\omega,N}.$$

Which means that $\mathcal{H}(\nu_{\omega,i}, \nu_{\omega,i}) = 0$ for all $i = 1, \ldots, N$, and since H only vanishes in the diagonal, it must hold that $\nu_{\omega,i} = \delta_{y_i}$. From (i) and the previous argument, any minimizer of $\mathcal{J}$ is of the form of (7.31).

With a dual reasoning, if γ_N has an atom, *i.e.* if there is a point where $\gamma_N(\{x_i, y_i\}) > 0$, and H explodes in the diagonal, the self interaction term gives $\mathcal{J}_{\omega,N}(\gamma_N) = +\infty$ and it cannot be a minimizer. $\square$

4. Γ-CONVERGENCE

In this Section we prove the Γ-convergence result for both the closed loop and the open loop formulation. Despite the more complicated topology of random probability measures used in the open loop formulation, the convergence proof is actually easier and hence we shall start with this case and then move on the closed loop formulation.

4.1. Γ-CONVERGENCE FOR THE OPEN LOOP FORMULATION

Before passing to the Γ convergence result, we will need to characterize the cluster points of random measures in the set $\mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X})$ when μ_N is a sequence of empirical measures.

Lemma 7.13. *Let $\boldsymbol{\mu}_N$ be a sequence of empirical measures of an i.i.d. sample of law μ. Let $\boldsymbol{\gamma}_N$ be a sequence of random measures such that $\boldsymbol{\gamma}_N \in \mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X}\times\mathcal{Y})$ for all $N\in\mathbb{N}$ and converging in the narrow convergence of probability measures to a random measure $\boldsymbol{\gamma}$. Then $\boldsymbol{\gamma}$ is a deterministic measure in the sense that there is a measure $\gamma \in \mathscr{P}_{\mu}(\mathcal{X}\times\mathcal{Y})$ such that $\boldsymbol{\gamma}(\omega)=\gamma$ almost surely.*

Remark 7.14. *The major difficulty of the following proof comes from the fact that the conditional expectation is not continuous w.r.t. weak convergence in general. In order words, if a sequence of measures $(\gamma_N)_{N\in\mathbb{N}}$ converging weakly to γ has the following disintegration representation $\gamma_N=\mu\otimes\nu_N^x$ and $\gamma=\mu\otimes\nu^x$, it does not hold in general that $\nu_N^x \xrightarrow[N\to\infty]{} \nu^x$, not even for a.e. x.*

Proof. If $\boldsymbol{\gamma}_N$ converges weakly to $\boldsymbol{\gamma}$, it is a priori just a random probability measure in $\mathscr{P}_{\Omega}(\mathcal{X}\times\mathcal{Y})$. Hence, we first need to show that $\boldsymbol{\gamma}(\omega)\in\mathscr{P}_{\mu}(\mathcal{X}\times\mathcal{Y})$ with probability 1. For any $f\in C_b(\mathcal{X})$, we have

$$\int_{\mathcal{X}} f(x)\mathrm{d}(\pi_{\mathcal{X}})_{\sharp}\boldsymbol{\gamma}(\omega)=\lim_{N\to\infty}\int_{\mathcal{X}} f\circ\pi_{\mathcal{X}}\mathrm{d}\boldsymbol{\gamma}_N(\omega)=\lim_{N\to\infty}\int_{\mathcal{X}} f\mathrm{d}\boldsymbol{\mu}_N(\omega)=\int_{\mathcal{X}} f\mathrm{d}\mu,$$

where the last limit is true almost surely from the Glivenko-Cantelli law of large numbers. As a consequence, by definition of $\mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X}\times\mathcal{Y})$ and disintegration applied to $\boldsymbol{\gamma}$ we have the following representations

$$\boldsymbol{\gamma}_N=\boldsymbol{\mu}_N\otimes\nu_N^x,\quad \boldsymbol{\gamma}=\mu\otimes\boldsymbol{\nu}^x,$$

where $(\nu_N^x)_{x\in\mathcal{X}}\subset\mathscr{P}(\mathcal{Y})$ is a sequence of measurable maps of deterministic measures and $(\boldsymbol{\nu}^y)_{y\in Y}$ is a family of random measures in $\mathscr{P}_{\Omega}(X)$. Notice that while the stochasticity of $\boldsymbol{\gamma}_N$ is concentrated in the $\mathcal{X}$-marginal, we cannot say for now that the same is true for $\boldsymbol{\gamma}$ and our goal is precisely to show that the disintegration family $(\boldsymbol{\nu}^x)_{x\in\mathcal{X}}$ is a family of non-random probability measures.

Recall the definition of expectation measure in (1.7), for any $\varphi \in C_b(\mathcal{X} \times \mathcal{Y})$, we have

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\mathbb{E}\boldsymbol{\gamma}_N = \mathbb{E}\left[\int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\boldsymbol{\gamma}_N\right] \xrightarrow[N\to\infty]{} \mathbb{E}\left[\int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\boldsymbol{\gamma}\right] = \int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\mathbb{E}\boldsymbol{\gamma},$$

so that $\mathbb{E}\boldsymbol{\gamma}_N \xrightharpoonup[N\to\infty]{} \mathbb{E}\boldsymbol{\gamma}$. In the sequel, we check that $\mathbb{E}\boldsymbol{\gamma}_N = \mu \otimes \nu_N^x$. Indeed, still using duality, for $\varphi \in C_b(\mathcal{X} \times \mathcal{Y})$ we have that

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\mathbb{E}\boldsymbol{\gamma}_N = \frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\int_{\mathcal{Y}} \varphi(X_i, y)\mathrm{d}\nu_N^{X_i}\right] = \int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}(\mu \otimes \nu_N^x).$$

To finish the proof it suffices to show that for any real valued, bounded and $(\Omega, \mathcal{F}, \mathbb{P})$-adapted random variable Θ and $\varphi \in C_b(\mathcal{X} \times \mathcal{Y})$, it holds that

$$\Delta_{N,\Theta} \stackrel{\text{def.}}{=} \left|\mathbb{E}\left[\Theta \int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\boldsymbol{\gamma}_N\right] - \mathbb{E}\left[\Theta\right] \int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\mathbb{E}\boldsymbol{\gamma}_N\right| \xrightarrow[N\to\infty]{} 0, \tag{7.32}$$

since then we will have that

$$\mathbb{E}\left[\Theta\left(\int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}(\boldsymbol{\gamma} - \mathbb{E}\boldsymbol{\gamma})\right)\right] = 0$$

for any bounded random variable Θ, meaning that $\boldsymbol{\gamma} = \mathbb{E}\boldsymbol{\gamma}$ almost surely.

For this, we will use *Hoeffding's inequality*, which states that if $Z_1, \dots, Z_N$ are i.i.d. real variables such that $a \le Z_i \le b$ for all i almost surely, then

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^N Z_i - \mathbb{E}[Z_1]\right| \ge \varepsilon\right) \le 2\exp\left(-\frac{2N\varepsilon^2}{(b-a)^2}\right). \tag{7.33}$$

Notice that we can rewrite

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\boldsymbol{\gamma}_N = \frac{1}{N}\sum_{i=1}^N \tilde{\varphi}_{i,N} \text{ where } \tilde{\varphi}_{i,N} \stackrel{\text{def.}}{=} \int_{\mathcal{Y}} \varphi(X_i, y)\mathrm{d}\nu_N^{X_i},$$

so that $(\tilde{\varphi}_{i,N})_{i=1}^N$ are i.i.d., $|\tilde{\varphi}_{i,N}| \le \|\varphi\|_{L^\infty}$ and

$$\mathbb{E}\tilde{\varphi}_{1,N} = \int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\mu \otimes \nu_N^x = \int_{\mathcal{X}\times\mathcal{Y}} \varphi \mathrm{d}\mathbb{E}\boldsymbol{\gamma}_N.$$

So setting

$$A_\varepsilon \stackrel{\text{def.}}{=} \left\{\left|\frac{1}{N}\sum_{i=1}^N \tilde{\varphi}_{i,N} - \mathbb{E}\tilde{\varphi}_{1,N}\right| \ge \varepsilon\right\},$$

we can use Hoeffding's inequality to bound the LHS of (7.32)

$$\begin{aligned}\Delta_{N,\Theta} &\le \mathbb{E}\left[|\Theta|\left|\frac{1}{N}\sum_{i=1}^{N}\tilde{\varphi}_{i,N} - \mathbb{E}\tilde{\varphi}_{1,N}\right|\right] \\ &\le \|\Theta\|_{L^\infty}\int_{A_\varepsilon}\left|\frac{1}{N}\sum_{i=1}^{N}\tilde{\varphi}_{i,N} - \mathbb{E}\tilde{\varphi}_{1,N}\right|\mathrm{d}\mathbb{P} + \|\Theta\|_{L^\infty}\varepsilon \\ &\le 2\|\Theta\|_{L^\infty}\|\varphi\|_{L^\infty}\mathbb{P}(A_\varepsilon) + \|\Theta\|_{L^\infty}\varepsilon \\ &\le 4\|\Theta\|_{L^\infty}\|\varphi\|_{L^\infty}\exp\left(-\frac{N\varepsilon^2}{2\|\varphi\|_{L^\infty}^2}\right) + \|\Theta\|_{L^\infty}\varepsilon\end{aligned}$$

Choosing $\varepsilon = N^{-1/3}$, we get that $\Delta_{N,\Theta} \xrightarrow[N\to\infty]{} 0$. We conclude that $\boldsymbol{\gamma} = \mathbb{E}\boldsymbol{\gamma}$. □

Remark 7.15. *In fact we have show that $\boldsymbol{\gamma}_N$ has a subsequence converging to γ in the much stronger topology of narrow converge $\mathbb{P}$ almost surely.*

The previous Lemma is the crucial observation that allows the passage of the limit of a sequence of stochastic variational problems to a deterministic one as we shall see in the following Γ-convergence result.

Theorem 7.16. *Given an i.i.d. sample $(X_i)_{i\in\mathbb{N}}$ with law μ, let $\boldsymbol{\mu}_N \in \mathscr{P}_\Omega(\mathcal{X})$ the associated sequence of empirical random measures. Let $\mathcal{J}_{\Omega,N}$ be the sequence of potential functionals defined in (7.29), then it holds thatw*

$$\mathcal{J}_{\Omega,N} \xrightarrow[N\to\infty]{\Gamma} \mathcal{J}_\Omega(\boldsymbol{\gamma}) \overset{\text{def.}}{=} \begin{cases}\mathcal{J}(\gamma), & \textit{if } \boldsymbol{\gamma} = \gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y}),\\ +\infty, & \textit{otherwise},\end{cases}$$

where the Γ-convergence is in $\mathscr{P}_\Omega(\mathcal{X}\times\mathcal{Y})$ equipped with the narrow topology of random probability measures.

Proof. Starting with $\Gamma - \liminf$, consider a sequence $(\boldsymbol{\gamma}_N)_{N\in\mathbb{N}}$ converging to $\boldsymbol{\gamma}$ in the narrow topology of random measures. From Lemma 7.13, it follows that $\boldsymbol{\gamma}$ is actually a non-random measure $\gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$. Without loss of generality we assume that

$$\liminf_{N\to\infty}\mathcal{J}_{\Omega,N}(\boldsymbol{\gamma}_N) < \infty,$$

otherwise there is nothing to prove. Then, up to taking a subsequence attaining the $\liminf$, one can assume that $\mathcal{J}_{\Omega,N}(\boldsymbol{\gamma}_N) \le C$ for all $N \in \mathbb{N}$, so in particular $\boldsymbol{\gamma}_N \in \mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X}\times\mathcal{Y})$.

For an arbitrary $M > 0$, define the truncated interaction energy as

$$\mathcal{H}^M(\nu,\nu) \overset{\text{def.}}{=} \int H^M \mathrm{d}\nu\otimes\nu, \text{ where } H^M \overset{\text{def.}}{=} H \wedge M$$

and the truncated total energies $\mathcal{J}^M$ and $\mathcal{J}^M_{\Omega,N}$ as in (7.18) and (7.29) by replacing $\mathcal{H}$ with $\mathcal{H}^M$. Then it follows from Fubini's Theorem that

$$\begin{aligned}\mathcal{J}^M_{\Omega,N}(\boldsymbol{\gamma}_N) &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \int_{\mathcal{Y}} c(X_i,y)+L(y)\mathrm{d}\nu_N^{X_i} + \frac{1}{N^2}\sum_{i\neq j}\int_{\mathcal{Y}\times\mathcal{Y}} \mathcal{H}^M(\nu_N^{X_i},\nu_N^{X_j})\right] \\ &= \mathbb{E}\left[\int_{\mathcal{X}\times\mathcal{Y}} c+L\mathrm{d}\boldsymbol{\gamma}_N + \int_{\mathcal{Y}\times\mathcal{Y}} H^M \mathrm{d}\boldsymbol{\gamma}_N\otimes\boldsymbol{\gamma}_N\right] - \frac{1}{N^2}\sum_{i=1}^N \mathbb{E}\left[\mathcal{H}^M\left(\nu_N^{X_i},\nu_N^{X_j}\right)\right] \\ &= \mathcal{J}^M(\mathbb{E}\boldsymbol{\gamma}_N) - \frac{1}{N^2}\sum_{i=1}^N \underbrace{\mathbb{E}\left[\mathcal{H}^M(\nu_N^{X_i},\nu_N^{X_i})\right]}_{\leq M} \geq \mathcal{J}^M(\mathbb{E}\boldsymbol{\gamma}_N) - \frac{M}{N},\end{aligned}$$

where the last inequality was obtained from the fact that $\mathcal{H}^M$ is bounded by M. For any fixed $M>0$, the sum on the right-hand side above vanishes as $N\to\infty$ and hence since $\mathbb{E}\boldsymbol{\gamma}_N \xrightharpoonup[N\to\infty]{} \gamma$, the lower semi-continuity of $\mathcal{J}$ gives that

$$\liminf_{N\to\infty} \mathcal{J}_{\Omega,N}(\boldsymbol{\gamma}_N) \geq \liminf_{N\to\infty} \mathcal{J}^M(\mathbb{E}\boldsymbol{\gamma}_N) \geq \mathcal{J}^M(\gamma).$$

Noticing that from the monotone convergence theorem $\mathcal{H}(\nu,\nu) = \sup_{M>0} \mathcal{H}^M(\nu,\nu)$, we get

$$\liminf_{N\to\infty} \mathcal{J}_{\Omega,N}(\boldsymbol{\gamma}_N) \geq \sup_{M>0} \mathcal{J}^M(\gamma) = \mathcal{J}(\gamma),$$

and the result follows.

To prove the Γ-limsup it suffices to construct recovery sequences only for non-random transportation plans $\gamma\in\mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$. For any such measure, consider its disintegration representation as $\gamma=\mu\otimes\nu^x$, and define a recovery sequence as

$$\boldsymbol{\gamma}_N \overset{\text{def.}}{=} \boldsymbol{\mu}_N\otimes\nu^x,$$

where $(\boldsymbol{\mu}_N)_{N\in\mathbb{N}}$ is the family of empirical random measures built from the i.i.d. sample $(X_i)_{i\in\mathbb{N}}$ of law μ. Let us show that $\boldsymbol{\gamma}_N \xrightharpoonup[N\to\infty]{} \gamma$. We know from Lemma 7.13 that for any cluster point $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}_N$ it holds that $\tilde{\boldsymbol{\gamma}}$ is a deterministic measure, so for any convergent subsequence we have

$$\boldsymbol{\gamma}_{N_k} \xrightharpoonup[k\to\infty]{} \tilde{\boldsymbol{\gamma}} = \lim_{N\to\infty} \mathbb{E}\boldsymbol{\gamma}_N = \gamma,$$

so that the whole sequence must converge to γ.

Next, for each $N \in \mathbb{N}$, a simple computation yields

$$
\begin{aligned}
\mathcal{J}_{\Omega,N}(\boldsymbol{\gamma}_N) =&\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\int_{\mathcal{Y}} c(X_i,y)\mathrm{d}\nu^{X_i}(y) + \int_{\mathcal{Y}} L(y)\mathrm{d}\nu^{X_i}(y)\right) + \frac{1}{N^2}\sum_{i\neq j}\int H\mathrm{d}\nu^{X_i}\otimes\nu^{X_j}\right]\\
=&\frac{1}{N}\sum_{i=1}^{N}\int_{\mathcal{X}}\left(\int_{\mathcal{Y}} c(x_i,y)\mathrm{d}\nu^{x_i}(y) + \int_{\mathcal{Y}} L(y)\mathrm{d}\nu^{x_i}(y)\right)\mathrm{d}\mu(x_i)\\
&+\frac{1}{N^2}\sum_{i\neq j}\int_{\mathcal{X}\times\mathcal{X}}\left(\int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu^{x_i}\otimes\nu^{x_j}\right)\mathrm{d}\mu\otimes\mu(x_i,x_j)\\
=&\frac{1}{N}\sum_{i=1}^{N}\int_{\mathcal{X}\times\mathcal{Y}}(c+L)\mathrm{d}\gamma + \frac{1}{N^2}\sum_{i\neq j}\int H\mathrm{d}\nu\otimes\nu\\
=&\mathcal{J}(\gamma) - \frac{1}{N}\int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu\otimes\nu \le \mathcal{J}(\gamma).
\end{aligned}
$$

Taking the $\limsup$ as $N \to \infty$, the result follows. □

Now we use the properties of Γ convergence along Prokhorov's compactness Theorem for random measures,Thm. 1.12 in Chap. 1, to show that cluster points of equilibria for the N-players game are Cournot-Nash equilibria in the sense of Definition 7.1.

Theorem 7.17. *Assume that $\inf_{\mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})}\mathcal{J} < \infty$, then if $(\boldsymbol{\gamma}_N)_{N\in\mathbb{N}}$ is a sequence of minimizers of $\mathcal{J}_{\Omega,N}$, then there exists a subsequence such that*

$$\boldsymbol{\gamma}_{N_k} \xrightarrow[N_k\to\infty]{} \gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y}),$$

in the narrow topology of $\mathscr{P}_\Omega(\mathcal{X}\times\mathcal{Y})$, and in addition γ is a Cournot-Nash equilibrium in the sense of Definition 7.1.

Assuming in addition that $H \in C_b(\mathcal{Y}\times\mathcal{Y})$, for any sequence of Nash equilibria $(\boldsymbol{\gamma}_N)_{N\in\mathbb{N}}$ from game (7.24), that is

$$\boldsymbol{\gamma}_N \overset{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^{N}\delta_{X_i}\otimes\nu^x_{i,N} \in \mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X}\times\mathcal{Y}), \tag{7.34}$$

converging to γ in the narrow topology of $\mathscr{P}_\Omega(\mathcal{X}\times\mathcal{Y})$, it holds that $\gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$, and it is a Cournot-Nash equilibrium in the sense of Definition 7.1.

Proof. To prove the first assertion, we know from the properties of Γ-convergence that

$$\inf_{\mathscr{P}_{\Omega,\boldsymbol{\mu}_N}(\mathcal{X}\times\mathcal{Y})}\mathcal{J}_{\Omega,N} \xrightarrow[N\to\infty]{} \inf_{\mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})}\mathcal{J} \overset{\text{def.}}{=} C < +\infty. \tag{7.35}$$

Hence, since the functionals $\mathcal{J}_{\Omega,N}$ are l.s.c. with compact level sets, for each $N \in \mathbb{N}$ it admits a minimizer $\boldsymbol{\gamma}_N$. So if this sequence has a cluster point, then it must also minimize

$\mathcal{J}$, from Theorem 7.16. Hence, to finish the proof, it suffices to obtain such cluster point. This will be done with the version of Prokhorov's Theorem for random measures, see Theorem 1.12, which states that a sequence of random measures is sequentially compact in the narrow topology if and only if it is tight.

As $\boldsymbol{\mu}_N \xrightarrow[N\to\infty]{} \mu$ in the narrow topology of random measures it is a tight family, from the Prokhorov's Theorem, so for any $\varepsilon > 0$ there exists a compact set $K_{\mathcal{X},\varepsilon} \subset \mathcal{X}$ such that

$$\mathbb{E}\left[\boldsymbol{\mu}_N(\mathcal{X} \setminus K_{\mathcal{X},\varepsilon})\right] < \frac{\varepsilon}{2}.$$

From (7.35) we get that, for N large enough,

$$\mathbb{E}\left[\int_{\mathcal{Y}} L \mathrm{d}\boldsymbol{\nu}_N\right] \leq 2C,$$

so that, for some $\varepsilon > 0$ we obtain from Markov's inequality that

$$\mathbb{E}\left[\boldsymbol{\nu}_N\left(\left\{L \leq \frac{4C}{\varepsilon}\right\}\right)\right] \leq \frac{\mathbb{E}\left[\int_{\mathcal{Y}} L \mathrm{d}\boldsymbol{\nu}_N\right]}{2C/\varepsilon} \leq \frac{\varepsilon}{2}.$$

Since L has compact level sets, we set $K_{\mathcal{Y},\varepsilon} = \{L \leq 4C/\varepsilon\}$ and set $K_\varepsilon \overset{\text{def.}}{=} K_{\mathcal{X},\varepsilon} \times K_{\mathcal{Y},\varepsilon}$, so that

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\gamma}_N\left(\mathcal{X} \times \mathcal{Y} \setminus K_\varepsilon\right)\right] &\leq \mathbb{E}\left[\boldsymbol{\gamma}_N\left((\mathcal{X} \setminus K_{\mathcal{X},\varepsilon}) \times \mathcal{Y}\right)\right] + \mathbb{E}\left[\boldsymbol{\gamma}_N\left(\mathcal{X} \times (\mathcal{Y} \setminus K_{\mathcal{Y},\varepsilon})\right)\right] \\ &= \mathbb{E}\left[\boldsymbol{\mu}_N\left((\mathcal{X} \setminus K_{\mathcal{X},\varepsilon}) \times \mathcal{Y}\right)\right] + \mathbb{E}\left[\boldsymbol{\nu}_N\left(\mathcal{X} \times (\mathcal{Y} \setminus K_{\mathcal{Y},\varepsilon})\right)\right] < \varepsilon.\end{aligned}$$

We conclude that the sequence of random measures $\boldsymbol{\gamma}_N$ is tight, and hence admits a convergent subsequence in the narrow topology of $\mathscr{P}_\Omega(\mathcal{X} \times \mathcal{Y})$. As discussed above, from Lemma 7.13 the limit of this subsequence belongs in $\mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$ and minimizes $\mathcal{J}$. From the variational characterization of equilibria given in Theorem 7.4, γ is a Cournot-Nash equilibrium in the sense of Definition 7.1.

To prove the second assertion, let $\boldsymbol{\gamma}_N$ be defined as in (7.34) and γ a limit point. From Lemma 7.13, $\gamma \in \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$, our goal is to verify that γ is a critical point of $\mathcal{J}$, *i.e.* for any $\bar{\gamma} \in \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$ we verify that

$$\left\langle \frac{\delta \mathcal{J}}{\delta \gamma}, \bar{\gamma} - \gamma \right\rangle = \int_{\mathcal{X}\times\mathcal{Y}} \left(c(x,y) + L(y) + 2\int_{\mathcal{Y}} H(y,\bar{y}) \mathrm{d}\nu(\bar{y})\right) \mathrm{d}(\bar{\gamma} - \gamma)(x,y) \geq 0,$$

where $\nu = (\pi_{\mathcal{Y}})_\sharp \gamma$. From Thm. 7.4, this will show that γ is a Cournot-Nash equilibrium.

Fix some $\bar{\gamma} \in \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$, and recall the recovery sequence obtained from the Γ-convergence proof; consider a disintegration family $\bar{\gamma} = \mu \otimes \bar{\nu}^x$ so that

$$\bar{\boldsymbol{\gamma}}_N \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^N \bar{\nu}^{X_i} \xrightarrow[N\to\infty]{} \gamma.$$

We consider a unilateral deviation of player i with the alternative strategy $\bar{\nu}^{X_i}$, to the profile $\left(\nu_{1,N}^{X_1}, \ldots, \nu_{N,N}^{X_N}\right)$. Since the latter is a Nash equilibrium in mixed strategies, we get that $\mathcal{J}_{\Omega,i}(\bar{\nu}^{X_i}, \nu_{-i,N}^{X_{-i}}) \geq \mathcal{J}_{\Omega,i}(\nu_{i,N}^{X_i}, \nu_{-i,N}^{X_{-i}})$, for $\mathcal{J}_{\Omega,i}$ defined in (7.24). This can be rewritten as

$$\begin{aligned}\mathbb{E}&\left[\int_{\mathcal{Y}} c(X_i, y) + L(y)\mathrm{d}\bar{\nu}^{X_i} + \frac{2}{N}\sum_{j\neq i}\int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\bar{\nu}^{X_i}\otimes\nu_{j,N}^{X_j}\right] \\ &\geq \mathbb{E}\left[\int_{\mathcal{Y}} c(X_i, y) + L(y)\mathrm{d}\nu_{i,N}^{X_i} + \frac{2}{N}\sum_{j\neq i}\int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu_{i,N}^{X_i}\otimes\nu_{j,N}^{X_j}\right].\end{aligned}$$

Let us define the measures

$$\boldsymbol{\gamma}_{N,-i} \overset{\text{def.}}{=} \frac{1}{N}\sum_{j\neq i}\delta_{X_j}\otimes\nu_{j,N}^{X_j}, \text{ and } \boldsymbol{\nu}_{N,-i} \overset{\text{def.}}{=} (\pi_{\mathcal{Y}})_\sharp\boldsymbol{\gamma}_{N,-i},$$

so that evaluating the expectations, using the definition of the expectation measure we obtain

$$\int_{\mathcal{X}\times\mathcal{Y}}(c+L)\mathrm{d}\bar{\gamma}+2\int H\mathrm{d}\bar{\gamma}\otimes\mathbb{E}\boldsymbol{\gamma}_{N,-i} \geq \mathbb{E}\left[\int_{\mathcal{Y}} c(X_i,y)+L(y)\mathrm{d}\nu_{i,N}^{X_i}\right]+2\int H\mathrm{d}\mathbb{E}[\nu_{i,N}^{X_i}]\otimes\mathbb{E}\boldsymbol{\gamma}_{N,-i}.$$

Rewriting $\boldsymbol{\gamma}_{N,-i} = \boldsymbol{\gamma}_N - \frac{1}{N}\delta_{X_i}\otimes\boldsymbol{\nu}_{i,N}^{X_i}$ and averaging over all i, we get that

$$\begin{aligned}&\int_{\mathcal{X}\times\mathcal{Y}}(c+L)\mathrm{d}\bar{\gamma} + 2\left(1-\frac{1}{N}\right)\int H\mathrm{d}\bar{\gamma}\otimes\mathbb{E}\boldsymbol{\gamma}_N \\ &\qquad \geq \int_{\mathcal{X}\times\mathcal{Y}}(c+L)\mathrm{d}\mathbb{E}\boldsymbol{\gamma}_N + 2\int H\mathrm{d}\mathbb{E}\boldsymbol{\gamma}_N\otimes\mathbb{E}\boldsymbol{\gamma}_N - \frac{2}{N^2}\sum_{i=1}^N\int H\mathrm{d}\mathbb{E}\boldsymbol{\nu}_{i,N}^{X_i}\otimes\mathbb{E}\boldsymbol{\nu}_{i,N}^{X_i}\end{aligned}$$

As $H \in C_b$, the last term is a $O(1/N)$ and hence vanishes as $N \to \infty$. In addition, since $\mathbb{E}\boldsymbol{\gamma}_N \xrightarrow[N\to\infty]{} \gamma$, from the convergence of $\boldsymbol{\gamma}_N$ and Lemma 7.13, we get that

$$\begin{aligned}0 &\geq \liminf_{N\to\infty}\int_{\mathcal{X}\times\mathcal{Y}} c + L\mathrm{d}(\mathbb{E}\boldsymbol{\gamma}_N - \bar{\gamma}) + \frac{2(N-1)}{N}\int H\mathrm{d}\mathbb{E}\boldsymbol{\gamma}_N\otimes(\mathbb{E}\boldsymbol{\gamma}_N - \bar{\gamma}) \\ &\geq \liminf_{N\to\infty}\int_{\mathcal{X}\times\mathcal{Y}} c + L\mathrm{d}(\gamma - \bar{\gamma}) + 2\int H\mathrm{d}\gamma\otimes(\gamma-\bar{\gamma}) = \left\langle\frac{\delta\mathcal{J}}{\delta\gamma}, \gamma - \bar{\gamma}\right\rangle.\end{aligned}$$

From Thm. 7.4, γ is a Cournot-Nash equilibrium. □

4.2. Γ-convergence for the closed loop formulation

Now we move on to the question of the convergence of a sequence of Nash equilibria for the games in closed loop (7.23). In this case we have a family of games indexed by the sample $\omega = (x_i)_{i\in\mathbb{N}}$ of the players' state variables, therefore we can only expect a Γ convergence to hold with $\mathbb{P}$-probability 1. We start by showing a general Lemma that gives Γ-convergence with full probability.

Lemma 7.18. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family of functionals $(\mathscr{F}_{\omega,N})_{\substack{N\in\mathbb{N}\\ \omega\in\Omega}}$ and a functional $\mathscr{F}$ over a Polish space $\mathcal{X}$ such that*

1. *there is a set Ω_0 with full $\mathbb{P}$-probability such that for any $x_N \to x$ the $\Gamma - \liminf$ inequality for $\mathscr{F}_{\omega,N}$ holds*

$$\mathscr{F}(x) \leq \liminf_{N\to\infty} \mathscr{F}_{\omega,N}(x_N), \text{ for all } \omega \in \Omega_0$$

2. *for each $x \in \mathcal{X}$ there is a set Ω_x with full $\mathbb{P}$-probability for which we can construct recovery sequences of $\mathscr{F}_{\omega,N}$*

$$\limsup_{N\to\infty} \mathscr{F}_{\omega,N}(x_N) \leq \mathscr{F}(x), \text{ for all } \omega \in \Omega_x.$$

Under these conditions, there is a set $\bar{\Omega}_0$ with full $\mathbb{P}$-probability such that for any $\omega \in \bar{\Omega}_0$ the sequence $\mathscr{F}_{\omega,N}$ Γ-converges to $\mathscr{F}$.

Proof. First we claim that there exists a countable and dense set $\mathscr{D} \subset \mathcal{X}$ which is dense in the energy $\mathscr{F}$, i.e. for each $x \in \mathcal{X}$ there is $(x_n)_{n\in\mathbb{N}} \subset \mathscr{D}$ such that

$$x_n \xrightarrow[n\to\infty]{} x \text{ and } \mathscr{F}(x_n) \xrightarrow[n\to\infty]{} \mathscr{F}(x). \tag{7.36}$$

See for instance [Ambrosio et al., 2021, Lemma 11.12] for a constructive argument, a simple proof comes from the fact that $\mathbb{R} \times \operatorname{dom} \mathscr{F}$ is separable as an (arbitrary) subset of the separable space $\mathbb{R} \times \mathcal{X}$, since subsets of second countable spaces are second countable.

Hence we can define the set $\bar{\Omega}_0$ as

$$\bar{\Omega}_0 \overset{\text{def.}}{=} \Omega_0 \cap \bigcap_{x\in\mathscr{D}} \Omega_x,$$

where Ω_0 denotes the set where the Γ-$\liminf$ holds for all points $x \in \mathcal{X}$ and Ω_x denotes the event in which we can construct recovery sequences for x. Since $\mathscr{D}$ is countable, it holds that $\mathbb{P}(\bar{\Omega}_0) = 1$.

To prove the Γ-convergence for each $\omega \in \bar{\Omega}_0$, we recall the notions of lower and upper Γ limits from Section 1.2 from Chapter 1, and to conclude it suffices to prove for all $\omega \in \bar{\Omega}_0$ that $\Gamma\text{-}\liminf \mathscr{F}_{\omega,N} = \Gamma\text{-}\limsup \mathscr{F}_{\omega,N} = \mathscr{F}$. Indeed, item (1) shows that

$$\mathscr{F} \leq \Gamma\text{-}\liminf \mathscr{F}_{\omega,N}, \text{ for all } \omega \in \bar{\Omega}_0.$$

On the other hand, from item (2), it follows for any $x \in \mathscr{D}$ that

$$\Gamma\text{-}\limsup \mathscr{F}_{\omega,N}(x) \leq \mathscr{F}(x), \text{ for all } \omega \in \bar{\Omega}_0.$$

Hence, for any $\omega \in \bar{\Omega}_0$ and an arbitrarily $x \in \mathcal{X}$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathscr{D}$ satisfying (7.36), so that using the lower semi-continuity of the Γ upper limit we have that

$$\Gamma\text{-}\limsup_{N\to\infty} \mathscr{F}_{\omega,N}(x) \leq \liminf_{n\to\infty} \left(\Gamma\text{-}\limsup_{N\to\infty} \mathscr{F}_{\omega,N}(x_n) \right) \leq \liminf_{n\to\infty} \mathscr{F}(x_n) = \mathscr{F}(x)$$

which gives the Γ-convergence with full probability. □

To apply this Lemma, we know from the Glivenko-Cantelli law of large numbers that empirical measures converge $\mathbb{P}$ almost surely. Hence, we consider the set

$$\Omega_0 \overset{\text{def.}}{=} \left\{ \omega = (x_i)_{i\in\mathbb{N}} \in \operatorname{supp}\mathbb{P} : \begin{array}{c} x_i \neq x_j, \text{ for } i \neq j \\ \mu_N(\omega) \xrightarrow[N\to\infty]{} \mu \end{array} \right\}. \tag{7.37}$$

The first condition above is so that the sequence of functionals $\mathcal{J}_{\omega,N}$ is well-defined for any $\omega \in \Omega_0$. From the fact that μ is atomless and the above discussion, $\mathbb{P}(\Omega_0) = 1$.

While the Γ-liminf argument will be similar to the open loop information structure, for the Γ-limsup we will use a construction depending on a sequence of random variations of the form

$$\begin{gathered} \frac{1}{N}\sum_{i=1}^{N} L_i + \frac{1}{N^2}\sum_{i\neq j} H_{i,j} \\ \text{where } L_i \overset{\text{def.}}{=} c(X_i, Y_i) + L(Y_i), \quad H_{i,j} \overset{\text{def.}}{=} H(Y_i, Y_j), \end{gathered} \tag{7.38}$$

where $(X_i, Y_i) \sim \gamma$. The first sum is fortunately an i.i.d. sequence, so that from the law of large numbers it must converge to its mean. The second term however is not i.i.d., but it is *exchangeable* as it can be written as a symmetric function of an i.i.d. sample. In the following Proposition, whose proof is a synthesis of the ideas from [Klenke, 2013, Chap. 12], we show that such families of random variables also enjoy a law of large numbers.

Proposition 7.19. *Let $(\bar{H}_n)_{n\in\mathbb{N}}$ be a sequence of random variables obtained as the symmetric image of an i.i.d. sample, that is let it be the enumeration of the family of random variables*

$$(\Phi(X_i, X_j))_{i\neq j\in\mathbb{N}},$$

where $\Phi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a symmetric function and $(X_i)_{i\in\mathbb{N}}$ is an i.i.d. sample. Then

$$\frac{1}{N}\sum_{n=1}^{N} \bar{H}_n \xrightarrow[N\to\infty]{} \mathbb{E}[\bar{H}_1], \text{ with probability } 1.$$

For the sake of readability of the main ideas employed to prove the Γ-convergence result, we include the proof of the previous proposition in Appendix A.

Theorem 7.20. *With full $\mathbb{P}$-probability, the sequence of functionals $\mathcal{J}_{\omega,N}$ convergence to $\mathcal{J}$ in the sense of Γ convergence in the narrow topology of $\mathscr{P}(\mathcal{X} \times \mathcal{Y})$.*

Proof. It suffices to verify the hypothesis of Lemma 7.18. To prove (1), consider $\omega \in \Omega_0$ defined above in (7.37), and let $(\gamma_N)_{N\in\mathbb{N}}$ be a sequence such that $\gamma_N \in \mathscr{P}_{\mu_N}(\mathcal{X} \times \mathcal{Y})$ and converging to γ. So we can assume that γ_N can be written as $\gamma_N = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \delta_{x_i} \otimes \nu^{x_i}$, where $\nu^{x_i} \in \mathscr{P}(\mathcal{Y})$ and for any $\omega \in \Omega_0$, it follows from the continuity w.r.t. convergence of marginals that $\gamma \in \mathscr{P}_\mu(\mathcal{X} \times \mathcal{Y})$.

For an arbitrary $M > 0$, define $\mathcal{H}^M(\nu,\nu) \overset{\text{def.}}{=} \int H \wedge M \mathrm{d}\nu \otimes \nu$, and it follows that

$$\begin{aligned}
\mathcal{J}_{\omega,N}(\gamma_N) &\geq \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\gamma_N + \underbrace{\frac{1}{N}\sum_{i=1}^N \mathcal{L}(\nu^{x_i})}_{=\mathcal{L}(\nu_N)} + \underbrace{\frac{1}{N^2}\sum_{i,j}\mathcal{H}^M(\nu^{x_i},\nu^{x_j})}_{=\mathcal{H}^M(\nu_N,\nu_N)} - \frac{1}{N^2}\sum_{i=1}^N \mathcal{H}^M(\nu^{x_i},\nu^{x_i}) \\
&= \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\gamma_N + \mathcal{L}(\nu_N) + \mathcal{H}^M(\nu_N,\nu_N) - \frac{1}{N^2}\sum_{i=1}^N \mathcal{H}^M(\nu^{x_i},\nu^{x_i}) \\
&\geq \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\gamma_N + \mathcal{L}(\nu_N) + \mathcal{H}^M(\nu_N,\nu_N) - \frac{M}{N}.
\end{aligned}$$

The sum on the RHS vanishes as $N \to \infty$ for each $M > 0$ and hence the lower semicontinuity of the remaining terms w.r.t. narrow convergence, as integrals of l.s.c. integrands, for every $M > 0$ gives

$$\begin{aligned}
\liminf_{N\to\infty} \mathcal{J}_{\omega,N}(\gamma_N) &\geq \liminf_{N\to\infty} \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\gamma_N + \mathcal{L}(\nu_N) + \mathcal{H}^M(\nu_N,\nu_N) \\
&\geq \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\gamma + \mathcal{L}(\nu) + \mathcal{H}^M(\nu,\nu).
\end{aligned}$$

Noticing that from the monotone convergence theorem $\mathcal{H}(\nu,\nu) = \sup_{M>0} \mathcal{H}^M(\nu,\nu)$, the Γ-$\liminf$ follows.

To verify property (2) from Lemma 7.18, given some $\gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$, let $\nu = (\pi_\mathcal{Y})_\sharp\gamma$. By an application of the disintegration theorem one can write $\gamma = \nu^x \otimes \mu$ for some Borel map $(\nu^x)_{x\in\mathcal{X}}$, *i.e.*

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi(x,y)\mathrm{d}\gamma = \int_\mathcal{X}\left(\int_\mathcal{Y}\varphi(x,y)\mathrm{d}\nu^x(y)\right)\mathrm{d}\mu(x), \text{ for all } \varphi \in C_b(\mathcal{X}\times\mathcal{Y}).$$

This disintegration family is only μ-a.e.uniquely defined, but we can fix one such family and define a new transportation plan as $\gamma_N \overset{\text{def.}}{=} \mu_N \otimes \nu^x$. Since we have fixed one disintegration family, $\gamma_N \in \mathscr{P}_{\mu_N}(\mathcal{X}\times\mathcal{Y})$ is well-defined for every event $\omega = (x_i)_{i\in\mathbb{N}}$. From the definition, it then holds that

$$\int_{\mathcal{X}\times\mathcal{Y}} \phi(x,y)\mathrm{d}\gamma_N \overset{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^N \int \phi(x_i,y)\mathrm{d}\nu^{x_i}, \text{ for all } \phi \in C_b(\mathcal{X}\times\mathcal{Y}).$$

Hence $\gamma_N \in \Pi(\mu_N,\nu_N)$ where $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, $\nu_N = \frac{1}{N}\sum_{i=1}^N \nu^{x_i}$.

Let us prove that γ_N converges narrowly to γ; indeed from Prop. 1.6 we know there is a countable set $\mathcal{K} \subset C_b(X)$, such that to prove narrow convergence it suffices to verify that

$$\int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\mathrm{d}\gamma_N \xrightarrow[N\to\infty]{} \int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\mathrm{d}\gamma, \text{ for all } f \in \mathcal{K}.$$

For each $f \in \mathcal{K}$, we compute

$$\int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\mathrm{d}\gamma_N = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathcal{X}\times\mathcal{Y}} f(x_i,y)d\nu^{x_i}(y).$$

Hence, each term of the sum on the right is a realization of the i.i.d. sequence of random variables $F_i \overset{\text{def.}}{=} \int_{\mathcal{Y}} f(X_i,\cdot)\mathrm{d}\nu^{X_i}$. From the strong law of large numbers, it holds with probability 1 that

$$\int_{\mathcal{X}\times\mathcal{Y}} f\mathrm{d}\gamma_N = \frac{1}{N}\sum_{i=1}^{N} F_i(\omega) \xrightarrow[N\to\infty]{} \mathbb{E}_{\mathbb{P}}\left[F_1\right] = \int_X f\mathrm{d}\gamma.$$

Let $\Omega_{\gamma,f}$ denote the set of probability 1, which depends on γ and f, where the above converge holds. Then defining

$$\tilde{\Omega}_\gamma = \bigcap_{f\in\mathcal{K}} \Omega_{\gamma,f},$$

we have that $\mathbb{P}(\tilde{\Omega}_\gamma) = 1$ and for any $\omega \in \tilde{\Omega}_\gamma$ it holds that $\gamma_N \xrightarrow[N\to\infty]{} \gamma$.

We now apply a similar argument to the convergence of the energies. Indeed, writing

$$\mathcal{J}_{\omega,N}(\gamma_N) = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathcal{Y}} \left(c(x_i,y)+L(y)\right)\mathrm{d}\nu^{x_i}(y) + \frac{1}{N^2}\sum_{j\neq i}\int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu^{x_i}\otimes\nu^{x_j}.$$

We see that the first sum is the empirical average of the i.i.d. sequence of random variables $L_i \overset{\text{def.}}{=} \int_{\mathcal{Y}} (c(X_i,y)+L(y))\mathrm{d}\nu^{X_i}(y)$ while the double sum can be written in terms of the sequence $H_{i,j} \overset{\text{def.}}{=} \int H\mathrm{d}\nu^{X_i}\otimes\nu^{X_j}$. As a consequence, applying once again the strong law of large numbers, there is a set $\Omega_{L,\gamma}$ with probability 1, such that for any $\omega \in \Omega_{L,\gamma}$ it holds that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} L_i(\omega) \xrightarrow[N\to\infty]{} \mathbb{E}_{\mathbb{P}}[L_1] &= \int_{\mathcal{X}}\left[\int_{\mathcal{Y}}(c(x,y)+L(y))\mathrm{d}\nu^x(y)\right]\mathrm{d}\mu(x)\\
&= \int_{\mathcal{X}\times\mathcal{Y}}(c(x,y)+L(y))\mathrm{d}\gamma = \int_{\mathcal{X}\times\mathcal{Y}} c\mathrm{d}\gamma + \mathcal{L}(\nu).
\end{aligned}$$

For the second term, the random variables $(H_{i,j})_{i\neq j}$ are no longer i.i.d., but satisfy the hypothesis of Thm. 7.19 with Φ given by

$$\Phi(x_1,x_2) \overset{\text{def.}}{=} \int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu^{x_1}\otimes\nu^{x_2},$$

which is symmetric and measurable from the measurability of the family $(\nu^x)_{x\in\mathcal{X}}$. We conclude that there is another set $\Omega_{H,\gamma}$ with probability 1 such that for all $\omega \in \Omega_{H,\gamma}$ it holds that

$$\frac{1}{N^2}\sum_{j\neq i}\int_{\mathcal{Y}\times\mathcal{Y}} H\mathrm{d}\nu^{x_i}\otimes\nu^{x_j} \xrightarrow[N\to\infty]{} \mathbb{E}_{\mathbb{P}}[H_{1,2}] = \mathcal{H}(\nu,\nu).$$

Finally, the set $\Omega_\gamma \overset{\text{def.}}{=} \tilde{\Omega}_\gamma \cap \Omega_{L,\gamma} \cap \Omega_{H,\gamma}$ has probability 1 and satisfies all the properties of item (2).

From the thesis of Lemma 7.18, the Γ convergence with full $\mathbb{P}$-probability follows $\square$

As in the closed loop case, with an analogous proof to the open loop case, we obtain a result assuring, with full $\mathbb{P}$-probability, the convergence of a particular sequence of Nash equilibria to equilibria of Cournot-Nash type, and whenever H is continuous the convergence of any sequence of Nash equilibria.

Theorem 7.21. *Assume that* $\inf_{\mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})} \mathcal{J} < \infty$, *then there are sequences of Nash equilibria for the game (7.23), described by transportation plans* $(\gamma_N)_{N\in\mathbb{N}}$ *such that, with full* $\mathbb{P}$*-probability, converge up to a subsequence in the narrow topology to a Cournot-Nash equilibrium* $\gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$, *in the sense of Definition 7.1.*

Assuming in addition that $H \in C_b(\mathcal{Y}\times\mathcal{Y})$, *with* $\mathbb{P}$*-full probability, for any sequence of Nash equilibria* $(\gamma_N)_{N\in\mathbb{N}}$ *from game (7.23), that is*

$$\gamma_N \overset{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i} \otimes \nu_{i,N} \in \mathscr{P}_{\mu_N(\omega)}(\mathcal{X}\times\mathcal{Y}), \tag{7.39}$$

converging to γ *in the narrow topology of* $\mathscr{P}(\mathcal{X}\times\mathcal{Y})$, *it holds that* $\gamma \in \mathscr{P}_\mu(\mathcal{X}\times\mathcal{Y})$, *and it is a Cournot-Nash equilibrium in the sense of Definition 7.1.*

A. The law of large numbers for symmetric functions of an i.i.d. sample

In this appendix we prove Proposition 7.19. The ideas are a minor modification of the presentation of [Klenke, 2013], hence our goal is to make it as self-contained as possible to readers less familiarized with probability theory, but we hope it can be useful in other contexts as well. We also observe that this proof remains true if one considers $\Phi : \mathcal{X}^{\otimes k} \to \mathbb{R}$, for any $k \in \mathbb{N}$. With this we can now proceed with our Γ-convergence type result.

Proposition 7.22. *Let* $(H_{i,j})_{i\neq j\in\mathbb{N}}$ *be a sequence of random variables obtained as*

$$H_{i,j} = (\Phi(X_i, X_j))_{i\neq j\in\mathbb{N}},$$

where $\Phi : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ *is a symmetric function and* $(X_i)_{i\in\mathbb{N}}$ *is an i.i.d. sample. Then*

$$\frac{1}{N^2}\sum_{\substack{1\leq i,j\leq N\\ i\neq j}} H_{i,j} \xrightarrow[N\to\infty]{} \mathbb{E}[H_{1,2}], \text{ with probability } 1.$$

Proof of Prop. 7.19. First, define the exchangeable σ-algebra as follows: we say a function $f : \mathbb{R}^{\otimes \mathbb{N}} \to \mathbb{R}$ is n-symmetric if it is symmetric w.r.t. permutations of at most n indexes. In other words, for any permutation $\sigma : \mathbb{N} \to \mathbb{N}$ swapping at most n indexes, then $f\left(\left(x_{\sigma(n)}\right)_{n\in\mathbb{N}}\right) = f\left(\left(x_n\right)_{n\in\mathbb{N}}\right)$. We then define the exchangeable σ-algebra as

$$\mathcal{E}_\infty \overset{\text{def.}}{=} \bigcap_{n\in\mathbb{N}} \mathcal{E}_n, \text{ where } \mathcal{E}_n \overset{\text{def.}}{=} \sigma\left(\left\{ f\left(\left(X_i\right)_{i\in\mathbb{N}}\right) : \begin{array}{c} f : \mathbb{R}^{\otimes\mathbb{N}} \to \mathbb{R} \\ \text{is } n\text{-symmetric and Borel} \end{array} \right\}\right),$$

where $\sigma\left(\{F_i\}_{i\in I}\right)$ is defined as the smallest σ-algebra that makes the hole family of random variables $(F_i)_{i\in I}$ measurable.

Take $g : \mathbb{R}^{\otimes\mathbb{N}} \to \mathbb{R}$, bounded and n-symmetric function, for all $i \leq n$, it holds from exchangeability that

$$\begin{aligned}\mathbb{E}\left[H_{i,j} g\left(X.\right)\right] &= \mathbb{E}\left[\Phi(X_i, X_j) g\left(X_1, X_2, X_3, \dots, X_{i-1}, X_i, X_{i+1}, \dots, X_{j-1}, X_j, X_{j+1}, \dots\right)\right] \\ &= \mathbb{E}\left[\Phi(X_1, X_2) g\left(X_i, X_j, X_3, \dots, X_{i-1}, X_1, X_{i+1}, \dots, X_{j-1}, X_2, X_{j+1}, \dots\right)\right] \\ &= \mathbb{E}\left[H_{1,2} g\left(X.\right)\right]\end{aligned}$$

In particular, taking $g = 1_A$ for an arbitrary set $A \in \mathcal{E}_N$ and averaging the above equality for all $1 \leq i, j \leq N$ with $i \neq j$, we obtain that

$$\frac{1}{N(N-1)}\mathbb{E}\left[\sum_{\substack{1\leq i,j\leq N \\ i\neq j}} H_{i,j} 1_A\right] = \mathbb{E}\left[H_{1,2} 1_A\right], \text{ so } \frac{1}{N(N-1)} \sum_{\substack{1\leq i,j\leq N \\ i\neq j}} H_{i,j} = \mathbb{E}\left[H_{1,2} \middle| \mathcal{E}_N\right],$$

by the definition of conditional expectation for L^1 random variables. This means that $\frac{1}{N(N-1)} \sum_{\substack{1\leq i,j\leq N \\ i\neq j}} H_{i,j}$ is a backwards martingale for the filtration $\left(\mathcal{E}_N\right)_{N\in\mathbb{N}}$ and a suitable martingale convergence Theorem, [Klenke, 2013, Thm. 12.14], gives that

$$\frac{1}{N(N-1)} \sum_{\substack{1\leq i,j\leq N \\ i\neq j}} H_{i,j} \xrightarrow[n\to\infty]{} \mathbb{E}\left[H_{1,2} \mathcal{E}_\infty\right] \text{ with convergence a.s. and in } L^1.$$

Since $(X_i)_{i\in\mathbb{N}}$ is i.i.d., the Hewitt-Savage $0-1$ law, see [Klenke, 2013, Cor. 12.19] and [Hewitt and Savage, 1955], states that $\mathcal{E}_\infty$ is a trivial σ-algebra, so that for any set $A \in \mathcal{E}_\infty$, $\mathbb{P}(A)$ is either 0 or 1. Hence, as $\mathbb{E}\left[H_{1,2} \middle| \mathcal{E}_\infty\right]$ is an $\mathcal{E}_\infty$-adapted random variable, it must be given by a constant given by its mean $\mathbb{E}\left[\mathbb{E}\left[H_{1,2} \middle| \mathcal{E}_\infty\right]\right] = \mathbb{E}\left[H_{1,2}\right]$, and the result follows. $\square$

Chapter 8

Conclusion and Perspectives

In this thesis we have studied three variational problems, the Wasserstein-$\mathscr{H}^1$ problem, the JKO scheme of the total variational functional and potential Cournot-Nash equilibria. We have employed a variety of methods from the Calculus of Variations. In Chapter 3, we used the relaxation strategy described in the Introduction to prove an existence result for the newly introduced Wasserstein-$\mathscr{H}^1$ problem. Afterwards, this relaxation has proven itself useful to prove qualitative properties, as it is easier to generate variations for it. One can also argue that the representation of strategy profiles as transportation plans, in Chapter 7, is another instance of the relaxation strategy which allowed us to exploit the compactness of spaces of probability measures to obtain a convergence result to Cournot-Nash equilibria.

Another major tool throughout the thesis was the notion of Γ-convergence. In Chapter 5, we have provided a "classical" application, in the form of a diffuse approximation result that enables numerical simulations for the Wasserstein-$\mathscr{H}^1$ problem, but it was also employed in less standard ways. The arguments of existence and absence of loops for the Wasserstein-$\mathscr{H}^1$ consist of a contradiction of the fundamental property of Γ-convergence, by constructing a better competitor to a Γ-limiting functional. On the other hand, in Chapter 7, it is used to prove a seemingly unrelated result, about the convergence of Nash to Cournot-Nash equilibria. In particular, in the closed loop formulation the full Γ-convergence is not even proved will full probability, instead a single recovery sequence of a minimizer of the limit problem allows to conclude that any sequence of minimizers will also be a minimizer.

We shall now discuss some unexplored directions and future perspectives.

Wasserstein-$\mathscr{H}^1$ problem

Since this problem was first introduced in this thesis, it was first necessary to develop a satisfactory existence theory for it, which was more difficult than the proofs of existence for other 1D shape optimization problems in the literature, since we can not simply apply the direct method. We were nonetheless able to derive a rich qualitative description of minimizers and provide a phase-field approximation for it that allows for extensive numerical simulations in the future. Next we list some open questions left concerning this problem.

- The length functional has proven itself difficult to systematically generate variations, a clear direction is to understand what are the Euler-Lagrange equations for this problem. Once this is done, what kind of qualitative properties could we derive from it?
- The proof of absence of loops discussed in Chapter 4 is very flexible, can we adapt it to other cases?
- What are the topological properties/configuration of minimizers? Can we expect only triple points as in other 1D shape optimization problems, such as the average distance minimizers? Can we completely characterize the blow-up limit of every point in an optimal network?
- Concerning the phase-field approximations from Chap. 5, extensive numerical simulations are in order. In addition, it would also be interesting if this theory would allow for the numerical resolution of the average distance minimizers problem.

TV-JKO

The Wasserstein gradient flow of the total variation functional has already a quite mature literature discussed in Chap. 6, however the convergence of the scheme to the limit PDE is not yet completely understood.

- A question that arises is if the further regularity obtained with our methods can be useful to understand this convergence. Unfortunately, the Lipschitz constant obtained via our method explodes as $\tau \to 0$. It would be interesting if a variation of our approach that relates the TV-JKO with a suitable (ROF) problem could eliminate the dependence on τ.
- In addition, we have capitalized on the back-n-forth method to obtain a scalable way of computing the proximal operator of the squared Wasserstein distance. It would be interesting to evaluate if it could be used as an off-the-shelf routine to minimize other variational problems involving a Wasserstein term.

Potential Cournot-Nash equilibria

If anything, the convergence of Nash to Cournot-Nash equilibria discussed in Chapter 7 demonstrates how difficult the convergence question in the context of Mean Field Games is. The Γ-convergence approach relies entirely on the fact that a variational description of equilibria in provided in Thm. 7.4 and is not useful to other games, for which we have only fixed point techniques at disposal. The following questions then present themselves:

- Can we consider other types of energy? The analysis seems very specific to an energy that is the sum of an individual and a pair-wise interaction costs.

  - Using the characterization of convex functions as the envelope of all linear functions below it, one could try to adapt the arguments of the linear term to the case of an individual convex energy.
  - In principle, the arguments treating the pair-wise interaction term could be extended to a k-wise interaction, as long as the number of players interacting remains uniformly bounded, as in this case an analogous law of large numbers from Appendix A should hold.

- The convergence result in the closed loop case does not really require a Γ-convergence argument with full probability. It would be an interesting improvement in itself if this could be proved with the tools at disposal, but also imperative to adapt the arguments from Theorem 7.17 and show that with full probability any convergent sequence of Nash equilibria in the closed loop formulation converges to a Cournot-Nash equilibrium.

- As discussed in the conclusion of Chapter 7, another direction would be to derive a large deviations principle for the Gibbs measures associated with the potential function of the N-player games, whose Nash equilibria converge to Cournot-Nash equilibria, in accordance with the statistical mechanics intuition that motivated the original name Mean Field Games.

References

[Alberti et al., 2019] Alberti, G., Buttazzo, G., Bianco, S. G. L., and Oudet, É. (2019). Optimal reinforcing networks for elastic membranes. *arXiv preprint arXiv:1904.01905*.

[Alberti and Ottolini, 2017] Alberti, G. and Ottolini, M. (2017). On the structure of continua with finite length and gołąb's semicontinuity theorem. *Nonlinear Analysis: Theory, Methods & Applications*, 153:35–55.

[Aliprantis and Border, 2006] Aliprantis, C. D. and Border, K. C. (2006). *Infinite dimensional analysis*. Springer, Berlin, third edition. A hitchhiker's guide.

[Ambrosio, 1992] Ambrosio, L. (1992). On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital., B*, pages 105–123.

[Ambrosio, 2003] Ambrosio, L. (2003). *Lecture Notes on Optimal Transport Problems*, pages 1–52. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Ambrosio et al., 2021] Ambrosio, L., Brué, E., and Semola, D. (2021). Lectures on optimal transport.

[Ambrosio et al., 2000] Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, Oxford, New York.

[Ambrosio et al., 2008] Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition.

[Ambrosio and Tilli, 2004] Ambrosio, L. and Tilli, P. (2004). *Topics on analysis in metric spaces*, volume 25. Oxford University Press on Demand.

[Attouch et al., 2014] Attouch, H., Buttazzo, G., and Michaille, G. (2014). *Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization*. SIAM.

[Aubin and Frankowska, 2009] Aubin, J.-P. and Frankowska, H. (2009). *Set-valued analysis*. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA. Reprint of the 1990 edition [MR1048347].

[Aumann, 1964] Aumann, R. J. (1964). Markets with a continuum of traders. *Econometrica: Journal of the Econometric Society*, pages 39–50.

[Aumann, 1966] Aumann, R. J. (1966). Existence of competitive equilibria in markets with a continuum of traders. *Econometrica: Journal of the Econometric Society*, pages 1–17.

[Benning et al., 2013] Benning, M., Calatroni, L., Düring, B., and Schönlieb, C. (2013). A primal-dual approach for a total variation wasserstein flow. In *Geometric Science of Information: First International Conference, GSI 2013, Paris, France, August 28-30, 2013. Proceedings*, pages 413–421. Springer.

[Blanchet and Carlier, 2014a] Blanchet, A. and Carlier, G. (2014a). From nash to cournot–nash equilibria via the monge-kantorovich problem. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2028):20130398.

[Blanchet and Carlier, 2014b] Blanchet, A. and Carlier, G. (2014b). Remarks on existence and uniqueness of cournot–nash equilibria in the non-potential case. *Mathematics and Financial Economics*, 8:417–433.

[Blanchet and Carlier, 2016] Blanchet, A. and Carlier, G. (2016). Optimal transport and cournot-nash equilibria. *Mathematics of Operations Research*, 41(1):125–145.

[Blanchet et al., 2018] Blanchet, A., Carlier, G., and Nenna, L. (2018). Computation of cournot–nash equilibria by entropic regularization. *Vietnam Journal of Mathematics*, 46:15–31.

[Bonnans et al., 2021] Bonnans, J. F., Gianatti, J., and Pfeiffer, L. (2021). A Lagrangian approach for aggregative mean field games of controls with mixed and final constraints. *arXiv preprint arXiv:2103.10743*.

[Bonnivard et al., 2020] Bonnivard, M., Bretin, E., and Lemenant, A. (2020). Numerical approximation of the steiner problem in dimension 2 and 3. *Mathematics of Computation*, 89(321):1–43.

[Bonnivard et al., 2018] Bonnivard, M., Lemenant, A., and Millot, V. (2018). On a phase field approximation of the planar steiner problem: existence, regularity, and asymptotic of minimizers. *Interfaces and free Boundaries*, 20(1):69–106.

[Bonnivard et al., 2015] Bonnivard, M., Lemenant, A., and Santambrogio, F. (2015). Approximation of length minimization problems among compact connected sets. *SIAM Journal on Mathematical Analysis*, 47(2):1489–1529.

[Bourbaki, 2003] Bourbaki, N. (2003). *Topological vector spaces: Chapters 1–5*. Springer Science & Business Media.

[Braides, 1998] Braides, A. (1998). *Approximation of free-discontinuity problems*. Number 1694. Springer Science & Business Media.

[Braides, 2002] Braides, A. (2002). *Gamma-convergence for beginners*. Number 22 in Oxford lecture series in mathematics and its applications. Oxford University Press, New York.

[Brazil et al., 2014] Brazil, M., Graham, R. L., Thomas, D. A., and Zachariasen, M. (2014). On the history of the euclidean steiner tree problem. *Archive for history of exact sciences*, 68(3):327–354.

[Bredies and Holler, 2016] Bredies, K. and Holler, M. (2016). A pointwise characterization of the subdifferential of the total variation functional. *arXiv preprint arXiv:1609.08918*.

[Brenier, 1987] Brenier, Y. (1987). Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808.

[Brenier, 1991] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417.

[Brezis, 2010] Brezis, H. (2010). *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media.

[Brown and Purves, 1973] Brown, L. D. and Purves, R. (1973). Measurable selections of extrema. *The Annals of Statistics*, 1(5):902–912.

[Burger et al., 2012] Burger, M., Franek, M., and Schönlieb, C. (2012). Regularized regression and density estimation based on optimal transport. *Applied Mathematics Research eXpress*, 2012(2):209–253.

[Buttazzo and Stepanov, 2003] Buttazzo, G. and Stepanov, E. (2003). Optimal transportation networks as free dirichlet regions for the monge-kantorovich problem. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 2(4):631–678.

[Cardaliaguet, 2010] Cardaliaguet, P. (2010). Notes on mean field games. Technical report, Technical report.

[Carlier, 2003] Carlier, G. (2003). Duality and existence for a class of mass transportation problems and economic applications. *Advances in mathematical economics*, pages 1–21.

[Carlier and Poon, 2019] Carlier, G. and Poon, C. (2019). On the total variation Wasserstein gradient flow and the TV-JKO scheme. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:42.

[Carrillo et al., 2003] Carrillo, J. A., McCann, R. J., and Villani, C. (2003). Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana*, 19(3):971–1018.

[Carrillo et al., 2006] Carrillo, J. A., McCann, R. J., and Villani, C. (2006). Contractions in the 2-wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, 179:217–263.

[Chambolle, 2004] Chambolle, A. (2004). An algorithm for mean curvature motion. *Interfaces and free Boundaries*, 6(2):195–218.

[Chambolle et al., 2010] Chambolle, A., Caselles, V., Cremers, D., Novaga, M., and Pock, T. (2010). An Introduction to Total Variation for Image Analysis. In *An Introduction to Total Variation for Image Analysis*, pages 263–340. De Gruyter.

[Chambolle et al., 2016] Chambolle, A., Duval, V., Peyré, G., and Poon, C. (2016). Geometric properties of solutions to the total variation denoising problem. *Inverse Problems*, 33(1):015002.

[Chambolle et al., 2019a] Chambolle, A., Ferrari, L. A., and Merlet, B. (2019a). Variational approximation of size-mass energies for k-dimensional currents. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:43.

[Chambolle et al., 2019b] Chambolle, A., Ferrari, L. A. D., and Merlet, B. (2019b). A phase-field approximation of the steiner problem in dimension two. *Advances in Calculus of Variations*, 12(2):157–179.

[Chambolle et al., 2017] Chambolle, A., Lamboley, J., Lemenant, A., and Stepanov, E. (2017). Regularity for the optimal compliance problem with length penalization. *SIAM Journal on Mathematical Analysis*, 49(2):1166–1224.

[Chambolle and Pock, 2021] Chambolle, A. and Pock, T. (2021). Learning consistent discretizations of the total variation. *SIAM Journal on Imaging Sciences*, 14(2):778–813.

[Chauffert et al., 2017] Chauffert, N., Ciuciu, P., Kahn, J., and Weiss, P. (2017). A projection method on measures sets. *Constructive Approximation*, 45(1):83–111.

[Choksi et al., 2017] Choksi, R., Muratov, C. B., and Topaloglu, I. (2017). An old problem resurfaces nonlocally: Gamow's liquid drops inspire today's research and applications. *Notices of the AMS*, 64(11):1275–1283.

[Clarke, 2013] Clarke, F. (2013). *Functional analysis, calculus of variations and optimal control*, volume 264. Springer.

[Combettes and Pesquet, 2011] Combettes, P. and Pesquet, J. (2011). Proximal splitting methods in signal processing. *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212.

[Condat, 2017] Condat, L. (2017). Discrete total variation: New definition and minimization. *SIAM Journal on Imaging Sciences*, 10(3):1258–1290.

[Contreras and Cominetti, 2022] Contreras, J. and Cominetti, R. (2022). Optimal error bounds for non-expansive fixed-point iterations in normed spaces. *Mathematical Programming*, pages 1–32.

[Crauel, 2002] Crauel, H. (2002). *Random probability measures on Polish spaces*, volume 11. CRC press.

[Cucker and Smale, 2007] Cucker, F. and Smale, S. (2007). Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52(5):852–862.

[Cuturi and Peyré, 2018] Cuturi, M. and Peyré, G. (2018). Semidual regularized optimal transport. *SIAM Review*, 60(4):941–965.

[Dacorogna, 2008] Dacorogna, B. (2008). *Direct method in the calculus of variations*. Springer.

[Dal Maso, 1993] Dal Maso, G. (1993). *An Introduction to* Γ*-convergence*. Progress in nonlinear differential equations and their applications. Birkhäuser, Boston, MA.

[David, 2006] David, G. (2006). *Singular sets of minimizers for the Mumford-Shah functional*, volume 233. Springer Science & Business Media.

[De Lellis, 2006] De Lellis, C. (2006). Lecture notes on rectifiable sets, densities, and tangent measures. *Preprint*, 23.

[Delattre and Fischer, 2020] Delattre, S. and Fischer, A. (2020). On principal curves with a length constraint. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 2108–2140. Institut Henri Poincaré.

[Diestel, 2017] Diestel, R. (2017). *Graph Theory*. Springer Berlin, Heidelberg, 5 edition.

[Dondl et al., 2018] Dondl, P., Novaga, M., Wirth, B., and Wojtowytsch, S. (2018). Approximation of the relaxed perimeter functional under a connectedness constraint by phase-fields. *arXiv preprint arXiv:1810.05787*.

[Dondl et al., 2021] Dondl, P., Novaga, M., Wojtowytsch, S., and Wolff-Vorbeck, S. (2021). Connected coulomb columns: analysis and numerics. *Nonlinearity*, 34(9):6120.

[Dondl and Wojtowytsch, 2021] Dondl, P. and Wojtowytsch, S. (2021). Keeping it together: A phase-field version of path-connectedness and its implementation. *Journal of Algorithms & Computational Technology*, 15:17483026211054342.

[Dondl et al., 2017] Dondl, P. W., Lemenant, A., and Wojtowytsch, S. (2017). Phase field models for thin elastic structures with topological constraint. *Archive for Rational Mechanics and Analysis*, 223(2):693–736.

[Dudley, 1969] Dudley, R. M. (1969). The speed of mean glivenko-cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50.

[Dudley, 2002] Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition.

[Ehler et al., 2021] Ehler, M., Gräf, M., Neumayer, S., and Steidl, G. (2021). Curve based approximation of measures on manifolds by discrepancy minimization. *Foundations of Computational Mathematics*, 21(6):1595–1642.

[Evans and Gariepy, 2015] Evans, L. and Gariepy, R. (2015). *Measure theory and fine properties of functions.* CRC press.

[Evans, 2022] Evans, L. C. (2022). *Partial differential equations*, volume 19. American Mathematical Society.

[Falconer, 2004] Falconer, K. (2004). *Fractal geometry: mathematical foundations and applications.* John Wiley & Sons.

[Falconer, 1986] Falconer, K. J. (1986). *The geometry of fractal sets.* Cambridge university press.

[Federer, 2014] Federer, H. (2014). *Geometric measure theory.* Springer.

[Figalli et al., 2010] Figalli, A., Maggi, F., and Pratelli, A. (2010). A mass transportation approach to quantitative isoperimetric inequalities. *Inventiones mathematicae*, 182(1):167–211.

[Fonseca and Leoni, 2007] Fonseca, I. and Leoni, G. (2007). *Modern methods in the calculus of variations: L^ p spaces.* Springer Science & Business Media.

[Frostman, 1935] Frostman, O. (1935). Potential d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions. *Meddel. Lunds Univ. Math. Sem.*, 3:1.

[Fusco, 2003] Fusco, N. (2003). An overview of the mumford-shah problem. *Milan Journal of Mathematics*, 71:95–119.

[Gangbo and McCann, 1996] Gangbo, W. and McCann, R. (1996). The geometry of optimal transportation. *Acta Math*, 177:113–161.

[Gołąb, 1928] Gołąb, S. (1928). Sur quelques points de la théorie de la longueur. *Annales de la Société Polonaise de Mathématiques*, VII:227–241. Rocznik polskiego tow.matematycznego.

[Goldstine, 1980] Goldstine, H. H. (1980). *A History of the Calculus of Variations from the 17th through the 19th Century.* Springer-Verlag New York Inc.

[Gray, 2015] Gray, J. (2015). *The real and the complex: a history of analysis in the 19th century.* Springer.

[Hart et al., 1974] Hart, S., Hildenbrand, W., and Kohlberg, E. (1974). On equilibrium allocations as distributions on the commodity space. *Journal of Mathematical Economics*, 1(2):159–166.

[Hastie and Stuetzle, 1989] Hastie, T. and Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84(406):502–516.

[Hausdorff, 1918] Hausdorff, F. (1918). Auberes mab and dimension. *Mathematische Annalen*, pages 157–179.

[Hewitt and Savage, 1955] Hewitt, E. and Savage, L. J. (1955). Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501.

[Jacobs et al., 2021] Jacobs, M., Lee, W., and Léger, F. (2021). The back-and-forth method for wasserstein gradient flows. *ESAIM: Control, Optimisation and Calculus of Variations*, 27:28.

[Jordan et al., 1998] Jordan, R., Kinderlehrer, D., and Otto, F. (1998). The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17.

[Kallenberg et al., 2017] Kallenberg, O. et al. (2017). *Random measures, theory and applications*, volume 1. Springer.

[Kantorovich, 1942] Kantorovich, L. V. (1942). On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201.

[Karp, 1972] Karp, R. M. (1972). *Reducibility among Combinatorial Problems*, pages 85–103. Springer US, Boston, MA.

[Kégl et al., 2000] Kégl, B., Krzyzak, A., Linder, T., and Zeger, K. (2000). Learning and design of principal curves. *IEEE transactions on pattern analysis and machine intelligence*, 22(3):281–297.

[Klenke, 2013] Klenke, A. (2013). *Probability theory: a comprehensive course*. Springer Science & Business Media.

[Lasry and Lions, 2006a] Lasry, J.-M. and Lions, P.-L. (2006a). Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*, 343(9):619–625.

[Lasry and Lions, 2006b] Lasry, J.-M. and Lions, P.-L. (2006b). Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*, 343(10):679–684.

[Lasry and Lions, 2007] Lasry, J.-M. and Lions, P.-L. (2007). Mean field games. *Jpn. J. Math.*, 2(1):229–260.

[Lebrat et al., 2019] Lebrat, L., de Gournay, F., Kahn, J., and Weiss, P. (2019). Optimal transport approximation of 2-dimensional measures. *SIAM Journal on Imaging Sciences*, 12(2):762–787.

[Lemenant, 2010] Lemenant, A. (2010). A presentation of the average distance minimizing problem.

[Lemenant, 2012] Lemenant, A. (2012). A presentation of the average distance minimizing problem. *Journal of Mathematical Sciences*, 181(6).

[Lemenant, 2016] Lemenant, A. (2016). A selective review on mumford–shah minimizers. *Bollettino dell'Unione Matematica Italiana*, 9(1):69–113.

[Lieu and Vese, 2008] Lieu, L. and Vese, L. (2008). Image restoration and decomposition via bounded total variation and negative hilbert-sobolev spaces. *Applied Mathematics and Optimization*, 58:167–193.

[Liu and Pfeiffer, 2023] Liu, K. and Pfeiffer, L. (2023). Mean field optimization problems: stability results and lagrangian discretization. *arXiv preprint arXiv:2310.20037*.

[Lu and Slepčev, 2016] Lu, X. Y. and Slepčev, D. (2016). Average-distance problem for parameterized curves. *ESAIM: Control, Optimisation and Calculus of Variations*, 22(2):404–416.

[Maggi, 2012] Maggi, F. (2012). *Sets of Finite Perimeter and Geometric Variational Problems: An Introduction to Geometric Measure Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[Mas-Colell, 1984] Mas-Colell, A. (1984). On a theorem of schmeidler. *Journal of Mathematical Economics*, 13(3):201–206.

[Mattila, 1995] Mattila, P. (1995). *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

[McCann, 1995] McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323.

[McCann, 2001] McCann, R. J. (2001). Polar factorization of maps on riemannian manifolds. *Geometric & Functional Analysis GAFA*, 11(3):589–608.

[Mercier, 2018] Mercier, G. (2018). Continuity results for tv-minimizers. *Indiana University Mathematics Journal*, pages 1499–1545.

[Meyer, 2001] Meyer, Y. (2001). *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, volume 22. American Mathematical Soc.

[Modica and Mortola, 1977] Modica, L. and Mortola, S. (1977). Il limite nella γ-convergenza di una famiglia di funzionali ellittici. *Boll. Unione Mat. Ital. V. Ser. A*, 14.

[Monge, 1781] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704.

[Morel and Solimini, 2012] Morel, J.-M. and Solimini, S. (2012). *Variational methods in image segmentation: with seven image processing experiments*, volume 14. Springer Science & Business Media.

[Morgan, 1994] Morgan, F. (1994). (m, ε, δ)-minimal curve regularity. *Proceedings of the American Mathematical Society*, pages 677–686.

[Munkres, 2017] Munkres, J. (2017). *Topology (Classic Version)*. Pearson, London, 2 edition.

[Nash, 1951] Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295.

[Neumayer and Steidl, 2021] Neumayer, S. and Steidl, G. (2021). From optimal transport to discrepancy. *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pages 1–36.

[Paolini and Stepanov, 2004] Paolini, E. and Stepanov, E. (2004). Qualitative properties of maximum distance minimizers and average distance minimizers in rn. *Journal of Mathematical Sciences*, 122(3):3290–3309.

[Paolini and Stepanov, 2013] Paolini, E. and Stepanov, E. (2013). Existence and regularity results for the steiner problem. *Calculus of Variations and Partial Differential Equations*, 46(3):837–860.

[Ponce, 2016] Ponce, A. C. (2016). Elliptic pdes, measures and capacities. *Tracts in Mathematics*, 23:10.

[Pratelli, 2007] Pratelli, A. (2007). On the equality between monge's infimum and kantorovich's minimum in optimal mass transportation. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 43(1):1–13.

[Rockafellar and Wets, 2009] Rockafellar, R. T. and Wets, R. J.-B. (2009). *Variational analysis*, volume 317. Springer Science & Business Media.

[Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268.

[Rudin, 1986] Rudin, W. (1986). *Real and Complex Analysis.* McGraw Hill, New York, 3rd edition edition.

[Sadeghi Arjmand, 2022] Sadeghi Arjmand, S. (2022). *Mean field games with free final time.* PhD thesis, École Polytechnique, CMAP.

[Santambrogio, 2015] Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94.

[Santambrogio and Shim, 2021] Santambrogio, F. and Shim, W. (2021). A Cucker-Smale inspired deterministic mean field game with velocity interactions. *SIAM J. Control Optim.*, 59(6):4155–4187.

[Santambrogio and Tilli, 2005] Santambrogio, F. and Tilli, P. (2005). Blow-up of optimal sets in the irrigation problem. *The Journal of Geometric Analysis*, 15:343–362.

[Schmeidler, 1973] Schmeidler, D. (1973). Equilibrium points of nonatomic games. *Journal of statistical Physics*, 7:295–300.

[Serfaty, 2015] Serfaty, S. (2015). *Coulomb gases and Ginzburg–Landau vortices*. European Mathematical Society (EMS) Zurich.

[Stroock and Varadhan, 1997] Stroock, D. W. and Varadhan, S. S. (1997). *Multidimensional diffusion processes*, volume 233. Springer Science & Business Media.

[Stuhmer et al., 2013] Stuhmer, J., Schroder, P., and Cremers, D. (2013). Tree shape priors with connectivity constraints using convex relaxation on general graphs. In *Proceedings of the IEEE International conference on Computer Vision*, pages 2336–2343.

[Villani, 2009] Villani, C. (2009). *Optimal Transport: Old and New*. Number 338 in Grundlehren Der Mathematischen Wissenschaften. Springer, Berlin.

[Voß, 1992] Voß, S. (1992). Steiner's problem in graphs: Heuristic methods. *Discrete Applied Mathematics*, 40(1):45–72.

[Willmore, 1992] Willmore, T. J. (1992). A survey on willmore immersions. *Geometry and topology of submanifolds, IV (Leuven, 1991)*, pages 11–16.

[Willmore, 1993] Willmore, T. J. (1993). *Riemannian geometry*. Oxford University Press.

Résumé

Dans cette thèse, nous étudions trois problèmes de Calcul des Variations. Bien que leurs thématiques semblent distantes (les problèmes variationnels géométriques, les flots de gradient Wasserstein et la théorie des jeux) ils sont étudiés dans le cadre de l'optimisation dans les espaces de mesures. Le premier consiste à approcher une image donnée avec un ensemble 1-dimensionnel. Pour cela, nous interprétons les images comme des mesures de probabilité et l'on cherche à minimiser la distance de Wasserstein entre la mesure initiale et toutes les mesures uniformément distribuées parmi les ensembles 1-dimensionnels et connexes. Nous démontrons l'existence de solutions pour ce problème, quelques propriétés qualitatives des minimiseurs et nous proposons un résultat d'approximation, en forme de Γ-convergence, qui permet son optimisation numérique. Ensuite nous regardons le flot de gradient de la fonctionnelle de variation totale dans l'espace de Wasserstein. En faisant un lien entre ceci et un problème classique nous utilisons la théorie bien connue de ce dernier pour en déduire les équations d'Euler-Lagrange et obtenir des résultats de régularité. Cette connecxion nous permet aussi de proposer un algorithme proximal pour son optimisation numérique. Dans le troisième problème étudié dans cette thèse, on s'intérèsse à la question "Quand est-ce que les équilibres de Nash d'un jeu à N-joueurs convergent vers une notion d'équilibre d'un jeu avec une infinité de joueurs?", une question centrale dans la théorie des jeux à champ moyen. Pour une classe assez générale de jeux qui possèdent une fonction de potentiel dans l'espace des mesures de probabilité, dont les minimiseurs sont des équilibres, nous démontrons cette convergence en définissant une famille appropriée des jeux à N-joueurs et démontrant que leurs fonctions de potentiel Γ-convergent vers la fonction de potentiel du jeu avec une infinité de joueurs.

Mots clés

Calcul des Variations, Transport Optimal, Théorie Géométrique de la Mesure, Γ-convergence

Abstract

In this thesis, we study three problems in the Calculus of Variations. Although their themes seem far apart, being geometric variational problems, optimal transportation and gradient flows, and game theory, we cast these problems in the unified framework of optimization in spaces of measures. The first consists in approximating a given image with 1 dimensional sets. For this we see images as probability measures and seek to minimize the Wasserstein distance between the given measure and all measures uniformly distributed over 1-dimensional and connected sets. We manage to prove existence of solutions to this problem, some qualitative properties of optimizers and we propose a diffuse approximation result, in the form of Γ-convergence, that enables its numerical optimization. Next we turn to the gradient flow of the total variation functional in the Wasserstein space. By relating it with a classical problem in the literature, the Rudin-Osher-Fatemi problem, we use the latter's properties to derive the Euler-Lagrange equations and regularity properties. This connection also allows us to propose a proximal splitting algorithm to solve it numerically. In the third problem, we study the question of "When to Nash equilibria to N-player games converge to a notion of equilibrium of a game with a continuum of players?" This question is in the heart of Mean Field Games theory and, for a fairly general class of games, which have a potential function on the space of probability measures, whose minimizers are equilibria, we prove this convergence by defining a family of N-player games whose potential function Γ-converge to the one of the game with infinite players.

Keywords

Calculus of Variations, Optimal Transport, Geometric Measure Theory, Γ-convergence