



HAL
open science

Optimal homotopy reconstruction results à la Niyogi, Smale, and Weinberger

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.. Optimal homotopy reconstruction results à la Niyogi, Smale, and Weinberger. 2022. hal-03721463v1

HAL Id: hal-03721463

<https://hal.science/hal-03721463v1>

Preprint submitted on 12 Jul 2022 (v1), last revised 7 Mar 2024 (v2)

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.

Optimal homotopy reconstruction results à la Niyogi, Smale, and Weinberger

Dominique Attali ✉

Université 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]

André Lieutier ✉

No affiliation

Elizabeth Stephenson

IST Austria
[Klosterneuburg, Austria]

Mathijs Wintraecken ✉ 

IST Austria
[Klosterneuburg, Austria]

Abstract

In this article we show that the proof of the homotopy reconstruction result by Niyogi, Smale, and Weinberger can be streamlined considerably using Federer’s work on the reach and several geometric observations. While Niyogi, Smale, and Weinberger restricted themselves to C^2 manifolds (\mathcal{M}) with positive reach, our proof extends to sets (\mathcal{S}) of positive reach.

The sample we consider does not have to lie directly on the set of positive reach. Instead, we assume that the two one-sided Hausdorff distances (δ and ε) — between the sample P to the set \mathcal{S} , are bounded. We provide explicit bounds in terms of ε and δ , that guarantee that there exists a parameter r such that the union of balls $\bigcup_{p \in P} B(p, r)$ deformation retracts to \mathcal{S} . We provide even better bounds for the manifold case. In both cases, our bounds improve considerably on the state-of-the-art in almost all settings. In fact the bounds are optimal.

2012 ACM Subject Classification Theory of computation → Computational geometry

Keywords and phrases Homotopy, Inference, Sets of positive reach

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.

Mathijs Wintraecken: Supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 754411, and the Austrian science fund (FWF) grant No. M-3073

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

1 Introduction

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

These questions were first motivated by the shape reconstruction of 3-dimensional objects. There, the partial geometric information was represented by a finite, in general noisy, set of points obtained from photogrammetric or lidar measurements [1, 4, 6, 7, 11].

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 [8, 10, 12]. Assuming the support is a manifold, one calls this process manifold learning [21].

In [20], Niyogi, Smale, and Weinberger showed that, given a C^2 manifold of positive reach and a sufficiently dense point sample on (or near) the manifold, the union of balls of certain radii centered on the point sample captures the homotopy type of the manifold. By the nerve theorem, the homotopy type of the union of balls is shared by the Čech complex of the point sample [5, 13]. From the Čech complex we can then infer the topological information such as the homology groups of the underlying manifold.

Niyogi, Smale, and Weinberger’s homotopy reconstruction result has led to generalizations including [3, 9, 15, 24].

Related work and contribution Niyogi, Smale, and Weinberger [20] seem to have been unaware of Federer’s theory of sets of positive reach [14]. In this article we discuss the extension of Niyogi, Smale, and Weinberger’s result to sets of positive reach and use Federer’s theory to streamline their proof (for both the manifold case and general sets of positive reach). We recall the relevant results from Federer’s work [14] in the appendix.

Attali *et al.*, Chazal *et al.*, and Kim *et al.* studied even more general settings [3, 9, 15]. Their proofs are, however, different from ours, and (necessarily) more involved.

While our proof is short and elementary, we provide optimal bounds on the distance between the manifold (or the set) of positive reach and its sample. The best previously known bounds (for some particular cases) can be found in [20] and [3], respectively.

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 R be a non-negative real number such that $R \leq \text{rch}(\mathcal{S})$ (resp. $R \leq \text{rch}(\mathcal{M})$). We denote the one-sided Hausdorff distance¹ from P to \mathcal{M} (resp. \mathcal{S}) by δ , and the one-sided Hausdorff distance from \mathcal{M} (resp. \mathcal{S}) to P by ε . Distinguishing the two one-sided Hausdorff distances seems natural to the authors, because in measurements one would expect the measurement error δ (with the exception of some small number of outliers) to be often smaller than the sampling density ε . Similar assumptions seem to be common in the learning community, see e.g. [16]. Niyogi, Smale, and Weinberger made similar assumptions on the support of the measure from which they sampled. We achieved the following conditions on ε and δ which, if satisfied, guarantee the existence of a radius $r > 0$ such that the union of balls $\bigcup_{p \in P} B(p, r)$ deformation-retracts onto \mathcal{M} (resp. \mathcal{S}):

- If \mathcal{S} has positive reach, we achieved the condition $\varepsilon + \sqrt{2}\delta \leq (\sqrt{2} - 1)R$.

¹ We recall that the one sided Hausdorff distance from X to Y , denoted by $d_H^o(X; Y)$, is the smallest ρ such that the union of balls of radius ρ centred at X covers Y .

- If moreover $\mathcal{S} = \mathcal{M}$ is a manifold, we achieved the condition $(R - \delta)^2 - \varepsilon^2 \geq (4\sqrt{2} - 5) R^2$ (in this case we assume that $\delta \leq \varepsilon \leq R$).

The set of pairs (ε, δ) that satisfy the conditions depicted in Figure 1.

In this article we only consider samples for which we have precise bounds on ε and δ . In [20], 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.

We stress that in [8, 20, 24], the authors use $\varepsilon/2$ instead of our ε .

As shown in Section 4, our conditions on ε and δ 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 $\bigcup_{p \in P} B(p, r)$ has the same homology as \mathcal{S} (resp. \mathcal{M}). In the case of 1-dimensional manifolds, we expect the optimal condition to be weaker.

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 favorably with the previously best known constant $\frac{1}{2}\sqrt{\frac{3}{5}}$ from [20] for manifolds.

In Proposition 7.1 of [20] one encounters the condition $\varepsilon < (3 - \sqrt{8})R$, for a particular case of the family of settings we consider, namely when the sampling condition is expressed through an upper bound ε on the Hausdorff distance, in other words $\delta = \varepsilon$ in our setting. The same constant $3 - \sqrt{8}$ was proven independently in [2, Theorem 4] for general sets of positive reach. While the optimality was not previously proven, this constant is in fact optimal when $\delta = \varepsilon$ both for general sets with positive reach and for manifolds. To contrast the two related results in [20], for $\delta = 0$ and $\delta = \varepsilon$ respectively, with our bounds, we portray them as black dots in Figure 1, that depicts our sufficient and necessary conditions on the pair (ε, δ) .

The case of manifolds with boundary has been studied in [24, Theorem 3.2], assuming lower bounds on both the reach of the manifold and the reach of its boundary. We improve also on this result by considering a manifold with boundary as a particular case of a set with positive reach, while our bounds only depend on the reach of the set itself and not the one of its boundary.

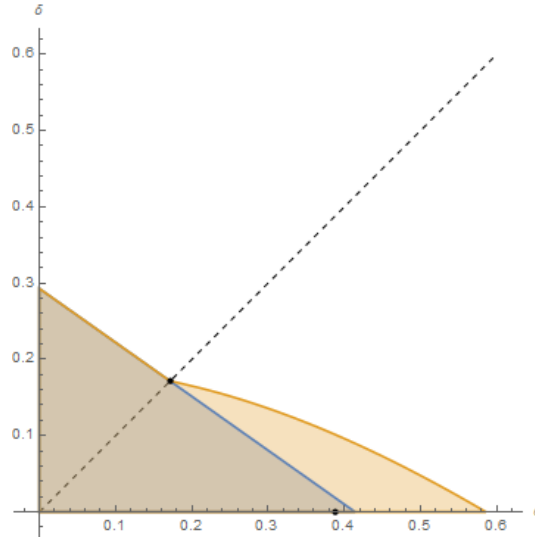
Outline, setting, and notation In Section 2, we reprove — and extend — the homotopy reconstruction result of Niyogi, Smale, and Weinberger. In Section 3, we determine the bounds on the quality parameters ε and δ of the sample. Finally, in Section 4, we show that these bounds are optimal.

We denote the closed ball in a Euclidean space centred at a point p with radius r by $B(p, r)$. Open balls are denoted by $B(p, r)^\circ$. If no centre is indicated we tacitly assume that the centre is at 0. The Minkowski sum of two sets A and B is denoted by $A \oplus B$. The normal cone of a closed set \mathcal{S} at a point $q \in \mathcal{S}$ is denoted by $\text{Nor}(q, \mathcal{S})$ (see also Definition 21). The closest point projection on a set \mathcal{S} is denoted by $\pi_{\mathcal{S}}$.

► **Universal Assumption 1.** *Throughout this article, we work with a closed set $\mathcal{S} \subseteq \mathbb{R}^d$ with positive reach $\text{rch}(\mathcal{S})$, and let $R > 0$ be a constant satisfying $R \leq \text{rch}(\mathcal{S})$. Furthermore, we consider a set $P \subseteq \mathbb{R}^d$, such that the one-sided Hausdorff distance from P to \mathcal{S} is δ , and the one-sided Hausdorff distance from \mathcal{S} to P by ε . That is,*

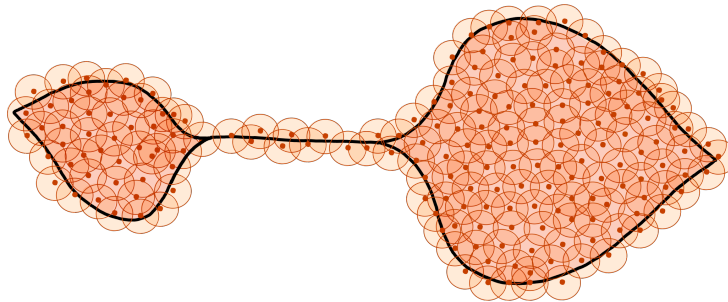
$$\mathcal{S} \subseteq P \oplus B(\varepsilon) \quad \text{and} \quad P \subseteq \mathcal{S} \oplus B(\delta).$$

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



■ **Figure 1** In blue-gray we depict the region in (ϵ, δ) -space for which there exists a radius r such that the union of balls $P \oplus B(r)$ captures the homotopy type of a set of positive reach. We do the same in yellow for a manifold. We stress that for $\delta \geq \epsilon$ the fact that a set of positive reach is also a manifold does not lead to better bounds. We have rescaled such that the reach is 1. The black points indicate the bounds that were known to Niyogi, Smale and Weinberger.

For most applications the assumption $\delta \leq \epsilon$ seems to be natural, but we do not need this for our analysis.



■ **Figure 2** A set \mathcal{S} of positive reach with a sample P . The set $\bigcup_{p \in P} B(p, r) = P \oplus B(r)$ is also indicated.

2 The geometric argument

In this section we show that if the union of balls $P \oplus B(r)$ covers a sufficiently large neighbourhood of \mathcal{S} , and the parameter r is not too large, then $P \oplus B(r)$ deformation-retracts to \mathcal{S} .

► **Theorem 2.** *Assume that a parameter $\alpha > 0$ is small enough, so that the α -neighbourhood $\mathcal{S} \oplus B(\alpha)$ of the set \mathcal{S} is contained in the union of balls $P \oplus B(r)$. In other words,*

$$\mathcal{S} \oplus B(\alpha) \subseteq P \oplus B(r). \quad (1)$$

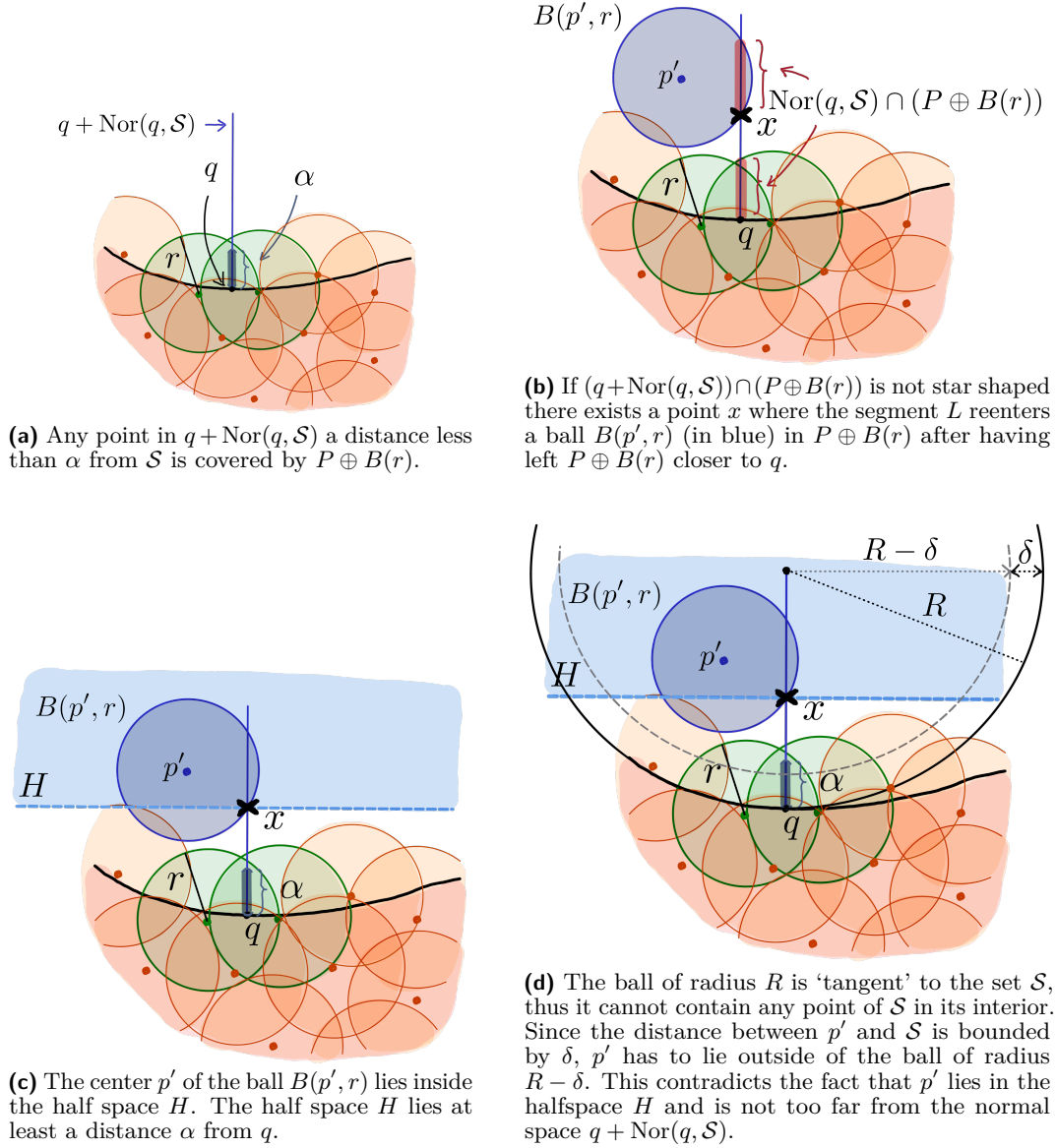
If, moreover,

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

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

► Remark 3. The statement of Theorem 2 does not use the hypothesis $\mathcal{S} \subseteq P \oplus B(\varepsilon)$ from the universal assumption.

We refer to Figure 3 for a pictorial overview of the proof of Theorem 2.



■ Figure 3 A pictorial overview of the proof.

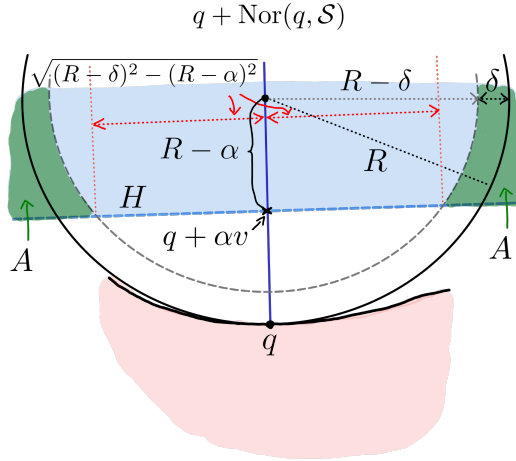
Proof of Theorem 2. We prove the claim by contradiction. For any point $q \in \mathcal{S}$, the set $(q + \text{Nor}(q, \mathcal{S})) \cap (\mathcal{S} \oplus B(\alpha))$ is contained in the union of balls $P \oplus B(r)$, as illustrated in Figure 3a. Assume that there exists a point $q \in \mathcal{S}$ and a vector $v \in \text{Nor}(q, \mathcal{S})$, with $\|v\| = 1$, such that the intersection of $P \oplus B(r)$ with the segment

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

consists of several connected components (as illustrated in Figure 3b). Thanks to Equation (1), the connected component that contains q has length at least α . Let x be first point along L , seen from q , lying inside a connected component of $(P \oplus B(r)) \cap L$ that does not contain q . Then x lies at the intersection of the segment L and a ball $B(p', r)$, with $p' \in P$. 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 $z = q + Rv$ be the open endpoint of L . Since, by [14, Theorem 4.8 (12)], the intersection $\mathcal{S} \cap B(z, R)^\circ$ is empty and the distance between p' and \mathcal{S} is bounded by δ , we know that $p' \notin B(z, R - \delta)^\circ$. Thus,

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



■ **Figure 4** The center of the ball creating a new connected component along the normal cone $q + (q, \mathcal{S})$ is constrained to belong to the set A .

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

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

as illustrated in Figure 4. Hence, if

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

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

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

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

Thanks to this, the fact that the set $(q + \text{Nor}(q, \mathcal{S})) \cap B(q, R) \cap (P \oplus B(r))$ is star-shaped with centre q , and [14, Theorem 4.8 (12)], the map

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

is well-defined.

Furthermore, thanks to [14, Theorem 4.8 (8)], the projection $\pi_{\mathcal{S}}$ is continuous, and even Lipschitz. Thus, the map \mathcal{H} is a deformation retract from the union of balls $P \oplus B(r)$ to the set \mathcal{S} . ◀

In Appendix B, we provide an alternative proof of Theorem 2, similar to an argument presented in [11].

3 Bounds

In this section we first compute the bounds on the size α of the neighbourhood $\mathcal{S} \oplus B(\alpha)$ covered by the union of balls $P \oplus B(r)$ in terms of ε , δ , and r . We then combine these bounds with Equation (2) to infer (optimal) upper bounds on ε and δ , for which there exists a radius r such that the deformation retract from $P \oplus B(r)$ to \mathcal{S} is possible. We do so first for sets of positive reach and then for manifolds. Somewhat counter-intuitively, it turns out to be easier to determine the bounds for sets of positive reach.

3.1 Sets of positive reach

For sets of positive reach, the bound on α is almost trivial. Nevertheless, it is tight, as we will see in Section 4.

► **Lemma 4.** *Suppose that $\mathcal{S} \subseteq P \oplus B(\varepsilon)$ for some $\varepsilon \geq 0$. Then, for all $\alpha \leq r - \varepsilon$, the α -neighbourhood $\mathcal{S} \oplus B(\alpha)$ of \mathcal{S} is contained in the finite union of balls $P \oplus B(r)$. That is,*

$$\mathcal{S} \oplus B(\alpha) \subseteq P \oplus B(r).$$

Proof. Indeed, $\mathcal{S} \oplus B(\alpha) \subseteq (P \oplus B(\varepsilon)) \oplus B(\alpha) \subseteq P \oplus B(\varepsilon + \alpha) \subseteq P \oplus B(r)$. ◀

From this, we derive the bounds on ε and δ (in terms of R).

► **Proposition 5.** *If ε and δ satisfy*

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

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

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

where

$$\Delta = 2(R - \delta)^2 - (R + \varepsilon)^2.$$

Proof. We combine the bound from Lemma 4 with the conditions of Theorem 2. More precisely, inserting $\alpha = r - \varepsilon$ in Equation (2) yields that

$$r^2 + (R - r + \varepsilon)^2 \leq (R - \delta)^2. \quad (4)$$

Using the abc-formula for quadratic equations, this is equivalent to

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

where

$$\Delta = 2\delta^2 + R^2 - 4\delta R - 2R\varepsilon - \varepsilon^2 = 2(R - \delta)^2 - (R + \varepsilon)^2$$

is the discriminant. This interval is non-empty if the discriminant is non-negative, that is, if $\varepsilon + \sqrt{2}\delta \leq (\sqrt{2} - 1)R$. \blacktriangleleft

► **Remark 6.** The parameter δ is not necessarily smaller than ε , even if this would be natural in most applications.

3.2 Manifolds with positive reach

In this section, we show that the bounds from Proposition 5 can be improved further if the set of positive reach is a manifold. In Lemma 4, we used the triangle inequality to set $\alpha = r - \varepsilon$. If \mathcal{S} is a manifold, however, the parameter α can be increased using more subtle arguments than the triangle inequality: Manifolds with positive reach are $C^{1,1}$ smooth², i.e., differentiable with Lipschitz derivative. Moreover, Federer's normal cone $\text{Nor}(q, \mathcal{M})$ (Definition 21) coincides with the 'classical' normal space $N_q\mathcal{M}$ of an n -dimensional submanifold \mathcal{M} of \mathbb{R}^d at every point $q \in \mathcal{M}$. The tangent and normal cones of manifolds of positive reach are n - and $(d - n)$ -dimensional linear spaces, respectively, that are not only dual, but also orthogonal.

In Lemma 7, we establish a lower bound for the parameter α in the case that $\mathcal{S} = \mathcal{M}$ is a manifold. This bound is tight, as we will see in Section 4.

► **Lemma 7.** *Suppose that $\mathcal{M} \subseteq P \oplus B(\varepsilon)$ for some $\varepsilon \geq 0$. Then, for any $r \geq \alpha \geq 0$ satisfying*

$$r^2 \geq \alpha^2 + \frac{\alpha}{R} (R^2 + \varepsilon^2 - (R - \delta)^2) + \varepsilon^2, \quad (5)$$

the α -neighbourhood $\mathcal{M} \oplus B(\alpha)$ of \mathcal{M} is contained in the finite union of balls $P \oplus B(r)$. That is,

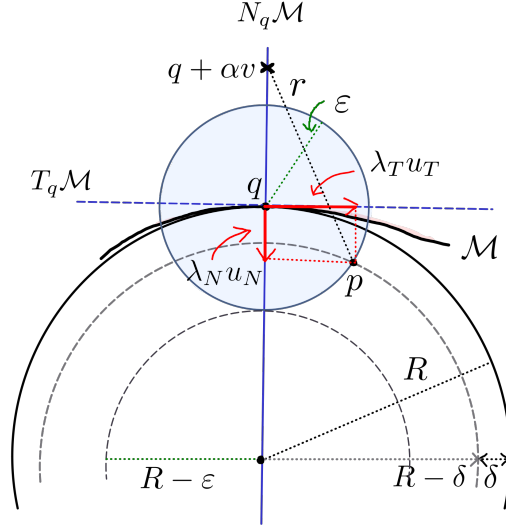
$$\mathcal{M} \oplus B(\alpha) \subseteq P \oplus B(r).$$

Proof. Given a point $q \in \mathcal{M}$, the tangent cone $T_q\mathcal{M}$ and the normal cone $N_q\mathcal{M}$ are orthogonal vector spaces satisfying $T_q\mathcal{M} \oplus N_q\mathcal{M} = \mathbb{R}^d$. Since $\mathcal{M} \subseteq P \oplus B(\varepsilon)$, the intersection $P \cap B(q, \varepsilon)$ is non-empty. Let $p \in P \cap B(q, \varepsilon)$.

The vector $p - q$ decomposes uniquely as

$$p - q = \lambda_T u_T + \lambda_N u_N,$$

² Topologically embedded manifolds with positive reach are $C^{1,1}$ embedded [14, 17, 18, 19, 22, 23].



■ **Figure 5** Overview of the notation used in the proof of Lemma 7.

with $u_T \in T_q \mathcal{M}$, $u_N \in N_q \mathcal{M}$, $\|u_T\| = \|u_N\| = 1$, and $\lambda_T, \lambda_N \geq 0$ (see Figure 5). Since $\|p - q\| \leq \varepsilon$,

$$\lambda_T^2 + \lambda_N^2 \leq \varepsilon^2. \quad (6)$$

Thanks to [14, Theorem 4.8 (12)] (Theorem 24 in Appendix A), the sets \mathcal{M} and $B(q + Ru_N, R)^\circ$ do not intersect, and thus:

$$p \notin B(q + Ru_N, R - \delta)^\circ.$$

Hence, $(p - q - Ru_N)^2 \geq (R - \delta)^2$. Applying the decomposition of $p - q$ we obtain

$$(\lambda_T u_T + (\lambda_N - R)u_N)^2 \geq (R - \delta)^2,$$

which implies that

$$\lambda_T^2 + (R - \lambda_N)^2 \geq (R - \delta)^2.$$

Combining this result with Equation (6) implies that

$$\varepsilon^2 - \lambda_N^2 + (R - \lambda_N)^2 \geq (R - \delta)^2,$$

which can be rewritten as

$$2R\lambda_N \leq R^2 + \varepsilon^2 - (R - \delta)^2. \quad (7)$$

Choose a vector $v \in \text{Nor}(q, \mathcal{M})$ with $\|v\| = 1$, and let $\alpha \geq 0$. Then,

$$\begin{aligned} (p - (q + \alpha v))^2 &= ((\lambda_N u_N - \alpha v) + \lambda_T u_T)^2 \\ &= (\lambda_N u_N - \alpha v)^2 + (\lambda_T u_T)^2 \\ &\leq (\lambda_N + \alpha)^2 + \lambda_T^2 \\ &\leq (\lambda_N + \alpha)^2 + \varepsilon^2 - \lambda_N^2 \\ &= \alpha^2 + 2\alpha\lambda_N + \varepsilon^2. \end{aligned} \quad (\text{by (6)})$$

Using inequality (7) to substitute $2\lambda_N$, we further obtain:

$$(p - (q + \alpha v))^2 \leq \alpha^2 + \frac{\alpha}{R} (R^2 + \varepsilon^2 - (R - \delta)^2) + \varepsilon^2.$$

Thus, if

$$r^2 \geq \alpha^2 + \frac{\alpha}{R} (R^2 + \varepsilon^2 - (R - \delta)^2) + \varepsilon^2,$$

then the point $q + \alpha v$ lies in $B(p, r) \subseteq P \oplus B(r)$. Since this inclusion holds for any $q \in \mathcal{M}$ and $v \in N_q \mathcal{M}$, $\mathcal{M} \oplus B(\alpha) \subseteq P \oplus B(r)$. \blacktriangleleft

As in Proposition 5, we now derive a bound on ε .

► **Proposition 8.** *If ε and δ satisfy*

$$(R - \delta)^2 - \varepsilon^2 \geq (4\sqrt{2} - 5) R^2 \tag{8}$$

and $\delta \leq \varepsilon$, there exists a radius $r > 0$ such that the union of balls $P \oplus B(r)$ deformation-retracts onto \mathcal{M} along the closest point projection. In particular, r can be chosen as in (11).

The bound is illustrated in Figure 1.

Proof. We combine the bound from Lemma 7 with the conditions of Theorem 2. More precisely, combining Equations (2) and (5) yields the following sufficient condition for $L \cap (P \oplus B(r))$ to be connected:

$$\alpha^2 + \frac{\alpha}{R} (R^2 + \varepsilon^2 - (R - \delta)^2) + \varepsilon^2 \leq r^2 \leq (R - \delta)^2 - (R - \alpha)^2. \tag{9}$$

The inequality between leftmost and rightmost members of (9), which needs to be satisfied for a non-empty range of values for r to exist, can be rearranged as:

$$0 \geq \varepsilon^2 - (R - \delta)^2 + R^2 + \alpha \frac{1}{R} (\varepsilon^2 - R^2 - (R - \delta)^2) + 2\alpha^2.$$

Using the abc-formula for quadratic equations, the above inequality is satisfied if $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, with

$$\begin{aligned} \alpha_{\min} &= \frac{1}{4} \left(\frac{(R - \delta)^2 + R^2 - \varepsilon^2}{R} - \sqrt{\Delta} \right), \\ \alpha_{\max} &= \frac{1}{4} \left(\frac{(R - \delta)^2 + R^2 - \varepsilon^2}{R} + \sqrt{\Delta} \right), \end{aligned} \tag{10}$$

where the discriminant Δ is

$$\Delta = \frac{1}{R^2} (\varepsilon^2 - (R - \delta)^2)^2 - 10 (\varepsilon^2 - (R - \delta)^2) - 7R^2.$$

The discriminant can be viewed as a polynomial in $y = \varepsilon^2 - (R - \delta)^2$. Solving $\Delta(y) = 0$ with respect to y yields $y = R^2 (5 \pm 4\sqrt{2})$. This in turn implies that Δ is non-negative if either $\varepsilon^2 - (R - \delta)^2 \leq R^2 (5 - 4\sqrt{2})$ or $\varepsilon^2 - (R - \delta)^2 \geq R^2 (5 + 4\sqrt{2})$. Thanks to Assumption 1, we are only interested in the case where $\varepsilon, \delta < R$, and thus we can ignore the second inequality. Hence the interval $[\alpha_{\min}, \alpha_{\max}]$ is non-empty if

$$\varepsilon^2 - (R - \delta)^2 \leq (5 - 4\sqrt{2}) R^2.$$

Plugging the bounds on α as given in (10) into Equations (2) and (5) yields

$$\left(1 + \frac{\alpha_{\min}}{R}\right) \varepsilon^2 + \alpha_{\min}^2 + \frac{\alpha_{\min}}{R} (R^2 - (R - \delta)^2) \leq r^2 \leq (R - \delta)^2 - (R - \alpha_{\max})^2. \quad (11)$$

◀

► **Remark 9.** We restricted ourselves to the case where $\delta \leq \varepsilon$, because if $\delta > \varepsilon$, the fact that the set is a set of positive reach is a manifold [instead of an arbitrary set] no longer helps. The geometric reason for this is that p in Figure 5 may lie in $N_q\mathcal{M}$.

4 Optimality

In this section we prove that the bounds provided in Section 3 are optimal in the following sense:

► **Proposition 10.** *Suppose that the dimension d of the ambient space \mathbb{R}^d satisfies $d \geq 2$, and the one-sided Hausdorff distances ε and δ do not satisfy the bound (3). Then there exists a set \mathcal{S} of positive reach and a sample P that satisfy Universal Assumption 1, while the homology of the union of balls $P \oplus B(r)$ does not equal the homology of \mathcal{S} for any r .*

► **Proposition 11.** *Suppose that the dimension d of the ambient space \mathbb{R}^d satisfies $d \geq 3$, the one-sided Hausdorff distances ε and δ do not satisfy the bound (8), and $\delta \leq \varepsilon$. Then there exists a manifold \mathcal{M} of positive reach and a sample P that satisfy Universal Assumption 1, while the homology of the union of balls $P \oplus B(r)$ does not equal the homology of \mathcal{M} for any r .*

To prove Propositions 10 and 11, we construct the set \mathcal{S} , the manifold \mathcal{M} , and the corresponding samples in Examples 15 and 16, respectively. Due to rescaling it suffices to construct sets of reach $R = 1$.

► **Remark 12.** The dimension of the set in both examples will be two-dimensional. As mentioned in the introduction, we expect that better bounds can be obtained for one-dimensional sets, i.e. curves, possibly with boundary.

► **Remark 13.** When $\delta \geq \varepsilon$, which in Figure 1 corresponds to the area above the diagonal $\delta = \varepsilon$, the same bound is optimal whether the set is assumed manifold or not. Indeed, in this case the union of annuli \mathcal{S} in Example 15 can be replaced by a union of circles, namely the inner boundaries of the annuli. Thus, the bound is tight for manifolds, including one-dimensional submanifolds in \mathbb{R}^2 .

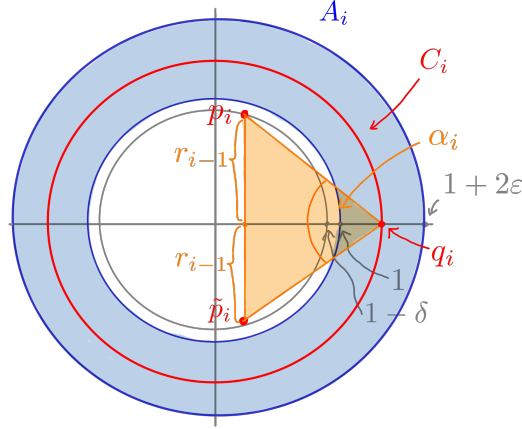
► **Remark 14.** For simplicity, the examples 15 and 16 are not connected. However, one could easily cut up our examples and glue the pieces together in a way that preserves the reach so that resulting examples still yield Propositions 10 and 11.

4.1 Sets of positive reach

The construction of the set proving Proposition 10 goes as follows:

► **Example 15.** We define \mathcal{S} to be the disjoint union of annuli A_i in \mathbb{R}^2 , each of which has inner radius 1 and outer radius $1 + 2\varepsilon$, and lies a distance at least 2 away from the other. We number the annuli from $i = 1$. A bound on their total number will be discussed later. The sample P consists of circles C_i of radius $1 + \varepsilon$ lying in the middle of the annuli ($C_i \subseteq A_i$), and a discrete set of points.

We describe the positioning of the discrete set of points in an inductive manner. Each pair of points, p_i and \tilde{p}_i , lies inside the annulus A_i , at a distance δ from A_i , and a distance $2r_{i-1}$ from each other. We define $r_0 = \frac{\delta + \varepsilon}{2}$. The distance r_i is defined as follows: Let q_i be the intersection point of the bisector of p_i and \tilde{p}_i , and the circle C_i . Then r_i is the circumradius of the triangle $p_i\tilde{p}_iq_i$. The situation is illustrated in Figure 6.



■ **Figure 6** Situation in one annulus of Example 15.

Denoting the angle at q_i in the triangle $p_i\tilde{p}_iq_i$ by α_i , we can express the inductive step as

$$r_{i+1} = \frac{r_i}{\sin \alpha_i},$$

as a consequence of the law of sines. It is important to observe that, since $r_{i+1} \geq r_i$ for each i , the sequence of angles satisfies $\alpha_{i+1} \leq \alpha_i$. The construction thus necessarily stops if either $r_{i+1} = r_i$ (i.e., $\alpha_i = \pi/2$) or $r_{i+1} > 1 - \delta$. In the former case, p_{i+1} and \tilde{p}_{i+1} coincide with p_i and \tilde{p}_i , respectively. In the latter, it is not possible to place p_{i+1} and \tilde{p}_{i+1} in the annulus according to our conditions. The latter case is reached after finitely many iterations if $r_{i+1} - r_i$ is lower bounded, or equivalently if α_i is bounded away from $\pi/2$. The proof of Proposition 10 focuses on establishing this bound.

Proof of Proposition 10. Assume that δ and ε fail to satisfy the bound (3), that is,

$$\sqrt{2}(1 - \delta) - 1 < \varepsilon. \quad (12)$$

We first show that each triangle $p_i\tilde{p}_iq_i$ is acute, that is, that $\alpha_i < \pi/2$. This is the case precisely when the height of the triangle, $1 + \varepsilon - \sqrt{(1 - \delta)^2 - r_i^2}$, is larger than r_i . This is a quadratic inequality in r_i ,

$$r_i < 1 + \varepsilon - \sqrt{(1 - \delta)^2 - r_i^2} \iff 0 < (1 + \varepsilon)^2 - (1 - \delta)^2 - 2(1 + \varepsilon)r_i + 2r_i^2.$$

The inequality holds for all r_i if

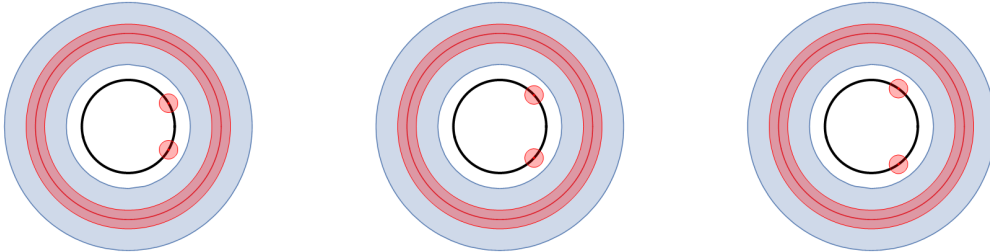
$$0 > 4(2(1 - \delta)^2 - (1 + \varepsilon)^2),$$

which is equivalent to bound (12). In particular, \mathcal{S} consists of finitely many annuli.

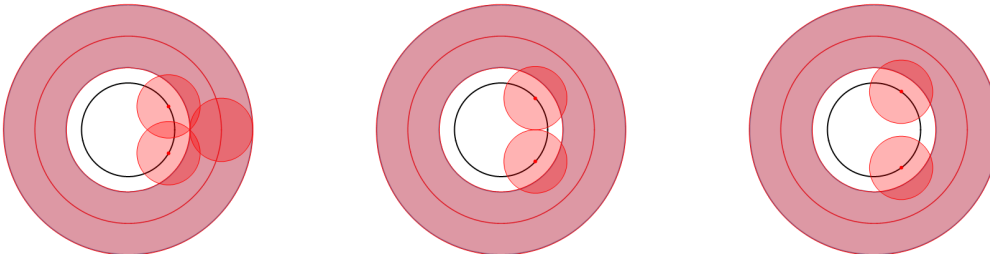
For any $r \in [0, 1 - \delta)$, the union of balls $P \oplus B(r)$ has different homology than the set \mathcal{S} . Indeed,

- for $r \in [0, r_0)$, each set $(C_i \cup \{p_i, \tilde{p}_i\}) \oplus B(r)$ has three connected components,
- for $r \in [r_{i-1}, r_i)$, the set $(C_i \cup \{p_i, \tilde{p}_i\}) \oplus B(r)$ has the homotopy type of two circles that share a point, that is, a wedge of two circles, also known as a bouquet.

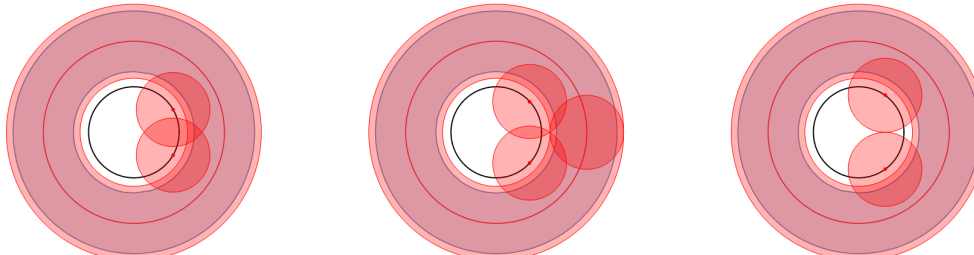
The situation is illustrated in Figure 7. At the same time, each annulus $A_i \subseteq \mathcal{S}$ has the homotopy type of a circle.



(a) For all $r \leq r_0$ the offset of the sample $(C_i \cup \{p_i, \tilde{p}_i\}) \oplus B(r)$ has three connected components for each of the three annuli.



(b) At the radius r_1 the cycle in the thickening of the sample of the first annulus dies, while at the same radius a cycle is created for the second annulus.



(c) At the radius r_2 the cycle in the thickening of the sample of the second annulus dies, while at the same radius a cycle is created for the third annulus.

■ **Figure 7** In the figure we only depict three annuli A_i sorted from left to right.



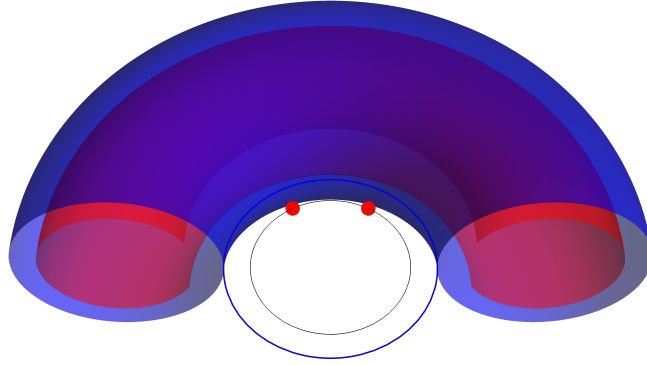
4.2 Manifolds

The construction of the manifold proving Proposition 11 goes as follows:

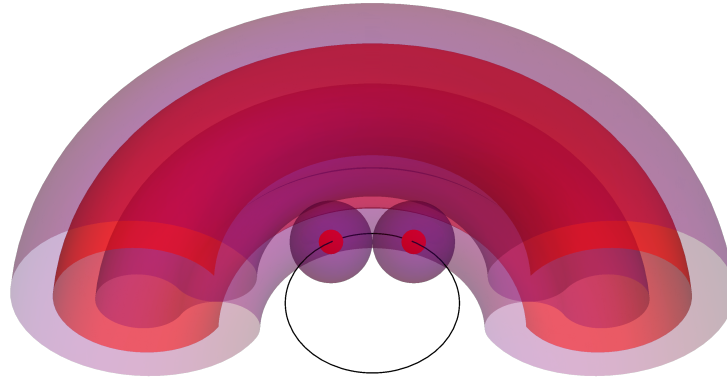
► **Example 16.** We define \mathcal{M} to be a finite union of tori of revolution T_i in \mathbb{R}^3 . Each of these tori is the 1-offset of a circle of radius 2 in \mathbb{R}^3 . Put differently, each T_i is — up to Euclidean transformations — the surface of revolution around the y -axis of a circle of radius 1 in the xz -plane, centred at $(2, 0, 0)$. In order for the reach of \mathcal{M} to be equal to 1, we assume that the tori lie at a distance at least 2 apart from each other. The set T_i is illustrated in Figure 8 (in blue).

The sample P consists of tori with parts cut out, denoted by C_i , and a discrete set of points. We construct each set C_i by taking a δ -offset of the T_i on its interior. In other words, we

consider a surface of revolution around the y -axis of a circle of radius $1 - \delta$ in the xz -plane centred at $(2, 0, 0)$, with an ε -neighbourhood of the circle found by revolving the point $(1, 0, 0)$ around the y -axis cut out. For the positioning of the discrete set of points, we let $C_{i,\delta}$ be the circle found by revolving the point $(1 - \delta, 0, 0)$ around the y -axis.



■ **Figure 8** The (half of the) torus T_i depicted in blue and the sample the set C_i together with the points p_i and \tilde{p}_i in red. In black we indicate the circle $C_{i,\delta}$. The closest point projection of this circle onto \mathcal{M} is indicated in blue.

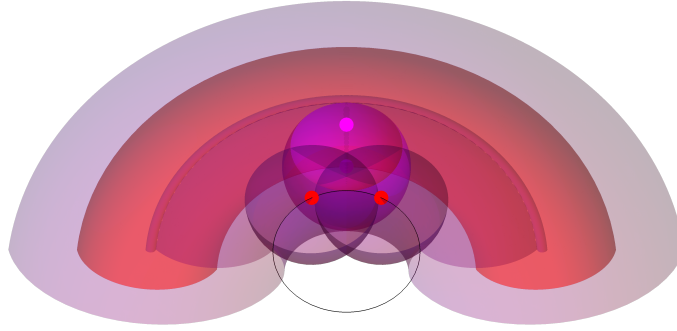


■ **Figure 9** We depict (half of) the sample P in red and its offset $P \oplus B(r)$ in purple.

As in Example 15, we describe the positioning of the discrete points in an inductive manner. Each pair of points, p_i and \tilde{p}_i , lies on $C_{i,\delta}$ at a distance $2r_{i-1}$ from each other. We define

$$r_0 = \frac{\sqrt{\delta^2 + \varepsilon^2 + \delta(\varepsilon^2 - \delta^2 + 2\delta)}}{2} = \frac{\sqrt{\varepsilon^2(1 + \delta) + \delta^2(3 - \delta)}}{2}.$$

The distance r_0 is chosen such that the balls $B(p_1, r)$ and $B(\tilde{p}_1, r)$ start to intersect at the same value of r as these balls start to intersect $C_1 \oplus B(r)$. The computation of this distance is based on Figure 11.



■ **Figure 10** We depict the points p and \tilde{p} in red as well as their offset $\{p, \tilde{p}\} \oplus B(r)$ in purple, the points q and \tilde{q} as well as their offset $\{q, \tilde{q}\} \oplus B(r)$ are indicated in magenta.

The consecutive r_i 's are defined as follows: Let q_i and \tilde{q}_i be the two points in the intersection of the bisector of p_i and \tilde{p}_i and the set C_i that lie closest to p_i and \tilde{p}_i . Then we define r_i to be the radius of the tetrahedron with vertices $p_i \tilde{p}_i q_i \tilde{q}_i$. Note that $\{q_1, \tilde{q}_1\} = \pi_{C_1} \left(\frac{p_1 + \tilde{p}_1}{2} \right)$. We continue this construction inductively until $r_i > 1 - \delta$.

In the proof of Proposition 10 acuteness played an essential role. We argued that if

$$\sqrt{2}(1 - \delta) - 1 < \varepsilon, \quad (12)$$

then any triangle $p_i \tilde{p}_i q_i$ (see Example 15) is acute.

A triangle is acute if and only if it contains its circumcentre. For the manifold case we therefore generalize acuteness to tetrahedra in the following manner:

► **Definition 17** (Self-centred simplices). *A simplex is called self-centred if it contains its circumcentre.*

▷ **Claim 18.** Let $C = C_i$ and $C_\delta = C_{i,\delta}$ be as defined in Example 16. Let $p, \tilde{p} \in C_\delta$, and define q, \tilde{q} to be the two points in the intersection of the bisector of p, \tilde{p} and C_δ that are closest to p (and thus \tilde{p}). If

$$\varepsilon^2 > (6 - 4\sqrt{2}) - 2\delta + \delta^2, \quad (13)$$

then the simplex $p\tilde{p}q\tilde{q}$ is self-centred.

In the proofs of Claim 18 and Proposition 11 we need a number of distances, namely $r_0 = \frac{w}{2}$, $r_{-1} = h$, and w' , as indicated in Figure 11. It is convenient to calculate these in one go. We first focus on the first two. The sample P contains sets C_i which are tori with a band cut out, r_{-1} denotes half the size of the gap. As we will discuss in the proof of Proposition 11 there is a topological transition at this radius. The distance w' is important in establishing self-centredness. From Figure 11 we immediately see that

$$1 = \ell_1 + \ell_2 \quad (14)$$

$$\ell_1^2 + h^2 = \varepsilon^2 \quad (15)$$

$$\ell_2^2 + h^2 = (1 - \delta)^2 \quad (16)$$

$$w^2 = (\delta + \ell_1)^2 + h^2. \quad (17)$$

Combining these yields

$$\ell_1 = \frac{\varepsilon^2 - \delta^2 + 2\delta}{2}$$

(by using (14) to substitute ℓ_1 for ℓ_2 in (16) and then subtracting the result from (15))

$$h^2 = \varepsilon^2 - \left(\frac{\varepsilon^2 - \delta^2 + 2\delta}{2}\right)^2 \quad (\text{plugging (18) into (15)})$$

$$w^2 = \delta^2 + \varepsilon^2 + \delta(\varepsilon^2 - \delta^2 + 2\delta). \quad (\text{by combining (17), (18), and (19)})$$

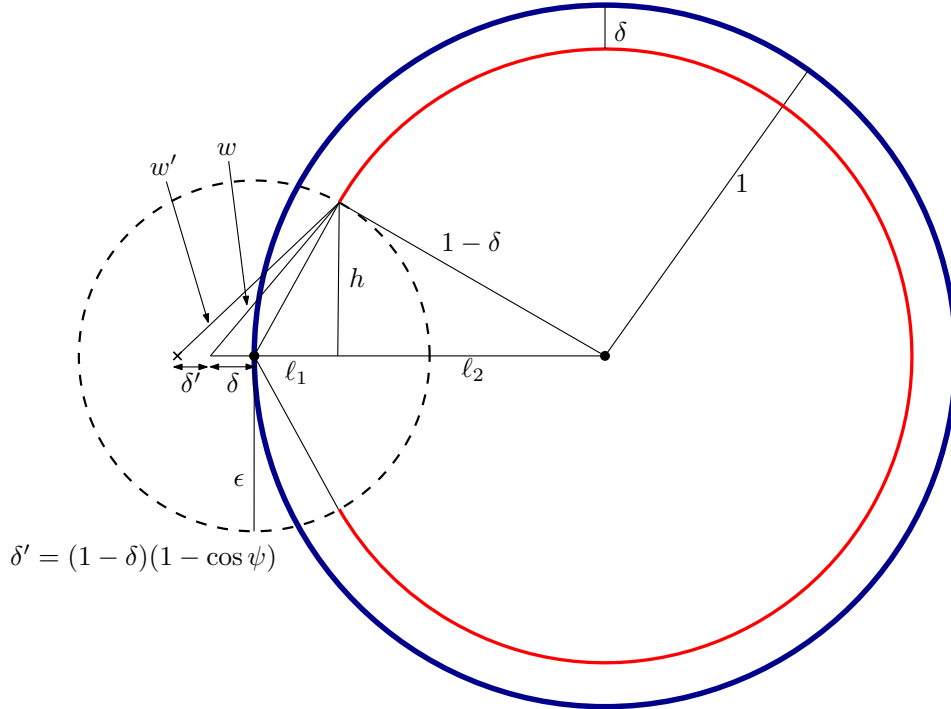
Figure 11 also yields

$$w'^2 = (\delta' + \delta + \ell_1)^2 + h^2,$$

from which one derives that

$$w'^2 = \varepsilon^2 + (1 - (1 - \delta) \cos \psi)^2 + (1 - (1 - \delta) \cos \psi)(\varepsilon^2 - \delta^2 + 2\delta), \quad (20)$$

using (18), (19), and the formula for δ' (see Figure 11 and (21) below)



■ **Figure 11** This figure is used for the computation of $r_0 = \frac{w}{2}$, $r_{-1} = h$, and w' .

Proof of Claim 18. By definition, the circumcentre of the simplex $p\tilde{p}q\tilde{q}$ lies on the bisector of p and \tilde{p} and the bisector of q and \tilde{q} . Thus, if the position of p and \tilde{p} on C_δ varies, then the simplex $p\tilde{p}q\tilde{q}$ transitions from being non-self-centred to self-centred exactly when its

circumcentre lies on the edge $p\tilde{p}$. At this point, the circumcentre of the face $p\tilde{p}q$ also lies on the edge $p\tilde{p}$. The same also holds for the face $p\tilde{p}\tilde{q}$.

This is again equivalent to the fact that the distance w' between the midpoint $\frac{p+\tilde{p}}{2}$ and q (and the midpoint $\frac{p+\tilde{p}}{2}$ and \tilde{q}) equals $\frac{\|p-\tilde{p}\|}{2}$. Because the points p and \tilde{p} lie on the circle C_δ , this distance can be written in terms of the angle $\psi = \angle p0\frac{p+\tilde{p}}{2}$, that is, $\frac{\|p-\tilde{p}\|}{2} = (1-\delta)\sin\psi$. The (one-sided Hausdorff) distance from $\frac{p+\tilde{p}}{2}$ to C_δ then equals

$$\delta' = (1-\delta)(1-\cos\psi). \quad (21)$$

The distance w' is given by (20). Hence the transition between self-centred and non-self-centred takes place when

$$(1-\delta)^2 \sin^2\psi = \varepsilon^2 + (1-(1-\delta)\cos\psi)^2 + (1-(1-\delta)\cos\psi)(\varepsilon^2 - \delta^2 + 2\delta).$$

Writing $x = \cos\psi$, this translates into

$$0 = 2(1-\delta)^2 x^2 - (1-\delta)(2 + \varepsilon^2 - \delta^2 + 2\delta)x + 2(\varepsilon^2 - \delta^2 + 2\delta).$$

The solution space of this quadratic equation is non-empty if the discriminant

$$\Delta = (1-\delta)^2(2 + \varepsilon^2 - \delta^2 + 2\delta)^2 - 4 \cdot 2(1-\delta)^2 \cdot 2(\varepsilon^2 - \delta^2 + 2\delta)$$

is non-negative. Introducing the variable $y = \varepsilon^2 - \delta^2 + 2\delta$ and applying that $\delta < 1$, the inequality $\Delta \geq 0$ reduces to

$$(2+y)^2 - 16y \geq 0,$$

and thus

$$y > 6 - 4\sqrt{2}.$$

The above expression is the same as Equation (13). In summary, this implies that there is no transition from self-centred to non-self centred simplices as the positions of p and \tilde{p} vary on C_δ if the inequality (13) is satisfied. ◀

► **Remark 19.** If $\varepsilon^2 = c + (6 - 4\sqrt{2}) - 2\delta + \delta^2$ with $c > 0$, then the circumcentre of the simplex $p\tilde{p}q\tilde{q}$ lies in its interior. This guarantees that the construction in Example 16 consists of a finite number of tori.

We can now finally prove the main statement of this section.

Proof of Proposition 11. We note that for $r \in [0, r_1)$ the set $(C_1 \cup \{p_1\tilde{p}_1\}) \oplus B(r)$ does not have the same homology as the torus T_1 , because the tetrahedron $p_1\tilde{p}_1q_1\tilde{q}_1$ is self-centred. The reason for this is that

- For $r \in [0, r_0)$, the set $(C_1 \cup \{p_1, \tilde{p}_1\}) \oplus B(r)$ has three connected components.
- For $r \in [r_0, r_1)$, the set $(C_1 \cup \{p_1, \tilde{p}_1\}) \oplus B(r)$ has homology type of a torus with either a circle or a 2-sphere attached (depending on if r is smaller or larger than the circumradius of the triangle $p_1\tilde{p}_1q_1$ -which is in turn equal to the circumradius of the triangle $p_1\tilde{p}_1\tilde{q}_1$).

We note there is also a topological transition at $r_{-1} = \sqrt{\varepsilon^2 - \left(\frac{\varepsilon^2 - \delta^2 + 2\delta}{2}\right)^2}$ (r_{-1} is h in Figure 11). The homotopy type of each $C_i \oplus B(r)$ changes from that of a circle to that of a torus. ◀

References

- 1 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.
- 2 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>.
- 3 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). URL: <https://www.sciencedirect.com/science/article/pii/S0925772112001423>, doi:<https://doi.org/10.1016/j.comgeo.2012.02.009>.
- 4 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.
- 5 Anders Björner. Topological methods. handbook of combinatorics, vol. 1, 2, 1819–1872, 1995.
- 6 Jean-Daniel Boissonnat. Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics (TOG)*, 3(4):266–286, 1984.
- 7 Jean-Daniel Boissonnat. Shape reconstruction from planar cross sections. *Computer vision, graphics, and image processing*, 44(1):1–29, 1988.
- 8 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.
- 9 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.
- 10 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.
- 11 Frédéric Chazal and André Lieutier. Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Computational Geometry*, 40(2):156–170, 2008.
- 12 Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- 13 Herbert Edelsbrunner and Nimish R Shah. Triangulating topological spaces. In *Proceedings of the tenth annual symposium on Computational geometry*, pages 285–292, 1994.
- 14 H. Federer. Curvature measures. *Transactions of the America mathematical Society*, 93:418–491, 1959.
- 15 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. URL: <https://drops.dagstuhl.de/opus/volltexte/2020/12212>, doi:10.4230/LIPIcs.SoCG.2020.54.
- 16 David Levin. The approximation power of moving least-squares. *Mathematics of computation*, 67(224):1517–1531, 1998.
- 17 André Lieutier and Mathijs Wintraecken. Embedded topological manifolds have positive reach if and only if they are $C^{1,1}$. To be submitted, 2022.
- 18 Alexander Lytchak. On the geometry of subsets of positive reach. *manuscripta mathematica*, 115(2):199–205, 2004.
- 19 Alexander Lytchak. Almost convex subsets. *Geometriae Dedicata*, 115(1):201–218, 2005.
- 20 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.

- 21 Robert Pless and Richard Souvenir. A survey of manifold learning for images. *IPSN Transactions on Computer Vision and Applications*, 1:83–94, 2009.
- 22 Jan Rataj and Martina Zähle. *Curvature measures of singular sets*. Springer, 2019.
- 23 Jan Rataj and Luděk Zajíček. On the structure of sets with positive reach. *Mathematische Nachrichten*, 290(11-12):1806–1829, 2017. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mana.201600237>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/mana.201600237>, doi:<https://doi.org/10.1002/mana.201600237>.
- 24 Yuan Wang and Bei Wang. Topological inference of manifolds with boundary. *Computational Geometry*, 88:101606, 2020. URL: <https://www.sciencedirect.com/science/article/pii/S0925772119301476>, doi:<https://doi.org/10.1016/j.comgeo.2019.101606>.

A Federer’s sets of positive reach

We recall the results from [14] that we used in the main text. Here we just assume that \mathcal{S} is a closed set.

► **Definition 20.** *The medial axis $\text{ax}(\mathcal{S})$ of a set \mathcal{S} is the set of points in the ambient space that do not have a unique closest point on \mathcal{S} . The (minimal) distance between $\text{ax}(\mathcal{S})$ and \mathcal{S} is the reach of \mathcal{S} . The distance from a point p to the medial axis is called the local feature size $\text{lfs}(p)$.*

► **Definition 21** (Definitions 4.3 and 4.4 of [14]). *If $\mathcal{S} \subseteq \mathbb{R}^d$ and $p \in \mathcal{S}$, then the generalized tangent space*

$$\text{Tan}(p, \mathcal{S})$$

is the set of all tangent vectors of \mathcal{S} at p . It consists of all those $u \in \mathbb{R}^d$, such that either $u = 0$ or for every $\varepsilon > 0$ there exists a point $q \in \mathcal{S}$ with

$$0 < \|q - p\| < \varepsilon \quad \text{and} \quad \left\| \frac{q - p}{\|q - p\|} - \frac{u}{\|u\|} \right\| < \varepsilon.$$

The set

$$\text{Nor}(p, \mathcal{S})$$

of all normal vectors of \mathcal{S} at p consists of all those $v \in \mathbb{R}^d$ such that $v \cdot u \leq 0$ for all $u \in \text{Tan}(p, \mathcal{S})$.

► **Definition and Remark 22** ([14, Remark 4.5]). *A subset $C \subseteq \mathbb{R}^d$ is a convex cone if and only if for all $x, y \in C$ and $\lambda > 0$ we have $x + y \in C$ and $\lambda x \in C$. For every $A \subseteq \mathbb{R}^d$,*

$$\text{Dual}(A) = \{v \mid \langle v, u \rangle \leq 0 \text{ for all } u \in A\}$$

is a closed convex cone and $\text{Dual}(\text{Dual}(A))$ is the smallest closed convex cone that contains A . The set $\text{Nor}(p, \mathcal{S})$ of all normal vectors of \mathcal{S} at p is therefore a cone and will be referred to as the normal cone. The generalized tangent space $\text{Tan}(p, \mathcal{S})$ on the other hand is only closed and positively homogeneous, i.e. if $v \in \text{Tan}(p, \mathcal{S})$ then $\lambda v \in \text{Tan}(p, \mathcal{S})$ for all $\lambda \in \mathbb{R}_{\geq 0}$, but not necessarily convex. Note that below we’ll see that $\text{Tan}(p, \mathcal{S})$ is a convex cone if \mathcal{S} has positive reach.

► **Theorem 23** ([14, Theorem 4.8 (8)]). *If $0 < \ell < R < \infty$, $x, y \notin \text{ax}(\mathcal{S})$, and*

$$d(x, \mathcal{S}) \leq \ell \quad d(y, \mathcal{S}) \leq \ell \quad \text{rch}(\mathcal{S}) \geq R,$$

then

$$\|\pi_{\mathcal{S}}(x) - \pi_{\mathcal{S}}(y)\| \leq \frac{R}{R - \ell} \|x - y\|.$$

► **Theorem 24** ([14, Theorem 4.8 (12)]). *If $p \in \mathcal{S}$ and $\text{lfs}(p) > \ell > 0$, then*

$$\text{Nor}(p, \mathcal{S}) = \{\lambda v \mid \lambda \geq 0, \|v\| = \ell, \pi_{\mathcal{S}}(p + v) = p\}.$$

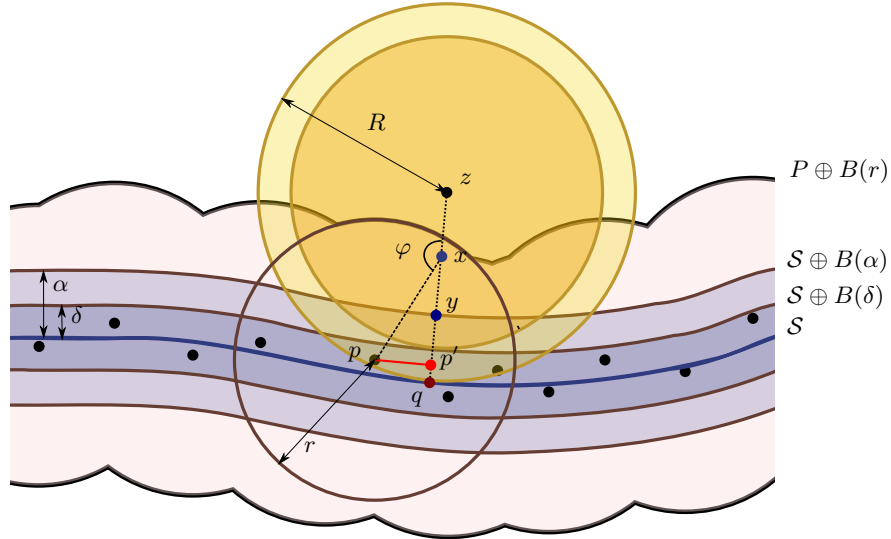
$\text{Tan}(p, \mathcal{S})$ is the convex cone dual to $\text{Nor}(p, \mathcal{S})$, and

$$\lim_{t \rightarrow 0^+} t^{-1} d_{\mathcal{S}}(p + tu) = 0,$$

for $u \in \text{Tan}(p, \mathcal{S})$.

B Alternative proof

Alternative proof of Theorem 2. Let us prove that the set $(q + \text{Nor}(q, \mathcal{S})) \cap B(q, R) \cap (P \oplus B(r))$ is star-shaped with respect to q . For this, consider a point $x \in (q + \text{Nor}(q, \mathcal{S})) \cap B(q, R) \cap (P \oplus B(r))$ and let us prove that the segment xq is also contained in $P \oplus B(r)$. We consider two cases. First, suppose that $\|x - q\| \leq \alpha$. In that case, $xq \subseteq B(q, \alpha) \subseteq \mathcal{S} \oplus B(\alpha) \subseteq P \oplus B(r)$ and we are done. Second, suppose that $\|x - q\| > \alpha$; see Figure 12.



■ **Figure 12** For the alternate proof of Theorem 2.

In that case, $x \neq q$ and the half-line with origin at q and passing through x is well-defined. Let y be the point on this half-line whose distance to q is α . Let z be the point on this half-line whose distance to q is R . Because $x \in B(q, R)$, we have that x lies on the segment qz . Let p be any point of P whose distance to x is smaller than or equal to r . It is this assumption that later gives $\|x - p\| \leq r$. Let p' be the projection of p onto the straight-line passing through q and x . We have that the five points x, y, z, q and p' are aligned and y lies between x and q . We claim that y also lies between x and p' . The claim is clearly true if q lies between x and p' . Let us assume that q does not lie between x and p' , in other words, let us assume that p' is on the half-line with origin at q and passing through x . Let φ be the internal angle of triangle xpz at x . The cosine rule gives:

$$\|z - p\|^2 = \|z - x\|^2 + \|x - p\|^2 - 2\|z - x\|\|x - p\| \cos \varphi. \quad (22)$$

By Theorem 24, the interior of $B(z, R)$ does not intersect \mathcal{S} and because $P \subseteq \mathcal{S} \oplus B(\delta)$, we have $\|z - p\| \geq R - \delta$. By construction, we have that $\|x - p\| \leq r$. Furthermore,

$\|z - x\| = \|z - q\| - \|q - x\| \leq R - \alpha$. It follows that

$$\begin{aligned} 2\|z - x\|\|x - p\| \cos \varphi &= \|z - x\|^2 + \|x - p\|^2 - \|z - p\|^2 && \text{(reshuffling (22))} \\ &\leq (R - \alpha)^2 + r^2 - (R - \delta)^2 \\ &\leq 0, \end{aligned}$$

showing that $\cos \varphi \leq 0$, or equivalently $\varphi \geq \frac{\pi}{2}$. Hence, p' lies on the segment qx . Let us show that $\|q - p'\| \leq \alpha$. Because p' belongs to the segment $qx \subseteq qz$, we have

$$\begin{aligned} \|q - p'\| &= \|q - z\| - \|p' - z\| \\ &= R - \sqrt{\|z - p\|^2 - \|p' - p\|^2} \\ &\leq R - \sqrt{\|z - p\|^2 - \|x - p\|^2} \\ &\leq R - \sqrt{(R - \delta)^2 - r^2} \\ &\leq \alpha = \|q - y\|. \end{aligned}$$

hence, y lies between x and p' . This shows that the distance to p decreases as we move on the segment xy , starting from x and going toward y . It follows that $xy \subseteq B(p, r) \subseteq P \oplus B(r)$. Since $yq \subseteq B(q, \alpha) \subseteq \mathcal{S} \oplus B(\alpha) \subseteq P \oplus B(r)$, we deduce that the whole segment xq belongs to $P \oplus B(r)$. The proof is completed by using the same deformation retract argument as in the first version of the proof. \blacktriangleleft