Inverse problems in non-imaging optics and generated jacobian equations

Anatole Gallouet

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et technologies de l'information, Informatique
Spécialité : Mathématiques Appliquées
Unité de recherche : Laboratoire Jean Kuntzmann

Problèmes inverses en optique anidolique et équations de jacobien généré

Inverse problems in non-imaging optics and generated jacobian equations

Présentée par :

Anatole GALLOUET

Direction de thèse :

Boris THIBERT PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Directeur de thèse
Quentin MERIGOT PROFESSEUR DES UNIVERSITES, UNIVERSITE PARIS-SACLAY	Co-directeur de thèse

Rapporteurs :

YOUNG-HEON KIM
PROFESSEUR, UNIVERSITY OF BRITISH COLUMBIA
JEAN-MARIE MIREBEAU
DIRECTEUR DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE SUD

Thèse soutenue publiquement le **18 octobre 2023**, devant le jury composé de :

BORIS THIBERT PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Directeur de thèse
JEAN-MARIE MIREBEAU DIRECTEUR DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE SUD	Rapporteur
QUENTIN MERIGOT PROFESSEUR DES UNIVERSITES, UNIVERSITE PARIS-SACLAY	Co-directeur de thèse
ALFRED GALICHON PROFESSEUR, NEW YORK UNIVERSITY PARIS	Examinateur
JULIE DIGNE CHARGEE DE RECHERCHE HDR, CNRS DELEGATION RHONE AUVERGNE	Examinatrice
VIRGINIE EHRLACHER INGENIEURE DES PONTS ET CHAUSSEES, ECOLE NATIONALE DES PONTS ET CHAUSSEES	Présidente

Optimal transport and Generated Jacobian equations, application to non-imaging optics.

Anatole Gallouët

November 16, 2023

Remerciements

Je voudrais d'abord remercier Boris et Quentin pour ces quatre années passées à travailler ensemble, pour leur apport sur le plan scientifique, mais également pour tous les moments passés ensemble, des bureaux d'Orsay à l'école de physique des Houches ;ces experiences m'ont beaucoup apporté, et je n'aurais pu rêver meilleurs directeurs de thèses.

Merci aux rapporteurs Jean-Marie Mirebeau et Young-Heon Kim pour avoir lu ce manuscrit et pour leurs retours pertinents, qui ont contribué à sa forme finale. Merci également aux membres du Jury, Virginie Ehrlacher, Alfred Galichon, Julie Digne et Emmanuel Maitre pour avoir écouté mon exposé et pour les échanges qui en ont suivi.

Merci à tous mes collègues de travail, à mes très chers co-bureau Hubert, Yu-Guan et Carlos. A Simon, Thibault, Hippolyte, Qiao, Mano, Sergei et Edouard pour nos séances de grimpe et les heures passées à en parler à la cafet. A Waïss, Eloi, Julien et Nils pour les discussions de maths plus ou moins pertinentes... A tous les doctorants et post-doctorants qui ont rendu mes pauses cafés bien plus agréables, Victor, Sélim, Margaux, Dima, Alexandre, Yannis, Sasila, Flo, Jean-Baptiste, Hélène, Manon, et plus globalement tous les membres du LJK pour ces quatre années passée ensemble.

Merci aussi à mes amis hors du cadre professionnel, Léo, Malek, Ulysse, Elie, Syssou, Louis, Jérem, Amandine et Marie pour les weekends et vacances passés ensemble. A mes compagons de ski Rémy, Bastien, Quentin, Pepe, Tom, Keke et Percy avec qui j'ai passé des moments inoubliables dans nos montagnes grenobloises. Et à tous ceux que je n'ai pas cités, pardonnez moi, je suis incapable de faire une liste exhaustive.

Finalement merci à ma famille, Manou, Thierry, Raphaèle, Emilie, Thomas, Capucine, Mathilde et Ysabel pour tout ce que vous m'avez apporté. Et à tous mes neveux et nièces pour l'animation des réunions familiales, peut-être qu'un jour ce sera votre tour...

Contents

Abstract

This thesis is motivated by non imaging optics problems. To begin with, we introduce the field of non-imaging optics. We see that some of the problems arising in this field can be recast as optimal transport problems or more generally as generated Jacobian equations (GJE). This work is divided in two parts.

The first part deals with the stability of solutions to optimal transport problems under variation of the measures, and is closely related to the convergence of numerical approaches to solve optimal transport problems and justifies many of the applications of optimal transport. We first introduce the notion of strong c-concavity, and we show that it plays an important role for proving stability results in optimal transport for general cost functions. We then introduce a differential criterion for proving that a function is strongly c-concave, under an hypothesis on the cost introduced originally by Ma-Trudinger-Wang for establishing regularity of optimal transport maps. Finally, we provide two examples where this stability result can be applied, for cost functions taking infinite value on the sphere: the reflector problem and the Gaussian curvature measure prescription problem.

The second part deals with Generated Jacobian equations, that have been introduced by Trudinger [Disc. cont. dyn. sys (2014), pp. 1663–1681] as a generalization of Monge-Ampère equations arising in optimal transport. We present and study a damped Newton algorithm for solving these equations in the *semi-discrete* setting, meaning that one of the two measures involved in the problem is finitely supported and the other one is absolutely continuous. We also present a numerical application of this algorithm to the near-field parallel reflector problem arising in non-imaging problems. Finally we also explore a method to approximate (GJE) using entropic regularization. We then present a stochastic algorithm to solve this approached problem.

Chapter 1

Introduction

The starting point for this thesis is non-imaging optics [70, 71, 50, 17, 32]. Non-imaging optics is a branch of optics that focuses on the design and the analysis of optical systems that do not rely on conventional imaging principles. Unlike traditional imaging systems, which aim to capture and reproduce precise images of objects, thus inducing a one-to-one mapping between source and target, non-imaging optics explores the manipulation and control of light for applications beyond imaging. In particular we are interested in the quantity of light transferred from the source to an area on the target but not where it finds its origin on the source. Non-imaging optics finds applications in various fields, for instance solar energy collection, where the goal is to redirect the light of the sun in the most efficient way possible to use its energy. It may also be used in lighting design and display technology for aesthetic purposes or practical ones such as public lighting. Industrial applications also exist, for example headlights for cars, trains and planes and fiber optics for optical communications.

Non-imaging optics. A non-imaging optics problem is an inverse problem. The direct problem is the following: given a light source represented by a measure μ on a topological space $\mathcal{X}$, and an optical component Σ (e.g. mirror or lens), compute by Snell's law the ray tracing map T_Σ associated to Σ. Then deduce the light distribution created by the reflection (or refraction) of the source $(\mathcal{X}, \mu)$ by Σ. This target distribution will also be represented by a measure ν on another topological space $\mathcal{Y}$, and we have that $\nu := T_{\Sigma\#}\mu$ is the image measure, or pushforward measure (see Def 1), of μ by the measurable map T_Σ. A draft representing the ray tracing map T_Σ is presented in Figure 1.

The data of the inverse problem is a light source $(\mathcal{X}, \mu)$ and a target distribution $(\mathcal{Y}, \nu)$, and the goal is to construct an optical component Σ that redirects the source toward the target. This means that we seek Σ satisfying $T_{\Sigma\#}\mu = \nu$. Assuming that the measures μ and ν have density ρ and σ, the equation $T_{\Sigma\#}\mu = \nu$ is equivalent to

$$\forall B \in \mathcal{T}(\mathcal{Y}), \int_{T_\Sigma^{-1}(B)} \rho(x)\mathrm{d}x = \int_B \sigma(y)\mathrm{d}y,$$

where $\mathcal{T}(\mathcal{Y})$ denotes the Borel set of $\mathcal{Y}$. Under some reasonable assumptions, a change of variable yields the following conservation equation

$$\forall x \in \mathcal{X}, \sigma(T_\Sigma(x)) \det(DT_\Sigma(x)) = \rho(x).$$

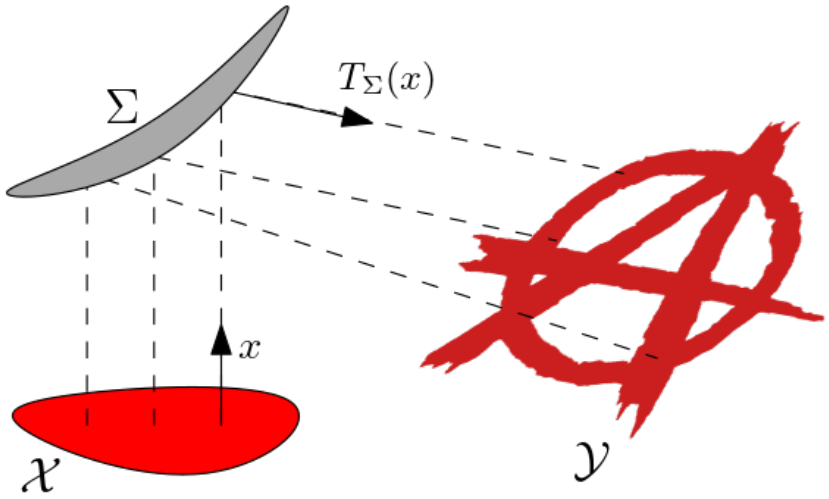

Figure 1: An example of non-imaging optics realization, the source $\mathcal{X}$ is transformed into a different image $\mathcal{Y}$ by a reflector Σ. Each ray x is reflected by Σ in the direction $T_\Sigma(x)$.

A Monge-Ampère type equation. Let us consider a particular non-imaging optics problem, the near-field parallel reflector (defined in equation (NF-par), in Section 2.4). In this problem, the source is collimated, meaning that all the rays are parallel going up. The intensity of the light emitted in each direction is represented by a measure μ on a domain $\mathcal{X} \subset \mathbb{R}^2$. The target is contained in an hyperplane of $\mathbb{R}^3$, meaning that we aim at points which are also represented by a measure ν on $\mathcal{Y} \subset \mathbb{R}^3$. The mirror $\Sigma := \Sigma_\varphi$ can be parametrized by the graph of a function $\varphi : \mathcal{X} \to \mathbb{R}$. Snell's law then gives a direct dependency between the ray tracing map $T_\varphi := T_{\Sigma_\varphi}$ and the normal vector to the mirror Σ. This normal vector is obtained using the first order derivative $\nabla\varphi$ of φ at each point $x \in \mathcal{X}$. One can then express $T_\varphi(x)$ by a function F of three variables: $T_\varphi(x) = F(x, \varphi(x), \nabla\varphi(x))$ (see for example [34] for details). Again if F is regular enough, the conservation equation thus becomes

$$\forall x \in \mathcal{X}, \quad \det\left(DT_\varphi(x)\right)) = \frac{\rho(x)}{\sigma(T_\varphi(x))}, \quad \text{with } T_\varphi(x) = F(x, \varphi(x), \nabla\varphi(x)).$$

Now assume that the derivative of F with respect to its third variable, denoted $\nabla_3 F$, is invertible. Then, multiplying the above equation by $\det(\nabla_3 F^{-1})$ gives the following Monge-Ampère type equation

$$\forall x \in \mathcal{X}, \quad \det(D^2\varphi(x) + g(x, \varphi(x), \nabla\varphi(x))) = f(x, \varphi(x), \nabla\varphi(x)),$$

where f and g are continuous functions. We used a non-imaging optics problem to find this equation, but it is actually very general and appears in many other problems. Among them optimal transport problems [68, 66] can be expressed as a Monge-Ampère equation. When $g = 0$, this equation corresponds to optimal transport in the quadratic case, i.e. for the cost $c(x, y) = \|x - y\|^2$. When the cost function is regular enough, other optimal transport problems can be written with $g(x, \varphi(x), \nabla\varphi(x)) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(x \mapsto c(x, T_\varphi(x))\right)$. The study of solutions to the Monge-Ampère equation in the optimal transport case has been a subject of interest for many years, and some results on the regularity of solutions have been obtained in this way, see for example [26, 51]. There exists a category of problems that are somewhat more general than optimal transport; these are called *generated*

Jacobian equations. They were introduced by Trudinger [66] and are sometimes named prescribed Jacobian equations. These equations may also be written as Monge-Ampère type equations, see [34, 66] for details. All the non-imaging optics problems we are interested in are either optimal transport or generated Jacobian, which implies that they are all Monge-Ampère equations. Due to the non linear terms in the Monge-Ampère equations, our approach is not focused on the PDE formulation but more on the geometry and calculus of variations. Note that generated Jacobian equations can be found in other fields than optics, for example economics [27], where they are mentioned as equilibrium matching problems.

Objectives of this thesis. This work focuses on the numerical computation of solutions to non-imaging optics problems. It is divided in two parts that are almost independent. In the first part, we are interested in the stability of solutions to optimal transport problems. These stability results are interesting from a numerical analysis point of view. For example, they guarantee the convergence of discrete solutions toward the continuous ones, thus validating the discretization techniques for numerical applications. In the second part, we focus on new algorithms to solve optimal transport and generated Jacobian equations. The idea is to adapt algorithms that already exist for optimal transport to our class of problems.

Numerical resolution. The numerical resolution of optimal transport problems has been trending in the last decade, resulting in substantial progress. The dynamical or Benamou-Brenier formulation opened the door to solvers for Lagrangian costs [8]. Other methods are based on finite differences [9] or finite volumes [18] schemes for the Monge-Ampère equation. The entropic regularization of optimal transport in the discrete case leads to solvers using the Sinkhorn algorithm [64]. Due to its simplicity, the Sinkhorn algorithm was widely used for numerical analysis, optimal transport and machine learning purposes [25, 28, 21, 60]. Entropic optimal transport can also be used in the semi-discrete setting, Genevay et. al [4] used it to develop a stochastic gradient descent. In the semi-discrete case, there also exists a Newton algorithm exploiting some computational geometry techniques [54]. Several other numerical methods exists, many of them are presented in the book of Cuturi and Peyré [60].

In Chapter 4 and 5 we extend these semi-discrete algorithms to generated Jacobian equations. We focus on the practical semi-discrete framework of [54], where we transport an absolutely continuous source measure toward a discrete target measure. This framework is known to be quite useful for non-usual cost functions, that appear often in non-imaging optics. The numerical resolution of generated Jacobian equations did not catch as much attention as optimal transport. An iterative algorithm has been proposed in [1] when the source measure is absolutely continuous and the target measure is discrete, which is also our framework. There also exists a least-squared minimization heuristic that has been proposed in the case of two absolutely continuous measures [62].

The damped Newton algorithm that was initially developed for optimal transport [54] is generalized to generated Jacobian equations (Chapter 4, published in [29]). However it is difficult to implement, this is why we also consider in Chapter 5 an entropic regularization to apply the stochastic algorithm of Genevay et. al [4] to generated Jacobian equations. The proof of convergence of this latter algorithm remains an open problem, but the convergence can be observed numerically on an example.

Stability of solutions. Many problems involve the comparison of point clouds between each other, for example in biology [38]. To this purpose, one can see these clouds as probability measures and can use the Wasserstein metric induced by optimal transport to compare them. Several solvers have thus been developed to solve efficiently discrete optimal transport problems. It may also happen that one wants to compare the distance between a point cloud and a density, or the distance between two densities. To do so, a natural idea would be to discretize the densities and compare the point clouds instead of the densities. The question is then, how good is the approximation ? From a numerical analysis point of view, stability is an answer to this question. Stability results allow to bound the distance between solutions by distances between the data, meaning that if the approximation of the measure is precise enough, the computed solution will be close to the exact one. This is why we study the stability of solutions to optimal transport problems with respect to small variations in the data. The statistical version of stability, aiming at approximating an optimal transport map between smooth density by an optimal map between sampled discrete measures, has been the object of several works, e.g. [39] and references therein.

There exist a few results of numerical analysis and stability in optimal transport, but only for the quadratic cost. The first stability result of optimal transport is due to Ambrosio and reported by Gigli [33]. Briefly, this result his a local 1/2-Holder behavior of the Monge embedding, which is the functional that associates an optimal transport map to a target measure. This stability result is thus applicable when the target measure is the only one that varies. Li and Nochetto [48] have a similar result for transport plans, but with respect to both source and target measures. Both these results are local, meaning that they are only true around "regular enough" solutions to optimal transport problems. Berman [11] has a global stability result that was extended by Mérigot and Delalande [24]. All these results are stated for the quadratic cost, i.e. for $c(x, y) = \|x - y\|^2$, and to the best of our knowledge there exists no stability results for optimal transport with other costs. The optimal transport problems arising from non-imaging optics are mostly stated with other cost functions, which lead to studying stability in a more general setup. In particular, this document contains a generalization of the local stability results of Ambrosio-Gigli [33] and Li-Nochetto [48] for other cost functions, provided that they satisfy a strong regularity assumption, namely the weak MTW hypothesis [51]. These results are submitted for publication [31].

After extending the stability to more general cost functions, a natural follow-up would be to check the stability for the generated Jacobian equation. This question is significantly more complicated due to the non-linear structure of the generated Jacobian equation. Even the question of uniqueness of the solutions, which is common knowledge in optimal transport (up to the addition of a constant to the potential), is not completely clear for generated Jacobian equations. There exist some partial uniqueness results in different setting, by Rankin [61] and Gutierrez–Huang [36]. A stability result that would allow to understand quantitatively the convergence of discrete solutions towards continuous ones is currently out of reach.

Content of the thesis

This work is divided in two parts; the first part presented in Chapter 2 and 3 is about optimal transport problems. In particular we study the stability of solution to optimal transport problems with respect to the data. This result is of importance because, as

already mentioned, it guarantees that approximating the measures that are transported yields reasonable solutions. The second part is contained in Chapter 4 and 5, it is dedicated to the numerical resolution of generated Jacobian equations, for which the literature is not as extensive as for optimal transport. The content of each chapter is summarized thereafter.

Chapter 2 is an introduction to the optimal transport theory and applications to non-imaging optics. In this chapter we start by introducing the Monge problem which consists in finding a transport map between $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(\mathcal{Y})$ realizing the following infimum

$$(\text{MP}) := \inf_{T_{\#}\mu=\nu} \int_{\mathcal{X}} c(x, T(x)) d\mu(x).$$

where $T_{\#}\mu$ is the pushforward measure of μ by T (see Def 1). We also introduce the Kantorovich relaxation where the map $T : \mathcal{X} \to \mathcal{Y}$ is replaced by a probability measure on the product space $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with constraint on its marginals

$$(\text{KP}) := \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) \mathrm{d}\gamma(x, y)$$

These are the two fundamental formulations of optimal transport problems. We then derive the Kantorovich dual formulation of the problem that leads to the introduction of Kantorovich potentials. These potential functions are fundamental throughout this work, as they allow to introduce a generalization of convex analysis notions to more global cost functions c. Roughly speaking they allow to generalize the notion of convexity to c-convexity (or c-concavity depending on conventions), where the scalar product $\langle \cdot | \cdot \rangle$ is replaced by the real valued cost function $c(\cdot, \cdot)$. The c-concavity of potentials is highly useful to solve optimal transport problems and study their solutions, it is the starting point of the next chapter to derive stability of transport maps with respect to the measures of the problem. An important remark is that from an optimal potential function ψ, one can reconstruct an optimal transport map T_ψ solution of (MP). We then introduce the semi-discrete setting of optimal transport, which is the framework that is considered here for the numerical resolution of both optimal transport and generated Jacobian equations throughout the document. In this framework, the space $\mathcal{X}$ is a domain while the space $\mathcal{Y}$ is finite of size N. This setting allows to work on a tessellation of the source space $\mathcal{X}$ composed of N *Laguerre cells* defined for each $y \in \mathcal{Y}$ by

$$\mathrm{Lag}_y(\psi) = \{x \in \mathcal{X} \mid \forall z \in \mathcal{Y}, c(x, y) + \psi(y) \leqslant c(x, z) + \psi(z)\}.$$

where $\psi : \mathcal{Y} \to \mathbb{R}$ is the Kantorovich potential. Under some assumptions on the cost function, the Laguerre cells are uniquely defined almost everywhere. In this case, the transport maps associated to ψ is defined by $\mathrm{T}_\psi(x) = y \iff x \in \mathrm{Lag}_y(\psi)$. If we enumerate $\mathcal{Y} = \{y_i\}_{1\leqslant i\leqslant N}$ and identify $\mathcal{Y} \to \mathbb{R}$ with $\mathbb{R}^N$ we can then define the mass function $H : \mathbb{R}^N \to \mathbb{R}^N$ by $H(\psi) = \big(\mu(\mathrm{Lag}_{y_i}(\psi))\big)_{1\leqslant i\leqslant N}$, and the semi-discrete optimal transport problem then amounts to the *mass prescription problem* of finding $\psi \in \mathbb{R}^N$ such that

$$H(\psi) = \nu.$$

The last section of this chapter is dedicated to non-imaging optics, including a global presentation of the field and some specific problems. We show that these problems in the semi-discrete setting have the same mass prescription formulation as above. Some non-imaging optics problems, in the "far-field" case, can then be expressed as the dual

formulation of an optimal transport problem. Other optics problems, in the "near-field" case, do not yield optimal transport equations but slightly more general ones, which happen to be generated Jacobian equations.

Chapter 3 contains stability results of optimal transport maps with respect to the measures. An important result of this chapter is the stability of optimal transport maps under some regularity assumptions mainly on the cost function c. The starting point is a result from Ambrosio and Gigli [33] recalled in Theorem 11. Formally, Ambrosio and Gigli showed that if we consider a transport map for which the Brenier potential is strongly convex, then the L^2 distance between this map and another one with a different target measure can be bounded by the Wasserstein distance between target measures, i.e.

$$\|T_{\mu\to\nu_0} - T_{\mu\to\nu_1}\|^2_{L^2(\mu)} \leqslant CW_1(\nu_0,\nu_1)$$

This result is true when $T_{\mu\to\nu}$ is an optimal transport map for the quadratic cost $c(x,y) = \|x-y\|^2$. In order to generalize this result to other costs, it was natural to extend the notion of c-concavity to strong c-concavity. This notion is a generalization in the sense that strong c-concavity is similar to strong concavity, replacing the supporting hyperplane of the concave function by levelsets of the cost function. To define this notion, we recall the notion of c-superdifferential. Let $\psi : \mathcal{Y} \to \mathbb{R}$, the c-superdifferential of ψ at a point $y \in \mathcal{Y}$ is defined by

$$\partial^c\psi(y) = \{x \in \mathcal{X} \mid \forall z \in \mathcal{Y}, \psi(z) - c(x,z) \leqslant \psi(y) - c(x,y)\}$$

Note that ψ is c-concave iff for any $y \in N$ its c-superdifferential $\partial^c\psi(y)$ is non-empty. If this is the case, then ψ is strongly c-concave with modulus ω if for all $y, z \in \mathcal{Y}$ and $x \in \partial^c\psi(y)$

$$\psi(z) - c(x,z) \leqslant \psi(y) - c(x,y) - \omega(\mathrm{d}(y,z))$$

where d is a distance on $\mathcal{Y}$. Once we introduced strong c-concavity, an important theorem can be summarized as follow.

Theorem 14 (Strong c-concavity implies stability)**.** *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu_0, \nu_1 \in \mathcal{P}(\mathcal{Y})$. We assume that the cost c is regular enough, and that there exists optimal transport maps T_i from μ to ν_i with associated potential $\psi_i : N \to \mathbb{R}$ $(i = 0, 1)$ such that:*

- *ψ_0 is Lipschitz on $\mathcal{Y}$ and c-concave.*
- *ψ_1 is Lipschitz on $\mathcal{Y}$ and strongly c-concave with modulus ω.*

Then,

$$\int_{\mathcal{X}} \omega(\mathrm{d}(T_0(x), T_1(x)))\mathrm{d}\mu(x) \leqslant (\mathrm{Lip}(\psi_0) + \mathrm{Lip}(\psi_1))W_1(\nu_0,\nu_1)$$

where W_1 is the 1-Wasserstein distance.

The full hypotheses are stated in Section 3.2. The chapter also treats other stability results for general cost functions, notably for transport plans with respect to both measures, as stated in the following proposition.

Proposition 19 (Stability with respect to both measures)**.** *Let $\mu, \tilde{\mu} \in \mathcal{P}(\mathcal{X})$ and $\nu, \tilde{\nu} \in \mathcal{P}(\mathcal{Y})$. Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a cost function which is Lipschitz on $\mathcal{X} \times \mathcal{Y}$. Let $T : \mathcal{X} \to \mathcal{Y}$ be an optimal transport map between μ and ν, and $\tilde{\gamma}$ be an optimal transport plan between*

$\tilde{\mu}$ and $\tilde{\nu}$ for the cost c. We assume that T is induced by a strongly c-concave potential $\psi : \mathcal{Y} \to \mathbb{R}$ with associated modulus $\omega(r) = Cr^2$. Then we have

$$\mathrm{W}_1(\gamma_T, \tilde{\gamma}) \leqslant \varepsilon + \sqrt{\frac{2\mathrm{Lip}(c)}{C}\varepsilon}, \quad \textit{where } \varepsilon := \mathrm{W}_1(\tilde{\mu}, \mu) + \mathrm{W}_1(\nu, \tilde{\nu})$$

where $\gamma_T = (\mathrm{Id}, T)_{\#}\mu$ is the transport plan induced by the map T.

The main issue of this result is that it requires the cost function to be Lipschitz on the whole product space $\mathcal{X} \times \mathcal{Y}$, which is often not satisfied. All these results are given for a strongly c-concave potential. The most technical part of the chapter is contained in Section 3.3. It gives a sufficient condition for strong c-concavity, under the Ma-Trudinger-Wang hypothesis on the cost function [51]. This hypothesis is not surprising as it is necessary to obtain the regularity of optimal transport maps [69], and in fact, strong c-concavity implies some regularity of the associated transport map. The result is a differential criterion for strong c-concavity. It is an adaptation of Villani's criterion for usual c-concavity [69, Th. 12.46]. Here we can work on a subset $D \subset \mathcal{X} \times \mathcal{Y}$ on which the cost function is regular and the MTW tensor $\mathfrak{S}_c$ (Def 20) is positive.

Theorem 24 (Characterization of strong c-concavity)**.** *Under some assumptions on $D \subseteq \mathcal{X} \times \mathcal{Y}$ (detailed later). We assume that the MTW hypothesis is satisfied on D. Let $\psi \in \mathcal{C}^2(\mathcal{Y}, \mathbb{R})$ be a c-concave function on D and such that there exists $\lambda > 0$ satisfying for any $x \in \partial^c\psi(y)$*

$$D^2_{yy}c(x, y) - D^2\psi(y) \geqslant \lambda Id$$

Then ψ is strongly c-concave on D with modulus $\omega(\mathrm{d}_N(\overline{y}, y)) = C\mathrm{d}_N(\overline{y}, y)^2$, where $C > 0$ is a constant depending on c, $\mathcal{X}$ and $\mathcal{Y}$. This means that we have

$$\psi(y) - c(\overline{x}, y) \leqslant \psi(\overline{y}) - c(\overline{x}, \overline{y}) - C\mathrm{d}_N(\overline{y}, y)^2$$

for the points $\overline{x} \in \mathcal{X}$, $\overline{y}, y \in \mathcal{Y}$ such that $\overline{x} \in \partial^c\psi(\overline{y})$, $(\overline{x}, \overline{y}) \in D$ and $(\overline{x}, y) \in D$.

From this theorem we can deduce the strong c-concavity of potentials associated to regular enough transport maps for MTW costs. This is stated in Corollary 25.

We finally show that the strong c-concavity of potentials and the stability results that follow can be applied to two optimal transport problems for different cost functions on the sphere. One of the applications is the far-field point reflector problem which is presented in Section 2.4. This problem is equivalent to an optimal transport problem on the sphere for the cost $c(x, y) = -\ln(1 - \langle x|y\rangle)$. The stability result can be stated as follows, with $M_\nu(\beta)$ being the maximum mass given by the measure ν on all balls of radius β.

Theorem 31 (Stability for the reflector cost)**.** *Let $c(x, y) = -\ln(1-\langle x|y\rangle)$ be the reflector cost on the sphere $M = \mathcal{S}^{d-1}$. Let $\mu, \nu_0 \in \mathcal{P}(M)$ be absolutely continuous with respect to the Lebesgue measure with strictly positive $\mathcal{C}^{1,1}$ densities. Then for all $\beta > 0$, there exists a constant $C > 0$ depending on μ, ν_0 and β such that*

$$\forall \nu_1 \in \mathcal{P}(M) \text{ s.t. } M_{\nu_1}(\beta) < 1/8, \quad \|\mathrm{d}_M(T_0, T_1)\|^2_{L^2(\mu)} \leqslant C\ W_1(\nu_0, \nu_1)$$

where d_M is the geodesic distance on M and T_i be optimal transport maps between μ and ν_i.

This result guarantees that computing a solution by discretizing the target measure yields a correct approximation of the reflection that would be expected in the continuous

case. It is not trivial because the reflector cost $c(x,y) = -\ln(1-\langle x|y\rangle)$ explodes when $x = y$, which makes it non differentiable on the diagonal. This issue is tackled in Section 3.5, the main argument is to show that the support of the optimal transport plan stays far from the diagonal because it is repulsive, i.e. $c(x,x) = +\infty$.

Chapter 4 is the main chapter on generated Jacobian equations. These equations are Monge-Ampère type equations that can be written as a highly non linear PDE, but we will only consider here the semi-discrete setting. This choice allows to present the problem directly as a mass prescription equation, as it is the case in optimal transport. The difference is that here, instead of working with a cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, we work with a scalar *generating function* $G : \mathcal{X} \times \mathcal{Y} \times \mathbb{R} \to \mathbb{R}$ which gives "generalized" Laguerre cells of the form

$$\mathrm{Lag}_y(\psi) = \{x \in \mathcal{X} \mid \forall z \in \mathcal{Y}, G(x,y,\psi(y)) \geqslant G(x,z,\psi(z))\}$$

where again, $\psi : \mathcal{Y} \to \mathbb{R}$ is a scalar function. We can thus define the mass function $H : \mathbb{R}^N \to \mathbb{R}^N$ in the same way, by $H(\psi) = \mu(\mathrm{Lag}_i(\psi))_{1\leqslant i\leqslant N}$, and the semi-discrete generated Jacobian equation amounts to finding $\psi : \mathcal{Y} \to \mathbb{R}$ such that

$$H(\psi) = \nu.$$

This equation amounts to finding the zero of a function, and one can consider trying to solve it using a Newton algorithm, as it has been done in optimal transport [54]. The advantage of optimal transport is that it can be shown that the map H is the gradient of a concave function, thus implying that its differential DH is actually a Hessian which makes it symmetric and negative semi-definite. Moreover it is easy to identify its kernel and image. For a generated Jacobian equation, it is not as easy, the Jacobian matrix DH is not symmetric in general; it is also more difficult to identify its kernel, but one can still deduce its structure. The complete study of this differential DH is presented in the chapter. These results allow to show that by using a correct damping parameter, the Newton algorithm converges in the case of generated Jacobian equations, as it is the case in optimal transport. The main theorem is the following.

Theorem 58 (Convergence of the Newton algorithm)**.** *Assume that the generating function is regular enough, that $\mathcal{X}$ is compact, and that the sets $\mathcal{Y}$ and $\partial\mathcal{X}$ are generic (see definition in Chapter 4). Then, there exists $\tau^* \in]0,1]$ such that the iterates of Algorithm 1 satisfy*

$$\|H(\psi^k) - \nu\| \leqslant \left(1 - \frac{\tau^*}{2}\right)^k \|H(\psi^0) - \nu\|.$$

In particular, Algorithm 1 converges.

The end of the chapter is dedicated to the implementation of the algorithm on an non-imaging optics problem, the near field parallel reflector problem (NF-par).

Chapter 5 is an ongoing work about the entropic regularization of generated Jacobian equations. Entropic regularization has been widely used in the numerical optimal transport community to obtain fast algorithms that are easy to implement [60, 21, 4]. We focus here on the semi-discrete entropic regularization of optimal transport presented by Genevay et. al [4]. The dual of the regularized semi-discrete problem can be expressed as a mass prescription problem just like regular optimal transport. The difference is that the Laguerre cells are replaced by "smoothed" Laguerre cells. The generated Jacobian equation can also be expressed as mass prescription of Laguerre cells, even though there

is no variational formulation. It is thus quite natural to try to adapt the smoothing of Laguerre cells in optimal transport to generated Jacobian equations. The regularization of the semi-discrete formulation leads to a stochastic gradient descent to solve regularized optimal transport. Formally, regularizing the semi-discrete problem amounts to replace the maximum by a softmax in the definition of Laguerre cells, with a scalar parameter $\varepsilon > 0$. We thus obtain "smoothed Laguerre cells" that are regular functions that form a partition of unity, and converge to the usual Laguerre cells when $\varepsilon \to 0$. Because of its formulation, the regularization easily adapts to the generated Jacobian equations. In this case, the smoothed Laguerre cells are defined for $\psi \in \mathbb{R}^N$ and $x \in \mathcal{X}$ by

$$\mathcal{L}_{\varepsilon,i}[\psi](x) = \frac{e^{G(x,y_i,\psi_i)/\varepsilon}}{\sum_k e^{G(x,y_k,\psi_k)/\varepsilon}}$$

and the mass function associated to cell $i \in [\![1, N]\!]$ is

$$H_i^\varepsilon(\psi) = \int_X \mathcal{L}_{\varepsilon,i}[\psi](x)\mathrm{d}\rho(x)$$

The stochastic gradient descent of [4] can be adapted into a stochastic fixed point to solve an approximation of the mass prescription equation $H^\varepsilon(\psi) - \nu = 0$. The fact that the mass function H is not a gradient makes the analysis of the algorithm harder, and the theoretical convergence is still an open problem. The algorithm was still tested on an example of generated Jacobian equation for which it converges.

Chapter 2

Optimal transport and applications to optics

Optimal transport theory [69, 63, 7] was introduced by Monge in 1781 with the objective of finding the most efficient way to transport mass from one location to another, given certain constraints and costs. This theory has had an important impact in applied mathematics, with a wide range of applications in various fields [14, 60, 12, 55], including machine learning, economics, physics, computer vision, image processing and optics. In economics, optimal transport can be used to analyze and optimize the flow of goods and services between regions or markets, while in physics, it can be used to model the movement of fluids and particles. In computer vision, optimal transport can be used to match images and estimate correspondences between different sets of data. In image processing, it can be used for tasks such as image registration and morphing. The applications we are interested in here are non-imaging optics, where the goal is to create an optical component (e.g. mirror or lens) that redirects a given light source toward a prescribed target distribution, without enforcing to have a one-to-one mapping between source and target.

2.1 Monge-Kantorovich problem and Wasserstein distances

The Monge problem consists in finding an optimal transportation map that assigns each item in the source location to a unique location in the destination while minimizing the total transportation cost. This problem can be challenging to solve, especially when dealing with large datasets. Kantorovich relaxation is a more flexible approach that allows for partial assignments between the source and destination locations, making it easier to find solutions to the optimal transport problem. We represent the source and the destination by Polish (separable completely metrizable) spaces $\mathcal{X}$ and $\mathcal{Y}$. We denote by $\mathcal{M}(A)$ and $\mathcal{P}(A)$ respectively the sets of signed measures and probability measures on a set A. The goal is to transport a mass represented by a measure $\mu \in \mathcal{P}(\mathcal{X})$ toward a measure $\nu \in \mathcal{P}(\mathcal{Y})$. The cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a lower semi-continuous function that is bounded from below such that $c(x, y)$ is the cost of transporting the point x toward y.

Definition 1 (Push-forward measure and transport map)**.** *Let $\mu \in \mathcal{M}(\mathcal{X})$ and $T : \mathcal{X} \to \mathcal{Y}$.*

- *The push-forward measure of μ by T is a measure on $\mathcal{Y}$ denoted by $\mathrm{T}_{\#}\mu \in \mathcal{M}(\mathcal{Y})$*

and defined by

$$\forall B \subset \mathcal{Y}, \quad T_{\#}\mu(B) = \mu(T^{-1}(B)).$$

- *A* transport map *between $\mu \in \mathcal{M}(\mathcal{X})$ and $\nu \in \mathcal{M}(\mathcal{Y})$ is a measurable map $T : \mathcal{X} \to \mathcal{Y}$ such that the image measure $T_{\#}\mu$ equals ν.*

Note that for a transport map to exists, the total mass needs to be the same for both measures, i.e. $\mu(\mathcal{X}) = \nu(\mathcal{Y})$. One can then normalize, making the transport maps being always defined between probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$.

Definition 2 (The Monge problem)**.** *Monge's optimal transport problem between $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ for the cost $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ amounts to finding a map $T : \mathcal{X} \to \mathcal{Y}$ that realizes the following infimum*

$$(\text{MP}) := \inf_{T_{\#}\mu=\nu} \int_{\mathcal{X}} c(x, T(x)) d\mu(x). \tag{2.1.1}$$

As mentioned earlier, the Monge problem (MP) is quite hard to solve. Kantorovich introduced a relaxed formulation of the problem [41], in which the mass emanating from a point $x \in \mathcal{X}$ is allowed to split and reach several positions in $\mathcal{Y}$. Thus, instead of minimizing over transport maps, one minimizes over transport plans between μ and ν, which we define hereafter. We first introduce the notion of marginal.

Definition 3 (Marginal)**.** *For a measure $\gamma \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$, the marginals $\Pi_x\#\gamma \in \mathcal{M}(\mathcal{X})$ and $\Pi_y\#\gamma \in \mathcal{M}(\mathcal{Y})$ are defined for every measurable sets $A \subset \mathcal{X}$ and $B \subset \mathcal{Y}$ by*

$$\Pi_x\#\gamma(A) = \gamma(A \times \mathcal{Y}), \quad \Pi_y\#\gamma(B) = \gamma(\mathcal{X} \times B)$$

A transport plan between two measures is a measure on the product space with constrained marginals.

Definition 4 (Transport plan)**.** *A transport plan between $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ is a probability measure $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ whose marginals are μ and ν. The set of all transport plans between μ and ν is denoted $\Gamma(\mu, \nu)$ i.e.*

$$\Gamma(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid \forall A \subset \mathcal{X}, \gamma(A \times \mathcal{Y}) = \mu(A), \forall B \subset \mathcal{Y}, \gamma(\mathcal{X} \times B) = \nu(B)\}$$

For a transport plan γ, if $A \times B \subset \mathcal{X} \times \mathcal{Y}$, then $\gamma(A \times B)$ represents the mass transported from A to B.

Definition 5. *The relaxed problem introduced by Kantorovich is*

$$(\text{KP}) := \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) \mathrm{d}\gamma(x, y) \tag{2.1.2}$$

It is clear that this problem is weaker than the Monge problem, in the sense that (KP) $\leqslant$ (MP). This can be seen by choosing a transport plan γ that never splits the mass of a point x. In this case, the transport plan $\gamma \in \Gamma(\mu, \nu)$ is said to be induced by a transport map T, it is given by the formula $\gamma = (\mathrm{Id}, T)_{\#}\mu$. There exist several situations where one can obtain equality between Monge and Kantorovich problems. It is actually the case whenever the minimizing transport plan is induced by a map (or a limit of such plans). We give below one general framework where there is equality.

Proposition 1 (Kantorovich as a relaxation of Monge)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be compact subsets of $\mathbb{R}^d$ and $c \in \mathcal{C}^0(\mathcal{X} \times \mathcal{Y})$. Let $\mu \in \mathcal{P}(\mathcal{X})$ be atomless and $\nu \in \mathcal{P}(\mathcal{Y})$ any probability measure, then* (KP) = (MP)*.*

A proof of this equality can be found in [63]. When $\mathcal{X}$ and $\mathcal{Y}$ live in the same Polish space Ω, the Kantorovich problem can be used to define a distance between probability measures. We denote by d the distance on Ω.

Definition 6 (Wasserstein distances)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be subsets of the same metric space (Ω, d). Let $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and $c(x, y) = d^p(x, y)$ for some $p \geqslant 1$. Then the p-Wasserstein distance is given by*

$$\mathrm{W}_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} d^p(x, y) \mathrm{d}\gamma(x, y) \right)^{1/p}$$

The Wasserstein distance is sometimes called the Earth Mover's distance as it measures the cost of moving the mass from μ to ν for the distance d. This distance has several advantages over other distance metrics, including its ability to capture structural differences between distributions and its robustness to outliers. It has been broadly studied and is used in a wide field of applications including machine learning, image processing or statistics.

2.2 Dual formulation and c-concavity

The Kantorovich Problem has a dual formulation that can be derived using Lagrange multipliers for the marginals constraints on the transport plan. In many cases, the dual problem can be solved more efficiently than the primal problem, yielding a practical alternative for solving optimal transport problems. It also allows to obtain a different formulation for the 1-Wasserstein distance W_1 which is called the Kantorovich-Rubinstein metric. It is detailed in Theorem 4. Finally the dual formulation leads to the notion of c-transform which can be seen as a generalization of the Legendre transform. One can then be inspired by convex analysis to introduce c-concave functions which are quite useful in optimal transport theory for general cost functions. We will use these notions in Chapter 3 to study the regularity and the stability of solutions to optimal transport problems. In this section we give a formal derivation of the dual problem based on [54]. A complete proof of Kantorovich duality can be found in [68]. Unless specified otherwise, we will consider for simplicity that the cost function is continuous on $\mathcal{X} \times \mathcal{Y}$. We denote by $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$ the set of positive measures on $\mathcal{X} \times \mathcal{Y}$. For any functions $\varphi : \mathcal{X} \to \mathbb{R}$ and $\psi : \mathcal{Y} \to \mathbb{R}$ we define the functions $(\varphi \oplus \psi)(x, y) = \varphi(x) + \psi(y)$ and $(\varphi \otimes \psi)(x, y) = \varphi(x)\psi(y)$. To lighten the notation, we denote by $\langle \varphi | \mu \rangle$ the integral $\int \varphi \mathrm{d}\mu$, and similarly for others functions and measures. Then one has

$$\sup_{\varphi \in \mathcal{C}^0(\mathcal{X})} \langle \varphi | \mu \rangle - \langle \varphi \otimes 1 | \gamma \rangle = \begin{cases} 0 & \text{if } \Pi_x \# \gamma = \mu \\ +\infty & \text{otherwise.} \end{cases}$$

and similarly

$$\sup_{\psi \in \mathcal{C}^0(\mathcal{Y})} \langle \psi | \nu \rangle - \langle 1 \otimes \psi | \gamma \rangle = \begin{cases} 0 & \text{if } \Pi_y \# \gamma = \nu \\ +\infty & \text{otherwise.} \end{cases}$$

where $\Pi_x \# \gamma \in \mathcal{M}(\mathcal{X})$ and $\Pi_y \# \gamma \in \mathcal{M}(\mathcal{Y})$ denotes the marginals of γ. Combining these two equalities gives

$$\sup_{\varphi \in \mathcal{C}^0(\mathcal{X}), \psi \in \mathcal{C}^0(\mathcal{Y})} \langle \varphi | \mu \rangle + \langle \psi | \nu \rangle - \langle (\varphi \oplus \psi) | \gamma \rangle = \begin{cases} 0 \text{ if } \gamma \in \Gamma(\mu, \nu) \\ +\infty \text{ otherwise.} \end{cases}$$

Using Lagrange multipliers φ and ψ for the marginal constraints, the Kantorovich problem then writes

$$(\mathrm{KP}) = \inf_{\gamma \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})} \sup_{\varphi \in \mathcal{C}^0(\mathcal{X}), \psi \in \mathcal{C}^0(\mathcal{Y})} \langle \varphi | \mu \rangle + \langle \psi | \nu \rangle + \langle c - (\varphi \oplus \psi) | \gamma \rangle.$$

The dual problem is then defined by inverting supremum and infimum, this gives

$$(\mathrm{DP}) := \sup_{\varphi, \psi} \inf_{\gamma \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})} \langle c - (\varphi \oplus \psi) | \gamma \rangle + \langle \varphi | \mu \rangle + \langle \psi | \nu \rangle.$$

Note that at this point, there is no reason to have (KP) = (DP), but one has always (KP) $\geqslant$ (DP). One can then simplify this dual formulation by remarking that

$$\inf_{\gamma \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})} \langle c - (\varphi \oplus \psi) | \gamma \rangle = \begin{cases} 0 \text{ if } \varphi \oplus \psi \leqslant c \\ -\infty \text{ otherwise.} \end{cases}$$

This allows to replace the infimum over γ by a condition on the functions φ and ψ.

Definition 7 (Kantorovich dual problem)**.** *Let $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, the dual formulation of* (KP) *is*

$$(\mathrm{DP}) = \sup_{\varphi \in \mathcal{C}^0(\mathcal{X}), \psi \in \mathcal{C}^0(\mathcal{Y}) | \varphi \oplus \psi \leqslant c} \langle \varphi | \mu \rangle + \langle \psi | \nu \rangle \tag{2.2.3}$$

The functions φ and ψ are called Kantorovich potentials. The global theory is called Kantorovich duality, by the name of the Russian mathematician Leonid Kantorovich who first introduced it in 1942 [41]. Under some mild assumptions strong duality holds.

Theorem 2 (Strong Kantorovich duality)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ Polish spaces. If the cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is l.s.c. and bounded below, then* (KP) = (DP) *and* (KP) *is a minimum, i.e. the infimum in* (KP) *is reached for some transport plan $\gamma \in \Gamma(\mu, \nu)$.*

One proof is given by Villani in his first book [68]. We can go further in the simplification of the dual problem, by remarking that the condition $\varphi \oplus \psi \leqslant c$ gives a strong relation between φ and ψ. Let us fix a function $\psi \in \mathcal{C}^0(\mathcal{Y})$. Since $\varphi \oplus \psi \leqslant c$, we have for any $y \in \mathcal{Y}$, $\varphi(x) \leqslant c(x, y) - \psi(y)$. The measures being positive, we want to maximize φ. The best φ possible is then the infimum over all $y \in \mathcal{Y}$ of $c(\cdot, y) - \psi(y)$, this quantity depending on c and ψ is called the c-transform of ψ.

Definition 8 (c-transform)**.** *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and $\psi : \mathcal{Y} \to \mathbb{R}$, the c-transform of ψ is a function $\psi^c : \mathcal{X} \to \mathbb{R}$ defined for $x \in \mathcal{X}$ by*

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

By symmetry, for $\varphi : \mathcal{X} \to \mathbb{R}$, $\varphi^c(y) = \inf_x c(x, y) - \varphi(x)$. When we have both $\varphi = \psi^c$ and $\psi = \varphi^c$ we say that φ and ψ are c-conjugate.

Remark 3. *When $\mathcal{X} = \mathcal{Y}$ lives in an Hilbert space and $c(x, y) = -\langle x|y\rangle$, the c-transform almost coincides with the convex conjugate [6], more precisely $-\psi^c$ is the convex conjugate of $-\psi$ or $-(-\psi)^c$ is the convex conjugate of ψ.*

The dual problem can thus be written

$$(\mathrm{DP}) = \sup_{\psi \in \mathcal{C}^0(\mathcal{Y})} \int_{\mathcal{X}} \psi^c \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu \tag{2.2.4}$$

This trick is used both in numerical application to build algorithms to solve optimal transport problem, and in theory to study the solutions of optimal transport problem. We will call *Kantorovich functional* the maximized function above.

Definition 9 (Kantorovich functional)**.** *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, Kantorovich functional $\mathcal{K}$ is defined for $\psi : \mathcal{Y} \to \mathbb{R}$ by*

$$\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^c \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu$$

To justify what follows, recall that a convex analysis result claims that a function is l.s.c. and convex if and only if it is the convex conjugate of some function. We use a generalization of this result to define the c-concavity.

Definition 10. *A function $\psi : \mathcal{Y} \to \mathbb{R}$ is c-concave if there exists a function $\varphi : \mathcal{X} \to \mathbb{R}$ such that $\psi = \varphi^c$.*

An equivalent definition for the c-concavity of ψ is if $\psi^{cc} = \psi$ (in that case, $\varphi = \psi^c$), see for instance [68]. C-concavity can be seen as a generalization of concavity where the supporting hyperplanes are replaced by level sets of the cost function c. Using this equivalence one can generalize the notion of superdifferential to c-superdifferential. The usual superdifferential of a function $f : \mathbb{R}^d \to \mathbb{R}$ at a point $y \in \mathbb{R}^d$ is

$$\partial^+ f(y) = \left\{ v \in \mathbb{R}^d \mid \forall x \in \mathbb{R}^d, \quad f(x) \leqslant f(y) + \langle v | x - y \rangle \right\}$$

The c-superdifferential of ψ at a point $y \in \mathcal{Y}$ is defined by

$$\partial^c \psi(y) = \{ x \in \mathcal{X} \mid \forall z \in \mathcal{Y}, \quad c(x, y) - \psi(y) \leqslant c(x, z) - \psi(z) \} \tag{2.2.5}$$

As it is the case for regular concavity (in finite dimension), a continuous function is c-concave if and only if its c-superdifferential is non-empty at every point. Note that when it exists, a maximizer ψ of (DP) is always c-concave. Indeed by definition of the c-transform we have for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that $\psi^c(x) + \psi(y) \leqslant c(x, y)$, and thus $\psi(y) \leqslant \inf_x c(x, y) - \psi^c(x)$ which means that $\psi \leqslant \psi^{cc}$. Recall that (DP) is a maximum with respect to ψ, thus if $\psi \neq \psi^{cc}$ on a non zero measure set with respect to ν, then ψ^{cc} being greater, it is a better choice than ψ. Note that such a maximizer does not necessarily exists. Though one can for example add to the hypothesis of Proposition 2 that the cost function satisfies $c(x, y) \leqslant c_1(x) + c_2(y)$ with $c_1 \in L^1(\mathrm{d}\mu)$ and $c_2 \in L^1(\mathrm{d}\nu)$; the problem (DP) would then admits a pair of c-conjugate functions (φ, ψ) as maximizer [68].

Distance cost. We focus here on a cost function that is a distance, i.e. $c(x, y) = \mathrm{d}(x, y)$. In that case the dual problem rewrites as a supremum over Lipschitz functions. This result is stated in the following theorem, a proof can be found in [68]. We denote by $\mathrm{Lip}(f)$ the best Lipschitz constant of a function f, i.e. $\mathrm{Lip}(f) = \sup_{x \neq y} |f(x) - f(y)| / \mathrm{d}(x, y)$.

Theorem 4 (Kantorovich-Rubinstein)**.** *Let $\mathcal{X} = \mathcal{Y}$ be a polish space and $c(x, y) = \mathrm{d}(x, y)$ where d is a lower semi-continuous distance on $\mathcal{X}$. Then the 1-Wasserstein distance $\mathrm{W}_1(\mu, \nu)$ between μ and ν in $\mathcal{P}(\mathcal{X})$ rewrites*

$$\mathrm{W}_1(\mu, \nu) = \sup_{\mathrm{Lip}(f) \leqslant 1} \int_X f \mathrm{d}(\mu - \nu)$$

We give below the main lines of the proof of this very well known theorem because it is a nice and direct application of Kantorovich duality. For simplicity we will assume here that $\mathcal{X}$ and $\mathcal{Y}$ are compact sets, though it is not necessary for the Theorem to hold. *Sketch of proof.* First recall that W_1 is defined by

$$\mathrm{W}_1(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \mathrm{d}(x, y) \mathrm{d}\gamma(x, y)$$

and by Kantorovich duality it rewrites

$$\mathrm{W}_1(\mu, \nu) = \sup_{\psi \in \mathcal{C}^0(\mathcal{Y})} \int_{\mathcal{X}} \psi^c \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu$$

for which there exists a maximizer ψ which is c-concave. What we have left to show is that when $c = \mathrm{d}$, a c-concave function ψ is 1-Lipschitz and its c-transform satisfies $\psi^c = -\psi$. First remark that if ψ is c-concave, then it writes $\psi(y) = \inf_x c(x, y) - \varphi(x)$ for some φ, and thus satisfies for any y, z

$$\psi(z) - \psi(y) \leqslant \mathrm{d}(x_y, z) - \mathrm{d}(x_y, y) \leqslant \mathrm{d}(y, z)$$

where x_y is the point that minimizes $c(x, y) - \varphi(x)$. It is thus 1-Lipschitz. Then we have

$$\psi^c(x) = \inf_y \mathrm{d}(x, y) - \psi(y) = \inf_y \mathrm{d}(x, y) - \psi(y) + \psi(x) - \psi(x) \geqslant -\psi(x).$$

By choosing $y = x$ in the first infimum we deduce $\psi^c = -\psi$. Since this result is true for any 1-Lipschitz function it is enough to conclude. □

Remark 5. *This definition of the W_1 distance can be extended to measures that does not have the same mass, or that are not even positive, by adding an L^∞ bound on the Lipschitz function f. Thus giving a metric on the whole space $\mathcal{M}(\mathcal{X})$ which can be quite useful, for example as a distance between seismograph [55].*

2.3 The semi-discrete case

Among the many formulations of optimal transport, the particular framework we will bring our interest on is the semi-discrete setting. The main point of this formulation is to transport a measure $\mu \in \mathcal{P}^{\mathrm{ac}}(\mathcal{X})$, that is absolutely continuous with respect to a reference measure, toward a discrete measure $\nu \in \mathcal{P}(\mathcal{Y})$ which takes the form of a weighted sum of diracs $\nu = \sum_i \nu_i \delta_{y_i}$. The semi-discrete setting of the optimal transport problem is useful for several applications [14, 23], for example it allows to compute the distance between a density and a point cloud. It is also an efficient way to solve optimal transport problems numerically, see for instance [54].

To define semi-discrete optimal transport, we need a reference measure on the set $\mathcal{X}$. It is very common to chose $\mathcal{X}$ to be a domain of $\mathbb{R}^d$ in which case the reference measure is the Lebesgue measure, or if $\mathcal{X}$ is a k-dimensional submanifold of $\mathbb{R}^d$, then the reference

measure is the k-Hausdorff measure. The target space $\mathcal{Y}$ is a finite set of size N. For simplicity we consider that both $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Y} \subset \mathbb{R}^d$, and the reference measure on $\mathcal{X}$ is the Lebesgue measure denoted vol^d. The semi-discrete optimal transport problem consists in finding a transport map (or plan) minimizing the transport cost from an absolutely continuous measure $\mu \in \mathcal{P}^{\mathrm{ac}}(\mathcal{X})$ to a discrete measure $\nu = \sum_{y\in\mathcal{Y}} \nu_y \delta_y$. We sometimes enumerate the set $Y = \{y_i\}_{1\leqslant i\leqslant N}$ to identify the maps going from $\mathcal{Y}$ to $\mathbb{R}$ with vectors in $\mathbb{R}^N$. In this case the target measure rewrites $\nu = \sum_{i=1}^N \nu_i \delta_{y_i}$. In the semi-discrete setting, the dual formulation can be rephrased using the notion of Laguerre tessellation, which is a generalization of Voronoi tessellation. This connection has been known for a long time, details can be found in [54]. From a pedagogical perspective, it is interesting to present here an economic metaphor that is quite common in the optimal transport community. Assume that the set $\mathcal{X}$ represents a city and the absolutely continuous measure $\mu \in \mathcal{P}(\mathcal{X})$ is the population density of that city. In this city we consider N bakeries that are located at positions $y \in \mathcal{Y}$. The cost function here is a function of the distance in the city, for differentiability we often choose the quadratic cost $c(x, y) = \mathrm{d}^2(x, y)$. At this point, it is natural to consider that each person of the city will buy their bread in the closest bakery, thus partitioning the city in the following Voronoi tessellation

$$\mathrm{Vor}_y = \{x \in \mathcal{X} \mid \forall z \in \mathcal{Y}, \mathrm{d}(x, y) \leqslant \mathrm{d}(x, z)\}$$

such that any person living at $x \in \mathrm{Vor}_y$ will buy their bread in bakery y, giving a total number of customers $\mu(\mathrm{Vor}_y)$ in each bakery y. Now assume that the bakery y has a capacity $\nu_y > 0$, and in order to have a well-posed problem we will assume that $\sum \nu_y = \mu(\mathcal{X}) = 1$ so the capacity of all the bakeries is matching exactly the population of the city (this is called the mass balance condition). Obviously the cells (Vor_y) depend on where are positioned the y, and there is no reason that $\mu(\mathrm{Vor}_y) = \nu_y$. To enforce this condition, one can then add a variable corresponding to the price of the bread for each bakery, taking the form of a function $\psi : \mathcal{Y} \to \mathbb{R}$. A customer $x \in \mathcal{X}$ will then go to bakery y if it is both close and with an attractive price, partitioning this time the space $\mathcal{X}$ in the following Laguerre tessellation such that this time, a customer $x \in \mathrm{Lag}_y(\psi)$ will go to bakery y.

Definition 11 (Laguerre cells)**.** *The Laguerre cell associated to a point $y \in \mathcal{Y}$ is a subset of $\mathcal{X}$ defined by*

$$\mathrm{Lag}_y(\psi) = \{x \in \mathcal{X} \mid \forall z \in \mathcal{Y}, c(x, y) + \psi(y) \leqslant c(x, z) + \psi(z)\}$$

For $y \neq z$ we also denote the intersection of two Laguerre cells by

$$\mathrm{Lag}_{yz}(\psi) = \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$$

One can then define the (possibly multi-valued) map T_ψ associated to a potential function $\psi : \mathcal{Y} \to \mathbb{R}$ by

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, y \in T_\psi(x) \iff x \in \mathrm{Lag}_y(\psi)$$

Remark 6 (Link with the c-superdifferential)**.** *The semi-discrete Laguerre cell* $\mathrm{Lag}_y(\psi)$ *correspond to the c-superdifferential $\partial_c\psi(y)$ of ψ at y. The formula for a transport map induced by a potential is not restricted to the semi-discrete setting. When it is well defined, we say that T is induced by ψ if $T = T_\psi = \partial^c\psi^{-1}$.*

In order to have non-overlapping Laguerre cells, and a well defined map T_ψ, we need the cost function $c \in \mathcal{C}^1(\mathcal{X} \times \mathcal{Y})$ to satisfy the classical twist condition.

Definition 12 (Twisted cost)**.** *The cost function $c \in \mathcal{C}^1(\mathcal{X} \times \mathcal{Y})$ is said to satisfy the twist condition if*

$$\forall x_0 \in \mathcal{X}, \textit{the map} \quad \begin{cases} \mathcal{Y} \to T_{x_0}\mathcal{X} \\ y \mapsto \nabla_x c(x_0, y) \end{cases} \quad \textit{is injective.}$$

Here $T_{x_0}\mathcal{X}$ denotes the tangent space of $\mathcal{X}$ at x_0.

Formally, this condition guarantees that for $y \neq z$, the level sets of the function $c(\cdot, y) - c(\cdot, z)$ are manifolds of dimension at most $d-1$. When the cost is twisted the intersection Lag_{yz} between two Laguerre cells is always of zero Lebesgue measure.

Proposition 7. *If the cost function c satisfies the twist condition, then for any $y \neq z$, $\mathrm{vol}^d(\mathrm{Lag}_{yz}(\psi)) = 0$ and the map T_ψ defined for $x \in \mathcal{X}$ by*

$$T_\psi(x) = \mathrm{argmin}_y \, c(x, y) + \psi(y)$$

is well defined Lebesgue almost everywhere. Moreover T_ψ is the optimal transport map between μ and $\nu_\psi = T_\psi \# \mu$ which is defined by

$$\nu_\psi = T_\psi \# \mu = \sum_{y \in \mathcal{Y}} \mu(\mathrm{Lag}_y(\psi))\delta_y$$

Proof. Note first that $\mathcal{Y}$ being finite, the argmin defining T_ψ always exists. We are now going to show that it is unique for almost every $x \in \mathcal{X}$. Let $f(x) = c(x, y) + \psi(y) - c(x, z) - \psi(z)$. Then by the twist condition we have $\nabla f(x) = \nabla_x c(x, y) - \nabla_x c(x, z) \neq 0$, which guarantee that $\mathrm{vol}^d(f^{-1}(\{0\})) = 0$. Since $\mathrm{Lag}_{yz} \subset f^{-1}(\{0\})$ we also have $\mathrm{vol}^d(\mathrm{Lag}_{yz}(\psi)) = 0$. It follows naturally that T_ψ is well defined Lebesgue almost everywhere. We now want to prove that T_ψ is optimal between μ and ν_ψ. By definition of T_ψ we have

$$\forall x, y \in \mathcal{X} \times \mathcal{Y}, c(x, T_\psi(x)) + \psi(T_\psi(x)) \leqslant c(x, y) + \psi(y)$$

Now let $\gamma \in \Gamma(\mu, \nu_\psi)$ be any transport plan between μ and ν_ψ, then integrating the previous inequality with respect to γ gives

$$\int_{\mathcal{X}} c(x, T_\psi(x)) + \psi(T_\psi(x)) \mathrm{d}\mu(x) \leqslant \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) + \psi(y) \mathrm{d}\gamma(x, y)$$

where the left hand side is obtained using $\Pi_x \# \gamma = \mu$ and $\Pi_y \# \gamma \in \mathcal{P}(\mathcal{Y})$. Now using that $T_{\psi\#}\mu = \Pi_y \# \gamma$, by making the change of variable $y = T_\psi(x)$ we get

$$\int_{\mathcal{X}} \psi(T_\psi(x)) \mathrm{d}\mu(x) = \int_{\mathcal{X} \times \mathcal{Y}} \psi(y) \mathrm{d}\gamma(x, y)$$

and finally

$$\int_{\mathcal{X}} c(x, T_\psi(x)) \mathrm{d}\mu(x) \leqslant \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\gamma(x, y) \qquad \square$$

Remark 8. *In the previous proposition we proved that T_ψ is always optimal between the appropriate measures, and that in the semi-discrete case under the same hypothesis Monge and Kantorovich problems are always equivalent.*

Optimal Transport as mass function prescription. Since T_ψ is always optimal between μ and ν_ψ, the Monge problem between μ and ν in the semi-discrete case then amounts to finding $\psi \in \mathbb{R}^{\mathcal{Y}}$ such that $\nu_\psi = \nu$. We define the Mass function H of a Laguerre tessellation associated to ψ by

$$\begin{aligned} H_y &: \mathbb{R}^{\mathcal{Y}} \to \mathbb{R}, \quad \psi \mapsto \mu\big(\mathrm{Lag}_y(\psi)\big) \\ H &: \mathbb{R}^{\mathcal{Y}} \to \mathbb{R}^{\mathcal{Y}}, \quad \psi \mapsto (y \mapsto H_y(\psi))\,. \end{aligned} \tag{2.3.6}$$

With these notations the measure ν_ψ defined before rewrites

$$\nu_\psi = \sum_{y \in \mathcal{Y}} H_y(\psi)\delta_y$$

and the semi-discrete optimal transport problem amounts to finding $\psi \in \mathbb{R}^{\mathcal{Y}}$ such that for any $y \in \mathcal{Y}$, $H_y(\psi) = \nu_y$. Since the set $\mathcal{Y} = \{y_i\}_{1 \leqslant i \leqslant N}$ is finite of size N, one can identify $\mathbb{R}^{\mathcal{Y}}$ and $\mathbb{R}^N$. Similarly the measure ν can be assimilated to the vector of the weights $(\nu_i)_{1 \leqslant i \leqslant N} \in \mathbb{R}^N$. The semi-discrete problems then becomes an equation in $\mathbb{R}^N$, that consists in finding $\psi \in \mathbb{R}^N$ such that

$$H(\psi) = \nu \tag{MA}$$

Equation (MA) can be seen as a discrete version of the Monge-Ampère type equation arising in optimal transport. The existence of solutions to Equation (MA) and numerical methods to solve it strongly relies on properties of the mass function H. In the following we denote by $(\mathbf{1}_y)_{y \in \mathcal{Y}}$ the canonical basis of $\mathbb{R}^{\mathcal{Y}}$, and $\mathbf{1}_{\mathcal{Y}}$ the constant equal to 1.

Proposition 9 (Properties of H, [54])**.** *Assume that the cost function c is twisted (Def 12) and that $\mu \in \mathcal{P}^{\mathrm{ac}}(\mathcal{X})$. Then the mass function H satisfies the following properties*

- $\forall y \in \mathcal{Y}, \forall t \geqslant 0, H_y(\psi + t\mathbf{1}_y) \leqslant H_y(\psi)$,
- $\forall y \neq z \in \mathcal{Y}, \forall t \geqslant 0, H_y(\psi + t\mathbf{1}_z) \geqslant H_y(\psi)$,
- $\forall y \in \mathcal{Y}, \forall t \geqslant 0, H_y(\psi + t\mathbf{1}_{\mathcal{Y}}) = H_y(\psi)$,
- $\forall y \in \mathcal{Y}$, $H(\psi) \in \mathcal{P}(\mathcal{Y})$,
- *H is continuous on $\mathbb{R}^{\mathcal{Y}}$.*

The proposition is proved in [54]. Using these properties, one can prove the existence of solution to the semi-discrete optimal transport problem for any target measure $\nu \in \mathcal{P}(\mathcal{Y})$, which is also done in [54].

Recall that equation (MA) amounts to maximizing Kantorovich functional (Definition 9) which is defined by

$$\mathcal{K}(\psi) = \int_{\mathcal{X}} \psi^c \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu = \sum_y \int_{\mathrm{Lag}_y(\psi)} c(x,y) + \psi(y)\mathrm{d}\mu(x) + \sum_y \psi(y)\nu_y$$

Theorem 10. *If $\mu \in \mathcal{P}^{\mathrm{ac}}(\mathcal{X})$, then Kantorovich functional $\mathcal{K}$ defined in 9 is concave, and if c is twisted (Def 12) then $\mathcal{K}$ is $\mathcal{C}^1$ with gradient*

$$\nabla\mathcal{K}(\psi) = H(\psi) - \nu$$

This has been shown by Aurenhammer, Hoffman, Aronov [5] in the quadratic case. Kantorovich functional $\mathcal{K}$ being concave, this result implies that maximizing the dual formulation of (KP) is equivalent to find a zero of its gradient, i.e. to solve (MA).

2.4 Non-imaging optics

Non-imaging optics (or anidolic optics) is a branch of optics where the optical components (e.g. mirrors or lenses) are designed to manipulate light in a different way than traditional imaging optics. Non-imaging optics focuses on controlling the distribution and concentration of light rather than forming an image. These kind of optics have many applications like solar collectors to produce energy, public lightning to reduce glare or increase efficient or optical communications like optical fibers. All these problems are inverse problems that can be posed by the following problem. Given a light source and a target distribution, the goal is to construct a mirror that concentrates the light energy of the source toward the target. Several variants of these optics problem exists, whether it is a reflector or a refractor, for a point or collimated source, and near-field or far-field target [70, 71, 22, 35, 34, 42]. In this section we focus on four non-imaging optics problems, and see how they fall into the scope of optimal transport and Monge-Ampere equations [57]. The first one, namely the far-field reflector problem can be recast as an optimal transport problems, while the second one, the near-field reflector problem [34], has a formulation that falls close to optimal transport but is not. In both problems we represent the light source by a measure $\mu \in \mathcal{P}^{\text{ac}}(\mathcal{X})$ where $\mathcal{X}$ is a domain embedded in $\mathbb{R}^3$. The target be a discrete measure $\nu \in \mathcal{P}(\mathcal{Y})$, with $\mathcal{Y}$ a finite set of points in $\mathbb{R}^3$. The goal is to build a mirror Σ that reflect the desired quantity of light $\nu(y)$ to each point $y \in \mathcal{Y}$ of the target.

2.4.1 Far-field parallel reflector: an optimal transport problem

The far-field parallel reflector problem considers a collimated light source, that is a plane source with parallel rays of light going upward, and a target light distribution at infinity, which means that we don't aim at points but at directions. More precisely, let $\mathcal{X} \subset \mathbb{R}^2 \times \{0\}$ be a plane embedded in $\mathbb{R}^3$, the light source $\mu \in \mathcal{P}^{\text{ac}}(\mathcal{X})$ is an absolutely continuous measure with respect to the Lebesgue measure. The light emitted at some position $x \in \mathcal{X}$ is a vertical ray of intensity $\mu(x)$. The target space $\mathcal{Y} \subset \mathcal{S}^2$ is a finite set of directions represented by unit vectors on the sphere, and the measure $\nu \in \mathcal{P}(\mathcal{Y})$ represents the quantity of light to be sent in each direction. The problem is then to find a mirror surface Σ placed above the space $\mathcal{X}$ that reflects the rays emitted by the source toward the desired target distribution.

Parametrization of the mirror. In some cases, many different mirrors Σ can be solutions of the problem. To restrict the set of mirrors and simplify the problem, we decide to consider mirrors that are plane by parts, which means that Σ is imposed to be the graph of an affine by parts function $u : \mathcal{X} \to \mathbb{R}$. This choice comes from the fact that if we want to reflect all the rays toward a direction y, by Snell's law we just have to chose a plane Σ with normal n_y satisfying $y = e_3 - 2\langle e_3 | n_y \rangle n_y$, where e_3 is the vertical upward vector, or third vector of the canonical basis of $\mathbb{R}^3$. Since the target space $\mathcal{Y}$ is of size N, we will then choose N planes on which the mirror Σ will be supported, each one being defined by its normal vector n_y. Each plane Π_y reflecting the light toward direction y thus have for equation $\langle x | n_y \rangle - \psi(y) = 0$ where $\psi : \mathcal{Y} \to \mathbb{R}$ will parametrize its height. An additional choice we make is to consider the mirror Σ to be the maximum of the planes $(\Pi_y)_{y \in \mathcal{Y}}$. This choice is completely arbitrary, one could have considered a minimum, or even some obscure way of constructing an affine by parts function, but maximum is nice because it yields to a convex mirror. It means that the function $u : \mathcal{X} \to \mathbb{R}$ is defined for

$x \in \mathcal{X}$ by

$$u(x) = \max_{y \in \mathcal{Y}} \langle x | n_y \rangle - \psi(y).$$

An example of such a mirror is given in Figure 2.

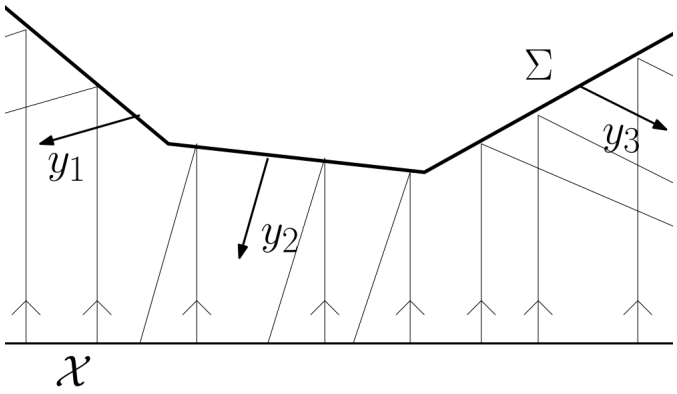

Figure 2: An example of reflector Σ composed of three planes.

The far-field parallel reflector as Optimal Transport. Let us define the cost function $c(x, y) = -\langle x | n_y \rangle$, which is twisted when all the $y \in \mathcal{Y}$ are distinct. Then $\Sigma = \{(x, u(x)) | x \in \mathcal{X}\}$ reflects x toward y if for any $z \in \mathcal{Y}$, $-c(x, y) - \psi(y) \geqslant -c(x, z) - \psi(z)$, or equivalently $x \in \mathrm{Lag}_y(\psi)$, where the Laguerre cell Lag_y is defined in Def 11. Another way of saying this is that the function ψ defines the height of each mirror at the origin, and the point x will be reflected by the highest plane above this point. The quantity of light reflected to y is the quantity of light in $\mathrm{Lag}_y(\psi)$, which is $\mu(\mathrm{Lag}_y(\psi))$.

Our parametrization thus allows to solve the Far-field parallel reflector problem by finding $\psi : \mathcal{Y} \to \mathbb{R}$ such that

$$\forall y \in \mathcal{Y}, \mu(\mathrm{Lag}_y(\psi)) = \nu(y) \tag{FF-Par}$$

which is exactly Equation (MA). The far-field parallel reflector problem can thus be reduced to a semi-discrete optimal transport problem for the cost $c(x, y) = -\langle x | y \rangle$ (or $-\langle x | n_y \rangle$ which is equivalent up to a redefinition of the target). The fact that this problem can be recast as the dual of an optimal transport problem is not trivial and quite surprising. Note that alike every optimal transport problem, this problem is invariant by addition of a constant. Indeed if one replace ψ by $\psi + C$, the Laguerre diagram will remain the same. Physically this means that if we raise or lower the mirror, the light will be reflected to the same direction. A few other non-imaging optics problems can be recast as optimal transport problems (see for example [70, 71, 22, 34]), but it is not always the case.

2.4.2 Near-field parallel reflector: a generated Jacobian equation

This section is about another non-imaging optics problem, the near-field parallel reflector [34]. We will see that unlike the previous one, this problem cannot be recast as an optimal transport, but can take the form of a slightly more general problem, namely a generated Jacobian equation. Like for the far-field parallel reflector, the source is collimated (all rays parallel and vertical) emitted from a plane $\mathcal{X} \subset \mathbb{R}^2 \times \{0\}$, and represented

by a measure $\mu \in \mathcal{P}^{\text{ac}}(\mathcal{X})$. This time the target set $\mathcal{Y}$ will be a finite set that also lives in $\mathbb{R}^2 \times \{0\}$, and the target measure $\nu \in \mathcal{P}(\mathcal{Y})$ is a discrete measure on $\mathcal{Y}$. We are aiming at points and not direction, hence the terminology "near-field". The problem is to construct a mirror Σ that reflects μ to ν.

Parametrization of the mirror. If one wants to reflect all the rays emitted upward from $\mathcal{X}$ toward the same point y, the solution is to chose a downward parabola P_y of focal point y that is above $\mathcal{X}$. A parabola P_y, of focal distance $\psi(y)$, is the graph of the function $x \mapsto 1/2\psi(y) - \psi(y)\|x-y\|^2/2$. Once again, we chose Σ to be the maximum of N downward paraboloids of focal point y and focal distance $\psi(y)$ for some $\psi : \mathcal{Y} \to \mathbb{R}$. The mirror Σ is the graph of the function

$$u(x) = \max_{y \in Y} \frac{1}{2\psi(y)} - \frac{\psi(y)}{2}\|x-y\|^2.$$

An example of paraboloid by parts mirror is given in Figure 3.

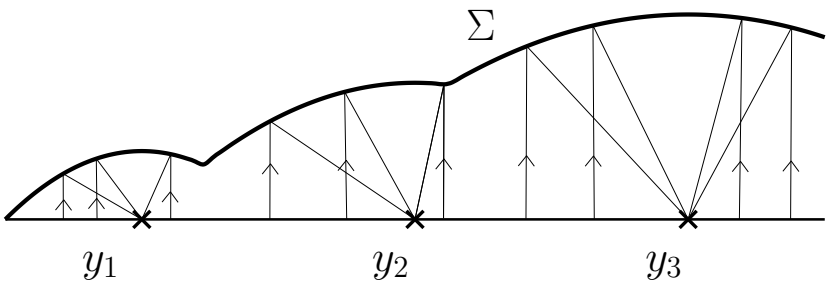

Figure 3: An example of reflector Σ composed of three paraboloids of focus y_i.

Near-field parallel reflector, a Generated Jacobian Equation. Let us consider the function $G : \mathcal{X} \times \mathcal{Y} \times \mathbb{R}_+^* \to \mathbb{R}$ defined by

$$G(x,y,v) = \frac{1}{2v} - \frac{v}{2}\|x-y\|^2$$

The light reflected toward a point y is the subset of $\mathcal{X}$ defined by

$$\text{Lag}_y(\psi) = \{x \in \mathcal{X} \,|\forall z \in \mathcal{Y}, G(x,y,\psi(y)) \geqslant G(x,z,\psi(z))\}$$

By a slight abuse of notation, we denoted the above cell Lag_y while it is not exactly a Laguerre cell, but a "generalized" Laguerre cell. This is due to the fact that the expression of $G(x,y,\psi(y))$ can not be split in a sum of the variable $\psi(y)$ and a cost function $c(x,y)$ independent of ψ. The function G is called the generating function, a precise definition is given in Chapter 4, Definition 25.

The near-field parallel reflector problem then amounts to finding $\psi : \mathcal{Y} \to \mathbb{R}$ such that

$$\forall y \in \mathcal{Y}, \mu(\text{Lag}_y(\psi)) = \nu(y) \qquad \text{(NF-par)}$$

This equation is not an optimal transport problem but is somehow quite similar, it is called a *generated Jacobian equation*. These equations were initially introduced in the continuous setting as Monge-Ampere type equations by Trudinger [66], a pedagogical

survey was made by Guillen [34]. It is stated here in its semi-discrete setting. Unlike optimal transport, the generated Jacobian equation has no variational formulation, which implies in particular that the mass function of the Laguerre cells H defined in Equation (2.3.6) has no reason to be a gradient. This problem is not invariant by addition of a constant, but the set of solutions is still a family of functions with one degree of freedom. Unlike optimal transport, the Laguerre tessellation corresponding to the solutions has no reason to be unique. The structures being similar, some of the algorithm to solve optimal transport problems can still be used to solve generated Jacobian equations. Details about this are given in Chapter 4.

2.4.3 Some other non-imaging optics problems

There exist other non-imaging optics problems that can be written as optimal transport or generated Jacobian equations. The approximation consisting in placing the target at infinity, i.e. the far-field case, seems to lead to optimal transport problems, while the near field case leads to generated Jacobian equations. We present two additional non-imaging optics problems [22, 71, 34]. The first one will be recast as optimal transport, while the second one will fall in the scope of generated Jacobian equations. Both are similar to what we presented before, the difference is that we chose a point light source instead of a collimated one, and thus represented by directions on the sphere.

Far-field point reflector

The far-field point reflector is very similar to the first problem. The target is located at infinity, so we are aiming at directions represented by unit vectors on the sphere, i.e. $\mathcal{Y} = (y_i)_{1\leqslant i\leqslant N} \subset \mathcal{S}^2$. The target intensity is represented by a discrete measure $\nu \in \mathcal{P}(\mathcal{Y})$. The light is emitted by a single point, which we consider at the origin $\mathcal{O} \in \mathbb{R}^3$. The source is thus characterized by a measure $\mu \in \mathcal{P}(\mathcal{S}^2)$, such that the ray emitted in direction x is of intensity $\mu(x)$. Following the technique of the two previous problems, it is natural here to consider a mirror shaped of N paraboloid of axis $(y_i)_{1\leqslant i\leqslant N}$ and sharing the origin $\mathcal{O}$ as focal point. Since the source is the origin, it makes sense to use the radial parametrization of the mirror Σ. A paraboloid of revolution of axis y and focal distance $2/v(y)$ is defined for $x \in \mathcal{S}^2$ by its radial function

$$\rho_y(x) = \frac{1}{v(y)} \frac{1}{1 - \langle x|y\rangle}.$$

This time, we consider the minimum of radial functions, meaning that we will choose the paraboloid that reflects the light toward y when its the closest of the origin (one can remark that this choice makes the mirror convex, which is nice in practice). We thus define the mirror by the formula $\Sigma = \{x\rho(x) \mid x \in \mathcal{S}^2\}$ where $\rho(x) = \min_{y\in\mathcal{Y}} \rho_y(x)$. Passing to the logarithm, we get $\log(\rho_y(x)) = -\log(v(y)) - \log(1 - \langle x|y\rangle)$, thus if we consider the variables $\psi(y) = -\log(v(y))$ and the cost function $c(x, y) = -\log(1 - \langle x|y\rangle)$, we get that the set of rays reflected in the direction y is the Laguerre cell

$$\mathrm{Lag}_y(\psi) = \{x \in \mathcal{S}^2 \mid \forall z \in \mathcal{Y}, -\ln(1 - \langle x|y\rangle) + \psi(y) \leqslant -\ln(1 - \langle x|z\rangle) + \psi(z)\}$$

and again, the far-field point reflector problem amounts to solving the optimal transport problem

$$\forall y \in \mathcal{Y}, \mu(\mathrm{Lag}_y(\psi)) = \nu(y) \qquad \text{(FF-point)}$$

This cost function is a bit more unconventional than what is usually studied, but it is still lower semi-continuous and bounded below. One can also take advantage that it is repulsive

when non differentiable, which means that the points (x, y) where c is not differentiable satisfy $c(x, y) = +\infty$, and it is possible to quantify the fact that any "decent" transport plan stays afar from such points. Some details on this strategy can be found at the end of Chapter 3.

Near-field point reflector

Finally we will briefly introduce the near-field point reflector [42, 34], to emphasize the fact that near-field reflector problems seems to always recast as generated Jacobian equations instead of optimal transport. Is is also interesting to notice that among the four problems presented in this section, this one is the most common in physics because of its point source and target at finite distance (far-field being an approximation). The framework is the same as the previous problem, except that here the target is at finite distance. This means that the target space $\mathcal{Y} \subset \mathbb{R}^3$ is a set of points in the space $\mathbb{R}^3$ instead of direction on the sphere $\mathcal{S}^2$. The source being at the origin $\mathcal{O} \in \mathbb{R}^3$, the mirror reflecting all the light from the source to a single point y is an ellipsoid which has $\mathcal{O}$ and y for focal points. The radial parametrization of such an ellipsoid is the function

$$e_y(x, t) = \frac{t^{-2} - \frac{1}{4}\|y\|^2}{t^{-1} - \frac{1}{2}\langle x|y\rangle}$$

with $t \in]0, 2/\|y\|]$ so that the eccentricity $\|y\|\, t/2 \in]0, 1]$. This formulation of the radial function of an ellipsoid is as written by Guillen [34]. It might not seem very natural but it is nice for us because it matches the definition of a generating function, refer to Chapter 4 for details. Let us define the function $G(x, y, t) = e_y(x, t)$. Then if one chooses the furthest ellipsoid from the origin, it amounts to pick the maximum among the radial functions e_y to parametrize the mirror Σ. Finally, the set of rays reflected toward the point y will be the "generalized" Laguerre cell defined by

$$\mathrm{Lag}_y(\psi) = \{x \in \mathcal{S}^2 \mid \forall z \in \mathcal{Y}, G(x, y, \psi(y)) \geqslant G(x, z, \psi(z))\}$$

and the near-field point reflector thus consists in finding $\psi : \mathcal{Y} \to \mathbb{R}$ such that

$$\forall y \in \mathcal{Y}, \mu(\mathrm{Lag}_y(\psi)) = \nu(y) \qquad \text{(NF-point)}$$

This problem cannot be solved using optimal transport, but it is a generated Jacobian equation. The regularity of the solutions has been studied in [42].

We now have introduced the necessary notions to present the research contained in this thesis. In the following chapters, we will show stability of optimal transport maps and numerical techniques to solve generated Jacobian equations. All these results can be applied to non-imaging optics problem which are the core of this work.

Chapter 3

Stability of optimal transport maps

The stability of solutions to optimal transport problems under variation of the measures is fundamental from a mathematical viewpoint: it is closely related to the convergence of numerical approaches to solve optimal transport problems and justifies many of the applications of optimal transport. In this chapter, we introduce the notion of *strong c-concavity*, and we show that it plays an important role for proving stability results in optimal transport for general cost functions c. We then introduce a differential criterion for proving that a function is strongly c-concave, under an hypothesis on the cost introduced originally by Ma-Trudinger-Wang for establishing regularity of optimal transport maps. Finally, we provide two examples introduced in Chapter 2 where this stability result can be applied, for cost functions taking value $+\infty$ on the sphere: the far-field point reflector problem and the Gaussian curvature measure prescription problem. This chapter originates from [31] written in collaboration with Quentin Mérigot and Boris Thibert.

3.1 Introduction

Numerical applications of optimal transport theory have been made possible thanks to the tremendous progress of optimal transport solvers in the last decade [60, 54, 7].

The stability of solutions to optimal transport problems under variation of the data is fundamental from a mathematical viewpoint, making optimal transport a "well-posed" problem in the terminology of Hadamard. The question of *quantitative stability* is also of prime importance. The first and most obvious reason is that it is strongly related to the convergence of many numerical approaches to solve optimal transport problems — both in statistical and in numerical analysis contexts — and explicitly or implicitly it justifies most of the applications of optimal transport. Quantitative stability is at the heart of several other applications, including the understanding of geometric embeddings of spaces of probability measures to Hilbert spaces used in statistics [24], the convergence analysis of numerical methods for evolution equations using optimal transport as a building block [14], the estimation of transport maps in high dimension [39] or the construction of precise asymptotics for random matching problems [3].

The stability of optimal transport plans can be established in a very general setting [69], under variations of the source and target measures, and even under variations of the cost. However, the question of *quantitative* stability has only been addressed rather recently, and most of the existing results deal with the cost function $c(x, y) = \|x - y\|^2$

[33, 11, 24, 48], or with the squared geodesic distance on a Riemannian manifold [3]. The aim of this chapter is to establish stability results for more general cost functions, namely those that satisfy the *strong Twist* and *Ma-Trudinger-Wang* conditions on manifolds. We also identify *strong c-concavity* of the Kantorovitch potential as a central notion to get stability results. In this chapter, the cost function $c : M \times N \to \mathbb{R} \cup \{+\infty\}$ will be defined on the product space $M \times N$ where M and N will be either domains in $\mathbb{R}^d$ or submanifolds of $\mathbb{R}^d$.

3.1.1 Existing stability results

The problem of stability of optimal transport maps can be expressed as a continuity property of the map $(\mu, \nu) \mapsto T_{\mu\to\nu}$, where $T_{\mu\to\nu}$ is the optimal transport map between a source probability measure μ and a target measure ν. In order to have a common space in which to consider the optimal transport map $T_{\mu\to\nu}$, we will mainly consider the problem of the stability of the map $T_\nu := T_{\mu\to\nu}$ for a fixed μ. As first noted by Li and Nochetto [48], the arguments implying quantitative stability of $\nu \mapsto T_{\mu\to\nu}$ sometimes also imply general stability results, where both the source and target measures can change.

To the best of our knowledge, the first quantitative stability result in optimal transport is of "local" nature, in the sense that it only holds near a configuration (μ, ν), and is established under strong assumptions on the data. It is due to Ambrosio and reported in an article of Gigli [33]. It can be phrased as follows.

Theorem 11 (Ambrosio-Gigli)**.** *Assume that M and N are compact subsets of $\mathbb{R}^d$, that $\mu \in \mathcal{P}(M)$ is absolutely continuous, and that for some $\nu_0 \in \mathcal{P}(N)$ the optimal transport map $T_{\mu\to\nu_0}$ for the quadratic cost $c(x, y) = \|x - y\|^2$ is Lipschitz. Then*

$$\forall \nu_1 \in \mathcal{P}(N), \quad \|T_{\mu\to\nu_0} - T_{\mu\to\nu_1}\|^2_{L^2(\mu)} \leqslant 2\,\mathrm{diam}(M)\mathrm{Lip}(T_{\mu\to\nu_0})W_1(\nu_0, \nu_1). \tag{3.1.1}$$

In the above statement, $\mathrm{Lip}(T)$ is the Lipschitz constant of the map T and $\mathrm{W}_1(\nu_0, \nu_1)$ is the Wasserstein distance between ν_0 and ν_1 with respect to the Euclidean distance on N. For pedagogical purpose, and because it is similar to the proof for more general cost, we give below a proof of this theorem that can be found in [24, Theorem 2.2].

Proof. We denote by $T_i = T_{\mu\to\nu_i}$ for $i \in \{0, 1\}$. By Brenier theorem [15], we know that T_i is the gradient of a convex function φ_i. A convex analysis result [6] shows that T_0 is K-Lipschitz if and only if the convex conjugate of φ_0 defined by $\varphi_0^*(y) = \psi_0(y) = \max\langle x|y\rangle + \varphi_0(x)$ is $\frac{1}{K}$-strongly convex. Let $A = \int_M \psi_0 \mathrm{d}(\nu_1 - \nu_0)$ and $B = \int_M \psi_1 \mathrm{d}(\nu_0 - \nu_1)$ Using that $\nabla\varphi_i(x)_\#\mu = \nu_i$ we get

$$\begin{aligned} A &= \int_M \psi_0(\nabla\varphi_1(x)) - \psi_0(\nabla\varphi_0(x))\mathrm{d}\mu(x) \\ &= \int_M \psi_0(\nabla\psi_1^*(x)) - \psi_0(\nabla\psi_0^*(x))\mathrm{d}\mu(x). \end{aligned}$$

By strong convexity of ψ_0 we have for any v in the subdifferential of $\partial\psi_0(z)$ that $\psi_0(y) - \psi_0(z) \geqslant \langle v|y - z\rangle + \frac{1}{2K}\|y - z\|^2$. Taking $z = \nabla\psi_0^*(x)$ and $y = \nabla\psi_1^*(x)$ one can check that $x \in \partial\psi_0(z)$ thus giving

$$A \geqslant \int_M \langle x|\nabla\psi_1^*(x) - \nabla\psi_0^*(x)\rangle + \frac{1}{2K}\|\nabla\psi_1^*(x) - \nabla\psi_0^*(x)\|^2 \,\mathrm{d}\mu(x)$$

Similarly using convexity of ψ_1 one has

$$B \geqslant \int_M \langle x | \nabla\psi_0^*(x) - \nabla\psi_1^*(x)\rangle \mathrm{d}\mu(x)$$

and summing these two inequalities one get

$$\int_N \psi_0 - \psi_1 \mathrm{d}(\nu_1 - \nu_0) \geqslant \frac{1}{2K} \left\|\nabla\psi_1^*(x) - \nabla\psi_0^*(x)\right\|^2 \mathrm{d}\mu(x) = \frac{1}{2K} \left\|T_0 - T_1\right\|^2_{L^2(\mu)}$$

Moreover it can be shown [24] that $\mathrm{Lip}(\psi_i) \leqslant \mathrm{diam}(M)$, which finally gives

$$\begin{aligned}
\left\|T_0 - T_1\right\|^2_{L^2(\mu)} &\leqslant 2K \int_N \psi_0 - \psi_1 \mathrm{d}(\nu_1 - \nu_0) \\
&\leqslant 2K \max_{\mathrm{Lip}(f)\leqslant \mathrm{diam}(M)} \int_N f \mathrm{d}(\nu_1 - \nu_0) \\
&= 2K\,\mathrm{diam}(M) \max_{\mathrm{Lip}(f)\leqslant 1} \int_N f \mathrm{d}(\nu_1 - \nu_0) \\
&= 2K\,\mathrm{diam}(M)\,\mathrm{W}_1(\nu_0, \nu_1)
\end{aligned}$$

where the last inequality is given by Kantorovich-Rubinstein theorem. □

Li and Nochetto [48] have a similar result but with respect to both measures.

Theorem 12 (Li-Nochetto)**.** *Assume the hypothesis of Theorem 11 (with $\nu_0 = \nu$), and denote by $\gamma \in \mathcal{P}(M \times N)$ the transport plan induced by the optimal map $T_{\mu\to\nu}$. Then for any two measure $\tilde{\mu} \in \mathcal{P}(M)$, $\tilde{\nu} \in \mathcal{P}(N)$, and* any *optimal transport plan $\tilde{\gamma}$ between $\tilde{\mu}$ and $\tilde{\nu}$, i.e. any solution to* (KP) *then*

$$\mathrm{W}_2(\gamma, \tilde{\gamma})^2 \leqslant C(\mathrm{W}_2(\mu, \tilde{\mu}) + \mathrm{W}_2(\nu, \tilde{\nu}))$$

where C is a constant that depends on $\mathrm{Lip}(T_{\mu\to\nu})$ and on the diameters of M and N. The Wasserstein distance W_2 in the left-hand side is with respect to a product metric on $M \times N$.

The "Euclidean" stability result of Theorem 11 can be extended to optimal transport problems on a compact Riemannian manifold with the squared geodesic distance [3]. We also mention the more "global" stability results of [11, 24], which do not make regularity assumptions on $T_{\mu\to\nu}$, but come with worse continuity estimates. For instance, the main theorem of [24] shows that if $\mu \in \mathcal{P}(\mathbb{R}^d)$ is a probability density on a compact convex subset of $\mathbb{R}^d$, which is bounded from above and below by a positive constant, then for any compact subset $Y \subseteq \mathbb{R}^d$, the map $\nu \mapsto T_{\mu\to\nu}$ is $\frac{1}{6}$-Hölder from $(\mathcal{P}(Y), \mathrm{W}_1)$ to $\mathrm{L}^2(\mu, \mathbb{R}^d)$, to be compared to the $\frac{1}{2}$ exponent in (3.1.1).

3.1.2 Strong c-concavity of the potential

A key ingredient in the stability results for the quadratic cost [33, 3] is the strong convexity of the Brenier potentials associated to the optimal transport maps. For general cost functions, Brenier theorem doesn't holds, but one can generalize it in some sense by introducing c-concavity using Kantorovich duality presented in Section 2.2. In order to get stability results for these general cost functions c, we introduce below the notion of *strong c-concavity*. To inscrease the readibility of this chapter, we recall a few notions that are already presented in Section 2.2.

About the support of transport maps. The cost function is often not regular on the whole product space. To be more general, we thus consider that the cost function is "regular enough" on a subset $D \subset M \times N$ that does not necessarily writes as a product. We denote by $\mathcal{X} = \mathrm{proj}_M(D)$ and $\mathcal{Y} = \mathrm{proj}_N(D)$. We will then consider that any optimal transport plan is supported on D, or that any optimal transport map $T : \mathcal{X} \to \mathcal{Y}$ satisfies for any $x \in \mathcal{X}$, $(x, T(x)) \in D$. Take for example the cost of the far-field point reflector (FF-point) presented in Section 2.4, which is defined on $M \times N$ by $c(x, y) = -\ln(1 - \langle x|y\rangle)$ with $M = N = \mathcal{S}^{d-1}$. Then $c(x, y) = +\infty \iff x = y$, meaning that c explodes on the diagonal $\Delta = \{(x, x)|x \in \mathcal{S}^{d-1}\}$ but is regular on $D = M \times N \setminus \Delta$. Since the cost is repulsive when it is not regular, we can show that any descent transport plan stays afar from the diagonal Δ, this is done in Section 3.5.

We denote by $\mathrm{d}_N : N \times N \to \mathbb{R}_+$ a distance on N. Recall that the p-Wasserstein distance on $\mathcal{P}(N)$ between two probability measures is defined by

$$W_p^p(\nu_0, \nu_1) = \inf_{\gamma \in \Gamma(\nu_0, \nu_1)} \int_{N \times N} d_N(y, z)^p d\gamma(y, z),$$

Definition 13 (Transport map induced by a potential)**.** *Let $T : M \to N$ be a measurable map, and $\psi : N \to \mathbb{R}$. We say that T is induced by ψ, or that ψ is a potential associated to T if*

$$\forall x \in M, \quad T(x) \in \mathrm{argmin}_{y \in N}\, c(x, y) - \psi(y)$$

We know by Kantorovich theory that if a transport map T from μ to ν is induced by a potential ψ then T is a solution to the Monge problem (MP). Such a potential ψ can be constructed by solving the dual problem

$$\sup_{\psi : N \to \mathbb{R}} \int_M \psi^c \mathrm{d}\mu + \int_N \psi \mathrm{d}\nu \tag{DP}$$

where $\psi^c : M \to \mathbb{R}$ is the c-transform of ψ (Def 8), see Section 2.2 for details. The dual problem (DP) has a maximizer, for instance, if the cost c is continuous on the compact $M \times N$, but existence also holds with weaker hypotheses on c, some of which can be found in [69]. When such a maximizer exists, and still by Kantorovich theory, we can assume that a map T solution of (MP) is induced by a c-concave potential ψ. We give here an equivalent notion for c-concavity.

Proposition 13 (Equivalent definition of c-concavity)**.** *The function $\psi : N \to \mathbb{R} \cup \{-\infty\}$ is c-concave (Def 10) if and only if for any $y \in N$ there exists $x \in M$ such that*

$$\forall z \in N, \quad \psi(z) - c(x, z) \leqslant \psi(y) - c(x, y)$$

Recall that the c-superdifferential of ψ at a point $y \in N$ is defined by

$$\partial^c \psi(y) = \{x \in M \mid \forall z \in N, \psi(z) - c(x, z) \leqslant \psi(y) - c(x, y)\}$$

Note that ψ is c-concave iff for any $y \in N$ its c-superdifferential $\partial^c \psi(y)$ is non-empty. We can now introduce the notion of strong c-concavity.

Definition 14 (strong c-concavity on D)**.** *We say that a c-concave function ψ is strongly c-concave on a set $D \subseteq M \times N$ with modulus $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ if for all x, y, z such that $(x, y) \in D, (x, z) \in D$ and $x \in \partial^c \psi(y)$ one has*

$$\psi(z) - c(x, z) \leqslant \psi(y) - c(x, y) - \omega(\mathrm{d}_N(y, z)) \tag{3.1.2}$$

In the above definition, the modulus $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function that satisfies $\omega(0) = 0$. One can check that when $c(x, y) = -\langle x|y\rangle$ and $\omega(r) = Cr^2$ the notion of strong concavity and strong c-concavity are equivalent. Moreover if a function $\psi : N \to \mathbb{R}$ is strongly c-concave, then for $y \neq z$ in N, $\partial^c\psi(y) \cap \partial^c\psi(z) = \emptyset$. Equivalently for any $x \in M$, when it exists, the minimizer of $y \mapsto c(x, y) - \psi(y)$. This implies that the transport map associated to ψ is uniquely defined by minimizing $c(x, \cdot) - \psi$:

$$\forall x \in M \quad T(x) = \operatorname{argmin}_{y\in N} c(x, y) - \psi(y)$$

The map T is actually not defined on the whole set M but only when the minimum exists, which is only guaranteed on the image of $\partial^c\psi$. In other words for the transport map $T = \partial^c\psi^{-1}$ to be well defined, we need $\partial^c\psi$ to be surjective.

3.1.3 Contribution

This chapter is concerned with stability problems in optimal transport. We introduce the notion of *strong c-concavity*, which is central to get stability results.

- We provide two stability results in Section 3.2 that depend on an assumption of strong c-concavity. First, we extend the 1/2-Hölder stability result of Ambrosio stated in [33] to general cost function c (Theorem 14). Our result is local around transport maps associated to *strongly c-concave* potential. Second, we generalize a result of Li and Nochetto [48] that estimates the distance of a transport plan to an optimal transport map (the source and target measures being fixed) in terms of the suboptimality gap (Proposition 17). We then use this result to obtain quantitative stability of the transport plan with respect to both measures (Proposition 19), following the strategy of Li-Nochetto [48] for the quadratic cost.
- We provide in Section 3.3 the central result of this work (Theorem 24), which is a differential criterion for a potential function ψ to be strongly c-concave. This result generalizes a sufficient condition for c-convexity proposed by Villani [69, Th. 12.46]. It requires that M, N are two smooth d-dimensional complete Riemannian manifolds. Similarly to Villani, we also require a local condition on the derivatives of the potential ψ and a weak Ma-Trudinger-Wang condition [51] . In Section 3.4, we combine Theorem 24 to the stability results of Section 3.2 to get local stability results for optimal transport maps.
- The last two sections are dedicated to the applications of our stability results to two optimal transport problems on the sphere, with cost functions taking the value $+\infty$. In Section 3.5 we consider the reflector antenna problem, which is a non-imaging optics problem that can be recast as an optimal transport problem [71]. Section 3.6 is dedicated to the prescription of the Gaussian curvature measure of a convex body, originally introduced by Alexandrov [2] which can also be rephrased as an optimal transport problem by Oliker [58].

3.2 Stability under strong c-concavity

In this section we assume that M and N are Polish spaces. We provide stability results in the neighborhood of transport maps that are associated to strongly c-concave Kantorovitch potential. The stability result of Section 3.2.1 is with respect to variations of the target measure, whereas the result in Section 3.2.3 is with respect to variations of

both the source and the target measures. This last result is a consequence of an error bound for a fixed optimal transport problem given in Section 3.2.2. As a side note, we also remark in the last section that strong c-concavity implies Hölder regularity of transport maps.

3.2.1 Stability with respect to the target measure

The following theorem extends to general cost functions a theorem of Ambrosio [33], using a reformulation proposed in [24]. The hypothesis that the transport map T is Lipschitz (in the formulation of [24]) is replaced by the assumption that the transport map is induced by a strongly c-concave potential ψ, i.e.

$$\forall x \in M \quad T(x) \in \operatorname{argmin}_{y \in N} c(x,y) - \psi(y).$$

Theorem 14. *Let $D \subseteq M \times N$ be a compact set and $c : M \times N \to \mathbb{R} \cup \{+\infty\}$ be a cost function of class $\mathcal{C}^1$ on D. Let $\mu \in \mathcal{P}(M)$ and $\nu_0, \nu_1 \in \mathcal{P}(N)$. We assume that there exists optimal transport maps T_i from μ to ν_i with associated potential $\psi_i : N \to \mathbb{R}$ $(i = 0, 1)$ such that:*

- *ψ_0 is Lipschitz on N and c-concave on D.*
- *ψ_1 is Lipschitz on N and strongly c-concave with modulus ω on D.*
- *The maps T_i satisfies for any $x \in M$, $(x, T_i(x)) \in D$.*

Then,

$$\int_M \omega(\mathrm{d}_N(T_0(x), T_1(x)))\mathrm{d}\mu(x) \leqslant (\mathrm{Lip}(\psi_0) + \mathrm{Lip}(\psi_1))W_1(\nu_0, \nu_1) \tag{3.2.3}$$

Remark 15. *The left hand side of inequality* (3.2.3) *measures the distance between transport maps T_0 and T_1. To see this let us consider a simpler case where M and N are domains of $\mathbb{R}^d$ and $\omega(r) = r^2$ then we get*

$$\int_M \omega(\mathrm{d}_N(T_0(x), T_1(x)))d\mu(x) = \|T_1 - T_0\|^2_{L^2(\mu)}$$

and in that case, Theorem 14 amounts to bounding the L^2 norm of the distance between transport maps. Note that unlike Equation (3.1.1)*, the Lipschitz constants of the right hand side of* (3.2.3) *are the ones of the potentials and not the transport map.*

Remark 16 (Discretization of the target measure)**.** *Assume that we have two absolutely continuous measures $\mu \in \mathcal{P}(M)$ and $\nu \in \mathcal{P}(N)$ and an optimal transport map T from μ to ν satisfying all the hypothesis of Theorem 14. One can pick a family of points $(y_i)_{1 \leqslant i \leqslant n}$ in the target space N and approximate the measure ν by a discrete measure ν_h of the form*

$$\nu_h = \sum_i \nu(V_i)\delta_{y_i}$$

where $(V_i)_{1 \leqslant i \leqslant n}$ is a Voronoi tesselation of N around the points $(y_i)_{1 \leqslant i \leqslant n}$ chosen in an appropriate way in the support of ν. The parameter h is given by $h = \max_{1 \leqslant i \leqslant n} \operatorname{diam}(V_i)$ so that $W_1(\nu, \nu_h) \leqslant h$. We can compute the optimal transport map T_h between μ and ν_h using semi-discrete methods such as [45]. Then, Theorem 14 implies

$$\int_M \omega(\mathrm{d}_N(T(x), T_h(x)))d\mu(x) \leqslant Ch$$

where the constant C depends on the Lipschitz constants of the potentials, which can be controlled explicitly in many cases. If the modulus $\omega(r)$ is quadratic, then the $\mathrm{L}^2(\mu)$ distance between T and T_h is controlled by $h^{1/2}$.

Proof of Theorem 14. We have

$$\langle \nu_1 - \nu_0 | \psi_1 - \psi_0 \rangle = \int_N \psi_1 \mathrm{d}(\nu_1 - \nu_0) + \int_N \psi_0 \mathrm{d}(\nu_0 - \nu_1)$$

Let $A = \int_N \psi_1 \mathrm{d}(\nu_1 - \nu_0)$ and $B = \int_N \psi_0 \mathrm{d}(\nu_0 - \nu_1)$. Since $T_{i\#}\mu = \nu_i$ we have

$$\begin{aligned} A &= \int_N \psi_1 \mathrm{d}\nu_1 - \int_N \psi_1 \mathrm{d}\nu_0 \\ &= \int_M \psi_1(T_1(x))\mathrm{d}\mu(x) - \int_M \psi_1(T_0(x))\mathrm{d}\mu(x) \end{aligned}$$

For $x \in M$ we have $x \in \partial^c \psi_i(T_i(x))$. Then the strong c-concavity of ψ_1 gives

$$\begin{aligned} A &= \int_M \psi_1(T_1(x)) - \psi_1(T_0(x))\mathrm{d}\mu(x) \\ &\geqslant \int_M c(x, T_1(x)) - c(x, T_0(x)) + \omega(\mathrm{d}_N(T_0(x), T_1(x)))\mathrm{d}\mu \end{aligned}$$

Now since ψ_0 is also c-concave, we have

$$B \geqslant \int_M -c(x, T_1(x)) + c(x, T_0(x))\mathrm{d}\mu$$

Summing these two inequalities gives

$$\int_M \omega(\mathrm{d}_N(T_0(x), T_1(x)))\mathrm{d}\mu(x) \leqslant \int_N \psi_1 - \psi_0 \mathrm{d}(\nu_1 - \nu_0)$$

Since ψ_0 and ψ_1 are Lipschitz, using Kantorovich-Rubinstein theorem we get

$$\int_N \psi_1 - \psi_0 \mathrm{d}(\nu_1 - \nu_0) \leqslant (\mathrm{Lip}(\psi_0) + \mathrm{Lip}(\psi_1)) W_1(\nu_0, \nu_1). \qquad \square$$

3.2.2 Error bounds for optimal transport problems

In this section, we generalize in Proposition 17 a stability result of Li and Nochetto [48] to general cost functions, using the notion of strong c-concavity. This result allows us to bound in Corollary 18 the Wasserstein distance between the optimal transport map and any transport plan with the same marginals by the suboptimality gap of the transport plan.

Proposition 17. *Let $\mu \in \mathcal{P}(M)$, $\nu \in \mathcal{P}(N)$ and $T : M \to N$ be an optimal transport map from μ to ν. We assume that T is induced by a strongly c-concave potential $\psi : N \to \mathbb{R}$ with modulus ω on a compact subset D of $M \times N$ which contains the graph of T. Then any transport plan $\gamma \in \Gamma(\mu, \nu)$ supported on D satisfies*

$$\int_{M\times N} \omega(\mathrm{d}_N(T(x), y)) d\gamma(x, y) \leqslant \int_{M\times N} c(x, y) d\gamma(x, y) - \int_M c(x, T(x)) d\mu(x)$$

The right hand side of this equation is called the suboptimality gap of γ, and measures how worse the transport plan γ behaves compared to the optimal transport map T.

Proof. The strong c-concavity of ψ implies that for any $x, y \in D$,

$$\psi(y) \leqslant \psi(T(x)) - c(x, T(x)) + c(x, y) - \omega(\mathrm{d}_N(T(x), y)).$$

Moreover since $T_{\#}\mu = \nu$, we have

$$\int_N \psi(y) d\nu(y) = \int_M \psi(T(x)) d\mu(x)$$

which combined with the strong c-concavity of ψ gives

$$\begin{aligned}
0 &= \int_N \psi(y) d\nu(y) - \int_M \psi(T(x)) d\mu(x) \\
&= \int_D \psi(y) - \psi(T(x)) d\gamma(x, y) \\
&\leqslant \int_D c(x, y) - c(x, T(x)) - \omega(\mathrm{d}_N(T(x), y)) d\gamma(x, y) \\
&= \int_D c(x, y) d\gamma(x, y) - \int_M c(x, T(x)) d\mu(x) - \int_D \omega(\mathrm{d}_N(T(x), y)) d\gamma(x, y)
\end{aligned}$$

Rearranging this inequality gives the desired conclusion. □

We can rephrase this proposition using the the 1-Wasserstein distance W_1 in $\mathcal{P}(M \times N)$ induced by the distance

$$\mathrm{d}_{M\times N}((x, y), (x', y')) = \mathrm{d}_M(x, x') + \mathrm{d}_N(y, y').$$

Corollary 18. *Under the assumptions of Proposition 17, if the modulus of the Kantorovitch potential ψ is $\omega(r) = Cr^2$, one has*

$$\mathrm{W}_1(\gamma, \gamma_T) \leqslant \frac{1}{\sqrt{C}} \left(\int_{M\times N} c(x, y) d\gamma(x, y) - \int_M c(x, T(x)) d\mu(x) \right)^{1/2}$$

where $\gamma_T = (Id, T)_{\#}\mu$.

Proof. Let $S : M \times N \to (M \times N)^2$ defined by

$$S(x, y) = (S_1(x, y), S_2(x, y))$$

where $S_1(x, y) = (x, T(x))$ and $S_2(x, y) = (x, y)$. Let $\pi = S_{\#}\gamma \in \mathcal{P}((M \times N)^2)$. One can check that $\pi \in \Gamma(\gamma_T, \gamma)$, which implies

$$\begin{aligned}
W_1(\gamma_T, \gamma) &\leqslant \int_{(M\times N)^2} \mathrm{d}_{M\times N}((x, y), (x', y')) \mathrm{d}\pi(x, y, x', y') \\
&= \int_{M\times N} \mathrm{d}_{M\times N}(S_1(x, y), S_2(x, y)) \mathrm{d}\gamma(x, y) \\
&= \int_{M\times N} \mathrm{d}_N(T(x), y) \mathrm{d}\gamma(x, y).
\end{aligned}$$

We use the Cauchy-Schwarz inequality in $L^2(M \times N, \gamma)$ and Proposition 17 to get the desired result. □

3.2.3 Stability with respect to both measures

Here we apply Corollary 18 to show stability results of transport plans with respect to both the source and the target measures. Our result holds for general cost functions and is inspired by a result of Li and Nochetto [48] that holds in the quadratic case. We recall that d_M is the distance on M and d_N is the distance on N. We also choose for distance on the product space $\mathrm{d}_{M\times N}((x,y),(x',y')) = \mathrm{d}_M(x,x') + \mathrm{d}_N(y,y')$. Throughout this section, we require the cost function c to be Lipschitz on the whole product space $M \times N$.

Proposition 19 (Stability with respect to both measures)**.** *Let $\mu, \tilde{\mu} \in \mathcal{P}(M)$ and $\nu, \tilde{\nu} \in \mathcal{P}(N)$. Let $c : M\times N \to \mathbb{R}$ be a cost function which is Lipschitz on $M\times N$. Let $T : M \to N$ be an optimal transport map between μ and ν, and $\tilde{\gamma}$ be an optimal transport plan between $\tilde{\mu}$ and $\tilde{\nu}$ for the cost c. We assume that T is induced by a strongly c-concave potential $\psi : N \to \mathbb{R}$ with associated modulus $\omega(r) = Cr^2$ on $D = M \times N$. Then we have*

$$\mathrm{W}_1(\gamma_T, \tilde{\gamma}) \leqslant \varepsilon + \sqrt{\frac{2\mathrm{Lip}(c)}{C}\varepsilon}, \quad \text{where } \varepsilon := \mathrm{W}_1(\tilde{\mu}, \mu) + \mathrm{W}_1(\nu, \tilde{\nu}).$$

The end of this section is devoted to the proof of this proposition. As in [48], we will use the gluing lemma [63, 69].

Lemma 20 (gluing of measures)**.** *Let (X_i, μ_i) be probability spaces for $i \in \{1, 2, 3\}$, and $\gamma_{12} \in \Gamma(\mu_1, \mu_2)$, $\gamma_{23} \in \Gamma(\mu_2, \mu_3)$. Then there exists $\pi \in \mathcal{P}(X_1 \times X_2 \times X_3)$ such that $\pi(\cdot, \cdot, X_3) = \gamma_{12}$ and $\pi(X_1, \cdot, \cdot) = \gamma_{23}$. Or equivalently*

$$p_{12\#}\pi = \gamma_{12} \quad p_{23\#}\pi = \gamma_{23}$$

where p_{ij} is the projection defined by $p_{ij}(x_1, x_2, x_3) = (x_i, x_j)$.

We also need the following (easy) lemma, showing that the transport cost

$$\mathcal{T}^c(\mu, \nu) := \min_{\gamma\in\Gamma(\mu,\nu)} \int c\mathrm{d}\gamma$$

is Lipschitz with respect to perturbations of the measures when c is Lipschitz.

Lemma 21. *Let $c : M \times N \to \mathbb{R}$ be a Lipschitz cost function. Let $\mu, \tilde{\mu} \in \mathcal{P}(M)$ and $\nu, \tilde{\nu} \in \mathcal{P}(N)$. Then we have*

$$|\mathcal{T}^c(\mu, \nu) - \mathcal{T}^c(\tilde{\mu}, \tilde{\nu})| \leqslant \mathrm{Lip}(c)(\mathrm{W}_1(\mu, \tilde{\mu}) + \mathrm{W}_1(\nu, \tilde{\nu})).$$

Proof. Kantorovich duality gives

$$\mathcal{T}^c(\mu, \nu) = \max_{\varphi\oplus\psi\leqslant c} \int_M \varphi\mathrm{d}\mu + \int_N \psi\mathrm{d}\nu.$$

Moreover, since the cost is Lipschitz, the maximum is attained in the dual problem; one can assume that the maximum is attained for two potentials φ, ψ satisfying $\varphi = \psi^c$ and $\psi = \varphi^c$. In particular both φ and ψ are Lipschitz continuous with Lipschitz constant lower than $\mathrm{Lip}(c)$. Kantorovitch (weak) duality applied to the two measures $\tilde{\mu}$ and $\tilde{\nu}$ gives

$$\mathcal{T}^c(\tilde{\mu}, \tilde{\nu}) \geqslant \int_M \varphi\mathrm{d}\tilde{\mu} + \int_N \psi\mathrm{d}\tilde{\nu}.$$

We thus get

$$\mathcal{T}^c(\mu, \nu) - \mathcal{T}^c(\tilde{\mu}, \tilde{\nu}) \leqslant \int_M \varphi\mathrm{d}(\mu - \tilde{\mu}) + \int_N \psi\mathrm{d}(\nu - \tilde{\nu}) \leqslant \mathrm{Lip}(c)(\mathrm{W}_1(\mu, \tilde{\mu}) + \mathrm{W}_1(\nu, \tilde{\nu}))$$

where the last inequality is given by Kantorovich-Rubinstein Theorem. By symmetry the same result holds when we exchange μ, ν and $\tilde{\mu}, \tilde{\nu}$. □

Proof of Proposition 19. Let $\alpha \in \Gamma(\mu, \tilde{\mu})$ and $\beta \in \Gamma(\tilde{\nu}, \nu)$ be optimal transport plans for the cost d_M and d_N. Let $\pi \in \mathcal{P}(M^2 \times N^2)$ be a gluing of $\alpha, \tilde{\gamma}$ and β, i.e.

$$p_{12\#}\pi = \alpha, \quad p_{23\#}\pi = \tilde{\gamma}, \quad p_{34\#}\pi = \beta$$

Defining $\gamma = p_{14\#}\pi \in \Gamma(\mu, \nu)$, π is a transport map between γ and $\tilde{\gamma}$, and we get

$$\begin{aligned} \mathrm{W}_1(\gamma, \tilde{\gamma}) &\leqslant \int_{M^2 \times N^2} \mathrm{d}_M(x, x') + \mathrm{d}_N(y, y') \mathrm{d}\pi(x, x', y, y') \\ &= \int_{M^2} \mathrm{d}_M(x, x') \mathrm{d}\alpha(x, x') + \int_{N^2} \mathrm{d}_N(y, y') \mathrm{d}\beta(y, y') \\ &= \mathrm{W}_1(\tilde{\mu}, \mu) + \mathrm{W}_1(\nu, \tilde{\nu}) \end{aligned} \tag{3.2.4}$$

We also have

$$\begin{aligned} &\int_{M \times N} c(x, y) \mathrm{d}\gamma \\ &= \int_{M^2 \times N^2} c(x, y) \mathrm{d}\pi(x, x', y', y) \\ &= \int_{M^2 \times N^2} c(x', y') + c(x, y) - c(x', y') \mathrm{d}\pi(x, x', y', y) \\ &\leqslant \int_{M^2 \times N^2} c(x', y') + \mathrm{Lip}(c)(\mathrm{d}_M(x, x') + \mathrm{d}_N(y, y')) \mathrm{d}\pi(x, x', y', y) \\ &= \int_{M \times N} c(x', y') \mathrm{d}\tilde{\gamma} + \mathrm{Lip}(c) \left(\int_{M^2} \mathrm{d}_M(x, x') \mathrm{d}\alpha + \int_{N^2} \mathrm{d}_N(y, y') \mathrm{d}\beta \right) \\ &\leqslant \int_{M \times N} c(x, y) \mathrm{d}\tilde{\gamma} + \mathrm{Lip}(c)(W_1(\mu, \tilde{\mu}) + W_1(\nu, \tilde{\nu})) \end{aligned} \tag{3.2.5}$$

The transport plans $\gamma_T = (Id, T)_{\#}\mu \in \Gamma(\mu, \nu)$ and $\tilde{\gamma} \in \Gamma(\tilde{\mu}, \tilde{\nu})$ are optimal, so that by Lemma 21,

$$\int_{M \times N} c(x, y) \mathrm{d}\tilde{\gamma} \leqslant \int_{M \times N} c(x, y) \mathrm{d}\gamma_T + \mathrm{Lip}(c)(W_1(\mu, \tilde{\mu}) + W_1(\nu, \tilde{\nu}))$$

which combined with (3.2.5) gives

$$\int_{M \times N} c(x, y) \mathrm{d}\gamma - \int_{M \times N} c(x, y) \mathrm{d}\gamma_T \leqslant 2\mathrm{Lip}(c)(W_1(\mu, \tilde{\mu}) + W_1(\nu, \tilde{\nu}))$$

Corollary 18 then implies that

$$\mathrm{W}_1(\gamma, \gamma_T) \leqslant \left[\frac{2\mathrm{Lip}(c)}{C} (\mathrm{W}_1(\tilde{\mu}, \mu) + \mathrm{W}_1(\nu, \tilde{\nu})) \right]^{1/2}$$

Finally, using the triangle inequality along with (3.2.4) we get

$$\begin{aligned} \mathrm{W}_1(\tilde{\gamma}, \gamma_T) &\leqslant \mathrm{W}_1(\tilde{\gamma}, \gamma) + \mathrm{W}_1(\gamma, \gamma_T) \\ &\leqslant \mathrm{W}_1(\tilde{\mu}, \mu) + \mathrm{W}_1(\nu, \tilde{\nu}) + \left(\frac{2\mathrm{Lip}(c)}{C} (\mathrm{W}_1(\tilde{\mu}, \mu) + \mathrm{W}_1(\nu, \tilde{\nu})) \right)^{1/2} \\ &= \varepsilon + \sqrt{\frac{2\mathrm{Lip}(c)}{C}} \varepsilon \end{aligned}$$

□

3.2.4 A remark on regularity

The above results show that the notion of strong c-concavity is sufficient to get stability results. In fact, this notion can also lead to regularity of the associated transport maps, as expressed in the following lemma.

Lemma 22 (Regularity under strong c-concavity)**.** *Let us assume that the cost function $c : M \times N \to \mathbb{R}$ is Lipschitz on $M \times N$ and let $T : M \to N$ be a transport map induced by a strongly c-concave potential $\psi : N \to \mathbb{R}$, with continuity modulus $\omega(r) = Cr^2$ on $M \times N$. Then T is $1/2$-Hölder:*

$$\mathrm{d}_N(T(x), T(x')) \leqslant \left(\frac{\mathrm{Lip}(c)}{C}\mathrm{d}_M(x, x')\right)^{1/2}$$

Proof. Let $x \in M$. Since T is induced by a strongly c-concave potential ψ we have $T(x) = \mathrm{argmin}_{y\in N}\, c(x, y) - \psi(y)$. The strong c-concavity of ψ implies that for every $y \in N$

$$c(x, y) - \psi(y) \geqslant c(x, T(x)) - \psi(T(x)) + \omega(\mathrm{d}_N(y, T(x)))$$

Now let $x' \in M$. By choosing $y = T(x')$ the above inequality becomes

$$c(x, T(x')) - \psi(T(x')) \geqslant c(x, T(x)) - \psi(T(x)) + \omega(\mathrm{d}_N(T(x'), T(x)))$$

This inequality still holds when we exchange x and x', summing the two gives

$$2\omega(\mathrm{d}_N(T(x), T(x'))) \leqslant c(x', T(x)) + c(x, T(x')) - c(x', T(x')) - c(x, T(x))$$

and since c Lipschitz we have

$$C\mathrm{d}_N(T(x), T(x'))^2 \leqslant \mathrm{Lip}(c)\mathrm{d}_M(x, x'). \qquad \square$$

Thus, strong c-concavity of the potential entails some regularity of the transport map, generalizing what is well-known in the convex setting (i.e. if ψ is strongly convex, then ψ^* is $\mathcal{C}^{1,1}$). The next section will show a partial converse statement, under strong assumptions on the cost function.

3.3 Sufficient condition for strong c-concavity

This section is about sufficient conditions for establishing strong c-concavity, which we used through the previous section to deduce stability results of optimal transport maps. From now on, we assume that M and N are smooth complete Riemannian manifolds.

It is well known that the notions of convexity and strong convexity can be easily characterized by conditions on the Hessian for smooth functions. The c-convexity is not that easy to study but for cost functions c that are regular enough in a certain sense, there exists a differential criterion for c-convexity, given by Villani [69]. In this section we extend Villani's statement for strong c-concavity, in other words we show that the strong c-concavity of a function can also be guaranteed by conditions on its derivatives. This result is presented in Corollary 25. To do so we need the cost function $c : M \times N \to \mathbb{R} \cup \{+\infty\}$ to satisfy the Ma-Trudinger-Wang (MTW) condition, which is a well known condition in the regularity theory of optimal transport.

3.3.1 The Ma-Trudinger-Wang tensor

We recall in this section the notion of MTW tensor [69]. Recall that we are working with two smooth complete Riemannian manifolds M and N, and a cost function $c : M \times N \to \mathbb{R} \cup \{+\infty\}$. We denote by $\mathrm{Dom}(\nabla_x c) \subseteq M \times N$ the domain of differentiability of the cost c and $\mathrm{Dom}'(\nabla_x c(x,\cdot)) = \mathrm{int}(\mathrm{Dom}(\nabla_x c(x,\cdot)))$ its interior, then

$$\mathrm{Dom}'(\nabla_x c) = \{(x,y) \mid x \in \mathrm{int}(M), y \in \mathrm{Dom}'(\nabla_x c(x,\cdot)))\} \tag{3.3.6}$$

Definition 15 (Twisted cost)**.** *The cost c satifies the (Twist) condition if $\nabla_x c(x,\cdot)$ is injective on its domain of definition, i.e. for any x, y, y' such that $(x,y) \in \mathrm{Dom}'(\nabla_x c)$ and $(x,y') \in \mathrm{Dom}'(\nabla_x c)$:*

$$\nabla_x c(x,y) = \nabla_x c(x,y') \implies y = y'$$

Definition 16 (STwist)**.** *The cost satisfies the strong Twist condition (STwist) if c is $\mathcal{C}^2$, $\nabla_x c$ is one-to-one and $D^2_{xy} c$ is non singular on* $\mathrm{Dom}'(\nabla_x c)$.

If the cost function satisfies (Twist), then for $x \in \mathrm{int}(M)$ the function $-\nabla_x c(x,\cdot)$ is invertible on its image, i.e.

$$-\nabla_x c(x,\cdot) : \mathrm{Dom}'(\nabla_x c(x,\cdot)) \subseteq N \to \mathfrak{I}_x \subseteq T_x M$$

is one-to-one, with $\mathfrak{I}_x = \{-\nabla_x c(x,y) | y \in \mathrm{Dom}'(\nabla_x c(x,\cdot))\}$.

Definition 17 (c-exponential)**.** *When the cost c satisfies the (Twist) condition, we can define the c-exponential map for $x \in M$ by* $\text{c-exp}_x = (-\nabla_x c(x,\cdot))^{-1}$*, giving for $p \in \mathfrak{I}_x$:*

$$\begin{aligned} \text{c-exp}_x(p) : \mathfrak{I}_x \subseteq T_x M &\to \mathrm{Dom}'(\nabla_x c(x,\cdot)) \subseteq N \\ p &\to \left(\nabla_x c(x,\cdot)\right)^{-1}(-p) \end{aligned}$$

Definition 18 (c-segment)**.** *A c-segment is the image of a usual segment in $\mathfrak{I}_x$ by the map* c-exp_x*. We denote $(y_t)_{0 \leqslant t \leqslant 1} = [y_0, y_1]_x$ the c-segment between y_0 and y_1 with base x defined for $p_0 = -\nabla_x c(x, y_0)$ and $p_1 = -\nabla_x c(x, y_1)$ by*

$$y_t = \text{c-exp}_x((1-t)p_0 + t p_1)$$

Definition 19 (c-convex set)**.** *Let $A \subseteq N$.*

- *We say that A is c-convex with respect to $x \in M$ if for any $y_0, y_1 \in A$, there is a c-segment $[y_0, y_1]_x$ entirely contained in A.*
- *The set A is said to be c-convex with respect to a set $B \subseteq M$ if A is c-convex with respect to any $x \in B$.*
- *A set $D \subseteq M \times N$ is said to be totally c-convex if for any two points $(x, y_0) \in D$ and $(x, y_1) \in D$, the c-segment $(y_t)_{0 \leqslant t \leqslant 1} = [y_0, y_1]_x$ satisfies for any t $(x, y_t) \in D$.*
- *We say that $D \subseteq M \times N$ is symmetrically c-convex if it is totally c-convex and if for any two points $(x_0, y) \in D$ and $(x_1, y) \in D$, $[x_0, x_1]_y \times \{y\} \subseteq D$.*

Definition 20 (MTW tensor)**.** *Assuming that c is of class $\mathcal{C}^4$ on* $\mathrm{Dom}'(\nabla_x c)$ *and satisfies the (STwist) condition, the Ma-Trudinger-Wang tensor is defined for $(x_0, y_0) \in \mathrm{Dom}'(\nabla_x c)$ and $(\eta, \zeta) \in T_{x_0} M \times T_{y_0} N$ by*

$$\mathfrak{S}_c(x_0, y_0)(\eta, \zeta) = -\frac{3}{2} \frac{\partial^2}{\partial q_{\tilde{\eta}}^2} \frac{\partial^2}{\partial y_\zeta^2} \big(c(\text{c-exp}_{y_0}(q), y)\big)\Big|_{y = y_0, q = -\nabla_y c(x_0, y_0)}$$

with $\tilde{\eta} = -\nabla^2_{xy} c(x_0, y_0)\eta \in T_{y_0} N$.

In the above definition, we use Villani's notation to indicate that the cost is differentiated twice with respect to q in the direction $\tilde{\eta}$ and twice with respect to y in the direction ζ. It is a slight abuse of notation since this differentiation should depend on the choice of coordinates. However the operator $\mathfrak{S}_c$ is independent of the choice of coordinates, even though it cannot be directly seen in the above formula. This formulation of the MTW tensor can be found in [69, Remarks 12.31 & 12.33]. In this definition $-\nabla^2_{xy}c(x_0,y_0) : T_{x_0}M \times T_{y_0}N \to \mathbb{R}$ is a bilinear form which is non singular since (STwist) is satisfied. We then identify for $\eta \in T_xM$ the linear form $-\nabla^2_{xy}c(x_0,y_0)\eta = \tilde{\eta} : T_{y_0}N \to \mathbb{R}$ with a vector of $T_{y_0}N$ using the Riemannian structure.

Definition 21 (weak MTW)**.** *We say that the weak MTW condition (MTWw) is satisfied on a compact set $D \subseteq M \times N$ if there exists a constant $C > 0$ such that for any $(x,y) \in D$ and $(\eta,\zeta) \in T_xM \times T_yN$ we have*

$$\mathfrak{S}_c(x,y)(\eta,\zeta) \geqslant -C|\langle\zeta|\tilde{\eta}\rangle|\,\|\zeta\|\,\|\eta\| \tag{MTWw}$$

This condition was introduced by Ma, Trudinger and Wang [51] and is often referred to as (A3w)*.*

There exists a geometric interpretation of the MTW hypothesis that is given in terms of curvature proposed by Kim–McCann [43].

3.3.2 Differential criterion for strong c-concavity

The goal here is to generalize Trudinger and Wang's differential criterion [65, 69] (detailed in the following theorem) for c-convexity to our definition of strong c-concavity. Our proof is highly inspired from Villani's one, in particular we study the same real valued function $h : [0,1] \to \mathbb{R}$ and show inequalities that are similar and also require positivity of the MTW tensor.

Theorem 23 (Differential criterion for c-convexity, [69, Th. 12.46][65])**.** *Let $D \subseteq M \times N$ be a closed symmetrically c-convex set and $c \in \mathcal{C}^4(D,\mathbb{R})$ such that c and $\check{c}$ satisfy (STwist) on D. Assume that the weak MTW condition* (MTWw) *is satisfied on D. Let $\mathcal{X} = \mathrm{proj}_M(D)$ and $\psi \in \mathcal{C}^2(\mathcal{X},\mathbb{R})$. If for any $x \in \mathcal{X}$ there exists $y \in N$ such that $(x,y) \in D$ and*

$$\begin{cases} \nabla\psi(x) + \nabla_x c(x,y) = 0 \\ D^2\psi(x) + D^2_{xx}c(x,y) \geqslant 0 \end{cases}$$

Then ψ is c-convex on D, or equivalently $-\psi$ is c-concave.

This theorem is given for a potential function ψ on $\mathcal{X} \subseteq M$ and gives a c-convexity result while we consider $\psi : N \to \mathbb{R}$ and work on c-concavity, but this is really just a matter of convention. Also Villani needs the Hessian $D^2\psi(x) + D^2_{xx}c(x,y)$ to be positive semi-definite to obtain c-convexity, while we are naturally going to need the Hessian $D^2_{yy}c(x,y) - D^2\psi(y)$ to have eigenvalues bounded from below by a positive constant to obtain strong c-concavity. A noticeable difference of c-convexity with respect to convexity is that it cannot be expressed locally, as we require the MTW tensor to be positive on the whole set D which is a global condition.

Theorem 24 (Differential criterion for strong c-concavity)**.** *We consider $D \subseteq \mathrm{Dom}'(\nabla_x c) \cap \mathrm{Dom}'(\nabla_y c)$ a symmetrically c-convex compact set and denote $\mathcal{X} = \mathrm{proj}_M(D)$, $\mathcal{Y} = \mathrm{proj}_N(D)$. We assume that $c \in \mathcal{C}^4(D,\mathbb{R})$, that c and $\check{c}$ satisfy (STwist) on D where*

$c(x,y) = \check{c}(y,x)$. *We also assume that the weak MTW condition is satisfied on* D. *Let* $\psi \in \mathcal{C}^2(\mathcal{Y}, \mathbb{R})$ *be a c-concave function on* D *and such that there exists* $\lambda > 0$ *satisfying for any* $x \in \partial^c \psi(y)$

$$D^2_{yy} c(x,y) - D^2\psi(y) \geqslant \lambda Id$$

Then ψ *is strongly c-concave on* D *with modulus* $\omega(\mathrm{d}_N(\overline{y},y)) = C\mathrm{d}_N(\overline{y},y)^2$, *where* $C > 0$ *is a constant depending on* λ, c, $\mathcal{X}$ *and* $\mathcal{Y}$. *This means that we have*

$$\psi(y) - c(\overline{x},y) \leqslant \psi(\overline{y}) - c(\overline{x},\overline{y}) - C\mathrm{d}_N(\overline{y},y)^2$$

for the points $\overline{x} \in \mathcal{X}$, $\overline{y}, y \in \mathcal{Y}$ *such that* $\overline{x} \in \partial^c\psi(\overline{y})$, $(\overline{x},\overline{y}) \in D$ *and* $(\overline{x},y) \in D$.

Corollary 25 (Strong c-concavity)**.** *We make the same hypothesis on* c *and* D, *and in addition assume* $\psi \in \mathcal{C}^2(\mathcal{Y},\mathbb{R})$. *We assume that the map* $T : \mathcal{X} \to \mathcal{Y}$ *defined by* $T(x) = \mathrm{argmin}_y\, c(x,y) - \psi(y)$ *is of class* $\mathcal{C}^1$ *and satisfies for any* $x \in \mathcal{X}$, $(x, T(x)) \in D$. *Then the function* ψ *is strongly c-concave on the set* D *with modulus* $\omega(\mathrm{d}_N(\overline{y},y)) = C\mathrm{d}_N(\overline{y},y)^2$.

Remark 26 (Restriction of c-concavity to D)**.** *In the Corollary 25, we assume that the graph of* T *is supported on a set* D *where the cost function is smooth enough. This can be an issue for transport maps* $T : M \to N$ *of the form* $T(x) = \mathrm{argmin}_{z\in N}\, c(x,y) - \psi(y)$, *since we cannot ensure that the argmin is obtained at a point* y *such that* $(x,y) \in D$. *This issue has to be treated independently for each application.*

3.3.3 Proof of Theorem 24.

We denote $\mathcal{Y}^x = \{y \in N \mid (x,y) \in D\}$ for any $x \in \mathcal{X}$. Let now fix $\overline{y} \in \mathcal{Y}$ and $\overline{x} \in \partial^c\psi(\overline{y})$ such that $(\overline{x},\overline{y}) \in D$. Note that $\overline{x}$ always exists by hypothesis. Let us fix $y \in \mathcal{Y}^{\overline{x}}$. We want to show that there exists a constant $C > 0$ independant of $\overline{x}, \overline{y}$ and y such that

$$c(\overline{x},y) - \psi(y) \geqslant c(\overline{x},\overline{y}) - \psi(\overline{y}) + C\mathrm{d}_N(y,\overline{y})^2 \tag{3.3.7}$$

We put $(y_t)_{0\leqslant t\leqslant 1} = [\overline{y}, y]_{\overline{x}}$ the c-segment between $\overline{y}$ and y with base $\overline{x}$. Remark that the c-convexity of D implies that for any t in $[0,1]$, $(\overline{x}, y_t) \in D$. We define the function h by

$$h(t) := c(\overline{x}, y_t) - \psi(y_t)$$

such that Equation (3.3.7) writes

$$h(1) \geqslant h(0) + C\mathrm{d}_N(y,\overline{y})^2 \tag{3.3.8}$$

The end of this section is devoted to the proof of Equation (3.3.8).

Notation. We first introduce some notations. Note that $A_{\overline{x}} := \nabla^2_{xy} c(\overline{x}, y_t) : T_{\overline{x}}M \times T_{y_t}N \to \mathbb{R}$ is a bilinear form which is assumed to be nonsingular. For any $X \in T_{\overline{x}}M$ and $Y \in T_{y_t}N$, we can write $\nabla^2_{xy}c(\overline{x},y_t)(X,Y) = \langle A_{\overline{x}}X|Y\rangle = \langle {}^tA_{\overline{x}}Y|X\rangle$ where in some local coordinates $A_{\overline{x}}$ is an invertible matrix and X and Y are column matrices. Then $\nabla^2_{xy}c(\overline{x},y_t)(X,\cdot)$ is a linear form on $T_{y_t}N$ which is identified to the vector $A_{\overline{x}}X \in T_{y_t}N$. Similarly ${}^tA_{\overline{x}}Y \in T_{\overline{x}}M$. We take the same notation for $A_{x_s} = \nabla^2_{xy}c(x_s,y_t)$.

Lemma 27 (Formula for the derivatives of h)**.**

$$h'(t) = \langle \zeta | \hat{\eta} \rangle$$

and

$$h''(t) = \Big(\mathrm{D}^2_{yy} c(x^t, y_t) - D^2\psi(y_t)\Big)(\hat{\eta}, \hat{\eta}) + \frac{2}{3}\int_0^1 \mathfrak{S}_c(x_s, y_t)(\overline{\zeta}, \hat{\eta})(1-s)\mathrm{d}s,$$

where $x^t \in \partial^c\psi(y_t)$ *and* $x_s = \text{c-exp}_{y_t}(\overline{q_t} + s\zeta)$.

$$\begin{array}{ll} \eta = \nabla_x c(\overline{x}, \overline{y}) - \nabla_x c(\overline{x}, y) \in T_{\overline{x}}M & \hat{\eta} = -{}^t A_{\overline{x}}^{-1}\eta \in T_{y_t}N \\ \zeta = \nabla_y c(\overline{x}, y_t) - \nabla\psi(y_t) \in T_{y_t}N & \hat{\zeta} = -A_{\overline{x}}^{-1}\zeta \in T_{\overline{x}}M \\ \overline{q_t} := -\nabla_y c(\overline{x}, y_t) \in T_{y_t}M & \overline{\zeta} = -A_{x_s}^{-1}\zeta \in T_{x_s}M \end{array}$$

Note that in the above lemma, x^t is not necessarily a c-segment, while x_s is the c-segment between $\overline{x}$ and $\text{c-exp}_{y_t}(-\nabla\psi(y_t)) = x^t$ with base y_t.

Proof of Lemma 27. Since D is symmetrically c-convex and $(\overline{x}, \overline{y}) \in D$, $(\overline{x}, y) \in D$, we can differentiate h as follows

$$h'(t) = \langle \nabla_y c(\overline{x}, y_t) - \nabla\psi(y_t) | \dot{y_t} \rangle$$

We also have by differentiating $-\nabla_x c(\overline{x}, y_t) = \overline{p} + t\eta$, where $\overline{p} = -\nabla_x c(\overline{x}, \overline{y})$:

$$\eta = -\nabla^2_{xy} c(\overline{x}, y_t)\dot{y_t} = -{}^t A_{\overline{x}}\dot{y_t}$$

So that $\hat{\eta} = -{}^t A_{\overline{x}}^{-1}\eta = \dot{y_t}$ and thus

$$h'(t) = \langle \zeta | \hat{\eta} \rangle.$$

Differentiating h' gives

$$h''(t) = \Big(\nabla^2_{yy} c(\overline{x}, y_t) - \nabla^2\psi(y_t)\Big)(\dot{y_t}, \dot{y_t}) + \langle \zeta | \ddot{y_t} \rangle.$$

By differentiating $-\eta = \nabla^2_{xy} c(\overline{x}, y_t)\dot{y_t}$, one gets

$$\nabla^3_{xyy} c(\overline{x}, y_t)(\dot{y_t}, \dot{y_t}) + \nabla^2_{xy} c(\overline{x}, y_t)\ddot{y_t} = 0$$

so that

$$\ddot{y_t} = -{}^t A_{\overline{x}}^{-1} \nabla^3_{xyy} c(\overline{x}, y_t)(\hat{\eta}, \hat{\eta})$$

and

$$\langle \zeta | \ddot{y_t} \rangle = \langle \zeta | -{}^t A_{\overline{x}}^{-1} \nabla^3_{xyy} c(\overline{x}, y_t)(\hat{\eta}, \hat{\eta}) \rangle = \langle -A_{\overline{x}}^{-1}\zeta | \nabla^3_{xyy} c(\overline{x}, y_t)(\hat{\eta}, \hat{\eta}) \rangle.$$

We therefore have

$$h''(t) = \Big(\nabla^2_{yy} c(\overline{x}, y_t) - \nabla^2\psi(y_t)\Big)(\hat{\eta}, \hat{\eta}) + \langle \hat{\zeta} | \nabla^3_{xyy} c(\overline{x}, y_t)(\hat{\eta}, \hat{\eta}) \rangle$$

We define $\Phi(x) := \Big(\nabla^2_{yy} c(x, y_t) - \nabla^2\psi(y_t)\Big)(\hat{\eta}, \hat{\eta})$. Then we have for $X \in T_x M$

$$D\Phi(x).X = \langle X | \nabla^3_{xyy} c(x, y_t)(\hat{\eta}, \hat{\eta}) \rangle,$$

so that

$$h''(t) = \Phi(\overline{x}) + D\Phi(\overline{x})\hat{\zeta}$$

We define $\tilde{\Phi}(q) = \Phi(\text{c-exp}_{y_t}(q))$ or equivalently $\tilde{\Phi}(-\nabla_y c(x, y_t)) = \Phi(x)$, so that for $X \in T_x M$

$$D\Phi(x)X = D\tilde{\Phi}(\overline{q_t})(-A_x X)$$

For $x = \overline{x}$ and $\zeta \in T_{y_t}N$

$$D\Phi(\overline{x})\hat{\zeta} = D\tilde{\Phi}(\overline{q_t})(-A_{\overline{x}}\hat{\zeta}) = D\tilde{\Phi}(\overline{q_t})\zeta$$

We set $q_t := -\nabla_y c(x^t, y_t)$. Since $x^t \in \partial^c\psi(y_t)$, we have $\nabla\psi(y_t) = \nabla_y c(x^t, y_t) = -q_t$. We recall that $\overline{q_t} = -\nabla_y c(\overline{x}, y_t)$ and therefore we get $\zeta = q_t - \overline{q_t}$. Since the set D is c-convex we can differentiate c at $(\text{c-exp}_{y_t}(\overline{q_t} + s\zeta), y_t) = (x_s, y_t)$, we get using a Taylor expansion of $\tilde{\Phi}$ at $\overline{q_t}$

$$h''(t) = \tilde{\Phi}(\overline{q_t}) + D\tilde{\Phi}(\overline{q_t})(q_t - \overline{q_t}) = \tilde{\Phi}(q_t) - \int_0^1 D^2_{qq}\tilde{\Phi}(\overline{q_t} + s\zeta)(\zeta,\zeta)(1-s)\mathrm{d}s$$

Using the change of variable $q = -\nabla_y c(x, y_t) \in T_{y_t}M$ (or equivalently $x = \text{c-exp}_{y_t}(q)$), we get

$$\tilde{\Phi}(q) = \Phi(\text{c-exp}_{y_t}(q)) = \Big(\nabla^2_{yy}c(\text{c-exp}_{y_t}(q), y_t) - \nabla^2\psi(y_t))\Big)(\hat{\eta},\hat{\eta}) \tag{3.3.9}$$

Since $q_t = -\nabla_y c(x^t, y_t)$ we have

$$\tilde{\Phi}(q_t) = \Phi(x^t) = \Big(\nabla^2_{yy}c(x^t, y_t) - \nabla^2\psi(y_t)\Big)(\hat{\eta},\hat{\eta})$$

Moreover $\nabla^2\psi(y_t)$ does not depend on q, so we have by differenciating (3.3.9) twice with respect to q in the direction ζ

$$D^2_{qq}\tilde{\Phi}(q)(\zeta,\zeta) = \frac{\partial^2}{\partial q^2_\zeta}\frac{\partial^2}{\partial y^2_{\hat{\eta}}}\big(c(\text{c-exp}_{y_t}(q), y_t)\big)$$

Finally, using $x_s = \text{c-exp}_{y_t}(\overline{q_t} + s\zeta)$, we get for any $s \in [0,1]$

$$D^2_{qq}\tilde{\Phi}(\overline{q_t} + s\zeta)(\zeta,\zeta) = \frac{\partial^2}{\partial q^2_\zeta}\frac{\partial^2}{\partial y^2_{\hat{\eta}}}\big(c(\text{c-exp}_{y_t}(\overline{q_t} + s\zeta), y_t)\big) = -\frac{2}{3}\mathfrak{S}_c(x_s, y_t)(\overline{\zeta}, \hat{\eta})$$

where we put $\overline{\zeta} := -A^{-1}_{x_s}\zeta$ so as to have $\overline{\tilde{\zeta}} = \zeta$.

□

To finish the proof we need the following elementary lemma.

Lemma 28. *Let $y \in \mathcal{C}^1([0,1],\mathbb{R})$ satisfying for $C > 0$,*

$$\begin{cases} y'(t) \geqslant -C|y(t)| \\ y(0) = 0 \end{cases}$$

then $y(t) \geqslant 0$ for any $t \in [0,1]$.

Proof. We remark that there exists $g \in \mathcal{C}^0([0,1],\mathbb{R}_+)$ such that y is solution of

$$\begin{cases} y'(t) = -C|y(t)| + g(t) \\ y(0) = 0 \end{cases}$$

Then by Cauchy-Lipschitz Theorem the unique solution of this equation is $t \mapsto \int_0^t g(s)e^{C(s-t)}ds \geqslant 0$.

□

Proposition 29. *Under hypothesis of Theorem 24,*

$$h''(t) \geqslant -Ch'(t) + \lambda\|\hat{\eta}\|^2$$

Proof. We have

$$|h'(t)| = |\langle\zeta|\hat{\eta}\rangle|.$$

We also have

$$h''(t) = \Big(\mathrm{D}^2_{yy}c(x^t, y_t) - D^2\psi(y_t)\Big)(\hat{\eta}, \hat{\eta}) + \frac{2}{3}\int_0^1 \mathfrak{S}_c(x_s, y_t)(\overline{\zeta}, \hat{\eta})(1-s)\mathrm{d}s,$$

where $x_s = \text{c-exp}_{y_t}(\overline{q_t} + s\zeta)$. By hypothesis we have

$$\Big(\mathrm{D}^2_{yy}c(x^t, y_t) - D^2\psi(y_t)\Big)(\hat{\eta}, \hat{\eta}) \geqslant \lambda\,\|\hat{\eta}\|^2$$

and (MTWw) gives

$$\mathfrak{S}_c(x_s, y_t)(\overline{\zeta}, \hat{\eta}) \geqslant -C|\langle\nabla^2_{xy}c(x_s, y_t)\hat{\eta}|\overline{\zeta}\rangle|\|\hat{\eta}\|\|\overline{\zeta}\|$$

The norms $\|\hat{\eta}\|$ and $\|\overline{\zeta}\|$ can be integrated in the constant by compactness, so we get

$$\mathfrak{S}_c(x_s, y_t)(\overline{\zeta}, \hat{\eta}) \geqslant -C|\langle\nabla^2_{xy}c(x_s, y_t))\hat{\eta}|\overline{\zeta}\rangle| = -C|\langle{}^tA_{x_s}\hat{\eta}|\overline{\zeta}\rangle|$$

Recall that $\overline{\zeta} = -\,A^{-1}_{x_s}\zeta$. Therefore we get

$$|\langle{}^tA_{x_s}\hat{\eta}|\overline{\zeta}\rangle| = |\langle{}^tA_{x_s}\hat{\eta}|A^{-1}_{x_s}\zeta\rangle| = |\langle\zeta|\hat{\eta}\rangle| = |h'(t)|,$$

We thus have $h''(t) \geqslant -C|h'(t)| + \lambda\|\hat{\eta}\|^2$. Note that $\zeta|_{t=0} = 0$ so $h'(0) = 0$. Then we can apply Lemma 28 to h', which gives $h'(t) \geqslant 0$, so we can drop the absolute value and we obtain $h''(t) \geqslant -Ch'(t) + \lambda\|\hat{\eta}\|^2$. □

Proof of Theorem 24. By compactness we have

$$C_1 := \inf_{(x,y)\in D, u\in T_xM, \|u\|=1} \left\|\nabla^2_{xy}c(x,y)^{-1}u\right\|^2 > 0$$

and

$$C_2 := \inf_{x\in\mathcal{X}, y,z\in\mathcal{Y}^x} \frac{\|\nabla_x c(x,y) - \nabla_x c(x,z)\|^2}{\mathrm{d}_N(y,z)^2} > 0$$

such that $\|\hat{\eta}\|^2 \geqslant C_1C_2\mathrm{d}_N(y,\overline{y})^2$. By Proposition 29, we get

$$h''(t) \geqslant -Ch'(t) + \lambda C_1C_2\mathrm{d}_N(y,\overline{y})^2$$

Using Grönwall's Lemma we then have that $h'(t) \geqslant g(t)$ with g solution of

$$\begin{cases} g'(t) = -Cg(t) + \lambda C_1C_2\mathrm{d}_N(y,\overline{y})^2 \\ g(0) = 0 \end{cases}$$

which immediatly gives $g(t) = \left(\frac{\lambda C_1C_2}{C}\mathrm{d}_N(y,\overline{y})^2\right)(1-e^{-Ct})$, so finally we have for $t\in[0,1]$, $h'(t) \geqslant \left(\frac{\lambda C_1C_2}{C}\mathrm{d}_N(y,\overline{y})^2\right)(1-e^{-Ct})$, and then by integrating for $t\in[0,1]$, there exists a constant $C_3 > 0$ such that

$$\int_0^1 h'(t)dt \geqslant C_3\mathrm{d}_N(y,\overline{y})^2$$

which is exactly what we wanted in Equation (3.3.8). □

Proof of Corollary 25. We want to show that under the hypothesis of Corollary 25, we have

$$\forall y \in \mathcal{Y}\ \forall x \in \partial^c \psi(y),\ D^2_{yy} c(x,y) - D^2\psi(y) \geqslant \lambda Id,$$

We recall that $T : \mathcal{X} \to \mathcal{Y}$ is of class $\mathcal{C}^1$. Let $x \in \mathcal{X}$, we first assume that $T(x) \in \text{int}(\mathcal{Y})$. Since $T(x)$ minimizes $c(x,\cdot) - \psi(\cdot)$ we have

$$\nabla_y c(x, T(x)) - \nabla\psi(T(x)) = 0 \tag{3.3.10}$$

and

$$D^2_{yy} c(x, T(x)) - D^2\psi(T(x)) \geqslant 0 \tag{3.3.11}$$

By differentiating (3.3.10) with respect to x, we get

$$\left(D^2_{yy} c(x, T(x)) - D^2\psi(T(x))\right) \circ DT(x) = -D^2_{xy} c(x, T(x)). \tag{3.3.12}$$

By (STwist) assumption, $D^2_{xy} c(x, T(x))$ is nonsingular, which implies that $D^2_{yy} c(x, T(x)) - D^2\psi(T(x))$ is also nonsingular. Since we also know that it is positive semi-definite from (3.3.11) we get that

$$D^2_{yy} c(x, T(x)) - D^2\psi(T(x)) > 0.$$

We now need to extend this inequality for any $T(x) \in \partial\mathcal{Y}$, including the boundary. By continuity, since ψ is $\mathcal{C}^2$ on $\mathcal{Y}$, c is $\mathcal{C}^2$ on D and T is $\mathcal{C}^1$ on $\mathcal{X}$, Equations (3.3.11) and (3.3.12) still hold when $T(x) \in \partial\mathcal{Y}$. Moreover (STwist) being satisfied on D, we have $D^2_{yy} c(x, T(x)) - D^2\psi(T(x)) > 0$ for any $x \in \mathcal{X}$. By compactness of $\mathcal{X}$, there exists $\lambda > 0$ such that

$$\forall x \in \mathcal{X} \quad D^2_{yy} c(x, T(x)) - D^2\psi(T(x)) \geqslant \lambda Id.$$

We conclude using that $T(x) = y$ is equivalent to $x \in \partial^c\psi(y)$. □

3.4 Stability of optimal transport map for MTW cost

In this section, we show that the stability results of Section 3.2 can be applied to optimal transport maps. We consider two compact Riemannian manifolds M and N in $\mathbb{R}^d$ and still denote by d_N the distance on N.

Theorem 30 (Stability in optimal transport)**.** *Let $\mu \in \mathcal{P}(M)$ and $\nu \in \mathcal{P}(N)$ be two probability measures. Let $c : M \times N \to \mathbb{R}$ be a cost function of class $\mathcal{C}^4$ that satisfies* (STwist) *and* (MTWw) *hypothesis. Let $T : M \to N$ be an optimal transport map between μ and ν of class $\mathcal{C}^1$ for the cost c and assume that its associated Kantorovich potential $\psi : N \to \mathbb{R}$ is of class $\mathcal{C}^2$.*

- *Let $\tilde{\nu} \in \mathcal{P}(N)$ be any probability measure, and $S : M \to N$ be an optimal transport map between μ and $\tilde{\nu}$. Then we have*

  $$\|\mathrm{d}_N(T, S)\|^2_{L^2(\mu)} \leqslant C W_1(\nu, \tilde{\nu})$$

  where W_1 denotes the 1-Wasserstein distance and C is a constant depending on the cost c, M and N.

- *Let $\tilde{\mu} \in \mathcal{P}(M)$, $\tilde{\nu} \in \mathcal{P}(N)$ and $\tilde{\gamma}$ be an optimal transport plan between $\tilde{\mu}$ and $\tilde{\nu}$. Then we have*

  $$\mathrm{W}_1(\tilde{\gamma}, \gamma_T) \leqslant C\left(\mathrm{W}_1(\tilde{\mu}, \mu) + \mathrm{W}_1(\nu, \tilde{\nu})\right)^{1/2}$$

  where $\gamma_T = (Id, T)_{\#}\mu$ and C is a constant depending on the cost c, M and N.

Proof. Since M and N are compact we have strong duality with a cost c that is Lipschitz on $M \times N$ so S is induced by a Lipschitz potential. Since $T \in \mathcal{C}^1$, $\psi \in \mathcal{C}^2$ and $c \in \mathcal{C}^4$ satisfies (STwist) and (MTWw), we can then apply Corollary 25 to ψ, which gives that it is strongly c-concave on N, with modulus $\omega(\mathrm{d}_N(y,z)) = C\mathrm{d}_N(y,z)^2$. Then the first result is given by Theorem 14 and the second is given by Proposition 19. □

For simplicity, the above theorem is stated in a restrictive way as it requires c to be smooth on the whole product space $M \times N$. It may happen that the regularity conditions such as (STwist) and (MTWw) are not satisfied on the whole product space $M \times N$, but only on a subset $D \subseteq M \times N$. In this case we can still obtain stability with respect to the target measure if we can show that optimal transports plans are supported on this subset D. This is treated independently on examples of Sections 3.5 and 3.6.

3.5 Stability for the reflector cost on the sphere

In this section, we apply a stability result of Section 3.2 to the far-field point reflector problem (FF-point) presented in Section 2.4. We have seen that this problem amounts to solving an optimal transport problem on the unit sphere $M = N = \mathcal{S}^{d-1}$ for the cost function $c(x,y) = -\ln(1 - \langle x|y\rangle)$ [71], extended by $+\infty$ on the diagonal $\{x = y\}$. One of the key element in the proof is to show that optimal transport maps are supported on compact sets that avoid the diagonal

$$D_\varepsilon = \{(x,y) \in M^2 \mid \mathrm{d}_M(x,y) \geqslant \varepsilon\} \tag{3.5.13}$$

where d_M is the geodesic distance on M. We first need the following definition.

Definition 22. *Given a probability measure $\mu \in \mathcal{P}(M)$, we put*

$$M_\mu(r) = \sup_{x \in M} \mu(\mathrm{B}(x,r)).$$

Theorem 31. *Let $c(x,y) = -\ln(1 - \langle x|y\rangle)$ be the reflector cost on the sphere $M = \mathcal{S}^{d-1}$. Let $\mu, \nu_0 \in \mathcal{P}(M)$ be absolutely continuous with respect to the Lebesgue measure with strictly positive $\mathcal{C}^{1,1}$ densities. Then for all $\beta > 0$, there exists a constant $C > 0$ depending on μ, ν_0 and β such that*

$$\forall \nu_1 \in \mathcal{P}(M) \text{ s.t. } M_{\nu_1}(\beta) < 1/8, \quad \|\mathrm{d}_M(T_0, T_1)\|^2_{L^2(\mu)} \leqslant C\ W_1(\nu_0, \nu_1)$$

where d_M is the geodesic distance on M and T_i be optimal transport maps between μ and ν_i.

The main difficulty to prove the previous theorem is to show that the optimal transport plan is supported on the compact set D_ε for some ε. This is done in the following subsection in a more general setting.

3.5.1 Support of the optimal transport plan

In this subsection, we show that optimal transport plans are supported on compact sets of the form D_ϵ. Since our result holds in a slightly more general context than the sphere, we consider that M can be any smooth complete Riemannian manifold. Let $c : M \times M \to \mathbb{R}$ be any cost bounded from below that satisfies $c(x,y) = h(\mathrm{d}_M(x,y))$ where $h : \mathbb{R}_+ \to \mathbb{R}$ is a continuous decreasing function such that $h(0) = +\infty$ and $h(t) < +\infty$ for $t > 0$.

Theorem 32. *Let $\mu, \nu \in \mathcal{P}(M)$ and $\beta > 0$ such that both $M_\mu(\beta) < 1/8$ and $M_\nu(\beta) < 1/8$, then there exists a constant $\varepsilon > 0$ such that any optimal transport plan $\gamma \in \Gamma(\mu, \nu)$ is concentrated on D_ε.*

Similar results have already been obtained in different settings [32, 16, 50], but none of them can be applied to discrete measures and therefore does not imply our result. W. Gangbo and V.Oliker [32] work with Borel measures that vanish on $(d-1)$-rectifiable sets. G. Buttazzo et al. [16] consider multimarginal optimal transport problems for constant measures. G.Loeper [50] considers two measures μ and ν such that $\mu \geqslant m dVol$ with $m > 0$ and ν satisfies for any $\varepsilon \geqslant 0$ and $x \in M$, $\nu(\mathrm{B}(x, \varepsilon)) \leqslant f(\varepsilon)\varepsilon^{n(1-1/n)}$ for some function $f : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\lim_{t\to 0} f(t) = 0$. These hypothesis imply that neither μ nor ν can be discrete.

Our proof is an adaptation of their proofs in a different context. Lemma 37 is inspired by [32] while Lemma 36 and the overall strategy of the proof come from [16]. The main difference is that here we work on any measure satisfying $M(\beta) < 1/8$, including discrete measures, which is useful for semi-discrete optimal transport.

Remark 33. *Our proof requires $M(\beta) < 1/8$ but we believe that the theoretical bound is $M(\beta) \leqslant 1/2$, which is enough to guarantee that there exists a transport plan with finite global cost, as showed in the following lemma. It is easy to show that we cannot expect a greater bound. Take for example $x \neq y$ in $\mathcal{S}^{d-1}$, $\varepsilon \in]0, 1/2[$, $\mu = 1/2(\delta_x + \delta_y)$ and $\nu = (1/2 + \varepsilon)\delta_x + (1/2 - \varepsilon)\delta_y$. Any transport plan between μ and ν will send a set of measure at least ε from x to itself for which the cost is infinite.*

The end of this section is mainly dedicated to the proof of Theorem 32, which is necessary to guarantee that the optimal transport plan is supported where the cost is regular enough and the MTW tensor is non-negative, and allow us to apply our strong c-concavity result in order to obtain the stability results of Theorem 31. We first show in the following lemma that there exists a transport plan with bounded total cost.

Lemma 34. *If $M_\mu(\beta) \leqslant 1/2$ and $M_\nu(\beta) \leqslant 1/2$ for some $\beta > 0$, then there exists $\gamma \in \Gamma(\mu, \nu)$ s.t.*

$$\int \mathrm{c}\mathrm{d}\gamma \leqslant h(\beta/2).$$

The proof of Lemma 34 relies on the following result, which can be seen as a continuous formulation of Hall's mariage Lemma. A proof is given in [68, Theorem 1.27].

Lemma 35 (Continuous Hall's marriage lemma)**.** *Let M, N be Polish spaces, and let P be a closed subset of $M \times N$. Given $\mu \in \mathcal{P}(M)$ and $\nu \in \mathcal{P}(N)$, the following propositions are equivalent:*

(i) $\exists \gamma \in \Gamma(\mu, \nu)$ *such that* $\mathrm{spt}(\gamma) \subseteq P$ *;*

(ii) for every Borel subset $B \subseteq M$,

$$\nu(\{y \in N \mid \exists x \in B \text{ s.t. } (x, y) \in P)\}) \geqslant \mu(B).$$

Proof of Lemma 34. We are going to apply the Continuous Hall's marriage lemma to the set $P = \{(x, y) \in M^2,\ \mathrm{d}_M(x, y) \geqslant \beta/2\}$. Let B be any Borel set of X. We first assume that the diameter of B is a most β so that $B \subseteq \mathrm{B}(x_0, \beta)$ for some $x_0 \in B$. Then,

$\mu(B) \leqslant \mu(\mathrm{B}(x_0, \beta)) \leqslant 1/2$ using $M_\mu(\beta) \leqslant 1/2$. having also $M_\nu(\beta) \leqslant 1/2$ we get

$$\begin{aligned}\nu(\{y \in M \mid \exists x \in B,\ \mathrm{d}_M(x,y) \geqslant \beta/2\}) &\geqslant \nu(\{y \in M \mid \mathrm{d}_M(x_0,y) \geqslant \beta/2\}) \\ &= 1 - \nu(\mathrm{B}(x_0, \beta/2) \\ &\geqslant 1/2 \\ &\geqslant \mu(B).\end{aligned}$$

Assume now that the diameter of B is greater than β. Then there exist $x, x' \in B$ such that $\mathrm{d}_M(x,x') \geqslant \beta$ and the left hand side of the previous inequation is equal to 1 and the condition is obviously satisfied. We can therefore apply Lemma 35, which implies the existence of a transport plan γ between μ and ν such that for any pair $(x,y) \in \mathrm{spt}(\gamma)$ one has $\mathrm{d}_M(x,y) \geqslant \beta/2$. Since h is decreasing, we have $c(x,y) \leqslant h(\beta/2)$ for every pair $(x,y) \in \mathrm{spt}(\gamma)$, which implies the desired result. □

Lemma 36. *Let γ be an optimal transport map between μ and ν for the cost c and let $\beta > 0$ such that $M_\mu(\beta) < 1/8$ and $M_\nu(\beta) < 1/8$. Then, for any optimal transport plan $\gamma \in \Gamma(\mu,\nu)$, there exists pairs $(x_0,y_0),(x_0',y_0') \in \mathrm{spt}(\gamma)$ such that the four points x_0, y_0, x_0', y_0' are at distance at least $\min(\varepsilon,\beta)$ with $\varepsilon := h^{-1}(4h(\beta/2))$ from each other.*

Proof. Since γ is an optimal transport plan, its cost is less than the cost of the transport plan constructed in Lemma 34. Since h is decreasing and by definition of Δ_ε, we have for any $\varepsilon > 0$,

$$h(\varepsilon)\gamma(\Delta_\varepsilon) \leqslant \int_{\Delta_\varepsilon} c\mathrm{d}\gamma \leqslant h(\beta/2)$$

Note that we can consider $h(\beta/2) > 0$, choosing a smaller β if necessary. Then if $\varepsilon = h^{-1}(4h(\beta/2))$ we get $\gamma(\Delta_\varepsilon) \leqslant \frac{1}{4}$, thus proving the existence of a pair $(x_0,y_0) \in \mathrm{spt}(\gamma) \setminus \Delta_\varepsilon$.

Since $M_\mu(\beta) < 1/8$, one has

$$\gamma((\mathrm{B}(x_0,\beta) \cup \mathrm{B}(y_0,\beta)) \times \mathcal{S}^{d-1}) \leqslant \mu(\mathrm{B}(x_0,\beta)) + \mu(\mathrm{B}(y_0,\beta)) < \frac{1}{4},$$

Similarly, $M_\nu(\beta) < 1/8$, gives

$$\gamma(\mathcal{S}^{d-1} \times (\mathrm{B}(x_0,\beta) \cup \mathrm{B}(y_0,\beta))) \leqslant \nu(\mathrm{B}(x_0,\beta)) + \nu(\mathrm{B}(y_0,\beta) < \frac{1}{4},$$

so that

$$\begin{aligned}&\gamma(\{(x,y) \in M^2 \mid \mathrm{d}_M(x,x_0) > \beta,\ \ \mathrm{d}_M(y,y_0) > \beta,\ \mathrm{d}_M(y,-x_0) > \beta, \\ &\qquad\qquad \mathrm{d}_M(x,y_0) > \beta \text{ and } \mathrm{d}_M(x,y) > \varepsilon\} \\ &= \gamma\Bigg(M^2 \setminus \Bigg[(\mathrm{B}(x_0,\beta) \cup \mathrm{B}(y_0,\beta)) \times \mathcal{S}^{d-1} \\ &\qquad\qquad \cup \mathcal{S}^{d-1} \times (\mathrm{B}(x_0,\beta) \cup \mathrm{B}(y_0,\beta)) \cup \Delta_\varepsilon\Bigg]\Bigg) \\ &\geqslant 1 - \gamma((\mathrm{B}(x_0,\beta) \cup \mathrm{B}(y_0,\beta)) \times \mathcal{S}^{d-1}) \\ &\qquad - \gamma(\mathcal{S}^{d-1} \times (\mathrm{B}(x_0,\beta) \cup \mathrm{B}(y_0,\beta))) - \gamma(\Delta_\varepsilon) \quad > 1/4.\end{aligned}$$

This proves the existence of $(x_0',y_0') \in \mathrm{spt}(\gamma)$ such that $\mathrm{d}_M(x_0,x_0') > \beta$ and $\mathrm{d}_M(y_0,y_0') > \beta$ and $\mathrm{d}_M(x_0',y_0') \geqslant \varepsilon$ and allows us to conclude. □

To state the next lemma, we first need to introduce the definition of c-cyclically monotone sets. Roughly speaking these are subsets of the product space $M \times N$ for which each pair has a lower costs than the other possible choices. It is well known that the support of an optimal transport plan is always c-cyclically monotone [68].

Definition 23 (c-cyclically monotone sets)**.** *We say that $S \subset M \times N$ is c-cyclically monotone for the cost function $c : M \times N \to \mathbb{R}$ if for any family of k couples $(x_1, y_1), \cdots, (x_k, y_k) \in S$ and any permutation σ we have*

$$\sum_i c(x_i, y_i) \leqslant \sum_i c(x_i, y_{\sigma(i)}).$$

Lemma 37. *Assume that c is bounded from below by a constant c_{min}. Let $S \subseteq M \times M$ be a c-cyclically monotone set , which contains two pairs $(x_0, y_0), (x_0', y_0')$ such that the pairwise distance between the points x_0, y_0, x_0', y_0' is at least $\varepsilon > 0$. Then,*

$$\forall (x, y) \in S, \quad c(x, y) \leqslant C_\varepsilon := h(\varepsilon) + 2h(\varepsilon/2) + 2|c_{min}|.$$

Proof. Using the c-cyclical monotonicity of S and $c \geqslant c_{min}$ one has

$$c(x, y) \leqslant c(x, y) + c(x_0, y_0) + c(x_0', y_0') - 2c_{min} \leqslant F(x, y) + 2|c_{min}|$$

where

$$\begin{cases} F(x, y) = \min(c(x, y_0) + R_1(y), c(x, y_0') + R_2(y)) \\ R_1(y) = \min(c(x_0, y) + c(x_0', y_0'), c(x_0, y_0') + c(x_0', y))) \\ R_2(y) = \min(c(x_0, y) + c(x_0', y_0), c(x_0, y_0) + c(x_0', y))). \end{cases}$$

By assumption, we have $\mathrm{d}_M(x_0, x_0') \geqslant \varepsilon$, thus $\max(\mathrm{d}_M(x_0, y), \mathrm{d}_M(x_0', y)) \geqslant \varepsilon/2$. Then, since h is decreasing, one has $\min(c(x_0, y), c(x_0', y)) \leqslant h(\varepsilon/2)$. We also have $c(x_0', y_0') \leqslant h(\varepsilon)$ and $c(x_0, y_0') \leqslant h(\varepsilon)$, which leaves us with

$$R_1(y) \leqslant h(\varepsilon) + \min(c(x_0, y), c(x_0', y)) \leqslant h(\varepsilon) + h(\varepsilon/2),$$

and the same bound holds for $R_2(y)$. Using the same argument we get $\min(c(x, y_0), c(x, y_0')) \leqslant h(\varepsilon/2)$ and thus,

$$F(x, y) \leqslant h(\varepsilon) + h(\varepsilon/2) + \min(c(x, y_0), c(x, y_0')) \leqslant h(\varepsilon) + 2h(\varepsilon/2). \qquad \square$$

Proof of Theorem 32. Let $\beta > 0$ such that $M(\beta) < 1/8$. Let γ be an optimal transport plan, and denote by S its support. By Lemma 34, the cost of this transport plan is finite. This implies that S is c-cyclically monotone. Recall that by assumption, the cost c is bounded from below. Therefore by Lemmas 36 and 37 one has

$$\forall (x, y) \in S, \quad c(x, y) \leqslant C_\varepsilon := h(\varepsilon) + 2h(\varepsilon/2) + 2|c_{min}|.$$

where $\varepsilon = \min(\beta, h^{-1}(4h(\beta/2)))$. This directly implies that $S \subseteq D_\delta$ with $\delta = h^{-1}(C_\varepsilon)$.

3.5.2 Proof of Theorem 31

Here, we come back to the sphere case, i.e. $M = \mathcal{S}^{d-1}$. We recall that the reflector cost is given on M^2 by $c(x, y) = -\ln(1 - \langle x|y\rangle)$. Note that on the unit sphere, $\mathrm{d}_M(x, y) = \arccos(\langle x|y\rangle)$, hence the reflector cost is of the form $c(x, y) = h(\mathrm{d}_M(x, y))$ with $h(t) = -\ln(1 - \cos(t))$ and satisfies the assumptions of Theorem 32.

Lemma 38. *For $\varepsilon < 2$, D_ε is symmetrically c-convex.*

Proof. A simple computation gives for $x \in M$, that $\nabla_x c(x,\cdot) : M \setminus \{x\} \to T_x M$ is one to one and given by

$$\nabla_x c(x,y) = \frac{y - \langle x|y\rangle x}{1 - \langle x|y\rangle}$$

and the inverse of $-\nabla_x c(x,\cdot)$ is

$$\text{c-exp}_x(p) = \left(1 - \frac{2}{1+\|p\|^2}\right) x - \frac{2}{1+\|p\|^2} p$$

Let (x,y_0) and (x,y_1) in D_ε, and define the c-segment $(y_t) = [y_0,y_1]_x$. For $p_0 = \nabla_x c(x,y_0)$ and $p_1 = \nabla_x c(x,y_1)$, we put $p_t = (1-t)p_0 + tp_1$, so that $y_t = \text{c-exp}_x(p_t)$. We want to show that $(x,y_t) \in D_\varepsilon$, hence we only have to show that $\mathrm{d}_M(x,y_t) \geqslant \varepsilon$. We have

$$x - y_t = \frac{2}{1+\|p_t\|^2} x + \frac{2}{1+\|p_t\|^2} p_t.$$

Since x is orthogonal to p_t and $\|x\| = 1$, we get

$$\mathrm{d}_M(x,y_t) = \arccos(\langle x|y_t\rangle) = \arccos\left(1 - \frac{2}{1+\|p_t\|^2}\right).$$

So $\mathrm{d}_M(x,y_t) \geqslant \varepsilon$ is satisfied if $1 - \frac{2}{1+\|p_t\|^2} \leqslant \cos(\varepsilon)$. Since $\cos(\varepsilon) \geqslant 1 - \varepsilon^2/2$ it is sufficient to show that

$$\frac{2}{1+\|p_t\|^2} \geqslant \varepsilon^2/2.$$

Since $\|p_t\| \leqslant \max(\|p_0\|, \|p_1\|)$, and by symmetry of p_0 and p_1 it is sufficient to show that $\|p_0\|^2 \leqslant \frac{4}{\varepsilon^2} - 1$. Again using that $\|x\| = \|y_0\| = 1$, we have

$$\|p_0\|^2 = \left\| \frac{y_0 - \langle x|y_0\rangle x}{1 - \langle x|y_0\rangle^2} \right\|^2 = \frac{1 - \langle x|y_0\rangle^2}{(1 - \langle x|y_0\rangle)^2} = \frac{1 + \langle x|y_0\rangle}{1 - \langle x|y_0\rangle}$$

Finally using the relation $\langle x|y_0\rangle = 1 - \|x - y_0\|^2/2$, we get

$$\|p_0\|^2 = \frac{4}{\|x-y_0\|^2} - 1 \leqslant \frac{4}{\varepsilon^2} - 1$$

and in conclusion, D_ε is c-convex. Note that by symmetry it is obviously symmetrically c-convex. □

End of proof of Theorem 31 Since μ and ν_0 are absolutely continuous there exists $\beta > 0$ such that $M_\mu(\beta) < 1/8$, $M_{\nu_0}(\beta) < 1/8$ and $M_{\nu_1}(\beta) < 1/8$. Therefore, by Theorem 32, there exists $\varepsilon > 0$ such that for every $x \in M$, $(x, T_i(x)) \in D_\varepsilon$. The set D_ε is a compact set and symmetrically c-convex by Lemma 38. Recall that the optimal transport map T_0 between μ and ν_0 is of the form $T_0(x) = \text{argmin}_{y\in N}\, c(x,y) - \psi_0(y)$, where $\psi_0 : N \to \mathbb{R}$ is a c-concave function. Since μ and ν_0 have $\mathcal{C}^{1,1}$ strictly positive densities, a result of Gregoire Loeper [50, Theorem 2.5] implies that ψ_0, ψ_0^c are of class $\mathcal{C}^3$ and that $T_0 : x \mapsto \text{c-exp}_x(\nabla\psi_0^c(x))$ is of class $\mathcal{C}^2$. As seen in the proof of Theorem 32, ψ_1 is c-concave for the truncated cost, which is Lipschitz, and is therefore also Lipschitz. Furthermore, it is known that the reflector cost satisfies MTW and (STwist) [50]. We can thus apply Corollary 25 which gives that ψ_0 is strongly c-concave on D_ε. We then conclude by applying Theorem 14.

3.6 Prescription of Gauss curvature measure

The problem of Gauss curvature measure prescription for a convex body has been introduced by A.D. Aleksandrov in 1950 [2] and has been shown to be equivalent to an optimal transport problem on the sphere [58, 12]. In this section we apply our stability result to this optimal transport problem.

To this purpose we define the Gauss curvature measure introduced in [2]. Let $K \subseteq \mathbb{R}^d$ be a closed bounded convex body such that $0 \in \mathrm{int}(K)$. We denote by $\rho_K : \mathcal{S}^{d-1} \to \mathbb{R}$ the radial parametrization of ∂K defined for any direction x in the sphere $\mathcal{S}^{d-1}$ by $\rho_K(x) = \sup\{r \in \mathbb{R} \mid rx \in K\}$. This induces a homeomorphism $\overrightarrow{\rho_K}$ from $\mathcal{S}^{d-1}$ to ∂K defined by

$$\begin{aligned}\overrightarrow{\rho_K} : \mathcal{S}^{d-1} &\to \partial K \\ x &\mapsto \rho_K(x)x\end{aligned}$$

We call (multivalued) Gauss map, the map $\mathcal{G}_K$ which maps a point $x \in \partial K$ to the set of unit exterior normals to K at x, namely

$$\mathcal{G}_K(x) = \{n \in \mathcal{S}^{d-1} \mid x \in \arg\max_K \langle n|\cdot\rangle\}.$$

Note that $\mathcal{G}_K(x)$ is a set when K is not smooth at x. Through this section, we denote by σ the uniform probability measure on the sphere $\mathcal{S}^{d-1}$, i.e. the normalized $(d-1)$-dimensional Hausdorff measure.

Definition 24 (Gauss curvature measure)**.** *Let K be a bounded convex body containing* 0 *in its interior. The* Gauss curvature measure *of K, denoted μ_K, is a probability measure over $\mathcal{S}^{d-1}$ defined for any Borel subset $A \subseteq \mathcal{S}^{d-1}$ by $\mu_K(A) = \sigma(\mathcal{G}_K \circ \overrightarrow{\rho_K}(A))$.*

The *Gauss curvature measure prescription problem* is the following inverse problem: given a measure $\mu \in \mathcal{P}(\mathcal{S}^{d-1})$, is it possible to find a convex body K such that $\mu = \mu_K$? It is well-known that convexity of K implies that for every non-empty spherical convex subset $\Theta \subsetneq \mathcal{S}^{d-1}$ – i.e. subsets Θ that contains any mimimizing geodesic between any pair of its points — we have

$$\mu_K(\Theta) < \sigma(\Theta_{\pi/2}) \tag{3.6.14}$$

with $\Theta_{\pi/2} = \{x \in \mathcal{S}^{d-1} \mid d_M(x, \Theta) < \pi/2\}$, and where where d_M is the geodesic distance on the sphere. Aleksandrov's theorem states that Equation (3.6.14) is in fact a sufficient condition for μ to be the Gauss curvature measure of a convex body.

Theorem 39 (Aleksandrov)**.** *Let $\mu \in \mathcal{P}(\mathcal{S}^{d-1})$ be a probability measure satisfying condition* (3.6.14)*, then there exists a unique (up to homotheties) convex body $K \subseteq \mathbb{R}^d$ with $0 \in \mathrm{int}(K)$ such that μ is the Gaussian curvature measure of K.*

3.6.1 An optimal transport problem

Following [58, 12] we briefly recall that this inverse problem can be recast as an optimal transport problem on the sphere for the cost $c(x, n) = -\ln(\max(0, \langle x|n\rangle))$, which takes value $+\infty$ when $\langle x|n\rangle \leqslant 0$. Let μ be any measure in $\mathcal{P}(\mathcal{S}^{d-1})$ satisfying condition (3.6.14). Note that the very same cost plays an important role in the theory of unbalanced optimal transport [20, 49, 30].

In the following proposition, we use the notion of *support function* of a convex set K, defined by

$$h_K(n) = \sup_{x \in \mathcal{S}^{d-1}} \rho_K(x)\langle x|n\rangle.$$

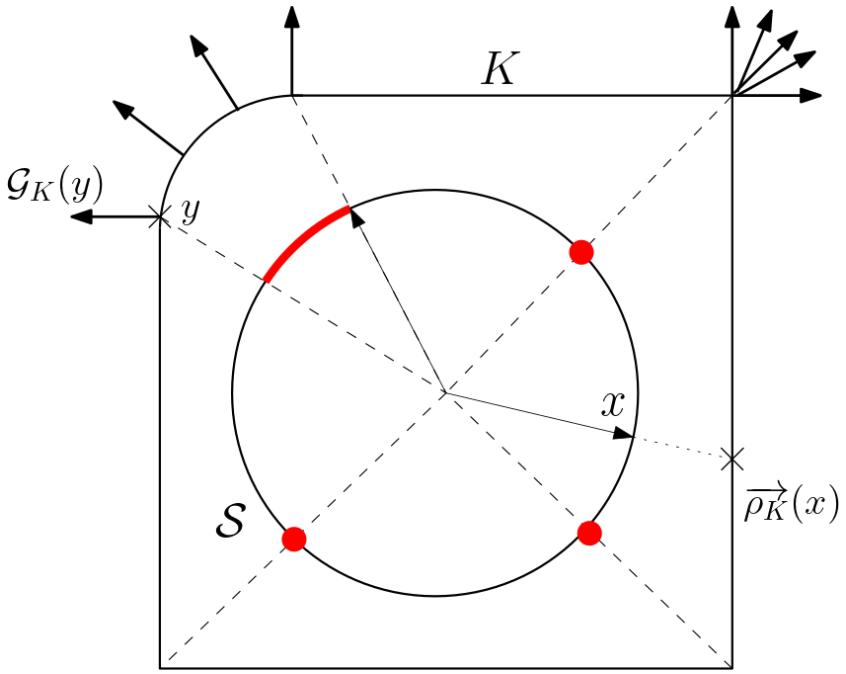

Figure 4: An example of Gauss curvature measure on a convex K. The support of the curvature measure is shown in red. It is composed of 3 diracs of mass $1/4$ for each angle, and a density on an arc of circle for the directions where the normals have smooth variations.

Proposition 40 ([58, 12])**.** *Let $\sigma \in \mathcal{P}(\mathcal{S}^{d-1})$ be the uniform measure over the sphere, let K be a compact convex body containing zero in its interior, and let $\mu = \mu_K$. Then,*

- *The map $T_K : \mathcal{S}^{d-1} \to \mathcal{S}^{d-1}$ defined σ-a.e by*

$$T_K(n) = (\mathcal{G}_K \circ \overrightarrow{\rho_K})^{-1}(n)$$

  is the optimal transport map between σ and μ for the cost c.

- *The functions $\varphi_K = -\ln(h_K)$ and $\psi_K = \ln(\rho_K)$ are maximizers of the Kantorovich dual problem. In particular we have*

$$\int_{\mathcal{S}^{d-1}} c(T_K(n), n)\mathrm{d}\sigma(n) = \int \varphi_K(n)\mathrm{d}\sigma(n) + \int \psi_K(x)\mathrm{d}\mu_K(x). \tag{3.6.15}$$

For the sake of completeness, we recall the proof of this proposition.

Proof. Let $(x, n) \in \mathcal{S}^{d-1} \times \mathcal{S}^{d-1}$ be such that $c(x, n) < +\infty$, i.e. $\langle x|n\rangle > 0$. Then,

$$h_K(n) = \max_{y\in K} \langle n|y\rangle \geqslant \langle n|\rho_K(x)x\rangle = \rho_K(x)\langle n|x\rangle, \tag{3.6.16}$$

with equality if and only if $n \in \mathcal{G}_K(x)$. Since all quantities are positive, taking the logarithm, we see that $\varphi_K(n) + \psi_K(x) \leqslant c(x, n)$, ensuring that (φ_K, ψ_K) are admissible for the dual Kantorovich problem.

Note that e.g. by [12] σ-a.e. direction $n \in \mathcal{S}^{-d-1}$ is normal to a unique point in ∂K. This implies that the map $T_K = (\mathcal{G}_K \circ \overrightarrow{\rho_K})^{-1}$ is well defined σ-a.e. The equality case of (3.6.16) gives

$$\varphi_K(n) + \psi_K(T_K(n)) \leqslant c(x, T_K(n)).$$

Integrating this equality with respect to σ directly gives (3.6.15). In turn, Kantorovich duality implies that T_K is an optimal transport between σ and μ, and that (φ_K, ψ_K) is a maximizer in the dual Kantorovich problem. □

3.6.2 Stability of transport maps

In this subsection we apply our stability result to the Gauss curvature measure prescription problem. We introduce the following notation:

$$\mathcal{K}(r, R) = \{K \subseteq \mathbb{R}^d \text{ convex, compact } \mid B(0, r) \subseteq K \subseteq B(0, R)\}.$$

Proposition 41. *Let K be a strictly convex and $\mathcal{C}^2$ compact body containing 0 in its interior. Then, for any $R > r > 0$, there exists a constant C depending on K, r and R such that*

$$\forall L \in \mathcal{K}(r, R), \quad \|\mathrm{d}_M(T_K, T_L)\|^2_{L^2(\sigma)} \leqslant C\, \mathrm{W}_1(\mu_K, \mu_L).$$

Note that in addition to the strict convexity and smoothness of K, the constant C also depends on the anisotropy of K — i.e. the radii $R_K \geqslant r_K > 0$ such that $K \in \mathcal{K}(r_K, R_K)$. The end of the section is devoted to the proof of Proposition 41. We need to check that the hypothesis of Corollary 25 are satisfied for the cost $c(x, n) = -\ln(\max(0, \langle x|n\rangle))$.

Lemma 42. *Given any $R > r > 0$, there exists $\varepsilon > 0$ such that for any set $K \in \mathcal{K}(r, R)$ and any c-optimal transport plan $\gamma \in \Gamma(\sigma, \mu_K)$, one has*

$$\mathrm{spt}(\gamma) \subseteq D_\varepsilon,$$

where $D_\varepsilon = \{(x, n) \in (\mathcal{S}^{d-1})^2 \mid d_M(x, n) \leqslant \pi/2 - \varepsilon\}$.

Proof. By hypothesis, $r \leqslant \rho_K(x) \leqslant R$ for all $x \in \mathcal{S}^{d-1}$, where ρ_K is the radial function of the convex K. Since

$$h_K(n) = \sup_{x \in \mathcal{S}^{d-1}} \rho_K(x) \langle x|n\rangle,$$

we also have $r < h_K(n) < R$. Hence the two Kantorovich potential $\varphi_K(n) = -\ln(h_K(n))$ and $\psi_K(x) = \ln(\rho_K(x))$ therefore satisfy

$$\varphi_K(n) + \psi_K(x) \leqslant -\ln(r) + \ln(R) = \ln(R/r),$$

By strong Kantorovich duality $\varphi_K(n) + \psi_K(x) = c(x, n)$ on $\mathrm{spt}(\gamma)$, which implies that c is bounded by $\ln(R/r)$ on $\mathrm{spt}(\gamma)$, i.e. for any $(x, n) \in \mathrm{spt}(\gamma)$, one has

$$c(x, n) = -\ln(\max(0, \langle x|n\rangle)) \leqslant \ln(R/r),$$

implying that $\langle x|n\rangle \geqslant r/R$ and $d_M(x, n) = \arccos(\langle x|n\rangle) \leqslant \arccos(r/R)$. Finally $(x, n) \in D_\varepsilon$ with $\varepsilon = \pi/2 - \arccos(r/R)$. □

Lemma 43. *The set $D_\varepsilon = \{(x, n) \in (\mathcal{S}^{d-1})^2 \mid d_M(x, n) \leqslant \pi/2 - \varepsilon\}$ is symmetrically c-convex for the cost $c(x, n) = -\ln(\max(0, \langle x|n\rangle))$.*

Proof. We have

$$\nabla_x c(x,n) = -\frac{n}{\langle x|n\rangle} + x$$

and by inverting $-\nabla_x c(x,\cdot)$ we get

$$\text{c-exp}_x(p) = \frac{p+x}{\sqrt{1+\|p\|^2}}$$

Let $(x,y_0) \in D_\varepsilon$ and $(x,y_1) \in \mathrm{D}_\varepsilon$ then we have $y_t = \text{c-exp}_x(p_t)$ where $p_0 = -\nabla_x c(x,y_0)$ and $p_1 = -\nabla_x c(x,y_1)$ and $p_t = (1-t)p_0 + tp_1$. By symmetry we can consider that $\|p_t\| \leqslant \|p_0\|$, which implies $\frac{1}{\sqrt{1+\|p_t\|^2}} \geqslant \frac{1}{\sqrt{1+\|p_0\|^2}}$ and thus

$$\begin{aligned} d_M(x,y_t) = \arccos(\langle x|y_t\rangle) &= \arccos\left(\frac{1}{\sqrt{1+\|p_t\|^2}}\right) \\ &\leqslant \arccos\left(\frac{1}{\sqrt{1+\|p_0\|^2}}\right) = d_M(x,y_0) \leqslant \frac{\pi}{2} - \varepsilon \end{aligned}$$

□

End of proof of Proposition 41. The map T_K (resp. T_L) is the optimal transport map between the uniform measure σ on $\mathcal{S}^{d-1}$ and μ_K (resp. μ_L) for the cost $c(x,n) = -\ln(\max(0,\langle x|n\rangle))$. From Lemma 42, for any $n \in \mathcal{S}^{d-1}$ we have $(T_K(n),n) \in D_\varepsilon$ and $(T_L(n),n) \in D_\varepsilon$. Note that for $(x,n) \in D_\varepsilon$, one has $\langle x|n\rangle > 0$ and therefore $c(x,n) = -\ln(\langle x|n\rangle) = -\ln(\cos(\mathrm{d}_M(x,n)))$. It has been shown in [30] that this cost satisfies (STwist) and (MTWw) on D_ε. By Lemma 43 the set D_ε is a symmetrically c-convex compact set.

Finally it remains to show that ψ_K is of class $\mathcal{C}^2$ and T_K is of class $\mathcal{C}^1$. Since ∂K is $\mathcal{C}^2$, its radial parametrization ρ_K is also $\mathcal{C}^2$, so $\psi_K = \ln(\rho_K)$ of class $\mathcal{C}^2$. Furthermore $\overrightarrow{\rho_K}(x) = \rho_K(x)x$ is a $\mathcal{C}^1$ diffeomorphism. Since K is stricly convex and ∂K is of class $\mathcal{C}^2$, its associated Gauss map $\mathcal{G}_K$ is a $\mathcal{C}^1$ diffeomorphism. We thus have that $T_K = (\mathcal{G}_K \circ \overrightarrow{\rho_K})^{-1}$ is of class $\mathcal{C}^1$. By Corollary 25, we know that ψ_K is strongly c-concave. We conclude by applying Theorem 14. □

Chapter 4

Generated Jacobian equations

Generated Jacobian equations have been introduced by Trudinger [66] as a generalization of Monge-Ampère equations arising in optimal transport. In this chapter, we introduce and study a damped Newton algorithm for solving these equations in the *semi-discrete* setting. We also present a numerical application of this algorithm to the near-field parallel refractor problem arising in non-imaging problems. This chapter comes from [29] written in collaboration with Quentin Mérigot and Boris Thibert.

4.1 Introduction

This chapter is concerned with the numerical resolution of *Generated Jacobian equations*, introduced by N. Trudinger [66] as a generalization of Monge-Ampère equations arising in optimal transport. Our goal is to generalize the damped Newton algorithm proposed in [46] to solve these equations in a semi-discrete setting. As mentioned earlier in Section 2.4, Generated Jacobian equations were originally motivated by inverse problems arising in non-imaging optics in the near-field case [47, 35, 37] but they also apply to problems arising in economy [56, 27]. A survey on these equations and their applications was recently written by N. Guillen [34]. The input for a generated Jacobian equations are two probability measures μ and ν over two spaces X and Y, and a *generating function* $G : X \times Y \times \mathbb{R} \to \mathbb{R}$. Loosely speaking, a scalar function ψ on Y is an Alexandrov solution to the generated Jacobian equation if the map T_ψ defined by

$$T_\psi(x) \in \operatorname*{arg\,max}_{y \in Y} G(x, y, \psi(y))$$

transports μ onto ν, i.e. ν is the image of the measure μ under T_ψ, denoted

$$T_{\psi\#}\mu = \nu.$$

Note that one needs to impose some conditions on μ and G ensuring that the map T_ψ is well-defined μ-almost everywhere. One can describe the meaning of this equation using an economic metaphor. We consider X as a set of customers, Y as a set of products and $G(x, y, \psi(y))$ corresponds to the utility of the product y for the customer x given a price $\psi(y)$. The probability measure μ and ν describe the distribution of customers and products. The map T_ψ can be described as the "best response" of customers given a price menu $\psi : Y \to \mathbb{R}$: each customer $x \in X$ tries to maximize its own utility $G(x, y, \psi(y))$ over all products $y \in Y$: the maximizer, if it exists and is unique, is denoted $T_\psi(x)$. Then, ψ is a solution to the generated jacobian equation if the best response map T_ψ pushes the distribution of customers to the distribution of available products ν.

To our knowledge, there are not many algorithms to numerically solve generated Jacobian equations. An iterative algorithm has been proposed in [1] when the source measure is absolutely continuous and the target measure is discrete, which is also our framework. More recently, a least-squared minimization algorithm has been proposed in the case of two absolutely continuous measures [62].

4.1.1 Contribution.

In this chapter, we are interested in algorithms for solving the semi-discrete case, where the source measure μ is absolutely continuous with respect to the Lebesgue measure on $X \subseteq \mathbb{R}^d$ and the target measure ν is finitely supported. Such discretization can be traced back to Minkowski, but have been used more recently to solve Monge-Ampère equations [59], problems from non-imaging optics [17], more general optimal transport problems [44], but also generated Jacobian equations [1]. In all the cited papers, the methods are coordinate-wise algorithms with minimal increment and are similar to the algorithm introduced by Oliker-Prussner [59]. The computational complexity of these algorithms scales more than cubicly [54] (N^3, where N is the size of the support of ν), making them limited to fairly small discretizations. More recently, Newton methods have been introduced to solve semi-discrete optimal transport problems [46, 53]. In this chapter, we show that newtonian techniques can also be applied to Generated Jacobian equations under mild conditions on the generating function G.

4.1.2 Semi-discrete optimal transport.

The semi-discrete optimal transport problem is presented in details in Section 2.3, we give here a quick reminder in order to see the link with the semi-discrete Generated Jacobian equations. This setting refers to the case where one is given an absolutely continuous probability measure μ (with respect to the Lebesgues measure) supported on a domain X of $\mathbb{R}^d$ and a discrete probability measure $\nu = \sum_y \nu_y \delta_y$ supported on a finite set Y. Given a cost function $c : X \times Y \to \mathbb{R}$, the optimal transport problem amounts to finding a function $T : X \to Y$ that minimizes the total cost $\int_X c(x, T(x)) d\mu(x)$ under the condition $\mu(T^{-1}(y)) = \nu_y$ for any $y \in Y$. We have seen in Chapter 2 that this problem amounts to finding a dual potential $\psi : Y \to \mathbb{R}$ that satisfies

$$\forall y \in Y \quad \mu(\mathrm{Lag}_y(\psi)) = \nu_y \tag{MA}$$

where $\mathrm{Lag}_y(\psi)$ are the Laguerre cells defined by

$$\mathrm{Lag}_y(\psi) = \{x \in X \mid \forall z \in Y, c(x, y) + \psi(y) \leqslant c(x, z) + \psi(z)\}.$$

This summarize the fact that the semi-discrete optimal transport problem can be recast as a mass prescription problem of the Laguerre cells, which is what is important here seen we are going to write the generated Jacobian equation the same way.

4.1.3 Generated Jacobian equation.

The Generated Jacobian equation in the semi-discrete setting has a very similar form. The problem also amounts to finding a function $\psi : Y \to \mathbb{R}$ that satisfies Equation (MA), but the Laguerre cells have a more general form and read

$$\mathrm{Lag}_y(\psi) = \{x \in X \mid \forall z \in Y, G(x, y, \psi(y)) \geqslant G(x, z, \psi(z))\}$$

where G is called a *generating function*. When G is linear in the last variable, i.e. when $G(x, y, v) = -c(x, y) - v$, one obviously recovers the Laguerre cells from optimal transport. If one choose $G(x, y, v) = \frac{1}{2v} - \frac{v}{2}\|x - y\|^2$, then we recover the near field parallel reflector problem (NF-par) presented in Section 2.4, which does not fall in the scope of optimal transport.

Note that the lack of linearity in the generating function G adds several theoretical and practical difficulties. To see this, consider the mass function

$$H : \mathbb{R}^Y \to \mathbb{R}^Y, \quad \psi \mapsto (\mu(\mathrm{Lag}_y(\psi)))_{y \in Y}.$$

In the optimal transport case, the function H is invariant under the addition of a constant (i.e. $H(\psi + c) = H(\psi)$ for any $c \in \mathbb{R}$), which entails under mild assumptions that the kernel of $\mathrm{D}H(\psi)$ has rank one and coincides with the vector space of constant functions on Y [46]. Furthermore, as a consequence of Kantorovitch duality, the function H is the gradient of a functional, called *Kantorovitch functional* in [46]. This implies that the differential $\mathrm{D}H(\psi)$ is symmetric. In the case of generated Jacobian equations, these two properties do not hold anymore: the differential $\mathrm{D}H(\psi)$ is not necessarily symmetric and its kernel is not reduced to the set of constant functions in general.

In this chapter, we generalize the damped Newton algorithm proposed in [46] to solve generated Jacobian equations. Note that unlike [46] we do not require any Ma-Trudinger-Wang type condition to prove the convergence of our algorithm. In Section 4.2 we recall the notion of generating function and its properties, and introduce the generated Jacobian equation in the semi-discrete setting. Section 4.3 is entirely dedicated to the numerical resolution of the generated Jacobian equation. In Section 4.4, we apply our algorithm to numerically solve the Near Field Parallel Reflector problem. Note that F. Abedin and C. Gutierrez also consider this problem [1], but their algorithm requires a strong condition, called *Visibility Condition*, that implies the *Twist condition* (defined hereafter) of the generating function G. We show that under a much weaker assumption, this twist condition holds for a subset of dual potential $\psi : Y \to \mathbb{R}$ on which we can apply our algorithm. It is very likely that our assumption could also be adapted to [1].

4.2 Semi-discrete generated Jacobian equation

In this section, we recall the notions introduced by N. Trudinger in order to define the generated Jacobian equation [66] in the semi-discrete setting. Let Ω be an open bounded domain of $\mathbb{R}^d$, let X be a compact subset of Ω and let Y be a finite set of $\mathbb{R}^d$. Let μ be a measure on Ω, which is absolutely continuous with respect to the Lebesgue measure, with non-negative density ρ supported on X (i.e. $\mathrm{spt}(\rho) \subset X$), and let $\nu = \sum_{y \in Y} \nu_y \delta_y$ be a measure on the finite set Y such that all ν_y are positive ($\nu_y > 0$). These two measures must satisfy the mass balance condition $\mu(X) = \nu(Y)$ and it is not restrictive to view them as probability measures:

$$\int_X \rho(x)dx = \sum_{y \in Y} \nu_y = 1$$

Notations. We denote by $\mathcal{H}^k$ the k-dimensional Hausdorff measure in $\mathbb{R}^d$. In particular $\mathcal{H}^d$ is the Lebesgue measure in $\mathbb{R}^d$. The set of functions from Y to $\mathbb{R}$ is denoted by $\mathbb{R}^Y$. We denote by $\langle \cdot | \cdot \rangle$ the Euclidean scalar product, by $\| \cdot \|$ the Euclidean norm, by $\mathrm{B}(x, r)$ the Euclidean ball of center x and radius r, by $\chi_A : \mathbb{R}^d \to \{0, 1\}$ the indicator function

of a set A. The image and kernel of a matrix M are respectively denoted by $\mathrm{im}(M)$ and $\ker(M)$. We denote by $\mathrm{span}(u)$ the linear space spanned by a vector u, by $\nabla_x G$ the gradient of a function G with respect to x and by $\partial_v G$ its scalar derivative with respect to v. Finally, for $N \in \mathbb{N}$, we denote $[\![1, N]\!] = \{1, \dots, N\}$.

4.2.1 Generating function

We recall below the notion of generating function and G-convexity in the semi-discrete setting [66, 1].

Definition 25 (Generating function)**.** *Let $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$ with $a < b$ and $I =]a, b[$. A function $G : \Omega \times Y \times I \to \mathbb{R}$ is called a generating function. We assume that it satisfies the following properties:*

- Regularity condition: $(x, y, v) \mapsto G(x, y, v)$ *is continuously differentiable in x and v, and*
$$\forall \alpha < \beta \in I, \sup_{(x,y,v) \in \Omega \times Y \times [\alpha, \beta]} |\nabla_x G(x, y, v)| < +\infty \tag{Reg}$$
- Monotonicity condition:
$$\forall (x, y, v) \in \Omega \times Y \times I : \partial_v G(x, y, v) < 0 \tag{Mono}$$
- Twist condition:
$$\forall x \in \Omega, (y, v) \mapsto (G(x, y, v), \nabla_x G(x, y, v)) \text{ is injective on } Y \times I \tag{Twist}$$
- Uniform Convergence condition:
$$\forall y \in Y, \lim_{v \to a} \inf_{x \in \Omega} G(x, y, v) = +\infty \tag{UC}$$

Remark 44 (Range of G)**.** *Without loss of generality we will consider that $I = \mathbb{R}$. Indeed suppose that $G : \Omega \times Y \times I \to \mathbb{R}$ satisfies the assumptions of the above definition. Considering a strictly increasing $\mathcal{C}^1$ diffeomorphism $\zeta : \mathbb{R} \to I$ and setting $\widetilde{G}(x, y, v) = G(x, y, \zeta(v))$, we get a generating function $\tilde{G} : \Omega \times Y \times \mathbb{R} \to \mathbb{R}$, which also satisfies the conditions above. Moreover, up to reparametrization, the generated Jacobian equations associated to G and $\tilde{G}$ are equivalent.*

We defined the c-convexity (or c-concavity) with respect to a cost c and the dual of an optimal transport problem in Section 2.2. One can use the same technique for a Generating function, giving G-convexity. Then the duality theory of optimal transport can be mostly imitated for generated Jacobian equation.

Definition 26 (G-convexity)**.** *Let $\varphi : \Omega \to \mathbb{R}$ be a function. If $\varphi \geqslant G(\cdot, y_0, \lambda_0)$ for all $x \in \Omega$ with equality at $x = x_0$, we say that the function $G(\cdot, y_0, \lambda_0)$* supports *φ at x_0. A function $\varphi : \Omega \to \mathbb{R}$ is said to be G-convex if it is supported at every point, i.e. for all $x_0 \in \Omega$,*
$$\exists (y_0, \lambda_0) \in Y \times \mathbb{R} \text{ s.t. } \begin{cases} \forall x \in \Omega, \varphi(x) \geqslant G(x, y_0, \lambda_0) \\ \varphi(x_0) = G(x_0, y_0, \lambda_0) \end{cases} \tag{4.2.1}$$

Remark 45 (Relation with convexity)**.** *The notion of G-convexity generalizes in a certain sense the notion of convexity. Intuitively, it amounts to replacing the supporting hyperplanes by functions of the form $G(\cdot, y, \lambda)$. If $G(\cdot, y, \lambda)$ is convex for any $y \in Y$ and any $\lambda \in \mathbb{R}$, then a G-convex function is always convex. Moreover, if the generating function G is affine (i.e $G(x, y, \lambda) = \langle x, y\rangle + \lambda$) and if $Y = \mathbb{R}^d$, then the notions of G-convexity and convexity are equivalent.*

Definition 27 (G-subdifferential)**.** *Let φ be a G-convex function and let $x_0 \in \Omega$. The G-subdifferential $\partial_G\varphi$ of φ at x_0 is defined by*

$$\partial_G\varphi(x_0) = \{y \in Y \mid \exists \lambda_0 \in \mathbb{R} s.t.\ G(\cdot, y, \lambda_0) \text{ supports } \varphi \text{ at } x_0\} \tag{4.2.2}$$

The following lemma (Lemma 2.1 in [1]) shows that the $\partial_G\varphi$ is single-valued almost everywhere, and induces a measurable map.

Lemma 46. *[1, Lemma 2.1] Let φ be G-convex with G satisfying* (Reg)*,* (Mono) *and* (Twist)*. Then, there exists a measurable map $S_\varphi : \Omega \to Y$ s.t.*

$$\text{for a.e. } x \in \Omega, \quad \partial_G\varphi(x) = \{S_\varphi(x)\}.$$

We can define the notion of generated Jacobian equation.

Definition 28 (Brenier solution to the GJE)**.** *A function $\varphi : X \to \mathbb{R}$ is a Brenier solution to the generated Jacobian equation between a probability density μ on Ω and a probability measure $\nu = \sum_{y\in Y} \nu_y\delta_y$ on Y if it satisfies*

$$\begin{cases} \varphi \text{ is G-convex} \\ \forall y \in Y, \mu(S_\varphi^{-1}(\{y\})) = \nu_y \end{cases} \tag{GenJac}$$

4.2.2 G-transform

The goal in this section is to write a dual formulation of the generated Jacobian equation, using the notion of G-transform introduced by Trudinger [66].

Definition 29. *The G-transform $\psi^G : \Omega \to \mathbb{R}$ of $\psi : Y \to \mathbb{R}$ is defined by*

$$\forall x \in \Omega, \quad \psi^G(x) = \max_{y\in Y} G(x, y, \psi(y)). \tag{4.2.3}$$

Proposition 47. *Assume G satisfies* (Reg)*,* (Mono) *and* (Twist) *and let $\varphi : \Omega \to \mathbb{R}$ be a G-convex function. Then there exists $\psi : S_\varphi(\Omega) \to \mathbb{R}$ s.t.*

$$\forall x \in \Omega, \quad \varphi(x) = \max_{y\in S_\varphi(\Omega)} G(x, y, \psi(y))$$

Proof. Let $y \in S_\varphi(\Omega)$, then for any $x_0 \in S_\varphi^{-1}(y)$ there exists $\lambda_0 \in \mathbb{R}$ such that $\varphi(x_0) = G(x_0, y, \lambda_0)$. Since φ is G-convex we also have for any $x \in \Omega$ that $\varphi(x) \geqslant G(x, y, \lambda_0)$. Specifically for $x_1 \in S_\varphi^{-1}(y)$, we get $\varphi(x_1) = G(x_1, y, \lambda_1) \geqslant G(x_1, y, \lambda_0)$ and since $\partial_v G(x, y, v) < 0$ then $\lambda_1 \leqslant \lambda_0$. By symmetry we have $\lambda_1 = \lambda_0$. We can deduce that there exists a unique $\psi(y) \in \mathbb{R}$ such that for any $x \in S_\varphi^{-1}(y), \varphi(x) = G(x, y, \psi(y))$. This defines a map $\psi : S_\varphi(\Omega) \to \mathbb{R}$ satisfying

$$\forall x \in \Omega, \ \begin{cases} \forall y \in S_\varphi(\Omega), \varphi(x) \geqslant G(x, y, \psi(y)) \\ \exists y \in S_\varphi(\Omega), \varphi(x) = G(x, y, \psi(y)) \end{cases}$$

As a conclusion we have $\varphi(x) = \max_{y\in S_\varphi(\Omega)} G(x, y, \psi(y))$. □

Corollary 48. *Let φ be a G-convex function such that $S_\varphi(\Omega) = Y$, then there exists $\psi : Y \to \mathbb{R}$ such that $\varphi = \psi^G$.*

Remark 49 (G-convex functions are not always G-transforms)**.** *Without any additional assumptions on the generating function, we cannot guarantee that any G-convex function φ on X is the G-transform of a function ψ on Y. Define for instance*

$$\Omega = (1,2), \quad Y = \{0,1\}, \quad G(x,y,v) = \begin{cases} xe^{-v} & \text{if } y = 0 \\ -xv & \text{if } y = 1 \end{cases}.$$

and consider the function φ on Ω defined by $\varphi(x) = G(x,1,1) = -x$, which is G-convex by definition. Yet for any $v \in \mathbb{R}$ and any $x \in \Omega$,

$$\max(G(x,0,v), G(x,1,1)) = \max(xe^{-v}, -x) = xe^{-v},$$

thus implying that that there does not exist any $\psi : Y \to \mathbb{R}$ such that φ is the G-transform of ψ.

Suppose that φ is a solution of (GenJac) and that for all $y \in Y$, the mass ν_y is positive. Then, for any $y \in Y$ one has $\mu(S_\varphi^{-1}(y)) = \nu_y > 0$, which guarantees that $S_\varphi(\Omega) = Y$. Therefore by Corollary 48 there exists a function ψ on Y such that $\varphi = \psi^G$. This means that we can reparametrize the problem (GenJac) by assuming that the solution φ is the G-transform of some function ψ. The sets $S_{\psi^G}^{-1}(\{y\})$, which appear in (GenJac) will be called generalized Laguerre cells.

Definition 30 (Generalized Laguerre cells)**.** *The* generalized Laguerre cells *associated to a function $\psi : Y \to \mathbb{R}$ are defined for every $y \in Y$ by*

$$\begin{aligned} \mathrm{Lag}_y(\psi) &:= S_{\psi^G}^{-1}(\{y\}) \\ &= \{x \in \Omega \mid \forall z \in Y, G(x,y,\psi(y)) \geqslant G(x,z,\psi(z))\}. \end{aligned} \tag{4.2.4}$$

Note that by Lemma 46, the intersection of two generalized Laguerre cells has zero Lebesgue measure, ensuring that the sets $\mathrm{Lag}_y(\psi)$ form a partition of Ω up to a μ-negligible set.

Definition 31 (Alexandrov solution to GJE)**.** *A function $\psi : Y \to \mathbb{R}$ is an* Alexandrov solution *to the generated Jacobian equation between generated Jacobian equation between a probability density μ on Ω and a probability measure $\nu = \sum_{y \in Y} \nu_y \delta_y$ on Y if ψ^G is a Brenier solution (Definition 28) to the same GJE, or equalently if*

$$\forall y \in Y, \quad H_y(\psi) = \nu_y, \qquad \text{where } H_y(\psi) = \mu(\mathrm{Lag}_y(\psi)).$$

Setting $H(\psi) = (H_y(\psi))_{y \in Y}$ and considering ν as a function over Y, we can even rewrite this equation as

$$H(\psi) = \nu. \tag{GJE}$$

4.3 Resolution of the generated Jacobian equation

The goal of this section is to introduce and study a Newton algorithm to solve the semi discrete generated Jacobian equation (GJE). Before doing so, we study the regularity of the mass function $H : \mathbb{R}^Y \to \mathbb{R}^Y$ in Section 4.3.1 and establish a non-degeneracy property

of its differential $\mathrm{D}H$ in Section 4.3.2, under a connectedness assumption on the support of the source measure. We present the algorithm and prove its convergence in Section 5.3.

For simplicity, we will number the points in Y, i.e. we assume that

$$Y = \{y_1, \dots, y_N\},$$

where the points y_i are distinct. This allows us to identify the set of functions $\mathbb{R}^Y$ with $\mathbb{R}^N$, by setting $\psi_i = \psi(y_i)$. We also denote $(e_i)_{1 \leqslant i \leqslant N}$ the canonical basis of $\mathbb{R}^N$. Finally, we introduce a shortened notation for Laguerre cells and intersections thereof

$$\mathrm{Lag}_i(\psi) = \mathrm{Lag}_{y_i}(\psi), \quad \mathrm{Lag}_{ij}(\psi) = \mathrm{Lag}_i(\psi) \cap \mathrm{Lag}_j(\psi).$$

Throughout this section, we assume that the generating function G satisfies all the conditions of Definition 25.

4.3.1 $\mathcal{C}^1$-regularity of H

The differentiability of H is established under a (mild) genericity hypothesis on the cost function, ensuring in particular that the intersection between three distinct Laguerre cells is negligible with respect to the $(d-1)$-dimensional Hausdorff measure, denoted $\mathcal{H}^{d-1}$. To write this hypothesis, we denote for three distinct indices i, j, k in $[\![1, N]\!]$,

$$\Gamma_{ij}(\psi) = \{x \in \Omega \mid G(x, y_i, \psi_i) = G(x, y_j, \psi_j)\}, \quad \Gamma_{ijk}(\psi) = \Gamma_{ij}(\psi) \cap \Gamma_{ik}(\psi).$$

Definition 32 (Genericity of the generating function.)**.** *The generating function G is generic with respect to Ω and Y if for any distinct indices i, j, k in $[\![1, N]\!]$ and any $\psi \in \mathbb{R}^N$ we have*

$$\mathcal{H}^{d-1}(\Gamma_{ijk}(\psi)) = 0. \tag{Gen^Y_Ω}$$

The generating function G is generic with respect to the boundary ∂X and Y if for any distinct indices i, j in $[\![1, N]\!]$ and any $\psi \in \mathbb{R}^N$ we have

$$\mathcal{H}^{d-1}(\Gamma_{ij}(\psi) \cap \partial X) = 0. \tag{$\mathrm{Gen}^Y_{\partial X}$}$$

Proposition 50. *Assume that*

- *$G \in \mathcal{C}^2(\Omega \times Y \times \mathbb{R})$ satisfies* (Reg), (Mono), (Twist), (Gen^Y_Ω), ($\mathrm{Gen}^Y_{\partial X}$),
- *$X \subseteq \Omega$ is compact and that ρ is a continuous probability density on X.*

Then the mass function $H : \mathbb{R}^N \to \mathbb{R}^N$ defined by $H(\psi) = (\mu(\mathrm{Lag}_i(\psi)))_{1 \leqslant i \leqslant N}$ has class $\mathcal{C}^1$. We have for $\psi \in \mathbb{R}^N$ and $i \in [\![1, N]\!]$

$$\begin{cases} \dfrac{\partial H_j}{\partial \psi_i}(\psi) = \displaystyle\int_{\mathrm{Lag}_{ij}(\psi)} \rho(x) \frac{|\partial_v G(x, y_i, \psi_i)|}{\|\nabla_x G(x, y_j, \psi_j) - \nabla_x G(x, y_i, \psi_i)\|} d\mathcal{H}^{d-1}(x) \geqslant 0 \text{ for } j \neq i \\ \dfrac{\partial H_i}{\partial \psi_i}(\psi) = -\displaystyle\sum_{j \neq i} \frac{\partial H_j}{\partial \psi_i}(\psi) \end{cases} \tag{4.3.5}$$

Proof. Let $\psi \in \mathbb{R}^N$ and $i, j \in [\![1, N]\!]$ be fixed indices such that $i \neq j$. We want to compute $\partial H_j / \partial \psi_i(\psi)$. For this purpose, we introduce $\psi^t = \psi + t e_i$ for $t \geqslant 0$. From (Mono), we obviously have $\mathrm{Lag}_j(\psi) \subseteq \mathrm{Lag}_j(\psi^t)$. Therefore

$$H_j(\psi^t) - H_j(\psi) = \mu(\mathrm{Lag}_j(\psi^t)) - \mu(\mathrm{Lag}_j(\psi)) = \mu(\mathrm{Lag}_j(\psi^t) \setminus \mathrm{Lag}_j(\psi))$$

We introduce the set L obtained by removing one inequality in the definition of the generalized Laguerre cell $\mathrm{Lag}_j(\psi)$:

$$L = \{x \in \Omega \mid \forall k \neq i, G(x, y_j, \psi_j) \geqslant G(x, y_k, \psi_k)\}.$$

We have in particular $\mathrm{Lag}_j(\psi) \subseteq L$ and more precisely

$$\mathrm{Lag}_j(\psi^t) \setminus \mathrm{Lag}_j(\psi) = \bigsqcup_{0<s\leqslant t} L \cap \Gamma_{ij}(\psi^t).$$

We will use this formula to get another expression of $H_j(\psi^t) - H_j(\psi)$.

Step 1. Construction of u_{ij} such that $\Gamma_{ij}(\psi^t) = u_{ij}^{-1}(\{t\})$. To construct such a function $u_{ij} : \Omega \to \mathbb{R}$, we first consider the function $f_{ij} : \Omega \times \mathbb{R} \to \mathbb{R}$ defined by

$$f_{ij}(x,t) = G(x, y_j, \psi_j) - G(x, y_i, \psi_i + t)$$

This function f_{ij} is of class $\mathcal{C}^1$ on $\Omega \times \mathbb{R}$ by hypothesis on G and we have

$$\forall (x,t) \in \Omega \times \mathbb{R}, \frac{\partial f_{ij}}{\partial t}(x,t) = -\partial_v G(x, y_i, \psi_i + t) > 0.$$

This implies that a fixed $x \in \Omega$, the function $f_{ij}(x, \cdot)$ is strictly increasing, so that equation $f_{ij}(x,t) = 0$ has at most one solution. Denoting

$$\mathcal{V}_{ij} = \{x \in \Omega \mid \exists t \in \mathbb{R}, f_{ij}(x,t) = 0\} = \bigcup_{t\in\mathbb{R}} \Gamma_{ij}(\psi^t),$$

one can therefore define a function $u_{ij} : \mathcal{V}_{ij} \to \mathbb{R}$ which satisfies

$$\forall x \in \mathcal{V}_{ij}, \quad f_{ij}(x,t) = 0 \iff u_{ij}(x) = t.$$

By the implicit function theorem, the set $\mathcal{V}_{ij}$ is open and the function u_{ij} is $\mathcal{C}^1$ on $\mathcal{V}_{ij}$. In order to apply the co-area formula, we need to compute the gradient of u_{ij}. For any point x in $\mathcal{V}_{ij}$, we have by definition

$$f_{ij}(x, u_{ij}(x)) = G(x, y_j, \psi_j) - G(x, y_i, \psi_i + u_{ij}(x)) = 0.$$

Differentiating this expression with respect to x, we obtain

$$\nabla u_{ij}(x) = \frac{\nabla_x G(x, y_j, \psi_j) - \nabla_x G(x, y_i, \psi_i + u_{ij}(x))}{\partial_v G(x, y_i, \psi_i + u_{ij}(x))}$$

which is well defined since $\partial_v G(x, y_i, v) < 0$ on $\Omega \times Y \times \mathbb{R}$ by the (Mono) hypothesis. The (Twist) condition guarantees that for all $x \in \mathcal{V}_{ij}$, the map $(y, v) \mapsto (G(x,y,v), \nabla_x G(x,y,v))$ is injective. By definition of u_{ij} we have $f_{ij}(x, u_{ij}(x))$, so that

$$G(x, y_j, \psi_j) = G(x, y_i, \psi_i + u_{ij}(x)).$$

The (Twist) condition then entails

$$\nabla_x G(x, y_j, \psi_j) \neq \nabla_x G(x, y_i, \psi_i + u_{ij}(x)),$$

implying that the gradient $\nabla u_{ij}(x)$ does not vanish.

Step 2. Computation of the partial derivatives. We can write the difference between Laguerre cells using the function u_{ij}:

$$\begin{aligned}\mathrm{Lag}_j(\psi^t) \setminus \mathrm{Lag}_j(\psi) &= \bigcup_{0<s\leqslant t} \mathrm{Lag}_{ij}(\psi^s)\\ &= \{x \in \Omega, \exists s \in]0,t], f_{ij}(x,s) = 0\} \cap L\\ &= \{x \in \Omega, \exists s \in]0,t], u_{ij}(x) = s\} \cap L\\ &= u_{ij}^{-1}(]0,t]) \cap L,\end{aligned}$$

giving directly

$$H_j(\psi^t) - H_j(\psi) = \mu(L \cap u_{ij}^{-1}(]0,t])) = \int_{L\cap u_{ij}^{-1}(]0,t])} \rho(x)dx.$$

Then the co-area formula gives us

$$\frac{H_j(\psi^t) - H_j(\psi)}{t} = \frac{1}{t}\int_{L\cap u_{ij}^{-1}(]0,t])} \rho(x)dx = \frac{1}{t}\int_0^t H_{ij}(\psi^s)ds,$$

where we introduced

$$H_{ij}(\psi) = \int_{\mathrm{Lag}_{ij}(\psi)} \frac{\rho(x)}{\|\nabla u_{ij}(x)\|} d\mathcal{H}^{d-1}(x). \tag{4.3.6}$$

Note that thanks to the computations above, we already know that the gradient $\nabla u_{ij}(x)$ does not vanish. Moreover, for any x in $\mathrm{Lag}_{ij}(\psi) \subseteq \Gamma_{ij}(\psi)$, one has $u_{ij}(x) = 0$. Thus,

$$\nabla u_{ij}(x) = (\nabla_x G(x, y_j, \psi_j) - \nabla_x G(x, y_i, \psi_i))/(\partial_v G(x, y_i, \psi_i)).$$

We can therefore rewrite

$$H_{ij}(\psi) = \int_{\mathrm{Lag}_{ij}(\psi)} \frac{\rho(x)|\partial_v G(x, y_i, \psi_i)|}{|\nabla_x G(x, y_j, \psi_j) - \nabla_x G(x, y_i, \psi_i)|} d\mathcal{H}^{d-1}(x). \tag{4.3.7}$$

As shown in Proposition 51 below, H_{ij} is continuous on $\mathbb{R}^N$. We deduce that

$$\frac{\partial H_j}{\partial \psi_i}(\psi) = \lim_{t\to 0, t>0} \frac{H_j(\psi^t) - H_j(\psi)}{t} = H_{ij}(\psi) \geqslant 0. \tag{4.3.8}$$

The case $t < 0$ can be treated similarly by replacing $\mathrm{Lag}_j(\psi) \subseteq \mathrm{Lag}_j(\psi^t)$ with $\mathrm{Lag}_j(\psi^t) \subseteq \mathrm{Lag}_j(\psi)$. We thus get the desired expression for the partial derivative $\partial H_j/\partial\psi_i$ for $i \neq j$.

To compute the partial derivative for $j = i$, we use the mass conservation property $\sum_{1\leqslant i\leqslant N} H_i(\psi) = 1$ to deduce that

$$\frac{\partial H_i}{\partial \psi_i}(\psi) = -\sum_{j\neq i} \frac{\partial H_j}{\partial \psi_i}(\psi). \qquad \square$$

It remains to show that the functions H_{ij} used in the proof of Proposition 50 are continuous.

Proposition 51. *Under the assumptions of Proposition 50, for every $i, j \in [\![1, N]\!]$, the function H_{ij} defined in* (4.3.6) *is continuous on $\mathbb{R}^N$.*

Proof. We introduce the function $g : \Omega \times \mathbb{R}^N \to \mathbb{R}$ defined by

$$g(x,\psi) = \overline{\rho}(x)\frac{|\partial_v G(x,y_i,\psi_i)|}{\|\nabla_x G(x,y_j,\psi_j) - \nabla_x G(x,y_i,\psi_i)\|}$$

where $\overline{\rho}$ is a continuous extension of the probability density $\rho_{|X}$ on Ω. For a given $\psi \in \mathbb{R}^N$, the (Twist) hypothesis guarantees that for any $x \in \Gamma_{ij}(\psi)$, $\nabla_x G(x,y_j,\psi_j) \neq \nabla_x G(x,y_i,\psi_i)$. This implies that g is continuous on a neighborhood of the set $\{(x,\psi) \in \Omega \times \mathbb{R}^N | x \in \Gamma_{ij}(\psi)\}$. We introduced in Proposition 50 the function

$$H_{ij}(\psi) = \int_{\mathrm{Lag}_{ij}(\psi)\cap X} g(x,\psi) d\mathcal{H}^{d-1}(x).$$

Let $\psi^\infty \in \mathbb{R}^N$ and ψ^n a sequence converging towards ψ^∞.The main difficulty for proving that $H_{ij}(\psi^n)$ converges to $H_{ij}(\psi^\infty)$ as $n \to +\infty$ is that the integrals in the definition of $H_{ij}(\psi^n)$ and $H_{ij}(\psi^\infty)$ are over different hypersurfaces, namely $\Gamma_{ij}(\psi^n)$ and $\Gamma_{ij}(\psi^\infty)$. Our first step will therefore be to construct a diffeomorphism between (subsets) of these hypersurfaces. We introduce $f : \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}$ the function defined by

$$f(a,b,x) = G(x,y_j,\psi_j^\infty + a) - G(x,y_i,\psi_i^\infty + b)$$

We put $a_n = \psi_j^n - \psi_j^\infty$ and $b_n = \psi_i^n - \psi_i^\infty$, so that $a_n \to 0$ and $b_n \to 0$ as n tends to $+\infty$. We also have

$$\Gamma_{ij}(\psi^\infty) = (f(0,0,\cdot))^{-1}(0), \quad \Gamma_{ij}(\psi^n) = (f(a_n,b_n,\cdot))^{-1}(0).$$

Step 1: Construction of a map F_n between $\Gamma_{ij}(\psi^\infty)$ and $\Gamma_{ij}(\psi^n)$.
This map is constructed using the composition of the flows associated to two vector fields X_a and X_b. Let $\widetilde{\Omega} \subset \Omega$ an open domain containing X. The (Twist) hypothesis guarantees that there exists a neighborhood $\widetilde{V}$ of the set $\{(a,b,x) \in \mathbb{R}^2 \times \widetilde{\Omega} |\ f(a,b,x) = 0\ \}$ such that we have for any $v \in \widetilde{V}$, $\nabla_x f(v) \neq 0$. We can then define two vector fields X_a, X_b on $\widetilde{V}$ by

$$X_a(a,b,x) = \left(1, 0, -\partial_a f(a,b,x)\frac{\nabla_x f(a,b,x)}{\|\nabla_x f(a,b,x)\|^2}\right)$$
$$X_b(a,b,x) = \left(0, 1, -\partial_b f(a,b,x)\frac{\nabla_x f(a,b,x)}{\|\nabla_x f(a,b,x)\|^2}\right)$$

Since f is of class $\mathcal{C}^2$, X_a and X_b are both of class $\mathcal{C}^1$ on $\widetilde{V}$. We then consider Φ_a and Φ_b the flows associated respectively to X_a and X_b defined for $(t,v) \in [-\varepsilon,\varepsilon]^2 \times \widetilde{V}$ by

$$\begin{cases} \Phi_a(0,v) = v \\ \partial_t \Phi_a(t,v) = X_a(\Phi(t,v)) \end{cases}$$

and

$$\begin{cases} \Phi_b(0,v) = v \\ \partial_t \Phi_b(t,v) = X_b(\Phi(t,v)) \end{cases}$$

The vector fields X_a and X_b are continuously differentiable on $\widetilde{V}$ which implies that both $\Phi_a(t,\cdot)$ and $\Phi_b(t,\cdot)$ converge pointwise in the $\mathcal{C}^1$ sense toward the identity as $t \to 0$. Let

$(t,v) \in [-\varepsilon,\varepsilon] \times \widetilde{V}$. Denoting $\nabla f(v) = (\partial_a f, \partial_b f, \nabla_x f)(v)$, we then have

$$\begin{aligned} f(\Phi_a(t,v)) &= f(\Phi_a(0,v)) + \int_0^t \frac{\partial}{\partial s}\big(s \mapsto f(\Phi_a(s,v))\big) ds \\ &= f(v) + \int_0^t \langle \nabla f(\Phi_a(s,v)) | \partial_t\ \Phi_a(s,v)\rangle ds \\ &= f(v) + \int_0^t \langle \nabla f(\Phi_a(s,v)) | X_a(\Phi_a(s,v))\rangle ds \\ &= f(v) \end{aligned}$$

Similarly one has $f(\Phi_b(t,v)) = f(v)$. Let $\Pi : \widetilde{V} \to \Omega$ the projection of $\widetilde{V} \subseteq \mathbb{R}^2 \times \Omega$ on Ω, and let $F_n : \Gamma_{ij}(\psi^\infty) \cap \widetilde{\Omega} \to \Omega$ be the function defined by

$$F_n(x) = \Pi\big(\Phi_a(a_n, \Phi_b(b_n, (0,0,x)))\big).$$

For $x \in \Gamma_{ij}(\psi^\infty)$ and $v = (0,0,x) \in \widetilde{V}$, we have

$$\Phi_a(a_n, \Phi_b(b_n, v)) = (a_n, b_n, F_n(x))$$

and from the previous equality we deduce that

$$f(\Phi_a(a_n, \Phi_b(b_n, v))) = f(v) = 0$$

This means that for $x \in \Gamma_{ij}(\psi^\infty)$, $F_n(x) \in \Gamma_{ij}(\psi^n)$. Moreover $\Phi_a(a_n, \cdot)$ and $\Phi_b(b_n, \cdot)$ are both invertible of inverse $\Phi_a(-a_n, \cdot)$ and $\Phi_b(-b_n, \cdot)$. Thus F_n is also invertible of inverse

$$F_n^{-1}(x) = \Pi\big(\Phi_b(-b_n, \Phi_a(-a_n, (a_n, b_n, x)))\big)$$

Since both $\Phi_a(a_n, \cdot)$ and $\Phi_b(b_n, \cdot)$ converge pointwise in the $\mathcal{C}^1$ toward the identity as $n \to +\infty$, we have for $x \in \Gamma_{ij}(\psi^\infty) \cap \widetilde{\Omega}$

$$\begin{cases} \lim\limits_{n \to +\infty} F_n(x) = x, \\ \lim\limits_{n \to +\infty} JF_n(x) = 1, \end{cases}$$

where JF_n is the absolute value of the determinant of the Jacobian matrix of F_n.

Step 2: Convergence of $H_{ij}(\psi^n)$ toward $H_{ij}(\psi^\infty)$.
We let $L_\infty = \mathrm{Lag}_{ij}(\psi^\infty)$ and $L_n = F_n^{-1}(\mathrm{Lag}_{ij}(\psi^n) \cap \widetilde{\Omega})$. Denoting by χ_A the indicator function of a set A, we have

$$\begin{aligned} H_{ij}(\psi^\infty) &= \int_{\Gamma_{ij}(\psi^\infty)} g(x, \psi^\infty) \chi_X(x) \chi_{L_\infty}(x) d\mathcal{H}^{d-1}(x) \\ &= \int_{\Gamma_{ij}(\psi^\infty) \cap \widetilde{\Omega}} g(x, \psi^\infty) \chi_X(x) \chi_{L_\infty}(x) d\mathcal{H}^{d-1}(x) \end{aligned}$$

because $\Gamma_{ij}(\psi^\infty) \cap X \subset \widetilde{\Omega}$. We also have

$$H_{ij}(\psi^n) = \int_{\Gamma_{ij}(\psi^n)} g(x, \psi^n) \chi_X(x) \chi_{\mathrm{Lag}_{ij}(\psi^n)}(x) d\mathcal{H}^{d-1}(x)$$

By a change of variable from x to $F_n(x)$, the latter equality becomes

$$H_{ij}(\psi^n) = \int_{\Gamma_{ij}(\psi^\infty)\cap\widetilde{\Omega}} g(F_n(x),\psi^n) JF_n(x)\chi_X(F_n(x))\chi_{L_n}(x) d\mathcal{H}^{d-1}(x)$$

where $JF_n(x)$ denotes the determinant of the Jacobian matrix of F_n. We already have the pointwise convergences $F_n(x) \to x$ and $JF_n(x) \to 1$ as $n \to \infty$. If we can show that

$$\lim_{n\to+\infty} \chi_X(F_n(x))\chi_{L_n}(x) = \chi_X(x)\chi_{L_\infty}(x)$$

for vol^{d-1} almost every point x, then using Lebesgue's dominated convergence theorem, we will obtain that $H_{ij}(\psi^n) \to H_{ij}(\psi^\infty)$.

We first show that $\lim_{n\to+\infty} \chi_{L_n}(x) \to \chi_{L_\infty}(x)$ vol^{d-1}-almost everywhere on $\Gamma_{ij}(\psi^\infty)\cap\widetilde{\Omega}$. We first consider the superior limit: given $x \in \Gamma_{ij}(\psi^\infty)\cap\widetilde{\Omega}$, we prove that $\limsup_{n\to\infty} \chi_{L_n}(x) \leqslant \chi_{L_\infty}(x)$. The limsup is non-zero if and only if there exists a subsequence $(\sigma(n))_{n\in\mathbb{N}}$ such that $\forall n \in \mathbb{N}, x \in L_{\sigma(n)}$. In this case we have $F_{\sigma(n)}(x) \in F_{\sigma(n)}(L_{\sigma(n)}) = \mathrm{Lag}_{ij}(\psi^{\sigma(n)}) \cap \widetilde{\Omega}$. This means that for any $k \neq i, j$

$$G(F_{\sigma(n)}(x), y_i, \psi_i^{\sigma(n)}) = G(F_{\sigma(n)}(x), y_j, \psi_j^{\sigma(n)}) \leqslant G(x, y_k, \psi_k^{\sigma(n)})$$

Since G is continuous the previous inequality passes to the limit $n \to \infty$, showing that $x \in L_\infty$, and that

$$\limsup_{n\to\infty} \chi_{L_n}(x) \leqslant \chi_{L_\infty}(x)$$

We now want to show $\liminf_{n\to\infty} \chi_{L_n}(x) \geqslant \chi_{L_\infty}(x)$. If $x \notin L_\infty$ the result is straightforward. Let us consider the set

$$S_{ij} = \left(\bigcup_{k\neq i,j} \Gamma_{ijk}(\psi^\infty) \right) \cup (\Gamma_{ij}(\psi^\infty) \cap \partial X) \tag{4.3.9}$$

By the genericity hypothesis (Definition 32) we have $\mathcal{H}^{d-1}(S_{ij}) = 0$. If $x \in L_\infty \setminus S_{ij}$, by definition we get for every $k \notin \{i, j\}$ that x does not belong to $\Gamma_{jk}(\psi^\infty)$. This implies a strict inequality

$$G(x, y_i, \psi_i^\infty) = G(x, y_j, \psi_j^\infty) < G(x, y_k, \psi_k^\infty).$$

Since $F_n(x)$ converges to x and since ψ^n converges to ψ^∞, we get for n large enough

$$\begin{cases} G(F_n(x), y_i, \psi_i^n) < G(F_n(x), y_k, \psi_k^n) \\ G(F_n(x), y_j, \psi_j^n) < G(F_n(x), y_k, \psi_k^n). \end{cases}$$

Moreover since $x \in \Gamma_{ij}(\psi^\infty)$, $F_n(x) \in \Gamma_{ij}(\psi^n)$. Combining the inequalities above, this shows that $F_n(x)$ belongs to $\mathrm{Lag}_{ij}(\psi^n) \cap \widetilde{\Omega} = F_n(L_n)$, i.e. $x \in L_n$. This gives us

$$\forall x \notin S_{ij}, \quad \liminf_{n\to\infty} \chi_{L_n}(x) \geqslant \chi_{L_\infty}(x).$$

Consider $x \notin S_{ij}$. For such x, we already know that $\chi_{L_n}(x) \to \chi_{L_\infty}(x)$ as $n \to +\infty$. Thus, if x does not belong to L_∞, we directly have

$$\lim_{n\to+\infty} \chi_{L_n}(x)\chi_X(F_n(x)) = \chi_{L_\infty}\chi_X(x) = 0.$$

We may now assume that x belongs to $L_\infty \setminus S_{ij}$. By definition of S_{ij}, this implies that $x \notin \partial X$. We can directly deduce that χ_X is continuous at x and that $\chi_X(F_n(x)) \to \chi_X(x)$ when $n \to +\infty$.

In conclusion we have that $H_{ij}(\psi^n) \to H_{ij}(\psi^\infty)$, so that H_{ij} is continuous. □

4.3.2 Kernel and image of $\mathrm{D}H$

The goal of this section is to prove Proposition 52 that gives properties on the differential of the mass function H. We consider the admissible set

$$\mathcal{S}^+ = \left\{\psi \in \mathbb{R}^N \mid \forall i \in [\![1, N]\!], H_i(\psi) > 0\right\}. \tag{4.3.10}$$

Proposition 52. *In addition to the assumptions of Proposition 50, we assume that*

$$\mathrm{int}(X) \cap \{\rho > 0\} \text{ is path-connected},$$

where $\mathrm{int}(X)$ *is the interior of* X*. Then we have for any* $\psi \in \mathcal{S}^+$

- *The differential* $DH(\psi)$ *has rank* $N-1$*;*
- *The image of* DH *is* $\mathrm{im}(DH(\psi)) = \mathbf{1}^\perp$ *where* $\mathbf{1} = (1, \cdots, 1) \in \mathbb{R}^N$*;*
- *For any* $w \in \ker(DH(\psi)) \setminus \{0\}$*, we have for all* $i \in [\![1, N]\!], w_i \neq 0$ *and all* w_i *have the same sign.*

The next two lemmas have already been included in the recent survey on optimal transport involving the second and third authors [54], but we include them here for completeness. The proof of Proposition 52 is different from the previous work in optimal transport because H is not symmetric.

Lemma 53. *Let* $U \subset \mathbb{R}^d$ *be a path-connected open set, and* $S \subset \mathbb{R}^d$ *be a closed set such that* $\mathcal{H}^{d-1}(S) = 0$*. Then,* $U \setminus S$ *is path-connected.*

Proof. It suffices to treat the case where U is an open ball, the general case will follow by standard connectedness arguments. Let $x, y \in U \setminus S$ be distinct points. Since $U \setminus S$ is open, there exists $r > 0$ such that $\mathrm{B}(x, r)$ and $\mathrm{B}(y, r)$ are included in $U \setminus S$. Consider the hyperplane H orthogonal to the segment $[x, y]$, and Π_H the projection on H. Then, since Π_H is 1-Lipschitz, $\mathcal{H}^{d-1}(\Pi_H S) \leqslant \mathcal{H}^{d-1}(S) = 0$, so that $H \setminus \Pi_H S$ is dense in the hyperplane H. In particular, there exists a point $z \in \Pi_H(\mathrm{B}(x, r)) \setminus S = \Pi_H(\mathrm{B}(y, r)) \setminus S$. By construction the line $z + \mathbb{R}(y - x)$ avoids S and passes through the balls $\mathrm{B}(x, r) \subset U \setminus S$ and $\mathrm{B}(y, r) \subset U \setminus S$. This shows that the points x, y can be connected in $U \setminus S$. □

We define for $\psi \in \mathbb{R}^N$ the graph $\mathcal{G}_\psi = (V, E)$ with vertex set $V = \{1, \ldots, N\}$ with edges

$$E = \left\{(i, j) \in V^2 \mid \frac{\partial H_i}{\partial \psi_j}(\psi) > 0\right\}$$

We have the following result.

Lemma 54. *Under the assumptions of Proposition 52 and for* $\psi \in \mathcal{S}^+$*, the graph* $\mathcal{G}_\psi$ *is connected.*

Proof. Let $Z = \mathrm{int}(X) \cap \{\rho > 0\}$, and $S = \bigcup_{ij} S_{ij}$ where S_{ij} is defined in (4.3.9). From Lemma 53 the set $Z \setminus S$ is path connected, we also have $\mu(Z \setminus S) = 1$ since $\mu(\partial X) = \mu(S) = 0$. Suppose that $\mathcal{G}_\psi$ is not connected. Let $i_0 \in [\![1, N]\!]$, and let I_0 be the connected component of i_0 in the graph $\mathcal{G}_\psi$. We thus have $i_0 \in I_0 \neq [\![1, N]\!]$. We consider the two non-empty sets

$$U_1 = \bigcup_{i \in I_0} \mathrm{Lag}_i(\psi) \cap (Z \setminus S) \text{ and } U_2 = \bigcup_{i \notin I_0} \mathrm{Lag}_i(\psi) \cap (Z \setminus S),$$

which partition $Z \setminus S$ up to a Lebesgue-negligible set. Moreover, since $\psi \in \mathcal{S}^+$,

$$\begin{cases} U_1 \cup U_2 = Z \setminus S, \\ 0 < \mu(U_1) < 1, \\ 0 < \mu(U_2) < 1. \end{cases}$$

By construction U_1 and U_2 are closed sets in $Z \setminus S$. Since $\mu(U_i) > 0$ we can pick x and y in $Z \setminus S$ such that $x \in U_1$ and $y \in U_2$. The $Z \setminus S$ being path-connected, we know that there exists a path $\gamma \in \mathcal{C}^0([0,1], Z \setminus S)$ satisfying $\gamma(0) = x$ and $\gamma(1) = y$. We let $t = \max\{s \in [0,1] | \gamma(s) \in U_1\}$ and we are going to show that $\gamma(t) \in U_1 \cap U_2$. By construction, $\gamma(t)$ obviously belongs to U_1. Now if $t = 1$ we have $\gamma(t) = y \in U_2$. If not, we have for all $\epsilon > 0$ that $\gamma(t+\epsilon) \in U_2$. Since U_2 is relatively closed in $Z\setminus$ and since γ is continuous, we have $\gamma(t) \in U_2$. Naming $z = \gamma(t)$, there exists $i \in I_0$, $j \notin I_0$ such that $z \in \mathrm{Lag}_i(\psi) \cap \mathrm{Lag}_j(\psi)$. Moreover, since $z \notin S$ we get that for any $k \notin \{i,j\}$,

$$G(z, y_i, \psi_i) = G(z, y_j, \psi_j) > G(z, y_k, \psi_k).$$

By continuity of G we can deduce that there exists an open ball of radius $r > 0$ such that

$$\forall x \in \mathrm{B}(z,r), \forall k \notin \{i,j\}, G(x, y_i, \psi_i) > G(x, y_k, \psi_k)$$

This implies that

$$\mathrm{B}(z,r) \cap \Gamma_{ij}(\psi) \subset \mathrm{Lag}_{ij}(\psi)$$

where $\Gamma_{ij}(\psi)$ is defined in Definition 32. By (Twist) condition and the inversion function theorem, we know that $\Gamma_{ij}(\psi)$ is a $d-1$ dimensional manifold and $z \in \Gamma_{ij}(\psi)$. Moreover we have $\rho(z) > 0$ because $z \in Z$ and ρ is continuous on $Z \subset X$ by hypothesis. We now have

$$\begin{aligned} \frac{\partial H_i}{\partial \psi_j}(\psi) &= \int_{\mathrm{Lag}_{ij}(\psi)} \rho(x) \frac{|\partial_v G(x, y_i, \psi_i)|}{\|\nabla_x G(x, y_j, \psi_j) - \nabla_x G(x, y_i, \psi_i)\|} d\mathcal{H}^{d-1}(x) \\ &\geqslant \int_{\mathrm{B}(z,r) \cap \Gamma_{ij}(\psi)} \rho(x) \frac{|\partial_v G(x, y_i, \psi_i)|}{\|\nabla_x G(x, y_j, \psi_j) - \nabla_x G(x, y_i, \psi_i)\|} d\mathcal{H}^{d-1}(x) > 0 \end{aligned}$$

which is a contradiction with the hypothesis that i and j are not connected in the graph $\mathcal{G}_\psi$. □

Proof of Proposition 52. We note the matrix $M = DH(\psi)$, with coefficients $m_{i,j} = \partial H_i / \partial \psi_j(\psi)$. We first show that $\ker(M^T) = \mathrm{span}(\mathbf{1})$. The inclusion $\mathbf{1} \in \ker(M^T)$ follows from

$$\sum_{i=1}^N m_{i,j} = \frac{\partial}{\partial \psi_j} \left(\sum_{i=1}^N H_i(\psi) \right) = 0.$$

Consider now $v \in \ker(M^T)$, and pick an index i_0 where v is maximum, i.e. $i_0 \in \arg\max_{1 \leqslant i \leqslant N} v_i$. We have

$$0 = (M^T v)_{i_0} = \sum_{i=1}^N m_{i,i_0} v_i = \sum_{i \neq i_0} m_{i,i_0} v_i + m_{i_0,i_0} v_{i_0} = \sum_{i \neq i_0} m_{i,i_0} (v_i - v_{i_0}).$$

Since $\psi \in \mathcal{S}^+$, we have by Proposition 50 that for $i \neq i_0$, $m_{i,i_0} \geqslant 0$. By definition of i_0 we also have $v_i - v_{i_0} \leqslant 0$. From all this we deduce that $v_i = v_{i_0}$ for any $i \neq i_0$ satisfying

$m_{i,i_0} > 0$, i.e. any vertex i adjacent to i_0 in the graph $\mathcal{G}_\psi$. By connectedness of $\mathcal{G}_\psi$, we conclude that $v = v_{i_0}\mathbf{1}$, thus showing $\ker(M^T) = \mathrm{span}(\mathbf{1})$.

We can deduce from this result that M is of rank $N-1$ because $\mathrm{rk}(M) = \mathrm{rk}(M^T) = N-1$. Moreover for any $u \in \mathbb{R}^N$,

$$\langle \mathbf{1}, Mu \rangle = (Mu)^T \mathbf{1} = u^T M^T \mathbf{1} = 0.$$

Since the spaces $\mathrm{im}(M)$ and $\mathbf{1}^\perp$ have the same dimension, we immediately get $\mathrm{im}(M) = \mathbf{1}^\perp$.

Let $w \in \ker(M) \setminus \{0\}$, we now want to show that for all $i \in [\![1, N]\!], w_i \neq 0$ and that all of the w_i have the same sign. The proof consists in two steps:

- Step 1: we show that $w \geqslant 0$ (or $-w \geqslant 0$).
- Step 2: we show that for $i \in [\![1, N]\!]$, $w_i > 0$.

We define $\lambda = \max_i |m_{i,i}|$ and $A = \lambda I + M$. With these definitions, v belongs to $\ker(M)$ if and only if $Av = \lambda v$. Moreover, for any $i, j \in [\![1, N]\!]$, one has $a_{i,j} \geqslant 0$ and

$$\sum_{k=1}^{N} a_{k,j} = \lambda.$$

Step 1: Assume that there exists $i_0 \in [\![1, N]\!]$ such that $w_{i_0} \geqslant 0$ (we can do this without loss on generality, by working on $-w$ otherwise). Suppose that there exists $j \neq i_0$ such that $a_{i_0,j} > 0$ and $w_j < 0$, then since $Aw = \lambda w$, we have $\lambda w_{i_0} = \sum_{j=1}^N a_{i_0,j} w_j$ and thus $\lambda |w_{i_0}| < \sum_{j=1}^N a_{i,j} |w_j|$. We also have for any $i \in [\![1, N]\!], \lambda |w_i| \leqslant \sum_{j=1}^N a_{i,j} |w_j|$. By summing this inequality on i and since the inequality is strict when $i = i_0$, we obtain

$$\sum_{i=1}^{N} \lambda |w_i| < \sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j} |w_j| = \sum_{j=1}^{N} |w_j| \sum_{i=1}^{N} a_{i,j} = \sum_{j=1}^{N} \lambda |w_j|,$$

which is a contradiction, so we can affirm that there exists no index $j \neq i_0$ such that $w_j < 0$ and $a_{i_0,j} > 0$. Since $A = M + \lambda I$, for $j \neq i_0, a_{i_0,j} = m_{i_0,j}$. We thus have $\forall j \in [\![1, N]\!], m_{i_0,j} > 0 \implies w_j \geqslant 0$. By connectedness of $\mathcal{G}$ we deduce $w \geqslant 0$.

Step 2: If there exists $i \in [\![1, N]\!]$ such that $w_i = 0$, then $\sum_j a_{i,j} w_j = 0$. Recall that by construction $a_{i,j} \geqslant 0$ and with step 1 $w_j \geqslant 0$, so we have $\forall j, a_{i,j} > 0 \implies w_j = 0$. Again by connectedness of $\mathcal{G}$ we have $w = 0$.

□

Remark 55. *Remark that a part of the proof of Proposition 52 could also be seen as a consequence of the Perron Frobenius theorem, using the notions of irreducible and stochastic matrices. The matrix $A = M + \lambda I$ can be written $A = \lambda S$ where S^T is a stochastic matrix. The matrix S is thus of spectral radius 1 and A is of spectral radius λ. Since M is irreducible, A is also irreducible. Perron Frobenius Theorem then implies that λ is a simple eigenvalue with an associated eigenvector w satisfying $w_i > 0$ for any $i \in [\![1, N]\!]$. Since $Av = \lambda v \iff Mv = 0$, we can deduce that $\mathrm{rk}(M) = N-1$ and $\ker(M) = \mathrm{span}(w)$. Moreover since $\mathbf{1} \in \ker(M^T)$, we have for any $u \in \mathbb{R}^N$, $\langle \mathbf{1}, Mu \rangle = (Mu)^T \mathbf{1} = u^T M^T \mathbf{1} = 0$ and $\mathrm{im}(M) = \mathbf{1}^\perp$.*

4.3.3 Damped Newton algorithm

In this section, we present a damped Newton algorithm to solve the generated Jacobian equation (GJE), namely $H(\psi) = \nu$. For this purpose we define in the following lemma an admissible set of variable that can be used in our algorithm.

Lemma 56 (Admissible set)**.** *Suppose that the hypothesis of Proposition 50 are satisfied. For any $\delta > 0$, there exists $\alpha \in \mathbb{R}$ such that the set*

$$\mathcal{S}^{\alpha,\delta} := \left\{\psi \in \mathbb{R}^N \mid \psi_1 = \alpha \text{ and } \forall i \in [\![1, N]\!], H_i(\psi) \geqslant \delta\right\} \subset \mathcal{S}^+ \tag{4.3.11}$$

is a compact subset of $\mathbb{R}^N$. Furthermore for δ small enough, the set (4.3.11) *is non-empty.*

Proof. Let $\gamma \in \mathbb{R}$ and $M = \max_{(x,y)\in X\times Y} G(x, y, \gamma)$, where M is finite thanks to the continuity of G and compactness of $X \times Y$. From the condition (UC), there exists $\alpha \in \mathbb{R}$ such that $\min_{x\in X} G(x, y_1, \alpha) > M$. If $\psi \in \mathbb{R}^N$ is such that $\psi_1 = \alpha$ and $\psi_i > \gamma$ for some $i \geqslant 2$, then using (Mono),

$$\forall x \in X, G(x, y_1, \alpha) > M \geqslant G(x, y_i, \gamma) \geqslant G(x, y_i, \psi_i),$$

thus implying that $\mathrm{Lag}_i(\psi) = \emptyset$, and in particular $\psi \notin \mathcal{S}^{\alpha,\delta}$. We argue similarly to show an upper bound on the elements of $\mathcal{S}^{\alpha,\delta}$: by (UC), there exists $\beta \in \mathbb{R}$ such that $\min_{(x,y)\in X\times Y} G(x, y, \beta) > \max_{x\in X} G(x, y_1, \alpha)$. If $\psi \in \mathbb{R}^N$ is such that $\psi_1 = \alpha$ and $\psi_i < \beta$ for some $i \geqslant 2$, then using (Mono), we get

$$\forall x \in X, G(x, y_i, \psi_i) \geqslant G(x, y_i, \beta) > G(x, y_1, \alpha),$$

thus showing that $\mathrm{Lag}_1(\psi) = \emptyset$, so that $\psi \notin \mathcal{S}^{\alpha,\delta}$. The set $\mathcal{S}^{\alpha,\delta}$ can be written as $\mathcal{S}^{\alpha,\delta} = \{\alpha\} \times \cap H^{-1}([\delta, 1]^N)$, and is therefore closed by continuity of H. The previous computations show that $\mathcal{S}^{\alpha,\delta} \subseteq \{\alpha\} \times [\beta, \gamma]^{N-1}$, proving that $\mathcal{S}^{\alpha,\delta}$ is compact.

Now suppose that $\delta \leqslant 1/2^{N-1}$, then we can iteratively construct a vector $\psi \in \mathcal{S}^{\alpha,\delta}$ in the following way. We start from $\psi = (\alpha, \gamma, \cdots, \gamma) \in \mathbb{R}^N$. We then have $H_1(\psi) = 1$ and for any $i \geqslant 2, H_i(\psi) = 0$. Then for all i from 2 to N can decrease ψ_i such that $H_i(\psi) = 1/2^{i-1}$. Then after iteration i we have

$$\forall k < i, H_k(\psi) \geqslant \frac{1}{2^{k-1}} - \sum_{k+1\leqslant j\leqslant i} \frac{1}{2^{j-1}} = \frac{1}{2^{i-1}}$$

After iteration N we thus have that for all $i \in [\![1, N]\!], H_i(\psi) \geqslant 1/2^{N-1} \geqslant \delta$, and since $\psi_1 = \alpha$ has not been changed during the process we have $\psi \in \mathcal{S}^{\alpha,\delta}$ and $\mathcal{S}^{\alpha,\delta} \neq \emptyset$. □

The differential of H is not invertible, but we can still define a Newton's direction by fixing one coordinate:

Proposition 57 (Newton's direction)**.** *Under the assumptions of Proposition 52, the system*

$$\begin{cases} DH(\psi)u = H(\psi) - \nu \\ u_1 = 0 \end{cases} \tag{4.3.12}$$

has a unique solution in $\mathbb{R}^N$.

Proof. Notice that from Proposition 52, $DH(\psi)$ is of rank $N-1$ and since $H(\psi)-\nu \in \mathbf{1}^\perp = \mathrm{im}(DH(\psi))$, the set $S = \{u \in \mathbb{R}^N | DH(\psi)u = H(\psi) - \nu\}$ is of dimension 1. For $u \in S$ and $w \in \ker(DH(\psi)) \setminus \{0\}$, $S = \{u + tw, t \in \mathbb{R}\}$. Since $w_1 \neq 0$ for $w \in \ker(DH(\psi)) \setminus \{0\}$, system (4.3.12) has a unique solution. □

Algorithm 1 Damped Newton algorithm to solve (GJE)

Require: $\epsilon > 0$; initialization $\psi^0 \in \mathcal{S}^{\alpha,\delta}$ where $\delta \leqslant \min_i \nu_i/2$
Ensure: ψ such that $\|H(\psi) - \nu\| \leqslant \epsilon$
1: $k \leftarrow 0$
2: **while** $\|H(\psi^k) - \nu\| > \epsilon$ **do**
3: Define u^k as the solution of the linear system

$$\begin{cases} DH(\psi^k)u = H(\psi^k) - \nu \\ u_1 = 0 \end{cases}$$

4: Compute τ^k by backtracking, i.e.

$$\tau^k = \max\{\tau \in 2^{-\mathbb{N}} | \psi^{k,\tau} = \psi^k - \tau u^k \in \mathcal{S}^{\alpha,\delta} \text{ and } \|H(\psi^{k,\tau}) - \nu\| \leqslant (1 - \frac{\tau}{2})\|H(\psi^k) - \nu\|\}$$

5: $\psi^{k+1} \leftarrow \psi^k - \tau^k u^k$ and $k \leftarrow k + 1$
6: **return** ψ^k

Theorem 58 (Linear convergence)**.** *Assume the following assumptions:*

- *the generating function $G \in \mathcal{C}^2(\Omega \times Y \times \mathbb{R})$ satisfies the assumptions* (Reg)*,* (Mono)*,* (Twist)*,* (UC)*,* (Gen^Y_Ω)*,* $(\text{Gen}^Y_{\partial X})$*,*
- *$X \subseteq \Omega$ is compact and ρ is a continuous probability density on X.*
- *$\text{int}(X) \cap \{\rho > 0\}$ is is path-connected.*

Then, there exists $\tau^* \in]0, 1]$ such that the iterates of Algorithm 1 satisfy

$$\|H(\psi^k) - \nu\| \leq \left(1 - \frac{\tau^*}{2}\right)^k \|H(\psi^0) - \nu\|. \tag{4.3.13}$$

In particular, Algorithm 1 terminates.

Proof. Let $\psi^0 \in \mathcal{S}^{\alpha,\delta}$, we define the set

$$K^\delta = \{\psi \in \mathcal{S}^{\alpha,\delta}, \|H(\psi) - \nu\| \leqslant \|H(\psi_0) - \nu\|\}$$

Since the function H is continuous, the set K^δ is non-empty and compact. Note that system (4.3.12) has $N + 1$ lines for N variables, and we know that the last line $u_1 = 0$, which can be written $e_1^T u = 0$, is linearly independent from the others. We can thus rewrite the system in the following form

$$M(\psi)u = H(\psi) - \nu \tag{4.3.14}$$

where $M(\psi) = DH(\psi) + e_1 e_1^T$. Obviously if u is a solution of (4.3.12) then it is also a solution of (4.3.14). Now if u is a solution of (4.3.14), since $e_1 \notin \text{im}(DH(\psi))$ and $H(\psi) - \nu \in \text{im}(DH(\psi))$, we have $e_1 e_1^T u = e_1^T u e_1 = 0$ which means that the scalar $e_1^T u = 0$ and thus, u satisfies (4.3.12). Since (4.3.14) has a unique solution, $M(\psi)$ is thus invertible. Let u_ψ solution of (4.3.14) for a given ψ. We have $u_\psi = M^{-1}(\psi)(H(\psi) - \nu)$. We thus have for any $\psi \in K^\delta$ that $\|u_\psi\| \leqslant \|M^{-1}(\psi)\|_{op}\|(H(\psi) - \nu)\|$ where $\|\cdot\|_{op}$ denotes the operator norm in $\mathcal{M}_N(\mathbb{R})$. The function $\psi \mapsto M(\psi)$ is continuous and M is invertible

so $\psi \mapsto \|M^{-1}(\psi)\|_{op}$ is also continuous and admits a maximum on the compact set K^δ. We note $C = \max_{\psi \in K^\delta} \|M^{-1}(\psi)\|_{op}$ so we have for any $\psi \in K^\delta$, $\|u_\psi\| \leqslant C\|H(\psi) - \nu\|$. Let $\psi \in K^\delta$ and $\psi^\tau = \psi - \tau u_\psi$ for $\tau \in [0,1]$. The first coordinate of ψ^τ satisfies $\psi_1^\tau = \alpha$. For a small τ we can write the Taylor expansion

$$\begin{aligned} H(\psi^\tau) &= H(\psi) - \tau DH(\psi)u_\psi + o(\tau\|u_\psi\|) \\ &= H(\psi) - \tau(H(\psi) - \nu) + o(\tau\|H(\psi) - \nu\|) \end{aligned}$$

it follows that

$$\|H(\psi^\tau) - \nu\| = (1-\tau)\|H(\psi) - \nu\| + o(\tau\|H(\psi) - \nu\|)$$

and thus there exists $\tau_\psi^1 > 0$ such that for all $\tau \in]0, \tau_\psi^1[$

$$\|H(\psi^\tau) - \nu\| \leqslant (1 - \frac{\tau}{2})\|H(\psi) - \nu\|$$

By compactness of K^δ, this property holds on an uniform open range $]0, \tau^1[$.
Moreover, coordinatewise we have for $i \in [\![1, N]\!]$,

$$H_i(\psi^\tau) = (1-\tau)H_i(\psi) + \tau\nu_i + o(\tau\|H(\psi) - \nu\|)$$

and since $\nu_i \geqslant 2\delta$ there exists $\tau_\psi^2 > 0$ such that

$$\forall \tau \in]0, \tau_\psi^2[,\ \forall i \in [\![1, N]\!],\ H_i(\psi^\tau) \geqslant (1 + \frac{\tau}{2})\delta.$$

Then again by compactness of K^δ, there exists $\tau^2 > 0$ such that for all $\psi \in K^\delta$ and $\tau \in]0, \tau^2[$, $\psi^\tau \in \mathcal{S}^{\alpha,\delta}$. This implies that the chosen τ^k in the algorithm will always be larger than

$$\tau^* = \frac{1}{2}\min(\tau^1, \tau^2).$$

By definition of the iterates, we deduce at one that

$$\|H(\psi^{k+1}) - \nu\| \leqslant (1 - \frac{\tau^*}{2})\|H(\psi^k) - \nu\|,$$

thus proving the desired convergence result. □

Remark 59 (Existence)**.** *Note that the convergence of the algorithm allows to recover the existence of a solution to the semi-discrete generated Jacobian equation. To obtain this result we need the set $\mathcal{S}^{\alpha,\delta}$ to be non-empty, which is the case by Lemma 56 if δ is small enough. Moreover, it can be shown that if $\mathcal{S}^{\alpha,\delta}$ is connected, then there is uniqueness of the solution when ψ_1 is fixed. Indeed, the set $\{\ \psi^0 \mid \lim_{k\to+\infty} \psi^k = \psi\}$ is open in $\mathcal{S}^{\alpha,\delta}$, which implies that if there exists two distinct solutions in $\mathcal{S}^{\alpha,\delta}$ then one can partition it into two open sets, which is impossible if it is connected.*

4.4 Application to the near field parallel reflector problem

The near field parallel reflector problem (NF-par) is a non-imaging optics problem presented in Section 2.4 that cannot be recast as an optimal transport problem [47, 57], but that can be written as a generated Jacobian equation [40, 1]. We show in this section that we can apply the Damped Newton algorithm to solve this problem.

4.4.1 Generated Jacobian equation.

To have a detailed presentation of the near field parallel reflector problem (NF-par), the reader is invited to refer to Section 2.4. We recall here that X is a domain of $\mathbb{R}^2$ and Y is a finite set of size N in $\mathbb{R}^2$, both embedded in $\mathbb{R}^2 \times \{0\}$. We search a mirror surface Σ that is the graph of the function

$$u(x) = \max_{y \in Y} \frac{1}{2\psi(y)} - \frac{\psi(y)}{2}\|x - y\|^2.$$

to which we associate the map $T_\Sigma : X \to Y$ such that any ray of light $x \in X$ is reflected toward the point $T_\Sigma(x) \in Y$. Then Equation (NF-par) amounts to finding $\psi : Y \to \mathbb{R}$ such that for any $y \in Y$, $\mu(T_\Sigma^{-1}(y)) = \nu(y)$.

We define $G : \Omega \times Y \times \mathbb{R}_+^* \to \mathbb{R}$ by

$$G(x, y, v) = \frac{1}{2v} - \frac{v}{2}\|x - y\|^2 \tag{4.4.15}$$

where Ω is a bounded open set containing X. Then for every $y \in Y$, one has $T_\Sigma^{-1}(y) = \mathrm{Lag}_y(\psi)$. In order to show that the semi-discrete version of Near Field problem (NF-par) can be solved using our algorithm, we need to show that the generating function G satisfies all the hypothesis of Definition 25.

The conditions (Reg), (Mono) and (UC) are easy to verify, as mentioned in [1]. This follows from the fact that $(x, y, v) \mapsto G(x, y, v)$ is continuously differentiable in x and v, that $\nabla_x G(x, y, v) = v(y - x)$ and that $\partial_v G(x, y, v) = -1/(2v^2) - v\|x - y\|^2/2$. The (UC) condition is satisfied because Ω is bounded. Concerning the Twist assumption, F. Abedin and C. Gutierrez [1] introduce a necessary condition that they call *Visibility condition*. This condition is that for any two point $y_i, y_j \in Y$ the line containing these two points does not intersect X. Since X and Y lie in the same plane $\mathbb{R}^2 \times \{0\}$, this condition is quite restrictive in practice. We show below that it is not necessary here, since it is sufficient to have the (Twist) Condition on some interval $]0, \gamma[$ with $\gamma \in \mathbb{R}_+$.

Proposition 60. *The function G satisfies the* (Twist) *condition on $X \times Y \times]0, \gamma[$ where γ satisfies*

$$\gamma < \inf_{(x,y) \in X \times Y} \frac{1}{\|x - y\|}$$

Proof. Let $x \in X$, and suppose that $G(x, y_1, v_1) = G(x, y_2, v_2)$ and that $\nabla_x G(x, y_1, v_1) = \nabla_x G(x, y_2, v_2)$, with $v_i \in]0, \gamma]$ and $y_i \in Y$. The second condition implies that $v_1(y_1 - x) = v_2(y_2 - x)$, which implies that x, y_1 and y_2 are collinear. We then have $y_1 - x = (v_2/v_1)(y_2 - x)$. Plugging this in the relation $G(x, y_1, v_1) = G(x, y_2, v_2)$ gives

$$\frac{1}{2v_1} - \frac{v_1}{2}\frac{v_2^2}{v_1^2}\|x - y_2\|^2 = \frac{1}{2v_2} - \frac{v_2}{2}\|x - y_2\|^2$$

which gives

$$\frac{1}{2v_1}(1 - v_2^2\|x - y_2\|^2) = \frac{1}{2v_2}(1 - v_2^2\|x - y_2\|^2),$$

thus we have either $(y_1, v_1) = (y_2, v_2)$ or $v_2 = 1/\|x - y_2\|$. The latter implying that $v_2 > \gamma$, which is not possible since by assumption $v_2 \leqslant \gamma$. It follows that $y, v \mapsto (G(x, y, v), \nabla_x G(x, y, v))$ is injective on $Y \times]0, \gamma]$ for any $x \in X$. □

Having the (Twist) hypothesis only satisfied on $]0, \gamma]$ instead of $\mathbb{R}^+$ is not restrictive numerically, as we can initialize the iterate at $\psi^0 = (\alpha, \gamma, \cdots, \gamma)$ with α arbitrarily small and then guarantee that each coordinates stays below γ. In the physical problem, this condition means that each piece of paraboloid has to be "high enough".

4.4.2 Laguerre and Möbius diagram.

In order to solve the Generated Jacobian equation (NF-par) with the Damped Newton algorithm, we study the Laguerre diagram induced by the generating function G. We observe that it is a particular instance of a Möbius diagram [13]. This will be useful to get a geometric condition that implies genericity (necessary to apply Algorithm 1) and it will also be used for the numerical computation of the Laguerre diagram.

Definition 33 (Möbius diagram)**.** *The Möbius diagram of a family $\omega = (\omega_i)_{1\leqslant i\leqslant N}$ of N triplets $\omega_i = (\lambda_i, \mu_i, p_i) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$ is the decomposition of the space into Möbius cells $M_i(\omega)$ defined by*

$$M_i(\omega) = \left\{ x \in \mathbb{R}^d |\ \forall j \in [\![1, N]\!], \lambda_i \|x - p_i\|^2 - \mu_i \leqslant \lambda_j \|x - p_j\|^2 - \mu_j \right\}$$

A simple calculation shows the boundary of Möbius cells is composed of arc of (possibly degenerated) circles [13].

Proposition 61. *For any $p_i \neq p_j$, the intersection $M_i(\omega) \cap M_j(\omega)$ between two Möbius cells is either empty, or an arc of circle whose center belong to the line passing through p_i and p_j, or the bisector of p_i and p_j.*

Note that if we define $\lambda_i = \psi_i/2$, $\mu_i = 1/2\psi_i$ and $p_i = y_i$, then the Laguerre cells are Möbius cells, namely

$$\mathrm{Lag}_i(\psi) = M_i(\omega) \cap \Omega.$$

This allows to show that the conditions (Gen^Y_Ω) and $(\mathrm{Gen}^Y_{\partial X})$ that are required to show the convergence of Algorithm 1 are not restrictive. Indeed, by the previous proposition, the interface $\Gamma_{ij}(\psi)$ between the two Laguerre cells associated to y_i and y_j is contained in a circle for which the center is on the line passing through y_i and y_j. This circle can degenerate into a line, in this case it is the bisector between y_i and y_j. Suppose that Y does not contain three collinear points, then for any distinct i, j, k, $\Gamma_{ijk}(\psi)$ is the intersection of two circles with different centers and (Gen^Y_Ω) is satisfied. Similarly if ∂X doesn't contain any circle arc, nor bisectors of any two points of Y, then $(\mathrm{Gen}^Y_{\partial X})$ is also satisfied. This allows to prove the following theorem.

Theorem 62. *Suppose that Y does not contain three aligned points, and that ∂X doesn't contain any circle arc, nor bisectors of any two points of Y. Assuming that the measures μ and ν satisfy the mass balance $\mu(X) = \nu(Y)$ and that μ is absolutely continuous with a continuous density ρ such that $\mathrm{int}(X) \cap \{\rho > 0\}$ is path-connected. Then the Damped newton Algorithm (Algorithm 1) converges toward a solution of* (NF-par)*.*

Remark 63. *The Generating function is defined on $\Omega \times Y \times]0, \gamma[$ instead of $\Omega \times Y \times \mathbb{R}$. As mentioned in Remark 44, if $\zeta : \mathbb{R} \to]0, \gamma[$ is a $\mathcal{C}^1$-diffeomorphism, then the function $\widetilde{G}$ defined by $\widetilde{G}(x, y, v) = G(x, y, \zeta(v))$ is a generating function defined on $\Omega \times Y \times \mathbb{R}$ and we can apply Algorithm 1 to $\widetilde{G}$.*

4.4.3 Implementation.

The main difficulty in the implementation of the Newton algorithm is the evaluation of the function H and of its differential DH, which requires an accurate computation of the Laguerre diagram. For this, we use the fact that a Möbius diagram can be obtained by intersecting a 3D Power diagram with a paraboloid [13].

Definition 34 (Power diagram)**.** *The power diagram of a set of N weighted points $\mathcal{P} = ((p_i, r_i))_{1 \leqslant i \leqslant N}$ where $p_i \in \mathbb{R}^d$ and $r_i \in \mathbb{R}$ is the decomposition of the space into Power cells given by*

$$\mathrm{Pow}_i(\mathcal{P}) = \{x \in \mathbb{R}^d | \forall j \in [\![1, N]\!] : \|x - p_i\|^2 - r_i \leqslant \|x - p_j\|^2 - r_j\}$$

Proposition 64. *The Laguerre cells associated to the generating function G defined in* (5.3.10) *are given for any i by*

$$\mathrm{Lag}_i(\psi) = \Pi(\mathrm{Pow}_i(\mathcal{P}) \cap P) \cap \Omega,$$

where P is the paraboloid in $\mathbb{R}^3$ parametrized by $x_3 = x_1^2 + x_2^2$, Π is the projection of $\mathbb{R}^3$ on $\mathbb{R}^2$ defined by $\Pi(x, y, z) = (x, y)$, and $(\mathrm{Pow}_i(\mathcal{P}))_{1 \leqslant i \leqslant N}$ is the Power diagram associated to the weighted points $\mathcal{P}$ given by

$$\forall i \in [\![1, N]\!] : \begin{cases} p_i = \left(\dfrac{\psi_i}{2} y_i, \dfrac{-\psi_i}{4}\right) \\ r_i = \dfrac{\psi_i^2}{16} + \dfrac{\psi_i^2 \|y_i\|^2}{4} - \dfrac{\psi_i \|y_i\|^2}{2} + \dfrac{1}{2\psi_i} \end{cases} \tag{4.4.16}$$

In our implementation of the algorithm, the intersection of power diagrams with a paraboloid is computed using an algorithm presented in [52]. Once the diagram is computed, the function H and its differential DH are computed using the trapezoidal rule. Numerical experiments are performed with $X = [-1, 1]^2$ and μ equal to (one fourth) of the restriction of the Lebesgue measure on X. The set Y is randomly generated in the square $[0, 1]^2$ for different values of N, associated with a discrete uniform measure ν. Figure 5 (left) shows the initial diagram $(\mathrm{Lag}_i(\psi))_{1 \leqslant i \leqslant N}$ with $N = 5000$ for some vector $\psi = \lambda \mathbf{1}$ with $\lambda > 0$. Figure 5 (right) is the same diagram after convergence of the algorithm, where ψ is an approximate solution of (NF-par). The graph of Figure 6 represents the error $\|H(\psi^k) - \nu\|_1$ as a function of iteration k. It shows superlinear convergence of the damped Newton method.

Remark 65 (Number of iterations in Algorithm 1)**.** *Using inequality* (4.3.13) *of Theorem 58 and $\|H(\psi) - \nu\| \leqslant 2$ for any $\psi \in \mathbb{R}^N$, we get an upper bound on the number of iterations of Algorithm 1:*

$$k_{\max} = \left\lceil \frac{\ln(\epsilon/2)}{\ln(1 - \tau^*/2)} \right\rceil,$$

where $\lceil x \rceil$ denotes the smallest integer greater than x. The τ^* parameter is obtained by compactness and might depend on N. However, for a fixed N, the number of iterations of the damped Newton algorithm is $\mathcal{O}(\log(1/\epsilon))$.
The most expensive part (in terms of time) of the algorithm is the evaluation of the function H, since it requires the computation of Laguerre cells. In all our experiments we observe that τ^k is equal to 1 when $k \geqslant 5$ iterations, so that at each iteration $k \geqslant 5$, there is only one evaluation of the function H. Overall, Figure 7 shows that the overall number of evaluations of H remains low with Algorithm 1. Numerically, we also observe in Figure 6 a quadratic convergence of the Newton's algorithm.

Comparison with the iterative method of [1]

It is not straightforward to theoretically compare the computational complexities of our algorithm with the iterative method of [1]. The main reason is that the two analyses

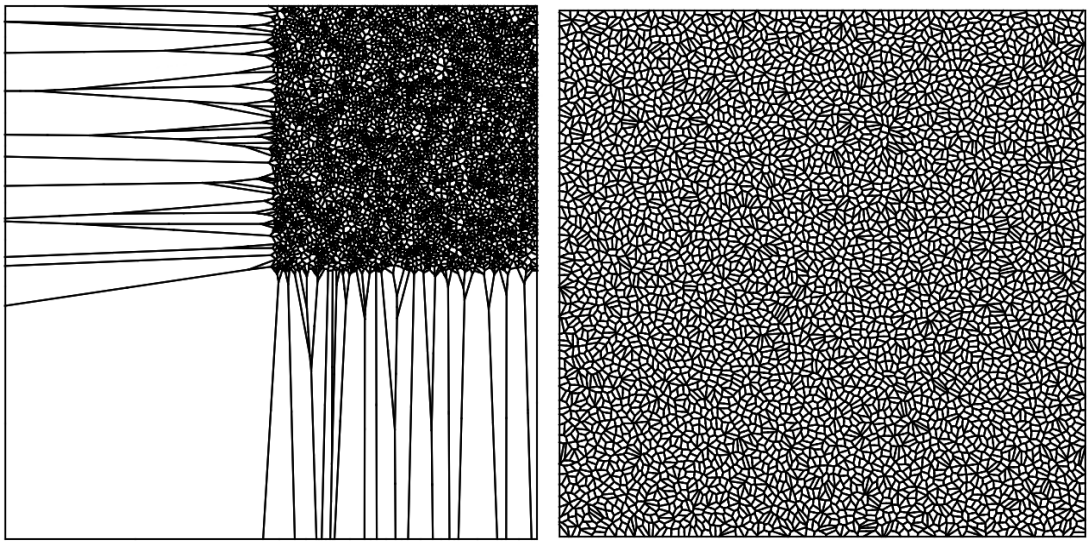

Figure 5: Initial diagram for $N = 5000$, and final diagram, after convergence of the algorithm.

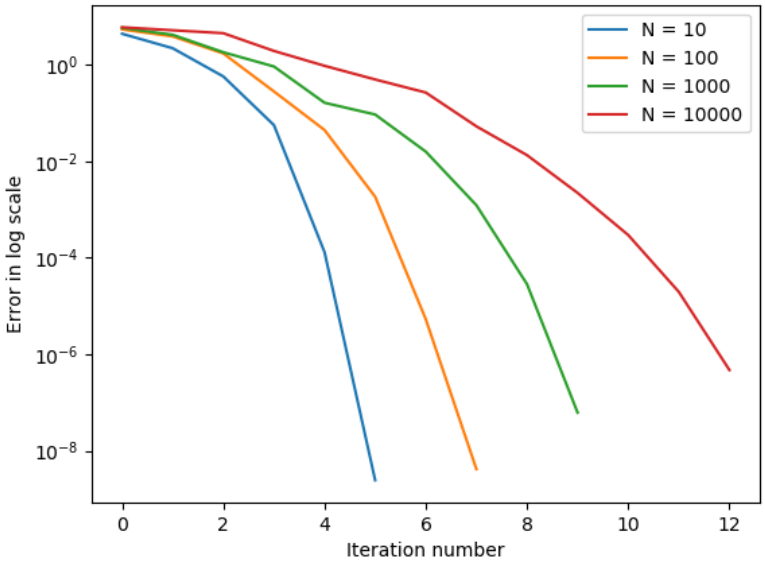

Figure 6: Numerical error $\|H(\psi^k) - \nu\|_1$ as a function of the iteration k.

Precision ϵ	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
# eval of H (our algorithm)	4	4	5	5	6	6
# eval of H (algorithm of [1])	1923	3729	5464	8317	10087	12696

Figure 7: Comparison of two algorithms for different values of ϵ and $N = 10$.

involve some constants that depend on N. The algorithm presented in [1] to solve the generated Jacobian equation at a precision ϵ requires $\mathcal{O}(N^3/\epsilon)$ evaluations of H, even disregarding the linear searches. However the constant hidden in the $\mathcal{O}(\cdot)$ notation depends on several quantities such as the (Twist) parameter λ in the Lipschitz constant of [1, Theorem 5.1], which typically depends on N. As explained above, our algorithm only requires $\mathcal{O}(\log(1/\epsilon))$ iterations, but the constant hidden in the notation $\mathcal{O}(\cdot)$ involves a constant τ^* that (a priori) also depends on N. We still note that for a fixed N, the number of iteration in our algorithm is $\mathcal{O}(\log(1/\epsilon))$ while it is $\mathcal{O}(1/\epsilon)$ for the one of [1]. Since the most computationally expensive part of both algorithms is the evaluation of the function H, we evaluate numerically their efficiency by comparing the number of evaluations of H. Figure 7 compares numerically the two algorithms when $N = 10$ and for different values of ϵ, and shows that for small values of ϵ, the number of evaluations of H can be several orders of magnitude lower with our algorithm.

On the limitations of the Newton algorithm.

The bottleneck of the algorithm is the computation of the Laguerre diagram. There exists no general algorithm to compute the Laguerre cells, each case has to be treated separately and can be quite difficult to implement. For instance, in our implementation of the Near-field reflector problem, we intersect a power diagram in $\mathbb{R}^3$ with a paraboloid to compute the Möbius diagram in $\mathbb{R}^2$. The complexity of the Möbius diagram in $\mathbb{R}^d$ is $\mathcal{O}(N\log(N) + N^{\lfloor d/2 \rfloor + 1})$ operations [13]. Still the constant hidden in the $\mathcal{O}$ notation is quite big. Another drawback is that the computation of the diagram scales exponentially badly with the dimension, which is true for any type of Laguerre diagram. This is not really a problem in optics since we always work in $\mathbb{R}^3$, but it can be in other applications of the generated Jacobian equations.

The other costly operation is to solve the linear system at each iteration. We used the LU decomposition provided by the numpy.linalg library. Even though is has a complexity of $\mathcal{O}(N^3)$ operations, it is still much quicker than the computation of the diagram. This is why we did not bother using the sparsity of the matrices to improve the computational speed of this step. Yet it is not very relevant to compare the time spent computing the diagram and solving the linear system in our experiments as we do not have an optimal implementation of all the steps. It is likely that for some big enough N the resolution of the system will take more time than the computation of the diagram in our implementation, but this N is too big for us to observe.

Chapter 5

Entropic regularization of generated Jacobian equations

In the literature, entropic regularization has been widely studied in optimal transport. It is an efficient way to modify the problem to make it somewhat more regular. Roughly speaking, entropic regularization of optimal transport consists in adding a small entropy (e.g. a Kullback-Liebler divergence) with parameter $\varepsilon > 0$ in the Kantorovich problem, that acts as a regularizer. By doing so, when passing to the dual, the maximum in the c-transform is replaced by a softmax, i.e. a regular function that converges toward a maximum as $\varepsilon \to 0$. In other words instead of partitioning the source space X in Laguerre cells, we have a partition of unity composed of smooth functions that we will name "smoothed" Laguerre cells. In the semi-discrete setting, Genevay et al [4] use these smoothed Laguerre cells to write a stochastic gradient descent algorithm for the entropic optimal transport problem.

In this chapter, we introduce an entropic regularization of generated Jacobian equations. These equations have no variational formulation, but can also be expressed using Laguerre cells. Our regularization is thus inspired of the dual formulation of regularized optimal transport, in the sense that we construct smoothed Laguerre cells for generated Jacobian by copying the formulation of the smoothed cells in optimal transport. We obtain a regularized problem for which we show that there exists solution. Once we have a regularized formulation, we can adapt the stochastic gradient descent (SGD)algorithm of Genevay et al [4] to generated Jacobian equations. Due to the nature of our equations, the algorithm is a stochastic fixed point instead of a SGD. To this day it is still a heuristic, the convergence is tested numerically on an example but there is no theoretical proof. Entropic regularization is also possible in the discrete generated Jacobian equations in economics, which in this field is known under the terminology of equilibrium matching [27].

5.1 Semi-discrete Entropic optimal transport

Before going into the details of regularized generated Jacobian equation, we will present briefly the entropic regularization of optimal transport. Through the whole section we will only consider again the semi-discrete setting, meaning that we want to transport an absolutely continuous measure $\rho \in \mathcal{P}^{\text{ac}}(X)$ toward a discrete measure $\nu \in \mathcal{P}(Y)$, where X is a compact subset of $\mathbb{R}^d$ and $Y = \{y_i\}_{1 \leqslant i \leqslant N}$ is a finite set of $\mathbb{R}^d$. Since Y is finite, the measure $\nu \in \mathcal{P}(\mathcal{Y})$ can be identified with a vector $\nu \in \mathbb{R}^N$, i.e. $\nu = \sum_i \nu_i \delta_{y_i}$ where

$\nu_i > 0$ and δ_{y_i} is the dirac measure at y_i.

5.1.1 Entropic regularization of optimal transport

The optimal transport problem is hard to solve numerically, especially because the computation of the Laguerre tessellation is quite demanding and very case dependent. To avoid these difficulties, one solution is to regularize the problem. This allows to approximate the problem by a smooth strictly convex optimization problem. This algorithm is easy to implement and can converge quickly in some cases [21]. The regularized optimal transport consists in adding an entropy term to (KP).

Definition 35 (Entropy of a measure)**.** *Let $\pi = \lambda_X \otimes \mathbf{1}_Y$ be the reference measure, where λ_X is the Lebesgue measure on X and $\mathbf{1}_Y = \sum_{y\in Y} \delta_y$ is the counting measure on Y. Then the entropy of a measure $\gamma \in \mathcal{P}(X \times Y)$ is defined by*

$$\mathcal{H}(\gamma) = \sum_{y\in Y} \int_X \gamma(x,y)(\log(\gamma(x,y)) - 1)\mathrm{d}\lambda_X(x).$$

where $\gamma(x,y)$ is the density of γ with respect to π. By convention, $H(\gamma) = +\infty$ if γ is not absolutely continuous with respect to π.

Let $\varepsilon > 0$. The semi discrete entropic optimal transport problem is the following:

$$(\mathrm{KP}^\varepsilon) := \inf_{\gamma\in\Gamma(\rho,\nu)} \int_{X\times Y} c(x,y)\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma) \qquad (\mathrm{KP}^\varepsilon)$$

As it is the case for regular optimal transport, the dual formulation of the regularized optimal transport amounts to maximizing a concave function which is called the *regularized Kantorovich function.*

Formal derivation of the dual formulation. Let $\mathcal{M}(X \times Y)$ be the set of all the measures on $X \times Y$. To write the dual problem of this formulation, we consider the Lagrangian defined for $\gamma \in \mathcal{M}(X \times Y)$ by

$$L(\gamma,\varphi,\psi) = \int_{X\times Y} c\mathrm{d}\gamma + \varepsilon\mathcal{H}(\gamma) + \int_X \varphi\mathrm{d}(\rho - \gamma(\cdot, Y)) + \int_Y \psi\mathrm{d}(\gamma(X,\cdot) - \nu)$$

where $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ are the Lagrange multipliers of the marginal constraints, and $\gamma(\cdot, Y)$ and $\gamma(X,\cdot)$ are the marginals of γ. Note that the sign conventions of the dual problem does not match the one of Chapters 2 and 3. This is done so that we recover the framework of Chapter 4 when generalizing to generated Jacobian equations. Then Equation $(\mathrm{KP}^\varepsilon)$ can be written as the unconstrained problem

$$(\mathrm{KP}^\varepsilon) = \inf_{\gamma\in\mathcal{M}(X\times Y)} \sup_{\varphi\in L^1(X),\psi\in\mathbb{R}^Y} L(\gamma,\varphi,\psi)$$

By definition the dual formulation is obtained by exchanging infimum and supremum

$$\begin{aligned}(\mathrm{DP}^\varepsilon) := \sup_{\varphi\in L^1(X),\psi\in\mathbb{R}^Y} \inf_{\gamma} \int_{X\times Y} & c(x,y) + \psi(y) - \varphi(x) \\ & + \varepsilon(\log(\gamma(x,y)) - 1)\mathrm{d}\gamma(x,y) + \int_X \varphi\mathrm{d}\rho - \sum_{y\in Y} \psi(y)\nu_y\end{aligned}$$

Since $\gamma \mapsto L(\gamma, \varphi, \psi)$ is strictly convex, we obtain by deriving with respect to γ the optimality condition

$$c(x,y) + \psi(y) - \varphi(x) + \varepsilon \log(\gamma) = 0$$

implying that for fixed functions $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$, the optimal γ^* must satisfy $\gamma^*(x,y) = e^{\frac{\varphi(x)-\psi(y)-c(x,y)}{\varepsilon}}$. Replacing γ with its optimal value we obtain the dual problem

$$(\mathrm{DP}^\varepsilon) = \sup_{\varphi,\psi} \int_X \varphi(x) \mathrm{d}\rho(x) - \sum_y \psi(y)\nu_y - \varepsilon \sum_y \int_X e^{-\frac{c(x,y)+\psi(y)-\varphi(x)}{\varepsilon}} \mathrm{d}\lambda_X(x) \qquad (\mathrm{DP}^\varepsilon)$$

For a fixed $\psi : Y \to \mathbb{R}$, the optimal $\varphi : X \to \mathbb{R}$ satisfies

$$\rho(x) - \sum_{y \in Y} e^{-\frac{c(x,y)+\psi(y)-\varphi(x)}{\varepsilon}} = 0$$

which gives

$$\varphi(x) = -\varepsilon \log \left(\sum_{y \in Y} e^{-\frac{c(x,y)+\psi(y)}{\varepsilon}} \right) + \varepsilon \log(\rho(x))$$

Plugging this in Equation $(\mathrm{DP}^\varepsilon)$ gives the following maximization problem.

Definition 36 (Dual formulation of regularized optimal transport.)**.** *The dual of* $(\mathrm{KP}^\varepsilon)$ *writes*

$$(\mathrm{DP}^\varepsilon) = \sup_{\psi \in \mathbb{R}^Y} \mathcal{K}^\varepsilon(\psi)$$

where $\mathcal{K}^\varepsilon$ *is the regularized Kantorovich function defined by*

$$\mathcal{K}^\varepsilon(\psi) = -\varepsilon \int_X \log \left(\sum_{y \in Y} e^{\frac{-c(x,y)-\psi(y)}{\varepsilon}} \right) \mathrm{d}\rho(x) - \sum_{y \in Y} \psi(y)\nu(y) + \varepsilon \mathcal{H}(\rho) \qquad (5.1.1)$$

Theorem 66 (Strong duality [19])**.** *Strong duality holds, meaning that the solutions to* $(\mathrm{KP}^\varepsilon)$ *and* $(\mathrm{DP}^\varepsilon)$ *are the same. In equation this means*

$$(\mathrm{KP}^\varepsilon) = (\mathrm{DP}^\varepsilon)$$

One way to prove strong duality is by applying the Fenchel-Rockafellar Theorem, as done in [19]. To compute the gradient of the regularized Kantorovich function $\mathcal{K}^\varepsilon$, we introduce the notion of regularized Laguerre cell. Recall $Y = \{y_i\}_{1 \leqslant i \leqslant N}$, and $\mathbb{R}^Y$ can be identified to $\mathbb{R}^N$.

Definition 37 (Regularized Laguerre cells)**.** *For* $1 \leqslant i \leqslant N$*, we define the* i*-th regularized Laguerre cells as a function of* $\psi \in \mathbb{R}^N$ *and* $x \in X$ *by*

$$\mathcal{L}_{\varepsilon,i}[\psi](x) = \frac{e^{\frac{-c(x,y_i)-\psi_i}{\varepsilon}}}{\sum_{j=1}^N e^{\frac{-c(x,y_j)-\psi_j}{\varepsilon}}}.$$

We will denote by $\mathcal{L}_\varepsilon[\psi](x) = (\mathcal{L}_{\varepsilon,i}[\psi](x))_{1 \leqslant i \leqslant N}$ *the vector of* $\mathbb{R}^N$ *composed by all the regularized Laguerre cells.*

Note that the regularized Laguerre cells forms a partition of unity, meaning that $\sum_i \mathcal{L}_{\varepsilon,i} = 1$. Remark also that the regularized Laguerre cells converge simply toward the Laguerre cells, i.e. for $\psi \in \mathbb{R}^N$ and $x \in X$,

$$\lim_{\varepsilon \to 0} \mathcal{L}_{\varepsilon,i}[\psi](x) = \mathbf{1}_{\mathrm{Lag}_i(\psi)}(x)$$

where $\mathbf{1}_A$ is the indicator function of a set A. So the functions $(\mathcal{L}_{\varepsilon,i})_i$ are indeed a regularized version of the indicator functions of the Laguerre cells.

Proposition 67 (Gradient of the regularized Kantorovich function)**.** *The Regularized Kantorovich function $\mathcal{K}^\varepsilon$ is concave and its gradient is given by*

$$\nabla \mathcal{K}^\varepsilon(\psi) = \int_X \mathcal{L}_\varepsilon[\psi](x) \mathrm{d}\rho(x) - \nu \tag{5.1.2}$$

The regularized optimal transport problem is thus reduced to finding ψ that maximizes $\mathcal{K}^\varepsilon$ or equivalently such that $\nabla \mathcal{K}^\varepsilon = 0$. The regularized semi-discrete optimal transport problem then amounts to find $\psi \in \mathbb{R}^N$ such that

$$\int_X \mathcal{L}_\varepsilon[\psi](x) \mathrm{d}\rho(x) = \nu \tag{MA$^\varepsilon$}$$

It is an unconstrained maximization of a concave function that can be solved by gradient ascent, numerical resolution of this equation is reviewed in Section 5.3.

5.2 Entropic regularization of generated Jacobian equation

In this section we propose a formulation for the entropic regularization of generated Jacobian equations, and prove the existence of solutions to this formulation. As detailed in Chapter 4, the generated Jacobian equation is a generalization of optimal transports problems that has several formulations. We recall briefly some definitions and rewrite the equation for better readability. Just as we did before we will focus here on the semi-discrete equation because of the parallel with the dual of an optimal transport problem. The framework is similar, we have an absolutely continuous measure $\rho \in \mathcal{P}^{\mathrm{ac}}(X)$ and a discrete measure $\nu \in \mathcal{P}(Y)$, where X is a compact subset of $\mathbb{R}^d$ and $Y = \{y_i\}_{1 \leqslant i \leqslant N}$ is a finite set of $\mathbb{R}^d$. The cost function is replaced by a generating function. Recall that a generating function is a map $G : X \times Y \times \mathbb{R} \to \mathbb{R}$ that satisfies several hypotheses, which are detailed in Section 4.2.1. Then we can partition the set X in generalized Laguerre cells in the same way it is done in the previous chapter. For $1 \leqslant i \leqslant N$ and $\psi \in \mathbb{R}^N$ the generalized Laguerre cells are defined by

$$\mathrm{Lag}_i(\psi) = \{x \in X \mid G(x, y_i, \psi_i) \geqslant G(x, y_j, \psi_j)\}$$

One can also define the strict Laguerre cells

$$\mathrm{SLag}_i(\psi) = \{x \in X \mid G(x, y_i, \psi_i) > G(x, y_j, \psi_j)\}$$

Recall that the (Twist) condition gives that $\rho(\mathrm{Lag}_i \cap \mathrm{Lag}_j) = 0$ for any $i \neq j$. It also guarantees that $\rho(\mathrm{Lag}_i(\psi) \setminus \mathrm{SLag}_i(\psi)) = 0$. The generated Jacobian equation consists in finding $\psi \in \mathbb{R}^N$ such that

$$H(\psi) = \nu \tag{GJE}$$

where $H : \mathbb{R}^N \to \mathbb{R}^N$ is the mass function of the Laguerre cells defined for $i \in [\![1, N]\!]$ by $H_i(\psi) = \rho(\mathrm{Lag}_i(\psi))$. As it is the case in optimal transport, the generalized Laguerre cells are complicated to compute. One can thus regularize the generalized Laguerre cells in the same way it is done for optimal transport, using a softmax for some $\varepsilon > 0$ instead of a maximum.

Definition 38 (Regularized Laguerre cells)**.** *Regularized Laguerre cells for generated Jacobian equation are a partition of unity defined for $1 \leqslant i \leqslant N$ and $\varepsilon > 0$ by the function*

$$\mathcal{L}_{\varepsilon,i}[\psi](x) = \frac{\mathcal{G}_{\varepsilon,i}(x)}{\sum_k \mathcal{G}_{k,\varepsilon}(x)}$$

with $\mathcal{G}_{\varepsilon,i}(x) = e^{G(x,y_i,\psi_i)/\varepsilon}$.

Lemma 68. *The regularized cell $\mathcal{L}_{\varepsilon,i}[\psi]$ converges simply toward $\chi_{\mathrm{Lag}_i(\psi)}$ almost everywhere on X for the Lebesgue measure. More precisely we have*

$$\forall x \in \mathrm{SLag}_i(\psi), \quad \lim_{\varepsilon \to 0} \mathcal{L}_{\varepsilon,i}[\psi](x) = 1.$$

Proof. Let $x \in \mathrm{SLag}_i(\psi)$, then we have

$$\mathcal{L}_{\varepsilon,i}[\psi](x) = \frac{1}{\sum_j \mathcal{G}_{\varepsilon,j}(x)/\mathcal{G}_{\varepsilon,i}(x)}$$

and for $j \neq i$, $\mathcal{G}_{\varepsilon,j}/\mathcal{G}_{\varepsilon,i} = e^{\frac{1}{\varepsilon}(G(x,y_j,\psi_j)-G(x,y_i,\psi_i))} \to 0$ as $\varepsilon \to +\infty$. We can then conclude since $\sum_i \mathcal{L}_{\varepsilon,i}[\psi] = 1$. □

The mass function H^ε of the regularized cells is defined by

$$H_i^\varepsilon(\psi) := \int_X \mathcal{L}_{\varepsilon,i}[\psi](x)\mathrm{d}\rho(x) \tag{5.2.3}$$

Since ρ is absolutely continuous, Lebesgue's dominated convergence theorem implies that

$$\lim_{\varepsilon \to 0} H_i^\varepsilon(\psi) = \int_X \lim_{\varepsilon \to 0} \mathcal{L}_{\varepsilon,i}[\psi](x)\mathrm{d}\rho(x) = \rho(\mathrm{Lag}_i(\psi)).$$

This allows to introduce the regularized Generated Jacobian equation

$$H^\varepsilon(\psi) = \nu \tag{GJE$^\varepsilon$}$$

Remark 69 (Link with optimal transport)**.** *Note that if we choose $G(x, y, v) = -c(x, y) - v$, then the function $H^\varepsilon - \nu$ is the gradient of the regularized Kantorovich function and Equation* (GJE$^\varepsilon$) *is exactly* (MA$^\varepsilon$)*.*

In the end of this part we show that the regularized mass function H^ε is similar to the exact one, especially it is monotone and has the same limits. We then use these results to show the existence of solutions to the regularized problem.

Lemma 70 (Properties of H^ε)**.** *Assume that X is a compact subset of $\mathbb{R}^d$.*

- *The mass function H^ε is of class $\mathcal{C}^1$ on $\mathbb{R}^N$.*
- *The function H_i^ε is decreasing in ψ_i and increasing in ψ_j for $j \neq i$.*

- *Let $\psi \in \mathbb{R}^N$ and $j \neq i$, then*

$$\lim_{\psi_i \to -\infty} H_i^\varepsilon(\psi) = 1 \quad \text{and} \quad \lim_{\psi_i \to -\infty} H_j^\varepsilon(\psi) = 0$$

Proof. The first limit comes from the uniform convergence hypothesis (UC) on the generating function G, which states that $G(x, y_i, \psi_i) \to +\infty$ as $\psi_i \to -\infty$. In that case the i-th regularized cell will be the one concentrating all the mass, and one can conclude using Lebesgue dominated convergence theorem. For the derivatives, recall that the generating function G is $\mathcal{C}^1$ in its third variable. Since X is compact, again by Lebesgue dominated convergence theorem we have

$$\frac{\partial H_i^\varepsilon}{\partial \psi_j}(\psi) = \frac{-1}{\varepsilon} \int_X \frac{\partial}{\partial \psi_j} G(x, y_j, \psi_j) \frac{e^{G(x,y_i,\psi_i)/\varepsilon} e^{G(x,y_j,\psi_j)/\varepsilon}}{\left(\sum_k e^{G(x,y_k,\psi_k)/\varepsilon}\right)^2} \mathrm{d}\rho(x)$$

which is continuous by regularity assumptions in G. Since $\sum H_j = 1$ we also have $\frac{\partial H_i^\varepsilon}{\partial \psi_i} = -\sum_{j \neq i} \frac{\partial H_j^\varepsilon}{\partial \psi_i}$ and the second limit. The monotonicity of H_i^ε then follows from the monotonicity hypothesis (Mono) on the generating function G and the formula for the partial derivatives. □

Proposition 71 (Existence of solution)**.** *Let $\varepsilon > 0$. Assume that G is a generating function (see Def 25). Then there exists $\psi \in \mathbb{R}^N$ solution of* (GJE$^\varepsilon$)*.*

Proof. Let δ such that $0 < \delta < \min_i \nu_i / N$, we are going to construct ψ^δ such that

$$\left\| H^\varepsilon(\psi^\delta) - \nu \right\|_\infty \leqslant (N-1)\delta \tag{5.2.4}$$

using the coordinate wise Oliker-Prussner algorithm [59]. First we want to initialize by putting all the mass in the first cell. Let $\gamma \in \mathbb{R}$ and $\psi = (\gamma, \cdots, \gamma)$. Since H^ε is continuous and $\lim_{\psi_1 \to -\infty} H_1^\varepsilon(\psi) = 1$, we can decrease the first coordinate of ψ to a number α such that $H_1^\varepsilon(\psi) = 1 - \min_i \nu_i / 2$. Initializing at $\psi^0 = (\alpha, \gamma, \cdots, \gamma)$, we build iteratively a sequence ψ^k by

$$\begin{cases} i^k = \min\{i \geqslant 2 \mid H_i^\varepsilon(\psi^k) \leqslant \nu_i - \delta\} \\ t^k = \min\{t \geqslant 0 \mid H_{i^k}^\varepsilon(\psi^k - t e_{i^k}) \geqslant \nu_{i^k}\} \\ \psi^{k+1} = \psi^k - t^k e_{i^k} \end{cases}$$

where $(e_i)_{1 \leqslant i \leqslant N}$ is the canonical basis of $\mathbb{R}^N$. When i^k does not exist, the algorithm terminates, in that case we put $\psi^\delta = \psi^k$ and claim that it satisfies (5.2.4). If i^k exists, then t^k always exists because of the monotonicity and limit of H_i^ε. By continuity and monotonicity properties of H^ε, all the iterates satisfy $H_i^\varepsilon(\psi^k) \leqslant \nu_i$ for $i \geqslant 2$. Since $\sum_i H_i^\varepsilon = 1$, we also have $H_1^\varepsilon(\psi^k) \geqslant \nu_1$. Through the algorithm, the first coordinate of ψ^k is fixed, so the limit $\lim_{\psi_i \to -\infty} H_1^\varepsilon(\psi) = 0$ implies that the other coordinates are bounded below by a constant $\beta \in \mathbb{R}$. All the ψ_i for $i \geqslant 2$ being decreased through the algorithm, we have that for any k, $\psi^k \in \{\alpha\} \times [\beta, \gamma]^{N-1}$. Note that the numbers α, β and γ do not depend on δ. By Lemma 70, we know that H^ε is $\mathcal{C}^1$ on $\mathbb{R}^N$ and thus Lipschitz on $\{\alpha\} \times [\beta, \gamma]^{N-1}$, let us denote by L its Lipschitz constant on this space. By construction of the sequence, at each step we have $\delta \leqslant \left\| H_i^\varepsilon(\psi^k) - H_i^\varepsilon(\psi^k - t^k e_{i^k}) \right\| \leqslant L t^k$ which implies that $t^k \geqslant \delta / L$. Since the ψ_i^k are all contained in $[\beta, \gamma]$ the sequence is constant when $k \geqslant (N-1)(\gamma - \beta)L/\delta$. At the end, for any $i \geqslant 2$, $\nu_i - \delta \leqslant H_i^\varepsilon(\psi^k) \leqslant \nu_i$, implying $\nu_1 \leqslant H_1^\varepsilon(\psi^k) \leqslant \nu_1 + (N-1)\delta$ and ψ^δ satisfies (5.2.4). This being true for any small enough $\delta > 0$, and by compactness of $\{\alpha\} \times [\beta, \gamma]^{N-1}$, we can deduce the existence of solution to the equation $H^\varepsilon(\psi) = \nu$.

□

The generated Jacobian equation (GJE$^\varepsilon$) has no variational formulation in the general case. In the next section we propose to solve it using a stochastic gradient descent, as it can be done for optimal transport.

5.3 The stochastic gradient descent algorithm

In this section we review the Stochastic Gradient Descent (SGD) algorithm, and how it can be applied to optimal transport and generated Jacobian equations. Though there exists a wide literature on the subject, this section is entirely based on a short self contained report by Turicini [67] on the subject.

5.3.1 The algorithm

Let $(\Omega, \mathcal{T}, \mu)$ be a probability space and $\mathbb{F} : \Omega \times \mathbb{R}^N \to \mathbb{R}$ be an operator depending on a random variable ω and a deterministic parameter X to be optimized. We denote

$$F(X) = \mathbb{E}[\mathbb{F}(\omega, X)]$$

where $\mathbb{E}$ is the expectation with respect to ω. We assume that the two following hypothesis are satisfied:

- The gradient of $\mathbb{F}$ is bounded:

$$\exists B > 0, \forall X \in \mathbb{R}^N \sup_{\omega \in \Omega} \|\nabla_X \mathbb{F}(\omega, X)\| \leqslant B \tag{5.3.5}$$

- There exists an $\alpha > 0$ such that F is α-strongly convex:

$$\forall X, Y \in \mathbb{R}^N, F(Y) - F(X) + \langle \nabla F(X) | Y - X \rangle \geqslant \frac{\alpha}{2} \|X - Y\|^2 \tag{5.3.6}$$

The goal of (SGD) is to find a minimizer X^* of F, which is equivalent to solving

$$\nabla F(X^*) = 0$$

The idea is to start with an $X^0 \in \mathbb{R}^N$, then to pick a random variable $\omega_k \sim \mu$ and a step size $\tau_k > 0$ at each iteration k, and iteratively compute

$$X^{k+1} = X^k - \tau_k \nabla \mathbb{F}(\omega_k, X^k)$$

The complete algorithm is detailed in Algorithm 2. Sufficient conditions for the conver-

Algorithm 2 Stochastic Gradient Descent

Require: $X^0 \in \mathbb{R}^N$, steps sizes $(\tau_k)_{k \in \mathbb{N}}$ and number of iteration $n \in \mathbb{N}$.
Ensure: X^n such that $\mathbb{E}(\|X^n - X^*\|) \leqslant C/n$

```
for k ∈ {1, ··· , n} do
    Sample ω_k from μ.
    X^{k+1} ← X^k − τ_k ∇F(ω_k, X^k)
return X^k
```

gence of SGD algorithm are given in Theorem 72, a short and self-contained proof of this classical theorem is given by Turicini [67].

Theorem 72. *Assume that $\mathbb{F}(\omega,\cdot)$ is differentiable for almost every $\omega\in\Omega$ and that hypothesis* (5.3.5) *and* (5.3.6) *are satisfied. Let $d_k=\mathbb{E}[\left\|X^k-X^*\right\|]$, then we have*

$$d_{k+1}\leqslant(1-\tau_k\alpha)d_k+\tau_k^2B$$

In particular if the sequence $(\tau_k)_{k\in\mathbb{N}}$ satisfies $\tau_k\to 0$ and $\sum\tau_k=+\infty$ then $d_k\to 0$ as $k\to+\infty$, which means that the random variable X^k converges toward X^* in the L^2 sense for random variables.

5.3.2 Application to regularized optimal transport

The regularized OT problem (MA^ε) is an unconstrained optimization problem which consists in maximizing $\mathcal{K}^\varepsilon$. We consider here a gradient ascent iterate of the type

$$\psi^{k+1}=\psi^k+\tau_k\nabla\mathcal{K}^\varepsilon(\psi^k)$$

where τ_k is a damping parameter. Since the function $\mathcal{K}^\varepsilon$ is strictly concave on $\mathbf{1}^\perp$ and has regular gradient, this algorithm is known to converge for a suitable choice of step τ_k. We have an explicit formula for $\nabla\mathcal{K}^\varepsilon$ given in Equation (5.1.2), but in order to compute it we need to integrate the regularized Laguerre cells $\mathcal{L}_{\varepsilon,i}$ against the measure ρ which can be quite costly, especially in high dimension since computation of integral scales exponentially in dimension. To get a faster algorithm, Genevay et al. applied the stochastic optimization theory to the semi-discrete entropic transport [4]. The (SGD) algorithm applies to regularized optimal transport with $F=K_\varepsilon$ and

$$\mathbb{F}(x,\psi)=-\varepsilon\log\left(\sum_{i=1}^N e^{\frac{-c(x,y_i)-\psi_i}{\varepsilon}}\right)-\sum_{i=1}^N\psi_i\nu_i+\varepsilon\mathcal{H}(\rho).$$

Note that one has

$$\nabla_\psi\mathbb{F}(x,\psi)=\mathcal{L}_\varepsilon[\psi](x)-\nu \tag{5.3.7}$$

If $x\sim\rho$ is a random variable sampled from the probability density ρ, then we have indeed $\mathbb{E}(\mathbb{F}(x,\psi))=F(\psi)$. The stochastic gradient ascent iterate for regularized optimal transport thus reads

$$\psi^{k+1}=\psi^k+\tau_k(\mathcal{L}_\varepsilon[\psi^k](x_k)-\nu) \tag{5.3.8}$$

where $(x_k)_{k\geqslant 1}$ is a sequence of iid random variables sampled from ρ. If one uses the right initialization and step size, this algorithm converges [10].

Theorem 73 (Convergence of SGD for regularized optimal transport)**.** *Let $\psi^0\in\mathbf{1}^\perp$ and $(\tau_k)_{k\in\mathbb{N}}$ be a positive sequence such that $\sum\tau_k=+\infty$ and $\sum\tau_k^2<+\infty$, then*

$$\lim_{k\to+\infty}\mathbb{E}(\mathcal{L}_\varepsilon[\psi^k](x)-\nu)=0$$

Remark 74. *The hypothesis on (τ_k) is sufficient but not necessary. Genevay et al [4] uses $\tau_k=C/\sqrt{k}$ and have a convergence rate of $O(1/\sqrt{k})$ thanks to an averaging step.*

The detailed proof is done in [10, Theorem 3.1]. We give here the main ideas. For the bounded part, it is easy to see from equation (5.3.7) that $\left\|\nabla_\psi\mathbb{F}(x,\psi)\right\|_1\leqslant 2$ and thus, hypothesis (5.3.5) is satisfied. The main issue is that the function $\mathcal{K}_\varepsilon$ is only concave and not strongly concave on $\mathbb{R}^N$. To tackle this issue one can use the following technique. First, note that both $\nabla F(\psi)=\nabla\mathcal{K}^\varepsilon(\psi)\in\mathbf{1}^\perp$ and $\nabla\mathbb{F}(x,\psi)\in\mathbf{1}^\perp$. So if we pick $\psi^0\in\mathbf{1}^\perp$,

all the iterates ψ^k stay in the subset $\mathbf{1}^\perp$. We denote by $\mathrm{diag}(v) \in \mathcal{M}_N(\mathbb{R})$ the matrix with diagonal components formed by a vector $v \in \mathbb{R}^N$, then one has

$$D^2F(\psi) = \mathbb{E}(D^2_\psi \mathbb{F}(x,\psi)) = \mathbb{E}\left[\frac{1}{\varepsilon}\left(\mathcal{L}_\varepsilon[\psi](x)\mathcal{L}_\varepsilon[\psi]^T(x) - \mathrm{diag}(\mathcal{L}_\varepsilon[\psi](x))\right)\right]$$

which has only strictly negatives eigenvalues on the set $\mathbf{1}^\perp$ (see for example [54]) and thus allows to prove convergence of the (SGD) algorithm on the subset $\mathbf{1}^\perp$.

5.3.3 Stochastic fixed point for GJE

We propose here a heuristic to numerically solve the regularized generated Jacobian equation. The idea is to replace $\nabla \mathcal{K}_\varepsilon$ by H^ε in the iterate, where H^ε is defined in (5.2.3). The algorithm for generated Jacobian equation is thus a stochastic fixed point instead of gradient descent. Recall that we want to find $\psi \in \mathbb{R}^N$ such that $H^\varepsilon(\psi) - \nu = 0$. Then starting $\psi^0 \in \mathbb{R}^N$ the fixed point algorithm reads

$$\psi^{k+1} = \psi^k + \tau^k(H^\varepsilon(\psi^k) - \nu) \tag{5.3.9}$$

If $x \sim \rho$, then we have $\mathbb{E}(\mathcal{L}_\varepsilon[\psi](x)) = H^\varepsilon(\psi)$. As in SGD, the idea is to pick an $x_k \sim \rho$ at each iteration k and apply iterate (5.3.8). This very similar to the SGD iterate, except that $H^\varepsilon(\psi)$ is not the gradient of a concave function anymore.

Algorithm 3 Stochastic fixed point for (GJE^ε)

Require: $\psi^0 \in \mathbb{R}^N$, step size $C > 0$ and number of iteration $K \in \mathbb{N}$.
Ensure: ψ such that $\mathbb{E}[\|H^\varepsilon(\psi) - \nu\|] \leqslant c/K$

```
k ← 0
for k ∈ {1, ··· , K} do
    Sample x_k from ρ.
    ψ^{k+1} ← ψ^k + (C/√k)(L_ε[ψ^k](x_k) − ν)
return ψ^k
```

The fact that Algorithm 3 is a stochastic fixed point and not a gradient descent, combined with the structure of the operator, makes the analysis of convergence more difficult. For example in optimal transport, H^ε is a gradient and thus its differential is symmetric, and it can be shown that it is diagonally dominant which makes it positive definite on the subspace $\mathbf{1}^\perp$. This is not the case for generated Jacobian equations. Nonetheless numerical experiments tend to show that the algorithm converges. One idea to prove convergence is to use some monotonicity properties of the differential DH^ε of the regularized mass function, since we still know that $\ker(DH^\varepsilon) = \mathbf{1}^\perp$.

5.3.4 Numerical results for GJE

Numerical experiments were performed in $\mathbb{R}^2$, with $X = [0,1]^2$ with the measure ρ uniform on the square, and $Y = \{y_i\}_{1 \leqslant i \leqslant 10}$ a set of 10 points chosen randomly in the square $[0,1]^2$. The discrete measure $\nu = \sum \nu_i \delta_{y_i}$ is chosen with either uniform or random weights. The considered generating function is corresponding to the near-field parallel reflector problem (NF-par), and is defined for $(x,y,v) \in X \times Y \times \mathbb{R}^*_+$ by

$$G(x,y,v) = \frac{1}{2v} - \frac{v}{2}\|x-y\|^2. \tag{5.3.10}$$

Figures 8 and 9 plots the L^1 error $\left\|H^\varepsilon(\psi^k) - \nu\right\|_1$ along the algorithm for values of k between 1 and 10^7. Figure 8 is for a uniform measure ν while Figures 9 is in the case where the ν_i are chosen randomly. The curves are plotted for different values of ε ($\varepsilon = 1$ in blue, $\varepsilon = 0.1$ in orange and $\varepsilon = 0.01$ in green). One can observe that the bigger ε is, the faster $H^\varepsilon(\psi^k)$ goes to ν as k goes to infinity. This is not surprising since it is also the case for optimal transport. When ε is bigger, the entropic term has more effect and the problem is more regular. Figure 10 shows the approximation error as an L^1-difference $\|H(\psi_\varepsilon^*) - H^\varepsilon(\psi_\varepsilon^*)\|_1$ for different values of ε. It is computed at the end of Algorithm 3, which means that ψ_ε^* is ψ^k with $k = 10^7$.

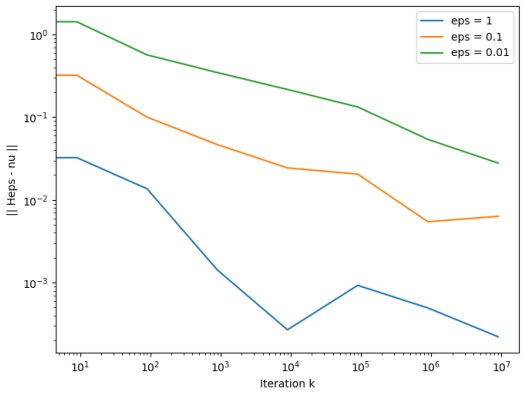

Figure 8: L^1 error $\left\|H^\varepsilon(\psi^k) - \nu\right\|_1$ at each iteration k for a uniform target measure ν.

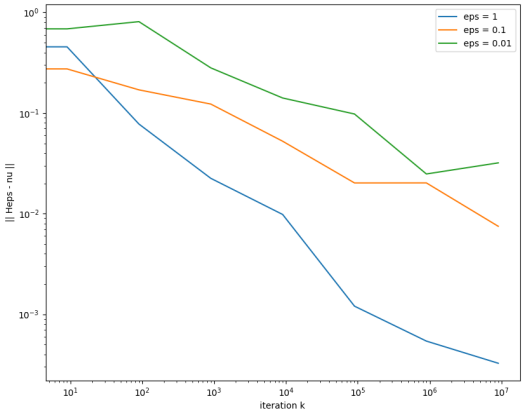

Figure 9: L^1 error $\left\|H^\varepsilon(\psi^k) - \nu\right\|_1$ at each iteration k for a random target measure ν.

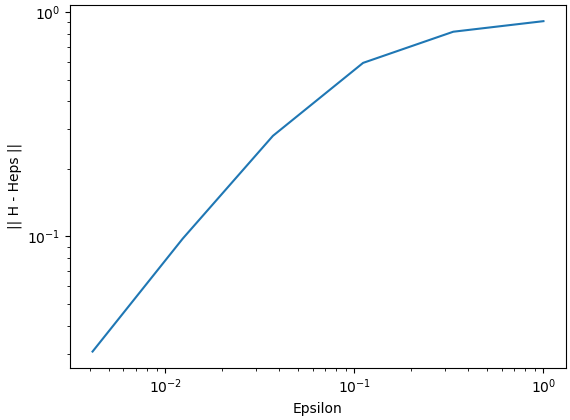

Figure 10: Convergence of $\|H(\psi_\varepsilon^*) - H^\varepsilon(\psi_\varepsilon^*)\|_1$ when $\varepsilon \to 0$. Here ψ_ε^* is computed using Algorithm 3.

Bibliography

[1] Farhan Abedin and Cristian E Gutiérrez. An iterative method for generated jacobian equations. *Calculus of Variations and Partial Differential Equations*, 56(4):101, 2017.

[2] AD Alexandrov. Convex polyhedra. springer monographs in mathematics. springer-verlag, berlin, 2005. 1950.

[3] Luigi Ambrosio, Federico Glaudo, and Dario Trevisan. On the optimal map in the 2-dimensional random matching problem. *Discrete & Continuous Dynamical Systems*, 39(12):7291–7308, 2019.

[4] Genevay Aude, Marco Cuturi, Gabriel Peyré, and Francis Bach. Stochastic optimization for large-scale optimal transport, 2016.

[5] Franz Aurenhammer, Friedrich Hoffmann, and Boris Aronov. Minkowski-type theorems and least-squares partitioning. In *SCG '92*, 1992.

[6] H.H. Bauschke and Patrick Louis Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.

[7] Jean-David Benamou. Optimal transportation, modelling and numerical simulation. *Acta Numerica*, 30:249–325, 2021.

[8] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[9] Jean-David Benamou, Brittany D Froese, and Adam M Oberman. Two numerical methods for the elliptic monge-ampere equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(4):737–758, 2010.

[10] Bernard Bercu and Jérémie Bigot. Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures, 2020.

[11] Robert J Berman. Convergence rates for discretized monge–ampère equations and quantitative stability of optimal transport. *Foundations of Computational Mathematics*, 21(4):1099–1140, 2021.

[12] Jérôme Bertrand. Prescription of gauss curvature using optimal mass transport. *Geometriae Dedicata*, 183, 08 2016.

[13] Jean-Daniel Boissonnat, Camille Wormser, and Mariette Yvinec. Curved voronoi diagrams. *Effective Computational Geometry for Curves and Surfaces*, 01 2007.

[14] David P Bourne, Charlie P Egan, Beatrice Pelloni, and Mark Wilkinson. Semi-discrete optimal transport methods for the semi-geostrophic equations. *Calculus of Variations and Partial Differential Equations*, 61(1):1–34, 2022.

[15] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[16] Giuseppe Buttazzo, Thierry Champion, and Luigi De Pascale. Continuity and estimates for multimarginal optimal transportation problems with singular costs. *Applied Mathematics & Optimization*, 78(1):185–200, 2018.

[17] Luis A Caffarelli, Sergey A Kochengin, and Vladimir I Oliker. Problem of reflector design with given far-field scattering data. *Monge Ampère equation: applications to geometry and optimization*, 226:13, 1999.

[18] Clément Cances, Thomas O Gallouët, and Gabriele Todeschi. A variational finite volume scheme for wasserstein gradient flows. *Numerische Mathematik*, 146:437–480, 2020.

[19] Lenaic Chizat. *Unbalanced Optimal Transport : Models, Numerical Methods, Applications*. Theses, Université Paris sciences et lettres, November 2017.

[20] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.

[21] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[22] Pedro Machado Manhães De Castro, Quentin Mérigot, and Boris Thibert. Far-field reflector problem and intersection of paraboloids. *Numerische Mathematik*, 134:389–411, 2016.

[23] F. de Goes, K. Breeden, V. Ostromoukhov, and M. Desbrun. Blue noise through optimal transport. *ACM Trans. Graph. (SIGGRAPH Asia)*, 31, 2012.

[24] Alex Delalande and Quentin Merigot. Quantitative stability of optimal transport maps under variations of the target measure. *arXiv preprint arXiv:2103.05934*, 2021.

[25] Sven Erlander and Neil F Stewart. *The gravity model in transportation analysis: theory and extensions*, volume 3. Vsp, 1990.

[26] Alessio Figalli and Grégoire Loeper. C 1 regularity of solutions of the monge–ampère equation for optimal transport in dimension two. *Calculus of Variations and Partial Differential Equations*, 35:537–550, 2009.

[27] Alfred Galichon, Scott Duke Kominers, and Simon Weber. Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Political Economy*, 127(6):2875–2925, 2019.

[28] Alfred Galichon and Bernard Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. 2009.

[29] Anatole Gallouët, Quentin Merigot, and Boris Thibert. A damped newton algorithm for generated jacobian equations. *Calculus of Variations and Partial Differential Equations*, 61(2):49, 2022.

[30] Thomas Gallouët, Roberta Ghezzi, and François-Xavier Vialard. Regularity theory and geometry of unbalanced optimal transport. *arXiv preprint arXiv:2112.11056*, 2021.

[31] Anatole Gallouët, Quentin Mérigot, and Boris Thibert. Strong c-concavity and stability in optimal transport, 2022.

[32] Wilfrid Gangbo and Vladimir Oliker. Existence of optimal maps in the reflector-type problems. *ESAIM Control Optim. Calc. Var.*, 13, 03 2007.

[33] Nicola Gigli. On hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proceedings of the Edinburgh Mathematical Society*, 54(2):401–409, 2011.

[34] Nestor Guillen. A primer on generated jacobian equations: Geometry, optics, economics. *Notices of the American Mathematical Society*, 66:1, 10 2019.

[35] Nestor Guillen and Jun Kitagawa. Pointwise estimates and regularity in geometric optics and other generated jacobian equations. *Communications on Pure and Applied Mathematics*, 70(6):1146–1220, 2017.

[36] Cristian Gutierrez and Qingbo Huang. The near field refractor. *arXiv preprint arXiv:1307.5709*, 2013.

[37] Cristian E Gutiérrez and Federico Tournier. Regularity for the near field parallel refractor and reflector problems. *Calculus of Variations and Partial Differential Equations*, 54(1):917–949, 2015.

[38] Geert-Jan Huizing, Gabriel Peyré, and Laura Cantini. Optimal transport improves cell–cell similarity inference in single-cell omics data. *Bioinformatics*, 38(8):2169–2177, 2022.

[39] Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166–1194, 2021.

[40] Feida Jiang and Neil S Trudinger. On pogorelov estimates in optimal transportation and geometric optics. *Bulletin of Mathematical Sciences*, 4(3):407–431, 2014.

[41] L. Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133, 03 2006.

[42] Aram Karakhanyan and Xu-Jia Wang. On the reflector shape design. *Journal of Differential Geometry*, 84, 03 2010.

[43] YH Kim and RJ McCann. Continuity, curvature, and the general covariance of optimal transportation, to appear in j. *Eur. Math. Soc*, 432:433–437, 2010.

[44] Jun Kitagawa. An iterative scheme for solving the optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 51(1-2):243–263, 2014.

[45] Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society*, 21(9):2603–2651, 2019.

[46] Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society*, 21(9):2603–2651, 2019.

[47] Sergey A Kochengin and Vladimir I Oliker. Determination of reflector surfaces from near-field scattering data. *Inverse Problems*, 13(2):363, 1997.

[48] Wenbo Li and Ricardo H Nochetto. Quantitative stability and error estimates for optimal transport plans. *IMA Journal of Numerical Analysis*, 41(3):1941–1965, 2021.

[49] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.

[50] Grégoire Loeper. Regularity of optimal maps on the sphere: The quadratic cost and the reflector antenna. *Archive for rational mechanics and analysis*, 199(1):269–289, 2011.

[51] Xi-Nan Ma, Neil S Trudinger, and Xu-Jia Wang. Regularity of potential functions of the optimal transportation problem. *Archive for rational mechanics and analysis*, 177(2):151–183, 2005.

[52] Pedro Machado Manhães De Castro, Quentin Mérigot, and Boris Thibert. Far-field reflector problem and intersection of paraboloids. *Numerische Mathematik*, 134(2):389–411, 2016.

[53] Quentin Mérigot, Jocelyn Meyron, and Boris Thibert. An algorithm for optimal transport between a simplex soup and a point cloud. *SIAM Journal on Imaging Sciences*, 11(2):1363–1389, 2018.

[54] Quentin Merigot and Boris Thibert. Optimal transport: discretization and algorithms. *Handbook of Numerical Analysis*, 22:134–212, 2021.

[55] Ludovic Métivier, Romain Brossier, Quentin Mérigot, Edouard Oudet, and Jean Virieux. Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 205(1):345–377, 2016.

[56] Georg Nöldeke and Larry Samuelson. The implementation duality. *Econometrica*, 86(4):1283–1324, 2018.

[57] Vladimir Oliker. Mathematical aspects of design of beam shaping surfaces in geometrical optics. In Markus Kirkilionis, Susanne Krömker, Rolf Rannacher, and Friedrich Tomi, editors, *Trends in Nonlinear Analysis*, pages 193–224, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[58] Vladimir Oliker. Embedding sn into rn+ 1 with given integral gauss curvature and optimal mass transport on sn. *Advances in Mathematics*, 213(2):600–620, 2007.

[59] Vladimir I Oliker and Laird D Prussner. On the numerical solution of the equation $\frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y}\right)^2 = f$ and its discretizations, i. *Numerische Mathematik*, 54(3):271–293, 1989.

[60] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[61] Cale Rankin. Distinct solutions to generated jacobian equations cannot intersect. *Bulletin of the Australian Mathematical Society*, 102(3):462–470, 2020.

[62] L. B. Romijn, J. H. M. ten Thije Boonkkamp, M. J. H. Anthonissen, and W. L. IJzerman. An iterative least-squares method for generated jacobian equations in freeform optical design. *SIAM Journal on Scientific Computing*, 43:2:B298–B322, 2021.

[63] Filippo Santambrogio. Optimal transport for applied mathematicians. calculus of variations, pdes and modeling. 2015.

[64] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[65] Neil Trudinger and Xu-Jia Wang. On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, Ser. 5, 8(1):143–174, 2009.

[66] Neil S Trudinger. On the local theory of prescribed jacobian equations. *Discrete & Continuous Dynamical Systems-A*, 34(4):1663–1681, 2014.

[67] Gabriel Turinici. The convergence of the stochastic gradient descent (sgd) : a self-contained proof. Technical report, 2021.

[68] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.

[69] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.

[70] Xu-Jia Wang. On the design of a reflector antenna. *Inverse Problems*, 12(3):351, jun 1996.

[71] Xu-Jia Wang. On the design of a reflector antenna ii. *Calculus of Variations and Partial Differential Equations*, 20(3):329–341, 2004.