

Slopes of Kantorovich potentials and existence of optimal transport maps in metric measure spaces

Luigi Ambrosio · Tapio Rajala

Received: 24 November 2011 / Accepted: 2 March 2012 / Published online: 14 March 2012
© Fondazione Annali di Matematica Pura ed Applicata and Springer-Verlag 2012

Abstract We study optimal transportation with the quadratic cost function in geodesic metric spaces satisfying suitable non-branching assumptions. We introduce and study the notions of slope along curves and along geodesics, and we apply the latter to prove suitable generalizations of Brenier’s theorem of existence of optimal maps.

Keywords Optimal transportation · Geodesic metric space · Non-branching · Upper gradient

Mathematics Subject Classification Primary 49Q20 · 53C23

1 Introduction

The problem of finding an optimal way to transport mass has a long history, starting from Monge’s seminal paper [14]. The optimality of a transport can be measured in many ways, depending on the choice of the cost function. In this paper, we focus on the case when the cost is the square of the distance.

Given two positive and finite measures μ and ν on some metric space (X, d) with the same total mass, which we may normalize to 1, our task is then to study whether the infimum

$$\inf_X \int d^2(x, T(x)) d\mu(x), \quad (1.1)$$

over all possible μ -measurable maps $T : X \rightarrow X$, which send the measure μ to ν , is attained. If such a minimizing map exists, we call it an optimal transport map between μ and ν .

L. Ambrosio (✉) · T. Rajala
Scuola Normale Superiore, Piazza dei Cavalieri 7, 56127 Pisa, Italy
e-mail: luigi.ambrosio@sns.it

T. Rajala
e-mail: tapio.rajala@sns.it

Existence of optimal maps or even of admissible ones is problematic, for instance, no admissible map exists when μ is a Dirac mass and ν is not a Dirac mass.

Kantorovich's relaxed [9, 10] formulation of the optimal transport problem consists in finding the infimum

$$\inf_{X \times X} \int d^2(x, y) \pi(x, y) \quad (1.2)$$

over all possible transport plans, that is probability measures π on $X \times X$ that have μ and ν as marginals. Again, if there is a measure that attains the infimum, it is called an optimal transport plan between μ and ν . Notice that transport plans can split measure, and so they avoid the problem faced by transport maps. In fact, not only the Kantorovich formulation of the problem is well-posed, but also the infimum is attained (possibly infinite) under the only assumption that (X, d) is complete and separable. Since we will be dealing with geodesic metric spaces, we will mostly work with the equivalent formulation in terms of geodesic transport plans, that is probability measures in the space $\text{Geo}(X)$ of constant speed geodesics parameterized on $[0, 1]$, with marginal conditions at $t = 0$ and $t = 1$.

In general, it seems to be a difficult problem to find necessary and sufficient conditions under which Monge's problem has a solution, but by now, several sufficient conditions are known. For the quadratic cost in the Euclidean setting, it was proved independently by Brenier [6] and Smith and Knott [16] that there exists a unique optimal map T , given by the gradient of a convex function, provided that μ is absolutely continuous with respect to the Lebesgue measure. This result was generalized to Riemannian manifolds by McCann [13], to Alexandrov spaces by Bertrand [5], to the Heisenberg group by Ambrosio and Rigot [4] and, very recently, to non-branching metric spaces with Ricci curvature bounded from below (in the sense of Lott, Sturm and Villani) by Gigli [8]. Notice that in all these results, a reference measure m (Lebesgue measure, Riemannian volume, Haar measure, etc.) plays a role, so the proper setting for this question is the family of metric measure spaces (X, d, m) .

In another recent paper [1], a metric Brenier theorem is proved under mild assumptions on (X, d, m) , see Theorem 10.3 and Remark 10.7 therein. In the case when (X, d) has bounded diameter and m is a finite measure, the main assumption is the existence of bounds on the relative entropy along geodesics (a condition weaker than the $CD(K, \infty)$ condition of Lott, Sturm and Villani), and the metric Brenier theorem states that, for any optimal geodesic plans π , it holds

$$|\nabla^+ \varphi|(\gamma_0) = d(\gamma_0, \gamma_1) \quad \pi \text{-a.e. in } \text{Geo}(X) \quad (1.3)$$

(here φ is any Kantorovich potential and $|\nabla^+ \varphi|$ is its ascending slope). In other words, the transportation distance depends μ -a.e. only on the initial point.

This result raises some questions that we plan to investigate in this paper: the first one is to understand under which additional assumptions one can really recover an optimal map, the second one is about the differentiability of φ along geodesics used by the optimal plan.

In connection with the first question, we start from this heuristic idea more or less implicit in many proofs: under appropriate structural assumptions on the space, (1.3) identifies the "initial velocity" of the geodesic. Indeed, assuming suitable non-branching assumptions on the space and on its tangent metric spaces, we can perform a suitable blow-up analysis that leads to the existence of optimal maps. The proof of this result requires a detailed analysis of the proof of the metric Brenier theorem in [1] and the introduction of a sharper notion of ascending slope, namely the ascending slope $|\nabla_g^+ \varphi|$ along geodesics. Since we believe that this concept has an independent interest, we compare this slope to the usual one and to the

slope along curves, we provide an example and we raise some open problems. Coming back to the existence of optimal maps, our result covers as a particular case the Euclidean, the Riemannian and the Alexandrov case, see also the paragraph immediately after Theorem 4.3 for a more detailed discussion.

In connection with the second question, it has already been proved in Theorem 10.4 of [1] a “differentiability in mean,” namely

$$\lim_{t \downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{d(\gamma_0, \gamma_t)} = |\nabla^+ \varphi|(\gamma_0) \quad \text{in } L^2(\pi). \tag{1.4}$$

This weak differentiability property plays an important role in the subsequent paper [2], for the computation of the derivative of the entropy along geodesics. Here, under an additional doubling assumption on m , we are able to improve (1.4) to a pointwise differentiability property, so that

$$\varphi(\gamma_t) = \varphi(\gamma_0) - t|\nabla^+ \varphi|^2(\gamma_0) + o(t)$$

for π -a.e. $\gamma \in \text{Geo}(X)$.

2 Non-branching metric spaces

Let us start by laying out the definitions for the metric spaces that will be used in this paper. First of all, we will be working exclusively in metric spaces (X, d) that are complete and separable. Second, by measure in (X, d) , we mean a non-negative Borel measure, finite on bounded sets. We will mainly consider metric spaces X equipped with a *doubling measure* m meaning that there exists a constant $0 < C < \infty$ so that for all $0 < r < \text{diam}(X)$ and $x \in X$ we have

$$m(B(x, 2r)) \leq Cm(B(x, r)).$$

A related notion for metric spaces where the measure has not been specified is that of a *doubling metric space*, which means that there exists an integer $N \geq 1$ so that, for all $0 < r < \infty$, any ball of radius $2r$ can be covered by N balls of radius r . It is obvious that if there exists a doubling measure on X , then the space X has to be doubling as well. The converse is also true for complete metric space, see for example [11, 12].

We call any absolutely continuous map $\gamma : [a, b] \rightarrow X$ a curve and use the abbreviation $\gamma_s = \gamma(s)$. The length of the curve γ is defined as

$$l(\gamma) = \sup \left\{ \sum_{i=1}^N d(\gamma_{t_i}, \gamma_{t_{i-1}}) : a \leq t_0 < t_1 < \dots < t_N \leq b, N \in \mathbb{N} \right\}.$$

We call the curve $\gamma : [a, b] \rightarrow X$ a *geodesic* if $l(\gamma) = d(\gamma_a, \gamma_b)$. The metric space X itself is called *geodesic* if any two points $x, y \in X$ can be connected with a geodesic, that is, there exists a geodesic $\gamma : [a, b] \rightarrow X$ with $\gamma_a = x$ and $\gamma_b = y$. Sometimes, when there is no danger of confusion, we also call the image of a geodesic a geodesic.

The speed of a curve γ is given by

$$|\dot{\gamma}|(t) = \lim_{s \rightarrow t} \frac{d(\gamma_s, \gamma_t)}{|s - t|}$$

whenever the limit exists. It is not hard to prove, see for instance Theorem 1.1.2 in [3], that it indeed exists at \mathcal{L}^1 -almost every point $t \in [a, b]$, where \mathcal{L}^1 is the Lebesgue measure on \mathbb{R} , and that $l(\gamma) = \int_a^b |\dot{\gamma}|(t) dt$.

We denote by $\text{Geo}(X)$ the set of all constant speed geodesics in X , which are parametrized by $[0, 1]$, namely $d(\gamma_s, \gamma_t) = |t - s|d(\gamma_0, \gamma_1)$ for all $s, t \in [0, 1]$. By a reparameterization argument, constant speed geodesics connecting any two given points exist in any geodesic space. We equip the space $\text{Geo}(X)$ with the distance

$$d^*(\gamma, \tilde{\gamma}) = \max_{t \in [0, 1]} d(\gamma_t, \tilde{\gamma}_t)$$

and note that $(\text{Geo}(X), d^*)$ is also complete and separable since the underlying metric space is. We will also use the convenient notation of *evaluation map* $e_t : \text{Geo}(X) \rightarrow X$, defined as $e_t(\gamma) = \gamma_t$ for all $t \in [0, 1]$.

With the basic notation related to geodesics now fixed, we are ready to introduce the two definitions of non-branching, which play a crucial role in our results.

Definition 2.1 We call a geodesic metric space (X, d) *non-branching* if for any two constant speed geodesics $\gamma, \gamma' : [0, 1] \rightarrow X$ with $\gamma_0 = \gamma'_0$ and $\gamma_s = \gamma'_s$ for some $s \in (0, 1)$ we have $\gamma_t = \gamma'_t$ for all $t \in [0, 1]$.

We would like to use non-branching on the level of the tangent spaces. However, two distinct geodesics of a metric space can collapse into a single geodesic of the tangent space in the blow-up. To control such collapsing, we will assume a stronger version of non-branching.

Definition 2.2 We call a geodesic metric space (X, d) *strongly non-branching* if for any two constant speed geodesics $\gamma, \gamma' : [0, 1] \rightarrow X$ with $\gamma_0 = \gamma'_0, \gamma_1 \neq \gamma'_1$ and $d(\gamma_0, \gamma_1) > 0$, we have

$$\liminf_{t \downarrow 0} \frac{d(\gamma_t, \gamma'_t)}{d(\gamma_0, \gamma_1)} > 0. \tag{2.1}$$

In our main theorem, Theorem 4.3, we assume that the space is strongly non-branching and that at almost every point, we have some non-branching tangent space. Before defining what we mean by tangent space, we recall the definitions of Hausdorff and Gromov-Hausdorff distance. The *Hausdorff distance* between closed sets $A, B \subset X$ is defined as

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}.$$

Using the Hausdorff distance, the *Gromov-Hausdorff distance* between two metric spaces (X, d_X) and (Y, d_Y) is then defined as

$$d_{GH}(X, Y) = \inf d_H(f(X), g(Y)),$$

where the infimum is taken over all metric spaces (Z, d_Z) and isometries $f : X \rightarrow Z, g : Y \rightarrow Z$. Finally, a sequence $(X_n, d_n, x_n)_{n=1}^\infty$ of metric spaces (X_n, d_n) and points $x_n \in X_n$ is said to converge to (X, d, x) in the *pointed Gromov-Hausdorff* sense if

$$\lim_{n \rightarrow \infty} d_{GH}(\overline{B}_{X_n}(x_n, r), \overline{B}_X(x, r)) = 0 \quad \forall r > 0,$$

where by \overline{B} we denote the closed ball. Given a metric space (X, d) and a scaling factor $r > 0$, we define a rescaled metric d_r on X by setting

$$d_r(x, y) = \frac{1}{r} d(x, y)$$

for all $x, y \in X$.

Definition 2.3 Let (X, d) be a metric space. We call a metric space (Y, ρ) *tangent* to (X, d) at $x \in X$ if there exist a sequence $(r_n) \downarrow 0$ and $y \in Y$ so that

$$(X, d_{r_n}, x) \xrightarrow[n \rightarrow \infty]{} (Y, \rho, y)$$

in the pointed Gromov-Hausdorff convergence.

Notice that our definition of a tangent space is weaker than Gromov’s original notion. He required the tangent space to be the full limit of the spaces (X, d_r, x) as $r \downarrow 0$, whereas in our definition, we only require convergence along a subsequence. As a consequence, with our definition, the space (X, d) can in principle have a huge collection of different tangent spaces at a single point.

We will use the following well-known result (see for instance [2, Proposition 2.7]) that allows us to move from the Gromov-Hausdorff convergence to Hausdorff convergence. Recall that a sequence (X_n, d_n) of metric spaces is equi-compact if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that any of the spaces (X_n) can be covered by at most N balls with radius ϵ .

Theorem 2.4 *If $(X_n, d_n) \rightarrow (X, d)$ in the Gromov-Hausdorff convergence, then there exist a space (Z, d_Z) and isometric embeddings*

$$i_n : (X_n, d_n) \rightarrow (Z, d_Z), \quad i : (X, d) \rightarrow (Z, d_Z)$$

so that $i_n(X_n) \rightarrow i(X)$ in the Hausdorff convergence.

In addition, if (X_n, d_n) are equi-compact, then (Z, d_Z) can be taken to be a compact metric space.

3 Gradients along geodesics

Let us recall some basic definitions in measure theory. The collection of *universally measurable sets* of the space X , denoted by $\mathcal{B}^*(X)$, is the σ -algebra of the sets, which are μ -measurable for all finite non-negative Borel measure μ of (X, d) . The collection of all Borel sets of (X, d) will be denoted by $\mathcal{B}(X)$.

Now we turn to our next set of definitions that concern metric differentials.

Definition 3.1 Given a function $f : X \rightarrow \mathbb{R}$, we define the *lower ascending slope along geodesics* of f at $x \in X$ as

$$|\nabla_g^+ f|(x) = \sup_{\gamma} \liminf_{s \downarrow 0} \frac{[f(\gamma_s) - f(x)]^+}{d(\gamma_s, x)},$$

where $^+$ denotes the positive part and the supremum is taken over all non-constant geodesics in X that start from the point x .

Here *ascending* refers to the fact that we are taking the positive part of the difference quotient, and *lower* refers to the fact that we are taking the \liminf , rather than the \limsup .

Proposition 3.2 *Suppose that $f : X \rightarrow \mathbb{R}$ is continuous, then $|\nabla_g^+ f|$ is universally measurable.*

Proof Let $T \in \mathbb{R}$ and consider the sublevel set $\{|\nabla_g^+ f| > T\} \subset X$. For the universal measurability, it is sufficient to show that this set is Suslin. Because X is complete and separable,

so are $\text{Geo}(X)$ and $X \times \text{Geo}(X)$. Therefore $\{|\nabla_g^+ f| > T\}$, being the projection of the set

$$\left\{ (x, \gamma) : \gamma_0 = x, \liminf_{s \downarrow 0} \frac{[f(\gamma_s) - f(x)]^+}{d(\gamma_s, x)} > T \right\} \subset X \times \text{Geo}(X)$$

to the space X is indeed Suslin since the projected set can be written as

$$\bigcap_{t \in \mathbb{Q} \cap (0, 1)} \bigcup_{s \in \mathbb{Q} \cap (0, t)} \left\{ (x, \gamma) : \gamma_0 = x, \frac{[f(\gamma_s) - f(x)]^+}{d(\gamma_s, x)} > T \right\}$$

and so it is Borel (countable intersection of countable unions of open sets). □

Definition 3.3 A function $g : X \rightarrow [0, \infty]$ is an *upper gradient along geodesics* of a function $f : X \rightarrow \mathbb{R}$, if for any $\gamma \in \text{Geo}(X)$, we have

$$|f(\gamma_0) - f(\gamma_1)| \leq \int_{\gamma} g, \tag{3.1}$$

where the integral along γ is understood as

$$\int_{\gamma} g = l(\gamma) \int_0^1 g(\gamma_s) ds.$$

The almost everywhere differentiability of Lipschitz functions on the real line implies that ascending slopes are upper gradients for Lipschitz functions. We include the easy proof of this fact here for the convenience of the reader. Recall that a function $f : X \rightarrow \mathbb{R}$ is called Lipschitz if there exists a constant $0 \leq L < \infty$ so that for any two points $x, y \in X$, we have

$$|f(x) - f(y)| \leq Ld(x, y).$$

Proposition 3.4 *Let $f : X \rightarrow \mathbb{R}$ be Lipschitz. Then the lower ascending slope along geodesics is an upper gradient along geodesics.*

Proof By Proposition 3.2, the function $|\nabla_g^+ f|$ is universally measurable, and it is easily seen that this implies the \mathcal{L}^1 -measurability of $|\nabla_g^+ f| \circ \gamma$ (just consider the push forward under \mathcal{L}^1 of γ , see also [1, Lemma 2.4]. Take $\gamma \in \text{Geo}(X)$. The function $f \circ \gamma : [0, 1] \rightarrow \mathbb{R}$ is Lipschitz and therefore differentiable \mathcal{L}^1 -almost everywhere. In particular, $|(f \circ \gamma)'(t)| \leq l(\gamma)|\nabla_g^+ f|(\gamma_t)$ holds, and both sides of the inequality are well defined at \mathcal{L}^1 -almost every point $t \in [0, 1]$. Thus

$$|f(\gamma_0) - f(\gamma_1)| = \left| \int_0^1 (f \circ \gamma)'(s) ds \right| \leq \int_0^1 |(f \circ \gamma)'(s)| ds \leq l(\gamma) \int_0^1 |\nabla_g^+ f|(\gamma_s) ds.$$

□

It is interesting to compare the ascending slope and the upper gradient defined along geodesics to the more commonly used versions. First of all, it is immediate that we always have

$$|\nabla_g^+ f|(x) \leq |\nabla_c^+ f|(x) \leq |\nabla^+ f|(x), \tag{3.2}$$

where the usual *ascending slope* $|\nabla^+ f|(x)$ of f at a point x is defined as

$$|\nabla^+ f|(x) = \limsup_{y \rightarrow x} \frac{[f(y) - f(x)]^+}{d(y, x)}$$

and the *lower ascending slope along curves* as

$$|\nabla_c^+ f|(x) = \sup_{\gamma} \liminf_{s \downarrow 0} \frac{[f(\gamma_s) - f(x)]^+}{d(\gamma_s, x)}$$

with the supremum taken over all curves (recall that by convention all curves we consider are absolutely continuous). Moreover, the inequalities in (3.2) can be strict. Notice also that the choice of the lower concept (i.e. with the \liminf_s) is justified by the fact that the upper concept is easily seen to coincide with $|\nabla^+ f|$.

Recall that we have the usual notion of an *upper gradient* $g: X \rightarrow [0, \infty]$ of a function $f: X \rightarrow \mathbb{R}$ if we require the inequality (3.1) to hold along all curves on $[0, 1]$, where this time $\int_{\gamma} g$ is understood as $\int_0^1 g(\gamma_s)|\dot{\gamma}_s|ds$. It is not difficult to show, following the same proof given in Proposition 3.2, that ascending slopes along curves are universally measurable. Moreover, as in Proposition 3.4, one can prove that ascending slopes along curves are upper gradients for Lipschitz functions.

Ascending slopes along geodesics could be thought, identifying geodesics to the tangent space as in the theory of Alexandrov spaces, as directional one-sided derivatives. Hence, it is natural to ask whether ascending slopes along geodesics are also upper gradients (in the usual sense) for Lipschitz functions. This is not true in general, as we will see in the next example.

Example 3.5 There exist a separable complete geodesic metric space (X, d) and a Lipschitz function $f: X \rightarrow \mathbb{R}$ so that $|\nabla_g^+ f|$ is not an upper gradient of f .

Let us first construct the metric space (X, d) . We start the construction by taking a unit line segment, which we simply denote by $[0, 1]$. This line segment will eventually have length 2 with the distance we construct. Next for all $n \in \mathbb{N}$ and $0 \leq k < 2^n$, we connect the points $k2^{-n}$ and $(k + 1)2^{-n}$ in $[0, 1]$ with an arc $A_{n,k}$ of length $(2 - 2^{-n})2^{-n}$. In Fig. 1, the arcs $A_{n,k}$ are drawn as half-circles. The space X is then the disjoint union of the arcs $A_{n,k}$ and the initial line segment $[0, 1]$.

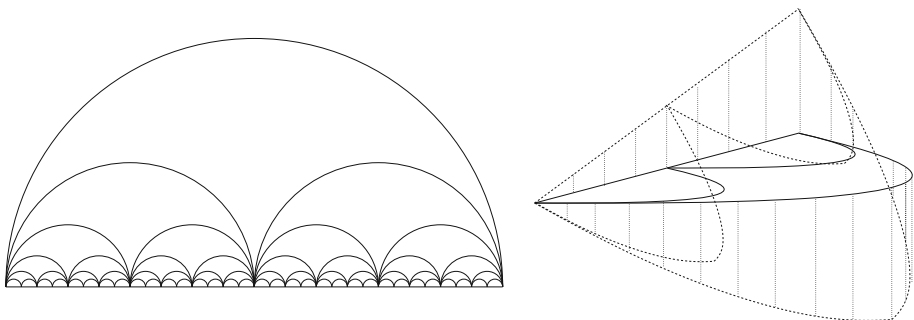


Fig. 1 On the *left* is an illustration of the space (X, d) . The lengths of the curves are chosen so that the geodesics between two points on the interval prefer going along the longest curves. On the *right* is the graph of the Lipschitz function f drawn along a part of the interval and along a couple of the construction curves joining the points of the interval

We define the distance d between two points $x, y \in X$ as

$$d(x, y) = \inf \sum_i l(E_i),$$

where the infimum is taken over all collections of E_i 's that connect the points x and y , E_i are subsets of the arcs, and $l(E_i)$ is the length of the piece determined by the length of the arc. This way on each arc $A_{n,k}$ the distance is given by the natural distance determined by the length of the arc. See the left part of Fig. 1 for an illustration of the space.

Let us check that (X, d) is geodesic. Let $x, y \in X$ be two distinct points. If it happens that x and y lie on the same arc then the segment of the arc joining the points is our geodesic. We may then assume that the points are not on the same arc. We may also assume that $x, y \in [0, 1]$. If this is not the case, for example $x \notin [0, 1]$, we simply notice that any curve connecting the points x and y must go via one of the end-points x' and x'' of the arc in which x lies in and that

$$d(x, y) = \min\{d(x, x') + d(x', y), d(x, x'') + d(x'', y)\}.$$

We can now find the geodesic between the points x and y with the following procedure. Let $(\gamma^i)_{i=0}^\infty$ be a sequence of curves joining x to y so that $\lim_{i \rightarrow \infty} l(\gamma^i) = d(x, y)$. Because the lengths of the arcs are chosen so that the shortest curve between points $k2^{-n}$ and $(k + 1)2^{-n}$ is the arc $A_{n,k}$, there exists $i_0 \in \mathbb{N}$ so that each $\gamma^i, i \geq i_0$, contains some $A_{n,k}$ with

$$n \leq \left\lceil \frac{\log |x - y|}{\log 2} \right\rceil + 1 \tag{3.3}$$

and with some k . So, taking a subsequence of (γ_i) , we may assume that all the curves contain the same arc $A_{n,k}$. Continuing inductively in the same way with the end-points of this arc and the points x and y , and finally using a diagonal argument, we obtain the geodesic.

We define the Lipschitz function $f: X \rightarrow \mathbb{R}$ first on $[0, 1]$ by setting $f|_{[0,1]}(x) = x$. This fixes the function on the end-points of all the arcs. We continue it inside the arcs by defining for all $n \in \mathbb{N}$ and $0 \leq k < 2^n$

$$f(\gamma_t) = \begin{cases} k2^{-n} - 2^{-n+1}t, & \text{for } t \in [0, 1/2]; \\ (k - 3)2^{-n} + 2^{-n+2}t, & \text{for } t \in [1/2, 1], \end{cases}$$

where $\gamma: [0, 1] \rightarrow X$ is the constant speed geodesic joining $k2^{-n}$ to $(k + 1)2^{-n}$ in $A_{n,k}$. See the right part of Fig. 1 for the graph of the function along a couple of the arcs and the line $[0, 1]$.

Let us now show that $|\nabla_g^+ f|(x) = 0$ for all $x \in [0, 1]$. To see this take a geodesic γ starting from a point $x \in [0, 1]$. If γ near the point x consists only of one piece of an arc, the equality

$$\liminf_{s \downarrow 0} \frac{[f(\gamma_s) - f(x)]^+}{d(\gamma_s, x)} = 0$$

is immediate. Suppose then that for every $\epsilon > 0$, there exists a point $y \in [0, 1]$, which is in the considered geodesic γ and $0 < d(x, y) < \epsilon$. As we have noted before, the part of γ that connects x to y must contain an arc $A_{n,k}$ with n bounded as in (3.3) and with some k . Moreover, we can take such an $A_{n,k}$ that $x > (k - 1)2^{-n}$. Let z be the middle point of $A_{n,k}$. Now $f(z) = (k - 1)2^{-n} < f(x)$, and so indeed $|\nabla_g^+ f|(x) = 0$.

On the other hand, the constant speed curve $\gamma: [0, 1] \rightarrow [0, 1]$ has length 2 and is therefore an admissible test curve for the upper gradient property.

It is easy to see that the space X in our previous example is not doubling and that it is extremely branching. In light of the example, one could still hope many natural conjectures to be true.

Question 3.6 Which assumptions are needed on the metric space (X, d) and on the measure m to ensure for any Lipschitz function $f : X \rightarrow \mathbb{R}$ that

- (i) the equality $|\nabla_g^+ f|(x) = |\nabla_c^+ f|(x)$ holds at m -almost every point $x \in X$?
- (ii) the function $|\nabla_g^+ f|(x)$ is an upper gradient of f ?

Notice that if we assume that m is a doubling measure on (X, d) , then $|\nabla^\pm f| = |\nabla f|$ m -a.e. in X (see [1, Proposition 2.5] for the simple proof of this fact) for any Lipschitz function f .

In addition, if (X, d, m) supports a $(1, 1)$ -Poincaré inequality, then we can apply Cheeger’s theory [7] to obtain $|\nabla f| \leq g$ m -a.e. in X for any (weak) upper gradient of f . Choosing $g = |\nabla_c^+ f|$ yields

$$|\nabla_c^+ f| = |\nabla^+ f| = |\nabla f| \quad m \text{-a.e. in } X.$$

For the same reason, under the same doubling and Poincaré assumptions, a positive answer to question (ii) implies a positive answer to question (i): indeed, choosing $g = |\nabla_g^+ f|$ one obtains

$$|\nabla_g^+ f| = |\nabla_c^+ f| = |\nabla^+ f| = |\nabla f| \quad m \text{-a.e. in } X.$$

4 Mass transportation in metric spaces

Before stating and proving our main results, we briefly discuss in the next subsection the basic properties of transport plans and maps in the general metric space setting. A comprehensive treatment of the theory can be found for example in [3, 17].

4.1 Basic properties of transport plans

Let $\mathcal{P}(X)$ denote the set of all Borel probability measures on X . The Wasserstein distance between two measures $\mu, \nu \in \mathcal{P}(X)$ is defined as

$$W_2(\mu, \nu) = \left(\inf_{\gamma} \int_{X \times X} d^2(x, y) d\gamma(x, y) \right)^{1/2}, \tag{4.1}$$

where the infimum is taken over all *transport plans* γ between μ and ν , that is measures $\gamma \in \mathcal{P}(X \times X)$ for which $\mathfrak{p}_\#^1 \gamma = \mu$ and $\mathfrak{p}_\#^2 \gamma = \nu$. Here the mappings \mathfrak{p}^1 and \mathfrak{p}^2 denote the projections to the first and second coordinate, respectively. The notation $f_\# \mu$ for a measure $\mu \in \mathcal{P}(X)$ and a μ -measurable mapping $f : X \rightarrow Y$ means the *push-forward* measure defined as $f_\# \mu(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{B}(Y)$. Notice that in general, $W_2(\mu, \nu)$ might be infinite. We call a transport plan γ_0 between two measures $\mu, \nu \in \mathcal{P}(X)$, for which $W_2(\mu, \nu) < \infty$, *optimal* if the infimum in (4.1) is attained at $\gamma = \gamma_0$.

Since we are dealing with geodesic spaces, we can equivalently consider geodesic transport plans. We define the set of *geodesic plans* between μ and ν as the set of all $\pi \in \mathcal{P}(\text{Geo}(X))$ for which $(e_0)_\# \pi = \mu, (e_1)_\# \pi = \nu$. We say that a geodesic plan is optimal, and write

$\pi \in \text{GeoOpt}(\mu, \nu)$, if

$$\int_{\text{Geo}(X)} d^2(\gamma_0, \gamma_1) d\pi(\gamma) = W_2^2(\mu, \nu) < \infty.$$

Given an optimal geodesic plan $\pi \in \text{GeoOpt}(\mu, \nu)$, it is clear that $(e_0, e_1)_\# \pi$ is an optimal plan. Conversely, making a measurable selection of constant speed geodesics γ^{xy} from x to y and considering the law of $(x, y) \mapsto \gamma^{xy}$ under γ , any optimal plan can be “lifted” to an optimal geodesic plan with the same cost.

The Kantorovich formulation of the transportation problem has also a very useful dual formulation: The minimum in (1.2) is equal to

$$2 \sup \left\{ \int_X \varphi(x) d\mu(x) + \int_X \psi(y) d\nu(y) \right\},$$

where the supremum is taken among all pairs $(\varphi, \psi) \in C_b^0(X) \times C_b^0(X)$ satisfying $\varphi(x) + \psi(y) \leq \frac{1}{2}d^2(x, y)$.

We define the *c-transform* of a function $\varphi: X \rightarrow \mathbb{R} \cup \{-\infty\}$ as

$$\varphi^c(x) = \inf_{y \in X} \left\{ \frac{d^2(x, y)}{2} - \varphi(y) \right\}.$$

A function ψ is called *c-concave* if $\psi = \varphi^c$ for some function φ . This terminology (*c-transform*, *c-concavity*) refers to a general cost function c . Here and in the sequel, the cost function c is given by the halved square of the distance.

Definition 4.1 Given an optimal geodesic plan $\pi \in \text{GeoOpt}(\mu, \nu)$, we call a Borel function $\varphi: X \rightarrow \mathbb{R} \cup \{-\infty\}$ a *Kantorovich potential* (relative to the optimal geodesic plan π) if it is *c-concave* and

$$\varphi(\gamma_0) + \varphi^c(\gamma_1) = \frac{d^2(\gamma_0, \gamma_1)}{2} \quad \text{for } \pi \text{-a.e. } \gamma \in \text{Geo}(X).$$

Notice that because the Kantorovich potential φ is *c-concave*, we have $\varphi = (\varphi^c)^c$ and that we make no integrability assumption on φ and φ^c .

A set $\Gamma \subset X \times X$ is called *cyclically monotone* if

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

for any $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ and permutation σ of $\{1, \dots, n\}$.

Suppose that $W_2(\mu, \nu) < \infty$ and $\pi \in \text{GeoOpt}(\mu, \nu)$, then $(e_0, e_1)_\# \pi$ is supported on a cyclically monotone set, and a Kantorovich potential relative to π exists.

4.2 Brenier theorem in metric spaces

As a starting point, we prove the following result that originates from [1, Theorem 10.3]. The difference compared to the original result is in the definition of the ascending slope, here replaced by the lower ascending slope. Also, we like to repeat the proof in our situation for the convenience of the reader.

Proposition 4.2 *Suppose that m is a finite measure on a bounded space (X, d) and $\mu = \rho m \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(X)$. Take $\pi \in \text{GeoOpt}(\mu, \nu)$ and let $\varphi: X \rightarrow \mathbb{R} \cup \{-\infty\}$ be a Kantorovich potential relative to π . If there exists $\bar{s} \in (0, 1)$ satisfying $(e_s)_\# \pi = \rho_s m$ for all $s \in (0, \bar{s})$ and*

$$\limsup_{s \downarrow 0} \int_X \rho_s \log \rho_s dm < \infty, \tag{4.2}$$

then

$$|\nabla_g^+ \varphi|(\gamma_0) = d(\gamma_0, \gamma_1) \text{ for } \pi \text{-a.e. } \gamma \in \text{Geo}(X).$$

Proof By the very definition of the Kantorovich potential, we have

$$\varphi(\gamma_0) = \frac{d^2(\gamma_0, \gamma_1)}{2} - \varphi^c(\gamma_1) \tag{4.3}$$

for π -a.e. $\gamma \in \text{Geo}(X)$. On the other hand, for any $z \in X$, we have

$$\varphi(z) \leq \frac{d^2(z, \gamma_1)}{2} - \varphi^c(\gamma_1). \tag{4.4}$$

Thus combining these two we get that for π -a.e. γ it holds

$$\begin{aligned} |\nabla_g^+ \varphi|(\gamma_0) &\leq \limsup_{z \rightarrow \gamma_0} \frac{[\varphi(z) - \varphi(\gamma_0)]^+}{d(z, \gamma_0)} \leq \limsup_{z \rightarrow \gamma_0} \frac{[d^2(z, \gamma_1) - d^2(\gamma_0, \gamma_1)]^+}{2d(z, \gamma_0)} \\ &\leq \limsup_{z \rightarrow \gamma_0} \frac{d^2(z, \gamma_0) + 2d(\gamma_0, \gamma_1)d(z, \gamma_0)}{2d(z, \gamma_0)} = d(\gamma_0, \gamma_1). \end{aligned}$$

Let us now prove the converse inequality in an integral form. Taking $z = \gamma_t$ in (4.4) and combining it with (4.3), we obtain

$$\varphi(\gamma_0) - \varphi(\gamma_t) \geq \frac{d^2(\gamma_0, \gamma_1)}{2} - \frac{d^2(\gamma_t, \gamma_1)}{2} = \frac{2t - t^2}{2} d^2(\gamma_0, \gamma_1). \tag{4.5}$$

Because X is bounded, φ is Lipschitz and so by Proposition 3.4, the function $|\nabla_g^+ \varphi|$ is an upper gradient of φ along geodesics. Thus we have

$$(\varphi(\gamma_0) - \varphi(\gamma_t))^2 \leq \left(\int_0^t |\nabla_g^+ \varphi|(\gamma_s) d(\gamma_0, \gamma_1) ds \right)^2 \leq t d^2(\gamma_0, \gamma_1) \int_0^t |\nabla_g^+ \varphi|^2(\gamma_s) ds.$$

Dividing this by $d^2(\gamma_0, \gamma_t)$ and integrating over $\text{Geo}(X)$ yields

$$\begin{aligned} \frac{1}{t} \int_0^t \int_X |\nabla_g^+ \varphi|^2(x) \rho_s dm(x) ds &= \frac{1}{t} \int_0^t \int_{\text{Geo}(X)} |\nabla_g^+ \varphi|^2(\gamma_s) d\pi(\gamma) ds \\ &\geq \int_{\text{Geo}(X)} \left(\frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{d(\gamma_0, \gamma_t)} \right)^2 d\pi(\gamma) \geq \frac{2-t}{2} \\ &\quad \times \int_{\text{Geo}(X)} d^2(\gamma_0, \gamma_1) d\pi(\gamma) \end{aligned}$$

Recall that m is finite. From the assumption (4.2), we know by the Dunford-Pettis theorem that ρ_s converges to ρ weakly in $L^1(X, m)$ and so

$$\int_X g \rho_s dm \rightarrow \int_X g \rho dm \quad \text{as } s \downarrow 0 \text{ for all } g \in L^\infty(X, m).$$

Since $|\nabla_g^+ \varphi|$ is bounded, this holds also for $g = |\nabla_g^+ \varphi|^2$. Therefore,

$$\begin{aligned} \int_X |\nabla_g^+ \varphi|^2 d\mu &= \lim_{t \downarrow 0} \frac{1}{t} \int_0^t \int_X |\nabla_g^+ \varphi|^2(x) \rho_s dm(x) ds \\ &\geq \lim_{t \downarrow 0} \frac{2-t}{2} \int_{\text{Geo}(X)} d^2(\gamma_0, \gamma_1) d\pi(\gamma). \end{aligned}$$

□

With the help of the Proposition 4.2, we are now able to prove a Brenier-type theorem in strongly non-branching metric spaces.

Theorem 4.3 *Assume that (X, d) is a strongly non-branching geodesic metric space equipped with a doubling measure m , and that $\mu = \rho m \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(X)$ satisfy $W_2(\mu, \nu) < \infty$. Assume also that:*

- (a) *for m -almost every point $x \in X$, the space X has a non-branching tangent space at x ;*
- (b) *there exists a transport plan $\pi \in \text{GeoOpt}(\mu, \nu)$ such that for all $s \in [0, 1)$ sufficiently small we have $(e_s)_\# \pi \ll m$, and the densities ρ_s satisfy*

$$\limsup_{s \downarrow 0} \int_X \rho_s \log \rho_s dm < \infty.$$

Then the optimal geodesic plan π is given by a mapping $T: X \rightarrow X$, i.e. $\gamma_1 = T(\gamma_0)$ for π -a.e. $\gamma \in \text{Geo}(X)$.

Proof Suppose that there is no such T . If we then fix a point $x_0 \in X$ and consider the restricted and rescaled measures

$$\pi_r = \frac{\pi|_{A(r)}}{\pi(A(r))}, \quad \text{where } A(r) = \{\gamma \in \text{Geo}(X), \gamma_t \in B(x_0, r) \text{ for all } t \in [0, 1]\},$$

we notice that for large enough $r > 0$, the assumptions of the theorem are satisfied, and still there exists no such T . Therefore, we may assume the space X to be bounded.

Let φ be the Kantorovich potential relative to π . Let $x \in X$ be a point where the space X has a non-branching tangent. Suppose that there are two geodesics $\gamma, \tilde{\gamma} \in \text{Geo}(X)$ so that $\gamma_0 = \tilde{\gamma}_0 = x$,

$$|\nabla_g^+ \varphi|(x) = d(x, \gamma_1) = d(x, \tilde{\gamma}_1) \tag{4.6}$$

and

$$\varphi^c(\tilde{\gamma}_1) = \varphi^c(\gamma_1) = \frac{d^2(x, \gamma_1)}{2} - \varphi(x).$$

Let (Y, d_Y) be a non-branching tangent space to X at x and $r_n \downarrow 0$ a sequence so that

$$(X, d_{r_n}, x) \rightarrow (Y, d_Y, 0)$$

in the pointed Gromov-Hausdorff convergence as $n \rightarrow \infty$. Since we assumed (X, d, m) to be doubling, the spaces $(\overline{B}_{(X, d_{r_n})}(x, 1), d_{r_n})$ are easily seen to be equi-compact. Indeed, for any $\epsilon > 0$, we can find a maximal disjoint family of balls of radius $r_n\epsilon/2$ contained in $\overline{B}(x, r_n)$, so that the family of balls with doubled radius covers $\overline{B}(x, 1)$, and use the doubling inequality

$$\mu(B_r(y)) \geq C \left(\frac{r}{R}\right)^\alpha \mu(B_R(x)) \quad \text{whenever } B_r(y) \subset B_R(x)$$

(here $C > 0$ and $\alpha > 0$ depend on the doubling constant only) with $r = r_n\epsilon/2$ and $R = r_n$ to estimate the number of these balls with a constant depending only on C, α and ϵ . We can then apply Theorem 2.4 to obtain a compact space (Z, d_Z) and isometric embeddings

$$i_n : (\overline{B}_{(X, d_{r_n})}(x, 1), d_{r_n}) \rightarrow (Z, d_Z), \quad i : (\overline{B}_{(Y, d_Y)}(0, 1), d_Y) \rightarrow (Z, d_Z)$$

so that $i_n(\overline{B}_{(X, d_{r_n})}(x, 1)) \rightarrow i(\overline{B}_{(Y, d_Y)}(0, 1))$ in the Hausdorff convergence.

From (4.6) we get for every $n \in \mathbb{N}$ a constant speed geodesic γ^n with $\gamma_0^n = x$ and a radius $R_n > 0$ so that

$$\frac{(\varphi(\gamma_s^n) - \varphi(x))^+}{d(\gamma_s^n, x)} > |\nabla_g^+ \varphi|(x) - \frac{1}{n}$$

for every $s \in (0, 1)$ for which $d(\gamma_s^n, x) < R_n$.

Now, possibly taking a subsequence of $(r_n)_{n=1}^\infty$ so that $r_n < R_n$, we get a sequence of points $(y_n)_{n=1}^\infty \subset X$ with $d(y_n, x) = r_n$ and

$$\frac{(\varphi(y_n) - \varphi(x))^+}{d(y_n, x)} > |\nabla_g^+ \varphi|(x) - \frac{1}{n} = d(x, \gamma_1) - \frac{1}{n}. \tag{4.7}$$

Notice also that for any $y \in X$ we have

$$\varphi(y) \leq \frac{d^2(y, \gamma_1)}{2} - \varphi^c(\gamma_1) = \frac{d^2(y, \gamma_1) - d^2(x, \gamma_1)}{2} + \varphi(x). \tag{4.8}$$

Writing $z_n = \gamma_s$ for the s for which $d(\gamma_s, x) = r_n$, triangle inequality and geodesic property yield

$$d(y_n, \gamma_1) - d(y_n, z_n) \leq d(z_n, \gamma_1) = d(x, \gamma_1) - d(x, z_n). \tag{4.9}$$

Hence, using first (4.9), then (4.8) and eventually (4.7), we have

$$\begin{aligned} 0 &\leq d(y_n, x) + d(x, z_n) - d(y_n, z_n) \leq d(y_n, x) + d(x, \gamma_1) - d(y_n, \gamma_1) \\ &= d(y_n, x) + \frac{d^2(x, \gamma_1) - d^2(y_n, \gamma_1)}{d(x, \gamma_1) + d(y_n, \gamma_1)} \\ &= d(y_n, x) + \frac{1}{2} \frac{d^2(x, \gamma_1) - d^2(y_n, \gamma_1)}{d(y_n, x)} \frac{2d(y_n, x)}{d(x, \gamma_1) + d(y_n, \gamma_1)} \\ &\leq d(y_n, x) - \frac{\varphi(y_n) - \varphi(x)}{d(y_n, x)} \frac{2d(y_n, x)}{d(x, \gamma_1) + d(y_n, \gamma_1)} \\ &< \left(1 - \frac{2(d(x, \gamma_1) - \frac{1}{n})}{d(x, \gamma_1) + d(y_n, \gamma_1)}\right) d(y_n, x) = o(r_n). \end{aligned}$$

By taking a subsequence we find points $y, z \in Y$ so that

$$i_n(y_n) \rightarrow i(y) \quad \text{and} \quad i_n(z_n) \rightarrow i(z).$$

In particular

$$d_Y(y, z) - d_Y(y, 0) - d_Y(0, z) = \lim_{n \rightarrow \infty} \frac{d(y_n, z_n) - d(y_n, x) - d(x, z_n)}{r_n} = 0$$

and so 0 lies on some constant speed geodesic η in Y joining y to z (obtained by the concatenation of the geodesics joining y to 0 and 0 to z).

With a similar argument, we can show that 0 lies on some constant speed geodesic $\tilde{\eta}$ in Y joining y to \tilde{z} , where \tilde{z} is obtained as the limit $i_n(\tilde{z}_n) \rightarrow i(\tilde{z})$ of the points \tilde{z}_n , which are taken from the geodesic $\tilde{\gamma}$ so that $d(\tilde{z}_n, x) = r_n$. Note that we might have to go to yet another subsequence to achieve the convergence to \tilde{z} .

Because the space X is strongly non-branching, we have

$$d_Y(z, \tilde{z}) \geq \liminf_{n \rightarrow \infty} \frac{d(z_n, \tilde{z}_n)}{r_n} > 0$$

and so the geodesics η and $\tilde{\eta}$ contradict the assumption that Y is non-branching.

This means that our assumptions on the geodesics γ and $\tilde{\gamma}$ can not be satisfied. Therefore, there exists a set $A \subset \text{Geo}(X)$ so that $\pi(\text{Geo}(X) \setminus A) = 0$, and for every $x \in X$, there is at most one $\gamma \in A$ with $x = \gamma_0$. Using the set A we can define the transport map T as

$$T(x) = \begin{cases} \gamma_1, & \text{when } x = \gamma_0 \text{ for some } \gamma \in A \\ x, & \text{otherwise.} \end{cases}$$

□

Remark 4.4 One could prove Theorem 4.3 also under slightly different assumptions. Namely by weakening the definition of strongly non-branching metric space by replacing the liminf in (2.1) to limsup, and then assuming that at almost every point all the tangent spaces to X are non-branching. This modified theorem is achieved by letting the sequence of radii in the blow-up be dictated by the weakened form of strong non-branching property, namely choosing r_n in such a way that $d(x, z_n) = d(x, \tilde{z}_n) = r_n$ and $\lim_n d(z_n, \tilde{z}_n)/r_n > 0$.

Theorem 4.3 applies, for example, when the space (X, d) is a finite dimensional Alexandrov space and m is the corresponding volume measure on X . The estimate (4.2) on the relative entropy follows in this case from the result of Petrunin [15], which shows that in Alexandrov spaces, the functional

$$\mu \mapsto \int_X \rho^{1-\frac{1}{N}} dm \tag{4.10}$$

is concave along Wasserstein geodesics. Notice that a different proof for the Brenier theorem in Alexandrov spaces was already given by Bertrand in [5].

Brenier theorem has been recently established by Gigli [8] in non-branching spaces with Ricci curvature bounded from below. This generalizes the previous result by Bertrand, and it covers, for example, the case where the functional (4.10) is concave along geodesics in the Wasserstein space $(\mathcal{P}(X), W_2)$ of a non-branching space (X, d) . Whereas our proof of Theorem 4.3 is based on the behavior of blow-ups and the Kantorovich potential, the proof by Gigli relies on the concavity of the functional and does not use the Kantorovich potential at all. Notice that because of the different techniques used in the proofs our geometric assumptions on the metric space X differ from those assumed by Gigli, and hence, the two theorems cover different collection of metric spaces.

It is also important to notice that our Theorem 4.3 by no means covers all the cases where the Brenier theorem is known to hold. For example, the Brenier theorem holds in the Heisenberg group [4], but it is not difficult to see that the Heisenberg group is not strongly non-branching.

We end this paper with an improvement of [1, Theorem 10.4.] in the case where the reference measure m is doubling. In [1], it was shown that without the assumption that m is doubling, we have the conclusion

$$\lim_{t \downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{d(\gamma_0, \gamma_t)} = d(\gamma_0, \gamma_1) \quad \text{in } L^2(\text{Geo}(X), \pi)$$

in the following theorem. Because our Proposition 4.2 was proved in the case where the space X is bounded, we will make the same boundedness assumption here. As in many of the results in [1], we could remove this assumption by requiring the density of the initial measure μ with respect to m to be uniformly bounded away from zero.

Theorem 4.5 *Let m be a doubling measure on a bounded metric space X and $\mu = \rho m \in \mathcal{P}(X)$, with $\rho > 0$ m -a.e. in X and $\nu \in \mathcal{P}(X)$. Let $\pi \in \text{GeoOpt}(\mu, \nu)$ and $\varphi: X \rightarrow \mathbb{R} \cup \{-\infty\}$ be a Kantorovich potential relative to π satisfying*

$$|\nabla_g^+ \varphi|(\gamma_0) = d(\gamma_0, \gamma_1) \quad \pi \text{-a.e. in } \text{Geo}(X). \tag{4.11}$$

Further assume that there exists $\bar{s} \in (0, 1]$ such that for all $s \in [0, \bar{s})$, we have $(e_s)_{\#}\pi \ll m$. Then for π -a.e. $\gamma \in \text{Geo}(X)$, we have

$$\lim_{t \downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{d(\gamma_0, \gamma_t)} = d(\gamma_0, \gamma_1).$$

Proof If the lower ascending slope along geodesics of the Kantorovich potential were continuous, the theorem would follow immediately from the fact that the lower ascending slope is an upper gradient. This is not true in general, but what we can prove using density points and cyclical monotonicity is that for π -almost every geodesic the lower ascending slope is continuous along the geodesic at its starting point.

As we have seen in the proof of Proposition 4.2, in (4.5), the inequality

$$\liminf_{t \downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{d(\gamma_0, \gamma_t)} \geq d(\gamma_0, \gamma_1)$$

holds for π -a.e. $\gamma \in \text{Geo}(X)$.

On the other hand, by Proposition 3.4, we know that $|\nabla_g^+ \varphi|$ is an upper gradient of φ along geodesics and so for all $\gamma \in \text{Geo}(X)$ the estimate

$$\frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{d(\gamma_0, \gamma_t)} \leq \frac{1}{d(\gamma_0, \gamma_t)} \int_{\gamma|_{[0,t]}} |\nabla_g^+ \varphi|$$

holds for all $t \in (0, 1)$. So, our claim follows if we can show with any $\delta > 0$ that for π -a.e. $\gamma \in \text{Geo}(X)$

$$|\nabla_g \varphi^+|(\gamma_s) \leq (1 + \delta)d(\gamma_0, \gamma_1) \quad \text{for } \mathcal{L}^1 \text{-a.e. } s \in (0, t) \tag{4.12}$$

when $t > 0$ (depending on δ and γ) is small enough.

Because $\rho > 0$, we know from (4.11) that for m -a.e. $x \in X$ there exists $\gamma^x \in \text{Geo}(X)$ with $\gamma_0^x = x$ and $|\nabla_g^+ \varphi|(x) = d(\gamma_0^x, \gamma_1^x)$. When we combine this with the assumption

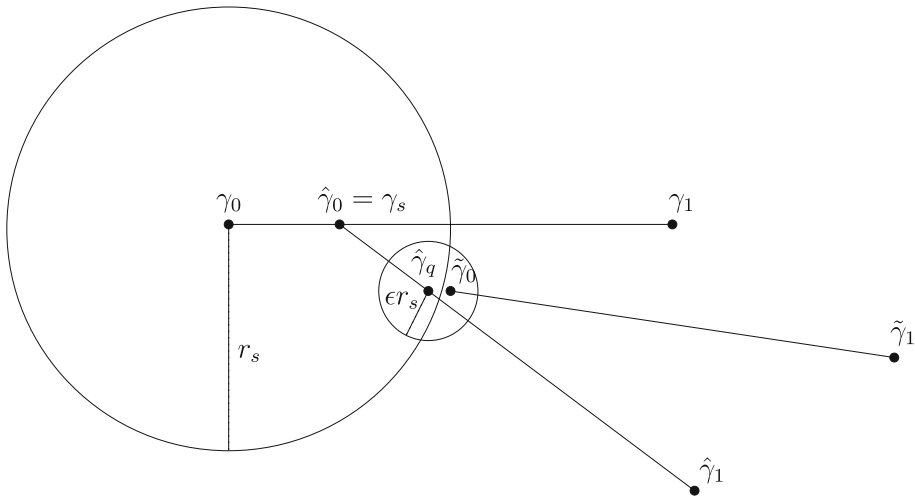


Fig. 2 The three curves γ , $\hat{\gamma}$ and $\tilde{\gamma}$, which are used in the proof

$(e_s)_\# \pi \ll m$, for any $s \in (0, \bar{s})$, we get for π -a.e. $\gamma \in \text{Geo}(X)$ a curve $\hat{\gamma} \in \text{Geo}(X)$ so that $\gamma_s = \hat{\gamma}_0$ and

$$|\nabla_g^+ \varphi|(\hat{\gamma}_0) = d(\hat{\gamma}_0, \hat{\gamma}_1).$$

Hence, by Fubini’s theorem we know that for π -a.e. $\gamma \in \text{Geo}(X)$, a curve $\hat{\gamma} \in \text{Geo}(X)$ with the above properties exists for \mathcal{L}^1 -a.e. $s \in (0, \bar{s})$.

Our task is now to estimate $d(\hat{\gamma}_0, \hat{\gamma}_1)$ from above. Because m is doubling, it is enough to prove this for $\gamma \in \text{Geo}(X)$ for which γ_0 is a Lebesgue point of $|\nabla_g^+ \varphi|$. Let $\epsilon > 0$ and take $t \in (0, 1)$ so small that for every $0 < s < t$ and $x \in B(\gamma_0, r_s)$, where $r_s = 2sd(\gamma_0, \gamma_1)$, there exists $\tilde{\gamma} \in \text{Geo}(X)$ with $\tilde{\gamma}_0 \in B(x, \epsilon r_s)$ and

$$d(\tilde{\gamma}_0, \tilde{\gamma}_1) = |\nabla_g^+ \varphi|(\tilde{\gamma}_0) < (1 + \epsilon)|\nabla_g^+ \varphi|(\gamma_0) = (1 + \epsilon)d(\gamma_0, \gamma_1). \tag{4.13}$$

Define $q = sd(\gamma_0, \gamma_1)/d(\hat{\gamma}_0, \hat{\gamma}_1)$ and let $x = \hat{\gamma}_q \in B(\gamma_0, r_s)$. With this choice of x let $\tilde{\gamma} \in \text{Geo}(X)$ be as above. Notice that we may assume $q < s$, as otherwise the upper bound on $d(\hat{\gamma}_0, \hat{\gamma}_1)$ immediately follows. The selected curves are illustrated in Fig. 2.

Now we are ready to estimate $d(\hat{\gamma}_0, \hat{\gamma}_1)$ from above. For this, we use cyclical monotonicity:

$$\begin{aligned} d^2(\hat{\gamma}_0, \hat{\gamma}_1) + d^2(\tilde{\gamma}_0, \tilde{\gamma}_1) &\leq d^2(\tilde{\gamma}_0, \hat{\gamma}_1) + d^2(\hat{\gamma}_0, \tilde{\gamma}_1) \\ &\leq (d(\tilde{\gamma}_0, \hat{\gamma}_q) + d(\hat{\gamma}_q, \tilde{\gamma}_1))^2 + (d(\hat{\gamma}_0, \hat{\gamma}_q) + d(\hat{\gamma}_q, \tilde{\gamma}_0) + d(\tilde{\gamma}_0, \tilde{\gamma}_1))^2 \\ &= d^2(\tilde{\gamma}_0, \hat{\gamma}_q) + 2d(\tilde{\gamma}_0, \hat{\gamma}_q)d(\hat{\gamma}_q, \tilde{\gamma}_1) + d^2(\hat{\gamma}_q, \tilde{\gamma}_1) + d^2(\hat{\gamma}_0, \hat{\gamma}_q) + 2d(\hat{\gamma}_0, \hat{\gamma}_q)d(\hat{\gamma}_q, \tilde{\gamma}_0) \\ &\quad + 2d(\hat{\gamma}_0, \hat{\gamma}_q)d(\tilde{\gamma}_0, \tilde{\gamma}_1) + d^2(\hat{\gamma}_q, \tilde{\gamma}_0) + 2d(\hat{\gamma}_q, \tilde{\gamma}_0)d(\tilde{\gamma}_0, \tilde{\gamma}_1) + d^2(\tilde{\gamma}_0, \tilde{\gamma}_1). \end{aligned}$$

Now, using the inequalities $d(\tilde{\gamma}_0, \hat{\gamma}_q) < \epsilon r_s$ and $d(\gamma_0, \hat{\gamma}_q) < r_s$, we get that $d^2(\hat{\gamma}_0, \hat{\gamma}_1) + d^2(\tilde{\gamma}_0, \tilde{\gamma}_1)$ is bounded from above by

$$\begin{aligned} &\epsilon^2 r_s^2 + 2\epsilon r_s(1 - q)d(\hat{\gamma}_0, \hat{\gamma}_1) + (1 - q)^2 d^2(\hat{\gamma}_0, \hat{\gamma}_1) + q^2 d^2(\hat{\gamma}_0, \hat{\gamma}_1) \\ &+ 2q\epsilon r_s d(\hat{\gamma}_0, \hat{\gamma}_1) + 2qd(\hat{\gamma}_0, \hat{\gamma}_1)d(\tilde{\gamma}_0, \tilde{\gamma}_1) + \epsilon^2 r_s^2 + 2\epsilon r_s d(\tilde{\gamma}_0, \tilde{\gamma}_1) + d^2(\tilde{\gamma}_0, \tilde{\gamma}_1) \\ &= 2\epsilon r_s (\epsilon r_s + d(\hat{\gamma}_0, \hat{\gamma}_1) + d(\tilde{\gamma}_0, \tilde{\gamma}_1)) + d^2(\hat{\gamma}_0, \hat{\gamma}_1) + d^2(\tilde{\gamma}_0, \tilde{\gamma}_1) \\ &\quad + 2qd(\hat{\gamma}_0, \hat{\gamma}_1)d(\tilde{\gamma}_0, \tilde{\gamma}_1) + 2(q - 1)qd^2(\hat{\gamma}_0, \hat{\gamma}_1) \end{aligned}$$

$$= 2\epsilon r_s (\epsilon r_s + d(\hat{\gamma}_0, \hat{\gamma}_1) + d(\tilde{\gamma}_0, \tilde{\gamma}_1)) + d^2(\hat{\gamma}_0, \hat{\gamma}_1) + d^2(\tilde{\gamma}_0, \tilde{\gamma}_1) \\ + r_s d(\tilde{\gamma}_0, \tilde{\gamma}_1) + (q-1)r_s d(\hat{\gamma}_0, \hat{\gamma}_1).$$

It follows that $2\epsilon r_s (\epsilon r_s + d(\hat{\gamma}_0, \hat{\gamma}_1) + d(\tilde{\gamma}_0, \tilde{\gamma}_1)) + r_s d(\tilde{\gamma}_0, \tilde{\gamma}_1) + (q-1)r_s d(\hat{\gamma}_0, \hat{\gamma}_1) \geq 0$, so that dividing by r_s and using (4.13) yields

$$d(\hat{\gamma}_0, \hat{\gamma}_1) \leq \frac{(1+2\epsilon)d(\tilde{\gamma}_0, \tilde{\gamma}_1) + 2\epsilon^2 r_s}{1-q-2\epsilon} \leq \frac{(1+2\epsilon)(1+\epsilon) + 4\epsilon^2 s}{1-s-2\epsilon} d(\gamma_0, \gamma_1).$$

Choosing s and ϵ small enough, depending on δ , we achieve (4.12) and conclude the proof. \square

Acknowledgments The authors acknowledge the support of the ERC ADG GeMeThNES. The second author also acknowledges the support of the Academy of Finland, project no. 137528.

References

1. Ambrosio, L., Gigli, N., Savaré, G.: Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. (2011) [Preprint]
2. Ambrosio, L., Gigli, N., Savaré, G.: Metric measure spaces with Riemannian Ricci curvature bounded from below. (2011) [Preprint]
3. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser Verlag, Basel (2008)
4. Ambrosio, L., Rigot, S.: Optimal mass transportation in the Heisenberg group. *J. Funct. Anal.* **208**(2), 261–301 (2004)
5. Bertrand, J.: Existence and uniqueness of optimal maps on Alexandrov spaces. *Adv. Math.* **219**(3), 838–851 (2008)
6. Brenier, Y.: Polar factorization and monotone rearrangement of vector valued functions. *Commun. Pure Appl. Math.* **64**, 375–417 (1991)
7. Cheeger, J.: Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.* **9**, 428–517 (1999)
8. Gigli, N.: Optimal maps in non branching spaces with Ricci curvature bounded from below. *Geom. Funct. Anal.* (to appear)
9. Kantorovich, L.V.: On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37**, 199–201 (1942)
10. Kantorovich, L.V.: On a problem of Monge (in Russian). *Uspekhi Mat. Nauk.* **3**, 225–226 (1948)
11. Käenmäki, A., Rajala, T., Suomala, V.: Existence of doubling measures via generalised nested cubes. *Proc. Am. Math. Soc.* (to appear)
12. Luukkainen, J., Saksman, E.: Every complete doubling metric space carries a doubling measure. *Proc. Am. Math. Soc.* **126**(2), 531–534 (1998)
13. McCann, R.: Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.* **11**, 589–608 (2001)
14. Monge, G.: Mémoire sur la théorie des déblais et remblais. In: *Histoire de l’Académie Royale des Sciences de Paris*, pp 666–704 (1781)
15. Petrunin, A.: Alexandrov meets Lott-Villani-Sturm. *Münster J. Math.* **4**, 53–64 (2011)
16. Smith, C., Knott, M.: On the optimal transportation of distributions. *J. Optim. Theory Appl.* **52**, 323–329 (1987)
17. Villani, C.: Optimal transport. Old and new, vol. 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Berlin (2009)