The Monge Problem in R^d: Variations on a Theme

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

THIERRY CHAMPION AND LUIGI DE PASCALE

Abstract. In a recent paper the authors proved that, under natural assumptions on the first marginal, the Monge problem in $\mathbb{R}^d$ for cost given by a general norm admits a solution. Although the basic idea of the proof is simple, it involves some complex technical results. Here we will give a proof of the result in the simpler case of uniformly convex norm and we will also use very recent results by other authors [1]. This allows 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 the particular case of this paper is original.

1. Introduction

The Monge problem has origin in the *Mémoire sur la théorie des déblais et remblais* written by G. Monge [15]. 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 minimizing the work done in the transportation process? 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 Borel probability measures μ and ν to model respectively the sand pile and 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 from μ to ν then the problem will take the form

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

where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^d$.

The natural interest of the problem and the many applications attracted the interest, through the years, to a generalization in which one considers a general norm $\|\cdot\|$ on $\mathbb{R}^d$, which leads to the formulation

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

Date: June 1, 2011.

2000 *Mathematics Subject Classification.* 49Q20, 49K30, 49J45.

Key words and phrases. Monge-Kantorovich problem, optimal transport problem, cyclical monotonicity.

A first strategy to prove that the problem (1.2) admits a solution was devised by Sudakov in [19]. 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 3. 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. This method unfortunately involved a crucial step on the disintegration of measures which was not completed correctly at that time, and has recently been justified in the case of a strictly convex norm $\|\cdot\|$ by Caravenna [8]. Meanwhile, the problem (1.1) has been solved by Evans *et al.* [12] 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.* [20] for μ and ν with integrable density. The more general problem (1.2) for C^2 uniformly convex norms has been solved independently by Caffarelli *et al.* [7] and Ambrosio *et al.* [4], and for crystalline norms in $\mathbb{R}^d$ and general norms in $\mathbb{R}^2$ by Ambrosio *et al.* [3]. As for the original proof of Sudakov, all the proofs of the above listed existence results are based on the reduction of the problem to a 1-dimensional problem via a change of variable or area-formula. In [9, 10], we introduced a different method to prove the existence of a solution for (1.2) which does not require the reduction to 1-dimensional settings.

1.1. **This paper.** Althought the aim of this paper is mainly expository we will try to introduce some technical novelty which should give easier access to non specialists. In Section §2 we introduce the general theory and classical facts developed to solve (1.2). Then in Section §3 we introduce the classical notion of transport sets. Finally in section §4 we will follow the strategy of proof of existence for (1.2) developed in [9, 10, 11], but with the additional assumption that the norm $\|\cdot\|$ is uniformly convex, i.e. $\|\cdot\|^2$ is of class $\mathcal{C}^2$ on $\mathbb{R}^d$ with

$$c\,\mathrm{id_d} \leq \mathrm{D}^2(\|\cdot\|^2) \leq \mathrm{C}\,\mathrm{id_d} \quad \text{for some} \;\; 0 < \mathrm{c} \leq \mathrm{C}. \tag{1.3}$$

This considerably reduces the technical burdens of our original proof while leaving intact the main ideas. Some more simplifications are introduced thanks to some technical novelty which appeared after [9, 10]. The proof of Proposition 4.4, which is a cornerstone of our main existence result Theorem 4.5, is original. We hope that this paper will make the problem accessible to a wider audience.

2. The main players and their basic properties

Most of the results of this section and the following section 3 are by now classical, and may be found for example in [2, 4, 21, 22], unless otherwise stated. We shall give some proofs for the convenience of the reader.

2.1. **Relaxation.** The first step to solve (1.2) consists in suitably relaxing the problem. This was done by Kantorovich [13, 14] who introduced the set

$$\Pi(\mu,\nu) = \{\gamma \ \text{ a probability on } \ \Omega\times\Omega \ \text{ such that } \ \pi^1_\sharp\gamma = \mu,\ \pi^2_\sharp\gamma = \nu\},$$

and the cost

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

The elements of $\Pi(\mu,\nu)$ are called transport plans and, as tools to transport μ to ν they allow the mass located at a point x to be split among many points y while a transport map T moves all the mass being at x to $T(x)$. There is a natural embedding of $\mathcal{T}(\mu,\nu)$ 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

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

The inescapable question is whether or not

$$\inf_{T\in\mathcal{T}(\mu,\nu)} \int_\Omega ||x-T(x)|| d\mu = \min_{\gamma\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \|x-y\| d\gamma. \tag{2.2}$$

Since the main result of this work, Theorem 4.5 below, asserts that if μ is absolutely continuous with respect to $\mathcal{L}^d$ then some optimal transport plan for (2.1) is induced by an optimal transport map, the equality follows in that case. However it can be proved the more general following result.

Proposition 2.1. *If μ has no atom then (2.2) holds.*

For a proof of Proposition 2.1 in wide generality we refer to [16] and reference therein.

The assumption that μ is non atomic cannot be removed since otherwise the set $\mathcal{T}(\mu,\nu)$ may be empty, on the other hand $\mu\otimes\nu$ always belongs to $\Pi(\mu,\nu)$.

Example 2.2. Let $\mu := \delta_0$ and $\nu := \frac{1}{2}(\delta_1+\delta_{-1})$, in this case the set $\mathcal{T}(\mu,\nu)$ is easily seen to be empty.

For general marginals μ and ν it may happen that $\mathcal{T}(\mu,\nu)$ is non empty but the inf in (1.2) is not attained, while the problem 2.1 (under the current assumptions) always has a minimizer.

Example 2.3. In $\mathbb{R}^2$, let $S_t = \{t\}\times[0,1]$ for $t\in\{-1;0;1\}$. Also set $\mu := \mathcal{H}^1\lfloor S_0$ 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. We consider the case where $\|\cdot\|$ is the Euclidean norm. Then the optimal transport plan for (2.1) will move half of the mass horizontally to the right and the other half horizontally to the left, for a total cost equal to 1. This cannot be achieved by any transport map in $\mathcal{T}(\mu,\nu)$.

Since by Proposition 2.1 the values of the original problem (1.2) and the relaxed one (2.1) coincide, a natural way to obtain the existence of a solution for (1.2) is to show that some minimizer for (2.1) is induced by a transport map. This is indeed the strategy of this paper, which reduces to prove that some minimizer of (2.1) is supported on a graph. Then a natural (although technical) question is whether a transport plan supported on a graph is induced by a transport map or not. This is the topic of Lemma 2.1 in [1], and the aim of this paper being partly expository we report the short proof below.

Lemma 2.4 ([1])**.** *Let X and Y be compact subsets of $\mathbb{R}^d$, 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. First notice that γ is a regular measure since it is σ-finite and Borel on a complete and separable metric space. Then since $\gamma(X \times Y \setminus (Graph(T))) = 0$ there exists an increasing sequence $(K_i)_i$ of compact subsets of $Graph(T)$ such that $K_\infty := \bigcup_i K_i \subset Graph(T)$ 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, and then it is μ-measurable. We now check that $\gamma = (id \times T_\infty)_\sharp \mu$. Indeed let $U \times V$ be any Borel *"rectangle"* then

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

And this implies the thesis. $\square$

2.2. **Dual problem and Kantorovich potentials.** Problem (2.1) is a linear minimization problem with convex constraints then an important tool to deal with it is duality theory. In our case, as the Proposition below shows, the classical convex dual problem for (2.1) is given by

$$\max_{u \in Lip_1(\Omega, \|\cdot\|)} \int_\Omega u(x)d\mu - \int_\Omega u(y)d\nu \tag{2.3}$$

where $Lip_1(\Omega, \|\cdot\|) = \{u \, : \, u(x) - u(y) \leq \|x - y\| \text{ for all } x, y \in \Omega\}$. We call a maximizer u of that dual problem a Kantorovich potential for (2.3). The existence of such maximizers and the link between (2.1) and (2.3) are formalized in the next Proposition.

Proposition 2.5. *Under the current assumptions one has*

$$\min(2.1) \;=\; \max(2.3)$$

Moreover if u is a maximizer for (2.3), then $\gamma \in \Pi(\mu,\nu)$ is optimal for (2.1) if and only if $u(x)-u(y)=\|x-y\|$ on the support of γ.

Proof. We first notice that the existence of a maximizer for (2.3) 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 of (2.1) and (2.3) we use the convex duality theory. We first write

$$\max(2.3) \;=\; \sup\left\{\int_\Omega u\,d\mu + \int_\Omega v\,d\nu \;:\; \forall x,y, \quad u(x)+v(y)\le \|x-y\|\right\}. \quad (2.4)$$

It is indeed clear that the maximum of (2.3) is lower than the sup of the right hand side. 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 (2.4) 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 of the right hand side of (2.4) as follows:

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

Notice in particular that $h(0)=-\max(2.3)$. 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 measure γ is not non-negative):

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

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 neighbourhood of 0 for the uniform convergence: since h is also convex, it follows that it is Lipschitz continuous in a neighbourhood of 0. The equality between the extremal values of (2.1) and (2.3) follows.

To prove the last part of the statement, take a maximizer u for (2.3) 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. $\square$

Remark 2.6. 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 couples of points $(x_1, y_1),\ (x_2, y_2) \in$ *support*(γ) it holds

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

Indeed, by the previous theorem if $u \in Lip_1(\Omega, \|\cdot\|)$ is a Kantorovich potential for (2.3) then the above characterization of optimality for γ yields

$$\begin{aligned} \|x_1 - y_1\| + \|x_2 - y_2\| &= u(x_1) - u(y_1) + u(x_2) - u(y_2) \\ &= u(x_1) - u(y_2) + u(x_2) - u(y_1) \leq \|x_1 - y_2\| + \|x_2 - y_1\|. \end{aligned}$$

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 [17].

2.3. **Selection of particular solutions.** It is a well known fact that the problem (1.2) may have several solutions, in which case the problem (2.1) admits solutions that are not induced by a transport map. We recall this in the following example.

Example 2.7. Consider the case where $\mu = \mathcal{L}^1 \lfloor [0, 1]$ and $\nu = \mathcal{L}^1 \lfloor [1, 2]$, then $T_+ : x \mapsto x + 1$ and $T_- : x \mapsto 2 - x$ are both solutions of (1.2), and then the transport plan $\gamma := \frac{1}{2}[(id \times T_+)_\sharp \mu + (id \times T_-)_\sharp \mu]$ is optimal for (2.1) but is not associated with any transport map. Notice that in this case an explicit Kantorovich potential may be computed and it is given by $u(x) = 1 - x$.

As a consequence of the above example, we can not expect to show that any solution of (2.1) is indeed induced by a transport map, so we have to select particular solutions of (2.1) that achieve this property. Here we adopt the strategy of [4] that consists in selecting those solutions of (2.1) that are monotone non-decreasing (in a sense that we precise below) through an auxiliary problem. In the above example, this reduces to isolate the monotone non-decreasing transport map T_+.

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

$$\min\left\{\int_{\Omega\times\Omega}|y-x|^2 d\gamma(x,y)\ :\ \gamma\in\mathcal{O}_1(\mu,\nu)\right\}, \tag{2.5}$$

where we remark the fact that the cost in consideration 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 [5] and [6]) . Notice that (2.5) admits solutions since the set $\mathcal{O}_1(\mu,\nu)$ is a weakly compact subset of $\Pi(\mu,\nu)$.

The fact that the solutions of (2.5) do satisfy some sort of non-decreasing monotony comes from the following. As a solution of (2.5), a transport plan γ enjoys a cyclical monotonicity property inherited from the cost $(x,y)\mapsto|y-x|^2$ (see remark 2.9 below), stated in the following proposition, whose proof may be derived from that of Lemma 4.1 in [3] and is given in [9] (see Proposition 3.2 therein).

Proposition 2.8. *Let γ be a solution of (2.5), then γ is concentrated on a Borel set Γ with the following property:*

$$\forall (x,y),(x',y')\in\Gamma,\qquad x\in[x',y']\ \Rightarrow\ (x-x')\cdot(y-y')\geq 0\,, \tag{2.6}$$

where $\cdot$ stands for the usual scalar product in $\mathbb{R}^d$.

Remark 2.9. To explain condition (2.6) above, we notice that for a minimizer λ of

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

the support of λ satisfies a 2-cyclical monotonicity condition with respect to the cost $|x-y|^2$, which states that for any couple of points (x_1,y_1), $(x_2,y_2)\in$ *support*(λ) one has

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

which is equivalent to

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

The measure γ involved in Proposition 2.8 is a minimizer for the constrained version (2.5) of (2.7) and the additional request $x\in[x',y']$ in (2.6) is a consequence of the constraint $\gamma\in\mathcal{O}_1(\mu,\nu)$.

3. Transport sets

3.1. **Definitions of transport sets.** Given a Kantorovich potential u for (2.3), the duality result given in Proposition 2.5 shows that an admissible transport plan γ is optimal for (2.1) if and only if

$$u(x) - u(y) \ = \ \|x - y\| \tag{3.1}$$

for all (x, y) in the support of γ. In this respect, the optimality of γ is characterized by its support, and a necessary condition for optimality may be written in the following way:

$$\mathcal{T}(\text{support}(\gamma)) \ \subset \ \mathcal{T}(\{(x, y) \,:\, x \neq y,\ u(x) - u(y) = \|x - y\|\}) \tag{3.2}$$

where the open transport set $\mathcal{T}(\Sigma)$ associated to a subset Σ of Ω^2 is given by

$$\mathcal{T}(\Sigma) \ := \ \bigcup_{(x,y)\in\Sigma}]x, y[\,.$$

Intuitively, $\mathcal{T}(\text{support}(\gamma))$ is the union of all the open segments $]x, y[$ along which some mass is transported by the transport plan γ (this obviously excludes the points $(x, x) \in \text{support}\,\gamma$ which correspond to the part of the mass that is not moved by γ), so that it is understood to be the transport set for γ. As we shall see in Section §4, a cornerstone for the existence proof of our main result Theorem 4.5 is that this transport set $\mathcal{T}(\text{support}(\gamma))$ satisfies some regularity property whenever γ is a solution of (2.5) (see Proposition 4.4 hereafter).

In fact, it is more convenient to study the regularity of the transport set

$$\mathcal{T}_u \ := \ \mathcal{T}(\{(x, y) \,:\, x \neq y,\ u(x) - u(y) = \|x - y\|\})$$

associated to a Kantorovich potential $u \in Lip_1(\Omega, \|\cdot\|)$. By (3.2), the set $\mathcal{T}_u$ contains the transport set $\mathcal{T}(\text{support}(\gamma))$ of any solution of (2.1).

The transport set $\mathcal{T}_u$ is also classically defined as the union of all the transport rays :

$$\mathcal{T}_u \ := \ \bigcup_{]x,y[\,\in\mathcal{R}_u}]x, y[$$

where $\mathcal{R}_u$ is the set of transport rays for u : following [12], a non-empty open segment $]x, y[$ is called transport ray for u if it is a maximal, open and oriented segment whose end-points satisfy the condition (3.1). In the following, we study the slightly larger set $\mathcal{T}_u^e$ (where e stands for end-points) :

$$\mathcal{T}_u^e \ := \ \bigcup_{]x,y[\,\in\mathcal{R}_u} [x, y]$$

for which we obtain Lemma 3.2 and Proposition 3.4 below.

3.2. **Dual mapping in** $(R^d, \|\cdot\|)$**.** In this paper, $\mathbb{R}^d$ is always considered as endowed with its classical Euclidean norm $|\cdot|$ and scalar product $\cdot$, nevertheless the definitions of the transport set above via the identity (3.1) indicates that the dual norm $\|\cdot\|_*$ for $\|\cdot\|$ shall also play a role. The dual norm is given by

$$\xi \;\mapsto\; \|\xi\|_* := \max\{\xi \cdot z \,:\, \|z\| \leq 1\}$$

where the maximum is uniquely attained (when $\xi \neq 0$) because of the strict convexity of the unit ball $\{\|\cdot\| \leq 1\}$ which follows from (1.3). Therefore we can define the duality mapping $\xi \mapsto \xi^*$ on the unit sphere $\{\|\cdot\|_* = 1\}$ which associates to ξ the unique ξ^* in $\{\|\cdot\| = 1\}$ such that $\xi \cdot \xi^* = 1$. We now state a useful regularity result for this mapping.

Lemma 3.1. *Under (1.3), the duality mapping* $\xi \mapsto \xi^*$ *is Lipschitz on* $\{\|\cdot\|_* = 1\}$.

Proof. For convenience, we shall denote $\|\cdot\|^2$ by f in the following. It then follows from (1.3) that for any x, y one has

$$c\,|y-x|^2 \;\leq\; [\nabla f(y) - \nabla f(x)]\cdot(y-x) \;\leq\; C\,|y-x|^2 \tag{3.3}$$

Moreover we notice that if $\xi \in \{\|\cdot\|_* = 1\}$ has image ξ^* by the duality mapping then ξ^* is the unique solution to

$$\max\{\xi \cdot z \,:\, f(z) \leq 1\}\,,$$

so that we infer

$$\xi \;=\; \frac{1}{\nabla f(\xi^*)\cdot \xi^*}\nabla f(\xi^*)\,.$$

Now take ξ and ζ in $\{\|\cdot\|_* = 1\}$, and assume that $\nabla f(\xi^*)\cdot\xi^* \geq \nabla f(\zeta^*)\cdot\zeta^*$ (both being positive by (3.3)), then we compute

$$\begin{aligned}
(\xi-\zeta)\cdot(\xi^*-\zeta^*) \;&=\; \frac{1}{\nabla f(\xi^*)\cdot\xi^*}(\nabla f(\xi^*) - \nabla f(\zeta^*))\cdot(\xi^*-\zeta^*)\\
&\qquad + \frac{\nabla f(\xi^*)\cdot\xi^* - \nabla f(\zeta^*)\cdot\zeta^*}{\nabla f(\xi^*)\cdot\xi^*\;\nabla f(\zeta^*)\cdot\zeta^*}\nabla f(\zeta^*)\cdot(\zeta^*-\xi^*)\\
&\geq\; \frac{c}{\nabla f(\xi^*)\cdot\xi^*}|\xi^*-\zeta^*|^2
\end{aligned}$$

because $\nabla f(\zeta^*)\cdot(\zeta^*-\xi^*) \geq f(\zeta^*) - f(\xi^*) = 0$. As a consequence

$$|\xi^*-\zeta^*| \;\leq\; \frac{\nabla f(\xi^*)\cdot\xi^*}{c}|\xi-\zeta| \;\leq\; \frac{C\,m^2}{c}\,|\xi-\zeta|$$

where m denotes the maximum of $|.|$ on the unit sphere $\{\|\cdot\| = 1\}$. □

3.3. **Regularity of u over $\mathcal{T}_u^e$.** The Lemma 3.2 below shows that whenever u is differentiable at some z in $\mathcal{T}_u^e$, the image $-[\nabla u(z)]^*$ of its gradient by the dual mapping indicates the direction of the transport ray whose closure contains z (this transport ray being then unique). The Proposition 3.4 is a regularity result for $[\nabla u]^*$ (which exists a.e. since u is Lipschitz) on $\mathcal{T}_u^e$.

Lemma 3.2. *Let $u \in Lip_1(\Omega, \|\cdot\|)$ and $z \in \mathcal{T}_u^e$. If u is differentiable at z then*

$$\|\nabla u(z)\|_* = 1 \qquad \text{and} \qquad -[\nabla u(z)]^* = \frac{y-x}{\|y-x\|}$$

for the unique transport ray $]x,y[\in \mathcal{R}_u$ such that $z \in [x,y]$.

Proof. Let $]x,y[$ be a transport ray for u such that $z \in [x,y]$. Without loss of generality we may assume that $z \in [x,y[$. Since (3.1) holds, we get that

$$\forall t \in [\,0, \|z-y\|\,], \qquad u(z) = u\left(z + t\frac{y-x}{\|y-x\|}\right) + t$$

and then $-\nabla u(z) \cdot \frac{y-x}{\|y-x\|} = 1$. In particular $-\nabla u(z) \neq 0$. Since also $u(z) \leq u(z+t\zeta)+t$ for all t sufficiently small and ζ in the unit ball $\{\|\cdot\| \leq 1\}$, we infer that $-\nabla u(z)\cdot\zeta \leq 1$ for any such ζ. The conclusion of the Lemma follows from the definition of $\|\cdot\|_*$ and (1.3). $\square$

Remark 3.3. Proposition 2.5 and Lemma 3.2 indicate that in the case of a strictly 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 the other authors cited in the introduction. When the norm is not strictly convex one needs to consider the regions on which the transport potential u is affine and these regions may be higher dimensional affine submanifolds of $\mathbb{R}^d$.

We now state that the direction of transport (which is individuated as $-\nabla u$ by Lemma 3.2) enjoys some regularity property. The proof follows [4].

Proposition 3.4. *Let $u \in Lip_1(\Omega, \|\cdot\|)$, then there exists a sequence of Borel sets F_h such that $\mathcal{L}^d(\mathcal{T}_u^e \setminus \bigcup_h F_h) = 0$ and such that the gradient map $[\nabla u]^*$ restricted to F_h is Lipschitz for any h.*

Proof. We first set

$$Z := \bigcup_{[x,y[\in \mathcal{R}_u} [x,y[$$

and we show that ∇u has the countable Lipschitz property claimed by the statement on Z.

Let $\xi \in \{|\cdot| = 1\}$ be a direction in $\mathbb{R}^d$ and $a \in \mathbb{R}$. We define the sets

$$Y_{\xi,a} := \{y \,:\, \xi\cdot y > a \quad \text{and} \quad \exists x \in \Omega,\ ]x,y[\in \mathcal{R}_u\}$$

and

$$Z_{\xi,a} := \{z \in Z \,:\, \xi \cdot z \,<\, a\,, \quad z \in [x,y[\ \ for\ some\ ]x,y[\in \mathcal{R}_u\ with\ y \in Y_{\xi,a}\},$$

that is $Y_{\xi,a}$ is the set of the right-ends of transport rays contained in the hyperplane $\{z : \xi \cdot z \,>\, a\}$, and $Z_{\xi,a}$ is the union of the parts of the half-closed transport rays which end in $Y_{\xi,a}$ and belong to the hyperplane $\{z : \xi \cdot z \,<\, a\}$. We first show the countable Lipschitz property for ∇u on $Z_{\xi,a}$.

Since BV_{loc} functions have this countable Lipschitz property, it is sufficient to prove that ∇u coincides a.e. with a function in $BV_{loc}(Z_{\xi,a}, \mathbb{R}^d)$. Consider 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 largest $1-$Lipschitz extension of $u_{|Y_{\xi,a}}$ we have $\tilde{u} \geq u$. On the other hand by definition of transport rays it holds

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

For $b < a$ there exists a constant $K_a(b)$ such that

$$\tilde{u} - K_a(b)\,|\cdot|^2$$

is concave in $Z_{\xi,b}$. Indeed in for every $z \in Z_{\xi,b}$ and $y \in Y_{\xi,a}$ we have $|z-y| \geq a-b$, and then it follows from (1.3) that for all $y \in Y_{\xi,a}$ the function

$$z \mapsto u(y) + \|z - y\| - K_a(b)|z|^2$$

is concave in $Z_{\xi,b}$ for $K_a(b)$ large enough. Since gradients of concave functions are BV_{loc} we obtain that

$$\nabla\tilde{u} = \nabla(\tilde{u} - K_a(b)|.|^2) + 2K_a(b)\nabla(|.|^2)$$

is BV_{loc} and then it enjoys the countable Lipschitz property in $Z_{\xi,b}$. Taking a sequence $b_n \to a^-$, we conclude that ∇u has the countable Lipschitz property in $Z_{\xi,a}$. Finally, if remains to consider countable and dense sequences of directions $(\xi_n)_n$ and of real numbers $(a_n)_n$ to obtain that ∇u has the countable Lipschitz property in Z.

By a similar construction, using the lowest Lipschitz extension instead of the largest one as above, one can take into account the right ends of transport rays that form $\mathcal{T}_u^e$ and then conclude that ∇u has the countable Lipschitz property in $\mathcal{T}_u^e$. It remains to notice that $[\nabla u]^*$ inherits this property from ∇u by Lemmas 3.1 and 3.2. □

4. Finer properties and proof of the main theorem

4.1. **Regular points of the support of a transport plan.** 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 [11] 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 Lemma. First we introduce some basic definition.

Definition 4.1. Let $\gamma \in \Pi(\mu,\nu)$ be a transport plan, and $\Gamma \subset \mathrm{support}(\gamma)$ be a Borel set on which it is concentrated. For $y \in \Omega$ and $r > 0$ we define

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

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 the restriction of γ to 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 [11] and we report it for the convenience of the reader.

Lemma 4.2. *Let $\gamma \in \Pi(\mu,\nu)$ and $\Gamma \subset \mathrm{support}(\gamma)$ be a Borel set on which it is concentrated. If $\mu << \mathcal{L}^d$, then γ is concentrated on a subset $R(\Gamma)$ of Γ 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

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

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 $(x,y) \in$ support(γ) 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 $R(\Gamma) = \Gamma \setminus A$ has the desired property. □

The above Lemma yields us to introduce the following notion:

Definition 4.3. The couple $(x, y) \in \mathrm{support}(\gamma)$ 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 4.2 is a Γ-regular point. Lemma 4.2 above therefore states that any transport plan γ is concentrated on its Γ-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 4.4 below illustrates.

4.2. **Regularity of the transport sets of the solutions of (2.5).** In the next proposition we prove that, for a solution γ of (2.5), if the mass lying at x_0 is partly moved to y_0 by the transport plan γ, then the transport set from a neighborhood of x_0 to a neighborhood of y_0 has positive density at x_0.

Proposition 4.4. *Let $\gamma \in \mathcal{O}_1(\mu, \nu)$ by a solution of (2.5) and $\Gamma \subset \mathrm{support}(\gamma)$ be a Borel set on which it is concentrated. Let u be a Kantorovich potential and let $\{F_h\}_h$ be the sets associated to the countable Lipschitz property of $[\nabla u]^*$ by Proposition 3.4.*

Assume that (x_0, y_0) is a Γ-regular point with $x_0 \neq y_0$ and $x_0 \in \mathrm{Leb}(F_h)$ for some h, then for all $s > 0$

$$\theta_*(T(\Gamma \cap [(F_h \cap B(x_0, s)) \times B(y_0, s)]), x_0) \ > \ 0.$$

Proof. We need to estimate from below the quantity

$$\liminf_{r\to 0} \frac{\mathcal{L}^d(T(\Gamma \cap [(F_h \cap B(x_0, s)) \times B(y_0, s)]) \cap B(x_0, r))}{\mathcal{L}^d(B(x_0, r))},$$

then without loss of generality we may assume $r < s$ and $B(x_0, s) \cap B(y_0, s) = \emptyset$.

We first set

$$P_h := F_h \bigcap \Gamma^{-1}(B(y_0, s)).$$

For any t such that $0 < t << \|x_0 - y_0\|$ and any $x \in P_h \cap B(x_0, s)$, it follows from Lemma 3.2 that

$$x - t[\nabla u(x)]^* \in T(\Gamma \cap [(F_h \cap B(x_0, s)) \times B(y_0, s)]).$$

Furthermore, denoting by m the maximum of $|.|$ on the unit sphere $\{\|\cdot\|=1\}$, if $t<\dfrac{r}{2m}$ and $x\in B(x_0,\frac{r}{2})$ it follows that $x-t[\nabla u(x)]^*\in B(x_0,r)$. At this step, we have obtained that for r sufficiently small and for any $t<\dfrac{r}{2m}$ it holds

$$\begin{aligned}\mathcal{L}^d(\{x-t[\nabla u(x)]^* \ : \ x\in P_h\cap B(x_0,\frac{r}{2})\})\\ \leq \ \mathcal{L}^d(T(\Gamma\cap[(F_h\cap B(x_0,s))\times B(y_0,s)])\cap B(x_0,r)))\end{aligned}$$

Now since on F_h the map $-[\nabla u]^*$ coincides with a Lipschitz map G_h (with Lipschitz constant denoted by L_h) we also have that for any t the map $x-t[\nabla u]^*(x)$ coincides with the Lipschitz map $Id+tG_h$ on F_h. For $t<\frac{1}{L_h}$ the Lipschitz function $Id+tG_h$ is injective, and we may also choose t sufficiently small so that $\dfrac{1}{2}\leq|det(Id+tDG_k)|$ on P_h. Then by the area formula

$$\begin{aligned}\frac{1}{2}\mathcal{L}^d(P_h\cap B(x_0,\frac{r}{2}))\leq\int_{P_h\cap B(x_0,\frac{r}{2})}|det(Id+tDG_h)|dx\\ = \ \mathcal{L}^d(\{x-t[\nabla u(x)]^* \ : \ x\in P_h\cap B(x_0,\frac{r}{2})\}).\end{aligned}$$

Finally, since x_0 is a Lebesgue point for both F_h and $\Gamma^{-1}(B(y_0,s))$ it is also a Lebesgue point for P_h, so that

$$\lim_{r\to 0}\frac{\mathcal{L}^d(P_h\cap B(x_0,\frac{r}{2}))}{\mathcal{L}^d(B(x_0,r))} \ = \ \frac{1}{2^d}.$$

Summing up the previous observations, we get that

$$\theta_*(T(\Gamma\cap[(F_h\cap B(x_0,s))\times B(y_0,s)]),x_0) \ \geq \ \frac{1}{2^{d+1}}.$$

□

4.3. **The main existence result.** We are now in position to state and prove our main result, from which it directly follows that (1.2) has at least one solution, to which is associated the unique solution of (2.5). Its proof follows the lines of that of Theorem 6.1 in [9].

Theorem 4.5. *Assume that $\mu<<\mathcal{L}^d$, then problem (2.5) admits a unique solution γ which is induced by a transport map T_γ, i.e. $\gamma=(id\times T_\gamma)_\sharp\mu$.*

Proof. We first show that any solution γ of (2.5) is induced by a transport map. By Lemma 2.4 it is enough to prove that γ is concentrated on a graph.

Fix a Kantorovich potential u, and let $(F_h)_h$ be the sequence of sets given by Proposition 3.4. Since $\pi^1(\text{support}(\gamma))$ is included in $\mathcal{T}_u^e$ and $\mu<<\mathcal{L}^d$, we infer that γ is concentrated on

$$D(\gamma) := R(\Gamma) \bigcap \left[\bigcup_h \mathrm{Leb}(F_h) \times \Omega \right]$$

where Γ is the set given in Proposition 2.8 and $R(\Gamma)$ is given in Lemma 4.2. We show that $D(\gamma)$ is included in a graph, that is if (x_0, y_0) and (x_0, y_1) both belong to $D(\gamma)$ 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

$$(y_0 - y_1) \cdot (y_1 - x_0) < 0, \tag{4.1}$$

which in particular implies $y_1 \neq x_0$.

We fix $s > 0$ small enough so that

$$\forall x \in B(x_0, s), \forall y \in \overline{B(y_0, s)}, \forall y' \in \overline{B(y_1, s)}, \qquad (y - y') \cdot (y' - x) < 0. \tag{4.2}$$

By definition of $D(\gamma)$ and by Proposition 4.4 we know that

- x_0 is a Lebesgue point for the set $\Gamma^{-1}(B(y_0, s))$,
- x_0 is a Lebesgue point for the set $\Gamma^{-1}(B(y_1, s))$,
- the set $T[\Gamma \cap ([F_h \cap B(x_0, s)] \times B(y_1, s))]$ has positive lower density at x_0.

As a consequence, for r small enough there exist $\tilde{x} \in B(x_0, r)$ which belongs to the intersection of these three sets. In other words there exists $(\overline{x}, \overline{y}_1)$, $(\tilde{x}, \tilde{y}_0)$ and $(\tilde{x}, \tilde{y}_1)$ in Γ such that $\overline{x} \in F_h \cap B(x_0, s)$, $\overline{y}_1 \in B(y_1, s)$, $\tilde{x} \in]\overline{x}, \overline{y}_1[$, $\tilde{y}_0 \in B(y_0, s)$ 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 (2.6) applied to $(\overline{x}, \overline{y}_1)$ and $(\tilde{x}, \tilde{y}_0)$ that

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

Since $\tilde{x}$ belongs to $]\overline{x}, \overline{y}_1[$ we have $\tilde{x} - \overline{x} = \frac{|\tilde{x} - \overline{x}|}{|\overline{y}_1 - \overline{x}|}(\overline{y}_1 - \overline{x})$ and thus get contradiction with (4.2). Thus $D(\gamma)$ is included in a graph, so that γ is induced by a transport map.

Finally we prove the uniqueness part by a standard method (see Step 5 of the proof of Theorem B in [3]): if γ_1 and γ_2 are two solutions of (2.5), then $\frac{1}{2}(\gamma_1 + \gamma_2)$ is also a solution of this convex problem. It follows from the preceding that these three plans are all induced by transport maps, which must then coincide μ almost everywhere. $\square$

References

[1] Ahmad, N., Kim, H.K., McCann, R., *Extremal doubly stochastic measures and optimal transportation*, 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., Kirchheim, B., Pratelli, A., *Existence of optimal transport maps for crystalline norms*, Duke Math. J., **125** (2004), no. 2, 207–241.

[4] Ambrosio, L., Pratelli, A., 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.

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

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

[7] Caffarelli, L.A., Feldman, M., McCann, R.J., *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.

[8] Caravenna, L., *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.

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

[10] Champion, T., De Pascale, L., *The Monge problem in* $\mathbb{R}^d$, Duke Math. J. (to appear), temporarily available on http://cvgmt.sns.it .

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

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

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

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

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

[16] Pratelli, A., *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.

[17] Pratelli, A., *On the sufficiency of c-cyclical monotonicity for optimality of transport plans*, Math. Z., **258** (2008), no. 3, 677–690

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

[19] Sudakov, V. N., 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.

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

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

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

T.C. Institut de Mathématiques de Toulon et du Var, U.F.R. des Sciences et Techniques, Université du Sud Toulon-Var, Avenue de l'Université, BP 20132, 83957 La Garde cedex, FRANCE

L.D.P. Dipartimento di Matematica Applicata, Universitá di Pisa, Via Buonarroti 1/c, 56127 Pisa, ITALY