



UNIVERSITÀ
DI PADOVA

DIPARTIMENTO DI MATEMATICA “TULLIO LEVI-CIVITA”

Corso di Laurea in Matematica

Caffarelli's contraction theorem
and related estimates on Brenier maps

Relatore:
Prof. Guido De Philippis

Laureando: Andrea Bidoia
Matricola: 2118454

Anno Accademico 2025/2026

17 Luglio 2026

Contents

Introduction	5
1 Overview of classical optimal transport	7
1.1 Monge-Kantorovich problems	7
1.2 Cyclical monotonicity	10
1.2.1 Convex analysis tools	11
1.3 Characterization of optimal plans	13
1.4 Stability	16
2 The Monge-Ampère equation	17
2.1 Classical and Brenier solutions	17
2.2 Aleksandrov solutions	19
2.3 Regularity	21
2.4 Stability	22
3 Global estimates on the Brenier map	25
3.1 Caffarelli's contraction theorem	25
3.2 The divergence of the Brenier map	29
3.3 Good estimates	32
Appendix	37
<i>References</i>	41

Introduction

The history of optimal transport began at the end of the eighteenth century, when the French geometer Gaspard Monge proposed the following problem in [24]: assume that you are extracting a fixed amount of soil from the ground, and you wish to transport it to places where it is incorporated in a construction. The source and target distributions of soil are prescribed, but you may choose how to assign each extraction spot to a destination. Monge assumed that the cost of transporting a unit of mass was proportional to the distance traveled. How should you assign each place to a destination to minimize the total cost?

The study of this natural problem has largely developed in the last century. An interesting question is one of regularity: one expects optimal maps to be more regular than general transport maps (informally, introducing irregularities is expensive). If so, optimal maps provide controlled examples of change of variables; this idea has many applications in functional and geometric inequalities (for example, see [18] for a quantitative version of the isoperimetric inequality).

In this thesis we focus on Caffarelli's contraction theorem, an important example of a global regularity estimate. The original statement asserts that the Brenier map from a Gaussian measure to a log-concave perturbation is 1-Lipschitz. As a result, many functional inequalities transfer from the Gaussian setting to the perturbed case with no loss in the constant. We aim to generalize this statement to different estimates on the derivative of the Brenier map (the Hessian of the Brenier potential, by Theorem 1.3.2). First, we present a more modern result due to Gozlan and Sylvestre, which can be viewed as an analogous estimate for the Laplacian of the Brenier potential, and give an original proof revisiting Caffarelli's argument. Next, we show that Caffarelli's theorem and its Laplacian counterpart are two examples of a single phenomenon, and extend these results to a wider class of estimates given by *good* functions of the Hessian, i.e. convex, increasing, and positively homogeneous functions (in the sense made precise in Section 3.3).

In **Chapter 1**, we give a brief overview of the classical theory of optimal transport, focusing on the case of quadratic cost in $\mathbb{R}^n$; the main result in this chapter is the existence and characterization of optimal maps as gradients of convex functions. In **Chapter 2**, we introduce the Monge-Ampère equation, a powerful tool throughout the study of the regularity of optimal maps, and mention the main ideas and results in this direction. In **Chapter 3**, we present the proofs of Caffarelli's theorem, its Laplacian counterpart, and the generalization to good functions of the Hessian. The second and third parts of this chapter are original contributions based on the author's preprint [4].

Chapter 1

Overview of classical optimal transport

1.1 Monge-Kantorovich problems

Given a topological space $\mathcal{X}$, we denote by $\mathcal{P}(\mathcal{X})$ the set of Borel probability measures on $\mathcal{X}$. The correct setting for the general problems in optimal transport is that of Polish spaces, i.e. complete, separable metric spaces.

Definition 1.1.1. Given $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and a Borel cost $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$, the Monge Problem is to minimize

$$\mathcal{C}(T) := \int_{\mathcal{X}} c(x, T(x)) \, d\mu(x) \quad (\text{MP})$$

among all transport maps T from μ to ν , i.e. all Borel maps $T: \mathcal{X} \rightarrow \mathcal{Y}$ such that $T_{\#}\mu = \nu$.

Remark 1.1.2. The condition $T_{\#}\mu = \nu$ needs only to be checked on open sets. In particular it is equivalent to

$$\int_{\mathcal{X}} \phi \circ T \, d\mu = \int_{\mathcal{Y}} \phi \, d\nu \quad \forall \phi \in C_b(\mathcal{Y})$$

since the characteristic functions of open sets are monotone non-decreasing limits of functions in $C_b(\mathcal{Y})$.

In general, Monge's problem is ill-posed because:

- the set of transport maps can be empty (for example if μ is a Dirac delta and ν is not);
- even when there are transport maps, the minimum in (MP) may not be attained.

As an example of the second phenomenon, we can consider $\mathcal{X} = \mathcal{Y} = \mathbb{R}^2$, with quadratic cost $c(x, y) := |x - y|^2$ and measures

$$\mu := \mathcal{H}^1 \llcorner (\{0\} \times [0, 1]) \quad \text{and} \quad \nu := (1/2) \mathcal{H}^1 \llcorner (\{-1, +1\} \times [0, 1]),$$

where $\mathcal{H}^1$ is the one-dimensional Hausdorff measure. The distance between $\text{supp } \mu$ and $\text{supp } \nu$ is 1, hence $\mathcal{C}(T) \geq 1$ for any transport map T , with equality if and only if T is a horizontal translation μ -almost everywhere. In that case, the set of points $\{0\} \times \mathcal{D}$ which get sent to the right segment $\{+1\} \times [0, 1]$ has μ -measure $1/2$, hence its left translation $\{-1\} \times \mathcal{D}$ has ν -measure $1/4$, which is a contradiction as $\{-1\} \times \mathcal{D}$ can only receive mass from points sent non-horizontally. A simple construction shows that $\mathcal{C}(T)$ can be arbitrarily close to 1.

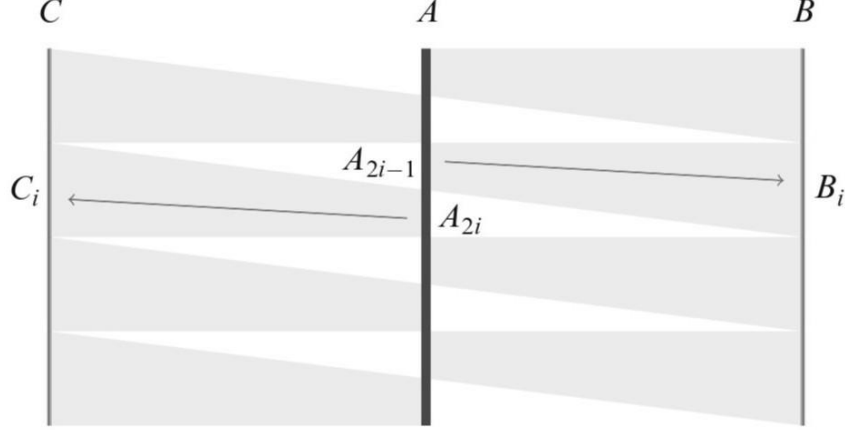


Figure 1.1: The filled regions represent the lines of transport for a map with cost $1 + o(1)$ as the number of segments goes to infinity. Reproduced from [27].

A way to overcome these issues is via the following relaxation to the original Monge formulation:

Definition 1.1.3. Given $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and a Borel cost $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$, the Kantorovich Problem is to minimize

$$\mathcal{C}(\pi) := \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi \quad (\text{KP})$$

among all *transport plans* π from μ to ν , i.e. all probability measures $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ that satisfy the *marginal conditions*

$$e_{1\#}\pi = \mu \quad \text{and} \quad e_{2\#}\pi = \nu,$$

where e_1, e_2 are the projections on the first and second coordinates respectively. We denote by $\Pi(\mu, \nu)$ the set of admissible transport plans.

This relaxation enjoys many advantages, in particular:

- $\mu \otimes \nu \in \Pi(\mu, \nu)$, hence the class of transport plans is not empty;
- by the linearity of $\pi \mapsto \int c \, d\pi$, this formulation is an instance of a linear programming problem;
- optimal plans exist under mild assumptions on the cost c (Theorem 1.1.9);
- transport plans generalize transport maps, in particular any transport map T induces a plan $\pi := (\text{Id} \times T)_{\#}\mu$ such that $\mathcal{C}(\pi) = \mathcal{C}(T)$.

The following lemma, along with the previous remark, gives a characterization of the plans induced by transport maps:

Lemma 1.1.4. *If a plan π is concentrated on a Borel graph Γ , then there exists a Borel transport map T such that $\pi = (\text{Id} \times T)_\# \mu$.*

Proof. If π is concentrated on $\Gamma = \text{graph}(T)$, then by inner-regularity we may assume $\Gamma = \bigcup_k \Gamma_k$ to be σ -compact; under this assumption, $e_1(\Gamma) = \text{dom}(T)$ is Borel, and the restriction of T to each $e_1(\Gamma_k)$ is continuous (its graph is compact). It follows that T is Borel. Since $y = T(x)$ $\pi - a.e.$, we conclude that

$$\int_{\mathcal{X} \times \mathcal{Y}} \phi(x, y) d\pi = \int_{\mathcal{X} \times \mathcal{Y}} \phi(x, T(x)) d\pi = \int_{\mathcal{X}} \phi(x, T(x)) d\mu \quad \forall \phi \text{ Borel},$$

therefore $\pi = (\text{Id} \times T)_\# \mu$. □

The strategy to find minimizers to (KP) is a simple application of the Direct Methods in the Calculus of Variations. First, we prove that $\Pi(\mu, \nu)$ is compact in the weak topology, then we check that the cost map is weakly lower semicontinuous; hence, it has a minimum over $\Pi(\mu, \nu)$.

Definition 1.1.5 (Weak convergence). Given $(\mu_k)_k \subset \mathcal{P}(\mathcal{X})$, we say that μ_k *converge weakly* to $\mu \in \mathcal{P}(\mathcal{X})$ if

$$\int_{\mathcal{X}} \phi d\mu_k \rightarrow \int_{\mathcal{X}} \phi d\mu \quad \text{for all } \phi \in C_b(\mathcal{X}),$$

and we write $\mu_k \rightharpoonup \mu$. A classical result known as the Portmanteau Theorem shows that the previous condition is equivalent to either of the following:

$$\begin{aligned} \limsup_n \mu_n(F) &\leq \mu(F) \quad \text{for all closed } F \subset \mathcal{X}, \\ \liminf_n \mu_n(U) &\geq \mu(U) \quad \text{for all open } U \subset \mathcal{X}. \end{aligned}$$

Theorem 1.1.6 (Prokhorov). *A set $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ is precompact for the weak topology if and only if it is tight, i.e. for any $\varepsilon > 0$, there is a compact set K_ε such that $\mu(\mathcal{X} \setminus K_\varepsilon) < \varepsilon$ for all $\mu \in \mathcal{A}$.*

Theorem 1.1.7. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$. Then the set of transport plans $\Pi(\mu, \nu)$ is compact for the weak topology.*

Proof. By writing the marginal condition in the form

$$\int_{\mathcal{X}} \phi d\mu = \int_{\mathcal{X} \times \mathcal{Y}} \phi d\pi \quad \forall \phi \in C_b(\mathcal{X}), \quad \int_{\mathcal{Y}} \psi d\nu = \int_{\mathcal{X} \times \mathcal{Y}} \psi d\pi \quad \forall \psi \in C_b(\mathcal{Y}),$$

it is immediate that $\Pi(\mu, \nu)$ is weakly closed; by Prokhorov's theorem, we shall prove its tightness. Let $\varepsilon > 0$ and choose $K \subset \mathcal{X}$ and $K' \subset \mathcal{Y}$ compact sets such that $\mu(\mathcal{X} \setminus K) < \varepsilon/2$ and $\nu(\mathcal{Y} \setminus K') < \varepsilon/2$, then for any $\pi \in \Pi(\mu, \nu)$:

$$\begin{aligned} \pi(\mathcal{X} \times \mathcal{Y} \setminus K \times K') &\leq \pi((\mathcal{X} \setminus K) \times \mathcal{Y}) + \pi(\mathcal{X} \times (\mathcal{Y} \setminus K')) \\ &= \mu(\mathcal{X} \setminus K) + \nu(\mathcal{Y} \setminus K') < \varepsilon, \end{aligned}$$

therefore, the compact set $K \times K'$ satisfies the tightness condition for ε . □

Theorem 1.1.8. *If the cost $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and bounded below, then $\pi \mapsto \mathcal{C}(\pi)$ is weakly lower semicontinuous and bounded below.*

Proof. Since c is lower semicontinuous and bounded below, it can be written as the pointwise limit of a non-decreasing family $(c_\ell)_\ell$ of continuous and bounded functions. If $\pi_k \rightharpoonup \pi$, then

$$\int c \, d\pi = \lim_{\ell \rightarrow \infty} \int c_\ell \, d\pi = \lim_{\ell \rightarrow \infty} \lim_{k \rightarrow \infty} \int c_\ell \, d\pi_k \leq \liminf_{k \rightarrow \infty} \int c \, d\pi_k.$$

Therefore, the function $\pi \mapsto \mathcal{C}(\pi)$ is weakly lower semicontinuous and bounded below by $\inf c$. $\square$

Corollary 1.1.9. *If the cost $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and bounded below, then the minimum in (KP) is attained.*

Once we establish the existence of minimizers for (KP), a few natural questions arise:

- are the optimal plans unique?
- can we characterize optimal plans?
- when are optimal plans induced by maps?
- while (MP) may not have a minimum, is $\inf(\text{MP}) = \min(\text{KP})$?

A sufficient condition for the last equality is given in the following lemma:

Lemma 1.1.10 (Pratelli, [25]). *If $(\mathcal{X}, \mu)$ is non-atomic and $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ is continuous and bounded below, then the infimum in (MP) is equal to the minimum in (KP).*

As for the other questions, without any prior information on the cost or the measures, one cannot expect a positive answer. We will focus our study on the quadratic setting, namely $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$, $c(x, y) := |x - y|^2$.

1.2 Cyclical monotonicity

In this section we introduce a geometric property called *cyclical monotonicity*, and prove that it is the right notion to characterize optimal plans in the quadratic setting.

Example 1.2.1. We start with a discrete example. Let us assume that

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

Then any transport plan π can be written as

$$\pi = \frac{1}{m} \sum_{i,j} \pi_{ij} \delta_{(x_i, y_j)}$$

where $(\pi_{ij})_{ij}$ is a doubly stochastic matrix, i.e. $\pi_{ij} \geq 0$ and

$$\sum_{j=1}^m \pi_{ij} = 1 \quad \forall i, \quad \sum_{i=1}^m \pi_{ij} = 1 \quad \forall j.$$

In this setting, (KP) is a finite-dimensional linear problem with a convex constraint, hence the minimum is achieved at one of the extremal points of the set of admissible solutions. Birkhoff's Theorem states that these extremal points are precisely the set of permutation matrices, which correspond to transport maps.

Among all the finitely many transport maps, assume that $T: x_i \mapsto y_i$ is optimal. Then all other maps are obtained by rearranging $(y_i)_i$; hence, the optimality condition reads as

$$\sum_{i=1}^m |x_i - y_i|^2 \leq \sum_{i=1}^m |x_i - y_{\sigma(i)}|^2 \quad \text{for all permutations } \sigma.$$

By cancellation of the terms fixed by σ , this condition is equivalent to the following: for all subsets $\{i_1, \dots, i_k\} \subset \{1, \dots, m\}$,

$$\sum_{\ell=1}^k |x_{i_\ell} - y_{i_\ell}|^2 \leq \sum_{\ell=1}^k |x_{i_\ell} - y_{i_{\ell+1}}|^2$$

where $i_{k+1} := i_1$. This example motivates the following definition:

Definition 1.2.2. A set $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is said to be *cyclically monotone* if for all $m \geq 1$ and $(x_1, y_1), \dots, (x_m, y_m) \in \Gamma$,

$$\sum_{i=1}^m |x_i - y_i|^2 \leq \sum_{i=1}^m |x_i - y_{i+1}|^2 \quad \text{or, equivalently,} \quad \sum_{i=1}^m \langle x_{i+1} - x_i, y_i \rangle \leq 0 \quad (1.1)$$

with indices mod m . A transport plan π is said to be cyclically monotone if it is concentrated on a cyclically monotone set.

Informally, a cyclically monotone plan is one that cannot be improved by local perturbations (as considered before, namely by rerouting mass along a cycle); therefore, it is intuitive that any optimal plan should be cyclically monotone. Under mild assumptions, we will see that the converse also holds true. Hence, we are interested in understanding cyclically monotone sets.

1.2.1 Convex analysis tools

A function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to be *convex* if its epigraph

$$\text{epi}(f) := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\}$$

is convex. Equivalently,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y), \quad \forall x, y \in \mathbb{R}^n, \quad \forall t \in [0, 1].$$

A convex function f is said to be *proper* if $f \not\equiv +\infty$, and *lower semicontinuous* if $\text{epi}(f)$ is closed. We denote by $\Gamma_0(\mathbb{R}^n)$ the set of proper convex and lower semicontinuous functions. We define the subdifferential of f at a point x as the (possibly empty) closed convex set

$$\partial f(x) := \{x^* : f(y) \geq f(x) + \langle y - x, x^* \rangle \forall y\}.$$

If $f(x) < +\infty$, the previous inequality has a simple geometric meaning, namely that the graph of the affine function $h(y) := f(x) + \langle y - x, x^* \rangle$ is a non-vertical hyperplane supporting $\text{epi}(f)$ at $(x, f(x))$. Moreover, if f is differentiable at x , then $\partial f(x) = \{\nabla f(x)\}$.

Given $f \in \Gamma_0(\mathbb{R}^n)$, we define its *Legendre transform* as

$$f^*(x^*) := \sup_{x \in \mathbb{R}^n} \{\langle x, x^* \rangle - f(x)\}.$$

One can prove that f^* is also proper, convex and lower semicontinuous, and that $f^{**} = f$ (hence f^* is also called *convex conjugate*). In addition, the subdifferentials of f and f^* are dual in the sense of this lemma:

Lemma 1.2.3. *Let $f \in \Gamma_0(\mathbb{R}^n)$, then for any pair of points x and x^* , the following conditions are equivalent:*

1. $x^* \in \partial f(x)$;
2. $y \mapsto \langle y, x^* \rangle - f(y)$ achieves its maximum at $y = x$;
3. $x \in \partial f^*(x^*)$;
4. $y^* \mapsto \langle x, y^* \rangle - f^*(y^*)$ achieves its maximum at $y^* = x^*$;
5. $f(x) + f^*(x^*) \leq \langle x, x^* \rangle$;
6. $f(x) + f^*(x^*) = \langle x, x^* \rangle$.

The following theorem due to Rockafellar gives a geometric characterization of cyclically monotone sets in terms of the tools described above:

Theorem 1.2.4. *A set $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone if and only if it is included in the subdifferential of some proper convex and lower semicontinuous function.*

Proof. Let $\phi \in \Gamma_0(\mathbb{R}^n)$; we show that $\partial \phi$ (hence all its subsets) is cyclically monotone. Let $(x_1, y_1), \dots, (x_m, y_m) \in \partial \phi$; then for all i ,

$$\phi(x_{i+1}) - \phi(x_i) \geq \langle x_{i+1} - x_i, y_i \rangle,$$

which summed up yields (1.1). Conversely, choose $(x_0, y_0) \in \Gamma$ and define

$$\phi(x) := \sup \{\langle x - x_m, y_m \rangle + \dots + \langle x_1 - x_0, y_0 \rangle : (x_1, y_1), \dots, (x_m, y_m) \in \Gamma\}.$$

Then ϕ is the supremum of affine functions, hence it is convex and lower semicontinuous; by cyclical monotonicity $\phi(x_0) = 0$ thus ϕ is proper. We now prove $\Gamma \subset \partial \phi$, i.e.

$$\phi(z) - \phi(x) \geq \langle z - x, y \rangle \quad \text{for all } (x, y) \in \Gamma \text{ and } z \in \mathbb{R}^n.$$

To see this, notice that for all $\varepsilon > 0$, there exist $(x_1, y_1), \dots, (x_m, y_m) \in \Gamma$ such that

$$\phi(x) + \langle z - x, y \rangle \leq \langle x - x_m, y_m \rangle + \dots + \langle x_1 - x_0, y_0 \rangle + \langle z - x, y \rangle + \varepsilon \leq \phi(z) + \varepsilon.$$

Letting $\varepsilon \rightarrow 0^+$ concludes the proof. □

1.3 Characterization of optimal plans

Using the theory of cyclically monotone sets, we answer the questions posed above. Notice that if the minimum cost in (KP) is infinite, any plan is optimal, which means that there is no information to gain from a characterization of optimal plans. To avoid this, we initially consider only probability measures with finite second moments, i.e.

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

and we denote by $\mathcal{P}_2(\mathbb{R}^n)$ the set of such measures. In this setting we can easily see that *any* plan has finite cost.

Theorem 1.3.1. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$, then a plan $\pi \in \Pi(\mu, \nu)$ is optimal for (KP) if and only if $\text{supp } \pi$ is cyclically monotone.*

Proof: Necessity. Assume that there exist $(x_1, y_1), \dots, (x_m, y_m) \in \text{supp } \pi$ such that

$$\sum_{i=1}^m |x_i - y_{i+1}|^2 < \sum_{i=1}^m |x_i - y_i|^2.$$

We want to show that moving mass from x_i to y_{i+1} instead of y_i is more convenient. Choose U_i, V_i neighborhoods of x_i, y_i such that $m_i := \pi(U_i \times V_i) > 0$ and

$$\sum_{i=1}^m |u_i - v_{i+1}|^2 < \sum_{i=1}^m |u_i - v_i|^2 \quad \text{for all } u_i \in U_i, v_i \in V_i.$$

Consider the probability space $(\Omega, \mathbf{P})$ with

$$\Omega := \prod_{i=1}^m U_i \times V_i \quad \text{and} \quad \mathbf{P} := \prod_{i=1}^m \frac{\pi \llcorner U_i \times V_i}{m_i}$$

and define the new plan

$$\tilde{\pi} := \pi + \frac{\min m_i}{m} \sum_{i=1}^m \left[(u_i, v_{i+1})_{\#} \mathbf{P} - (u_i, v_i)_{\#} \mathbf{P} \right],$$

where $u_i(\omega), v_i(\omega)$ are the natural projections of $\omega \in \Omega$ to U_i, V_i respectively. We now show the following:

1. $\tilde{\pi}$ is non-negative: it is sufficient to show $\pi \geq \frac{\min m_i}{m} \sum_{i=1}^m (u_i, v_i)_{\#} \mathbf{P}$. For any $\psi \geq 0$,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \psi d\pi \geq \int_{U_i \times V_i} \psi d\pi \geq \frac{1}{m} \sum_{i=1}^m m_i \int_{\Omega} \psi(u_i, v_i) d\mathbf{P} \geq \frac{\min m_i}{m} \sum_{i=1}^m \int_{\Omega} \psi(u_i, v_i) d\mathbf{P}.$$

2. $\tilde{\pi} \in \Pi(\mu, \nu)$:

$$\begin{aligned} e_{1\#} \tilde{\pi} &= \mu + \frac{\min m_i}{m} \sum_{i=1}^m \left[u_{i\#} \mathbf{P} - u_{i\#} \mathbf{P} \right] = \mu, \\ e_{2\#} \tilde{\pi} &= \nu + \frac{\min m_i}{m} \sum_{i=1}^m \left[v_{i+1\#} \mathbf{P} - v_{i\#} \mathbf{P} \right] = \nu. \end{aligned}$$

3. $\mathcal{C}(\tilde{\pi}) < \mathcal{C}(\pi)$:

$$\sum_{i=1}^m \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d((u_i, v_{i+1})_{\#} \mathbf{P}) = \int_{\Omega} \sum_{i=1}^m |u_i - v_{i+1}|^2 d\mathbf{P} < \int_{\Omega} \sum_{i=1}^m |u_i - v_i|^2 d\mathbf{P}.$$

It follows that π is not optimal.

Sufficiency. Assume that $\text{supp } \pi$ is cyclically monotone; by Theorem 1.2.4, there is a proper convex and lower semicontinuous function ϕ such that $\text{supp } \pi \subset \partial\phi$. Let ϕ^* be the convex conjugate of ϕ , thus

$$\phi(x) + \phi^*(y) \geq \langle x, y \rangle, \quad (1.2)$$

with equality on $\partial\phi$. Since ϕ and ϕ^* are proper, (1.2) implies

$$\phi(x) \geq c_1 - c_2|x|^2, \quad \text{and} \quad \phi^*(y) \geq c'_1 - c'_2|y|^2$$

for some $c_2, c'_2 > 0$; hence, for any $\tilde{\pi} \in \Pi(\mu, \nu)$, one has $\phi_-(x), \phi^*(y) \in L^1(\tilde{\pi})$. Integrating (1.2) and using the previous relations to divide the integral, we obtain

$$\begin{aligned} \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, y \rangle d\tilde{\pi} &\leq \int_{\mathbb{R}^n \times \mathbb{R}^n} (\phi(x) + \phi^*(y)) d\tilde{\pi} = \int_{\mathbb{R}^n} \phi d\mu + \int_{\mathbb{R}^n} \phi^* d\nu \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} (\phi(x) + \phi^*(y)) d\pi = \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, y \rangle d\pi. \end{aligned}$$

Since μ, ν have finite second moments, it is easily seen that the previous inequality is equivalent to $\mathcal{C}(\tilde{\pi}) \geq \mathcal{C}(\pi)$; hence, π is optimal. $\square$

The following theorem gives a complete answer to the existence and the uniqueness questions in (MP), and will be an important starting point for the discussion of the Monge Problem in terms of convex potentials instead of optimal maps.

Theorem 1.3.2 (Brenier, Knott-Smith). *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$. Suppose that μ is absolutely continuous with respect to the Lebesgue measure. Then*

1. (KP) has a unique minimizer π . In addition, π is induced by a transport map T (which is the unique μ -a.e. minimizer of (MP)) and $T = \nabla\phi$, where $\phi: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex and lower semicontinuous function differentiable μ -almost everywhere;
2. conversely, if ϕ is convex, lower semicontinuous, differentiable μ -almost everywhere and $|\nabla\phi| \in L^2(\mu)$, then $T := \nabla\phi$ is optimal from μ to $T_{\#}\mu \in \mathcal{P}_2(\mathbb{R}^n)$;
3. moreover, if ν is absolutely continuous with respect to the Lebesgue measure, then $S := \nabla\phi^*$ is the unique optimal map from ν to μ , and

$$\nabla\phi^*(\nabla\phi(x)) = x \text{ } \mu\text{-a.e.} \quad \text{and} \quad \nabla\phi(\nabla\phi^*(y)) = y \text{ } \nu\text{-a.e.}$$

Proof.

1. Combining the existence of a minimum (Theorem 1.1.9), its cyclical monotonicity (Theorem 1.3.1) and the characterization of cyclically monotone sets (Theorem 1.2.4) we obtain that (KP) has a minimizer π , and $\text{supp } \pi \subset \partial\phi$ for some proper convex and lower semicontinuous function ϕ . Since

$$\partial\phi \subset \{\phi < +\infty\} \times \mathbb{R}^n \quad \text{and} \quad \partial\{\phi < +\infty\} \text{ is Lebesgue-negligible,}$$

μ is concentrated on $\{\phi < +\infty\}$. Let $E := \{\text{points of non differentiability of } \phi\}$, then E is Lebesgue-negligible, hence

$$\pi(E \times \mathbb{R}^n) = \mu(E) = 0.$$

In particular, π is concentrated on $\partial\phi \setminus (E \times \mathbb{R}^n) = \text{graph}(\nabla\phi)$ and by Lemma 1.1.4, we conclude that $\pi = (\text{Id} \times \nabla\phi)_\# \mu$. If π' is another optimal plan, then the same argument yields $\pi' = (\text{Id} \times \nabla\psi)_\# \mu$, but by linearity of the cost functional $\pi'' := \frac{1}{2}\pi + \frac{1}{2}\pi'$ is also an optimal plan, and it is concentrated on a graph if and only if $\nabla\phi = \nabla\psi$ μ -a.e.

2. Our assumptions guarantee $T_\# \mu \in \mathcal{P}_2(\mathbb{R}^n)$. By Theorem 1.2.4, $\text{graph}(\nabla\phi) \subset \partial\phi$ is cyclically monotone, hence $(\text{Id} \times T)_\# \mu$ is a minimizer for (KP).
3. Observe that if a plan is optimal from μ to ν , its “inversion”

$$i_\# \pi, \quad \text{where } i(x, y) := (y, x)$$

is optimal from ν to μ . But $\partial\phi^*$ is the (set) inversion of $\partial\phi$, hence

$$\pi = (\nabla\phi^* \times \text{Id})_\# \nu \quad \text{and} \quad S = \nabla\phi^* \quad \nu\text{-a.e.}$$

By a change of variables

$$0 = \int_{\mathbb{R}^n \times \mathbb{R}^n} |y - \nabla\phi(x)|^2 d\pi = \int_{\mathbb{R}^n} |y - \nabla\phi(\nabla\phi^*(y))|^2 d\nu,$$

therefore, $\nabla\phi(\nabla\phi^*(y)) = y$ ν -a.e., and similarly $\nabla\phi^*(\nabla\phi(x)) = x$ μ -a.e.

□

Observe that a sufficient condition on μ is that the sets

$$\{\text{points of non differentiability of } \phi\} \quad \text{and} \quad \partial\{\phi < +\infty\}$$

are μ -negligible. It can be shown that both of those sets have Hausdorff dimension less than or equal to $n - 1$; the following theorem due to McCann makes use of this weaker hypothesis and removes the assumptions of bounded second moments.

Theorem 1.3.3 (McCann). *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$. Suppose that μ vanishes on Borel sets of Hausdorff dimension less than or equal to $n - 1$. Then there exists a proper convex and lower semicontinuous function ϕ differentiable μ -almost everywhere such that $\nabla\phi_\# \mu = \nu$. If ψ is another function as above, then $\nabla\psi = \nabla\phi$ μ -almost everywhere.*

Remark 1.3.4. While uniqueness of optimal plans fails if the minimum in (KP) is infinite, under the hypothesis of Theorem 1.3.3 it is the only case in which it does. Indeed, the proof of Theorem 1.3.1 shows that a finite cost is sufficient to guarantee that every optimal plan π is cyclically monotone and thus is supported on the subdifferential of a convex function $\phi \in \Gamma_0(\mathbb{R}^n)$. As in the proof of Theorem 1.3.2, $\pi = (\text{Id} \times \nabla\phi)_\# \mu$ follows from the assumptions on μ . For this reason we refer to $T = \nabla\phi$ as the Brenier map even if the cost may be infinite.

1.4 Stability

We conclude this chapter with a weak stability result:

Theorem 1.4.1. *Let $(\mu_k)_k$ and $(\nu_k)_k$ be sequences of probability measures in $\mathbb{R}^n$. Assume that $\mu_k \rightharpoonup \mu$ and $\nu_k \rightharpoonup \nu$. For each k , let π_k be a cyclically monotone transport plan between μ_k and ν_k . Then (up to subsequence) π_k converges weakly to a cyclically monotone transport plan π between μ and ν . Moreover, if $\mu_k = \mu$ for all k , $\pi_k = (\text{Id} \times T_k)_\# \mu$ and $\pi = (\text{Id} \times T)_\# \mu$, then $T_k \rightarrow T$ in μ -measure.*

Proof. Since $(\mu_k)_k$ and $(\nu_k)_k$ are weakly convergent, by Prokhorov's theorem they form tight sets. By an analogous argument to the proof of Theorem 1.1.7, this implies that $\{\pi_k\}_k$ is tight, and that $\pi_k \rightharpoonup \pi \in \Pi(\mu, \nu)$ up to subsequence. For each k and m , the cyclical monotonicity of π_k implies that for $\pi_k^{\otimes m}$ -almost all $(x_1, y_1), \dots, (x_m, y_m)$, one has

$$\sum_{i=1}^m |x_i - y_i|^2 \leq \sum_{i=1}^m |x_i - y_{i+1}|^2. \quad (1.3)$$

In other words, $\pi_k^{\otimes m}$ is concentrated on the set $\mathcal{C}_m$ of all $(x_1, y_1), \dots, (x_m, y_m)$ satisfying (1.3). Since $\mathcal{C}_m$ is closed, it follows that the weak limit $\pi^{\otimes m}$ of $\pi_k^{\otimes m}$ is also concentrated on $\mathcal{C}_m$; this holds for all m , hence π is cyclically monotone.

Assume further that $\pi_k = (\text{Id} \times T_k)_\# \mu$ and $\pi = (\text{Id} \times T)_\# \mu$. Let $\varepsilon > 0$, then by Lusin's theorem, for each $\delta > 0$ there is a closed set $K \subset \mathbb{R}^n$ such that $\mu(\mathbb{R}^n \setminus K) < \delta$ and the restriction of T to K is continuous. It follows that the set

$$C_{\varepsilon\delta} := \{(x, y) \in K \times \mathbb{R}^n : |T(x) - y| \geq \varepsilon\}$$

is closed, which implies

$$\begin{aligned} 0 = \pi(C_{\varepsilon\delta}) &\geq \limsup_{k \rightarrow \infty} \pi_k(C_{\varepsilon\delta}) \\ &= \limsup_{k \rightarrow \infty} \mu(\{x \in K : |T(x) - T_k(x)| \geq \varepsilon\}) \\ &\geq \limsup_{k \rightarrow \infty} \mu(\{|T - T_k| \geq \varepsilon\}) - \delta. \end{aligned}$$

Letting $\delta \rightarrow 0$ concludes the proof. $\square$

Notice that in the case of a unique cyclically monotone plan between μ and ν , the convergence holds for the whole sequence $(\pi_k)_k$.

Remark 1.4.2. In general, if the measure μ is replaced by the sequence $(\mu_k)_k$, the maps T_k may not converge to T , even in μ_k -measure. An example is

$$\mu_k = \delta_{1/k}, \quad \mu = \delta_0, \quad \nu_k = \nu = \delta_0, \quad T_k(x) = 0, \quad T(x) = \mathbf{1}_{x \neq 0}.$$

However, if we assume that T is continuous in $\mathbb{R}^n$, then

$$\mu_k(\{|T - T_k| \geq \varepsilon\}) \rightarrow 0 \quad \text{for all } \varepsilon > 0,$$

as we can directly take the closed set

$$C_\varepsilon := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : |T(x) - y| \geq \varepsilon\}$$

and notice that

$$0 = \pi(C_\varepsilon) \geq \limsup_{k \rightarrow \infty} \pi_k(C_\varepsilon) = \limsup_{k \rightarrow \infty} \mu_k(\{|T - T_k| \geq \varepsilon\}).$$

Chapter 2

The Monge-Ampère equation

Once we have established the existence and uniqueness of Brenier maps, a natural problem is to deduce some regularity information from the regularity of μ, ν . Again, without prior information on the measures, the answer is negative, as the following examples show.

Example A. Let μ be the uniform probability on the unit ball B in $\mathbb{R}^2$, and let ν be the uniform probability on a set C obtained by separating B into two halves B_1 and B_2 and moving them apart symmetrically. Then both μ and ν have smooth densities on their supports, but $\text{supp } \nu$ is disconnected; thus, there is no continuous transport map from μ to ν .

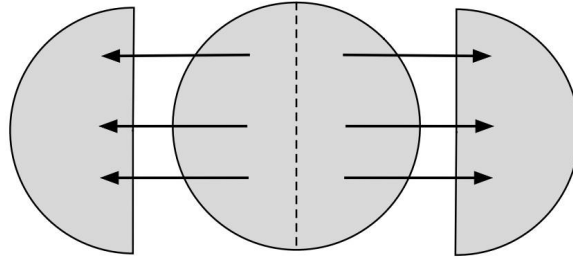


Figure 2.1: A discontinuous optimal map. Reproduced from [14].

A more interesting variation of this first example shows that assuming $\text{supp } \nu$ to be connected is still not enough to guarantee the continuity of the Brenier map.

Example B. Let μ be as in the previous example, and let ν_ε be the uniform probability on the set C_ε obtained by connecting the set C with a thin strip of width $O(\varepsilon)$. By weak stability (Theorem 1.4.1), the optimal maps T_ε converge to the previous optimal map in μ -measure as $\varepsilon \rightarrow 0$. Using this, one can easily show that for $\text{graph}(T_\varepsilon)$ to be cyclically monotone, T_ε has to be discontinuous for small enough ε .

As we will later see, Caffarelli understood that the right assumption on $\text{supp } \nu$ is *convexity*.

2.1 Classical and Brenier solutions

In this section we introduce the Monge-Ampère equation, an important tool throughout the study of regularity and stability of Brenier maps. We assume that both μ and ν are

absolutely continuous with respect to the Lebesgue measure, in particular

$$d\mu = f dx, \quad d\nu = g dx,$$

where

$$\text{supp } \mu = \overline{\Omega_1}, \quad \text{supp } \nu = \overline{\Omega_2}$$

and $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ are open, bounded and with Lebesgue-negligible boundary. We also assume that f, g are bounded above and below by positive constants, namely

$$0 < \lambda \leq f, g \leq 1/\lambda$$

on their supports. First, notice that *any* transport map T from μ to ν that is C^1 and injective satisfies

$$|\det \nabla T| = \frac{f}{g \circ T}$$

from a simple change of variables. In the case $T = \nabla \phi$ with ϕ convex, this becomes the so-called Monge-Ampère equation

$$\det \nabla^2 \phi = \frac{f}{g \circ \nabla \phi}. \quad (2.1)$$

The above equation should be considered with the “boundary condition”

$$\nabla \phi(\Omega_1) \subset \Omega_2, \quad (2.2)$$

and we call a convex C^2 function satisfying (2.1) and (2.2) a *classical solution* to the Monge-Ampère equation. As we have seen from the examples above, the Brenier map is generally not C^1 ; hence, we are interested in weak solutions. Recall the following version of the change of variables formula [14,17].

Proposition 2.1.1. *Let T be differentiable almost everywhere, and let $\Sigma := \text{dom } \nabla T$. Then for every Borel and bounded function ψ ,*

$$\int_{\Sigma} \psi |\det \nabla T| dx = \int_{\mathbb{R}^n} \left(\sum_{x \in \Sigma \cap T^{-1}(y)} \psi(x) \right) dy.$$

In this case, if $T := \nabla \phi$ is a Brenier map, then by Aleksandrov’s Theorem [1] $\nabla \phi$ is differentiable almost everywhere in $\text{dom } \nabla \phi$. Since $\nabla \phi^* \circ \nabla \phi = \text{Id}$ almost everywhere on Ω_1 , $\nabla \phi_{\#} \mu = \nu$ implies

$$\int_{\Omega_1} \psi(x) f(x) dx = \int_{\Omega_2} \psi(\nabla \phi^*(y)) g(y) dy,$$

for all Borel and bounded functions ψ . Applying Proposition 2.1.1 to T restricted to $\text{dom } \nabla \phi$ and using $\nabla \phi \circ \nabla \phi^* = \text{Id}$ almost everywhere on Ω_2 , we obtain

$$\int_{\Omega_1} \psi(x) f(x) dx = \int_{\Omega_1} \psi(x) \det \nabla^2 \phi(x) g(\nabla \phi(x)) dx.$$

It follows that

$$\det \nabla^2 \phi = \frac{f}{g \circ \nabla \phi} \quad \text{a.e. in } \Omega_1, \quad (2.3)$$

and we call a convex function satisfying (2.3) and (2.2) a *Brenier solution*. While partial regularity results are known for Brenier solutions (see [15,19]), by the previous examples we cannot hope for continuity on the whole domain.

2.2 Aleksandrov solutions

Let $\Omega \subset \mathbb{R}^n$ be an open convex set, and let $u: \Omega \rightarrow \mathbb{R}$ be a convex function. We define the *Monge-Ampère measure* of u as follows:

$$\mu_u(E) := |\partial u(E)| \quad \text{for every Borel set } E \subset \Omega, \quad (2.4)$$

where

$$\partial u(E) := \bigcup_{x \in E} \partial u(x),$$

and $|\cdot|$ denotes the Lebesgue outer measure. Note that if $u \in C^2(\Omega)$, the change of variable formula gives

$$\mu_u(E) = |\nabla u(E)| = \int_E \det \nabla^2 u \, dx \quad \text{for every Borel set } E \subset \Omega,$$

therefore, $d\mu_u = \det \nabla^2 u \, dx$ inside Ω . In general, the following proposition holds.

Proposition 2.2.1. *Let $u: \Omega \rightarrow \mathbb{R}$ be a convex and continuous function. Then μ_u is a Borel measure, and its absolutely continuous part with respect to the Lebesgue measure has density $\det \nabla^2 u$.*

Definition 2.2.2. Let μ be a Borel measure defined in Ω , an open and convex subset of $\mathbb{R}^n$. A convex function $u \in C(\Omega)$ is an *Aleksandrov solution* to the Monge-Ampère equation

$$\det D^2 u = \mu$$

if $\mu_u = \mu$ as Borel measures. When $d\mu = h \, dx$ we will simply write $\det D^2 u = h$.

In the previous setting, Proposition 2.2.1 shows that any Aleksandrov solution $\phi \in C(\mathbb{R}^n)$ to

$$\det D^2 \phi = \frac{f}{g \circ \nabla \phi} \quad (2.5)$$

restricts to a Brenier solution on Ω_1 (where we consider f to be vanishing outside Ω_1). The converse is generally false, as the following example shows.

Example 2.2.3. Let $d\mu = f \, dx$ and $d\nu = g \, dx$ be as in Example A. Then the Brenier map can be computed,

$$T = \nabla \phi, \quad \text{where} \quad \phi(x) = \frac{1}{2}|x|^2 + |x_1|.$$

From the previous considerations, ϕ is a Brenier solution to

$$\det \nabla^2 \phi = \frac{f}{g \circ \nabla \phi} = 1 \quad \text{a.e. in } B.$$

Meanwhile,

$$\partial \phi(x) = \begin{cases} \{x + e_1\} & x_1 > 0 \\ \{x - e_1\} & x_1 < 0, \\ [-1, 1] \times \{x_2\} & x_1 = 0 \end{cases}$$

hence, $\det D^2\phi = \mathcal{L}^2 \llcorner B + \mathcal{H}^1 \llcorner (\{x_1 = 0\} \cap B)$.

The reason for the lack of regularity is that for Aleksandrov solutions with f, g bounded away from zero and infinity, (2.5) implies $|E| \approx |\partial\phi(E)|$, i.e.

$$c|E| \leq |\partial\phi(E)| \leq C|E| \quad \text{for some } c, C > 0.$$

Instead, for Brenier solutions we have $\mu(E) = \nu(\partial\phi(E))$; hence, only $|E| \approx |\partial\phi(E) \cap \Omega_2|$ holds, which means that we lose full control of the singular part of μ_ϕ .

The following theorem shows that convexity of Ω_2 is enough to pass from a Brenier solution to an Aleksandrov one.

Theorem 2.2.4. *Let μ, ν satisfy the assumptions at the beginning of Section 2.1. Assume further that Ω_2 is convex. Let $T := \nabla\phi$ be the Brenier map from μ to ν , then ϕ coincides on Ω_1 with an Aleksandrov solution of*

$$\det D^2\phi = \frac{f}{g \circ \nabla\phi}.$$

Proof. The proof is the composition of the following two steps:

1. We construct a Brenier potential $\tilde{\phi}$ that coincides with ϕ on Ω_1 , such that $\nabla\tilde{\phi}(\mathbb{R}^n) \subset \overline{\Omega}_2$. Note that in this step the convexity of Ω_2 is not necessary.
2. Assuming without loss of generality that $\nabla\phi(\mathbb{R}^n) \subset \overline{\Omega}_2$, we show that ϕ is an Aleksandrov solution, provided that Ω_2 is convex.

Step 1. Define

$$\tilde{\phi}(x) := \sup_{y \in \overline{\Omega}_2} (\langle x, y \rangle - \phi^*(y)). \quad (2.6)$$

Recalling that

$$\phi(x) = \sup_{y \in \mathbb{R}^n} (\langle x, y \rangle - \phi^*(y)),$$

we get $\tilde{\phi} \leq \phi$. Since $\overline{\Omega}_2$ is compact and ϕ^* is lower semicontinuous, the maximum in (2.6) is attained; thus $\tilde{\phi}$ is convex and finite on the entire $\mathbb{R}^n$, and hence continuous [26, Corollary 10.1.1]. If ϕ is differentiable at x and $\nabla\phi(x) \in \overline{\Omega}_2$, then

$$\begin{aligned} \phi(x) &= \langle x, \nabla\phi(x) \rangle - \phi^*(\nabla\phi(x)), \\ \tilde{\phi}(x) &\geq \phi(x). \end{aligned}$$

Since the previous condition holds in a dense subset of Ω_1 , it follows that $\tilde{\phi} = \phi$ in Ω_1 , and in particular $\nabla\tilde{\phi} \# \mu = \nu$. To show $\nabla\tilde{\phi}(\mathbb{R}^n) \subset \overline{\Omega}_2$, let $x_0 \in \text{dom } \nabla\tilde{\phi}$ and $y_0 \in \overline{\Omega}_2$ be a maximizer for (2.6). Then

$$\begin{aligned} \langle x, y_0 \rangle - \phi^*(y_0) &\leq \tilde{\phi}(x) = \langle x_0, y_0 \rangle - \phi^*(y_0) + \langle x - x_0, \nabla\phi(x_0) \rangle + o(|x - x_0|), \\ \langle x - x_0, y_0 \rangle &\leq \langle x - x_0, \nabla\tilde{\phi}(x_0) \rangle + o(|x - x_0|), \end{aligned}$$

which implies $\nabla\tilde{\phi}(x_0) = y_0 \in \overline{\Omega}_2$.

Step 2. By Proposition 2.2.1, the absolutely continuous part of $\det D^2\phi$ has density $\det \nabla^2\phi$, hence

$$\det D^2\phi \geq \frac{f}{g \circ \nabla\phi}.$$

To prove the other inequality, we only need to show

$$|\partial\phi(\mathbb{R}^n)| \leq \int_{\Omega_1} \frac{f}{g \circ \nabla\phi} dx.$$

Recall that for any $x \in \text{int}(\text{dom } \phi)$,

$$\partial\phi(x) = \overline{\text{co}(\nabla_*\phi(x))}$$

where

$$\nabla_*\phi(x) := \{\text{limit points of } \nabla\phi(x_k) \text{ as } \text{dom } \nabla\phi \ni x_k \rightarrow x\}$$

is the set of *reachable gradients* and $\text{co}(\cdot)$ denotes the convex hull of the set [26, Theorem 25.6]. The convexity of Ω_2 yields

$$\partial\phi(\mathbb{R}^n) \subset \overline{\text{co}(\nabla\phi(\mathbb{R}^n))} \subset \overline{\Omega_2},$$

hence,

$$|\partial\phi(\mathbb{R}^n)| \leq |\Omega_2| = \int_{\Omega_2} \frac{g(y)}{g(y)} dy = \int_{\Omega_1} \frac{f}{g \circ \nabla\phi} dx,$$

as desired. $\square$

2.3 Regularity

In this section we will give a brief overview (omitting proofs) of the main regularity results for Aleksandrov solutions to the Monge-Ampère equation, focusing on the optimal transport setting.

Remark 2.3.1. Let $u: \Omega \rightarrow \mathbb{R}$ be a smooth solution of

$$\det \nabla^2 u = h \quad \text{in } \Omega, \tag{2.7}$$

with $h = h(x) > 0$ smooth. To prove regularity estimates on u , a standard technique when dealing with nonlinear equations is to differentiate the equation solved by u to obtain a linear second-order equation on its first derivatives. More precisely, if we differentiate (2.7) in direction $\alpha \in \mathbb{S}^{n-1}$, we get

$$(\det H) \text{Tr} (H^{-1} \partial_\alpha H) = \partial_\alpha h \quad \text{in } \Omega,$$

where $H(x) := \nabla^2 u(x)$. Substituting $\det H = h$, we obtain

$$\text{Tr}(H^{-1} \partial_\alpha H) = \frac{\partial_\alpha h}{h} \quad \text{in } \Omega. \tag{2.8}$$

To apply elliptic regularity theory for linear equations and obtain some regularity estimates on $\partial_\alpha u$, one would like the matrix H^{-1} to be positive semidefinite (which corresponds to the convexity of u). Still, without any a priori bound on $\nabla^2 u$, the matrix H^{-1} may have arbitrarily small eigenvalues, and this is why one says that (2.7) is “degenerate elliptic”. Proving a bound of the type $c \text{Id} \preceq \nabla^2 u \preceq C \text{Id}$ on Ω makes the linearized equation (2.8) uniformly elliptic; in particular, the *uniform convexity* of u will be a determining factor for the regularity of solutions.

An important starting point for the regularity theory is the following result due to Caffarelli, which establishes Hölder regularity of the gradients of strictly convex Aleksandrov solutions (i.e. such that the supporting hyperplanes intersect the graph at exactly one point).

Theorem 2.3.2 ([5]). *Let $u: \Omega \rightarrow \mathbb{R}$ be a strictly convex Aleksandrov solution of*

$$0 < \lambda \leq \det D^2 u \leq 1/\lambda.$$

Then $u \in C_{\text{loc}}^{1,\alpha}(\Omega)$ for some $\alpha > 0$ depending only on n and λ .

In the context of optimal transport, the convexity of the target support allows us to obtain Aleksandrov solutions, as shown in Theorem 2.2.4. In this setting, one can prove that the Brenier potential is strictly convex, which combined with the previous result yields Hölder regularity of the Brenier map. Higher-order regularity is obtained classically via a bootstrap argument.

Theorem 2.3.3 ([6]). *Let μ, ν satisfy the assumptions at the beginning of Section 2.1. Assume further that Ω_2 is convex. Let $T := \nabla \phi$ be the Brenier map from μ to ν , then ϕ is strictly convex and $C_{\text{loc}}^{1,\alpha}(\Omega_1)$. Moreover, if $f \in C_{\text{loc}}^{k,\beta}(\Omega_1)$ and $g \in C_{\text{loc}}^{k,\beta}(\Omega_2)$ for some $\beta \in (0, 1)$, then $\phi \in C_{\text{loc}}^{k+2,\beta}(\Omega_1)$.*

While Caffarelli's setting deals with bounded domains and densities globally bounded away from zero and infinity on their respective supports, these conditions have been relaxed (see [2, 12]). In particular, we mention the case where $\text{supp } \mu = \mathbb{R}^n$ and $\text{supp } \nu$ is (possibly) unbounded, which will be useful in Chapter 3.

Theorem 2.3.4 ([12]). *Let $Y \subset \mathbb{R}^n$ be a convex open set, and μ, ν be probability measures with densities $f: \mathbb{R}^n \rightarrow \mathbb{R}^+$ and $g: Y \rightarrow \mathbb{R}^+$ respectively. Assume that $\log f \in L^\infty(\mathbb{R}^n \cap B_R)$ and $\log g \in L^\infty(Y \cap B_R)$ for any $R > 0$. Let $T := \nabla \phi$ be the Brenier map from μ to ν , then ϕ is strictly convex, and for any compact set $X' \Subset \mathbb{R}^n$, there exists $\alpha' > 0$ such that $\phi \in C^{1,\alpha'}(X')$. Moreover, if $f \in C_{\text{loc}}^{k,\beta}(\mathbb{R}^n)$ and $g \in C_{\text{loc}}^{k,\beta}(Y)$ for some $\beta \in (0, 1)$, then $\phi \in C_{\text{loc}}^{k+2,\beta}(\mathbb{R}^n)$.*

2.4 Stability

We conclude the chapter with a stronger stability result, which holds in the unbounded setting of Theorem 2.3.4.

Theorem 2.4.1. *Let $Y \subset \mathbb{R}^n$ be a convex open set, and μ_k, ν_k be probability measures with densities $f_k: \mathbb{R}^n \rightarrow \mathbb{R}^+$ and $g_k: Y \rightarrow \mathbb{R}^+$ respectively. Assume that $\log f_k \in L^\infty(\mathbb{R}^n \cap B_R)$ and $\log g_k \in L^\infty(Y \cap B_R)$ uniformly in k for k large enough, for any $R > 0$. Assume that $f_k \rightarrow f$ and $g_k \rightarrow g$ in L_{loc}^1 , where f, g are probability densities on $\mathbb{R}^n$ and Y respectively. Then, the associated Brenier potentials $\phi_k \rightarrow \phi$ locally uniformly (up to translations), and $\nabla \phi_k \rightarrow \nabla \phi$ locally in measure.*

Proof. By L_{loc}^1 convergence, we have $\log f \in L^\infty(\mathbb{R}^n \cap B_R)$ and $\log g \in L^\infty(Y \cap B_R)$ for any $R > 0$; hence, by Theorem 2.3.4, the Brenier maps $\nabla \phi_k, \nabla \phi$ are continuous. The L_{loc}^1 convergence of the densities also implies weak-* convergence and tightness of the

measures, which combined yield weak convergence. By Theorem 1.4.1 and Remark 1.4.2, it follows that

$$\mu_k(\{|\nabla\phi - \nabla\phi_k| \geq \varepsilon\}) \rightarrow 0 \quad \text{for all } \varepsilon > 0.$$

Using the local uniform boundedness, for all compact sets $K \Subset \mathbb{R}^n$ one has

$$\lambda_K \mathcal{L}^n(\{|\nabla\phi - \nabla\phi_k| \geq \varepsilon\} \cap K) \leq \mu_k(\{|\nabla\phi_k - \nabla\phi| \geq \varepsilon\}) \rightarrow 0,$$

hence, $\nabla\phi_k \rightarrow \nabla\phi$ locally in measure. Without loss of generality, assume that

$$\phi_k(0) = \phi(0) = 0 \quad \forall k.$$

If $\nabla\phi_k$ are locally uniformly bounded, then ϕ_k are locally equi-Lipschitz and locally equibounded. By the Ascoli-Arzelà Theorem and a diagonal argument, we may extract a subsequence $(\phi_{k_j})_j$ such that $\phi_{k_j} \rightarrow \psi$ locally uniformly, and ψ is another convex function. This implies $\nabla\phi_{k_j} \rightarrow \nabla\psi$ as distributions, but one has $\nabla\phi_k \rightarrow \nabla\phi$ locally in measure (thus in L^1_{loc} and as distributions); it follows that $\nabla\phi = \nabla\psi$ a.e., which yields $\phi = \psi$ (the difference is locally Lipschitz, with gradient equal to zero a.e.; hence, it is constant). Repeating the same argument for every subsequence, we obtain $\phi_k \rightarrow \phi$ locally uniformly. It remains to prove that $\nabla\phi_k$ are locally uniformly bounded.

Let $K \Subset \mathbb{R}^n$ and $K_\delta := \{x : d(x, K) \leq \delta\}$; suppose to the contrary that there exists a sequence $(x_j)_j \subset K$ and a subsequence $(\nabla\phi_{k_j})_j$ such that $|\nabla\phi_{k_j}(x_j)| \rightarrow \infty$. Consider the cone with vertex at x_j pointing in the direction of $\nabla\phi_{k_j}(x_j)$

$$\Gamma_j := \{x \in \mathbb{R}^n : \angle(x - x_j, \nabla\phi_{k_j}(x_j)) \leq \pi/4\}.$$

By the monotonicity of the gradients, we have

$$\langle \nabla\phi_{k_j}(x) - \nabla\phi_{k_j}(x_j), x - x_j \rangle \geq 0,$$

hence, for all $x \in \Gamma_j$, one has

$$\begin{aligned} |\nabla\phi_{k_j}(x)| |x - x_j| &\geq \langle \nabla\phi_{k_j}(x), x - x_j \rangle \\ &\geq \langle \nabla\phi_{k_j}(x_j), x - x_j \rangle \\ &\geq |\nabla\phi_{k_j}(x_j)| |x - x_j| \cos(\pi/4), \end{aligned}$$

which implies

$$|\nabla\phi_{k_j}(x)| \geq |\nabla\phi_{k_j}(x_j)| \cos(\pi/4) \quad \forall x \in \Gamma_j.$$

Notice that $|\Gamma_j \cap K_\delta| \geq \eta > 0$ where η depends only on δ ; therefore,

$$\mathcal{L}^n(\{|\nabla\phi_{k_j}| \geq M\} \cap K_\delta) \geq |\Gamma_j \cap K_\delta| \geq \eta \quad \text{for } j \text{ large enough,}$$

in contradiction with the convergence in measure on K_δ . □

Chapter 3

Global estimates on the Brenier map

In [7], Caffarelli built Lipschitz changes of variables between a Gaussian measure and a log-concave perturbation. More precisely, he proved the following result:

Theorem (Caffarelli). *Let γ be a Gaussian measure in $\mathbb{R}^n$, and $d\nu = e^{-F} d\gamma$ be a probability measure with F convex. Then the Brenier map from γ to ν is 1-Lipschitz.*

This theorem allows one to deduce several functional inequalities (such as log-Sobolev or Poincaré-type inequalities) in log-concave perturbed settings, starting from their Gaussian versions. For example, denote by γ_n the standard Gaussian measure on $\mathbb{R}^n$; then the Gaussian log-Sobolev inequality reads as

$$\int_{\mathbb{R}^n} f^2 \log f^2 d\gamma_n \leq 2 \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n + \left(\int_{\mathbb{R}^n} f^2 d\gamma_n \right) \log \left(\int_{\mathbb{R}^n} f^2 d\gamma_n \right),$$

for $f \in W^{1,2}(\mathbb{R}^n)$. For a log-concave perturbation ν , let T be the Brenier map from γ_n to ν . Then, applying the change of variables formula, we get

$$\begin{aligned} \int_{\mathbb{R}^n} f^2 \log f^2 d\nu &= \int_{\mathbb{R}^n} f(T)^2 \log f(T)^2 d\gamma_n \\ &\leq 2 \int_{\mathbb{R}^n} |\nabla[f(T)]|^2 d\gamma_n + \left(\int_{\mathbb{R}^n} f(T)^2 d\gamma_n \right) \log \left(\int_{\mathbb{R}^n} f(T)^2 d\gamma_n \right) \\ &\leq 2 \|\nabla T\|_{L^\infty}^2 \int_{\mathbb{R}^n} |\nabla f(T)|^2 d\gamma_n + \left(\int_{\mathbb{R}^n} f(T)^2 d\gamma_n \right) \log \left(\int_{\mathbb{R}^n} f(T)^2 d\gamma_n \right) \\ &= 2 \|\nabla T\|_{L^\infty}^2 \int_{\mathbb{R}^n} |\nabla f|^2 d\nu + \left(\int_{\mathbb{R}^n} f^2 d\nu \right) \log \left(\int_{\mathbb{R}^n} f^2 d\nu \right). \end{aligned}$$

Hence, by Caffarelli's theorem, ν satisfies a log-Sobolev inequality with an optimal constant less than or equal to two. Similar methods have been used to transfer isoperimetric inequalities, to obtain correlation inequalities and more (see [10,11,21,22]).

3.1 Caffarelli's contraction theorem

We will prove a more general formulation of Caffarelli's theorem, given in [9,23]:

Theorem 3.1.1. Let $d\mu = e^{-V} dx$, $d\nu = e^{-W} dx$ be probability measures on $\mathbb{R}^n$ with W convex and $V, W, W^* \in C_{\text{loc}}^{1,1}(\mathbb{R}^n)$. Assume that there are constants $\Lambda_V, \lambda_W > 0$ such that

$$\nabla^2 V \preceq \Lambda_V \text{Id}, \quad \nabla^2 W^* \preceq \frac{1}{\lambda_W} \text{Id} \quad \text{a.e. in } \mathbb{R}^n.$$

Let $T = \nabla\phi$ be the Brenier map transporting μ to ν . Then,

$$\|\nabla^2\phi\|_{L^\infty} \leq \sqrt{\Lambda_V/\lambda_W}.$$

Remark 3.1.2. The hypotheses on W are more commonly $W \in C_{\text{loc}}^{1,1}$ and

$$\nabla^2 W \succeq \lambda_W \text{Id} \quad \text{a.e. in } \mathbb{R}^n.$$

The reason we choose to express the estimate on W^* instead of W will be more evident in the next sections; for now, we limit ourselves to proving that the hypotheses on W imply those on W^* . Indeed, the differentiability of W^* follows from the strict convexity of W [26, Theorem 26.3], while the uniform convexity of W yields a Lipschitz estimate on ∇W^* . Moreover, let $\mathcal{N}$ be the Lebesgue-negligible set of points x such that either W is not twice differentiable at x , or $\nabla^2 W(x) \not\succeq \lambda_W \text{Id}$. Then $\nabla W(\mathcal{N})$ is Lebesgue-negligible, and at almost every point $y \in \mathbb{R}^n \setminus \nabla W(\mathcal{N})$, one has $\nabla W^*(y) \in \mathbb{R}^n \setminus \mathcal{N}$ and

$$\nabla^2 W^*(y) = [\nabla^2 W(\nabla W^*(y))]^{-1} \preceq \frac{1}{\lambda_W} \text{Id}.$$

Proof (informal). Assume that V, W are smooth. Then, by the regularity of Monge-Ampère, ϕ is smooth and satisfies the equation

$$V = W(\nabla\phi) - \log \det \nabla^2\phi.$$

Differentiating twice in direction y yields

$$\begin{aligned} \partial_{yy}^2 V &= \langle \nabla W(\nabla\phi), \nabla \partial_{yy}^2 \phi \rangle + \langle \nabla^2 W(\nabla\phi) \nabla \partial_y \phi, \nabla \partial_y \phi \rangle \\ &\quad + \text{Tr} [(H^{-1} \partial_y H)^2] - \text{Tr}(H^{-1} \partial_{yy}^2 H) \end{aligned} \quad (3.1)$$

where $H := \nabla^2\phi$. Assume further that $\partial_{yy}^2 \phi$ attains a maximum at $(\bar{x}, \alpha) \in \mathbb{R}^n \times \mathbb{S}^{n-1}$, then the optimality conditions give

$$\begin{aligned} \partial_{\alpha\alpha}^2 \nabla\phi(\bar{x}) &= 0, \quad \partial_{\alpha\alpha}^2 H(\bar{x}) \preceq 0, \\ \partial_\beta(\partial_{yy}^2 \phi(\bar{x}))|_{y=\alpha} &= 2\partial_{\alpha\beta}^2 \phi(\bar{x}) = 0 \quad \text{for all } \beta \perp \alpha. \end{aligned} \quad (3.2)$$

Combining (3.1) and (3.2) yields

$$\Lambda_V \geq \partial_{\alpha\alpha}^2 V \geq \langle \nabla^2 W(\nabla\phi(\bar{x})) \nabla \partial_\alpha \phi, \nabla \partial_\alpha \phi \rangle \geq \lambda_W |\nabla \partial_\alpha \phi(\bar{x})|^2 = \lambda_W (\partial_{\alpha\alpha}^2 \phi(\bar{x}))^2 \quad (3.3)$$

where we used

$$\text{Tr} [(H^{-1} \partial_\alpha H)^2] = \text{Tr} [(H^{-1/2} \partial_\alpha H H^{-1/2})^T (H^{-1/2} \partial_\alpha H H^{-1/2})] \geq 0.$$

Rearranging (3.3) yields

$$\|\nabla^2\phi\|_{L^\infty} = \partial_{\alpha\alpha}^2 \phi(\bar{x}) \leq \sqrt{\Lambda_V/\lambda_W},$$

which concludes the proof. $\square$

While the regularity assumptions can be met by approximation, the main obstacle to this argument is the existence of a maximum for $\partial_{yy}^2 \phi$. The following lemma due to Caffarelli (a short proof of which we include in the Appendix) motivates the use of incremental quotients.

Lemma 3.1.3. *Let $d\mu = f dx$, $d\nu = g dx$ be probability measures on $\mathbb{R}^n$ and $T = \nabla \phi$ be the Brenier map transporting μ to ν . Assume that $\log f \in L_{\text{loc}}^\infty(\mathbb{R}^n)$, $\log g \in L^\infty(B_R)$ for some $R > 0$ and that g vanishes outside B_R . Then,*

$$T(x) \rightarrow R \frac{x}{|x|} \quad \text{uniformly as } |x| \rightarrow \infty.$$

In particular, for any fixed $\varepsilon > 0$, the function $\phi(x + \varepsilon y) + \phi(x - \varepsilon y) - 2\phi(x) \rightarrow 0$ as $|x| \rightarrow \infty$, uniformly for $y \in \mathbb{S}^{n-1}$.

Definition 3.1.4. Let $y \in \mathbb{R}^n$, and define the increments:

$$(\delta_y f)(x) := f(x + y) - f(x), \quad (\delta_y^2 f)(x) := f(x + y) + f(x - y) - 2f(x).$$

We leave the proof of some approximation properties of $\delta_{\varepsilon y}^2$ in the Appendix.

Lemma 3.1.5. *Let $\varepsilon > 0$, $y \in \mathbb{S}^{n-1}$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Denote by $\partial_{yy}^2 f$ the distributional second derivative along y . Then,*

$$\lim_{\varepsilon \rightarrow 0} \frac{\delta_{\varepsilon y}^2 f}{\varepsilon^2} = \partial_{yy}^2 f \text{ in the sense of distributions.}$$

Further, if $\partial_{yy}^2 f \leq \ell$, then

$$\delta_{\varepsilon y}^2 f \leq \ell \varepsilon^2, \quad \forall \varepsilon > 0.$$

Proof of Theorem 3.1.1. By approximation, we may assume that W is equal to infinity outside a ball B_R for some $R > 0$. Indeed, if we define

$$W_j(x) := \begin{cases} W(x) + c_j & \text{in } B_j \\ +\infty & \text{in } \mathbb{R}^n \setminus B_j, \end{cases}$$

we have

$$e^{c_j} = \int_{B_j} e^{-W} dx \rightarrow 1.$$

Hence, $e^{-W_j} \rightarrow e^{-W}$ locally uniformly, and we may apply Theorem 2.4.1 to pass the result to the limit. By the regularity of Monge-Ampère, ϕ satisfies the equation

$$V = W(\nabla \phi) - \log \det H.$$

Taking the increments $\delta_{\varepsilon y}^2$ on both sides yields

$$\varepsilon^2 \Lambda_V \geq \delta_{\varepsilon y}^2 V = \delta_{\varepsilon y}^2 (W(\nabla \phi)) - \delta_{\varepsilon y}^2 (\log \det H). \quad (3.4)$$

By Lemma 3.1.3, $0 \leq \delta_{\varepsilon y}^2 \phi(x) \rightarrow 0$ uniformly as $|x| \rightarrow \infty$, therefore $\delta_{\varepsilon y}^2 \phi$ attains a global maximum at some $(\bar{x}, \alpha) \in \mathbb{R}^n \times \mathbb{S}^{n-1}$, where the optimality conditions read

$$\begin{aligned} \delta_{\varepsilon \alpha}^2 \nabla \phi(\bar{x}) &= 0, \quad \delta_{\varepsilon \alpha}^2 H(\bar{x}) \preceq 0, \\ \partial_\beta \delta_{\varepsilon \alpha}^2 \phi(\bar{x}) &= \langle \nabla \phi(\bar{x} + \varepsilon \alpha) - \nabla \phi(\bar{x} - \varepsilon \alpha), \beta \rangle = 0 \quad \text{for all } \beta \perp \alpha. \end{aligned} \quad (3.5)$$

In particular, (3.5) implies

$$\begin{aligned} \nabla \phi(\bar{x} \pm \varepsilon \alpha) &= \nabla \phi(\bar{x}) \pm \eta \alpha \quad \text{for some } \eta > 0, \\ \delta_{\varepsilon \alpha}^2 (W(\nabla \phi)) &= (\delta_{\eta \alpha}^2 W)(\nabla \phi) \geq \eta^2 \lambda_W. \end{aligned} \quad (3.6)$$

For the other term on the right-hand side, the concavity of $\log \det$ implies

$$\begin{aligned} \log \det H(\bar{x} \pm \varepsilon \alpha) &\leq \log \det H(\bar{x}) + \langle H(\bar{x})^{-1}, H(\bar{x} \pm \varepsilon \alpha) - H(\bar{x}) \rangle, \\ \delta_{\varepsilon \alpha}^2 (\log \det H)(\bar{x}) &\leq \langle H(\bar{x})^{-1}, \delta_{\varepsilon \alpha}^2 H(\bar{x}) \rangle \leq 0. \end{aligned} \quad (3.7)$$

Combining (3.4), (3.6) and (3.7) yields

$$\eta \leq \varepsilon \sqrt{\Lambda_V / \lambda_W} =: \varepsilon C. \quad (3.8)$$

To relate this quantity to $\delta_{\varepsilon \alpha}^2 \phi(\bar{x})$, notice that the convexity of ϕ yields

$$\begin{aligned} \phi(\bar{x}) &\geq \phi(\bar{x} \pm \varepsilon \alpha) \mp \varepsilon \langle \nabla \phi(\bar{x} \pm \varepsilon \alpha), \alpha \rangle, \\ \delta_{\varepsilon \alpha}^2 \phi(\bar{x}) &\leq \varepsilon \langle \nabla \phi(\bar{x} + \varepsilon \alpha) - \nabla \phi(\bar{x} - \varepsilon \alpha), \alpha \rangle = 2\varepsilon \eta. \end{aligned} \quad (3.9)$$

In particular, from (3.8) and (3.9) we have

$$\delta_{\varepsilon \alpha}^2 \phi(\bar{x}) \leq 2\varepsilon^2 C \implies \|\nabla^2 \phi\|_{L^\infty} \leq 2C. \quad (3.10)$$

This first estimate is the desired one up to a factor of 2; we use a bootstrap argument to remove this factor. Suppose we have obtained $\|\nabla^2 \phi\|_{L^\infty} \leq aC$ for some $1 < a \leq 2$, then

$$\delta_{\varepsilon \alpha}^2 \phi(\bar{x}) = \int_0^\varepsilon \langle \nabla \phi(\bar{x} + t\alpha) - \nabla \phi(\bar{x} - t\alpha), \alpha \rangle dt.$$

By the convexity of ϕ , we have

$$\langle \nabla \phi(\bar{x} + t\alpha) - \nabla \phi(\bar{x} - t\alpha), \alpha \rangle \leq \langle \nabla \phi(\bar{x} + \varepsilon \alpha) - \nabla \phi(\bar{x} - \varepsilon \alpha), \alpha \rangle \leq 2\varepsilon C$$

but also

$$\langle \nabla \phi(\bar{x} + t\alpha) - \nabla \phi(\bar{x} - t\alpha), \alpha \rangle = \int_0^t \langle \alpha, (H(\bar{x} + h\alpha) + H(\bar{x} - h\alpha)) \alpha \rangle dh \leq 2aCt$$

for all $0 \leq t \leq \varepsilon$. Using the first upper bound for $t \geq \varepsilon/a$ and the second for $t < \varepsilon/a$ we derive

$$\delta_{\varepsilon \alpha}^2 \phi(\bar{x}) \leq 2C \left[\int_0^{\varepsilon/a} at dt + \int_{\varepsilon/a}^\varepsilon \varepsilon dt \right] = \varepsilon^2 C(2 - 1/a),$$

and by Lemma 3.1.5,

$$\|\nabla^2 \phi\|_{L^\infty} \leq C(2 - 1/a).$$

Starting from $a_0 := 2$ and repeating the same procedure with $a_{n+1} := 2 - 1/a_n$ an infinite number of times, we conclude the proof as a_n decreases to 1. $\square$

3.2 The divergence of the Brenier map

A more recent Caffarelli-type result was found at the level of the Laplacians:

Theorem 3.2.1. *Let $d\mu = e^{-V} dx$, $d\nu = e^{-W} dx$ be probability measures on $\mathbb{R}^n$ with W convex and $V, W, W^* \in C_{\text{loc}}^{1,1}(\mathbb{R}^n)$. Assume that there are constants $\Lambda_V, \lambda_W > 0$ such that*

$$\Delta V \leq \Lambda_V, \quad \Delta W^* \leq 1/\lambda_W \quad \text{a.e. in } \mathbb{R}^n.$$

Let $T = \nabla\phi$ be the Brenier map transporting μ to ν . Then,

$$\|\Delta\phi\|_{L^\infty} \leq \sqrt{\Lambda_V/\lambda_W}. \quad (3.11)$$

More precisely, De Philippis and Shenfeld [16] used the same L^p approach of Kolesnikov [23] to prove Theorem 3.2.1 under the stronger assumption $\nabla^2 W \succeq n\lambda_W \text{Id}$. This has been weakened by Gozlan and Sylvestre [20] using a different method based on entropic regularization. In this section we give a different proof of Theorem 3.2.1, following a strategy analogous to Caffarelli's: first we show how the desired estimate follows formally from a maximum principle argument applied to the Monge-Ampère equation, then we replicate the procedure in a rigorous way using incremental quotient approximations to the Laplacian.

Proof (informal). Assume V, W, ϕ are smooth, then ϕ satisfies the (log-)Monge-Ampère equation

$$V = W(\nabla\phi) - \log \det \nabla^2 \phi.$$

Differentiating twice in direction e_i yields

$$\partial_{ii} V = \langle \nabla W(\nabla\phi), \nabla \partial_{ii} \phi \rangle + \langle \nabla^2 W(\nabla\phi) \nabla \partial_i \phi, \nabla \partial_i \phi \rangle + \text{Tr}[(H^{-1} \partial_i H)^2] - \text{Tr}(H^{-1} \partial_{ii} H)$$

where $H := \nabla^2 \phi$. By summing over the basis $\{e_i\}$ we get

$$\begin{aligned} \Delta V &= \langle \nabla W(\nabla\phi), \nabla \Delta\phi \rangle + \text{Tr}(H \nabla^2 W(\nabla\phi) H) + \sum_i \text{Tr}[(H^{-1} \partial_i H)^2] - \text{Tr}(H^{-1} \nabla^2 \Delta\phi) \\ &\geq \langle \nabla W(\nabla\phi), \nabla \Delta\phi \rangle + \text{Tr}(H \nabla^2 W(\nabla\phi) H) - \text{Tr}(H^{-1} \nabla^2 \Delta\phi). \end{aligned} \quad (3.12)$$

where we used

$$\text{Tr}[(H^{-1} \partial_i H)^2] = \text{Tr}[(H^{-1/2} \partial_i H H^{-1/2})^T (H^{-1/2} \partial_i H H^{-1/2})] \geq 0.$$

Assume further that $\Delta\phi$ attains a maximum at x_0 , then the optimality conditions give $\nabla \Delta\phi = 0$, $\nabla^2 \Delta\phi \preceq 0$ at x_0 . Hence (3.12) implies

$$\Lambda_V \geq \Delta V \geq \text{Tr}(H \nabla^2 W(\nabla\phi) H) \geq \frac{\text{Tr}(H)^2}{\text{Tr}[\nabla^2 W(\nabla\phi)^{-1}]} \geq \lambda_W (\Delta\phi)^2 \quad \text{at } x_0 \quad (3.13)$$

where we used $\nabla^2 W(x)^{-1} = \nabla^2 W^*(y)$ and the Cauchy-Schwarz inequality for the Frobenius scalar product $\langle A, B \rangle := \text{Tr}(B^T A)$:

$$\text{Tr}(H X H) \text{Tr}(X^{-1}) = \langle H X^{1/2}, H X^{1/2} \rangle \langle X^{-1/2}, X^{-1/2} \rangle \geq \langle H X^{1/2}, X^{-1/2} \rangle^2 = \text{Tr}(H)^2. \quad (3.14)$$

Rearranging (3.13) yields

$$\|\Delta\phi\|_{L^\infty} = \Delta\phi(x_0) \leq \sqrt{\Lambda_V/\lambda_W},$$

which concludes the proof. $\square$

Similarly to Caffarelli's theorem, the main obstacle to this formal argument is the existence of a maximum for $\Delta\phi$. Again, we will bypass this issue with ε -approximations to the Laplacian.

Definition 3.2.2. Let $\varepsilon > 0$, define the increment:

$$(\Delta_\varepsilon f)(x) := \int_{\mathbb{S}^{n-1}} (\delta_{\varepsilon y} f)(x) \, dy = \frac{1}{2} \int_{\mathbb{S}^{n-1}} (\delta_{\varepsilon y}^2 f)(x) \, dy.$$

Recall the definition of δ_y , δ_y^2 :

$$(\delta_y f)(x) := f(x+y) - f(x), \quad (\delta_y^2 f)(x) := f(x+y) + f(x-y) - 2f(x).$$

We leave the proof of some approximation properties of Δ_ε in the Appendix.

Lemma 3.2.3. Let $\varepsilon > 0$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}$, denote by Δf its distributional Laplacian. Then,

$$\lim_{\varepsilon \rightarrow 0} \frac{\Delta_\varepsilon f}{\varepsilon^2} = \frac{\Delta f}{2n} \text{ in the sense of distributions.}$$

Further, if $\Delta f \leq \ell$, then

$$\Delta_\varepsilon f \leq \frac{\ell \varepsilon^2}{2n}, \quad \forall \varepsilon > 0.$$

Using Theorem 2.4.1, we may work with regularized measures such that ν is supported in a ball of radius R . We include the proof of this step in the Appendix.

Lemma 3.2.4. By approximation, we may assume that V is smooth and $V \geq C_1 + C_2|x|^2$ for some $C_2 > 0$, that W is equal to infinity outside the ball $\overline{B_R}$ and that W is smooth in $\overline{B_R}$ for some $R > 0$.

In this setting we are ready for the proof of Theorem 3.2.1:

Proof. By Lemma 3.2.4 and regularity of the optimal transport, the map $T = \nabla\phi$ is smooth and satisfies the (log-)Monge-Ampère equation

$$V = W(\nabla\phi) - \log \det H.$$

Taking the increments $2\Delta_\varepsilon$ on both sides yields

$$\frac{\varepsilon^2}{n} \Lambda_V \geq 2\Delta_\varepsilon V = 2\Delta_\varepsilon(W(\nabla\phi)) - 2\Delta_\varepsilon(\log \det H). \quad (3.15)$$

By Lemma 3.1.3, $0 \leq \Delta_\varepsilon\phi(x) \rightarrow 0$ uniformly as $|x| \rightarrow \infty$, therefore $\Delta_\varepsilon\phi$ attains a global maximum at some $\bar{x} \in \mathbb{R}^n$ where the optimality conditions read $\Delta_\varepsilon\nabla\phi(\bar{x}) = 0$, $\Delta_\varepsilon H(\bar{x}) \preceq 0$. We estimate each term on the right-hand side: the concavity of $\log \det$ implies

$$\begin{aligned} \delta_{\varepsilon y}(\log \det H)(\bar{x}) &\leq \langle H(\bar{x})^{-1}, \delta_{\varepsilon y} H(\bar{x}) \rangle, \\ \Delta_\varepsilon(\log \det H)(\bar{x}) &\leq \langle H(\bar{x})^{-1}, \Delta_\varepsilon H(\bar{x}) \rangle \leq 0. \end{aligned} \quad (3.16)$$

The following lemma deals with the term $W(\nabla\phi)$:

Lemma 3.2.5. *With the previous notation, $\forall \eta > 0$, $\exists \bar{\varepsilon} > 0$ such that for all $0 < \varepsilon \leq \bar{\varepsilon}$, the following estimate holds:*

$$2\Delta_\varepsilon(W(\nabla\phi))(\bar{x}) \geq (1 - \eta) \int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), \nabla^2 W(\nabla\phi(\bar{x})) \delta_{\varepsilon y} \nabla\phi(\bar{x}) \rangle dy. \quad (3.17)$$

Proof. We can write

$$2\delta_{\varepsilon y}(W(\nabla\phi))(\bar{x}) = 2\langle \nabla W(\nabla\phi(\bar{x})), \delta_{\varepsilon y} \nabla\phi(\bar{x}) \rangle + \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), X \delta_{\varepsilon y} \nabla\phi(\bar{x}) \rangle + R_\varepsilon(y) \quad (3.18)$$

where $X := \nabla^2 W(\nabla\phi(\bar{x}))$, and estimate the remainder $|R_\varepsilon(y)| \leq K|\delta_{\varepsilon y} \nabla\phi(\bar{x})|^3$ as $\nabla^3 W$ is bounded. Now if our point of maximum $\bar{x}$ stays in a bounded ball as $\varepsilon \rightarrow 0$, then $|\delta_{\varepsilon y} \nabla\phi(\bar{x})| \lesssim \varepsilon$. If instead (up to a subsequence) $|\bar{x}| \rightarrow \infty$, then by Lemma 3.1.3 $|\delta_{\varepsilon y} \nabla\phi(\bar{x})| \rightarrow 0$ uniformly for $y \in \mathbb{S}^{n-1}$. Either way $\forall \eta > 0$, for ε small enough we obtain

$$|R_\varepsilon(y)| \leq K|\delta_{\varepsilon y} \nabla\phi(\bar{x})|^3 \leq K\eta|\delta_{\varepsilon y} \nabla\phi(\bar{x})|^2 \leq \frac{K\eta}{\lambda_W} \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), X \delta_{\varepsilon y} \nabla\phi(\bar{x}) \rangle.$$

Up to renormalizing η we can integrate (3.18) to obtain the desired estimate

$$2\Delta_\varepsilon(W(\nabla\phi))(\bar{x}) \geq (1 - \eta) \int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), X \delta_{\varepsilon y} \nabla\phi(\bar{x}) \rangle dy.$$

□

Combining (3.15), (3.16) and (3.17) yields

$$\frac{\varepsilon^2}{n} \frac{\Lambda_V}{1 - \eta} \geq \int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), X \delta_{\varepsilon y} \nabla\phi(\bar{x}) \rangle dy. \quad (3.19)$$

Now define $F(y) := X^{1/2} \delta_{\varepsilon y} \nabla\phi(\bar{x})$ and $G(y) := X^{-1/2} y$, then we have the following Cauchy-Schwarz inequality analogous to (3.14):

$$\begin{aligned} \left(\int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), y \rangle dy \right)^2 &= \left(\int_{\mathbb{S}^{n-1}} \langle F(y), G(y) \rangle dy \right)^2 \leq \int_{\mathbb{S}^{n-1}} |F|^2 dy \int_{\mathbb{S}^{n-1}} |G|^2 dy \\ &= \int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), X \delta_{\varepsilon y} \nabla\phi(\bar{x}) \rangle dy \int_{\mathbb{S}^{n-1}} \langle y, X^{-1} y \rangle dy. \end{aligned} \quad (3.20)$$

By combining (3.19), (3.20) and using

$$\int_{\mathbb{S}^{n-1}} \langle y, X^{-1} y \rangle dy = \frac{\text{Tr}(X^{-1})}{n} \leq \frac{1}{n\lambda_W}$$

we obtain the estimate:

$$\int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla\phi(\bar{x}), y \rangle dy \leq \frac{\varepsilon}{n} \frac{C}{\sqrt{1 - \eta}} \quad \text{where } C := \sqrt{\Lambda_V / \lambda_W}. \quad (3.21)$$

To relate this quantity to $\Delta_\varepsilon \phi(\bar{x})$, notice that the convexity of ϕ yields

$$\phi(\bar{x}) \geq \phi(\bar{x} \pm \varepsilon y) \mp \varepsilon \langle \nabla \phi(\bar{x} \pm \varepsilon y), y \rangle$$

and

$$\delta_{\varepsilon y}^2 \phi(\bar{x}) \leq \varepsilon \langle \nabla \phi(\bar{x} + \varepsilon y) - \nabla \phi(\bar{x} - \varepsilon y), y \rangle,$$

which integrated gives

$$\Delta_\varepsilon \phi(\bar{x}) \leq \frac{\varepsilon}{2} \int_{\mathbb{S}^{n-1}} \langle \nabla \phi(\bar{x} + \varepsilon y) - \nabla \phi(\bar{x} - \varepsilon y), y \rangle dy = \varepsilon \int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla \phi(\bar{x}), y \rangle dy. \quad (3.22)$$

In particular, from (3.21) and (3.22) we have

$$\Delta_\varepsilon \phi(\bar{x}) \leq \frac{\varepsilon^2}{2n} \frac{2C}{\sqrt{1-\eta}} \implies \|\Delta \phi\|_{L^\infty} \leq \frac{2C}{\sqrt{1-\eta}}. \quad (3.23)$$

This first estimate is the desired one up to a factor of 2; to remove this factor, we use a bootstrap argument similar to Caffarelli's. Suppose we have obtained $\Delta \phi \leq \frac{aC}{\sqrt{1-\eta}}$ for some $1 < a \leq 2$, then

$$\delta_{\varepsilon y}^2 \phi(\bar{x}) = \int_0^\varepsilon \langle \nabla \phi(\bar{x} + ty) - \nabla \phi(\bar{x} - ty), y \rangle dt \implies \Delta_\varepsilon \phi(\bar{x}) = \int_{\mathbb{S}^{n-1}} \int_0^\varepsilon \langle \delta_{ty} \nabla \phi(\bar{x}), y \rangle dt dy.$$

By the convexity of ϕ , we have $\langle \delta_{ty} \nabla \phi(\bar{x}), y \rangle \leq \langle \delta_{\varepsilon y} \nabla \phi(\bar{x}), y \rangle$ for all $0 \leq t \leq \varepsilon$ thus

$$\int_{\mathbb{S}^{n-1}} \langle \delta_{ty} \nabla \phi(\bar{x}), y \rangle dy \leq \int_{\mathbb{S}^{n-1}} \langle \delta_{\varepsilon y} \nabla \phi(\bar{x}), y \rangle dy \leq \frac{\varepsilon}{n} \frac{C}{\sqrt{1-\eta}} \quad \text{by (3.21)}$$

but also

$$\int_{\mathbb{S}^{n-1}} \langle \delta_{ty} \nabla \phi(\bar{x}), y \rangle dy = \frac{t}{n} \int_{B_1} \Delta \phi(\bar{x} + tz) dz \leq \frac{t}{n} \frac{aC}{\sqrt{1-\eta}}.$$

Using the first upper bound for $t \geq \varepsilon/a$ and the second for $t < \varepsilon/a$ we derive

$$\Delta_\varepsilon \phi(\bar{x}) \leq \frac{C}{n\sqrt{1-\eta}} \left[\int_0^{\varepsilon/a} at dt + \int_{\varepsilon/a}^\varepsilon \varepsilon dt \right] = \frac{\varepsilon^2}{2n} \frac{C(2-1/a)}{\sqrt{1-\eta}}$$

and by Lemma 3.2.3,

$$\|\Delta \phi\|_{L^\infty} \leq \frac{C(2-1/a)}{\sqrt{1-\eta}}.$$

Starting from $a_0 := 2$ and repeating the same procedure with $a_{n+1} := 2 - 1/a_n$ an infinite number of times, then letting $\eta \rightarrow 0$ we conclude the proof, as a_n decreases to 1. $\square$

3.3 Good estimates

Both Caffarelli's and Gozlan-Sylvestre's theorems are examples of how upper bounds for $f(\nabla^2 V)$ and $f(\nabla^2 W^*)$ provide upper bounds for $f(\nabla^2 \phi)$, for some estimating function f equal to $\lambda_{\max}$, Tr respectively. In this section we use elementary convex analysis to generalize this result to a broader class of functions $\mathbf{G}_n$.

Definition 3.3.1. Denote by $\mathbb{S}_n$ the set of $n \times n$ symmetric real matrices; a continuous function $f: \mathbb{S}_n \rightarrow \mathbb{R}$ is called *good* if it is

- convex,
- increasing in $\mathbb{S}_n^+$: $0 \preceq X \preceq X' \implies f(X) \leq f(X')$, and $0 \not\preceq X \implies f(X) > 0$,
- positively homogeneous: $f(tX) = tf(X) \forall t > 0$,

and we denote by $\mathbf{G}_n$ the set of *good* functions.

Theorem 3.3.2. Let $f \in \mathbf{G}_n$, and $d\mu = e^{-V} dx$, $d\nu = e^{-W} dx$ be probability measures on $\mathbb{R}^n$ with W convex and $V, W, W^* \in C_{\text{loc}}^{1,1}(\mathbb{R}^n)$. Assume that there are constants $\Lambda_V, \lambda_W > 0$ such that

$$f(\nabla^2 V) \leq \Lambda_V, \quad f(\nabla^2 W^*) \leq 1/\lambda_W \quad \text{a.e. in } \mathbb{R}^n.$$

Let $T = \nabla\phi$ be the Brenier map transporting μ to ν . Then,

$$\|f(\nabla^2\phi)\|_{L^\infty} \leq \sqrt{\Lambda_V/\lambda_W}.$$

The previous bound is sharp as can be seen by taking

$$\mu = \mathcal{N}(0, \alpha^2 \text{Id}), \quad \nu = \mathcal{N}(0, \beta^2 \text{Id}); \quad \Lambda_V = \frac{f(\text{Id})}{\alpha^2}, \quad 1/\lambda_W = \beta^2 f(\text{Id})$$

as in this case $\nabla\phi(x) = \frac{\beta x}{\alpha}$, which yields $f(\nabla^2\phi) = \frac{\beta}{\alpha} f(\text{Id})$.

Remark. By the characterization of convex unitarily invariant functions on $\mathbb{S}_n$ [13], if $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and symmetric, then $f(X) := g(\lambda(X))$ is convex, which lets us build some interesting examples of *good* functions: denote by $\lambda_1 \geq \dots \geq \lambda_n$ the eigenvalues of X and by λ^+ their positive parts, then

$$S_k(X) := \lambda_1 + \dots + \lambda_k, \quad H_k(X)^{1/k} := \left(\prod_{1 \leq i_1 \leq \dots \leq i_k \leq n} \lambda_{i_1}^+ \dots \lambda_{i_k}^+ \right)^{1/k}, \quad N_p(X) := \|\lambda^+\|_p \quad p \geq 1$$

are *good*, as well as all norms increasing on $\mathbb{S}_n^+$. In particular, for $S_1 = \lambda_{\max}$ and $S_n = \text{Tr}$ we recover Caffarelli's and Gozlan-Sylvestre's theorems respectively. Although most useful estimates involve unitarily invariant functions, we do not require this property.

Denote by $\mathbb{S}_n^+$ the set of $n \times n$ positive semidefinite symmetric real matrices, and by $\mathbb{S}_n^{++}$ the positive definite ones. Our main observations are that a change of variables given by a matrix $A \in \mathbb{S}_n^{++}$ gives us an anisotropic version of the Laplacian estimate, and that the sup of a (countable) family of valid estimates continues to be a valid estimate.

Lemma 3.3.3. Let $Y \in \mathbb{S}_n^{++}$, then the statement of Theorem 3.3.2 holds for $f(X) = \langle X, Y \rangle$.

Proof. Let $T = \nabla\phi$ be the Brenier map transporting μ to ν , and $A := Y^{1/2}$. Define the following:

$$\begin{aligned} \phi_A(x) &:= \phi(Ax), \\ V_A(x) &:= V(Ax) - \log \det A, \\ W_A(x) &:= W(A^{-1}x) + \log \det A. \end{aligned}$$

Then the probability measures $d\mu_A := e^{-V_A} dx$, $d\nu_A := e^{-W_A} dx$ satisfy the Laplacian estimates:

$$\Delta V_A(x) = \langle \nabla^2 V(Ax), Y \rangle \leq \Lambda_V, \quad \Delta W_A^*(y) = \langle \nabla^2 W^*(Ay), Y \rangle \leq 1/\lambda_W \quad \text{a.e. in } \mathbb{R}^n.$$

We can easily verify that ϕ_A is convex and $(\nabla \phi_A)_\# \mu_A = \nu_A$, therefore $\nabla \phi_A$ is the Brenier map and Theorem 3.2.1 yields

$$\|\langle \nabla^2 \phi, Y \rangle\|_{L^\infty} = \|\Delta \phi_A\|_{L^\infty} \leq \sqrt{\Lambda_V/\lambda_W}.$$

□

Proof of Theorem 3.3.2. Denote by

$$\partial f(X) := \{Y \in \mathbb{S}_n : f(Z) \geq f(X) + \langle Z - X, Y \rangle \quad \forall Z \in \mathbb{S}_n\}$$

the subdifferential of f at X . By convexity and homogeneity [26, Theorem 13.2], we can write

$$f(X) = \sup_{Y \in \partial f(0)} \langle X, Y \rangle \quad \text{with equality for } Y \in \partial f(X) \subset \partial f(0).$$

Observe that for any $X \in \mathbb{S}_n^{++}$, the subdifferential satisfies $\partial f(X) \subset \mathbb{S}_n^+$. Indeed, for any $Y \in \partial f(X)$ and $v \in \mathbb{R}^n$, choose $\alpha > 0$ such that $0 \preceq X - \alpha v \otimes v \preceq X$. Then,

$$\langle X - \alpha v \otimes v, Y \rangle \leq f(X - \alpha v \otimes v) \leq f(X) = \langle X, Y \rangle,$$

which implies $\langle v, Yv \rangle \geq 0$. By the monotonicity of f ,

$$\beta := \min\{f(X) : X \succeq 0, \text{Tr}(X) = 1\} > 0,$$

hence for any $X \succeq 0$, the estimate $f(X) \geq \beta \text{Tr}(X)$ holds. Our next goal is to obtain a positive definite matrix $E \in \partial f(0)$, and we will do so in two steps:

1. first we prove that there is no non-zero vector v in the intersection of $\ker Y$ for all $0 \preceq Y \in \partial f(0)$;
2. second, we iteratively construct a positive definite matrix $E \in \partial f(0)$.

Step 1. Suppose that there exists a vector $v \neq 0$ such that

$$Yv = 0 \quad \text{for all } 0 \preceq Y \in \partial f(0).$$

Then by writing

$$f(X) = \sup_{0 \preceq Y \in \partial f(0)} \langle X, Y \rangle \quad \text{for any } X \succ 0,$$

we obtain

$$f(X) = f(X + \alpha v \otimes v) \geq \beta \text{Tr}(X) + \alpha \beta |v|^2 \quad \forall \alpha > 0$$

which yields a contradiction as $\alpha \rightarrow +\infty$.

Step 2. Let $v \neq 0$ and denote by Y_v a matrix satisfying $Y_v v \neq 0$ and $0 \preceq Y_v \in \partial f(0)$. If $Y_1 := Y_v$ is non-singular, we choose $E := Y_1$. Otherwise, let $0 \neq v_2 \in \ker Y_1$, and define

$$0 \preceq Y_2 := \frac{Y_1 + Y_{v_2}}{2} \in \partial f(0).$$

Notice that if $Y_2 w = 0$ for some vector w , then

$$\langle w, Y_1 w \rangle + \langle w, Y_{v_2} w \rangle = |Y_1^{1/2} w|^2 + |Y_{v_2}^{1/2} w|^2 = 0,$$

which implies $Y_1 w = 0$ and $Y_{v_2} w = 0$. Hence $\ker Y_2 \not\subseteq \ker Y_1$, and repeating the same construction on Y_2 yields a positive definite matrix $E \in \partial f(0)$ after a finite number of steps.

Let $0 \preceq Y \in \partial f(0)$ and $\varepsilon > 0$, and notice that

$$\begin{aligned} \langle \nabla^2 V, Y + \varepsilon E \rangle &\leq (1 + \varepsilon) f(\nabla^2 V) \leq (1 + \varepsilon) \Lambda_V, \\ \langle \nabla^2 W^*, Y + \varepsilon E \rangle &\leq (1 + \varepsilon) f(\nabla^2 W^*) \leq (1 + \varepsilon) / \lambda_W \quad \text{a.e. in } \mathbb{R}^n. \end{aligned}$$

Hence, applying Lemma 3.3.3 to $Y + \varepsilon E \in \mathbb{S}_n^{++}$ yields

$$\|\langle \nabla^2 \phi, Y \rangle\|_{L^\infty} \leq (1 + \varepsilon) \sqrt{\Lambda_V / \lambda_W}.$$

Taking Y in a countable dense subset of $\partial f(0) \cap \mathbb{S}_n^+$, we obtain

$$\langle \nabla^2 \phi(x), Y \rangle \leq \sqrt{\Lambda_V / \lambda_W} \quad \text{for all } Y \in \partial f(0) \cap \mathbb{S}_n^+, \quad x\text{-a.e.}$$

In particular, up to a Lebesgue-negligible set we have

$$\begin{aligned} f(\nabla^2 \phi(x)) &\leq f(\nabla^2 \phi(x) + \varepsilon \text{Id}) = \langle \nabla^2 \phi(x) + \varepsilon \text{Id}, Y_\varepsilon \rangle \\ &= \langle \nabla^2 \phi(x), Y_\varepsilon \rangle + \langle \varepsilon \text{Id}, Y_\varepsilon \rangle \leq \sqrt{\Lambda_V / \lambda_W} + \varepsilon f(\text{Id}) \end{aligned}$$

where $\varepsilon > 0$ and $0 \preceq Y_\varepsilon \in \partial f(\nabla^2 \phi(x) + \varepsilon \text{Id})$. Taking a countable sequence of $\varepsilon \rightarrow 0^+$ concludes the proof. $\square$

Appendix A

In this appendix we prove a few technical results, namely Caffarelli's lemma on the decay of the incremental quotient $\delta_{\varepsilon y}^2 \phi$, the approximation properties of the operators $\delta_{\varepsilon y}^2$, Δ_ε , and the approximation step in the proof of Theorem 3.2.1.

Lemma A.1. *Let $d\mu = f dx$, $d\nu = g dx$ be probability measures on $\mathbb{R}^n$ and $T = \nabla \phi$ be the Brenier map transporting μ to ν . Assume that $\log f \in L_{\text{loc}}^\infty(\mathbb{R}^n)$, $\log g \in L^\infty(B_R)$ for some $R > 0$ and that g vanishes outside B_R . Then,*

$$T(x) \rightarrow R \frac{x}{|x|} \quad \text{uniformly as } |x| \rightarrow \infty.$$

In particular, for any fixed $\varepsilon > 0$, the function $\phi(x + \varepsilon y) + \phi(x - \varepsilon y) - 2\phi(x) \rightarrow 0$ as $|x| \rightarrow \infty$, uniformly for $y \in \mathbb{S}^{n-1}$.

Proof. We begin by noticing that, as a consequence of Theorem 2.3.4, T is continuous on $\mathbb{R}^n$ and, in particular, the map T is well defined at every point.

Let $x_0 \in \mathbb{R}^n$ and $\theta \in (0, \pi/4)$ be fixed, and consider the cone with vertex at $T(x_0)$ and pointing in the x_0 -direction

$$\Gamma := \left\{ y \in \mathbb{R}^n : \angle(x_0, y - T(x_0)) \leq \frac{\pi}{2} - \theta \right\}.$$

By the convexity of ϕ we see that

$$\angle(x - x_0, T(x) - T(x_0)) \leq \frac{\pi}{2},$$

hence

$$\angle(x - x_0, x_0) \leq \angle(x - x_0, T(x) - T(x_0)) + \angle(x_0, T(x) - T(x_0)) \leq \pi - \theta \quad \forall x \text{ s.t. } T(x) \in \Gamma,$$

and so, up to a set of measure zero, the preimage of Γ under T is contained in the (concave) cone

$$\Omega := \{x \in \mathbb{R}^n : \angle(x_0, x - x_0) \leq \pi - \theta\}.$$

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

$$\inf_{x \in B_R} g(x) \mathcal{L}^n(\Gamma \cap B_R) \leq \nu(\Gamma \cap B_R) = \nu(\Gamma) \leq \mu(\Omega).$$

Let $B = B_{(|x_0| \tan \theta)/2}$, and notice that $\Omega \subseteq \mathbb{R}^n \setminus B$. This proves that $\mu(\Omega) \leq \mu(\mathbb{R}^n \setminus B)$.

Now, $\mu(\mathbb{R}^n \setminus B) \rightarrow 0$ as $|x_0| \rightarrow \infty$ since B covers $\mathbb{R}^n$ as $|x_0| \rightarrow \infty$. Recalling that g is bounded away from zero in B_R , we have that

$$\lim_{|x_0| \rightarrow \infty} \mathcal{L}^n(\Gamma \cap B_R) = 0.$$

Letting $\theta \rightarrow 0$, we see that $T(x_0) \rightarrow R \frac{x_0}{|x_0|}$. As the point x_0 was fixed arbitrarily,

$$\nabla \phi(x) = T(x) \rightarrow R \frac{x}{|x|} \text{ uniformly as } |x| \rightarrow \infty.$$

Thus, for any fixed $\varepsilon > 0$,

$$\begin{aligned} |\phi(x + \varepsilon y) + \phi(x - \varepsilon y) - 2\phi(x)| &= \varepsilon |\langle \nabla \phi(x + \eta y) - \nabla \phi(x - \eta y), y \rangle| \quad \text{for some } \eta \in [0, \varepsilon] \\ &\leq \varepsilon R \left[\left| \frac{x + \eta y}{|x + \eta y|} - \frac{x - \eta y}{|x - \eta y|} \right| + o(1) \right] \rightarrow 0 \text{ as } |x| \rightarrow \infty, \text{ uniformly for } y \in \mathbb{S}^{n-1}. \end{aligned}$$

□

Lemma A.2. *Let $\varepsilon > 0$, $y \in \mathbb{S}^{n-1}$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Denote by $\partial_{yy}^2 f$ the distributional second derivative along y . Then,*

$$\lim_{\varepsilon \rightarrow 0} \frac{\delta_{\varepsilon y}^2 f}{\varepsilon^2} = \partial_{yy}^2 f \text{ in the sense of distributions.}$$

Further, if $\partial_{yy}^2 f \leq \ell$, then

$$\delta_{\varepsilon y}^2 f \leq \ell \varepsilon^2, \quad \forall \varepsilon > 0. \quad (\text{A.1})$$

Proof. First notice that

$$\left\langle \frac{\delta_{\varepsilon y}^2 f}{\varepsilon^2}, \eta \right\rangle = \left\langle f, \frac{\delta_{\varepsilon y}^2 \eta}{\varepsilon^2} \right\rangle$$

for test functions η . Then the convergence follows from writing

$$\delta_{\varepsilon y}^2 \eta = \varepsilon^2 \partial_{yy}^2 \eta + O(\varepsilon^3).$$

Next we prove (A.1): we have

$$\begin{aligned} \delta_{\varepsilon y}^2 f(x) &= \int_0^\varepsilon \frac{d}{dr} [f(x + ry) + f(x - ry) - 2f(x)] dr = \int_0^\varepsilon \langle \nabla f(x + ry) - \nabla f(x - ry), y \rangle dr \\ &= 2 \int_0^\varepsilon \langle \nabla f(x + ry) - \nabla f(x), y \rangle dr = 2 \int_0^\varepsilon \int_0^r \partial_{yy}^2 f(x + sy) ds dr \leq \ell \varepsilon^2, \end{aligned}$$

where the last inequality used $\partial_{yy}^2 f \leq \ell$. □

Lemma A.3. *Let $\varepsilon > 0$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}$, denote by Δf its distributional Laplacian. Then,*

$$\lim_{\varepsilon \rightarrow 0} \frac{\Delta_\varepsilon f}{\varepsilon^2} = \frac{\Delta f}{2n} \text{ in the sense of distributions.}$$

Further, if $\Delta f \leq \ell$, then

$$\Delta_\varepsilon f \leq \frac{\ell \varepsilon^2}{2n}, \quad \forall \varepsilon > 0. \quad (\text{A.2})$$

Proof. Notice that

$$\lim_{\varepsilon \rightarrow 0} \frac{\Delta_\varepsilon f}{\varepsilon^2} = \frac{1}{2} \lim_{\varepsilon \rightarrow 0} \left[\int_{\mathbb{S}^{n-1}} \frac{\delta_{\varepsilon y}^2 f}{\varepsilon^2} dy \right] = \frac{1}{2} \int_{\mathbb{S}^{n-1}} \partial_{yy}^2 f dy.$$

For each $i, j \in \{1, \dots, n\}$, by the symmetry of $\mathbb{S}^{n-1}$ under $y_j \mapsto -y_j$,

$$\int_{\mathbb{S}^{n-1}} y_i y_j dy = \delta_{ij} \int_{\mathbb{S}^{n-1}} y_i^2 dy = \delta_{ij} \frac{1}{n}.$$

It follows that

$$\lim_{\varepsilon \rightarrow 0} \frac{\Delta_\varepsilon f}{\varepsilon^2} = \frac{1}{2} \sum_{i=1}^n \partial_{ii}^2 f \int_{\mathbb{S}^{n-1}} y_i^2 dy = \frac{1}{2n} \sum_{i=1}^n \partial_{ii}^2 f = \frac{\Delta f}{2n}.$$

Next we prove (A.2): we have

$$\Delta_\varepsilon f(x) = \int_0^\varepsilon \frac{d}{dr} \left[\int_{\mathbb{S}^{n-1}} (f(x + ry) - f(x)) dy \right] dr.$$

Integration by parts yields

$$\begin{aligned} \frac{d}{dr} \left[\int_{\mathbb{S}^{n-1}} (f(x + ry) - f(x)) dy \right] &= \int_{\mathbb{S}^{n-1}} \frac{\partial}{\partial r} (f(x + ry) - f(x)) dy = \int_{\mathbb{S}^{n-1}} \langle \nabla f(x + ry), y \rangle dy \\ &= \int_{\mathbb{S}^{n-1}} \frac{1}{r} \langle \nabla_y [f(x + ry)], y \rangle dy = \frac{1}{n} \int_{B_1} \frac{1}{r} \Delta_y [f(x + ry)] dy = \frac{r}{n} \int_{B_1} \Delta f(x + ry) dy \leq \frac{\ell r}{n}, \end{aligned}$$

where the last inequality used $\Delta f \leq \ell$. It follows that

$$\Delta_\varepsilon f(x) \leq \frac{\ell}{n} \int_0^\varepsilon r dr = \frac{\ell \varepsilon^2}{2n}.$$

□

Lemma A.4. *In the proof of Theorem 3.2.1, we may assume that V is smooth and satisfies $V \geq C_1 + C_2|x|^2$ for some $C_2 > 0$, that W is equal to infinity outside the ball $\overline{B_R}$ and that W is smooth in $\overline{B_R}$ for some $R > 0$.*

Proof. Let η be a standard convolution kernel (radially symmetric, non-negative, compactly supported), and define $e^{-V_t} := e^{-V} * \eta_t$, then we obtain

$$V_t = -\log(e^{-V} * \eta_t), \quad \nabla V_t = \mathbb{E}[\nabla V] \quad \text{where} \quad \mathbb{E}[X](x) := \frac{(X e^{-V}) * \eta_t}{e^{-V} * \eta_t}(x),$$

and by Cauchy-Schwarz inequality:

$$\nabla^2 V_t = \mathbb{E}[\nabla^2 V] - \mathbb{E}[\nabla V \otimes \nabla V] + \mathbb{E}[\nabla V] \otimes \mathbb{E}[\nabla V] \preceq \mathbb{E}[\nabla^2 V]$$

which implies $\Delta V_t \leq \mathbb{E}[\Delta V] \leq \Lambda_V$. Taking $\text{supp } \eta_t \subset B_{\delta_t}$ yields

$$e^{-V_t(x)} \lesssim \int_{B(x, \delta_t)} e^{-V} dy \rightarrow 0 \text{ as } |x| \rightarrow \infty,$$

and in particular, V_t is positive outside a bounded set. Clearly $(e^{-V_t})_{t \rightarrow 0}$ satisfies the necessary stability conditions, therefore we may assume V is smooth and positive outside a bounded set. Up to adjusting $V_t := (1-t)V + \frac{t\Lambda_V}{n} \frac{|x|^2}{2} + c_t$ we may also assume $V \geq C_1 + C_2|x|^2$ for some $C_2 > 0$. More precisely,

$$\Delta V_t = (1-t)\Delta V + t\Lambda_V \leq \Lambda_V, \quad e^{-(1-t)V - \frac{t\Lambda_V}{n} \frac{|x|^2}{2}} \rightarrow e^{-V} \text{ locally uniformly,}$$

and by dominated convergence,

$$e^{c_t} = \int_{\mathbb{R}^n} e^{-(1-t)V - \frac{t\Lambda_V}{n} \frac{|x|^2}{2}} dx \rightarrow 1$$

where we used $(1-t)V + \frac{t\Lambda_V}{n} \frac{|x|^2}{2} \geq \min\{V, \frac{\Lambda_V}{n} \frac{|x|^2}{2}\}$, which together imply $e^{-V_t} \rightarrow e^{-V}$ locally uniformly. Next if we define

$$W_t(x) := \begin{cases} (W * \eta_t)(x) + c_t & \text{in } \overline{B_{1/t}} \\ +\infty & \text{in } \mathbb{R}^n \setminus \overline{B_{1/t}}, \end{cases}$$

we obtain, by Jensen's inequality and the convexity of the map $\mathbb{S}_n^{++} \ni A \mapsto A^{-1}$ [3, p. V.1.15],

$$\nabla^2 W_t^*(y) \preceq (\nabla^2 W * \eta_t)(x)^{-1} \preceq ((\nabla^2 W)^{-1} * \eta_t)(x) \text{ for } x \in \partial W_t^*(y)$$

which yields $\Delta W_t^* \leq \text{Tr} [(\nabla^2 W)^{-1}] * \eta_t \leq 1/\lambda_W$. The estimate $\Delta W^* \leq 1/\lambda_W$ also implies $\nabla^2 W \succeq \lambda_W \text{Id}$, therefore

$$W(x) \geq W(0) + \langle \nabla W(0), x \rangle + \frac{\lambda_W}{2} |x|^2 \geq C'_1 + C'_2 |x|^2$$

for some $C'_2 > 0$. Similarly, for some $C''_2 > 0$ we have

$$W * \eta_t(x) \geq \inf_{B(x, \delta_t)} W \geq C'_1 + C'_2 (|x| - \delta_t)^2 \geq C''_1 + C''_2 |x|^2.$$

Lastly, $e^{-W * \eta_t} \rightarrow e^{-W}$ locally uniformly, and by dominated convergence we have

$$e^{c_t} = \int_{B_{1/t}} e^{-W * \eta_t} dx \rightarrow 1,$$

which together imply $e^{-W_t} \rightarrow e^{-W}$ locally uniformly. □

References

- [1] A. D. Aleksandrov. “Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it.” *Leningrad State Univ. Annals [Uchenye Zapiski] Math. Ser.*, **6**, 1939, pp. 3–35.
- [2] S. Alesker, S. Dar, and V. Milman. “A remarkable measure preserving diffeomorphism between two convex bodies in $\mathbb{R}^n$.” *Geometriae Dedicata*, **74**(2), 1999, pp. 201–212.
- [3] R. Bhatia. *Matrix analysis*. Springer Science & Business Media, 2013.
- [4] A. Bidoia. “Global estimates on the Brenier map.” 2026. URL: <https://arxiv.org/abs/2605.23731>.
- [5] L. A. Caffarelli. “Some regularity properties of solutions of Monge Ampère equation.” *Commun. Pure Appl. Math.*, **44**(8-9), 1991, pp. 965–969. ISSN: 0010-3640.
- [6] L. A. Caffarelli. “The regularity of mappings with a convex potential.” *J. Am. Math. Soc.*, **5**(1), 1992, pp. 99–104. ISSN: 0894-0347.
- [7] L. A. Caffarelli. “Monotonicity properties of optimal transportation and the FKG and related inequalities.” *Communications in Mathematical Physics*, **214**(3), 2000, pp. 547–563.
- [8] L. A. Caffarelli. “Erratum: Monotonicity properties of optimal transportation and the FKG and related inequalities.” *Communications in Mathematical Physics*, **225**(2), 2002, pp. 449–450.
- [9] M. Colombo, A. Figalli, and Y. Jhaveri. “Lipschitz changes of variables between perturbations of log-concave measures.” *Ann. Sc. Norm. Super. Pisa, Cl. Sci. (5)*, **17**(4), 2017, pp. 1491–1519. ISSN: 0391-173X.
- [10] D. Cordero-Erausquin, M. Fradelizi, and B. Maurey. “The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related problems.” *J. Funct. Anal.*, **214**(2), 2004, pp. 410–427. ISSN: 0022-1236.
- [11] D. Cordero-Erausquin. “Some applications of mass transport to Gaussian-type inequalities.” *Arch. Ration. Mech. Anal.*, **161**(3), 2002, pp. 257–269. ISSN: 0003-9527.
- [12] D. Cordero-Erausquin and A. Figalli. “Regularity of monotone transport maps between unbounded domains.” *Discrete Contin. Dyn. Syst.*, **39**(12), 2019, pp. 7101–7112. ISSN: 1078-0947.
- [13] C. Davis. “All convex invariant functions of hermitian matrices.” *Archiv der Mathematik*, **8**(4), 1957, pp. 276–278.
- [14] G. De Philippis. *Regularity of optimal transport maps and applications*. Vol. 17. Tesi, Sc. Norm. Super. Pisa (N.S.) Pisa: Edizioni della Normale; Pisa: Scuola Normale Superiore (Diss. 2012), 2013. ISBN: 978-88-7642-456-4; 978-88-7642-458-8.

- [15] G. De Philippis and A. Figalli. “Partial regularity for optimal transport maps.” *Publ. Math., Inst. Hautes Étud. Sci.*, **121**, 2015, pp. 81–112. ISSN: 0073-8301.
- [16] G. De Philippis and Y. Shenfeld. “Optimal transport maps, majorization, and log-subharmonic measures.” *Ann. Henri Lebesgue*, **8**, 2025, pp. 925–963. ISSN: 2644-9463.
- [17] H. Federer. *Geometric measure theory*. Vol. 153. Grundlehren Math. Wiss. Springer, Cham, 1969.
- [18] A. Figalli, F. Maggi, and A. Pratelli. “A mass transportation approach to quantitative isoperimetric inequalities.” *Invent. Math.*, **182**(1), 2010, pp. 167–211. ISSN: 0020-9910.
- [19] A. Figalli and Y.-H. Kim. “Partial regularity of Brenier solutions of the Monge-Ampère equation.” *Discrete Contin. Dyn. Syst.*, **28**(2), 2010, pp. 559–565. ISSN: 1078-0947.
- [20] N. Gozlan and M. Sylvestre. “Global Regularity Estimates for Optimal Transport via Entropic Regularisation.” 2025. URL: <https://arxiv.org/abs/2501.11382>.
- [21] G. Hargé. “A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces.” *Probab. Theory Relat. Fields*, **130**(3), 2004, pp. 415–440. ISSN: 0178-8051.
- [22] B. Klartag. “Marginals of geometric inequalities.” In: *Geometric aspects of functional analysis. Proceedings of the Israel seminar (GAFA) 2004-2005*. Berlin: Springer, 2007, pp. 133–166. ISBN: 978-3-540-72052-2.
- [23] A. V. Kolesnikov. “Mass transportation and contractions.” 2011. URL: <https://arxiv.org/abs/1103.1479>.
- [24] G. Monge. “Mémoire sur la théorie des déblais et des remblais.” *Histoire de l’Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, 1781, pp. 666–704.
- [25] A. Pratelli. “On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation.” *Ann. Inst. Henri Poincaré, Probab. Stat.*, **43**(1), 2007, pp. 1–13. ISSN: 0246-0203. URL: <https://eudml.org/doc/77922>.
- [26] R. T. Rockafellar. *Convex analysis*. Vol. 28. Princeton University Press, 1997.
- [27] F. Santambrogio. *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*. Vol. 87. Prog. Nonlinear Differ. Equ. Appl. Cham: Birkhäuser/Springer, 2015. ISBN: 978-3-319-20827-5; 978-3-319-20828-2.
- [28] C. Villani. *Optimal transport. Old and new*. Vol. 338. Grundlehren Math. Wiss. Berlin: Springer, 2009. ISBN: 978-3-540-71049-3.