Existence and stability results in the L^1 theory of optimal transportation

Luigi Ambrosio[1] and Aldo Pratelli[2]

[1] Scuola Normale Superiore
Piazza dei Cavalieri, 56100 Pisa, Italy
luigi@ambrosio.sns.it

[2] Scuola Normale Superiore
Piazza dei Cavalieri, 56100 Pisa Italy
a.pratelli@sns.it

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. Denoting by h_0, $h_1 : \mathbf{R}^2 \to [0, +\infty)$ the Borel functions describing the initial and final distribution of matter, there is obviously a compatibility condition, that the total mass is the same:

$$\int_{\mathbb{R}^2} h_0(x)\,dx = \int_{\mathbb{R}^2} h_1(y)\,dy. \tag{1}$$

Assuming with no loss of generality that the total mass is 1, we say that a Borel map $t : \mathbf{R}^2 \to \mathbf{R}^2$ is a *transport* if a local version of the balance of mass condition holds, namely

$$\int_{t^{-1}(E)} h_0(x)\,dx = \int_E h_1(y)\,dy \qquad \text{for any } E \subset \mathbb{R}^2 \text{ Borel.} \tag{2}$$

Then, the Monge problem consists in minimizing the work of transportation in the class of transports, i.e.

$$\min\left\{\int_{\mathbb{R}^2} |t(x) - x| h_0(x)\,dx : \ t \text{ transport}\right\}. \tag{3}$$

The Monge transport problem can be easily generalized in many directions, and all these generalizations have proved to be quite useful:

- General measurable spaces X, Y, with measurable maps $t : X \to Y$;

L. Ambrosio et al.: LNM 1813, L.A. Caffarelli and S. Salsa (Eds.), pp. 123–160, 2003.

• General probability measures μ in X and ν in Y. In this case the local balance of mass condition (2) reads as follows:

$$\nu(E) = \mu(t^{-1}(E)) \qquad \text{for any } E \subset Y \text{ measurable.} \tag{4}$$

This means that the push-forward operator $t_\#$ induced by t, mapping probability measures in X into probability measures in Y, maps μ into ν.
• General cost functions: a measurable map $c : X \times Y \to [0, +\infty]$. In this case the cost to be minimized is

$$W(t) := \int_X c(x, t(x))\, d\mu(x).$$

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 [3], [23], [36], [43], [44] 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 [31], [32] a notion of weak solution of the transport problem. He suggested to look for *plannings* instead of transports, i.e. probability measures γ in $X \times Y$ whose marginals are μ and ν. Formally this means that $\pi_{X\#}\gamma = \mu$ and $\pi_{Y\#}\gamma = \nu$, where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are the canonical projections. Denoting by $\Pi(\mu, \nu)$ the class of plannings, he wrote the following minimization problem

$$\min \left\{ \int_{X \times Y} c(x, y)\, d\gamma : \ \gamma \in \Pi(\mu, \nu) \right\}. \tag{5}$$

Notice that $\Pi(\mu, \nu)$ is not empty, as the product $\mu \otimes \nu$ has μ and ν as marginals. Due to the convexity 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 (5): this happens for instance whenever X and Y are Polish spaces and c is lower semicontinuous (see for instance [36]).

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 planning γ, defined by $(Id \times t)_\#\mu$. This planning is concentrated on the graph of t in $X \times Y$ 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, it turns out that

$$\inf(3) \geq \min(5).$$

Moreover, by approximating any planning by plannings induced by transports, it can be shown that equality holds under fairly general assumptions (see for instance [3]). 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 the Appendix) provides a very useful representation of plannings, and more generally of probability measures γ in $X \times Y$ whose first marginal is μ: there exist probability measures γ_x in Y such that $\gamma = \gamma_x \otimes \mu$, i.e.

$$\int_{X\times Y} \varphi(x,y)\, d\gamma(x,y) = \int_X \left(\int_Y \varphi(x,y)\, d\gamma_x(y) \right) d\mu(x).$$

for any bounded measurable function φ. In this sense we can consider a planning 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 μ-a.e. $x \in X$. This representation of plannings also shows the close connection between the ideas of Kantorovich and of L.C. Young, who developed in the same years his theory of generalized controls (see [45, 46, 47]).

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 [36], [24], [39], [40], [44]). 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: for instance the measure $\gamma := \gamma_x \otimes \mathcal{L}^1 \llcorner [0,1]$ with $\gamma_x := \frac{1}{2}(\delta_x + \delta_{2-x})$ is not induced by a transport $y = t(x)$ but it is extremal in $\Pi(\mu,\nu)$ with $\nu := \frac{1}{2}\mathcal{L}^1 \llcorner [0,2]$, due to the fact that it is induced by the transport $x = t(y) = |y-1|$. It turns out that the existence of optimal transport maps depends not only on the geometry of $\Pi(\mu,\nu)$, but also on the choice of the cost function c.

When $X = Y = \mathbb{R}^n$ and $c(x,y) = h(x-y)$ with h *strictly* convex and μ absolutely continuous with respect to $\mathcal{L}^n$, the duality methods yield that any optimal planning is induced by a transport; as a consequence, the optimal transport map exists and is unique (see [12], [13], [14], [24], [39], [40]).

The same results can be shown directly making a first variation in the dual formulation, bypassing the Kantorovich formulation (see [29], [15]). See also [34] for the extension of these results to a Riemannian setting.

The case when h is not strictly convex, corresponding to the original problem raised by Monge, is more subtle. Indeed, if $c(x,y)=|x-y|$ (the euclidean distance) then the standard duality methods provide information on the direction of transportation but no information on the distance $|x-y|$, at least when μ is absolutely continuous with respect to Lebesgue measure: to be precise one can show the existence of a 1-Lipschitz map $u:\mathbb{R}^n\to\mathbb{R}$ such that

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

for μ-a.e. x. A similar result holds if the Euclidean norm is replaced by any strictly convex norm. Moreover, when $c(x,y)=\|x-y\|$ with $\|\cdot\|$ *not* strictly convex, then we have an even more dramatic loss of information about the location of y:

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

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 [41], 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, where he states that, if $\mu\ll\mathcal{L}^n$, then the conditional measures induced by the decomposition are absolutely continuous with respect to the Lebesgue measure (of the correct dimension). However, it turns out that this property is not true in general even for the simplest decomposition, i.e. the decomposition in segments: G.Alberti, B.Kirchheim and D.Preiss [4] found an example of a compact faily of pairwise disjoint open segments in $\mathbb{R}^3$ such that the family M of their midpoints has strictly positive Lebesgue measure (the construction is a variant of previous examples due to A.S.Besicovitch and D.G.Larman [33]). In this case, choosing $\mu=\mathcal{L}^3\llcorner M$, the conditional measures induced by the decomposition are Dirac masses. Therefore it is clear that this kind of counterexamples should be ruled out by some kind of additional "regularity" property of the decomposition. In this way the Sudakov strategy would be fully rigorous.

Several years later, Evans and Gangbo made a remarkable progress in [24], showing by differential methods 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 $\mathcal{L}^n$ and that their densities are Lipschitz functions with compact support. The missing piece of information about the length of transportation is recovered by a p-laplacian approximation

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

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

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

The measure $\sigma := a\mathcal{L}^n$, the so-called *transport density*, plays an important rôle in the theory: its uniqueness and its regularity are discussed, under more general assumptions on μ and ν, in [27], [20], [21]. This measure appears in the *scalar mass optimization* problem studied in [9, 10, 11], and in [3] several equivalent representation of the transport density and its uniqueness have been studied.

Coming back to the transport problem with Euclidean distance (or, more generally, with a distance induced by a C^2 and uniformly convex norm), 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 [16] and by N.Trudinger and L.Wang in [42]. Afterwards, the first author estabilished in [3] the existence of an optimal transport map assuming only that the initial measure μ is absolutely continuous, and the results of [16] and [42] have been extended to a Riemannian setting in [28]. All these proofs involve basically a Sudakov decomposition in transport rays, but the technical implementation of the idea is different from paper to paper: for instance in [16] a local change of variable is made, so that transport rays become parallel and Fubini theorem, in place of abstract disintegration theorems for measures, can be used. The proof in [3], instead, uses the co-area formula to show that absolute continuity with respect to Lebesgue measure is stable under disintegration.

In this paper we are particularly interested to the strategy pursued in [16], based on the approximation of the cost function $c(x, y) = \|x - y\|$ by the cost functions $c_\epsilon(x, y) = \|x - y\|^{1+\varepsilon}$. The approximation is used in that paper to build a special Kantorovich potential u, by taking limits as $\varepsilon \downarrow 0$ of the potentials $(u_\varepsilon, u_\varepsilon^{c_\varepsilon})$ in the dual formulation (see Section 3 for a precise description of the dual formulation). The potential u is used to prescribe the geometry of transport rays and to build, by a 1-dimensional reduction, an optimal transport map. Here we give a new variational interpretation of the Caffarelli-Feldman-McCann approximation, based on the theory of asymptotic developments by Γ-convergence, developed by G.Anzellotti and S.Baldo in [7] (see also [8]). This new interpretation provides stronger results and, in particular, allows us to show that the family of optimal maps t_ε relative to the costs c_ε converges in measure as $\varepsilon \downarrow 0$ to the map built in [16]. However, since we don't assume a priori the existence of this map, our strategy provides at the same time an existence and a stability result for the Monge problem.

The underlying variational principle in our argument is that any limit of the optimal plannings γ_ε associated to t_ε is not only optimal for the Kantorovich problem, but optimal for the *secondary* variational problem

$$\min_{\gamma\in\Pi_1(\mu,\nu)}\int_{\mathbb{R}^n\times\mathbb{R}^n}\|x-y\|\ln(\|x-y\|)\,d\gamma, \tag{6}$$

where $\Pi_1(\mu,\nu)$ denotes the class of all optimal plannings for the Kantorovich problem (the entropy function in (6) comes from the Taylor expansion of c_ε around $\varepsilon = 0$, see also [11]). Then, we show that the secondary variational problem has a unique minimizer, and that this minimizer is induced by a transport map (a posteriori, the map built in [16]).

It is very likely, as the authors themselves of [16] suggest in the introduction of their paper, that the convergence of t_ε can still be proved working in the dual formulation, without appealing to our variational argument. However, we discovered that this new principle can be used in some situations to provide a "variational" decomposition in transport rays, bypassing the above mentioned difficulties in Sudakov's argument: in the forthcoming paper [5] we will show existence of optimal transport maps for distances induced by any "crystalline" norm $\|\cdot\|$ (whose unit ball is contained in finitely many hyperplanes and therefore not strictly convex) by looking at the approximation

$$c_\varepsilon(x,y) := \|x-y\| + \varepsilon|x-y| + \varepsilon^2|x-y|\ln(|x-y|).$$

Quite surprisingly, also in this situation we get full convergence as $\varepsilon\downarrow 0$ of t_ε to an optimal map t. Moreover, we will obtain existence of optimal transports for distances induced by *any* norm in the planar case $n = 2$.

We close this introduction by an analytic description of the content of this paper, conceived as a survey paper but also with original results, some of which are necessary for the forthcoming work [5].

In Section 3 we develop the duality theory for the Kantorovich problem. In view of the applications we have in mind (see Remark 7.1) we allow lower semicontinuous and possibly infinite cost functions, showing that also in this situation the c-monotonicity is a necessary condition for minimality (this is one of the key technical ingredients in [5]). We also discuss the problem of sufficiency of c-monotonocity: a general answer to this problem is not known, but we find a very general sufficient condition which seems not to be available in the literature (see Remark 3.1). We also provide a counterexample, but with a $+\infty$ valued cost function.

Section 4 contains the basic facts about the theory of Γ-asymptotic developments.

Section 5 reviews the theory of optimal transportation on the real line, for convex and nondecreasing cost functions. In this situation it is well known that, if μ has no atom, the unique nondecreasing map t pushing μ into ν is optimal, and it is the unique optimal map (even among plannings) when the cost function is nondecreasing and strictly convex. We show also a simple variant of this result (Theorem 5.2) where we drop the monotonicity assumption on the cost, to allow the entropy function as in (6).

Section 6 contains an abstract version of Sudakov's argument, based on a decomposition of the space in 1-dimensional transport rays, see Theorem 6.2.

We will apply this result to solve the transport problem in particular situations, see [5], [6]. In view of the counterexample [4], we make the assumption that the family of (maximal) rays is countably Lipschitz to ensure that any absolutely continuous measure μ with respect to $\mathcal{L}^n$ produces, after disintegration, a family of measures concentrated on 1-dimensional rays and absolutely continuous with respect to $\mathcal{H}^1$, therefore with no atom. As a consequence the 1-dimensional theory of the previous section applies to these measures. Notice also that, in view of the example in [33], the countable Lipschitz property seems to be necessary also to show that the family of estreme points of rays is Lebesgue negligible.

Section 7 contains the existence and stability result for Monge optimal transports mentioned above, under the same assumptions on the norm made in [16].

In Section 8 we show by a counterexample (or, rather, a class of counterexamples) that the absolute continuity assumption on μ is necessary. This is a distinctive feature of the linear case, since for strictly convex cost functions we have existence and uniqueness of optimal transport maps whenever μ has dimension strictly greater than $n-1$ (see [30]).

Finally, the Appendix contains all basic facts about disintegration of measures needed in this paper. Of particular interest is Theorem 9.4, taken from [3], where we show stability of absolute continuity under disintegration, and the measurability criterion stated in Theorem 9.2.

2 Notation

In this section we fix our main notation and the terminology. We shall always assume that the measurable spaces we deal with are metric spaces endowed with the Borel σ-algebra, although this assumption could be easily weakened in some situations. Given a Borel map $f : X \to Y$, and given a positive and finite measure μ on X, we denote by $f_\#\mu$ its image, defined by $f_\#\mu(B) = \mu\left(f^{-1}(B)\right)$ for any Borel set $B \subset Y$. According to the change of variable formula we have

$$\int_Y \phi \, d\nu = \int_X \phi \circ f \, d\mu \qquad \text{for any bounded Borel function } \phi : Y \to \mathbb{R}.$$

We denote by $\operatorname{spt}\mu$ the support of μ, i.e. the closed set of all points $x \in X$ such that $\mu(B_r(x)) > 0$ for any $r > 0$. We say that μ is *concentrated* on a Borel set B if $\mu(X \setminus B) = 0$. If X is separable then μ is concentrated on $\operatorname{spt}\mu$ and the support is the minimal closed set on which μ is concentrated. On the other hand, a measure can be concentrated on a set much smaller than the support: for instance the probability measure

$$\mu := \sum_{n=0}^{\infty} 2^{-1-n} \delta_{q_n},$$

where $\{q_n\}$ is an enumeration of the rational numbers, has $\mathbb{R}$ as support but it is concentrated on $\mathbf{Q}$.

In the following table we resume the notation used without further explaination throughout the text:

$\mathcal{L}^n$	Lebesgue measure in $\mathbb{R}^n$
$\mathcal{H}^k$	Hausdorff k-dimensional measure in $\mathbb{R}^n$
$\mathbf{S}^{n-1}$	unit sphere in $\mathbb{R}^n$
$\mathcal{B}(X)$	Borel σ-algebra of X
$\mathrm{Lip}(X)$	real valued Lipschitz functions defined on X
$\mathrm{Lip}_1(X)$	functions in $\mathrm{Lip}(X)$ with Lipschitz constant not greater than 1
π_0, π_X	projection $X \times Y \ni (x,y) \mapsto x \in X$
π_1, π_Y	projection $X \times Y \ni (x,y) \mapsto y \in Y$
$\mathcal{S}_o(X)$	open oriented segments $]\!]x,y[\![$ with $x,\, y \in X,\, x \neq y$
$\mathcal{S}_c(X)$	closed oriented segments $[\![x,y]\!]$ with $x,\, y \in X,\, x \neq y$
$\mathcal{M}_+(X)$	positive and finite Radon measures in X
$\mathcal{P}(X)$	probability measures in X
$\mu \llcorner B$	restriction of μ to B, defined by $\chi_B \mu$.

3 Duality and optimality conditions

In this section we look for general necessary and sufficient optimality conditions for the Kantorovich problem (5). We make fairly standard assumptions on the spaces X, Y, assuming them to be locally compact and separable (these assumptions can be easily relaxed, see for instance [36]), but we look for general lower semicontinuous cost functions $c : X \times Y \to [0, +\infty]$, allowing in some cases the value $+\infty$. This extension is important in view of the applications we have in mind (see Remark 7.1 and [5]). See also [36] for more general versions of the duality formula.

Theorem 3.1 (Duality formula). *The minimum of the Kantorovich problem is equal to*

$$\sup \left\{ \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \right\} \tag{7}$$

where the supremum runs among all pairs $(\varphi, \psi) \in L^1(X, \mu) \times L^1(Y, \nu)$ such that $\varphi(x) + \psi(y) \leq c(x,y)$.

Proof. This identity is well-known if c is bounded, see for instance [36], [44]. In the general case it suffices to approximate c from below by an increasing sequence of bounded continuous functions c_h, defined for instance by

$$c_h(x,y) := \min \left\{ c(x',y') \wedge h + h d_X(x,x') + h d_Y(y,y') \right\},$$

noticing that a simple compactness argument gives

$$\min\left\{\int_{X\times Y} c_h\,d\gamma:\ \gamma\in\Pi(\mu,\nu)\right\}\quad\uparrow\quad\min\left\{\int_{X\times Y} c\,d\gamma:\ \gamma\in\Pi(\mu,\nu)\right\}$$

and that any pair (φ,ψ) such that $\varphi+\psi\le c_h$ is admissible in (7).

We recall briefly the definitions of c-transform, c-concavity and c-cyclical monotonicity, referring to the papers [22], [37], [30] and to the book [36] for a more detailed analysis.

For $u: X\to\overline{\mathbf{R}}$, the *$c$-transform* $u^c: Y\to\overline{\mathbf{R}}$ is defined by

$$u^c(y):=\inf_{x\in X} c(x,y)-u(x)$$

with the convention that the sum is $+\infty$ whenever $c(x,y)=+\infty$ and $u(x)=+\infty$. Analogously, for $v: Y\to\overline{\mathbf{R}}$, the *$c$-transform* $v^c: X\to\overline{\mathbf{R}}$ is defined by

$$v^c(x):=\inf_{y\in Y} c(x,y)-v(y)$$

with the same convention when an indetermination of the sum is present.

We say that $u: X\to\overline{\mathbf{R}}$ is *c-concave* if $u=v^c$ for some v; equivalently, u is c-concave if there is some family $\{(y_i,t_i)\}_{i\in I}\subset Y\times\overline{\mathbf{R}}$ such that

$$u(x)=\inf_{i\in I} c(x,y_i)+t_i\qquad\forall x\in X.$$

An analogous definition can be given for functions $v: Y\to\overline{\mathbf{R}}$.

It is not hard to show that $u^{cc}\ge u$ and that equality holds if and only if u is c-concave. Analogously, $v^{cc}\ge v$ and equality holds if and only if v is c-concave.

Finally, we say that $\Gamma\subset X\times Y$ is *c-monotone* if

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

whenever $(x_1,y_1),\ldots,(x_n,y_n)\in\Gamma$ and σ is a permutation of $\{1,\ldots,n\}$.

Theorem 3.2 (Necessary and sufficient optimality conditions).
(Necessity) If $\gamma\in\Pi(\mu,\nu)$ is optimal and $\int_{X\times Y} c\,d\gamma<+\infty$, then γ is concentrated on a c-monotone Borel subset of $X\times Y$.
(Sufficiency) Assume that c is real-valued, γ is concentrated on a c-monotone Borel subset of $X\times Y$ and

$$\mu\left(\left\{x\in X:\ \int_Y c(x,y)\,d\nu(y)<+\infty\right\}\right)>0,\tag{8}$$

$$\nu\left(\left\{y\in Y:\ \int_X c(x,y)\,d\mu(x)<+\infty\right\}\right)>0.\tag{9}$$

Then γ is optimal, $\int_{X\times Y} c\,d\gamma<+\infty$ and there exists a maximizing pair (φ,ψ) in (7) with φ c-concave and $\psi=\varphi^c$.

Proof. Let (φ_n, ψ_n) be a maximizing sequence in (7) and let $c_n = c - \varphi_n - \psi_n$. Since

$$\int_{X\times Y} c_n\,d\gamma = \int_{X\times Y} c\,d\gamma - \int_X \varphi_n\,d\mu - \int_Y \psi_n\,d\nu \to 0$$

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

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

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

Now we show the converse implication, assuming that (8) and (9) hold. We denote by Γ a Borel and c-monotone set on which γ is concentrated; without loss of generality we can assume that $\Gamma = \cup_k \Gamma_k$ with Γ_k compact and $c|_{\Gamma_k}$ continuous. We choose continuous functions c_l such that $c_l \uparrow c$ and split the proof in several steps.

Step 1. There exists a c-concave Borel function $\varphi : X \to [-\infty, +\infty)$ such that $\varphi(x) > -\infty$ for μ-a.e. $x \in X$ and

$$\varphi(x') \leq \varphi(x) + c(x', y) - c(x, y) \qquad \forall x' \in X,\ \ (x, y) \in \Gamma. \tag{10}$$

To this aim, we use the explicit construction given in the generalized Rockafellar theorem in [37], setting

$$\begin{aligned}\varphi(x) := \inf\{&c(x, y_p) - c(x_p, y_p) + c(x_p, y_{p-1}) - c(x_{p-1}, y_{p-1}) \\ &+ \cdots + c(x_1, y_0) - c(x_0, y_0)\}\end{aligned}$$

where $(x_0, y_0) \in \Gamma_1$ is fixed and the infimum runs among all integers p and collections $\{(x_i, y_i)\}_{1\leq i\leq p} \subset \Gamma$.

It can be easily checked that

$$\varphi = \lim_{p\to\infty} \lim_{m\to\infty} \lim_{l\to\infty} \varphi_{p,m,l},$$

where

$$\begin{aligned}\varphi_{p,m,l}(x) := \inf\{&c_l(x, y_p) - c(x_p, y_p) + c_l(x_p, y_{p-1}) - c(x_{p-1}, y_{p-1}) \\ &+ \cdots + c_l(x_1, y_0) - c(x_0, y_0)\}\end{aligned}$$

and the infimum is made among all collections $\{(x_i, y_i)\}_{1\leq i\leq p} \subset \Gamma_m$. As all functions $\varphi_{p,m,l}$ are upper semicontinuous we obtain that φ is a Borel function.

Arguing as in [37] it is straightforward to check that $\varphi(x_0) = 0$ and (10) holds. Choosing $x' = x_0$ we obtain that $\varphi > -\infty$ on $\pi_X(\Gamma)$ (here we use the

assumption that c is real-valued). But since γ is concentrated on Γ the Borel set $\pi_X(\Gamma)$ has full measure with respect to $\mu = \pi_{X\#}\gamma$, hence $\varphi \in \mathbb{R}$ μ-a.e.
Step 2. Now we show that $\psi := \varphi^c$ is ν-measurable, real-valued ν-a.e. and that

$$\varphi + \psi = c \quad \text{on } \Gamma. \tag{11}$$

It suffices to study ψ on $\pi_Y(\Gamma)$: indeed, as γ is concentrated on Γ, the Borel set $\pi_Y(\Gamma)$ has full measure with respect to $\nu = \pi_{Y\#}\gamma$. For $y \in \pi_Y(\Gamma)$ we notice that (10) gives

$$\psi(y) = c(x,y) - \varphi(x) \in \mathbb{R} \qquad \forall x \in \Gamma_y := \{x : \ (x,y) \in \Gamma\}.$$

In order to show that ψ is ν-measurable we use the disintegration $\gamma = \gamma_y \otimes \nu$ of γ with respect to y (see the appendix) and notice that the probability measure γ_y is concentrated on Γ_y for ν-a.e. y, therefore

$$\psi(y) = \int_X c(x,y) - \varphi(x)\, d\gamma_y(x) \qquad \text{for } \nu\text{-a.e. } y.$$

Since $y \mapsto \gamma_y$ is a Borel measure-valued map we obtain that ψ is ν-measurable.
Step 3. We show that φ^+ and ψ^+ are integrable with respect to μ and ν respectively (here we use (8) and (9)). By (8) we can choose x in such a way that $\int_Y c(x,y)\, d\nu(y)$ is finite and $\varphi(x) \in \mathbb{R}$, so that by integrating on Y the inequality $\psi^+ \le c(x,\cdot) + \varphi^-(x)$ we obtain that $\psi^+ \in L^1(Y,\nu)$. The argument for φ^+ uses (9) and is similar.
Step 4. Conclusion. The semi-integrability of φ and ψ gives the null-lagrangian identity

$$\int_{X\times Y} (\varphi + \psi)\, d\tilde\gamma = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \in \mathbb{R} \cup \{-\infty\} \qquad \forall \tilde\gamma \in \Pi(\mu,\nu),$$

so that choosing $\tilde\gamma = \gamma$ we obtain from (11) that $\int_{X\times Y} c\, d\gamma < +\infty$ and $\varphi \in L^1(X,\mu)$, $\psi \in L^1(Y,\nu)$. Moreover, for any $\tilde\gamma \in \Pi(\mu,\nu)$ we get

$$\begin{aligned}\int_{X\times Y} c\, d\tilde\gamma &\ge \int_{X\times Y} (\varphi+\psi)\, d\tilde\gamma = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \\ &= \int_{X\times Y} (\varphi+\psi)\, d\gamma = \int_\Gamma (\varphi+\psi)\, d\gamma = \int_{X\times Y} c\, d\gamma.\end{aligned}$$

This chain of inequalities gives that γ is optimal and, at the same time, that (φ,ψ) is optimal in (7).

We say that a Borel function $\varphi \in L^1(X,\mu)$ is a *maximal Kantorovich potential* if (φ,φ^c) is a maximizing pair in (7). In many applications it is useful to write the optimality conditions using a maximal Kantorovich potential, instead of the cyclical monotonicity.

Theorem 3.3. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, assume that (8) and (9) hold, that c is real-valued and the* sup *in (7) is finite. Then there exists a maximizing pair (φ, φ^c) in (7) and $\gamma \in \Pi(\mu, \nu)$ is optimal if and only if*

$$\varphi(x) + \varphi^c(y) = c(x, y) \qquad \gamma\text{-a.e. in } X \times Y.$$

Proof. The existence of a maximizing pair is a direct consequence of the sufficiency part of the previous theorem, choosing an optimal γ and (by the necessity part of the statement) a c-monotone set on which γ is concentrated.

If γ is optimal then

$$\int_{X\times Y} (c - \varphi - \varphi^c)\, d\gamma = \int_{X\times Y} c\, d\gamma - \int_X \varphi\, d\mu - \int_Y \varphi^c\, d\nu = 0.$$

As the integrand is nonnegative, it must vanish γ-a.e. The converse implication is analogous.

Remark 3.1. The assumptions (8), (9) are implied by

$$\int_{X\times Y} c(x, y)\, d\mu \otimes \nu(x, y) < +\infty. \tag{12}$$

In turn, (12) is weaker than the condition

$$c(x, y) \le a(x) + b(y) \quad \text{with} \quad a \in L^1(\mu),\ b \in L^1(\nu)$$

considered in Theorem 2.3.12 of Part I of [36].

The following example shows that some kind of finiteness/integrability condition seems to be necessary in order to infer minimality from cyclical monotonicity. It is interesting to notice that in very specific cases (but important for the applications) like $X = Y = \mathbb{R}^n$ and $c(x, y) = |x - y|^2$ it is not presently clear whether cyclical monotonicity implies minimality *without* additional conditions, e.g. the finiteness of the moments $\int |x|^2\, d\mu$, $\int |y|^2\, d\nu$ (see the Open problem 16 in Chapter 3 of [44]).

Example 3.1. Given $\alpha \in [0, 1] \setminus \mathbf{Q}$, let $\varphi : [0, 1] \times [0, 1] \to [0, +\infty]$ be defined as follows:

$$\varphi(x, y) := \begin{cases} 1 & \text{if } y = x \\ 2 & \text{if } y = x \oplus \alpha \\ +\infty & \text{otherwise,} \end{cases}$$

where $\oplus : [0, 1] \times [0, 1] \to [0, 1]$ is the sum modulo 1: Figure 1 shows this function. Let us then consider the transport problem in $\Omega = [0, 2] \subset \mathbb{R}$ with $\mu = \mathcal{L}^1 \llcorner [0, 1]$, $\nu = \mathcal{L}^1 \llcorner [1, 2]$ and with any cost c such that $c(x, y) = \varphi(x, y - 1)$ whenever $0 \le x \le 1 \le y \le 2$. Clearly the unique optimal plan of transport is

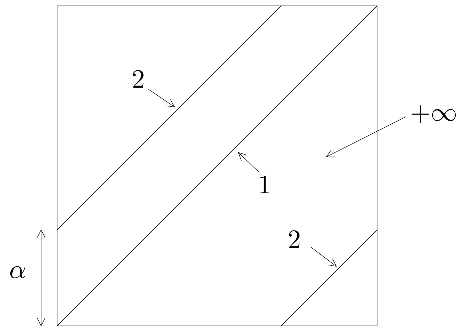

Fig. 1. The function φ in Example 3.1

$\bar{\gamma} = (Id, 1 + Id)_\# \mathcal{L}^1 \llcorner [0,1]$, while we will prove that also the support of the non-optimal plan $\gamma = \big(Id, 1 + (Id \oplus \alpha)\big)_\# \mathcal{L}^1 \llcorner [0,1]$ is c-monotone.

If not, there would be a minimal set of couples (x_1, y_1), $(x_2, y_2), \dots,$ (x_n, y_n) in the support of γ with the property that

$$d(x_1, y_1) + \cdots + d(x_n, y_n) > d(x_2, y_1) + d(x_3, y_2) + \cdots + d(x_1, y_n);$$

but $(x_i, y_i) \in \operatorname{spt}\gamma$ means $y_i = 1 + (x_i \oplus \alpha)$: moreover, since the preceding inequality assures $d(x_{i+1}, y_i) < +\infty$, we can infer that $y_i = 1 + x_{i+1}$ or $y_i = 1 + (x_{i+1} \oplus \alpha)$, and then $x_{i+1} = x_i \oplus \alpha$ or $x_{i+1} = x_i$. Since the second possibility is incompatible with the minimality of the set, it must be $x_{i+1} = x_i \oplus \alpha$; applying this equality n times, we find $x_1 = x_1 \oplus n\alpha$, which is impossible since α is not a rational number.

4 Γ-convergence and Γ-asymptotic expansions

In this section we recall some basic facts about Γ-convergence and we present the essential aspects of the theory of Γ-asymptotic expansions, first introduced in [7] (see also [8] for an application of this theory in elasticity). A general reference for the theory of Γ-convergence is [18].

Let X be a compact metric space and let us denote by $S_-(X)$ the collection of all lower semicontinuous functions $f : X \to \overline{\mathbf{R}}$. We say that a sequence $(f_h) \subset S_-(X)$ *Γ-converges* to $f \in S_-(X)$ if the following two properties hold for any $x \in X$:

(a) $\liminf f_h(x_h) \geq f(x)$ for any sequence $x_h \to x$;
(b) there exists a sequence $x_h \to x$ such that $f_h(x_h) \to f(x)$.

It can be shown that the Γ-convergence is induced by a metric d_Γ and that $(S_-(X), d_\Gamma)$ is a compact metric space.

If (f_h) Γ-converges to f then $m_h := \min_X f_h \to m := \min_X f$ and, in addition,

$$\limsup_{h\to\infty} \operatorname{Argmin}(f_h) \subset \operatorname{Argmin}(f). \tag{13}$$

In words, any limit point of minimizers of f_h minimizes f. The same is true for sequences (x_h) which are asymptotically minimizing, i.e. such that $f_h(x_h) - m_h \to 0$.

However, the converse is not true in general: for instance the functions $f_h(x) = 1/h \wedge |x|$ Γ-converge to $f \equiv 0$ in $X = [-1, 1]$ but $x = 0$ is the only minimizer of f that can be approximated by minimizers of f_h.

In order to improve the inclusion above, G.Anzellotti and S.Baldo proposed the following procedure: assuming that m is a real number, they proposed to consider the new functions

$$f'_h := \frac{f_h - m}{\varepsilon_h}$$

for suitable positive infinitesimals ε_h. Assuming that f'_h Γ-converge to f' (this is not really restrictive, by the compactness of $S_-(X)$), the following result holds:

Theorem 4.1. *The functional f' is equal $+\infty$ out of* $\operatorname{Argmin} f$, *hence* $\operatorname{Argmin} f' \subset \operatorname{Argmin} f$. *Moreover*

$$\limsup_{h\to\infty} \operatorname{Argmin}(f_h) \subset \operatorname{Argmin}(f'). \tag{14}$$

Proof. If $f'(x) < +\infty$ there exists a sequence (x_h) converging to x, by condition (b), such that $f'_h(x_h)$ is bounded above. As

$$f_h(x_h) = m + \varepsilon_h f'_h(x_h) \le m_h + o(1)$$

the sequence (x_h) is asymptotically minimizing and therefore $x \in \operatorname{Argmin} f$. Finally (14) is a direct consequence of (13), noticing that $\operatorname{Argmin} f_h = \operatorname{Argmin} f'_h$.

If ϵ_h have been properly chosen, so that $m' := \min_X f' \in \mathbb{R}$, then the convergence of $\min_X f'_h = (m_h - m)/\varepsilon_h$ to m' gives the expansion

$$m_h = m + m'\varepsilon_h + o(\varepsilon_h).$$

This procedure can of course be iterated, giving further restrictions on the set of limit points of minimizers of f_h and higher order expansions of the difference $m_h - m$.

5 1-dimensional theory

In this section we recall some aspects of the theory of optimal transportation in $\mathbb{R}$. In this case, at least when μ has no atom, there is a canonical transport map obtained by monotone rearrangement. This map is optimal whenever the cost is a nondecreasing and convex function of the distance.

Theorem 5.1. *Let μ, $\nu \in \mathcal{P}(\mathbb{R})$, μ without atoms, and let*

$$G(x) := \mu\left((-\infty, x)\right), \qquad F(y) := \nu\left((-\infty, y)\right)$$

be respectively the distribution functions of μ, ν. Then

(i) The nondecreasing function $t : \mathbb{R} \to \overline{\mathbf{R}}$ defined by

$$t(x) := \sup\left\{y \in \mathbb{R} : \ F(y) \le G(x)\right\}$$

(with the convention $\sup\emptyset = -\infty$) maps μ into ν. Any other nondecreasing map t' such that $t'_{\#}\mu = \nu$ coincides with t on $\operatorname{spt}\mu$ up to a countable set.

(ii) If $\phi : [0, +\infty) \to \mathbb{R}$ is nondecreasing and convex, then t is an optimal transport between μ and ν relative to the cost $c(x, y) = \phi(|x-y|)$. Moreover t is the unique optimal transport map if ϕ is strictly convex.

For the proof the reader may consult [3], [30], [44]).

Notice that the monotonicity constraint forces us to take $\overline{\mathbf{R}}$ as range of t (for instance this happens when μ has compact support and $\operatorname{spt}\nu = \mathbb{R}$). However, since $t_{\#}\mu = \nu$, the half-lines $\{t = \pm\infty\}$ are μ-negligible.

In view of the applications we have in mind (where $\phi(t)$ could be $t\ln t$) we are interested in dropping the assumption that ϕ is nondecreasing. This can be done under a suitable compatibility condition between μ and ν, expressed in (15) below, by restricting the class of competitors γ.

Theorem 5.2. *Let μ, $\nu \in \mathcal{P}(\mathbb{R})$, μ without atoms and let t be the map in Theorem 5.1. Assume that μ and ν have finite first order moments and*

$$\mathcal{A} := \left\{\gamma \in \varPi(\mu, \nu) : \ \operatorname{spt}\gamma \subset \left\{(x, y) : \ y \ge x\right\}\right\} \ne \emptyset. \tag{15}$$

Then $t(x) \ge x$ for μ-a.e. $x \in \mathbb{R}$ and $\gamma_t = (Id \times t)_{\#}\mu$ is a solution of the problem

$$\min_{\gamma \in \mathcal{A}} \int_{\mathbb{R}\times\mathbb{R}} \phi(|y - x|)\, d\gamma \tag{16}$$

whenever $\phi : [0, +\infty) \to \mathbb{R}$ is a convex function bounded from below. If ϕ is strictly convex and the minimum in (16) is finite, then γ_t is the unique solution.

Proof. We argue as in [30] and we consider the strictly convex case only (the general case follows by a simple perturbation argument). Since $\mathcal{A}$ is weakly compact, we can find an optimal $\gamma \in \mathcal{A}$ for (16). Assuming that γ has finite energy, by the construction in [30] one can show that $\varGamma := \operatorname{spt}\gamma$ satisfies the restricted monotonicity condition

$$\phi(y_1 - x_1) + \phi(y_2 - x_2) \le \phi(y_2 - x_1) + \phi(y_1 - x_2)$$

whenever (x_1, y_1), $(x_2, y_2) \in \varGamma$ and $x_1 < y_2$, $x_2 < y_1$ (indeed, in this case the additional constraint that competitors must be in $\mathcal{A}$ is not effective).

Now we show the implication:

$$(x_1, y_1) \in \Gamma, \quad (x_2, y_2) \in \Gamma, \quad x_1 < x_2 \qquad \Longrightarrow \qquad y_1 \leq y_2. \tag{17}$$

Assuming by contradiction that $y_1 > y_2$, since $x_i \leq y_i$, $i = 1, 2$, we obtain $x_1 < x_2 \leq y_2 < y_1$. In this case, setting $a = x_2 - x_1$, $b = y_2 - x_2$, $c = y_1 - y_2$, the cyclical monotonicity of Γ gives

$$\phi(a + b + c) + \phi(b) \leq \phi(a + b) + \phi(b + c).$$

On the other hand, since $c > 0$ the strict convexity of ϕ gives

$$\phi(a + b + c) - \phi(b + c) > \phi(a + b) - \phi(b)$$

and therefore a contradiction.

By (17) we obtain that the vertical sections Γ_x of Γ are ordered, i.e. $y_1 \in \Gamma_{x_1} \leq y_2 \in \Gamma_{x_2}$ whenever $x_1 < x_2$. As a consequence the set of all x such that Γ_x is not a singleton is at most countable (since we can index with this set a family of pairwise disjoint open intervals), and therefore μ-negligible. As $\operatorname{spt} \gamma_x \subset \Gamma_x$, it follows that γ_x is a Dirac mass for μ-a.e. x. Setting $\gamma_x = \delta_{t'(x)}$ the map t' is nondecreasing by (17), satisfies $t'(x) \geq x$ because $\gamma \in \mathcal{A}$ and maps μ into ν because $\gamma = (Id \times t')_\# \mu$. By Theorem 5.1(i) we obtain $t' = t$ μ-a.e.

The proof is finished by showing that problem (16) is non trivial (i.e. the minimum is finite) for at least one stricly convex ϕ. This follows by the assumption on the finiteness of first moments, choosing $\phi(t) := \sqrt{1 + t^2} - 1$.

6 Transport rays and transport set

Given $\Gamma \subset \mathbf{R}^n \times \mathbf{R}^n$, we define the notions of *transport ray*, *ray direction*, *transport set*, *fixed point*, *maximal transport ray.* The terminology is close to the one adopted in [24] and [16], with the difference that our definitions depend on Γ rather than a Kantorovich potential. We reconcile with the other approaches in (24) and in Theorem 6.2.

(Transport ray) We say that $]\!]x, y[\![$ 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 sets) We denote by T_Γ the union of all transport rays relative to Γ, i.e.

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

We define also T_Γ^l as the union of all sets $[\![x,y[\![$, as $(x,y) \in \Gamma$, and T_Γ^r as the union of all sets $]\!]x,y]\!]$, as $(x,y) \in \Gamma$ (by convention $[\![x,y[\![=]\!]x,y]\!] = \{x\}$ if $x = y$).

(Fixed points) We say that x is a fixed point if $(x,x) \in \Gamma$ and $(x,y) \notin \Gamma$ for any $y \neq x$. We denote by F_Γ the set of all fixed points.

(Maximal transport ray) We say that an open interval $S \subset \mathbf{R}^n$ (possibly unbounded) is a maximal transport ray if

(a) for any $z \in S$ there exists $(x,y) \in \Gamma$ with $z \in]\!]x,y[\![$;

(b) any open interval containing S and satisfying (a) coincides with S.

Notice that maximal transport rays need not be transport rays: for instance if $\Gamma = \{(1/(k+2), 1/k)\}$, for $k \geq 1$ integer, then $(0,1)$ is a maximal transport ray but not a transport ray.

In the following we shall always assume that Γ is a σ-compact set. This ensures that all sets T_Γ, T_Γ^l, T_Γ^r, F_Γ associated to Γ are Borel sets. Notice also that

$$F_\Gamma = \pi_0(\Gamma \cap \Delta) \setminus \pi_0(\Gamma \setminus \Delta),$$

where Δ is the diagonal of $\mathbb{R}^n \times \mathbb{R}^n$.

Clearly any transport ray is contained in a unique maximal transport ray. As a consequence, any point in T_Γ is contained in a maximal transport ray. Under the no-crossing condition

$$[\![x,y]\!] \cap [\![x',y']\!] \neq \emptyset \quad \Longrightarrow \quad x = x' \text{ or } y = y' \qquad \text{whenever } \tau(x,y) \neq \tau(x',y') \tag{18}$$

(meaning that two closed rays with different orientations can meet only at a common endpoint) it is also immediate to check that this maximal transport ray is unique. Therefore Γ induces a map $\pi_\Gamma : T_\Gamma \to \mathcal{S}_o(\mathbb{R}^n)$ which associates to any point the maximal transport ray containing it. We also denote by $\tau_\Gamma : T_\Gamma \to \mathbf{S}^{n-1}$ the map which gives the direction of the transport ray.

Since we used the open segments to define the transport set and the maximal transport ray, we need to take into account also the extreme points (which are not in T_Γ) of the maximal transport rays.

The following proposition shows that only points in $T_\Gamma^l \cup T_\Gamma^r$ can carry some mass, and therefore are relevant for the transport problem.

Proposition 6.1. *(i) Any point in $T_\Gamma^l \setminus (T_\Gamma \cup F_\Gamma)$ (respectively $T_\Gamma^r \setminus (T_\Gamma \cup F_\Gamma)$) is a left (resp. right) extreme point of a maximal transport ray.*
(ii) If $\gamma \in \Pi(\mu,\nu)$ is concentrated on Γ, then μ is concentrated on T_Γ^l and ν is concentrated on T_Γ^r.

Proof. (i) If $x \in T_\Gamma^l \setminus F_\Gamma$ there exists $y \neq x$ such that $(x,y) \in \Gamma$. If $x \notin T_\Gamma$ the maximal transport ray containing $]\!]x,y[\![$ must have x as left extreme point.
(ii) Clearly μ is concentrated on L, the projection of Γ on the first factor. If $x \in L$ there exists $y \in \mathbb{R}^n$ such that $(x,y) \in \Gamma$, and therefore $x \in T_\Gamma^l$. The argument for ν is similar.

The converse implication in Proposition 6.1(i) does not hold, as shown by the previous example of a maximal transport ray which is not a transport ray.

Definition 6.1 (Metric on $\mathcal{S}_c(\mathbb{R}^n)$ and $\mathcal{S}_o(\mathbb{R}^n)$). In the following we need to define a metric structure (actually we would need only a measurable one) on the spaces $\mathcal{S}_c(\mathbb{R}^n)$ and $\mathcal{S}_o(\mathbb{R}^n)$. When one considers only bounded oriented segments the natural metric comes from the embedding into $\mathbb{R}^n \times \mathbb{R}^n$, by looking at the distances between the two left extreme points and the two right extreme points. In our case, since we allow halflines and lines as segments, we define the metric in $\mathcal{S}_c(\mathbb{R}^n)$ as

$$d\left(S,S'\right) := \sum_{R=1}^{\infty} 2^{-R} \frac{|x_R - x'_R| + |y_R - y'_R|}{1 + |x_R - x'_R| + |y_R - y'_R|},$$

where $[\![x_R, y_R]\!]$ (respectively $[\![x'_R, y'_R]\!]$) is the intersection of S (resp. S') with the closed ball $\overline{B}_R$. Since any open segment is in one to one correspondence with a closed segment (remember that singletons do not belong to $\mathcal{S}_c(\mathbb{R}^n)$) we define a metric in $\mathcal{S}_o(\mathbb{R}^n)$ in such a way that this correspondence is an isometry.

It is not hard to check that $\mathcal{S}_o(\mathbb{R}^n)$ and $\mathcal{S}_c(\mathbb{R}^n)$ are locally compact and separable metric spaces.

Now we prove a measurability result about the map π_Γ.

Lemma 6.1. *If Γ is $\sigma-$compact and (18) holds, then the map π_Γ which associates to any point in T_Γ the maximal transport ray in $\mathcal{S}_o(\mathbb{R}^n)$ containing it is Borel.*

Proof. Since Γ is $\sigma-$compact, let us write $\Gamma = \bigcup_{i\in\mathbb{N}} \Gamma_i$, where each Γ_i is compact; we define also $\varphi : \mathrm{P}(\mathcal{S}_o) \to \mathrm{P}(\mathbb{R}^n)$, denoting by $\mathrm{P}(X)$ the subsets of X, as follows:

$$\varphi(C) := \{x \in \mathbb{R}^n : \exists S \in C \, s.t. \, x \in S\} \qquad \forall C \subset \mathcal{S}_o.$$

The first thing we can note is the following
Claim: If $C \subset \mathcal{S}_o(\mathbb{R}^n)$ is closed, then $\varphi(C) \subset \mathbb{R}^n$ is Borel.
Let us first of all consider the case when C is compact: the assert follows directly, recalling Definition 6.1: if we would work with the closed segments, i.e. with $\mathcal{S}_c(\mathbb{R}^n)$ instead of $\mathcal{S}_o(\mathbb{R}^n)$, then the image of C were easily seen to be compact, and then $\varphi(C)$ is a countable union of compact sets, and then a Borel set. In general, if C is closed, it is a countable union of compact sets, and then $\varphi(C)$ is a countable union of Borel sets, thus Borel.

To prove the thesis, given a compact $C \subset \mathcal{S}_o(\mathbb{R}^n)$, it suffices to show that ${\pi_\Gamma}^{-1}(C)$ is a Borel set: to do this, first of all we define

$$C_i := \left\{ \left]sx + (1-s)y, ty + (1-t)x\right[\;\; s.t. \,]\!]x, y[\![\in C, \, 0 \le s, t \le \frac{1}{2^i} \right\},$$

which is easily seen to be closed (in fact, it is compact). Given any $p > 0$, we will moreover denote by G_p the (closed) set of all the open segments in $\mathbb{R}^n$ of lenght p. We need to define now the sets $\Gamma_{i,j} \subset \mathcal{S}_o(\mathbb{R}^n)$ with $i, j \in \mathbb{N}$; to begin, for any $X \subset \mathcal{S}_o(\mathbb{R}^n)$ we will denote by $Sub(X)$ the set of all its open subsegments: in other words, $]\!]x, y[\![\in Sub(X)$ if and only if there exists $]\!]z, w[\![\in X$ such that $]\!]x, y[\![\subset]\!]z, w[\![$. For $j = 1$, then, we fix $\Gamma_{i,1} := Sub\,(\Gamma_i)$, while for $j = 2$, it will be

$$\Gamma_{i,2} := Sub\Big(\Gamma_{i,1} \bigcup \{]\!]x, y[\![=]\!]x, w[\![\cup]\!]z, y[\![\text{ s.t. }]\!]x, w[\![,]\!]z, y[\![\in \Gamma_{i,1}, |z - w| \geq 1\}\Big) .$$

In words, we add to $\Gamma_{i,1}$ some open segments which are union of two segments in $\Gamma_{i,1}$, and then consider again all the possible subsegments; it is easy to note that $\Gamma_{i,2}$ is closed thanks to the request $|z - w| \geq 1$ in the last definition (in fact, we made that assumption only to ensure the closedness of $\Gamma_{i,2}$). In general, we will write

$$\Gamma_{i,j+1} := Sub\big(\Gamma_{i,j} \cup \big\{]\!]x, y[\![=]\!]x, w[\![\cup]\!]z, y[\![\text{ s.t. }]\!]x, w[\![,]\!]z, y[\![\in \Gamma_{i,j}, |z - w| \geq 1/j\big\}\big)$$

which generalizes the definition of $\Gamma_{i,2}$. One can note that $\Gamma_{i,i}$ is an increasing sequence of subsets of $\mathcal{S}_o(\mathbb{R}^n)$ and that

$$\Gamma_{i,i} \longrightarrow Sub\left(\pi_\Gamma(T_\Gamma)\right) \qquad \textit{for } i \to +\infty.$$

The last step to reach the thesis is then to note that

$$\pi_\Gamma{}^{-1}(C) = \bigcap_{i\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcup_{p\in\mathbb{Q}^+} \bigcap_{n\geq m} \Big(\big(\varphi\big(\Gamma_n \cap C_i \cap G_p\big)\big) \Big\backslash \big(\varphi\big(\Gamma_n \cap G_{p+1/2^{i-1}}\big)\big)\Big),$$

which assures $\pi_\Gamma{}^{-1}(C)$ to be Borel. To be convinced of the last equation, let us restrict our attention to the case when Γ is associated to a single maximal transport ray $]\!]x, y[\![$ of lenght l: then we can note that, for n sufficiently large,

$$\big(\varphi\big(\Gamma_n \cap G_p\big)\big) \setminus \big(\varphi\big(\Gamma_n \cap G_{p+1/2^{i-1}}\big)\big)$$

is empty unless $p \leq l \leq p + 1/2^{i-1}$. Thus

$$\bigcup_{m\in\mathbb{N}} \bigcup_{p\in\mathbb{Q}^+} \bigcap_{n\geq m} \Big(\big(\varphi\big(\Gamma_n \cap C_i \cap G_p\big)\big) \Big\backslash \big(\varphi\big(\Gamma_n \cap G_{p+1/2^{i-1}}\big)\big)\Big) \tag{19}$$

is empty if C_i does not contain segments "close" to $]\!]x, y[\![$, and then the intersection among all the integers i is empty if C does not contain $]\!]x, y[\![$; on the other hand, if C contains $]\!]x, y[\![$ then the set in (19) is exactly $\varphi(]\!]x, y[\![)$. The intersection for all the integers i, then, gives the thesis. We can now note that the same argument works in the general case with many maximal transport rays, since we can consider separately each maximal ray and apply the argument to it; note also that the role of the no-crossing condition (18) is only to ensure that each point in T_Γ is contained in a unique maximal transport ray, which allows to define the map π_Γ.

Now we state an "abstract" existence result on optimal transport maps. The result is valid under some regularity conditions on the decomposition in transport rays induced by Γ. Notice that assumption (ii) below on the μ-negligibility of left extreme points which are not fixed points is necessary for $n \geq 3$ even for absolutely continuous measures μ, in view of the counterexample in [33]. We will show in [5] that (ii), (iii) hold for $n = 2$ whenever μ is absolutely continuous.

Theorem 6.1. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ with finite first order moments and let $\gamma \in \Pi(\mu, \nu)$ be concentrated on Γ. Assume that (18) holds and that*

(i) μ is absolutely continuous with respect to $\mathcal{L}^n$;
(ii) $T^l_\Gamma \setminus (T_\Gamma \cup F_\Gamma)$ is μ-negligible;
(iii) there exists a μ-negligible set $N \subset T_\Gamma$ and an increasing sequence of compact sets K_h such that $\tau_\Gamma|_{K_h}$ is a Lipschitz map and the union of K_h is $T_\Gamma \setminus N$.

Then there exists $\gamma_\# \in \Pi(\mu, \nu)$ such that

(a) $\gamma_\#$ is induced by a transport map t and t is optimal whenever Γ is c-monotone.
(b) $\gamma_\#$ is concentrated on the set of pairs (x, y) such that either $x = y$ or $[\![x, y]\!]$ is contained in the closure of a maximal transport ray of Γ.
(c) For any convex function $\phi : [0, +\infty) \to \mathbb{R}$ bounded from below we have

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \phi(|x - y|)\, d\gamma_\# \leq \int_{\mathbb{R}^n \times \mathbb{R}^n} \phi(|x - y|)\, d\gamma.$$

If ϕ is strictly convex the inequality above is strict unless $\gamma = \gamma_\#$.

Proof. We split the proof in five steps. In the first four steps we assume that $F_\Gamma \subset T_\Gamma$, i.e. that any fixed point is contained in a transport ray. In the following, points in the closure of the same maximal transport rays will be ordered according to the orientation of the ray.

Step 1. The map r which associates to a pair $(x, y) \in \Gamma$ the closure of the maximal transport ray containing $]\!]x, y[\![$ when $x \neq y$ and containing $\{x\}$ when $x = y$ is well defined on Γ, hence defined γ-a.e. According to Theorem 9.2 we can represent

$$\gamma = \gamma_C \otimes \sigma \qquad \text{with} \qquad \sigma := r_\# \gamma \in \mathcal{P}(\mathcal{S}_c(\mathbb{R}^n))$$

where γ_C are probability measures in $\mathbb{R}^n \times \mathbb{R}^n$ concentrated on $r^{-1}(C) \subset C \times C$, and therefore satisfying

$$\gamma_C\left(\{(x, y) \in C \times C : \; y < x\}\right) = 0. \tag{20}$$

We define $\mu_C := \pi_{0\#}\gamma_C$ and $\nu_C := \pi_{1\#}\gamma_C$, so that $\gamma_C \in \Pi(\mu_C, \nu_C)$ and (37) yields

$$\mu = \mu_C \otimes \sigma, \qquad \nu = \nu_C \otimes \sigma. \tag{21}$$

Since $r^{-1}(C) \subset C \times C$ the probability measures μ_C, ν_C are concentrated on C. More precisely, we will use in the following the fact that μ_C is concentrated on $\pi_0\left(r^{-1}(C)\right)$. Notice also that μ_C, ν_C have finite first order moments for σ-a.e. C because μ and ν have finite first order moments.

Step 2. We claim that μ_C has no atom for σ-a.e. C.

Taking into account Proposition 6.1(ii), the inclusion $F_\Gamma \subset T_\Gamma$ and the assumption (ii) we obtain that μ is concentrated on T_Γ, therefore μ_C is concentrated on $T_\Gamma \cap \pi_0(r^{-1}(C))$ for σ-a.e. C. Now we check that, due to condition (18), $T_\Gamma \cap \pi_0(r^{-1}(C))$ is contained in the relative interior of C. Indeed, let $x \in T_\Gamma \cap \pi_0(r^{-1}(C))$, let $(x', y') \in \Gamma$ such that $x \in]\!]x', y'[\![$ and let C_0 be the closure of the maximal transport ray containing $]\!]x', y'[\![$. If $C = C_0$ then x is in the relative interior of C; if, on the other hand, $C \neq C_0$ then there exists $y \in \mathbb{R}^n$ such that $r(x, y) = C$, thus condition (18) is violated because $[\![x, y]\!]$ and $[\![x', y']\!]$ intersect at x', in the relative interior of $[\![x, y]\!]$.

Denoting by $\pi_\Gamma : T_\Gamma \mapsto \mathcal{S}_o(\mathbb{R}^n)$ the map which associates to a point the maximal transport ray containing it, by Theorem 9.4, Remark 9.1 and assumption (iii) we have $\mu = \mu'_A \otimes \theta$ with $\theta = \pi_{\Gamma\#}\mu$, μ'_A concentrated on $\pi_\Gamma^{-1}(A) \subset A$, and $\mu'_A \ll \mathcal{H}^1 \llcorner A$ (and in particolar has no atom) for θ-a.e. A.

Denoting by $\mathrm{cl} : \mathcal{S}_o(\mathbb{R}^n) \to \mathcal{S}_c(\mathbb{R}^n)$ the bijection which associates to an open segment its closure, we have

$$\mu = \mu_C \otimes \sigma(C) = \mu_{\mathrm{cl}(A)} \otimes \mathrm{cl}_\#^{-1}\sigma(A)$$

and since $\mu_{\mathrm{cl}(A)}$ are probability measures concentrated on A the uniqueness Theorem 9.2 gives $\mathrm{cl}_\#^{-1}\sigma = \theta$ and $\mu_{\mathrm{cl}(A)} = \mu'_A$ for θ-a.e. A.

As $\sigma = \mathrm{cl}_\#\theta$, this proves that μ_C has no atom for σ-a.e. C.

Step 3. According to Theorem 5.1 we can find a non-decreasing map t_C defined on the relative interior of C (i.e. the maximal transport ray relative to C) and with values in C, such that $t_{C\#}\mu_C = \nu_C$. Moreover, by Theorem 5.2 and (20), for any convex function $\phi : [0, +\infty) \to \mathbb{R}$ bounded from below we have

$$\int_C \phi(|x - t_C(x)|)\, d\mu_C \leq \int_{C \times C} \phi(|x - y|)\, d\gamma_C \tag{22}$$

with strict inequality if ϕ is strictly convex and $(Id \times t_C)_\#\mu_C \neq \gamma_C$.

Step 4. We define t on T_Γ by gluing the maps t_C. Since the map $C \mapsto \gamma_C = (Id \times t_C)_\#\mu_C$ is Borel (as a measure-valued map, see the Appendix) by Theorem 9.3 we infer the existence of a Borel map t such that $t = t_C$ μ_C-a.e. for σ-a.e. C. As t_C and t map μ_C in ν_C, it follows immediately from (21) that t maps μ into ν.

Setting $\gamma_\# := (Id \times t)_\#\mu \in \Pi(\mu, \nu)$, conditions (a) and (b) are satisfied by construction, as the segments $[\![x, t(x)]\!]$ are contained in the closure of a maximal transport ray of Γ. Condition (c) follows by (22) after an integration on $\mathcal{S}_c(\mathbb{R}^n)$ with respect to σ.

Step 5. Now we remove the assumption that $F_\Gamma \subset T_\Gamma$. Let $L := F_\Gamma \setminus T_\Gamma$ and let $\Gamma' = \Gamma \setminus \{(x,x):\ x \in L\}$. By applying the first four steps to $\gamma' := \gamma \llcorner \Gamma'$ we obtain a transport map t defined on $T_{\Gamma'}$ mapping $\pi_{0\#}\gamma'$ to $\pi_{1\#}\gamma'$. Noticing that $T_\Gamma = T_{\Gamma'}$, it suffices to extend t to L setting $t = Id$ on L and to set $\gamma_\# := (Id \times t)_\#\mu$.

Now we assume that $\|\cdot\|$ is a norm in $\mathbb{R}^n$ satisfying the regularity and uniform convexity conditions

$$c_1 \le \frac{\partial^2}{\partial\xi\partial\xi}\|\cdot\|^2 \le c_2 \quad \forall \xi \in \mathbf{S}^{n-1} \qquad \text{for some } c_2 \ge c_1 > 0. \tag{23}$$

We consider a σ-compact set $\Gamma \subset \Gamma_u$, where

$$\Gamma_u := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n :\ \|x-y\| = u(x) - u(y)\} \tag{24}$$

for some function $u : \mathbb{R}^n \to \mathbb{R}$ which is 1-Lipschitz relative to the distance induced in $\mathbb{R}^n$ by $\|\cdot\|$.

Theorem 6.2. *With the choice of Γ above, (18) and the following properties hold:*

(i) For any maximal transport ray S we have $u(x')-u(y') = \|x'-y'\|$ whenever x', $y' \in S$ and $x' \le y'$.
(ii) The set $T^l_\Gamma \setminus (T_\Gamma \cup F_\Gamma)$ is Lebesgue negligible.
(iii) Condition (iii) in Theorem 6.1 holds.

Proof. Arguing as in [16] (or [24] for the euclidean norm) one can show that at any point $z \in]\!]x,y[\![$ inside T_Γ the function u is differentiable, $\|du(z)\|^* = 1$, and $(du(z))^*$ is the direction of the ray $]\!]x,y[\![$. This immediately leads to the fact that (18) holds.
(i) We first show the property for transport rays. If $x', y' \in]\!]x,y[\![$ and $x' \le y'$ then

$$u(x) - u(x') \le \|x - x'\| \qquad \text{and} \qquad u(y') - u(y) \le \|y - y'\|,$$

so that, taking into account that $u(x) - u(y) = \|x-y\|$ and

$$\|x-y\| = \|x-x'\| + \|x'-y'\| + \|y'-y\|$$

we obtain that $u(x') - u(y') \ge \|x'-y'\|$. The extension to maximal transport rays is analogous.
Properties (ii), (iii) with $\Gamma = \Gamma_u$ and $N = \emptyset$ are shown in [16]. A fortiori (iii) holds when $\Gamma \subset \Gamma_u$, as $T_\Gamma \subset T_{\Gamma_u}$, while the set in (ii) is more sensitive to the choice of Γ. For the reader's convenience we give a different proof of both (ii) and (iii), in the spirit of [3] and based on semiconcavity estimates, in the general situation when Γ is contained in Γ_u. In the following we denote by T the set $T^l_{\Gamma_u} \setminus \Sigma$, where Σ is the $\mathcal{L}^n$-negligible Borel set where u is not differentiable.

Step 1. We show that $x \mapsto du(x)$ is $\mathcal{L}^n$-countably Lipschitz on T. Given a direction $\xi \in \mathbf{S}^{n-1}$ and $a \in \mathbb{R}$, let R be the union of the half closed maximal transport rays $[\![x, y[\![$ with $\langle y - x, \xi \rangle \geq 0$ and $\langle y, \xi \rangle \geq a$. It suffices to prove that the restriction of du to

$$R_a := R \cap \{x: \ x \cdot \xi < a\}$$

has the countable Lipschitz property stated in the theorem. To this aim, since $BV_{\rm loc}$ funtions have this property (see for instance Theorem 5.34 of [2] or [26]), it suffices to prove that ∇u coincides $\mathcal{L}^n$-a.e. in R_a with a suitable function $w \in [BV_{\rm loc}(S_a)]^n$, where $S_a = \{x: \ x \cdot \xi < a\}$. To this aim we define

$$\tilde{u}(x) := \min\{u(y) + \|x - y\|: \ y \in Y_a\}$$

where Y_a is the collection of all right endpoints of maximal transport rays with $y \cdot \xi \geq a$. By construction $\tilde{u} \geq u$ and equality holds on R_a.

We claim that, for $b < a$, $\tilde{u} - C|x|^2$ is concave in S_b for $C = C(b)$ large enough. Indeed, since $\|x - y\| \geq a - b > 0$ for any $y \in Y_a$ and any $x \in S_b$, the functions

$$x \mapsto u(y) + \|x - y\| - C|x|^2, \qquad\qquad y \in Y_a$$

are all concave in S_b for C large enough depending on $a - b$ (here we use the upper estimate in (23)). In particular, as gradients of real valued concave functions are $BV_{\rm loc}$ (see for instance [1]), we obtain that

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

is a $BV_{\rm loc}$ function in S_a. Since $\nabla u = w$ $\mathcal{L}^n$-a.e. in R_a the proof is achieved.

Now we notice that the duality map which associates to a unit vector $L \in (\mathbb{R}^n)^*$ the unique unit vector $v \in \mathbb{R}^n$ such that $L(v) = 1$ is Lipschitz (here we use the lower bound in (23)): indeed, setting $\phi(v) = \|v\|^2/2$, by the Lagrange multiplier rule we have $L + \lambda\nabla\phi(v) = 0$ for some $\lambda \in \mathbb{R}$ and evaluation at v gives $\lambda = -1/\langle\nabla\phi(v), v\rangle$ (because $\langle L, v\rangle = 1$), therefore

$$L = \frac{\nabla\phi(v)}{\langle\nabla\phi(v), v\rangle} = \frac{\nabla\phi(v)}{2\phi(v)} = \nabla\phi(v).$$

Since $v \mapsto \nabla\phi(v)$ is a strictly monotone operator its inverse is a Lipschitz map. It follows that $x \mapsto (du(x))^*$ is $\mathcal{L}^n$-countably Lipschitz on T.

Step 2. Now we show that the family L of left extreme points of maximal transport rays of Γ_u is Lebesgue negligible. As Γ_u is closed and

$$L = \bigcup_{(x,y)\in\Gamma_u} [\![x, y[\![\Big\backslash \bigcup_{(x,y)\in\Gamma_u}]\!]x, y[\![$$

we have that L is a Borel set. For any $x \in L \setminus \Sigma$ there is a unique maximal transport ray emanating from x, with direction $(du(x))^*$ and with length

$l(x)$. It is immediate to check that l is upper semicontinuous on $L \setminus \Sigma$, and therefore l is a Borel function. By Lusin theorem it will be sufficient to show that $\mathcal{L}^n(K) = 0$ for any compact set $K \subset L \setminus \Sigma$ where l and $(du)^*$ are continuous. We define

$$B := \bigcup_{x \in K} [\![x, x + \frac{l(x)}{2}(du(x))^*]\!] \setminus \Sigma$$

(B is Borel due to the continuity of l and of $(du)^*$) and we apply Theorem 9.4 with $\lambda = \mathcal{L}^n \llcorner K$ and $\tau(x) = (du(x))^*$ (notice that condition (iii) of the theorem holds because of Step 1 and the inclusion $B \subset T$) to get

$$\mathcal{L}^n(K) = \int_{\mathcal{S}_c(\mathbb{R}^n)} \lambda_C(K)\, d\mu(C) = 0$$

because $\lambda_C \ll \mathcal{H}^1 \llcorner C$ for μ-a.e. C and $K \cap C$ contains only one point for any closed maximal transport ray C.

Step 3. We show that $T^l_\Gamma \setminus (T_\Gamma \cup F_\Gamma)$ is Lebesgue negligible. Any point in this set is either a left extreme point of a maximal transport ray of Γ_u or is contained in T_{Γ_u}. Therefore, by Step 2, it suffices to consider only the set

$$R := T^l_\Gamma \cap T_{\Gamma_u} \setminus (T_\Gamma \cup F_\Gamma).$$

Since the intersection of R with any maximal transport ray of Γ_u is at most countable, by Remark 9.1 with $B = R$ and $\lambda = \mathcal{L}^n \llcorner R$ we obtain as in Step 2 that $\mathcal{L}^n(R) = 0$.

7 A stability result

In this section we assume that:
(I) μ, ν are probability measures in $\mathbb{R}^n$ with finite first order moments and $\mu \ll \mathcal{L}^n$;
(II) $c(x, y) = \|x - y\|$, where the norm $\|\cdot\|$ satisfies the regularity and uniform convexity conditions (23).

The main results of this section is the following existence and stability theorem. For $\varepsilon > 0$, we consider nondecreasing and strictly convex maps $\phi_\varepsilon : [0, +\infty) \to [0, +\infty)$ satisfying the following conditions:
(a) $\varepsilon \to \phi_\varepsilon(d)$ is convex and $\phi_\varepsilon(d) \to d$ as $\varepsilon \downarrow 0$ for any $d \geq 0$;
(b) the right derivative

$$\phi(d) := \lim_{\varepsilon \to 0^+} \frac{\phi_\varepsilon(d) - d}{\varepsilon}$$

exists and is a real-valued strictly convex function in $[0, +\infty)$ bounded from below.

Two model cases are $\phi_\varepsilon(d) = d + \varepsilon\phi(d)$, with ϕ strictly convex and bounded from below, or $\phi_\varepsilon(d) = d^{1+\varepsilon}$. In the latter case, $\phi(d) = d \ln d$.

Theorem 7.1. *Assume (I) and (II). Then:*

(i) The problem

$$\min\left\{\int_{\mathbb{R}^n} \|t(x)-x\|\,d\mu(x):\ t_{\#}\mu=\nu\right\} \tag{25}$$

has a solution.

(ii) Assume that

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} \phi_{\varepsilon_0}(\|x-y\|)\,d\gamma<\infty \qquad \text{for any } \gamma\in\Pi(\mu,\nu) \tag{26}$$

for some $\varepsilon_0>0$ and let, for $\varepsilon\in(0,\varepsilon_0)$, t_ε be the optimal maps in the problem

$$\min\left\{\int_{\mathbb{R}^n} \phi_\varepsilon\left(\|t(x)-x\|\right)\,d\mu(x):\ t_{\#}\mu=\nu\right\}. \tag{27}$$

Then $t_\varepsilon\to t$ in measure as $\varepsilon\to 0^+$ and t solves (25).

Proof. We need only to prove statement (ii). Indeed, choosing $\phi(t)+\sqrt{1+t^2}-1$, (26) holds for any $\varepsilon_0>0$ with $\phi_\varepsilon=Id+\varepsilon\phi$. Therefore we can apply statement (ii) to obtain an optimal transport map as limit as $\varepsilon\to 0^+$ of maps t_ε.

The proof of (ii) relies essentially on Theorem 7.2 below. Indeed, the variational argument of Proposition 7.1, based on the theory of Γ-asymptotic expansions, shows that any γ_0, limit point of $\gamma_\varepsilon=(Id\times t_\varepsilon)_{\#}\mu$ as $\varepsilon\to 0^+$, is an optimal planning for the Kantorovich problem and, in addition, minimizes the secondary variational problem

$$\gamma\mapsto\int_{\mathbb{R}^n\times\mathbb{R}^n}\phi(\|x-y\|)\,d\gamma$$

among all optimal plannings for the primary variational problem. Theorem 7.2 says that there is a unique such minimizer induced by a transport map t, i.e. $\gamma_0=(Id\times t)_{\#}\mu$.

This shows that problem (25) has a solution and that $(Id\times t_\varepsilon)_{\#}\mu\to(Id\times t)_{\#}\mu$ weakly as $\varepsilon\to 0^+$. Let now $\delta>0$ and choose a compact set $K\subset\mathbb{R}^n$ such that $t|_K$ is continuous and $\mu(\mathbb{R}^n\setminus K)<\delta$. Denoting by $\tilde t$ a continuous extension of $t|_K$ and choosing as test function

$$\varphi(x,y)=\chi_K(x)\times\psi\left(y-\tilde t(x)\right)$$

with $\psi\in C(\mathbb{R}^n,[0,1])$, $\psi(0)=0$ and $\psi(z)=1$ for $|z|\ge\delta$, we obtain

$$\begin{aligned}\limsup_{\varepsilon\to 0^+}\mu\left(\{|t_\varepsilon-t|>\delta\}\right)&\le\delta+\limsup_{\varepsilon\to 0^+}\mu\left(K\cap\{|t_\varepsilon-t|>\delta\}\right)\\&\le\delta+\limsup_{\varepsilon\to 0^+}\int\varphi\,d\gamma_\varepsilon\\&\le\delta+\int\varphi\,d(Id\times t)_{\#}\mu=\delta.\end{aligned}$$

Since $\delta > 0$ is arbitrary this proves the convergence in measure of t_ε to t.

Theorem 7.2. *Assume (I) and (II). Let $\Pi_1(\mu,\nu)$ be the collection of all optimal plannings in the primary problem*

$$\min\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\|\,d\gamma:\ \gamma\in\Pi(\mu,\nu)\right\} \tag{28}$$

and let $\phi:[0,+\infty)\to\mathbb{R}$ be a strictly convex function bounded from below. Let us consider the secondary variational problem

$$\min\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n} \phi(\|x-y\|)\,d\gamma:\ \gamma\in\Pi_1(\mu,\nu)\right\} \tag{29}$$

and let us assume that the minimum is finite. Then (29) has a unique solution and this solution is induced by a transport map t.

Proof. The existence of a solution γ_0 of the secondary variational problem is a direct consequence of the weak compactness of the class $\Pi_1(\mu,\nu)$. As c-concavity reduces to 1-Lipschitz continuity when the cost function is a distance, according to Theorem 3.3 there exists a function $u:\mathbb{R}^n\to\mathbb{R}$ which is 1-Lipschitz with respect to the distance induced by $\|\cdot\|$ and such that $\gamma\in\Pi_1(\mu,\nu)$ if and only if

$$\operatorname{spt}\gamma\subset\{(x,y)\in\mathbb{R}^n\times\mathbb{R}^n:\ u(x)-u(y)=\|x-y\|\}\,.$$

We define $\Gamma=\Gamma_u$ as in (24) and we wish to apply Theorem 6.1 to γ_0. Obviously assumption (i) of the theorem is satisfied, while assumptions (ii), (iii) follow by Theorem 6.2(ii) and Theorem 6.2(iii).

By Theorem 6.1 we obtain $\gamma_\#=(Id\times t)_\#\mu\in\Pi(\mu,\nu)$ such that

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}\psi(|x-y|)\,d\gamma_\#\le\int_{\mathbb{R}^n\times\mathbb{R}^n}\psi(|x-y|)\,d\gamma$$

for any convex function $\psi:[0,+\infty)\to\mathbb{R}$ bounded from below, with strict inequality if ψ is strictly convex. Choosing $\psi(t)=t$ we obtain that $\gamma_\#\in\Pi_1(\mu,\nu)$. Choosing $\psi=\phi$ the minimality of γ_0 in (29) gives $\gamma_0=\gamma_\#$, and therefore γ_0 is induced by a transport map.

The uniqueness of γ_0 is an easy consequence of the linear structure of the variational problems (28), (29): if $\gamma_0=(Id\times t)_\#\mu$ and $\gamma_0'=(Id\times t')_\#\mu$ are both optimal then $(\gamma_0+\gamma_0')/2$ is still optimal and therefore is induced by a transport map. This is possible only if $t=t'$ μ-a.e.

Remark 7.1. The secondary variational problem (29) can also be rephrased as follows: minimize

$$\gamma\mapsto\int_{\mathbb{R}^n\times\mathbb{R}^n}\tilde c(x,y)\,d\gamma$$

in $\Pi(\mu,\nu)$, where $\tilde c:\mathbb{R}^n\times\mathbb{R}^n\to[0,+\infty]$ is given by

$$\tilde{c}(x,y) = \begin{cases} \phi(\|x-y\|) & \text{if } \|x-y\| \le u(x)-u(y) \\ +\infty & \text{if } \|x-y\| > u(x)-u(y). \end{cases}$$

Indeed, the duality theory for the primary variational problem says that $\gamma \in \Pi_1(\mu,\nu)$ if and only if $\operatorname{spt}\gamma \subset \Gamma_u$.

Proposition 7.1. *Assume (I) and (26). For $\varepsilon \in (0,\varepsilon_0)$, let t_ε be the optimal maps in (27) and let $\gamma_\varepsilon = (Id \times t_\varepsilon)_\# \mu$ be the optimal plannings associated to t_ε. Then any limit point γ_0 of γ_ε is a minimizer of the secondary variational problem (29) and the minimum if finite.*

Proof. For $\gamma \in \Pi(\mu,\nu)$ and $\varepsilon \in (0,\varepsilon_0)$ we define

$$F_\varepsilon(\gamma) := \int_{\mathbb{R}^n\times\mathbb{R}^n} \phi_\varepsilon\left(\|x-y\|\right)\, d\gamma, \qquad F(\gamma) := \int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\|\, d\gamma.$$

Let $m = \min F$ and $F'_\varepsilon := (F_\varepsilon - m)/\varepsilon$. According to Theorem 4.1 it suffices to show that F_ε Γ-converge in $\Pi(\mu,\nu)$ to F and F'_ε Γ-converge in $\Pi(\mu,\nu)$ to

$$F'(\gamma) := \begin{cases} \displaystyle\int_{\mathbb{R}^n\times\mathbb{R}^n} \phi\left(\|x-y\|\right)\, d\gamma & \text{if } \gamma \in \Pi_1(\mu,\nu) \\ +\infty & \text{otherwise} \end{cases}$$

(here we consider any metric in $\Pi(\mu,\nu)$ inducing the convergence in the duality with $C_b(\mathbb{R}^n\times\mathbb{R}^n)$).

In order to show the Γ-convergence of F_ε to F, we notice that the convexity of $\varepsilon \mapsto \phi_\varepsilon(d)$ gives

$$\liminf_{\varepsilon\to 0^+} F_\varepsilon(\gamma_\varepsilon) \ge \liminf_{\varepsilon\to 0^+} \int_B \|y-x\| - \varepsilon\phi\left(\|y-x\|\right)\, d\gamma_\varepsilon \ge \int_B \|y-x\|\, d\gamma$$

for any family γ_ε weakly converging to γ and any bounded open set $B \subset \mathbb{R}^n\times\mathbb{R}^n$. Letting $B \uparrow \mathbb{R}^n\times\mathbb{R}^n$ the lim inf inequality follows.

The lim sup inequality with $\gamma_\varepsilon = \gamma$ is again a direct consequence of a convexity argument, which provides the estimate

$$F_\varepsilon(\gamma) - F(\gamma) \le \frac{\varepsilon}{\varepsilon_0}\left(F_{\varepsilon_0}(\gamma) - F(\gamma)\right).$$

Now we show the Γ-convergence of F'_ε, starting from the lim inf inequality. Let $\gamma_\varepsilon \to \gamma$ weakly and assume with no loss of generality that $\liminf_\varepsilon F'_\varepsilon(\gamma_\varepsilon)$ is finite. Then, the Γ-convergence of F_ε to F gives that $\gamma \in \Pi_1(\mu,\nu)$ and the convexity of the map $\varepsilon \mapsto \phi_\varepsilon(d)$ gives $\phi_\varepsilon(d) \ge d + \varepsilon\phi(d)$, so that

$$F_\varepsilon(\gamma_\varepsilon) - m \ge F_\varepsilon(\gamma_\varepsilon) - F(\gamma_\varepsilon) \ge \varepsilon F'(\gamma_\varepsilon).$$

As ϕ is continuous and bounded from below, dividing both sides by ε we obtain

$$\liminf_{\varepsilon\to 0^+} F'_\varepsilon(\gamma_\varepsilon) \geq \liminf_{\varepsilon\to 0^+} F'(\gamma_\varepsilon) \geq F'(\gamma).$$

In order to show the lim sup inequality we can assume with no loss of generality that $\gamma \in \Pi_1(\mu,\nu)$. Then

$$\frac{F_\varepsilon(\gamma)-m}{\varepsilon} = \int_{\mathbb{R}^n\times\mathbb{R}^n} \frac{\phi_\varepsilon(\|y-x\|)-\|y-x\|}{\varepsilon}\,d\gamma$$

and since $\varepsilon \mapsto (\phi_\varepsilon(d)-d)/\varepsilon$ is nondecreasing the dominated convergence theorem gives

$$\lim_{\varepsilon\to 0^+} \frac{F_\varepsilon(\gamma)-m}{\varepsilon} = F'(\gamma).$$

8 A counterexample

In the transport problem in the Euclidean space $\mathbb{R}^n$, the condition

$$\mu(B) = 0 \qquad \text{whenever} \qquad B \in \mathcal{B}(\mathbb{R}^n) \text{ and } \mathcal{H}^{n-1}(B) < \infty \tag{30}$$

ensures existence of optimal transport maps whenever ν has compact support and $c(x,y) = h(y-x)$, with h *strictly* convex and locally $C^{1,1}$ (see [30]). Condition (30) is sharp, as the following simple and well-known example shows:

Example 8.1. Let $I = [0,1]$, $\mu = \mathcal{H}^1 \llcorner \{0\} \times I$, $2\nu = \mathcal{H}^1 \llcorner \{-1\} \times I + \mathcal{H}^1 \llcorner \{1\} \times I$. Then, choosing $c(x,y) = |x-y|^\alpha$, with $\alpha > 0$, it is easy to check that $\gamma_x := (\delta_{(1,x_2)} + \delta_{(-1,x_2)})/2$ is the unique solution of the Kantorovich problem. Therefore the classical transport problem has no solution.

In this section we show that (30) does not provide in general existence of optimal transport maps when the cost function is the euclidean distance, building measures μ with dimension arbitrarily close to n such that the transport problem has no solution. This basically happens because in this case different maximal transport rays cannot cross in their interior (see (18)). Our construction provides also a counterexample to the statement made in the last page of [41] about the existence of optimal transport maps for measures μ such that $\mu(B_r(x)) = o(r^{n-1})$.

Lemma 8.1 (Horizontal transport rays). *Let* $I = [0,1]$ *and let* μ, ν *be probability measures in* $\mathbb{R}^2$ *with support respectively in* $I \times I$ *and* $[5,6] \times I$*. We assume that*

$$\mu([0,1]\times[0,t]) = \nu([5,6]\times[0,t]) \qquad \forall t \in I. \tag{31}$$

Then the optimal plannings move mass only along horizontal rays.

Proof. Let γ be an optimal planning relative to μ, ν. Assuming by contradiction that mass is not transported along horizontal rays, the number

$$\varepsilon := \sup \{|y_2 - x_2| : \ (x, y) \in \operatorname{spt} \gamma\}$$

is strictly positive and we can choose $(x, y) \in \operatorname{spt} \gamma$ such that $|y_2 - x_2| > \varepsilon/2$. We assume with no loss of generality (up to a reflection) that $y_2 < x_2$ and we prove by an elementary geometric argument the existence of $(x', y') \in \operatorname{spt} \gamma$ with $y'_2 - x'_2 > \varepsilon$, thus reaching a contradiction.

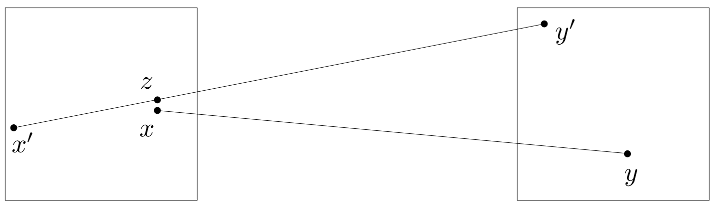

Fig. 2. Location of rays in Lemma 8.1

Applying the mass balance condition (31) with $t = (x_2 + y_2)/2$, we can find $(x', y') \in \operatorname{spt} \gamma$ such that $y'_2 \geq t \geq x'_2$ and, since the rays $[\![x, y]\!]$ and $[\![x', y']\!]$ cannot cross, there is $z \in [\![x', y']\!]$ with $z_1 = x_1$ and $z_2 > x_2$ as in Figure 2 (possibly exchanging the roles of μ and ν). By easy geometric arguments the following inequalities hold:

$$z_1 - x'_1 \leq 1, \qquad y'_1 - x'_1 \geq 4, \qquad z_2 - x'_2 \geq x_2 - t = \frac{x_2 - y_2}{2} > \frac{\varepsilon}{4}.$$

Then, since by similitude $\dfrac{y'_2 - x'_2}{z_2 - x'_2} = \dfrac{y'_1 - x'_1}{z_1 - x'_1}$, it follows $y'_2 - x'_2 > \varepsilon$, the searched contradiction.

Theorem 8.1. *For any $s \in (1, 2)$ there exists a continuous function $f : I \to I$ such that, setting*

$$\Gamma := \{(x_1, x_2) : \ x_1 = f(x_2)\}, \quad \mu := (f \times Id)_\# \mathcal{L}^1 \llcorner I, \quad \nu := \mathcal{L}^2 \llcorner ([5, 6] \times I)$$

the following properties hold:

(i) The Hausdorff dimension of Γ is s and $\mu \ll \mathcal{H}^t$ for any $t < s$;
(ii) the Kantorovich problem with data μ, ν and $c(x, y) = |x - y|$ has the unique solution $\gamma_x = \mathcal{H}^1 \llcorner \big([5, 6] \times \{x_2\}\big)$.

In particular μ satisfies (30) but the classical optimal transport problem has no solution.

Proof. (i) We use the classical construction given in Theorem 8.2 of [25]. Let $g : \mathbb{R} \to \mathbb{R}$ be the 4-periodic sawtooth function defined by

$$g(x) = \begin{cases} x & \text{if } 0 \le x \le 1 \\ 2 - x & \text{if } 1 \le x \le 3 \\ x - 4 & \text{if } 3 \le x \le 4 \end{cases} \tag{32}$$

and set

$$f(x) := \kappa + \sum_{i=1}^{\infty} \lambda_i^{s-2} g(\lambda_i x) \qquad x \in I$$

where $(\lambda_i) \subset (0, +\infty)$ are such that $\lambda_{i+1}/\lambda_i \to +\infty$ and $\ln(\lambda_{i+1})/\ln(\lambda_i) \to 1$ (for instance $\lambda_i = i!$). We choose $\kappa \in \mathbb{R}$ and normalize λ_i so that $0 \le f \le 1$. In [25] it is shown that there exists a constant $\delta > 0$ such that

$$|f(x) - f(y)| \le 6|x - y|^{2-s} \qquad \text{for } |x - y| \le \delta. \tag{33}$$

As a consequence, a simple covering argument (see Theorem 8.1 in [25]) gives

$$\mathcal{H}^s\left(\Gamma \cap Q_r(x)\right) \le c r^s \qquad \forall r \in (0, 2), \tag{34}$$

with $c = c(s, \delta)$, for any cube $Q_r(x)$ with side length r centered at $x \in \Gamma$. In particular $\mathcal{H}^s(\Gamma) < \infty$.

Another estimate still proved in [25] (see (8.12) on page 117) gives for any $t < s$ the existence of a constant $c_1 = c_1(t)$ such that

$$\mu\left(\Gamma \cap Q_r(x)\right) \le c_1 r^t \qquad \forall x \in \Gamma,\ r \in (0, 2). \tag{35}$$

It follows that $\mu \ll \mathcal{H}^t$ for any $t < s$. If $\mathcal{H}^t(\Gamma)$ were finite for some $t < s$ then $\mathcal{H}^{t'}(\Gamma)$ would be equal to 0 for $t' = (s + t)/2$, hence $\mu(\Gamma)$ would be zero. This contradiction proves that $\mathcal{H}^t(\Gamma) = +\infty$ for any $t < s$, hence the Hausdorff dimension of Γ is s.

(ii) The measures μ, ν satisfy by construction the identity (31). By Lemma 8.1 the support of γ_x is contained in $[5, 6] \times \{x_2\}$ for μ-a.e. x. Since

$$\nu = \int_{\mathbb{R}^2} \gamma_x \, d\mu(x) = \int_I \gamma_{(f(t),t)} \, dt$$

and the measures $\gamma_{(f(t),t)}$ are supported on $\mathbb{R} \times \{t\}$, the uniqueness of the disintegration of ν with respect to $t = x_2$ yields $\gamma_{(f(t),t)} = \mathcal{H}^1 \llcorner \big([5, 6] \times \{t\}\big)$ for a.e. $t \in I$.

9 Appendix: disintegration of measures

In this appendix we recall some basic facts about disintegration of measures, focussing for simplicity on the case of positive measures. In this section, unless otherwise stated, all spaces X, Y, Z we consider are locally compact and separable metric spaces.

Theorem 9.1 (Existence). *Let $\pi : X \to Y$ be a Borel map, let $\lambda \in \mathcal{M}_+(X)$ and set $\mu = \pi_\# \lambda \in \mathcal{M}_+(Y)$. Then there exist measures $\lambda_y \in \mathcal{M}_+(X)$ such that*

(i) $y \mapsto \lambda_y$ is a Borel map and $\lambda_y \in \mathcal{P}(X)$ for μ-a.e. $y \in Y$;
(ii) $\lambda = \lambda_y \otimes \mu$, i.e.

$$\lambda(A) = \int_Y \lambda_y(A)\, d\mu(y) \qquad \forall A \in \mathcal{B}(X); \tag{36}$$

(iii) λ_y is concentrated on $\pi^{-1}(y)$ for μ-a.e. $y \in Y$.

According to our terminology (which maybe is not canonical), a map $y \mapsto \lambda_y$ is Borel if $y \mapsto \lambda_y(B)$ is a Borel map in Y for any $B \in \mathcal{B}(X)$. This is equivalent (see for instance [2]) to the property that

$$y \mapsto \int_X \varphi(x, y)\, d\lambda_y(x)$$

is a Borel map in Y for any bounded Borel function $\varphi : X \times Y \to \mathbb{R}$. Our terminology is also justified by the observation that $y \mapsto \lambda_y$ is a Borel map in the conventional sense if we view $\mathcal{P}(X)$ as a subset of the compact metric of all positive Radon measures with total mass less than 1, endowed with the weak* topology coming from the duality with continuous and compactly supported functions in X. We will always make this embedding when we need to consider the space of probability measures as a measurable space.

The representation provided by Theorem 9.1 of λ can be used sometimes to compute the push forward of λ. Indeed,

$$f_\#(\lambda_y \otimes \mu) = f_\# \lambda_y \otimes \mu \tag{37}$$

for any Borel map $f : X \to Z$, where Z is any other metric space. Notice also that if $T : Y \to Z$ is a Borel and 1-1 map, $\mu' := T_\# \mu$ and $\lambda = \eta_z \otimes \mu'$, then

$$\lambda_y = \eta_{T(y)} \quad \text{for } \mu\text{-a.e. } y \in Y. \tag{38}$$

This is a simple consequence of the uniqueness of the disintegration, see Theorem 9.2 below.

The proof of Theorem 9.1 is available in many textbooks of measure theory or probability (in this case λ_y are the the so-called conditional probabilities induced by the random variable π, see for instance [2, 19]). In the case when $X = Y \times Z$ is a product space and $\pi(y, z) = y$ is the projection on the first variable the measures λ_y are concentrated on $\pi^{-1}(y) = \{y\} \times Z$, therefore it is often convenient to consider them as measures on Z, rather than measures on X, writing (36) in the form

$$\lambda(B) = \int_Y \lambda_y\left(\{z :\ (y, z) \in B\}\right) d\mu(y) \qquad \forall B \in \mathcal{B}(X). \tag{39}$$

We will always use this convention in the Kantorovich problem, writing each $\gamma \in \Pi(\mu,\nu) \subset \mathcal{P}(X \times Y)$ as $\gamma_x \otimes \mu$ with γ_x probability measures in Y.

Once the decomposition theorem is known in the special case $X = Y \times Z$ and $\pi(y,z) = z$ the general case can be easily recovered: it suffices to embed X into the product $Y \times X$ through the map $f(x) = (\pi(x), x)$ and to apply the decomposition theorem to $\tilde{\lambda} = f_\# \lambda$.

Now we discuss the uniqueness of λ_y and μ in the representation $\lambda = \lambda_y \otimes \mu$.

Theorem 9.2 (Uniqueness). *Let X, Y and π be as in Theorem 9.1; let $\lambda \in \mathcal{M}_+(X)$, $\mu \in \mathcal{M}_+(Y)$ and let $y \mapsto \eta_y$ be a Borel $\mathcal{M}_+(X)$-valued map defined on Y such that*

(i) $\lambda = \eta_y \otimes \mu$, i.e. $\lambda(A) = \int_Y \eta_y(A)\, d\mu(y)$ for any $A \in \mathcal{B}(X)$;
(ii) η_y is concentrated on $\pi^{-1}(y)$ for μ-a.e. $y \in Y$.

Then the η_y are uniquely determined μ-a.e. in Y by (i), (ii) and moreover, setting $C = \{y : \ \eta_y(X) > 0\}$, the measure $\mu \llcorner C$ is absolutely continuous with respect to $\pi_\# \lambda$. In particular

$$\frac{\mu \llcorner B}{\pi_\# \lambda} \eta_y = \lambda_y \qquad \text{for } \pi_\# \lambda\text{-a.e. } y \in Y \tag{40}$$

where λ_y are as in Theorem 9.1.

Proof. Let η_y, η'_y be satisfying (i), (ii). We have to show that $\eta_y = \eta'_y$ for μ-a.e. y. Let (A_n) be a sequence of open sets stable by finite intersection which generates the Borel σ-algebra of X. Choosing $A = A_n \cap \pi^{-1}(B)$, with $B \in \mathcal{B}(Y)$, in (i) gives

$$\int_B \eta_y(A_n)\, d\mu(y) = \int_B \eta'_y(A_n)\, d\mu(y).$$

Being B arbitrary, we infer that $\eta_y(A_n) = \eta'_y(A_n)$ for μ-a.e. y, and therefore there exists a μ-negligible set N such that $\eta_y(A_n) = \eta'_y(A_n)$ for any $n \in \mathbf{N}$ and any $y \in Y \setminus N$. By a well-know coincidence criterion for measures (see for instance Proposition 1.8 of [2]) we obtain that $\eta_y = \eta'_y$ for any $y \in Y \setminus N$.

Let $B' \subset B$ be any $\pi_\# \lambda$-negligible set; then $\pi^{-1}(B')$ is λ-negligible and therefore (ii) gives

$$0 = \int_Y \eta_y \left(\pi^{-1}(B')\right)\, d\mu(y) = \int_{B'} \eta_y(X)\, d\mu(y).$$

As $\eta_y(X) > 0$ on $B \supset B'$ this implies that $\mu(B') = 0$. Writing $\mu \llcorner B = h \pi_\# \lambda$ we obtain $\lambda = h\eta_y \otimes \pi_\# \lambda$ and $\lambda = \lambda_y \otimes \pi_\# \lambda$. As a consequence (40) holds.

In the following proposition we address the delicate problem, which occurs in optimal transport problems, of the measurability of maps f obtained by "gluing" different transport maps defined on the level sets of π. Simple examples show that the λ_y-measurability of f, though necessary, it is not sufficient:

for instance when $X = Y \times Z$ is a product space, π is the projection on the first factor and λ is concentrated on the graph of $\phi : Y \to Z$, then λ_y are Dirac masses concentrated at $(y, \phi(y))$ for μ-a.e. y, therefore λ_y-measurability provides no information on λ-measurability or, rather, on the existence of a Borel map g such that $g = f$ λ_y-a.e. for μ-a.e. $y \in Y$.

In order to state our measurability criterion we need some more terminology. Given a λ_y-measurable function $f : X \to Z$, we canonically associate to f the measure $\gamma_y(f) := (Id \times f)_\# \lambda_y$, a probability measure in $X \times Z$. It turns out that the measurability of the map $y \mapsto \gamma_y$ is sufficient to provide a Borel map g equivalent to f, i.e. such that $g = f$ λ_y-a.e. for μ-a.e. y.

Theorem 9.3 (Measurability criterion). *Keeping the notation of Theorem 9.1, let $f : X \to Z$ be satisfying the following two conditions:*

(i) f is λ_y-measurable for μ-a.e. y;
(ii)$y \mapsto \gamma_y(f)$ is a μ-measurable map between Y and $\mathcal{P}(X \times Z)$.

Then there exists a Borel map $g : X \to Z$ such that $g = f$ λ_y-a.e. for μ-a.e. y.

Proof. For the sake of simplicity we consider only the case when X, Y, Z are compact. Given $\nu \in \mathcal{M}_+(X)$ we define a metric on the space $L(X, \nu; Z)$ of ν-measurable maps between X and Z by

$$d_\nu(f, g) := \int_X d_Z\left(f(x), g(x)\right) d\nu(x).$$

It is well known that this metric induces the convergence in ν-measure and that $L(X, \nu; Z)$ is a complete metric space (with the canonical equivalence relation between ν-measurable maps).

Without loss of generality (by Lusin's theorem, see Theorem 2.3.5 in [26]) we can assume that:
(a) π is continuous (so that $\operatorname{spt} \lambda_y \subset \pi^{-1}(y)$ are pairwise disjoint).
(b) $y \mapsto \operatorname{spt} \lambda_y$ is continuous between Y and the class $\mathcal{K}$ of closed subsets of X, endowed with the Hausdorff metric. Indeed, the σ-algebra of the Borel subsets of $\mathcal{K}$ is generated by the sets $\{K : \ K \cap U \neq \emptyset\}$, for $U \subset X$ open, and

$$\{y : \ \operatorname{spt} \lambda_y \cap U \neq \emptyset\} = \{y : \ \lambda_y(U) > 0\} \in \mathcal{B}(Y).$$

(c) $y \mapsto \lambda_y$ is continuous.
Step 1. We assume first that the restriction of f to $\operatorname{spt} \lambda_y$ is a M-Lipschitz function with values in Z for some $M \geq 0$ independent of y. By Lusin's theorem again we can find an increasing family of compact sets $Y_h \subset Y$ whose union covers μ-almost all of Y and such that $\gamma_y(f)$ restricted to Y_h is continuous for any h.

We claim now that the restriction of f to the compact set (compactness comes from assumption (b))

$$X_h := \bigcup_{y \in Y_h} \operatorname{spt} \lambda_y$$

is continuous. Indeed, assume that $x_k \in \operatorname{spt} \lambda_{y_k}$ converge to $x \in \operatorname{spt} \lambda_y$ (with $y_k,\, y \in Y_h$) but $d_Y\left(f(x_k), f(x)\right) \geq \epsilon$ for some $\epsilon > 0$. Then the equi-Lipschitz condition provides $r > 0$ such that $d_Y\left(f(z), f(w)\right) \geq \epsilon/2$ for any choice of $z \in B_r(x) \cap \operatorname{spt} \lambda_{y_k}$, $w \in B_r(x) \cap \operatorname{spt} \lambda_y$. Choosing a test function χ of the form $\chi_1(x)\chi_2(z)$, with $\operatorname{spt} \chi_1 \subset B_r(x)$, $\operatorname{spt} \chi_2 \subset B_{\epsilon/2}(f(x))$ and $\chi_2(f(x)) = 1$ we find

$$\lim_{k \to +\infty} \gamma_{y_k}(\chi) = 0 < \gamma_y(\chi),$$

contradicting the continuity of $y \mapsto \gamma_y$ on Y_h.

As the union of X_h covers λ-almost all of X we obtain that f is λ-measurable.

Step 2. Now we attack the general case. For any h and any $y \in Y$ we consider the set $K_h(y)$ of all probability measures in $X \times Z$ of the form $(Id \times f)_\# \lambda_y$, with $f : X \to Z$ with Lipschitz constant less than h. By assumption (c) the multifunction $K_h(y)$ has a closed graph in $Y \times \mathcal{P}(X \times Z)$ and therefore, according to Proposition 9.1, we can find a μ-measurable map $y \mapsto (Id \times f^h_y)_\# \lambda_y$ such that

$$d\left((Id \times f)_\# \lambda_y, (Id \times f^h_y)_\# \lambda_y\right) = \operatorname{dist}\left((Id \times f)_\# \lambda_y, K_h(y)\right) \qquad \text{for } \mu\text{-a.e. } y.$$

Defining f^h in such a way that $f^h = f^h_y$ on $\operatorname{spt} \lambda_y$, by Step 1 we have that f^h is λ-measurable. Moreover

$$\lim_{h \to +\infty} (Id \times f^h)_\# \lambda_y = (Id \times f)_\# \lambda_y$$

for μ-a.e. y, hence (see the simple argument in the end of the proof of Theorem 7.1) we obtain that (f^h) converges in $L(X, \lambda_y; Z)$ to f for μ-a.e. y. As $\lambda = \lambda_y \otimes \mu$ we obtain that (f^h) is a Cauchy sequence in $L(X, \lambda; Z)$. Denoting by g a Borel limit function, we can find a subsequence $h(k)$ such that

$$\int_Y \sum_{k=1}^{\infty} d_{\lambda_y}\left(f^{h(k)}, g\right)\, d\mu = \sum_{k=1}^{\infty} d_\lambda\left(f^{h(k)}, g\right) < +\infty.$$

Therefore $f^{h(k)}$ converge in $L(X, \lambda_y; Z)$ to g for μ-a.e. y and $g = f$ λ_y-a.e. for μ-a.e. y.

The proof of the following measurable selection result is available for instance in [17].

Proposition 9.1 (Measurable selection). *Let $\lambda \in \mathcal{M}_+(X)$ and let $f : X \to Y$ be λ-measurable. Assume that $x \mapsto \Gamma(x)$ is a multifunction which associates to any $x \in X$ a compact and nonempty subset of Y. If the graph of Γ (i.e. $\{(x, y) :\ y \in \Gamma(x)\}$) is closed, there exists a λ-measurable map $g : X \to Y$ such that $g(x) \in \Gamma(x)$ and*

$$d_Y\left(f(x), g(x)\right) = \operatorname{dist}_Y\left(f(x), \Gamma(x)\right) \quad \text{for } \lambda\text{-a.e. } x \in X.$$

In the applications to transport problems in Euclidean spaces the typical situation occurs with $B \in \mathcal{B}(\mathbf{R}^n)$ and $\pi : B \to \mathcal{S}_c(\mathbf{R}^n)$, where π satisfies with $x \in \pi(x)$ for any x. By disintegrating a measure λ concentrated on B along the level sets $\pi^{-1}(C)$ (contained in C and therefore 1-dimensional), in order to apply the 1-dimensional theory we would like to find conditions ensuring that the disintegrated measures λ_C have no atom. Although no sharp condition seems to be known, it can be shown that the absolute continuity of λ_C is inherited from λ provided the family of segments $\pi(B)$ is countably Lipschitz. The proof below is taken from [3], where this problem is discussed more in detail (see Remark 6.1 therein).

Theorem 9.4 (Absolute continuity). *Let $B \in \mathcal{B}(\mathbf{R}^n)$, let $Y = \mathcal{S}_c(\mathbf{R}^n)$ and let $\pi : B \to Y$ be a Borel map satisfying the conditions*

(i) If $\pi(x) \neq \pi(x')$ then the intersection $\pi(x) \cap \pi(x')$ can contain at most the initial point of $\pi(x)$ and of $\pi(x')$ and this point is not in B.
(ii) $x \in \pi(x)$ for any $x \in B$.
(iii) The direction $\tau(x)$ of $\pi(x)$ is a $\mathbf{S}^{n-1}$-valued countably Lipschitz map on B, i.e. there exist sets $B_h \subset B$ whose union contains B and such that $\tau|_{B_h}$ is a Lipschitz map for any h.

Then, for any measure $\lambda \in \mathcal{M}_+(\mathbf{R}^n)$ absolutely continuous with respect to $\mathcal{L}^n \llcorner B$, setting $\mu = \pi_\# \lambda \in \mathcal{M}_+(Y)$, the measures λ_C of Theorem 9.1 are absolutely continuous with respect to $\mathcal{H}^1 \llcorner C$ for μ-a.e. $C \in Y$.

Proof. Being the property stated stable under countable disjoint unions we may assume that

(a) there exists a unit vector ξ such that $\tau(x) \cdot \xi \geq \frac{1}{2}$ for any $x \in B$;
(b) $\tau(x)$ is a Lipschitz map on B;
(c) B is contained in a strip

$$\{x : \ a - b \leq x \cdot \xi \leq a\}$$

with $b > 0$ sufficiently small (depending only on the Lipschitz constant of ν) and $\pi(x)$ intersects the hyperplane $\{x : \ x \cdot \xi = a\}$.

Assuming with no loss of generality $\xi = e_n$ and $a = 0$, we write $x = (y, z)$ with $y \in \mathbf{R}^{n-1}$ and $z < 0$. Under assumption (a), the map $T : \pi(B) \to \mathbf{R}^{n-1}$ which associates to any segment $\pi(x)$ the vector $y \in \mathbf{R}^{n-1}$ such that $(y, 0) \in \pi(x)$ is well defined. Moreover, by condition (i), T is one to one. Hence, setting $f = T \circ \pi : B \to \mathbf{R}^{n-1}$,

$$\nu := T_\# \mu = f_\# \lambda, \qquad\qquad C(y) := T^{-1}(y) \supset f^{-1}(y)$$

and representing $\lambda = \eta_y \otimes \nu$ with $\eta_y = \lambda_{C(y)} \in \mathcal{M}_1(f^{-1}(y))$ (see (38)), we need only to prove that $\eta_y \ll \mathcal{H}^1 \llcorner C(y)$ for ν-a.e. y.

To this aim we examine the Jacobian, in the y variables, of the map $f(y, t)$. Writing $\tau = (\tau_y, \tau_t)$, we have

$$f(y,t) = y + d(y,t)\tau_y(y,t) \qquad \text{with} \qquad d(y,t) = -\frac{t}{\tau_t(y,t)}.$$

Since $\tau_t \geq 1/2$ and $d \leq 2b$ on B we have

$$\det\left(\nabla_y f(y,t)\right) = \det\left(Id + d\nabla_y \tau_y + \frac{t}{\tau_t^2}\nabla_y \tau_t \otimes \tau_y\right) > 0$$

if b is small enough, depending only on the Lipschitz constant of τ.

Therefore, the coarea factor

$$\mathbf{C}f := \sqrt{\sum_A \det{}^2 A}$$

(where the sum runs on all $(n-1)\times(n-1)$ minors A of ∇f) of f is strictly positive on B and, writing $\lambda = g\mathcal{L}^n$ with $g = 0$ out of B, Federer's coarea formula (see for instance [2], [26], [38]) gives

$$\lambda = \frac{g}{\mathbf{C}f}\mathbf{C}f\mathcal{L}^n = \frac{g}{\mathbf{C}f}\mathcal{H}^1 \llcorner f^{-1}(y) \otimes \mathcal{L}^{n-1} = \eta'_y \otimes \nu'$$

and

$$\eta'_y := \frac{\frac{g}{\mathbf{C}f}\mathcal{H}^1 \llcorner f^{-1}(y)}{\int_{f^{-1}(y)} g/\mathbf{C}f\, d\mathcal{H}^1}, \qquad \nu' := \left(\int_{f^{-1}(y)} \frac{g}{\mathbf{C}f}\, d\mathcal{H}^1\right)\mathcal{L}^{n-1} \llcorner L$$

with $L := \left\{y \in \mathbf{R}^{n-1} :\ \mathcal{H}^1(f^{-1}(y)) > 0\right\}$.

By Theorem 9.2 we obtain $\nu = \nu'$ and $\eta_y = \eta'_y$ for ν-a.e. y, and this concludes the proof.

Remark 9.1. As the proof clearly shows, the statement is still valid for maps $\pi : B \to \mathcal{S}_o(\mathbb{R}^n)$ provided condition (i) is replaced by the simpler condition that $\pi(x) \cap \pi(x') = \emptyset$ whenever $\pi(x) \neq \pi(x')$.

References

1. G. Alberti & L. Ambrosio: *A geometric approach to monotone functions in* $\mathbb{R}^n$. Math. Z., **230** (1999), 259–316.
2. L. Ambrosio, N. Fusco & D. Pallara: *Functions of Bounded Variation and Free Discontinuity Problems.* Oxford University Press, 2000.
3. L. Ambrosio: *Lecture Notes on the Optimal Transport Problems.* Notes of a CIME Course given in Madeira (2000), to be published in the CIME Springer Lecture Notes.
4. G. Alberti, B. Kirchheim & D. Preiss: Personal communication.
5. L. Ambrosio, B. Kirchheim & A. Pratelli: *Existence of optimal transports with crystalline norms.* In preparation.

6. L.Ambrosio & S.Rigot: *Optimal mass transportation in the Heisenberg group.* In preparation.
7. G.Anzellotti & S.Baldo: *Asymptotic development by Γ-convergence.* Appl. Math. Optim., **27** (1993), 105–123.
8. G.Anzellotti, D.Percivale & S.Baldo: *Dimension reduction in variational problems, asymptotic developments in Γ-convergence and thin structures in elasticity.* Asymptotic Anal., **9** (1994), 61–100.
9. G.Bouchitté, G.Buttazzo & P.Seppecher: *Shape optimization via Monge–Kantorovich equation.* C.R. Acad. Sci. Paris, **324-I** (1997), 1185–1191.
10. G.Bouchitté & G.Buttazzo: *Characterization of optimal shapes and masses through Monge–Kantorovich equation.* J. Eur. Math. Soc., **3** (2001), 139–168.
11. G.Bouchitté & G.Buttazzo: In preparation.
12. Y.Brenier: *Décomposition polaire et réarrangement monotone des champs de vecteurs.* C.R. Acad. Sci. Paris, Sér I Math., **305** (1987), 805–808.
13. Y.Brenier: *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math., **44** (1991), 375–417.
14. Y.Brenier: *Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations.* Comm. Pure Appl. Math., **52** (1999), 411–452.
15. L.Caffarelli: *Boundary regularity of maps with a convex potential.* Commun. Pure Appl. Math., **45** (1992), 1141–1151.
16. 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.
17. C.Castaing & M.Valadier: *Convex analysis and measurable multifunctions.* Lecture Notes in Mathematics **580**, Springer, 1977.
18. G.Dal Maso: *An Introduction to Γ-Convergence*, Birkhäuser, 1993.
19. C.Dellacherie & P.Meyer: *Probabilities and potential.* Mathematical Studies **29**, North Holland, 1978.
20. 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.
21. L.De Pascale, L.C. Evans & A.Pratelli: In preparation.
22. H.Dietrich: *Zur c-konvexität und c-Subdifferenzierbarkeit von Funktionalen.* Optimization, **19** (1988), 355–371.
23. L.C.Evans: *Partial Differential Equations and Monge–Kantorovich Mass Transfer.* Current Developments in Mathematics, 1997, 65–126.
24. L.C.Evans & W.Gangbo: *Differential Equation Methods for the Monge-Kantorovich Mass Transfer Problem.* Memoirs AMS, **653**, 1999.
25. K.J.Falconer: *The geometry of fractal sets*, Cambridge University Press, 1985.
26. H.Federer: *Geometric measure theory.* Springer, 1969.
27. M.Feldman & R.McCann: *Uniqueness and transport density in Monge's mass transportation problem.* 2000, to appear on Calc. Var.
28. M.Feldman & R.McCann: *Monge's transport problem on a Riemannian manifold.* Trans. Amer. Mat. Soc., **354** (2002), 1667–1697.
29. W.Gangbo: *An elementary proof of the polar factorization theorem for functions.* Arch. Rat. Mech. Anal., **128** (1994), 381–399.
30. W.Gangbo & R.J.McCann: *The geometry of optimal transportation.* Acta Math., **177** (1996), 113–161.

31. L.V.Kantorovich: *On the transfer of masses.* Dokl. Akad. Nauk. SSSR, **37** (1942), 227–229.
32. L.V.Kantorovich: *On a problem of Monge.* Uspekhi Mat. Nauk., **3** (1948), 225–226.
33. 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.
34. R.McCann: *Polar factorization of maps on Riemannian manifolds.* Geom. Funct. Anal., **11** (2001), 589–608.
35. G.Monge: *Memoire sur la Theorie des Déblais et des Remblais.* Histoire de l'Acad. des Sciences de Paris, 1781.
36. S.T.Rachev & L.Rüschendorf: *Mass transportation problems.* Vol I: Theory, Vol. II: Applications. Probability and its applications, Springer, 1998.
37. L.Rüschendorf: *On c-optimal random variables.* Statistics & Probability Letters, **27** (1996), 267–270.
38. L.Simon: *Lectures on geometric measure theory*, Proc. Centre for Math. Anal., Australian Nat. Univ., **3**, 1983.
39. C.Smith & M.Knott: *On the optimal transportation of distributions.* J. Optim. Theory Appl., **52** (1987), 323–329.
40. C.Smith & M.Knott: *On Hoeffding–Fréchet bounds and cyclic monotone relations.* J. Multivariate Anal., **40** (1992), 328–334.
41. V.N.Sudakov: *Geometric problems in the theory of infinite dimensional distributions.* Proc. Steklov Inst. Math., **141** (1979), 1–178.
42. N.S.Trudinger & X.J.Wang: *On the Monge mass transfer problem.* Calc. Var. PDE, **13** (2001), 19–31.
43. J.Urbas: *Mass transfer problems*, 1998.
44. C.Villani: *Topics in mass transportation.* Forthcoming book by AMS.
45. L.C.Young: *Generalized curves and the existence of an attained absolute minimum in the calculus of variations.* Comptes Rendus Soc. Sci. et Lettres Varsovie, **30** (1937), 212–234.
46. L.C.Young: *Generalized surfaces in the Calculus of Variations.* Ann. of Math., **43** (1942), 84–103.
47. L.C.Young: *Generalized surfaces in the Calculus of Variations II.* Ann. of Math., **43** (1942), 530-544.