AREA SPACES: FIRST STEPS

D. BURAGO AND S. IVANOV, WITH AN APPENDIX BY N. HIGSON

ABSTRACT. We show that a Riemannian manifold of dimension at least 3 can be recovered from the space of boundaries of rectifiable integral 2-currents (the "loops") equipped with the filling area distance, and discuss possible approaches to "spaces with area structures". The first question we asked ourselves was

1. INTRODUCTION

This paper is motivated by the idea of studying spaces using areas instead of lengths. In particular, one can think of a space formed by closed curves equipped with a function that mimics the "minimal filling area". Hence the first question we asked ourselves was: "How much information does this space capture in case of a Riemannian manifold?". The main result of this paper asserts that in dimensions ≥ 3 both the topology and geometry of the manifold in question can be recovered from this "loop space". After proving this result, we took the liberty of including a short section (Section 5) with speculations about possible approaches to "areas spaces" in general, even though we have not gotten almost anywhere in this direction (yet).

Let M be a compact Riemannian manifold. For an integer $m \geq 0$, let $\mathcal{S}_m(M)$ denote the abelian group of all m-dimensional Lipschitz chains with integer coefficients and $\mathcal{B}_m(M)$ the image of the boundary map $\partial : \mathcal{S}_{m+1}(M) \to \mathcal{S}_m(M)$.

The group $\mathcal{B}_1(M)$ is equipped with a (possibly non-homogeneous) semi-norm

$$|\gamma|_F = \inf\{area(s) : s \in \mathcal{S}_2(M), \partial s = \gamma\}$$

In other words, $|\gamma|_F$ is the filling area of γ in M. This yields a semi-metric d_F on $\mathcal{B}_1(M)$:

$$d_F(\gamma_1, \gamma_2) = |\gamma_1 - \gamma_2|_F.$$

Note that the semi-norm $|\cdot|_F$ is not necessarily homogeneous: in general, $|2\gamma|_F \neq 2|\gamma|_F$.

It is possible that $d_F(\gamma_1, \gamma_2) = 0$ for different $\gamma_1, \gamma_2 \in \mathcal{B}_1(M)$ (for instance, a segment traversed back and forth, and a constant curve).

The first author is partially supported by the NSF Grant DMS-0412166. The second author is partially supported by the Dynasty foundation and RFBR grant 08-01-00079-a.

Definition 1.1. Let $\Omega(M)$ denote the associated metric space $\mathcal{B}_1(M)/d_F$, that is, we identify γ_1 and γ_2 if $d_F(\gamma_1, \gamma_2) = 0$. We refer to $\Omega(M)$ as the *loop space* of M.

This space is not complete. This is an Abelian group equipped with an invariant intrinsic metric induced by the filling norm.

Definition 1.2. Let $\overline{\Omega}(M)$ denote the completion of $\Omega(M)$, which is thus also a geodesic space. Note that the norm $|\cdot|_F$ and the abelian group structure naturally extend to $\overline{\Omega}(M)$.

To picture a "nasty" element of $\overline{\Omega}(M)$, consider a smooth surface in M and a Cantor set of positive area (or, more generally, any Borel measurable set) in this surface. The boundary of the corresponding current is an element from $\overline{\Omega}$. In terms of geometric measure theory, $\overline{\Omega}(M)$ is a subset of the space $\mathcal{F}_1(M)$ of integral flat 1-chains in M. More precisely, $\overline{\Omega}(M)$ is the set of all integral flat chains representable as boundaries of rectifiable 2-currents (see Section 2).

Two curves which are located far apart in M may be very close when viewed as elements in $\overline{\Omega}(M)$: For instance, for two far-apart small circles in Euclidean space, the filling distance between them is achieved by a pair of flat discs filling each in.

If $\phi : M_1 \to M_2$ is a smooth map then ϕ pushes forward curves and currents, and so induces a map $\phi_* : \overline{\Omega}(M) \to \overline{\Omega}(M)$. If ϕ is an isometry then ϕ_* is an isometry.

Furthermore, any area preserving diffeomorphism between higher dimensional Riemannian manifolds is an isometry, as can be seen by examining the eigenvalues of the differential of the map. With a little bit more care one can show that if a diffeomorphism between M and M' induces an isomorphism between $\overline{\Omega}(M)$ and $\overline{\Omega}(M')$, then the diffeomorphism is an isometry. The following theorem, which is the main result of this paper, is a far-stretched generalization of this simple observation:

Theorem 1.3. *Let M, M' be compact Riemannian manifolds (possibly with boundaries) of dimension ≥ 3. Suppose that the metric spaces $\overline{\Omega}(M)$ and $\overline{\Omega}(M')$ are isometric via a homogeneous isometry. (An isometry Φ is homogeneous if $\Phi(ks) = k\Phi(s)$ for all $k \in \mathbb{Z}$.) Then M and M' are isometric. Moreover, every homogeneous isometry $\Phi : \overline{\Omega}(M) \to \overline{\Omega}(M')$ is induced by a Riemannian isometry $\phi : M \to M'$ so that $\phi_* = \Phi$.*

Let us emphasize that we do not assume that the homogeneous isometry in the formulation of the Theorem is not assumed to be coming from some map between manifolds, it is a map between abstract Abelian groups with norms.

Note that Theorem 1.3 does not hold in dimension 2. Indeed, any two simply connected surfaces, M_1, M_2, of the same area have an area preserving diffeomorphism between them $f : M_1 \to M_2$ [Moser]. This diffeomorphism induces an isometry between $\overline{\Omega}(M_1)$ and $\overline{\Omega}(M_2)$. Furthermore, one can actually show that the only information about M that can be recovered from $\overline{\Omega}(M)$ is the area of M.

Theorem 1.3 asserts that when two loop spaces are isometric (via a *homogeneous* isometry; this seems to be a minor assumption which we however cannot remove) then the manifolds are isometric. It would have been easier to prove than Theorem 1.3 because everything would have been Lipschitz and rectifiable. However it is important for feasible applications of these results that we only assume in the hypothesis of Theorem 1.3 that the completions of the loop spaces are isometric.

The proof of Theorem 1.3 is contained in Sections 2–4. In the proof, there are notions and constructions of two different types, and it is very important to distinguish them. Namely, there are objects and constructions that use the underlying manifold M, and there are ones formulated entirely in terms referring to $\overline{\Omega}$ as an Abelian group with a (non-homogeneous) norm $|\cdot|_F$ that is a complete geodesic space. Then, for instance, the notions of $sub(C)$, *tube*, *width*, *TubeLength*, *TubeDist*, and *TubeDiam* are well defined on an Abelian groups with an invariant intrinsic metric. On the other hand, such notions as $span(C)$, and results relating $TubeDist$ to Riemannian metric heavily use the Riemannian structure of the manifold M and its relationship with $\overline{\Omega}(M)$.

One might ask whether something can be said about two manifolds with isometric spaces of **real** flat chains. It is clear that two proportional metrics give rise to the same space of real flat chains. One could even guess that, alike the case of L_1-spaces, the space of real currents is simply the same for all manifolds. We do not know if this is true or not.

Acknowledgements. This work has been started in collaboration with *Christina Sormani*. Regretfully, Christina decided to quit the collaboration when this work was nearly completed. We are very grateful to her for many inspiring discussions, without which this paper would hardly appear at all. We are grateful to *Frank Morgan* for many useful remarks. In particular, Frank pointed out to as the existence of minimizing currents spanning non-rectifiable boundaries of rectifiable currents.

2. Geometric Measure Theory Preliminaries

Of course, the language of Lipschitz chains used in the formulation is rather inconvenient. Following the basic techniques of the Geometric Measure Theory, we will work with rectifiable currents and their boundaries. More precisely, we identify $\overline{\Omega}(M)$ with the set of boundaries of rectifiable 2-currents (which actually lies in the space of 1-dimensional integral flat chains) (see [Federer, 4.1.7, 4.1.24] for definitions and details). For a Lipschitz boundary α, $|\alpha|_F$ is the same as the least mass of rectifiable 2-currents spanning α (as follows from [Federer, 4.2.20]). Now we consider $|\cdot|_F$ on the space of all boundaries of rectifiable 2-currents.

The space of rectifiable m-currents in M is denoted by $\mathcal{R}_m(M)$, the space of m-dimensional integral flat chains (that is, the space generated by the rectifiable m-currents and the boundaries of rectifiable $(m+1)$-currents) by $\mathcal{F}_m(M)$.

Note that the traditionally used *flat norm*, introduced to work with currents with boundaries, is different from $|\cdot|_F$. Recall that the *flat norm* $\mathcal{F}$ on $\mathcal{F}_m(M)$ is defined by

$$\mathcal{F}(T) = \inf\{\mathbf{M}(A) + \mathbf{M}(B) : \ T = A + \partial B, \ A \in \mathcal{R}_m(M), \ B \in \mathcal{R}_{m+1}(M)\}$$

(cf. [Federer, 4.1.24]).

We use our filling area norm because it depends only on area (not length) and so Theorem 1.3 takes information purely about areas and produces information about lengths. On the other hand, it is obvious that $\mathcal{F}(T) \le |T|_F$. Furthermore, if T is the boundary of a rectifiable 2-current, then $|T|_F \le C \cdot \mathcal{F}(T)$ for some constant C depending only on M. (This follows from the following fact: if A is a boundary of a rectifiable current, then it can be filled by a current of mass not exceeding $C \cdot \mathbf{M}(A)$.) Therefore, the flat norm $\mathcal{F}$ restricted to boundaries and $|\cdot|_F$ are equivalent before taking the completion, and sequences which converge with respect to our filling area norm $|\cdot|_F$ converge if and only if they converge with respect to the usual flat norm; hence the flat norm compactness theorem and other similar theorems hold for our filling norm as well.

Our loop space $\overline{\Omega}(M)$ is isometric to the $\mathcal{F}$-closure of the set of currents corresponding to Lipschitz chains from $\mathfrak{B}_1(M)$. This set is the set of boundaries of rectifiable 2-chains. This follows from the Approximation Theorem ([Federer, 4.2.20], [Morgan, 7.1]) and the fact that $\mathbf{E}_2$ is $\mathbf{M}$-dense in $\mathcal{R}_2$ (a remark in [Federer, 4.1.24], an exercise in [Morgan, 4.10]).

It is well-known that for every rectifiable boundary there exist a mass-minimizing current filling this boundary. As it was pointed out to us by Frank Morgan, this is also true for non-rectifiable boundaries:

Lemma 2.1 ([Morgan, 5.7], [Federer, 5.1.6])**.** *If a current $T \in \mathcal{F}_1(M)$ is a boundary of a rectifiable 2-current, then $|T|_F$ is realizable by rectifiable 2-currents, namely there exists a $S \in \mathcal{R}_2(M)$ such that*

$$\mathbf{M}(S) = |T|_F$$

We say that S is a minimizing current spanning T.

The existence of minimizers is used throughout the proof.

2.1. **Recovering the Addition.** In this section we study our space, $\overline{\Omega}(M)$ as an Abelian group with a (non-homogeneous) norm ($|\cdot|_F$). Our aim is to show that we can recover the addition of the currents as soon as we have a homogeneous isometry between the spaces:

Proposition 2.2. *If $I : \overline{\Omega}(M) \to \overline{\Omega}(M')$ is a homogeneous isometry, then I is a homomorphism.*

Proof. We need the following lemma.

Lemma 2.3. *Let A and B be torsion-free abelian groups equipped with homogeneous norms $|\cdot|_A$ and $|\cdot|_B$, $I : A \to B$ satisfy $|I(a)|_B = |a|_A$. Then I is a homomorphism.*

Proof. The Lemma follows immediately from Theorem A.1 (see Appendix). □

The proposition follows by applying the above lemma to $A = \overline{\Omega}(M)$, $B = \overline{\Omega}(M')$ and the norms defined as the stabilizations of $|\cdot|_F$ given by $|\gamma| = \lim_{n\to\infty} \frac{|n\gamma|_F}{n}$. One can easily check that this is actually a norm, that is, it does not vanish on non-zero elements. Indeed, let $\gamma \in \overline{\Omega}(M) \setminus \{0\}$. Consider γ as a 1-current. Since $\gamma \neq 0$, there exist a differential 1-form ω on M such that $\gamma \cdot \omega = a \neq 0$. Then, for every 2-current s such that $\partial s = n\gamma$, one has $s \cdot d\omega = \partial s \cdot \omega = na$, hence $\mathbf{M}(s) \geq \frac{|na|}{\|d\omega\|_\infty}$ where $\mathbf{M}$ denotes the mass. Therefore $|\gamma| \geq \frac{|a|}{\|d\omega\|_\infty} > 0$. □

2.2. **Subsets and Spans.** We begin with the following definition:

Definition 2.4. Let $C \subset \overline{\Omega}$. We denote by $sub(C) \subset \overline{\Omega}$ the set of all $C' \in \overline{\Omega}$ such that $|C'|_F + |C - C'|_F = |C|_F$.

Observe that

$$C' \in sub(C) \text{ iff } sub(C') \subset sub(C).$$

Intuitively, when $\overline{\Omega} = \overline{\Omega}(M)$ and C is represented by a closed loop, then $C' \in sub(C)$ iff C' lies in a minimal surface spanned by C.

Definition 2.5. Let $C \in \overline{\Omega}(M)$. We define $span(C) \subset M$ as the union of supports of all minimizing 2-currents spanning C.

Note that, if C is a boundary of a unique minimal surface S, then then $span(C)$ is (the image of) S. The following proposition intuitively says that if a curve, C', lies on a minimal surface filling another curve, then its minimal filling also lies on that minimal surface. The proposition is in fact stronger than this statement since $\overline{\Omega}(M)$ includes a large class of objects:

Proposition 2.6. *If $C, C' \in \overline{\Omega}(M)$ and $C' \subset sub(C)$, then $span(C') \subset span(C)$.*

Proof. Let S' be a minimizing 2-current spanning C'. It suffices to prove that there exist a minimizing 2-current S spanning C such that $spt(S') \subset spt(S)$. Let S_1 be a minimizing 2-current spanning $C - C'$ (whose existence is guaranteed by Lemma 2.1). Define $S = S' + S_1$. Then $\partial S = C$ and

$$\mathbf{M}(S) \leq \mathbf{M}(S') + \mathbf{M}(S_1) = |C|_F + |C - C'|_F = |C|_F,$$

because $C' \in sub(C)$. Hence S is mass-minimizing and $\mathbf{M}(S) = \mathbf{M}(S') + \mathbf{M}(S_1)$. The latter implies that $spt(S) = spt(S') \cup spt(S_1)$, hence $spt(S') \subset spt(S)$. □

Proposition 2.7. *There exists an* $\varepsilon_0 > 0$ *such that, if* $C \in \overline{\Omega}(M)$ *and the support* $spt(C)$ *lies in a closed metric ball* B *of radius* $\varepsilon < \varepsilon_0$*, then* $span(C) \subset B$*.*

Proof. Let ε_0 be one third of the convexity radius of M (that is, every geodesic ball of radius $< 3\varepsilon_0$ is convex). Let $\varepsilon < \varepsilon_0$, B a ball of radius ε centered at o. We want to prove that if S is a minimizing 2-current and $spt(\partial S) \subset B$, then $spt(S) \subset B$.

It suffices to construct a 1-Lipschitz map $f : M \to B$ such that $F|_B = id_B$ and F is strictly contracting (that is, has a Lipschitz constant < 1) outside any neighborhood of B.

To construct such an f, introduce normal spherical coordinates (ρ, θ) in 3ε-neighborhood of o ($\rho \in [0, 3\varepsilon]$, $\theta \in S^2$) and define a map $f : B_{3\varepsilon}(o) \to M$ sending a point (ρ, θ) to itself if $\rho \le \varepsilon$ and to $(\frac{1}{2}(3\varepsilon - \rho), \theta)$ if $\varepsilon \le \rho \le 3\varepsilon$. Define $f(x) = o$ for all $x \in M \setminus B_{3\varepsilon}(o)$. The resulting map $f : M \to M$ has the desired properties. □

3. Areas to Distances

In this section we will work to convert information about areas into information about lengths. Let us describe the intuitive idea in dimension 3. Note that in a Riemannian manifold we can examine the boundary of a tubular neighborhood, $\partial T_\varepsilon(\gamma)$, of a geodesic γ. The length of $\partial T_\varepsilon(\gamma)$ will be close to the length of γ and can be computed using its area and its radius. The radius ε can be computed using the cut area of $T_\varepsilon(\gamma)$ which is close to $\pi\varepsilon^2$.

Since we do not have means of defining tubular neighborhoods without a distance, we need to work towards this concept using minimal surfaces. Instead of using the tubular neighborhood, we could have used a large collection of circular loops C_i lying in $\partial T_\varepsilon(\gamma)$ running around γ which lie very close together so that the minimal surfaces running from each C_i to C_{i+1} looks like a cylinder (a wide catenoid) and building a tube from these minimal surfaces. We then measure the length of this tube using its width and cut area. As a matter of fact, we only "mimic" this geometric picture, for our objects can be defined for an abstract Abelian group $\overline{\Omega}$ with an invariant intrinsic metric.

3.1. Tubes and Cylinders.

Here we begin with a rigorous construction of a tube in an abstract space $\overline{\Omega}$ and define its length in that setting, so that an isometry between $\overline{\Omega}_1$ and $\overline{\Omega}_2$ maps tubes to tubes and preserves their lengths. In the next section we will relate their lengths to Riemannian lengths when $\overline{\Omega} = \overline{\Omega}(M)$.

Definition 3.1. A cylinder is a pair (C_0, C_1) of elements $C_0, C_1 \in \overline{\Omega}$.

Definition 3.2. The width of a cylinder is defined as

$$W(C_0, C_1) = \inf\{|C|_F : |C_0 - C|_F + |C - C_1|_F = |C_0 - C_1|_F\}$$

Example 3.3. For instance, in Euclidean three space, given two circular loops, $C_0(t) = (0, sin(t), cos(t))$ and $C_1(t) = (1, sin(t), cos(t))$, then the flat norm between them is achieved by a standard catenoid, and the width, $W(C_0, C_1)$ is the area of the disk filling the intersection of the catenoid with the plane $z = \frac{1}{2}$.

Things do not always appear as cylindrical as in the catenoid example. For instance, if C_0 and C_1 are two small circles far away from each other, the width $w(C_0, C_1) = 0$ because one can take $C = 0$. It is also possible that C_0 is formed by two circles and C_1 is a single circle, so that the minimal surface between them looks like a pair of pants, or that both C_i consist of pairs of circles and the minimal surface between them is diffeomorphic to a pair of cylinders.

Definition 3.4. A *tube* is a finite sequence $T = (C_0, C_1, \ldots, C_n)$ in $\overline{\Omega}$. We refer to C_0 and C_n as the *bases* of T. The area of T is defined by $area(T) = \sum |C_i - C_{i+1}|_F$.

The reader should think of a tube as a hose made of pieces (cylinders) each looking like a short catenoids. In actuality a tube may be built of pants or its “top” and “bottom” could be completely disconnected. The width of a tube is, intuitively, the size of a smallest cut of the hose. If the hose is disconnected the width will be zero. The smallest slice of a hose could cut across more than one cylinder: imagine a cylinder made of slanted sections. Hence we need the following formal definition, which at first glance may look a bit complicated:

Definition 3.5. We say that a loop $C \in \overline{\Omega}(M)$ is a *cut* of a cylinder (C_0, C_1) if

$$|C_0 - C_1|_F = |C_0 - C|_F + |C - C_1|_F.$$

A loop $C \in \overline{\Omega}(M)$ is said to be a *cut* of a tube $T = (C_0, C_1, \ldots, C_n)$ if the loop $C + C_1 + C_2 + \cdots + C_{n-1}$ can be represented as a sum $C_1' + C_2' + \cdots + C_n'$ where each C_i' is a cut of the cylinder (C_{i-1}, C_i).

An interested reader can see the geometric meaning of this definition by picturing a tube C_0, C_1, C_2 made of two cylinders and with a cut crossing both of them. Dividing this cut into two segments and adding appropriate segments of the middle loop C_2, one gets two cuts of the cylinders. We will however use this definition in a rather formal way.

Definition 3.6. The width $w(T)$ of a tube T is defined by

$$w(T) = \inf\{|C|_F : \ C \text{ is a cut of } T\}$$

Note that the width of $T = (C_1, \ldots, C_n)$ could be significantly smaller than the smallest of the individual widths of cylinders (C_i, C_{i+1}) (picture a slanted slicing of a cylinder).

If a tube were actually a standard cylinder of length L, surface area A, circumference R, and cut area W then:

$$L = A/C = A/(2\pi r) = A/(2\pi\sqrt{W/\pi}) = A/(2\sqrt{\pi}\sqrt{W}).$$

So we define the following *length of a tube*:

Definition 3.7. The length of T is defined by

$$TubeLength(T) = \frac{area(T)}{2\sqrt{\pi}\sqrt{w(T)}}.$$

Since none of "real tubes" in a Riemannian manifold are cylinders, even if the manifold is flat and they are made of catenoids, this length is distinct from the geometric length of the tube.

Example 3.8. If we create a tube (C_0, C_1) from two circles as in the above "catenoid example" then both the area and the width of the tube are less than the corresponding circle, and $TubeLength$ is not 1. On the other hand if we create a tube with the same bases C_0 and C_1 and many many identical circles lying on the cylinder between them, then the area of the tube approaches the area of the cylinder and the width approaches the width of the cylinder, and thus the length of the tube approaches the length of the cylinder.

Later on in the proof of Theorem 1.3 we will be taking limits of areas, width, and lengths of tubes. To prepare for this we first examine the stability of the width:

Lemma 3.9. *For every* $C_0, C_1, \dots, C_n \in \overline{\Omega}(M)$ *one has*

$$w(C_0, \dots, C_n) \ge w(C_0, \dots, C_{n-1}) - |C_{n-1} - C_n|_F.$$

Proof. Let C be a cut of $(C_0, \dots, C_n)$, then by definition

$$C = C_1' + \dots + C_n' - C_1 - \dots - C_{n-1}$$

where each C_i' is a cut of the cylinder (C_{i-1}, C_i). Define

$$A = C_1' + \dots + C_{n-1}' - C_1 - \dots - C_{n-2}.$$

Then A is a cut of $(C_1, \dots, C_k)$, hence $|A|_F \ge w(C_1, \dots, C_{n-1})$. On the other hand, $C - A = C_n' - C_{n-1}$, hence

$$|C - A|_F = |C_n' - C_{n-1}|_F = |C_n - C_{n-1}|_F - |C_n' - C_n|_F \le |C_n - C_{n-1}|_F$$

(the second equality follows from the fact that C_n' is a cut of (C_{n-1}, C_n)). The inequalities $|A|_F \ge w(C_1, \dots, C_{n-1})$ and $|C-A|_F \le |C_n - C_{n-1}|_F$ imply that $|C|_F \ge w(C_0, \dots, C_{n-1}) - |C_{n-1} - C_n|_F$. $\square$

3.2. **TubeLength vs Riemannian Distance.** In this section we use the co-area formula to prove the following relationship between the tube lengths in $\overline{\Omega}(M)$ and the Riemannian distances on M:

Proposition 3.10. *For every tube* $T = (C_0, \dots, C_n)$*,*

$$TubeLength(T) \geq \rho_0(1 - \psi(w(T))),$$

where $\lim_{t\to 0} \psi(t) = 0$*,* ρ_0 *is the Riemannian distance between the supports of* $span(C_0)$ *and* $span(C_n)$*.*

Proof. Let S_i be a minimizing 2-current spanning $C_i - C_{i-1}$, $i = 1, \dots, n$. Define $S = S_1 + \cdots + S_n$. Then $area(T) \geq \mathbf{M}(T)$, and it suffices to prove that $\mathbf{M}(T) \geq 2\sqrt{\pi}\sqrt{w(T)}\rho_0(1 - \psi(w(T)))$.

Let $f : M \to \mathbb{R}$ be the distance function of the set $span(S_0)$. This is a Lipschitz-1 function with respect to the Riemannian structure. Define $\tilde{S} = S + S_0 + S_{n+1}$, where S_0 and S_{n+1} are minimizing 2-currents spanning C_0 and $-C_{n+1}$ (we added the "top" and "bottom" S_0 and S_{n+1} to work with a closed current $\tilde{S}$ rather than S which may have non-rectifiable boundary). We slice $\tilde{S}$ by the level sets of f, cf. [Morgan, 4.11] and [Federer, 4.2.1], namely, for each $0 < r < \rho_0$, we consider the slice $\langle S, f, r+\rangle$ which is defined as the boundary of the restriction of $\tilde{S}$ to the sublevel set of f:

$$\langle S, f, r+\rangle = \partial(S \llcorner \{x : f(x) \leq r\}).$$

Note that the Lipschits constant of f is 1. Then by the co-area inequality, we have

$$\mathbf{M}(S) \geq \mathbf{M}(\tilde{S} \llcorner \{x : 0 < f(x) < \rho_0\} \geq \int_0^{\rho_0} \mathbf{M}(\langle S, f, r+\rangle)\, dr.$$

Our goal is to estimate the integral from below. First we need the following geometrically obvious fact:

Lemma 3.11. *For every* $0 < r < \rho_0$*, the slice* $\langle S, f, r+\rangle$ *is a cut of* T*.*

Proof. We need the following properties of the restriction operator $\llcorner$: for every rectifiable current S and every Borel measurable set $A \subset M$, the restriction $S \llcorner A$ is a rectifiable current and

(0) $S \llcorner M = S$;

(1) if $A \cap spt(S) = \emptyset$, then $S \llcorner A = 0$;

(2) if $A, B \subset M$ are Borel measurable and $A \cap B = \emptyset$, then $S \llcorner (A \cup B) = S \llcorner A + S \llcorner B$ and $\mathbf{M}(S \llcorner (A \cup B)) = \mathbf{M}(S \llcorner A) + \mathbf{M}(S \llcorner B)$;

(3) $(S_1 + S_2) \llcorner A = S_1 \llcorner A + S_2 \llcorner A$ if S_1 and S_2 are rectifiable currents.

Denote $A = \{x : f(x) \leq r\}$, $B = \{x : f(x) > r\}$. Note that $\langle S, f, r+\rangle$ belongs to our space $\overline{\Omega}(M)$ since it is the boundary of a rectifiable 2-current $\tilde{S} \llcorner A$. For $i = 1, \dots, n$ define

$$C_i' = \partial(S_i \llcorner A) + C_{i-1} = C_i - \partial(S_i \llcorner B).$$

The second equality here follows from the fact that $\partial S_i = C_i - C_{i-1}$ and the above properties of $\llcorner$. Then

$$|C_i' - C_{i-1}|_F = |\partial(S_i \llcorner A)|_F \le \mathbf{M}(S_i \llcorner A)$$
$$|C_i - C_i'|_F = |\partial(S_i \llcorner B)|_F \le \mathbf{M}(S_i \llcorner B),$$

hence

$$|C_i' - C_{i-1}|_F + |C_i - C_i'|_F \le \mathbf{M}(S_i \llcorner A) + \mathbf{M}(S_i \llcorner B) = \mathbf{M}(S_i) = |C_i - C_{i-1}|_F.$$

Therefore C_i' is a cut of (C_{i-1}, C_i). Furthermore,

$$C_1' + \cdots + C_n' = \partial(S \llcorner A) + C_0 + C_1 + \ldots C_{n-1}.$$

Observe that

$$\partial(S \llcorner A) + C_0 = \partial(S \llcorner A) + \partial(S_0 \llcorner A) + \partial(S_{n+1} \llcorner A) = \partial(\tilde{S} \llcorner A) = \langle \tilde{S}, f, r+\rangle$$

since $S_0 \llcorner A = S_0$ and $S_{n+1} \llcorner A = 0$ (recall that $spt(S_0) \subset span(C_0) \subset A$, $spt(S_n) \subset span(C_n) \subset B$). Thus

$$C_1' + \cdots + C_n' = \langle \tilde{S}, f, r+\rangle + C_1 + \ldots C_{n-1},$$

hence $\langle \tilde{S}, f, r+\rangle$ is a cut by definition. □

We now continue with the proof of Proposition 3.10.

By Almgren's isoperimetric inequality [Almgren], for every closed rectifiable 1-current C is $\mathbb{R}^n$, $|C|_F \le \frac{1}{4\pi}\mathbf{M}(C)^2$. It follows that, in a compact Riemannian manifold M, a rectifiable 1-current $C \in \overline{\Omega}(M)$ satisfies $|C|_F \le \frac{1}{4\pi}\mathbf{M}(C)^2(1 + \varphi(\mathbf{M}(C)))$ where $\varphi(t) \to 0$ as $t \to 0$ (the correction term φ depends on Riemannian metric). For $C = \langle \tilde{S}, f, r+\rangle$, we have $|C|_F \ge w(T)$ by the definition of width, hence

$$\mathbf{M}(\langle S, f, r+\rangle) \ge 2\sqrt{\pi}\sqrt{w(T)}(1 - \psi(w(T))$$

where $\psi(t) = 1 - (1 + \varphi(t))^{-1/2}$. Applying the above coarea inequality finishes the proof of the proposition. □

Example 3.12. Recall Example 3.8. It was pointed out there that a tower of catenoids formed a tube whose length did not agree with its height but that if the number of catenoids between fixed curves were increased so that the height of each catenoid approached zero then the length approached the total height. By Proposition 3.10 we now know that if we not only decreased the height of the catenoids but also the radii, the limit could not be less than the total height. If we decrease the radii too fast compared to the heights then the tube could break and have width 0 sending the length to infinity.

3.3. **TubeDistance vs Riemannian Distance.**

Definition 3.13. For $C, C' \in \overline{\Omega}(M)$, we say that a tube $T = (C_0, \ldots, C_n)$ connects C and C' if $C_0 \in sub(C), C_n \in sub(C')$. We define

$$TubeDist(C, C') = \lim_{\varepsilon \to 0} \inf\{TubeLength(T) : w(T) < \varepsilon\}$$

where the infimum is taken over tubes T connecting C and C'.

Intuitively the tubular distance measures the smallest tubular length of thinner and thinner tubes with bases lying within two minimal surfaces.

Now we want to show that the Riemannian distance between the spans is actually equal to $TubeDist$. The results of the previous section imply that the Riemannian distance is less than the tubular distance between two spans:

Corollary 3.14. *For all $C, C' \in \overline{\Omega}(M)$, we have*

$$TubeDist(C,C') \geq d_M(span(C), span(C'))$$

Proof. Proposition 3.10 implies that

$$TubeDist(C,C') \geq \lim_{\varepsilon \to 0} \inf\{d_M(span(C_0), span(C_k))(1-\psi(w(T))) : \\ w(C_0,...C_k) < \varepsilon, C_0 \in span(C), C_k \in span(C')\},$$

where $\lim_{t\to 0}\psi(t) = 0$. Now Proposition 2.6 tells us that $span(C_0) \subset span(C)$ and $span(C_k) \subset span(C')$, and the Corollary follows. □

To prove the opposite inequality we need to construct tubes approaching the infimum. The idea is that we can always run a geodesic achieving the distance between $span(C)$ and $span(C')$ and then find $C_0 \in sub(C)$ and $C_1 \in sub(C')$ lying arbitrarily near the ends of this geodesic. Then we build a thin tube around the geodesic. This is intuitively easy and has already been described at the beginning of this section. However, to complete the construction given the large class of limiting objects included in our completion space $\overline{\Omega}(M)$ we first need following lemma and the concept of a *density point*:

Lemma 3.15. *For a rectifiable 2-current $S \neq 0$, there exists a point $x \in spt(S)$ and a C^1 surface Σ passing through x such that, for all sufficiently small r, the disc in Σ of radius r centered at x is $o(r^2)$-close (in $|\cdot|_F$) to a sub-current S_r of S (S_r is said to be a sub-current of S if $\mathbf{M}(C_r)+\mathbf{M}(C-C_r) = \mathbf{M}(C)$). We call such a point x a density point.*

Furthermore, the set of density points x is dense in $spt(S)$.

Proof. By [Federer, 4.1.28, 3.2.29], every rectifiable 2-current S can be represented as a countable sum $S = \sum S_i$ where each S_i is a current associated with a Borel measurable subset of a 2-dimensional oriented C^1 submanifold Σ_i of M. We may assume that there is no cancellations between the S_i's, that is, $\mathbf{M}(S) = \sum \mathbf{M}(S_i)$. Then every S_i is a sub-current of S. We abuse notation and use the same letters S_i for the currents and the respective subsets of surfaces Σ_i. Assuming $S_1 \neq 0$, let $\Sigma = \Sigma_1$ and x a density point of S_1 (here S_1 is regarded as a subset of Σ). Let B_r denote the metric ball of radius r in Σ centered at x, then

$$|B_r - (S_1 \cap B_r)|_F \leq \mathbf{M}(B_r \setminus S_i) = o(r^2).$$

Then $S_1 \cap B_r$ is a desired sub-current.

To show that the set of such points x is dense in $spt(S)$, we just apply the first statement to $S \llcorner U$ where U is an arbitrary open set intersecting $spt(S)$. □

Proposition 3.16. *Let $C, C' \in \Omega(M)$, and let ρ be the Riemannian distance between $span(C)$ and $span(C')$. Then $TubeDist(C, C') = \rho$.*

Proof. By Corollary 3.14, we have $TubeDist(C, C') \geq \rho$. It remains to prove the opposite inequality. Let S and S' be minimizing 2-currents spanning C and C' respectively and such that the distance between them is close to ρ. We need to prove the following: for every $\varepsilon > 0$, there exist a tube $T = (C_0, \dots, C_n)$ such that $C_0 \in sub(C)$, $C_n \in sub(C')$, $w(T) < \varepsilon$ and $TubeLength(T) < \rho + \varepsilon$.

To facilitate understanding, we first consider the case when S and S' are smooth embedded surfaces. To construct a desired tube, consider a simple smooth curve γ connecting two interior points of S and S' such that $L = length(\gamma) < \rho + \varepsilon/5$ and γ is orthogonal to S and S' at endpoints. Let $r > 0$ be so small that the $(5r)$-neighborhood U of γ is bi-Lipschitz diffeomorphic, with bi-Lipschitz constant very close to 1, to the $(5r)$-neighborhood of a straight line segment $[0, Le_1]$ in $\mathbb{R}^m$ where $m = \dim M$. Furthermore, the diffeomorphism can be chosen so that it maps $S \cap U$ and $S' \cap U$ to affine subspaces parallel to the coordinate (e_2, e_3)-plane.

Now consider a solid cylinder $[0, L] \times B_r \subset \mathbb{R}^3 \subset \mathbb{R}^m$ where B_r is the r-ball in $\mathbb{R}^2$. Pick $h \ll r$ and consider loops $C_i = \{hi\} \times \partial B_r$ in $\mathbb{R}^3$, for $i = 0, 2, \dots, n = L/h$. In R^m, the resulting tube $(C_0, \dots, C_n)$ has width $\pi r^2 + o(1)$ and area $2\pi r L + o(1)$ as $h \to 0$, and hence its $TubeLength$ is close L. Since the corresponding Riemannian tube is bi-Lipschitz close to the Euclidean one, its $TubeLength$ is also close to L.

To prove the general case, we choose density points in S, S' such that the Riemannian distance between them is less than $\rho + \frac{1}{5}\varepsilon$ and simply apply the above argument for smooth surfaces guaranteed by Lemma 3.15. The bases of the resulting tube $T = (C_0, \dots, C_n)$ may not lie exactly in S, S'. They however lie within small $|\cdot|_F$-distance from sub-currents C_0', C_n' of S, S' (by the definition of density points), so we can replace the tube by $T' = (C_0', C_0, C_1, \dots, C_n, C_n')$. Then Lemma 3.9 tells us that the widths of T and T' differ by no more than $o(r^2)$, which concludes the proof. □

Recall that for $A \subset M$, the diameter of A is

$$diam(A) = \sup\{d(x, y) : x, y \in A\}.$$

Definition 3.17. Let $C \in \overline{\Omega}(M)$. We define

$$TubeDiam(C) = \sup\{TubeDist(C', C'') : C', C'' \in sub(C), C', C'' \neq 0\}.$$

The following lemma is a trivial corollary of the previous proposition:

Lemma 3.18. *Let $C \in \overline{\Omega}(M)$, then*

$$TubeDiam(C) = diam(span(C)).$$

4. Proof of Theorem 1.3

Recall that we have an isometry $\Phi : \overline{\Omega}(M) \to \overline{\Omega}(M')$ (which is also a homomorphism). We want to construct a corresponding isometry $\phi : M \to M'$.

Definition 4.1. A *loops to a point sequence* is a sequence $\{C_1, C_2, \dots\}$ of elements of $\overline{\Omega}$ such that

$$\lim_{i\to\infty} TubeDiam(C_i) = 0$$

and

$$\lim_{i,j\to\infty} TubeDist(C_i, C_j) = 0.$$

Results of the previous section immediately imply that, given a loops to a point sequence $\{C_i\}$ in $\overline{\Omega}(M)$, the sets $\{span(C_i)\}$ form a Cauchy sequence (with respect to the Hausdorff distance on M), and their diameter goes to zero. Hence $\lim_{i\to\infty} span(C_i)$ is a single point in M.

Definition 4.2. We say the "M-limit" $MLim\{C_i\}$ of a loops to a point sequence $\{C_i\}$ in $\Omega(M)$ is the point $p \in M$ where $\{p\} = \lim_{i\to\infty} span(C_i)$.

Naturally the limit of this sequence with respect to $|\cdot|_F$ is $0 \in \overline{\Omega}(M)$.

Had we not required $TubeDiam(C_i) \to 0$ and just assumed $|C_i|_F \to 0$ then the C_i would not necessarily be localized (say, we could get a sequence of loops formed by pairs of circles shrinking to two fixed points).

Proposition 2.7 implies that for every $p \in M$ there is a loops to a point sequence $\{C_i\}$ in $\overline{\Omega}(M)$ such that $MLim\{C_i\} = p$.

Now we are ready to define ϕ from the formulation of Theorem 1.3. Let $MLim\{C_i\} = p$. The images $\Phi(C_i)$ is a loop to a point sequence for M', and the spans of the loops converge to some point $q \in M'$. We set $\phi(p) = q$. To see that ϕ is correctly defined note that, for two loop to a point sequences $\{C_i\}$ and $\{\tilde{C}_i\}$, one has $MLim\{C_i\} = MLim\{\tilde{C}_i\}$ if and only if $\lim TubeDist(C_i, \tilde{C}_i) = 0$.

Note that, more generally, the Riemannian distance between $MLim\{C_i\}$ and $MLim\{\tilde{C}_i\}$ is equal to $\lim TubeDist(C_i, \tilde{C}_i)$. This immediately implies that ϕ is a distance preserving map. Applying the same argument to Φ^{-1} we conclude that ϕ is a surjective map and hence is an isometry between M and M'.

To conclude the proof, we need to show that the map ϕ_* between $\overline{\Omega}(M)$ and $\overline{\Omega}(M')$ induced by ϕ is the same as Φ. First note that points p lying in $span(C)$ can be characterized in terms of $\overline{\Omega}(M)$ as follows: $p \in span(C)$ if and only if $p = MLim(C_i)$ for some sequence $C_i \in sub(C)$. This implies that $span(\Phi(C)) = \phi(span(C))$ for all $C \in \overline{\Omega}(M)$. Therefore $\Phi(C) = \phi_*(C)$ if C is a smooth loop spanning a unique area-minimizing surface (indeed, in this case $span(C)$ uniquely determines C). Such loops generate a dense subgroup of $\overline{\Omega}(M)$ (since boundaries of smooth minimizing currents are dense in $\overline{\Omega}(M)$, and a sub-current of a smooth minimizing current is a unique minimizer if it has sufficiently small diameter). This implies that $\Phi = \phi_*$. □

5. Area Spaces: Where to Proceed?

We now take the liberty to include a short speculation on prospects of this approach. This work was motivated by a rather naive idea to introduce "area spaces", with an ultimate goal of obtaining some compactness theorems, which could be helpful in proving inequalities involving areas in situations without a good control over lengths.

We will not even attempt to define an area space here: we have not studies enough examples yet. Vaguely speaking, we expect that, similarly to length structures, an area structure could be a function on a certain class of closed curves in a topological space X. Of course, the function should satisfy certain axioms, and examples we looked at suggest that we better keep some class of metrics or length structures on X too.

Furthermore, one could run the recovery construction for an abstract Abelian group with an invariant metric. Of course, this metric should be a length metric, so one does not want to look at examples like $\mathbb{Z}$. Furthermore, as we mentioned earlier, it is crucial that we work with currents with integer coefficients, so perhaps one should think of an Abelian group which sits "as a lattice" in the tensor product of this group and $\mathbb{R}$ equipped with an invariant intrinsic metric. We have not tried studying any examples yet.

Our observation that $\overline{\Omega}(M)$ uniquely determines M for higher dimensional manifold suggests that convergence of "area spaces" could be defined via (some sort of) convergence of corresponding spaces $\overline{\Omega}$. Note that it is very unlikely that the recovery result Theorem 1.3 is stable. Spaces of this type even tend to be universal (like in two dimensions) rather than unique for each manifold. Theorem 1.3 is based on certain rigidity. Perhaps this (plausible) lack of stability could even be good, for whatever convergence one defines for "area spaces', it should not imply metric convergence. An interesting example can already be obtained by looking at two different surfaces Σ_1 and Σ_2 of the same area multiplied by a circle or length ε when $\varepsilon \to 0$ (recall that $\overline{\Omega}(\Sigma_1) = \overline{\Omega}(\Sigma_2)$). It is absolutely not clear what one can conclude from an assumption that $\overline{\Omega}(M)$ admits a short map on $\overline{\Omega}(M')$, and what one should assume to guarantee the existence of an area non-increasing map with certain topological properties.

$\overline{\Omega}(M)$ is full of strange "garbage". For instance, consider a circle of radius $\frac{1}{5}$ centered at every integer point of a huge 3-D cube of size n, and re-scale the cube down to size $n^{-\frac{1}{2}}$. One gets a "loop" whose filling area is $\pi/25$ contained in a microscopic chunk of space. There are "loops" of huge filling norm formed by "dust" of microscopic circles spread all over the manifold. These ugly loops have nothing to do with objects we may be ultimately interested in, but they caused a lot of trouble when we were working on the proof of the main theorem, and they cause a lot of problems in our attempts to define convergence.

Even though $\overline{\Omega}(M)$ uniquely determines M, it would be nice to recover some properties of M (such as the stable norm on second homologies) directly from $\overline{\Omega}(M)$. Such a description would suggest that one could try to work with convergence where the limit is just some "loop space" $\overline{\Omega}$ without any underlying manifold M.

Appendix A. Remark on the Mazur-Ulam Theorem by Nigel Higson

A.1. **Statement of the Theorem.** The Mazur-Ulam theorem asserts that a distance preserving map from a real normed space onto itself that maps zero to itself is necessarily linear. I shall repeat the original argument of [Mazur-Ulam] and observe that it also proves the following generalization:

Theorem A.1. *Let A be an additive subgroup of a real normed space V such that $\mathbb{R} \cdot A = V$. If ϕ is a distance-preserving map from A onto itself that maps zero onto itself, then ϕ is the restriction to A of an isometric linear transformation from V onto itself.*

A.2. **Midpoints.**

Definition A.2. Let X be any metric space. If $x, y \in X$, then define the *midset* of x and y to be

$$M(x, y) = \{\, z \in X : d(x, z) = d(y, z) = \tfrac{1}{2} d(x, y) \,\}.$$

This is obviously a bounded subset of X. It may be empty.

Definition A.3. If M is any bounded subset of X, then define its *sequence of central subsets* $C_0(M), C_1(M), \ldots$ by $C_0(M) = M$ and

$$C_k(M) = \big\{\, z \in C_{k-1}(M) : d(z, w) \le \tfrac{1}{2} \operatorname{diam}\big(C_{k-1}(M)\big) \ \forall w \in C_{k-1}(M) \,\big\}$$

for $k \ge 1$.

The sets $C_k(M)$ may once again be empty. But in any case the diameter of $C_k(M)$ is at most half the diameter of $C_{k-1}(M)$, and as a result the intersection of all the $C_k(M)$ is either empty or consists of a single point.

Definition A.4. If M is any bounded subset of A, then its *center*, if it exists, is the unique point in $\cap_{k \ge 0} C_k(M)$. The *midpoint* of a pair of points $x, y \in A$, if it exists, is the center of the midset $M(x, y)$.

A.3. **Proof of the Theorem.** The proof of the theorem is based on the previous definitions and the following trivial observation:

Lemma A.5. *If ϕ is an isometry of a metric space X onto itself, and if $x, y \in X$, then the midpoint of x and y exists if and only if the midpoint of $\phi(x)$ and $\phi(y)$ exists. If both do exist, then ϕ maps the first onto the second. In particular, if ϕ maps the set $\{x, y\}$ onto itself, and if the midpoint of x and y exists, then ϕ fixes the midpoint.* □

Mazur and Ulam noted that if X is a normed space, then the midpoint of x and y exists and is the average $\frac{1}{2}(x+y)$. As a result, surjective isometries of normed spaces preserve averages, and hence, by a small additional argument, are affine.

To prove Theorem A.1 one needs to be just a little more careful in dealing with the average operation, which is not generally available in an abelian group. From now on let A be an abelian group equipped with a distance function such that

$$d(x+z, y+z) = d(x,y), \quad d(2x,2y) = 2d(x,y) \quad \text{and} \quad d(x,y) = d(-x,-y),$$

for all $x, y, z \in A$.

Lemma A.6. *Let $x, y \in A$. The midpoint of x and y exists if and only if there is an element $z \in A$ such that $2z = x + y$. If such an element z exists, then it is the midpoint of x and y.*

Proof. Define an isometry $\phi_{x,y}\colon A \to A$ by the formula

$$\phi_{x,y}(z) = x + y - z.$$

This isometry is its own inverse, it exchanges x and y, and its fixed point set consists of those $z \in A$ such that $2z = x + y$. By Lemma A.5, if the midpoint z of x and y exists, then $\phi_{x,y}$ fixes it. We see that if the midpoint z exists, then $z \in M(x,y)$ and $2z = x + y$.

Conversely, suppose $z \in A$ and $2z = x + y$. We should like to show that z is the midpoint of x and y; that is, we should like to show that z belongs to each midset $C_k(M(x,y))$. Certainly $z \in M(x,y)$ because

$$x - y = 2x - 2z \quad \text{and} \quad y - x = 2y - 2z,$$

so that by our assumptions on the distance function

$$d(x,y) = 2d(x,z) = 2d(y,z).$$

Assume that $z \in C_{k-1}(M(x,y))$. If $w \in C_{k-1}(M(x,y))$, then the element

$$v = \phi_{x,y}(w) = x + y - w$$

also belongs to $C_{k-1}(M(x,y))$. But

$$d(v,w) = d(x+y-w, w) = d(x+y, 2w) = d(2z, 2w) = 2d(z,w),$$

from which it follows that

$$d(z,w) \le \tfrac{1}{2} \operatorname{diam}\bigl(C_{k-1}(M(x,y))\bigr),$$

and so $z \in C_k(M(x,y))$. □

Proof of Theorem A.1. Let A and V be as in the statement of the theorem and let ϕ be an isometry of A onto itself that fixes $0 \in A$. Let $x \in A$. By Lemma A.6, the element x is the midpoint of the elements 0 and $2x$. It follows from Lemma A.5 that $\phi(x)$ is the midpoint of 0 and $\phi(2x)$, and therefore by Lemma A.6 again,

$$2\phi(x) = \phi(2x).$$

Now let $x, y \in A$. By Lemma A.6, the element $x + y$ is the midpoint of $2x$ and $2y$, and so $\phi(x + y)$ is the midpoint of $\phi(2x)$ and $\phi(2y)$. One final application of Lemma A.6 tells us that

$$2\phi(x + y) = \phi(2x) + \phi(2y) = 2\phi(x) + 2\phi(y),$$

and so $\phi(x + y) = \phi(x) + \phi(y)$. Thus ϕ is an automorphism of the abelian group A. Since ϕ is a group automorphism, and since the norm on V is positive-homogeneous, we may extend ϕ to an isometric group automorphism of each group $\frac{1}{n!} \cdot A \subseteq V$, and hence of $\mathbb{Q} \cdot A \subseteq V$, by means of the formula $\phi(ax) = a\phi(x)$. Since $\mathbb{R} \cdot A = V$, this extension now extends further by continuity to an isometric automorphism of V. □

References

[Almgren] Almgren, F. J., Jr. *Optimal isoperimetric inequalities.* Indiana Univ. Math. J. **35** (1986), 451–547.

[Federer] Federer, H. *Geometric Measure Theory.* Springer-Verlag, New York, 1969.

[Mazur-Ulam] Mazur, S. and Ulam, S. *Sur les transformations isométriques d'éspaces vectoriels normés.* C. R. Acad. Sci. Paris **194** (1932) 946–948.

[Morgan] Morgan, F. *Geometric Measure Theory. A Beginner's Guide*, Third edition. Academic press, 2000.

[Moser] Moser, J. *On the volume elements on a manifold*, Transactions of the American Mathematical Society **120** 1965 286-294.

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802

E-mail address: burago@math.psu.edu

Steklov Math. Institute, St. Petersburg, Russia

E-mail address: svivanov@pdmi.ras.ru