



**HAL**  
open science

# Variational and optimal transport problems in $L^\infty$

Camilla Brizzi

► **To cite this version:**

Camilla Brizzi. Variational and optimal transport problems in  $L^\infty$ . Optimization and Control [math.OC]. Université Paris sciences et lettres; Università degli Studi di Firenze, 2023. English. NNT : 2023UPSLD002 . tel-04200007

**HAL Id: tel-04200007**

**<https://theses.hal.science/tel-04200007v1>**

Submitted on 8 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à l'Université Paris Dauphine  
Dans le cadre d'une cotutelle avec Università degli Studi di Firenze

**Problèmes variationnels et de transport optimal  $L^\infty$**

Variational and optimal transport problems in  $L^\infty$

Soutenue par

**Camilla BRIZZI**

Le 14 Avril 2023

École doctorale n°543

**École doctorale SDOSE**

Spécialité

**Mathématiques**

Composition du jury :

Guillaume CARLIER Professeur des Universités Université Paris Dauphine PSL	<i>Directeur de thèse</i>
Luigi DE PASCALE Associate Professor Università degli Studi di Firenze	<i>Directeur de thèse</i>
Maria Stella GELLI Associate Professor Università di Pisa	<i>Membre du jury</i>
Nicolas JUILLET Professeur des Universités Université Haute-Alsace	<i>Président du jury</i>
Nikos KATZOURAKIS Associate Professor University of Reading	<i>Rapporteur</i>



# Abstract

In this thesis we investigate some properties of solutions of  $L^\infty$ -variational and transport problems. This manuscript is divided into two parts.

The first part, made up of Chapter 2 and Chapter 3, deals with a supremal variational problem. Supremal variational problems appeared for the first time in the late 60s in the pioneering works of Aronsson [7, 8, 9]. Due to the nature of the  $L^\infty$ -norm, the interesting minimizers are the so-called absolute minimizers (AM), which often happen to be solutions of an associated PDE and to have uniqueness and regularity properties. In Chapter 2 we investigate the problem associated with a quasiconvex continuous supremand  $(x, p) \mapsto H(x, p)$ . Notably, we show a new optimality property for  $u \in \text{AM}$  and prove a structure result for the set of points  $x$  where  $H(x, Du(x)) = \max_x H(x, Du(x))$ . In Chapter 3 we consider the supremal variational problem in the framework of problems with constraints on the gradient, proving  $C^1$  regularity of absolute minimizers on the above mentioned set.

In the second part, which consists of Chapter 4 and Chapter 5, we are interested in the  $L^\infty$ -optimal transport problem ( $L^\infty$ -OT), studied for the first time by Champion, De Pascale, and Juutinen in 2007 [47]. An original contribution, presented in Chapter 4, is the proof that the restrictable (optimal) plans (the analogous of AM) are concentrated on a graph, if the cost function is strictly quasiconvex and satisfies a property similar to the classical twist condition. Moreover, we prove uniqueness in the case of a discrete target measure.

The  $L^\infty$ -OT is a non-convex problem, considered to be more difficult than the classical OT. In order to have a better understanding, seeking a generalization to this setting of the entropic approximation seems quite natural. With this intention, in Chapter 5, we provide a regularization which guarantees the  $\Gamma$ -convergence to the non-regularized  $L^\infty$ -OT problem. Remarkably, we show that minimizers of the approximating functionals select restrictable optimal plans. Finally we prove some estimates on the speed of convergence and present some numerical illustrations performed with Sinkhorn's algorithm.

The results in Chapter 2 are based on a joint work with L. De Pascale [32] and the ones of Chapter 3 refer to the preprint [30]. Chapter 4 concerns a joint paper with L. De Pascale and A. Kausamo [33], while Chapter 5 is inspired by a preprint with G. Carlier and L. De Pascale [31].

---

# Riassunto

In questa tesi investighiamo alcune proprietà delle soluzioni dei problemi variazionali e di trasporto in  $L^\infty$ . Il manoscritto è diviso in due parti.

La prima parte, composta dai Capitoli 2 e 3, tratta di un problema variazionale di tipo supremale. I problemi variazionali supremali sono apparsi per la prima volta nei tardi anni '60 nei lavori pionieristici di Aronsson [7, 8, 9]. Considerata la natura della norma  $L^\infty$ , i minimizzanti interessanti sono i cosiddetti *minimizzanti assoluti* (AM), che spesso risultano essere soluzioni di una PDE associata e hanno proprietà di unicità e regolarità. Nel Capitolo 2, analizziamo il problema associato ad un funzionale  $(x, p) \mapsto H(x, p)$  continuo e quasiconvesso. Nello specifico, mostriamo un'ulteriore proprietà di ottimalità per  $u \in \text{AM}$  e dimostriamo un risultato di struttura per l'insieme dei punti  $x$  dove  $H(x, Du(x)) = \max_x H(x, Du(x))$ . Nel Capitolo 3, inseriamo il problema variazionale nel contesto dei problemi con vincoli sul gradiente, provando la  $C^1$  regolarità dei minimizzanti assoluti sull'insieme sopraccitato.

Nella seconda parte, costituita dal Capitolo 4 e dal Capitolo 5, ci interessiamo al problema di trasporto ottimo in  $L^\infty$ , studiato per la prima volta da Champion, De Pascale e Juutinen nel 2007 [47]. Un contributo originale, presentato nel Capitolo 4, è la dimostrazione che i piani ottimali *restrictable* (l'analogo degli AM) sono concentrati in un grafico, se la funzione costo è strettamente quasiconvessa e soddisfa una proprietà simile alla condizione di *twist* classica. Inoltre, mostriamo l'unicità nel caso di una misura *target* discreta.

Il problema OT in  $L^\infty$  è non-convesso e presumibilmente più difficile di quello classico. Al fine di ottenere una migliore comprensione, sembra naturale cercare una generalizzazione a questo ambito delle tecniche dell'approssimazione entropica. A questo scopo, nel Capitolo 5, definiamo una regolarizzazione che garantisce la  $\Gamma$ -convergenza al problema OT in  $L^\infty$  non regolarizzato. In particolare proviamo che i minimi dei funzionali approssimanti selezionano piani *restrictable*. Infine, forniamo alcune stime sulla velocità di convergenza e presentiamo alcuni esempi numerici realizzati con l'algoritmo di Sinkhorn.

I risultati nel Capitolo 2 sono basati su un lavoro con L. De Pascale e quelli nel Capitolo 3 si riferiscono al preprint [30]. Il Capitolo 4 riguarda un articolo con L. De Pascale e A. Kausamo, mentre il Capitolo 5 è ispirato ad un lavoro con G. Carlier e L. De Pascale [31].

---

# Résumé

Dans cette thèse on étudie certaines propriétés des solutions de problèmes variationnels et de transport  $L^\infty$ . Ce manuscrit est divisé en deux parties.

La première partie, composée du Chapitre 2 et du Chapitre 3, traite d'un problème variationnel suprémal. Les problèmes variationnels suprémaux sont apparus pour la première fois à la fin des années 60 dans les travaux pionniers d'Aronsson [7, 8, 9]. En raison de la nature de la norme  $L^\infty$ , les minimiseurs intéressants sont les minimiseurs dits absolus (AM), qui sont souvent solutions d'une EDP associée et ont des propriétés d'unicité et de régularité. À la lumière de ces considérations, dans le Chapitre 2 nous analysons le problème associé à une fonctionnelle continue quasiconvexe  $(x, p) \mapsto H(x, p)$ . Nous montrons notamment une nouvelle propriété d'optimalité pour  $u \in \text{AM}$  et prouvons un résultat de structure pour l'ensemble des points  $x$  où  $H(x, Du(x)) = \max_x H(x, Du(x))$ .

Dans le Chapitre 3, nous resituons le problème variationnel dans le cadre des problèmes avec contraintes sur le gradient, en prouvant la régularité  $C^1$  des minimiseurs absolus sur l'ensemble mentionné ci-dessus.

Dans la deuxième partie, qui comprend le Chapitre 4 et le Chapitre 5, on s'intéresse au problème de transport optimal  $L^\infty$ , étudié pour la première fois par Champion, De Pascale, et Juutinen en 2007 [47]. Une contribution originale, présentée dans le Chapitre 4, est la preuve que les plans optimaux dits restreignables (*restrictable*) (l'analogue de AM) sont concentrés sur un graphe, si la fonction de coût est strictement quasiconvexe et satisfait une propriété similaire à la condition classique de *twist*. De plus, nous prouvons l'unicité dans le cas d'une mesure cible discrète.

Le problème de transport optimal  $L^\infty$  est non convexe, donc vraisemblablement plus complexe que le problème de transport classique. Afin d'avoir une meilleure compréhension, il semble naturel de chercher une généralisation à ce cadre de l'approximation entropique. Dans ce but, dans le Chapitre 5, nous introduisons une régularisation qui garantit la  $\Gamma$ -convergence vers le problème de transport  $L^\infty$ . En particulier, nous montrons que les minimiseurs des fonctionnelles régularisées sélectionnent des plans optimaux *restrictable*. Enfin, nous prouvons quelques estimations sur la vitesse de convergence et présentons quelques illustrations numériques réalisées avec l'algorithme de Sinkhorn.

Les résultats du Chapitre 2 sont basés sur une collaboration avec L. De Pascale [32] et ceux du Chapitre 3 font référence à la pré-publication [30]. Le Chapitre 4 concerne un travail commun avec L. De Pascale et A. Kausamo [33], tandis que le Chapitre 5 est inspiré d'un travail avec G. Carlier et L. De Pascale [31].

---

# Acknowledgements

Desidero prima di tutto ringraziare Luigi De Pascale, il mio relatore già dai tempi della preparazione della tesi magistrale, ormai più di quattro anni fa. A lui va la mia più profonda gratitudine per le lunghe ore di lavoro, sempre molto istruttive, e per la gentilezza e la pazienza dimostratemi in ogni contesto e per ogni necessità. Sono sinceramente grata anche per la presenza costante, in particolare nei periodi più difficili, come i due anni di restrizioni a causa della pandemia, che hanno segnato il mio dottorato. Ma soprattutto, desidero ringraziarlo per la fiducia che ha riposto in me e l'interesse sincero che ha sempre mostrato di avere riguardo la mia opinione.

Mes remerciements vont également à mon autre directeur de thèse, Guillaume Carlier, qui m'a tout de suite accueilli avec beaucoup de gentillesse et a toujours été d'une grande aide, m'introduisant immédiatement dans le groupe. Inutile de dire que les discussions avec lui sont toujours intéressantes, stimulantes et souvent éclairantes. Merci donc, car même si à cause du covid je n'ai pu passer que quelques mois à Paris, j'ai beaucoup grandi et appris pendant ce séjour.

I would like to thank also the two referees Nicolas Julliet and Nikos Katzourakis to have accepted to correct my thesis. Thanks to their remarks and corrections I had the opportunity to improve my work and learn a lot.

Un remerciement particulier à Nicolas, pour tous les commentaires qu'il a partagés avec moi et qui m'ont permis de réfléchir à chaque détail. J'ai beaucoup apprécié l'attention et le soin avec lesquels il a lu la thèse.

Thanks also to Thierry Champion and Francesca Prinari to have accepted, with such a short notice, to attend my defense and to be part of the jury.

Non posso poi che ringraziare Anna Kausamo, la mia "sorella maggiore" in matematica, con cui ho avuto il piacere di collaborare per più di un anno, imparando molto, ma anche la fortuna di poter passare del tempo insieme, di poterla conoscere e di poter diventare sua amica.

Un grazie particolare ai miei compagni di ufficio fiorentini, Ettore, Giada e Lapo, testimoni di tutti i momenti di stress. È stato un piacere condividere questi anni con voi, dai momenti spensierati a quelli più pesanti, agli scleri perché il fancoil non funzionava mai.

Et un grand merci aux doctorants du CEREMADE et de l'Inria, pour avoir été, en plus de collègues, des amis de la première heure. La période, bien que brève, que j'ai passée à Paris a été l'une des meilleures de ce doctorat et c'est principalement grâce à vous : pour les discussions mathématiques et les séminaires des jeunes chercheurs, pour les bières, les conférences, l'escalade et tous les moments de travail et de loisir passés ensemble. Vous me manquez.

Merci également à toute l'équipe de MokaPlan, qui m'a toujours fait sentir que je faisais

---

partie du groupe.

A special thanks goes to Kathi, to whom I owe so much: she has been from the beginning a reference point, a super smart colleague, a true friend. I am so glad for the deep esteem and affection between us.

Ci tengo poi a ringraziare i miei amici. Quelli conosciuti al Dini, compagni di ore infinite e pranzi a Sgabuzzini, momenti che ho rimpianto durante il dottorato, ma anche amici di serate e vacanze. Siete stati una vera e propria famiglia.

Grazie anche a quelli di “Stasera”, a cui devo splendidi momenti, giornate di svago e cammini che hanno accompagnato questi anni, spesso faticosi, di dottorato.

Un grazie gigante alle mie amiche Emma e Meri, per essere pronte a sostenermi e a consigliarmi. Grazie per non essere mai mancate e per continuare ad esserci sempre, ovunque. Grazie per essere una certezza. Vi voglio bene.

Grazie ai miei amici storici, Elisa, Claudio, Matteo e Filippo, per essere stati in questi anni – e da sempre - il mio rifugio e il mio porto sicuro. Ogni volta rivedervi è “tornare a casa”. Grazie a Federica, una splendida scoperta. Grazie a Giulio, Silvia e Morgana che ho conosciuto grazie al federalismo e che, ognuno a modo suo, sono diventati anche tanto altro e tanto importanti, segnandomi profondamente. E grazie a tutte le splendide persone della GFE e della JEF, con cui ho condiviso e condivido la passione per una causa, arricchendomi immensamente. Grazie ad Emanuele, coinquilino fedele e presenza rassicurante. Grazie alle donzelle volanti, per tutti i momenti condivisi, densi e leggeri, per il tempo per noi.

Grazie poi alla mia famiglia, Baldaccini-Brizzi, per il bene profondo che mi dimostrano sempre. Grazie ai miei nonni, enorme fonte di ispirazione. Mi mancate tanto.

Grazie infinite ai miei genitori. Grazie perché nonostante ormai da un po’ di anni io sia (o dovrei essere) un’adulta “fatta e finita”, sono in prima linea nel farmi scudo e ad aiutarmi, a supportarmi in ogni mia fragilità, a sopportarmi. Ma soprattutto sono pronti ad accogliere con entusiasmo qualsiasi mia scelta, assecondandomi senza riserve.

Grazie a mio fratello Filippo, per il bene mai ostentato, ma sempre presente, per la generosità infinita. Sei il mio primo punto di riferimento e di confronto.

Il grazie più grande va infine a Luca, il mio compagno in tutto, un matematico brillante al momento del bisogno, la mia ancora nei momenti di difficoltà, il mio complice nei momenti di gioia. Grazie per tutto l’amore. A te devo ogni riga di questa tesi.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	$L^\infty$ -variational problems . . . . .	1
1.1.1	Constraints on the gradient . . . . .	5
1.2	$L^\infty$ -Optimal Transport problem . . . . .	7
1.2.1	$L^\infty$ -Entropic Optimal Transport . . . . .	11
<b>2</b>	<b>An <math>L^\infty</math>-variational problem and absolute minimizers</b>	<b>17</b>
2.1	Preliminaries . . . . .	17
2.1.1	Some basics of convex analysis . . . . .	17
2.1.2	Some basics about Finsler manifolds . . . . .	19
2.2	Some properties of absolute minimizers of supremal functionals . . . . .	27
2.2.1	The supremand $H$ . . . . .	28
2.2.2	The quasiconvex conjugate of $H$ . . . . .	30
2.2.3	The family of pseudo-distances and a Lipschitz characterization of minimizers . . . . .	32
2.2.4	Absolute minimizers and the property of comparison with Distance Functions . . . . .	37
2.2.5	Point-wise definition of $H(x, Du(x))$ . . . . .	41
2.2.6	The attainment set and its minimality . . . . .	43
<b>3</b>	<b>On functions with convex constraints on the gradient</b>	<b>49</b>
3.1	Preliminaries . . . . .	49
3.2	The uniqueness set . . . . .	56
3.2.1	The structure of the uniqueness set . . . . .	56
3.2.2	Regularity of solutions on the uniqueness set . . . . .	62
3.2.3	The attainment set as uniqueness set . . . . .	67
<b>4</b>	<b><math>L^\infty</math>- Optimal Transport</b>	<b>71</b>
4.1	Preliminaries . . . . .	71
4.1.1	A brief survey of Optimal Transport . . . . .	71
4.1.2	A brief survey of $L^\infty$ -Optimal Transport . . . . .	75
4.1.3	Discrete Optimal Transport . . . . .	80
4.2	$L^\infty$ -Optimal Transport for a class of strictly quasiconvex cost functions . . . . .	82
4.2.1	Basic definitions and main theorem . . . . .	82
4.2.2	On the uniqueness of ICM optimal transport plans . . . . .	90
4.2.3	Some examples . . . . .	93

---

<b>5</b>	<b>Entropic regularization of <math>L^\infty</math>- Optimal Transport</b>	<b>97</b>
5.1	Preliminaries . . . . .	97
5.1.1	A brief survey of Entropic Optimal Transport . . . . .	97
5.1.2	Discrete Entropic Optimal Transport and Sinkhorn Algorithm . . . . .	99
5.2	$L^\infty$ -Entropic Optimal Transport . . . . .	103
5.2.1	$\Gamma$ -convergence . . . . .	104
5.2.2	Selection of plans with $\infty$ -cyclically monotone support . . . . .	107
5.2.3	Some estimates on the speed of convergence . . . . .	110
5.2.4	A large deviations upper bound . . . . .	113
5.2.5	Numerical examples . . . . .	116

# Chapter 1

## Introduction

The work of this thesis is devoted to the study of  $L^\infty$ -variational and optimal transport problems. A general definition of variational problem is minimizing a functional  $\mathcal{E}$  defined on a set  $U$  with values in  $[-\infty, +\infty]$ . Classically the set  $U$  is some suitable set of functions  $u : \Omega \rightarrow \mathbb{R}^n$ , where  $\Omega$  is an open subset of  $\mathbb{R}^d$ .

However, this concept has been widely extended: for instance in Shape Optimization one is interested in minimizing among a collections of “shapes”, usually bounded subsets of  $\mathbb{R}^d$  with some regularity assumptions on the boundary.

In the case of Optimal Transport problems, which is a framework of interest in this thesis, the set  $U$  is a set of Borel probability measures.

### 1.1 $L^\infty$ -variational problems

The first part of this manuscript is dedicated to some aspects related to absolute minimizers of the so-called supremal or  $L^\infty$  - variational problems.

Classically, a variational problem is the problem of minimizing the integral functional

$$\mathcal{F}(u, \Omega) := \int_{\Omega} H(x, u(x), Du(x)) dx, \quad (1.1.1)$$

where  $\Omega$  is an open set of  $\mathbb{R}^d$  and  $H$  is a Borel function defined on  $\Omega \times \mathbb{R}^n \times \mathbb{R}^{dn}$  with values in  $[0, +\infty]$ , among  $u : \Omega \rightarrow \mathbb{R}^n$  belonging to a suitable class of functions.

As the name suggests, an  $L^\infty$ -variational problem is instead the problem of minimizing the functional

$$\mathcal{F}_\infty(u, \Omega) := \|H(\cdot, u, Du)\|_{L^\infty(\Omega)} = \operatorname{ess\,sup}_{x \in \Omega} H(x, u(x), Du(x)), \quad (1.1.2)$$

that is the  $L^\infty$ -norm of the density function  $H$ .

$L^\infty$  variational problems arise naturally in many physical problems and not only. For example, in the problem of the deflection of a transversally loaded beam one seeks to minimize the maximum deflection. In nonlinear elasticity, one wants to minimize the maximum stored energy of a deformation gradient of the reference configuration<sup>1</sup>. In chemotherapy,

---

<sup>1</sup>In physics and continuum mechanics, deformation is the transformation of a body from a reference configuration to a current configuration. A configuration is a set containing the positions of all particles of the body.

one seeks to minimize the maximum tumor load. Also the problem of image reconstruction and enhancement has been formulated as an  $L^\infty$  problem (see [40]).

The existence of minimizers for the functional (1.1.2), for  $u \in W^{1,\infty}(\Omega, \mathbb{R}^n)$ , holds with assumptions comparable with the ones of the integral case. As proved in [16], sufficient conditions for weak\* lower-semicontinuity of  $\mathcal{F}_\infty$  are for instance lower semicontinuity of  $H$  with respect to the last two variables and *quasiconvexity* (i.e. convex sublevel sets) with respect to the last one.

In contrast to the integral case, the essential supremum is a global operator, hence the functional (1.1.2) might not depend locally on  $x$  and might not be sensitive to small perturbations of  $u$ , often leading to high non-uniqueness of minimizers. For this reason, while dealing with supremal variational problems it makes sense to restrict our attention, rather than to general minimizers, to “local” ones, known in the literature as *absolute minimizers*. We say that  $u$  is an absolute minimizer for the functional (1.1.2) if for any  $V \subset\subset \Omega$ ,

$$\mathcal{F}_\infty(u, V) \leq \mathcal{F}_\infty(v, V),$$

for any  $v$  which coincide with  $u$  on the boundary of  $V$ .

$L^\infty$ -variational problems first appear in the pioneering works [7], [8] and [11], where the author also showed for the scalar case  $u : \Omega \rightarrow \mathbb{R}$ , where  $\Omega \subset \mathbb{R}^d$ , that in analogy with the classical variational problems there exists a PDE

$$D(H(\cdot, u, Du)) \cdot H_p(\cdot, u, Du) = 0, \tag{A-E}$$

which can be thought as the Euler-Lagrange equation for this class of problems. This PDE has then become known as Aronsson-Euler equation in his honor. Contrary to the classical case, where being a minimizer is a sufficient condition to solve the associated Euler-Lagrange equation, in this framework the right condition is to be an absolute minimizer, as showed by Barron, Jensen and Wang in [17], where they prove that if  $H$  is  $C^2$  and strictly quasiconvex, any absolute minimizer is a solution of (A-E) in the sense of viscosity (we refer to [55] for more information about viscosity solutions). Their result holds both in the scalar case  $u : \Omega \rightarrow \mathbb{R}$ , with  $\Omega \subset \mathbb{R}^d$  and in the vectorial case  $u : \Omega \rightarrow \mathbb{R}^n$ , where  $\Omega \subset \mathbb{R}^d$ .

The study of the general vectorial case of the problem (1.1.2) is much more complicated and for this topic we refer to some of the pioneering works of Katzourakis [83, 84, 85].

A special case of interest, and the first one to be studied, is the one of the functional

$$u \mapsto \|Du\|_{L^\infty(\Omega)}, \tag{A_\infty}$$

where  $u \in W^{1,\infty}(\Omega, \mathbb{R})$  and the supremand  $H(x, r, p) = |p|$  depends only on the last variable. This functional has been treated first by Aronsson in [9], where he also proves existence of absolute minimizers. The notoriety of the functional (A<sub>∞</sub>) is due to the fact that the associated Aronsson-Euler equation

$$-\Delta_\infty u = -Du^\top \cdot D^2 u Du = 0 \tag{\Delta_\infty}$$

is the  $\infty$ -laplacian equation, which can be interpreted (see for instance [24]) as the limit for  $p \rightarrow \infty$  of the more famous  $p$ -laplacian equation

$$-\Delta_p u = -\operatorname{div}(|Du|^{p-2} Du) = 0, \quad (\Delta_p)$$

which is in turn the Euler-Lagrange equation of the functional

$$u \mapsto \frac{1}{p} \int_{\Omega} |Du(x)|^p dx.$$

Still in [9] it has been shown that, for  $u \in C^1(\Omega)$ , being an *infinity harmonic function*, that is a solution of the equation  $(\Delta_\infty)$ , is not only a necessary but also a sufficient condition to be an absolute minimizer of  $(A_\infty)$ . This result has been generalized to the functionals  $H = H(p) \in C^2(\mathbb{R}^d)$  by Gariepy, Wang, and Yu in [70]. Finally we want to underline the fact that this characterization does not hold in the vectorial case for absolute minimizers in the sense of Aronsson, even if we consider the functional  $\|Du\|_{L^\infty}$ , as showed by Katzourakis and Shaw in [87].

Among many outstanding aspects relating to the equation (A-E), uniqueness and regularity issues of its viscosity solutions are the most challenging.

In his seminal work [75], Jensen proves uniqueness for the viscosity solutions of the Dirichlet problem associated to  $(\Delta_\infty)$ . In [76] this result has been extended to  $p \mapsto H(p)$  which is  $C^2$  and convex.

For the regularity, Savin [109] and Evans and Savin [65] have proved that any infinity harmonic function in  $\mathbb{R}^2$  is  $C^{1,\alpha}$  for some  $0 < \alpha < 1$ . Wang and Yu [121], based on some extensions of the ideas of Savin [109] and the technique of discrete gradient flow for Aronsson's equation, have proved  $C^1$ -regularity for viscosity solutions to Aronsson's equation in  $\mathbb{R}^2$  for any uniformly convex  $H \in C^2(\mathbb{R}^2)$ . The techniques used by Savin in [109] exploit in turn a very interesting result appearing in [53].

In [53] Crandall, Evans, and Gariepy extend the result of [9], proving that every absolute minimizer  $u$  of  $(A_\infty)$  (dropping Aronsson's assumption  $u \in C^1(\Omega)$ ) is characterized not only by being a viscosity solution of  $(\Delta_\infty)$ , but also by what they call *comparison with cones principle*, which states that for any  $V \subset\subset \Omega$ , for any  $a, b \in \mathbb{R}$  and for any  $x_0 \in \Omega$ , if

$$u(x) \geq (\leq) a|x - x_0| + b, \quad \text{on } \partial(V \setminus \{x_0\}),$$

then

$$u(x) \geq (\leq) a|x - x_0| + b, \quad \text{in } V.$$

The success of this characterization led Gariepy, Wang, and Yu [70] to develop the *comparison principle of generalized cones* for absolute minimizers of  $H = H(p) \in C^2(\mathbb{R}^d)$  and the year later Champion, and De Pascale [43] to further generalize it with the so-called *comparison with distance functions principle*, which characterizes the absolute minimizers when  $H = H(x, p)$  is supposed to be only lower semicontinuous, in addition to satisfying the natural assumptions of quasiconvexity and coercivity with respect to the second variable and the assumption of a locally uniform "non-emptiness" of the sublevel sets of  $H$ .

More precisely, they prove that for any  $V \subset\subset \Omega$ , any  $x_0 \in \Omega$ , any  $\lambda \geq 0$ , and  $\alpha \in \mathbb{R}$ , the inequality

$$u \leq d_\lambda(x_0, \cdot) + \alpha, \quad \text{on } \partial(V \setminus \{x_0\})$$

implies

$$u \leq d_\lambda(x_0, \cdot) + \alpha, \quad \text{in } \bar{V},$$

where for every  $\lambda \geq 0$ ,  $d_\lambda$  is a (pseudo) distance which depends on the sublevel sets  $\{p : H(x, p) \leq \lambda\}$  of the Hamiltonian  $H$ . Thanks to this property, in [98] the authors prove the uniqueness of absolute minimizers when  $H$  is convex.

It is possible to show that this distance coincides with the distance  $\delta_H^\lambda$  associated to  $H$  defined by (1.3) in [73]

$$\delta_H^\lambda(x, y) = \sup \{u(y) - u(x) : u \in W^{1,\infty}, \|H(\cdot, Du)\|_{L^\infty(\Omega)} \leq \lambda\}.$$

The importance of the relation between supremal functionals and distances associated with  $H$  has been observed by several author, for instance in [72], under the assumptions of 1-homogeneity of  $H$  with respect to the gradient variable, the authors use the intrinsic distance function to characterize the relaxation of a supremal functional (for an extension of this result see the more recent [73]).

The main idea behind the definition of the family of distances  $(d_\lambda)_\lambda \geq 0$  lies on the fact that for  $H(x, p) = |p|$ , if  $\Omega$  is convex and the boundary datum  $g$  is Lipschitz, then the Dirichlet problem associated to  $(A_\infty)$

$$\min \{ \|Du\|_{L^\infty} : u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega}) \text{ and } u = g \text{ on } \partial\Omega \}$$

admits solutions and an example of minimizers is provided by the so-called *maximal* and *minimal extensions* (or *McShane-Whitney extensions*) of  $g$

$$\begin{aligned} S^-(x) &:= \sup_{y \in \partial\Omega} \{g(y) - \text{Lip}(g, \partial\Omega)|x - y|\} \\ S^+(x) &:= \inf_{y \in \partial\Omega} \{g(y) + \text{Lip}(g, \partial\Omega)|x - y|\}, \end{aligned}$$

where  $\text{Lip}(g, \partial\Omega)$  is the Lipschitz constant of  $g$ .

The goal thus is to define a new metric on  $\Omega$  which depends on  $H$  in a way that McShane-Whitney extensions for this metric are still solutions of the more general problem

$$\min \{ \|H(\cdot, Dv)\|_{L^\infty(\Omega)} : v \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega}), v = g \text{ on } \partial\Omega \}. \quad (\text{H})$$

If we think of  $\Omega$  as a manifold with only one chart, the idea is to define for every  $x$  a substitute for the metric tensor on  $\mathbb{R}^d = T_x\Omega$  which depends on the  $\lambda$ -sublevel sets of  $H$  at  $x$ . By the fact that  $\{p : H(x, p) \leq \lambda\}$  is convex and compact for every  $\lambda$ , the natural candidate is the so-called *support function* of the sublevel sets (see Definition 2.1.3). Since the support function is only positively 1-homogeneous one cannot expect to have a Riemannian structure. However, as showed in Proposition 2.2.11, we have a Finslerian structure, as the one defined in 2.1.11 (for more about Finsler metric see [60]), which allows for the definition of the required family of pseudo-distances, which are also geodesic distances (see Proposition 2.1.27).

In **Chapter 2**, which is based on a joint work with L. De Pascale [32], after giving some preliminary results and definitions, we explore the characterization with the comparison

with distance functions in order to prove further optimality results for the absolute minimizers of the problem (H), where  $H$  is as in [43], with the additional assumptions of continuity and emptiness of the interior of the level sets.

In Section 2.2.5, we show that is possible to give a pointwise definition of

$$H(x, Du(x)),$$

which is denoted by  $H(\cdot, Dv)(x)$  to distinguish it from the classical value which is only almost everywhere defined. This definition is well-posed because, as proved in Proposition 2.2.39 and Proposition 2.2.40d if  $u$  is differentiable at  $x_0$ , then  $H(x_0, Du(x_0)) = H(\cdot, Du)(x_0)$ . Moreover, for any minimizer  $v$  we introduce the *attainment set* (see Definition 2.2.46)

$$\mathcal{A}(v) := \{x \in \Omega \mid H(\cdot, Dv)(x) = \operatorname{ess\,sup}_{x \in \Omega} H(x, Dv(x))\} \quad (1.1.3)$$

and in Theorem 2.2.47 we prove the following optimality property:

if  $u \in W^{1,\infty}(\Omega) \cap C(\overline{\Omega})$  is an absolute minimizer, then

$$\mathcal{A}(u) \subset \mathcal{A}(v) \quad (1.1.4)$$

for all minimizers  $v$  of (H).

In the same theorem we also provide a qualitative description of  $\mathcal{A}(u)$ . More precisely if  $x_0 \in \mathcal{A}(u)$ , for some absolute minimizer  $u$ , there exists a geodesic for  $d_\mu$  passing through  $x_0$  and connecting two different points of  $\partial\Omega$ , with  $\mu$  equal to the minimum of (H).

### 1.1.1 Constraints on the gradient

We observe that solving the problem (H) is equivalent to find  $u \in W^{1,\infty}(\Omega) \cap C(\overline{\Omega})$ , such that

$$\begin{cases} u = g & \text{on } \partial\Omega, \\ Du(x) \in \{p : H(x, p) \leq \mu\} & \text{for a.e. } x \in \Omega, \end{cases}$$

where

$$\mu := \min \{ \|H(\cdot, Dv)\|_{L^\infty(\Omega)} : v \in g + W^{1,\infty}(\Omega) \cap C_0(\Omega) \}.$$

If in addition  $H \geq 0$  satisfies the natural assumptions of quasiconvexity, coercivity, and lower semicontinuity, the sublevel sets are convex bodies (i.e. compact and convex sets) of  $\mathbb{R}^d$  and the system above is the Dirichlet formulation of the problem known in the literature as *convex constraints on the gradient*, i.e. the study of the differential inclusion

$$\{u \in W^{1,\infty}(\Omega) : Du(x) \in K(x) \text{ for a.e. } x \in \Omega\}, \quad (1.1.5)$$

where  $K(x)$  is a convex body for every  $x \in \overline{\Omega}$ .

The interest on problems like (1.1.5), especially for  $K(x) = \overline{B(0, f(x))}$ , for some non-negative function  $f$ , is motivated also by some physical issues, such as the problem of the elastic-plastic torsion of a cylindrical bar (see for instance [116, 63]), polycrystal plasticity, or torsional creep problems and the problem of dielectric-breakdown, in which a body, subject to an electric field  $Du$ , behaves as an insulator if  $Du(x)$  belongs to a convex set

$K(x)$  for a.e.  $x$ , otherwise the dielectric breakdown occurs and the body starts to conduct. Moreover the dielectric-breakdown (see for instance [71, 41] and references therein) can be mathematically interpreted as limiting case of the power-law model, which is known to give a adequate description for some classes of conductors (see for instance [26] and reference therein).

Problems with convex constraints on the gradient appear also in [42], where the authors study the homogenization of some Dirichlet problems as a limit of variational problems with convex bounded constraints on the gradient, providing a proof of a conjecture present in [20]. Finally, convex constraints on the gradient appear also in the Monge-Kantorovich transportation problem, in which one seeks to maximize

$$\int u(x)d\mu(x) - \int u(x)d\nu(x),$$

among all the  $u$ 's such that  $Du(x) \in K(x)$  a.e. (see for instance [120]).

**Chapter 3** is devoted to the presentation of the results present in the preprint [30]. Given a connected and bounded open set  $\Omega \subset \mathbb{R}^d$ , we are interested in the Dirichlet problem

$$\begin{cases} u = g & \text{on } \partial\Omega, \\ Du(x) \in K(x) & \text{for a.e. } x \in \Omega, \end{cases} \quad (\text{P})$$

where  $g \in C(\partial\Omega)$ ,  $u \in W^{1,\infty}(\Omega) \cap C(\overline{\Omega})$  and  $K(x)$  belongs to  $\mathcal{K}$ , the family of convex bodies of  $\mathbb{R}^d$ ,  $K(x)$  is continuous w.r.t. the Hausdorff distance and there exist  $0 < \alpha < M$  such that  $B(0, \alpha) \subset K(x) \subset B(0, M)$  for every  $x$ . This is a classical problem and the case  $K(x) := B(0, f(x))$ , with  $f$  bounded from below by a positive constant, has been studied by Aronsson in [12].

It is straightforward that if  $\Omega$  is convex and  $K(x) = B(0, R)$  for every  $x$ , the problem (P) admits a solution if and only if  $\text{Lip}(g, \partial\Omega) \leq R$ . A similar fact, proved also in [11], still holds for  $K(x) = \overline{B(0, f(x))}$ . We then expect to find a suitable metric such that the same result holds for more general  $K(x)$ . Given the analogy with  $L^\infty$ -variational problems like (H), for every  $x$  we define the *support function*  $\varphi^0(x, \cdot)$  of  $K(x)$  (see Definition 2.1.3), which is in fact a Finsler metric and allows for the definition of a (pseudo) distance  $(x, y) \mapsto d(x, y)$  (see Definition 2.1.14) and it turns out to be

$$d(x, y) = \sup \{u(y) - u(x) : u \in W^{1,\infty}(\Omega) \cap C(\overline{\Omega}), Du(x) \in K(x) \text{ a.e. in } \Omega\},$$

where the left hand side of the equality is a distance introduced in [60] (see also [72, 73]). The case  $K(x) = \{H(x, \cdot) \leq 0\}$  has been treated by Lions in Theorem 5.3 of [92]. However he does not use any Finsler structure and proves this result just for  $H(x, \cdot)$  continuous and convex.

As expected, the result (inspired by [12, 43]) is that the problem (P) admits a solution if and only if the boundary datum  $g$  is 1-Lipschitz w.r.t. to the new distance  $d$  (see Theorem 3.1.14) and examples of solutions are provided by the so-called *maximal* and *minimal extensions*, defined for any  $x \in \overline{\Omega}$  by:

$$\begin{aligned} S^-(x) &= \sup \{g(y) - d(x, y) : y \in \partial\Omega\}, \\ S^+(x) &= \inf \{g(y) + d(y, x) : y \in \partial\Omega\}. \end{aligned}$$

Since for every solution  $u$  of (P),  $S^-(x) \leq u(x) \leq S^+(x)$ , it is natural to investigate the structure of the *uniqueness set*, that is the set of points where  $S^- = S^+$  and thus all the solutions coincides. Theorem 3.2.8 states that for every point  $x$  of the uniqueness set there exists an optimal Lipschitz curve (optimal in the sense of the definition of the Finsler distance see 3.2.2) that passes through  $x$ , it is all contained in  $\Omega$  except for the extreme points, it connects two points of the boundary and  $S^+$ ,  $S^-$  and every solution  $u$  of (P) coincide along it. Moreover, all the solutions of (P) are derivable  $\mathcal{H}^1$ -a.e. along that curve with curve derivative equal to  $\varphi^0$ .

Under suitable assumptions on  $\varphi^0$  or on  $d$  it is also possible to prove (see Proposition 3.2.11 and Proposition 3.2.14) that the maximal and minimal extensions are respectively locally semiconvex and semiconcave. Thanks to this fact in Theorem 3.2.17 we are able to prove a further regularity result: every solution  $u$  of (P) is continuously differentiable on the uniqueness set and if the interior of the uniqueness set is not empty,  $u \in C_{loc}^{1,1}$  on the interior.

Finally, if one inserts the supremal problem (H) of Chapter 2, in the framework of problems like (P), it follows that for every solution  $u$  of (P) the uniqueness set is included in the the attainment set  $\mathcal{A}(u)$ , as shown in Theorem 3.2.22. Moreover in Theorem 3.2.23 is proved that if  $u$  is an absolute minimizers the other inclusion also occurs and the uniqueness and the attainment sets coincides.

A byproduct of this fact is (Corollary 3.2.24) the continuous differentiability of the absolute minimizers on the attainment set under suitable regularity assumptions on  $H$ .

## 1.2 $L^\infty$ -Optimal Transport problem

The second part of this manuscript is dedicated to the optimal mass transport problem in  $L^\infty$ .

The Optimal Transport problem (OT) has its origin in 1871 with Monge [99], who wanted to find the best way to move a pile of sand from a region to another, minimizing the work. In rigorous mathematical terms, if  $(X, d_X)$  and  $(Y, d_Y)$  are two Polish spaces, i.e. separable completely metrizable topological spaces,  $\mu$  and  $\nu$  are two Borel probability measures respectively of  $X$  and  $Y$  (we write  $\mu \in \mathcal{P}(X)$  and  $\nu \in \mathcal{P}(Y)$ ) and  $c : X \times Y \rightarrow \overline{\mathbb{R}}$  is a Borel cost function, he was interested in solving the problem <sup>2</sup>

$$\inf_{\mathcal{T}(\mu, \nu)} \int_X c(x, T(x)) d\mu, \tag{1.2.1}$$

where

$$\mathcal{T}(\mu, \nu) := \{T : X \rightarrow Y \text{ Borel} : T_{\#}\mu = \nu\}.$$

However, since it is hard and sometimes not possible to find a solution of the above formulation, the problem remained unsolved for a long time, until 1942 when Kantorovich ([82]) formulated a relaxed version dropping the deterministic assumption that all the infinitesimal mass in  $x$  is forced to be “sent” by a map to “only” a  $y$ , in favor of the weaker assumption of the mass being transported by a measure and therefore allowing the mass

---

<sup>2</sup>To be more precise in Monge’s original formulation  $X = Y = \mathbb{R}^3$  and  $c(x, y) = |x - y|$ , the Euclidean distance, so that we can think of  $\mu$  and  $\nu$  as the two distributions of sand’s mass and - under the assumption of proportionality between force and mass - the cost is proportional to the work done.

to “be split”. The new problem then becomes

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

where

$$\Pi(\mu, \nu) := \{ \gamma \in \mathcal{P}(X \times Y) : \pi_{\#}^1 \gamma = \mu \text{ and } \pi_{\#}^2 \gamma = \nu \},$$

with  $\pi^1$  and  $\pi^2$  being respectively the projections on the first and the second coordinate of  $X \times Y$ . As it will be discussed in Section 4.1.1, the set  $\mathcal{T}(\mu, \nu)$  is “contained” in  $\Pi(\mu, \nu)$  and Kantorovich’s formulation is actually a relaxation of the Monge’s problem (for a rigorous proof see for instance Section 1.5 in [108]). Moreover, Kantorovich’s version admits a solution - known in the literature as *optimal plan* - under the mild assumptions of lower semicontinuity and boundness from below on  $c$  (see for instance [3, 108]) and if  $c$  is continuous and non-negative and  $\mu$  has no atom (see [105]),

$$\min_{\Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma = \inf_{\mathcal{T}(\mu, \nu)} \int_X c(x, T(x)) d\mu.$$

In contrast to the classical optimal transport in which one considers the integral and thus the overall cost, in the  $L^\infty$ -optimal transport one is interested in the  $L^\infty$ -norm of the cost function with the aim of minimizing the “maximal cost”, i.e.

$$\min_{\gamma \in \Pi(\mu, \nu)} C_\infty(\gamma) := \min_{\gamma \in \Pi(\mu, \nu)} \left( \gamma - \operatorname{ess\,sup}_{(x, y) \in X \times Y} c(x, y) \right) = \min_{\gamma \in \Pi(\mu, \nu)} \|c\|_{L^\infty(X \times Y, \gamma)}. \quad (\text{P}_\infty)$$

Since  $\gamma$  is a probability measure, given a cost function  $c \geq 0$ ,  $\|c\|_{L^\infty(\gamma)}$  is equal to the limit as  $p \rightarrow \infty$  of the  $p$ -norm of  $c$ ,  $\|c\|_{L^p(\gamma)}$ , and therefore the  $L^\infty$ -OT can be seen as the limit problem of

$$\min_{\gamma \in \Pi(\mu, \nu)} C_p^{\frac{1}{p}}(\gamma) := \left( \int_{X \times Y} c^p(x, y) d\gamma \right)^{\frac{1}{p}}. \quad (\text{P}_p)$$

As in the classical setting, minimizers of  $P_\infty$  exist under mild assumptions on  $c$  such as lower-semicontinuity (see [79]).

The first who studied the optimal transport problem in  $L^\infty$  where Champion, De Pascale, and Juutinen, who in [47] carried out a comprehensive study in the case  $X = Y = \mathbb{R}^d$ ,  $d \geq 1$ , and the cost of transporting a point  $x$  to a point  $y$  is given by their Euclidean distance:  $c(x, y) = |x - y|$  and  $\mu$  and  $\nu$  are concentrated on compact sets of  $\mathbb{R}^d$ .

When  $X = Y$  and  $c = d_X$ , the minimal value of  $(\text{P}_\infty)$  is called  $\infty$ -Wasserstein distance of the measures  $\mu$  and  $\nu$  and denoted by  $W_\infty(\mu, \nu)$  and by definition it provides a natural control from above of the  $W_p$  distances, which are the minimal value of the problem  $(\text{P}_p)$  with  $c = d_X$ . The space  $(\mathcal{P}_p(X), W_p)$  is a metric space on the probability measures with finite  $p$ -moment  $\mathcal{P}_p(X)$  for any  $1 \leq p < \infty$  and on the probability measures with bounded support when  $p = \infty$ . When  $1 < p < \infty$ , especially  $p = 2$ , this distance and the associated metric space has been widely studied, leading to a considerable improvement in the theory of optimal transport and its range of application: among the many others we recall the definition of displacement interpolation given for the first time by McCann in [94] or of Wasserstein barycenter due to Agueh, and Carlier appeared for the first time in [2], but

also the breakthrough of the ‘‘Riemannian nature’’ of the Wasserstein spaces, which was pointed out first in [102, 103].

Even though the case  $p = \infty$  is more delicate and less understood, the distance  $W_\infty$  appears naturally in many modelling issues. One of the first paper using such a distance was [96], devoted to the shape of rotating stars, in which it turns out that the right local minimizers of a functional  $F$  to consider are the one with respect to  $W_\infty$ , because they happen to be characterized by the fact that every point of their support satisfies a local minimality condition for a function  $f$ , which is the first variation of  $F$ .

Finally it is interesting to mention that for  $X = \mathbb{R}^d$  and  $c$  the Euclidean distance, Bouchitté, Jimenez and Rajesh provide in [25] a reverse inequality between  $W_\infty$  and  $W_p$ , of the form

$$W_\infty(\mu, \nu) \leq CW_p(\mu, \nu)^{\frac{p}{p+d}},$$

where  $C$  is a constant depending on  $p$ , on the dimension  $d$ , on  $\Omega$  and on a lower bound on the density of  $\mu$ . The result was also generalized to more general costs by Jylhä and Rajala in [80] establishing necessary and sufficient conditions for the existence of  $W_\infty$ -lower bounds for the integral optimal transportation costs.

Since the optimal transport problem can be seen as a variational problem in which one minimizes the functional among a set of measures, it seems reasonable to expect similarities between the  $L^\infty$ -Monge-Kantorovich problem and the  $L^\infty$ -variational problems presented before. As for the minimizers of the  $L^\infty$  variational problems, in general we can not expect any local property of optimal  $L^\infty$ -transportation plans. Indeed, differently from the standard integral Monge-Kantorovich problem, in which the functional  $\gamma \mapsto \int cd\gamma$  is linear, if  $\gamma$  is an optimal plan for the  $L^\infty$ -OT problem and  $B$  is a Borel subset of  $X \times Y$ ,  $\gamma|_B$  is not necessarily optimal with respect to its marginals (see for instance Example 4.1.22). To address this problem, Champion, De Pascale, and Juutinen invoked the concept of *restrictability*. An optimal  $L^\infty$ -transportation plan is restrictable if, loosely speaking, its restrictions are also optimal (for the rigorous definition see Definition 4.1.32) and it can be seen as the the analogous of the *absolute minimizers* in the  $L^\infty$ -calculus of variations problems. Moreover, they introduced the concept of  $\infty$ -*c-cyclical monotonicity*, which carries the well-known notion of  $c$ -cyclical monotonicity of optimal transportation plans (see [3, 108, 120] for a detailed discussion) to the  $L^\infty$ -case. A plan  $\gamma$  is  $\infty$ - $c$ -cyclically monotone if for any finite set of points  $\{(x_i, y_i)\}_{i=1, \dots, N}$  in the support of  $\gamma$ , assigning  $x_i$  to  $y_i$  is the best choice, in the sense that if we permute the  $y_i$ 's with any permutation  $\sigma$ , the maximum of  $c$  over  $\{(x_i, y_{\sigma(i)})\}_{i=1, \dots, N}$  is never smaller than the maximum of  $c$  over  $\{(x_i, y_i)\}_{i=1, \dots, N}$ . The interesting fact is that they showed that restrictability and  $\infty$ -cyclical monotonicity are equivalent. This equivalence also holds for more general cost functions at least if they are continuous, as it was proven by Jylhä in Theorem 2.19 of [79].

In light of the fact that  $\Pi(\mu, \nu)$  is a convex set and the functional  $\gamma \mapsto \int cd\gamma$  is by linearity a convex functional, the classical Kantorovich optimal transport problem is a convex optimization problem. It is well known in optimization that associated with such a problem there is a dual problem, which provides useful information about the primal one. More precisely the Monge-Kantorovich problem is equivalent to the dual formulation (see also (DualOT)) given by

$$\sup \left\{ \int_X \varphi(x) d\mu + \int_Y \psi(y) d\nu : \varphi \oplus \psi \leq c, (\varphi, \psi) \in C_b(X) \times C_b(Y) \right\}.$$

The above dual formulation is a key aspect of the theory in optimal transport, especially for deriving fundamental results of optimal transport like Brenier's theorem ([27]) and the Gangbo-McCann's theorem ([69]), which not only prove the existence of *optimal maps* for the Monge's formulation (1.2.1) but also the uniqueness and the deterministic nature of optimal plans for some cost functions  $c$ , i.e. that optimal plans are concentrated on graphs.

Since the functional  $\gamma \mapsto C_\infty(\gamma)$  is not convex, we are not dealing with a convex optimization problem and the question on whether there exists a dual formulation for the  $L^\infty$ -optimal transport is more difficult. Contrary to the primal formulation it is not possible to consider the problem

$$\sup \left\{ \left( \int_X \varphi_p(x) d\mu + \int_Y \psi_p(y) d\nu \right)^{\frac{1}{p}} : \varphi_p \oplus \psi_p \leq c^p, (\varphi, \psi) \in C_b(X) \times C_b(Y) \right\}$$

and send  $p \rightarrow \infty$ . Indeed even in the unlikely case in which  $\varphi \oplus \psi \geq 0$  and the power  $\frac{1}{p}$  makes sense, there is no reason why the limit  $p \rightarrow \infty$  should give any information.

Yet, although not convex, analogously to the supremal functionals, the functional  $\gamma \mapsto C_\infty(\gamma)$  is *quasiconvex*, which means that the level sets are convex. This allows for a (sort of)  $L^\infty$ -duality theory introduced and investigated first by Barron, Bocea, and Jensen in [15]. However, the existence of trivial solutions in the dual formulation they provide prevents it to be a useful and immediate tool.

The theory was further developed in the 1-dimensional case for the cost function  $c$  equal to the Euclidean distance  $|x - y|$  by De Pascale and Louet in [61], where they construct an example of a non trivial pair  $(\varphi, \psi)$ , solution of the  $L^\infty$ -dual formulation (4.1.2). As an application, they prove a result (Theorem 4.2 in [61]) that share similarities with Theorem 2.2.47 in Chapter 2, about the properties of the attainment set of an absolute minimizer for the supremal variational problem (H). Indeed they prove that there exists a set of points of the support of the source measure which are displaced at maximal distance by all the optimal transport plans and on this set, which is shared by all the  $\infty$ -cyclically monotone transport plans, any optimal transport plan must exactly coincide with the  $\infty$ -cyclically monotone one.

Even though the absence of a satisfying duality theory, Champion, De Pascale, and Juuinen were able to introduce in [47] a duality-free technique to show that for  $X = Y = \mathbb{R}^d$  and  $c(x, y) = |x - y|$ , under natural assumptions on  $\mu$ ,  $\infty$ -cyclical monotone transport plans - or, even less,  $\infty$ -monotone transport plans (i.e. when the monotonicity property holds only for sets of two points) - are actually induced by a transport map. The novelty of their work is not only the result itself but also the fact that the technique they developed turned out to be very useful to prove existence of solutions of the Monge problem also for some integral transport costs, in which it is not evident how to exploit duality (see for instance [44, 45]).

Still in [47], they prove that if the target measure  $\nu$  is atomic,  $\infty$ -cyclical monotone - and thus restrictable - transport plans are also unique. More precisely, if  $\nu$  contains an atom, say  $y_0$ , and  $T$  and  $S$  are two optimal transport maps, both corresponding to infinitely cyclically monotone transport plans, then  $\mu(T^{-1}(\{y_0\}) \setminus S^{-1}(\{y_0\})) = 0$ .

Jylhä in [79] generalizes the result of [47] to cost functions of the form  $c(x, y) = h(y - x)$ , where  $h : \mathbb{R}^d \rightarrow \mathbb{R}_+$  is strictly quasiconvex in the sense that for all  $t \in (0, 1)$  and  $x, \bar{x} \in \mathbb{R}^d$  with  $x \neq \bar{x}$

$$h((1 - t)x + t\bar{x}) < \max\{h(x), h(\bar{x})\}.$$

In **Chapter 4**, after all the needed preliminaries, we present an original version of the results discussed above about the deterministic nature of  $\infty$ - $c$ -cyclically monotone plans and their uniqueness, based on a joint work with L. De Pascale, and A. Kausamo [33]. In particular consider a different cost function  $c$  which is strictly quasiconvex with respect to one coordinate, but not translation invariant, and satisfies a special condition on the norm of the gradient (see Theorem 4.2.10).

The relevant notion in the Monge-Kantorovich problem for integral costs is an invertibility condition on the partial gradient of  $c$  ([46, 67] and [108] and references therein). This condition is commonly used in dynamical systems, it is called the *twist condition* and it goes very well along with convexity properties. There is no analogue of the twist condition for  $L^\infty$ -optimal transport problem. In Theorem 4.2.8 we introduce a property which could serve as twist condition in this setting and we give some examples of costs which have this property (see Section 4.2.3). The property we introduce is invariant (as one could expect) by composition of  $c$  with a strictly increasing, differentiable function. The proof of the “Brenier type” result, that is the concentration of  $\infty$ - $c$ -cyclically monotone plans on a graph (and the consequent existence optimal maps), is provided by Theorem 4.2.10.

Although in Theorem 4.2.13 we do not change the essence of the proof of uniqueness, we give a new structure to it which, in our opinion, makes the arguments more transparent and may allow a better understanding of the problem.

The proof of the main results is based on measure theoretic considerations and on the construction of certain specific cones. The same happens in [79] as well as in previous works in which these techniques are applied. What is different is that the construction of [79] is based on the translation invariance of the cost while our construction relies on the notion of normal cone to the boundary of a convex set. This points to the possibility of a general construction which may be adapted to different costs.

A different technique to prove the existence of an optimal transport map, which applies to several costs of the form  $c(x, y) = h(x - y)$ , for  $h$  strictly convex, was devised by C. Jimenez and F. Santambrogio in [78]. They consider the convex set

$$\Pi_\infty(\mu, \nu) := \left\{ \gamma \in \arg \min_{\Pi(\mu, \nu)} C_\infty \right\}$$

and define the secondary variational problem

$$\min \left\{ \int_{X \times Y} |x - y|^2 d\gamma : \gamma \in \Pi_\infty(\mu, \nu) \right\}.$$

They prove that if  $\gamma$  is optimal for the secondary problem then it is deterministic. However the optimal map of [78] is not proved to be  $\infty$ - $c$ -cyclically monotone.

### 1.2.1 $L^\infty$ -Entropic Optimal Transport

The main idea of entropic regularization of the optimal transport problem in the integral setting is to “relax” the Monge-Kantorovich problem by adding a noise. This allows to replace the Monge-Kantorovich problem by a strongly convex problem which admits a unique solution.

More precisely, given two probability measure  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  and a cost function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ , the Entropic Optimal Transport problem (also called entropy-regularized optimal transport problem, EOT for short) reads as:

$$v_\varepsilon := \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c d\gamma + \varepsilon H(\gamma | \mu \otimes \nu) \right\},$$

where the penalization term

$$H(\gamma | \mu \otimes \nu) := \begin{cases} \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \left( \frac{d\gamma}{d\mu \otimes \nu} \right) d\gamma & \text{if } \gamma \ll \mu \otimes \nu, \\ +\infty & \text{otherwise,} \end{cases}$$

is the Boltzmann-Shannon relative entropy (also called Kullback-Leibler divergence) and  $\varepsilon > 0$  can be interpreted as a temperature parameter. Heuristically, this consists in moving  $\mu$  onto  $\nu$  in the cheapest and (at the same time) most “diffuse” way, since the minimizer  $\gamma_\varepsilon$  of  $v_\varepsilon$  is forced to be absolutely continuous with respect to  $\mu \otimes \nu$  because of the entropy term, which is  $+\infty$  otherwise, and so its mass has to be “spread out”, in contrast with solutions to the unperturbed transport problem, which happen to be concentrated on graphs of Borel maps. Clearly, the “spreadness” of the approximate plan will depend on the magnitude of the regularization parameter.

As highlighted by C. Léonard in many works, see for instance [89, 90], when  $c(x, y) = \frac{|x-y|^2}{2}$  the entropic regularization of the optimal transport problem is deeply connected to the problem posed by Schrödinger in his seminal works [111, 112]. Suppose that you know that at initial time  $t = 0$  a very large number of non-interacting indistinguishable particles are approximately distributed as a probability measure  $\mu$  and that the dynamics of each individual particle is driven by a Brownian motion with diffusion parameter  $\varepsilon$ . Suppose also that at a final time  $t = 1$ , you observe the system in a configuration close to some probability measure  $\nu$  which is far from the one expected by the law of large numbers. Then Schrödinger’s question is: “*Conditionally on this very rare event, what is the most likely path of the whole system between the times  $t = 0$  and  $t = 1$ ?*”

Since EOT is a perturbed transport problem it is natural to investigate its behaviour as the parameter  $\varepsilon$  approaches 0 and if and how it approximates the Monge-Kantorovich problem. In the pioneering works of Léonard [89] (which tackled the question from the Schrödinger problem’s viewpoint) and of Carlier *et al.* in [38] the  $\Gamma$ -convergence was proved in the case  $c(x, y) = \frac{|x-y|^2}{2}$  (but it can be easily generalized to any  $1 \leq p < +\infty$ ). The same result has been then extended to a very large class of continuous cost functions. As a direct consequence, the optimal value  $v_\varepsilon$  of the EOT problem converges to the optimal value of the OT problem -  $W_2^2(\mu, \nu)$  in the case of quadratic costs - raising naturally the question about a possible quantitative rate (see for instance [1, 64, 104, 50, 49, 39]). Also the convergence of optimal plans and potentials - i.e. the solutions of the dual formulation of the EOT problem - has been largely studied (see [100] as a comprehensive survey and the references therein) and the question of quantitative estimates has been addressed in [22, 101] by leveraging on a large deviations interpretation and on the notion of  $(c, \varepsilon)$ -invariance, defined for the first time in [22]

In the last decade, the entropic optimal transport problem has witnessed an increasing

interest and is now an extremely active research topic, because it has found numerous applications and proved to be an efficient way to approximate OT problems, especially from a computational viewpoint.

The Monge-Kantorovich problem is a linear programming, which means a linear optimization problem under linear constraints. In the discrete case this is even more evident and this property has allowed for the use of the simplex method algorithm which provides exact solutions, but it is computationally prohibitive in big dimensions ( $N^2$  unknown with  $2N$  constraints). Other algorithms were later found, such as the Hungarian (see [88]) and the Auction (see [23]) methods, but the computational cost still remains high. For a complete overview on numerical methods for optimal transport see [58, 19, 97].

On the contrary, one of the greatest advantage of EOT lies on the fact that we can introduce a simple algorithm to compute the solution, which is performed by alternating Kullback-Leibler projections on the two marginal constraints. By the algebraic properties of the entropy such iterative projections correspond to the celebrated Sinkhorn's algorithm [114], which was applied in this framework for the first time by Cuturi in the pioneering work [57]. Beside being simple, this method turns out to converge much better (see [68, 62]) compared to the algorithms used for the Monge-Kantorovich problem, determining the success of EOT for applications in machine learning, statistics, image processing, language processing and other areas (see the monograph [58] and references therein).

As already discussed, the optimal transport problem in  $L^\infty$ -OT is a nonconvex and presumably much harder problem than OT and many aspects remain still unsolved and not fully understood, such as the question of finding the best assumptions on  $c$  in order to have Brenier type results as the one of Theorem 4.1.37 and Theorem 4.2.10 or of extending the result on uniqueness (Theorem 4.1.38 and Theorem 4.2.13) to more general target measures  $\nu$ .

Therefore due to the success of entropic approximation of the integral optimal transport just recalled and in order to gain a better understanding, it makes sense to try to investigate, theoretically and numerically, whether the entropic approximation strategy can be used for  $L^\infty$ -optimal transport as well.

In **Chapter 5** we present this investigation, based on an ongoing work with G. Carlier and L. De Pascale [31].

The approximated problem we provide reads

$$v_{\varepsilon,p} := \min_{\gamma \in \Pi(\mu, \nu)} \left\{ \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)^p d\gamma(x, y) + \varepsilon H(\gamma | \mu \otimes \nu) \right)^{\frac{1}{p}} \right\}.$$

Since  $H(\gamma | \mu \otimes \nu) \geq 0$  with an equality exactly when  $\gamma = \mu \otimes \nu$ , we can see that

$$\left( \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)^p d\gamma(x, y) + \varepsilon H(\gamma | \mu \otimes \nu) \right)^{\frac{1}{p}} \geq \|c\|_{L^p(\gamma)}$$

but

$$\|c\|_{L^p(\gamma)} \leq \|c\|_{L^\infty(\gamma)},$$

so roughly speaking these approximations play in opposite directions: adding the entropic term is an approximation *from above* but approximating  $\|c\|_{L^\infty(\gamma)}$  by  $\|c\|_{L^p(\gamma)}$  is an approximation *from below*.

Under suitable assumptions on the convergence of  $\varepsilon \rightarrow 0$  - which has to be related to  $p$  - or

on the cost function  $c$ , in Theorem 5.2.4 we show the  $\Gamma$ -convergence towards the supremal cost functional when  $p \rightarrow \infty$  and  $\varepsilon(p) \rightarrow 0$ .

Moreover, what is particularly interesting in our opinion is that the results of [22] can be used to show in Theorem 5.2.9 that  $\Gamma$ -convergence selects at the limit the distinguished restrictable  $\infty$ - $c$ -cyclically monotone minimizers, i.e. any cluster point  $\gamma_\infty$  of the minimizers  $\gamma_{\varepsilon,p}$  of  $v_{\varepsilon,p}$  is  $\infty$ - $c$ -cyclically monotone and therefore, in some cases, concentrated on a graph of a map.

Also the question of quantitative estimates on the rate of convergence of  $v_{\varepsilon,p}$  has been addressed. In Proposition 5.2.11 and Remark 5.2.12, under the assumption of  $\|c\|_{L^\infty(\gamma_\infty)}$  greater than 1, we have an exponential upper bound, while providing a lower bound is more delicate and we manage to have a result just in the discrete case (Proposition 5.2.14). In Section (5.2.4), we also discuss a partial extension of the large deviations result proved in [22]

Finally, by using the Sinkhorn's algorithm, we were able to give numerical examples, which we think can well illustrate what has been proved and are presented in Section 5.2.5.

# Notation

i.e.	<i>namely;</i>
w.r.t	<i>with respect to;</i>
a.e.	<i>almost everywhere;</i>
s.t.	<i>such that;</i>
$\mathbb{N}, \mathbb{N}^*$	set of non-negative integers, set of positive integers;
$\mathbb{R}$	set of real numbers;
$\overline{\mathbb{R}}$	the set $\mathbb{R} \cup \{+\infty\}$
$\mathbb{R}_+$	set $[0, +\infty)$ of non-negative real numbers;
$\mathbb{R}^d$	vector space of dimension $d > 1$ on the set of real numbers;
$p \cdot q$	standard scalar product between two vectors $p, q \in \mathbb{R}^n$
$\text{int } A$	the interior of a set $A$ ;
$\overline{A}$	the closure of a set $A$ ;
$B(x_0, r)$	open ball of $\mathbb{R}^d$ with radius $r$ and center $x_0 \in \mathbb{R}^d$ ;
$\text{dist}(x, A)$	distance function, i.e. $\inf_{y \in A}  x - y $ ;
$C(X)$	space of continuous functions defined on the space $X$ with real values;
$C_b(X)$	space of bounded continuous functions defined on the space $X$ with real values;
$\text{Lip}(X)$	space of Lipschitz continuous functions;
$\text{Lip}(f, X)$	Lipschitz constant of a function $f$ on the space $X$ ;
$\mathcal{M}(X)$	space of Borel measures on $X$ ;
$\mathcal{P}(X)$	space of Borel probability measures on $X$ ;
$\text{supp } \mu$	support of the measure $\mu$ ;
$L^p(X, \mu), L^p(\mu)$	space of measurable functions for which $p$ th power of absolute value is integrable w.r.t. the measure $\mu$ , $p \geq 1$ ;
$L^\infty(X, \mu)$	space of measurable functions with finite essential supremum w.r.t. $\mu$ ;
$\Omega$	open subset of $\mathbb{R}^d$ ;
$V \subset\subset \Omega$	open subset such that $V \subset \overline{V} \subset \Omega$ ;
$C_c(\Omega)$	space of continuous functions on $\Omega$ with compact support;
$C_0(\Omega)$	space of continuous functions on $\Omega$ vanishing at the boundary;
$C^k(\Omega)$	space of $k \in \mathbb{N}$ times continuously differentiable functions on $\Omega$ ;
$C^k(\overline{\Omega})$	restriction on $\overline{\Omega}$ of $C^k(\mathbb{R}^d)$ ;
$C^{k,\alpha}(\Omega)$	space of $k \in \mathbb{N}$ times continuously differentiable functions on $\Omega$ , whose $k$ th derivative is Hölder continuous with exponent $\alpha \in (0, 1]$ ;

$\mathcal{L}^d$	Lebesgue measure on $\mathbb{R}^d$ ;
$\mu \ll \mathcal{L}^d$	Borel measure $\mu$ absolutely continuous w.r.t. to $\mathcal{L}^d$ ;
$L^p(\Omega), L^\infty(\Omega)$	$L^p, L^\infty$ space on $\Omega$ w.r.t. the Lebesgue measure;
$W^{k,p}(\Omega, \mathbb{R}^d)$	Sobolev of order $k \in \mathbb{N}^*$ and integrability $p \geq 1$ with values in $\mathbb{R}^d$ ;
$W^{k,p}(\Omega)$	Sobolev of order $k \in \mathbb{N}^*$ and integrability $p \geq 1$ with values in $\mathbb{R}$ ;
$W^{k,\infty}(\Omega)$	Sobolev of order $k \in \mathbb{N}^*$ and finite essential supremum;
$g + W^{1,\infty}(\Omega) \cap C_0(\Omega)$	$u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ s.t. $u = g$ on $\partial\Omega$ ;
$\nabla u$	classical gradient of a function $u$ defined on $\Omega$ ;
$Du$	weak gradient of $u$ (in case of $u \in W^{1,\infty}(\Omega)$ , $Du(x) = \nabla u(x)$ a.e.).

**Disclaimer.** Following the usual notation present in the literature:

- in Chapter 2 and 3 we denote by  $\gamma$  a continuous curve  $\gamma : [a, b] \rightarrow A$ , where  $[a, b]$  is an interval of  $\mathbb{R}$  and  $A \subset \mathbb{R}^d$ ;
- in Chapter 4 and 5 we denote by  $\gamma$  a probability measure belonging to the transport plans from two probability measures  $\mu$  and  $\nu$ .

## Chapter 2

# An $L^\infty$ -variational problem and absolute minimizers

### 2.1 Preliminaries

#### 2.1.1 Some basics of convex analysis

From now on we will refer to a compact convex set as *convex body* and we define the set  $\mathcal{K}$  as

$$\mathcal{K} := \{K \subset \mathbb{R}^d : K \text{ convex body}\}. \quad (2.1.1)$$

For an exhaustive reference about convex bodies, see for instance [110].

**Definition 2.1.1.** Let  $K$  be a convex body such that  $0 \in \text{int } K$ . The function  $\varphi : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$  defined by

$$\varphi(p) := \inf \left\{ t > 0 : \frac{p}{t} \in K \right\},$$

is called *gauge function* of  $K$ .

*Remark 2.1.2.* One can easily prove the following properties:

- (i) non-negativity:  $\varphi(p) \geq 0$  for every  $p \in \mathbb{R}^d$  and  $\varphi(p) = 0$  if and only if  $p = 0$ ;
- (ii) positive 1-homogeneity:  $\varphi(rp) = r\varphi(p)$  for all  $r > 0$ ;
- (iii) subadditivity:  $\varphi(p_1 + p_2) \leq \varphi(p_1) + \varphi(p_2)$ , for all  $p_1, p_2 \in \mathbb{R}^d$ .

We note that  $\varphi$  is not exactly a norm: the property (ii) tells that  $\varphi$  is positively 1-homogeneous but not absolutely. Indeed it can happen that  $\varphi(p) \neq \varphi(-p)$ . Therefore  $\varphi$  is a norm (and one can write  $\varphi(p) = \|p\|_K$ ) if and only if  $K$  is symmetric with respect to the origin  $0$ . However, thanks to (ii) and (iii) we still have the convexity of  $\varphi(\cdot)$ .

**Definition 2.1.3.** Let a convex body  $K \subset \mathbb{R}^d$  such that  $0 \in \text{int } K$ , the *support function* is the function  $\varphi^0 : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ , defined by

$$\varphi^0(q) := \sup\{p \cdot q : p \in K\}.$$

*Remark 2.1.4.* Directly by the definition of support function, we have the same properties of  $\varphi$ :

- (i) non-negativity:  $\varphi^0(q) \geq 0$  for every  $q \in \mathbb{R}^d$  and  $\varphi^0(q) = 0$  if and only if  $q = 0$ ;
- (ii) positive 1-homogeneity:  $\varphi^0(rq) = r\varphi(q)$  for all  $r > 0$ ;
- (iii) subadditivity:  $\varphi^0(q_1 + q_2) \leq \varphi(q_1) + \varphi(q_2)$ , for all  $q_1, q_2 \in \mathbb{R}^d$ .

A before, (ii) and (iii) imply convexity.

**Definition 2.1.5.** Let  $K \subset \mathbb{R}^d$  a convex body such that  $0 \in \text{int } K$ , then the *polar set* of  $K$  is defined by

$$K^0 := \left\{ q \in \mathbb{R}^d : p \cdot q \leq 1 \text{ for all } p \in K \right\}.$$

*Remark 2.1.6.* Clearly  $K^0$  is a convex body with  $0 \in \text{int } K$ . Moreover we observe that

$$\begin{aligned} B_\varphi(0, 1) &:= \{p \in \mathbb{R}^d : \varphi(p) \leq 1\} = K, \\ \partial B_\varphi(0, 1) &:= \{p \in \mathbb{R}^d : \varphi(p) = 1\} = \partial K \end{aligned} \tag{2.1.2}$$

and

$$\begin{aligned} B_{\varphi^0}(0, 1) &:= \{q \in \mathbb{R}^d : \varphi^0(q) \leq 1\} = K^0, \\ \partial B_{\varphi^0}(0, 1) &:= \{q \in \mathbb{R}^d : \varphi^0(q) = 1\} = \partial K^0 \end{aligned} \tag{2.1.3}$$

**Proposition 2.1.7.**  $K^{00} = K$ .

*Proof.* Since

$$K^{00} := \{w : w \cdot q \leq 1 \text{ for all } q \text{ s.t. } q \cdot p \leq 1, \text{ for all } p \in K\},$$

the inclusion  $K \subset K^{00}$  is straightforward.

Let us assume by contradiction that there exists  $p \in K^{00}$  such that  $p \notin K$ . By the Hahn-Banach Theorem, there exists a closed hyperplane  $\{v : f \cdot v = \alpha\}$ , with  $\alpha > 0$  since  $0 \in K$ , such that

$$f \cdot v < \alpha < f \cdot p, \quad \text{for all } v \in K.$$

If we take  $\bar{f} = \varphi^0(f)^{-1}f$  ( $f \neq 0$  because  $\alpha > 0$ ), then  $\varphi^0(\bar{f}) = 1$  and we get

$$\bar{f} \cdot v < \frac{\alpha}{\varphi^0(f)} < \bar{f} \cdot p \leq 1, \quad \text{for all } v \in K,$$

so that  $\bar{f} \cdot v < 1$  for all  $v \in K$ . On the other hand, by compactness we know that there exists  $\bar{v} \in K$  such that  $\bar{f} \cdot \bar{v} = \varphi^0(\bar{f}) = 1$  and we get a contradiction.  $\square$

**Definition 2.1.8.** We say that a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is *quasiconvex* if for all  $x, y \in \mathbb{R}^d$  and for all  $t \in [0, 1]$  we have

$$f((1-t)x + ty) \leq \max\{f(x), f(y)\}.$$

We say that  $f$  is *strictly quasiconvex* if for all  $x \neq y$  and for all  $t \in (0, 1)$

$$f((1-t)x + ty) < \max\{f(x), f(y)\}.$$

*Remark 2.1.9.* Immediate properties are the following

1.  $f$  is quasiconvex if and only if for all  $\lambda \in \mathbb{R}$  the sublevel set  $C_\lambda = \{x : f(x) \leq \lambda\}$  is convex (possibly empty);
2. if  $f$  is continuous,  $f$  is strictly quasiconvex if and only if for all  $\lambda \in \mathbb{R}$  the sublevel set  $C_\lambda = \{x : f(x) \leq \lambda\}$  is strictly convex (possibly empty);
3. if  $f$  is continuous and strictly quasiconvex, given  $m \in \mathbb{R}$ , the only sublevel sets  $C_m$  with empty interior are either the sets  $C_m$  with  $m < \inf f$ , which are empty, or the set  $C_m$  with  $m = \inf f$ , which is either empty or a singleton;
4. if  $f$  is continuous and strictly quasiconvex, then the level set  $\{x : f(x) = \lambda\}$  has always empty interior.

### 2.1.2 Some basics about Finsler manifolds

Let us now consider an open and connected subset  $\Omega$  of  $\mathbb{R}^d$ . The definition of (weak) Finsler metric is given as follows (see also Definition 1.7 in [59], (1.1)-(1.5) in [60] or Definition 1.3 in [72]).

**Definition 2.1.10.** A Finsler metric on a differentiable manifold  $M$  is a map  $\psi : TM \rightarrow \mathbb{R}_+$  such that  $\psi(x, \cdot)$  is convex and positively 1-homogeneous on  $T_x M$  for every  $x \in M$ .

Usually the Finsler metric is assumed to be smooth. However it is possible to weaken this hypothesis by defining a (weak) Finsler metric.

**Definition 2.1.11.** A non-negative Borel-measurable function  $\psi : \overline{\Omega} \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  is said to be a (weak) Finsler metric on  $\overline{\Omega}$  if

- (i)  $\psi(x, \cdot)$  is positively 1-homogeneous for every  $x \in \overline{\Omega}$ ;
- (ii)  $\psi(x, \cdot)$  is convex for  $\mathcal{L}^d$ -a.e.  $x \in \overline{\Omega}$ ;
- (iii) for every compact subset  $B$  of  $\overline{\Omega}$ , there exist  $0 < \alpha_B < M_B$ , such that

$$\alpha_B |q| \leq \psi(x, q) \leq M_B |q|, \quad \text{for all } (x, q) \in B \times \mathbb{R}^d.$$

*Remark 2.1.12.* The restriction of  $\psi$  to each tangent space  $T_x M$  gives rise to the point-wise *Minkowski norm*, which differs from that of a usual norm by the fact that in general is not an even function, that is  $\psi(x, q)$  can be different from  $\psi(x, -q)$ .

Finsler metrics generalize the notion of Riemannian ones, which correspond to the case where  $\psi(x, \cdot)$  is the square root of a positive quadratic form,  $\psi(x, q) = \left( \sum_{i,j} a_{ij}(x) q_i q_j \right)^{\frac{1}{2}}$ .

**Definition 2.1.13.** Given a Finsler metric  $\psi$ , we can define the associated *Finslerian length functional*  $L$ :

$$L(\gamma) := \int_0^1 \psi(\gamma(t), \dot{\gamma}(t)) dt,$$

with  $\gamma \in \text{Lip}([0, 1]; \overline{\Omega})$ , where if  $U$  is a connected subset of  $\mathbb{R}^d$

$$\text{Lip}([0, 1]; U) := \{\gamma : [0, 1] \rightarrow U : \gamma \text{ Lipschitz}\}.$$

Notice that  $L$  is well defined. Indeed, the map  $t \mapsto (\gamma(t), \dot{\gamma}(t))$  is Lebesgue measurable on  $[0, 1]$  and  $\psi$  is Borel measurable on  $\overline{\Omega} \times \mathbb{R}^d$ , hence their composition  $\psi(\gamma(t), \dot{\gamma}(t))$  is Lebesgue measurable.

The Finsler length functional induces a distance on  $\Omega$  as follows

**Definition 2.1.14.** Let  $x, y \in \Omega$ , then we define

$$d(x, y) := \inf \left\{ \int_0^1 \psi(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}_\Omega(x, y) \right\}, \quad (2.1.4)$$

where, if  $U$  is a connected subset of  $\mathbb{R}^d$ ,

$$\text{path}_U(x, y) := \{ \gamma \in \text{Lip}([0, 1]; U) : \gamma(0) = x, \gamma(1) = y \}. \quad (2.1.5)$$

The above definition can be extended to all  $x, y \in \overline{\Omega}$ , in the following way:

$$d(x, y) := \inf \left\{ \liminf_{n \rightarrow \infty} d(x_n, y_n) : (x_n), (y_n) \in \Omega^\mathbb{N} \text{ and } x_n \rightarrow x, y_n \rightarrow y \right\}. \quad (2.1.6)$$

**Notation.** When the set  $U$  in the definition of  $\text{path}_U$  is clear from the context, we will simply write  $\text{path}$ .

*Remark 2.1.15.* Thanks to the positive 1-homogeneity of  $\psi$ ,  $L(\gamma)$  does not depend on the chosen parametrization of  $\gamma$ . Indeed if  $\rho : [0, 1] \rightarrow [a, b]$ , with  $a < b$ , is an order preserving Lipschitz diffeomorphism, if  $\tilde{\gamma} : [a, b] \rightarrow \Omega$  is s.t.  $\tilde{\gamma}(\rho(t)) = \gamma(t)$ , then if  $\tau = \rho(t)$

$$\begin{aligned} L(\gamma) &= \int_0^1 \psi(\gamma(t), \dot{\gamma}(t)) dt = \int_0^1 \psi \left( \tilde{\gamma}(\rho(t)), \dot{\tilde{\gamma}}(\rho(t)) \frac{d}{dt} \rho(t) \right) dt \\ &= \int_a^b \psi(\tilde{\gamma}(\tau), \dot{\tilde{\gamma}}(\tau)) d\tau. \end{aligned}$$

This is the reason why in the (2.1.4) one can consider the infimum on  $\text{path}_\Omega$ , defined by (2.1.5), where the curves are defined on a fixed interval (in this case  $[0, 1]$ ).

*Remark 2.1.16.* We observe that with the definition given by (2.1.6), the function  $d : \overline{\Omega} \times \overline{\Omega} \rightarrow [0, +\infty]$  is the *lower semicontinuous envelope* or *relaxed functional* of the function

$$\tilde{d}(x, y) := \begin{cases} d(x, y) & \text{if } x, y \in \Omega, \\ +\infty & \text{otherwise,} \end{cases}$$

and it is, therefore, lower semicontinuous (see for instance Proposition 1.31 and 1.33 in [35] for the proof that the lower semicontinuous envelope can be written as in (2.1.6)).

**Definition 2.1.17.** Given a connected open subset  $U$  of  $\mathbb{R}^d$ , we define the *intrinsic distance* in  $U$  w.r.t. to the Euclidean length by

$$|x - y|_U := \inf \left\{ \int_0^1 |\dot{\gamma}(t)| dt : \gamma \in \text{path}_U(x, y) \right\}.$$

Clearly, if  $U$  is convex  $|\cdot|_U$  coincides with the Euclidean distance.

The above definition can be extended to all  $x, y \in \partial U$ , in the following way:

$$|x - y|_U := \inf \left\{ \liminf_{n \rightarrow \infty} |x_n - y_n|_U : (x_n), (y_n) \in \Omega^\mathbb{N} \text{ and } x_n \rightarrow x, y_n \rightarrow y \right\}.$$

*Remark 2.1.18.* We observe that  $|x - y|_U < \infty$  for every  $x, y \in U$ . Indeed, since  $U$  is connected and open, it is also path-connected. We can choose a continuous curve  $\gamma$  which connects the two points. If it is not Lipschitz then, since the  $\gamma([0, 1])$  is well contained in  $\Omega$ , there exist  $N \in \mathbb{N}$  and  $x_0, \dots, x_N \in \gamma([0, 1])$  such that  $x_0 = x$  and  $x_N = y$  and the piece-wise curve  $\eta$  obtained by gluing the segments  $(1 - t)x_{i-1} + tx_i$  is contained in  $\Omega$ . Clearly  $l(\eta) = \int_0^1 |\dot{\eta}(t)| dt < +\infty$ .

Moreover if  $x, y$  are close enough that the segment connecting them is contained in  $U$ , for instance if there exists  $r > 0$  and  $z \in \Omega$  such that  $B(z, r) \subset\subset \Omega$  and  $x, y \in B(z, r)$ , then  $|x - y|_\Omega = |x - y|$ .

*Remark 2.1.19.* Note that the pseudo-distance defined in Definition 2.1.14 is not a distance in general: as  $\psi(x, q)$  is not necessarily even in  $q$ , the distance  $d$  may be non-symmetric. However, from now on, we will refer to  $d$  as “distance”.

*Remark 2.1.20.* Let us notice the following properties of  $d$ :

1.  $d(x, y) < +\infty$  for every  $x, y \in \Omega$ . Indeed, by Remark 2.1.18 we can choose  $\eta \in \text{path}(x, y)$ , obtaining

$$d(x, y) \leq \int_0^1 \psi(\eta, \dot{\eta}) dt \leq M_B \int_0^1 |\dot{\eta}(t)| dt < +\infty,$$

for some  $M_B$  which depends on some compact set  $B \subset \bar{\Omega}$  such that  $x, y \in B$ , whose existence is insured by property (iii) of Definition (2.1.11).

2. Since the boundary of  $\Omega$  is not required to be regular, one may have  $d_\lambda(\tilde{x}, y) = +\infty$  for some  $\tilde{x} \in \partial\Omega$  and  $y \in \Omega$ : in this case,  $d_\lambda(\tilde{x}, y) = +\infty$  for any  $y \in \Omega$  due to the connectedness of  $\Omega$ ;
3.  $d(x, y) = 0$  if and only if  $x = y$ . Indeed, if  $x = y$  it is enough to take the constant curve  $\eta(t) \equiv x$  for every  $t$ , obtaining  $\psi(\eta(t), \dot{\eta}(t)) = \psi(x, 0) = 0$  for every  $t$ . If  $x \neq y$ , let  $r > 0$  such that  $B(x, r) \subset \Omega$  and  $y \notin B(x, r)$ . Then for every  $\gamma \in \text{path}_\Omega(x, y)$  there exists  $\bar{t} \in (0, 1)$  such that  $\gamma([0, \bar{t})) \subset B(x, r)$  and  $\gamma(\bar{t}) \in \partial B(x, r)$ . Hence we get

$$\int_0^1 \psi(\gamma, \dot{\gamma}) dt \geq \int_0^{\bar{t}} \psi(\gamma, \dot{\gamma}) dt \geq \alpha_{B(x, r)} \int_0^{\bar{t}} |\dot{\gamma}| dt \geq \alpha_{B(x, r)} r,$$

where  $\alpha_{B(x, r)} > 0$  is the one of (iii) of Definition (2.1.11). The proof is concluded by taking the infimum over all  $\gamma \in \text{path}_\Omega(x, y)$ .

4. with this definition of  $d$ ,

$$d(x, y) \leq d(x, z) + d(z, y)$$

for every  $x, y \in \bar{\Omega}$  and  $z \in \Omega$ , but, as shown in the Example 2.1.21 below, it may fail if the interpolating point  $z$  belongs to the boundary. However, as proved in Proposition 2.1.24 if we assume  $\partial\Omega$  to be Lipschitz (see Definition 2.1.22 below) the triangle inequality holds for every  $x, y, z \in \bar{\Omega}$ .

*Example 2.1.21.* Let  $\Omega \subset \mathbb{R}^2$ , defined as  $\Omega := B(0, 1) \setminus \{[0, y) : 0 < y < 1\}$  and  $K(x) \equiv B_1(0)$  for all  $x \in \Omega$ . Then  $\varphi^0(x, q) = |q|$  for any  $q \in \mathbb{R}^d$  and for any  $x \in \Omega$ . Let's take  $x = (x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$ ,  $y = (y_1, y_2) = (-\frac{1}{2}, \frac{1}{2})$  and  $z = (z_1, z_2) = (0, \frac{1}{2})$  then

$$d(x, y) = \inf \left\{ \int_0^1 |\dot{\gamma}(t)| dt : \gamma \in \text{path}(x, y) \right\} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} = \sqrt{2},$$

while

$$d(x, z) + d(z, y) = \frac{1}{2} + \frac{1}{2} = 1.$$

**Definition 2.1.22** (see for instance Definition 12.9 in [91]). The boundary  $\partial\Omega$  of an open set  $\Omega \subset \mathbb{R}^d$  is said to be *locally Lipschitz* if for every point  $x \in \partial\Omega$  there exists a neighborhood  $A$  of  $x$ , local coordinates  $y = (y', y_d) \in \mathbb{R}^{d-1} \times \mathbb{R}$  with origin  $y = (\mathbf{0}, 0)$  at  $x$ , a Lipschitz function  $\phi : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ , with  $\text{Lip}(\phi) = L$  and  $\phi(\mathbf{0}) = 0$ , and  $\rho, h > 0$  such that

$$\begin{aligned} \partial\Omega \cap A &= \{(y', \phi(y')) : y' \in B(\mathbf{0}, \rho)\} \\ \Omega \cap A &= \{(y', y_d) : y' \in B^{d-1}(\mathbf{0}, \rho), h > y_d > \phi(y')\}. \end{aligned}$$

If  $\partial\Omega$  is bounded, then we refer to locally Lipschitz boundaries simply as *Lipschitz*.

**Lemma 2.1.23.** *If  $\Omega$  is bounded and  $\partial\Omega$  is Lipschitz, there exists  $C_\Omega > 0$ , such that*

$$|x - y| \leq |x - y|_\Omega \leq C_\Omega |x - y|, \quad \text{for any } x, y \in \Omega,$$

and in particular  $|x - y|_\Omega$  is always finite.

*Proof.* We want to prove that there exists a constant  $C = C_\Omega$  such that for every  $x, y \in \Omega$  there exists a curve  $\gamma \in \text{path}_\Omega(x, y)$  such that  $\int_0^1 |\dot{\gamma}| dt \leq C|x - y|$ . Since  $\Omega$  is bounded,  $\partial\Omega$  is a compact set. Let  $x_0 \in \partial\Omega$  and let us consider  $L_0, \rho_0, r_0$  and as the ones in Definition 2.1.22. Clearly  $\partial\Omega \subset \bigcup_{x_0 \in \partial\Omega} B(x_0, \frac{r_0}{2})$ . By compactness there exist  $N \in \mathbb{N}$  and  $x_1, \dots, x_N$  such that  $\partial\Omega \subset \bigcup_{i=1, \dots, N} B(x_i, \frac{r_i}{2})$ . We take  $L = \max_{i=1, \dots, N} L_i$  for some  $\varepsilon > 0$  and  $r := \min_{i=1, \dots, N} \frac{r_i}{2}$ .

Let  $x, y \in \Omega$ .

Case 1.: The segment connecting  $x$  and  $y$  is contained in  $\Omega$ . In this case  $|x - y|_\Omega = |x - y|$ .

Case 2.:  $|x - y| < r$  and the segment intersects the boundary. We call  $\tilde{x}$  and  $\tilde{y}$  respectively the closest point to  $x$  and  $y$  given by the intersection of the segment  $[x, y]$  with  $\partial\Omega$ . Let  $i$  such that  $\tilde{x} \in B(x_i, \frac{r_i}{2})$ . Since  $|\tilde{x} - \tilde{y}| < r \leq \frac{r_i}{2}$ , then  $\tilde{x}, \tilde{y} \in B(x_i, r_i)$ . By our choice of  $r_i$ , we have that

$$\partial\Omega \cap B(x_i, r_i) = \{(y', \phi_i(y')) : y' \in B(\mathbf{0}, \rho_i)\}$$

for some  $\phi_i$  such that  $\text{Lip}(\phi_i) = L_i \leq L$ . Therefore one can take  $\tilde{\gamma} \in \text{path}_{\overline{\Omega}}(\tilde{x}, \tilde{y})$  such that  $\tilde{\gamma}([0, 1]) \subset \partial\Omega$ ,  $\tilde{\gamma}(0) = \tilde{x} = (\tilde{x}', \phi_i(\tilde{x}'))$  and  $\tilde{\gamma}(1) = \tilde{y} = (\tilde{y}', \phi_i(\tilde{y}'))$  and for every  $t$ ,  $\tilde{\gamma}(t) = (\tilde{z}', \phi_i(\tilde{z}'))$ , where  $\tilde{z}' = (1-t)\tilde{x}' + t\tilde{y}'$ . Then, if  $\tilde{w}' = (1-s)\tilde{x}' + s\tilde{y}'$ , it holds

$$\begin{aligned} |\tilde{\gamma}(t) - \tilde{\gamma}(s)| &\leq \sqrt{|\tilde{z}' - \tilde{w}'|^2 + |\phi_i(\tilde{z}') - \phi_i(\tilde{w}')|^2} \\ &\leq \sqrt{|\tilde{z}' - \tilde{w}'|^2 + L_i |\tilde{z}' - \tilde{w}'|^2} \\ &\leq (1+L) |\tilde{z}' - \tilde{w}'| \\ &\leq (1+L) |t-s| |\tilde{x}' - \tilde{y}'| \\ &\leq (1+L) |t-s| |\tilde{x} - \tilde{y}|. \end{aligned}$$

Therefore  $\text{Lip}(\tilde{\gamma}) \leq (1+L)|\tilde{x} - \tilde{y}|$ . Now, thanks to Lemma 3.2.6 below, for every  $\varepsilon > 0$ , there exists  $\tilde{\gamma}_\varepsilon$  such that  $\tilde{\gamma}_\varepsilon(0) = \tilde{x}_\varepsilon \in \Omega$ ,  $\tilde{\gamma}_\varepsilon(1) = \tilde{y}_\varepsilon \in \Omega$ ,  $\text{Lip}(\tilde{\gamma}) < \text{Lip}(\tilde{\gamma}_\varepsilon) + \varepsilon \leq (1+L)|\tilde{x} - \tilde{y}| + \varepsilon$ ,  $|\tilde{\gamma}_\varepsilon(0) - \tilde{\gamma}(0)| = |\tilde{x}_\varepsilon - \tilde{x}| < \varepsilon$  and  $|\tilde{\gamma}_\varepsilon(1) - \tilde{\gamma}(1)| < \varepsilon$ . Then, if we consider the curve  $\eta_\varepsilon$  obtained by gluing the segment  $[x, \tilde{x}_\varepsilon]$ ,  $\tilde{\gamma}_\varepsilon$  and  $[y, \tilde{y}_\varepsilon]$ , we have

$$\begin{aligned} |x - y|_\Omega &\leq |x - \tilde{x}_\varepsilon| + \int_0^1 |\dot{\tilde{\gamma}}_\varepsilon| dt + |y - \tilde{y}_\varepsilon| \\ &\leq |x - \tilde{x}| + \varepsilon + (1+L)|\tilde{x} - \tilde{y}| + \varepsilon + |y - \tilde{y}| + \varepsilon. \end{aligned}$$

Since we can choose  $\varepsilon < |x - y|$  and since  $|x - \tilde{x}|, |y - \tilde{y}| \leq |x - y|$  ( $x, \tilde{x}, \tilde{y}, y$  all belong to the same segment), we get the following bound from above

$$|x - y|_\Omega \leq 5|x - y| + (1 + L)|\tilde{x} - \tilde{y}| \leq (6 + L)|x - y|.$$

Case 3.:  $|x - y| \geq r$  and the segment intersects the boundary. Since  $\Omega$  is compact, we can "add balls" to the covering of  $\partial\Omega$  in order to have a covering of  $\bar{\Omega}$ . More precisely, there exists  $M \geq N, x_{N+1}, \dots, x_M \in \Omega$  and  $\frac{r_{N+1}}{2}, \dots, \frac{r_M}{2}$  such that  $\bar{\Omega} \subset \bigcup_{i=1}^M B(x_i, \frac{r_i}{2})$  and  $x_i, \frac{r_i}{2}$  are the same as before for every  $1 \leq i \leq N$ . With an iterative procedure we start from  $x$  which belongs to a ball, let's say (up to a change in the order of the indexes)  $B_1 := B(x_1, \frac{r_1}{2})$ . If  $B_1$  intersects the ball which contains  $y$ , then take  $z_2$  in the intersection of the two balls and call that ball  $B_2$ . If this is not the case we consider another ball, different from  $B_1$ , which intersects  $B_1$  and call it  $B_2$  (such a  $B_2$  always exists by connectedness of  $\Omega$ ). Take a point  $z_2 \in B_2 \cap B_1$ . If  $B_1 \cap \partial\Omega = \emptyset$ , we are in Case 1. and we connect  $x := z_1$  to  $z_2$  with a segment. If  $B_1 \cap \partial\Omega \neq \emptyset$  we connect  $x := z_1$  to  $z_2$  with the curve  $\eta_\varepsilon^1$ , given by Case 2., with  $\varepsilon$  small enough. If  $y \in B_2$  we connect  $z_2$  to  $y$  either with a segment or with some  $\eta_\varepsilon^2$  given by Case 2. and we stop. Otherwise, we consider a ball  $B_3$  different from  $B_1$  and  $B_2$  such that  $B_3 \cap B_2 \neq \emptyset$ , we take  $z_3 \in B_2 \cap B_3$  and we connect  $z_2$  with  $z_3$  either with a segment or with  $\eta_\varepsilon^2$ . We continue like this, trying to avoid, at every step, the balls already intersected. If this is not possible, we come back until we find a ball which has never been considered. By connectedness of  $\Omega$  and since the number of balls is finite, we reach  $y$  in a finite number of steps. If we pass for two times from the same segment or from the same piece  $\eta_\varepsilon^i$  for some  $i$  we erase that piece. We then consider a curve  $\eta$  obtained by gluing all the segments and the  $\eta_\varepsilon^i$  (taken at most one time). So at the end we have that  $\eta$  is composed by at most  $M$  pieces. Thus, if  $R := \max_{i=1, \dots, M} \frac{r_i}{2}$

$$|x - y|_\Omega \leq \int_0^1 |\eta| dt \leq N(6 + L)R \leq \frac{N(6 + L)R}{r} |x - y|.$$

□

**Proposition 2.1.24.** *The distance function  $d : \bar{\Omega} \times \bar{\Omega} \rightarrow [0, +\infty]$  is lower semicontinuous and  $d : \Omega \times \Omega \rightarrow \mathbb{R}_+$  is continuous.*

*If  $\partial\Omega$  is Lipschitz and  $\Omega$  is bounded, the distance function  $d : \bar{\Omega} \times \bar{\Omega} \rightarrow \mathbb{R}_+$  is continuous and the triangle inequality holds for every  $x, y, z \in \bar{\Omega}$ .*

*Proof.* By Remark 2.1.16 we already know that the function  $d : \bar{\Omega} \times \bar{\Omega} \rightarrow [0, +\infty]$  is lower semicontinuous. Thus, in order to prove the continuity of  $d : \Omega \times \Omega \rightarrow \mathbb{R}_+$  it is sufficient to prove the upper semicontinuity. Let  $x, y \in \Omega$  and  $r > 0$  such that  $B(x, r), B(y, r) \subset \subset \Omega$  and set  $B := \overline{B(x, r)} \cup \overline{B(y, r)}$ . If  $(x_n)_n, (y_n)_n \subset \Omega$  are two sequences converging respectively to  $x$  and  $y$ , we can assume without loss of generality that  $(x_n)_n \subset B(x, r)$  and  $(y_n)_n \subset B(y, r)$ . Then if  $\gamma \in \text{path}_{B(x, r)}(x_n, x)$  we have

$$d(x_n, x) \leq \int_0^1 \psi(\gamma, \dot{\gamma}) dt \leq M_B \int_0^1 |\dot{\gamma}(t)| dt$$

and analogously if  $\gamma \in \text{path}_{B(y, r)}(y_n, y)$

$$d(y_n, y) \leq M_B \int_0^1 |\dot{\gamma}(t)| dt,$$

where  $M_B$  is given by property (iii) of Definition 2.1.11. Thus, passing to the infimum over  $\text{path}_{B(x,r)}(x_n, x)$  and  $\text{path}_{B(y,r)}(y_n, y)$  we obtain

$$\begin{aligned} d(x_n, x) &\leq M_B|x_n - x|_{B(x,r)} = M_B|x_n - x| \quad \text{and} \\ d(y_n, y) &\leq M_B|y_n - y|_{B(x,r)} = M_B|y_n - y|, \end{aligned}$$

which implies

$$\limsup_{n \rightarrow +\infty} d(x_n, y_n) \leq \limsup_{n \rightarrow +\infty} d(x_n, x) + d(x, y) + \limsup_{n \rightarrow +\infty} d(y_n, y) = d(x, y).$$

Finally, if  $\partial\Omega$  is Lipschitz, thanks to Lemma 2.1.23, there exists  $C_\Omega > 0$ , such that

$$|x - y| \leq |x - y|_\Omega \leq C_\Omega|x - y|, \quad \text{for any } x, y \in \bar{\Omega}, \quad (2.1.7)$$

and in particular  $|x - y|_\Omega$  is always finite. The (2.1.7) allows us to extend  $d$  continuously to the boundary, defining, for every  $x, y \in \partial\Omega$ :

$$d(x, y) := \lim_{n \rightarrow \infty} d(x_n, y_n), \quad \text{for some } x_n \rightarrow x, y_n \rightarrow y \text{ and } (x_n)_n, (y_n)_n \subset \Omega.$$

This definition is well posed and equivalent to the one given by (2.1.6), indeed, if  $(x_n)_n, (x'_n)_n \subset \Omega$  are two sequences converging to  $x$  and  $(y_n)_n, (y'_n)_n \subset \Omega$  are two sequences converging to  $y$ , by using the triangular inequality for points in  $\Omega$  it holds

$$\begin{aligned} |d(x_n, y_n) - d(x'_n, y'_n)| &= |d(x_n, y_n) - d(x'_n, y_n) + d(x'_n, y_n) - d(x'_n, y'_n)| \\ &\leq \max\{d(x_n, x'_n), d(x'_n, x_n)\} + \max\{d(y_n, y'_n), d(y'_n, y_n)\} \\ &\leq M_B|x_n - x'_n|_\Omega + M_B|y_n - y'_n|_\Omega \end{aligned} \quad (2.1.8)$$

$$\leq C_\Omega M_B|x_n - x'_n| + C_\Omega M_B|y_n - y'_n| \rightarrow 0, \quad (2.1.9)$$

where, as above, for any  $n$  large enough,  $B \subset \Omega$  is a compact set such that  $x_n, x'_n, y_n, y'_n, x, y \in B$ .  $\square$

**Proposition 2.1.25.** *Let  $V \subset \Omega$  such that  $V$  is bounded. Then there exist two constants  $0 < \alpha_V < M_V$  such that*

$$\alpha_V|x - y|_V \leq d^V(x, y) \leq M_V|x - y|_V, \quad \text{for every } x, y \in V.$$

*If  $V \subset\subset \Omega$ , there exist a compact set  $B \subset \bar{\Omega}$  and two constant  $0 < \alpha_B < M_B$  such that*

$$d(x, y) \leq M_B|x - y|_\Omega \quad \text{for every } x, y \in V$$

*and, for any fixed  $x \in V$ ,*

$$d(x, y) \geq \alpha_B|x - y|_\Omega, \quad \text{for every } y \text{ s.t. } |x - y| < \frac{\alpha}{M} \text{dist}(x, \partial V).$$

*Proof.* Let  $x, y \in V \subset \Omega$  and  $V$  bounded. Thanks to property (iii) in Definition 2.1.11 of the Finsler metric  $\psi$ , there exist  $0 < \alpha_V < M_V$  such that

$$\alpha_V|q| \leq \psi(z, q) \leq M_V|q|, \quad \text{for all } z \in V.$$

Hence, we have that for any  $\gamma \in \mathbf{path}_V(x, y)$

$$\alpha_{\overline{V}} \int_0^1 |\dot{\gamma}(t)| dt \leq \int_0^1 \psi(\gamma(t), \dot{\gamma}(t)) dt \leq M_{\overline{V}} \int_0^1 |\dot{\gamma}(t)| dt$$

and then

$$0 \leq \alpha_{\overline{V}} |x - y|_V \leq d^V(x, y) \leq M_{\overline{V}} |x - y|_V.$$

If  $V \subset\subset \Omega$  we consider  $A \subset \Omega$  bounded, such that

$$|x - y|_\Omega = \inf \left\{ \int_0^1 |\dot{\gamma}(t)| dt : \gamma \in \mathbf{path}_\Omega(x, y) \right\} = \inf \left\{ \int_0^1 |\dot{\gamma}(t)| dt : \gamma \in \mathbf{path}_A(x, y) \right\},$$

for any  $x, y \in V$ . We stress that such set  $A$  exists. Indeed, by continuity  $|x - y|_\Omega < C$ , for some  $C \geq 0$ , for any  $x, y \in V$ . Let  $x, y \in V$ . If  $\gamma \in \mathbf{path}_\Omega(x, y)$  such that  $l(\gamma) < |x - y|_\Omega + 1 < C + 1$ , then  $\gamma([0, 1]) \subset B(x, 2(C + 1)) \cap \Omega$ , otherwise  $l(\gamma) > 2(C + 1)$ . Thus we consider  $A = V + B(0, 2(C + 1))$ . If  $V + B(0, 2(C + 1))$  is not connected we consider the connected component containing  $V$ . Then we consider the set  $\overline{A}$  (the closure of  $A$ ). Again, thanks to property (iii) in Definition 2.1.11 of the Finsler metric  $\psi$ , there exist  $0 < \alpha_{\overline{A}} < M_{\overline{A}}$  such that

$$\alpha_{\overline{A}} |q| \leq \psi(z, q) \leq M_{\overline{A}} |q|, \quad \text{for all } z \in \overline{A}.$$

Hence, we have that for any  $\gamma \in \mathbf{path}_A(x, y)$

$$\alpha_{\overline{A}} \int_0^1 |\dot{\gamma}(t)| dt \leq \int_0^1 \psi(\gamma(t), \dot{\gamma}(t)) dt \leq M_{\overline{A}} \int_0^1 |\dot{\gamma}(t)| dt \quad (2.1.10)$$

and then

$$d(x, y) \leq M_{\overline{A}} |x - y|_\Omega, \quad \text{for every } x, y \in V. \quad (2.1.11)$$

Finally, let  $x \in V$  and  $y$  such that  $|x - y| \leq \frac{\alpha_{\overline{A}}}{M_{\overline{A}}} \text{dist}(x, \partial V)$ . Then, from the inequality above, we infer that

$$d(x, y) \leq M_{\overline{A}} |x - y|_\Omega = M_{\overline{A}} |x - y| < \alpha_{\overline{A}} \text{dist}(x, \partial V) \leq \alpha_{\overline{A}} \text{dist}(x, \partial A),$$

where the equality is due to the fact that by definition the segment which joins  $x$  and  $y$  is contained in  $\Omega$ . Possibly modifying  $A$  we consider  $\varepsilon > 0$  such that  $\text{dist}(x, \partial A) > \text{dist}(x, \partial V) + \varepsilon$ . We claim that  $d(x, y) = d^A(x, y)$ . Let  $\gamma \in \mathbf{path}_\Omega(x, y)$  and let us assume that there exists  $s \in (0, 1)$  such that  $|\gamma(s) - x| \geq \text{dist}(x, \partial A)$ . We consider  $\bar{s} = \inf \{s : |\gamma(s) - x| \geq \text{dist}(x, \partial A)\}$ . Then

$$\begin{aligned} \int_0^1 \psi(\gamma, \dot{\gamma}) dt &\geq \int_0^{\bar{s}} \psi(\gamma, \dot{\gamma}) dt \geq \alpha_{\overline{A}} \int_0^{\bar{s}} |\dot{\gamma}| dt \\ &\geq \alpha_{\overline{A}} |\gamma(s) - x| \geq \alpha_{\overline{A}} \text{dist}(x, \partial A) \\ &> \alpha_{\overline{A}} \text{dist}(x, \partial V) + \varepsilon > d(x, y) + \varepsilon. \end{aligned}$$

Therefore  $d(x, y) = d^A(x, y)$ . We conclude the last part of the proof observing that (2.1.10) implies

$$\alpha_{\overline{A}} |x - y|_\Omega = \alpha_{\overline{A}} |x - y|_A \leq d^A(x, y) = d(x, y).$$

□

**Corollary 2.1.26.** *If property (iii) is global, i.e. there exist  $0 < \alpha < M$  such that*

$$\alpha|q| \leq \psi(x, q) \leq M|q|, \quad \text{for all } (x, q) \in \bar{\Omega} \times \mathbb{R}^d,$$

*then  $d : \bar{\Omega} \times \bar{\Omega} \rightarrow \mathbb{R}_+$  is equivalent to the intrinsic distance, i.e.*

$$\alpha|x - y|_\Omega \leq d(x, y) \leq M|x - y|_\Omega, \quad \text{for all } x, y \in \bar{\Omega}, \quad (2.1.12)$$

*and if  $\partial\Omega$  is Lipschitz,  $d$  is equivalent to the Euclidean distance.*

**Proposition 2.1.27.** *The space  $(\Omega, d)$ , where  $d$  is the (pseudo) distance defined in Definition 2.1.14 is a length space, i.e. a metric space such that for every  $x, y \in \Omega$ , it holds*

$$d(x, y) = \inf\{L_d(\gamma) : \gamma \in \text{path}_\Omega(x, y)\}, \quad (2.1.13)$$

where

$$L_d(\gamma) := \sup \left\{ \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) : 0 = t_0 < t_1 < \dots < t_n = 1 \right\},$$

is the  $d$ -length of  $\gamma$ .

*Proof.* For all  $x, y \in \Omega$ , we define the intrinsic distance  $d'$  associated to  $d$ , as follows

$$d'(x, y) := \inf\{L_d(\gamma) : \gamma \in \text{path}_\Omega(x, y)\},$$

By definition  $d \leq d'$ . We want to prove that  $d' \leq d$ .

Let's take  $\gamma \in \text{path}$ , we show that

$$L_d(\gamma) \leq \int_0^1 \psi(\gamma(t), \dot{\gamma}(t)) dt.$$

Let  $0 = t_0 < t_1 < \dots < t_n = 1$  be a partition of  $[0, 1]$  and let  $\gamma_i \in \text{path}(\gamma(t_i), \gamma(t_{i+1}))$  be defined by

$$\gamma_i(t) := \gamma(t), \quad t \in [t_i, t_{i+1}].$$

By definition of  $d$  we have

$$\sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \leq \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \psi(\gamma_i(t), \dot{\gamma}_i(t)) dt = \int_0^1 \psi(\gamma(t), \dot{\gamma}(t)) dt. \quad (2.1.14)$$

□

**Corollary 2.1.28.** *If  $\Omega$  is bounded and property (2.1.2) of Definition 2.1.11 is global, i.e. there exist  $0 < \alpha < M$  such that*

$$\alpha|q| \leq \psi(x, q) \leq M|q|, \quad \text{for all } (x, q) \in \bar{\Omega} \times \mathbb{R}^d,$$

*for every  $x, y \in \Omega$  there exists  $\gamma \in \text{path}_{\bar{\Omega}}(x, y)$  such that*

$$d(x, y) = L_d(\gamma).$$

*We will call such a  $\gamma$ , curve of minimal  $d$ -length.*

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

*Proof.* Let  $x, y \in \Omega$ . First of all we notice that  $L_d(\gamma)$  is lower semicontinuous with respect to the uniform convergence, indeed it is the supremum of a family of continuous functions ( $d$  is continuous thanks to Proposition 2.1.24).

We also observe that for every  $\gamma \in \text{path}_\Omega(x, y)$  and for every partition  $t_0, \dots, t_N$  such that  $|\gamma(t_{i-1}) - \gamma(t_i)| = r_k$  and  $B(\gamma(t_i), r_k) \subset \subset \Omega$  for every  $i$ ,

$$\sum_{i=1}^N d(\gamma(t_{i-1}), \gamma(t_i)) \geq \alpha \sum_{i=1}^N |\gamma(t_{i-1}) - \gamma(t_i)|_\Omega = \alpha \sum_{i=1}^N |\gamma(t_{i-1}) - \gamma(t_i)|,$$

where the inequality is due to Corollary 2.1.26 and the equality to the fact that with our partition  $|\gamma(t_{i-1}) - \gamma(t_i)|_\Omega = |\gamma(t_{i-1}) - \gamma(t_i)|$  (see Remark 2.1.18). Hence, by passing to the supremum we obtain  $L_d(\gamma) \geq \alpha \int_0^1 |\dot{\gamma}(t)| dt$ .

Let

$$\ell := d(x, y) = \inf\{L_d(\gamma) : \gamma \in \text{path}_\Omega(x, y)\}$$

and consider a sequence of curves  $(\gamma_n)$  such that  $\ell_n := L_d(\gamma_n)$  converges to  $\ell$ . Without loss of generality we can assume that  $L_d(\gamma_n) \leq \ell + 1$  for all  $n$ . For the previous observation we can thus reparametrize the curves  $\gamma_n : [0, 1] \rightarrow \Omega$  such that  $|\dot{\gamma}_n| = \int_0^1 |\dot{\gamma}_n(t)| dt \leq \frac{\ell+1}{\alpha}$ . Since  $\Omega$  is bounded we also have that  $|\gamma_n| \leq C$  for every  $n$ , for some  $C \geq 0$ , and we can apply Arzelà-Ascoli theorem and find a subsequence, that we will still call  $(\gamma_n)$ , such that uniformly converges to a curve  $\gamma$ . Clearly  $\gamma$  belongs to  $\text{path}_{\overline{\Omega}}(x, y)$ . Moreover

$$\ell \leq L_d(\gamma) \leq \liminf \ell_n = \ell.$$

So we have that

$$d(x, y) = d'(x, y) = L_d(\gamma).$$

□

## 2.2 Some properties of absolute minimizers of supremal functionals

For the reader's convenience and to fix the notation, we recall the following definition.

**Definition 2.2.1.** Given a positive measure  $\mu$  on a space  $X$ , i.e.  $\mu \in \mathcal{M}(X)$  and a Borel function  $f : X \rightarrow \overline{\mathbb{R}}$ , the essential supremum of  $f$  with respect to  $\mu$  is defined by

$$\mu - \text{ess sup}_{x \in X} f(x) := \inf\{t \in \mathbb{R} : \mu(f^{-1}(t, +\infty)) = 0\}.$$

If  $X = \Omega \subset \mathbb{R}^d$  and  $\mu = \mathcal{L}^d$  we simply write  $\text{ess sup}_{x \in \Omega} f(x)$ .

*Remark 2.2.2.* If  $f : X \rightarrow \overline{\mathbb{R}}$  is lower semicontinuous, then

$$\mu - \text{ess sup}_{x \in X} f(x) = \sup\{f(x) : x \in \text{supp } \mu\}.$$

Let  $\Omega \subset \mathbb{R}^d$  an open set and  $H : \Omega \times \mathbb{R}^n \times \mathbb{R}^{dn} \rightarrow \overline{\mathbb{R}}$  a Borel function,  $H \geq 0$ , a *supremal variational problem* or *variational problem in  $L^\infty$*  is the problem of minimizing

$$\mathcal{F}_\infty(u, \Omega) := \|H(x, u(x), Du(x))\|_{L^\infty(\Omega)} = \text{ess sup}_{x \in \Omega} H(x, u(x), Du(x)),$$

among a suitable class of functions  $u$ . The function  $H$  is called *supremand*.

### 2.2.1 The supremand $H$

In this chapter we will consider  $\Omega$  to be connected and bounded and we will be interested in the following problem:

$$\min \left\{ \mathcal{F}_H(v, \Omega) := \operatorname{ess\,sup}_{x \in \Omega} H(x, Dv(x)) : v \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega}), v = g \text{ on } \partial\Omega \right\}, \quad (\text{H})$$

where  $H : \bar{\Omega} \times \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$  is a Borel function,  $H \geq 0$ , and the boundary datum  $g$  belongs to  $W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ .

We recall that the set  $W^{1,\infty}(\Omega)$  coincides with the set of locally Lipschitz continuous functions on  $\Omega$ , i.e. the set of functions  $u : \Omega \rightarrow \mathbb{R}$  such that for every  $x \in \Omega$  there exists a neighborhood  $A \subset \Omega$  of  $x$  such that for every  $z, y \in A$ ,  $|u(z) - u(y)| \leq L|z - y|$ , for some  $L \geq 0$  (see for instance Proposition 9.3 in [28]). Since the Lipschitz continuity holds only locally, it might be not possible to extend  $u \in W^{1,\infty}(\Omega)$  continuously up to the boundary, that is why we consider the functions  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ .

**Proposition 2.2.3.** *If  $H$  is quasiconvex, then  $\mathcal{F}_H(\cdot, \Omega)$  is quasiconvex.*

*Proof.* Let  $u, v \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ ,  $u = v = g$  on  $\partial\Omega$ , then for a.e.  $x \in \Omega$

$$\begin{aligned} H(x, (1-t)Du + tDv) &\leq \max\{H(x, Du(x)), H(x, Dv(x))\} \\ &\leq \max\{\operatorname{ess\,sup}_{x \in \Omega} H(x, Du(x)), \operatorname{ess\,sup}_{x \in \Omega} H(x, Dv(x))\} \\ &= \max\{\mathcal{F}_H(u, \Omega), \mathcal{F}_H(v, \Omega)\}. \end{aligned}$$

Thus

$$\mathcal{F}_H((1-t)u + tv, \Omega) \leq \max\{\mathcal{F}_H(u, \Omega), \mathcal{F}_H(v, \Omega)\}.$$

□

**Theorem 2.2.4.** *(Theorem 2.9 and 3.4 in [16]) If  $H(x, \cdot) : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$  is lower semicontinuous and quasiconvex for a.e.  $x \in \Omega$ , then  $\mathcal{F}_H(\cdot, \Omega)$  is sequentially weak\* lower semicontinuous on  $W^{1,\infty}(\Omega)$ . Moreover if  $H(x, \cdot)$  is coercive uniformly w.r.t.  $x$  and  $H(\cdot, p)$  is uniformly continuous on  $\Omega$  locally uniformly w.r.t.  $p$ , then problem (H) admits a solution.*

**Notation.** We will write  $\{H(x, \cdot) \leq \lambda\} := \{p : H(x, p) \leq \lambda\}$ .

Throughout this chapter we will assume that the function  $H : \bar{\Omega} \times \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$  satisfies the following assumptions:

- (A)  $H \geq 0$ ,  $H(\cdot, 0) = 0$  and  $H(x, \cdot)$  is quasiconvex;
- (B) the map  $(x, p) \mapsto H(x, p)$  satisfies the following coercivity assumptions in  $p$ , uniformly with respect to  $x$ : for every  $\lambda \geq 0$  there exists  $M \geq 0$  such that

$$H(x, p) \leq \lambda \implies |p| \leq M;$$

- (C) the map  $(x, p) \mapsto H(x, p)$  is continuous in  $\Omega \times \mathbb{R}^d$ ;
- (D) For all  $\lambda > \mu \geq 0$  there exists  $\alpha > 0$  such that

$$\{H(x, \cdot) \leq \mu\} + B(0, \alpha) \subset \{H(x, \cdot) \leq \lambda\},$$

for every  $x \in \bar{\Omega}$ ;

2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

---

(E) For all  $\beta > 0$ ,  $\lambda \geq 0$  and  $V \subset\subset \Omega$ , there exists  $\delta$  such that, for any  $\bar{\lambda} \geq 0$  with  $|\lambda - \bar{\lambda}| < \delta$

$$\{H(x, \cdot) \leq \lambda\} \subset \{H(x, \cdot) \leq \bar{\lambda}\} + B(0, \beta),$$

for every  $x \in V$ .

*Remark 2.2.5.* We observe that:

- (i) by continuity (and thus lower semicontinuity) and coercivity (assumption (B)) of  $p \mapsto H(x, p)$ , the sublevel sets  $\{p : H(x, p)\}$  are compact;
- (ii) the continuity of  $H$ , given by the property (C), allows for the local uniform continuity. This means that for every  $x_0 \in \Omega$  and for every  $K, r > 0$ , there exists a non decreasing function  $\omega : [0, +\infty) \rightarrow [0, +\infty)$ , such that  $\lim_{t \rightarrow 0} \omega(t) = 0$

$$|H(x, p) - H(y, p')| \leq \omega(|x - y| + |p - p'|),$$

for all  $(x, p), (y, p') \in B(x_0, r) \times B(0, K)$ ;

- (iii) by property (D), since  $0 \in \{H(x, \cdot) \leq 0\}$ , for any  $\lambda > 0$  there exists  $\alpha > 0$  such that

$$B(0, \alpha) \subset \{H(x, \cdot) \leq \lambda\}, \text{ for any } x \in \bar{\Omega};$$

- (iv) property (E) implies that the interior part of the level set  $\{H(x, p) = \lambda\}$  is empty for every  $\lambda \geq 0$ . Indeed, let us assume by contradiction that there exists  $\lambda \geq 0$  and  $p \in \{H(x, \cdot) = \lambda\}$  such that  $B(p, r) \subset \{H(x, \cdot) = \lambda\}$ , for some  $r > 0$ . Let  $\beta = \frac{r}{2}$  and, for every  $n$ , let  $\lambda_n = \lambda - \frac{1}{n}$ . We claim that

$$\{H(x, \cdot) \leq \lambda\} \not\subset \{H(x, \cdot) \leq \lambda_n\} + B(0, \beta)$$

for every  $n$ . Since  $\lambda > \lambda_n$  for every  $n$ , it is enough to prove that

$$\{H(x, \cdot) \leq \lambda\} \not\subset \{H(x, \cdot) < \lambda\} + B(0, \beta).$$

If  $p \in \partial\{H(x, \cdot) < \lambda\}$  there exists a unitary vector  $w$  such that the point  $z = p + \frac{3}{4}rw$  doesn't belong to  $\{H(x, \cdot) < \lambda\} + B(0, \beta)$ . If  $\text{dist}(p, \{H(x, \cdot) < \lambda\}) > 0$  the existence of such  $w$  is a fortiori true.

*Remark 2.2.6.*  $\{H(x, \cdot) = \lambda\} = \partial\{H(x, \cdot) \leq \lambda\}$  and  $\{H(x, \cdot) < \lambda\} = \text{int}\{H(x, \cdot) \leq \lambda\}$ .

*Proof.* Let  $p \in \partial\{H(x, \cdot) \leq \lambda\}$ , then, by continuity of  $H$ ,  $p \in \{H(x, \cdot) = \lambda\}$ . We now prove the inclusion  $\{H(x, \cdot) = \lambda\} \subset \partial\{H(x, \cdot) \leq \lambda\}$ . Let us assume by contradiction that there exists  $p \in \{H(x, \cdot) = \lambda\}$  such that  $B(p, r) \subset \{H(x, \cdot) \leq \lambda\}$  for some  $r > 0$ . By property (E) we have that there exists  $p_1 \in B(p, r)$  such that  $H(x, p_1) < \lambda$ . By continuity there exists  $r_1$  such that  $H(x, q) < \lambda$  for any  $q \in B(p_1, r_1)$ . We choose  $r_1$  small enough such that  $B(p_1, r_1) \subset\subset B(p, r)$ . Let  $p_2$  such that  $p = \frac{1}{2}(p_1 + p_2)$ . By symmetry of  $B(p, r)$ ,  $B(p_2, r_1) \subset B(p, r)$ . Again by property (E) we have that there exists  $q_2 \in B(p_2, r_1)$  such that  $H(x, q_2) < \lambda$ . Let  $q_1 \in B(p_1, r_1)$  such that  $p = \frac{1}{2}(q_1 + q_2)$ . Then  $q_1, q_2 \in \{H(x, \cdot) \leq \lambda'\}$  for some  $\lambda' < \lambda$ , which implies, by convexity,  $p \in \{H(x, \cdot) \leq \lambda'\}$ , leading us to a contradiction since  $H(x, p) = \lambda > \lambda'$ .

Finally, since  $\{H(x, \cdot) \leq \lambda\}$  is closed,

$$\{H(x, \cdot) = \lambda\} = \partial\{H(x, \cdot) \leq \lambda\} = \{H(x, \cdot) \leq \lambda\} \setminus \text{int}\{H(x, \cdot) \leq \lambda\}$$

and therefore  $\text{int}\{H(x, \cdot) \leq \lambda\} = \{H(x, \cdot) < \lambda\}$ . □

*Example 2.2.7.* The functional  $H : \overline{\Omega} \times \mathbb{R} \rightarrow \overline{\mathbb{R}}$ , defined by

$$H(x, p) := \begin{cases} \frac{|p|}{\text{dist}(x, \partial\Omega)}, & \text{if } x \in \Omega. \\ +\infty & \text{if } x \in \partial\Omega, \end{cases}$$

satisfies all the properties (A)-(E), but the property (D). However it satisfies a local version of that property, that is for all  $\lambda > \mu \geq 0$  and for every  $V \subset\subset \Omega$  there exists  $\alpha_V > 0$  such that

$$\{H(x, \cdot) \leq \mu\} + B(0, \alpha_V) \subset \{H(x, \cdot) \leq \lambda\},$$

for every  $x \in V$ .

**Proposition 2.2.8.** *Let  $\lambda > 0$  and  $x \in \Omega$ . Then for every sequence  $(x_n)_n$  converging to  $x$*

$$\{H(x_n, \cdot) \leq \lambda\} \xrightarrow{H} \{H(x, \cdot) \leq \lambda\}, \quad \text{as } n \rightarrow \infty,$$

where with  $\xrightarrow{H}$  we mean convergence with respect to the Hausdorff distance<sup>1</sup>  $d_{\mathcal{H}}$ .

*Proof.* First of all we observe that since  $\{H(x_n, \cdot) \leq \lambda\} \subset B(0, M)$  by assumption (B) of  $H$ , by compactness of the metric  $d_{\mathcal{H}}$  we can extract a converging subsequence, that we will still call  $\{H(x_n, \cdot) \leq \lambda\}$ , to a certain compact set  $K$ . Therefore for any  $p \in K$  there exists  $p_n \in \{H(x_n, \cdot) \leq \lambda\}$  such that  $p_n \rightarrow p$  and then by continuity of  $H$ ,  $p \in \{H(x, \cdot) \leq \lambda\}$ . Let us assume by contradiction that there exists  $p \in \{H(x, \cdot) \leq \lambda\}$  such that  $p \notin K$ . Let  $\alpha = \text{dist}(p, K)$  and  $n_\alpha$  such that  $\{H(x_n, \cdot) \leq \lambda\} \subset K + B(0, \frac{\alpha}{2})$  for any  $n \geq n_\alpha$ , then  $p \in (K + B(0, \frac{\alpha}{2}))^c \subset \cap_{n \geq n_\alpha} \{H(x_n, \cdot) \leq \lambda\}$ . We now take  $r > 0$  such that  $B(p, r) \subset (K + B(0, \frac{\alpha}{2}))^c$  and we consider  $z \in B(p, r) \cap \text{int} \{H(x, \cdot) \leq \lambda\}$ . Notice that the intersection is not empty by property (D), which insure for  $\lambda > 0$  that  $\text{int} \{H(x, \cdot) \leq \lambda\} \neq \emptyset$ , and by convexity of the sublevel sets. Then  $H(x_n, z) > \lambda$  for every  $n \geq n_\alpha$ , but also  $H(x, z) < \lambda$  (by Remark 2.2.6,  $H(x, z)$  has to be strictly smaller than  $\lambda$ ) contradicting the continuity of  $H$ .  $\square$

## 2.2.2 The quasiconvex conjugate of $H$

There exists a generalization of convex conjugate of a convex function for quasiconvex functions (see for instance in [43]).

**Definition 2.2.9.** For every  $x \in \overline{\Omega}$  and  $\lambda \geq 0$ , we define  $L_\lambda(x, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$  by

$$L_\lambda(x, q) := \sup\{p \cdot q : p \in \mathbb{R}^d, H(x, p) \leq \lambda\}. \quad (2.2.1)$$

*Remark 2.2.10.* For every  $x \in \overline{\Omega}$  and  $\lambda \geq 0$  the function  $L_\lambda(x, \cdot)$  is the support function of the convex body  $\{H(x, \cdot) \leq \lambda\}$  defined in Definition 2.1.3, therefore

- $L_\lambda(x, \cdot) \geq 0$  and  $L_\lambda(x, q) = 0$  if and only if  $q = 0$ ;
- $L_\lambda(x, \cdot)$  is positively 1-homogeneous for every  $x \in \Omega$ ;
- $L_\lambda(x, \cdot)$  is convex for every  $x \in \Omega$ ;

Moreover by compactness there exists  $p \in \{H(x, \cdot) \leq \lambda\}$  such that  $L_\lambda(x, q) = p \cdot q$ . By convexity  $p \in \partial\{H(x, \cdot) \leq \lambda\}$ .

<sup>1</sup>For more details about Hausdorff distance see e.g. [74].

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

**Proposition 2.2.11.** *For every fixed  $\lambda \geq 0$ ,  $L_\lambda : \bar{\Omega} \times \mathbb{R}^d \rightarrow \mathbb{R}$  is a Finsler metric (see Definition 2.1.11).*

*Proof.* As pointed out in Remark 2.2.10,  $L_\lambda(x, \cdot)$  is non-negative, positively 1-homogeneous and convex for every  $x \in \bar{\Omega}$ . Moreover, since property (D) and (B) imply e that for every fixed  $\lambda$  there exist  $0 < \alpha < M$  such that for every  $x$

$$B(0, \alpha) \subset \{H(x, \cdot) \leq \lambda\} \subset B(0, M),$$

and thus

$$\sup\{p \cdot q : p \in B(0, \alpha)\} \leq \sup\{p \cdot q : p \in \mathbb{R}^d, H(x, p) \leq \lambda\} \leq \sup\{p \cdot q : p \in B(0, M)\},$$

we have that

$$\alpha|q| \leq L_\lambda(x, q) \leq M|q|,$$

for every  $x$  in  $\Omega$ , proving property (iii). Finally, the Borel measurability is a direct consequence of the upper semicontinuity showed in Lemma 2.2.12 below.  $\square$

**Lemma 2.2.12.** *The function  $(\lambda, x, q) \mapsto L_\lambda(x, q)$  is upper semicontinuous.*

*Proof.* Let  $(\lambda_n, x_n, q_n)_n$  a sequence converging to  $(\lambda, x, q)$ . Without loss of generality we can assume that  $\lim_{n \rightarrow \infty} L_{\lambda_n}(x_n, q_n) = \limsup_{n \rightarrow \infty} L_{\lambda_n}(x_n, q_n)$ . For every  $n$  we take  $p_n$  such that  $L_\lambda(x_n, q_n) = p_n \cdot q_n$  (see Remark 2.2.10). Since  $\lambda_n \leq \lambda + \varepsilon$  for  $n$  large enough, by the coercivity assumption (B) on  $H$ , it is possible to choose a subsequence of  $(p_n)_n$  converging to a certain  $p$ , which we will still call  $(p_n)_n$ . By continuity of  $(x, p) \rightarrow H(x, p)$ , we have that

$$H(x, p) = \lim_n H(x_n, p_n) \leq \lim_n \lambda_n = \lambda.$$

Therefore

$$\limsup_n L_{\lambda_n}(x_n, q_n) = \lim_n L_{\lambda_n}(x_n, q_n) = \lim_n p_n \cdot q_n = p \cdot q \leq L_\lambda(x, q).$$

$\square$

**Lemma 2.2.13.** *The map  $\lambda \mapsto L_\lambda(x, q)$  is lower semicontinuous for any  $x \in \Omega$  and  $q \in \mathbb{R}^d$ .*

*Proof.* Let  $\lambda \geq 0$  and  $(\lambda_n)_n$  converging to  $\lambda$ . Then by assumption (E), for any  $\beta > 0$  we have  $\{H(x, \cdot) \leq \lambda\} \subset \{H(x, \cdot) \leq \lambda_n\} + B(0, \beta)$  for  $n$  large enough. Let  $p \in \{H(x, \cdot) \leq \lambda\}$  such that  $L_\lambda(x, q) = p \cdot q$ , then  $p = p_n + w_n$ , with  $p_n \in \{H(x, \cdot) \leq \lambda_n\}$  and  $w_n \in B(0, \beta)$ . Thus

$$p \cdot q \leq \liminf_n p_n \cdot q + \beta|q| \leq \liminf_n L_{\lambda_n}(x, q) + \beta|q|.$$

The thesis follows by the arbitrariness of  $\beta$ .  $\square$

**Proposition 2.2.14.** *For any  $\lambda \geq 0$ , the unitary ball of  $L_\lambda(x, \cdot)$  is the polar set of the convex sublevel set  $\{H(x, \cdot) \leq \lambda\}$  of  $H$ , i.e. it holds*

$$p \in \{H(x, \cdot) \leq \lambda\} \iff \sup\{p \cdot q : L_\lambda(x, q) \leq 1\} \leq 1. \quad (2.2.2)$$

*Proof.* The thesis follows directly by Proposition 2.1.7. Indeed for any  $x \in \Omega$  and  $\lambda \geq 0$ , the set  $\{L_\lambda(x, \cdot) \leq 1\}$  is the polar set of the convex body  $\{H(x, \cdot) \leq \lambda\}$  (see Remark 2.1.6).  $\square$

**Lemma 2.2.15.** *Let  $\lambda \geq 0$ , then the map  $(x, q) \mapsto L_\lambda(x, q)$  is continuous in  $\Omega \times \mathbb{R}^d$ .*

*Proof.* In light of Lemma 2.2.12 it is enough to show the lower semicontinuity of the map. Let us fix  $(x, q)$  and consider a converging sequence  $((x_n, q_n))_n$ . Let  $p \in \partial\{H(x, \cdot) \leq \lambda\}$  such that  $L_\lambda(x, q) = p \cdot q$  (see Remark 2.2.10) and take  $0 < t < 1$ . By convexity  $tp \in \text{int}\{H(x, \cdot) \leq \lambda\}$  and by Proposition 2.2.8 we have that  $tp \in \{H(x_n, \cdot) \leq \lambda\}$  indefinitely for  $n$  large and thus

$$\liminf_n L_\lambda(x_n, q_n) \geq \liminf_n tp \cdot q_n.$$

The thesis follows sending  $t \rightarrow 1$ .  $\square$

We also have the following result of lower semicontinuity in  $x$  of  $L_\lambda$ , uniform with respect to  $q$ .

**Lemma 2.2.16.** *Let  $\lambda > 0$ ,  $x_0 \in \Omega$ ,  $K > 0$ . Then for all  $\eta > 0$  there exists  $\delta > 0$  such that*

$$|q| \leq K, |x - x_0| \leq \delta \quad \Rightarrow \quad L_\lambda(x, q) \geq L_\lambda(x_0, q) - \eta.$$

*Proof.* Let us assume by contradiction that there exists  $\bar{\eta} > 0$  such that for all  $n \geq 0$ , there exists  $q_n$ , with  $q_n \leq K$ , and  $x_n \in \Omega$ , with  $|x_n - x| < 2^{-n}$ , such that

$$L_\lambda(x_n, q_n) - L_\lambda(x_0, q_n) < -\bar{\eta}.$$

Since  $K$  is compact, up to the choice of a subsequence, we have that  $q_n$  converges to a certain point  $q \in K$ . Then we get the following contradiction:

$$L_\lambda(x_0, q) = \lim_{n \rightarrow \infty} L_\lambda(x_n, q_n) \leq \liminf_{n \rightarrow \infty} L_\lambda(x_0, q_n) - \bar{\eta} = L_\lambda(x_0, q) - \bar{\eta},$$

where the equalities are due to Lemma 2.2.15.  $\square$

### 2.2.3 The family of pseudo-distances and a Lipschitz characterization of minimizers

As pointed out in the Introduction 1.1, it makes sense to define a new metric on  $\Omega$  which depends on the sublevel sets  $\{H(x, \cdot) \leq \lambda\}$  in order to extend the notion of Lipschitz function. As showed in Proposition 2.2.11,  $L_\lambda(x, \cdot)$  is a Finsler metric for every  $x \in \bar{\Omega}$  and  $\lambda \geq 0$ , thus we can use it to define a family of (pseudo) distances on  $\Omega$  (and on the connected open subsets of  $\Omega$ ) associated to  $H$  (see 2.1.11). The fact that we talk about “pseudo” distances is due to the lack of symmetry (see Remark 2.1.19).

This family of distances is useful to characterize the minimizers and to prove a property for the absolute minimizers as we will see in the sequel of this chapter.

For every  $\lambda \geq 0$  and any  $x, y \in \Omega$  and any  $\lambda \geq 0$ , we set

$$d_\lambda(x, y) := \inf \left\{ \int_0^1 L_\lambda(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}(x, y) \right\}, \quad (2.2.3)$$

where  $\text{path}(x, y) := \text{path}_\Omega(x, y)$ , defined in (2.1.5).

For any  $x, y \in \bar{\Omega}$  we set

$$d_\lambda(x, y) := \inf \left\{ \liminf_{n \rightarrow +\infty} d_\lambda(x_n, y_n) : (x_n)_n, (y_n)_n \in \Omega^\mathbb{N} \text{ and } x_n \rightarrow x, y_n \rightarrow y \right\}.$$

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

*Remark 2.2.17.* Since for every  $\lambda \geq 0$ ,  $L_\lambda$  is a Finsler metric,  $d_\lambda$  has all the properties pointed out in Section 2.1.2, in particular in Remark 2.1.20. Moreover by Proposition 2.1.24  $d : \Omega \times \Omega \rightarrow \mathbb{R}$  is continuous and by Corollary 2.1.26 it is equivalent to  $|\cdot|_\Omega$ .

*Remark 2.2.18.* By definition, the function  $\lambda \mapsto d_\lambda$  is non-decreasing. In particular the property (D) of  $H$  yields the strict monotonicity.

*Remark 2.2.19.* Sometimes it could be useful to restrict Definition 2.1.14 to connected open subsets  $V$  of  $\Omega$ , that are connected and well contained in  $\Omega$ . For example when property (D) holds only locally (see Example 2.2.7). In this case the distance between two points  $x, y \in V$  will be

$$d_\lambda^V(x, y) := \inf \left\{ \int_0^1 L_\lambda(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}_V(x, y) \right\},$$

where

$$\text{path}_V(x, y) := \{ \gamma \in \text{Lip}([0, 1], V) : \gamma(0) = x, \gamma(1) = y \}.$$

We point out that all the properties listed in Remark 2.1.20 and all the results in this chapter hold also when we restrict to  $V$ .

**Proposition 2.2.20.** *The map  $\lambda \mapsto d_\lambda(x, y)$  is left continuous on  $\mathbb{R}_+$  for every  $(x, y) \in \overline{\Omega} \times \overline{\Omega}$ .*

*Proof.* Let us consider  $x, y \in \Omega$  and  $V \subset\subset \Omega$  such that  $x, y \in V$ . Fix  $\lambda \geq 0$  and consider a sequence  $(\lambda_n)_n \subset \mathbb{R}_{\geq 0}$  such that  $\lambda_n \nearrow \lambda$ . Let us then take  $\beta > 0$ . From the property (E) of  $H$  we know that there exists  $\delta$  such that

$$|\lambda - \bar{\lambda}| < \delta \text{ implies } \{H(x, \cdot) < \lambda\} \subset \{H(x, \cdot) < \bar{\lambda}\} + B(0, \beta).$$

So, if we take  $\bar{n}$  such that  $|\lambda_{\bar{n}} - \lambda| < \delta$ , we obtain

$$\begin{aligned} L_\lambda(x, q) &:= \sup\{p \cdot q : H(x, p) \leq \lambda\} \\ &\leq \sup\{(\tilde{p} + w_\beta) \cdot q : H(x, \tilde{p}) \leq \lambda_{\bar{n}}, w_\beta \in B(0, \beta)\} \\ &= L_{\lambda_{\bar{n}}}(x, q) + \beta|q|, \end{aligned} \tag{2.2.4}$$

for every  $x \in V$ . From (2.2.4) and from the monotonicity of  $\lambda \mapsto d_\lambda$  we infer that, for any  $(x, y) \in \Omega \times \Omega$ ,

$$d_{\lambda_n}(x, y) < d_\lambda(x, y) \leq d_{\lambda_n}(x, y) + \beta|x - y|_\Omega. \tag{2.2.5}$$

proving the left continuity in  $\Omega \times \Omega$ . The extension of this result to any two points  $x, y \in \overline{\Omega} \times \overline{\Omega}$  follows by applying (2.2.5) to any two converging sequences  $x_n \rightarrow x$  and  $y_n \rightarrow y$  and then considering the infimum.  $\square$

We are now ready to state and prove the above mentioned characterization of the minimizers of (H), proving a Lipschitz property w.r.t.  $d_\lambda$ .

*Remark 2.2.21. Comment on the assumptions.*

The results of the present section (Section 2.2.3) and of Section 2.2.4 below are present in [43] and they all hold for milder assumptions on  $H$ , in particular assumption (C) can be weakened in

(C') the map  $(x, p) \mapsto H(x, p)$  is lower semicontinuous in  $\Omega \times \mathbb{R}^d$

and the assumption (E) can be dropped.

**Theorem 2.2.22.** *Let  $\lambda \geq 0$  and  $u : \bar{\Omega} \rightarrow \mathbb{R}$  be such that*

$$u(y) - u(x) \leq d_\lambda(x, y), \quad \text{for all } x, y \in \Omega.$$

*Then  $u \in W^{1,\infty}(\Omega)$  and*

$$H(\cdot, Du(\cdot)) \leq \lambda$$

*a.e. in  $\Omega$ .*

*Viceversa, if  $u \in W^{1,\infty}(\Omega)$  and*

$$H(\cdot, Du(\cdot)) \leq \lambda$$

*a.e. in  $\Omega$ , then*

$$u(y) - u(x) \leq d_\lambda(x, y), \quad \text{for all } x, y \in \Omega.$$

*If  $u \in C(\bar{\Omega})$  then  $H(\cdot, Du(\cdot)) \leq \lambda$  a.e. in  $\Omega$  if and only if  $u(y) - u(x) \leq d_\lambda(x, y)$  for any  $x, y \in \bar{\Omega}$ .*

The same result as the one stated in Theorem 2.2.22 can be found in [43].

*Proof.* Let us assume that  $u(y) - u(x) \leq d_\lambda(x, y)$ , for any  $x, y \in \Omega$ . We take  $x, y \in \Omega$  such that the segment joining them is all contained in  $\Omega$ . Then, by assumption (D) on  $H$ , as pointed out in Proposition 2.2.11, there exists  $M \geq 0$

$$d_\lambda(x, y) \leq \int_0^1 L_\lambda(x + t(y-x), y-x) dt \leq M \int_0^1 |x-y| = M|x-y|.$$

Then

$$|u(y) - u(x)| \leq \max\{d_\lambda(x, y), d_\lambda(y, x)\} \leq M|x-y|,$$

implying that  $u \in W^{1,\infty}(\Omega) \cap C(\Omega)$ .

Since  $u$  is differentiable a.e. in  $\Omega$ , we will show now that

$$H(x_0, Du(x_0)) \leq \lambda$$

for any  $x_0$  that is a point of differentiability. We observe that, by (2.2.2) it is sufficient to show that  $Du(x_0) \cdot q \leq 1$  for every  $q$  such that  $L_\lambda(x, q) \leq 1$ .

$$Du(x_0) \cdot q = \lim_{h \rightarrow 0} \frac{u(x_0 + hq) - u(x_0)}{h} \leq \lim_{h \rightarrow 0} \frac{d_\lambda(x_0, x_0 + hq)}{h}.$$

Moreover, for  $h$  small enough, by the 1-homogeneity of  $L_\lambda$  we have

$$\frac{1}{h} d_\lambda(x_0, x_0 + hq) \leq \frac{1}{h} \int_0^1 L_\lambda(x_0 + thq, hq) dt = \int_0^1 L_\lambda(x_0 + thq, q) dt.$$

Finally, by the upper semicontinuity of  $L_\lambda$  it follows that

$$\limsup_{h \rightarrow 0} \frac{d_\lambda(x_0, x_0 + hq)}{h} \leq \limsup_{h \rightarrow 0} \int_0^1 L_\lambda(x_0 + thq, q) dt \leq \int_0^1 L_\lambda(x_0, q) dt \leq 1.$$

By (2.2.2), we have  $H(x_0, Du(x_0)) \leq \lambda$ .

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

Suppose now that  $H(\cdot, Du(\cdot)) \leq \lambda$  a.e. in  $\Omega$  for some  $\lambda \geq 0$ .

Let  $N := \{x \in \Omega : u \text{ is not differentiable at } x \text{ or } H(x, Du(x)) > \lambda\}$ . Then

$$u(y) - u(x) \leq \inf \left\{ \int_0^1 L_\lambda(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}(x, y), \gamma \text{ trasversal to } N \right\},$$

where trasversal means that  $\mathcal{H}^1(\gamma((0, 1)) \cap N) = 0$ . Indeed, for any  $\gamma$  trasversal to  $N$ , we have

$$u(y) - u(x) = \int_0^1 (u \circ \gamma)' dt = \int_0^1 Du(\gamma(t)) \cdot \dot{\gamma}(t) dt \leq \int_0^1 L_\lambda(\gamma(t), \dot{\gamma}(t)) dt,$$

where the inequality follows from the definition of  $L_\lambda$ , since  $H(\cdot, Du) \leq \lambda$ . Let's now take  $\gamma \in \text{path}(x, y)$ . It is possible to approximate  $\gamma$  in  $W^{1, \infty}((0, 1), \Omega)$  by a sequence  $(\gamma_k)_k \subset \text{path}(x, y)$ , with  $\gamma_k$  trasversal to  $N$  for any  $k \in \mathbb{N}$  (Lemma 2.2.23 below). Then it follows that

$$u(y) - u(x) \leq \inf \left\{ \int_0^1 L_\lambda(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}(x, y) \right\} = d_\lambda(x, y). \quad (2.2.6)$$

If  $u \in C(\bar{\Omega})$ , then, by (2.2.6) and continuity of  $u$ , one has

$$u(y) - u(x) \leq \liminf_{n \rightarrow +\infty} d_\lambda(x_n, y_n),$$

for any  $(x_n)_n, (y_n)_n$  such that  $x_n \rightarrow x$  and  $y_n \rightarrow y$ . Then  $u(y) - u(x) \leq d_\lambda(x, y)$  for any  $x, y \in \bar{\Omega}$ .  $\square$

**Lemma 2.2.23.** *Let  $x, y \in \Omega$ ,  $\gamma \in \text{Lip}([0, 1], \Omega)$  and  $E$  such that  $\mathcal{L}^d(E) = 0$ . Then for every  $\varepsilon > 0$  there exists a curve  $\gamma_\varepsilon \in \text{Lip}([0, 1], \Omega)$  transversal to  $E$  (i.e.  $\mathcal{H}^1(\gamma_\varepsilon((0, 1)) \cap E) = 0$ ) such that*

$$\|\gamma_\varepsilon - \gamma\|_{W^{1, \infty}((0, 1))} < \varepsilon.$$

*Proof.* Let  $g(t) \in C^1([0, 1])$  be such that  $g(0) = g(1) = 0$  and  $g(t) > 0$ , for every  $t \in (0, 1)$ . For every  $v \in \mathbb{R}^d$ , we define the curve  $\gamma_v(t) = \gamma(t) + vg(t)$ . Let  $A$  be the set of the points  $(t, v) \in [0, 1] \times \mathbb{R}^d$  such that  $\gamma_v(t) \in E$  and  $A_t := \{v \in \mathbb{R}^d : (t, v) \in A\}$ . Since  $\mathcal{L}^d(E) = 0$ ,  $\mathcal{L}^d(A_t) = 0$  for every fixed  $t \in [0, 1]$  and therefore also  $\mathcal{L}^{d+1}(A) = 0$ . This implies that  $A_v := \{t \in [0, 1] : (t, v) \in A\}$  is such that  $\mathcal{L}^1(A_v) = 0$  for a.e.  $v \in \mathbb{R}^d$ . Let  $v$  such that  $\mathcal{L}^1(A_v) = 0$ , then  $\gamma_v$  is transversal to  $E$ . Indeed by Lipschitzianity of  $\gamma_v$  we have that  $\mathcal{H}^1(\gamma_v(A_v)) = 0$  and we conclude by observing that  $\gamma_v(A_v) = \gamma_v([0, 1]) \cap E$ .

Finally, we choose  $v$  such that  $|v| < \varepsilon / \|g(t)\|_{W^{1, \infty}([0, 1])}$ , so to obtain

$$\|\gamma_v - \gamma\|_{W^{1, \infty}((0, 1))} \leq |v| \|g\|_{W^{1, \infty}([0, 1])} < \varepsilon. \quad \square$$

*Remark 2.2.24.* A consequence of Theorem 2.2.22 is that for every  $\lambda \geq 0$ ,

$$d_\lambda(x, y) = \delta_H^\lambda(x, y)$$

for every  $x, y \in \Omega$ , where

$$\delta_H^\lambda(x, y) = \sup \left\{ u(y) - u(x) : u \in W^{1, \infty}(\Omega), \text{ess sup}_{x \in \Omega} H(x, Du(x)) \leq \lambda \right\}$$

which is a natural way to define a distance associated to the supremand  $H$  in a supremal variational problem (see [72, 73]).

*Proof.* The inequality “ $\geq$ ” is trivial using Theorem 2.2.22. To recover the converse inequality we use again Theorem 2.2.22 to  $u(\cdot) = d_\lambda(z, \cdot)$ , for some  $z \in \Omega$  (indeed for every  $x, y \in \Omega$   $d_\lambda(z, y) - d_\lambda(z, x) \leq d_\lambda(x, y)$ ).  $\square$

**Definition 2.2.25.** Let  $V$  a connected open subset of  $\Omega$  and  $g$  be a function in  $W^{1,\infty}(V) \cap C(\bar{V})$  and  $\lambda \geq 0$ . We define the functions  $S_\lambda^-(g, V)$  and  $S_\lambda^+(g, V)$  given on  $\bar{V}$  by:

$$\begin{aligned} \forall x \in \bar{V} \quad S_\lambda^-(g, V)(x) &= \sup \{g(y) - d_\lambda^V(x, y) : y \in \partial\Omega\}, \\ \forall x \in \bar{V} \quad S_\lambda^+(g, V)(x) &= \inf \{g(y) + d_\lambda^V(y, x) : y \in \partial\Omega\}. \end{aligned}$$

When  $g$  and  $V$  are clear from the context, we will write  $S_\lambda^\pm(x)$  instead of  $S_\lambda^\pm(g, V)(x)$ .

*Remark 2.2.26.* As stressed in the Introduction 1.1, the functions  $S^-$  and  $S^+$  are obtained by analogy with the MacShane-Whitney operator (we refer to the introduction of [13] for more about this operator). We also underline the fact that in the expression of  $S^-$  it appears  $d_\lambda^V(x, y)$  while in the one of  $S^+$  one has  $d_\lambda(y, x)$ . Since  $d_\lambda^V$  is not symmetric this is an important fact.

**Theorem 2.2.27.** *Let  $g$  be a function of  $W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$  and assume that there exists  $\lambda \geq 0$  such that  $g$  is “ $\lambda$ -Lipschitz” on  $\partial\Omega$ , i.e.*

$$g(y) - g(x) \leq d_\lambda(x, y), \quad \text{for any } x, y \in \partial\Omega.$$

Let

$$\mu := \inf\{\lambda : g(y) - g(x) \leq d_\lambda(x, y) \text{ for any } x, y \in \partial\Omega\}, \quad (2.2.7)$$

then  $\mu$  is the minimal value for the problem (H) and

$$\mu = \min\{\lambda : g(y) - g(x) \leq d_\lambda(x, y) \text{ for any } x, y \in \partial\Omega\} < +\infty.$$

Moreover, the functions  $S_\mu^-(g, \Omega) = S_\mu^-$  and  $S_\mu^+(g, \Omega) = S_\mu^+$  are optimal solution of (H) and for any optimal solution  $u$  of (H) one has

$$S_\mu^-(x) \leq u(x) \leq S_\mu^+(x),$$

for all  $x \in \bar{\Omega}$ .

*Remark 2.2.28.* The same result holds if we consider a connected open subset  $V$  of  $\Omega$  and the associated  $d_\lambda^V$ .

*Proof.* We now notice that the minimum  $\mu$  does not need *a priori* to be attained, so we first set:

$$\mu := \inf\{\lambda : g(y) - g(x) \leq d_\lambda(x, y) \text{ for any } x, y \in \partial V\} \quad (2.2.8)$$

and for every  $x \in \bar{\Omega}$ ,

$$\begin{aligned} S_\mu^-(x) &:= \sup \{g(y) - d_\lambda(x, y) : \lambda > \mu, y \in \partial\Omega\}, \\ S_\mu^+(x) &:= \inf \{g(y) + d_\lambda(y, x) : \lambda > \mu, y \in \partial\Omega\}. \end{aligned}$$

We now claim that  $S_\mu^-(x) = g(x)$  for any  $x \in \partial\Omega$ . Indeed, taking  $y = x$  in the definition of  $S_\mu^-$  yields  $S_\mu^-(x) \geq g(x)$ , while by definition of  $\mu$  one has  $g(y) - d_\lambda(x, y) \leq g(x)$ , for any

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

$\lambda > \mu$  and  $y \in \partial\Omega$ , so that  $S_\mu^-(x) \leq g(x)$ , which in turn proves the claim. The same holds for  $S_\mu^+$ .

The second claim is that, for any  $\lambda > \mu$  and for any  $x, y \in \bar{\Omega}$ , one has:

$$S_\mu^-(y) - S_\mu^-(x) \leq d_\lambda(x, y). \quad (2.2.9)$$

Indeed, take  $\lambda > \mu$ ,  $x \in \bar{\Omega}$  and  $y \in \Omega$ . We notice that since  $\lambda \mapsto d_\lambda$  is not decreasing, the supremum in the definition of  $S_\mu^-$  can be taken for  $\sigma \in (\mu, \lambda]$ , so that

$$\begin{aligned} S_\mu^-(y) - S_\mu^-(x) &= \sup_{z \in \partial\Omega, \mu < \sigma \leq \lambda} \inf_{z' \in \partial\Omega, \mu < \sigma' \leq \lambda} \{g(z) - d_\sigma(y, z) - g(z') + d_{\sigma'}(x, z')\} \\ &\leq \sup_{z \in \partial\Omega, \mu < \sigma \leq \lambda} \{g(z) - d_\mu(y, z) - g(z) + d_\sigma(x, z)\} \leq \sup_{\mu < \sigma \leq \lambda} d_\sigma(x, y) = d_\lambda(x, y), \end{aligned}$$

where the last inequality is due to the triangular inequality that holds since  $y \in \Omega$ .

When  $y \in \partial\Omega$ , we notice that

$$S_\mu^-(y) = g(y) \text{ and } S_\mu^-(x) \geq g(y) - d_\lambda(x, y) \text{ for every } \lambda > \mu.$$

In a similar way the estimate (2.2.9) can be proved for  $S_\mu^+$ .

By Theorem 2.2.22 we infer that  $\mathcal{F}_H(S_\mu^-), \mathcal{F}_H(S_\mu^+) \leq \lambda$  for every  $\lambda > \mu$ . It follows that  $\mathcal{F}_H(S_\mu^-), \mathcal{F}_H(S_\mu^+) \leq \mu$ . In particular, by Theorem 2.2.22, we have that for every  $x, y \in \partial\Omega$

$$S_\mu^-(x) - S_\mu^-(y) \leq d_\mu(x, y),$$

and, since  $S_\mu^- = g$  on  $\partial\Omega$  this implies that the minimum in the definition of  $\mu$  (2.2.8) is attained.

Moreover, we show that the minimal value for the problem (H) is equal to  $\mu$ . By contradiction assume that there exists a function  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ ,  $u = g$  on  $\partial\Omega$ , such that  $\mathcal{F}_H(u) \leq \lambda$  for some  $\lambda < \mu$ . Then, by Theorem 2.2.22 again, we would have  $u(y) - u(x) \leq d_\lambda(x, y)$  for all  $x, y \in \bar{\Omega}$ . Since  $u = g$  on  $\partial\Omega$ , this contradicts the minimality of  $\mu$ .

Finally, let  $u$  be an optimal solution of (H), i.e.  $H(\cdot, Du(\cdot)) \leq \mu$  a.e. in  $\Omega$ . Then, by Theorem 2.2.22,  $u(y) - u(x) \leq d_\mu(x, y)$  for any  $x, y \in \bar{\Omega}$ . If  $y \in \partial\Omega$ , this yields  $g(y) - d_\mu(x, y) \leq u(x)$  and then  $S_\mu^-(x) \leq u(x)$ . With similar arguments can be shown that  $u \leq S_\mu^+$  in  $\bar{\Omega}$ .  $\square$

### 2.2.4 Absolute minimizers and the property of comparison with Distance Functions

**Definition 2.2.29.** An absolute minimizer (AM) for (H) is a function  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$  and for all open subset  $V \subset\subset \Omega$  one has

$$\operatorname{ess\,sup}_{x \in V} H(x, Du(x)) \leq \operatorname{ess\,sup}_{x \in V} H(x, Dv(x))$$

for all  $v$  in  $W^{1,\infty}(V) \cap C(\bar{V})$  such that  $u = v$  on  $\partial V$ .

Existence of absolute minimizers holds under our (or even milder) assumptions (see Theorem 4.1 or Theorem 4.7 in [48]).

Notice that in the above definition, we restrict ourselves to the open subsets  $V$  which are

relatively compact in  $\Omega$ , which is the common definition for absolute minimizers (see [13]), but we do not assume that  $u$  is a solution of (H) which is in fact a consequence of this definition and it is proved in the following lemma.

We recall that as stated in Remark 2.2.21 assumption (C) on  $H$  can be weakened in (C').

**Lemma 2.2.30.** *If  $u \in W^{1,\infty}(\Omega) \cap C(\overline{\Omega})$  is an absolute minimizer for the problem (H) such that  $u = g$  on  $\partial\Omega$ , then it is also a solution of (H).*

*Proof.* Let  $v \in W^{1,\infty}(\Omega) \cap C(\overline{\Omega})$ , such that  $u = g$  on  $\partial\Omega$ . We show that

$$\operatorname{ess\,sup}_\Omega H(\cdot, Du) \leq \operatorname{ess\,sup}_\Omega H(\cdot, Dv).$$

Let  $\delta > 0$ , we define  $V_\delta^+ := \{x : u > v + \delta\}$ . Since  $\partial V_\delta^+ = \{x : u = v + \delta\}$  and  $u = g = v$  on  $\partial\Omega$ , we have that  $V_\delta^+ \subset\subset \Omega$ . Thus, by the fact that  $u$  is an absolute minimizer we have that

$$\operatorname{ess\,sup}_{V_\delta^+} H(\cdot, Du) \leq \operatorname{ess\,sup}_{V_\delta^+} H(\cdot, Dv).$$

By letting  $\delta \rightarrow 0$  we obtain

$$\operatorname{ess\,sup}_{u>v} H(\cdot, Du) \leq \operatorname{ess\,sup}_{u>v} H(\cdot, Dv).$$

We prove that

$$\operatorname{ess\,sup}_{u<v} H(\cdot, Du) \leq \operatorname{ess\,sup}_{u<v} H(\cdot, Dv)$$

in a similar way by defining  $V_\delta^- := \{x : u < v - \delta\}$ .

We conclude by observing that  $Du(x) = Dv(x)$  for a.e.  $x \in \{u = v \cap \Omega\}$ .  $\square$

**Proposition 2.2.31.** *Let  $V \subset\subset \Omega$ , then for any  $\lambda \geq 0$ ,  $\alpha \in \mathbb{R}$  and  $x_0 \in \Omega$  one has*

$$\begin{aligned} d_\lambda(x_0, x) + \alpha &\geq S_\mu^+(d_\lambda(x_0, x) + \alpha, V) \text{ on } \overline{V}, \\ -d_\lambda(x, x_0) + \alpha &\leq S_\mu^-(d_\lambda(x, x_0) + \alpha, V) \text{ on } \overline{V}, \end{aligned}$$

where

$$\mu := \min\{\sigma : d_\lambda(x_0, y) - d_\lambda(x_0, x) \leq d_\sigma^V(x, y) \text{ for any } x, y \in \partial V\},$$

which is well defined thanks to Theorem 2.2.27 and Remark 2.2.28.

*Proof.* First of all we observe that  $d_\lambda(x, y) \leq d_\lambda^V(x, y)$  for every  $x, y \in \Omega$ , since  $\operatorname{path}_V(x, y) \subset \operatorname{path}(x, y)$  and that, by (4) of Remark 2.1.20,

$$d_\lambda(x_0, x) - d_\lambda(x_0, y) \leq d_\lambda(x, y),$$

for every  $x, y \in \Omega$ . We thus obtain  $\mu \leq \lambda$  and consequently  $d_\mu^V \leq d_\lambda^V$ . Therefore, recalling that,

$$S_\mu^+(d_\lambda(x_0, \cdot) + \alpha, V)(x) := \inf\{d_\lambda(x_0, y) + \alpha + d_\mu^V(y, x)\}$$

we have

$$S_\mu^+(d_\lambda(x_0, \cdot) + \alpha, V)(x) \leq \inf\{d_\lambda(x_0, y) + \alpha + d_\lambda^V(y, x)\}. \quad (2.2.10)$$

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

---

Let  $\delta > 0$  and  $\gamma \in \text{path}(x_0, x)$  such that

$$d_\lambda(x_0, x) \geq \int_0^1 L_\lambda(\gamma(t), \dot{\gamma}(t)) dt - \delta.$$

Let  $\bar{t} \in [0, 1)$  such that  $\gamma(t) \in V$  for every  $t > \bar{t}$ . Then  $\gamma(\bar{t})$  it holds

$$\begin{aligned} d_\lambda(x_0, x) &\geq \int_0^{\bar{t}} + \int_{\bar{t}}^{\bar{t} + \frac{1}{n}} + \int_{\bar{t} + \frac{1}{n}}^1 L_\lambda(\gamma(t), \dot{\gamma}(t)) dt - \delta \\ &\geq d_\lambda(x_0, \gamma(\bar{t})) + 0 + d_\lambda^V\left(\gamma\left(\bar{t} + \frac{1}{n}\right), x\right) - \delta, \end{aligned}$$

for any  $n > 0$  such that  $t + \frac{1}{n} < 1$ . By passing to the  $\liminf$  for  $n \rightarrow \infty$ , we have

$$d(x_0, x) \geq d_\lambda(x_0, \gamma(\bar{t})) + d_\lambda^V(\gamma(\bar{t}), x) - \delta.$$

We conclude by taking  $\gamma(\bar{t})$  as a competitor in (2.2.10). Indeed

$$S_\mu^+(d_\lambda(x_0, \cdot) + \alpha, V)(x) \leq d_\lambda(x_0, \gamma(\bar{t})) + \alpha + d_\lambda^V(\gamma(\bar{t}), x) \leq d(x_0, x) + \alpha + \delta,$$

for any  $\delta > 0$ . □

**Definition 2.2.32** (Comparison with Distance Functions). We say that a continuous function  $u : \bar{\Omega} \rightarrow \mathbb{R}$  satisfies the Comparison with Distance Functions (CDF) property from above if for any connected open subset  $V \subset\subset \Omega$ , any  $x_0 \in \Omega$ , any  $\lambda \geq 0$  and  $\alpha \in \mathbb{R}$ , the inequality

$$u \leq d_\lambda(x_0, \cdot) + \alpha, \quad \text{on } \partial(V \setminus \{x_0\})$$

implies

$$u \leq d_\lambda(x_0, \cdot) + \alpha, \quad \text{in } \bar{V}.$$

Analogously  $u : \bar{\Omega} \rightarrow \mathbb{R}$  satisfies the Comparison with Distance Functions (CDF) property from below if for any connected open subset  $V \subset\subset \Omega$ , any  $x_0 \in \Omega$ , any  $\lambda \geq 0$  and  $\alpha \in \mathbb{R}$ , the inequality

$$u \geq -d_\lambda(x_0, \cdot) + \alpha, \quad \text{on } \partial(V \setminus \{x_0\})$$

implies

$$u \geq -d_\lambda(x_0, \cdot) + \alpha, \quad \text{in } \bar{V}.$$

Finally a continuous function  $u : \bar{\Omega} \rightarrow \mathbb{R}$  satisfies the Comparison with Distance Functions (CDF) property if it satisfies CDF both from above and from below.

*Remark 2.2.33.* In case assumption (B) and (D) on  $H$  hold only locally, i.e. for every  $\lambda > \lambda' \geq 0$  and every  $V \subset\subset \Omega$  there exist  $0 < \alpha_V < M_V$  such that

$$B(0, \alpha) + \{H(x, \cdot) \leq \lambda'\} + B(0, \alpha_V) \subset \{H(x, \cdot) \leq \lambda\} \subset B(0, M_V),$$

for every  $x \in V$  (see Example 2.2.7), it makes sense to give a local definition. We say that  $u$  satisfies the CDF locally (from above and from below) if for any connected open subset  $V \subset\subset \Omega$ , any  $x_0 \in \bar{V}$  any  $\lambda \geq 0$  and  $\alpha \in \mathbb{R}$ , the inequality

$$\begin{aligned} u \leq d_\lambda^V(x_0, \cdot) + \alpha, \quad \text{on } \partial(V \setminus \{x_0\}) &\implies u \leq d_\lambda^V(x_0, \cdot) + \alpha, \quad \text{in } \bar{V} \quad \text{and} \\ u \geq -d_\lambda^V(x_0, \cdot) + \alpha, \quad \text{on } \partial(V \setminus \{x_0\}) &\implies u \geq -d_\lambda^V(x_0, \cdot) + \alpha, \quad \text{in } \bar{V}. \end{aligned} \quad (2.2.11)$$

**Theorem 2.2.34.** *Let  $g \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ . Then  $u = g$  on  $\partial\Omega$  is an absolute minimizer for (H) if and only if  $u$  satisfies the Comparison with Distance Functions property.*

*Proof.* We start with the *only if* part. Let  $u$  be an absolute minimizer and let us consider a connected open subset  $V \subset\subset \Omega$ ,  $x_0 \in \bar{V}$  and  $\alpha \in \mathbb{R}$  such that

$$u \leq d_\lambda(x_0, \cdot) + \alpha, \quad \text{on } \partial(V \setminus \{x_0\}).$$

Let  $\alpha_k := \alpha + \frac{1}{2^k}$  and

$$U_k := \{x \in V : u > d_\lambda(x_0, \cdot) + \alpha_k\}.$$

We observe that  $U_k \subset\subset V$ . Indeed if there exists a sequence  $(x_n) \in U_k$  converging to a point  $x \in \partial V$ , then

$$u(x) = \lim_n u(x_n) \geq \lim_n d_\lambda(x_0, x_n) + \alpha_k = d_\lambda(x_0, x) + \alpha_k > d_\lambda(x_0, x) + \alpha.$$

Then we have that  $u = d_\lambda(x_0, \cdot) + \alpha_k$  on  $\partial U_k$ . Let us assume that  $U_k$  is connected, otherwise the same reasoning holds in all the connected components. Since it is an absolute minimizer,  $u$  is a solution of

$$\min \{ \mathcal{F}_H(v, U_k) : v \in d_\lambda(x_0, \cdot) + \alpha_k + W^{1,\infty}(\Omega) \cap C_0(\Omega) \}$$

and, by Proposition 2.2.27

$$u(x) \leq S_\mu^+(d_\lambda(x_0, \cdot) + \alpha_k, U_k), \quad \text{for every } x \in U_k.$$

By Proposition 2.2.31 we infer that

$$u(x) \leq d_\lambda(x_0, \cdot) + \alpha_k, \quad \text{for every } x \in U_k,$$

proving that  $U_k$  is empty for every  $k$ . Thus  $u(x) \leq d_\lambda(x_0, x) + \alpha_k$  for every  $k$  and for every  $x \in V$ . The proof is concluded by letting  $k \rightarrow \infty$ .

We pass now to the *if* part. For simplicity we provide here the proof of the fact that if  $u$  satisfies the CDF locally (in the sense of (2.2.11) in Remark (2.2.33)) then  $u$  is an absolute minimizer for (H)<sup>2</sup>. The same proof in the case  $u$  satisfies the global property of CDF is in Theorem 4.3 in [43]. We omit it here because is much more technical and we present some of those techniques in the proof of Proposition 2.2.42. See also Remark 2.2.35

Let  $V$  a connected open subset  $V \subset\subset \Omega$ . We want to prove that

$$\mathcal{F}_H(u, V) := \min \{ \mathcal{F}_H(v, V) : v \in u + W^{1,\infty}(V) \cap C_0(\bar{V}) \}.$$

We know by Theorem 2.2.27 that this is true if and only if

$$\mathcal{F}_H(u, V) = \mu := \min \{ \lambda \geq 0 : u(y) - u(x) \leq d_\lambda^V(x, y) \text{ for any } x, y \in \partial V \}.$$

By definition of  $\mu$  we have that for every  $x \in \partial V$

$$u(y) \leq u(x) + d_\mu^V(x, y) \text{ for any } y \in \partial(V \setminus \{x\})$$

and since  $u$  satisfies the comparison with distance functions from above we have that the inequality holds for every  $y \in \bar{V}$ . With the same reasoning

$$u(x) \geq u(y) - d_\mu^V(x, y) \text{ for any } x \in \partial(V \setminus \{y\})$$

---

<sup>2</sup>This proof can be found in Theorem 3.5 in [43]

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

---

and since  $u$  satisfies the comparison with distance functions from below we have that the inequality holds for every  $x \in \bar{V}$ . Therefore

$$u(x) - u(y) \leq d_\mu^V(x, y) \text{ for any } x, y \in \bar{V}$$

and we conclude thanks to Theorem 2.2.22 which implies  $\mathcal{F}_H(u, V) \leq \mu$ . □

*Remark 2.2.35.* The main difference in the case  $u$  satisfies the CDF property globally is the fact that this property holds considering  $d_\lambda$ , while in the definition of  $\mu$  we have  $d_\lambda^V$ . In this case in [43] the authors show in Lemma B.4 that if  $x, y \in V \subset\subset \Omega$  are close enough, then  $d_\lambda^V(x, y) = d_\lambda^\Omega(x, y)$ . They then use this fact to apply the local CDF property to a suitable choice of a subset of  $V$ .

### 2.2.5 Point-wise definition of $H(x, Du(x))$

Since  $u \in W^{1,\infty}(\Omega)$  the quantity  $H(x, Du(x))$  is, a priori, defined only for a.e.  $x \in \Omega$ . In this section we show that there exists a natural point-wise definition of  $H(x, Du(x))$  which will be denoted by  $H(x, Du)(x)$ .

**Definition 2.2.36.** Let  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ . For any  $x_0 \in \Omega$  and for any  $r > 0$  such that  $r < \text{dist}(x_0, \partial\Omega)$ , we set

$$\mu(x_0, r) := \inf\{\lambda : u(x) - u(x_0) \leq d_\lambda(x_0, x) \text{ for any } x \in B(x_0, r)\}. \quad (2.2.12)$$

We observe that  $\mu(x_0, r)$  is not decreasing in  $r$ . This allows for the following definition:

$$H(\cdot, Du)(x_0) := \lim_{r \rightarrow 0} \mu(x_0, r) = \inf_r \mu(x_0, r).$$

**Lemma 2.2.37.** Let  $r > 0$  then  $x \mapsto \mu(x, r)$  is upper semicontinuous in  $\Omega$ .

*Proof.* Let  $x_n \rightarrow x$  and we may assume, without loss of generality that

$$\mu(x_n, r) \rightarrow \nu.$$

We want to prove that  $\mu(x, r) \geq \nu$ .

Let  $\alpha < \nu$ . For  $n$  such that  $\mu(x_n, r) > \alpha$  there exists  $y_n \in B(x_n, r)$  such that

$$u(y_n) - u(x_n) > d_\alpha(x_n, y_n),$$

and we may assume, up to extraction of a subsequence, that  $y_n \rightarrow y \in B(x, r)$ . Using the continuity of  $u$  and the continuity of  $d_\alpha$  (see Proposition 2.1.24), we have

$$u(y) - u(x) = \lim_{n \rightarrow +\infty} u(y_n) - u(x_n) \geq \lim_{n \rightarrow +\infty} d_\alpha(x_n, y_n) = d_\alpha(x, y),$$

which implies  $\mu(x, r) \geq \alpha$  for every  $\alpha < \nu$ . □

**Corollary 2.2.38.** The function  $x \mapsto H(\cdot, Du)(x)$  is upper semicontinuous in  $\Omega$ .

*Proof.* This follows from Lemma 2.2.37 above since the infimum of upper semicontinuous functions is upper semicontinuous. □

**Proposition 2.2.39.** *Let  $u \in W^{1,\infty}(\Omega)$  be such that  $u$  is differentiable at  $x_0 \in \Omega$ . Then*

$$H(x_0, Du(x_0)) \leq H(\cdot, Du)(x_0).$$

*Proof.* Let  $(\varepsilon_r)_{r>0}$  a sequence of positive real numbers that decreases to zero. We want to show that  $H(x_0, Du(x_0)) \leq \mu(x_0, r) + \varepsilon_r$ , for every  $r > 0$ . If we fix  $\mu_{\varepsilon_r} := \mu(x_0, r) + \varepsilon_r$  and we consider  $q \in \mathbb{R}^d$  such that  $L_{\mu_{\varepsilon_r}}(x_0, q) \leq 1$ , by the characterization (2.2.2) it is sufficient to show that

$$Du(x_0) \cdot q \leq 1.$$

Since  $u$  is differentiable at  $x_0$  we have:

$$\begin{aligned} Du(x_0) \cdot q &= \lim_{h \rightarrow 0} \frac{u(x_0 + hq) - u(x_0)}{h} \leq \lim_{h \rightarrow 0} \frac{1}{h} d_{\mu_{\varepsilon_r}}(x_0, x_0 + hq) \leq \\ &\leq \lim_{h \rightarrow 0} \frac{1}{h} \int_0^1 L_{\mu_{\varepsilon_r}}(x_0 + thq, hq) dt = \lim_{h \rightarrow 0} \int_0^1 L_{\mu_{\varepsilon_r}}(x_0 + thq, q) dt = \\ &= L_{\mu_{\varepsilon_r}}(x_0, q) \leq 1, \end{aligned}$$

where the first inequality follows from the definition of  $\mu(x_0, r)$ . The proof is then concluded.  $\square$

**Proposition 2.2.40.** *Let  $u \in W^{1,\infty}(\Omega)$  be differentiable at  $x_0$ . Then*

$$H(x_0, Du(x_0)) \geq H(x_0, Du)(x_0).$$

*Proof.* For every  $r > 0$  such that  $B(x_0, r) \subset \subset \Omega$  denote by  $\mu_r = \mu(x_0, r)$  and by  $\mu := H(x_0, Du)(x_0)$ . Let  $\varepsilon > 0$  and  $x_r \in B(x_0, r)$  be such that

$$u(x_r) - u(x_0) \geq d_{\mu_r - \varepsilon}(x_0, x_r).$$

Consider a sequence  $\frac{x_{r_n} - x_0}{s_n}$ , where  $s_n := |x_{r_n} - x_0|$ . By compactness up to the choice of a subsequence we have that

$$\frac{x_{r_n} - x_0}{s_n} \rightarrow q,$$

for some unitary vector  $q$ . Let  $(\varepsilon_n)$  be sequence of real numbers decreasing to 0. For every  $n \geq 0$ , we denote by  $\mu_n = \mu_{r_n}$ . Let  $\gamma_n \in \text{path}(x_0, x_{r_n})$ , suitably parametrized, such that (up to an increasing reparametrization)  $d_{\mu_n - \varepsilon}(x_0, x_{r_n}) \geq \int_0^{s_n} L_{\mu_n - \varepsilon}(\gamma_n, \dot{\gamma}_n) dt - \frac{\varepsilon_n}{2}$ . Then

$$\begin{aligned} \frac{u(x_{r_n}) - u(x_0)}{s_n} &\geq \frac{d_{\mu_n - \varepsilon}(x_0, x_{r_n})}{s_n} \\ &\geq \frac{1}{s_n} \int_0^{s_n} L_{\mu_n - \varepsilon}(\gamma_n, \dot{\gamma}_n) dt - \frac{\varepsilon_n}{2} \\ &\geq \frac{1}{s_n} \int_0^{s_n} L_{\mu - \varepsilon}(\gamma_n, \dot{\gamma}_n) dt - \frac{\varepsilon_n}{2} \\ &\geq \frac{1}{s_n} \int_0^{s_n} L_{\mu - \varepsilon}(x_0, \dot{\gamma}_n) dt - \frac{\varepsilon_n}{2} - \frac{\varepsilon_n}{2} \\ &\geq L_{\mu - \varepsilon}(x_0, \frac{1}{s_n} \int_0^{s_n} \dot{\gamma}_n dt) - \varepsilon_n \xrightarrow{\liminf} L_{\mu - \varepsilon}(x_0, q), \end{aligned}$$

obtaining that  $Du(x_0) \cdot q \geq L_{\mu - \varepsilon}(x_0, q)$ . The inequality in the third line above follows from Lemma 2.2.16 with  $\eta = \frac{\varepsilon_n}{2}$ . It follows by the definition of  $L_\lambda$  that

$$H(x_0, Du(x_0)) \geq \mu - \varepsilon.$$

The claim follows by the arbitrariness of  $\varepsilon$ .  $\square$

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

---

**Corollary 2.2.41.** For all  $u \in W^{1,\infty}(\Omega)$

$$\operatorname{ess\,sup}_{x \in \Omega} H(x, Du(x)) = \operatorname{ess\,sup}_{x \in \Omega} H(\cdot, Du)(x).$$

### 2.2.6 The attainment set and its minimality

**Proposition 2.2.42.** Let  $u$  be an absolute minimizer for the problem (H). Let  $x_0$  be such that  $H(x_0, Du)(x_0) = \operatorname{ess\,sup}_{x \in \Omega} H(x, Du(x)) = \mu$  and let  $\alpha$  and  $M$  be such that  $B(0, \alpha) \subset \{H(x, \cdot) \leq \lambda\} \subset B(0, M)$  for all  $\lambda$  in a neighborhood of  $\mu$ . Then there exists  $y$  such that  $d_\mu(x_0, y) = \frac{\alpha^2}{2M} d(x_0, \partial\Omega)$  such that  $u(y) - u(x_0) = d_\mu(x_0, y)$  and  $H(y, Du)(y) = \mu$ .

*Proof.* Let  $\lambda < \mu$ . By Theorem 2.2.27,  $\mu$  is the one defined by (2.2.7). Then for all  $r > 0$  there exists  $y_r \in B(x_0, r)$  such that

$$u(y_r) - u(x_0) > d_\lambda(x_0, y_r)$$

or, equivalently, for every open set  $A \subset\subset \Omega$  there exists  $y_A \in A$  such that

$$u(y_A) - d_\lambda(x_0, y_A) > u(x_0).$$

Let  $R = \frac{\alpha^2}{2M} \operatorname{dist}(x_0, \partial\Omega)$ . Consider the open set  $V_R = \{z \in \Omega : d_\lambda(x_0, z) < R\}$ . By definition of  $R$ ,  $V_R \subset\subset \Omega^3$ . Let

$$a = \max_{\overline{V}_R} \{u(z) - d_\lambda(x_0, z)\}.$$

We claim that

$$a = \max_{\partial V_R} \{u(z) - d_\lambda(x_0, z)\}.$$

Indeed, let

$$a_b = \max_{\partial(V_R \setminus \{x_0\})} \{u(z) - d_\lambda(x_0, z)\},$$

we have that the distance function

$$a_b + d_\lambda(x_0, z)$$

satisfies

$$a_b + d_\lambda(x_0, z) \geq u(z) \quad \text{on} \quad \partial(V_R \setminus \{x_0\}),$$

and then, since  $u$  is an absolute minimizer, thanks to Theorem 2.2.34  $u$  satisfies the comparison with distance functions property and thus

$$a_b + d_\lambda(x_0, z) \geq u(z) \quad \text{on} \quad \overline{V}_R.$$

Moreover, by the choice of  $x_0$  and the fact that  $\lambda < \mu$ , the maximum in  $a_b$  can not be reached in  $x_0$ . Thus we have  $a = a_b$ . Let  $y_\lambda \in \partial V_R$  be such that  $a = u(y_\lambda) - d_\lambda(x_0, y_\lambda)$  let  $\varepsilon \ll 1$  and let  $y_\varepsilon \in B(y_\lambda, \varepsilon)$  be a point of the  $d_\lambda$ -geodesic between  $x_0$  and  $y_\lambda$ , which exists by Corollary 2.1.28. By the definition of  $V_R$ ,  $y_\varepsilon \in V_R$  so that

$$\begin{aligned} u(y_\varepsilon) - d_\lambda(x_0, y_\varepsilon) &\leq a \leq u(y_\lambda) - d_\lambda(x_0, y_\lambda) \\ &= u(y_\lambda) - d_\lambda(x_0, y_\varepsilon) - d_\lambda(y_\varepsilon, y_\lambda). \end{aligned}$$

---

<sup>3</sup>Indeed if  $z \in V_R$  then  $|x_0 - z| \leq \frac{1}{\alpha} d_\lambda(x_0, z) < \frac{1}{\alpha} R = \frac{\alpha}{2M} \operatorname{dist}(x_0, \partial\Omega) < \operatorname{dist}(x_0, \partial\Omega)$ .

We obtain

$$u(y_\lambda) - u(y_\varepsilon) \geq d_\lambda(y_\varepsilon, y_\lambda),$$

and then for small  $\delta > 0$

$$\mu(y_\varepsilon, \varepsilon) > \lambda - \delta$$

and when  $\varepsilon < \delta$

$$\mu(y_\varepsilon, \delta) \geq \mu(y_\varepsilon, \varepsilon) > \lambda - \delta.$$

By the upper semicontinuity (see Lemma 2.2.37) we have for fixed  $\delta$

$$\mu(y_\lambda, \delta) \geq \limsup_{\varepsilon \rightarrow 0} \mu(y_\varepsilon, \delta) \geq \lambda - \delta,$$

and letting  $\delta \rightarrow 0$ ,

$$H(y_\lambda, Du)(y_\lambda) \geq \lambda.$$

Consider  $\lambda_n \nearrow \mu$  and the corresponding  $y_{\lambda_n} \in \partial V_{R_n}$  as constructed above. Up to subsequences  $y_{\lambda_n} \rightarrow y$  and, by the upper semicontinuity of  $H(x, Du)(x)$  we obtain

$$H(y, Du)(y) \geq \mu.$$

Moreover

$$u(y) - u(x_0) = \lim_{n \rightarrow \infty} u(y_{\lambda_n}) - u(x_0) = \lim_{n \rightarrow \infty} d_{\lambda_n}(x_0, y_n) = d_\mu(x_0, y),$$

where the last equality is due to Lemma 2.2.20.  $\square$

*Remark 2.2.43.* We observe that from Proposition 2.2.42 one can infer that for every  $V \subset\subset \Omega$  that contains  $x_0$ , there exists  $y \in \partial V$  such that  $H(y, Du)(y) = \mu$ .

**Theorem 2.2.44.** *Let  $u$  be an absolute minimizer for the problem (H) and let  $x_0 \in \Omega$  be such that*

$$H(\cdot, Du)(x_0) = \operatorname{ess\,sup}_{x \in \Omega} H(x, Du)(x) := \mu. \quad (2.2.13)$$

*Then there exist  $x_{+\infty}, x_{-\infty} \in \partial\Omega$ , such that*

$$\begin{aligned} u(x_0) &= u(x_{-\infty}) + d_\mu(x_{-\infty}, x_0) = g(x_{-\infty}) + d_\mu(x_{-\infty}, x_0), \\ u(x_0) &= u(x_{+\infty}) - d_\mu(x_0, x_{+\infty}) = g(x_{+\infty}) - d_\mu(x_0, x_{+\infty}) \text{ and} \\ u(x_{+\infty}) - u(x_{-\infty}) &= d_\mu(x_{-\infty}, x_{+\infty}). \end{aligned}$$

*Proof.* We first claim that there exists  $x_{+\infty} \in \partial\Omega$ , such that

$$u(x_{+\infty}) - u(x_0) \geq d_\mu(x_0, x_{+\infty}). \quad (2.2.14)$$

Starting from  $x_0$  and following Proposition 2.2.42 we build a sequence  $(x_n) \subset \Omega$  such that for all  $n \in \mathbb{N}$

$$\frac{\alpha^2}{2M^2} \operatorname{dist}(x_n, \partial\Omega) = \frac{d_\mu(x_n, x_{n+1})}{M} \leq |x_{n+1} - x_n| \leq \frac{d_\mu(x_n, x_{n+1})}{\alpha} = \frac{\alpha}{2M} \operatorname{dist}(x_n, \partial\Omega)$$

and

$$u(x_{n+1}) - u(x_n) = d_\mu(x_n, x_{n+1}).$$

## 2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

---

From this it follows that for any  $n \in \mathbb{N}$

$$u(x_n) - u(x_0) = \sum_{i=0}^{n-1} d_\mu(x_i, x_{i+1}) \geq \sum_{i=0}^{n-1} \frac{\alpha^3}{2M^2} \text{dist}(x_i, \partial\Omega)$$

Since  $u$  is continuous in  $\bar{\Omega}$  and  $\Omega$  is bounded, both series on the right of the equation above converge and then

$$\lim_{n \rightarrow \infty} \text{dist}(x_n, \partial\Omega) = 0,$$

and  $(x_n)$  is a Cauchy sequence converging to some point  $x_{+\infty} \in \partial\Omega$ . For this point it holds

$$\begin{aligned} u(x_{+\infty}) - u(x_0) &= \lim_{n \rightarrow \infty} u(x_n) - u(x_0) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} d_\mu(x_i, x_{i+1}) \\ &\geq \lim_{n \rightarrow \infty} d_\mu(x_0, x_n) = d_\mu(x_0, x_{+\infty}). \end{aligned}$$

In a similar way we can find  $x_{-\infty} \in \partial\Omega$ , such that

$$u(x_0) - u(x_{-\infty}) \geq d_\mu(x_{-\infty}, x_0),$$

and so we get

$$u(x_{+\infty}) - u(x_{-\infty}) \geq d_\mu(x_{-\infty}, x_{+\infty}).$$

One also deduces that

$$\begin{aligned} &d_\mu(x_{-\infty}, x_0) + d_\mu(x_0, x_{+\infty}) \\ &= u(x_{+\infty}) - u(x_0) + u(x_0) - u(x_{-\infty}) \\ &= u(x_{+\infty}) - u(x_{-\infty}) = d_\mu(x_{-\infty}, x_{+\infty}), \end{aligned}$$

that is  $x_0$  belongs to a geodesics for  $d_\mu$  connecting  $x_{-\infty}$  to  $x_{+\infty}$ . □

In the following example we show that assumption (E) on the functional  $H$  and the consequent left continuity of the map  $\lambda \mapsto d_\lambda$  (Prop. 2.2.20), are essential for the validity of Proposition 2.2.42 and the consequent proof of the Theorem 2.2.44.

*Example 2.2.45.* Let's take  $\Omega = B(0, 2)$  and

$$H(x, p) = \begin{cases} |p| & \text{if } |p| < \frac{1}{2} \\ \frac{1}{2} & \text{if } \frac{1}{2} \leq |p| \leq \frac{3}{4} \\ |p| - \frac{1}{4} & \text{if } |p| > \frac{3}{4} \end{cases} .$$

Let  $u : B(0, 2) \rightarrow \mathbb{R}$  be such that  $u(x) = a \cdot x$ , where  $\frac{1}{2} < |a| < \frac{3}{4}$ . Then  $u$  is an absolute minimizer for  $H$  and  $H(x, Du(x)) = \text{ess sup}_{x \in \Omega} H(x, Du(x)) = \frac{1}{2}$  for all  $x \in \Omega$ .

Calling  $x_0 = 0$  and taking  $V = B(0, 1)$  we have that, for any  $y \in \partial V$ ,

$$u(y) - u(x_0) = a \cdot y \leq |a| < \frac{3}{4}.$$

Moreover

$$L_{\frac{1}{2}}(x, q) = \frac{3}{4}|q| \quad \text{for any } x \in \Omega$$

and

$$d_{\frac{1}{2}}^V(x_0, y) = \inf \left\{ \frac{3}{4} \int_0^1 |\dot{\gamma}(t)| dt \right\} = \frac{3}{4} \int_0^1 |y| dt = \frac{3}{4} \quad (2.2.15)$$

That is

$$\max_{x \in \partial V} \{u(y) - d_{\frac{1}{2}}^V(x, y)\} < 0.$$

**Definition 2.2.46.** We call *attainment set* of a function  $u \in W^{1,\infty} \cap C(\bar{\Omega})$ , the set  $\mathcal{A}(u)$  defined by

$$\mathcal{A}(u) := \{x \in \Omega : H(x, Du)(x) = \operatorname{ess\,sup}_{\Omega} H(x, Du(x))\}.$$

We finally prove the minimality property announced at the end of the introduction

**Theorem 2.2.47.** *Let  $\Omega$  be a bounded open set of  $\mathbb{R}^d$  and  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$  be an absolute minimizer for (H) and let  $v$  be any other minimizer for (H), then*

$$\mathcal{A}(u) \subset \mathcal{A}(v). \quad (2.2.16)$$

Moreover  $\mathcal{A}(u)$  is the union of Lipschitz curves which are geodesics for  $d_\mu$  where  $\mu$  is the minimal value in (H)<sup>4</sup>.

*Proof.* Let  $x_0 \in \mathcal{A}(u)$  then, by Theorem 2.2.44, there exist  $x_{+\infty}, x_{-\infty} \in \partial\Omega$ , such that

$$u(x_{+\infty}) - u(x_{-\infty}) = d_\mu(x_{-\infty}, x_{+\infty}).$$

$$u(x_{+\infty}) - u(x_0) = d_\mu(x_0, x_{+\infty}), \text{ and } u(x_0) - u(x_{-\infty}) = d_\mu(x_{-\infty}, x_0)$$

This implies that

$$d_\mu(x_{-\infty}, x_0) + d_\mu(x_0, x_{+\infty}) = d_\mu(x_{-\infty}, x_{+\infty}).$$

Then  $x_0$  belongs to a curve  $\gamma \in \operatorname{Lip}([0, 1])$  of minimal length for  $d_\mu$  (obtained, for example, joining a curve of minimal length from  $x_{-\infty}$  to  $x_0$  and one from  $x_0$  to  $x_{+\infty}$ ). We now prove that

$$\gamma([0, 1]) \subset \mathcal{A}(u) \text{ and } \gamma([0, 1]) \subset \mathcal{A}(v).$$

Let  $t_0 \in [0, 1]$  such that  $x_0 = \gamma(t_0)$  and let  $0 \leq t_0 < s \leq 1$ , then we have

$$u(\gamma(s)) - u(x_0) = d_\mu(x_0, \gamma(s)).$$

Indeed

$$\begin{aligned} d_\mu(x_0, x_{+\infty}) &= u(x_{+\infty}) - u(x_0) = u(x_{+\infty}) - u(\gamma(s)) + u(\gamma(s)) - u(x_0) \\ &\leq d_\mu(\gamma(s), x_{+\infty}) + d_\mu(x_0, \gamma(s)) = d_\mu(x_0, x_{+\infty}), \end{aligned}$$

by minimality of  $\gamma$ . Analogously, if  $0 \leq s < t_0 \leq 1$ , then

$$u(x_0) - u(\gamma(s)) \leq d_\mu(\gamma(s), x_0).$$

Therefore, for any  $0 \leq t < s \leq 1$

$$u(\gamma(s)) - u(\gamma(t)) = d_\mu(\gamma(t), \gamma(s)).$$

---

<sup>4</sup>We recall that existence of geodesics, that is of minimal  $d_\mu$  length, is ensured by Proposition 2.1.28.

2.2. SOME PROPERTIES OF ABSOLUTE MINIMIZERS OF SUPREMAL FUNCTIONALS

---

Then, since we can choose  $s$  arbitrarily close to  $t$ ,

$$H(\gamma(t), Du)(\gamma(t)) \geq \mu,$$

and, since the other inequality holds everywhere, equality holds. A consequence of this proof is that if  $u(\gamma(t)) = v(\gamma(t))$  for any  $t \in [0, 1]$  then

$$H(\gamma(t), Dv)(\gamma(t)) \geq \mu.$$

We now prove that for any  $t$

$$S_{\mu}^{-}(u, \Omega)(\gamma(t)) = S_{\mu}^{+}(u, \Omega)(\gamma(t)) = u(\gamma(t)),$$

By definition of  $S_{\mu}^{+}(u, \Omega)$  and  $S_{\mu}^{-}(u, \Omega)$ , for any  $t \in [-1, 1]$ , it holds

$$\begin{aligned} S_{\mu}^{-}(u, \Omega)(\gamma(t)) &\geq u(x_{+\infty}) - d(\gamma(t), x_{+\infty}) = u(\gamma(t)) \quad \text{and} \\ S_{\mu}^{+}(u, \Omega)(\gamma(t)) &\leq u(x_{-\infty}) + d(x_{-\infty}, \gamma(t)) = u(\gamma(t)), \end{aligned} \tag{2.2.17}$$

where the equality follows from the (2.2.6). The converse inequality is provided by Theorem 2.2.27, for which

$$S_{\mu}^{-}(u, \Omega)(x) \leq u(x) \leq S_{\mu}^{+}(u, \Omega)(x).$$

We then conclude recalling that again Theorem 2.2.27 implies that  $u(\gamma(t)) = v(\gamma(t))$  for any  $t \in [0, 1]$  for any other minimizer  $v$ .  $\square$



## Chapter 3

# On functions with convex constraints on the gradient

### 3.1 Preliminaries

**Definition 3.1.1.** Given an open set  $\Omega$  and  $p \in [1, +\infty]$ , we say that  $u$  satisfies a convex constraint on the gradient if it belongs to the set

$$\{u \in W^{1,p}(\Omega) : Du(x) \in C(x) \text{ for a.e. } x \in \Omega\}, \quad (3.1.1)$$

where, for a.e.  $x \in \Omega$ ,  $C(x)$  is a convex closed subset of  $\mathbb{R}^d$ .

Given a bounded and connected open set  $\Omega \subset \mathbb{R}^d$ , we consider the Dirichlet problem associated to the problem of convex constraints on the gradient, that is: given a function  $g \in C(\partial\Omega)$ , we study the solutions  $u \in W^{1,\infty}(\Omega) \cap C(\overline{\Omega})$  of the following system

$$\begin{cases} u = g & \text{on } \partial\Omega, \\ Du(x) \in K(x) & \text{for a.e. } x \in \Omega, \end{cases} \quad (\text{P})$$

where  $K(x)$  belongs to the set  $\mathcal{K}$  of the convex bodies of  $\mathbb{R}^d$  (see (2.1.1)) for every  $x \in \overline{\Omega}$ . We will also assume that  $0 \in K(x)$  for every  $x \in \overline{\Omega}$ .

The results in this chapter extend the work of Arronson [12] where he considers the more classical case  $K(x) := \overline{B(0, f(x))}$ ,  $f$  being a nonnegative function, which is equivalent to asking that  $|Du(x)| \leq f(x)$  for a.e.  $x \in \Omega$ .

For any  $x \in \overline{\Omega}$ , we consider the *gauge functional*  $\varphi(x, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}_+$  associated to  $K(x)$  (see Definition 2.1.1), i.e.

$$\varphi(x, p) := \inf \left\{ t > 0 : \frac{p}{t} \in K(x) \right\}.$$

Then  $\varphi(x, \cdot)$  satisfies the properties (i), (ii), (iii) in Remark 2.1.2 for every  $x \in \overline{\Omega}$ . Moreover, by (2.1.2) in Remark 2.1.6, it holds:

$$\begin{aligned} p \in K(x) &\Leftrightarrow \varphi(x, p) \leq 1 \\ p \in \partial K(x) &\Leftrightarrow \varphi(x, p) = 1. \end{aligned}$$

**Assumptions.** From now on we will also requires that:

(a) there exist  $\alpha, M > 0$  such that  $B(0, \alpha) \subset K(x) \subset B(0, M)$ , for every  $x \in \bar{\Omega}$ ;

(b)  $\varphi(\cdot, p)$  is continuous on  $\Omega$  for every  $p \in \mathbb{R}^d$ .

*Remark 3.1.2.* Assuming the continuity of  $\varphi(\cdot, p)$  in (b) is equivalent to asking for the continuity of  $K : \Omega \rightarrow \mathcal{K}$  w.r.t. the Hausdorff distance, as shown in Proposition 3.1.6 below.

*Remark 3.1.3.* With assumptions (a) (which implies Equation (3.1.2) below) and (b) we have that  $\varphi : \bar{\Omega} \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  is a Finsler metric, defined in Definition 2.1.11.

First of all, we prove that  $\varphi$  is continuous w.r.t. the second variable as well.

**Proposition 3.1.4.** *There exist  $\alpha, M > 0$  such that*

$$\frac{|p|}{M} \leq \varphi(x, p) \leq \frac{|p|}{\alpha}, \quad \text{for all } x \in \bar{\Omega}. \quad (3.1.2)$$

*In particular,  $\varphi(x, \cdot)$  is continuous w.r.t the Euclidean distance, uniformly w.r.t.  $x \in \bar{\Omega}$ .*

*Proof.* For every  $t < \frac{|p|}{M}$ ,  $\frac{p}{t} \notin K(x)$ , because  $|\frac{p}{t}|$  is greater than  $M$ , that is not possible by assumption (a). Then the left inequality is proved in (3.1.2). The proof of the right one, follows easily by the fact that again by the assumption (a),

$$\alpha \frac{p}{|p|} \in K(x),$$

and in view the positive 1-homogeneity of  $\varphi$ .

Finally, we show that the upper bound on  $\varphi(x, p)$  in the (3.1.4) implies continuity uniformly w.r.t.  $x \in \bar{\Omega}$ . Indeed for every  $q \in \mathbb{R}^d$

$$|\varphi(x, q) - \varphi(x, p)| \leq \max\{\varphi(x, q - p), \varphi(x, p - q)\} \leq \frac{|q - p|}{\alpha},$$

where the first inequality holds thanks to the subadditivity property of  $\varphi$  (property (iii) in Remark 2.1.2). □

**Corollary 3.1.5.** *The functional  $\varphi : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$  is continuous.*

*Proof.* The proof is a direct consequence of Proposition 3.1.4 and assumption (b). □

**Proposition 3.1.6.** *Let  $x \in \Omega$  and  $(x_n)_{n \in \mathbb{N}}$  a sequence converging to  $x$ . Then, for all  $p \in \mathbb{R}^d$*

$$\varphi(x_n, p) \longrightarrow \varphi(x, p) \iff K(x_n) \xrightarrow{H} K(x),$$

*where on the right side we mean the convergence w.r.t. to the Hausdorff distance for sets<sup>1</sup>.*

*Proof.* In order to lighten the notation, we write  $K := K(x)$  and  $K_n := K(x_n)$ .

We first assume that  $\varphi$  is continuous w.r.t  $x$ . Let  $\varepsilon > 0$ , then there exists  $n_0 \in \mathbb{N}$  such that

$$|\varphi(x_n, p) - \varphi(x, p)| < \varepsilon, \quad \text{for all } n \geq n_0. \quad (3.1.3)$$

We recall that  $d_H(K_n, K) = \max\{\rho(K_n, K), \rho(K, K_n)\}$ , where

$$\rho(K_n, K) := \sup_{p \in K_n} \text{dist}(p, K) \quad \text{and} \quad \rho(K, K_n) := \sup_{p \in K} \text{dist}(p, K_n).$$

---

<sup>1</sup>For more details about Hausdorff distance see e.g. [74].

Let  $n \geq n_0$ . Then for every  $p \in K$  and  $\frac{p}{1+\varepsilon} \in K_n$ , since, by (3.1.3),

$$\varphi(x_n, p) < \varphi(x, p) + \varepsilon \leq 1 + \varepsilon.$$

That means that

$$\text{dist}(p, K_n) \leq \left| p - \frac{p}{1+\varepsilon} \right| = \frac{\varepsilon|p|}{1+\varepsilon} \leq \frac{\varepsilon}{1+\varepsilon}M,$$

and thus  $\rho(K, K_n) \leq \frac{\varepsilon}{1+\varepsilon}M$ . Let us now consider  $p \in K_n$ , then as above  $\frac{p}{1+\varepsilon} \in K$  and  $\rho(K_n, K) \leq \frac{\varepsilon}{1+\varepsilon}M$ .

We now prove the converse implication. Let  $p \in \mathbb{R}^d$ . Proving that

$$\lim_{n \rightarrow +\infty} \varphi(x_n, p) = \varphi(x, p)$$

is equivalent to show that

$$\lim_{n \rightarrow \infty} \varphi(x_n, p) = \varphi(x, p) = 1, \quad \text{for every } p \in \partial K.$$

Indeed for a generic  $p \in \mathbb{R}^d$  such that  $\varphi(x, p) = \bar{t}$ , for some  $\bar{t} > 0$ , the positive 1-homogeneity of  $\varphi$  implies

$$\varphi(x, p) = \bar{t} \Leftrightarrow \varphi(x, \frac{p}{\bar{t}}) = 1 \Leftrightarrow \frac{p}{\bar{t}} \in \partial K.$$

Thus we restrict ourselves to  $p \in \partial K$ . Then, by convexity of  $K$ ,

$$\text{dist}(\beta p, K) > 0, \quad \text{for every } \beta > 1.$$

Let  $(\beta_k)_k$  a decreasing sequence of real numbers converging to 1. Then for every fixed  $k$ , there exists  $\eta_k > 0$  such that

$$\text{dist}(\beta_k p, K) > \eta_k > 0.$$

Since  $K_n \xrightarrow{H} K$ , there exists  $N_{\eta_k} \in \mathbb{N}$ , such that<sup>2</sup>

$$d(\beta_k p, K_n) > \eta_k > 0, \quad \text{for every } n \geq N_{\eta_k},$$

which implies  $\varphi(x_n, \beta_k p) > 1$ .

By the triangular inequality,

$$\varphi(x_n, p) \geq \varphi(x_n, \beta_k p) - \varphi(x_n, \beta_k p - p) > 1 - \frac{|\beta_k p - p|}{\alpha},$$

where the last inequality is due to Proposition 3.1.4. We infer that

$$\liminf_{n \rightarrow +\infty} \varphi(x_n, p) \geq 1 - \frac{|\beta_k - 1||p|}{\alpha}.$$

By sending  $k \rightarrow \infty$ , we get  $\liminf_{n \rightarrow \infty} \varphi(x_n, p) \geq 1$ .

We conclude the proof showing that  $\limsup_{n \rightarrow \infty} \varphi(x_n, p) \leq 1$ . Let  $n \in \mathbb{N}$ , if  $p \in K_n$ , then  $\varphi(x_n, p) \leq 1$ , if, instead,  $p \notin K_n$ , we want to estimate  $\varphi(x_n, p)$  from above. Let  $p^n \in \partial K_n$ , such that  $\text{dist}(p, K_n) = |p - p^n|$ , then:

$$\varphi(x_n, p) \leq \varphi(x_n, p^n) + \varphi(x_n, p - p^n) \leq 1 + \frac{|p - p^n|}{\alpha} \leq 1 + \frac{d(p, K_n)}{\alpha}.$$

The proof is concluded recalling that  $\text{dist}(p, K_n) \rightarrow \text{dist}(p, K) = 0$ . □

---

<sup>2</sup>In order to see this, one can use the fact that

$$d_H(K_n, K) = \sup_{p \in \mathbb{R}^d} |\text{dist}(p, K_n) - \text{dist}(p, K)|.$$

**Definition 3.1.7.** For any  $x \in \overline{\Omega}$  and for any  $q \in \mathbb{R}^d$ , we define  $\varphi^0(x, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ , the support function (see Definition 2.1.3) of  $K(x)$ , that is

$$\varphi^0(x, q) := \sup \{p \cdot q : p \in K(x)\} = \sup \{p \cdot q : \varphi(x, p) \leq 1\}.$$

*Remark 3.1.8.* For every  $x \in \overline{\Omega}$ ,  $\varphi^0(x, \cdot)$  satisfies all the properties (i), (ii), (ii). Moreover  $\varphi^0$  is a Finsler metric. Indeed, property (iii) in the definition of Finsler metric is satisfied thanks to assumption (a) and the measurability of  $\varphi^0(\cdot, q)$  follows from the continuity of  $\varphi^0(\cdot, q)$  for every  $q \in \mathbb{R}^d$  (see Proposition 3.1.10 below) and the measurability of  $\varphi(x, \cdot)$  for every  $x \in \overline{\Omega}$  is a direct consequence of the convexity.

**Lemma 3.1.9.** For every  $x \in \overline{\Omega}$  and  $q \in \mathbb{R}^d$  there exists  $p = p(x, q) \in K(x)$  such that

$$\varphi^0(x, q) = p \cdot q.$$

If  $q \in \mathbb{R}^d \setminus \{0\}$ , then  $p \in \partial K(x)$ .

Moreover, if  $K(x)$  is strictly convex and  $q \in \mathbb{R}^d \setminus \{0\}$ , then  $p = p(x, q)$  is unique.

*Proof.* The existence of  $p = p(x, q)$  follows directly from the compactness of  $K(x)$ , while the fact that  $p \in \partial K(x)$  is a consequence of the convexity.

If  $\partial K(x)$  is strictly convex, let us assume by contradiction that there exists  $p_1 \neq p_2$  such that  $\varphi^0(x, q) = p_1 \cdot q = p_2 \cdot q$ . Then, for all  $t \in (0, 1)$  we have

$$(tp_1 + (1-t)p_2) \cdot q = tp_1 \cdot q + (1-t)p_2 \cdot q = \varphi^0(x, q),$$

that means that  $tp_1 + (1-t)p_2$  belongs to  $\partial K(x)$ , contradicting the assumption of strictly convexity.  $\square$

**Proposition 3.1.10.**  $\varphi^0 : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$  is continuous. Moreover if  $K(x)$  is strictly convex for every  $x \in \Omega$ , also the function  $p : \Omega \times (\mathbb{R}^d \setminus \{0\}) \rightarrow \mathbb{R}^d$  such that  $p(x, q) \in K(x)$  and  $p(x, q) \cdot q = \varphi^0(x, q)$  is continuous.

*Proof.* Let  $((x_n, q_n))_n$  converging to  $(x, q)$ . By Lemma 3.1.9, there exists a sequence  $(p_n)_n$ , such that

$$\varphi^0(x_n, q_n) = p_n \cdot q_n.$$

Since  $K(x) \subset B(0, M)$  for all  $x \in \overline{\Omega}$ , up to the choice of a subsequence we can assume that there exists  $p \in \mathbb{R}^d$  such that  $p_n \rightarrow p$  when  $n \rightarrow +\infty$ . Clearly we have that  $p_n \cdot q_n \rightarrow p \cdot q$ . If  $q = 0$  the proof is concluded. If  $q \neq 0$ , we want to prove that

$$\varphi^0(x, q) = p \cdot q.$$

Again by Lemma 3.1.9  $p_n \in \partial K(x_n)$ . By continuity of  $\varphi$  (Corollary 3.1.5) we infer that

$$1 = \lim_{n \rightarrow +\infty} \varphi(x_n, p_n) = \varphi(x, p),$$

which implies that  $p \in \partial K(x)$  and thus that  $p \cdot q \leq \varphi^0(x, q)$ . We now claim that

$$p \cdot q \geq w \cdot q, \quad \text{for all } w \in \partial K(x).$$

Let  $w \in \partial K(x)$ , by convexity  $tw \in K(x)$  for all  $0 \leq t \leq 1$ , with  $tw \in \text{int } K$ , for  $t < 1$ . Then by Proposition 3.1.6 there exists  $\nu \in \mathbb{N}$ , such that  $tw \in K(x_n)$  for all  $n \geq \nu$ . Therefore

$\varphi^0(x_n, q) = p_n \cdot q > tw \cdot q$  for all  $n \geq \nu$ , which implies  $p \cdot q \geq tw \cdot q$ . Letting  $t$  to 1, we get the claim.

Finally if  $K(x)$  is strictly convex for every  $x$ , by the uniqueness of  $p = p(x, q) = \arg \max \{w \cdot q : w \in K(x)\}$  proved in Lemma 3.1.7 before, we have the convergence of the whole  $(p_n)$  to  $p$  which implies continuity of  $p(x, q)$ .  $\square$

**Remark 3.1.11. Supremal variational problems as problems with constraints on the gradient.**

If we consider the supremal variational problem (H), considered in Chapter 2

$$\min \left\{ \mathcal{F}_H(v, \Omega) := \operatorname{ess\,sup}_{x \in \Omega} H(x, Dv(x)) : v \in g + W^{1,\infty}(\Omega) \cap C_0(\Omega) \right\}, \quad (\text{H})$$

where  $H$  satisfies the assumptions (A), (B), defined at the beginning of Section 2.2.1, and the assumption (C'), defined in Remark 2.2.21, and such that

$$\mu := \min \left\{ \operatorname{ess\,sup}_{x \in \Omega} H(x, Dv(x)) : v \in g + W^{1,\infty}(\Omega) \cap C_0(\Omega) \right\} < +\infty,$$

then, if  $\mu > 0$ , solving (H) is equivalent to solve

$$\begin{cases} u = g & \text{on } \partial\Omega, \\ Du(x) \in K(x) = \{p : H(x, p) \leq \mu\} & \text{for a.e. } x \in \Omega. \end{cases}$$

Indeed by quasiconvexity, lower semicontinuity and coercivity

$$K(x) = \{H(x, \cdot) \leq \mu\}$$

is convex and compact.

Moreover, if  $H$  satisfies the assumption (C) (see again beginning of Section 2.2.1), we recover exactly the case where

$$K(x) = \{H(x, \cdot) \leq \mu\} \text{ and } \varphi(x, p) := \inf \left\{ t > 0 : \frac{p}{t} \in \{H(x, \cdot) \leq \mu\} \right\}$$

satisfy assumption (a) and (b). In particular the continuity of  $\varphi$  is given by Proposition 3.1.6 and Proposition 2.2.8, which ensures the convergence w.r.t. Hausdorff distance of the sublevel sets of  $H$ .

In this setting we also have  $\varphi^0 = L_\mu$  and  $d = d_\mu$ , where for every  $\lambda \geq 0$ ,  $L_\lambda$  and  $d_\lambda$  are defined by (2.2.1) and (2.2.3).

Problem (P) admits a solution if and only if the boundary values are compatible with the gradient bound. If  $K(x) = \overline{B(0, f(x))}$  this is a known fact, which was proved also by Aronsson in [11] and that we state in the following theorem.

**Theorem 3.1.12** (Lemma 2.1 in [12]). *Given a boundary datum  $g \in C(\partial\Omega)$ , if  $K(x) := \overline{B(0, f(x))}$  for every  $x \in \overline{\Omega}$ , with  $f \in C_b(\Omega)$  and  $f \geq 0$ . Then the problem (P) admits a solution  $u$ , if and only if*

$$g(y) - g(x) \leq \int_0^1 f(\gamma(t)) \left| \frac{d\gamma}{dt}(t) \right| dt,$$

for every  $\gamma \in \operatorname{path}_\Omega(x, y)$ , where  $\operatorname{path}_\Omega(x, y)$  is the set of Lipschitz curves connecting  $x$  to  $y$  defined by (2.1.5).

When  $K(x) = \overline{B(0, f(x))}$ ,  $\varphi^0(x, q) = f(x)|q|$  so it is natural to expect a similar result for a more general  $K(x)$ .

As we did in Chapter 2, we define a Finslerian structure.

For every  $x, y \in \Omega$ , we set

$$d(x, y) := \inf \left\{ \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}(x, y) \right\}, \quad (3.1.4)$$

where  $\text{path}(x, y) := \text{path}_\Omega(x, y)$  defined in (2.1.5). Moreover and for any  $x, y \in \partial\Omega$  we set

$$d(x, y) := \inf \left\{ \liminf_{n \rightarrow +\infty} d(x_n, y_n) : (x_n)_n, (y_n)_n \in \Omega^{\mathbb{N}} \text{ and } x_n \rightarrow x, y_n \rightarrow y \right\}.$$

*Remark 3.1.13.* Since  $\varphi^0$  is a Finsler metric,  $d : \overline{\Omega} \times \overline{\Omega} \rightarrow [0, +\infty]$  defined by (3.1.4) has all the properties pointed out in Section 2.1.2, in particular in Remark 2.1.20, Proposition 2.1.24, and Corollary 2.1.26.

The extended version of Theorem 3.1.12 is provided by the following theorem.

**Theorem 3.1.14.** *Given a boundary datum  $g \in C(\partial\Omega)$ , there exists a solution  $u$  of (P) if and only if*

$$g(y_2) - g(y_1) \leq d(y_1, y_2) \quad \text{for all } y_1, y_2 \in \partial\Omega \quad (1\text{-Lip})$$

In analogy with Definition 2.2.25, we consider the maximal and minimal extensions defined respectively by

$$S^-(x) = \sup \{g(y) - d(x, y) : y \in \partial\Omega\}, \quad (3.1.5)$$

$$S^+(x) = \inf \{g(y) + d(y, x) : y \in \partial\Omega\}. \quad (3.1.6)$$

**Proposition 3.1.15.** *Let us assume that the boundary datum  $g$  of (P) satisfies (1-Lip). Then the following facts hold true:*

1.  $S^+(y) = S^-(y) = g(y)$  for every  $y \in \partial\Omega$ ;
2. For every  $x, y \in \overline{\Omega}$

$$S^+(y) - S^+(x) \leq d(x, y) \text{ and } S^-(y) - S^-(x) \leq d(x, y);$$

*Proof.* The proof of (1) follows directly from the definition of  $S^+$  and  $S^-$  and by  $g$  satisfying (1-Lip).

Let us prove (2). Let  $x \in \overline{\Omega}$ . Since  $g$  is continuous,  $d$  is lower semicontinuous (see Proposition 2.1.24) and  $\overline{\Omega}$  is a compact set, there exists  $x_1 \in \partial\Omega$ , such that  $S^+(x) = g(x_1) + d(x_1, x)$ . By definition of  $S^+$ , we obtain:

$$S^+(y) - S^+(x) \leq g(x_1) + d(x_1, y) - g(x_1) - d(x_1, x) = d(x_1, y) - d(x_1, x).$$

If  $x \in \partial\Omega$ , thanks to point (1), we have

$$S^+(y) - S^+(x) = S^+(y) - g(x) \leq d(x, y),$$

where the second inequality is a direct consequence of the definition of  $S^+$ .

If  $x \in \Omega$ , we can apply the triangular inequality (see (4) in Remark 2.1.20) to have  $d(x_1, y) \leq d(x_1, x) + d(x, y)$ . □

**Theorem 3.1.16.** *Let  $u : \bar{\Omega} \rightarrow \mathbb{R}$  be such that*

$$u(y) - u(x) \leq d(x, y) \quad \text{for all } x, y \in \Omega.$$

*Then  $u \in W^{1,\infty}(\Omega) \cap C(\Omega)$  and  $Du(x) \in K(x)$  for a.e.  $x \in \Omega$  (in particular  $Du(x) \in K(x)$  for any point  $x$  at which  $u$  is differentiable).*

*If, in addition,  $u = g$  on  $\partial\Omega$ , then  $u \in C(\bar{\Omega})$  and  $u(y) - u(x) \leq d(x, y)$  for all  $x, y \in \bar{\Omega}$ . Viceversa, if  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$  such that  $Du(x) \in K(x)$  a.e. in  $\Omega$ , then*

$$u(y) - u(x) \leq d(x, y) \quad \text{for all } x, y \in \bar{\Omega}.$$

*Proof.* This result is analogous to Theorem 2.2.22 and the proof is the same. □

**Corollary 3.1.17.** *If  $g$  satisfies (1-Lip),  $S^+, S^-$  are solutions of (P).*

*Proof.* The proof follows directly from Proposition 3.1.15 (1) and (2) and Theorem 3.1.16. □

*Proof of Theorem 3.1.14.* The proof of Theorem 3.1.14 is a direct consequence of Corollary 3.1.17 and Theorem 3.1.16. □

**Proposition 3.1.18.** *If  $u$  is a solution of (P), then*

$$S^-(x) \leq u(x) \leq S^+(x), \quad \text{for every } x \in \bar{\Omega}.$$

*Proof.* Let us prove that  $u(x) \leq S^+(x)$  for all  $x \in \Omega$ . The case  $S^-(x) \leq u(x)$  is similar. Since by Theorem 3.1.16  $u$  is 1-Lipschitz w.r.t.  $d$ , then the thesis is straightforward. Indeed we have

$$u(x) - g(y) = u(x) - u(y) \leq d(y, x) \quad \text{for every } y \in \partial\Omega,$$

that implies  $u(x) \leq \inf\{g(y) + d(y, x)\}$ . □

*Remark 3.1.19.* Exactly as in Remark 2.2.24, a consequence of Theorem 3.1.16 is that

$$d(x, y) = \delta_{\varphi^0}(x, y)$$

for every  $x, y \in \Omega$ , where

$$\begin{aligned} \delta_{\varphi^0}(x, y) &= \sup \left\{ u(y) - u(x) : u \in W^{1,\infty}(\Omega), \operatorname{ess\,sup}_{x \in \Omega} \varphi^{00}(x, Du(x)) \leq 1 \right\} \\ &= \sup \left\{ u(y) - u(x) : u \in W^{1,\infty}(\Omega), \operatorname{ess\,sup}_{x \in \Omega} \varphi(x, Du(x)) \leq 1 \right\}, \end{aligned} \quad (3.1.7)$$

which is a natural way to define a distance starting from a Finsler metric (see [60, 72]).

*Proof.* The inequality “ $\geq$ ” is trivial using Theorem 3.1.16 and the definition of  $\varphi$ . To recover the converse inequality it is enough  $u(\cdot) = d(z, \cdot)$ , which is 1-Lipschitz thanks to the triangular inequality, and apply again Theorem 3.1.16. □

## 3.2 The uniqueness set

**Assumptions.** In this section we will assume that

(a\*) the map  $(x, p) \mapsto \varphi(x, p)$  is continuous on  $\overline{\Omega} \times \mathbb{R}^d$ .

Assumption a\* implies the continuity of  $K : \overline{\Omega} \rightarrow \mathcal{K}$  thanks to Proposition 3.1.6 and of  $\varphi^0 : \overline{\Omega} \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  thanks to Proposition 3.1.10. Indeed in both cases, it is enough to do the proofs with  $x \in \overline{\Omega}$ .

### 3.2.1 The structure of the uniqueness set

We start with some results about the (pseudo) distance  $d$ .

**Proposition 3.2.1.** *Let  $x, y \in \Omega$ . Then there exists  $\gamma \in \text{path}_{\overline{\Omega}}(x, y)$ , such that*

$$d(x, y) = \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt.$$

We say that such a  $\gamma$  is a  $d$ -geodesic connecting  $x$  to  $y$ .

*Proof.* We first observe that  $|x - y|_{\Omega} < +\infty$  (see Remark 2.1.18) implies, thanks to Corollary 2.1.26, that  $d(x, y) < +\infty$ . We then consider a sequence  $(\gamma_n)_{n \in \mathbb{N}} \subset \text{path}(x, y)$ , converging to the infimum, i.e. such that

$$\lim_{n \rightarrow \infty} \int_0^1 \varphi^0(\gamma_n(t), \dot{\gamma}_n(t)) dt = d(x, y), \quad (3.2.1)$$

Thanks to the fact that

$$\int_0^1 \varphi^0(\gamma_n(t), \dot{\gamma}_n(t)) dt \geq \alpha \int_0^1 |\dot{\gamma}_n(t)| dt,$$

for every  $n$  and for some  $\alpha > 0$  (which exists thanks to assumption (a) on  $K(x)$ ), we can also assume without loss of generality that  $|l(\gamma_n)| \leq L$ , for all  $n$ , for some  $L \geq 0$ , where  $l(\gamma_n)$  is the Euclidian length of each  $\gamma_n$ . Then there exists a subsequence of  $(\gamma_n)_n$ , that we will still call  $(\gamma_n)$ , such that

1.  $|\gamma_n(t)| \leq C$ , for every  $s \in [0, 1]$ , for some  $C \geq 0$ , since  $\Omega$  is bounded;
2.  $|\dot{\gamma}_n(t)| \leq L$  a.e. in  $[0, 1]$ .

By Arzelà-Ascoli Theorem, there exists a subsequence such that  $\gamma_n$  converges uniformly to some curve  $\gamma$ , such that  $\gamma([0, 1] \subset \overline{\Omega})$  and  $|\dot{\gamma}(t)| \leq L$  a.e. in  $(0, 1)$ . Point (b) implies that  $\|\dot{\gamma}_n\|_{L^2((0,1))}$  are equibounded and therefore that  $\dot{\gamma}_n \rightharpoonup \dot{\gamma}$  in  $L^2((0, 1))$ .

In order to prove that  $\gamma$  is a minimizing curve, we observe that by definition of  $\varphi^0$  and thanks, again, to assumption (a) on  $K(x)$ , i.e. there exist  $0 < \alpha < M$  such that  $B(0, \alpha) \subset K(x) \subset B(0, M)$  for every  $x \in \overline{\Omega}$ , given  $q_1, q_2 \in \mathbb{R}^d$ , we have that

$$\begin{aligned} \varphi^0(x, q_1) &\leq M|q_1| \quad \text{and} \\ \varphi^0(x, q_2) &\geq \alpha \frac{q_1}{|q_1|} \cdot q_2, \end{aligned}$$

for any,  $x \in \bar{\Omega}$  and hence

$$\varphi^0(x, q_2) - \varphi^0(x, q_1) \geq \alpha \frac{q_1}{|q_1|} \cdot q_2 - M|q_1| = (\alpha - M) \frac{q_1}{|q_1|} \cdot (q_2 - q_1).$$

We thus get the following estimate

$$\int_0^1 \varphi^0(\gamma_n(t), \dot{\gamma}_n(t)) dt \geq \int_0^1 \varphi^0(\gamma_n(t), \dot{\gamma}(t)) dt + \int_0^1 (\alpha - M) \frac{\dot{\gamma}(t)}{|\dot{\gamma}(t)|} \cdot (\dot{\gamma} - \dot{\gamma}_n) dt,$$

thanks to which we can conclude

$$\liminf_{n \rightarrow \infty} \int_0^1 \varphi^0(\gamma_n(t), \dot{\gamma}_n(t)) dt \geq \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt.$$

Indeed

$$\lim_{n \rightarrow \infty} \int_0^1 (\alpha - M) \frac{\dot{\gamma}(t)}{|\dot{\gamma}(t)|} \cdot (\dot{\gamma} - \dot{\gamma}_n) dt = 0,$$

by the fact that  $\dot{\gamma}_n \rightharpoonup \dot{\gamma}$  in  $L^2((0, 1))$ , and

$$\lim_{n \rightarrow \infty} \int_0^1 \varphi^0(\gamma_n(t), \dot{\gamma}(t)) dt = \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt, \quad (3.2.2)$$

by (uniform) convergence of  $\gamma_n$  toward  $\gamma$  and by continuity of  $\varphi^0$ .  $\square$

**Proposition 3.2.2.** *Let  $x, y \in \bar{\Omega}$  such that  $d(x, y) < +\infty$ . Then there exists  $\gamma \in \text{path}_{\bar{\Omega}}(x, y)$ , such that*

$$d(x, y) = \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt.$$

*We say that such a  $\gamma$  is a  $d$ -geodesic connecting  $x$  to  $y$ .*

*Proof.* If  $x, y \in \Omega$ , we conclude thanks to Proposition 3.2.1. Let  $x \in \partial\Omega, y \in \bar{\Omega}$  such that  $d(x, y) < +\infty$  and let  $(x_n)_n, (y_n)_n$  two sequences (they can be constructed for instance by a diagonal argument) converging respectively to  $x$  and  $y$  such that

$$d(x, y) = \lim_{n \rightarrow +\infty} d(x_n, y_n).$$

For every  $n$  we consider  $\gamma_n \in \text{path}_{\bar{\Omega}}(x_n, y_n)$ , reparametrized by arc length, such that

$$d(x_n, y_n) = \int_0^{l(\gamma_n)} \varphi^0(\gamma_n(s), \dot{\gamma}_n(s)) ds,$$

obtaining that

$$d(x, y) = \lim_{n \rightarrow \infty} \int_0^{l(\gamma_n)} \varphi^0(\gamma_n(s), \dot{\gamma}_n(s)) ds. \quad (3.2.3)$$

Finally we apply the same reasoning as in the proof of Proposition 3.2.1 to  $(\gamma_n)_n$  and, up to the choice of subsequences, we get a curve  $\gamma \in \text{path}_{\bar{\Omega}}(x, y)$ , such that

$$\lim_{n \rightarrow \infty} \int_0^{l(\gamma_n)} \varphi^0(\gamma_n(s), \dot{\gamma}_n(s)) ds = \int_0^{l(\gamma)} \varphi^0(\gamma(s), \dot{\gamma}(s)) ds.$$

$\square$

*Remark 3.2.3.* From Proposition 3.2.1 and Proposition 3.2.2 we infer that, for every  $x, y \in \overline{\Omega}$  such that  $d(x, y) < +\infty$

$$d(x, y) \geq \inf \left\{ \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}_{\overline{\Omega}}(x, y) \right\}.$$

We point out, as shown in the example below, that it might happen

$$d(x, y) > \inf \left\{ \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}_{\overline{\Omega}}(x, y) \right\}.$$

*Example 3.2.4.* Let  $\Omega, K(x), x = (x_1, x_2)$  and  $y = (y_1, y_2)$  be as in Example 2.1.21. Then

$$\begin{aligned} d(x, y) &= \inf \left\{ \int_0^1 |\dot{\gamma}(t)| dt : \gamma \in \text{path}_{\Omega}(x, y) \right\} \\ &= \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} = \sqrt{2}, \end{aligned}$$

while

$$\inf \left\{ \int_0^1 |\dot{\gamma}(t)| dt : \gamma \in \text{path}_{\overline{\Omega}}(x, y) \right\} = 1.$$

**Proposition 3.2.5.** *If  $\Omega$  has a Lipschitz boundary then, for every  $x, y \in \overline{\Omega}$ ,  $d(x, y) < +\infty$  and*

$$d(x, y) = \inf \left\{ \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}_{\overline{\Omega}}(x, y) \right\}. \quad (3.2.4)$$

*Proof.* The first part of the statement is a consequence of the second part of Proposition 2.2.5.

As for the second part, since  $\partial\Omega$  is Lipschitz (see Definition 2.1.22) by Lemma 3.2.6 below, for any given  $\gamma \in \text{path}_{\overline{\Omega}}(x, y)$  it is possible to locally approximate  $\gamma([0, 1]) \cap \partial\Omega$  by a sequence of Lipschitz curves  $(\gamma_n)_n$  such that  $\gamma_n([0, 1]) \subset \Omega$  and therefore, by compactness it is possible to approximate the entire curve  $\gamma$  by a sequence of Lipschitz curves  $(\gamma_n)_n$  such that  $\gamma_n([0, 1]) \subset \Omega$ .

Hence, given  $x, y \in \overline{\Omega}$ , if  $\gamma$  is such that

$$\int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt \leq \inf \left\{ \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}_{\overline{\Omega}}(x, y) \right\} + \varepsilon$$

and  $(\gamma_n)_n$  is a converging sequence, we have

$$\int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt = \lim_n \int_0^1 \varphi^0(\gamma_n(t), \dot{\gamma}_n(t)) dt \geq \lim d(x_n, y_n) = d(x, y).$$

By the arbitrariness of  $\varepsilon$  we get

$$d(x, y) \leq \inf \left\{ \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt : \gamma \in \text{path}_{\overline{\Omega}}(x, y) \right\}.$$

The other inequality follows by Proposition 3.2.1 and Proposition 3.2.2 and it is pointed out in Remark 3.2.3. □

### 3.2. THE UNIQUENESS SET

---

**Lemma 3.2.6.** *Let  $\Omega$  bounded and  $\partial\Omega$  be Lipschitz (see Definition 2.1.22). Let  $x \in \partial\Omega$  and let  $A \subset \Omega$  be a neighborhood of  $x$ ,  $\rho, h > 0$  and  $\phi : B^{d-1}(\mathbf{0}, \rho) \rightarrow \mathbb{R}$  be a Lipschitz function such that*

$$\begin{aligned}\partial\Omega \cap A &= \{(y', \phi(y')) : y' \in B(\mathbf{0}, \rho)\} \\ \Omega \cap A &= \{(y', y_d) : y' \in B^{d-1}(\mathbf{0}, \rho), h > y_d > \phi(x')\}.\end{aligned}$$

*Then for every Lipschitz curve  $\gamma : [0, 1] \rightarrow \partial\Omega \cap B(x, r)$  there exists a sequence of curves  $(\gamma_n)$ , such that*

1.  $\gamma_n([0, 1]) \subset \Omega \cap A$ ,
2.  $\gamma_n \rightarrow \gamma$  uniformly,
3.  $\dot{\gamma}_n \rightarrow \dot{\gamma}$  a.e. in  $(0, 1)$ .

*Proof.* We take a sequence of functions  $\phi_n : B^{d-1}(\mathbf{0}, \rho) \rightarrow \mathbb{R}$  such that  $\phi_n(x) = \phi(x) + \frac{1}{n}$ . Let  $\pi((x', x_d) = x'$  the projection of  $\mathbb{R}^d$  in  $\mathbb{R}^{d-1}$ . We define

$$\gamma_n(s) := \gamma(s) + (\mathbf{0}, \phi_n(\pi(\gamma(s)))) - (\mathbf{0}, \phi(\pi(\gamma(s)))).$$

Clearly, for  $n$  large enough,  $\gamma_n([0, 1]) \subset \Omega \cap A$  and  $\gamma_n \rightarrow \gamma$  uniformly. Moreover

$$\begin{aligned}\dot{\gamma}_n(s) &= \dot{\gamma}(s) + (\mathbf{0}, D\phi_n(\pi \circ \gamma)^\top \cdot D\pi(\gamma(s))\dot{\gamma}(s)) - (\mathbf{0}, D\phi(\pi \circ \gamma)^\top \cdot D\pi(\gamma(s))\dot{\gamma}(s)) = \\ &= \dot{\gamma}(s) + (\mathbf{0}, (D\phi_n(\pi \circ \gamma) - D\phi(\pi \circ \gamma))^\top \cdot \pi(\dot{\gamma}(s))).\end{aligned}$$

Then

$$|\gamma_n(s) - \gamma(s)| \leq \|D\phi_n - D\phi\|_{L^\infty} |\dot{\gamma}(s)|.$$

□

**Proposition 3.2.7.** *Let us assume that the boundary datum  $g$  of (P) satisfies (1-Lip). Then*

1. *for every  $x \in \Omega$  there exist  $y_1, y_2 \in \partial\Omega$  and  $\gamma_1 \in \text{path}_{\overline{\Omega}}(y_1, x), \gamma_2 \in \text{path}_{\overline{\Omega}}(x, y_2)$  with*

$$S^+(x) = g(y_1) + \int_0^1 \varphi^0(\gamma_1(t), \dot{\gamma}_1(t)) dt,$$

*such that  $\gamma_1(t) \in \Omega$  for every  $t \in (0, 1]$  and*

$$S^-(x) = g(y_2) - \int_0^1 \varphi^0(\gamma_2(t), \dot{\gamma}_2(t)) dt,$$

*such that  $\gamma_2(t) \in \Omega$  for every  $t \in [0, 1)$ . We will refer to such  $y_1, \gamma_1$  and  $y_2, \gamma_2$  as optimal respectively for  $S^+(x)$  and  $S^-(x)$ ;*

2. *if  $y_1 \in \partial\Omega$  and  $\gamma_1 \in \text{path}_{\overline{\Omega}}(y_1, x)$  are optimal for  $S^+(x)$ , then for every  $\hat{t} \in (0, 1)$ , if  $z_1 = \gamma_1(\hat{t})$ , we have*

$$S^+(z_1) = g(y_1) + \int_0^{\hat{t}} \varphi^0(\gamma_1(t), \dot{\gamma}_1(t)) dt, \quad (3.2.5)$$

*and if  $y_2 \in \partial\Omega$  and  $\gamma_2 \in \text{path}_{\overline{\Omega}}(y_2, x)$  are optimal for  $S^-(x)$ , then for every  $\hat{t} \in (0, 1)$ , if  $z_2 = \gamma_2(\hat{t})$ ,*

$$S^-(z_2) = g(y_2) + \int_0^{\hat{t}} \varphi^0(\gamma_2(t), \dot{\gamma}_2(t)) dt. \quad (3.2.6)$$

*Proof.* We start with the proof of (1). We first claim that for any  $x \in \Omega$  there exist  $y \in \partial\Omega$  and  $\gamma \in \text{path}_{\overline{\Omega}}(x, y)$ , such that

$$S^+(x) = g(y) + \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt. \quad (3.2.7)$$

Since  $g$  is continuous,  $d$  is lower semicontinuous (see Proposition 2.1.24) and  $\overline{\Omega}$  is a compact set, there exists  $y \in \partial\Omega$  such that

$$S^+(x) = g(y) + d(y, x).$$

Now, the existence of an optimal curve is given by Proposition 3.2.2 and (3.2.7) is proven. We then show that it is possible to find a curve that solves (3.2.7) and it is all contained in  $\Omega$  except from the first point. Let  $\hat{y} = \gamma(\hat{t})$  for some  $\hat{t} \in (0, 1]$ , be the last point of the boundary touched by  $\gamma$ , then  $\gamma : [\hat{t}, 1] \rightarrow \overline{\Omega}$  touches the boundary only in  $\hat{t}$ . We show that

$$g(y) + \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt = g(\hat{y}) + \int_{\hat{t}}^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt.$$

By definition of  $S^+$  we know that “ $\leq$ ” holds. Let us assume by contradiction that the inequality is strict. Then

$$d(y, \hat{y}) \leq \int_0^{\hat{t}} \varphi^0(\gamma(t), \dot{\gamma}(t)) dt < g(\hat{y}) - g(y), \quad (3.2.8)$$

which contradicts (1-Lip).

We conclude by showing (2). Let  $y_1 \in \partial\Omega$  and  $\gamma_1$  be optimal for the definition of  $S^+(x)$  and let  $z_1 = \gamma(\hat{t})$ , for some  $\hat{t} \in (0, 1)$ . We prove that

$$S^+(z_1) = g(y_1) + \int_0^{\hat{t}} \varphi^0(\gamma_1(t), \dot{\gamma}_1(t)) dt = g(y_1) + d(y_1, z).$$

Assume by contradiction that there exist  $\hat{y} \in \partial\Omega$  and  $\eta \in \text{path}_{\Omega}(\hat{y}, z)$  such that

$$g(\hat{y}) + \int_0^1 \varphi^0(\eta(t), \dot{\eta}(t)) ds < g(y_1) + \int_0^{\hat{t}} \varphi^0(\gamma_1(t), \dot{\gamma}_1(t)) dt. \quad (3.2.9)$$

Then if we call  $\hat{\gamma}$  the piece of  $\gamma$  connecting  $z_1$  to  $x$  and  $\hat{\eta}$  the reparametrized version of the curve obtained by gluing  $\eta$  and  $\hat{\gamma}$ , we have

$$g(\hat{y}) + \int_0^1 \varphi^0(\hat{\eta}(t), \dot{\hat{\eta}}(t)) dt < g(y_1) + \int_0^1 \varphi^0(\gamma_1(t), \dot{\gamma}_1(t)) dt,$$

contradicting the optimality of  $y$  and  $\gamma$ .

The proof of (1) and (2) for  $S^-$  is analogous. □

**Theorem 3.2.8.** *Let us assume that the boundary datum  $g$  of (P) satisfies (1-Lip) and let  $x \in \Omega$  such that  $S^+(x) = S^-(x)$ . Then for every  $y_1, y_2 \in \partial\Omega$  and  $\gamma_1 \in \text{path}_{\overline{\Omega}}(y_1, x), \gamma_2 \in$*

### 3.2. THE UNIQUENESS SET

---

$\text{path}_{\overline{\Omega}}(x, y_2)$  which are respectively optimal for  $S^+(x)$  and  $S^-(x)$  in the sense of Proposition 3.2.7, if one defines the curve given by the concatenation of  $\gamma_1$  and  $\gamma_2$ ,

$$\gamma(t) := \begin{cases} \gamma_1(2t) & \text{for } 0 \leq t \leq \frac{1}{2} \\ \gamma_2(2t-1) & \text{for } \frac{1}{2} < t \leq 1 \end{cases}, \quad (3.2.10)$$

it holds that  $\gamma \in \text{path}_{\overline{\Omega}}(y_1, y_2)$  and  $\gamma$  is a  $d$ -geodesic.

For every  $z \in \gamma$ <sup>3</sup> and for every solution  $u$  of (P), it also holds

$$S^-(z) = u(z) = S^+(z). \quad (3.2.11)$$

Finally,  $S^+, S^-$  and every solution  $u$  of (P) are derivable along  $\gamma$  for  $\mathcal{H}^1$  a.e. point of  $\gamma$ , and  $\nabla_{\gamma} S^+ = \nabla_{\gamma} S^- = \nabla_{\gamma} u = \varphi^0(\gamma, \dot{\gamma})$ .

*Proof.* Let  $x, y_1, y_2, \gamma_1$  and  $\gamma_2$  be as in the assumptions. Then  $S^-(x) = S^+(x)$  can be rewritten as

$$g(y_1) + \int_0^1 \varphi^0(\gamma_1(t), \dot{\gamma}_1(t)) dt = g(y_2) - \int_0^1 \varphi^0(\gamma_2(t), \dot{\gamma}_2(t)) dt,$$

which implies

$$\begin{aligned} d(y_1, y_2) &\geq g(y_2) - g(y_1) = \\ &= \int_0^1 \varphi^0(\gamma_1(t), \dot{\gamma}_1(t)) dt + \int_0^1 \varphi^0(\gamma_2(t), \dot{\gamma}_2(t)) dt = \\ &= \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt \geq d(y_1, y_2), \end{aligned} \quad (3.2.12)$$

proving that  $\gamma$  is a  $d$ -geodesic connecting  $y_1$  to  $y_2$ .

For every  $z \in \gamma_2$ , thanks to Proposition 3.2.7 (2)

$$S^-(z) - S^-(x) = g(y_2) - d(z, y_2) - g(y_2) + d(x, y_2) = d(x, z). \quad (3.2.13)$$

Analogously, if  $z \in \gamma_1$ ,  $S^+(x) - S^+(z) = d(z, x)$ .

Let now  $z_1 \in \gamma_1$  and  $z_2 \in \gamma_2$ , then: .

$$\begin{aligned} d(z_1, z_2) &\geq S^-(z_2) - S^-(z_1) \geq S^-(z_2) - S^+(z_1) = \\ &= S^-(z_2) - S^-(x) + S^+(x) - S^+(z_1) = d(x, z_2) + d(z_1, x) \geq d(z_1, z_2). \end{aligned}$$

This proves that  $S^+(z_1) = S^-(z_1)$ , switching the role of  $z_1$  and  $z_2$  we get  $S^+(z_2) = S^-(z_2)$  and thus that  $S^+$  and  $S^-$  coincide along  $\gamma$ . Therefore

$$S^- = u = S^+ \text{ along } \gamma,$$

for every solution  $u$  of (P), by Proposition 3.1.18.

Moreover, if  $\tau_1$  is such that  $\gamma_1 \ni z_1 = \gamma(\tau_1)$  and  $\tau_2$  such that  $\gamma_2 \ni z_2 = \gamma(\tau_2)$ , we claim that

$$S^-(z_1) = g(y_2) - \int_{\tau_1}^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt \quad \text{and} \quad (3.2.14)$$

$$S^+(z_2) = g(y_1) + \int_0^{\tau_2} \varphi^0(\gamma(t), \dot{\gamma}(t)) dt. \quad (3.2.15)$$

---

<sup>3</sup>with a little abuse of notation we identify  $\gamma([0, 1])$  with  $\gamma$ .

Let us prove (3.2.14), (3.2.15) can be proven similarly. By the definition of  $S^-(z_1)$  it is sufficient to prove that

$$g(y_2) - \int_{\tau_1}^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt \geq S^-(z_1). \quad (3.2.16)$$

Since we know that

$$S^-(z_1) = S^+(z_1) = g(y_1) + \int_0^{\tau_1} \varphi^0(\gamma(t), \dot{\gamma}(t)) dt,$$

where the last equality holds again by Proposition (3.2.7) (2), the inequality (3.2.16) holds if and only if

$$g(y_2) - g(y_1) \geq \int_0^1 \varphi^0(\gamma(t), \dot{\gamma}(t)) dt,$$

that is true by (3.2.12). We point out that this result extends at all the points of  $\gamma$  what we proved in Proposition (3.2.7) (2) for the points of  $\gamma_1$  and  $\gamma_2$ .

We conclude showing that the derivative along  $\gamma$  of  $S^+$  and  $S^-$  (and also of any solution  $u$  of (P)) exists at every point at which the curve is differentiable and it is equal to  $\varphi^0$ . Let  $\bar{t} \in (0, 1)$  a Lebesgue point for  $\varphi^0(\gamma(t), \dot{\gamma}(t))$ , such that the curve is differentiable at  $\bar{t}$ , then for all what we proved above we can write

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{S^+(\gamma(\bar{t} + h)) - S^+(\gamma(\bar{t}))}{h} = \\ & = \lim_{h \rightarrow 0} \frac{1}{h} \left( g(y_1) + \int_0^{\bar{t}+h} \varphi^0(\gamma(t), \dot{\gamma}(t)) dt - g(y_1) - \int_0^{\bar{t}} \varphi^0(\gamma(t), \dot{\gamma}(t)) dt \right). \end{aligned}$$

Which is equal to

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_{\bar{t}}^{\bar{t}+h} \varphi^0(\gamma(t), \dot{\gamma}(t)) dt = \varphi^0(\gamma(\bar{t}), \dot{\gamma}(\bar{t})).$$

□

**Definition 3.2.9.** We will refer to the set of points

$$\mathcal{U} := \{x \in \Omega : S^+(x) = S^-(x)\}$$

as *uniqueness set*. What we have proved in Theorem 3.2.8, is that for every point  $x \in \mathcal{U}$  there exists a Lipschitz curve all contained in  $\Omega$  except for its extreme points that is a  $d$ -geodesic in the sense of Proposition 3.2.2.

### 3.2.2 Regularity of solutions on the uniqueness set

In this section we will study the regularity of solutions of (P) in the uniqueness set. More precisely, we will show that such solutions of (P) are both locally semiconcave and locally semiconvex at each point of the uniqueness set and therefore differentiable. Moreover, if the interior part of the uniqueness set is not empty, every solution is locally  $C^{1,1}$ .

**Assumptions.** In this section we will assume that the boundary datum  $g$  of (P) satisfies

$$g(y_2) - g(y_1) \leq d(y_1, y_2) \quad \text{for all } y_1, y_2 \in \partial\Omega. \quad (1\text{-Lip})$$

Let us recall some definitions (see [37]) .

**Definition 3.2.10** (Semiconcavity and semiconvexity (see Def 1.1.1 [37])). We say that  $u_1, u_2 : \Omega \rightarrow \mathbb{R}^d$  are respectively semiconcave and semiconvex, if there exist  $C_1, C_2 \geq 0$  such that

$$u(x+h) + u(x-h) - 2u(x) \leq C_1|h|^2, \quad (3.2.17)$$

$$u(x+h) + u(x-h) - 2u(x) \geq C_2|h|^2, \quad (3.2.18)$$

for every  $x \in \Omega$ ,  $h \in \mathbb{R}^d$  such that the segment  $[x-h, x+h] \subset \Omega$ .

We say that  $u_1, u_2$  are respectively locally semiconcave and locally semiconvex, if for every  $V \subset\subset \Omega$ , there exist  $C_1(V), C_2(V) \geq 0$  such that (3.2.17) and (3.2.18) hold for every  $x \in V$ ,  $h \in \mathbb{R}^d$  such that the segment  $[x-h, x+h] \subset V$ .

**Proposition 3.2.11.** *Let us assume that  $\varphi^0 \in C^2(\Omega \times (\mathbb{R}^d \setminus \{0\}))$ . Then  $S^+$  and  $S^-$  are respectively locally semiconcave and locally semiconvex.*

*Proof.* The proof of this fact is strongly inspired to the proof of Lemma 5.1 in [12]. We prove (3.2.17) for  $S^+$ . The proof of (3.2.18) for  $S^-$  is analogous. Let  $V \subset\subset \Omega$  and  $\delta > 0$  such that  $\text{dist}(x, \partial\Omega) > 8\delta$  for all  $x \in V$ . Since  $S^+$  is bounded, if  $|h| \geq \delta$  the inequality (3.2.17) is easily verified. Let us then assume that  $|h| < \delta$ . Let  $x$  in  $V$  and  $y_1 \in \partial\Omega$ ,  $\gamma_1 \in \text{path}_{\bar{\Omega}}(y_1, x)$ , reparametrized by arc length, such that

$$S^+(x) = g(y_1) + \int_0^{l(\gamma_1)} \varphi^0(\gamma_1(s), \dot{\gamma}_1(s)) ds.$$

We define  $a := \sup\{t \in [0, l(\gamma_1)] : \text{dist}(\gamma_1(t), \partial\Omega) \leq 4\delta\}$ . We then consider

$$y(s) := \frac{s-a}{l(\gamma_1)-a} \frac{h}{|h|} \quad \text{for } a \leq s \leq l(\gamma_1)$$

and

$$\bar{\gamma}(s) := \begin{cases} \gamma_1(s) & \text{for } 0 \leq s \leq a \\ \gamma_1(s) + ty(s) & \text{for } a \leq s \leq l(\gamma_1), \end{cases}$$

where  $t \in [-|h|, |h|]$  is some parameter at our disposal. Notice that by construction  $l(\gamma_1) - a > 4\delta$  and that  $\bar{\gamma}(s) \in \Omega$  for every  $a \leq s \leq l(\gamma_1)$ , indeed  $\text{dist}(\bar{\gamma}(s), \partial\Omega) > 3\delta$  for  $t \in [-|h|, |h|]$ .

We define

$$\begin{aligned} F(t) &:= \int_0^{l(\gamma_1)} \varphi^0(\bar{\gamma}(s), \dot{\bar{\gamma}}(s)) ds \\ &= \int_0^a \varphi^0(\gamma_1(s), \dot{\gamma}_1(s)) ds + \int_a^{l(\gamma_1)} \varphi^0(\gamma_1(s) + ty(s), \dot{\gamma}_1(s) + t\dot{y}(s)) ds \end{aligned}$$

and we observe that  $F(0) = S^+(x)$  and  $F^+(\pm|h|) \geq S^+(x \pm h)$ . In order to prove (3.2.17) it is then enough to prove that there exists  $C_1 \leq 0$  such that

$$F(|h|) + F(-|h|) - 2F(0) \leq |h|^2 C_1.$$

Thanks to the assumptions on  $\varphi^0$  we have

$$F'(t) = \int_a^{l(\gamma_1)} \nabla_x \varphi^0(\bar{\gamma}(s), \dot{\bar{\gamma}}(s)) \cdot y(s) + \nabla_p \varphi^0(\bar{\gamma}(s), \dot{\bar{\gamma}}(s)) \cdot \dot{y}(s) ds$$

and

$$F''(t) = \int_a^{l(\gamma_1)} (y(s), \dot{y}(s))^\top \nabla^2 \varphi^0(\bar{\gamma}(s), \dot{\bar{\gamma}}(s)) \cdot (y(s), \dot{y}(s)) ds, \quad (3.2.19)$$

where  $|F''(t)| \leq \frac{C_1}{2}$  for some  $C_1 \geq 0$  for every  $t \in [-|h|, |h|]$ , by continuity of  $\nabla^2 \varphi^0$ . The proof is concluded thanks to the Taylor's formula with the rest of Lagrange, for which there exist  $\theta_1 \in [0, |h|]$  and  $\theta_2 \in [-|h|, 0]$  such that

$$F(|h|) + F(-|h|) - 2F(0) = (F''(\theta_1) + F''(\theta_2)) |h|^2.$$

□

The following is a partial regularity result on  $\varphi^0$ , which states that continuity of  $\nabla_x \varphi$  implies continuity of  $\nabla_x \varphi^0$ .

**Proposition 3.2.12.** *If  $K(x)$  is strictly convex for all  $x \in \bar{\Omega}$ ,  $\varphi(\cdot, q)$  differentiable w.r.t.  $x$  for every  $q \in \mathbb{R}^d \setminus \{0\}$  and  $\nabla_x \varphi(x, p) \in C((\Omega \times (\mathbb{R}^d \setminus \{0\})); \mathbb{R}^d)$ , then also  $\nabla_x \varphi^0(x, q) \in C((\Omega \times (\mathbb{R}^d \setminus \{0\})); \mathbb{R}^d)$ .*

*Proof.* We first want to prove that the limit

$$\lim_{h \rightarrow 0} \frac{\varphi^0(x + he_i, q) - \varphi^0(x, q)}{h} \quad (3.2.20)$$

exists and is finite for any  $i = 1, \dots, d$ . We prove it for  $i = 1$  and we fix  $\bar{h} = he_1$ , in order to simplify the notation. By Lemma(3.1.9), there exist unique  $p(x + \bar{h}, q) = p(x + \bar{h})$  and  $p(x, q) = p(x)$ , such that

$$\begin{aligned} \varphi^0(x + \bar{h}, q) &= \max\{p \cdot q \mid \varphi(x + \bar{h}, p) = 1\} = p(x + \bar{h}) \cdot q, \\ \varphi^0(x, q) &= \max\{p \cdot q \mid \varphi(x, p) = 1\} = p(x) \cdot q. \end{aligned}$$

The limit in (3.2.20) can be rewritten as

$$\lim_{h \rightarrow 0} \frac{p(x + \bar{h}) \cdot q - p(x) \cdot q}{h}. \quad (3.2.21)$$

Since  $\varphi$  is continuously differentiable w.r.t.  $x$  we can write

$$\begin{aligned} \varphi(x, p(x + \bar{h})) &= \varphi(x + \bar{h}, p(x + \bar{h})) - \partial_{x_1} \varphi(x, p(x + \bar{h}))h + o(h) \\ &= 1 - \partial_{x_1} \varphi(x, p(x + \bar{h}))h + o(h). \end{aligned}$$

and

$$\begin{aligned} \varphi(x + \bar{h}, p(x)) &= \varphi(x, p(x)) + \partial_{x_1} \varphi(x, p(x))h + o(h) \\ &= 1 + \partial_{x_1} \varphi(x, p(x))h + o(h), \end{aligned}$$

which implies that

$$\begin{aligned} \varphi\left(x, \frac{p(x + \bar{h})}{1 - \partial_{x_1} \varphi(x, p(x + \bar{h}))h + o(h)}\right) &= 1 \quad \text{and} \\ \varphi\left(x + \bar{h}, \frac{p(x)}{1 + \partial_{x_1} \varphi(x, p(x))h + o(h)}\right) &= 1. \end{aligned}$$

and therefore that

$$\begin{aligned} p(x) \cdot q &\geq \frac{p(x + \bar{h})}{1 - \partial_{x_1} \varphi(x, p(x + \bar{h}))h + o(h)} \cdot q, \quad \text{and} \\ p(x + \bar{h}) \cdot q &\geq \frac{p(x)}{1 + \partial_{x_1} \varphi(x, p(x))h + o(h)} \cdot q, \end{aligned}$$

Then the quotient in the limit (3.2.21) can be estimated from above and below, by:

$$\begin{aligned} \frac{p(x + \bar{h}) \cdot q - p(x) \cdot q}{h} &\leq \frac{\left(1 - \frac{1}{1 - \partial_{x_1} \varphi(x, p(x + \bar{h}))h + o(h)}\right) p(x + \bar{h}) \cdot q}{h} \\ &= \frac{1}{1 - \partial_{x_1} \varphi(x, p(x + \bar{h}))h + o(h)} \left(-\partial_{x_1} \varphi(x, p(x + \bar{h})) + \frac{o(h)}{h}\right) p(x + \bar{h}) \cdot q \end{aligned}$$

and

$$\begin{aligned} \frac{p(x + \bar{h}) \cdot q - p(x) \cdot q}{h} &\geq \frac{\left(\frac{1}{1 + \partial_{x_1} \varphi(x, p(x))h + o(h)} - 1\right) p(x) \cdot q}{h} \\ &= \frac{1}{1 + \partial_{x_1} \varphi(x, p(x))h + o(h)} \left(-\partial_{x_1} \varphi(x, p(x)) + \frac{o(h)}{h}\right) p(x) \cdot q. \end{aligned}$$

Letting  $h \rightarrow 0$  in both inequalities, by continuity of  $p(x)$  and  $\partial_{x_1} \varphi$  we get

$$\partial_{x_1} \varphi^0(x, q) = \lim_{h \rightarrow 0} \frac{p(x + \bar{h}) \cdot q - p(x) \cdot q}{h} = -\partial_{x_1} \varphi(x, p(x)) p(x) \cdot q,$$

which shows also that  $\partial_{x_1} \varphi^0$  is continuous.  $\square$

**Proposition 3.2.13.** *Let  $u : V \rightarrow \mathbb{R}^d$  be a continuous function. Then the following facts are equivalent:*

- (a)  $u : V \rightarrow \mathbb{R}$  is semiconcave (semiconvex);
- (b) there exists  $C \geq 0$  such that  $u(x) - \frac{C}{2}|x|^2$  is concave (convex) in  $V$ ;
- (c)  $u$  satisfies

$$(1 - \lambda)u(x) + \lambda u(y) - u((1 - \lambda)x + \lambda y) \leq (\geq) \lambda(1 - \lambda)C|y - x|^2, \quad (3.2.22)$$

for every  $x, y \in V$  such that the segment  $[x, y] \subset V$  and  $\lambda \in [0, 1]$ .

More details about the above result can be found in [37].

**Proposition 3.2.14.** *If  $d(y, \cdot) \in C_{\text{loc}}^{1,1}(\bar{\Omega})$ , uniformly w.r.t.  $y$ , then  $S^+$  is locally semiconcave. If  $d(\cdot, y) \in C_{\text{loc}}^{1,1}(\bar{\Omega})$  then  $S^-$  is locally semiconvex.*

*Proof.* We start proving that  $d(y, \cdot)$  is locally semiconcave. Let  $V \subset\subset \Omega$ . Then by Proposition 3.2.13  $d(y, \cdot)$  is semiconcave in  $V$  uniformly in  $y$  if there exists  $C \geq 0$  such that  $d_C(y, x) := d(y, x) - \frac{C}{2}|x|^2$  is concave for every  $x \in V$  and  $y \in \bar{\Omega}$ . By assumption there exists  $C > 0$  such that  $|\nabla d(y, x) - \nabla_x d(y, z)| \leq C|x - z|$ . This implies that

$$\begin{aligned} 0 &\geq (\nabla_x d(y, x) - \nabla_x d(y, z)) \cdot (x - z) - C|x - z|^2 \\ &= \left( \nabla_x \left( d(y, x) - \frac{C}{2}|x|^2 \right) - \nabla_x \left( d(y, z) - \frac{C}{2}|z|^2 \right) \right) \cdot (x - z) \\ &= (\nabla_x d_C(y, x) - \nabla_x d_C(y, z)) \cdot (x - z). \end{aligned}$$

The monotonicity relation and the differentiability of  $d_C(y, \cdot)$  imply that  $d_C(y, \cdot)$  is concave for every  $x \in V$  and  $y \in \bar{\Omega}$ . The semiconcavity of  $S^+$  is proved just observing that by continuity of  $g$ , lower semicontinuity of  $d$  (Proposition 2.1.24) and compactness of  $\partial\Omega$ , there exists  $y_1 \in \partial\Omega$  such that  $S^+(x) = g(y_1) + d(y_1, x)$ . The local semiconvexity of  $S^-$  can be proved analogously.  $\square$

**Definition 3.2.15** (see [37] Definition 3.1.1). For any  $x \in \Omega$ , the sets

$$D^-u(x) := \left\{ p \in \mathbb{R}^d : \liminf_{y \rightarrow x} \frac{u(y) - u(x) - p \cdot (y - x)}{|y - x|} \geq 0 \right\} \quad (3.2.23)$$

$$D^+u(x) := \left\{ p \in \mathbb{R}^d : \limsup_{y \rightarrow x} \frac{u(y) - u(x) - p \cdot (y - x)}{|y - x|} \leq 0 \right\} \quad (3.2.24)$$

are called respectively superdifferential and subdifferential of  $u$  at  $x$ .

**Proposition 3.2.16** (see [37] Proposition 3.3.1). *If  $u : \Omega \rightarrow \mathbb{R}^d$  is locally semiconcave, then  $p$  belongs to  $D^+u(x)$  if and only if for every  $V \subset\subset \Omega$  there exists  $C = C(V)$*

$$u(y) - u(x) - p \cdot (y - x) \leq C|x - y|^2, \quad (3.2.25)$$

for every  $x, y \in V$  such that the segment  $[x, y] \subset V$ .

*Proof.* If (3.2.25) holds, then clearly  $p$  belongs to  $D^+u(x)$ .

Viceversa, let us assume that  $p \in D^+u(x)$ . Dividing the inequality in (3.2.22) by  $(1 - \lambda)|x - y|$  we obtain

$$\begin{aligned} \frac{u(y) - u(x)}{|y - x|} &\leq \frac{u((1 - \lambda)y + \lambda x) - u(x)}{(1 - \lambda)|y - x|} + C\lambda|y - x| \\ &\leq \frac{u(x + (1 - \lambda)(x - y)) - u(x)}{(1 - \lambda)|y - x|} + C\lambda|y - x| \end{aligned}$$

The thesis follows passing to the lim sup for  $\lambda \rightarrow 1^-$  and using (3.2.24).  $\square$

**Theorem 3.2.17.** *Every solution  $u$  of (P) is differentiable at each point  $x$  of  $\mathcal{U}$ , where  $\mathcal{U}$  is the uniqueness set. Moreover  $DS^+ = Du = DS^-$  on  $\mathcal{U}$  and if  $(x_k) \subset \mathcal{U}$  is a sequence of points that converges to  $x \in \mathcal{U}$ , then  $Du(x_k) \rightarrow Du(x)$ . Finally, if  $\text{int}\mathcal{U} \neq \emptyset$  then  $u \in C^{1,1}(\omega)$ , for every  $\omega$  subdomain of  $\mathcal{U}$ .*

*Proof.* We provide here a sketch of the proof of the first part. For the last part we refer to Corollary 3.3.8 of [37].

Claim 1:  $D^+S^+(x) \neq \emptyset$  and  $D^-S^-(x) \neq \emptyset$  for every  $x \in \Omega$ .

*Proof 1.* First of all we recall that if a function is locally semiconcave in  $\Omega$ , then it is locally Lipschitz (see for example Proposition 2.1.7 in [37]). Thus by Rademacher's theorem it follows that  $S^+$  is differentiable for a.e.  $x$  in  $\Omega$ . This means that for any  $x \in \Omega$  there exists a sequence of points of differentiability  $(x_k)_k$  that converges to  $x$ . By definition of subdifferential we have that  $DS^+(x_k) \in D^+S^+(x_k)$  for every  $k$ . Moreover, thanks to Proposition 3.2.16, we have that

$$S^+(y) - S^+(x_k) - DS^+(x_k) \cdot (y - x_k) \leq C|x_k - y|^2,$$

for every  $y \in \Omega$  such that  $[x_k, y] \subset \Omega$ . Then up to the choice of a subsequence  $DS^+(x_k) \rightarrow p$ , for some  $p \in \mathbb{R}^d$ , so passing to the limit and using again Proposition 3.2.16 we obtain

### 3.2. THE UNIQUENESS SET

$p \in D^+S^+(x)$ . In order to prove that  $D^-S^-(x) \neq \emptyset$  it is enough to observe that  $-S^-$  is locally semiconcave in  $\Omega$  and that for any function  $u$ ,  $D^+(-u) = -D^-(u)$ .

Claim 2: If  $u$  is a solution of (P), then  $u$  is differentiable at  $x$  for every  $x \in \mathcal{U}$  and  $DS^+(x) = DS^-u(x) = Du(x)$ .

*Proof 2.* If  $u$  is a solution of (P), then by Theorem 3.2.8  $u(x) = S^+(x) = S^-(x)$  for every  $x \in \mathcal{U}$ . Thus for every  $V \subset\subset \Omega$  and  $y \in \Omega$  with  $[x, y] \subset V$  there exists  $C = C(V)$  such that

$$u(y) - u(x) \leq S^+(y) - S^+(x) \leq DS^+(x) \cdot (y - x) + C|x - y|^2,$$

which by Proposition 3.2.16 implies that  $DS^+(x) \in D^+u(x)$ . Analogously one can prove that  $DS^-(x) \in D^-u(x) \neq \emptyset$ . Let  $\theta$  be any unitary vector and  $p^+ \in D^+u(x)$ ,  $p^- \in D^-u(x)$ , then

$$p^- \cdot \theta \leq \liminf_{h \rightarrow 0^+} \frac{u(x + h\theta) - u(x)}{h} \leq \limsup_{h \rightarrow 0^+} \frac{u(x + h\theta) - u(x)}{h} \leq p^+ \cdot \theta, \quad (3.2.26)$$

that implies  $(p^- - p^+) \cdot \theta \leq 0$ . By the arbitrariness of the unitary vector  $\theta$  we have that  $p^+ = p^- = Du(x) = DS^+(x) = DS^-(x)$ .

Claim 3: If  $u$  is a solution of (P), then  $Du|_{\mathcal{U}}$  is continuous.

*Proof 3.* Let  $x \in \mathcal{U}$ . We consider a sequence  $(x_k) \subset \mathcal{U}$  such that  $x_k \rightarrow x$ , then since  $Du(x_k)$  belongs both to  $D^+u(x_k)$  and  $D^-u(x_k)$  we have by Proposition 3.2.16 that

$$\begin{aligned} u(y) - u(x_k) - Du(x_k) \cdot (y - x_k) &\leq C_1|x_k - y|^2, \\ u(y) - u(x_k) - Du(x_k) \cdot (y - x_k) &\geq C_2|x_k - y|^2, \end{aligned}$$

for every  $y \in V$  such that  $[x_k, y] \subset V$  and thus that  $(Du(x_k))_k$  is bounded and, arguing as before,  $(Du(x_k))_k$  must converge to  $Du(x)$ .  $\square$

#### 3.2.3 The attainment set as uniqueness set

As pointed out in Remark 3.1.11, if we consider a Borel function  $H : \bar{\Omega} \times \mathbb{R}^d \rightarrow \mathbb{R}$  satisfying all the assumptions (A)–(D) defined in Section 2.2.1 then if

$$\mu := \min \left\{ \operatorname{ess\,sup}_{x \in \Omega} H(x, Dv(x)) : v \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega}), u = g \text{ on } \partial\Omega \right\} < +\infty,$$

the problem (H), can be rewritten in the form of (P) with  $K(x) = \{H(x, \cdot) \leq \mu\}$ .

**Assumptions.** In this section we will assume that

$$(C^*) \quad \text{the map } (x, p) \mapsto H(x, p) \text{ is continuous in } \bar{\Omega} \times \mathbb{R}^d.$$

Thanks to assumption (C\*), we have that  $\varphi(x, p) := \inf \{t > 0 : \frac{p}{t} \in \{H(x, \cdot) \leq \mu\}\}$  satisfies assumption (a\*) and thus all the results of Section 3.2.1 hold. Moreover, by Theorem 2.2.4 we infer that  $\mu < +\infty$ .

Finally, we recall that by Theorem 2.2.27

$$\mu = \min\{\lambda : g(y) - g(x) \leq d_\lambda(x, y) \text{ for any } x, y \in \partial\Omega\}$$

and therefore  $g$  satisfies (1-Lip), i.e. it is 1 Lipschitz w.r.t.  $d$ .

**Proposition 3.2.18.** *Let  $H : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\mu$  as above. Given a boundary datum  $g \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$ , if there exists a point  $x \in \Omega$  where  $S_\mu^+(x) = S_\mu^-(x)$  then there exists a Lipschitz curve  $\gamma$  passing through  $x$  such that  $\gamma$  belongs to the uniqueness set.*

*Proof.* We recall that in this setting  $S_\mu^+(g, \Omega) = S^+$  and  $S_\mu^-(g, \Omega) = S^-$ . Then proof follows directly by Theorem 3.2.8.  $\square$

The following result is not directly useful for our purpose, but it may be interesting for a better understanding of the functional  $\varphi$  and its regularity depending on  $H$ .

**Proposition 3.2.19.** *If  $H(x, p)$  is positively 1-homogeneous w.r.t.  $p$  and  $\nabla_x H(x, p) \in C(\Omega \times (\mathbb{R}^d \setminus \{0\}))$ , then  $\nabla_x \varphi(x, p) \in C(\Omega \times (\mathbb{R}^d \setminus \{0\}))$ .*

*Proof.* We first consider  $p \in \mathbb{R}^d \setminus \{0\}$  such that  $\varphi(x, p) = 1$  (that means  $H(x, p) = \mu$ ). We have to prove that

$$\lim_{|h| \rightarrow 0} \frac{\varphi(x+h, p) - \varphi(x, p) - \nabla_x \varphi(x, p) \cdot h}{|h|} = 0 \quad (3.2.27)$$

and  $\nabla_x \varphi(x, p)$  is continuous. Since  $\nabla_x H$  exists and is continuous, it holds

$$H(x+h, p) = H(x, p) + \nabla_x H(x, p) \cdot h + o(|h|).$$

By 1-homogeneity, we have

$$\frac{\mu}{\mu + \nabla_x H(x, p) \cdot h + o(|h|)} H(x+h, p) = H\left(x+h, p \frac{\mu}{\mu + \nabla_x H(x, p) \cdot h + o(|h|)}\right) = \mu.$$

That implies

$$1 = \varphi\left(x+h, p \frac{\mu}{\mu + \nabla_x H(x, p) \cdot h + o(|h|)}\right) = \varphi(x+h, p) \frac{\mu}{\mu + \nabla_x H(x, p) \cdot h + o(|h|)}.$$

From which we get

$$\varphi(x+h, p) = 1 + \frac{\nabla_x H(x, p) \cdot h + o(|h|)}{\mu} = \varphi(x, p) + \frac{\nabla_x H(x, p) \cdot h + o(|h|)}{\mu}.$$

The (3.2.27) and the regularity of  $\nabla_x \varphi$  are proved just observing that  $\nabla_x \varphi(x, p) = \mu^{-1} \nabla_x H(x, p)$ . If  $\varphi(x, p) \neq 1$ , we can use the same reasoning with  $\frac{p}{\varphi(x, p)}$  and, exploiting the 1-homogeneity of  $\varphi$  w.r.t. to  $p$ , we get that  $\nabla_x \varphi(x, p) = \mu^{-1} \varphi(x, p) \nabla_x H(x, p)$ .  $\square$

**Corollary 3.2.20.** *If  $H(x, p)$  is positively 1-homogeneous w.r.t.  $p$ ,  $\nabla_x H : \Omega \times (\mathbb{R}^d \setminus \{0\}) \rightarrow \mathbb{R}$  is continuous and the sublevel sets  $\{H(x, p) \leq \mu\}$  are strictly convex for every  $x \in \Omega$ , then  $\varphi^0(\cdot, q) \in C^1(\Omega)$  for any  $q \in \mathbb{R}^d \setminus \{0\}$ .*

*Proof.* The proof follows directly by Proposition 3.2.19 and Proposition 3.2.12.  $\square$

The aim now is to show for this case further properties of the uniqueness set  $U$  defined in Section 3.2.

**Proposition 3.2.21.** *Let  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$  be an optimal solution of (H). If  $x$  belongs to the uniqueness set  $\mathcal{U}$  and  $u$  is differentiable at  $x$  then*

$$H(x, Du(x)) = \operatorname{ess\,sup}_{x \in \Omega} H(x, Du(x)).$$

### 3.2. THE UNIQUENESS SET

---

*Proof.* By Theorem 3.2.17  $u$  is differentiable at  $x$ . Since  $u$  is a solution of (H), we know that  $Du(x) \in \{H(x, p) \leq \mu\}$ . Let  $\gamma$  be a curve of the uniqueness set passing through  $x$ . Let  $s > 0$  such that  $x = \gamma(s)$ . By Theorem 3.2.8 we know that

$$Du(x) \cdot \dot{\gamma}(s) = \lim_{t \rightarrow 0} \frac{u(x + t\dot{\gamma}(s)) - u(x)}{t} = \varphi^0(x_0, \dot{\gamma}(s)),$$

as stated in the last part of Theorem 3.2.8. Then by Lemma 3.1.9 we have that  $Du(x)$  belongs to the boundary of  $K(x) = \{H(x, p) \leq \mu\}$ , that means  $H(x, Du(x)) = \mu$ .  $\square$

As we saw in Section 2.2.5, without further regularity assumptions, it is also possible to give a point-wise definition (see Definition 2.2.36) of the functional  $x \mapsto H(x, Du)(x)$  for every  $x \in \Omega$ , in such a way that  $H(x, Du(x)) = H(x, Du)(x)$  if  $u$  is differentiable at  $x$ . Therefore we are able to prove the following improvement of the Proposition 3.2.21:

**Theorem 3.2.22.** *Let  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$  be an optimal solution of the (H). If  $x$  belongs to the uniqueness set  $\mathcal{U}$  then*

$$H(x, Du)(x) = \operatorname{ess\,sup}_{x \in \Omega} H(x, Du(x)) = \mu.$$

*In particular*

$$\mathcal{U} \subset \mathcal{A}(u),$$

*where  $\mathcal{A}(u)$  is the attainment set of  $u$  defined in Definition 2.2.46.*

*Proof.* We have to prove that  $\lim_{r \rightarrow 0} \mu(x, r) = \mu$ . For every  $r > 0$ ,  $\mu(x, r) \leq \mu$ , by Theorem 3.1.16. Moreover, if we consider a curve  $\gamma \subset \mathcal{U}$  passing through  $x$  given by Theorem 3.2.8, we have that  $u(\gamma(s)) - u(x) = d(x, \gamma(s))$ , proving that  $\mu(x, r) = \mu$  for every  $r$ .  $\square$

In the next theorem we show that if we consider an absolute minimizer  $u$ , the converse inclusion also holds, implying  $\mathcal{A}(u) = \mathcal{U}$ .

**Theorem 3.2.23.** *Let us assume that  $H$  satisfies also assumption (E). Let  $u \in W^{1,\infty}(\Omega) \cap C(\bar{\Omega})$  be an absolute minimizer for (H) and let  $x \in \Omega$  be such that  $H(x, Du)(x) = \operatorname{ess\,sup}_{x \in \Omega} H(x, Du(x))$ . Then  $x$  belongs to the uniqueness set, that is*

$$\mathcal{A}(u) \subset \mathcal{U}.$$

*Proof.* Let  $x$  be as in the statement. By Theorem 2.2.44 we know that there exist  $y_1, y_2 \in \partial\Omega$  such that  $u(x) = g(y_1) - d(x, y_1) = g(y_2) + d(y_2, x)$ . Then  $S^+(x) \leq u(x) \leq S^-$ , by definition of  $S^+$  and  $S^-$ . Moreover, by Proposition 3.1.18,  $S^+ \geq u(x) \geq S^-$ .  $\square$

**Corollary 3.2.24.** *Let us assume that  $H$  satisfies also assumption (E). Under the same assumptions of Proposition 3.2.11 on  $\varphi^0$  or of Proposition 3.2.14 on  $d$ , any absolute minimizer  $u$  for the problem (H) is differentiable on  $\mathcal{A}(u)$  and  $Du|_{\mathcal{A}(u)}$  is continuous. Moreover if  $\operatorname{int} \mathcal{A}(u) \neq \emptyset$ , then  $u \in C_{loc}^{1,1}(\mathcal{A}(u))$ .*

*Proof.* The proof follows by Theorem 3.2.23 and Theorem 3.2.17.  $\square$



# Chapter 4

## $L^\infty$ - Optimal Transport

### 4.1 Preliminaries

As pointed out in the Introduction, the so-called optimal transport problem has been widely studied in the last decades and its theory has witnessed a huge development in many directions. The scope of this section is to provide a brief survey of the results which are useful for our purpose. More detailed and comprehensive references are, for instance, [3, 108, 120].

#### 4.1.1 A brief survey of Optimal Transport

Let  $(X, d_X)$  and  $(Y, d_Y)$  be Polish spaces, i.e. complete and separable metric spaces, and  $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ , where  $\mathcal{P}(Z)$  denotes the space of Borel probability measures on a Polish space  $Z$ .

**Definition 4.1.1** (Weak - or narrow - topology). If  $(Z, \tau)$  is a topological space, we say that a sequence of probability measure  $(\rho_n)_n$  weakly converges or converges w.r.t. to the weak (or narrow) topology of measure to a probability measure  $\rho$  if

$$\lim_{n \rightarrow +\infty} \int_Z f d\rho_n = \int_Z f d\rho,$$

for every  $f \in C_b(Z)$ .

**Definition 4.1.2** (Transport plans). We say that  $\gamma$  is a transport plan if it belongs to the set

$$\Pi(\mu, \nu) := \{ \gamma \in \mathcal{P}(X \times Y) : \pi_{\#}^1 \gamma = \mu \text{ and } \pi_{\#}^2 \gamma = \nu \},$$

where  $\pi^1$  and  $\pi^2$  represent respectively the projections on the spaces  $X$  and  $Y$ . In other words,  $\gamma \in \Pi(\mu, \nu)$  if and only if

$$\gamma(A \times Y) = \mu(A) \quad \text{and} \quad \gamma(X \times B) = \nu(B),$$

for any Borel sets  $A \subset X, B \subset Y$ .

**Definition 4.1.3** (Transport maps). We say that  $T$  is a transport map if it belongs to the set

$$\mathcal{T}(\mu, \nu) := \{ T : X \rightarrow Y \text{ Borel} : T_{\#} \mu = \nu \},$$

where  $T_{\#}\mu = \nu$  means that

$$\mu(T^{-1}(B)) = \nu(B),$$

for any Borel set  $B \subset Y$ , and  $T_{\#}\mu$  is called the *push-forward* of  $\mu$  under  $T$ .

*Remark 4.1.4.* We observe that

1.  $\mathcal{T}(\mu, \nu)$  can be empty: consider for instance

$$\mu = \delta_x \text{ and } \nu = \frac{1}{2}(\delta_{y_1} + \delta_{y_2}),$$

for some  $x \in X$ ,  $y_1 \neq y_2 \in Y$ ;

2. " $\mathcal{T}(\mu, \nu) \subset \Pi(\mu, \nu)$ ": for every  $T \in \mathcal{T}(\mu, \nu)$ ,

$$\gamma := (Id \times T)_{\#}\mu \in \Pi(\mu, \nu).$$

**The Optimal Transport Problem: Monge's formulation [99]**

Let  $c : X \times Y \rightarrow [0, +\infty]$  be the so-called *cost function* (note that in the original work  $c(x, y) = d(x, y)$  and  $X = Y$ ), the Monge Optimal Transport Problem is

$$\inf_{T \in \mathcal{T}(\mu, \nu)} \int_X c(x, T(x)) d\mu(x). \tag{M-OT}$$

*Remark 4.1.5.* The problem (M-OT) does not always admits a solution. First of, all because  $\mathcal{T}(\mu, \nu)$  can be empty (Remark 4.1.4 (1)), but also because it can happen, as showed in Example 1.14 in [3], that transport maps exist but the infimum is not a minimum. In general, it is not possible to find a topology such that  $T \rightarrow \int c(x, T(x))$  is lower semicontinuous and  $\mathcal{T}(\mu, \nu)$  is closed and compact.

**The Optimal Transport Problem: Kantorovich's formulation [82]**

Let the Borel function  $c : X \times Y \rightarrow [0, +\infty]$  be the so-called *cost function*, then the Kantorovich Optimal Transport Problem is

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y). \tag{OT}$$

*Remark 4.1.6.* Some considerations:

1. for what we have observed in (2) of Remark 4.1.4,

$$\min_{\Pi(\mu, \nu)} (\text{OT}) \leq \inf_{\Pi(\mu, \nu)} (\text{M-OT})$$

and, moreover, the Kantorovich problem can be seen as the relaxation (for a detailed explanation see for instance Chapter 1.5 in [108]) of the original problem by Monge in which we allow for *mass splitting*;

2. since  $\Pi(\mu, \nu)$  is convex and the functional  $\gamma \rightarrow \int c d\gamma$  is linear, the problem (OT) is a linear optimization problem under linear constraints.

In this case we have the following result.

**Theorem 4.1.7.** *Let  $c : X \times Y \rightarrow [0, +\infty]$  lower semicontinuous. Then the infimum in (OT) is attained.*

*Proof.* The proof makes use of the classic direct method in the calculus of variations. Indeed it is possible to prove that

- the functional  $\gamma \rightarrow \int c d\gamma$  is lower semicontinuous with respect to the weak topology of measures if  $c$  is lower semicontinuous,
- $\Pi(\mu, \nu)$  is compact w.r.t. the weak topology of measures, thanks to Prokhorov's theorem.

□

**Definition 4.1.8.** A set  $\Gamma \subset X \times Y$  is  $c$ -cyclically monotone ( $c$ -cm) if for every finite set of points  $\{(x_i, y_i)\}_{i=1}^k \subset \Gamma$  and for every permutation  $\sigma$  of the set  $\{1, \dots, k\}$  we have

$$\sum_{1 \leq i \leq k} c(x_i, y_i) \leq \sum_{1 \leq i \leq k} c(x_i, y_{\sigma(i)}), \quad (c\text{-cm})$$

or, equivalently,

$$\sum_{1 \leq i \leq k} c(x_i, y_i) \leq \sum_{1 \leq i \leq k} c(x_i, y_{i+1}).$$

with  $y_{k+1} := y_1$ . The equivalence is due to the fact that every permutation  $\sigma$  can be obtained as composition of cycles acting on disjoint subsets of  $\{1, \dots, k\}$  and trivial cycles on fixed points.

We say that a transport plan  $\gamma \in \Pi(\mu, \nu)$  is  $c$ -cyclically monotone ( $c$ -cm), if it is concentrated on  $c$ -cyclically monotone set.

*Remark 4.1.9.* If  $X = Y = \mathbb{R}^d$  and  $c(x, y) = |x - y|^2$ , the inequality ( $c$ -cm) is equivalent to:

$$\sum_{1 \leq i \leq k} \langle x_i, y_i \rangle \geq \sum_{1 \leq i \leq k} \langle x_i, y_{\sigma(i)} \rangle.$$

If the cost function is continuous the inequality in (1) is actually an equality, as stated in the following theorem.

**Theorem 4.1.10** (see A. Pratelli in [106]). *Assume that the cost function  $c : X \times Y \rightarrow \mathbb{R}_+$  is continuous. Then a transport plan  $\gamma$  is optimal if and only if  $\gamma$  is  $c$ -cyclically monotone.*

**Theorem 4.1.11** (Duality, see for instance [3]). *Let  $c : X \times Y \rightarrow \overline{\mathbb{R}}$  lower semicontinuous and bounded from below then*

$$\min(\text{OT}) = \sup_{(\varphi, \psi) \in I_c} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \right\}, \quad (\text{DualOT})$$

where

$$I_c = \{(\varphi, \psi) \in C_b(X) \times C_b(Y) \text{ and } \varphi(x) + \psi(y) \leq c(x, y)\}.$$

If  $c(x, y) \leq a(x) + b(y)$  for some  $a \in L^1(\mu)$ ,  $b \in L^1(\nu)$ , then there exists an optimal pair  $(\varphi, \psi) \in L^1(X, \mu) \times L^1(Y, \nu)$  and in particular

$$\min_{\Pi(\mu, \nu)}(\text{OT}) = \max_{I_c}(\text{DualOT}).$$

$\varphi$  and  $\psi$  are called Kantorovich potentials.

**Definition 4.1.12** (2-Wasserstein distance). If  $X = Y$  and  $c(x, y) = d^2(x, y)$ , where  $d$  is the distance between  $x$  and  $y$ , the minimum value

$$\min_{\Pi(\mu, \nu)} \left( \int d^2(x, y) d\gamma \right)^{\frac{1}{2}},$$

is known as 2-Wasserstein distance between  $\mu$  and  $\nu$  and denoted by  $W_2(\mu, \nu)$ .

We have that  $W_2$  defines a distance on the set of probability measures with finite 2nd moment, i.e.

$$\mathcal{P}_2(X) := \left\{ \mu \in \mathcal{P}(X) : \int_X d^2(x_0, x) d\mu < +\infty, \text{ for some } x_0 \in X \right\}.$$

In particular the topology induced by the distance of the space  $(\mathcal{P}_2(X), W_2)$  inherits completeness, compactness and separability from  $X$  and, if  $X$  is compact,  $W_2(\mu_n, \mu) \rightarrow 0$  for  $n \rightarrow \infty$  if and only if  $\mu_n \rightarrow \mu$ . Moreover all these results hold also for  $(\mathcal{P}_p(X), W_p)$ , with any  $1 < p < +\infty$ . For more details we refer to [3, 108].

The following are results on the sufficient conditions for an optimal plan  $\gamma$  to be concentrated on a graph, i.e. to be equal to  $(Id \times T)_\# \mu$ , for some  $T \in (\mu, \nu)$ . The consequence is the existence and uniqueness of solutions of (M-OT)=(OT). The first and most famous one is Brenier's theorem:

**Theorem 4.1.13** (Brenier, 1987 [27]). *Let  $X = Y = \mathbb{R}^d$ ,  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  and  $c(x, y) = \frac{|x-y|^2}{2}$ . Suppose that  $\int |x|^2 d\mu, \int |y|^2 d\nu < +\infty$ , which implies  $\min(\text{OT}) < +\infty$ , and suppose that  $\mu$  gives no mass to  $(d-1)$ -surfaces of class  $C^2$ . Then, there exists a unique optimal transport plan  $\gamma_K$ . Moreover,  $\gamma_K$  is induced by a transport map  $T$  from  $\mu$  to  $\nu$ , which is the unique solution of (M-OT), and it is of the form  $T = \nabla u$  for a convex function  $u$ . In particular,  $u(x) = \frac{1}{2}|x|^2 - \varphi(x)$ , where  $\varphi$  is a Kantorovich potential.*

**Theorem 4.1.14** (Theorem 1.17 [108], first proved by Gangbo and McCann in [69]). *Let  $\mu, \nu \in \mathcal{P}(\Omega)$  with  $\Omega$  a compact domain of  $\mathbb{R}^d$ ,  $\mu \ll \mathcal{L}^d$  and  $\partial\Omega$  negligible. Let the cost function be  $c(x, y) = h(x-y)$  with  $h$  strictly convex. Then there exists a unique optimal transport plan  $\gamma_K$  for the cost, which is unique and of the form  $(Id \times T)_\# \mu$ , for some transport map  $T$  from  $\mu$  to  $\nu$ . Moreover, there exists a Kantorovich potential  $\varphi$ , such that  $T$  and the potential  $\varphi$  are linked by*

$$T(x) = x - (\nabla h)^{-1}(\nabla \varphi(x)),$$

for  $\mu$ -a.e.  $x$ .

*Remark 4.1.15.* Let us point out the key ingredients of the proof of Theorem 4.1.13 and Theorem 4.1.14. The first one is the differentiability  $\mu$ -almost everywhere of  $\varphi$ , while the second one is the existence of a unique  $y$  solving the equation

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

**Definition 4.1.16** (Twist condition). For  $c : \Omega \times \Omega \rightarrow \mathbb{R}$  with  $\Omega \subset \mathbb{R}^d$  we say that  $c$  satisfies the *twist condition* whenever  $c$  is differentiable w.r.t.  $x$  at every point, and the map  $y \mapsto \nabla_x c(x_0, y)$  is injective for every  $x_0$ . For "nice" domains and cost functions, it corresponds to

$$\det \nabla_{x,y}^2 c(x, y) \neq 0.$$

**Proposition 4.1.17.** *If  $c$  is  $C^1$ ,  $\varphi$  is a Kantorovich potential for the cost  $c$  in the transport from  $\mu$  to  $\nu$ , and  $(x_0, y_0)$  belongs to the support of an optimal transport plan  $\gamma$ , then*

$$\nabla\varphi(x_0) = \nabla_x c(x_0, y_0),$$

*provided  $\varphi$  is differentiable at  $x_0$ . In particular, the gradients of two different Kantorovich potentials coincide on every point  $x_0$  of  $\text{supp } \mu$  where both the potentials are differentiable. Moreover if  $c$  satisfies the twist condition*

$$y_0 = (\nabla_x c(x_0, y_0))^{-1} \nabla\varphi(x_0).$$

Theorem 4.1.13, Theorem 4.1.14 and Prop 4.1.17 can be generalized to Riemannian manifold (see for instance [95, 21, 67] and also [120] and the references therein) and to Sub-riemannian manifolds (see for instance [5]). Moreover some Monge type results have been found also in cases where the twist condition does not hold, for example in case where the cost function is equal to the Euclidean norm (see [115, 117] and also [3] and the references therein) or equal to any norm in  $\mathbb{R}^d$  (see [45]).

#### 4.1.2 A brief survey of $L^\infty$ -Optimal Transport

Let still consider  $(X, d_X)$  and  $(Y, d_Y)$  to be Polish spaces,  $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$  and  $c : X \times Y \rightarrow [0, +\infty]$  a Borel function.

The  $L^\infty$ -Optimal Transport problem is the problem

$$\min_{\gamma \in \Pi(\mu, \nu)} C_\infty(\gamma) := \min_{\gamma \in \Pi(\mu, \nu)} \left( \gamma - \text{ess sup}_{(x, y) \in X \times Y} c(x, y) \right) = \min_{\gamma \in \Pi(\mu, \nu)} \|c\|_{L^\infty(\gamma, X \times Y)}, \quad (\text{OT}_\infty)$$

where the  $\gamma$ -essential supremum of the function  $c$  is defined in Definition 2.2.1. See also Remark 2.2 for the case  $c$  is lower semicontinuous.

Since we are considering probability measures  $\gamma$ , we have that  $C_p(\gamma) \nearrow C_\infty(\gamma)$  as  $p \rightarrow \infty$ , where

$$\min_{\gamma \in \Pi(\mu, \nu)} C_p(\gamma) := \min_{\gamma \in \Pi(\mu, \nu)} \left( \int_{X \times Y} c^p(x, y) d\gamma(x, y) \right)^{\frac{1}{p}} = \min_{\gamma \in \Pi(\mu, \nu)} \|c\|_{L^p(\gamma, X \times Y)}. \quad (\text{OT}_p)$$

*Remark 4.1.18.* The problem  $(\text{OT}_p)$  is equivalent to the Monge-Kantorovich problem (OT), where the cost function is  $c^p$ , since the power  $\frac{1}{p}$  does not affect the set of minimizers. Moreover, in addition to the known property of  $L^p$  norms in case of finite measure spaces,

$$\|c\|_{L^p(\gamma)} \rightarrow \|c\|_{L^\infty(\gamma)},$$

it is also true (see Lemma 2.13 in [79]) that

$$\min_{\gamma \in \Pi(\mu, \nu)} C_p(\gamma) \rightarrow \min_{\gamma \in \Pi(\mu, \nu)} C_\infty(\gamma)$$

**Definition 4.1.19** ( $\infty$ -Wasserstein distance). If  $X = Y$  the cost function and  $c(x, y) = d(x, y)$  is the distance between  $x$  and  $y$ , the  $(\text{OT}_\infty)$  becomes,

$$\min_{\Pi(\mu, \nu)} \|d\|_{L^\infty(\gamma)},$$

which is the so-called  $\infty$ -Wasserstein distance between  $\mu$  and  $\nu$ , denoted by  $W_\infty(\mu, \nu)$ .

In order to have  $W_\infty < +\infty$ , we usually consider  $\mu$  and  $\nu$  with compact support.

Exactly as for the 2-Wasserstein distance defined in Definition 4.1.12 and for the more general  $p$ -Wasserstein distance, denoted by  $W_p$ ,

$$\min_{\Pi(\mu,\nu)} \left( \int d^p(x,y) d\gamma \right)^{\frac{1}{p}},$$

we have a new distance on probability measures on compact sets, which measures the minimal maximal displacement that should be done to move particles from one distribution to the other. However the topology induced is not the same: even though  $X$  is compact,  $(\mathcal{P}(X), W_\infty)$  is neither compact, nor separable, as soon as the cardinality of  $X$  is larger than 1. Consider for instance

$$W_\infty((1-t)\delta_x + t\delta_y, \delta_x) = |x - y|,$$

which tells us that  $(1-t)\delta_x + t\delta_y$  doesn't tend to  $\delta_x$  for  $t \rightarrow 0$ .

*Remark 4.1.20.* Exactly as the supremal functional  $\mathcal{F}_H$  studied in Chapter 2 (see (H)), the functional  $C_\infty : \mathcal{P}(X \times Y) \rightarrow \overline{\mathbb{R}}$  is quasiconvex, i.e. for any  $\gamma_1, \gamma_2 \in \mathcal{P}(X \times Y)$

$$C_\infty(t\gamma_1 + (1-t)\gamma_2) \leq \max\{C_\infty(\gamma_1), C_\infty(\gamma_2)\}$$

In particular, by definition of essential supremum one can see that

$$C_\infty(t\gamma_1 + (1-t)\gamma_2) = \max\{C_\infty(\gamma_1), C_\infty(\gamma_2)\} \tag{4.1.1}$$

The first important result is the existence of optimal plans for the problem  $(OT_\infty)$ , which was first established in [47] and then generalized in [79].

**Theorem 4.1.21.** *[Theorem 2.6 in [79]] Let  $c$  be lower semicontinuous. Then the problem  $(OT_\infty)$  has at least one minimizer, i.e. there exists  $\gamma_\infty$  such that*

$$\gamma_\infty \in \arg \min_{\Pi(\mu,\nu)} \|c\|_{L^\infty(\gamma, X \times Y)}.$$

*Proof.* The proof is similar to the one of Theorem 4.1.7. The keys ideas are:

- $\gamma \mapsto C_\infty(\gamma)$  is lower semicontinuous with respect to the weak topology of measures. Indeed if  $\gamma_n \rightharpoonup \gamma$  for any  $z \in \text{supp } \gamma$  there exists  $(z_n)_n$ , such that  $z_n \in \text{supp } \gamma_n$  and  $z_n \rightarrow z$  and we conclude with the lower semicontinuity of  $c$ ;
- $\Pi(\mu, \nu)$  is compact w.r.t. the weak topology of measures, thanks to Prokhorov's theorem.

□

In general, there can be a high level of non-uniqueness of the solutions of  $(OT_\infty)$ : we are only optimizing the "worst case" and more locally the situation can be far from optimal. We provide below some easy examples to better illustrate this fact.

*Example 4.1.22.* Let us take  $\mu := \frac{1}{3}(\delta_{x_1} + \delta_{x_2} + \delta_{x_3})$ ,  $\nu := \frac{1}{3}(\delta_{y_1} + \delta_{y_2} + \delta_{y_3})$  and  $\gamma_1 := \frac{1}{3}(\delta_{x_1, y_3} + \delta_{x_2, y_1} + \delta_{x_3, y_2})$ ,  $\gamma_2 := \frac{1}{3}(\delta_{x_1, y_3} + \delta_{x_2, y_2} + \delta_{x_3, y_1})$ , as illustrated in Figure 4.1.1. We can easily see that both the plans are optimal since the furthest points have been optimized by sending  $x_1$  to  $y_3$ .

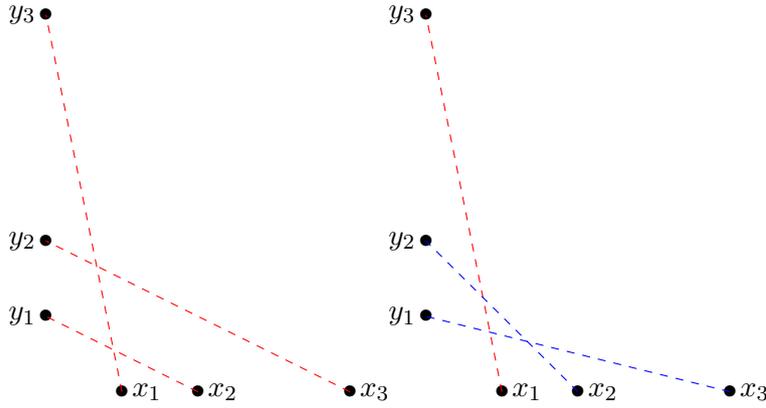


Figure 4.1.1: On the left  $\gamma_1$ , on the right  $\gamma_2$  obtained sending  $x_2$  to  $y_2$  and  $x_3$  to  $y_1$ .

*Example 4.1.23.* This is an example in which every transport plan is optimal. This example was illustrated in [47]. Let us consider  $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$  defined by  $\mu := \mathcal{L}^2_{[0,1]^2}$  and  $\nu := \frac{1}{2}(\delta_{(2,1)} + \delta_{(1,2)})$  and let  $c(x, y) = |x - y|$ . Then

$$C_\infty(\gamma) = \sup\{|x - y| : (x, y) \in \text{supp } \gamma\} = \sqrt{5},$$

for any  $\gamma \in \Pi(\mu, \nu)$ . Indeed,

$$C_\infty(\gamma) = \sup\{c(x, y) : (x, y) \in \text{supp } \gamma\} = |(0, 0) - (1, 2)| = |(0, 0) - (2, 1)| = \sqrt{5}$$

for every  $\gamma \in \Pi(\mu, \nu)$ .

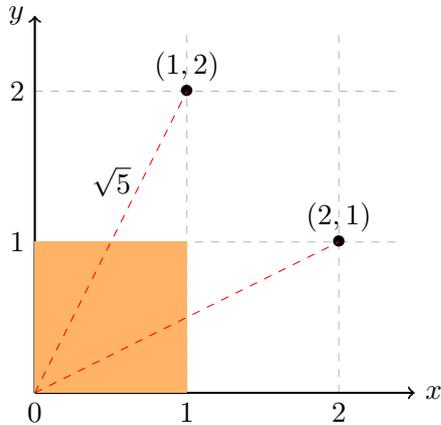


Figure 4.1.2: In the figure  $\text{supp } \mu$  is the orange square and the two points  $\text{supp } \nu : \{(1, 2), (2, 1)\}$ .

It is often useful to consider a subset of better-behaving minimizers, the so-called  $\infty$ -*cyclically monotone* transport plans. The idea is that on the support of such a plan, there is no rearrangement of destinations of a fixed finite set of points that would improve the highest cost coupling of that set. To our purpose it is also useful to have a definition for a plan that does not improve under the interchange of any two destinations of a given pair of initial points. The exact definition of these propriety appears first in [47] and it can be seen as a generalization to the  $L^\infty$  setting of the Definition 4.1.8.

**Definition 4.1.24.** A set  $\Gamma \subset X \times Y$  is  $\infty$ - $c$ -cyclically monotone ( $\infty$ -cm) if for every finite set of points  $\{(x_i, y_i)\}_{i=1}^k \subset \Gamma$  and for every permutation  $\sigma$  of the set  $\{1, \dots, k\}$  we have

$$\max_{1 \leq i \leq k} c(x_i, y_i) \leq \max_{1 \leq i \leq k} c(x_i, y_{\sigma(i)}), \quad (\infty\text{-cm})$$

or, equivalently,

$$\max_{1 \leq i \leq k} c(x_i, y_i) \leq \max_{1 \leq i \leq k} c(x_i, y_{i+1}),$$

with  $y_{k+1} := y_1$ . As in Definition 4.1.8, the equivalence is due to the fact that every permutation  $\sigma$  can be obtained as composition of cycles acting on disjoint subsets of  $\{1, \dots, k\}$  and trivial cycles on fixed points.

We say that a transport plan  $\gamma \in \Pi(\mu, \nu)$  is  $\infty$ - $c$ -cyclically monotone ( $\infty$ -cm), if it is concentrated on  $\infty$ - $c$ -cyclically monotone set.

**Definition 4.1.25.** The set  $\Gamma \subset X \times Y$  is  $\infty$ - $c$ -monotone ( $\infty$ -m) if for any  $(x, y), (x', y') \in \Gamma$  we have

$$\max\{c(x, y), c(x', y')\} \leq \max\{c(x, y'), c(x', y)\}. \quad (\infty\text{-m})$$

We say that a transport plan  $\gamma \in \Pi(\mu, \nu)$  is  $\infty$ - $c$ -monotone ( $\infty$ -m), if it is concentrated on an  $\infty$ - $c$ -monotone set.

**Notation:** When the cost function  $c$  is the Euclidean distance (see original definition in [47]) or  $c$  is clear from the context, we will omit it and we will simply write that a plan  $\gamma$  is  $\infty$ -cyclically monotone or  $\infty$ -monotone. Concerning this point it is also interesting to consider the following remark.

*Remark 4.1.26.* If a set  $\Gamma \subset X \times Y$  is  $\infty$ - $c$ -cyclically monotone ( $\infty$ -monotone) for some cost function  $c$ , then  $\Gamma$  is also  $\infty$ - $f(c)$ -cyclically monotone ( $\infty$ - $f(c)$ -monotone), where  $f$  is any nondecreasing real function.

**Lemma 4.1.27** (Lemma 2.11 in [79]). *If the cost function  $c$  is continuous then a transport plan  $\gamma$  is  $\infty$ - $c$ -cyclically monotone ( $\infty$ - $c$ -monotone) if and only if  $\text{supp } \gamma$  is  $\infty$ - $c$ -cyclically monotone ( $\infty$ - $c$ -monotone).*

**Theorem 4.1.28** (Theorem 2.17 in [79]). *If the transport cost  $c$  is continuous, then any  $\infty$ - $c$ -cyclically monotone plan is optimal.*

*Remark 4.1.29.* Every  $\infty$ -cyclically monotone plan is  $\infty$ -monotone but the reverse is not true in general. Consider, for example, the marginal measures  $\mu = \nu$  to be the uniform probability measure on the circle  $\mathbb{S}^1$ , the cost function  $c(x, y) = |x - y|$  (the Euclidean distance), and the set  $\Gamma = \text{Graph}(R_\theta)$ , where  $R_\theta$  is the rotation of a vector  $v \in \mathbb{S}^1$  by an angle  $\theta < \pi/2$ . Now  $\Gamma$  is  $\infty$ -monotone, but since the transport plan supported by  $\Gamma$  is not optimal (the optimal cost being 0, given by the identity map) the set  $\Gamma$  cannot be  $\infty$ -cyclically.

The existence is ensured by the following lemma, proved first by [47] and then generalized by H. Jylhä in [79].

**Lemma 4.1.30** (Lemma 2.12 in [79]). *Let  $c$  be lower semicontinuous. Let  $\gamma_p$  be a solution of the problem  $(P_p)$ ,  $p \geq 1$ . Then cluster points of  $(\gamma_p)_{p \geq 1}$  w.r.t. to the weak convergence exist and each of them is a solution of the problem  $(P_\infty)$ . If, in addition,  $c$  is continuous and the minimum of each problem  $(P_p)$  is finite, then the cluster points are  $\infty$ - $c$ -cyclically monotone.*

**Theorem 4.1.31.** (Theorem 2.16 in [79]) *If  $c$  is continuous, then the problem  $(\text{OT}_\infty)$  has a solution  $\gamma$  that is  $\infty$ -cyclically monotone.*

Theorem 4.1.31 is a consequence of Lemma 4.1.30 and it is proved by approximating  $c$  with bounded and continuous functions  $c_m$ .

As expected,  $\infty$ -cyclically monotone transport plans are much better-behaved than arbitrary solutions to the problem  $(\text{OT}_\infty)$ . For example, they are restrictable in the sense of the following definition.

**Definition 4.1.32** (Restrictable minimizers). A transport plan  $\gamma \in \Pi(\mu, \nu)$  is said to be a restrictable minimizer of the problem  $(\text{P}_\infty)$  if it satisfies the following condition: for any  $\gamma' \leq \gamma$ , i.e.  $\gamma'(B) \leq \gamma(B)$  for every Borel sets  $B$ , and  $\gamma'(X \times Y) > 0$  we have

$$C_\infty(\bar{\gamma}) = \min \{ C_\infty(\lambda) : \lambda \in \Pi(\pi_{\sharp}^1 \bar{\gamma}, \pi_{\sharp}^2 \bar{\gamma}) \},$$

where  $\bar{\gamma} = \frac{\gamma'}{\gamma'(X \times Y)}$ .

*Remark 4.1.33.* We point out how the definition of restrictable plans reminds the definition given in Chapter (2) of absolute minimizers (see Definition 2.2.29).

In the Example 4.1.22 one can easily check that the plan  $\gamma_2$  obtained by switching the role of  $y_1$  and  $y_2$  is  $\infty$ -cyclically monotone. It is interesting to notice that it is the one which optimizes how to connect also the other points and not only the worst case and it is therefore restrictable in the sense of Definition 4.1.32. Differently from  $\gamma_1$  which is neither  $\infty$ -monotone nor restrictable. This elementary example could give an intuition of the theorem stated below.

**Theorem 4.1.34** (Theorem 2.19 in [79]). *Let the transport cost  $c$  be continuous, then a transport plan  $\gamma$  is a restrictable minimizer of  $(\text{OT}_\infty)$  if and only if it is  $\infty$ -c-cyclically monotone.*

As explained in Section 4.1.1, a crucial question in optimal transport theory is whether an optimal solution of  $(\text{OT})$  is deterministic, i.e. it is also a solution of  $(\text{M-OT})$ . In the  $L^\infty$ -setting this problem is even more complicated since the lack of convexity implies a lack of a duality tool, which, as pointed out in Section 1.2, has a key role in proving Brenier's theorem (Theorem 4.1.13), Gangbo and McCann's theorem (Theorem 4.1.14) and Proposition 4.1.17. The only result regarding duality for  $L^\infty$  is the following, which has been proved in [15].

**Theorem 4.1.35.** *Let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  and  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$  lower semicontinuous. Then if  $C_\infty(\gamma) < +\infty$  for some  $\gamma$ , the problem  $(\text{OT}_\infty)$  is equivalent to*

$$\inf_{\lambda \geq 0} \sup \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu + \int_{\mathbb{R}^d} \psi(y) d\nu : (\varphi, \psi) \in I_c^\lambda \right\}, \quad (4.1.2)$$

where for any  $\lambda \geq 0$

$$I_c^\lambda := \left\{ (\varphi, \psi) \in C_b(\mathbb{R}^d) \times C_b(\mathbb{R}^d) : \varphi(x) + \psi(y) \leq \lambda, \text{ if } c(x, y) \leq \lambda \right\}.$$

Moreover the infimum in (4.1.2) is achieved for

$$\lambda_\infty = \min_{\Pi(\mu, \nu)} C_\infty(\gamma).$$

*Remark 4.1.36.* Theorem 4.1.35 above implies that

$$\min_{\Pi(\mu, \nu)} (\text{OT}_\infty) = \sup \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu + \int_{\mathbb{R}^d} \psi(y) d\nu : \varphi(x) + \psi(y) \leq \lambda_\infty \right\}.$$

Unfortunately the supremum with respect to  $(\varphi, \psi)$  always admits the trivial solution  $\varphi = \psi = \frac{\lambda_\infty}{2}$ , making hard a successful application of this theorem.

The first result regarding existence of Monge-type solutions in  $L^\infty$  was found by De Pascale, Champion and Juutinen in [47] who proved that for  $X = Y = \mathbb{R}^d$  and  $c(x, y) = |x - y|$ , not only we have existence of optimal maps but also that among the optimal transport plans for  $(\text{OT}_\infty)$ , the  $\infty$ -cm plans are actually concentrated on a graph of a measurable function. We state here (Theorem 4.1.37 below) the generalized version of that result proved by Jylä in [79].

**Theorem 4.1.37** (Theorem 3.5 in [79]). *Let the cost function  $c$  be of the form  $c(x, y) := h(x - y)$ , where*

$$h : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+ \text{ is continuous and strictly quasiconvex}$$

and let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ , with  $\mu$  such that

$$\mu(A) = 0 \text{ for every countably } (n - 1)\text{-Lipschitz set } A \subset \mathbb{R}^d.$$

Then every  $\infty$ -monotone plan  $\gamma$  is induced by a transport maps, i.e. there exists  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  Borel such that  $\gamma = (\text{Id} \times T)_\# \mu$ .

In [47] the authors also prove a uniqueness result, generalized in [79], when the target measure has atoms.

**Theorem 4.1.38** (Theorem 3.13 in [79]). *Let us assume that  $c$  and  $\mu$  are as in Theorem 4.1.37 and that  $\nu(\{y_0\}) > 0$  for some  $y_0 \in \mathbb{R}^d$ . Let  $T$  and  $\tilde{T}$  be optimal transport maps corresponding respectively to  $\infty$ -c-cyclically monotone transport plans  $\gamma$  and  $\tilde{\gamma}$ . Then*

$$\mu \left( T^{-1}(\{y_0\}) \setminus \tilde{T}^{-1}(\{y_0\}) \right) = 0.$$

**Corollary 4.1.39.** *Let us assume that  $c$ ,  $\mu$  are as in Theorem 4.1.37 and that  $\nu$  is a discrete measure, then there exists a unique  $\infty$ -c-cyclically monotone transport plan.*

The aim of the following section is to provide a further step in the direction of Brenier's type theorems in the  $L^\infty$ -optimal transport, by considering different cost functions, which are not translation invariant. Our contribution in this sense is presented in Theorem 4.2.10 and Theorem 4.2.13. The attempt is to get rid of the translation invariance trying to find a generalization of the notion of *twist condition* defined in Definition 4.1.16.

### 4.1.3 Discrete Optimal Transport

Regarding this part we refer to the book by Cuturi and Peyré ([58]), which provides an exhaustive reference for the computational part of the optimal transport problem.

With discrete measure  $\mu$  we mean a finite sum of Dirac masses

$$\mu = \sum_{i=1}^n \mu_i \delta_{x_i},$$

where  $\{x_i\}_{i=1}^n$  are the points of  $\mathbb{R}^d$  where it is concentrated on and  $\{\mu_i\}_{i=1}^n$  are the weights associated to such points. If  $\mu$  is a probability measure then  $\sum_1^n \mu_i = 1$ .

If  $\mu$  and  $\nu$  are two discrete probability measures, given a cost function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  one can associate a matrix  $(c_{i,j})_{i,j} \in M^{nm}$ , where  $c_{i,j} := c(x_i, y_j)$  is the cost needed to transport mass from  $x_i$  to  $y_j$ . In this setting the optimal transport problem and the  $L^\infty$ -optimal transport problem can be written as

$$\min \left\{ \sum_{i,j} \gamma_{i,j} c_{i,j} : \gamma \in \Pi(\mu, \nu) \right\}, \quad (\text{DOT})$$

$$\min_{\gamma \in \Pi(\mu, \nu)} \max \{c_{i,j} : \gamma_{i,j} \neq 0\}, \quad (\infty\text{-DOT})$$

where  $\gamma = \sum_{i,j} \gamma_{i,j} \delta_{(x_i, y_j)}$  satisfies the mass conservation constraint  $\gamma \in \Pi(\mu, \nu)$  if<sup>1</sup>

$$\gamma \mathbb{1}_m = \left( \sum_j \gamma_{i,j} \right) = \mu \quad \text{and} \quad \gamma^\top \mathbb{1}_n = \left( \sum_i \gamma_{i,j} \right) = \nu, \quad (4.1.3)$$

where  $\mathbb{1}_n$  is the vector of size  $n$  with all entries equal to 1. In particular  $\Pi(\mu, \nu)$  is bounded and defined by  $n + m$  equality constraints and therefore is a convex polytope, i.e. the convex hull of its (finite) extreme points (see Theorem 8.8 in [34]).

If  $n = m$  and  $\mu_i = \nu_j = \frac{1}{n}$ , by (4.1.3) we have that the set  $\Pi(\mu, \nu)$  is the set of the so-called *bi-stochastic* matrices, whose extreme point, by Birkhoff's theorem, are the matrices belonging to the set

$$\Sigma_n := \{ \gamma : \text{there exists } \sigma \in \mathcal{S}_n \text{ such that } \gamma_{i,j} = \frac{1}{n} \delta_{j, \sigma(i)} \}, \quad (4.1.4)$$

where  $\delta_{j, \sigma(i)}$  is the Kronecker delta. Being  $\gamma \rightarrow \sum \gamma_{i,j} c_{i,j}$  an affine function, the minimum is attained on the extreme points and so in this case optimal plans are permutations.

Even though the problem  $\gamma \rightarrow C_\infty(\gamma) = \max \{c_{i,j} : \gamma_{i,j} \neq 0\}$  is not linear we have a similar result.

**Proposition 4.1.40.** *If  $n = m$ , then*

$$v_\infty = \min_{\sigma \in \mathcal{S}_n} \max_i c_{i, \sigma(i)}.$$

*Proof.* From the considerations above, for any  $\gamma \in \Pi(\mu, \nu)$  there exists a finite set of indexes  $S$ , such that  $\gamma = \sum_{s \in S} a_s \gamma^s$ , with  $\sum a_s = 1$  and  $\gamma^s$  an extreme point. By Remark 4.1.20 and equation (4.1.1), we have that  $C_\infty(\gamma) = \max_{s \in S} C_\infty(\gamma^s)$ , proving that the minimum is attained on the extreme points, which belong to the set  $\Sigma_n$  defined by (4.1.4).  $\square$

---

<sup>1</sup>Here with a small abuse of notation in (4.1.3)  $\gamma$  indicates the matrix with the weights  $\gamma_{i,j}$  as entries and  $\mu, \nu$  the vectors of the weights  $\mu_i$  and  $\nu_j$  respectively.

## 4.2 $L^\infty$ -Optimal Transport for a class of strictly quasiconvex cost functions

### 4.2.1 Basic definitions and main theorem

In the following we state some definitions and results which are necessary for the proof of Theorem 4.2.10. They can also be found in [47].

**Definition 4.2.1.** Let  $y \in \mathbb{R}^d$ ,  $r > 0$  and let  $\gamma \in \Pi(\mu, \nu)$  be a transport plan. We define

$$\gamma^{-1}(B(y, r)) := \pi^1 \left( (\mathbb{R}^d \times B(y, r)) \cap \text{supp } \gamma \right)$$

In other words,  $\gamma^{-1}(B(y, r))$  is the set of points whose mass is partially or completely transported to  $B(y, r)$  by  $\gamma$ , and  $\gamma^{-1}$  corresponds to the inverse of the multimap induced by  $\text{supp } \gamma$ . Notice also that  $\gamma^{-1}(B(y, r))$  is a Borel set.

Since this notion is important in the sequel, we recall that when  $U$  is a Borel set, one has

$$\lim_{r \rightarrow 0^+} \frac{\mathcal{L}^d(U \cap B(x, r))}{\mathcal{L}^d(B(x, r))} = 1,$$

for almost every  $x$  in  $U$ : we shall call such a point  $x$  a Lebesgue point of  $U$  and we will denote by  $\text{Leb}(U)$  the Borel set of Lebesgue points.

**Definition 4.2.2.** We say that the couple  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$  is a  $\gamma$ -regular point if  $x \in \text{Leb}(\gamma^{-1}(B(y, r)))$  for any positive  $r$ . We denote

$$R(\gamma) := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : (x, y) \text{ is a } \gamma\text{-regular point}\}. \quad (4.2.1)$$

*Remark 4.2.3.* Notice first that by the closedness of  $\text{supp } \gamma$  we have that  $R(\gamma) \subset \text{supp } \gamma$ . Moreover it is not difficult to show that  $R(\gamma)$  is a Borel set. Indeed, if for every fixed  $n \in \mathbb{N}$  we consider a countable covering  $\mathbb{R}^d \subset \bigcup_{i \in \mathbb{N}} B(y_{i,n}, 2^{-n})$  of balls of radius  $2^{-n}$ , then

$$R(\gamma) = \bigcap_{n \geq 1} \bigcup_{i \in \mathbb{N}} (\text{Leb}(\gamma^{-1}(B(y_{i,n}, 2^{-n}))) \times B(y_{i,n}, 2^{-n})). \quad (4.2.2)$$

**Lemma 4.2.4.** Let  $\gamma \in \Pi(\mu, \nu)$  and assume that  $\mu \ll \mathcal{L}^d$ . Then  $\gamma$  vanishes outside the set  $R(\gamma)$  of  $\gamma$ -regular points.

*Proof.* Let us denote by  $S := \text{supp } \gamma \setminus R(\gamma)$  the complement of the  $\gamma$ -regular points, intersected with the support of  $\gamma$ , i.e.

$$S := \{(x, y) \in \text{supp } \gamma : x \notin \text{Leb}(\gamma^{-1}(B(y, r))) \text{ for some positive } r\}.$$

One can prove that  $S$  is actually equal to the set

$$\bigcup_{n \geq 1} \bigcup_{i \in \mathbb{N}} ((\gamma^{-1}(B(y_{i,n}, 2^{-n})) \setminus \text{Leb}(\gamma^{-1}(B(y_{i,n}, 2^{-n})))) \times B(y_{i,n}, 2^{-n})) \cap \text{supp } \gamma.$$

In particular we have that

$$\pi^1(S) = \bigcup_{n \geq 1} \bigcup_{i \in \mathbb{N}} (\gamma^{-1}(B(y_{i,n}, 2^{-n})) \setminus \text{Leb}(\gamma^{-1}(B(y_{i,n}, 2^{-n})))) ,$$

## 4.2. $L^\infty$ -OPTIMAL TRANSPORT FOR A CLASS OF STRICTLY QUASICONVEX COST FUNCTIONS

---

and the set on the right-hand side has Lebesgue measure 0. Therefore, by absolute continuity of  $\mu$  with respect to  $\mathcal{L}^d$  we have  $\mu(\pi^1(S)) = 0$ . Finally

$$\gamma(S) \leq \gamma(\pi^1(S) \times \mathbb{R}^d) = \mu(\pi^1(S)) = 0.$$

□

**Notation.** At this point, it is natural to introduce a more refined definition

$$\gamma_*^{-1}(B(y, r)) := \pi^1 \left( (\mathbb{R}^d \times B(y, r)) \cap R(\gamma) \right).$$

For future use we also introduce a suitable notation for a cone: let  $x_0, \xi \in \mathbb{R}^d$  such that  $|\xi| = 1$  and  $\delta \in [0, 2]$  then

$$K(x_0, \xi, \delta) := \left\{ x \in \mathbb{R}^d \setminus \{x_0\} : \frac{x - x_0}{|x - x_0|} \cdot \xi \geq 1 - \delta \right\} \cup \{x_0\},$$

and for a “truncated” cone

$$K(x_0, \xi, \delta, s) := K(x_0, \xi, \delta) \cap B(x_0, s). \quad (4.2.3)$$

**Proposition 4.2.5.** *Let  $\mu \ll \mathcal{L}^d$ , let  $(x_0, y_0) \in R(\gamma)$  and let  $r > 0$ , then for every  $\xi \in \partial B(0, 1)$ ,  $\delta \in (0, 2]$  and  $s > 0$  it holds:*

$$\mathcal{L}^d(\gamma_*^{-1}(B(y_0, r)) \cap K(x_0, \xi, \delta, s)) > 0.$$

*In general, if  $A$  is a Borel set and  $x_0 \in \text{Leb}(A)$  it holds:*

$$\mathcal{L}^d(A \cap \gamma_*^{-1}(B(y_0, r)) \cap K(x_0, \xi, \delta, s)) > 0.$$

*Proof.* By definition of  $R(\gamma)$ ,  $x_0$  is a Lebesgue point of  $\gamma^{-1}(B(y_0, r))$  which implies

$$\mathcal{L}^d(\gamma^{-1}(B(y_0, r)) \cap K(x_0, \xi, \delta, s)) > 0$$

for every  $\xi \in \partial B(0, 1)$ ,  $\delta \in (0, 2]$ ,  $s > 0$ . Thus there exists

$$(x, y) \in (K(x_0, \xi, \delta, s) \times B(y_0, r)) \cap \text{supp } \gamma$$

such that  $x \neq x_0$ . We take  $\rho > 0$  small enough so that  $B(x, \rho) \subset K(x_0, \xi, \delta, s)$  and  $B(y, \rho) \subset B(y_0, r)$ . We know that  $\gamma(B(x, \rho) \times B(y, \rho)) > 0$  because  $(x, y) \in \text{supp } \gamma$ , and since  $\gamma$  is concentrated in  $R(\gamma)$  we have:

$$\begin{aligned} 0 &< \gamma((B(x, \rho) \times B(y, \rho)) \cap R(\gamma)) \\ &\leq \gamma((K(x_0, \xi, \delta, s) \times B(y_0, r)) \cap R(\gamma)) \\ &\leq \gamma(K(x_0, \xi, \delta, s) \cap \gamma_*^{-1}(B(y_0, r)) \times \mathbb{R}^d) \\ &= \mu(K(x_0, \xi, \delta, s) \cap \gamma_*^{-1}(B(y_0, r))). \end{aligned}$$

Which in turn implies  $\mathcal{L}^d(K(x_0, \xi, \delta, s) \cap \gamma_*^{-1}(B(y_0, r))) > 0$  by the fact that  $\mu \ll \mathcal{L}^d$ . □

**Corollary 4.2.6.** *Let  $(x_0, y_0) \in R(\gamma)$  and let  $C$  be a convex set with  $\text{int } C \neq \emptyset$  such that  $x_0 \in C$ . Then for every  $r > 0$*

$$\mathcal{L}^d(\gamma_*^{-1}(B(y_0, r) \cap \text{int } C)) > 0.$$

Moreover, if  $A$  is a Borel set and  $x_0 \in \text{Leb}(A)$

$$\mathcal{L}^d(A \cap \gamma_*^{-1}(B(y_0, r) \cap \text{int } C)) > 0.$$

*Proof.* We observe that by convexity of  $C$  there exists a cone  $K(x_0, \xi, \delta)$ , for some  $\xi \in \partial B(0, 1)$  and  $\delta \in (0, 1)$ , such that for  $s$  sufficiently small

$$K(x_0, \xi, \delta, s) \subset \text{int } C \cup \{x_0\}.$$

By the monotonicity of the Lebesgue measure, the claim now follows from Proposition 4.2.5.  $\square$

Before the first main result of this paper we present a technical lemma on convex sets that will be useful in the sequel.

**Lemma 4.2.7.** *Let  $B, C \subset \mathbb{R}^d$  be two closed convex sets with nonempty interiors, differentiable boundaries, and such that  $B \cap C \neq \emptyset$ . Let  $x \in \partial B \cap \partial C$  be such that  $n_B(x) \neq n_C(x)$ , where  $n_B(x)$  and  $n_C(x)$  are the unit outer normals of  $B$  and  $C$  at  $x$ , respectively. Then there exists a point  $a \in \text{int } B \cap \{w : (w - x) \cdot n_C(x) > 0\}$  and  $\delta, s > 0$  such that the intersection of the cone of direction  $a - x$  and amplitude  $\delta$  with the ball centered at  $x$  and of radius  $s$  is all contained in  $\text{int } B$ , that is*

$$K\left(x, \frac{a - x}{|a - x|}, \delta, s\right) \subset \text{int } B \cup \{x\}. \quad (4.2.4)$$

Moreover for the reverse cone we have

$$K\left(x, -\frac{a - x}{|a - x|}, \delta, s\right) \subset \text{int } C \cup \{x\}. \quad (4.2.5)$$

*Proof.* By the differentiability of the boundaries of sets  $B$  and  $C$  we have that  $n_B(x)$  and  $n_C(x)$  are well-defined elements of  $\mathbb{R}^d$ . In particular, by convexity  $n_B(x)$  is the only unit vector such that

$$(b - x) \cdot n_B(x) \leq 0, \quad \text{for every } b \in B.$$

Hence there exists  $a \in \text{int } B$  such that

$$(a - x) \cdot n_C(x) > 0. \quad (4.2.6)$$

By choosing a suitable  $s > 0$  and  $\delta$  close enough to 0, by the continuity of the scalar product and the convexity of  $B$  we can fix the cone  $K\left(x, \frac{a - x}{|a - x|}, \delta, s\right)$  that satisfies the first claim (4.2.4).

Moreover, by Condition (4.2.6) we have  $-\frac{a - x}{|a - x|} \cdot n_C(x) < 0$ . Thus for a suitable choice of

## 4.2. $L^\infty$ -OPTIMAL TRANSPORT FOR A CLASS OF STRICTLY QUASICONVEX COST FUNCTIONS

$\tau > 0$ , we have<sup>2</sup>  $x - \frac{a-x}{|a-x|}\tau \in \text{int } C$ . Arguing as before, there exists  $s' < \left| x - \frac{a-x}{|a-x|}\tau - x \right| = \tau$  and  $\delta'$  close enough to 0, such that  $K\left(x, -\frac{a-x}{|a-x|}, \delta', s'\right)$  satisfies the (4.2.5). Therefore, possibly substituting  $s$  with  $\min\{s, s'\}$  and  $\delta$  with  $\min\{\delta, \delta'\}$ , (4.2.4) and (4.2.5) hold simultaneously.  $\square$

**Notation.** Given  $y \in \mathbb{R}^d$  and  $\lambda \in \mathbb{R}$  we denote, whenever it is possible, by

$$n_{c(\cdot, y)}(x)$$

the unit outer normal to the sublevel set  $C_\lambda = \{z : c(z, y) \leq \lambda\}$  at the point  $x \in \partial C_\lambda$ .

Before stating the following theorem we recall that the definition of quasiconvex and strictly quasiconvex function is given by Definition 2.1.8. We also recall Remark 2.1.9 about the properties of the sublevel sets.

**Theorem 4.2.8.** *Let  $\mu, \nu$  be two Borel probability measures on  $\mathbb{R}^d$  with compact supports and  $\mu \ll \mathcal{L}^d$ . Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a continuous function satisfying the following properties:*

- (i) *For all  $x$  there exists a unique  $y \in \mathbb{R}^d$  such that  $c(x, y) = 0$ , and for every  $y \in \mathbb{R}^d$  we have<sup>3</sup>  $\inf_{x \in \mathbb{R}^d} c(x, y) = 0$ ;*
- (ii)  *$c(\cdot, y)$  is strictly quasiconvex for every  $y$ , i.e.*

$$c((1-t)x + t\bar{x}, y) < \max\{c(x, y), c(\bar{x}, y)\}$$

*for all  $t \in (0, 1)$  and for every  $x, \bar{x} \in \mathbb{R}^d$ ,  $x \neq \bar{x}$ ;*

- (iii) *for all  $\lambda > 0$  and  $y \in \mathbb{R}^d$  the convex set  $C_\lambda = \{z : c(z, y) \leq \lambda\}$  has differentiable ( $C^1$ ) boundary;*
- (iv)  *$c$  satisfies a “twist kind” condition, that is: for every  $x, y, \tilde{y} \in \mathbb{R}^d$  and for all  $\lambda > 0$*

$$\begin{cases} c(x, y) = c(x, \tilde{y}) = \lambda \\ n_{c(\cdot, y)}(x) = n_{c(\cdot, \tilde{y})}(x) \end{cases} \implies y = \tilde{y}.$$

*Let  $\gamma \in \Pi(\mu, \nu)$  and  $(x, \tilde{y}) \in R(\gamma)$ . Then for every  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$  with  $y \neq \tilde{y}$ , there exists  $(x', y') \in R(\gamma)$  such that*

$$\max\{c(x', y), c(x, y')\} < \max\{c(x, y), c(x', y')\}. \quad (4.2.7)$$

*Moreover, if  $x \in \text{Leb}(A)$ , for some Borel set  $A$ , then  $(x', y')$  above can be taken with  $x' \in A$ .*

<sup>2</sup>There exist  $\varepsilon > 0$  and a concave function  $\phi : B^{d-1}(0, \rho) \rightarrow \mathbb{R}$  such that  $x = (\mathbf{0}, \phi(\mathbf{0}))$  and such that  $\partial C$  is the graph of  $\phi$  in a sufficiently small neighborhood of  $x$ . In this setting  $n_C(x) = \frac{1}{\sqrt{1 + |\nabla \phi(\mathbf{0})|^2}}(-\nabla \phi(\mathbf{0}), 1)$ . We use the notation  $(\mathbf{v}, v_d) = v = -\frac{a-x}{|a-x|}$ , where  $\mathbf{v} \in \mathbb{R}^{d-1}$  and  $v_d \in \mathbb{R}$ . The condition  $v \cdot n_C(x) < 0$  becomes  $\mathbf{v} \cdot \nabla \phi(\mathbf{0}) > v_d$ . Proving that  $x + \tau v \in \text{int } C$ , for some  $\tau > 0$ , can be done by showing that  $\phi(\tau \mathbf{v}) > \phi(\mathbf{0}) + \tau v_d$ . The latter inequality follows directly by the fact that  $\lim_{t \rightarrow 0} \frac{\phi(t\mathbf{v}) - \phi(\mathbf{0})}{t} > v_d$ .

<sup>3</sup>By the strict quasiconvexity of  $c(\cdot, y)$  (Assumption (ii)), if there exists a point  $x$  such that  $c(x, y) = \inf_{z \in \mathbb{R}^d} c(z, y)$ , then such a point is also unique.

*Remark 4.2.9.* We call condition (iv) above “twist kind” because whenever  $c$  is differentiable with respect to  $x$ , if  $\nabla_x c(x, \tilde{y}) \neq 0$  and  $\nabla_x c(x, y) \neq 0$ , then the condition becomes

$$\begin{cases} c(x, y) = c(x, \tilde{y}) \\ \frac{\nabla_x c(x, y)}{|\nabla_x c(x, y)|} = \frac{\nabla_x c(x, \tilde{y})}{|\nabla_x c(x, \tilde{y})|} \end{cases} \implies y = \tilde{y}.$$

This looks like the twist condition in the classical integral optimal transport.

*Proof. Case 1:*  $c(x, y) > c(x, \tilde{y})$ .

First of all we fix  $\lambda = c(x, y)$  and  $C_\lambda = \{z : c(z, y) \leq \lambda\}$ . Notice that  $\text{int } C_\lambda$  is nonempty by part (2) of Remark 2.1.9. Indeed  $c(x, y) = \lambda > 0 = \inf_{w \in \mathbb{R}^d} c(w, \tilde{y})$ , by Assumption (i). Now, the continuity of the cost function allows us to fix  $r > 0$  such that

$$c(x, y') < c(x, y) \quad \text{for all } y' \in B(\tilde{y}, r). \quad (4.2.8)$$

Moreover, since  $(x, \tilde{y}) \in R(\gamma)$ , thanks to Corollary 4.2.6, we know that there exists  $(x', y') \in R(\gamma)$  with  $x' \in \text{int } C_\lambda = \{z : c(z, y) < \lambda\}$  (so that  $x \neq x'$ ) and  $y' \in B(\tilde{y}, r)$  such that

$$(x', y') \in (\text{int } C_\lambda \times B(\tilde{y}, r)) \cap R(\gamma).$$

Notice that such a  $y'$  exists by the definition of  $\gamma_*^{-1}(B(\tilde{y}, r))$ .

By Condition (4.2.8) we have

$$c(x, y') < c(x, y)$$

We also have that

$$c(x', y) < c(x, y),$$

because  $x' \in \text{int } C_\lambda$ . These two inequalities imply that

$$\max\{c(x', y), c(x, y')\} < c(x, y) \leq \max\{c(x, y), c(x', y')\},$$

concluding the proof for the Case 1.

Case 1 bis: If  $x \in \text{Leb}(A)$ , for  $A$  being a Borel set, then by the second part of Corollary 4.2.6, we can find  $(x', y') \in R(\gamma)$ , such that  $x' \in \text{int } C_\lambda \cap A$  and  $y \in B(\tilde{y}, r)$ . As before, such  $(x', y')$  satisfies the (4.2.7).

Case 2:  $c(x, y) < c(x, \tilde{y})$ .

For this case, we follow a path similar to that of [79] but with a different construction of cones.

This time we fix  $\lambda = c(x, \tilde{y})$ , so that  $x \in \partial \tilde{C}_\lambda$ , where  $\tilde{C}_\lambda = \{z : c(z, \tilde{y}) \leq \lambda\}$ . Again we denote by  $n_{c(\cdot, \tilde{y})}(x)$  the unit outer normal to the set  $\tilde{C}_\lambda$  at  $x$ .

Let  $s$  be a small, positive real number such that  $\tilde{a} := x - sn_{c(\cdot, \tilde{y})}(x) \in \text{int } \tilde{C}_\lambda$ , which is not empty thanks to (2) of Remark 2.1.9. We fix

$$\varepsilon = \frac{1}{4} \min \{c(x, \tilde{y}) - c(\tilde{a}, \tilde{y}), c(x, \tilde{y}) - c(x, y)\}.$$

Thanks to the continuity of  $c$  it is possible to find a positive radius  $r$  such that  $B(\tilde{a}, r) \subset \tilde{C}_\lambda$  and for every  $(x', y') \in B(\tilde{a}, r) \times B(\tilde{y}, r)$  we have

$$c(x', y') < c(\tilde{a}, \tilde{y}) + \varepsilon \quad (4.2.9)$$

4.2.  $L^\infty$ -OPTIMAL TRANSPORT FOR A CLASS OF STRICTLY QUASICONVEX COST FUNCTIONS

---

and for every  $y' \in B(\tilde{y}, r)$  it holds

$$c(x, \tilde{y}) - \varepsilon < c(x, y'). \quad (4.2.10)$$

Consider the set  ${}^4 \tilde{K} = \{(1-t)x + tB(\tilde{a}, r) : t \in (0, 1)\}$ . Thanks to the convexity,  $\tilde{K}$  is contained in  $\tilde{C}_\lambda$ . We claim that

$$\tilde{K} \subset \{z : c(z, y') < c(x, y')\} \quad \text{for every } y' \in B(\tilde{y}, r).$$

Let  $z \in \tilde{K}$ . Then  $z = (1-t)x + t\tilde{z}$  for some  $\tilde{z} \in B(\tilde{a}, r)$  and some  $t \in (0, 1)$ . We observe that  $c(\tilde{z}, y') < c(x, y')$ . Indeed

$$c(\tilde{z}, y') < c(\tilde{a}, \tilde{y}) + \varepsilon \leq \frac{1}{4}c(x, \tilde{y}) + \frac{3}{4}c(\tilde{a}, \tilde{y}) < c(x, \tilde{y}) - 3\varepsilon < c(x, y'), \quad (4.2.11)$$

where the first inequality is due to Condition (4.2.9), the second and the third ones to the definition of  $\varepsilon$  and the last one to (4.2.10). We conclude by the strict quasi-convexity of  $c(\cdot, y')$ , which implies

$$c(z, y') < \max\{c(x, y'), c(\tilde{z}, y')\} = c(x, y').$$

On the other hand, if we consider the reflected cone  $K = (1-t)x + tB(\tilde{a}, r)$  with  $t \in (-1, 0)$ , again by the strict quasiconvexity of  $c(\cdot, y')$  we show that

$$K \subset \{z : c(z, y') > c(x, y')\} \quad \text{for every } y' \in B(\tilde{y}, r).$$

Indeed, let  $z \in K$ , then  $z = (1-t)x + t\tilde{z}$ , for some  $\tilde{z} \in B(\tilde{a}, r)$  and  $t \in (-1, 0)$ . So  $x = sz + (1-s)\tilde{z}$ , with  $s = \frac{1}{1-t}$ . Thus  $c(x, y') < \max\{c(z, y'), c(\tilde{z}, y')\} = c(z, y')$ , where the last equality is due to (4.2.11), and therefore for every  $x' \in K$  and  $y' \in B(\tilde{y}, r)$  it holds

$$c(x, y') < c(x', y').$$

Let now  $\rho > 0$  such that  $c(x', y) < c(x, y) + \varepsilon$  for every  $x' \in K \cap B(x, \rho)$ . Then for every  $x' \in K \cap B(x, \rho)$  and  $y' \in B(\tilde{y}, r)$  it holds:

$$c(x', y) < c(x, y) + \varepsilon < \frac{1}{4}c(x, \tilde{y}) + \frac{3}{4}c(x, y) \leq c(x, \tilde{y}) - 3\varepsilon < c(x, y') < c(x', y').$$

By choosing  $\rho$  small enough such that  $K \cap B(x, \rho)$  is a “truncated cone” of the form (4.2.3), the proof of Case 2 is concluded thanks to Proposition 4.2.5, for which it is possible to take  $x'$  in the set

$$(B(x, \rho) \cap K \cap \gamma_*^{-1}(B(\tilde{y}, r)))$$

and therefore  $(x', y')$  in the set

$$((K \cap B(x, \rho)) \times B(\tilde{y}, r)) \cap R(\tilde{\gamma}).$$

Case 2 bis: If  $x \in \text{Leb}(A)$  for some Borel set  $A$ , by the last part of Proposition 4.2.5, we can choose  $(x', y')$  satisfying the (4.2.7) in the set

$$((A \cap K \cap B(x, \rho)) \times B(\tilde{y}, r)) \cap R(\tilde{\gamma}).$$

**Case 3:**  $c(x, y) = c(x, \tilde{y}) = \lambda > 0$  (if  $c(x, y) = c(x, \tilde{y}) = 0$  then  $y = \tilde{y}$  by Assumption (i)) and  $n_{c(\cdot, y)}(x) \neq n_{c(\cdot, \tilde{y})}(x)$ .

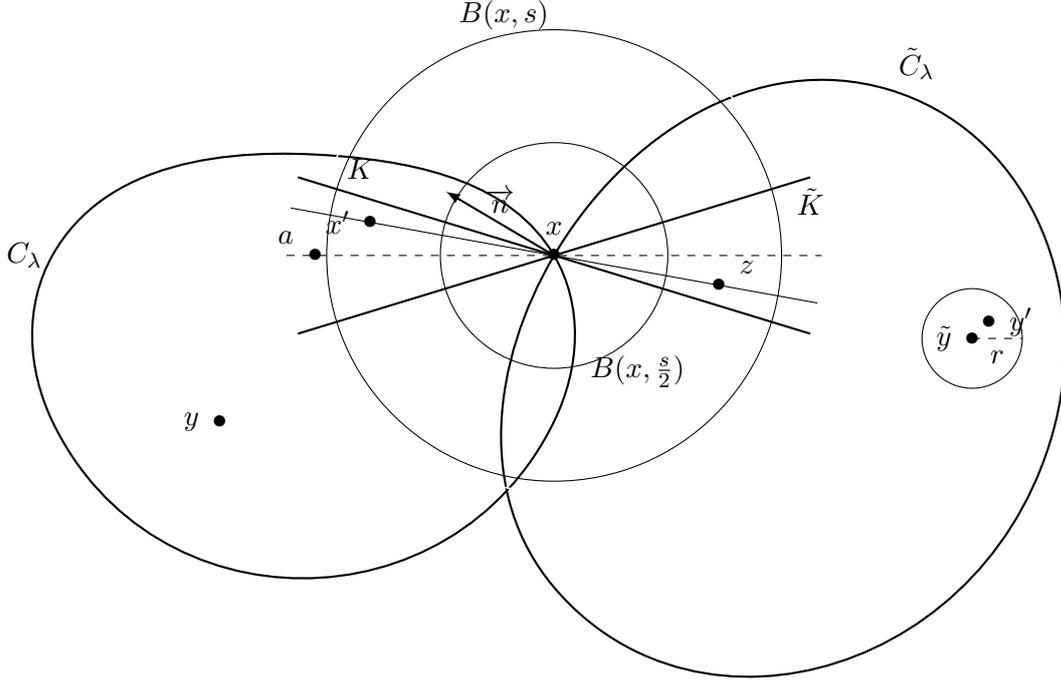


Figure 4.2.1: Case 2 - Theorem 4.2.10

In Figure 4.2.1 we provide a visual description of the proof of this case, where we have denoted the vector  $n_{c(\cdot, \tilde{y})}(x)$  by  $\vec{n}$ .

Since  $x \in \partial C_\lambda \cap \partial \tilde{C}_\lambda$  and since by part (2) of Remark 2.1.9 these sublevel sets have nonempty interiors, the assumptions of Lemma 4.2.7 above are satisfied. Therefore we can construct two “truncated” cones, one inside  $\text{int } C_\lambda \cup \{x\}$  and the “reverse” one contained in  $\text{int } \tilde{C}_\lambda$ . More precisely there exist  $a \in \text{int } C_\lambda$  and  $\delta, s > 0$  such that

$$K \left( x, \frac{a-x}{|a-x|}, \delta, s \right) \subset \text{int } C_\lambda \cup \{x\} \text{ and}$$

$$K \left( x, -\frac{a-x}{|a-x|}, \delta, s \right) \subset \text{int } \tilde{C}_\lambda \cup \{x\}.$$

In order to simplify the notation, let us denote by

$$K := K \left( x, \frac{a-x}{|a-x|}, \delta, s \right) \quad \text{and} \quad \tilde{K} := K \left( x, -\frac{a-x}{|a-x|}, \delta, s \right).$$

We observe that  $c(x, \tilde{y}) > c(z, \tilde{y})$  for every  $z \in \text{int } \tilde{C}_\lambda$ , so if we take  $z$  in a set well-contained in  $\text{int } \tilde{C}_\lambda$ , for instance the portion of annulus  $\tilde{K} \setminus B(x, \frac{s}{2})$ , since  $\tilde{K} \setminus B(x, \frac{s}{2})$  is compact and  $c(\cdot, \tilde{y})$  is continuous,  $c(x, \tilde{y}) > \min\{c(z, \tilde{y}) : z \in \tilde{K} \setminus B(x, \frac{s}{2})\}$ . Moreover

<sup>4</sup>We use the notation  $\tilde{K}$  because also this set is a cone, even if slightly different from the cones defined by (4.2.3).

4.2.  $L^\infty$ -OPTIMAL TRANSPORT FOR A CLASS OF STRICTLY QUASICONVEX COST FUNCTIONS

---

$y \mapsto \min\{c(z, y) : z \in \tilde{K} \setminus B(x, \frac{s}{2})\}$  is continuous for every  $y$ . Therefore, there exists  $r > 0$  such that

$$c(x, y') > c(z, y'), \quad \text{for all } y' \in B(\tilde{y}, r), \text{ for all } z \in \tilde{K} \setminus B(x, \frac{s}{2}). \quad (4.2.12)$$

Let  $(x', y') \in (K \times B(\tilde{y}, r)) \cap R(\gamma)$  with  $x' \neq x$ , which exists by Proposition 4.2.5 (and definition of  $\gamma_*^{-1}(\tilde{y}, r)$ ). Now, if we take  $z \in \tilde{K} \setminus B(x, \frac{s}{2})$  on the line passing from  $x'$  and  $x$  (the order of the points on the line being  $x', x, z$ ), by the strict quasiconvexity of  $c$  with respect to the first variable we have

$$c(x, y') < \max\{c(x', y'), c(z, y')\}.$$

By Condition (4.2.12), we infer that  $c(x, y') < c(x', y')$ .

Finally, since  $x' \in \text{int } C_\lambda$ , we have  $c(x', y) < c(x, y)$  (again we use that, by the continuity and the strict quasiconvexity of  $c$ ,  $\partial C_\lambda = \{z : c(z, y) = \lambda\}$ ). Hence Condition (4.2.7) holds.

Case 3 bis: Also in this case thanks to the second part of Proposition 4.2.5, we can choose our  $(x', y') \in ((A \cap K) \times B(\tilde{y}, r)) \cap R(\gamma)$  if  $x \in \text{Leb}(A)$ , for some Borel set  $A$ .

Case 4:  $c(x, y) = c(x, \tilde{y}) = \lambda > 0$  and

$$n_{c(\cdot, y)}(x) = n_{c(\cdot, \tilde{y})}(x).$$

If this case occurs, by the assumption (iv) (the “twist kind” condition) we know that  $y = \tilde{y}$  against our assumption.  $\square$

We are now ready to state and prove the main result of this section, which follows directly by Theorem 4.2.10.

**Theorem 4.2.10.** *Let  $\mu, \nu$  be two Borel probability measures on  $\mathbb{R}^d$  with compact supports and  $\mu \ll \mathcal{L}^d$  and let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a continuous function satisfying the assumptions (i)-(iv) of Theorem 4.2.8. If  $\gamma \in \Pi(\mu, \nu)$  is  $\infty$ -monotone, then  $\gamma$  vanishes outside the graph of a Borel map  $T \in \mathcal{T}(\mu, \nu)$ .*

*Proof.* Thanks to Lemma 4.2.4 we know that  $\gamma$  is concentrated on the set of  $\gamma$ -regular points  $R(\gamma)$  and that the set  $R(\gamma)$  is a Borel set. Therefore, it suffices to show that the set  $R(\gamma)$  is the graph of a function  $T \in \mathcal{T}(\mu, \nu)$ . More precisely it is enough to prove that the set  $R(\gamma)$  is contained in the graph of some function  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . Indeed, Theorem 2.3 of [3] ensures that  $T$  is a Borel map and that  $\gamma = (Id \times T)_\# \mu$ . Let us assume, by contradiction, that there exist  $(x, y), (x, \tilde{y}) \in R(\gamma)$  with  $y \neq \tilde{y}$ . Then we can apply Theorem 4.2.8, thanks to which it holds that there exists  $(x', y') \in R(\gamma)$  such that

$$\max\{c(x', y), c(x, y')\} < \max\{c(x, y), c(x', y')\}.$$

The above inequality implies that  $R(\gamma)$  is not  $\infty$ -monotone and, since  $R(\gamma) \subset \text{supp } \gamma$ , neither is  $\text{supp } \gamma$ . This fact, together with Lemma 4.1.27 leads us to a contradiction.  $\square$

### 4.2.2 On the uniqueness of ICM optimal transport plans

The next lemma is slightly less general than its equivalent in [79] but we believe it makes more transparent the proof of uniqueness that will follow.

**Lemma 4.2.11.** *Let  $\mu$  and  $\nu$  be two probability measures on  $\mathbb{R}^d$ , with compact supports and  $\mu \ll \mathcal{L}^d$ . Assume that the function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  satisfies the assumptions of Theorem 4.2.8. Let  $T, \tilde{T} \in \mathcal{T}(\mu, \nu)$  be two transport maps associated to two  $\infty$ -monotone transport plans  $\gamma$  and  $\tilde{\gamma}$ , i.e.,  $\gamma = (Id \times T)_\# \mu$  and  $\tilde{\gamma} = (Id \times \tilde{T})_\# \mu$ . Let  $A \subset \mathbb{R}^d$  be a Borel set such that for all  $x \in A$ , one has  $T(x) \neq \tilde{T}(x)$ . Define*

$$\mathcal{R}(A) = \{z \in \mathbb{R}^d : \exists x \in A \text{ s.t.} \\ \max\{c(x, T(z)), c(z, \tilde{T}(x))\} < \max\{c(x, \tilde{T}(x)), c(z, T(z))\}\}.$$

Then  $\mu(A \setminus \mathcal{R}(A)) = 0$ .

*Remark 4.2.12.* The definition of  $\mathcal{R}(A)$  is not symmetric with respect to the interchange of roles of  $T$  and  $\tilde{T}$ , in the sense that

$$\tilde{\mathcal{R}}(A) = \{z \in \mathbb{R}^d : \exists x \in A \text{ s.t.} \\ \max\{c(x, \tilde{T}(z)), c(z, T(x))\} < \max\{c(x, T(x)), c(z, \tilde{T}(z))\}\}.$$

is not necessarily equal to  $\mathcal{R}(A)$ . However, the same result can also be proven for  $\tilde{\mathcal{R}}(A)$ .

*Proof.* First of all we observe that

$$\begin{aligned} \mu(A \setminus \mathcal{R}(A)) &= \mu(\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap A \setminus \mathcal{R}(A)) \\ &= \mu(\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap \text{Leb}(A) \setminus \mathcal{R}(A)), \end{aligned}$$

since  $\gamma$  and  $\tilde{\gamma}$  are respectively concentrated on  $R(\gamma)$  and  $R(\tilde{\gamma})$  and  $\mu \ll \mathcal{L}^d$ . We will prove that

$$\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap \text{Leb}(A) \setminus \mathcal{R}(A) = \emptyset.$$

Let us assume by contradiction that there exists  $x \in \pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap \text{Leb}(A) \setminus \mathcal{R}(A)$  and consider  $(x, T(x)) \in R(\gamma)$  and  $(x, \tilde{T}(x)) \in R(\tilde{\gamma})$ . By definition of  $A$ , we have  $T(x) \neq \tilde{T}(x)$ . Thanks to the second part of Theorem 4.2.8 applied to  $R(\tilde{\gamma})$ , we can find  $(x', y') = (x', \tilde{T}(x')) \in R(\tilde{\gamma})$  such that  $x' \in A$  and

$$\max\{c(x, \tilde{T}(x')), c(x', T(x))\} < \max\{c(x, T(x)), c(x', \tilde{T}(x'))\}. \quad (4.2.13)$$

We conclude noticing that the inequality (4.2.13) implies that  $x \in \mathcal{R}(A)$ , contradicting the fact that  $x \in A \setminus \mathcal{R}(A)$ .  $\square$

**Theorem 4.2.13.** *Let  $\mu$  and  $\nu$  be two probability measures on  $\mathbb{R}^d$ , with compact support and  $\mu \ll \mathcal{L}^d$ . Assume that the function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  satisfies the assumptions of Theorem 4.2.8 and that  $\nu(\{y_0\}) > 0$  for some  $y_0 \in \mathbb{R}^d$ . Let  $T$  and  $\tilde{T}$  be optimal transport maps corresponding respectively to  $\infty$ -c-cyclically monotone transport plans  $\gamma$  and  $\tilde{\gamma}$ . Then*

$$\mu\left(T^{-1}(\{y_0\}) \setminus \tilde{T}^{-1}(\{y_0\})\right) = 0.$$

4.2.  $L^\infty$ -OPTIMAL TRANSPORT FOR A CLASS OF STRICTLY QUASICONVEX COST FUNCTIONS

---

*Proof.* Assume, by contradiction, that

$$\mu\left(T^{-1}(\{y_0\}) \setminus \tilde{T}^{-1}(\{y_0\})\right) > 0.$$

We first restrict to the set of the full  $\mu$ -measure where both the graphs of  $T$  and  $\tilde{T}$  are  $\infty$ - $c$ -cyclically monotone (by Theorem 4.2.10 we know that we may consider the set  $\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma}))$ ). We may also assume, restricting our attention to the set of interest, that

$$T(x) \neq \tilde{T}(x) \quad \text{for all } x \in \pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})). \quad (4.2.14)$$

In this proof we will apply Lemma 4.2.11 repeatedly to different subsets  $A$  of  $\text{supp } \mu$ , this is always possible by Assumption (4.2.14).

We define  $B_0 = \{y_0\}$ ,  $A_1 = T^{-1}(B_0)$ ,  $A'_1 = \mathcal{R}(A_1)$ , and  $B_1 = \tilde{T}(A'_1)$ . We continue recursively: assuming that for all  $j \in \{1, \dots, k-1\}$  the sets  $A_j, A'_j$ , and  $B_j$  have already been defined we set

$$\begin{aligned} A_k &= T^{-1}(B_{k-1}), & A'_k &= \mathcal{R}(A_k), & \text{and} \\ B_k &= \tilde{T}(A'_k). \end{aligned}$$

We observe that by construction  $\mu(A_j) > 0$  and, by Lemma 4.2.11, also  $\mu(A'_j) > 0$  for every  $j$ . We continue by defining

$$P_1 = \bigcup_{k \geq 1} A_k \quad \text{and} \quad P_2 = \bigcup_{k \geq 1} B_k.$$

We prove the following intermediate claim:

**Claim**  $y_0 \in P_2$ .

*Proof.* Let us first show that

$$\mu(P_1) = \nu(P_2). \quad (4.2.15)$$

We prove that  $\nu(P_2) \leq \mu(P_1)$ . Right from the definitions of the sets  $P_1, P_2, A_k$ , and  $B_k$  we see that

$$T^{-1}(B_0 \cup P_2) = P_1.$$

Since  $\nu = T_{\#}\mu$ , we get

$$\nu(P_2) \leq \nu(B_0 \cup P_2) = \mu(T^{-1}(B_0 \cup P_2)) = \mu(P_1). \quad (4.2.16)$$

In order to prove the other direction we first observe that, by using Lemma 4.2.11 with  $A = A_k$ , we have

$$\mu(A_k \setminus \mathcal{R}(A_k)) = 0 \quad \text{for all } k \geq 1; \quad (4.2.17)$$

remember that in our construction  $A'_k = \mathcal{R}(A_k)$ . Now

$$\begin{aligned} \mu(P_1) &= \mu\left(\bigcup_{k \geq 1} A_k\right) \stackrel{a)}{\leq} \mu\left(\bigcup_{k \geq 1} A'_k\right) \\ &\leq \mu\left(\bigcup_{k \geq 1} \tilde{T}^{-1}(\tilde{T}(A'_k))\right) = \mu\left(\tilde{T}^{-1}\left(\bigcup_{k \geq 1} \tilde{T}(A'_k)\right)\right) \\ &= \nu\left(\bigcup_{k \geq 1} \tilde{T}(A'_k)\right) = \nu\left(\bigcup_{k \geq 1} B_k\right) = \nu(P_2) \end{aligned}$$

where inequality a) is due to Condition (4.2.17). So we have  $\mu(P_1) \leq \nu(P_2)$ , which completes the proof of Condition (4.2.15).

If by contradiction  $y_0 \notin P_2$ , since  $\nu(\{y_0\}) > 0$  (remember that in our construction  $\{y_0\} = B_0$ ), then

$$\nu(P_2) < \nu(B_0 \cup P_2) = \mu(P_1), \quad (4.2.18)$$

where in the last equality we have used Condition (4.2.16). Inequality (4.2.18) is in contradiction with (4.2.15) and the **Claim** is proven.

By the inclusion  $y_0 \in P_2$  we now know that  $y_0$  belongs to  $B_k$  for some  $k \geq 1$ . Therefore there exist  $x'_k \in A'_k$  (remember that  $B_k = \tilde{T}(A'_k)$ ) such that  $y_0 = \tilde{T}(x'_k)$ . Next we choose  $x_k \in A_k$  such that

$$\max\{c(x_k, \tilde{T}(x'_k)), c(x'_k, T(x_k))\} < \max\{c(x_k, T(x_k)), c(x'_k, \tilde{T}(x'_k))\}.$$

We recall that this is possible since  $A'_k = \mathcal{R}(A_k)$ .

Because by construction  $A_k = T^{-1}(B_{k-1})$  and  $x_k \in A_k$ , we have that  $T(x_k) \in B_{k-1} = \tilde{T}(A'_{k-1})$ , so there exists  $x'_{k-1} \in A'_{k-1}$  such that  $\tilde{T}(x'_{k-1}) \in B_{k-1}$  and  $T(x_k) = \tilde{T}(x'_{k-1})$ . And so we continue, decreasing with the indices down until having defined  $x'_1$  and  $x_1$ . Since  $x_1 \in A_1 = T^{-1}(B_0)$  we have that  $T(x_1) = y_0 = \tilde{T}(x'_k)$ . Thus, we have constructed two  $k$ -uples of points  $(x_j)_{j=1}^k$  and  $(x'_j)_{j=1}^k$  such that for all  $j$ ,  $x_j \in A_j$  and  $x'_j \in A'_j$ .

$$\begin{cases} T(x_j) = \tilde{T}(x'_{j-1}) & \text{for all } 2 \leq j \leq k; \\ T(x_1) = \tilde{T}(x'_k) = y_0. \end{cases}$$

and that for all  $1 \leq j \leq k$

$$\max\{c(x'_j, T(x_j)), c(x_j, \tilde{T}(x'_j))\} < \max\{c(x'_j, \tilde{T}(x'_j)), c(x_j, T(x_j))\}. \quad (4.2.19)$$

We observe that by Condition (4.2.19),  $x_j \neq x'_j$ . We now apply to the set  $\{x_j\}_{j=1}^k$  the cyclical permutation

$$\sigma(j) = \begin{cases} j+1 & \text{if } j \in \{1, \dots, k-1\} \\ 1 & \text{if } j = k \end{cases}.$$

We have

$$\begin{aligned} \max_{1 \leq j \leq k} c(x_j, T(x_j)) &\stackrel{a)}{\leq} \max_{1 \leq j \leq k} c(x_j, T(x_{\sigma(j)})) \stackrel{b)}{=} \max_{1 \leq j \leq k} c(x_j, \tilde{T}(x'_j)) \\ &\stackrel{c)}{<} \max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} \end{aligned} \quad (4.2.20)$$

where inequality a) follows from the  $\infty$ -cm assumed on  $T$ . Equality b) follows from the fact that by construction  $T(x_{\sigma(j)}) = \tilde{T}(x'_j)$  for all  $j \in \{1, \dots, k\}$ . Estimate c) follows from Condition (4.2.19). Concerning the last term in (4.2.20), we have two possibilities: either

$$\max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} = c(x_m, T(x_m)) \quad \text{for some } m$$

or

$$\max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} = \max_{1 \leq j \leq k} c(x'_j, \tilde{T}(x'_j)).$$

## 4.2. $L^\infty$ -OPTIMAL TRANSPORT FOR A CLASS OF STRICTLY QUASICONVEX COST FUNCTIONS

The first case leads immediately to a contradiction. So we are left with the latter case. We apply to the set  $\{x'_j\}_{j=1}^k$  the cyclical permutation

$$\tau(j) = \begin{cases} j-1 & \text{if } j \in \{2, \dots, k\} \\ k & \text{if } j = 1 \end{cases}.$$

We get

$$\begin{aligned} \max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} &= \max_{1 \leq j \leq k} c(x'_j, \tilde{T}(x'_j)) \\ &\stackrel{a)}{\leq} \max_{1 \leq j \leq k} c(x'_j, \tilde{T}(x'_{\tau(j)})) \stackrel{b)}{=} \max_{1 \leq j \leq k} c(x'_j, T(x_j)) \end{aligned} \quad (4.2.21)$$

$$\stackrel{c)}{<} \max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\}. \quad (4.2.22)$$

Above, in a) we have used the  $\infty$ -cm property of  $\tilde{T}$ , in b) the fact that, by construction,  $T(x_j) = \tilde{T}(x'_{\tau(j)})$  for all  $j \in \{1, \dots, k\}$ , and in c) again Condition (4.2.19). Estimate (4.2.21) is impossible, so we get a contradiction and we are done. □

**Corollary 4.2.14.** *Let us assume that  $c$ ,  $\mu$  are as in Theorem 4.2.13 and that  $\nu$  is a discrete measure, then there exists a unique  $\infty$ -c-cyclically monotone transport plan.*

*Example 4.2.15.* Let us consider  $c$ ,  $\mu$  and  $\nu$  be the same as in Example 4.1.23. Even though in this case all plans are optimal, by Theorem 4.2.13 we know that there exists just one which is  $\infty$ -cyclically monotone and it is the one which splits the unitary square in two triangles and sends the triangle of vertices  $\{(0, 0), (0, 1), (1, 1)\}$  to the point  $(1, 2)$  and the other triangle to the point  $(2, 1)$ .

### 4.2.3 Some examples

In this section we give some examples of cost functions which satisfy the assumptions of Theorem 4.2.10. We also present a counterexample which shows that the assumption of the strict quasiconvexity of  $c(\cdot, y)$  is necessary.

*Example 4.2.16.* Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a cost function of the type introduced in [79], defined by

$$c(x, y) := h(y - x),$$

where  $h : \mathbb{R}^d \rightarrow \mathbb{R}_+$  is a continuous and strictly quasiconvex function. Here, we also assume that there exists a unique  $p \in \mathbb{R}^d$  such that  $h(p) = 0$ , that  $h$  is differentiable, and that  $\nabla h(v) \neq 0$  for every  $v \neq p$ . This cost function satisfies the assumptions of Theorem 4.2.10 from which we infer that the  $\infty$ -m transportation plans corresponding to  $c$  are of the Monge type. The fact that  $c$  fulfils the condition (iv) can be seen as a special case of the proof of the example below.

*Example 4.2.17.* Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  be defined by

$$c(x, y) := h(G(y) - F(x)),$$

where  $h : \mathbb{R}^d \rightarrow \mathbb{R}_+$  is defined as in the Example 4.2.16,  $F, G : \mathbb{R}^d \rightarrow \mathbb{R}^d$  affines transformations of the form

$$F(x) = Ax + b \quad \text{and} \quad G(y) = A'y + b',$$

where  $A, A' \in M^{d \times d}$  are invertible matrices and  $b, b' \in \mathbb{R}^d$ .

We observe that the cost function  $c$  satisfies the assumptions of Theorem 4.2.10. By the assumptions on  $h$  and the invertibility of  $F$  and  $G$ , Assumption (i) is satisfied. The strict quasiconvexity of  $h$  and the linearity of  $F$  and  $G$  ensure the strict quasiconvexity of  $c(\cdot, y)$  for every  $y$ , and, for the same reason,  $c(x, \cdot)$  is strictly quasiconvex for every  $x$ .

Let us prove that the “twist kind” condition holds. Let us assume that there exist  $y, \tilde{y} \in \mathbb{R}^d$  and  $\lambda > 0$  such that  $c(x, y) = c(x, \tilde{y}) = \lambda$  and  $n_{c(\cdot, y)}(x) = n_{c(\cdot, \tilde{y})}(x)$ . We first observe that, by the fact that

$$\nabla_x c(x, y) = -\nabla F(x)^T \cdot \nabla h(G(y) - F(x)) = -A^T \cdot \nabla h(G(y) - F(x)),$$

$\nabla_x c(x, y) \neq 0$  for every  $x, y$  such that  $G(y) - F(x) \neq p$  (that is, for every  $x, y$  such that  $c(x, y) \neq 0$ ). Thus, the equality between the two unit normals becomes

$$\frac{\nabla_x c(x, y)}{|\nabla_x c(x, y)|} = \frac{\nabla_x c(x, \tilde{y})}{|\nabla_x c(x, \tilde{y})|}.$$

Since

$$\nabla_y c(x, y) = A'^T \cdot \nabla h(G(y) - F(x)),$$

thanks to the invertibility of  $A'$  we have that

$$\nabla_x c(x, y) = -A^T (A'^T)^{-1} \nabla_y c(x, y).$$

The computation above, the invertibility assumptions on  $A$  and  $A'$  and the condition on the normalized gradients imply

$$\frac{\nabla_y c(x, y)}{|\nabla_y c(x, y)|} = \frac{\nabla_y c(x, \tilde{y})}{|\nabla_y c(x, \tilde{y})|}.$$

We observe that both  $y$  and  $\tilde{y}$  belong to the boundary of the sublevel set with respect to  $x$ ,  $C_\lambda^x := \{z \in \mathbb{R}^d : c(x, z) \leq \lambda\}$ . Then  $y$  and  $\tilde{y}$  must coincide, since they are two points on the smooth boundary of a strictly convex set  $C_\lambda^x$  with the same normal.

The following is an example of an  $\infty$ -cm plan that is not given by a map for a cost function that doesn't satisfy the assumptions of Theorem 4.2.10.

*Example 4.2.18.* We consider the space  $X = \mathbb{R}^2$  with the cost function given by the  $\infty$ -distance

$$d((x_1, y_1), (x_2, y_2)) = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

We denote by  $Q := [0, 1] \times [0, 1]$ , the unit square centered at  $(\frac{1}{2}, \frac{1}{2})$ . We take  $\mu = \mathcal{L}^2|_Q$ , the 2-dimensional Lebesgue measure restricted to  $Q$ , and  $\nu = \mathcal{L}^2|_{Q+(10,0)}$ , the 2-dimensional Lebesgue measure restricted to the translation of  $Q$  by the vector  $(10, 0)$ . Let  $\gamma \in \Pi(\mu, \nu)$  be defined by

$$\gamma = \mu \otimes \mathcal{H}^1|_{\{x_1+10\} \times [0,1]},$$

that is, the transport plan which distributes evenly every point  $(x_1, x_2) \in \text{supp } \mu$  to the vertical line segment  $\{(x_1 + 10, y_2) \mid y_2 \in \mathbb{R}\} \cap (Q + (10, 0))$ . This plan is clearly not given by a map, but it is  $\infty$ -cm.

4.2.  $L^\infty$ -OPTIMAL TRANSPORT FOR A CLASS OF STRICTLY QUASICONVEX COST FUNCTIONS

---

To prove that  $\gamma$  is  $\infty$ -c-cyclically monotone we fix points  $\{(x_1^i, x_2^i), (x_1^i + 10, y_2^i)\}_{i=1}^N$  in the support of  $\gamma$  and a permutation  $\sigma$  of the set  $\{1, \dots, N\}$ . Now

$$\max_{1 \leq i \leq N} \{d((x_1^i, x_2^i), (x_1^i + 10, y_2^i))\} = \max_{1 \leq i \leq N} \{\max\{10, |x_2^i - y_2^i|\}\} = 10$$

where in the last equality we have used the fact that

$$|x_2^i - y_2^i| \leq 1 < 10.$$

Let us prove that

$$\max_{1 \leq i \leq N} \{d((x_1^i, x_2^i), (x_1^{\sigma(i)} + 10, y_2^{\sigma(i)}))\} \geq 10.$$

It is enough to show that there exists an index  $k \in \{1, \dots, N\}$  such that

$$|x_1^k - (x_1^{\sigma(k)} + 10)| \geq 10.$$

We may assume that  $\sigma$  is not the identity: for the identity permutation the  $\infty$ -cm condition holds as an equality. Let us denote  $I = \{1, \dots, N\}$  and let  $k_0$  be such that

$$x_1^{k_0} = \min\{x_1^k \mid k \in I\}.$$

The point  $x_1^{k_0}$  satisfies  $x_1^{k_0} \leq x_1^{\sigma(k_0)}$ . Therefore,

$$|x_1^{k_0} - (x_1^{\sigma(k_0)} + 10)| = 10 + (x_1^{\sigma(k_0)} - x_1^{k_0}) \geq 10$$

and we are done.



## Chapter 5

# Entropic regularization of $L^\infty$ - Optimal Transport

### 5.1 Preliminaries

#### 5.1.1 A brief survey of Entropic Optimal Transport

**Definition 5.1.1.** Given a measurable space  $(X, \mathcal{A})$  and  $\gamma, \eta \in \mathcal{P}(X)$ , the relative entropy of  $\gamma$  with respect to  $\eta$  is defined as

$$H(\gamma|\eta) := \begin{cases} \int_{\mathbb{R}^d \times \mathbb{R}^d} \log\left(\frac{d\gamma}{d\eta}\right) d\gamma = \int_{\mathbb{R}^d \times \mathbb{R}^d} \log\left(\frac{d\gamma}{d\eta}\right) d\left(\frac{d\gamma}{d\eta}\right) d\eta & \text{if } \gamma \ll \eta, \\ +\infty & \text{otherwise.} \end{cases}$$

*Remark 5.1.2.* Noting that the function  $h : [0, +\infty] \rightarrow [-e^{-1}, +\infty]$  defined by  $h(t) = t \log t$  is strictly convex, we see that the functional  $\gamma \mapsto H(\gamma|\eta)$  is convex e nonnegative (by Jensen's inequality) and strictly convex on the set where it is finite. Clearly  $H(\gamma|\eta) = 0$  if and only if  $\gamma = \eta$ .

**Assumptions.** In this chapter:

- $X = \mathbb{R}^d$ ,  $d \geq 1$ ;
- $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ ;
- $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ .

**Definition 5.1.3** (Entropic Optimal Transport Problem). Given  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ , the entropic optimal transport problem with regularization parameter  $\varepsilon$  is

$$\inf_{\Pi(\mu, \nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} cd\gamma + \varepsilon H(\gamma|\mu \otimes \nu) \right\}. \quad (\varepsilon\text{EOT})$$

This problem admits a solution under mild assumptions on  $c$ , as we will see in Theorem 5.1.9. Let us also underline the fact that  $(\varepsilon\text{EOT})$  provides a good approximation of (OT): among other convergence results (see for instance [90, 100]) we state here the  $\Gamma$ -convergence result established by Carlier *et al.* in [38].

Let us define the following functionals:

$$J_\varepsilon := \begin{cases} \int_{\mathbb{R}^d \times \mathbb{R}^d} cd\gamma + \varepsilon H(\gamma|\mu \otimes \nu) & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise,} \end{cases} \quad (5.1.1)$$

and

$$J := \begin{cases} \int_{\mathbb{R}^d \times \mathbb{R}^d} c d\gamma & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise.} \end{cases} \quad (5.1.2)$$

**Theorem 5.1.4** ( $\Gamma$ -convergence). *Given two probability measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  with finite entropy and finite moment of order  $p > 1$  (i.e.  $\int_{\mathbb{R}^d} |x|^p d\mu < +\infty$ ) and a cost function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $c(x, y) = h(|x - y|)$  and  $|x|^p \leq h(x) \leq 1 + |x|^p$ . Then, when  $\varepsilon \rightarrow 0$ ,  $J_\varepsilon$   $\Gamma$ -converges to  $J$  with respect to the weak convergence of measures on  $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ .*

*Remark 5.1.5.* If  $\mu$  and  $\nu$  have compact support then the same result holds for any  $c$  continuous and nonnegative.

Let  $\eta \ll \mu \otimes \nu$  and consider the problem

$$\inf_{\Pi(\mu, \nu)} H(\gamma|\eta). \quad (\text{S})$$

This is a generalized version (see [100]) of the so-called *static Schrödinger problem* (see [90]), where *static* refers to the fact that one is interested only in the initial and final configuration  $\mu$  and  $\nu$ .

We state here an important result of existence and uniqueness of optimal solution for the problem (S). The proof of this result can be found for example in [100](Theorem 2.1) or in [90]. The interest of the following result relies also on the fact that it provides a characterization of the density with respect to  $\mu \otimes \nu$  of the optimal solution  $\gamma_*$ .

**Theorem 5.1.6.** *Let  $\eta \ll \mu \otimes \nu$ , such that  $\frac{d\eta}{d\mu \otimes \nu} > 0$   $\mu \otimes \nu$  a.e., and let us assume that there exists  $\gamma \in \Pi(\mu, \nu)$  such that<sup>1</sup>  $H(\gamma|\eta) < +\infty$ , then there exists a unique coupling  $\gamma_*$ , such that*

$$\gamma_* = \arg \min_{\Pi(\mu, \nu)} H(\gamma|\eta),$$

called the (static) Schrödinger bridge from  $\mu$  to  $\nu$ . Moreover:

- (a) *If  $\gamma_* \in \Pi(\mu, \nu)$  is the Schrödinger bridge, then there exist two functions  $\varphi \in L^1_\mu(\mathbb{R}^d)$  and  $\psi \in L^1_\nu(\mathbb{R}^d)$ , called Schrödinger potentials, such that*

$$\frac{d\gamma_*}{d\eta} = e^{\varphi(x) + \psi(y)} \quad \eta\text{-a.e.}$$

and  $\varphi$  and  $\psi$  are unique up to an additive constant<sup>2</sup>.

- (b) *Conversely, if  $\gamma_0 \in \Pi(\mu, \nu)$  and there exist two measurable functions  $\varphi, \psi : \mathbb{R}^d \rightarrow [-\infty, \infty)$  such that*

$$\frac{d\gamma_0}{d\eta} = e^{\varphi(x) + \psi(y)} \quad \eta\text{-a.e.},$$

then  $\gamma_0 = \gamma_*$  and  $\varphi \in L^1_\mu(\mathbb{R}^d)$  and  $\psi \in L^1_\nu(\mathbb{R}^d)$ .

Thanks to the convexity, the problem (S) has a dual formulation, as stated in the following theorem.

<sup>1</sup>for instance  $\log(\frac{d\mu \otimes \nu}{d\eta}) \in L^1(\mu \otimes \nu)$

<sup>2</sup>i.e. if  $\varphi'$  and  $\psi'$  are potentials, then  $\varphi' = \varphi + a$   $\mu$  a.e. and  $\psi' = \psi - a$   $\nu$  a.e., for some  $a \in \mathbb{R}$

**Theorem 5.1.7** (Duality). *Under the same assumptions on  $\eta$  of Theorem 5.1.6, let us consider the problem*

$$\sup_{\varphi \in L^1_\mu, \psi \in L^1_\nu} \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu + \int_{\mathbb{R}^d} \psi(y) d\nu - \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{\varphi(x) + \psi(y)} d\eta + 1 \right\}. \quad (\text{D})$$

Then (S)=(D), Moreover the supremum is attained by the Schrödinger potentials  $(\varphi_*, \psi_*)$  and the maximizers are unique up to an additive constant. Therefore if  $\gamma_*$  is the Schrödinger bridge, the (D) becomes

$$H(\gamma_* | \eta) = \int_{\mathbb{R}^d} \varphi_*(x) d\mu + \int_{\mathbb{R}^d} \psi_*(y) d\nu.$$

*Remark 5.1.8.* Let  $\eta = \alpha_\varepsilon^{-1} e^{-\frac{c}{\varepsilon}} \mu \otimes \nu$ , where  $\alpha_\varepsilon := \int e^{-\frac{c}{\varepsilon}} d\mu \otimes \nu$ , then

$$\begin{aligned} H(\gamma | \mu \otimes \nu) &= \int \log \left( \frac{d\gamma}{d\eta} \frac{d\eta}{d\mu \otimes \nu} \alpha_\varepsilon^{-1} \right) d\gamma \\ &= H(\gamma | \eta) - \int \frac{c}{\varepsilon} d\gamma - \log(\alpha_\varepsilon) \end{aligned}$$

and thus

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} cd\gamma + \varepsilon H(\gamma | \mu \otimes \nu) = \varepsilon H(\gamma | \eta) - \varepsilon \log \alpha_\varepsilon, \quad (5.1.3)$$

that means that for this choice of  $\eta$  and for every fixed parameter  $\varepsilon$ , the ( $\varepsilon$ EOT) problem is equivalent to the static Schrödinger problem (S). Therefore if (for instance)  $c \in L^1(\mu \otimes \nu)$  we can apply Theorem 5.1.6 and 5.1.7, obtaining the following result.

**Theorem 5.1.9.** *Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \in L^1(\mu \otimes \nu)$  and  $\varepsilon > 0$ , then there exists a unique coupling  $\gamma_\varepsilon \in \Pi(\mu, \nu)$ , such that*

$$\gamma_\varepsilon = \arg \min_{\Pi(\mu, \nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} cd\gamma + \varepsilon H(\gamma | \mu \otimes \nu) \right\}$$

and

$$\frac{d\gamma}{d\mu \otimes \nu} = e^{\varphi_\varepsilon(x) + \psi_\varepsilon(y) - \frac{c(x,y)}{\varepsilon}},$$

where  $\varphi_\varepsilon = \varepsilon(\varphi_* - \frac{\log \alpha_\varepsilon}{2})$  and  $\psi_\varepsilon = \varepsilon(\psi_* - \frac{\log \alpha_\varepsilon}{2})$ , and  $\varphi_*, \psi_*$  are the Schrödinger potentials defined in Theorem 5.1.6, with  $\eta = \alpha_\varepsilon^{-1} e^{-\frac{c}{\varepsilon}} \mu \otimes \nu$ .

Moreover

$$\begin{aligned} &\max_{\varphi \in L^1_\mu, \psi \in L^1_\nu} \int_{\mathbb{R}^d} \varphi(x) d\mu + \int_{\mathbb{R}^d} \psi(y) d\nu - \varepsilon \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{\frac{\varphi(x) + \psi(y) - c(x,y)}{\varepsilon}} d\mu \otimes \nu + 1 \\ &= \int_{\mathbb{R}^d} \varphi_\varepsilon(x) d\mu + \int_{\mathbb{R}^d} \psi_\varepsilon(y) d\nu = (\varepsilon\text{EOT}). \end{aligned}$$

### 5.1.2 Discrete Entropic Optimal Transport and Sinkhorn Algorithm

For every fixed  $\varepsilon > 0$ , the discrete version of the entropic optimal transport problem at scale  $\varepsilon$  will be

$$\min_{\gamma \in \Pi(\mu, \nu)} \left\{ \sum_{i,j} \gamma_{i,j} c_{i,j} + \varepsilon H(\gamma | \mu \otimes \nu) \right\}, \quad (\varepsilon\text{DEOT})$$

where the discrete relative entropy of  $\gamma$  with respect to  $\mu \otimes \nu = (\mu_i \nu_j)_{i,j}$  becomes (assuming without loss of generality that  $\mu_i, \nu_j \neq 0$  for any  $i, j$ )

$$H(\gamma | \mu \otimes \nu) = \sum_{i,j} \gamma_{i,j} \log \left( \frac{\gamma_{i,j}}{\mu_i \nu_j} \right).$$

According to Remark 5.1.8 (see (5.1.3)), if  $\eta_{i,j} = e^{-\frac{c_{i,j}}{\varepsilon}} \mu_i \nu_j$ , the problem  $(\varepsilon\text{DEOT})$  is equivalent to

$$\min_{\gamma \in \Pi(\mu, \nu)} \varepsilon H(\gamma | \eta) - \varepsilon \log a_\varepsilon, \quad (\text{DS})$$

with  $a_\varepsilon = \sum e^{-\frac{c_{i,j}}{\varepsilon}} \mu_i \nu_j$ .

We observe that

$$\begin{aligned} H(\gamma | \eta) &= \sum_{i,j} \gamma_{i,j} \log \left( \frac{\gamma_{i,j}}{e^{-\frac{c_{i,j}}{\varepsilon}} \mu_i \nu_j} \right) \\ &= \sum_{i,j} \gamma_{i,j} \log \left( \frac{\gamma_{i,j}}{e^{-\frac{c_{i,j}}{\varepsilon}}} \right) - \sum_i \mu_i \log(\mu_i) - \sum_j \nu_j \log(\nu_j), \end{aligned}$$

where  $\sum_i \mu_i \log(\mu_i)$  and  $\sum_j \nu_j \log(\nu_j)$  are given constants. So we have that the problem of finding a minimizer for  $(\varepsilon\text{DEOT})$  (and thus for (DS)) is equivalent to solving

$$\arg \min \text{KL}(\gamma | K) := \arg \min_{\Pi(\mu, \nu)} \sum_{i,j} \gamma_{i,j} \log \left( \frac{\gamma_{i,j}}{K_{i,j}} \right) - \gamma_{i,j} + K_{i,j}$$

where  $K_{i,j} = e^{-\frac{c_{i,j}}{\varepsilon}}$  is known as *Gibbs kernel* associated to the cost matrix  $(c_{i,j})_{i,j}$  and  $\text{KL}(\gamma | K)$  is the Kullback–Leibler divergence between couplings. In other words, the unique solution  $\gamma_\varepsilon$  of the  $(\varepsilon\text{DEOT})$  is the projection with respect to the K-L divergence of the Gibbs kernel into  $\Pi(\mu, \nu)$ ,

$$\gamma_\varepsilon := \arg \min_{\Pi(\mu, \nu)} \text{P}_{\Pi(\mu, \nu)}^{\text{KL}}(K).$$

### The Iterative Fitting Procedure or Sinkhorn Algorithm - Primal version

In the discrete case we have also another similar version of Theorem 5.1.9 (see Chapter 4 in [58]): if  $\gamma_\varepsilon$  is the solution of  $(\varepsilon\text{DEOT})$ , there exists  $(f^\varepsilon, g^\varepsilon) \in \mathbb{R}^n \times \mathbb{R}^m$  such that

$$(\gamma_\varepsilon)_{i,j} = e^{f_i^\varepsilon} K_{i,j} e^{g_j^\varepsilon} = u_i^\varepsilon K_{i,j} v_j^\varepsilon, \quad (5.1.4)$$

with  $(u^\varepsilon, v^\varepsilon) = (e^{f^\varepsilon}, e^{g^\varepsilon})$ . One can see that  $f^\varepsilon = \varphi_\varepsilon + \varepsilon \log \mu$  and  $g^\varepsilon = \psi_\varepsilon + \varepsilon \log \nu$ , where  $\varphi_\varepsilon$  and  $\psi_\varepsilon$  are the potentials (unique up to an additive constant) defined by Theorem 5.1.9. Since (5.1.4) means that  $\gamma_\varepsilon = \text{diag}(u^\varepsilon) K \text{diag}(v^\varepsilon) \in \Pi(\mu, \nu)$ , we have that  $(u^\varepsilon, v^\varepsilon)$  must satisfies

$$\text{diag}(u^\varepsilon) K \text{diag}(v^\varepsilon) \mathbb{1}_m = u^\varepsilon \odot K v^\varepsilon = \mu \quad (5.1.5)$$

$$\text{diag}(v^\varepsilon) K^\top \text{diag}(u^\varepsilon) \mathbb{1}_n = v^\varepsilon \odot K^\top u^\varepsilon = \nu, \quad (5.1.6)$$

where  $\mathbb{1}_n$  is the vector of size  $n$  with all entries equal to 1,  $\text{diag}(u)$  is the diagonal matrix with diagonal equal the vector  $u$  and the symbol  $\odot$  corresponds to entrywise multiplication

of vectors. The algorithm which solves the equation (5.1.5), (5.1.6) has been firstly introduced by Sinkhorn in [114] and it is also known as Iterative Proportional Fitting Procedure (IPFP), because it proceeds iteratively, by modifying first  $u$  so that it satisfies Equation (5.1.5) and then  $v$  to satisfy (5.1.6).

### Sinkhorn algorithm - primal version

Initialize  $v^{(0)} := \mathbb{1}_m$ , then the algorithm is the following

$$u^{(l)} := \frac{\mu}{Kv^{(l)}} \quad \text{and} \quad \gamma^{(2l)} := \text{diag}(u^{(l)})K\text{diag}(v^{(l)}) \quad (5.1.7)$$

$$v^{(l+1)} := \frac{\nu}{K\gamma u^{(l)}} \quad \text{and} \quad \gamma^{(2l+1)} := \text{diag}(u^{(l)})K\text{diag}(v^{(l+1)}) \quad (5.1.8)$$

where the division is to be intended entrywise. We remark that a different initialization leads to different  $(u, v)$  which are only defined up to a multiplicative constant, but the solution converges in any case to  $\text{diag}(u)K\text{diag}(v)$ , as it has been showed by Franklin and Lorenz in [68] who proved, using the Hilbert projective metric on the space of positive vectors of  $\mathbb{R}^n$ , the linear convergence of the Sinkhorn's iterations.

### Sinkhorn algorithm - dual version

There exists also a dual - equivalent to the primal - version of the Sinkhorn algorithm. Recalling Theorem 5.1.9 the dual formulation of the entropic optimal transport in the discrete case looks

$$\max_{(f,g) \in \mathbb{R}^n \times \mathbb{R}^m} \left\{ \sum_i f_i \mu_i + \sum_j g_j \nu_j - \varepsilon \sum_{i,j} e^{\frac{f_i}{\varepsilon}} e^{-\frac{c_{i,j}}{\varepsilon}} e^{\frac{g_j}{\varepsilon}} \mu_i \nu_j \right\} \quad (\varepsilon\text{DD})$$

and of course  $(\varepsilon\text{DEOT})=(\varepsilon\text{DD})$ . We observe that

$$\sum_i f_i \mu_i + \sum_j g_j \nu_j - \varepsilon \sum_{i,j} e^{\frac{f_i}{\varepsilon}} e^{-\frac{c_{i,j}}{\varepsilon}} e^{\frac{g_j}{\varepsilon}} \mu_i \nu_j$$

is equal to

$$\begin{aligned} & \sum_i f_i \mu_i + \sum_j g_j \nu_j - \varepsilon \sum_{i,j} e^{\frac{f_i}{\varepsilon}} e^{-\frac{c_{i,j}}{\varepsilon}} e^{\frac{g_j}{\varepsilon}} e^{\log(\mu_i) + \log(\nu_j)} \\ &= \sum_i (f_i + \log(\mu_i)) \mu_i + \sum_j (g_j + \log(\nu_j)) \nu_j - \varepsilon \sum_{i,j} e^{\frac{f_i + \log(\mu_i) + g_j + \log(\nu_j) - c_{i,j}}{\varepsilon}} \\ & - \log(\mu_i) - \log(\nu_j) \\ &= \sum_i f_i \mu_i + \sum_j g_j \nu_j - \varepsilon \sum_{i,j} e^{\frac{f_i}{\varepsilon}} e^{-\frac{c_{i,j}}{\varepsilon}} e^{\frac{g_j}{\varepsilon}} - \log(\mu_i) - \log(\nu_j). \end{aligned}$$

Therefore

$$(\varepsilon\text{DD}) = \max_{(f,g) \in \mathbb{R}^n \times \mathbb{R}^m} \left\{ \sum_i f_i \mu_i + \sum_j g_j \nu_j - \varepsilon \sum_{i,j} e^{\frac{f_i}{\varepsilon}} e^{-\frac{c_{i,j}}{\varepsilon}} e^{\frac{g_j}{\varepsilon}} - \log(\mu_i) - \log(\nu_j) \right\} \quad (5.1.9)$$

In order to obtain the optimal couple  $(f, g)$  one can compute the gradients of the argument in the right-hand side of the equation above, first with respect to  $f$  and then with respect to  $g$  and impose them to be equal to 0, obtaining, for every component,

$$\begin{aligned}\mu_i - e^{\frac{f_i}{\varepsilon}} \sum_j e^{-\frac{c_{i,j}}{\varepsilon}} e^{\frac{g_j}{\varepsilon}} &= 0, \\ \nu_j - e^{\frac{g_j}{\varepsilon}} \sum_i e^{-\frac{c_{i,j}}{\varepsilon}} e^{\frac{f_i}{\varepsilon}} &= 0\end{aligned}$$

and thus

$$\begin{aligned}f &= \varepsilon \log(\mu) - \varepsilon \log(K e^{\frac{g}{\varepsilon}}), \\ g &= \varepsilon \log(\nu) - \varepsilon \log(K^\top e^{\frac{f}{\varepsilon}}),\end{aligned}$$

where  $\log$  has to be intended entry-wise.

Therefore a simple approach to solving computationally the unconstrained maximization problem ( $\varepsilon$ DD) is to use an exact block coordinate ascent strategy, namely to update alternatively  $f$  and  $g$  to cancel the respective gradients.

#### Sinkhorn algorithm - dual version

Initialize with  $g^{(0)}$ , then the algorithm will proceed as follows. For  $l \geq 0$ :

$$f^{(l+1)} = \varepsilon \log(\mu) - \varepsilon \log \left( K e^{\frac{g^{(l)}}{\varepsilon}} \right) = \varepsilon \log(\mu) - \varepsilon \log \left( \sum_j K_{i,j} e^{\frac{g_j^{(l)}}{\varepsilon}} \right), \quad (5.1.10)$$

$$g^{(l+1)} = \varepsilon \log(\nu) - \varepsilon \log \left( K^\top e^{\frac{f^{(l+1)}}{\varepsilon}} \right) = \varepsilon \log(\nu) - \varepsilon \log \left( \sum_i K_{i,j}^\top e^{\frac{f_i^{(l+1)}}{\varepsilon}} \right). \quad (5.1.11)$$

We observe that iterations (5.1.10) and (5.1.11) are mathematically equivalent to the ones of the primal version of Sinkhorn ((5.1.7),(5.1.8)). Indeed, we recover that at any iteration

$$(f^{(l)}, g^{(l)}) = \varepsilon \left( \log \left( u^{(l)} \right), \log \left( v^{(l)} \right) \right).$$

**Definition 5.1.10.** Given  $z \in \mathbb{R}^n$  the soft-minimum of  $z$  at scale  $\varepsilon > 0$  is defined by

$$\min_\varepsilon(z) := -\varepsilon \log \left( \sum_i^n e^{-\frac{z_i}{\varepsilon}} \right).$$

*Remark 5.1.11.* First of all we notice that  $\min_\varepsilon(z) \leq \min(z) := \min_{i=1,\dots,n} \{z_i\}$  and that  $\min_\varepsilon(z) \rightarrow \min(z)$  when  $\varepsilon \rightarrow 0$ . Moreover for any  $w \in \mathbb{R}$

$$\min_\varepsilon(z) = w - \varepsilon \log \left( \sum_i^n e^{-\frac{z_i - w}{\varepsilon}} \right).$$

We observe that each components of the column vectors  $f^{(l+1)}$  and  $g^{(l+1)}$  in the iterations (5.1.10) and (5.1.11) can be rewritten as

$$f_i^{(l+1)} = \min_\varepsilon \left( c_{i,j} - g_j^{(l)} \right)_j + \varepsilon \log \mu_i, \quad (5.1.12)$$

$$g_j^{(l+1)} = \min_\varepsilon \left( c_{i,j} - f_i^{(l+1)} \right)_i + \varepsilon \log \nu_j, \quad (5.1.13)$$

where the term  $\min_\varepsilon \left( c_{i,j} - g_j^{(l)} \right)_j$  denotes the soft-min of the  $i$ -th row of the matrix  $C - \mathbb{1}_m (g^{(l)})^\top$  and  $\min_\varepsilon \left( c_{i,j} - f_i^{(l+1)} \right)_i$  denotes the soft-min of the  $j$ -th column of the matrix  $C - f^{(l+1)} \mathbb{1}_n^\top$ .

A strategy to avoid underflow for small values of  $\varepsilon$  is to use the *log-sum-exp* trick in the iterations (5.1.12) and (5.1.13), which means rewriting the soft minimum as suggested in the Remark 5.1.11.

### Log-Domain Sinkhorn

Let  $G_i^{(l)} := \min_{j=1,\dots,m} \left\{ c_{i,j} - g_j^{(l)} \right\}_j$  and  $F_j^{(l)} := \min_{i=1,\dots,n} \left\{ c_{i,j} - f_i^{(l)} \right\}_i$ , and initialize  $g^{(0)}$ , then the Sinkhorn algorithm is equivalent to

$$f_i^{(l+1)} = \min_\varepsilon \left( c_{i,j} - g_j^{(l)} - G_i^{(l)} \right)_j + \varepsilon \log \mu_i + G_i^{(l)}, \quad (5.1.14)$$

$$g_j^{(l+1)} = \min_\varepsilon \left( c_{i,j} - f_i^{(l+1)} - F_j^{(l+1)} \right)_i + \varepsilon \log \nu_j + F_j^{(l)}, \quad (5.1.15)$$

In contrast to the original iterations (5.1.10) and (5.1.11), these log-domain iterations (5.1.14) and (5.1.15) are stable for arbitrary  $\varepsilon > 0$ . Indeed for any vector  $z$ ,

$$\sum_i^n e^{-\frac{z_i - \bar{z}}{\varepsilon}} \geq e^{-\frac{\bar{z} - \bar{z}}{\varepsilon}} = 1,$$

so that in the iterations (5.1.14) and (5.1.15) the argument of the logarithm is never smaller than 1.

## 5.2 $L^\infty$ -Entropic Optimal Transport

In the sequel, we will always assume that

1. the transportation cost  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  is continuous,
2. the fixed marginals  $\mu$  and  $\nu$  of the problem are two Borel probability measures on  $\mathbb{R}^d$  with compact support.

Note that every  $\gamma$  in  $\Pi(\mu, \nu)$  has its support in  $\text{supp } \mu \times \text{supp } \nu$  and that  $c$  is uniformly continuous on  $\text{supp } \mu \times \text{supp } \nu$ .

In this setting, it seems natural to introduce, for  $\varepsilon > 0$  and exponent  $p \geq 1$  the functionals  $J_{p,\varepsilon}, J_\infty : \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \rightarrow \overline{\mathbb{R}}$

$$J_{p,\varepsilon}(\gamma) := \begin{cases} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)^p d\gamma(x,y) + \varepsilon H(\gamma | \mu \otimes \nu) \right)^{\frac{1}{p}} & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise,} \end{cases}$$

and

$$J_\infty(\gamma) := \begin{cases} \|c\|_{L^\infty(\gamma)} & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise.} \end{cases}$$

Then we define

$$\gamma_{p,\varepsilon} := \arg \min_{\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)} J_{p,\varepsilon}(\gamma) = \arg \min_{\Pi(\mu,\nu)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)^p d\gamma(x,y) + \varepsilon H(\gamma|\mu \otimes \nu) \right)^{\frac{1}{p}}, \quad (5.2.1)$$

which exists and is unique by Theorem 5.1.9, and

$$\gamma_\infty \in \arg \min_{\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)} J_\infty(\gamma) = \arg \min_{\Pi(\mu,\nu)} C_\infty(\gamma) = \arg \min_{\Pi(\mu,\nu)} \|c\|_{L^\infty(\gamma)}, \quad (5.2.2)$$

whose existence is guaranteed by Theorem 4.1.21.

### 5.2.1 $\Gamma$ -convergence

In this section we will present a result regarding the  $\Gamma$ -convergence of  $J_{p,\varepsilon}$  to  $J_\infty$ .

However just letting  $p \rightarrow \infty$  and  $\varepsilon \rightarrow 0$  is not enough to ensure  $\Gamma$ -convergence. Indeed, without additional assumptions, it can happen, as shown in the Example (5.2.1) below, that minimizers of  $J_{p,\varepsilon}$  do not converge to a minimizer of  $J_\infty$  (i.e. a solution of (OT $_\infty$ )).

*Example 5.2.1.* Let the cost  $c$  such that  $c \leq \frac{1}{2}$  and  $\varepsilon = \frac{1}{p}$ . Then the minimizer  $\gamma_p$  of  $J_{p,\frac{1}{p}}$  satisfies

$$\int c^p d\gamma_p + \frac{1}{p} H(\gamma_p|\mu \otimes \nu) \leq \int c^p d\mu \otimes \nu \leq 2^{-p}$$

and thus

$$H(\gamma_p|\mu \otimes \nu) \leq p2^{-p}.$$

Hence  $\gamma_p$  converges (actually strongly by Pinsker's inequality, see e.g. Lemma 2.5 in [118]) to  $\mu \otimes \nu$  which in general is not a minimizer of  $J_\infty$ .

On the one hand, the example above suggests that  $\Gamma$ -convergence of the regularizations  $J_{p,\varepsilon}$  to  $J_\infty$  require conditions relating  $\varepsilon$  to  $p$ . On the other hand, in the previous example, we see that the range of  $c^p$  compared to the size of the entropic penalization  $\varepsilon$  is crucial. With this in mind, it is worth to remarking that solution of the problem

$$\arg \min_{\Pi(\mu,\nu)} \|c\|_{L^\infty(\gamma)}$$

are invariant when replacing  $c$  with an increasing function of  $c$ . In particular, as shown in Theorem 5.2.4 below, one can replace  $c$  by  $c + C$ , with  $C > 0$  such that  $c^p$  is large enough to dominate the entropic term and to have  $\Gamma$ -convergence as  $p \rightarrow \infty$  for a fixed (or even large) value of  $\varepsilon$ .

For the  $\Gamma$ -limsup inequality, we have used the so-called *block approximation* technique introduced in [38], which is defined as follows:

**Definition 5.2.2.** Let  $\gamma \in \Pi(\mu,\nu)$ . For  $\delta > 0$  and  $k \in \mathbb{Z}^d$ , we denote by  $Q_k^\delta$  the cube  $\delta(k + [0,1]^d)$ . The block approximation of  $\gamma$  at scale  $\delta \in (0,1)$  is then defined by

$$\gamma^\delta := \sum_{k,l \in \mathbb{Z}^d : \mu(Q_k^\delta) > 0, \nu(Q_l^\delta) > 0} \frac{\gamma(Q_k^\delta \times Q_l^\delta)}{\mu(Q_k^\delta)\nu(Q_l^\delta)} \mu_k^\delta \otimes \nu_l^\delta \quad (5.2.3)$$

where  $\mu_k^\delta$  and  $\nu_l^\delta$  are defined by

$$\mu_k^\delta(A) = \mu(Q_k^\delta \cap A), \quad \nu_l^\delta(A) = \nu(Q_l^\delta \cap A)$$

for every Borel subset  $A$  of  $\mathbb{R}^d$ .

For the sake of completeness, we provide a short proof of the properties of the block approximation we will use in the proof of Theorem 5.2.4. These properties have been first proved by Carlier *et al.* in [38] and an interesting improvement is provided in [39], where the bound on  $H(\gamma^\delta|\mu \otimes \nu)$  is given by the upper entropy (or Rény) dimension of  $\mu$  and  $\nu$  and this allows to cover the support of  $\mu \otimes \nu$  with general partitions rather than cubes and, if  $\mu$  has finite log-moment one can also drop the compactness assumption on  $\text{supp } \mu$ .

**Lemma 5.2.3.** *Let  $\gamma \in \Pi(\mu, \nu)$  and  $\gamma^\delta$  be the block approximation of  $\gamma$  at scale  $\delta \in (0, 1)$ , then  $\gamma^\delta \in \Pi(\mu, \nu)$  and*

$$W_\infty(\gamma^\delta, \gamma) \leq \sqrt{2d}\delta, \quad (5.2.4)$$

$$H(\gamma^\delta|\mu \otimes \nu) \leq d \log \left( \frac{L}{\delta} \right), \quad (5.2.5)$$

where  $L$  is a constant depending only on  $\text{supp } \mu$ .

*Proof.* The fact that  $\gamma^\delta \in \Pi(\mu, \nu)$  is easy to check by construction (see [38]). Now observe that by (5.2.3) the density of  $\gamma^\delta$  with respect to  $\mu \otimes \nu$  is

$$\frac{d\gamma^\delta}{d\mu \otimes \nu}(x, y) = \begin{cases} \frac{\gamma(Q_k^\delta \times Q_l^\delta)}{\mu(Q_k^\delta)\nu(Q_l^\delta)} & \text{if } (x, y) \in Q_k^\delta \times Q_l^\delta, \text{ and } \mu(Q_k^\delta), \nu(Q_l^\delta) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

$$\begin{aligned} H(\gamma^\delta|\mu \otimes \nu) &= \sum_{k, l \in \mathbb{Z}^d : \mu(Q_k^\delta) > 0, \nu(Q_l^\delta) > 0} \int_{Q_k^\delta \times Q_l^\delta} \log \left( \frac{\gamma(Q_k^\delta \times Q_l^\delta)}{\mu(Q_k^\delta)\nu(Q_l^\delta)} \right) d\gamma^\delta \\ &\leq \sum_{k, l \in \mathbb{Z}^d : \mu(Q_k^\delta) > 0, \nu(Q_l^\delta) > 0} \int_{Q_k^\delta \times Q_l^\delta} \log \left( \frac{1}{\mu(Q_k^\delta)} \right) d\gamma^\delta \\ &= \sum_{k \in \mathbb{Z}^d : \mu(Q_k^\delta) > 0} \mu(Q_k^\delta) \log \left( \frac{1}{\mu(Q_k^\delta)} \right), \end{aligned}$$

where the inequality is due to the fact that  $\frac{\gamma(Q_k^\delta \times Q_l^\delta)}{\nu(Q_l^\delta)} \leq 1$ , while the last equality is obtained summing over  $l$ . If  $L \geq 1$  is such that  $\text{supp } \mu$  is contained in a cube of side  $L - 1$ , the number of cubes  $Q_k^\delta$  with positive  $\mu$ -measure is not greater than  $N_\delta := \left(\frac{L}{\delta}\right)^d$ . Therefore, applying Jensen's inequality to the concave function  $f(z) = z \log\left(\frac{1}{z}\right)$ , we have

$$\begin{aligned} H(\gamma^\delta|\mu \otimes \nu) &\leq \sum_{k=1}^{N_\delta} \mu(Q_k^\delta) \log \left( \frac{1}{\mu(Q_k^\delta)} \right) \\ &\leq N_\delta \left( \left( \frac{1}{N_\delta} \sum_{k=1}^{N_\delta} \mu(Q_k^\delta) \right) \log \left( \frac{1}{\sum_{k=1}^{N_\delta} \frac{1}{N_\delta} \mu(Q_k^\delta)} \right) \right) \\ &= \log(N_\delta) = d \log(L) - d \log(\delta), \end{aligned}$$

which proves (5.2.5).

By construction  $\gamma(Q_k^\delta \times Q_l^\delta) = \gamma^\delta(Q_k^\delta \times Q_l^\delta)$ , for any  $k, l$ . Let  $J$  be the set of pairs of

indices  $(k, l)$  such that  $\gamma^\delta(Q_k^\delta \times Q_l^\delta) > 0$  and set  $\bar{Q}_j = Q_k^\delta \times Q_l^\delta$ , for any  $j = (k, l) \in J$ . We define

$$\eta^\delta := \sum_j \gamma(\bar{Q}_j) \gamma_j \otimes \gamma_j^\delta,$$

where  $\gamma_j(A) := \frac{\gamma(A \cap \bar{Q}_j)}{\gamma(\bar{Q}_j)}$  and  $\gamma_j^\delta(A) := \frac{\gamma^\delta(A \cap \bar{Q}_j)}{\gamma^\delta(\bar{Q}_j)}$ . By construction  $\eta^\delta \in \Pi(\gamma, \gamma^\delta)$ , thus

$$W_\infty(\gamma, \gamma^\delta) \leq \|x - y\|_{L^\infty(\eta^\delta)} \leq \text{diam}(\bar{Q}_j) = \sqrt{2d}\delta.$$

□

**Theorem 5.2.4.** *Under the general assumptions (1) and (2), we have:*

1.  $J_{p, \varepsilon_p}$   $\Gamma$ -converges (for the weak topology<sup>3</sup> on  $\mathcal{P}(\text{supp } \mu \times \text{supp } \nu)$ ) to  $J_\infty$  as  $p \rightarrow \infty$  provided  $\varepsilon_p^{\frac{1}{p}} \rightarrow 0$  as  $p \rightarrow \infty$ ,
2. if in addition,  $c \geq 1 + \lambda$  with  $\lambda \geq 0$ , then  $J_{p, \varepsilon_p}$   $\Gamma$ -converges to  $J_\infty$  as  $p \rightarrow \infty$  provided

$$\lim_{p \rightarrow \infty} \frac{1}{p} \log \left( 1 + \varepsilon_p \frac{\log(p)}{(1 + \lambda)^p} \right) = 0. \quad (5.2.6)$$

In particular  $J_{p,1}$  and  $J_{p,p}$   $\Gamma$ -converge to  $J_\infty$  as  $p \rightarrow \infty$ .

*Proof.* 1. Let  $\gamma_p \in \Pi(\mu, \nu)$  converge weakly star to  $\gamma$ . By nonnegativity of  $H(\gamma_p | \mu \otimes \nu)$ , we have

$$\liminf_p J_{p, \varepsilon_p}(\gamma_p) \geq \liminf_p \|c\|_{L^p(\gamma_p)}$$

hence, for fixed  $q$  since  $\|c\|_{L^p(\gamma_p)} \geq \|c\|_{L^q(\gamma_p)}$  for  $p \geq q$ , we have

$$\liminf_p J_{p, \varepsilon_p}(\gamma_p) \geq \liminf_p \|c\|_{L^q(\gamma_p)} = \|c\|_{L^q(\gamma)}$$

taking the supremum with respect to  $q$  thus yields the desired  $\Gamma$ -liminf inequality

$$\liminf_p J_{p, \varepsilon_p}(\gamma_p) \geq \|c\|_{L^\infty(\gamma)} = J_\infty(\gamma).$$

Let us now prove the  $\Gamma$ -limsup inequality. For any  $\gamma \in \Pi(\mu, \nu)$  we consider  $\gamma^\delta$ , the block approximation of  $\gamma$  at scale  $\delta \in (0, 1)$  defined by (5.2.3) below, whose convergence to  $\gamma$  is insured by (5.2.4)<sup>4</sup>. By concavity, for  $p \geq 1$ ,

$$\begin{aligned} J_{p, \varepsilon_p}(\gamma^\delta) &\leq \|c\|_{L^p(\gamma^\delta)} + \varepsilon_p^{\frac{1}{p}} H(\gamma^\delta | \mu \otimes \nu)^{\frac{1}{p}} \\ &\leq \|c\|_{L^\infty(\gamma^\delta)} + \varepsilon_p^{\frac{1}{p}} H(\gamma^\delta | \mu \otimes \nu)^{\frac{1}{p}}. \end{aligned}$$

Denoting by  $\omega$  a modulus of continuity of  $c$  on  $\text{supp } \mu \times \text{supp } \nu$ , thanks to by (5.2.4), we have

$$\|c\|_{L^\infty(\gamma^\delta)} \leq \|c\|_{L^\infty(\gamma)} + \omega(\sqrt{2d}\delta),$$

<sup>3</sup>Note that, since  $\text{supp } \mu \times \text{supp } \nu$  is compact, the weak topology defined in Definition 4.1.1 is equivalent to what is usually called in literature *weak\* topology* of measures, i.e. convergence tested against  $C_c$  functions, which, in case of a compact space, are all the continuous functions.

<sup>4</sup> $W_\infty(\gamma, \gamma^\delta) \geq W_2(\gamma, \gamma^\delta)$  and, as pointed out in Definition 4.1.12,  $W_2(\gamma, \gamma^\delta) \rightarrow 0$  implies  $\gamma^\delta \rightarrow \gamma$  (see also Theorem 8.8 in [3])

being  $\sqrt{2d}\delta$  the diameter of the cubes of the approximation. Moreover, by (5.2.5), we have

$$H(\gamma^\delta | \mu \otimes \nu)^{\frac{1}{p}} \leq d^{\frac{1}{p}} \log(L/\delta)^{\frac{1}{p}}$$

so if we define  $\gamma_p$  as the block approximation of  $\gamma$  at scale (say)  $\delta = \frac{1}{p}$ , we obtain

$$\limsup_p J_{p,\varepsilon_p}(\gamma_p) \leq J_\infty(\gamma) + \limsup_p \left( \omega\left(\frac{\sqrt{2d}}{p}\right) + d^{\frac{1}{p}} \varepsilon_p^{\frac{1}{p}} \log(Lp)^{\frac{1}{p}} \right) = J_\infty(\gamma),$$

since we have assumed that  $\varepsilon_p^{\frac{1}{p}} \rightarrow 0$  as  $p \rightarrow +\infty$ .

2. Let us now assume that  $c \geq 1 + \lambda$ , the proof of the  $\Gamma$ -liminf inequality for  $J_{p,\varepsilon_p}$  is exactly as above. For  $\gamma \in \Pi(\mu, \nu)$  and  $\gamma_p$  the block approximation of  $\gamma$  at scale  $\frac{1}{p}$ , we have

$$\begin{aligned} J_{p,\varepsilon_p}(\gamma_p) &\leq \|c\|_{L^\infty(\gamma_p)} \left( 1 + \frac{d\varepsilon_p \log(Lp)}{(1+\lambda)^p} \right)^{\frac{1}{p}} \\ &\leq \left( J_\infty(\gamma) + \omega\left(\frac{\sqrt{2d}}{p}\right) \right) \left( 1 + \frac{d\varepsilon_p \log(Lp)}{(1+\lambda)^p} \right)^{\frac{1}{p}} \end{aligned} \quad (5.2.7)$$

so that, as soon as (5.2.6) holds, one has

$$\limsup_p J_{p,\varepsilon_p}(\gamma_p) \leq J_\infty(\gamma).$$

□

*Remark 5.2.5.* Notice that in case  $c \geq 1 + \lambda$  for some  $\lambda > 0$ ,  $\Gamma$ -convergence of  $J_{p,\varepsilon_p}$  to  $J_\infty$  is guaranteed even for fastly increasing  $\varepsilon_p$  like  $\varepsilon_p = p^m (1 + \lambda)^p$  with  $m \geq 0$ . On the contrary, in the general case, the condition  $\varepsilon_p^{\frac{1}{p}} \rightarrow 0$  requires to choose values of  $\varepsilon$  way too small to be used in practice for numerical computations. This suggests in practice to rescale the cost so that it is bounded from below by 1.

*Remark 5.2.6.* We observe that in (5.2.7) it is sufficient that  $\|c\|_{L^\infty(\gamma_p)} \geq 1 + \lambda$ , for some  $\lambda \geq 0$ , therefore the conclusion of case 2. in Theorem 5.2.4 remains valid under the weaker assumption that  $v_\infty = \min_{\Pi(\mu, \nu)} J_\infty \geq 1 + \lambda$ .

### 5.2.2 Selection of plans with $\infty$ -cyclically monotone support

In Section 4.1.2, Definition 4.1.24, we have defined  $\infty$ - $c$ -cyclically monotone plans ( $\infty$ -cm), which are special solutions of  $(OT_\infty)$ , indeed in addition to be characterized by restrictability (Theorem 4.1.34), under suitable assumptions on  $\mu$  and the cost function  $c$  (see Theorem 4.1.37 and Theorem 4.2.10), they are also solution of the  $L^\infty$  Monge problem, i.e. there exists a Borel map  $T$  such that  $\gamma_\infty = (Id \times T)_\# \mu$ .

**Notation.** As pointed out in Section 4.1.2, when the cost function is clear from the context we will simply write  $\infty$ -cyclically monotone plans, omitting the dependence on  $c$ . We also recall Remark 4.1.26, for the invariance of this notion under increasing functions of  $c$ .

Thanks to the  $\Gamma$ -convergence result (Theorem 5.2.4) proved in the previous section, we have convergence up to subsequences of the minimizers  $\gamma_{p,\varepsilon}$  of  $J_{p,\varepsilon}$ . Indeed,  $J_{p,\varepsilon}$  are equi-coercive since all the pre images are contained in  $\Pi(\mu, \nu)$ , which is compact by Prokhorov's theorem. In particular it is sequentially compact, as  $\mathcal{P}(\mathbb{R}^d)$  is a Polish space (see Remark 5.1.1 in [4]). The goal of this section is to show that entropic approximation selects exactly  $\infty$ - $c$ -cyclically monotone plans.

We shall make use of the analysis of the landmark recent article [22]. Let us first recall the notion of  $(c, \varepsilon)$ -cyclically invariance introduced in [22]:

**Definition 5.2.7.** Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow (0, \infty)$  be a measurable function. A coupling  $\gamma \in \Pi(\mu, \nu)$  is called  $(c, \varepsilon)$ -cyclically invariant if  $\gamma \ll \mu \otimes \nu$  and its density admits a representative  $\frac{d\gamma}{d\mu \otimes \nu} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow (0, \infty)$  such that

$$\prod_{i=1}^k \frac{d\gamma}{d\mu \otimes \nu}(x_i, y_i) = \exp\left(-\frac{1}{\varepsilon} \left[ \sum_{i=1}^k (c(x_i, y_i) - c(x_i, y_{i+1})) \right]\right) \prod_{i=1}^k \frac{d\gamma}{d\mu \otimes \nu}(x_i, y_{i+1}),$$

for all  $k \in \mathbb{N}^*$  and  $\{(x_i, y_i)\}_{i=1}^k \subset \mathbb{R}^d \times \mathbb{R}^d$ , where  $y_{k+1} = y_1$ .

In [22] (Proposition 2.2), it is shown that whenever

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c \, d\gamma + \varepsilon H(\gamma | \mu \otimes \nu) \quad (5.2.8)$$

is finite, the (unique) solution  $\gamma_\varepsilon$  of (5.2.8) is characterized by being  $(c, \varepsilon)$ -cyclically invariant. The next lemma, which is a part of Lemma 3.1 in [22], provides an estimate for  $(c, \varepsilon)$ -cyclically invariant couplings, which will be useful for our purpose. For the reader's convenience we provide also here the proof.

**Lemma 5.2.8.** *Let  $\varepsilon > 0$  and  $\gamma_\varepsilon \in \Pi(\mu, \nu)$  be  $(c, \varepsilon)$ -cyclical invariant. For every fixed  $k \geq 2$ ,  $k \in \mathbb{N}$ , and  $\delta \geq 0$ , let  $A_{k,c}(\delta)$  be the set defined by*

$$A_{k,c}(\delta) := \left\{ (x_i, y_i)_{i=1}^k \in (\mathbb{R}^d \times \mathbb{R}^d)^k : \sum_{i=1}^k c(x_i, y_i) - \sum_{i=1}^k c(x_i, y_{i+1}) \geq \delta \right\} \quad (5.2.9)$$

where  $y_{k+1} = y_1$ . Let  $A \subset A_{k,c}(\delta)$  be Borel. Then  $\gamma_\varepsilon^k := \prod_{i=1}^k \gamma_\varepsilon(dx_i, dy_i)$  satisfies

$$\gamma_\varepsilon^k(A) \leq e^{-\frac{\delta}{\varepsilon}}.$$

*Proof.* By Definition 5.2.7 of  $(c, \varepsilon)$ -cyclical invariance, for  $\gamma_\varepsilon^k$  a.e.  $(x_i, y_i)_{i=1}^k \in A$  we have that

$$\prod_{i=1}^k \frac{d\gamma_\varepsilon}{d\mu \otimes \nu}(x_i, y_i) \leq e^{-\frac{\delta}{\varepsilon}} \prod_{i=1}^k \frac{d\gamma_\varepsilon}{d\mu \otimes \nu}(x_i, y_{i+1}).$$

In one defines the set  $\bar{A} := \{(x_i, y_{i+1})_{i=1}^k : (x_i, y_i) \in A\}$ , by integrating over  $A$  with respect to  $\gamma_\varepsilon^k = \prod \gamma_\varepsilon(x_i, y_i) = \prod \gamma_\varepsilon(x_i, y_{i+1})$  we obtain

$$\gamma_\varepsilon^k(A) \leq e^{-\frac{\delta}{\varepsilon}} \gamma_\varepsilon^k(\bar{A}) \leq e^{-\frac{\delta}{\varepsilon}}.$$

□

The fact that the entropic approximation procedure selects  $\infty$ -cyclically monotone plans is then ensured by the following:

**Theorem 5.2.9.** *Let us assume that  $c > 0$ , and let  $\gamma_{p,\varepsilon_p}$  be the minimizer of  $J_{p,\varepsilon_p}$ . Then, as  $p \rightarrow \infty$ , any cluster point  $\gamma_\infty$  of the family  $(\gamma_{p,\varepsilon_p})$  is  $\infty$ -cyclically monotone, provided*

1.  $\varepsilon_p^{\frac{1}{p}} \rightarrow 0$  as  $p \rightarrow \infty$ ,
2.  $\varepsilon_p = o(p(1+\lambda)^p)$  if, in addition,  $c \geq 1 + \lambda$  with  $\lambda \geq 0$ .

*Proof.* Up to extracting a subsequence, let us assume that  $\gamma_{p,\varepsilon_p}$  weakly star converges to  $\gamma_\infty$ . We proceed by contradiction assuming that there exists  $\delta > 0$  and a finite sequence of points  $(x_i, y_i)_{i=1}^k$  contained in  $\text{supp } \gamma_\infty$ , such that

$$\max_{i=1,\dots,k} \{c(x_i, y_i)\} > \max_{i=1,\dots,k} \{c(x_i, y_{i+1})\} + \delta.$$

By the continuity of the cost function  $c$  and by the uniform convergence of  $\left(\sum_{i=1}^k c(x'_i, y'_i)^p\right)^{\frac{1}{p}}$  to  $\max_{i=1,\dots,k} \{c(x'_i, y'_i)\}$ , as  $p \rightarrow +\infty$ , we deduce that for every  $i = 1, \dots, k$  there exists an open neighborhood  $U_i$  of  $(x_i, y_i)$  and  $p(\delta) > 0$ , such that

$$\left(\sum_{i=1}^k c(x'_i, y'_i)^p\right)^{\frac{1}{p}} > \left(\sum_{i=1}^k c(x'_i, y'_{i+1})^p\right)^{\frac{1}{p}} + \delta,$$

for every  $(x'_i, y'_i) \in U_i$  (again with the convention that  $y'_{k+1} = y'_1$ ), for every  $i = 1, \dots, k$  and  $p \geq p(\delta)$ . We now observe that

$$\begin{aligned} \sum_{i=1}^k c(x'_i, y'_i)^p &> \left( \left( \sum_{i=1}^k c(x'_i, y'_{i+1})^p \right)^{\frac{1}{p}} + \delta \right)^p \\ &\geq \sum_{i=1}^k c(x'_i, y'_{i+1})^p + p \left( \sum_{i=1}^k c(x'_i, y'_{i+1})^p \right)^{\frac{p-1}{p}} \delta, \end{aligned} \quad (5.2.10)$$

where the last inequality follows from the convexity of  $t \mapsto t^p$ , with  $p > 1$ . Since  $c > 0$  and since, without loss of generality, we can assume  $U_i$  to be bounded for every  $i = 1, \dots, k$ , there exists some  $b > 0$  such that  $c \geq b$  on each  $U_i$ ,  $i = 1, \dots, k$ , hence, for every  $(x'_i, y'_i) \in U_i$  and  $p \geq p(\delta)$

$$\sum_{i=1}^k c(x'_i, y'_i)^p > \sum_{i=1}^k c(x'_i, y'_{i+1})^p + p\delta b^{p-1}. \quad (5.2.11)$$

We thus have  $U_1 \times \dots \times U_k \subset A_{k,c^p}(p\delta b^{p-1})$ , where  $A_{k,c^p}(p\delta b^{p-1})$  is defined as in (5.2.9) with  $c$  replaced by  $c^p$ . Applying Lemma 5.2.8, we thus get:

$$\begin{aligned} \gamma_\infty^k(U_1 \times \dots \times U_k) &:= \prod_{i=1}^k \gamma_\infty(U_i) \\ &\leq \liminf_p \gamma_{p,\varepsilon_p}^k(U_1 \times \dots \times U_k) := \prod_{i=1}^k \gamma_{p,\varepsilon_p}(U_i) \\ &\leq \liminf_p e^{-\frac{p\delta b^{p-1}}{\varepsilon_p}} \end{aligned} \quad (5.2.12)$$

so that if  $\varepsilon_p^{\frac{1}{p}} \rightarrow 0$  as  $p \rightarrow \infty$ , for large enough  $p$  one has  $\varepsilon_p \leq b^p$ , which yields

$$\liminf_p e^{-\frac{p\delta b^{p-1}}{\varepsilon_p}} = 0.$$

On the other hand, since the points  $(x_i, y_i)$  belong to  $\text{supp } \gamma_\infty$ , we have that  $\gamma_\infty^k(U_1 \times \cdots \times U_k) > 0$ , which yields the desired contradiction. This shows the first assertion.

Now, if  $c \geq (1 + \lambda)$  with  $\lambda \geq 0$ , we can replace  $b$  by  $(1 + \lambda)$  in (5.2.12) and the same conclusion will be reached as soon as  $\varepsilon_p = o(p(1 + \lambda)^p)$ , proving the second assertion.  $\square$

*Remark 5.2.10.* Despite what we observed in Remark 5.2.6 regarding Theorem 5.2.4, in the proof of the second assertion of Theorem 5.2.9, it does not seem that the condition  $c(x, y) \geq 1$  for every  $(x, y)$  can be weakened to  $J_\infty \geq 1$ . Note also that the condition  $\varepsilon_p = o(p(1 + \lambda)^p)$  is stronger than condition (5.2.6) that guarantees  $\Gamma$ -convergence when  $c \geq 1 + \lambda$ , for some  $\lambda \geq 0$ .

### 5.2.3 Some estimates on the speed of convergence

Our aim in this section is to give some error estimates for  $v_p - v_\infty$  where

$$v_p := \min_{\gamma \in \Pi(\mu, \nu)} J_p \quad \text{and} \quad v_\infty := \min_{\gamma \in \Pi(\mu, \nu)} J_\infty. \quad (5.2.13)$$

and we recall that  $J_p = J_{p,1}$  (i.e. for the sake of simplicity we take  $\varepsilon_p = 1$  as entropic penalization parameter).

#### Upper bounds

**Proposition 5.2.11** (Upper bounds). *Let  $c \in C^{0,\alpha}(\mathbb{R}^d \times \mathbb{R}^d)$ , with  $\alpha \in (0, 1]$  and let us assume that  $v_\infty \geq 1 + \lambda$  for some  $\lambda \geq 0$ . Then we have<sup>5</sup>*

$$v_p - v_\infty \leq \begin{cases} O(e^{-\beta p}), & \text{with } \beta = \min\{\alpha, \log(1 + \lambda)\} & \text{if } \lambda > 0 \\ O\left(\frac{\log(\log p)}{p}\right) & & \text{if } \lambda = 0. \end{cases}$$

*Proof.* Let  $\gamma_\infty$  be a minimizer of  $J_\infty$  and  $\gamma^\delta$  be the block approximation of  $\gamma_\infty$  at scale  $\delta \in (0, 1)$ , as defined in (5.2.3). We observe that, by construction and by the Hölder condition on  $c$ , denoting by  $A$  the  $C^{0,\alpha}$  semi-norm of  $c$ , we first have

$$\|c\|_{L^\infty(\gamma^\delta)} \leq \|c\|_{L^\infty(\gamma_\infty)} + A\delta^\alpha.$$

Then

$$\begin{aligned} v_p &\leq \left( \int c^p d\gamma^\delta + H(\gamma^\delta | \mu \otimes \nu) \right)^{\frac{1}{p}} \leq \left( \|c\|_{L^\infty(\gamma^\delta)}^p + H(\gamma^\delta | \mu \otimes \nu) \right)^{\frac{1}{p}} \\ &\leq \left( \|c\|_{L^\infty(\gamma_\infty)} + A\delta^\alpha \right) \left( 1 + \frac{H(\gamma^\delta | \mu \otimes \nu)}{(1 + \lambda)^p} \right)^{\frac{1}{p}} \\ &\leq (v_\infty + A\delta^\alpha) \left( 1 + \frac{d \log(L/\delta)}{(1 + \lambda)^p} \right)^{\frac{1}{p}}, \end{aligned} \quad (5.2.14)$$

<sup>5</sup>We recall that  $v_p - v_\infty$  might be either greater or smaller than 0. Indeed the  $p$ -norm is smaller than the  $\infty$ -norm, but the entropy term is nonnegative.

where the last inequality follows from Lemma 5.2.3. For  $\lambda > 0$ , choosing  $\delta := e^{-p}$ , (5.2.14) becomes (setting  $C = d \log(L)$ )

$$v_p \leq (v_\infty + Ae^{-\alpha p}) \left(1 + \frac{C + dp}{(1 + \lambda)^p}\right)^{\frac{1}{p}},$$

then for large  $p$ , one has

$$\left(1 + \frac{C + dp}{(1 + \lambda)^p}\right)^{\frac{1}{p}} = 1 + \frac{d}{(1 + \lambda)^p} + o\left(\frac{1}{(1 + \lambda)^p}\right).$$

Therefore, for  $p$  large enough,

$$v_p \leq v_\infty + Be^{-\beta p},$$

for some  $B > 0$  and  $\beta = \min\{\alpha, \log(1 + \lambda)\}$ .

Now if  $\lambda = 0$ , we choose  $\delta = p^{-1/\alpha}$  in (5.2.14) which gives

$$\begin{aligned} v_p &\leq \left(v_\infty + \frac{A}{p}\right) \exp\left(\frac{1}{p} \log(1 + d \log(Lp^{1/\alpha}))\right) \\ &= v_\infty + \frac{v_\infty}{p} \log(\log(p)) + o\left(\frac{\log(\log(p))}{p}\right) \end{aligned}$$

which ends the proof. □

### Upper and lower bounds in the discrete case

Let us now consider the discrete case where for some points  $x_1, \dots, x_n$  and  $y_1, \dots, y_m$  in  $\mathbb{R}^d$

$$\mu = \sum_{i=1}^n \mu^i \delta_{x_i} \quad \text{and} \quad \nu = \sum_{j=1}^m \nu^j \delta_{y_j} \quad (5.2.15)$$

with (strictly, without loss of generality) positive weights  $\mu_i$  and  $\nu_j$  summing to 1. As in Section 4.1.3 and 5.1.2,  $c_{i,j} := c(x_i, y_j) \geq 0$  and  $\gamma$  will denote, with a small abuse of notation, both the transport plan and the  $n \times m$  matrix with entries  $\gamma_{i,j}$ .

*Remark 5.2.12.* In the discrete setting transport plans have a finite entropy with respect to  $\mu \otimes \nu$ , with the (crude) bound

$$H(\gamma | \mu \otimes \nu) \leq M := - \sum_{i=1}^n \mu_i \log(\mu_i) - \sum_{j=1}^m \nu_j \log(\nu_j)$$

for every  $\gamma \in \Pi(\mu, \nu)$ . So if  $v_\infty \geq 1 + \lambda$  with  $\lambda \geq 0$ , taking  $\gamma_\infty$  a minimizer of  $J_\infty$ , we obtain

$$\begin{aligned} v_p &\leq J_p(\gamma_\infty) \leq v_\infty \left(1 + \frac{M}{(1 + \lambda)^p}\right)^{\frac{1}{p}} \\ &\leq v_\infty \left(1 + \frac{M}{p(1 + \lambda)^p} + o\left(\frac{M}{p(1 + \lambda)^p}\right)\right) \end{aligned}$$

which gives an exponentially decaying upper bound for  $v_p - v_\infty$  for  $\lambda > 0$  and an algebraic upper bound  $v_p - v_\infty \leq O(1/p)$  if  $\lambda = 0$ . The fact that  $v_\infty \geq 1$  therefore ensures that  $p(v_p - v_\infty)$  is bounded from above.

It turns out, that in the discrete setting, the condition  $v_\infty \geq 1$  guarantees that we also have an algebraically decaying lower bound for the error. To see this, we first need the following:

**Lemma 5.2.13.** *Let  $\mu$  and  $\nu$  be discrete measures i.e. of the form (5.2.15) and define*

$$F_\infty := \arg \min_{\Pi(\mu, \nu)} \|c\|_{L^\infty(\gamma)} = \{\gamma \in \Pi(\mu, \nu) : J_\infty(\gamma) = v_\infty\}$$

and for every  $\gamma \in F_\infty$ ,

$$m(\gamma) := \max\{\gamma_{ij} : \gamma_{ij} > 0, c_{ij} = v_\infty\}.$$

Then there is some  $\theta > 0$  such that  $m(\gamma) \geq \theta$ , for every  $\gamma \in F_\infty$ .

*Proof.* Since  $v_\infty$  is the minimum of  $J_\infty$  over  $\Pi(\mu, \nu)$ , one can write  $F_\infty$  as the set of transport plans for which

$$\gamma_{ij} > 0 \Rightarrow c_{ij} - v_\infty \leq 0$$

or equivalently

$$l(\gamma) := \sum_{ij} \gamma_{ij}(c_{ij} - v_\infty)_+ = 0.$$

In other words,  $F_\infty$  is the facet of  $\Pi(\mu, \nu)$  where the linear form  $l$  (which is nonnegative on  $\Pi(\mu, \nu)$ ) achieves its minimum, it is therefore a convex polytope, whose extreme points belong to the (finite) set of extreme points of  $F_\infty$ . Let us then denote the set of extreme points of  $F_\infty$  by  $\{\gamma^s, s \in S\}$  with  $S$  a finite index set. Thanks to Minkowski's theorem, we can write any  $\gamma \in F_\infty$  as<sup>6</sup>

$$\gamma := \sum_{s \in S} \alpha_s \gamma^s, \tag{5.2.16}$$

for some weights  $\alpha_s \geq 0$  summing to 1. In particular we may pick  $s_0 \in S$  with  $\alpha_{s_0} \geq \frac{1}{|S|}$  (with  $|S|$  denoting the cardinality of  $S$ ). Then we have

$$m(\gamma) \geq \frac{m(\gamma^{s_0})}{|S|} \geq \theta := \min_{s \in S} \frac{m(\gamma^s)}{|S|}$$

the strict positivity of  $\theta$  then follows from the fact that  $S$  is finite and  $m(\gamma^s) > 0$  for every  $s \in S$ .  $\square$

We are now ready to prove the announced lower bound.

**Proposition 5.2.14** (Lower bound, discrete case). *Assume that  $\mu$  and  $\nu$  are discrete measures i.e. of the form (5.2.15) and that  $v_\infty \geq 1$ , then  $p(v_p - v_\infty)$  is bounded from below. Hence since  $p(v_p - v_\infty)$  is also bounded from above (see Remark 5.2.12), we have*

$$v_p - v_\infty = O\left(\frac{1}{p}\right).$$

---

<sup>6</sup>Another way to show (5.2.16) is to observe that if  $\gamma \in F_\infty$  then there exists  $\gamma^s \in \Pi(\mu, \nu)$ , with  $s \in S$ , a finite set of indexes, such that  $\gamma := \sum_{s \in S} \alpha_s \gamma^s$ . By Remark 4.1.20,  $J_\infty(\gamma) = C_\infty(\gamma) = \max_s C_\infty(\gamma^s)$  and thus  $\gamma^s \in F_\infty$  for every  $s \in S$ .

*Proof.* Let us argue by contradiction and assume that  $p(v_p - v_\infty)$  is unbounded from below, then there is a sequence  $p_n \rightarrow \infty$  as  $n \rightarrow \infty$  such that

$$\lim_n p_n(v_{p_n} - v_\infty) = -\infty. \quad (5.2.17)$$

Letting  $\gamma_n$  be the minimizer of  $J_{p_n}$ , passing to a subsequence if necessary, we may assume that  $\gamma_n$  converges to some  $\gamma_\infty$  which belongs to  $F_\infty$  (as defined in Lemma 5.2.13) since  $v_\infty \geq 1$ . In particular, there exists  $i_0, j_0$  such that

$$c_{i_0, j_0} = v_\infty \text{ and } \gamma_{\infty, i_0, j_0} =: \gamma_\infty^{i_0 j_0} \geq \theta > 0,$$

where  $\theta$  is the lower bound from Lemma 5.2.13. But since  $\gamma_n^{i_0 j_0}$  converges to  $\gamma_\infty^{i_0 j_0}$  we have, for large enough  $n$ ,  $\gamma_n^{i_0 j_0} \geq \frac{\theta}{2}$ , hence, using the fact that  $c_{i_0 j_0} = v_\infty$  and the nonnegativity of the entropy

$$\begin{aligned} v_{p_n} &\geq v_\infty \left(\frac{\theta}{2}\right)^{\frac{1}{p_n}} = v_\infty \exp\left(\frac{1}{p_n} \log \frac{\theta}{2}\right) \\ &\geq v_\infty \left(1 + \frac{1}{p_n} \log \frac{\theta}{2}\right) \end{aligned}$$

which is the desired contradiction to (5.2.17). □

### 5.2.4 A large deviations upper bound

In this (somehow independent) paragraph, our goal is to discuss a (partial) extension of the large deviations results of [22] to the  $L^\infty$ -optimal transport framework. Considering the Monge-Kantorovich problem (OT)

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y)$$

it is well-known (as we have seen in Theorem 4.1.10) that the optimal plans  $\gamma \in \Pi(\mu, \nu)$  for (OT) are characterized by a property of  $c$ -cyclical monotonicity of  $\Gamma := \text{supp } \gamma$  (see Definition 4.1.8).

To analyze fine convergence properties of the entropic approximation ( $\varepsilon$ EOT) of (OT), assuming convergence (taking a subsequence if necessary) as  $\varepsilon \rightarrow 0$ , of the minimizer  $\gamma_\varepsilon$  of ( $\varepsilon$ EOT) to some  $\gamma$  and denoting by  $\Gamma$  the  $c$ -cyclically monotone set  $\text{supp } \gamma$ , the authors of [22] introduced

$$I(x, y) := \sup_{k \geq 2} \sup_{(x_i, y_i)_{i=2}^k \subset \Gamma} \sup_{\sigma \in \Sigma(k)} \left\{ \sum_{i=1}^k c(x_i, y_i) - \sum_{i=1}^k c(x_i, y_{\sigma(i)}) \right\}, \quad (x, y) \in \mathbb{R}^d \times \mathbb{R}^d,$$

with  $(x_1, y_1) = (x, y)$ . They proved that  $I$  is a good rate function for the family of optimal entropic plans,  $\{\gamma_\varepsilon\}_{\varepsilon > 0}$  in the sense that (see for instance [119])  $I$  is non-negative and not identical equal to  $+\infty$ ,  $I$  is lower semicontinuous and it obeys, under very general conditions, the large deviations principle

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log(\gamma_\varepsilon(C)) &\leq - \inf_{(x, y) \in C} I(x, y) \quad \text{and} \\ \liminf_{\varepsilon \rightarrow 0} \varepsilon \log(\gamma_\varepsilon(U)) &\geq - \inf_{(x, y) \in U} I(x, y), \end{aligned}$$

for every compact  $C$  and every open  $U$  included in  $\text{supp } \mu \times \text{supp } \nu$

Denoting by  $\gamma_{p,\varepsilon}$  the minimizer of  $J_{p,\varepsilon}$ , the results of [22] (using  $c^p$  instead of  $c$ ) of course apply to the convergence of  $\gamma_{p,\varepsilon}$  as  $\varepsilon \rightarrow 0^+$  for a *fixed* exponent  $p$ . For  $L^\infty$  optimal transport, it makes more sense to rather consider the situation where  $\varepsilon > 0$  is fixed and  $p$  tends to  $\infty$ . More precisely, we know from Theorem 5.2.9, that if  $c \geq 1$ ,  $\varepsilon > 0$  is fixed, as  $p \rightarrow \infty$  the family  $\{\gamma_{p,\varepsilon}\}_{p \geq 1}$  weakly converges (again possibly after an extraction) to some  $\gamma_\infty$ , such that  $\Gamma_\infty := \text{supp}(\gamma_\infty)$  is  $\infty$ -cyclically monotone.

Therefore we will assume throughout this paragraph that

- $c \geq 1$ ,
- $\varepsilon > 0$  being fixed, the sequence of minimizers  $(\gamma_{p,\varepsilon})_{p \geq 1}$  weakly converges as  $p \rightarrow \infty$  to some  $\gamma_\infty$ , with ( $\infty$ -cyclically monotone) support  $\Gamma_\infty$ .

Let us define for every  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$

$$I_\infty(x, y) := \sup_{k \geq 2} \sup_{(x_i, y_i)_{i=2}^k \subset \Gamma_\infty} \sup_{\sigma \in \Sigma(k)} \left\{ \max_{1 \leq i \leq k} \{c(x_i, y_i)\} - \max_{1 \leq i \leq k} \{c(x_i, y_{\sigma(i)})\} \right\},$$

where  $(x_1, y_1) = (x, y)$ . Also define

$$\tilde{I}_\infty(x, y) := \sup_{k \geq 2} \sup_{(x_i, y_i)_{i=2}^k \subset \Gamma_\infty} \left\{ \max_{1 \leq i \leq k} \{c(x_i, y_i)\} - \max_{1 \leq i \leq k} \{c(x_i, y_{i+1})\} \right\},$$

where  $(x_1, y_1) = (x, y)$  and  $y_{k+1} = y_1$ . In our supremal optimal transport setting, we cannot really expect that  $I_\infty$  is a good rate function for  $(\gamma_{p,\varepsilon})_{p \geq 1}$ ; indeed,  $\arg \min_{\Pi(\mu, \nu)} J_\infty$  is unchanged when replacing  $c$  with a strictly increasing function of  $c$ , while the same does not hold for the function  $I_\infty$ . However it can be interesting to have a better understanding of the function  $I_\infty$ , which still provides an upper bound for the family  $(\gamma_{p,\varepsilon})$ , as we will see in Proposition 5.2.17.

**Lemma 5.2.15.** *Let  $I_\infty$  and  $\tilde{I}_\infty$  be defined as above, then*

- $I_\infty$  and  $\tilde{I}_\infty$  are related by  $I_\infty = \max(0, \tilde{I}_\infty)$ ,
- $I_\infty$  and  $\tilde{I}_\infty$  are lower semicontinuous,  $I_\infty \geq 0$ ,  $I_\infty = 0$  on  $\Gamma_\infty$ ,
- $I_\infty$  and  $\tilde{I}_\infty$  coincide on  $(\text{supp } \mu \times \mathbb{R}^d) \cup (\mathbb{R}^d \times \text{supp } \nu)$ .

*Proof.* The fact that  $I_\infty \geq \max(0, \tilde{I}_\infty)$  is obvious as well as the fact that  $\tilde{I}_\infty = 0$  on  $\Gamma_\infty$ . Fix now  $(x, y) = (x_1, y_1) \in \mathbb{R}^d \times \mathbb{R}^d$ ,  $k \geq 2$ ,  $(x_2, y_2), \dots, (x_k, y_k)$  in  $\Gamma_\infty$  and  $\sigma \in \Sigma(k)$ . We can then partition  $\{1, \dots, k\}$  into  $I_0$  the (possibly empty) set of fixed-points of  $\sigma$  and disjoint (empty if  $\sigma$  is the identity) orbits  $I_1, \dots, I_l$  on each of which  $\sigma$  is a cycle, this means that for  $j = 1, \dots, l$ , we may denote  $(x_i, y_i)_{i \in I_j}$  as  $(\tilde{x}_r^j, \tilde{y}_r^j)_{r=1, \dots, |I_j|}$  and  $(x_i, y_{\sigma(i)})_{i \in I_j}$  as  $(\tilde{x}_r^j, \tilde{y}_{r+1}^j)_{r=1, \dots, |I_j|}$  with the convention  $\tilde{y}_{|I_j|+1}^j = \tilde{y}_1^j$ . We now observe that

$$\max_{1 \leq i \leq k} \{c(x_i, y_i)\} - \max_{1 \leq i \leq k} \{c(x_i, y_{\sigma(i)})\} \leq \max_j \left\{ \max_{i \in I_j} c(x_i, y_i) - \max_{i \in I_j} c(x_i, y_{\sigma(i)}) \right\}.$$

where the max with respect to  $j$  is taken on indices for which  $I_j$  is nonempty. To shorten notations, for such a  $j$  let us set

$$\beta_j := \max_{i \in I_j} c(x_i, y_i) - \max_{i \in I_j} c(x_i, y_{\sigma(i)}).$$

Of course if  $I_0$  is nonempty,  $\beta_0 = 0$ . Now if  $j \geq 1$  and  $I_j$  is nonempty

$$\beta_j = \max_{r=1, \dots, |I_j|} c(\tilde{x}_r^j, \tilde{y}_r^j) - \max_{r=1, \dots, |I_j|} c(\tilde{x}_r^j, \tilde{y}_{r+1}^j) \leq \tilde{I}_\infty(\tilde{x}_1^j, \tilde{y}_1^j).$$

So, if  $(\tilde{x}_1^j, \tilde{y}_1^j) = (x_1, y_1)$ ,  $\beta_j \leq \tilde{I}_\infty(x, y)$  and if  $(\tilde{x}_1^j, \tilde{y}_1^j) \neq (x_1, y_1)$ , then  $(\tilde{x}_1^j, \tilde{y}_1^j) \in \Gamma_\infty$ , hence  $\tilde{I}_\infty(\tilde{x}_1^j, \tilde{y}_1^j) = 0$  by the definition of  $\tilde{I}_\infty$  and the fact that  $\Gamma_\infty$  is  $\infty$ -cyclically monotone. In other words, we can bound from above each  $\beta_j$  by  $\max(0, \tilde{I}_\infty(x, y))$ . Taking suprema with respect to  $k$ ,  $(x_2, y_2), \dots, (x_k, y_k)$  in  $\Gamma_\infty$  and  $\sigma \in \Sigma(k)$ , we thus get  $I_\infty \leq \max(0, \tilde{I}_\infty)$ . Lower semi continuity of  $I_\infty$  and  $\tilde{I}_\infty$  follows from the continuity of  $c$ , we have already observed that  $\tilde{I}_\infty \leq 0$  on  $\Gamma_\infty$  which implies that  $I_\infty = \max(0, \tilde{I}_\infty) = 0$  on  $\Gamma_\infty$ . Finally assume that  $x \in \text{supp } \mu$  and  $y \in \mathbb{R}^d$ , since  $\Gamma_\infty = \text{supp } \gamma_\infty$  is compact and  $\gamma_\infty \in \Pi(\mu, \nu)$ , there exists  $y' \in \mathbb{R}^d$  such that  $(x, y') \in \Gamma_\infty$ , taking  $(x_1, y_1) = (x, y)$ ,  $(x_2, y_2) = (x, y')$  as a competitor in the definition of  $\tilde{I}_\infty(x, y)$  we see that  $\tilde{I}_\infty(x, y) \geq 0$  hence  $I_\infty(x, y) = \tilde{I}_\infty(x, y)$ . The same argument shows that  $I_\infty$  and  $\tilde{I}_\infty$  coincide on  $\mathbb{R}^d \times \text{supp } (\nu)$ .  $\square$

**Lemma 5.2.16.** *Let us fix  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ . Suppose that for some  $\delta \in \mathbb{R}$ ,  $k \in \mathbb{N}$ ,  $k \geq 2$  and  $(x_i, y_i)_{i=2}^k \subset \text{supp } \gamma_\infty$ , we have*

$$\max_{1 \leq i \leq k} \{c(x_i, y_i)\} - \max_{1 \leq i \leq k} \{c(x_i, y_{i+1})\} > \delta, \text{ where } (x_1, y_1) := (x, y).$$

Then there exist  $\alpha > 0$ ,  $r > 0$  and  $p_0 \geq 1$  such that

$$\gamma_{p,\varepsilon}(B_r(x, y)) \leq \alpha e^{-\frac{p\delta}{\varepsilon}}, \quad \forall p \geq p_0,$$

where  $\gamma_{p,\varepsilon}$  is the minimizer of  $J_{p,\varepsilon}$ .

*Proof.* Of course if  $\delta \leq 0$ , one can just take  $\alpha = 1$  so we may assume that  $\delta > 0$ . Reasoning as in the proof of Theorem 5.2.9 (recall that we have assumed  $c \geq 1$ ), we know that there exist  $p_0$  and  $r > 0$  such that

$$\sum_{i=1}^k c^p(x'_i, y'_i) - \sum_{i=1}^k c^p(x'_i, y'_{i+1}) > p\delta,$$

for every  $p \geq p_0$  and  $(x'_i, y'_i)_{i=1}^k \subset B_r(x_1, y_1) \times \dots \times B_r(x_k, y_k)$ . Then  $B_r(x_1, y_1) \times \dots \times B_r(x_k, y_k) \subset A_{k,c^p}(p\delta)$  so, thanks to Lemma 5.2.8,

$$\gamma_{p,\varepsilon}^k(B_r(x_1, y_1) \times \dots \times B_r(x_k, y_k)) \leq e^{-\frac{p\delta}{\varepsilon}}.$$

Moreover  $\liminf_{p \rightarrow \infty} \gamma_{p,\varepsilon}(B_r(x_i, y_i)) \geq \gamma_\infty(B_r(x_i, y_i)) > \beta$ , for all  $2 \leq i \leq k$ , for some  $\beta > 0$  since  $(x_i, y_i)_{i=2}^k \subset \text{supp } \gamma_\infty$ , then

$$\gamma_{p,\varepsilon}(B_r(x, y)) \leq \left(\frac{\beta}{2}\right)^{1-k} e^{-\frac{p\delta}{\varepsilon}},$$

for all  $p \geq p_0$  (possibly replacing  $p_0$  with a larger one).  $\square$

**Proposition 5.2.17.** *Under the assumptions of this paragraph, for any compact set  $C \subset \mathbb{R}^d \times \mathbb{R}^d$  s.t.  $C \cap \text{supp } \gamma \neq \emptyset$ , one has*

$$\limsup_{p \rightarrow \infty} \frac{\varepsilon}{p} \log \gamma_{p,\varepsilon}(C) \leq - \inf_{C \cap (\text{supp } \mu \times \text{supp } \nu)} \tilde{I}_\infty \leq - \inf_C I_\infty.$$

*Proof.* First note that since  $\gamma_{p,\varepsilon}$  is supported on  $\text{supp } \mu \times \text{supp } \nu$ ,

$$\gamma_{p,\varepsilon}(C) = \gamma_{p,\varepsilon}(C \cap (\text{supp } \mu \times \text{supp } \nu))$$

and there is nothing to prove if  $C$  is disjoint from  $\text{supp } \mu \times \text{supp } \nu$ . Therefore we can assume that  $C \cap (\text{supp } \mu \times \text{supp } \nu) \neq \emptyset$ . It then follows from Lemma 5.2.15 that

$$\inf_{C \cap (\text{supp } \mu \times \text{supp } \nu)} \tilde{I}_\infty = \inf_{C \cap (\text{supp } \mu \times \text{supp } \nu)} I_\infty \geq \inf_C I_\infty.$$

Now let  $\eta > 0$  and  $(x, y) \in C \cap (\text{supp } \mu \times \text{supp } \nu)$ . By definition of  $\tilde{I}_\infty(x, y)$ , given  $\eta > 0$ , there exist  $k \geq 2$  and  $(x_i, y_i)_{i=2}^k \subset \Gamma_\infty$ , such that (setting as usual  $(x_1, y_1) = (x, y)$  and  $y_{k+1} = y_1 = y$ )

$$\max_{1 \leq i \leq k} \{c(x_i, y_i)\} - \max_{1 \leq i \leq k} \{c(x_i, y_{i+1})\} > \min(\eta^{-1}, \tilde{I}_\infty(x, y)) - \eta.$$

Note that the truncation is used to handle the case where  $\tilde{I}_\infty(x, y) = +\infty$ . By Lemma 5.2.16 we know that there exist  $\alpha, r > 0$  such that

$$\gamma_{p,\varepsilon}(B_r(x, y)) \leq \alpha \exp\left(\frac{-p(\min(\eta^{-1}, \tilde{I}_\infty(x, y)) - \eta)}{\varepsilon}\right).$$

Then

$$\limsup_{p \rightarrow \infty} \frac{\varepsilon}{p} \log \gamma_{p,\varepsilon}(B_r(x, y)) \leq -\min(\eta^{-1}, \tilde{I}_\infty(x, y)) + \eta$$

and, by compactness of  $C$ ,

$$\limsup_{p \rightarrow \infty} \frac{\varepsilon}{p} \log \gamma_{p,\varepsilon}(C) \leq -\inf_{C \cap (\text{supp } \mu \times \text{supp } \nu)} \min(\eta^{-1}, \tilde{I}_\infty) + \eta$$

which, letting  $\eta \rightarrow 0^+$ , yields the desired upper bound.  $\square$

### 5.2.5 Numerical examples

In this section we will provide some - in our opinion interesting - numerical examples of what we have discussed and proved theoretically in the previous sections.

In all the examples we have used the Sinkhorn algorithm (which has been discussed in Section 5.1.2) to find a good approximation (with error smaller than  $10^{-5}$ ) of the solution of  $(\varepsilon\text{DEOT})$  with  $c$  replaced by  $c^p$ :

$$\gamma_{p,\varepsilon} = \arg \min_{\Pi(\mu, \nu)} \sum_{i,j} \gamma_{i,j} c_{i,j}^p + \varepsilon H(\gamma | \mu \otimes \nu). \quad ((p, \varepsilon)\text{DEOT})$$

Clearly such a  $\gamma_{p,\varepsilon}$  minimizes also  $J_{p,\varepsilon} = \left(\sum_{i,j} \gamma_{i,j} c_{i,j}^p + \varepsilon H(\gamma | \mu \otimes \nu)\right)^{\frac{1}{p}}$ .

In light of Theorem 5.2.4 we expect to find, for suitable  $p$  and  $\varepsilon$ , a good approximation of an optimal plan for the problem

$$\min_{\gamma \in \Pi(\mu, \nu)} \max \{c_{i,j} : \gamma_{i,j} \neq 0\}.$$

Furthermore, thanks to Theorem 5.2.9, if  $c \geq 1$  we expect to find a plan close to an  $\infty$ -cyclically monotone one.

**Notation.** In the following, given  $\mu$  and  $\nu$ , two discrete probability measures,  $a := (\mu_i)_i$  and  $b := (\nu_j)_j$  will denote respectively the vectors of the weights  $\mu_i$  of  $\mu$  and  $\nu_j$  of  $\nu$  and  $P = (\gamma_{ij})_{i,j}$  will indicate the output matrix of the Sinkhorn algorithm, associated to the plan  $\gamma$  which approximates the optimal solution  $\gamma_{p,\varepsilon}$  of  $((p, \varepsilon)$ DEOT).

All the examples will be in dimension  $d = 2$ .

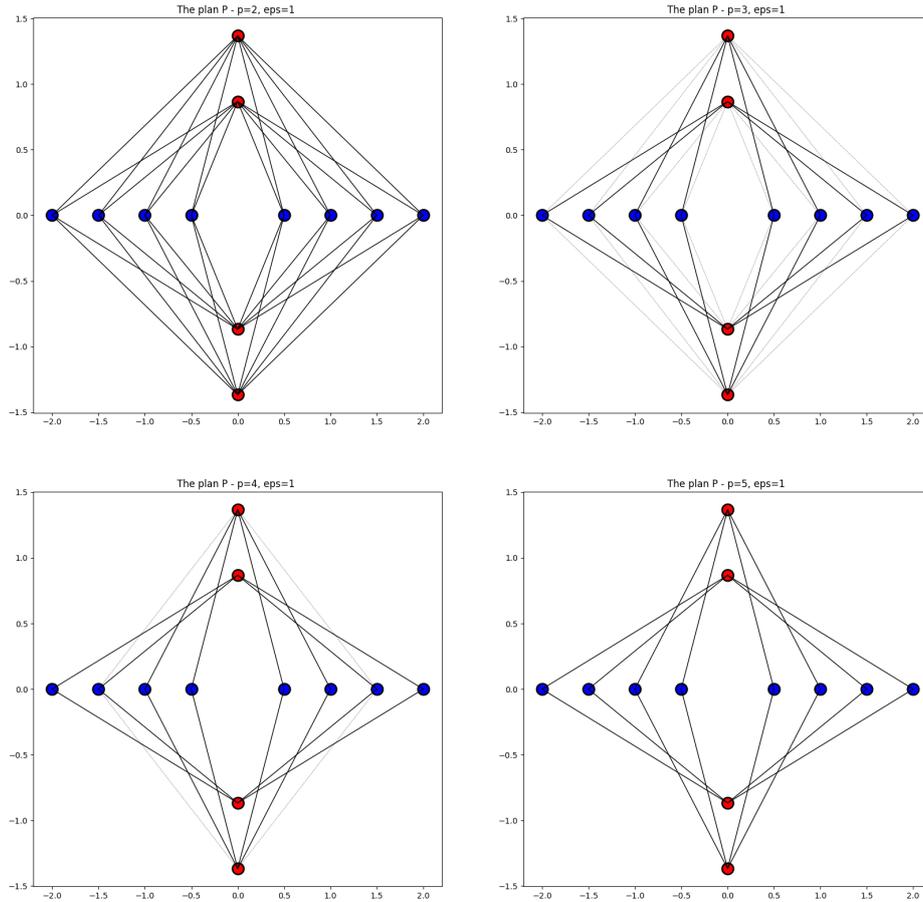


Figure 5.2.1: Example of convergence of the plan to the  $\infty$ -cm plan, for  $c(x, y) = |x - y|^p$ , for  $p \in \{2, 3, 4, 5\}$ ,  $\varepsilon = 1$  and  $\mu$  (blue),  $\nu$  (red) having orthogonal supports.

In the first example, as shown by Figure 5.2.1, we consider  $c^p = |x - y|^p$ , for  $p \in \{2, 3, 4, 5\}$ ,  $\mu$  which is uniformly concentrated on the blue points

$$\{(-2, 0), (-1.5, 0), (-1, 0), (-0.5, 0), (0.5, 0), (1, 0), (1.5, 0), (2, 0)\}$$

and  $\nu$  on the red points

$$\{(0, -1.367), (0, -0.867), (0, 0.867), (0, 1.367)\}.$$

Note that with this choice of  $\text{supp } \mu$  and  $\text{supp } \nu$ ,  $c \geq 1$  everywhere and therefore, thanks to Theorem 5.2.4 and Theorem 5.2.9,  $\Gamma$ -convergence and convergence of the outputs towards

$\infty$ -cm plans holds choosing  $\varepsilon = 1$ .

The simplicity of the problem and the big value of  $c$  allows for the use of the primal version of the Sinkhorn algorithm, discussed in Section 5.1.2. The plan given by the algorithm is represented by arrows: the black ones indicate that a blue point is sent to a red point with high probability, while the gray ones indicate that a blue point is sent to a red point with lower probability (but still not negligible).

We observe that for  $p = 2$ , by the orthogonality of the two supports, every transport plan  $\gamma$  is concentrated on a cyclically monotone set (see Remark 4.1.9) and it is thus optimal (see Theorem 4.1.10) for the Kantorovich problem. Here, since we look for a plan which minimizes also the entropy, the Sinkhorn algorithm selects the most diffuse one. Convergence towards an  $\infty$ -cm plan is really fast and it occurs already for  $p = 5$ .

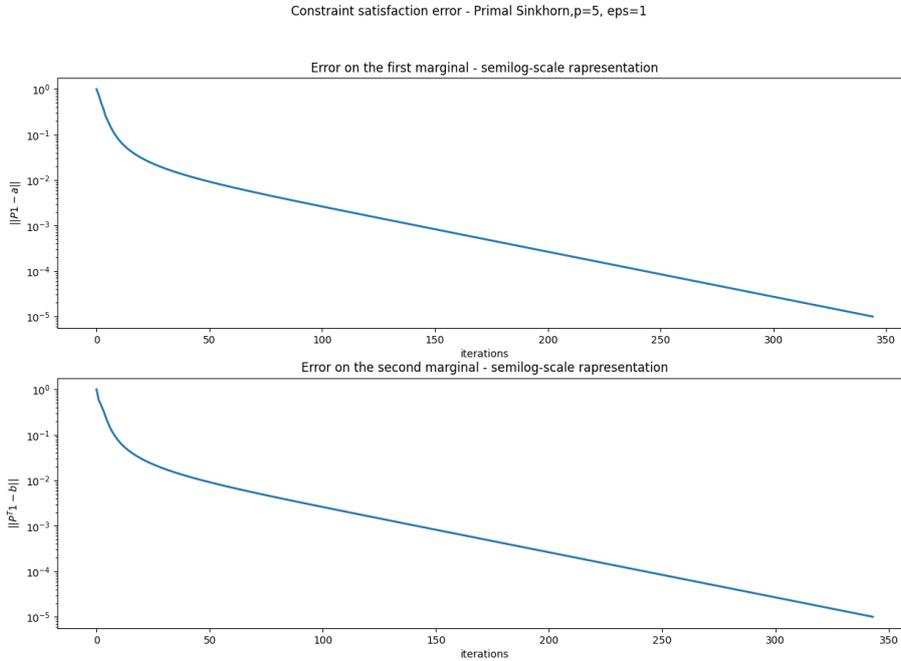


Figure 5.2.2: Error on the marginals: the first image shows the error  $|P\mathbb{1}_4 - a|$  of the output  $P$  on the first marginal and the second one the error  $|P^T\mathbb{1}_8 - b|$  on the second marginal.

Regarding the accuracy, Figure 5.2.2 shows that for  $p = 5$  and  $\varepsilon = 1$  the distance  $|P\mathbb{1}_4 - a|$  between the first marginals of the output  $\gamma$  and the distance  $|P^T\mathbb{1}_8 - b|$  between the second marginal of  $\gamma$  and  $\nu$  is of the order of  $10^{-5}$  after only 350 iterations.

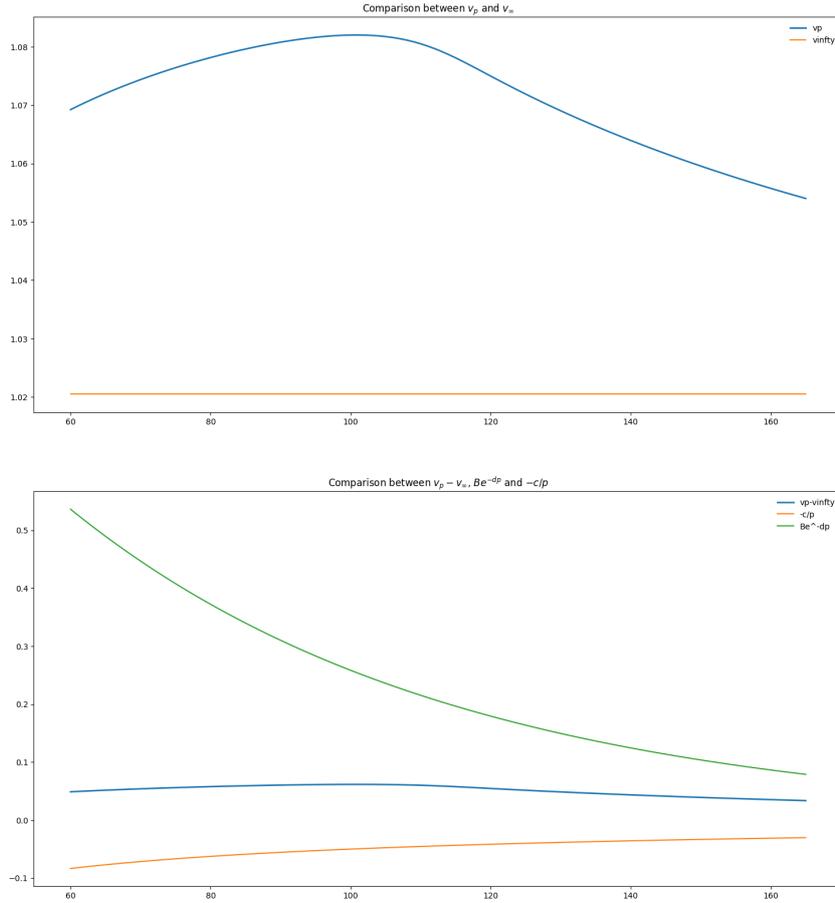


Figure 5.2.3: Comparison among the speed of convergence for  $p \in [65, 160]$ . On the top:  $v_p$  in blue and  $v_\infty$  in orange. On the bottom:  $Be^{-dp}$  in green,  $-\frac{c}{p}$  in orange and  $v_p - v_\infty$  in blue. Here  $c := \alpha$  and  $d := \beta$ .

Finally, Figure 5.2.3 shows the asymptotic behavior of  $v_p := \min_{\Pi(\mu, \nu)} J_p$ . In particular we want to numerically represent the upper and lower bounds on the speed of convergence of  $v_p$  towards  $v_\infty := \min_{\Pi(\mu, \nu)} J_\infty$  proved in Proposition 5.2.11 and Proposition 5.2.14. In light of the fact that in Proposition 5.2.11 and Proposition 5.2.14 is enough to assume a lower bound on  $v_\infty$  and not a point-wise one on  $c$  (as required in order to have convergence to  $\infty$ -cm plans), we have re-scaled the points of the supports of  $\mu$  and  $\nu$ , obtaining  $v_\infty \simeq 1.020469 > 1$ . Even though this example is still not fully satisfying, having a smaller value of  $c$  was needed in order to increase the value of  $p$ , as discussed in Remark 5.2.18 below. The image on the top of Figure 5.2.3 shows in blue how  $v_p$  changes varying  $p$ , which is taken in the interval  $[60, 165]$ . For  $p$  in the same interval, in the image on the bottom we have represented in blue  $v_p - v_\infty$ , in green the upper bound  $Be^{-\beta p}$  and in orange the lower bound  $-\frac{\alpha}{p}$ , where  $\beta = v_\infty$  comes by Proposition 5.2.11 (in this case  $c$  is Lipschitz so  $\alpha = 1 > \log(v_\infty)$ ) and  $\alpha, B$  have been calculated using the mean square method (in Figure 5.2.3,  $c = \alpha$  and  $d = \beta$ ).

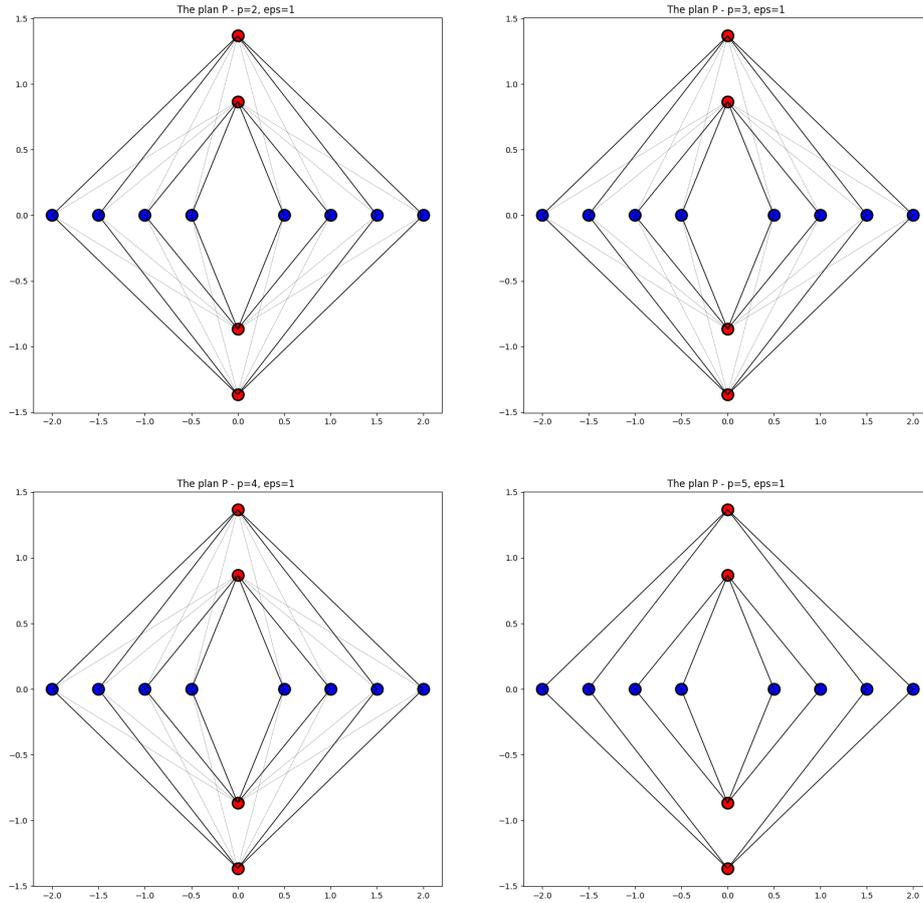


Figure 5.2.4: Example of convergence of the plan to the  $\infty$ -cm plan for  $c(x, y) = (\max\{|x_1 - y_1|, |x_2 - y_2|\})^p$ , for  $p \in \{2, 3, 4, 5\}$ ,  $\varepsilon = 1$  and  $\mu$  (blue),  $\nu$  (red) having orthogonal supports.

We have also considered the same example (see Figure 5.2.4) with the cost function  $c^p(x, y) := (\max\{|x_1 - y_1|, |x_2 - y_2|\})^p$ . In this case the convergence is still fast and the error is small after few iterations (of order  $10^{-5}$  after about 180 iterations).

*Remark 5.2.18.* When  $c > 1$  we have the advantage that not only we don't need small  $\varepsilon$ , but we can even increase it as much as needed (by case 2. in Theorem 5.2.4 we can even choose for instance  $\varepsilon_p = (1 + \lambda)^p$ ), on the other side we can encounter some difficulties when computing the Gibbs kernel  $K_{ij} = e^{-\frac{c_{ij}^p}{\varepsilon}}$ : if  $p$  is large it can happen that, for some  $i, j$ ,  $K_{i,j} = 0$  making impossible to perform the division in the iterations of the primal version of the Sinkhorn algorithm ((5.1.7),(5.1.8)). Fortunately, this problem can be overcome using the Log-Domain version as discussed in the second part of Section 5.1.2.

Also the computation of  $v_p$  can be delicate. Indeed when  $c$  is big the convergence to  $\gamma_\infty$  occurs for relatively small  $p$  and therefore - since  $\gamma_\infty$  is usually sparse - it is likely that the output matrix  $P$  has some 0 entries, preventing the possibility of computing the logarithm in the computation of  $H(\gamma|\mu \otimes \nu)$ . This is why when looking for the asymptotic behavior is important to consider  $\mu, \nu$  and  $c$  such that  $v_\infty > 1$  but  $c \leq 1 + \eta$ , for some suitable  $\eta > 0$ .

## 5.2. $L^\infty$ -ENTROPIC OPTIMAL TRANSPORT

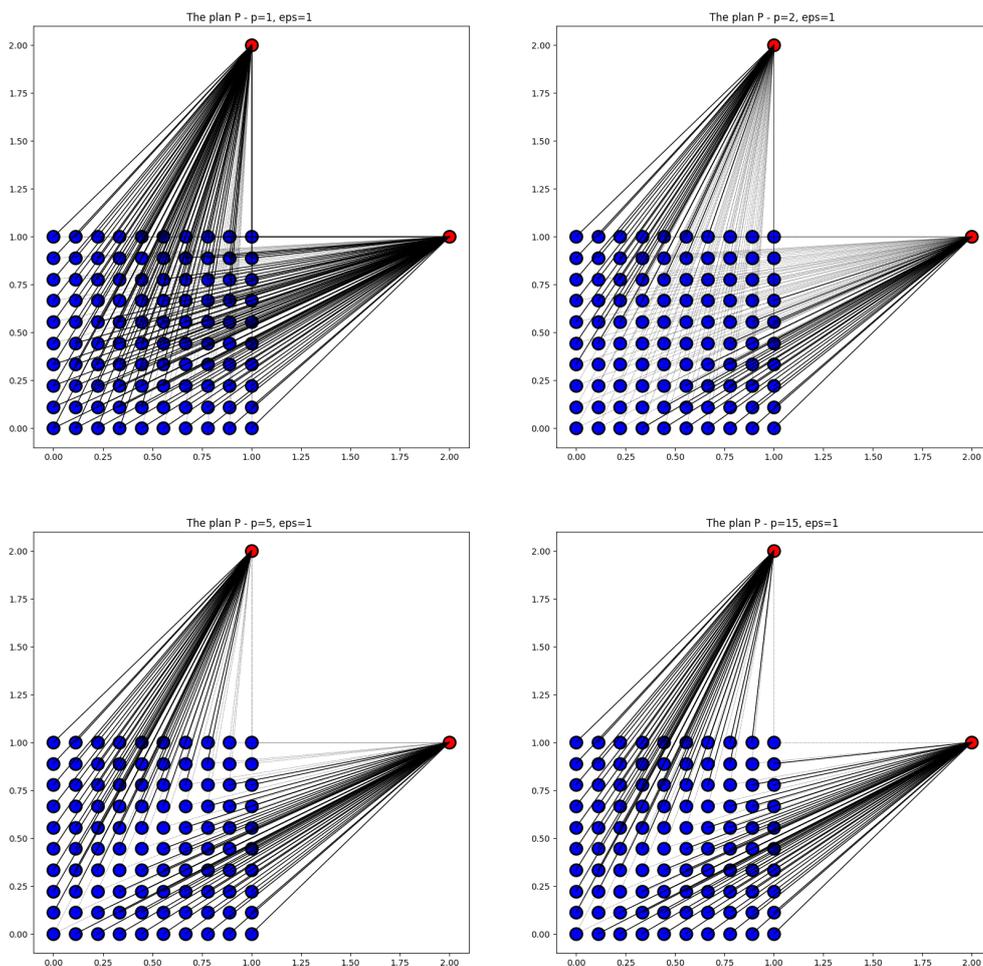


Figure 5.2.5:  $c(x, y) = |x - y|^p$ , for  $p \in \{1, 2, 5, 15\}$ ,  $\varepsilon = 1$ ,  $\mu$  (blue) a uniform discretization of the unitary square and  $\nu$  (red) uniformly concentrated on the points  $(1, 2)$  and  $(2, 1)$ .

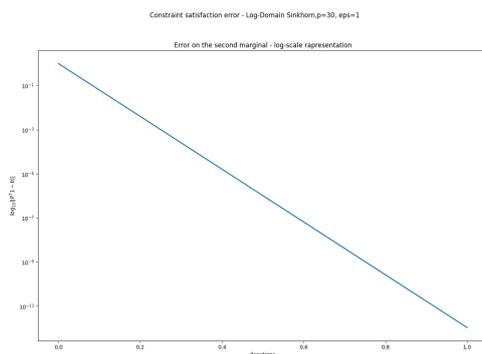


Figure 5.2.6: Error on the second marginal: for  $p = 30$  and  $\varepsilon = 1$  the distance between  $P^\top \mathbb{1}_m$  and  $\nu$  is of order  $10^{-15}$  after about 100 iterations.

Another interesting example is represented by Figure 5.2.5, which shows a numerical implementation of Example 4.1.23:  $\nu$  is concentrated on the points  $(1, 2)$  and  $(2, 1)$  as in

the original version, while  $\mu$  is uniformly concentrated on 100 points which discretize the unitary square. Since every plan is optimal, when  $p$  is smaller, the role of the entropy is more important and the selected plan is more diffuse. While increasing the value of  $p$  the entropy becomes more and more negligible and the plan given by the algorithm is more sparse and closer to the  $\infty$ -cyclically monotone plan. In this example, in light of Remark 5.2.18 we have used Sinkhorn algorithm in Log-Domain and we have considered  $p \in \{1, 2, 5, 15\}$ . Already for  $p = 15$  one can notice a very good approximation of the  $\infty$ -cyclically monotone plan, which in this case is unique (see Theorem 4.2.13 and Example 4.2.15).

Figure 5.2.7 represents the asymptotic behavior of  $v_p - v_\infty$ , studied re-scaling the supports in order to have  $v_\infty \simeq 1.08166$  and considering  $p \in [10, 206]$  and  $\varepsilon = 15$ .

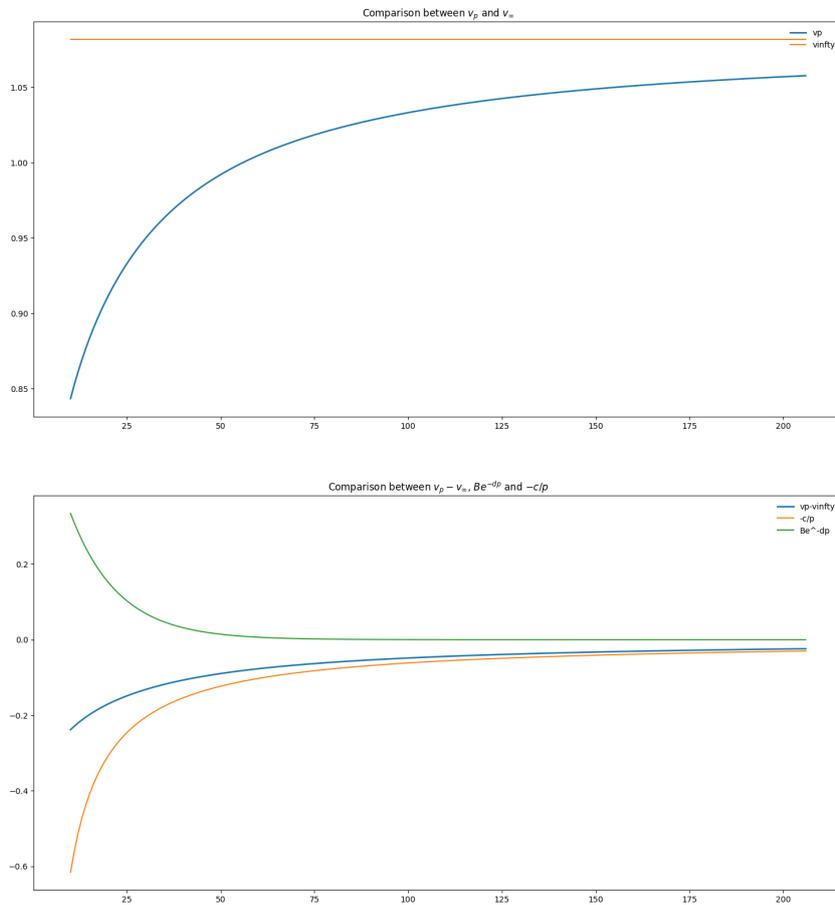


Figure 5.2.7: Comparison among the speed of convergence of  $v_p - v_\infty$ ,  $Be^{dp}$  and  $-\frac{c}{p}$  for  $p \in [10, 206]$ . On the top:  $v_p$  in blue and  $v_\infty$  in orange. On the bottom:  $Be^{dp}$  in green,  $-\frac{c}{p}$  in orange and  $v_p - v_\infty$  in blue. Here  $c, B$  obtained with the square mean method and  $d = v_\infty$  ( $d = \beta$  of Proposition 5.2.11).

A small variation is to consider  $\nu$  which is not uniformly concentrated on the points  $(1, 2)$  and  $(2, 1)$  in order to have an idea of the shape of the  $\infty$ -cyclically monotone plan. In Figure 5.2.8 and Figure 5.2.9, for  $p = 15$  and  $\varepsilon = 1$ , we have considered progressively

## 5.2. $L^\infty$ -ENTROPIC OPTIMAL TRANSPORT

$\nu_1 = 0.4\delta_{(1,2)} + 0.6\delta_{(2,1)}$ ,  $\nu_2 = 0.3\delta_{(1,2)} + 0.7\delta_{(2,1)}$ ,  $\nu_3 = 0.2\delta_{(1,2)} + 0.8\delta_{(2,1)}$  and  $\nu_4 = 0.1\delta_{(1,2)} + 0.9\delta_{(2,1)}$ . Figure 5.2.9 shows which points of the support of  $\mu$  are sent to  $(1, 2)$  (dark blue) and which ones are sent to  $(2, 1)$  (yellow).

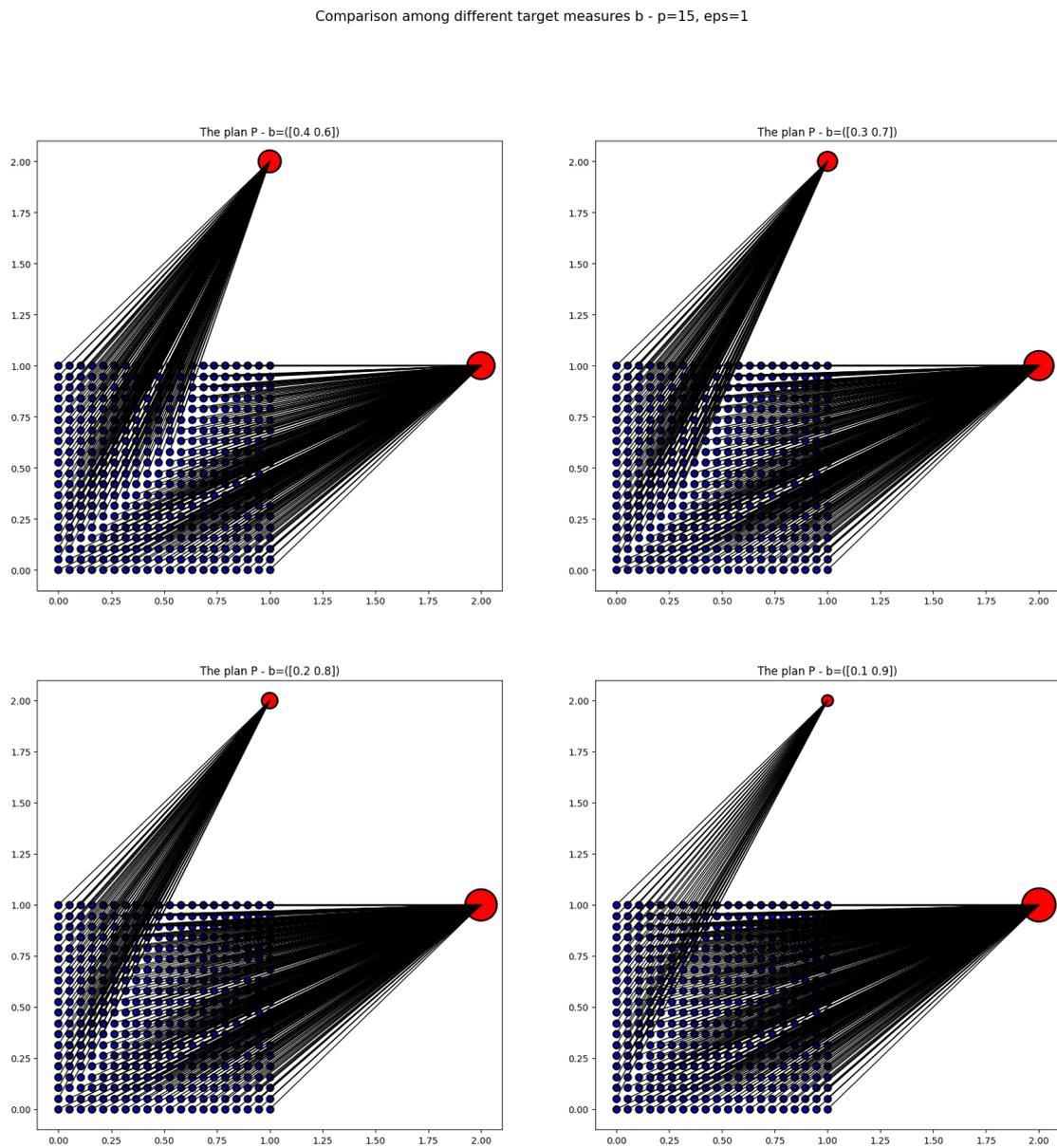


Figure 5.2.8: Comparison among different not uniform target measures.

Comparison among different target measures  $b$  -  $p=15$ ,  $\text{eps}=1$

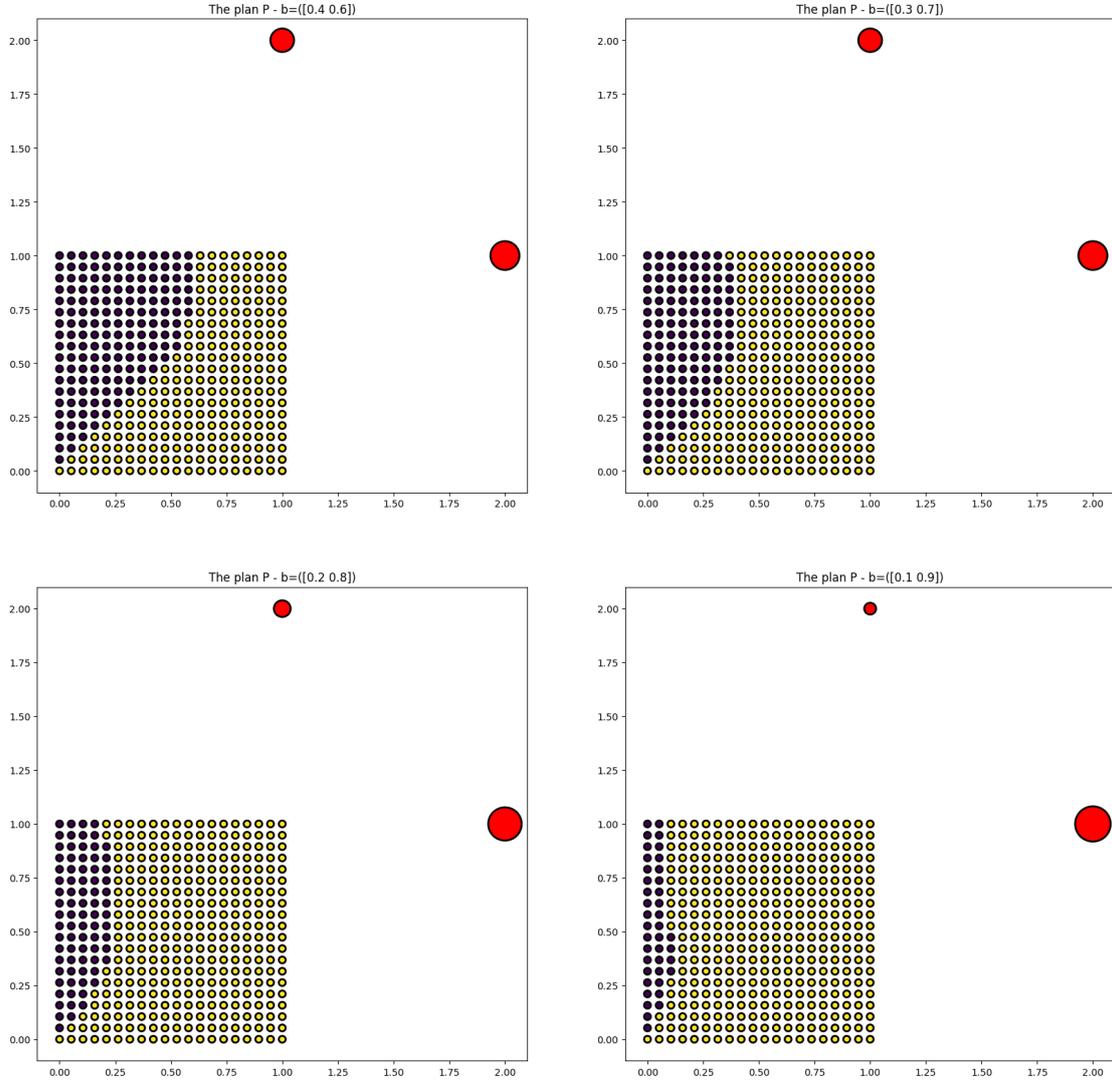


Figure 5.2.9: Comparison among different not uniform target measures: in dark blue points of  $\text{supp } \mu$  which are sent to  $(1, 2)$  and in yellow points of  $\text{supp } \mu$  which are sent to  $(2, 1)$ .

An example of both  $\mu$  and  $\nu$  being the discretization of the Lebesgue measure restricted to a portion of space is given by Figure 5.2.10, in which  $\mu$  is supported on the square  $[-0.25, 0.25] \times [-0.25, 0.25]$  and  $\nu$  is supported on the rectangle  $[1.25, 1.5] \times [-0.5, 0.5]$  and both measures are discretized with 100 points. The figure shows a comparison between the output  $\gamma$  for a fixed  $\varepsilon = 1$  when  $p = 2$  and  $p = 20$ . One can immediately notice that for  $p = 2$  the entropy plays an important role and the algorithm selects the most diffuse plan. Looking at the picture one could try to make a guess of the value of  $v_\infty$ , but in this case we are not able to calculate it exactly. Indeed, even if the size of  $\mu$  and  $\nu$  is the same, to apply what we observed in Proposition 4.1.40 we should be able to perform  $100!$  iterations, which is infeasible in practice!

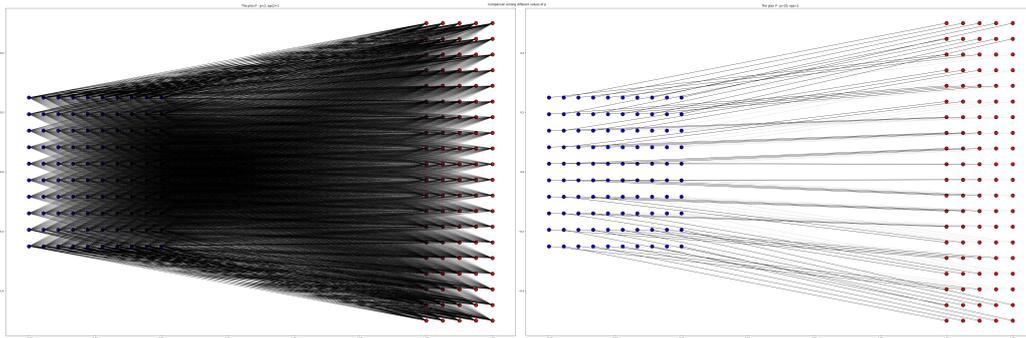


Figure 5.2.10: Comparison among different values of  $p$ :  $\mu$  (blue) discretization of the square  $[-0.25, 0.25] \times [-0.25, 0.25]$ ,  $\nu$  (red) discretization of the rectangle  $[1.25, 1.5] \times [-0.5, 0.5]$ ,  $\varepsilon = 1$ . On the left  $p = 2$  while on the right  $p = 20$ .

An example in which it is possible (even if it is really slow!) to calculate  $v_\infty$  “by hands” is provided by Figure 5.2.11. Here  $\mu$  is concentrated on 8 points, given by

$$\{(x_1, x_2) : x_1 = -0.25 + 0.125 \cdot i, i = 1, \dots, 4, x_2 \in \{-0.1, 0.1\}\}$$

and  $\nu$  is concentrated on 8 equidistant points of the segment starting from the point  $(0.625, 1.25)$  to the point  $(1.25, 0)$  of the line  $y_2 = -2y_1 + 2.5$ . We have computed  $v_\infty$  for the cost  $c(x, y) = |x - y|$  applying Proposition 4.1.40, obtaining  $v_\infty = 1.386473466558011$  and that the points which are at the minimal-maximal distance are  $x_* = (-0.25, -0.1)$  and  $y_* = (0.98214286, 0.53571429)$ , the one connected by the yellow segment in the picture.

Regarding the speed of convergence we rescaled the problem in order to decrease further  $v_\infty = 1.0524606002472119$ . As showed in Figure 5.2.12,  $v_p$  is calculated varying  $p$  in the interval  $[10, 172]$ , with  $\varepsilon = 500^2$ . We observe that in this case, as shown in the picture on the top,  $v_p$  is initially smaller than  $v_\infty$ , then it increases becoming bigger and finally it starts decreasing converging to  $v_\infty$ .

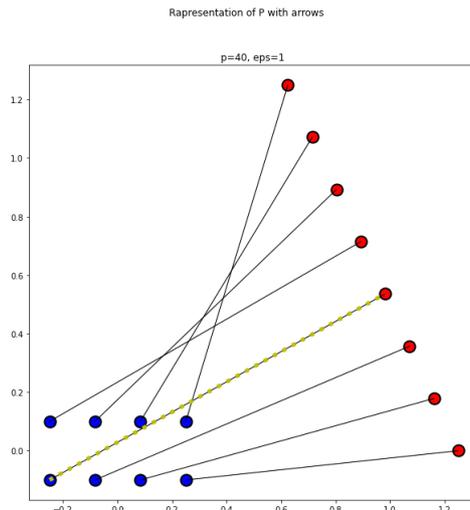


Figure 5.2.11:  $\mu$  and  $\nu$  uniformly distributed both concentrated on 8 points. The value of  $v_\infty$  is  $1.386473466558011$  and it is obtained transporting mass between the two points connected by the yellow segment.

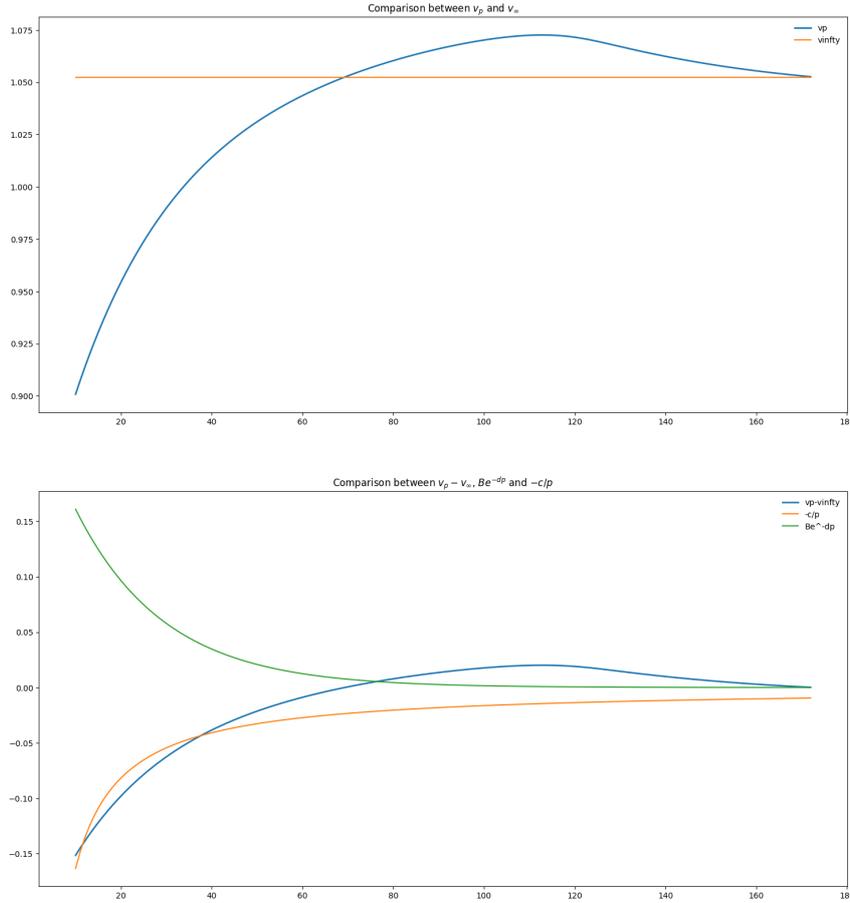


Figure 5.2.12: Comparison among the speed of convergence of  $v_p - v_\infty, Be^{dp}$  and  $-\frac{c}{p}$  for  $p \in [10, 172]$  and  $\varepsilon = 500^2$ . On the top:  $v_p$  in blue and  $v_\infty$  in orange. On the bottom:  $Be^{dp}$  in green,  $-\frac{c}{p}$  in orange and  $v_p - v_\infty$  in blue. Here  $c, B$  obtained with the square mean method and  $d = v_\infty$  ( $d = \beta$  of Proposition 5.2.11).

The last example that we provide is presented in Figure 5.2.13 and Figure 5.2.14 in which we have taken  $\mu$  supported on 31 points which discretize a circle, centered in  $(0, 0)$  and radius  $r = 0.2$ . The mass over those points is distributed as a Gaussian  $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{|x-x_0|^2}{2\sigma^2}}$ , where  $\sigma = 0.2$  and  $x_0 = (0.2, -0.1)$ . The target measure  $\nu$  is instead concentrated on 62 points which discretize two different circles both of radius  $r = 0.2$ , one centered in  $(1.2001, 0)$  and the other centered in  $(0, 1.2001)$ . The mass is distributed as two different Gaussians:  $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{|y-y_0^1|^2}{2\sigma^2}}$ , with  $y_0^1 = (1.1, 0.05)$  and  $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{|y-y_0^2|^2}{2\sigma^2}}$ , with  $y_0^2 = (-0.05, 101)$ . Figure 5.2.13 shows the distribution of the mass in 3-dimension, while Figure 5.2.14 shows the distribution with colors: as explained in the bar, the brighter is the color the more mass is concentrated at the point. In the same picture there is also the representation by arrows of the output  $\gamma$  of the algorithm for  $p = 45$  and  $\varepsilon = 1$ . Finally, in Figure 5.2.15 we can see the error on the second marginal, which is of order  $10^{-5}$  after less than 2000 iterations.

## 5.2. $L^\infty$ -ENTROPIC OPTIMAL TRANSPORT

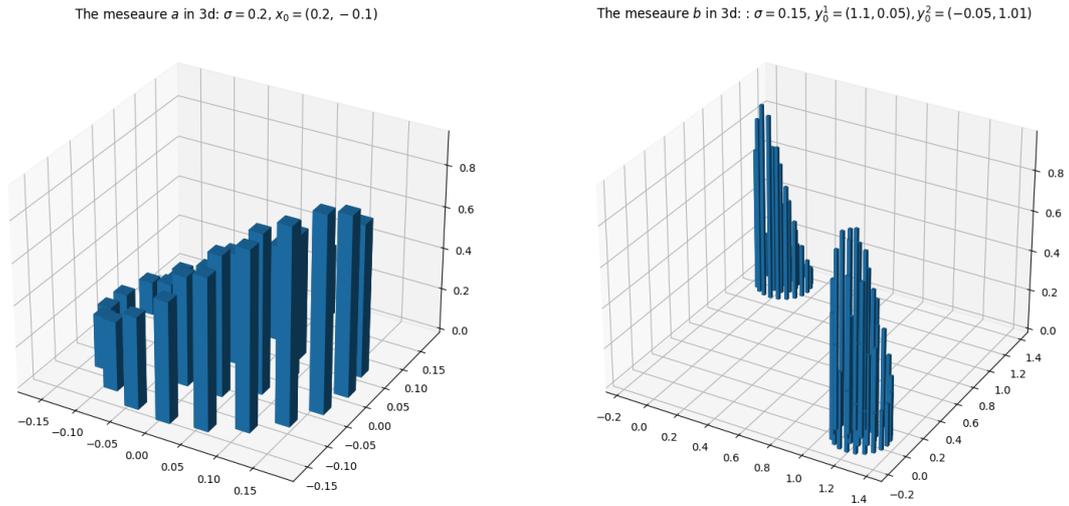


Figure 5.2.13: 3d representation of the normalization of  $\mu = \sum_{i=1}^{31} e^{-\frac{|x_i - x_0|^2}{2\sigma^2}} \delta_{x_i}$  (on the left) and the normalization of  $\nu = \sum_{j=1}^{31} e^{-\frac{|y_j - y_0^1|^2}{2\sigma^2}} \delta_{y_j} + \sum_{j=32}^{62} e^{-\frac{|y_j - y_0^2|^2}{2\sigma^2}} \delta_{y_j}$  (on the right).

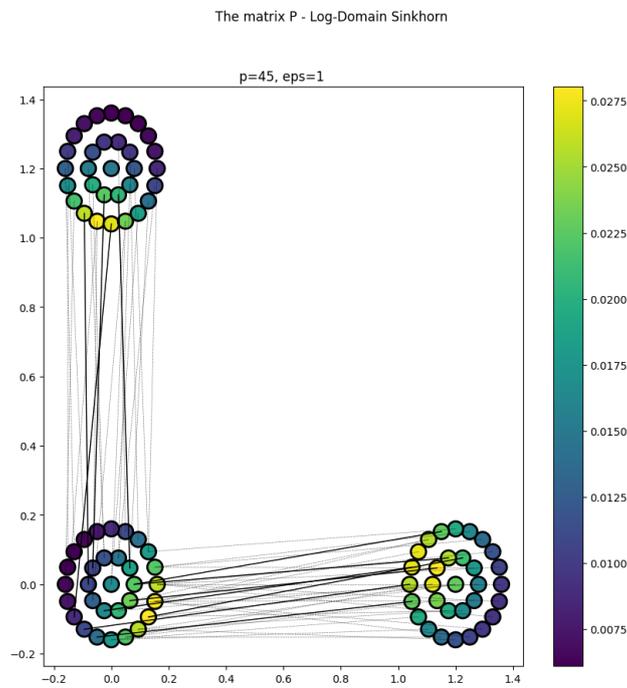


Figure 5.2.14: Representation by colors of  $\mu$  and  $\nu$  and representation by arrows of the output  $\gamma$  for  $p = 45$  and  $\varepsilon = 1$

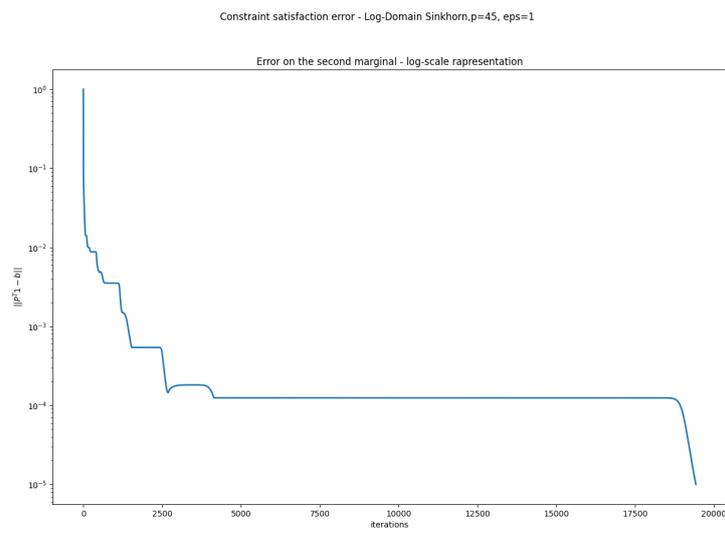


Figure 5.2.15: Error on the second marginal:  $\|P^T \mathbb{1}_{31} - b\|$

# Bibliography

- [1] S. Adams, M. Dirr, M. A. Peletier, and J. Zimmer. From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage, *Comm. Math. Phys.* 307(3):791–815, 2011.
- [2] M. Agueh, and G. Carlier. Barycenters in the Wasserstein space, *SIAM J. Math. Anal.*, 43(2):904–924, 2011.
- [3] L. Ambrosio, E. Brué, and D. Semola. Lectures on optimal transport, *Springer, Cham - UNITEXT* 130:ix+250, 2021.
- [4] L. Ambrosio, N. Gigli, and G. Savaré. Gradient flows in metric spaces and in the space of probability measures, *Birkhäuser Verlag, Basel - Lectures in Mathematics ETH Zürich*, x+334, 2008.
- [5] L. Ambrosio, and S. Rigot. Optimal mass transportation in the Heisenberg group, *J. Funct. Anal.*, 208(2):261–301, 2004.
- [6] L. Ambrosio, and P. Tilli. Topics on analysis in metric spaces. *Oxford University Press, Oxford - Oxford Lecture Series in Mathematics and its Applications* (25):viii+133, 2004.
- [7] G. Aronsson. Minimization problems for the functional  $\sup_x F(x, f(x), f'(x))$ . *Ark. Mat.*, 6:33–53, 1965.
- [8] G. Aronsson. Minimization problems for the functional  $\sup_x F(x, f(x), f'(x))$ . II. *Ark. Mat.*, 6:409–431, 1966.
- [9] G. Aronsson. Extension of functions satisfying Lipschitz conditions. *Ark. Mat.*, 6:551–561, 1967.
- [10] G. Aronsson. On the partial differential equation  $u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy} = 0$ . *Ark. Mat.*, 7:395–425, 1968.
- [11] G. Aronsson. Minimization problems for the functional  $\sup_x F(x, f(x), f'(x))$ , III. *Ark. Mat.*, 7(6): 509 – 512, 1969.
- [12] G. Aronsson. Interpolation under a gradient bound, *J. Aust. Math. Soc.* 87(1), 19–35, 2009.
- [13] G. Aronsson, M. G. Crandall, and P. Juutinen. A tour of the theory of absolutely minimizing functions, *Bull. Amer. Math. Soc. (N.S.)*, 41(4):439–505, 2004.

- 
- [14] M. Bansil, and J. Kitagawa.  $W_\infty$ -transport with discrete target as a combinatorial matching problem *Arch. Math. (Basel)*, 117(2):189–202, 2021.
- [15] E. N. Barron, M. Bocea, and R. R. and Jensen. Duality for the  $L^\infty$  optimal transport problem, *Trans. Amer. Math. Soc.*, 369(5):3289–3323, 2017.
- [16] E. N. Barron, R. R. Jensen, and C. Y. Wang. Lower semicontinuity of  $L^\infty$  functionals. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 18(4):495–517, 2001.
- [17] E. N. Barron, R. R. Jensen, and C. Y. Wang. The Euler equation and absolute minimizers of  $L^\infty$  functionals, *Arch. Ration. Mech. Anal.* 157(4):255–283, 2001.
- [18] D. Bao, S.-S. Chern, and Z. Shen. An introduction to Riemann-Finsler geometry. *Springer-Verlag, New York - Graduate Texts in Mathematics* 200:223–237, 2000.
- [19] J-D. Benamou. Optimal transportation, modelling and numerical simulation, *Acta Numer.* 22:133–212, 2021.
- [20] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. Asymptotic analysis for periodic structures, *Studies in Mathematics and its Applications -North-Holland Publishing Co., Amsterdam-New York*, 5:xxiv+700, 1978.
- [21] P. Bernard, and B. Buffoni. Optimal mass transportation and Mather theory, *J. Eur. Math. Soc. (JEMS)*, 9(1):85–121, 2007.
- [22] E. Bernton, P. Ghosal, and M. Nutz. Entropic optimal transport: Geometry and large deviations, *Duke Math. J.* 171(16):3363–3400, 2022.
- [23] D.P. Bertsekas. The auction algorithm: a distributed relaxation method for the assignment problem, *Ann. Oper. Res.* 14(1-4):105–123, 1988.
- [24] T. Bhattacharya, E. Di Benedetto, and J. Manfredi. Y. Limits as  $p \rightarrow \infty$  of  $\Delta_p u_p = f$  and related extremal problems, *Rend. Sem. Mat. Univ. Politec. Torino - Some topics in nonlinear PDEs (Turin, 1989)*, 15–68, 1989.
- [25] G. Bouchitté, C. Jimenez, and M. Rajesh. A new  $L^\infty$  estimate in optimal mass transport, *Proc. Amer. Math. Soc.*, 135(11):3525–3535, 2007.
- [26] D. J. Bregman, and D. Stroud. Physical Properties of Macroscopically Inhomogeneous Media, *Solid State Physics - Academic Press*, 46:147-269, 1992.
- [27] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs, *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.
- [28] H. Brezis. Functional analysis, Sobolev spaces and partial differential equations, *Springer, New York - Universitext* , xiv+599, 2011.
- [29] M. R. Bridson, and A. Haefliger. Metric spaces of non-positive curvature, *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, *Springer-Verlag, Berlin*, 319:xxii+643, 1999.
- [30] C. Brizzi. On functions with given boundary data and convex constraints on the gradient, *arXiv:2209.01462*, 2022.

- [31] C. Brizzi, G. Carlier, and L. De Pascale. Entropic approximations of  $L^\infty$ -optimal transport problems, *arXiv:2302.11896v1*, 2023.
- [32] C. Brizzi, C., and L. De Pascale. A property of absolute minimizers in  $L^\infty$  calculus of variations and of solutions of the Aronsson-Euler equation, *Adv. Differential Equations*, 28(3-4):287–310, 2023.
- [33] C. Brizzi, L. De Pascale, and A. Kausamo.  $L^\infty$ -optimal transport for a class of quasiconvex cost functions, *arXiv:2104.08074*, 2022.
- [34] R. A. Brualdi. Combinatorial matrix classes, *Encyclopedia of Mathematics and its Applications - Cambridge University Press, Cambridge*, 108:x+544, 2006.
- [35] G. Buttazzo. Semicontinuity, relaxation and integral representation in the calculus of variations, *Pitman Research Notes in Mathematics Series - ELongman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York*, 207:iv+2224, 1989.
- [36] F. Camilli and A. Siconolfi. Nonconvex degenerate Hamilton-Jacobi equations, *Math. Z.*, 242(1):1–21, 2002.
- [37] P. Cannarsa, and C. Sinestrari, C. Semiconcave functions, Hamilton-Jacobi equations, and optimal control. *Birkhäuser Boston, Inc., Boston, MA - Progress in Nonlinear Differential Equations and their Applications* 58, 2004.
- [38] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418,2017.
- [39] G. Carlier, P. Pegon, and L. Tamanini Convergence rate of general entropic optimal transport costs *arXiv:2206.03347* , 2022.
- [40] V. Caselles, J.-M. Morel, Jean-Michel, and C. Sbert. An axiomatic approach to image interpolation, *IEEE Trans. Image Process.*, 7(3):376–386, 1998.
- [41] T. Champion, and L. De Pascale. On the power law asymptotics for a model of dielectric breakdown, *Homogenization, 2001 (Naples) - GAKUTO Internat. Ser. Math. Sci. Appl. - Gakkōtoshō, Tokyo* , 18:231–235, 2003.
- [42] T. Champion, and L. De Pascale. Homogenization of Dirichlet problems with convex bounded constraints on the gradient, *Z. Anal. Anwendungen* , 22(3):591–608, 2003.
- [43] T. Champion, and L. De Pascale. Principles of comparison with distance functions for absolute minimizers, *J. Convex Anal.*, 14(3):515-541, 2007.
- [44] T. Champion, and L. De Pascale. The Monge problem for strictly convex norms in  $\mathbb{R}^d$ , *J. Eur. Math. Soc. (JEMS)*, 12(6):1355–1369, 2010.
- [45] T. Champion, and L. De Pascale. The Monge problem in  $\mathbb{R}^d$ , *Duke Math. J.*, 157(3):551–572, 2011.
- [46] T. Champion, and L. De Pascale. On the twist condition and  $c$ -monotone transport plans, *Discrete Contin. Dyn. Syst.*, 34 (4): 1339–1353 2014.

- 
- [47] T. Champion, L. De Pascale, and P. Juutinen. The  $\infty$ -Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.*, 40(1):1-20, 2008.
- [48] T. Champion, L. De Pascale, and F. Prinari.  $\Gamma$ -convergence and absolute minimizers for supremal functionals, *ESAIM Control Optim. Calc. Var.*, 10(1):14–27, 2004.
- [49] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard, and G. Peyré. Faster Wasserstein Distance Estimation with the Sinkhorn Divergence. *arXiv:2006.08172*, 2020.
- [50] G. Conforti, and L. Tamanini. A formula for the time derivative of the entropic cost and applications, *J. Funct. Anal.* 280(11), 2021.
- [51] M. G. Crandall. An efficient derivation of the Aronsson equation. *Arch. Ration. Mech. Anal.*, 167(4):271–279, 2003.
- [52] M. G. Crandall, and L. C. Evans. A remark on infinity harmonic functions. *Electron. J. Differ. Equ., Conf. 06*, pages 123–129, 2001.
- [53] M. G. Crandall, L. C. Evans, L. C. and R. F. Gariepy. Optimal Lipschitz extensions and the infinity Laplacian. *Calc. Var. Partial Differential Equations*13(2):123–139, 2001.
- [54] M. G. Crandall, G. Gunnarsson, and P. Wang. User’s guide to viscosity solutions of second order partial differential equations. *Comm. Partial Differential Equations*, 32(10):1587–1615, 2007.
- [55] M. G. Crandall, H. Ishii, and P.-L. Lions. Uniqueness of  $\infty$ -harmonic functions and the eikonal equation. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.
- [56] M. G. Crandall, C. Wang, and Y. Yu, Y. Derivation of the Aronsson equation for  $C^1$  Hamiltonians, *Trans. Amer. Math. Soc.* 36(1):103–124, 2009.
- [57] M. Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport, *Adv. in Neural Inf. Proc. Syst.*, 26, 2013.
- [58] M. Cuturi, and G. Peyré. Computational Optimal Transport, *arXiv:1803.00567*, 2020.
- [59] A. Davini. Finsler metrics in optimization problems and Hamilton-Jacobi equation, *University of Pisa, ph.d. dissertation in mathematics*, 2004.
- [60] G. De Cecco, and G. Palmieri, LIP manifolds: from metric to Finslerian structure. *Math. Z.* 218(2):223–237, 1995.
- [61] L. De Pascale, and J. Louet. A study of the dual problem of the one-dimensional  $L^\infty$ -optimal transport problem with applications, *C. J. Funct. Anal.*, 276(11): 3304–3324, 2019.
- [62] S. Di Marino, and A. Gerolin. An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm, *J. Sci. Comput.* 85(2), 2020.
- [63] G. Duvaut, and J.-L. Lions. Inequalities in mechanics and physics (Translated from the French by C. W. John), *Grundlehren der Mathematischen Wissenschaften - Springer-Verlag, Berlin-New York*, 219:xvi+397, 1976.

- [64] M. Erbar, and J. Maas, D. R. M. Renger. From large deviations to Wasserstein gradient flows in multiple dimensions, *Electron. Commun. Probab.* 20, no. 89, 12, 2015.
- [65] L. C. Evans, and O. Savin.  $C^{1,\alpha}$  regularity for infinity harmonic functions in two dimensions, *Calc. Var. Partial Differential Equations*, 32(3):325–347, 2008.
- [66] L. C. Evans, and Y. Yu. Various properties of solutions of the infinity-laplacian equation. *Communications in Partial Differential Equations*, 30(9):1401–1428, 2005.
- [67] A. Fathi, and A. Figalli. Optimal transportation on non-compact manifolds, *Israel J. Math.*, 175, 1–59, 2010.
- [68] J. Franklin, and J. Lorenz. On the scaling of multidimensional matrices, *Linear Algebra Appl.* 114/115 717–735, 1989.
- [69] W. Gangbo, and R. J. McCann. Optimal maps in Monge’s mass transport problem, *C. R. Acad. Sci. Paris Sér. I Math.*, 321(12) 1653–1658, 1995.
- [70] R. Gariepy, C. Wang, and Y. Yu. Generalized cone comparison principle for viscosity solutions of the Aronsson equation and absolute minimizers. *Comm. Par. Diff. Eq.* 31:1027–1046, 2006.
- [71] A. Garroni, V. Nesi, V. and M. Ponsiglione. Dielectric breakdown: optimal bounds. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* 457: 515–541, 2001.
- [72] A. Garroni, M. Ponsiglione, and F. Prinari. From 1-homogeneous supremal functionals to difference quotients: relaxation and  $\Gamma$ -convergence, *Calc. Var. Partial Differential Equations* 27(4):397–420, 2006.
- [73] M. S. Gelli, and F. Prinari. The role of intrinsic distances in the relaxation of  $L^\infty$ -functionals, *Nonlinear Anal.*, 204, 2021.
- [74] A. Henrot, and M. Pierre. Shape variation and optimization. European Mathematical Society (EMS), Zürich - EMS Tracts in Mathematics 87(1):19–35, 2018.
- [75] R. Jensen. Uniqueness of Lipschitz extensions: minimizing the sup norm of the gradient. *Arch. Rational Mech. Anal.*, 123(1):51–74, 1993.
- [76] R. Jensen, C. Wang, and Y. Yu. Uniqueness and nonuniqueness of viscosity solutions to Aronsson’s equation, *Arch. Ration. Mech. Anal.*, 190 (2):347–370, 2008.
- [77] R. Jordan, D. Kinderlehrer and F. Otto. The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.*, 29(1) 1–17, 1998.
- [78] C. Jimenez, and F. Santambrogio. Optimal transportation for a quadratic cost with convex constraints and applications, *J. Math. Pures Appl.*(9), 98(1):103–113, 2012
- [79] H. Jylhä. The  $L^\infty$  optimal transport: infinite cyclical monotonicity and the existence of optimal transport maps. *Calc. Var. Partial Differential Equations*, 52(1-2):303–326, 2015.
- [80] H. Jylhä, T. Rajala.  $L^\infty$  estimates in optimal mass transportation, *J. Funct. Anal.*, 270(11) 4297–4321, 2016.

- 
- [81] P. Juutinen. Absolutely minimizing Lipschitz extensions on a metric space. *Ann. Acad. Sci. Fenn. Math.*, 27(1):57–67, 2002.
- [82] K. L. Kantorovich. On mass transportation (Russian), *Dokl. Acad. Nauk, USSR*, 37:7–8, 1942.
- [83] N. Katzourakis.  $L^\infty$  variational problems for maps and the Aronsson PDE system, *J. Differential Equations*, 253(7):2123–2139, 2012.
- [84] N. Katzourakis.  $\infty$ -minimal submanifolds, *Proc. Amer. Math. Soc.*, 142(8):2797–2811, 2014.
- [85] N. Katzourakis. On the structure of  $\infty$ -harmonic maps, *Comm. Partial Differential Equations*, 39(11):2091–2124, 2014.
- [86] N. Katzourakis, T. Pryer. Second-order  $L^\infty$  variational problems and the  $\infty$ -polylaplacian, *Adv. in Calc. Var.*, 13(2):115–140, 2020.
- [87] N. Katzourakis, and G. Shaw. Counterexamples in Calculus of Variations in  $L^\infty$  through the vectorial Eikonal equation, *Comptes Rendus Mathematique*, 356(5):498–502, 2018.
- [88] H.W. Kuhn. The Hungarian method for the assignment problem, *Naval Res. Logist. Quart.* 2(1-2)83–97, 1955.
- [89] C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(4)1879–1920, 2012.
- [90] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4)1533–1574, 2014.
- [91] G. Leoni. A first course in Sobolev spaces. *American Mathematical Society, Providence, RI - Graduate Studies in Mathematics*, 105:xvi+607, 2009.
- [92] L. Lions. Generalized solutions of Hamilton-Jacobi equations, *Research Notes in Mathematics - Pitman (Advanced Publishing Program), Boston, Mass.-London*, 69:iv+317, 1982.
- [93] J. W. Macki, and A. Strauss. Introduction to optimal control theory. *Springer-Verlag, New York-Berlin-Undergraduate Texts in Mathematics*, 1982.
- [94] R. J. McCann. A convexity principle for interacting gases, *Adv. Math.*, 128(1):153–179, 1997.
- [95] R. J. McCann. Polar factorization of maps on Riemannian manifolds, *Geom. Funct. Anal.*, 11(3):589–608, 2001.
- [96] R. J. McCann. Stable rotating binary stars and fluid in a tube. *Houston J. Math.*, 32(2)603–631, 2006.
- [97] Q. Mérigot, B. Thibert. Optimal transport: discretization and algorithms, *Geometric partial differential equations. Part II, Handb. Numer. Anal.* 30:249–325, 2021.

- [98] Q. Miao, C. Wang, and Y. Zhou. Uniqueness of absolute minimizers for  $L^\infty$ -functionals involving Hamiltonians  $H(x, p)$ , *Arch. Ration. Mech. Anal.* 223(1):141–198, 2017.
- [99] G. Monge. Mémoire sur la théorie des déblais et des remblais, *Histoire de l'Académie Royale des Sciences de Paris*, 3666–704, 1781.
- [100] M. Nutz. Introduction to Entropic Optimal Transport. *Lecture notes*, 2021.
- [101] M. Nutz, and J. Wiesel. Stability of Schrödinger Potentials and Convergence of Sinkhorn's Algorithm, *arXiv:2201.10059*, 2022.
- [102] F. Otto. Doubly degenerate diffusion equations as steepest descent, *Manuscript*, 1996.
- [103] F. Otto. The geometry of dissipative evolution equations: the porous medium equation, *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
- [104] S. Pal. On the Difference between Entropic Cost and the Optimal Transport Cost, *arXiv:1905.12206*.
- [105] A. Pratelli. On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation, *Ann. Inst. H. Poincaré Probab. Statist.*, 43(1):1–13, 2007
- [106] A. Pratelli. On the sufficiency of  $c$ -cyclical monotonicity for optimality of transport plans, *Math. Z.*, 258(3):677–690, 2008.
- [107] F. Prinari. *Semicontinuity and supremal representation in the calculus of variations*. Appl. Math. Optim. **58**(1): 111–145, 2008.
- [108] F. Santambrogio. Optimal transport for applied mathematicians, *Birkhäuser/Springer, Cham - Progress in Nonlinear Differential Equations and their Applications* 87:xxvii+353, 2015.
- [109] O. Savin.  $C^1$  regularity for infinity harmonic functions in two dimensions. *Arch. Ration. Mech. Anal.*, 176(3):351–361, 2005.
- [110] R. Schneider. Convex bodies: the Brunn-Minkowski theory, *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge , 44:xiv+490, 1993.
- [111] E. Schrödinger. Über die umkehrung der naturgesetze. *Akademie der wissenschaften in kommission bei Walter de Gruyter u. Company*, 1931
- [112] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique., *Ann. Inst. H. Poincaré* 2(4):269–310, 1932.
- [113] A. Siconolfi. Metric character of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, 355(5):1987–2009, 2003.
- [114] R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums, *Amer. Math. Monthly*, 74:402–405, 1967.

- [115] V. N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions, *Proc. Steklov Inst. Math.*, (2):v+178, 1979.
- [116] T. W. Ting. Elastic-plastic torsion of convex cylindrical bars, *J. Math. Mech.* , 19:531–551, 1969/1970.
- [117] N. S. Trudinger, and X.-J. Wang. On the Monge mass transfer problem, *Calc. Var. Partial Differential Equations*, 13(1):19–31, 2001.
- [118] A. B. Tsybakov. Introduction to nonparametric estimation, *Springer, New York - Springer Series in Statistics*, xii+214, 2009.
- [119] S. R. S. Varadhan. Large deviations, *Courant Ins. of Math. Sci., New York; Am. Math. Soc., Providence, RI - Lecture Notes in Mathematics*, vii+104, 2016.
- [120] C. Villani. Optimal transport, Old and New. *Springer-Verlag, Berlin*, 2009.
- [121] C. Wang and Y. Yu.  $C^1$ -regularity of the aronsson equation in  $\mathbb{R}^2$ . *Annales de l'IHP Analyse non linéaire*, 25:659–678, 2008.
- [122] C. Wang and Y. Yu.  $C^1$ - boundary regularity of planar infinity harmonic functions. *Mathematical Research Letters*, 19(4):823–835, 2012.
- [123] Y. Yu. A remark on infinity-harmonic functions. *Electronic Journal of Differential Equations (EJDE)[electronic only]*, 2006.
- [124] Yu, Y.  $L^\infty$  variational problems and Aronsson equations, *Arch. Ration. Mech. Anal.*182(1):153–180, 2006.



## RÉSUMÉ

---

Dans cette thèse on étudie certaines propriétés des solutions de problèmes variationnels et de transport  $L^\infty$ . Ce manuscrit est divisé en deux parties.

La première partie, composée du Chapitre 2 et du Chapitre 3, traite d'un problème variationnel suprémal. Les problèmes variationnels suprémaux sont apparus pour la première fois à la fin des années 60 dans les travaux pionniers d'Aronsson [7, 8, 9]. En raison de la nature de la norme  $L^\infty$ , les minimiseurs intéressants sont les minimiseurs dits absolus (AM), qui sont souvent solutions d'une EDP associée et ont des propriétés d'unicité et de régularité. À la lumière de ces considérations, dans le Chapitre 2 nous analysons le problème associé à une fonctionnelle continue quasiconvexe  $(x, p) \mapsto H(x, p)$ . Nous montrons notamment une nouvelle propriété d'optimalité pour  $u \in \text{AM}$  et prouvons un résultat de structure pour l'ensemble des points  $x$  où  $H(x, Du(x)) = \max H(x, Du(x))$ .

Dans le Chapitre 3, nous resituons le problème variationnel dans le cadre des problèmes avec contraintes sur le gradient, en prouvant la régularité  $C^1$  des minimiseurs absolus sur l'ensemble mentionné ci-dessus.

Dans la deuxième partie, qui comprend le Chapitre 4 et le Chapitre 5, on s'intéresse au problème de transport optimal  $L^\infty$ , étudié pour la première fois par Champion, De Pascale, et Juutinen en 2007 [47]. Une contribution originale, présentée dans le Chapitre 4, est la preuve que les plans optimaux dits restreignables (*restrictable*) (l'analogue de AM) sont concentrés sur un graphe, si la fonction de coût est strictement quasiconvexe et satisfait une propriété similaire à la condition classique de *twist*. De plus, nous prouvons l'unicité dans le cas d'une mesure cible discrète.

Le problème de transport optimal  $L^\infty$  est non convexe, donc vraisemblablement plus complexe que le problème de transport classique. Afin d'avoir une meilleure compréhension, il semble naturel de chercher une généralisation à ce cadre de l'approximation entropique. Dans ce but, dans le Chapitre 5, nous introduisons une régularisation qui garantit la  $\Gamma$ -convergence vers le problème de transport  $L^\infty$ . En particulier, nous montrons que les minimiseurs des fonctionnelles régularisées sélectionnent des plans optimaux *restrictable*. Enfin, nous prouvons quelques estimations sur la vitesse de convergence et présentons quelques illustrations numériques réalisées avec l'algorithme de Sinkhorn.

## MOTS CLÉS

---

variationnel, transport optimal, minimiseurs absolus, supérieure essentielle, cycliquement monotone

## ABSTRACT

---

In this thesis we investigate some properties of solutions of  $L^\infty$ -variational and transport problems. This manuscript is divided into two parts.

The first part, made up of Chapter 2 and Chapter 3, deals with a supremal variational problem. Supremal variational problems appeared for the first time in the late 60s in the pioneering works of Aronsson [7, 8, 9]. Due to the nature of the  $L^\infty$ -norm, the interesting minimizers are the so-called absolute minimizers (AM), which often happen to be solutions of an associated PDE and to have uniqueness and regularity properties. In Chapter 2 we investigate the problem associated to a quasiconvex continuous supremand  $(x, p) \mapsto H(x, p)$ . Notably, we show a new optimality property for  $u \in \text{AM}$  and prove a structure result for the set of points  $x$  where  $H(x, Du(x)) = \max H(x, Du(x))$ . In Chapter 3 we insert the supremal variational problem in the framework of problems with constraints on the gradient, proving  $C^1$  regularity of the absolute minimizers on the above mentioned set.

In the second part, which consists of Chapter 4 and Chapter 5, we are interested in the  $L^\infty$ -optimal transport problem ( $L^\infty$ -OT), studied for the first time by Champion, De Pascale, and Juutinen in 2007 [47]. An original contribution, presented in Chapter 4, is the proof that the restrictable (optimal) plans (the analogous of AM) are concentrated on a graph, if the cost function is strictly quasiconvex and satisfies a property similar to the classical the twist condition. Moreover, we prove uniqueness in the case of a discrete target measure.

The  $L^\infty$ -OT is a non-convex problem, presumably more difficult than the classical OT. In order to have a better understanding, seeking for a generalization to this setting of the entropic approximation seems quite natural. With this intent, in Chapter 5, we provide a regularization which guarantees the  $\Gamma$ -convergence to the not regularized  $L^\infty$ -OT problem. Remarkably, we show that minimizers of the approximating functionals select restrictable optimal plans. Finally we prove some estimates on the speed of convergence and present some numerical illustrations performed with Sinkhorn's algorithm.

## KEYWORDS

---

variational, optimal transport, absolute minimizers, essential supremum, cyclically monotone