KANTOROVICH POTENTIALS AND CONTINUITY OF TOTAL COST FOR RELATIVISTIC COST FUNCTIONS

JEROME BERTRAND, ALDO PRATELLI, AND MARJOLAINE PUEL

Abstract. In this paper we consider the optimal mass transport problem for relativistic cost functions, introduced in [12] as a generalization of the relativistic heat cost. A typical example of such a cost function is $c_t(x,y) = h(\frac{y-x}{t})$, h being a strictly convex function when the variable lies on a given ball, and infinite otherwise. It has been already proved that, for every t larger than some critical time $T > 0$, existence and uniqueness of optimal maps hold; nonetheless, the existence of a Kantorovich potential is known only under quite restrictive assumptions. Moreover, the total cost corresponding to time t has been only proved to be a decreasing right-continuous function of t. In this paper, we extend the existence of Kantorovich potentials to a much broader setting, and we show that the total cost is a continuous function. To obtain both results the two main crucial steps are a refined "chain lemma" and the result that, for $t > T$, the points moving at maximal distance are negligible for the optimal plan.

1. Introduction

In this paper, we consider the classical mass transport problem with a particular class of cost functions c. While the case when c is strictly convex and real-valued is well understood (we refer to [34] for a survey), our goal is to continue the study of a wide class of non real-valued cost functions, called "relativistic costs", which were introduced in [12] as a generalization to the "relativistic heat cost".

More precisely, the *relativistic heat cost*, introduced by Brenier in [14], is defined as $c(x,y) = h(y-x)$, where h is given by the formula

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

As explained by Brenier in the aforementioned paper, this cost function can be used in order to study a relativistic heat equation, where the region where the cost is infinite comes from the fact that the heat has finite speed propagation. Afterwards, the corresponding relativistic heat equation has also been studied by McCann and Puel in [28] by means of the Jordan-Kinderlehrer-Otto approach introduced in [25], and by Andreu, Caselles and Mazón via PDE methods in a series of papers [3, 5, 4, 6, 7, 8, 9, 16]. We refer to [14, 12] and the references therein for more on this topic.

Notice that, from the mathematical point of view, a quite interesting feature of this cost function is that it is strictly convex and bounded on its domain, hence in particular it is not continuous on $\mathbb{R}^n \times \mathbb{R}^n$ (while the non real-valued cost functions have been often assumed to be continuous, see for instance [23, 22, 29, 30, 33, 11]).

More in general, one can consider costs "of relativistic type", meaning that they are strictly convex and bounded in a strictly convex domain $\mathscr{C}$ and $+\infty$ outside. The corresponding optimal mass transport problem has been studied in [12], see also [24].

In the classical case (strictly convex and real-valued cost function), a lot can be said; in particular, an optimal transport map exists, is unique, and it is induced by a *Kantorovich potential* φ (for instance the optimal transport map is simply the gradient of φ when the cost is the Euclidean norm squared, see [27, 13]). This potential is also related to the so-called Kantorovich dual problem. For non-negative and lower semi-continuous cost functions, the following duality result is well known (see [21]).

Theorem 1.1 (Kantorovich duality)**.** *Let μ and ν be two probability measures with finite second order moment on $\mathbb{R}^n$. Then, the following equality holds:*

$$\sup_{\mathcal{A}} \left\{ \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(y) d\nu(y) \right\} = \min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y)\, d\,\gamma(x,y) < +\infty.$$

Here and in the following, $\Pi(\mu,\nu)$ denotes the set of positive measures on $\mathbb{R}^n \times \mathbb{R}^n$ whose first (resp., second) marginal is μ (resp., ν), while $\mathcal{A}$ is the set of pairs (φ,ψ) of Lipschitz functions defined on $\mathbb{R}^n$ that satisfy $\varphi(x) + \psi(y) \leq c(x,y)$ for all $x, y \in \mathbb{R}^n$.

However, it should be emphasized that in general, Kantorovich's dual problem (on the left hand side) has no solutions, counterexamples can be found for instance in [10]. A sufficient condition for maximisers to exist is that the cost function is a bounded and uniformly continuous function, see [34]; a more general criterion for real-valued cost functions is given in [2]. In the case of non real-valued cost functions, the existence of maximisers is proved for reflector-type problems [23, 22] and for Alexandrov's Gauss curvature prescription problem [29, 11]; in both cases the cost function is a continuous non real-valued function on the unit sphere in Euclidean space. Existence of maximisers for the dual problem has also been proved in the case of the Wiener space for the quadratic cost [20].

One of the main results of the present paper is the existence of a Kantorovich potential for the mass transport problem relative to a relativistic cost. Beside a usual hypothesis on the initial measure, the main assumption is that of a "supercritical regime" which means, loosely speaking, that the mass transport problem is already well-posed even if the set of admissible motions is shrunk by a homothety of ratio $1-\varepsilon$. The precise statement is in Theorem B.

To conclude this introductory part, we recall that for real-valued cost functions, the enhanced formula involving a potential is the starting point of the study of the optimal map regularity. This topic, initiated by Caffarelli [15], has gone under tremendous development, we refer to [34, Chapter 12] for more on the subject. The question about the regularity of the optimal map for the relativistic heat cost in a very general setting (where only part of the mass can be moved at finite cost) is highlighted in [14].

1.1. **Previous results.** We briefly recall here some results already known, to highlight the novelty of the results of the present paper. To do so, we need to introduce a parameter t, called the "speed", which takes into account the relativistic behaviour of the cost functions we consider. To be precise, whenever $h : \mathbb{R}^n \to [0, +\infty]$ is a strictly convex function, bounded in some strictly convex set $\mathscr{C}$ and $+\infty$ outside, then for every $t > 0$ we call

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

a *relativistic cost function* (see the formal Definition 3.1). Now, given two probability measures μ and ν with compact support, we can see what happens when the speed t varies: to understand it, let us define the *total cost* as

$$\mathcal{C}(t) = \min_{\pi\in\Pi(\mu,\nu)} \mathcal{C}_t(\gamma) = \min_{\pi\in\Pi(\mu,\nu)} \iint_{\mathbb{R}^{2n}} c_t(x, y)\, d\gamma(x, y)\,. \tag{1.2}$$

It is easy to guess that the total cost is infinite when t is small (at least if $\mu \neq \nu$), while it becomes finite when t is large, and finally it does not depend on t if it is sufficiently big. The following properties are proved in [12].

Proposition 1.2. *Given two probability measures $\mu \neq \nu$ with compact support and a relativistic cost function, the following holds.*

a) The function $t \mapsto \mathcal{C}(t)$ is decreasing.
b) The speed $T := \inf\{t\ ;\ \mathcal{C}(t) < +\infty\}$ is positive and finite (T is called the critical speed*).*
c) The total cost $\mathcal{C}(T)$ is finite.
d) The function $\mathcal{C}(t)$ is right continuous.

Moreover, one can prove the existence of an optimal transport map for (super)critical speed.

Theorem 1.3. *Let μ and ν be two probability measures with compact support on $\mathbb{R}^n$ and c_t be a relativistic cost. Assume that μ is absolutely continuous with respect to the Lebesgue measure and that $t \geq T$. Then, there exists a unique optimal transport plan γ_t for the cost c_t, and this plan is induced by a map F_t.*

1.2. **Statement of the main results.** Let us briefly describe the main achievements of this paper. First of all, we can observe that the total cost is a continuous function of t when the cost is "highly relativistic": basically, a relativistic cost function is called "highly relativistic" when the slope of h explodes at the boundary $\partial\mathscr{C}$, as it happens for the original relativistic heat cost (1.1) and most of the important examples (see the formal Definition 3.2).

Theorem A (Continuity of the total cost). *Let μ and ν be two probability measures with compact support on $\mathbb{R}^n$, c_t be a highly relativistic cost function, and assume that $\mu \in L^\infty(\mathbb{R}^n)$. Then, the total cost function $t \mapsto \mathcal{C}(t)$ is continuous on $[T, +\infty)$. More precisely, for every $\rho \ll 1$ and every t, there exists $\overline{C} = \overline{C}(n, \mu, \nu, t, \rho)$ such that, if $\delta \ll 1$,*

$$\mathcal{C}(t) \leq \mathcal{C}(t - \delta t) \leq \mathcal{C}(t)\big(1 + \mathrm{exc}(\delta)\big) + \frac{\overline{C}}{\kappa_\delta(\rho)}\,, \tag{1.3}$$

where $\mathrm{exc}(\delta)$ *and $\kappa_\delta(\rho)$ converge to 0 and $+\infty$ respectively when $\delta \to 0$.*

The precise expressions of $\mathrm{exc}(\delta)$ and $\kappa_\delta(\rho)$ are introduced in Definition 4.1 and in (3.11) respectively. However, knowing that $\mathrm{exc}(\delta) \to 0$ and $\kappa_\delta(\rho) \to \infty$ when $\delta \to 0$ is enough to infer the continuity of the total cost from (1.3). Of course, Theorem A extends the plain right continuity stated in Proposition 1.2.

Our second main result is the existence of a Kantorovich potential.

Theorem B (Existence of a Kantorovich potential). *Let μ and ν be two probability measures with compact support on $\mathbb{R}^n$, c_t be a highly relativistic cost function, and assume that μ is absolutely continuous with respect to the Lebesgue measure and has a connected support. Then, for any supercritical speed $t > T$ there exists a Kantorovich potential φ_t. In particular, there is a unique optimal transport plan, induced by an optimal transport map F_t defined, μ-almost everywhere, as*

$$F_t(x) = x + t\nabla h^*\big(-t\widetilde{\nabla}\varphi_t(x)\big)\,,$$

where h^* denotes the Legendre transform of h and $\widetilde{\nabla}\varphi_t$ is the approximate gradient of φ_t.

The notions of Kantorovich potential and of approximate gradient are recalled in Definitions 5.3 and 5.4. Let us recall that the existence of a Kantorovich potential was already proved in [12], but only under a regularity assumption on μ, and only for almost every $t > T$.

We underline that all our proofs are self-contained, in particular they do not rely on Theorem 1.3.

There are two ingredients which are crucial for the proof of both our main theorems. The first one is the “finite chain Lemma” 2.11, which generalizes similar results already used in the literature, and which basically uses the cyclical monotonicity in order to modify a transport plan to get a competitor. We will also use a discrete version of this finite chain result, given in Lemma 5.5.

The second ingredient can be stated as an independent interesting result. It says that, for any supercritical time $t > T$, the optimal transport plan γ moves no point at “maximal distance”, that is, the pairs (x, y) such that $y - x$ belongs to the boundary of $t\mathscr{C}$ are γ-negligible.

Theorem C (No points move at maximal distance). *Let μ and ν be two probability measures, c_t a highly relativistic cost, $\mu \ll \mathscr{L}^n$, $t > T$ and γ an optimal plan with respect to c_t. Then*

$$\gamma\Big(\Big\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n :\, y - x \in t\,\partial\mathscr{C}\Big\}\Big) = 0\,.$$

In all our results above we use the assumption that the cost is highly relativistic: actually, we can provide a counterexample showing that for most of our claims the sole relativistic assumption is not enough.

The paper is organised as follows. In Section 2 we prove the finite chain Lemma 2.11. In Section 3 we prove Theorem C. In Section 4 we prove Theorem A, and we present an example showing that most of our claims fail without the highly relativistic assumption. Finally, in Section 5 we prove Theorem B.

2. Preliminaries and the finite chain lemma

The goal of this section is to show the “finite chain Lemma” 2.11. We start with a section containing most of our technical definitions and preliminary results.

2.1. **General definitions and introductory lemmas.** Throughout this paper, X and Y will be arbitrary Polish spaces. We will denote by $\mathcal{M}^+(X)$ the space of positive Borel measures, and by $\mathcal{P}(X)$ the subspace of probability measures; usually, μ and ν will be two measures in $\mathcal{M}^+(X)$ and $\mathcal{M}^+(Y)$ respectively, having the same total mass, i.e. $\|\mu\|_{\mathcal{M}^+(X)} = \|\nu\|_{\mathcal{M}^+(Y)}$. We will denote by $\Pi(\mu,\nu)$ the space of *transport plans* between μ and ν, that is, an element of $\Pi(\mu,\nu)$ is some measure $\gamma \in \mathcal{M}^+(X\times Y)$ whose marginals are μ and ν. Given a map $g : X \to Y$, we will denote by $g_\# : \mathcal{M}^+(X) \to \mathcal{M}^+(Y)$ the *push-forward* operator, defined by

$$\int_Y \varphi(y)\, dg_\#\mu(y) := \int_X \varphi(g(x))\, d\mu(x) \qquad\qquad \forall\, \varphi \in \mathrm{C}_b(Y)\,.$$

The projections of $X\times Y$ onto X and Y will be denoted as π_1 and π_2; for simplicity of notations, the corresponding projections from $\mathcal{M}^+(X\times Y)$ onto $\mathcal{M}^+(X)$ and $\mathcal{M}^+(Y)$ are also denoted by π_1 and π_2, instead of $\pi_{1\#}$ and $\pi_{2\#}$. In particular, $\gamma \in \Pi(\mu,\nu)$ means $\pi_1\gamma = \mu$, $\pi_2\gamma = \nu$. When α and μ are two measures such that $\alpha \ll \mu$, we denote by d_α the density of α with respect to μ, that is, $\alpha = \mathrm{d}_\alpha\mu$.

Definition 2.1 ($\alpha \le \mu$ and its relative transport plan)**.** *Given $\alpha \ll \mu \in \mathcal{M}^+(X)$, we write that $\alpha \le \mu$ if $\mathrm{d}_\alpha \le 1$, and we will denote by $\mathcal{M}^+_\mu(X)$ the set of the Borel measures $\alpha \in \mathcal{M}^+(X)$ such that $\alpha \le \mu$. If $\alpha \in \mathcal{M}^+_\mu(X)$, for any transport plan $\gamma \in \Pi(\mu,\nu)$ we denote by $\mathrm{d}_\alpha\gamma$ the transport plan defined by*

$$\int_{X\times Y} \varphi(x,y)\; d\,\mathrm{d}_\alpha\gamma(x,y) = \iint_{X\times Y} \varphi(x,y)\mathrm{d}_\alpha(x)\, d\gamma(x,y) \qquad\qquad \forall\, \varphi \in \mathrm{C}_b(X\times Y)\,.$$

Notice that the first marginal of $\mathrm{d}_\alpha\gamma$ is precisely α. Similarly, we define $\mathrm{d}^\beta\gamma$ if $\beta \le \nu$.

Definition 2.2. *For any two measures μ_1, $\mu_2 \in \mathcal{M}^+(X)$, we define $\mu_1 \wedge \mu_2$ as*

$$\mu_1 \wedge \mu_2(B) := \inf\Big\{\mu_1(B_1) + \mu_2(B_2) :\ B_1 \cap B_2 = \emptyset,\ B_1 \cup B_2 = B\Big\}\,,$$

where B, B_1 and B_2 are Borel subsets of X. Notice that, if $\mu_1 \le \mu$ and $\mu_2 \le \mu$ for some $\mu \in \mathcal{M}^+(X)$, then also $\mu_1 \wedge \mu_2 \le \mu$. In particular, one has

$$\mu_1 \wedge \mu_2 = \min\{\mathrm{d}_1, \mathrm{d}_2\}\,\mu\,. \tag{2.1}$$

Roughly speaking, when γ is a transport plan between μ and ν, for any $\alpha \le \mu$ the relative transport plan $\mathrm{d}_\alpha\gamma$ is the "portion" of the transport plan γ relative to α, that is, a smaller transport plan which transports the portion α of μ onto some portion of ν. With the next definition, we "keep track" of what is precisely the latter portion of ν.

Definition 2.3. *For any transport plan $\gamma \in \Pi(\mu,\nu)$ we define $\overrightarrow{\Phi} : \mathcal{M}^+_\mu(X) \to \mathcal{M}^+_\nu(Y)$ and $\overleftarrow{\Phi} : \mathcal{M}^+_\nu(Y) \to \mathcal{M}^+_\mu(X)$ as*

$$\overrightarrow{\Phi}(\alpha) := \pi_2(\mathrm{d}_\alpha\gamma)\,, \qquad\qquad \overleftarrow{\Phi}(\beta) := \pi_1(\mathrm{d}^\beta\gamma)\,.$$

Remark 2.4. *Note that the mappings $\overleftarrow{\Phi}$ and $\overrightarrow{\Phi}$ are non-decreasing. Namely, given $\alpha \le \tilde\alpha \in \mathcal{M}^+_\mu(X)$, the following inequality holds*

$$\overrightarrow{\Phi}(\alpha) \le \overrightarrow{\Phi}(\tilde\alpha)$$

as well as the corresponding inequality with $\overleftarrow{\Phi}$ *instead of* $\overrightarrow{\Phi}$.

Through the paper, we will make extensive use of the disintegration Theorem (see [18, 1] for a proof).

Theorem 2.5 (Disintegration Theorem). *Let* X, Y *be two Polish spaces and* $\gamma \in \mathcal{P}(X \times Y)$. *Then there exists a measurable family* $(\gamma_x)_{x\in X}$ *of probability measures on* Y *such that* $\gamma = \pi_1\gamma \otimes \gamma_x$, *namely,*

$$\int_{X\times Y} \varphi(x,y)d\gamma(x,y) = \int_X \left(\int_Y \varphi(x,y)d\gamma_x(y)\right)d\pi_1\gamma(x)$$

for all $\varphi \in \mathrm{C}_b(X \times Y)$.

Thanks to the Disintegration Theorem, we can easily compose two transport plans.

Definition 2.6 (Composition of plans). *Let* X, Y, *and* Z *be three Polish spaces, and let* $\gamma \in \mathcal{M}(X \times Z)$ *and* $\theta \in \mathcal{M}(Z \times Y)$ *be such that*

$$\pi_1\gamma = \mu\,, \qquad\qquad \pi_2\gamma = \pi_1\theta = \omega\,, \qquad\qquad \pi_2\theta = \nu\,.$$

Then, disintegrating $\gamma = \gamma_z \otimes \omega$ *and* $\theta = \omega \otimes \theta_z$, *we define the plan* $\theta \circ \gamma \in \Pi(\mu,\nu)$ *as*

$$\iint_{X\times Y} \varphi(x,y)\, d\gamma\circ\theta(x,y) := \int_Z \left(\int_{X\times Y} \varphi(x,y)\, d\gamma_z(x)d\theta_z(y)\right) d\omega(z)\,,$$

for every $\varphi \in \mathrm{C}_b(X \times Y)$.

We can now show the first properties concerning the above definitions.

Lemma 2.7. *If* $\gamma \in \Pi(\mu,\nu)$ *and* $\theta \ll \gamma$, *then* $\theta \ll \mathrm{d}_{\pi_1\theta}\gamma$ *and, analogously,* $\theta \ll \mathrm{d}^{\pi_2\theta}\gamma$.

Proof. Since characteristic functions of rectangles form a basis for the Borel sets, it is enough to show that, for any rectangle $A \times B \subseteq X \times Y$ such that $\mathrm{d}_{\pi_1\theta}\gamma(A \times B) = 0$, there holds $\theta(A \times B) = 0$.

Let us take such a rectangle and, for any $j \in \mathbb{N}$, set $A_j := \{x \in A : \mathrm{d}_{\pi_1\theta}(x) \geq 1/j\}$. Then,

$$0 = \mathrm{d}_{\pi_1\theta}\gamma(A\times B) = \iint_{A\times B} \mathrm{d}_{\pi_1\theta}(x)\, d\gamma(x,y) \geq \frac{1}{j} \iint_{A_j\times B} d\gamma(x,y) = \frac{\gamma(A_j\times B)}{j}\,,$$

which letting $j \in \mathbb{N}$ vary implies that $\gamma(A^+ \times B) = 0$, being $A^+ = \cup_{j\in\mathbb{N}} A_j = A \setminus A_0$, where $A_0 = \{x \in A : \mathrm{d}_{\pi_1\theta}(x) = 0\}$. Since $\theta \ll \gamma$, this implies that also $\theta(A^+ \times B) = 0$, so that

$$\theta(A\times B) = \theta(A_0\times B) \leq \theta(A_0 \times Y) = \pi_1\theta(A_0) = \mathrm{d}_{\pi_1\theta}\mu(A_0) = \int_{A_0} \mathrm{d}_{\pi_1\theta}(x)\, d\mu(x) = 0\,,$$

and the proof is concluded. □

In the above lemma, it is not possible to replace "$\ll$" by "$\leq$" both in the assumptions and in the conclusion, as the next example shows.

Example 2.8. *Let $\gamma = \mathcal{L}([0,1]\times[0,1])$ and $\theta = \mathcal{L}([0,a]\times[0,a])$ for some $a\in(0,1)$, so that $\theta\le\gamma$. Then, $\pi_1\theta = a\,\mathcal{L}([0,a])$, so that $\mathrm{d}_{\pi_1\theta}(x)$ equals a in $[0,a]$ and 0 in $(a,1]$; as a consequence, $\mathrm{d}_{\pi_1\theta}\gamma = a\,\mathcal{L}([0,a]\times[0,1])$, hence we have $\theta \ll \mathrm{d}_{\pi_1\theta}\gamma$, according to Lemma 2.7, but not $\theta\le \mathrm{d}_{\pi_1\theta}\gamma$. Actually, we can observe that in general, in the assumptions of Lemma 2.7, one has*

$$\|\mathrm{d}_{\pi_1\theta}\gamma\|_{X\times Y} = \|\pi_1(\mathrm{d}_{\pi_1\theta}\gamma)\|_X = \|\pi_1\theta\|_X = \|\theta\|_{X\times Y}\,,$$

hence one never has $\theta\le\mathrm{d}_{\pi_1\theta}\gamma$, unless the two measures coincide.

The next two lemmas deal with properties of the mappings $\overrightarrow{\Phi}$ and $\overleftarrow{\Phi}$; recall that these operators depend on the choice of a transport plan $\gamma\in\Pi(\mu,\nu)$, however the results in these lemmas hold for any such γ. Note that, in general, $\overrightarrow{\Phi}\big(\overleftarrow{\Phi}(\beta)\big)\neq\beta$ and $\overleftarrow{\Phi}\big(\overrightarrow{\Phi}(\alpha)\big)\neq\alpha$. Nevertheless, the following holds.

Lemma 2.9. *With the notations of Definition 2.3, if $0\neq\alpha\in\mathcal{M}^+_\mu(X)$, then $\alpha\ll\overleftarrow{\Phi}\big(\overrightarrow{\Phi}(\alpha)\big)$. In particular, $\alpha\wedge\overleftarrow{\Phi}\big(\overrightarrow{\Phi}(\alpha)\big)>0$.*

Proof. The second property follows by the first one – recall (2.1)– and in turn the first property is a direct consequence of Lemma 2.7. Indeed, setting $\theta=\mathrm{d}_\alpha\gamma\ll\gamma$ and recalling Definition 2.3, we get by Lemma 2.7 that

$$\mathrm{d}_\alpha\gamma\ll\mathrm{d}^{\pi_2(\mathrm{d}_\alpha\gamma)}\gamma=\mathrm{d}^{\overrightarrow{\Phi}(\alpha)}\gamma\,,$$

and the result follows because

$$\pi_1(\mathrm{d}_\alpha\gamma)=\alpha\,,\qquad\qquad \pi_1(\mathrm{d}^{\overrightarrow{\Phi}(\alpha)}\gamma)=\overleftarrow{\Phi}\big(\overrightarrow{\Phi}(\alpha)\big)\,.$$

□

The last property that we list here will be quite useful in the following.

Lemma 2.10. *Given $\alpha\le\mu$ and $\beta\le\nu$, one has*

$$\alpha\wedge\overleftarrow{\Phi}(\beta)>0\iff\beta\wedge\overrightarrow{\Phi}(\alpha)>0\,.$$

Proof. By symmetry, the conclusion is obtained as soon as we show that

$$\alpha\wedge\overleftarrow{\Phi}(\beta)=0\iff\gamma(A\times B)=0\,,\tag{2.2}$$

where the sets $A\subseteq X$ and $B\subseteq Y$ are defined as $A=\{x:\mathrm{d}_\alpha(x)>0\}$ and $B=\{y:\mathrm{d}^\beta(y)>0\}$.

First of all, let us assume that $\gamma(A\times B)=0$, and observe that by construction $\alpha\le\mu\llcorner A$ and $\beta\le\nu\llcorner B$, so that

$$\overleftarrow{\Phi}(\beta)\le\overleftarrow{\Phi}(\nu\llcorner B)=\pi_1(\mathrm{d}^{\nu\llcorner B}\gamma)=\pi_1\big(\gamma\llcorner(X\times B)\big)\,.$$

As a consequence, by definition

$$\begin{aligned}\alpha\wedge\overleftarrow{\Phi}(\beta)(X)&\le\Big(\mu\llcorner A\wedge\pi_1\big(\gamma\llcorner(X\times B)\big)\Big)(X)\le\mu\llcorner A(X\setminus A)+\pi_1\big(\gamma\llcorner(X\times B)\big)(A)\\&=\gamma\llcorner(X\times B)\big(A\times Y\big)=\gamma\big((X\times B)\cap(A\times Y)\big)=\gamma(A\times B)=0\,,\end{aligned}$$

by assumption. Hence, the left implication in (2.2) is proved.

On the other hand, assume that $\gamma(A\times B)>0$. Thus, there exists $\varepsilon>0$ such that $\gamma(A_\varepsilon\times B_\varepsilon)>0$, being $A_\varepsilon=\{x:\mathrm{d}_\alpha(x)>\varepsilon\}$, $B_\varepsilon=\{y:\mathrm{d}^\beta(y)>\varepsilon\}$. Therefore, we get

$$\alpha\geq\varepsilon\mu\llcorner A_\varepsilon\,,\qquad\qquad \beta\geq\varepsilon\nu\llcorner B_\varepsilon\,.$$

The first inequality yields the bound

$$\alpha\geq\varepsilon\pi_1\big(\gamma\llcorner(A_\varepsilon\times Y)\big)\geq\varepsilon\pi_1\big(\gamma\llcorner(A_\varepsilon\times B_\varepsilon)\big)\,,$$

while the second inequality implies $\mathrm{d}^\beta\gamma\geq\varepsilon\gamma\llcorner(X\times B_\varepsilon)\geq\varepsilon\gamma\llcorner(A_\varepsilon\times B_\varepsilon)$, so that

$$\overleftarrow{\Phi}(\beta)\geq\varepsilon\pi_1\big(\gamma\llcorner(A_\varepsilon\times B_\varepsilon)\big)\,.$$

The last two estimates, together, ensure that $\alpha\wedge\overleftarrow{\Phi}(\beta)>0$, so also the right implication in (2.2) is proved and the proof is concluded. □

2.2. **The Finite Chain Lemma.** This section is devoted to the proof of our main technical tool, which is a general abstract result on measures. Let us first explain its meaning.

Let γ and γ' be two transport plans between μ and ν, and assume for simplicity that all the measures are purely atomic. Assume that we can find $N+1$ pairs $(x_i,y_i)\in X\times Y$, for $0\leq i\leq N$, with the property that each (x_i,y_i) belongs to the support of γ, and each (x_i,y_{i+1}) belongs to the support of γ', with the usual convention that $y_{N+1}\equiv y_0$. We can then modify the plan γ' by making use of γ, deciding that a small quantity ε of the mass which was originally moved from every x_i to y_{i+1} should be instead moved from x_i to y_i. This modified plan is clearly better than the original one if and only if

$$\sum_{i=0}^N c(x_i,y_i)\leq\sum_{i=0}^N c(x_i,y_{i+1})\,,$$

and this is a crucial property in mass transportation, called *cyclical monotonicity*, strongly related with the optimality of plans (see for instance [19, 21, 30]). This property is actually also fundamental in the construction of the Kantorovich potential, and we will then use it in Section 5, see Definition 5.1.

Notice that the strategy discussed above makes sense only if we can find a "chain" of pairs as the (x_i,y_i) defined above, because this is the only way to guarantee that the modified measure is indeed a transport plan. Actually, the only problem is in "closing" the chain: we can arbitrarily choose (x_0,y_0) in the support of γ; this implies that x_0 is in the support of μ, thus we can find $y_1\in Y$ such that (x_0,y_1) is in the support of γ' (to do this formally, we make use of the map $\overrightarrow{\Phi}$ of Definition 2.3). In turn, this ensures that y_1 is in the support of ν, hence (with the map $\overleftarrow{\Phi}$) we find some $x_1\in X$ such that (x_1,y_1) is in the support of γ, and so on. This procedure will be concluded as soon as some x_N has the property that (x_N,y_0) is in the support of γ', therefore everything works if we can prove that this will happen sooner or later. This is precisely the goal of our "chain Lemma" below. Actually, we do not make the assumption that the measures are purely atomic, so our procedure is technically a bit more complicated than what we just described, but the underlying idea is the same.

Lemma 2.11 (finite chain Lemma)**.** *Let γ, $\gamma' \in \Pi(\mu,\nu)$, let $\gamma_0 \leq \gamma$ be a non zero measure, and define $\mu_0 = \pi_1\gamma_0 \leq \mu$ and $\nu_0 = \pi_2\gamma_0 \leq \nu$ the marginals of γ_0. Define recursively the measures*

$$\nu_{i+1} = \overrightarrow{\Phi}'(\mu_i)\,, \qquad\qquad \mu_{i+1} = \overleftarrow{\Phi}(\nu_{i+1})\,,$$

where $\overleftarrow{\Phi}$ and $\overrightarrow{\Phi}'$ are the functions of Definition 2.3 relative to γ and to γ' respectively. Then, there exists $j > 0$ such that $\nu_j \wedge \nu_0 > 0$.

Proof. Let us start by introducing the sets (notice the slight differences!)

$$\mathcal{B} := \Big\{\tilde\nu \leq \nu :\ \text{if } \exists\, i > 0,\ \nu_i \wedge \tilde\nu > 0, \text{ then } \exists\, j \geq i :\ \nu_j \wedge \nu_0 > 0\Big\}\,,$$
$$\mathcal{A} := \Big\{\tilde\mu \leq \mu :\ \text{if } \exists\, i \geq 0,\ \mu_i \wedge \tilde\mu > 0, \text{ then } \exists\, j > i :\ \nu_j \wedge \nu_0 > 0\Big\}\,.$$

Notice that both sets are non-empty since $\mathcal{B}$ contains $\tilde\nu = \nu_0$, while $\mathcal{A}$ contains $\tilde\mu = 0$. Set now

$$\beta := \sup\{\tilde\nu :\ \tilde\nu \in \mathcal{B}\}\,, \qquad\qquad \alpha := \sup\{\tilde\mu :\ \tilde\mu \in \mathcal{A}\}\,.$$

Note that both suprema are actually maxima, namely $\beta \in \mathcal{B}$ and $\alpha \in \mathcal{A}$. To show the first property, assume that for some $i > 0$ one has $\nu_i \wedge \beta > 0$: by definition of sup of measures, this means that there exists some $\tilde\nu \in \mathcal{B}$ such that $\nu_i \wedge \tilde\nu > 0$; hence, by definition of $\mathcal{B}$ there is some $j \geq i$ such that $\nu_j \wedge \nu_0 > 0$, and finally this means exactly $\beta \in \mathcal{B}$. The proof that $\alpha \in \mathcal{A}$ is identical.

Assume that, for some $i \geq 0$, $\mu_i \wedge \overleftarrow{\Phi}'(\beta) > 0$; hence, Lemma 2.10 implies that $\beta \wedge \overrightarrow{\Phi}'(\mu_i) > 0$, that is, $\beta \wedge \nu_{i+1} > 0$. Since $\beta \in \mathcal{B}$, this implies the existence of some $j \geq i+1$ such that $\nu_j \wedge \nu_0 > 0$, and in turn this establishes that $\overleftarrow{\Phi}'(\beta) \in \mathcal{A}$, so $\overleftarrow{\Phi}'(\beta) \leq \alpha$. The very same argument gives that $\overrightarrow{\Phi}(\alpha) \leq \beta$, thus we obtain $\overrightarrow{\Phi}\big(\overleftarrow{\Phi}'(\beta)\big) \leq \beta$. But the maps $\overrightarrow{\Phi}$ and $\overleftarrow{\Phi}'$ do not change the total mass, so the last inequality must be an equality, and this implies

$$\alpha = \overleftarrow{\Phi}'(\beta)\,, \qquad\qquad \beta = \overrightarrow{\Phi}(\alpha)\,,$$

hence in particular

$$\|\alpha\| = \|\beta\|\,. \tag{2.3}$$

Let us now recall that $\alpha \leq \mu$ and $\beta \leq \nu$, so we have $\alpha = \mathrm{d}_\alpha\mu$ and $\beta = \mathrm{d}^\beta\nu$; if we now define

$$A := \{x \in X :\ \mathrm{d}_\alpha(x) > 0\}\,, \qquad B := \{y \in Y :\ \mathrm{d}^\beta(y) > 0\}\,, \qquad \overline{\mu} := \mu \llcorner A\,, \qquad \overline{\nu} := \nu \llcorner B\,,$$

we have that for any $i \geq 0$

$$\mu_i \wedge \overline{\mu} > 0 \Longleftrightarrow \mu_i \wedge \alpha > 0\,,$$

and since $\alpha \in \mathcal{A}$ we deduce that also $\overline{\mu} \in \mathcal{A}$, hence $\overline{\mu} \leq \alpha$. On the other hand, $\alpha \leq \overline{\mu}$ by definition, hence $\alpha = \overline{\mu}$ and then $\mathrm{d}_\alpha \equiv 1$ on A; similarly, $\beta = \overline{\nu}$ and $\mathrm{d}^\beta \equiv 1$ on B. Observe that

$$\nu \llcorner B = \beta = \overrightarrow{\Phi}(\alpha) = \pi_2(\mathrm{d}_\alpha\gamma) = \pi_2\big(\gamma \llcorner (A \times Y)\big)\,,$$

which implies

$$\gamma\big(A \times (Y \setminus B)\big) = \pi_2\big(\gamma \llcorner (A \times Y)\big)(Y \setminus B) = \nu \llcorner B(Y \setminus B) = 0\,.$$

Then, from (2.3) we readily deduce

$$\nu(B) = \|\beta\| = \|\alpha\| = \mu(A) = \gamma(A \times Y) = \gamma(A \times B) \leq \gamma(X \times B) = \nu(B)\,,$$

which also implies that

$$\gamma\big((X \setminus A) \times B\big) = 0\,.$$

We can then deduce that

$$\overleftarrow{\Phi}(\beta) = \pi_1(\mathrm{d}^\beta \gamma) = \pi_1\big(\gamma \llcorner (X \times B)\big) = \pi_1\big(\gamma \llcorner (A \times B)\big) \leq \pi_1\big(\gamma \llcorner (A \times Y)\big) = \mu \llcorner A = \alpha\,,$$

which again since $\|\alpha\| = \|\beta\| = \|\overleftarrow{\Phi}(\beta)\|$ allows us to infer

$$\alpha = \overleftarrow{\Phi}(\beta)\,. \tag{2.4}$$

Notice that previously we had found that $\alpha = \overleftarrow{\Phi}'(\beta)$. Now, since of course $\nu_0 \in \mathcal{B}$, then $\nu_0 \leq \beta$ and so by (2.4) $\overleftarrow{\Phi}(\nu_0) \leq \overleftarrow{\Phi}(\beta) = \alpha$. Moreover, $\gamma_0 \ll \gamma$ and then Lemma 2.7 ensures that $\gamma_0 \ll \mathrm{d}_{\pi_2\gamma_0}\gamma$, so we get

$$\mu_0 = \pi_1\gamma_0 \ll \pi_1(\mathrm{d}^{\pi_2\gamma_0}\gamma) = \pi_1(\mathrm{d}^{\nu_0}\gamma) = \overleftarrow{\Phi}(\nu_0) \leq \alpha\,,$$

which implies $\mu_0 \wedge \alpha > 0$. Finally, since $\alpha \in \mathcal{A}$, the last estimate implies the existence of some $j > 0$ for which $\nu_j \wedge \nu_0 > 0$, which is the thesis. □

We can use Lemma 2.11 to make the following decomposition which is the formal way to define the modification that we described at the beginning of the section.

Proposition 2.12. *Let γ, $\gamma' \in \Pi(\mu,\nu)$, $0 \neq \gamma_0 \leq \gamma$, 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*

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

In particular, $\tilde\gamma$ can be decomposed as $\tilde\gamma = \tilde\gamma_0 + \tilde\gamma_\infty$, where

$$\tilde\gamma_0 \leq \gamma_0\,, \qquad \tilde\gamma_\infty \leq \gamma - \gamma_0\,, \qquad \pi_1\tilde\gamma_0 = \mu_A\,, \qquad \pi_2\tilde\gamma_0 = \nu_A\,, \qquad \pi_1\tilde\gamma_\infty = \tilde\mu\,, \qquad \pi_2\tilde\gamma_\infty = \tilde\nu\,.$$

Proof. Let us use the same notations as in the proof of Lemma 2.11, and let us define N as the smallest integer such that $\bar\varepsilon := \|\nu_{N+1} \wedge \nu_0\| > 0$. For any $\varepsilon \leq \bar\varepsilon$, we fix arbitrarily a measure $\tilde\nu_{N+1} \leq \nu_{N+1} \wedge \nu_0$ such that $\|\tilde\nu_{N+1}\| = \varepsilon$. Now, for any $0 < i \leq N$, let us call $\gamma_i = \mathrm{d}^{\nu_i}\gamma$, so that

$$\pi_1\gamma_i = \pi_1\mathrm{d}^{\nu_i}\gamma = \overleftarrow{\Phi}(\nu_i) = \mu_i\,, \qquad\qquad \pi_2\gamma_i = \nu_i\,. \tag{2.5}$$

Analogously, for any $0 \leq i \leq N$ we let $\gamma'_i = \mathrm{d}_{\mu_i}\gamma'$, and as before

$$\pi_1\gamma'_i = \mu_i\,, \qquad\qquad \pi_2\gamma'_i = \pi_2\mathrm{d}_{\mu_i}\gamma' = \overrightarrow{\Phi}'(\mu_i) = \nu_{i+1}\,,.$$

Observe that γ_0 is already defined, the result of (2.5) holds for $i = 0$ however. We claim now that

$$\gamma_0 + \gamma_1 + \cdots + \gamma_N \leq \gamma\,, \qquad\qquad \gamma'_0 + \gamma'_1 + \cdots + \gamma'_N \leq \gamma'\,. \tag{2.6}$$

If $N = 0$, there is nothing to prove. Otherwise, by definition of N we know that $\nu_0 \wedge \nu_1 = 0$, which implies a fortiori that $\gamma_0 \wedge \gamma_1 = 0$, thus $\gamma_0 + \gamma_1 \leq \gamma$. As a consequence, $\mu \geq \mu_0 + \mu_1$, and this gives by definition that $\gamma' \geq \gamma'_0 + \gamma'_1$, which by taking the second marginals ensures

$\nu \geq \nu_1 + \nu_2$. If $N = 1$, we have then concluded (2.6). If $N > 1$, we have also that $\nu_0 \wedge \nu_2 = 0$, and exactly as before this implies first that $\gamma_0 \wedge (\gamma_1 + \gamma_2) = 0$, then that $\gamma_0 + \gamma_1 + \gamma_2 \leq \gamma$, then that $\mu \geq \mu_0 + \mu_1 + \mu_2$, and finally that $\gamma'_0 + \gamma'_1 + \gamma'_2 \leq \gamma'$. This proves (2.6) if $N = 2$, and with an obvious induction argument we actually derive the validity of (2.6) whatever N is.

We can now easily conclude the proof. Since $\tilde\nu_{N+1} \leq \nu_{N+1} = \pi_2\gamma'_N$, we define $\tilde\gamma'_N = \mathrm{d}^{\tilde\nu_{N+1}}\gamma'_N$, and by construction $\tilde\gamma'_N \leq \gamma'_N$ and $\pi_2\tilde\gamma'_N = \tilde\nu_{N+1}$. We set then $\tilde\mu_N = \pi_1\tilde\gamma'_N$, and notice that $\tilde\mu_N \leq \pi_1\gamma'_N = \mu_N$. Recalling then that $\mu_N = \pi_1\gamma_N$, we define $\tilde\gamma_N = \mathrm{d}_{\tilde\mu_N}\gamma_N \leq \gamma_N$, and $\tilde\nu_N = \pi_2\tilde\gamma_N \leq \nu_N$. We can now define $\tilde\gamma'_{N-1} = \mathrm{d}^{\tilde\nu_N}\gamma'_{N-1}$ and continue this construction backward in the obvious way. In the end, for every $i \in \{0, 1, \cdots, N\}$ we have built measures $\tilde\gamma_i$, $\tilde\gamma'_i$, $\tilde\mu_i$ and $\tilde\nu_i$, with the properties that

$$\tilde\mu_i = \pi_1\tilde\gamma'_i = \pi_1\tilde\gamma_i\,, \qquad\qquad \tilde\nu_i = \pi_2\tilde\gamma'_{i-1} = \pi_2\tilde\gamma_i\,.$$

We define now

$$\tilde\gamma = \sum_{i=0}^N \tilde\gamma_i\,, \quad \tilde\gamma' = \sum_{i=0}^N \tilde\gamma'_i\,, \quad \tilde\nu = \sum_{i=1}^N \tilde\nu_i\,, \quad \tilde\mu = \sum_{i=1}^N \tilde\mu_i\,, \quad \mu_A = \tilde\mu_0\,, \quad \nu_A = \tilde\nu_0\,, \quad \nu_B = \tilde\nu_{N+1}\,.$$

Observe that $\tilde\gamma \leq \gamma$ and $\tilde\gamma' \leq \gamma'$ thanks to (2.6): then, by setting $\tilde\gamma_\infty = \sum_{i=1}^N \tilde\gamma_i$, it is immediate to check all the properties and the proof is concluded. □

3. Proof of Theorem C

This section is devoted to show our first main result, Theorem C, which says that for supercritical speeds only a negligible set of points move at maximal speed. We will show this under the assumption that the cost is highly relativistic (see Definition 3.2) and that $\mu \ll \mathscr{L}^n$. Example 4.2 in the next section will show that both these assumptions are sharp. After that, we will give a precise quantitative estimate of how many points are moved at almost maximal speed, in Proposition 3.3.

First of all, we give the formal definition of the (highly) relativistic cost functions, together with some notation.

Definition 3.1 (Relativistic cost function). *Let $\mathscr{C}$ be a strictly convex closed subset of $\mathbb{R}^n$, containing the origin in its interior, and let $h : \mathbb{R}^n \to [0, +\infty]$ be a strictly convex function with $h(0) = 0$, continuous and bounded in $\mathscr{C}$, and identically $+\infty$ in $\mathbb{R}^n \setminus \mathscr{C}$. Then, for every $t > 0$ we call* relativistic cost function *the function*

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

Definition 3.2 (Highly relativistic cost function). *Let c_t be a relativistic cost function. For every η, $\rho \ll 1$ we define*

$$\kappa(\eta, \rho) := \inf\left\{\frac{h(v) - h(v - \lambda w)}{\lambda} : v \in \partial\mathscr{C},\, |w - v| \leq \rho,\, 0 < \lambda < \eta\right\}. \tag{3.1}$$

We say that c_t *is a* highly relativistic cost function *if, for* ρ *small enough* [1], *one has* $\kappa(\eta,\rho)\to\infty$ *for* $\eta\searrow 0$. *Notice that this only depends on* h, *and not on* t.

Notice that the above definition has the following meaning: we are asking that the (radial) derivative of h explode at $\partial\mathscr{C}$, and we write this by introducing the notation $\kappa(\eta,r)$ only in order not to require differentiability of the function h. For instance, Brenier's relativistic heat cost (1.1) satisfies this assumption. We are now able to prove Theorem C. Let us first briefly explain our strategy. The idea is to take an optimal plan γ relative to the cost c_t and assume that a positive quantity of mass moves exactly of maximal distance; for simplicity, let us think that there is a small cube $Q_{\overline{x}}$ of side r and mass ε which is translated by γ by a vector $v\in t\partial C$ onto another cube $Q_{\overline{y}}$; let us also think that μ has constant density on this cube. We can now take the right half of $Q_{\overline{x}}$ and decide to move it to the left half of $Q_{\overline{y}}$: in this way, the points are not moving of a distance $|v|$, but only of a distance about $|v|-r/2$. Thanks to the fact that c_t is a *highly* relativistic cost, the gain is much bigger than εr (and then in the end it appears evident why the Theorem fails for just relativistic costs). Now, we cannot send the left part of $Q_{\overline{x}}$ onto the right part of $Q_{\overline{y}}$, because the points would move of a distance $|v|+r/2$, and this would have an infinite cost. Here the assumption $t>T$ comes into play: indeed, we can take a plan γ' relative to some $t'\in(T,t)$, and use it to modify γ according to Proposition 2.12. Since $t'<t$, the modification given by γ' has finite cost for $\mathcal{C}_t$ even after a change of order $r\ll t-t'$, and then this gives an admissible plan which pays something of order $N\varepsilon$ for the modification. Since the loss $N\varepsilon$ is much smaller than the gain that we had before, our modified transport plan has a cost strictly smaller than γ, and the contradiction will end the proof. Note however that the actual proof is more complicated because of all the simplifications we assume in the above discussion.

Proof (of Theorem C). To keep the proof simple to read, we split it in four steps.
Step I. Setting of the construction and definitions.
Let $t>T$, and assume that there exists an optimal plan γ for which

$$\gamma\Big(\Big\{(x,y):\,y-x\in t\partial\mathscr{C}\Big\}\Big)>0\,.$$

Let then $(\overline{x},\overline{y})$ be a Lebesgue point (with respect to γ) of the set $\{y-x\in t\partial\mathscr{C}\}$, and let $r\ll 1$ to be specified during the proof. Without loss of generality, and up to rescaling, we can assume that $\overline{y}-\overline{x}=\mathrm{e}_1$; let then $Q_{\overline{x}}$ and $Q_{\overline{y}}$ be two cubes, centered at $\overline{x}$ and $\overline{y}$ respectively, with sides parallel to the coordinate axes and of length r. Since $(\overline{x},\overline{y})$ is a γ-Lebesgue point of the pairs $\{(x,y):y-x\in t\mathscr{C}\}$, and since $\mu\ll\mathscr{L}^n$, there is a big constant K such that the measure

$$\gamma_0:=\gamma\llcorner\Big\{(x,y)\in Q_{\overline{x}}\times Q_{\overline{y}}:\,y-x\in t\partial\mathscr{C},\,\rho(x)\leq K\Big\}$$

is a non-zero measure, being ρ the density of μ with respect to the Lebesgue measure. As usual, we call $\mu_0\leq\mu$ and $\nu_0\leq\nu$ the marginals of γ_0.

Let us denote the generic point of $\mathbb{R}^n$ as $x=(\sigma,\tau)\in\mathbb{R}\times\mathbb{R}^{n-1}$, and call $\pi_\tau:\mathbb{R}^n\to\mathbb{R}^{n-1}$ the projection on the τ variable; for simplicity of notation, let us assume that the σ-coordinates

[1] so that $\kappa(\eta,\rho)>-\infty$

of the points of $Q_{\overline{x}}$ are between 0 and r. We disintegrate now $\mu_0 = \mu^\tau \otimes \alpha$, being $\alpha = \pi_{\tau\#}\mu_0$; recall that each μ^τ is a probability measure concentrated in the segment $[0,r]$, and that by Fubini Theorem α-a.e. measure μ^τ is absolutely continuous with respect to $\mathscr{L}^1$. Let us call now $m = \|\mu_0\|$ and let $\delta = \delta(r) \ll 1$ be a small quantity, to be specified later; for α-a.e. τ, we can define the function

$$f_\tau : \Big\{\sigma \in [0,r] : \mu^\tau\big([0,\sigma]\big) \leq 1-\delta\Big\} \longrightarrow \Big\{\sigma \in [0,r] : \mu^\tau\big([0,\sigma]\big) \geq \delta\Big\}$$

as $f_\tau(\sigma) = \sigma'$ where

$$\mu^\tau\big([0,\sigma']\big) = \mu^\tau\big([0,\sigma]\big) + \delta\,.$$

In words, this means that we are taking all the points of the segment $\{(\sigma,\tau) : 0 \leq \sigma \leq r\}$ except the last portion δ on the right, and we are calling f_τ the map associating to each σ a point $f_\tau(\sigma)$ "after" a portion δ of mass. In average, we should expect points to be translated by f_τ of a quantity δr to the right: of course this would be true only if μ^τ were constant, and actually all what we can say is that every point is moved to the right because $\mu^\tau\big([0,\sigma']\big) > \mu^\tau\big([0,\sigma]\big)$ implies $\sigma' > \sigma$; nevertheless, it is useful to keep in mind this "average" idea. In fact, let us fix a big constant $M = M(r)$, to be specified later, and let us set

$$Z := \left\{\tau : \exists\,\sigma \in [0,r] : f_\tau(\sigma) < \sigma + \frac{\delta r}{M}\right\}.$$

So, the elements τ of Z are those for which at least some point σ of the horizontal segment moves only very little to the right compared with the "average". We can expect Z to contain only few points if M is big enough, and this is precisely the content of next step.

Step II. *One has $\alpha(Z) \leq m/3$.*

By the Measurable Selection Theorem (see for instance [17, 18]), there exists a measurable function $\tau \mapsto \sigma(\tau)$ on Z such that, for any $\tau \in Z$,

$$f_\tau\big(\sigma(\tau)\big) - \sigma(\tau) < \frac{\delta r}{M}\,.$$

Define then the box

$$\Gamma := \left\{(\sigma,\tau) \in Q_{\overline{x}} : \tau \in Z,\, \sigma(\tau) \leq \sigma \leq \sigma(\tau) + \frac{\delta r}{M}\right\}.$$

Then, on the one hand we have

$$\mu_0(\Gamma) \leq K\mathscr{L}^n(\Gamma) = K\,\frac{\delta r}{M}\,\mathscr{L}^{n-1}(Z) \leq \frac{K\delta r^n}{M}\,,$$

and on the other hand

$$\mu_0(\Gamma) = \int_Z \bigg(\int_{\sigma(\tau)}^{\sigma(\tau)+\frac{\delta r}{M}} 1\, d\mu^\tau(\sigma)\bigg)\, d\alpha(\tau) \geq \int_Z \bigg(\int_{\sigma(\tau)}^{f_\tau(\sigma(\tau))} 1\, d\mu^\tau(\sigma)\bigg)\, d\alpha(\tau) = \delta\alpha(Z)\,.$$

Putting together the last two estimates, it is clear that the claim holds true if we choose, for instance,

$$M = \frac{3Kr^n}{m}\,. \tag{3.2}$$

Note that such M depends on r but not on δ.

Step III. *The measures ξ_1 and ξ_2.*
Let us start to build the competitor $\xi \in \Pi(\mu,\nu)$ for γ. First of all, we set $W = \{\tau \notin Z\}$ and

$$L := \Big\{(\sigma,\tau) :\, \tau \in W,\, \mu^\tau\big([0,\sigma]\big) \leq 1-\delta\Big\}, \qquad R := \Big\{(\sigma,\tau) :\, \tau \in W,\, \mu^\tau\big([0,\sigma]\big) \geq \delta\Big\}.$$

We define now the functions $\tilde g : L \to Q_{\overline{x}}$ and $g : L \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ as

$$\tilde g(\sigma,\tau) := \big(f_\tau(\sigma),\tau\big)\,, \qquad\qquad g(x,y) := \big(\tilde g(x), y\big)\,.$$

Notice that $\tilde g$ is a μ_0-essentially bijective function from L to R. We can now define the first two pieces of ξ, namely,

$$\xi_1 := \gamma_0 \llcorner \Big\{(x,y) \in Q_{\overline{x}} \times Q_{\overline{y}} :\, x \in [0,r]\times Z\Big\}, \qquad \xi_2 := g_\#\Big(\gamma_0 \llcorner \big(L \times \mathbb{R}^n\big)\Big)\,.$$

The meaning of the last two definitions is the following: for pairs $(x,y) \in Q_{\overline{x}} \times Q_{\overline{y}}$ for which $x = (\sigma,\tau)$ and $\tau \in Z$, we decide not to do any change, so ξ_1 equals γ_0 for those points. For the points for which $\tau \in W$, instead, we want to shorten the path: ξ_2 takes the "right points" which belong to R, and move them as the corresponding "left points" in L were doing; it is easy to imagine that this will shorten the paths, and that in the end we will remain with a very left part of the cube $Q_{\overline{x}}$ which must go onto a very right part of the cube $Q_{\overline{y}}$: we will take care of this in the next step with ξ_3 and ξ_4. Let us now calculate the projections of ξ_2 (since the projections of ξ_1 are obvious). For the first one, for any Borel set $A \subseteq \mathbb{R}^n$ we have

$$\begin{aligned}
\pi_1\xi_2(A) &= \xi_2(A\times\mathbb{R}^n) = \gamma_0\Big((L\times\mathbb{R}^n)\cap g^{-1}(A\times\mathbb{R}^n)\Big) = \gamma_0\big(g^{-1}(A\times\mathbb{R}^n)\big) = \gamma_0\big(\tilde g^{-1}(A)\times\mathbb{R}^n\big)\\
&= \mu_0\big(\tilde g^{-1}(A)\big) = \iint_{(\sigma,\tau)\in Q_{\overline{x}}} \chi_{\tilde g^{-1}(A)}(\sigma,\tau)\,d\mu_0(\tau,\sigma)\\
&= \int_{\tau\in W}\int_{\sigma\in(0,r)} \chi_{\tilde g^{-1}(A)}(\sigma,\tau)\,d\mu^\tau(\sigma)\,d\alpha(\tau) = \int_{\tau\in W} \mu^\tau\big(\big\{\sigma :\, (\sigma,\tau)\in\tilde g^{-1}(A)\big\}\big)\,d\alpha(\tau)\,.
\end{aligned}$$

Now we observe that, by definition of f_τ and of g, for any τ one has

$$\mu^\tau\big(\big\{\sigma :\, (\sigma,\tau)\in\tilde g^{-1}(A)\big\}\big) = \mu^\tau\big(\big\{\sigma :\, (\sigma,\tau)\in A\cap R\big\}\big)\,.$$

Hence, from last estimate we finally get

$$\pi_1\xi_2(A) = \int_{\tau\in W}\mu^\tau\big(\big\{\sigma :\, (\sigma,\tau)\in A\cap R\big\}\big)\,d\alpha(\tau) = \mu_0(A\cap R)\,.$$

In the very same way, using the fact the the second component of g is the identity, one obtains that

$$\pi_2\xi_2(B) = \gamma_0(L\times B)\,.$$

Since ξ_1 coincides with the restriction of γ_0 to the points (σ,τ) with $\tau\in Z$, we simply have

$$\pi_1\big(\xi_1+\xi_2\big) = \mu_0 - \mu_{\rm rem}\,, \qquad\qquad \pi_2\big(\xi_1+\xi_2\big) = \nu_0 - \nu_{\rm rem}\,, \tag{3.3}$$

where the "remaining measures" are

$$\begin{aligned}
\mu_{\rm rem} &= \mu_0\llcorner\Big(\big(Q_{\overline{x}}\setminus R\big)\cap\big\{(\sigma,\tau) :\, \tau\in W\big\}\Big)\,,\\
\nu_{\rm rem} &= \pi_2\bigg(\gamma_0\llcorner\Big[\Big(\big(Q_{\overline{x}}\setminus L\big)\cap\big\{(\sigma,\tau) :\, \tau\in W\big\}\Big)\times\mathbb{R}^n\Big]\bigg)\,.
\end{aligned}$$

Let us call $\varepsilon := \|\mu_{\mathrm{rem}}\| = \|\nu_{\mathrm{rem}}\|$, and notice that by Step II, we have

$$\varepsilon = \delta\alpha(W) \in \left[\frac{2}{3}\,m\delta, m\delta\right]. \tag{3.4}$$

To conclude this step, we need to evaluate the costs of ξ_1 and ξ_2. To do so, take any $x = (\sigma,\tau) \in L$, and let $y \in Q_{\overline{y}}$ be a point such that (x,y) is in the support of γ_0. Call now $v = (y-x)/t \in \partial\mathscr{C}$ and $w = \mathrm{e}_1/t$, and notice that by construction $\tilde{g}(x) - x = \lambda wt$ for some

$$\frac{\delta r}{M} \leq \lambda \leq r\,,$$

where the second inequality comes from geometric arguments, while the first one is true because $x \in L$ implies $\tau \in W$; observe also that $|w - v| \leq \sqrt{n}r/t$ (keep in mind that $|\overline{y} - \overline{x}| = 1$). Notice also that, since h is strictly convex, we readily obtain that the function $\lambda \mapsto h(v - \lambda w)$ is decreasing for $\lambda \in (\delta r/M, r)$ as soon as r is small enough. We can then evaluate

$$\begin{aligned} c_t\big(g(x,y)\big) - c_t(x,y) &= c_t\big(\tilde{g}(x),y\big) - c_t(x,y) = h\bigg(\frac{y - \tilde{g}(x)}{t}\bigg) - h\bigg(\frac{y-x}{t}\bigg) \\ &= h(v - \lambda w) - h(v) \leq h\Big(v - \frac{\delta r}{M}\,w\Big) - h(v) \leq -\frac{\delta r}{M}\,\kappa\bigg(\frac{\delta r}{M}, \frac{\sqrt{n}r}{t}\bigg). \end{aligned} \tag{3.5}$$

As a consequence, the cost of $\xi_1 + \xi_2$ is

$$\begin{aligned} \mathcal{C}_t(\xi_1+\xi_2) &= \iint_{\mathbb{R}^n\times\mathbb{R}^n} c_t(x,y)\,d\xi_1(x,y) + \iint_{\mathbb{R}^n\times\mathbb{R}^n} c_t(x,y)\,d\xi_2(x,y) \\ &= \int_{[0,r]\times Z\times Q_{\overline{y}}} c_t(x,y)\,d\gamma_0(x,y) + \iint_{L\times\mathbb{R}^n} c_t\big(g(x,y)\big)\,d\gamma_0(x,y) \\ &\leq \int_{[0,r]\times Z\times Q_{\overline{y}}} c_t(x,y)\,d\gamma_0(x,y) + \iint_{L\times\mathbb{R}^n} c_t(x,y) - \frac{\delta r}{M}\,\kappa\bigg(\frac{\delta r}{M}, \frac{\sqrt{n}r}{t}\bigg)\,d\gamma_0(x,y) \\ &\leq \mathcal{C}_t(\gamma_0) - \frac{\delta r}{M}\,\kappa\bigg(\frac{\delta r}{M}, \frac{\sqrt{n}r}{t}\bigg)\gamma_0(L\times\mathbb{R}^n) = \mathcal{C}_t(\gamma_0) - \frac{\delta r}{M}\,\kappa\bigg(\frac{\delta r}{M}, \frac{\sqrt{n}r}{t}\bigg)\mu_0(L) \\ &= \mathcal{C}_t(\gamma_0) - \frac{\delta r}{M}\,\kappa\bigg(\frac{\delta r}{M}, \frac{\sqrt{n}r}{t}\bigg)(1-\delta)\alpha(W) = \mathcal{C}_t(\gamma_0) - \frac{\varepsilon r}{2M}\,\kappa\bigg(\frac{\delta r}{M}, \frac{\sqrt{n}r}{t}\bigg), \end{aligned} \tag{3.6}$$

where we have used also (3.4) and the fact that $\delta \leq 1/2$.

Step IV. *The measures ξ_3 and ξ_4.*

We now present that last two pieces of the competitor $\xi \in \Pi(\mu,\nu)$. Let us start by recalling that by (3.3) and (3.4) the marginals of the measure $\xi_1 + \xi_2$ cover almost all μ_0 and ν_0, and then it would be easy to conclude the competitor simply by adding the whole $\gamma - \gamma_0$ to a plan transporting μ_{rem} onto ν_{rem}. Unfortunately, this would have an infinite cost, because points of μ_{rem} and point of ν_{rem} have distance more or less $\mathrm{e}_1 + r$, which is outside $t\mathscr{C}$. To overcome this problem, we use the assumption that $t > T$, then we select some $T < t' < t$ and a transport plan $\gamma' \in \Pi(\mu,\nu)$ which has finite $\mathcal{C}_{t'}$ cost (we could take, for instance, some optimal plan for $\mathcal{C}_{t'}$, but the finiteness of the cost is enough). Then, we apply Proposition 2.12 and we get some $N \in \mathbb{N}$ and $\overline{\varepsilon} > 0$, depending on γ, γ' and γ_0, also on r, but not on δ; if $\varepsilon \leq \overline{\varepsilon}$, then we get also plans $\tilde{\gamma}$ and $\tilde{\gamma}'$ satisfying the claim of the proposition. But actually, by (3.4) we can assume $\varepsilon \leq \overline{\varepsilon}$ as soon as we choose δ small enough; thus, Proposition 2.12 provides us with two plans

$\tilde\gamma = \tilde\gamma_0 + \tilde\gamma_\infty$ and $\tilde\gamma'$, being $\tilde\gamma_0 \le \gamma_0$ and $\tilde\gamma_\infty \le \gamma - \gamma_0$. We can then define $\xi_3 = \gamma - \gamma_0 - \tilde\gamma_\infty$, which is a positive measure whose marginals are

$$\pi_1\xi_3 = \mu - \mu_0 - \tilde\mu\,, \qquad\qquad \pi_2\xi_3 = \nu - \nu_0 - \tilde\nu\,,$$

so that by (3.3) we have now

$$\pi_1(\xi_1+\xi_2+\xi_3) = \mu - \mu_{\rm rem} - \tilde\mu\,, \qquad\qquad \pi_2(\xi_1+\xi_2+\xi_3) = \nu - \nu_{\rm rem} - \tilde\nu\,. \tag{3.7}$$

Since $\xi = \xi_1+\xi_2+\xi_3+\xi_4$ must be in $\Pi(\mu,\nu)$, we need then the marginals of ξ_4 to be $\tilde\mu + \mu_{\rm rem}$ and $\tilde\nu + \nu_{\rm rem}$. Since the marginals of $\tilde\gamma'$ are $\tilde\mu + \mu_A$ and $\tilde\nu + \nu_B$, the plan $\tilde\gamma'$ does the job almost perfectly, the only mistake being the presence of μ_A and ν_B instead of $\mu_{\rm rem}$ and $\nu_{\rm rem}$. However, notice that both μ_A and $\mu_{\rm rem}$ are measures smaller than μ_0 and of mass ε, and similarly both ν_B and $\nu_{\rm rem}$ are measures smaller than ν_0 and of mass ε. Let us then introduce two auxiliary transport plans, namely

$$\alpha = ({\rm Id},{\rm Id})_\#\tilde\mu + \mu_{\rm rem}\otimes\mu_A\,, \qquad\qquad \beta = ({\rm Id},{\rm Id})_\#\tilde\nu + \nu_B\otimes\nu_{\rm rem}\,,$$

whose marginals are precisely $\tilde\mu + \mu_{\rm rem}$ and $\tilde\mu + \mu_A$, and $\tilde\nu + \nu_B$ and $\tilde\nu + \nu_{\rm rem}$ respectively. Let us finally define $\xi_4 = \beta\circ\tilde\gamma'\circ\alpha$, so that

$$\pi_1\xi_4 = \tilde\mu + \mu_{\rm rem}\,, \qquad\qquad \pi_2\xi_4 = \tilde\nu + \nu_{\rm rem}\,,$$

so that by (3.7) we finally get that $\xi \in \Pi(\mu,\nu)$, so ξ is an admissible competitor. To conclude, we need to estimate the costs of ξ_3 and ξ_4. While for ξ_3 we have

$$\mathcal{C}_t(\xi_3) = \mathcal{C}_t(\gamma) - \mathcal{C}_t(\gamma_0) - \mathcal{C}_t(\tilde\gamma_\infty) \le \mathcal{C}_t(\gamma) - \mathcal{C}_t(\gamma_0)\,, \tag{3.8}$$

concerning ξ_4 we must observe what follows: since $\tilde\gamma' \le \gamma'$, and γ' has a finite $\mathcal{C}_{t'}$ cost, we know that $y - x \in t'\mathscr{C}$ for γ'-a.e. (x,y). Hence, since μ_A and $\mu_{\rm rem}$ are both supported in $Q_{\overline{x}}$ and ν_B and $\nu_{\rm rem}$ are both supported in $Q_{\overline{y}}$, we derive that α and β move points at most of a distance $\sqrt{n}r$, and then for ξ_4-a.e. (x,y) one has $y - x \in t'\mathscr{C} + 2\sqrt{n}r$. Since we are free to chose r as small as we wish, and $t' < t$, it is admissible to assume that $t'\mathscr{C} + 2\sqrt{n}r \subseteq t\mathscr{C}$, thus ξ_4 has a finite $\mathcal{C}_t$ cost (and maybe an infinite $\mathcal{C}_{t'}$ cost, but this is not a problem); notice that formally we are *first* choosing t', *then* r, *then* M, and *only at the end* δ. Recalling that h is bounded where it is finite, and calling C the maximum of h in $\mathscr{C}$, we get then

$$\mathcal{C}_t(\xi_4) \le C\|\xi_4\| = C(N+1)\varepsilon\,. \tag{3.9}$$

Putting together (3.6), (3.8) and (3.9), we finally get

$$\mathcal{C}_t(\xi) \le \mathcal{C}_t(\gamma) + \varepsilon\bigg(C(N+1) - \frac{r}{2M}\,\kappa\bigg(\frac{\delta r}{M}, \frac{\sqrt{n}r}{t}\bigg)\bigg)\,. \tag{3.10}$$

Let us finally use the assumption that c_t is a *highly* relativistic cost function: when we choose r only t' has been fixed, hence we are allowed to chose it so small that $\kappa(\eta, \sqrt{n}r/t) \to \infty$ for $\eta \searrow 0$. Having chosen r, we get M by Step II, and N by Proposition 2.12. Thus, we are free to choose δ depending on N and r; since c_t is highly relativistic, we can do this in such a way that the term in parentheses in (3.10) is strictly negative. We have then found a competitor with cost strictly smaller than the optimal plan γ, the contradiction gives the thesis. ☐

We can now give a "quantitative" version of Theorem C, where we state precisely how many pairs (x, y) can be moved from an optimal plan γ of a vector which is inside $t\mathscr{C}$ but very close to $t\partial\mathscr{C}$; it is important to notice that the next result holds for any relativistic cost function, regardless whether or not it is also highly relativistic. To state our result, we have to generalize the definition (3.1) of $\kappa(\eta, \rho)$ as follows: for every $s \ll 1$, we set

$$\kappa_s(\rho) := \lim_{\eta \searrow 0} \kappa_s(\eta, \rho) = \lim_{\eta \searrow 0} \inf \left\{ \frac{h(v) - h(v - \lambda w)}{\lambda} : v \in \overline{\mathscr{C}} \setminus (1-s)\mathscr{C}, |w - v| \leq \rho, 0 < \lambda < \eta \right\}. \tag{3.11}$$

Proposition 3.3. *Let c_t be a relativistic cost function, $t > T$, and assume that $\mu \in L^\infty$ and the supports of μ and ν are bounded. Then, for any $t > T$ and any ρ small enough, there exists a constant $\overline{C} = \overline{C}(\mu, \nu, t, \rho)$ such that, for s small enough, any optimal plan γ for $\mathcal{C}_t$ satisfies*

$$\gamma\Big(\Big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : \frac{y - x}{t} \in \overline{\mathscr{C}} \setminus (1-s)\mathscr{C}\Big\}\Big) \leq \frac{\overline{C}}{\kappa_s(\rho)}\,. \tag{3.12}$$

Proof. The scheme of proof is the same as that of Theorem C, there are some important modifications however. Let us start from the beginning: first of all, we fix once for all $T < t' < t$, a plan γ' with finite $\mathcal{C}_{t'}$ cost, and a constant r such that the inclusion $t'\mathscr{C} + 2\sqrt{n}r \subseteq t\mathscr{C}$ holds.

We can find a finite number points x_i and y_j such that the cubes centered at x_i (resp., y_i) with side-length r cover the support of μ (resp., of ν): here we use the fact that both the supports are bounded. The support of γ is then covered by a finite number C_1 of products of cubes $Q_{x_i} \times Q_{y_j}$. Let us fix such a product, say $Q_{x_i} \times Q_{y_j}$ and consider the measure $\gamma_1 = \gamma_1(i, j)$ given by

$$\gamma_1 := \gamma \llcorner \big(Q_{x_i} \times Q_{y_j}\big)\,.$$

If this measure is non-zero, we apply Proposition 2.12 finding constants $N(i, j)$ and $\overline{\varepsilon}(i, j)$. We call simply N and $\overline{\varepsilon}$ the biggest and smallest of the constants $N(i, j)$ and $\overline{\varepsilon}(i, j)$. For any $s < 1$, now, we define

$$\gamma_s := \gamma \llcorner \left\{(x, y) \in Q_{x_i} \times Q_{y_j},\, \frac{y - x}{t} \in \overline{\mathscr{C}} \setminus (1-s)\mathscr{C}\right\}.$$

Notice that $\gamma_s \leq \gamma_1$. We will now repeat the very same construction as in the proof of Theorem C, using γ_s in place of γ_0 and (x_i, y_j) in place of $(\overline{x}, \overline{y})$. In particular, we give the same definition of f_τ and Z as in Step I there, with γ_0 replaced by γ_s; then, we repeat exactly Step II, so we find that $\alpha(Z) \leq m_{ij}/3$ as soon as we have chosen

$$M = \frac{3\|\mu\|_{L^\infty} r^n}{m_{ij}}\,, \tag{3.13}$$

where $m_{ij} = \|\gamma_s\|$ –compare with the estimate (3.2). Observe that the choice of M depends on m_{ij}, which in turn depends also on s. Notice also that the assumption that $\mu \in L^\infty$ is used in (3.13) only.

Now, we repeat Step III, defining the measures ξ_1 and ξ_2; keep in mind that, in the whole proof of Theorem C, we used only once the fact that the measure γ_0 was concentrated on pairs (x, y) with $y - x \in t\partial\mathscr{C}$, that is, to obtain the estimate (3.5) on the gain in the cost of a single pair, and in turn (3.5) was used only immediately after, to obtain (3.6). Since now the measure

γ_s is concentrated in pairs (x,y) such that $(y-x)/t$ does not necessarily belong to $\partial\mathscr{C}$, but to $\overline{\mathscr{C}}\setminus(1-s)\mathscr{C}$, we just have to substitute κ with κ_s, so this time instead of (3.5) we get

$$c_t\big(g(x,y)\big)-c_t(x,y)\leq -\frac{\delta r}{M}\,\kappa_s\big(\delta r/M,\rho\big)\,,$$

where we write for brevity $\rho=r\sqrt{n}/t$, and then instead of (3.6) we get

$$\mathcal{C}_t(\xi_1+\xi_2)\leq \mathcal{C}_t(\gamma_s)-\frac{\varepsilon r}{2M}\,\kappa_s\big(\delta r/M,\rho\big)\,. \tag{3.14}$$

We have now to repeat Step IV: first recall that, in the construction of ξ_1 and ξ_2, we have used the measure γ_s in place of γ_0; nevertheless, we have applied Proposition 2.12 to γ_1: this is of primary importance, since in this way N and $\overline{\varepsilon}$ do not depend on s. Then, this time we have that the two projections of $\xi_1+\xi_2$ cover almost completely the two projections of γ_s; hence, as in Step IV of the last proof, we set now $\xi_3=\gamma-\gamma_s-\tilde\gamma_\infty$, which is positive since $\tilde\gamma_\infty\leq\gamma-\gamma_1\leq\gamma-\gamma_s$. Finally ξ_4 is, as before, a plan having the correct marginals so that $\xi\in\Pi(\mu,\nu)$. Again, for almost every pair (x,y) in the support of ξ_4 we have that $y-x\in t'\mathscr{C}+2\sqrt{n}r$, thus belongs to $t\mathscr{C}$ (and again, the choice of s has no influence on this). Therefore, (3.8) and (3.9) generalize to

$$\mathcal{C}_t(\xi_3)=\mathcal{C}_t(\gamma)-\mathcal{C}_t(\gamma_s)\,,\qquad\qquad \mathcal{C}_t(\xi_4)\leq C_2(N+1)\varepsilon\,,$$

which together with (3.14) tell us that (3.10) becomes now

$$\mathcal{C}_t(\xi)\leq\mathcal{C}_t(\gamma)+\varepsilon\bigg(C_2(N+1)-\frac{r}{2M}\,\kappa_s\big(\delta r/M,\rho\big)\bigg)\,.$$

Since γ is an optimal measure, we get that the term in parentheses must be positive, which by letting $\delta\searrow 0$ means

$$M\geq\frac{r}{2C_2(N+1)}\,\kappa_s(\rho)\,.$$

Recalling now the choice of M in (3.13), we deduce that

$$\|\gamma_s\|=m_{ij}\leq\frac{6C_2(N+1)\|\mu\|_{L^\infty}r^{n-1}}{\kappa_s(\rho)}=\frac{C_3}{\kappa_s(\rho)}\,.$$

Basically, we have found that the pairs in $C_{x_i}\times C_{y_j}$ which are moved of (t times) a vector in $\overline{\mathscr{C}}\setminus(1-s)\mathscr{C}$ are only few. Putting now together all the pairs of cubes covering the support of γ, we precisely find (3.12), with $\overline{C}=C_1C_3$, and the proof is then concluded. □

Remark 3.4. *Observe that the above result is valid for any relativistic cost, not necessarily highly relativistic: actually, in the proof of Theorem C we use the highly relativistic assumption only at the very end, to infer that $\kappa(\eta,\rho)\to\infty$ for $\eta\searrow 0$; instead, this time the constant $\kappa_s(\rho)$ is surely finite for every $s>0$. Nevertheless, if the cost is highly relativistic, then $\kappa_s(\rho)\to+\infty$ for $s\searrow 0$ if ρ is small enough; hence, in the particular case of a highly relativistic cost the result of Proposition 3.3 is stronger than Theorem C. Notice also that, in the particular case of the* relativistic heat cost *(1.1), formula (3.12) takes the nice form*

$$\gamma\bigg(\Big\{(x,y)\in\mathbb{R}^n\times\mathbb{R}^n:\,\frac{y-x}{t}\in\overline{\mathscr{C}}\setminus(1-s)\mathscr{C}\Big\}\bigg)\leq\overline{C}\sqrt{s}\,.$$

4. Continuity of the total cost function

In this section, we prove our result on the continuity of the total cost function $\mathcal{C}(t)$, that is, Theorem A. First of all, we need the following auxiliary definition.

Definition 4.1. *Let c_t be a relativistic cost function. Then, for every $\delta > 0$ we call*

$$\operatorname{exc}(\delta) := \sup\left\{\frac{h(v)}{h\big((1-\delta)v\big)} \,:\, v \in \mathscr{C}\right\} - 1\,.$$

Notice that $\operatorname{exc}(\delta) \searrow 0$ *when* $\delta \searrow 0$.

Proof (of Theorem A). The proof of this Theorem is quite similar to that of Theorem C and Proposition 3.3, and in particular it makes use of Proposition 2.12.

Notice first that the function c_t decreases with t, so the inequality $\mathcal{C}(t) \leq \mathcal{C}(t-\delta)$ is obvious for every δ and we only have to show the second inequality in (1.3). Let us then fix t and t' such that $T < t' < t$, call γ and γ' two optimal plans corresponding to $\mathcal{C}_t$ and $\mathcal{C}_{t'}$, and let r be a small side-length with the property that the inclusion $t'\mathscr{C} + 2\sqrt{n}r \subset\subset t\mathscr{C}$ holds. Then, let us fix some $\overline{\delta} > 0$ with the property that

$$t'\mathscr{C} + 2\sqrt{n}r \subseteq (t-\overline{\delta})\mathscr{C}\,. \tag{4.1}$$

For every small $\delta > 0$, we define

$$S_\delta := \left\{(x,y) \in \mathbb{R}^{2n} \,:\, \frac{y-x}{t} \in (1-\delta)\mathscr{C}\right\}.$$

We can now cover the support of $\gamma \llcorner \big\{(x,y) \notin S_{\overline{\delta}}\big\}$ with C_1 cubes Q_i in $\mathbb{R}^{2n}$ with side-length r; for every $1 \leq i \leq C_1$ and for every $\delta < \overline{\delta}$, we set then

$$\gamma^i_{\overline{\delta}} := \gamma \llcorner Q_i \setminus S_{\overline{\delta}}\,, \qquad\qquad \gamma^i_\delta := \gamma \llcorner Q_i \setminus S_\delta \leq \gamma^i_{\overline{\delta}}\,,$$

and let us call $\mu^i_{\overline{\delta}}$, $\nu^i_{\overline{\delta}}$, μ^i_δ and ν^i_δ their projections. Up to discard useless cubes, we can assume that each $\gamma^i_{\overline{\delta}}$ is a non-zero measure. Proposition 2.12 applied to each of the measures $\gamma^i_{\overline{\delta}}$ gives us constants $N(i)$ and $\overline{\varepsilon}(i)$; we then set $N = \max\{N(i)\}$ and $\overline{\varepsilon} = \min\{\overline{\varepsilon}(i)\}$. Moreover, up to decrease the value of $\overline{\varepsilon}$, we have

$$\min\left\{\|\gamma^i_{\overline{\delta}}\|,\, 1 \leq i \leq C_1\right\} > 2\overline{\varepsilon}\,. \tag{4.2}$$

Since c_t is highly relativistic, we know that $\kappa_s(\rho) \to \infty$ when $s \to 0$, as soon as $\rho = r\sqrt{n}/t$ has been chosen small enough. Then Proposition 3.3 ensures us that

$$\sum_{i=1}^{C_1} \varepsilon_i := \sum_{i=1}^{C_1} \|\gamma^i_\delta\| \leq \frac{\overline{\varepsilon}}{N}\,, \tag{4.3}$$

as soon as $\delta \ll 1$, depending on $\overline{\varepsilon}$ and on $\overline{\delta}$.

For each $1 \leq i \leq C_1$, we apply now Proposition 2.12, relative to the measure $\gamma^i_{\overline{\delta}}$, to the constant $C_1\varepsilon_i$, and we find measures $\tilde{\gamma}^i \leq \gamma$ and $\tilde{\gamma}'^i \leq \gamma'$ with total mass $(N(i)+1)C_1\varepsilon_i$ satisfying all the properties listed in the proposition. By definition, each measure $\tilde{\gamma}^i_\infty$ is smaller than $\gamma - \gamma^i_{\overline{\delta}}$, hence it is orthogonal to the measure γ^i_δ; nevertheless, it could have some common part with γ_δ^j for some $j \neq i$. As a consequence, we need now to slightly modify each γ_δ^i: more precisely, we call $\gamma_{\delta,AD}^i$ the "already done" part, that is,

$$\gamma_{\delta,AD}^i := \left(\sum_{j=1}^{C_1} \frac{\tilde\gamma_\infty^j}{C_1}\right) \llcorner Q_i \setminus S_\delta \leq \gamma_\delta^i\,,$$

we define arbitrarily a "replacement part" $\gamma_{\delta,R}^i$ satisfying

$$\|\gamma_{\delta,R}^i\| = \|\gamma_{\delta,AD}^i\|\,, \qquad\qquad \gamma_{\delta,R}^i \leq \left(\gamma - \sum_{j=1}^{C_1} \frac{\tilde\gamma_\infty^j}{C_1}\right) \llcorner \Big((Q_i \setminus S_{\bar\delta}) \cap S_\delta \Big)\,, \tag{4.4}$$

so that $\gamma_{\delta,R}^i \leq \gamma_{\bar\delta}^i$ but $\gamma_{\delta,R}^i \wedge \gamma_\delta^i = 0$, and we finally set the "modified version" of γ_δ^i as

$$\widehat\gamma_\delta^i := \gamma_\delta^i - \gamma_{\delta,AD}^i + \gamma_{\delta,R}^i\,,$$

calling $\widehat\mu_\delta^i$ and $\widehat\nu_\delta^i$ its projections. Notice that the existence of some replacement part as in (4.4) is ensured by (4.2) and (4.3), and by construction we have that $\|\widehat\gamma_\delta^i\| = \|\gamma_\delta^i\| = \varepsilon_i$. We can now set

$$\xi_1 := \gamma - \sum_{i=1}^{C_1} \widehat\gamma_\delta^i - \sum_{i=1}^{C_1} \frac{\tilde\gamma_\infty^i}{C_1}\,,$$

which is a positive measure by construction (this could have been false if we had used γ_δ^i instead of $\widehat\gamma_\delta^i$). Notice carefully that the measure ξ_1 is concentrated on S_δ. The projections of ξ_1 are then

$$\pi_1\xi_1 = \mu - \sum_{i=1}^{C_1} \widehat\mu_\delta^i - \sum_{i=1}^{C_1} \frac{\tilde\mu^i}{C_1}\,, \qquad\qquad \pi_2\xi_1 = \nu - \sum_{i=1}^{C_1} \widehat\nu_\delta^i - \sum_{i=1}^{C_1} \frac{\tilde\nu^i}{C_1}\,,$$

and we look now for a measure ξ_2^i satisfying

$$\pi_1\xi_2^i = \widehat\mu_\delta^i + \frac{\tilde\mu^i}{C_1}\,, \qquad\qquad \pi_2\xi_2^i = \widehat\nu_\delta^i + \frac{\tilde\nu^i}{C_1}\,. \tag{4.5}$$

Notice that the measure $\tilde\gamma'^i/C_1$ has projections $\tilde\mu^i/C_1 + \mu_A/C_1$ and $\tilde\nu^i/C_1 + \nu_B/C_1$, and the measures μ_A/C_1 and ν_B/C_1 have mass ε_i, exactly as $\widehat\mu_\delta^i$ and $\widehat\nu_\delta^i$. Arguing as in the proof of Theorem C, we define $\xi_2^i = \beta \circ \tilde\gamma'^i/C_1 \circ \alpha$ so that (4.5) is satisfied. Notice that, as in Theorem C and thanks to (4.1), the measure ξ_2^i is concentrated on S_δ: since the same was already observed for ξ_1, we derive that the plan $\xi = \xi_1 + \sum_i \xi_2^i$ belongs to $\Pi(\mu,\nu)$, and it has finite $\mathcal{C}_{t-\delta t}$-cost. More precisely, setting $C = \max\{h(v) : v \in \mathscr{C}\}$, we have

$$\sum_{i=1}^{C_1} \mathcal{C}_{t-\delta t}(\xi_2^i) \leq C \sum_{i=1}^{C_1} \|\xi_2^i\| = C \sum_{i=1}^{C_1} N(i)\varepsilon_i \leq CN \sum_{i=1}^{C_1} \varepsilon_i\,.$$

Concerning ξ_1, we have

$$\begin{aligned}
\mathscr{C}_{t-\delta t}(\xi_1) &\leq \mathscr{C}_{t-\delta t}\Big(\gamma - \textstyle\sum_i \gamma_\delta^i\Big) = \displaystyle\iint_{S_\delta} c_{t-\delta t}(x,y)\,d\gamma(x,y) = \iint_{S_\delta} h\left(\frac{y-x}{t(1-\delta)}\right) d\gamma(x,y)\\
&\leq \iint_{S_\delta} (1+\mathrm{exc}(\delta))h\left(\frac{y-x}{t}\right) d\gamma(x,y) \leq \big(1+\mathrm{exc}(\delta)\big)\mathcal{C}_t(\gamma) = \big(1+\mathrm{exc}(\delta)\big)\mathcal{C}(t)\,.
\end{aligned}$$

Putting together the last two estimates (recall (4.3)) and applying Proposition 3.3, we get

$$\begin{aligned}\mathcal{C}(t-\delta t) &\leq \mathcal{C}_{t-\delta t}(\xi) \leq \big(1+\operatorname{exc}(\delta)\big)\mathcal{C}(t) + CN\sum\nolimits_i \varepsilon_i \\ &= \big(1+\operatorname{exc}(\delta)\big)\mathcal{C}(t) + CN\gamma\big(\mathbb{R}^{2n}\setminus S_\delta\big) \leq \big(1+\operatorname{exc}(\delta)\big)\mathcal{C}(t) + \frac{CN\overline{C}}{\kappa_\delta(\rho)}\,.\end{aligned}$$

This gives (1.3) which in turn, since $\kappa_\delta(\rho)\to\infty$ for $\delta\searrow 0$, also implies the continuity of the map $t\mapsto\mathcal{C}(t)$. $\square$

We end this section with a counterexample showing that, basically, all our assumptions are sharp.

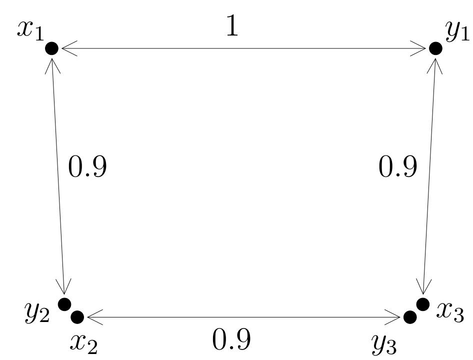

FIGURE 1. Situation in Example 4.2

Example 4.2. *Let us consider the situation drawn in Figure 1, where μ (resp., ν) is the sum of three Dirac masses of weight $1/3$ at the points x_1, x_2 and x_3 (resp., y_1, y_2 and y_3). Assume that the distance between x_1 and y_1 is 1, the distances between x_1 and y_2, x_2 and y_3, and x_3 and y_1 are all 0.9, and the distances between x_2 and y_2 and between x_3 and y_3 are some small $\varepsilon \ll 1$. Let h be a relativistic cost function such that $h(v)=f(|v|)$, with $f(0)=0$, $f(\varepsilon)\approx\varepsilon^2$, $f(1)=1$ and $f(r)=+\infty$ for $r>1$. We can assume that*

$$f(1)+2f(\varepsilon) < 3f(0.9)\,. \tag{4.6}$$

There are basically two possible ways of transporting μ onto ν: either sending each x_i onto the corresponding y_i, or sending each x_i onto y_{i+1}, where we denote for simplicity $y_4\equiv y_1$. These two possibilities correspond to paying once the price of moving of a distance 1 and twice of a distance ε, or to paying thrice the price of moving of a distance 0.9. It is obvious that $T=0.9$, and that for $T\leq t<1$ the second transport plan is the unique having a finite cost; instead, for $t\geq 1$ both transports (and all their convex combinations) have finite cost; by keeping in mind (4.6), this gives that the total cost is given by

$$\mathcal{C}(t) = \begin{cases} f(0.9/t) & \text{if } 0.9\leq t<1\,,\\ \frac{1}{3}f(1/t)+\frac{2}{3}f(\varepsilon/t) & \text{if } t\geq 1\,.\end{cases}$$

Hence, the total cost function $t\mapsto\mathcal{C}(t)$ can be discontinuous if μ is a singular measure, and this gives a counterexample to Theorem A without the assumption that $\mu\ll\mathscr{L}^n$ (while the *additional fact that μ is in L^∞ is basically only needed to get (1.3)). The same example also gives a counterexample to Theorem C without the assumption $\mu \ll \mathscr{L}^n$: indeed, if we take $t = 1$, then $t > T = 0.9$, and nevertheless there is a positive amount of mass moved with directions in $t\partial\mathscr{C}$ by the optimal plan relative to $\mathcal{C}_t$.*

Finally, a simple modification of the same example gives also a counterexample to the validity of Theorem C for costs which are relativistic but not highly relativistic. Indeed, substitute the Dirac masses in the example above with uniform measures concentrated in very small balls, the centers being the points x_i and y_i. It is then true that $T = 0.9$, and we want to show that for $t = 1$ (and with a good choice of f) the tranport plan γ_1 translating each ball centered at x_i onto the ball centered at y_i is optimal. If we do so, then again we have a counterexample to Theorem C for μ absolutely continuous but with relativistic and not highly relativistic cost functions. We just sketch how one can prove this: arguing as in the proof of Theorem C, if another transport plan $\tilde\gamma$ moves a portion $1-\delta$ of the points of the ball at x_1 onto the ball at y_1 by making a distance $1-\delta$ instead of 1, then the remaining portion δ of the ball at x_1 must move of a distance 0.9 onto the ball at y_2, then a portion δ of the ball at x_2 must go onto the ball at y_3, and finally a portion δ of the ball at x_3 must fill the remaining portion δ of the ball at y_1. The costs of the plans γ_1 and $\tilde\gamma$ are then

$$3\mathcal{C}_1(\gamma_1) = f(1) + 2f(\varepsilon)\,, \qquad\qquad 3\mathcal{C}_1(\tilde\gamma) \approx (1-\delta)\big(f(1-\delta) + 2f(\varepsilon)\big) + 3\delta f(0.9)\,.$$

A simple calculation, then, shows that basically $\tilde\gamma$ is better than γ_1 if

$$f'(1) > 3f(0.9) - f(1)$$

(recall that the right term is positive by (4.6)). Now, if the cost is highly *relativistic, then this means $f'(1) = +\infty$, and thus the above inequality is surely true and then γ_1 is* not *optimal, according with Theorem C. Instead, for a cost which is relativistic but not highly relativistic, it is possible that the above inequality fails, which implies the optimality of γ_1, and thus the claim of Theorem C is seen to be false without the highly relativistic assumption.*

Remark 4.3. *Roughly speaking, Theorems C and A have two main assumptions: c_t is* highly *relativistic and $\mu \ll \mathscr{L}^n$. The above example shows that the absolute continuity assumption on μ is essential for both results, and that the highly relativistic assumption on c_t is needed for Theorem C. On the contrary, we do not know whether the highly relativistic assumption is necessary for the continuity of the total cost function to hold true.*

5. Existence of Kantorovich potentials

The goal of this section is to prove the existence of a Kantorovich potential, namely, Theorem B. To do so, we now recall some classical concepts. The first definition is the *c-cyclical monotonicity*, which was introduced by Knott and Smith in [26, 27], generalizing the classical cyclical monotonicity defined by Rockafellar in [31], and later used by several authors.

Definition 5.1 (c-cyclically monotone set)**.** *A set $S \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is said* c-cyclically monotone *if for any $m \in \mathbb{N}$ and any choice of $(x_i, y_i) \in S$, $0 \le i \le m$, one has*

$$\sum_{i=0}^{m} c(x_i, y_i) \le \sum_{i=0}^{m} c(x_i, y_{i+1}) \,,$$

where $y_{m+1} = y_0$.

A classical result asserts that, whenever a mass transport problem with a l.s.c. cost c is considered, if the total cost is finite then any optimal transport plan is concentrated on a c-cyclical monotone set, see for instance [34]. The second notation that we will use is that of c-transform, c-concavity and c-subdifferential.

Definition 5.2 (c-transform, c-concave function and c-subdifferential)**.** *Let $\varphi : \operatorname{supp}\mu \to \mathbb{R} \cup \{-\infty\}$ (resp., $\psi : \operatorname{supp}\nu \to \mathbb{R} \cup \{-\infty\}$) be a Borel function; its* c-transform $\varphi^c : \operatorname{supp}\nu \to \mathbb{R} \cup \{\pm\infty\}$ *(resp., $\psi^c : \operatorname{supp}\mu \to \mathbb{R} \cup \{\pm\infty\}$) is defined as*

$$\varphi^c(y) := \inf_{x \in \operatorname{supp}\mu} c(x,y) - \varphi(x) \,, \qquad\qquad \psi^c(x) := \inf_{y \in \operatorname{supp}\nu} c(x,y) - \psi(y) \,.$$

The function φ is said c-concave *if $\varphi^c(y) < +\infty$ for every $y \in \operatorname{supp}\nu$ and $\varphi^{cc} = \varphi$. Finally, for every c-concave function φ, we define the* c-subdifferential *as the set*

$$\partial_c \varphi = \Big\{ (x,y) \in \operatorname{supp}\mu \times \operatorname{supp}\nu : \ \varphi(x) + \varphi^c(y) = c(x,y) \Big\} \,.$$

It is well-known that the c-subdifferential of any c-concave function is a c-monotone set, see for instance [34]. Now, we recall the definition of Kantorovich potential, and that of the approximate gradient.

Definition 5.3 (Kantorovich potential)**.** *A* Kantorovich potential *is a c-concave map $\varphi : \operatorname{supp}\mu \to \mathbb{R} \cup \{-\infty\}$ such that any optimal transport plan $\gamma \in \Pi(\mu,\nu)$ is concentrated on $\partial_c\varphi$.*

Definition 5.4 (Approximate gradient)**.** *Given a function $f : X \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ and a point x of density 1 in X, we say that $v \in \mathbb{R}^n$ is the* approximate gradient of f at x *if there exists a map $g : \mathbb{R}^n \to \mathbb{R}^n$ differentiable at x, with $\nabla g(x) = v$ and such that x has density 1 in the set $\{f = g\}$. If there exists such a vector, then it is necessarily unique, and we denote it by $\widetilde{\nabla} f(x)$; moreover, the function f is said to be* approximately differentiable *at x.*

The existence of a Kantorovich potential is an important issue in the mass transportation problem, also because it is helpful in showing the existence of an optimal transport map. It is known that a Kantorovich potential exists in the most common cases of transport problems, though not always.

To show Theorem B, we start with another "chain" lemma inspired by a result in [12] which is a sort of discrete version of the Chain Lemma 2.11.

Lemma 5.5. *Under the assumptions of Theorem B, there exists $M \in \mathbb{N}$ such that the following holds. Let γ_t be an optimal transport plan for the cost c_t, let $\Gamma \subseteq \operatorname{supp}\gamma_t$ be a set on which*

γ_t is concentrated, and fix $(x_0, y_0) \in \operatorname{supp}\gamma_t$. Then, for every $(x, y) \in \operatorname{supp}\gamma_t$, there is some $0 \le k \le M$ and points $(x_i, y_i) \in \Gamma$ for $0 < i \le k$ such that, calling $(x_{k+1}, y_{k+1}) = (x, y)$, for every $0 \le i \le k$ one has $y_i - x_{i+1} \in t\mathscr{C}$.

Proof. The proof is somehow similar to that of the chain Lemma 2.11, but in a discretized setting. For simplicity, we divide it in some steps.

Step I. *The discretized setting and the cycles.*

We start by selecting some $T < t' < t'' < t$ and a constant $\varepsilon > 0$, much smaller than $t - t''$, $t'' - t'$, and $t' - T$, and we cover $\operatorname{supp}\mu$ (resp., $\operatorname{supp}\nu$) with finitely many essentially disjoint cubes Q^1_j (resp., Q^2_k) of side-length $\varepsilon > 0$. We set $N + 1$ the number of pairs (j, k) such that $\gamma_t(Q^1_j \times Q^2_k) > 0$, and enumerate these cubes $Q^1_j \times Q^2_k$ as $Q_0, Q_1, Q_2, \ldots, Q_N \subseteq \mathbb{R}^n \times \mathbb{R}^n$. Notice that the number of cubes can be bounded only in terms of the supports of μ and ν, and of the value of ε, thus in turn only in terms of μ, ν and t; hence, there is some $M \in \mathbb{N}$, depending only on μ, ν and t, such that $N \le M$. For every $0 \le i \le N$, writing $Q_i = Q^1_j \times Q^2_k$ we call z_i and w_i the centers of Q^1_j and Q^2_k. Notice that the pairs (z_i, w_i) are all different, but of course $z_i = z_m$ (resp., $w_i = w_m$) whenever the first (resp., the second) projection of the $2n$-dimensional cubes Q_i and Q_m is the same n-dimensional cube.

Let us now set the "discretized problem" as follows: we define

$$\gamma_d = \sum_{i=0}^{N} \delta_{(z_i, w_i)}\gamma_t(Q_i)\,, \qquad \mu_d = \pi_1\gamma_d, \qquad \nu_d = \pi_2\gamma_d\,,$$

and we observe that the critical speed T_d of the discretized problem corresponding to μ_d and ν_d is at most t'. Indeed, for every j, k let us call

$$\alpha_j = \mu_d \llcorner Q^1_j \otimes \mu \llcorner Q^1_j\,, \qquad \beta_k = \nu \llcorner Q^2_k \otimes \nu_d \llcorner Q^2_k\,,$$

so that $\alpha = \sum \alpha_j$ (resp., $\beta = \sum \beta_k$) is a transport plan between μ_d and μ (resp., ν and ν_d); by construction, for every $(x', x'') \in \operatorname{supp}\alpha$ (resp., for every $(y', y'') \in \operatorname{supp}\beta$) one has $|x'' - x'| \le \varepsilon\sqrt{n}$ (resp., $|y' - y''| \le \varepsilon\sqrt{n}$). As a consequence, if we call γ_T an optimal transport plan corresponding to c_T, we have that $\gamma' := \beta \circ \gamma_T \circ \alpha$ is a transport plan between μ_d and ν_d, and for every $(x, y) \in \operatorname{supp}\gamma'$ one has $y - x \in T\mathscr{C} + \mathcal{B}(2\varepsilon\sqrt{n}) \subseteq t'\mathscr{C}$, where $\mathcal{B}(r)$ denotes the ball of radius r centered at the origin, and the last inclusion holds as soon as ε is small enough.

Let us now take any two indices $0 \le i, i' \le N$: we say that i' *follows* i if $(z_{i'}, w_i) \in \operatorname{supp}\gamma'$; notice that for every index i there is at least one i' which follows i, possibly i itself. Then, we call *cycle* any finite sequence of not necessarily distinct indices $I = \{i_1, i_2, \ldots, i_p\}$ such that every i_{m+1} follows i_m, and i_1 follows i_p; in particular, if i follows itself, then the singleton $I = \{i\}$ is a cycle.

Step II. *Every index $0 \le i \le N$ belongs to at least a cycle.*

For every $0 \le i \le N$, we can start a sequence of indices from i, so that each one follows the preceding one, until a cycle is found; notice that this cycle does not necessarily contain i (otherwise, the claim would be trivial); notice also that a cycle might contain twice the same index, but any cycle contains a subcycle with all distinct indices. Take then an arbitrary cycle

$I = \{i_1,\, i_2,\, \dots\, ,\, i_p\}$, containing all distinct indices, and set

$$\eta = \min_{\ell=1,\,\dots\,,\,p} \gamma_d\big(\{(z_{i_\ell}, w_{i_\ell})\}\big) \wedge \min_{\ell=1,\,\dots\,,\,p} \gamma'\big(\{(z_{i_{\ell+1}}, w_{i_\ell})\}\big)\,,$$

considering as usual $i_{p+1} = i_1$; notice that η is strictly positive by construction. Now, define the "reduced transport plans"

$$\gamma_{\rm red} = \gamma_d - \eta \sum_{\ell=1}^{p} \delta_{(z_{i_\ell}, w_{i_\ell})}\,, \qquad\qquad \gamma'_{\rm red} = \gamma' - \eta \sum_{\ell=1}^{p} \delta_{(z_{i_{\ell+1}}, w_{i_\ell})}\,.$$

Observe that these are two positive measures with the same marginals; observe also that the number of the pairs in the support of $\gamma_{\rm red}$ and $\gamma'_{\rm red}$ is at most the corresponding number for γ_d and γ', and at least one of them has strictly decreased. We can repeat the same construction with $\gamma_{\rm red}$ and $\gamma'_{\rm red}$ in place of γ_d and γ', so to find again a cycle: notice that, since $\gamma_{\rm red} \le \gamma_d$ and $\gamma'_{\rm red} \le \gamma'$, then any cycle with respect to $\gamma_{\rm red}$ and $\gamma'_{\rm red}$ is *a fortiori* a cycle with respect to γ_d and γ'. Going on in the same way, after at most $2M + 2$ steps the reduced transport plans become null: this means that every index belongs to some of the cycles which have been found, so this step is concluded.

Step III. The "regions" and the "networks".

For any index $0 \le i \le N$, we call *region* the union $\mathcal{R}$ of all the indices which belong to some cycle containing i. It is clear that for any two elements $m, p \in \mathcal{R}$ there is a sequence of indices, each following the preceding one, which starts with m and ends in p; it is also clear that, if we had started with some $m \in \mathcal{R}$, instead than with i, we would have found the same region. In other words, the set of all the indices is subdivided in finitely many disjoint regions. Take now two indices i and m such that $z_i = z_m$; then, m and i belong to the same region: indeed, for any index p one has that i follows p if and only if m follows p, and then we can trivially use two cycles, one containing i and the other containing m, to build another cycle containing both i and m. As a consequences, the first projections of the regions form also a disjoint cover of $\operatorname{supp} \mu_d$.

Let us now introduce another piece of notation: given two indices m, p we say that *m can be hooked to p* if $w_p - z_m \in t''\mathscr{C}$. By definition, if m follows p then it is also true that m can be hooked to p, but the converse is not necessarily true. Moreover, we say that *the region $\mathcal{R}'$ can be hooked to the region $\mathcal{R}$* if there exist two indices $m \in \mathcal{R}'$, $p \in \mathcal{R}$ such that m can be hooked to p. Finally, for any region $\mathcal{R}$ we call its *network* the union of all the regions $\mathcal{R}'$ such that there is a finite sequence of regions, each one hooked to the preceding one, starting with $\mathcal{R}$ and ending with $\mathcal{R}'$. Notice that, in principle, the fact that $\mathcal{R}'$ can be hooked to $\mathcal{R}$ does not imply that $\mathcal{R}$ can be hooked to $\mathcal{R}'$; more in general, if the network of $\mathcal{R}$ contains $\mathcal{R}'$, in principle it is possible that the network of $\mathcal{R}'$ does not contain $\mathcal{R}$: in other words, the networks are not automatically a partition of the indices. Nevertheless, in the next step we see that something much stronger is actually true.

Step IV. Every network contains all the regions.

Let us fix a region $\mathcal{R}$, and call $\mathcal{N}$ the network of $\mathcal{R}$: we aim to prove that $\mathcal{N}$ contains all the regions. Since we already observed, in Step III, that whenever $z_i = z_m$ the two indices i and m

are in the same region, it is enough to show that the union $\mathcal{N}_1$ of all the cubes centered at some z_i with $i \in \mathcal{N}$ contains the whole support of μ_d.

Suppose that it is not so: this means that the measure μ is not concentrated on $\mathcal{N}_1$. Since the support of μ is connected, there must be some cube Q^1_j which does not belong to $\mathcal{N}_1$ but is adjacent to one of the cubes of $\mathcal{N}_1$, say Q^1_l, such that $\mu(Q^1_j) > 0$. Assuming for simplicity of notations that Q^1_j and Q^1_l are contained in the $2n$-dimensional cubes Q_j and Q_l, we have then that $|z_j - z_l| \leq \varepsilon\sqrt{n}$; recall that by construction we have $l \in \mathcal{N}$ and $j \notin \mathcal{N}$. Let then p be an index such that l follows p: this implies that $(z_l, w_p) \in \operatorname{supp}\gamma'$, hence $w_p - z_l \in t'\mathscr{C}$, so in turn $w_p - z_j \in t'\mathscr{C} + \mathcal{B}(\varepsilon\sqrt{n}) \subseteq t''\mathscr{C}$, where again the last inclusion holds if ε has been chosen small enough. Since this means that j can be hooked to p, and since the fact that l follows p implies that l and p are in the same region, thus also p belongs to $\mathcal{N}$, we derive that the index j is an element of $\mathcal{N}$, which gives the desired contradiction. Hence, the network of each region is made of all the indices, and this concludes this step.

Step V. Conclusion.
We are finally in position to conclude the proof of this lemma. Fix a pair (x_0, y_0) in $\operatorname{supp}\gamma_t$, and take any other pair $(x, y) \in \operatorname{supp}\gamma_t$. Thanks to Step IV, we can find $k \geq 0$ and indices $j(0), j(1), \ldots, j(k+1)$ such that $(x_0, y_0) \in Q_{j_0}$, $(x, y) \in Q_{j_{k+1}}$, and for every $0 \leq i \leq k$ the index j_{i+1} can be hooked to j_i; of course, we can do this in such a way that all the indices j_i are distinct, except for j_0 and j_{k+1} which could coincide. As a consequence, since there exist $N+1$ indices, for sure we have $k \leq N$, hence $k \leq M$ since we know that $N \leq M$. Now, set $x_{k+1} = x$ and $y_{k+1} = y$, and for every $0 < i \leq k$ choose arbitrarily some pair $(x_i, y_i) \in \Gamma \cap Q_i$: this is possible because, by construction, $\gamma_t(Q_i) > 0$, and γ_t is concentrated on Γ. We claim that the sequence (x_i, y_i) for $0 < i \leq k$ fulfills the requirements of the Lemma.

Indeed, the fact that every pair (x_i, y_i) with $0 < i \leq k$ belongs to Γ is true by construction. Moreover, for every i we have that x_i (resp., y_i) is in a cube of side ε centered at z_i (resp., w_i). As a consequence, for every $0 \leq i \leq k$ we have

$$y_i - x_{i+1} \in w_i - z_{i+1} + \mathcal{B}(2\varepsilon\sqrt{n}) \subseteq t''\mathscr{C} + \mathcal{B}(2\varepsilon\sqrt{n}) \subseteq t\mathscr{C}\,,$$

where as usual the last inclusion is true if ε was chosen small enough. $\square$

The next step to prove Theorem B is to show the existence of a Kantorovich potential for c_t whenever $t > T$. This is the content of the next result. We underline that the same result was proven in [12] only for almost every $t > T$, and under additional assumption on the measure μ.

Proposition 5.6. *Under the assumptions of Theorem B, for any supercritical speed $t > T$ there exists a Kantorovich potential φ_t for c_t, which is approximately differentiable μ-almost everywhere.*

Proof. Let us fix a supercritical speed $t > T$, and let us assume without loss of generality –and just for the simplicity of notations– that $t = 1$. Let us call γ an optimal transport plan relative to the cost $c = c_1$, let $\Gamma \subseteq \operatorname{supp}\gamma$ be a c-cyclical monotone set on which γ is concentrated, and let us fix arbitrarily a point $(x_0, y_0) \in \Gamma$: since the time $t = 1$ is supercritical and γ is an optimal transport plan, we know that $y_0 - x_0 \in \mathscr{C}$. Let us then define the map $\varphi : \operatorname{supp}\mu \to \overline{\mathbb{R}}$ as

$$\varphi(x) = \inf_{k\in\mathbb{N}} \inf \left\{ \sum_{i=1}^{k} c(x_i, y_{i-1}) + c(x, y_k) - \sum_{i=0}^{k} c(x_i, y_i) : (x_i, y_i)_{i=1}^{k} \in \Gamma^k \right\}. \tag{5.1}$$

The proof will by concluded once we show that this map is a μ-almost everywhere approximately differentiable Kantorovich potential.

Step I. The map φ is well-defined and bounded from above and from below.

First of all notice that, since the set $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : c(x, y) = +\infty\}$ is open, then $c(x, y) < +\infty$ for every $(x, y) \in \operatorname{supp}\gamma$, hence in particular for every $(x, y) \in \Gamma$. As a consequence, for every $k \in \mathbb{N}$ and every choice of the sequence $(x_i, y_i)_{i=1}^{k} \in \Gamma^k$, the second sum in the braces above is real valued, while the first sum is possibly $+\infty$. Hence, the term in braces is well-defined for every sequence and then the function φ is well-defined.

We aim now to show that φ is bounded from above and from below. First of all, take any $x \in \operatorname{supp}\mu$, and let y be a point such that $(x, y) \in \operatorname{supp}\gamma$. By Lemma 5.5, we can find a finite chain $(x_i, y_i)_{i=1}^{k} \in \Gamma^k$ such that, writing $x_{k+1} = x$, for every $0 \le i \le k$ one has $y_i - x_{i+1} \in \mathscr{C}$, hence $c(x_{i+1}, y_i) \le \|h\|_{L^\infty}$. This implies that $\varphi(x) \le (k + 1)\|h\|_{L^\infty}$. Lemma 5.5 and the fact that x is arbitrary give $k \le M$ for some M depending only on μ, ν and t. Thus, we obtain that $\varphi \le (M + 1)\|h\|_{L^\infty}$ and the upper bound is proved.

We now deal with the lower bound. Set $x \in \operatorname{supp}\mu$, $k \in \mathbb{N}$, and pairs $(x_i, y_i)_{i=1}^{k} \in \Gamma^k$. Let us apply Lemma 5.5 with starting point (x_k, y_k) and final point (x_0, y_0). We get some $0 \le \ell \le M$ and points $(x_i, y_i) \in \Gamma$ for $k < i \le k + \ell$ such that $t\mathscr{C}$ contains $y_{i-1} - x_i$ for every $k < i \le k + \ell$, as well as $y_{k+\ell} - x_0$. As a consequence $c(x_i, y_{i-1})$ is bounded by $\|h\|_{L^\infty}$ for every $k < i \le k + \ell$, as well as $c(x_0, y_{k+\ell})$. Therefore, we can estimate

$$\begin{aligned}
&\sum_{i=1}^{k} c(x_i,y_{i-1}) + c(x, y_k) - \sum_{i=0}^{k} c(x_i, y_i)\\
&\qquad\ge \sum_{i=1}^{k+\ell} c(x_i, y_{i-1}) - \sum_{i=0}^{k+\ell} c(x_i, y_i) - \sum_{i=k+1}^{k+\ell} c(x_i, y_{i-1}) + c(x_0, y_{k+\ell}) - c(x_0, y_{k+\ell})\\
&\qquad\ge \sum_{i=1}^{k+\ell} c(x_i, y_{i-1}) + c(x_0, y_{k+\ell}) - \sum_{i=0}^{k+\ell} c(x_i, y_i) - (M + 1)\|h\|_{L^\infty} \ge -(M + 1)\|h\|_{L^\infty} ,
\end{aligned}$$

where the last inequality holds because all the points (x_i, y_i) are in the c-cyclically monotone set Γ. Keeping in mind the definition (5.1) of φ and the arbitrariness of $x \in \operatorname{supp}\mu$ and of the sequence $(x_i, y_i)_{i=1}^{k}$, we obtain that $\varphi \ge -(M + 1)\|h\|_{L^\infty}$.

Step II. φ is c-concave.

In this step we want to show that φ is c-concave, that is, $\varphi^c(y) < +\infty$ for every $y \in \operatorname{supp}\nu$, and $\varphi^{cc} = \varphi$. First of all, take any $y \in \operatorname{supp}\nu$ and let $x \in \operatorname{supp}\mu$ be such that $(x, y) \in \operatorname{supp}\gamma$, hence $c(x, y) \le \|h\|_{L^\infty}$. Thus, also by the estimates of Step I,

$$\varphi^c(y) \le c(x, y) - \varphi(x) \le (M + 2)\|h\|_{L^\infty} .$$

We have proved that φ^c is bounded from above, so the function φ^{cc} is well defined. We have to check that $\varphi^{cc} = \varphi$. Actually, since the inequality $\varphi^{cc} \geq \varphi$ is always true, we only have to prove the reverse inequality. Let us now define the function

$$\psi(y) = \sup_{k\in\mathbb{N}} \sup \left\{ \sum_{i=0}^{k} c(x_i, y_i) - \sum_{i=1}^{k} c(x_i, y_{i-1}) :\ (x_i, y_i)_{i=1}^k \in \Gamma^k,\ y_k = y \right\}, \tag{5.2}$$

which is well-defined and Borel. We claim that $\psi(y) < +\infty$ for every $y \in \operatorname{supp}\nu$. If this is the case, then the c-transform ψ^c is well-defined, and comparing (5.2) with (5.1) it is immediate to realize that $\psi^c = \varphi$. Since we know that $\varphi = \psi^c$ is bounded, we obtain then that $\psi \leq \psi^{cc} = \varphi^c$, and this implies that $\psi^c \geq \varphi^{cc}$, that is, the inequality $\varphi \geq \varphi^{cc}$ follows. Summarizing, to prove that φ is c-concave we just have to check that $\psi(y) < +\infty$ for every $y \in \operatorname{supp}\nu$.

First, notice that $\psi(y_0) = c(x_0, y_0)$: indeed, the inequality $\psi(y_0) \geq c(x_0, y_0)$ comes by choosing $k = 1$ and $(x_1, y_1) = (x_0, y_0)$ in (5.2); on the other hand, for every sequence $(x_i, y_i)_{i=1}^k$ in Γ^k with $y_k = y_0$, one has

$$\sum_{i=0}^{k} c(x_i, y_i) - \sum_{i=1}^{k} c(x_i, y_{i-1}) = \sum_{i=0}^{k-1} c(x_i, y_i) + c(x_k, y_0) - \sum_{i=1}^{k} c(x_i, y_{i-1}) \leq c(x_0, y_0)$$

by the c-cyclical monotonicity of Γ, and then we get $\psi(y_0) \leq c(x_0, y_0)$ by definition.

Let us now take any $y \in \operatorname{supp}\nu$: we must show that $\psi(y) < +\infty$. If y does not belong to the second projection $\pi_2(\Gamma)$ of Γ, then (5.2) readily gives $\psi(y) = -\infty$, so there is nothing to prove. Otherwise, let $x \in \operatorname{supp}\mu$ be such that $(x, y) \in \Gamma$. First of all, consider any other $(\tilde{x}, \tilde{y}) \in \Gamma$: for each sequence $(x_i, y_i)_{i=1}^k \in \Gamma^k$ with $y_k = y$, we can then add $(x_{k+1}, y_{k+1}) = (\tilde{x}, \tilde{y})$ to get a sequence of length $k + 1$ which is admissible in the definition of $\psi(\tilde{y})$, and we then $\psi(\tilde{y}) \geq \psi(y) + c(\tilde{x}, \tilde{y}) - c(\tilde{x}, y)$, which implies, for any choice of $(\tilde{x}, \tilde{y}) \in \Gamma$, the inequality

$$\psi(y) \leq \psi(\tilde{y}) + c(\tilde{x}, y) - c(\tilde{x}, \tilde{y}) \leq \psi(\tilde{y}) + c(\tilde{x}, y)\,. \tag{5.3}$$

Let us now apply once again Lemma 5.5, starting from (x, y) and reaching (x_0, y_0). We get some $0 \leq k \leq M$ and point (x_i, y_i) in Γ for $1 \leq i \leq k$ in such a way that $\mathscr{C}$ contains all the vectors

$$y - x_1\,,\ y_1 - x_2\,,\ \dots\,,\ y_{k-1} - x_k\,,\ y_k - x_0\,.$$

Therefore, we can apply several times (5.3) to get

$$\begin{aligned}
\psi(y) &\leq \psi(y_1) + c(x_1, y) \leq \psi(y_2) + c(x_2, y_1) + c(x_1, y) \leq \ \cdots \\
&\leq \psi(y_k) + c(x_k, y_{k-1}) + \cdots + c(x_1, y) \leq \psi(y_0) + c(x_0, y_k) + c(x_k, y_{k-1}) + \cdots + c(x_1, y) \\
&\leq c(x_0, y_0) + (M + 1)\|h\|_{L^\infty} \leq (M + 2)\|h\|_{L^\infty}\,.
\end{aligned}$$

As a consequence, we have proved that ψ is bounded from above, so this step is concluded.

Step III. *γ is concentrated on $\partial_c\varphi$.*

To conclude the proof that φ is a Kantorovich potential, we have to check that every optimal transport plan is concentrated on $\partial_c\varphi$; since Theorem 1.3 gives the uniqueness of an optimal transport plan, we just have to check that the plan γ is concentrated on $\partial_c\varphi$ (actually, we do not need to rely on Theorem 1.3, see Remark 5.7 at the end). Since we know that γ is concentrated on Γ, we will just check that $\Gamma \subseteq \partial_c\varphi$.

To do so, let us fix $(\overline{x},\overline{y})\in\Gamma$: by (5.3), for any other $y\in\pi_2(\Gamma)$ we have the inequality

$$\psi(\overline{y})-c(\overline{x},\overline{y})\geq\psi(y)-c(\overline{x},y)=-\Big(c(\overline{x},y)-\psi(y)\Big)\,. \tag{5.4}$$

We notice that the same inequality holds also for every $y\in\operatorname{supp}\nu$, because if $y\in\operatorname{supp}\nu\setminus\pi_2(\Gamma)$ then we have already noticed that $\psi(y)=-\infty$, so (5.4) is true. Passing then to the supremum in $y\in\operatorname{supp}\nu$, we get $\psi(\overline{y})-c(\overline{x},\overline{y})\geq-\psi^c(\overline{x})$. We now observe that, since $\psi^c=\varphi$, then $\varphi^c=\psi^{cc}\geq\psi$, so the last estimate gives

$$\varphi(\overline{x})+\varphi^c(\overline{y})\geq\varphi(\overline{x})+\psi(\overline{y})\geq c(\overline{x},\overline{y})\,,$$

and since the inequality $\varphi(x)+\psi(y)\leq c(x,y)$ is always true for any $x\in\operatorname{supp}\mu$ and $y\in\operatorname{supp}\nu$, we have proved that $(\overline{x},\overline{y})\in\partial_c\varphi$, and this step is concluded.

Step IV. *φ is approximatively differentiable μ-almost everywhere.*

To conclude the proof, it remains to show that φ is Borel and approximately differentiable μ-almost everywhere. To do so, let us define the sets

$$\Theta_n:=\Big\{x\in\operatorname{supp}\mu:\,\exists\,y\in\operatorname{supp}\nu,\,(x,y)\in\Gamma,\,y-x\in\big(1-n^{-1}\big)\mathscr{C}\Big\}.$$

Observe that the union of the sets Θ_n has full μ-measure in $\operatorname{supp}\mu$ since γ is concentrated in Γ and by Theorem C. Let now x, z be two points in Θ_n, let y be such that $(x,y)\in\Gamma$ with $y-x\in(1-n^{-1})\mathscr{C}$, and observe that, since

$$\varphi(x)+\psi(y)=c(x,y)\,,\qquad\qquad\varphi(z)+\psi(y)\leq c(z,y)\,,$$

then we have

$$\varphi(z)-\varphi(x)\leq c(z,y)-c(x,y)\,. \tag{5.5}$$

Let now $\varepsilon_n>0$ be a geometrical constant such that $B(0,2\varepsilon_n)\subseteq(1/n)\mathscr{C}$. Since c is a relativistic cost, it is a convex function of the Euclidean distance; thus (5.5) implies the existence of a constant K_n such that, for $x,z\in\Theta_n$, $|z-x|\leq\varepsilon_n$,

$$|\varphi(z)-\varphi(x)|\leq K(n)|z-x|\,.$$

In other words, $\varphi_{|\Theta_n}$ is locally Lipschitz and, since it is also bounded by Step I, we get that it is actually Lipschitz. By Kirszbraun's Theorem, there is a Lipschitz extension φ_n of $\varphi_{|\Theta_n}$. According to Definition 5.4, we obtain that φ is approximately differentiable at x for every point x of density 1 in Θ_n where φ_n is differentiable, hence for μ-almost every $x\in\Theta_n$ (recall that μ is absolutely continuous with respect to the Lebesgue measure). Since this holds for every $n\in\mathbb{N}$, we have obtained that φ is Borel and approximatively differentiable μ-almost everywhere. □

Having the existence of the Kantorovich potential at hand, the existence of an optimal transport map is classical. Let us give a proof for the sake of completeness.

Proof (of Theorem B). Let φ_t be the Kantorovich potential given by Proposition 5.6, Θ_n the sets defined in Step IV of the proof of that proposition. According to Theorem C, for μ-a.e. x, there exists some y such that $(x,y)\in\operatorname{supp}\gamma_t$, $y-x\in t\mathring{\mathscr{C}}$, and φ_t is approximately differentiable at x. For every $z \in \cup_n \Theta_n$ close enough to x we know the validity of (5.5), which can be rewritten as

$$h\left(\frac{y-z}{t}\right) - h\left(\frac{y-x}{t}\right) \geq \varphi_t(z) - \varphi_t(x) = \widetilde{\nabla}\varphi_t(x) \cdot (z-x) + o(|z-x|)\,.$$

As a consequence, since x has density 1 in $\cup_n \Theta_n$, we get

$$-t\,\widetilde{\nabla}\varphi_t(x) \in \partial h\left(\frac{y-x}{t}\right),$$

hence

$$y = x + t\nabla h^*\big(-t\,\widetilde{\nabla}\varphi_t(x)\big)\,,$$

h^* being the Legendre transform of the strictly convex function h. This shows that the optimal transport plan γ_t is actually an optimal transport map, in particular $\gamma = (Id, F_t)_{\#}\mu$ for the optimal transport map $F_t(x) = x + t\nabla h^*\big(-t\widetilde{\nabla}\varphi_t(x)\big)$. The proof is then complete. □

Remark 5.7. *As underlined in the Introduction, our construction does not rely on the result of Theorem 1.3; in particular, we do not need to know* a priori *that there is a unique optimal transport plan and that it is given by a map, since we are able to show it in Theorem B. Nevertheless, we have seemingly used this existence during the proof of Proposition 5.6, at the beginning of Step III. Therefore, we have now to explain how the proof really works if one does not want to use the uniqueness given by Theorem 1.3. In Step III of the proof of Proposition 5.6, we are only able to check that the optimal plan γ used to build φ is concentrated on $\partial_c\varphi$; hence, we are still not sure that φ is a Kantorovich potential, because we would not know how to deal with other optimal plans, if there were. Thus, the proof of Theorem B still ensures that the plan γ is actually a transport map, but it gives no information on possible other plans. However, we can now conclude in a classical way: indeed, by the arbitrariness of γ, we have proved that every optimal transport plan is a transport map. If now γ_1 and γ_2 are two optimal transport plans, then so is also $\gamma = (\gamma_1 + \gamma_2)/2$, thus it must be a map. But it is immediate to realize that γ can be a map only if $\gamma_1 = \gamma_2$: this shows the uniqueness of the optimal transport plan, and in turn only now we can conclude that the map φ built before is actually a Kantorovich potential.*

References

[1] L. Ambrosio, N. Fusco & D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford University Press (2000).

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

[3] Andreu, Fuensanta; Caselles, Vicent; Mazón, José M. *Existence and uniqueness of solution for a parabolic quasilinear problem for linear growth functionals with L^1 data.* Math. Ann., 322, 139–206, (2002).

[4] Andreu, Fuensanta; Caselles, Vicent; Mazón, José M. *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals.* Progress in Mathematics vol 223, Birkhäuser Verlag (2004).

[5] Andreu, Fuensanta; Caselles, Vicent; Mazón, José M. *A strongly degenerate quasilinear equation: the elliptic case.* Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) 3, no. 3, 555–587 (2004).

[6] Andreu, Fuensanta; Caselles, Vicent; Mazón, José M. *A strong degenerate quasilinear equation: the parabolic case* Arch Rational Mech. Anal., 176, 415-453 (2005).

[7] Andreu, Fuensanta; Caselles, Vicent; Mazón, José M. *A strongly degenerate quasilinear elliptic equation.* Non Linear analysis, 61, 637–669 (2005).

[8] Andreu, Fuensanta; Caselles, Vicent; Mazón, José M. *The Cauchy Problem for a Strong Degenerate Quasilinear Equation.* J. Europ. Math.Soc., 7 , 361–393 (2005).

[9] Andreu, Fuensanta; Caselles, Vicent; Mazón, José M.; Moll, Salvador. *Finite propagation speed for limited flux diffusion equations.* Arch. Ration. Mech. Anal. 182, no. 2, 269–297 (2006).

[10] Beiglböck Mathias; Schachermayer, Walter. *Duality for Borel Measurable Cost Functions.* Trans. Amer. Math. Soc., 363, no. 8, 4203–4224 (2011).

[11] Bertrand, Jérôme. *Prescription of Gauss curvature using optimal mass transport.* To appear in Geometriae Dedicata.

[12] Bertrand, Jérôme; Puel, Marjolaine. *The optimal mass transport problem for relativistic costs.*, Calc. Var. and PDE **46**, no. 1–2 (2013), 353–374.

[13] Brenier, Yann. *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math. 44, 4 (1991), 375–417.

[14] Brenier, Yann. *Extended Monge-Kantorovich theory. Optimal transportation and applications* (Martina Franca, 2001), 91–121, Lecture Notes in Math., 1813, Springer, Berlin, (2003).

[15] Caffarelli, Luis A. *Boundary regularity of maps with convex potentials.* II. Ann. of Math. (2) 144 (1996), no. 3, 453–496.

[16] Caselles, Vicent. *Convergence of the "relativistic" heat equation to the heat equation as $c \to \infty$.* Publicaciones Matemàtiques. Vol 51, Nùm. 1, 121-142 (2007).

[17] C. Castaing & M. Valadier, Convex analysis and measurable multifunctions, Lecture Notes in Mathematics, **580**, Springer (1977).

[18] Dellacherie, Claude; Meyer, Paul-André. *Probabilities and potential* North-Holland Mathematics Studies, no. 29, North-Holland Publishing Co. Amsterdam (1978).

[19] L.C. Evans & W. Gangbo, Differential Equations Methods for the Monge–Kantorovich Mass Transfer Problem, Memoirs of the A.M.S., Vol. **137**, Number 653, (1999).

[20] Feyel, Denis; Üstunel, Suleyman. *Solution of the MongeAmpère equation on Wiener space for general log-concave measures.* J.F.A. 232, no. 1, 29–55 (2006).

[21] Gangbo, Wilfrid; McCann, Robert J. *The geometry of optimal transportation.* Acta Math. 177, no. 2, 113–161 (1996).

[22] Gangbo, Wilfrid; Oliker, Vladimir. *Existence of optimal maps in the reflector-type problems.* ESAIM COCV 13, 93–106 (2007).

[23] Glimm, Tilmann; Oliker, Vladimir. *Optical design of two-reflector systems, the Monge-Kantorovich mass transfer problem and Fermat's principle.* Indiana Univ. Math. J. 53, 1255–1277 (2004).

[24] Jimenez, Chloé; Santambrogio, Filippo. *Optimal transportation for a quadratic cost with convex constraints and applications.* To appear in Journal de Mathématiques Pures et Appliquées.

[25] Jordan, Richard; Kinderlehrer, David; Otto, Felix. *The variational formulation of the Fokker-Planck equation.* SIAM J. Math. Anal. 29, no. 1, 1–17 (1998).

[26] M. Knott & C. Smith, On the optimal mapping of distributions, J. Optim. Theory Appl., **43** no. 1 (1984), 39–49.

[27] Knott, Martin; Smith, Cyril S. *Note on the optimal transportation of distributions.* J. Optim. Theory Appl., 52 (2), 323–329, (1987).

[28] McCann, Robert; Puel, Marjolaine. *Constructing a relativistic heat flow by transport time steps.* Ann. Inst. H. Poincaré Anal. Non Linéaire 26, no. 6, 2539–2580 (2009).

[29] Oliker, Vladimir. *Embedding $\mathbb{S}^n$ into $\mathbb{R}^{n+1}$ with given integral Gauss curvature and optimal mass transport on $\mathbb{S}^n$.* Adv. Math. 213, no. 2, 600–620 (2007).

[30] Pratelli, Aldo. *On the sufficiency of c-cyclical monotonicity for optimality of transport plans*. Math. Z. 258 (2008), no. 3, 677–690.

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

[32] Rockafellar R.T. *Convex analysis.* Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970.

[33] Schachermayer, Walter; Teichmann, Josef. *Characterization of optimal transport plans for the Monge-Kantorovich problem.* Proc. Amer. Math. Soc. 137 (2009), no. 2, 519–529.

[34] Villani, Cédric. *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, Old and new, 2009.