



**HAL**  
open science

# On the continuity of the total cost in the mass transport problem with relativistic cost functions

Jean Louet, Aldo Pratelli, Florian Zeisler

## ► To cite this version:

Jean Louet, Aldo Pratelli, Florian Zeisler. On the continuity of the total cost in the mass transport problem with relativistic cost functions. 2016. hal-01419526

**HAL Id: hal-01419526**

**<https://hal.science/hal-01419526>**

Preprint submitted on 19 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ON THE CONTINUITY OF THE TOTAL COST IN THE MASS TRANSPORT PROBLEM WITH RELATIVISTIC COST FUNCTIONS

JEAN LOUET, ALDO PRATELLI, AND FLORIAN ZEISLER

ABSTRACT. In this paper we consider the mass transport problem in the case of a relativistic cost; we can establish the continuity of the total cost, together with a general estimate about the directions in which the mass can actually move, under mild assumptions. These results generalize those of the recent paper [2], also positively answering some of the open questions there.

## 1. INTRODUCTION

In this paper we concentrate on a particular question in the mass transport problem. The general mass transport problem in  $\mathbb{R}^d$ , which is now widely known (for a wide source on that, refer on the book [7]), consists in considering two probability measures  $\mu \neq \nu$  in  $\mathbb{R}^d$ , and trying to minimize the cost of the transport plans between  $\mu$  and  $\nu$ . More precisely, the set  $\Pi(\mu, \nu)$  of the transport plans is given by all the measures  $\gamma$  on  $\mathbb{R}^d \times \mathbb{R}^d$  whose marginals on the two copies of  $\mathbb{R}^d$  are  $\mu$  and  $\nu$  respectively. And the cost of a transport plan  $\gamma$  is given by

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y),$$

where  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$  is some given l.s.c. function, called *cost function*. It is immediate to show that the set of transport plans is never empty, and in particular there exist always minimizers of the cost, which are called *optimal transport plans*.

A quite interesting example of a cost function is the so-called *relativistic heat cost*, first introduced by Brenier in [4], and which is defined as  $c(x, y) = h(y - x)$ , where

$$h(z) = \begin{cases} 1 - \sqrt{1 - |z|^2} & |z| \leq 1, \\ +\infty & |z| > 1. \end{cases} \quad (1.1)$$

Notice that the function  $h$  is strictly convex in a strictly convex subset of  $\mathbb{R}^d$  (the closed unit ball), and  $+\infty$  outside. The study of the transport with relativistic heat cost was again studied in [5], and generalized to the case of *relativistic cost functions* in [3]. The relativistic cost functions are defined again as  $c(x, y) = h(y - x)$ , but this time  $h$  is a generic function which is strictly convex and bounded in a strictly convex and bounded subset  $\mathcal{C}$  of  $\mathbb{R}^d$ , and  $+\infty$  outside, see Definition 2.1. Actually, when speaking about the transport problem with a relativistic cost, there is an additional parameter  $t > 0$ , corresponding to the time; more precisely, the relativistic cost functions are the functions  $c_t$ , for every  $t > 0$ , which are defined as

$$c_t(x, y) = h\left(\frac{y - x}{t}\right).$$

Notice that, if  $t$  is very small, then  $c_t(x, y) = +\infty$  for all the points  $x, y$  which are not extremely close to each other, so it is easy to guess that all the transport plans have infinite cost: this corresponds to the fact that, in a relativistic context, particles cannot move faster than a given maximal velocity, hence in a very short time it is simply impossible to transport the density  $\mu$  onto  $\nu$ ; instead, for  $t$  bigger and bigger,

not only it is possible to transform  $\mu$  onto  $\nu$ , but it becomes also cheaper and cheaper. In particular, the following results were proved in [3]; here, and through the whole paper, by  $\mathcal{C}_t : \Pi(\mu, \nu) \rightarrow [0, \infty]$  we denote the *cost relative to the time  $t$* , and by  $\mathcal{C} : (0, +\infty) \rightarrow [0, +\infty]$  the *minimal cost* corresponding to the time  $t$ , which are given by

$$\mathcal{C}_t(\gamma) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} c_t(x, y) d\gamma(x, y), \quad \mathcal{C}(t) = \min \left\{ \mathcal{C}_t(\gamma) : \gamma \in \Pi(\mu, \nu) \right\}.$$

**Theorem 1.1.** *Let  $\mu \neq \nu$  be two probability measures with compact support. Then there exists a time  $T > 0$ , called critical time, such that  $\mathcal{C}(t) = +\infty$  for every  $0 < t < T$ , while  $\mathcal{C}(t)$  is bounded for every  $t \geq T$ ; and moreover, the function  $\mathcal{C}$  is non-increasing and right-continuous on the interval  $[T, +\infty)$ . Finally, there exists a unique optimal transport plan  $\gamma_t$  for every  $t > T$ .*

The study of the transport problem in a relativistic context has been then continued in the very recent paper [2]. There, the authors have introduced a subclass of the relativistic cost functions, namely, the *highly relativistic cost functions*, see Definition 2.6. Basically, a relativistic cost is called “highly relativistic” if the slope of  $h$  explodes on the boundary  $\partial\mathcal{C}$  of the convex set where  $h < \infty$ . The reason to introduce this subclass is simple: observe that the original relativistic heat cost defined in (1.1) is highly relativistic; one can also notice that some of the nice properties which hold in this model case actually depend on the infinite slope of  $h$  at the boundary. In fact, in the paper [2] the following results were proved.

**Theorem 1.2.** *Let  $\mu \neq \nu$  be two probability measures with compact support in  $\mathbb{R}^d$ , and assume that  $c_t$  is a highly relativistic cost function, and that  $\mu \ll \mathcal{L}$ , being  $\mathcal{L}$  the Lebesgue measure on  $\mathbb{R}^d$ . Then,*

- (i) *If  $\mu \in L^\infty$ , then the function  $t \mapsto \mathcal{C}(t)$  is continuous on  $[T, +\infty)$ .*
- (ii) *For every supercritical time  $t > T$ , the optimal plan  $\gamma_t$  satisfies*

$$\gamma_t \left( \left\{ (x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \frac{y-x}{t} \in \partial\mathcal{C} \right\} \right) = 0.$$

Let us now briefly discuss the above result, and then its assumptions. The first claim simply states the continuity of the total cost; the second one is basically saying that the optimal transport does not move the points “with maximal distance”, that is, the vectors lying in the boundary of  $\mathcal{C}$  are almost never used: notice that the cost, on the boundary of  $\mathcal{C}$ , is still bounded, so it is not obvious that it should not be convenient to use also such vectors.

Let us now pass to discuss the assumptions of the two claims in Theorem 1.2: while the fact that  $\mu$  is bounded for the first claim is only a technical assumption, just helping to simplify the notations, the “serious” assumptions are that  $\mu$  must be absolutely continuous with respect to the Lebesgue measure, and that the cost is not just relativistic, but also highly relativistic. The counterexamples in [2] ensure that both the claims are false without the absolute continuity, as well as the second claim without the assumption that  $c_t$  is highly relativistic. Instead, the question whether the highly relativistic assumption is essential also for (i) was set as an open question at the end of that paper.

The aim of this paper is to generalize Theorem 1.2 as much as possible. In particular, we will show that for the relativistic cost functions (not necessarily highly relativistic, then) the first claim is still valid (even removing the boundedness assumption for  $\mu$ ), as well as a generalisation of the second one, which basically says that an optimal transport plan does not use those vectors in the boundary  $\mathcal{C}$  where  $h$  has infinite slope (which form the whole  $\partial\mathcal{C}$  if the cost is highly relativistic). More precisely, our main result reads as follows.

**Theorem A.** *Let  $\mu \neq \nu$  be two probability measures with compact support in  $\mathbb{R}^d$ , and assume that  $c_t$  is a relativistic cost function, and that  $\mu \ll \mathcal{L}$ . Then,*

- (i) *The function  $t \mapsto \mathcal{C}(t)$  is continuous on  $[T, +\infty)$ .*
- (ii) *For every supercritical time  $t > T$ , the optimal plan  $\gamma_t$  satisfies  $\gamma_t(\{(x, y) : \frac{y-x}{t} \in \Theta\}) = 0$ , where*

$$\Theta := \left\{ v \in \partial\mathcal{C} : D_{-v}h(v) = -\infty \right\},$$

*and  $D_{-v}h(v) \in [-\infty, +\infty)$  is the slope of  $h$  at the point  $v$  in the direction  $-v$ .*

Our constructions are reminiscent of those made in [2], but there are some fundamental differences and new ideas, which are necessary in order to deal with the quite weaker assumption of Theorem A with respect to Theorem 1.2, in particular with the fact of considering relativistic, but not highly relativistic cost functions.

The plan of the paper is simple: in Section 2 we list all the relevant definitions, notations, and the known technical facts, among which the so-called Chain Lemma (Lemma 2.7). Then, in the Sections 3 and 4 we prove the two parts of our Theorem A.

## 2. NOTATION AND PRELIMINARY RESULTS

In this section we collect some standard notation about the mass transport problem, and we give the relevant definitions about relativistic and highly relativistic cost functions.

Let  $X$  and  $Y$  be two Polish spaces (through the paper, we will only be interested in the case  $X = Y = \mathbb{R}^d$ ), and let  $\mu$  and  $\nu$  be two probabilities on  $X$  and  $Y$  respectively. A probability measure  $\gamma$  on  $X \times Y$  is called a *transport plan* if its two marginals on  $X$  and  $Y$  coincide with  $\mu$  and  $\nu$ ; the collection of the transport plans is denoted by  $\Pi(\mu, \nu)$ . Given a l.s.c. function  $c : X \times Y \rightarrow [0, +\infty]$ , the *cost* of the plan  $\gamma$  is given by

$$\iint_{X \times Y} c(x, y) d\gamma(x, y),$$

and  $\gamma$  is called an *optimal transport plan* if it minimizes the cost among elements of  $\Pi(\mu, \nu)$ . A particular case of cost functions, namely, the relativistic ones, is now introduced.

**Definition 2.1.** *Let  $\mathcal{C}$  be a closed, bounded, convex set in  $\mathbb{R}^d$ , containing the origin in its interior, and let  $h : \mathbb{R}^d \rightarrow [0, +\infty]$  be a function which is strictly convex and bounded on  $\mathcal{C}$ , constantly  $+\infty$  on  $\mathbb{R}^d \setminus \mathcal{C}$ , and such that  $h(0) = 0$ . Then, for any  $t > 0$  we define the function  $c_t : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$  as*

$$c_t(x, y) = h\left(\frac{y-x}{t}\right).$$

*Such functions  $c_t$  are called relativistic cost functions, and for any  $t > 0$  we denote by  $\mathcal{C}_t(\gamma)$  the cost of a plan  $\gamma \in \Pi(\mu, \nu)$  with respect to  $c_t$ , and by  $\mathcal{C}(t)$  the minimum of these costs. For simplicity of notations, we will denote by  $\|h\|_{L^\infty}$  the maximum of  $h$  on  $\mathcal{C}$ .*

A useful concept in mass transportation is the composition of plans, which can be simply defined thanks to a disintegration of the plans (for the definition of disintegration of measures, one can refer for instance to [1]).

**Definition 2.2** (Composition of transport plans). *Let  $\mu$ ,  $\alpha$  and  $\nu$  be three probability measures on the Polish spaces  $X$ ,  $Y$  and  $Z$  respectively, and  $\gamma_1 \in \Pi(\mu, \alpha)$  and  $\gamma_2 \in \Pi(\alpha, \nu)$  be two transport plans. Let*

us disintegrate  $\gamma_1$  and  $\gamma_2$  with respect to the projection on  $Y$ , so writing  $\gamma_1 = \gamma_1^y \otimes \alpha$  and  $\gamma_2 = \alpha \otimes \gamma_2^y$ . Then, the composition of  $\gamma_1$  and  $\gamma_2$  is the transport plan  $\gamma_2 \circ \gamma_1 \in \Pi(\mu, \nu)$ , defined as

$$\iint_{X \times Z} \varphi(x, z) d\gamma_2 \circ \gamma_1(x, z) := \int_Y \left( \iint_{X \times Z} \varphi(x, z) d\gamma_1^y(x) d\gamma_2^y(z) \right) d\alpha(y)$$

for every  $\varphi \in C_b(X \times Z)$ . It is immediate to check that  $\gamma_2 \circ \gamma_1$  is a transport plan with marginals  $\mu$  and  $\nu$ , as well as that  $(x, z) \in \text{spt}(\gamma_2 \circ \gamma_1)$  if and only if there exists some  $y \in Y$  such that  $(x, y) \in \text{spt} \gamma_1$  and  $(y, z) \in \text{spt} \gamma_2$ .

We present here two elementary results about convex sets and convex functions.

**Lemma 2.3.** *Let  $c_t$  be a relativistic cost function, let  $A$  be the minimum of  $|v|$  for vectors  $v \in \partial\mathcal{C}$ , and let  $\bar{\delta}$  and  $r^+$  be positive numbers, both much smaller than  $A$ . Then, there exists an increasing function  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , with  $\lim_{\eta \rightarrow 0} \phi(\eta) = 0$ , such that*

$$\left| h(v) - h\left(\frac{v - \eta w}{t}\right) \right| < \phi(\eta)$$

holds for every  $v, w \in \mathbb{R}^d$  and  $t, \eta > 0$  satisfying

$$v \in \mathcal{C} \setminus (1 - \bar{\delta})\mathcal{C}, \quad |w - v| \leq r^+ \sqrt{d}, \quad |1 - t| < \frac{\eta}{2}.$$

*Proof.* Since the function  $h$  is uniformly continuous on  $\mathcal{C}$ , we only have to show that, for  $v, w$  and  $t$  as in the claim, one has

$$\frac{v - \eta w}{t} \in \mathcal{C},$$

and this is in turn a trivial geometric property, since  $\mathcal{C}$  is a bounded, convex set.  $\square$

Let us now give the following definition of “directional derivative” for a relativistic cost.

**Definition 2.4.** *Let  $\mathcal{C}$  and  $h$  be as in Definition 2.1, let  $P \in \mathcal{C}$ , and let  $v \in \mathbb{S}^{d-1}$  be an internal direction, that is, a vector such that  $P + \varepsilon v$  belongs to the interior of  $\mathcal{C}$  for every  $0 < \varepsilon \ll 1$ . We define then directional derivative of  $h$  at  $P$  in the direction  $v$  the number*

$$D_v h(P) := \lim_{\varepsilon \searrow 0} \frac{h(P + \varepsilon v) - h(P)}{\varepsilon}.$$

Notice that  $D_v h(P)$  is the right derivative at 0 of a real one-dimensional convex function defined in a right neighborhood of 0, hence it always exists and it belongs to  $[-\infty, +\infty)$ . Notice also that the set  $\mathbb{S}_P$  of the internal directions at  $P$  is the whole  $\mathbb{S}^{d-1}$  if  $P$  belongs to the interior of  $\mathcal{C}$ , while for  $P \in \partial\mathcal{C}$  it is an open subset of  $\mathbb{S}^{d-1}$ , in particular an open half-sphere if  $\partial\mathcal{C}$  admits a normal vector at  $P$ .

We can now observe that the directional derivatives are either all finite or all infinite.

**Lemma 2.5.** *Let  $\mathcal{C}$  and  $h$  be as in Definition 2.1, and  $P$  and  $\mathbb{S}_P$  as in the Definition 2.4. Then, the function  $v \mapsto D_v h(P) \in [-\infty, +\infty)$ , defined on  $\mathbb{S}_P$ , is continuous. In addition, it is either real on the whole  $\mathbb{S}_P$ , or constantly  $-\infty$ .*

*Proof.* The only interesting case is when  $P \in \partial\mathcal{C}$ , since otherwise everything is trivial by the convexity of  $h$  and  $\mathcal{C}$ , and in particular  $D_v h(P)$  is always real. Hence, we suppose from now on that  $P \in \partial\mathcal{C}$ ; let us first prove that the map  $v \mapsto D_v$  is either always real or constantly  $-\infty$ , and then the continuity.

Step I. *The realness (or constant infiniteness) of the directional derivatives.*

Let us start by taking some internal direction  $v \in \mathbb{S}_P$ , and let us assume for a moment that  $D_v h(P) > -\infty$ , so  $D_v h(P) \in \mathbb{R}$ . Let now  $\omega$  be another internal direction; an immediate geometrical consideration ensures

that, if  $\omega$  is close enough to  $v$ , then there exists some positive constant  $\ell > 0$  with the following property. For every  $0 < \delta \ll 1$ , if we define

$$Q = P + \delta v, \quad R = P + \frac{\delta}{2} \omega, \quad d = |Q - R| \leq \frac{3}{2} \delta \quad (2.1)$$

and we let  $S$  be the point on the half-line starting at  $R$  and passing through  $Q$  having distance  $d + \ell$  from  $R$ , then  $S \in \mathcal{C}$ . Notice carefully that a possible value of  $\ell$ , as well as of the necessary closeness between  $\omega$  and  $v$ , can be obtained independently on  $v$ , and only depending on the exact form of  $\mathcal{C}$ , as well as on the distance between  $v$  and the boundary of  $\mathbb{S}_P$ .

We can now write  $Q$  as a convex combination of  $R$  and  $S$ , in fact by construction one has

$$Q = \frac{\ell}{\ell + d} R + \frac{d}{\ell + d} S. \quad (2.2)$$

As a consequence, the convexity of  $h$  ensure

$$h(P) + \delta D_v h(P) \leq h(Q) \leq \frac{\ell}{\ell + d} h(R) + \frac{d}{\ell + d} h(S) \leq h(R) + \frac{d}{\ell + d} \|h\|_{L^\infty} \leq h(R) + \frac{3\delta}{2\ell} \|h\|_{L^\infty},$$

from which we get

$$h\left(P + \frac{\delta}{2} \omega\right) - h(P) \geq \delta \left( D_v h(P) - \frac{3}{2\ell} \|h\|_{L^\infty} \right)$$

and hence, sending  $\delta \searrow 0$ ,

$$D_\omega h(P) \geq 2D_v h(P) - \frac{3}{\ell} \|h\|_{L^\infty}. \quad (2.3)$$

This estimate immediately ensures that the subset of  $\mathbb{S}_P$  made by the directions along which the directional derivative is not  $-\infty$  form an open subset of  $\mathbb{S}_P$ . But actually, since, as underlined above, the amplitude of the neighborhood of  $v$  in which the estimate (2.3) holds only depends on the distance of  $v$  from the boundary of  $\mathbb{S}_P$ , then we directly get that, if at some direction  $v \in \mathbb{S}_P$  one has  $D_v h(P) > -\infty$ , then the same inequality holds for all the directions  $v \in \mathbb{S}_P$ .

Summarizing, we have proved that  $D_v h(P)$  is never  $+\infty$ , and actually it is either real for every  $v \in \mathbb{S}_P$ , or equal to  $-\infty$  for every  $v \in \mathbb{S}_P$ . This concludes the first step of the proof.

Step II. The continuity.

In this step we prove the continuity, which does not come from (2.3). Thanks to Step I, we only have to consider the case when  $D_v h(P)$  is real for every  $v \in \mathbb{S}_P$ .

Let us then take a direction  $v \in \mathbb{S}_P$ , select some  $0 < \varepsilon \ll 1$ , and take some  $\ell > 0$  such that  $V = P + \ell v$  belongs to the interior of  $\mathcal{C}$  and satisfies  $h(V) \leq h(P) + \ell(D_v h(P) + \varepsilon)$ . Let now  $\omega \in \mathbb{S}_P$  be another direction, and for  $\delta \ll \ell$  define again  $Q$  and  $R$  as in (2.1), and  $S$  as in the following line. The point  $S$  is arbitrarily close to  $V$  (thus also in the interior of  $\mathcal{C}$ ) as soon as  $\delta$  is small enough and  $\omega$  is close enough to  $v$ , hence by the continuity of  $h$  we can assume

$$h(S) \leq h(P) + \ell(D_v h(P) + 2\varepsilon).$$

Notice that, differently with what happened in Step I, this time the value of  $\ell$  and the needed closeness of  $\omega$  to  $v$  really depend on  $v$  (this is why the argument of the present step could not prove the realness that, instead, we got in Step I). Formula (2.2) is still valid, with  $d = |Q - R|$  being this time arbitrarily close to  $\delta/2$ , again up to select  $\omega$  close enough to  $v$ . Hence we can evaluate, similarly with what we have done in Step I,

$$\begin{aligned} h(P) + \delta D_v h(P) &\leq h(Q) \leq \frac{\ell}{\ell + d} h(R) + \frac{d}{\ell + d} h(S) \\ &\leq \left(1 - \frac{d}{\ell + d}\right) h\left(P + \frac{\delta}{2} \omega\right) + \frac{d}{\ell + d} h(P) + \frac{d\ell}{\ell + d} (D_v h(P) + 2\varepsilon), \end{aligned}$$

from which it readily follows

$$\lim_{\delta \searrow 0} \frac{h(P + \frac{\delta}{2}\omega) - h(P)}{\frac{\delta}{2}} \geq D_v h(P) - 2\varepsilon. \quad (2.4)$$

The opposite inequality, namely,

$$\lim_{\delta \searrow 0} \frac{h(P + \frac{\delta}{2}\omega) - h(P)}{\frac{\delta}{2}} \leq D_v h(P) + 2\varepsilon, \quad (2.5)$$

can be obtained in the very same way, just exchanging the role of  $Q$  and  $R$ ; more precisely, we define again  $Q = P + \delta v$ , but this time  $R = P + 2\delta\omega$ , and this time  $S$  has distance  $\ell$  from  $R$ , on the half-line starting at  $Q$  and passing through  $R$ . Hence, this time we can write  $R$  as a convex combination of  $Q$  and  $S$ , and then the very same calculation which brought to (2.4) now bring to (2.5). Since  $\varepsilon$  is arbitrary, the continuity is then established and the proof is concluded.  $\square$

Thanks to the above result, it is now very simple to introduce the ‘‘highly relativistic’’ cost functions.

**Definition 2.6.** *Let  $h$  be as in Definition 2.1. The functions  $c_t$  are called highly relativistic cost functions if  $D_v h(P)$  is constantly  $-\infty$  on  $\mathbb{S}_P$  for every  $P \in \partial\mathcal{C}$ .*

To conclude this section, we give the claim of the Chain Lemma (for a proof, see [2, Proposition 2.11]).

**Lemma 2.7** (Chain Lemma). *Let  $\gamma, \gamma' \in \Pi(\mu, \nu)$ ,  $0 \neq \gamma_0 \leq \gamma$ , and set  $\mu_0 = \pi_1 \gamma_0$ ,  $\nu_0 = \pi_2 \gamma_0$ . Then, there exist  $N \in \mathbb{N}$  and  $\bar{\varepsilon} > 0$  such that, for every  $\varepsilon \leq \bar{\varepsilon}$ , there are plans  $\tilde{\gamma} \leq \gamma$  and  $\tilde{\gamma}' \leq \gamma'$  satisfying*

$$\begin{aligned} \pi_1 \tilde{\gamma} &= \pi_1 \tilde{\gamma}' = \tilde{\mu} + \mu_A, & \pi_2 \tilde{\gamma} &= \tilde{\nu} + \nu_A, & \pi_2 \tilde{\gamma}' &= \tilde{\nu} + \nu_B, \\ \mu_A &\leq \mu_0, \quad \tilde{\mu} \leq \mu - \mu_0, & \nu_A, \nu_B &\leq \nu_0, & \tilde{\nu} \wedge \nu_0 &= 0, \\ \|\tilde{\gamma}\| &= \|\tilde{\gamma}'\| = (N+1)\varepsilon, & \|\tilde{\mu}\| &= \|\tilde{\nu}\| = N\varepsilon, & \|\mu_A\| &= \|\nu_A\| = \|\nu_B\| = \varepsilon. \end{aligned}$$

In particular,  $\tilde{\gamma}$  can be decomposed as  $\tilde{\gamma} = \tilde{\gamma}_0 + \tilde{\gamma}_\infty$ , where

$$\tilde{\gamma}_0 \leq \gamma_0, \quad \tilde{\gamma}_\infty \leq \gamma - \gamma_0, \quad \pi_1 \tilde{\gamma}_0 = \mu_A, \quad \pi_2 \tilde{\gamma}_0 = \nu_A, \quad \pi_1 \tilde{\gamma}_\infty = \tilde{\mu}, \quad \pi_2 \tilde{\gamma}_\infty = \tilde{\nu}.$$

### 3. CONTINUITY OF THE TOTAL COST

This section is devoted to prove the continuity of the total cost, that is, part (i) of Theorem A.

*Proof of Theorem A, part (i).* Since the right continuity of the function  $t \mapsto \mathcal{C}(t)$  is obvious (and it was also proved in the literature, see Theorem 1.1), we only have to deal with the left continuity. Up to rescaling, we can assume that  $T < 1$  and we aim to prove the left continuity at  $t = 1$ . Let us fix any  $T < t' < 1$ , and let us call  $\gamma$  and  $\gamma'$  the optimal transport plans corresponding to  $t = 1$  and  $t = t'$ . Moreover, let us define

$$A := \min \{|v| : v \in \partial\mathcal{C}\},$$

let us fix two small constants  $r^+ \ll \bar{\delta} \ll 1 - t'$  such that

$$r^+ \sqrt{\bar{d}} \ll \bar{\delta} A, \quad (3.1)$$

and let  $r$  be a third positive constant, much smaller than  $r^+$ . Define then the set

$$S_{\bar{\delta}} := \left\{ (x, y) \in \mathbb{R}^d \times \mathbb{R}^d : y - x \in (1 - \bar{\delta}) \mathcal{C} \right\}$$

and the measure

$$\gamma_0 := \gamma \mathbf{L} \left( \mathbb{R}^d \times \mathbb{R}^d \setminus S_{\bar{\delta}} \right).$$

Notice that  $\gamma_0$  is the part of the transport made by the points which “do not move too much”, that is,  $y - x$  is not too close to the boundary of  $\mathcal{C}$ ; as a consequence, it is immediate to observe that  $\mathcal{C}_t(\gamma_0)$  converges to  $\mathcal{C}_1(\gamma_0)$  for  $t \nearrow 1$ . Instead,  $\gamma - \gamma_0$  is the part of the transport with which we have to deal carefully. Notice also that we do not even know that  $\|\gamma - \gamma_0\|$  is small if  $\bar{\delta}$  is small enough: this would be the case only for highly relativistic transport costs. In order to prove the claim, we will build a transport plan  $\xi \in \Pi(\mu, \nu)$  whose cost satisfies

$$\limsup_{t \nearrow 1} \mathcal{C}_t(\xi) \leq \mathcal{C}_1(\gamma) + 2\phi\left(\frac{2r}{A}\right), \quad (3.2)$$

where  $\phi$  is the function given by Lemma 2.3. Notice that the function  $\phi$  depends on  $\bar{\delta}$  and on  $r^+$ , but not on  $A$  (which is a geometrical constant, only depending on  $\mathcal{C}$ ), neither on  $r$  (which is an arbitrarily small constant, in turn depending on  $\bar{\delta}$  and  $r^+$ ). Since the function  $\phi$  is infinitesimal for  $r \searrow 0$ , the searched left continuity of  $t \mapsto \mathcal{C}(t)$  at  $t = 1$  will be established once we prove (3.2). Unfortunately, the trivial choice  $\xi = \gamma$  does not work, one could even have  $\mathcal{C}_t(\gamma) = +\infty$  for every  $t < 1$ . For the sake of clarity, we divide our construction in some steps.

*Step I. Definition of the cubes  $Q_i^1$  and  $Q_i^2$ , and of the measures  $\gamma_i$ ,  $\mu_i$  and  $\nu_i$ , and  $\gamma_i^K$ ,  $\mu_i^K$  and  $\nu_i^K$ .*

We start by covering the support of  $\gamma - \gamma_0$  with finitely many cubes of side  $r$  in  $\mathbb{R}^d \times \mathbb{R}^d$ , and we call  $M$  their number. More precisely, for every  $1 \leq i \leq M$  we select two cubes  $Q_i^1$  and  $Q_i^2$  of side  $r$  in  $\mathbb{R}^d$ , and a positive measure  $\gamma_i$  concentrated in  $Q_i^1 \times Q_i^2$ , in such a way that

- $\gamma - \gamma_0 = \sum_{i=1}^M \gamma_i$ ;
- $\gamma_i \wedge \gamma_j = 0$  for every  $i \neq j$ ;
- for every  $i$ , the cubes  $Q_i^1$  and  $Q_i^2$  are parallel; more precisely,  $Q_i^2 = Q_i^1 + v_i$ , and the vector  $v_i$  is parallel to one of the sides of  $Q_i^1$  (hence, also of  $Q_i^2$ ).

Notice that the existence of such cubes and measures is obvious (in fact, we do not require the products  $Q_i^1 \times Q_i^2$  to be disjoint, but only the measures  $\gamma_i$  to be mutually singular). We will call  $\mu_i$  and  $\nu_i$  the two marginals of  $\gamma_i$  and, up to remove useless cubes, we will assume that

$$m_i := \|\gamma_i\| = \|\mu_i\| = \|\nu_i\| > 0$$

for every  $i$ . We apply now the Chain Lemma 2.7 to the measures  $\gamma$ ,  $\gamma'$  and  $\gamma_i$ , obtaining some constants  $N_i$  and  $\bar{\varepsilon}_i$ , and we call  $N = \max_i N_i$  and  $\bar{\varepsilon} = \min_i \bar{\varepsilon}_i$ . Then, we fix an arbitrarily small constant  $\varepsilon \ll \min_i \{m_i\} \bar{\varepsilon} / MN$ , also satisfying

$$\varepsilon N \ll \phi\left(\frac{2r}{A}\right), \quad (3.3)$$

and we find  $K \gg 1$  such that for every  $i$  one has

$$\mu_i\left(\{x : \rho(x) \geq K\}\right) \leq \varepsilon m_i, \quad (3.4)$$

being  $\rho$  the density of  $\mu$  with respect to the Lebesgue measure  $\mathcal{L}$ . Then, we call

$$\gamma_i^K = \gamma_i \llcorner \left\{ (x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \rho(x) \leq K \right\},$$

and we let  $\mu_i^K$  and  $\nu_i^K$  be its two marginals.

*Step II. Definition of the transport plan  $\xi_i^1$  and its cost.*

In this step, we provide the first part of the “competitor” transport plan  $\xi$ , namely, a transport plan  $\xi_i^1$  for every  $1 \leq i \leq M$ . Since this will be done for each  $i$  independently, we concentrate ourselves, only within this step, on a given  $1 \leq i \leq M$ . For further clarity, this step will be further subdivided in two substeps. Let us consider the cubes  $Q_i^1$  and  $Q_i^2$ . For simplicity of notations, and up to a rotation and



translation, we can assume that  $Q_i^1 = [0, r]^d$ , while  $Q_i^2 = [a, a+r] \times [0, r]^{d-1}$ . Notice that, for the ease of notation, we call “ $a$ ”, and not “ $a_i$ ”, the distance between the cubes; more in general, since in this step we only work with a fixed index  $i$ , we will not put a subscript “ $i$ ” to all the quantities which are used only inside this step. Keep in mind that the measure  $\gamma_i$  is concentrated in  $\mathbb{R}^d \times \mathbb{R}^d \setminus S_{\bar{\delta}}$ , hence by (3.1) and the fact that  $a + r\sqrt{d} > (1 - \bar{\delta})A$ , we have  $r\sqrt{d} \ll a$ .

*Step IIa. Definition of the relevant sets and of the plan  $\xi_i^1$ .*

In this first substep, we give the definition of the plan  $\xi_i^1$ . Let us use the notation  $x = (\sigma, \tau) \in \mathbb{R} \times \mathbb{R}^{d-1}$  for points in  $\mathbb{R}^d$ , denote by  $\pi_\tau : \mathbb{R}^d \rightarrow \mathbb{R}^{d-1}$  the projection on the variable  $\tau$ , and disintegrate the measure  $\mu_i^K$  with respect to  $\pi_\tau$ , obtaining the decomposition

$$\mu_i^K = \mu_\tau \otimes \alpha,$$

where  $\alpha = (\pi_\tau)_\# \mu_i^K$  and the measure  $\mu_\tau$  is a probability measure concentrated in  $[0, r]$  for  $\alpha$ -a.e.  $\tau$ . Let us now fix an arbitrarily small constant  $\delta$ , much smaller than  $\varepsilon$ , and for  $\alpha$ -a.e.  $\tau \in [0, r]^{d-1}$  let

$$f_\tau : \left\{ (\sigma, \tau) \in [0, r]^d : \mu_\tau([0, \sigma]) \leq 1 - \delta \right\} \rightarrow \left\{ (\sigma, \tau) \in [0, r]^d : \mu_\tau([0, \sigma]) \geq \delta \right\}$$

be the measurable function given by

$$\mu_\tau((\sigma, f_\tau(\sigma))) = \delta.$$

Notice that these functions are well-defined because  $\mu_i^K$  is absolutely continuous with respect to the  $d$ -dimensional Lebesgue measure, and thus for  $\alpha$ -a.e.  $\tau$  the measure  $\mu_\tau$  is absolutely continuous with respect to the 1-dimensional Lebesgue measure. If the density of  $\mu_i^K$  is constant, then the functions  $f_\tau$  are nothing else than the right translation of a distance  $r\delta$ ; therefore, we can expect that most of the functions  $f_\tau$  move points to the right more or less of a distance comparable with  $r\delta$ . More precisely, we define a large constant  $H$  as

$$H = \frac{Kr^d}{\varepsilon m_i}; \quad (3.5)$$

notice that  $H$  depends on  $r$ , on  $\varepsilon$ , on the measures  $\gamma_i$  (thus on  $\bar{\delta}$ ) and on  $K$  (so, again on  $\varepsilon$  and on  $\bar{\delta}$ ), but not on  $\delta$ : in particular,  $\delta/H$  is arbitrarily small. We define now

$$Z = \left\{ \tau \in [0, r]^{d-1} : \exists \sigma, f_\tau(\sigma) - \sigma < \frac{r\delta}{H} \right\},$$

and we claim that

$$\alpha(Z) \leq \varepsilon m_i. \quad (3.6)$$

Indeed, by the Measurable Selection Theorem we can select a measurable function  $\tau \mapsto \sigma(\tau)$ , which associates to every  $\tau \in Z$  some  $\sigma(\tau)$  with the property that  $f_\tau(\sigma(\tau)) - \sigma(\tau) < r\delta/H$ , and we can also define the “box”

$$\Gamma = \left\{ (\sigma, \tau) \in [0, r]^d : \tau \in Z, \sigma(\tau) < \sigma < \sigma(\tau) + \frac{r\delta}{H} \right\}.$$

Thus, by Fubini Theorem and recalling the decomposition  $\mu_i^K = \mu_\tau \otimes \alpha$ , on one side we have that

$$\mu_i^K(\Gamma) \leq K \mathcal{L}(\Gamma) \leq K \frac{r\delta}{H} \mathcal{H}^{d-1}(Z) \leq K \frac{r^d \delta}{H},$$

and on the other side that

$$\mu_i^K(\Gamma) = \int_{\tau \in Z} \mu_\tau \left( \left( \sigma(\tau), \sigma(\tau) + \frac{r\delta}{H} \right) \right) d\alpha(\tau) \geq \int_{\tau \in Z} \mu_\tau \left( (\sigma(\tau), f_\tau(\sigma(\tau))) \right) d\alpha(\tau) = \delta \alpha(Z),$$

thus by the choice (3.5) the estimate (3.6) follows.

We can now go into the definition of the plan  $\xi_i^1$ ; the very rough idea is to “copy” the original plan  $\gamma_i$ , but instead of sending a generic point  $x = (\sigma, \tau)$  onto  $y$ , we send the corresponding point  $(f_\tau(\sigma), \tau)$

onto  $y$ . If  $\tau \notin Z$ , then we are sure that  $f_\tau(\sigma)$  is at least a given bit more on the right, with respect to  $\sigma$ , and then the distance has been decreased and it is reasonable to hope that the cost has been lowered. Let us now make it formal: we define the sets

$$L = \left\{ (\sigma, \tau) \in [0, r]^d : \tau \notin Z, \mu_\tau([0, \sigma]) \leq 1 - \delta \right\}, \quad R = \left\{ (\sigma, \tau) \in [0, r]^d : \tau \notin Z, \mu_\tau([0, \sigma]) \geq \delta \right\},$$

the functions  $\tilde{g} : L \rightarrow [0, r]^d$  and  $g : L \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$  as

$$\tilde{g}(\sigma, \tau) = (f_\tau(\sigma), \tau), \quad g(x, y) = (\tilde{g}(x), y),$$

and then the plan  $\xi_i^1$  as

$$\xi_i^1 = g_{\#} \left( \gamma_i^K \llcorner L \times \mathbb{R}^d \right).$$

It is not difficult to check that the two marginals of  $\xi_i^1$  are given by

$$\pi_1 \xi_i^1(A) = \mu_i^K(A \cap R) \leq \mu_i(A), \quad \pi_2 \xi_i^1(B) = \gamma_i^K(L \times B) \leq \nu_i(B) :$$

here and in the following, for simplicity of notations, we denote the two marginals of a generic plan  $\zeta$  as  $\pi_1 \zeta$  and  $\pi_2 \zeta$ , instead than  $\pi_{1\#} \zeta$  and  $\pi_{2\#} \zeta$ . As a consequence, if we write

$$\mu_i = \pi_1 \xi_i^1 + \mu_{i,\text{rem}}, \quad \nu_i = \pi_2 \xi_i^1 + \nu_{i,\text{rem}},$$

we can evaluate by (3.4) and (3.6)

$$\|\nu_{i,\text{rem}}\| = \|\mu_{i,\text{rem}}\| = \|\mu_i - \mu_i^K\| + \mu_i^K(Q_i^1 \setminus R) = \|\mu_i - \mu_i^K\| + \delta \alpha([0, r]^{d-1} \setminus Z) + \alpha(Z) \leq (2\varepsilon + \delta) m_i. \quad (3.7)$$

Step IIIb. Estimate on the cost of  $\xi_i^1$ .

In this substep, we obtain an estimate on the cost of the plan  $\xi_i^1$ . Let us take  $(z, y) \in \text{spt } \xi_i^1$ : this means that  $z = \tilde{g}(x)$  for some  $x \in L$ , with  $(x, y) \in \text{spt } \gamma_i^K$ ; in particular,  $y - x \in S_{\bar{\delta}}$ , and if we write  $x = (\sigma, \tau)$ , then  $z = (f_\tau(\sigma), \tau)$ . Hence, we have that  $z - x = (f_\tau(\sigma) - \sigma) e_1$ ; let us then call  $v = y - x$ ,  $w = a e_1$  and  $\eta = (f_\tau(\sigma) - \sigma)/a$ , so that  $y - z = v - \eta w$ . Since  $(x, y) \in \text{spt } \gamma_i^K$ , we have that  $v = y - x \in \mathcal{C} \setminus (1 - \bar{\delta}) \mathcal{C}$ , and by construction we have that  $|w - v| \leq r \sqrt{d} \ll r^+ \sqrt{d}$ . Moreover,  $x \in L$ , thus

$$\frac{r\delta}{aH} \leq \eta \leq \frac{r}{a} < \frac{2r}{A}.$$

As a consequence, if

$$1 - \frac{r\delta}{2H} < t < 1,$$

then for sure  $|1 - t| < \eta/2$ , hence we can apply Lemma 2.3 to find that

$$\begin{aligned} |c_1(x, y) - c_t(g(x, y))| &= |c_1(x, y) - c_t(z, y)| = \left| h(y - x) - h\left(\frac{y - z}{t}\right) \right| = \left| h(v) - h\left(\frac{v - \eta w}{t}\right) \right| < \phi(\eta) \\ &< \phi\left(\frac{2r}{A}\right). \end{aligned}$$

By the definition of  $\xi_i^1$ , we get then

$$\begin{aligned} \limsup_{t \nearrow 1} \mathcal{C}_t(\xi_i^1) &= \limsup_{t \nearrow 1} \iint_{\mathbb{R}^d \times \mathbb{R}^d} c_t(z, y) d\xi_i^1(z, y) = \limsup_{t \nearrow 1} \iint_{\mathbb{R}^d \times \mathbb{R}^d} c_t(g(x, y)) d\gamma_i^K(x, y) \\ &\leq \mathcal{C}_1(\gamma_i^K) + \|\gamma_i^K\| \phi\left(\frac{2r}{A}\right). \end{aligned} \quad (3.8)$$

Step III. Definition of the transport plans  $\xi^2$  and  $\xi^3$ .

In the preceding step, we have found a transport plan  $\xi_i^1$  which is sending ‘‘almost all’’ of  $\mu_i$  onto ‘‘almost all’’ of  $\nu_i$ . To complete the construction of our competitor transport plan  $\xi$ , we have then to fix the remaining parts of the  $\mu_i$  and  $\nu_i$ , as well as to send  $\mu_0$  onto  $\nu_0$ . To do so, we will make use of the Chain

Lemma 2.7. More precisely, for every  $i$  we apply the Chain Lemma with constant  $M\|\mu_{i,\text{rem}}\|$ : notice that this is possible only if the constant is smaller than  $\varepsilon_i$ , but in fact by (3.7)

$$M\|\mu_{i,\text{rem}}\| \leq M(2\varepsilon + \delta)m_i \leq 3\varepsilon M \ll \bar{\varepsilon} \leq \bar{\varepsilon}_i.$$

The Chain Lemma then provides us with measures, which we call for simplicity  $M\tilde{\gamma}_i$  and  $M\tilde{\gamma}'_i$ ; then,  $\tilde{\gamma}_i = \tilde{\gamma}_i^A + \tilde{\gamma}_i^\infty$  with

$$M\tilde{\gamma}_i^A \leq \gamma_i, \quad M\tilde{\gamma}_i^\infty \leq \gamma - \gamma_i, \quad \|\tilde{\gamma}_i^A\| = \|\mu_{i,\text{rem}}\|, \quad \|\tilde{\gamma}_i^\infty\| = N_i\|\mu_{i,\text{rem}}\| \leq N\|\mu_{i,\text{rem}}\|. \quad (3.9)$$

Notice that we have the inequality  $\tilde{\gamma}_i^\infty \leq \gamma - \gamma_i$ , but this does not mean  $\tilde{\gamma}_i^\infty \leq \gamma_0$ , since  $\tilde{\gamma}_i^\infty$  might have parts in common with  $\gamma_j$  for some  $j \neq i$ . We can then further subdivide  $\tilde{\gamma}_i^\infty = \tilde{\gamma}_i^{\text{OUT}} + \tilde{\gamma}_i^{\text{NO}}$ , with

$$M\tilde{\gamma}_i^{\text{OUT}} \leq \gamma_0, \quad M\tilde{\gamma}_i^{\text{NO}} \leq \gamma - \gamma_0 - \gamma_i. \quad (3.10)$$

The marginals of these measures are

$$\begin{aligned} \pi_1\tilde{\gamma}_i^A &= \mu_i^A \leq \mu_i, & M\pi_1\tilde{\gamma}_i^{\text{OUT}} &\leq \mu_0, & \pi_1\tilde{\gamma}_i^{\text{NO}} &\leq \mu - \mu_0 - \mu_i, \\ \pi_2\tilde{\gamma}_i^A &= \nu_i^A \leq \nu_i, & M\pi_2\tilde{\gamma}_i^{\text{OUT}} &\leq \nu_0, & \pi_2\tilde{\gamma}_i^{\text{NO}} &\leq \nu - \nu_0 - \nu_i. \end{aligned}$$

Instead, the marginals of  $\tilde{\gamma}'_i$  are given by

$$\pi_1\tilde{\gamma}'_i = \pi_1\tilde{\gamma}_i = \mu_i^A + \pi_1(\tilde{\gamma}_i^{\text{OUT}} + \tilde{\gamma}_i^{\text{NO}}), \quad \pi_2\tilde{\gamma}'_i = \nu_i^B + \pi_2(\tilde{\gamma}_i^{\text{OUT}} + \tilde{\gamma}_i^{\text{NO}}), \quad (3.11)$$

where  $\nu_i^B$  does not necessarily coincide with  $\nu_i^A$ , but they are both measures of norm  $\|\mu_{i,\text{rem}}\|$  smaller than  $\nu_i$ . We are then ready to define the plan  $\xi^2$  as

$$\xi^2 = \gamma_0 - \gamma_0^{\text{OUT}}, \quad \text{where} \quad \gamma_0^{\text{OUT}} := \sum_{i=1}^M \tilde{\gamma}_i^{\text{OUT}}.$$

Notice that  $\xi^2$  is a positive measure because, according to (3.10), we do not have just  $\tilde{\gamma}_i^{\text{OUT}} \leq \gamma_0$ , but also  $\tilde{\gamma}_i^{\text{OUT}} \leq \gamma_0/M$ : in fact, the reason why we have applied the Chain Lemma with constants  $M\|\mu_{i,\text{rem}}\|$  instead of just  $\|\mu_{i,\text{rem}}\|$  was precisely to be sure to get, at this point, a *positive* measure  $\xi^2$ . We also call

$$\mu_0^{\text{OUT}} = \pi_1\gamma_0^{\text{OUT}} \leq \mu_0, \quad \nu_0^{\text{OUT}} = \pi_2\gamma_0^{\text{OUT}} \leq \nu_0,$$

so that the marginals of  $\xi^2$  are

$$\pi_1\xi^2 = \mu_0 - \mu_0^{\text{OUT}}, \quad \pi_2\xi^2 = \nu_0 - \nu_0^{\text{OUT}}. \quad (3.12)$$

Let us now set  $\gamma^{\text{NO}} = \sum_{i=1}^M \tilde{\gamma}_i^{\text{NO}}$ ; since again by (3.10)  $\gamma^{\text{NO}} \leq \gamma - \gamma_0$ , we can decompose it as

$$\gamma^{\text{NO}} = \sum_{i=1}^M \gamma_i^{\text{NO}}, \quad \text{with} \quad \gamma_i^{\text{NO}} \leq \gamma_i.$$

Notice that  $\gamma_i^{\text{NO}}$  does not coincide with  $\tilde{\gamma}_i^{\text{NO}}$ ; on the contrary,  $\gamma_i^{\text{NO}}$  comes from the measures  $\tilde{\gamma}_j^{\text{NO}}$  for all  $j \neq i$ . Let us finally call  $\mu_i^{\text{NO}}$  and  $\nu_i^{\text{NO}}$  the two marginals of  $\gamma_i^{\text{NO}}$ , and notice that by (3.9) and (3.7)

$$\|\mu_i^{\text{NO}}\| = \|\gamma_i^{\text{NO}}\| \leq \|\gamma^{\text{NO}}\| = \sum_{i=1}^M \|\tilde{\gamma}_i^{\text{NO}}\| \leq \sum_{i=1}^M \|\tilde{\gamma}_i^\infty\| \leq \sum_{i=1}^M N\|\mu_{i,\text{rem}}\| \leq N \sum_{i=1}^M 3\varepsilon m_i \leq 3\varepsilon N \ll m_i,$$

while

$$\|\xi_i^1\| = \|\mu_i - \mu_{i,\text{rem}}\| \geq (1 - 3\varepsilon)m_i.$$

As a consequence, all the constants  $\lambda_i$  given by

$$\lambda_i = 1 - \frac{\|\mu_i^{\text{NO}}\|}{\|\xi_i^1\|}$$

are only slightly smaller than 1; define then

$$\mu_{i,\text{rem}}^+ = \mu_i - \pi_1(\lambda_i \xi_i^1) \geq \mu_{i,\text{rem}}, \quad \nu_{i,\text{rem}}^+ = \nu_i - \pi_2(\lambda_i \xi_i^1) \geq \nu_{i,\text{rem}},$$

and observe that, by the definition of  $\lambda_i$ , we have

$$\|\mu_{i,\text{rem}}^+\| = \|\nu_{i,\text{rem}}^+\| = \|\mu_{i,\text{rem}}\| + (1 - \lambda_i)\|\xi_i^1\| = \|\mu_{i,\text{rem}}\| + \|\mu_i^{\text{NO}}\| = \|\mu_i^A\| + \|\mu_i^{\text{NO}}\|. \quad (3.13)$$

Finally, we can set

$$\xi^1 := \sum_{i=1}^M \lambda_i \xi_i^1,$$

whose marginals are

$$\pi_1 \xi^1 = \sum_{i=1}^M \mu_i - \mu_{i,\text{rem}}^+, \quad \pi_2 \xi^1 = \sum_{i=1}^M \nu_i - \nu_{i,\text{rem}}^+. \quad (3.14)$$

We aim to define our competitor plan as  $\xi = \xi^1 + \xi^2 + \xi^3$ : by (3.12) and (3.14), the marginals of  $\xi^3$  must satisfy

$$\pi_1 \xi^3 = \mu_0^{\text{OUT}} + \sum_{i=1}^M \mu_{i,\text{rem}}^+, \quad \pi_2 \xi^3 = \nu_0^{\text{OUT}} + \sum_{i=1}^M \nu_{i,\text{rem}}^+. \quad (3.15)$$

Let us consider the measure  $\xi_{\text{TEMP}}^3 = \sum_{i=1}^M \tilde{\gamma}_i$ : keeping in mind (3.11), we have

$$\begin{aligned} \pi_1 \xi_{\text{TEMP}}^3 &= \sum_{i=1}^M \mu_i^A + \pi_1(\tilde{\gamma}_i^{\text{OUT}} + \tilde{\gamma}_i^{\text{NO}}) = \mu_0^{\text{OUT}} + \sum_{i=1}^M \mu_i^A + \mu_i^{\text{NO}}, \\ \pi_2 \xi_{\text{TEMP}}^3 &= \sum_{i=1}^M \nu_i^B + \pi_2(\tilde{\gamma}_i^{\text{OUT}} + \tilde{\gamma}_i^{\text{NO}}) = \nu_0^{\text{OUT}} + \sum_{i=1}^M \nu_i^B + \nu_i^{\text{NO}}. \end{aligned}$$

Notice that the marginals of  $\xi_{\text{TEMP}}^3$  are almost exactly those required for  $\xi^3$  in (3.15), the only ‘‘mistake’’ being that for each  $1 \leq i \leq M$  in place of the measures  $\mu_{i,\text{rem}}^+ \leq \mu_i$  and  $\nu_{i,\text{rem}}^+ \leq \nu_i$  one has the measures  $\mu_i^A + \mu_i^{\text{NO}} \leq \mu_i$  and  $\nu_i^B + \nu_i^{\text{NO}} \leq \nu_i$ , which have anyway the same mass thanks to the estimate (3.13). It is then easy to adjust the measure  $\xi_{\text{TEMP}}^3$ : we define the two auxiliary transport plans

$$\beta_1 = (\text{Id}, \text{Id})_{\#} \mu_0^{\text{OUT}} + \sum_{i=1}^M \mu_{i,\text{rem}}^+ \otimes (\mu_i^A + \mu_i^{\text{NO}}), \quad \beta_2 = (\text{Id}, \text{Id})_{\#} \nu_0^{\text{OUT}} + \sum_{i=1}^M (\nu_i^B + \nu_i^{\text{NO}}) \otimes \nu_{i,\text{rem}}^+.$$

Notice that  $\beta_1$  has first marginal  $\mu_0^{\text{OUT}} + \sum_i \mu_{i,\text{rem}}^+$  and second marginal  $\mu_0^{\text{OUT}} + \sum_i (\mu_i^A + \mu_i^{\text{NO}})$ , while  $\beta_2$  has first marginal  $\nu_0^{\text{OUT}} + \sum_i (\nu_i^B + \nu_i^{\text{NO}})$  and second marginal  $\nu_0^{\text{OUT}} + \sum_i \nu_{i,\text{rem}}^+$ . Therefore, if we finally define the composition  $\xi^3 = \beta_2 \circ \xi_{\text{TEMP}}^3 \circ \beta_1$  in the sense of Definition 2.2, then  $\xi_3$  is a positive measure whose marginals satisfy (3.15), thus the plan  $\xi = \xi^1 + \xi^2 + \xi^3$  is an admissible transport plan.

*Step IV. Estimate on the cost of the transport plan  $\xi$ .*

In this last step we want to estimate the cost of the transport plan  $\xi$ : in particular, we will establish (3.2), so concluding the proof. By linearity of the cost, we have of course  $\mathcal{C}_t(\xi) = \mathcal{C}_t(\xi^1) + \mathcal{C}_t(\xi^2) + \mathcal{C}_t(\xi^3)$ , hence we will consider the three terms separately. Concerning  $\xi^1$ , it is enough to recall (3.8) and the fact that the constants  $\lambda_i$  from Step III are smaller than 1, so to get

$$\begin{aligned} \limsup_{t \nearrow 1} \mathcal{C}_t(\xi^1) &= \limsup_{t \nearrow 1} \sum_{i=1}^M \lambda_i \mathcal{C}_t(\xi_i^1) \leq \sum_{i=1}^M \limsup_{t \nearrow 1} \mathcal{C}_t(\xi_i^1) \leq \sum_{i=1}^M \mathcal{C}_1(\gamma_i^K) + \|\gamma_i^K\| \phi\left(\frac{2r}{A}\right) \\ &\leq \mathcal{C}_1(\gamma - \gamma_0) + \phi\left(\frac{2r}{A}\right). \end{aligned} \quad (3.16)$$

Let us now consider  $\xi^2$ : since  $\xi^2 \leq \gamma_0$ , then of course  $\mathcal{C}_t(\xi^2) \leq \mathcal{C}_t(\gamma_0)$ . On the other hand, the transport plan  $\gamma_0$  is concentrated by definition in  $S_{\bar{\delta}}$ ; that is, for  $\gamma_0$ -a.e.  $(x, y)$ , one has  $y - x \in (1 - \bar{\delta})\mathcal{C}$ . Since the function  $h$  is strictly convex in the whole  $\mathcal{C}$ , it is uniformly Lipschitz in a neighborhood of  $(1 - \bar{\delta})\mathcal{C}$ , which implies

$$\sup_{(x,y) \in S_{\bar{\delta}}} |c_t(x, y) - c_1(x, y)| \xrightarrow{t \rightarrow 1} 0.$$

Consequently, we can simply estimate

$$\limsup_{t \nearrow 1} \mathcal{C}_t(\xi^2) \leq \limsup_{t \nearrow 1} \mathcal{C}_t(\gamma_0) \leq \mathcal{C}_1(\gamma_0). \quad (3.17)$$

Finally, let us pass to consider  $\xi^3$ , which was defined as  $\xi^3 = \beta_2 \circ \xi_{\text{TEMP}}^3 \circ \beta_1$ . Keep in mind that  $\xi_{\text{TEMP}}^3 \leq \gamma'$  by definition, hence a pair  $(x, y) \in \text{spt}(\xi_{\text{TEMP}}^3)$  must satisfy  $y - x \in t'\mathcal{C} \subseteq (1 - \bar{\delta})\mathcal{C}$ . Moreover, the plans  $\beta_1$  and  $\beta_2$  are only moving points inside given squares; more precisely, if  $(\tilde{x}, x) \in \text{spt} \beta_1$  then necessarily  $|x - \tilde{x}| \leq r\sqrt{d}$ , and similarly if  $(y, \tilde{y}) \in \text{spt} \beta_2$  then  $|\tilde{y} - y| \leq r\sqrt{d}$ . As a consequence, keeping in mind (3.1) we have that for every  $(\tilde{x}, \tilde{y}) \in \text{spt} \xi^3$  it is  $\tilde{y} - \tilde{x} \in (1 - \frac{\bar{\delta}}{2})\mathcal{C}$ , and then  $c_t(\tilde{x}, \tilde{y}) < +\infty$  for every  $t$  close enough to 1. Summarizing, for any such  $t$  we have, also recalling (3.7) and (3.3),

$$\mathcal{C}_t(\xi^3) \leq \|h\|_{L^\infty} \|\xi^3\| = \|h\|_{L^\infty} \|\xi_{\text{TEMP}}^3\| \leq (N+1) \sum_{i=1}^M \|\mu_{i, \text{rem}}\| \leq 3\varepsilon(N+1) \sum_{i=1}^M m_i \leq 3\varepsilon(N+1) \leq \phi\left(\frac{2r}{A}\right).$$

Putting this last estimate together with (3.16) and (3.17), we have finally established (3.2), and the proof is complete.  $\square$

#### 4. DIRECTIONS IN THE BOUNDARY OF $\mathcal{C}$ WITH INFINITE SLOPE ARE NOT USED BY OPTIMAL PLANS

In this section we prove the claim (ii) of Theorem A, which says that, for every supercritical time  $t > T$ , the optimal transport plan  $\gamma_t$  does not use vectors in the boundary of  $\mathcal{C}$  at which the directional derivative of  $h$  is  $-\infty$ . Since this happens at all the vectors in the boundary of  $\mathcal{C}$  when the cost is highly relativistic (this is indeed the definition of highly relativistic costs), then this claim generalizes the analogous one in Theorem 1.2. The construction needed to prove this part is very similar to the one that we performed to prove the first part of Theorem A; in fact, the situation this time is much simpler, because we need to use only a single square in  $\mathbb{R}^d \times \mathbb{R}^d$ .

*Proof of Theorem A, part (ii).* Up to rescaling, we can assume for simplicity that  $T < t = 1$ . Exactly as in the proof of part (i), let us arbitrarily fix some  $T < t' < 1$  and call  $\gamma$  and  $\gamma'$  the optimal transport plans corresponding to the times  $t = 1$  and  $t = t'$ . Recall that we have to show that  $\gamma$  does not charge the pairs  $(x, y)$  with  $y - x \in \Theta$ , where the set  $\Theta$  is defined as

$$\Theta = \left\{ v \in \partial\mathcal{C} : D_{-v}h(v) = -\infty \right\}.$$

Assume, instead, that the measure  $\gamma \llcorner \{(x, y) : y - x \in \Theta\}$  is non trivial, and let  $(\bar{x}, \bar{y})$  belong to its support. Up to a rotation and a rescaling, we can assume that  $\bar{y} - \bar{x} = e_1$ . Let us now fix a small constant  $r \ll 1 - t'$ , call  $Q_1$  and  $Q_2$  the two squares centered at  $\bar{x}$  and  $\bar{y}$  with sides parallel to the coordinate axes and of length  $r$ , and define

$$\gamma_0 = \gamma \llcorner \left\{ (x, y) \in Q_1 \times Q_2 : y - x \in \Theta \right\}.$$

Notice that  $\gamma_0$  is not the trivial measure by construction; call also  $\mu_0$  and  $\nu_0$  the marginals of  $\gamma_0$ ; up to a last translation, we assume for simplicity that  $Q_1 = [0, r]^d$  and  $Q_2 = Q_1 + e_1$ . Let us apply the Chain

Lemma 2.7 to the measures  $\gamma$ ,  $\gamma'$  and  $\gamma_0$ , finding the constants  $\bar{\varepsilon} > 0$  and  $N \in \mathbb{N}$ . Now, we select some positive constant  $K$  such that

$$\gamma_0(\{(x, y) \in Q_1 \times Q_2 : \rho(x) < K\}) > 0,$$

calling again  $\rho$  the density of  $\mu$  with respect to the Lebesgue measure. Let us call  $\gamma_{\text{TEMP}}^K$  the restriction of  $\gamma_0$  to the points  $(x, y) \in Q_1 \times Q_2$  with  $\rho(x) < K$ , let  $m_{\text{TEMP}} \ll \bar{\varepsilon}$  be its mass, and let  $\mu_{\text{TEMP}}^K$  and  $\nu_{\text{TEMP}}^K$  be its two marginals. Now, we can repeat verbatim the construction of Step II in the proof of part (i) of Theorem A (with  $\varepsilon = 1/2$ , which is enough for this proof), disintegrating  $\mu_{\text{TEMP}}^K = \mu_\tau \otimes \alpha$ , and defining first the constant  $\delta \ll 1$  (to be sent to 0 at the end) and the function  $f_\tau$  for  $\tau \in [0, r]^{d-1}$ , then the constant  $H = 2Kr^d/m_{\text{TEMP}}$  and the set  $Z$  with  $\alpha(Z) \leq m_{\text{TEMP}}/2$ , and finally the sets  $L$  and  $R$  and the functions  $\tilde{g}$  and  $g$ . We now call

$$\gamma^K = \gamma_{\text{TEMP}}^K \mathbf{L} \left\{ (x, y) \in Q_1 \times Q_2 : x = (\sigma, \tau), \tau \notin Z \right\},$$

and let as usual  $\mu^K$  and  $\nu^K$  be its marginals, and  $m \in [m_{\text{TEMP}}/2, m_{\text{TEMP}}]$  its total mass. Define now

$$\xi^1 = g_{\#} \gamma^K \mathbf{L} (L \times Q_2),$$

and notice carefully that its two marginals are given by

$$\pi_1 \xi^1(A) = \mu_{\text{TEMP}}^K(A \cap R) \leq \mu^K(A), \quad \pi_2 \xi^1(B) = \gamma_{\text{TEMP}}^K(L \times B) \leq \nu^K(B);$$

as a consequence, we can define the ‘‘remaining measures’’

$$\mu_{\text{rem}} = \mu^K - \pi_1 \xi^1, \quad \nu_{\text{rem}} = \nu^K - \pi_2 \xi^1,$$

whose measure is  $\eta := \|\mu_{\text{rem}}\| = \|\nu_{\text{rem}}\| = \delta m$  by construction. Since  $\delta$  can be taken arbitrarily small, we can assume that  $\eta < \bar{\varepsilon}$ , so that the Chain Lemma 2.7 provides us with measures  $\tilde{\gamma} = \tilde{\gamma}_0 + \tilde{\gamma}_\infty$ ,  $\tilde{\gamma}'$ ,  $\tilde{\mu}$ ,  $\tilde{\nu}$ ,  $\mu_A$ ,  $\nu_A$  and  $\nu_B$  with  $\|\tilde{\gamma}_0\| = \eta$  and  $\|\tilde{\gamma}_\infty\| = N\eta$  satisfying

$$\begin{aligned} \mu_A &\leq \mu_0, & \nu_A, \nu_B &\leq \nu_0, & \pi_1 \tilde{\gamma}_0 &= \mu_A, & \pi_2 \tilde{\gamma}_0 &= \nu_A, \\ \pi_1 \tilde{\gamma}_\infty &= \tilde{\mu}, & \pi_2 \tilde{\gamma}_\infty &= \tilde{\nu}, & \pi_1 \tilde{\gamma}' &= \tilde{\mu} + \mu_A, & \pi_2 \tilde{\gamma}' &= \tilde{\nu} + \nu_B. \end{aligned}$$

We can then immediately define  $\xi^2 = \gamma - \gamma^K - \tilde{\gamma}_\infty$ , and observe that the two marginals of  $\xi^1 + \xi^2$  are

$$\pi_1(\xi^1 + \xi^2) = \mu - \mu_{\text{rem}} - \tilde{\mu}, \quad \pi_2(\xi^1 + \xi^2) = \nu - \nu_{\text{rem}} - \tilde{\nu}.$$

As a consequence, in order to get a competitor transport plan  $\xi = \xi^1 + \xi^2 + \xi^3$ , we need a plan  $\xi^3$  with marginals  $\tilde{\mu} + \mu_{\text{rem}}$  and  $\tilde{\nu} + \nu_{\text{rem}}$ , so exactly as in the proof of last section we simply define  $\xi^3 = \beta_2 \circ \tilde{\gamma}' \circ \beta_1$ , being the auxiliary transport plans  $\beta_1$  and  $\beta_2$  defined as

$$\beta_1 = (\text{Id}, \text{Id})_{\#} \tilde{\mu} + \mu_{\text{rem}} \otimes \mu_A, \quad \beta_2 = (\text{Id}, \text{Id})_{\#} \tilde{\nu} + \nu_B \otimes \nu_{\text{rem}}.$$

Estimating the cost of  $\xi^3$  is very simple: as in last section, we only have to observe that every pair  $(x, y)$  in the support of  $\tilde{\gamma}'$  satisfies  $(y - x) \in t'\mathcal{C}$ , while the auxiliary plans  $\beta_1$  and  $\beta_2$  only move points at most of a distance at most  $2r \ll 1 - t'$ , so we get that for all the pairs  $(x, y) \in \text{spt } \xi^3$  it is  $y - x \in \mathcal{C}$ . Thus, since  $\gamma$  is an optimal transport plan for time  $t = 1$ , we can estimate

$$\begin{aligned} \mathcal{C}_1(\gamma) &\leq \mathcal{C}_1(\xi) = \mathcal{C}_1(\xi^1) + \mathcal{C}_1(\xi^2) + \mathcal{C}_1(\xi^3) \leq \mathcal{C}_1(\xi^1) + \mathcal{C}_1(\gamma - \gamma^K) + \|h\|_{L^\infty} \|\xi^3\| \\ &= \mathcal{C}_1(\xi^1) + \mathcal{C}_1(\gamma - \gamma^K) + \|h\|_{L^\infty} (N + 1) \delta m, \end{aligned}$$

which implies

$$\mathcal{C}_1\left(\gamma^K \mathbf{L} (L \times Q_2)\right) - \mathcal{C}_1\left(g_{\#}\left(\gamma^K \mathbf{L} (L \times Q_2)\right)\right) \leq \|h\|_{L^\infty} (N + 1) \delta m, \quad (4.1)$$

Let us now estimate the left term in the last inequality as

$$\begin{aligned} \mathcal{C}_1\left(\gamma^K \llcorner (L \times Q_2)\right) - \mathcal{C}_1\left(g_{\#}\left(\gamma^K \llcorner (L \times Q_2)\right)\right) &= \iint_{L \times Q_2} c(x, y) - c(\tilde{g}(x), y) d\gamma^K(x, y) \\ &= \iint_{L \times Q_2} h(y - x) - h(y - \tilde{g}(x)) d\gamma^K(x, y) = \iint_{Q_1 \times Q_2} \chi_L(x) (h(y - x) - h(y - \tilde{g}(x))) d\gamma^K(x, y), \end{aligned} \quad (4.2)$$

and notice that for  $\gamma^K$ -a.e.  $(x, y)$  we have  $y - \tilde{g}(x) = y - x - \varphi(x)e_1$ , with

$$\frac{m\delta}{2Kr^{d-1}} \leq \varphi(x) \leq r.$$

Observe now that  $y - x$  is in a very small neighborhood of  $e_1$ , hence by convexity of  $h$ , recalling also that  $h(0) = 0 < h(e_1)$ , we immediately get that  $h$  strictly decreases in the direction  $-e_1$  in a small neighborhood of  $e_1$ . Thus, we can evaluate

$$h(y - \tilde{g}(x)) \leq h(y - x - c\delta e_1),$$

setting  $c = m/(2Kr^{d-1})$ . Putting this estimate together with (4.1) and (4.2), we obtain

$$\iint_{Q_1 \times Q_2} \chi_L(x) \frac{h(y - x) - h(y - x - c\delta e_1)}{\delta} d\gamma^K(x, y) \leq \|h\|_{L^\infty}(N + 1)m.$$

Now, keep in mind that the set  $L$  actually depends on the choice of  $\delta$ ; nevertheless, it is obvious from the definition that the set  $L$  increases when  $\delta$  decreases, and the union of all the sets  $L$  for  $\delta \rightarrow 0$  covers  $\gamma^K$  all of  $Q_1$ . As a consequence, for  $\gamma^K$ -a.e.  $(x, y)$  the function

$$\delta \mapsto \chi_L(x) \frac{h(y - x) - h(y - x - c\delta e_1)}{\delta}$$

is increasing for  $\delta \searrow 0$ , and it converges to  $-cD_{-e_1}h(y - x)$ . The Monotone Convergence Theorem gives then

$$\iint_{Q_1 \times Q_2} -D_{-e_1}h(y - x) d\gamma^K(x, y) \leq \|h\|_{L^\infty}(N + 1)2Kr^{d-1}.$$

And finally,  $\gamma^K$  is concentrated on pairs  $(x, y)$  with  $y - x \in \Theta$ , so with  $D_{-(y-x)}h(y - x) = -\infty$ , thus Lemma 2.5 implies  $-D_{-e_1}h(y - x) = +\infty$  for  $\gamma^K$ -a.e.  $(x, y)$ , and this gives the searched contradiction, so concluding the proof.  $\square$

**Acknowledgements.** This work has been mostly done while the first author was post-doctoral fellow at the Friederich-Alexander Universität Erlangen-Nürnberg during the academic year 2014-15, funded by the ERC grant 258685 “AnOptSetCon”.

## REFERENCES

- [1] L. Ambrosio, N. Fusco, D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press (2000).
- [2] J. Bertrand, A. Pratelli, M. Puel, *Kantorovich potentials and continuity of total cost for relativistic cost functions*, preprint (2015).
- [3] J. Bertrand, M. Puel, *The optimal mass transport problem for relativistic costs*, *Calc. Var. PDEs* **46** (2013), no. 1–2, 353–374.
- [4] Y. Brenier, *Extended Monge-Kantorovich theory*, in “*Optimal Transportation and Applications*”, *Lecture Notes in Mathematics*, LNM **1813**, Springer (2003), 91–121.
- [5] R. McCann, M. Puel, *Constructing a relativistic heat flow by transport time steps*, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26** (2009), no. 6, 2539–2580.
- [6] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Birkhäuser, (2015).
- [7] C. Villani, *Optimal transport: Old and New*, Springer-Verlag, Berlin (2009).

CEREMADE, UNIVERSITÉ PARIS-DAUPHINE, PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY, 75 775 PARIS CEDEX  
16, FRANCE

INRIA, MOKAPLAN, 2 RUE SIMONE IFF, 75 012 PARIS, FRANCE

*E-mail address:* `louet@ceremade.dauphine.fr`

DEPARTMENT MATHEMATIK, FRIEDERICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG, CAUERSTRASSE,11, 91058  
ERLANGEN (GERMANY)

*E-mail address:* `pratelli@math.fau.de`, `zeisler@math.fau.de`