Local regularity result for an optimal transportation problem with rough measures in the plane

P.-E. Jabin[*], A. Mellet[†], M. Molina-Fructuoso [‡]

Abstract

We investigate the properties of convex functions in $\mathbb{R}^2$ that satisfy a local inequality which generalizes the notion of sub-solution of Monge-Ampère equation for a Monge-Kantorovich problem with quadratic cost between *non-absolutely* continuous measures. For each measure, we introduce a discrete scale so that the measure behaves as an absolutely continuous measure up to that scale. Our main theorem then proves that such convex functions cannot exhibit any flat part at a scale larger than the corresponding discrete scales on the measures. This, in turn, implies a C^1 regularity result up to the discrete scale for the Legendre transform. Our result applies in particular to any Kantorovich potential associated to an optimal transportation problem between two measures that are (possibly only locally) sums of uniformly distributed Dirac masses. The proof relies on novel explicit estimates directly based on the optimal transportation problem, instead of the Monge-Ampère equation.

1 Introduction

1.1 Generalized Monge-Ampère equation for rough measures

In this paper, we investigate the properties of convex functions in $\mathbb{R}^2$ that can be seen as *local* one-sided Kantorovich potentials. More specifically, we consider a (continuous) convex function $\psi : \mathbb{R}^2 \to \mathbb{R}$ that satisfies

$$\mu(A) \leq \nu(\partial\psi(A)) \quad \text{for all Borel sets } A \subset \Omega \tag{1}$$

where μ and ν are two non-negative Radon measures on $\mathbb{R}^2$ and Ω is an open bounded subset of $\mathbb{R}^2$.

We recall that the subdifferential $\partial\psi$ of the function ψ is given by

$$\partial\psi(x) = \left\{ z \in \mathbb{R}^2 \,;\, \psi(y) \geq \psi(x) + z \cdot (y - x) \text{ for all } y \in \mathbb{R}^2 \right\}$$

and that $\partial\psi(A) = \cup_{x\in A}\partial\psi(x)$.

[*]pejabin@psu.edu, Pennsylvania State University, Department of Mathematics and Huck Institutes, State College, PA 16802. Partially supported by NSF DMS Grant 161453, 1908739, 2049020 and NSF Grant RNMS (Ki-Net) 1107444.

[†]mellet@umd.edu, University of Maryland, Department of Mathematics and CSCAMM, College Park, MD 20742 USA. Partially supported by NSF Grant NSF Grant DMS-1501067 and DMS-2009236.

[‡]mmolina2@ncsu.edu, North Carolina State University, Department of Mathematics, Raleigh, NC 27695

We note that (1) is a *local condition* as it is only posed in a subset $\Omega \subset \mathbb{R}^2$. As we will explain shortly, Inequality (1) appears naturally when ψ is a Kantorovich potential associated to the optimal transportation problem between two probability measures (although we do not need to require μ and ν have the same mass in our paper).

When μ and ν satisfy

$$d\mu(x) \geq \frac{1}{\lambda} dx \quad \text{in } \Omega, \qquad d\nu(x) \leq \lambda dx \quad \text{in } \partial\psi(\Omega) \tag{2}$$

for some $\lambda > 0$, then (1) implies that

$$\frac{1}{\lambda^2}|A| \leq |\partial\psi(A)| \quad \text{for all Borel sets } A \subset \Omega$$

which says that ψ is a solution (in the *Alexandrov sense*) of the one-sided Monge-Ampère equation

$$\det(D^2\psi(x)) \geq \frac{1}{\lambda^2} \quad \text{in } \Omega. \tag{3}$$

In the case of absolutely continuous measures, inequality (1) is hence a reformulation of sub-solutions to the Monge-Ampère equation in terms of optimal transportation.

In dimension 2, it is known that (3) implies that ψ is strictly convex in Ω (see Theorem 1.5 below) and the strict convexity of ψ is the first step in the regularity theory for the solutions of the full Monge-Ampère equation (17) and the corresponding optimal transportation problem (see Section 1.4 for further presentation of relevant results and references). In dimension 2 and only in dimension 2 (see below again), this is a *local property* in the sense that it does not require any boundary condition on ψ: Whenever ψ satisfies (3) on any subdomain Ω, it is strictly convex in the interior of that subdomain, independently of what occurs outside of Ω.

The goal of this paper is to extend this result when ψ satisfies (1) with measures μ and ν that are not necessarily absolutely continuous with respect to the Lebesgue measure and satisfy some lower and upper bounds only up to a certain scale (Assumption 1 below). This includes the case where μ and ν are sums of uniformly distributed Dirac masses. Our main theorem (Theorem 1.2) implies in particular that ψ is then strictly convex up to a certain scale (in Corollary 1.3, we derive a bound on the diameter of the flat parts of ψ). Equivalently, our result says that the Legendre transform of ψ is C^1 regular up to a certain scale (see Corollary 1.4) on a subset of Ω.

Outline of the paper: Our first result, Theorem 1.1 in Section 1.2 below, makes precise the relation between inequality (1) and optimal transportation theory. It is proved in Section 5 and is of independent interest. The strict convexity "up to a certain scale", which is the main topic of this paper is then presented in Section 1.3 together with several consequences. This is followed by an overview of the existing literature and results related to our work (Section 1.4). The proof of our main theorem is given in Section 2, while Sections 3 and 4 are devoted to the proofs of Corollaries 1.3 and 1.4.

1.2 Relation to Optimal transportation problem

Inequality (1) is natural in the context of optimal transportation theory (in any dimension $n \geq 1$). To explain this connection, we first recall that given two probability measures $\mu \in \mathscr{P}(\mathbb{R}^n)$, $\nu \in \mathscr{P}(\mathbb{R}^n)$

the *Monge-Kantorovich Problem* with quadratic cost is concerned with the minimization problem

$$\min_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 d\pi(x,y) \tag{4}$$

where $\Pi(\mu,\nu)$ denotes the set of all probability measures $\pi \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with marginals μ and ν, i.e. such that $\pi(A\times\mathbb{R}^n) = \mu(A)$ for all $A \subset \mathbb{R}^n$ and $\pi(\mathbb{R}^n \times B) = \nu(B)$ for all $B \subset \mathbb{R}^n$.

The existence of a minimizer π for problem (4) (which will be called an *optimal transport plan* between the measures μ and ν) is a classical result, see for example [23]. Moreover, it is known that $\pi \in \Pi(\mu,\nu)$ is a minimizer for (4) *if and only if* it is supported on the graph of the subdifferential of a lower semi-continuous convex function ψ:

$$\mathrm{supp}(\pi) \subset \mathrm{Graph}(\partial\psi) := \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n \mid z \in \partial\psi(x)\}\,. \tag{5}$$

The function ψ is called a Kantorovich potential for problem (4) and we can easily check that it satisfies (1) (globally). Indeed, for all Borel sets $A \subset \mathbb{R}^n$, we can write

$$\mu(A) = \int_A \int_{\mathbb{R}^n} d\pi = \int_A \int_{\partial\psi(A)} d\pi \leq \int_{\mathbb{R}^n} \int_{\partial\psi(A)} d\pi = \nu(\partial\psi(A)).$$

We note that this inequality might be strict in general but that we can similarly prove the inequality

$$\mu(\partial\psi^*(A)) \geq \nu(A) \tag{6}$$

where ψ^* denotes the Legendre transform defined by

$$\psi^*(z) = \sup_{x\in\mathbb{R}^n} \big(x\cdot z - \psi(x)\big).$$

(indeed, the duality relation $z \in \partial\psi(x) \Leftrightarrow x \in \partial\psi^*(z)$ together with (5) implies $\mathrm{supp}(\pi) \subset \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n \mid x \in \partial\psi^*(z)\}$.)

The purpose of this paper is to derive strict convexity estimates on ψ when we assume only that (1) *holds locally* (and we do not require (6)). If we assume that both (1) and (6) hold (locally), then we get some convexity estimates on ψ^*, which are equivalent to C^1 estimates on ψ.

A question that arises naturally is whether assuming (1) and (6) is equivalent to assuming that ψ is associated to an optimal transport plan between μ and ν. The answer is of course straightforward for the Monge-Ampère equation since a function that is a sub-solution and a super-solution is necessarily a solution. It is hence natural for (1) and (6) to imply a similar result, but the situation is more delicate. We recall in particular that the Monge-Kantorovich potential is not unique when the measures μ and ν are not absolutely continuous with respect to the Lebesgue Measure.

To the authors' knowledge, this question has not been previously addressed in the literature and we will show that if ψ satisfies (1) and (6) globally (so in $\mathbb{R}^n$) and if μ and ν have finite second moment, then ψ must indeed be a Kantorovich potential associated to the optimal transportation problem (4). More precisely, we prove the following result:

Theorem 1.1. *Let μ, ν be two probability measures on $\mathbb{R}^n$ with finite second moment:*

$$\int_{\mathbb{R}^n} |x|^2 d\mu(x) + \int_{\mathbb{R}^n} |y|^2 d\nu(y) < \infty$$

and let ψ be a proper lower semi-continuous convex function such that (1) *and* (6) *hold for all measurable sets A. Then there exists an optimal plan $\pi \in \Pi(\mu,\nu)$ (minimizer of* (4)*) supported on $\Gamma = \operatorname{Graph} \partial\psi$ and the pair (ψ,ψ^*) is then a minimizer for the dual problem*

$$\inf\left\{\int \psi d\mu + \int \varphi\, d\nu\,;\, x\cdot y \le \psi(x)+\varphi(y) \quad \forall (x,y)\right\}.$$

This result is completely independent from the regularity theory that we develop in the rest of this paper, but it clarifies the relation between equation (1) and the optimal transportation problem. The proof (see Section 5) relies on the approximation of the measures μ and ν by sums of Dirac masses to construct a plan $\pi \in \Pi(\mu,\nu)$ supported on $\Gamma = \operatorname{Graph}\partial\psi$ (the optimality of such a plan then follows from classical results from optimal transportation theory).

1.3 Local convexity and regularity

The main goal of this paper is to develop a regularity theory when we do not assume that μ and ν are absolutely continuous with respect to the Lebesgue measure, but that they satisfy the following assumption:

Assumption 1. *Assume that there are constants $h_1, h_2 > 0$ and $\lambda_1, \lambda_2 > 0$, such that the measures μ and ν satisfy*

$$\mu(R) \ge \frac{|R|}{\lambda_1} \;\text{ and }\; \nu(R' \cap \partial\psi(\Omega)) \le \lambda_2 |R'| \tag{7}$$

for any rectangles $R \subset \Omega$, $R' \subset \mathbb{R}^2$ with dimensions at least h_1 and h_2 in every direction for R and R' respectively.

This assumption is less restrictive than (2) and is relevant in the framework of optimal transportation. In fact, the original problem considered by Kantorovich in [18] included measures μ and ν that are sums of Dirac masses rather than absolutely continuous measures. This setting is also important for numerical applications: Measures satisfying Assumption 1 appear naturally when introducing discrete approximations of absolutely continuous measures with bounded densities, as is often done for computational purposes.

In order to state our main result, we introduce the set

$$\Omega^\delta = \{x \in \Omega\,;\, \operatorname{dist}(x,\partial\Omega) > \delta\}.$$

The main result of this paper is the derivation of the technical inequality (10) below, which quantify the strict convexity of ψ up to a scale depending on h_1 and h_2:

Theorem 1.2 (Strict convexity of Kantorovich potentials ψ)**.** *Let $\psi : \mathbb{R}^2 \to \mathbb{R}$ be a convex function satisfying* (1) *for some measures μ and ν satisfying Assumption 1. Given $\delta > 0$ and $(x,y) \in \Omega^\delta \times \Omega^\delta$, let K be any constant satisfying*

$$K \ge \operatorname{diam} \partial\psi(U_\delta) \tag{8}$$

where U_δ is a δ-neighborhood of the segment $[x,y]$ (i.e. $U_\delta = \cup_{z\in[x,y]} B_\delta(z)$) and define

$$\varepsilon = -\min_{t\in[0,1]} \psi((1-t)x+ty) - [(1-t)\psi(x)+t\psi(y)] \ge 0.$$

There exists a universal constant C such that if the length $\ell := |x-y|/2$ satisfies

$$\ell \geq 2\,h_1, \quad \ell^2 \geq C\,K\,\lambda_1\,\lambda_2\,h_2\,, \tag{9}$$

then the following inequality holds:

$$\ell^8\,\log\left(1+\frac{\delta}{\gamma}\right) \leq C\,\lambda_1^4\,\lambda_2^4\,K^8, \tag{10}$$

provided

$$\gamma := \max\left\{\frac{\varepsilon}{K}, 2h_1, \frac{\ell h_2}{C\,K}\right\} \leq \delta/2.$$

We immediately make the following remarks:

1. The logarithm in the left hand side of (10) goes to infinity when γ goes to zero. So Theorem 1.2 provides a lower bound on γ depending on the quantity $\frac{\ell^2}{C\lambda_1\lambda_2 K^2}$. Indeed we have that either $\gamma > \delta/2$ or inequality (10) provides a lower bound for γ. So with the notations of the theorem, we see that as long as (9) holds, we have

$$\gamma \geq \delta \min\left\{\frac{1}{\exp\left(\frac{C^4\lambda_1^4\lambda_2^4K^8}{\ell^8}\right)-1}, \frac{1}{2}\right\}. \tag{11}$$

   Note that we can take $K = \operatorname{diam}\partial\psi(\Omega)$ which does not depends on ℓ.

2. The conditions (9) are clearly satisfied in the absolutely continuous case $h_1 = h_2 = 0$. In that case, (11) gives a lower bound on $\gamma = \varepsilon/K$ and implies the strict convexity of ψ. We actually recover a classical result, see Theorem 1.5 below.

3. The proof will make it clear that the assumption $(x,y) \in \Omega^\delta \times \Omega^\delta$ in the theorem is not necessary. The result holds for $(x,y) \in \Omega \times \Omega$ provided there is a rectangle $R_\delta(x,y)$, with base equal to the line segment $[x,y]$ and height equal to δ which is contained in Ω. In this setting we can also take $K = \operatorname{diam}(\partial\psi(R_\delta(x,y)))$.

4. As mentioned above, the conditions (9) are trivially satisfied when $h_1 = h_2 = 0$. When $h_1, h_2 \neq 0$, it is clear that we need some conditions on ℓ since we expect the potential ψ to have flat parts and so $\varepsilon = 0$ if ℓ is small enough. In the simple case where μ and ν are uniformly distributed Dirac masses (on lattices of characteristic length h_1 and h_2), then the first condition in (9) is necessary to have several lattice points in the set U_δ, while the second condition in (9) will guarantee that all those points cannot be sent onto a thin rectangle (of height h_2).

5. The result is consistent with the natural scaling of the problem. For example, if we replace the measure ν by the new measure $\tilde{\nu}$ defined by $\tilde{\nu}(R) := \nu(\tau R)$ for some fixed $\tau > 0$, then $\tilde{\nu}$ satisfies the conditions of Assumption 1 with $\tilde{h}_2 = \tau^{-1}h_2$ and $\tilde{\lambda}_2 = \tau^2\lambda_2$. Furthermore, the function $\tilde{\psi} = \tau^{-1}\psi$ is a Kantorovich potential associated to the measures μ and $\tilde{\nu}$ which satisfies inequality (1) (with $\tilde{\nu}$ instead of ν). One can then check that the conditions (9) and the inequality (10) are unchanged by these transformations.

Theorem 1.2 provides a way to quantify how close ψ is to being strictly convex. For instance, we can use Theorem 1.2 to estimate the largest possible length of a "flat part" of ψ by assuming that $\varepsilon = 0$ and using (10) to get an upper bound on ℓ. We get:

Corollary 1.3. *Under the conditions of Theorem 1.2, assume furthermore that $\varepsilon = 0$ (that is, ψ is affine on the segment $[x, y]$).*

If $h_1 \le \delta/4$, $\sqrt{C\lambda_1\lambda_2}h_2 \le \delta$ and

$$\frac{\ell h_2}{K} \le \frac{C\delta}{2} \tag{12}$$

then

$$\ell \le \max\left\{ 2h_1, \sqrt{C\lambda_1\lambda_2}\, Kh_2, \frac{K\sqrt{C\lambda_1\lambda_2}}{\left[\ln\left(\frac{\delta}{2h_1}\right)\right]^{1/8}}, \frac{K\sqrt{C\lambda_1\lambda_2}}{\left[\ln\left(\frac{\delta}{\sqrt{C\lambda_1\lambda_2}h_2}\right)\right]^{1/8}} \right\}. \tag{13}$$

We recall that we can take $K = \operatorname{diam}\partial\psi(\Omega)$ in which case (12) reads $\ell h_2 \le C\operatorname{diam}\partial\psi(\Omega)$ and (13) gives an upper bound on ℓ which only depends on the data of the problem and goes to zero when $\max\{h_1, h_2\} \to 0$. When $h_1 = h_2 = 0$, Corollary 1.3 gives $\ell = 0$, so we recover the classical result that ψ must be strictly convex in that case (no flat parts).

We can also take $K = \operatorname{diam}\partial\psi(U_\delta)$ (so that (13) is sharper) in which case we note that if $h_2^2 \le \frac{\delta\ell}{\lambda_1\lambda_2}$ then we can use the estimate (29), derived further in the proof, to replace the condition (12) with the following condition that does not depend on ℓ:

$$\sqrt{C\lambda_1\lambda_2}h_2 \le \frac{3\delta^{3/2}}{(\operatorname{diam}\Omega_2)^{1/2}}.$$

Going back to Theorem 1.2, we observe that the control it provides on the convexity of ψ should imply some C^1 regularity up to some length scale depending on h_1, h_2 on the Legendre dual or conjugate (see [20]) defined for all $z \in \mathbb{R}^2$ by

$$\psi^*(z) = \sup_{x\in\Omega}\big(x\cdot z - \psi(x)\big). \tag{14}$$

Indeed, we can show:

Corollary 1.4 (C^1 regularity of ψ^*)**.** *Under the conditions of Theorem 1.2 and given $\delta > 0$, there exists some functions $\rho(\ell)$, $\rho_1(\ell)$ and $\rho_2(\ell)$ monotone increasing, with limit 0 when $\ell \to 0^+$, and depending only on δ, $\lambda_1\lambda_2$, $D = \operatorname{diam}\Omega$ and K such that for all $z, z' \in \partial\psi(\Omega^\delta)$, we have*

$$|x - x'| \le \max\left(\rho(|z - z'|), \rho_1(h_1), \rho_2(h_2),\right) \qquad \forall x \in \partial\psi^*(z),\ x' \in \partial\psi^*(z'). \tag{15}$$

In particular if $h_1 = h_2 = 0$ then ψ^* is C^1 in $\partial\psi(\Omega^\delta)$ with the explicit estimate on the modulus of continuity of $\nabla\psi^*$,

$$|\nabla\psi^*(z) - \nabla\psi^*(z')| \le C\,\frac{\sqrt{\lambda_1\,\lambda_2}\,L_\infty}{\left(\log\left(1 + \frac{1}{C\sqrt{\lambda_1\,\lambda_2}\,|z-z'|}\right)\right)^{1/8}} \tag{16}$$

where L_∞ now denotes the Lipschitz bound of ψ over Ω.

We conclude this presentation of our result by observing that in Assumption 1 we only require a lower bound on μ and an upper bound on ν. Such bounds are all that we need to study the strict convexity of ψ. Opposite bounds would be required to prove the C^1 regularity of ψ up to a certain

length scale (recall that ψ^* is associated to an optimal transportation problem in which the roles of μ and ν are inverted). More precisely, if we assume that

$$\mu(R \cap \Omega) \leq \lambda_2 |R| \quad \text{and} \quad \nu(R') \geq \frac{1}{\lambda_1} |R'|$$

for $R \subset \mathbb{R}^2$ and $R' \subset \Lambda$ (up to a certain scale), then using inequality (6) instead of (1) our analysis yields the C^1 regularity up to a certain scale for the potential ψ on the set

$$\Omega' = \bigcup_{\delta > 0} \partial\psi^*(\Lambda^\delta)$$

where $\Lambda^\delta = \{x \in \Lambda \,;\, \mathrm{dist}(x, \partial\Lambda) > \delta\}$.

1.4 Brief overview of the literature

When the measures $\mu = f\,dx$ and $\nu = g\,dx$ are absolutely continuous and concentrated on the open sets Ω and Λ, a classical result due to Brenier ([3, 4]) states that the solution of the minimization problem (4) is unique and is given by $\pi = (Id \times \nabla\psi)_{\#}\mu$, where $\psi : \mathbb{R}^n \to \mathbb{R}$ is a globally Lipschitz convex function such that $\nabla\psi_{\#}\mu = \nu$.

If furthermore, there exists $\lambda > 0$ such that $1/\lambda \leq f, g \leq \lambda$ on their respective supports, then ψ satisfies

$$\frac{1}{\lambda^2}\chi_\Omega \leq \det D^2\psi \leq \lambda^2 \chi_\Omega \tag{17}$$

in a weak sense (the *Brenier sense*) together with the boundary condition $\nabla\psi(\mathbb{R}^n) \subset \Lambda$ (see for instance [8, 10]).

Even in that case, it is classical that the regularity of ψ requires some condition on the support of g (for example if Ω is connected but Λ is not, then the map $\nabla\psi$ must be discontinuous). Caffarelli proved in [8] that if we further assume that Λ is *convex*, then ψ is a strictly convex solution of the Monge-Ampère equation (17) in the following *Alexandrov sense*:

$$\frac{1}{\lambda^2}|A \cap \Omega| \leq |\partial\psi(A)| \leq \lambda^2 |A \cap \Omega| \qquad \text{for any Borel set } A \subset \mathbb{R}^n. \tag{18}$$

The regularity theory for Monge-Ampère equation developed by Caffarelli [5, 6, 7, 8, 9] for strictly convex solutions of (18) then implies that ψ is $C^{1,\alpha}_{loc}$.

Even in this absolutely continuous framework, our result requires only inequality (1) - which is equivalent to the lower bound in (18) - and *no assumption on* Λ. We note that this inequality is always satisfied by Brenier's potential, while the upper bound in (18) requires further assumptions on Λ (e.g. convexity) to hold. When Λ is non-convex, the convex potential ψ cannot be expected to be C^1 everywhere. However, partial regularity results have been derived that offer a useful comparison, first in dimension 2 by Yu [24] and Figalli [12] and then generalized to higher dimension in [13] and to more general cost functions in [11]. These results show in particular that there exists an open subset of Ω of full measure in which ψ is $C^{1,\alpha}$ and strictly convex. In dimension 2, a precise geometric description of the singular set can be given [12]. Our argument provides explicit quantitative estimates in that sense that extend to non absolutely continuous measures.

In this paper, we do not need to assume that ψ is associated to an optimal transportation problem with nice properties globally. We only require inequality (1) to hold for measures μ and ν that satisfy

some lower and upper bounds in some subsets of their supports. In dimension 3 and higher, functions that satisfy (17) in Ω might not be strictly convex, as shown by Pogorelov's classical counterexample [16] and the regularity theory for such solutions requires appropriate assumptions on the boundary conditions [5, 6, 7, 9]. However, in dimension 2 (which is the setting of our main result), there is a simple proof of the strict convexity of smooth local solutions of (17), which was originally proved in [1] and [17] by Aleksandrov and Heinz independently (see also [21]). That result can be formulated as follows:

Theorem 1.5 ([1], [17]). *For $n = 2$, let $\psi \in C^2_{loc}(\Omega)$ satisfy*

$$\det D^2\psi \geq \lambda^{-2} > 0 \qquad \text{in } \Omega, \tag{19}$$

and assume that $\psi \geq 0$ in Ω and $\psi(x_0) = 0$ for some x_0 in the interior of Ω. Denote $\delta := dist(x_0, \partial\Omega) > 0$ and let H be any line passing through x_0. Then for all $\ell \leq \frac{\delta}{2}$, the quantity

$$\gamma = \frac{\sup_{x\in B_\ell(x_0)\cap H} \psi(x)}{\|\nabla\psi\|_{L^\infty(\Omega)}}.$$

satisfies

$$\ell^2 \ln\left(1 + \frac{\delta}{\gamma}\right) \leq 8\lambda^2 \|\nabla\psi\|^2_{L^\infty(\Omega)}. \tag{20}$$

Inequality (20) implies the following estimate:

$$\sup_{x\in B_\ell(x_0)\cap H} \psi(x) \geq \frac{\|\nabla\psi\|_\infty \delta}{\exp\left(\dfrac{8\lambda^2\|\nabla\psi\|^2_\infty}{\ell^2}\right) - 1} > 0 \tag{21}$$

for all $\ell \leq \frac{\delta}{2}$.

Our Theorem 1.2 with $h_1 = h_2 = 0$ gives a proof of this result, and in particular estimate (20), when we assume only that ψ is an *Alexandrov* solution of (19), that is a convex function satisfying

$$|\partial\psi(A)| \geq \frac{1}{\lambda^2}|A \cap \Omega| \quad \text{ for all Borel sets in } \Omega.$$

But the main interest of our result is that we consider measures that may not be absolutely continuous with respect to the Lebesgue measure. In this case, the Kantorovich potential ψ (which still exists but may not be unique) does not solve the Monge-Ampère equation (either (17) or (18)). To our knowledge, no quantitative estimates on the convex function ψ are known in this setting. Brenier's result does not apply (there might not be any measurable map T such that $T_\#\mu = \nu$), and Kantorovich potentials are not expected to be either convex (they will have 'flat parts') nor C^1 (they will have 'corners').

Theorem 1.2 is the derivation of an inequality similar to (20) when (19) is replaced by (1) with measures satisfying Assumption 1 with $\lambda^2 = \lambda_1\lambda_2$ and γ replaced by

$$\max\left\{\gamma, 2h_1, \frac{\ell h_2}{C\,K}\right\}.$$

This implies in particular that (in dimension 2) any Kantorovich potential ψ is strictly convex up to some scale depending on h_1 and h_2 in any open set in which the lower and upper bounds (7) hold.

Although our result is similar to Theorem 1.5, the proof is completely different since we cannot use the Monge-Ampère equation (19) in this non absolutely continuous setting and instead we must rely solely on the measure inequality (1). In the groundbreaking work of Caffarelli as well as in the partial regularity theory of [24, 12, 13], a key tool is the use of some variants of the maximum principle for the Monge-Ampère equation and the use of appropriate barriers. This is not possible in our framework. Instead, the proof of our Theorem 1.2 relies on the derivation of upper and lower bounds for an integral quantity defined in (31)-(32). Note that a variational approach (relying on optimal transportation arguments rather than using some barriers for Monge-Ampère equation) to the partial regularity theory of [13] was recently developed in [15, 14].

It is natural to ask whether our result could be extended to dimension $n \geq 3$. It turns out that even in the absolutely continuous case, the result of Theorem 1.5 does not hold in dimension 3 and higher. Indeed, a classical example by Pogorelov shows that ψ can have a flat part and is thus not necessarily strictly convex (see [16]). A natural extension of Theorem 1.5 can however be found in [2, Theorem 2.34] : Under conditions similar to Theorem 1.5 but in dimension $n \geq 3$, the convex function ψ cannot be affine on a set of dimension larger than or equal to $n/2$. For the sake of completeness, we present in Appendix B a short proof, based on the ideas of [2], of the following quantitative estimate

Theorem 1.6. *Let $n \geq 3$ and let $\psi \in C^2$, $\psi \geq 0$ satisfy $\det D^2\psi \geq \lambda^{-2} > 0$ and assume that $\psi(x_0) = 0$ with $\delta := dist(x_0, \partial\Omega) > 0$. Let H be an affine surface of dimension d passing through x_0, then for all $\ell \leq \frac{\delta}{2}$, the quantity*

$$\gamma = \frac{\sup_{x \in B^n_\ell(x_0) \cap H} \psi(x)}{\|\nabla \psi\|_{L^\infty(\Omega_1)}}$$

satisfies

$$\ell^{2d} \varphi(\delta/\gamma) \leq C\lambda^2 \|\nabla\psi\|^n_{L^\infty(\Omega_1)} \delta^{2d-n} \tag{22}$$

with $\varphi(s) := s^{2d-n} \int_0^s \frac{r^{n-d-1}}{(r+1)^d}\, dr$.

We note that φ satisfies $\lim_{s\to\infty} \varphi(s) = \infty$ if and only if $d \geq n/2$ and so (22) implies the following lower bound:

$$\sup_{x \in B^n_\ell(x_0) \cap H} \psi(x) \geq \begin{cases} \min\left\{ \delta\|\nabla\psi\|_\infty, \left(\frac{\ell^{2d}}{C\lambda^2\|\nabla\psi\|_\infty^{2n-2d}}\right)^{\frac{1}{2d-n}} \right\} & \text{if } d > n/2 \\ \delta\|\nabla\psi\|_\infty \exp\left(-C\frac{\lambda^2\|\nabla\psi\|_\infty^n}{\ell^n}\right) & \text{if } d = n/2. \end{cases} \tag{23}$$

In view of this result, it seems that the main result of this paper (Theorem 1.2) could be extended to higher dimensions, provided one considers hypersurfaces of dimension $n/2$. However, the basic tool of our proof, the integral quantity (31)-(32), is not well suited for such a generalization, and a new quantity would need to be introduced. This question will thus be addressed in a future work.

2 Proof of Theorem 1.2

2.1 Preliminaries

The Kantorovich problem with the quadratic cost function is invariant under rigid motions. Up to a translation and a rotation of Ω, we can thus assume that the points x, y in Theorem 1.2 are

$a := (-\ell, 0)$ and $b := (\ell, 0)$ and that the rectangle $[-\ell, \ell] \times [0,\ \delta]$ is contained in Ω.

Up to subtracting an affine function, we can also assume that ψ satisfies

$$\psi(-\ell, 0) = \psi(\ell, 0) = 0 \quad \text{and} \quad 0 \in \partial\psi([a, b]). \tag{24}$$

Throughout the proofs, $x = (x_\parallel, x_\perp)$ or $y = (y_\parallel, y_\perp)$ will denote points in $\Omega \subset \mathbb{R}^2$ with $x_\parallel,\ y_\parallel$ the first coordinate parallel to the segment $[a, b]$. Similarly $z = (z_\parallel, z_\perp)$ will denote a point in $\partial\psi(\Omega) \subset \mathbb{R}^2$.

We will also use the following notation:

$$R_\delta = \{(x_\parallel,\ x_\perp) \,|\, |x_\parallel| \le \ell/2,\ 0 \le x_\perp \le \delta\}. \tag{25}$$

Furthermore, (24) implies that $0 \in \partial\psi(U_\delta)$ and so for any $K \ge \operatorname{diam} \partial\psi(U_\delta)$ we have

$$K \ge \|\partial\psi\|_{L^\infty(R_\delta)} = \sup_{y \in R_\delta} \sup_{z \in \partial\psi(y)} |z|. \tag{26}$$

We also note that

$$\varepsilon := -\min_{t \in [0,1]} \psi(ta + (1-t)b) \ge 0. \tag{27}$$

Throughout the proofs, C denotes a numerical constant, which depends only on the dimension $d = 2$ and whose value may change from line to line in the calculations.

Before moving to the heart of the proof, we state the following simple lemma which we will use repeatedly,

Lemma 2.1. *Let $\psi : [-\ell, \ell] \times [0, 2\delta] \to \mathbb{R}$ be a convex function satisfying* (24) *and* (27)*. Then for all $y \in \Omega$ such that $|y_\parallel| \le \ell/2$ we have*

$$|z_\parallel| \le \frac{2}{\ell}\,(K\,|y_\perp| + \varepsilon)\,, \qquad \forall z \in \partial\psi(y).$$

Proof. Consider any $y \in \Omega$ with $|y_\parallel| \le \ell/2$, $0 \le y_\perp \le 2\,\delta$ and any $z \in \partial\psi(y)$. Then we have by the definition of subdifferential

$$\psi(b) \ge \psi(y) + z \cdot (b - y) = \psi(y) + z_\parallel \cdot (b_\parallel - y_\parallel) + z_\perp \cdot (b_\perp - y_\perp).$$

Since $b_\parallel - y_\parallel = \ell - y_\parallel \ge \ell/2$, and $a_\perp = 0$, this lets us deduce that:

$$\begin{aligned} |z_\parallel| &\le \frac{1}{b_\parallel - y_\parallel}\,[z_\perp \cdot (y_\perp - a_\perp) + (\psi(b) - \psi(y))] \\ &\le \frac{2}{\ell}\,[z_\perp \cdot y_\perp + (\psi(b) - \psi(y))] \\ &\le \frac{2}{\ell}\,(K\,|y_\perp| + \varepsilon)\,, \end{aligned}$$

where we have used (24) so $\psi(b) = 0$, (27) so $\psi(y) \ge -\varepsilon$ and the fact that $|z_\perp| \le K$ (by (26)). This completes the proof of Lemma 2.1. □

We conclude these preliminaries by noting that the quantity $\operatorname{diam} \partial\psi(U_\delta)$ *a priori* depends on ℓ. We obviously have

$$\operatorname{diam} \partial\psi(U_\delta) \leq \operatorname{diam} \partial\psi(\Omega) \tag{28}$$

and we can show the following lower bound:

Lemma 2.2. *If* $h_1 \leq \min\{\delta, \ell\}$ *and* $h_2^2 < \frac{\delta\ell}{\lambda_1\lambda_2}$, *then*

$$\operatorname{diam} \partial\psi(U_\delta) \geq \left(\frac{\delta\ell}{\lambda_1\lambda_2}\right)^{1/2}. \tag{29}$$

Proof. Inequality (1) gives

$$\mu(U_\delta) \leq \nu(\partial\psi(U_\delta)).$$

Since the dimensions of U_δ satisfy $\min\{\delta, \ell\} \geq h_1$, Assumption 1 implies

$$\mu(U_\delta) \geq \frac{\ell\delta}{\lambda_1}, \quad \text{and} \quad \nu(\partial\psi(U_\delta)) \leq \lambda_2 \max\{(\operatorname{diam} \partial\psi(U_\delta))^2, h_2^2\}.$$

We deduce

$$\frac{\ell\delta}{\lambda_1\lambda_2} \leq \max\{(\operatorname{diam} \partial\psi(U_\delta))^2, h_2^2\}$$

and the condition $h_2^2 < \frac{\delta\ell}{\lambda_1\lambda_2}$ implies (29). □

2.2 Proof of Theorem 1.2

We now describe our strategy for proving Theorem 1.2. First, we note that since ψ is a convex function in Ω, it is differentiable in a subset $\widetilde{\Omega} \subset \Omega$ of full measure ($|\Omega \setminus \widetilde{\Omega}| = 0$), see for instance [19].

We can thus define a map $T : \Omega \mapsto \mathbb{R}^2$ which satisfies

$$T(x) := \nabla\psi(x) \qquad \forall x \in \widetilde{\Omega}. \tag{30}$$

Our proof of Theorem 1.2 relies on some careful estimates (upper and lower bounds) of the integral quantity

$$\int_{R_\delta \times R_\delta} |T_\perp(y) - T_\perp(x)|\, \varphi(x, y)\, dy\, dx \tag{31}$$

where the weight function $\varphi(x, y)$ is given by

$$\varphi(x, y) = \frac{1}{(x_\perp + \gamma)^2} \mathbf{1}_{\{\frac{1}{2}x_\perp \leq y_\perp \leq 2\, x_\perp\}}, \tag{32}$$

for some $\gamma > 0$. The exponent 2 is chosen to obtain the right logarithmic divergence in the estimates.

Using the notations from Theorem 1.2, we will first prove the following upper bound which does not require (1):

Proposition 2.3. *Assume that $\psi : [-\ell, \ell] \times [0, 2\delta] \to \mathbb{R}$ is a convex function satisfying* (24)*. Then there exists a universal constant $C > 0$ s.t. the following inequality holds for all $\gamma \geq \varepsilon/K$*

$$\int_{R_\delta \times R_\delta} |T_\perp(y) - T_\perp(x)| \, \varphi(x,y) \, dy \, dx \leq C \, K \, \ell^2 \left(\left[\log\left(1 + \frac{\delta}{\gamma}\right) \right]^{1/2} + 1 \right), \tag{33}$$

where we recall that K and ε satisfy (26) *and* (27)*.*

The proof of this upper bound is fairly straightforward (see Section 2.3) and only makes use of the convexity of ψ and Lemma 2.1.

Next, we will prove the following lower bound for (31):

Proposition 2.4. *Let ψ be a convex function satisfying* (1) *for some measure μ and ν satisfying Assumption 1. Assume further than ψ satisfies* (24)*. There exists a universal constant C s.t. assuming that ℓ satisfies* (9)*, which we recall is*

$$\ell \geq 2\, h_1, \quad \ell^2 \geq C \, K \, \lambda_1 \, \lambda_2 \, h_2,$$

and defining

$$\gamma := \max\left(\frac{\varepsilon}{K}, \, 2\, h_1, \, \frac{\ell h_2}{C\, K} \right), \tag{34}$$

then the following inequality holds

$$\begin{aligned} &\int_{R_\delta \times R_\delta} |T_\perp(y) - T_\perp(x)| \, \varphi(x,y) \, dy \, dx \\ &\qquad \geq \frac{\ell^4}{C \, \lambda_1 \, \lambda_2 \, K} \left(1 \wedge \frac{\ell^2}{\lambda_1 \, \lambda_2 \, K^2} \right) \log\left(\frac{1}{2} + \frac{\delta}{2\,\gamma} \right), \end{aligned} \tag{35}$$

provided $\gamma < \delta$ and where we recall the notation $a \wedge b = \min(a, b)$.

The proof of this proposition, which is presented in Section 2.4, is more delicate. This is where we use the fact that ψ satisfy the Monge-Ampère like condition (1) with measures μ and ν satisfying (7).

Proof of Theorem 1.2. The key to conclude the proof of Theorem 1.2 is that the bounds provided by Propositions 2.3 and 2.4 scale differently in ℓ and γ. Combining the two will hence naturally lead either to an upper bound on ℓ or to a lower bound on γ. More precisely we directly obtain from (33) and (35) that

$$\frac{\ell^2}{\lambda_1 \, \lambda_2 \, K^2} \left(1 \wedge \frac{\ell^2}{\lambda_1 \, \lambda_2 \, K^2} \right) \log\left(\frac{1}{2} + \frac{\delta}{2\,\gamma} \right) \leq C \left(\left[\log\left(1 + \frac{\delta}{\gamma}\right) \right]^{1/2} + 1 \right).$$

Since we assumed in the theorem that $\delta \geq 2\,\gamma$, we have $\log\left(1 + \frac{\delta}{\gamma}\right) \leq C \log\left(\frac{1}{2} + \frac{\delta}{2\,\gamma}\right)$ so that we can simplify the inequality above to:

$$\frac{\ell^2}{\lambda_1 \, \lambda_2 \, K^2} \left(1 \wedge \frac{\ell^2}{\lambda_1 \, \lambda_2 \, K^2} \right) \left[\log\left(1 + \frac{\delta}{\gamma}\right) \right]^{1/2} \leq C \left(1 + \left[\log\left(1 + \frac{\delta}{\gamma}\right) \right]^{-1/2} \right) \leq C. \tag{36}$$

Moreover we also get $\log\left(1+\frac{\delta}{\gamma}\right) \geq \log 3$ (still using the assumption that $\delta \geq 2\,\gamma$) so (36) gives

$$\frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\left(1 \wedge \frac{\ell^2}{\lambda_1\lambda_2 K^2}\right) \leq C[\log 3]^{-1/2},$$

which can be used to show that

$$\left(\frac{\ell^2}{\lambda_1\lambda_2 K^2} \wedge \left(\frac{\ell^2}{\lambda_1\lambda_2 K^2}\right)^2\right) \geq C\left(\frac{\ell^2}{\lambda_1\lambda_2 K^2}\right)^2.$$

for some (different) constant C. Together with (36), this finally implies

$$\left(\frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right)^2\left[\log\left(1+\frac{\delta}{\gamma}\right)\right]^{1/2} \leq C$$

which completes the proof of Theorem 1.2. □

2.3 Upper bound: Proof of Proposition 2.3

Proof of Proposition 2.3. We first assume that ψ is C^2 so that all the computations below make sense. We can write

$$\begin{aligned}
&\int_{R_\delta\times R_\delta} |T_\perp(y)-T_\perp(x)|\,\varphi(x,y)\,dy\,dx \\
&\qquad = \int_{R_\delta\times R_\delta}\left|\int_0^1 \nabla T_\perp(x+t(y-x))\cdot(y-x)\,dt\right|\varphi(x,y)\,dy\,dx \\
&\qquad \leq \int_{R_\delta\times R_\delta}\int_0^1 \left|\partial_\parallel T_\perp(x+t(y-x))\cdot(y_\parallel-x_\parallel)\right|\,dt\,\varphi(x,y)\,dy\,dx \\
&\qquad\quad + \int_{R_\delta\times R_\delta}\int_0^1 |\partial_\perp T_\perp(x+t(y-x))\cdot(y_\perp-x_\perp)|\,dt\,\varphi(x,y)\,dy\,dx,
\end{aligned}$$

where $\partial_\parallel$ denotes the derivative with respect to the first component and $\partial_\perp$ is the derivative in the orthogonal direction. Using the symmetry of the expression in x and y, we have

$$\begin{aligned}
&\int_{R_\delta\times R_\delta} |T_\perp(y)-T_\perp(x)|\varphi(x,y)\,dy\,dx \\
&\leq 2\int_{R_\delta\times R_\delta}\int_{1/2}^1 \left|\partial_\parallel T_\perp(x+t(y-x))\cdot(y_\parallel-x_\parallel)\right|\,dt\,\varphi(x,y)\,dy\,dx \\
&\qquad + 2\int_{R_\delta\times R_\delta}\int_{1/2}^1 |\partial_\perp T_\perp(x+t(y-x))\cdot(y_\perp-x_\perp)|\,dt\,\varphi(x,y)\,dy\,dx, \qquad (37)
\end{aligned}$$

To bound the first term in the right-hand side, we note that by definition of R_δ, $|y_\| - x_\|| \le \ell$ so that using the change of variable $y \to z = x + t(y-x)$

$$\begin{aligned}
&\int_{R_\delta \times R_\delta} \int_{1/2}^1 \left|\partial_\| T_\perp(x + t(y-x)) \cdot (y_\| - x_\|)\right| \, \varphi(x,y)\, dt\, dy\, dx \\
&\qquad \le \ell \int_{R_\delta} \int_{1/2}^1 \int_{R_\delta} \left|\partial_\| T_\perp(x + t(y-x))\right| \, \varphi(x,y)\, dy\, dt\, dx \\
&\qquad \le \ell \int_{R_\delta} \int_{1/2}^1 \int_{R_\delta} \left|\partial_\| T_\perp(z)\right| \, \mathbf{1}_{x + \frac{z-x}{t} \in R_\delta} \varphi\left(x, x + \frac{z-x}{t}\right) \frac{1}{t^d}\, dz\, dt\, dx \\
&\qquad \le \ell \int_{R_\delta} \left|\partial_\| T_\perp(z)\right| J_1(z)\, dz. \qquad (38)
\end{aligned}$$

Using the definition of $\varphi(x,y)$ (see (32)) and the notation

$$\Omega_{x_\perp} = \left\{ y \in [\ell/2, \ell/2] \times [0,\delta];\ \frac{1}{2}\, x_\perp \le y_\perp \le 2\, x_\perp \right\},$$

we get that the weight $J_1(z)$ is equal to

$$\begin{aligned}
J_1(z) &= 2 \int_{1/2}^1 \int_{R_\delta} \mathbf{1}_{x + \frac{z-x}{t} \in R_\delta} \varphi\left(x, x + \frac{z-x}{t}\right) dx\, dt \\
&= 2 \int_{1/2}^1 \int_{R_\delta} \frac{1}{(|x_\perp| + \gamma)^2} \mathbf{1}_{x + \frac{z-x}{t} \in \Omega_{x_\perp}}\, dx\, dt.
\end{aligned}$$

Observe that the definition of $\Omega_{x_\perp}$ is actually symmetric on R_δ: $y \in \Omega_{x_\perp}$ iff $x \in \Omega_{y_\perp}$ since $x_\perp \ge 0$. Consequently $z \in \Omega_{x_\perp}$ implies that $x \in \Omega_{z_\perp}$ as $x \in R_\delta$ and

$$\begin{aligned}
J_1(z) &\le \frac{C}{(|z_\perp| + \gamma)^2} \int_{1/2}^1 \int_{R_\delta} \mathbf{1}_{x + \frac{z-x}{t} \in \Omega_{x_\perp}}\, dx\, dt \le \frac{C}{(|z_\perp| + \gamma)^2} \int_{\Omega_{z_\perp}} dx \\
&\le \frac{C}{(|z_\perp| + \gamma)^2}\, \ell\, |z_\perp| \le \frac{C\,\ell}{(|z_\perp| + \gamma)}.
\end{aligned}$$

Going back to (38), we find

$$\int_{R_\delta \times R_\delta} \int_{1/2}^1 \left|\partial_\| T_\perp(x + t(y-x)) \cdot (y_\| - x_\|)\right| \, \varphi(x,y) dt\, dy\, dx \le C\, \ell^2 \int_{R_\delta} \frac{\left|\partial_\| T_\perp(z)\right|}{(|z_\perp| + \gamma)}\, dz. \qquad (39)$$

Next, we note that the convexity of ψ implies that the matrix

$$\begin{bmatrix} \partial_\| T_\| & \partial_\perp T_\| \\ \partial_\| T_\perp & \partial_\perp T_\perp \end{bmatrix}$$

is symmetric and non-negative with a non-negative determinant: $\partial_\| T_\|(z) \partial_\perp T_\perp(z) - \partial_\| T_\perp \partial_\perp T_\| \ge 0$, which implies that $|\partial_\| T_\perp(z)| \le |\partial_\| T_\|(z)|^{1/2}\, |\partial_\perp T_\perp(z)|^{1/2}$. This lets us deduce that

$$\begin{aligned}
\int_{R_\delta} \frac{|\partial_\| T_\perp(z)|}{|z_d| + \gamma} dz &\le \int_{R_\delta} \frac{|\partial_\| T_\||^{1/2}}{(|z_\perp| + \gamma)} |\partial_\perp T_\perp(z)|^{1/2}\, dz \\
&\le \left[\int_{R_\delta} \frac{|\partial_\| T_\|(z)|}{(|z_\perp| + \gamma)^2} dz\right]^{1/2} \left[\int_{R_\delta} |\partial_\perp T_\perp(z)|\, dz\right]^{1/2}. \qquad (40)
\end{aligned}$$

Using the fact that $\partial_\parallel T_\parallel \geq 0$ from the convexity of ψ,

$$\int_{R_\delta} \frac{|\partial_\parallel T_\parallel(z)|}{(z_\perp+\gamma)^2} dz_\perp = \int_0^\delta \frac{T_\parallel(\ell/2, z_\perp) - T_\parallel(-\ell/2, z_\perp)}{(z_\perp+\gamma)^2} dz_\perp \leq \frac{C}{\ell} \int_0^\delta \frac{(Kz_\perp+\varepsilon)}{(z_\perp+\gamma)^2} dz_\perp \tag{41}$$

$$\leq \frac{CK}{\ell} \int_0^\delta \frac{1}{(z_\perp+\gamma)} dz_\perp = CK\ell^{-1} \int_0^\delta \frac{1}{z_\perp+\gamma} dz_\perp, \tag{42}$$

by using Lemma 2.1 and the fact that $\gamma \geq \varepsilon/K$.

Similarly, we have that $\partial_\perp T_\perp \geq 0$ so that

$$\int_{R_\delta} |\partial_\perp T_\perp(z)| dz \leq \int_{-\ell/2}^{\ell/2} [T_\perp(z_\parallel, \delta) - T_\perp(z_\parallel, 0)] dz_\parallel \leq 2K\ell. \tag{43}$$

Combining (42) and (43) into (40) and inserting the result into (39), we conclude that

$$\int_{R_\delta \times R_\delta} \int_{1/2}^1 \left|\partial_\parallel T_\perp(x+t(y-x))(y_1-x_1)\right| \varphi(x,y) dt\, dy\, dx \leq C\, \ell^2\, K \left[\int_0^{2\delta} \frac{1}{z_\perp+\gamma} dz_\perp\right]^{1/2}. \tag{44}$$

which gives a bound for the first term in the right hand side of (37).

We now proceed similarly to bound the second term in the right-hand side of (37). First we write, recalling that $\partial_\perp T_\perp \geq 0$,

$$\int_{R_\delta \times R_\delta} \int_{1/2}^1 \left|\partial_\perp T_\perp(x+t(y-x))\,(y_\perp - x_\perp)\right| \, dt\, \varphi(x,y)\, dy\, dx$$

$$\leq \int_{R_\delta} \int_{1/2}^1 \int_{R_\delta} \partial_\perp T_\perp(x+t(y-x))\, |y_\perp - x_\perp|\, \varphi(x,y)\, dy\, dt\, dx.$$

Note that the definition of φ in (32) implies that

$$|y_\perp - x_\perp|\, \varphi(x,y) = \frac{|y_\perp - x_\perp|}{(x_\perp+\gamma)^2} \mathbf{1}_{\{\frac{1}{2}x_\perp \leq y_\perp \leq 2x_\perp\}}$$

$$\leq \frac{1}{x_\perp+\gamma} \mathbf{1}_{\{\frac{1}{2}x_\perp \leq y_\perp \leq 2x_\perp\}}.$$

We perform the same change of variable $z = x + t(y-x)$ as we used after (38)) to find that

$$\int_{R_\delta \times R_\delta} \int_{1/2}^1 \left|\partial_\perp T_\perp(x+t(y-x))\,(y_\perp - x_\perp)\right| \, dt\, \varphi(x,y)\, dy\, dx$$

$$\leq \int_{R_\delta} \int_{1/2}^1 \int_{R_\delta} \partial_\perp T_\perp(z)\, \frac{1}{x_\perp+\gamma}\, \mathbf{1}_{x+\frac{z-x}{t} \in \Omega_{x_\perp}}\, dz\, dt\, dx$$

$$\leq \int_{R_\delta} \partial_\perp T_\perp(z)\, J_2(z)\, dz, \tag{45}$$

with

$$J_2(z) = 2 \int_{1/2}^1 \int_{R_\delta} \frac{1}{x_\perp+\gamma} \mathbf{1}_{x+\frac{z-x}{t} \in \Omega_{x_\perp}}\, dx\, dt.$$

Proceeding as with the weight $J_1(z)$ above (the only difference lies in the power of $(x_\perp + \gamma)$), we find that $\mathbf{1}_{x+\frac{z-x}{t}\in\Omega_{x_\perp}} \leq \mathbf{1}_{x\in\Omega_{z_\perp}}$

$$J_2(z) \leq \frac{C}{z_\perp + \gamma} |\Omega_{z_\perp}| \leq C\,\ell.$$

Inserting this bound in (45), we obtain

$$\begin{aligned}\int_{R_\delta\times R_\delta} \int_{1/2}^1 |\partial_\perp T_\perp(x + t(y-x))\,(y_\perp - x_\perp)|\; dt\,\varphi(x,y)\,dy\,dx \\ \leq C\,\ell \int_{R_\delta} \partial_\perp T_\perp(z)\,dz = C\,\ell \int_{-\ell/2}^{\ell/2} (T_\perp(z_\parallel,\delta) - T_\perp(z_\parallel,0))\,dz_\parallel \leq C\,K\,\ell^2. \end{aligned} \tag{46}$$

Combining (46) and (44) in (37), we obtain that

$$\int_{R_\delta\times R_{2\delta}} |T_\perp(x) - T_\perp(y)|\,\varphi(x,y)\,dy\,dx \leq C\,K\,\ell^2 \left(\left[\log\left(1+\frac{\delta}{\gamma}\right)\right]^{1/2} + 1\right), \tag{47}$$

which proves the proposition if T is C^1 and hence ψ is C^2.

When ψ is only convex but not C^2, we naturally introduce the convex function $\psi^\eta = \psi \star_x \rho_\eta$, where ρ_η is a standard mollifier. We may then apply (47) to ψ^η and find for $T^\eta = \nabla\psi^\eta$

$$\int_{R_{2\delta}\times R_{2\delta}} |T^\eta_\perp(x) - T^\eta_\perp(y)|\,\varphi(x,y)\,dy\,dx \leq C\,K_\eta\,\ell^2 \left(\left[\log\left(1+\frac{\delta}{\gamma}\right)\right]^{1/2} + 1\right),$$

where we observe that, in this case, since we only integrate over R_δ, K_η is given by

$$K_\eta = \sup_{R_\delta} |\nabla\psi_\eta| \leq \|\partial\psi\|_{L^\infty(R_\delta)} \leq K,$$

for $\eta < \delta$. At the same time, since ψ is convex then $T = \nabla\psi$ belongs to $BV(R_\delta)$ and therefore $\|T^\eta - T\|_{L^1(R_\delta)} \to 0$ as $\eta \to 0$. Since φ is bounded for any fixed $\gamma > 0$, we may directly pass to the limit $\eta \to 0$ and obtain (47) on T. □

2.4 Lower bound: Proof of Proposition 2.4

We now turn to the proof of the lower bound (35). Given $x_\perp \in (0,\delta)$, we recall for convenience the definition of the set $\Omega_{x_\perp}$, the following sets

$$\Omega_{x_\perp} = \left\{ y \in [-\ell/2,\ell/2]\times[0,\delta]\,;\;\; \frac{1}{2}\,x_\perp \leq y_\perp \leq 2\,x_\perp \right\},$$

together with the more restricted set

$$\Lambda_{x_\perp} = \left\{ y \in [-\ell/4,\ell/4]\times[0,\delta]\,;\;\; x_\perp \leq y_\perp \leq \frac{3}{2}\,x_\perp \right\}.$$

Since we are trying to show that $T_\perp(y)$ cannot be concentrated, instead of looking at $|T_\perp(y) - T_\perp(x)|$, we define, for $\xi \in \mathbb{R}$ and $\eta > 0$, the more general set

$$\Omega_{x_\perp,\eta} = \{ y \in \Omega_{x_\perp}\,;\, |z_\perp - \xi| > \eta \text{ for all } z \in \partial\psi(y) \}. \tag{48}$$

Our first task, in Lemma 2.5 below, is to show that for an appropriate value of η and for all $\xi \in \mathbb{R}$, the set $\Omega_{x_\perp,\eta}$ is non empty, and more precisely $\Lambda_{x_\perp} \cap \Omega_{x_\perp,\eta} \neq \emptyset$. This will allow us to construct a half-cone within $\Omega_{x_\perp,\eta}$ in Lemma 2.6 and finally to obtain a lower bound for $|\Omega_{x_\perp,\eta}|$ in Lemma 2.7. This will finally let us conclude the proof of Prop. 2.4 and obtain the lower bound (35).

2.4.1 Non-emptyness of the set $\Omega_{x_\perp,\eta}$

First we have the following lemma, which implies in particular that the set $\Omega_{x_\perp,\eta}(\xi)$ is not empty.

Lemma 2.5. *Let ψ be a convex function satisfying (1) for some measure μ and ν satisfying Assumption 1 and let K satisfy (8). There exists a universal constant C such that defining*

$$\eta := \frac{1}{C\,\lambda_1\,\lambda_2}\frac{\ell^2}{K}, \tag{49}$$

and assuming furthermore that ℓ satisfies

$$\ell \geq 2\,h_1, \quad \ell^2 \geq C\,K\,\lambda_1\,\lambda_2\,h_2,$$

then for all $x_\perp > \gamma = \max(\frac{\varepsilon}{K},\, 2h_1,\, \frac{\ell h_2}{C\,K})$ and for all $\xi \in \mathbb{R}$, there is at least one point $y^* \in \Lambda_{x_\perp}$ such that for some $z \in \partial\psi(y^*)$ we have $|z_\perp - \xi| \geq 3\eta$.

The idea of the proof is to look at the image of the set $\Lambda_{x_\perp}$ by the subdifferential $\partial\psi$. By Lemma 2.1 this image is bounded in the horizontal (i.e. $z_\parallel$) directions. However, (1) together with Assumption 1 gives a lower bound on the measure of this image, which is where the fact that ψ is the Kantorovich potential for an optimal transportation problem is crucial. Therefore the image cannot be too small in the vertical (i.e. $z_\perp$) directions, which is essentially the statement of Lemma 2.5. The lower bounds on ℓ and $x_\perp$ in Lemma 2.5 are necessary so that we can use (7) on the measures μ and ν.

Proof of Lemma 2.5. We start by noticing that for all $z \in \partial\psi(\Lambda_{x_\perp})$, Lemma 2.1 implies that

$$|z_\parallel| \leq \frac{2}{\ell}\left(K\,\frac{3}{2}\,x_\perp + \varepsilon\right).$$

This leads to defining the rectangle

$$\tilde{R} = \left\{ z \in \mathbb{R}^2\,;\, |z_\parallel| \leq \max\left(\frac{2\,K}{\ell}\,\left(\frac{3}{2}\,x_\perp + \frac{\varepsilon}{K}\right), h_2\right) \text{ and } |z_\perp - \xi| \leq \max(3\,\eta, h_2)\right\}.$$

We show the existence of y^* as in Lemma 2.5 by contradiction: Suppose there is no such point y^*, then we must have that $\partial\psi(\Lambda_{x_\perp}) \subset \tilde{R}$ and inequality (1) gives

$$\mu(\Lambda_{x_\perp}) \leq \nu(\partial\psi(\Lambda_{x_\perp})) \leq \nu(\tilde{R} \cap \partial\psi(\Omega)). \tag{50}$$

We now want to use Assumption 1 to estimate the left and right hand side of (50). The rectangle $\Lambda_{x_\perp}$ has size $\left(\frac{\ell}{2}\right) \times \frac{x_\perp}{2}$. Since $\ell \geq 2\,h_1$ and $x_\perp \geq \gamma \geq 2h_1$, the rectangle $\Lambda_{x_\perp}$ has size at least h_1 in all directions and Assumption 1 (see (7)) implies that $\mu(\Lambda_{x_\perp}) \geq |\Lambda_{x_\perp}|/\lambda_1$.

Similarly, the definition of the set $\tilde{R}$ guarantees that $\tilde{R}$ has size at least h_2 in all directions and so $\nu(\tilde{R} \cap \partial\psi(\Omega)) \leq \lambda_2|\tilde{R}|$. Equation (50) thus yields

$$|\Lambda_{x_\perp}| \leq \lambda_1\,\lambda_2\,|\tilde{R}|. \tag{51}$$

We now note that

$$|\Lambda_{x_\perp}| = C\,\ell\,x_\perp, \tag{52}$$

while the assumption $x_\perp > \gamma$ with $\gamma \geq \frac{\varepsilon}{K}$ and $\gamma \geq \ell\, h_2/C\,K$ implies

$$\begin{aligned} |\tilde{R}| &= C\,\max\left(\frac{2K}{\ell}\left(\frac{3}{2}\,x_\perp + \frac{\varepsilon}{K}\right),\, h_2\right)\,(\max(3\,\eta, h_2)) \\ &\leq C\,\left(\frac{K}{\ell}\,x_\perp\right)\,(\max(3\,\eta, h_2))\,. \end{aligned} \tag{53}$$

Together with the definition of η this shows that

$$|\tilde{R}| \leq C\left(\frac{K}{\ell}x_\perp\right)\max\left(\frac{1}{C\lambda_1\lambda_2}\frac{\ell^2}{K}, h_2\right) \tag{54}$$

Equation (51) then proves that

$$C\ell x_\perp \leq \frac{\ell x_\perp}{C} \tag{55}$$

which is a contraction by taking C large enough and concludes the second part of the proof. □

2.4.2 Lower Bound on $|\Omega_{x_\perp,\eta}|$

We now show that the measure $|\Omega_{x_\perp,\eta}|$ is bounded from below. We first need, as an intermediary result, the following lemma which only relies on the convexity of the function ψ. This lemma mostly states that if the subdifferentials corresponding to two points y' and y'' is concentrated in the vertical direction and the segment $[y',\ y'']$ is almost vertical, then the subdifferential corresponding to any point in that segment also has to be concentrated.

We will later use this lemma together with Lemma 2.5 to obtain contradictions and ensures the absence of concentration in the subdifferential over half a cone.

Lemma 2.6. *Let ψ satisfy (24), consider any $x_\perp \geq \gamma = \max\left(\frac{\varepsilon}{K},\, 2\,h_1,\, \frac{\ell\,h_2}{C\,K}\right)$ and fix any $\xi \in \mathbb{R}$. Assume that $y',\ y'' \in \Omega_{x_\perp}$ are such that $y' \neq y''$ and*

$$\partial\psi(y') \cap \{z \in \Omega_2\,;\, |z_\perp - \xi| \leq \eta\} \neq \emptyset, \qquad \partial\psi(y'') \cap \{z \in \Omega_2\,;\, |z_\perp - \xi| \leq \eta\} \neq \emptyset.$$

There exists a universal constant C s.t., if

$$|\tan((y',\ y''),\ e_\perp)| \leq \frac{\ell\,\eta}{C\,K x_\perp},$$

where $(y',\ y''),\ e_\perp)$ is the angle between the vertical direction $e_\perp$ and the segment $[y',\ y'']$, then for all $y = s\,y' + (1-s)\,y''$ with $s \in (0,1)$ we have

$$\partial\psi(y) \subset \{z\,;\, |z_\perp - \xi| \leq 2\,\eta\}.$$

Proof. Take $z' \in \partial\psi(y') \cap \{|z_\perp - \xi| \leq \eta\}$ and $y = s\,y' + (1-s)\,y''$ for some fixed $s \in (0,1)$. We can assume (without loss of generality) that $y'_\perp - y_\perp > 0$ and $y''_\perp - y_\perp < 0$. For any $z \in \partial\psi(y)$, the convexity of ψ implies (cyclical monotonicity of the sub-differential):

$$(z' - z)\cdot(y' - y) \geq 0,$$

and therefore

$$(z'_\perp - z_\perp)\,(y'_\perp - y_\perp) \geq -(z'_\parallel - z_\parallel)(y'_\parallel - y_\parallel).$$

We hence deduce that

$$z_\perp \leq z'_\perp + \frac{(z'_\parallel - z_\parallel)(y'_\parallel - y_\parallel)}{y'_\perp - y_\perp} \leq \xi + \eta + (|z'_\parallel| + |z_\parallel|)\,\frac{|y'_\parallel - y_\parallel|}{y'_\perp - y_\perp},$$

since $|\overline{z}_\perp - \xi| \leq \eta$. Using now Lemma 2.1, we then get that

$$\begin{aligned} z_\perp &\leq \xi + \eta + \left[\frac{2}{\ell}\,(K\,|y'_\perp| + \varepsilon) + \frac{2}{\ell}\,(K\,|y_\perp| + \varepsilon)\right]\,\frac{|y' - y|}{y'_\perp - y_\perp} \\ &\leq \xi + \eta + \frac{C}{\ell}\,K\,x_\perp\,|\tan((y'', y'),\,e_\perp)| \leq \xi + 2\eta, \end{aligned}$$

by the definition of the tangent and where we used the fact that $x_\perp \geq \gamma \geq \varepsilon/K$, that y', $y'' \in \Omega_{x_\perp}$ so $y \in \Omega_{x_\perp}$ and as a consequence $y'_\perp,\ y_\perp \leq 2\,x_\perp$.

Proceeding similarly using y'' instead of y', we can get the inequality $z_\perp \geq \xi - 2\,\eta$ and the result follows. □

Using Lemmas 2.5 and 2.6, we can now get a lower bound on the measure of the set $\Omega_{x_\perp,\eta}(\xi)$ (which we recall is defined by (48)). This will be the key estimate in the proof of Proposition 2.4.

Lemma 2.7. *Let ψ be a convex function satisfying* (1) *for some measure μ and ν satisfying Assumption 1. Assume further that ψ satisfies* (24)*. Recall that K satisfies* (8) *and that η is defined by* (49)*. Assume furthermore that ℓ satisfies, for an appropriate universal constant C,*

$$\ell \geq 2h_1, \quad \ell^2 \geq C\,K\,\lambda_1\,\lambda_2\,h_2.$$

Then, for all $x_\perp > \gamma = \max\left(\frac{\varepsilon}{K},\ 2\,h_1,\ \frac{\ell h_2}{C\,K}\right)$, and for all $\xi \in \mathbb{R}$, one has the lower bound

$$|\Omega_{x_\perp,\eta}(\xi)| \geq \frac{\ell\,x_\perp}{C}\,\min\left(1,\ \frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right). \tag{56}$$

Proof. Start by using Lemma 2.5 to obtain the existence of one $\tilde{y} \in \Lambda_{x_\perp}$ be such that for some $\tilde{z} \in \partial\psi(y)$ we have $|\tilde{z}_\perp - \xi| \geq 3\,\eta$. Define now C_θ as the cone (see figure 1) with vertex $\tilde{y}$ and angle θ with the vertical direction $e_\perp$, such that

$$\tan\theta = \min\left(\frac{\ell}{2x_\perp},\ \frac{\ell\,\eta}{C\,K x_\perp}\right).$$

Define furthermore the truncated cone $S_\theta = \{y \in C_\theta \,|\, |y_\perp - \tilde{y}_\perp| \leq x_\perp/2\}$.

We first observe that $S_\theta \subset \Omega_{x_\perp}$ as for any $y \in S_\theta$, we have first $\tilde{y}_\perp - x_\perp/2 \leq y_\perp \leq \tilde{y}_\perp + x_\perp/2$ and hence $\frac{x_\perp}{2} \leq y_\perp \leq 2\,x_\perp$ since $x_\perp \leq \tilde{y}_\perp \leq \frac{3}{2}\,x_\perp$. Second, since $|\tilde{y}_\parallel| \leq \ell/4$, we have that

$$|y_\parallel| \leq |\tilde{y}_\parallel| + |\tan\theta|\,|y_\perp - \tilde{y}_\perp| \leq \frac{\ell}{4} + |\tan\theta|\,\frac{x_\perp}{2} \leq \frac{\ell}{2},$$

which is the reason for the condition $\tan\theta \leq \ell/(2\,x_\perp)$.

The next step is to use Lemma 2.6 to prove that $|\Omega_{x_\perp,\eta} \cap S_\theta| \geq |S_\theta|/2$. For this consider any segment in the truncated cone S_θ with hence angle $\theta' \leq \theta$ with $e_\perp$. Denote by $L^1_{\theta'}$ and $L^2_{\theta'}$ the two half-parts of the segment from $\tilde{y}$.

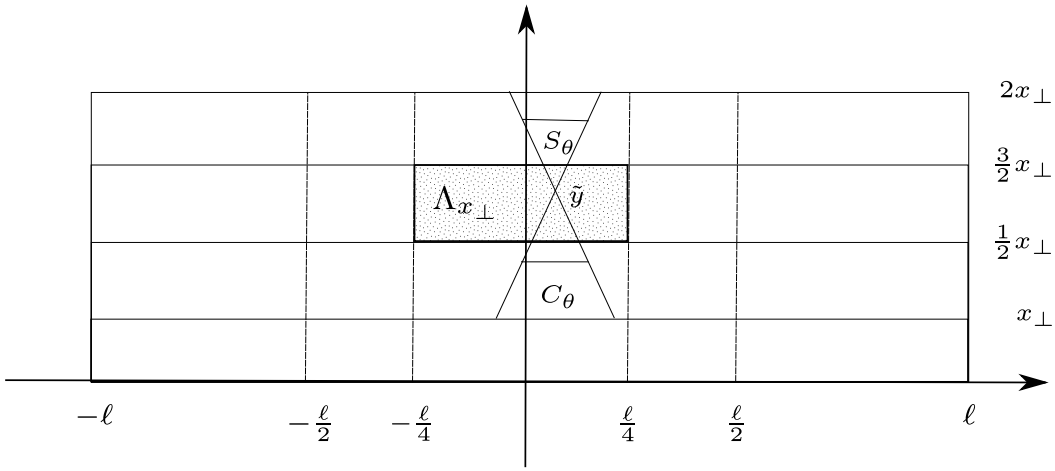

Figure 1: Cones C_θ and S_θ

We can show that either $L^1_{\theta'} \subset \Omega_{x_\perp,\eta}$ or $L^2_{\theta'} \subset \Omega_{x_\perp,\eta}$. Indeed by contradiction, if this was not the case we would have some $y' \in L^1_{\theta'} \setminus \Omega_{x_\perp,\eta}$, $y'' \in L^2_{\theta'} \setminus \Omega_{x_\perp,\eta}$. By the definition of $\Omega_{x_\perp,\eta}$, there exists $z' \in \partial\psi(y')$ with $|z'_\perp - \xi| \leq \eta$ and similarly $z'' \in \partial\psi(y'')$ with $|z''_\perp - \xi| \leq \eta$.

Of course by definition

$$\tan((y',\, y''),\, e_\perp) = \tan\theta' \leq \tan\theta \leq \frac{\ell\,\eta}{C\,K\,|x_\perp|},$$

and we can directly apply Lemma 2.6. As $\tilde{y}$ is a convex combination of y', y'', this implies that $\partial\psi(\tilde{y}) \subset \{z \,|\, |z_\perp - \xi| \leq 2\,\eta\}$ contradicting the fact that $\tilde{z} \in \partial\psi(\tilde{y})$ but $|\tilde{z}_\perp - \xi| \geq 3\,\eta$.

This proves as claimed that either $L^1_{\theta'} \subset \Omega_{x_\perp,\eta}$ or $L^2_{\theta'} \subset \Omega_{x_\perp,\eta}$ and integrating over all possible segments with all possible angles that $|\Omega_{x_\perp,\eta} \cap S_\theta| \geq |S_\theta|/2$.

To conclude the proof, it is hence enough to bound from below $|S_\theta|$,

$$|S_\theta| = C\,x_\perp\,(x_\perp\,\tan\theta) \geq \frac{1}{C}\,x_\perp\,\ell\,\left(\min\left(1,\,\frac{\eta}{K}\right)\right).$$

Using the definition of η in (49), we eventually obtain that

$$|S_\theta| \geq \frac{\ell\,x_\perp}{C}\,\min\left(1,\,\frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right).$$

□

2.4.3 Proof of Proposition 2.4

We now have all the estimates required to prove Proposition 2.4:

Proof of Proposition 2.4. Using the definition of φ as given in (32), and the set $\Omega_{x_\perp,\eta}(\xi)$ introduced in (48), we get

$$\int_{R_\delta\times R_\delta} |T_\perp(y)-T_\perp(x)|\,\varphi(x,y)\,dy\,dx = \int_{R_\delta} \frac{1}{(x_\perp+\gamma)^2}\int_{\Omega_{x_\perp}} |T_\perp(y)-T_\perp(x)|\,dy\,dx.$$

Now fix $\xi = T_\perp(x)$ and calculate

$$\int_{\Omega_{x_\perp}} |T_\perp(y)-T_\perp(x)|\,dy \geq \eta \int_{\Omega_{x_\perp,\eta}} dy = \eta\,|\Omega_{x_\perp,\eta}|.$$

Observe that the assumptions on ℓ and the definition of γ in Proposition 2.4 exactly coincide with Lemma 2.7. Hence we may apply the lemma whenever $x_\perp > \gamma$ to find

$$\begin{aligned}\int_{\Omega_{x_\perp}} |T_\perp(y)-T_\perp(x)|\,dy &\geq \frac{\ell\,x_\perp\,\eta}{C}\,\min\left(1,\ \frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right)\\ &= \frac{\ell^3\,x_\perp}{C\,\lambda_1\,\lambda_2\,K}\,\min\left(1,\ \frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right),\end{aligned}$$

by the definition of η. This leads to

$$\begin{aligned}\int_{R_\delta\times R_\delta} |T_\perp(y)-T_\perp(x)|\,\varphi(x,y)\,dy\,dx &\geq \int_{R_\delta\cap\{x_\perp\geq\gamma\}} |T_\perp(y)-T_\perp(x)|\,\varphi(x,y)\,dy\,dx\\ &\geq \frac{\ell^3}{C\,\lambda_1\,\lambda_2\,K}\,\min\left(1,\ \frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right)\int_{R_\delta\cap\{x_\perp\geq\gamma\}}\frac{dx_\perp}{(x_\perp+\gamma)}\\ &\geq \frac{\ell^4}{C\,\lambda_1\,\lambda_2\,K}\,\min\left(1,\ \frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right)\int_\gamma^\delta \frac{dr}{r+\gamma},\end{aligned}$$

and we may conclude that

$$\begin{aligned}&\int_{R_\delta\times R_\delta} |T_\perp(y)-T_\perp(x)|\,\varphi(x,y)\,dy\,dx\\ &\quad\geq \frac{\ell^4}{C\,\lambda_1\,\lambda_2\,K}\,\min\left(1,\ \frac{\ell^2}{\lambda_1\,\lambda_2\,K^2}\right)\log\left(\frac{1}{2}+\frac{\delta}{2\,\gamma}\right),\end{aligned}$$

which completes the proof. □

3 Proof of Corollary 1.3

Proof of Corollary 1.3. We now have that $\varepsilon = 0$ and so

$$\gamma = \max\left\{2h_1, \frac{\ell h_2}{CK}\right\}.$$

If the length ℓ does not satisfy (9) then we have

$$\text{either}\quad \ell < 2h_1, \quad \text{or}\quad \ell^2 < C\lambda_1\lambda_2\,K h_2, \tag{57}$$

which gives the first two terms in (13).

If ℓ satisfies (9), then we note that since $h_1 \leq \delta/4$, condition (12) implies $\gamma < \delta/2$ and so we can use Theorem 1.2 to find

$$\left(\frac{\ell^2}{C\lambda_1\lambda_2 K^2}\right)^4 \ln\left(\frac{\gamma+\delta}{\gamma}\right) \leq 1$$

Setting $u := \frac{\ell}{K\sqrt{C\lambda_1\lambda_2}}$, we rewrite this inequality as

$$u^8 \ln\left(\frac{\gamma(u)+\delta}{\gamma(u)}\right) \leq 1, \tag{58}$$

where $\gamma(u) = \max\left\{2h_1, \sqrt{\frac{\lambda_1\lambda_2}{C}}h_2 u\right\}$.

When $\gamma(u) = 2h_1 \leq \delta/2$, then (58) implies

$$u \leq \left[\ln\left(\frac{\delta}{2h_1}\right)\right]^{-1/8}. \tag{59}$$

When $\gamma(u) = \sqrt{\frac{\lambda_1\lambda_2}{C}}h_2 u$, the assumption $\frac{\sqrt{C\lambda_1\lambda_2}h_2}{\delta} \leq 1$ implies $\gamma(u) \leq \frac{\delta u}{C}$ and so the inequality (58) gives

$$u^8 \ln\left(1+\frac{C}{u}\right) \leq u^8 \ln\left(1+\frac{\delta}{\gamma(u)}\right) \leq 1.$$

Thus we obtain $u \leq C$ (since $C \geq 1$ and $u \mapsto u^8 \ln\left(1+\frac{C}{u}\right) \leq u^8$ is increasing). It follows that $\gamma(u) = \sqrt{\frac{\lambda_1\lambda_2}{C}}h_2 u \leq \sqrt{C\lambda_1\lambda_2}h_2$ and Inequality (58) then yields:

$$u^8 \leq \left[\ln\left(1+\frac{\delta}{\gamma(u)}\right)\right]^{-1} \leq \left[\ln\left(\frac{\delta}{\sqrt{C\lambda_1\lambda_2}h_2}\right)\right]^{-1}. \tag{60}$$

Inequalities (59) and (60) gives the last two terms in (13) and conclude the proof of this first corollary. □

4 Proof of corollary 1.4

Before proving Corollary 1.4, we state the following lemma which is proved at the end of this section.

Lemma 4.1. *Let ψ be a convex function on Ω and let $x, x' \in \Omega \times \Omega$. Denote $\ell = |x - x'|$ and*

$$\varepsilon = -\min_{t\in[0,1]} \psi((1-t)x + tx') - [(1-t)\psi(x) + t\psi(x')]. \tag{61}$$

Then, for any $z \in \partial\psi(x)$ and $z' \in \partial\psi(x')$, we have that

$$|z - z'| \geq 2\frac{\varepsilon}{\ell}. \tag{62}$$

Proof of Corollary 1.4. We recall that

$$x \in \partial\psi^*(z) \Longleftrightarrow z \in \partial\psi(x)$$

so we want to use (11) to prove (15). But in order to apply (11), we first need to prove that the conditions of Theorem 1.2 are satisfied. We will first prove that

$$\partial\psi^*(z) \subset \Omega^{\delta/2} \quad \forall z \in \partial\psi(\Omega^\delta). \tag{63}$$

This is not obvious, since the definition of $\partial\psi(\Omega^\delta)$ only guarantees that there exists at least one $\bar{x} \in \Omega^\delta$ such that $z \in \partial\psi(\bar{x})$ (in other words $\partial\psi^*(z) \cap \Omega^{\delta/2} \neq \emptyset$). We will prove (63) by contradiction: Assume that there exists also $x \in \Omega \setminus \Omega^{\delta/2}$ such that $z \in \partial\psi(x)$. Since ψ is convex, this implies that ψ must have a flat part along the segment $[\bar{x}, x]$. Indeed, the definition of the subdifferential implies that

$$\psi(tx + (1-t)\bar{x}) \geq \psi(\bar{x}) + tz \cdot (x - \bar{x}) \qquad \forall t \in [0,1]$$

and

$$\psi(tx + (1-t)\bar{x}) \geq \psi(x) + (1-t)z \cdot (\bar{x} - x) \qquad \forall t \in [0,1]$$

and a linear combination of those inequality yields

$$\psi(tx + (1-t)\bar{x}) \geq (1-t)\psi(\bar{x}) + t\psi(x) \qquad \forall t \in [0,1].$$

The convexity of ψ implies that we must have equality in this inequality.

After possibly replacing x with the point $[\bar{x}, x] \cap \partial\Omega^{\delta/2}$, we deduce (since $\bar{x} \in \Omega^\delta$) that ψ has a flat part of size at least $\delta/2$ in the set $\Omega^{\delta/2}$. By Corollary 1.3, this is impossible if $h_1 \leq k_1(\delta)$ and $h_2 \leq k_2(\delta)$ for some functions k_1, k_2 depending only on $\lambda_1\lambda_2$ and L_∞. This proves that (63) must hold.

Next, we use Theorem 1.2. We denote $\ell = |x - x'|$ and assume that $h_1 \leq \delta$ and that ℓ satisfies:

$$\ell \geq 2h_1, \quad \ell \geq \max\left\{\sqrt{C\lambda_1\lambda_2\, K h_2}, \frac{\lambda_1\lambda_2}{\delta} h_2\right\} \tag{64}$$

Then h_1 and h_2 satisfy

$$h_1 \leq \min\{\delta, \ell/2\} \quad h_2 \leq \min\left\{\sqrt{\frac{\delta\ell}{\lambda_1\lambda_2}}, \frac{\ell^2}{C\lambda_1\lambda_2\, K}\right\}. \tag{65}$$

In particular, h_1 and h_2 satisfies (9) and so we can apply Theorem 1.2 to get (see (11))

$$\max\{\varepsilon, h_1 K, \ell h_2\} \geq \delta K \min\left\{\frac{1}{\exp\left(\frac{C^4\lambda_1^4\lambda_2^4 K^8}{\ell^8}\right) - 1}, 1\right\}$$

Furthermore, under conditions (65) we can use Lemma 2.2 to write

$$K \geq \left(\frac{\delta\ell}{\lambda_1\lambda_2}\right)^{1/2}.$$

It follows that (recall that $D = \operatorname{diam}\Omega_1$),

$$\frac{C^4\lambda_1^4\lambda_2^4 K^8}{\ell^8} \geq C^4\left(\frac{\delta}{\ell}\right)^4 \geq C^4\left(\frac{\delta}{D}\right)^4$$

and so

$$\frac{1}{\exp\left(\frac{C^4\lambda_1^4\lambda_2^4K^8}{\ell^8}\right)-1} \leq \frac{1}{\exp\left(C^4\left(\frac{\delta}{D}\right)^4\right)-1}.$$

We deduce (using (4)) that there exists a constant C_0, depending on D, δ and the dimension such that

$$\max\{\varepsilon, Kh_1, Dh_2\} \geq \frac{1}{C_0}\frac{K}{\exp\left(\frac{C^4\lambda_1^4\lambda_2^4K^8}{\ell^8}\right)-1}.$$

We now observe the following elementary fact:

$$\forall a>0, \qquad u \mapsto \frac{u}{\exp(au^8)-1} \text{ is monotone decreasing in } u. \tag{66}$$

This implies in particular that for all ℓ we have

$$\frac{K}{\exp\left(\frac{C^4\lambda_1^4\lambda_2^4K^8}{\ell^8}\right)-1} \geq \frac{L_\infty}{\exp\left(\frac{C^4\lambda_1^4\lambda_2^4L_\infty^8}{\ell^8}\right)-1},$$

and so

$$\max\{\varepsilon, Kh_1, Dh_2\} \geq \sigma(\ell) := \frac{1}{C_0}\frac{L_\infty}{\exp\left(\frac{C^4\lambda_1^4\lambda_2^4L_\infty^8}{\ell^8}\right)-1}, \tag{67}$$

where the function $\sigma(\ell)$ is monotone increasing and satisfies $\lim_{\ell\to 0^+}\sigma(\ell)=0$. In other words

$$\text{either } \frac{\varepsilon}{\ell} \geq \frac{\sigma(\ell)}{\ell} \text{ or } Kh_1 \geq \sigma(\ell) \text{ or } Dh_2 \geq \sigma(\ell).$$

Since both functions $\ell \mapsto \sigma(\ell)$ and $\ell \mapsto \frac{\sigma(\ell)}{\ell}$ are monotone increasing (for the second one, this is a consequence of (66) again), we can introduce their inverses σ_1 and σ_2. The conditions above are then equivalent to

$$\ell \leq \max\left\{\sigma_2\left(\frac{\varepsilon}{\ell}\right), \sigma_1(Kh_1), \sigma_1(Dh_2)\right\}.$$

Combining this with (64), we deduce that for all $\ell>0$ we have

$$\ell \leq \max\left\{\sigma_2\left(\frac{\varepsilon}{\ell}\right), \max\{\sigma_1(Kh_1), 2h_1\}, \max\left\{\sigma_1(Dh_2), \sqrt{C\lambda_1\lambda_2\,L_\infty h_2}, \frac{\lambda_1\lambda_2}{\delta}h_2\right\}\right\}$$

and the general result follows, recalling that $\ell=|x-x'|$ and that Lemma 4.1 gives $|z-z'|\geq 2\frac{\varepsilon}{\ell}$.

It remains to treat the special case $h_1=h_2=0$, where we immediately obtain that

$$|x-x'| \leq \sigma_2(|z-z'|),$$

proving that for any given z the sub-differential of ψ^* is always reduced to one point (take $z=z'$ and any $x,\ x'\in\partial\psi^*(z)$). Consequently ψ^* is C^1 as claimed.

To bound σ_2, we trivially observe that

$$\frac{u}{\exp(a\,u^8)-1} \leq 2\,\frac{a^{-1/8}}{\exp(a\,u^8/2)-1}.$$

Consequently for some numerical constant $\tilde{C}$

$$\sigma_2^{-1}(\ell) = 2\,\frac{\sigma(\ell)}{\ell} \geq \frac{1}{\tilde{C}}\frac{\lambda_1^{-1/2}\,\lambda_2^{-1/2}}{\exp\left(\frac{\tilde{C}^4\lambda_1^4\lambda_2^4L_\infty^8}{\ell^8}\right)-1}.$$

Therefore for some $\tilde{C}$

$$\sigma_2(w) \leq \tilde{C}\, \frac{\sqrt{\lambda_1\,\lambda_2}\, L_\infty}{\left(\log\left(1+\frac{1}{\tilde{C}\,\sqrt{\lambda_1\,\lambda_2}\,w}\right)\right)^{1/8}},$$

which concludes the proof. □

Proof of lemma 4.1. By definition of ε, there exists $y \in (x, x')$, with $y = tx + (1-t)x'$ for some $t \in (0,1)$ such that

$$\psi(y) = t\psi(x) + (1-t)\psi(x') - \varepsilon. \tag{68}$$

By the definition of subdifferential we have that

$$\begin{aligned} \psi(z) &\geq \psi(x') + y' \cdot (z - x') \\ \psi(z) &\geq \psi(x') + y'' \cdot (z - x''). \end{aligned} \tag{69}$$

Plugging (68) into the inequalities (69) yields

$$\begin{aligned} y' \cdot (x' - x'') &\geq \psi(x') - \psi(x'') + \frac{\varepsilon}{1-t} \\ -y'' \cdot (x' - x'') &\geq \psi(x'') - \psi(x') + \frac{\varepsilon}{t}, \end{aligned}$$

so that finally, by adding both inequalities, we get

$$2\ell|y' - y''| \geq (y' - y'') \cdot (x' - x'') \geq \frac{\varepsilon}{1-t} + \frac{\varepsilon}{t} \geq 4\varepsilon,$$

which concludes the proof. □

5 Proof of Theorem 1.1

Theorem 1.1 follows from the following result together with well known facts from the theory of optimal transportation:

Theorem 5.1. *Let μ, ν be two probability measures on $\mathbb{R}^n$. Assume $\Gamma \subset \mathbb{R}^{2n}$ is a closed set such that for any Borel set O there holds:*

$$\begin{aligned} \mu(O) &\leq \nu\left(\{y \in \mathbb{R}^n \mid \exists x \in O,\ (x,y) \in \Gamma\}\right), \\ \nu(O) &\leq \mu\left(\{x \in \mathbb{R}^n \mid \exists y \in O,\ (x,y) \in \Gamma\}\right). \end{aligned} \tag{70}$$

Then there exists $\pi \in \Pi(\mu,\nu)$ concentrated on Γ (that is $\mathrm{Supp}(\pi) \subset \Gamma$*).*

Proof of Theorem 1.1. We note that when $\Gamma = \mathrm{Graph}\,\partial\psi$ for some convex function ψ, then (70) is equivalent to the conditions (1) and (6). Theorem 5.1 thus implies that there exists $\pi \in \Pi(\mu,\nu)$ such that $\mathrm{Supp}(\pi) \subset \Gamma$. The result then follows from a classical result of measure theory (see for example Theorem 2.12 in [22]). □

We now turn to the proof of Theorem 5.1. The first step is to prove the result when μ and ν are both sums of Dirac masses with identical mass:

$$\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \quad \nu = \frac{1}{N}\sum_{j=1}^N \delta_{x_j}.$$

In that case, we define the $N \times N$ matrix $A = (A_{ij})$ by

$$A_{ij} = \begin{cases} 1 & \text{if } (x_i, y_j) \in \Gamma, \\ 0 & \text{otherwise} \end{cases}$$

and Theorem 5.1 is equivalent to

Proposition 5.2. *Let A be a $N \times N$ matrix with $A_{ij} \in \{0,\ 1\}$ and such that for any I, J subsets of $\{1, \dots, N\}$, we have*

$$|I| \leq |\{j \mid \sum_{i \in I} A_{ij} > 0\}|, \quad |J| \leq |\{i \mid \sum_{j \in J} A_{ij} > 0\}|. \tag{71}$$

Then there exists a stochastic matrix $\pi = (\pi_{ij})$ such that $\pi_{ij} \in \{0, 1\}$, $\sum_j \pi_{ij} = 1$, $\sum_i \pi_{ij} = 1$ and satisfying $\pi_{ij} = 0$ whenever $A_{ij} = 0$.

Proof of Proposition 5.2. The proof proceeds by induction on N, the result being obvious for $N = 1$. We distinguish two cases: Whether there are strict subsets I_0 or J_0 for which there is equality in (71) or not.

Case 1: We assume that for all strict subsets I or J of $\{1, \dots, N\}$, we have a strict inequality in (71), that is

$$|I| < |\{j \mid \sum_{i \in I} A_{ij} > 0\}|, \quad |J| < |\{i \mid \sum_{j \in J} A_{ij} > 0\}|.$$

We then choose any i_0, j_0 such that $A_{i_0 j_0} = 1$. Up to relabeling the rows and columns of A, we can always assume that we can take $i_0 = j_0 = N$ and we consider the $N - 1 \times N - 1$ matrix $\tilde{A}$ consisting of the first $N - 1$ rows and columns of A. We claim that $\tilde{A}$ satisfies (71): Indeed, for any $I \subset \{1, \dots, N-1\}$, by applying (71) to A, we have that

$$\begin{aligned} |I| &\leq |\{j \in \{1, \dots, N\} \mid \sum_{i \in I} A_{ij} > 0\}| - 1 \\ &\leq |\{j \in \{1, \dots, N-1\} \mid \sum_{i \in I} A_{ij} > 0\}| \\ &= |\{j \mid \sum_{i \in I} \tilde{A}_{ij} > 0\}|. \end{aligned}$$

and a similar inequality for $J \subset \{1, \dots, N-1\}$.

Therefore by induction, there exists a stochastic $(N-1) \times (N-1)$ matrix $\tilde{\pi}$ with $\tilde{\pi} \in \{0, 1\}$, $\tilde{\pi}_{ij} = 0$ if $\tilde{A}_{ij} = 0$ and $\sum_i \tilde{\pi}_{ij} = 1$, $\sum_j \tilde{\pi}_{ij} = 1$. We can then define the matrix π by setting $\pi_{ij} = \tilde{\pi}_{ij}$ if $i \leq N - 1$ and $j \leq N - 1$, $\pi_{NN} = 1$ and $\pi_{ij} = 0$ otherwise. It is straightforward to check that π satisfies all requirements.

Case 2: We assume that there exists a strict subset I_0 or J_0 of $\{1, \dots, N\}$ for which we have equality in (71). For example, we assume that there is a strict subset I_0 such that

$$|I_0| = |\{j \mid \sum_{i \in I_0} A_{ij} > 0\}| \tag{72}$$

and we denote $J_0 = \{j \mid \sum_{i \in I_0} A_{ij} > 0\}$.

Since $|I_0| = |J_0|$, we can define the square matrices P and Q by

$$P_{ij} = A_{ij}\,\mathbb{I}_{i\in I_0, j\in J_0}, \quad Q_{ij} = A_{ij}\,\mathbb{I}_{i\in I_0^c, j\in J_0^c}.$$

and we are going to show that P and Q satisfy (71) on $I_0 \times J_0$, and $I_0^c \times J_0^c$ respectively.

We start with P: for any $I \subset I_0$, (71) implies:

$$|I| \leq |\{j \mid \sum_{i\in I} A_{ij} > 0\}|.$$

The definition of J_0 implies that $A_{ij} = 0$ if $i \in I_0$ and $j \notin J_0$. Since $I \subset I_0$, we can thus write:

$$\{j \mid \sum_{i\in I} A_{ij} > 0\} = \{j \in J_0 \mid \sum_{i\in I} A_{ij} > 0\} = \{j \in J_0 \mid \sum_{i\in I} P_{ij} > 0\},$$

and we deduce that

$$|I| \leq |\{j \in J_0 \mid \sum_{i\in I} P_{ij} > 0\}| \qquad \text{for all } I \subset I_0 \tag{73}$$

which shows that P satisfies the first inequality in (71). To prove the second inequality, we proceed by contradiction: Assume that there is a subset $J \subset J_0$ such that

$$|J| > |\{i \in I_0 \mid \sum_{j\in J} P_{ij} > 0\}| = |\{i \in I_0 \mid \sum_{j\in J} A_{ij} > 0\}|,$$

then denote $\tilde{I} = \{i \in I_0 \mid \sum_{j\in J} A_{ij} > 0\}$ and define $I' = I_0 \setminus \tilde{I}$, $J' = J_0 \setminus J$. Since $|J| > |\tilde{I}|$, we have that

$$|I'| = |I_0| - |\tilde{I}| > |I_0| - |J| = |J_0| - |J| = |J'|. \tag{74}$$

Consider any j' s.t. $\sum_{i\in I'} A_{ij'} > 0$. Since $A_{ij'} = 0$ for $i \in I_0$ and $j' \notin J_0$, we have that $j' \in J_0$. Next, let $i' \in I'$ be such that $A_{i'j'} > 0$ (which exists by the choice of j'). Since $i' \in I' = I_0 \setminus \tilde{I}$, the definition of $\tilde{I}$ implies that $A_{i'j} = 0$ for all $j \in J$. So we must have $j' \notin J$. We thus have $j' \in J'$. We just proved that

$$\{j' \mid \sum_{i'\in I'} A_{i'j'} > 0\} = \{j' \in J_0 \mid \sum_{i'\in I'} A_{i'j'} > 0\} \subset J'.$$

Together with (74), this implies that

$$|I'| > |J'| \geq |\{j' \mid \sum_{i'\in I'} A_{i'j'} > 0\}|,$$

which contradict (71). We can then conclude that P satisfies (71) on $I_0 \times J_0$.

We now turn to Q and start by considering any $J \subset J_0^c$. We recall that $A_{ij} = 0$ for $i \in I_0$ and $j \in J \subset J_0^c$ and since $\{1, \ldots, N\} = I_0 \cup I_0^c$ we have:

$$\{i \mid \sum_{j\in J} A_{ij} > 0\} = \{i \in I_0^c \mid \sum_{j\in J} A_{ij} > 0\}.$$

Applying (71) on A for J, we get

$$|J| \leq |\{i \mid \sum_{j\in J} A_{ij} > 0\}| = |\{i \in I_0^c \mid \sum_{j\in J} Q_{ij} > 0\}|,$$

proving the corresponding half of (71) for Q.

Next, for any $I \subset I_0^c$, we apply (71) with $\bar{I} = I_0 \cup I$:

$$|I_0| + |I| = |\bar{I}| \leq |\bar{J}|, \quad \bar{J} = \{j \mid \sum_{i \in \bar{I}} A_{ij} > 0\}.$$

We recall that $J_0 = \{j \mid \sum_{i \in I_0} A_{ij} > 0\}$ and denote $J = \{j \in J_0^c \mid \sum_{i \in I} A_{ij} > 0\}$. We have $J \cup J_0 \supset \{j \mid \sum_{i \in I} A_{ij} > 0\}$, and since

$$\bar{J} = \{j \mid \sum_{i \in I_0} A_{ij} > 0\} \bigcup \{j \mid \sum_{i \in I} A_{ij} > 0\}.$$

we have $\bar{J} = J_0 \cup J$. This implies that $|I_0| + |I| \leq |\bar{J}| \leq |J_0| + |J|$, that is

$$|I| \leq |J| = |\{j \in J_0^c \mid \sum_{i \in I} A_{ij} > 0\}|,$$

which completes the proof that Q satisfies (71).

We can now complete the proof: Since $I_0 \neq \emptyset$ and $I_0 \neq \{1, \dots, N\}$, P and Q have dimensions strictly less than N and we may apply our induction assumption. This gives us p_{ij} on $I_0 \times J_0$ and q_{ij} on $I_0^c \times J_0^c$, stochastic matrices,

$$\sum_j p_{ij} = 1 \quad \forall i \in I_0, \qquad \sum_j q_{ij} = 1 \quad \forall i \in I_0^c,$$

$$\sum_j p_{ij} = 1 \quad \forall j \in J_0, \qquad \sum_j q_{ij} = 1 \quad \forall j \in J_0^c,$$

with $p_{ij} = 0$ if $P_{ij} = 0$ and $q_{ij} = 0$ if $Q_{ij} = 0$. We simply extend p and q by 0 on the whole $\{1, \dots, N\}^2$ and define $\pi = p + q$.

□

Proposition 5.2 proves Theorem 5.1 when the measures μ and ν are both sums of Dirac masses with identical mass. Our next step is to extend that result to general sums of Dirac masses.

Proposition 5.3. *Assume that*

$$\mu = \sum_{i=1}^{M_1} m_i \, \delta_{x_i}, \quad \nu = \sum_{j=1}^{M_2} n_j \, \delta_{y_j}$$

for some points $x_1, \dots, x_{M_1}$ and $y_1, \dots, y_{M_2}$ in $\mathbb{R}^n$ and some (positive) masses $m_1, \dots, m_{M_1}$, and $n_1, \dots, n_{M_2}$. Then Theorem 5.1 holds.

Proof of Proposition 5.3. By splitting points if needed, we can always assume that $M_1 = M_2 = M$ and that $m_i, n_j > 0$ for all i, j. We can also assume that Γ is concentrated on $\bigcup_{i,j} \{(x_i, y_j)\}$. For this reason, we define γ as the set of indices (i, j) s.t. $(x_i, x_j) \in \Gamma$. In that discrete setting, Assumption (70) is then equivalent to

$$\begin{aligned} &\forall I \subset \{1, \dots, M\}, \quad \sum_{i \in I} m_i \leq \sum_{\{j \mid \exists i \in I \text{ s.t. } (i,j) \in \gamma\}} n_j, \\ &\forall J \subset \{1, \dots, M\}, \quad \sum_{i \in J} n_i \leq \sum_{\{i \mid \exists j \in J \text{ s.t. } (i,j) \in \gamma\}} m_j. \end{aligned} \tag{75}$$

In order to use the result of Proposition 5.2, we want to approximate μ and ν by measures μ_N, ν_N that are sums of Dirac masses with identical mass. To do this, given $N \geq 2\,M$, we replace the mass m_i at x_i (respectively the mass n_j at y_j) into $k(i)$ masses (respectively $l(j)$ masses) $\frac{1}{N}$ all located at that same point, with $k(i)$ and $l(j)$ such that

$$m_i - \frac{1}{N} \leq \frac{k(i)}{N} \leq m_i, \quad n_j - \frac{1}{N} \leq \frac{l(j)}{N} \leq n_j$$

(we can always assume that $1/N \leq \inf(\inf_i m_i,\ \inf_j n_j)$ so that $k(i), l(j) > 0$). We note that we have some left over mass $m_i - \frac{k(i)}{N}$ at each points. By summing over all i (and all j), the total left over masses are

$$1 - \sum_i \frac{k(i)}{N} = \frac{k(0)}{N} \in \left(0, \frac{M}{N}\right), \quad 1 - \sum_j \frac{l(j)}{N} = \frac{l(0)}{N} \leq 1 \in \left(0, \frac{M}{N}\right).$$

where $k(0) = N - \sum_i k(i) \leq M$ and $l(0) = N - \sum_j l(j) \leq M$. We thus add $k(0)$ (resp. $l(0)$) masses $1/N$ at some point $x_0 \neq x_i$ (resp. $y_0 \neq y_j$).

We can write

$$\mu_N = \frac{1}{N} \sum_{k=1}^{N} \delta_{\bar{x}_k}, \quad \nu_N = \frac{1}{N} \sum_{l=1}^{N} \delta_{\bar{y}_l}$$

where the points $\bar{x}_k$ (resp. $\bar{y}_l$) are the same points as the x_i or the additional distinct point x_0 (resp. y_i and y_0) which we just subdivide. It is useful to introduce $K(i)$ (resp. $L(j)$), the set of indices k such that $\bar{x}_k = x_i$ (resp. $\bar{y}_k = y_j$). We have in particular $|K(i)| = k(i)$ and $|L(j)| = l(j)$.

Observe that μ_N and ν_N converge strongly to μ and ν when $N \to \infty$ since

$$\int d|\mu_N - \mu| \leq \frac{2M}{N}, \quad \int d|\nu_N - \nu| \leq \frac{2M}{N}$$

(since there is a mass discrepancy of at most $1/N$ at each of the M points $x_1, \ldots, x_M$ and an additional mass of at most M/N at x_0).

We now define γ_N as

$$\gamma_N = \left(\bigcup_{(i,j)\in\gamma} K(i) \times L(j) \right) \bigcup (K(0) \times (\{1, \ldots, N\} \setminus L(0))) \\ \bigcup ((\{1, \ldots, N\} \setminus K(0)) \times L(0)) \,.$$

The first component of γ_N is the natural extension from γ: If x_i and y_j were connected, then any $\bar{x}_k$ s.t. $\bar{x}_k = x_i$ is still connected to any $\bar{y}_l$ with $\bar{y}_l = y_j$. Because we lose a bit of mass on the points $\bar{x}_k$ and $\bar{y}_l$, this may not be enough to ensure that (75) holds on μ_N and ν_N. For this reason, $K(0)$ and $L(0)$ serve as a mass reservoir: $K(0)$ is connected to all $l \geq 1$ and $L(0)$ to all $k \geq 1$, but of course $K(0)$ is not connected with $L(0)$.

We now check that (75) holds for this γ_N. Let I be a subset of $\{1, \ldots, N\}$. We consider three cases:

If $I \subset K(0)$ then $\{l \,|\, \exists k \in I \ s.t.\ (k,l) \in \gamma_N\} = \{1, \ldots, N\} \setminus L(0)$. Therefore

$$\mu_N(I) \leq \mu_N(K(0)) \leq \frac{M}{N} \leq 1 - \frac{M}{N} = \nu_N(\{l,\ \exists k \in I \ s.t.\ (k,l) \in \gamma_N\}),$$

since $N \geq 2\,M$.

If $I \cap K(0) \neq \emptyset$ but $I \not\subset K(0)$, then $\{l \,|\, \exists k \in I \ s.t. \ (k,l) \in \gamma_N\} = \{1, \ldots, N\}$. Trivially

$$\mu_N(I) \leq 1 = \nu_N(\{l, \ \exists k \in I \ s.t. \ (k,l) \in \gamma_N\}).$$

If $I \cap K(0) = \emptyset$, then denote $\bar{I} = \{i \,|\ K(i) \cap I \neq \emptyset\}$. Observe that if $(k,l) \in \gamma_N$ and $k \in K(i)$, $l \in L(j)$ then for any $k' \in K(i)$, $l' \in L(j)$ one also has that $(k',l') \in \gamma_N$ by the definition of γ_N. Therefore

$$\{l \,|\, \exists k \in I \ s.t. \ (k,l) \in \gamma_N\} = L(0) \bigcup \left(\bigcup_{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma} L(j) \right). \tag{76}$$

Consequently by the definition of $\bar{I}$ we have

$$\mu_N(I) = \frac{|I|}{N} \leq \sum_{i \in \bar{I}} \frac{k(i)}{N}$$

and since $k(i) \leq N\, m_i$ from the construction of μ_N we deduce

$$\mu_N(I) \leq \sum_{i \in \bar{I}} \frac{k(i)}{N} \leq \sum_{i \in \bar{I}} m_i = \mu(\bar{I}).$$

Applying (70) to μ, we get

$$\mu_N(I) \leq \mu(\bar{I}) \leq \nu(\{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma\}) = \sum_{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma} n_j.$$

From the construction of ν_N, we have $n_j \geq \frac{l(j)}{N}$ and

$$\begin{aligned}
\mu_N(I) &\leq \sum_{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma} \frac{l(j)}{N} + \sum_{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma} \left(n_j - \frac{l(j)}{N} \right) \\
&\leq \sum_{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma} \frac{l(j)}{N} + \sum_{j=1}^{M} \left(n_j - \frac{l(j)}{N} \right) \\
&\leq \sum_{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma} \frac{l(j)}{N} + 1 - \sum_{j=1}^{M} \frac{l(j)}{N} \\
&\leq \sum_{j, \ \exists i \in \bar{I} \ s.t. \ (i,j) \in \gamma} \frac{l(j)}{N} + \frac{l(0)}{N}
\end{aligned}$$

Using (76), we deduce

$$\mu_N(I) \leq \nu_N(\{l, \ \exists k \in I \ s.t. \ (k,l) \in \gamma_N\}),$$

which proves that (70) holds for the measures μ_N, ν_N and the set γ_N.

We now apply Proposition 5.2 to μ_N and ν_N. We obtain π_N a transference plan

$$\pi_N = \sum_{k,l} \frac{\pi^N_{k,l}}{N} \delta_{(\bar{x}_k, \, \bar{y}_l)}, \quad \sum_l \pi^N_{k,l} = 1, \quad \sum_k \pi^N_{k,l} = 1, \quad \pi^N_{k,l} = 0 \text{ if } (k,l) \notin \gamma_N.$$

Because the $\bar{x}_k$ and $\bar{y}_k$ are equal to the $x_i,\ y_j$ or to $x_0,\ y_0$, we can also express π_N as

$$\pi_N = \sum_{i,j\geq 1} \bar{\pi}^N_{i,j}\, \delta_{(x_i,\, y_j)} + \sum_{j\geq 1} \bar{\pi}^N_{0,j}\, \delta_{(x_0,\, y_j)} + \sum_{i\geq 1} \bar{\pi}^N_{0,j}\, \delta_{(x_i,\, y_0)}. \tag{77}$$

Moreover for any $i \geq 1$, or any $j \geq 1$

$$\sum_{j\geq 1} \bar{\pi}^N_{ij} + \pi^N_{i0} = \frac{k(i)}{N}, \quad \sum_{j\geq 1} \bar{\pi}^N_{ij} + \pi^N_{i0} = \frac{l(i)}{N}.$$

By the construction of $\mu_N,\ \nu_N$, this yields that

$$m_i - \frac{1}{N} \leq \sum_{j\geq 1} \bar{\pi}^N_{ij} + \pi^N_{i0} \leq m_i, \quad n_i - \frac{1}{N} \leq \sum_{i\geq 1} \bar{\pi}^N_{ij} + \pi^N_{0j} \leq n_i. \tag{78}$$

On the other hand,

$$\sum_{j} \bar{\pi}^N_{0j} = \frac{k(0)}{N} \leq \frac{M}{N}, \quad \sum_{i} \bar{\pi}^N_{i0} = \frac{l(0)}{N} \leq \frac{M}{N}. \tag{79}$$

Since the points $x_i,\ y_j$ together with $x_0,\ y_0$ are fixed, we may simply pass to the limit in $\pi_N \to \pi$ using (77) by extracting subsequences such that all $\bar{\pi}^N_{ij} \to \bar{\pi}_{ij}$ for all $i,\ j \geq 0$. By (79), we have that $\bar{\pi}_{0j} = \bar{\pi}_{i0} = 0$ so that

$$\pi = \sum_{i,j\geq 1} \bar{\pi}_{i,j}\, \delta_{(x_i,\, y_j)}.$$

By (78) and given that $\bar{\pi}_{i0} = \bar{\pi}_{0j} = 0$, we get

$$\sum_{j\geq 1} \bar{\pi}_{ij} = m_i, \quad \sum_{i\geq 1} \bar{\pi}_{ij} = n_j,$$

and so π is a transference plan between μ and ν. It only remains to check that π is concentrated on Γ: Given any $(i,j) \notin \gamma$, then for any N and any $k \in K(i),\ l \in L(j)$, we have that $(k,l) \notin \gamma_N$. Therefore $\pi^N_{k,l} = 0$ and

$$\bar{\pi}^N_{ij} = \sum_{k\in K(i),\ l\in L(j)} \frac{\pi^N_{kl}}{N} = 0,$$

so that we also have that $\bar{\pi}_{ij} = 0$, thus proving that π is indeed concentrated on Γ.

□

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. First of all by density, we can assume that both μ and ν are compactly supported on some large ball $B(0,R)$.

Define Γ_x and Γ_y the projections of Γ,

$$\Gamma_x = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^n,\ (x,y) \in \Gamma\}, \quad \Gamma_y = \{y \in \mathbb{R}^n \mid \exists x \in \mathbb{R}^n,\ (x,y) \in \Gamma\}.$$

Of course $\Gamma_x,\ \Gamma_y$ are closed. Moreover μ is supported on Γ_x: For any open set O with $O \cap \Gamma_x = \emptyset$ then for any $y \in \mathbb{R}^n$ and any $x \in O$, $(x,\, y) \notin \Gamma$ so

$$\{y \in \mathbb{R}^n \mid \exists x \in O,\ (x,y) \in \Gamma\} = \emptyset,$$

and therefore $\mu(O) = 0$ by assumption (70). Similarly ν is supported on Γ_y.

For any $k \in \mathbb{N}$, we define the hypercubes C_i^k of diameter 2^{-k}, centered at points $x_i \in 2^{-k}\mathbb{Z}^n$ and that cover a fixed selected hypercube C_0 s.t. $B(0,R) \subset C_0$. This decomposition is obviously hierarchical since C_i^k is composed of exactly 2^n small hypercubes C_j^{k+1}.

By shifting the hypercubes if necessary, we may assume that $\mu\left(\bigcup_i \partial C_i^k\right) = 0$. For any k, we define an approximation μ_N with $N = C\,2^{kn}$ points,

$$\mu_N = \sum_{i=1}^{N} m_i^N\, \delta_{x_i}, \quad m_i^N = \mu(C_i^k).$$

We also define ν_N in the same manner. Both μ_N and ν_N remain probability measures since $\sum_i m_i^N = \sum_i \mu(C_i^k) = \mu(\bigcup C_i^k)$ as $\mu\left(\bigcup_i \partial C_i^k\right) = 0$.

Finally we define Γ_N as the union $\bigcup_{(i,j)\in\gamma_N} C_i^k \times C_j^k$ where

$$\gamma_N = \{(i,j) \mid \exists x \in C_i^k,\ \exists y \in C_j^k,\ (x,y) \in \Gamma\}.$$

We observe of course that $\Gamma \subset \Gamma_N$ and that $d(\Gamma_N, \Gamma) \le C\,2^{-k} \to 0$ as $N \to \infty$.

Consider now any subset I of indices i and define $O = \bigcup_{i\in I} C_i^k$. By our construction and assumption (70)

$$\mu(O) = \sum_{i\in I} m_i^N \le \nu\left(\{y \in \mathbb{R}^n \mid \exists x \in O,\ (x,y) \in \Gamma\}\right).$$

By the definition of O, we have

$$\{y \in \mathbb{R}^n \mid \exists x \in O,\ (x,y) \in \Gamma\} = \bigcup_{i\in I} \{y \in \mathbb{R}^n \mid \exists x \in C_i^k,\ (x,y) \in \Gamma\}.$$

And since $\Gamma \subset \Gamma_N$, we deduce

$$\begin{aligned}\{y \in \mathbb{R}^n \mid \exists x \in O,\ (x,y) \in \Gamma\} &\subset \bigcup_{i\in I} \{y \in \mathbb{R}^n \mid \exists x \in C_i^k,\ (x,y) \in \Gamma_N\} \\ &\subset \bigcup_{\{j \mid \exists i \in I \text{ s.t. } (i,j)\in\gamma_N\}} C_j^k,\end{aligned}$$

by the definition of γ_N. Hence, since $\nu\left(\bigcup_i \partial C_i^k\right) = 0$

$$\nu\left(\{y \in \mathbb{R}^n \mid \exists x \in O,\ (x,y) \in \Gamma\}\right) \le \sum_{\{j \mid \exists i \in I \text{ s.t. } (i,j)\in\gamma_N\}} \nu(C_j^k) = \sum_{\{j \mid \exists i \in I \text{ s.t. } (i,j)\in\gamma_N\}} n_j^N$$

Therefore

$$\sum_{i\in I} m_i^N \le \sum_{\{j \mid \exists i \text{ s.t. } (i,j)\in\gamma_N\}} n_j^N$$

which is the first inequality in (75). The proof is similar when reversing the roles of μ_N and ν_N and this allows us to conclude that μ_N and ν_N satisfy (70) with Γ_N.

We can thus apply Proposition 5.3 to get the existence of a transference plan π_N concentrated on Γ_N and with marginals μ_N and ν_N.

By the tightness of π_N, we can extract a converging subsequence (still denoted by N for simplicity) s.t. $\pi_N \to \pi$ for the weak-* topology of measures. Trivially π has marginals μ and ν.

To conclude the proof, we need to show that π is concentrated on Γ: Consider any open set O with $O \cap \Gamma = \emptyset$ and any continuous function ϕ on $\mathbb{R}^{2n}$ with compact support on O.

We claim that $\sup_{\Gamma_N} |\phi| \to 0$ as N tends to ∞. Indeed assume, by contradiction, that there exists $(x_N,\ y_N) \in \Gamma_N$ and $\eta > 0$ s.t. $\phi(x_N, y_N) > \eta$. Since $\Gamma_N \subset B(0, 2R)$ then we can extract converging subsequences $x_N \to x$ and $y_N \to y$. Since $d(\Gamma_N, \Gamma) \to 0$ then $(x, y) \in \Gamma$ but, since ϕ is continuous, we have that $\phi(x, y) > \eta$. Recalling that ϕ has compact support in O with $O \cap \Gamma = \emptyset$ gives a contradiction.

We thus have

$$\int \phi \, d\pi_N = \int_{\Gamma_N} \phi \, d\pi_N \leq \sup_{\Gamma_N} |\phi| \longrightarrow 0, \quad \text{as } N \to \infty,$$

which gives $\int \phi \, d\pi = 0$, proving that π is concentrated on Γ and completing the proof of Theorem 5.1. □

A Proof of Theorem 1.5

As in the proof of our main theorem, we denote by $x = (x_\parallel, x_\perp)$ the points in $\Omega \subset \mathbb{R}^2$ where $x_\parallel$ is the coordinate along the line H and $x_\perp$ the orthogonal coordinate. We then have (the proof is similar to that of Lemma 2.1) that for all $x \in \Omega_1$ such that $|x_\parallel| \leq \ell/2$ there holds

$$|\partial_{x_\parallel} \psi(x)| \leq \frac{2K}{\ell} \, (|x_\perp| + \gamma) \, . \tag{80}$$

Next, we note that the fact that $\det D^2\psi \geq \lambda^{-2}$ implies that

$$\partial_{x_\parallel x_\parallel} \psi \, \partial_{x_\perp x_\perp} \psi \geq \lambda^{-2},$$

and the convexity of ψ gives $\partial_{x_\parallel x_\parallel} \psi$, $\partial_{x_\perp x_\perp} \psi \geq 0$. We deduce

$$\begin{aligned}
\left(\int_{-\frac{\ell}{2}}^{\frac{\ell}{2}} \lambda^{-1} \, dx_\parallel \right)^2 &\leq \left(\int_{-\frac{\ell}{2}}^{\frac{\ell}{2}} |\partial_{x_\parallel x_\parallel} \psi|^{1/2} |\partial_{x_\perp x_\perp} \psi|^{1/2} dx_\parallel \right)^2 \\
&\leq \int_{-\frac{\ell}{2}}^{\frac{\ell}{2}} \partial_{x_\parallel x_\parallel} \psi \, dx_\parallel \int_{-\frac{\ell}{2}}^{\frac{\ell}{2}} \partial_{x_\perp x_\perp} \psi \, dx_\parallel \\
&\leq \left[\partial_{x_\parallel} \psi \left(\frac{\ell}{2}, x_\perp \right) - \partial_{x_\parallel} \psi \left(-\frac{\ell}{2}, x_\perp \right) \right] \int_{-\frac{\ell}{2}}^{\frac{\ell}{2}} \partial_{x_\perp x_\perp} \psi dx_\parallel,
\end{aligned}$$

which implies (using (80))

$$\frac{\ell^3}{4\lambda^2 K (|x_\perp| + \gamma)} \leq \int_{-\frac{\ell}{2}}^{\frac{\ell}{2}} \partial_{x_\perp x_\perp} \psi dx_\parallel.$$

Finally, integrating with respect to $x_\perp$ we get

$$\begin{aligned}\frac{\ell^3}{4\lambda^2 K}\int_0^\delta \frac{dx_\perp}{x_\perp+\gamma}dx_\perp &\le \int_{-\frac{\ell}{2}}^{\frac{\ell}{2}}\int_0^\delta \partial_{x_\perp x_\perp}\psi\, dx_\perp\, dx_\parallel\\ &\le \int_{-\frac{\ell}{2}}^{\frac{\ell}{2}}\left[\partial_{x_\perp}\psi(x_\parallel,\delta)-\partial_{x_\perp}\psi(x_\parallel,0)\right]dx_\parallel\\ &\le 2K\ell,\end{aligned}$$

and (20) follows.

B Proof of Theorem 1.6

We have $1 \le d < n$ and we choose a system of coordinates

$$x=(x_\parallel,x_\perp)\in\mathbb{R}^d\times\mathbb{R}^{n-d} \quad \text{with } x_\parallel=(x_1,\dots x_d) \text{ and } x_\perp=(x_{d+1},\dots x_n),$$

so that $H=\{x_\perp=0\}$. For $\ell<\delta/2$, we have $B^d_\ell(0)\times B^{n-d}_{\delta/2}(0)\subset\Omega$, and the following lemma is the equivalent of Lemma 2.1 in this higher dimensional setting (the proof is similar),

Lemma B.1. *For all $x_\parallel\in B^d_{\ell/2}(0)$ and for all $x_\perp\in B^{n-d}_{\delta/2}$ we have*

$$|\nabla_{x_\parallel}\psi(x_\parallel,x_\perp)|\le\frac{2}{\ell}K(|x_\perp|+\gamma). \tag{81}$$

The starting point of the proof of Theorem 1.6 is the following consequence of Fischer's inequality

$$\det\left(D^2_{x_\parallel}\psi\right)\det\left(D^2_{x_\perp}\psi\right)\ge\det(D^2\psi)\ge\lambda^{-2}. \tag{82}$$

Integrating (82) with respect to $x_\parallel$ after taking the square root, we get for all $x_\perp\in B^{n-d}_{\delta/2}$

$$\begin{aligned}\lambda^{-1}\mathcal{L}^d(B^d_{\ell/2}) &\le \int_{B^d_{\ell/2}}\left(\det D^2_{x_\parallel}\psi\right)^{1/2}\left(\det D^2_{x_\perp}\psi\right)^{1/2}dx_\parallel\\ &\le \left(\int_{B^d_{\ell/2}}\det D^2_{x_\parallel}\psi\,dx_\parallel\right)^{1/2}\left(\int_{B^d_{\ell/2}}\det D^2_{x_\perp}\psi\,dx_\parallel\right)^{1/2}\\ &\le \left(\mathcal{L}^d\left(\nabla_{x_\parallel}\psi(B^d_{\ell/2}\times\{x_\perp\}\right)\right)^{1/2}\left(\int_{B^d_{\ell/2}}\det D^2_{x_\perp}\psi(x_\parallel,x_\perp)\,dx_\parallel\right)^{1/2},\end{aligned}$$

where we used the fact that for a convex function ϕ, the integral $\int_U\det D^2\phi\,dx$ is the volume of the image of U under $\nabla\phi$.

Using (81) we deduce

$$\lambda^{-2}\ell^{2d}\le C\left(\frac{1}{\ell}K(|x_\perp|+\gamma)\right)^d\int_{B^d_{\ell/2}}\det D^2_{x_\perp}\psi(x_\parallel,x_\perp)\,dx_\parallel,$$

which implies in particular

$$\begin{aligned}\int_{B^{n-d}_{\delta/2}} \frac{\lambda^{-2}\ell^{3d}}{K^d(|x_\perp|+\gamma)^d}\,dx_\perp &\le C\int_{B^{n-d}_{\delta/2}}\int_{B^d_\ell} \det D^2_{x_\perp}\psi\,dx_\parallel dx_\perp\\ &\le C\int_{B^d_\ell}\int_{B^{n-d}_{\delta/2}} \det D^2_{x''}\psi\,dx_\perp dx_\parallel\\ &\le C\int_{B^d_\ell}\mathcal{L}^{n-d}\left(\nabla_{x_\perp}\psi(\{x_\parallel\}\times B^{n-d}_{\delta/2})\right)\,dx_\parallel\\ &\le C\ell^d K^{n-d}.\end{aligned}$$

We finally obtain that

$$\ell^{2d}\int_{B^{n-d}_{\delta/2}} \frac{1}{(|x_\perp|+\gamma)^d}\,dx_\perp \le C\lambda^2 K^n, \tag{83}$$

where we can write

$$\int_{B^{n-d}_{\delta/2}} \frac{1}{(|x_\perp|+\gamma)^d}\,dx_\perp = \int_0^{\delta/2}\frac{r^{n-d-1}}{(r+\gamma)^d}\,dr = \gamma^{n-2d}\int_0^{\delta/2\gamma}\frac{r^{n-d-1}}{(r+1)^d}\,dr$$

and (22) follows.

References

[1] AD Aleksandrov. Smoothness of the convex surface of bounded gaussian curvature. In *Dokl. Akad. Nauk SSSR*, volume 36, pages 211–216, 1942.

[2] Francesco Bonsante and François Fillastre. The equivariant minkowski problem in minkowski space. In *Annales de l'Institut Fourier*, volume 67, pages 1035–1113, 2017.

[3] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[4] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[5] Luis A Caffarelli. A localization property of viscosity solutions to the monge-ampère equation and their strict convexity. *Annals of Mathematics*, 131(1):129–134, 1990.

[6] Luis A Caffarelli. Some regularity properties of solutions of monge-ampère equation. *Communications on pure and applied mathematics*, 44(8-9):965–969, 1991.

[7] Luis A Caffarelli. Boundary regularity of maps with convex potentials. *Communications on pure and applied mathematics*, 45(9):1141–1151, 1992.

[8] Luis A Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992.

[9] Luis A Caffarelli. Boundary regularity of maps with convex potentials–ii. *Annals of mathematics*, 144(3):453–496, 1996.

[10] Guido De Philippis. *Regularity of optimal transport maps and applications*, volume 17 of *Tesi. Scuola Normale Superiore di Pisa (Nuova Series) [Theses of Scuola Normale Superiore di Pisa (New Series)]*. Edizioni della Normale, Pisa, 2013.

[11] Guido De Philippis and Alessio Figalli. Partial regularity for optimal transport maps. *Publ. Math. Inst. Hautes Études Sci.*, 121:81–112, 2015.

[12] Alessio Figalli. Regularity properties of optimal maps between nonconvex domains in the plane. *Comm. Partial Differential Equations*, 35(3):465–479, 2010.

[13] Alessio Figalli and Young-Heon Kim. Partial regularity of Brenier solutions of the Monge-Ampère equation. *Discrete Contin. Dyn. Syst.*, 28(2):559–565, 2010.

[14] M. Goldman, F. Otto, and M. Huesmann. A large-scale regularity theory for the monge-ampere equation with rough data and application to the optimal matching problem. *arXiv:1808.09250*, 2018.

[15] Michael P Goldman and Felix Otto. A variational proof of partial regularity for optimal transportation maps. *arXiv:1704.05339*, 2017.

[16] Cristian E Gutiérrez. *The Monge-Ampère equation*, volume 44. Springer, 2001.

[17] Erhard Heinz. Über die differentialungleichung $0 < \alpha \leq rt - s^2 \leq \beta < \infty$. *Mathematische Zeitschrift*, 72(1):107–126, 1959.

[18] Leonid Kantorovitch. On the translocation of masses. *Management Science*, 5(1):1–4, 1958.

[19] R Tyrrell Rockafellar. *Convex analysis*, volume 28. Princeton university press, 1970.

[20] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[21] Neil S Trudinger and Xu-Jia Wang. The monge-ampère equation and its geometric applications. *Handbook of geometric analysis*, 1:467–524, 2008.

[22] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[23] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[24] Yifeng Yu. Singular set of a convex potential in two dimensions. *Comm. Partial Differential Equations*, 32(10-12):1883–1894, 2007.