

# EXISTENCE AND UNIQUENESS OF OPTIMAL TRANSPORT MAPS

FABIO CAVALLETTI AND MARTIN HUESMANN

**ABSTRACT.** Let  $(X, d, m)$  be a proper, non-branching, metric measure space. We show existence and uniqueness of optimal transport maps for cost written as non-decreasing and strictly convex functions of the distance, provided  $(X, d, m)$  satisfies a new weak property concerning the behavior of  $m$  under the shrinking of sets to points, see Assumption 1. This in particular covers spaces satisfying the measure contraction property.

We also prove a stability property for Assumption 1: If  $(X, d, m)$  satisfies Assumption 1 and  $\tilde{m} = g \cdot m$ , for some continuous function  $g > 0$ , then also  $(X, d, \tilde{m})$  verifies Assumption 1. Since these changes in the reference measures do not preserve any Ricci type curvature bounds, this shows that our condition is strictly weaker than measure contraction property.

## 1. INTRODUCTION

In [10], Gaspard Monge studied the by now famous minimization problem

$$(1.1) \quad \inf_{T_{\#}\mu_0=\mu_1} \int d(x, T(x))\mu_0(dx),$$

on Euclidean space, where  $\mu_0$  and  $\mu_1$  are two given probability measures and the minimum is taken over all maps pushing  $\mu_0$  forward to  $\mu_1$ . This problem turned out to be very difficult because the functional is non-linear and the constraint set maybe empty. 70 years ago, Kantorovich [8] came up with a relaxation of the minimization problem (1.1). He allowed arbitrary couplings  $q$  of the two measures  $\mu_0$  and  $\mu_1$ , which we denote by the set  $\Pi(\mu_0, \mu_1)$ , and also more general cost functions  $c : X \times X \rightarrow \mathbb{R}$ :

$$(1.2) \quad \inf_{q \in \Pi(\mu_0, \mu_1)} \int c(x, y)q(dx, dy).$$

Minimizers of (1.2) are called optimal couplings and, therefore, this family of problems is commonly called optimal transport problems. A natural and interesting question is when do these two minimization problems coincide, i.e. when is the or an optimal coupling given by a transportation map. In [5], Brenier showed using ideas from fluid dynamics that on Euclidean space with cost function  $c(x, y) = |x - y|^2$  there is always a unique optimal transportation map as soon as  $\mu_0$  is absolutely continuous with respect to the Lebesgue measure. Soon after, McCann [9] generalized this result to Riemannian manifolds with more general cost functions including convex functions of the distance. By now, this result is shown in a wide class of settings, for instance for non-decreasing strictly convex functions of the distance in Alexandrov spaces [3], for squared distance on the Heisenberg group [2], and recently for the squared distance on  $CD(K, N)$  and  $CD(K, \infty)$  spaces by Gigli [7] and for squared distance cost by Rajala and Ambrosio in a metric Riemannian like framework [1].

In this paper we show existence and uniqueness of optimal transport maps on proper, non-branching, metric measure spaces satisfying a new condition, Assumption 1, for cost functions of the form  $c(x, y) = h(d(x, y))$ , with  $h$  strictly convex and non-decreasing.

---

*Key words and phrases.* optimal transport; existence of maps; uniqueness of maps; measure contraction property.  
MH gratefully acknowledges funding through CRC 1060.

Assumption 1 does not imply any lower curvature bounds in the sense of Lott, Sturm and Villani. In particular in Section 3 we prove that Assumption 1 cannot imply the measure contraction property, MCP. On the other hand the measure contraction property implies Assumption 1. Therefore our result applies to spaces enjoying MCP, recovers most of the previously mentioned results and in many cases also extends them.

To our knowledge this is the first existence result of optimal maps in metric spaces for  $c(x, y) = h(d(x, y))$ , with  $h$  strictly convex and non-decreasing with no assumption on a lower bound on the Ricci curvature of the space. For  $h = id$ , existence of optimal maps, again with no assumption on the curvature of the metric space, has been obtained in [4].

The crucial idea for the proof of the main result is to *approximate the  $c$ -cyclically monotone set* on which the optimal measure is concentrated by means of a suitably chosen sequence of  $c$ -cyclically monotone sets representing transports into a discrete target.

We conclude this Introduction by describing the structure of the paper. In Section 2 we introduce the general setting of the paper, define Assumption 1 and state the two main results: the existence of optimal transport maps (Theorem 2.1) and the stability under changes in the reference measure of  $(X, d, m)$  of Assumption 1 (Theorem 2.2). In Section 3 we prove Theorem 2.2 while Section 4 and Section 5 are devoted to the proof of Theorem 2.1.

## 2. NOTATION AND MAIN RESULT

We now introduce the setting of this article. If not explicitly stated otherwise we will always assume to work in this framework.

Let  $(X, d, m)$  be a proper, non-branching, metric measure space, that is

- $(X, d)$  is a proper, complete and separable metric space with a non-branching geodesic structure;
- $m$  is a positive Borel measure, finite over compact sets whose support coincides with  $X$ .

In case we drop the proper assumption, we will refer to  $(X, d, m)$  just as non-branching metric measure space. Let  $\mu_0, \mu_1$  be probability measures over  $X$  and let  $h : [0, \infty) \rightarrow [0, \infty)$  be a strictly convex and non-decreasing map.

We study the following minimization problem

$$(2.1) \quad \min_{T_{\#}\mu_0 = \mu_1} \int h(d(x, T(x)))\mu_0(dx),$$

where  $T_{\#}\mu_0$  denotes the push forward of  $\mu_0$  under the map  $T$ . In the sequel, we will often denote the cost function  $h \circ d$  just with  $c$ . To get hands on the minimization problem (2.1) we also study its relaxed form, the Kantorovich problem. Let  $\Pi(\mu_0, \mu_1)$  be the set of transference plans, i.e.

$$\Pi(\mu_0, \mu_1) := \{\pi \in \mathcal{P}(X \times X) : (P_1)_{\#}\pi = \mu_0, (P_2)_{\#}\pi = \mu_1\},$$

where  $P_i : X \times X \rightarrow X$  is the projection map onto the  $i$ -th component,  $P_i(x_1, x_2) = x_i$  for  $i = 1, 2$ .

We will *always assume that  $\mu_0$  and  $\mu_1$  have finite  $c$ -transport distance* in the sense that

$$\inf \left\{ \int_{X \times X} h(d(x, y))\pi(dxdy) : \pi \in \Pi(\mu_0, \mu_1) \right\} < \infty.$$

Recall that a transference plan  $\pi \in \Pi(\mu_0, \mu_1)$  is said to be  $c$ -cyclically monotone if there exists  $\Gamma$  so that  $\pi(\Gamma) = 1$  and for every  $N \in \mathbb{N}$  and every  $(x_1, y_1), \dots, (x_N, y_N) \in \Gamma$  it holds

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

with  $x_{N+1} = x_1$ .

We also introduce a few objects connecting geodesics of the space  $X$  to optimal transport plans. Let

$$\mathcal{G}(X) \subset C([0, 1]; X),$$

be the set of geodesics endowed with the uniform topology inherited from  $C([0, 1]; X)$ . Being a closed subset of  $C([0, 1]; X)$ , it is Polish. For any  $t \in [0, 1]$  consider the map

$$e_t : \mathcal{G}(X) \rightarrow X,$$

the evaluation at time  $t$  defined by  $e_t(\gamma) = \gamma_t$ . For a subset  $A \subset X$  and a point  $x \in X$  the  $t$ -intermediate points between  $A$  and  $x$  are defined as

$$(2.2) \quad A_{t,x} := e_t(\{\gamma \in \mathcal{G}(X) : \gamma_0 \in A, \gamma_1 = x\}).$$

Assuming  $A$  compact, in a general non-branching metric measure space, the set  $A_{t,x}$  is closed. If we also assume the space to be proper, as we do here, the set  $A_{t,x}$ , being bounded, is indeed compact.

This evolution defined as (2.2) will play a fundamental role in our analysis. In particular we make the following

**Assumption 1.** A non-branching, metric measure space  $(X, d, m)$  verifies Assumption 1 if for every compact set  $K \subset X$  there exists a measurable function  $f : [0, 1] \rightarrow (0, 1]$  with

$$\limsup_{t \rightarrow 0} f(t) > \frac{1}{2},$$

and a positive  $\delta \leq 1$  such that

$$m(A_{t,x}) \geq f(t) \cdot m(A), \quad \forall 0 \leq t \leq \delta,$$

for any compact set  $A \subset K$  and any base point  $x \in K$ .

We can now state the main result of this paper.

**Theorem 2.1.** *Let  $(X, d, m)$  be a proper, non-branching, metric measure space verifying Assumption 1. Let  $\mu_0$  and  $\mu_1$  be two probability measures over  $X$  with finite  $c$ -transport distance. If  $\mu_0 \ll m$  and  $h$  is strictly-convex and non-decreasing, the optimal transport problem associated to (2.1) has a unique solution induced by a map.*

In detail we will prove that if  $\mu_0 \ll m$  then any  $c$ -cyclically monotone plan  $\pi$  is induced by a map  $T : X \rightarrow X$ . With  $\pi$  induced by a map we mean that  $\pi = (id, T)_\# \mu_0$ . This implies that the two minimization problems (1.1) and (1.2) coincide. Then a direct Corollary of this result is the uniqueness of the optimal coupling. We will prove the claim by showing that branching at starting points does not happen almost surely.

Regarding Assumption 1, we will prove the following result, that can be understood as a stability property. Here the space is not needed to be proper.

**Theorem 2.2.** *Let  $(X, d, m)$  be a non-branching metric measure space verifying Assumption 1. Consider a continuous function  $g : X \rightarrow (0, \infty)$  and the measure  $\tilde{m} := g \cdot m$ . Then  $(X, d, \tilde{m})$  is a non-branching metric measure space verifying Assumption 1.*

## 3. ON ASSUMPTION 1

It is clear that spaces satisfying the *measure contraction property* – for a definition we refer to [11, 13] – also satisfy Assumption 1. However, as we will prove in this section, Assumption 1 does not imply the measure contraction property or, more in general, any synthetic Ricci curvature bounds.

In detail, we will show that if  $(X, d, m)$  is a non-branching metric measure space verifying Assumption 1 and  $\tilde{m} = gm$ , with  $g$  continuous and strictly positive, then also  $(X, d, \tilde{m})$  verifies Assumption 1. Since this kind of changes in the measure destroy Ricci lower bounds, Assumption 1 cannot imply any of them. See [13], Theorem 1.7.

The setting of this subsection is slightly different from the remaining of this note, so we will specify all the assumptions needed in each statement. We start with two simple lemmas.

**Lemma 3.1.** *Let  $(X, d, m)$  be a metric measure space. For any compact set  $K$  and any  $\varepsilon > 0$  there exists  $n \in \mathbb{N}$  and  $K_i \subset K$  compact for  $i = 1, \dots, n$  such that*

$$\text{diam}(K_i) \leq \varepsilon, \quad m\left(K \setminus \bigcup_{i=1}^n K_i\right) \leq \varepsilon,$$

and  $K_i \cap K_j = \emptyset$  for  $i \neq j$ .

*Proof.* So let  $\varepsilon > 0$  be given. Then consider the open covering of  $K$  given by  $\{B_\varepsilon(x)\}_{x \in K}$ . By compactness, there exists finitely many  $\{x_i\}_{i \leq n}$  so that every point of  $K$  is at distance less than  $\varepsilon$  for some  $x_i$ . Then consider the compact sets  $H_i := K \cap \overline{B}_\varepsilon(x_i)$  for  $i = 1, \dots, n$ . Clearly the union of all  $H_i$  covers  $K$  and each of  $H_i$  has diameter less than  $\varepsilon$ . Taking differences we can pass to a family of Borel sets  $\hat{H}_i$  so that

$$\text{diam}(\hat{H}_i) \leq \varepsilon, \quad \bigcup_{i=1}^n \hat{H}_i = K.$$

with  $\hat{H}_i \cap \hat{H}_j = \emptyset$  if  $i \neq j$ . Then by inner regularity with compact sets, choose for each  $i \leq n$  a compact set  $K_i \subset \hat{H}_i$  so that  $m(\hat{H}_i \setminus K_i) \leq \varepsilon/n$ . The claim follows.  $\square$

**Lemma 3.2.** *Let  $(X, d, m)$  be a non-branching metric measure space. Suppose that for each  $K \subset X$  compact there exist  $\delta, \varepsilon > 0$  and a measurable function  $f : [0, \delta] \rightarrow (0, \infty)$  with  $\limsup_{t \rightarrow 0} f(t) > 1/2$ , so that*

$$m(A_{t,x}) \geq f(t)m(A), \quad \forall t \in [0, \delta],$$

for any  $x \in K$  and  $A \subset K$  compact with  $\text{diam}(A) \leq \varepsilon$ . Then  $(X, d, m)$  verifies Assumption 1.

*Proof.* Consider  $K \subset X$  compact set. Let  $\delta, \varepsilon > 0$  and the measurable map  $f$  given by the hypothesis. Fix also  $x \in K$ . Let  $A \subset K$  be any compact set. Now for any  $\eta < \varepsilon$  consider the finite family of disjoint compact  $\{A_i\}_{i \leq n(\eta)}$  sets given by Lemma 3.1. Then since the space is non-branching and  $\text{diam}(A_i) \leq \varepsilon$  it follows that

$$\begin{aligned} m(A_{t,x}) &= \sum_{i \leq n(\eta)} m((A_i)_{t,x}) \\ &\geq f(t) \sum_{i \leq n(\eta)} m(A_i) \\ &\geq f(t)m(A) - \eta f(t), \end{aligned}$$

for all  $t \in [0, \delta]$ . Since  $\eta$  was any positive number less than  $\varepsilon$  and  $\delta$  depends only on  $K$  and  $\varepsilon$ , the claim follows.  $\square$

It follows from Lemma 3.2 that to verify Assumption 1 it is sufficient to consider compact sets of small diameter. This already suggests that Assumption 1 is stable under continuous changes of the measure as the one we proposed few lines above. We now state and prove this stability property.

**Theorem 3.3.** *Let  $(X, d, m)$  be a non-branching metric measure space verifying Assumption 1. Consider a continuous function  $g : X \rightarrow (0, \infty)$  and the measure  $\tilde{m} := g \cdot m$ . Then  $(X, d, \tilde{m})$  is a non-branching metric measure space verifying Assumption 1.*

*Proof. Step 1.* Note first, that by continuity of  $g$ ,  $\tilde{m}$  is finite over compact sets and therefore  $(X, d, \tilde{m})$  is a non-branching, metric measure space. Let  $K \subset X$  be any compact set and  $\delta > 0$  and  $f$  measurable be given by Assumption 1 for  $(X, d, m)$ . Note that  $\text{diam}(K)$  is bounded, say by  $M > 0$ . Then for any  $A \subset K$  compact,  $x \in K$  and  $t \in [0, \delta]$  the following chain of inequalities holds:

$$\begin{aligned} \tilde{m}(A_{t,x}) &= \int_{A_{t,x}} g(x)m(dx) \\ &\geq \inf\{g(x) : x \in A_{t,x}\} m(A_{t,x}) \\ &\geq \inf\{g(x) : x \in A_{t,x}\} f(t)m(A) \\ &\geq \frac{\inf\{g(x) : x \in A_{t,x}\}}{\max\{g(x) : x \in A\}} f(t) \int_A g(x)m(dx) \\ &= \frac{\inf\{g(x) : x \in A_{t,x}\}}{\max\{g(x) : x \in A\}} f(t)\tilde{m}(A). \end{aligned}$$

Moreover from Lemma 3.2 it follows that we can focus only on compact  $A$  with arbitrarily small diameter.

*Step 2.* Then we reason as follows: consider  $\eta > 0$  so that

$$\left(1 - \frac{\eta}{\alpha}\right) \limsup_{t \rightarrow 0} f(t) > \frac{1}{2},$$

where  $\alpha > 0$  is so that  $g(x) > \alpha$  for all  $x \in K$ . Then since  $g$  is uniformly continuous over  $K$ , there exists  $\varepsilon > 0$  so that  $|g(z) - g(w)| \leq \eta$  whenever  $d(z, w) \leq 2\varepsilon$  for  $z, w \in K$ .

Let now  $A \subset K$  be any compact set with  $\text{diam}(A) \leq \varepsilon$  and take  $t \leq \min\{\delta, \varepsilon/M\}$ . Then if  $z \in A_{t,x}$  and  $w \in A$ , it follows that  $d(z, w) \leq 2\varepsilon$ : indeed there exists a geodesic  $\gamma$  so that  $\gamma_0 \in A$ ,  $\gamma_1 = x$  and  $\gamma_t = z$ , then

$$d(z, w) \leq d(z, \gamma_0) + d(\gamma_0, w) \leq t \cdot \text{diam}(K) + \varepsilon \leq 2\varepsilon.$$

Then if  $A \subset K$  compact set with  $\text{diam}(A) \leq \varepsilon$ ,  $x \in K$  and  $t \leq \min\{\delta, \varepsilon/M\}$  and  $x_M \in A$  so that  $g(x_M) = \max\{g(x) : x \in A\}$ , we have:

$$1 - \frac{\inf\{g(x) : x \in A_{t,x}\}}{\max\{g(x) : x \in A\}} = \frac{\sup\{g(x_M) - g(z) : z \in A_{t,x}\}}{g(x_M)} \leq \frac{\eta}{\alpha},$$

and therefore

$$\tilde{m}(A_{t,x}) \geq \left(1 - \frac{\eta}{\alpha}\right) f(t)\tilde{m}(A).$$

By the choice of  $\eta$  we have proved Assumption 1 for all compact sets with diameter smaller than  $\varepsilon$ . Lemma 3.2 gives the claim.  $\square$

Nevertheless if  $(X, d, m)$  is also proper, Assumption 1 carries some geometric property of the space.

**Proposition 3.4.** *Any proper, non-branching, metric measure space  $(X, d, m)$  satisfying Assumption 1 is locally doubling.*

*Proof.* Take any ball  $B_{2r}$  of radius  $2r$ . Fix  $0 < t \leq \delta$  and  $n$  such that  $(1-t)^n \leq 1/2$ . Contracting  $B_{2r}$  to its center yields  $(B_{2r})_t = B_{(1-t)2r}$ . Contracting  $B_{(1-t)2r}$  to its center yields  $(B_{(1-t)2r})_t = B_{(1-t)^2 2r}$ . Since  $(X, d, m)$  is proper we can use Assumption 1 and estimate

$$m(B_{(1-t)^2 2r}) \geq f(t)m(B_{(1-t)2r}) \geq f(t)^2 m(B_{2r}).$$

Repeating this another  $n-2$  times yields

$$m(B_r) \geq m(B_{(1-t)^n 2r}) \geq f(t)^n m(B_{2r}).$$

□

**Remark 3.5.** Assume that  $(X, d, m)$  is locally doubling and for any compact set  $K \subset X$  there exists  $0 < \delta \leq 1$  such that for any  $t \leq \delta$  there exists a map  $F_t : K \times K \rightarrow X$  such that

$$d(x, F_t(x, y)) = td(x, y), \quad \frac{1}{L(t)} d(x, z) \leq d(F_t(x, y), F_t(z, y)) \leq L(t) d(x, z)$$

and  $L(t) \rightarrow 1$  as  $t$  goes to 0. Moreover, assume that  $F_t$  varies continuously in time and for all compact sets  $K$ :

$$\limsup_{t \rightarrow 0} \inf \left\{ \frac{m(B_r(F_t(x, y)))}{m(B_r(x))} : x, y \in K, r > 0 \right\} > \frac{1}{2}.$$

Then it is not hard to show using covering theorems that  $(X, d, m)$  verifies Assumption 1.

This says that a certain type of Ahlfors regularity together with a bi-Lipschitz selection of  $t$ -intermediate points implies Assumption 1.

#### 4. EVOLUTION ESTIMATES

Following Section 2, we fix once for all  $(X, d, m)$  a proper, non-branching, metric measure space verifying Assumption 1, two probability measures  $\mu_0, \mu_1$  with  $\mu_0 \ll m$  and  $h : [0, \infty) \rightarrow [0, \infty)$  strictly convex and non-decreasing.

Since we have to prove a local property, we can assume that  $\text{supp}(\mu_0), \text{supp}(\mu_1) \subset K$  with  $K$  compact. Then by standard results in optimal transportation, there exists a couple of Kantorovich potentials  $(\varphi, \varphi^c)$  such that if

$$\Gamma := \{(x, y) \in X \times X : \varphi(x) + \varphi^c(y) = c(x, y)\},$$

then the transport plan  $\pi$  is optimal iff  $\pi(\Gamma) = 1$  (e.g. see Theorem 5.10 in [14]). Note also that the set  $\Gamma$  is  $c$ -cyclically monotone. So also  $K, \varphi, \varphi^c$  and  $\Gamma$  are fixed.

We start by proving the standard property of geodesics belonging to the support of the optimal dynamical transference plan  $\pi$ : they cannot meet at the same time  $t$  if  $t \neq 0, 1$ . For existence results and details on dynamical transference plans we refer to [14] Chapter 7.

**Lemma 4.1.** *Let  $(x_0, y_0), (x_1, y_1) \in \Gamma$  be two distinct points. Then for any  $t \in (0, 1)$ ,*

$$d(x_0(t), x_1(t)) > 0,$$

where  $x_i(t)$  is any  $t$ -intermediate point between  $x_i$  and  $y_i$ , for  $i = 0, 1$ .

*Proof.* Assume by contradiction the existence of  $x_0(t) = x_1(t) \in X$ ,  $t$ -intermediate points of  $(x_0, y_0)$  and  $(x_1, y_1)$ , i.e.

$$d(x_0, x_0(t)) = td(x_0, y_0), \quad d(x_0(t), y_0) = (1-t)d(x_0, y_0),$$

and

$$d(x_1, x_1(t)) = td(x_1, y_1), \quad d(x_1(t), y_1) = (1-t)d(x_1, y_1).$$

Case 1:  $d(x_0, y_0) \neq d(x_1, y_1)$ . Then

$$\begin{aligned} h(d(x_0, y_1)) + h(d(x_1, y_0)) &\leq h(d(x_0, x_0(t)) + d(x_1(t), y_1)) + h(d(x_1, x_1(t)) + d(x_0(t), y_0)) \\ &< th(d(x_0, y_0)) + (1-t)h(d(x_1, y_1)) \\ &\quad + th(d(x_1, y_1)) + (1-t)h(d(x_0, y_0)) \\ &= h(d(x_0, y_0)) + h(d(x_1, y_1)). \end{aligned}$$

Where between the first and the second line we have used the strict convexity of  $h$ . From  $c$ -cyclical monotonicity we have a contradiction.

Case 2.  $d(x_0, y_0) = d(x_1, y_1)$ . Let  $\gamma^0, \gamma^1 \in \mathcal{G}(X)$  be such that

$$\gamma_0^0 = x_0, \quad \gamma_t^0 = x_0(t), \quad \gamma_1^0 = y_0, \quad \gamma_0^1 = x_1, \quad \gamma_t^1 = x_1(t), \quad \gamma_1^1 = y_1,$$

and define the curve  $\gamma : [0, 1] \rightarrow X$  by

$$\gamma_t := \begin{cases} \gamma_s^0, & s \in [0, t] \\ \gamma_s^1, & s \in [t, 1]. \end{cases}$$

Then  $\gamma$  is a geodesic different from  $\gamma^0$  but coinciding with it on the non trivial interval  $[0, t]$ . Since this is a contradiction with the non-branching assumption, the claim is proved.  $\square$

**Remark 4.2.** In the framework of metric measure spaces enjoying synthetic Ricci curvature bounds, like  $\text{CD}(K, N)$ , see [13] for its definition, it has recently been shown by Rajala that assuming the convexity of the entropy along all  $L^2$ -Wasserstein geodesics implies that any optimal transport plan is concentrated on a family of non-branching geodesics, even if the space is not assumed to be non-branching. Unfortunately, in our framework such a technique cannot be used, at least for now. Indeed while all the curvature information are stated in terms of  $L^2$ -Wasserstein geodesics, here we would need a non-branching property of the geodesics of the space  $X$  with final and initial points forming a  $c$ -cyclically monotone set. The latter property cannot be deduced straightforwardly by  $d^2$ -monotonicity. For the moment the only result going in this direction is for  $h = id$  and it is proven in [6].

For any compact set  $\Lambda \subset X \times X$  we can now consider the associated evolution map. For every  $t \in [0, 1]$  and every  $A \subset X$  compact set

$$A_{t, \Lambda} := e_t((e_0, e_1)^{-1}((A \times X) \cap \Lambda)).$$

It is easily seen that  $A_{t, \Lambda}$  is a closed and bounded set. Hence since  $(X, d, m)$  is proper we also obtain compactness of  $A_{t, \Lambda}$ . Moreover we will use the following notation: to any  $\Lambda \subset X \times X$  we associate the following set:

$$(4.1) \quad \hat{\Lambda} := (P_1(\Lambda) \times P_2(\Lambda)) \cap \Gamma.$$

We are now ready to prove the main consequence of Assumption 1.

**Proposition 4.3.** *For any  $\Lambda \subset \Gamma$  compact the following inequality holds:*

$$(4.2) \quad m(A_{t, \hat{\Lambda}}) \geq f(t)m(A), \quad t \in [0, \delta],$$

for any  $A \subset P_1(\Lambda)$ .

*Proof. Step 1.* Let  $\{y_i\}_{i \in \mathbb{N}} \subset P_2(\Lambda)$  be a dense set in  $P_2(\Lambda)$ .

Consider the following family of sets: for  $n \in \mathbb{N}$  and  $i \leq n$

$$E_n(i) := \{x \in P_1(\Lambda) : c(x, y_i) - \varphi^c(y_i) \leq c(x, y_j) - \varphi^c(y_j), j = 1, \dots, n\}.$$

If we now consider

$$\Lambda_n := \bigcup_{i=1}^n E_n(i) \times \{y_i\},$$

it is straightforward to check that  $P_1(\Lambda_n) = P_1(\Lambda)$  and  $\Lambda_n$  is  $c$ -cyclically monotone. Indeed, for any  $(x_1, y_1), \dots, (x_m, y_m) \in \Lambda_n$ , by definition it holds that

$$c(x_i, y_i) - \varphi^c(y_i) \leq c(x_i, y_{i+1}) - \varphi^c(y_{i+1}), \quad i = 1, \dots, m.$$

Taking the sum over  $i$ , the property follows.

By Assumption 1 there exists  $f : [0, 1] \rightarrow \mathbb{R}$  measurable with  $f(0) > 1/2$ , independent of the sequence  $\{y_i\}_{i \in \mathbb{N}}$  and of  $n$ , such that for any  $A \subset P_1(\Lambda)$  compact it holds that

$$m\left((A \cap E_n(i))_{t, y_i}\right) \geq f(t)m(A \cap E_n(i)), \quad \forall t \in [0, \delta],$$

where  $(A \cap E_n(i))_{t, y_i} = (A \cap E_n(i))_{t, E_n(i) \times \{y_i\}}$ . Note that since  $A = \bigcup_{i \leq n} A \cap E_n(i)$  it follows that

$$\begin{aligned} A_{t, \Lambda_n} &= e_t((e_0, e_1)^{-1}((A \times X) \cap \Lambda_n)) \\ &= \bigcup_{i \leq n} e_t((e_0, e_1)^{-1}((A \cap E_n(i)) \times X) \cap \Lambda_n) \\ &= \bigcup_{i \leq n} (A \cap E_n(i))_{t, \Lambda_n} \\ &\supset \bigcup_{i \leq n} (A \cap E_n(i))_{t, E_n(i) \times \{y_i\}}. \end{aligned}$$

Moreover, Lemma 4.1 implies

$$(A \cap E_n(i))_{t, y_i} \cap (A \cap E_n(j))_{t, y_j} = \emptyset, \quad i \neq j,$$

for all  $t \in (0, 1)$ .

Then it holds for all  $t \in [0, \delta]$ :

$$\begin{aligned} m(A_{t, \Lambda_n}) &\geq m\left(\bigcup_{i=1}^n (A \cap E_n(i))_{t, E_n(i) \times \{y_i\}}\right) = \sum_{i=1}^n m\left((A \cap E_n(i))_{t, y_i}\right) \\ &\geq f(t) \sum_{i=1}^n m(A \cap E_n(i)) \\ &\geq f(t)m\left(\bigcup_{i=1}^n A \cap E_n(i)\right) \\ (4.3) \quad &= f(t)m(A). \end{aligned}$$

*Step 2.* Note that for every  $n \in \mathbb{N}$ ,  $\Lambda_n \subset \text{supp}(\mu_0) \times \text{supp}(\mu_1)$  and the latter, by assumption, is a subset of  $K \times K$ . Since the space of closed subsets of  $K \times K$  endowed with the Hausdorff metric  $(\mathcal{C}(K \times K), d_{\mathcal{H}})$  is a compact space, there exists a subsequence  $\{\Lambda_{n_k}\}_{k \in \mathbb{N}}$  and  $\Theta \subset K \times K$  compact such that

$$\lim_{k \rightarrow \infty} d_{\mathcal{H}}(\Lambda_{n_k}, \Theta) = 0.$$

Since the sequence  $\{y_i\}_{i \in \mathbb{N}}$  is dense in  $P_2(\Lambda)$  and  $\Lambda \subset \Gamma$  is compact, by definition of  $E_n(i)$ , necessarily for every  $(x, y) \in \Theta$  it holds

$$\varphi(x) + \varphi^c(y) = c(x, y), \quad x \in P_1(\Lambda), \quad y \in P_2(\Lambda).$$

Hence  $\Theta \subset (P_1(\Lambda) \times P_2(\Lambda)) \cap \Gamma = \hat{\Lambda}$ . To conclude the proof we have to observe



$$m(A_{t,\Theta}) \geq \limsup_{k \rightarrow \infty} m(A_{t,\Lambda_{n_k}}).$$

Indeed, since  $A_{t,\Theta}$  is a compact set, it follows that if  $A_{t,\Theta}^\varepsilon = \{x \in X : d(x, A_{t,\Theta}) \leq \varepsilon\}$ , then for  $k$  sufficiently big  $A_{t,\Lambda_{n_k}} \subset A_{t,\Theta}^\varepsilon$  and  $m(A_{t,\Theta}^\varepsilon)$  converges to  $m(A_{t,\Theta})$ .

Then

$$m(A_{t,\hat{\Lambda}}) \geq \limsup_{k \rightarrow \infty} m(A_{t,\Lambda_{n_k}}) \geq f(t)m(A),$$

and the claim follows.  $\square$

## 5. EXISTENCE OF OPTIMAL MAPS

In this section we show that branching at starting points does not happen almost surely. Recall that

$$\Gamma = \{(x, y) \in X \times X : \varphi(x) + \varphi^c(y) = c(x, y)\}$$

and any optimal transport plan is concentrated on  $\Gamma$ .

**Lemma 5.1.** *Let  $\Lambda_1, \Lambda_2 \subset \Gamma$  be compact sets such that*

- i)  $P_1(\Lambda_1) = P_1(\Lambda_2)$ ;*
- ii)  $P_2(\Lambda_1) \cap P_2(\Lambda_2) = \emptyset$ .*

*Then  $m(P_1(\Lambda_1)) = m(P_1(\Lambda_2)) = 0$ .*

*Proof.* Note that since  $P_2(\Lambda_1) \cap P_2(\Lambda_2) = \emptyset$ , necessarily  $\hat{\Lambda}_1 \cap \hat{\Lambda}_2 = \emptyset$ , where  $\hat{\Lambda}_i$  are defined by (4.1), for  $i = 1, 2$ . Hence from Lemma 4.1, for every  $A \subset P_1(\Lambda_1) = P_1(\Lambda_2)$

$$A_{t,\hat{\Lambda}_1} \cap A_{t,\hat{\Lambda}_2} = \emptyset,$$

for every  $t \in (0, 1)$ . Then let  $A := P_1(\Lambda_1) = P_1(\Lambda_2)$  and recall that as  $t \rightarrow 0$  the sets  $A_{t,\hat{\Lambda}_1}$  and  $A_{t,\hat{\Lambda}_2}$  both converge in Hausdorff topology to  $A$ . Put  $A^\varepsilon = \{x : d(x, A) \leq \varepsilon\}$ . Then it follows from Proposition 4.3 that

$$\begin{aligned} m(A) &= \limsup_{\varepsilon \rightarrow 0} m(A^\varepsilon) \geq \limsup_{t \rightarrow 0} m(A_{t,\hat{\Lambda}_1} \cup A_{t,\hat{\Lambda}_2}) \\ &= \limsup_{t \rightarrow 0} (m(A_{t,\hat{\Lambda}_1}) + m(A_{t,\hat{\Lambda}_2})) \\ &\geq m(A) \limsup_{t \rightarrow 0} 2f(t) = \alpha \cdot m(A), \end{aligned}$$

with  $\alpha > 1$ . Hence, necessarily  $m(P_1(\Lambda_1)) = m(P_1(\Lambda_2)) = m(A) = 0$ , and the claim follows.  $\square$

We will use the following notation:  $\Gamma(x) := (\{x\} \times X) \cap \Gamma$  and given a set  $\Theta \subset X \times X$  we say that  $T$  is a selection of  $\Theta$  if  $T : P_1(\Theta) \rightarrow X$  is  $m$ -measurable and  $\text{graph}(T) \subset \Theta$ .

**Proposition 5.2.** *Consider the sets*

$$E := \{x \in P_1(\Gamma) : \Gamma(x) \text{ is not a singleton}\}, \quad \Gamma_E := \Gamma \cap (E \times X).$$

*Then for any selection  $T$  of  $\Gamma_E$  and every  $\pi \in \Pi(\mu_0, \mu_1)$  with  $\pi(\Gamma) = 1$  it holds*

$$\pi(\Gamma_E \setminus \text{graph}(T)) = 0.$$

*Proof. Step 1.* Suppose by contradiction the existence of  $\pi \in \Pi(\mu_0, \mu_1)$  with  $\pi(\Gamma) = 1$  and of a selection  $T$  of  $\Gamma_E$  such that

$$\pi(\Gamma_E \setminus \text{graph}(T)) = \beta > 0.$$

By inner regularity, to prove the complete statement it is enough to prove it under the additional assumptions that  $E$  is compact and  $T$  is continuous.

Note that

$$\Gamma_E \setminus \text{graph}(T) = \bigcup_{n=1}^{\infty} \{(x, y) \in \Gamma_E : d(y, T(x)) \geq 1/n\}.$$

Hence, there exists  $n \in \mathbb{N}$  such that

$$\pi(\{(x, y) \in \Gamma_E : d(y, T(x)) \geq 1/n\}) \geq \beta' > 0.$$

Put  $\Lambda := \{(x, y) \in \Gamma_E : d(y, T(x)) \geq 1/n\}$ . Note that  $m(P_1(\Lambda)) > 0$ .

*Step 2.* From the continuity of  $T$  it follows the existence of  $\eta > 0$  so that if  $d(x, z) \leq \eta$  then  $d(T(x), T(z)) \leq 1/2n$ . Clearly we can take  $x \in P_1(\Lambda)$  so that

$$m(P_1(\Lambda) \cap \bar{B}_\eta(x)) > 0.$$

where  $\bar{B}_\eta(x)$  denotes the closed ball of radius  $\eta$  around  $x$ . So consider the two sets

$$\Xi_1 := \text{graph}(T) \cap ((\bar{B}_\eta(x) \cap P_1(\Lambda)) \times X), \quad \Xi_2 := (\bar{B}_\eta(x) \times X) \cap \Lambda.$$

By construction  $\Xi_1, \Xi_2 \subset \Gamma$  and

$$P_1(\Xi_1) = P_1(\Xi_2) = P_1(\Lambda) \cap \bar{B}_\eta(x),$$

therefore  $m(P_1(\Xi_1)) > 0$ .

Moreover for any  $y \in P_2(\Xi_2)$  there exists  $w \in \bar{B}_\eta(x)$  so that

$$d(y, T(w)) \geq \frac{1}{n}.$$

Hence for any  $z \in \bar{B}_\eta(x)$  it holds

$$d(y, T(z)) \geq d(y, T(w)) - d(T(w), T(z)) \geq \frac{1}{n} - \frac{1}{2n} = \frac{1}{2n}.$$

Hence

$$P_2(\Xi_1) \cap P_2(\Xi_2) = \emptyset.$$

Since this is in contradiction with Lemma 5.1, the claim is proved.  $\square$

We can now state the main result of this paper whose proof now follows as a straightforward corollary of what we proved so far.

**Theorem 5.3.** *Let  $(X, d, m)$  be a non-branching metric measure space verifying Assumption 1. Let  $\mu_0$  and  $\mu_1$  be two probability measures over  $X$  with finite  $c$ -transport distance. If  $\mu_0 \ll m$  and  $h$  is strictly-convex and non-decreasing, for any  $\pi \in \Pi(\mu_0, \mu_1)$  such that  $\pi(\Gamma) = 1$  there exists an  $m$ -measurable map  $T : X \rightarrow X$  such that*

$$\pi(\text{graph}(T)) = 1.$$

*Proof.* Let  $\pi \in \Pi(\mu_0, \mu_1)$  be any transference plan so that  $\pi(\Gamma) = 1$ . As for Proposition 5.2, consider the sets

$$E := \{x \in P_1(\Gamma) : \Gamma(x) \text{ is not a singleton}\}, \quad \Gamma_E := \Gamma \cap (E \times X).$$

Since

$$\Gamma_E = P_{12}(\{(x, y, z, w) \in \Gamma \times \Gamma : d(x, z) = 0, d(y, w) > 0\}),$$

the set  $\Gamma_E$  is an analytic set. For the definition of analytic set, see Chapter 4 of [12]. We can then use the Von Neumann Selection Theorem for analytic sets, see Theorem 5.5.2 of [12], to obtain a map  $T : E \rightarrow X$ ,  $\mathcal{A}$ -measurable, where  $\mathcal{A}$  is the  $\sigma$ -algebra generated by analytic sets, so that  $(x, T(x)) \in \Gamma_E$ .

Then Proposition 5.2 implies that

$$\pi \llcorner_{\Gamma_E} = (Id, T) \# \mu_{0 \llcorner E}.$$

Since on  $\Gamma \setminus \Gamma_E$   $\pi$  is already supported on a graph, the claim follows.  $\square$

This directly implies

**Corollary 5.4.** *Under the assumptions of Theorem 5.3, there is a unique optimal transport map.*

*Proof.* The last theorem shows that every optimal coupling is induced by a transport map. As the set of all optimal couplings is convex this directly implies the uniqueness.  $\square$

#### REFERENCES

- [1] L. Ambrosio and T. Rajala. Slopes of Kantorovich potentials and existence of optimal transport maps in metric measure spaces. *Annali di Matematica Pura ed Applicata*, pages 1–17, 2011.
- [2] L. Ambrosio and S. Rigot. Optimal mass transportation in the Heisenberg group. *Journal of Functional Analysis*, 208(2):261–301, 2004.
- [3] J. Bertrand. Existence and uniqueness of optimal maps on Alexandrov spaces. *Advances in Mathematics*, 219(3):838–851, 2008.
- [4] S. Bianchini and F. Cavalletti. The Monge problem for distance cost in geodesic spaces. *Commun. Math. Phys.*, 318:615–673, 2013.
- [5] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.
- [6] F. Cavalletti. Monge problem in metric measure spaces with Riemannian curvature-dimension condition. *Nonlinear Analysis*, 99:136–151, 2014.
- [7] N. Gigli. Optimal maps in non branching spaces with Ricci curvature bounded from below. *Geometric And Functional Analysis*, pages 1–10, 2011.
- [8] LV Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133(4):1381–1382, 2006.
- [9] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geometric And Functional Analysis*, 11(3):589–608, 2001.
- [10] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. De l’Imprimerie Royale, 1781.
- [11] S.-I. Ohta. On the measure contraction property of metric measure spaces. *Comment. Math. Helv.*, 82:805–828, 2007.
- [12] A. M. Srivastava. *A course on Borel sets*. Springer, 1998.
- [13] K.T. Sturm. On the geometry of metric measure spaces.II. *Acta Math.*, 196(1):133–177, 2006.
- [14] C. Villani. *Optimal transport, old and new*. Springer, 2008.

RWTH, DEPARTMENT OF MATHEMATICS, TEMPLERGRABEN 64, D-52062 AACHEN (GERMANY)  
*E-mail address:* `cavalletti@instmath.rwth-aachen.de`

UNIVERSITÄT BONN, INSTITUT FÜR ANGEWANDTE MATHEMATIK, ENDENICHER ALLEE 60, D-53115 BONN (GERMANY)  
*E-mail address:* `huesmann@iam.uni-bonn.de`