

FULL CHARACTERIZATION OF OPTIMAL TRANSPORT PLANS FOR CONCAVE COSTS

PAUL PEGON, DAVIDE PIAZZOLI, FILIPPO SANTAMBROGIO

ABSTRACT. This paper slightly improves a classical result by Gangbo and McCann (1996) about the structure of optimal transport plans for costs that are strictly concave and increasing functions of the Euclidean distance. Since the main difficulty for proving the existence of an optimal map comes from the possible singularity of the cost at 0, everything is quite easy if the supports of the two measures are disjoint; Gangbo and McCann proved the result under the assumption $\mu(\text{supp}(\nu)) = 0$; in this paper we replace this assumption with the fact that the two measures are singular to each other. In this case it is possible to prove the existence of an optimal transport map, provided the starting measure μ does not give mass to small sets (i.e. $(d-1)$ -rectifiable sets). When the measures are not singular the optimal transport plan decomposes into two parts, one concentrated on the diagonal and the other being a transport map between mutually singular measures.

MSC 2010 : 49J45, 49K21, 49Q20, 28A75

Keywords : Monge-Kantorovich, transport maps, approximate gradient, rectifiable sets, density points

CONTENTS

1. Introduction	2
2. Tools	5
2.1. General facts on optimal transportation	5
2.2. The approximate gradient	9
3. The absolutely continuous case	10
4. The general case: μ does not give mass to small sets	12
4.1. A GMT lemma for density points	12
4.2. How to handle non-absolutely continuous measures μ	14
5. Appendix	17
5.1. Concave costs and translations	17
5.2. On the definition of a weak μ -approximated gradient	18
References	20

1. INTRODUCTION

Optimal transport is nowadays a very powerful and widely studied theory for many applications and connections with other pieces of mathematics. The minimization problem

$$(M) \quad \min \left\{ \int c(x, T(x)) d\mu(x) : T_{\#}\mu = \nu \right\}$$

proposed by Monge in 1781 (see [13]) has been deeply understood thanks to the relaxation proposed by Kantorovich in [10] in the form of a linear programming problem

$$(1.1) \quad (K) \quad \min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c \, d\gamma : \gamma \in \Pi(\mu, \nu) \right\}.$$

Here μ and ν are two probability measures on \mathbb{R}^d and c a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$. The set $\Pi(\mu, \nu)$ is the set of the so-called *transport plans*, i.e. $\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : (\pi_x)_{\#}\gamma = \mu, (\pi_y)_{\#}\gamma = \nu\}$ where π_x and π_y are the two projections of $\mathbb{R}^d \times \mathbb{R}^d$ onto \mathbb{R}^d . These probability measures over $\mathbb{R}^d \times \mathbb{R}^d$ are an alternative way to describe the displacement of the particles of μ : instead of saying, for each x , which is the destination $T(x)$ of the particle originally located at x , we say for each pair (x, y) how many particles go from x to y . It is clear that this description allows for more general movements, since from a single point x particles can a priori move to different destinations y . If multiple destinations really occur, then this movement cannot be described through a map T .

The Kantorovich problem is interesting in itself and carries many of the features of Monge's one. Since it can be rigorously proven to be its relaxation in the sense of l.s.c. envelopes, the minimal value of the two problems is the same, provided c is continuous. For many applications, dealing with the optimum of (K) is enough. Yet, a very classical question is whether the optimizer γ of (K) is such that for almost every x only one point y can be such that $(x, y) \in \text{supp}(\gamma)$. In this case γ will be of the form $(id \times T)_{\#}\mu$ and will provide an optimal transport map for (M).

For several different applications, from fluid mechanics to differential geometry, the case which has been studied the most is the quadratic one, $c(x, y) = |x - y|^2$, first solved by Brenier in [2]. Other costs which are strictly convex functions of $x - y$, for instance all the powers $|x - y|^p$, $p > 1$, can be dealt with in a similar way. Next Section will give a general strategy to prove the existence of a transport map which fits very well this case. The limit case $c(x, y) = |x - y|$, which was by the way the original interest of Monge, has also received much attention.

Yet, another very natural case is that of concave costs, more precisely $c(x, y) = \ell(|x - y|)$ where $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a strictly concave and increasing function. From the economical and modelization point of view, this is the most natural choice: moving a mass has a cost which is proportionally less if the distance increases, as everybody can notice from travel fares. In many practical cases, moving two masses on a distance d each is more expensive than moving one at distance $2d$ and keeping at rest the other. The typical example is the power cost $|x - y|^\alpha$, $\alpha < 1$. Notice that all these costs satisfy the triangle inequality and are thus distances on \mathbb{R}^d . Among the other interesting features of these costs, let us mention two. From the theoretical point of view, there is the fact that all power costs $|x - y|^\alpha$ with $\alpha < 1$ satisfy Ma-Trudinger-Wang assumption for regularity, see [12]. From

the computational point of view, the subadditivity properties of these costs allow for some efficient algorithms using local indicators, at least in the one dimensional discrete case, see [7].

Moreover, under strict convexity assumptions, these costs satisfy a strict triangle inequality (see Lemma 2.1). This last fact implies (see Proposition 2.2, but it is a classical fact) that the common mass between μ and ν must stay at rest. This gives a first constraint on how to build optimal plans γ : look at μ and ν , take the common part $\mu \wedge \nu$, leave it on place, subtract it from the rest, and then build an optimal transport between the two remainders, which will have no mass in common. Notice that when the cost c is linear in the Euclidean distance, then the common mass *may* stay at rest but is not forced to do so (the very well known example is the transport from $\mu = \mathcal{L}^1 \llcorner [0, 1]$ and $\nu = \mathcal{L}^1 \llcorner [\frac{1}{2}, \frac{3}{2}]$, where both $T(x) = x + \frac{1}{2}$ and $T(x) = x + 1$ on $[0, \frac{1}{2}]$ and $T(x) = x$ on $[\frac{1}{2}, 1]$ are optimal); on the contrary, when the cost is strictly convex in the Euclidean distance, in general the common mass *does not* stay at rest (in the previous example only the translation is optimal for $c(x, y) = |x - y|^p$, $p > 1$). Notice that the fact that the common mass stays at rest implies that in general there is no optimal map T , since whenever there is a set A with $\mu(A) > (\mu \wedge \nu)(A) = \nu(A)$ then almost all the points of A must have two images: themselves, and another point outside A .

Yet, this suggests to study the case where μ and ν are mutually singular, and the best one can do would be proving the existence of an optimal map in this case. This is a good point, since the singularity of the function $(x, y) \mapsto \ell(|x - y|)$ is mainly concentrated on the diagonal $\{x = y\}$ (look at the example $|x - y|^\alpha$), and when the two measures have no common mass almost no point x is transported to $y = x$. Yet, exploiting this fact needs some attention.

First, a typical assumption on the starting measure is required: we need to suppose that μ does not give mass to $(d-1)$ -rectifiable sets. This is standard and common with other costs, such as the quadratic one. From the technical point of view, this is needed in order to guarantee μ -a.e. differentiability of the Kantorovich potential, and counter-examples are known without this assumption (see next Section both for Kantorovich potentials and for counter-examples).

Hence, if we add this assumption on μ , the easiest case is when μ and ν have disjoint supports, since in this case there is a lower bound on $|x - y|$ and this allow to stay away from the singularity. Yet, $\text{supp } \mu \cap \text{supp } \nu = \emptyset$ is too restrictive, since even in the case where μ and ν have smooth densities f and g it may happen that, after subtracting the common mass, the two supports meet on the region $\{f = g\}$.

The problem has actually been solved in one of the first papers about optimal transportation, written by Gangbo and McCann in 1996, [9], where they choose the slightly less restrictive assumption $\mu(\text{supp}(\nu)) = 0$. This assumption covers the example above of two continuous densities, but does not cover other cases such as μ being the Lebesgue measure on a bounded domain Ω and ν being an atomic measure with an atom at each rational point, or other examples that one can build with fully supported absolutely continuous measures concentrated on disjoint sets A and $\Omega \setminus A$ (see Section 5.3). The present paper completes the proof by Gangbo and McCann, making use of recent ideas on optimal transportation to tackle the general case; i.e. we solve the problem under the only assumption that μ and ν are singular to each other (and that μ does not give mass to “small” (i.e. $(d-1)$ -rectifiable) sets). From the case of mutually singular measures we can deduce how to deal with the case of measures with a common . The title of the paper exactly refers to this

fact: by “characterization of optimal transport plans” we mean “understanding their structure, composed by a diagonal part and a transport map out of the diagonal”, the word “full” stands for the fact that we arrive to the minimal set of assumptions with respect to previous works, and by “concave costs” we indeed mean “costs which are strictly concave and increasing functions of the euclidean distance”.

In Section 2 we present the main tools that we need. Section 2.1 is devoted to well-known facts from optimal transport theory: we recall the usual strategy to prove the existence of an optimal T based on the Kantorovich potential and, after proving that the common mass stays at rest via a c -cyclical monotonicity argument, we adapt them to the concave case. Section 2.2 recalls the notion and some properties of approximate gradients, which we will use in the following Section.

In Section 3 we start generalizing Gangbo and McCann’s result, in the case where μ is absolutely continuous. In this case we use the fact that almost every point x is sent to a different point $y \neq x$ together with density points argument in order to prove that the Kantorovich potential is approximately differentiable almost everywhere, according to the notions that we presented in Section 2.2. Notice that this strategy is very much linked to many arguments recently used in optimal transportation in [4, 6], where the existence of an optimal map is proven by restricting the transport plan γ to a suitable set of Lebesgue points which is c -cyclically monotone. Here we do not need to address explicitly such a construction but the idea is very much similar. The result we obtain in this section is not contained in [9] but it does not contain their result neither.

It is in Section 4 that we consider the most general case of an arbitrary measure μ not giving mass to small sets. The approximate differentiability of the potential in Section 3 was based on Lebesgue points arguments which require to be adapted to this new framework. This is why we present an interesting Geometric Measure Theory Lemma (Lemma 4.1) which states that, whenever μ does not give mass to small sets, then μ -almost every point x is such that every cone with vertex at x , of arbitrary size, direction and opening, has positive mass for μ . This lemma is not new, but it is surprisingly not so well-known, at least in this very formulation, in the geometric measure theory community (a weaker version of this lemma is actually contained in the classical book [8]). On the contrary, it starts being popular in the optimal transport community (see [5], where the authors prove it and say that it has been presented to them by T. Rajala), which is strange if we think that it is really a GMT statement. Yet, even if not concerned directly with optimal transport, it has been popularized thanks to its applications in optimal transport theory.

For the sake of completeness, we present the proof of this lemma in Section 4.1, even if we stress that the contribution of the paper is not in such a proof; but in the applications of this result to the differentiability of the Kantorovich potential that we face. This is what we do in Section 4.2, where we define an ad-hoc notion of gradient for the Kantorovich potential, by using this density result to prove that it is well-defined μ -a.e. and that it satisfies all the properties we need. Finally, we prove that the optimal γ is concentrated on a graph.

In this way we can now state the main theorem of this paper. Define $\mu \wedge \nu$ as the maximal positive measure which is both less or equal than μ and than ν , and $(\mu - \nu)_+ = \mu - \mu \wedge \nu$, so that the two measures μ and ν uniquely decompose into a common part $\mu \wedge \nu$ and two mutually singular parts $(\mu - \nu)_+$ and $(\nu - \mu)_+$.

Theorem 1.1. *Suppose that μ and ν are probability measures on \mathbb{R}^d such that $(\mu - \nu)_+$ gives no mass to all $(d-1)$ -rectifiable sets, and take the cost $c(x, y) = \ell(|x-y|)$,*

for $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ strictly concave and increasing. Then there exists a unique optimal transport plan γ , and it has the form $(id, id)_\#(\mu \wedge \nu) + (id, T)_\#(\mu - \nu)_+$.

This theorem is obtained as a corollary of Theorem 4.6 which concerns measures with no common mass, whereas Theorem 3.3 is a simplified version of the same statement, where the assumption that μ is absolutely continuous plays an important role.

The paper ends with an appendix in two parts. One explains that, differently from convex costs, in the case of concave costs translations are never optimal, while the second one presents a discussion about the possibility of defining a sort of approximate gradient adapted to the measure μ .

2. TOOLS

2.1. General facts on optimal transportation. We start the preliminaries of this paper with some important and well-known facts about Kantorovich linear programming problem

$$(2.1) \quad (K) \quad \min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c \, d\gamma : \gamma \in \Pi(\mu, \nu) \right\},$$

Here the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ is supposed to be continuous. We can suppose the supports of the two measures to be compact, for simplicity, but we do not need it (yet, in this case it is better to suppose c to be uniformly continuous).

We need to underline some main aspects of this problem (K). First, as any linear programming problem, it admits a dual problem, which reads

$$(2.2) \quad (D) \quad \max \left\{ \int \phi \, d\mu + \int \psi \, d\nu : \phi(x) + \psi(y) \leq c(x, y) \text{ for all } (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \right\},$$

where the supremum is computed over all pairs (ϕ, ψ) of continuous functions on \mathbb{R}^d . We refer to [15] for this duality relation. Here are some of the properties of (K), (D) and the connection between them:

- (K) admits at least a solution γ , called *optimal transport plan*;
- (D) also has a solution (ϕ, ψ) ;
- The functions (ϕ, ψ) are such that

$$\psi(y) = \inf_x c(x, y) - \phi(x) \quad \text{and} \quad \phi(x) = \inf_y c(x, y) - \psi(y)$$

(we say that they are conjugate to each other and $\phi = \psi^c$ is the c -transform of ψ and $\psi = \phi^c$ is the c -transform of ϕ). Hence, the dual problem can be expressed in terms of one only function ϕ , taking $\psi = \phi^c$ (which automatically implies the constraint $\phi(x) + \phi^c(y) \leq c(x, y)$). Any optimal ϕ is called *Kantorovich potential*.

- Given γ and a pair (ϕ, ψ) then γ is optimal for (K) and (ϕ, ψ) is optimal for (D) if and only if the equality $\phi(x) + \psi(y) = c(x, y)$ holds for all $(x, y) \in \text{supp } \gamma$.

An interesting consequence of this last property is the fact that the support Γ of any optimal γ is c -cyclically monotone.

Definition 1. Given a function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, we say that a set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is c -cyclically monotone (briefly c -CM) if, for every $k \in \mathbb{N}$, every permutation σ

of k elements and every finite family of points $(x_1, y_1), \dots, (x_k, y_k) \in \Gamma$ we have

$$\sum_{i=1}^k c(x_i, y_i) \leq \sum_{i=1}^k c(x_i, y_{\sigma(i)}).$$

The word ‘‘cyclical’’ refers to the fact that we can restrict our attention to cyclical permutations. The word ‘‘monotone’’ is a left-over from the case $c(x, y) = -x \cdot y$.

Indeed, it is easy to check that Γ is c -CM from the fact that $\sum_{i=1}^k c(x_i, y_i) = \sum_{i=1}^k \phi(x_i) + \psi(y_i) = \sum_{i=1}^k \phi(x_i) + \psi(y_{\sigma(i)}) \leq \sum_{i=1}^k c(x_i, y_{\sigma(i)})$. This fact has a very interesting consequence in the case we consider in this paper, i.e. $c(x, y) = \ell(|x - y|)$ with ℓ increasing and strictly concave.

Lemma 2.1. *Let $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a strictly concave and increasing function with $\ell(0) = 0$. Then ℓ is strictly subadditive, i.e. $\ell(s + t) < \ell(s) + \ell(t)$ for every $s, t > 0$. Also, if x, y, z are points in \mathbb{R}^n with $x \neq y$ and $y \neq z$, then $\ell(|x - z|) < \ell(|x - y|) + \ell(|y - z|)$.*

Proof. The subadditivity of positive concave functions is a classical fact which can be proven in the following way. Take $t, s > 0$ and consider the function $g : [0, t + s] \rightarrow \mathbb{R}_+$ defined through $g(r) = \ell(t + s - r) + \ell(r)$. Then g is strictly concave (as a sum of two strictly concave functions), and hence its minimal value is attained (only) on the boundary of the interval. Since $g(0) = g(t + s) = \ell(t + s) + \ell(0) = \ell(t + s)$ we get $\ell(s) + \ell(t) = g(t) > \min g = g(0) = \ell(t + s)$.

The second part of the statement is an easy consequence of the triangle inequality and the monotonicity of ℓ :

$$\ell(|x - z|) \leq \ell(|x - y| + |y - z|) < \ell(|x - y|) + \ell(|y - z|). \quad \square$$

This can be applied to the study of optimal transport plans.

Theorem 2.2. *Let γ be an optimal transport plan for the cost $c(x, y) = \ell(|x - y|)$ with $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ strictly concave, increasing, and such that $\ell(0) = 0$. Let $\gamma = \gamma_D + \gamma_O$, where γ_D is the restriction of γ to the diagonal $D = \{(x, x) : x \in \mathbb{R}^d\}$ and γ_O is the part outside the diagonal, i.e. the restriction to $D^c = (\mathbb{R}^d \times \mathbb{R}^d) \setminus D$. Then this decomposition is such that $(\pi_x)_\# \gamma_O$ and $(\pi_y)_\# \gamma_O$ are mutually singular measures.*

Proof. It is clear that γ_O is concentrated on $\text{supp } \gamma \setminus D$ and hence $(\pi_x)_\# \gamma_O$ is concentrated on $\pi_x(\text{supp } \gamma \setminus D)$ and $(\pi_y)_\# \gamma_O$ is concentrated on $\pi_y(\text{supp } \gamma \setminus D)$. We claim that these two sets are disjoint. Indeed suppose that a common point z belongs to both. Then, by definition, there exists y such that $(z, y) \in \text{supp } \gamma \setminus D$ and x such that $(x, z) \in \text{supp } \gamma \setminus D$. This means that we can apply c -cyclical monotonicity to the points (x, z) and (z, y) and get

$$\ell(|x - z|) + \ell(|z - y|) \leq \ell(|x - y|) + \ell(|z - z|) = \ell(|x - y|) < \ell(|x - z|) + \ell(|z - y|),$$

where the last strict inequality, justified by Lemma 2.1, gives a contradiction. \square

As we said in the introduction, the theorem above has some important consequences. In particular, it states that for this class of ‘‘concave’’ costs the common mass between μ and ν must stay at rest, which we can explain in details. Indeed, we have $(\pi_x)_\# \gamma_D = (\pi_y)_\# \gamma_D$ and since the remaining parts $(\pi_x)_\# \gamma_O$ and $(\pi_y)_\# \gamma_O$ are mutually singular, the decompositions $\mu = (\pi_x)_\# \gamma_D + (\pi_x)_\# \gamma_O$ and $\nu = (\pi_y)_\# \gamma_D + (\pi_y)_\# \gamma_O$ imply $(\pi_x)_\# \gamma_D = (\pi_y)_\# \gamma_D = \mu \wedge \nu$, $(\pi_x)_\# \gamma_O = (\mu - \nu)_+$

and $(\pi_y)_\# \gamma_O = (\nu - \mu)_+$. Hence, we must look in particular at the optimal transport problem between the two mutually singular measures $(\mu - \nu)_+$ and $(\mu - \nu)_-$. We will show that, under some natural regularity assumptions on the starting measure, this problem admits the existence of an optimal map.

From now on, we will just assume w.l.o.g. that μ and ν are mutually singular. If not, just remove the common part, since we can deal with it separately.

Let us see which is the general strategy for proving existence of an optimal T when the cost c is of the form $c(x, y) = h(x - y)$ (not necessarily depending only on the norm $|x - y|$). Here as well the duality plays an important role. Indeed, if we consider an optimal transport plan γ and a Kantorovich potential ϕ with $\psi = \phi^c$, we may write

$$\phi(x) + \psi(y) \leq c(x, y) \text{ on } \mathbb{R}^d \times \mathbb{R}^d \text{ and } \phi(x) + \psi(y) = c(x, y) \text{ on } \text{supp } \gamma.$$

Once we have that, let us fix a point $(x_0, y_0) \in \text{supp } \gamma$. One may deduce from the previous computations that

$$(2.3) \quad x \mapsto \phi(x) - h(x - y_0) \quad \text{is maximal at } x = x_0$$

and, if ϕ is differentiable at x_0 and x_0 is a interior point, one gets $\nabla \phi(x_0) = \nabla h(x_0 - y_0)$ (we also assume that h is differentiable at $x_0 - y_0$). The easiest case is that of a strictly convex function h , since one may inverse the relation passing to ∇h^* thus getting

$$x_0 - y_0 = \nabla h^*(\nabla \phi(x_0)).$$

This allows to express y_0 as a function of x_0 , thus proving that there is only one point $(x_0, y) \in \text{supp } \gamma$ and hence that γ comes from a transport $T(x) = x - \nabla h^*(\nabla \phi(x))$. It is possible to see that strict convexity of h is not the important assumption, but we need ∇h to be injective, which is also the case if $h(z) = \ell(|z|)$ since $\nabla h(z) = \ell'(|z|) \frac{z}{|z|}$ and the modulus of this vector identifies the modulus $|z|$ (since ℓ' is strictly increasing) and the direction gives also the direction of z . We will deal later with the case where ℓ is not differentiable at some points, and it will not be so difficult. The only difficult point is 0 since h would be highly singular at the origin, but fortunately if μ and ν have no mass in common then the case $x_0 = y_0$ will be negligible.

However, we need to guarantee that ϕ is differentiable a.e. with respect to μ . This is usually guaranteed by requiring μ to be absolutely continuous with respect to the Lebesgue measure, and using the fact that $\phi(x) = \inf_y h(|x - y|) - \psi(y)$ allows to prove Lipschitz continuity of ϕ if h is Lipschitz continuous. This is indeed the main difficulty: concave functions on \mathbb{R}^+ may have an infinite slope at 0 and be non-Lipschitz. The fact that the case of distance 0 is negligible is not enough to restrict h to a set where it is Lipschitz, if we do not have a lower bound on $|x - y|$. This can be easily obtained if one supposes that $\text{supp } \mu \cap \text{supp } \nu = \emptyset$, but this is in general not the case when μ and ν are obtained from the two original measures by removing the common mass. The paper by Gangbo and McCann makes a slightly less restrictive assumption, i.e. that $\mu(\text{supp } \nu) = 0$. This allows to say that μ -almost any point is far from the support of ν and, since we only need local Lipschitz continuity, this is enough.

In the next two sessions we will develop two increasingly hard arguments to provide μ -a.e. differentiability for ϕ . The next session will assume μ to be absolutely continuous and will make use of the notion of approximate gradient. Then, we will weaken the assumptions on μ supposing only that it does not give mass to $(d-1)$ -rectifiable sets. This will require a different ad-hoc notion of gradient and a

geometric measure theory lemma which replaces the notion of Lebesgue points. The first proof in the case $\mu \ll \mathcal{L}^d$ is given for the sake of simplicity and completeness. A consequence of this all is the existence of a transport map for $c(x, y) = \ell(|x - y|)$ every time that

- μ and ν are mutually singular
- μ does not give mass to “small sets”.

This map can also be proven to be unique, as it is the case every time that we can prove that every optimal plan is indeed induced by a map. Moreover, if we only suppose b) and we admit the existence of a common mass between μ and ν (actually it is even enough to suppose that $\mu - \mu \wedge \nu$ does not give mass to “small sets”), then the optimal γ is unique and composed of two parts: one on the diagonal and one induced by a transport map, which implies that every point has at most two images, one of the two being the point itself.

We finish this section with a - standard - counterexample which shows that the assumption that μ does not give mass to “small sets” is natural. Indeed, one can consider

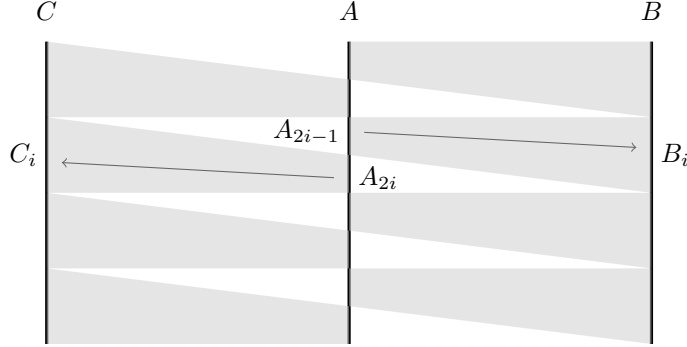
$$\mu = \mathcal{H}^1 \llcorner A \quad \text{and} \quad \nu = \frac{\mathcal{H}^1 \llcorner B + \mathcal{H}^1 \llcorner C}{2}$$

where A , B and C are three vertical parallel segments in \mathbb{R}^2 whose vertexes lie on the two line $y = 0$ and $y = 1$ and the abscissas are 0, 1 and -1 , respectively, and \mathcal{H}^1 is the 1-dimensional Hausdorff measure. It is clear that no transport plan may realize a cost better than 1 since, horizontally, every point needs to be displaced of a distance 1. Moreover, one can get a sequence of maps $T_n : A \rightarrow B \cup C$ by dividing A into $2n$ equal segments $(A_i)_{i=1, \dots, 2n}$ and B and C into n segments each, $(B_i)_{i=1, \dots, n}$ and $(C_i)_{i=1, \dots, n}$ (all ordered downwards). Then define T_n as a piecewise affine map which sends A_{2i-1} onto B_i and A_{2i} onto C_i . In this way the cost of the map T_n is less than $\ell(1 + 1/n)$, which implies that the infimum of the Kantorovich problem is $\ell(1)$, as well as the infimum on transport maps only. Yet, no map T may obtain a cost $\ell(1)$, as this would imply that all points are sent horizontally, but this cannot respect the push-forward constraint. On the other hand, the transport plan associated to T_n weakly converge to the transport plan $\frac{1}{2}T_{\#}^+ \mu + \frac{1}{2}T_{\#}^- \mu$, where $T^{\pm}(x) = x \pm e$ and $e = (1, 0)$. This transport plan turns out to be the only optimal transport plan and its cost is $\ell(1)$.

In this example we have two measures with disjoint supports and no optimal transport map (every point x is forced to have two images $x \pm e$), but we can add a common mass, for instance taking

$$\mu = \mathcal{H}^1 \llcorner A \quad \text{and} \quad \nu = \frac{\mu}{2} + \frac{\mathcal{H}^1 \llcorner B + \mathcal{H}^1 \llcorner C}{4}$$

and in such a case every point should have three images (part of the mass at x is sent at $x \pm e$ and part stays at x).



2.2. The approximate gradient. In this section we recall some notions about a measure-theoretical notion replacing the gradient for less regular functions. The interested reader can find many details in [3].

Let us start from the following observation: given a function $f : \Omega \rightarrow \mathbb{R}$ and a point $x_0 \in \Omega$, we say that f is differentiable at $x_0 \in \Omega$ and that its gradient is $\nabla f(x_0) \in \mathbb{R}^d$ if for every $\epsilon > 0$ the set

$$\{x \in \Omega : |f(x) - f(x_0) - \nabla f(x_0) \cdot (x - x_0)| > \epsilon |x - x_0|\}$$

is at positive distance from x_0 , i.e. if there exist a ball around x_0 which does not meet it. Instead of this requirement, we could ask for a weaker condition, namely that x_0 is a zero-density point for the same set (i.e. a Lebesgue point of its complement). More precisely, if there exists a vector v such that

$$\lim_{\delta \rightarrow 0} \frac{|\{x \in \Omega : |f(x) - f(x_0) - v \cdot (x - x_0)| > \epsilon |x - x_0|\}|}{|B(x_0, \delta)|} = 0$$

then we say that f is approximately differentiable at x_0 and its approximate gradient is v . The approximate gradient will be denoted by $\nabla_{app} f(x_0)$. As one can expect, it enjoys several of the properties of $v(x_0)$, that we list here.

- The approximate gradient, provided it exists, is unique.
- The approximate gradient is nothing but the usual gradient if f is differentiable.
- The approximate gradient shares the usual algebraic properties of gradients, in particular $\nabla_{app}(f + g)(x_0) = \nabla_{app} f(x_0) + \nabla_{app} g(x_0)$.
- If x_0 is a local minimum or local maximum for f , and if $\nabla_{app} f(x_0)$ exists, then $\nabla_{app} f(x_0) = 0$.

these four properties are quite easy to check.

Another very important property that we need is a consequence of the well-know Rademacher theorem, which states that Lipschitz functions are almost everywhere differentiable.

Proposition 2.3. *Let $f, g : \Omega \rightarrow \mathbb{R}$ be two functions defined on a same domain Ω with g Lipschitz continuous. Let $A \subset \Omega$ be a Borel set such that $f = g$ on A . Then f is approximately differentiable almost everywhere on A and $\nabla_{app} f(x) = \nabla g(x)$ for a.e. $x \in A$.*

Proof. It is enough to consider all the points in A which are Lebesgue points of A and at the same time differentiability points of g . These points cover almost all A . It is easy to check that the definition of approximate gradient of f at a point x_0 is satisfied if we take $v = \nabla g(x_0)$. \square

3. THE ABSOLUTELY CONTINUOUS CASE

Suppose now $\mu \ll \text{Leb}$, and proceed according to the strategy presented in Section 2.1. Take $x_0 \in \text{supp}(\mu)$: there exists a point $y_0 \in \mathbb{R}^d$ such that $(x_0, y_0) \in \text{supp}(\gamma)$. Denote by ϕ a Kantorovich potential and by ϕ^c its conjugate function. From Equation (2.3), provided we can differentiate, we get

$$(3.1) \quad 0 = \nabla\phi(x_0) - \nabla\ell(|x_0 - y_0|) = \nabla\phi(x_0) - \ell'(|x_0 - y_0|) \frac{x_0 - y_0}{|x_0 - y_0|},$$

so that x_0 uniquely determines y_0 , but unfortunately neither ϕ nor ℓ are smooth enough to differentiate. To comply with this, we first want to prove ϕ admits an approximate gradient Lebesgue-a.e. From what we saw in Section 2.2 this would imply (if ℓ is differentiable, we will see later how to handle the case where it is not) that Equation (3.1) is satisfied if we replace the gradient with the approximate gradient.

Recall that we may suppose

$$\phi(x) = \phi^{cc}(x) = \inf_{y \in \mathbb{R}^d} \ell(|x - y|) - \phi^c(y).$$

Now consider a countable family of closed balls B_i generating the topology of \mathbb{R}^d , and for every i consider the function defined as

$$\phi_i(x) := \inf_{y \in B_i} \ell(|x - y|) - \phi^c(y).$$

for $x \in \mathbb{R}^n$. One cannot provide straight Lipschitz properties for ϕ_i , since a priori y is arbitrarily close to x and in general ℓ is not Lipschitz close to 0. However ϕ_i is Lipschitz on every B_j such that $\text{dist}(B_i, B_j) > 0$. Indeed if $x \in B_j$, $y \in B_i$ one has $|x - y| \geq d > 0$, therefore the Lipschitz constant of $\ell(|\cdot - y|) - \phi^c(y)$ does not exceed $\ell'(d)$ (or the right derivative of ℓ at d). It follows that ϕ_i is Lipschitz on B_j , and its constant does not exceed $\ell'(d)$.

Then, by Prop. 2.3, ϕ has an approximate gradient almost everywhere on $\{\phi = \phi_i\} \cap B_j$. By countable union, ϕ admits an approximate gradient a.e. on

$$\bigcup_{\substack{i,j \\ d(B_i, B_j) > 0}} [\{\phi_i = \phi\} \cap B_j].$$

As a consequence of this and of the absolute continuity of μ , in order to prove that ϕ has an approximate gradient μ -almost everywhere, it is enough to prove that

$$\mu \left(\bigcup_{\substack{i,j \\ d(B_i, B_j) > 0}} \{\phi_i = \phi\} \cap B_j \right) = 1.$$

Lemma 3.1. *For every i and j*

$$\pi_x(\text{supp } \gamma \cap (B_j \times B_i)) \subset \{\phi = \phi_i\} \cap B_j.$$

Proof. Let $(x, y) \in \text{supp } \gamma \cap (B_j \times B_i)$. Then $\phi(x) + \phi^c(y) = \ell(|x - y|)$. It follows that

$$\phi_i(x) = \inf_{y' \in B_i} \ell(|x - y'|) - \phi^c(y') \leq \ell(|x - y|) - \phi^c(y) = \phi(x).$$

On the other hand, for every $x \in \mathbb{R}^n$

$$\phi_i(x) = \inf_{y \in B_i} \ell(|x - y|) - \phi^c(y) \geq \inf_{y \in \mathbb{R}^n} \ell(|x - y|) - \phi^c(y) = \phi(x). \quad \square$$

As a consequence of this,

$$\begin{aligned}
\mu \left(\bigcup_{\substack{i,j \\ d(B_i, B_j) > 0}} \{\phi_i = \phi\} \cap B_j \right) &\geq \mu \left(\bigcup_{\substack{i,j \\ d(B_i, B_j) > 0}} \pi^x(\text{supp } \gamma \cap (B_j \times B_i)) \right) \\
&= \mu \left(\pi^x \left(\text{supp } \gamma \cap \bigcup_{\substack{i,j \\ d(B_i, B_j) > 0}} B_j \times B_i \right) \right) \\
&= \mu(\pi^x(\text{supp } \gamma \setminus D)) \\
&= \gamma [(\pi^x)^{-1}(\pi^x(\text{supp } \gamma \setminus D))] \\
&\geq \gamma(\text{supp } \gamma \setminus D) = 1
\end{aligned}$$

since the diagonal is γ -negligible. In other words the following theorem is proved

Theorem 3.2. *Let $\ell : [0, +\infty) \rightarrow \mathbb{R}_+$ be a concave function, and suppose μ and ν are two probability measures on \mathbb{R}^n , with $\mu \ll \text{Leb}$. Call ϕ a Kantorovitch potential for the transport problem with cost $\ell(|x - y|)$. Then ϕ admits an approximated gradient μ -a.e.*

We now come back to the proof of the main theorem of the paper, under the additional assumption that μ is absolutely continuous.

Theorem 3.3. *Suppose that μ and ν are two mutually singular probability measures on \mathbb{R}^d such that $\mu \ll \mathcal{L}^d$, and take the cost $c(x, y) = \ell(|x - y|)$, for $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ strictly concave and increasing. Then there exists a unique optimal transport plan γ and it is induced by a transport map.*

Proof. We will prove that any optimal transport plan γ is induced by a transport map, which also implies uniqueness by standard techniques (see [15]). From the strategy that has been presented in Section 2.1, and since we know that γ is concentrated outside the diagonal D , it is enough to prove that, if $(x_0, y_0) \in \text{supp } \gamma \setminus D$, then y_0 is uniquely determined by x_0

- Case 1: ℓ differentiable at $|x_0 - y_0|$
Let $(x_0, y_0) \in \text{supp } \gamma \setminus D$, and suppose that ℓ is differentiable at $|x_0 - y_0|$. Then, if ϕ is approximately differentiable at x_0 (which is true μ -a.e.)

$$0 = \nabla_{app} \phi(x_0) - \ell'(|x_0 - y_0|) \frac{x_0 - y_0}{|x_0 - y_0|}.$$

Thus, since ℓ was supposed strictly increasing and strictly concave,

$$|x_0 - y_0| = (\ell')^{-1}(\nabla_{app} \phi(x_0)),$$

and

$$(3.2) \quad \frac{y_0 - x_0}{|y_0 - x_0|} = -\frac{\nabla_{app} \phi(x_0)}{\ell'(|x_0 - y_0|)}.$$

In other words y_0 is uniquely determined by x_0 .

- Case 2: ℓ not differentiable at $|x_0 - y_0|$
This second case is even more striking. Consider the values of the right and left derivatives $\ell'_r(|x_0 - y_0|)$ and $\ell'_l(|x_0 - y_0|)$ and pick a value $p \in I := [\ell'_r(|x_0 - y_0|), \ell'_l(|x_0 - y_0|)]$. We know by concavity that we have

$$l(t) - l(|x_0 - y_0|) \leq p(t - |x_0 - y_0|) \quad \text{for all } t \in \mathbb{R}.$$

Consider the function

$$\phi(\cdot) - [l(|x_0 - y_0|) + p(|\cdot - y_0| - |x_0 - y_0|)]$$

defined on \mathbb{R}^n . Then, for every $x \in \mathbb{R}^n$,

$$\begin{aligned} \phi(x) - [l(|x_0 - y_0|) + p(|x - y_0| - |x_0 - y_0|)] \\ \leq \phi(x) - l(|x - y_0|) \leq \phi(x_0) - l(|x_0 - y_0|) \\ = \phi(x_0) - [l(|x_0 - y_0|) + p(|x_0 - y_0| - |x_0 - y_0|)]. \end{aligned}$$

In other words, $\phi(\cdot) - [l(|x_0 - y_0|) + p(|\cdot - y_0| - |x_0 - y_0|)]$ has a maximum in x_0 . Since $x_0 \neq y_0$,

$$0 = \nabla_{app}\phi(x_0) - p \frac{x_0 - y_0}{|x_0 - y_0|}.$$

It follows that $|\nabla_{app}\phi(x_0)| = p$ for every $p \in I$, and this leads to a contradiction.

This means that this second case can only occur on a negligible set of points x_0 (those where ϕ is not approximately differentiable). \square

Remark 1. The present paper is fully written under the assumption that ℓ is strictly concave and increasing, which implies $\ell' > 0$. Yet, most of the analysis could be adapted to the case where ℓ is only supposed to be strictly concave, even if monotonicity is indeed very much reasonable from the modelization point of view. One of the problem in case ℓ' is not strictly positive is the fact that we risk to divide by 0 in equation (3.2). This can be fixed since it only happens at a maximum point for ℓ , and c -cyclical monotonicity can be used to prove that $\gamma(\{(x, y) : |x - y| = m\}) = 0$ when $m = \arg \max \ell$ together with a density points argument. Another difficulty to be fixed is the fact that with ℓ not increasing the cost c is not anymore bounded from below, and we would need to assume μ and ν compactly supported. Yet, this is not in the scopes of the present paper.

4. THE GENERAL CASE: μ DOES NOT GIVE MASS TO SMALL SETS

Let us now consider weaker assumptions, i.e. suppose that μ does not give mass to small sets. Namely suppose that, for every $A \subset \mathbb{R}^d$ which is \mathcal{H}^{d-1} -rectifiable, we have $\mu(A) = 0$.

4.1. A GMT lemma for density points. The following lemma is an interesting result from Geometric Measure Theory that can be used instead of Lebesgue points-type results when we face a measure which is not absolutely continuous but “does not give mass to small sets”. It states that, in such a case, μ -a.e. point x is such that every cone exiting from x , even if very small, has positive μ mass. In particular it means that we can find points of $\text{supp } \mu$ in almost arbitrary directions close to x .

We give the proof of this lemma for the sake of completeness, and because it is not easy to find hints or references for it in the literature about GMT. Indeed, the result is known in a part of the optimal transport community, where density points (in particular Lebesgue points) play an important role in some strategies for the existence of optimal maps (see [4, 6]), and it has recently been detailed in [5]. It is also part of the folklore of some branches of the GMT community, but, also in this case, there is no written evidence of it in this form and under these assumptions. Indeed, Lemma 3.3.5 in [8] proves the $(d-1)$ -rectifiability of any set such that every point admits the existence of a small two-sided cone containing no other point of the set, and this would allow to prove the statement we need, up

to the fact that in our case we only consider one-sided cones. The proof follows anyway the same main ideas, and the one that we propose here is especially written in terms of μ -negligibility.

Lemma 4.1. *Let μ be a Borel measure on \mathbb{R}^d , and suppose that μ does not charge small sets. Then μ is concentrated on the set*

$$\{x : \forall \epsilon > 0, \forall \delta > 0, \forall u \in \mathbb{S}^{d-1}, \mu(C(x, u, \delta, \epsilon)) > 0\} ,$$

where

$$C(x, u, \delta, \epsilon) = C(x, u, \delta) \cap \overline{B(x, \epsilon)} := \{y : \langle y - x, u \rangle \geq (1 - \delta) |y - x| \} \cap \overline{B(x, \epsilon)}$$

Proof. Equivalently we will prove that

$$\mu \left(\bigcup_{u, \delta, \epsilon} \{x : \mu(C(x, u, \delta, \epsilon)) = 0\} \right) = 0 .$$

First notice that u , δ and ϵ may be taken each in a countable set. This means that it is enough to fix then *una tantum* u , δ and ϵ and then prove that

$$\mu(\{x : \mu(C(x, u, \delta, \epsilon)) = 0\}) = 0 .$$

Moreover, for sake of simplicity, suppose $u = (0, \dots, -1)$. Take now all cubes Q with sides parallel to the coordinate hyperplanes, centered in a point of \mathbb{Q}^d and with sidelength belonging to \mathbb{Q}_+ , and call the set of such cubes $\{Q_n\}$. We can see that

$$\{x : \mu(C(x, u, \delta, \epsilon)) = 0\} \subset \bigcup_n \{x \in Q_n : \mu(Q_n \cap C(x, u, \delta)) = 0\} .$$

Therefore we will show, for a fixed cube Q , that

$$\mu(\{x \in Q : \mu(Q \cap C(x, u, \delta)) = 0\}) = 0 .$$

Now write every $y \in \mathbb{R}^d$ as $y = (y', y_d)$, with $y' \in \mathbb{R}^{d-1}$. Then a quick computation shows that $(y', y_d) \in C((x', x_d), u, \delta)$ if and only if

$$y_d \leq x_d - \frac{1 - \delta}{\sqrt{\delta(2 - \delta)}} |y' - x'| ,$$

and we set $k(\delta) := \frac{1 - \delta}{\sqrt{\delta(2 - \delta)}}$.

Define now X as the projection of Q along the first $d - 1$ coordinates, and for every $x' \in X$

$$z(x') := \sup\{z \in \mathbb{R} : \mu(Q \cap C((x', z), u, \delta)) = 0\} .$$

Notice that the set in the supremum is never empty, by taking for instance $z \leq \min\{x_d : x \in Q\}$.

Let us study the function $x' \mapsto z(x')$ and notice that it is $k(\delta)$ -Lipschitz continuous: indeed, if we had $z(x'_1) < z(x'_0) - k(\delta)|x'_1 - x'_0|$, then, the cone $C((x'_1, z(x'_1)), u, \delta)$ would be included in the interior of $C((x'_0, z(x'_0)), u, \delta)$ and hence the same would be true for $C((x'_1, z(x'_1) + t), u, \delta)$ for small $t > 0$. Yet, this implies $\mu(C((x'_1, z(x'_1) + t), u, \delta)) = 0$ and hence the supremum defining $z(x'_1)$ should be at least $z(x'_1) + t$, which is obviously a contradiction. By interchanging the role of x'_1 and x'_0 we get $|z(x'_1) - z(x'_0)| \leq k(\delta)|x'_1 - x'_0|$.

Now we take

$$\begin{aligned} \{x \in Q : \mu(Q \cap C(x, \delta, u)) = 0\} &= \{(x', x_d) : x_d \leq z(x')\} \\ &= \{(x', x_d) : x_d < z(x')\} \cup \{(x', x_d) : x_d = z(x')\}. \end{aligned}$$

The second set on the right hand side is μ -negligible since it is the graph of a Lipschitz function of $(d-1)$ variables and μ is supposed not to give mass to these sets. The first set is also negligible since it is contained in the complement of $\text{supp } \mu$. Indeed, if a point $x = (x', x_d)$ satisfies $x_d < z(x')$, then we also have $x_d + t < z(x')$ for small $t > 0$ and the interior of the cone $C((x', z(x') + t), u, \delta)$ is μ -negligible. This means that x has a neighborhood where there is no mass of μ , and hence that it does not belong to $\text{supp } \mu$.

This ends the proof. \square

Notice that the above result is obviously false if we withdraw the hypothesis on μ : take a measure concentrated on a $(d-1)$ -manifold (for instance, an hyperplane). Then, for small ε and δ , the measure $\mu(C(x, u, \delta, \varepsilon))$ will be zero if u does not belong to the tangent space to the manifold at x .

4.2. How to handle non-absolutely continuous measures μ . Consider again the transport problem, and suppose μ does not charge small sets. Call γ the minimizer of the Kantorovitch problem, ϕ a maximizer of the dual problem. As in Section 3, we take a countable family of balls B_i generating the topology of \mathbb{R}^n , and for each i define

$$\phi_i(x) := \inf_{y \in B_i} l(|x - y|) - \phi^c(y), \quad x \in \mathbb{R}^d.$$

Lemma 4.2. *If $B \subset \mathbb{R}^d$ is such that $d(0, B) = d_0 > 0$ and ℓ is concave and increasing, then the function $B \ni x \mapsto \ell(|x|)$ is semi-concave, and, more precisely, $x \mapsto \ell(|x|) - \frac{1}{2}\ell'(d_0)|x|^2$ is concave (where ℓ' denotes the derivative, or right derivative in case ℓ is not differentiable at d_0).*

Proof. We just need to give an upper bound on the second derivatives of $g(x) := \ell(|x|)$, which we will do in the case where ℓ is C^2 . The general result will follow by approximation. Compute

$$\partial_i g(x) = \ell'(|x|) \frac{x_i}{|x|}; \quad \partial_{ij} g(x) = \ell''(|x|) \frac{x_i x_j}{|x|^2} + \ell'(|x|) \frac{|x|^2 \delta_{ij} - x_i x_j}{|x|^3}.$$

Hence, if we denote by \hat{x} the unit vector $x/|x|$, the Hessian of g is composed of two parts: the positive matrix $\hat{x} \otimes \hat{x}$ times the non-positive factor $\ell''(|x|)$, and the positive matrix $(\text{Id} - \hat{x} \otimes \hat{x})$ times the positive factor $\ell'(|x|)$. Since $\text{Id} - \hat{x} \otimes \hat{x} \leq \text{Id}$ (in the sense of positive-definite symmetric matrices), this gives

$$D^2 g \leq \ell'(|x|) \text{Id} \leq \ell'(d_0) \text{Id}$$

and thesis follows. \square

Corollary 4.3. *For every i and j such that $d(B_i, B_j) > 0$, the function ϕ_i is semi-concave and hence μ -a.e. differentiable on B_j .*

Proof. The semiconcavity of ϕ_i follows from its definition as an infimum of semi-concave functions, with the same semicontinuity constant which is $\ell'(d(B_i, B_j))$.

This implies that ϕ_i has the same regularity and differentiability points of convex functions. It is well-known that convex functions are differentiable everywhere but on a $(d-1)$ -rectifiable set: this is a consequence of the more general fact that the set

where a convex-valued monotone multifunction in \mathbb{R}^d takes values of dimension at least k can be covered by countably many $(d - k)$ -dimensional graphs of Lipschitz functions (see [1]), applied to the subdifferential multifunction.

Since μ does not give mass to small sets, the set of non-differentiability points of ϕ_i is also negligible. \square

Remark 2. For each pair i, i' consider the measure

$$\mu_{i,i'} := \mu \llcorner \{\phi = \phi_i = \phi_{i'}\},$$

that is, for every A Borel, $\mu_{i,i'}(A) = \mu(A \cap \{x : \phi(x) = \phi_i(x) = \phi_{i'}(x)\})$. Since for every small set A one has $\mu_{i,i'}(A) \leq \mu(A) = 0$, Lemma 4.1 yields the existence of a μ -negligible set $N_{i,i'} \subset \{\phi = \phi_i = \phi_{i'}\}$ such that for every $x \in \{\phi = \phi_i = \phi_{i'}\} \setminus N_{i,i'}$ and for every $u \in S^{d-1}$, $\delta > 0$, $\epsilon > 0$, one has $\mu_{i,i'}(C(x, u, \delta, \epsilon)) > 0$. In particular it follows that $C(x, u, \delta, \epsilon) \cap \{\phi = \phi_i = \phi_{i'}\} \neq \emptyset$.

This isotropy in the structure of $\{\phi = \phi_i = \phi_{i'}\}$ implies the following

Proposition 4.4. *Consider*

$$N := \bigcup_{i,i'} N_{i,i'} \cup \bigcup_{\substack{i,j \\ d(B_i, B_j) > 0}} \{x \in B_j : \phi_i \text{ not differentiable at } x\},$$

then N is μ -negligible. Moreover, consider $A := (\pi_x)(\text{supp } \gamma \setminus D)$, which is a set with $\mu(A) = 1$: then, for every $x \in A \setminus N$ there exists i such that $\phi(x) = \phi_i(x)$.

Moreover, if $\phi(x) = \phi_i(x) = \phi_{i'}(x)$, then

$$\nabla \phi_i(x) = \nabla \phi_{i'}(x).$$

Proof. The fact that N is negligible follows from Remark 2 and Corollary 4.3.

Let $x \in A \setminus N$. There exists y such that $(x, y) \in \text{supp } \gamma$ and $x \neq y$. It follows that there exist two balls B_i and B_j with $x \in B_j$, $y \in B_i$ and $d(B_i, B_j) > 0$. Then, by Lemma 3.1, $\phi(x) = \phi_i(x)$.

Suppose now, for the sake of simplicity, $\phi(x) = \phi_1(x) = \phi_2(x)$. Then, since $x \notin N$, $v_1 := \nabla \phi_1(x)$ and $v_2 := \nabla \phi_2(x)$ are both well-defined. By contradiction suppose $v_1 \neq v_2$. Then for every $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that for every $\tilde{x} \in B(x, \delta)$

$$|\phi_1(\tilde{x}) - \phi(x) - v_1 \cdot (\tilde{x} - x)| \leq \epsilon |\tilde{x} - x|$$

and

$$|\phi_2(\tilde{x}) - \phi(x) - v_2 \cdot (\tilde{x} - x)| \leq \epsilon |\tilde{x} - x|.$$

Therefore, for such \tilde{x} ,

$$\begin{aligned} & |\phi_1(\tilde{x}) - \phi_2(\tilde{x}) + (v_2 - v_1) \cdot (\tilde{x} - x)| \\ &= |\phi_1(\tilde{x}) - \phi(x) - v_1 \cdot (\tilde{x} - x) - [\phi_2(\tilde{x}) - \phi(x) - v_2 \cdot (\tilde{x} - x)]| \leq 2\epsilon |\tilde{x} - x| \end{aligned}$$

In order to have a contradiction, it is enough to choose \tilde{x} such that

$$\phi_1(\tilde{x}) = \phi_2(\tilde{x})$$

and

$$(v_2 - v_1) \cdot (\tilde{x} - x) > 2\epsilon |\tilde{x} - x|.$$

The latter may be expressed as

$$\frac{(\tilde{x} - x)}{|\tilde{x} - x|} \cdot \frac{v_2 - v_1}{|v_2 - v_1|} > \frac{2\epsilon}{|v_2 - v_1|}.$$

In order to guarantee that this is possible, choose $\epsilon < \frac{|v_2 - v_1|}{2}$. Then, by Remark 2 with $C\left(x, \frac{|v_2 - v_1|}{2}, 1 - \frac{2\epsilon}{|v_2 - v_1|}, \delta(\epsilon)\right)$, we are done. \square

To go on with our proof we need the following lemma.

Proposition 4.5. *Let $f : \Omega \rightarrow \mathbb{R}$ be differentiable at $x_0 \in \Omega$. Suppose $B \subset \Omega$ is such that $x_0 \in B$ and that for every $u \in S^{n-1}$, $\delta > 0$ and $\epsilon > 0$ one has $B \cap C(x_0, u, \delta, \epsilon) \neq \emptyset$. Moreover suppose that x_0 is a local maximum for f on B . Then*

$$\nabla f(x_0) = 0 .$$

Proof. Write $v := \nabla f(x_0)$ and suppose by contradiction $v \neq 0$. Then for every $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that if $|x - x_0| \leq \delta(\epsilon)$

$$f(x) - f(x_0) \geq v \cdot (x - x_0) - \epsilon |x - x_0| .$$

Moreover, from the assumption on B , we may suppose $x \in B \setminus \{x_0\}$,

$$\frac{x - x_0}{|x - x_0|} \cdot \frac{v}{|v|} \geq \frac{1}{2}$$

and

$$|x - x_0| \leq \delta(\epsilon) .$$

Then we have

$$f(x) - f(x_0) \geq |x - x_0| (|v|/2 - \epsilon) .$$

Now choose $\epsilon < \frac{|v|}{2}$. It follows that for every $\eta \leq \delta(\epsilon)$ we may choose $x \in B \setminus x_0$ such that $|x - x_0| \leq \eta$ and

$$f(x) > f(x_0) ,$$

which is a contradiction. \square

Theorem 4.6. *Let μ and ν be Borel probability measures on \mathbb{R}^d , and suppose that μ does not charge small sets. Suppose moreover that μ and ν are mutually singular. Then, if $\ell : [0, +\infty) \rightarrow \mathbb{R}$ is strictly concave and increasing, there exists a unique optimal transport map for the cost $\ell(|x - y|)$.*

Proof. We argue as in Section 3 and accordingly to the usual strategy.

Let us take $x_0 \in A \setminus N$, with $(x_0, y_0) \in \text{supp } \gamma$ and $x_0 \neq y_0$. We just need to show that y_0 is uniquely determined by x_0 . By Proposition 4.4 $\phi(x_0) = \phi_i(x_0)$ for some i . Since

$$x_0 \in \arg \max[\phi(\cdot) - \ell(|\cdot - y_0|)] ,$$

in particular

$$x_0 \in \arg \max[\phi_i(\cdot) - \ell(|\cdot - y_0|)]|_{\{\phi = \phi_i\}} .$$

- Case 1: ℓ differentiable at $|x_0 - y_0|$.

It follows that $\phi_i(\cdot) - \ell(|\cdot - y_0|)$ is differentiable at x_0 . From Lemma 4.5 we have $\nabla \phi_i(x_0) - \nabla \ell(|\cdot - y_0|)|_{x_0} = 0$. Notice that the value of the vector $\nabla \phi_i(x_0)$ only depends on x_0 and not on y_0 , and does not depend on i .

By Remark 2 with $i = i'$ we may apply Proposition 4.5 to yield that

$$\nabla \phi_i(x_0) = \ell'(|x_0 - y_0|) \frac{x_0 - y_0}{|x_0 - y_0|} .$$

It follows that

$$|y_0 - x_0| = (\ell')^{-1}(|\nabla \phi_i(x_0)|)$$

and

$$\frac{x_0 - y_0}{|x_0 - y_0|} = \frac{\nabla \phi_i(x_0)}{\ell'(|x_0 - y_0|)} .$$

- Case 2: ℓ not differentiable at $|x_0 - y_0|$

Here as well we argue as in Section 3: pick a value $p \in I := [\ell'_r(|x_0 - y_0|), \ell'_l(|x_0 - y_0|)]$. Suppose $\phi(x_0) = \phi_i(x_0)$ and consider the function

$$\phi(\cdot) - [l(|x_0 - y_0|) + p(|\cdot - y_0| - |x_0 - y_0|)] .$$

defined on \mathbb{R}^n . Then, $\phi_i(\cdot) - [l(|x_0 - y_0|) + p(|\cdot - y_0| - |x_0 - y_0|)]$ has a maximum point at x_0 . It follows that $|\nabla \phi_i(x_0)| = p$ for every choice of p , and this leads to a contradiction. \square

5. APPENDIX

5.1. Concave costs and translations. Here we wish to analyse a feature of strictly concave costs, namely that they penalize equal displacements.

Proposition 5.1. *Let μ and ν be mutually singular Borel measures on \mathbb{R}^d , and suppose μ does not charge small sets. Let $l : [0, +\infty) \rightarrow \mathbb{R}$ be a C^2 , increasing, strictly concave function, and suppose $\ell''(x) < 0$ for every $x > 0$. Call γ the optimal transport plan with respect to $\ell(|\cdot|)$. Then for every $e \in \mathbb{R}^d \setminus \{0\}$*

$$\gamma(\{(x, x + e) : x \in \mathbb{R}^d\}) = 0 .$$

Proof. For $e \in \mathbb{R}^d$ write $\gamma_e := \gamma \llcorner \{(x, x + e) : x \in \mathbb{R}^d\}$. Suppose by contradiction that for some e the measure γ_e does not vanish. It follows that $\mu_e := (\pi^x)_\# \gamma_e$ is nontrivial as well. Moreover it does not charge small sets, since $\mu_e \leq \mu$. Therefore Theorem 4.1 implies that μ_e is concentrated on

$$(5.1) \quad \{x : \forall \epsilon > 0, \forall \delta > 0, \forall u \in \mathbb{R}^d \text{ such that } |u| = 1, \mu_e(C(x, u, \delta, \epsilon)) > 0\} ,$$

which we will refer to as A_e . Clearly we have

$$A_e \subset \text{supp } \mu_e = \pi_x(\text{supp } \gamma_e) .$$

Now take x_1 and x_2 in $\text{supp } \mu_e$. Then $(x_1, x_1 + e), (x_2, x_2 + e) \in \text{supp } \gamma_e$, which is l -cyclically monotone. If we call $\xi := x_2 - x_1$ it follows that

$$2\ell(|e|) \leq \ell(|e + \xi|) + \ell(|e - \xi|) .$$

Write then a second order Taylor expansion, to yield

$$2\ell(|e|) \leq 2\ell(|e|) + \sum_{i,j} \xi_i \left[\ell''(|e|) \frac{e_i e_j}{|e|^2} + \ell'(|e|) \left(\frac{\delta_{ij}}{|e|} - \frac{e_i e_j}{|e|^3} \right) \right] \xi_j + o(|\xi|^2)$$

or, in other words,

$$\begin{aligned} 0 &\leq \sum_{i,j} \xi_i \left[\ell''(|e|) \frac{e_i e_j}{|e|} + \ell'(|e|) \left(\delta_{ij} - \frac{e_i e_j}{|e|^2} \right) \right] \xi_j + o(|\xi|^2) \\ &= \frac{\ell''(|e|)}{|e|} (e \cdot \xi)^2 + \ell'(|e|) |\xi|^2 - \frac{\ell'(|e|)}{|e|^2} (e \cdot \xi)^2 + o(|\xi|^2) . \end{aligned}$$

Now take any $x_1 \in A_e$, which is nonempty by our hypothesis. Then there exists $x_2 \in \text{supp } \mu_e$ such that $\xi \cdot e \geq (1 - \delta) |\xi| |e|$ with δ arbitrarily small. Then by plugging this into the previous equation, and letting $\xi \rightarrow 0$, we obtain a contradiction. \square

The conclusion of the previous proposition is strongly different from the convex case. Indeed, consider any Borel measure μ on \mathbb{R}^d with compact support, and fix a vector $e \in \mathbb{R}^d \setminus \{0\}$. Define τ_e to be the translation by e and

$$\mu_e := (\tau_e)_\# \mu$$

as the translated μ . Then one may prove that τ_e is an optimal transportation of μ onto μ_e if the cost is convex, while Proposition 5.1 implies that translations are never optimal for concave costs.

In order to prove that τ_e is an optimal transportation of μ onto μ_e , consider any map T which pushes forward μ to μ_e . Then, by Jensen inequality, if $x \mapsto \ell(|x|)$ is convex, we have

$$\begin{aligned} \int c(x, Tx) d\mu &= \int \ell(|x - Tx|) d\mu \geq \ell\left(\left|\int x - Tx d\mu\right|\right) \\ &= \ell\left(\left|\int x d\mu - \int x d\mu_e\right|\right) = \ell(|e|) \\ &= \int \ell(|x - \tau_e(x)|) d\mu. \end{aligned}$$

5.2. On the definition of a weak μ -approximated gradient. Even if not strictly necessary for the sake of the paper, in this section we discuss the possibility of using Lemma 4.1 in order to define a sort of approximated gradient with respect to a measure μ which gives no mass to small sets. The idea is that in Section 4 we produced an ad-hoc choice of gradient, can we extend it to more general frameworks in order to use it in other situations? The answer will be both yes and no.

Before entering into details, let us set the following language.

Definition 2. Let $\mu \in \mathcal{P}(\mathbb{R}^n)$. We define the isotropic set of μ by

$$\text{isot } \mu = \{x : \forall \delta, u, \epsilon, \mu(C(x, u, \delta, \epsilon)) > 0\}.$$

If A is a Borel subset, the μ -isotropic set A is by definition

$$\text{isot}_\mu(A) = A \cap \text{isot}(\mu \llcorner A).$$

We know that, if μ does not charge small sets, then for all Borel subset A (by Lemma 4.1) we have

$$\mu(A \setminus \text{isot}_\mu(A)) = 0.$$

We now pass to a naive definition of gradient.

Definition 3. We say that $L \in \mathbb{R}^n$ is a μ -isotropic gradient of $f : A \rightarrow \mathbb{R}$ at x if for all $\epsilon > 0$, x belongs to the set

$$\text{isot}_\mu(\{y : |f(y) - (f(x) + \langle L, y - x \rangle)| \leq \epsilon |y - x|\}).$$

This definition only means that we can find points $y \in \text{supp } \mu \llcorner D_\epsilon$ where

$$D_\epsilon = \{y : |f(y) - (f(x) + \langle L, y - x \rangle)| \leq \epsilon |y - x|\}$$

in almost arbitrary directions and arbitrary close to x . It satisfies some properties, for instance

Proposition 5.2. *If L is a μ -isotropic gradient of f at x , where f has a local extremum, then $L = 0$.*

We do not give the proof, which is quite similar to that of Proposition 4.5 (indeed, most proofs will be skipped in this appendix subsection, since they simply recall the proofs of Section 4). Yet, this definition is not enough to guarantee uniqueness. Indeed, take two disjoint sets $A, B \subset \mathbb{S}^1$ with $A \cup B = \mathbb{S}^1$, such that the supports of $\mathcal{L}^1 \llcorner A$ and $\mathcal{L}^1 \llcorner B$ are both the whole \mathbb{S}^1 (finding two such sets is a non-trivial, but classical exercise). Now take two different vectors L_A and L_B and define

$$f(x) = \begin{cases} 0 & \text{if } x = 0, \\ L_A \cdot x & \text{if } \frac{x}{|x|} \in A, \\ L_B \cdot x & \text{if } \frac{x}{|x|} \in B. \end{cases}$$

It follows that both L_A and L_B are μ -isotropic gradient of f at any $x \in \mathbb{R}^d$ such that $L_A \cdot x = L_B \cdot x$. Notice that this is a “small set”: it leaves open the question of a possible μ -a.e. uniqueness, but the situation is anyway worse than what happens for the approximate gradient (Section 2), where the gradient is necessarily unique at any point where it exists.

Hence, we try to switch to a better definition of gradient. We rely on the following observation, which is essentially contained in the proof of Proposition 4.4

Proposition 5.3. *If $A \subseteq \Omega \subseteq \mathbb{R}^n$ where A is a Borel subset, Ω open set, and $\phi, \psi : \Omega \rightarrow \mathbb{R}$, then $\nabla \phi(x) = \nabla \psi(x)$ on the set*

$$\text{isot}_\mu(\{x : \phi(x) = \psi(x), \nabla \phi(x) \text{ and } \nabla \psi(x) \text{ exist}\}).$$

The following proposition is an attempt at defining a uniquely determined μ -approximated gradient under hypotheses satisfied by the Kantorovitch potential in Section 4.

Proposition 5.4. *Given $A \subseteq \Omega \subseteq \mathbb{R}^n$, with A Borel and Ω open, given $\mu \in \mathcal{P}(\mathbb{R}^n)$ which does not charge small sets, and $f : A \rightarrow \mathbb{R}^n$ measurable, we say that f is μ -differentiable on A if there exists a countable family $(\phi_n)_{n \in \mathbb{N}}$, $\phi_n : \Omega \rightarrow \mathbb{R}$ such that*

$$\forall x \in A, \exists n \in \mathbb{N}, \quad \phi_n(x) = f(x) \text{ and } \nabla \phi_n(x) \text{ exists.}$$

Then there exists a μ -a.e. unique function $\nabla_\mu f : A \rightarrow \mathbb{R}^n$ enjoying the property

$$(5.2) \quad \begin{aligned} &\text{for all } \phi : \Omega \rightarrow \mathbb{R}, \nabla_\mu f \text{ coincides with } \nabla \phi \text{ on } \mu\text{-a.e. point } x \\ &\text{such that } f(x) = \phi(x) \text{ and } \nabla \phi(x) \text{ exists.} \end{aligned}$$

In order to prove the statement above, one can build $\nabla_\mu f$ as being equal to $\nabla \phi_n$ on almost all the set $\{x : f(x) = \phi_n(x) \text{ and } \nabla \phi_n(x) \text{ exists}\}$, which is well-defined a.e. by Proposition 5.3. The same proposition allows us to check easily that it satisfies the desired property, and is as such unique.

The connection between the two notions we defined is contained in the following statement.

Proposition 5.5. *If $f : A \rightarrow \mathbb{R}$ is μ -differentiable, then for a.e $x \in A$, $\nabla_\mu f(x)$ is a μ -isotropic gradient of f .*

Together with Proposition 5.2, this would allow us to follow the usual strategy described in Section 2.1 to prove the existence of an optimal T based on the Kantorovich potential ϕ and $\nabla_\mu \phi$.

Acknowledgments This paper is part of the work of the ANR project ANR-12-BS01-0014-01 GEOMETRYA, whose support is gratefully acknowledged by the

third author. The work started when the second author was master student at Paris-Sud, financed by *Fondation Mathématique Jacques Hadamard*, which is also gratefully acknowledged for its support.

REFERENCES

- [1] G. ALBERTI AND L. AMBROSIO, A geometrical approach to monotone functions in \mathbb{R}^n , *Math. Z.*, 230 (1999), pp. 259–316.
- [2] Y. BRENIER, Polar factorization and monotone rearrangement of vector-valued functions, *Communications on Pure and Applied Mathematics* 44, 375-417, 1991.
- [3] L. C. EVANS AND R. F. GARIEPY, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.
- [4] T. CHAMPION AND L. DE PASCALE, The Monge problem in \mathbb{R}^d , *Duke Math. J.* Volume 157, Number 3 (2011), 551-572.
- [5] T. CHAMPION AND L. DE PASCALE, On the twist condition and c -monotone transport plans, *Discr. Cont. Dyn. Syst. Ser. A*, Vol 34, No 4, 2014, 1339–1353
- [6] T. CHAMPION, L. DE PASCALE AND P. JUUTINEN, The ∞ -Wasserstein distance: local solutions and existence of optimal transport maps, *SIAM J. of Mathematical Analysis*, 40, (2008), no. 1, 1-20.
- [7] J. DELON, J. SALOMON, A. SOBOLEVSKII, Local matching indicators for transport problems with concave costs, *SIAM J. Disc. Math*, 26 (2), pp. 801-827 (2012).
- [8] H. FEDERER, *Geometric Measure Theory*, Classics in Mathematics, Springer, 1996 (reprint of the 1st ed. Berlin, Heidelberg, New York 1969 edition)
- [9] W. GANGBO, R. MCCANN, The geometry of optimal transportation, *Acta Math.*, 177 (1996), 113–161.
- [10] L.V. KANTOROVICH, On the translocation of masses, *C.R. (Dokl.) Acad. Sci. URSS*, 37 (1942), 199-201.
- [11] L.V. KANTOROVICH, , On a problem of Monge (in Russian), *Uspekhi Mat. Nauk.*, 3 (1948), 225-226.
- [12] X.-N. MA, N. S. TRUDINGER, ET X.-J. WANG Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.*, 177(2) :151–183, 2005.
- [13] G. MONGE, Mémoire sur la théorie des Déblais et des Remblais, *Histoire de l'Académie des Sciences de Paris*, 1781.
- [14] A. PRATELLI, On the sufficiency of c -cyclical monotonicity for optimality of transport plans, *Math. Z.*, 258 (2008), no. 3, 677–690
- [15] C. VILLANI, *Topics in Optimal Transportation*. Graduate Studies in Mathematics, AMS, 2003.

Paul Pegon – Filippo Santambrogio,
 Laboratoire de Mathématiques d’Orsay,
 Université Paris-Sud,
 91405 Orsay cedex, FRANCE,
 paul.pegon@math.u-psud.fr
 filippo.santambrogio@math.u-psud.fr

Davide Piazzoli,
 Cambridge Centre for Analysis,
 University of Cambridge,
 Wilberforce Road,
 Cambridge, CB3 0WB, UK,
 D.Piazzoli@dpmmms.cam.ac.uk