THE MONGE PROBLEM IN $\mathbb{R}^d$

THIERRY CHAMPION AND LUIGI DE PASCALE

Abstract. We first consider the Monge problem in a convex bounded subset of $\mathbb{R}^d$. The cost is given by a general norm, and we prove the existence of an optimal transport map under the classical assumption that the first marginal is absolutely continuous with respect to the Lebesgue measure. In the final part of the paper we show how to extend this existence result to a general open subset of $\mathbb{R}^d$.

1. Introduction

The Monge problem has origin in the *Mémoire sur la théorie des déblais et remblais* written by G. Monge [16], and may be stated as follows:

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

where Ω is the closure of a bounded convex open subset of $\mathbb{R}^d$, $|\cdot|$ denotes the usual Euclidean norm of $\mathbb{R}^d$, μ,ν are Borel probability measures on Ω and $\mathcal{T}(\mu,\nu)$ denotes the set of transport maps from μ to ν, i.e. the class of Borel maps T such that $T_\sharp\mu=\nu$ (where $T_\sharp\mu(B):=\mu(T^{-1}(B))$ for each Borel set B). A detailed account of the more relevant variants of problem (1.1) may be found in the recent books [21, 22]. In this paper we prove the following existence result for a generalization of the problem, where the Euclidean norm $|\cdot|$ is replaced by a general norm on $\mathbb{R}^d$.

Theorem 1.1. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$ and assume that μ is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^d$, then the problem*

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

has at least one solution.

We emphasize the fact that we make no regularity assumption on the norm $\|\cdot\|$. On the other hand, the assumption that the first marginal μ should be absolutely continuous with respect to the Lebesgue measure is classical and may be justified by Theorem 8.3 in Ambrosio *et al.* [3], which states that for any $s<d$ there exists a measure $\mu<<\mathcal{H}^s$ for which (1.1) does not have any solution.

The main difficulties in (1.2) are due to the facts that the objective functional is non-linear in T and the set $\mathcal{T}(\mu,\nu)$ does not possess the right compactness properties to apply the direct methods of the Calculus of Variations. A suitable relaxation was

Date: 25 May 2010.
2000 *Mathematics Subject Classification.* 49Q20, 49K30, 49J45.
Key words and phrases. Monge-Kantorovich problem, optimal transport problem, cyclical monotonicity.

introduced by Kantorovich [14, 15] and it proved to be a decisive tool to deal with this problem. Define the set of transport plans from μ to ν as

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

where $\mathcal{P}(\Omega \times \Omega)$ denotes the set of Borel probability measures on $\Omega \times \Omega$ and π^i denotes the standard projection in the Cartesian product. The set $\Pi(\mu,\nu)$ is always non-empty as it contains at least $\mu \otimes \nu$. Then Kantorovich proposed to study the problem

$$\min \left\{ \int_{\Omega\times\Omega} \|x-y\| d\gamma(x,y) \ : \ \gamma \in \Pi(\mu,\nu) \right\}. \tag{1.3}$$

Problem (1.3) is convex and linear in γ, then the existence of a minimizer may be obtained by the direct method of the Calculus of Variations. Moreover the minimal value of problem (1.3) is equal to the infimum of problem (1.2) in a very general setting (see for example [17]). To obtain the existence of a minimizer for (1.2) it is then sufficient to prove that some solution $\gamma \in \Pi(\mu,\nu)$ of (1.3) is in fact induced by a transport $T \in \mathcal{T}(\mu,\nu)$, i.e. may be written as $\gamma = (id \times T)_\sharp \mu$.

Before describing the present work, let us review briefly other existence results for (1.2). Sudakov [19] first proposed an efficient strategy to solve (1.2) for a general norm $\|\cdot\|$ on $\mathbb{R}^d$. However this method involved a crucial step on the disintegration of an optimal measure γ for (1.3) which was not completed correctly at that time, and has recently been justified in the case of a strictly convex norm by Caravenna [7]. Meanwhile, the problem (1.1) has been solved by Evans *et al.* [13] with the additional regularity assumption that μ and ν have Lipschitz-continuous densities with respect to $\mathcal{L}^d$, and then by Ambrosio [1] 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 by Caffarelli *et al.* [6] and Ambrosio *et al.* [3], and finally for crystalline norms in $\mathbb{R}^d$ and general norms in $\mathbb{R}^2$ by Ambrosio *et al.* [2]. 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 [8], we designed a different method which does not require the reduction to 1-dimensional settings. However, we were able to carry on one of the steps of our proof only in the case of strictly convex norms.

In this paper, we prove the existence of a solution to (1.2) for a general norm $\|\cdot\|$ on $\mathbb{R}^d$. The originality of our method for the proof of Theorem 1.1 above is that it does not require disintegration of measures and relies on a simple but powerful regularity result (see Lemma 3.3 below), which is inspired by a previous regularity result obtained in the study of an optimal transportation problem with cost functional in non-integral form in [9]. In section §2, we introduce a variational approximation to select solutions of (1.3) that have a particular monotonicity property. Section §3 is devoted to the notion of density-regular points of a transport γ and in particular to Lemma 3.3, which states that a transport map $\gamma \in \Pi(\mu,\nu)$ is concentrated on such points. In the following section §4, we infer from the preceding some technical regularity result for the particular solutions of (1.3) previously selected. The proof of our main result Theorem 1.1 is finally derived in §5, while some final comments are collected in §6. Finally, in section §7, we also shortly discuss the generalization of our approach to the unbounded setting.

2. Variational approximation to select monotone transport plans

Following the line of [2, 6, 18], we introduce a variational approximation to select optimal transport plans for (1.3) which have some additional properties, and in the next sections we shall prove that these particular optimal transport plans are induced by transport maps. This procedure of choosing particular minimizers is the root of the idea of asymptotic development by Γ-convergence (see [4] and [5]) .

We denote by $\mathcal{O}_1(\mu,\nu)$ the set of optimal transport plans for (1.3), and consider the auxiliary problem:

$$\min\left\{\int_{\Omega\times\Omega}|y-x|^2 d\gamma(x,y)\ :\ \gamma\in\mathcal{O}_1(\mu,\nu)\right\}, \tag{2.1}$$

where we remark the fact that the cost in consideration involves the Euclidean norm $|\cdot|$ of $\mathbb{R}^d$. Following §3.1 in [18], we introduce an approximating procedure for some particular solutions of (2.1) (see Lemma 2.3 below). Given two Borel probability measures α and β on Ω, we denote by

$$\mathcal{W}_1(\alpha,\beta) := \min\left\{\int_{\Omega\times\Omega}\|x-y\|d\gamma\ :\ \gamma\in\Pi(\alpha,\beta)\right\}$$

the usual $1-$Wasserstein distance associated to the norm $\|\cdot\|$. Notice that problem (1.3) then corresponds to $\mathcal{W}_1(\mu,\nu)$. For $\varepsilon>0$, we also set

$$C_\varepsilon(\gamma;\nu) := \frac{1}{\varepsilon}\mathcal{W}_1(\pi^2_\sharp\gamma,\nu)+\int_{\Omega\times\Omega}\|x-y\|d\gamma+\varepsilon\int_{\Omega\times\Omega}|x-y|^2d\gamma+\varepsilon^{3d+2}\mathrm{Card}(\pi^2_\sharp\gamma)$$

for any $\gamma\in\mathcal{P}(\Omega\times\Omega)$, where $\mathrm{Card}(\cdot)$ denotes the cardinality of the support of the measure. We emphasize the fact that the norm $\|\cdot\|$ appears in the two first terms of C_ε while the Euclidean norm $|\cdot|$ appears only in the third term. We then consider the following family of minimization problems $(D_\varepsilon)_{\varepsilon>0}$ associated to (1.3) and (2.1):

$$(D_\varepsilon)\qquad \min\{C_\varepsilon(\gamma;\nu)\ :\gamma\in\mathcal{P}(\Omega\times\Omega),\ \pi^1_\sharp\gamma=\mu\}.$$

For any $\varepsilon>0$ the problem (D_ε) admits at least one solution γ_ε, with discrete second marginal $\pi^2_\sharp\gamma_\varepsilon$.

We finally introduce the standard family of interpolated projections.

Definition 2.1. For $t\in[0,1]$ we will denote by P^t the map

$$\begin{array}{rrcl} P^t: & \Omega\times\Omega & \to & \Omega \\ & (x,y) & \mapsto & (1-t)x+ty. \end{array}$$

The following Proposition collects, for later use, some properties of the minimizers of (D_ε). This proposition is mainly inspired from [18] where some results of [10, 11, 12] are improved and simplified by the use of the above approximation process and time interpolant.

Proposition 2.2. *Let B be a Borel subset of $\Omega\times\Omega$. Let $\varepsilon>0$ and γ_ε be a solution for (D_ε), and set $\mu_{\varepsilon,B}:=\pi^1_\sharp\gamma_\varepsilon\lfloor B$ and $\nu_{\varepsilon,B}:=\pi^2_\sharp\gamma_\varepsilon\lfloor B$. Then it holds*

(1) *the measure $\gamma_\varepsilon \lfloor B$ is a solution of the problem*

$$(D_{\varepsilon,B}) \qquad \min\left\{ \int_{\Omega\times\Omega} (\|x-y\| + \varepsilon|x-y|^2) d\gamma \ : \gamma \in \Pi(\mu_{\varepsilon,B}, \nu_{\varepsilon,B}) \right\}$$

where $\Pi(\mu_{\varepsilon,B}, \nu_{\varepsilon,B})$ denotes the set of non-negative Borel measures with marginals $\mu_{\varepsilon,B}$ and $\nu_{\varepsilon,B}$;

(2) *if $\mu_{\varepsilon,B} \in L^\infty(\Omega)$ then for any $t \in (0,1)$ it holds*

$$\|{P^t}_\sharp(\gamma_\varepsilon\lfloor B)\|_{L^\infty} \leq (1-t)^{-d} \|\mu_{\varepsilon,B}\|_{L^\infty}.$$

Proof. Since γ_ε is a solution of (D_ε), it is a solution of

$$\min\left\{ \int_{\Omega\times\Omega} (\|x-y\| + \varepsilon|x-y|^2) d\gamma \ : \gamma \in \Pi(\mu, \pi^2_\sharp \gamma_\varepsilon) \right\}. \tag{2.2}$$

The claim (1) then follows from the linearity of the functional in problem (2.2) (e.g. see the proof of Lemma 4.2 in [2]).

The claim (2) is a direct application of Lemma 2 in §3.2 of [18], since by (1) the measure $\gamma_\varepsilon\lfloor B$ is an optimal transport plan between $\mu_{\varepsilon,B}$, which is absolutely continuous with respect to $\mathcal{L}^d$, and the discrete measure $\nu_{\varepsilon,B}$ for the strictly convex cost $(x,y) \mapsto \|x-y\| + \varepsilon|x-y|^2$ (see also the Appendix below). □

The link between the family of problems (D_ε) and (2.1) is given in the following lemma, whose proof coincides with that of Lemma 1 in §3.1 of [18] and will be given in the Appendix for sake of completeness.

Lemma 2.3. *For any $\varepsilon > 0$ let γ_ε be a solution of (D_ε), then the sequence $(\pi^2_\sharp\gamma_\varepsilon)$ w^*-converges to ν as $\varepsilon \to 0$. Moreover, any w^*-limit as $\varepsilon_k \to 0$ of a subsequence of solutions $(\gamma_{\varepsilon_k})_{k\in\mathbb{N}}$ is a solution of (2.1).*

The above Lemma suggests to introduce the following set of optimal transport plans for (1.3).

Definition 2.4. We shall denote by $\mathcal{O}_2(\mu,\nu)$ the minimizers for (2.1) which are w^*-limits as $\varepsilon_k \to 0$ of a subsequence $(\gamma_{\varepsilon_k})_{k\in\mathbb{N}}$ of minimizers of (D_{ε_k}).

We observe that, by definition, the minimizers γ_ε of problem (D_ε) are all probability measures on $\Omega\times\Omega$, and since their marginals converge as $\varepsilon \to 0$ to μ and ν, we infer that $\mathcal{O}_2(\mu,\nu)$ is not empty.

It is an important fact in the following that the local properties stated in Proposition 2.2 pass to the limit and are still valid for the elements of $\mathcal{O}_2(\mu,\nu)$. Notice that, in general, the restrictions of a sequence of weakly converging measures does not converge without additional assumptions. The following lemma states that this is the case when considering a sequence of transport plans with the same first marginals.

Lemma 2.5. *Let $(\gamma_\varepsilon)_\varepsilon$ a sequence in $\mathcal{P}(\Omega\times\Omega)$ with w^*-limit $\gamma \in \mathcal{P}(\Omega\times\Omega)$ as $\varepsilon \to 0$, and such that $\pi^1_\sharp\gamma_\varepsilon = \pi^1_\sharp\gamma = \mu$ for any $\varepsilon > 0$, with $\mu << \mathcal{L}^d$. Then for any Borel set $G \subset \Omega$ it holds $\gamma_\varepsilon\lfloor G\times\Omega \overset{*}{\rightharpoonup} \gamma\lfloor G\times\Omega$.*

Proof. We have to prove that for all $\varphi \in \mathcal{C}_b(\Omega \times \Omega)$ it holds

$$\int_{\Omega\times\Omega} \chi_G(x)\varphi(x,y)d\gamma_\varepsilon(x,y) \to \int_{\Omega\times\Omega} \chi_G(x)\varphi(x,y)d\gamma(x,y) \qquad as\ \varepsilon \to 0. \tag{2.3}$$

Fix $\varphi \in \mathcal{C}_b(\Omega \times \Omega)$. For all $\delta > 0$ there exists a closed set F_δ and an open set U_δ such that

$$F_\delta \subset G \subset U_\delta \ \text{ and } \ \mu(U_\delta \setminus F_\delta) \le \delta.$$

Since $\pi^1_\sharp \gamma_\varepsilon = \pi^1_\sharp \gamma = \mu$ we have

$$\gamma_\varepsilon((U_\delta \setminus F_\delta) \times \Omega) \le \delta \ \text{ and } \ \gamma((U_\delta \setminus F_\delta) \times \Omega) \le \delta.$$

Using the Uryson's lemma we can find $\psi_\delta \in C_b(\Omega \times \Omega)$ which coincides with φ on $F_\delta \times \Omega$, is 0 outside $U_\delta \times \Omega$ and $\|\psi_\delta\|_\infty \le \|\varphi\|_\infty$. It follows that

$$\left| \int_{\Omega\times\Omega} \chi_G(x)\varphi(x,y)d\gamma_\varepsilon(x,y) - \int_{\Omega\times\Omega} \psi_\delta(x,y)d\gamma_\varepsilon(x,y) \right| \le 2\|\varphi\|_\infty \delta,$$

and

$$\left| \int_{\Omega\times\Omega} \chi_G(x)\varphi(x,y)d\gamma(x,y) - \int_{\Omega\times\Omega} \psi_\delta(x,y)d\gamma(x,y) \right| \le 2\|\varphi\|_\infty \delta.$$

Since ψ_δ is continuous, passing to the $\limsup$ as $\varepsilon \to 0$ and then taking the limit for $\delta \to 0$ yields the conclusion. □

Finally, since an element of $\mathcal{O}_2(\mu,\nu)$ is a solution of (2.1), it enjoys a cyclical monotonicity property inherited from the cost $(x,y) \mapsto |y-x|^2$ (see remark 2.7 below), stated in the following proposition, whose proof may be derived from that of Lemma 4.1 in [2] and is given in [8] (see Proposition 3.2 therein).

Proposition 2.6. *Let γ be a solution of (2.1), then γ is concentrated on a σ-compact set Γ with the following property:*

$$\forall (x,y),(x',y') \in \Gamma, \qquad x \in [x',y'] \ \Rightarrow \ (x-x')\cdot(y-y') \ge 0, \tag{2.4}$$

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

Remark 2.7. A solution γ of the classical transport problem associated to $|\cdot|^2$:

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

is known to be concentrated on a $|\cdot|^2$-cyclically monotone set Γ, that is:

$$\forall (x,y),(x',y') \in \Gamma, \qquad (x-x')\cdot(y-y') \ge 0.$$

In (2.4), the restriction that x should be in $[x',y']$ to get the inequality has its origin in the fact that the constraint in (2.1) is $\mathcal{O}_1(\mu,\nu)$ in place of $\Pi(\mu,\nu)$.

Remark 2.8. The reason to deal with σ-compact sets Γ, in the above proposition as well as in the following, is that the projection $\pi^1(\Gamma)$ is also σ-compact, and in particular is a Borel set.

3. A property of transport plans

We begin by considering some general properties of transport plans. This section is independent of the transport problem (1.3), and some of the techniques detailed below are refinements of similar ones which were first applied in [9] in the framework of non-classical transportation problems involving cost functionals not in integral form.

Definition 3.1. Let Γ be a σ-compact set. For $y \in \Omega$ and $r > 0$ we define

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

If γ is a transport plan and Γ is a σ-compact set on which γ is concentrated, the set $\Gamma^{-1}(\overline{B(y,r)})$ contains the set of those points whose mass (with respect to μ) is partially or completely transported to $\overline{B(y,r)}$ by the restriction of γ to Γ. We may justify this slight abuse of notations by the fact that γ should be thought as a device that transports mass. Notice also that $\Gamma^{-1}(\overline{B(y,r)})$ is a σ-compact set.

Since this notion is important in the sequel, we recall that when a function f is *locally integrable* for the Lebesgue measure $\mathcal{L}^d$, one has

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^d(B(x,r))} \int_{B(x,r)} |f(z) - f(x)|dz \;=\; 0$$

for almost every x in Ω. These points x are usually called Lebesgue points of f. When A is an $\mathcal{L}^d$-measurable subset of Ω, we shall call Lebesgue point of A any element $x \in A$ which is a Lebesgue point of the characteristic function $f = \chi_A$ of A, and then satisfies

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

In the following, we shall denote by Leb(f) (resp. Leb(A)) the set of points $x \in \Omega$ (resp. $x \in A$) which are Lebesgue points of f (resp. A). Moreover we will denote by *support*(f) the smallest closed set out of which f equals 0 a.e.. For a non-negative function f we will use the equivalent characterization that $x \in$ *support*(f) if and only if $\int_{B(x,r)} f(z)dz > 0$ for any $r > 0$.

Definition 3.2. We will call density of an absolutely continuous measure λ the function

$$g(x) = \limsup_{r\to 0} \frac{\lambda(B(x,r))}{\mathcal{L}^d(B(x,r))}.$$

Then the set Leb(g) of the Lebesgue points of the density g of λ are uniquely determined as well as the value of g at those points.

The following Lemma is an essential step in the proof of Proposition 4.2 and Theorem 5.1 below. This result is a refinement of Lemma 5.2 from [9] and Lemma 4.3 in [8], and its proof follows the lines of those Lemmas. It in fact encompasses those results, as Lemma 3.5 below shows.

Lemma 3.3. *Assume that $\mu << \mathcal{L}^d$ with density denoted by f. Let $\gamma \in \Pi(\mu,\nu)$, and Γ a set on which γ is concentrated. Then there exists a σ-compact subset $D(\Gamma)$ of $\Gamma \cap$ support(γ) on which γ is concentrated, and such that for any $(x,y) \in D(\Gamma)$ and $r > 0$, there exist $\tilde{y} \in \Omega$ and $\tilde{r} > 0$ such that*

$$y \in B(\tilde{y}, \tilde{r}) \subset B(y,r), \quad x \in \mathrm{Leb}(f) \cap \mathrm{Leb}(\tilde{f}), \quad f(x) < +\infty \quad \textit{and} \quad \tilde{f}(x) > 0 \tag{3.1}$$

where $\tilde{f}$ *is the density of* $\pi^1_\sharp \gamma \lfloor \Omega \times B(\tilde{y}, \tilde{r})$ *with respect to* $\mathcal{L}^d$.

Proof. Let $(y_n)_n$ be a dense sequence in Ω. For each $(n,k) \in \mathbb{N}^2$ we set $\gamma_{n,k} := \gamma \lfloor \Omega \times B(y_n, \frac{1}{k+1})$ and define $f_{n,k}$ to be the density of $\pi^1_\sharp \gamma_{n,k}$ with respect to $\mathcal{L}^d$.

For all $n, k \in \mathbb{N}$ we set

$$C_{n,k} := [\Omega \setminus (\mathrm{Leb}(f) \cap \mathrm{Leb}(f_{n,k}) \cap \{f < +\infty\})] \times \Omega$$

and

$$D_{n,k} := [\Omega \setminus \{f_{n,k} > 0\}] \times B(y_n, \frac{1}{k+1}).$$

We claim that $\gamma[\cup_{n,k}(C_{n,k} \cup D_{n,k})] = 0$. Indeed fix $n, k \in \mathbb{N}$, then the set $\Omega \setminus (\mathrm{Leb}(f) \cap \mathrm{Leb}(f_{n,k}) \cap \{f < +\infty\})$ has $\mathcal{L}^d$ measure 0, so that it also has μ-measure 0 and then

$$\gamma(C_{n,k}) \;=\; \mu[\Omega \setminus (\mathrm{Leb}(f) \cap \mathrm{Leb}(f_{n,k}) \cap \{f < +\infty\})] \;=\; 0.$$

On the other hand, one has

$$\gamma(D_{n,k}) \;=\; \gamma_{n,k}([\Omega \setminus \{f_{n,k} > 0\}] \times \Omega) \;=\; \int_{\Omega \setminus \{f_{n,k} > 0\}} f_{n,k}(z) dz \;=\; 0.$$

As a consequence $\gamma[\cup_{n,k}(C_{n,k} \cup D_{n,k})] = 0$. Then γ is concentrated on the set $D = \mathit{support}(\gamma) \cap \Gamma \setminus [\cup_{n,k}(C_{n,k} \cup D_{n,k})]$. This set D has all the desired properties but the $\sigma-$compactness. Indeed, for every $(x,y) \in D$ and $r > 0$ there exists n, k such that $y \in B(y_n, \frac{1}{k+1}) \subset B(y,r)$ and (3.1) holds for the choice $(\tilde{y}, \tilde{r}) = (y_n, \frac{1}{k+1})$.

Since γ is a Borel probability measure, by inner regularity it is concentrated on a $\sigma-$compact set $D(\Gamma)$ included in D. □

The above discussion and Lemma yield us to introduce the following notions:

Definition 3.4. The couple $(x,y) \in \Gamma$ is a Γ-*regular point* if x is a Lebesgue point of $\Gamma^{-1}(\overline{B(y,r)})$ for any positive r. If $\gamma \in \Pi(\mu,\nu)$, the couple $(x,y) \in \Gamma$ is a (γ,Γ)-*density-regular point* if for any $r > 0$ there exists $(\tilde{y}, \tilde{r})$ such that (3.1) holds.

Lemma 3.3 above therefore states that if the transport plan γ is concentrated on a set Γ, then it is also concentrated on the $\sigma-$compact set $D(\Gamma)$ which consists of (γ,Γ)-density-regular points. The following Lemma shows that the elements of $D(\Gamma)$ are also Γ-regular points.

Lemma 3.5. *Assume that* $\mu << \mathcal{L}^d$*, let* $\gamma \in \Pi(\mu,\nu)$ *be concentrated on a set* Γ*, and* $D(\Gamma)$ *given by Lemma 3.3. Then any element of* $D(\Gamma)$ *is a* Γ*-regular point.*

Proof. Let $(x,y) \in D(\Gamma)$ and $r > 0$, we check that x is a Lebesgue point for $\Gamma^{-1}(\overline{B(y,r)})$. First there exists $(\tilde{y}, \tilde{r})$ such that (3.1) holds with the density $\tilde{f}$ of $\pi^1_\sharp \gamma \lfloor \Omega \times B(\tilde{y}, \tilde{r})$ with respect to $\mathcal{L}^d$. Since one has $\tilde{f}(x) > 0$ and $x \in \mathrm{Leb}(\tilde{f})$ it follows that x belongs to $\mathrm{Leb}(\{\tilde{f} > 0\})$. Moreover $B(\tilde{y}, \tilde{r}) \subset \overline{B(y,r)}$ so that

$$\left[\Omega \setminus \Gamma^{-1}(\overline{B(y,r)})\right] \times B(\tilde{y}, \tilde{r}) \;\subset\; \Omega^2 \setminus \Gamma. \tag{3.2}$$

Then by the definition of $\tilde{f}$,

$$\int_{\Omega\setminus\Gamma^{-1}(\overline{B(y,r)})} \tilde{f}(z)dz = \gamma\left(\left[\Omega \setminus \Gamma^{-1}(\overline{B(y,r)})\right] \times B(\tilde{y},\tilde{r})\right) = 0$$

where the last equality follows from (3.2) and the fact that γ is supported on Γ. Then the non-negative function $\tilde{f}$ is a.e. 0 on $\Omega\setminus\Gamma^{-1}(\overline{B(y,r)})$ so that $\mathcal{L}^d(\{\tilde{f} > 0\}\setminus\Gamma^{-1}(\overline{B(y,r)})) = 0$. Since $x \in \Gamma^{-1}(\overline{B(y,r)})$ is a Lebesgue point of $\{\tilde{f} > 0\}$, we conclude that x is a Lebesgue point of $\Gamma^{-1}(\overline{B(y,r)})$. □

4. A property of the selected optimal transport plans

In this section, we obtain a regularity result (Proposition 4.2 below) for the transport plans which belong to $\mathcal{O}_2(\mu,\nu)$ (see Definition 2.4). Following the formalism of [3] we introduce the notion of transport set related to a subset Γ of $\mathbb{R}^d \times \mathbb{R}^d$.

Definition 4.1. Let Γ be a subset of $\mathbb{R}^d \times \mathbb{R}^d$, the transport set $T(\Gamma)$ of Γ is

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

Notice that if Γ is σ-compact then $T(\Gamma)$ is also σ-compact.

Proposition 4.2. *Assume that $\mu << \mathcal{L}^d$ and let $\gamma \in \mathcal{O}_2(\mu,\nu)$ be concentrated on a σ-compact set Γ. Then for any $(x,y) \in D(\Gamma)$ (obtained by Lemma 3.3) with $x \neq y$ and for $r > 0$ it holds*

$$\liminf_{\delta\to 0^+} \frac{\mathcal{L}^d\left[T\left(\Gamma \cap \left[B(x,\frac{\delta}{2}) \times B(y,r)\right]\right) \cap B(x,\delta)\right]}{\mathcal{L}^d(B(x,\delta))} \ > \ 0. \tag{4.1}$$

Proof. We denote by f the density of μ. Consider $(x,y) \in D(\Gamma)$ with $x \neq y$ and $r > 0$. Let $\tilde{y}$ and $\tilde{r}$ be as in Lemma 3.3; we recall that $\pi^1_\sharp\gamma\lfloor\Omega \times B(\tilde{y},\tilde{r})$ is absolutely continuous with respect to $\mathcal{L}^d$, with density denoted by $\tilde{f}$, that $\tilde{f}(x) > 0$, that $x \in \mathrm{Leb}(\tilde{f})$ so that

$$\lim_{s\to 0} \frac{1}{\mathcal{L}^d(B(x,s))} \int_{B(x,s)} |\tilde{f}(z) - \tilde{f}(x)|dz \ = \ 0, \tag{4.2}$$

and that $f(x) < +\infty$ with $x \in \mathrm{Leb}(f)$ so that

$$\lim_{s\to 0} \frac{1}{\mathcal{L}^d(B(x,s))} \int_{B(x,s)} |f(z) - f(x)|dz \ = \ 0. \tag{4.3}$$

Let $G := \{z \in \Omega \mid \frac{1}{2}\tilde{f}(x) \le \tilde{f}(z) \text{ and } f(z) \le f(x) + 1\}$. Possibly subtracting a set of $\mathcal{L}^d$-measure 0 we may consider G a Borel set, and it follows from (4.2) and (4.3) that

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

Fix $\delta > 0$ so that

$$\frac{\delta}{|x-y| + r} < 1 \quad \text{and} \quad \forall s \in (0,\delta),\ \mathcal{L}^d(G \cap B(x,s)) \ge \frac{1}{2}\mathcal{L}^d(B(x,s)) \tag{4.4}$$

and fix $t \in (0, \frac{\delta}{2(|x-y|+r)})$. Then for every $z \in B(x,\frac{\delta}{2})$ and every $w \in B(y,r)$ it holds

$$(1-t)z + tw \in B(x,\delta). \tag{4.5}$$

For such a choice of δ define the subset $G_\delta := B(x, \frac{\delta}{2}) \cap G$ of G and notice that

$$\mathcal{L}^d(G_\delta) \geq \frac{1}{2}\mathcal{L}^d(B(x, \frac{\delta}{2})). \tag{4.6}$$

Let $A_\delta := G_\delta \times B(\tilde{y}, \tilde{r})$ and consider the measure $\gamma_{A_\delta} := \gamma \lfloor A_\delta$. We observe that $\pi^1_\sharp \gamma_{A_\delta}$ is absolutely continuous with respect to $\mathcal{L}^d$ and we denote by f_{A_δ} its density. Then, according to definition 3.2, for all $z \in \Omega$ the density f_{A_δ} satisfies

$$\begin{aligned} f_{A_\delta}(z) &= \limsup_{s\to 0} \frac{\pi^1_\sharp \gamma_{A_\delta}(B(z,s))}{\mathcal{L}^d(B(z,s))} = \limsup_{s\to 0} \frac{\gamma[(B(z,s)\cap G_\delta)\times B(\tilde{y},\tilde{r})]}{\mathcal{L}^d(B(z,s))} \\ &= \limsup_{s\to 0} \frac{1}{\mathcal{L}^d(B(z,s))} \int_{B(z,s)\cap G_\delta} \tilde{f}(w)dw \\ &\geq \frac{\tilde{f}(x)}{2} \limsup_{s\to 0} \frac{\mathcal{L}^d(B(z,s)\cap G_\delta)}{\mathcal{L}^d(B(z,s))}. \end{aligned}$$

For almost every z in G_δ (namely the Lebesgue points of G_δ), the limsup on the right hand side is equal to 1, so that

$$\frac{1}{2}\tilde{f}(x) \leq f_{A_\delta} \qquad \text{a.e. on } G_\delta. \tag{4.7}$$

It then follows from (4.5), (4.6) and (4.7) that

$$\begin{aligned} \frac{\tilde{f}(x)}{4}\mathcal{L}^d(B(x,\frac{\delta}{2})) \leq \frac{\tilde{f}(x)}{2}\mathcal{L}^d(G_\delta) \leq \pi^1_\sharp\gamma_{A_\delta}(G_\delta) \\ \leq \pi^1_\sharp\gamma_{A_\delta}(B(x,\frac{\delta}{2})) \leq P^t_\sharp\gamma_{A_\delta}(B(x,\delta)). \end{aligned} \tag{4.8}$$

Since γ belongs to $\mathcal{O}_2(\mu,\nu)$, it is a w^*-limit of a subsequence $(\gamma_{\varepsilon_k})_k$ of minimizers of (D_{ε_k}). We notice that claim (2) of Proposition 2.2 holds for $\gamma_{\varepsilon_k}\lfloor G_\delta \times \Omega$, so that

$$\|P^t_\sharp \gamma_{\varepsilon_k}\lfloor G_\delta\times\Omega\|_{L^\infty} \leq (1-t)^{-d}\|\pi^1_\sharp\gamma_{\varepsilon_k}\lfloor G_\delta\times\Omega\|_\infty$$

and then

$$\|P^t_\sharp \gamma_{\varepsilon_k}\lfloor G_\delta\times\Omega\|_{L^\infty} \leq (1-t)^{-d}\|f\lfloor G_\delta\|_\infty \leq 2^d(f(x)+1). \tag{4.9}$$

By Lemma 2.5 it follows that $\gamma\lfloor G_\delta\times\Omega$ is the w^*-limit of the subsequence $(\gamma_{\varepsilon_k}\lfloor G_\delta\times\Omega)_k$. The sequence $(P^t_\sharp\gamma_{\varepsilon_k}\lfloor G_\delta\times\Omega)_k$ then converges weakly in $L^\infty(\Omega)$ to $P^t_\sharp\gamma\lfloor G_\delta\times\Omega$, and in particular letting $k\to+\infty$ in the above estimate yields

$$\|P^t_\sharp\gamma_{A_\delta}\|_{L^\infty} \leq \|P^t_\sharp\gamma\lfloor G_\delta\times\Omega\|_{L^\infty} \leq 2^d(f(x)+1). \tag{4.10}$$

On the other hand we claim that whenever a measure $\lambda\in\mathcal{M}(\Omega\times\Omega)$ is concentrated on a Borel set Λ the measure $P^t_\sharp\lambda$ is concentrated on $T(\Lambda)$. Indeed

$$P^t_\sharp\lambda(\Omega\setminus T(\Lambda)) = \lambda((P^t)^{-1}(\Omega\setminus T(\Lambda))) \leq \lambda(\Omega^2\setminus\Lambda) = 0.$$

As a consequence $P^t_\sharp\gamma_{A_\delta}$ is concentrated on $T(\Gamma\cap\big[B(x,\frac{\delta}{2})\times B(y,r)\big])$.

Then again the choice of t and (4.10) imply that

$$P^t_\sharp \gamma_{A_\delta}(B(x,\delta)) = P^t_\sharp \gamma_{A_\delta}\left(T(\Gamma\cap[B(x,\frac{\delta}{2})\times B(y,r)])\cap B(x,\delta)\right)$$
$$\leq 2^d(f(x)+1)\mathcal{L}^d\left(T(\Gamma\cap[B(x,\frac{\delta}{2})\times B(y,r)])\cap B(x,\delta)\right). \quad (4.11)$$

The proof is now complete since (4.8) and (4.11) yield

$$\mathcal{L}^d(T(\Gamma\cap[B(x,\frac{\delta}{2})\times B(y,r)])\cap B(x,\delta)) \geq \frac{\tilde{f}(x)}{2^{d+2}(f(x)+1)}\mathcal{L}^d(B(x,\frac{\delta}{2}))$$

for any $\delta > 0$ small enough for (4.4) to hold. □

5. Proof of the main theorem

We now conclude with the proof of Theorem 1.1, which is a consequence of the following result.

Theorem 5.1. *Assume that $\mu << \mathcal{L}^d$. Then $\mathcal{O}_2(\mu,\nu)$ has a unique element γ which is induced by a transport map $T_\gamma \in \mathcal{T}(\mu,\nu)$, i.e. $\gamma = (id\times T_\gamma)_\sharp\mu$.*

Proof. We first prove that if $\gamma \in \mathcal{O}_2(\mu,\nu)$ then it is induced by a transport map $T_\gamma \in \mathcal{T}(\mu,\nu)$. By Proposition 2.1 in [1], it is sufficient to prove that γ is concentrated on a Borel graph.

It follows from Proposition 2.6 that γ is concentrated on a σ-compact set Γ satisfying (2.4). We then apply Proposition 4.2 to get that γ is concentrated on a σ-compact subset $D(\Gamma)$ of $\Gamma\cap \mathit{support}(\gamma)$ and on which (4.1) is satisfied.

We claim that $D(\Gamma)$ is contained in a graph. To prove this, we show that if (x_0,y_0) and (x_0,y_1) both belong to $D(\Gamma)$ then $y_0 = y_1$. We argue by contradiction, and assume that $y_1 \neq y_0$. As a consequence, 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

$$(y_1-y_0)\cdot(y_0-x_0) \ < \ 0. \quad (5.1)$$

We fix $r > 0$ small enough so that

$$\forall x\in B(x_0,r),\ \forall y'\in\overline{B(y_0,r)},\ \forall y\in\overline{B(y_1,r)}, \qquad (y-y')\cdot(y'-x) \ < \ 0. \quad (5.2)$$

Since $(x_0,y_1)\in D(\Gamma)$, we infer from Lemma 3.5 that x_0 is a Lebesgue point for $\Gamma^{-1}(\overline{B(y_1,r)})$, so that

$$\lim_{\delta\to 0^+}\frac{\mathcal{L}^d\left(\pi^1[\Gamma\cap(\Omega\times\overline{B(y_1,r)})]\cap B(x_0,\delta)\right)}{\mathcal{L}^d(B(x_0,\delta))} \ = \ 1.$$

Moreover $x_0\neq y_0$ by (5.1), then we get from $(x_0,y_0)\in D(\Gamma)$ and (4.1) that

$$\liminf_{\delta\to 0^+}\frac{\mathcal{L}^d\left(T\left(\Gamma\cap\left[B(x_0,\frac{\delta}{2})\times B(y_0,r)\right]\right)\cap B(x_0,\delta)\right)}{\mathcal{L}^d(B(x_0,\delta))} \ > \ 0.$$

As a consequence, for δ small enough there exist (x',y') and (x,y) in Γ such that

$$x'\in B(x_0,\frac{\delta}{2}),\quad y'\in\overline{B(y_0,r)},\quad x\in[x',y']\cap B(x_0,\delta)\quad \mathit{and}\quad y\in\overline{B(y_1,r)}.$$

It follows from (2.4) applied to (x', y') and (x, y) that

$$(y - y') \cdot (x - x') \geq 0$$

but since $x \in [x', y']$ one also has $x - x' = \frac{|x-x'|}{|y'-x|}(y' - x)$ which contradicts (5.2). This concludes the proof for the fact that any element $\gamma \in \mathcal{O}_2(\mu, \nu)$ is induced by a transport map $T_\gamma \in \mathcal{T}(\mu, \nu)$.

We now prove that $\mathcal{O}_2(\mu, \nu)$ has a unique element, using the same type of uniqueness argument as in the Step 5 of the proof of Theorem B in [2].

Let γ_1 and γ_2 be two elements of $\mathcal{O}_2(\mu, \nu)$, then by the preceding there exist $T_1, T_2 \in \mathcal{T}(\mu, \nu)$ such that $\gamma_i = (id \times T_i)_\sharp \mu$ for $i = 1, 2$. Now consider $\gamma := \frac{1}{2}(\gamma_1 + \gamma_2)$; we claim that the above arguments also apply to γ, so that γ is also induced by a transport map $T_\gamma \in \mathcal{T}(\mu, \nu)$. On the one hand, γ belongs to the convex set $\mathcal{O}_1(\mu, \nu)$ so that γ satisfies the hypotheses of Proposition 2.6. On the other hand, the result of Proposition 4.2 also holds for γ: in the proof of that result, the hypothesis that γ should belong to $\mathcal{O}_2(\mu, \nu)$ is only used to derive the inequality (4.10). In our case, γ is the $w*$-limit of a sequence of the form $(\frac{1}{2}(\gamma_{1,\varepsilon_k} + \gamma_{2,\eta_k}))_k$ for which (4.9) holds, and then (4.10) is also valid. As a consequence, γ is induced by a transport map $T_\gamma \in \mathcal{T}(\mu, \nu)$. This implies that $T_1 = T_2 = T_\gamma$ holds $\mu-$a.e., so that in fact $\gamma_1 = \gamma_2$. □

Remark 5.2. It follows the uniqueness part of Theorem 5.1 above that the set $\mathcal{O}_2(\mu, \nu)$ is reduced to a single element. A natural question is whether the set of minimizers of problem (2.1) itself is a singleton. This is indeed true in the case of a strictly convex norm (see [8]) so in that case one has that $\mathcal{O}_2(\mu, \nu)$ coincides with the set of minimizers of(2.1). In the general case considered in this paper this question is still open.

6. Comments

The strategy for proving Theorem 5.1 above relies on two fundamental ingredients: the cyclical-monotonicity for particular solutions of (1.3) obtained in Proposition 2.6, and the density result for the set of transport rays obtained in Proposition 4.2. This strategy was already that developed in [8] for the special case of a strictly convex norm.

The originality in the use of Proposition 2.6 is that, since the norm $\|\cdot\|$ is not assumed to be strictly convex, it may happen that the points x, x', y, y' in consideration are not aligned. In the strictly convex case this property of alignment is fundamental since it basically allows to reduce the problem (1.3) to a family of one-dimensional problems, on which one can use the property of monotonicity of the selected optimal transport plan (solution of (2.1)). In the general - not necessarily strictly convex - case, we need to use the full information that the selected particular solution is concentrated on a set which is cyclically monotone in the classical sense of convex analysis.

The fact that the result stated in Proposition 4.2, although quite natural, happens to be somewhat difficult to obtain (and in particular was not derived in its full generality in Proposition 5.2 of [8]), may be illustrated by the following example. Let us first recall the following result in [2]:

Theorem 6.1 (Theorem A of [2]). *There exist a Borel set $M \subset [-1, 1]^3$ with $\mathcal{L}^3(M) = 8$ and two Borel maps $f_i : M \to [-2, 2] \times [-2, 2]$ for $i = 1, 2$ such that the following holds. For $x \in M$ denote by l_x the segment connecting $(f_1(x), -2)$ to $(f_2(x), 2)$ then*

(1) $\{x\} = l_x \cap M$ *for all* $x \in M$,
(2) $l_x \cap l_y = \emptyset$ *for all* $x, y \in M$ *different.*

If one considers $\Gamma := \{(x, F(x)) \,:\, x \in M\}$ with $F(x) := (f_2(x), 2)$, then we observe that the transport set $T(\Gamma)$ has density 0 at every point of $\pi^1(\Gamma) = M$ (although M has full measure in $[-1,1]^3$). We notice that Γ supports the transport plan $(id \times F)_\sharp(\mathcal{L}^3 \lfloor M)$ which is an optimal transport plan between its marginals for the cost $\|x\| := \max\{|x_1|, |x_2|, 3|x_3|\}$. The Lemma 3.3 (and the notion of Γ-density-regular points) as well as the approximating procedure provided in [18] (and recalled in §2) then appear as the necessary cornerstones to derive Proposition 4.2. In fact, it had been noticed in section §7 of [8] that using some estimate for the so called "transport density" may allow to obtain some technical result analogous to Proposition 4.2. Although this is not exactly what we did in the present paper, the inequality (4.8) in the proof of Proposition 4.2 contains that type of estimate.

We now discuss further possible extensions of the methodology developed here to prove Theorem 5.1. The above example first indicates that for some very bad cases, the open transport set $T(\Gamma)$ may have density 0 at any point of $\pi^1(\Gamma)$ when the norm is not strictly convex. This may be a limit of the definition of the open transport set that we use: a possible alternative would be to consider the set of all geodesics joining two points instead of considering only the segments. This would give a "fat" transport set. For the moment, our approach cannot be extended to this kind of transport sets without some substantial addition. We also notice that the construction we make in this paper does not make explicit use of the geometry of the segments, but it is based on some property of segments which may be enjoyed by more general family of curves. Then we believe that there is a possibility that the same approach could contribute to the proof of existence of optimal transports also in other geometric settings where this result is currently out of sight.

We conclude with a final remark for the readers with a broader knowledge of the literature on the Monge problem. The strategy presented in this paper provides a very efficient way to recover the existence result for an optimal transport map for the classical case of the Euclidean norm (or a C^2 strictly convex norm). Indeed, in that case the approximating procedure of §2 is useless and Proposition 4.2 holds for any solution γ of (1.3) by the following direct arguments: the transport rays do not cross, and if u is a potential for (1.3) (*i.e.* a solution of the classical dual problem for (1.3)) then there exists a countable union of sets $\cup_i T_i$ on which μ is concentrated and such that the gradient ∇u is Lipschitz-continuous on each T_i (for instance see [1, 6, 20]).

7. The unbounded case

Most of the arguments of the previous sections are of local nature, and may be carried out in the case of an unbounded set Ω. However the assumption that the set Ω is bounded plays a role in two crucial points of the methodology of the present work.

The first point is the choice of the secondary variational problem (2.1):

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

In the unbounded case, in order to write this problem one would be led to assume that

$$\int_\Omega |x|^2 d\mu + \int_\Omega |y|^2 d\nu < \infty$$

which is not a natural assumption in the setting of problem (1.2). To avoid this, one may for example replace (2.1) by

$$\min\left\{\int_{\Omega\times\Omega} \sqrt{1+|y-x|^2}\, d\gamma(x,y) \;:\; \gamma\in\mathcal{O}_1(\mu,\nu)\right\}. \tag{7.1}$$

Since for any $\gamma\in\Pi(\mu,\nu)$ one has

$$\int_{\Omega\times\Omega} \sqrt{1+|y-x|^2}\, d\gamma(x,y) \le 1 + \int_{\Omega\times\Omega} |y-x|\, d\gamma(x,y)$$

the value of (2.1) is finite whenever that of (1.2) is finite. Then a natural assumption on the measures μ and ν for (1.2) and (7.1) to have finite values is:

$$\int_\Omega |x|\, d\mu(x) + \int_\Omega |y|\, d\nu(y) < \infty. \tag{7.2}$$

In fact condition (7.2) amounts to the fact that the trivial transport $\mu\otimes\nu$ has finite cost, so that any admissible transport plan has finite cost. The cost $c:\xi\mapsto\sqrt{1+|\xi|^2}$ being smooth and convex, an analogue of Proposition 2.6 holds, where (2.4) is replaced by

$$\forall (x,y),(x',y')\in\Gamma, \qquad x\in[x',y'] \;\Rightarrow\; (x-x')\cdot\big(\nabla c(y-x')-\nabla c(y'-x)\big)\ge 0. \tag{7.3}$$

Remark 7.1. Notice that for the cost $\xi\mapsto|\xi|^2$ the monotonicity property (2.4) is much simpler than (7.3) since it is obtained by development of the scalar product rather than via a convexity argument.

By strict convexity of c, it also comes that if $y_0\neq y_1$ then

$$((y_1-x_0)-(y_0-x_0))\cdot(\nabla c(y_1-x_0)-\nabla c(y_0-x_0)) \;>\; 0,$$

so that one may replace (5.1) in the proof of Theorem 5.1 by the following assumption:

$$(\nabla c(y_1-x_0)-\nabla c(y_0-x_0))\cdot(y_0-x_0)<0.$$

The contradiction in the proof of Theorem 5.1 then follows with analogue arguments.

Of course the previous discussion also requires to change the approximation procedure described in Section 2, and that is the second point where the boundedness of Ω is explicitly used in our proof. Indeed, one should replace the functional C_ε by the following:

$$C'_\varepsilon(\gamma;\nu) := \frac{1}{\varepsilon}\mathcal{W}_1(\pi^2_\sharp\gamma,\nu) + \int_{\Omega\times\Omega}\left[\|x-y\| + \varepsilon\sqrt{1+|x-y|^2}\right] d\gamma + \theta(\varepsilon)\mathrm{Card}(\pi^2_\sharp\gamma)$$

for some function $\theta:\mathbb{R}^*_+\to\mathbb{R}^*_+$. Then the arguments in the proof of Proposition 2.2 stay valid, but the proof of Lemma 2.3 given in the appendix has to be adapted. In particular, it is no more possible to explicitly propose an approximation of ν by discrete measures $(p_{n\sharp}\nu)_{n\in\mathbb{N}}$, but such an approximation exists thanks to hypothesis (7.2). Then the function θ has to be chosen so that to have

$$\frac{1}{\varepsilon^2}\mathcal{W}_1(p_{n(\varepsilon)\sharp}\nu,\nu) + \frac{\theta(\varepsilon)}{\varepsilon}\,\mathrm{Card}(p_{n(\varepsilon)\sharp}\nu) \;\to\; 0$$

for some sequence $n(\varepsilon)$ converging to $+\infty$ as $\varepsilon \to 0$ (this is needed in the last step of the proof of Lemma 2.3).

Appendix

For the sake of completeness, we give some details of the arguments of the proofs of Proposition 2.2 as well as Lemma 2.3. These proofs are adapted from that of Theorem 1 and Lemmas 1 and 2 of [18].

Proof of Proposition 2.2 (2). Fix $\varepsilon > 0$ and $t > 0$. Let $\{y_i\}_{i\in I}$ be the finite support of $\nu_{\varepsilon,B}$. For $i \in I$ we set $\Omega_i := support(\gamma_\varepsilon \lfloor \Omega \times \{y_i\})$ and $\Omega_i(t) := P_t(\Omega_i \times \{y_i\})$. Then if A is a Borel subset of Ω we have

$$\begin{aligned} P^t_\sharp(\gamma_\varepsilon \lfloor B)(A) &\leq \sum_{i\in I} (\gamma_\varepsilon \lfloor B)((P^t)^{-1}(A \cap \Omega_i(t))) \\ &= \sum_{i\in I} \mu_{\varepsilon,B} \left(\frac{1}{1-t}(A \cap \Omega_i(t) - t\, y_i) \right) \\ &\leq \sum_{i\in I} (1-t)^{-d} \|\mu_{\varepsilon,B}\|_{L^\infty} \mathcal{L}^d(A \cap \Omega_i(t)). \end{aligned}$$

The conclusion then follows whenever

$$\sum_{i\in I} \mathcal{L}^d(A \cap \Omega_i(t)) \;=\; \mathcal{L}^d\left(\bigcup_{i\in I} A \cap \Omega_i(t) \right) \quad (\leq\; \mathcal{L}^d(A)).$$

This equality indeed follows from the fact that the sets $\Omega_i(t)$ and $\Omega_j(t)$ are disjoint when $i \neq j$. We prove this by contradiction, and assume that $(1-t)x_i + ty_i = (1-t)x_j + ty_j$ for some $x_i \in \Omega_i$, $x_j \in \Omega_j$ with $i \neq j$. Notice that since $y_i \neq y_j$, one also has $y_i - x_i \neq y_j - x_j$. The cost $c : (x, y) \mapsto \|x-y\| + \varepsilon|x-y|^2$ being continuous, the support of γ_ε is a c-cyclically monotone set, and thus one has

$$c(y_i - x_i) + c(y_j - x_j) \;\leq\; c(y_j - x_i) + c(y_i - x_j).$$

Since $y_j - x_i = t(y_i - x_i) + (1-t)(y_j - x_j)$ and $y_i - x_j = (1-t)(y_i - x_i) + t(y_j - x_j)$, we conclude from the strict convexity of c that

$$c(y_j - x_i) + c(y_i - x_j) \;<\; c(y_i - x_i) + c(y_j - x_j)$$

which is a contradiction. $\square$

Proof of Lemma 2.3. Since Ω is bounded, we may assume that $\Omega \subset B(0, R)$. For $n \geq 1$ let p_n be a measurable map from Ω to a grid of at most $(2Rn)^d$ points with the property that $|p_n(x) - x| \leq \frac{1}{n}$ for any $x \in \Omega$. Let γ be a solution of (2.1), for every $n \geq 1$ we set $\gamma^n := (id \times p_n)_\sharp \gamma$.

We now write the optimality of γ_ε for (D_ε) so that for any $\varepsilon > 0$ and $n \geq 1$ it holds

$$\begin{aligned} C_\varepsilon(\gamma_\varepsilon;\nu) &= \frac{1}{\varepsilon}\mathcal{W}_1(\pi^2_\sharp\gamma_\varepsilon,\nu) + \int_{\Omega\times\Omega} \|x-y\| d\gamma_\varepsilon + \varepsilon\int_{\Omega\times\Omega}|x-y|^2 d\gamma_\varepsilon + \varepsilon^{3d+2}\mathrm{Card}(\pi^2_\sharp\gamma_\varepsilon) \\ &\leq C_\varepsilon(\gamma^n;\nu) \\ &= \frac{1}{\varepsilon}\mathcal{W}_1(p_{n\sharp}\nu,\nu) + \int_{\Omega\times\Omega}\|x-y\|d\gamma^n + \varepsilon\int_{\Omega\times\Omega}|x-y|^2 d\gamma^n + \varepsilon^{3d+2}\mathrm{Card}(p_{n\sharp}\nu) \\ &\leq \frac{1}{n\,\varepsilon} + \int_{\Omega\times\Omega}\|x-y\|d\gamma^n + \varepsilon\int_{\Omega\times\Omega}|x-y|^2 d\gamma^n + \varepsilon^{3d+2}(2Rn)^d. \end{aligned}$$

Keeping the first term in $C_\varepsilon(\gamma_\varepsilon;\nu)$, multiplying by ε and letting $\varepsilon\to 0$ then yields

$$\forall n\geq 1,\quad \limsup_{\varepsilon\to 0}\mathcal{W}_1(\pi^2_\sharp\gamma_\varepsilon,\nu)\leq\frac{1}{n}.$$

Letting $n\to+\infty$ we get the w^*-convergence of $\pi^2_\sharp\gamma_\varepsilon$ to ν. As a consequence, any w^*-cluster point of $(\gamma_\varepsilon)_\varepsilon$ as $\varepsilon\to 0$ belongs to $\Pi(\mu,\nu)$.

Keeping the second term in $C_\varepsilon(\gamma_\varepsilon,\nu)$ and taking $n(\varepsilon)\approx\varepsilon^{-2}$ yields

$$\int_{\Omega\times\Omega}\|x-y\|d\gamma_\varepsilon \leq \varepsilon + \int_{\Omega\times\Omega}\|x-y\|d\gamma^{n(\varepsilon)} + \varepsilon\int_{\Omega\times\Omega}|x-y|^2 d\gamma^{n(\varepsilon)} + \varepsilon^{d+2}(2R)^d.$$

We let $\varepsilon\to 0$ and notice that

$$\int_{\Omega\times\Omega}\|x-y\|d\gamma^{n(\varepsilon)}\to\int_{\Omega\times\Omega}\|x-y\|d\gamma = \mathcal{W}_1(\mu,\nu),$$

so that any w^*-cluster point of $(\gamma_\varepsilon)_\varepsilon$ is a solution of (1.3).

We now notice that

$$\int_{\Omega\times\Omega}\|x-y\|d\gamma_\varepsilon \geq \mathcal{W}_1(\mu,\pi^2_\sharp\gamma_\varepsilon)\geq\mathcal{W}_1(\mu,\nu)-\mathcal{W}_1(\nu,\pi^2_\sharp\gamma_\varepsilon)$$

and

$$\int_{\Omega\times\Omega}\|x-y\|d\gamma^n \leq \int_{\Omega\times\Omega}\|x-y\|d\gamma + \int_{\Omega\times\Omega}\|p_n(y)-y\|d\gamma^n \leq \mathcal{W}_1(\mu,\nu)+\frac{1}{n}$$

where we used the optimality of γ for (1.3). Keeping the three first terms in $C_\varepsilon(\gamma_\varepsilon,\nu)$, we then obtain that

$$(\frac{1}{\varepsilon}-1)\mathcal{W}_1(\nu,\pi^2_\sharp\gamma_\varepsilon)+\varepsilon\int_{\Omega\times\Omega}|x-y|^2 d\gamma_\varepsilon \leq \frac{1+\varepsilon}{n\,\varepsilon}+\varepsilon\int_{\Omega\times\Omega}|x-y|^2 d\gamma^n+\varepsilon^{3d+2}(2Rn)^d.$$

The first term on the right hand side is non-negative for ε small enough, then dividing by ε and taking $n(\varepsilon)\approx\varepsilon^{-3}$ yield

$$\int_{\Omega\times\Omega}|x-y|^2 d\gamma_\varepsilon \leq (1+\varepsilon)\varepsilon + \int_{\Omega\times\Omega}|x-y|^2 d\gamma^{n(\varepsilon)} + \varepsilon(2R)^d.$$

so that any w^*-cluster point of $(\gamma_\varepsilon)_\varepsilon$ is a solution of (2.1). □

Acknowledgments

The research of the first author was partially supported by the Centro de Modelamiento Matemático (UMI CNRS 2807, Universidad de Chile).

The research of the second author was partially supported by the project *"Metodi variazionali nella teoria del trasporto ottimo di massa e nella teoria geometrica della misura"* of the program PRIN 2006 of the Italian Ministry of the University and by the *"Fondi di ricerca di ateneo"* of the University of Pisa.

Visits of both authors have been partially supported by the universities of Pisa and Toulon.

Both authors are currently supported by the Project GALILEO 2009 "Optimal transport problem and their interactions with geometry and Calculus of Variations" of the Università Italo-Francese-EGIDE.

We thank Pierre Cardaliaguet and Chloé Jimenez for usefull comments and discussions.

This paper was strongly improved by the, often detailed, criticism and suggestions of Luigi Ambrosio, Severine Rigot and the anonymous referees.

References

[1] 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.

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

[3] 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.

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

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

[6] 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.

[7] Caravenna, L.. *A partial proof of Sudakov theorem via disintegration of measures* Preprint SISSA (2008) available at http://cvgmt.sns.it .

[8] Champion, T. De Pascale *The Monge problem for strictly convex norms in* $\mathbb{R}^d$, J. Eur. Math. Soc. (to appear), temporarily available on http://cvgmt.sns.it .

[9] 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.

[10] De Pascale,L., Evans, L.C., Pratelli, A., *Integral Estimates for Transport Densities*, Bulletin of the London Mathematical Society, **36** (2004), no.3, 383-396.

[11] De Pascale, L., Pratelli, A., *Regularity properties for Monge transport density and for solutions of some shape optimization problem*, Calc. Var. Partial Differ. Equ., **14** (2002), no. 3, 249–274.

[12] De Pascale, L., Pratelli, A. *Interpolation and sharp summability for Monge Transport density*, ESAIM Control, Optimization and Calculus of Variations, **10** (2004), no. 4, 549-552.

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

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

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

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

[17] 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.

[18] Santambrogio, F., *Absolute continuity and summability of optimal transport densities: simpler proofs and new estimates* Calc. Var. Partial Differential Equations (to appear), temporarily available on http://cvgmt.sns.it .

[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