

# NON-BRANCHING GEODESICS AND OPTIMAL MAPS IN STRONG $CD(K, \infty)$ -SPACES

TAPIO RAJALA AND KARL-THEODOR STURM

ABSTRACT. We prove that in metric measure spaces where the entropy functional is  $K$ -convex along every Wasserstein geodesic any optimal transport between two absolutely continuous measures with finite second moments lives on a non-branching set of geodesics. As a corollary we obtain that in these spaces there exists only one optimal transport plan between any two absolutely continuous measures with finite second moments and this plan is given by a map.

The results are applicable in metric measure spaces having Riemannian Ricci curvature bounded below, and in particular they hold also for Gromov-Hausdorff limits of Riemannian manifolds with Ricci curvature bounded from below by some constant.

## 1. INTRODUCTION

Ricci curvature lower bounds in general metric measure spaces were studied by the second author in [16, 17] and at the same time with a similar approach by Lott and Villani in [12]. There the lower bound  $K \in \mathbb{R}$  on the Ricci curvature without a reference to the dimension of the metric measure space was defined as  $K$ -convexity of the entropy functional along Wasserstein geodesics (see Section 2 for details). These spaces are called (weak)  $CD(K, \infty)$ -spaces. The word *weak* is sometimes used to emphasize that the convexity is required only along one geodesic between any two given probability measures.

A property of the  $CD(K, \infty)$ -spaces which complicates the theory is the possibility to have branching geodesics. For example the space  $\mathbb{R}^2$  with the  $l_\infty$ -norm is a  $CD(0, \infty)$ -space and it has lots of branching geodesics. Being the limit as  $p \rightarrow \infty$  (in any reasonable sense) of the spaces  $\mathbb{R}^2$  with the  $l_p$ -norm – which are non-branching  $CD(K, \infty)$ -spaces – this example in particular illustrates that being non-branching is not a stable property. A number of results in  $CD(K, \infty)$ -spaces have been proven only under the extra assumption that there are no branching geodesics in the space. Although in some of the results this assumption has recently been removed (see for example [15, 14]), in many it still remains.

Because branching geodesics are hard to deal with, it is reasonable to consider more restrictive definitions that exclude spaces with branching geodesics or at least limit the amount of branching that can occur. One of the essential properties of the definitions of Ricci curvature lower bounds in metric measure spaces is the stability under the measured Gromov-Hausdorff convergence. Therefore any stable definition that extends the Riemannian case should include the limit spaces of Riemannian manifolds with uniform Ricci curvature lower bounds. While Riemannian manifolds are known to be non-branching, to our knowledge it is not known whether this holds for their limit spaces. The spaces we consider in this paper cover also these limit spaces. Although our result does not rule out the possibility to have some

---

*Date:* July 16, 2013.

*2000 Mathematics Subject Classification.* Primary 53C23. Secondary 28A33, 49Q20.

*Key words and phrases.* Ricci curvature, metric measure spaces, branching geodesics.

branching geodesics, it still says that there has to be so few branching geodesics that optimal transports between any two absolutely continuous measures do not see them.

Before stating our main result we fix some terminology. In this paper we will always assume  $(X, d)$  to be a complete separable geodesic metric space and  $\mathfrak{m}$  to be a locally finite Borel measure. We refer to Section 2 for some details on optimal mass transportation and  $CD(K, \infty)$  condition in such spaces. We call a space  $(X, d, \mathfrak{m})$  *essentially non-branching* if for every  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$  which are absolutely continuous with respect to  $\mathfrak{m}$  we have that any  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  is concentrated on a set of non-branching geodesics. The space  $(X, d, \mathfrak{m})$  is said to be a strong  $CD(K, \infty)$ -space if the entropy  $\text{Ent}_{\mathfrak{m}}$  is  $K$ -convex along every  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  for every  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ .

**Theorem 1.1.** *Every strong  $CD(K, \infty)$ -space is essentially non-branching.*

One of the cases which is covered by Theorem 1.1 are the  $RCD(K, \infty)$ -spaces, that is, spaces with Riemannian Ricci curvature bounded from below by some constant  $K \in \mathbb{R}$ . These spaces were recently defined and studied in [3, 4], see also [2]. They were obtained by reinforcing the  $CD(K, \infty)$ -spaces with a requirement that the local structure of the space must be Hilbertian. The Hilbertian structure immediately rules out spaces like the above mentioned  $\mathbb{R}^2$  with the  $l_\infty$ -norm. In [4] it was shown that one of the equivalent formulations of  $RCD(K, \infty)$ -spaces is that any probability measure with finite second moment is the starting point of an  $\text{EVI}_K$ -gradient flow of the entropy. This condition is known to imply  $K$ -convexity of the entropy along every geodesic [8].

In [4] it was also proven that the definition of  $RCD(K, \infty)$  with a finite reference measure is stable under the measured Gromov-Hausdorff convergence (or under the  $\mathbb{D}$ -convergence introduced in [16]). Later in [10] the stability was proven with more general reference measures. When we combine the stability with Theorem 1.1 we arrive at the following corollary which in fact was our main motivation to write this paper.

**Corollary 1.2.** *The  $RCD(K, \infty)$  condition is stable under measured Gromov-Hausdorff convergence (or under the  $\mathbb{D}$ -convergence) and it implies essential non-branching.*

As we already mentioned, we are not aware of any other non-branching results even for the Gromov-Hausdorff limits of Riemannian manifolds with Ricci curvature bounded from below by some constant.

Another corollary of our result is that the  $RCD(K, \infty)$  condition implies the formulation of  $CD(K, \infty)$  that was used by Lott and Villani [12]. They required convexity type inequalities for a class of functionals instead of just  $\text{Ent}_{\mathfrak{m}}$ , and hence their definition was at least a priori more restrictive than the definition of  $CD(K, \infty)$  we use here, following [16]. Corollary 1.3 is proven exactly as in the non-branching  $CD(K, \infty)$ -spaces. For the proof, see for instance [18, Theorem 30.32] or [17, Proposition 4.2].

**Corollary 1.3.** *The  $RCD(K, \infty)$  condition implies the  $CD(K, \infty)$  condition by Lott and Villani.*

By inspecting the proof of [9, Theorem 3.3] where the existence of optimal maps in non-branching  $CD(K, \infty)$ -spaces was shown we can make a refinement of the statement of Theorem 1.1. Regarding this refinement, stated in Corollary 1.4, we note that in [9] there was an extra assumption for the measures  $\mu_0$  and  $\mu_1$  to have finite entropy. This assumption was needed for showing that all absolutely continuous measures with respect to a measure  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  satisfying (2.2) also satisfy (2.2). Here this conclusion is already as an assumption so finiteness of the initial and final entropies are not needed.

**Corollary 1.4.** *Let  $(X, d, \mathbf{m})$  be a strong  $CD(K, \infty)$ -space. Then for every  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$  that are absolutely continuous with respect to  $\mathbf{m}$  there is a unique  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  and it is induced by a map.*

Although Corollary 1.4 follows quite easily from the proof of [9, Theorem 3.3] together with Theorem 1.1 we will give at the end of the paper an outline of the proof together with some details that are different from the proof by Gigli.

In the classical situation of Euclidean spaces the existence of optimal transport maps was proven by Brenier in [7]. Since then there have been several generalizations of this result. Most relevant in the context of this paper are the generalizations of McCann [13] for Riemannian manifolds, of Bertrand [6] for Alexandrov spaces, and the most recent results of Gigli [9] in non-branching  $CD(K, N)$  and  $CD(K, \infty)$ -spaces and of Ambrosio and the first author [5] in strongly non-branching metric spaces. Notice that as in the approach by Gigli [9] our proof for the existence of optimal transport maps does not use Kantorovich potentials.

## 2. PRELIMINARIES

Let us first recall some definitions and results related to optimal mass transportation and Ricci curvature lower bounds using optimal mass transportation. More details on this subject can be found for example in the book by Villani [18].

In this paper we always work in a complete separable geodesic metric space  $(X, d)$  equipped with a locally finite Borel measure  $\mathbf{m}$ .

**2.1. Optimal mass transportation.** We write as  $\mathcal{P}(X)$  the set of all the probability measures defined on the  $\sigma$ -algebra consisting of universally measurable sets of  $X$ . We denote by  $\mathcal{P}_2(X)$  the subset of  $\mathcal{P}(X)$  consisting of probability measures with finite second moments. We equip the space  $\mathcal{P}_2(X)$  with the Wasserstein 2-distance which for any two measures  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$  is defined as

$$W_2(\mu_0, \mu_1) = \left( \inf \left\{ \int_{X \times X} d(x, y)^2 d\sigma(x, y) \right\} \right)^{1/2}, \quad (2.1)$$

where the infimum is taken over all  $\sigma \in \mathcal{P}_2(X \times X)$  with  $\mu_0 = (\mathbf{p}^1)_\# \sigma$  and  $\mu_1 = (\mathbf{p}^2)_\# \sigma$ . Here, and later on,  $\mathbf{p}^k$  denotes the projection to the  $k$ -th coordinate.

We call a plan  $\sigma \in \mathcal{P}_2(X \times X)$  that minimizes (2.1) an optimal plan. Optimal plans exist under very mild assumptions. In contrast, the existence of optimal maps is rare. By an optimal map we mean Borel  $T: X \rightarrow X$  for which the plan  $G_\# \mu_0$  given by the graph  $G(x) = (x, T(x))$  of  $T$  is optimal. For the arguments in this paper, it is crucial to notice that any subplan of an optimal plan is also optimal in the sense that for  $\sigma$  optimal between  $\mu_0$  and  $\mu_1$  any  $\tilde{\sigma} \ll \sigma$  is optimal between  $(\mathbf{p}^1)_\# \tilde{\sigma}$  and  $\mu_1 = (\mathbf{p}^2)_\# \tilde{\sigma}$ .

Any optimal plan is concentrated on a cyclically monotone set  $M \subset X \times X$ . This means that for any family  $(x_1, y_1), \dots, (x_n, y_n) \in M$  and any permutation  $p: \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  we have

$$\sum_{i=1}^n d(x_i, y_i)^2 \leq \sum_{i=1}^n d(p(x_i), y_i)^2.$$

Intuitively this just means that optimal plans can not be improved.

**2.2. Geodesics.** We denote by  $C([0, 1]; X)$  the space of continuous curves  $\gamma: [0, 1] \rightarrow X$ . All the geodesics we consider in this paper are constant speed geodesics parametrized by the unit interval  $[0, 1]$ . We denote the set of all such geodesics in  $X$  by  $\text{Geo}(X) \subset C([0, 1]; X)$ , and for  $\gamma \in C([0, 1]; X)$  and  $t \in [0, 1]$  we use the abbreviation  $\gamma_t = \gamma(t)$ . The distance between  $\gamma^1, \gamma^2 \in C([0, 1]; X)$  is given by  $d^*(\gamma^1, \gamma^2) = \max\{d(\gamma_t^1, \gamma_t^2) : 0 \leq t \leq 1\}$ . Notice that the subspace  $(\text{Geo}(X), d^*)$ , with  $d^*$  restricted to  $\text{Geo}(X) \times \text{Geo}(X)$ , is complete and separable because the space  $(X, d)$  is. (This is because  $\text{Geo}(X)$  is a  $d^*$ -closed subset of the separable space  $C([0, 1]; X)$ .) We write  $l(\gamma) = d(\gamma_0, \gamma_1)$  for any geodesic  $\gamma \in \text{Geo}(X)$ .

We define for all  $s, t \in [0, 1]$  the restriction map  $\text{restr}_s^t: C([0, 1]; X) \rightarrow C([0, 1]; X): \gamma \mapsto \gamma \circ f_s^t$  with  $f_s^t: [0, 1] \rightarrow [0, 1]: x \mapsto (t - s)x + s$ . Notice that  $\text{restr}_s^t(\text{Geo}(X)) \subset \text{Geo}(X)$ . We call a set  $\Gamma \subset \text{Geo}(X)$  *non-branching* if for any  $\gamma, \tilde{\gamma} \in \Gamma$  we have: if there exists  $t \in (0, 1)$  such that  $\text{restr}_0^t \gamma = \text{restr}_0^t \tilde{\gamma}$ , then  $\gamma = \tilde{\gamma}$ . A measure  $\pi \in \mathcal{P}(\text{Geo}(X))$  is said to be concentrated on a set of non-branching geodesics, if there exists a non-branching Borel set  $\Gamma \subset \text{Geo}(X)$  so that  $\pi(\Gamma) = 1$ . The space consisting of all measures  $\pi \in \mathcal{P}(\text{Geo}(X))$  for which the mapping  $t \mapsto (e_t)_\# \pi$  is a geodesic in  $\mathcal{P}(X)$  from  $\mu_0 = (e_0)_\# \pi$  to  $\mu_1 = (e_1)_\# \pi$  is denoted by  $\text{OptGeo}(\mu_0, \mu_1)$ . Here the evaluation map is defined as  $e_t: C([0, 1]; X) \rightarrow X: \gamma \mapsto \gamma_t$ .

For every geodesic  $(\mu_t) \in \text{Geo}(\mathcal{P}_2(X))$  joining  $\mu_0$  to  $\mu_1$ , there also exists a measure  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  with  $(e_t)_\# \pi = \mu_t$ . See for instance [1, Theorem 2.10] for the proof of this fact. Notice that the measure  $\pi$  is not necessarily unique.

Our proof is heavily based on restricting a given  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$ . Without mentioning it every time, we use the fact that for a Borel  $f: \text{Geo}(X) \rightarrow [0, \infty)$  with  $\int_{\text{Geo}(X)} f d\pi = 1$  we have  $f\pi \in \text{OptGeo}((e_0)_\# f\pi, (e_1)_\# f\pi)$ , analogously to the restrictions of optimal plans. Another fact which we will repeatedly use is that  $(\text{restr}_i^s)_\# \pi \in \text{OptGeo}((e_t)_\# \pi, (e_s)_\# \pi)$ .

**2.3. Ricci curvature lower bounds.** The Ricci curvature lower bounds are defined using the entropy functional  $\text{Ent}_m: \mathcal{P}(X) \rightarrow [-\infty, \infty]$ . It is defined as

$$\text{Ent}_m(\mu) = \int_X \rho \log \rho d\mathbf{m}$$

for any absolutely continuous measure  $\mu = \rho\mathbf{m} \in \mathcal{P}(X)$  for which the positive part of  $\rho \log \rho$  is integrable. For other measures in  $\mathcal{P}(X)$  we define  $\text{Ent}_m(\mu) = \infty$ . Notice that for an absolutely continuous measure with support in a set of finite  $\mathbf{m}$ -measure the negative part of  $\rho \log \rho$  is always integrable. As we will later see we may always assume the measures to be supported in a set of finite  $\mathbf{m}$ -measure in the proof of Theorem 1.1.

Following the definition in [16] we call a metric measure space  $(X, d, \mathbf{m})$  a (weak)  $CD(K, \infty)$ -space, with some  $K \in \mathbb{R}$ , provided that for any  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$  that are absolutely continuous with respect to  $\mathbf{m}$  there exists a geodesic  $(\mu_t) \in \text{Geo}(\mathcal{P}_2(X))$  along which  $\text{Ent}_m$  is  $K$ -convex, that is

$$\text{Ent}_m(\mu_t) \leq (1 - t)\text{Ent}_m(\mu_0) + t\text{Ent}_m(\mu_1) - \frac{K}{2}t(1 - t)W_2^2(\mu_0, \mu_1) \quad (2.2)$$

holds for all  $t \in [0, 1]$ . We say that a functional is  $K$ -convex along  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  if it is  $K$ -convex along the corresponding geodesic  $(e_t)_\# \pi$ . If the inequality (2.2) is required to hold for all geodesics  $(\mu_t) \in \text{Geo}(\mathcal{P}_2(X))$  with endpoints absolutely continuous with respect to  $\mathbf{m}$ , the space is called a strong  $CD(K, \infty)$ -space.

In the proofs we will need  $\text{Ent}_m$  to be convex along restrictions  $f\pi$ . If the space  $(X, d, \mathbf{m})$  would only be a weak  $CD(K, \infty)$ -space,  $K$ -convexity along restrictions would not be guaranteed. However, in strong  $CD(K, \infty)$ -spaces it follows directly from the definition.

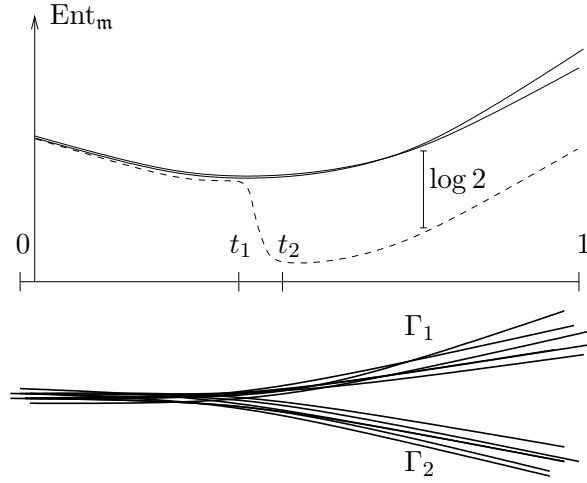


FIGURE 1. The idea in the proof of Theorem 1.1 is to find two sets of geodesics  $\Gamma_1$  and  $\Gamma_2$  so that the transport supported on the union of the geodesics is a measure  $\pi$  satisfying the assumption of Theorem 1.1 and so that the two sets of geodesics agree until time  $t_1$  and then branch out so that they become totally separated after time  $t_2$ . The entropies along the set  $\Gamma_1$  and along the set  $\Gamma_2$  are illustrated by the solid  $K$ -convex graphs. The entropy along  $\Gamma_1 \cup \Gamma_2$  is illustrated by the dashed graph. The non- $K$ -convexity of this graph due to the drop of  $\log 2$  in the entropy contradicts the assumption of the theorem.

### 3. PROOF OF THEOREM 1.1

The idea of the proof is similar to the proof of [15, Theorem 4]. We prove the claim by contradiction. First we find two geodesics in the Wasserstein space which start as the same geodesic and then branch out to two completely disjoint ones. By comparing the two geodesics separately and on the other hand their sum as one geodesic, we notice that there will be an extra drop of  $\log 2$  in the entropy during the time interval when the branching occurs. In order to arrive at a contradiction this branching has to happen in a small enough time interval as in the Figure 1.

There are two difficult steps in the proof before we arrive at the contradiction mentioned above. First of all we have to find the two geodesics that branch out to two completely separate ones. Once we have found them, we have to restrict the measures so that branching happens in a small enough interval. Here the difficulty is that when we restrict the measures, also their marginals change. By choosing the correct restrictions we can overcome this problem.

Let us start the proof with a simple lemma. It allows us to select the two disjoint geodesics needed for the contradiction. The geodesics will be selected using a probability measure on the product space  $\text{Geo}(X) \times \text{Geo}(X)$ . Notice that the constant  $1/5$  in the lemma is not sharp, however it is sufficient for our use.

**Lemma 3.1.** *Let  $(X, d)$  be a separable metric space. Then for any  $\sigma \in \mathcal{P}(X \times X)$  for which  $\sigma(\{(x, x) : x \in X\}) = 0$  there exists  $E \subset X$  so that  $\sigma(E \times (X \setminus E)) > 1/5$ .*

*Proof.* We will first reduce the general case to the case where  $X$  is a finite set. Take  $\epsilon > 0$ . Because the diagonal  $\{(x, x) : x \in X\}$  has zero  $\sigma$ -measure, there exists  $\delta > 0$  so that

$$\sigma(\{(x, y) \in X \times X : d(x, y) < \delta\}) < \epsilon. \quad (3.1)$$

Partition the space  $X$  into a countable collection of Borel sets  $(Q_i)_i$  with diameter at most  $\delta/2$ . Now there exists some  $n \in \mathbb{N}$  so that

$$\sum_{i=1}^n \left( (\mathbb{P}^k)_{\#} \sigma \right) (Q_i) > 1 - \epsilon, \quad \text{for } k = 1, 2. \quad (3.2)$$

Therefore, by combining (3.1) and (3.2) we have

$$\sigma \left( \bigcup_{1 \leq i, j \leq n, i \neq j} Q_i \times Q_j \right) \geq 1 - 3\epsilon$$

and so by forgetting a part of the space with arbitrarily small measure we may assume the space  $X$  to consist of  $n$  points.

The existence of the set  $E$  in the case where  $X$  consists of  $n$  points follows easily: There are a total number of  $2^n - 2$  ways to select a non-empty set  $E \subsetneq X$  and for any pair of points  $(i, j)$  with  $i \neq j$  there are  $2^{n-2}$  sets  $E \subset X$  with  $i \in E$  and  $j \notin E$ . As a consequence

$$\sum_{E \subset X} \sigma(E \times (X \setminus E)) = 2^{n-2}$$

and so there has to be a set  $E \subset X$  with

$$\sigma(E \times (X \setminus E)) \geq \frac{2^{n-2}}{2^n - 2} > \frac{1}{4}.$$

Taking  $\epsilon > 0$  sufficiently small finishes the proof.  $\square$

Now we are ready to continue with the proof of Theorem 1.1. The first reduction steps contain ideas which were used in the proofs of [9, Theorem 3.3] and [15, Theorem 4]. The rest of the proof is then close to that of [15, Theorem 4].

*Proof of Theorem 1.1.* Let us give the idea behind each step of the proof. We will prove the claim by contradiction, so we have a measure  $\pi$  that does not live on a non-branching set of geodesics. The first step is to restrict the measures in time and space to live inside a sufficiently small ball. This will help in the last step of the proof in estimates involving the extra term coming from  $K$ -convexity when  $K \neq 0$ . In the second step we produce via disintegration a measure  $\sigma$  on the product space  $\text{Geo}(X) \times \text{Geo}(X)$  which gives positive measure to pairs of geodesics which start as the same but after some time  $T$  branch out.

In the third step we pushforward the measure  $\sigma$  on the product space to live on more suitable pairs of geodesics. The pairs of geodesics which we want to avoid are the ones that branch out and come back together infinitely often. This step will be needed only to achieve the reductions in the next step. In the fourth step of the proof we restrict the measure to geodesics which stay disjoint at least for some time  $\delta > 0$  (independent of the geodesics) immediately after the branch out for the first time. We also restrict the measure so that its marginals have bounded densities.

In the fifth step we find, for a given  $\epsilon > 0$ , a time  $t$  so that during the time interval  $[t, t + \epsilon]$  the measure sees lots of branching. In the sixth step we first use Lemma 3.1 to obtain two disjoint sets of geodesics so that their product has large measure. After this we

further restrict the measure so that we can project the product measure to two measures  $\pi^{\text{up}}$  and  $\pi^{\text{down}}$  which have disjoint supports at time  $t + \epsilon$ . Finally in the last step, step seven we compare the entropies along  $\pi^{\text{up}}$ ,  $\pi^{\text{down}}$  and  $\pi^{\text{up}} + \pi^{\text{down}}$  to obtain a contradiction which we already mentioned in the beginning of this section and in the Figure 1.

**Step 1: Localization to a small ball.**

Assume that Theorem 1.1 is not true. Then there exist  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$  and an optimal  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  that is not concentrated on non-branching geodesics. Because the space  $(X, \mathbf{d})$  is separable and the measure  $\mathbf{m}$  locally finite, we can cover  $(X, \mathbf{d})$  with a countable collection of balls  $B(x_i, l_i/4)$  so that

$$l_i \leq \sqrt{\frac{\log 2}{6|K| + 1}} \quad (3.3)$$

and  $\mathbf{m}(B(x_i, l_i)) < \infty$ . Since  $\pi$  is not concentrated on non-branching geodesics there exist some  $i \in \mathbb{N}$  and  $L > l_i$  so that  $\pi$  has some branching inside  $B(x_i, l_i/4)$  along geodesics with length at most  $L$ . That is,  $\pi(\Gamma) < 1$  for every  $\Gamma \subset \text{Geo}(X)$  satisfying the following: if  $\gamma^1, \gamma^2 \in \Gamma$  so that  $\gamma_s^1 = \gamma_s^2$  for all  $s \in [0, t_0]$  with  $\gamma_{t_0}^1 \in B(x_i, l_i/4)$  and  $l(\gamma^1) \leq L$ , then  $\gamma_s^1 = \gamma_s^2$  also for all  $s$  for which  $\gamma_s^1 \in B(x_i, l_i/4)$  or  $\gamma_s^2 \in B(x_i, l_i/4)$ .

Therefore there exists  $t_1 \in (0, 1 - l_i/(4L))$  so that

$$\pi^r = (\text{restr}_{t_1}^{l_i/(4L)})_{\#} \left( \frac{1}{\pi(\Gamma^r)} \pi|_{\Gamma^r} \right),$$

with  $\pi(\Gamma^r) > 0$  defined as

$$\Gamma^r = \{ \gamma \in \text{Geo}(X) : l(\gamma) \leq L \text{ and } \gamma_{t_1 + l_i/(8L)} \in B(x_i, l_i/4) \},$$

is not concentrated on non-branching geodesics. Now the measure  $\pi^r$  is supported on a set of geodesics that live inside the ball  $B(x_i, l_i/2)$ . Thus without loss of generality, we may assume from the beginning that the original measure  $\pi$  is concentrated on geodesics living inside some ball  $B(x, l/2) \subset X$  with  $l$  having the same bound from above as  $l_i$  in (3.3) and  $\mathbf{m}(B(x, l)) < \infty$ .

**Step 2: Disintegrating the branching measure.**

We claim that from the assumption that  $\pi$  is not concentrated on a non-branching set of geodesics we know that there exists some  $T \in (0, 1)$  so that the measure  $\pi_\gamma$  is not a Dirac mass for a  $(\text{restr}_0^T)_{\#} \pi$ -positive set of curves  $\gamma \in \text{Geo}(X)$ , where  $\{\pi_\gamma\} \subset \mathcal{P}(\text{Geo}(X))$  is the disintegration of  $\pi$  with respect to  $\text{restr}_0^T$ . The measure  $\pi_\gamma$  not being a Dirac mass means that it is concentrated on geodesics which coincide with  $\gamma$  in  $[0, T]$  and that a set of  $\pi_\gamma$ -positive measures of geodesics do branch after time  $T$ . We prove the claim by contradiction. Assume that it is not true so that for any  $T \in (0, 1)$  the measures  $\pi_\gamma$  are Dirac for  $(\text{restr}_0^T)_{\#} \pi$ -almost every curve  $\gamma \in \text{Geo}(X)$ . Take  $\epsilon > 0$ . Since  $\text{Geo}(X)$  is separable there exists a countable Borel decomposition of  $\text{Geo}(X)$  into disjointed sets  $(A_{i,\epsilon})_{i \in I}$  with  $\text{diam}(A_{i,\epsilon}) < \epsilon$ . For each  $i \in I$  the map  $M_{i,\epsilon,T}: \text{Geo}(X) \rightarrow [0, 1]: \gamma \mapsto \pi_\gamma(A_{i,\epsilon})$  is Borel measurable and so  $M_{i,\epsilon,T}^{-1}(\{1\})$  is a Borel set. Since  $\pi_\gamma$  are probability measures,  $(M_{i,\epsilon,T}^{-1}(\{1\}))_{i \in I}$  are disjointed. By the assumption that almost every  $\pi_\gamma$  is a Dirac mass the Borel set

$$\Gamma_{\epsilon,T} = \bigcup_{i \in I} \left( (\text{restr}_0^T)^{-1}(M_{i,\epsilon,T}^{-1}(\{1\})) \cap A_{i,\epsilon} \right)$$

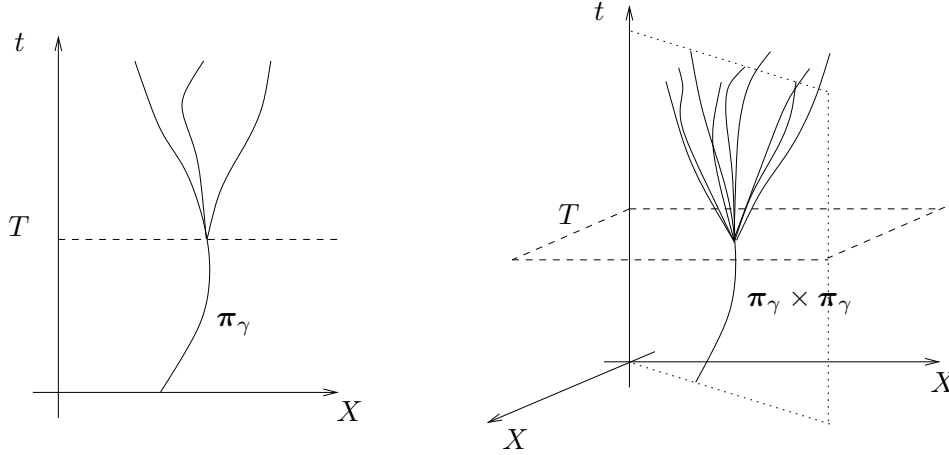


FIGURE 2. The measure  $\sigma$  is constructed by integrating up the measures  $\pi_\gamma \times \pi_\gamma$ . Since branching for  $\pi_\gamma$  occurs only after time  $T$ , all the measures  $\pi_\gamma \times \pi_\gamma$  and hence also the measure  $\sigma$  live on the diagonal of  $X \times X$  until time  $T$ .

has full  $\pi$ -measure. Therefore also the Borel set

$$\Gamma_T = \bigcap_{\epsilon \in (0,1) \cap \mathbb{Q}} \Gamma_{\epsilon, T}$$

has full  $\pi$ -measure. Notice that for  $\gamma^1, \gamma^2 \in \Gamma_T$  with  $\gamma^1 \neq \gamma^2$  there exists  $\epsilon \in (0,1) \cap \mathbb{Q}$  so that  $\gamma^1 \in A_{i,\epsilon}$  and  $\gamma^2 \in A_{j,\epsilon}$  with  $i \neq j$ . Therefore  $\text{restr}_0^T(\gamma^1) \in M_{i,\epsilon,T}^{-1}(\{1\})$  and  $\text{restr}_0^T(\gamma^2) \in M_{j,\epsilon,T}^{-1}(\{1\})$  giving  $\text{restr}_0^T(\gamma^1) \neq \text{restr}_0^T(\gamma^2)$ . This means that the set  $\Gamma_T$  does not contain geodesics that branch after time  $T$ . Thus the Borel set

$$\bigcap_{T \in (0,1) \cap \mathbb{Q}} \Gamma_T$$

is non-branching and has full  $\pi$ -measure, giving a contradiction.

Consider a measure  $\sigma \in \mathcal{P}(\text{Geo}(X) \times \text{Geo}(X))$  obtained by integrating up the disintegrated measure  $\pi_\gamma$  as a product measure  $\pi_\gamma \times \pi_\gamma$ . That is, for any Borel measurable  $f: \text{Geo}(X) \times \text{Geo}(X) \rightarrow [0, \infty]$  we have

$$\begin{aligned} & \int_{\text{Geo}(X) \times \text{Geo}(X)} f(\gamma^1, \gamma^2) d\sigma(\gamma^1, \gamma^2) \\ &= \int_{\text{Geo}(X)} \int_{(\text{restr}_0^T)^{-1}(\gamma) \times (\text{restr}_0^T)^{-1}(\gamma)} f(\gamma^1, \gamma^2) d(\pi_\gamma \times \pi_\gamma)(\gamma^1, \gamma^2) d(\text{restr}_0^T)_\# \pi(\gamma). \end{aligned}$$

The measures  $\pi_\gamma \times \pi_\gamma$  are illustrated in Figure 2.

**Step 3: Reducing the number of branching points of pairs of geodesics.**

It might happen that part of the measure  $\sigma$  lives on pairs of geodesics  $(\gamma^1, \gamma^2)$  for which the set  $\{t \in [0,1] : \gamma_t^1 = \gamma_t^2\}$  consists of infinitely many intervals. We want to exclude this behaviour so that branching geodesics stay disjoint for a positive time immediately after branching. We will do this by a measurable selection.



We start by defining the subsets  $\mathfrak{D}, \mathfrak{B}, \mathfrak{B}^1$  of  $\text{Geo}(X)^2$ . The first one is the diagonal set

$$\mathfrak{D} := \{(\gamma^1, \gamma^2) \in \text{Geo}(X)^2 : \gamma^1 = \gamma^2\},$$

the second one the set of pairs of geodesics branching after time  $T$

$$\mathfrak{B} := \{(\gamma^1, \gamma^2) \in \text{Geo}(X)^2 \setminus \mathfrak{D} : \text{restr}_0^T \gamma^1 = \text{restr}_0^T \gamma^2\}$$

and the final one the set of pairs of geodesics that branch exactly once after time  $T$

$$\mathfrak{B}^1 := \{(\gamma^1, \gamma^2) \in \mathfrak{B} : \text{the set } \{t \in [0, 1] : \gamma_t^1 \neq \gamma_t^2\} \text{ is an interval}\}.$$

The subsets  $\mathfrak{D}$  and  $\mathfrak{B}$  are clearly Borel. Let us show that  $\mathfrak{B}^1$  is also Borel. For every  $\epsilon, t_1, t_2 > 0$  define the closed sets

$$\mathfrak{B}_{\epsilon, t_1, t_2} = \{(\gamma^1, \gamma^2) \in \mathfrak{B} : \text{there exist } s \in [t_1, t_2] \text{ with } \gamma_s^1 = \gamma_s^2 \text{ and} \\ r_1 \in [T, t_1], r_2 \in [t_2, 1] \text{ such that } d(\gamma_{r_i}^1, \gamma_{r_i}^2) \geq \epsilon \text{ for } i = 1, 2\}.$$

Now

$$\mathfrak{B}^1 = \mathfrak{B} \setminus \left( \bigcup_{\epsilon, t_1, t_2 \in (0, 1] \cap \mathbb{Q}} \mathfrak{B}_{\epsilon, t_1, t_2} \right)$$

and so  $\mathfrak{B}^1$  is Borel.

Now define  $\text{End}: \mathfrak{B}^1 \rightarrow X^3: (\gamma^1, \gamma^2) \mapsto (\gamma_0^1, \gamma_1^1, \gamma_1^2)$  and let  $\text{End}^{-1}: \text{End}(\mathfrak{B}^1) \rightarrow \mathfrak{B}^1$  be its Suslin measurable right-inverse given by the Jankoff theorem [11]. (The set  $\mathfrak{B}^1$  is a Suslin space as a Borel subset of a Polish space. The mapping  $\text{End}$  is continuous and thus  $\text{End}(\mathfrak{B}^1)$  is also Suslin space. Suslin subsets of a Polish space are universally measurable and so Suslin-measurability suffices for our considerations.)

Now consider the Suslin measurable map  $\text{Br}: \text{Geo}(X)^2 \rightarrow \text{Geo}(X)^2$  given by

$$\text{Br}(\gamma^1, \gamma^2) = \begin{cases} \text{End}^{-1}(\gamma_0^1, \gamma_1^1, \gamma_1^2) & \text{if } (\gamma_0^1, \gamma_1^1, \gamma_1^2) \in \text{End}(\mathfrak{B}^1), \\ (\gamma^1, \gamma^2) & \text{otherwise.} \end{cases}$$

The role of  $\text{Br}$  is to select an element from  $\mathfrak{B}^1$  with the same starting point  $\gamma_0^1$  and endpoints  $\gamma_1^1$  and  $\gamma_1^2$ , if there exists one, and otherwise to give the original pair. For any branching pair of geodesics it gives a pair of geodesics that branches exactly once, like in Figure 3. (Such geodesics exist: If  $\gamma_1^1 \neq \gamma_1^2$ , let  $s \in (0, 1)$  be the last time when  $\gamma_s^1 = \gamma_s^2$ . Then defining  $\gamma^3|_{[0, s]} = \gamma^4|_{[0, s]} = \gamma^1|_{[0, s]}$ ,  $\gamma^3|_{[s, 1]} = \gamma^1|_{[s, 1]}$  and  $\gamma^4|_{[s, 1]} = \gamma^2|_{[s, 1]}$  gives you the desired pair of geodesics  $(\gamma^3, \gamma^4)$ . If  $\gamma_1^1 = \gamma_1^2$ , take  $0 < s_1 < s_2 \leq 1$  such that  $\gamma_{s_1}^1 = \gamma_{s_2}^2$ ,  $\gamma_{s_2}^1 = \gamma_{s_2}^2$  and  $\gamma_s^1 \neq \gamma_s^2$  for all  $s \in (s_1, s_2)$ . Then define  $\gamma^3|_{[0, s_1] \cup [s_2, 1]} = \gamma^4|_{[0, s_1] \cup [s_2, 1]} = \gamma^1|_{[0, s_1] \cup [s_2, 1]}$ ,  $\gamma^3|_{[s_1, s_2]} = \gamma^1|_{[s_1, s_2]}$  and  $\gamma^4|_{[s_1, s_2]} = \gamma^2|_{[s_1, s_2]}$ , and again you have the desired pair of geodesics  $(\gamma^3, \gamma^4)$ .) In other words  $\text{Br}(\mathfrak{B}) \subset \mathfrak{B}^1$ . Now the pushforward measure  $\text{Br}_\# \sigma$  gives  $\mathfrak{B}^1$  positive measure and it has the same marginals as  $\sigma$ , in other words,  $(e_0, e_0)_\#(\text{Br}_\# \sigma) = (e_0, e_0)_\# \sigma$  and  $(e_1, e_1)_\#(\text{Br}_\# \sigma) = (e_1, e_1)_\# \sigma$ . Therefore, after pushing forward the measure by  $\text{Br}$  if necessary, we may assume that  $\sigma(\mathfrak{B}^1) > 0$ .

**Step 4: Further restrictions of the measure  $\sigma$ .**

Next we will show that by restricting and rescaling the measure  $\sigma$  we may assume that there exist  $S \in (T, 1)$ ,  $\delta \in (0, (1 - S)/2)$  and  $C > 0$  so that

$$\frac{d(e_0)_\#(\mathfrak{p}^1 \sigma)}{d\mathfrak{m}}, \frac{d(e_1)_\#(\mathfrak{p}^1 \sigma)}{d\mathfrak{m}}, \frac{d(e_1)_\#(\mathfrak{p}^2 \sigma)}{d\mathfrak{m}} < C \quad (3.4)$$

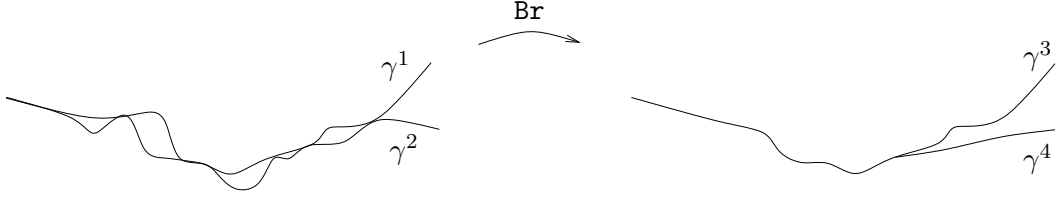


FIGURE 3. The mapping  $\text{Br}$  takes a branching pair of geodesics  $(\gamma^1, \gamma^2)$  to another pair of geodesics  $(\gamma^3, \gamma^4)$  that have the same endpoints, but branch only once.

and for  $\sigma$ -almost every  $(\gamma^1, \gamma^2)$  there exists  $t \in [T, S]$  such that

$$\text{restr}_0^t(\gamma^1) = \text{restr}_0^t(\gamma^2) \text{ and } \gamma_s^1 \neq \gamma_s^2 \text{ for all } s \in (t, t + \delta). \quad (3.5)$$

First of all we may select  $S \in (T, 1)$  such that

$$\sigma(\{(\gamma^1, \gamma^2) \in \mathfrak{B}^1 : \text{restr}_0^S(\gamma^1) \neq \text{restr}_0^S(\gamma^2)\}) > 0.$$

On the other hand the set

$$\{(\gamma^1, \gamma^2) \in \mathfrak{B}^1 : \text{restr}_0^S(\gamma^1) \neq \text{restr}_0^S(\gamma^2)\}$$

can be covered by a countable union of sets of the form

$$A_q = \{(\gamma^1, \gamma^2) \in \mathfrak{B}^1 : \text{there exists } t \in [T, S] \text{ such that } \text{restr}_0^t(\gamma^1) = \text{restr}_0^t(\gamma^2) \\ \text{and } \gamma_s^1 \neq \gamma_s^2 \text{ for all } s \in (t, t + q)\}.$$

Therefore there exists a  $\delta \in (0, (1 - S)/2)$  so that

$$\sigma(A_\delta) > 0.$$

Notice that the sets  $A_q$  are Borel because they can be written as

$$A_q = \bigcap_{z \in (q, 1) \cap \mathbb{Q}} \bigcup_{t, \epsilon \in (0, 1 - z] \cap \mathbb{Q}} \{(\gamma^1, \gamma^2) \in \mathfrak{B}^1 : d(\gamma_s^1, \gamma_s^2) \geq \epsilon \text{ for all } s \in [t, t + z]\}.$$

By restricting  $\sigma$  to  $A_\delta$  we have then obtained (3.5). Since  $\mu_0$  and  $\mu_1$  are absolutely continuous with respect to  $\mathfrak{m}$  a further restriction and finally rescaling yields (3.4).

**Step 5: Selecting a time  $t$  where sufficient amount of branching occurs.**

Define the function  $f: [0, 1] \rightarrow [0, 1]$  as

$$f(t) = \sigma(\{(\gamma^1, \gamma^2) : \text{restr}_0^t \gamma^1 = \text{restr}_0^t \gamma^2\}).$$

Notice that  $f$  is decreasing,  $f = 1$  on  $[0, T]$  and  $f = 0$  on  $[S, 1]$ . Roughly speaking  $f(t)$  gives the amount of branching that will occur after time  $t$ .

Take a small  $\epsilon \in (0, \delta)$  and let  $t \in [T, S]$  be such that  $f(t) - f(t + \epsilon) > \epsilon$ . Such  $t$  exists because  $f$  decreases by 1 on the interval  $[T, S]$ . Define

$$F = \{(\gamma^1, \gamma^2) : \text{restr}_0^t \gamma^1 = \text{restr}_0^t \gamma^2 \text{ and } \text{restr}_0^{t+\epsilon} \gamma^1 \neq \text{restr}_0^{t+\epsilon} \gamma^2\}.$$

Now  $\sigma(F) > \epsilon$  by the choice of  $t$  and (3.5).

**Step 6: Finding two measures completely branching out during  $[t, t + \epsilon]$ .**

By Lemma 3.1 there exists  $E \subset \text{Geo}(X)$  so that

$$\sigma\left((E \times (\text{Geo}(X) \setminus E)) \cap F\right) > \frac{\sigma(F)}{5} > \frac{\epsilon}{5}.$$

Let  $\eta > 0$  be such that

$$M := \sigma(A) > \frac{\sigma\left(\left(E \times (\text{Geo}(X) \setminus E)\right) \cap F\right)}{2} > \frac{\epsilon}{10}, \quad (3.6)$$

where  $A = \{(\gamma^1, \gamma^2) \in (E \times (\text{Geo}(X) \setminus E)) \cap F : d(\gamma_{t+\epsilon}^1, \gamma_{t+\epsilon}^2) > \eta\}$ . Such  $\eta$  exists since  $\gamma_{t+\epsilon}^1 \neq \gamma_{t+\epsilon}^2$  for  $\sigma$ -almost every  $(\gamma^1, \gamma^2)$  by (3.5) and the fact that  $\epsilon < \delta$ .

Now define  $\tilde{\sigma} = \sigma|_A$  and using it  $\tilde{\pi} = (\mathbf{p}^1)_\# \tilde{\sigma}$ . Write  $\tilde{\rho}_s \mathbf{m} = (e_s)_\# \tilde{\pi}$  for all  $s \in [0, 1]$ . By Jensen's inequality and (3.6) we get

$$\int \tilde{\rho}_{t+\epsilon} \log \tilde{\rho}_{t+\epsilon} \, d\mathbf{m} \geq M \log \frac{M}{\mathbf{m}(B(x, l/2))} \geq M \log \frac{\epsilon}{10\mathbf{m}(B(x, l/2))}. \quad (3.7)$$

Since  $X$  is separable we can decompose it into a countable collection  $\{Q_i\}$  of disjointed Borel sets with diameter less than  $\eta$ . We claim that for at least one  $Q \in \{Q_i\}$  we have

$$w := \int_Q \tilde{\rho}_{t+\epsilon} \, d\mathbf{m} = \tilde{\pi}(G) > 0$$

for  $G = \{\gamma \in \text{Geo}(X) : \gamma_{t+\epsilon} \in Q\}$  and

$$\int \rho_{t+\epsilon}^{\text{up}} \log \rho_{t+\epsilon}^{\text{up}} \, d\mathbf{m} \geq w \log \frac{\epsilon}{10\mathbf{m}(B(x, l/2))}, \quad (3.8)$$

where  $\pi^{\text{up}} = \tilde{\pi}|_G = (\mathbf{p}^1)_\# \tilde{\sigma}|_{G \times \text{Geo}(X)}$  and  $\mu_s^{\text{up}} = \rho_s^{\text{up}} \mathbf{m} = (e_s)_\# \pi^{\text{up}}$  for all  $s \in [0, 1]$ . For if this were not the case we would have

$$\begin{aligned} \int \tilde{\rho}_{t+\epsilon} \log \tilde{\rho}_{t+\epsilon} \, d\mathbf{m} &= \sum_i \int_{Q_i} \tilde{\rho}_{t+\epsilon} \log \tilde{\rho}_{t+\epsilon} \, d\mathbf{m} < \sum_i \int_{Q_i} \tilde{\rho}_{t+\epsilon} \, d\mathbf{m} \log \frac{\epsilon}{10\mathbf{m}(B(x, l/2))} \\ &= M \log \frac{\epsilon}{10\mathbf{m}(B(x, l/2))} \end{aligned}$$

contradicting (3.7).

Define also  $\pi^{\text{down}} = (\mathbf{p}^2)_\# \tilde{\sigma}|_{G \times \text{Geo}(X)}$  and  $\mu_s^{\text{down}} = \rho_s^{\text{down}} \mathbf{m} = (e_s)_\# \pi^{\text{down}}$  for all  $s \in [0, 1]$ .

Notice that  $(\text{restr}_0^t)_\# \pi^{\text{up}} = (\text{restr}_0^t)_\# \pi^{\text{down}}$  and  $\mu_{t+\epsilon}^{\text{up}} \perp \mu_{t+\epsilon}^{\text{down}}$ .

By (3.4) the estimate

$$\int \rho \log \rho \, d\mathbf{m} \leq w \log C \quad (3.9)$$

holds for all  $\rho \in \{\rho_0^{\text{down}}, \rho_0^{\text{up}}, \rho_1^{\text{down}}, \rho_1^{\text{up}}\}$ .

**Step 7: Comparing entropies along different geodesics.**

Now, similarly as in [15] we can estimate using the  $K$ -convexity first along the measure  $(\pi^{\text{up}} + \pi^{\text{down}})/(2w)$  between times 0,  $t$  and  $t + \epsilon$ , and then separately along  $\pi^{\text{up}}/w$  and

$\pi^{\text{down}}/w$  between times  $t$ ,  $t + \epsilon$  and 1 to get

$$\begin{aligned}
& \int \frac{\rho_t^{\text{down}}}{w} \log \frac{\rho_t^{\text{down}}}{w} \, \text{d}\mathbf{m} \\
& \leq \frac{\epsilon}{t + \epsilon} \int \frac{\rho_0^{\text{down}}}{w} \log \frac{\rho_0^{\text{down}}}{w} \, \text{d}\mathbf{m} + \frac{t}{t + \epsilon} \int \frac{\rho_{t+\epsilon}^{\text{down}} + \rho_{t+\epsilon}^{\text{up}}}{2w} \log \frac{\rho_{t+\epsilon}^{\text{down}} + \rho_{t+\epsilon}^{\text{up}}}{2w} \, \text{d}\mathbf{m} + \frac{|K|}{2} \frac{t}{t + \epsilon} l^2 \\
& \leq \frac{\epsilon}{t + \epsilon} \log \frac{C}{w} - \frac{t}{t + \epsilon} \log 2 \\
& \quad + \frac{t}{2(t + \epsilon)} \left( \int \frac{\rho_{t+\epsilon}^{\text{down}}}{w} \log \frac{\rho_{t+\epsilon}^{\text{down}}}{w} \, \text{d}\mathbf{m} + \int \frac{\rho_{t+\epsilon}^{\text{up}}}{w} \log \frac{\rho_{t+\epsilon}^{\text{up}}}{w} \, \text{d}\mathbf{m} \right) + \frac{|K|}{2} \frac{t}{t + \epsilon} l^2 \\
& \leq \frac{\epsilon}{t + \epsilon} \log \frac{C}{w} - \frac{t}{t + \epsilon} \log 2 \\
& \quad + \frac{t}{2(t + \epsilon)} \left( \frac{\epsilon}{1-t} \int \frac{\rho_1^{\text{down}}}{w} \log \frac{\rho_1^{\text{down}}}{w} \, \text{d}\mathbf{m} + \frac{1-t-\epsilon}{1-t} \int \frac{\rho_t^{\text{down}}}{w} \log \frac{\rho_t^{\text{down}}}{w} \, \text{d}\mathbf{m} \right) \\
& \quad + \frac{t}{2(t + \epsilon)} \left( \frac{\epsilon}{1-t} \int \frac{\rho_1^{\text{up}}}{w} \log \frac{\rho_1^{\text{up}}}{w} \, \text{d}\mathbf{m} + \frac{1-t-\epsilon}{1-t} \int \frac{\rho_t^{\text{up}}}{w} \log \frac{\rho_t^{\text{up}}}{w} \, \text{d}\mathbf{m} \right) + |K| \frac{t}{t + \epsilon} l^2 \\
& \leq \left( \frac{\epsilon}{t + \epsilon} + \frac{t\epsilon}{(1-t)(t + \epsilon)} \right) \log \frac{C}{w} - \frac{t}{t + \epsilon} \log 2 \\
& \quad + \frac{t(1-t-\epsilon)}{(t + \epsilon)(1-t)} \int \frac{\rho_t^{\text{down}}}{w} \log \frac{\rho_t^{\text{down}}}{w} \, \text{d}\mathbf{m} + |K| \frac{t}{t + \epsilon} l^2.
\end{aligned}$$

Moving the integral term from the right-hand side to the left and using (3.3) gives

$$\begin{aligned}
& \frac{\epsilon}{(t + \epsilon)(1-t)} \int \frac{\rho_t^{\text{down}}}{w} \log \frac{\rho_t^{\text{down}}}{w} \, \text{d}\mathbf{m} \\
& \leq \left( \frac{\epsilon}{t + \epsilon} + \frac{t\epsilon}{(1-t)(t + \epsilon)} \right) \log \frac{C}{w} - \frac{t}{t + \epsilon} \log 2 + |K| \frac{t}{t + \epsilon} l^2 \\
& \leq \left( \frac{\epsilon}{t + \epsilon} + \frac{t\epsilon}{(1-t)(t + \epsilon)} \right) \log \frac{C}{w} - \frac{2t}{3(t + \epsilon)} \log 2
\end{aligned}$$

which is the same as

$$\int \frac{\rho_t^{\text{down}}}{w} \log \frac{\rho_t^{\text{down}}}{w} \, \text{d}\mathbf{m} \leq \log \frac{C}{w} - \frac{2t(1-t)}{3\epsilon} \log 2. \quad (3.10)$$

Using convexity along  $\pi^{\text{up}}/w$  once more together with (3.10), (3.3) and the fact that  $\epsilon \leq t$  yields

$$\begin{aligned}
& \int \frac{\rho_{t+\epsilon}^{\text{up}}}{w} \log \frac{\rho_{t+\epsilon}^{\text{up}}}{w} \, \text{d}\mathbf{m} \leq \frac{\epsilon}{1-t} \log \frac{C}{w} + \frac{1-t-\epsilon}{1-t} \int \frac{\rho_t^{\text{down}}}{w} \log \frac{\rho_t^{\text{down}}}{w} \, \text{d}\mathbf{m} + \frac{|K|}{2} \epsilon(1-t-\epsilon) l^2 \\
& \leq \log \frac{C}{w} - \frac{t(1-t-\epsilon)}{3\epsilon} \log 2.
\end{aligned}$$

When we combine this with (3.8) we have

$$3\epsilon \left( \log \frac{\epsilon}{10\mathbf{m}(B(x, l/2))} - \log C \right) \leq -t(1-t-\epsilon) \log 2 \leq -\frac{TS}{2} \log 2$$

which is a contradiction as the left-hand side goes to 0 when  $\epsilon \downarrow 0$ .  $\square$

**Remark 3.2.** *In the proof of Theorem 1.1 we worked with two fixed marginals  $\mu_0, \mu_1$ . Therefore the assumption in Theorem 1.1 of being a strong  $CD(K, \infty)$ -space could be weakened. The slightly stronger (but more complicated-looking) version of Theorem 1.1 would be the following.*

*Let  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$  be absolutely continuous. Suppose that for every  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$ , every  $t, s \in [0, 1]$  and any Borel function  $f: \text{Geo}(X) \rightarrow [0, \infty)$  with  $\int_{\text{Geo}(X)} f d\pi = 1$  the entropy  $\text{Ent}_{\mathfrak{m}}$  is  $K$ -convex along  $(\text{restr}_s^t)_\#(f\pi)$ . Then any  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  is concentrated on a set of non-branching geodesics.*

#### 4. PROOF OF COROLLARY 1.4

Let us now outline the proof of Corollary 1.4. As usual, it suffices to prove that every optimal plan is given by a map. This implies uniqueness of optimal plans because for any  $\pi^1, \pi^2 \in \text{OptGeo}(\mu_0, \mu_1)$  also  $\frac{1}{2}(\pi^1 + \pi^2) \in \text{OptGeo}(\mu_0, \mu_1)$ . Suppose then that there are two  $\mu_0, \mu_1 \in \mathcal{P}_2(X)$  which are absolutely continuous with respect to  $\mathfrak{m}$  and that there is  $\pi \in \text{OptGeo}(\mu_0, \mu_1)$  which is not induced by an optimal map.

Because  $\mu_0 = \rho_0 \mathfrak{m}$  is absolutely continuous, the union

$$\bigcup_{C \geq 0} \Gamma_C, \quad \Gamma_C = \{\gamma \in \text{Geo}(X) : \rho_0(\gamma_0) \leq C\}$$

has full  $\pi$ -measure. Therefore for some  $C \geq 0$  the measure  $\pi|_{\Gamma_C}$  is not induced by a map. Therefore we may assume  $\mu_0$  to have bounded density. Similarly we may assume  $\mu_1$  to have bounded density. Emptying the space  $X$  with larger and larger balls we may also assume  $\mu_0$  and  $\mu_1$  to have bounded supports. Therefore we may assume  $\text{Ent}_{\mathfrak{m}}(\mu_0), \text{Ent}_{\mathfrak{m}}(\mu_1) \in \mathbb{R}$ .

In the proof of [9, Theorem 3.3] Gigli finds two probability measures  $\pi^1, \pi^2 \ll \pi$  with  $\pi^1 \perp \pi^2$  and  $(e_0)_\# \pi^1 = (e_0)_\# \pi^2 = \mathfrak{m}|_D$  for a compact  $D \subset X$  with  $\mathfrak{m}(D) > 0$ . But by [9, Lemma 3.2], which holds also under the strong  $CD(K, \infty)$  assumption, we have

$$\liminf_{t \searrow 0} \mathfrak{m}(\{\rho_t^i > 0\}) \geq \mathfrak{m}(D), \quad i = 1, 2$$

for the densities  $\rho_t^i$  of  $(e_t)_\# \pi^i$ . Therefore for some small time  $t \in (0, 1)$  the sets  $\{\rho_t^1 > 0\}$  and  $\{\rho_t^2 > 0\}$  must intersect in a set  $E$  of positive  $\mathfrak{m}$ -measure.

So far we have followed the proof of [9, Theorem 3.3]. Now the final step in the case of non-branching  $CD(K, \infty)$ -spaces would be to say that no two different geodesics in the support of an optimal transport can intersect. In the essentially non-branching spaces this conclusion is not so clear, so we argue differently.

The heuristic idea is to mix the measures  $\pi^1$  and  $\pi^2$  so that at time  $t$  we are allowed to change from the geodesics where  $\pi^1$  lives to the geodesics where  $\pi^2$  lives, and vice versa.

To write this more rigorously we first pushforward the combined measure  $\pi^1 + \pi^2$  to the left and right part from time  $t$

$$\pi^{\text{left}} = \frac{1}{2} \left( (\text{restr}_0^t)_\# \pi^1 + (\text{restr}_0^t)_\# \pi^2 \right), \quad \pi^{\text{right}} = \frac{1}{2} \left( (\text{restr}_t^1)_\# \pi^1 + (\text{restr}_t^1)_\# \pi^2 \right).$$

Now let  $\{\pi_x^{\text{left}}\}$  be the disintegration of  $\pi^{\text{left}}$  with respect to  $e_1$  and let  $\{\pi_x^{\text{right}}\}$  be the disintegration of  $\pi^{\text{right}}$  with respect to  $e_0$ .

Observe that the mapping

$$\text{Sp}: C([0, 1]; X) \rightarrow \{(\gamma^1, \gamma^2) \in C([0, 1]; X) \times C([0, 1]; X) : \gamma_1^1 = \gamma_0^2\} : \gamma \mapsto (\text{restr}_0^t \gamma, \text{restr}_t^1 \gamma)$$

is bi-Lipschitz. We define the collection of measures  $\{\pi_x\}$  on  $C([0, 1]; X)$  as pullbacks of the product measures  $\pi_x^{\text{left}} \times \pi_x^{\text{right}}$  under  $\text{Sp}$ , in other words as the pushforwards

$$\pi_x = (\text{Sp}^{-1})_{\#}(\pi_x^{\text{left}} \times \pi_x^{\text{right}}).$$

Finally we integrate these up to a mixed measure  $\pi^{\text{mix}}$  by requiring for all Borel  $f: C([0, 1]; X) \rightarrow [0, \infty]$  that

$$\int_{C([0,1];X)} f(\gamma) d\pi^{\text{mix}}(\gamma) = \int_{C([0,1];X)} \int_{(e_t)^{-1}(x)} f(\gamma) d\pi_x(\gamma) d((e_1)_{\#}\pi^{\text{left}})(x).$$

Since  $\pi$  is optimal, there exists a set  $\Gamma \subset \text{Geo}(X)$  such that  $\pi(\Gamma) = 1$  and the set  $\{(\gamma_0, \gamma_1) : \gamma \in \Gamma\}$  is cyclically monotone. Because  $\pi^1, \pi^2 \ll \pi$ , also  $\pi^1(\Gamma) = \pi^2(\Gamma) = 1$ . For any pair  $\gamma^1, \gamma^2 \in \Gamma$  with  $\gamma_t^1 = \gamma_t^2$  we have by the cyclical monotonicity and the triangle inequality that

$$\begin{aligned} d^2(\gamma_0^1, \gamma_1^1) + d^2(\gamma_0^2, \gamma_1^2) &\leq d^2(\gamma_0^1, \gamma_1^2) + d^2(\gamma_0^2, \gamma_1^1) \\ &\leq (tl(\gamma^1) + (1-t)l(\gamma^2))^2 + (tl(\gamma^2) + (1-t)l(\gamma^1))^2 \\ &= l(\gamma^1)^2 + l(\gamma^2)^2 - 2t(1-t)(l(\gamma^1) - l(\gamma^2))^2 \\ &\leq l(\gamma^1)^2 + l(\gamma^2)^2 = d^2(\gamma_0^1, \gamma_1^1) + d^2(\gamma_0^2, \gamma_1^2), \end{aligned} \tag{4.1}$$

and so all the inequalities in the above chain (4.1) are equalities. Consequently  $l(\gamma^1) = l(\gamma^2)$  meaning that for  $\mathfrak{m}$ -almost every  $x \in X$  there exists  $l_x$  such that  $\pi_x^{\text{left}}$  is concentrated on geodesics of length  $tl_x$  and  $\pi_x^{\text{right}}$  is concentrated on geodesics of length  $(1-t)l_x$ . The measure  $\pi^{\text{mix}}$  is concentrated on

$$\{\gamma \in C([0, 1]; X) : \text{there exist } \gamma^1, \gamma^2 \in \Gamma \text{ s.t. } \text{restr}_0^t \gamma = \text{restr}_0^t \gamma^1 \text{ and } \text{restr}_t^1 \gamma = \text{restr}_t^1 \gamma^2\},$$

which by the equalities in (4.1) is a subset of  $\text{Geo}(X)$ .

Furthermore, since  $(e_i)_{\#}\pi^{\text{mix}} = (e_i)_{\#}(\frac{1}{2}(\pi^1 + \pi^2))$  for  $i = 0, 1$  and

$$\begin{aligned} \int_{\text{Geo}(X)} \frac{1}{2} d^2(\gamma_0, \gamma_1) d(\pi^1 + \pi^2)(\gamma) &= \int_X \frac{1}{2} l_x^2 d((e_t)_{\#}(\pi^1 + \pi^2))(x) \\ &= \int_X l_x^2 d((e_t)_{\#}\pi^{\text{mix}})(x) = \int_{\text{Geo}(X)} d^2(\gamma_0, \gamma_1) d\pi^{\text{mix}}(\gamma), \end{aligned}$$

the measure  $\pi^{\text{mix}}$  is optimal.

Because  $\pi^1 \perp \pi^2$  and because  $\{\rho_t^1 > 0\}$  and  $\{\rho_t^2 > 0\}$  intersect in the set  $E$  of positive  $\mathfrak{m}$ -measure, we know that for  $\mathfrak{m}$ -almost every  $x \in E$  at least one of the measures  $\pi_x^{\text{left}}, \pi_x^{\text{right}}$  is not a Dirac mass. Therefore  $\pi^{\text{mix}}$  is not essentially non-branching or the time-inverse  $I_{\#}\pi^{\text{mix}}$  defined via the mapping

$$I: \text{Geo}(X) \rightarrow \text{Geo}(X): \gamma \mapsto (\gamma': [0, 1] \rightarrow X: t \mapsto \gamma_{1-t})$$

is not essentially non-branching. This contradicts Theorem 1.1.

**Acknowledgments.** *T.R. acknowledges the financial support of the European Project ERC AdG \*GeMeThNES\* and the Academy of Finland project no. 137528. The authors also thank the anonymous referee for carefully reading the paper.*

## REFERENCES

- [1] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, in Modelling and Optimisation of Flows on Networks (B. Piccoli and M. Rascle eds.), Lecture Notes in Mathematics, Springer, Berlin, 2013, pp. 1–155.
- [2] L. Ambrosio, N. Gigli, A. Mondino and T. Rajala, *Riemannian Ricci curvature lower bounds in metric measure spaces with  $\sigma$ -finite measure*, Trans. Amer. Math. Soc., to appear, arXiv:1207.4924.
- [3] L. Ambrosio, N. Gigli and G. Savaré, *Calculus and heat flows in metric measure spaces with Ricci curvature bounded from below*, Invent. Math., to appear, arXiv:1106.2090.
- [4] L. Ambrosio, N. Gigli and G. Savaré, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, preprint, arXiv:1109.0222.
- [5] L. Ambrosio and T. Rajala, *Slopes of Kantorovich potentials and existence of optimal transport maps in metric measure spaces*, Ann. Mat. Pura Appl., to appear, arXiv:1111.5119.
- [6] J. Bertrand, *Existence and uniqueness of optimal maps on Alexandrov spaces*, Adv. Math., **219** (2008), no. 3, 838–851.
- [7] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, C. R. Acad. Sci. Paris Sér. I Math., **305** (1987), 805–808.
- [8] S. Daneri and G. Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal. **40** (2008), no. 3, 1104–1122.
- [9] N. Gigli, *Optimal maps in non branching spaces with Ricci curvature bounded from below*, Geom. Funct. Anal. **22** (2012), no. 4, 990–999.
- [10] N. Gigli, A. Mondino and G. Savaré, *A notion of pointed convergence of non-compact metric measure spaces and stability of Ricci curvature bounds and heat flows*, preprint, (2012).
- [11] W. Jankoff, *Sur l'uniformisation des ensembles A*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **30** (1941), 597–598.
- [12] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. **169** (2009), no. 3, 903–991.
- [13] R. J. McCann, *Polar factorization of maps on riemannian manifolds*, Geom. Funct. Anal. **11** (2001), 589–608.
- [14] T. Rajala, *Interpolated measures with bounded density in metric spaces satisfying the curvature-dimension conditions of Sturm*, J. Funct. Anal., **263** (2012), no. 4, 896–924.
- [15] T. Rajala, *Local Poincaré inequalities from stable curvature conditions on metric spaces*, Calc. Var. Partial Differential Equations, **44** (2012), no. 3–4, 477–494.
- [16] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), no. 1, 65–131.
- [17] K.-T. Sturm, *On the geometry of metric measure spaces. II*, Acta Math. **196** (2006), no. 1, 133–177.
- [18] C. Villani, *Optimal transport. Old and new*, vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.

DEPARTMENT OF MATHEMATICS AND STATISTICS, P.O. BOX 35 (MAD), FI-40014 UNIVERSITY OF JYVÄSKYLÄ,  
FINLAND

*E-mail address:* `tapio.m.rajala@jyu.fi`

INSTITUT FÜR ANGEWANDTE MATHEMATIK, UNIVERSITÄT BONN, ENDENICHER ALLEE 60, D-53115 BONN,  
GERMANY

*E-mail address:* `sturm@uni-bonn.de`