ON THE EQUALITY BETWEEN MONGE'S INFIMUM AND KANTOROVICH'S MINIMUM IN OPTIMAL MASS TRANSPORTATION

ALDO PRATELLI

Abstract. This paper concerns the Monge's transport problem in a general Polish space. We find optimal conditions to establish the existence of classic transport maps and the equality between the infimum of Monge's problem and the minimum of the Kantorovich's relaxed version of the problem. As a byproduct, we can prove a stronger version of the classic Isomorphism Theorem of measure rings. A preliminary version of the results of this paper is contained in the Ph.D. thesis [11].

1. Introduction

1.1. **Setting of the problem and statement of the main results.** In 1781, G. Monge [10] first proposed the mass transport problem. Today, one can restate his idea as follows: we are given two Polish spaces X and Y and two measures $f^+ \in \mathcal{M}^+(X)$ and $f^- \in \mathcal{M}^-(Y)$ of the same total mass $\|f^+\| = \|f^-\|$ (we will recall all the standard definitions in Appendix). A *transport map* from f^+ to f^- is any function $t : X \to Y$ such that $t_\# f^+ = f^-$ (of course, t need to be measurable with respect to the $\sigma-$algebra of the f^+-measurable sets on X and to the Borel $\sigma-$algebra on Y); we are also given a l.s.c. function $c : X \times Y \to \overline{R}^+ := \mathbb{R}^+ \cup \{+\infty\}$, and to each transport map one associates the cost

$$\mathfrak{C}(t) := \int_X c\bigl(x, t(x)\bigr)\, df^+(x). \tag{1.1}$$

The interpretation of this problem as a transportation of mass is very easy: if f^+ and f^- measure the height of some soil and the depth of some hole of the same volume, a transport map $t : X \to Y$ says how to move the mass into the hole, and $\mathfrak{C}(t)$ represents the cost of carrying out this transportation if $c(x, y)$ is the cost of moving a unitary mass from the point $x \in X$ to the point $y \in Y$.

In general, it may happen that there are no optimal transport maps (i.e. minimizers of $\mathfrak{C}$ among the transport maps): for instance, if $X = Y = \mathbb{R}^N$ and $c(x, y) = |y - x|$ is the Euclidean distance, the existence of optimal transport maps is true only under the hypothesis that f^+ is absolutely continuous w.r.t. the Lebesgue measure (see [1], or also [5], [3], [14], [2], [12] or [11] for general references). Even worse, it may happen that there are not transport maps at all: for instance, if f^+ is a Dirac mass in the point $x \in X$, then $t_\# f^+$ is the Dirac mass in the point $t(x) \in Y$, therefore there is no transport map at all if f^- is not a Dirac mass.

In 1940's, Kantorovich proposed [7, 8] a relaxed version of the problem, that is of fundamental importance to understand it and to find solutions. His idea, roughly speaking, is to allow the splitting of mass, removing the obstacle in the example with the Dirac mass we mentioned above: to do that, one defines *transport plan* between f^+ and f^- any probability measure $\gamma \in \mathcal{M}^+(X \times Y)$ whose marginals $\pi_X \gamma$ and $\pi_Y \gamma$ are f^+ and f^- respectively. Any such measure can be interpreted as a way to transport mass if $\gamma(A \times B)$ represents the amount of mass contained in the set $A \subseteq X$ and to be moved inside

$B \subseteq Y$. Analogously to (1.1), one defines a cost for any transport plan as

$$\mathfrak{C}(\gamma) := \iint_{X\times Y} c(x,y)\, d\gamma(x,y). \tag{1.2}$$

It is to be noticed that the transport plans are a generalization of the transport maps: indeed, the map t is "naturally" associated with the plan $\gamma_t := (1,t)_\# f^+$, and also the cost is the same. More in general, a transport plan γ "is" a transport map –i.e. there is a transport plan t such that $\gamma = \gamma_t$– if and only if γ is concentrated in a graph: formally, the following result holds.

Lemma 1.1. *A transport plan γ is induced by a transport map if and only if γ is concentrated in a $\sigma-$compact set S such that for f^+-a.e. x the set $\{y:\,(x,y)\in S\}$ consists of exactly a point.*

Of course an implication is trivial, the other one is quite easy but not completely straightforward: for a formal proof of this fact one can refer to [1], or to Lemma 1.6 in [11] for the general case of the Polish spaces. Since the transport maps are a particular case of the transport plans, of course one has

$$\min(1.2) \leq \inf(1.1) \tag{1.3}$$

(the fact that the first is a minimum will be discussed in a moment). One can easily notice that the relaxed problem is much simpler: in fact, while the cost (1.1) depends in an involved way by the map t and the set of the transport maps has no good structure, the cost (1.2) depends linearly by the plan γ, and the transport plans are a bounded a weak* closed subset of $\mathcal{M}^+(X\times Y)$. Moreover, since the function $c: X\times Y$ is l.s.c., so is the function $\gamma \mapsto \mathfrak{C}(\gamma)$: therefore, while –as we noticed before– there could be no optimal transport maps or no transport maps at all, the following is true for the transport plans.

Lemma 1.2. *There always exits transport plans; moreover, there always exist optimal transport plans.*

Concerning the existence of transport plans, just notice that the tensor product $f^+ \boxtimes f^-$ –see the definition (A.1) in the Appendix– is clearly a transport plan. For the existence of optimal transport plans, it is a consequence of the boundedness and the weak* closure of the set of the transport plans, as well as of the lower semicontinuity of $\mathfrak{C}$: this is immediately obtained if X and Y are compact, the general case of the Polish spaces is standard and well known, see for example Theorem 1.11 in [11] or [12].

In this paper we will give an answer to two very natural questions. The example above with the Dirac mass tells that the existence of transport maps is not always true, then the first question is to establish sufficient conditions for the existence of some transport maps: since it is evident that, as it happened in the example above, the presence of Dirac masses can prevent this existence, the following result is clearly sharp.

Theorem A. *If f^+ is non-atomic there are transport maps. In addition, we can find a transport map t such that*

$$\mathfrak{C}(t) \leq \mathfrak{C}(f^+ \boxtimes f^-) \quad (\leq +\infty). \tag{1.4}$$

The second question concerns the inequality (1.3): in fact, it is of great importance to know whether or not it is an equality, also because of course otherwise one cannot hope the relaxed problem to give some help in studying the original one. The second main result we will prove (that is also sharp thanks to the examples we will provide at the beginning of Section 4) states that a sufficient condition is the non-atomicity of f^+ together with the continuity of c.

Theorem B. *Assume that f^+ is non atomic and c is continuous (but still possibly $+\infty-$valued somewhere): then inf(1.1)=min(1.2) (possibly $+\infty$).*

Two particular cases of our Theorems have been proved by Gangbo in [6] and by Ambrosio in [1]:

Theorem 1.3 (Gangbo)**.** *If X and Y are Polish spaces and both f^+ and f^- are non-atomic, then there exist transport maps.*

Theorem 1.4 (Ambrosio)**.** *If $X=Y$ is a convex and compact subset of $\mathbb{R}^N$ and f^+ is non-atomic, then there exist transport maps. In particular, if the cost function $c : X\times Y \to \overline{R}^+$ is continuous and bounded, then the equality inf(1.1)=min(1.2) holds.*

Let us briefly discuss the main differences between Theorems A and B and the previous Theorems 1.3 and 1.4: concerning the existence of a trasport map, of course it does not depend on the particular choice of a cost function; indeed, the most important part of Theorem A is the estimate (1.4). Concerning the equality inf(1.1)=min(1.2), we remark that Theorem 1.4 was stated in $\mathbb{R}^N$ only for simplicity; however, the proof works in a more general compact Polish space. Therefore, the main new feature of Theorem B is to allow non-compact spaces and a cost function that is not bounded, and also possibly assuming the value $+\infty$ somewhere.

The strategy to prove Theorem 1.4 was to show that, under the hypotheses of the claim, the subset of the transport plans made by the transport maps is dense with respect to the weak* convergence of measures; notice that this fact does not depend on the choice of a cost function. As a consequence, to any transport plan one can associate a weak*-converging sequence of transport maps: therefore, given any continuous and bounded cost function c, the costs of the transport maps converge to the cost of the transport plan, and this gives the thesis. On the other hand, to show Theorem B one needs a more subtle construction which depends in a substantial way on the choice of the cost function: given a transport plan γ and given the cost function c, we will build transport maps t_ε with $\mathfrak{C}(t_\varepsilon)\leq \mathfrak{C}(\gamma)+\varepsilon$; but the same transport maps may not work with another cost function.

The plan of the paper is the following: in Subsection 1.2 we discuss the relation between Theorem B and the Isomorphism Theorem of measure rings; in Section 2 we prove a general result valid in Polish spaces that is the main ingredient for our following constructions. In Sections 3 and 4 we show Theorems A and B; finally, in Section 5 we will prove a stronger version of the Isomorphism Theorem.

1.2. **The Isomorphism Theorem.** In [6], with the aim of establishing existence of transport maps, W. Gangbo suggested the idea of making use of the classic Isomorphism Theorem of measure rings. We rewrite the statement of this result (see for instance the classical book [13], Theorem 9 in Chapter 15):

Theorem 1.5 (Isomorphism Theorem of measure rings)**.** *Let X be a Polish space and μ a non-atomic probability measure on X. Then (X,μ) is isomorphic to $([0,1],\mathscr{L})$: we mean that there are a Borel subset $\widetilde{X}$ of X of full $\mu-$measure, a Borel subset $\widetilde{[0,1]}$ of $[0,1]$ of full $\mathscr{L}-$measure, and a bijective function $\varphi:\widetilde{X}\to\widetilde{[0,1]}$ with the property that both φ and φ^{-1} map Borel sets in Borel sets, and $\varphi_\#\mu=\mathscr{L}$, $\varphi^{-1}_\#\mathscr{L}=\mu$.*

The proposal of Gangbo was to use this result to reduce oneself to a known setting: in other words, to show the existence of a transport map from (X,f^+) to (Y,f^-), one could use the isomorphisms given by the above Theorem to reduce to the case when X and Y are $[0,1]$ (or any other canonical space), apply the well known (and trivial) existence result in $[0,1]$ to find a transport map between $[0,1]$ and itself, and then use the isomorphisms again to come back to a transport map between X and Y. The proof of Theorem 1.3 is basically obtained through this idea.

Unfortunately, the same argument cannot be used to derive also the particular case of Theorem B when both f^+ and f^- are non-atomic: indeed, the result of equality between inf(1.1) and min(1.2) is true only under a continuity assumption on the cost function (the reason for this fact will be extensively discussed in the beginning of Section 4). But, since the map φ in Theorem 1.5 tipically fails to be continuous, passing from X and Y to $[0,1]$ the continuous cost function $c : X \times Y \to \overline{R}^+$ may become a discontinuous cost function from $[0,1] \times [0,1]$ to $\overline{R}^+$.

As a byproduct of our general Proposition 2.2, we will be able in Section 5 to give a simple proof of Theorem C, a stronger version of the Isomorphism Theorem which states also the bicontinuity of a map φ as in Theorem 1.5. However, we like to point the readers' attention to the fact that also this stronger claim cannot be use to derive a particular case of Theorem B: indeed, the map φ is not bicontinuous between X and $[0,1]$ (which is of course not possible in general), but only between $\widetilde{X}$ and $\widetilde{[0,1]}$; therefore, since we have to deal with continuous cost functions, one could not reduce himself to the easy case of the ambient space $[0,1]$, but only to the very general case of a subset of $[0,1]$.

2. A general result about Polish spaces

This section is devoted to show how one can decompose a Polish space endowed with a probability measure; this decomposition helps to regard the Polish space as $\{0,1\}^{\mathbb{N}}$, that is a very standard reduction in probability (see for instance [9]). First, we fix some notations for the binary sequences that we will use extensively throughout the paper.

Definition 2.1. *A* finite sequence *(resp. a* finite binary sequence*) is any $I \in \mathbb{N}^k$ (resp. $I \in \{0,1\}^k$) with $k \in \mathbb{N}$. A* countable sequence *(resp. a* countable binary sequence*) is any $I \in \mathbb{N}^{\mathbb{N}}$ (resp. $I \in \{0,1\}^{\mathbb{N}}$). The length (finite or $+\infty$) of any sequence is denoted by $l(I)$, and whenever $s \leq l(I)$ we denote by I_s the subsequence of I made by the first s terms. If $I = (i_1, i_2, \dots, i_k)$ is a finite sequence of length $k \in \mathbb{N}$, we write (I,n) to denote the sequence $(i_1, i_2, \dots, i_k, n)$, of length $k+1$. We give also a partial order to the sequences saying that $I \leq J$ if J extends I, that is if $l(I) \leq l(J)$ and $J_{l(I)} = I$. To give a formal consistence to our constructions, we will need to consider also the set $\{0,1\}^0$ of the finite binary sequences of length 0, whose unique element is $\varnothing$.*

Now, we can state and proof our result.

Proposition 2.2. *Let Z be a Polish space and let $\mu \in \mathcal{P}(Z)$ be a probability measure. Then there are measures μ_I and sets $Z_I \subseteq Z$ for the finite binary sequences I with the following properties:*

(i) *for any finite sequence I of length $k = l(I)$, $\|\mu_I\| = 2^{-k}$;*
(ii) *for any finite sequence I, $\mu_I = \mu_{(I,0)} + \mu_{(I,1)}$; hence, for any $k \in \mathbb{N}$ it is $\mu = \sum \{\mu_I : \ l(I) = k\}$;*
(iii) *for any finite sequence I, μ_I is concentrated in Z_I;*
(iv) *for any $I \leq J$, $Z_J \subseteq Z_I$;*
(v) *for any countable binary sequence I except countably many, $\operatorname{diam}(Z_{I_k}) \xrightarrow[k\to+\infty]{} 0$;*
(vi) *if, in addition, μ is non-atomic, then for any $k \in \mathbb{N}$ the sets Z_I with $I \in \{0,1\}^k$ are 2^k disjoint open sets, and $\mu_I = \mu \llcorner Z_I$.*

Proof. Fix a metric on Z and take a countable dense subset $\{z_n, \, n \in \mathbb{N}\}$ of Z. Let now j_1 be the least integer such that $B(z_{j_1}, 1/4)$ is not $\mu-$negligible and consider the sets $\{z \in Z : \ d(z, z_{j_1}) = r\}$ with $1/4 \leq$

$r \leq 1/2$: since they are more than countable many disjoint sets and μ is a finite measure, we can clearly choose $1/4 \leq r_1 \leq 1/2$ such that the boundary of $B\big(z_{j_1}, r_1\big)$ (that is contained in $\{z : d(z, z_{j_1}) = r_1\}$) is $\mu-$negligible; define then $B_1 := B\big(z_{j_1}, r_1\big)$. Let now $j_2 \geq 2$ be the least integer number –if there are any– such that $B\big(z_{j_2}, 1/4\big) \setminus \overline{B}_1$ has strictly positive $\mu-$measure: as before we can choose $1/4 \leq r_2 \leq 1/2$ such that, defining $B_2 := B\big(z_{j_2}, r_2\big) \setminus \overline{B}_1$, ∂B_2 is $\mu-$negligible. Let now j_3 be the least number with $B\big(z_{j_3}, 1/4\big) \setminus \big(\overline{B}_1 \cup \overline{B}_2\big)$ of strictly positive $\mu-$measure and choose $1/4 \leq r_3 \leq 1/2$ such that the boundary of $B_3 := B\big(z_{j_3}, r_3\big) \setminus \big(\overline{B}_1 \cup \overline{B}_2\big)$ is $\mu-$negligible. Iterating, we find a sequence $\{B_n\}$ of disjoint open sets of strictly positive measure, each of them with diameter less than 1 and $\mu-$negligible boundary. The union of the B_n is dense, since it contains $\{z_n\}$; moreover we claim that

$$\partial\Big(\bigcup\nolimits_{n\in\mathbb{N}} B_n\Big) = \bigcup\nolimits_{n\in\mathbb{N}} \partial B_n. \tag{2.1}$$

Indeed, take $z \in \partial\big(\cup B_n\big) = Z \setminus \cup_n B_n$, let n be the least integer such that $z \in B\big(z_n, 1/4\big)$, and let k be the least integer such that $j_k \geq n$: by construction, since $z \notin \cup B_n$ we infer that $z \in \partial B_1 \cup \cdots \cup \partial B_{k-1}$, then the non-trivial inclusion of (2.1) is shown; notice that the set in the right in (2.1) is $\mu-$negligible by construction, so we deduce that also the boundary of $\cup B_n$ is $\mu-$negligible: this holds thanks to the uniform lower bound on the radii r_n. As a consequence, $\cup B_n$ is of full measure.

The same construction, replacing Z with B_j, μ with $\mu \llcorner B_j$ and $1/4$ and $1/2$ with $1/8$ and $1/4$, allows to define disjoint open sets $B_{j,n} \subseteq B_j$ for $n \in \mathbb{N}$, each of them of strictly positive $\mu-$measure, $\mu-$negligible boundary and diameter less than $1/2$, and in such a way that the union $\cup_{n\in\mathbb{N}} B_{j,n}$ is of full measure in B_j. Iterating, we can define the open sets B_I for any finite sequence I, with the following properties (to be consistent, we also define $B_\varnothing := Z$):

- for each finite sequence I, B_I is an open set with strictly positive $\mu-$measure, $\mu-$negligible boundary and (unless $I = \varnothing$) diameter less than $1/2^{l(I)-1}$;
- if $I \leq J$, then $B_J \subseteq B_I$; on the other hand, if $I \not\leq J$ and $J \not\leq I$ then $\overline{B}_I \cap B_J = B_I \cap \overline{B}_J = \emptyset$;
- for each finite sequence I, the set $\bigcup\big\{B_J : I \leq J,\, l(J) = l(I) + 1\big\}$ is of full measure in B_I.

We see now how to define the sets Z_I and the measures μ_I for the binary sequences starting from the sets B_J already built. First of all, let j_1 be the integer such that

$$\mu\big(B_1 \cup B_2 \cup \cdots \cup B_{j_1-1}\big) \leq 1/2 < \mu\big(B_1 \cup B_2 \cup \cdots \cup B_{j_1}\big),$$

and set

$$Z_0^1 := B_1 \cup B_2 \cup \cdots \cup B_{j_1-1} \qquad\qquad Z_1^1 := B_{j_1+1} \cup B_{j_1+2} \cup B_{j_1+3} \cdots$$

Hence, Z_0^1 and Z_1^1 are disjoint open sets with $\mu(Z_0^1) \leq 1/2$ and $\mu(Z_1^1) \leq 1/2$. Let now j_2 be the integer such that

$$\mu\big(Z_0^1 \cup B_{j_1,1} \cup B_{j_1,2} \cup \cdots \cup B_{j_1,j_2-1}\big) \leq 1/2 < \mu\big(Z_0^1 \cup B_{j_1,1} \cup B_{j_1,2} \cup \cdots \cup B_{j_1,j_2}\big),$$

and set

$$Z_0^2 := Z_0^1 \cup B_{j_1,1} \cup B_{j_1,2} \cup \cdots \cup B_{j_1,j_2-1}, \qquad Z_1^2 := Z_1^1 \cup B_{j_1,j_2+1} \cup B_{j_1,j_2+2} \cup B_{j_1,j_2+3} \cup \cdots,$$

so that Z_0^2 and Z_1^2 are disjoint open sets with $\mu(Z_0^2)$, $\mu(Z_1^2) \leq 1/2$. Iterating, we find two increasing sequences of disjoint sets Z_0^n and Z_1^n with $\mu(Z_0^n)$, $\mu(Z_1^n) \leq 1/2$; define now $\widehat{Z}_0 = \cup_n Z_0^n$ and $\widehat{Z}_1 = \cup_n Z_1^n$. Notice that by the properties of the B_I and by construction the measure $\mu \llcorner Z \setminus \big(\widehat{Z}_0 \cup \widehat{Z}_1\big)$ is concentrated in the intersection of the sets $B_{j_1,j_2,\ldots,j_k}$, which has diameter 0 and then consists of at most a single point z. If μ is non-atomic –or in general if $\mu(\{z\}) = 0$– we let $Z_0 := \widehat{Z}_0$ and $Z_1 := \widehat{Z}_1$, so that Z_0 and Z_1

are two disjoint open sets of measure $1/2$, and we let $\mu_0 := \mu \llcorner Z_0$ and $\mu_1 := \mu \llcorner Z_1$. Otherwise, we let $Z_0 := \widehat{Z}_0 \cup \{z\}$ and $Z_1 := \widehat{Z}_1 \cup \{z\}$ and $\mu_0 := \mu \llcorner Z_0 + \big(1/2 - \mu(\widehat{Z}_0)\big)\delta_z$, $\mu_1 := \mu \llcorner Z_1 + \big(1/2 - \mu(\widehat{Z}_1)\big)\delta_z$. In both cases, $\mu = \mu_0 + \mu_1$.

The same procedure can be used recursively to define μ_I and Z_I for any finite binary sequence I: if μ_J and Z_J are given, one can repeat the construction above replacing $1/2$ by $1/2^{l(J)+1}$, μ by μ_J and all the open sets B_I by $B_I \cap Z_J$; in this way one finds two subsets of Z_J, that we call respectively $Z_{J,0}$ and $Z_{J,1}$ and that are disjoint and open if μ is non-atomic, and two measures of total mass $\|\mu_J\|/2$, denoted by $\mu_{J,0}$ and $\mu_{J,1}$ respectively, such that $\mu_J = \mu_{J,0} + \mu_{J,1}$.

In this way we defined the sets Z_I and the measures μ_I for all the finite binary sequences I; the properties (i), (ii), (iii), (iv) and (vi) follow directly by construction. Concerning the property (v), pick $i \neq j \in \mathbb{N}$: since by construction for any finite binary sequence J there is exactly a single n such that B_n intersects both $Z_{J,0}$ and $Z_{J,1}$, it is clear that there can be at most a unique countable binary sequence I such that, for any $k \in \mathbb{N}$, Z_{I_k} intersects both B_i and B_j. Since the pairs of integer numbers are countable many, it follows that for each countable binary sequence I except at most countable many, Z_{I_k} is definitively contained in a single B_n, therefore $\operatorname{diam}(Z_{I_k})$ is definitively less than 1. In the very same way, for each countable binary sequence I except at most countable many, Z_{I_k} is definitively contained in a single $B_{n,m}$, then it has definitively diameter less than $1/2$. In general, given any countable binary sequence I except at most countable many, for any $l \in \mathbb{N}$ there is a finite sequence $J \in \mathbb{N}^l$ of length l such that Z_{I_k} is definitively contained in B_J, thus it has diameter less than 2^{1-l}. Hence, also (v) holds and the thesis follows. □

3. Proof of Theorem A

In this section we will prove the existence Theorem A, which asserts that, provided f^+ is non-atomic (that is, $f^+(\{x\}) = 0$ for each $x \in X$), there always exists some transport map; in particular, we will be able to show the existence of a transport map having cost not greater that $f^+ \boxtimes f^-$ (recall that the existence is trivial for the transport plans, since $f^+ \boxtimes f^-$ is always a transport plan). This comparison with $f^+ \boxtimes f^-$ will be essential in next section to show the equality between inf(1.1) and min(1.2). Since we already noticed in the introduction that Dirac masses cannot be split by a transport map and then they can prevent the existence of a transport map, this result is sharp.

Proof. (of Theorem A): We divide the proof in three steps.

Step I. *Definition of the sequence $\{\gamma_k\}$.*
We assume without loss of generality that $\|f^+\| = \|f^-\| = 1$; moreover, we apply Proposition 2.2 to X and f^+, and to Y and f^-, finding the sets X_I and Y_I and the measures f_I^+ and f_I^-. To show the existence of a transport map, we will build a sequence $\{\gamma_k\}$ of transport plans between f^+ and f^- weakly* converging to a transport map (that is, weakly* converging to a plan induced by a transport map).

To define γ_k, we will take a suitable bijective function $\varphi_k : \{0,1\}^k \to \{0,1\}^k$, and set

$$\gamma_k := \sum\nolimits_{I \in \{0,1\}^k} f_I^+ \boxtimes f_{\varphi_k(I)}^- ; \tag{3.1}$$

notice that the above definition makes sense and defines a transport plan for any choice of φ_k thanks to (i) in Proposition 2.2. We choose the functions φ_k subject to the following two requirements:

$$\mathfrak{C}(\gamma_k) \leq \mathfrak{C}(\gamma_{k-1}) \leq \cdots \leq \mathfrak{C}(\gamma_1) \leq \mathfrak{C}(\gamma_0) = \mathfrak{C}(f^+ \otimes f^-), \tag{3.2}$$

$$\forall\, 0 \leq r \leq s \leq k,\ \forall I \in \{0,1\}^s \text{ one has } \varphi_r(I_r) = \big(\varphi_s(I)\big)_r. \tag{3.3}$$

The first condition is taken, of course, to ensure the bound on the cost of the transport map we will define; concerning the second one, we remark that it means that the different functions φ_k are "compatible", or in other words that any one refines the preceding one. Hence, the sequence $\{\gamma_k\}$ is defined by (3.1) once given the functions φ_k satisfying (3.2) and (3.3).

Step II. *Induction proof of the existence of the functions φ_k.*
We show now that it is possible to define the functions φ_k subject to the constraints (3.2) and (3.3). For $k = 0$, the unique function between $\{0,1\}^0$ and itself is $\varphi_0(\varnothing) = \varnothing$, and of course (3.2) and (3.3) hold emptily: indeed, keeping in mind our assumption that f^+ and f^- have unitary mass, it is $\gamma_0 = f^+ \boxtimes f^- = f^+ \otimes f^-$. For $k > 0$, we argue by induction: once $\varphi_0, \varphi_1, \dots, \varphi_k$ are given, we have to define φ_{k+1}. Since we want (3.3) to hold, to define φ_{k+1} we have 2^k decisions to take: for any $I \in \{0,1\}^k$, we can set

$$\text{either} \qquad \left\{ \begin{array}{l} \varphi_{k+1}(I,0) = \big(\varphi_k(I),0\big) \\ \varphi_{k+1}(I,1) = \big(\varphi_k(I),1\big) \end{array} \right. \qquad \text{or} \qquad \left\{ \begin{array}{l} \varphi_{k+1}(I,0) = \big(\varphi_k(I),1\big) \\ \varphi_{k+1}(I,1) = \big(\varphi_k(I),0\big). \end{array} \right.$$

These 2^k choices are clearly independent, and they correspond to all the possible φ_{k+1} satisfying (3.3): therefore, we can give a bijective correspondence between the functions satisfying (3.3) and the functions $\tau : \{0,1\}^k \to \{0,1\}$, associating to any such τ the function φ^τ given by

$$\varphi^\tau\big(I,i\big) = \left\{ \begin{array}{ll} \big(\varphi_k(I),i\big) & \text{if } \tau(I) = 0, \\ \big(\varphi_k(I),1-i\big) & \text{if } \tau(I) = 1. \end{array} \right. \tag{3.4}$$

Now, we have to care about (3.2). Notice that any function $\tau : \{0,1\}^k \to \{0,1\}$ can be associated to a function $\varphi^\tau : \{0,1\}^{k+1} \to \{0,1\}^{k+1}$ via (3.4), therefore to a transport plan γ^τ given –remind (3.1)– by

$$\gamma^\tau := \sum\nolimits_{I\in\{0,1\}^{k+1}} f_I^+ \boxtimes f_{\varphi^\tau(I)}^-.$$

By the construction, and recalling property (ii) in Proposition 2.2, it is immediately seen that the "mean" of the different γ^τ is γ_k, that is

$$\gamma_k = \frac{1}{2^{2^k}} \sum\nolimits_{\tau:\{0,1\}^k\to\{0,1\}} \gamma^\tau :$$

by the linearity of the cost $\mathfrak{C}$ in (1.2), it follows the existence of at least a $\tau : \{0,1\}^k \to \{0,1\}$ with $\mathfrak{C}(\gamma^\tau) \leq \mathfrak{C}(\gamma_k)$: setting $\varphi_{k+1} := \varphi^\tau$, then, both (3.3) and (3.2) are satisfied and our induction argument is completed. Having the functions φ_k and thanks to (3.3), we can also define $\varphi : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ the unique function extending all the φ_k.

Step III. *Proof that γ_k weakly* converges to a transport map t with $\mathfrak{C}(t) \leq \mathfrak{C}(f^+ \boxtimes f^-)$.*
Thanks to Prokhorov's Theorem (see the Appendix), the sequence $\{\gamma_k\}$ of transport plans is weakly* sequentially compact, thus there is a weak* limit γ of some subsequence (recall that all the measures γ_k have the same marginals f^+ and f^-, thus Prokhorov's Theorem can be trivially applied). Our aim is to show that γ is in fact induced by a transport map t, and our argument will also show as a byproduct the full weak* convergence of the sequence $\{\gamma_k\}$ to γ; notice that this will give the thesis, since $\mathfrak{C}(\gamma) \leq \mathfrak{C}(f^+ \boxtimes f^-)$ follows by the lower semicontinuity of $\mathfrak{C}$ and by (3.2). Recalling Lemma 1.1, it is sufficient to find a set $S \subseteq X \times Y$ of full measure w.r.t. γ and to check that for f^+-a.e. $x \in X$ the set $\{y \in Y : (x,y) \in S\}$ has diameter 0.

Define $X_k := \bigcup\{X_I : l(I) = k\}$, that is a set of full f^+-measure by (ii) and (iii) in Proposition 2.2; hence, also the set $\widetilde{X} := \bigcap\{X_k, k \in \mathbb{N}\}$ is of full measure. By (iv) and (vi) in Proposition 2.2 we can define a function $\psi : \widetilde{X} \to \{0,1\}^{\mathbb{N}}$ which associates to any $x \in \widetilde{X}$ the unique countable binary sequence $I = \psi(x)$ with the property that $x \in X_{I_k}$ for any $k \in \mathbb{N}$. According to (v) in Proposition 2.2, we define now A the subset of $\{0,1\}^{\mathbb{N}}$ of those countable binary sequences I such that $\operatorname{diam}(Y_{I_k})$ does not converge to 0; since A is countable, φ is a bijection and by (i) in Proposition 2.2, the set

$$\overline{X} := \left\{x \in \widetilde{X} :\ \varphi\big(\psi(x)\big) \notin A\right\}$$

has full f^+-measure. For $k \in \mathbb{N}$, we define now

$$S_k := \bigcup \left\{X_I \times \overline{Y}_{\varphi_k(I)} :\ I \in \{0,1\}^k\right\} :$$

notice that by definition (3.1) the plan γ_k is concentrated in S_k, and thanks to (3.3) also each γ_n with $n > k$ is concentrated in S_k. We want to infer that also γ is concentrated in S_k: to this aim, recall that by (vi) in Proposition 2.2 and by the non-atomicity of f^+ the sets X_I with $I \in \{0,1\}^k$ are open and disjoint and their union is of full f^+-measure: therefore, for any $I \in \{0,1\}^k$ one has $f^+(X_I) = f^+(\overline{X}_I)$. Consequently,

$$\begin{aligned}\gamma\big(X_I \times \overline{Y}_{\varphi_k(I)}\big) &= \gamma\big(\overline{X}_I \times \overline{Y}_{\varphi_k(I)}\big) \geq \limsup_{n\to\infty} \gamma_n\big(\overline{X}_I \times \overline{Y}_{\varphi_k(I)}\big) = \limsup_{n\to\infty} \gamma_n\big(\overline{X}_I \times Y\big) = f^+(X_I)\\ &= \gamma\big(X_I \times Y\big) \geq \gamma\big(X_I \times \overline{Y}_{\varphi_k(I)}\big),\end{aligned}$$

and this ensures that $\gamma \llcorner \big(X_I \times Y\big)$ is concentrated in $X_I \times \overline{Y}_{\varphi_k(I)}$: since the union of the different X_I for $I \in \{0,1\}^k$ has full f^+-measure –keep in mind that $\pi_1\gamma = f^+$– we infer that in fact γ is concentrated in S_k. Hence, γ is concentrated in $S := \cap_{k\in\mathbb{N}} S_k$: as pointed out in the beginning of this step, the thesis will be reached showing that for any $x \in \overline{X}$ –thus, for f^+-a.e. x– the set $\{y \in Y : (x,y) \in S\}$ has diameter 0. Indeed, if $(x,y) \in S$ and $x \in \overline{X}$ then $y \in Y_{\varphi\left(\psi(x)\right)}$ and, since $\varphi\big(\psi(x)\big) \notin A$, $Y_{\varphi(\psi(x))}$ has diameter 0. This concludes the proof; notice that the transport map can be defined almost everywhere by $t(x) := Y_{\varphi(\psi(x))}$ (our argument shows that the set in the right is exactly a point for a.e. x). □

4. Proof of Theorem B

In this section we will prove Theorem B, which shows the equality inf(1.1)=min(1.2) in the case when f^+ is non-atomic and the cost function c is continuous. First of all, we briefly discuss the hypotheses to convince the reader that they are sharp.

The meaning of the continuity assumption on c is easy to understand: indeed, there are situations when the splitting of masses (not allowed for the transport maps) is necessary due to the discontinuities of c, as the next example shows.

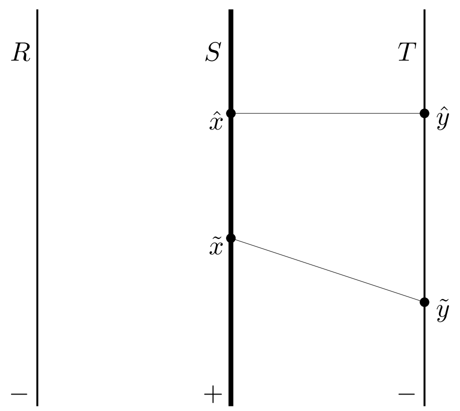

Figure 1. Example 4.1

Example 4.1. *Take, as in Figure 1, three parallel segments R, S and T of length 2 and with $\operatorname{dist}(R,S) = \operatorname{dist}(S,T) = 1$; then, let f^- be the one-dimensional Hausdorff measure on R and T, and let f^+ be the Hausdorff measure on S, multiplied by 2 to get $\|f^+\| = \|f^-\| = 4$. Finally, set the non-continuous cost*

$$c(x,y) := \begin{cases} 1 & \text{if } \operatorname{dist}(x,y) = 1; \\ 2 & \text{otherwise:} \end{cases}$$

for instance, in the figure one has $c(\hat{x},\hat{y}) = 1$ while $c(\tilde{x},\tilde{y}) = 2$. It is clear that there are many transport plans and also, according with Theorem A, transport maps. We notice that the minimum of the Kantorovich problem is $\min(1.2) = 4$, and it is achieved by the transport plan splitting the central segment in two parts and translating them on the left and on the right; on the other hand, we claim that the infimum of the Monge problem is $\inf(1.1) = 8 > \min(1.2)$ (and it is also a minimum). To do that, it clearly suffices to take a transport map t and to show that

$$f^+(E) = f^+\big(\{x \in S : \operatorname{dist}(x,t(x)) = 1\}\big) = 0 : \tag{4.1}$$

indeed, $c(x,t(x)) = 1$ if $x \in E$, and $c(x,t(x)) = 2$ otherwise. Notice also that E is the set of points $x \in S$ such that $t(x)$ is "at the same height" as x (according to the figure we consider the segments as vertical); therefore, we define E_l and E_r as the subsets of E made by the points which are moved on the left or on the right respectively, so that $E = E_l \cup E_r$ and $E_l \cap E_r = \emptyset$. Defining $F := t(E)$, by construction one has $\mathscr{H}^1(F) = \mathscr{H}^1(F \cap R) + \mathscr{H}^1(F \cap T) = \mathscr{H}^1(E_l) + \mathscr{H}^1(E_r) = \mathscr{H}^1(E)$; therefore, $f^+(E) = 2f^-(F)$. But since t is a transport map, by definition one has

$$f^-(F) = f^+\big(t^{-1}(F)\big) \geq f^+(E) = 2f^-(F),$$

that implies $f^+(E) = 0$: thus (4.1) is proved and our claim follows.

In the example above, it is clear how the non-continuity of c has the effect of making $\inf(1.1)$ strictly greater than $\min(1.2)$, and then why the hypothesis of continuity of c in Theorem B is sharp. It could seem less clear the reason to ask also the non-atomicity of f^+: indeed, by Theorem A we already know that without this assumption there could be no transport maps, so one could think that the non-atomicity of f^+ is needed only to get existence of transport maps. In particular, one could ask if the equality between the infimum and the minimum is still true if c is continuous and there exists some transport map; as the example below shows, this is not true, since the non-atomicity of f^+ plays a crucial role: roughly speaking, if there are Dirac masses, the fact that they cannot be split by a transport map is a too heavy obstacle.

Figure 2. Example 4.2

Example 4.2. *Let x^+ and x^- be two points in $X = Y = \mathbb{R}^3$, let B^+ and B^- be two balls of unitary volume, take the Euclidean cost function $c(x,y) := |y - x|$, and define $f^+ := \delta_{x^+} + \mathscr{L} \llcorner B^+$ and $f^- := \delta_{x^-} + \mathscr{L} \llcorner B^-$. If, as in the figure above, the point x^+ is close to the ball B^- and the point x^- is close to the ball B^+, but x^+ and B^- are very far from x^- and B^+, then the equality $\inf(1.1) = \min(1.2)$ does* *not hold true. Indeed, the transport plan that distributes the mass in x^+ on B^- and moves the mass in B^+ on x^- has a low cost; on the other hand, a transport map must necessarily move x^+ on x^- and B^+ on B^-, therefore it has a very high cost.*

In the example above, the equality between the Monge infimum and the Kantorovich minimum does not hold, even though c is continuous and the existence of transport maps is true. This explains how also the non-atomicity condition on f^+ in Theorem B is sharp. Now, we turn to the proof of Theorem B.

Proof. *(of Theorem B):* The proof of this result is a quite involved construction, so we will make our best to explain it step by step. First of all, take an optimal transport plan γ, which exists thanks to Lemma 1.2; we can assume that $\mathfrak{C}(\gamma) = \min(1.2) < +\infty$, since otherwise by (1.3) the claim is immediate. By the definition of Polish spaces we fix a metric on X and on Y; moreover we take $\varepsilon > 0$, and the thesis will be reached finding a transport map t such that $\mathfrak{C}(t) \leq \mathfrak{C}(\gamma) + \varepsilon\|\gamma\|$.

Step I. *Definition of A_I, B_I, M_I, m_I.*
Our first step consists in defining sets $A_I \subseteq X$ and $B_I \subseteq Y$ for the finite sequences I, which must be so small that $c(x,y)$ is almost constant for $(x,y) \in A_I \times B_I$, but not too small in order to fill γ−a.a. $X \times Y$. Moreover we want that A_I and A_J do not to intersect unless $I \leq J$ or $J \leq I$, while $A_J \subseteq A_I$ if $I \leq J$: this will be needed to glue the different "local transport maps" that we will find.

First of all, we take two sequences $\{x_i\}$ and $\{y_i\}$ dense in X and in Y respectively; then, take any $(i,j) \in \mathbb{N}^2$: if $c(x_i, y_j) = +\infty$, define $C_{i,j} := \emptyset \subseteq X$ and $D_{i,j} := \emptyset \subseteq Y$. On the other hand, if $c(x_i, y_j) < +\infty$, we set

$$C_{i,j} := B_X(x_i, r) \qquad\qquad D_{i,j} := B_Y(y_j, r),$$

where r is the maximum real number such that

$$\sup\Big\{c(x,y) - c(z,w) : (x,y),\, (z,w) \in B_X(x_i,r) \times B_Y(y_j,r)\Big\} \leq \varepsilon;$$

notice that, thanks to the continuity of c, the number r is well-defined and strictly positive. By the density of the sequences and again by the continuity of c, we have

$$\bigcup_{(i,j)\in\mathbb{N}^2} C_{i,j} \times D_{i,j} = X \times Y \setminus \Big\{(x,y) \in X \times Y : c(x,y) = +\infty\Big\};$$

therefore, recalling that $C(\gamma) < +\infty$, we have

$$\gamma\Big((X \times Y) \setminus \bigcup_{i,j} C_{i,j} \times D_{i,j}\Big) = 0. \tag{4.2}$$

We want now to define the sets A_I and B_I for the finite sequences I, in such a way that the following properties hold:

$$\forall I,\ \sup\Big\{c(x,y) - c(z,w) : (x,y),\, (z,w) \in A_I \times B_I\Big\} \leq \varepsilon; \tag{4.3}$$

$$I \leq J \Longrightarrow A_I \supseteq A_J; \tag{4.4}$$

$$A_I \cap A_J = \emptyset \text{ unless } I \leq J \text{ or } J \leq I; \tag{4.5}$$

$$B_I \cap B_J = \emptyset \text{ if } I < J \text{ or } J < I; \tag{4.6}$$

$$\gamma\Big(X \times Y \setminus \bigcup_I (A_I \times B_I)\Big) = 0. \tag{4.7}$$

Let us start with the sequences of length one: we set $A_1 := C_{i,j}$ and $B_1 := D_{i,j}$, where the pair (i,j) is chosen so that $\gamma(A_1 \times B_1)$ is "almost maximal": this means, $\gamma(A_1 \times B_1) \geq \gamma\big(C_{m,n} \times D_{m,n}\big)/2$ for any

$(m,n)\in\mathbb{N}^2$. Then, we proceed by induction: if A_m and B_m have been chosen for $1\le m<n$, we set

$$A_n := C_{i,j}\setminus \bigcup_{m=1}^{n-1} A_m, \qquad\qquad B_n := D_{i,j},$$

where the pair $(i,j)\in\mathbb{N}^2$ is again chosen in such a way that $\gamma(A_n\times B_n)$ is almost maximal. In this way, we have defined A_n and B_n for any integer $n\in\mathbb{N}$, and by construction we know that $A_i\cap A_j=\emptyset$ whenever $i\neq j$. We also claim that the sets A_i cover f^+−a.a. X, that is

$$f^+(X\setminus\cup_n A_n)=\gamma\big((X\setminus\cup_n A_n)\times Y\big)=0: \tag{4.8}$$

indeed, if not there would be, thanks to (4.2), a pair (i,j) such that $\gamma\big((C_{i,j}\times D_{i,j})\setminus(\cup_{n\in\mathbb{N}}A_n\times Y)\big)=\delta>0$; by construction, this would imply that $\gamma(A_n\times B_n)\ge\delta/2$ for any $n\in\mathbb{N}$, since otherwise A_n and B_n would have been chosen differently. Since the sets A_n are disjoint by construction, this would give $\gamma(\cup_n A_n\times B_n)=+\infty$, and the absurd yields (4.8).

Let us consider now the sequences of length 2: first we introduce the measure

$$\gamma_1 := \gamma \llcorner \Big(\big(X\times Y\big)\setminus\big(\cup_n A_n\times B_n\big)\Big);$$

then, for any $m\in\mathbb{N}$ we define

$$A_{m,1} := C_{i,j}\cap A_m \qquad\qquad B_{m,1} := D_{i,j}\setminus B_m$$

where, as before, the pair (i,j) is chosen almost maximizing $\gamma_1(A_{m,1}\times B_{m,1})$. Again, by induction we set also, choosing (i,j) to maximize $\gamma_1(A_{m,n}\times B_{m,n})$,

$$A_{m,n} := \big(C_{i,j}\cap A_m\big)\setminus\bigcup_{i=1}^{n-1} A_{m,i} \qquad\qquad B_{m,n} := D_{i,j}\setminus B_m,$$

so that $A_{m,n}\subseteq A_m$ and $B_{m,n}\cap B_m=\emptyset$. In this way we define all the sets $A_{m,n}$ and $B_{m,n}$, and the property (4.8) can be generalized as

$$\pi_1\gamma_1(X\setminus\cup_{m,n}A_{m,n})=0: \tag{4.9}$$

indeed, as before, otherwise there would be by (4.2) –recall that $\gamma_1\le\gamma$– a pair (i,j) such that $\gamma_1\big((C_{i,j}\setminus \cup_{m,n}A_{m,n})\times D_{i,j}\big)>0$. By (4.8), we infer the existence of some $\widehat{m}\in\mathbb{N}$ such that $\gamma_1\big((C_{i,j}\cap A_{\widehat{m}}\setminus \cup_{m,n}A_{m,n})\times D_{i,j}\big)=\delta>0$. Recall that, by definition of γ_1, $\gamma_1(A_{\widehat{m}}\times B_{\widehat{m}})=0$, therefore for any $n\in\mathbb{N}$ one has

$$\gamma_1\Big(\Big(\big(C_{i,j}\cap A_{\widehat{m}}\big)\setminus\bigcup_{i=1}^{n-1}A_{\widehat{m},i}\Big)\times\Big(D_{i,j}\setminus B_{\widehat{m}}\Big)\Big)\ge\delta:$$

this implies, by the definition of $A_{m,n}$ and $B_{m,n}$, that $\gamma_1\big(A_{\widehat{m},n}\times B_{\widehat{m},n}\big)\ge\delta/2$ for any $n\in\mathbb{N}$. As before, since the $A_{\widehat{m},n}$ are, varying n, disjoint sets, this would imply $\|\gamma_1\|=+\infty$ and the absurd shows (4.9).

Iterating the same procedure, we can define the measures

$$\gamma_p := \gamma\llcorner\Big(\big(X\times Y\big)\setminus\Big(\bigcup_{I:\,l(I)\le p}A_I\times B_I\Big)\Big) \tag{4.10}$$

and the sets

$$A_{(I,n)} := \big(C_{i,j}\cap A_I\big)\setminus\bigcup_{i=1}^{n-1}A_{(I,i)} \qquad\qquad B_{(I,n)} := D_{i,j}\setminus B_I$$

almost maximizing $\gamma_{l(I)}\big(A_{(I,n)}\times B_{(I,n)}\big)$, and it is shown exactly as before that

$$\pi_1\gamma_{p-1}(X\setminus\cup_{l(I)=p}A_I)=0. \tag{4.11}$$

We have now to check the validity of (4.3)–(4.7): (4.3), (4.4), (4.5) and (4.6) are obvious from the construction; concerning (4.7), if it were not true then there would exist (i,j) such that $\gamma(\Lambda)=\delta>0$, where

$$\Lambda := \big(C_{i,j}\times D_{i,j}\big)\setminus\Big(\bigcup_I A_I\times B_I\Big),$$

using again (4.2). For any $p \in \mathbb{N}$ and any I with $l(I) = p$, call now

$$\delta_I := \gamma\big(\Lambda \cap (A_I \times Y)\big) = \gamma_{p-1}\big(\Lambda \cap (A_I \times Y)\big);$$

thanks to (4.11), we have then $\sum_{l(I)=p} \delta_I = \delta$. Arguing as before, it is clear that $\gamma_p(A_{(I,1)} \times B_{(I,1)}) \geq \delta_I/2$ for any $I \in \mathbb{N}^p$, so that a fortiori

$$\gamma_p\Big(\bigcup_{J\in\mathbb{N}^{p+1}} A_J \times B_J\Big) \geq \delta/2.$$

And since this would give $\|\gamma\| = +\infty$, also (4.7) is shown.

We conclude now this first step by giving the following definitions:

$$M_I := f^+(A_I), \qquad\qquad m_I := \gamma(A_I \times B_I); \tag{4.12}$$

it is clear that for any finite sequence I one has $m_I \leq M_I$; more precisely, one has that

$$M_I \geq \sum\{m_J : J \geq I\} \text{ (in particular, } M_I = \sum\{m_J : J \geq I\} \text{ if } l(I) = 1), \tag{4.13}$$

$$\sum_I m_I = \|\gamma\|. \tag{4.14}$$

These two properties follow immediately by the following easy but remarkable consequence of (4.5) and (4.6):

$$A_I \times B_I \cap A_J \times B_J = \emptyset \qquad \forall I \neq J \text{ finite sequences.} \tag{4.15}$$

Step II. *Definition of the sets X_I.*
In this second step we want to define sets X_I, essentially disjoint w.r.t. f^+, with the property that $X_I \subseteq A_I$ and that $f^+(X_I) = m_I$.

To do that, we will define sets X_I^p for any finite sequence I and for any $p \geq l(I)$, and each set X_I will be found as the limit of the sequence $\{X_I^p\}$ as p goes to $+\infty$. More precisely, we will define arbitrarily the set $X_I^{l(I)} \subseteq A_I$, we will build sets $X_I^{s,+}$ and $X_I^{s,-}$ for $s \geq l(I)$, and we will take as inductive definition for X_I^p with any $p > l(I)$ the following:

$$X_I^{p+1} := X_I^p \cup X_I^{p,+} \setminus X_I^{p,-}; \tag{4.16}$$

in words, the sequence of sets X_I^p is obtained adding the sets $X_I^{s,+}$ and subtracting the sets $X_I^{s,-}$. The sets X_I^p and the sets $X_I^{s,\pm}$ will satisfy the following number of properties (the choice of the index $p+1$ instead of p in some of the equations below may seem a bit odd at first glance, but this will be needed to show them in the same induction step).

For any $p \geq 0$, the sets X_I^{p+1} with $l(I) \leq p+1$ are disjoint w.r.t. f^+; (4.17)

for any $1 \leq l(I) \leq p+1$, $X_I^{p+1} \subseteq A_I$; (4.18)

for any $1 \leq l(I) \leq p+1$, $f^+(X_I^{p+1}) = m_I$; (4.19)

for any $1 \leq l(I) \leq p$, $X_I^{p,+} \cap X_I^p = \emptyset$, while $X_I^{p,-} \subseteq X_I^p$; (4.20)

for any $1 \leq l(I) \leq p$, $f^+(X_I^{p,+}) = f^+(X_I^{p,-})$; (4.21)

for any $1 \leq l(I) \leq p$, $f^+(X_I^{p,-}) \leq \sum_{l(J)=p+1} m_J$. (4.22)

We leave to the next step the proof that it is possible to choice sets X_I^p, $X_I^{s,\pm}$ in such a way; now, we show that as a consequence we can obtain the sets X_I as claimed in the beginning of this step.

Consider any fixed finite sequence I, and the sequence $\{X_I^p\}$ with $p \geq l(I)$; it is a sequence of sets, each of them with measure m_I according with (4.19): notice that, in fact, the property (4.19) is a straightforward consequence of the definition (4.16) together with (4.20) and (4.21). It is clear from the construction that the sequence X_I^p converges to some set (in the sense of the measure f^+) if the sequence $\sigma_m := \sum_{p=m}^{+\infty} f^+(X_I^{p,-})$ converges to 0 when m goes to ∞. Write now $\lambda_p := \sum_{l(I)=p} m_I$: from the property (4.15) one obtains, directly from the definitions (4.10) and (4.12), that $\lambda_p = \|\gamma_p - \gamma_{p-1}\|$. Therefore, by (4.22) we have

$$\sigma_m = \sum\nolimits_{p=m}^{+\infty} \lambda_{p+1} = \sum\nolimits_{p=m}^{+\infty} \|\gamma_{p+1} - \gamma_p\| = \|\gamma_m\| :$$

indeed, $\{\gamma_p\}$ is a decreasing sequence of measures by (4.10), and it strongly converges to 0 thanks to (4.7). It follows that σ_m converges to 0, and then that the sequence $\{X_I^p\}$ converge in the sense of the measure f^+ to a set, that we denote by X_I. The fact that the different sets X_I are disjoint and that each X_I is contained in the corresponding set A_I and has mass m_I is clear from (4.17), (4.18) and (4.19), so this step is concluded.

Step III. *Proof of (4.17)–(4.22).*
This step is devoted to build the sets X_I^p and $X_I^{s,\pm}$ fulfilling the requirements (4.17)–(4.22); we will do it via an inductive construction on p.

The starting case $p = 0$ is trivial, since we have only to take care of (4.17), (4.18) and (4.19). To do that, it suffices to choose arbitrarily for any $i \in \mathbb{N}$ a set $X_i^1 \subseteq A_i$ with $f^+(X_i) = m_i$: this is possible since $f^+(A_i) = M_i \geq m_i$ by (4.13), and in this way the three properties are clearly satisfied.

Before to face the induction in the general case, we briefly show the particular case $p = 1$, to help the reader to understand the construction. First of all, we choose arbitrarily, for i, $j \in \mathbb{N}$, sets $X_{i,j}^2 \subseteq A_{i,j}$ with mass $f^+(X_{i,j}^2) = m_{i,j}$: this is possible, as before, because $f^+(A_{i,j}) = M_{i,j} \geq m_{i,j}$ by (4.13). Now, we have to define the sets X_i^2; in order to make the sets disjoint, we subtract to X_i^1 the part now occupied by the sets $X_{i,j}^2$: more precisely, we set

$$X_i^{1,-} := X_i^1 \cap \big(\cup_{j \in \mathbb{N}} X_{i,j}^2 \big).$$

Notice that we do not need to subtract from X_i^1 the sets $X_{m,j}^2$ with $m \neq i$, in view of (4.5). To complete the case $p = 1$, we need to define suitable sets $X_i^{1,+}$ –and apply the definition (4.16). Notice that if we can define a set $X_i^{1,+} \subseteq A_i$ with $f^+(X_i^{1,+}) = f^+(X_i^{1,-})$ not intersecting X_i^1 nor any of the $X_{i,j}^2$, then all the properties (4.17)–(4.22) are trivially satisfied. And this is possible if "there is enough space", that is, if $f^+(A_i) \geq \sum_j f^+(X_{i,j}^2) + m_i$: but this means $M_i \geq m_i + \sum_j m_{i,j}$, that is true thanks to (4.13).

Now, we consider the induction in the general case: take $n \geq 1$ and assume by hypothesis that the sets $X_I^{l(I)}$ for $1 \leq l(I) \leq n+1$ and the sets $X_I^{p,\pm}$ for $1 \leq l(I) \leq p \leq n$ have been chosen in such a way that the properties (4.17)–(4.22) hold for any $0 \leq p \leq n$. Then, we need to define X_I^{n+2} for the sequences I of length $l(I) = n+2$ and $X_I^{n+1,\pm}$ for the sequences I of length $l(I) \leq n+1$ fulfilling the properties (4.17)–(4.22) with $p = n+1$. First of all, for any sequence I of length $n+2$ choose arbitrarily $X_I^{n+2} \subseteq A_I$ with mass $f^+(X_I^{n+2}) = m_I$. Then, to define the sets $X_I^{n+1,\pm}$ for the sequences I of length at most $n+1$, we argue recursively backward: we start from the sequences of length $s = n+1$ and we go back to those of length $s = 1$.

Let us do that taking care only that (4.18), (4.21) and the property

$$X_I^{n+1,+} \cap \Big(X_I^{n+1} \cup \bigcup \big\{ X_J^{n+2} :\ l(I) < l(J) \leq n+2 \big\} \Big) = \emptyset \tag{4.23}$$

hold, and taking as definition for $X_I^{n+1,-}$ the following

$$X_I^{n+1,-} := X_I^{n+1} \cap \bigcup \big\{ X_J^{n+2} :\ l(I) < l(J) \leq n+2 \big\}. \tag{4.24}$$

Notice that, since we start from the sequences of length $s = n+1$ and we go back till $s = 1$, the definition above makes sense: indeed, when one defines $X_I^{n+1,\pm}$ he has already set X_J^{n+2} for $l(J) > l(I)$.

Assume then that we have already defined $X_J^{n+1,\pm}$ for $1 \leq s < l(J) \leq n+1$ so that (4.18), (4.21) and (4.23) hold with $p = n+1$; so, take any sequence I of length $l(I) = s$, use (4.24) as definition for $X_I^{n+1,-}$, and consider the problem to define $X_I^{n+1,+}$ fulfilling (4.18), (4.21) and (4.23). This would mean to select a subset of A_I of f^+-measure equal to $f^+(X_I^{n+1,-})$ and non intersecting X_I^{n+1} nor any of the X_J^{n+2} with $l(I) < l(J) \leq n+2$; this is clearly possible if (and only if) "there is enough space", i.e., if

$$f^+\left(A_I \setminus \left(X_I^{n+1} \cup \bigcup \left\{X_J^{n+2} :\ l(I) < l(J) \leq n+2\right\}\right)\right) \geq f^+(X_I^{n+1,-}). \tag{4.25}$$

Since by hypothesis we know the validity of (4.18) for $l(J) > l(I)$ and by (4.5), $A_I \cap X_J^{n+2} = \emptyset$ for each J with $l(J) > l(I)$ and $J \not\geq I$; moreover we already have that $f^+(X_J^{n+2}) \leq m_J$: indeed, this is true by construction if $l(J) = n+2$, while otherwise it is true by (4.16) and (4.21), recalling that $f^+(X_J^{n+1}) = m_J$ because (4.19) is assumed by induction in the case $p = n$. Therefore, keeping in mind the definition (4.24), we have that

$$\begin{aligned} f^+\left(A_I \setminus \left(X_I^{n+1} \cup \bigcup\{X_J^{n+2} :\ l(I) < l(J) \leq n+2\}\right)\right) \geq \\ \geq M_I - \left(\sum \left\{m_J :\ J \geq I,\ l(J) \leq n+2\right\} - f^+(X_I^{n+1,-})\right). \end{aligned}$$

Hence, the validity of the sufficient condition (4.25) is directly given by (4.13), and then our recursive construction works (notice also that the different finite sequences of the same length do not interact with each other, since the different sets A_I with fixed length $l(I)$ are disjoint). Therefore, we have defined all the sets $X_I^{n+1,\pm}$ with $1 \leq l(I) \leq n+1$ and the properties (4.18), (4.21) and (4.23) hold.

To finish the induction and then conclude this step, we have therefore only to check the validity of (4.17), (4.19), (4.20) and (4.22). First, we notice that the property (4.19) is clear by (4.16), (4.21) and (4.23), while the property (4.20) is given by (4.23) and (4.24).

Concerning (4.17), if $I \neq J$ and $l(I) = l(J)$ then the disjointness of X_I^{n+2} and X_J^{n+2} is immediate by (4.18) and (4.5); on the other hand we can assume $l(I) < l(J)$ and then the disjointness follows immediately by the definition (4.24) and the property (4.23).

Finally, concerning (4.22), rewrite (4.24) as

$$X_I^{n+1,-} = X_I^{n+1} \cap \left(\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+2}\right); \tag{4.26}$$

then recall by (4.16) that, whenever $l(J) < n+2$, one has $X_J^{n+2} = X_J^{n+1} \cup X_J^{n+1,+} \setminus X_J^{n+1,-}$; on the other hand, each $X_J^{n+1,-}$ is contained in

$$\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(J)<l(K)<n+2} X_K^{n+1,+},$$

and then (4.26) can be rewritten as

$$X_I^{n+1,-} = X_I^{n+1} \cap \left(\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1,+}\right).$$

Since (4.17) is already known for $p = n$, it is $X_I^{n+1} \cap X_J^{n+1} = \emptyset$ for $I \neq J$, then we can also write

$$X_I^{n+1,-} = X_I^{n+1} \cap \left(\bigcup\nolimits_{l(J)=n+2} X_J^{n+2} \cup \bigcup\nolimits_{l(I)<l(J)<n+2} X_J^{n+1,+}\right).$$

Moreover, the intersection between X_I^{n+1} and each $X_J^{n+1,-}$ with $l(I)<l(J)<n+2$ is empty since X_I^{n+1} does not intersect X_J^{n+1} which contains $X_J^{n+1,-}$. Then we can further change (4.26) writing

$$\begin{aligned} X_I^{n+1,-} &= X_I^{n+1} \cap \left(\bigcup_{l(J)=n+2} X_J^{n+2} \cup \bigcup_{l(I)<l(J)<n+2} X_J^{n+1,+} \Big\backslash \bigcup_{l(I)<l(J)<n+2} X_J^{n+1,-} \right) = \\ &= X_I^{n+1} \cap (A \cup B \setminus C) = X_I^{n+1} \cap D \end{aligned}$$

We conclude noticing that A and B are disjoint sets by the property (4.23), while C is contained in $A \cup B$ by (4.24), thus $f^+(D) = f^+(A) + f^+(B) - f^+(C)$. Finally, $f^+(B) = f^+(C)$ thanks to (4.21) and since both B and C are union of disjoint sets by construction; keeping in mind (4.19), we have

$$f^+\left(X_I^{n+1,-}\right) \leq f^+(D) = f^+(A) = \sum_{l(J)=n+2} m_J,$$

thus the validity of (4.22) holds and this step is concluded.

Step IV. *Conclusion.*

In this last step we show how the sets X_I defined in the second step can be used to get the thesis. First of all we write $\gamma_I := \gamma \llcorner (A_I \times B_I)$, so that by (4.15) and (4.7) it is $\gamma = \sum_I \gamma_I$; then we define the measure

$$\tilde{\gamma}_I := f^+ \llcorner X_I \boxtimes \pi_2 \gamma_I.$$

This definition makes sense since $\|\pi_2\gamma_I\|_Y = \|\gamma_I\|_{X\times Y} = m_I = f^+(X_I)$, therefore $\tilde{\gamma}_I$ is a transport plan between $f^+ \llcorner X_I$ and $\pi_2\gamma_I$; on the other hand, γ_I is a transport plan between $\pi_1\gamma_I$ and $\pi_2\gamma_I$. Keeping in mind that $\|\tilde{\gamma}_I\| = \|\gamma_I\| = m_I$, that both the measures are concentrated in $A_I \times B_I$, and (4.3), one immediately gets

$$\left|\mathfrak{C}(\tilde{\gamma}_I) - \mathfrak{C}(\gamma_I)\right| \leq m_I \varepsilon. \tag{4.27}$$

We apply now Theorem A to the marginals of $\tilde{\gamma}_I$, obtaining a transport map t_I from $f^+ \llcorner X_I$ to $\pi_2\gamma_I$ having cost $\mathfrak{C}(t_I) \leq \mathfrak{C}(\tilde{\gamma}_I)$. Since the different sets X_I are essentially disjoint w.r.t. f^+, we can glue the maps t_I to get a map t; recalling now (4.14) to make sure that $\cup X_I$ is of full f^+-measure, it follows that t is a transport map from f^+ to $\sum_I \pi_2\gamma_I = \pi_2 \sum_I \gamma_I = \pi_2\gamma = f^-$, thus a transport map for the original problem. By (4.27), the linearity of the cost and (4.14) again, we derive

$$\mathfrak{C}(t) = \sum_I \mathfrak{C}(t_I) \leq \sum_I \mathfrak{C}(\tilde{\gamma}_I) \leq \sum_I \mathfrak{C}(\gamma_I) + m_I\varepsilon = \mathfrak{C}(\gamma) + \varepsilon\|\gamma\| :$$

such a transport map is what we needed to find, so the proof is complete. □

5. A stronger version of the Isomorphism Theorem

In this section we will see how the classical statement of the Isomorphism Theorem can be strengthened stating also the bicontinuity property of the isomorphism φ. As far as we know, the statement below is not known, even though it could have also be obtained starting from the ingredients of the classical proofs of the Isomorphism Theorem. In the following proof we will not make use of the Isomorphism Theorem, so that ours is also a new proof of the classical result.

Theorem C. *Let, as in the Isomorphism Theorem 1.5, X be a Polish space and μ a non-atomic probability measure on X. Then, there are two sets $\widetilde{X} \subseteq X$ and $\widetilde{[0,1]} \subseteq [0,1]$ and a bijective function $\varphi : \widetilde{X} \to \widetilde{[0,1]}$ as in Theorem 1.5, with the additional property that φ is bicontinuous (so, φ is an homeomorphism).*

Proof. Apply Proposition 2.2 to (X,μ) finding the sets X_I and the measures μ_I, for the finite binary sequences I, fulfilling (i)–(vi). Consider now $x \in X$: either there exists $n \in \mathbb{N}$ such that $x \notin \bigcup_{l(I)=n} X_I$, or –by (iv) and (vi)– there is a countable binary sequence $I^x \in \{0,1\}^{\mathbb{N}}$ such that $x \in \bigcap_{n\in\mathbb{N}} X_{I^x_n}$. Thanks to (ii) and (iii), the set X_1 of the points for which the second possibility happens is of full $\mu-$measure.

Take now any fixed countable binary sequence I: the set $\{x\in X_1 : I^x = I\}$ is $\mu-$negligible thanks to (i), therefore by (v) the set $\widetilde{X} := \{x\in X_1 : x = \bigcap_{n\in\mathbb{N}} X_{I^x_n}\}$ is still of full $\mu-$measure. Notice now that, since the sets X_I are open, the set X_1 is a countable intersection of open sets, thus Borel; on the other hand, for any countable I one clearly has that $\{x\in X_1 : I^x = I\} = \bigcap_{n\in N} X_{I_n}$ is a countable intersection of open sets, thus Borel. Finally, $\widetilde{X}$ is a Borel set minus a countable union of Borel sets, so $\widetilde{X}$ is a Borel subset of X.

Consider now the map $\varphi : \widetilde{X} \to [0,1]$ defined by

$$\varphi(x) := \sum_{n\in\mathbb{N}} \frac{i^x_n}{2^n},$$

where i^x_n is the $n-$th element of the sequence I^x. It is clear that φ is injective, because by definition $\varphi(x) = \varphi(z)$ implies $I^x = I^z$ and so (inside $\widetilde{X}$) $x = z$. We claim now that $\varphi_\#\mu = \mathscr{L}$: indeed, take any finite sequence I and define the open interval

$$T_I := \left(\frac{\sum_{n=1}^{l(I)} i_n 2^{l(I)-n}}{2^{l(I)}}, \frac{1+\sum_{n=1}^{l(I)} i_n 2^{l(I)-n}}{2^{l(I)}} \right) \subseteq [0,1];$$

one has $\mathscr{L}(T_I) = 2^{-l(I)}$, and on the other hand $\varphi^{-1}(T_I) = X_I \cap \widetilde{X}$ and so by (i)

$$\mu\big(\varphi^{-1}(T_I)\big) = \mu\big(X_I\cap\widetilde{X}\big) = \mu(X_I) = 2^{-l(I)} = \mathscr{L}(T_I);$$

since the sets T_I are a basis for the topology of $[0,1]$, the claimed equality $\varphi_\#\mu = \mathscr{L}$ holds. Define now $\widetilde{[0,1]} := \varphi\big(\widetilde{X}\big)$: arguing exactly as before, this is seen to be a Borel subset of $[0,1]$ because it is obtained from the open sets T_I by countable unions and intersections, and of course φ is a bijective function between $\widetilde{X}$ and $\widetilde{[0,1]}$. As a consequence, of course $\varphi^{-1}_\#\mathscr{L} = \mu$, so that we have shown the Isomorphism Theorem 1.5: we have then only to check the bicontinuity of φ to show the claim of Theorem C (of course, the bicontinuity will imply also that Borel sets are mapped in Borel sets).

Let then $x\in\widetilde{X}$ and $\varepsilon>0$: taken an integer n such that $2^{-n}<\varepsilon$, for any $z\in\widetilde{X}\cap X_{I^x_n}$ for definition one has that both $\varphi(x)$ and $\varphi(z)$ belong to $T_{I^x_n}$, so $|\varphi(x)-\varphi(z)| < 2^{-n} < \varepsilon$; since $X_{I^x_n}$ is open in X, $X_{I^x_n}\cap\widetilde{X}$ is an open neighborhood of x, and then the continuity of φ is shown.

Finally, to show the continuity of φ^{-1}, take $\varphi(x)\in\widetilde{[0,1]}$ and take $\varepsilon>0$; by construction, $\operatorname{diam}(X_{I^x_n}) \xrightarrow[n\to+\infty]{} 0$, so there is $\bar n$ such that $\operatorname{diam}(X_{I^x_{\bar n}}) < \varepsilon$. Then, $T_{I^x_{\bar n}}\cap\widetilde{[0,1]}$ is an open neighborhood of $\varphi(x)$ in $\widetilde{X}$, whose image under the map φ^{-1} is contained in the ball centered in x at with radius ε: the continuity of φ^{-1} and then the thesis follow. $\square$

An immediate corollary is the following.

Corollary 5.1. *Given the Polish spaces X and Y and the non-atomic probability measures $\mu\in\mathcal{P}(X)$ and $\nu\in\mathcal{P}(Y)$, there are a Borel subset $\widetilde{X}$ (resp. $\widetilde{Y}$) of X (resp. Y) of full $\mu-$measure (resp. $\nu-$measure) and an homeomorphism $\varphi : \widetilde{X}\to\widetilde{Y}$ such that $\varphi_\#\mu = \nu$ and $\varphi^{-1}_\#\nu = \mu$.*

Proof. It follows immediately applying Theorem C both to X and Y and recalling that the intersection of two Borel sets of full measure is still a Borel set of full measure. $\square$

Appendix A. Some basic tools on Polish spaces

Here we briefly recall some basic definitions and properties about Polish spaces (for a more exhaustive but elementary discussion on this subject, see the Appendix in [11]); first of all, we need the definition:

Definition A.1. *A topological space X is said a Polish space if it is metrizable with a distance making it separable and complete.*

Recall that a space is said metrizable if there is some distance on it inducing the same topology; note also that the completeness, but not the separability, depends on the distance. In this paper, we always consider each Polish space endowed with the Borel $\sigma-$algebra, denoted by $\mathscr{B}(X)$.

Given a Polish space X, we consider the space $\mathcal{M}^+(X)$ of the finite positive measures on it, that is defined, as usual, as the set of the countably additive functions from $\mathscr{B}(X)$ to $\mathbb{R}^+$; it is well known that $\mathcal{M}^+(X)$ is a (strongly) closed subset of $\big(C_b(X)\big)'$, the dual space of the space $C_b(X)$ of the continuous and bounded real functions on X, endowed with the norm

$$\|u\|_{C_b(X)} := \sup\big\{|u(x)|,\ x \in X\big\};$$

we denote by $\langle\cdot,\cdot\rangle$ the duality between $C_b(X)'$ and $C_b(X)$, and also –in particular– the duality between $\mathcal{M}^+(X)$ and $C_b(X)$. Even though $\mathcal{M}^+(X)$ is not a linear space because we deal only with positive measures, we consider on it the norm induced by the inclusion in $C_b(X)'$, that is

$$\|\mu\|_{\mathcal{M}^+(X)} := \sup\big\{\langle\mu,u\rangle,\ \|u\|_{C_b(X)} = 1\big\} = \mu(X).$$

We can also consider on $\mathcal{M}^+(X)$ the weak* convergence induced by $C_b(X)$: we say that a sequence $\mu_n \in \mathcal{M}^+(X)$ weak* converge to μ if

$$\langle\mu_n,u\rangle \longrightarrow \langle\mu,u\rangle \qquad \text{for each } u \in C_b(X).$$

A fundamental result on Polish spaces, which is used in this paper, and which can be found for example in the book of Dellacherie and Meyer [4], is the following:

Theorem A.2 (Prokhorov)**.** *A bounded set $\{\mu_i,\ i\in I\}$ in $\mathcal{M}^+(X)$ is weak* sequentially relatively compact if and only if for any ε there is a compact set K_ε such that*

$$\mu_i(X\setminus K_\varepsilon) \le \varepsilon \qquad \textit{for any } i \in I.$$

A.1. **Tensor product of measures.** Given two measures $\mu \in \mathcal{M}^+(X)$ and $\nu \in \mathcal{M}^+(Y)$, we define the *tensor product* of μ and ν, $\mu\otimes\nu \in \mathcal{M}^+(X\times Y)$, as the unique measure on $X \times Y$ satisfying

$$\mu\otimes\nu(A\times B) := \mu(A)\cdot\nu(B) \qquad \forall\, A \in \mathscr{B}(X),\ B \in \mathscr{B}(Y):$$

in particular, the marginals of $\mu\otimes\nu$ are

$$\pi_1(\mu\otimes\nu) = \nu(Y)\mu, \qquad\qquad \pi_2(\nu\otimes\mu) = \mu(X)\nu$$

writing, with a slight abuse, π_1 and π_2 in place of $\pi_{1\#}$ and $\pi_{2\#}$.

An interesting particular case is when μ and ν are probability measures; so, the projections of $\mu\otimes\nu$ are precisely μ and ν: this is particularly useful in the study of the mass transportation, since if $\|f^+\| = \|f^-\| = 1$ then $f^+\otimes f^-$ is always a transport plan between f^+ and f^-, and it is the trivial transport plan which moves each part of the mass f^+ of X in Y proportionally to f^-. However, it is often useful to have a compact symbol to denote this trivial transport plan also when the measures f^+ and f^- are not probability measures. Therefore, whenever the measures μ and ν are so that $\|\mu\| = \|\nu\|$, we define the *rescaled tensor product* of μ and ν as the unique measure $\mu\boxtimes\nu \in \mathcal{M}^+(X\times Y)$ such that

$$\mu\boxtimes\nu(A\times B) = \frac{\mu(A)\cdot\nu(B)}{\|\mu\|} = \frac{\mu(A)\cdot\nu(B)}{\|\nu\|} \qquad \forall\, A \in \mathscr{B}(X),\ \forall\, B \in \mathscr{B}(Y), \tag{A.1}$$

so that

$$\mu \boxtimes \nu = \frac{\mu \otimes \nu}{\|\mu\|} = \frac{\mu \otimes \nu}{\|\nu\|}:$$

the projections of $\mu \boxtimes \nu$ are precisely μ and ν and therefore whenever $\|f^+\| = \|f^-\|$ the measure $f^+ \boxtimes f^-$ is a transport plan between f^+ and f^-.

Acknowledgments

References

[1] L. Ambrosio, Lecture Notes on Optimal Transport Problems, in "Mathematical Aspects of Evolving Interfaces", Lecture Notes in Mathematics, LNM **1812**, Springer (2003), 1–52.

[2] L. Ambrosio & A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation, in "Optimal Transportation and Applications", Lecture Notes in Mathematics, LNM **1813**, Springer (2003), 123–160.

[3] 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.

[4] C. Dellacherie & P.A. Meyer, Probabilities and potential, North-Holland Publ. Co. (1978).

[5] L.C. Evans & W. Gangbo, Differential equations methods for the Monge–Kantorovich mass transfer problem, Memoirs of the A.M.S., Vol. **137**, Number 653, (1999).

[6] W. Gangbo, The Monge mass transfer problem and its applications, Contemp. Math., **226** (1999), 79–104.

[7] L.V. Kantorovich, On the transfer of masses, Dokl. Akad. Nauk. SSSR, **37** (1942), 227–229.

[8] L.V. Kantorovich, On a problem of Monge, Uspekhi Mat. Nauk., **3** (1948), 225–226.

[9] G. Letta, ?

[10] G. Monge, Memoire sur la Theorie des Déblais et des Remblais, Hist. de l'Acad. des Sciences de Paris (1781).

[11] A. Pratelli, Existence of optimal transport maps and regularity of the transport density in mass transportation problems, Ph.D. Thesis, Scuola Normale Superiore, Pisa, Italy (2003). Available on http://cvgmt.sns.it/ .

[12] S.T. Rachev & L. Rüschendorf, Mass Transportation Problems, Springer–Verlag (1998).

[13] H.L. Royden, Real Analysis (2nd edition), Macmillan, (1968).

[14] N.S. Trudinger & X.J. Wang, On the Monge mass transfer problem, Calc. Var. P.D.E., **13** (2001), 19–31.

Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy

E-mail address: `a.pratelli@sns.it`