Existence of optimal maps for the L^1 and the L^∞ versions of the Kantorovitch problem

March 7, 2011

This document presents the proof for the existence of an optimal transport plan for the L^1 and L^∞ costs, both by secondary variational techniques, with decomposition or density arguments. In all the document we will suppose that $\Omega \subset \mathbb{R}^d$ is compact and $\mu << \mathcal{L}^d$.

The Monge case, with cost $|x-y|$

The starting point for the L^1 case is the following. The duality formula, in the case of a distance cost function, gives

$$\min\left\{\int |x-y|\,d\gamma,\ \gamma \in \Pi(\mu,\nu)\right\} = \max\left\{\int ud(\mu-\nu)\ :\ u \in \mathrm{Lip}_1\right\}. \tag{0.1}$$

We can now pick a maximizer for the dual problem, which is a Lip_1 function u called *Kantorovitch potential*, that we will consider as fixed from now on. Let us call $O(\mu,\nu)$ the set of optimal transport plans for the cost $|x-y|$. For notational simplicity we will also denote by c_p the functional associating to $\gamma \in \mathcal{P}(\Omega\times\Omega)$ the quantity $\int |x-y|^p d\gamma$, and m_p its minimal value on $\Pi(\mu,\nu)$. In this language $O(\mu,\nu) = \mathrm{argmin}_{\gamma\in\Pi(\mu,\nu)}\, c_1(\gamma) = \{\gamma \in \Pi(\mu,\nu)\ :\ c_1(\gamma) \le m_1\}$. Notice that $O(\mu,\nu)$ is a closed subset (w.r.t. the weak convergence of measures) of $\Gamma(\mu,\nu)$, which is compact. This is a general fact whenever we minimize a semicontinuous functional of γ In this case c_1 is also continuous (w.r.t. the same convergence), and hence the set $\{\gamma \in \Pi(\mu,\nu)\ :\ c_1(\gamma) \le m_p\}$ is closed, but lower semicontinuituy would have been enough. We will use this fact several times.

First, let us notice that it holds

$$\gamma \in O(\mu,\nu) \Leftrightarrow \mathrm{spt}(\gamma) \subset \{(x,y) :\ u(x)-u(y) = |x-y|\}.$$

This is true because optimality implies $\int(u(x)-u(y))d\gamma = \int |x-y|d\gamma$ and the global inequality $u(x)-u(y) \le |x-y|$ gives equality γ−a.e. All these functions being continuous, the equality finally holds on the whole support. Viceversa, equality on the support allows to integrate it and prove that $c_1(\gamma)$ equals the value of the dual problem, which is the same of the primal, hence one gets optimality.

First geometric properties of transport rays.

Let us consider for a while the role played by the function u. We collect some properties.

Lemma 0.1. *If $x, y \in \Omega$ are such that $u(x) - u(y) = |x - y|$, then u is affine on the whole segment $[x, y] := \{z = (1 - t)x + ty,\ t \in [0, 1]\}$.*

Proof. Take $z = (1 - t)x + ty$. Just consider that the Lip_1 condition implies

$$u(x) - u(z) \leq |x - z| = t|x - y|, \quad u(z) - u(y) \leq |z - y| = (1 - t)|x - y|.$$

Summing up the two inequalities we get $u(x) - u(y) \leq |x - y|$, but the assumption is that this should be an equality. Hence we can infer that both inequalities are equalities, and in particular $u(z) = u(x) - t|x - y|$. $\square$

Lemma 0.2. *If $z \in]x, y[:= \{z = (1 - t)x + ty,\ t \in]0, 1[\}$, for a pair of points $x \neq y$ such that $u(x) - u(y) = |x - y|$, then u is differentiable at z and $\nabla u(z) = \frac{x-y}{|x-y|}$.*

The proof can be found in [1].

Definition 1. We call transport ray any non-trivial (i.e. different from a singleton) segment $[x, y]$ such that $u(x) - u(y) = |x - y|$ which is maximal for the inclusion among segments of this form. The corresponding open segment $]x, y[$ is called the interior of the transport ray and x and y its boundary points. We call T_u the union of all non degenerate transport rays, $T_u^{(b)}$ the union of their boundary points and $T_u^{(i)}$ the union of their interiors. Moreover, let $T_u^{(b+)}$ be the set of upper boundary points of non-degenerate transport rays (i.e. those where u is maximal on the transport ray, say the points x in the definition $u(x) - u(y) = |x - y|$) and $T_u^{(b-)}$ the set of lower boundary points of non-degenerate transport rays (where u is minimal, i.e. the points y).

Corollary 0.3. *Two different transport ray can only meet in a point in a point z which is a boundary point for both of them, and in such a case u is not differentiable at z. In particular, if one removes a suitable negligible set (that of non-differentiability points of u), the transport rays are disjoint.*

Proof. Suppose that two transport rays meet at a point z which is internal for both rays. In such a case u must have two different gradients at z, following both the directions of the rays, which is impossible (recall that the different rays meeting at one point must have different directions, since otherwise they are the same transport ray, by maximality).

Suppose now that they meet at z, which is in the interior of a transport ray whose direction is $e = \frac{x-y}{|x-y|}$ but is a boundary point for another ray, with direction $e' \neq e$. In such a case u should be differentiable at z and $\nabla u(z) = e$. Hence, we should have

$$t = u(z + te') - u(z) = e \cdot te' + o(t)$$

(the first equality coming from the behavior of u on the ray stemming from z with direction e' and the second from the definition of $\nabla u(z)$). Yet, this implies $e \cdot e' = 1$ and hence $e = e'$ (since $|e| = |e'| = 1$), which is a contradiction.

Suppose now that the intersection point z is a boundary point for both segments, one with direction e and the other one with direction $e' \neq e$. In this case there is no contradiction but if one supposes that there exists $\nabla u(z) = w$ one gets

$$t = u(z + te) - u(z) = w \cdot te + o(t), \quad t = u(z + te') - u(z) = w \cdot te' + o(t)$$

This implies $w \cdot e = w \cdot e' = 1$ but, since $|w| \leq 1$ (due to $u \in \mathrm{Lip}_1$) and $|e| = |e'| = 1$, we should get $w = e = e'$, which is, again, a contradiction. $\square$

We will see later on that we need to say something more on the direction of the transport rays.

Secondary variational problem.

Since we know that in general there is no uniqueness for the minimizers of c_1 and that they are not all induced by transport maps, we will select a special minimizer, better than the others, and prove that it is actually induced by a map.

Let us consider the problem

$$\min\{c_2(\gamma) \,:\, \gamma \in O(\mu, \nu)\}.$$

This problem has a solution $\bar{\gamma}$ since c_2 is continuous for the weak convergence and $O(\mu, \nu)$ is compact for the same convergence. We do not know a priori about the uniqueness of such a minimizer. It is interesting to notice that the solution of this problem may also be obtained as the limits of solutions γ_ε of the transport problem

$$\min\{c_1(\gamma) + \varepsilon c_2(\gamma), \colon \gamma \in \Pi(\mu, \nu)\},$$

but we will not exploit this fact here.

The goal is now to characterize this plan $\bar{\gamma}$ and prove that it is induced y a transport map.

The fact that the condition $\gamma \in O(\mu, \nu)$ may be rewritten as a condition on the support of γ is really useful since it allows to state that $\bar{\gamma}$ also solves

$$\min \int c \, d\gamma \,:\, \gamma \in \Pi(\mu, \nu), \qquad \text{where } c(x, y) = \begin{cases} |x - y|^2 & \text{if } u(x) - u(y) = |x - y| \\ +\infty & \text{otherwise.} \end{cases}$$

Actually, minimizing this new cost implies being concentrated on the set where $u(x) - u(y) = |x - y|$ (i.e. belonging to $O(\mu, \nu)$) and minimizing the quadratic cost among those plans γ concentrated on the same sets (i.e. among those $\gamma \in O(\mu, \nu)$).

Let us spend some words more in general on costs of the form

$$c(x, y) = \begin{cases} |x - y|^2 & \text{if } (x, y) \in A \\ +\infty & \text{otherwise,} \end{cases}$$

where A is a given closed subset of $\Omega \times \Omega$. First of all we notice that such a cost is l.s.c. on $\Omega \times \Omega$. To prove it, just take a sequence $(x_k, y_k) \to (x, y)$. We want to prove $\liminf_k c(x_k, y_k) \geq c(x, y)$, and there is nothing to prove if the liminf on the left is $+\infty$. We can suppose hence that the liminf is finite, which means that, up to a subsequence, $(x_{k_j}, y_{k_j}) \in A$. We can choose such a subsequence so that $\lim_j c(x_{k_j}, y_{k_j}) = \liminf_k c(x_k, y_k)$. Since A is closed we also have $(x, y) \in A$. Then, we can replace every occurrence of c with the expression it has on A (i.e. the quadratic cost, which is continuous), and since we have $|x_{k_j} - y_{k_j}|^2 \to |x - y|^2$ we have $\liminf_k c(x_k, y_k) = c(x, y)$ (the fact that c is only l.s.c. and not continuous comes from the possibility of getting $+\infty$ on the left).

Semi-continuity of the cost implies that optimal plans are concentrated on a set which is $c-CM$. What does it mean in such a case? $c-$cyclical monotonicity is a condition which imposes an inequality for every k, every σ and every family $(x_1, y_1), \dots, (x_k, y_k)$; here we only need to use the condition for $k = 2$. This means that, if $\Gamma \subset \Omega \times \Omega$ is $c - CM$ we have

$$(x_1, y_1), (x_2, y_2) \in \Gamma \Rightarrow c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1).$$

For costs c of this form, this is only useful when both (x_1, y_2) and (x_2, y_1) belong to A (otherwise we have $+\infty$ at the right had side of the inequality). If we also use the fact that

$$|x_1 - y_1|^2 + |x_2 - y_2|^2 \leq |x_1 - y_2|^2 + |x_2 - y_1|^2 \Leftrightarrow (x_1 - x_2) \cdot (y_1 - y_2) \geq 0,$$

(which is easy to get by developing the squares), this means that if Γ is $c-$CM, then

$$(x_1, y_1), (x_2, y_2) \in \Gamma, (x_1, y_2), (x_2, y_1) \in A \Rightarrow (x_1 - x_2) \cdot (y_1 - y_2) \geq 0.$$

Let us now use the fact that our optimal plan $\bar{\gamma}$ is actually concentrated on a set Γ which is $c-$CM and see how this interacts with transport rays. We can suppose $\Gamma \subset A = \{(x, y) \,:\, u(x) - u(y) = |x - y|\}$, since anyway γ must be concentrated on such a set, so as to have a finite value for $\int c\, d\gamma$. Wa want to say that γ behaves, on each transport ray, as the monotone increasing transport. More precisely, the following is true.

Lemma 0.4. *Suppose that x_1, x_2, y_1 and y_2 are all points of a transport ray $[x, y]$ with direction $e = \frac{x-y}{|x-y|}$ and that $(x_1, y_1), (x_2, y_2) \in \Gamma$. Define a order relation on such a transport ray through $x \leq x' \Leftrightarrow x \cdot e \leq x' \cdot e$. Then if $x_1 < x_2$ we also have $y_1 \leq y_2$.*

Proof. We already know that $x_1 \geq y_1$ and $x_2 \geq y_2$ (thanks to $\Gamma \subset A$). Hence, the only case to be considered is the case where we have $x_2 > x_1 \geq y_1 > y_2$. If we prove that this is not possible than we have proven the thesis. And this is not possible due to the fat that Γ is $c-$CM, since on this transport ray, due to the order relationship we have supposed and to the behavior of u on such a segment, the condition $(x_1, y_2), (x_2, y_1) \in A$ is guaranteed. This implies $(x_1 - x_2) \cdot (y_1 - y_2) \geq 0$. But this is the scalar product of two vectors parallel to e, and on the segment this simply means that y_1 and y_2 must be ordered exactly as x_1 and x_2 are. □

From what we have seen when we discussed the one-dimensional situation, we know that when s is a segment and $\Gamma \subset s \times s$ is such that $(x_1, y_1), (x_2, y_2) \in \Gamma$ and $x_1 < x_2$ imply $y_1 \leq y_2$ (for the order relation on s), then Γ is contained in the graph of a monotone increasing multivalued function, which associates to every point either a point or a segment. Yet, the interiors of these segments being disjoints, there is at most a countable number of points where the image is not a singleton. This means that Γ is contained in a graph over s, up to a countable number of points of s.

If we combine what we get on every transport ray, we have obtained the following:

Proposition 0.5. *The optimal transport plan $\bar{\gamma}$ is concentrated on a set Γ with the following properties:*

- *if $(x, y) \in \Gamma$, then either $x = y$ or $x \in T_u$ and y belongs to the same transport ray of x (which is unique up to a negligible set of points x, i.e. that where u is not differentiable)*
- *on each transport ray s, $\Gamma \cap s \times s$ is contained in the graph of a monotone increasing multivalued function*
- *on each transport ray s, the set $N_s = \{x \in s : \#(\{y : (x, y) \in \Gamma\} > 1\}$ is countable.*

It is clear that $\bar{\gamma}$ is induced by a transport map if $\mathcal{L}^d(\bigcup_s N_s) = 0$, i.e. if we can get rid of a countable number of points on every transport ray. It is also clear in this construction that all the points which belong to a degenerate transport ray, composed by such a point itself only, are not a problem: the map T will be the identity on these points.

This could also be expressed in terms of disintegration of measures (if μ is absolutely continuous, then all the measures μ_s given by the disintegrations of μ along the rays s are atomless), but we will try to avoid such an argument for the sake of simplicity. The only point that we need is the following (*Property N*, for negligibility): if $B \subset \Omega$ is such that

- B does not contain any point belonging to a degenerate transport ray,
- $B \cap [x, y]$ is at most countable for every transport ray $[x, y]$,

then $\mu(B) = 0$. Since $\mu << \mathcal{L}^d$, it is sufficient to guarantee $\mathcal{L}^d(B) = 0$.

This property is not always satisfied by any disjoint family of segments in $\mathbb{R}^d$ and there is an example (by Alberti, Kirchheim and Preiss, later improved by Ambrosio and Pratelli) where a disjoint family of segments contained in a cube is such that the collection of their middle points has positive measure. We will prove that the direction of the transport rays satisfy additional properties, which guarantee the one we need.

Just a last remark: we are ignoring here measurability issues of the transport map T that we are constructing. Actually, such map is obtained by gluing the monotone maps on every segment, but this should be done in a measurable way. It is possible to prove that this is the case, either by restricting to a $\sigma-$compact set Γ or by considering the disintegrations μ_s and ν_s and using the fact that, on each s, T is the monotone map sending μ_s onto ν_s (and hence it inherits some measurability properties of the dependence of μ_s and ν_s w.r.t. s, which are guaranteed by abstract disintegration theorems).

Directions of the transport rays

It appears that the main tool to prove Property N is the Lipschitz regularity of the directions of the transport rays (which is the same as the direction of ∇u).

Theorem 0.6. *Property N holds if ∇u is Lipschitz continuous or if there exists a countable family of sets E_h such that ∇u is Lipschitz continuous when restricted to each E_h and $\mathcal{L}^d(Tu \setminus \bigcup_h E_h) = 0$.*

Proof. First, suppose that ∇u is Lipschitz. Consider all the hyperplanes parallel to $d-1$ coordinate axes and with rational coordinates on the last coordinate. Since B is only made of points belonging to non-degenerate transport rays, every point of B belongs to a transport ray that meets at least one of these hyperplanes at exactly one point of its interior. Since these hyperplanes are a countable quantity, we can suppose that B is included in a collection S_Y of transport rays all meeting the same hyperplane Y. If we can prove that B is negligible under this assumption, then it will be negligible even if we withdraw it, by countable union. Not only, we can also suppose that B does not contain any point which is a boundary point of two different transport rays, since we already know that those points are negligible. Now, let us fix such an hyperplane Y and let us consider a map $f : Y \times \mathbb{R} \to \mathbb{R}^d$ of the following form: for $y \in Y$ and $t \in \mathbb{R}$ the point $f(y,t)$ is defined as $y + t\nabla u(y)$. This map is well-defined and injective on a set $\omega \subset Y \times \mathbb{R}$ which is the one we are interested in. This set ω is defined as those pairs (y,t) where y is in the interior of a transport ray, which gives that u is differentiable at such a point, and $y + t\nabla u(y)$ belongs to a transport ray of S_Y and it is not the boundary point of more than one transport ray. f is injective, since getting the same point as the image of (y,t) and of (y',t') would mean that two different transport rays cross at such a point. The map f is also Lipschitz continuous, as a consequence of the Lipschitz behavior of ∇u. We can hence consider $B' := f^{-1}(B)$. This set is a subset of $Y \times \mathbb{R}$ containing at most countably many points on every line $\{y\} \times \mathbb{R}$. By Fubini's theorem, this implies $\mathcal{L}^d(B') = 0$. Then we have also $\mathcal{L}^d(B) = \mathcal{L}^d(f(B')) \le \mathrm{Lip}(f)^d \mathcal{L}^d(B')$, which implies $\mathcal{L}^d(B) = 0$.

It is clear that the property is also true when ∇u is not Lipschitz but is Lipschitz continuous on each set E_h of a partition covering almost all the points of T_u, since one can apply the same kind of arguments to all the sets $B \cap E_h$ and then use countable unions. □

We now need to prove that ∇u is Lipschitz continuous, at least on a countable decomposition.

Definition 2. A function $f : \omega \to \mathbb{R}^d$ is said to be countably Lipschitz if there exist a family of sets E_h such that f is Lipschitz continuous on each E_h and $\mathcal{L}^d(\omega \setminus \bigcup_h E_h) = 0$. Notice that "being Lipschitz continuous on a set E" or "being the restriction to E of a Lipschitz continuous function defined on the whole $\mathbb{R}^d$" are actually the same property, due to Lipschitz extension theorems.

We want to prove that ∇u is countably Lipschitz. We will first prove that it coincides with some $\lambda-$convex or $\lambda-$concave functions on a sequence of sets covering almost everything. This requires a definition.

Definition 3. A function $f : \Omega \to \mathbb{R}$ is said to be $\lambda-$convex if $x \mapsto f(x) + \frac{\lambda}{2}|x|^2$ is convex, and $\lambda-$concave if $x \mapsto f(x) - \frac{\lambda}{2}|x|^2$ is concave. Notice that the number λ is not required to be positive, so that $\lambda-$convex functions for $\lambda > 0$ are strictly convex, and if $\lambda < 0$ they just have second derivatives bounded from below. Analogous considerations for $\lambda-$concave functions.

Proposition 0.7. *There exist some sets E_h^+ and E_h^- such that*

- *u coincides with a $\lambda-$concave function on each E_h^+ (for a value of λ depending on h),*
- *u coincides with a $\lambda-$convex function on each E_h^- (again, for a value of λ depending on h),*
- *$\bigcup_h E_h^+ = T_u^{(i)} \cup T_u^{(b+)}$, $\bigcup_h E_h^- = T_u^{(i)} \cup T_u^{(b-)}$,*
- *on $E_h := E_h^+ \cap E_h^-$ u is both λ convex and $-\lambda-$concave, and $\bigcup_h E_h^+ \cap E_h^- = T_u^{(i)}$.*

Proof. Let us define

$$E_h^+ = \left\{ x \in T_u \,:\, \exists z \in T_u \text{ with } |x-z| > \frac{1}{h},\, u(z) - u(x) = |x-z| \right\},$$

which is roughly speaking made of those points in the transport rays that are at least at a distance $\frac{1}{h}$ apart from the upper boundary point of the ray. Analogously, set

$$E_h^- = \left\{ x \in T_u \,:\, \exists z \in T_u \text{ with } |x-z| > \frac{1}{h},\, u(x) - u(z) = |x-z| \right\}.$$

It is clear that $E_h^+ \subset T_u \setminus T_u^{(b+)}$ and $E_h^- \subset T_u \setminus T_u^{(b-)}$. Moreover, if we set $E_h := E_h^+ \cap E_h^-$ we have $E_h \subset T_u^{(i)}$ and $\bigcup_h E_h = T_u$.

Let us fix a function $c_h : \mathbb{R}^d \to \mathbb{R}$ with the following properties: $c_h \in C^2(\mathbb{R}^d)$, $nablac_h| \leq 1$, $c_h(z) \geq |z|$ for all $z \in \mathbb{R}^n$, $c_h(z) = |z|$ for all $z \notin B(0, 1/h)$. It is easy to check that, if $x \in E_h^+$, one has

$$u(x) = \inf_{y \in \mathbb{R}^d} |x-y| + u(y) \leq \inf_{y \in \mathbb{R}^d} c_h(x-y) + u(y) \leq \inf_{y \notin B(x,1/h)} |x-y| + u(y) = u(x),$$

where the first inequality is a consequence of $|z| \leq c_h(z)$ and the second is due to the restriction to $y \notin B(x, 1/h)$. The last equality is justified by the definition of E_h^+. This implies that all the inequalities are actually equalities, and that $u(x) = u_h^+(x)$ for all $x \in E_h^+$, where

$$u_h^+(x) := \inf_{y \in \mathbb{R}^d} c_h(x-y) + u(y).$$

It is important to notice that u_h^+ is a $\lambda-$ *concave* function.

Let us justify that u_h^+ is $\lambda-$concave, for $\lambda \approx -\frac{1}{h}$. Actually, it is possible to choose c_h so that $D^2 c_h \leq -\lambda I$, for $\lambda = -\frac{2}{h}$ (and anyway the C^2 regularity is enough to bound the derivatives of c_h on bounded sets). This means that c_h is $\lambda-$concave. Consider

$$u_h^+(x) - \frac{\lambda}{2}|x|^2 = \inf_{y \in \mathbb{R}^d} c_h(x-y) - \frac{\lambda}{2}|x|^2 + u(y) = \inf_{y \in \mathbb{R}^d} c_h(x-y) - \frac{\lambda}{2}|x-y|^2 - \lambda x \cdot y + \frac{\lambda}{2}|y|^2 + u(y).$$

This last expression shows that $u_h^+(x) - \frac{\lambda}{2}|x|^2$ is concave in x, since it is expressed as the infimum of concave functions ($c_h(x-y) - \frac{\lambda}{2}|x-y|^2$ is concave and $\lambda x \cdot y$ is linear, the other terms being constant in x). Hence $u_h^+(x)$ is $\lambda-$concave.

Similarly, one can define

$$u_h^-(x) := \sup_{y\in\mathbb{R}^d} -c_h(x-y) + u(y).$$

The same argument proves that u_h^- is $\lambda-$convex. Moreover, for $x \in E_h^-$, one has

$$u(x) = \sup_{y\in\mathbb{R}^d} -|x-y| + u(y) \geq \inf_{y\in\mathbb{R}^d} -c_h(x-y) + u(y) \geq \sup_{y\notin B(x,1/h)} |x-y| + u(y) = u(x).$$

This proves that for $x \in E_h$ one has $u(x) = u_h^+(x) = u_h^-(x)$. □

The very last property stated by the previous proposition could be used together with the following theorem.

Theorem 0.8. *Take $\lambda > 0$. Then a function $f : \mathbb{R}^d \to \mathbb{R}$ is at the same time $\lambda-$convex and $-\lambda-$concave (for $\lambda < 0$), if and only if it is differentiable everywhere and $\nabla f \in \mathrm{Lip}_{|\lambda|}$.*

The main point in the proof of this theorem is the following; if f is both $\lambda-$convex and $-\lambda-$concave, then

$$\lambda|x-y|^2 \leq (x-y)\cdot(\nabla f(x) - \nabla f(y)) \leq -\lambda|x-y|^2,$$

and one would like to deduce that this implies the Lipschitz behavior of ∇f. For $f \in C^2$ this is easy, since letting $x = y + \varepsilon v$ and letting $\varepsilon \to 0$ one gets bounds on $D^2 f(y)v \cdot v$ and, $D^2 f$ being symmetric, this is enough to bound all the components of the matrix $D^2 f$, which gives Lipschitz bounds on ∇f. And, for $f \notin C^2$, a simple convolution is enough to prove the same bound. Yet, in our case there would some extra difficulties due to the fact that f satisfies this assumption on a subset $E_h \subset \mathbb{R}^d$ and not on the whole space; moreover, this result would not be enough in any case, since the union of all the sets E_h is only $T_u^{(i)}$ and not T_u. Hence, we need to evoke the following theorem.

Theorem 0.9. *If f is a convex function, then ∇f is countably Lipschitz.*

The proof of this theorem may be found in [2], in Theorem 5.34. It is true for any BV functions, and this is the framework that one finds in [2]. Here we only recall that BV functions are those whose derivatives in distributional sense are measures. This is the case for gradients of convex (or concave) functions, since the second derivatives of these functions are positive distributions, and hence measures.

As a consequence of Proposition 0.7 and Theorem 0.9 one has.

Proposition 0.10. *If u is a Kantorovitch potential, then $\nabla u : T_u \to \mathbb{R}^d$ is countably Lipschitz.*

Proof. It is clear that the countably Lipschitz regularity of Theorem 0.9 is also true for the gradients of $\lambda-$convex and $\lambda-$concave functions. This means that this is true for ∇u_h^+ and ∇u_h^- and, by countable union, for ∇u. □

Finally, we get

Theorem 0.11. *Under the usual assumption $\mu << \mathcal{L}^d$, the secondary variational problem admits a unique solution $\bar{\gamma}$, which is induced by a transport map T, monotone on very transport ray.*

Proof. Proposition 0.10 together with Theorem 0.6 guarantee that Property N holds. Hence, Proposition 0.5 may be applied to get $\bar{\gamma} = \gamma_T$. The uniqueness follows in the usual way: if two plans $\bar{\gamma}' = \gamma_{T'}$ and $\bar{\gamma}'' = \gamma_{T''}$ optimize the secondary variational problem, the same should be true for $\frac{1}{2}\gamma_{T'} + \frac{1}{2}\gamma_{T''}$. Yet, for this measure to be induced by a ma, it is necessary to have $T' = T''$ a.e. $\square$

Notice that as a byproduct of this analysis we also obtain $\mathcal{L}^d(T_u^{(b)}) = 0$, since $T_u^{(b)}$ is a set meeting every transport ray in two points, and it is hence negligible by Property N. But we could not have proven it before, so unfortunately every strategy based on a decomposition of $T_u^{(i)}$ is not complete.

The supremal case, L^∞

We consider now a different problem: instead of minimizing $c_p(\gamma) = \int |x-y|^p d\gamma$, we want to minimize the maximal displacement, i.e. its L^∞ norm.

Let us define

$$c_\infty(\gamma) := ||x-y||_{L^\infty(\gamma)} = \inf\{m \in \mathbb{R} \,:\, |x-y| \le m \text{ for } \gamma-a.e.(x,y)\} = \max\{|x-y| \,:\, (x,y) \in \mathrm{spt}(\gamma)\}$$

(where the last equality, between an L^∞ norm and a maximum on the support, is justified by the continuity of the function $|x-y|$).

Lemma 0.12. *For every $\gamma \in \mathcal{P}(\Omega \times \Omega)$ the quantities $c_p(\gamma)^{1/p}$ increasingly converge to $c_\infty(\gamma)$ as $p \to +\infty$. In particular $c_\infty(\gamma) = \sup_{p\ge 1} c_p(\gamma)^{1/p}$ and c_∞ is l.s.c. for the weak convergence in $\Pi(\mu,\nu)$. Thus, it admits a minimizer over $\Pi(\mu,\nu)$, which is compact.*

Proof. It is well known that, on any finite measure space, L^p norms converge to the L^∞ norm, and we will not reprove it here. This may be applied to the function $(x,y) \mapsto |x-y|$ on $\Omega \times \Omega$, endowed with the measure γ, thus getting $c_p(\gamma)^{1/p} \to c_\infty(\gamma)$. Yet, it is important that this convergence is monotone here, and this is true when the measure is a probability. In such a case, we have for $p < q$, using Hölder (or Jensen) inequality

$$\int |f|^p d\gamma \le \left(\int |f|^q d\gamma\right)^{p/q} \left(\int 1\, d\gamma\right)^{1-p/q} = \left(\int |f|^q d\gamma\right)^{p/q},$$

for every $f \in L^q(\gamma)$. This implies, by taking the $p-$th root, $||f||_{L^p} \le ||f||_{L^q}$. Applied to $f(x,y) = |x-y|$ this gives the desired monotonicity.

From that we infer that c_∞ is the supremum of a family of functional which are continuous for the weak convergence (since c_p is the integral of a bounded continuous function, Ω being compact, and taking the $p-$th root does not break continuity). As a supremum of continuous functionals, it is l.s.c. and the conclusion follows. $\square$

The goal now is to analyze the solution of

$$\min c_\infty(\gamma) \ : \ \gamma \in \Pi(\mu,\nu)$$

ant to prove that there is at least a minimizer γ induced by a transport map. This map would minimize

$$\min ||T(x) - x||_{L^\infty(\mu)} \ , \ T_\# \mu = \nu.$$

Here as well there will be no uniqueness (it is almost always the case when we minimize an L^∞ criterion), hence we define $O_\infty(\mu,\nu) = \mathrm{argmin}_{\gamma\in\Pi(\mu,\nu)} c_\infty(\gamma)$, the set of optimal transport plans for this L^∞ cost. Notice that $O_\infty(\mu,\nu)$, since c_∞ is l.s.c. (as for $O(\mu,\nu)$). Suppose now that $\min\{c_\infty(\gamma) \ : \ \gamma \in \Pi(\mu,\nu)\} = L$: notice also that we have

$$\gamma \in O_\infty(\mu,\nu) \Leftrightarrow \mathrm{spt}(\gamma) \subset \{(x,y) : \, |x-y| \le L\}$$

(since any transport plan γ concentrated on the pairs where $|x-y| \le L$ satisfies $c_\infty(\gamma) \le L$ and is hence optimal). We will suppose $L > 0$ otherwise this means that it is possible to obtain ν from μ with no displacement, i.e. $\mu = \nu$ and the optimal displacement is the identity.

Consequently, exactly as for the L^1 case, we can define a secondary variational problem:

$$\min\{c_2(\gamma) \, : \, \gamma \in O_\infty(\mu,\nu)\}.$$

This problem has a solution $\bar\gamma$ since c_2 is continuous for the weak convergence and $O_\infty(\mu,\nu)$ is compact. Also for this minimizer, we do not know a priori any uniqueness. Again, it is possible to say that $\bar\gamma$ also solves

$$\min \int c \, d\gamma \ : \ \gamma \in \Pi(\mu,\nu), \qquad \text{where } c(x,y) = \begin{cases} |x-y|^2 & \text{if } |x-y| \le L \\ +\infty & \text{otherwise.} \end{cases}$$

The arguments are the same as in the L^1 case. Moreover, also the form of the cost c is similar, and this cost is l.s.c. as well. Hence, $\bar\gamma$ is concentrated on a set $\Gamma \subset \Omega \times \Omega$ which is $c-$CM. This means

$$(x_1,y_1),(x_2,y_2) \in \Gamma, |x_1 - y_2|, |x_2 - y_1| \le L \Rightarrow (x_1 - x_2)\cdot(y_1 - y_2) \ge 0. \tag{0.2}$$

We can also suppose $\Gamma \subset \{(x,y) \ : \ |x-y| \le L\}$. We will try to improve a little bit the set Γ, by removing negligible sets and getting better properties. Then, we will show that the remaining set $\tilde\Gamma$ will be contained in the graph of a map T, thus obtaining the result.

First, we need to introduce the concept of Lebesgue points:

Definition 4. For a measurable set $E \subset \mathbb{R}^d$ we call Lebesgue point of E a point $x \in E$ such that

$$\lim_{r\to 0} \frac{\mathcal{L}^d(E \cap B(x,r))}{\mathcal{L}^d(B(x,r))} = 1.$$

The set of Lebesgue points of E is denoted by $\mathrm{Leb}(E)$ and it is well-known that $\mathcal{L}^d(E \setminus \mathrm{Leb}(E)) = 0$.

Improvement of Γ

Let B_i be a countable family of balls in $\mathbb{R}^d$, generating the topology of $\mathbb{R}^d$ (for instance, all the balls with rational radius and rational center). Consider now

$$A_i := (\pi_x)(\Gamma \cap (\Omega \times B_i)),$$

i.e. the set of all points x such that at least a point y with $(x, y) \in \Gamma$ belongs to B_i (the points that have at least one "image" in B_i). Then set $N_i := A_i \setminus \mathrm{Leb}(A_i)$. This set has zero Lebesgue measure, and hence it is $\mu-$negligible. Also $\mu(\bigcup_i N_i) = 0$. As a consequence, one can define $\tilde{\Gamma} := \Gamma \setminus ((bigcup_i N_i) \times \Omega)$. The plan $\bar{\gamma}$ is still concentrated on $\tilde{\Gamma}$, since we only removed $\mu-$negligible points. Moreover, $\tilde{\Gamma}$ has the following property: if $(x_0, y_0) \in \tilde{\Gamma}$, then, for every $\varepsilon, \delta > 0$, every unit vector ξ and every sufficiently small $r > 0$, there are a point $x \in \big(B(x_0, r) \setminus B(x_0, \frac{r}{2})\big) \cap C(x_0, \xi, \delta)$ and a point $y \in B(y_0, \varepsilon)$ such that $(x, y) \in \tilde{\Gamma}$, where $C(x_0, \xi, \delta)$ is the following convex cone:

$$\{x : (x - x_0) \cdot \xi > (1 - \delta)|x - x_0|\}.$$

This means that one can choose a point x close to x_0, which is sent to at least a point y close to y_0, also imposing the direction $x - x_0$ up to an error δ as small as we want. This is true since there is at least one of the ball B_i containing y_0 and contained in $B(y_0, \varepsilon)$. Since $x_0 \in A_i$ and we have removed N_i, this means that x_0 is a Lebesgue point for A_i. Since the region $\big(B(x_0, r) \setminus B(x_0, \frac{r}{2})\big) \cap C(x_0, \xi, \delta)$ is a portion of the ball $B(x_0, r)$ which takes a fixed proportion (depending on δ) of volume of the whole ball, for $r \to 0$ it is clear that A_i (and also $\mathrm{Leb}(A_i)$) must meet it (otherwise x_0 would not be a Lebesgue point). It is hen sufficient to pick a point in $\mathrm{Leb}(A_i) \cap \big(B(x_0, r) \setminus B(x_0, \frac{r}{2})\big) \cap C(x_0, \xi, \delta)$ and we are done.

Moreover, $\tilde{\Gamma}$ stays $c-$CM and enjoys property (0.2).

$\tilde{\Gamma}$ is a graph

Lemma 0.13. *If (x_0, y_0) and (x_0, z_0) belong to $\tilde{\Gamma}$, then $y_0 = z_0$.*

Proof. Suppose by contradiction $y_0 \neq z_0$. In order to fix the ideas, let us suppose $|x_0 - z_0| \leq |x_0 - y_0|$ (and in particular $y_0 \neq x_0$, since otherwise $|x_0 - z_0| = |x_0 - y_0| = 0$ and $z_0 = y_0$).

Set $\eta := \frac{z_0 - y_0}{|z_0 - y_0|}$ (which is possible to do since $z_0 \neq y_0$).

At least one of the angles $x_0 y_0 z_0$ or $x_0 z_0 y_0$ is acute, i.e. either $(x_0 - y_0) \cdot (z_0 - y_0) > 0$ or $(x_0 - z_0) \cdot (y_0 - z_0) > 0$. Let us suppose that $(x_0 - z_0) \cdot (z_0 - y_0) \geq 0$, for simplicity. The other case is completely symmetrical.

Now, use the property of $\tilde{\Gamma}$ and find $(x, y) \in \tilde{\Gamma}$ with $y \in B(y_0, \varepsilon)$ and $x \in \big(B(x_0, r) \setminus B(x_0, \frac{r}{2})\big) \cap C(x_0, \xi, \delta)$, for a vector ξ to be determined later. Use now the fact that $\tilde{\Gamma}$ is $c-$CM applied to (x_0, z_0) and (x, y).

If we can prove that $|x - z_0|, |x_0 - y| \leq L$, then we should have

$$(x - x_0) \cdot (y - z_0) \geq 0.$$

Yet, the direction of $x-x_0$ is almost that of ξ (up to an error of δ) and that of $y-z_0$ is almost that of y_0-z_0 (up to an error of the order of $\varepsilon/|z_0-y_0|$). If we choose ξ such that $\xi\cdot(y_0-z_0)<0$, this means that for small δ and ε we would get a contradiction.

Moreover, we need to guarantee $|x-z_0|, |x_0-y|\leq L$, in order to prove the thesis. Let us compute $|x-z_0|^2$: we have

$$|x-z_0|^2=|x_0-z_0|^2+|x-x_0|^2+2(x-x_0)\cdot(x_0-z_0).$$

In this sum we have

$$|x_0-z_0|^2\leq L^2;\quad |x-x_0|^2\leq r^2;\quad 2(x-x_0)\cdot(x_0-z_0)=|x-x_0|(\xi\cdot(x_0-z_0)+O(\delta)).$$

Suppose now that we are in one of the following situations: either choose ξ such that $\xi\cdot(x_0-z_0)<0$ or $|x_0-z_0|^2<L^2$. In both cases we get $|x-z_0|\leq L$ for r and δ small enough. In the fist case, since $|x-x_0|\geq \frac{r}{2}$, we have a negative term of the order of r and a positive one of the order of r^2; in the second we add to $|x_0-z_0|^2<L^2$ some terms of the order of r or r^2.

Analogously, for $|x_0-y|$ we have

$$|x_0-y|^2=|x_0-x|^2+|x-y|^2+2(x_0-x)\cdot(x-y).$$

The three terms satisfy

$$|x_0-x|^2\leq r^2;\quad |x-y|^2\leq L^2;\quad 2(x_0-x)\cdot(x-y)=|x-x_0|(-\xi\cdot(x_0-y_0)+O(\delta+\varepsilon+r)).$$

In this case, we need to impose $\xi\cdot(x_0-y_0)>0$ so as to guarantee $|x_0-y|\leq L$ for r,ε and δ small enough.

All in all, we are done if we can select a vector ξ such that

$$\xi\cdot(y_0-z_0)<0;\quad \xi\cdot(x_0-z_0)<0;\quad \xi\cdot(x_0-y_0)>0,$$

i.e. $\xi\cdot z_0>\xi\cdot x_0>\xi\cdot y_0$. Thanks to the assumption $y_0\neq z_0$ the only case when this is not possible is when $z_0=x_0$ and we have three different points (we already know that $y_0\neq x_0$). Actually, to visualize an easier situation, suppose $x_0=0$ - that we could obtain by translation - and look for a vector ξ such that $\xi\cdot z_0>0>\xi\cdot y_0$.

The only case which is left is when $z_0=x_0$ but in such a case one has $|x_0-z_0|^2<L^2$. In such a case the condition $\xi\cdot(x_0-z_0)<0$ is not necessary, and one only needs to guarantee $\xi\cdot z_0=\xi\cdot x_0>\xi\cdot y_0$, which is obviously possible (for instance for $\xi=(z_0-y_0)/|z_0-y_0|$). □

Theorem 0.14. *The secondary variational problem* $\min c_2(\gamma)\ :\ \gamma\in O_\infty(\mu,\nu)$ *admits a unique solution* $\bar\gamma$*, it is induced by a transport map* T*, and such a map is an optimal transport for the problem*

$$\min ||T(x)-x||_{L^\infty(\mu)}\ ,\ T_\#\mu=\nu.$$

Proof. We have already seen that $\bar{\gamma}$ is concentrated on a set $\tilde{\Gamma}$ satisfying some useful properties. Lemma 0.13 shows that $\tilde{\gamma}$ is contained in a graph, since for any x_0 there is no more than one possible point y_0 such that $(x_0; y_0) \in \tilde{\Gamma}$. Let us consider such a point y_0 as the image of x_0 and call it $T(x_0)$. Then $\bar{\gamma} = \gamma_T$. The optimality of T in the "Monge" version of this L^∞ problem comes from the usual comparison with the Kantorovitch version on plans γ. The uniqueness comes from the standard arguments (since it is true as well that convex combinations of minimizers should be minimizers, and this allows to perform the usual proof). $\square$

We conclude this document by stressing that many recent researches have tried to extend similar results to the case of alternative norms, different from the euclidean one (in more than one point we used here, both for the L^1 and for the L^∞ case, that the norm was the euclidean one, coming from a scalar product). In this framework, this density-based proof of the L^∞ case has also been used to tackle the L^1 problem. It gives a proof that does not use decomposition into rays, but some properties on transport rays like the one that we proved here are to be studied anyway.

References

[1] L. Ambrosio, *Lecture Notes on Optimal Transport Problems, Mathematical aspects of evolving interfaces,* CIME Summer School in Madeira, vol. 1812, Springer, 2003.

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.* Oxford mathematical monographs. Oxford University Press, Oxford, 2000.