

# On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation

Aldo Pratelli

*Dipartimento di Matematica "F. Casorati", via Ferrata, 1, 27100 Pavia, Italy*

Received 18 January 2005; received in revised form 20 September 2005; accepted 9 December 2005

Available online 7 July 2006

---

## Abstract

This paper concerns the Monge's transport problem in a general Polish space. We find optimal conditions to establish the equality between the infimum of Monge's problem and the minimum of the Kantorovich's relaxed version of the problem. A preliminary version of the results of this paper is contained in the Ph.D. thesis [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/>].

© 2006 Elsevier Masson SAS. All rights reserved.

## Résumé

Ce papier considère le problème du transport de Monge dans un espace Polonais général. Nous trouvons des conditions optimales pour établir l'égalité entre l'infimum du problème de Monge et le minimum de la version relâchée du problème de Kantorovich. Une version préliminaire des résultats de ce papier est contenue dans la thèse de Doctorat [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/>].

© 2006 Elsevier Masson SAS. All rights reserved.

MSC: 49N15; 28A50

Keywords: Monge problem; Optimal transport

---

## 1. Introduction

### 1.1. Setting of the problem and statement of the main results

In 1781, G. Monge [8] 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  $f^0 \in \mathcal{M}^+(X)$  and  $f^1 \in \mathcal{M}^+(Y)$  of the same total mass  $\|f^0\| = \|f^1\|$ . A *transport map* from  $f^0$  to  $f^1$  is any function  $t : X \rightarrow Y$  such that  $t_{\#}f^0 = f^1$  (of course,  $t$  need to be measurable with respect to the  $\sigma$ -algebra of the  $f^0$ -measurable sets on  $X$  and to the Borel  $\sigma$ -algebra  $\mathcal{B}(Y)$  on  $Y$ ); we are also given a l.s.c. function  $c : X \times Y \rightarrow \bar{\mathbb{R}}^+ := \mathbb{R}^+ \cup \{+\infty\}$ , and to each transport map one associates the cost

---

*E-mail address:* [aldo.pratelli@unipv.it](mailto:aldo.pratelli@unipv.it) (A. Pratelli).

$$\mathfrak{C}(t) := \int_X c(x, t(x)) \, d f^0(x). \quad (1.1)$$

The interpretation of this problem as a transportation of mass is very easy: if  $f^0$  and  $f^1$  measure the height of some soil and the depth of some hole of the same volume, a transport map  $t : X \rightarrow Y$  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  $f^0$  is absolutely continuous w.r.t. the Lebesgue measure (see [1], or also [4,3,13, 2,11] or [10] for general references). Even worse, it may happen that there are no transport maps at all: for instance, if  $f^0$  is a Dirac mass in the point  $x \in X$ , then  $t_{\#} f^0$  is the Dirac mass in the point  $t(x) \in Y$ , therefore there is no transport map if  $f^1$  is not a Dirac mass.

In 1940's, Kantorovich proposed [6,7] 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  $f^0$  and  $f^1$  any probability measure  $\gamma \in \mathcal{M}^+(X \times Y)$  whose marginals  $\pi_X \gamma$  and  $\pi_Y \gamma$  are  $f^0$  and  $f^1$  respectively. Any such measure can be interpreted as a way to transport mass, where  $\gamma(A \times B)$  represents the amount of mass originally 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

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

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  $\gamma_t := (1, t)_{\#} f^0$ , and also the cost is the same. More in general, a transport plan  $\gamma$  “is” a transport map – i.e. there is a transport plan  $t$  such that  $\gamma = \gamma_t$  – if and only if  $\gamma$  is concentrated in a graph: formally, the following result holds.

**Lemma 1.1.** *A transport plan  $\gamma$  is induced by a transport map if and only if  $\gamma$  is concentrated in a  $\sigma$ -compact set  $S$  such that for  $f^0$ -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 [1,2], or to Lemma 1.6 in [10] for the general case of the Polish spaces. Since the transport maps are a particular case of the transport plans, of course one has

$$\min(1.2) \leq \inf(1.1) \quad (1.3)$$

(the fact that the first one 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  $\gamma$ , and the transport plans are a bounded and 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 exist optimal transport plans.*

Concerning the existence of transport plans, just notice that the rescaled tensor product  $f^0 \boxtimes f^1$  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 [10] or [11].

In this paper we give an answer to a very natural question concerning 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. First of all, notice that the example above with the Dirac mass tells that

the existence of transport maps is not always true, and in particular that Dirac masses in  $f^0$  can prevent this existence. On the other hand, if  $f^0$  is non-atomic then there exist transport maps, as it is well known that a non-atomic measure can be transformed into any probability measure (see Theorem 1.4). Our first result is the following one.

**Theorem A.** *If  $f^0$  is non-atomic, there exists a transport map  $t$  such that*

$$\mathfrak{C}(t) \leq \mathfrak{C}(f^0 \boxtimes f^1) \quad (\leq +\infty).$$

Once established this first property, we will prove our main result – that is sharp thanks to the examples we will provide at the beginning of Section 4 – stating that a sufficient condition for the validity of (1.3) is the non-atomicity of  $f^0$  together with the continuity of  $c$ .

**Theorem B.** *Assume that  $f^0$  is non-atomic and  $c$  is continuous (but still possibly  $+\infty$ -valued somewhere): then  $\inf(1.1) = \min(1.2)$  (possibly  $+\infty$ ).*

A particular case of our theorems has been proved by Ambrosio in [1]:

**Theorem 1.3 (Ambrosio).** *If  $X = Y$  is a convex and compact subset of  $\mathbb{R}^N$ ,  $f^0$  is non-atomic, and the cost function  $c: X \times Y \rightarrow \bar{\mathbb{R}}^+$  is continuous and bounded, then the equality  $\inf(1.1) = \min(1.2)$  holds.*

Let us briefly discuss the main differences between Theorem B and the previous Theorem 1.3: we remark that the equality  $\inf(1.1) = \min(1.2)$  in Theorem 1.3 was stated in  $\mathbb{R}^N$  only for simplicity, but the proof can be easily adapted to work in a more general compact Polish space. Moreover, the relevant property is the boundedness of the cost function  $c$ , which is ensured by the compactness of the space since we consider continuous cost functions: if  $c$  is bounded, then the equality  $\inf(1.1) = \min(1.2)$  is true also for a non-compact Polish space. Therefore, the new feature of Theorem B is to allow a cost function that is not bounded, and also possibly assuming the value  $+\infty$  somewhere.

The strategy to prove Theorem 1.3 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  $\gamma$  and given the cost function  $c$ , we will build transport maps  $t_\varepsilon$  with  $\mathfrak{C}(t_\varepsilon) \leq \mathfrak{C}(\gamma) + \varepsilon$ ; but the same transport maps  $t_\varepsilon$  may not work with another cost function.

Throughout the paper, we will use the following notation: given two Polish spaces  $X$  and  $Y$  and two measures  $\mu \in \mathcal{M}^+(X)$  and  $\nu \in \mathcal{M}^+(Y)$  with  $\|\mu\| = \|\nu\| > 0$ , we define the *rescaled tensor product* of  $\mu$  and  $\nu$  as the unique measure  $\mu \boxtimes \nu \in \mathcal{M}^+(X \times Y)$  such that

$$\mu \boxtimes \nu(A \times B) = \frac{\mu(A) \cdot \nu(B)}{\|\mu\|} = \frac{\mu(A) \cdot \nu(B)}{\|\nu\|} \quad \forall A \in \mathcal{B}(X), \forall B \in \mathcal{B}(Y),$$

or in other words  $\mu \boxtimes \nu = (\mu \otimes \nu) / \|\mu\| = (\mu \otimes \nu) / \|\nu\|$ : the projections of  $\mu \boxtimes \nu$  are precisely  $\mu$  and  $\nu$ , and in particular  $f^0 \boxtimes f^1$  is the trivial transport plan between  $f^0$  and  $f^1$ .

The plan of the paper is the following: in Subsection 1.2 we discuss the relation between Theorem B and the classical results of isomorphisms of measure spaces; 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.

### 1.2. Isomorphism results and Theorem B

In [5], concerning the problem of proving the equality between  $\inf(1.1)$  and  $\min(1.2)$ , W. Gangbo suggested the idea of making use of the classical isomorphism Theorems of measure spaces. We rewrite here one of such results, due to Oxtoby (Theorem 2 in [9], see also [12], Theorem 9 in Chapter 15); in the sequel, we will denote by  $\mathcal{I} = [0, 1] \setminus \mathbb{Q}$  the set of the irrational numbers in  $[0, 1]$ , and by  $\mathcal{L}$  the Lebesgue measure on  $\mathcal{I}$ .

**Theorem 1.4.** *Let  $Z$  be a Polish space and  $\mu$  a non-atomic probability measure on  $Z$ . Then  $(Z, \mu)$  is isomorphic to  $(\mathcal{I}, \mathcal{L})$ : this means that there is a Borel subset  $\tilde{Z}$  of  $Z$  of full  $\mu$ -measure and an homeomorphism  $\varphi: \tilde{Z} \rightarrow \mathcal{I}$  with the property that  $\varphi_{\#}\mu = \mathcal{L}$ ,  $\varphi_{\#}^{-1}\mathcal{L} = \mu$ .*

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  $(X, f^0)$  to  $(Y, f^1)$ , one could use the isomorphisms given by the above theorem to reduce to the case when  $X$  and  $Y$  are  $[0, 1]$  and  $f^0$  and  $f^1$  are  $\mathcal{L}$ , apply the well known (and trivial) existence result in  $[0, 1]$  to find transport maps between  $[0, 1]$  and itself, and then use the isomorphisms again to come back to transport maps between  $X$  and  $Y$ .

Unfortunately, this argument cannot be used to derive a proof for Theorem B: indeed, the 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 discussed in the beginning of Section 4). But the continuous map  $\varphi$  in Theorem 1.4 cannot – of course – be extended to a continuous map between  $\tilde{X}$  and  $[0, 1]$ , and more in general it is not possible to give an homeomorphism between a full-measure subset of  $X$  and the whole  $[0, 1]$ . Therefore, it is possible to transform the general problem in a problem with  $X = Y = [0, 1]$ , but the continuous cost function  $c: X \times Y \rightarrow \bar{\mathbb{R}}^+$  may be transformed in a discontinuous cost function from  $[0, 1] \times [0, 1]$  to  $\bar{\mathbb{R}}^+$ , for which the searched equality is not true.

However, our approach will be in the same direction of Gangbo's proposal: we will use a clever decomposition of the Polish spaces to provide an explicit construction of almost optimal transport maps.

## 2. A general result about Polish spaces

This section is devoted to show a simple decomposition of a Polish space endowed with a probability measure; this decomposition helps to regard the Polish space as (a subset of)  $\{0, 1\}^{\mathbb{N}}$ , that is a very standard reduction in probability. 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  $\emptyset$ .*

Now, we can state and prove 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  $\mu_I$  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  $\mu = \sum_{\{l(I)=k\}} \mu_I$ ;*
- (iii) *for any finite sequence  $I$ ,  $\mu_I$  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,  $\bigcap_{k \in \mathbb{N}} Z_{I_k}$  consists of at most a point;*
- (vi) *if, in addition,  $\mu$  is non-atomic, then for any  $k \in \mathbb{N}$  the sets  $Z_I$  with  $I \in \{0, 1\}^k$  are  $2^k$  disjoint sets such that  $\mu(Z_I) = \mu(\bar{Z}_I)$  and  $\mu_I = \mu \llcorner Z_I$ .*

**Remark 2.3.** With a more elaborate proof, one can strengthen the property (vi) as follows – notice that of course (vi') is stronger than (vi):

- (vi') *if, in addition,  $\mu$  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 of Proposition 2.2.** Let us first consider the easier case of a non-atomic measure  $\mu$ : in this case, by Theorem 1.4 there is a full-measure set  $\tilde{Z} \subseteq Z$  and an homeomorphism  $\varphi: \tilde{Z} \rightarrow \mathcal{I}$ . The decomposition is trivial for  $\mathcal{I}$ : indeed it suffices to define, for any finite binary sequence  $I = (i_1, i_2, \dots, i_k)$  of length  $l(I) = k \in \mathbb{N}$ ,

$$\mathcal{I}_I := \left( \frac{i_1 2^{k-1} + i_2 2^{k-2} + \dots + i_k}{2^k}, \frac{i_1 2^{k-1} + i_2 2^{k-2} + \dots + i_k + 1}{2^k} \right) \cap \mathcal{I}, \quad \mathcal{L}_I := \mathcal{L} \llcorner \mathcal{I}_I.$$

Hence, we define  $Z_I = \varphi^{-1}(\mathcal{I}_I)$  and  $\mu_I = \mu \llcorner Z_I = \varphi_{\#}^{-1}(\mathcal{L}_I)$ : it is then obvious that (i)–(v) hold; in particular (v) hold for *all* the countable sequences instead of all but countably many. Concerning (vi), by construction the sets  $Z_I$  with fixed length  $l(I) = k$  are  $2^k$  disjoint sets, which are open inside  $\tilde{Z}$  since  $\varphi$  is continuous. Therefore, for each  $I$  one has  $Z_I \subseteq \overline{Z_I} \subseteq Z_I \cup (Z \setminus \tilde{Z})$  and, since  $Z \setminus \tilde{Z}$  is  $\mu$ -negligible, also (vi) follows.

Let us turn now to the case of a general measure, and let us write  $\mu = \mu^{na} + \mu^{pa}$ , where  $\mu^{na}$  is non-atomic while  $\mu^{pa} = \sum \alpha_n \delta_{z_n}$  is purely atomic ( $\alpha_n > 0$  and  $z_n \in Z$ ). By Theorem 1.4 we can find an isomorphism  $\varphi$  between a subset  $\tilde{Z}$  of  $Z$  of full  $\mu^{na}$ -measure and  $\mathcal{I}^{na} := (0, \|\mu^{na}\|) \cap \mathcal{I}$  such that  $\varphi_{\#} \mu^{na} = \mathcal{L}^{na} := \mathcal{L} \llcorner \mathcal{I}^{na}$  and  $\varphi_{\#}^{-1}(\mathcal{L}^{na}) = \mu^{na}$ . Let us define now a map  $\psi: \mathcal{I} \rightarrow X$  as follows: if  $x \leq \|\mu^{na}\|$  then  $\psi(x) = \varphi^{-1}(x)$ ; otherwise  $\psi(x) = z_m$ , where  $m$  is the integer such that

$$\|\mu^{na}\| + \sum_{m < n} a_m < x \leq \|\mu^{na}\| + \sum_{m \leq n} a_m;$$

finally we define, for all the finite binary sequences  $I$ ,  $Z_I := \psi(\mathcal{I}_I)$  and  $\mu_I := \psi_{\#} \mathcal{L}_I$ .

It is then obvious that all the properties (i)–(v) hold; notice that in the general case (v) may not hold for all the countable sequences, since for some  $I \in \{0, 1\}^{\mathbb{N}}$  the intersection  $\bigcap_{k \in \mathbb{N}} Z_{I_k}$  might contain two different points  $z_m$  and  $z_l$ ; however, it holds for all but countably many sequences, thus (v) is true. Notice also that, again due to the purely atomic part of  $\mu$ , it is not always true neither that  $\mu_I = \mu \llcorner Z_I$ , nor that  $\mu(\overline{Z_I}) = \mu(Z_I)$ , nor that the sets  $Z_I$  with fixed length are disjoint.  $\square$

### 3. Proof of Theorem A

In this section we will prove the existence Theorem A which asserts that, provided  $f^0$  is non-atomic, there exists some transport map having cost not greater than  $f^0 \boxtimes f^1$  (recall that  $f^0 \boxtimes f^1$  is always a transport plan). This comparison with  $f^0 \boxtimes f^1$  will be essential in next section to show the equality between  $\inf(1.1)$  and  $\min(1.2)$ .

**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  $\|f^0\| = \|f^1\| = 1$ ; then, we apply Proposition 2.2 to  $X$  and  $f^0$ , and to  $Y$  and  $f^1$ , finding the sets  $X_I$  and  $Y_I$  and the measures  $f_I^0$  and  $f_I^1$ . To show the existence of a transport map, we will build a sequence  $\{\gamma_k\}$  of transport plans between  $f^0$  and  $f^1$  weakly\* converging to a transport map (that is, weakly\* converging to a plan induced by a transport map).

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

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

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

$$\mathfrak{C}(\gamma_k) \leq \mathfrak{C}(\gamma_{k-1}) \leq \dots \leq \mathfrak{C}(\gamma_1) \leq \mathfrak{C}(\gamma_0) = \mathfrak{C}(f^0 \otimes f^1), \tag{3.2}$$

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

The first condition is taken, of course, to ensure the bound on the cost of the transport map that we will eventually define; concerning the second one, we remark that it means that the different functions  $\varphi_k$  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  $\varphi_k$  satisfying (3.2) and (3.3).

*Step II. Induction proof of the existence of the functions  $\varphi_k$ .*

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

$$\text{either } \begin{cases} \varphi_{k+1}(I, 0) = (\varphi_k(I), 0), \\ \varphi_{k+1}(I, 1) = (\varphi_k(I), 1) \end{cases} \quad \text{or} \quad \begin{cases} \varphi_{k+1}(I, 0) = (\varphi_k(I), 1) \\ \varphi_{k+1}(I, 1) = (\varphi_k(I), 0). \end{cases}$$

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

$$\varphi^\tau(I, i) = \begin{cases} (\varphi_k(I), i) & \text{if } \tau(I) = 0, \\ (\varphi_k(I), 1 - i) & \text{if } \tau(I) = 1. \end{cases} \quad (3.4)$$

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

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

By the construction, and recalling property (ii) in Proposition 2.2, it is immediately seen that the “mean” of the different  $\gamma^\tau$  is  $\gamma_k$ , that is

$$\gamma_k = \frac{1}{2^{2^k}} \sum_{\tau : \{0, 1\}^k \rightarrow \{0, 1\}} \gamma^\tau :$$

by the linearity of the cost  $\mathfrak{C}$  in (1.2), it follows the existence of at least a  $\tau : \{0, 1\}^k \rightarrow \{0, 1\}$  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  $\varphi_k$  and thanks to (3.3), we can also define  $\varphi : \{0, 1\}^{\mathbb{N}} \rightarrow \{0, 1\}^{\mathbb{N}}$  the unique function extending all the  $\varphi_k$ .

*Step III. Proof that  $\gamma_k$  weakly\* converges to a transport map  $t$  with  $\mathfrak{C}(t) \leq \mathfrak{C}(f^0 \boxtimes f^1)$ .*

Thanks to Prokhorov’s Theorem, the sequence  $\{\gamma_k\}$  of transport plans is weakly\* sequentially compact, thus there is a weak\* limit  $\gamma$  of some subsequence (recall that all the measures  $\gamma_k$  have the same marginals  $f^0$  and  $f^1$ , thus Prokhorov’s Theorem can be applied). Our aim is to show that  $\gamma$  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  $\gamma$ ; notice that this will give the thesis, since  $\mathfrak{C}(\gamma) \leq \mathfrak{C}(f^0 \boxtimes f^1)$  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.  $\gamma$  and to check that for  $f^0$ -a.e.  $x \in X$  the set  $\{y \in Y : (x, y) \in S\}$  has diameter 0.

Define  $X_k := \bigcup \{X_I : I(I) = k\}$ , that is a set of full  $f^0$ -measure by (ii) and (iii) in Proposition 2.2; hence, also the set  $\tilde{X} := \bigcap \{X_k, k \in \mathbb{N}\}$  is of full measure. By (iv) and (vi) in Proposition 2.2 we can define a function  $\psi : \tilde{X} \rightarrow \{0, 1\}^{\mathbb{N}}$  which associates to any  $x \in \tilde{X}$  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  $\text{diam}(Y_{I_k})$  does not converge to 0; since  $A$  is countable and  $\varphi$  is a bijection, and by (i) in Proposition 2.2, the set

$$\bar{X} := \{x \in \tilde{X} : \varphi(\psi(x)) \notin A\}$$

has full  $f^0$ -measure. For  $k \in \mathbb{N}$ , we define now

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

notice that by definition (3.1) the plan  $\gamma_k$  is concentrated in  $S_k$ , and thanks to (3.3) also each  $\gamma_n$  with  $n > k$  is concentrated in  $S_k$ . We want to infer that also  $\gamma$  is concentrated in  $S_k$ : to this aim recall that, by (vi) in Proposition 2.2, for any  $I \in \{0, 1\}^k$  one has  $f^0(X_I) = f^0(\bar{X}_I)$ . Consequently,

$$\begin{aligned} \gamma(X_I \times \bar{Y}_{\varphi_k(I)}) &= \gamma(\bar{X}_I \times \bar{Y}_{\varphi_k(I)}) \geq \limsup_{n \rightarrow \infty} \gamma_n(\bar{X}_I \times \bar{Y}_{\varphi_k(I)}) = \limsup_{n \rightarrow \infty} \gamma_n(\bar{X}_I \times Y) = f^0(\bar{X}_I) \\ &= f^0(X_I) = \gamma(X_I \times Y) \geq \gamma(X_I \times \bar{Y}_{\varphi_k(I)}), \end{aligned}$$

and this ensures that  $\gamma \llcorner (X_I \times Y)$  is concentrated in  $X_I \times \bar{Y}_{\varphi_k(I)}$ : since the union of the different  $X_I$  for  $I \in \{0, 1\}^k$  has full  $f^0$ -measure, we infer that in fact  $\gamma$  is concentrated in  $S_k$  – keep in mind that  $\pi_1 \gamma = f^0$ . Hence,  $\gamma$  is concentrated in  $S := \bigcap_{k \in \mathbb{N}} S_k$ : as pointed out in the beginning of this step, the thesis will be reached showing that for any  $x \in \bar{X}$  – thus, for  $f^0$ -a.e.  $x$  – the set  $\{y \in Y: (x, y) \in S\}$  has diameter 0. Indeed, if  $(x, y) \in S$  and  $x \in \bar{X}$  then  $y \in Y_{\varphi(\psi(x))}$  and, since  $\varphi(\psi(x)) \notin A$ ,  $Y_{\varphi(\psi(x))}$  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 consists of exactly a point for a.e.  $x$ ).  $\square$

#### 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  $f^0$  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.

**Example 4.1.** Let  $R, S$  and  $T$  be three parallel segments of length 2 and with  $\text{dist}(R, S) = \text{dist}(S, T) = 1$ ; then, let  $f^1$  be the one-dimensional Hausdorff measure on  $R$  and  $T$ , and let  $f^0$  be the Hausdorff measure on  $S$ , multiplied by 2 to get  $\|f^0\| = \|f^1\| = 4$ . Finally, set the non-continuous cost

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

It is clear that the minimum of the Kantorovich problem is  $\min(1.2) = 4$ , and it is achieved only 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, the infimum of the Monge problem is easily shown to be  $\inf(1.1) = 8 > \min(1.2)$  (and it is also a minimum)

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  $f^0$ : 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  $f^0$  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  $f^0$  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.

**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  $f^0 := \delta_{x^+} + \mathcal{L} \llcorner B^+$  and  $f^1 := \delta_{x^-} + \mathcal{L} \llcorner B^-$ . If 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, any 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  $f^0$  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 divide it in several steps. First of all, take an optimal transport plan  $\gamma$ , which exists thanks to Lemma 1.2; we can assume that  $\mathcal{C}(\gamma) = \min(1.2) < +\infty$ ,

since otherwise by (1.3) the claim is immediate. We fix a metric on  $X$  and on  $Y$  according to the definition of Polish spaces; 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  $\gamma$ -a.a.  $X \times Y$ . Moreover we want that  $A_I$  and  $A_J$  do not 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

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

where  $r$  is the maximum real number such that

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

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

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

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

$$\gamma\left((X \times Y) \setminus \bigcup_{i,j} C_{i,j} \times D_{i,j}\right) = 0. \quad (4.1)$$

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:

$$\forall I, \sup\{c(x, y) - c(z, w) : (x, y), (z, w) \in A_I \times B_I\} \leq \varepsilon; \quad (4.2)$$

$$I \leq J \implies A_I \supseteq A_J; \quad (4.3)$$

$$A_I \cap A_J = \emptyset \quad \text{unless } I \leq J \text{ or } J \leq I; \quad (4.4)$$

$$B_I \cap B_J = \emptyset \quad \text{if } I < J \text{ or } J < I; \quad (4.5)$$

$$\gamma\left((X \times Y) \setminus \bigcup_I (A_I \times B_I)\right) = 0. \quad (4.6)$$

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,  $\gamma(A_1 \times B_1) \geq \gamma(C_{m,n} \times D_{m,n})/2$  for any  $(m, n) \in \mathbb{N}^2$ . Then, we proceed by induction: if  $A_m$  and  $B_m$  have been chosen for  $1 \leq m < n$ , we set

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

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  $f^0$ -a.a.  $X$ , that is

$$f^0\left(X \setminus \bigcup_n A_n\right) = \gamma\left(\left(X \setminus \bigcup_n A_n\right) \times Y\right) = 0: \quad (4.7)$$

indeed, if not there would be, thanks to (4.1), a pair  $(i, j)$  such that  $\gamma((C_{i,j} \times D_{i,j}) \setminus (\bigcup_{n \in \mathbb{N}} A_n \times Y)) = \delta > 0$ ; by construction, this would imply that  $\gamma(A_n \times B_n) \geq \delta/2$  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  $\gamma(\bigcup_n A_n \times B_n) = +\infty$ , and the absurd yields (4.7).



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

$$\gamma_1 := \gamma \llcorner \left( (X \times Y) \setminus \left( \bigcup_n A_n \times B_n \right) \right);$$

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

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

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  $\gamma_1(A_{m,n} \times B_{m,n})$ ,

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

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 (4.7) can be generalized as

$$\pi_1 \gamma_1 \left( X \setminus \bigcup_{m,n} A_{m,n} \right) = 0: \tag{4.8}$$

indeed, as before, otherwise there would be by (4.1) – recall that  $\gamma_1 \leq \gamma$  – a pair  $(i, j)$  such that  $\gamma_1((C_{i,j} \setminus \bigcup_{m,n} A_{m,n}) \times D_{i,j}) > 0$ . By (4.7), we would infer the existence of some  $\widehat{m} \in \mathbb{N}$  such that  $\gamma_1((C_{i,j} \cap A_{\widehat{m}} \setminus \bigcup_{m,n} A_{m,n}) \times D_{i,j}) = \delta > 0$ . Recall that, by definition of  $\gamma_1$ ,  $\gamma_1(A_{\widehat{m}} \times B_{\widehat{m}}) = 0$ , therefore for any  $n \in \mathbb{N}$  one has

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

this implies, by the definition of  $A_{m,n}$  and  $B_{m,n}$ , that  $\gamma_1(A_{\widehat{m},n} \times B_{\widehat{m},n}) \geq \delta/2$  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 (4.8).

Iterating the same procedure, we can define the measures

$$\gamma_p := \gamma \llcorner \left( (X \times Y) \setminus \left( \bigcup_{I: l(I) \leq p} A_I \times B_I \right) \right) \tag{4.9}$$

and the sets

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

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

$$\pi_1 \gamma_{p-1} \left( X \setminus \bigcup_{l(I)=p} A_I \right) = 0. \tag{4.10}$$

We have now to check the validity of (4.2)–(4.6): (4.2)–(4.5) are obvious from the construction; concerning (4.6), if it were not true then there would exist  $(i, j)$  such that  $\gamma(\Lambda) = \delta > 0$ , where

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

using again (4.1). For any  $p \in \mathbb{N}$  and any  $I$  with  $l(I) = p$ , call now

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

thanks to (4.10), 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

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

Since  $\gamma_p \leq \gamma \forall p$  and recalling (4.9), this would give the absurd  $\|\gamma\| = +\infty$ : hence, also (4.6) is shown.

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

$$M_I := f^0(A_I), \quad m_I := \gamma(A_I \times B_I); \quad (4.11)$$

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

$$M_I \geq \sum \{m_J : J \supseteq I\} \quad (\text{in particular, } M_I = \sum \{m_J : J \supseteq I\} \text{ if } l(I) = 1), \quad (4.12)$$

$$\sum_I m_I = \|\gamma\|. \quad (4.13)$$

These two properties follow immediately by the following easy but remarkable consequence of (4.4) and (4.5):

$$A_I \times B_I \cap A_J \times B_J = \emptyset \quad \forall I \neq J \text{ finite sequences.} \quad (4.14)$$

*Step II. Definition of the sets  $X_I$ .*

In this second step we want to define sets  $X_I$ , essentially disjoint w.r.t.  $f^0$ , with the property that  $X_I \subseteq A_I$  and that  $f^0(X_I) = m_I$ .

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:

$$X_I^{p+1} := X_I^p \cup X_I^{p,+} \setminus X_I^{p,-}; \quad (4.15)$$

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).

$$\text{For any } p \geq 0, \text{ the sets } X_I^{p+1} \text{ with } l(I) \leq p+1 \text{ are disjoint w.r.t. } f^0; \quad (4.16)$$

$$\text{for any } 1 \leq l(I) \leq p+1, X_I^{p+1} \subseteq A_I; \quad (4.17)$$

$$\text{for any } 1 \leq l(I) \leq p+1, f^0(X_I^{p+1}) = m_I; \quad (4.18)$$

$$\text{for any } 1 \leq l(I) \leq p, X_I^{p,+} \cap X_I^p = \emptyset, \text{ while } X_I^{p,-} \subseteq X_I^p; \quad (4.19)$$

$$\text{for any } 1 \leq l(I) \leq p, f^0(X_I^{p,+}) = f^0(X_I^{p,-}); \quad (4.20)$$

$$\text{for any } 1 \leq l(I) \leq p, f^0(X_I^{p,-}) \leq \sum_{l(J)=p+1} m_J. \quad (4.21)$$

We leave to the next step the proof that it is possible to choose sets  $X_I^p, X_I^{s,\pm}$  in such a way; now, we only show that once fulfilled these properties 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 (4.18): notice that, in fact, the property (4.18) is a straightforward consequence of the definition (4.15) together with (4.19) and (4.20). It is clear from the construction that the sequence  $X_I^p$  converges to some set (in the sense of the measure  $f^0$ ) if the sequence  $\sigma_m := \sum_{p=m}^{+\infty} f^0(X_I^{p,-})$  converges to 0 when  $m$  goes to  $\infty$ . Write now  $\lambda_p := \sum_{l(I)=p} m_I$ : from the property (4.14) one obtains, directly from the definitions (4.9) and (4.11), that  $\lambda_p = \|\gamma_p - \gamma_{p-1}\|$ . Therefore, by (4.21) we have

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

indeed,  $\{\gamma_p\}$  is a decreasing sequence of measures by (4.9), and it strongly converges to 0 thanks to (4.6). It follows that  $\sigma_m$  converges to 0, and then that the sequence  $\{X_I^p\}$  converge in the sense of the measure  $f^0$  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 (4.16), (4.17) and (4.18), so this step is concluded.

*Step III. Construction of sets fulfilling (4.16)–(4.21).*

This step is devoted to build the sets  $X_I^p$  and  $X_I^{s,\pm}$  fulfilling the requirements (4.16)–(4.21); 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 (4.16), (4.17) and (4.18). To do that, it suffices to choose arbitrarily for any  $i \in \mathbb{N}$  a set  $X_i^1 \subseteq A_i$  with  $f^0(X_i^1) = m_i$ : this is possible since  $f^0(A_i) = M_i \geq m_i$  by (4.12), 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  $f^0(X_{i,j}^2) = m_{i,j}$ : this is possible, as before, because  $f^0(A_{i,j}) = M_{i,j} \geq m_{i,j}$  by (4.12). 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

$$X_i^{1,-} := X_i^1 \cap \left( \bigcup_{j \in \mathbb{N}} X_{i,j}^2 \right)^c.$$

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 (4.4). To complete the case  $p = 1$ , we need to define suitable sets  $X_i^{1,+}$  – and apply the definition (4.15). Notice that if we can define a set  $X_i^{1,+} \subseteq A_i$  with  $f^0(X_i^{1,+}) = f^0(X_i^{1,-})$  not intersecting  $X_i^1$  nor any of the  $X_{i,j}^2$ , then all the properties (4.16)–(4.21) are trivially satisfied. And this is possible if “there is enough space”, that is, if  $f^0(A_i) \geq \sum_j f^0(X_{i,j}^2) + m_i$ : but this means  $M_i \geq m_i + \sum_j m_{i,j}$ , that is true thanks to (4.12).

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 (4.16)–(4.21) 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 (4.16)–(4.21) 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  $f^0(X_I^{n+2}) = m_I$ . 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 (4.17), (4.20) and the property

$$X_I^{n+1,+} \cap \left( X_I^{n+1} \cup \bigcup \{X_J^{n+2} : l(I) < l(J) \leq n + 2\} \right) = \emptyset \tag{4.22}$$

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

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

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 any  $J$  with  $n + 1 \geq l(J) > s$  and a given  $1 \leq s \leq n + 1$ , in such a way that (4.17), (4.20) and (4.22) hold with  $p = n + 1$ ; so, take any sequence  $I$  of length  $l(I) = s$ , use (4.23) as definition for  $X_I^{n+1,-}$ , and consider the problem to define  $X_I^{n+1,+}$  fulfilling (4.17), (4.20) and (4.22). This would mean to select a subset of  $A_I$  of  $f^0$ -measure equal to  $f^0(X_I^{n+1,-})$  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

$$f^0\left(A_I \setminus \left(X_I^{n+1} \cup \bigcup \{X_J^{n+2} : l(I) < l(J) \leq n + 2\}\right)\right) \geq f^0(X_I^{n+1,-}). \tag{4.24}$$

Since by hypothesis we know the validity of (4.17) for  $l(J) > l(I)$  and by (4.4),  $A_I \cap X_J^{n+2} = \emptyset$  for each  $J$  with  $l(J) > l(I)$  and  $J \not\geq I$ ; moreover we already have that  $f^0(X_J^{n+2}) \leq m_J$ : indeed, this is true by construction if  $l(J) = n + 2$ , while otherwise it is true by (4.15) and (4.20), recalling that  $f^0(X_J^{n+1}) = m_J$  because (4.18) is assumed by induction in the case  $p = n$ . Therefore, keeping in mind the definition (4.23), we have that

$$\begin{aligned} & f^0\left(A_I \setminus \left(X_I^{n+1} \cup \bigcup \{X_J^{n+2} : l(I) < l(J) \leq n + 2\}\right)\right) \\ & \geq M_I - \left(\sum \{m_J : J \geq I, l(J) \leq n + 2\} - f^0(X_I^{n+1,-})\right). \end{aligned}$$

Hence, the validity of the sufficient condition (4.24) is directly given by (4.12), 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 (4.17), (4.20) and (4.22) hold.

To finish the induction and then conclude this step, we have therefore only to check the validity of (4.16), (4.18), (4.19) and (4.21). First, we notice that the property (4.18) is clear by (4.15), (4.20) and (4.22), while the property (4.19) is given by (4.22) and (4.23).

Concerning (4.16), 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 (4.17) and (4.4); otherwise, we can assume  $l(I) < l(J)$  and then the disjointness follows immediately by the definition (4.23) and the property (4.22).

Finally, concerning (4.21), rewrite (4.23) as

$$X_I^{n+1, -} = X_I^{n+1} \cap \left( \bigcup_{l(J)=n+2} X_J^{n+2} \cup \bigcup_{l(I) < l(J) < n+2} X_J^{n+2} \right); \quad (4.25)$$

then recall by (4.15) 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

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

and then (4.25) can be rewritten as

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

Since (4.16) 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

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

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 (4.25) writing

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

We conclude noticing that  $A$  and  $B$  are disjoint sets by the property (4.22), while  $C$  is contained in  $A \cup B$  by (4.23), thus  $f^0(D) = f^0(A) + f^0(B) - f^0(C)$ . Finally,  $f^0(B) = f^0(C)$  thanks to (4.20) and since both  $B$  and  $C$  are union of disjoint sets by construction; keeping in mind (4.18), we have

$$f^0(X_I^{n+1, -}) \leq f^0(D) = f^0(A) = \sum_{l(J)=n+2} m_J,$$

thus the validity of (4.21) 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 (4.14) and (4.6) it is  $\gamma = \sum_I \gamma_I$ ; then we define the measure

$$\tilde{\gamma}_I := f^0 \llcorner X_I \boxtimes \pi_2 \gamma_I.$$

This definition makes sense since  $\|\pi_2 \gamma_I\|_Y = \|\gamma_I\|_{X \times Y} = m_I = f^0(X_I)$ , therefore  $\tilde{\gamma}_I$  is a transport plan between  $f^0 \llcorner X_I$  and  $\pi_2 \gamma_I$ ; on the other hand,  $\gamma_I$  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 (4.2), one immediately gets

$$|\mathfrak{C}(\tilde{\gamma}_I) - \mathfrak{C}(\gamma_I)| \leq m_I \varepsilon. \quad (4.26)$$

We apply now Theorem A to the marginals of  $\tilde{\gamma}_I$ , obtaining a transport map  $t_I$  from  $f^0 \llcorner X_I$  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.  $f^0$ , we can glue the maps  $t_I$  to get a map  $t$ ; recalling now (4.13) to make sure that  $\cup X_I$  is of full  $f^0$ -measure, it follows that  $t$  is a transport map from  $f^0$  to  $\sum_I \pi_2 \gamma_I = \pi_2 \sum_I \gamma_I = \pi_2 \gamma = f^1$ , thus a transport map for the original problem. By (4.26), the linearity of the cost and (4.13) 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.  $\square$

## Acknowledgements

I wish to thank B. Kirchheim for pointing me out the questions originating the paper, and L. Ambrosio for his interest and the discussions we had during the preparation. I am also deeply indebted with G. Letta, who read very carefully the proofs and gave me many hints about them, for his precious and friendly help.

## References

- [1] L. Ambrosio, Lecture notes on optimal transport problems, in: *Mathematical Aspects of Evolving Interfaces*, in: *Lecture Notes in Math.*, vol. 1812, Springer, 2003, pp. 1–52.
- [2] L. Ambrosio, A. Pratelli, Existence and stability results in the  $L^1$  theory of optimal transportation, in: *Optimal Transportation and Applications*, in: *Lecture Notes in Math.*, vol. 1813, Springer, 2003, pp. 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] L.C. Evans, W. Gangbo, Differential equations methods for the Monge–Kantorovich mass transfer problem, *Mem. Amer. Math. Soc.* 137 (653) (1999).
- [5] W. Gangbo, The Monge mass transfer problem and its applications, *Contemp. Math.* 226 (1999) 79–104.
- [6] L.V. Kantorovich, On the transfer of masses, *Dokl. Akad. Nauk SSSR* 37 (1942) 227–229.
- [7] L.V. Kantorovich, On a problem of Monge, *Uspekhi Mat. Nauk* 3 (1948) 225–226.
- [8] G. Monge, *Memoire sur la Theorie des Déblais et des Remblais*, *Hist. de l’Acad. des Sciences de Paris*, 1781.
- [9] J.C. Oxtoby, Homeomorphic measures in metric spaces, *Proc. Amer. Math. Soc.* 24 (1970) 419–423.
- [10] 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/>.
- [11] S.T. Rachev, L. Rüschendorf, *Mass Transportation Problems*, Springer-Verlag, 1998.
- [12] H.L. Royden, *Real Analysis*, second ed., Macmillan, 1968.
- [13] N.S. Trudinger, X.J. Wang, On the Monge mass transfer problem, *Calc. Var. Partial Differential Equations* 13 (2001) 19–31.