LECTURE NOTES ON OPTIMAL TRANSPORTATION - WITH AN EYE TOWARDS COMPLEX GEOMETRY

ROBERT J. BERMAN

Abstract. These are very preliminary lecture notes for an ongoing graduate course on Optimal Transportation that I teach at Chalmers University of Technology. As I add new material I also go through the previous pages, revising them. The current version of the notes can be found on the course web page. The notes follow to a large extent [18] and [3], with an emphasizes on the quadratic setting and the connections to real Monge-Ampère equations. Some particular aspects of the quadratic theory are emphasized, which admit a "complex" counterpart in the realm of complex geometry and pluripotential theory (compare [6]). Be aware of a proliferation of typos! For example "cr" refers to a cross reference to a section to be written and "ref" refers to a reference that will eventually be added...

Contents

1. Existence of optimal transport plans and Kantorovich duality

1.1. **Notation.** In the following the spaces X, Y etc that we will work with will always be subsets of $\mathbb{R}^n$ and the word *measure* on such a space X will refer to a Borel measure (i.e. a measure which is compatible with the standard topology). All functions will be assumed to be (Borel) measureable.

Given a map T from X to Y and a measure μ on X we write $T_*\mu$, the *push-forward of μ under T*, is the measure on Y defined by

$$(T_*\mu)(F) := \mu(T^{-1}(F))$$

if F is a subset in Y. Recall that, under the usual pairing between measures and functions

$$\langle f, \mu \rangle := \int f d\mu$$

the push-forward on measures is dual to the *pull-back* on functions:

$$\langle T^* g, \mu \rangle = \langle g, T_*\mu \rangle$$

(where $(T^*g)(x) := g(T(x))$). In other words, $T_*\mu$ satisfies the following "change of variables formula":

$$\int_X g(T(x))\mu = \int_Y g(y) T_*\mu$$

1.2. **Formulation of the transport problem according to Monge and Kantorovich.** Let us start by the formulation of the transport problem according to Monge: given two probability measures μ and ν on $\mathbb{R}^n$ a map T from $\mathbb{R}^n$ to $\mathbb{R}^n$ is said to be a *transport map* (transporting μ to ν) if

$$T_*\mu = \nu$$

Given a *cost function* $c(x, y)$, i.e. a function on $\mathbb{R}^n \times \mathbb{R}^n$ one can define the (total) cost $C(T)$ of a transport map by

$$C(T) := \int c(x, T(x)) d\mu(x)$$

Accordingly, a transport map T is said to be *optimal* (wrt the given cost function) if it minimizes $C(T)$ over all transport maps (transporting μ to ν).

The "Monge problem" is to prove the existence of an optimal T (and to "characterize" T) under suitable regularity assumptions. However, there are two main difficulties:

- *Non-linearity:* The functional C is non-linear in T
- *Lack of compactness:* if T_i is a sequence of transport maps there seems to be no useful topology ensuring that it has a subsequence convirging to some transport map T

In Kantorovich's formulation of the transport problems both these difficulties disappear! The point is to consider a "relaxed" version of the optimization problem, where we enlarge the domain of the functional C that we want to optimize.

Definition 1.1. *A transport plan* γ (between μ and ν) is a probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ whose first and second marginals are equal to μ and ν, respectively, i.e. $(\pi_1)_*\gamma = \mu$ and $(\pi_2)_*\gamma = \nu$ where π_1 and π_2 denote the projections from $\mathbb{R}^n \times \mathbb{R}^n$ onto the first and second factors, respectively.

Given a cost function $c(x,y)$ one then defines

$$C(\gamma) := \int c(x,y)d\gamma,$$

which is clearly linear. Moreover, under suitable assumptions the desired compactness property above holds and the existence of a transport plan is then a routine matter (see below).

Lemma 1.2. *If T is a transport map (between μ and ν), then*

$$\gamma_T := (I \times T)_*\mu$$

is a transport plan (supported on the graph of T) and $C(T) = C(\gamma)$. Conversly, if γ is suppored on the graph of a transport map T then $\gamma = \gamma_T$.

Proof. [Exercise] (hint: use the the push-forward behaves well under composition of maps). □

Example 1.3. The "standard" or "quadratic" setting refers to the case when $c(x,y) = |x-y|^2$

1.3. **Existence of optimal transport plans.** To simplify the exposition we will in the following mainly focus our attention on the case when the supports X and Y of μ and ν, respectively, are compact.

Proposition 1.4. *Suppose that the cost functional $c(x,y)$ is continuous. Then, given μ and ν, there exists an optimal transport plan (between μ and ν). More generally, the result holds as long as $c(x,y)$ is lower semi-continuous (lsc) and bounded from below.*

Proof. Set $M := X \times Y$ and assume first that X and Y are compact. The key fact we want to use is that if M is a compact space then the space $\mathcal{P}(M)$ of all probability measures on M is compact, when equipped with the weak topology. Recall that the weak topology is defined by $\gamma_j \to \gamma$ iff

$$\int u d\gamma_j \to \int u d\gamma$$

for all "test functions", i.e. all continuous functions u on M. Let now $\Gamma(\mu,\nu)$ denotes the subspace of $\mathcal{P}(X \times Y)$ consisting of all transport plans (between μ and ν) and observe that is closed in the weak topology [exercise: prove this using that push-forward is dual to pull-back]. The space $\Gamma(\mu,\nu)$ is non-empty, as it contains $\mu \times \nu$. Also, since a closed subspace of a compact space is closed it follows that $\Gamma(\mu,\nu)$ is *compact.* Finally, note that, by the very definition of the weak topology, the functional C is continuous on $\Gamma(\mu,\nu)$ (just take u to be equal to c). But a continuous function on a compact space always admits a minimizer and that concludes the proof.

Finally, let us briefly explain how to prove the general case. The point is to equip M with the weak topology defined with respect to test functions that are *bounded* continuous functions. Due to the non-compactness of M it is not true in general that a sequence of probability measures γ_i on M has a subsequence converging to another probability measure γ. The problem is that "mass can be pushed out to infinity" (for example: take γ_i be translations of a fixed probability measure ν with compact support, so that $\gamma_i \to 0$, which is not a probability measure). However, if we can rule out this phenomen then relative compactness indeed holds. More precisely, suppose that the sequence γ_i is *tight,* i.e. for any $\epsilon>0$ there exists a compact subset K_ϵ such that $\int_{K_\epsilon^c} d\gamma_j < \epsilon$ for all j, then, after perhaps passing to a subsequence, $\gamma_i \to \gamma$ for some probability measure γ (this is the content of Prokhorov's theorem, which holds in the general setting of Polish spaces). In the present setting tightness is ensured by the condition on the marginals [exercise].For $c(x,y)$ continuous and bounded below one can then take γ_i to be a sequence converging to the minimium of the functional C (which is finite since c is bounded from below). As explained above we may assume that $\gamma_i \to \gamma$, where γ is a probability measure (and as before one also sees that $\gamma \in \Gamma(\mu,\nu)$). To conclude the proof it is enough to check that the functional C is lower-semi continuous along γ_i, i.e. that

$$\liminf_{j\to\infty} C(\gamma_j) \geq C(\gamma).$$

Indeed, then γ must be a minimizer of C on $\Gamma(\mu,\nu)$). In the case when c is a bounded continuous function the inequality above follows immediately from the definition of the weak topology. The general case is then reduced to the previous case by an approximation/truncation argument, using that a lower-semi continuous function can be written as an increasing limit of continuous functions c_k which are bounded from below. Indeed, as above we get

$$\liminf_{j\to\infty} C(\gamma_j) \geq \liminf_{j\to\infty} C_k(\gamma_j) \geq C_k(\gamma).$$

for any fixed k. Finally, letting $k \to \infty$ concludes the proof [warning: the argument does not exclude that C is identically equal to ∞, but then any γ is a minimizer] . □

In order to make contact with the Monge problem we need to adress two problems:

- When can an optimal transport plan be realized by a transport plan?
- How to characterize optimal transport plans (maps)?

It turns out that both these problems have satisfactory answers (under suitable regularity assumptions). We will start with the second point, focusing on the "quadratic setting". In a nut shell we will show that a map T is optimal if and only if it can be written *as a "gradient map" of a convex function:*

$$T" = "\nabla\phi$$

However, this needs to be made more precise as the gradient of a convex function is not everywhere defined, unless ϕ is differentiable. There is also a generalization of the previous statement to transport plans, saying that a transport plan γ is optimal iff the *support of γ is contained in the graph of the sub-gradient of a convex function* ϕ. To make this statements precise we start with:

1.4. **Recap of basic convexity I (gradients).**

Definition 1.5. A function $\phi : \mathbb{R}^n \to]-\infty,\infty]$ is said to be convex if it is "sub-affine", i.e. for any two points x and y and $t \in]0,1[$

$$\phi(tx+(1-t)y) \leq t\phi(x) + (1-t)\phi(y)$$

(excluding the case when $\phi = \infty$ everwhere). Similarly, a finite function ϕ is *strictly convex* if strict inequality holds above for $x \neq y$. A general reference for convexity is [17].

The convex subset $\{\phi < \infty\}$ is called the the *domain* of ϕ. Conversely, given a finite convex function ϕ on a convex subset A of $\mathbb{R}^n$ one obtains a canonical convex extension of ϕ to all of $\mathbb{R}^n$ by declaring that $\phi = \infty$ on $\mathbb{R}^n - A$ (i.e. the domain of the extension is A). The set of all convex functions is closed under the max operatation. In particular, the sup of a family of affine functions is always convex. It is not hard to check that any finite convex function is automatically continuous [exercise!], but, of course, it need not be differentiable.

Proposition 1.6. *Let ϕ be a smooth function on $\mathbb{R}^n$. Then ϕ is convex iff any of the following two conditions holds:*

- *Its Hessian is semi-positive in the sense of matrices: $(\partial^2\phi) \geq 0$*
- *The graph of ϕ lies over any of its tangents: given $x_0 \in X$ we have that for any $x \in \mathbb{R}^n$*

$$\phi(x_0) + \langle y_0, x - x_0 \rangle \leq \phi(x) \tag{1.1}$$

if $y_0 = \nabla\phi(x_0)$.

Conversly, if y_0 satisfies the previous inequality, then $y_0 = \nabla\phi(x_0)$.

For a general, possibly non-differentiable, convex function one may define a multi-valed generalization of the gradient called the *sub-gradient:*

Definition 1.7. If ϕ is a convex function, then its *sub-gradient* $(\partial\phi)(x_0)$ at x_0 is defined as the set of all vectors $y \in \mathbb{R}^n$ satisfying the last inequality in the previous proposition (in particular, if $\phi(x_0) = \infty$, then $(\partial\phi)(x_0)$ is empty). The *graph* $\Gamma_{\partial\phi}$ is defined as the graph of the multivalued function $\partial\phi : x \mapsto (\partial\phi)(x)$, i.e. the set of all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $y \in (\partial\phi)(x)$.

Note that $(\partial\phi)(x)$ is always a closed convex set (warning: $\partial\phi)(x)$ may be non-empty even if ϕ is finite at x!)

Example 1.8. If $\phi(x) = |x|$, then ϕ is convex (as it is the sup of the affine functions x and $-x$) and its sub-gradient at the point $x = 0$ is equal to $[-1, 1]$.

Proposition 1.9. *A convex function ϕ is differentiable at x_0 iff $(\partial\phi)(x_0)$ contains a single point y_0. In that case y_0 coincides with the classical gradient at x_0.*

Using properties of the sub-gradient it can be checked that any convex function is locally Lipschitz continuous. It then follows from the classical Rademacher's theorem that *ϕ is differentiable almost everywhere.* More precisely, it can be shown that the the non-differentiability locus of ϕ is a *small set,* i.e. it has Hausdorff dimension at most $n - 1$ (this is geometrically clear in the special case when ϕ is the max of a finite number of affine functions).

Definition 1.10. If ϕ is a convex function then the corresponding *Brenier map (or "gradient map in the sense of Brenier")* is the $L^\infty_{loc}-$map from the domain of ϕ to $\mathbb{R}^n$ defined by the almost everywhere defined gradient $\nabla\phi$.

1.5. **Characterization of optimal transport plans and maps for a quadratic cost (formulation).** Recall that by Prop 1.4 an optimal transport plan always exists. However, it could be that for any transport plan γ the cost $C(\gamma)$ is infinite. Accordingly, to exclude this scenario we will first, following ref ref, focus on the case when the given marginals μ and ν have finite second moments. The first main results that we shall discuss is:

Theorem 1.11. (the Knott-Smith optimality criterion). *Assume that the given probability measures μ and ν have finite second moments. Then a transport plan γ is optimal iff the* support of γ is contained in the graph of the sub-gradient of a convex function ϕ.

The relation to the original Monge problem is given by the second result to be discussed:

Theorem 1.12. (Brenier) *Assume that the given probability measures μ and ν have finite second moments. If the measure μ does not charge small sets, then*

- *There exists an optimal (measurable) map T which can be realized as a gradient map, $x \mapsto \nabla\phi(x)$ (in the sense of Brenier) of a convex function ϕ.*
- *The optimal map T is uniquely determined up to a small set*
- *The closeure of the image of the support of μ under the map $\nabla\phi$ is equal to the support of ν*
- *If moreover ν does not charge small sets, then for $d\mu-$almost every x and $d\nu-$almost every y*
$$\nabla\phi^* \circ \nabla\phi(x) = x, \quad \nabla\phi \circ \nabla\phi^*(y) = y$$
and the map $y \mapsto \nabla\phi^*(y)$ is the (almost $\nu-$) unique optimal map transporting ν to μ (where ϕ^* is the Legendre transform defined below).

Recall that a set is *small* if it has Hausdorff dimension at most $n-1$ and in particular the assumption in the previous theorem is satisfied for any μ which has a density, i.e. which is absolutely continuous wrt Lesbegue measure.

The relation to gradients of convex functions can be established directly using the notion of *cyclic monotonicity* [15, 3], which leads to the following general result, which bypasses the need for finite moment assumptions:

Theorem 1.13. *(McCann [15]). Given any two probability measures μ, such that μ does not charge small sets, there exists a convex function ϕ such that the $L^\infty_{loc}-$map $\nabla\phi$ pushes forward μ to ν. Moreover, the map $\nabla\phi$ is determined almost everywhere wrt μ.*

However, here we will follow a different route which passes via Kantorovich duality (compare [7, 18]). This notion turns out to be closely related to the classical notion of *Legendre transforms* in convex analysis.

1.6. **Recap of basic convexity II (the Legendre transform).** Given a function $\phi(x)$ on $\mathbb{R}^n$ its Legendre transform is defined by

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

(in more invariant notation we should really replace $\mathbb{R}^n_x$ by a vector space V and $\mathbb{R}^n_y$ by its dual V^*, but the previous concrete notation will be adequate for the moment). The function $\phi^*(y)$ is always convex and lower semi-continuous, or *lsc,* for short (since it is a sup of affine functions). If moreover ϕ is convex then the following fundamental duality relation holds:

Proposition 1.14. *Let ϕ be lower semi-continuous function on $\mathbb{R}^n$. Then*

$$(\phi^*)^* = \phi \tag{1.2}$$

This relation becomes almost obvious once we recall a basic motivation for introducing Legendre transforms. The starting point is the fact that any given lsc convex function ϕ can be written as an envelope (i.e. the point-wise sup) of affine functions f which are smaller than ϕ

$$\phi(x) = \sup_{f \leq \phi} f(x) \tag{1.3}$$

(this is geometrically very reasonable, so let's accept it for the moment). Now, since $f(x)$ is affine we can write it as

$$f(x) = \langle y, x \rangle - a$$

for unique $y \in \mathbb{R}^n$ and $a \in \mathbb{R}$. Given y there is an optimal choice of a which satisfies $f \leq \phi$, namely $a = \phi^*(y)$ [check this!]. Hence, setting

$$f_y := \langle y, x \rangle - \phi^*(y)$$

allows us to remove the constraint $f \leq \phi$ to get

$$\phi(x) = \sup_{f \leq \phi} f(x) = \sup_{y} f_y(x) = (\phi^*)^*(x),$$

which proves the desired duality relation 1.2. Coming back to the statement 1.3 it will follow once one has check that, given a convex function $\phi(x)$ and a fixed point x_0 there exists an affine function $f(x)$ such that $f(x_0) = f(x)$ and $f \leq \phi$. Geometrically, this means that we need to find a hyperplane passing through x_0 which stays on the lower side of the graph of ϕ. In other words, its normal vector y is an element of $\partial\phi(x)$, so we just need to know that the letter set is non-empty. This is indeed the case as long as ϕ is finite in a neighbourhood of x (and in general one can apply the Hahn-Banach separation theorem which applies to the infinite dimensional setting of locally convex topological vector spaces). Anyway, we may as well assume that ϕ is smooth (by a simple regularization argument) and then simply take $y = \nabla\phi(x)$.

The previous discussion establishes a link between ϕ^* and the sub-gradient of ϕ which is made precise in the following

Proposition 1.15. *If ϕ is convex then*

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

with equality iff $y \in (\partial\phi)(x)$ (and if ϕ is lsc then equality holds iff $x \in \partial\phi^*$)

Proof. First, the proof of the general inequality follows directly from the definition of ϕ^*. Next, the opposite inequality holds precisely when $y \in (\partial\phi)(x)$, by the very definition of a sub-gradient. Indeed,

$$\phi(x) + \phi^*(y) \leq \langle x, y \rangle \iff \phi(z) - \phi(x) \geq \langle z - x, y \rangle$$

for all z, which, by definition, means that $y \in (\partial\phi)(x)$. The last statement follows immediatly from $\phi^{**} = \phi$ and symmetry. □

The equality case in the previous proposition has the following classical formulation in the case when ϕ is smooth and strictly convex (which for example appears in the formalism of thermodynamics): consider the change of variables $y = \nabla\phi(x)$ (which is one-to-one by the strict convexity). Then the Legendre transform may be defined by

$$\phi^*(y) = \langle x, \nabla\phi(x) \rangle - \phi(x)$$

Note that it also follows from the condition for equality in the previous proposition and th the fact that $\phi^{**} = \phi$ that the inverse of the gradient map $\nabla\phi(x)$ is $\nabla\phi^*(y)$ if ϕ is lsc. In the general case these relations are made precise by the following lemma which will play an important role in section 2.

Lemma 1.16. *The Legendre transform $\phi^*(y)$ is differentiable at y iff there exists a unique x such that $y = \nabla\phi(x)$. Moreover, in that case $\nabla\phi^*(y) = x$, if ϕ is lsc.*

Remark 1.17. The previous lemma in particular implies a useful duality relation, saying that "$\phi^*(y)$ is essentially differentiable iff $\phi(x)$ is essentially strictly convex" (the relaxtion to strict convexity comes from the fact that it can be phrased in terms of the injectivity of the gradient). Here the adjective *essentially* has a precise meaning: a convex function ϕ is essentially differentiable if it is differentiable on its domain and $|\nabla\phi|(x) \to \infty$ as x converges to a point in the boundary of the domain of ϕ. Similarly, ϕ is strictly convex if its strictly convex on the domain of $\partial\phi$, i.e. on the set where $\partial\phi$ is non-empty [see ref R section 23].

Example 1.18. The following simple example illustrates the need to allow the value ∞ when working with convex functions. Let $\phi(x) = \log(e^{-x} + e^{x})$ (or any smooth strictly convex regularization, in the sense of convolution, of the function $|x|$. Then $\nabla\phi$ maps $\mathbb{R}$ diffemorphically onto $]-1,1[$ (since $\nabla\phi(x) \to \pm\infty$ as $x \to \pm\infty$). In particular, the graph of Γ of $\nabla\phi(x)$ looks as follows: [add picture]. Conversely, if we swap the role of x and y and set $\psi(x) := \phi^*(x)$, then $\phi^* = \infty$ on the complement of $]-1,1[$ and the graph of $\nabla\psi$ is a subset of $]-1,1[\times\mathbb{R}$. Note that if one fixes a probability measure of the form $\mu = f dx$ for $f > 0$ and defines $\nu := (\nabla\phi)_*\mu$ (whose support is equal to $[-1,1]$) then, by Theorem 1.12 $T = \nabla\phi$ is the unique continuous optimal map tranporting μ to ν.

Proposition 1.19. *Let ϕ be a finite convex function on $\mathbb{R}^n$. Then*

$$\partial\phi(\mathbb{R}^n) = \{\phi^* < \infty\}$$

In particular, if $\partial\phi(\mathbb{R}^n) \subset F$ where F is a closed convex set, then

$$\phi = (\chi_F \phi^*)^* \tag{1.4}$$

(where χ_F is the indicator function defined below)

Proof. First observe that $\partial\phi(\mathbb{R}^n) \subset \{\phi^* < \infty\}$ as follows immediateley from Prop 1.15. Next, we observe that the discussion above explaining the proof of the relation $\phi^{**} = \phi$ gives that $\phi(x) = \sup_{y \in \partial\phi(\mathbb{R}^n)} \langle y, x \rangle - \phi^*(y)$ (since we can always take the maximizer $y(x)$ as $\nabla\phi(x)$ and hence $\phi = (\chi_{\partial\phi(\mathbb{R}^n)}\phi^*)^*$. But $\partial\phi(\mathbb{R}^n)$ is closed and convex [add why] and hence $u := \chi_{\partial\phi(\mathbb{R}^n)}\phi^*$ is convex and lsc. In particular, $u^{**} = u$ and hence $\phi^* = \chi_{\partial\phi(\mathbb{R}^n)}\phi^*$,which $\partial\phi(\mathbb{R}^n)^c \subset \{\phi^* < \infty\}^c$, proving the first statement of the proposition. The same argument also proves the last formula in the proposition. Finally, a general finite convex function may be written as a decreasing limit of differentiable smooth convex functions ϕ_j (with the same sub-gradient image). In particular, since $\phi_j \geq \phi$ we get $\phi_j^* \leq \phi$. But by the previous discussion $\phi_j^* = \infty$ on F^c and hence $\phi^* = \infty$ on F^c, which concludes the proof of the general case. □

1.6.1. *Indicator and support functions.* To a given a convex subset $A \subset \mathbb{R}^n$ one can attach two canonical convex functions χ_A and h_A defined on $\mathbb{R}^n_x$ and its dual $\mathbb{R}^n_y$, respectively:

- The *indicator function* $\chi_A(x)$ is defined $\chi_A(x) = 0$ on A and cr$\chi_A(x) = \infty$ on $\mathbb{R}^n - A$ (i.e. $e^{-\chi_A}$ is the characterisitic function of A).
- The *support function* $h_A(y)$ is defined by $h_A(y) := \sup_{x \in A} \langle y, x \rangle$

Note that h_A is always a *one-homogeneuous* convex function, i.e. $\lambda^{-1} h_A(\lambda \cdot) = h_A(\cdot)$ for any $\lambda \in \mathbb{R}$ (and conversely any such function is the support function of a convex set A [exercise]).

The indicator and support function of a closed convex set A are intertwined under Legendre transformation. Indeed, it follows immediately from the definitions that

$$h_A = \chi_A^*$$

This basic duality relation plays a leading role in the proof of Kantorovich duality to which we next turn.

1.7. **Kantorovich duality and linear programming.** Recall that the Kantorovich problem amounts to optimizing a certain linear functional $C(\gamma)$ defined on a convex set of *measures* γ. The Kantorovich duality, which is the subject of this section, relates this problem to a *dual problem* of optimizing another functional $J(\phi, \psi)$ defined on pairs of *functions* $\phi(x)$ and $\psi(y)$, where

$$J(\phi, \psi) := \int \phi(x) d\mu + \int \psi(y) d\nu$$

Theorem 1.20. *(Kantorovich duality) Assume that the cost function $c(x, y)$ is the standard quadratic one and that the μ and ν are supported on compact sets, denoted by X and Y, respectively.*

- *Then,*
$$\inf_{\gamma} C(\gamma) = \sup_{\phi, \psi} J(\phi, \psi)$$
where γ ranges over all transport plans from μ to ν and $\phi(x)$ and $\psi(y)$ over all continuous functions on $X \times Y$ such that $\phi(x) + \psi(y) \leq c(x, y)$.
- *The infimum in the rhs above is attained for $(\phi, \psi) = (\phi, \phi^*)$, where ϕ can be taken to be convex on X (i.e. the restriction to X of a convex function on $\mathbb{R}^n$).*

The previous theorem can be seen as an infinite dimensional version of a standard duality result for the finite dimensional setting of "linear programming" or "linear optimization" (this is no coincidence as Kantorovich is the father of the latter theory). To see this consider the following finite dimensional problem. Let V be a finite dimensional vector space and denote by V^* its dual (defined as the set of linear functions on V). Given an element in $w \in V^*$ and $v \in V$ we write, as usual, the value of w at v as $\langle w, v \rangle$. More concretely, identifying V and V^* with $\mathbb{R}^N$ we may identitify $\langle w, v \rangle$ with the scalar product between w and v. This identification also induces a partial order $\leq$ on V, where, by definition, $v \leq v'$ if the inequality holds coordinate-wise (and similarly on V^*). We can then consider the convex subspace $V_+ = \{v \in V : v \geq 0\}$ of V (the "positive octant" in V). In this setting a linear program amounts to the problem of minimizing a linear function C on the convex set V_+, subject to affine contraints. The duality result in question relates this problem to a dual optimization problem where, loosely speaking, *the constraints determine the new function to optimize and the original function determines the new contraints:*

Proposition 1.21. *(linear progamming duality) Assume given $w_0 \in V^*$, a linear subspace $W \subset V^*$ and $v_0 \in V$. Then the infimum*

$$\inf_{v \in V_+} \langle w_0, v \rangle$$

over all $v \in V_+$, subject to the affine constraints $\langle w, v \rangle = \langle w, v_0 \rangle$ for all $w \in W$ is equal to

$$\sup_{w \in W} \langle w, v_0 \rangle ,$$

subject to the constraint $w \leq w_0$.

To see the relation to the previous setting first note that we can view the elements $v = (v_1, ..., v_N)$ of V as *(signed) measures on the finite set $I = \{1, 2, ..., N\}$*, simply by letting v_i be the mass of v at i. In other words, we identify v with the measure

$$v := \sum v_i \delta_i,$$

where δ_i is the Dirac mass at i. Then $v \geq 0$ iff the corresponding signed measure is in fact a bona fide measure, i.e. non-negative. As for the dual space V^* we can identify it with the set of all *functions on I*, simply by setting $w(i) := w_i$. Then $w \geq 0$ iff the corresponding function is non-negative. Moreover, we can write the corresponding pairing as

$$\langle w, v \rangle := \int_I w dv,$$

Going on to the infinite dimensional setting the idea is to replace the set I with the set $M := X \times Y$. Then we let V be space of all singed measures γ on X (equipped with the weak topology) and V^* its topological linear dual, which may be identfied with the space $C^0(X)$ of all continous functions f on M. The corresponding pairing is thus taken as

$$\langle f, \gamma \rangle := \int f d\gamma$$

Now the cost-function $c(x, y)$ may be identified with an element $w_0 \in V^*$, but what about the affine constraints on γ (i.e. the conditions that the marginals of γ are given by μ and ν respectively)? The point is that this condition may be formulated as the condition that

$$\int (\phi(x) + \psi(y))\, d\gamma = \int \phi(x) d\mu + \int \psi(y) d\nu (= \int (\phi(x) + \psi(y))\, d(\mu \times \nu)$$

for any pair of functions $\phi(x)$ and $\psi(y)$ [exercise!]. Accordingly, the role of the linear subspace $W \subset V^*$ is here played by the subset of all continuous functions on $X \times Y$ of the form $(\phi(x) + \psi(y))$ and the role of v_0 is played by the measure $\mu \times \nu$. Then, formally applying the previous proposition gives the Kantorovich duality in Theorem 1.20 (note that we do not need to assume a priori that γ has total unit mass, as it follows from the constraints). More precisely, in section cr below we will give a proof of Prop 1.21 which also applies to the infinite dimensional setting of the the Kantorovich theorem.

Remark 1.22. The standard duality in linear programming is usually formulated in a slightly more general form (the "canonical form") involving a matrix A : the infimum

$$\inf_{v \in V_+} \langle w_0, v \rangle$$

over all $v \in V_+$, subject to the affine constraints $A^T v = v_0$ is equal to

$$\sup \langle w, v_0 \rangle$$

over all w subject to the constraint $Aw \leq w_0$. Exercise: show that this general form may be reduced to the special one appearing in the previous proposition, by relating the linear subspace W to the matrix A.

1.8. **Proof of Theorem 1.11 using Kantorivich duality.** To see the relation to Legendre transforms one first observes that the cost $c(x,y) = |x-y|^2$ may be replaced by $-\langle x, y \rangle$ without changing the optimizers. Indeed, expanding $|x-y|^2 = -2\langle x, y \rangle + |x|^2 + |y|^2$ reveals that

$$C(\gamma) := \int -\left(2\langle x, y \rangle + |x|^2 + |y|^2\right) d\gamma(x,y) = -2\int \langle x, y \rangle \, d\gamma(x,y) + A + B,$$

where the constants A and B are the second moments of μ and ν respectively. To ger rid of the minus sign it is convenient to reformulate the Kantorivich duality as

$$\sup_{\gamma}(-C(\mu,\nu) = \inf_{\phi,\psi} \int \phi(x) d\mu + \int \psi(y) d\nu$$

(where we have changed (ϕ, ψ) to $(-\phi, -\psi)$) where now $\phi(x) + \psi(x) \geq -c(x,y) = \langle x, y \rangle$). In other words,

$$\sup_{\gamma}(\int \langle x, y \rangle \, d\gamma = \inf_{\phi,\psi} \int \phi(x) d\mu + \int \psi(y) d\nu), \quad \phi(x) + \psi(y) \geq \langle x, y \rangle$$

Fixing ϕ we note that, by the very definition of the Legendre transform, $\psi(y) \geq \phi^*(y)$ and hence

$$J(\phi, \psi) \geq J(\phi, \phi^*) := J(\phi)$$

Next one observes that ϕ may as well be taken to be convex by replacing ϕ with

$$P(\phi) := \phi^{**}$$

- note that this essentially amounts to applying the previous argument again (this is refered to as the "double convexification trick" in [18]). The point is that, by definition, $P(\phi) \leq \phi$(ensuring that $P(\phi)$ is finite and continuous on X) and $P(\phi)^* = \phi^*$, which implies that

$$J(\phi) \geq J(P\phi)$$

Hence,

$$\sup_{\gamma}(\int \langle x, y \rangle \, d\gamma = \inf_{\phi} J(\phi)$$

where ϕ ranges over all *convex* continuous functions on X. Moreover, the existence of a maximizer ϕ is not hard to establish (for example it follows from the second point in Theorem 1.23). Now, by construction, we have, if ϕ is a minimizer of J and γ an optimal transport plan that

$$\int \langle x, y \rangle \, d\gamma = \int \phi d\mu + \int \phi^* d\nu = \int (\phi(x) + \phi^*(y)) \, d\gamma$$

But this means that

$$\langle x, y \rangle = (\phi(x) + \phi^*(y))$$

a.e. wrt γ. Finally, by the variational property of the Legendre transform (Prop 1.15) this means that the support of γ is contanied in the sub-gradient of ϕ, as desired. The converse is proved in a similar manner [exercise].

1.8.1. *Proof of Theorem 1.12.* By Theorem 1.11 there exists an optimal transport plan γ which is supported in the graph $\Gamma_{\partial\phi}$ of the subgradient of a convex function ϕ. In particular, $C(\gamma) = \int_{\Gamma_{\partial\phi}} c(x,y)d\gamma(x,y)$. Next, let T be the gradient map of ϕ which is well-defined on the set $X - S$, where, by assumption, $\mu(S) = 0$. We write $\Gamma_{\partial\phi}$ as the disjoint union $\Gamma_1 \cup \Gamma_2$, where Γ_1 is the gradient of T over the set $X - S$. By definition $\int_{\Gamma_2} f(x,y)d\gamma(x,y) = \int_{X-S} f(x,T(x))d\mu$ for any continuous function $f(x,y)$ on $X \times Y$ and hence $\int_{\Gamma_1} d\gamma = \mu(X-S) = 1$ (since $\mu(S) = 0$). It follows that γ puts no mass on Γ_2 and hence $C(\gamma) = \int_{X-S} f(x,T(x))d\mu$ for any $f \in C^0(X \times Y)$, i.e. $\gamma = (I \times T)_*(\mu)$, where T is the Brenier gradient map of ϕ. The minimizing property of γ thus gives $C(T) = C(\gamma_T) \leq C(\gamma_{T'}) = C(T')$ for any transport map T' , which concludes the existence proof of an optimal transport map realized as a Brenier map.

Finally, to prove uniqueness we assume that the gradients of ϕ and ψ both have the prescribed push-forward property. We can then take $\gamma = \gamma_\phi$ to the optimal transport plan defined by ϕ and use that ψ optimizes the functional $J(\psi)$ appearing in the Kantorovich duality. Repeating the argument in the proof of the characterization of the support of γ gives that the support of γ is contained in the graph of $\partial\psi$. But by definition it is also contained in the graph of $\partial\phi$ and since ϕ and ψ are differentiable $\mu-$almost everywhere we deduce that $\nabla\psi(x) = \nabla\phi(x)$ a.e. wrt μ as desired.

1.9. **Proofs of the Kantorovich (linear programming) duality.**

1.9.1. *Proof of the linear pro1.21gamming duality using the min-max principle.* The proof of Proposition 1.21 can be reduced to a general result about Legendre transforms (see below). Alternatively, more directly, it follows from the so called *min-max principle,* saying that if $F(v,w)$ is a function which is convex in v and concave in w, then

$$\inf_v(\sup_w F(v,w)) = (\sup_w \inf_v F(v,w))$$

Indeed, the point is to first remove the constraints by adding a "penalizing term" in the function to optimize, i.e. replacing $\langle w_0, v\rangle$ with

$$\langle w_0, v\rangle + \chi_A(v),$$

where $\chi_A(v)$ is the indicator function of the affine space $A \subset V$ defined by the affine constraints above, i.e. the set of all $v \in V$ such that $\langle w, v\rangle = \langle w, v_0\rangle$ for all $w \in W$. More precisely, we can write

$$\chi_A(v) = \sup_{w\in W} \langle w, v_0 - v\rangle \tag{1.5}$$

(note that the rhs clearly vanishes on A and if it doesn't vanish then we can, by the one-homogenity wrt w, replace w with λw to see that $\chi_A(v) = \infty$). The original constrained infimum may thus be written as

$$\inf_{v\in V_+}(\langle w_0, v\rangle + \chi(v)) := \inf_{v\in V_+}\left(\langle w_0, v\rangle + \sup_{w_i}\langle w_i, v - v_0\rangle\right) := \inf_{v\in V_+}\sup_{w\in W} F(v,w)$$

where

$$F(v,w) = \langle w_0, v\rangle + \langle w, v_0 - v\rangle = \langle w_0 - w, v\rangle + \langle w, v_0\rangle$$

and $w \in W$. Applying the min-max principle allows us to swap the inf and the sup to get $\sup_W \inf_{V_+} F(v,w) := \sup_W f(w)$, where

$$f(w) := \inf_{v \in V_+} F(v,w) := \inf_{v \in V_+} \langle w_0 - w, v\rangle + \langle w, v_0\rangle$$

Now, if we add the constraint that $w \leq w_0$ then clearly $f(w) = 0 + \langle w, v_0\rangle$ and otherwise $f(w) = -\infty$ (indeed, if it is not the case that $w \leq w_0$ then we can find $v \geq 0$ such that $\langle w_0 - w, v\rangle < 0$ and hence replacing v with λv for $\lambda >> 1$ shows that the infimum above is $-\infty$. Accordingly, when taking the sup of $f(w)$ we may as well restrict to the subset where $w \leq w_0$ giving

$$\sup_W \inf_{V_+} F(v,w) := \sup_W f(w) = \sup_{w \leq w_0} \langle w, v_0\rangle$$

as desired. There are also infinite dimensional generalizations of the min-max principle which can be used to prove the Kantorovich duality, but here we will instead give a proof which uses Legendre transforms in infinite dimensions.

1.9.2. *Proof of linear progamming and Kantorovich duality using Legendre transforms.* Let us first consider the finite dimensional setting for linear programming described above. Set $E(v) := \langle v, w_0\rangle$ if $v \geq 0$ and $E(v) = \infty$ otherwise and $D(v) := \chi_A(v)$ where $A \subset V$ is the affine space defined by the affine constraints. Then the primal problem is to minimize the function

$$E(v) + D(v)$$

on the vector space V. Note that (by formula 1.5) D is the Legendre transform of $w \mapsto \langle w, v_0\rangle + \chi_W(w)$. To prove the duality in question we will invoke the following general result:

Theorem 1.23. *(Fenchel-Rockafeller). Let V be a normed (possibly infinite dimensional) topological vector space. Then*

$$\inf_{v \in V}(E + D) = \sup_{w \in V^*}(-E^*(w) - D^*(-w))$$

as long as there exists some $w \in V^*$ such that $E(w) \neq \infty, D(w) \neq \infty$ and D is continuous at w. Moreover, the sup above is attained.

To deduce Prop 1.21 from the previous theorem we also need to compute the Legendre transform of E, which by definition is given by

$$E^*(w) = \sup_{v \in V_+}(\langle v, w\rangle - \langle v, w_0\rangle) = \sup_{v \in V_+}\langle v, w - w_0\rangle = \chi_{\{w \leq w_0\}}$$

Since $D^*(w)$ is clearly one-homogenous we get $-D^*(-w) = D^*(w)$. Hence by the previous theorem

$$\inf_V(E + \chi_A) = \sup_{w \in V^*}(-\chi_{\{w \leq w_0\}} + (\langle w, v_0\rangle + \chi_W(w))) = \sup_{w \in W, w \leq w_0}\langle w, v_0\rangle,$$

which proves the linear programming duality in Prop 1.21. Similarly, in the infinite dimensional setting we can take V to be the space of all signed (Borel) measures on $X \times Y$ etc and proceed as in the discussion following the statement of Prop 1.21.

Remark 1.24. More generally, the previous argument reveals that if $E(v)$ is a linear function, i.e. $E(v) = \langle v, w_0\rangle$ and $A \subset V$ a convex subspace, then

$$\inf_{v \in V_+,\ v \in A} E(v) = \sup_{w \in V^*, w \leq w_0} h_A(w)$$

where $h_A(w) := \sup_{v \in A} \langle v, w \rangle$ is the support function of A. In the previous setting A was affine and determinedT by $v_0 \in V$ and the subspace $W \subset V^*$, which gives $h_A(w) = \langle w, v_0 \rangle + \chi_W(w)$, as before.

Finally, the proof in the infinite dimensional setting of Kantorovich duality proceeds in a similar way, taking V to be the space of all signed (Borel) measures on $X \times Y$ etc. Note also that the existence of a minimizer (ϕ, ψ) also follows from the Fenchel-Rockafeller theorem.

1.10. **Complements on the case of non-compactly supported measures.** Theorem 1.20 is valid as long as the moments of the marginals μ and ν are finite (which is needed to make sure that the functionals C and J are both finite). The proof can be reduced to the special case above by a suitable truncation procedure. The easiest step in the generalization is to allow one of the measures, say μ to have non-compact support. The point is that one can then exploit that if $P\phi_j$ is a sequence of convex functions minimizing $J(\phi, \phi^*)(:= J(\phi))$ then the image of $\nabla\phi_j$ is contained in the support of ν and hence uniformly bounded. The Arzela-Ascoli compactness result can then be invoked to deduce the existence of a limiting minimizer which is a finite (and even Lipschitz continuous) convex function on $\mathbb{R}^n$ (compare section cr).

However, if ν has non-compact support then the limiting convex function ϕ may be non-finite at some points. Typically, this happens if μ has convex compact support while the support of ν is non-compact, say all of $\mathbb{R}^n$. This is may be well illustrated by a one-dimensional example. Start with ϕ a smooth and strictly convex function on $]-1,1[$ with the property that $T := \nabla\phi$ maps $]-1,1[$ diffeomorphically onto $\mathbb{R}$. Then extending ϕ as $\chi_{[-1,1]}\phi$ gives a (lsc) convex function on all of $\mathbb{R}$ such that $\nabla\phi$ pushes forward a fixed probability measure $\mu := fdx$ with support $[-1,1]$ to a measure $\nu := gdy$ with support equal to all of $\mathbb{R}$. Moreover, since $|\nabla\phi(x)| \to \infty$ as x approaches the boundary of $[-1,1]$ (and since ϕ is uniquely determined on the support of μ) the extension by ∞ gives the only possible convex potential of the corresponding optimal transport map T (note that this example corresponds to swapping the roles of μ and ν in example 1.18).

Finally, it should be pointed out that the proof of the existence part of McCann's Theorem 1.13 in [15] also proceeds by truncation and is based on the observation that a sequence of transport plans whose support is cyclically monotone converges (after passing to a subsequence) weakly to another transport plan whose support is also cyclically monotone.

1.11. **Complements on uniqueness of the potential ϕ.** By Theorem 1.12 the optimal transport map T between μ and ν is uniquely determined almost everywhere wrt μ. But unless the support of μ is all of $\mathbb{R}^n$ this does not imply that a correponding "potential" ϕ, i.e. a convex function ϕ on $\mathbb{R}^n$ satisfying $\nabla\phi = T$, is uniquely determined (modulo constants). Indeed, we can modify ϕ as we wish on the complement of the support of μ (as long as it stays convex). On the other hand, if the support of g is convex and we add the extra global condition that the sub-gradient of ϕ maps all of $\mathbb{R}^n$ into the support of g, then uniqueness indeed holds:

Proposition 1.25. *Assume that μ and ν are probability measures not charging small sets (say, compactly supported) and that the support of ν is convex. Then any finite convex function Φ on $\mathbb{R}^n$ such that $(\nabla\Phi)_*\mu = \nu$ is uniquely determined* *(modulo constants) by the further condition that the image of its sub-gradient be contained in the support Y of g. Such a Φ will be called a* canonical potential.

Proof. As recalled above $\nabla\phi$ is uniquely determined a.e. wrt μ.. Similarly, by symmetry $\nabla\phi^*$ is uniquely determined a.e. wrt ν. But the support of ν is assumed convex and in particular it is the closure of its interior, which is an open convex set G. But this means that if ψ is another transport plan then $\nabla(\phi^*-\psi^*)=0$ on G and hence, by continuity $\phi^*-\psi^*$ is equal to a constant on G (which after subtraction may assumed to vanish). Now, by assumption the images of the sub-gradients of ϕ and ψ are contained in the closure of G and hence it follows from Prop 1.19 that $\phi=\psi$ everywhere. □

Note that the "canonical potential" Φ (as above) indeed exist: starting with any potential ϕ we can simply set

$$\Phi := (\chi_Y(\phi^*))^*,$$

which means that Φ is determined by the condition equivalently: $\Phi^* = \chi_Y(\phi^*)$ (this point of view will be futher developed in section cr, cr). The point is that the restriction of ϕ^* to Y is always uniquely determined and the convexity of Y is then needed to make sure that the extension $\chi_Y(\phi^*)$ is convex (or equivalently, $\Phi^* = \chi_Y(\phi^*)$).

Remark 1.26. If one removes the assumption that the support of g be convex then it may seem natural to ask if the uniquness still holds if one instead demands that the image of its sub-gradient be contained in the convex hull of the support of g (recall that the subgradient image is always convex). But this is not true as illustrated by the following simple counter-example: let $\phi(x)$ be as in example 1.18 and take μ to be supported on two intervals obtained by removing a small neighbourhood of 0 from a small interval. Let ν be the push-forward of μ under $\nabla\phi$. But the push-forward property of ϕ and the condition on the sub-gradient image are unchanged under suitable deformations of ϕ close to zero, so there can be no uniqueness.

2. From optimal transport to real Monge-Ampere equations and back again

In this section the optimal transport problem will be rephrased in terms of the solvability of certain fully non-linear PDEs of Monge-Ampère type. A direct variational approach to the latter equations will be explained, which turns out to be essentially equivalent to Kantorovich duality (in particular, it yields yet another proof of the latter duality in the present setting). We also explain how the direct variational approach naturally leads to the setting of the second boundary value problem for Monge-Ampère equations in $\mathbb{R}^n$. Some complements on the regularity theory of Monge-Ampère equations are also given and how they apply to give regularity results for optimal transport maps.

2.1. **The smooth setting.** For a smooth function ϕ on $\mathbb{R}^n$ one defines the *Monge-Ampère operator* as the following partial differential operator:

$$\phi \mapsto \det(\partial^2\phi),$$

where the symmetric matrix $\partial^2\phi$ is the Hessian of ϕ. Note that for $n>1$ this is a *non-linear* partial differential operator and hence it is not a priori clear how to define its action on non-smooth functions. The relation to (quadratic) optimal transport comes from the following observation: assume given two probability measures

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

with smooth densities f and g and ϕ a smooth and strictly convex function. Let

$$T := \nabla \phi$$

be the corresponding gradient map and assume also the map $\nabla\phi$ is surjective. Then

$$T_* f dx = g dy \tag{2.1}$$

iff ϕ solves the following Monge-Ampère equation

$$\det(\partial^2 \phi) g(\nabla \phi) = f \tag{2.2}$$

Indeed, by assumption $y := \nabla\phi(x)$ defines a diffeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$ and differentiating gives

$$dy = \det(\partial^2 \phi)(x) dx \tag{2.3}$$

But since $\nabla\phi$ is a diffeomorphism the push-forward relation 2.1 holds iff the form $f dx$ is the pull-back of the form $g dy$ under the map $T = \nabla\phi$ i.e. iff

$$f(x)dx = g(y(x))dy(x) = g(\nabla\phi(x)) \det(\partial^2\phi)(x)dx$$

as desired.

2.2. **Brenier solutions.** For a non-smooth ϕ one makes the following

Definition 2.1. Given two function f and g in $L^1(\mathbb{R}^n)$ a convex function ϕ is said to be a *Brenier solution* (or a weak solution in the sense of Brenier) to the Monge-Ampère equation 2.2, if $g dy$ is the push-forward of $f dx$ under the Brenier map $x \mapsto \nabla\phi(x)$ Equivalently, for any $v(y) \in C_0(\mathbb{R}^n)$

$$\int v(y) g(y) dy = \int v(y) f((\nabla\phi)(x)) dx$$

The definition is made so that the Brenier map $\nabla\phi$ provided by Theorem 1.12 furnishes a Brenier solution to the Monge-Ampère equation 2.2. When f and g are smooth it can be shown that, under some further assumptions, any Brenier solution is in fact smooth (compare section cr below). But here we will mainly discuss what can be said in the case of general densities f and g in $L^1(\mathbb{R}^n)$. The first observation is that if ϕ is a Brenier solution then the Monge-Ampère equation 2.2 in fact holds point-wise almost everywhere. The point is that any convex function ϕ is two times differentiable almost everywhere, i.e. $\partial^2\phi$ is well-defined almost everywhere (according to a classical result of Alexandrov).

Proposition 2.2. *(McCann* [16]*). Suppose that ϕ is a a Brenier solution to the Monge-Ampère equation 2.2. Then ϕ satisfies the equation 2.2 almost everywhere.*

See ref for the elementary proof, which uses the Lesbegue density theorem and the notion of Alexandrov solution recalled below. In fact, as shown by McCann the density $\det(\partial^2\phi)$ coincides the absolutely continuous part of the Monge-Ampère measure of ϕ discussed in the following section.

2.3. **Relations to Monge-Ampère measures and Alexandrov solutions.** The geometric content of the Monge-Ampère operator on smooth functions is encoded in the corresponding *Monge-Ampère measure* (compare formula 2.3):

$$MA(\phi) = \det(\partial^2\phi)(x)dx$$

The general definition of the Monge-Ampère measure goes back to the work of Alexandrov:

Definition 2.3. For a general (finite) convex function ϕ the *Monge-Ampère measure* $MA(\phi)$ (called the *Hessian measure* in ref) is the (Borel) measure defined by

$$MA(\phi)(E) := |(\partial\phi)(E)|$$

(using the Lebegue measure in the rhs above)

This is indeed a bona fide measure and in particular additive (see below). More generally, for any given $g \in L^1(\mathbb{R}^n)$ (or even more general, given a measure ν not charging small sets) one can define the following generalization of the Monge-Ampère measure:

$$MA_g(\phi)(E) := \int_{(\partial\phi)(E)} g(y)dy$$

One can then consider the following Monge-Ampère equation for a convex function ϕ, formulated on the level of measures:

$$MA_g(\phi) = \mu \tag{2.4}$$

given probability measures μ and gdy. As will be explained in section cr, if the support of μ is not all of $\mathbb{R}^n$ it is also natural to supplement the previous equation with a condition on the sub-gradient image of ϕ, but for the moment we do not impose any further conditions.

Remark 2.4. The original motivation for Alexandrov's definition of the Monge-Ampère measure $MA_g(\phi)$ came from his work on the classical *Minkowski problem* in differential geometry, i.e. the problem of finding a convex hypersurface S with prescribed Gauss curvature. Analytically, this problem amounts to solving the Monge-Ampère equation 2.2 with $g(y) := (1+|y|^2)^{(n+2)/2}$ and f the prescribed Gauss curvature. Indeed, if ϕ denotes the "height" of the smooth hypersurface S in $\mathbb{R}^{n+1}$, i.e. the map $F : x \mapsto (x, \phi(x))$ embedds S as the graph of ϕ in Euclidean $\mathbb{R}^{n+1}$, then the Gauss curvature $\kappa(x)$ of S at the point in S corresponding to $x \in \mathbb{R}^n$ is given by the density of $MA_g(\phi)$ (using that $\kappa(s)$ if the determinant of the second fundamental form of S). As observed by Alexandrov, applying $MA_g(\phi)$ to a piecewise affine convex function allows one to extend the notion of Gauss curvature to the singular setting of convex polyhedra.

Lemma 2.5. *The set-function MA_g defines a (Borel) measure.*

- *The measure MA_g may be equivalently represented as*

$$MA_g(\phi) = (\nabla\phi^*)_*(gdy), \tag{2.5}$$

where $\nabla\phi^*$ is the Brenier map (in the "reversed direction") of the Legendre transform ϕ^*.

- *In particular, the measure equation 2.4 with the correspondning target condition holds iff μ is the push-forward of gdy under the Brenier map $y \mapsto \nabla\phi^*(y)$ iff*

$$\int u(x)MA_g(\phi) = \int g(y)u(\nabla\phi^*(y))dy$$

for any $u \in C_b(\mathbb{R}^n)$
- *ϕ is a Brenier solution to the equation 2.2 iff ϕ solves equation 2.4 for $\mu = fdx$*

Proof. In order to make sure that $MA_g(\phi)$ is a well-defined measure (and in particular additive) we need to verify that the multi-valued map $\partial\phi$ is invertible almost everywhere wrt gdy, i.e. wrt the Lesbegue measure dy, in the sense that for almost any y in the image of $\partial\phi$ there exists a unique x such that $y \in (\partial\phi)(x)$. By Lemma 1.16 this holds iff the Legendre transform ϕ^* is differentiable at y. But since $\phi^*$is convex it is indeed differentiable almost everywhere (on its domain). This shows that $\nabla\phi^*$ defines an inverse to $\nabla\phi$ in the almost everywhere sense and the formula 2.5 then follows immediately from the definition of the push-forward. The second point then follows directly from the definitions. To prove the final point we observe that in the general case that μ and ν do not charge small sets, arguing as above reveals that

$$(\nabla\phi)_*\mu = \nu \Longleftrightarrow (\nabla\phi^*)_*\nu = \mu$$

(compare the last point in Brenier's theorem). The third point now follows from the first one. Indeed, if ϕ is a Brenier solution then $(\nabla\phi^*)_*(gdy) = fdx$. □

Remark 2.6. Note that it follows immediately from the definition that if g is a probability density (which we will always assume) then $MA_g(\phi)$ is a probability measure iff $\phi^* < \infty$ a.e. wrt $g(y)dy$. For example, if ϕ is a constant function, then $MA(\phi) = 0$ and $\phi^* = \infty$ almost everywhere [exercise: check this using the various equivalent definitions of $MA(\phi)$].

The Monge-Ampère measure has good continuity properties:

Proposition 2.7. *Let ϕ_j be a sequence of (finite) convex functions converging uniformly to $\phi(x)$ on an open set U. Then $MA_g(\phi_j) \to MA_g(\phi)$ weakly on U.*

Proof. This can, for example, be shown using formula 2.5 [add proof]. □

Since any (finite) convex function can be locally regularized (for example, using convolution) the previous proposition shows that the general Monge-Ampère measure $MA(\phi)$ be obtained in terms of a limit of smooth ones.

Example 2.8. Let $\phi(x) = |x|/2$. Then $MA(\phi) = \mu$ for $\mu = \delta_0$. More generally, if $g(y)$ is any continuous function on $\mathbb{R}$ then $MA_g(\phi) = c\delta_0$, for $c = \int_{-1}^{1} g(y)dy$. [exercise: give two proofs of this fact: one using the definition of MA_g and the other using the equivalent Legendre transform definition. Also: regularize $\phi(x)$, illustrating the validity of the previous proposition].

Definition 2.9. A (finite) convex function ϕ is an *Alexandrov solution* of equation 2.2 if the Monge-Ampère measure $MA(\phi)$ has no singular part and equation 2.4 holds with $\mu = fdx$. In other words, ϕ is a Brenier solution such that $MA(\phi)$ has no singular part.

A Brenier solution need not be an Alexandrov solution, but this is the case if the support of ν is convex (see below).

Example 2.10. Consider first the case when $n = 1$ and assume that the support of fdx is an interval, while the support of gdy is the union of two different intervals. Then an optimal map has to be discontinuous and hence $T = \nabla\phi$ "jumps" at a point of discontinuity, which produces a a Dirac measure in $MA(\phi)$ a the corresponding point. For example, this situation appears if one takes the function $\phi(x) = |x|/2$ appearing the previous exercise and slightly change it by making it smooth and strictly convex on the complement of a small neigbourhood of the origin. On can then take f to be any probability measure on $\mathbb{R}$ containing a sufficently large interval centered at the origin and define gdy as the push-forward of μ under $\nabla\phi$. Since ϕ coincides with $|x|/2$ close to 0 the measure $MA(\phi)$ aquires a Dirac mass at 0.

Proposition 2.11. [11] *Assume that the support Y of $\nu = gdy$ is convex and $g(y) > 0$ for almost all y in Y. Then any Brenier solution of equation 2.2 is an Alexandrov solution.*

Proof. We have to show that the Monge-Ampère measure $MA(\phi)$ has no singular part, i.e. that it is absoluletely continuous wrt dx. The starting point is that, by Lemma 2.5, a Brenier solution satifies $MA_g(\phi) = fdx$, i.e. the following "change of variables formula" holds:

$$\int_E f(x)dx = \int_{(\partial\phi)(E)} g(y)dy$$

We need to prove that $MA(\phi)(E) = 0$ for any null set E, i.e. that for a null set

$$\int_{(\partial\phi)(E)} dy = 0$$

By the previous formula this is certainly the case if $g > 0$ on $(\partial\phi)(E)$. However, in general $(\partial\phi)(\mathrm{supp}(f))$ is only known to be contained in the convex hull of Y [add]. But if Y is convex we can indeed conclude the argument. □

2.4. **A direct variational approach to solving Monge-Ampère equations.** Given probability measures μ and ν, consider the following functional

$$J(\phi) := \int \phi(x)d\mu + \int \phi^*(y)d\nu$$

defined on the space $\mathcal{C}(\mathbb{R}^n)$ of all continuous functions ϕ on $\mathbb{R}^n$. This functional appeared naturally in the Kantorovich duality formulation of the optimal transport problem, but here we will show that it can be used *directly* to produce a solution to the Monge-Ampère equation

$$MA_\nu(\phi) = \mu \tag{2.6}$$

and hence an optimal transport plan (and as a byproduct this will give another proof of Kantorovich duality). In fact, the proof will automatically produce a solution ϕ in the following subspace of $C(\mathbb{R}^n)$:

$$\mathcal{C}_Y(\mathbb{R}^n) := \{\phi \in C(\mathbb{R}^n),\ \phi \text{ is convex and } (\partial\phi)(\mathbb{R}^n) \subset Y\} \tag{2.7}$$

where Y denotes the convex hull of the support of ν (which in this section will always be assumed compact). More precisely we will show the following

Theorem 2.12. *Assume given two probability measures $\mu$and ν such thatν does not charge small sets and has compact support, whose convex hull is denoted by $Y.$ If the measure μ has finite first moments, then there exists a finite convex function ϕ minimizing the functional J (with $J(\phi)$ finite) satisfying*

$$MA_\nu(\phi) = \mu$$

Moreover, ϕ may be taken to be in the space $\mathcal{C}_Y(\mathbb{R}^n)$ (then ϕ is called a canonial potential) and if the support of g is convex, then such a function ϕ is uniquely determined modulo constants.

The existence proof proceeds in two steps:

Step one: The functional J is differentiable and a convex function $\phi$is a critical point of the functional J on $C^0(\mathbb{R}^n)$ (i.e. the differential $dJ_{|\phi}$ vanishes) iff it satisfies the Monge-Ampère equation

$$MA_\nu(\phi) = \mu$$

(in the week sense introduced above).

Step two: there exists a convex function ϕ (with the further property that its sub-gradient image is contained in the convex hull of the support of ν) minimizing the functional J on $C^0(\mathbb{R}^n)$.

2.4.1. *The proof of step one (the critical point equation).* To prove step one we need to calculate the differential $dJ.$ By definition this is a one-form on $C^0(\mathbb{R}^n)$, i.e. for each point ϕ in $C^0(\mathbb{R}^n)$ $dJ_{|\phi}$ is a continuous linear functional on the tangent space of the vector space $C^0(\mathbb{R}^n)$. Conretely, in the present setting this means that we can identify $dJ_{|\phi}$ with a measure defined by its action on a continuous function u (which for our pourposes may be assumed to be compactly supported):

$$\left\langle dJ_{|\phi}, u \right\rangle := \frac{d}{dt}_{|t=0} J(\phi + tu)$$

(which can be interpreted as the "directional derivative" of J at ϕ along the tangent vector u). To calculate the differential we write

$$J = L_\mu - \mathcal{F}_\nu$$

where

$$L_\mu(\phi) := \int \phi d\mu \tag{2.8}$$

is the linear piece of J and

$$\mathcal{F}_\nu(\phi) := -\int \phi^*(y) d\nu(u) \tag{2.9}$$

is (minus) the non-linear piece (to simplify the notation we will often write $\mathcal{F}_\nu = \mathcal{F}$)

Proposition 2.13. *The functional $\mathcal{F}$ is differentiable on $C(\mathbb{R}^n)$. More precisely, it is differentiable wrt bounded tangent vectors on the subspace where it is finite and*

$$d\mathcal{F}_{|\phi} := MA_\nu(\phi^{**})$$

Alternatatively,

$$d\mathcal{F}_{|\phi} := MA_\nu(P\phi)$$

where P is the following projection operator from $C(\mathbb{R}^n)$ to $\mathcal{C}_Y$:

$$P\phi := (\chi_Y \phi^*)^*, \tag{2.10}$$

and where Y is the convex hull of the support of ν.

Proof. The idea of the proof is to use the following variational property of the Legendre transform: if ϕ is a convex function such that $\mathcal{F}(\phi)$ is finte and u is continuous and bounded, then

$$\frac{d}{dt}_{|t=0}(\phi+tv)^*(y) = -v(\nabla\phi^*(y))$$

almost everywhere, or more precisely: on the complement of a small set. This can be shown by a direct calculus argument in the case when ϕ is smooth and strictly convex [exercise]. See [5] (Prop 2.13) for the proof in the general case, which is based on the elementary Lemma 2.15 below. Next, using the dominated convergence theorem and convexity one can differentiate under the integral sign (compare [5]) to get

$$\frac{d}{dt}_{|t=0}\mathcal{F}(\phi+tv) := \int v(\nabla\phi^*)d\nu(y)$$

Using the push-forward formula for MA_g in Lemma 2.5 this proves that $d\mathcal{F}_{|\phi} := MA(\phi)$ for any finite convex function as above (which, in fact, will be enough for our purposes). Finally, in the general case we observe that $\mathcal{F}(\phi) = \mathcal{F}(P\phi)$ and apply the previous argument to the function ϕ^{**} (or $P\phi$) (anyway, the case when ϕ is convex will be enough for our purposes). □

Remark 2.14. Note that even if ϕ is convex the convex function $P(\phi)$ differs from ϕ^{**} unless the sub-gradient image of ϕ is contained in Y. But according to the previous proposition the corresponding Monge-Ampère measures MA_ν are the same (the point is that ν is supported in Y).

Lemma 2.15. *(differentiation of perturbed optima)* Let G_0 be a proper upper semi-continuous function on $\mathbb{R}^n$ with a unique maximizer x_0 and let $G_t(x) := G_0(x) + tv(x)$ for a bounded continuous function v. Then $g(t) := \sup_{x\in\mathbb{R}^n} G_t(x)$ is differentiable at $t=0$ and

$$\frac{dg(t)}{dt}_{t=0} = v(x_0)$$

Proof. A simple formal proof goes as follows: write $g(t) = G(x(t),t)$, where $dx(t)_{|t=0} = 0$. Applying the chain rule gives, at $t=0$, $dg/dt = 0 + \frac{\partial G}{\partial t}(x(0),0) = v(x(0))$ as desired. Of course, the problem is that it is precisely the differentiability which is at stake. But the rigorous proof proceeds by explicitely writing out the derivative as a difference quotient and estimating all terms properly (exploiting that $g(t)$ is convex in t); see the appendix of [5]. □

2.4.2. *The proof of step two (existence of minimizers).* Recall that Y denotes the convex hull of the support of ν. Recall that $\mathcal{C}_Y$ denotes the convex subspace of $C(\mathbb{R}^n)$ consisting of all convex functions ϕ whose sub-gradient image is contained in Y. Note that the operator P from $C(\mathbb{R}^n)$ to $\mathcal{C}_Y$ (formula cr) satisfies $P\phi \leq \phi$ and $P(\phi)^* = \phi^*$ on the support of g. Hence, $J(P\phi) \leq J(\phi)$ and in particular

$$\inf_{C(\mathbb{R}^n)} J = \inf_{\mathcal{C}_Y} J.$$

This means that it will be enough to find a minimizer of the functional J restricted to C_Y. Now, by assumption, Y is bounded and hence the functions in C_Y are uniformly Lipschitz in $\mathbb{R}^n$ (since $|\partial\phi| \leq C$ where C only depends on the diameter of Y). Moreover, since J is invariant under $\phi \mapsto \phi + c$, for $c \in \mathbb{R}^n$ we may as well restict J further to the subspace $C_{Y,0}$ of functions satisfying the normalization condition $\phi(0) = 0$ holds). But, by the Arzela-Ascoli theorem, the space $C_{Y,0}$ is compact wrt the topology defined by uniform convergence on compact subset of $\mathbb{R}^n$. Accordingly, it will be enough to verify that J is lsc wrt the latter topology on $C_{Y,0}$. For the linear piece $(:= L_\mu)$ of J (depending on μ) this follows imediately from Fatou's lemma and the dominated convergence theorem (using that $\phi(x) \leq C|x|$ and by assumption $L_\mu(|x|) := \int |x|\mu < \infty$). As for the non-linear piece $\mathcal{F}_\nu$ the lower semi-continuity follows from the following observation: if ϕ_i is a sequence in C_Y converging point-wise to ϕ then

$$(i) \ \liminf_{i \to \infty} \phi_i^* \geq \phi^*, \quad (ii) \ \phi_i^* \geq -C$$

Indeed, the lower semi-continuity in question then follows immediately from Fatou's lemma. Finally, note that the inequality (i) follows from the variational definition of ϕ_i : for any fixed y we have $\phi_i^*(y) \geq \langle y, x \rangle - \phi_i(y)$ and hence first letting $i \to \infty$ and then taking the sup over y proves (i). To prove (ii) note that since $\partial\phi \subset Y$ we have $\phi_i(x) \leq \phi_i(0) + \sup_{y \in Y} \langle y, x \rangle \leq C + \sup_{y \in Y} \langle y, x \rangle$ and optimizing over x then proves (ii).

2.4.3. *End of the proof of Theorem 2.12.* Since μ has finite first moments we $J(\phi)$ is finite for any $\phi \in C_Y$ such that $\mathcal{F}(\phi) > -\infty$). Let$\phi$ be a minimizer of J (which by the second step above may be taken in the subspace C_Y). For any fixed continuous function u with compact support the function $t \mapsto J(\phi + tu)$ has a local minimum at $t = 0$ and hence differentiating wrt t gives

$$0 = \frac{d}{dt}_{|t=0} J(\phi + tu) =: \langle dJ_{|\phi}, u \rangle = \langle (\mu - MA(\phi)), u \rangle$$

Since the vanishing holds for any such u this concludes the existence proof. As for the uniqueness it was shown in Prop 1.25 (and in this formulation μ can be any probability measure). Indeed, first, using Kantorovich duality (which as explained below is a consequence of Theorem 2.12) one sees that $\nabla\phi^* = \nabla\psi^*$ a.e. wrt ν. But then it follows as in the proof of Prop 1.25 that (after perhaps shifting ψ by a constant) $\phi^* = \psi^*$ a.e. wrt ν and hence $\phi = \psi$ on $\mathbb{R}^n$ (using that ϕ and ψ are in $\mathcal{C}_Y$).

2.4.4. *Remarks on the general structure of the proof (with an eye towards the complex setting).* In order to highlight the similarities with the complex setting we next spell out the general structure of the previous variational proof of Theorem 2.12. The starting point is the problem of solving an equation in a closed convex convex subspace $\mathcal{C}$ of a vector space V. At least formally, the equation in question can be realized as the critical point equation for a functional J on the convex subspace $\mathcal{C}$ of a vector space V :

$$dJ_{|\phi} = 0$$

Accordingly, to find a critical point ϕ of J in $\mathcal{C}$ it is natural to try to minimize the functional J on $\mathcal{C}$. But since $\mathcal{C}$ is not open it is not a priori clear that a minimizer of J on $\mathcal{C}$ will satisfy the critical point equation above. Indeed, a priori the minimizer may (and typically will) be on the boundary of $\mathcal{C}$. To circumvent this problem we note that there exists a natural extension of J to V and a "projection map" P which decreases J :

$$J : V \to \mathcal{C}, \quad J(P\phi) \leq J(\phi) \tag{2.11}$$

and which has the crucial property of beeing differentiable (at least when tested on a dense subset of perturbations in V) and such that critical points of J in $\mathcal{C}$ still satisfy the original equation. Moreover, the space $\mathcal{C}$ is (modulo a normalization) *compact* which leads to the existence of a minimizer in V (which, by construction is in $\mathcal{C}$) and thanks to its minimization property on all of V it satisfies the critical point equation.

In conclusion, one interesting feature of this approach is the interplay between the "intrinsic" property of $\mathcal{C}$ of beeing compact (whis gives the existence of a minimizer of J) and the "extrinsic" property of beeing embedded as a closed subset of V (which together with the existence of a suitable projection map is exploited to pass from a minimizer to a solution of the critical point equation).

Another interesting feature in the present setting (which also appears in the complex setting) is that the original functional $J(\phi)$ on $\mathcal{C}$ is a "local functional" in the sense that it only depends on the derivatives of ϕ (which follows form the fact that for $\phi \in \mathcal{C}$ the differential $d\mathcal{F}_{|\phi}$ is the Monge-Ampère measure of ϕ which clearly is a local operator). However, the extension of J (implicitely) involves the projection operator P which is certainly *not* a local operator. More precisely, the restriction of J to $\mathcal{C}$ may be written as

$$J(\phi) = L_\mu(\phi) - \mathcal{E}(\phi),$$

where $\mathcal{E}(\phi)$ has the property that $d\mathcal{E}_{|\phi} = MA(\phi)$ (on the convex space $\mathcal{C}$). Next, the extension of J to V may be defined by

$$J(\phi) = L_\mu(\phi) - \mathcal{F}(\phi),$$

where

$$\mathcal{F}(\phi) = \mathcal{E}(P\phi).$$

The fact that $P\phi \leq \phi$ and $P(P\phi) = \phi$ then gives the important decreasing property $J(P\phi) \leq J(\phi)$. Moreover, by Prop 2.13

$$d\mathcal{F}_{|\phi} = MA(P\phi) \tag{2.12}$$

and in particular $d\mathcal{F}_{|\phi} = MA(\phi)$ for $\phi \in \mathcal{C}$, which is crucial for the existence proof. Finally, note that it may be temptating to try to prove formula 2.12 by invoking the chain rule to get

$$d\mathcal{F}_{|\phi} := d(\mathcal{E} \circ P)_{|\phi} = d\mathcal{E}_{|P\phi}(dP_{|\phi}) \tag{2.13}$$

But this is not really the relation we are looking for and moreover the projection operator P is not even differentiable! Still the argument can be made to work using that the following "orthogonality relation" holds in the present setting:

$$d\mathcal{E}_{|P\phi}(P\phi - \phi) = 0 \tag{2.14}$$

(which can be used to replace dP_ϕ in formula 2.13 with the identity map as desired). A crucial ingredient in the variational approach in the complex setting introduced in [4] is that the corresponding ortogonality relation can be established directly, using the complex analoge of the following alternative expression for the projection operator P :

Lemma 2.16. *Let Y be a closed convex set in $\mathbb{R}^n$. For any lsc function ϕ on $\mathbb{R}^n$ the following formula holds*

$$P\phi := (\chi_Y \phi^*)^* = \sup_{\psi} \{\psi(\cdot) : \ \psi \leq \phi\}$$

where ψranges over all convex functions on $\mathbb{R}^n$ whose sub-gradient image is contained in Y.

Proof. Denote by Φ the rhs in the formula to be proved. First observe that, for any $y \in Y$ we have

$$\Phi^*(y) = \phi^*(y)$$

Indeed, this is a special case of the following relation (with $\psi(x) = \langle x, y \rangle$) which holds holds for any $\psi \in \mathcal{C}_Y$:

$$\sup_{x \in \mathbb{R}^n} (\psi(x) - \Phi(x)) = \sup_{x \in \mathbb{R}^n} (\psi(x) - \phi(x))$$

which follows immediately form the extremal definition of Φ (since $\psi \leq \phi + c$ iff $\psi \leq \Phi + c$). In other words, we have shown that $\chi_Y \Phi^* = \chi_Y \phi^*$. But then taking the Legendre transform of both sides and applying formula 1.4 concludes the proof. $\square$

2.4.5. *A direct proof of Kantorovich duality.*

Proposition 2.17. *Let μ and ν be as in Theorem 2.12. Then*

$$\sup_{\gamma} -C(\gamma) = \inf_{\phi,\psi} J(\phi, \psi) = \inf_{\phi \in \mathcal{C}(\mathbb{R}^n)} J(\phi)$$

where $\phi(x)$ and $\psi(y)$ are functions such that $\phi(x) + \psi(y) \geq \langle x, y \rangle$ (in an almost everywhere sense)

Proof. The inequalities $-C(\gamma) \leq J(\phi, \phi^*) \leq J(\phi, \psi)$ follow immediately from the inequality in Prop 1.15. Indeed,

$$-C(\gamma) := \int \langle x, y \rangle \, d\gamma \leq \int (\phi(x) + \phi^*(y)) d\gamma = J(\phi, \phi^*) \leq J(\phi, \psi)$$

Hence, maximizing over γ and minimizing over (ϕ, ψ) gives $lhs \leq rhs$. To conclude the proof it will (by logic) be enough to find a ϕ and a corresponding transport plan γ such that $-C(\gamma) = J(\phi, \phi^*)$. To this end we take ϕ to be the minimizer provided by Theorem 2.12 and γ the transport plan defined by the "reversed" Brenier map $\nabla\phi^*$, i.e. $\gamma := (\nabla\phi^* \times I)_* \nu$ (the fact that γ has the marginals μ and ν follows from the fact that the Monge-Ampère equation satisfied by the minimizer ϕ is equivalent to $\mu = (\nabla\phi^*)\nu$ (by Lemma 2.5)). In particular,

$$J(\phi, \phi^*) = \int (\phi(x) + \phi^*(y)) d\gamma = \int \langle x, y \rangle \, d\gamma$$

where the last equality follows from the equality case in Prop 1.15 (using that γ is supported in the graph of $\partial\phi^*(y)$, i.e. $x \in \partial\phi^*(y)$ a.e. wrt γ). Hence, $J(\phi, \phi^*) = -C(\gamma)$, as desired. $\square$

2.5. **The optimal cost functional, infinite dimensional Legendre duality and weighted transport theory.** In this section we will briefly discuss a weighted variant of the quadratic setting, which appeares naturally in [6] and which turns out to give a natural bridge to the complex setting as studied in [4]. The discussion will hopefully also shed some light on the study of the geometry of the Wasserstein space (which corresponds to having a quadratic weight function) that will be taken up in section cr.

Given a cost function $c(x,y)$ we define the *optimal cost* $C(\mu,\nu)$ as the optimal cost of transporting μ to ν :

$$C(\mu,\nu)=\inf_{\gamma\in\Gamma(\mu,\nu)}C(\gamma) \tag{2.15}$$

Coming back to the previous quadratic setting we set $c(x,y)=-\langle x,y\rangle$ and fix a probability measure ν which does not charge small sets, denoting the convex hull of its support by Y. Then we can view $C(\mu):=C(\mu,\nu)$ as a functional on the space $\mathcal{P}(\mathbb{R}^n)$ of all probability measures on $\mathbb{R}^n$. Interestingly, as will be next explained $C(\mu)$ may be viewed as the Legendre transform of the functional $\mathcal{F}$ (formula 2.9).

Lemma 2.18. *The funcional $C(\mu)$ can realized as the Legendre transform of $\mathcal{F}(\phi)$ in the following sense: if μ has finite first moments, then*

$$C(\mu)=\sup_{\phi\in C(\mathbb{R}^n)}\left(-\int\phi d\mu+\mathcal{F}(\phi)\right),$$

where the sup may as well be taken over the subspace $\mathcal{C}_Y$.

Proof. The first formula is an immediate consequence of Kantorovich duality, or alternatively Theorem 2.12. The second statement was proved in the course of the proof of Theorem 2.12. □

Some precisions are in order. First of all, $C(\mu)$ is really the Legendre transform of the convex functional $-\mathcal{F}(-\phi)$, but this is just a matter of sign conventions. Secondly, one has to be a bit careful when specifying precisely which function spaces are used in the definition of the Legendre transform (or more precisely, the behaviour of the functions at infinity in $\mathbb{R}^n$). Finally, the Legendre transform should really be defined on the whole vector space of all signed measures on $\mathbb{R}^n$, but this can be arranged by setting $C(\mu)=\infty$ if μ is not a probability measure. We will combe back to these subtles issues in the weighted setting introduced below, where they can be dealt with in a satisfactory manner.

But before turning to the weighted setting we first observe that the previous lemma, or more precisely Theorem 2.12, allows one to decide when $C(\mu)$ is differentiable. To make this precise first recall that a functional $C(\mu)$ on a convex space (here, the space of all probability measures $\mathcal{P}(\mathbb{R}^n)$) is said to be *(Gateuax) differentiable* if it is differentiable along all affine lines, i.e. given any two elements μ_0 and μ_1, setting $\eta:=\mu_1-\mu_0$, the function $t\mapsto C(\mu+t\eta)$ is differentiable wrt t. Moreover, in the present setting we will say that the differential $dC_{|\mu}$ at μ along $\mathcal{P}(\mathbb{R}^n)$ exists if there is a funcion $\psi_\mu\in C(\mathbb{R}^n)$ such that

$$\frac{d}{dt}_{|t=0}C(\mu+t\eta)=\int\psi_\mu\eta \tag{2.16}$$

(this is non-standard terminology). Note that if ψ_μ is only determined mod $\mathbb{R}$, i.e. up to addition with a constant (since $\int\psi_\mu\eta=0$).

Remark 2.19. One has to be a bit careful when talking about the differential of $C(\mu)$. The point is that we can either view $C(\mu)$ as intrinsically defined only on the space of all probability measures or as defined on the whole vector space of all signed measures (extending by ∞). In the latter case $C(\mu)$ will never be differentiable. Indeed, the function $C(\mu)$ becomes infinity when leaving the subspace $\mathcal{P}(\mathbb{R}^n)$, which in a sense has codimension one and can thus be thought of as a hypersurface in an infinite dimensional vector space. Here we have the intrinsic setting in mind when talking about differentiability and then we use duality to (at least formally) identify the tangent space at μ in $\mathcal{P}(\mathbb{R}^n)$ with a subset of $C(\mathbb{R}^n)/\mathbb{R}$.

Proposition 2.20. *Suppose that the probability measure ν has convex support :(=Y). Then the functional $C(\mu)$ is differentiable on the subspace in $\mathcal{P}(\mathbb{R}^n)$ consisting of all measures with finite first moments and the differential of $C(\mu)$ at μ along $\mathcal{P}(\mathbb{R}^n)$ may be represented by the canonical potential$-\phi_\mu$ of μ (furnished by Theorem 2.12):*

$$dC_{|\mu} = -\phi_\mu$$

(in sense that formula 2.16 holds with $\psi_\mu = -\phi_\mu$).

Proof. This is shown as in the finite dimensional setting of Lemma 2.15, using the Legendre transform type formula in the previous lemma. □

Remark 2.21. If the support of ν is not convex, then the corresponding functional $C(\mu)$ will, in general, *not* be differentiable. This can be shown using that the corresponding canonical potential need not be unique (mod $\mathbb{R}$).

Next we turn to the weighted setting. Given a continuous functions ϕ_0 on $\mathbb{R}^n$ (refereded to as the *weight function*) we consider the following weighted cost functional:

$$c_{\phi_0}(x,y) := c(x,y) + \phi_0(x)$$

It follows immediately that the corresponding optimal cost functional $C_{\phi_0}(\mu)$ is simply a linear perturbation of $C(\mu)$:

$$C_{\phi_0}(\mu) = C(\mu) + \int \phi_0 d\mu$$

if all terms are finite. In the present quadratic setting we recall that $c(x,y) = -\langle x,y\rangle$ and hence

$$c_{\phi_0}(x,y) := -\langle x,y\rangle + \phi_0(x)$$

The main analytical advantage of introducing the weight ϕ_0 is that if ϕ_0 has sufficent growth at infinity in $\mathbb{R}^n$ then $c_{\phi_0}(x,y)$ will be bounded from below (when $y \in Y$ which we recall is supposed to be compact). For example, the weight $\phi_0(x) := |x|^2/2$ certainly has sufficent growth in the previous sense. But in this section we will focus on another natural class of weights with (asymptotically) *linear* growth which are singled out by the fixed target measure ν or more precisely by its support Y, namely the weights ϕ_0 such that $\phi_0 - \phi_Y$ is bounded:

$$\phi_0 = \phi_Y + O(1), \quad \phi_Y(x) := \sup_{y\in Y} \langle x,y\rangle\,, \tag{2.17}$$

where ϕ_Y denotes the support function of Y (which is finite since Y is compact). To simplify the notation we will simply take the canonical choice $\phi_0 := \phi_Y$, which has the convenient property that $c_{\phi_0}(x,y) \geq 0$ when $y \in Y$.

One interesting feature of the weighted setting is that one does not need to impose any assumptions (such as finite moments) on the measure μ as illustrated by the following Theorem.

Theorem 2.22. *Assume that $\nu = g1_Y dy$ where Y is a compact convex set and g is bounded from above and below by positive constants and ϕ_0 is a weight functional with linear growth as in formula 2.17. Then the following statements are equivalent:*

- $C_{\phi_0}(\mu) < \infty$
- *There exists a solution ϕ to the equation $MA_\nu(\phi) = \mu$ such that $\phi \in \mathcal{C}_Y$ and $\phi - \phi_0 \in L^1(\mu)$, i.e.*
$$\int -(\phi - \phi_0) MA_\nu(\phi) < \infty$$
- *There exists a convex function ψ on Y such that $\mu = (\nabla\psi)_*\nu$ and such that $\psi \in L^1(Y, d\nu)$ (namely, $\psi = \phi^*$ where ϕ is the solution appearing in the previous point)*

Proof. We will not go into details concerning the proof which can be adapted from the complex setting in [4]. The hard part is to show that first point implies the second one (or the third one). The condition that $C_{\phi_0}(\mu) < \infty$ means that

$$\int (\phi - \phi_0)\mu + \int \phi^* \nu \geq -C$$

for a uniform constant C (or more precisely, $C = C_{\phi_0}(\mu)$). Normalizing we may as well assume that $(\phi - \phi_0) \leq 0$ and hence $\phi^* \geq 0$ (indeed, ...). Hence, it is not a proiri clear that the previous lower bound implies a bounded on both individual terms appearing in the left hand side above. But in fact, it can be shown that there exists a constant A such that

$$\int -(\phi - \phi_0)\mu \leq A(\int \phi^* \nu)^{1/2}$$

and hence it must be that both terms are indeed under control. In particular, any minimizer ϕ satisfies $\int -(\phi - \phi_0)\mu < \infty$ and $\int(\int \phi^* \nu < \infty$ which proves the second point (since $\mu = MA_\nu(\phi)$) and the third point (setting $\psi = \phi^*$). □

Remark 2.23. If μ has finite first moment then moments, then $C_{\phi_Y}(\mu) < \infty$ (since $c_{\phi_y}(x, y) \leq A|x|$). However, the converse is not true. Indeed, by the previous theorem $C_{\phi_Y}(\mu) < \infty$ iff $\mu = (\nabla\psi)_*\nu$ for some convex $\psi \in L^1(Y, d\nu)$. But $\int |x| d\mu = \int |\nabla\psi| \, d\nu$ which may be infinite [exercise: construct a one-dimensional counter example].

Corollary 2.24. *(weighted Kantorovich duality) Under the assumptions in the previous theorem, the following holds for any probability measure μ :*

$$\inf_{\gamma \in \Gamma(\mu,\nu)} C_{\phi_0}(\gamma) = \sup_{\phi \in C(\mathbb{R}^n)} (\mathcal{E}_\nu(\phi) - \int (\phi - \phi_0) d\mu)$$

In particular, the inf in the lhs is finite iff the sup in the rhs is (alternatively, $\phi \in C(\mathbb{R}^n)$ can be replaced with $\phi_0 + C_b(\mathbb{R}^n)$).

Remark 2.25. This corollary is also a consequence of the very general Kantorovich duality formula in Prop 1.22 [18], which holds for any non-negative cost function.

Again, we can reformulate the previous corollary in terms of Legendre transforms in finite dimensions. This time no moment assumptions or needed. To make this precise, we set $F(u) := -\mathcal{F}(\phi_0 + u)$, viewed as a functional on the space$C_b(\mathbb{R}^n)$ of all bounded continuous functions. Denote by $\langle \cdot, \cdot \rangle$ the standard pairing between $C_b(\mathbb{R}^n)$ and the space $\mathcal{M}(\mathbb{R}^n)$ of all signed (Borel) measures on $\mathbb{R}^n$. Extending by ∞ we can view the optimal cost functional $C(\mu)$ as a convex functional on $\mathcal{M}(\mathbb{R}^n)$.

Corollary 2.26. *(same assumptions as in the previous theorem). The convex functional $C(\mu)$ on $\mathcal{M}(\mathbb{R}^n)$ is the Legendre transform of the convex functional $F(-u)$ on $C_b(\mathbb{R}^n)$ (and conversely). In particular,*

$$\mathcal{F}(\phi_0 + u) = \inf_{\mu \in \mathcal{P}(\mathbb{R}^n)} \left(C_{\phi_0}(\mu) + \int u d\mu \right) \tag{2.18}$$

Moreover, the unique minimizer of the functioal appearing in the rhs above is the measure $\mu_{\phi_0} := MA_\nu(P(\phi_0 + u))$ (and the differential....)

Proof. The first statement follows immediately from the previous corollary, when $\mu \in \mathcal{P}(\mathbb{R}^n)$ and the general case can be reduced to this case by general convexity considerations (for example using that the differential of C_{ϕ_0} takes values in $\mathcal{P}(\mathbb{R}^n)$; compare section in ref for the complex setting). It is tempting to deduce the converse Legendre relations by saying that $F^{**} = F$, but then one has to be a bit careful when applying a suitable version of the Hahn-Banach theorem in topological vector spaces. But the point here is that the Hahn-Banach theorem is not needed as we know that F is Gateaux differentiable (by Prop 2.13). Indeed, for any functional F on $C_b(\mathbb{R}^n)$ we have that, by definition, $F^{**}(u) \leq F(u)$ and we claim that if F is Gateaux differentiable at u then $F^{**}(u) = u$ and moreover

$$F(u) = \sup_{\mu} \langle \mu, u \rangle - C(\mu), \quad C(\mu) := F^*(\mu) \tag{2.19}$$

where the sup is attained precisely for $\mu = \mu_u := dF_{|u}$ (this argument already appeared in the discussion preceeding...) Indeed, by convexity μ_u is a sub-gradient at u for F and hence

$$F(u) + C(\mu) = \langle \mu, u \rangle$$

for $\mu = \mu_u$ (as in Prop 1.15). Moreover, as in the latter proposition equality holds above iff μ is a sub-gradient for F at u. But since F is Gateaux differentiable with differential μ_u it follows that any sub-gradient coincides with μ_u [exercise: prove this using that $\mu$is determined by its action on $C_b(\mathbb{R}^n)$] and hence equality above holds iff $\mu = \mu_u$, which proves the uniqueness claim in the corollary (since $\mu_u = MA_\nu(P(\phi_0 + u))$) according to Prop 2.13). Finally, writing out formula 2.19 and paying attention to the signs gives formula 2.18. □

[C11, regularity, absolutely continuous, moving boundary]

2.6. **The regularity problem for Monge-Ampère equations.** The deep regularity theory of the real Monge-Ampère operator is beyond the scope of the present lecture notes and we will just make a few remarks to illustrate the difficulties that appear and what kind of results do hold (for precise references see the survey [19]).

First recall that for a general partial differential operator $M(\phi)$ the regularity theory is centered around the equation

$$M(\phi) = f \quad \text{on}\, \Omega \tag{2.20}$$

for a given function f defined on a given domain Ω. The corresponding Dirichlet problem is when (continuous) vanishing boundary values are imposed. The classical situation is when $M(\phi)$ is a second order linear elliptic operator, the prototype beeing the case when $M(\phi)$ is the Euclidean Laplacian Δ on a domain Ω with smooth boundary (or more generally, the Laplacian Δ_g defined with respect to a given Riemannian metric g on $\bar{\Omega}$). Then a solution ϕ is always more regular than f. More precisely, by standard linear elliptic estimates ϕ is "two derivatives smoother than f"; for example, if f is in the Hölder space $C^{k,\alpha}_{loc}(\Omega)$ (for $\alpha > 1$) then ϕ is in $C^{k+2,\alpha}_{loc}(\Omega)$ and if f is in the Sobolev space $W^{k,p}_{loc}$ then ϕ is in $W^{k+2,p}_{loc}$. However, for the Monge-Ampère operator these results are, in general, no longer true (for $n > 1$).

Example 2.27. For $n > 1$ any function of the form $\phi(x) = \phi_1(x_1)$ for ϕ_1 a given convex function on $\mathbb{R}$ is a weak solution of the Monge-Ampère equation corresponding to equation 2.20, with $f = 0$. In particular, the solution is in general not differentiable, or even in $C^{\alpha}_{loc}(\Omega)$ for $\alpha > 1$.

In a nutshell, the problem is that the Monge-Ampère operator only prescribes the *product* of the eigenvalues of the Hessian matrix $\partial^2\phi$, while, in order to control the second order regularity of ϕ all eigenvalues need to be controled (in the previous example the second eigenvalue always vanishes, ensuring that $MA(\phi) = 0$, but giving no control of the first eigenvalue. Even if one demands that f be smooth and strictly positive, ϕ need not be smooth in any way. The problem is that (say when $n = 2$) the first eigenvalue can still blow-up if there is a compensatation coming from the second eigenvalue:

Example 2.28. For $n = 2$ set $f = 1$. Then the coresponding Monge-Ampère equation is invariant under dilations, i.e. under the transformation $\phi(x_1, x_2) \mapsto \phi_\epsilon(x) := \phi(\epsilon_1 x, \epsilon^{-1} x_2)$. But the derivatives wrt x_2 of the function $\phi_\epsilon(x)$ blow-up as $\epsilon \to 0$. In particular, there can be no, so called, *interior regularity estimates* of the form $\left\|\partial^k\phi\right\|_B \leq C_B \left\|M(\phi)\right\|_\Omega$ when B is a domain compactly included in Ω.

However, as illustrated by the following deep results of Caffarelli these blow-up phenomena can be ruled out if one imposes some boundary control on the solution, for example if one demands that ϕ vanishes on the boundary $\bar{\Omega}$ (and is continuous up to the boundary).

Theorem 2.29. *(Caffarelli* [11, 9]*). Assume that $f > 0$. Then any (convex) solution ϕ on Ω of the Monge-Ampère equation 2.20 which is continuous on $\bar{\Omega}$ and vanishes on $\partial\Omega$ has the following properties: ϕ is*

- *"Better than C^1", i.e. strictly convex and locally $C^{1,\alpha}$ for some $\alpha > 0$, if there exists a constant C such that $1/C \leq f \leq C$.*
- *"Almost C^2", i.e. in the class $W^{2,p}_{loc}$ for any $p > 1$, if f is continuous*
- *"Better than C^2", i.e. in $C^{2,\alpha}_{loc}$ if f is in $C^{0,\alpha}_{loc}$ and smooth if f is smooth*

Proof. (extreme sketch). The starting point of the proof is the following regularity result from [10]: if ϕ is a convex function on a convex domain Ω such that the Monge-Ampère measure $MA(\phi)$ satisfies the following inequalities:

$$1/Cdx \leq MA(\phi) \leq Cdx \tag{2.21}$$

and ϕ is strictly convex, then ϕ is $C^{1,\alpha}_{loc}$ for some $\alpha > 0$. But, as shown in [11] if the inequalities above hold and $\phi = 0$ on $\partial\Omega$, then ϕ is indeed strictly convex in Ω

[11].[1] As for the second point is contained in the main result in [9]. The final point then follows from the Evans-Krylov theory for fully non-linear elliptic operators and standard linear bootstrapping (all of which is purely local, i.e. independent of boundary assumptions). Briefly, the point is that by Evans-Krylov theory the following local result holds: if $MA(\phi) = fdx$ locally for $f > 0$ and ϕ is "almost C^2" then $f \in C^{\alpha}_{loc}$ implies that $\phi \in C^{2,\alpha}_{loc}$. Next, once we know that $\phi \in C^{2,\alpha}$ we can view the equation $MA(\phi) = fdx$ as a *linear* Laplace equation $\Delta_g \phi = \tilde{f}$ for ϕ with respect the $\phi-$dependent metric $g := \partial^2 \phi$ (with $\tilde{f} \in C^{\alpha}$for some $\alpha > 0$) and by the previous step g is in C^{α}_{loc}. But then it follows from classical elliptic theory that $\tilde{f} \in C^{k,\alpha}_{loc}$ implies that $\phi \in C^{k+2,\alpha}_{loc}$ for any $k > 0$. □

Let us briefly recall the general setting for Evans-Krylov theory refered to in the proof of the last point of the previous theorem (see the survey [8] and references therein). In the general theory of non-linear partial differential operators an operator of the form

$$\phi \mapsto F(\partial^2 \phi),$$

where F is a function on the space of all symmetric $n \times n$ matrices, is said to be *(uniformly(elliptic* at ϕ if the linearization of F at ϕ is (uniformly) positive definite (where for simplicity we assume that F is smooth close to $\partial^2 \phi$). Equivalently, this means that the linearization of the opererator $F(\partial^2 \phi)$ is a *linear* (uniformly) elliptic operator. Concretely, if L denotes the linearization at a given ϕ then the ellipticity means that

$$L(u) = \sum_{i,j} a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + ...$$

(where the dots indicate lower order terms) for a positive definite symmetric matrix $A := A_{ij}$ and uniform ellipticity amounts to the eigenvalues of A being uniformly bounded from above and below by a positive constant (the "modulus of the ellipticity"). In other words, ellipticity holds if $L = \Delta_g + ...$ for some metric g (corresponding to the inverse of A above). The function F is usually assumed to be concave (which simplifies the theory even if it is expected that concavity is not essential [8]). Here, when considering the Monge-Ampère equation 2.20 for f positive we can take

$$F(A) := \log \det A$$

and then the linearized operator at ϕ is precisely the Laplacian Δ_g defined with respect to the (degenerate) Riemanniam metric g determined by the Hessian $\partial^2 \phi$ (i.e. $g_{ij} = \partial^2 \phi / \partial x_i \partial x_j$) [exercise] which is thus elliptic precisely when ϕ is smooth and strictly convex. In fact, the strict convexity is automatic for any C^2-smooth solution of the equation in question (since f is assumed positive) but here it is precisely the regularity of ϕ which is at stake! According to the general Evans-Krylov theory for fully non-linear elliptic operators (which is purely local and thus independent of the boundary behaviour the solution) it is enough with a slightly weaker a priori control than C^2-regularity, namely $C^{1,1}-$regularity or even $W^{2,p}_{loc}-$regularity for p sufficently large (which for example holds if the Laplacian is bounded). In the

[1] More generally, it is enough to know that ϕ is $C^{1,\beta}$ on $\partial\Omega$ for some $\beta > 1 - 2/n$ (which rules out the scenario in example 2.28 where all k th derivatives blow up even though $MA(\phi)$ is under control). However, $\beta = 0$, i.e. Lipschitz continiuity, is not enough in general, as illustrated by $\phi(x_1, x') = |x'| + |x'|^{n/2}(1 + x_1^2)$ with Ω as small ball centered at 0 (see the discussion in section 5.5 in ref and the references therein).

present setting such a priori regularity for the corresponding Dirichlet problem is furnished by the second point in the previous theorem, as soon as f is positive and continuous.

Remark 2.30. In the light of the previous discussion the main analytical difficulty in obtaining higher regularity of a solution ϕ of Monge-Ampère equations is the a priori estimate of the Laplacian $\Delta\phi$. The first result in this direction is an explicit Laplacian a priori estimate of Pogorelov, which appeared in his seminal work on the Minkowski problem, later developed by Cheng-Yau and others. See [19] and references therein. Interestingly, Pogorelov type estimates also hold for the *complex* Monge-Ampère operator, where they are due to Aubin and Yau (in the case where there is no boundary) and the estimate then admit a geometric formulation in terms of Ricci and bisectional curvatures. However, in the complex setting there is nothing as precise as the second point in the previous theorem.

Corollary 2.31. *(Regularity of optimal maps) Let μ and ν be two probability measures with densities f and g respectively.*

- *If f and g are smooth and strictly positive, then the unique optimal map T transporting μ to ν is smooth (defining a diffeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$)*
- [12] *If the support of μ is a domain X such that ∂X has measure zero and the support Y of ν is convex and $1/C \leq g \leq C$ on Y, then the optimal map T (which is uniquely determined on X) is continuous, or more precisely in C^{α}_{loc} for some $\alpha > 0$. Moreover, if f and g are smooth on the interiour of X and Y, respectilvey, then T smooth on the interior of X.*

Proof. By Brenier's theorem $T = \nabla\phi$, where ϕ is a Brenier solution of the Monge-Ampère equation corresponding to μ and ν. Let us first see how to deduce the first point from the previous theorem (see [1] for a slightly different proof). First, invoking Lemma 2.32 below shows that ϕ is proper and hence the sublevel sets $\Omega_R := \{\phi < R\}$ are bounded convex domains exhausting $\mathbb{R}^n$. Fixing R, writing $\Omega := \Omega_R$ and replacing ϕ with $\phi - R$ we then have that $\phi = 0$ on $\partial\Omega$ and $1/Cdx \leq MA(\phi) \leq Cdx$ on Ω for some positive constant C [exercise: prove this using that g is smooth and strictly positive]. Hence, it follows from the first point in Theorem 2.29 that ϕ is in the Hölder class $\mathcal{C}^{1;\alpha}_{loc}$ for some $\alpha > 0$. In particular, the gradient $\nabla\phi$ is a single-valued Hölder continuous function and hence $MA(\phi) = f$ for a Hölder continuous function f such that $1/C' \leq f \leq C'$ in Ω. Applying the third point in Theorem 2.29 thus shows that ϕ is smooth as desired.

Turning to the proof of the second point, it first follows from Prop 2.11 that ϕ is an Alexandrov solution (using that Y is convex). More precisely, the proof of reveals that the inequalities 2.21 for $MA(\phi)$ hold. To conclude it is then enough to prove that ϕ is strictly convex (so that [10] can be invoked as in the proof of the first point in the previous theorem). The new difficulty is that the solution ϕ is not zero (or constant) on the boundary. Yet, as shown in [12] it is indeed strictly convex also in the present setting. Briefly, the point is to use that, since Y is compact, ϕ is globally Lipshitz continuous on $\mathbb{R}^n$ (and thus its graph admits an "asymptotic cone at ∞"). □

Lemma 2.32. *Let ϕ be a convex function (finite) convex function ϕ on $\mathbb{R}^n$ such that $MA_g(\phi) = \mu$ where gdy and μ are probability measures and the support Y of g is convex with g strictly positive on Y. If moreover 0 is contained in the interior of Y, then ϕ is proper.*

Proof. First observe that it will be enough to show that $\psi := \phi^*$ is finite on a closed small ball B_ϵ of radius ϵ centered at 0, Indeed, since ϕ^* is continuous there if follows that $|\phi^*| \geq C$ on B_ϵ. Hence, $\phi(x) = \psi^*(x) \geq \sup_{p \in B_\epsilon} \langle p, x \rangle - C = \epsilon |x| - C$, showing that ϕ is proper. Next, by Prop 1.19 it will thus be enough to show that 0 is an interior point of $(\partial\phi)(\mathbb{R}^n)$. But since Y is convex the MA-equation forces $(\partial\phi)(\mathbb{R}^n) \subset Y$ (compare the proof of Prop 2.11) and the MA-equation gives $\int_{(\partial\phi)(\mathbb{R}^n)} g dy = \int_{\mathbb{R}^n} \mu = \int_Y g dy$. Hence, $(\partial\phi)(\mathbb{R}^n)$ only misses a set in Y of measure zero wrt gdY, i.e. a null set for Lesbegue measure (since $g > 0$ in Y). But then if follows from a simple convexity argument that all of the interior of Y is contained in $(\partial\phi)(\mathbb{R}^n)$. In particular, 0 is an interior point of $(\partial\phi)(\mathbb{R}^n)$ as desired. □

2.7. **The second boundary value problem for the Monge-Ampère operator.** Let us briefly come back to the setting considered in section 2.4 in the special case when the target measure $\nu = 1_Y dy$ is the Lesbegue measure on a closed bounded convex set Y with unit-volume (i.e. Y is a convex body with normalized volume). Given any probability measure μ there is then a unique (mod $\mathbb{R}$) function ϕ on $\mathbb{R}^n$ such that

$$MA_\nu(\phi) = \mu, \quad \phi \in \mathcal{C}_Y,$$

where we recall that $\mathcal{C}_Y$ is the space of all convex functions on $\mathbb{R}^n$ whose subgradient image in Y. Since, ν coincides with dy on the image of $\partial\phi$ this means that we can remove the sub-script ν in the previous equation, i.e. ϕ is equivalently a solution to

$$(i)\, MA(\phi) = \mu, \quad (ii)\, (\partial\phi)(\mathbb{R}^n) \subset Y$$

The previous equation (i) with the "target condition" (ii) is often refered to as the second boundary value problem for the Monge-Ampère operator in the PDE litterature, where the convex body Y is called the target convex set (note that since the total mass of μ coincides with the volume of Y the condition (ii) above equivalently means that image of $(\partial\phi)(\mathbb{R}^n)$ is dense in Y). The rational for calling this a boundary value problem is that the target condition (ii) above only depends on the behaviour of ϕ at infinity in $\mathbb{R}^n$. Indeed, we have the following simple observation:

$$\phi \in \mathcal{C}_Y \Longleftrightarrow \phi(x) \leq \phi_Y(x) + C,$$

where ϕ_Y is the support function of Y and C is a constant (depending on ϕ) [exercise]. Geometrically, the target Y thus determines an asymptotic cone for the graph of the solution ϕ at infinity.

Theorem 2.33. *Let μ be a probability measure on $\mathbb{R}^n$ and denote by ϕ the corresponding convex function (unique mod $\mathbb{R}$) solving the corresponding second boundary value problem. Then*

- *If μ has finite weighted cost, with respect ϕ_Y, then $\int -(\phi - \phi_Y) MA(\phi) < \infty$ (equivalently, $\phi^* \in L^1(Y, dy)$)*
- *If μ has a density with finite q th moments, for some $q > n$, then $\phi - \phi_Y$ is bounded (equivalently, $\phi^* \in L^\infty(Y)$)*
- *If μ has a smooth density, then ϕ is smooth (equivalently, $\phi^* \in C^\infty_{loc}(Y)$)*

Proof. The first point is the content of Theorem 2.22. The second point follows from the Sobolev inequality applied on Y, using that $\int |x|\mu = \int |\nabla\phi^*|^q dy$ (compare [5]) and the last point follows from Theorem 2.29. □

Note that the target condition (ii) above is needed for the uniqueness property to hold. This is already clear in the case when $n = 1$. Indeed, then we may write $Y = [a, a + 1]$ for some $a \in \mathbb{R}$. But the equation (i) only determines $\phi(x)$ up to an additive affine term $Ax+B$, where the constant A is fixed by the gradient condition (ii) (since the term Ax shifts a to $a + A$).

References

[1] Alesker, S., Dar, S., Milman, V., A remarkable measure preserving diffeomorphism between two convex bodies in R n . Geom. Dedicata, Vol. 74, No. 2, (1999), 201-212.
[2] Alexandrov, A.D: Existence and uniqueness of a convex surface with a given integral curvature CR Doklady Acad Sci URSS NS (1942)
[3] Ambrosio, L; Gigli, N: A user's guide to optimal transport. http://cvgmt.sns.it/media/doc/paper/195/users_guide-final.pdf
[4] Berman, R.J; Boucksom, S; Guedj,V; Zeriahi: A variational approach to complex Monge-Ampere equations. Publications math. de l'IHÉS (2012): 1-67 , November 14, 2012
[5] Berman, R. J., Berndtsson, B., Real Monge-Amp ' ere equations and K¨ ahler-Ricci solitons on toric log Fano varieties. Preprint, arXiv:1207.6128
[6] Berman, R. J: Statistical mechanics of permanents, real Monge-Ampère equations and optimal transport. ArXiv.
[7] Brenier, Y: Polar factorization and monotone rearrangement of vector valued functions. Communications on pure and applied mathematics, 1991
[8] Cabre, X: Elliptic PDEs in Probability and Geometry. Symmetry and regularity of solutions. Discrete Contin. Dyn. Syst. 20 (2008), 425-457.
[9] Caffarelli, L.A: Interior \$W^{2,p}\$ estimates for solutions of the Monge-Ampère equation. Ann. of Math. (2) 131 (1990), no. 1, 135–150.
[10] Caffarelli, L.A: Some regularity properties of solutions of Monge Ampère equation. Comm. Pure Appl. Math. 44 (1991), no. 8-9, 965–969.
[11] Caffarelli, L. A. A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity. Ann. of Math. (2) 131 (1990), no. 1, 129–134.
[12] Caffarelli, L. A: The regularity of mappings with a convex potential. J. Amer. Math. Soc. 5 (1992), no. 1, 99–104
[13] Knott, M; Smith, C. S: On the optimal mapping of distributions. J. Optim. Theory Appl. 43 (1984), no. 1, 39–49.
[14] Knott, M; Smith, C. S: Note on the optimal transportation of distributions. Journal of Optimization Theory and Applications, 1987 - Springer
[15] McCann, Robert J. Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. 80 (1995), no. 2, 309–323
[16] McCann, Robert J: A convexity principle for interacting gases. Adv. Math. 128 (1997), no. 1, 153–179.
[17] Rockafellar, R. T: Convex analysis. Reprint of the 1970 original. Princeton Landmarks in Mathematics. Princeton Paperbacks. Princeton University Press, Princeton, NJ, 1997.
[18] Villani, C: Topics in optimal transportation. Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003. xvi+370 pp
[19] Trudinger, N.S: Wang, X.J: The Monge-Ampere equation and its geometric applications. Handbook of Geometric Analysis, 2008, pp. 467-524.