



**HAL**  
open science

# Tight Bounds for the Learning of Homotopy à la Niyogi, Smale, and Weinberger for Subsets of Euclidean Spaces and of Riemannian Manifolds

Dominique Attali, Hana Dal Poz Kouřimská, Christopher Fillmore, Ishika Ghosh, André Lieutier, Elizabeth Stephenson, Mathijs Wintraecken

## ► To cite this version:

Dominique Attali, Hana Dal Poz Kouřimská, Christopher Fillmore, Ishika Ghosh, André Lieutier, et al.. Tight Bounds for the Learning of Homotopy à la Niyogi, Smale, and Weinberger for Subsets of Euclidean Spaces and of Riemannian Manifolds. SoCG 2024 - 40th International Symposium on Computational Geometry, Jun 2024, Athènes, Greece. 10.4230/LIPIcs.SoCG.2024.11 . hal-04628805

**HAL Id: hal-04628805**

**<https://hal.science/hal-04628805>**

Submitted on 1 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Tight Bounds for the Learning of Homotopy à la Niyogi, Smale, and Weinberger for Subsets of Euclidean Spaces and of Riemannian Manifolds

Dominique Attali ✉

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

Hana Dal Poz Kouřimská ✉ 

IST Austria, Klosterneuburg, Austria

Christopher Fillmore ✉ 

IST Austria, Klosterneuburg, Austria

Ishika Ghosh ✉ 

IST Austria, Klosterneuburg, Austria


Michigan State University, East Lansing, MI, USA

André Lieutier ✉

Aix-en-Provence, France

Elizabeth Stephenson ✉ 

IST Austria, Klosterneuburg, Austria

Mathijs Wintraecken ✉ 

Inria Sophia Antipolis, Université Côte d'Azur, Sophia Antipolis, France

---

## Abstract

In this article we extend and strengthen the seminal work by Niyogi, Smale, and Weinberger on the learning of the homotopy type from a sample of an underlying space. In their work, Niyogi, Smale, and Weinberger studied samples of  $C^2$  manifolds with positive reach embedded in  $\mathbb{R}^d$ . We extend their results in the following ways:

- As the ambient space we consider both  $\mathbb{R}^d$  and Riemannian manifolds with lower bounded sectional curvature.
- In both types of ambient spaces, we study sets of positive reach – a significantly more general setting than  $C^2$  manifolds – as well as general manifolds of positive reach.
- The sample  $P$  of a set (or a manifold)  $\mathcal{S}$  of positive reach may be noisy. We work with two one-sided Hausdorff distances –  $\varepsilon$  and  $\delta$  – between  $P$  and  $\mathcal{S}$ . We provide tight bounds in terms of  $\varepsilon$  and  $\delta$ , that guarantee that there exists a parameter  $r$  such that the union of balls of radius  $r$  centred at the sample  $P$  deformation-retracts to  $\mathcal{S}$ . We exhibit their tightness by an explicit construction.

We carefully distinguish the roles of  $\delta$  and  $\varepsilon$ . This is not only essential to achieve tight bounds, but also sensible in practical situations, since it allows one to adapt the bound according to sample density and the amount of noise present in the sample separately.

**2012 ACM Subject Classification** Theory of computation → Computational geometry

**Keywords and phrases** Homotopy, Inference, Sets of positive reach

**Digital Object Identifier** 10.4230/LIPIcs.SoCG.2024.11

**Related Version** *Full Version*: <https://arxiv.org/abs/2206.10485> [13]

**Funding** This research has been supported by the European Research Council (ERC), grant No. 788183, by the Wittgenstein Prize, Austrian Science Fund (FWF), grant No. Z 342-N31, and by the DFG Collaborative Research Center TRR 109, Austrian Science Fund (FWF), grant No. I 02979-N35.



© Dominique Attali, Hana Dal Poz Kouřimská, Christopher Fillmore, Ishika Ghosh, André Lieutier, Elizabeth Stephenson, and Mathijs Wintraecken; licensed under Creative Commons License CC-BY 4.0

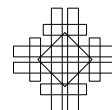
40th International Symposium on Computational Geometry (SoCG 2024).

Editors: Wolfgang Mulzer and Jeff M. Phillips; Article No. 11; pp. 11:1–11:19



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



*Mathijs Wintraecken*: Supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 754411, the Austrian science fund (FWF) grant No. M-3073, and the welcome package from IDEX of the Université Côte d’Azur.

**Acknowledgements** We thank Jean-Daniel Boissonnat, Herbert Edelsbrunner, and Mariette Yvinec for discussion.

## 1 Introduction

Can we infer the topology of a set if we are only given partial geometric information about it? Under which conditions is such inference possible?

These questions were first motivated by the shape reconstruction of objects in 3-dimensional Euclidean space. There, the partial geometric information was represented by a finite, in general noisy, set of points obtained from photogrammetric or lidar measurements [10, 19, 21, 22, 32].

More recently, the same questions have arisen in the context of learning and topological data analysis (TDA). In these fields, one seeks to recover a (relatively) low-dimensional support of a probability measure in a high-dimensional space, given a (finite) data set drawn from this probability measure [23, 29, 40, 38]. Assuming the support is a manifold, one calls this process manifold learning [62].

In [60], Niyogi, Smale, and Weinberger showed that, given a  $C^2$  manifold of positive reach<sup>1</sup> embedded in Euclidean space and a sufficiently dense point sample on (or near) the manifold, the union of balls of certain radii centred on the point sample captures the homotopy type of the manifold. By the nerve theorem [40], the homotopy type of the union of balls is shared by the Čech complex [20, 41] and  $\alpha$ -complex [39] of the point sample. From these complexes we can then learn the topological information such as the homology groups of the underlying manifold. Niyogi, Smale, and Weinberger’s homotopy learning result has led to numerous generalizations including [11, 15, 27, 52, 69].

In this article, we revisit the work of Niyogi, Smale, and Weinberger, generalizing the settings of their work in various ways.

The first generalization is in terms of ambient space – we consider both the Euclidean space  $\mathbb{R}^d$  and Riemannian manifolds with bounded sectional curvature. To this end, we introduce a new version of the reach in the Riemannian setting inspired by the cut locus (see Definition 13).

The second generalization lies in the types of sets we study – we consider sets of positive reach and manifolds of positive reach. Sets of positive reach need not be manifolds – in fact, they can have varying dimensions (see for example Figure 1). Manifolds with positive reach are  $C^{1,1}$  smooth<sup>2</sup>, i.e., differentiable with Lipschitz derivative. This is a significantly larger family of sets in comparison to  $C^2$  manifolds with positive reach, considered by Niyogi, Smale, and Weinberger.

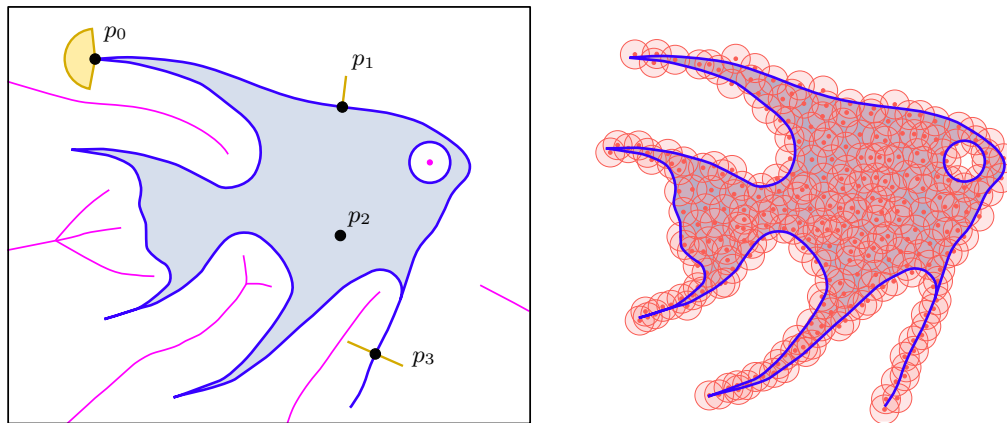
As in the work of Niyogi, Smale, and Weinberger, our settings consist of a set (or a manifold)  $\mathcal{S}$  of positive reach and its sample  $P$ . We distinguish two sample quality parameters – sample density  $\varepsilon$  and sample noisiness  $\delta$ , which we encode using one-sided Hausdorff distances between  $P$  and  $\mathcal{S}$ . We provide explicit conditions on  $\varepsilon$  and  $\delta$ , under which there exists a

<sup>1</sup> We recall that the reach of a closed subset in Euclidean space is the distance from the set to its medial axis. In turn, the medial axis of a set consists of those points in Euclidean space that do not have a unique closest point on the set. Both notions are defined in [13, Definition 18].

<sup>2</sup> Topologically embedded manifolds with positive reach are  $C^{1,1}$  embedded [42, 58, 59, 63, 64].

parameter  $r$  such that the union of balls of radius  $r$  centred at the sample  $P$  deformation-retracts to  $\mathcal{S}$ . This result expands on the work of Niyogi, Smale, and Weinberger, who considered the cases  $\delta = 0$  and  $\delta = \varepsilon$  only, and only achieved tight bounds in the latter case (see Figure 2).

Furthermore, given a set of positive reach  $\mathcal{S}$  and its sample  $P$ , we identify an interval of radii  $r$  (equation (4)) for which the union of balls of radius  $r$  centred at the sample  $P$  deformation-retracts to  $\mathcal{S}$ . Thus, we provide a *guarantee* for a successful homotopy inference of the set  $\mathcal{S}$  from the sample  $P$ . Moreover, we show that for a specific choice of  $\mathcal{S}$  and  $P$  (see Propositions 8, 9, [13, Proposition 47], and [13, Proposition 48]), the homotopy of  $\mathcal{S}$  is not inferable from  $P$  if our conditions on  $\varepsilon$  and  $\delta$  are not satisfied, proving that our bounds are, in terms of  $\varepsilon$  and  $\delta$ , tight.



■ **Figure 1** Left: A fish shaped set  $\mathcal{S}$  of positive reach (in blue). Its medial axis (in purple) is at a positive distance. For  $0 \leq i \leq 3$ , we also represent the normal cone of  $p_i$  with respect to  $\mathcal{S}$  (after an intersection with a small disk and a translation to  $p_i$ ). The normal cone of the point  $p_2$  is  $p_2$  itself. Right: The set  $\mathcal{S}$  with a sample  $P$  and a thickening of  $P$ . We see that the thickening has the same homotopy type as  $\mathcal{S}$ .

This article provides a solid overview of our results. We concentrate on carefully explaining the setting and only state our results. We shifted most proofs and technical details, and in particular the vast machinery we used to prove our statements for subsets and submanifolds of Riemannian manifolds, into the full version of the article [13].

## 2 State-of-the-art

### 2.1 Sets of positive reach

Our extension of Niyogi, Smale, and Weinberger’s result to sets of positive reach – as well as improvement of their results on manifolds – relies on the work of Federer [42], which Niyogi, Smale, and Weinberger have not cited. In particular, we use Federer’s generalization of normal spaces to normal cones (see Figure 1 (left) for a pictorial introduction and [13, Appendix A.1] for a full definition) and his different characterizations of the normal cone as a key building block. We recall the relevant results from Federer’s work in [13, Appendix A.1].

We note that the reach can be estimated from a sample [2, 3, 17, 35, 37].

Subsets of positive reach of Riemannian manifolds were studied extensively by Kleinjohann [53, 54] and Bangert [16] in generalization of Federer’s theory [42] for subsets of Euclidean space. Boissonnat and Wintraecken investigated yet another definition of the reach for subsets of Riemannian manifolds in [24].

## 2.2 Homotopy learning

For some particular cases, the best previously known bounds on the distance between a manifold (or a set) of positive reach and its sample that guarantee successful homotopy inference, can be found in [15] and [60]. Attali *et al.* [15], Chazal *et al.* [27], and Kim *et al.* [52] expanded homotopy learning to even more general subsets of Euclidean space, such as subsets with positive  $\mu$ -reach. Their proofs are, however, different from ours, more involved, and their bounds are not shown to be tight.

## 2.3 Manifold and stratification learning

Although this article focuses on homotopy learning, our work should also be seen as part of recent developments in manifold learning [4, 5, 43, 44, 45, 66]. The goal of this field is to reconstruct a manifold from a “reasonable” sample lying on or near it – at least up to a homeomorphism, but usually an ambient isotopy.

At the moment work is ongoing to expand this strategy to more general spaces – see for example the work of Aamari *et al.* [1] on manifolds with boundary.

Although inferring the homotopy of a manifold is simpler than manifold learning, the sets we consider are more general than manifolds or manifolds with boundary. The extension of learning from subsets of Euclidean space to subsets of Riemannian manifolds also departs from the usual track. We are only aware of one work in computational geometry and topology which operates within this context, namely [30]. These are the first steps in the developing field of stratification learning. Homotopy inference in the hyperbolic space was considered in [11].

## 3 Contribution

### 3.1 Subsets of Euclidean space

Let  $\mathcal{M}$  denote a manifold of positive reach,  $\mathcal{S}$  a set of positive reach and let  $P$  be a sample. All sets are assumed to be compact unless stated otherwise. We denote the reach of a set  $\mathcal{X}$  by  $\text{rch}(\mathcal{X})$  and let  $\mathcal{R}$  be a non-negative real number such that  $\mathcal{R} \leq \text{rch}(\mathcal{S})$  (resp.  $\mathcal{R} \leq \text{rch}(\mathcal{M})$ ).

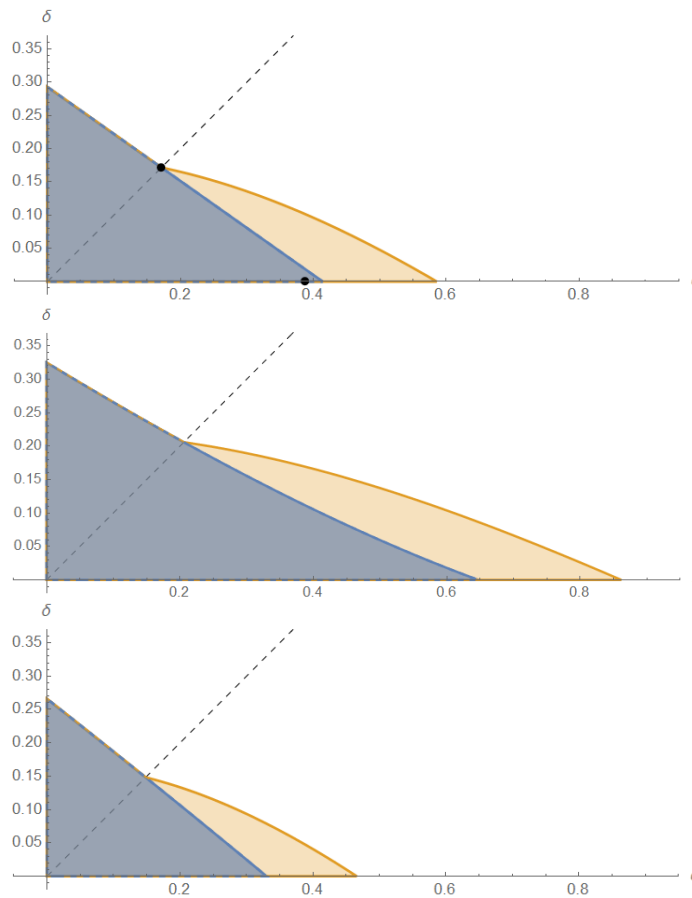
We denote the bound on the one-sided Hausdorff distance<sup>3</sup> from  $P$  to  $\mathcal{S}$  (resp.  $\mathcal{M}$ ) by  $\varepsilon$ , and the one-sided Hausdorff distance from  $\mathcal{S}$  (resp.  $\mathcal{M}$ ) to  $P$  by  $\delta$ .

In this article we establish conditions on  $\varepsilon$  and  $\delta$  which, if satisfied, guarantee the existence of a radius  $r > 0$  such that the union of balls of radius  $r$  centred at the sample  $P$  deformation-retracts onto  $\mathcal{M}$  (resp.  $\mathcal{S}$ ). The set of pairs  $(\varepsilon, \delta)$  that satisfy these conditions is depicted in Figure 2 on the left. The precise conditions are given in Propositions 5 and 7.

Distinguishing the two one-sided Hausdorff distances seems natural to the authors, because in measurements one would expect the measurement error  $\delta$  (with the exception of some small number of outliers) to be often smaller than the sampling density  $\varepsilon$ . Similar assumptions seem to be common in the learning community, see e.g. [56]. Niyogi, Smale, and Weinberger [60] also made similar assumptions on the support of the measure from which they sampled.

We only consider samples for which we have precise bounds on  $\varepsilon$  and  $\delta$ . In [60], the authors also consider a setting where the point sample is drawn from a distribution centred on the manifold. They still recover the homotopy type of the underlying manifold with high probability. Our results can be applied to improve the bounds also in this context. However, we have not discussed this in detail, since combining both results is straightforward.

<sup>3</sup> We recall that the one sided Hausdorff distance from  $X$  to  $Y$ , denoted by  $d_H^o(X; Y)$ , is the smallest  $\rho$  such that  $Y$  is covered by the union of balls of radius  $\rho$  centred at  $X$ , that is,  $Y \subseteq \bigcup_{x \in X} B(x, \rho)$ .



■ **Figure 2** The blue-gray region bounded by the blue dashed curve represents the set of pairs  $(\epsilon, \delta)$  for which there exists a radius  $r$  such that the union of balls of radius  $r$  centred at  $P$  captures the homotopy type of a set of positive reach  $\mathcal{R} = 1$ . The equivalent region for a manifold of reach  $\mathcal{R} = 1$  is depicted in yellow and is a superset of the previous one. The two regions coincide above the diagonal  $\delta = \epsilon$ . The bounds for the Euclidean setting are indicated on top, for an ambient manifold with positive curvature bound ( $\Lambda_\ell = +2$ ) in the middle, and for an ambient manifold with negative curvature bound ( $\Lambda_\ell = -2$ ) bottom. In the top picture, the black points indicate the bounds that were known to Niyogi, Smale, and Weinberger.

We stress that in [23, 60], and [69], the authors use  $\epsilon/2$  instead of our  $\epsilon$ . We also stress that  $\epsilon$  and  $\delta$  have precisely opposite meanings in [52] compared to this paper.

Our conditions on  $\epsilon$  and  $\delta$  are optimal for sets of dimension at least 2 in the following sense: if the conditions are not satisfied, we can construct a set of positive reach  $\mathcal{S}$  (resp. manifold  $\mathcal{M}$ ) and a sample  $P$ , such that there is no  $r \geq 0$  for which the union of balls of radius  $r$  centred at  $P$  would have the same homology as  $\mathcal{S}$  (resp.  $\mathcal{M}$ ). These constructions are explained in Section 4.4.

We would like to emphasize that for noiseless samples, (that is, when  $\delta = 0$ ,) both the constant  $(\sqrt{2} - 1)$  (for general sets of positive reach), and the constant  $(2 - \sqrt{2})$  (for manifolds) compare favourably with the previously best known constant  $\frac{1}{2}\sqrt{\frac{3}{5}}$  from [60] for manifolds.<sup>4</sup>

<sup>4</sup> It should be noted that in [60]  $r$  was not considered as a variable, but set equal to  $2\epsilon$ , which (at least partially) explains the suboptimal result in that paper.

In Proposition 7.1 of [60], one encounters the condition  $\varepsilon < (3 - \sqrt{8})\mathcal{R}$  for a particular case of the setting we consider, namely when the sampling condition is expressed through an upper bound  $\varepsilon$  on the Hausdorff distance ( $\delta = \varepsilon$  in our setting). The same constant  $3 - \sqrt{8}$  appears independently in [14, Theorem 4] for general sets of positive reach. Our results (Propositions 8 and 9) show that this bound is optimal when  $\delta = \varepsilon$ , both for general sets of positive reach and for manifolds.

To contrast the two related results in [60], for  $\delta = 0$  and  $\delta = \varepsilon$  respectively, with our bounds, we portray them as black dots in Figure 2.

Homotopy reconstruction of manifolds with boundary has been studied in [69, Theorem 3.2], assuming lower bounds on both the reach of the manifold and the reach of its boundary. We also improve on this result by treating a manifold with boundary as a particular case of a set of positive reach, while our bounds only depend on the reach of the set itself and not the one of its boundary.

### 3.2 Subsets of Riemannian manifolds

In the second part of this article we extend the homotopy reconstruction results to sets  $\mathcal{S}$  and manifolds  $\mathcal{M}$  of positive reach embedded in a Riemannian manifold whose sectional curvatures<sup>5</sup> are bounded.

Also in this Riemannian setting we find tight<sup>6</sup> bounds on the one-sided Hausdorff distances  $\varepsilon$  and  $\delta$  between  $\mathcal{S}$  (resp.  $\mathcal{M}$ ) and its sample  $P$ . The set of pairs  $(\varepsilon, \delta)$  that satisfy these conditions is depicted in Figure 2 (centre and right). The precise bounds are given in Propositions 15 and 16.

The main pillar of this part of our work is comparison theory. We recall the most essential definitions and results in [13, Appendix C], and refer to [18, 25, 26, 34, 47, 51] for further reading.

For the extension to the Riemannian setting we also formulate a new generalization of the reach. To establish some of its properties, we use results on the gradient of the distance function [9], see also [57]. These results in turn require non-smooth analysis [36] and semi-concave functions [8]. We refer to [13, Appendix G] for discussion.

In computer vision, many papers have argued in favour of using Riemannian manifolds as the main setting without embedding the Riemannian manifold in Euclidean space. In particular, symmetric positive definite matrices and Grassmannians form the natural stage for some data [68, 71]. Symmetric positive definite matrices occur as diffusion tensors [61] (used in e.g. magnetic resonance imaging), in image segmentation [46, 65], and in texture classification [67], while Grassmannians are used in image matching and recognition [48, 49]. Although it is possible to embed these manifolds in Euclidean space, it is not natural and would increase the dimensionality significantly. In [70], time-series obtained from observations of dynamical systems are encoded as positive semi-definite matrices, produced by forming Hankel matrices and taking their Gram matrices. Thus, the problem of analysing time-series data is transformed into the problem of analysing point set data on a Riemannian manifold, namely the one formed by semi-positive definite matrices.

<sup>5</sup> We recall (one of) the (equivalent) definition(s) of sectional curvatures of the Riemannian manifold  $\mathcal{N}$ : For a point  $p \in \mathcal{N}$  let  $\Pi \subseteq T_p\mathcal{N}$  be a two dimensional plane in the tangent space to  $p$  at  $\mathcal{N}$ . If  $U \subseteq \Pi$  is a sufficiently small neighbourhood of  $p$  in  $\Pi$ , then  $\exp_p(U)$  is a surface. The Gauss curvature of this surface at  $p$  is the sectional curvature of  $\mathcal{N}$  at  $p$  for the directions that span  $\Pi$ .

<sup>6</sup> When the curvature of the ambient manifold is positive we face a subtle issue because the manifold has a small volume. In that case, the meaning of optimality becomes less straightforward.

## 4 Results for subsets of the Euclidean space

### 4.1 Setting

We denote the closed ball in Euclidean space centred at a point  $p$  with radius  $r$  by  $B(p, r)$ .

The thickening of a set  $A \subseteq \mathbb{R}^d$  by parameter  $r > 0$  is denoted by  $A^{\boxplus r}$ , that is,

$$A^{\boxplus r} := \bigcup_{a \in A} B(a, r).$$

► **Remark 1.** We use the notation  $A^{\boxplus r}$  to remind the reader of the Minkowski sum. It is indeed true that in  $\mathbb{R}^d$ ,  $A^{\boxplus r} = A \oplus B(0, r)$ . However, the above notation is also well-defined for subsets of manifolds, whereas the Minkowski sum is not.

While working with subsets of the Euclidean space (Section 4 and [13, Appendix A]) we assume the following:

► **Universal assumption in the Euclidean setting 2.** We work with a closed set  $\mathcal{S} \subseteq \mathbb{R}^d$  with positive reach  $\text{rch}(\mathcal{S})$ , and let  $\mathcal{R} > 0$  be a constant satisfying  $\mathcal{R} \leq \text{rch}(\mathcal{S})$ . Furthermore, we consider a set  $P \subseteq \mathbb{R}^d$ , such that the one-sided Hausdorff distance from  $P$  to  $\mathcal{S}$  is at most  $\delta$ , and the one-sided Hausdorff distance from  $\mathcal{S}$  to  $P$  is at most  $\varepsilon$ . That is,

$$\mathcal{S} \subseteq P^{\boxplus \varepsilon} \quad \text{and} \quad P \subseteq \mathcal{S}^{\boxplus \delta}.$$

We assume that  $\delta, \varepsilon < \mathcal{R}$ . If the set  $\mathcal{S}$  is a submanifold of  $\mathbb{R}^d$ , we denote it by  $\mathcal{M}$ .

For most applications the assumption  $\delta \leq \varepsilon$  seems natural, but we do not need this. However, when  $\mathcal{S} = \mathcal{M}$ , we achieve better bounds when  $\delta \leq \varepsilon$ . See [13, Remark 29] for more details.

### 4.2 The geometric argument

We show that if the thickening  $P^{\boxplus r} = \bigcup_{p \in P} B(p, r)$  covers a sufficiently large thickening of  $\mathcal{S}$  – quantified by parameter  $\alpha$  – and the parameter  $r$  is not too big,  $P^{\boxplus r}$  deformation-retracts to  $\mathcal{S}$ .

We start by recalling that the normal cone at a point  $p$  of a set of positive reach is the set of directions such that if you move from  $p$  in that direction the closest point projection will remain  $p$ . For a definition we refer to [13, Definition 19].

► **Theorem 3.** Assume that a parameter  $\alpha > 0$  is small enough, so that the  $\alpha$ -neighbourhood  $\mathcal{S}^{\boxplus \alpha}$  of the set  $\mathcal{S}$  is contained in  $P^{\boxplus r}$ . In other words,

$$\mathcal{S}^{\boxplus \alpha} \subseteq P^{\boxplus r}. \tag{1}$$

If, moreover,

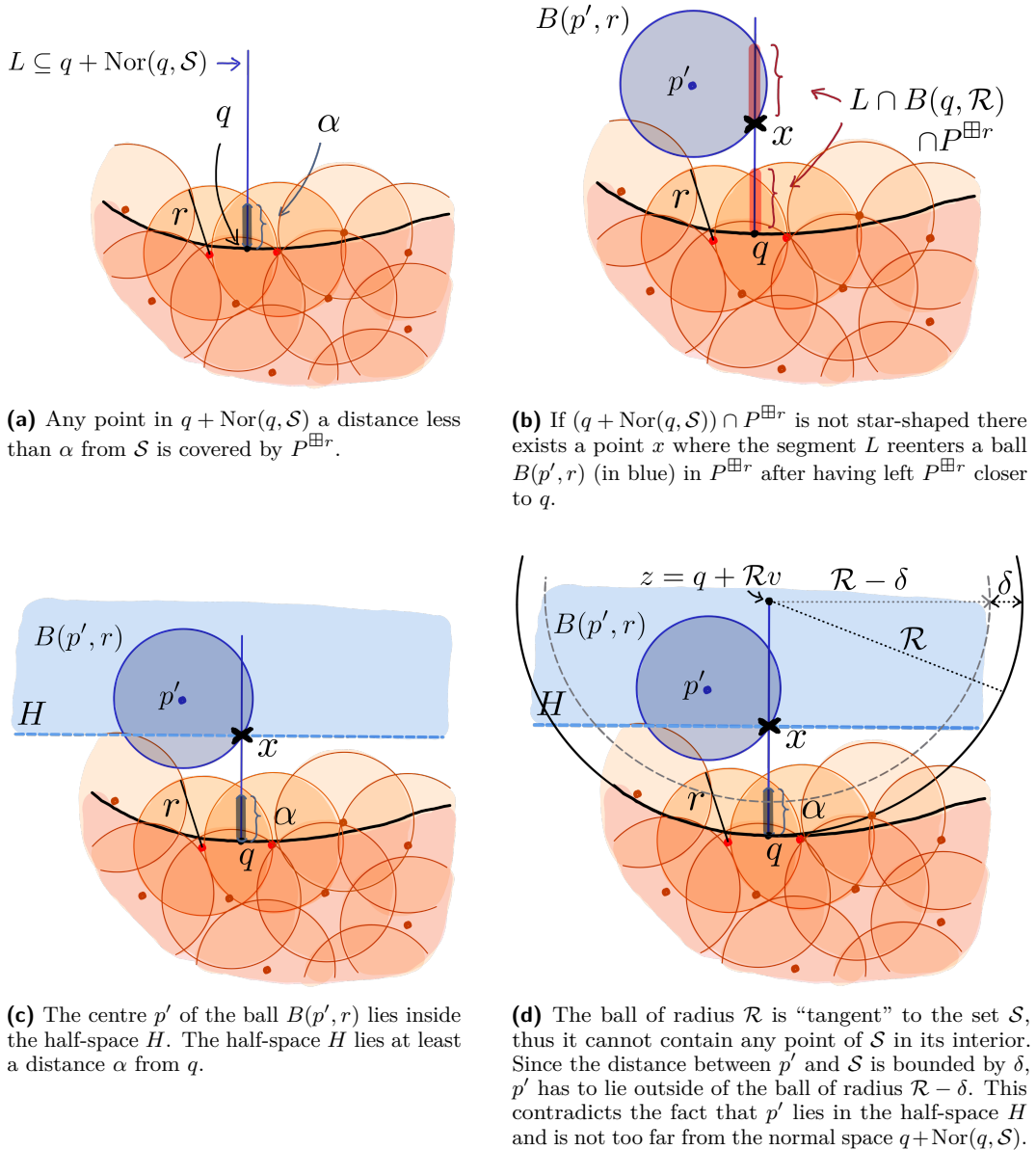
$$r^2 \leq (\mathcal{R} - \delta)^2 - (\mathcal{R} - \alpha)^2, \tag{2}$$

then, for any point  $q \in \mathcal{S}$ , the intersection  $(q + \text{Nor}(q, \mathcal{S})) \cap B(q, \mathcal{R}) \cap P^{\boxplus r}$  of the normal cone  $q + \text{Nor}(q, \mathcal{S})$ , the ball  $B(q, \mathcal{R})$ , and the union of balls  $P^{\boxplus r}$ , is star-shaped, with the point  $q$  as its “centre”. Furthermore,  $P^{\boxplus r}$  deformation-retracts onto  $\mathcal{S}$  along the closest point projection.



► Remark 4. The statement of Theorem 3 does not use the hypothesis  $\mathcal{S} \subseteq P^{\boxplus \varepsilon}$  from the Universal Assumption 2.

We refer to Figure 3 for a pictorial overview of the proof of Theorem 3. Further in the paper, we express the parameter  $\alpha$  in terms of  $r$  and the quality parameters  $\varepsilon$  and  $\delta$ . The expression differs depending on whether  $\mathcal{S}$  is a set or a manifold of positive reach. Inserting the appropriate expression into bound (2) yields the final bounds on  $\varepsilon$  and  $\delta$  (see Propositions 5 and 7).



■ **Figure 3** A pictorial overview of the proof. The pink shaded region represents a part of the set  $\mathcal{S}$ , the union of balls  $P^{\boxplus r}$  is coloured orange. The thickened blue segment shows those points of the segment  $L$  that lie a distance less than  $\alpha$  from  $\mathcal{S}$ . Per assumption, this segment is contained in the union of balls  $P^{\boxplus r}$ .

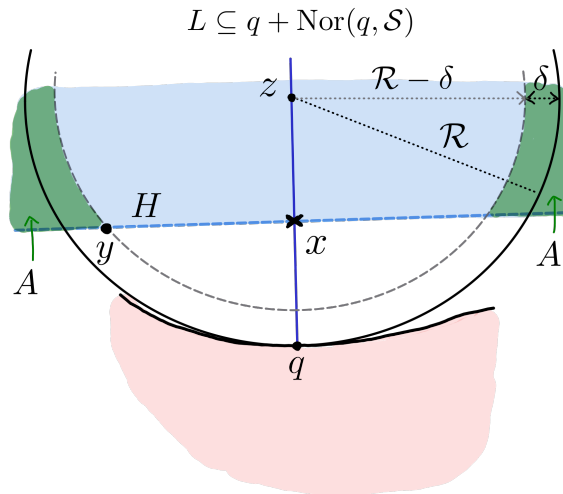
**Proof of Theorem 3.** We prove the claim by contradiction. For any point  $q \in \mathcal{S}$ , the set  $(q + \text{Nor}(q, \mathcal{S})) \cap (\mathcal{S}^{\boxplus\alpha})$  is contained in the union of balls  $P^{\boxplus r}$ . In Figure 3a, we illustrate this for the case where the set  $q + \text{Nor}(q, \mathcal{S})$  consists of one ray. Assume that there exists a point  $q \in \mathcal{S}$  and a vector  $v \in \text{Nor}(q, \mathcal{S})$ , with  $\|v\| = 1$ , such that the intersection of  $P^{\boxplus r}$  with the segment

$$L \stackrel{\text{def.}}{=} \{q + \lambda v \mid \lambda \in [0, \mathcal{R}]\}$$

consists of several connected components (as illustrated in Figure 3b). Thanks to Equation (1), the connected component that contains  $q$  has length at least  $\alpha$ . Let  $x$  be first point along  $L$ , seen from  $q$ , lying inside a connected component of  $(P^{\boxplus r}) \cap L$  that does not contain  $q$ . Then  $x$  lies at the intersection of the segment  $L$  and a ball  $B(p', r)$ , with  $p' \in P$ . We have  $\|x - q\| \geq \alpha$ . Furthermore, the point  $p'$  is contained in the open half-space  $H$  orthogonal to the vector  $v$ , that does not contain  $q$ , and whose boundary contains  $x$ . We stress that if  $p'$  lies on the boundary of  $H$  then the line  $L$  is tangent to the sphere  $\partial B(p', r)$ , which is still compatible with star-shapedness. The situation is illustrated in Figure 3c.

Let  $z = q + \mathcal{R}v$  be the open endpoint of  $L$ . Since, by Federer's result [42, Theorem 4.8 (12)] (recalled in [13, Theorem 22]), the intersection  $\mathcal{S} \cap B(z, \mathcal{R})^\circ$  is empty and the distance between  $p'$  and  $\mathcal{S}$  is bounded by  $\delta$ , we know that  $p' \notin B(z, \mathcal{R} - \delta)^\circ$ . Thus,

$$p' \in A \stackrel{\text{def.}}{=} H \cap (\mathbb{R}^d \setminus B(z, \mathcal{R} - \delta)^\circ).$$



■ **Figure 4** The centre of the ball creating a new connected component along one direction in the normal cone  $q + \text{Nor}(q, \mathcal{S})$  (in blue) is constrained to belong to the set  $A$  (in green). The set  $\mathcal{S}$  is coloured pink, the half-plane  $H$  in light blue.

The sphere  $\partial B(z, \mathcal{R} - \delta)$  has a non-empty intersection with the plane  $\partial H$ . Indeed, the sphere passes through point  $q + \delta v$  which does not belong to  $H$  while its centre  $z$  belongs to  $H$ ; see Figure 3d. We can thus pick a point  $y$  in the intersection  $\partial H \cap \partial B(z, \mathcal{R} - \delta)$ . By the Pythagorean theorem, the minimal squared distance between  $A$  and  $L$  is:

$$\inf_{\substack{a \in A \\ \ell \in L}} \|a - \ell\|^2 = \|x - y\|^2 = \|z - y\|^2 - (\|z - q\| - \|x - q\|)^2 \geq (\mathcal{R} - \delta)^2 - (\mathcal{R} - \alpha)^2,$$

## 11:10 Learning Homotopy in Euclidean Spaces and Riemannian Manifolds

as illustrated in Figure 4. Hence, if

$$r^2 \leq (\mathcal{R} - \delta)^2 - (\mathcal{R} - \alpha)^2, \quad (2)$$

the ball  $B(p', r)$  does not intersect  $L$ . Therefore,  $L \cap (P^{\boxplus r})$  cannot have more than one connected component. The set  $(q + \text{Nor}(q, \mathcal{S})) \cap B(q, \mathcal{R}) \cap (P^{\boxplus r})$  is thus star-shaped with centre  $q$ .

Since  $r$  satisfies Equation (2), we deduce that  $\delta + r < \mathcal{R}$ , and thus

$$P^{\boxplus r} \subseteq (\mathcal{S}^{\boxplus \mathcal{R}})^\circ.$$

Thanks to this, the fact that the set  $(q + \text{Nor}(q, \mathcal{S})) \cap B(q, \mathcal{R}) \cap (P^{\boxplus r})$  is star-shaped with centre  $q$ , and Federer's result [42, Theorem 4.8 (12)] (recalled in [13, Theorem 22]), the map

$$\begin{aligned} \mathcal{H} : [0, 1] \times (P^{\boxplus r}) &\rightarrow P^{\boxplus r}, \\ (t, x) &\mapsto (1 - t)x + t\pi_{\mathcal{S}}(x), \end{aligned}$$

is well-defined.

Furthermore, since  $\mathcal{S}$  has positive reach, then, thanks to Federer's theorem [42, Theorem 4.8 (12)] (recalled in [13, Theorem 21]), the projection  $\pi_{\mathcal{S}}$  is (Lipschitz) continuous. Thus, the map  $\mathcal{H}$  is a deformation retract from the union of balls  $P^{\boxplus r}$  to the set  $\mathcal{S}$ . ◀

In [13, Appendix E], we provide an alternative proof of Theorem 3, similar to an argument presented in [32].

### 4.3 Bounds on the sampling parameters

Recall that throughout the paper we assume the Universal Assumption 2. For sets of positive reach, we obtain the following bounds on the quality parameters  $\varepsilon$  and  $\delta$ :

► **Proposition 5.** *If  $\varepsilon$  and  $\delta$  satisfy*

$$\varepsilon + \sqrt{2}\delta \leq (\sqrt{2} - 1)\mathcal{R}, \quad (3)$$

*there exists a radius  $r > 0$  such that the union of balls  $P^{\boxplus r} = \bigcup_{p \in P} B(p, r)$  deformation-retracts onto  $\mathcal{S}$  along the closest point projection. In particular,  $r$  can be chosen as:*

$$r \in \left[ \frac{1}{2} (\mathcal{R} + \varepsilon - \sqrt{\Delta}), \frac{1}{2} (\mathcal{R} + \varepsilon + \sqrt{\Delta}) \right], \quad (4)$$

where  $\Delta = 2(\mathcal{R} - \delta)^2 - (\mathcal{R} + \varepsilon)^2$ .

► **Remark 6.** The interval for  $r$  as given in (4) can be slightly extended to

$$r \in \left[ \frac{1}{2} (\mathcal{R} + \varepsilon - \sqrt{\Delta}), \sqrt{\frac{1}{2}(\mathcal{R} - \delta)^2 + \frac{1}{2}(\mathcal{R} + \varepsilon)\sqrt{\Delta}} \right], \quad (5)$$

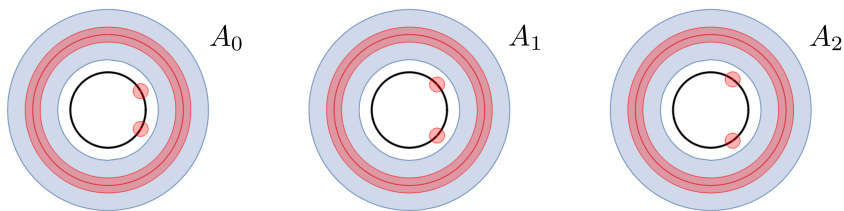
as we show in an alternative proof of Proposition 5 in [13, Appendix E]. It is not obvious that even this improved bound is tight.

If the set is a manifold, the bounds on  $\varepsilon$  and  $\delta$  can be improved as follows:

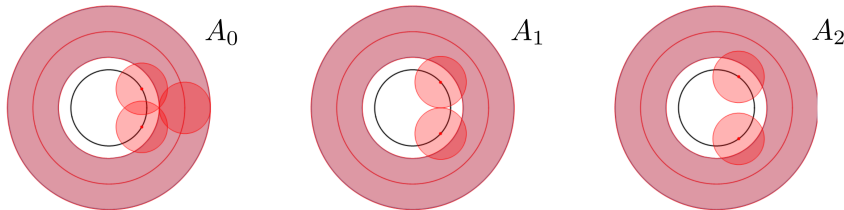
► **Proposition 7.** *If  $\varepsilon$  and  $\delta$  satisfy*

$$(\mathcal{R} - \delta)^2 - \varepsilon^2 \geq (4\sqrt{2} - 5) \mathcal{R}^2 \tag{6}$$

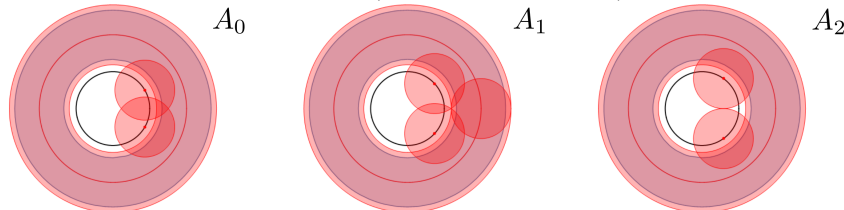
and  $\delta \leq \varepsilon$ , there exists a radius  $r > 0$  such that the union of balls  $P^{\boxplus r}$  deformation-retracts onto  $M$  along the closest point projection. The radius  $r$  can be chosen as in [13, equation (18)].



(a) At first, the thickening of the sample has three connected components per annulus. The thickening thus has three times as many connected components as the set  $\mathcal{S}$ .



(b) As the radius of the thickening grows, the connected components merge. However, at all times there exists an additional cycle at one of the annuli (annulus  $A_1$  in this case).



(c) At the moment when the cycle at annulus  $A_1$  vanishes, another cycle is formed at annulus  $A_2$ .

■ **Figure 5** A pictorial explanation of why  $P^{\boxplus r}$  never has the homotopy type of the set  $\mathcal{S}$ . We only depict three annuli in the sequence of  $A_i$ s. The set  $\mathcal{S}$  is in blue, the sample  $P$  in red, and the thickening of  $P$  in pink. The black circles indicate the location of the two isolated sample points of  $P$  associated to each annulus.

Both in Propositions 5 and 7, the interval for  $r$  tends to  $[0, \mathcal{R}]$  as  $\varepsilon$  and  $\delta$  tend to zero.

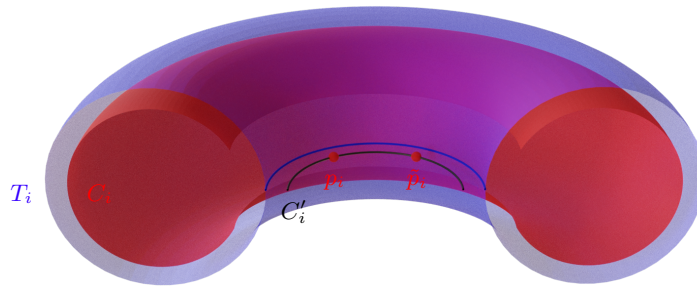
#### 4.4 Tightness of the bounds on the sampling parameters

Our sampling criteria for homotopy inference of sets of positive reach are tight in the following sense:

## 11:12 Learning Homotopy in Euclidean Spaces and Riemannian Manifolds

► **Proposition 8.** *Suppose that the dimension  $d$  of the ambient space  $\mathbb{R}^d$  satisfies  $d \geq 2$ , and the one-sided Hausdorff distances  $\varepsilon$  and  $\delta$  fail to satisfy bound (3). Then there exists a set  $\mathcal{S}$  of positive reach and a sample  $P$  that satisfy Universal Assumption 2, while the homology of the union of balls  $P^{\boxplus r}$  does not equal the homology of  $\mathcal{S}$  for any  $r$ .*

We construct the set  $\mathcal{S}$  and the sample  $P$  explicitly in  $\mathbb{R}^2$ . The set  $\mathcal{S}$  consists of a finite family of annuli  $A_i$ , the first three of which are depicted in Figure 5. The sample  $P$  is the union of a circle and two points for every annulus. In Figure 5, we illustrate that the thickening of the sample never captures the homotopy type of the set  $\mathcal{S}$ . The details of the construction and the proof of Proposition 8 are provided in [13, Section A.3.1].



■ **Figure 6** The (half of the) torus  $T_i$  depicted in blue; the sample – the set  $C_i$  and the points  $p_i$  and  $\tilde{p}_i$  – in red. In black we indicate the circle  $C'_i$  on which the points  $p_i$  and  $\tilde{p}_i$  lie. The closest point projection of this circle onto  $\mathcal{M}$  is indicated in blue.

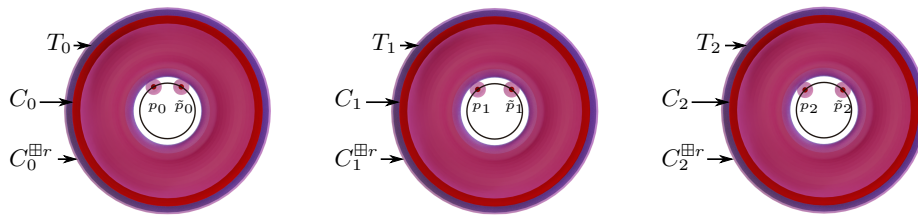
► **Proposition 9.** *Suppose that the dimension  $d$  of the ambient space  $\mathbb{R}^d$  satisfies  $d \geq 3$ , the one-sided Hausdorff distances  $\varepsilon$  and  $\delta$  fail to satisfy bound (6), and  $\delta \leq \varepsilon$ . Then there exists a manifold  $\mathcal{M}$  of positive reach and a sample  $P$  that satisfy Universal Assumption 2, while the homology of the union of balls  $P^{\boxplus r}$  does not equal the homology of  $\mathcal{M}$  for any  $r$ .*

We again construct the manifold  $\mathcal{M}$  and the sample  $P$  explicitly, this time in  $\mathbb{R}^3$ . The manifold  $\mathcal{M}$  is the union of a finite family of tori  $T_i$ . The sample  $P$  consists of one set  $C_i$  and one pair of points  $\{p_i, \tilde{p}_i\}$  for each torus  $T_i$ . The set  $C_i$  is constructed by taking a copy of  $T_i$ , decreasing the minor radius and cutting out a part close to the axis of revolution. We illustrate the manifold  $\mathcal{M} = \bigcup_i T_i$  together with the sample  $P = \bigcup_i C_i \cup \{p_i, \tilde{p}_i\}$  in Figure 6, and sketch why the underlying homology is not captured in Figure 7. The proof of Proposition 9 as well as details on the construction are provided in [13, Section A.3.2].

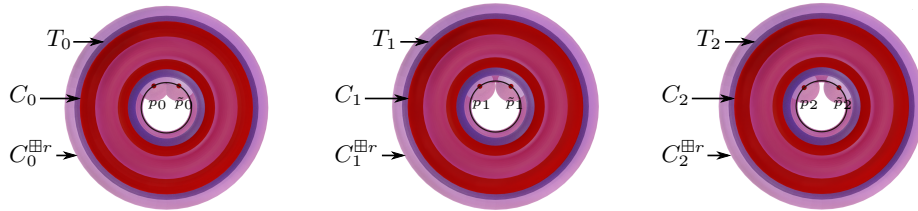
A video animating our construction has been submitted to the Media Exposition at Computational Geometry Week 2024 [12].

► **Remark 10.** For simplicity, the sets constructed, see Figures 7 and 5 (or [13, Example 31] and [13, Example 34] for details), are not connected. However, in each construction one can glue the connected components together in a way that preserves the reach, and the resulting examples still yield Propositions 8 and 9. See Figure 8 for a sketch of the modification needed.

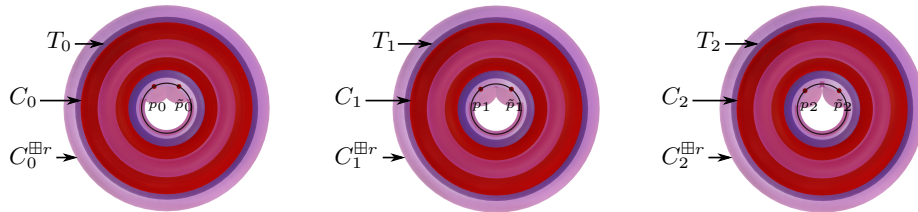
► **Remark 11.** Propositions 8 and 9 show that the bounds (3) and (6) are tight in (ambient) dimensions  $d \geq 2$ , resp.  $d \geq 3$ . We did not construct similar examples in lower dimensions. Nevertheless, our intuition is that, in these cases, the bounds (3) and (6) can be improved further.



(a) At first, the balls around the points  $p_i$  and  $\tilde{p}_i$  do not intersect the thickening of the set  $C_i$ , and thus the number of connected components of the thickening (in pink) of  $P$  is different from the number of components of the manifold.



(b) Then we create a (or possibly multiple) spurious cycle(s) for the first torus in the sequence (on the left).



(c) By the time the spurious cycles at the first torus have disappeared, others have been created at the second torus. This process is then repeated for all tori in the sequence as  $r$  increases.

■ **Figure 7** The construction for manifolds imitates the construction for general sets of positive reach as much as possible. The manifold  $\mathcal{M}$  is depicted in blue, the sample  $P$  in red, and the thickening in pink. We only display the part of objects below a horizontal clipping plane.



■ **Figure 8** The connected variants of our sets  $\mathcal{S}$  and  $\mathcal{M}$  are a topological disc with  $k$  holes and a genus  $k$  surface. On the left we sketch both the sample and the set of positive reach, on the right we only give the sample for the manifold setting because of visualization constraints.

## 5 Results for subsets of Riemannian manifolds

### 5.1 Setting

In the second part of this paper we consider subsets of a  $(C^2)$  Riemannian manifold  $\mathcal{N}$ . In this Riemannian setting we denote (geodesic) balls with radius  $r > 0$  centred at a point  $p \in \mathcal{N}$  by  $B(p, r)$ , and write  $A^{\boxplus r} = \bigcup_{a \in A} B(a, r)$  for the union of (geodesic) balls of radius  $r$  centred at a subset  $A \subseteq \mathcal{N}$ . Similarly, the one-sided Hausdorff distance from  $X \subseteq \mathcal{N}$  to  $Y \subseteq \mathcal{N}$  is defined as the smallest  $\rho$  such that the union of (geodesic) balls of radius  $\rho$  centred at  $X$  covers  $Y$ .

## 11:14 Learning Homotopy in Euclidean Spaces and Riemannian Manifolds

To be able to proceed as in the Euclidean setting and state tight bounds on the sampling parameters, we need a notion of the reach in the Riemannian setting. To this end, we introduce a new definition, inspired by the cut locus (which is defined for example in [18]):

► **Definition 12** (Cut locus). *Given a closed subset  $\mathcal{S} \subseteq \mathcal{N}$ , the cut locus of  $\mathcal{S}$  is the set  $\text{cl}_{\mathcal{N}}(\mathcal{S})$  of points  $p \in \mathcal{N}$  for which there are at least 2 geodesics of minimal length from  $p$  to some point in  $\mathcal{S}$ .*

► **Definition 13** (Cut locus reach). *The cut locus reach  $\text{rch}_{\mathcal{N}}^{\text{cl}}(\mathcal{S})$  of a closed set  $\mathcal{S} \subseteq \mathcal{N}$  is the infimum of distances between  $\mathcal{S}$  and its cut locus  $\text{cl}_{\mathcal{N}}(\mathcal{S})$ :*

$$\text{rch}_{\mathcal{N}}^{\text{cl}}(\mathcal{S}) \stackrel{\text{def.}}{=} \inf_{\substack{p \in \mathcal{S}, \\ q \in \text{cl}_{\mathcal{N}}(\mathcal{S})}} d_{\mathcal{N}}(p, q).$$

Our definition is a refinement of the notion used by Bangert and Kleinjohann [16, 53, 54], as well as the reach defined in [24]. We explain why our new definition is appropriate for the learning of topological features in [13, Appendix F]. Using the new extension of the reach we assume the following conditions, which resemble the ones in the Euclidean setting closely:

► **Universal assumption in the Riemannian setting 14.** *We work with a closed set  $\mathcal{S} \subseteq \mathcal{N}$  with positive cut locus reach  $\text{rch}_{\mathcal{N}}^{\text{cl}}(\mathcal{S})$ , and let  $\mathcal{R} > 0$  be a constant satisfying  $\mathcal{R} \leq \text{rch}_{\mathcal{N}}^{\text{cl}}(\mathcal{S})$ . Furthermore, we consider a set  $P \subseteq \mathcal{N}$ , such that the one-sided Hausdorff distance from  $P$  to  $\mathcal{S}$  is at most  $\delta$ , and the one-sided Hausdorff distance from  $\mathcal{S}$  to  $P$  is at most  $\varepsilon$ . That is,  $\mathcal{S} \subseteq P^{\boxplus \varepsilon}$  and  $P \subseteq \mathcal{S}^{\boxplus \delta}$ . We assume that  $\delta, \varepsilon < \mathcal{R}$ . We also assume that the sectional curvatures of the manifold  $\mathcal{N}$  are lower bounded by a constant  $\Lambda_{\ell} \in \mathbb{R}$ . When  $\Lambda_{\ell} > 0$  and  $\mathcal{S} = \mathcal{M}$  is a manifold, we can safely assume, thanks to [13, Lemma 62], that  $\mathcal{R} \leq \frac{\pi}{2\sqrt{\Lambda_{\ell}}}$ .*

This assumption is used in Section 5 and [13, Appendix B].

### 5.2 Bounds on the sampling parameters

Also in the Riemannian setting we provide (tight) bounds that the sample  $P$  needs to satisfy in order to be able to infer homotopy. For sets of positive (cut locus) reach, we obtain the following bounds on  $\varepsilon$  and  $\delta$ :

► **Proposition 15.** *If  $\varepsilon$  and  $\delta$  satisfy*

$$\begin{aligned} 2 \cos\left(\sqrt{\Lambda_{\ell}}(\mathcal{R} - \delta)\right) - \cos\left(\sqrt{\Lambda_{\ell}}(\mathcal{R} + \varepsilon)\right) &\leq 1 && \text{if } \Lambda_{\ell} > 0, \\ \sqrt{2}(\mathcal{R} - \delta) - (\mathcal{R} + \varepsilon) &\leq 0 && \text{if } \Lambda_{\ell} = 0, \\ 2 \cosh\left(\sqrt{|\Lambda_{\ell}|}(\mathcal{R} - \delta)\right) - \cosh\left(\sqrt{|\Lambda_{\ell}|}(\mathcal{R} + \varepsilon)\right) &\geq 1 && \text{if } \Lambda_{\ell} < 0, \end{aligned} \tag{7}$$

*there exists a radius  $r > 0$  such that the union of balls  $P^{\boxplus r}$  deformation-retracts onto  $\mathcal{S}$  along the closest point projection. In particular,  $r$  can be chosen as:*

$$r = \frac{1}{2}(\mathcal{R} + \varepsilon). \tag{8}$$

If the set is a manifold, the bounds on  $\varepsilon$  and  $\delta$  can be improved as follows:

► **Proposition 16.** Let  $\tilde{x} = \sqrt{|\Lambda_\ell|}x$ . For  $\delta \leq \varepsilon$  satisfying

$$(2 \cos \tilde{\varepsilon} \cos \tilde{\mathcal{R}} - 3 \cos (\tilde{\mathcal{R}} - \tilde{\delta}))^2 \leq \left( \frac{\cos \tilde{\varepsilon} - \cos (\tilde{\mathcal{R}} - \tilde{\delta}) \cos \tilde{\mathcal{R}}}{\sin \tilde{\mathcal{R}}} \right)^2 + \cos^2 (\tilde{\mathcal{R}} - \tilde{\delta})$$

if  $\Lambda_\ell > 0$ , (9)

$$(\mathcal{R} - \delta)^2 - \varepsilon^2 \geq (4\sqrt{2} - 5) \mathcal{R}^2$$

if  $\Lambda_\ell = 0$ , (6)

and

$$2 \cosh \tilde{\varepsilon} \cosh \tilde{\mathcal{R}} \leq 3 \cosh (\tilde{\mathcal{R}} - \tilde{\delta})$$

$$\cosh^2 (\tilde{\mathcal{R}} - \tilde{\delta}) \leq \left( \frac{\cosh \tilde{\varepsilon} - \cosh (\tilde{\mathcal{R}} - \tilde{\delta}) \cosh \tilde{\mathcal{R}}}{\sinh \tilde{\mathcal{R}}} \right)^2 + (2 \cosh \tilde{\varepsilon} \cosh \tilde{\mathcal{R}} - 3 \cosh (\tilde{\mathcal{R}} - \tilde{\delta}))^2$$

if  $\Lambda_\ell < 0$ , (10)

there exists a radius  $r > 0$  such that  $P^{\boxplus r}$  deformation-retracts onto  $\mathcal{M}$  along the (geodesic) closest point projection  $\pi_{\mathcal{M}}$ . The interval from which  $r$  can be chosen can be recovered from [13, equations (42), (18), and (45)] respectively.

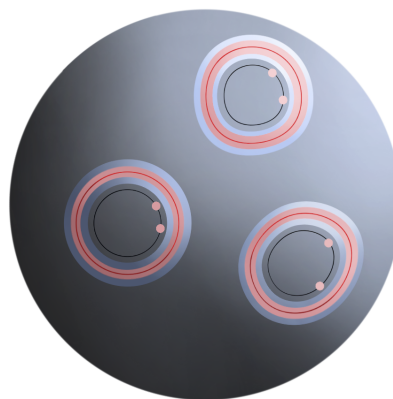
The computation of Čech complexes in a Riemannian manifold can be difficult (depending on the manifold). Fortunately, we can avoid this step and still recover the homology:

► **Remark 17.** The results of Chazal and co-authors [33] on the interleaving between the Čech and Rips complexes extend to the Riemannian setting. By combining their results with the results of this paper, one can recover the homology type of a subset of positive reach of a Riemannian manifold using persistent homology of Rips complexes.

The Rips complex is easier to calculate than the Čech complex, since the calculation only involves distances between pairs of points.

### 5.3 Tightness of the bounds on the sampling parameters

We exhibit the tightness of the bounds on  $\varepsilon$  and  $\delta$  from Propositions 15 and 16 by constructions of examples in (simply connected) spaces of constant curvature. These constructions are similar to the Euclidean setting – they also consist of annuli and tori, see Figure 9. However, due to the curvature of the ambient manifold, the proof of the tightness of the bounds is significantly more involved (see [13, Appendix B.4]).



■ **Figure 9** The construction for sets of positive reach on a manifold with (constant) positive curvature (the sphere). For a detailed version of the figure see [13, Figure 23].



## 6 Future work

This article leaves several important questions unanswered. We mention three.

First of all, we consider the union of balls centered on a sample  $P$  whose homotopy type is equal to that of the Čech complex of  $P$  and, when the ambient space is a Riemannian manifold, the radius of balls is smaller than the convexity radius.

It would be interesting to see if our work would help understanding the same question for Rips complexes. For related work see e.g. [6, 7, 50, 55].

Second, we consider sets embedded in Riemannian manifolds whose sectional curvature is lower bounded. A natural question is under which conditions do our results generalize to a larger class of metric spaces with lower bounded curvatures.

The generalized gradient of the distance function and its flow have been used to generalize results on subsets of positive reach in Euclidean space to subsets with positive  $\mu$ -reach and weak feature size [27, 28, 31, 33]. Our work on the cut locus reach makes it possible to extend the notations of positive  $\mu$ -reach and weak feature size to Riemannian manifolds. It is expected that many of the main results from the Euclidean setting still hold with minor modifications in this more general context.

---

### References

- 1 Eddie Aamari, Catherine Aaron, and Clément Levrard. Minimax boundary estimation and estimation with boundary, 2021. doi:10.48550/arXiv.2108.03135.
- 2 Eddie Aamari, Clément Berenfeld, and Clément Levrard. Optimal reach estimation and metric learning, 2022. doi:10.48550/arXiv.2207.06074.
- 3 Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electronic Journal of Statistics*, 13(1):1359–1399, 2019. doi:10.1214/19-EJS1551.
- 4 Eddie Aamari and Alexander Knop. Statistical query complexity of manifold estimation. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 116–122, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451135.
- 5 Eddie Aamari and Clément Levrard. Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. *Discrete & Computational Geometry*, 59:923–971, 2018.
- 6 Michał Adamaszek and Henry Adams. The Vietoris–Rips complexes of a circle. *Pacific Journal of Mathematics*, 290(1):1–40, 2017.
- 7 Michał Adamaszek, Henry Adams, and Samadwara Reddy. On Vietoris–Rips complexes of ellipses. *Journal of Topology and Analysis*, 11(03):661–690, 2019. doi:10.1142/S1793525319500274.
- 8 Paolo Albano and Piermarco Cannarsa. Structural properties of singularities of semiconcave functions. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 28(4):719–740, 1999.
- 9 Paolo Albano, Piermarco Cannarsa, Khai T Nguyen, and Carlo Sinestrari. Singular gradient flow of the distance function and homotopy equivalence. *Mathematische Annalen*, 356(1):23–43, 2013.
- 10 Nina Amenta, Sunghee Choi, Tamal K Dey, and Naveen Leekha. A simple algorithm for homeomorphic surface reconstruction. In *Proceedings of the sixteenth annual symposium on Computational geometry*, pages 213–222, 2000.
- 11 Aleksander Antasik. Sampling  $C^1$ -submanifolds of  $\mathbb{H}^n$ . In *Colloquium Mathematicum*, volume 168, pages 211–228. Instytut Matematyczny Polskiej Akademii Nauk, 2022.

- 12 Dominique Attali, Hana Dal Poz Kouřimská, Christopher Fillmore, Ishika Ghosh, André Lieutier, Elizabeth Stephenson, and Mathijs Wintraecken. The ultimate frontier: An optimality construction for homotopy inference. In Xavier Goaoc and Michael Kerber, editors, *Medial exposition 40th International Symposium on Computational Geometry (SoCG 2024)*, Leibniz International Proceedings in Informatics (LIPIcs), Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- 13 Dominique Attali, Hana Dal Poz Kouřimská, Christopher Fillmore, Ishika Ghosh, André Lieutier, Elizabeth Stephenson, and Mathijs Wintraecken. Tight bounds for the learning of homotopy à la Niyogi, Smale, and Weinberger for subsets of Euclidean spaces and of Riemannian manifolds. *arXiv preprint*, 2024. [arXiv:2206.10485](https://arxiv.org/abs/2206.10485).
- 14 Dominique Attali and André Lieutier. Reconstructing shapes with guarantees by unions of convex sets. 33 pages, December 2009. URL: <https://hal.archives-ouvertes.fr/hal-00427035v2>.
- 15 Dominique Attali, André Lieutier, and David Salinas. Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry*, 46(4):448–465, 2013. 27th Annual Symposium on Computational Geometry (SoCG 2011). doi:10.1016/j.comgeo.2012.02.009.
- 16 Victor Bangert. Sets with positive reach. *Archiv der Mathematik*, 38(1):54–57, 1982.
- 17 Clément Berenfeld, John Harvey, Marc Hoffmann, and Krishnan Shankar. Estimating the reach of a manifold via its convexity defect function. *Discrete & Computational Geometry*, 67(2):403–438, 2022.
- 18 Marcel Berger. *A panoramic view of Riemannian geometry*. Springer, 2003.
- 19 Matthew Berger, Andrea Tagliasacchi, Lee M Seversky, Pierre Alliez, Gael Guennebaud, Joshua A Levine, Andrei Sharf, and Claudio T Silva. A survey of surface reconstruction from point clouds. In *Computer Graphics Forum*, volume 36, pages 301–329. Wiley Online Library, 2017.
- 20 Anders Björner. Topological methods. *handbook of combinatorics*, vol. 1, 2, 1819–1872, 1995.
- 21 Jean-Daniel Boissonnat. Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics (TOG)*, 3(4):266–286, 1984.
- 22 Jean-Daniel Boissonnat. Shape reconstruction from planar cross sections. *Computer vision, graphics, and image processing*, 44(1):1–29, 1988.
- 23 Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and Topological Inference*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2018. doi:10.1017/9781108297806.
- 24 Jean-Daniel Boissonnat and Mathijs Wintraecken. The reach of subsets of manifolds. *Journal of Applied and Computational Topology*, pages 1–23, 2023.
- 25 P. Buser and H. Karcher. *Gromov’s almost flat manifolds*, volume 81 of *Astérisque*. Société mathématique de France, 1981.
- 26 Isaac Chavel. *Riemannian geometry: a modern introduction*, volume 98. Cambridge university press, 2006.
- 27 F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in Euclidean space. *Discrete and Computational Geometry*, 41(3):461–479, 2009.
- 28 F. Chazal and A. Lieutier. The  $\lambda$ -medial axis. *Graphical Models*, 67(4):304–331, 2005.
- 29 Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for measures based on distance functions. *Foundations of computational mathematics*, 11(6):733–751, 2011.
- 30 Frédéric Chazal, Leonidas J Guibas, Steve Y Oudot, and Primoz Skraba. Persistence-based clustering in Riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):1–38, 2013.
- 31 Frédéric Chazal and André Lieutier. Weak feature size and persistent homology: computing homology of solids in  $\mathbb{R}^n$  from noisy data samples. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 255–262, 2005.
- 32 Frédéric Chazal and André Lieutier. Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Computational Geometry*, 40(2):156–170, 2008.

- 33 Frédéric Chazal and Steve Yann Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 232–241, 2008.
- 34 Jeff Cheeger and David G. Ebin. *Comparison Theorems in Riemannian Geometry*, volume 365. American Mathematical Soc., 2008.
- 35 Alejandro Cholaquidis, Ricardo Fraiman, and Leonardo Moreno. Universally consistent estimation of the reach. *Journal of Statistical Planning and Inference*, 225:110–120, 2023.
- 36 Frank H. Clarke. *Optimization and Nonsmooth Analysis*, volume 5 of *Classics in applied mathematics*. SIAM, 1990.
- 37 Ryan Cotsakis. Computable bounds for the reach and  $r$ -convexity of subsets of  $\mathbb{R}^d$ . *Discrete & Computational Geometry*, pages 1–37, 2024.
- 38 Tamal Krishna Dey and Yusu Wang. *Computational topology for data analysis*. Cambridge University Press, 2022.
- 39 Herbert Edelsbrunner. Alpha shapes—a survey. In *Tessellations in the Sciences: Virtues, Techniques and Applications of Geometric Tilings*. Springer, 2011.
- 40 Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- 41 Herbert Edelsbrunner and Nimish R Shah. Triangulating topological spaces. In *Proceedings of the tenth annual symposium on Computational geometry*, pages 285–292, 1994.
- 42 H. Federer. Curvature measures. *Transactions of the American mathematical Society*, 93:418–491, 1959.
- 43 Charles Fefferman, Sergei Ivanov, Yaroslav Kurylev, Matti Lassas, and Hariharan Narayanan. Fitting a putative manifold to noisy data. In *Conference On Learning Theory*, pages 688–720. PMLR, 2018.
- 44 Charles Fefferman, Sergei Ivanov, Matti Lassas, and Hariharan Narayanan. Fitting a manifold of large reach to noisy data. *arXiv preprint*, 2019. [arXiv:1910.05084](https://arxiv.org/abs/1910.05084).
- 45 Charles Fefferman, Sergei Ivanov, Matti Lassas, and Hariharan Narayanan. Reconstruction of a Riemannian manifold from noisy intrinsic distances. *SIAM Journal on Mathematics of Data Science*, 2(3):770–808, 2020.
- 46 Alvina Goh and René Vidal. Clustering and dimensionality reduction on Riemannian manifolds. In *2008 IEEE Conference on computer vision and pattern recognition*, pages 1–7. IEEE, 2008.
- 47 Detlef Gromoll, Wilhelm Klingenberg, and Wolfgang Meyer. *Riemannsche Geometrie im Großen*. Lecture Notes in Mathematics. Springer, 1975. [doi:10.1007/BFb0079185](https://doi.org/10.1007/BFb0079185).
- 48 Jihun Hamm and Daniel D Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proceedings of the 25th international conference on Machine learning*, pages 376–383, 2008.
- 49 Mehrtash T Harandi, Conrad Sanderson, Sareh Shirazi, and Brian C Lovell. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching. In *CVPR 2011*, pages 2705–2712. IEEE, 2011.
- 50 Jean-Claude Hausmann et al. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. *Annals of Mathematics Studies*, 138:175–188, 1995.
- 51 H. Karcher. Riemannian comparison constructions. In S.S. Chern, editor, *Global Differential Geometry*, pages 170–222. The mathematical association of America, 1989.
- 52 Jisu Kim, Jaehyeok Shin, Frédéric Chazal, Alessandro Rinaldo, and Larry Wasserman. Homotopy Reconstruction via the Čech Complex and the Vietoris-Rips Complex. In Sergio Cabello and Danny Z. Chen, editors, *36th International Symposium on Computational Geometry (SoCG 2020)*, volume 164 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 54:1–54:19, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. [doi:10.4230/LIPIcs.SoCG.2020.54](https://doi.org/10.4230/LIPIcs.SoCG.2020.54).
- 53 Norbert Kleinjohann. Convexity and the unique footpoint property in Riemannian geometry. *Archiv der Mathematik*, 35(1):574–582, 1980.

- 54 Norbert Kleinjohann. Nächste Punkte in der Riemannschen Geometrie. *Mathematische Zeitschrift*, 176(3):327–344, 1981.
- 55 Janko Latschev. Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold. *Archiv der Mathematik*, 77(6):522–528, 2001.
- 56 David Levin. The approximation power of moving least-squares. *Mathematics of computation*, 67(224):1517–1531, 1998.
- 57 André Lieutier. Any open bounded subset of  $\mathbb{R}^n$  has the same homotopy type as its medial axis. *Computer-Aided Design*, 36(11):1029–1046, 2004. Solid Modeling Theory and Applications. doi:10.1016/j.cad.2004.01.011.
- 58 Alexander Lytchak. On the geometry of subsets of positive reach. *manuscripta mathematica*, 115(2):199–205, 2004.
- 59 Alexander Lytchak. Almost convex subsets. *Geometriae Dedicata*, 115(1):201–218, 2005.
- 60 P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.
- 61 Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of computer vision*, 66:41–66, 2006.
- 62 Robert Pless and Richard Souvenir. A survey of manifold learning for images. *IPSN Transactions on Computer Vision and Applications*, 1:83–94, 2009.
- 63 Jan Rataj and Martina Zähle. *Curvature measures of singular sets*. Springer, 2019.
- 64 Jan Rataj and Luděk Zajíček. On the structure of sets with positive reach. *Mathematische Nachrichten*, 290(11-12):1806–1829, 2017. doi:10.1002/mana.201600237.
- 65 Yogesh Rathi, Allen Tannenbaum, and Oleg Michailovich. Segmenting images on the tensor manifold. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- 66 Barak Sober and David Levin. Manifold approximation by moving least-squares projection (MMLS). *Constructive Approximation*, 52(3):433–478, 2020.
- 67 Oncel Tuzel, Fatih Porikli, and Peter Meer. Region covariance: A fast descriptor for detection and classification. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006. Proceedings, Part II 9*, pages 589–600. Springer, 2006.
- 68 Raviteja Vemulapalli and David W Jacobs. Riemannian metric learning for symmetric positive definite matrices. *arXiv preprint*, 2015. arXiv:1501.02393.
- 69 Yuan Wang and Bei Wang. Topological inference of manifolds with boundary. *Computational Geometry*, 88:101606, 2020. doi:10.1016/j.comgeo.2019.101606.
- 70 Xikang Zhang, Yin Wang, Mengran Gou, Mario Sznaiar, and Octavia Camps. Efficient temporal sequence comparison and classification using gram matrix embeddings on a riemannian manifold. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4498–4507, 2016.
- 71 Pengfei Zhu, Hao Cheng, Qinghua Hu, Qilong Wang, and Changqing Zhang. Towards generalized and efficient metric learning on riemannian manifold. In *International Joint Conference on Artificial Intelligence*, pages 3235–3241, 2018.