EXISTENCE OF OPTIMAL TRANSPORT MAPS FOR CRYSTALLINE NORMS

L. AMBROSIO, B. KIRCHHEIM, AND A. PRATELLI

ABSTRACT. We show the existence of optimal transport maps in the case when the cost function is the distance induced by a crystalline norm in $\mathbb{R}^n$, assuming that the initial distribution of mass is absolutely continuous with respect to $\mathscr{L}^n$. The proof is based on a careful decomposition of the space in transport rays induced by a secondary variational problem having as cost function the Euclidean distance. Moreover, improving a construction by Larman, we show the existence of a Nikodym set in $\mathbb{R}^3$ having full measure in the unit cube, intersecting each element of a family of pairwise disjoint open lines only in one point. This example can be used to show that the regularity of the decomposition in transport rays plays an essential role in Sudakov-type arguments for proving the existence of optimal transport maps.

1. INTRODUCTION

In 1781, G.Monge raised the problem of transporting a given distribution of matter (a pile of sand for instance) into another (an excavation for instance) in such a way that the work done is minimal. His problem, rephrased and generalized in modern mathematical terms, consists in minimizing the work of transportation in the class of transports, i.e.

$$\min\left\{\int_{\mathbb{R}^n} c(x,t(x))\,d\mu(x):\ t \text{ transport}\right\}. \tag{1.1}$$

Here μ, ν are given probability measures in $\mathbb{R}^n$ and the class of transports is given by all Borel maps $t:\mathbb{R}^n\to\mathbb{R}^n$ such that $t_\#\mu=\nu$ (i.e. $\mu(t^{-1}(E))=\nu(E)$ for any Borel set E); the cost function $c(x,y)$ represents the work done in shipping a unit mass from x to y, and was the Euclidean distance $|y-x|$ in the original Monge problem. The transport problem has by now an impressive number of applications, covering Non-linear PDE's, Calculus of Variations, Probability, Economics, Statistical Mechanics and many other fields. We refer to the surveys/books [1], [16], [26], [32], [33] for more informations on this wide topic.

Even in Euclidean spaces, the problem of existence of optimal transport maps is far from being trivial, mainly due to the non-linearity with respect to t of the condition $t_\#\mu=\nu$. In particular the class of transports is not closed with respect to any reasonable weak topology. Furthermore, it is easy to build examples where the Monge problem is ill-posed simply because there is no transport map: this happens for instance when μ is a Dirac mass and ν is not a Dirac mass.

In order to overcome these difficulties, in 1942 L.V.Kantorovich proposed in [21], [22] a notion of weak solution of the transport problem. He suggested to look for *transport plans* instead of transports, that is probability measures γ in $\mathbb{R}^n\times\mathbb{R}^n$ whose marginals are μ and ν. Formally this means that $\pi_{1\#}\gamma=\mu$ and $\pi_{2\#}\gamma=\nu$, where π_1, π_2 are the canonical projections. Denoting by $\Pi(\mu,\nu)$ the class of plannings, he formulated the following minimization problem

$$\min\left\{\iint_{\mathbb{R}^n\times\mathbb{R}^n} c(x,y)\,d\gamma:\ \gamma\in\Pi(\mu,\nu)\right\}. \tag{1.2}$$

Notice that $\Pi(\mu,\nu)$ is not empty, as the product $\mu\otimes\nu$ has μ and ν as marginals. Due to the linearity of the new constraint $\gamma\in\Pi(\mu,\nu)$ it turns out that weak topologies can be effectively used to provide existence of solutions to (1.2): this happens for instance whenever c is lower semicontinuous (e.g., see [26]).

The connection between the Kantorovich formulation of the transport problem and Monge's original one can be seen noticing that any transport map t induces a plan γ, defined by $(Id\times t)_\#\mu$. This plan is concentrated on the graph of t in $\mathbb{R}^n\times\mathbb{R}^n$ and it is easy to show that the converse holds, i.e. whenever γ is concentrated on a graph, then γ is induced by a transport map. Since any transport induces a planning with the same cost, we have

$$\inf (1.1) \geq \min (1.2).$$

Moreover, by approximating any planning by plannings induced by transports, it can be shown that equality holds under fairly general assumptions (see for instance [1]). Therefore we can really consider the Kantorovich formulation of the transport problem as a weak formulation of the original problem.

The theory of disintegration of measures (see for instance [12]) provides a very useful representation of transport plans: there exist probability measures γ_x in $\mathbb{R}^n$ such that $\gamma=\gamma_x\otimes\mu$, i.e.

$$\iint_{\mathbb{R}^n\times\mathbb{R}^n}\varphi(x,y)\,d\gamma(x,y)=\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\varphi(x,y)\,d\gamma_x(y)\right)\,d\mu(x).$$

for any bounded measurable function φ. In this sense we can consider a transport plan γ as a "stochastic" transport map $x\mapsto\gamma_x$, allowing the splitting of mass, and corresponding to a "deterministic" transport map only if γ_x is a Dirac mass for $\mu-$a.e. $x\in\mathbb{R}^n$.

Kantorovich's weak solutions are by now considered as the "natural" solutions of the problem in Probability and in some related fields and, besides the general existence theorem mentioned above, general necessary and sufficient conditions for optimality, based on a duality formulation, have been found (see for instance [26], [17], [28], [29], [33]). Notice that, by the Choquet theorem, the linearity of the functional and the convexity of the constraint $\gamma\in\Pi(\mu,\nu)$ ensure that the minimum is achieved on an *extremal* point of $\Pi(\mu,\nu)$. Therefore, if extremal points were induced by transports one would get existence of transport maps directly from the Kantorovich formulation. It is not difficult to show that plannings γ induced by transports are extremal in $\Pi(\mu,\nu)$, since the disintegrated measures γ_x are Dirac masses; the converse holds in some particular cases, as

$$\mu=\sum_{i=1}^N\frac{1}{N}\delta_{x_i},\qquad \nu=\sum_{i=1}^N\frac{1}{N}\delta_{y_i}$$

(by the well-known Birkhoff theorem) but unfortunately it is not true in general. It turns out that the existence of optimal transport maps seems to depend not only on the geometry of $\Pi(\mu,\nu)$, but also on the choice of the cost function c.

There are several results in the literature (see [5], [6], [7], [17], [19], [20], [24], [28], [29]) about the existence of optimal transport maps in the case when $c(x,y)=h(y-x)$, with h *strictly* convex.

The case when h is not strictly convex, corresponding to the original problem raised by Monge, is more subtle. Indeed, if $c(x,y)=|y-x|$ (the Euclidean distance) then, due to the loss of strict convexity in the radial direction, the standard duality methods provide information on the direction of transportation but no information on the distance $|y-x|$: to be precise one can show the existence of a $1-$Lipschitz map $\varphi:\mathbb{R}^n\to\mathbb{R}$ such that

$$(x,y)\in\operatorname{spt}\gamma\quad\Longrightarrow\quad y\in\{x-t\nabla\varphi(x):\ t\geq 0\}$$

for any differentiability point x of φ. Moreover, when $c(x,y)=\|y-x\|$ with $\|\cdot\|$ *not* strictly convex, then we have an even more severe loss of information about the possible location of y (lying a priori in a cone, and not a halfline): for a suitable $1-$Lipschitz function $\varphi : (\mathbb{R}^n, \|\cdot\|) \to \mathbb{R}$, we have

$$(x,y) \in \operatorname{spt}\gamma \quad \Longrightarrow \quad y \in \{x - t(d\varphi(x))^* : \ t \geq 0\},$$

where, for L in the unit ball of $(\mathbb{R}^n)^*$ (the dual of $\mathbb{R}^n$), the set L^* consists of all vectors $v \in \mathbb{R}^n$ such that $\|v\| = 1$ and $L(v) = 1$.

The first attempt to bypass these difficulties came with the work of Sudakov [30], who claimed to have a solution for any distance cost function induced by a norm. Sudakov's approach is based on a clever decomposition of the space $\mathbb{R}^n$ in affine regions with variable dimension where the Kantorovich potential associated to the transport problem is an affine function. His strategy is to solve the transport problem in any of these regions, eventually getting an optimal transport map just by gluing all these transport maps. An essential ingredient in his proof is Proposition 78 of [30], where it is stated that, if $\mu \ll \mathscr{L}^n$, then the conditional μ_C measures induced by the decomposition are absolutely continuous with respect to the Lebesgue measure (of the correct dimension). More precisely, when the decomposition is made in lines, than one needs to know at least that μ_C, supported on the line C, has no atom: in fact, this condition is sufficient to solve the optimal transportation problem in one space variable, via monotone rearrangement (see for instance [1], [33]). But, it turns out that when $n > 2$ the property claimed by Sudakov is not true even for the decomposition in segments: G.Alberti, the second author and D.Preiss [2] found an example of a compact family of pairwise disjoint open segments in $\mathbb{R}^3$ such that the family M of their midpoints has strictly positive Lebesgue measure. Therefore, by applying the disintegration theorem to $\mu = \mathscr{L}^3 \llcorner M$, we find that the conditional measures μ_C are Dirac masses! In Theorem A of §2 we improve their original construction building a "Nikodym set" in $\mathbb{R}^3$, i.e. a Borel set having full measure in the cube $[-1,1]^3$, intersecting each element of a family of pairwise disjoint open lines only in one point. This example also improves a previous one, still in $\mathbb{R}^3$, due to D.G.Larman [23].

Coming back to the problem of the existence of optimal transport maps, it is clear that this kind of pathological counterexamples should be ruled out by some kind of additional "regularity" property of the decomposition. In this way Sudakov's dimension reduction argument could be well implemented.

Several years later (and before the discovery, which came in 2000, that the proof in [30] was not correct), Evans and Gangbo attacked in [17] the problem by PDE methods, showing the existence of a transport map, under the assumption that $\operatorname{spt}\mu \cap \operatorname{spt}\nu = \emptyset$, that the two measures are absolutely continuous with respect to $\mathscr{L}^n$ and that their densities are Lipschitz functions with compact support. The missing piece of information about the length of transportation is recovered by the $p-$laplacian approximation

$$-\operatorname{div}\left(|\nabla\varphi|^{p-2}\nabla\varphi\right) = \mu - \nu, \qquad \varphi \in H^1_0(B_R), \qquad R \gg 1$$

obtaining in the limit as $p \to +\infty$ a nonnegative function $a \in L^\infty(\mathbb{R}^n)$ and a $1-$Lipschitz function φ solving

$$-\operatorname{div}(a\nabla\varphi) = \mu - \nu, \qquad |\nabla\varphi| = 1 \ \mathscr{L}^n\text{−a.e. on } \{a > 0\}.$$

The measure $\sigma = a\,\mathscr{L}^n$, the so-called *transport density*, plays an important role in the theory: its uniqueness and its regularity are discussed, under more general assumptions on μ and ν, in [1], [18], [13], [14].

The first existence results for general absolutely continuous measures μ, ν with compact support have been independently obtained by L.Caffarelli, M.Feldman and R.Mc Cann in [8] and by N.Trudinger and L.Wang in [31]. Afterwards, the first author established in [1] the existence of an optimal transport map assuming only that the initial measure μ is absolutely continuous. All these proofs basically involve a Sudakov-type decomposition in transport rays, but the technical implementation of the idea is different from paper to paper: for instance in [8] a local change of variable is made, so that transport rays become parallel and the Fubini theorem, in place of abstract disintegration theorems for measures, can be used. The proof in [1], instead, uses the co-area formula to show that absolute continuity with respect to Lebesgue measure is stable under disintegration (and this is where the "regularity" of the decomposition plays a role).

The strategy pursued in [8] is based on the approximation of the cost function $c(x,y)=\|y-x\|$, with $\|\cdot\|^2$ of class C^2 and *uniformly convex*, by the cost functions $c_\varepsilon(x,y)=\|y-x\|^{1+\varepsilon}$. The approximation is used in that paper to build a special Kantorovich potential φ, by taking limits as $\varepsilon\downarrow 0$ of the potentials $(\varphi_\varepsilon,\psi_\varepsilon)$ in the dual formulation (see (3.1)). The potential φ is used to prescribe the geometry of transport rays and to build, by a $1-$dimensional reduction, an optimal transport map. In [3] the first and third author improved and gave a variational interpretation of the Caffarelli-Feldman-McCann approximation, showing convergence in measure of the optimal transport maps t_ε relative to the costs c_ε to an optimal transport map for the limit problem.

The underlying variational principle in [3] (based on the theory of asymptotic developments by $\Gamma-$convergence, developed by G.Anzellotti and S.Baldo in [4]), is that any limit of the optimal plannings γ_ε associated to t_ε is not only optimal for the Kantorovich problem, but also optimal for the *secondary* variational problem

$$\min_{\gamma\in\Pi'(\mu,\nu)}\iint_{\mathbb{R}^n\times\mathbb{R}^n}\|y-x\|\ln\|y-x\|\,d\gamma, \tag{1.3}$$

where $\Pi'(\mu,\nu)$ denotes the class of all optimal plannings for the Kantorovich problem (the entropy function in (1.3) comes from the Taylor expansion of c_ε around $\varepsilon=0$). Then, it is shown that the secondary variational problem has a unique minimizer, and that this minimizer is induced by a transport map (a posteriori, the same map built in [8]).

In this paper we show that this principle can be used also in other situations to provide a "variational" decomposition in transport rays, bypassing the above mentioned difficulties in Sudakov's argument. In fact, we show in Theorem B of §5 existence of optimal transport maps for cost functions c induced by any "crystalline" norm $\|\cdot\|$ (whose unit sphere is contained in finitely many hyperplanes and which therefore is not strictly convex), assuming that μ and ν have finite first order moments and $\mu\ll\mathscr{L}^n$ (in view of the counterexample in [3] this assumption is close to be sharp even in the case $c(x,y)=|y-x|$).

In this case we look at the approximation

$$c_\varepsilon(x,y):=\|y-x\|+\varepsilon|y-x|+\varepsilon^2\alpha(|y-x|) \tag{1.4}$$

where $\alpha(t)$ is any strictly convex and nondecreasing function such that $\iint\alpha(|y-x|)d\mu\otimes\nu<+\infty$ (for instance $\alpha(t)=\sqrt{1+t^2}$, in view of the moment condition). Roughly speaking, the first perturbation solves the ambiguity about the direction of transport (due to the lack of strict convexity of the crystalline norm), while the higher order perturbation solves the ambiguity about the length of transportation. Therefore, the optimal transport maps t_ε relative to the cost function c_ε converge in measure as $\varepsilon\downarrow 0$ to an optimal transport map t for the limit problem (see the remark made after the end of the proof of Theorem B). In this way we prove existence of optimal transports with crystalline norms.

Our proof could easily be adapted to handle more general norms whose unit ball is in part uniformly convex and in part polyhedral, but there are some borderline cases that apparently can't be attacked by any of the existing techniques. For instance, if the norm is strictly convex but not uniformly convex, then the geometry of transport rays is prescribed, but we don't know whether the resulting decomposition is regular enough to produce, after disintegration, non-atomic measures. Therefore the general problem is definitely open, at least when $n\geq 3$: in Theorem C of §6 we show that this "regularity" property is always true in dimension 2, and therefore optimal transport maps exist for any norm.

2. A Nikodym set in $\mathbb{R}^3$

The main result of this section is the following.

Theorem A. *There exist a Borel set $M_N \subset [-1,1]^3$ with $|[-1,1]^3 \setminus M_N| = 0$ and a Borel map $f : M_N \to [-2,2]^2 \times [-2,2]^2$ such that the following holds. If we define for $x \in M_N$ the open segment l_x connecting $(f_1(x), -2)$ to $(f_2(x), 2)$, then*

- $\{x\} = l_x \cap M_N$ *for all* $x \in M_N$,
- $l_x \cap l_y = \emptyset$ *for all* $x, y \in M_N$ *different.*

The main ingredient in the construction is the following remarkable result due to Davies [11].

Theorem 2.1 (Davies)**.** *Let $Q = \overline{Q_{10}Q_{11}Q_{20}Q_{21}}$ be a parallelogram in $\mathbb{R}^3$ and let $K \subset Q$ be closed. Then for any $\varepsilon > 0$ there are "substripes" $\{\omega_i\}_{i=1}^k$, i.e. parallelograms $\omega_i = \overline{Q_{10}^iQ_{11}^iQ_{20}^iQ_{21}^i}$, which cover up to an error ε exactly K. This means*

a) $\overline{Q_{10}^iQ_{11}^i} \subset \overline{Q_{10}Q_{11}}$, $\overline{Q_{20}^iQ_{21}^i} \subset \overline{Q_{20}Q_{21}}$,
b) $K \subset \bigcup_1^k \omega_i$,
c) $\mathscr{H}^2\left(\bigcup_1^k \omega_i \setminus K\right) < \varepsilon$.

In our approach, the two dimensional stripes will be replaced by their three dimensional counterparts which will be called beams.

Definition 2.2 (Beams)**.** *A beam B is the image of the unit cube $[0,1]^3$ under an affine map, so should formally be identified with the pair (P_B, a_B) where $P_B \in \mathbb{M}^{3\times 3}$ is nonsingular, $a_B \in \mathbb{R}^3$ and $B = \{P_B(x) + a_B \ : \ x \in [0,1]^3\}$.*

Then we denote

- $axis(B) = P_B(e_3)$,
- $bottom(B) = P_B([0,1]^2 \times \{0\}) + a_B$,
- $top(B) = P_B([0,1]^2 \times \{1\}) + a_B$.

The ambiguity present in the definition will be ruled out in all situations to be dealt with by the fact that the beams will be very much stretched out into one of the three directions of their edges. This one will then considered to be the axis of the beam, and its orientation will be inherited from some larger beam containing the one considered.

The main new feature of the three dimensional situation is that we can replace any finite collection of finite families of beams by families of subbeams such that the new beams from different families are disjoint but the measure of the entire configuration decreases only as little as desired. To formulate this more precisely we first give the following

Definition 2.3. *Let B_1, B_2 be two beams. We write $B_1 \prec B_2$ provided $bottom(B_1) \subset bottom(B_2)$ and $top(B_1) \subset top(B_2)$. Similarly, we say that a family $\mathcal{F}_1$ of beams refines the family $\mathcal{F}_2$ of beams, and write $\mathcal{F}_1 \prec \mathcal{F}_2$, if for each $B \in \mathcal{F}_1$ there is a $B' \in \mathcal{F}_2$ with $B \prec B'$. Moreover we will write $\bigcup \mathcal{F}$ to denote $\bigcup_{B\in\mathcal{F}} B$.*

We will also use the following criterion ensuring when an intersection of beams is a beam again.

Lemma 2.4. *Let B be a beam and let T_B and B_B be its top and bottom. Moreover, let F_B and F'_B be two other closed opposite faces of B. Let S be the unbounded three dimensional stripe $\{x \ : \ \langle x, d\rangle \in [a,b]\}$ and suppose*

$$int(S) \cap \partial B \subset T_B \cup B_B \cup F_B \cup F'_B \ \text{and}\ int(S \cap B) \neq \emptyset.$$

Then $S \cap B$ is a beam and, properly oriented, we have $S \cap B \prec B$.

Proof. First, we assume $B = [0,1]^3$, $P_B = \mathrm{Id}$, $a_B = 0$ and $F_B = \{0\} \times [0,1]^2$, $F'_B = \{1\} \times [0,1]^2$. Then we claim that

$$S \cap B = \{x \ : \ \langle x, e_1 \rangle, \langle x, e_3 \rangle \in [0,1] \text{ and } \langle x, d \rangle \in [a,b]\} \tag{2.1}$$

Else there would be an x in the set on the right hand side which satisfies $|\langle x, e_2 \rangle| > 1$. Let $x_0 \in \mathrm{int}(S \cap B)$, then clearly for all y in the open segment (x, x_0) we have $\langle y, e_1 \rangle, \langle y, e_3 \rangle \in (0,1)$ and $y \in \mathrm{int}(S)$. Therefore, our assumption ensures $(x, x_0) \cap \partial B = \emptyset$. This obviously contradicts $x_0 \in \mathrm{int}(B)$, $x \notin B$.

Once (2.1) is established, we set

$$A = e_1 \otimes e_1 + e_2 \otimes \tilde{d} + e_3 \otimes e_3$$

where $\tilde{d} = d/(b-a)$ and $\Delta x = (a / \langle d, e_2 \rangle)\, e_2$. This is possible, since (2.1) also shows that e_1, d, e_3 form a basis of $\mathbb{R}^3$. It is now easily verified that

$$x \in S \cap B \Longleftrightarrow A(x - \Delta x) \in [0,1]^3.$$

Thus we conclude that $S \cap B = A^{-1}([0,1]^3) + \Delta x$ and in this parametrization it follows also that $\mathrm{bottom}(S \cap B) \subset \{x \in S \cap B \ : \ x_3 = 0\} \subset B_B$, similar for $\mathrm{top}(S \cap B)$.

Finally, given a general $B \simeq (P_B, a_B)$ we simply transform the problem to the beam $(Id, 0)$ and the stripe

$$\tilde{S} = P_B^{-1}(S - a_B) = \{x \ : \ \langle x, P_B^T d \rangle \in [a,b] - \langle a_B, d \rangle\}.$$

□

Proposition 2.5. *Let B_0 and $\mathcal{F}$ be respectively a beam and a finite family of beams. Then for each $\varepsilon > 0$ there are finite families*

$$\mathcal{F}'_0 \prec \{B_0\} \text{ and } \mathcal{F}' \prec \mathcal{F}$$

such that

(i) $\bigcup \mathcal{F}'_0 \cap \bigcup \mathcal{F}' = \emptyset$,

(ii) $|(B_0 \cup \bigcup \mathcal{F}) \setminus (\bigcup \mathcal{F}'_0 \cup \bigcup \mathcal{F}')| < \varepsilon$.

Proof. Without loss of generality, we can suppose that moreover for each $B \in \mathcal{F}$

$$\mathrm{axis}(B) \not\parallel \mathrm{axis}(B_0), \tag{2.2}$$

$$P_B(e_2) \in \mathrm{span}\{\mathrm{axis}(B_0), \mathrm{axis}(B)\}, \tag{2.3}$$

and that

$$\sum_{B \in \mathcal{F}} \Big| \Big((B_0 \cap B) + \mathbb{R}\,\mathrm{axis}(B_0) \Big) \cap (B \setminus B_0) \Big| < \frac{\varepsilon}{2}. \tag{2.4}$$

Indeed, (2.2) can be easily ensured replacing B_0 with $B'_0 \prec B_0$ having a slightly turned axis but almost the same measure as B_0. Once the "good" B_0 is fixed, it is now possible to achieve (2.3) for each $B \in \mathcal{F}$. For this purpose we choose a direction $v \in P_B(\mathbb{R}^2 \times \{0\}) \cap \mathrm{span}\{\mathrm{axis}(B), \mathrm{axis}(B_0)\}$ and cut B into finitely many $\hat{B}_i \prec B$ with the same axis but one other edge running in direction v, as showed in Figure 1. If we exhaust $\mathrm{bottom}(B)$ with the bottoms of the $\hat{B}_i$'s sufficiently well, we have also satisfied (2.3) but again lost just arbitrarily little measure –so (ii) will still be true.

Finally, also (2.4) can now be established splitting each of the by now chosen $B \in \mathcal{F}$ into sufficiently many $\tilde{B}_i \prec B$ with the same direction of edges. Indeed, let π_B be the projection onto the plane containing $\mathrm{bottom}(B)$ in direction $\mathrm{axis}(B)$. If $B' \prec B$ satisfies $\mathrm{bottom}(B') \cap \pi_B\big(\mathrm{bottom}(B_0) \cup \mathrm{top}(B_0)\big) = \emptyset$, then

$$\big((B_0 \cap B') + \mathbb{R}\,\mathrm{axis}(B_0)\big) \cap \big(B' \setminus B_0\big) = \emptyset.$$

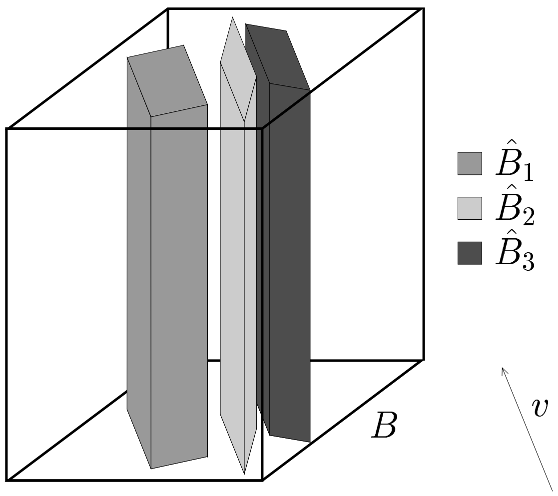

FIGURE 1. How to achieve (2.3)

Hence, if $\mathscr{H}^2\big(\pi_B(\mathrm{bottom}(B_0)\cup\mathrm{top}(B_0))\big)=0$ (equivalently, if $\mathrm{axis}(B)\in\mathrm{span}\{P_{B_0}(e_1),P_{B_0}(e_2)\}$) then it is clear how to choose the $\tilde{B}_i$'s. Else $\mathrm{axis}(B)$ intersects $\mathrm{top}(B_0)$ transversally and then it is easy to see that there is a constant $c_1<\infty$ such that for each $B'\prec B$ with the same direction of the edges and any $x\in B'$

$$\mathrm{diam}\Big(\pi_B^{-1}(x)\cap\big((B_0\cap B')+[0,\infty)\,\mathrm{axis}(B_0)\big)\cap(B'\setminus B_0)\Big)\le c_1\mathrm{diam}(\mathrm{bottom}(B')).$$

Because the shape of the $\mathrm{bottom}(B')$ is fixed to be that of $\mathrm{bottom}(B)$, and since the same argument applies if we replace $[0,\infty)$ by $(-\infty,0]$, we obtain

$$\big|\big((B_0\cap B')+\mathbb{R}\mathrm{axis}(B_0)\big)\cap\big(B'\setminus B_0\big)\big|\le c_2\,\mathrm{diam}(B')^3.$$

This means any sufficiently fine splitting of B into $\tilde{B}_i$'s will ensure (2.4).

The geometric idea behind our further procedure is simply to consider the intersection of the fixed beam B_0 with the "front" face $F_B=P_B([0,1]\times\{0\}\times[0,1))+a_B$ of a chosen B. In this planar setting we would use Davies' result to cover $F_B\setminus\big(B_0+\mathbb{R}\mathrm{axis}(B_0)\big)$ with stripes that have only a very small intersection with $F_B\cap B_0$. Then we would like to extend this pattern into the direction $\mathrm{axis}(B_0)$ as this should yield beams exhausting $B\setminus B_0$, but their "extensions" produce only little holes in B_0 itself. But since $\mathrm{axis}(B_0)$ is not necessarily precisely parallel to a "bottom" edge of B, some care is needed to ensure that the intersection of our extended stripes with B is a beam again.

Let us fix the notation (refer to Figure 2 for some help with it). Given a $B\in\mathcal{F}$, we know from (2.3) that there are $\lambda,\mu\in\mathbb{R}$ with $P_B(e_2)=\lambda\,\mathrm{axis}(B_0)+\mu\,\mathrm{axis}(B)$. For $t\in[0,1]$ let $I_t=[0,1]+t\mu$ and $S_t=P_B([0,1]\times\{0\}\times I_t)+a_B$. So S_t is just the projection of $P_B([0,1]\times\{t\}\times[0,1])+a_B$ in direction $\mathrm{axis}(B_0)$ into the plane spanned by $S_0=F_B$. We consider the parallelogram Q_B with corners

$$Q_{1i}=P_B((i,0,\min(0,\mu)))+a_B,\quad Q_{2i}=P_B((i,0,\max(0,\mu)+1))+a_B$$

for $i=0,1$ (note that $Q_B=\bigcup_{t\in[0,1]}S_t$), the convex set

$$K_B=Q_B\cap\big((B_0\cap B)+\mathbb{R}\mathrm{axis}(B_0)\big)$$

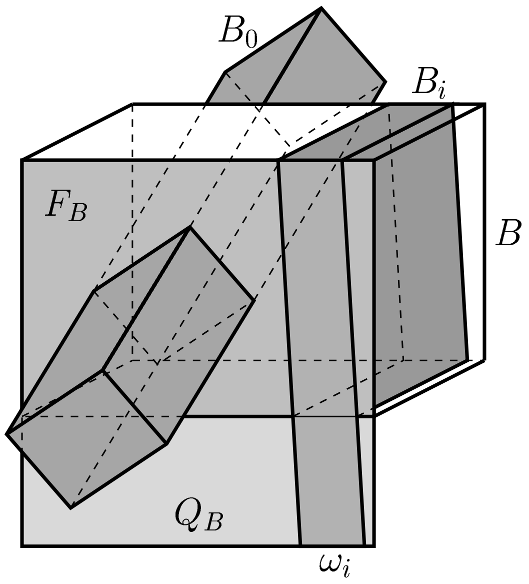

FIGURE 2. Construction of the family $\mathcal{F}_B$

and fix a small $\varepsilon_B > 0$ whose precise size will be specified in a moment.

Now we use Theorem 2.1 with $K = cl(Q_B \setminus K_B)$ to select $\{\omega_i\}_{i=1}^k$ refining Q_B with tops and bottoms in the interior of $\overline{Q_{20}Q_{21}}$ and $\overline{Q_{10}Q_{11}}$ respectively, such that

$$cl(Q_B \setminus K_B) \subset \bigcup_{i=1}^k \omega_i, \tag{2.5}$$

$$\mathscr{H}^2\Big(\Big(Q_B \cap \bigcup_{i=1}^k \omega_i\Big) \setminus cl(Q_B \setminus K_B)\Big) < \varepsilon_B. \tag{2.6}$$

Let us define ω_i^∞ as the unbounded stripe in the plane spanned by F_B whose sides contain the lateral sides of ω_i: the crucial observation is now that for each i and $B_i^\infty = \omega_i^\infty + \mathbb{R}\mathrm{axis}(B_0)$, $B_i = B_i^\infty \cap B$ we have, choosing a suitable orientation, $B_i \prec B$. This immediately follows from Lemma 2.4 because clearly $int(B_i^\infty \cap B) \neq \emptyset$ and moreover

$$int(B_i^\infty) \cap \partial B = \big(int(\omega_i^\infty) + \mathbb{R}\mathrm{axis}(B_0)\big) \cap \partial B \subset T_B \cup B_B \cup F_B \cup F'_B,$$

again recalling (2.3). Let us note now that from (2.5) it follows

$$Q_B \subset K_B \cup \bigcup \omega_i \subset K_B \cup \bigcup \omega_i^\infty.$$

We derive

$$B \subset Q_B + \mathbb{R}\mathrm{axis}(B_0) \subset K_B + \mathbb{R}\mathrm{axis}(B_0) \cup \bigcup B_i^\infty$$

and recalling $B_i = B_i^\infty \cap B$ it is then

$$B \subset K_B + \mathbb{R}\mathrm{axis}(B_0) \cup \bigcup B_i. \tag{2.7}$$

We then collect all the beams B_i constructed in this way into the family $\mathcal{F}_B \prec \{B\}$.

Repeating this procedure for all $B \in \mathcal{F}$ and finally setting $\mathcal{F}' = \bigcup_{B\in\mathcal{F}} \mathcal{F}_B$, obviously $\mathcal{F}' \prec \mathcal{F}$. To obtain now the family $\mathcal{F}'_0 \prec \{B_0\}$, we use that due to (2.6)

$$\Big|B_0 \cap \bigcup_{B\in\mathcal{F}} \bigcup_{B_i\in\mathcal{F}_B} B_i^\infty\Big| \le c_{B_0} \sum_{B\in\mathcal{F}} \varepsilon_B.$$

Because the ε_B's can be chosen arbitrarily small, we can assume $c_{B_0}\sum_{B\in\mathcal{F}}\varepsilon_B < \varepsilon/2$ and since all B_i^∞ have an edge parallel to axis(B_0), we can easily approximate the remainder $B_0 \setminus \bigcup\{B_i^\infty : B_i \in \mathcal{F}_B \text{ and } B \in \mathcal{F}\}$ from the inside with finitely many disjoint beams forming $\mathcal{F}_0' \prec \{B_0\}$ and satisfying

$$\big|B_0 \setminus \bigcup \mathcal{F}_0'\big| < \varepsilon/2. \tag{2.8}$$

Since $\mathcal{F}_0'$ approximates $B_0 \setminus \bigcup_{B\in\mathcal{F}}\bigcup_{B_i\in\mathcal{F}_B} B_i = B_0 \setminus \mathcal{F}'$ from the inside, of course (i) is satisfied. Finally, (ii) follows from (2.7), (2.4) and (2.8): in fact, (2.7) gives

$$\begin{aligned} B \setminus \big(B_0 \cup \bigcup \mathcal{F}_B\big) &\subset B \setminus \big(K_B + \mathbb{R}\mathrm{axis}(B_0)\big) \setminus \bigcup \mathcal{F}_B \cup \big(B \cap K_B + \mathbb{R}\mathrm{axis}(B_0)\big) \setminus B_0 \\ &\subset \big(B \cap K_B + \mathbb{R}\mathrm{axis}(B_0)\big) \setminus B_0. \end{aligned}$$

On the other hand, from the definition we have $K_B + \mathbb{R}\mathrm{axis}(B_0) = (B_0 \cap B) + \mathbb{R}\mathrm{axis}(B_0)$ and then

$$\begin{aligned} \big(B \cap K_B + \mathbb{R}\mathrm{axis}(B_0)\big) \setminus B_0 &= \Big((B_0 \cap B) + \mathbb{R}\mathrm{axis}(B_0) \cap B\Big) \setminus B_0 \\ &= (B_0 \cap B) + \mathbb{R}\mathrm{axis}(B_0) \cap \Big(B \setminus B_0\Big). \end{aligned}$$

Using then (2.4) we obtain

$$\Big|\bigcup \mathcal{F} \setminus \Big(B_0 \cup \bigcup \mathcal{F}'\Big)\Big| \le \sum\nolimits_{B\in\mathcal{F}} \Big|B \setminus \Big(\bigcup \mathcal{F}_B \cup B_0\Big)\Big| < \frac{\varepsilon}{2},$$

and finally, recalling (2.8), (ii) follows. □

Corollary 2.6. *Let $\mathcal{F}_1, \dots, \mathcal{F}_n$ be finite families of beams. Then for each $\varepsilon > 0$ there are finite families $\mathcal{F}_i' \prec \mathcal{F}_i$, $i = 1, \dots, n$ of beams such that*

(i) $\Big|\bigcup\limits_{i=1}^{n} \bigcup \mathcal{F}_i \setminus \bigcup\limits_{i=1}^{n} \bigcup \mathcal{F}_i'\Big| < \varepsilon,$

(ii) $\bigcup \mathcal{F}_i' \cap \bigcup \mathcal{F}_j' = \emptyset$ *for* $i \neq j$.

Proof. This is a simple induction argument, so suppose among all counterexamples our n is minimal (but at least 2 of course), and among all these with the minimal n also card$(\mathcal{F}_1)$ is minimal, and at least 1.

So we can choose $B_0 \in \mathcal{F}_1$ and set $\mathcal{F} = \bigcup_2^n \mathcal{F}_k$. We apply Proposition 2.5 to obtain $\mathcal{F}_0' \prec \{B_0\}$ and $\mathcal{F}' \prec \mathcal{F}$ with $\bigcup \mathcal{F}_0' \cap \bigcup \mathcal{F}' = \emptyset$ and $|(B_0 \cup \bigcup_2^n \mathcal{F}_k) \setminus \bigcup(\mathcal{F}_0' \cap \mathcal{F}')| < \varepsilon/2$. Obviously, we find $\hat{\mathcal{F}}_k \prec \mathcal{F}_k$ such that $\mathcal{F}' = \bigcup_2^n \hat{\mathcal{F}}_k$. Now it is possible to apply the induction hypothesis for $\hat{\varepsilon} = \varepsilon/2$ to $\mathcal{F}_1 \setminus \{B_0\}, \hat{\mathcal{F}}_2, \dots, \hat{\mathcal{F}}_n$ getting families $\mathcal{F}_1', \dots, \mathcal{F}_n'$ satisfying the corresponding conditions (i) and (ii). Clearly, $(\mathcal{F}_0' \cup \mathcal{F}_1'), \mathcal{F}_2', \dots, \mathcal{F}_n'$ now fulfill the original (i) and (ii). Because this can be done for any $\varepsilon > 0$, we obtained a contradiction proving the Corollary. □

Now we have at our disposal all the tools to prove Theorem A. We set $Q_u = [-1,1]^3$ and $B_0 \simeq (4 \cdot \mathrm{Id}, (-2,-2,-2))$, corresponding to the larger cube $[-2,2]^3$.

Let $\mathcal{D}_0 = \{B_0\}$ and $\mathcal{C}_0 = \{Q_u\}$. We choose $\mathcal{F}_0 \prec \{B_0\}$ which covers Q_u, all beams in $\mathcal{F}_0$ have width (i.e the diameter of top or bottom) at most 1 and the tails of these beams are small, i.e. $|\bigcup \mathcal{F}_0 \setminus Q_u| < 2^{-1}$. This is easily done using Davies' result stated in Theorem 2.1. Finally we set the empty family $\mathcal{N}_0 = \emptyset$ and the empty set T_0, and note the trivial estimate $|Q_u \setminus \bigcup(\mathcal{F}_0 \cup \mathcal{N}_0)| < 2^{-0}$.

To iterate our constructions, suppose now that finite families of beams $\mathcal{F}_k \prec \{B_0\}$, $\mathcal{N}_k \prec \{B_0\}$, $\mathcal{C}_k \subset Q_u$ and a set $T_k \subset B_0$, $k = 0, \dots, n$ are already given and satisfy for each k

$$\Big|Q_u \setminus \bigcup(\mathcal{F}_k \cup \mathcal{N}_k)\Big| < 2^{-k-1}, \tag{2.9}$$

$$\left|Q_u \setminus \bigcup \mathcal{C}_k\right| < 2^{-k+1}, \tag{2.10}$$

$$\left|\bigcup \mathcal{N}_k\right| < 3 \cdot 2^{-k-1}, \tag{2.11}$$

$$\bigcup \mathcal{F}_k \cap \bigcup \mathcal{N}_k = \emptyset, \tag{2.12}$$

$$\mathcal{F}_k = \bigcup\nolimits_{C \in \mathcal{C}_k} \mathcal{F}_k(C) \text{ with } C \subset \bigcup \mathcal{F}_k(C) \text{ and } \left|\bigcup \mathcal{F}_k(C) \setminus C\right| < \frac{2^{-k-1}}{\operatorname{card}(\mathcal{C}_k)}, \tag{2.13}$$

$$\text{the width of all beams in } \mathcal{F}_k \cup \mathcal{N}_k \text{ is at most } (k+1)^{-1}. \tag{2.14}$$

Splitting $\mathcal{F}_n \cup \mathcal{N}_n$ into finitely many families such that the direction of the axes varies in each of the families at most $(n+2)^{-1}$, we can use Corollary 2.6 to find $\mathcal{D}_{n+1} \prec \mathcal{F}_n \cup \mathcal{N}_n$ (we write $\mathcal{D}_{n+1} = \{D_i : i = 1, 2, \ldots, M\}$) such that

$$|Q_u \setminus \bigcup \mathcal{D}_{n+1}| < 2^{-n-1} \tag{2.15}$$

$$\begin{array}{l}\text{if } B, B' \in \mathcal{D}_{n+1} \text{ and } B \cap B' \neq \emptyset, \text{ then the angle} \\ \text{between axis}(B) \text{ and axis}(B') \text{ is at most } (n+2)^{-1}.\end{array} \tag{2.16}$$

Now we want to select a finite family $\mathcal{C}_{n+1}$ of disjoint beams with the property that each one is in the interior of some $D_i \in \mathcal{D}_{n+1}$: we claim that this can be done in such a way that the following holds:

$$\bigcup \mathcal{C}_{n+1} \subset \bigcup \mathcal{D}_{n+1} \cap \operatorname{int}(Q_u), \tag{2.17}$$

$$\begin{array}{l}\text{for each } C \in \mathcal{C}_{n+1} \text{ there is a } B = B_{n+1}(C) \in \mathcal{D}_{n+1} \text{ such that} \\ C \subset \operatorname{int}(B) \text{ and } C = s \cdot B + x \text{ for some } s \in (0,1), x \in \mathbb{R}^3,\end{array} \tag{2.18}$$

$$\text{if } C \in \mathcal{C}_{n+1} \text{ and } C \cap \bigcup \mathcal{C}_n \neq \emptyset, \text{ then there is a } C' = C_n(C) \in \mathcal{C}_n \text{ satisfying } C \subset \operatorname{int}(C') \tag{2.19}$$

$$|Q_u \setminus \bigcup \mathcal{C}_{n+1}| < 2^{-n} \text{ (then we exhaust almost all } Q_u \cap \bigcup \mathcal{D}_{n+1}). \tag{2.20}$$

To do this, let us consider separately each $D_i \in \mathcal{D}_{n+1}$: we can fill the interior of $Q_u \cap D_i \setminus \cup_{j<i} D_j$ with disjointed beams satisfying (2.18) with the property that, if $D \prec \mathcal{F}_n(C)$, then all the beams are contained in the interior of C. The union of all these beams is $\mathcal{C}_{n+1}$. We note now that (2.17) and (2.18) clearly hold by construction, as the disjointness of the elements of $\mathcal{C}_{n+1}$. Concerning (2.19), if $C \in \mathcal{C}_{n+1}$ is in the interior of some $D \prec \mathcal{F}_n$, then $D \prec \mathcal{F}_n(\widetilde{C})$ for some $\widetilde{C} \in C_n$ thanks to (2.13) and then by our construction $C \subset \operatorname{int}(\widetilde{C})$; on the other hand, if C is in the interior of some $D \prec \mathcal{N}_n$ then C does not intersect $\bigcup \mathcal{F}_n$ thanks to (2.12) and then it does not intersect $\mathcal{C}_n$ since, again by (2.13), $\mathcal{C}_n \subset \bigcup \mathcal{F}_n$. Finally (2.20) can be obtained if we fill almost completely $Q_u \cap D_i \setminus \cup_{j<i} D_j$ whenever $D_i \prec \mathcal{N}_n$ and $Q_u \cap C \cap D_i \setminus \cup_{j<i} D_j$ when $D_i \prec \mathcal{F}_n(C)$, using (2.15) and (2.13). Note that the condition (2.19) says that the new beams in $\mathcal{C}_{n+1}$ respect the tree-like structure which is already existing on the part of $\bigcup \mathcal{D}_{n+1}$ that comes from $\mathcal{F}_n$ and hence $\mathcal{C}_n$.

Because the shapes of C and $B_{n+1}(C)$ agree, we can easily use Davies' result (as we did to build $\mathcal{F}_0$) to find for each $C \in \mathcal{C}_{n+1}$ a finite family $\mathcal{F}_{n+1}(C) \prec \{B_{n+1}(C)\}$ covering C with narrow beams having small tails. This means

$$C \subset \bigcup \mathcal{F}_{n+1}(C),$$

$$\left|\bigcup \mathcal{F}_{n+1}(C) \setminus C\right| < 2^{-n-2}/\operatorname{card}(\mathcal{C}_{n+1}), \tag{2.21}$$

$$\text{the width of } F \text{ is at most } (n+2)^{-1} \text{ for all } F \in \mathcal{F}_{n+1}(C). \tag{2.22}$$

Setting

$$\mathcal{F}_{n+1} := \bigcup_{C\in\mathcal{C}_{n+1}} \mathcal{F}_{n+1}(C) \qquad\qquad T_{n+1} := \bigcup_{C\in\mathcal{C}_{n+1}} \Big(\bigcup \mathcal{F}_{n+1}(C)\setminus C\Big)$$

we have therefore (thanks to (2.20) and (2.21))

$$|Q_u \setminus \bigcup \mathcal{F}_{n+1}| < 2^{-n} \qquad\qquad \text{and} \qquad\qquad |T_{n+1}| < 2^{-n-2}. \tag{2.23}$$

At the end we select a finite family $\mathcal{N}_{n+1} \prec \{B_0\}$ with

$$\bigcup \mathcal{N}_{n+1} \cap \bigcup \mathcal{F}_{n+1} = \emptyset \tag{2.24}$$

to fill the holes our disjointification process created and improve estimate (2.9). For this purpose we require

$$\Big|\bigcup \mathcal{N}_{n+1}\Big| < 3\cdot 2^{-n-2} \tag{2.25}$$

$$\Big|Q_u \setminus \bigcup \Big(\mathcal{N}_{n+1}\cup \mathcal{F}_{n+1}\Big)\Big| < 2^{-n-2} \tag{2.26}$$

$$\text{each beam in } \mathcal{N}_{n+1} \text{ is of width at most } (n+2)^{-1}. \tag{2.27}$$

In this way we defined $\mathcal{F}_{n+1}$, $\mathcal{N}_{n+1}$, $\mathcal{C}_{n+1}$ and T_{n+1}. It follows from (2.26), (2.20), (2.25), (2.24), (2.21) and (2.22), (2.27) that (2.9), (2.10), (2.11), (2.12), (2.13) and (2.14) hold also for $k = n+1$.

Therefore we can keep iterating this procedure. Once these sequences of families of beams are given, we define our desired set

$$M_N := \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\Big(\bigcup \mathcal{C}_k \setminus \Big(\bigcup \mathcal{N}_k \cup T_k\Big)\Big).$$

By (2.17), $M_N \subset Q_u$ and $|Q_u \setminus M_N| = 0$ by (2.10), (2.11) and the second statement in (2.23). Let us consider an arbitrary $x \in M_N$, then $x \in \bigcup \mathcal{C}_k \setminus (\bigcup \mathcal{N}_k \cup T_k)$ for all $k \geq k_x$. Hence, since each $\mathcal{C}_k$ is a disjoint family, there is a unique sequence of $C_k \in \mathcal{C}_k$, $k \geq k_x$, with $x \in C_k$, and moreover $C_k = C_k(C_{k+1})$ –compare with (2.19). By (2.18) there are also $B_k(C_k) \in \mathcal{D}_k$ defined, and we claim now that also

$$B_{k+1}(C_{k+1}) \subset B_k(C_k) \text{ for } k \geq k_x. \tag{2.28}$$

Indeed, we know $B_{k+1}(C_{k+1}) \prec F \in \mathcal{F}_k \cup \mathcal{N}_k$, but now $x \notin \bigcup \mathcal{N}_k$ implies $F \in \mathcal{F}_k$. Thus $F \in \mathcal{F}_k(C')$ for some $C' \in \mathcal{C}_k$. If $C' \neq C_k$ then $C' \cap C_k = \emptyset$ and then $x \in \bigcup \mathcal{F}_k(C') \setminus C' \subset T_k$, contradiction. Therefore $C' = C_k$ and $B_{k+1}(C_{k+1}) \subset F \subset \bigcup \mathcal{F}_k(C_k) \subset B_k(C_k)$, and (2.28) is established. Because all beams in $\mathcal{D}_{k+1}$ have due to (2.14) width at most $1/k$, we see that we can define a segment l_x containing x and connecting $[-2,2]^2 \times \{-2\}$ to $[-2,2]^2 \times \{2\}$ by setting $l_x = \bigcap_{k=k_x}^{\infty} B_k(C_k)$. We also know that $l_x \subset B_{k+1}(C_{k+1}) \subset \bigcup \mathcal{F}_k(C_k) \subset C_k \cup T_k$. Thus $l_x \cap (\bigcup \mathcal{C}_k \setminus (\bigcup \mathcal{N}_k \cup T_k)) \subset C_k$: it follows $l_x \cap M_N \subset \bigcup_{m \geq k_x} \bigcap_m^{\infty} C_k$; on the other hand, $\bigcap_m^{\infty} C_k = \{x\}$ because $x \in C_k$ for any k and both the width and the height of C_k go to 0: then $l_x \cap M_N \subset \{x\}$ and we derive $l_x \cap M_N = \{x\}$ by the construction. This establishes the first conclusion of the theorem.

Moreover, from (2.16) and (2.28) it is clear that l_x, l_y with $x, y \in M_N$ can not intersect provided they are not parallel –and in that case obviously $l_x = l_y$. Hence, if $l_x \cap l_y \neq \emptyset$, then $l_x = l_y$ and so $\{x\} = \{y\} = M_n \cap l_x$, proving the second claim; the statement about the Borel measurability is obvious for M_N itself. If we define f as a pointwise limit of functions f_k which are on each of the closed $C \in \mathcal{C}_k$ constantly equal to a value corresponding to an "edge" axis of $B_k(C)$, then f is Borel measurable and $f(x)$ corresponds to the segment l_x for each $x \in M_N$. This finishes our proof.

3. Preliminary results on optimal transportation

Let μ, ν be probability measures in $\mathbb{R}^n$ and let $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty]$ be a lower semicontinuous function. It is well known (see for instance [3], [26], [33]) that the following duality formula for the Monge–Kantorovich problem holds:

$$\min(1.2) = \sup\left\{\int_{\mathbb{R}^n} \varphi(x)\, d\mu(x) + \int_{\mathbb{R}^n} \psi(y)\, d\nu(y)\right\} \tag{3.1}$$

where the supremum runs among all pairs $(\varphi, \psi) \in L^1(\mu) \times L^1(\nu)$ with $\varphi(x) + \psi(y) \leq c(x, y)$. Furthermore, under the assumption (see [3] for more general conditions)

$$\iint_{\mathbb{R}^n\times\mathbb{R}^n} c(x, y)\, d\mu \otimes \nu < +\infty \tag{3.2}$$

the supremum is attained. In the special case when c is a distance the maximizing pair is of the form $(\varphi, -\varphi)$, with φ 1−Lipschitz with respect to c. We will call *Kantorovich potential* any such maximizer φ. From (3.1) and the inequality $\varphi(x) - \varphi(y) \leq c(x, y)$ it follows that $\gamma \in \Pi(\mu, \nu)$ is optimal if and only if $\varphi(x) - \varphi(y) = c(x, y)$ γ−a.e. in $\mathbb{R}^n \times \mathbb{R}^n$. We will use this optimality condition several times throughout the paper.

Given $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$, we define the associated notions of *transport ray*, *ray direction*, *transport set*, *fixed point*, *maximal transport ray*, *extreme point*. The terminology is close to the one adopted in [17] and [8], with the difference that our definitions depend on Γ rather than a Kantorovich potential (see §6 of [3] for a more detailed discussion).

(Transport ray) We say that the open segment $]\!]x, y[\![$ is a transport ray if $x \neq y$ and $(x, y) \in \Gamma$.
(Ray direction) Given a transport ray $]\!]x, y[\![$, we denote its direction by

$$\tau(x, y) := \frac{(y - x)}{|y - x|}.$$

(Transport set and its orientation) We denote by T_Γ the union of all transport rays relative to Γ, i.e.

$$\bigcup_{(x,y)\in\Gamma}]\!]x, y[\![.$$

Assuming that distinct transport rays of Γ do not intersect unless they are parallel, the map $\tau_\Gamma : T_\Gamma \to \mathbb{S}^{n-1}$ which associates to any x the direction of transport rays containing x is well defined.
(Fixed and extreme points) We say that x is a fixed point if $(x, x) \in \Gamma$. We denote by F_Γ the set of all fixed points.

We say that x is a (left) extreme point if $x \in \pi_1(\Gamma) \setminus (F_\Gamma \cup T_\Gamma)$.
(Maximal transport ray) We say that an open oriented interval $S \subset \mathbb{R}^n$ (possibly unbounded) is a maximal transport ray if

(a) for any $z \in S$ there exists $(x, y) \in \Gamma$ with $z \in]\!]x, y[\![\subset S$;
(b) any open oriented interval containing S and satisfying (a) coincides with S.

In the following we shall always assume that Γ is a σ−compact set. This ensures that the set F_Γ, T_Γ and the set of extreme points are Borel (see Lemma 6.3 in [3]).

We recall that a set $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is said to be *c−cyclically monotone* if

$$\sum_{i=1}^{p} c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^{p} c(x_i, y_i)$$

whenever $(x_i, y_i) \in \Gamma$, $1 \leq i \leq p$, and σ is a permutation of $\{1, \ldots, p\}$.

Our set Γ will be chosen in a such a way that an optimal plan γ is concentrated on it and a suitable (and crucial) cyclical monotonicity property holds. Specifically, we will use the following result from [3]:

Theorem 3.1 (Necessary minimality condition). *Let $\beta : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty]$ be a lower semicontinuous cost function and let γ be an optimal transport plan with $C(\gamma) < \infty$. Then there exists a $\beta-$monotone Borel set Γ on which $\beta < +\infty$ and γ is concentrated.*

Proof. For the reader's convenience we include the simple proof from [3]. Let (φ_m, ψ_m) be a maximizing sequence in the duality formula (3.1) and let $\beta_m = \beta - \varphi_m - \psi_m$. Since

$$\iint_{X\times Y} \beta_m \, d\gamma = \iint_{X\times Y} \beta \, d\gamma - \int_X \varphi_m \, d\mu - \int_Y \psi_m \, d\nu \to 0$$

and $\beta_m \geq 0$ we can find a subsequence $\beta_{m(k)}$ and a Borel set Γ on which γ is concentrated and β is finite, such that $\beta_{m(k)} \to 0$ on Γ. If $\{(x_i, y_i)\}_{1\leq i\leq p} \subset \Gamma$ and σ is a permutation of $\{1, \ldots, p\}$ we get

$$\begin{aligned}\sum_{i=1}^{p} \beta(x_i, y_{\sigma(i)}) &\geq \sum_{i=1}^{p} \varphi_{m(k)}(x_i) + \psi_{m(k)}(y_{\sigma(i)}) \\ &= \sum_{i=1}^{p} \varphi_{m(k)}(x_i) + \psi_{m(k)}(y_i) = \sum_{i=1}^{p} \beta(x_i, y_i) - \beta_{m(k)}(x_i, y_i)\end{aligned}$$

for any k. Letting $k \to \infty$ the $\beta-$monotonicity of Γ follows. □

The following result, taken from Theorem 7.2 and Theorem 6.5 in [3], is concerned with the case when the cost function is the Euclidean distance, and is a refinement of [8].

Theorem 3.2. *Let $\alpha(t) = \sqrt{1+t^2}$. If μ, ν have finite first order moments and $\mu \ll \mathscr{L}^n$ there exists a unique plan of transport $\gamma \in \Pi(\mu, \nu)$ which is optimal for the Euclidean cost $c(x, y) = |y-x|$ and which minimizes*

$$\iint_{\mathbb{R}^n\times\mathbb{R}^n} \alpha(|y-x|) \, d\eta(x, y)$$

among all the optimal plans η for the Kantorovich problem with Euclidean cost.
Moreover, this transport plan is induced by a transport and, since γ is concentrated on a $\sigma-$compact set Γ, the family of extreme points of Γ is Lebesgue negligible.

Notice that α plays only an auxiliary role: any strictly convex and non decreasing function $\alpha : [0, +\infty) \to [0, +\infty)$ such that

$$\iint_{\mathbb{R}^n\times\mathbb{R}^n} \alpha(|y-x|) \, d\mu \otimes \nu(x, y) < +\infty$$

would give the same result, and select the same transport map. On the other hand, the C^2 regularity of c^2 and its uniform convexity play an important role in the proof. For this reason in this theorem the assumption was made that c is the Euclidean distance.

4. The secondary variational problem

In this section, we will consider the case when the cost is a distance induced by a norm, i.e.

$$c(x, y) = \|y - x\|,$$

where $\|\cdot\|$ is a norm in $\mathbb{R}^n$. We will apply Theorem 3.1 with the cost function defined by

$$\beta(x, y) := \begin{cases} |y-x| & \text{if } c(x, y) \leq \varphi(y) - \varphi(x) \\ +\infty & \text{if } c(x, y) > \varphi(y) - \varphi(x). \end{cases} \tag{4.1}$$

where φ is a Kantorovich potential. Indeed, $\beta-$monotonicity implies the following useful property of transport rays (here and in the following Δ is the diagonal of $\mathbb{R}^n \times \mathbb{R}^n$):

Lemma 4.1. *If Γ is $\beta-$cyclically monotone and $\beta < +\infty$ on Γ, then for any pair of points (x,y), $(\tilde{x},\tilde{y}) \in \Gamma \setminus \Delta$ the condition $]\!]x,y[\![\cap [\![\tilde{x},\tilde{y}]\!] \neq \emptyset$ implies that the two segments are parallel (i.e. have the same oriented direction).*

Proof. Assume that the transport rays $]\!]x,y[\![$ and $[\![\tilde{x},\tilde{y}]\!]$ intersect at s: then, denoting by φ the function in (4.1), since $(x,y) \in \Gamma$ and c is induced by a norm we have $c(x,y) = \varphi(y) - \varphi(x)$ and $c(x,s) + c(s,y) = c(x,y)$, therefore $c(x,s) = \varphi(s) - \varphi(x)$; in the same way $c(s,\tilde{y}) = \varphi(\tilde{y}) - \varphi(s)$ and then the two inequalities give

$$c(x,\tilde{y}) \leq c(x,s) + c(s,\tilde{y}) = \varphi(\tilde{y}) - \varphi(x).$$

In the same way we have also $c(\tilde{x},y) \leq \varphi(y) - \varphi(\tilde{x})$. Now, using the $\beta-$cyclical monotonicity of Γ we infer that

$$|y-x| + |\tilde{y}-\tilde{x}| \leq |\tilde{y}-x| + |y-\tilde{x}|,$$

an inequality that cannot hold when the segments $]\!]x,y[\![$ and $[\![\tilde{x},\tilde{y}]\!]$ intersect, unless they are parallel. □

Let now $\Pi'(\mu,\nu)$ be the set of the optimal transport plans relative to the cost c and consider the variational problem

$$\inf\left\{ C_{\mathrm{eucl}}(\gamma) := \iint_{\mathbb{R}^n\times\mathbb{R}^n} |y-x|\,d\gamma(x,y) :\ \gamma \in \Pi'(\mu,\nu) \right\}. \tag{4.2}$$

Define then $\Pi''(\mu,\nu) \subset \Pi'(\mu,\nu)$ as the set of the minimizers and finally consider the problem

$$\inf\left\{ C_{\mathrm{conv}}(\gamma) := \iint_{\mathbb{R}^n\times\mathbb{R}^n} \alpha(|y-x|)\,d\gamma(x,y) :\ \gamma \in \Pi''(\mu,\nu) \right\}, \tag{4.3}$$

with $\alpha(t) = \sqrt{1+t^2}$. A standard compactness and lower semicontinuity argument shows that the sets $\Pi'(\mu,\nu)$ and $\Pi''(\mu,\nu)$ are not empty and that there exists a solution of (4.3).

Thanks to the necessary and sufficient optimality conditions and our assumption (3.2), there exists a Kantorovich potential φ such that $\gamma \in \Pi'(\mu,\nu)$ if and only if $c(x,y) = \varphi(y) - \varphi(x)$ $\gamma-$a.e. Therefore, defining β as in (4.1), the problem (4.2) can be rephrased as

$$\inf\left\{ \iint_{\mathbb{R}^n\times\mathbb{R}^n} \beta(x,y)\,d\gamma(x,y) :\ \gamma \in \Pi(\mu,\nu) \right\}.$$

Hence, according to Theorem 3.1, we can find a $\beta-$cyclically monotone Borel set Γ on which β is finite and $\gamma \in \Pi''(\mu,\nu)$ is concentrated. This implies, by Lemma 4.1, that distinct transport rays induced by Γ do not intersect, unless they are parallel. Using the minimality in (4.3) we can actually recover stronger "geometric" properties; first we note the

Lemma 4.2. *Solving (4.3) is a local property, i.e. if γ solves (4.3) then for any Borel subset U of $\mathbb{R}^n \times \mathbb{R}^n$ the plan $\gamma \llcorner U$ solves problem (4.3) with the marginals of $\gamma \llcorner U$ in place of μ and ν.*

Proof. This is a standard fact based on the linearity of the energy with respect to γ: let μ_U, ν_U be the marginals of $\gamma \llcorner U$ and $\eta \in \Pi(\mu_U,\nu_U)$. Comparing γ with

$$\hat{\gamma} := \gamma \llcorner (\mathbb{R}^n \times \mathbb{R}^n \setminus U) + \eta$$

we obtain $C(\eta) \geq C(\gamma \llcorner U)$. Therefore $\gamma \llcorner U \in \Pi'(\mu_U,\nu_U)$. Choosing now $\eta \in \Pi'(\mu_U,\nu_U)$, the same comparison argument gives $C_{\mathrm{eucl}}(\eta) \geq C_{\mathrm{eucl}}(\gamma \llcorner U)$, therefore $\gamma \llcorner U \in \Pi''(\mu_U,\nu_U)$. By applying once more the comparison argument we obtain that $\gamma \llcorner U$ is optimal in (4.3). □

More generally we can prove the following result (actually the preceding Lemma could be obtained from the next one with $\pi = \chi_U$ and $Z = \{0,1\}$):

Lemma 4.3. *Let Z be a subset of an Euclidean space, let $\pi : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow Z$ be a Borel function and let γ be a solution of (4.3). If $\gamma = \gamma_z \otimes \pi_\# \gamma$ is the disintegration of γ induced by π, and μ_z and ν_z are the marginals of γ_z, then γ_z solves problem (4.3) with μ_z and ν_z in place of μ and ν for $\pi_\#\gamma-$a.e. $z \in Z$.*

Proof. Let E be the space of all nonnegative measures on $\mathbb{R}^n \times \mathbb{R}^n$ with mass at most 1, endowed with a compact metric inducing the w^*-topology. Let $\eta = \pi_\#\gamma$; we recall that a multifunction $\Pi : Z \to \mathcal{P}(E)$ is said to be $\eta-$measurable if its graph $\{(z,\rho) : \rho \in \Pi(z)\}$ is measurable with respect to the product $\sigma-$algebra $\mathcal{B}_\eta(Z) \otimes \mathcal{B}(E)$ (here $\mathcal{B}_\eta(Z)$ is the completion of the Borel $\sigma-$algebra of Z with respect to η).

It is not hard to show, using the $\eta-$measurability of the map $z \mapsto \gamma_z$ (ensured by the disintegration theorem), that $z \mapsto \Pi(\mu_z, \nu_z)$ is $\eta-$measurable. We claim that the same is true for the multifunctions $z \mapsto \Pi'(\mu_z,\nu_z)$, $z \mapsto \Pi''(\mu_z,\nu_z)$ and finally for the multifunction $z \mapsto \Pi'''(\mu_z,\nu_z)$ giving the solutions of (4.3) with marginals μ_z, ν_z. Indeed, let us first notice that the value function

$$v(z) := \min\{C(\rho) : \ \rho \in \Pi(\mu_z,\nu_z)\}$$

is $\eta-$measurable as a consequence of the projection theorem (see [9]), because for any $t \in \mathbb{R}$ the set $\{v < t\}$ is the projection on Z of the $\mathcal{B}_\eta(Z) \otimes \mathcal{B}(E)-$measurable set

$$\{(z,\rho) : \ \rho \in \Pi(\mu_z,\nu_z), \ C(\rho) < t\}.$$

Since $\Pi'(\mu_z,\nu_z) = \Pi(\mu_z,\nu_z) \cap \{C = v(z)\}$ the $\eta-$measurability of $z \mapsto \Pi'(\mu_z,\nu_z)$ is proved. Repeating this argument twice more with C_{eucl} and C_{conv} we find that $z \mapsto \Pi''(\mu_z,\nu_z)$ and $z \mapsto \Pi'''(\mu_z,\nu_z)$ are $\eta-$measurable as well.

By the previous claim and by the measurable selection theorem (see [9] again) we can find a $\eta-$measurable map $z \mapsto \tilde{\gamma}_z$ such that $\tilde{\gamma}_z$ solves (4.3) with marginals (μ_z,ν_z) for any $z \in Z$. Therefore we can define a "competitor" $\tilde{\gamma}$ to γ setting $\tilde{\gamma} = \tilde{\gamma}_z \otimes \eta$. The properties of disintegration give

$$C(\gamma) = \iint_{\mathbb{R}^n\times\mathbb{R}^n} c(x,y)\,d\gamma(x,y) = \int_Z \left(\iint_{\mathbb{R}^n\times\mathbb{R}^n} c(x,y)\,d\gamma_z(x,y)\right) d\eta(z) = \int_Z C(\gamma_z)\,d\pi_\#\gamma(z),$$

and the same is true replacing C by C_{eucl} or C_{conv} or γ by $\tilde{\gamma}$. Since for any $z \in Z$ we have $C(\tilde{\gamma}_z) \le C(\gamma_z)$, the optimality of γ ensures that $\gamma_z \in \Pi'(\mu_z,\nu_z)$ for $\eta-$a.e. $z \in Z$. In the same way we find that $\gamma_z \in \Pi''(\mu_z,\nu_z)$ and finally that γ_z is a solution of (4.3) for $\eta-$a.e. $z \in Z$. ☐

The last property we need will ensure that a transport ray cannot be strictly contained in another; in the proof of this fact we need the following definition.

Definition 4.4 (Non decreasing transport plans)**.** *Let $C \subset \mathbb{R}$ be an open interval and let γ be a probability measure in $C \times C$. We say that γ is* non decreasing *if there exists a non decreasing function $f : C \to C$ such that $\operatorname{spt}\gamma$ is concentrated on the extended graph of f, defined by*

$$\Gamma(f) := \left\{(t,s) \in C \times C : \ f(t^-) \le s \le f(t^+)\right\}, \tag{4.4}$$

with the notation $f(t^-) = \lim_{\tau\uparrow t} f(\tau)$ and $f(t^+) = \lim_{\tau\downarrow t} f(\tau)$ for any non decreasing function f.

Notice that $\Gamma(f) \cap \{t\} \times C$ contains only $f(t)$ for any continuity point of t, therefore the disintegration of $\gamma = \gamma_t \otimes \mu$ with respect to the projection on the first coordinate is made up of Dirac masses with at most countably many exceptions. Therefore $\gamma = (Id, f)_\#\mu$ if μ has no atom and it is well known that f is the unique non decreasing map pushing μ into the second marginal of γ. The following more general result is presumably known, but we include it for the reader's convenience.

Proposition 4.5. *Let $C \subset \mathbb{R}$ be an open interval (possibly unbounded) and let μ, ν be probability measures in C. Then the class $\Pi(\mu,\nu)$ contains only one non decreasing transport plan γ. Moreover, if μ and ν have finite first order moments then γ is a solution of the problem*

$$\min\left\{\iint_{C\times C} |y-x|\,d\eta :\ \eta\in\Pi(\mu,\nu)\right\} \tag{4.5}$$

and is the unique solution of the problem

$$\min\left\{\iint_{C\times C} \alpha(|y-x|)\,d\eta :\ \eta\in\Pi(\mu,\nu)\right\}. \tag{4.6}$$

Proof. **Uniqueness:** First of all, let us write $C =]a,b[$ and define $\bar f : C \longrightarrow \overline{C}$ as follows

$$\bar f(s) := \inf\left\{y : \mu(]a,s[) \le \nu(]a,y[)\right\}.$$

Let now γ be any non decreasing transport plan: we claim that $\operatorname{spt}\gamma \subset \Gamma(\bar f)$. In fact, if $(x,y)\in \operatorname{spt}\gamma$ and f is as in the Definition 4.4, we have

$$\nu(]a,y[) \le \mu(]a,x]) \qquad \text{and} \qquad \mu(]a,x[) \le \nu(]a,y]). \tag{4.7}$$

To show these inequalities, note that if $t > x$ and $(t,s) \in \operatorname{spt}\gamma$ then (4.4) gives $s \ge f(t^-) \ge f(x^+) \ge y$. Then

$$\nu(]a,y[) = \gamma(]a,b[\times]a,y[) = \gamma(]a,x]\times]a,y[) \le \mu(]a,x])$$

so that the first inequality is proved; the proof of the other one is similar. Thus (4.7) holds and the claimed inclusion $\operatorname{spt}\gamma \subset \Gamma(\bar f)$ follows immediately.

Now, set $y_x = \bar f(x^-)$, $z_x = \bar f(x^+)$ and define the at most countable set $A = \{x : z_x > y_x\}$; finally, let B be the union of all open intervals $]y_x, z_x[$, as x varies in A, and let $g : B \longrightarrow C$ be the function whose value is x in $]y_x, z_x[$. We can then conclude the uniqueness from the explicit formula

$$\begin{aligned}\gamma = (Id,\bar f)_{\#}\mu \llcorner (C\setminus A) + (g, Id)_{\#}\nu \llcorner B + \sum_{x\in A} \delta_x\otimes\delta_{y_x}\big(\nu(]a,y_x]) - \mu(]a,x[)\big)\\ + \delta_x\otimes\delta_{z_x}\big(\mu(]a,x]) - \nu(]a,z_x[)\big),\end{aligned}$$

simply noticing that all the mass in $x \in C\setminus A$ must be moved in $\bar f(x)$ and all the mass in $y \in B$ must come from $g(y)$ –which explains the first two terms– and recalling the balance of the mass following from the fact that γ is a non decreasing transport plan –which explains the last two terms.

Existence: Let us first assume that μ and ν have finite first moments and let us notice that problem (4.6) is simply the transport problem with the l.s.c. cost $\tilde c(x,y) = \alpha(|y-x|)$: thus, thanks to Theorem 3.1, there exist a solution γ of (4.6) which is supported on a $\tilde c$–monotone set. Note now that Definition 4.4 assures that any plan is a non decreasing transport plan unless there exist (x_1,y_1) and (x_2,y_2) in its support with $x_1 < x_2$ and $y_1 > y_2$. Since this cannot happen for γ due to the $\tilde c$–monotonicity of its support, we infer that γ is non decreasing. The uniqueness implies that γ solves (4.6) for any strictly convex and nondecreasing function α, thus by approximation we infer that γ solves also (4.5), and then the existence is proved.

To show the existence of non decreasing plans also for general probability measures μ and ν, it suffices now to take probability measures μ_n and ν_n with compact support (and then with finite first order moments) weakly converging to μ and ν. There exist then non decreasing transport plans γ_n between μ_n and ν_n, and finally the fact that any limit point γ of γ_n as $n\to\infty$ is a non decreasing transport plan between μ and ν is easy to check. □

Lemma 4.6. *Let γ be a solution of (4.3). Then γ is concentrated on a Borel set Γ such that, for any pair of points $(x,y) \in \Gamma \setminus \Delta$ and $(\tilde{x},\tilde{y}) \in \Gamma$, the segment $[\![\tilde{x},\tilde{y}]\!]$ is not contained in $]\!]x,y[\![$.*

Proof. Let us define $\Delta' := \{(x,x)\ s.t.\ x \in F_\Gamma \setminus T_\Gamma\}$. Then, possibly replacing γ by $\gamma \llcorner (\mathbb{R}^n \times \mathbb{R}^n \setminus \Delta')$ (still a solution of (4.3) relatively to its marginals) we can assume with no loss of generality that $\gamma \llcorner \Delta' = 0$. According to Theorem 3.1, we can find a $\beta-$cyclically monotone Borel set $G \subset \mathbb{R}^n \times \mathbb{R}^n \setminus \Delta'$ on which β is finite and γ is concentrated. This implies, by Lemma 4.1, that distinct transport rays induced by G do not intersect, unless they are parallel.

We denote by Z the space of open oriented segments in $\mathbb{R}^n$ (see [3] for a precise definition of the metric of this space) and we consider the map $\pi : G \to Z$ which associates to any pair (x,y) the maximal transport ray of G containing $]\!]x,y[\![$; this map is well defined by the above mentioned property of transport rays and one can show by standard but not completely trivial arguments, that π is a Borel map (see Lemma 6.3 of [3]).

Let $\gamma = \gamma_C \otimes \eta$, with $\eta = \pi_\#\gamma$, be the disintegration of γ induced by π and let μ_C, ν_C be respectively the first and second marginals of γ_C. By Proposition 4.5 we can consider the "non decreasing rearrangements" $\hat{\gamma}_C$ of γ_C, which are the unique solutions of problem (4.3) with marginals μ_C, ν_C. By Lemma 4.3 we infer that $\gamma_C = \hat{\gamma}_C$ for $\eta-$a.e. C, i.e. γ_C is non decreasing for $\eta-$a.e. C.

Let now ρ_n be a sequence of mollifiers, let $Z' \subset \pi(G)$ be a Borel set of full $\eta-$measure such that γ_C is non decreasing for any $C \in Z'$ and define

$$L := \left\{ ((x,y),C) : \{x,y\} \subset C \in Z' \quad \text{and} \quad \iint \rho_n(x'-x,y'-y)\, d\gamma_C(x',y') > 0 \ \ \forall n \in \mathbb{N} \right\}.$$

Since the functions $((x,y),C) \mapsto \int \rho_n(x'-x,y'-y)\, d\gamma_C(x',y')$ are continuous in (x,y) and Borel in C, a simple approximation argument shows that the functions are Borel measurable with respect to $((x,y),C)$, and therefore L is a Borel set. The projection theorem (see [9]) then gives that the projection of L on $\mathbb{R}^n \times \mathbb{R}^n$, i.e. the set

$$L' = \bigcup_{C \in Z'} (C \times C) \cap \operatorname{supp} \gamma_C$$

is $\eta-$measurable (actually a universally measurable set). The set L' has clearly the property stated in the theorem, as all supports of γ_C are contained in the extended graphs of non decreasing functions and $C \cap C' = \emptyset$ whenever C, $C' \in Z'$ and $C \neq C'$. Therefore, since $\gamma(L') = 1$ we need only to find a Borel set $\Gamma \subset L'$ with $\gamma(\Gamma) = 1$. □

5. The main result

We consider a "crystalline" norm $\|\cdot\|$ in $\mathbb{R}^n$, defined by

$$\|w\| := \max\{|v_i \cdot w| \ : \ i = 1, \ldots, N\}$$

for suitable vectors $v_1, \ldots, v_N$ in $\mathbb{R}^n$ spanning the whole space (we keep the notation $|\cdot|$ for the Euclidean norm).

We will consider in the following the optimal transport problem with $c(x,y) = \|y-x\|$ and the associated problems (4.2) and (4.3).

Theorem B. *Assume that μ, ν have finite first moments and $\mu \ll \mathscr{L}^n$. Then problem (4.3) has a unique solution, and this solution is induced by a transport. In particular there is an optimal transport, relative to the cost c, between μ and ν.*

Proof. The proof will be split in 5 steps. We fix a solution γ of (4.3) and a Borel set Γ on which γ is concentrated which is $\beta-$cyclically monotone and has the property stated in Lemma 4.6. Without loss of generality we can assume that Γ is $\sigma-$compact. First we introduce some notation.

Set $v_{-i} = -v_i$ for $i = 1, \dots, N$ and call "multiindex" any unordered subset of $\{\pm 1, \dots, \pm N\}$. Given a multiindex $I = \{i_1, \dots, i_m\}$, we will say that the couple $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ *is of the type* I if

$$c(x,y) = v_{i_1} \cdot (y-x) = \cdots = v_{i_m} \cdot (y-x) > v_j \cdot (y-x) \quad \forall j \notin I.$$

By definition, each couple (x,y) with $x \neq y$ is of a unique type, so that the types induce a partition of $\big(\mathbb{R}^n \times \mathbb{R}^n\big) \setminus \Delta$; we set then

$$\Gamma_I := \Gamma \cap \big\{(x,y) \in \big(\mathbb{R}^n \times \mathbb{R}^n\big) \setminus \Delta : \, (x,y) \text{ is of the type } I\big\}$$

and notice also that all sets Γ_I are $\sigma-$compact. We call *simple types* the ones corresponding to a multiindex with only an index. Notice that if (x,y) is of the simple type $I = \{i\}$, the same is true for a neighborhood of (x,y). Indeed, for each $j \neq i$ it holds $v_i \cdot (y-x) > v_j \cdot (y-x)$, and then the same inequalities continue to hold in a neighborhood of (x,y).

For each multiindex I we define the transport plan

$$\gamma_I := \gamma \llcorner \{(x,y) : \, (x,y) \text{ is of the type } I\},$$

writing also $\mu_I = \pi_{1\#}\gamma_I$ and $\nu_I = \pi_{2\#}\gamma_I$. In view of Lemma 4.2, it turns out that each γ_I is an optimal transport plan for the problem given by its marginals and, of course, γ_I is concentrated on the set Γ_I.

Step 1. In the first two steps we will prove that γ_I is induced by a transport for each simple type $I = \{i\}$; we fix I and set for simplicity $\gamma_i = \gamma_I$ and $\Gamma_i = \Gamma_I$. In this step we show that γ_i is locally a transport: given $(x,y) \in \Gamma_i$ there exists an open neighborhood U such that $\gamma_i \llcorner U$ is induced by a transport. Indeed, let $U = Z \times W$ be a product of two open balls such that only points of type I are contained in U, let η be a transport plan between $\pi_{1\#}\gamma_i \llcorner U$ and $\pi_{2\#}\gamma_i \llcorner U$. Clearly η is concentrated on U and

$$\begin{aligned} C(\eta) &= \iint_U c(z,w)\, d\eta(z,w) = \iint_U v_i \cdot (w-z)\, d\eta(z,w) \\ &= \int_W v_i \cdot w \, d\pi_{2\#}\eta - \int_Z v_i \cdot z \, d\pi_{1\#}\eta = \int_W v_i \cdot w \, d\pi_{2\#}\gamma_i \llcorner U - \int_Z v_i \cdot z \, d\pi_{1\#}\gamma_i \llcorner U. \end{aligned}$$

This shows that the cost is independent of η and then that each plan of transport between $\pi_{1\#}\gamma_i \llcorner U$ and $\pi_{2\#}\gamma_i \llcorner U$ is optimal. Since, as noted in Lemma 4.2, solving (4.3) is a local property, it follows that $\gamma \llcorner U$ is a transport plan that, among all the plans connecting $\pi_{1\#}\gamma_i \llcorner U$ and $\pi_{2\#}\gamma_i \llcorner U$, minimizes the Euclidean cost and, among all the plans η minimizing the Euclidean cost, minimizes the quantity

$$\iint_{\mathbb{R}^n \times \mathbb{R}^n} \alpha(|y-x|)\, d\eta(x,y).$$

By Theorem 3.2 we infer that $\gamma_i \llcorner U$ is induced by a transport.

Step 2. Now we claim that γ_i is induced by a transport defined on T_{Γ_i} and that the family of extreme points of Γ_i is Lebesgue negligible. By $\sigma-$compactness we can cover Γ_i with countably many products $U_j = Z_j \times W_j$ such that $\gamma_i \llcorner U_j$ is induced by a transport t_j; we also assume that the family U_j is so large that for any $(x,y) \in \Gamma_i$ and any $\varepsilon > 0$ we can find j such that $(x,y) \in U_j$ and the diameter of W_j is less than ε. By the second part of the statement of Theorem 3.2 the extreme points Θ_j of $\Gamma_i \cap U_j$ are Lebesgue negligible.

We define Θ as the union of all sets Θ_j. Since Θ is $\mathscr{L}^n-$negligible and then $\mu-$negligible, the proof of the claim will be achieved simply gluing all transport maps t_j if we show that for any $x \in \pi_1(\Gamma_i) \setminus \Theta$ there is only one y such that $(x,y) \in \Gamma_i$.

Assume that $y_1 \neq y_2$, that (x, y_1) and (x, y_2) both belong to Γ_i and that $|y_1 - x| \leq |y_2 - x|$ (this is not restrictive, up to a permutation of the two points). We can choose j such that $(x, y_2) \in U_j$ and the diameter of W_j is less than $|y_1 - y_2|$, thus $(x, y_1) \notin U_j$. Since $x \notin \Theta_j$ we can find $(\tilde{x}, \tilde{y}) \in \Gamma_i \cap U_j$ such that the segment $]\tilde{x}, \tilde{y}[$ contains x in its interior. Lemma 4.1 then gives that both $]x, y_1[$ and $]x, y_2[$ are parallel to $]\tilde{x}, \tilde{y}[$ and therefore are parallel.

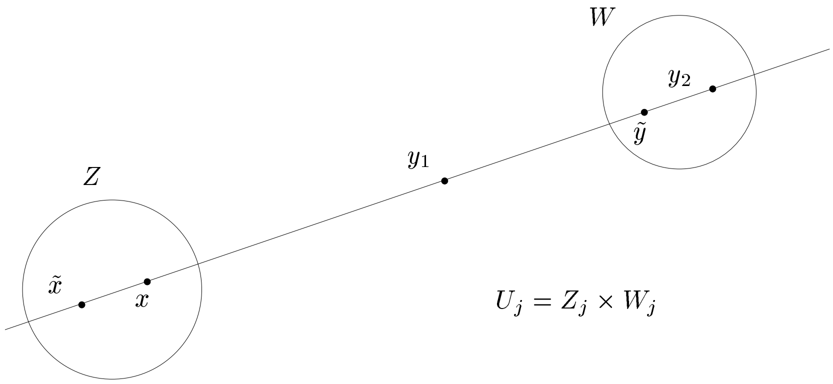

FIGURE 3. Geometric picture for Step 2

So we have only to exclude the case when x, y_1 and y_2 are aligned. But then, see Figure 3, our choices imply that also $y_1 \in]\tilde{x}, \tilde{y}[$: the contradiction is obtained by applying Lemma 4.6 to the points $(\tilde{x}, \tilde{y})$ and (x, y_1). This proves that γ_i is induced by a transport.

Let now x be an extreme point of Γ_i. Since $x \in \pi_1(\Gamma_i \cap U_j)$ for some j, it follows that $x \in \Theta_j$, otherwise we would have been able to find $(\tilde{x}, \tilde{y}) \in \Gamma_i \cap U_j$ such that $]\tilde{x}, \tilde{y}[$ contains x in its interior. This shows that the family of extreme points of Γ_i is contained in the union of all sets Θ_j and so is Lebesgue negligible.

Step 3. In this step we show that that γ_I is induced by a transport defined on T_{Γ_I} and that the extreme points of Γ_I are Lebesgue negligible even when I is not of simple type. If $\mu_I \neq 0$, let

$$P_I := \{x : x \cdot (v_i - v_j) = 0 \quad \forall i, j \in I\}, \tag{5.1}$$

$m := \dim(P_I) \geq 1$ and $K := P_I^{\perp}$. For $\omega \in K$ we set $\mathrm{P}_I^{\omega} = P_I + \omega$. We want to project the transport plan γ_I on the hyperplanes P_I^{ω}: to this aim, denoting by Proj_K the orthogonal projection on K, let us first define $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to K$ as

$$\varphi(x, y) := \mathrm{Proj}_K(x).$$

Then, consider the disintegration of γ_I with respect to φ: it associates to γ_I the family of measures γ_ω, concentrated on the set $\{(x, y) \in \Gamma_I : \varphi(x, y) = \omega\}$. But $\varphi(x, y) = \omega$ means $\mathrm{Proj}_K(x) = \omega$ or, in other words, $x \in \mathrm{P}_I^{\omega}$; on the other hand, since (x, y) is of the type I it is clear from (5.1) that $y - x \in P_I$ and then also $y \in \mathrm{P}_I^{\omega}$. So we proved that γ_ω is supported on $\mathrm{P}_I^{\omega} \times \mathrm{P}_I^{\omega}$, and then we can consider γ_ω as probability measures in $\mathrm{P}_I^{\omega} \times \mathrm{P}_I^{\omega}$.

Lemma 4.3 gives that for $\varphi_\# \gamma_I$–a.e. ω the transport plan γ_ω is a solution of the problem (4.3) considered on the plane P_I^{ω} and with marginals $\pi_{1\#}\gamma_\omega$ and $\pi_{2\#}\gamma_\omega$; moreover $\pi_{1\#}\gamma_\omega = \mu_\omega$, the disintegration of μ_I with respect to Proj_K, and then for $\mathrm{Proj}_{K\#}\mu_I$–a.e. ω (or, equivalently, $\varphi_\# \gamma_I$–a.e. ω) it is absolutely continuous with respect to the Lebesgue measure $\mathscr{L}^m$ on P_I^{ω} (recall that obviously the decomposition $\mathscr{L}^n = \mathscr{L}^n_\omega \otimes \mathrm{Proj}_{K\#}\mathscr{L}^n$ of $\mathscr{L}^n$ with respect to Proj_K is made with each $\mathscr{L}^n_\omega$ being the Lebesgue measure $\mathscr{L}^m$ on P_I^{ω}). Observe now that the distance on P_I^{ω} is induced by the crystalline norm associated to the projections $\tilde{v}_i$ of the vectors v_i on P_I: but the vectors $\tilde{v}_i$ with $i \in I$ are all equal, and then we call $\tilde{v}_I$ that unique vector; in other words, the crystalline norm on P_I^ω is induced by the vectors $\{\tilde{v}_j : j \notin I\} \cup \{\tilde{v}_I\}$, so the multiindex I has been contracted to a single index. Moreover each (x,y) in $\Gamma_I \cap (\mathrm{P}_I^\omega \times \mathrm{P}_I^\omega)$ is, on the plane P_I^ω, of type $\{I\}$, which is now a simple type: then we infer that γ_ω is induced by a transport from μ_ω to $\pi_{2\#}\gamma_\omega$ for $\mathscr{L}^{n-m}$–a.e. ω using the argument of Step 2 on P_I^ω. Since the different planes are disjoint, it follows that the whole transport plan γ_I is in fact induced by a transport.

Let E be the family of extreme points of Γ_I. Notice that for $\mathscr{L}^{n-m}$–a.e. ω the family E_ω of extreme points of $\Gamma_I \cap (\mathrm{P}_I^\omega \times \mathrm{P}_I^\omega)$ is Lebesgue negligible by Step 2. Since $E \cap \mathrm{P}_I^\omega = E_\omega$, Fubini's theorem gives that E is Lebesgue negligible.

Step 4. In this step we show that γ is induced by a transport defined on $T_\Gamma \cup F_\Gamma$. Assume first that the intersection of Γ with the diagonal Δ of $\mathbb{R}^n \times \mathbb{R}^n$ is empty, so that $F_\Gamma = \emptyset$ and $\gamma = \sum_I \gamma_I$. We will obtain a transport by gluing the different "partial" transports $t_I : T_{\Gamma_I} \to \mathbb{R}^n$, given by Step 3, such that $\gamma_I = (Id, t_I)_\# \mu_I$. Since $(x, t_I(x)) \in \Gamma_I$ for μ_I–a.e. $x \in \mathbb{R}^n$ (because γ_I is concentrated on Γ_I), the gluing construction would work if we knew that the set L of points x such that there exist $J \neq K$ and y_J, y_K such that $(x, y_J) \in \Gamma_J$ and $(x, y_K) \in \Gamma_K$ is μ–negligible. We know from Step 3 that the union Θ of all extreme points relative to the sets Γ_I is μ–negligible. If $x \in L \setminus \Theta$ and J, K, y_J, y_K are as above we obtain from Lemma 4.1 that the segments $]x, y_J[$ and $]x, y_K[$ are parallel. This implies that (x, y_J) and (x, y_K) are of the same type and gives a contradiction. Hence $L \subset \Theta$ and $\mu(L) = 0$.

In the general case we set $\Gamma' = \Gamma \setminus \Delta$. By applying the previous argument to $\gamma' := \gamma \llcorner \Gamma'$ we can show that γ' is induced by a transport t defined on $T_{\Gamma'}$. But since $T_\Gamma = T_{\Gamma'}$ and (by Lemma 4.6) $F_\Gamma \cap T_\Gamma = \emptyset$, it suffices to extend t to F_Γ setting $t = Id$ to obtain $\gamma = (Id \times t)_\# \mu$.

Step 5. Here we show the uniqueness of γ. Let us take two solutions γ_1 and γ_2, respectively induced by optimal transports t_1 and t_2, and define

$$\gamma := \frac{\gamma_1 + \gamma_2}{2}.$$

Clearly also γ is a solution of (4.3), and so it is induced by a transport. But since the disintegration of γ with respect to x is $(\delta_{t_1(x)} + \delta_{t_2(x)})/2$, this can happen only if $t_1 = t_2$ μ–a.e. in $\mathbb{R}^n$. □

As we mentioned in the introduction, the existence and uniqueness result for problem (4.3) can also be interpreted as follows: the optimal transport maps t_ε relative to the strictly convex costs c_ε in (1.4) converge in measure as $\varepsilon \downarrow 0$ to the map t of Theorem B. Indeed, arguing as in [3] one can show that any limit point of $\gamma_\varepsilon = (Id, t_\varepsilon)_\# \mu$ solves (4.3), therefore $(Id, t_\varepsilon)_\# \mu$ weakly converge to $(Id, t)_\# \mu$. This easily implies, arguing as in [3], that the corresponding functions converge in measure.

6. Existence in the two-dimensional case

In this section we show that in the two dimensional case the Sudakov method, coupled with the secondary variational problem (4.2), gives existence of optimal transport maps for cost functions $c(x,y)$ induced by an arbitrary norm $\|\cdot\|$ in $\mathbb{R}^2$:

Theorem C. *Let μ, ν be probability measures in $\mathbb{R}^2$ with finite first order moments and $\mu \ll \mathscr{L}^2$ and let c as above. Then there exists an optimal transport map t mapping μ to ν.*

The proof of Theorem C is based on the following two results. The first one is an "abstract" existence result on optimal transport maps based on Sudakov decomposition argument (see Theorem 6.4 of [3]). The result is valid in any dimension under some regularity conditions on the decomposition in transport rays induced by Γ.

Theorem 6.1. *Let μ, ν be probability measures in $\mathbb{R}^n$ with finite first order moments and let $\gamma \in \Pi(\mu,\nu)$ be concentrated on a $\sigma-$compact set Γ. Assume that*

(i) *for any pair of points (x,y), $(\tilde{x},\tilde{y}) \in \Gamma \setminus \Delta$ the condition $]\!]x,y[\![\cap [\![\tilde{x},\tilde{y}]\!] \neq \emptyset$ implies that the two segments are parallel;*
(ii) *μ is absolutely continuous with respect to $\mathscr{L}^n$;*
(iii) *the family of extreme points of Γ is $\mu-$negligible;*
(iv) *there exist a $\mu-$negligible set $N \subset T_\Gamma$ and an increasing sequence of compact sets K_h such that the orientation $\tau_\Gamma|_{K_h}$ is a Lipschitz map and the union of K_h is $T_\Gamma \setminus N$.*

Then there exists a Borel map $t : \mathbb{R}^n \to \mathbb{R}^n$ such that $t_\#\mu = \nu$, and $]\!]x,t(x)[\![$ is contained in a maximal transport ray of Γ for $\mu-$a.e. x such that $x \neq t(x)$.

The second result we need to prove Theorem C is specific of dimension 2: conditions (iii) and (iv) in Theorem 6.1 are implied by (i) and (ii).

Lemma 6.2. *Let $\Gamma \subset \mathbb{R}^2 \times \mathbb{R}^2$ be a $\sigma-$compact set satisfying condition (i) in Theorem 6.1. Then the family of extreme points of Γ is $\mathscr{L}^2-$negligible and we can cover T_Γ with a sequence of compact sets K_m such that $\tau_\Gamma|_{K_m}$ is a Lipschitz map for any m.*

Proof. Let E be the family of extreme points of Γ, and assume by contradiction that $\mathscr{L}^2(E) > 0$. Given $h > 0$ and a unit vector p, we consider the family $\Gamma' \subset \Gamma$ made by all points (x,y) with $|y-x| \geq h$ and such that the segment $[\![x,y]\!]$ makes an angle with the direction p that belongs to the interval $[-\pi/4,\pi/4]$. Letting h, p vary in countable subsets, we have that the collection F of extreme points of Γ' satisfies $\mathscr{L}^2(F) > 0$ for suitable values of the parameters h, p. In the following we assume (up to a rigid motion) that $p = e_1$ is the horizontal unit vector.

Let F_1 be the set of all points of F where F has density 1. By Fubini theorem there exists $b \in \mathbb{R}$ such that the set

$$S := \{t \in \mathbb{R} : \ (b,t) \in F_1\}$$

has strictly positive Lebesgue measure. There is a well defined angle function $\theta(t)$, with values into $[-\pi/4,\pi/4]$, which associates to any $t \in S$ the direction of a transport ray $((b,t),y) \in \Gamma'$ (in case there is more than one, we simply choose one).

We claim that $t \mapsto \tan\theta(t) + \sqrt{2}t/h$ is non-decreasing. Indeed, if t_1, $t_2 \in S$ and $t_1 < t_2$ we have

$$t_1 + \delta\tan\theta(t_1) < t_2 + \delta\tan\theta(t_2) \qquad \forall \delta \in [0, h/\sqrt{2})$$

because the two segments emanating from $x_1 = (t_1, b)$, $x_2 = (t_2, b)$, with length at least h, do not intersect in their interior. Choosing $\delta = h/\sqrt{2}$ yields the claim.

As a consequence of the previous claim, the derivative of $\tan\theta$ is finite $\mathscr{L}^1-$a.e. on S and we can find a compact set $S' \subset S$ with $\mathscr{L}^1(S') > 0$ such that $\tan\theta|_{S'}$ is a Lipschitz function. Finally we choose a point $t \in S'$ where the density of S' is 1 and we will achieve a contradiction by showing that F has not density 1 at $x = (b,t)$.

For $\delta > 0$, let $\Phi_\delta : S' \to \{x \in \mathbb{R}^2 : \ x_1 = b + \delta\}$ be the map defined by

$$\Phi_\delta(t) := (b+\delta, t + \delta\tan\theta(t)) .$$

Notice that Φ_δ is one to one and that the image of Φ_δ is disjoint with F for $\delta < h/\sqrt{2}$. Moreover, for $\delta \leq \delta_0$ (with $\delta_0 > 0$ depending only on the Lipschitz constant of $\tan\theta$) we have $|\Phi'_\delta| \geq 1/2$ on S', therefore

$$\mathscr{H}^1\Big(\Phi_\delta\big(t-r,t+r\big)\Big) \geq \frac{1}{2}\mathscr{L}^1\Big(S' \cap \big(t-r,t+r\big)\Big) \geq \frac{r}{2}$$

for r sufficiently small and independent on $\delta \leq \delta_0$. Moreover, as $|\tan\theta| \leq 1$, we have

$$\Phi_\delta\big(t-r,t+r\big) \subset \{b+\delta\} \times \big(t-2r,t+2r\big)$$

for $\delta \le r$. In particular

$$\mathscr{L}^2\Big(\big(b,b+r\big)\times\big(t-2r,t+2r\big)\setminus F\Big)\ge\int_0^r\mathscr{H}^1\Big(\Phi_\delta\big(t-r,t+r\big)\Big)\,d\delta\ge\frac{r^2}{2}$$

for r sufficiently small, contradicting the fact that $(b,t)\in F_1$.

The proof of the second part of the statement follows by elementary geometric considerations, writing $\Gamma\setminus\Delta$ as the union of compact sets Γ_m such that

$$a_m := \min_{(x,y)\in\Gamma_m}|y-x|>0 \qquad \forall m\in\mathbb{N}$$

and considering the compact sets K_m made by all points $z\in T_{\Gamma_m}$ contained in a ray $]\!]x,y[\![$ of Γ_m with $\min\{|x-z|,|y-z|\}\ge a_m/m$. □

Now we can proceed with the proof of Theorem C. We consider a Kantorovich potential φ, the function β defined in (4.1) and a solution $\gamma\in\Pi'(\mu,\nu)$ of problem (4.2). According to Theorem 3.1 and Lemma 4.1 we can find a $\sigma-$compact set Γ on which γ is concentrated, condition (i) in Theorem 6.1 is fulfilled and

$$\|y-x\|=\varphi(y)-\varphi(x) \qquad \forall (x,y)\in\Gamma. \tag{6.1}$$

By Lemma 6.2 conditions (iii), (iv) are fulfilled as well, therefore there exists a Borel map $t:\mathbb{R}^2\to\mathbb{R}^2$ such that $t_\#\mu=\nu$, and $]\!]x,t(x)[\![$ is contained in a maximal transport ray of Γ for $\mu-$a.e. x such that $t(x)\neq x$. It is immediate to check that (6.1) implies $\|y-x\|=\varphi(y)-\varphi(x)$ for any pair of points (x,y) in a maximal transport ray of Γ with $x\le y$, therefore

$$\|x-t(x)\|=\varphi(t(x))-\varphi(x) \qquad \text{for } \mu-\text{a.e. } x.$$

This proves that t is an optimal transport map relative to the cost $c(x,y)=\|y-x\|$.

References

[1] L. Ambrosio: *Lecture Notes on Optimal Transport Problems.* Notes of a CIME Course in Madeira, Portugal (2000), to be published in the CIME Springer Lecture Notes.

[2] G.Alberti, B.Kirchheim & D.Preiss: Personal communication.

[3] L.Ambrosio & A.Pratelli: *Existence and stability results in the L^1 theory of optimal transportation.* Notes of a CIME Course in Martina Franca, Italy (2001), to be published in the CIME Springer Lecture Notes.

[4] G.Anzellotti & S.Baldo: *Asymptotic development by $\Gamma-$convergence.* Appl. Math. Optim., **27** (1993), 105–123.

[5] Y.Brenier: *Décomposition polaire et réarrangement monotone des champs de vecteurs.* C.R. Acad. Sci. Paris, Sér I Math., **305** (1987), 805–808.

[6] Y.Brenier: *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math., **44** (1991), 375–417.

[7] L.Caffarelli: *Boundary regularity of maps with a convex potential.* Commun. Pure Appl. Math., **45** (1992), 1141–1151.

[8] L.Caffarelli, M.Feldman & R.J.McCann: *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs.* J. Amer. Math. Soc., **15** (2002), 1–26.

[9] C.Castaing & M.Valadier: *Convex analysis and measurable multifunctions.* Lecture Notes in Mathematics **580**, Springer, 1977.

[10] G.Dal Maso: *An Introduction to $\Gamma-$Convergence.* Birkhäuser, 1993.

[11] R.O. Davies: *On accessibility of plane sets and differentiation of functions of two real variables.* Proc. Cambridge Philos. Soc., **48** (1952), 215–232.

[12] C.Dellacherie & P.Meyer: *Probabilities and potential.* Mathematical Studies **29**, North Holland, 1978.

[13] L.De Pascale & A.Pratelli: *Regularity properties for Monge transport density and for solutions of some shape optimization problem.* Calc. Var., **14** (2002), 249–274.

[14] L.De Pascale, L.C. Evans & A.Pratelli: *Integral Estimates for Transport Densities.* preprint (Pisa, 2002).

[15] H.Dietrich: *Zur $c-$konvexität und $c-$Subdifferenzierbarkeit von Funktionalen.* Optimization, **19** (1988), 355–371.

[16] L.C.Evans: *Partial Differential Equations and Monge–Kantorovich Mass Transfer.* Current Developments in Mathematics, 1997, 65–126.
[17] L.C.Evans & W.Gangbo: *Differential Equation Methods for the Monge-Kantorovich Mass Transfer Problem.* Memoirs AMS, **653**, 1999.
[18] M.Feldman & R.McCann: *Uniqueness and transport density in Monge's mass transportation problem.* Calc. Var., **15** (2002), 81–113.
[19] W.Gangbo: *An elementary proof of the polar factorization theorem for functions.* Arch. Rat. Mech. Anal., **128** (1994), 381–399.
[20] W.Gangbo & R.J.McCann: *The geometry of optimal transportation.* Acta Math., **177** (1996), 113–161.
[21] L.V.Kantorovich: *On the transfer of masses.* Dokl. Akad. Nauk. SSSR, **37** (1942), 227–229.
[22] L.V.Kantorovich: *On a problem of Monge.* Uspekhi Mat. Nauk., **3** (1948), 225–226.
[23] D.G.Larman: *A compact set of disjoint line segments in* $\mathbf{R}^3$ *whose end set has positive measure.* Mathematatika, **18** (1971), 112–125.
[24] R.McCann: *Polar factorization of maps on Riemannian manifolds.* Geom. Funct. Anal., **11** (2001), 589–608.
[25] G.Monge: *Memoire sur la Theorie des Déblais et des Remblais.* Histoire de l'Acad. des Sciences de Paris, 1781.
[26] S.T.Rachev & L.Rüschendorf: *Mass transportation problems.* Vol I: Theory, Vol. II: Applications. Probability and its applications, Springer, 1998.
[27] L.Rüschendorf: *On* $c-$*optimal random variables.* Statistics & Probability Letters, **27** (1996), 267–270.
[28] C.Smith & M.Knott: *On the optimal transportation of distributions.* J. Optim. Theory Appl., **52** (1987), 323–329.
[29] C.Smith & M.Knott: *On Hoeffding–Fréchet bounds and cyclic monotone relations.* J. Multivariate Anal., **40** (1992), 328–334.
[30] V.N.Sudakov: *Geometric problems in the theory of infinite dimensional distributions.* Proc. Steklov Inst. Math., **141** (1979), 1–178.
[31] N.S.Trudinger & X.J.Wang: *On the Monge mass transfer problem.* Calc. Var. PDE, **13** (2001), 19–31.
[32] J.Urbas: *Mass transfer problems.* 1998.
[33] C.Villani: *Topics in mass transportation.* Forthcoming book by AMS.

(L. Ambrosio) Scuola Normale Superiore, Piazza dei Cavalieri, 56126 Pisa, Italy
E-mail address: `l.ambrosio@sns.it`

(B. Kirchheim) Mathematical Institute, University of Oxford, 24-29 St. Giles Oxford OX1 3LB
E-mail address: `kirchhei@maths.ox.ac.uk`

(A. Pratelli) Scuola Normale Superiore, Piazza dei Cavalieri, 56126 Pisa, Italy
E-mail address: `a.pratelli@sns.it`