ON THE SUFFICIENCY OF $c-$CYCLICAL MONOTONICITY FOR OPTIMALITY OF TRANSPORT PLANS

A. PRATELLI

Abstract. We consider a mass transport problem among Polish spaces, and we show that a transport plan concentrated in a $c-$cyclical monotone set is optimal if the cost function is continuous and possibly $+\infty$ valued; the same result is proved for a generic l.s.c. cost function in the case when the measures are purely atomic. This generalizes the previously known results.

1. Introduction

The mass transport problem, first proposed by Monge [12] in 1781, can be stated in its relaxed version (due to Kantorovich [9, 10]) as follows: we are given two Polish spaces X and Y and two positive measures $f^+ \in \mathcal{M}^+(X)$ and $f^- \in \mathcal{M}^+(Y)$ of the same total mass $\|f^+\|_X = \|f^-\|_Y$. A *transport plan* between f^+ and f^- is any measure $\gamma \in \mathcal{M}^+(X \times Y)$ whose marginals are $\pi_1\gamma = f^+$ and $\pi_2\gamma = f^-$. We are also given a l.s.c. function $c : X \times Y \to [0, +\infty]$, usually called *cost function*, and to each transport plan γ we associate a cost $\mathfrak{C}(\gamma)$ given by

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

The problem is to find a transport plan γ minimizing the cost $\mathfrak{C}(\gamma)$, and it is well-known that a solution always exists. The literature about mass transport problems is extremely wide, the interested reader may look for instance at [8, 5, 1, 19, 15, 3, 2, 4].

The fact that the values of the cost function range on $[0, +\infty]$, and so $+\infty$ is an admissible value, is to be underlined: indeed, this assumption might seem a not very important generalization of the real-valued case. However, this is not true; in fact, since all the transport plans with finite cost must be concentrated in the set $\{(x, y) : c(x, y) < +\infty\}$, the presence of a non-empty set on which $c(x, y) = +\infty$ is a strong constraint when looking for optimal transport plans. Hence, the general problem with the non real-valued cost function is quite different –and, of course, harder– than the case of the real-valued one. Moreover, some differences are known between the results which hold in the two cases: for instance, the counterexample given in [3] (which we will discuss later) makes sense only in the general case of the possibly $+\infty-$valued cost function. Lastly, and most important, there are situations for which $+\infty-$valued cost functions appear in a very reasonable way: this happens for instance when $X = Y$ is a linear space and $\Delta \subseteq X$ is a normed subspace, if one sets

$$c(x, y) := \begin{cases} \|y - x\| & \text{if } y - x \in \Delta; \\ +\infty & \text{otherwise.} \end{cases}$$

A important particular case of this setting is when $X = Y = W$ and $\Delta = H$ for an abstract Wiener space (W, μ, H), a situation studied in [7] (see also [6]): in this case, since usually the Cameron-Martin space H is $\mu-$negligible, the value of the cost function is almost always $+\infty$!

Strictly related to the transport problems is the notion of $c-$cyclical monotonicity given below, proposed by Knott and Smith [11], which generalizes the cyclical monotonicity introduced by Rockafellar [16].

Definition 1.1. *A set* $\Gamma \subseteq X \times Y$ *is said* $c-$cyclical monotone *if, for any* $n \geq 1$*, any permutation* $\sigma \in \mathfrak{S}^n$*, and any choice of* n *pairs* (x_1, y_1)*,* (x_2, y_2)*,* $\dots$*,* (x_n, y_n) *in* Γ*, one has*

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

Detailed discussions about the $c-$cyclical monotonicity can be found for instance in [11, 8, 3, 7]. The relation between this property and the transport problems is given by the fact that, *often*, a transport plan γ is optimal if and only if it is concentrated in a $c-$cyclical monotone set. Let us be now more precise.

Theorem 1.2. *If* γ *is an optimal transport plan and* $\mathfrak{C}(\gamma) < +\infty$*, then* γ *is concentrated in some* $c-$*cyclical monotone set.*

This result is well-known; the first proof was given by Rüschendorf [17] in a less general setting, then it has been generalized in [8, 3], essentially following the original construction of Rüschendorf. Notice that the claim is sharp, that is, the hypothesis $\mathfrak{C}(\gamma) < +\infty$ cannot be avoided: indeed, if for an optimal transport plan γ one has $\mathfrak{C}(\gamma) = +\infty$, this means that all the transport plans have the same cost, so any transport plan is optimal, regardless of whether it is concentrated in a $c-$cyclical monotone set or not.

Concerning the opposite implication, namely that a transport plan concentrated in a $c-$cyclical monotone set is optimal, it is known since a long time if the cost function c is continuous and the ambient space is a compact set $X = Y \subseteq \mathbb{R}^n$, see for instance [8, 1, 15]; recently, this result has been extended in [3, 13] as follows.

Theorem 1.3. *Let* γ *be a transport plan concentrated in some* $c-$*cyclical monotone set. Assume also that*

$$\begin{aligned} f^+\left(\left\{x \in X : \int_Y c(x,y)\, df^-(y) < +\infty\right\}\right) > 0 \\ f^-\left(\left\{y \in Y : \int_X c(x,y)\, df^+(x) < +\infty\right\}\right) > 0. \end{aligned} \tag{1.1}$$

Then γ *is optimal.*

In the proof of this result, as well as of the previously known versions of it, a crucial point of the strategy is to define suitable $c-$concave functions and to gain informations from a dual problem. In fact, the hypothesis (1.1) is made to obtain the well-posedness and the desired (one-sided) integrability of these $c-$concave functions.

The purpose of this paper is to find a stronger version of Theorem 1.3, needing weaker assumptions than (1.1): this seems an interesting question, and it was also proposed as an open problem by Villani in his recent book [19]. It is known that some hypothesis has to be made, because the completely general result –claiming that any transport plan concentrated in a $c-$cyclical monotone set is optimal– is false, due to a counterexample given in [3]. However, it appears convincing that the assumption (1.1) is merely technical and excessively strong; by the way, in the counterexample we have that for each $x \in X$ the function $y \mapsto c(x, y)$ has f^-–a.e. the value $+\infty$, and similarly for each $y \in Y$ one has that $c(x, y) = +\infty$ for f^+–a.e. x, hence the hypothesis (1.1) is very far from being satisfied.

In this paper we give two new results, which extend Theorem 1.3 giving weaker hypotheses under which the claim, that a transport plan concentrated in a $c-$cyclical monotone set must be optimal, is true. First of all, we show that the claim holds if f^+ and f^- are purely atomic (that is, countable sums of Dirac masses). The second result, obtained from the first one via a thorough approximation, shows that the claim is true –without assumptions on $f^\pm$– if the cost function c is continuous (but still possibly $+\infty-$valued). Therefore, it remains open the case when the cost function is not continuous and $f^\pm$ are not purely atomic. The constructions of this paper are made with a new strategy, and do not use $c-$concave functions, nor the dual version of the transport problem, nor the previously known result of Theorem 1.3.

We have recently received a very interesting preprint by Schachermayer and Teichmann [18]*, who are also interested in finding conditions under which transport plans concentrated in c−monotone sets are optimal. They are able to show that this is true in the general case of a l.s.c. cost function c which is real-valued. Therefore, their results together with the results of this paper show that the only open case remains when $f^\pm$ are not purely atomic and the cost function is neither continuous nor real-valued.*

2. Some technical definitions

In this section we briefly introduce some useful tools that we will use in the sequel. First of all, given a measure $\gamma \in \mathcal{M}^+(X \times Y)$, we call *projections* the measures $\pi_1\gamma \in \mathcal{M}^+(X)$ and $\pi_2\gamma \in \mathcal{M}^+(Y)$ given by

$$\pi_1\gamma(A) := \gamma(A \times Y), \qquad\qquad \pi_2\gamma(B) := \gamma(X \times B)$$

for any Borel set $A \subseteq X$ and $B \subseteq Y$. Notice that, of course,

$$\|\gamma\|_{X\times Y} = \|\pi_1\gamma\|_X = \|\pi_2\gamma\|_Y. \tag{2.1}$$

A measure $\mu \in \mathcal{M}^+(X)$ is said to be *concentrated in* $\Delta \subseteq X$ if $\mu(X \setminus \Delta) = 0$; in particular the *support of* μ, $\mathrm{spt}\,(\mu)$, is defined as the smallest closed set in which μ is concentrated. One can easily notice that

$$\mathrm{spt}\,(\mu) = \{x \in X :\ \forall U |\ U \text{ is a neighborhood of } x,\, \mu(U) > 0\}.$$

We underline that the support is a uniquely-defined set, while there may be many sets in which the measure is concentrated. However, it is often useful to consider some of these sets instead of the support: for instance, in Theorem 1.2 it is essential to say that γ is concentrated in some $c-$cyclical monotone set, while in general the support of an optimal transport plan need not to be $c-$cyclical monotone.

Given two measures $\mu \in \mathcal{M}^+(X)$ and $\nu \in \mathcal{M}^+(Y)$, we write $\mu \otimes \nu$ to denote the tensor product, i.e. the unique measure on $X \times Y$ such that for any Borel sets $A \subseteq X$, $B \subseteq Y$ one has

$$\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B); \tag{2.2}$$

notice that $\pi_1(\mu \otimes \nu) = \|\nu\|\mu$ and $\pi_2(\mu \otimes \nu) = \|\mu\|\nu$.

A measure $\mu \in \mathcal{M}^+(X)$ such that $\mu(\{x\}) = 0$ for any $x \in X$ is said *non-atomic*. Conversely, μ is said *purely atomic* if it is concentrated in a countable set, that is, there exists a set $\Delta \subseteq X$

containing countably many elements such that $\mu(X \setminus \Delta) = 0$. Equivalently, μ is purely atomic if and only if there are two sequences $\{x_n\} \subseteq X$ and $\{\alpha_n\} \in \mathbb{R}^+$ with

$$\mu = \sum \alpha_n \delta_{x_n}.$$

Given a purely atomic measure $\mu \in \mathcal{M}^+(X)$, there is a distinguished set on which μ is concentrated, the set $\Gamma(\mu) \subseteq \mathcal{M}^+(X)$ given by

$$\Gamma(\mu) := \big\{x \in X : \ \mu(\{x\}) > 0\big\}.$$

Remark 2.1. *Notice that $\Gamma(\mu) \subseteq \operatorname{spt}(\mu)$, and the inclusion may in general be strict. Notice also that $\Gamma(\mu)$ is a countable set, and it is the smallest set on which μ is concentrated: therefore, for any set $\Delta \subseteq X$ one has that μ is concentrated in Δ if and only if $\Gamma(\mu) \subseteq \Delta$.*

Let us consider now two purely atomic measures $f^+ \in \mathcal{M}^+(X)$ and $f^- \in \mathcal{M}^+(Y)$, and take a transport plan γ between f^+ and f^-. It is obvious that γ is concentrated in the countable set $\Gamma(f^+) \times \Gamma(f^-) \subseteq X \times Y$; hence, γ is also purely atomic, so we can consider the set $\Gamma(\gamma)$, and moreover $\Gamma(\gamma) \subseteq \Gamma(f^+) \times \Gamma(f^-)$. In addition, one can easily derive that

$$\pi_1 \Gamma(\gamma) = \Gamma(\pi_1\gamma)) = \Gamma(f^+), \qquad\qquad \pi_2 \Gamma(\gamma) = \Gamma(\pi_2\gamma)) = \Gamma(f^-). \tag{2.3}$$

The last tool that we give, taken from [14], is the notion of restriction of transport plans.

Definition 2.2. *Let γ be a transport plan between f^+ and f^-, and let $\mu \leq f^+$ and $\nu \leq f^-$; write then $\mu = \varphi f^+$ with $\varphi : X \to [0,1]$ and $\nu = \psi f^-$ with $\psi : Y \to [0,1]$. We call* restrictions *of γ the measures $\gamma_{|\mu}$ and $\gamma^{|\nu}$ given by*

$$\gamma_{|\mu} = \varphi\gamma, \qquad\qquad \gamma^{|\nu} = \psi\gamma.$$

Notice that $\gamma_{|\mu}$ is a transport plan between μ and $\pi_2\big(\gamma_{|\mu}\big) \leq f^-$, and $\gamma^{|\nu}$ is a transport plan between $\pi_1\big(\gamma^{|\nu}\big) \leq f^+$ and ν. Notice also that $\|\gamma_{|\mu}\| = \|\mu\|$ and $\|\gamma^{|\nu}\| = \|\nu\|$, as a particular case of (2.1).

3. The purely atomic case

In this section, we show that if the measures f^+ and f^- are purely atomic, then the $c-$cyclical monotonicity is a sufficient condition for the optimality. In other words, if a transport plan γ between f^+ and f^- is concentrated in a $c-$cyclical monotone set, we are able to show that γ is optimal. In the purely atomic case we do not need the hypothesis that c is continuous.

Theorem A. *Let $f^+ \in \mathcal{M}^+(X)$, $f^- \in \mathcal{M}^+(Y)$ be purely atomic, and let $\gamma \in \mathcal{M}^+(X \times Y)$ be a transport plan. Then γ is optimal if and only if it is concentrated in a $c-$cyclical monotone set. More precisely, γ is optimal if and only if $\Gamma(\gamma)$ is $c-$cyclical monotone.*

Proof. One implication is trivial by Theorem 1.2 recalling Remark 2.1; to show the other one, we suppose that γ is *not* optimal, and we must show that $\Gamma(\gamma)$ is *not* $c-$cyclical monotone. We divide the proof of this fact in five steps.

Step I. *Preliminary definitions.*
Since γ is not optimal, we can select another transport plan $\bar\gamma$ between f^+ and f^- such that

$$\mathfrak{C}(\bar\gamma) < \mathfrak{C}(\gamma).$$

Assume for the moment that

$$\gamma \wedge \bar\gamma = 0; \tag{3.1}$$

we will remove this assumption in the last step of the proof. According to the discussion after Remark 2.1, $\bar\gamma$ is concentrated in the set $\Gamma(\bar\gamma)$; we fix now a point $(x_1,y_1)\in\Gamma(\bar\gamma)$, and define the sets

$$\begin{aligned}B_1 &:= \{1\},\\ A_1 &:= \big\{j\in\mathbb{N}:\ (x_1,y_j)\in\Gamma(\gamma)\big\},\end{aligned}$$

and recursively, for any $n\geq 2$,

$$\begin{aligned}B_n &:= \big\{i\in\mathbb{N}:\ \exists\, j\in A_{n-1}:\ (x_i,y_j)\in\Gamma(\bar\gamma)\big\},\\ A_n &:= \big\{j\in\mathbb{N}:\ \exists\, i\in B_n:\ (x_i,y_j)\in\Gamma(\gamma)\big\}.\end{aligned}$$

Step II. *There exists $n\in\mathbb{N}$ such that $1\in A_n$.*

To show this claim we perform a recursive construction of measures $f_m^+\in\mathcal{M}^+(X)$, $f_m^-\in\mathcal{M}^+(Y)$, $\gamma_m\in\mathcal{M}^+(X\times Y)$ and $\bar\gamma_m\in\mathcal{M}^+(X\times Y)$ such that the following properties hold for $m\in\mathbb{N}$:

$$\sum\nolimits_{h=1}^m f_h^+\leq f^+ \tag{3.2}$$

$$\sum\nolimits_{h=1}^m f_h^-\leq f^- \tag{3.3}$$

$$f_1^-\wedge f_m^-=0\quad \text{if } m\geq 2 \tag{3.4}$$

$$\Gamma(f_m^+)=\big\{x_i:\ i\in B_m\big\} \tag{3.5}$$

$$\Gamma(f_m^-)=\big\{y_j:\ j\in A_{m-1}\big\}\quad \text{if } m\geq 2 \tag{3.6}$$

$$\begin{cases}\gamma_m=\gamma_{|f_{m-1}^+} & \text{if } m\geq 2\\ \pi_2\gamma_m=f_m^- & \text{if } m\geq 2\end{cases} \tag{3.7}$$

$$\begin{cases}\bar\gamma_m=\gamma^{|f_m^-} & \text{if } m\geq 2\\ \pi_1\bar\gamma_m=f_m^+. & \end{cases} \tag{3.8}$$

The first step is to define $\gamma_1=0$, $\bar\gamma_1=\bar\gamma\llcorner\{(x_1,y_1)\}$, $f_1^+=\pi_1\bar\gamma_1$ and $f_1^-=\pi_2\bar\gamma_1$: with $m=1$, (3.2), (3.3), (3.5) and (3.8) are immediate from the definition, while (3.4), (3.6) and (3.7) are emptily true. Let us then assume that $f_m^\pm$, γ_m and $\bar\gamma_m$ have been already defined for $1\leq m\leq n$, and that the properties (3.2)–(3.8) hold for any $1\leq m\leq n$; we must carry out the $(n+1)$−th step.

To begin, we set

$$\gamma_{n+1}:=\gamma_{|f_n^+},\qquad f_{n+1}^-:=\pi_2\gamma_{n+1},\qquad \bar\gamma_{n+1}:=\bar\gamma^{|f_{n+1}^-},\qquad f_{n+1}^+:=\pi_1\bar\gamma_{n+1};$$

notice that the definitions of γ_{n+1} and $\bar\gamma_{n+1}$ make sense because $f_n^+\leq f^+=\pi_1\gamma$ by (3.2) with $m=n$, and $f_{n+1}^-\leq\pi_2\gamma=f^-=\pi_2\bar\gamma$ by definition. We must now check whether the properties (3.2)–(3.8) hold with $m=n+1$. The properties (3.7) and (3.8) are satisfied by construction; concerning (3.6), we first underline that

$$\text{if } i\in B_n,\text{ then}\qquad \Big((x_i,y_j)\in\Gamma(\gamma_{n+1})\Longleftrightarrow(x_i,y_j)\in\Gamma(\gamma)\Big); \tag{3.9}$$

indeed, $\Longrightarrow$ is obvious since $\gamma_{n+1}\leq\gamma$; on the other hand, it immediately follows from the definition of the restriction that, given a purely atomic measure γ and $f_n^+\leq\pi_1\gamma$,

$$\Gamma(\gamma_{|f_n^+})=\Gamma(\gamma)\cap\big(\Gamma(f_n^+)\times Y\big);$$

therefore, also $\Longleftarrow$ is true recalling that if $i \in B_n$ then $x_i \in \Gamma(f_n^+)$ thanks to (3.5) with $m = n$. Hence, (3.6) is proved because, using the definition of γ_{n+1}, (2.3), property (3.5) with $m = n$, (3.9) and the definition of A_n, one has

$$\begin{aligned} y_j \in \Gamma(f_{n+1}^-) &\Longleftrightarrow y_j \in \Gamma(\pi_2\gamma_{n+1}) \Longleftrightarrow y_j \in \pi_2\big(\Gamma(\gamma_{n+1})\big) \Longleftrightarrow \exists\, x_i : (x_i, y_j) \in \Gamma(\gamma_{n+1}) \\ &\Longleftrightarrow \exists\, x_i \in \Gamma(f_n^+) : (x_i, y_j) \in \Gamma(\gamma_{n+1}) \Longleftrightarrow \exists\, i \in B_n : (x_i, y_j) \in \Gamma(\gamma_{n+1}) \\ &\Longleftrightarrow \exists\, i \in B_n : (x_i, y_j) \in \Gamma(\gamma) \Longleftrightarrow j \in A_n. \end{aligned}$$

Take now the property (3.4): if it does not hold, then $f_1^- \wedge f_{n+1}^- > 0$; but since $\Gamma(f_1^-) = \{y_1\}$ by definition, this would mean that $y_1 \in \Gamma(f_{n+1}^-)$, i.e. that $1 \in A_n$ thanks to (3.6) with $m = n+1$ that we have just proved. Hence, either the claim of this step is achieved (and we break this construction), or also (3.4) holds true. In this last case, the property (3.3) is easy to prove: indeed, using (3.7) with $2 \leq m \leq n+1$ and (3.2) with $m = n$ we obtain that

$$\sum\nolimits_{h=2}^{n+1} \gamma_h = \gamma_{|f_1^+ + \cdots + f_n^+} \leq \gamma,$$

so that, again by (3.7),

$$f^- = \pi_2\gamma \geq \pi_2\Big(\sum\nolimits_{h=2}^{n+1} \gamma_h\Big) = \sum\nolimits_{h=2}^{n+1} \pi_2\gamma_h = \sum\nolimits_{h=2}^{n+1} f_h^-;$$

we infer the validity of (3.3) since it is also $f^- \geq f_1^-$ and

$$f_1^- \wedge \Big(\sum\nolimits_{h=2}^{n+1} f_h^-\Big) = 0$$

by (3.4) with $2 \leq m \leq n+1$. Observe now that by (3.3) and (3.8) we have

$$\sum\nolimits_{h=2}^{n+1} \bar\gamma_h = \gamma^{|f_2^- + \cdots + f_{n+1}^-} \leq \bar\gamma;$$

but since $\bar\gamma_1$ is concentrated in $(\{x_1, y_1\})$ while $y_1 \notin \Gamma(\pi_2\bar\gamma_h)$ for $2 \leq h \leq n+1$ by (3.4), we deduce that in fact

$$\sum\nolimits_{h=1}^{n+1} \bar\gamma_h \leq \bar\gamma$$

so that projecting both sides on X we find also (3.2). Finally, checking (3.5) is almost identical to what we done for (3.6): in the very same way of (3.9) one can show

$$\text{if } j \in A_n, \text{ then } \qquad \Big((x_i, y_j) \in \Gamma(\bar\gamma_{n+1}) \Longleftrightarrow (x_i, y_j) \in \Gamma(\bar\gamma)\Big); \tag{3.10}$$

and then, using the definition of $\bar\gamma_{n+1}$, (2.3), property (3.6) with $m = n+1$, (3.10) and the definition of B_{n+1}, one has

$$\begin{aligned} x_i \in \Gamma(f_{n+1}^+) &\Longleftrightarrow x_i \in \Gamma(\pi_1\bar\gamma_{n+1}) \Longleftrightarrow x_i \in \pi_1\big(\Gamma(\bar\gamma_{n+1})\big) \Longleftrightarrow \exists\, y_j : (x_i, y_j) \in \Gamma(\bar\gamma_{n+1}) \\ &\Longleftrightarrow \exists\, y_j \in \Gamma(f_{n+1}^-) : (x_i, y_j) \in \Gamma(\bar\gamma_{n+1}) \Longleftrightarrow \exists\, j \in A_n : (x_i, y_j) \in \Gamma(\bar\gamma_{n+1}) \\ &\Longleftrightarrow \exists\, j \in A_n : (x_i, y_j) \in \Gamma(\bar\gamma) \Longleftrightarrow j \in B_{n+1}. \end{aligned}$$

So, also (3.5) is shown.

Summarizing, either we find $n \in \mathbb{N}$ such that $1 \in A_n$ –thus this step is concluded– or we can go on iterating our construction. Now, let us define $\delta := \bar\gamma(\{(x_1, y_1\}) > 0$: one has that $\|f_n^+\| = \delta$ for any n: indeed, we can argue again recursively noticing that $\|f_1^+\| = \|\bar\gamma_1\| = \delta$ by definition, and if $\|f_n^+\| = \delta$ then by (3.8) and (3.7) one has

$$\|f_{n+1}^+\| = \|\bar\gamma_{n+1}\| = \|f_{n+1}^-\| = \|\gamma_{n+1}\| = \|f_n^+\| = \delta.$$

By (3.2), our construction must stop at some $n \leq \|f^+\|/\delta$, and as noticed an interruption in the construction means that the claim of this step is shown.

Step III. *Existence of a "cyclical pair".*
Thanks to Step II, we know that $1 \in A_n$ for some n; we can exclude $n = 1$ because this would mean that $(x_1, y_1) \in \Gamma(\gamma)$, while this is not possible since $(x_1, y_1) \in \Gamma(\bar\gamma)$ and by (3.1). So, for some $n > 1$ we have $1 \in A_n$ which implies the existence of some $i(n) \in B_n$ with $(x_{i(n)}, y_1) \in \Gamma(\gamma)$; the fact that $i(n) \in B_n$ ensures the existence of $j(n-1) \in A_{n-1}$ such that $(x_{i(n)}, y_{j(n-1)}) \in \Gamma(\bar\gamma)$. Iterating, we find $i(h) \in B_h$ with $1 \leq h \leq n$ and $j(h) \in A_h$ with $1 \leq h \leq n-1$ such that

$$(x_{i(h)}, y_{j(h)}) \in \Gamma(\gamma) \quad \forall\, 1 \leq h \leq n-1, \qquad\qquad (x_{i(n)}, y_1) \in \Gamma(\gamma),$$

and

$$(x_{i(h)}, y_{j(h-1)}) \in \Gamma(\bar\gamma) \quad \forall\, 2 \leq h \leq n;$$

but $i_1 \in B_1 = \{1\}$, so $x_{i(1)} = x_1$; therefore, since $(x_1, y_1) \in \Gamma(\bar\gamma)$ by construction, setting $j(n) = 1$ and re-indexing for semplicity the elements X and Y so that $i(h) = j(h) = h$ for any $1 \leq h \leq n$, we found two finite sequences $(x_1,\, x_2,\, \dots,\, x_n) \in X$ and $(y_1,\, y_2,\, \dots,\, y_n) \in Y$ with the property that for any $1 \leq i \leq n$

$$(x_i, y_i) \in \Gamma(\gamma), \qquad\qquad (x_i, y_{\sigma(i)}) \in \Gamma(\bar\gamma),$$

being $\sigma \in \mathfrak{S}^n$ the cyclical permutation $\sigma(i) = i - 1 \pmod n$. Let then

$$\varepsilon := \min\{\gamma(\{(x_i, y_i)\})\} \wedge \min\{\bar\gamma(\{(x_i, y_{\sigma(i)})\})\} > 0,$$

and notice that, defining

$$\mu = \varepsilon \sum\nolimits_{i=1}^n \delta_{(x_i, y_i)}, \qquad\qquad \bar\mu = \varepsilon \sum\nolimits_{i=1}^n \delta_{(x_i, y_{\sigma(i)})}, \tag{3.11}$$

one has $\mu \leq \gamma$ and $\bar\mu \leq \bar\gamma$. We call a pair $(\mu, \bar\mu)$ as in (3.11), of course with $\varepsilon > 0$, a *cyclical pair of mass* ε; so, together with Step I and Step II we have shown that, for any two measures γ and $\bar\gamma$ with the same marginals f^+, f^- purely atomic and satisfying (3.1), there exists a cyclical pair $(\mu, \bar\mu)$ with $\mu \leq \gamma$ and $\bar\mu \leq \bar\gamma$.

Step IV. *Conclusion.*
Consider γ and $\bar\gamma$, and let $(\mu_1 \leq \gamma, \bar\mu_1 \leq \bar\gamma)$ be a cyclical pair with almost maximal mass ε_1; by "almost optimal" we mean that there exist no cyclical pairs of mass greater than $2\varepsilon_1$. If $\mu_1 = \gamma$ and $\bar\mu_1 = \bar\gamma$ we stop here, otherwise we apply the same argument to $\gamma - \mu_1$ and $\bar\gamma - \bar\mu_1$ (which have the same marginals), finding a cyclical pair $(\mu_2 \leq \gamma - \mu_1, \bar\mu_2 \leq \bar\gamma - \bar\mu_1)$ of almost optimal mass ε_2. Iterating this process, we find a finite or countable sequence of cyclical pairs $(\mu_i, \bar\mu_i)$ of masses ε_i such that

$$\sum \mu_i \leq \gamma, \qquad\qquad \sum \bar\mu_i \leq \bar\gamma.$$

In fact, these inequalities are both equalities: otherwise, there would be a cyclical pair $(\mu, \bar\mu)$ of mass $\varepsilon > 0$ with $\mu \leq \gamma - \sum \mu_i$ and $\bar\mu \leq \bar\gamma - \sum \bar\mu_i$; since of course $\varepsilon_i \to 0$, it would be $\varepsilon > 2\varepsilon_i$ for some i, then the choice of the pair $(\mu_i, \bar\mu_i)$ would have not been almost optimal. So, we have

$$\gamma = \sum \mu_i, \qquad\qquad \bar\gamma = \sum \bar\mu_i. \tag{3.12}$$

By the linearity of the cost and since $\mathfrak{C}(\bar\gamma) < \mathfrak{C}(\gamma)$, there is $m \in \mathbb{N}$ such that $\mathfrak{C}(\bar\mu_m) < \mathfrak{C}(\mu_m)$; recalling (3.11), this gives us the existence of two finite sequences $(x_1,\, x_2,\, \dots,\, x_n) \in X$ and $(y_1,\, y_2,\, \dots,\, y_n) \in Y$ with

$$\sum\nolimits_{i=1}^n c(x_i, y_i) = \frac{\mathfrak{C}(\mu_m)}{\varepsilon} > \frac{\mathfrak{C}(\bar\mu_m)}{\varepsilon} = \sum\nolimits_{i=1}^n c(x_i, y_{\sigma(i)}).$$

But $\mu_m \leq \gamma$, then $(x_i, y_i) \in \Gamma(\gamma)$ for any $1 \leq i \leq n$; so, by the above inequality we finally found that $\Gamma(\gamma)$ is *not* $c-$cyclical monotone and we have concluded the proof under the assumption (3.1).

Step V. *Removing assumption (3.1).*
To conclude, we take two generic transport plans γ and $\bar\gamma$ between f^+ and f^- with $\mathfrak{C}(\bar\gamma) < \mathfrak{C}(\gamma)$ and we must check that $\Gamma(\gamma)$ is not $c-$cyclical monotone. Define

$$\eta := \gamma \wedge \bar\gamma, \qquad \gamma' := \gamma - \eta, \qquad \bar\gamma' := \bar\gamma - \eta;$$

clearly, γ' and $\bar\gamma'$ are two transport plans with the same marginals, for which the assumption (3.1) holds, and such that $\mathfrak{C}(\bar\gamma') = \mathfrak{C}(\bar\gamma) - \mathfrak{C}(\eta) < \mathfrak{C}(\gamma) - \mathfrak{C}(\eta) = \mathfrak{C}(\gamma')$ by the linearity of the cost (notice that $\mathfrak{C}(\eta) < +\infty$ because $\eta \leq \bar\gamma$ and $\mathfrak{C}(\bar\gamma) < +\infty$ because $\mathfrak{C}(\bar\gamma) < \mathfrak{C}(\gamma)$). The first four steps ensure then us that $\Gamma(\gamma')$ is not $c-$cyclical monotone; to finish the proof, it suffices to recall that since $\gamma' \leq \gamma$ one has $\Gamma(\gamma') \subseteq \Gamma(\gamma)$, so a fortiori we derive that $\Gamma(\gamma)$ is not $c-$cyclical monotone. □

4. The general case

In this section, we will prove our main result.

Theorem B. *Assume that the cost function $c : X \times Y \to [0, +\infty]$ is continuous. Then a transport plan γ is optimal if and only* $\operatorname{spt}(\gamma)$ *is $c-$cyclical monotone.*

As for Theorem A, one implication will be given by the well-known Theorem 1.2; to show the other one, we will again take two transport plans γ and $\bar\gamma$ with $\mathfrak{C}(\bar\gamma) < \mathfrak{C}(\gamma)$, and we will show that $\operatorname{spt}(\gamma)$ is not $c-$cyclical monotone. In fact, as the next lemma shows, in the case when c is continuous it is admissible to consider the support of γ instead of a generic set on which γ is concentrated.

Lemma 4.1. *If c is continuous, there exists a $c-$cyclical monotone set Γ on which γ is concentrated if and only if* $\operatorname{spt}(\gamma)$ *is $c-$cyclical monotone.*

Proof. Of course one implication is trivial, since any measure is concentrated on its support. Suppose then that γ is concentrated in the $c-$cyclical monotone set Γ and take $n \in \mathbb{N}$, a permutation $\sigma \in \mathfrak{S}^n$ and n pairs $(x_i, y_i) \in \operatorname{spt}(\gamma)$ with $1 \leq i \leq n$. By definition of support, for any $1 \leq i \leq n$ we can take a sequence $\{(x_i^j, y_i^j)\}_{j\in\mathbb{N}}$ converging to (x_i, y_i) for $j \to +\infty$ and so that $(x_i^j, y_i^j) \in \Gamma$ for any $1 \leq i \leq n$ and $j \in \mathbb{N}$. Since Γ is $c-$cyclical monotone one has, for any $j \in \mathbb{N}$,

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

moreover, by the continuity of c one has $c(x_i^j, y_i^j) \to c(x_i, y_i)$ and $c(x_i^j, y_{\sigma(i)}^j) \to c(x_i, y_{\sigma(i)})$, so we can pass to the limit in the preceding inequality and the assert is shown. □

In order to show Theorem B, we need a couple of definitions.

Definition 4.2. *Let $\gamma, \bar\gamma \in \mathcal{M}^+(X \times Y)$ be two measures with the same marginals (then, two transport plans relative to the same measures). We call $\varepsilon-$*step *a new pair of measueres*

$$\gamma' = \gamma + \gamma^+ - \gamma^-, \qquad \bar\gamma' = \bar\gamma + \bar\gamma^+ - \bar\gamma^-$$

such that:

(1) *$\gamma^\pm$ and $\bar\gamma^\pm$ are positive measures on $X \times Y$, with $\gamma^- \leq \gamma$ and $\bar\gamma^- \leq \bar\gamma$;*

(2) $\|\gamma^+\| = \|\gamma^-\| = \alpha$, $\|\bar\gamma^+\| = \|\bar\gamma^-\| = 2\alpha$;

(3) γ^+ *is a Dirac mass, while* γ^- *is non-atomic;*

(4) $\pi_i\gamma' = \pi_i\bar\gamma'$ *with* $i = 1, 2$*;*

(5) $\mathrm{spt}\,(\gamma') \subseteq \mathrm{spt}\,(\gamma)$;

(6) $\mathfrak{C}(\gamma') \geq \mathfrak{C}(\gamma)$, $\mathfrak{C}(\bar\gamma') \leq \mathfrak{C}(\bar\gamma) + 2\varepsilon\alpha$.

The number $\alpha > 0$ *is called* power *of the step.*

Definition 4.3. *Given a measure* $\mu \in \mathcal{M}^+(X)$*, the* atomic part *of* μ *is the biggest purely atomic measure* $\mu_{\mathrm{AT}} \leq \mu$*.*

Remark 4.4. *For any measure* $\mu \in \mathcal{M}^+(X)$*,* $\mu - \mu_{\mathrm{AT}}$ *is non-atomic. Hence, any measure* μ *can be (uniquely) decomposed writing* $\mu = \mu_{\mathrm{AT}} + (\mu - \mu_{\mathrm{AT}})$ *as the sum of a purely atomic part and of a non-atomic one.*

Remark 4.5. *Any* $\varepsilon-$*step increases the atomic part of the measure* γ*; that is,* $\gamma'_{\mathrm{AT}} = \gamma_{\mathrm{AT}} + \gamma^+ \geq \gamma_{\mathrm{AT}}$*. In particular,* $\|\gamma'_{\mathrm{AT}}\| - \|\gamma_{\mathrm{AT}}\|$ *coincides with the power of the step.*

Let us show now the key ingredient of the proof of Theorem B, ensuring that it is always possible to find some $\varepsilon-$step.

Lemma 4.6. *Let* γ, $\bar\gamma$ *be two measures with the same marginals and with* $\mathfrak{C}(\bar\gamma) < +\infty$*, and let* $c : X \times Y \to [0, +\infty]$ *be continuous. Then, for any* $\varepsilon > 0$ *there exists an* $\varepsilon-$*step, unless* γ *is purely atomic.*

Proof. Since γ is not purely atomic, $\nu_1 = \gamma - \gamma_{\mathrm{AT}} > 0$; therefore $\pi_1\nu_1 \leq f^+$, thus

$$\mu_1 := \frac{\bar\gamma_{|\pi_1\nu_1}}{2}$$

is a positive measure less than $\bar\gamma$, so $\mathfrak{C}(\mu_1) \leq \mathfrak{C}(\bar\gamma) < +\infty$. We can then find a positive measure $\mu_2 \leq \mu_1$ concentrated in the rectangle $B_1 \times C_1$, where $B_1 \subseteq X$ and $C_1 \subseteq Y$ are two closed balls such that the cost function c is real-valued and almost constant in $B_1 \times C_1$, namely

$$\mathrm{diam}\big(c(B_1 \times C_1)\big) \leq \varepsilon. \tag{4.1}$$

By definition of restriction, $\pi_1\mu_2 \leq \pi_1\mu_1 = \pi_1\nu_1/2$, so we can find

$$0 < \nu_2 \leq \nu_{1|\pi_1\mu_2}$$

concentrated in some rectangle $B_2 \times C_2$; by construction we have $B_2 \subseteq B_1$, while C_1 and C_2 are generic: just for simplicity, we drawn them disjoint in Figure 1 (to help following the construction, the different stages are drawn so that they are made from the lighter to the darker.). Let us consider now $\bar\gamma^{|\pi_2\nu_2}$, which is defined since of course $\pi_2\nu_2 \leq f^-$: it is again a positive measure less than $\bar\gamma$, then with finite cost. Hence, as done before, we can find a measure

$$0 < \beta \leq \frac{\bar\gamma^{|\pi_2\nu_2}}{2}$$

concentrated in the rectangle $B_3 \times C_3$ with

$$\mathrm{diam}\big(c(B_3 \times C_3)\big) \leq \varepsilon. \tag{4.2}$$

Similarly as before, we have $C_3 \subseteq C_2$, while B_2 and B_3 are generic. Finally, we can set

$$\nu := \nu_2{}^{|\pi_2\beta}, \qquad\qquad \mu := \mu_{2|\pi_1\nu},$$

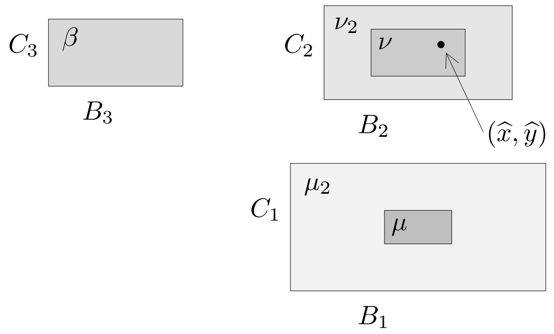

FIGURE 1. Construction of the measures μ, ν and β

and consider the three measures $\beta \leq \bar{\gamma}$, $\nu \leq \gamma$ and $\mu \leq \bar{\gamma}$. First of all, by construction we have $\mu + \beta \leq \bar{\gamma}$, since $\mu \leq \mu_2 \leq \mu_1 \leq \bar{\gamma}/2$ and analogously $\beta \leq \bar{\gamma}/2$. Moreover,

$$\pi_1 \nu = \pi_1 \mu, \qquad\qquad \pi_2 \nu = \pi_2 \beta. \tag{4.3}$$

Select now a point $(\widehat{x}, \widehat{y}) \in \operatorname{spt}(\nu)$ such that

$$c(\widehat{x}, \widehat{y}) \geq \fint c(x, y)\, d\nu \tag{4.4}$$

(both sides of this inequality may be $+\infty$), set $\alpha = \|\nu\|$, and finally define the desired measures

$$\gamma^+ := \alpha \delta_{(\widehat{x}, \widehat{y})}, \qquad \gamma^- := \nu, \qquad \bar{\gamma}^- := \mu + \beta, \qquad \bar{\gamma}^+ := (\delta_{\widehat{x}} \otimes \pi_2 \mu) + (\pi_1 \beta \otimes \delta_{\widehat{y}}).$$

We claim that $\gamma' = \gamma + \gamma^+ - \gamma^-$ and $\bar{\gamma}' = \bar{\gamma} + \bar{\gamma}^+ - \bar{\gamma}^-$ are an $\varepsilon-$step of power α, so to conclude we need to check the validity of properties (1)–(6) of Definition 4.2.

The property (1) is obvious from the construction. Concerning the property (2), $\|\gamma^-\| = \|\nu\| = \alpha$ is the definition of α, and $\|\gamma^+\| = \alpha$ since γ^+ is a Dirac mass of mass α; moreover, by (4.3) and recalling (2.1) we have that $\|\beta\| = \|\mu\| = \|\nu\| = \alpha$, so that $\|\bar{\gamma}^-\| = 2\alpha$ follows; finally,

$$\|\bar{\gamma}^+\| = \|\delta_{\widehat{x}} \otimes \pi_2 \mu\| + \|\pi_1 \beta \otimes \delta_{\widehat{y}}\| = \|\pi_2 \mu\| + \|\pi_1 \beta\| = \|\mu\| + \|\beta\| = 2\alpha.$$

The property (3) is satisfied since γ^+ is a Dirac mass by definition, while $\gamma^- \leq \nu_2 \leq \nu_1 = \gamma - \gamma_{\rm AT}$ is non-atomic in view of Remark 4.4. The property (4) is easily verified since, recalling the properties of the restriction, (2.2) and (4.3),

$$\begin{aligned}
\pi_1 \gamma' &= \pi_1 \gamma + \pi_1 \gamma^+ - \pi_1 \gamma^- = f^+ + \alpha \delta_{\widehat{x}} - \pi_1 \nu, \\
\pi_1 \bar{\gamma}' &= \pi_1 \bar{\gamma} + \pi_1 \bar{\gamma}^+ - \pi_1 \bar{\gamma}^- = f^+ + \alpha \delta_{\widehat{x}} + \pi_1 \beta - \pi_1 \mu - \pi_1 \beta = f^+ + \alpha \delta_{\widehat{x}} - \pi_1 \nu, \\
\pi_2 \gamma' &= \pi_2 \gamma + \pi_2 \gamma^+ - \pi_2 \gamma^- = f^- + \alpha \delta_{\widehat{y}} - \pi_2 \nu, \\
\pi_2 \bar{\gamma}' &= \pi_2 \bar{\gamma} + \pi_2 \bar{\gamma}^+ - \pi_2 \bar{\gamma}^- = f^- + \pi_2 \mu + \alpha \delta_{\widehat{y}} - \pi_2 \mu - \pi_2 \beta = f^- + \alpha \delta_{\widehat{y}} - \pi_2 \nu.
\end{aligned}$$

The property (5) is true since we selected the point $(\widehat{x}, \widehat{y})$ inside $\operatorname{spt}(\nu) \subseteq \operatorname{spt}(\gamma)$, and of course

$$\operatorname{spt}(\gamma') \subseteq \operatorname{spt}(\gamma) \cup \operatorname{spt}(\gamma^+) = \operatorname{spt}(\gamma) \cup \{(\widehat{x}, \widehat{y})\} = \operatorname{spt}(\gamma).$$

Finally, concerning (6), the fact that $\mathfrak{C}(\gamma') \geq \mathfrak{C}(\gamma)$ is immediate by (4.4), so let us consider the other estimate. Of course $\mathfrak{C}(\bar{\gamma}') = \mathfrak{C}(\bar{\gamma}) + \mathfrak{C}(\bar{\gamma}^+) - \mathfrak{C}(\bar{\gamma}^-)$, and

$$\begin{aligned}
&\mathfrak{C}(\bar{\gamma}^+) \leq \alpha \Big(\max c\big(\widehat{x} \times \operatorname{spt}(\pi_2 \mu)\big) + \max c\big(\operatorname{spt}(\pi_1 \beta) \times \widehat{y}\big) \Big); \\
&\mathfrak{C}(\bar{\gamma}^-) \geq \alpha \Big(\min c\big(\operatorname{spt}(\mu)\big) + \min c\big(\operatorname{spt}(\beta)\big) \Big);
\end{aligned}$$

recall now that $\widehat{x} \in \pi_1(\mathrm{spt}\,(\nu)) \subseteq \pi_1(\mathrm{spt}\,(\nu_2)) \subseteq B_2 \subseteq B_1$, while $\mathrm{spt}\,(\pi_2\mu) \subseteq \mathrm{spt}\,(\pi_2\mu_2) \subseteq C_1$: therefore,

$$\max c\big(\widehat{x} \times \mathrm{spt}\,(\pi_2\mu)\big) \leq \max c(B_1 \times C_1),$$

and analogously

$$\max c\big(\mathrm{spt}\,(\pi_1\beta) \times \widehat{y})\big) \leq \max c(B_3 \times C_3);$$

on the other hand, by construction μ and β are concentrated in $B_1 \times C_1$ and $B_3 \times C_3$ respectively, so that

$$\min c\big(\mathrm{spt}\,(\mu)\big) \geq \min c(B_1 \times C_1), \qquad \min c\big(\mathrm{spt}\,(\beta)\big) \geq \min c(B_3 \times C_3).$$

Thanks to (4.1) and (4.2), we derive $\mathfrak{C}(\bar\gamma') \leq \mathfrak{C}(\bar\gamma) + 2\alpha\varepsilon$ and so also (6) is established and the proof is completed. $\square$

We are now ready to prove the main theorem.

Proof. of Theorem B. One implication is immediate from Theorem 1.2 in view of Lemma 4.1. To show the other one, as done in Theorem A we take a transport plan γ which is not optimal and we show that $\mathrm{spt}\,(\gamma)$ is not $c-$cyclical monotone. To do that, we take another transport plan $\bar\gamma$ with $\mathfrak{C}(\bar\gamma) < \mathfrak{C}(\gamma)$, thus in particular $\bar\gamma$ has a finite cost.

Our aim is to pass from $(\gamma, \bar\gamma)$ to $(\gamma', \bar\gamma')$ via $\varepsilon-$steps in such a way that γ' is purely atomic, $\mathfrak{C}(\gamma') > \mathfrak{C}(\bar\gamma')$ and $\mathrm{spt}\,(\gamma') \subseteq \mathrm{spt}\,(\gamma)$; thanks to Theorem A we will obtain that $\mathrm{spt}\,(\gamma')$ is not $c-$cyclical monotone and thus we will derive the same for $\mathrm{spt}\,(\gamma)$. To do that, we will make use of Lemma 4.6 and of Zorn's Lemma.

Let us define ε so that $\mathfrak{C}(\bar\gamma) + 2\varepsilon\|\gamma\| < \mathfrak{C}(\gamma)$, the set Δ as

$$\Delta := \Big\{(\eta, \bar\eta) \in \big(\mathcal{M}^+(X \times Y)\big)^2 : \ \pi_i\eta = \pi_i\bar\eta \text{ for } i = 1,\, 2, \\ \mathfrak{C}(\eta) \geq \mathfrak{C}(\gamma),\ \mathfrak{C}(\bar\eta) \leq \mathfrak{C}(\bar\gamma) + 2\varepsilon\|\eta_{\mathrm{AT}}\|,\ \mathrm{spt}\,(\eta) \subseteq \mathrm{spt}\,(\gamma)\Big\},$$

and the partial order $\leq_\Delta$ on Δ for which $(\eta_1, \bar\eta_1) \leq_\Delta (\eta_2, \bar\eta_2)$ whenever, for some positive measures $\eta^\pm$ and $\bar\eta^\pm$,

$$\begin{cases} \eta_2 = \eta_1 + \eta^+ - \eta^- \\ \bar\eta_2 = \bar\eta_1 + \bar\eta^+ - \bar\eta^- \end{cases}; \tag{P.O. 1}$$

$$\|\eta^+\| = \|\eta^-\| = \frac{\|\bar\eta^+\|}{2} = \frac{\|\bar\eta^-\|}{2}; \tag{P.O. 2}$$

$$\eta_{2\,\mathrm{AT}} = \eta_{1\,\mathrm{AT}} + \eta^+; \tag{P.O. 3}$$

$$\mathfrak{C}(\eta^+) \geq \mathfrak{C}(\eta^-). \tag{P.O. 4}$$

In order to use Zorn's Lemma, we claim that $(\Delta, \leq_\Delta)$ is inductive; to show that, let $(I, \leq_I)$ be a totally ordered set of indexes and $\Pi = \{(\eta_i, \bar\eta_i) : \, i \in I\}$ a chain. We first notice that, by (P.O. 3), $\{\eta_{i\mathrm{AT}}\}$ is an increasing (with respect to $\leq_I$) set of measures, and it is bounded in norm since all the elements of Π have the same norm by (P.O. 1) and (P.O. 2). This immediately implies that $\eta_{i\mathrm{AT}}$ converges in norm to some measure $\eta_{\infty\mathrm{AT}}$ as $i \to \sup I$; hence, we can find a sequence $\{i_j\}_{j\in\mathbb{N}} \subseteq I$ such that

$$\eta_{j\,\mathrm{AT}} \xrightarrow[j\to+\infty]{} \eta_{\infty\,\mathrm{AT}}$$

in norm, denoting for simplicity by j the element i_j of I. Since Π is a totally ordered chain, we can write

$$\begin{cases} \eta_{j+1} = \eta_j + \eta_j^+ - \eta_j^-\,, \\ \bar\eta_{j+1} = \bar\eta_j + \bar\eta_j^+ - \bar\eta_j^-\,, \end{cases}$$

and define $\alpha_j := \|\eta_j^+\|$. Since $\eta_{j+1\,\mathrm{AT}} = \eta_{j\,\mathrm{AT}} + \eta_j^+$, we deduce that $\sum\alpha_j < +\infty$: this implies that $j \to \sup I$ with the order $\leq_I$, and also that –using again (P.O. 2)–

$$\eta_j \xrightarrow[j\to+\infty]{} \eta_\infty\,, \qquad\qquad \bar\eta_j \xrightarrow[j\to+\infty]{} \bar\eta_\infty$$

in norm for some measures $(\eta_\infty, \bar\eta_\infty)$. We want to prove now that $(\eta_\infty,\bar\eta_\infty)$ belongs to Δ and that it is an upper bound of the chain Π. Since $\sum\alpha_j < +\infty$, we have

$$\eta_\infty = \eta_0 + \sum\eta_j^+ - \sum\eta_j^-\,, \qquad\qquad \bar\eta_\infty = \bar\eta_0 + \sum\bar\eta_j^+ - \sum\bar\eta_j^-\,; \tag{4.5}$$

let now $i \in I$, and let $j = i_j \in I$ be an index such that $i \leq_I j$ (this is possible because, as already pointed out, $j \to \sup I$ w.r.t. $\leq_I$). Since Π is a chain we have $\eta_j = \eta_i + \eta_i^+ - \eta_i^-$ and $\bar\eta_j = \bar\eta_i + \bar\eta_i^+ - \bar\eta_i^-$ according to the definition of $\leq_\Delta$; therefore, defining

$$\eta^+ = \eta_i^+ + \sum_{n=j}^{+\infty}\eta_n^+\,, \qquad \eta^- = \eta_i^- + \sum_{n=j}^{+\infty}\eta_n^-\,, \qquad \bar\eta^+ = \bar\eta_i^+ + \sum_{n=j}^{+\infty}\bar\eta_n^+\,, \qquad \bar\eta^- = \bar\eta_i^- + \sum_{n=j}^{+\infty}\bar\eta_n^-\,,$$

it is obvious, also recalling (4.5), that the properties (P.O. 1)–(P.O. 4) are satisfied for $(\eta_\infty,\bar\eta_\infty)$ and $(\eta_i,\bar\eta_i)$: thus, as soon as we will check that $(\eta_\infty,\bar\eta_\infty) \in \Delta$, we will have that $(\eta_i,\bar\eta_i) \leq_\Delta (\eta_\infty,\bar\eta_\infty)$, hence that $(\eta_\infty,\bar\eta_\infty)$ is an upper bound for Π, and so finally that Δ is inductive.

Let us now show that $(\eta_\infty,\bar\eta_\infty)$ belongs to Δ: since the measures η_j and $\bar\eta_j$ strongly converge to η_∞ and $\bar\eta_\infty$ and the pairs $(\eta_j,\bar\eta_j)$ belong to Δ, it is clear that the two projections of η_∞ and those of $\bar\eta_\infty$ are the same, as well as the fact that $\operatorname{spt}(\eta_\infty) \subseteq \operatorname{spt}(\gamma)$. Moreover, since c is continuous the map $\eta \mapsto \mathfrak{C}(\eta)$ is l.s.c., then

$$\mathfrak{C}(\bar\eta_\infty) \leq \liminf \mathfrak{C}(\bar\eta_j) \leq \mathfrak{C}(\bar\gamma) + 2\varepsilon \liminf \|\eta_{j\,\mathrm{AT}}\| = \mathfrak{C}(\bar\gamma) + 2\varepsilon\|\eta_{\infty\,\mathrm{AT}}\|.$$

To conclude, we need to check $\mathfrak{C}(\eta_\infty) \geq \mathfrak{C}(\gamma)$, which is not completely trivial since the map $\eta \mapsto \mathfrak{C}(\eta)$ is not u.s.c.. Let us then denote by $\mu_{\mathrm{NA}} = \mu - \mu_{\mathrm{AT}}$ the non-atomic part of any measure μ, and notice that

$$\eta_{\infty\,\mathrm{AT}} = \eta_{0\,\mathrm{AT}} + \sum\eta_j^+\,, \qquad\qquad \eta_{\infty\,\mathrm{NA}} = \eta_{0\,\mathrm{NA}} - \sum\eta_j^-\,,$$

where both the sums make sense since $\sum\alpha_j < +\infty$. There are now two possibilities: if $\sum\mathfrak{C}(\eta_j^+) = +\infty$, then $\mathfrak{C}(\eta_\infty) \geq \mathfrak{C}(\eta_{\infty\,\mathrm{AT}}) = +\infty$, then the inequality $\mathfrak{C}(\eta_\infty) \geq \mathfrak{C}(\gamma)$ is emptily true. Otherwise, if $\sum\mathfrak{C}(\eta_j^+) < +\infty$, first of all by (P.O. 4) we have $\sum\mathfrak{C}(\eta_j^-) \leq \sum\mathfrak{C}(\eta_j^+) < +\infty$, and then

$$\begin{aligned} \mathfrak{C}(\eta_\infty) &= \mathfrak{C}(\eta_{\infty\,\mathrm{AT}}) + \mathfrak{C}(\eta_{\infty\,\mathrm{NA}}) = \mathfrak{C}(\eta_{0\,\mathrm{AT}}) + \sum\mathfrak{C}(\eta_j^+) + \mathfrak{C}(\eta_{\infty\,\mathrm{NA}}) \\ &\geq \mathfrak{C}(\eta_{0\,\mathrm{AT}}) + \sum\mathfrak{C}(\eta_j^-) + \mathfrak{C}(\eta_{\infty\,\mathrm{NA}}) = \mathfrak{C}(\eta_{0\,\mathrm{AT}}) + \mathfrak{C}(\eta_{0\,\mathrm{NA}}) = \mathfrak{C}(\eta_0) \geq \mathfrak{C}(\gamma). \end{aligned}$$

In both cases, we have $\mathfrak{C}(\eta_\infty) \geq \mathfrak{C}(\gamma)$, and then $(\eta_\infty,\bar\eta_\infty) \in \Delta$, so finally we proved that $(\Delta,\leq_\Delta)$ is inductive.

By Zorn's Lemma, we can take a maximal element $(\gamma',\bar\gamma') \in \Delta$ such that $(\gamma,\bar\gamma) \leq_\Delta (\gamma',\bar\gamma')$; note that, by definition, γ' and $\bar\gamma'$ are two transport plans associated to the same marginals. One has that the measure γ' is purely atomic: indeed, otherwise one could use Lemma 4.6 to obtain an $\varepsilon-$step from $(\gamma',\bar\gamma')$, and this would immediately give a contradiction with the fact that $(\gamma', \bar\gamma')$ is maximal with respect to $\leq_\Delta$. Now, define $\tilde f^+$ and $\tilde f^-$ the marginals of γ': since γ' is purely atomic it is clear that also $\tilde f^\pm$ are purely atomic; moreover, by definition of ε and recalling that $\|\gamma'\| = \|\gamma\|$ since $(\gamma, \bar\gamma) \leq_\Delta (\gamma', \bar\gamma')$, we have

$$\mathfrak{C}(\bar\gamma') \leq \mathfrak{C}(\bar\gamma) + 2\varepsilon\|\gamma'\| = \mathfrak{C}(\bar\gamma) + 2\varepsilon\|\gamma\| < \mathfrak{C}(\gamma) \leq \mathfrak{C}(\gamma').$$

Applying then Theorem A, we obtain that $\operatorname{spt}(\gamma')$ is not $c-$cyclical monotone. A fortiori, the bigger set $\operatorname{spt}(\gamma) \supseteq \operatorname{spt}(\gamma')$ is not $c-$cyclical monotone, and the proof is completed. $\square$

Remark 4.7. *In the proof of the Theorem above, to reach the suitable final measures $(\gamma', \bar\gamma')$ one could argue in a different way, more similar to the strategy of Theorem A, instead of using Zorn's Lemma. In fact, one can start from $(\gamma, \bar\gamma)$, make a sequence $(\gamma_n, \bar\gamma_n)$ of $\varepsilon-$steps of almost maximal power, and reach the final measures $(\gamma', \bar\gamma')$. But showing that γ' is purely atomic is not so easy as it was showing (3.12) in Theorem A. Indeed, if γ' is not purely atomic one can find by Lemma 4.6 the $\varepsilon-$step*

$$\gamma'' = \gamma' + \gamma^+ - \gamma^-, \qquad \bar\gamma'' = \bar\gamma' + \bar\gamma^+ - \bar\gamma^- \tag{4.6}$$

of power $\alpha > 0$. Now, it would be easy to conclude as in Theorem A if $\gamma^\pm$ and $\bar\gamma^\pm$ define an admissible $\varepsilon-$step also for $(\gamma_n, \bar\gamma_n)$, since this would give a contradiction with the maximality of the power in the recursive construction; but this cannot be done, since it is not true that $\bar\gamma^- \leq \bar\gamma_n$ (while the inequality $\gamma^- \leq \gamma_n$ holds). Therefore, to conclude one must follow again the construction of Lemma 4.6 proving that, using the $\varepsilon-$step of power α found in (4.6), there are $\varepsilon-$steps of power $\alpha - o(1/n)$ suitable for $(\gamma_n, \bar\gamma_n)$, so reaching anyway the desired contradiction.

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

[3] 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.

[4] P. Bernard & B. Buffoni, Optimal mass transportation and Mather theory, to appear on Journal of the European Mathematical Society.

[5] L.C. Evans, Partial Differential Equations and Monge–Kantorovich Mass Transfer, Current Developments in Mathematics (1997), 65–126.

[6] D. Feyel & A.S. Üstünel, Transport of measures on Wiener space and the Girsanov theorem, Comptes Rendus Mathématiques, **334**, no. 1 (2002), 1025–1028.

[7] D. Feyel & A.S. Üstünel, Monge–Kantorovich Measure Transportation and Monge–Ampère Equation on Wiener Space, Probab. Theory Relat. Fields, **128** (2004), 347–385.

[8] W. Gangbo & R.J. McCann, The geometry of optimal transportation, Acta Math., **177** (1996), 113–161.

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

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

[11] M. Knott & C. Smith, On Hoeffding–Frechet bounds and cyclice monotone relations, J. Multivariate Anal., **40** (1992), 328–334.

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

[13] 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/ .

[14] A. Pratelli, Equivalence between some definitions for the optimal mass transport problem and for the transport density on manifolds, Ann. Mat. Pura Appl., **184** (2005), no. 2, 215–238.

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

[16] T. Rockafellar, Characterization of the subdifferential of convex functions, Pacific Journal of Math **17** (1966), 497–510.

[17] L. Rüschendorf, On $c-$optimal random variables, Statistics & Probability Letters, **27** (1996), 267–270.

[18] W. Schachermayer & J. Teichmann, Solution of a problem in Villani's book, preprint (2005).

[19] C. Villani, Topics in optimal transportation, Graduate Studies in Mathematics, Vol. **58**, American Mathematical Society, Providence, RI (2003).

Dipartimento di Matematica, via Ferrata 1, 27100 Pavia
E-mail address: `aldo.pratelli@unipv.it`