Homotopy Reconstruction via the Cech Complex and the Rips Complex

Jisu Kim, Jaehyeok Shin, Frédéric Chazal, Alessandro Rinaldo, Larry Wasserman

Homotopy Reconstruction via the Cech Complex and the Rips Complex

Jisu Kim[1], Jaehyeok Shin[2], Frédéric Chazal[1], Alessandro Rinaldo[2], and Larry Wasserman[2]

[1]Inria Saclay – Île-de-France
[2]Department of Statistics and Data Science, Carnegie Mellon University

Abstract

We derive conditions under which the reconstruction of a target space is topologically correct via the Čech complex or the Rips complex obtained from possibly noisy point cloud data. We provide two novel theoretical results. First, we describe sufficient conditions under which any non-empty intersection of finitely many Euclidean balls intersected with a positive reach set is contractible, so that the Nerve theorem applies for the restricted Čech complex. Second, we demonstrate the homotopic equivalence of a positive μ-reach set and its offsets. Applying these results to the restricted Čech complex and using the interleaving relations with the Čech complex (or the Rips complex), we formulate conditions guaranteeing that the target space is homotopy equivalent to the Čech complex (or the Rips complex), in terms of the μ-reach. Our results sharpen existing results.

1 Introduction

A fundamental task in topological data analysis, geometric inference, and computational geometry is that of estimating the topology of a set $\mathbb{X} \subset \mathbb{R}^d$ based on a finite collection of data points $\mathscr{X}$ that lie on it or on its proximity. This problem naturally occurs in many applications area, such as cosmology Starck et al. [2005], time series data Robins et al. [2000], machine learning Dey [2007], and so on.

A natural way to approximate the target space is to consider an r-offset of the data points, that is, to take the union of the open balls of radius $r > 0$ centered at the data points. Under approriate conditions, by the Nerve theorem this offset is topologically equivalent to the target space $\mathbb{X}$ via Čech complex Björner [1995], Hatcher [2002]. For computational resons, the Alpha shape complex may be used instead, which is homotopy equivalent to the Čech complex Edelsbrunner and Mücke [1994]. To further speed up the calculation, and in particular if the data are high dimensional, the Vietoris-Rips complex may be preferable, where only the pairwise distances between the data points are used and boosts up the calculation.

To guarantee that the topological approximation based on the data points recovers correctly the homotopy type of $\mathbb{X}$, it is necessary that the data points are dense and close to the target space, and that the radius parameter used for constructing the Čech complex or the Rips complex be of appropriate size. The condition is usually expressed as the parameter being lower bounded by a constant times the Hausdorff distance between the target space and the data points, and upper bounded by another constant times a measure of the size of the topological features of the target space. This condition has been extensively studied. Originally, the topological feature size was expressed in terms of the reach of $\mathbb{X}$: such condition for the Čech complex is studied in Chazal and Lieutier [2008], Niyogi et al. [2008]. Subsequently, the notion of μ-reach was put forward to allow for more general target spaces: the condition for the Čech complex is studied in Attali et al. [2013], Chazal et al. [2009], and the condition for the Rips complex is studied in Attali et al. [2013]. Also, the radii parameters are generalized to vary across the data points in Chazal and Lieutier [2008].

In this paper, we derive conditions under which the homotopy type of the target space is correctly recovered via the Čech complex or the Rips complex, in terms of the Hausdorff distance and the μ-reach of the target space. To tackle this problem, we provide two novel theoretical results. First, we describe sufficient conditions under which any non-empty intersection of finitely many Euclidean balls intersected with a set of positive reach is contractible, so that the Nerve theorem applies for the restricted Čech complex. Second, we demonstrate the homotopic equivalence of a positive μ-reach set and its offsets. These results are new and interesting in its own as well.

Overall, our new bounds offer significant improvements over the previous results in Niyogi et al. [2008], Attali et al. [2013] and are sharp: in particular, they achieve the optimal upper bound for the parameter of the Čech complex and the Rips complex under the positive reach condition. We will provide a detailed comparison of our results with existing ones in Section 6.

2 Background

A natural way to approximate $\mathbb{X}$ with $\mathscr{X}$ is to take the union of open balls centered at the data points. In detail, let $r = \{r_x, x \in \mathscr{X}\} \in \mathbb{R}_+^{\mathscr{X}}$ be pre-specified radii and consider the set

$$\bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{X}}(x, r_x), \tag{1}$$

where $\mathbb{B}_{\mathbb{X}}(x,r)$ is the open *restricted ball* centered at $x \in \mathbb{R}^d$ of radius $r > 0$, defined as

$$\mathbb{B}_{\mathbb{X}}(x,r) := \{y \in \mathbb{X} : d(x,y) < r\}, \tag{2}$$

where $d(x,y) := \|x-y\|$, with $\|\cdot\|$ denoting the Euclidean norm. Even though we allow for the points in $\mathscr{X}$ to lie outside $\mathbb{X}$, we will assume throughout that $\mathbb{B}_{\mathbb{X}}(x,r_x) \neq \emptyset$ for all $x \in \mathscr{X}$.

The topological properties of the union of restricted balls given above in (1) are encoded in the corresponding weighted *Čech complex*, which is the abstract simplicial complex on $\mathscr{X}$ consisting of all subsets $\{x_1,\dots,x_k\}$ of $\mathscr{X}$ such that the intersection $\cap_{j=1}^k \mathbb{B}_{\mathbb{X}}(x_j, r_{x_j})$ is non-empty.

Definition 1 (Čech complex)**.** *Let $\mathscr{X}$, $\mathbb{X}$ be two subsets and $r \in \mathbb{R}_+^{\mathscr{X}}$. The (weighted) Čech complex is the simplicial complex*

$$\check{C}ech_{\mathbb{X}}(\mathscr{X}, r) := \left\{ \sigma = [x_1, \ldots, x_k] \subset \mathscr{X} : \bigcap_{j=1}^{k} \mathbb{B}_{\mathbb{X}}(x_j, r_{x_j}) \neq \emptyset, k = 1, \ldots, n \right\}. \tag{3}$$

Under appropriate conditions, the homology of the resulting Čech complex in fact coincides with the homology of $\mathbb{X}$ itself. This remarkable result is precisely the renown nerve theorem Björner [1995], Hatcher [2002], which we recall next.

Definition 2 (Nerve)**.** *Let $\mathscr{U} = \{U_\alpha\}$ be an open cover of a given topological space $\mathbb{X}$. The nerve of $\mathscr{U}$, denoted by $\mathscr{N}(\mathscr{U})$, is the abstract simplicial complex defined as*

$$\mathscr{N}(\mathscr{U}) = \left\{ [U_1, \ldots, U_k] \subset \mathscr{U} : \bigcap_{j=1}^{k} U_j \neq \emptyset \right\}.$$

The nerve theorem prescribes conditions under which the nerve of an open cover of $\mathbb{X}$ is homotopy equivalent to $\mathbb{X}$ itself.

Theorem 1 (Nerve theorem)**.** *Let $\mathscr{U}$ be its open cover of a paracompact space $\mathbb{X}$. If every nonempty intersections of finitely many sets in $\mathscr{U}$ is contractible, then $\mathbb{X}$ is homotopy equivalent to the nerve $\mathscr{N}(\mathscr{U})$.*

Thus, in order to conclude that $\check{C}ech_{\mathbb{X}}(\mathscr{X}, r)$ in (3) has the same homology as $\mathbb{X}$, it is enough to show, by the nerve theorem, that the union of restricted balls $\bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{X}}(x, r_x)$ covers the target space $\mathbb{X}$ and that any arbitrary non-empty intersection of restricted balls is contractible. The difficulty in establishing the latter, more technical, condition lies in the fact that it is not clear a priori what properties of $\mathbb{X}$ will imply it. If $\mathbb{X}$ is a convex set, then the nerve theorem applies straightforwardly. But for more general spaces, such as smooth lower-dimensional manifolds, it is not obvious how contractibility may be guaranteed. One of the main result of this note, given below in Theorem 5, asserts that if $\mathbb{X}$ has positive reach and the radii of the restricted balls are small compared to the reach, then any non-empty intersections of restricted balls is contractible.

2.1 The reach

First introduced by Federer [1959], the reach is a quantity expressing the degree of geometric regularity of a set. In detail, given a closed subset $A \subset \mathbb{R}^d$, the medial axis of A, denoted by $\mathrm{Med}(A)$, is the subset of $\mathbb{R}^d$ consisting of all the points that have at least two nearest neighbors in A. Formally,

$$\mathrm{Med}(A) = \left\{ x \in \mathbb{R}^d \setminus A \colon \exists q_1 \neq q_2 \in A, ||q_1 - x|| = ||q_2 - x|| = d(x, A) \right\}, \tag{4}$$

where $d(x, A) = \inf_{q \in A} ||q - x||$ denotes the distance from a generic point $x \in \mathbb{R}^d$ to A. The reach of A is then defined as the minimal distance from A to $\mathrm{Med}(A)$.

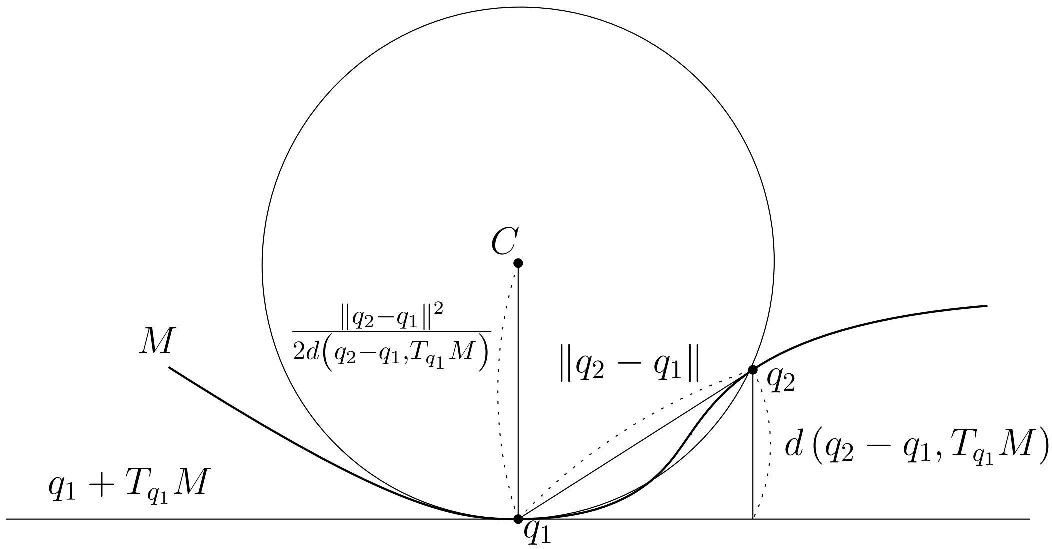

Figure 1: Geometric interpretation of the quantities involved in (6).

Definition 3. *The reach of a closed subset $A \subset \mathbb{R}^d$ is defined as*

$$\tau_A = \inf_{q\in A} d\left(q, \mathrm{Med}(A)\right) = \inf_{q\in A, x\in \mathrm{Med}(A)} ||q - x||. \tag{5}$$

Some authors [see, e.g. Niyogi et al., 2008, Singer and Wu, 2012] refer to τ_A^{-1} as the *condition number.* From the definition of the medial axis in (4), the projection $\pi_A(x) = \arg\min_{p\in A} \|p - x\|$ onto A is well defined (i.e. unique) outside $Med(A)$. In fact, the reach is the largest distance $\rho \geq 0$ such that π_A is well defined on the ρ-offset $\left\{x \in \mathbb{R}^d | d(x,A) < \rho\right\}$. Hence, assuming the set A has positive reach can be seen as a generalization or weakening of convexity, since a set $A \subset \mathbb{R}^d$ is convex if and only if $\tau_A = \infty$.

In the case of submanifolds, one can reformulate the definition of the reach in the following manner.

Proposition 2 (Theorem 4.18 in Federer [1959])**.** *For all submanifold $M \subset \mathbb{R}^d$,*

$$\tau_M = \inf_{q_1\neq q_2 \in M} \frac{\|q_1 - q_2\|_2^2}{2d(q_2 - q_1, T_{q_1}M)}. \tag{6}$$

Above, T_qM denotes the tangent space of M at $q \in M$. This formulation has the advantage of involving only points on M and its tangent spaces, while (5) uses the distance to the medial axis $Med(M)$, which is a global quantity. The ratio appearing in (6) has a clear geometric meaning, as it corresponds to the radius of the ambient ball tangent to M at q_1 and passing through q_2. See Figure 1. Hence, the reach gives a lower bound on the radii of curvature of M. Equivalently, τ_M^{-1} is an upper bound on the directional curvature of M.

Proposition 3 (Proposition 6.1 in Niyogi et al. [2008])**.** *Let $M \subset \mathbb{R}^d$ be a submanifold, and $\gamma_{p,v}$ an arc-length parametrized geodesic of M. Then for all $t > 0$,*

$$\left\|\gamma''_{p,v}(t)\right\| \leq 1/\tau_M.$$

The reach further provides an upper bound on the injectivity radius and the sectional curvature of M; see Proposition A.1. part (ii) and (iii) respectively, in Aamari et al. [2017]. Hence the reach is a quantity that characterizes the overall structure of M and, as argued in Aamari et al. [2017], captures structural properties of M of both global and local nature. In particular, assuming a lower bound on the reach prevents the manifold from being nearly self-intersecting or from having portions with very high curvature [Aamari et al., 2017, Theorem 3.4]. In the next section, we describe how to use the reach condition to ensure that the union of restricted balls is contractible, which in turn allows us to apply the Nerve theorem to recover the homology of the target space $\mathbb{X}$.

For non-smooth target spaces, the reach of the space can be equal to zero. In this case, we can deploy a more general notion of feature size, called μ-reach, introduced by Chazal et al. [2009]. For any point $x \in \mathbb{R}^d \setminus \mathbb{X}$, let $\Gamma_{\mathbb{X}}(x)$ be the set of points in $\mathbb{X}$ closest to x. Let $\Theta_{\mathbb{X}}(x)$ be the center of $\Gamma_{\mathbb{X}}(x)$. Then, for $x \in \mathbb{R}^d \setminus \mathbb{X}$, the generalized gradient of the distance function $x \mapsto d(x, \mathbb{X})$ is defined as

$$\nabla_{\mathbb{X}}(x) = \frac{x - \Theta_{\mathbb{X}}(x)}{d(x, \mathbb{X})}. \tag{7}$$

Then, the μ-medial axis of $\mathbb{X}$ is defined as

$$\mathrm{Med}_{\mu}(\mathbb{X}) = \left\{ x \in \mathbb{R}^d \setminus \mathbb{X} : \|\nabla_{\mathbb{X}}(x)\| < \mu \right\}, \tag{8}$$

Finally, the μ-reach of $\mathbb{X}$ is defined as the minimal distance from $\mathbb{X}$ to $\mathrm{Med}_{\mu}(\mathbb{X})$.

Definition 4. *The μ-reach of a compact set $\mathbb{X} \subset \mathbb{R}^d$ is defined as*

$$\tau_{\mathbb{X}}^{\mu} = \inf_{q \in \mathbb{X}} d\left(q, \mathrm{Med}_{\mu}(A)\right) = \inf_{q \in \mathbb{X}, x \in \mathrm{Med}_{\mu}(A)} ||q - x||. \tag{9}$$

Note that if $\mu = 1$, the corresponding μ-reach is equal to the reach of $\mathbb{X}$.

2.2 Restricted versus Ambient balls

It is important to point out that the nerve theorem needs not apply to the Čech complex built using ambient, as opposed, to restricted balls. In particular, the homology of $\mathbb{X}$, may not be correctly recovered using unions of ambient balls even if the point cloud is dense in $\mathbb{X}$ and the radii of the balls all vanish. We elucidate this point in the next example. Below, $\mathbb{B}_{\mathbb{R}^d}(x, r)$ denotes the open ambient ball in $\mathbb{R}^d$ centered at x and of radius $r > 0$.

Example 4. *Let $\mathbb{X} = (\partial \mathbb{B}_{\mathbb{R}^2}(0,1)) \cap \{x \in \mathbb{R}^2 : x_2 \geq 0\}$ be a semicircle in $\mathbb{R}^2$. Let $\varepsilon \in (0,1)$ be fixed, and let X_1, X_2 be points on $\mathbb{X}$ satisfying $\|X_1 - X_2\| \in \left(\varepsilon\sqrt{4-\varepsilon^2}, 2\varepsilon\right)$. Then, $\mathbb{B}_{\mathbb{R}^2}(X_1, \varepsilon) \cap \mathbb{B}_{\mathbb{R}^2}(X_2, \varepsilon)$ is nonempty but has an empty intersection with $\mathbb{X}$. Now, choose $\rho < d(\mathbb{X}, \mathbb{B}_{\mathbb{R}^2}(X_1, \varepsilon) \cap \mathbb{B}_{\mathbb{R}^2}(X_2, \varepsilon))$ and choose $\mathscr{X}_0 \subset \mathbb{X}$ be dense enough so that $\bigcup_{x \in \mathscr{X}_0} \mathbb{B}_{\mathbb{R}^2}(x, \rho)$ covers $\mathbb{X}$. Now, consider the union of ambient balls*

$$\left(\mathbb{B}_{\mathbb{R}^2}(X_1, \varepsilon) \bigcup \mathbb{B}_{\mathbb{R}^2}(X_2, \varepsilon)\right) \bigcup \left(\bigcup_{x \in \mathscr{X}_0} \mathbb{B}_{\mathbb{R}^2}(x, \rho)\right). \tag{10}$$

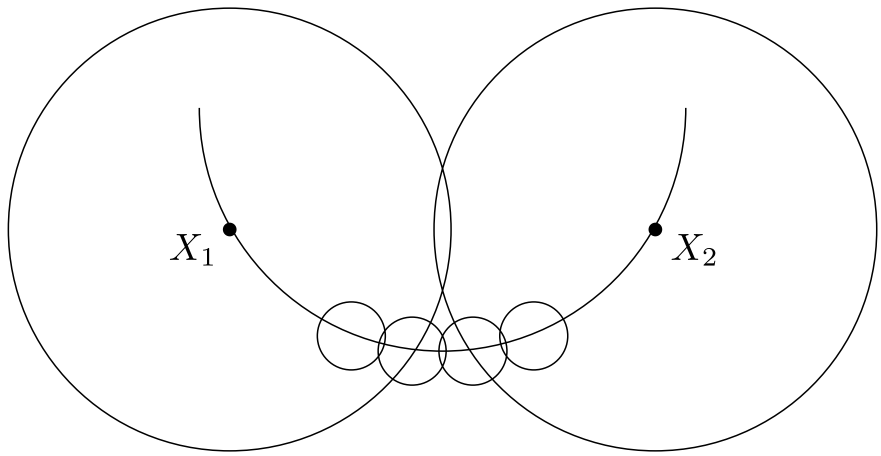

Figure 2: *An example in which the union of balls is different from the underlying space in terms of homotopy and homology. In the figure, the union of balls deformation retracts to a circle, hence its homotopy and homology is different from the underlying semicircle.*

Then from the fact $\rho < d\left(\mathbb{X}, \mathbb{B}_{\mathbb{R}^2}(X_1, \varepsilon) \cap \mathbb{B}_{\mathbb{R}^2}(X_2, \varepsilon)\right)$ *and* $\bigcup_{x \in \mathscr{X}_0} \mathbb{B}_{\mathbb{R}^2}(x, \rho)$ *covering* $\mathbb{X}$*, we have that the union of balls in* (10) *is homotopic to a circle, hence its homotopy and homology is different from the semicircle* $\mathbb{X}$*. Note that the above construction union of balls holds for all choices of* $\varepsilon \in (0,1)$*. Since* $\rho \to 0$ *as* $\varepsilon \to 0$*,* $\mathscr{X}_0$ *can be arbitrary dense in* $\mathbb{X}$*. See Figure 2.*

3 The nerve theorem for Euclidean sets of positive reach

In order to apply the nerve theorem to the Čech complex built on restricted balls, it is enough to check whether any finite intersection of the restricted balls $\bigcap_{j=1}^{k} \mathbb{B}_{\mathbb{X}}(x_j, r_{x_j})$ is contractible (since $\mathbb{X}$ is a subset of $\mathbb{R}^d$ and is endowed with the subspace topology, it is paracompact.)

Theorem 5 is the main statement of this paper and shows that if a subset $\mathbb{X} \subset \mathbb{R}^d$ has a positive reach $\tau > 0$, any non-empty intersection of restricted balls is contractible if the radii are small enough compared to the reach τ.

Theorem 5. *Let* $\mathbb{X} \subset \mathbb{R}^d$ *be a subset with reach* $\tau > 0$ *and let* $\mathscr{X} \subset \mathbb{R}^d$ *be a set of points. Let* $\{r_x > 0 : x \in \mathscr{X}\}$ *be a set of radii indexed by* $x \in \mathscr{X}$*. Then, if* $r_x \leq \sqrt{\tau^2 + (\tau - d_{\mathbb{X}}(x))^2}$ *for all* $x \in \mathscr{X}$*, any nonempty intersection of restricted balls* $\bigcap_{x \in I} \mathbb{B}_{\mathbb{X}}(x, r_x)$ *for* $I \subset \mathscr{X}$ *is contractible.*

Therefore, by combining Theorem 5 and the Nerve Theorem (Theorem 1), we can establish that the topology of the subspace $\mathbb{X}$ can be recovered by the corresponding restricted Čech complex $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}, r)$, provided the radii of the balls are not too large with respect to the reach. This result is summarized in the following corollary.

Corollary 6 (Nerve Theorem on the restricted balls)**.** *Under the same condition in Theorem 5, suppose the union of restricted balls covers the target space* $\mathbb{X}$*, that is,*

$$\mathbb{X} \subset \bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{X}}(x, r_x). \tag{11}$$

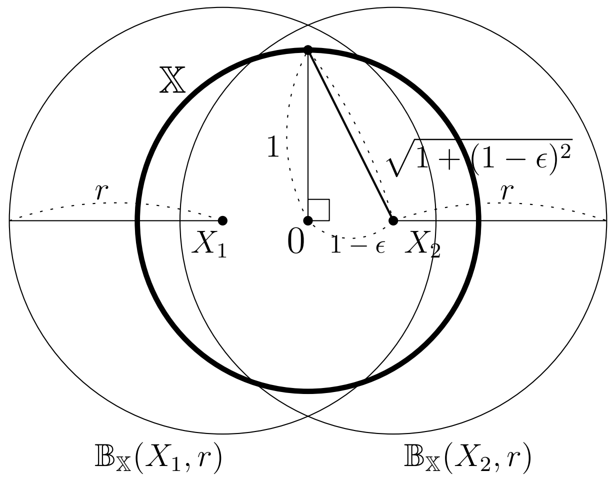

Figure 3: *An example in which $\mathbb{B}_{\mathbb{X}}(X_1,r)\bigcup\mathbb{B}_{\mathbb{X}}(X_2,r)$ is not homotopic equivalent to $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}_n,r)$ where $\mathbb{X}=\left\{x\in\mathbb{R}^2:\|x\|_2=1\right\}$, $X_1=(-1+\varepsilon,0)$, $X_2=(1-\varepsilon,0)$ and $r>\sqrt{1+(1-\varepsilon)^2}$, for any $\varepsilon>0$.*

If $r_x\leq\sqrt{\tau^2+(\tau-d_{\mathbb{X}}(x))^2}$ for all $x\in\mathscr{X}$, then $\mathbb{X}$ is homotopy equivalent to the restricted Čech complex $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r)$.

The reach condition $r_x\leq\sqrt{\tau^2+(\tau-d_{\mathbb{X}}(x))^2}$ is tight as the following example shows that the target space $\mathbb{X}$ is not homotopy equivalent to the Čech complex when $\max_{x\in\mathscr{X}} r_x>\sqrt{\tau^2+(\tau-d_{\mathbb{X}}(x))^2}$.

Example 7. *Let $\mathbb{X}$ be a unit Euclidean sphere in $\mathbb{R}^d$, and fix $\varepsilon>0$. Let $X_1:=(1-\varepsilon,0,\ldots,0),X_2:=(-1+\varepsilon,0,\ldots,0)\in\mathbb{R}^d$ be two points in $\mathbb{R}^d$. For a unit Euclidean sphere, the reach is equal to its radius 1. Therefore, if $r=(r_1,r_2)\in\left(0,\sqrt{1+(1-\varepsilon)^2}\right]^2$ and $\mathbb{X}\subset\mathbb{B}_{\mathbb{X}}(X_1,r_1)\bigcup\mathbb{B}_{\mathbb{X}}(X_2,r_2)$, $\mathbb{B}_{\mathbb{X}}(X_1,r_1)\bigcup\mathbb{B}_{\mathbb{X}}(X_2,r_2)$ is homotopic equivalent to $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}_n,r)$ by Corollary 6. However, if $r_1,r_2>\sqrt{1+(1-\varepsilon)^2}$, $\mathbb{B}_{\mathbb{X}}(X_1,r_1)\bigcup\mathbb{B}_{\mathbb{X}}(X_2,r_2)\simeq\mathbb{X}$ but $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}_n,r)\simeq 0$. Figure 3 illustrate the 2-dimensional case.*

4 Deformation retraction on positive μ-reach

The positive reach condition is critical for the nerve theorem on the restricted Čech complex. And it is not easily generalized to the positive μ-reach condition. Instead, we find a positive reach set that approximates the positive μ-reach set.

The set we will use to approximate positive μ-reach set is the double offset Chazal et al. [2007]. For $r>0$, an r-offset $\mathbb{X}^r$ of a set $\mathbb{X}$ is the collection of all points that are within r distance to $\mathbb{X}$, that is, $\mathbb{X}^r:=\bigcup_{x\in\mathbb{X}}\mathbb{B}_{\mathbb{R}^d}(x,r)$. The double offset is to take offset, take complement, take offset, and take complement, that is, $\mathbb{X}^{r,s}:=(((\mathbb{X}^r)^{\complement})^s)^{\complement}$. Roughly speaking, it is to inflate your set first, and then deflate your set, so that sharp corners become smooth. See Chazal et al. [2007] for more details.

To see the topological relation between the target space $\mathbb{X}$ of positive μ-reach and its double offset, we first need to understand the topological relation between the positive μ-reach set and its offset. Homotopic relationship between offsets $\mathbb{X}^r$ and $\mathbb{X}^s$ are well known for positive μ-reach set (see e.g., Grove [1993] or Proposition 3.4 in Chazal and Lieutier [2005]), but the correspondence to the original set $\mathbb{X}$ is not known before. Theorem 8, which is one of main theorems in our paper, asserts that the homotopy equivalence exists between the positive μ-reach set and its offset when the offset size is not large.

Theorem 8. *Let $\mathbb{X}$ be a set with positive μ-reach $\tau_\mu > 0$. For $r \le \tau_\mu$, $\mathbb{X}_r := \bigcup_{x\in\mathbb{X}} B(x,r)$ deformation retracts to $\mathbb{X}$. In particular, $\mathbb{X}$ and $\mathbb{X}_r$ are homotopy equivalent.*

To set up the homotopy equivalent of the positive μ-reach set and its double offset, we need another tool for finding the homotopy equivalence of the complement set.

Lemma 9. *Let $\mathbb{X}$ be a set with positive reach $\tau > 0$. For $r \le \tau$, $\mathbb{X}^\complement$ deformation retracts to $(\mathbb{X}^r)^\complement$. In particular, $\mathbb{X}^\complement$ and $(\mathbb{X}^r)^\complement$ are homotopic.*

Now, combining Theorem 8 and Lemma 9 gives the desired homotopy equivalent between the target set of positive μ-reach and its double offset, where the double offset has a positive reach.

Corollary 10. *Let $\mathbb{X}$ be a set with positive μ-reach $\tau_\mu > 0$. For $r,s > 0$ with $s \le r$, let $\mathbb{X}^{r,s} := (((\mathbb{X}^r)^\complement)^s)^\complement$ be a double offset of $\mathbb{X}$. If $s \le \mu r$, then $\mathbb{X}^{r,s}$ deformation retracts to $\mathbb{X}$, with a positive reach as*

$$reach(\mathbb{X}^{r,s}) \ge s.$$

5 Homotopy Reconstruction via Cech complex and Rips complex

We first reprove the well-known interleaving relationship of the ambient Čech complex and the Rips complex. The interleaving result when all radii are the same is well known (e.g., see Theorem 2.5 in de Silva and Ghrist [2007]). With the same proof technique, we are extending to the different radii case.

Lemma 11. *Let $\mathscr{X} \subset \mathbb{R}^d$ be a set of points. Let $\{r_x > 0 : x \in \mathscr{X}\}$ be a set of radii indexed by $x \in \mathscr{X}$. Then,*

$$\check{C}ech_{\mathbb{R}^d}(\mathscr{X}, r) \subset Rips(\mathscr{X}, r) \subset \check{C}ech_{\mathbb{R}^d}\left(\mathscr{X}, \sqrt{\frac{2d}{d+1}}r\right).$$

To recover the homotopy of the target set via the ambient Čech complex and the Rips complex, we utilize the restricted Čech complex. Hence, we set up the interleaving relationship between the the restricted Čech complex and the ambient Čech complex, and between the

Lemma 12. *Let $\mathbb{X} \subset \mathbb{R}^d$ be a subset with reach $\tau > 0$ and let $\mathscr{X} \subset \mathbb{R}^d$ be a set of points. Let $\{r_x > 0 : x \in \mathscr{X}\}$ be a set of radii indexed by $x \in \mathscr{X}$. Then,*

$$\check{C}ech_{\mathbb{X}}(\mathscr{X}, r) \subset \check{C}ech_{\mathbb{R}^d}(\mathscr{X}, r) \subset \check{C}ech_{\mathbb{X}}(\mathscr{X}, r'),$$

where $r' = \{r'_x > 0 : x \in \mathscr{X}\}$ *with* $r'_x = \sqrt{2r_x^2 + d_{\mathbb{X}}(x)(2\tau - d_{\mathbb{X}}(x))}$. *And also,*

$$\check{C}ech_{\mathbb{X}}(\mathscr{X}, r) \subset Rips(\mathscr{X}, r) \subset \check{C}ech_{\mathbb{X}}(\mathscr{X}, r''),$$

where $r'' = \{r''_x > 0 : x \in \mathscr{X}\}$ *with* $r''_x = \sqrt{\frac{4d}{d+1} r_x^2 + d_{\mathbb{X}}(x)(2\tau - d_{\mathbb{X}}(x))}$.

Combining Lemma 11 and 12 together with Nerve Theorem on a positive reach set (Corollary 6) gives the homotopy equivalece between the target space and the ambient Čech complex. Theorem 13 asserts that when the target space is densely covered by the data points and if they are not too far apart, the ambient Čech complex can be used to recover the homotopy type.

Theorem 13. *Let* $\mathbb{X} \subset \mathbb{R}^d$ *be a subset with reach* $\tau > 0$ *and let* $\mathscr{X} \subset \mathbb{R}^d$ *be a set of points. Let* $\{r_x > 0 : x \in \mathscr{X}\}$ *be a set of radii indexed by* $x \in \mathscr{X}$, *and let* $\varepsilon := \max\{d_{\mathbb{X}}(x) : x \in \mathscr{X}\}$. *Suppose* $\mathbb{X}$ *is covered by the union of balls centered at* $x \in \mathscr{X}$ *and radius* δ *as*

$$\mathbb{X} \subset \bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{R}}(x, \delta).$$

Suppose that for all $x \in \mathscr{X}$, *the radius* r_x *is bounded as*

$$r_x \leq \tau - d_{\mathbb{X}}(x). \tag{12}$$

Also, suppose δ *satisfies the following condition:*

$$\begin{aligned} &\delta + \sqrt{r_{\max}^2 + \varepsilon(2\tau - \varepsilon) - \frac{1}{4}\left(r_{\min} - \frac{r_{\max}^2}{\tau - \varepsilon + \sqrt{(\tau - \varepsilon)^2 - r_{\max}^2}} - \varepsilon - \delta\right)^2} \leq r_{\min}, \\ &\delta \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{r_{\min}^2 - \varepsilon(2\tau - \varepsilon)}{2}} - \sqrt{\frac{2\tau(r_{\min}^2 + \varepsilon(2\tau - \varepsilon))}{2\tau + \sqrt{4\tau^2 - 2\left(r_{\min}^2 + \varepsilon(2\tau - \varepsilon)\right)}}}\right). \end{aligned} \tag{13}$$

Then $\mathbb{X}$ *is homotopy equivalent to the ambient Čech complex* $\check{C}ech_{\mathbb{R}^d}(\mathscr{X}, r)$.

Similar approach also gives the homotopy equivalece between the target space and the Rips complex, as in Theorem 14.

Theorem 14. *Let* $\mathbb{X} \subset \mathbb{R}^d$ *be a subset with reach* $\tau > 0$ *and let* $\mathscr{X} \subset \mathbb{R}^d$ *be a set of points. Let* $\{r_x > 0 : x \in \mathscr{X}\}$ *be a set of radii indexed by* $x \in \mathscr{X}$, *and let* $\varepsilon := \max\{d_{\mathbb{X}}(x) : x \in \mathscr{X}\}$. *Suppose* $\mathbb{X}$ *is covered by the union of balls centered at* $x \in \mathscr{X}$ *and radius* δ *as*

$$\mathbb{X} \subset \bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{R}}(x, \delta).$$

Suppose that for all $x \in \mathscr{X}$, *the radius* r_x *is bounded as*

$$r_x \leq \sqrt{\frac{d+1}{2d}}\left(\tau - d_{\mathbb{X}}(x)\right). \tag{14}$$

Also, suppose δ satisfies the following condition:

$$\delta \leq r_{\min} - \frac{1}{2}\sqrt{\frac{2\tau\left(\frac{d}{2(d+1)}r_{\max}^2 + \varepsilon(2\tau-\varepsilon)\right)}{\tau + \sqrt{\tau^2 - \left(\frac{d}{2(d+1)}r_{\max}^2 + \varepsilon(2\tau-\varepsilon)\right)}}},$$

$$\delta \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{r_{\min}^2 - \varepsilon(2\tau-\varepsilon)}{2}} - \sqrt{\frac{2\tau(r_{\min}^2 + \varepsilon(2\tau-\varepsilon))}{2\tau + \sqrt{4\tau^2 - 2\left(r_{\min}^2 + \varepsilon(2\tau-\varepsilon)\right)}}}\right). \tag{15}$$

Then $\mathbb{X}$ is homotopy equivalent to the Rips complex $Rips(\mathscr{X}, r)$.

Example 15. *Let $\mathbb{X} \subset \mathbb{R}^d$ be the unit sphere in $\mathbb{R}^d$, and let $\mathscr{X} = \{X_1, \ldots, X_{d+1}\} \subset (1-\varepsilon)\mathbb{X}$ be $d+1$ points on the sphere centered at 0 and of radius $1-\varepsilon$. Further assume that $\|X_i - X_j\|$'s are all equal, so that X_i's form the regular d-dimensional simplex. Then $\|X_i - X_j\| = \sqrt{\frac{2(d+1)}{d}}(1-\varepsilon)$. For a unit Euclidean sphere $\mathbb{X}$, the reach is equal to its radius 1.*

For the ambient Čech complex, if $r \in (0, 1-\varepsilon]^{d+1}$ and $\mathbb{X} \subset \bigcup_{i=1}^{d+1} \mathbb{B}_{\mathbb{X}}(X_i, r_i)$, then $\bigcup_{i=1}^{d+1} \mathbb{B}_{\mathbb{X}}(X_i, r_i)$ is homotopic equivalent to $Čech_{\mathbb{X}}(\mathscr{X}, r)$ by Theorem 13. However, if $r_{x_I} > 1-\varepsilon$ for all i, then $0 \in \bigcap_{i=1}^{d+1} \mathbb{B}_{\mathbb{R}^d}(X_i, r_{X_i})$, and hence $Čech_{\mathbb{R}^d}(\mathscr{X}, r) = 2^{\mathscr{X}}$ and $Čech_{\mathbb{R}^d}(\mathscr{X}, r) \simeq 0$, while $d-1$-th homology group of $\mathbb{X}$ is $H_{d-1}(\mathbb{X}) = \mathbb{Z}$, so $\mathbb{X}$ and $Čech_{\mathbb{R}^d}(\mathscr{X}, r)$ are not homotopy equivalent.

For the Rips complex, if $r \in \left(0, \sqrt{\frac{d+1}{2d}}(1-\varepsilon)\right]^{d+1}$ and $\mathbb{X} \subset \bigcup_{i=1}^{d+1} \mathbb{B}_{\mathbb{X}}(X_i, r_i)$, then $\bigcup_{i=1}^{d+1} \mathbb{B}_{\mathbb{X}}(X_i, r_i)$ is homotopic equivalent to $Čech_{\mathbb{X}}(\mathscr{X}, r)$ by Theorem 13. However, if $r_{x_I} > \sqrt{\frac{d+1}{2d}}(1-\varepsilon)$ for all i, then $\|X_i - X_j\| = \sqrt{\frac{2(d+1)}{d}}(1-\varepsilon) < r_{X_i} + r_{X_j}$, and hence $Rips(\mathscr{X}, r) = 2^{\mathscr{X}}$ and $Rips(\mathscr{X}, r) \simeq 0$, while $d-1$-th homology group of $\mathbb{X}$ is $H_{d-1}(\mathbb{X}) = \mathbb{Z}$, so $\mathbb{X}$ and $Rips(\mathscr{X}, r)$ are not homotopy equivalent.

Corollary 16. *Let $\mathbb{X} \subset \mathbb{R}^d$ be a subset with positive μ-reach $\tau_\mu > 0$ and let $\mathscr{X} \subset \mathbb{R}^d$ be a set of points. Let $\{r_x > 0 : x \in \mathscr{X}\}$ be a set of radii indexed by $x \in \mathscr{X}$. Let $\mathbb{Y} := (((\mathbb{X}^{\frac{r_{\max}+\varepsilon}{\mu}})^{\complement})^{r_{\max}+\varepsilon})^{\complement}$ be the double offset, and let and let $\varepsilon := \max\{d_{\mathbb{X}}(x) : x \in \mathscr{X}\}$. Suppose $\mathbb{Y}$ is covered by the union of balls centered at $x \in \mathscr{X}$ and radius δ as*

$$\mathbb{Y} \subset \bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{R}}(x, \delta),$$

and also covered by the union of balls centered at $x \in \mathscr{X}$ and radius r_x as

$$\mathbb{Y} \subset \bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{R}}(x, r_x).$$

1. Suppose δ satisfies the following condition:

$$\delta + \sqrt{r_{\max}^2 + \varepsilon(2r_{\max}+\varepsilon) - \frac{1}{4}\left(r_{\min} - r_{\max} - \varepsilon - \delta\right)^2} \leq r_{\min},$$

$$\delta \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{r_{\min}^2 - \varepsilon(2r_{\max}+\varepsilon)}{2}} - \sqrt{\frac{2\tau(r_{\min}^2+\varepsilon(2\tau-\varepsilon))}{2\tau+\sqrt{4(r_{\max}+\varepsilon)^2 - 2\left(r_{\min}^2+\varepsilon(2\tau-\varepsilon)\right)}}}\right).$$

Then $\mathbb{X}$ is homotopy equivalent to the ambient Čech complex Čech$_{\mathbb{R}^d}(\mathscr{X}, r)$.

2. Suppose δ satisfies the following condition:

$$\delta \leq r_{\min} - \frac{1}{2}\sqrt{\frac{2(r_{\max}+\varepsilon)\left(\frac{d}{2(d+1)}r_{\max}^2 + \varepsilon(2r_{\max}+\varepsilon)\right)}{(r_{\max}+\varepsilon) + \sqrt{(r_{\max}+\varepsilon)^2 - \left(\frac{d}{2(d+1)}r_{\max}^2 + \varepsilon(2r_{\max}+\varepsilon)\right)}}},$$

$$\delta \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{r_{\min}^2 - \varepsilon(2r_{\max}+\varepsilon)}{2}} - \sqrt{\frac{2\tau(r_{\min}^2+\varepsilon(2r_{\max}+\varepsilon))}{2\tau+\sqrt{4\tau^2 - 2\left(r_{\min}^2+\varepsilon(2r_{\max}+\varepsilon)\right)}}}\right).$$

Then $\mathbb{X}$ is homotopy equivalent to the Rips complex Rips$(\mathscr{X}, r)$.

We end this section by introducing a sampling condition in which we can guarantee the covering conditions in Corollary 6 and Theorem 13, 14 are satisfied. Let P be the sampling distribution on $\mathbb{X}$. We assume that there exist positive constants a, b and ε_0 such that, for all $x \in \mathbb{X}$, the following inequality holds:

$$P(\mathbb{B}_{\mathbb{R}^d}(x,\varepsilon)) \geq a\varepsilon^b, \quad \forall \varepsilon \in (0, \varepsilon_0). \tag{16}$$

This condition on P is also known as the (a,b)-condition or the standard condition Cuevas and Rodríguez-Casal [2004], Cuevas [2009], Chazal et al. [2018]. It is satisfied, for example, if $\mathbb{X}$ is a smooth manifold of dimension b and P has a density with respect to the Hausdorff measure on it bounded from below by a.

Under this condition, we have the following covering lemma.

Lemma 17. *Let $\{r_n = (r_{n,1}, \ldots, r_{n,n})\}_{n\in\mathbb{N}}$ be a triangular array of positive numbers. For each n, suppose the minimum of radii can be bounded as*

$$2\left(\frac{\log n}{an}\right)^{1/b} \leq \min_i r_{n,i} \leq 2\varepsilon_0. \tag{17}$$

Then the probability of the samples forming an r_n-covering of $\mathbb{X}$ is bounded as

$$\mathbb{P}\left(\mathbb{X} \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^d}(X_i, r_{n,i})\right) \geq 1 - \frac{1}{2^b \log n}. \tag{18}$$

6 Discussion and Conclusion

So far, when the target space $\mathbb{X}$ has positive μ-reach τ^μ and the data points $\mathscr{X}$ being contained in the ε-offset $\mathbb{X}^\varepsilon$ of $\mathbb{X}$, this paper investigated the conditions under which the ambient Čech complex $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ and the Rips complex $\mathrm{Rips}(\mathscr{X}, r)$ are homotopic to the target space $\mathbb{X}$. In this section, we further discuss our result and compare it with previous results. For comparison purpose, we consider the case when all the radii r_x's are equal, and we denote that as r. When all the radii's are equal, an analogous homotopic relation between the ambient Čech complex $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ and the target space $\mathbb{X}$ were presented in Attali et al. [2013] and Niyogi et al. [2008].

First, we compare the upper bound for the maximum parameter value r in $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ or $\mathrm{Rips}(\mathscr{X}, r)$. When $\mu = 1$ so that we have the positive reach condition $\tau^\mu = \tau$, our result suggests that the homotopic relations hold when $r \leq \tau - \varepsilon$ for $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ and $r \leq \sqrt{\frac{d+1}{2d}}(\tau - \varepsilon)$ for $\mathrm{Rips}(\mathscr{X}, r)$. As we have seen, these bounds are optimal bounds. In Niyogi et al. [2008], such a bound for $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ in Proposition 7.1 is $\frac{(\tau+\varepsilon)+\sqrt{\tau^2+\varepsilon^2-6\tau\varepsilon}}{2}$. Then our bound is strictly sharper than this when $\varepsilon > 0$ since

$$\frac{(\tau+\varepsilon)+\sqrt{\tau^2+\varepsilon^2-6\tau\varepsilon}}{2} < \frac{(\tau+\varepsilon)+\sqrt{\tau^2+9\varepsilon^2-6\tau\varepsilon}}{2} = \tau - \varepsilon.$$

In Attali et al. [2013], a necessary condition for $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ in Section 5.3 is $r \leq \tau - 3\varepsilon$, so our upper bound is strictly better when $\varepsilon > 0$.

Second, we compare the condition for the maximum possible ratio of Hausdorff distance $d_H(\mathbb{X}, \mathscr{X})$ and the μ-reach τ^μ. In our notation, $d_H(\mathbb{X}, \mathscr{X}) = \max\{\delta, \varepsilon\}$. For our case, the upper bound for the ratio $\rho := \frac{d_H(\mathbb{X}, \mathscr{X})}{\tau}$ is the root coming from the bound: for $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}, r)$, it is the maximum value ρ that satisfies both conditions below:

$$\rho + \sqrt{(\frac{r}{\tau})^2 + \rho(2-\rho) - \frac{1}{4}\left((\frac{r}{\tau}) - \frac{(\frac{r}{\tau})^2}{1-\rho+\sqrt{(1-\rho)^2-(\frac{r}{\tau})^2}} - 2\rho\right)^2} \leq \frac{r}{\tau}$$

for some $\frac{r}{\tau} \in (0, 1]$,

$$\rho \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{(1-\rho)^2-\rho(2\tau-\rho)}{2}} - \sqrt{\frac{2((1-\rho)^2+\rho(2-\rho))}{2+\sqrt{4-2\left((1-\rho)^2+\rho(2-\rho)\right)}}}\right).$$

For $\mathrm{Rips}(\mathscr{X},r)$, it is the maximum value ρ that satisfies both conditions below:

$$\rho \leq \sqrt{\frac{d+1}{2d}}(1-\rho) - \frac{1}{2}\sqrt{\frac{2\left(\frac{1}{4}(1-\rho)^2+\rho(2-\rho)\right)}{1+\sqrt{1-\left(\frac{1}{4}(1-\rho)^2+\rho(2-\rho)\right)}}},$$

$$\rho \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{\frac{d+1}{2d}(1-\rho)^2-\rho(2\tau-\rho)}{2}} - \sqrt{\frac{2(\frac{d+1}{2d}(1-\rho)^2+\rho(2-\rho))}{2+\sqrt{4-2\left(\frac{d+1}{2d}(1-\rho)^2+\rho(2-\rho)\right)}}}\right).$$

With the aid of computer program, we can check that $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$ is homotopy equivalent to $\mathbb{X}$ when $\frac{d_H(\mathbb{X},\mathscr{X})}{\tau} \leq 0.01126\cdots$. This result is worse than $3-\sqrt{8} \approx 0.1716\cdots$ in Proposition 7.1 in Niyogi et al. [2008] or $\frac{-3+\sqrt{22}}{13} \approx 0.1300\cdots$ in Section 5.3 in Attali et al. [2013]. Again with the aid of computer program, we can check that $\mathrm{Rips}(\mathscr{X},r)$ is homotopy equivalent to $\mathbb{X}$ when $\frac{d_H(\mathbb{X},\mathscr{X})}{\tau} \leq 0.03982\cdots$. This result is better than $\frac{2\sqrt{2-\sqrt{2}}-\sqrt{2}}{2+\sqrt{2}} \approx 0.03412\cdots$ in Section 5.3 in Attali et al. [2013].

In the asymptotics, as we sample more and more points from the target space, $\mathscr{X}$ forms a dense cover on $\mathbb{X}$, i.e. $\sup_{y\in\mathbb{X}}\inf_{x\in\mathscr{X}}\|x-y\| \to 0$. Still, the Hausdorff distance $d_H(\mathbb{X},\mathscr{X})$ need not goes to 0 when we have noisy sample distribution, i.e. $\sup_{x\in\mathscr{X}}\inf_{y\in\mathbb{X}}\|x-y\| \nrightarrow 0$. In our notation, $\delta \to 0$ but ε does not converges to 0. For this case, the upper bound for the ratio is again coming from the bound: for $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r)$, it is the maximum value ρ that satisfies both conditions below:

$$\sqrt{(\frac{r}{\tau})^2+\rho(2-\rho)-\frac{1}{4}\left((\frac{r}{\tau})-\frac{(\frac{r}{\tau})^2}{1-\rho+\sqrt{(1-\rho)^2-(\frac{r}{\tau})^2}}-\rho\right)^2} \leq \frac{r}{\tau}$$

for some $\frac{r}{\tau} \in (0,1]$,

$$0 \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{(1-\rho)^2-\rho(2\tau-\rho)}{2}} - \sqrt{\frac{2((1-\rho)^2+\rho(2-\rho))}{2+\sqrt{4-2((1-\rho)^2+\rho(2-\rho))}}}\right)$$

For $\mathrm{Rips}(\mathscr{X}, r)$, it is the maximum value ρ that satisfies both conditions below:

$$0 \leq \sqrt{\frac{d+1}{2d}}(1-\rho) - \frac{1}{2}\sqrt{\frac{2\left(\frac{1}{4}(1-\rho)^2 + \rho(2-\rho)\right)}{1+\sqrt{1-\left(\frac{1}{4}(1-\rho)^2+\rho(2-\rho)\right)}}},$$

$$0 \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{\frac{d+1}{2d}(1-\rho)^2 - \rho(2\tau-\rho)}{2}} - \sqrt{\frac{2(\frac{d+1}{2d}(1-\rho)^2+\rho(2-\rho))}{2+\sqrt{4-2\left(\frac{d+1}{2d}(1-\rho)^2+\rho(2-\rho)\right)}}}\right).$$

With the aid of computer program, we can check that $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ is homotopy equivalent to $\mathbb{X}$ when $\frac{d_H(\mathbb{X},\mathscr{X})}{\tau} \leq 0.1096\cdots$, and $\mathrm{Rips}(\mathscr{X}, r)$ is homotopy equivalent to $\mathbb{X}$ when $\frac{d_H(\mathbb{X},\mathscr{X})}{\tau} \leq 0.06700\cdots$.

Finally, we emphasize that our result also allows the radii $\{r_x\}_{x\in\mathscr{X}}$ to vary across the observation points $x \in \mathscr{X}$. Considering different radii is practically of interest if each data point has different importance. For example, one might want to use large radius on the flat and sparse region, while to use small radius on the spiky and dense region. However, there are significant technical difficulties to allow different radius per each data point. As can be seen in Figure 2, too uneven distribution of radii might lead to nonhomotopic between the Čech complex or the Rips complex with the target space. It has been studied in Chazal and Lieutier [2008] for the union of balls under the reach condition, but not the Rips complex or under μ-reach case. Theorem 14 or Corollary 16 first tackles this homotopy reconstruction problem with different radii. for the Rips complex or under the *mu*-reach condition.

References

E. Aamari, J. Kim, F. Chazal, B. Michel, A. Rinaldo, and L. Wasserman. Estimating the Reach of a Manifold. *ArXiv e-prints*, May 2017.

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. ISSN 0925-7721. doi: https://doi.org/10.1016/j.comgeo.2012.02.009. URL `http://www.sciencedirect.com/science/article/pii/S0925772112001423`. 27th Annual Symposium on Computational Geometry (SoCG 2011).

A. Björner. Handbook of combinatorics (vol. 2). chapter Topological Methods, pages 1819–1872. MIT Press, Cambridge, MA, USA, 1995.

Frédéric Chazal and André Lieutier. Weak feature size and persistant homology: computing homology of solids in $\mathbb{R}^n$ from noisy data samples. In *Computational geometry (SCG'05)*, pages 255–262. ACM, New York, 2005. doi: 10.1145/1064092.1064132. URL https://doi.org/10.1145/1064092.1064132.

Frédéric Chazal and André Lieutier. Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Comput. Geom.*, 40(2):156–170, 2008. ISSN 0925-7721. doi: 10.1016/j.comgeo.2007.07.001. URL https://doi.org/10.1016/j.comgeo.2007.07.001.

Frédéric Chazal and Steve Yann Oudot. Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the Twenty-fourth Annual Symposium on Computational Geometry*, SCG '08, pages 232–241, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-071-5. doi: 10.1145/1377676.1377719. URL http://doi.acm.org/10.1145/1377676.1377719.

Frédéric Chazal, David Cohen-Steiner, André Lieutier, and Boris Thibert. Shape smoothing using double offsets. In Bruno Levy and Dinesh Manocha, editors, *Solid and Physical Modeling, June, 2007*, pages 183–192, Beijing, Chine, June 2007. ACM.

Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in euclidean space. *Discrete & Computational Geometry*, 41(3):461–479, 2009.

Frédéric Chazal, Brittany T Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance to a measure and kernel distance. *Journal of Machine Learning Research*, 18(159):1–40, 2018.

Antonio Cuevas. Set estimation: Another bridge between statistics and geometry. *Bol. Estad. Investig. Oper*, 25(2):71–85, 2009.

Antonio Cuevas and Alberto Rodríguez-Casal. On boundary estimation. *Advances in Applied Probability*, 36(2):340–354, 2004.

Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.*, 7:339–358, 2007. ISSN 1472-2747. doi: 10.2140/agt.2007.7.339. URL https://doi.org/10.2140/agt.2007.7.339.

Tamal K. Dey. *Curve and surface reconstruction: algorithms with mathematical analysis*, volume 23 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2007. ISBN 978-0-521-86370-4; 0-521-86370-8.

Herbert Edelsbrunner and Ernst P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, January 1994. ISSN 0730-0301. doi: 10.1145/174462.156635. URL http://doi.acm.org/10.1145/174462.156635.

Herbert Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959. ISSN 0002-9947.

Karsten Grove. Critical point theory for distance functions. In *Differential geometry: Riemannian geometry (Los Angeles, CA, 1990)*, volume 54 of *Proc. Sympos. Pure Math.*, pages 357–385. Amer. Math. Soc., Providence, RI, 1993.

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008. ISSN 0179-5376. doi: 10.1007/s00454-008-9053-2. URL http://dx.doi.org/10.1007/s00454-008-9053-2.

V. Robins, J. D. Meiss, and E. Bradley. Computing connectedness: disconnectedness and discreteness. *Phys. D*, 139(3-4):276–300, 2000. ISSN 0167-2789. doi: 10.1016/S0167-2789(99)00228-6. URL https://doi.org/10.1016/S0167-2789(99)00228-6.

A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.*, 65(8):1067–1144, 2012. ISSN 0010-3640. doi: 10.1002/cpa.21395. URL http://dx.doi.org/10.1002/cpa.21395.

J.-L. Starck, V. J. Martínez, D. L. Donoho, O. Levi, P. Querre, and E. Saar. Analysis of the spatial distribution of galaxies by multiscale methods. *EURASIP J. Appl. Signal Process.*, (15):2455–2469, 2005. ISSN 1110-8657.

A Projection on a Positive Reach set

We first start with a simple calculation of the distance from one point of a triangle to another point lying on the opposite side.

Claim 18. *Let $x, x_1, \dots, x_k \in \mathbb{R}^d$ and $\lambda_1, \dots, \lambda_k \in [0,1]$ with $\sum \lambda_i = 1$. Then*

$$\left\|\sum \lambda_i x_i - x\right\| = \sqrt{\sum \lambda_i \|x_i - x\|^2 - \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2}.$$

Proof of Claim 18. The distance from $\sum \lambda_i x_i$ to x can be expanded as

$$\begin{aligned} &\left\|\sum \lambda_i x_i - x\right\|^2 \\ &= \left\|\sum \lambda_i (x_i - x)\right\|^2 \\ &= \sum \lambda_i^2 \|x_i - x\|^2 + 2 \sum \lambda_i \lambda_j \left\langle x_i - x, x_j - x \right\rangle. \end{aligned} \tag{19}$$

Then applying the identity $2\left\langle x_i - x, x_j - x\right\rangle = \|x_i - x\|^2 + \left\|x_j - x\right\|^2 - \left\|x_i - x_j\right\|^2$ to (19) gives

$$\left\|\sum \lambda_i x_i - x\right\|^2 = \sum \lambda_i \|x_i - x\|^2 - \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2,$$

and the claim directly follows. □

Given a line segment whose endpoints are on A, Lemma 19 gives a bound on the distance from any point on that segment to its projection on A. See Figure 4.

Lemma 19. *Let $A \subset \mathbb{R}^d$ be a set with reach $\tau > 0$, and let $x_1, \dots, x_k \in \mathbb{R}^d$ with $d_A(x_i) < \tau$. Let $\lambda_1, \dots, \lambda_k \in [0,1]$ with $\sum \lambda_i = 1$, and let $u := \sum \lambda_i x_i$ be such that $d(u, A) < \tau$. Then*

$$\|u - \pi_A(u)\| \le \tau - \sqrt{\left(\sum \lambda_i (\tau - d_A(x_i))^2 - \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2\right)_+}.$$

Proof of Lemma 19. If $\pi_A(u) = u$, then there is nothing to prove. Now, suppose $\pi_A(u) \neq u$, and let $w := \pi_A(u) + \tau \frac{u - \pi_A(u)}{\|u - \pi_A(u)\|}$. Then, we have that $\|w - \pi_A(u)\| = \tau$ and $w - u = \left(\frac{\tau - \|\pi_A(u) - u\|}{\|\pi_A(u) - u\|}\right)(u - \pi_A(u))$. Since $\|u - \pi_A(u)\| < \tau$, it follows that $\langle w - u, u - \pi_A(u)\rangle = \|w - u\| \, \|u - \pi_A(u)\|$ and $\|u - \pi_A(u)\| + \|w - u\| = \|w - \pi_A(u)\|$, as in Figure 4. Since $\pi_A\left(\pi_A(u) + \|u - \pi_A(u)\| \frac{u - \pi_A(u)}{\|u - \pi_A(u)\|}\right) = \pi_A(u)$ and $\pi_A(u) + r\frac{u - \pi_A(u)}{\|u - \pi_A(u)\|} \notin Med(A)$ for all $r < \tau$, Theorem 4.8 (2) and (6) in Federer [1959] imply that

$$\pi_A\left(\pi_A(u) + r\frac{u - \pi_A(u)}{\|u - \pi_A(u)\|}\right) = \pi_A(u)$$

for all $r < \tau$. Thus, $\mathbb{B}(w, \tau) \cap A = \emptyset$ and we can conclude that $\|w - x_i\| \ge \tau - d_A(x_i)$. Applying Claim 18 to $\|w - u\|$ yields that

$$\begin{aligned} \|w - u\| &= \sqrt{\sum \lambda_i \|x_i - w\|^2 - \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2} \\ &\ge \sqrt{\left(\sum \lambda_i (\tau - d_A(x_i))^2 - \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2\right)_+}. \end{aligned}$$

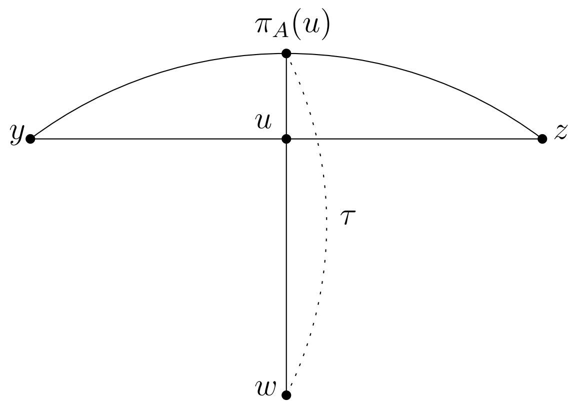

Figure 4: Bound on the distance from any point on the segment to its projection on A, as in Lemma 19.

Then, $\|u-\pi_A(u)\| = \|w-\pi_A(u)\| - \|w-u\|$ implies that

$$\|u-\pi_A(u)\| \le \tau - \sqrt{\left(\sum \lambda_i(\tau - d_A(x_i))^2 - \sum \lambda_i\lambda_j \left\|x_i - x_j\right\|^2\right)_+}.$$

□

In order to prove the contractibility, we will show that if two points in A belong to a ball of sufficiently small radius then the projection onto A of any point along the path connecting them also lies inside the ball. In particular, we will show that when two points $y,z \in A$ are in a ball with the center x, then for any point u on the line segment connecting y and z, its projection $\pi_A(u)$ to A is also in the ball by bounding its distance to x as

$$\|x-\pi_A(u)\| \le \sqrt{\lambda \left\|y-x\right\|^2 + (1-\lambda)\left\|z-x\right\|^2}, \tag{20}$$

where $\lambda \in [0,1]$ is a number satisfying $u := \lambda y + (1-\lambda)z$.

Extension of Theorem 4.8 (7) in Federer [1959].

Claim 20. *Let $A \subset \mathbb{R}^d$ be a set, and $x, y \in \mathbb{R}^d$ with $x, y \in Unp(A)$, and $d_A(x) < \tau_y := \mathrm{reach}(A, \pi_A(y))$. Then*

$$\langle y-\pi_A(y), \pi_A(y)-x\rangle \ge -\frac{\left\|x-\pi_A(y)\right\|^2 d_A(x)}{2\tau_y} - d_A(x)d_A(y)\left(1-\frac{d_A(x)}{2\tau_y}\right).$$

Proof. Assume $\pi_A(y) \ne y$ and let

$$v = \frac{y-\pi_A(y)}{d_A(y)}, \quad S = \{t > 0 : \pi_A(\pi_A(y)+tv) = \pi_A(y)\}.$$

Then $\|y-\pi_A(y)\| \in S$ implies $\sup S > 0$, and then Theorem 4.8 (6) in Federer [1959] implies that

$$\sup S \ge \tau_y.$$

Now, if $t \in S$, then $\pi_A(\pi_A(y)+tv) = \pi_A(y)$ implies

$\|\pi_A(y)+tv-x\| \geq \|\pi_A(y)+tv-\pi_A(x)\| - \|x-\pi_A(x)\| \geq \|\pi_A(y)+tv-\pi_A(y)\| - d_A(x) = t - d_A(x).$

Then additionally under $t \geq d_A(x)$, squaring and expanding gives

$$\|\pi_A(y)-x\|^2 + 2t\langle v, \pi_A(y)-x\rangle + t^2 \geq (t-d_A(x))^2.$$

Rearranging this gives

$$\langle v, \pi_A(y)-x\rangle \geq -\frac{\|\pi_A(y)-x\|^2}{2t} - d_A(x)\left(1-\frac{d_A(x)}{2t}\right).$$

Hence under the condition $\tau_y > d_A(x)$, applying $v = \frac{y-\pi_A(y)}{\|y-\pi_A(y)\|}$ and sending $t \to \tau_y$ gives

$$\langle y-\pi_A(y), \pi_A(y)-x\rangle \geq -\frac{\|\pi_A(y)-x\|^2 d_A(y)}{2\tau_y} - d_A(x)d_A(y)\left(1-\frac{d_A(x)}{2\tau_y}\right).$$

□

The situation in which the center of the ball lies on the set A, we will rely on a slight modification of Theorem 4.8 (8) in Federer [1959].

Lemma 21. *Let $A \subset \mathbb{R}^d$ be a set, and $x, y \in \mathbb{R}^d$ with $x, y \in Unp(A)$, and $d_A(x) < \tau_y := \mathrm{reach}(A, \pi_A(y))$. Then*

$$\|x-\pi_A(y)\| \leq \sqrt{\frac{\tau_y}{\tau_y - d_A(y)}\left(\|x-y\|^2 - d_A(y)\left(d_A(y) - 2d_A(x) + \frac{d_A(x)^2}{\tau_y}\right)\right)}.$$

Proof of Lemma 21. From above claim,

$$\langle y-\pi_A(y), \pi_A(y)-x\rangle \geq -\frac{\|\pi_A(y)-x\|^2 d_A(y)}{2\tau_y} - d_A(x)d_A(y)\left(1-\frac{d_A(x)}{2\tau_y}\right).$$

Hence $\|x-y\|^2$ can be expanded and lower bounded as

$$\begin{aligned}\|x-y\|^2 &= \|y-\pi_A(y)\|^2 + \|\pi_A(y)-x\|^2 + 2\langle y-\pi_A(y), \pi_A(y)-x\rangle \\ &\geq d_A(y)^2 + \|\pi_A(y)-x\|^2\left(1-\frac{d_A(y)}{\tau_y}\right) - 2d_A(x)d_A(y)\left(1-\frac{d_A(x)}{2\tau_y}\right).\end{aligned}$$

By rearranging, we obtain that

$$\|x-\pi_A(y)\| \leq \sqrt{\frac{\tau_y}{\tau_y - d_A(y)}\left(\|x-y\|^2 - d_A(y)\left(d_A(y) - 2d_A(x) + \frac{d_A(x)^2}{\tau_y}\right)\right)}.$$

□

Lemma 22. *Let $A \subset \mathbb{R}^d$ be a set, and $x, u \in \mathbb{R}^d$ with $x, u \in Unp(A)$, and $d_A(x) < \tau_u := \mathrm{reach}(A, \pi_A(u))$. Suppose that $\|x-u\| \leq \tau_u - d_A(x)$. Then*

$$
\begin{aligned}
&\|x-\pi_A(u)\| \\
&\leq \sqrt{\frac{2\tau\left(\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))\right)}{\tau + \sqrt{\tau^2 - \left(\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))\right)}} - d_A(x)\,(2\tau - d_A(x))} \\
&\leq \min\left\{\sqrt{\frac{\tau_u\left(2\,\|u-x\|^2 + d_A(x)\,(2\tau_u - d_A(x))\right)}{\tau_u + \sqrt{\tau_u^2 - \left(\|u-x\|^2 + d_A(x)\,(2\tau_u - d_A(x))\right)}}},\right. \\
&\quad \left(\left(\sqrt{\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))} + \frac{\sqrt{2}-1}{\tau}\left(\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))\right)\right)^2\right. \\
&\quad \left.\left. - d_A(x)\,(2\tau - d_A(x))\right)^{\frac{1}{2}}\right\}.
\end{aligned}
$$

Proof of Lemma 22. First, considering Lemma 21 gives

$$
\|x-\pi_A(u)\| \leq \sqrt{\frac{\tau}{\tau - d_A(u)}\left(\|u-x\|^2 - d_A(u)\left(d_A(u) - 2d_A(x) + \frac{d_A(x)^2}{\tau}\right)\right)}. \tag{21}
$$

For further bounding (21), let $\tilde{r}_u := \|u-x\|$ and $\tilde{r}_A := 2d_A(x) - \frac{d_A(x)^2}{\tau}$ for convenience. Now, consider the function

$$
t \in [0,\tau) \mapsto f(t) := \frac{\tau(\tilde{r}_u^2 - t^2 + \tilde{r}_A t)}{\tau - t}.
$$

Then its derivative is

$$
f'(t) = \frac{\tau(t^2 - 2\tau t + \tilde{r}_u^2 + \tilde{r}_A \tau)}{(\tau - t)^2}.
$$

Hence $f'(t) \geq 0$ if and only if $t \leq \tau - \sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A \tau)}$, and f attains its maximum at $t = \tau - \sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A \tau)}$. Hence, for all $t \in [0,\tau)$,

$$
f(t) \leq f\left(\tau - \sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A \tau)}\right) = \frac{\tau\left(2\tilde{r}_u^2 - 2\tau^2 + 2\tilde{r}_A \tau + (2\tau - \tilde{r}_A)\sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A \tau)}\right)}{\sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A \tau)}}.
$$

Hence (21) is correspondingly further upper bounded as

$$\begin{aligned}
\|x-\pi_A(u)\| &\leq \sqrt{\frac{\tau}{\tau-d_A(u)}\left(\|u-x\|^2-d_A(u)\left(d_A(u)-2d_A(x)+\frac{d_A(x)^2}{\tau}\right)\right)} \\
&= \sqrt{\frac{\tau\left(2\tilde{r}_u^2-2\tau^2+2\tilde{r}_A\tau+(2\tau-\tilde{r}_A)\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}\right)}{\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}}} \\
&= \sqrt{\tau\left(2\tau-\tilde{r}_A-2\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}\right)} \\
&= \sqrt{\frac{2\tau(\tilde{r}_u^2+\tilde{r}_A\tau)}{\tau+\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}}-\tilde{r}_A\tau} \\
&= \sqrt{\frac{2\tau\left(\|u-x\|^2+d_A(x)\,(2\tau-d_A(x))\right)}{\tau+\sqrt{\tau^2-\left(\|u-x\|^2+d_A(x)\,(2\tau-d_A(x))\right)}}-d_A(x)\,(2\tau-d_A(x))}.
\end{aligned}$$

And this is again further bounded in twofold. First, we bound as

$$\begin{aligned}
\|x-\pi_A(u)\| &\leq \sqrt{\frac{2\tau(\tilde{r}_u^2+\tilde{r}_A\tau)}{\tau+\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}}-\tilde{r}_A\tau} \\
&= \sqrt{\frac{2\tau\tilde{r}_u^2+\left(\tau-\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}\right)\tilde{r}_A\tau}{\tau+\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}}} \\
&\leq \sqrt{\frac{\tau(2\tilde{r}_u^2+\tilde{r}_A\tau)}{\tau+\sqrt{\tau^2-(\tilde{r}_u^2+\tilde{r}_A\tau)}}} \\
&= \sqrt{\frac{\tau\left(2\|u-x\|^2+d_A(x)\,(2\tau-d_A(x))\right)}{\tau+\sqrt{\tau^2-\left(\|u-x\|^2+d_A(x)\,(2\tau-d_A(x))\right)}}}.
\end{aligned}$$

Second, let $\tilde{r}:=\sqrt{\|u-x\|^2+d_A(x)\,(2\tau-d_A(x))}$, then $\tilde{r}\leq\tau$ and

$$\begin{aligned}
\sqrt{\frac{2\tau\left(\|u-x\|^2+d_A(x)\,(2\tau-d_A(x))\right)}{\tau+\sqrt{\tau^2-\left(\|u-x\|^2+d_A(x)\,(2\tau-d_A(x))\right)}}} &= \sqrt{\frac{2\tau\tilde{r}^2}{\tau+\sqrt{\tau^2-\tilde{r}^2}}} = \sqrt{2\tau(\tau-\sqrt{\tau^2-\tilde{r}^2})} \\
&= \sqrt{\tau}\left(\sqrt{\tau+\tilde{r}}-\sqrt{\tau-\tilde{r}}\right).
\end{aligned}$$

Now, let $f(t):=t\left(1+\frac{\sqrt{2}-1}{\tau}t\right)-\sqrt{\tau}\left(\sqrt{\tau+t}-\sqrt{\tau-t}\right)$ on $t\in[0,\tau]$, then $f'(t)=\left(1+\frac{2(\sqrt{2}-1)}{\tau}t\right)-\frac{\sqrt{\tau}}{2}\left(\frac{1}{\sqrt{\tau+t}}+\frac{1}{\sqrt{\tau-t}}\right)$, $f''(t)=\frac{2(\sqrt{2}-1)}{\tau}-\frac{\sqrt{\tau}}{4}\left((\tau-t)^{-\frac{3}{2}}-(\tau+t)^{-\frac{3}{2}}\right)$, and $f'''(t)=-\frac{3\sqrt{\tau}}{8}\left((\tau-t)^{-\frac{5}{2}}+\right.$

$(\tau+t)^{-\frac{5}{2}}\Big) < 0$. Hence $f''(t)$ is monotonically decreasing on $[0,\tau)$, and $f''(0) > 0$, $\lim_{t\to\tau} f''(t) < 0$ implies that there exists $t_2 \in (0,\tau)$ such that $f''(t) \geq 0$ if $t \leq t_2$ and $f''(t) \leq 0$ if $t \geq t_2$. Then, $f'(t)$ is increasing on $[0,t_2]$ and is decreasing on $[t_2,\tau)$. Then $f'(0) = 0$ and $\lim_{t\to\tau} f'(t) < 0$ implies that there exists $t_1 \in (t_2,\tau)$ such that $f'(t) \geq 0$ if $t \leq t_1$ and $f'(t) \leq 0$ if $t \geq t_1$. Since $f(0) = f(\tau) = 0$, this implies that $f(t) \geq 0$ for all $t \in [0,\tau]$. Hence

$$\sqrt{\frac{2\tau\left(\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))\right)}{\tau + \sqrt{\tau^2 - \left(\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))\right)}}} = \sqrt{\tau}\left(\sqrt{\tau+\tilde{r}} - \sqrt{\tau-\tilde{r}}\right) \leq \left(\tilde{r} + \frac{\sqrt{2}-1}{\tau}\tilde{r}^2\right),$$

and applying this gives

$$\begin{aligned}
&\|x - \pi_A(u)\| \\
&\leq \sqrt{\frac{2\tau(\tilde{r}_u^2 + \tilde{r}_A\tau)}{\tau + \sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A\tau)}} - \tilde{r}_A\tau} \\
&\leq \sqrt{\left(\sqrt{\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))} + \frac{\sqrt{2}-1}{\tau}\left(\|u-x\|^2 + d_A(x)\,(2\tau - d_A(x))\right)\right)^2 - d_A(x)\,(2\tau - d_A(x))}.
\end{aligned}$$

□

Lemma 23. *Let $A \subset \mathbb{R}^d$ be a set with reach $\tau > 0$. Let $x_i \in \mathbb{R}^d$, $\lambda_i \in [0,1]$ with $\sum \lambda_i = 1$, and let $u := \sum \lambda_i x_i$. Let $x \in \mathbb{R}^d$ with $\|x - x_i\| < \sqrt{(\tau - d_A(x))^2 + (\tau - d_A(x_i))^2}$. Then*

$$\|x - \pi_A(u)\| \leq \min\left\{\sqrt{\sum \lambda_i\left(\|x_i - x\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right)},\right.$$
$$\left.\sqrt{2\,\|x-u\|^2 + d_A(x)\,(2\tau - d_A(x))}\right\}.$$

Suppose further that $\sum \lambda_i\left(\|x_i - u\|^2 + 2\tau d_A(x_i) - d_A(x_i)^2\right) \leq \|x-u\|^2 + 2\tau d_A(x) - d_A(x)^2$, then

$$\begin{aligned}
&\|x - \pi_A(u)\| \\
&\leq \Bigg(\sum \lambda_i\left(\|x_i - x\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right) \\
&- \left((\tau - d_A(x))^2 + \sum \lambda_i\left((\tau - d_A(x_i))^2 - \|x_i - x\|^2\right)\right)\left(\frac{\tau}{\sqrt{\sum \lambda_i(\tau - d_A(x_i))^2 - \sum \lambda_i\,\|x_i - u\|^2}} - 1\right)\Bigg)^{\frac{1}{2}}.
\end{aligned}$$

Proof of Lemma 23. Let $r := \sqrt{\sum \lambda_i\,\|x_i - x\|^2}$, then $r < \sqrt{(\tau - d_A(x))^2 + \sum \lambda_i\,(\tau - d_A(x_i))^2}$. Then from Claim 18,

$$\begin{aligned}\|x-u\| &= \sqrt{\sum \lambda_i \|x_i - x\|^2 - \sum \lambda_i \|x_i - u\|^2} \\ &= \sqrt{r^2 - \sum \lambda_i \|x_i - u\|^2}.\end{aligned} \tag{22}$$

Again applying Claim 18 gives that $\sum \lambda_i \|x_i - u\|^2 = \sum \lambda_i \|x_i - u\|^2$, and hence

$$\|x-u\|^2 + \sum \lambda_i \|x_i - u\|^2 = r^2.$$

Then $r < \sqrt{(\tau - d_A(x))^2 + \sum \lambda_i (\tau - d_A(x_i))^2}$ implies that either $\|x-u\| < \tau - d_A(x)$ or $\|x_i - u\| < \tau - d_A(x_i)$ for some i,and hence $d_A(u) < \tau$ holds for either case. Hence applying Lemma 19 implies that

$$\|u - \pi_A(u)\| \le \tau - \sqrt{\left(\tau^2 - \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) - \sum \lambda_i \|x_i - u\|^2\right)_+}. \tag{23}$$

Now, considering Lemma 21 gives

$$\|x - \pi_A(u)\| \le \sqrt{\frac{\tau}{\tau - d_A(u)} \left(\|u - x\|^2 - d_A(u) \left(d_A(u) - 2d_A(x) + \frac{d_A(x)^2}{\tau} \right) \right)}. \tag{24}$$

For further bounding (24), let

$$\begin{aligned}\tilde{r}_u &:= \|u - x\|, \\ \tilde{r}_A &:= 2d_A(x) - \frac{d_A(x)^2}{\tau}, \\ \tilde{r}_{yz} &:= \sqrt{\sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) + \sum \lambda_i \|x_i - u\|^2},\end{aligned}$$

for convenience. Now, consider the function

$$t \in \left[0, \tau - \sqrt{\left(\tau^2 - \tilde{r}_{yz}^2\right)_+}\right] \mapsto f(t) := \frac{\tau(\tilde{r}_u^2 - t^2 + \tilde{r}_A t)}{\tau - t}.$$

Then its derivative is

$$f'(t) = \frac{\tau(t^2 - 2\tau t + \tilde{r}_u^2 + \tilde{r}_A \tau)}{(\tau - t)^2}.$$

Hence if $\tilde{r}_u^2 + \tilde{r}_A \tau \ge \tau^2$ then $f'(t) \ge 0$ holds for all $t \in \left[0, \tau - \sqrt{\left(\tau^2 - \tilde{r}_{yz}^2\right)_+}\right]$, and if $\tilde{r}_u^2 + \tilde{r}_A \tau \le \tau^2$ then $f'(t) \ge 0$ if and only if $t \le \tau - \sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A \tau)}$. Hence if $\tilde{r}_{yz}^2 \le \tau^2, \tilde{r}_u^2 + \tilde{r}_A \tau$ then f attains its maximum at $t = \tau - \sqrt{\tau^2 - \tilde{r}_{yz}^2}$, and if $\tilde{r}_u^2 + \tilde{r}_A \tau \le \tau^2, \tilde{r}_{yz}^2$ then f attains its maximum at $t = \tau - \sqrt{\tau^2 - (\tilde{r}_u^2 + \tilde{r}_A \tau)}$. Hence by letting $\tilde{r} := \min\{\sqrt{\tilde{r}_u^2 + \tilde{r}_A \tau}, \tilde{r}_{yz}\}$, for all $t \in \left[0, \tau - \sqrt{\left(\tau^2 - \tilde{r}_{yz}^2\right)_+}\right]$,

$$f(t) \le f\left(\tau - \sqrt{\tau^2 - \tilde{r}^2}\right) = \frac{\tau\left(\tilde{r}_u^2 + \tilde{r}^2 - 2\tau^2 + \tilde{r}_A \tau + (2\tau - \tilde{r}_A)\sqrt{\tau^2 - \tilde{r}^2}\right)}{\sqrt{\tau^2 - \tilde{r}^2}}.$$

Hence with (23), (24) is correspondingly further upper bounded as

$$\begin{aligned}
\|x-\pi_A(u)\| &\leq \sqrt{\frac{\tau}{\tau - d_A(u)}\left(\|u-x\|^2 - d_A(u)\left(d_A(u) - 2d_A(x) + \frac{d_A(x)^2}{\tau}\right)\right)} \\
&\leq \sqrt{\frac{\tau\left(\tilde{r}_u^2 + \tilde{r}^2 - 2\tau^2 + \tilde{r}_A\tau + (2\tau - \tilde{r}_A)\sqrt{\tau^2 - \tilde{r}^2}\right)}{\sqrt{\tau^2 - \tilde{r}^2}}} \\
&= \sqrt{\tilde{r}_u^2 + \tilde{r}^2 - (2\tau^2 - \tilde{r}_u^2 - \tilde{r}^2 - \tilde{r}_A\tau)\left(\frac{\tau}{\sqrt{\tau^2 - \tilde{r}^2}} - 1\right)}. \qquad (25)
\end{aligned}$$

Now, $\tilde{r}_u^2 + \tilde{r}^2$ is upper bounded as

$$\begin{aligned}
\tilde{r}_u^2 + \tilde{r}^2 &\leq \tilde{r}_u^2 + \tilde{r}_{yz}^2 \\
&= \|u-x\|^2 + \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) + \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2 \\
&= \left(r^2 - \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2\right) + \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) + \sum \lambda_i \lambda_j \left\|x_i - x_j\right\|^2 \\
&= r^2 + \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) \\
&= \sum \lambda_i \left(\|x_i - x\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right). \\
&\leq \sum \lambda_i (\tau - d_A(x_i))^2
\end{aligned}$$

And hence

$$\begin{aligned}
&\tilde{r}_u^2 + \tilde{r}^2 + \tilde{r}_A\tau \\
&\leq r^2 + \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) + d_A(x)(2\tau - d_A(x)) \\
&\leq (\tau - d_A(x))^2 + \sum \lambda_i (\tau - d_A(x_i))^2 + \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) + d_A(x)(2\tau - d_A(x)) \\
&\leq 2\tau^2.
\end{aligned}$$

applying this to (25) gives the desired upper bound as

$$\begin{aligned}
\|x-\pi_A(u)\| \leq \sqrt{\tilde{r}_u^2 + \tilde{r}^2} &\leq \sqrt{2\tilde{r}_u^2 + \tilde{r}_A\tau} \\
&\leq \sqrt{2\|u-x\|^2 + d_A(x)(2\tau - d_A(x))}.
\end{aligned}$$

And also,

$$\begin{aligned}
\|x-\pi_A(u)\| &\leq \sqrt{\tilde{r}_u^2 + \tilde{r}^2} \\
&\leq \sqrt{r^2 + \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i))} \\
&= \sqrt{\sum \lambda_i \left(\|x_i - x\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right)}.
\end{aligned}$$

Now, under further assumption that $\sum\lambda_i\left(\|u-x_i\|^2+2\tau d_A(x_i)-d_A(x_i)^2\right)\le\|u-x\|^2+2\tau d_A(x)-d_A(x)^2$, this is equivalent to $\tilde{r}_{yz}^2\le\tilde{r}_u^2+\tilde{r}_A\tau$, so

$$\tilde{r}_u^2+\tilde{r}^2=\tilde{r}_u^2+\tilde{r}_{yz}^2=\sum\lambda_i\left(\|x_i-x\|^2+d_A(x_i)(2\tau-d_A(x_i))\right)$$

and

$$\begin{aligned}
&2\tau^2-\tilde{r}_u^2-\tilde{r}^2-\tilde{r}_A\tau\\
&=2\tau^2-\tilde{r}_u^2-\tilde{r}_{yz}^2-\tilde{r}_A\tau\\
&=2\tau^2-\sum\lambda_i\left(\|x_i-x\|^2+d_A(x_i)(2\tau-d_A(x_i))\right)-\sum d_A(x)(2\tau-d_A(x))\\
&=(\tau-d_A(x))^2+\sum\lambda_i(\tau-d_A(x_i))^2-\sum\lambda_i\|x_i-x\|^2.
\end{aligned}$$

And hence (20) is expanded as

$$\begin{aligned}
&\|x-\pi_A(u)\|\\
&\le\sqrt{\tilde{r}_u^2+\tilde{r}_{yz}^2-(\tau^2-\tilde{r}_u^2-\tilde{r}_{yz}^2)\left(\frac{\tau}{\sqrt{\tau^2-\tilde{r}_{yz}^2}}-1\right)}\\
&=\Bigg(\sum\lambda_i\left(\|x_i-x\|^2+d_A(x_i)(2\tau-d_A(x_i))\right)\\
&\quad-\left((\tau-d_A(x))^2+\sum\lambda_i\left((\tau-d_A(x_i))^2-\|x_i-x\|^2\right)\right)\left(\frac{\tau}{\sqrt{\sum\lambda_i(\tau-d_A(x_i))^2-\sum\lambda_i\|x_i-u\|^2}}-1\right)\Bigg)^{\frac{1}{2}}.
\end{aligned}$$

$$\begin{aligned}
&\|x-\pi_A(u)\|\\
&\le\sqrt{\tilde{r}_u^2+\tilde{r}_{yz}^2-(\tau^2-\tilde{r}_u^2-\tilde{r}_{yz}^2)\left(\frac{\tau}{\sqrt{\tau^2-\tilde{r}_{yz}^2}}-1\right)}\\
&=\sqrt{r^2+\sum\lambda_i d_A(x_i)(2\tau-d_A(x_i))-(\tau^2-\tilde{r}_u^2-\tilde{r}^2)\frac{\tilde{r}^2}{\sqrt{\tau^2-\tilde{r}^2}\left(\tau+\sqrt{\tau^2-\tilde{r}^2}\right)}}\\
&=\sqrt{\tilde{r}_u^2+\frac{\left(\tau+\frac{\tilde{r}_u^2}{\sqrt{\tau^2-\tilde{r}^2}}\right)\tilde{r}^2}{\tau+\sqrt{\tau^2-\tilde{r}^2}}}.
\end{aligned}$$

Then $f(t)=\tilde{r}_u^2+\frac{\left(\tau+\frac{\tilde{r}_u^2}{\sqrt{\tau^2-t^2}}\right)t^2}{\tau+\sqrt{\tau^2-t^2}}$ is an increasing function on $t\in[0,\tau)$, so $\tilde{r}\le\tilde{r}_{yz}$ implies

Now, let $f(t) = \tilde{r}_u^2 + t^2 - (\tau^2 - \tilde{r}_u^2 - t^2)\left(\frac{\tau}{\sqrt{\tau^2-t^2}} - 1\right)$ for $t \in [0, \tau)$, then

$$\begin{aligned} f(t) &= \sqrt{\tilde{r}_u^2 + t^2 - (\tau^2 - \tilde{r}_u^2 - t^2)\frac{t^2}{\sqrt{\tau^2-t^2}\left(\tau + \sqrt{\tau^2-t^2}\right)}} \\ &= \sqrt{\tilde{r}_u^2 + \frac{\left(\tau + \frac{\tilde{r}_u^2}{\sqrt{\tau^2-t^2}}\right)t^2}{\tau + \sqrt{\tau^2-t^2}}}. \end{aligned}$$

Then it is apparent that $f(t)$ is an increasing function of t, so $f(\tilde{r}) \le f(\tilde{r}_{yz})$ holds. Hence applying this gives

$$\begin{aligned} &\|x - \pi_A(u)\| \\ &\le \sqrt{\tilde{r}_u^2 + \tilde{r}_{yz}^2 - (\tau^2 - \tilde{r}_u^2 - \tilde{r}_{yz}^2)\left(\frac{\tau}{\sqrt{\tau^2 - \tilde{r}_{yz}^2}} - 1\right)} \\ &= \left(\sum \lambda_i \left(\|x_i - x\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right)\right. \\ &\quad - \left(\tau^2 - \sum \lambda_i \left(\|x_i - x\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right)\right) \\ &\quad \times \left.\left(\frac{\tau}{\sqrt{\tau^2 - \sum \lambda_i d_A(x_i)(2\tau - d_A(x_i)) - \sum \lambda_i \lambda_j \|x_i - x_j\|^2}} - 1\right)\right)^{\frac{1}{2}}. \end{aligned}$$

□

B Proofs for Section 3

Theorem 5. *Let $\mathbb{X} \subset \mathbb{R}^d$ be a subset with reach $\tau > 0$ and let $\mathscr{X} \subset \mathbb{R}^d$ be a set of points. Let $\{r_x > 0 : x \in \mathscr{X}\}$ be a set of radii indexed by $x \in \mathscr{X}$. Then, if $r_x \le \sqrt{\tau^2 + (\tau - d_{\mathbb{X}}(x))^2}$ for all $x \in \mathscr{X}$, any nonempty intersection of restricted balls $\bigcap_{x \in I} \mathbb{B}_{\mathbb{X}}(x, r_x)$ for $I \subset \mathscr{X}$ is contractible.*

Proof of Theorem 5. Fix $x \in I$ and write $B_x := \mathbb{B}_{\mathbb{R}^d}(x, r_x)$. Let $y_1, y_2 \in B_x \cap \mathbb{X}$. Let $l : [0,1] \to B_x$ with $l(t) = ty_1 + (1-t)y_2$ be the line segment joining y_1 and y_2, and define the curve $\gamma_{y_1,y_2} : [0,1] \to \mathbb{X}$ as $\gamma(t) = \pi_{\mathbb{X}}(l(t))$. Theorem 4.8 (4) in Federer [1959] implies that γ is continuous.

Now we argue that $\gamma_{y_1,y_2}(t) \in B_x$ for $t \in [0,1]$. For notational convenience, let $\gamma = \gamma_{y_1,y_2}$ here. Then since $y_1, y_2 \in B_x \cap \mathbb{X}$, $d_{\mathbb{X}}(y_1) = d_{\mathbb{X}}(y_2) = 0$ and $\|x - y_1\|, \|x - y_2\| < r_x$, hence applying Lemma 21 gives,

$$\|x - \gamma(t)\| \le \sqrt{t\|x - y_1\|^2 + (1-t)\|x - y_2\|^2} < r_x.$$

Hence $\gamma(t) = \gamma_{y_1,y_2}(t) \in B_x$.

Now, fix $y_0 \in \bigcap_{x\in I} B_x \cap \mathbb{X}$, and define the map $F : \left(\bigcap_{x\in I} B_x \cap \mathbb{X}\right) \times [0,1] \to \left(\bigcap_{x\in I} B_x \cap \mathbb{X}\right)$ as $F(y,t) = \gamma_{y_0,y}(t)$. Since $\gamma_{y_0,y}$ is continuous and, as we have shown above, $\gamma_{y_0,y}(t) \in \bigcap_{x\in I} B_x \cap \mathbb{X}$ for all t, F is a well-defined, continuous map. As $F(y,0) = y$ and $F(y,y_0) = y_0$ for all $y \in \bigcap_{x\in I} B_x \cap A$, the intersection $\bigcap_{x\in I} B_x \cap \mathbb{X}$ is contractible. □

C Proofs for Section 4

Proof of Theorem 8. For each $x \in \mathbb{X}_r \backslash \mathbb{X}$, there exists a vector field $W_x : U_x \to \mathbb{R}^d$ defined on a neighborhood U_x of x such that for any $y \in U_x$,

$$\langle W_x(y), \nabla(y)\rangle \leq -\frac{\mu}{2} \|\nabla(y)\|.$$

From the open cover $\{U_x\} \supset \mathbb{X}_r \backslash \mathbb{X}$, take a locally finite covering $\{U_{x_i}\}$, and take a smooth partition of unity $\{\rho_i\}$ subordinate to $\{U_{x_i}\}$. Then we define a vector field $W : \mathbb{X}_r \backslash \mathbb{X} \to \mathbb{R}^d$ as

$$W = \sum \rho_i W_i.$$

Then note that for any $x \in \mathbb{X}_r \backslash \mathbb{X}$,

$$\langle W(x), \nabla(x)\rangle = \left\langle \sum \rho_i W_i(x), \nabla(x)\right\rangle \leq -\frac{\mu}{2} \|\nabla(x)\|.$$

Now, let $\mathbb{D} \subset (\mathbb{X}_r \backslash \mathbb{X}) \times [0,\infty)$ be the set of points (x,s) such that an integral curve ψ^x, i.e. $\psi^x(0) = x$ and $(\psi^x)'(t) = W(\psi^x(t))$, exists on an interval containing $[0,s]$. let $\psi : \mathbb{D} \subset (\mathbb{X}_r \backslash \mathbb{X}) \times [0,\infty) \to \mathbb{X}_r \backslash \mathbb{X}$ be the flow of W on $\mathbb{D}$, i.e. $\psi(x,s) = \psi^x(s)$. From W being smooth, $\mathbb{D}$ is an open set in $(\mathbb{X}_r \backslash \mathbb{X}) \times [0,\infty)$ and ψ is smooth on $\mathbb{D}$ from the fundamental theorem on flows.

Now, note that

$$\begin{aligned} \frac{d}{ds} d_{\mathbb{X}}(\psi(x,s)) &\leq \left\langle \nabla(\psi(x,s)), \frac{d}{ds}\psi(x,t)\right\rangle = \langle \nabla(\psi(x,s)), W(\psi(x,s))\rangle \\ &\leq -\frac{\mu}{2} \|\nabla(x)\| \leq -\frac{\mu^2}{2}. \end{aligned}$$

Hence $d_{\mathbb{X}}$ is strictly decreasing along the integral curve ψ^x.

Let $\mathbb{D}_x := \{s \in [0,\infty) : (x,s) \in \mathbb{D}\}$, then $\mathbb{D}_x$ is a connected open set in $[0,\infty)$, and above implies that $s_x := \sup \mathbb{D}_x \leq \frac{d_{\mathbb{X}}(x)}{\mu^2/2} \leq \frac{2\tau_\mu}{\mu^2} < \infty$, so $\mathbb{D}_x = [0,s_x)$. And since $\|(\psi^x)'(s)\| = \|W(\psi^x(s))\| < \infty$, $\psi^x([0,s_x))$ is a curve of finite length. Hence there exists a limit $\psi^x(s_x) := \lim_{s\to s_x} \psi^x(s)$, and we can extend ψ^x continuously on $[0,\infty)$ by $\psi^x(s) = \psi^x(s_x)$ if $s \geq s_x$. Using this, we extend ψ on $\mathbb{X}_r \times [0,\infty)$ as

$$\psi(x,s) = \begin{cases} \psi^x(s), & \text{if } x \in \mathbb{X}_r \backslash \mathbb{X}, \\ x, & \text{if } x \in \mathbb{X}. \end{cases}$$

Now we show that ψ is continuous on $\mathbb{X}_r \times [0,\infty)$. If $(x_0,s_0) \in \mathbb{D}$, then ψ is smooth on an open set $\mathbb{D}$, so ψ is continuous at (x_0,s_0). When $x_0 \in \mathbb{X}_r \backslash \mathbb{X}$ and $s_0 \notin \mathbb{D}_x$, let $\rho_x > 0$ be small enough so that $B(x_0,\rho_x) \times \left[0, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right] \subset \mathbb{D}$ and $|\psi(x,s) - \psi(x_0,s_0)| \leq \frac{\mu^2\varepsilon}{8}$ for all $(x,s) \in B(x_0,\rho_x) \times \left[0, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right]$. Then for any $x \in B(x_0,\rho_x)$,

$$\begin{aligned}
& d_{\mathbb{X}}\left(\psi\left(x, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right)\right) \\
& \leq \left| d_{\mathbb{X}}\left(\psi\left(x, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right)\right) - d_{\mathbb{X}}\left(\psi\left(x_0, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right)\right)\right| \\
& \quad + \left| d_{\mathbb{X}}\left(\psi\left(x_0, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right)\right) - d_{\mathbb{X}}(\psi(x_0,s_{x_0}))\right| + d_{\mathbb{X}}(\psi(x_0,s_{x_0})) \\
& \leq \frac{\mu^2\varepsilon}{4},
\end{aligned}$$

and hence $\left| s_x - \left(s_{x_0} - \frac{\mu^2}{8}\varepsilon\right)\right| \leq \frac{\varepsilon}{2}$. Hence for $\|x - x_0\| < \rho_x$ and $\|s - s_0\| < \frac{\mu^2}{8}\varepsilon$,

$$\begin{aligned}
& \|\psi(x,s) - \psi(x_0,s_0)\| \\
& \leq \left\|\psi(x,s) - \psi\left(x, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right)\right\| + \left\|\psi\left(x, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right) - \psi\left(x_0, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right)\right\| \\
& \quad + \left\|\psi\left(x_0, s_{x_0} - \frac{\mu^2}{8}\varepsilon\right) - \psi(x_0,s_0)\right\| \\
& \leq \frac{\varepsilon}{2} + \frac{\mu^2\varepsilon}{8} + \frac{\mu^2\varepsilon}{8} < \varepsilon.
\end{aligned}$$

Hence ψ is continuous at (x_0,s_0). When $x_0 \in \mathbb{X}$, let $x \in B(x_0, \frac{\mu^2}{4}\varepsilon)$. Then $d_{\mathbb{X}}(x) < \frac{\mu^2}{4}\varepsilon$, so $s_x < \frac{\varepsilon}{2}$. Then for any $x \in B(x_0, \frac{\mu^2}{4}\varepsilon)$ and for any $s \geq 0$,

$$\begin{aligned}
\|\psi(x,s) - \psi(x_0,s)\| &= \|\psi(x,s) - \psi(x,0)\| + \|x - x_0\| \\
&< \frac{\varepsilon}{2} + \frac{\mu^2\varepsilon}{4} \leq \varepsilon.
\end{aligned}$$

Hence ψ is continuous at (x_0,s_0).

Now, we define the deformation retract $H : \mathbb{X}_r \times [0,1] \to \mathbb{X}_r$ as $H(x,t) = \psi\left(x, \frac{2}{\mu^2}t\right)$. Then, $H(x,0) = x$ for all $x \in \mathbb{X}_r$ and $H(x,t) = x$ for all $x \in \mathbb{X}$. Also, since $s_x \leq \frac{2}{\mu^2}$ for all x, $H(x,1) = \psi\left(x, \frac{2}{\mu^2}\right) \in \mathbb{X}$ for all $x \in \mathbb{X}_r$. Also, H is continuous since ψ is continuous. Hence H gives the deformation retract from $\mathbb{X}_r$ to $\mathbb{X}$.

$\square$

Proof of Lemma 9. For $x \in \mathbb{X}^{\complement}$, let $\tau_x := \mathrm{reach}(\mathbb{X}, \pi_{\mathbb{X}}(x))$, and let the function $\psi : \mathbb{X}^{\complement} \to (\mathbb{X}^r)^{\complement}$ be

$$\psi(x) = x + \frac{(r - \|x - \pi_{\mathbb{X}}(x)\|)_+}{\|x - \pi_{\mathbb{X}}(x)\|}(x - \pi_{\mathbb{X}}(x)).$$

Then, note that ψ is continuous on $\mathbb{X}^{\complement}$. Also,

$$
\begin{aligned}
\|\psi(x)-\pi_{\mathbb{X}}(x)\| &= \left(1+\frac{(r-\|x-\pi_{\mathbb{X}}(x)\|)_{+}}{\|x-\pi_{\mathbb{X}}(x)\|}\right)\|x-\pi_{\mathbb{X}}(x)\| \\
&= \max\{r, \|x-\pi_{\mathbb{X}}(x)\|\} \geq r,
\end{aligned}
$$

so $\psi(x) \notin \mathbb{X}^{r}$, i.e. $\psi(x) \in (\mathbb{X}^{r})^{\complement}$. And if $x \in (\mathbb{X}^{r})^{\complement}$, then $\|x-\pi_{\mathbb{X}}(x)\| \geq r$, so $\psi(x) = x$ on $(\mathbb{X}^{r})^{\complement}$.

Now, we define the deformation retract $H : \mathbb{X}^{\complement} \times [0,1] \to \mathbb{X}^{\complement}$ as $H(x,t) = (1-t)x + t\psi(x)$. Then, $H(x,0) = x$ for all $x \in \mathbb{X}^{\complement}$, $H(x,t) = x$ for all $x \in (\mathbb{X}^{r})^{\complement}$, and $H(x,1) \in (\mathbb{X}^{r})^{\complement}$ for all $x \in \mathbb{X}^{r}$. Also, H is continuous since ψ is continuous. Hence H gives the deformation retract from $\mathbb{X}^{\complement}$ to $(\mathbb{X}^{r})^{\complement}$.

$\square$

Proof of Corollary 10. Since $\mathbb{X}$ has a positive μ-reach τ^{μ} and $r \leq \tau^{\mu}$, Theorem 8 implies that $\mathbb{X}$ and $\mathbb{X}^{r}$ are homotopy equivalent. Then, by Theorem 4.1 in Chazal et al. [2007],

$$
\mathrm{reach}((\mathbb{X}^{r})^{\complement}) \geq \mu r.
$$

Then, from Lemma 9 and $s \leq \mu r$, $\mathbb{X}^{r} = ((\mathbb{X}^{r})^{\complement})^{\complement}$ and $\mathbb{X}^{r,s} = (((\mathbb{X}^{r})^{\complement})^{s})^{\complement}$ are homotopy equivalent. And hence, $\mathbb{X}$ and $\mathbb{X}^{r,s}$ are homotopy equivalent as well. Also, by Theorem 4.1 in Chazal et al. [2007],

$$
\mathrm{reach}((((\mathbb{X}^{r})^{\complement})^{s})^{\complement}) \geq s.
$$

$\square$

D Proofs for Section 5

Lemma 24. *Let X be a paracompact space, and let $\mathscr{U} = \{U_i\}_{i\in I}$, $\mathscr{U}' = \{U_i'\}_{i\in I'}$, be good covers of X, with $I \subset I'$ and $U_i \subset U_i'$ for all $i \in I$. Let $\mathscr{S}$ be a simplicial complex satisfying $\mathscr{N}\mathscr{U} \subset \mathscr{S} \subset \mathscr{N}\mathscr{U}'$. Suppose there exists $\rho : \mathscr{S} \to \mathscr{N}\mathscr{U}$ such that $\iota_{\mathscr{N}\mathscr{U}\to\mathscr{S}} \circ \rho : S \to S$ is homotopy equivalent to $id_{\mathscr{S}} : \mathscr{S} \to \mathscr{S}$. Then X and $\mathscr{S}$ are homotopy equivalent.*

Proof of Lemma 24. Let $\phi : X \to \mathscr{N}\mathscr{U}$ and $\psi : \mathscr{N}\mathscr{U}' \to X$ be the maps giving homotopy equivalence of the Nerve Theorem. Then from Lemma 3.3 in Chazal and Oudot [2008], the following diagram commutes at homotopy level:

$$
\begin{tikzcd}
 & X \arrow[dl, "\phi"'] & \\
\mathscr{N}\mathscr{U} \arrow[rr, "\iota_{\mathscr{N}\mathscr{U}\to\mathscr{N}\mathscr{U}'}"'] & & \mathscr{N}\mathscr{U}' \arrow[ul, "\psi'"']
\end{tikzcd}
\quad . \tag{26}
$$

To show homotopic equivalence, we define maps $\Phi : X \to \mathscr{S}$ and $\Psi : \mathscr{S} \to X$ by $\Phi = \iota_{\mathscr{N}\mathscr{U}\to\mathscr{S}} \circ \phi$ and $\Psi = \psi' \circ \iota_{\mathscr{S}\to\mathscr{N}\mathscr{U}'}$. Then, we need to show that $\Psi \circ \Phi \simeq id_X$ and $\Phi \circ \Psi \simeq id_{\mathscr{S}}$ on homotopy level.

For $\Psi \circ \Phi \simeq id_X$, consider the diagram below:

$$
\begin{tikzcd}
& X \arrow[dl, "\phi"] & \\
\mathcal{N}\mathcal{U} \arrow[rr, "\iota_{\mathcal{N}\mathcal{U}\to\mathcal{N}\mathcal{U}'}"] \arrow[dr, "\iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}}"'] & & \mathcal{N}\mathcal{U}' \arrow[ul, "\psi'"'] \\
& \mathcal{S} \arrow[ur, "\iota_{\mathcal{S}\to\mathcal{N}\mathcal{U}'}"] &
\end{tikzcd}
.
$$

Then from (26), $\psi' \circ \iota_{\mathcal{N}\mathcal{U}\to\mathcal{N}\mathcal{U}'} \circ \phi \simeq id_X$ holds, and hence

$$
\begin{aligned}
\Psi \circ \Phi &= \psi' \circ \iota_{\mathcal{S}\to\mathcal{N}\mathcal{U}'} \circ \iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}} \circ \phi \\
&= \psi' \circ \iota_{\mathcal{N}\mathcal{U}\to\mathcal{N}\mathcal{U}'} \circ \phi \\
&\simeq id_X.
\end{aligned}
$$

And hence $\Psi \circ \Phi \simeq id_X$ on homotopy level.

For $\Phi \circ \Psi \simeq id_{\mathcal{S}}$, consider the diagram below:

$$
\begin{tikzcd}
& X \arrow[dl, "\phi"] & \\
\mathcal{N}\mathcal{U} \arrow[rr, "\iota_{\mathcal{N}\mathcal{U}\to\mathcal{N}\mathcal{U}'}"] \arrow[dr, shift right, "\iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}}"'] & & \mathcal{N}\mathcal{U}' \arrow[ul, "\psi'"'] \\
& \mathcal{S} \arrow[ur, "\iota_{\mathcal{S}\to\mathcal{N}\mathcal{U}'}"] \arrow[ul, shift right, "\rho"'] &
\end{tikzcd}
.
$$

Then $\iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}} \circ \rho \simeq id_{\mathcal{S}}$ from the condition, and from (26), $\phi \circ \psi' \circ \iota_{\mathcal{N}\mathcal{U}\to\mathcal{N}\mathcal{U}'} \simeq id_{\mathcal{N}\mathcal{U}}$ holds, and hence

$$
\begin{aligned}
\Phi \circ \Psi &= \iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}} \circ \phi \circ \psi' \circ \iota_{\mathcal{S}\to\mathcal{N}\mathcal{U}'} \\
&\simeq \iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}} \circ \phi \circ \psi' \circ \iota_{\mathcal{S}\to\mathcal{N}\mathcal{U}'} \circ \iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}} \circ \rho \\
&= \iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}} \circ \phi \circ \psi' \circ \iota_{\mathcal{N}\mathcal{U}\to\mathcal{N}\mathcal{U}'} \circ \rho \\
&\simeq \iota_{\mathcal{N}\mathcal{U}\to\mathcal{S}} \circ \rho \\
&\simeq id_{\mathcal{S}}.
\end{aligned}
$$

And hence $\Phi \circ \Psi \simeq id_{\mathcal{S}}$ on homotopy level.

□

Lemma 11. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a set of points. Let $\{r_x > 0 : x \in \mathcal{X}\}$ be a set of radii indexed by $x \in \mathcal{X}$. Then,*

$$
\check{C}ech_{\mathbb{R}^d}(\mathcal{X}, r) \subset Rips(\mathcal{X}, r) \subset \check{C}ech_{\mathbb{R}^d}\left(\mathcal{X}, \sqrt{\frac{2d}{d+1}}r\right).
$$

Proof of Lemma 11. The first inclusion $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r) \subset \mathrm{Rips}(\mathscr{X},r)$ is trivial.

For the second inclusion, the equivalent statement is as follows: if $\mathscr{X} = \{x_0,\ldots,x_k\} \subset \mathbb{R}^d$ satisfies that $\left\|x_i - x_j\right\| < r_i + r_j$ for all $0 \le i,j \le k$, then the intersection of the balls $\bigcap_{i=0}^k \mathbb{B}_{\mathbb{R}^d}\left(x_i, \sqrt{\frac{2d}{d+1}} r_{x_i}\right)$ is nonempty.

We first prove this for the case $k \le d$. Consider a function $f : \mathbb{R}^d \to \mathbb{R}$ defined as

$$f(y) = \max_{0 \le i \le k} \frac{\|x_i - y\|}{r_{x_i}}.$$

This is continuous, and $f(y) \to \infty$ as $y \to \infty$, so f has a global minimum $f(y_0)$. Then $y_0 \in \mathscr{C}(\mathscr{X})$, since otherwise $\left\|x_i - \pi_{\mathscr{C}(\mathscr{X})}(y_0)\right\| < \|x_i - y_0\|$ for each $x_i \in \mathscr{X}$, contradicting the minimality of y_0.

Let $\hat{x}_i := x_i - y_0$ be the translated x_i's. We can find a convex combination $(a_0/r_{x_0})\hat{x}_0 + \cdots + (a_k/r_{x_k})\hat{x}_k = 0$ for some $j \le k$, after relabeling so that $a_0 > 0$ is the largest among $a_0,\ldots,a_k$ and all a_i's are nonnegative. Then $-\hat{x}_0 = \sum_{i=1}^k \frac{a_i/r_{x_i}}{a_0/r_{x_0}} \hat{x}_i$, and so

$$-r_{x_0}^2 f(y_0)^2 = -\|\hat{x}_0\|^2 = \sum_{i=1}^k \frac{a_i/r_{x_i}}{a_0/r_{x_0}} \langle \hat{x}_0, \hat{x}_i \rangle .$$

Then at least one i should satisfy $(a_i/a_0)\langle \hat{x}_0, \hat{x}_i \rangle \le -\frac{r_{x_0} r_{x_i} f(y_0)^2}{k}$,which can be weakened to $\langle \hat{x}_0, \hat{x}_i \rangle \le -\frac{r_{x_0} r_{x_i} f(y_0)^2}{d}$. Putting these together gives

$$\begin{aligned} f(y_0)^2 (r_{x_0}^2 + \frac{2}{d} r_{x_0} r_{x_i} + r_{x_i}^2) &\le \|\hat{x}_0\|^2 - 2\langle \hat{x}_0, \hat{x}_i \rangle + \|\hat{x}_i\|^2 \\ &= \|\hat{x}_0 - \hat{x}_i\|^2 = (r_{x_0} + r_{x_i})^2. \end{aligned}$$

Then from AM-GM inequality,

$$\begin{aligned} r_{x_0}^2 + \frac{2}{d} r_{x_0} r_{x_i} + r_{x_i}^2 &= \frac{d-1}{2d}(r_{x_0}^2 + r_{x_i}^2) + \left(\frac{d+1}{2d}(r_{x_0}^2 + r_{x_i}^2) + \frac{2}{d} r_{x_0} r_{x_i}\right) \\ &\ge \frac{d-1}{d} r_{x_0} r_{x_i} + \left(\frac{d+1}{2d}(r_{x_0}^2 + r_{x_i}^2) + \frac{2}{d} r_{x_0} r_{x_i}\right) = \frac{d+1}{2d}(r_{x_0} + r_{x_i})^2. \end{aligned}$$

Hence combining these gives

$$f(y_0)^2 \frac{d+1}{2d}(r_{x_0} + r_{x_i})^2 \le f(y_0)^2 (r_{x_0}^2 + \frac{2}{d} r_{x_0} r_{x_i} + r_{x_i}^2) \le (r_{x_0} + r_{x_i})^2,$$

and hence

$$f(y_0) \le \sqrt{\frac{2d}{d+1}}.$$

Therefore $y_0 \in \bigcap_{i=0}^k \mathbb{B}_{\mathbb{R}^d}\left(x_i, \sqrt{\frac{2d}{d+1}} r_{x_i}\right)$, i.e. $\bigcap_{i=0}^k \mathbb{B}_{\mathbb{R}^d}\left(x_i, \sqrt{\frac{2d}{d+1}} r_{x_i}\right)$ is nonempty.

For the case $k > d$, the result follows by the Helly's theorem. $\square$

Lemma 12. *Let $\mathbb{X} \subset \mathbb{R}^d$ be a subset with reach $\tau > 0$ and let $\mathscr{X} \subset \mathbb{R}^d$ be a set of points. Let $\{r_x > 0 : x \in \mathscr{X}\}$ be a set of radii indexed by $x \in \mathscr{X}$. Then,*

$$\check{C}ech_{\mathbb{X}}(\mathscr{X}, r) \subset \check{C}ech_{\mathbb{R}^d}(\mathscr{X}, r) \subset \check{C}ech_{\mathbb{X}}(\mathscr{X}, r'),$$

where $r' = \{r'_x > 0 : x \in \mathscr{X}\}$ with $r'_x = \sqrt{2r_x^2 + d_{\mathbb{X}}(x)(2\tau - d_{\mathbb{X}}(x))}$.

Proof of Lemma 12. For $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}, r) \subset \check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$, this is implied from $\mathbb{B}_{\mathbb{X}}(x, r_x) \subset \mathbb{B}_{\mathbb{R}^d}(x, r_x)$.

For $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r) \subset \check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}, r')$, let $[x_1, \ldots, x_k] \in \check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$, then there exists $\lambda_1, \ldots, \lambda_k \in [0, 1]$ with $\sum \lambda_i = 1$ such that

$$u := \sum \lambda_i x_i \in \bigcap_{i=1}^{k} \mathbb{B}_{\mathbb{R}^d}(x_i, r_{x_i}).$$

Then from Lemma 21,

$$\begin{aligned}
\|x_i - \pi_{\mathbb{X}}(u)\| &\le \sqrt{2\|u - x_i\|^2 + d_{\mathbb{X}}(x_i)(2\tau - d_{\mathbb{X}}(x_i))} \\
&< \sqrt{2r_{x_i}^2 + d_{\mathbb{X}}(x_i)(2\tau - d_{\mathbb{X}}(x_i))}.
\end{aligned}$$

And hence

$$\pi_{\mathbb{X}}(u) \in \bigcap_{j=1}^{k} \mathbb{B}_{\mathbb{X}}\left(x_j, \sqrt{2r_{x_j}^2 + d_{\mathbb{X}}(x_i)(2\tau - d_{\mathbb{X}}(x_i))}\right).$$

Therefore, $[x_1, \ldots, x_k] \in \check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}, r')$ as well, and hence $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r) \subset \check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}, r')$ holds. $\square$

Claim 25. *Let $\mathbb{X} \subset \mathbb{R}^d$ be a subset with reach $\tau > 0$ and let $\mathscr{X} \subset \mathbb{R}^d$ be a set of points. Let $\{r_x > 0 : x \in \mathscr{X}\}$ be a set of radii indexed by $x \in \mathscr{X}$. For each $\sigma \subset \mathscr{X}$, let $r_\sigma := Rad(\sigma)$ and $\varepsilon_\sigma := \max\{d_{\mathbb{X}}(x) : x \in \mathscr{X}\}$. Suppose $\mathbb{X} \subset \bigcup_{x \in \mathscr{X}} \mathbb{B}_{\mathbb{X}}(x, \varepsilon)$. Then for each $\sigma \subset \mathscr{X}$, there exists $b_\sigma \in \mathscr{X}$ and $y_\sigma \in \mathscr{X}$ such that*

1. *If $r_\sigma < \tau - \varepsilon_\sigma$, then*

$$\begin{aligned}
\|x_i - y_\sigma\| &< \sqrt{\frac{2\tau\left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right)}{\tau + \sqrt{\tau^2 - \left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right)}} - \varepsilon_\sigma(2\tau - \varepsilon_\sigma) + \varepsilon} \\
&\le \sqrt{\left(\sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)} + \frac{\sqrt{2} - 1}{\tau}\left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right)\right)^2 - \varepsilon_\sigma(2\tau - \varepsilon_\sigma) + \varepsilon}.
\end{aligned}$$

2. *If $r_\sigma < \tau - \varepsilon_\sigma$ and suppose $\varsigma \subset \sigma$, then*

$$\begin{aligned}
\left\|y_\varsigma - y_\sigma\right\| &< \sqrt{\frac{2\tau\left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right)}{\tau + \sqrt{\tau^2 - \left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right)}}} + 2\varepsilon \\
&\le \sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)} + \frac{\sqrt{2} - 1}{\tau}\left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right) + \varepsilon.
\end{aligned}$$

3. *If $r_\sigma < \tau - \varepsilon_\sigma$ and suppose $\varepsilon + \sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma) - \frac{1}{4}\left(r_{\min} - \frac{r_\sigma^2}{\tau - \varepsilon_\sigma + \sqrt{(\tau-\varepsilon_\sigma)^2 - r_\sigma^2}} - \varepsilon_\sigma - \varepsilon\right)^2} \leq r_{\min}$, then $\sigma \in \check{C}ech_{\mathbb{R}^d}(\mathscr{X}, r)$ implies that $[x_1 \cdots x_j y_{[x_1\cdots x_j]} y_{[x_1\cdots x_{j+1}]} \cdots y_{[x_1\cdots x_k]}] \in \check{C}ech_{\mathbb{R}^d}(\mathscr{X}, r)$.*

4. *If $r_\sigma < \tau - \varepsilon_\sigma$ and suppose $\varepsilon \leq r_{\min} - \frac{1}{2}\sqrt{\frac{2\tau\left(\frac{d}{2(d+1)}r_{\max}^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau + \sqrt{\tau^2 - \left(\frac{d}{2(d+1)}r_{\max}^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}}}$, then $\sigma \in Rips(\mathscr{X}, r)$ implies that $[x_1 \cdots x_j y_{[x_1\cdots x_j]} y_{[x_1\cdots x_{j+1}]} \cdots y_{[x_1\cdots x_k]}] \in Rips(\mathscr{X}, r)$.*

5. *If $r_\sigma < \tau - \varepsilon_\sigma$ and suppose $\varepsilon \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}r_\sigma - \sqrt{\frac{\tau\left(2r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau + \sqrt{\tau^2 - \left(r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}} + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)}\right)$, then $r_\sigma = Rad(\sigma)$ implies that $[x_1 y_{[x_1 x_2]} \cdots y_{[x_1\cdots x_k]}] \in Rips\left(\mathscr{X}, \sqrt{\frac{(d+1)}{2d}}r_\sigma\right)$, i.e., $\|x_1 - y_\sigma\| < \sqrt{\frac{2(d+1)}{d}}r_\sigma$ and $\|y_\varsigma - y_\sigma\| < \sqrt{\frac{2(d+1)}{d}}r_\sigma$ if $\varsigma \subset \sigma$.*

Proof. We choose $b_\sigma \in \mathbb{R}^d$ be the center of the smallest enclosing ball of σ, so that $\max_i \|b_\sigma - x_i\| = r_\sigma$. And by covering condition, we can choose $y_\sigma \in \mathscr{X}$ be satisfying $\|y_\sigma - \pi_{\mathbb{X}}(b_\sigma)\| < \varepsilon$.

1.

Note that $\|x_i - b_\sigma\| \leq \tau - \varepsilon_\sigma \leq \tau - d_A(x_i)$, so applying Lemma 22 gives

$$\begin{aligned}
&\|x_i - \pi_{\mathbb{X}}(b_\sigma)\| \\
&\leq \sqrt{\frac{2\tau\left(\|x_i - b_\sigma\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right)}{\tau + \sqrt{\tau^2 - \left(\|x_i - b_\sigma\|^2 + d_A(x_i)(2\tau - d_A(x_i))\right)}} - d_A(x_i)(2\tau - d_A(x_i))} \\
&\leq \sqrt{\frac{\tau\,(2r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma))}{\tau + \sqrt{\tau^2 - (r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma))}} - \varepsilon_\sigma(2\tau - \varepsilon_\sigma)}. \\
&\leq \sqrt{\left(\sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\,(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma))\right)^2 - \varepsilon_\sigma(2\tau - \varepsilon_\sigma)}
\end{aligned}$$

And hence

$$\begin{aligned}
\|x_i - y_\sigma\| &\leq \|x_i - \pi_{\mathbb{X}}(b_\sigma)\| + \|\pi_{\mathbb{X}}(b_\sigma) - y_\sigma\| \\
&< \sqrt{\frac{2\tau\,(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma))}{\tau + \sqrt{\tau^2 - (r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma))}} - \varepsilon_\sigma(2\tau - \varepsilon_\sigma)} + \varepsilon \\
&\leq \sqrt{\left(\sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\,(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma))\right)^2 - \varepsilon_\sigma(2\tau - \varepsilon_\sigma)} + \varepsilon.
\end{aligned}$$

2.

Without loss of generality assume that $\varsigma = [x_1 \cdots x_j]$ and $\sigma = [x_1 \cdots x_k]$ with $j \leq k$. Then above argument implies that $\|x_i - \pi_{\mathbb{X}}(b_\sigma)\| \leq \sqrt{\frac{2\tau\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau+\sqrt{\tau^2-\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}} - \varepsilon_\sigma(2\tau-\varepsilon_\sigma)}$ for each $i = 1, \ldots, j$. Also,

$$\begin{aligned}
\|\pi_{\mathbb{X}}(b_\sigma) - x_i\| &< \sqrt{\frac{2\tau\left((\tau-\varepsilon_\sigma)^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau+\sqrt{\tau^2-\left((\tau-\varepsilon_\sigma)^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}} - \varepsilon_\sigma(2\tau-\varepsilon_\sigma)} \\
&= \sqrt{\tau^2+(\tau-\varepsilon_\sigma)^2} \leq \sqrt{(\tau - d_{\mathbb{X}}(\pi_{\mathbb{X}}(b_\sigma)))^2 + (\tau - d_{\mathbb{X}}(x_i))^2}.
\end{aligned}$$

Hence Lemma 23 is applicable and gives

$$\begin{aligned}
\left\|\pi_{\mathbb{X}}(b_\sigma) - \pi_{\mathbb{X}}(b_\varsigma)\right\| &\leq \sqrt{\sum \lambda_i \left(\|x_i - \pi_{\mathbb{X}}(b_\sigma)\|^2 + d_{\mathbb{X}}(x_i)(2\tau - d_{\mathbb{X}}(x_i))\right)} \\
&\leq \sqrt{\sum \lambda_i \left(\|x_i - \pi_{\mathbb{X}}(b_\sigma)\|^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)} \\
&< \sqrt{\frac{2\tau\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau+\sqrt{\tau^2-\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}}} \\
&\leq \sqrt{r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right).
\end{aligned}$$

And hence

$$\begin{aligned}
\left\|y_\varsigma - y_\sigma\right\| &\leq \left\|y_\varsigma - \pi_{\mathbb{X}}(b_\varsigma)\right\| + \left\|\pi_{\mathbb{X}}(b_\varsigma) - \pi_{\mathbb{X}}(b_\sigma)\right\| + \|\pi_{\mathbb{X}}(b_\sigma) - y_\sigma\| \\
&< \sqrt{\frac{2\tau\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau+\sqrt{\tau^2-\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}}} + 2\varepsilon \\
&\leq \sqrt{r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right) + 2\varepsilon.
\end{aligned}$$

3.

We need to find $\tilde{b} \in \mathbb{R}^d$ and check that $\left\|x_i - \tilde{b}\right\| < r_{x_i}$ and $\left\|y_\varsigma - \tilde{b}\right\| < r_{y_\varsigma}$. First, note that for any $\varsigma \subset \sigma$, $\left\|b_\varsigma - \pi_{\mathbb{X}}(b_\varsigma)\right\|$ is bounded as

$$\left\|b_\varsigma - \pi_{\mathbb{X}}(b_\varsigma)\right\| \leq \tau - \sqrt{\left((\tau-\varepsilon_\sigma)^2 - r_\sigma^2\right)_+} = \frac{r_\sigma^2}{\tau - \varepsilon_\sigma + \sqrt{(\tau-\varepsilon_\sigma)^2 - r_\sigma^2}} + \varepsilon_\sigma.$$

Now, let $t_0 := \frac{1}{2}\left(r_{\min} - \frac{r_\sigma^2}{\tau-\varepsilon_\sigma+\sqrt{(\tau-\varepsilon_\sigma)^2-r_\sigma^2}} - \varepsilon_\sigma - \varepsilon\right)$. We divide into two cases. First, if all ς with $\{x_1, \ldots, x_j\} \subset \varsigma \subset \sigma = \{x_1, \ldots, x_k\}$ satisfy that $\left\|b_\varsigma - b_\sigma\right\| \leq t_0$, then we can set $\tilde{b} := b_\sigma$. For this case, $\|x_i - \tilde{b}\| < r_{x_i}$ and

$$\begin{aligned}
&\|y_\varsigma - \tilde{b}\| \\
&\leq \|y_\varsigma - \pi_{\mathbb{X}}(b_\varsigma)\| + \|\pi_{\mathbb{X}}(b_\varsigma) - b_\varsigma\| + \|b_\varsigma - b_\sigma\| \\
&< \delta + \left(\frac{r_\sigma^2}{\tau - \varepsilon_\sigma + \sqrt{(\tau - \varepsilon_\sigma)^2 - r_\sigma^2}} + \varepsilon_\sigma \right) + \frac{1}{2} \left(r_{\min} - \frac{r_\sigma^2}{\tau - \varepsilon_\sigma + \sqrt{(\tau - \varepsilon_\sigma)^2 - r_\sigma^2}} - \varepsilon_\sigma - \delta \right) \\
&\leq r_{\min} \leq r_{y_\varsigma}.
\end{aligned}$$

For the other case, let $\hat{k} := \max\left\{ j : \left\| b_{[x_1 \cdots x_j]} - b_\sigma \right\| > t_0 \right\}$ be the largest integer that $\left\| b_{[x_1 \cdots x_j]} - b_\sigma \right\| > t_0$, and let $\hat{\sigma} := [x_1 \cdots x_{\hat{k}}]$. Let $\tilde{b} := \frac{t_0}{\|b_\sigma - b_{\hat{\sigma}}\|} b_{\hat{\sigma}} + \left(1 - \frac{t_0}{\|b_\sigma - b_{\hat{\sigma}}\|}\right) b_\sigma$, so that $\|\tilde{b} - b_\sigma\| = t_0$. Then for all ς with $\hat{\sigma} \subsetneq \varsigma \subset \sigma$, $\|\tilde{b} - b_\varsigma\| \leq \|\tilde{b} - b_\sigma\| + \|b_\sigma - b_\varsigma\| \leq 2t_0$ and

$$\begin{aligned}
&\|y_\varsigma - \tilde{b}\| \\
&\leq \|y_\varsigma - \pi_{\mathbb{X}}(b_\varsigma)\| + \|\pi_{\mathbb{X}}(b_\varsigma) - b_\varsigma\| + \|b_\varsigma - b_\sigma\| \\
&< \delta + \left(\frac{r_\sigma^2}{\tau - \varepsilon_\sigma + \sqrt{(\tau - \varepsilon_\sigma)^2 - r_\sigma^2}} + \varepsilon_\sigma \right) + \left(r_{\min} - \frac{r_\sigma^2}{\tau - \varepsilon_\sigma + \sqrt{(\tau - \varepsilon_\sigma)^2 - r_\sigma^2}} - \varepsilon_\sigma - \delta \right) \\
&= r_{\min} < r_{y_\varsigma}.
\end{aligned}$$

Now, note that $r_{\hat{\sigma}} \leq \sqrt{r_\sigma^2 - \|b_\sigma - b_{\hat{\sigma}}\|^2}$, so $x_i \in \mathbb{B}_{\mathbb{R}^d}(b_\sigma, r_\sigma) \cap \mathbb{B}_{\mathbb{R}^d}(b_{\hat{\sigma}}, r_{\hat{\sigma}}) \subset \mathbb{B}_{\mathbb{R}^d}(\tilde{b}, \sqrt{r_\sigma^2 - t_0^2})$ holds. Hence for all $x_1, \ldots, x_j$,

$$\|\tilde{b} - x_i\| < \sqrt{r_\sigma^2 - t_0^2} < r_\sigma.$$

Then Lemma 23 is applicable and gives

$$\begin{aligned}
\|\pi_{\mathbb{X}}(b_\varsigma) - \tilde{b}\| &\leq \sqrt{\sum \lambda_i \left(\|x_i - \tilde{b}\|^2 + d_{\mathbb{X}}(x_i)(2\tau - d_{\mathbb{X}}(x_i)) \right)} \\
&\leq \sqrt{r_\sigma^2 - t_0^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)}.
\end{aligned}$$

Then,

$$\begin{aligned}
\|y_\varsigma - \tilde{b}\| &\leq \|y_\varsigma - \pi_{\mathbb{X}}(b_\varsigma)\| + \|\pi_{\mathbb{X}}(b_\varsigma) - \tilde{b}\| \\
&< \varepsilon + \sqrt{r_\sigma^2 - t_0^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)} \\
&= \varepsilon + \sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma) - \frac{1}{4}\left(r_{\min} - \frac{r_\sigma^2}{\tau - \varepsilon_\sigma + \sqrt{(\tau - \varepsilon_\sigma)^2 - r_\sigma^2}} - \varepsilon_\sigma - \varepsilon \right)^2} \\
&\leq r_{\min} \leq r_\sigma.
\end{aligned}$$

4.

We need to check that $\|x_i - y_\sigma\| < r_{x_i} + r_{y_\sigma}$ and $\|y_\varsigma - y_\sigma\| < r_{y_\varsigma} + r_{y_\sigma}$.
For $\|x_i - y_\sigma\| < r_{x_i} + r_{y_\sigma}$, note that (1) and the condition on ε gives

$$\begin{aligned}\|x_1 - y_\sigma\| &< \sqrt{\frac{\tau\left(2r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau + \sqrt{\tau^2 - (r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma))}} - \varepsilon_\sigma(2\tau-\varepsilon_\sigma) + \varepsilon} \\ &\le 2r_{\min} \le r_{x_i} + r_{y_0}.\end{aligned}$$

For $\|y_\varsigma - y_\sigma\| < r_{y_\varsigma} + r_{y_\sigma}$, note that (2) and the condition on ε gives

$$\begin{aligned}\|y_\varsigma - y_\sigma\| &< \sqrt{\frac{2\tau\left(r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau + \sqrt{\tau^2 - (r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma))}}} + 2\varepsilon \\ &\le 2r_{\min} \le r_{y_\varsigma} + r_{y_0}.\end{aligned}$$

5.

We will show that $[x_1 y_{[x_1x_2]} \cdots y_{[x_1\cdots x_k]}] \in \text{Rips}\left(\mathscr{X}, \sqrt{\frac{2(d+1)}{d}} r_\sigma\right)$, i.e. we will show that if $\varsigma \subset \sigma$, then

$$\begin{aligned}\|x_1 - y_\sigma\| &< \sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\left(r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right) + 2\varepsilon, \\ \|y_\varsigma - y_\sigma\| &< \sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\left(r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right) + 2\varepsilon.\end{aligned}$$

For $\|x_1 - y_\sigma\| < \sqrt{\frac{2(d+1)}{d}} r_\sigma$, note that (1) and the condition on ε gives

$$\begin{aligned}\|x_1 - y_\sigma\| &< \sqrt{\left(\sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\left(r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)\right)^2 - \varepsilon_\sigma(2\tau-\varepsilon_\sigma) + \varepsilon} \\ &< \sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\left(r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right) + 2\varepsilon.\end{aligned}$$

For $\|y_\varsigma - y_\sigma\| < \sqrt{\frac{2(d+1)}{d}} r_\sigma$, note that (2) and the condition on ε gives

$$\|y_\varsigma - y_\sigma\| < \sqrt{r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)} + \frac{\sqrt{2}-1}{\tau}\left(r_\sigma^2 + \varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right) + 2\varepsilon.$$

□

Proof of Theorem 13. For a simplex $\sigma = [x_1, \ldots, x_k] \in \check{\text{C}}\text{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$, let $r_\sigma := Rad(\sigma)$ and $\varepsilon_\sigma := \min\{d_{\mathbb{X}}(x_i) : x_i \in \sigma\}$. Then as long as $r_\sigma < \tau - \varepsilon_\sigma$, Claim 25 (3) asserts that $[x_1, \ldots, x_k y_{[x_1\cdots x_k]}] \in \check{\text{C}}\text{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ holds. Now, define the homotopy map $F_1 : [x_1, \ldots, x_k] \times [0,1] \to [x_1, \ldots, x_k y_{[x_1\cdots x_k]}]$ as

$$F_1(\sum \lambda_i x_i, t) = \sum(\lambda_i - \min \lambda_i) x_i + k \min \lambda_i (t y_\sigma + (1-t)\frac{1}{k}\sum x_i).$$

Then F_1 gives homotopy between $\iota_{\sigma\to\sigma*y_\sigma}$ and $f_1:\sigma\to\sigma*y_\sigma$ defined as $f_1(\sum\lambda_i x_i)=\sum(\lambda_i-\min\lambda_i)x_i+(k\min\lambda_i)y_\sigma$, giving homotopic equivalence between $[x_1,\ldots,x_k]$ and $\sum[x_1\cdots\hat{x}_i\cdots x_k y_{[x_1\cdots x_k]}]$ in $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$. And again, Claim 25 (3) asserts that $[x_1\cdots x_{k-1}y_{[x_1\cdots x_k]}]\in\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$ holds. Now, define the homotopy map $F_2:[x_1\cdots x_{k-1}y_{[x_1\cdots x_k]}]\times[0,1]\to[x_1\cdots x_{k-1}y_{[x_1\cdots x_{k-1}]}y_{[x_1\cdots x_k]}]$ as

$$F_2(\sum\lambda_i x_i+\lambda_k y_{[x_1,\cdots,x_k]},t)=\lambda_k y_{[x_1,\cdots,x_k]}+\sum(\lambda_i-\min\lambda_i)x_i+(k-1)\min\lambda_i(ty_{[x_1\cdots x_{k-1}]}+(1-t)\frac{1}{k-1}\sum x_i).$$

Then F_2 gives homotopy between $\iota_{[x_1\cdots x_{k-1}y_{[x_1\cdots x_k]}]\to[x_1\cdots x_{k-1}y_{[x_1\cdots x_{k-1}]}y_{[x_1\cdots x_k]}]}$ and $f_2:[x_1\cdots x_{k-1}y_{[x_1\cdots x_k]}]\to[x_1\cdots x_{k-1}y_{[x_1\cdots x_{k-1}]}y_{[x_1\cdots x_k]}]$ defined as

$$f_2(\sum\lambda_i x_i+\lambda_k y_{[x_1,\cdots,x_k]})=\lambda_k y_{[x_1,\cdots,x_k]}+\sum(\lambda_i-\min\lambda_i)x_i+(k-1)\min\lambda_i y_{[x_1\cdots x_{k-1}]},$$

giving the homotopic equivalence between $[x_1\cdots x_{k-1}y_{[x_1\cdots x_k]}]$ and $\sum[x_1\cdots\hat{x}_i\cdots x_{k-1}y_{[x_1\cdots x_{k-1}]}y_{[x_1\cdots x_k]}]$ in $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$. By repeating this and concatnating the homotopy maps, we have the homotopy map $F_\sigma:\sigma\times[0,1]\to\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$ giving homotopy between $\iota_{\sigma\to\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)}$ and $f_\sigma:\sigma\to\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$ with $f_\sigma(\sigma)=\sum[x_{\tau(1)}y_{[x_{\tau(1)}x_{\tau(2)}]}\cdots y_{[x_1\cdots x_k]}]$, i.e. F_σ is giving homotopic equivalence between $[x_1,\ldots,x_k]$ and $\sum[x_{\tau(1)}y_{[x_{\tau(1)}x_{\tau(2)}]}\cdots y_{[x_1\cdots x_k]}]$ in $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$. Now, note that for $\sigma,\tau\in\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$, F_σ and F_τ coincides on $\sigma\cap\tau\times[0,1]$, hence by induction, the homotopy map can be extended to F on the enitre Rips complex $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)$ giving homotopy between $id_{\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)}$ and f.

Now, note that $f(\sigma)=\sum[x_{\tau(1)}y_{[x_{\tau(1)}x_{\tau(2)}]}\cdots y_{[x_1\cdots x_k]}]$ consists of simplices with diameters at most $\sqrt{\frac{2\tau\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}{\tau+\sqrt{\tau^2-\left(r_\sigma^2+\varepsilon_\sigma(2\tau-\varepsilon_\sigma)\right)}}}+2\varepsilon$, hence applying Lemma 11 gives that

$$Rad([x_{\tau(1)}y_{[x_{\tau(1)}x_{\tau(2)}]}\cdots y_{[x_1\cdots x_k]}])\leq\sqrt{\frac{d}{2(d+1)}}\left(\sqrt{\frac{2\tau\,(r_\sigma^2+\varepsilon(2\tau-\varepsilon))}{\tau+\sqrt{\tau^2-(r_\sigma^2+\varepsilon(2\tau-\varepsilon))}}}+2\varepsilon\right).$$

Hence by repeating this sufficiently many times (say N), we can guarantee that $f^{(N)}(\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r))$ consists of simplices with their radii (i.e. $Rad(\sigma)$) at most $\tilde{r}_\sigma$, where $\tilde{r}_\sigma$ is the solution of $f(t)=\sqrt{\frac{d}{2(d+1)}}\left(\sqrt{\frac{2\tau\left(t^2+\varepsilon(2\tau-\varepsilon)\right)}{\tau+\sqrt{\tau^2-\left(t^2+\varepsilon(2\tau-\varepsilon)\right)}}}+2\varepsilon\right)-t=0$. Then $f\left(\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}\right)\geq 0$ implies that $\tilde{r}_\sigma\leq\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}$. Hence ε satisfying $\varepsilon\leq\frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}-\sqrt{\frac{2\tau(r_{\min}^2+\varepsilon(2\tau-\varepsilon))}{2\tau+\sqrt{4\tau^2-2\left(r_{\min}^2+\varepsilon(2\tau-\varepsilon)\right)}}}\right)$ implies that $f^{(N)}(\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r))\subset\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}\left(\mathscr{X},\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}\right)$. And Lemma 12 implies that

$$f^{(N)}(\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r))\subset\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}\left(\mathscr{X},\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}\right)\subset\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r_{\min})\subset\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r),$$

i.e. $f^{(N)}: \check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r) \to \check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X}, r)$. Then by construction, $\iota_{\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r)\to\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)} \circ f^{(N)}$ is homotopy equivalent to $id_{\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X},r)}$. Hence by applying Lemma 24, $\mathbb{X}$ and $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}(\mathscr{X}, r)$ are homotopy equivalent.

□

Proof of Theorem 14. For a simplex $\sigma = [x_1, \ldots, x_k] \in \mathrm{Rips}(\mathscr{X}, r)$, let $\rho_\sigma := \frac{1}{2}\max\{\|x_i - x_j\|\}$ and $\varepsilon_\sigma := \min\{d_{\mathbb{X}}(x_i) : x_i \in \sigma\}$. From Lemma 11, we can find $b_\sigma \in conv(x_1, \ldots, x_k)$ such that $\|b_\sigma - x_j\| < \sqrt{\frac{2d}{d+1}}\rho_\sigma =: r_\sigma$ for all $x_j \in \sigma$. Then as long as $r_\sigma < \tau - \varepsilon_\sigma$, Claim 25 (4) asserts that $[x_1, \ldots, x_k y_{[x_1\cdots x_k]}] \in \mathrm{Rips}(\mathscr{X}, r)$ holds. Now, define the homotopy map $F_1 : [x_1, \ldots, x_k] \times [0,1] \to [x_1, \ldots, x_k y_{[x_1\cdots x_k]}]$ as

$$F_1(\sum \lambda_i x_i, t) = \sum(\lambda_i - \min\lambda_i)x_i + k\min\lambda_i(ty_\sigma + (1-t)\frac{1}{k}\sum x_i).$$

Then F_1 gives homotopy between $\iota_{\sigma\to\sigma * y_\sigma}$ and $f_1 : \sigma \to \sigma * y_\sigma$ defined as $f_1(\sum\lambda_i x_i) = \sum(\lambda_i - \min\lambda_i)x_i + (k\min\lambda_i)y_\sigma$, giving homotopic equivalence between $[x_1, \ldots, x_k]$ and $\sum[x_1 \cdots \hat{x}_i \cdots x_k y_{[x_1\cdots x_k]}]$ in $\mathrm{Rips}(\mathscr{X}, r)$. And again, Claim 25 (4) asserts that $[x_1 \cdots x_{k-1} y_{[x_1\cdots x_k]}] \in \mathrm{Rips}(\mathscr{X}, r)$ holds. Now, define the homotopy map $F_2 : [x_1 \cdots x_{k-1} y_{[x_1\cdots x_k]}] \times [0,1] \to [x_1 \cdots x_{k-1} y_{[x_1\cdots x_{k-1}]} y_{[x_1\cdots x_k]}]$ as

$$F_2(\sum\lambda_i x_i + \lambda_k y_{[x_1,\cdots,x_k]}, t) = \lambda_k y_{[x_1,\cdots,x_k]} + \sum(\lambda_i - \min\lambda_i)x_i + (k-1)\min\lambda_i(ty_{[x_1\cdots x_{k-1}]} + (1-t)\frac{1}{k-1}\sum x_i).$$

Then F_2 gives homotopy between $\iota_{[x_1\cdots x_{k-1}y_{[x_1\cdots x_k]}]\to[x_1\cdots x_{k-1}y_{[x_1\cdots x_{k-1}]}y_{[x_1\cdots x_k]}]}$ and $f_2 : [x_1 \cdots x_{k-1} y_{[x_1\cdots x_k]}] \to [x_1 \cdots x_{k-1} y_{[x_1\cdots x_{k-1}]} y_{[x_1\cdots x_k]}]$ defined as

$$f_2(\sum\lambda_i x_i + \lambda_k y_{[x_1,\cdots,x_k]}) = \lambda_k y_{[x_1,\cdots,x_k]} + \sum(\lambda_i - \min\lambda_i)x_i + (k-1)\min\lambda_i y_{[x_1\cdots x_{k-1}]},$$

giving the homotopic equivalence between $[x_1 \cdots x_{k-1} y_{[x_1\cdots x_k]}]$ and $\sum[x_1 \cdots \hat{x}_i \cdots x_{k-1} y_{[x_1\cdots x_{k-1}]} y_{[x_1\cdots x_k]}]$ in $\mathrm{Rips}(\mathscr{X}, r)$. By repeating this and concatnating the homotopy maps, we have the homotopy map $F_\sigma : \sigma \times [0,1] \to \mathrm{Rips}(\mathscr{X}, r)$ giving homotopy between $\iota_{\sigma\to\mathrm{Rips}(\mathscr{X},r)}$ and $f_\sigma : \sigma \to \mathrm{Rips}(\mathscr{X}, r)$ with $f_\sigma(\sigma) = \sum[x_{\tau(1)} y_{[x_{\tau(1)}x_{\tau(2)}]} \cdots y_{[x_1\cdots x_k]}]$, i.e. F_σ is giving homotopic equivalence between $[x_1, \ldots, x_k]$ and $\sum[x_{\tau(1)} y_{[x_{\tau(1)}x_{\tau(2)}]} \cdots y_{[x_1\cdots x_k]}]$ in $\mathrm{Rips}(\mathscr{X}, r)$. Now, note that for $\sigma, \tau \in \mathrm{Rips}(\mathscr{X}, r)$, F_σ and F_τ coincides on $\sigma \cap \tau \times [0,1]$, hence by induction, the homotopy map can be extended to F on the enitre Rips complex $\mathrm{Rips}(\mathscr{X}, r)$ giving homotopy between $id_{\mathrm{Rips}(\mathscr{X},r)}$ and f.

Now, note that $f(\sigma) = \sum[x_{\tau(1)} y_{[x_{\tau(1)}x_{\tau(2)}]} \cdots y_{[x_1\cdots x_k]}]$ consists of simplices with diameters at most $\sqrt{\frac{2\tau\left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right)}{\tau + \sqrt{\tau^2 - \left(r_\sigma^2 + \varepsilon_\sigma(2\tau - \varepsilon_\sigma)\right)}}} + 2\varepsilon$, hence applying Lemma 11 gives that

$$Rad([x_{\tau(1)} y_{[x_{\tau(1)}x_{\tau(2)}]} \cdots y_{[x_1\cdots x_k]}]) \leq \sqrt{\frac{d}{2(d+1)}}\left(\sqrt{\frac{2\tau\,(r_\sigma^2 + \varepsilon(2\tau - \varepsilon))}{\tau + \sqrt{\tau^2 - (r_\sigma^2 + \varepsilon(2\tau - \varepsilon))}}} + 2\varepsilon\right).$$

Hence by repeating this sufficiently many times (say N), we can guarantee that $f^{(N)}(\mathrm{Rips}(\mathscr{X}, r))$ consists of simplices with their radii (i.e. $Rad(\sigma)$) at most $\tilde{r}_\sigma$, where $\tilde{r}_\sigma$ is the solution of $f(t) =$

$\sqrt{\frac{d}{2(d+1)}}\left(\sqrt{\frac{2\tau\left(t^2+\varepsilon(2\tau-\varepsilon)\right)}{\tau+\sqrt{\tau^2-\left(t^2+\varepsilon(2\tau-\varepsilon)\right)}}}+2\varepsilon\right)-t=0$. Then $f\left(\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}\right)\geq 0$ implies that $\tilde{r}_\sigma \leq \sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}$. Hence ε satisfying $\varepsilon \leq \frac{1}{2}\left(\sqrt{\frac{2(d+1)}{d}}\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}-\sqrt{\frac{2\tau(r_{\min}^2+\varepsilon(2\tau-\varepsilon))}{2\tau+\sqrt{4\tau^2-2\left(r_{\min}^2+\varepsilon(2\tau-\varepsilon)\right)}}}\right)$ implies that $f^{(N)}(\mathrm{Rips}(\mathscr{X},r))\subset \check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}\left(\mathscr{X},\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}\right)$. And Lemma 12 implies that

$$f^{(N)}(\mathrm{Rips}(\mathscr{X},r))\subset \check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^d}\left(\mathscr{X},\sqrt{\frac{r_{\min}^2-\varepsilon(2\tau-\varepsilon)}{2}}\right)\subset \check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r_{\min})\subset \check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r),$$

i.e. $f^{(N)}:\mathrm{Rips}(\mathscr{X},r)\to \check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r)$. Then by construction, $\iota_{\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathscr{X},r)\to\mathrm{Rips}(\mathscr{X},r)}\circ f^{(N)}$ is homotopy equivalent to $id_{\mathrm{Rips}(\mathscr{X},r)}$. Hence by applying Lemma 24, $\mathbb{X}$ and $\mathrm{Rips}(\mathscr{X},r)$ are homotopy equivalent.

□

Proof of Corollary 16. Consider the double offset $\mathbb{Y}:=(((\mathbb{X}^{\frac{r_{\max}+\varepsilon}{\mu}})^{\complement})^{r_{\max}+\varepsilon})^{\complement}$. Since $\mathbb{X}$ has a positive μ-reach τ^{μ}, Corollary 10 implies that $\tau_{\mathbb{Y}}:=\mathrm{reach}(\mathbb{Y})\geq r_{\max}+\varepsilon$.

Then $r_x\leq \tau_{\mathbb{Y}}-\varepsilon$ holds, so Theorem 13 applies under the appropriate condition of δ.

Also, $r_x\leq\sqrt{\frac{d+1}{2d}}(\tau_{\mathbb{Y}}-\varepsilon)$ as well, so Theorem 14 applies under the appropriate condition of δ.

□

Proof of Lemma 17. To prove the covering lemma 17, we first show the following upper bound on the covering number holds.

Claim 26. *Suppose the distribution P satisfies the (a,b)-condition in* (16). *Then for all $\varepsilon<\varepsilon_0$, the covering number $\mathscr{N}(\mathbb{X},\|\cdot\|,2\varepsilon)$ is bounded as*

$$\mathscr{N}(\mathbb{X},\|\cdot\|,2\varepsilon)\leq\frac{1}{a}\varepsilon^{-b}.$$

Proof. Let $x_1,\dots,x_M$ be a maximal 2ε-packing of $\mathbb{X}$, with $M=\mathscr{M}(\mathbb{X},\|\cdot\|,2\varepsilon)$. Then $\mathscr{B}(x_i,\varepsilon)$ and $\mathscr{B}(x_j,\varepsilon)$ do not intersect for any i,j, and hence

$$\sum_{i=1}^{M}P(\mathscr{B}(x_i,\varepsilon))\leq P(\mathbb{R}^d)=1. \tag{27}$$

Then for all $\varepsilon<\varepsilon_0$, the (a,b)-condition implies

$$P(\mathscr{B}(x_i,\varepsilon))\geq a\varepsilon^b,$$

hence applying this to (27) gives that

$$\mathscr{M}(\mathbb{X},\|\cdot\|,2\varepsilon)\leq\frac{1}{a}\varepsilon^{-b}.$$

Then from the relationship between covering number and packing number,

$$\mathcal{N}(\mathbb{X}, \|\cdot\|, 2\varepsilon) \leq \mathcal{M}(\mathbb{X}, \|\cdot\|, 2\varepsilon) \leq \frac{1}{a}\varepsilon^{-b}.$$

$\square$

Now, to prove Lemma 17, set $\varepsilon := \frac{1}{4}\min_i r_{n,i}$. Under the (a,b)-condition, the previous claim 26 implies that there exists $x_1, \ldots, x_N$ with $N \leq a^{-1}\varepsilon^{-b}$ satisfying

$$\mathbb{X} \subset \bigcup_{j=1}^{N} \mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon).$$

Let E be the event that all $\mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon)$ have intersections with $\{X_1, \ldots, X_n\}$, that is, for each $1 \leq j \leq N$, there exists $1 \leq i \leq n$ with $X_i \in \mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon)$. Then note that $4\varepsilon = \min_i r_{n,i} \leq r_{n,i}$, and hence we have the following relations between balls:

$$\mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon) \subset \mathbb{B}_{\mathbb{R}^d}(X_i, 4\varepsilon) \subset \mathbb{B}_{\mathbb{R}^d}(X_i, r_{n,i}).$$

Therefore, under the event E, we have

$$\mathbb{X} \subset \bigcup_{j=1}^{N} \mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon) \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^d}(X_i, r_{n,i}),$$

which implies

$$\mathbb{P}\left(\mathbb{X} \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^d}(X_i, r_{n,i})\right) \geq \mathbb{P}(E). \tag{28}$$

Now, $P(E)$ can be lower bounded as

$$\begin{aligned}
\mathbb{P}(E) &= \mathbb{P}\left(\bigcap_{j=1}^{N}\bigcup_{j=1}^{n}\{X_i \in \mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon)\}\right) \\
&= 1 - \mathbb{P}\left(\bigcup_{j=1}^{N}\bigcap_{i=1}^{n}\{X_i \notin \mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon)\}\right) \\
&\geq 1 - \sum_{j=1}^{N}\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \notin \mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon)\}\right) \\
&= 1 - \sum_{j=1}^{N}\prod_{i=1}^{n}(1 - P(\mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon)) \\
&\geq 1 - \sum_{j=1}^{N}\exp\left(-\sum_{i=1}^{n}P(\mathbb{B}_{\mathbb{R}^d}(x_j, 2\varepsilon))\right),
\end{aligned}$$

where the last line is from that $1-t \leq \exp(-t)$ for all $t \in \mathbb{R}$. Now, from the covering number bound $N \leq a^{-1}\varepsilon^{-b}$ with the condition, $2\varepsilon = \frac{1}{2}\min_i r_{n,i} \geq \left(\frac{\log n}{an}\right)^{1/b}$, we can further lower bound $P(E)$ as following:

$$\begin{aligned} P(E) &\geq 1 - N\exp\left(-an(2\varepsilon)^b\right) \\ &\geq 1 - a^{-1}\varepsilon^{-b}\exp\left(-an(2\varepsilon)^b\right) \\ & 1 - \frac{n}{2^b \log n}\exp\left(-\log n\right) \\ &= 1 - \frac{1}{2^b \log n}, \end{aligned}$$

which implies

$$\mathbb{P}\left(\mathbb{X} \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^d}(X_i, r_{n,i})\right) \geq 1 - \frac{1}{2^b \log n},$$

as desired. □