Geometry & Topology 23 (2019) 3111–3139

Metric-minimizing surfaces revisited

ANTON PETRUNIN
STEPHAN STADLER

A surface that does not admit a length nonincreasing deformation is called *metric-minimizing*. We show that metric-minimizing surfaces in CAT(0) spaces are locally CAT(0) with respect to their length metrics.

53C23, 53C43, 53C45; 30L05

1 Introduction

Main result Assume s is a Lipschitz embedding of the disc $\mathbb{D}$ into the Euclidean space $\mathbb{R}^3$. We say s is *metric-minimizing* if its intrinsic metric is minimal. Explicitly, this means that if a map $s'\colon \mathbb{D} \to \mathbb{R}^3$ agrees with s on $\partial\mathbb{D} = \mathbb{S}^1$ and fulfills

$$\operatorname{length} s \circ \gamma \geqslant \operatorname{length} s' \circ \gamma$$

for all curves γ in $\mathbb{D}$, then equality holds for all curves γ.

As follows from the main theorem below (Theorem 1.1), the induced metric on the disc for a metric-minimizing embedding is $\mathrm{CAT}(0)$.

To formulate the theorem in full generality we need to extend the definition of metric-minimizing maps in two ways. First, we will only assume continuity of s, so the map s might not be an embedding and not necessary Lipschitz; this part is tricky — a straightforward generalization produces a too weak condition. Secondly, we also have to define it with arbitrary metric spaces as targets; this part is straightforward.

Let Y be a metric space and $s\colon \mathbb{D} \to Y$ be a continuous map. Let $\gamma\colon \mathbb{S}^1 \to Y$ be a closed rectifiable curve. We say that $s\colon \mathbb{D} \to Y$ *spans* γ if s is an extension of γ; that is, $s|_{\mathbb{S}^1} = \gamma$.

Consider the induced length pseudometric on $\mathbb{D}$ defined as

$$\langle x - y\rangle_s = \inf_{\alpha}\{\operatorname{length} s \circ \alpha\},$$

Published: 1 December 2019 DOI: 10.2140/gt.2019.23.3111

where the greatest lower bound is taken over all paths α from x to y in $\mathbb{D}$. The distance $\langle x-y\rangle_s$ might take infinite values. We denote by $\langle\mathbb{D}\rangle_s$ the corresponding metric space; see the next section for a precise definition.

The space $\langle\mathbb{D}\rangle_s$ comes with the projections $\mathbb{D}\xrightarrow{\hat\pi_s}\langle\mathbb{D}\rangle_s\xrightarrow{\hat s}Y$; the restriction $\hat\pi_s|_{\mathbb{S}^1}$ will be denoted by δ_s.

Assume that s and s' are two maps spanning the same curve. We write $s\succcurlyeq s'$ if there is a *majorization* $\mu\colon\langle\mathbb{D}\rangle_s\to\langle\mathbb{D}\rangle_{s'}$, meaning that μ is a *short* map and $\delta_{s'}=\mu\circ\delta_s$; here and below *short* stands for *distance-nonincreasing*.

(Note that if s and s' are Lipschitz embeddings, then any majorization $\langle\mathbb{D}\rangle_s\to\langle\mathbb{D}\rangle_{s'}$ admits a continuous lifting $\mathbb{D}\to\mathbb{D}$; in general such a lifting may not exist.)

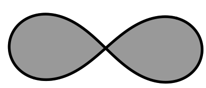
A map $s\colon\mathbb{D}\to Y$ will be called a *metric-minimizing disc* if $s\succcurlyeq s'$ implies that the corresponding majorization μ is an isometry.

A topological space W together with a choice of a closed curve $\delta\colon\mathbb{S}^1\to W$ is called a *disc retract* if the mapping cylinder of δ

$$W_\delta=W\sqcup_{\delta(u)\sim(u,0)}\mathbb{S}^1\times[0,1]$$

is homeomorphic to the disc $\mathbb{D}$. The curve δ will be called the *boundary curve* of the disc retract W.

An example of a disc retract that is not a disc is shown:



Note that it has essentially different boundary curves, in the sense that one is not a reparametrization of the other. The definition is motivated by the following observation: if the boundary curve of a disc retract W is a simple closed curve, then W is homeomorphic to $\mathbb{D}$.

1.1 Main theorem *Assume Y is a* CAT(0) *space and $s\colon\mathbb{D}\to Y$ is a metric-minimizing disc. Then $\langle\mathbb{D}\rangle_s$ is a* CAT(0) *disc retract with boundary curve δ_s.*

In particular, if s spans a simple closed curve, then $\langle\mathbb{D}\rangle_s$ is a CAT(0) *disc.*

Note that apart from continuity we did not make any regularity assumptions on the map s; in particular, beforehand, the space $\langle\mathbb{D}\rangle_s$ might have wild topology.

If we remove the condition that the boundary curve is rectifiable, then the space $\langle\mathbb{D}\rangle_s$ might have points at infinite distance from each other. However, our proof shows that all triangles with finite sides in $\langle\mathbb{D}\rangle_s$ are still thin; in particular, each metric component of $\langle\mathbb{D}\rangle_s$ is a $\mathrm{CAT}(0)$ space. So, in this case, one could consider $\langle\mathbb{D}\rangle_s$ as a $\mathrm{CAT}(0)$ space where infinite distances between points are legal.

The class of metric-minimizing discs is huge — if the ambient space Y is $\mathrm{CAT}(0)$, then one can find a metric-minimizing disc $\succcurlyeq$*–below* for any map $s\colon \mathbb{D}\to Y$ such that the metric $(x,y)\mapsto\langle x,y\rangle_s$ is continuous (this can be proved by applying the ultralimit/projection construction described below). Metric-minimizing discs include many well-studied maps from $\mathbb{D}$ to metric spaces. The ruled discs considered by Alexandr Alexandrov [1] are evidently metric-minimizing since one cannot shorten a geodesic. In addition, harmonic maps in the sense of Korevaar and Schoen [10] from the disc to $\mathrm{CAT}(0)$ spaces are metric-minimizing. Indeed, harmonic discs are solutions to the Dirichlet problem, that is, energy-minimizing fillings of given loops. Since the Korevaar–Schoen energy is convex, such solutions are unique. But decreasing the intrinsic metric will only decrease the energy of the corresponding map. As a consequence, minimal discs in the sense of Lytchak and Wenger [15] are metric-minimizing; these are Douglas–Rado solutions to the Plateau problem and therefore harmonic.

As intended, the main theorem, Theorem 1.1, subsumes and generalizes several previously known results, including Alexandrov's theorem about ruled surfaces in [1] and the main theorem by the first author announced in [20], assuming it is formulated correctly (see [22]); see Section 9. It is closely related to the classical Gauss formula and results on saddle surfaces by Samuel Shefel in [25; 24]. It also generalizes the result on minimal surfaces by Alexander Lytchak and Stefan Wenger [17, Theorem 1.2] and an earlier result of Chikako Mese [18]; see also Alexander, Kapovitch and Petrunin [3, Chapter 4] and Petrunin and Stadler [23]. Despite that some special cases of Theorem 1.1 were known, the result is new even for harmonic discs.

Let us list a few applications of metric-minimizing surfaces. They were used by Lytchak [13] to study sets of positive reach (ruled surfaces). In [2], Stephanie Alexander and Richard Bishop used them to generalize the Gauss equation to nonsmooth spaces (ruled surfaces). In [8], Mikhael Gromov used them to bound the complexity of smooth maps (general metric-minimizing surfaces). In [14], Lytchak and the second author used them to deform general $\mathrm{CAT}(0)$ spaces (minimal discs). In [26], the second author used them in the proof a $\mathrm{CAT}(0)$ version of the Fary–Milnor theorem to control the mapping behavior of minimal surfaces (minimal discs).

In general it is hard to check whether a given surface is metric-minimizing. Note that any smooth metric-minimizing disc in a Euclidean space is saddle. The converse fails in general (see Section 10), but the following statement gives a local converse in dimension three. We expect that in the four-dimensional case even the local converse does not hold.

1.2 Proposition *Any smooth strictly saddle surface in* $\mathbb{R}^3$ *is locally metric-minimizing.*

Structure of the paper In Section 2 we define a couple of metrics induced by a given continuous map. This naturally leads to the monotone–light factorization theorem in a metric context.

In Section 3 we obtain topological control on the length space associated to a metric-minimizing disc. This part follows from Moore's quotient theorem on cell-like maps.

Section 4 establishes compactness of a certain class of nonpositively curved surfaces. This will be used later, when we approximate metric-minimizing maps.

In Section 6 we prove the key lemma, which says that the restriction to a finite set of a metric-minimizing disc factorizes as a composition of short maps over a nonpositively curved surface. The proof uses the properties of *metric-minimizing graphs* discussed in Section 5.

In Section 7 we prove an extension result which is of independent interest and logically detached from the rest of the paper. It gives a criterion that allows us to extend maps from subsets to the whole space, once extensions on finite subsets are guaranteed.

In Section 8 we assemble the proof of the main theorem. We show that a metric-minimizing disc factorizes as a composition of short maps over a nonpositively curved surface. The proof is finished by showing that the first map of this factorization induces an isometry.

In Section 9, we give corrected formulations of the theorem in the old paper and show that they follow from the main theorem.

In the last section, we discuss the relation between saddle surfaces and metric-minimizing discs.

Acknowledgements We want to thank Sergei Ivanov, Nina Lebedeva, Carlo Sinestrari, Peter Topping and Burkhard Wilking for help. We also would like to thank Alexander Lytchak for explaining his recent work with Stefan Wenger and for several helpful discussions.

Petrunin was partially supported by NSF grant DMS 1309340. Stadler was supported by DFG grants STA 1511/1-1 and SPP 2026.

2 Definitions

Metrics and pseudometrics Let X be a set. A *pseudometric* on X is a function $X \times X \to [0,\infty]$ denoted as $(x,y) \mapsto |x-y|$ such that

- $|x-x| = 0$, for any $x \in X$;
- $|x-y| = |y-x|$, for any $x, y \in X$;
- $|x-y| + |y-z| \geqslant |x-z|$ for any $x, y, z \in X$.

If in addition $|x-y| = 0$ implies $x = y$, then the pseudometric $|*-*|$ is called a *metric*; some authors prefer to call it an ∞*–metric* to emphasize that the distance between points might be infinite. The value $|x-y|$ will also be called the *distance* from x to y.

A (pseudo)metric space X is the underlying set, which is also denoted by X, equipped with a (pseudo)metric, which often will be denoted by $|*-*|_X$. We will use X as an index if we want to emphasize that we are working in the space X; for example, the ball of radius R centered at z in X can be denoted as

$$B(z,R)_X = \{x \in X : |z-x|_X < R\}.$$

For any pseudometric on a set there is an equivalence relation "$\sim$" such that

$$x \sim y \iff |x-y| = 0.$$

The pseudometric induced on the set of equivalence classes

$$[x] = \{x' \in X : x' \sim x\}$$

becomes a metric. The obtained metric space will be denoted as $[X]$; it comes with the projection map $X \to [X]$ defined as $x \mapsto [x]$.

For a metric space we can consider the equivalence relation "$\approx$" defined as

$$x \approx y \iff |x-y| < \infty.$$

Its equivalence classes are called *metric components*. Note that by definition each metric component is a *genuine metric space*, meaning that distances between points are finite. Consequently, any metric space is a disjoint union of genuine metric spaces.

Pseudometrics induced by a map Assume X is a topological space and Y is a metric space. Let $f\colon X \to Y$ be a continuous map.

Let us define the *length pseudometric* on X induced by f as

$$\langle x-y\rangle_f = \inf\{\operatorname{length}(f\circ\gamma)_Y : \gamma \text{ a path in } X \text{ from } x \text{ to } y\}.$$

Denote by $\langle X\rangle_f$ the corresponding metric space; that is,

$$\langle X\rangle_f = [(X, \langle *-*\rangle_f)].$$

Similarly, define a *connecting pseudometric* $|*-*|_f$ on X by

$$|x-y|_f = \inf\{\operatorname{diam} f(K)\},$$

where the greatest lower bound is taken over all connected sets $K \subset X$ that contain x and y; if there is no such set, we set $|x-y|_f = \infty$. The associated metric space will be denoted as $|X|_f$; that is,

$$|X|_f = [(X, |*-*|_f)].$$

For the projections

$$\bar\pi_f\colon X \to |X|_f \quad\text{and}\quad \hat\pi_f\colon X \to \langle X\rangle_f$$

we will also use the shortcut notations

$$\bar x = \bar\pi_f(x) \quad\text{and}\quad \hat x = \hat\pi_f(x).$$

2.1 Lemma *Let X be a locally connected topological space and Y a metric space. Assume that $f\colon X \to Y$ is a continuous map. Then $\bar\pi_f\colon X \to |X|_f$ is continuous.*

In particular, if in addition X is compact, then so is $|X|_f$.

Proof For a point $x \in X$ and $\varepsilon > 0$ we denote by U the connected component of $f^{-1}[B(f(x),\varepsilon)_Y]$ that contains x. Since X is locally connected, the set U is open.

Note that $\bar\pi_f(U) \subset B(\bar x, 2\cdot\varepsilon)_{|X|_f}$, whence the result. □

Note that $\tau_f\colon \hat{x} \to \bar{x}$ defines a map $\tau_f\colon \langle X\rangle_f \to |X|_f$ and by construction it preserves the lengths of all curves coming from X. Since $\langle X\rangle_f$ is a length space, the latter implies that τ_f is short. The map τ_f might not induce an isometry

$$\langle X\rangle_f \to \langle |X|_f\rangle_{\bar{f}}.$$

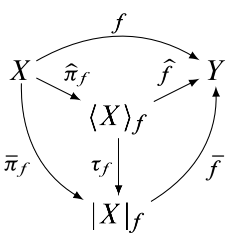
Moreover, τ_f does not have to be injective, an example is given in [21, Example 4.2]. However, for metric-minimizing discs $f\colon \mathbb{D} \to Y$ both statements hold true; see Proposition 3.1.

The space $\langle |X|_f\rangle_{\bar{f}}$ is the intrinsic metric on $|X|_f$, which is $\langle |X|_f\rangle$ in the notation defined below. We will denote it briefly by $\langle |X|\rangle_f$. The corresponding pseudometric will be called the *intrinsic pseudometric* on X induced by f; it will be denoted by $\langle |{*}-{*}|\rangle_f$. This is a more natural way to pull back intrinsic metric to X. If X is compact, then $\langle |{*}-{*}|\rangle_f$ coincides with the pseudometric pull_f defined in [21]. It will show up in Section 9.

The maps $\bar{f}\colon |X|_f \to Y$ and $\hat{f}\colon \langle X\rangle_f \to Y$ are uniquely defined by the identity

$$f(x) = \bar{f}(\bar{x}) = \hat{f}(\hat{x})$$

for any $x \in X$. By construction, the diagram



commutes.

Moreover, if X is compact, then $\bar{\pi}_f$ has connected fibers; see Lemma 2.4.

2.2 Lemma *Let X be a compact metric space. Then $\bar{f}\colon |X|_f \to Y$ preserves the length of every curve.*

Proof Since $\bar{f}$ is short, we have $\mathrm{length}(\bar{f}\circ\gamma) \le \mathrm{length}(\gamma)$ for every curve γ in $|X|_f$. Let a rectifiable curve γ be given and let $\varepsilon > 0$. Choose points x_i on γ such that $\mathrm{length}(\gamma) \le \sum_i |x_i, x_{i+1}|_{|X|_f} + \varepsilon$. Denote by γ_i the piece of γ between x_i and x_{i+1}. Since $\bar{\pi}_f$ has connected fibers, each set $\bar{\pi}_f^{-1}(\gamma_i)$ is connected. Hence

$$\mathrm{length}(\gamma) \le \sum_i \mathrm{diam}\, f(\bar{\pi}_f^{-1}(\gamma_i)) + \varepsilon = \sum_i \mathrm{diam}(\bar{f}(\gamma_i)) + \varepsilon \le \mathrm{length}(\bar{f}\circ\gamma) + \varepsilon.$$

The claim follows since ε was arbitrary. □

Metrics induced by metrics If X is a metric space, the two constructions above can be applied to the identity map id: $X \to X$. In this case the obtained spaces $\langle X\rangle_{\mathrm{id}}$ and $|X|_{\mathrm{id}}$ will be denoted by $\langle X\rangle$ and $|X|$, respectively. The space $\langle X\rangle$ is X equipped with induced length metric. All three spaces $\langle X\rangle$, $|X|$ and X have the same underlying set; in other words they can be considered as a single space with different metrics and tautological maps between them. Both tautological maps

$$\langle X\rangle \to |X| \to X$$

are short and length-preserving.

As a consequence of Lemma 2.2, the tautological map $\langle |X|_f\rangle_{\bar f} \to \langle |X|_f\rangle$ is an isometry; that is, for any continuous map $f\colon X \to Y$ the induced length metric on $|X|_f$ coincides with the length metric induced by $\bar f\colon |X|_f \to Y$.

Recall that a geodesic in a metric space is a curve whose length coincides with the distance between its endpoints. A metric space is called *geodesic* if any two points at finite distance can be joined by a geodesic.

2.3 Lemma *Let X be a compact metric space. Then $\langle X\rangle$ is a complete geodesic space.*

The second assertion is classical (see for example [9, Section II-8, Theorem 3]) but we were not able to find the first one in the literature.

Proof Assume that $\langle X\rangle$ is not complete. Fix a Cauchy sequence (x_n) in $\langle X\rangle$ that is not converging in $\langle X\rangle$. After passing to a subsequence, we can assume that the points of the sequence appear on a rectifiable curve $\hat\gamma\colon [0,1) \to \langle X\rangle$ in the same order.

The corresponding curve $\gamma\colon [0,1) \to X$ has the same length. Since X is compact we can extend it to a path $\gamma_+\colon [0,1] \to X$. The curve

$$\hat\gamma_+ = \hat\pi \circ \gamma_+\colon [0,1] \to \langle X\rangle$$

has the same length. Therefore $\hat\gamma_+(1)$ is the limit of (x_n), a contradiction.

It remains to show that $\langle X\rangle$ is geodesic. Assume γ_n is a sequence of constant speed paths from x to y in X such that $\operatorname{length}(\hat\gamma_n) \to \langle x - y\rangle$ as $n \to \infty$. Since X is compact, we can pass to a partial limit γ of γ_n. The corresponding curve $\hat\gamma = \hat\pi \circ \gamma$ is the needed geodesic from $\hat x$ to $\hat y$ in $\langle X\rangle$. □

Monotone–light factorization Let $f\colon X\to Y$ be a map between topological spaces. Recall that

- f is called *monotone* if the inverse image of each point is connected,
- f is called *light* if the inverse image of any point is totally disconnected.

Since a connected set is nonempty by definition, any monotone map is onto.

2.4 Lemma *Assume X is a locally connected compact metric space and Y is a metric space. Let $f\colon X\to Y$ be a continuous map. Then the map $\bar\pi_f$ is monotone and $\bar f$ is light. In particular,*

$$f=\bar f\circ\bar\pi_f$$

is a monotone–light factorization.

Proof First we prove the monotonicity of $\bar\pi_f$.

Assume the contrary; that is, for some $x\in X$ the equivalence class

$$K=\bar\pi_f^{-1}(\bar x)=\{x'\in X:|x-x'|_f=0\}$$

is not connected. Since X is normal, we can cover K by disjoint open sets $U,V\subset X$ such that both intersections $K\cap U$ and $K\cap V$ are nonempty.

By Lemma 2.1, K is closed.

Suppose that $x\in U$ and pick $x'\in K\cap V$. Then there is a sequence of connected sets $K_n\ni x,x'$ such that $\operatorname{diam} f(K_n)<\frac{1}{n}$. For each n we choose a point $k_n\in K_n\setminus(U\cup V)$. Let k be a partial limit of the sequence (k_n). It follows that $k\in K\setminus(U\cup V)$, a contradiction.

Assume $\bar f$ is not light; that is, the inverse image of some $y\in Y$ contains a closed connected set $C\subset|X|_f$ with more than one point. Note that the inverse image $Z:=\bar\pi_f^{-1}C$ is connected due to monotonicity of $\bar\pi_f$. It follows that $|z-z'|_f=0$ for any two points $z,z'\in Z$, a contradiction. □

3 Disc retracts

A disc retract as defined above is nothing but the image of strong deformation retraction of $\mathbb{D}$; the restriction of retraction to the boundary can be taken as the corresponding boundary curve. We will not need this statement, but it follows from Moore's quotient theorem quoted below.

3.1 Proposition *Let Y be a metric space and $s\colon \mathbb{D}\to Y$ be a metric-minimizing disc. Then $|\mathbb{D}|_s$ is a disc retract with boundary curve $\delta_s=\bar\pi_s|_{\mathbb{S}^1}$. Moreover, the map $\tau_s\colon \langle\mathbb{D}\rangle_s\to|\mathbb{D}|_s$ is injective and defines an isometry $\langle\mathbb{D}\rangle_s\to\langle|\mathbb{D}|\rangle_s$; that is, τ_s is an isometry from $\langle\mathbb{D}\rangle_s$ to $|\mathbb{D}|_s$ equipped with induced length metric.*

We need a little preparation before giving the proof.

Let Y be a metric space and $s\colon \mathbb{D}\to Y$ be a continuous map. We say that s has *no bubbles* if for any point $p\in Y$ every connected component of the complement $\mathbb{D}\setminus s^{-1}\{p\}$ contains a point from $\partial\mathbb{D}$.

3.2 Lemma *Let Y be a metric space and $s\colon \mathbb{D}\to Y$ be a metric-minimizing disc. Then s has no bubbles.*

Proof Assume the contrary; that is, there is $y\in Y$ such that the complement $\mathbb{D}\setminus s^{-1}(y)$ contains a connected component Ω with $\partial\mathbb{D}\cap\Omega=\varnothing$.

Let us define a new map $s'\colon \mathbb{D}\to Y$ by setting $s'(z)=y$ for any $x\in\Omega$ and $s'(x)=s(x)$ for any $x\notin\Omega$.

By construction, s' and s agree on $\partial\mathbb{D}$. Moreover, $s\succcurlyeq s'$ because of the majorization $\mu\colon \hat\pi_s(x)\mapsto\hat\pi_{s'}(x)$.

Note that

$$\langle x-x'\rangle_s>0=\langle x-x'\rangle_{s'}$$

for a pair of distinct points $x,x'\in\Omega$. In particular, μ is not an isometry, a contradiction. □

3.3 Lemma *Let Y be a metric space and assume that a map $f\colon \mathbb{D}\to Y$ has no bubbles. Then $|\mathbb{D}|_f$ is homeomorphic to a disc retract with boundary curve $\bar\pi_f|_{\mathbb{S}^1}$.*

This lemma is nearly identical to [16, Corollary 7.12]; it could be considered a disc version of Moore's quotient theorem [19; 7] which states that if a continuous map f from the sphere $\mathbb{S}^2$ to a Hausdorff space X has acyclic fibers, then f can be approximated by a homeomorphism; in particular, X is homeomorphic to $\mathbb{S}^2$.

Proof From Lemma 2.1 we know that $\bar\pi_f$ is continuous and hence $|\mathbb{D}|_f$ is a compact metric space.

The mapping cone over $\mathbb{D}$ along its boundary is homeomorphic to the sphere $\mathbb{S}^2$; denote by Σ the mapping cone over $|\mathbb{D}|_f$ with respect to $\bar{\pi}_s|_{\mathbb{S}^1}$. Let us extend the map $\bar{\pi}_f$ to a map between the mapping cones $\mathbb{S}^2 \to \Sigma$. Note that this map satisfies Moore's quotient theorem, hence the statement follows. □

Proof of Proposition 3.1 The first two statements follow from Lemmas 3.2 and 3.3.

Since $|\mathbb{D}|_s$ is a disc retract, the mapping cylinder over the boundary curve of $|\mathbb{D}|_s$ is homeomorphic to $\mathbb{D}$. Denote by $r\colon \mathbb{D} \to |\mathbb{D}|_s$ the corresponding retraction.

Note that $\langle\mathbb{D}\rangle_{\bar{s}\circ r}$ is isometric to $|\mathbb{D}|_s$ equipped with the induced length metric. Recall that the map $\tau_s\colon \langle\mathbb{D}\rangle_s \to |\mathbb{D}|_s$ is short and the induced map $\mu\colon \langle\mathbb{D}\rangle_s \to \langle\mathbb{D}\rangle_{\bar{s}\circ r}$ is a majorization. Since s is metric-minimizing, μ is an isometry. Hence the statement follows. □

Assume W is a disc retract with a boundary curve δ. Recall that a point p in a connected space W is a *cut point* if the complement $W \setminus \{p\}$ is not connected.

3.4 Lemma *Suppose that W is a disc retract. Let $\Delta \subset W$ be a maximal connected subset that contains no cut points. Assume Δ has at least two points. Then the closure of Δ is homeomorphic to $\mathbb{D}$.*

Proof Let δ be a boundary curve of W, so the mapping cylinder W_δ is homeomorphic to $\mathbb{D}$.

Note that p is a cut point of W if and only if $\delta^{-1}\{p\}$ has at least two connected components. In particular, any cut point of W lies on its boundary curve.

Denote by $\bar{\Delta}$ the closure of Δ. Note that for any $x \notin \bar{\Delta}$ there is a cut point $p \in \bar{\Delta}$ that cuts x from Δ. Moreover, the map $\sigma\colon x \mapsto p$ is uniquely defined on $W \setminus \bar{\Delta}$; extend this map to whole W by identity in $\bar{\Delta}$. By Moore's theorem, $\bar{\Delta}$ is a disc retract with a boundary curve $\sigma \circ \delta$. Namely we apply Moore's theorem to $\mathbb{S}^2 = W_\delta/(\mathbb{S}^1 \times 1)$ and the quotient map $\mathbb{S}^2 \to \mathbb{S}^2/{\sim}$ for the minimal equivalence relation such that $x \sim y$ if $\sigma(x) = \sigma(y)$.

The space $\bar{\Delta}$ has no cut points; in other words, $\sigma \circ \delta$ is monotonic. It follows that $\sigma \circ \delta$ can be reparametrized into a simple closed curve. By the Jordan–Schoenflies theorem, the statement follows. □

The following lemma will be used in the final step in the proof of the main theorem, Section 8.

3.5 Lemma *Let Y be a metric space and $s\colon \mathbb{D} \to Y$ be a metric-minimizing map. Assume that there is a* CAT(0) *disc retract W with boundary curve δ and a short map $f\colon \langle\mathbb{D}\rangle_s \to W$ such that $f\circ\delta_s = \delta$. If there exists a short map $q\colon W \to Y$ with $q\circ\delta = s|_{\partial\mathbb{D}}$, then the map f is an isometry.*

Proof Let $r\colon \mathbb{D} \to W$ be the projection from the mapping cylinder $\mathbb{D} = W_\delta$. Note that r is a retraction, $r|_{\partial\mathbb{D}} = \delta$ and the composition $\mathbb{D} \xrightarrow{r} W \xrightarrow{q} Y$ fulfills

$$q\circ r|_{\partial\mathbb{D}} = s|_{\partial\mathbb{D}}.$$

Note that $\langle W\rangle_q = \langle\mathbb{D}\rangle_{q\circ r}$ and the natural projection $\rho\colon W \to \langle W\rangle_q$ is short. It follows that $\rho\circ f\colon \langle\mathbb{D}\rangle_s \to \langle\mathbb{D}\rangle_{q\circ r}$ is a majorization. Since s is metric-minimizing, $\rho\circ f$ is an isometry.

Therefore f is an isometric embedding that contains δ in its image. By Lemma 2.3 and Proposition 3.1, $\langle\mathbb{D}\rangle_s$ is a complete geodesic space. So f has to be surjective and therefore an isometry. □

4 Compactness lemma

A sequence of pairs (X_n, γ_n), where X_n is a metric space and $\gamma_n\colon \mathbb{S}^1 \to X_n$ is a closed curve is said to *converge* to $(X_\infty, \gamma_\infty)$ if there is a convergence of X_n to X_∞ in the sense of Gromov–Hausdorff for which γ_n converges to γ_∞ pointwise.

More precisely, we ask the following:

(i) There is a metric ρ on the disjoint union

$$\boldsymbol{X} = X_\infty \sqcup X_1 \sqcup X_2 \sqcup \cdots$$

that restricts to the given metric on each X_α for $\alpha \in \{1, 2, \dots, \infty\}$, and such that X_n converge to X_∞ in the sense of Hausdorff as subsets in $(\boldsymbol{X}, \rho)$.

(ii) The sequence of compositions $\gamma_n\colon \mathbb{S}^1 \to X_n \hookrightarrow \boldsymbol{X}$ converges to $\gamma_\infty\colon \mathbb{S}^1 \to X_\infty \hookrightarrow \boldsymbol{X}$ pointwise.

Consider the class $\mathcal{K}_\ell$ of CAT(0) disc retracts whose marked boundary curves have Lipschitz constant ℓ.

4.1 Compactness lemma *$\mathcal{K}_\ell$ is compact in the topology described above.*

The lemma follows from the two lemmas below.

4.2 Lemma *$\mathcal{K}_\ell$ is precompact in the topology described above.*

Proof Let K be a metric space with the isometry class in $\mathcal{K}_\ell$.

Denote by area A the two-dimensional Hausdorff measure of $A \subset K$. By the Euclidean isoperimetric inequality we have

$$\operatorname{area} K \leqslant \pi \cdot \ell^2.$$

Fix $\varepsilon > 0$. Set $m = \lceil 10 \cdot \frac{\ell}{\varepsilon} \rceil$. Choose m points $y_1, \dots, y_m$ on ∂K that divide ∂K into arcs of equal length.

Consider the maximal set of points $\{x_1, \dots, x_n\}$ such that

$$d(x_i, x_j) > \varepsilon \quad \text{and} \quad d(x_i, y_j) > \varepsilon.$$

Note that the set $\{x_1, \dots, x_n, y_1, \dots, y_m\}$ is an ε–net in (K, d). Further note that the balls $B_i = B_{\varepsilon/2}(x_i)$ do not overlap.

By comparison,

$$\operatorname{area} B_i \geqslant \tfrac{1}{4}\pi \cdot \varepsilon^2.$$

It follows that $n \leqslant 4 \cdot \left(\frac{\ell}{\varepsilon}\right)^2$. In particular, there is a integer-valued function $N(\varepsilon)$ such that any K as above contains an ε–net with at most $N(\varepsilon)$ points.

The latter means that the class $\mathcal{K}_\ell$ is uniformly totally bounded. By the selection theorem [6, Theorem 7.4.15], the class of metrics with this property is precompact in the Gromov–Hausdorff topology.

Since the set of ℓ–Lipschitz maps defined on $\mathbb{S}^1$ with compact target is compact with respect to pointwise convergence, we conclude that $\mathcal{K}_\ell$ is precompact in the topology defined above. □

4.3 Lemma *$\mathcal{K}_\ell$ is closed in the topology described above.*

Proof Let (X_n, γ_n) be a sequence in $\mathcal{K}_\ell$. Assume $X_n \to X$ and $\gamma_n \to \gamma$. Choose a point $o_n \in \partial X_n$ and define $f_n : \mathbb{D} \to X_n$ by sending the geodesic $[0, \theta]$ for $\theta \in \partial\mathbb{D} = \mathbb{S}^1$ to the geodesic path $[o_n, \gamma_n(\theta)]$ with constant speed.

By comparison, f_n is a $(5 \cdot \ell)$–Lipschitz continuous ruled disc. The limit map $f : \mathbb{D} \to X$ is also a $(5 \cdot \ell)$–Lipschitz continuous ruled disc. In particular, f is metric-minimizing. By Lemma 3.2, f has no bubbles. Therefore, by Lemma 3.3, X is a disc retract with boundary curve γ. □

5 Metric-minimizing graphs

Metric-minimizing graphs are defined analogously to metric-minimizing discs.

Namely, let Y be a metric space, Γ be a finite graph and A be a subset of its vertices. Given two maps $f, f'\colon \Gamma \to Y$, we write $f \succcurlyeq f'$ (rel A) if f and f' agree on A and there is a majorization $\mu\colon \langle\Gamma\rangle_f \to \langle\Gamma\rangle_{f'}$ such that $f(a) = f'(\mu(a))$ for any $a \in A$.

A map $f\colon \Gamma \to Y$ is called *metric-minimizing relative to* A if $f \succcurlyeq f'$ (rel A) implies that the majorization μ is an isometry.

5.1 Proposition *Let Y be a* $\mathrm{CAT}(0)$ *space, Γ be a finite graph and A be a subset of its vertices.*

Given a continuous map $f\colon \Gamma \to Y$ there is a map $h\colon \Gamma \to Y$ that is metric-minimizing relative to A and $f \succcurlyeq h$ (rel A).

Proof Let us parametrize each edge of Γ by $[0,1]$. A map $h\colon \Gamma \to Y$ will be called *straight* if it sends each edge of Γ to a constant-speed geodesic path in Y.

If $h\colon \Gamma \to Y$ is straight, then $f \succcurlyeq h$ (rel A) if and only if

$$|f(v) - f(w)|_Y \geqslant |h(v) - h(w)|_Y$$

for any two adjacent vertices v and w in Γ. In particular, we can assume that the given map f is straight.

By finiteness of the number of vertices and Zorn's lemma, it is sufficient to prove that for any ordered sequence of straight maps $f_1 \succcurlyeq f_2 \succcurlyeq \cdots$ there exists a map $f \preccurlyeq f_n$ for all n.

Assume to the contrary; let us apply the *ultralimit/projection construction*.

Namely, fix an ultrafilter ω; denote by Y^ω the ultrapower of Y. Then Y^ω is a $\mathrm{CAT}(0)$ space that contains Y as a closed convex subset. The ω–limit $f_\omega\colon \Gamma \to Y^\omega$ is well defined since all f_n are Lipschitz continuous. Denote by f' the composition of f_ω with the nearest-point projection $Y^\omega \to Y$. The nearest-point projection to a closed convex set in $\mathrm{CAT}(0)$ space is short. Therefore $f_n \succcurlyeq f'$ for any n. Let f'' denote the straightening of f'. Then $f' \succcurlyeq f''$ and the claim follows. □

5.2 Proposition *Let Y be a* $\mathrm{CAT}(0)$ *space, Γ be a finite graph and A be a subset of its vertices. Assume that the assignment $v \mapsto v'$ is a metric-minimizing map $\Gamma \to Y$ relative to A. Then*

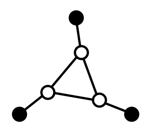
(a) *each edge of Γ maps to a geodesic*;

(b) *for any vertex* $v \notin A$ *and any* $x \neq v'$ *there is an edge* $[v, w]$ *in* Γ *such that* $\measuredangle\big[v'\,{}^{w'}_{x}\big] \geqslant \frac{\pi}{2}$;

(c) *for any vertex* $v \notin A$ *and any cyclic order* $w_1, \dots, w_n$ *of adjacent vertices we have*

$$\measuredangle\big[v'\,{}^{w_1'}_{w_2'}\big] + \dots + \measuredangle\big[v'\,{}^{w_{n-1}'}_{w_n'}\big] + \measuredangle\big[v'\,{}^{w_n'}_{w_1'}\big] \geqslant 2\cdot\pi.$$

Remark The conditions in the proposition do not guarantee that the map f is metric-minimizing. An example can be guessed from the diagram



where the solid points form the set A.

Proof The first condition is evident.

Assume the second condition does not hold at a vertex $v \notin A$; that is, there is a point $x \in Y$ such that $\measuredangle\big[v'\,{}^{w'}_{x}\big] < \frac{\pi}{2}$ for any adjacent vertex w. In this case moving v' toward x along $[v', x]$ decreases the lengths of all edges adjacent to v, a contradiction.

Assume the third condition does not hold; that is, the sum of the angles around a fixed interior vertex v' is less than $2\cdot\pi$.

Recall that the space of directions $\Sigma_{v'}$ is a CAT(1) space. Denote by $\xi_1, \dots, \xi_n$ the directions of $[v', w_1'], \dots, [v', w_n']$ in $\Sigma_{v'}$. By assumption, we have

$$|\xi_1 - \xi_2|_{\Sigma_{v'}} + \dots + |\xi_n - \xi_1|_{\Sigma_{v'}} < 2\cdot\pi.$$

By Reshetnyak's majorization theorem, the closed broken line $[\xi_1, \dots, \xi_k]$ is majorized by a convex spherical polygon P.

Note that P lies in an open hemisphere with pole at some point in P. Choose $x \in Y$ so that the direction from v' to x coincides with the image of the pole in $\Sigma_{f(v)}$. This choice of x contradicts (b). □

6 Key lemma

6.1 Lemma *Let* Y *be a* CAT(0) *space and* $s\colon \mathbb{D} \to Y$ *be a metric-minimizing disc. Assume* $F \subset \mathbb{D}$ *is a finite set such that* $\hat{\pi}_s(F)$ *has finite diameter in* $\langle\mathbb{D}\rangle_s$. *Then there exists a finite piecewise geodesic graph* Γ *embedded in* $\langle\mathbb{D}\rangle_s$ *that contains a geodesic between any pair of points in* $\hat{\pi}_s(F)$.

Proof For any pair $x, y \in F$, connect $\hat{x}$ to $\hat{y}$ by a minimizing geodesic in $\langle\mathbb{D}\rangle_s$. We can assume that the constructed geodesics are either disjoint or their intersection is formed by finite collections of arcs and points.

Indeed, if some number of geodesics $\gamma_1, \dots, \gamma_n$ already has this property and we are given points x and y, then we choose a minimizing geodesic γ_{n+1} from x to y that maximizes the time it spends in $\gamma_1, \dots, \gamma_n$ in the order of importance. Namely,

- among all minimizing geodesics connecting x to y choose one that spends maximal time in γ_1 — in this case γ_{n+1} intersects γ_1 along the empty set, a one-point set or a closed arc;
- among all minimizing geodesics as above choose one that spends maximal time in γ_2 — in this case γ_{n+1} intersects γ_2 along at most two arcs and points;
- and so on.

It follows that together the constructed geodesics form a finite graph Γ as required. □

6.2 Key lemma *Let Y be a* CAT(0) *space and $s\colon \mathbb{D} \to Y$ be a metric-minimizing disc. Given a finite set $F \subset \mathbb{D}$ there is*

(a) *a* CAT(0) *disc retract W with boundary curve δ;*

(b) *a map $p\colon F \to W$ such that*

$$|p(x) - p(y)|_W \leqslant \langle x - y\rangle_s$$

for $x, y \in F$ and $p(x) = \delta(x)$ for $x \in F \cap \partial\mathbb{D}$;

(c) *a short map $q\colon W \to Y$ such that*

$$s(x) = q \circ p(x)$$

for any $x \in \partial\mathbb{D} \cap F$.

Proof If $\partial\mathbb{D} \cap F = \varnothing$, then one can take a one-point space as W and arbitrary maps $p\colon F \to W$ and $q\colon W \to Y$. So suppose $\partial\mathbb{D} \cap F \neq \varnothing$.

Without loss of generality we may assume that the distance $\langle x - y\rangle_s$ between any pair of points $x, y \in F$ is finite. Indeed, since the boundary curve $s|_{\partial\mathbb{D}}$ is rectifiable, this always holds for pairs of points in $\partial\mathbb{D} \cap F$. Consider the subset $F' \subset F$ that lies at finite $\langle * - * \rangle_s$–distance from one (and therefore any) point in $\partial\mathbb{D} \cap F$. Suppose $p'\colon F' \to W$ and $q\colon W \to Y$ are maps satisfying the proposition for F'. Extend p'

to F by sending $F \setminus F'$ to one point in W. The resulting map p together with q will then satisfy the proposition for F.

By Lemma 6.1, there exists a finite piecewise geodesic graph Γ embedded in $\langle\mathbb{D}\rangle_s$ that contains F as a subset of its vertices. According to Proposition 3.1, τ_s embeds Γ in $|\mathbb{D}|_s$. By Proposition 3.1, $|\mathbb{D}|_s$ is a disc retract. Therefore Γ can be (and will be) considered as a graph embedded into the plane.

By Proposition 5.1, there is a map $u\colon \Gamma \to Y$ metric-minimizing relative to $A = F \cap \partial\mathbb{D}$ such that

$$s|_\Gamma \succcurlyeq u \text{ (rel } A). \tag{1}$$

Fix an open disc Δ cut by Γ from $|\mathbb{D}|_s$. By Reshetnyak's theorem, the closed curve $u|_{\partial\Delta}$ is majorized by a convex plane polygon, possibly degenerate to a point or a line segment. Note that the angle of the majorizing polygon cannot be smaller than the angle between the corresponding edges in $u(\Gamma) \subset Y$.

Let us glue the majorizing polygons into $\langle\Gamma\rangle_u$; denote by W the resulting space. According to Proposition 5.2(c), the angle around each inner vertex has to be at least $2\cdot\pi$. Clearly W is a disc retract; in particular, it is simply connected. It follows that W is a $\mathrm{CAT}(0)$ space.

The short map $q\colon W \to Y$ is constructed by gluing together the maps provided by Reshetnyak's majorization theorem. The space W comes with a natural short map $\langle\Gamma\rangle_u \to W$.

Define $p(x)$ for $x \in F$ as the image of the corresponding vertex of Γ in W. By (1),

$$|p(x) - p(y)|_W \leqslant \langle x - y\rangle_s$$

for any $x, y \in F$.

By construction the pair of maps p and q meet all the conditions. □

The following establishes a connection between the key lemma, Lemma 6.2, and the extension lemma, Lemma 7.1.

6.3 Lemma *Let Y be a $\mathrm{CAT}(0)$ space and $s\colon \mathbb{D} \to Y$ be a metric-minimizing disc. Let W be a $\mathrm{CAT}(0)$ disc retract with boundary curve δ. For a given finite set $F \subset \mathbb{D}$ we define $\mathfrak{S}_F$ to be the family of maps $p\colon F \to W$ such that*

$$|p(x) - p(y)|_W \leqslant \langle x - y\rangle_s$$

for $x, y \in F$ *with* $p(x) = \delta(x)$ *for* $x \in F \cap \partial\mathbb{D}$ *and such that there exists a short map* $q\colon W \to Y$ *with*

$$s(x) = q \circ p(x)$$

for any $x \in \partial\mathbb{D} \cap F$*. Then* $\mathfrak{S}_F$ *is closed under pointwise convergence.*

The proof is an application of the ultralimit/projection construction; we used it once before and will use it again later.

Proof Consider a converging sequence $p_n \in \mathfrak{S}_F$; denote by p_∞ its limit. For each p_n there is a short map $q_n\colon W \to Y$ satisfying the condition above. Pass to its ultralimit $q_\omega\colon W \to Y^\omega$. Recall that Y is a closed convex set in Y^ω. In particular, the nearest-point projection $\nu\colon Y^\omega \to Y$ is well defined and short. Therefore, the composition $q = \nu \circ q_\omega$ is short. Finally note that the maps $p_\infty\colon F \to W$ and $q\colon W \to Y$ satisfy the condition above. □

7 Extension lemma

7.1 Extension lemma *Suppose that* X *is a set and* Y *is a compact topological space. Assume that for any finite set* $F \subset X$ *a nonempty set* $\mathfrak{S}_F$ *of maps* $F \to Y$ *is given such that*

- $\mathfrak{S}_F$ *is closed under pointwise convergence*;
- *for any subset* $F' \subset F$ *and any map* $h \in \mathfrak{S}_F$ *the restriction* $h|_{F'}$ *belongs to* $\mathfrak{S}_{F'}$*.*

Then there is a map $h\colon X \to Y$ *such that* $h|_F \in \mathfrak{S}_F$ *for any finite set* $F \subset X$*.*

Proof Consider the space Y^X of all maps $X \to Y$ equipped with the product topology. Denote by $\overline{\mathfrak{S}}_F$ the set of maps $h \in Y^X$ such that its restriction $h|_F$ belongs to $\mathfrak{S}_F$. By assumption, the sets $\overline{\mathfrak{S}}_F \subset Y^X$ are closed and any finite intersection of these sets is nonempty.

According to Tikhonov's theorem, Y^X is compact. By the finite intersection property, the intersection $\bigcap_F \overline{\mathfrak{S}}_F$ for all finite sets $F \subset X$ is nonempty. Hence the statement follows. □

Note that if X and Y are metric spaces and A is a subset in X then one can take as $\mathfrak{S}_F$ the short maps $F \to Y$ that coincide with a given short map $A \to Y$ on $A \cap F$. This way we obtain the following corollary; it is closely related to [11, Proposition 5.2]. In a similar fashion, we will use the lemma in the proof of our main theorem.

7.2 Corollary *Let X and Y be metric spaces, $A \subset X$ and $f\colon A \to Y$ a short map. Assume Y is compact and for any finite set $F \subset X$ there is a short map $F \to Y$ that agrees with f in $F \cap A$. Then there is a short map $X \to Y$ that agrees with f in A.*

8 Proof assembling

Proof of the main theorem Given a finite set $F \subset \mathbb{D}$, denote by $\mathcal{W}_F$ the set of isometry classes of spaces W that meet the conditions of Lemma 6.2 for F; according to this lemma, $\mathcal{W}_F \neq \varnothing$. Note that for two finite sets $F \subset F'$ in $\mathbb{D}$, we have $\mathcal{W}_F \supset \mathcal{W}_{F'}$.

According to the compactness lemma (Lemma 4.1) $\mathcal{W}_F$ is compact. Therefore

$$\mathcal{W} = \bigcap_F \mathcal{W}_F \neq \varnothing,$$

where the intersection is taken over all finite subsets F in $\mathbb{D}$.

Fix a space W from $\mathcal{W}$; the space W is a CAT(0) disc retract such that given a finite set $F \subset \mathbb{D}$ there is a map $h_F\colon F \to W$ that is short with respect to $\langle * - * \rangle_s$ and a short map $q_F\colon W \to Y$ such that $q_F \circ h_F$ agrees with s on $\partial\mathbb{D} \cap F$.

Given a finite set $F \subset \mathbb{D}$, denote by $\mathfrak{S}_F$ the set of all maps $h_F\colon F \to W$ described above.

By Lemma 6.3, $\mathfrak{S}_F$ is closed. The condition on the restriction of $h_F \in \mathfrak{S}_F$ in the extension lemma, Lemma 7.1, is evident. Applying the lemma, we get a map $h\colon \mathbb{D} \to W$ such that $h|_F \in \mathfrak{S}_F$ for any finite set $F \subset \mathbb{D}$.

Our next aim is to show that there is a single map q such that for all finite sets F the composition $q \circ h|_F$ agrees with s on $\partial\mathbb{D} \cap F$. This is done by applying the ultralimit/projection construction:

Choose a sequence of finite sets F_n such that F_n get denser and denser in $\mathbb{D}$ and the intersections $F_n \cap \partial\mathbb{D}$ get denser and denser in $\partial\mathbb{D}$; denote by q_n the corresponding maps. Let $q_\omega\colon W \to Y^\omega$ be the ultralimit of q_n and set $q = \nu \circ q_\omega$, where $\nu\colon Y^\omega \to Y$ is the nearest-point projection. By construction $q\colon W \to Y$ is short and $q \circ h$ agrees with s on $\partial\mathbb{D}$. Note that we cannot conclude $s \succcurlyeq q \circ h$ because h might not be continuous.

By construction, the map h induces a short map $\hat{h}\colon \langle \mathbb{D} \rangle_s \to W$ such that $\hat{h} \circ \delta_s$ is the boundary curve of W. By Lemma 3.5, $\hat{h}$ is an isometry and the statement follows. □

9 About the old theorem

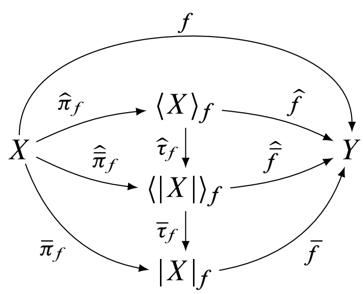
In this section we formulate and prove two versions of the main theorem in [20] with corrections as in [22].

Note that the meaning of the term *metric-minimizing* in the present paper differs from its meaning in [20] — the old paper used a weaker definition and the theorem requires an additional assumption.

In Section 2, we introduced three ways to pull back the metric along a map $f\colon X \to Y$ from a topological space X to a metric space Y: the induced length metric $\langle x - y\rangle_f$, the induced intrinsic metric $\langle |x - y| \rangle_f$ and the induced connecting metric $|x - y|_f$. From the definitions we have that

$$\langle x - y\rangle_f \geqslant \langle |x - y| \rangle_f \geqslant |x - y|_f$$

for any $x, y \in X$. The second inequality is strict for generic maps; an example of a map f with strict first inequality is given in [21, Example 4.2]. (If f is an embedding then equality holds [12, Proposition 4.5].) The diagram



is an extension of the diagram on page 3117 that includes $\langle |X| \rangle_f$. From above, both maps $\widehat{\tau}_f$ and $\overline{\tau}_f$ are short and $\tau_f = \overline{\tau}_f \circ \widehat{\tau}_f$. Note that τ_f might be not injective while $\overline{\tau}_f$ is always injective.

Our first formulation uses the $\langle |{*}-{*}| \rangle$–metric instead of $\langle * - * \rangle$, which was used in the old formulation.

9.1 Old theorem for $\langle |* - *| \rangle$ *Let Y be a* CAT(0) *space and $s\colon \mathbb{D} \to Y$ a continuous map that satisfies the following property: if a continuous map $s'\colon \mathbb{D} \to Y$ agrees with s on $\partial\mathbb{D}$ and*

$$\langle |x - y| \rangle_{s'} \leqslant \langle |x - y| \rangle_s$$

for any $x, y \in \mathbb{D}$ then the equality holds for all pairs of x and y. Assume that the space $\langle |\mathbb{D}| \rangle_s$ is compact. Then $\langle |\mathbb{D}| \rangle_s$ is CAT(0).

Proof Note that s has no bubbles; it can be proved the same way as Lemma 3.2. By Lemma 3.3, $|\mathbb{D}|_s$ is a disc retract.

Note that the natural map $\overline{\tau}_s\colon \langle|\mathbb{D}|\rangle_s \to |\mathbb{D}|_s$ is injective and continuous. Since $\langle|\mathbb{D}|\rangle_s$ is compact and $|\mathbb{D}|_s$ is Hausdorff, the map $\overline{\tau}_s$ is a homeomorphism. Since $\overline{\pi}_s$ is continuous, so is $\overline{\overline{\pi}}_s$.

In particular, $\langle|\mathbb{D}|\rangle_s$ is a disc retract as well as $|\mathbb{D}|_s$. Therefore the mapping cylinder of the boundary curve in $|\mathbb{D}|_s$ is homeomorphic to $\mathbb{D}$. Let us identify $\mathbb{D}$ with the mapping cylinder of the boundary curve in $|\mathbb{D}|_s$. Denote by $h\colon \mathbb{D} \to |\mathbb{D}|_s$ the natural projection; it maps the cylinder to the boundary curve and does not move the points in $|\mathbb{D}|_s$. (If $|\mathbb{D}|_s$ is a disc, we can assume instead that h is a homeomorphism.) Note that $\langle\mathbb{D}\rangle_{\overline{s}\circ h}$ is isometric to $\langle|\mathbb{D}|\rangle_s$.

If s is not metric-minimizing, then there is another map $h'\colon \mathbb{D} \to Y$ such that $\overline{s}\circ h \succcurlyeq h'$ with nonisometric majorization $\mu\colon \langle|\mathbb{D}|\rangle_{\overline{s}\circ h} = \langle|\mathbb{D}|\rangle_s \to \langle\mathbb{D}\rangle_{h'}$. For the composition $s' = \widehat{h}\circ\mu\circ\overline{\overline{\pi}}_s$, we have that

$$\langle x-y\rangle_{s'} \leqslant \langle|x-y|\rangle_s$$

and therefore

$$\langle|x-y|\rangle_{s'} \leqslant \langle|x-y|\rangle_s \tag{2}$$

for any $x, y \in \mathbb{D}$.

By the assumption, equality holds in (2) for any x and y. Since $\langle|\mathbb{D}|\rangle_s$ is compact, applying ultralimit/projection construction, we can assume that s' is metric-minimizing. By the main theorem, $\langle|\mathbb{D}|\rangle_{s'}$ is $\mathrm{CAT}(0)$ disc retract.

Note that we can assume that $\mu\colon \langle|\mathbb{D}|\rangle_s \to \langle\mathbb{D}\rangle_{h'}$ is saddle; otherwise it can be shortened, which would lead to a strict inequality in (2) for some x and y. From the main theorem in [23] it follows that μ is monotonic (we need to apply the theorem for each disc provided by Lemma 3.4 in the disc retract $\langle\mathbb{D}\rangle_{h'}$). Since equality holds in (2), μ has to be an isometry — a contradiction. □

In the second formulation use the metric $\langle * - *\rangle$ as in the original formulation.

9.2 Old theorem for $\langle * - *\rangle$ *Let Y be a $\mathrm{CAT}(0)$ space and $s\colon \mathbb{D} \to Y$ a continuous map that satisfies the following property: if a continuous map $s'\colon \mathbb{D} \to Y$ agrees with s on $\partial\mathbb{D}$ and*

$$\langle x-y\rangle_{s'} \leqslant \langle x-y\rangle_s$$

for any $x, y \in \mathbb{D}$*, then the equality holds for all pairs of* x *and* y*. Assume that the function* $(x, y) \mapsto \langle x - y \rangle_s$ *is continuous. Then* $\langle \mathbb{D} \rangle_s$ *is* CAT(0)*.*

Note that continuity of the function $(x, y) \mapsto \langle x - y \rangle_s$ implies that $\langle \mathbb{D} \rangle_s$ is compact. Therefore the former condition is stronger than the latter. The sketch of the proof given in [20] implicitly used that the metric $(x, y) \mapsto \langle x - y \rangle_s$ is continuous. (We do not know if compactness of $\langle \mathbb{D} \rangle_s$ alone is sufficient.)

Theorem 9.2 follows from the following proposition and Theorem 9.1.

9.3 Proposition *Let* Y *be a metric space and* $s\colon \mathbb{D} \to Y$ *is a continuous map without bubbles. Assume that the function* $(x, y) \mapsto \langle x - y \rangle_s$ *is continuous. Then*

$$\langle x - y \rangle_s = \langle |x - y| \rangle_s$$

for any $x, y \in \mathbb{D}$*.*

Before going into the proof, let us give an example showing that the proposition is not trivial.

Consider a pseudoarc $P \subset \mathbb{D}$ and let s be the quotient map $\mathbb{D} \to \mathbb{D}/P$. Evidently

$$\langle |x - y| \rangle_s = 0$$

for any $x, y \in P$. However, since P contains no curves, it is not at all evident that

$$\langle x - y \rangle_s = 0$$

for any $x, y \in P$.

We present an argument of Taras Banakh [4]; it works for two-dimensional disc and we do not know a generalization of the proposition to higher dimensions.

9.4 Lemma *Let* Y *be a metric space and* $s\colon \mathbb{D} \to Y$ *a continuous map without bubbles. Assume that the function* $(x, y) \mapsto \langle x - y \rangle_s$ *is continuous. Then there is a finite collection of curves in* $\mathbb{D}$ *with finite total* $\langle * - * \rangle$*–length that divide* $\mathbb{D}$ *into subsets with arbitrary small* $\langle * - * \rangle$*–diameter.*

Proof Note that given $\varepsilon > 0$ there is $\delta > 0$ such that a set of diameter δ in $\langle \mathbb{D} \rangle_s$ cannot separate a set of diameter at least ε from the boundary curve. If this is not the case, then sets of arbitrary small diameter can separate a set of diameter at least ε from the boundary. Passing to a limit we get a one-point set that separates a set from the boundary curve; that is, s has a bubble — a contradiction.

Let us subdivide $\mathbb{D}$ into small pieces by curves, say by vertical and horizontal lines. Since the function $(x,y)\mapsto\langle x-y\rangle_s$ is continuous, we can assume that all pieces have small $\langle *{-}*\rangle_s$–diameter; that is, given $\varepsilon>0$ we can assume that $\langle x-y\rangle_s<\varepsilon$ for any two points x and y in one piece.

It remains to modify the decomposition to make the boundary curves $\langle *{-}*\rangle_s$–rectifiable.

Subdivide the curves into arcs with $\langle *{-}*\rangle_s$–diameter smaller than $\frac{\delta}{5}$ and exchange this piece by a curve of $\langle *{-}*\rangle_s$–length smaller than $\frac{\delta}{5}$. The new arc together with the old one form a set of $\langle *{-}*\rangle_s$–diameter at most δ.[1] Therefore we might add to a piece a subset of diameter at most ε and the total $\langle *{-}*\rangle_s$–diameter of each piece remains below $3\cdot\varepsilon$. The curves might cut more pieces from $\mathbb{D}$, but by the same argument each of these pieces will have $\langle *{-}*\rangle_s$–diameter below $3\cdot\varepsilon$. □

Proof of Proposition 9.3 Fix points $x,y\in\mathbb{D}$. It is sufficient to construct a path α from x to y in $\mathbb{D}$ such that the length of $s\circ\alpha$ is arbitrarily close to $\langle|x-y|\rangle_s$.

Since $(x,y)\mapsto\langle x-y\rangle_s$ is continuous, $\langle\mathbb{D}\rangle_s$ and therefore $\langle|\mathbb{D}|\rangle_s$ are compact. In particular, there is a minimizing geodesic γ from $\hat{\hat{x}}=\hat{\hat{\pi}}(x)$ to $\hat{\hat{y}}=\hat{\hat{\pi}}(y)$ in $\langle|\mathbb{D}|\rangle_s$. Denote by Γ the inverse image of γ in $\mathbb{D}$; this is a connected compact set which does not have to be path connected.

To construct the needed path α, it is sufficient to prove the following claim:

($\star$) Given $\varepsilon>0$ there is a set $\Gamma'\subset\Gamma$ and a collection of paths $\alpha_0,\dots,\alpha_n$ such that
- (a) the total length of $s(\alpha_i\setminus\Gamma)$ is at most ε,
- (b) the set Γ' is a union of a finite collection of closed connected sets $\Gamma_0,\dots,\Gamma_n$,
- (c) the diameter of each Γ_i is at most ε,
- (d) $x\in\alpha_0$ and $y\in\alpha_n$, and
- (e) the union of $\Gamma'\cup\alpha_0\cup\dots\cup\alpha_n$ is connected.

Indeed, once the claim ($\star$) is proved, one can apply it recursively for a sequence of ε_n that converge to zero very fast. Namely we can apply the claim to each of the subsets Γ_n and take as Γ'' the union of all closed subsets provided by the claim. This way we obtain a nested sequence of closed sets $\Gamma\supset\Gamma'\supset\Gamma''\supset\cdots$ which break into a finite union of closed connected subsets of arbitrary small diameter and a countable collection of arcs with total length at most $\varepsilon_1+\varepsilon_2+\cdots$ outside of Γ. Set

$$\Phi=\Gamma\cap\Gamma'\cap\Gamma''\cap\cdots.$$

[1] This arc might travel far in the Euclidean metric on $\mathbb{D}$.

Note that there is a simple curve from x to y that runs in the constructed arcs and Φ. The part of the curve in Γ contributes at most $\langle|x-y|\rangle_s$ to its $\langle *-*\rangle_s$–length. Therefore the total length of the curve cannot exceed

$$\langle|x-y|\rangle_s+\varepsilon_1+\varepsilon_2+\cdots;$$

hence the result will follow.

It remains to prove ($\star$).

Fix a subdivision $\Upsilon_1,\dots,\Upsilon_k$ of $\mathbb{D}$ provided by Lemma 9.4 for the given ε. Denote by Δ the union of all the cutting curves.

By the regularity of $\langle *-*\rangle_s$–length, we may cover $\Delta\cap\Gamma$ by a finite collection of arcs with total $\langle *-*\rangle_s$–length arbitrary close to $\langle *-*\rangle_s$–length of $\Delta\cap\Gamma$. Denote these arcs by $\alpha_0,\dots,\alpha_n$. Without loss of generality we may assume that $x\in\alpha_0$ and $y\in\alpha_n$.

Consider a finite graph with the vertices labeled by $\alpha_0,\dots,\alpha_n$; two vertices α_i and α_j are connected by an edge if there is a connected set $\Theta\subset\Gamma\cap\Upsilon_k$ for some k such that Θ intersects α_i and α_j. Note that the graph is connected; therefore, we may choose a path from α_0 to α_n in the graph.

The path corresponds to a sequence of arcs α_i and a sequence of Θ–sets. The Θ–sets that correspond to the edges in the path can be taken as Γ_i in the claim. Hence the claim and therefore the proposition follow. □

10 Saddle surfaces

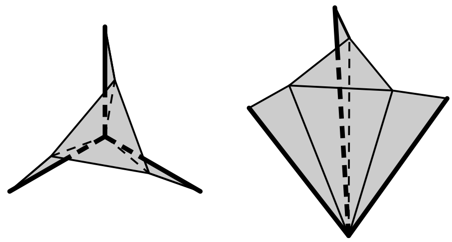
In this section we will discuss the relation between metric-minimizing discs and saddle discs.

Recall that a map $s\colon\mathbb{D}\to\mathbb{R}^m$ is called *saddle* if for any hyperplane $\Pi\subset\mathbb{R}^m$ each of the connected components of $\mathbb{D}\setminus s^{-1}\Pi$ meets the boundary.

If s is a smooth embedding in $\mathbb{R}^3$, then it is saddle if and only if the obtained surface has nonpositive Gauss curvature. An old conjecture of Shefel states that any saddle disc in a $\mathbb{R}^3$ is $\mathrm{CAT}(0)$ with respect to its length metric; see [24].

It is evident that any metric-minimizing disc s in a Euclidean space is saddle.

Three-dimensional case In general a saddle disc may not be globally metric-minimizing. An example is



It is a saddle polyhedral disc made from 10 triangles with a hexagon boundary. The boundary curve goes along the Y–shape marked with bold lines; each segment of the Y–shape is traveled twice back and forth.

The picture is rotationally symmetric by the angle $\frac{2}{3}\cdot\pi$. A shortening deformation can be obtained by rotating the central triangle slightly counterclockwise and extending the map on the remaining 9 triangles linearly.

By smoothing this example one can produce a smooth saddle disc that is not metric-minimizing.

Proposition 1.2 states that there are no local examples of that type that are smooth and *strictly saddle*, meaning that the principal curvatures at each interior point have opposite signs. To prove this proposition we need to introduce a certain energy functional.

Let $s\colon \mathbb{D}\to\mathbb{R}^3$ be a smooth map.

Fix an array of vector fields $\boldsymbol{v}=(v_1,\dots,v_k)$ on $\mathbb{D}$. Assume that each integral curve of vector fields v_i goes from boundary to boundary yielding sweep-outs of the whole disc $\mathbb{D}$.

Consider the energy functional

$$E_{\boldsymbol{v}}s:=\sum_i\int\limits_{\mathbb{D}}|v_is|^2,$$

where $vs=ds(v)$ denotes the derivative of s in the direction of the field v. Set

$$\Delta_{\boldsymbol{v}}s=\sum_i v_i(v_is).$$

It is convenient to think of the operator $s\mapsto\Delta_{\boldsymbol{v}}s$ as an analog of the Laplacian.

Note that:

(i) $E_{\boldsymbol{v}}$ is well defined for any Lipschitz map s.

(ii) $E_{\boldsymbol{v}}$ is convex; that is,

$$E_{\boldsymbol{v}} s_t \leqslant (1-t)\cdot E_{\boldsymbol{v}} s_0 + t\cdot E_{\boldsymbol{v}} s_1,$$

where $s_t = (1-t)\cdot s_0 + t\cdot s_1$ and $0 \leqslant t \leqslant 1$. Moreover, the equality holds for any t if and only if for any i we have $v_i s_0 = v_i s_1$ almost everywhere.

(iii) If s_0 is a smooth $E_{\boldsymbol{v}}$–minimizing map in the class of Lipschitz maps with given boundary data, then s_0 is metric-minimizing.

Indeed, if $s_1 \preccurlyeq s_0$, then s_1 has to be Lipschitz. It follows that $E_{\boldsymbol{v}} s_1 \leqslant E_{\boldsymbol{v}} s_0$ and from convexity $E_{\boldsymbol{v}} s_t \leqslant E_{\boldsymbol{v}} s_0$ if $0 \leqslant t \leqslant 1$. Since s_0 is $E_{\boldsymbol{v}}$–minimizing, $E_{\boldsymbol{v}} s_t = E_{\boldsymbol{v}} s_0$ for any t. Hence $v_i s_0 = v_i s_1$ almost everywhere. Since $\mathbb{D}$ is swept out by arcs of integral curves of v_i, the latter implies $s_0 = s_1$.

(iv) A smooth map $s\colon \mathbb{D} \to \mathbb{R}^3$ is a $E_{\boldsymbol{v}}$–minimizing map among the class of Lipschitz maps with given boundary if and only if

$$\Delta_{\boldsymbol{v}} s = 0.$$

The discussion above reduces Proposition 1.2 to the following:

10.1 Claim *Assume $s\colon \mathbb{D} \to \mathbb{R}^3$ is a smooth strictly saddle surface. Then for any interior point $p \in \mathbb{D}$ there is an array of 4 vector fields $\boldsymbol{v} = (v_1, v_2, v_3, v_4)$ such that the equation*

$$\Delta_{\boldsymbol{v}} s = 0 \tag{3}$$

holds in an open neighborhood of p.

Proof Denote by κ_1 and κ_2 the principal curvatures, and by e_1 and e_2 the corresponding unit principal vectors. Further, denote by a_1 and a_2 a pair of asymptotic vectors; that is, the normal curvatures in these directions vanish. We can assume that a_1 and a_2 form coordinate vector fields in a neighborhood of p.

Set $v_1 = (1/\sqrt{|\kappa_1|})\cdot e_1$ and $v_2 = (1/\sqrt{|\kappa_2|})\cdot e_2$. It remains to show that one can choose smooth functions λ_1 and λ_2 so that (3) holds in a neighborhood of p for $v_3 = \lambda_1\cdot a_1$ and $v_4 = \lambda_1\cdot a_1$.

Note that the sum $v_1(v_1 s) + v_2(v_2 s)$ has vanishing normal part. That is,

$$v_1(v_1 s) + v_2(v_2 s)$$

is a tangent vector to the surface.

Since the a_i are asymptotic, the vectors $a_1(a_1 s)$ and $a_2(a_2 s)$ have vanishing normal part. Therefore, for any choice of λ_i, the following two vectors are also tangent:

$$v_3(v_3 s) = \lambda_1^2 \cdot a_1(a_1 s) + \tfrac{1}{2} \cdot a_1 \lambda_1^2 \cdot a_1 s,$$
$$v_4(v_4 s) = \lambda_2^2 \cdot a_2(a_2 s) + \tfrac{1}{2} \cdot a_2 \lambda_2^2 \cdot a_2 s.$$

Set $w = (\lambda_1^2, \lambda_2^2)$. Note that the system (3) can be rewritten as

$$\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) w_x + \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) w_y = h(x, y, w),$$

where $h\colon \mathbb{R}^3 \to \mathbb{R}^2$ is a smooth function.

Change coordinates by setting $x = t + z$ and $y = t - z$. Then the system takes the form

$$w_t + \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) w_z = h(t + z, t - z, w),$$

which is a semilinear hyperbolic system. According to [5, Theorem 3.6], it can be solved locally for smooth initial data at $t = 0$.

It remains to choose v_3 and v_4 for a solution such that $\lambda_1, \lambda_2 > 0$ in a small neighborhood of p. $\square$

Four-dimensional case Except for constructing an energy as we did above, we do not see any way to show that a given smooth surface is metric-minimizing. Locally, the appropriate energy functional can be described by three functions defined on the disc. These three functions are subject to certain differential equations. Straightforward computations show that on generic smooth saddle surfaces in $\mathbb{R}^4$ there is no solution even locally.

For that reason we expect that generic smooth saddle surfaces in $\mathbb{R}^4$ are not locally metric-minimizing. That is, arbitrary small neighborhoods of any point admit deformations that shrink the length metric and keep the boundary fixed. On the other hand, we do not have an example of a saddle surface for which this condition would hold at a single point.

References

[1] **A D Aleksandrov**, *Ruled surfaces in metric spaces*, Vestnik Leningrad. Univ. 12 (1957) 5–26 MR In Russian

[2] **S B Alexander**, **R L Bishop**, *Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above*, Geom. Dedicata 117 (2006) 65–84 MR

[3] **S Alexander**, **V Kapovitch**, **A Petrunin**, *An invitation to Alexandrov geometry:* CAT(0) *spaces*, Springer (2019) MR

[4] **T Banakh**, *Running most of the time in a connected set*, MathOverflow post (2018) Available at `https://mathoverflow.net/q/308172`

[5] **A Bressan**, *Hyperbolic systems of conservation laws: the one-dimensional Cauchy problem*, Oxford Lecture Series in Mathematics and its Applications 20, Oxford Univ. Press (2000) MR

[6] **D Burago**, **Y Burago**, **S Ivanov**, *A course in metric geometry*, Graduate Studies in Mathematics 33, Amer. Math. Soc., Providence, RI (2001) MR

[7] **R J Daverman**, *Decompositions of manifolds*, Pure and Applied Mathematics 124, Academic, Orlando, FL (1986) MR

[8] **M Gromov**, *Singularities, expanders and topology of maps, I: Homology versus volume in the spaces of cycles*, Geom. Funct. Anal. 19 (2009) 743–841 MR

[9] **A N Kolmogorov**, **S V Fomin**, Элементы теории функций и функционального анализа, 4th edition, Izdat. "Nauka", Moscow (1976) MR Translated as *Elements of the theory of functions and functional analysis*, Dover, New York (1996)

[10] **N J Korevaar**, **R M Schoen**, *Sobolev spaces and harmonic maps for metric space targets*, Comm. Anal. Geom. 1 (1993) 561–659 MR

[11] **U Lang**, **V Schroeder**, *Kirszbraun's theorem and metric spaces of bounded curvature*, Geom. Funct. Anal. 7 (1997) 535–560 MR

[12] **E Le Donne**, *Lipschitz and path isometric embeddings of metric spaces*, Geom. Dedicata 166 (2013) 47–66 MR

[13] **A Lytchak**, *On the geometry of subsets of positive reach*, Manuscripta Math. 115 (2004) 199–205 MR

[14] **A Lytchak**, **S Stadler**, *Conformal deformations of* CAT(0) *spaces*, Math. Ann. 373 (2019) 155–163 MR

[15] **A Lytchak**, **S Wenger**, *Area minimizing discs in metric spaces*, Arch. Ration. Mech. Anal. 223 (2017) 1123–1182 MR

[16] **A Lytchak**, **S Wenger**, *Intrinsic structure of minimal discs in metric spaces*, Geom. Topol. 22 (2018) 591–644 MR

[17] **A Lytchak**, **S Wenger**, *Isoperimetric characterization of upper curvature bounds*, Acta Math. 221 (2018) 159–202 MR

[18] **C Mese**, *The curvature of minimal surfaces in singular spaces*, Comm. Anal. Geom. 9 (2001) 3–34 MR

[19] **R L Moore**, *Concerning upper semi-continuous collections of continua*, Trans. Amer. Math. Soc. 27 (1925) 416–428 MR

[20] **A Petrunin**, *Metric minimizing surfaces*, Electron. Res. Announc. Amer. Math. Soc. 5 (1999) 47–54 MR

[21] **A Petrunin**, *Intrinsic isometries in Euclidean space*, Algebra i Analiz 22 (2010) 140–153 MR In Russian; translated in St. Petersburg Math. J. 22 (2011) 803–812

[22] **A Petrunin**, *Correction to: Metric minimizing surfaces*, Electron. Res. Announc. Math. Sci. 25 (2018) 96 MR

[23] **A Petrunin**, **S Stadler**, *Monotonicity of saddle maps*, Geom. Dedicata 198 (2019) 181–188 MR

[24] **S Z Shefel**, *On the intrinsic geometry of saddle surfaces*, Sibirsk. Mat. Ž. 5 (1964) 1382–1396 MR In Russian

[25] **S Z Shefel**, *On saddle surfaces bounded by a rectifiable curve*, Dokl. Akad. Nauk SSSR 162 (1965) 294–296 MR In Russian

[26] **S Stadler**, *The structure of minimal surfaces in* CAT(0) *spaces*, preprint (2018) arXiv

Department of Mathematics, Pennsylvania State University
University Park, PA, United States

Mathematisches Institut der Universität München
München, Germany

petrunin@math.psu.edu, stadler@math.lmu.de

Proposed: Dmitri Burago
Seconded: Bruce Kleiner, Tobias H Colding

Received: 4 April 2018
Revised: 7 October 2018