Communications in Applied Analysis **15** (2011), no. 2, 3 and 4, 325–340

THE MONGE PROBLEM IN $\mathbb{R}^d$: VARIATIONS ON A THEME II

LUIGI DE PASCALE[1]

[1]Dipartimento di Matematica Applicata "U. Dini"
Università di Pisa
Via F.Buonarroti 1/C
I-56127 Pisa, ITALY
E-mail: depascal@dm.unipi.it

Dedicated to J.R.L. Webb
in occasion of his retirement.

ABSTRACT. In some recent papers [11, 12] it is proved that, under natural assumptions on the first marginal, the Monge problem in the metric space $\mathbb{R}^d$ equipped with a general norm admits a solution. Although the basic idea of the solution is simple the proof involves some very complex technical results. Here we will report a proof of the result in the simpler case of uniformly convex norms. Uniform convexity allow us to reduce the technical burdens while still giving the main ideas of the general proof. The proof of the density of the transport set given in this paper is original.

AMS (MOS) Subject Classification. 49Q20, 49K30, 49J4

1. INTRODUCTION

The Monge problem has origin in the *Mémoire sur la théorie des déblais et remblais* written by G. Monge [18]. The problem was stated, more or less, as follows: given a sand pile and an embankment with the same volume as the sand pile is there a way to transport the sand in the embankment which minimizes the work? We consider the closure Ω of an open, bounded and convex subset of $\mathbb{R}^d$ as ambient space for the model. Then, if we use a probability measure μ to model the sand pile and a probability measure ν to model the embankment, a transport map T from μ to ν will be a Borel map such that $T_\sharp\mu = \nu$ (i.e. $\nu(B) = \mu(T^{-1}(B))$ for all Borel sets $B \subset \Omega$). If we denote by $\mathcal{T}(\mu,\nu)$ the set of transport maps of μ to ν then the problem will take the form

$$\inf\left\{\int_\Omega |x - T(x)|d\mu(x) \,:\, T \in \mathcal{T}(\mu,\nu)\right\}. \tag{1.1}$$

Received November 10, 2010

The natural appeal of the problem and the many applications attracted the interest to a generalization in which one consider a general norm

$$\inf\left\{\int_\Omega ||x-T(x)||d\mu(x)\,:\,T\in\mathcal{T}(\mu,\nu)\right\},\tag{1.2}$$

A first strategy to solve the problem was devised by Sudakov in [23]. The basic idea in that paper was to reduce the problem to lower dimensional affine regions. This is quite natural as we will explain in Section 2. Reducing the problem to lower dimensional affine spaces requires to consider the restrictions (or conditional probability) of μ and ν to the regions of interest. However this method involved a crucial step on the disintegration of a measure which was not completed correctly at that time, and has recently been justified in the case of a strictly convex norm by Caravenna [10]. Meanwhile, the problem (1.1) has been solved by Evans *et al.* [14] with the additional regularity assumption that μ and ν have Lipschitz-continuous densities with respect to $\mathcal{L}^d$, and then by Ambrosio [2] and Trudinger *et al.* [24] for μ and ν with integrable density. The more general problem (1.2) for C^2 uniformly convex norms has been solved by Caffarelli *et al.* [9] and Ambrosio *et al.* [5], and for crystalline norms in $\mathbb{R}^d$ and general norms in $\mathbb{R}^2$ by Ambrosio *et al.* [4]. The original proof of Sudakov was based on the reduction of the transport problems to affine regions of smaller dimension, and all the proofs we listed above are based on the reduction of the problem to a 1-dimensional problem via a change of variable or area-formula. In [11, 12], we introduced a different method which does not require the reduction to 1-dimensional settings.

This paper. The aim of this paper is mainly expository. We will focus on the particular case of uniformly convex norms to illustrate the strategy underlying the proofs of [11, 12]. The uniform convexity assumption

$$c\le\frac{\partial^2}{\partial\xi\partial\xi}\|\cdot\|^2\le C\quad\text{for all }\xi\in\mathcal{S}^1\text{ and for some }\;0<c\le C.\tag{1.3}$$

will considerably reduce the technical burdens of the proof while leaving intact the main ideas.

One of the main steps in [11, 12] is the proof of the density of the transport set (see definition below) which involves some constants depending on the dimension of the ambient space. In particular the constants vanish when the dimension $d\to\infty$ thus preventing finite dimensional approximation in the spirit of [15]. As a minor original contribution we will revisit the proof of the density of the transport set first obtaining an estimate of the density instead of the lower density and then obtaining estimates which do not depends on the dimension. We hope that this paper will make the problem accessible to a non-specialist audience.

2. THE MAIN PLAYERS AND THEIR BASIC PROPERTIES

The first step consists in suitably relaxing the problem. This was done by Kantorovich [16, 17] who introduced the set

$$\Pi(\mu,\nu) = \{\gamma \in \mathcal{P}(\Omega \times \Omega) \ : \ \pi^1_\sharp\gamma = \mu, \ \pi^2_\sharp\gamma = \nu\},$$

and the cost

$$\int_{\Omega\times\Omega} \|x-y\| d\gamma.$$

The elements of $\Pi(\mu,\nu)$ are called transport plans and as tools to transport μ to ν they allow the mass sitting at a point x to be split among many point y while a transport map T move all of the mass sitting at x to $T(x)$. There is a natural embedding of $T(\mu,\nu)$ in $\Pi(\mu,\nu)$ which associates to a transport map T the transport plan $\gamma_T = (id \times T)_\sharp\mu$, which has the same cost

$$\int_{\Omega\times\Omega} \|x-y\| d\gamma_T = \int_{\Omega} \|x-T(x)\| d\mu.$$

Then the new problem is

$$\min_{\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \|x-y\| d\gamma. \tag{2.1}$$

The inescapable question is whether

$$\inf_{\mathcal{T}(\mu,\nu)} \int_{\Omega} ||x-T(x)|| d\mu(x) = \min_{\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \|x-y\| d\gamma.$$

Since we will prove that if μ is absolutely continuous with respect to $\mathcal{L}^d$ then some of the optimal transport plans are induced by optimal transport maps, the equality will follow. However it can be proved the more general result

Theorem 2.1. *If μ has no atoms then*

$$\inf_{\mathcal{T}(\mu,\nu)} \int_{\Omega} ||x-T(x)|| d\mu(x) = \min_{\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \|x-y\| d\gamma.$$

For a proof of Theorem 2.1 in wide generality we refer to [19] and reference therein.

The assumption that μ is non atomic cannot be removed since in that case the set $\mathcal{T}(\mu,\nu)$ may be empty.

Example 2.2. In $\mathbb{R}$ consider $\mu := \delta_0$ and $\nu := \frac{1}{2}(\delta_1+\delta_{-1})$. In this case the set $\mathcal{T}(\mu,\nu)$ is easily seen to be empty. In general it may happens that $\mathcal{T}(\mu,\nu)$ is non empty but the left-hand-side is an inf while the right-hand-side (under the current assumptions) is always a minimum.

Example 2.3. In $\mathbb{R}^2$ let $S_0 = \{(0,t) \ : \ t \in [0,1]\}$, $S_1 = \{(1,t) \ : \ t \in [0,1]\}$ and $S_{-1} = \{(-1,t) \ : \ t \in [0,1]\}$. Let $\mu := \mathcal{H}^1\lfloor S_0$ and $\nu := \frac{1}{2}(\mathcal{H}^1\lfloor S_1 + \mathcal{H}^1\lfloor S_{-1})$ where by $\mathcal{H}^1$ we denote the one-dimensional Hausdorff measure. In this case the optimal transport plan will move half of the mass horizontally to the right and the other half horizontally to the left. This cannot be achieved by any transport map.

A natural (although technical) question is whether a transport plan supported on a graph is induced by a transport map or not. The answer is relevant for this paper and is the topic of Lemma 3.1 in [1] . Since the aim of this paper is partly expository we report the proof below.

Lemma 2.4 ([1])**.** *Let X and Y be subsets of complete separable metric spaces, and $\gamma \geq 0$ a σ-finite Borel measure on the product space $X \times Y$. Denote the X-marginal of γ by μ. If γ vanishes outside the graph of $T : X \to Y$ (in the sense that the outer measure of $(X \times Y) \setminus Graph(T) = 0$) , then T is μ measurable and $\gamma = (id \times T)_\sharp \mu$.*

Proof. To start let us assume that X and Y are closed (and then complete and separable). In this case γ is a regular measure since it is σ-finite and Borel on a complete and separable metric space. Then, since $\gamma(X \times T \setminus (Graph(T))) = 0$, there exists an increasing sequence of compact sets $K_i \subset K_{i+1} \subset \cdots \subset Graph(T)$ such that $K_\infty \subset Graph(T)$ has full measure or equivalently $\gamma(X \times Y \setminus K_\infty) = 0$. Since K_i is compact the restriction of T to the compact set $\pi^X(K_i)$ is continuous and then the restriction T_∞ of T to $\pi^X(K_\infty)$ is a Borel map. We will now show that $\gamma = (id \times T_\infty)_\sharp \mu$. Let $U \times V$ be any Borel *"rectangle"* then

$$
\begin{aligned}
\gamma(U \times V) &= \gamma(U \cap T_\infty^{-1}(V) \times Y) \\
&= \mu(U \cap T_\infty^{-1}(V)) \\
&= (id \times T_\infty)_\sharp \mu(U \times V).
\end{aligned}
$$

And this implies the thesis. To conclude we show that the closure assumption on X and Y do not hurt the generality. Indeed we may consider $\overline{X}$ and $\overline{Y}$ and extend γ to $\overline{\gamma}$ by setting $\overline{\gamma}(B) = \gamma(B \cap (X \times Y))$. Clearly $\overline{\gamma}$ vanishes outside the graph of T and if the statement holds for $\overline{\gamma}$ then necessarily holds for γ too. □

Remark 2.5. In the original papers [11, 12] we used Proposition 2.1 of [2] instead of Lemma 2.4 above. Although the two are equivalent the formulation of Lemma 2.4 is more convenient. In fact Proposition 2.1 of [2] requires that the graph of T is γ-measurable. Checking this measurability may be technically non trivial.

Problem (2.1) is a linear minimization problem with convex constraints and then it allows the use of duality theory. The dual problem takes the following form

$$
\max_{u \in Lip_1(\Omega)} \int_\Omega u(x) d\mu - \int_\Omega u(y) d\nu. \tag{2.2}
$$

Theorem 2.6. *Under the current assumptions*

$$
\min_{\Pi(\mu,\nu)} \int_{\Omega \times \Omega} \|x - y\| d\gamma = \max\{\int_\Omega u(x) d\mu - \int_\Omega u(y) d\nu \mid u(x) - u(y) \leq \|x - y\|\}.
$$

Moreover if u is a maximizer for the right-hand-side then γ is optimal for the left-hand-side if and only if $u(x) - u(y) = \|x - y\|$ γ-a.e.

Proof. We first notice that the existence of a maximizer for the right hand side is easily obtained by the direct method of the Calculus of Variations. In fact, after observing that adding a constant to an admissible u does not change the value of the functional, one can apply the Ascoli-Arzelà theorem to a bounded, maximizing sequence.

To prove the equality between the extremal values we use the convex duality theory. We first rewrite the right-hand-side as follows

$$\sup\left\{\int_\Omega u\, d\mu + \int_\Omega v\, d\nu \ : \ \forall x, y, \quad u(x) + v(y) \leq \|x - y\|\right\}. \tag{2.3}$$

It is indeed clear that the right-hand-side is lower than the sup in (2.3). The proof of the reverse inequality is as follows: if one associates to a function u the function $\tilde{u} : y \mapsto \inf_x\{\|x-y\| - u(x)\}$, then the sup of the right hand side of (2.3) is also realized with couples of functions of the form $(u, \tilde{u})$, and then of the form $(\tilde{\tilde{u}}, \tilde{u}) = (-\tilde{u}, \tilde{u})$, from which the reverse inequality follows.

Then we consider $p \in \mathcal{C}(\Omega \times \Omega)$ and we perturb the problem (2.3) as follows:

$$h(p) = \inf\{-\int_\Omega u\, d\mu - \int_\Omega v\, d\nu \ : \ u(x) + v(y) + p(x, y) \leq \|x - y\|\}.$$

Notice in particular that $h(0) = -(2.3)$. Moreover the function h is convex. Let us compute the Moreau-Fenchel conjugate $h^*(\gamma)$ for $\gamma \in \mathcal{M}_+(\Omega \times \Omega)$ (we notice that $h^*(\gamma) = +\infty$ if the negative part γ^- of the measure γ is not 0):

$$\begin{aligned}
h^*(\gamma) &= \sup_p\{\langle\gamma, p\rangle - h(p)\} \\
&= \sup_{u,v,p}\left\{\langle\gamma, p\rangle + \int_\Omega u\, d\mu + \int_\Omega v d\nu \ : \ u(x) + v(y) + p(x, y) \leq \|x - y\|\right\} \\
&= \sup_{u,v}\left\{\int_{\Omega\times\Omega} \|x - y\| d\gamma + \langle\gamma, -u - v\rangle + \int_\Omega u\, d\mu + \int_\Omega v d\nu\right\} \\
&= \begin{cases} \displaystyle\int_{\Omega\times\Omega} \|x - y\| d\gamma & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Then

$$h^{**}(0) \ = \ -\min_{\gamma\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \|x - y\| d\gamma \ = \ -\min(2.1).$$

We just need to prove that h is lower semicontinuous at 0 and it will follow that $h^{**}(0) = h(0)$. Since Ω is compact, then h is bounded in a neighborhood of 0 for the uniform convergence: since h is also convex, it follows that it is Lipschitz continuous in a neighborhood of 0. The equality between the extremal values of (2.1) and (2.2) follows.

To prove the last part of the statement, take a maximizer u for (2.2) and $\gamma \in \Pi(\mu, \nu)$ optimal for (2.1). Since $u(x) - u(y) \leq \|x - y\|$ for all x and y then, by definition of marginal measures, the equality

$$\int_\Omega u(x)d\mu - \int_\Omega u(y)d\nu = \int_{\Omega\times\Omega} \|x - y\| d\gamma$$

holds if and only if $u(x) - u(y) = \|x - y\|$ for γ-a.e (x, y). A final remark is that by continuity of u this last equality is also satisfied on the support of γ. □

Remark 2.7. The optimality for the Kantorovich problem has a remarkable consequence on the structure of the support of an optimal measure γ which we may call 2-points cyclical monotonicity i.e. for any couple of points (x_1, y_1), $(x_2, y_2) \in spt(\gamma)$ then

$$\|x_1 - y_1\| + \|x_2 - y_2\| \leq \|x_1 - y_2\| + \|x_2 - y_1\|.$$

Infact, by the previous theorem,

$$\|x_1 - y_1\| + \|x_2 - y_2\| = u(x_1) - u(y_1) + u(x_2) - u(y_2) \leq \|x_1 - y_2\| + \|x_2 - y_1\|.$$

The monotonicity property illustrated in the previous remark is a particular case of the so called cyclical monotonicity. Here we will not discuss cyclical monotonicity in its full generality. It is worth to note that in a quite general setting cyclical monotonicity characterizes the optimality of γ (see [8, 20, 21]).

The duality theorem above brings us to introduce a relevant set which is called transport set. The transport set is the set in which the transport actually happens. We will now introduce two notions of transport set and transport rays one associated to a transport potential and the other to a transport plan. We will then make the natural comparison between the two.

Definition 2.8. Let $u \in Lip_1(\Omega)$, an open segment $]x, y[$ is called transport ray if it is a maximal, open, oriented segment whose end points satisfy the condition

$$u(x) - u(y) = \|x - y\|. \tag{2.4}$$

The transport set $\mathcal{T}_u$ is the union of all transport rays. The union $\mathcal{T}_u^e$ (e stands for end-points) of all the closed transport rays will also play a role.

Definition 2.9. The transport set associated to a set $\Gamma \in \Omega \times \Omega$ is defined as the set

$$T(\Gamma) = \{(1 - t)x + ty \mid (x, y) \in \Gamma \text{ and } t \in (0, 1)\}.$$

Remark 2.10. Let γ be an optimal plan of transport, by theorem 2.6 for any subset $R \subset spt(\gamma)$ and any Kantorovich potential u

$$T(R) \setminus \Delta \subset \mathcal{T}_u.$$

In the equation above Δ denotes the diagonal of the product space. This set has to be considered separatelly since we defined transport rays as open segments which requires distinct end-points.

In everything we said up to this point the properties of the norm did not play role. In what follows we will need some definition and elementary properties. Let $\|\cdot\|$ be a norm in $\mathbb{R}^d$, the dual norm $\|\cdot\|_*$ is defined as follows

$$\|p\|_* := \max\{\langle p, x\rangle \ : \ \|x\| = 1\}. \tag{2.5}$$

Since the function maximized on the right-hand-side of (2.5) is linear, whenever the norm $\|\cdot\|$ is strictly convex, for $p \neq 0$ the maximum is achieved at a unique point which we denote by p^*. The map which associate

$$p \mapsto p^*, \qquad (\text{for } p \neq 0)$$

is called duality map. If the norm $\|\cdot\|$ satisfies the estimate below in 1.3 then the duality map is Lipschitz on the set $\{p \ : \ \|p\|_* = 1\}$.

The transport potential u is affine on the transport rays and the following lemma holds.

Lemma 2.11. *Let $u \in Lip_1(\Omega)$, if u is differentiable at a point $z \in \mathcal{T}_u$ then $\|\nabla u\|_* = 1$ and the vector $-\nabla u(z)^*$ is parallel to the transport ray containing z.*

Proof. First we remark that since u is $1-$Lipschitz we have $\|\nabla u\|_* \leq 1$. Since $z \in \mathcal{T}_u$ let us denote by (x, y) the extremes of the transport ray to which z belongs. For every $s \in (-\|x-z\|, \|z-y\|)$ we have that $u(z) = u(z + s\frac{y-x}{\|y-x\|}) + s$ then

$$\lim_{s\to 0} \frac{u(z + s\frac{y-x}{\|y-x\|}) - u(z)}{s} = \nabla u(z) \cdot \frac{y-x}{\|y-x\|} = -1$$

which is, by definition, equivalent to the thesis. $\square$

Another property which follows from the strict convexity of the norm is the so called "no-crossing" property, i.e. different transport rays do not cross. We discuss this property and a consequence in the following remark.

Remark 2.12. Let $]x_1, y_1[$ and $]x_2, y_2[$ be two transport rays. Then either

$$]x_1, y_1[\,\cap\,]x_2, y_2[= \emptyset$$

or

$$]x_1, y_1[\,=\,]x_2, y_2[.$$

In fact if $z \in]x_1, y_1[\cap]x_2, y_2[$ then $u(x_1) - u(z) = \|x_1 - z\|$ and $u(z) - u(y_2) = \|z - y_2\|$ and this, together with the Lipschitz property of u implies

$$\|x_1 - z\| + \|z - y_2\| = u(x_1) - u(y_2) \leq \|x_1 - y_2\|.$$

By the strict convexity of the norm x_1, z and y_2 are on the same line and then by the maximality of the transport rays $]x_1, y_1[=]x_2, y_2[$.

A consequence of this property and of lemma 2.11 is the following. Let z_1 and z_2 be two different elements of the transport set $\mathcal{T}_u$, assume that u is differentiable at both points and let $t \in \mathbb{R}$ be sufficiently small, so that for $i = 1, 2$ $z_i - t\nabla u(z_i)^*$ belongs to the same transport ray as z_i. Then

$$z_1 - t\nabla u(z_1)^* \neq z_2 - t\nabla u(z_2)^*. \tag{2.6}$$

In fact if the two transport rays are different they cannot intersect. If both z_1 and z_2 belongs to the transport ray $]x, y[$ then $\nabla u(z_1)^* = \nabla u(z_2)^*$ and since $z_1 \neq z_2$ (2.6) holds

Remark 2.13. Theorem 2.6 and Lemma 2.11 indicate that in the case of a uniformly convex norm the transport happens along lines of maximal slope for a transport potential u. This is at the root of the $1-$dimensional decomposition strategies followed by other authors and cited in the introduction. When the norm is not strictly convex one need to consider the regions on which the transport potential u is affine and these regions may be higher dimensional affine submanifolds of $\mathbb{R}^d$.

Since one of the main ideas here is to select a *"better"* transport plan let us introduce the tools to achieve this aim.

We denote by $\mathcal{O}_1(\mu, \nu)$ the set of optimal transport plans for (2.1), and consider the auxiliary problem:

$$\min \left\{ \int_{\Omega \times \Omega} |y - x|^2 d\gamma(x, y) \ : \ \gamma \in \mathcal{O}_1(\mu, \nu) \right\}, \tag{2.7}$$

where we remark the fact that the cost considered in (2.7) involves the Euclidean norm $|\cdot|$ of $\mathbb{R}^d$. This procedure of choosing particular minimizers is the root of the idea of asymptotic development by Γ-convergence (see [6] and [7]) .

Finally, since an element of $\mathcal{O}_2(\mu, \nu)$ is a solution of (2.7), it enjoys a cyclical monotonicity property inherited from the cost $(x, y) \mapsto |y - x|^2$ (see remark 2.15 below), stated in the following proposition, whose proof may be derived from that of Lemma 4.1 in [4] and is given in [11] (see Proposition 3.2 therein).

Proposition 2.14. *Let γ be a solution of (2.7), then γ is concentrated on a σ-compact set Γ with the following property:*

$$\forall (x, y), (x', y') \in \Gamma, \qquad x \in [x', y'] \ \Rightarrow \ (x - x') \cdot (y - y') \geq 0, \tag{2.8}$$

where $\cdot$ denotes the usual scalar product on $\mathbb{R}^d$.

Remark 2.15. To explain condition (2.8) above we recall that if γ is a minimizer for

$$\min \left\{ \int_{\Omega \times \Omega} |y - x|^2 d\gamma(x, y) \ : \ \gamma \in \Pi(\mu, \nu) \right\} \tag{2.9}$$

then $spt(\gamma)$ satisfies a 2-points cyclical monotonicity condition related to the cost $|x-y|^2$ which says that for any couple of points $(x_1, y_1),\ (x_2, y_2) \in spt(\gamma)$ then

$$|x_1 - y_1|^2 + |x_2 - y_2|^2 \leq |x_1 - y_2|^2 + |x_2 - y_1|^2$$

which is equivalent to

$$(x_2 - x_1) \cdot (y_2 - y_1) \geq 0.$$

The measure γ involved in Proposition 2.14 is a minimizer for a the constrained version (2.7) of (2.9) and infact the additional request $x \in [x', y']$ in (2.8) is a consequence of the constraint $\gamma \in \mathcal{O}_1(\mu, \nu)$.

3. FINER PROPERTIES AND PROOF OF THE MAIN THEOREM

Beside the "*functional analytic*" properties studied in the previous section, optimal transport plans and Kantorovich potentials enjoy some finer properties which belong to the realm of Geometric Measure Theory. The properties of the transport plan we introduce below were first applied in [13] to deal with some optimal transport problem with cost in non integral form. When considered as multivalued maps, transport plans (not necessarily optimal) are measurable, then one expect some approximate continuity property to hold. And in fact this is the content of the next proposition. First we introduce some basic definition.

Definition 3.1. Let $\gamma \in \Pi(\mu, \nu)$ be a transport plan. For $y \in \Omega$ and $r > 0$ we define

$$\gamma^{-1}(B(y, r)) := \pi^1(spt(\gamma) \cap (\Omega \times B(y, r))).$$

In other words $\gamma^{-1}(B(y, r))$ is the set of those points whose mass (with respect to μ) is partially or completely transported to $B(y, r)$ by γ. We may justify this slight abuse of notations by the fact that γ should be thought of as a device that transports mass.

Since this notion is important in the sequel, we recall that when A is $\mathcal{L}^d$-measurable, one has

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

for almost every x in A: we shall call such a point x a Lebesgue point of A, this terminology deriving from the fact that such a point may also be considered as a Lebesgue point of χ_A. In the following, we shall denote by $\mathrm{Leb}(A)$ the set of points $x \in A$ which are Lebesgue points of A. We also define the lower density of A at x as:

$$\theta_*(A, x) := \liminf_{r \to 0} \frac{\mathcal{L}^d(A \cap B(x, r))}{\mathcal{L}^d(B(x, r))}.$$

The following Lemma details the meaning of approximate continuity for a transport plan. Its statement and proof are simplifications of that of Lemma 5.2 from [13] and we report it for the convenience of the reader.

Lemma 3.2. *Let $\gamma \in \Pi(\mu, \nu)$. If $\mu \ll \mathcal{L}^d$, then γ is concentrated on a set $R(\gamma)$ such that for all $(x, y) \in R(\gamma)$ the point x is a Lebesgue point of $\gamma^{-1}(B(y, r))$ for all $r > 0$.*

Proof. Let

$$A := \{(x, y) \in \ spt(\gamma) \ : \ \ x \notin \mathrm{Leb}(\gamma^{-1}(B(y, r))) \text{ for some } r > 0\};$$

we intend to show that $\gamma(A) = 0$. To this end, for each positive integer n we consider a finite covering $\Omega \subset \bigcup_{i \in I(n)} B(y_i^n, r_n)$ by balls of radius $r_n := \frac{1}{2n}$. We notice that if $(x, y) \in \ spt(\gamma)$ and x is not a Lebesgue point of $\gamma^{-1}(B(y, r))$ for some $r > 0$, then for any $n \geq \frac{1}{r}$ and y_i^n such that $|y_i^n - y| < r_n$ the point x belongs to $\gamma^{-1}(B(y_i^n, r_n))$ but is not a Lebesgue point of this set. Then

$$\pi^1(A) \ \subset \bigcup_{n \geq 1} \ \bigcup_{i \in I(n)} \left(\gamma^{-1}(B(y_i^n, r_n)) \setminus \mathrm{Leb}(\gamma^{-1}(B(y_i^n, r_n)))\right).$$

Notice that the set on the right hand side has Lebesgue measure 0, and thus μ-measure 0. It follows that $\gamma(A) \leq \gamma(\pi^1(A) \times \Omega) = \mu(\pi^1(A)) = 0$. In conclusion the set $R(\gamma) = spt(\gamma) \setminus A$ has the desired property. $\square$

The above Lemma yields us to introduce the following notion:

Definition 3.3. The couple $(x, y) \in spt(\gamma)$ is a *γ-regular point* if x is a Lebesgue point of $\gamma^{-1}(B(y, r))$ for any positive r.

Notice that any element of the set $R(\gamma)$ of Lemma 3.2 is a γ-regular point. Lemma 3.2 above therefore states that any transport plan γ is concentrated on a Borel set consisting of regular points: this regularity property turns out to be a powerful tool in the study of the support of optimal transport plans for problem (2.1), as the proof of Proposition 3.6 below illustrates.

Also the direction of transport (which is individuated as $-\nabla u(z)^*$) by Lemma 2.11 enjoys an additional regularity property. Let us first introduce the notion of *countable Lipschitz property*.

Definition 3.4. Let $S \subset \Omega$ be measurable, we say that a function $v : S \to \mathbb{R}^n$ has the countable Lipschitz property if there exists a sequence $\{S_k\}_k$ of measurable subsets of S such that $\mathcal{L}^d(S \setminus \bigcup_k S_k) = 0$ and the restriction of v to each S_k is Lipschitz.

The countable Lipschitz property is related to the concept of “approximate differentiability” and it is, in particular enjoyed by BV functions (see, for example, th. 5.34 of [3]).

The proof follows [5].

Proposition 3.5. *Let u be a Kantorovich potential then there exists a sequence of Borel sets F_k such that $\mathcal{L}^d(\mathcal{T}_u^e \setminus \bigcup_k F_k) = 0$ and such that the map $x \mapsto \nabla u(x)^*$ restricted to F_k is Lipschitz.*

Proof. Let $\xi \in \mathcal{S}^{d-1}$ be a direction and $a \in \mathbb{R}$. Let R be the union of half closed transport rays $[x, y[$ and $R_{\xi,a}$ the union of half closed transport rays $[x, y[$ whose direction satisfies $\langle \xi, \frac{y-x}{\|y-x\|} \rangle > 0$ and which ends above the hyperplane individuated by ξ and a, i.e. $\langle \xi, y \rangle \geq a$. We first prove that ∇u has the countable Lipschitz property claimed by the statement on the set $S_{\xi,a} = R_{\xi,a} \cap \{x \ : \ \langle \xi, x \rangle < a\}$. Since BV_{loc} functions have the countable Lipschitz property, to prove this one can show that ∇u coincide a.e. on $S_{\xi,a}$ with a function $v \in BV_{loc}(\{x \ : \ \langle \xi, x \rangle < a\}, \mathbb{R}^d)$. Consider the set $Y_{\xi,a}$ of right-ends of transport rays involved in $R_{\xi,a}$ and the function

$$\tilde{u}(x) := \min_{y \in Y_{\xi,a}} u(y) + \|x - y\|.$$

Since u and $\tilde{u}$ coincide on $Y_{\xi,a}$ and $\tilde{u}$ is the bigger $1-$Lipschitz extension of $u_{|Y_{\xi,a}}$

$$\tilde{u} \geq u$$

on the other hand by definition of transport ray

$$u = \tilde{u} \quad \text{on } S_{\xi,a}.$$

For $b < a$, thanks to the control above in (1.3), there exists a constant $C(b)$ such that

$$\tilde{u} - C(b)|x|^2$$

is concave in $\{x \ : \ \langle \xi, x \rangle < b\}$. Indeed in for every $x \in \{x \ : \ \langle \xi, x \rangle < b\}$ and $y \in Y_{\xi,a}$, we have $\|x - y\| \geq a - b$ and then for all $y \in Y_{\xi,a}$ the function

$$x \mapsto u(y) + \|x - y\| - C(b)|x|^2$$

is concave in $\{x \ : \ \langle \xi, x \rangle < b\}$. Since gradients of concave functions are BV_{loc} we obtain that

$$v := \nabla \tilde{u} = \nabla(\tilde{u} - C(b)|x|^2) + 2C(b)x$$

is BV_{loc} and then it enjoys the countable Lipschitz property in $\{x \ : \ \langle \xi, x \rangle < b\}$ and, as consequence in $S_{\xi,a} \cap \{x \ : \ \langle \xi, x \rangle < b\}$. Consider a sequence $b_n \to a^-$ since $S_{\xi,a} = \cup_n (S_{\xi,a} \cap \{x \ : \ \langle \xi, x \rangle < b\}$ we conclude that ∇u has the countable Lipschitz property in $S_{\xi,a}$. Considering countable and dense sets of directions ξ_n and real numbers a_n one obtain that ∇u has the countable Lipschitz property in R which includes the open transport set and only the starting points of transport rays. In order to include also the final points of transport rays, we make a similar construction using the lower Lipschitz extension and we conclude that ∇u has the countable Lipschitz property in $\mathcal{T}_u^e$. Finally it is enough to recall that, thanks to the control from below in (1.3), the duality map which associate to a unit vector $L \in (\mathbb{R}^d)^*$ the unique unit vector $L^* \in \mathbb{R}^d$ such that $L(L^*) = 1$ is Lipschitz and then also ∇u^* enjoy the countable Lipschitz property. $\square$

In the next proposition we prove that (γ-essentially) if x_0 is partly moved to y_0 the transport set from a neighborhood of x_0 to a neighborhood of $y_0$has positive density at x_0. Since the definition of transport ray and Proposition 3.5 require that $x \neq y$ we will need to assume this in the proposition.

Proposition 3.6. *Let $\gamma \in \mathcal{O}_1(\mu, \nu)$, let u be a Kantorovich potential and let $\{F_k\}_k$ be the sets associated to the countable Lipschitz property of ∇u^*. Let $(x_0, y_0) \in R(\gamma)$ with $x_0 \neq y_0$ and $x_0 \in \mathrm{Leb}(F_k)$ (for some k) then for all $s > 0$*

$$\theta_*(T(\{(x,y) \in spt(\gamma) \ : \ x \in F_k \cap B(x_0, s), \ y \in B(y_0, s)\}), x_0) = 1.$$

Then x_0 is a Lebesgue point for the transport set.

Proof. We need to estimate from below the

$$\liminf_{r \to 0} \frac{\mathcal{L}^d(T(\{(x,y) \in spt(\gamma) \ : \ x \in F_k \cap B(x_0, s), \ y \in B(y_0, s)\}) \cap B(x_0, r))}{\mathcal{L}^d(B(x_0, r))},$$

then without loss of generality we may assume $r < s$ and $B(x_0, s) \cap B(y_0, s) = \emptyset$. For any t such that $0 < t \ll \|x_0 - y_0\|$ and any $(x, y) \in spt(\gamma)$ such that $x \in F_k \cap B(x_0, s)$ and $y \in B(y_0, s)$ we have that

$$x - t\nabla u(x)^* \in T(\{(x,y) \in spt(\gamma) \ : \ x \in F_k \cap B(x_0, s), \ y \in B(y_0, s)\}).$$

If moreover, for some positive integer n, $t < \dfrac{r}{n}$ and $x \in B(x_0, \dfrac{(n-1)r}{n})$ then $x - t\nabla u(x)^* \in B(x_0, r)$. Since x_0 is a Lebesgue point for both F_k and $\gamma^{-1}(B(y_0, s))$ it is also a Lebesgue point for the intersection

$$P_k := F_k \bigcap \gamma^{-1}(B(y_0, s)).$$

Summing up the last three observations we obtain that for r sufficiently small and for any $t < \dfrac{r}{n}$

$$\begin{aligned}\mathcal{L}^d(\{x - t\nabla u(x)^* \ : \ x \in P_k \cap B(x_0, \frac{(n-1)r}{n})\}) \leq \\ \mathcal{L}^d(T(\{(x,y) \in spt(\gamma) \ : \ x \in F_k \cap B(x_0, s), \ y \in B(y_0, s)\}) \cap B(x_0, r))\end{aligned}$$

Since on F_k the map $-\nabla u^*$ coincide with a Lipschitz map G_k of Lipschitz constant L_k we also have that for any t the map $x - t\nabla u^*(x)$ coincide with the Lipschitz map $Id + tG_k$ on F_k. Moreover we may choose t sufficiently small so that $\dfrac{n-1}{n} \leq$

$|det(Id+tDG_k)|$ then by the area formula (applied to the Lipschitz function $Id+tG_k$ which is, by equation (2.6), injective in P_k thanks to the cyclical monotonicity)

$$\frac{n-1}{n}\mathcal{L}^d(P_k\cap B(x_0,\frac{(n-1)r}{n}))\leq\int_{P_k\cap B(x_0,\frac{(n-1)r}{n})}|det(Id+tDG_k)|dx=$$
$$\mathcal{L}^d(\{x-t\nabla u(x)^*\ :\ x\in P_k\cap B(x_0,\frac{(n-1)r}{n})\}).$$

Now we can estimate

$$\liminf_{r\to 0}\frac{\mathcal{L}^d(T(\{(x,y)\in spt(\gamma)\ :\ x\in F_k\cap B(x_0,s),\ y\in B(y_0,s)\})\cap B(x_0,r))}{\mathcal{L}^d(B(x_0,r))}$$
$$\geq\liminf_{r\to 0}\frac{\mathcal{L}^d(\{x-t\nabla u^*(x)\ :\ x\in P_k\cap B(x_0,\frac{(n-1)r}{n})\}\cap B(x_0,r))}{\mathcal{L}^d(B(x_0,r))}$$
$$\geq\lim_{r\to 0}(\frac{n-1}{n})\frac{\mathcal{L}^d(P_k\cap B(x_0,\frac{(n-1)r}{n}))}{\mathcal{L}^d(B(x_0,r))}=(\frac{n-1}{n})^{d+1}.$$

The conclusion follows since n can be chosen arbitrarily big. $\square$

Definition 3.7. Let $\gamma\in\mathcal{O}_1(\mu,\nu)$ and let u be a Kantorovich potential. Denote by $F=\cup_k\text{Leb}(F_k)$ where the sets F_k are the sets which appear in the countable Lipschitz property of ∇u^*. We will denote by

$$D(\gamma,u)=R(\gamma)\cap\{(x,y)\ :\ x\in F\}$$

Notice that $D(\gamma,u)$ adds a constraint only on the first coordinate of points in $R(\gamma)$.

Proposition 3.8. *Assume that $\mu\ll\mathcal{L}^d$. Then every $\gamma\in\mathcal{O}_2(\mu,\nu)$ is induced by a transport map T_γ, i.e. $\gamma=(id\times T_\gamma)_\sharp\mu$.*

Proof. Let $\gamma\in\mathcal{O}_2(\mu,\nu)$, by Lemma 2.4 it is enough to prove that γ is supported on a graph. Fix a Kantorovich potential u. We show that if (x_0,y_0) and (x_0,y_1) both belong to $D(\gamma,u)$ then $y_0=y_1$. By contradiction assume that $y_0\neq y_1$. Then one either has $(y_1-y_0)\cdot(y_0-x_0)<0$ or $(y_0-y_1)\cdot(y_1-x_0)<0$. Without loss of generality, we assume that

$$(y_0-y_1)\cdot(y_1-x_0)\ <\ 0,\tag{3.1}$$

which in particular implies $y_1\neq x_0$.

We fix $s>0$ small enough so that

$$\forall x\in B(x_0,s),\forall y\in\overline{B(y_0,s)},\forall y'\in\overline{B(y_1,s)},\qquad (y-y')\cdot(y'-x)\ <\ 0.\tag{3.2}$$

Since, by definition of $R(\gamma)$, the sets $\gamma^{-1}(B(y_0,s))$ and $\gamma^{-1}(B(y_1,s))$ both have density 1 at x_0 and, by Proposition 3.6, the set $T(\{(x,y)\in spt(\gamma)\ :\ x\in F_k\cap B(x_0,s),\ y\in B(y_1,s)\})$ has also density 1 at x_0 we infer that for r small enough there exist $\tilde{x}\in B(x_0,r)$ which belongs to the intersection of the three sets. In other words there exists $(\overline{x},\overline{y}_1)$, $(\tilde{x},\tilde{y}_0)$, $(\tilde{x},\tilde{y}_1)\in support(\gamma)$ such that $\overline{x}\in F_k\cap B(x_0,s)$, $\overline{y}_1\in B(y_1,s)$ $\tilde{x}=\overline{x}+t(\overline{y}_1-\overline{x})$ for some $t\in(0,1)$, $\tilde{y}_0\in B(y_0,s)$ and $\tilde{y}_1\in B(y_1,s)$. Since $\tilde{x}$ lies on the segment between $\overline{x}$ and $\overline{y}_1$, it follows from (2.8) applied to $(\overline{x},\overline{y}_1)$ and $(\tilde{x},\tilde{y}_0)$ that

$$(\tilde{y}_0-\overline{y}_1)\cdot(\tilde{x}-\overline{x})\geq 0$$

but since $\tilde{x}-\overline{x}=t(\overline{y}_1-\overline{x})$ this contradicts (3.2). Now we consider the set

$$B=\{(x,y)\in R(\gamma)\ :\ 1<\sharp(\{x\}\times\mathbb{R}^d\cap R(\gamma)),\ x\notin F\}.$$

We prove that $\gamma(B)=0$. In fact $\gamma(B)\leq\mu(\pi_1(B))$ and if $(x,y)\in B$ then there exists at least one $y_1\neq x$ such that $(x,y_1)\in R(\gamma)$ and then $x\in\mathcal{T}_u^e$ but, by definition $x\notin F$ and since μ is absolutely continuous, using the countable Lipschitz property of ∇u^*, we obtain $\mu(\mathcal{T}_u^e\setminus F)=0$. It remains to remark that γ is supported on $R(\gamma)$ and then to decompose first

$$R(\gamma)=D(\gamma,u)\cup[R(\gamma)\setminus D(\gamma,u)],$$

and then

$$R(\gamma)\setminus D(\gamma,u)=[(R(\gamma)\setminus D(\gamma,u))\cap B]\cup[(R(\gamma)\setminus D(\gamma,u))\setminus B].$$

We already proved that $D(\gamma,u)$ is contained in a graph and since, as we already observed, a point in $D(\gamma,u)$ and a point $R(\gamma)\setminus D(\gamma,u))$ cannot have the same first coordinate it is enough that now we study $R(\gamma)\setminus D(\gamma,u))$. We divided this last set in two parts, for the first

$$\gamma((R(\gamma)\setminus D(\gamma,u))\cap B)=0$$

since it is a subset of B, the second $(R(\gamma)\setminus D(\gamma,u))\setminus B$ is contained in a graph by definition of B. This concludes the proof of the fact that any element $\gamma\in\mathcal{O}_2(\mu,\nu)$ is induced by a transport map $T_\gamma\in\mathcal{T}(\mu,\nu)$. □

Finally we prove by a standard method that under the assumptions of the theorem above $\mathcal{O}_2(\mu,\nu)$ has only one element.

Proposition 3.9. *Assume that $\mu\ll\mathcal{L}^d$. Then there is only one $\gamma\in\Pi(\mu,\nu)\cap\mathcal{O}_2(\mu,\nu)$.*

Proof. The uniqueness of $\gamma\in\Pi(\mu,\nu)\cap\mathcal{O}_2(\mu,\nu)$ is obtained as in Step 5 of the proof of Theorem B in [4]: if γ_1 and γ_2 are two such transport plans, then $\frac{\gamma_1+\gamma_2}{2}$ also belongs to $\Pi(\mu,\nu)\cap\mathcal{O}_2(\mu,\nu)$. It follows from the Proposition 3.8 that these plans are all induced by transport maps, which then coincide μ almost everywhere. □

REFERENCES

[1] N. Ahnad, H. K. Kim, R. McCann, Optimal transportation, topology and uniqueness. To appear on *Bull. Math. Sci.*. Preprint, temporarily available on http://www.math.toronto.edu/mccann/

[2] Ambrosio, L., Lecture Notes on Optimal Transportation Problems, *Mathematical aspects of evolving interfaces (Funchal, 2000), Lecture Notes in Math.*, **1812**, Springer, Berlin, 2003, 1–52.

[3] Ambrosio, L., Fusco, N., Pallara, D., Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.

[4] L. Ambrosio, B. Kirchheim, A. Pratelli, Existence of optimal transport maps for crystalline norms, *Duke Math. J.*, **125** (2004), no. 2, 207–241.

[5] L. Ambrosio, A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation, *Optimal transportation and applications (Martina Franca, 2001), Lecture Notes in Math.*, **1813**, Springer, Berlin, 2003, 123–160.

[6] G. Anzellotti, S. Baldo, Asymptotic development by Γ-convergence, *Appl. Math. Optim.* **27** (1993), no. 2, 105–123.

[7] H. Attouch, Viscosity solutions of minimization problems, *SIAM J. Optim.*, **6** (1996), no. 3, 769–806.

[8] S.Bianchini, L. Caravenna, On optimality of $c-$cyclically monotone transference plans, *C. R. Math. Acad. Sci. Paris*, **348** (2010), no. 11-12, 613618.

[9] L.A. Caffarelli, M. Feldman, R. J. McCann, Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs, *J. Amer. Math. Soc.*, **15** (2002), no. 1, 1–26.

[10] L. Caravenna, A partial proof of Sudakov theorem via disintegration of measures, *To appear on Mathematische Zeitschrift, published on line DOI: 10.1007/s00209-010-0677-6*, Preprint SISSA (2008) available at http://cvgmt.sns.it.

[11] T. Champion, L. De Pascale, The Monge problem for strictly convex norms in $\mathbb{R}^d$, *J.Europ.Math.Soc.*, **12** (2010), no.6, 1355–1369.

[12] T. Champion, L. De Pascale, The Monge problem in $\mathbb{R}^d$, *Duke Math. J.*, **157** (2011), no.3, 551–572.

[13] T. Champion, L. De Pascale, P. Juutinen, The ∞-Wasserstein distance: local solutions and existence of optimal transport maps, *SIAM J. of Mathematical Analysis*, **40** (2008), no. 1, 1-20.

[14] L. C. Evans, W. Gangbo, Differential Equations Methods for the Monge–Kantorovich Mass Transfer Problem, *Mem. Amer. Math. Soc.*, **137** (1999).

[15] D. Feyel, A. S. Üstünel, Monge-Kantorovitch measure transportation and Monge-Ampre equation on Wiener space. *Probab. Theory Related Fields*, **128** (2004), no. 3, 347–385.

[16] L.V. Kantorovich, On the translocation of masses, *C.R. (Dokl.) Acad. Sci. URSS*, **37** (1942), 199-201.

[17] L.V. Kantorovich, On a problem of Monge (in Russian), *Uspekhi Mat. Nauk.*, **3** (1948), 225-226.

[18] G. Monge, Mémoire sur la théorie des Déblais et des Remblais, *Histoire de l'Académie des Sciences de Paris*, 1781.

[19] A. Pratelli, On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation, *Ann. Inst. H. Poincaré Probab. Statist.*, **43** (2007), no. 1, 1–13.

[20] A. Pratelli, On the sufficiency of $c-$cyclical monotonicity for optimality of transport plans. *Math. Z.*, **258** (2008), no. 3, 677690.

[21] L. Rüschendorf, On $c-$optimal random variables. *Statist. Probab. Lett.*, **27** (1996), no. 3, 267270.

[22] F. Santambrogio, Absolute continuity and summability of optimal transport densities: simpler proofs and new estimates, *Calc. Var. Partial Differential Equations* , **36** (2009), no. 3, 343–354.

[23] V. N. Sudakov, Geometric problems in the theory of infinite-dimensional probability distributions. *Cover to cover translation of Trudy Mat. Inst. Steklov* **141** (1976). *Proc. Steklov Inst. Math.* 1979, no. 2, i–v, 1–178.

[24] N. S. Trudinger, X. J.Wang, On the Monge mass transfer problem, *Calc. Var. Partial Differential Equations*, **13** (2001), no. 1, 19–31.

[25] C. Villani, Topics in optimal transportation. *Graduate Studies in Mathematics*, **58**, American Mathematical Society (2003)

[26] C. Villani, Optimal Transport, Old and New. *Grundlehren der mathematischen Wissenschaften*, **338**, Springer Berlin Heidelberg (2008)