Statistical Inference for Geometric Data

Jisu KIM

September 19, 2018

Carnegie Mellon University

Thesis Committee:

Larry Wasserman (Co-Chair)
Alessandro Rinaldo (Co-Chair)
Sivaraman Balakrishnan
Frédéric Chazal (INRIA Saclay)
Jessi Cisewski (Yale University)

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Abstract

Geometric structures can aid statistics in several ways. In high dimensional statistics, geometric structures can be used to reduce dimensionality. High dimensional data entails the curse of dimensionality, which can be avoided by if there are low dimensional geometric structures. On the other hand, geometric structures also provide useful information. Structures may carry scientific meaning about the data and can be used as features to enhance supervised or unsupervised learning.

In this defense, I will explore how statistical inference can be done on geometric structures. First, I will explore the minimax rates of dimension estimator and reach estimator. Second, I will investigate inference on cluster trees and persistent homology of density filtration on rips complex. Third, I will extend and improve R package TDA for computing topological data analysis.

Contents

Chapter 1

Introduction

In high dimensional statistics, geometrical structures can be used to reduce dimensionality. High dimensional data suffers from the "curse of dimensionality"[Bellman, 1961, Lee and Verleysen, 2007a, Hastie et al., 2009], which refers to the fact that the number of data samples for an inference with the desired accuracy grows exponentially with dimensions. The curse of dimensionality is mitigated if the data are to form geometrical structures. The assumed geometrical structures can both lower the dimensionality of the data and approximate complicated structure of the data.

On the other hand, geometrical structures of the data also provide information on data. First, geometrical structures carry scientific meaning about data in many scientific applications. For example, geometrical structures of galaxies, gas, and dark matter in the universe give clues on the initial state of the universe before the big bang. Also, geometrical structures of an enzyme determine its function. Second, the geometrical structures are used to enhance supervised or unsupervised learning. For this case, the interpretation of geometrical structures is unclear, but geometrical structures are extracted from data for higher performance in learning.

Lastly, geometry is also used in data visualization to provide insights on data through visual intuition. Some geometrical structures in data visualization such as size, orientation, shape are basic visual attributes that are perceived without conscious effort. Hence those geometrical structures are perceived in parallel and hence fast[Few, 2004]. Nonquantitative information can be also conveyed by geometric structures[Few, 2013]. For example, a graph in 2d representing network data gives an immediate interpretation about which nodes are clustered or which nodes are influential.

In this thesis, I will explore how statistical inference can be done on geometrical structures. First, I will explore the minimax rates of dimension estimator (Chapter 2) and reach estimator (Chapter 3). Second, I will investigate inference on cluster trees (Chapter 4) and persistent homology of density filtration on rips complex (Chapter 5). Third, I will extend and improve R package TDA for computing topological data analysis (Chapter 6).

1.1 Minimax

The minimax rate is the risk of an estimator that performs best in the worst case, as a function of the sample size [see, e.g. Tsybakov, 2008]. Let $\mathcal{P}$ be a collection of probability distributions over the same sample space $\mathbb{X}$ and let $\theta : \mathcal{P} \to \Theta$ be a function over $\mathcal{P}$ taking values in some space Θ, the parameter space. We can think of $\theta(P)$ as the feature of interest of the probability distribution P, such as its mean. For the fixed sample size n, suppose $X = (X_1, \ldots, X_n)$ is an i.i.d. (independent and identically distributed) sample drawn from a fixed probability distribution $P \in \mathcal{P}$. Thus X takes values in the

n-fold product space $\mathbb{X}^n = \mathbb{X} \times \cdots \times \mathbb{X}$ and is distributed as $P^{(n)}$, the n-fold product measure. An estimator $\hat{\theta}_n : \mathbb{R}^n \to \Theta$ is any measurable function that maps the observation X into the parameter space Θ. Let $\ell : \Theta \times \Theta \to \mathbb{R}$ be a loss function, a non-negative bounded function that measures how different two parameters are. Then for a fixed estimator $\hat{\theta}_n$ and a fixed distribution P, the risk of $\hat{\theta}_n$ is defined as

$$\mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{\theta}_n(X), \theta(P)\right)\right].$$

Then for a fixed estimator $\hat{\theta}_n$, its maximum risk is the supremum of its risk over every distribution $P \in \mathcal{P}$, that is,

$$\sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{\theta}_n(X), \theta(P)\right)\right]. \tag{1.1}$$

The minimax risk associated to $\mathcal{P}$, θ, ℓ and n is the maximal risk of any estimator that performs the best under the worst possible choice of P. Formally, the *minimax risk* is

$$R_n = \inf_{\hat{\theta}_n} \sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{\theta}_n(X), \theta(P)\right)\right]. \tag{1.2}$$

The minimax risk R_n in (1.2) is often viewed as a function of the sample size n, in which case any positive sequence ψ_n such that $\lim_{n\to\infty} R_n/\psi_n$ remains bounded away from 0 and ∞ is called a *minimax rate.* Notice that minimax rates are unique up to constants and lower order terms.

To define a meaningful minimax risk, it is essential to have some constraint on the set of distributions $\mathcal{P}$ in (1.1) and (1.2). If $\mathcal{P}$ is too large, then the minimax rate R_n in (1.2) will not converge to 0 as n goes to ∞: this means that the problem is statistically ill-posed. If $\mathcal{P}$ is too small, the minimax estimator depends too much on the specific distributions in $\mathcal{P}$ and is not a useful measure of a statistical difficulty.

Determining the value of the minimax risk R_n in (1.2) for a given problem requires two separate calculations: an upper bound on R_n and a lower bound. In order to derive an upper bound, one analyzes the asymptotic risk of a specific estimator $\hat{\theta}_n$. This will in turn yield an upper bound on the minimax risk R_n, since

$$R_n = \inf_{\hat{\theta}_n} \sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{\theta}_n(X), \theta(P)\right)\right] \leq \sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{\theta}_n(X), \theta(P)\right)\right]. \tag{1.3}$$

Naturally, choosing an appropriate estimator is critical to get a sharp bound.

Lower bounds are instead usually computed by measuring the difficulty of a multiple hypothesis testing problem that entails identifying finitely many distributions in $\mathcal{P}$ that are maximally difficult to discriminate [see, e.g. Tsybakov, 2008, Section 2.2].

One method for to compute the lower bound from those distributions is Le Cam's lemma [Yu, 1997, Chapter 29.2, Lemma 1].

Lemma 1. (Le Cam's Lemma) *Let $\mathcal{P}$ be a set of probability measures on $(\Omega, \mathcal{F})$, and $\mathcal{P}_1, \mathcal{P}_2 \subset \mathcal{P}$ be such that for all $P \in \mathcal{P}_i$, $\theta(P) = \theta_i$ for $i = 1, 2$. For any $Q_i \in co(\mathcal{P}_i)$, where $co(\mathcal{P}_i)$ is the convex hull of $\mathcal{P}_i$, let q_i be the density of Q_i with respect to a measure ν. Then*

$$\inf_{\hat{\theta}} \sup_{P\in\mathcal{P}} \mathbb{E}_P[\ell(\hat{\theta}, \theta(P))] \geq \frac{\Delta}{2} \int [q_1(x) \wedge q_2(x)] d\nu(x), \tag{1.4}$$

where $\Delta = \ell(\theta_1, \theta_2)$.

In above Le Cam's lemma, considering the convex hull of distributions $co(\mathcal{P}_i)$ is sometimes critical for getting the nontrivial lower bound. Sometimes P_1 from $\mathcal{P}_1$ and P_2 from $\mathcal{P}_2$ are always mutually singular, resulting in 0 as a lower bound in (1.4). However, Q_1 from $co(\mathcal{P}_1)$ and Q_2 from $co(\mathcal{P}_2)$ can be mutually nonsingular, resulting nontrivial lower bound in (1.4). This technique is used in Section 2.3.

1.2 Differential Geometry

We briefly review some notation from differential geometry. A topological manifold of dimension d is a topological space M and a family of homeomorphisms $\varphi_\alpha : U_\alpha \subset \mathbb{R}^d \to V_\alpha \subset M$ from an open subset of $\mathbb{R}^d$ to an open subset of M such that $\bigcup_\alpha \varphi_\alpha(U_\alpha) = M$. Such d is unique and is called the dimension of a manifold. If, for any pair α, β, with $\varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta) \neq \emptyset$, $\varphi_\beta^{-1} \circ \varphi_\alpha : U_\alpha \cap U_\beta \to U_\alpha \cap U_\beta$ is C^k, then M is a C^k-manifold.

We assume that the topological manifold M is embedded in $\mathbb{R}^m$, i.e. $M \subset \mathbb{R}^m$, and the metric is inherited from the metric of $\mathbb{R}^m$. For a topological manifold $M \subset \mathbb{R}^m$ and for any $q, r \in M$, a path joining q_1 to q_2 is a map $\gamma : [a, b] \to M$ for some $a, b \in \mathbb{R}$ such that $\gamma(a) = q_1$, $\gamma(b) = q_2$. The length of the curve γ is defined as $\text{Length}(\gamma) = \int_a^b ||\gamma'(t)||_2 dt$. A topological manifold M is equipped with the distance $dist_M : M \times M \to \mathbb{R}$ as $dist_M(q_1, q_2) = \inf_{\gamma\text{: path joining } q_1 \text{ and } q_2} \text{Length}(\gamma)$. A path $\gamma : [a, b] \to M$ is a geodesic if for all $t, t' \in [a, b]$, $dist_M(\gamma(t), \gamma(t')) = |t - t'|$.

Let T_qM denote the tangent space to M at q. Given $q \in M$, there exist a set $0 \in \mathcal{E} \subset T_q(M)$ and a mapping $\exp_q : \mathcal{E} \subset T_qM \to M$ such that $t \to \exp_q(tv)$, $t \in (-1, 1)$, is the unique geodesic of M which, at $t = 0$, passes through q with velocity v, for all $v \in \mathcal{E}$. The map $\exp_q : \mathcal{E} \subset T_qM \to M$ is called the exponential map on q.

1.3 Reach

First introduced by Federer [Federer, 1959], the reach is a regularity parameter defined as follows. Given a closed subset $A \subset \mathbb{R}^m$, the medial axis of A, denoted by $\text{Med}(A)$, is the subset of $\mathbb{R}^m$ composed of the points that have at least two nearest neighbors on A. Namely, denoting by $d(x, A) = \inf_{q \in A} ||q - x||$ the distance function to A,

$$\text{Med}(A) = \{x \in \mathbb{R}^m | \exists q_1 \neq q_2 \in A, ||q_1 - x|| = ||q_2 - x|| = d(x, A)\}. \tag{1.5}$$

The reach of A is then defined as the minimal distance from A to $Med(A)$.

Definition 2. The reach of a closed subset $A \subset \mathbb{R}^m$ is defined as

$$\tau_A = \inf_{q \in A} d(q, \text{Med}(A)) = \inf_{q \in A, x \in \text{Med}(A)} ||q - x||. \tag{1.6}$$

Some authors refer to τ_A^{-1} as the *condition number* Niyogi et al. [2008], Singer and Wu [2012]. From the definition of the medial axis in (1.5), the projection $\pi_A(x) = \arg\min_{p \in A} \|p - x\|$ onto A is well defined outside $Med(A)$. The reach is the largest distance $\rho \geq 0$ such that π_A is well defined on the ρ-offset $\{x \in \mathbb{R}^m | d(x, A) < \rho\}$. Hence, the reach condition can be seen as a generalization of convexity, since a set $A \subset \mathbb{R}^m$ 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.

Theorem 3. *[Federer, 1959, Theorem 4.18] For all submanifold $M \subset \mathbb{R}^m$,*

$$\tau_M = \inf_{q_1 \neq q_2 \in M} \frac{\|q_1 - q_2\|_2^2}{2d(q_2 - q_1, T_{q_1}M)}. \tag{1.7}$$

This formulation has the advantage of involving only points on M and its tangent spaces, while (1.6) uses the distance to the medial axis $Med(M)$, which is a global quantity. The formula (1.7) will be the starting point of the estimator proposed in Chapter 3 (see Section 3.2.1).

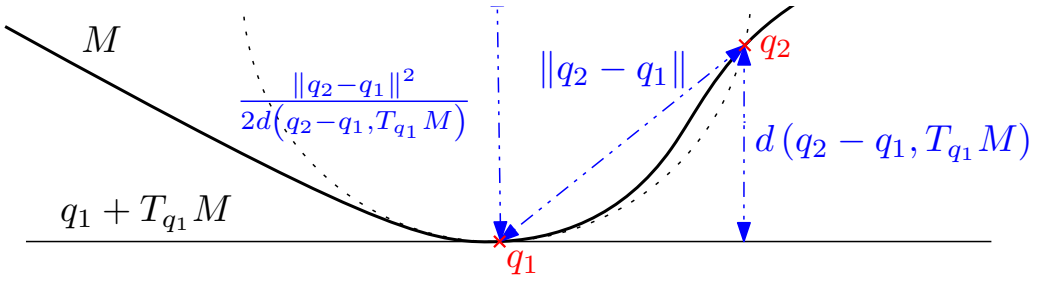


Figure 1.1: Geometric interpretation of quantities involved in (1.7).

The ratio appearing in (1.7) can be interpreted geometrically, as suggested in Figure 1.1. This ratio is the radius of an ambient ball, tangent to M at q_1 and passing through q_2. Hence, at a differential level, the reach gives a lower bound on the radii of curvature of M. Equivalently, τ_M^{-1} is an upper bound on the curvature of M.

Proposition 4 (Proposition 6.1 in Niyogi et al. [2008])**.** *Let $M \subset \mathbb{R}^m$ be a submanifold, and $\gamma_{p,v}$ an arc-length parametrized geodesic of M. Then for all t,*

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

In analogy with function spaces, the class $\{M \subset \mathbb{R}^m | \tau_M \geq \tau_{min} > 0\}$ can be interpreted as the Hölder space $\mathcal{C}^2(1/\tau_{min})$. In addition, as illustrated in Figure 1.2, the condition $\tau_M \geq \tau_{min} > 0$ also prevents bottleneck structures where M is nearly self-intersecting. This idea will be made rigorous in Section 3.2.

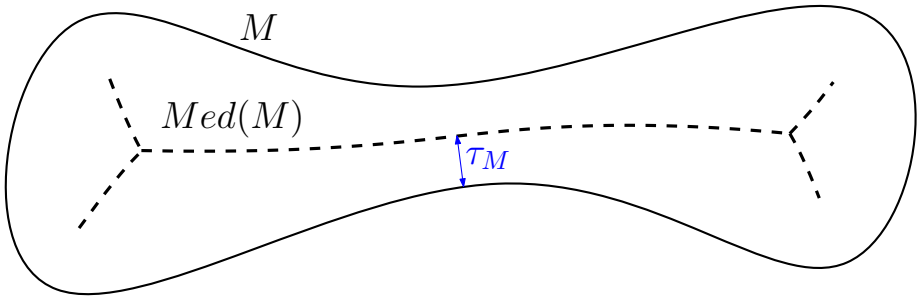


Figure 1.2: A narrow bottleneck structure yields a small reach τ_M.

1.4 Algebraic Topology

1.4.1 Simplicial complex

A simplicial complex can be seen as a high dimensional generalization of a graph. Given a set V, an *(abstract) simplicial complex* is a set K of finite subsets of V such that $\alpha \in K$ and $\beta \subset \alpha$ implies $\beta \in K$. Each set $\alpha \in K$ is called its *simpex*. The *dimension* of a simplex α is $\dim \alpha = \mathrm{card}\alpha - 1$, and the dimension of the simplicial complex is the maximum dimension of any of its simplices. Note that a simplicial complex of dimension 1 is a graph.

When we are to infer topological information of a metric space $(\mathbb{X}, d)$ from a finite sample points $\mathcal{X} = \{X_1, \ldots, X_n\} \subset \mathbb{X}$, we use several simplicial complexes built on the sample points $\mathcal{X}$. For $x \in \mathbb{X}$

and $r > 0$, let $\mathbb{B}_{\mathbb{X}}(x, r)$ be the ball centered at x and radius $r > 0$, i.e. $\mathbb{B}_{\mathbb{X}}(x, r) = \{y \in \mathbb{X} : d(x, y) < r\}$.

For a set of positive numbers $r \in (0, \infty)^n$, the simplicial complex on $\mathcal{X}$ consisting of all simplices $[X_{i_1}, \ldots, X_{i_k}]$ such that the intersection $\cap_{j=1}^{k} \mathbb{B}_{\mathbb{X}}(X_{i_j}, r_{i_j})$ is non-empty is known as the (weighted) *Čech complex.*

Definition 5 (Čech complex)**.** Let $\mathcal{X} = \{X_1, \ldots, X_n\} \subset \mathbb{X}$ and $r \in (0, \infty)^n$. The (weighted) Čech complex is the simplicial complex

$$\check{\text{C}}\text{ech}_{\mathbb{X}}(\mathcal{X}, r) := \left\{ \sigma = [X_{i_1}, \ldots, X_{i_k}] \subset \mathcal{X} : \bigcap_{j=1}^{k} \mathbb{B}_{\mathbb{X}}(X_{i_j}, r_{i_j}) \neq \emptyset \right\}, \tag{1.8}$$

We will drop the subscript $\mathbb{X}$ when it is clear from the context.

The topology of the Čech complex is linked to underlying continuous spaces via Nerve Theorem. Let $r = (r_1, \ldots, r_n) \in (0, \infty)^n$ and consider the union of balls

$$\bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{1.9}$$

Then the union of balls in (1.9) is homotopic equivalent to the Čech complex by the following Nerve Theorem.

Lemma 6 (Nerve Theorem)**.** *Let $\mathcal{X}_n \subset \mathbb{X}$ and $r = (r_1, \ldots, r_n) \in (0, \infty)^n$. If, for each $k = 1, \ldots, n$ and $i_1 < i_2, \ldots, < i_k$, the intersection $\bigcap_{j=1}^{k} \mathbb{B}_{\mathbb{X}}(X_{i_j}, r_{i_j})$ is either empty or contractibe, then the Čech complex $\check{\text{C}}\text{ech}_{\mathbb{X}}(\mathcal{X}_n, r)$ is homotopy equivalent to the union of balls $\bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{X}}(X_i, r_i)$.*

Computing the Čech complex requires computing all the intersections of the balls. To save on computation time, we may instead add a simplex whenever pairwise distances of its vertices are close. This leads to the *Vietoris-Rips complex*, also known as the Rips complex.

Definition 7 (Vietoris-Rips complex)**.** The (weighted) Vietoris-Rips complex $R(\mathcal{X}_n, r)$ is defined by

$$R(\mathcal{X}_n, r) := \left\{ \sigma = [X_{i_1}, \ldots, X_{i_k}] : d\left(X_{i_j}, X_{i_l}\right) < r_{i_j} + r_{i_l}, \forall j \neq l, k = 1, \ldots, n \right\}. \tag{1.10}$$

Note that the Čech complex and Rips complex have following interleaving inclusion relationship

$$\check{\text{C}}\text{ech}(\mathcal{X}_n, r) \subset R(\mathcal{X}_n, r) \subset \check{\text{C}}\text{ech}(\mathcal{X}_n, 2r). \tag{1.11}$$

In particular, when r_i's are all the same and $\mathbb{X}$ is a Euclidean space, then the constant 2 can be tightened to $\sqrt{2}$:

$$\check{\text{C}}\text{ech}(\mathcal{X}_n, r) \subset R(\mathcal{X}_n, r) \subset \check{\text{C}}\text{ech}(\mathcal{X}_n, \sqrt{2}r). \tag{1.12}$$

Hence both Čech complex (1.8) and Rips complex (1.10) are both topologically approximating the union of balls (1.9) via Nerve Theorem (Lemma 6) and interleaving relation between Čech complex and Rips complex ((1.11) or (1.12)).

1.4.2 Persistent Homology

Persistent homology is a multiscale approach to analyze topological features in data.

Suppose $X \subset \mathbb{X}$ be an observed data points. A filtration $\mathcal{F}$ is a collection of subspaces in $\mathbb{X}$ that approximates the data points in different resolutions. Define a partial order on $\mathbb{R}^D$ by taking $(a_1, \cdots, a_D) \preceq (b_1, \cdots, b_D)$ if and only if $a_i \leq b_i$ for all i.

Definition 8. A (D-dimensional) *filtration* $\mathcal{F} = \left\{\mathcal{F}_a \subset \mathbb{X} : a \in \mathbb{R}^D\right\}$ is a collection of subspaces in $\mathbb{X}$ satisfying that $a \preceq b$ implies $\mathcal{F}_a \subset \mathcal{F}_b$.

For a filtration $\mathcal{F}$ and for each $k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, associated persistent homology $H_k\mathcal{F}$ is a collection of k-th dimensional homology of each subset in $\mathcal{F}$.

Definition 9. Let $\mathcal{F}$ be a D-dimensional filtration and let $k \in \mathbb{N}_0$. Associated (D-dimensional) k-th *persistent homology* $PH_k\mathcal{F}$ is a collection of vector spaces $\{H_k\mathcal{F}_a\}_{a\in\mathbb{R}^D}$ equipped with homomorphisms $\left\{\imath_k^{a,b}\right\}_{a\preceq b}$, where $H_k\mathcal{F}_a$ is a k-th dimensional homology of $\mathcal{F}_a$ and $\imath_k^{a,b}$ is the homomorphism induced from the inclusion $\mathcal{F}_a \subset \mathcal{F}_b$.

For 1-dimensional persistent homology, its structure is completely represented as its decomposition. For k-th persistent homology $PH_k\mathcal{F}$, the set of filtration values that a specific homology appears is always an interval $[b, d) \subset [-\infty, \infty]$, i.e. a specific homology is formed at some filtration value $b \in [-\infty, \infty]$ and dies when the inside hole is filled at some filtration value $d \in [-\infty, \infty]$.

Definition 10. Let $\mathcal{F}$ be a 1-dimensional filtration and let $k \in \mathbb{N}_0$. Associated k-th *persistent diagram* $Dgm_k(\mathcal{F})$ is a finite multi-set of $(\mathbb{R} \cup \{\infty\})^2$, consisting of all pairs (b, d) where $[b, d)$ is the set of filtration values that a specific homology appears in $PH_k\mathcal{F}$. b is called a birth time and d is called a death time.

1.4.3 Stability and Statistical Inference of Persistent Homology

Stability theorems and statistical inference have been developed for 1-dimensional filtrations, in particular when the filtration $\mathcal{F}$ is generated from sub-level sets or super-level sets of a function. Let $f : \mathbb{X} \subset \mathbb{R}^m \to \mathbb{R}$ be a function that approximates the data points in different resolutions. The associated filtration $\mathcal{F}$ can be constructed from sub-level sets $\mathcal{F}_a = \{x \in \mathbb{R}^m : f(x) \le a\}$ or super-level sets $\mathcal{F}_a = \{x \in \mathbb{X} : f(x) \ge a\}$. Common choices for the filtration function f are as follows: (1) sub-level sets of distance function $f(x) = d(x, X) = \inf_{y\in X} d(x, y)$, (2) super-level set of density function $f(x) = \hat{p}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^m} K\left(\frac{||x-X_i||}{h}\right)$, with any kernel K and a positive number h. Super-level sets of function f corresponds to sub-level sets of function $-f$, hence the same theory can be used. For each $k \in \mathbb{N}_0$, let $Dgm_k(f)$ be k-th persistent diagram from either sub-level sets or super-level sets of f.

Let $f, g : \mathbb{X} \subset \mathbb{R}^m \to \mathbb{R}$ be two functions, and let $PH_*(f)$ and $PH_*(g)$ be the corresponding persistent homologies of the upper level set filtrations $\{f \le L\}_{L\in\mathbb{R}}$ and $\{g \le L\}_{L\in\mathbb{R}}$.

To impose stability, we first endow the space of persistence diagrams with a metric. The most fundamental one is the *bottleneck distance*.

Definition 11. The bottleneck distance between the persistent homology of the filtrations $PH_*(f)$ and $PH_*(g)$ is defined by

$$d_B(PH_k(f), PH_k(g)) = \inf_{\gamma\in\Gamma} \sup_{x\in Dgm_k(f)} \|p - \gamma(p)\|_\infty,$$

where the set Γ consists of all the bijections $\gamma : Dgm_k(f) \cup Diag \to Dgm_k(g) \cup Diag$, and $Diag$ is the diagonal line $\{(x, x) : x \in \mathbb{R}\} \subset \mathbb{R}^2$ with infinite multiplicity.

We will impose a standard regularity condition for the functions f and g, which is *tameness*.

Definition 12. (Chazal et al. [2009], Bobrowski et al. [2014]) Let $f : \mathbb{X} \to \mathbb{R}$. Then f is *tame* if $H_k(f^{-1}(-\infty, L])$ is of finite rank for all $k \in \mathbb{N} \cup \{0\}$ and $L \in \mathbb{R}$.

For two tame functions f and g, their bottleneck distance is bounded by their ℓ_∞ distance, an important and useful fact known as the stability theorem.

Theorem 13 (Stability Theorem)**.** *(Cohen-Steiner et al. [2007], Chazal et al. [2009]) For two tame functions $f, g : \mathbb{X} \to \mathbb{R}$,*

$$d_B(PH_k(f), PH_k(g)) \leq \|f - g\|_\infty.$$

Statistical inference have been developed for persistent homology in [Fasy et al., 2014b]. When points of birth and death are close to the diagonal in the persistence diagram, corresponding homologies are not significant, since corresponding holes will be soon filled out right after when they are born. With detailed statistical analysis, a $1-\alpha$ confidence band c_n for persistent homology can be calculated. Precisely, c_n satisfies

$$\liminf_{n\to\infty} \mathbb{P}\left(W_\infty(\widehat{Dgm_k(f)}, Dgm_k(f)) \in [0, c_n]\right) \geq 1 - \alpha,$$

where $Dgm_k(f)$ is persistence diagram for the true distribution of data, $\widehat{Dgm_k(f)}$ is persistence diagram computed on data, and $W_\infty(X, Y)$ is the bottleneck distance between two diagrams X and Y defined as $W_\infty(X, Y) = \inf_{\eta: X\to Y} \sup_{x\in X} \|x - \eta(x)\|_\infty$. Those holes above the confidence band are simultaneously statistically significant.

Sublevel sets of the distance to measure (DTM) [Caillerie et al., 2011] is considered to approximate holes in the data points in different resolutions. The DTM is a robustified version of the distance function. More precisely, the DTM d_{μ,m_0} for a probability distribution μ with parameter $m_0 \in [0, 1]$ is defined by

$$d_{\mu,m_0} : \mathbb{R}^m \to \mathbb{R}^+,\ x \mapsto \sqrt{\frac{1}{m_0}\int_0^{m_0} (\delta_{\mu,m}(x))^2 dm},$$

where $\delta_{\mu,m}(x) = \inf\{r > 0 :\ \mu(\mathbb{B}(x, r)) > m\}$. When μ is an empirical measure $P_n(x) = \frac{1}{n}\sum_{i=1}^n I_{X_i}(x)$, the empirical DTM is

$$\hat{d}_{\mu,m_0(x)} = d_{P_n,m_0}(x) = \sqrt{\frac{1}{m_0 n}\sum_{i\leq \lfloor m_0 n\rfloor} \|X_{(i)} - x\|_2^2 + \left(1 - \frac{\lfloor m_0 n\rfloor}{m_0 n}\right)\|X_{(\lceil m_0 n\rceil)} - x\|_2^2}, \tag{1.13}$$

where for each x, $X_{(1)}, \cdots, X_{(n)}$ is ordered so that $\|X_{(1)} - x\|_2 \leq \cdots \leq \|X_{(n)} - x\|_2$. Hence the empirical DTM behaves similarly to the k-nearest distance with $k = \lfloor m_0 n\rfloor$. The DTM is preferred choice for the filtration function, since the persistence diagram computed on the DTM is robust to noise.

Chapter 2

Minimax Rates for Estimating the Dimension of a Manifold

This chapter presents the work in [Kim et al., 2016].

Suppose that $X_1, \ldots, X_n$ is an i.i.d. sample from a distribution P whose support is an unknown, well behaved, manifold M of dimension d in $\mathbb{R}^m$, where $1 \leq d \leq m$. Manifold learning refers broadly to a suite of techniques from statistics and machine learning aimed at estimating M or some of its features based on the data.

Manifold learning procedures are widely used in high dimensional data analysis, mainly to alleviate the curse of dimensionality. Such algorithms map the data to a new, lower dimensional coordinate system [Bellman, 1961, Lee and Verleysen, 2007a, Hastie et al., 2009], with little loss in accuracy. Manifold learning can greatly reduce the dimensionality of the data.

Most manifold learning techniques require, as input, the intrinsic dimension of the manifold. However, this quantity is almost never known in advance and therefore has to be estimated from the data.

Various intrinsic dimension estimators have been proposed and analyzed; [see, e.g., Lee and Verleysen, 2007b, Koltchinskii, 2000, Kégl, 2003, Levina et al., 2004, Hein and Audibert, 2005, Raginsky and Lazebnik, 2005, Little et al., 2009, 2011, Sricharan et al., 2010, Rozza et al., 2012, Camastra and Staiano, 2016] However, characterizing the intrinsic statistical hardness of estimating the dimension remains an open problem.

The traditional way of measuring the difficulty of a statistical problem is to bound its *minimax risk,* which in the present setting is loosely described as the worst possible statistical performance of an optimal dimension estimator. Formally, given a class of probability distribution $\mathcal{P}$, the minimax risk $R_n = R_n(\mathcal{P})$ is defined as

$$R_n = \inf_{\hat{d}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[1(\hat{d} \neq d(P)) \right]. \tag{2.1}$$

In (2.1), $d(P)$ is the dimension of the support of P, $\mathbb{E}_P$ denotes the expectation with respect to the distribution P, $1(\cdot)$ is the indicator function, and the infimum is over all estimators (measurable functions of the data) $\hat{d} = \hat{d}(X_1, \ldots, X_n)$ of the dimension $d(P)$. The risk $\mathbb{E}_P[1(\hat{d} \neq d(P))]$ of a dimension estimator $\hat{d}$ is the probability that $\hat{d}$ differs from the true dimension $d(P)$ of the support of the data generating distribution P. The minimax risk $R_n(\mathcal{P})$, which is a function of both the sample size n and the class $\mathcal{P}$, quantifies the intrinsic hardness of the dimension estimation problem, in the sense that *any dimension estimator* cannot have a risk smaller than R_n uniformly over every $P \in \mathcal{P}$.

The purpose of this chapter is to obtain upper and lower bounds on the minimax risk R_n in (2.1). We impose several regularity conditions on the set of manifolds supporting the distribution in the class

$\mathcal{P}$, in order to make the problem analytically tractable and also to avoid pathological cases, such as space-filling manifolds. We first assume that the manifold supporting the data generating distribution P has two possible dimensions, d_1 and d_2. This assumption is then relaxed to any dimension $d(P)$ between 1 and the embedding dimension m. Our main result is the following theorem. See Section 2.1 for the definition of the class $\mathcal{P}$ of probability distributions supported on well-behaved manifolds in $\mathbb{R}^m$.

Theorem 14. *The minimax risk R_n in* (2.1) *satisfies, $a_n \leq R_n \leq b_n$, where*

$$a_n = (C^{(29)}_{K_I})^n \min\{\tau_\ell^{-4} n^{-2}, 1\}^n, \tag{2.2}$$

$$b_n = (C^{(28)}_{K_I,K_p,K_v,m})^n (1 + \tau_g^{-(m^2-m)n}) n^{-\frac{n}{m-1}}, \tag{2.3}$$

and the constants τ_ℓ, τ_g, $C^{(29)}_{K_I}$ and $C^{(28)}_{K_I,K_p,K_v,m}$ depend on $\mathcal{P}$ and are defined in Section 2.4.

This chapter is organized as follows. In Section 2.1, we formulate and discuss regularity conditions on distributions and their supporting manifolds. In Section 2.2, we provide an upper bound on the minimax rate by considering the traveling salesman path through the points. In Section 2.3, we derive a lower bound on the minimax rate by applying Le Cam's lemma with a specific set of d_1-dimensional and d_2-dimensional probability distributions. In Section 2.4, we extend our upper bound and lower bound for the case where the intrinsic dimension varies from 1 to m. For the readability, all the proofs are postponed to Appendix A.

2.1 Regularity conditions

In this section, we define the set $\mathcal{P}$ of probability distributions that we consider in bounding the minimax risk R_n in (2.1). Such distributions are supported on manifolds whose dimension d is between 1 and m, where m is the dimension of the embedding space. In particular, we require that the supporting manifolds have a uniform lower bound on their reach parameters τ_g and τ_l. The resulting class of distributions is denoted by

$$\mathcal{P} = \bigcup_{d=1}^{m} \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}. \tag{2.4}$$

In the rest of this subsection, we will make the definition $\mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}$ precise. Readers who are not interested in the details may skip the rest of the section. All the proofs for this section are in Section A.1.

In our analysis we require various regularity conditions on the class $\mathcal{P}$ of probability distributions appearing in the minimax risk (2.1). Most of these conditions are of a geometric nature and concern the properties of the manifolds supporting the probability distributions in $\mathcal{P}$. Altogether, our assumptions rule out manifolds that are so complicated to make the dimension estimation problem unsolvable and, therefore, guarantee that the minimax risk R_n in (2.1) converges to 0 as n goes to ∞. Such regularity assumptions are quite mild, and in fact allow for virtually all types of manifolds usually encountered in manifold learning problems.

Our first assumption is that the probability distributions in $\mathcal{P}$ are supported over manifold contained inside a compact set, which, without loss of generality, we take to be the cube $I := [-K_I, K_I]^m$, for some $K_I > 0$. See Figure 2.1.

Second, to exclude manifolds that are arbitrarily complicated in the sense of having unbounded curvatures or of being nearly self intersecting, we assume that the reach is uniformly bounded from below. More precisely, we will constrain both the global reach and the local reach as follows. Fix

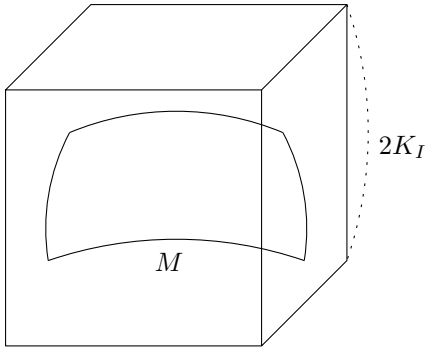


Figure 2.1: A manifold M is assumed to be contained inside the cube $I = [-K_I, K_I]^m$, for some $K_I > 0$. See Definition 15.

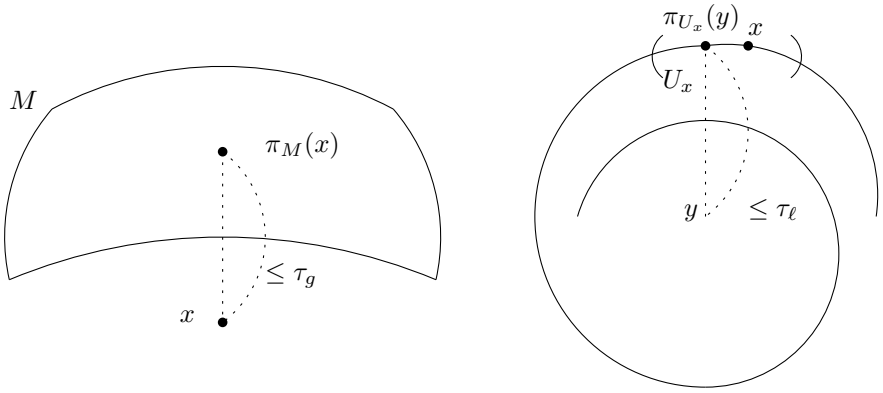


Figure 2.2: A manifold M with *global reach* at least τ_g, or *local reach* at least τ_ℓ. See Definition 15.

$\tau_g, \tau_\ell \in (0, \infty]$ with $\tau_g \leq \tau_\ell$. The global reach condition for a manifold M is that the usual reach $\tau(M)$ in (1.6) is lower bounded by τ_g as in Figure 2.2, and the local reach condition is that M can be covered by small patches whose reaches are lower bounded by τ_ℓ, as in Figure 2.2. (See Definition 15 below for more details.)

Third, we assume that the data are generated from a distribution P supported on a manifold M having a density with respect to the (restriction of the) Hausdorff measure on M bounded from above by some positive constant K_p.

For manifolds without boundary, the above conditions suffice for our analysis. However, to deal with manifolds with boundary, we need further assumptions, namely local geodesic completeness and essential dimension. A manifold M is said to be complete if any geodesic can be extended arbitrarily farther, i.e. for any geodesic path $\gamma : [a, b] \to M$, there exists a geodesic $\tilde{\gamma} : \mathbb{R} \to M$ that satisfies $\tilde{\gamma}|_{[a,b]} = \gamma$. [see, e.g., Lee, 2000, 2003, Petersen, 2006, do Carmo, 1992]. Accordingly, we define a manifold M to be locally (geodesically) complete, if any two points inside a geodesic ball of small enough radius in the interior of M can be joined by a geodesic whose image also lies on the interior of M.

Fifth, we assume the manifold M is of essential dimension d, in volume sense. If we fix any point p in the d-dimensional manifold M, then the volume of a ball of radius r grows in order of r^d when r is small. By extending this, fix $K_v \in (0, 2^{-m}]$, and we say that the manifold M is of essential volume dimension d, if the volume of a geodesic ball of radius r around any point in M is lower bounded by $K_v r^d \omega_d$, for some positive constant K_v and all r small enough.

We are now ready to formally define the class $\mathcal{P}$ of probability distributions that we will consider in our analysis of the minimax problem (2.1).

Definition 15. Fix $\tau_g,\ \tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, with $\tau_g \le \tau_\ell$. Let $\mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$ be the set of compact d-dimensional manifolds M such that:

(1) $M \subset I := [-K_I, K_I]^m \subset \mathbb{R}^m$;
(2) M is of *global reach* at least τ_g, i.e. $\tau(M) \ge \tau_g$, and M is of *local reach* at least τ_ℓ, i.e. for all $p \in M$, there exists a neighborhood U_p in M such that $\tau(U_p) \ge \tau_\ell$;
(3) M is *locally (geodesically) complete* (with respect to τ_g): for all $p \in \mathrm{int}(M)$ and for all $q_1, q_2 \in \mathbb{B}_M(p,\ 2\sqrt{3}\tau_g)$, there exists a geodesic γ joining q_1 and q_2 whose image lies on $intM$;
(4) M is of *essential volume dimension* d (with respect to K_v and τ_g): if for all $p \in M$ and for all $r \le \sqrt{3}\tau_g$, $vol_M(\mathbb{B}_M(p,r)) \ge K_v r^d \omega_d$.

Let $\mathcal{P} = \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}$ be the set of Borel probability distributions P such that:

(5) P is supported on a d-dimensional manifold $M \in \mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$;
(6) P is absolutely continuous with respect to the restriction vol_M of the d-dimensional Hausdorff measure on the supporting manifold M and such that $\sup_{x\in M} \frac{dP}{dvol_M}(x) \le K_p$.

For every $P \in \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}$, denote the dimension of its distribution as $d(P)$.

The regularity conditions in Definition 15 imply further constraints on both the distributions in $\mathcal{P}$ and their supporting manifolds, in Lemma 16, 17, and 18. Such properties are exploited in Section 2.2 and 2.3. The proofs for Lemma 16, 17, and 18 are in Appendix A.1.

Lemma 16. *Fix $\tau_g,\ \tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, with $\tau_g \le \tau_\ell$. For $M \in \mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$ and $r \in (0,\tau_g)$, let $M_r := \{x \in \mathbb{R}^m :\ dist_{\mathbb{R}^m}(x,M) < r\}$ be a r-neighborhood of M in $\mathbb{R}^m$. Then, the volume of M is upper bounded as:*

$$\begin{aligned} vol_M(M) &\le \frac{m!}{d!} r^{d-m} vol_{\mathbb{R}^m}(M_r) \\ &\le C^{(16)}_{K_I,d,m}\left(1 + \tau_g^{d-m}\right), \end{aligned}$$

where $C^{(16)}_{K_I,d,m}$ is a constant depending only on K_I, d and m.

Lemma 17. *Fix $\tau_g,\ \tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, with $\tau_g \le \tau_\ell$. Let $M \in \mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$ and $r \in (0, 2\sqrt{3}\tau_g]$. Then M can be covered by N radius r balls $\mathbb{B}_M(p_1,r)$, ..., $\mathbb{B}_M(p_N,r)$, with*

$$N \le \left\lfloor \frac{2^d vol(M)}{K_v r^d \omega_d} \right\rfloor.$$

Lemma 18. *Fix $\tau_g,\ \tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, with $\tau_g \le \tau_\ell$. Let $M \in \mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$ and let $\exp_{p_k} : \mathcal{E}_k \subset \mathbb{R}^m \to \mathcal{M}$ be an exponential map, where $\mathcal{E}_k$ is the domain of the exponential map $\exp_{p_k}$ and $T_{p_k}M$ is identified with $\mathbb{R}^d$. For all $v, w \in \mathcal{E}_k$, let $R_k := \max\{||v||, ||w||\}$. Then*

$$\|\exp_{p_k}(v) - \exp_{p_k}(w)\|_{\mathbb{R}^m} \le \frac{\sinh(\sqrt{2}R_k/\tau_\ell)}{\sqrt{2}R_k/\tau_\ell}\|v - w\|_{\mathbb{R}^d}.$$

Under these regularity conditions, the minimax risk R_n is defined as

$$R_n = \inf_{\hat{d}_n} \sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}\left[1\left(\hat{d}_n(X) \ne d(P)\right)\right], \tag{2.5}$$

where in Section 2.2 and 2.3 we fix $d_1, d_2 \in \mathbb{N}$ with $1 \le d_1 < d_2 \le m$ and define

$$\mathcal{P} = \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p} \bigcup \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}, \tag{2.6}$$

and in Section 2.4 we set instead

$$\mathcal{P} = \bigcup_{d=1}^{m} \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}. \tag{2.7}$$

In (2.5), $\hat{d}_n$ is any dimension estimator based on data $X = (X_1, \ldots, X_n)$, and the loss function $\ell(\cdot,\cdot)$ is $0-1$ loss, so for all $x, y \in \mathbb{R}$, $\ell(x,y) = 1(x = y)$.

2.2 Upper Bound for Choosing Between Two Dimensions

In this section we provide an upper bound on the minimax rate R_n in (2.5) when $d(P)$ can only take two known values. Fix $d_1,\ d_2 \in \mathbb{N}$ with $1 \le d_1 < d_2 \le m$, and assume that the data are generated from a distribution $P \in \mathcal{P}$ such that either $d(P) = d_1$ or $d(P) = d_2$ as in (2.6). In this case, the minimax risk quantifies the statistical hardness of the hypothesis testing problem of deciding whether the data originate from a d_1 or d_2-dimensional distribution. In Section 2.4 we will relax this assumption and allow for the intrinsic dimension $d(P)$ to be any integer between 1 and m as in (2.7). All the proofs for this section are in Section A.2.

Our strategy to derive an upper bound on R_n is to choose a particular estimator $\hat{d}_n$ and then derive a uniform upper bound on its risk over the class $\mathcal{P}$ in (2.6), i.e. an upper bound for the quantity

$$\sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}} \left[1\left(\hat{d}_n(X) \neq d(P) \right) \right], \tag{2.8}$$

where $P^{(n)}$ denotes the n-fold product of P. This will in turn yield an upper bound on the minimax risk R_n, as explained in (1.3). In Section 2.2.1, we define our dimension estimator $\hat{d}_n$ and analyze its risk. From that analysis, we derive an upper bound on the minimax risk R_n in (2.5) in Section 2.2.2.

2.2.1 Dimension Estimator and its Analysis

Our dimension estimator $\hat{d}_n$ is based on the d_1-squared length of the TSP (Traveling Salesman Path) generated by the data. The d_1-squared length of the TSP generated by the data is the minimal d_1-squared length of all possible paths passing through each sample point X_i once, which is

$$\min_{\sigma\in S_n} \left\{ \sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|^{d_1}_{\mathbb{R}^m} \right\}. \tag{2.9}$$

Then, $\hat{d}_n = d_1$ if and only if the d_1-squared length of the TSP is below a certain threshold; that is

$$\hat{d}_n(X) := \begin{cases} d_1, & if \min\limits_{\sigma\in S_n} \left\{ \sum\limits_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|^{d_1}_{\mathbb{R}^m} \right\} \le C^{(20)}_{K_I,K_v,d_1,m} \left(1 + \tau_g^{d_1-m}\right), \\ d_2, & \text{otherwise.} \end{cases} \tag{2.10}$$

where $C^{(20)}_{K_I,K_v,d_1,m}$ is a constant to be defined later.

We begin our analysis of the estimator $\widehat{d}_n$ with Lemma 19, which shows that $\hat{d}_n$ makes an error with probability of order $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$ if the correct dimension is d_2. Specifically, we demonstrate that, for any positive value L, the d_1-squared length of a piecewise linear path from X_1 to X_n, $\sum\limits_{i=1}^{n-1} \|X_{i+1} -$

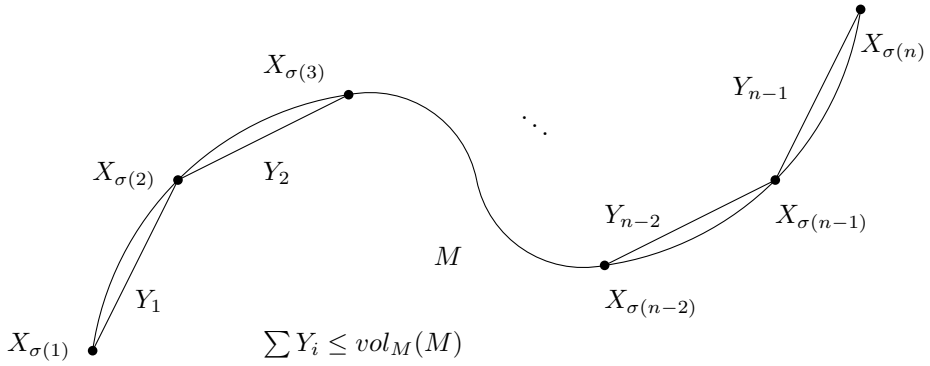


Figure 2.3: When the manifold is a curve, the length of the TSP path $\min_{\sigma \in S_n} \left\{ \sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m} \right\}$ in (2.9) is upper bounded by the length of the curve $vol_M(M)$.

$X_i\|_{\mathbb{R}^m}^{d_1}$, is upper bounded by L with a very small probability of order $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$, as in (2.11). Hence the d_1-squared length of the path is not likely to be bounded by any such threshold L.

Lemma 19. *Fix τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \leq \tau_\ell$ and $1 \leq d_1 < d_2 \leq m$. Let $X_1, \ldots, X_n \sim P \in \mathcal{P}^{d_2}_{\tau_g, \tau_\ell, K_I, K_v, K_p}$. Then for all $L > 0$,*

$$P^{(n)}\left[\sum_{i=1}^{n-1} \|X_{i+1} - X_i\|^{d_1} \leq L\right] \leq \frac{\left(C^{(19)}_{K_I, K_p, d_1, d_2, m}\right)^{n-1} L^{\frac{d_2}{d_1}(n-1)} \left(1 + \tau_g^{(d_2-m)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}(n-1)!}, \tag{2.11}$$

where $C^{(19)}_{K_I, K_p, d_1, d_2, m}$ is a constant depending only on K_I, K_p, d_1, d_2, m.

Next, Lemma 20 shows that the estimator $\hat{d}_n$ in (2.10) is always correct when the intrinsic dimension is d_1, as in (2.12). Specifically, the d_1-squared length of the TSP path in (2.9) is bounded by some positive threshold $C^{(20)}_{K_I, K_v, d_1, m}\left(1 + \tau_g^{d_1 - m}\right)$. We take note that, when $d_1 = 1$, Lemma 20 is straightforward: the length of the TSP path in (2.9) is upper bounded by the length of curve $vol_M(M)$, as in Figure 2.3. This fact, combined with Lemma 16, which shows that $vol_M(M) \leq C^{(16)}_{K_I, 1, m}\left(1 + \tau_g^{1-m}\right)$, yields the result. In particular, the constant $C^{(20)}_{K_I, K_v, d_1, m}$ can be set as $C^{(20)}_{K_I, K_v, d_1, m} = C^{(16)}_{K_I, 1, m}$.

When $d_1 > 1$, Lemma 20 is proved using Lemma 16, 17 and 18, along with the Hölder continuity of a d_1-dimensional space-filling curve [Steele, 1997, Buchin, 2008].

Lemma 20. *Fix τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, $d_1 \in \mathbb{N}$, with $\tau_g \leq \tau_\ell$. Let $M \in \mathcal{M}^{d_1}_{\tau_g, \tau_\ell, K_p, K_v}$ and $X_1, \ldots, X_n \in M$. Then*

$$\min_{\sigma \in S_n} \sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|^{d_1}_{\mathbb{R}^m} \leq C^{(20)}_{K_I, K_v, d_1, m}\left(1 + \tau_g^{d_1 - m}\right), \tag{2.12}$$

where $C^{(20)}_{K_I, K_v, d_1, m}$ is a constant depending only on m, d_1, K_v, and K_I.

Proposition 21 below is the main result of this subsection and follows directly from Lemma 19 and Lemma 20 above. Indeed, when the intrinsic dimension is d_2, the risk of our estimator $\hat{d}_n$, is of order $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$ by Lemma 19 and the union bound. On the other hand, when the intrinsic dimension is d_1, the risk of our estimator $\hat{d}_n$ is 0, because of Lemma 20.

Proposition 21. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \le \tau_\ell$ and $1 \le d_1 < d_2 \le m$. Let $\hat{d}_n$ be in* (2.10)*. Then either for $d = d_1$ or $d = d_2$,*

$$\sup_{P \in \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \le 1(d = d_2)\left(C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}\right)^n \left(1 + \tau_g^{-\left(\frac{d_2}{d_1}m+m-2d_2\right)n}\right) n^{-\left(\frac{d_2}{d_1}-1\right)n},$$

where $C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m} \in (0,\infty)$ is a constant depending only on $K_I, K_p, K_v, d_1, d_2, m$.

As described so far, the convergence analysis of our dimension estimator is probable. This is enough for our purpose, which is to quantify the statistical difficulties, in particular the minimax rate, of the dimension estimation problem. However, our $\widehat{d}_n$ in (2.10) is not completely data-driven but depends on the model parameters τ_g, K_I, and K_v. Hence the model on which our convergence analysis is valid depends on the model parameters. When it comes to applying our dimension estimator $\widehat{d}_n$ to real data, we need to estimate the constant $C^{(20)}_{K_I,K_v,d_1,m}$. Proofs of Lemma 19 and 20 suggest that overestimating $C^{(20)}_{K_I,K_v,d_1,m}$ by some constant factor doesn't deteriorate the convergence rate, so the constants $C^{(20)}_{K_I,K_v,d_1,m}$ and τ_g can be replaced by any consistent estimators. Still, we have the difficulty of tuning the constant $C^{(20)}_{K_I,K_v,d_1,m}$ and τ_g. Also, the constant $C^{(20)}_{K_I,K_v,d_1,m}$ is tuned to work for the worst case, so the practical performance of our dimension estimator is questionable.

2.2.2 Minimax Upper Bound

As noted in the beginning of Section 2.2, the maximum risk of our estimator $\hat{d}_n$ in (2.8) serves as an upper bound on the minimax risk R_n in (2.5). Since we assume that the intrinsic dimension is either d_1 or d_2, Proposition 21 yields that the maximum risk of our estimator $\hat{d}_n$ is of order $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$. This also serves as an upper bound of the minimax risk R_n, as in Proposition 22.

Proposition 22. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \le \tau_\ell$ and $1 \le d_1 < d_2 \le m$. Then*

$$\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \le \left(C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}\right)^n \left(1 + \tau_g^{-\left(\frac{d_2}{d_1}m+m-2d_2\right)n}\right) n^{-\left(\frac{d_2}{d_1}-1\right)n},$$

where $C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}$ is from Proposition 21 and

$$\mathcal{P}_1 = \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p}, \quad \mathcal{P}_2 = \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}.$$

2.3 Lower bound for Choosing Between Two Dimensions

The goal of this section is to derive a lower bound for the minimax rate R_n. As in Section 2.2, we fix d_1, $d_2 \in \mathbb{N}$ with $1 \le d_1 < d_2 \le m$, and assume that the intrinsic dimension of data is either d_1 or d_2 as in (2.6). This assumption is relaxed in Section 2.4. All the proofs for this section are in Section A.3.

Our strategy is to find a subset $T \subset I^n \subset (\mathbb{R}^d)^n$ and two sets of distributions $\mathcal{P}_1^{d_1}$ and $\mathcal{P}_2^{d_2}$ with dimensions d_1 and d_2, such that $\mathcal{P}_1^{d_1}$ and $\mathcal{P}_2^{d_2}$ satisfy the regularity conditions in Definition 15, and whenever the sample $X = (X_1, \ldots, X_n)$ lies on T, one cannot easily distinguish whether the underlying distribution is from $\mathcal{P}_1^{d_1}$ or $\mathcal{P}_2^{d_2}$.

After constructing T, $\mathcal{P}_1^{d_1}$ and $\mathcal{P}_2^{d_2}$, we derive the lower bound using Lemma 1 (Le Cam's Lemma).

In Lemma 1 (Le Cam's Lemma), considering the convex hull of distributions $co(\mathcal{P}_i)$ is critical for getting the nontrivial lower bound. Suppose we are using the basic version of Le Cam's lemma where the convex hull is not considered, i.e. $Q_i \in \mathcal{P}_i$. Then for two distributions Q_1 and Q_2 respectively from our d_1 and d_2 dimensional model $\mathcal{P}^{d_1}_{\tau_g,\tau_l,K_I,K_v,K_p}$ and $\mathcal{P}^{d_2}_{\tau_g,\tau_l,K_I,K_v,K_p}$, Q_1 and Q_2 are singular to each other; i.e. $q_1(x) \wedge q_2(x) = 0$ for all x. Hence no matter which subset $\mathcal{P}_1$ and $\mathcal{P}_2$ we choose with $d(\mathcal{P}_1) = d_1$ and $d(\mathcal{P}_2) = d_2$, the lower bound in (1.4) will be always 0. This trivial bound can be improved by considering the convex hull of distributions $co(\mathcal{P}_i)$ in Le Cam's lemma.

Our construction for T, $\mathcal{P}_1^{d_1}$, and $\mathcal{P}_2^{d_2}$ is based on mimicking a space-filling curve. Intuitively, this gives the lower bound since it is difficult to differentiate a space-filling curve and a higher dimensional cube. In detail, we set

$$\begin{aligned}\mathcal{P}_1^{d_1} = \{&\text{distributions supported on} \\ &\text{a space-filling-curve like } d_1\text{-dimensional manifold}\},\end{aligned} \tag{2.13}$$

and

$$\mathcal{P}_2^{d_2} = \{\text{uniform distributions on } [-K_I,\, K_I]^{d_2}\}. \tag{2.14}$$

To apply Le Cam's lemma, we construct a set $T \subset I^n$ so that, whenever $X = (X_1, \ldots, X_n) \in T$, we cannot distinguish whether X is from $\mathcal{P}_1^{d_1}$ in (2.13) or $\mathcal{P}_1^{d_2}$ in (2.14). Then, for an appropriately chosen distribution Q_1 in the convex hull of $\mathcal{P}_1^{d_1}$ with density q_1 with respect to Lebesgue measure λ on the cube $[-K_I, K_I]^{d_2}$, and a density q_2 from the class $\mathcal{P}_2^{d_2}$, $\int_T [q_1(x) \wedge q_2(x)] d\lambda(x)$ is a lower bound on the minimax rate R_n in (2.5). Indeed, from Le Cam's Lemma 1, we have that

$$\begin{aligned}\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\ell(\hat{\theta}, \theta(P))] &\geq \frac{1}{2} \int [q_1(x) \wedge q_2(x)] d\lambda(x) \\ &\geq \frac{1}{2} \int_T [q_1(x) \wedge q_2(x)] d\lambda(x).\end{aligned} \tag{2.15}$$

For constructing the class $\mathcal{P}_1^{d_1}$ in (2.13), it will be sufficient to consider the case $d_1 = 1$. In fact, Lemma 23 states that the regularity conditions in Definition 15 are still preserved when the manifold M is a Cartesian product with a cube $[-K_I,\, K_I]^{\Delta d}$, as in Figure 2.4. Hence for constructing a d-dimensional "space-filling" manifold, we first construct a 1-dimensional space-filling curve satisfying the required regularity conditions, and then we form a Cartesian product with a cube of dimension $d-1$, which becomes a d-dimensional manifold satisfying the same regularity conditions by Lemma 23.

Lemma 23. *Fix τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, d, $\Delta d \in \mathbb{N}$, with $\tau_g \leq \tau_\ell$ and $1 \leq d + \Delta d \leq m$. Let $M \in \mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$ be a d-dimensional manifold of global reach $\geq \tau_g$, local reach $\geq \tau_\ell$, which is embedded in $\mathbb{R}^{m-\Delta d}$. Then*

$$M \times [-K_I, K_I]^{\Delta d} \in \mathcal{M}^{d+\Delta d}_{\tau_g,\tau_\ell,K_I,K_v},$$

which is embedded in $\mathbb{R}^m$.

The precise construction of $\mathcal{P}_1^{d_1}$ in (2.13) and T is detailed in Lemma 24. As in Figure 2.5, we construct T_i's that are cylinder sets aligned as a zigzag in $[-K_I, K_I]^{d_2}$, and then T is constructed as

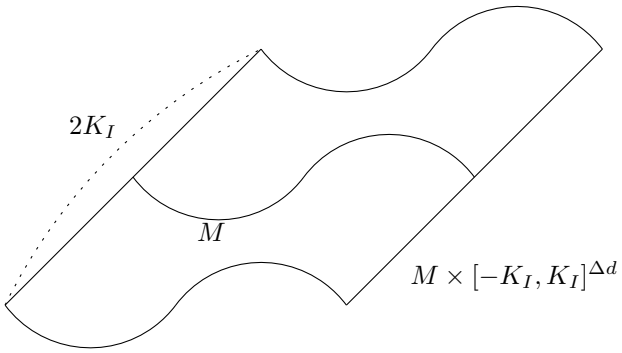


Figure 2.4: The regularity conditions in Definition 15 are still preserved under the Cartesian product with a cube $[-K_I,\ K_I]^{\Delta d}$. Detailed explanations are in Figure A.3.

$T = S_n \prod_{i=1}^{n} T_i$, where the permutation group S_n acts on $\prod_{i=1}^{n} T_i$ as a coordinate change. Then, we show below that, for any $x \in \prod T_i$, there exists a manifold $M \in \mathcal{M}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v}$ that passes through $x_1, \ldots, x_n$. The class $\mathcal{P}_1^{d_1}$ in (2.13) is finally defined as the set of distributions that are supported on such a manifold.

Lemma 24. *Fix $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $d_1,\ d_2 \in \mathbb{N}$, with $1 \le d_1 \le d_2$, and suppose $\tau_\ell < K_I$. Then there exist $T_1, \ldots, T_n \subset [-K_I, K_I]^{d_2}$ such that:*

(1) The T_i's are distinct.

(2) For each T_i, there exists an isometry Φ_i such that

$$T_i = \Phi_i\left([-K_I, K_I]^{d_1-1} \times [0,a] \times \mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0,w)\right),$$

where $c = \left\lceil \frac{K_I+\tau_\ell}{2\tau_\ell} \right\rceil$, $a = \frac{K_I-\tau_\ell}{\left(d_2-d_1+\frac{1}{2}\right)\left\lceil \frac{n}{c^{d_2-d_1}} \right\rceil}$, and $w = \min\left\{ \tau_\ell,\ \frac{(d_2-d_1)^2(K_I-\tau_\ell)^2}{2\tau_\ell\left(d_2-d_1+\frac{1}{2}\right)^2\left(\left\lceil \frac{n}{c^{d_2-d_1}} \right\rceil+1\right)^2} \right\}$.

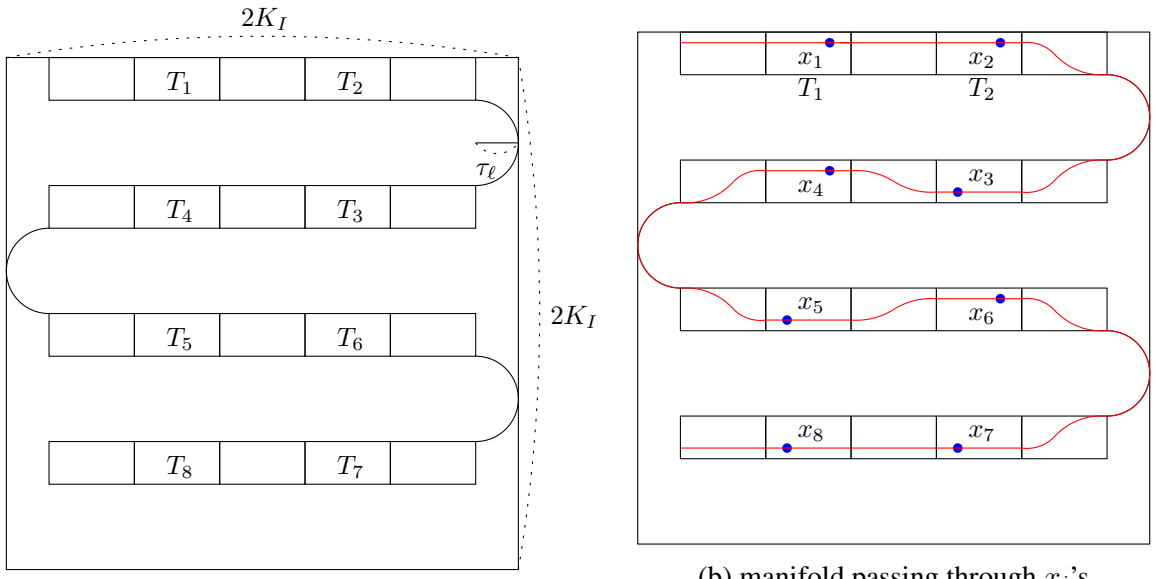
(3)There exists $\mathscr{M} : \left(\mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0,w)\right)^n \to \mathcal{M}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v}$ one-to-one such that for each $y_i \in \mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0,w)$, $1 \le i \le n$, $\mathscr{M}(y_1,\ldots,y_n) \cap T_i = \Phi_i([-K_I,K_I]^{d_1-1} \times [0,a] \times \{y_i\})$. Hence for any $x_1 \in T_1, \ldots, x_n \in T_n$, $\mathscr{M}(\{\Pi^{-1}_{(d_1+1):d_2}\Phi_i^{-1}(x_i)\}_{1\le i\le n})$ passes through $x_1, \ldots, x_n$.

Next we show that whenever $x = (x_1, \ldots, x_n) \in T$, it is difficult to tell whether the data originated from $P \in \mathcal{P}_1^{d_1}$ or $P \in \mathcal{P}_2^{d_2}$. Let Q_1 be in the convex hull of $\mathcal{P}_1^{d_1}$ and let q_2 be the density function of the uniform distribution on $[-K_I, K_I]^{d_2}$, then from (2.15), we know that a lower bound is given by $\int_T [q_1(x) \wedge q_2(x)] d\lambda(x)$. Hence if we can choose Q_1 such that $q_1(x) \ge C q_2(x)$ for every $x \in T$ with $C < 1$, then $q_1(x) \wedge q_2(x) \ge C q_2(x)$, so that $C \int_T q_2(x)$ can serve as lower bound of minimax rate. Such existence of Q_1 and the inequality $q_1(x) \ge C q_2(x)$ is shown in Claim 25.

Claim 25. Let $T = S_n \prod_{i=1}^{n} T_i$ where the T_i's are from Lemma 24. Let Q_2 be the uniform distribution on $[-K_I, K_I]^{d_2}$, and let $\mathcal{P}_1^{d_1}$ be as in (2.13). Then there exists $Q_1 \in co(\mathcal{P}_1^{d_1})$ satisfying that for all $x \in intT$, there exists $r_x > 0$ such that for all $r < r_x$,

$$Q_1\left(\prod_{i=1}^{n} \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)\right) \ge 2^{-n} Q_2\left(\prod_{i=1}^{n} \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)\right).$$

The following lower bound is than a consequence of Le Cam's lemma, Lemma 24, and the previous claim.



(a) alignment of T_i

(b) manifold passing through x_i's

Figure 2.5: This figure illustrates the case where $d_1 = 1$ and $d_2 = 2$. a shows how T_i's are aligned in a zigzag. b shows for given $x_1 \in T_1, \ldots, x_n \in T_n$(represented as blue points), how a manifold with regularity conditions(represented as a red curve) passes through $x_1, \ldots, x_n$. Detailed constructions in Figure A.4.

Proposition 26. *Fix τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \le \tau_\ell$ and $1 \le d_1 < d_2 \le m$, and suppose that $\tau_\ell < K_I$. Then*

$$\inf_{\hat{d}_n} \sup_{P \in \mathcal{Q}} \mathbb{E}_{P^{(n)}}[\ell(\hat{d}_n, d(P))] \geq \left(C^{(26)}_{d_1,d_2,K_I}\right)^n \min\left\{\tau_\ell^{-2(d_2-d_1+1)} n^{-2}, 1\right\}^{(d_2-d_1)n},$$

where $C^{(26)}_{d_1,d_2,K_I} \in (0, \infty)$ is a constant depending only on d_1, d_2, and K_I and

$$\mathcal{Q} = \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p} \bigcup \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}.$$

2.4 Upper Bound and Lower Bound for the General Case

Now we generalize our results to allow the intrinsic dimension d to be any integer between 1 and m. Thus the model is $\mathcal{P} = \bigcup_{d=1}^{m} \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}$ as in (2.7). For the upper bound, we extend the dimension estimator $\hat{d}_n$ in (2.10) and compute its maximum risk. And for the lower bound, we simply use the lower bound derived in Section 2.3 with $d_1 = 1$ and $d_2 = 2$. All the proofs for this section are in Section A.4.

For the model $\mathcal{P}$ in (2.7), our dimension estimator $\hat{d}_n$ estimates the dimension as the smallest integer

$1 \le d \le m$ that the d-squared length of the TSP is below a certain threshold, i.e. (2.12) holds; that is,

$$\hat{d}_n(X) := \min\Bigg\{ d \in [1,m] : \min_{\sigma \in S_n} \Bigg\{ \sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^d \Bigg\} \le C^{(20)}_{K_I,K_v,d,m} \left(1 + \tau_g^{d-m}\right) \Bigg\}. \tag{2.16}$$

As a generalized result of Proposition 21, Proposition 27 gives an upper bound for the risk of our estimator $\hat{d}_n$ in (2.16). When the intrinsic dimension is d, our estimator $\hat{d}_n$ makes an error with probability of order $O\left(n^{-\frac{1}{d-1}n}\right)$.

Proposition 27. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, with $\tau_g \le \tau_\ell$. Let $\hat{d}_n$ be in* (2.16)*. Then:*

$$\sup_{P \in \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}} \left[\ell\left(\hat{d}_n, d(P)\right)\right]$$

$$\begin{cases} = 0, & d = 1, \\ \le \left(C^{(27)}_{K_I,K_p,K_v,d,m}\right)^n \left(1 + \tau_g^{-(dm+m-2d)n}\right) n^{-\frac{1}{d-1}n}, & d > 1. \end{cases}$$

where $C^{(27)}_{K_I,K_p,K_v,d,m} \in (0,\infty)$ is a constant depending only on K_I, K_p, K_v, d, m.

Then similarly to Section 2.2.2, the maximum risk of our estimator $\hat{d}_n$ in (2.16) serves as an upper bound on the minimax risk R_n in (2.5). The maximum of the upper bound in Proposition 27 over d ranging from 1 to m should serve as the upper bound for the maximum risk, hence we get the upper bound of the minimax risk R_n in Proposition 28 as a generalized result of Proposition 22.

Proposition 28. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, with $\tau_g \le \tau_\ell$. Then:*

$$\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell\left(\hat{d}_n, d(P)\right)\right] \le \left(C^{(28)}_{K_I,K_p,K_v,m}\right)^n \left(1 + \tau_g^{-(m^2-m)n}\right) n^{-\frac{1}{m-1}n}$$

where $C^{(28)}_{K_I,K_p,K_v,m} \in (0,\infty)$ is a constant depending only on K_I, K_p, K_v, m.

Proposition 29 provides a lower bound for minimax rate R_n in (2.5), in multi-dimensions. It can be viewed of a generalization for the binary dimension case in Proposition 26.

Proposition 29. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, with $\tau_g \le \tau_\ell$, and suppose that $\tau_\ell < K_I$. Then,*

$$\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}[\ell(\hat{d}_n, d(P))] \ge \left(C^{(29)}_{K_I}\right)^n \min\left\{\tau_\ell^{-4} n^{-2}, 1\right\}^n$$

where $C^{(29)}_{K_I} \in (0,\infty)$ is a constant depending only on K_I.

Chapter 3

The Origin of the Reach: Better Understanding Regularity Through Minimax Estimation Theory

This chapter presents the work in [Aamari et al., 2017].

Complexity and regularity notions play a central role in estimation topics. When dealing with high dimensional data, a classical assumption is that a low dimensional curved structure underlies the studied phenomenon. This setting gave birth to global geometric methods among which manifold learning and topological data analysis. As in other fields of data analysis, regularity and scale parameters often remain to be tuned by the user when dealing with real data. In such frameworks, what arise naturally are intrinsic geometric quantities. Indeed, usual differential regularity notions are not relevant as they are very dependent to a specific coordinate system or parametrization.

First introduced by Federer Federer [1959] , the reach τ_M of $M \subset \mathbb{R}^m$ is the largest length such that any point at distance less than τ_M of M has a unique nearest neighbor on M. For a set, having reach greater than $\tau_{min} > 0$ roughly means that one can roll freely a ball of radius τ_{min} around it Cuevas et al. [2012]. The reach informs on maximal directional curvature and on the width of possible narrow bottleneck structures on the shape. It corresponds to a minimal size of features M contains. In a view to inference, this gives a minimal scale at which look at data. In statistical settings, such a scale corresponds to the least sampling density needed to recover geometric information.

Positive reach has been the minimal regularity assumption on sets in geometric measure theory Federer [1969], Thäle [2008]. Sets with positive reach enjoy a structure close to be differential — the so-called tangent and normal cones — and behave well in integral geometry. Since sets with positive reach enjoy good geometric Federer [1969], Thäle [2008] and statistical properties Cuevas et al. [2012], it has recently grown popular in the literature. In manifold reconstruction, the reach helps formalizing in a simple way models on which minimax rates are well posed Genovese et al. [2012], Kim and Zhou [2015]. The effective optimal estimators of Boissonnat and Ghosh [2014], Aamari and Levrard [2015] implicitly use it as a scale parameter in their construction. In homology inference Niyogi et al. [2008], Balakrishnan et al. [2013b], the reach drives the minimal sample size required to consistently estimate topological invariants, and their recovery probability. It emerges in Cuevas et al. [2007] as a regularity parameter in the estimation of Minkovski boundary lengths and surface areas. The reach has been explicitly used in geometric inference, volume estimation Arias-Castro et al. [2016] and manifold clustering Arias-Castro et al. [2013]. It is also a good regularity notion for dimension reduction techniques such as vector diffusions maps Singer and Wu [2012]. Computational geometry also makes use of it in

deterministic settings Boissonnat and Ghosh [2014].

This chapter gives new geometric results on what the reach relates to, and tackles the question of its estimation, in both deterministic and minimax frameworks. Formally, given a class of probability distribution $\mathcal{P}$, the minimax risk $R_n = R_n(\mathcal{P})$ is defined as

$$R_n = \inf_{\hat{\tau}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\left| \frac{1}{\tau(P)} - \frac{1}{\hat{\tau}_n} \right|^r \right]. \tag{3.1}$$

In (3.1), $\tau(P)$ is the reach of the support of P, $\mathbb{E}_P$ denotes the expectation with respect to the distribution P, and the infimum is over all estimators (measurable functions of the data) $\hat{\tau} = \hat{\tau}(X_1, \ldots, X_n)$ of the reach $\tau(P)$. The minimax risk $R_n(\mathcal{P})$ has an interpretation that *any reach estimator* cannot have a risk smaller than R_n uniformly over every $P \in \mathcal{P}$.

In our model, we assumed that tangent spaces are observed at all the sample points. In other words, we assume that when $X_1, \ldots, X_n$ are observed, $T_{X_1}M, \ldots, T_{X_n}M$ are observed as well.

3.1 Statistical Model and Loss

Let us now describe the regularity assumptions we will use throughout. To avoid arbitrarily irregular shapes, we consider submanifolds M with their reach lower bounded by $\tau_{min} > 0$. Since the parameter of interest τ_M is a $\mathcal{C}^2$-like quantity, it is natural — and actually necessary, as we shall see in Proposition 33 — to require an extra degree of smoothness. For example, by imposing an upper bound on the third order derivatives of geodesics.

Definition 30. We let $\mathcal{M}^{d,m}_{\tau_{min},L}$ denote the set of compact connected d-dimensional submanifolds $M \subset \mathbb{R}^m$ without boundary such that $\tau_M \geq \tau_{min}$, and for which every arc-length parametrized geodesic $\gamma_{p,v}$ is $\mathcal{C}^3$ and satisfies

$$\left\| \gamma'''_{p,v}(0) \right\| \leq L. \tag{3.2}$$

The regularity bounds τ_{min} and L are assumed to exist only for the purpose of deriving uniform estimation bounds. However, we emphasize the fact that the forthcoming estimator $\hat{\tau}$ (3.4) does not require them in its construction.

It is important to note that any compact d-dimensional $\mathcal{C}^3$-submanifold $M \subset \mathbb{R}^m$ belongs to such a class $\mathcal{M}^{d,m}_{\tau_{min},L}$, provided that $\tau_{min} \leq \tau_M$ and that L is large enough. Note also that since the third order condition $\left\| \gamma'''_{p,v}(0) \right\| \leq L$ needs to hold for all (p, v), we have in particular that $\left\| \gamma'''_{p,v}(t) \right\| \leq L$ for all $t \in \mathbb{R}$. To our knowledge, such a quantitative $\mathcal{C}^3$ assumption on the geodesic trajectories has not been considered in the computational geometry literature.

Any submanifold $M \subset \mathbb{R}^m$ of dimension d inherits a natural measure vol_M from the d-dimensional Hausdorff measure $\mathcal{H}^d$ on $\mathbb{R}^m$ [Federer, 1959, p. 171]. We will consider distributions Q that have densities with respect to vol_M that are bounded away from zero.

Definition 31. We let $\mathcal{Q}^{d,m}_{\tau_{\min},L,f_{\min}}$ denote the set of distributions Q having support $M \in \mathcal{M}^{d,m}_{\tau_{min},L}$ and with a Hausdorff density $f = \frac{dQ}{dvol_M}$ satisfying $\inf_{x \in M} f(x) \geq f_{\min} > 0$ on M.

As for τ_{min} and L, the knowledge of f_{min} will not be required in the construction of the estimator $\hat{\tau}$ (3.4) described below.

In order to focus on the geometric aspects of the reach, we will first consider the case where tangent spaces are observed at all the sample points. As mentioned in the introduction, the knowledge of tangent spaces is a reasonable assumption in digital imaging Klette and Rosenfeld [2004].

We let $\mathbb{G}^{d,m}$ denote the Grassmanian of dimension d of $\mathbb{R}^m$, that is the set of all d-dimensional linear subspaces of $\mathbb{R}^m$.

Definition 32. For any distribution $Q \in \mathcal{Q}^{d,m}_{\tau_{\min},L,f_{\min}}$ with support M we associate the distribution P of the random variable $(X, T_X M)$ on $\mathbb{R}^m \times \mathbb{G}^{d,m}$, where X has distribution Q. We let $\mathcal{P}^{d,m}_{\tau_{\min},L,f_{\min}}$ denote the set of all such distributions.

Formally, one can write $P(dx\, dT) = \delta_{T_x M}(dT)Q(dx)$, where $\delta_{\cdot}$ denotes the Dirac measure. An i.i.d. n-sample of P is of the form $(X_1, T_1), \ldots, (X_n, T_n) \in \mathbb{R}^m \times \mathbb{G}^{d,m}$, where $X_1, \ldots, X_n$ is an i.i.d. n-sample of Q and $T_i = T_{X_i} M$ with $M = supp(Q)$. For a distribution Q with support M and associated distribution P on $\mathbb{R}^m \times \mathbb{G}^{d,m}$, we will write $\tau_P = \tau_Q = \tau_M$, with a slight abuse of notation.

To simplify the statements and the proofs, we focus on a loss involving the condition number. Namely, we measure the error with the loss

$$\ell(\tau, \tau') = \left| \frac{1}{\tau} - \frac{1}{\tau'} \right|^p, \quad p \geq 1. \tag{3.3}$$

In other words, we will consider the estimation of the condition number τ_M^{-1} instead of the reach τ_M.

With the statistical framework developed above, we can now see explicitly why the third order condition $\|\gamma'''\| \leq L < \infty$ is necessary. Indeed, the following Proposition 33 demonstrates that relaxing this constraint — *i.e.* setting $L = \infty$ — renders the problem of reach estimation intractable. Below, σ_d stands for the volume of the d-dimensional unit sphere $\mathcal{S}^d$.

Proposition 33. *Given $\tau_{min} > 0$, provided that $f_{min} \leq (2^{d+1}\tau_{min}^d \sigma_d)^{-1}$, we have for all $n \geq 1$,*

$$\inf_{\hat{\tau}_n} \sup_{P \in \mathcal{P}^{d,m}_{\tau_{min},L=\infty,f_{min}}} \mathbb{E}_{P^n} \left| \frac{1}{\tau_P} - \frac{1}{\hat{\tau}_n} \right|^p \geq \frac{c_p}{\tau_{min}^p} > 0,$$

where the infimum is taken over the estimators $\hat{\tau}_n = \hat{\tau}_n(X_1, T_1, \ldots, X_n, T_n)$.

Thus, one cannot expect to derive consistent uniform approximation bounds for the reach solely under the condition $\tau_M \geq \tau_{min}$. This result is natural, since the problem at stake is to estimate a differential quantity of order two. Therefore, some notion of uniform $\mathcal{C}^3$ regularity is needed.

3.2 Geometry of the Reach

In this section, we give a precise geometric description of how the reach arises. In particular, below we will show that the reach is determined either by a bottleneck structure or an area of high curvature (Theorem 37). These two cases are referred to as *global* reach and *local* reach, respectively. All the proofs for this section are to be found in Section B.2.

Consider the formulation (1.6) of the reach as the infimum of the distance between M and its medial axis $Med(M)$. By definition of the medial axis (1.5), if the infimum is attained it corresponds to a point z_0 in $Med(M)$ at distance τ_M from M, which we call an *axis point*. Since z_0 belongs to the medial axis of M, it has at least two nearest neighbors q_1, q_2 on M, which we call a *reach attaining pair* (see Figure 3.1b). By definition, q_1 and q_2 belong to $\mathbb{B}(z_0, \tau_M)$ and cannot be farther than $2\tau_M$ from each other. We say that (q_1, q_2) is a *bottleneck* of M in the extremal case $\|q_2 - q_1\| = 2\tau_M$ of antipodal points of $\mathbb{B}(z_0, \tau_M)$ (see Figure 3.1a). Note that the ball $\mathbb{B}(z_0, \tau_M)$ meets M only on its boundary $\partial\mathbb{B}(z_0, \tau_M)$.

Definition 34. Let $M \subset \mathbb{R}^m$ be a submanifold with reach $\tau_M > 0$.

- A pair of points (q_1, q_2) in M is called *reach attaining* if there exists $z_0 \in Med(M)$ such that $q_1, q_2 \in \mathbb{B}(z_0, \tau_M)$. We call z_0 the *axis point* of (q_1, q_2), and $\|q_1 - q_2\| \in (0, 2\tau_M]$ its *size*.

- A reach attaining pair $(q_1, q_2) \in M^2$ is said to be a *bottleneck* of M if its size is $2\tau_M$, that is $\|q_1 - q_2\| = 2\tau_M$.

As stated in the following Lemma 35, if a reach attaining pair is not a bottleneck — that is $\|q_1 - q_2\| < 2\tau_M$, as in Figure 3.1b —, then M contains an arc of a circle of radius τ_M. In this sense, this "semi-local" case — when $\|q_1 - q_2\|$ can be arbitrarily small — is not generic. Though, we do not exclude this case in the analysis.

Lemma 35. *Let $M \subset \mathbb{R}^m$ be a compact submanifold with reach $\tau_M > 0$. Assume that M has a reach attaining pair $(q_1, q_2) \in M^2$ with size $\|q_1 - q_2\| < 2\tau_M$. Let $z_0 \in Med(M)$ be their associated axis point, and write $c_{z_0}(q_1, q_2)$ for the arc of the circle with center z_0 and endpoints as q_1 and q_2.*

Then $c_{z_0}(q_1, q_2) \subset M$, and this arc (which has constant curvature $1/\tau_M$) is the geodesic joining q_1 and q_2.

In particular, in this "semi-local" situation, since τ_M^{-1} is the norm of the second derivative of a geodesic of M (the exhibited arc of the circle of radius τ_M), the reach can be viewed as arising from directional curvature.

Now consider the case where the infimum (1.6) is not attained. In this case, the following Lemma 36 asserts that τ_M is created by curvature.

Lemma 36. *Let $M \subset \mathbb{R}^m$ be a compact submanifold with reach $\tau_M > 0$. Assume that for all $z \in Med(M)$, $d(z, M) > \tau_M$. Then there exists $q_0 \in M$ and a geodesic γ_0 such that $\gamma_0(0) = q_0$ and $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$.*

To summarize, there are three distinct geometric instances in which the reach may be realized:

- (See Figure 3.1a) M has a bottleneck: by definition, τ_M originates from a structure having scale $2\tau_M$.
- (See Figure 3.1b) M has a reach attaining pair but no bottleneck: then M contains an arc of a circle of radius τ_M (Lemma 35), so that M actually contains a zone with radius of curvature τ_M.
- (See Figure 3.1c) M does not have a reach attaining pair: then τ_M comes from a curvature-attaining point (Lemma 36), that is a point with radius of curvature τ_M.

From now on, we will treat the first case separately from the other two. We are now in a position to state the main result of this section. It is a straightforward consequence of Lemma 35 and Lemma 36.

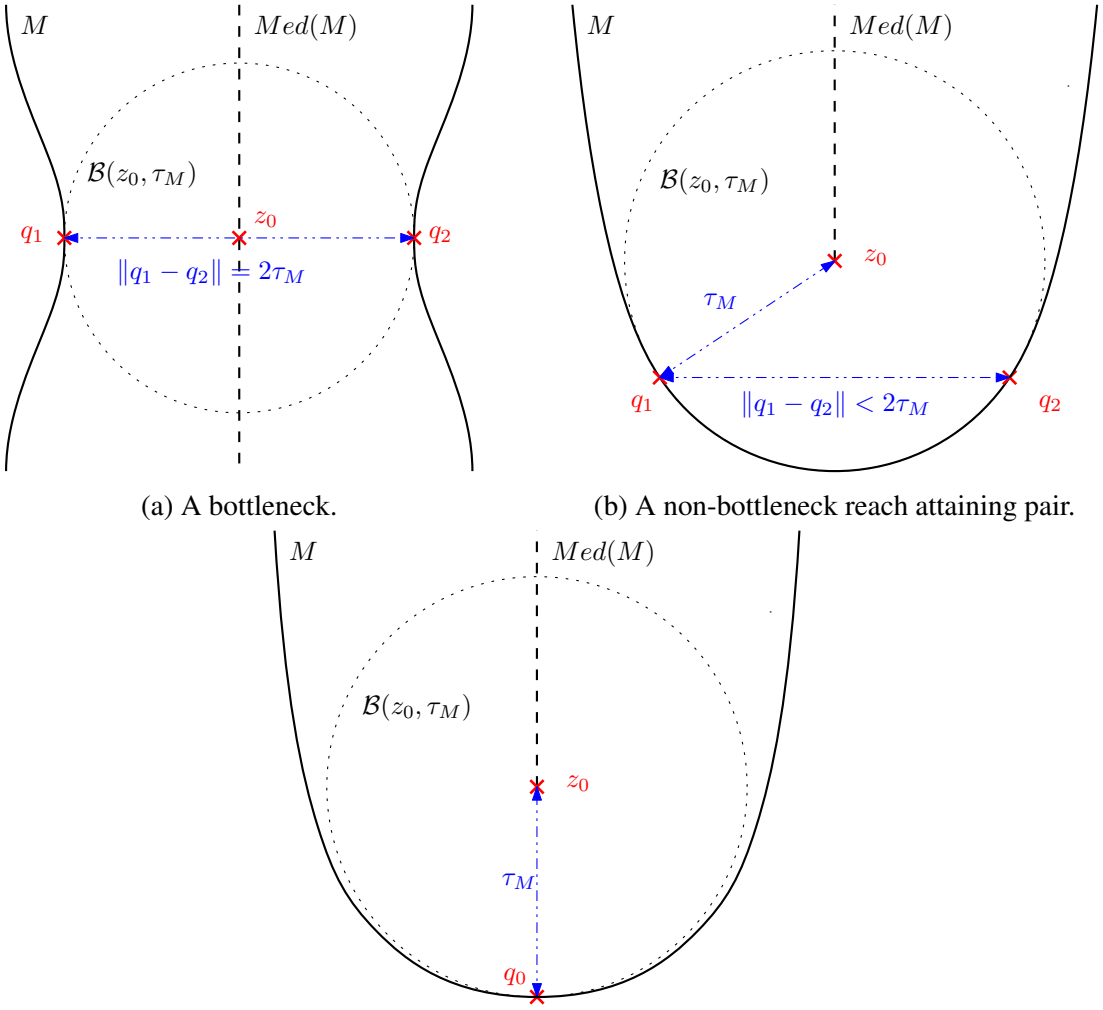
Theorem 37. *Let $M \subset \mathbb{R}^m$ be a compact submanifold with reach $\tau_M > 0$. At least one of the following two assertions holds.*

- (Global Case) *M has a bottleneck $(q_1, q_2) \in M^2$, that is, there exists $z_0 \in Med(M)$ such that $q_1, q_2 \in \partial\mathbb{B}(z_0, \tau_M)$ and $\|q_1 - q_2\| = 2\tau_M$.*
- (Local Case) *There exists $q_0 \in M$ and an arc-length parametrized geodesic γ_0 such that $\gamma_0(0) = q_0$ and $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$.*

Let us emphasize the fact that the global case and the local case of Theorem 37 are not mutually exclusive. Theorem 37 provides a description of the reach as arising from global and local geometric structures that, to the best of our knowledge, is new. Such a distinction is especially important in our problem. Indeed, the global and local cases may yield different approximation properties and require different statistical analyses. However, since one does not know a priori whether the reach arises from a global or a local structure, an estimator of τ_M should be able to handle both cases simultaneously.

3.2.1 Reach Estimator and its Analysis

In this section, we propose an estimator $\hat{\tau}(\cdot)$ for the reach and demonstrate its properties and rate of consistency under the loss (3.3). For the sake of clarity in the analysis, we assume the tangent spaces



(a) A bottleneck.

(b) A non-bottleneck reach attaining pair.

(c) Curvature-attaining point.

Figure 3.1: The different ways for the reach to be attained, as described in Lemma 35 and Lemma 36.

to be known at every sample point.

We rely on the formulation of the reach given in (1.7) (see also Figure 1.1), and define $\hat{\tau}$ as a plugin estimator as follows: given a point cloud $\mathcal{X} \subset M$,

$$\hat{\tau}(\mathcal{X}) = \inf_{x \neq y \in \mathcal{X}} \frac{\|y - x\|^2}{2d(y - x, T_x M)}. \tag{3.4}$$

In particular, we have $\hat{\tau}(M) = \tau_M$. Since the infimum (3.4) is taken over a set $\mathcal{X}$ smaller than M, $\hat{\tau}(\mathcal{X})$ always overestimates τ_M. In fact, $\hat{\tau}(\mathcal{X})$ is decreasing in the number of distinct points in $\mathcal{X}$, a useful property that we formalize in the following result, whose proof is immediate.

Corollary 38. *Let M be a submanifold with reach τ_M and $\mathcal{Y} \subset \mathcal{X} \subset M$ be two nested subsets. Then $\hat{\tau}(\mathcal{Y}) \geq \hat{\tau}(\mathcal{X}) \geq \tau_M$.*

We now derive the rate of convergence of $\hat{\tau}$. We analyze the global case (Section 3.2.2) and the local case (Section 3.2.3) separately. In both cases, we first determine the performance of the estimator in a deterministic framework, and then derive an expected loss bounds when $\hat{\tau}$ is applied to a random sample.

Respectively, the proofs for Section 3.2.2 and Section 3.2.3 are to be found in Section B.3.1 and Section B.3.2.

3.2.2 Global Case

Consider the global case, that is, M has a bottleneck structure (Theorem 37). Then the infimum (1.7) is achieved at a bottleneck pair $(q_1, q_2) \in M^2$. When $\mathcal{X}$ contains points that are close to q_1 and q_2, one may expect that the infimum over the sample points should also be close to (1.7): that is, that $\hat{\tau}(\mathcal{X})$ should be close to τ_M.

Proposition 39. *Let $M \subset \mathbb{R}^m$ be a submanifold with reach $\tau_M > 0$ that has a bottleneck $(q_1, q_2) \in M^2$ (see Definition 34), and $\mathcal{X} \subset M$. If there exist $x, y \in \mathcal{X}$ with $\|q_1 - x\| < \tau_M$ and $\|q_2 - y\| < \tau_M$, then*

$$0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X})} \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x, y\})} \leq \frac{9}{2\tau_M^2} \max\left\{d_M(q_1, x), d_M(q_2, y)\right\}.$$

The error made by $\hat{\tau}(\mathcal{X})$ decreases linearly in the maximum of the distances to the critical points q_1 and q_2. In other words, the radius of the tangent sphere in Figure 1.1 grows at most linearly in t when we perturb by $t < \tau_M$ its basis point $p = q_1$ and the point $q = q_2$ it passes through.

Based on the deterministic bound in Proposition 39, we can now give an upper bound on the expected loss under the model $\mathcal{P}^{d,m}_{\tau_{\min}, L, f_{\min}}$. We recall that, throughout this chapter, $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ is an i.i.d. sample with common distribution Q associated to P (see Definition 32).

Proposition 40. *Let $P \in \mathcal{P}^{d,m}_{\tau_{\min}, L, f_{\min}}$ and $M = supp(P)$. Assume that M has a bottleneck $(q_1, q_2) \in M^2$ (see Definition 34). Then,*

$$\mathbb{E}_{P^n}\left[\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right|^p\right] \leq C_{p,d,\tau_M,f_{\min}} n^{-\frac{p}{d}},$$

where $C_{p,d,\tau_M,f_{\min}}$ depends only on p,d,τ_M and $f_{\min}$, and is a decreasing function of τ_M.

Proposition 40 follows straightforwardly from Proposition 39 combined with the fact that with high probability, the balls centered at the bottleneck points q_1 and q_2 with radii $\mathcal{O}(n^{-1/d})$ both contain a sample point of $\mathcal{X}_n$.

3.2.3 Local Case

Consider now the local case, that is, there exists $q_0 \in M$ and $v_0 \in T_{q_0}M$ such that the geodesic $\gamma_0 = \gamma_{q_0,v_0}$ has second derivative $\|\gamma_0''(0)\| = 1/\tau_M$ (Theorem 37). Estimating τ_M boils down to estimating the curvature of M at q_0 in the direction v_0.

We first relate directional curvature to the increment $\frac{\|y-x\|^2}{2d(y-x,T_xM)}$ involved in the estimator $\hat{\tau}$ (3.4). Indeed, since the latter quantity is the radius of a sphere tangent at x and passing through y (Figure 1.1), it approximates the radius of curvature in the direction $y-x$ when x and y are close. For $x, y \in M$, we let $\gamma_{x\to y}$ denote the arc-length parametrized geodesic joining x and y, with the convention $\gamma_{x\to y}(0) = x$.

Lemma 41. *Let $M \in \mathcal{M}^{d,m}_{\tau_{\min},L}$ with reach τ_M and $\mathcal{X} \subset M$ be a subset. Let $x, y \in \mathcal{X}$ with $d_M(x,y) < \pi\tau_M$. Then,*

$$0 \le \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X})} \le \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x,y\})} \le \frac{1}{\tau_M} - \left\|\gamma''_{x\to y}(0)\right\| + \frac{2}{3}Ld_M(x,y).$$

Let us now state how directional curvatures are stable with respect to perturbations of the base point and the direction. We let κ_p denote the maximal directional curvature of M at $p \in M$, that is,

$$\kappa_p = \sup_{v \in \mathbb{B}_{T_pM}(0,1)} \left\|\gamma''_{p,v}(0)\right\|.$$

Lemma 42. *Let $M \in \mathcal{M}^{d,m}_{\tau_{\min},L}$ with reach τ_M and $q_0, x, y \in M$ be such that $x, y \in \mathbb{B}_M\left(q_0, \frac{\pi\tau_M}{2}\right)$. Let γ_0 be a geodesic such that $\gamma_0(0) = q_0$ and $\|\gamma_0''(0)\| = \kappa_{q_0}$. Write*

$$\theta_x := \angle(\gamma_0'(0), \gamma'_{q_0\to x}(0)), \quad \theta_y := \angle(\gamma_0'(0), \gamma'_{q_0\to y}(0)),$$

and suppose that $|\theta_x - \theta_y| \ge \frac{\pi}{2}$. Then,

$$\begin{aligned}&\left\|\gamma''_{x\to y}(0)\right\| \\ &\ge \kappa_{q_0} - \frac{1}{\sqrt{2}-1}\left(\kappa_x - \kappa_{q_0} + \sqrt{2}(3\kappa_{q_0}+\kappa_x)\sin^2(|\theta_x-\theta_y|) + \sqrt{2}Ld_M(q_0,x)\right).\end{aligned}$$

In particular, geodesics in a neighborhood of q_0 with directions close to v_0 have curvature close to $\frac{1}{\tau_M}$. A point cloud $\mathcal{X}$ sampled densely enough in M would contain points in this neighborhood. Hence combining Lemma 41 and Lemma 42 yields the following deterministic bound in the local case.

Proposition 43. *Under the same conditions as Lemma 42,*

$$\begin{aligned}0 \le \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X})} &\le \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x,y\})} \\ &\le \frac{4\sqrt{2}\sin^2(|\theta_x-\theta_y|)}{(\sqrt{2}-1)\tau_M} + L\left(\frac{2}{3}d_M(x,y) + \frac{\sqrt{2}}{\sqrt{2}-1}d_M(q_0,x)\right).\end{aligned}$$

In other words, since the reach boils down to directional curvature in the local case, $\hat{\tau}$ performs well if it is given as input a pair of points x, y which are close to the point q_0 realizing the reach, and almost aligned with the direction of interest v_0. Note that the error bound in the local case (Proposition 43) is very similar to that of the global case (Proposition 39) with an extra alignment term $\sin^2(|\theta_x - \theta_y|)$. This alignment term appears since, in the local case, the reach arises from directional curvature $\tau_M = \left\|\gamma''_{q_0,v_0}(0)\right\|$ (Theorem 37). Hence, it is natural that the accuracy of $\hat{\tau}(\mathcal{X})$ depends on how precisely $\mathcal{X}$ samples the neighborhood of q_0 in the particular direction v_0.

Similarly to the analysis of the global case, the deterministic bound in Proposition 43 yields a bound on the risk of $\hat{\tau}(\mathcal{X}_n)$ when $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ is random.

Proposition 44. *Let $P \in \mathcal{P}^{d,m}_{\tau_{\min},L,f_{\min}}$ and $M = supp(P)$. Suppose there exists $q_0 \in M$ and a geodesic γ_0 with $\gamma_0(0) = q_0$ and $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$. Then,*

$$\mathbb{E}_{P^n}\left[\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right|^p\right] \leq C_{\tau_{\min},d,L,f_{\min},p} n^{-\frac{2p}{3d-1}},$$

where $C_{\tau_{\min},d,L,f_{\min},p}$ depends only on $\tau_{\min}, d, L, f_{\min}$ and p.

This statement follows from Proposition 43 together with the estimate of the probability of two points being drawn in a neighborhood of q_0 and subject to an alignment constraint.

Proposition 40 and 44 yield a convergence rate of $\hat{\tau}(\mathcal{X}_n)$ which is slower in the local case than in the global case. Recall that from Theorem 37, the reach pertains to the size of a bottleneck structure in the global case, and to maximum directional curvature in the local case. To estimate the size of a bottleneck, observing two points close to each point in the bottleneck gives a good approximation. However, for approximating maximal directional curvature, observing two points close to the curvature attaining point is not enough, but they should also be aligned with the highly curved direction. Hence, estimating the reach may be more difficult in the local case, and the difference in the convergence rates of Proposition 40 and 44 accords with this intuition.

Finally, let us point out that in both cases, neither the convergence rates nor the constants depend on the ambient dimension D.

3.3 Minimax Estimates

In this section we derive bounds on the minimax risk R_n of the estimation of the reach over the class $\mathcal{P}^{d,m}_{\tau_{\min},L,f_{\min}}$, that is

$$R_n = \inf_{\hat{\tau}_n} \sup_{P \in \mathcal{P}^{d,m}_{\tau_{\min},L,f_{min}}} \mathbb{E}_{P^n}\left|\frac{1}{\tau_P} - \frac{1}{\hat{\tau}_n}\right|^p, \tag{3.5}$$

where the infimum ranges over all estimators $\hat{\tau}_n\big((X_1, T_{X_1}), \ldots, (X_n, T_{X_n})\big)$ based on an i.i.d. sample of size n with the knowledge of the tangent spaces at sample points.

The rate of convergence of the plugin estimator $\hat{\tau}(\mathcal{X}_n)$ studied in the previous section leads to an upper bound on R_n as explained in (1.3), which we state here for completeness.

Theorem 45. *For all $n \geq 1$,*

$$R_n \leq C_{\tau_{\min},d,L,f_{\min},p} n^{-\frac{2p}{3d-1}},$$

for some constant $C_{\tau_{\min},d,L,f_{\min},p}$ depending only on $\tau_{\min}, d, L, f_{\min}$ and p.

We now focus on deriving a lower bound on the minimax risk R_n. The method relies on an application of Le Cam's Lemma Yu [1997]. In what follows, let

$$TV(P, P') = \frac{1}{2}\int |dP - dP'|$$

denote the total variation distance between P and P', where dP, dP' denote the respective densities of P, P' with respect to any dominating measure. Since $|x - z|^p + |z - y|^p \geq 2^{1-p}|x - y|^p$, the following version of Le Cam's lemma results from Lemma 1 in Yu [1997] and $(1 - TV(P^n, P'^n)) \geq (1 - TV(P, P'))^n$.

Lemma 46 (Le Cam's Lemma)**.** *Let $P, P' \in \mathcal{P}^{d,m}_{\tau_{\min},L,f_{\min}}$ with respective supports M and M'. Then for all $n \geq 1$,*

$$R_n \geq \frac{1}{2^p}\left|\frac{1}{\tau_M} - \frac{1}{\tau_{M'}}\right|^p (1 - TV(P, P'))^n.$$

Lemma 46 states that in order to derive a lower bound on R_n one needs to consider distributions (hypotheses) in the model that are stochastically close to each other — i.e. with small total variation distance — but for which the associated reaches are as different as possible. A lower bound on the minimax risk over $\mathcal{P}^{d,m}_{\tau_{min},L,f_{min}}$ requires the hypotheses to belong to the class. Luckily, in our problem it will be enough to construct hypotheses from the simpler class $\mathcal{Q}^{d,m}_{\tau_{\min},L,f_{\min}}$. Indeed, we have the following isometry result between $\mathcal{Q}^{d,m}_{\tau_{\min},L,f_{\min}}$ and $\mathcal{P}^{d,m}_{\tau_{min},L,f_{min}}$ for the total variation distance, as proved in Section B.4.2.

Lemma 47. *In accordance with the notation of Definition 32, let $Q, Q' \in \mathcal{Q}^{d,m}_{\tau_{\min},L,f_{\min}}$ be distributions on $\mathbb{R}^m$ with associated distributions $P, P' \in \mathcal{P}^{d,m}_{\tau_{\min},L,f_{\min},}$ on $\mathbb{R}^m \times \mathbb{G}^{d,m}$. Then,*

$$TV\left(P, P'\right) = TV\left(Q, Q'\right).$$

In order to construct hypotheses in $\mathcal{Q}^{d,m}_{\tau_{\min},L,f_{\min}}$ we take advantage of the fact that the class $\mathcal{M}^{d,m}_{\tau_{min},L}$ has good stability properties, which we now describe. Here, since submanifolds do not have natural parametrizations, the notion of perturbation can be well formalized using diffeomorphisms of the ambient space $\mathbb{R}^m \supset M$. Given a smooth map $\Phi : \mathbb{R}^m \to \mathbb{R}^m$, we denote by $d^i_x\Phi$ its differential of order i at x. Given a tensor field A between Euclidean spaces, let $\|A\|_{op} = \sup_x \|A_x\|_{op}$, where $\|A_x\|_{op}$ is the operator norm induced by the Euclidean norm. The next result states, informally, that the reach and geodesics third derivatives of a submanifold that is perturbed by a diffeomorphism that is $\mathcal{C}^3$-close to the identity map do not change much. The proof of Proposition 48 can be found in Section B.4.3.

Proposition 48. *Let $M \in \mathcal{M}^{d,m}_{\tau_{min}L}$ be fixed, and let $\Phi : \mathbb{R}^m \to \mathbb{R}^m$ be a global $\mathcal{C}^3$-diffeomorphism. If $\|I_D - d\Phi\|_{op}$, $\|d^2\Phi\|_{op}$ and $\|d^3\Phi\|_{op}$ are small enough, then $M' = \Phi(M) \in \mathcal{M}^{d,m}_{\frac{\tau_{min}}{2},2L}$.*

Now we construct the two hypotheses Q, Q' as follows (see Figure 3.2). Take M to be a d-dimensional sphere and Q to be the uniform distribution on it. Let $M' = \Phi(M)$, where Φ is a bump-like diffeomorphism having the curvature of M' to be different of that of M in some small neighborhood. Finally, let Q' be the uniform distribution on M'. The proof of Proposition 49 is to be found in Section B.4.3.

Proposition 49. *Assume that $L \geq (2\tau^2_{min})^{-1}$ and $f_{min} \leq (2^{d+1}\tau^d_{min}\sigma_d)^{-1}$. Then for $\ell > 0$ small enough, there exist $Q, Q' \in \mathcal{Q}^{d,m}_{\tau_{\min},L,f_{min}}$ with respective supports M and M' such that*

$$\left|\frac{1}{\tau_M} - \frac{1}{\tau_{M'}}\right| \geq c_d \frac{\ell}{\tau^2_{min}} \quad \textit{and} \quad TV\left(Q, Q'\right) \leq 12\left(\frac{\ell}{2\tau_{min}}\right)^d.$$

Hence, applying Lemma 46 with the hypotheses P, P' associated to Q, Q' of Proposition 49, and taking $12\left(\ell/2\tau_{min}\right)^d = 1/n$, together with Lemma 47, yields the following lower bound.

Proposition 50. *Assume that $L \geq (2\tau^2_{min})^{-1}$ and $f_{min} \leq (2^{d+1}\tau^d_{min}\sigma_d)^{-1}$. Then for n large enough,*

$$R_n \geq \frac{c_{d,p}}{\tau^p_{min}} n^{-p/d},$$

where $c_{d,p}$ depends only on d and p.

Here, the assumptions on the parameters L and f_{min} are necessary for the model to be rich enough. Roughly speaking, they ensure at least that a sphere of radius $2\tau_{min}$ belongs to the model.

From Proposition 50, the plugin estimation $\hat{\tau}(\mathcal{X}_n)$ provably achieves the optimal rate in the global case (Theorem 40) up to numerical constants. In the local case (Theorem 44) the rate obtained presents a gap, yielding a gap in the overall rate. As explained above (Section 3.2.3), the slower rate in the local case is a consequence of the alignment required in order to estimate directional curvature. Though, let

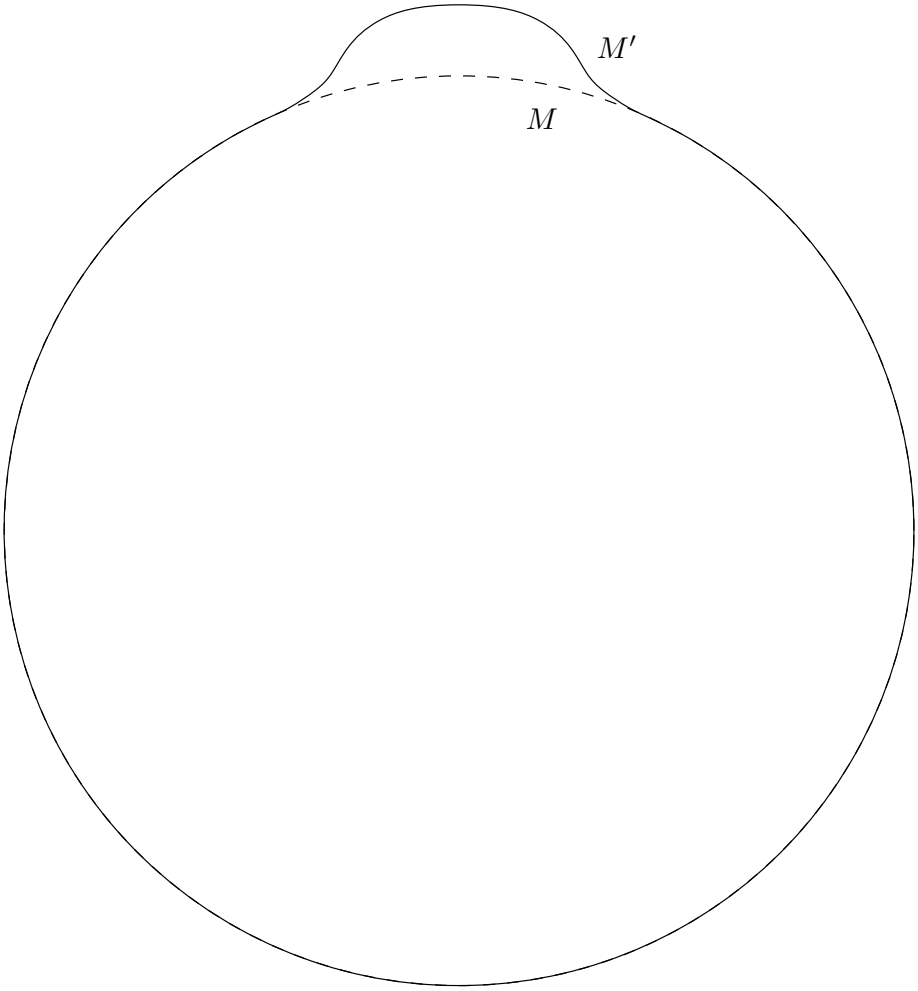


Figure 3.2: Hypotheses of Proposition 49.

us note that in the one-dimensional case $d = 1$, the rate of Proposition 50 matches the convergence rate of $\hat{\tau}(\mathcal{X}_n)$ (Theorem 45). Indeed, for curves, the alignment requirement is always fulfilled. Hence, the rate is exactly n^{-p} for $d = 1$, and $\hat{\tau}(\mathcal{X}_n)$ is minimax optimal.

Here, again, neither the convergence rate nor the constant depend on the ambient dimension m.

Chapter 4

Statistical Inference for Cluster Trees

This chapter presents the work in [Kim et al., 2016].

Clustering is a central problem in the analysis and exploration of data. It is a broad topic, with several existing distinct formulations, objectives, and methods. Despite the extensive literature on the topic, a common aspect of the clustering methodologies that has hindered its widespread scientific adoption is the dearth of methods for statistical inference in the context of clustering. Methods for inference broadly allow us to quantify our uncertainty, to discern "true" clusters from finite-sample artifacts, as well as to rigorously test hypotheses related to the estimated cluster structure.

In this chapter, we study statistical inference for the *cluster tree* of an unknown density. We assume that we observe an i.i.d. sample $\{X_1, \ldots, X_n\}$ from a distribution $\mathbb{P}_0$ with unknown density p_0. Here, $X_i \in \mathbb{X} \subset \mathbb{R}^m$. The connected components $\mathcal{C}(\lambda)$, of the upper level set $\{x : p_0(x) \geq \lambda\}$, are called *high-density clusters*. The set of high-density clusters forms a nested hierarchy which is referred to as the *cluster tree*[1] of p_0, which we denote as T_{p_0}.

Methods for density clustering fall broadly in the space of hierarchical clustering algorithms, and inherit several of their advantages: they allow for extremely general cluster shapes and sizes, and in general do not require the pre-specification of the number of clusters. Furthermore, unlike flat clustering methods, hierarchical methods are able to provide a multi-resolution summary of the underlying density. The cluster tree, irrespective of the dimensionality of the input random variable, is displayed as a two-dimensional object and this makes it an ideal tool to visualize data. In the context of statistical inference, density clustering has another important advantage over other clustering methods: the object of inference, the cluster tree of the unknown density p_0, is clearly specified.

In practice, the cluster tree is estimated from a finite sample, $\{X_1, \ldots, X_n\} \sim p_0$. In a scientific application, we are often most interested in reliably distinguishing topological features genuinely present in the cluster tree of the unknown p_0, from topological features that arise due to random fluctuations in the finite sample $\{X_1, \ldots, X_n\}$. In this chapter, we focus our inference on the cluster tree of the kernel density estimator, $T_{\hat{p}_h}$, where $\hat{p}_h$ is the kernel density estimator,

$$\hat{p}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\|x - X_i\|}{h}\right), \tag{4.1}$$

where K is a kernel and h is an appropriately chosen bandwidth [2].

To develop methods for statistical inference on cluster trees, we construct a confidence set for T_{p_0}, i.e. a collection of trees that will include T_{p_0} with some (pre-specified) probability. A confidence set can

[1] It is also referred to as the density tree or the level-set tree.

[2] We address computing the tree $T_{\hat{p}_h}$, and the choice of bandwidth in more detail in what follows.

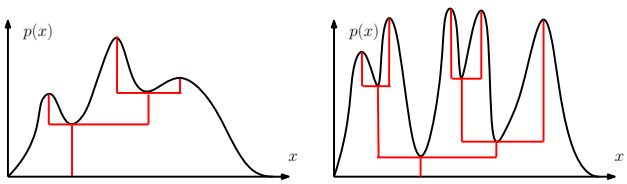


Figure 4.1: Examples of density trees. Black curves are the original density functions and the red trees are the associated density trees.

be converted to a hypothesis test, and a confidence set shows both statistical and scientific significances while a hypothesis test can only show statistical significances [Wasserman, 2010, p.155].

To construct and understand the confidence set, we need to solve a few technical and conceptual issues. The first issue is that we need a *metric* on trees, in order to quantify the collection of trees that are in some sense "close enough" to $T_{\hat{p}_h}$ to be statistically indistinguishable from it. We use the bootstrap to construct tight data-driven confidence sets. However, only some metrics are sufficiently "regular" to be amenable to bootstrap inference, which guides our choice of a suitable metric on trees.

On the basis of a finite sample, the true density is indistinguishable from a density with additional infinitesimal perturbations. This leads to the second technical issue which is that our confidence set invariably contains infinitely complex trees. Inspired by the idea of one-sided inference Donoho [1988], we propose a partial ordering on the set of all density trees to define simple trees. To find simple representative trees in the confidence set, we prune the empirical cluster tree by removing statistically insignificant features. These pruned trees are valid with statistical guarantees that are simpler than the empirical cluster tree in the proposed partial ordering.

4.1 Background and Definitions

We work with densities defined on a subset $\mathbb{X} \subset \mathbb{R}^m$, and denote by $\|.\|$ the Euclidean norm on $\mathbb{X}$. Throughout this chapter we restrict our attention to cluster tree estimators that are specified in terms of a function $f : \mathbb{X} \mapsto [0, \infty)$, i.e. we have the following definition:

Definition 51. For any $f : \mathbb{X} \mapsto [0, \infty)$ the *cluster tree* of f is a function $T_f : \mathbb{R} \mapsto 2^{\mathbb{X}}$, where $2^{\mathbb{X}}$ is the set of all subsets of $\mathbb{X}$, and $T_f(\lambda)$ is the set of the connected components of the upper-level set $\{x \in \mathbb{X} : f(x) \geq \lambda\}$. We define the collection of connected components $\{T_f\}$, as $\{T_f\} = \bigcup_{\lambda} T_f(\lambda)$.

As will be clearer in what follows, working only with cluster trees defined via a function f simplifies our search for metrics on trees, allowing us to use metrics specified in terms of the function f. With a slight abuse of notation, we will use T_f to denote also $\{T_f\}$, and write $C \in T_f$ to signify $C \in \{T_f\}$. The cluster tree T_f indeed has a tree structure, since for every pair $C_1, C_2 \in T_f$, either $C_1 \subset C_2$, $C_2 \subset C_1$, or $C_1 \cap C_2 = \emptyset$ holds. See Figure 4.1 for a graphical illustration of a cluster tree. The formal definition of the tree requires some topological theory; these details are in Appendix C.2.

In the context of hierarchical clustering, we are often interested in the "height" at which two points or two clusters merge in the clustering. We introduce the merge height from [Eldridge et al., 2015b, Definition 6]:

Definition 52. For any two points $x, y \in \mathbb{X}$, any $f : \mathbb{X} \mapsto [0, \infty)$, and its tree T_f, their **merge height** $m_f(x, y)$ is defined as the largest λ such that x and y are in the same density cluster at level λ, i.e.

$$m_f(x, y) = \sup \{\lambda \in \mathbb{R} : \text{ there exists } C \in T_f(\lambda) \text{ such that } x, y \in C\} .$$

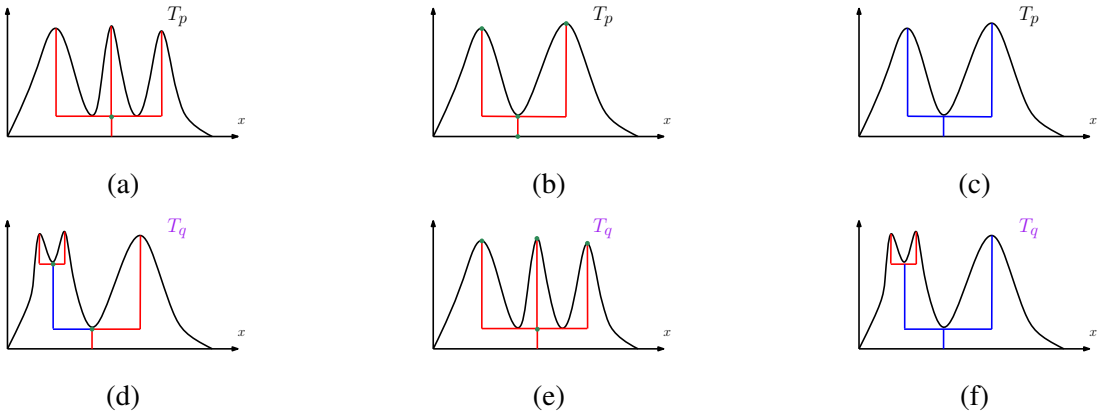


Figure 4.2: Three illustrations of the partial order $\preceq$ in Definition 54. In each case, in agreement with our intuitive notion of simplicity, the tree on the top (a, b, and c) is lower than the corresponding tree on the bottom(d, e, and f) in the partial order, i.e. for each example $T_p \preceq T_q$.

We refer to the function $m_f : \mathbb{X} \times \mathbb{X} \mapsto \mathbb{R}$ as the merge height function. For any two clusters $C_1, C_2 \in \{T_f\}$, their merge height $m_f(C_1, C_2)$ is defined analogously,

$$m_f(C_1, C_2) = \sup \{\lambda \in \mathbb{R} : \text{ there exists } C \in T_f(\lambda) \text{ such that } C_1, C_2 \subset C\}.$$

One of the contributions of this chapter is to construct valid confidence sets for the unknown true tree and to develop methods for visualizing the trees contained in this confidence set. Formally, we assume that we have samples $\{X_1, \ldots, X_n\}$ from a distribution $\mathbb{P}_0$ with density p_0.

Definition 53. An asymptotic $(1-\alpha)$ confidence set, C_α, is a collection of trees with the property that

$$\mathbb{P}_0(T_{p_0} \in C_\alpha) = 1 - \alpha + o(1).$$

We also provide non-asymptotic upper bounds on the $o(1)$ term in the above definition. Additionally, we provide methods to summarize the confidence set above. In order to summarize the confidence set, we define a partial ordering on trees.

Definition 54. For any $f, g : \mathbb{X} \mapsto [0, \infty)$ and their trees T_f, T_g, we say $T_f \preceq T_g$ if there exists a map $\Phi : \{T_f\} \to \{T_g\}$ such that for any $C_1, C_2 \in T_f$, we have $C_1 \subset C_2$ if and only if $\Phi(C_1) \subset \Phi(C_2)$.

With Definition 53 and 54, we describe the confidence set succinctly via some of the *simplest* trees in the confidence set in Section 4.3. Intuitively, these are trees without statistically insignificant splits.

It is easy to check that the partial order $\preceq$ in Definition 54 is reflexive (i.e. $T_f \preceq T_f$) and transitive (i.e. that $T_{f_1} \preceq T_{f_2}$ and $T_{f_2} \preceq T_{f_3}$ implies $T_{f_1} \preceq T_{f_3}$). However, to argue that $\preceq$ is a partial order, we need to show the antisymmetry, i.e. $T_f \preceq T_g$ and $T_g \preceq T_f$ implies that T_f and T_g are equivalent in some sense. In Appendices C.1 and C.2, we show an important result: for an appropriate topology on trees, $T_f \preceq T_g$ and $T_g \preceq T_f$ implies that T_f and T_f are *topologically equivalent*.

The partial order $\preceq$ in Definition 54 matches intuitive notions of the complexity of the tree for several reasons (see Figure 4.2). Firstly, $T_f \preceq T_g$ implies (number of edges of T_f) $\leq$ (number of edges of T_g) (compare Figure 4.2a and d, and see Lemma 103 in Appendix C.2). Secondly, if T_g is obtained from T_f by adding edges, then $T_f \preceq T_g$ (compare Figure 4.2b and e, and see Lemma 104 in Appendix C.2). Finally, the existence of a topology preserving embedding from $\{T_f\}$ to $\{T_g\}$ implies the relationship $T_f \preceq T_g$ (compare Figure 4.2c and f, and see Lemma 105 in Appendix C.2).

4.2 Tree Metrics

In this section, we introduce some natural metrics on cluster trees and study some of their properties that determine their suitability for statistical inference. We let $p, q : \mathbb{X} \to [0, \infty)$ be nonnegative functions and let T_p and T_q be the corresponding trees.

4.2.1 Metrics

We consider three metrics on cluster trees, the first is the standard ℓ_∞ metric, while the second and third are metrics that appear in the work of Eldridge et al. Eldridge et al. [2015b].

ℓ_∞ metric: The simplest metric is $d_\infty(T_p, T_q) = \|p - q\|_\infty = \sup_{x \in \mathbb{X}} |p(x) - q(x)|$. We will show in what follows that, in the context of statistical inference, this metric has several advantages over other metrics.

Merge distortion metric: The merge distortion metric intuitively measures the discrepancy in the merge height functions of two trees in Definition 52. We consider the *merge distortion metric* [Eldridge et al., 2015b, Definition 11] defined by

$$d_{\mathrm{M}}(T_p, T_q) = \sup_{x,y \in \mathbb{X}} |m_p(x, y) - m_q(x, y)|.$$

The merge distortion metric we consider is a special case of the metric introduced by Eldridge et al. [2015b][3]. The merge distortion metric was introduced by Eldridge et al. [2015b] to study the convergence of cluster tree estimators. They establish several interesting properties of the merge distortion metric: in particular, the metric is stable to perturbations in ℓ_∞, and further, that convergence in the merge distortion metric strengthens previous notions of convergence of the cluster trees.

Modified merge distortion metric: We also consider the *modified merge distortion metric* given by

$$d_{\mathrm{MM}}(T_p, T_q) = \sup_{x,y \in \mathbb{X}} |d_{T_p}(x, y) - d_{T_q}(x, y)|,$$

where $d_{T_p}(x, y) = p(x) + p(y) - 2m_p(x, y)$, which corresponds to the (pseudo)-distance between x and y *along* the tree. The metric d_{MM} is used in various proofs in the work of Eldridge et al. [2015b]. It is sensitive to both distortions of the merge heights in Definition 52, as well as of the underlying densities. Since the metric captures the distortion of distances between points along the tree, it is in some sense most closely aligned with the cluster tree. Finally, it is worth noting that unlike the interleaving distance and the functional distortion metric Bauer et al. [2015], Morozov et al. [2013], the three metrics we consider in this chapter are quite simple to approximate to a high-precision.

4.2.2 Properties of the Metrics

The following Lemma gives some basic relationships between the three metrics d_∞, d_{M} and d_{MM}. We define $p_{\inf} = \inf_{x \in \mathbb{X}} p(x)$, and $q_{\inf}$ analogously, and $a = \inf_{x \in \mathbb{X}} \{p(x) + q(x)\} - 2 \min\{p_{\inf}, q_{\inf}\}$. Note that when the Lebesgue measure $\mu(\mathbb{X})$ is infinite, then $p_{\inf} = q_{\inf} = a = 0$.

Lemma 55. *For any densities p and q, the following relationships hold: (i) When p and q are continuous, then $d_\infty(T_p, T_q) = d_{\mathrm{M}}(T_p, T_q)$. (ii) $d_{\mathrm{MM}}(T_p, T_q) \leq 4 d_\infty(T_p, T_q)$. (iii) $d_{\mathrm{MM}}(T_p, T_q) \geq d_\infty(T_p, T_q) - a$, where a is defined as above. Additionally when $\mu(\mathbb{X}) = \infty$, then $d_{\mathrm{MM}}(T_p, T_q) \geq d_\infty(T_p, T_q)$.*

[3] They further allow flexibility in taking a sup over a subset of $\mathbb{X}$.

The proof is in Appendix C.6. From Lemma 55, we can see that under a mild assumption (continuity of the densities), d_∞ and d_{M} are equivalent. We note again that the work of Eldridge et al. [2015b] actually defines a family of merge distortion metrics, while we restrict our attention to a canonical one. We can also see from Lemma 55 that while the modified merge metric is not equivalent to d_∞, it is usually multiplicatively sandwiched by d_∞.

Our next line of investigation is aimed at assessing the suitability of the three metrics for the task of statistical inference. Given the strong equivalence of d_∞ and d_{M} we focus our attention on d_∞ and d_{MM}. Based on prior work (see Chen et al. [2015], Chernozhukov et al. [2016]), the large sample behavior of d_∞ is well understood. In particular, $d_\infty(T_{\hat{p}_h}, T_{p_0})$ converges to the supremum of an appropriate Gaussian process, on the basis of which we can construct confidence intervals for the d_∞ metric.

The situation for the metric d_{MM} is substantially more subtle. One of our eventual goals is to use the non-parametric bootstrap to construct valid estimates of the confidence set. In general, a way to assess the amenability of a functional to the bootstrap is via *Hadamard differentiability* Wellner [2013]. Roughly speaking, Hadamard-differentiability is a type of *statistical stability*, that ensures that the functional under consideration is stable to perturbations in the input distribution. In Appendix C.3, we formally define Hadamard differentiability and prove that d_{MM} is *not* point-wise Hadamard differentiable. This does not completely rule out the possibility of finding a way to construct confidence sets based on d_{MM}, but doing so would be difficult and so far we know of no way to do it.

In summary, based on computational considerations we eliminate the interleaving distance and the functional distortion metric Bauer et al. [2015], Morozov et al. [2013], we eliminate the d_{MM} metric based on its unsuitability for statistical inference and focus the rest of this chapter on the d_∞ (or equivalently d_{M}) metric which is both computationally tractable and has well understood statistical behavior.

4.3 Confidence Sets

In this section, we consider the construction of valid confidence intervals centered around the kernel density estimator, defined in Equation (4.1). We first observe that a fixed bandwidth for the KDE gives a dimension-free rate of convergence for estimating a cluster tree. For estimating a density in high dimensions, the KDE has a poor rate of convergence, due to a decreasing bandwidth for simultaneously optimizing the bias and the variance of the KDE.

When estimating a cluster tree, the bias of the KDE does not affect its cluster tree. Intuitively, the cluster tree is a shape characteristic of a function, which is not affected by the bias. Defining the *biased* density, $p_h(x) = \mathbb{E}[\hat{p}_h(x)]$, two cluster trees from p_h and the true density p_0 are equivalent with respect to the topology in Appendix C.1, if h is small enough and p_0 is regular enough:

Lemma 56. *Suppose that the true unknown density p_0, has no non-degenerate critical points [4], then there exists a constant $h_0 > 0$ such that for all $0 < h \leq h_0$, the two cluster trees, T_{p_0} and T_{p_h} have the same topology in Appendix C.1.*

From Lemma 56, a fixed bandwidth for the KDE can be applied to give a dimension-free rate of convergence for estimating the cluster tree. Instead of decreasing bandwidth h and inferring the cluster tree of the true density T_{p_0} at rate $O_P(n^{-2/(4+d)})$, Lemma 56 implies that we can fix $h > 0$ and infer the cluster tree of the biased density T_{p_h} at rate $O_P(n^{-1/2})$ *independently of the dimension*. Hence a fixed bandwidth crucially enhances the convergence rate of the proposed methods in high-dimensional settings.

[4]The Hessian of p_0 at every critical point is non-degenerate. Such functions are known as Morse functions.

4.3.1 A data-driven confidence set

We recall that we base our inference on the d_∞ metric, and we recall the definition of a valid confidence set (see Definition 53). As a conceptual first step, suppose that for a specified value α we could compute the $1-\alpha$ quantile of the distribution of $d_\infty(T_{\hat{p}_h}, T_{p_h})$, and denote this value t_α. Then a valid confidence set for the unknown T_{p_h} is $C_\alpha = \{T : d_\infty(T, T_{\hat{p}_h}) \leq t_\alpha\}$. To estimate t_α, we use the bootstrap. Specifically, we generate B bootstrap samples, $\{\tilde{X}_1^1, \cdots, \tilde{X}_n^1\}, \ldots, \{\tilde{X}_1^B, \cdots, \tilde{X}_n^B\}$, by sampling with replacement from the original sample. On each bootstrap sample, we compute the KDE, and the associated cluster tree. We denote the cluster trees $\{\tilde{T}_{p_h}^1, \ldots, \tilde{T}_{p_h}^B\}$. Finally, we estimate t_α by

$$\hat{t}_\alpha = \hat{F}^{-1}(1-\alpha), \text{ where } \quad \hat{F}(s) = \frac{1}{B}\sum_{i=1}^{n} \mathbb{I}(d_\infty(\tilde{T}_{p_h}^i, T_{\hat{p}_h}) < s).$$

Then the data-driven confidence set is $\hat{C}_\alpha = \{T : d_\infty(T, \hat{T}_h) \leq \hat{t}_\alpha\}$. Using techniques from Chernozhukov et al. [2016], Chen et al. [2015], the following can be shown (proof omitted):

Theorem 57. *Under mild regularity conditions on the kernel*[5]*, we have that the constructed confidence set is asymptotically valid and satisfies,*

$$\mathbb{P}\left(T_h \in \hat{C}_\alpha\right) = 1 - \alpha + O\Big(\Big(\frac{\log^7 n}{nh^d}\Big)^{1/6}\Big).$$

Hence our data-driven confidence set is consistent at dimension independent rate. When h is a fixed small constant, Lemma 56 implies that T_{p_0} and T_{p_h} have the same topology, and Theorem 57 guarantees that the non-parametric bootstrap is consistent at a dimension independent $O(((\log n)^7/n)^{1/6})$ rate. For reasons explained in Chernozhukov et al. [2016], this rate is believed to be optimal.

4.3.2 Probing the Confidence Set

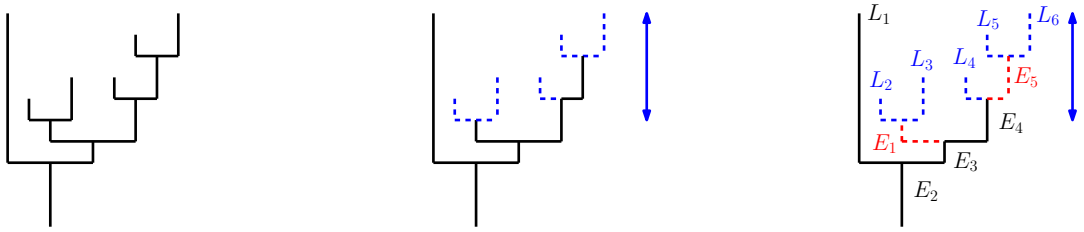
The confidence set $\hat{C}_\alpha$ is an infinite set with a complex structure. Infinitesimal perturbations of the density estimate are in our confidence set and so this set contains very complex trees. One way to understand the structure of the confidence set is to focus attention on simple trees in the confidence set. Intuitively, these trees only contain topological features (splits and branches) that are sufficiently strongly supported by the data.

We propose two *pruning* schemes to find trees, that are simpler than the empirical tree $T_{\hat{p}_h}$ that are in the confidence set. Pruning the empirical tree aids visualization as well as de-noises the empirical tree by eliminating some features that arise solely due to the stochastic variability of the finite-sample. The algorithms are (see Figure 4.3):

1. **Pruning only leaves:** Remove all leaves of length less than $2\hat{t}_\alpha$ (Figure 4.3b).
2. **Pruning leaves and internal branches:** In this case, we first prune the leaves as above. This yields a new tree. Now we again prune (using cumulative length) any leaf of length less than $2\hat{t}_\alpha$. We continue iteratively until all remaining leaves are of cumulative length larger than $2\hat{t}_\alpha$ (Figure 4.3c).

In Appendix C.4.2 we formally define the pruning operation and show the following. The remaining tree $\tilde{T}$ after either of the above pruning operations satisfies: (i) $\tilde{T} \preceq T_{\hat{p}_h}$, (ii) there exists a function f whose tree is $\tilde{T}$, and (iii) $\tilde{T} \in \hat{C}_\alpha$ (see Lemma 109 in Appendix C.4.2). In other words, we identified a valid tree with a statistical guarantee that is simpler than the original estimate $T_{\hat{p}_h}$. Intuitively, some of the statistically insignificant features have been removed from $T_{\hat{p}_h}$. We should point out, however, that

[5] See Appendix C.4.1 for details.



(a) The empirical tree. (b) Pruning only leaves. (c) Pruning leaves and branches.

Figure 4.3: Illustrations of our two pruning strategies. a shows the empirical tree. In b, leaves that are insignificant are pruned, while in c, insignificant internal branches are further pruned top-down.

there may exist other trees that are simpler than $T_{\hat{p}_h}$ that are in $\hat{C}_\alpha$. Ideally, we would like to have an algorithm that identifies all trees in the confidence set that are minimal with respect to the partial order $\preceq$ in Definition 54. This is an open question that we will address in future work.

4.4 Experiments

In this section, we demonstrate the techniques we have developed for inference on synthetic data, as well as on a real dataset.

4.4.1 Simulated data

We consider three simulations: the ring data (Figure 4.4a and d), the Mickey Mouse data (Figure 4.4b and e), and the yingyang data (Figure 4.4c and f). The smoothing bandwidth is chosen by the Silverman reference rule Silverman [1986] and we pick the significance level $\alpha = 0.05$.

Example 1: The ring data. (Figure 4.4a and d) The ring data consists of two structures: an outer ring and a center node. The outer circle consists of 1000 points and the central node contains 200 points. To construct the tree, we used $h = 0.202$.

Example 2: The Mickey Mouse data. (Figure 4.4b and e) The Mickey Mouse data has three components: the top left and right uniform circle (400 points each) and the center circle (1200 points). In this case, we select $h = 0.200$.

Example 3: The yingyang data. (Figure 4.4c and f) This data has 5 connected components: outer ring (2000 points), the two moon-shape regions (400 points each), and the two nodes (200 points each). We choose $h = 0.385$.

Figure 4.4 shows those data (a, b, and c) along with the pruned density trees (solid parts in d, e, and f). Before pruning the tree (both solid and dashed parts), there are more leaves than the actual number of connected components. But after pruning (only the solid parts), every leaf corresponds to an actual connected component. This demonstrates the power of a good pruning procedure.

4.4.2 GvHD dataset

Now we apply our method to the GvHD (Graft-versus-Host Disease) dataset Brinkman et al. [2007]. GvHD is a complication that may occur when transplanting bone marrow or stem cells from one subject to another Brinkman et al. [2007]. We obtained the GvHD dataset from R package 'mclust'. There are

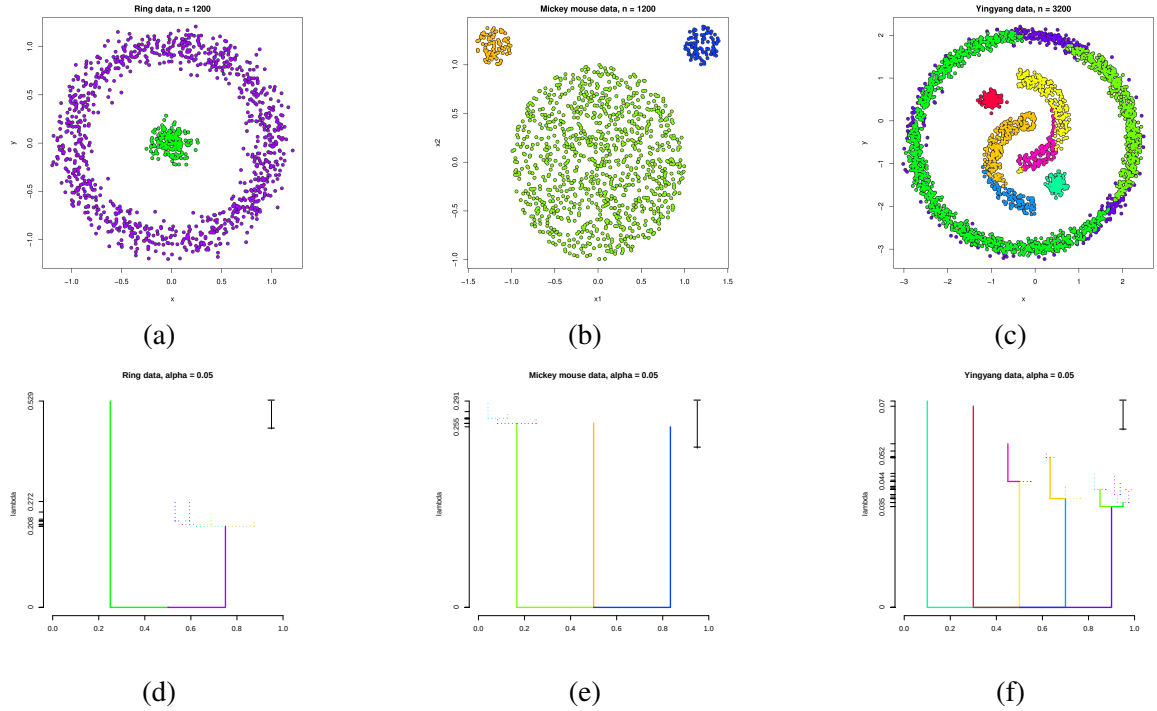
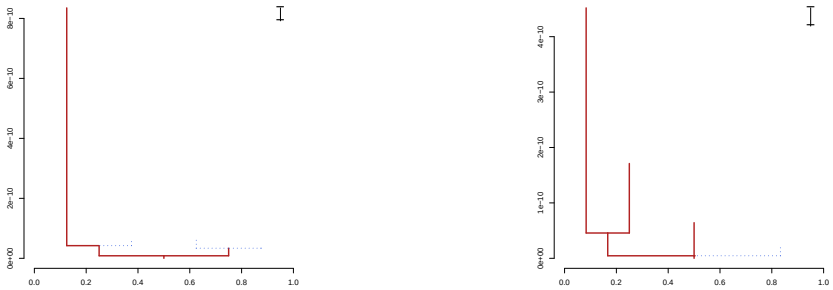


Figure 4.4: Simulation examples. a and d are the ring data; b and e are the mickey mouse data; c and f are the yingyang data. The solid lines are the pruned trees; the dashed lines are leaves (and edges) removed by the pruning procedure. A bar of length $2\hat{t}_\alpha$ is at the top right corner. The pruned trees recover the actual structure of connected components.



(a) The positive treatment data. (b) The control data.

Figure 4.5: The GvHD data. The solid brown lines are the remaining branches after pruning; the blue dashed lines are the pruned leaves (or edges). A bar of length $2\hat{t}_\alpha$ is at the top right corner.

two subsamples: the control sample and the positive (treatment) sample. The control sample consists of 9083 observations and the positive sample contains 6809 observations on 4 biomarker measurements ($d = 4$). By the normal reference rule Silverman [1986], we pick $h = 39.1$ for the positive sample and $h = 42.2$ for the control sample. We set the significance level $\alpha = 0.05$.

Figure 4.5 shows the density trees in both samples. The solid brown parts are the remaining components of density trees after pruning and the dashed blue parts are the branches removed by pruning. As can be seen, the pruned density tree of the positive sample (Figure 4.5a) is quite different from the pruned tree of the control sample (Figure 4.5b). The density function of the positive sample has fewer bumps (2 significant leaves) than the control sample (3 significant leaves). By comparing the pruned trees, we can see how the two distributions differ from each other.

Chapter 5

Persistent homology of KDE filtration on Rips complex

This chapter presents the work in Shin, Kim, Rinaldo, Wasserman, Persistent homology of KDE filtration on Rips complex.

When we observe data from a distribution P, the upper level sets $D_L := \{x \in \mathbb{R}^m : p(x) \geq L\}$ of the density function p reveal important topological features of the data generating distribution. For instance, density-based clustering methods [Hartigan, 1975, 1981, Cadre, 2006, Rinaldo and Wasserman, 2010] use the information about connected components of a level set to group data points in the hope that points in the same connected component share common characteristics. Rather than choosing a fixed level, a cluster tree [Chaudhuri and Dasgupta, 2010, Balakrishnan et al., 2013a, Eldridge et al., 2015a, Kim et al., 2016] summarizes the hierarchy of high-density clusters at all levels simultaneously.

We can investigate topological features of level sets by their corresponding homology groups. For example, the 0-th homology group of a level set contains information about connected components in the level set. By using higher order homology groups, we can further characterize each connected components. For instance, the rank of the 1st homology group of each connected component counts the number of one-dimensional holes.

Since different level sets could show different aspects of the data generating distribution, analyzing a fixed level set might be not enough to understand the overall shape of the distribution. Alternatively, as cluster trees show clusters at all levels, we can investigate changes in shapes by looking at all possible level sets simultaneously,

$$\{D_L\}_{L>0}. \tag{5.1}$$

Note that $D_{L_1} \subset D_{L_2}$ for any $L_1 \geq L_2$. Thus (5.1) is called the level sets filtration of the density function.

The persistent homology [Zomorodian and Carlsson, 2005, Edelsbrunner and Harer, 2008, 2010] quantifies topological features at multiple scales by analyzing a filtration of topological spaces. The persistent homology captures changes of homologies in filtrations simultaneously, see [Chung et al., 2009, Phillips et al., 2013, Fasy et al., 2014b, Bobrowski et al., 2014, Bubenik, 2015].

Since the density function is unknown, the persistent homology of the density function needs to be estimated. One approach, as in Fasy et al. [2014b], is to replace the level sets of the unknown density function by level sets of the kernel density estimator (KDE) computed on a grid of points. Another approach, as in Chazal et al. [2011b, 2013], Bobrowski et al. [2014], is to use level sets of the KDE computed on Rips complexes which can be viewed as an approximation of the union of balls centered at data points.

The goal of this chapter is to demonstrate the advantage and validity of the persistent homology of the KDE filtration on Rips complexes and show how to construct a bootstrap-based confidence set. The rest of this chapter is organized as follows: In Section 5.1, we discuss how to approximate a persistent homology of upper level set filtration of a general scalar function from noisy and finite number of observations by using Rips complex filtrations. In Section 5.2, we focus on how to use the persistent homology as a tool to extract the topological information of the data-generating distribution. After introducing a novel target quantity which can be viewed as a simplified but still useful version of the persistent homology of the upper level sets filtration of the density, we show consistency results for both the persistent homology of the upper level sets filtration of the density and the new target quantity we proposed. We also describe a novel methodology to construct an asymptotic confidence set based on the bootstrap procedure. In Section 5.3, we illustrate how we can use the proposed methods to do statistical inference on topological features of the underlying distribution by using toy examples. We also conduct numerical experiments to demonstrate the computational efficiency of the proposed method in Section 5.4. For the sake of readability, all proofs and technical details are postponed to Appendix D.

5.1 Persistent homology of Rips complex filtration and Stability

In this section, we discuss how to approximate a persistent homology of upper level set filtration of a scalar function from noisy and finite number of observations by using Rips complex filtrations. All the proofs for this section are in Section D.3.

Formally, let $f : \mathbb{X} \subset \mathbb{R}^m \to (0, \infty)$ be a scalar function of interest. The upper level set filtration of f on $\mathbb{X}$ is defined by $\{D_L\}_{L>0}$ where

$$D_L := \{x \in \mathbb{X} : f(x) \geq L\}, \ \ \forall L > 0. \tag{5.2}$$

Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be an i.i.d. sample from a sampling distribution P on $\mathbb{X}$. Let $\hat{f}$ be a fixed functional estimator of f. One natural way to approximate D_L is to use an union of closed balls around the sample points with higher function values. In detail, for any $L \in \mathbb{R}$ and $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, the upper level set estimator is defined by

$$\hat{D}_L(r) := \bigcup_{\{X_i : \hat{f}(X_i) \geq L\}} \mathbb{B}_{\mathbb{X}}(X_i, r_i), \tag{5.3}$$

where

$$\mathbb{B}_{\mathbb{X}}(x, r) := \{y \in \mathbb{X} : d(x, y) < r\}, \quad r > 0.$$

Let $\mathrm{PH}_*^{\mathbb{X}}(f)$ and $\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r)$ be persistent homologies of filtrations $\{D_L\}_{L>0}$ in (5.2) and $\{\hat{D}_L\}_{L>0}$ in (5.3), respectively. The following lemma shows how to bound the bottleneck distance between $\mathrm{PH}_*^{\mathbb{X}}(f)$ and $\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r)$ by controlling the estimation error (the difference between f and $\hat{f}$), and the geometrical approximation error (the difference between upper level set and the union of balls around high function value samples).

Lemma 58. *Suppose either f or $\hat{f}$ is M-Lipschitz continuous. For any given $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, suppose the samples form an r-covering of $\mathbb{X}$, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{5.4}$$

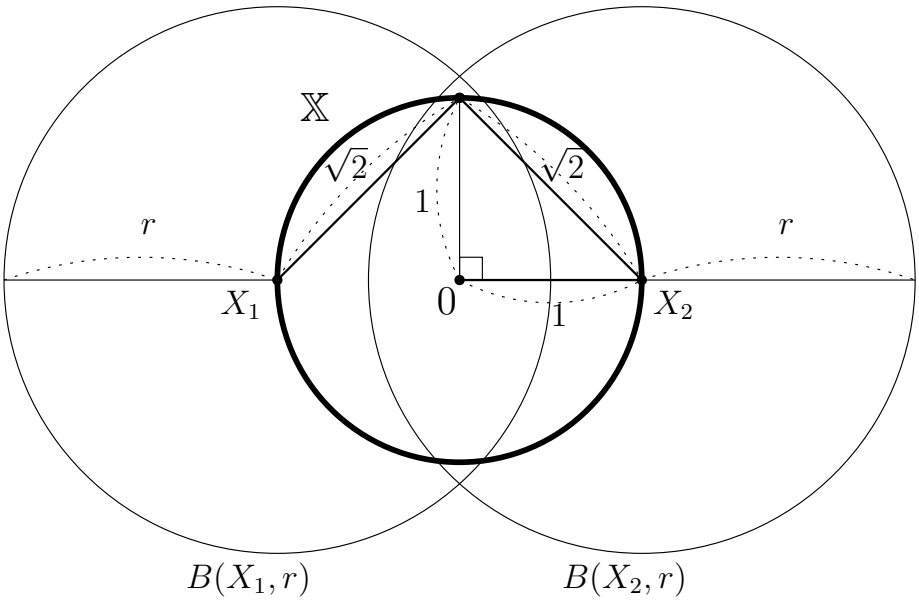


Figure 5.1: *An example in which* $\mathbb{B}_{\mathbb{X}}(X_1, r) \bigcup \mathbb{B}_{\mathbb{X}}(X_2, r)$ *is not homotopic equivalent to* $\check{C}ech_{\mathbb{X}}(\mathcal{X}_n, r)$ *where* $\mathbb{X} = \left\{x \in \mathbb{R}^2 : \|x\|_2 = 1\right\}$, $X_1 = (-1, 0)$, $X_2 = (0, 1)$ *and* $r > \sqrt{2}$.

Then the bottleneck distance between $\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r)$ *and* $\mathrm{PH}_*^{\mathbb{X}}(f)$ *is upper bounded as*

$$d_B\left(\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r), \mathrm{PH}_*^{\mathbb{X}}(f)\right) \leq \|\hat{f} - f\|_\infty + M\|r\|_\infty. \tag{5.5}$$

The persistent homology $\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r)$ is an oracle estimator, as it requires knowledge of $\mathbb{X}$. However, if the maximum radii of balls are smaller than the reach of $\mathbb{X}$ in (1.5), we can produce a computable estimator based on the Čech complexes over sample points. Precisely, let assume $\mathbb{X}$ has positive reach $\tau > 0$. The positive reach assumption is crucial in many parts of our analysis and cannot be dispensed of. In particular, one of the key implications is the fact that the homology of the union of balls (1.9) built on a sample $\mathcal{X}_n$ from P can be recovered using the corresponding Čech complex $\check{C}ech_{\mathbb{X}}(\mathcal{X}_n, r)$ in (1.8), provided the radii of the balls are all smaller than $\sqrt{2}$ times the reach.

Proposition 59. *Let* $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{X}$. *Suppose* $\mathbb{X}$ *has a positive reach* $\tau > 0$ *Then, for any* $r = (r_1, \ldots, r_n) \in (0, \sqrt{2}\tau]^n$, *the union of balls* $\bigcup_{i=1}^n \mathbb{B}_{\mathbb{X}}(X_i, r_i)$ *is homotopic equivalent to the Čech complex* $\check{C}ech_{\mathbb{X}}(\mathcal{X}_n, r)$.

The previous result provides the theoretical underpinning for the methodology developed in this chapter. Its proof is a direct consequence of the Lemma 6 (Nerve Theorem) and of Proposition 119 in Appendix D.2, a simple geometric result that appears to be new and may be of independent interest.

The following example shows that the reach condition $\|r\|_\infty \leq \sqrt{2}\tau$ is tight in the sense that there exists cases where Proposition 59 does not hold when $\|r\|_\infty > \sqrt{2}\tau$.

Example 60. Let $\mathbb{X}$ be a unit Euclidean sphere. Let X_1, X_2 be an antipodal pair of points on $\mathbb{X}$. For a unit Euclidean sphere, the reach is equal to its radius 1. Therefore, if $r = (r_1, r_2) \in (0, \sqrt{2}]^2$, $\mathbb{B}_{\mathbb{X}}(X_1, r_1) \bigcup \mathbb{B}_{\mathbb{X}}(X_2, r_2)$ is homotopic equivalent to $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathcal{X}_n, r)$ by Proposition 59. However, if $r_1, r_2 > \sqrt{2}$, $\mathbb{B}_{\mathbb{X}}(X_1, r_1) \bigcup \mathbb{B}_{\mathbb{X}}(X_2, r_2) \simeq \mathbb{X}$ but $\check{\mathrm{C}}\mathrm{ech}_{\mathbb{X}}(\mathcal{X}_n, r) \simeq 0$. Figure 5.1 illustrate the 2-dimensional case.

Even if $\check{C}ech_{\mathbb{X}}(\mathcal{X}_n, r)$ is more easily computable than $\bigcup_{i=1}^n \mathbb{B}_{\mathbb{X}}(X_i, r_i)$, it still requires knowledge of $\mathbb{X}$ to compute. Instead, we introduce a computable persistent homology estimator based on

$\check{C}ech_{\mathbb{R}^m}(\mathcal{X}_n, r)$, where the intersections of the balls in (1.8) are computed on $\mathbb{R}^m$ instead of the unknown space $\mathbb{X}$.

Definition 61. Let $\mathrm{PH}_*^{\check{\mathrm{C}}}(\hat{p}_h, r)$ be the persistent homology of the filtrations of Čech complexes in (1.8) as

$$\left\{\check{\mathrm{C}}\mathrm{ech}_{\mathbb{R}^m}\left(\mathcal{X}_{n,L}^{\hat{f}}\right)\right\}_{L>0},$$

where

$$\mathcal{X}_{n,L}^{\hat{f}} := \left\{X_i \in \mathcal{X}_n : \hat{f}(X_i) \geq L\right\}.$$

In general, $\check{C}ech_{\mathbb{R}^m}(\mathcal{X}_n, r)$ is not homotopic equivalent to $\check{C}ech_{\mathbb{X}}(\mathcal{X}_n, r)$. However its persistent homology is close to the one built up on $\check{C}ech_{\mathbb{X}}(\mathcal{X}_n, r)$ in terms of the bottleneck distance. Based on this fact, bounds on the bottleneck distance between $\mathrm{PH}_*^{Cech_{\mathbb{R}^m}}(\hat{p}_h, r)$ and the target persistent homology $\mathrm{PH}_*^{\mathbb{X}}(f)$ are derived in the following theorem.

Theorem 62. *Let τ be the reach of $\mathbb{X}$. Suppose either f or $\hat{f}$ is M-Lipschitz continuous. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \tau/\sqrt{2}]^n$, suppose the samples form an r-covering of $\mathbb{X}$, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{5.6}$$

Then the bottleneck distance between $\mathrm{PH}_*^{\check{\mathrm{C}}}(\hat{f}, r)$ and $\mathrm{PH}_*^{\mathbb{X}}(f)$ is upper bounded as

$$d_B\left(\mathrm{PH}_*^{\check{\mathrm{C}}}(\hat{f}, r), \mathrm{PH}_*^{\mathbb{X}}(f)\right) \leq \|\hat{f} - f\|_\infty + 2M\|r\|_\infty \tag{5.7}$$

$\mathrm{PH}_*^{\check{\mathrm{C}}}(\hat{f}, r)$ in Definition 61 is a computable estimator of $\mathrm{PH}_*^{\mathbb{X}}(f)$, since it does not require any knowledge of $\mathbb{X}$ (other than an upper bound on the reach). However, it is computationally expensive, as building the Čech complex rapidly becomes unfeasible when the sample size n (and the dimension d) gets large. Instead, we consider an analogous estimator based on Rips complexes, which can be more easily computed as it only needs as input the set of all pairwise Euclidean distances among the sample points. This is the main estimator of this chapter.

Definition 63. Let $\mathrm{PH}_*^R\left(\hat{f}, r\right)$ be the persistent homology of the filtrations of Rips complexes in (1.10) as

$$\left\{R\left(\mathcal{X}_{n,L}^{\hat{f}}, r\right)\right\}_{L>0}. \tag{5.8}$$

The next result shows that, not surprisingly, the performance of $\mathrm{PH}_*^R\left(\hat{f}, r\right)$ is at most worse than the performance of the computationally prohibitive estimator $\mathrm{PH}_*^{\check{\mathrm{C}}}(\hat{f}, r)$ only by a constant factor.

Theorem 64. *Let τ be the reach of $\mathbb{X}$. Suppose either f or $\hat{f}$ is M-Lipschitz continuous. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \tau/\sqrt{2}]^n$, suppose the samples form an r-covering of $\mathbb{X}$, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{5.9}$$

Then the bottleneck distance between $\mathrm{PH}_*^R(\hat{f}, r)$ and $\mathrm{PH}_*^{\mathbb{X}}(f)$ is upper bounded as

$$d_B\left(\mathrm{PH}_*^R(\hat{f}, r), \mathrm{PH}_*^{\mathbb{X}}(f)\right) \leq \|\hat{f} - f\|_\infty + 2M\|r\|_\infty. \tag{5.10}$$

Remark 65. If $r_i = r \ \ \forall i \in [n]$ and $\mathbb{X}$ is a Euclidean space then Theorem 64 holds under the weaker condition $r \leq \tau$ instead of $\sqrt{2}\|r\|_\infty \leq \tau$, and the terms in the bounds $2M\|r\|_\infty$ can be replaced with $\sqrt{2}Mr$.

5.2 Consistency and Confidence sets for Persistent homology of Density filtration

In this section, we discuss how to use the persistent homology as a tool to extract the topological information of a probability distribution P. After defining the target persistent homology, we propose two computable estimators based on a finite number of observations from P in the same way we did in the previous section. With a high probability, both estimators are close to the target persistent homology in terms of the bottleneck distance. Finally, we discuss how to construct bootstrap based asymptotic confidence sets which can be used to identify significant topological features of the distribution P. All the proofs for this section are in Section D.4.

5.2.1 Target Persistent Homology and Assumptions

Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be i.i.d. observations from a probability distribution P on $\mathbb{R}^m$ whose support $\mathrm{supp}(P)$ plays the role of the set $\mathbb{X}$ in the previous section.

We will impose the following assumptions on P:

Assumption 66. *The probability measure P is such that:*

1. $\mathrm{supp}(P)$ *is bounded and has positive reach $\tau_P > 0$, and*
2. *there exist positive constants $\nu_{\max}$, $a_{\min}$ and ϵ_0 such that, for all $x \in \mathrm{supp}(P)$,*

$$P(\mathbb{B}_{\mathbb{R}^m}(x, \epsilon)) \geq a_{\min}\epsilon^{\nu_{\max}}, \quad \forall \epsilon \in (0, \epsilon_0).$$

The above assumptions on P are fairly standard. In particular, the last condition is also known as the (a, b)-condition or the standard condition [Cuevas and Rodríguez-Casal, 2004, Cuevas, 2009, Chazal et al., 2014a]. It is satisfied, for example, if $\mathrm{supp}(P)$ is a smooth manifold of dimension $\nu_{\max}$ and P has a density with respect to the Hausdorff measure on it bounded from below by $a_{\min}$.

In order to extract topological information of the distribution P, we rely on the kernel density estimator (KDE), which smooths out the empirical measure by an appropriate kernel function K satisfying the following, standard, assumptions.

Assumption 67. *The kernel function $K : \mathbb{R}^m \to \mathbb{R}$ is a nonnegative function with the following conditions:*

1. $\int K(x)\mathrm{d}x = 1$.
2. $\int \|x\|K(x)\mathrm{d}x < \infty$ *and* $\sup_{x\in\mathbb{R}^m} K(x) < \infty$.
3. *K is Lipschitz continuous with the constant $M_K > 0$.*

For a fixed value $h > 0$ of the bandwidth parameter, the corresponding kernel density estimator is defined as

$$\hat{p}_h(x) := \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{5.11}$$

Let $p_h : \mathbb{R}^m \to \mathbb{R}$ be the pointwise average of the kernel density estimator, i.e. $p_h(x) := \mathbb{E}[\hat{p}_h(x)]$, for all $x \in \mathbb{R}^m$. It is easy to see that p_h is a density function (with respect to the Lebesgue measure). Throughout this chapter, we assume p_h is tame for any $h > 0$.

When the underlying distribution P admits a density p, the persistent homology $\mathrm{PH}_*(p)$ of the upper level set filtration $\{x \in \mathbb{R}^m : p(x) \geq L\}_{L>0}$ of p is a natural target quantity for understanding the topology of P. However, as discussed in Fasy et al. [2014b], the persistent homology of the upper level sets filtration of p_h, with fixed h, would also serve a similar purpose while offering several advantages. This is because:

1. the density p_h and the persitent homology of its upper level set flitration is always well-defined even if the Lebesgue density p does not exist;
2. the function p_h can be viewed as a topologically simplified version of p. The level sets of p_h may miss tiny topological features in p but can still capture significant ones.
3. The kernel density estimator $\hat{p}_h$ is a point-wise unbiased estimator of p_h and concentrates around it exponentially fast in the sup-norm (again h is fixed) at a parametric rate: see ?? below. In contrast, $\hat{p}_h$ is a biased estimator of p, and the bias can only be removed by letting $h \to 0$, in which case $\hat{p}_h$ converges to p at rates that depend on the dimension. Hence inference for p_h is more precise.

However, a potential complication arises when we target the persistent homology of the smoothed density p_h instead of the underlying density p (assuming it exists). Indeed, p_h remain positive even outside the support of P. As a result, the persistent homology of p_h may exhibit topological properties in regions that are of no interest. This issue can be avoided by considering only the persistent homology of the upper level set filtration of p_h restricted to $\mathrm{supp}(P)$ rather than the larger set $\mathrm{supp}(p_h)$.

Formally, for each $L \geq 0$, let

$$D_L := \{x \in \mathrm{supp}(P) : \ p_h(x) \geq L\}, \tag{5.12}$$

denote the corresponding upper level set of p_h intersected with $\mathrm{supp}(P)$. Let $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ be the persistent homology of the corresponding level sets filtration $\{D_L\}_{L>0}$. The usual persistent homology of the upper level sets filtration of p_h will be denoted by $\mathrm{PH}_*^{\mathbb{R}^m}(p_h)$ or, more conveniently, $\mathrm{PH}_*(p_h)$.

We first describe how the newly defined persistent homology $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ relates to the persistent homologies $\mathrm{PH}_*(p_h)$ and $\mathrm{PH}_*(p)$.

Proposition 68. *Let P be a probability measure on $\mathbb{R}^m$ and K be a kernel function satisfying Assumption 66 and 67. Let p be the Lebesgue density of P, and assume p is Lipschitz continuous. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, the following hold :*

(a) $d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*^{\mathbb{R}^m}(p_h)\right) \leq \sup_{x \notin \mathrm{supp}(P)} |p_h(x)| \leq C_K M_P h$,

(b) $d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*(p)\right) \leq \sup_{x \in \mathrm{supp}(P)} |p_h(x) - p(x)| \leq C_K M_P h$,

where $C_K = \int \|x\| K(x) \mathrm{d}x$ and $M_P > 0$ is the Lipschitz constant of p.

The following simple examples demonstrates that there exists a density p and a kernel K such that

$$d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*(p)\right) = 0 \quad \text{and} \quad d_B\left(\mathrm{PH}_*^{\mathbb{R}^m}(p_h), \mathrm{PH}_*(p)\right) > 0.$$

Thus, in this particular instance, the persistence homology $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ more accurately approximates the persistent homology $\mathrm{PH}_*(p)$.

Example 69. Let P be a mixture of uniform distributions in $\mathbb{R}$ with the density function

$$p(x) = \frac{1}{4}\mathbf{1}\left(|x| \in [\frac{1}{2}, \frac{5}{2}]\right).$$

If we use the triangular kernel, $K(x) = (1 - |x|)\,\mathbf{1}\,(|x| \leq 1)$, the pointwise average of the kernel density estimator, $p_h(x)$, become a combination of quadratic functions. Figure 5.2 illustrates the densities p and p_h for $h = 1$. In this case, the persistent homologies $\mathrm{PH}_*(p)$ and $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ both consist of two 0-th order homology classes that are born at $\frac{1}{4}$ and die at 0. On the other hand, the $\mathrm{PH}_*^{\mathbb{R}^m}(p_h)$ consists of two 0-th order homology classes that are born at $\frac{1}{4}$ and die at $\frac{1}{16}$. Therefore,

$$d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*(p)\right) = 0 \quad \text{but} \quad d_B\left(\mathrm{PH}_*^{\mathbb{R}^m}(p_h), \mathrm{PH}_*(p)\right) = \frac{1}{16} > 0$$

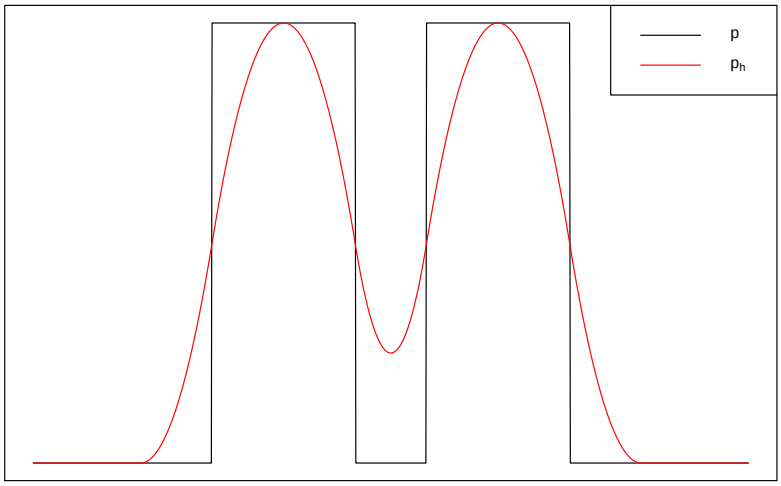


Figure 5.2: *The density function of a mixture of uniform distributions, and the pointwise average of the kernel density estimator with the triangular kernel ($h = 1$)*

5.2.2 Consistency and Confidence sets for Persistent homology of Density filtration

In Theorem 64 in Section 5.1, it was shown that for any function f, the persistent homology of upper level set filtration $\mathrm{PH}_*^{\mathbb{X}}(f)$ can be approximated by the persistent homology of Rips complexes built upon finite number of observations $\mathrm{PH}_*^R(\hat{f}, r)$. As a special case for the smoothed density function p_h, we define an estimator using the KDE filtration on Rips complexes $\mathrm{PH}_*^R(\hat{p}_h, r)$ for the persistent homology of upper level set filtration of the smoothed density function $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ as following :

Definition 70. The persistent homology of KDE filtrations on Rips complexes, $\mathrm{PH}_*^R(\hat{p}_h, r)$ is defined as the persistent homology of the filtration of Rips complexes in (1.10) as

$$\left\{ R\left(\mathcal{X}_{n,L}^{\hat{p}_h}, r\right)\right\}_{L>0}, \tag{5.13}$$

where

$$\mathcal{X}_{n,L}^{\hat{p}_h} := \{X_i \in \mathcal{X}_n : \hat{p}_h(X_i) \geq L\}.$$

Recall that, under the proper conditions described in Theorem 64, the bottleneck distance between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ and its estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ is upper bounded by $\|\hat{p}_h - p_h\|_\infty + 2M\|r\|_\infty$ where M is the Lipschitz constant of either $\hat{p}_h$ or p_h. Since we use M_K-Lipschitz continuous kernel, both $\hat{p}_h$ and p_h are $\frac{M_K}{h^{d+1}}$-Lipschitz continuous for any fixed $h > 0$. If the underlying distribution P is more "smooth", p_h can have better Lipschitz constant depending on P. For example, if P has M_P-Lipschitz continuous Lebesgue density p, p_h is also M_P-Lipschitz continuous regardless of the choice of the bandwidth h and the kernel function K satisfying Assumption

67. However, the assumption of Lipschitz continuous Lebesgue density could be too restrictive for many TDA applications. Instead, we introduce a weaker smoothness condition on P which would be more suitable for TDA purposes, and investigate the statistical performance of our estimator under both conditions.

Assumption 71. *The probability measure P satisfies the following: there exists $\nu_{\min}, a_{\max} > 0$ so that for all $r > 0$ and for all $x \in \mathbb{R}^m$, $P(\mathbb{B}_{\mathbb{R}^m}(x, r)) \leq a_{\max} r^{\nu_{\min}}$. Also, the support of the kernel function K is bounded by a unit ball centered around 0, i.e., supp$(K) \subset \mathbb{B}_{\mathbb{R}^m}(0, 1)$.*

Assumption 72. *The probability measure P has a density $p : \mathbb{R}^m \to \mathbb{R}$ with respect to the Lebesgue measure that is M_P-Lipschitz, for some $M_P > 0$*

If $\mathrm{supp}(P)$ is a well-behaved sets, such as a smooth manifold of dimension $\nu_{\min}$ (possibly smaller than d) and P has a bounded density with respect to the restriction of the Haudorff measure of dimension $\nu_{\min}$ on it, then Assumption 71 is satisfied, with $a_{\max}$ depending on the maximal value of the density.

The following proposition is a direct application of Theorem 64.

Proposition 73. *Let P be a probability measure on $\mathbb{R}^m$ and K be a kernel function satisfying Assumption 66 and 67. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \infty)^n$ with $\sqrt{2}\|r\|_\infty \leq \tau$, suppose the samples form an r-covering of the support of P, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i).$$

Then the bottleneck distance between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ and its estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ is upper bounded as, under Assumption 71,

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq \|\hat{p}_h - p_h\|_\infty + \frac{2a_{\max} M_K \|r\|_\infty}{h^{d+1-\nu_{\min}}}, \tag{5.14}$$

while, under Assumption 72,

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq \|\hat{p}_h - p_h\|_\infty + 2M_P \|r\|_\infty. \tag{5.15}$$

Proposition 73 shows that the bottleneck distance between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ and its estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ can be upper bounded by the statistical estimation error term, $\|\hat{p}_h - p_h\|_\infty$, and the geometrical error terms depending on smoothness assumptions on the underlying distribution P. Based on it, the following theorem shows that the proposed estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ is consistent for the persistent homology of the smoothed density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ with properly chosen sequences of r_n and h_n.

Theorem 74. *Suppose Assumption 66 and 67 holds. Let $\{r_n = (r_{n,1}, \ldots, r_{n,n})\}_{n \in \mathbb{N}}$ be a triangular array of positive numbers such that*

$$\min_i r_{n,i} \geq C_P \left(\frac{\log n}{n}\right)^{1/\nu_{\max}},$$

with a constant C_P depending only on $a_{\min}$. Let also assume $\sqrt{2}\|r_n\|_\infty \leq \tau$ for all sufficiently large n. Then, under Assumption 71, for a fixed $h > 0$, there exists a positive constant $C_{K,P}$ depends only on $\|K\|_\infty$, $\|K\|_2$, $\nu_{\min}$, $\nu_{\max}$, $a_{\min}$, $a_{\max}$ such that with probability at least $1 - \delta$, the bottleneck distance

between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ *and its estimator* $\mathrm{PH}_*^R(\hat{p}_h, r_n)$ *is upper bounded as*

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq C_{K,P}\left(\sqrt{\frac{\log(1/\delta)}{n}} + \|r_n\|_\infty\right), \tag{5.16}$$

for $\forall n$ *with* $\sqrt{2}\|r_n\|_\infty \leq \tau$.

Under Assumption 72, suppose $h_n \leq h_0$ *for some fixed* $h_0 \in (0,1)$ *for sufficiently large* n *and* $h_n^{-d}\log(1/h_n) \leq C_{h_0} n$ *for some constant* C_{h_0}. *Then there exists a positive constant* C_{K,P,h_0} *depends only on* $\|K\|_\infty$, $\|K\|_2$, d, $a_{\min}$, $\|p\|_\infty$, h_0 *such that with probability at least* $1-\delta$, *the bottleneck distance between the persistent homology of the density filtration* $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})$ *and its estimator* $\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n)$ *is upper bounded as*

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})\right) \leq C_{K,P,h_0}\left(\sqrt{\frac{\log(1/\delta)}{nh_n^d}} + \sqrt{\frac{\log(1/h_n)}{nh_n^d}} + \|r_n\|_\infty\right). \tag{5.17}$$

for $\forall n$ *with* $\sqrt{2}\|r_n\|_\infty \leq \tau$.

Furthermore, combining Proposition 68 (b) and Theorem 74 shows that the proposed estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ is consistent for the persistent homology of the true density filtration $\mathrm{PH}_*(p)$ as well with properly chosen sequences of r_n and h_n, as in Corollary 75.

Corollary 75. *Suppose Assumption 66, 67 and 72 holds. Let* $\{r_n = (r_{n,1}, \ldots, r_{n,n})\}_{n\in\mathbb{N}}$ *be a triangular array of positive numbers such that*

$$\min_i r_{n,i} \geq C_P\left(\frac{\log n}{n}\right)^{1/\nu_{\max}}$$

with a constant C_P *depending only on* $a_{\min}$. *Then, if* $\|r_n\|_\infty = o(1)$ *and* $\frac{\log(1/h_n)}{nh_n^d} = O(1)$, *then the bottleneck distance between the persistent homology of the true density filtration* $\mathrm{PH}_*(p)$ *and the proposed estimator* $\mathrm{PH}_*^R(\hat{p}_h, r_n)$ *is upper bounded as*

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*(p)\right) = O_P\left(\sqrt{\frac{\log(1/h_n)}{nh_n^d}} + \|r_n\|_\infty + h_n\right). \tag{5.18}$$

Remark 76. By using the same argument, we can show the consistency of the Čech complex based estimator $\mathrm{PH}_*^{Cech_{\mathbb{R}^m}}(\hat{p}_h, r)$ under the same assumptions.

Although Theorem 74 and Corollary 75 show the estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ and target quantities $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ and $\mathrm{PH}_*(p)$ are close to each other with high probability, the upper bounds for the bottleneck distances depend on unknown quantities of the underlying probability measure P. In the remaining part of this section, we build a computable confidence set for the persistent homology $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ of the level sets filtration of the smoothed density p_h on the support $\mathrm{supp}(P)$.

A confidence set of the persistent homology $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ is a random set of persistent homologies that contains $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ with some probability. Specifically, for given $\alpha \in (0,1)$, a valid $1-\alpha$ level asymptotic confidence set of $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ is a random set $\hat{C}_\alpha$ satisfying

$$\liminf_{n\to\infty} \mathbb{P}(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) \in \hat{C}_\alpha) \geq 1-\alpha.$$

We construct the confidence set $\hat{C}_\alpha$ by considering all persistent homologies within c_n bottleneck distance from the computable estimator $\mathrm{PH}_*^{Cech_{\mathbb{R}^m}}(\hat{p}_h, r)$ or $\mathrm{PH}_*^R(\hat{p}_h, r)$ for some $c_n > 0$. Let $\hat{\mathrm{PH}}_*(p_h)$ be one of the estimators. Then, the confidence set has the following form,

$$\hat{C}_\alpha = \left\{ \mathcal{P} : \; d_B \left(\mathcal{P}, \hat{\mathrm{PH}}_*(p_h) \right) \leq c_n \right\},$$

where both $\hat{\mathrm{PH}}_*(p_h)$ and radius c_n are functions of $X_1, \ldots, X_n$. Note that $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) \in \hat{C}_\alpha$ holds if and only if

$$d_B \left(\hat{\mathrm{PH}}_*(p_h), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) \right) \leq c_n.$$

Therefore $\hat{C}_\alpha$ is a valid $1-\alpha$ asymptotic confidence set if and only if

$$\liminf_{n\to\infty} \mathbb{P} \left(d_B \left(\hat{\mathrm{PH}}_*(p_h), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) \right) \leq c_n \right) \geq 1 - \alpha.$$

Proposition 73 cannot be directly used to build a confidence set because the covering condition is not checkable and bound terms are not computable without the knowledge of the data-generating distribution P, which is typically unavailable. Instead, we can split the filtration in two parts : $(0, \epsilon] \cup (\epsilon, \infty)$ for some $\epsilon > 0$ satisfying

$$\{x : \hat{p}_h(x) \geq \epsilon\} \subset \bigcup_i \mathbb{B}_{\mathbb{R}^m}(X_i, r_i). \tag{5.19}$$

Roughly speaking, when filtration values are restricted to $(0, \epsilon]$, the bottleneck distance between $\mathrm{PH}_*^{\mathrm{supp}(P)}(\hat{p}_h, r)$ and $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ is upper bounded by ϵ. For filtration values in (ϵ, ∞), due to the covering condition (5.19), the bottleneck distance can be bounded by the maximal possible difference between the value of $\hat{p}_h$ at a sample point X_i and its value at any points within an r_i-neighbor of X_i, for $\forall i = 1 \ldots, n$, which is given by

$$\max_i \sup_{\|x - X_i\| \leq r_i} |\hat{p}_h(x) - \hat{p}_h(X_i)|.$$

The following result formally shows how to combine these quantities to bound the distance between $\mathrm{PH}_*^{\mathrm{supp}(P)}(\hat{p}_h, r)$ and $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$.

Lemma 77. *Let P be a probability measure on $\mathbb{R}^m$ and K be a kernel function satisfying Assumption 66 and 67. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, set*

$$\mathcal{E}_r = \left\{ \epsilon \in \mathbb{R}_+ : \{x : \hat{p}_h(x) \geq \epsilon\} \subset \bigcup_i \mathbb{B}_{\mathbb{R}^m}(X_i, r_i) \right\}. \tag{5.20}$$

Then the bottleneck distance between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ and its estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ is upper bounded as,

$$d_B \left(\mathrm{PH}_*^{\mathrm{supp}(P)}(\hat{p}_h, r), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) \right) \leq \|\hat{p}_h - p_h\|_\infty + \hat{c}_r, \tag{5.21}$$

where

$$\hat{c}_r := \inf\{\epsilon \in \mathcal{E}_r\} \vee \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)} |\hat{p}_h(X_i) - \hat{p}_h(x)|. \tag{5.22}$$

It is important to realize that, since $\mathcal{E}_r$ in (5.20) is defined based on sample points and the values of the KDE only, the quantity $\hat{c}_r$ in (5.22) is computable without any knowledge about the underlying distribution P. From a statistical standpoint, this is key, as it makes it possible to build confidence sets for $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$.

As we did in Section 5.1, Rips complexes can be used to build computable estimators $\mathrm{PH}_*^R(\hat{p}_h, r)$ instead of $\mathrm{PH}_*^{\mathrm{supp}(P)}(\hat{p}_h, r)$, and Lemma 77 can be extended to $\mathrm{PH}_*^R(\hat{p}_h, r)$ by replacing $\hat{c}_r$ with $\hat{c}_r \vee \hat{c}_{2r}$. Since $\hat{c}_r$ and $\hat{c}_{2r}$ are numerically computable, once we get a confidence set for $\|\hat{p}_h - p_h\|_\infty$, we can easily convert it into the confidence set for our target quantity, $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$. In this chapter, we use the standard bootstrap based approach. We refer Chazal et al. [2014d] for the detailed discussion about the validity of the bootstrap procedure and its TDA applications.

First, we generate B bootstrap samples $\{\tilde{X}_1^1, \ldots, \tilde{X}_n^1\}, \ldots, \{\tilde{X}_1^B, \ldots, \tilde{X}_n^B\}$, by sampling with replacement from the original sample. On each bootstrap sample, let $T_i = \sqrt{nh^d}\|\hat{p}_h - \hat{p}_h^i\|_\infty$, where $\hat{p}_h^i$ is the kernel density estimator computed on ith bootstrap samples $\{\tilde{X}_1^i, \ldots, \tilde{X}_n^i\}$. Let the bootstrap quantile $\hat{z}_\alpha$ be

$$\hat{z}_\alpha = \inf\left\{ z : \ \frac{1}{B}\sum_{i=1}^{B} I(T_i > z) \leq \alpha \right\}. \tag{5.23}$$

Then, for large enough B, we have the following inequality which gives a $1-\alpha$ asymptotic confidence set for $\|p_h - \hat{p}_h\|_\infty$ with fixed $h > 0$.

$$\mathbb{P}\left(\sqrt{nh^d}\|\hat{p}_h - p_h\|_\infty) \leq \hat{z}_\alpha\right) = 1 - \alpha + O\left(\frac{1}{\sqrt{n}}\right). \tag{5.24}$$

Based on the (5.24), we get the following asymptotic confidence sets for the persistent homology $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$,

$$\hat{C}_\alpha^R := \left\{ \mathcal{P} : \ d_B\left(\mathcal{P}, \mathrm{PH}_*^R(\hat{p}_h, r)\right) \leq \frac{\hat{\hat{z}}_\alpha}{\sqrt{nh^d}} + \hat{c}_r \vee \hat{c}_{2r} \right\}. \tag{5.25}$$

$\hat{C}_\alpha^R$ is a valid asymptotic $1-\alpha$ confidence set for $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ as in the following theorem:

Theorem 78. *Suppose Assumption 66 and 67 holds. Let $\{r_n = (r_{n,1}, \ldots, r_{n,n})\}_{n\in\mathbb{N}}$ be a triangular array of positive numbers such that $\sqrt{2}\|r_n\|_\infty \leq \tau$ for all sufficiently large n. Then, the confidence set $\hat{C}_\alpha^R$ in* (5.25) *is asymptotically valid and satisfies*

$$\mathbb{P}\left(d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq \frac{\hat{\hat{z}}_\alpha}{\sqrt{nh^d}} + \hat{c}_{r_n} \vee \hat{c}_{2r_n} \right) \geq 1 - \alpha + O\left(\frac{1}{\sqrt{n}}\right).$$

Remark 79. If $r_{n,1} = \cdots = r_{n,n}$ and $\mathbb{X}$ is a Euclidean space, we can replace $\hat{c}_{2r_n}$ with $\hat{c}_{\sqrt{2}r_n}$.

5.3 Examples

To illustrate how one can use the methods in the previous section to do statistical inference on topological features of data generating distributions, we calculate persistence diagrams of our proposed estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ in Definition 70 and their confidence sets in (5.25) on toy examples. We make 2 synthetic data sets with circular shapes which are described in the left side of Figure 5.3 and 5.4. The right side shows persistence diagrams of $\mathrm{PH}_*^R(\hat{p}_h, r)$. Each black dot indicates the birth and death of each 0-th homology class corresponding to each connected component. Similarly, each red triangle represents the birth and death of each 1-st homology class related to each one-dimensional hole. For all

diagrams, the shaded banded regions correspond to 90% confidence sets in the sense that any homology class contained in the bands cannot be distinguished from the diagonal lines within the confidence sets. In other words, homology classes outside of band illustrate significant topological features of the underlying distribution. We refer to Fasy et al. [2014b] for the detailed interpretation. In Figure 5.3c and 5.4c, we can check there are a black dot and a red triangle outside of band which coincide to the fact that most of the data are distributed around a circle with a hole.

Persistence diagrams of $\mathrm{PH}_*^R(\hat{p}_h, r)$ depend on choices of parameters h and $r = (r_1, \ldots, r_n)$. In all examples, $r_i = r, \forall i = 1, \ldots, n$ are chosen to minimize $\hat{c}_r \vee \hat{c}_{\sqrt{2}r}$ for given h. To choose appropriate h, we can select the parameter that maximizes the total number of significant homology classes which is a generally adopted strategy in TDA [Chazal et al., 2014a].

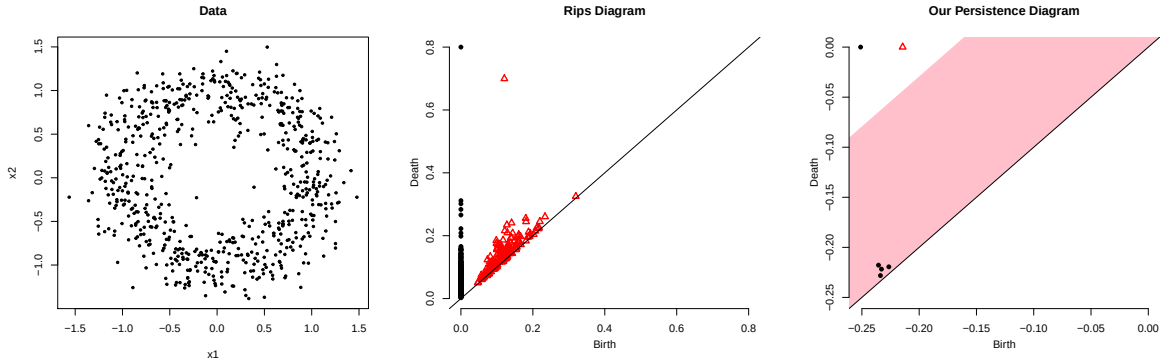
Remark 80. For our methods, we can also use another heuristic but intuitive parameter selection method based on the diagram of the Rips complex filtration

$$\{R(\mathcal{X}, r)\}_{r>0}. \tag{5.26}$$

Recall that $\mathrm{PH}_*^R(\hat{p}_h, r)$ in Definition 70 is the persistent homology of the filtration

$$\left\{R\left(\mathcal{X}_{n,L}^{\hat{p}_h}, r\right)\right\}_{L>0}.$$

Since it is based on Rips complex with radius r, $\mathrm{PH}_*^R(\hat{p}_h, r)$ can only capture the homology classes whose birth time is smaller than r and death time is greater than r in the usual Rips persistence diagram of the filtration in (5.26). Therefore, once the Rips persistence diagram in (5.26) reveals some seemingly significant homology classes whose lifetimes are longer than the others, we can choose appropriate h and r to make sure the base line Rips complex $R(\mathcal{X}, r)$ contain the seemingly significant homology groups.

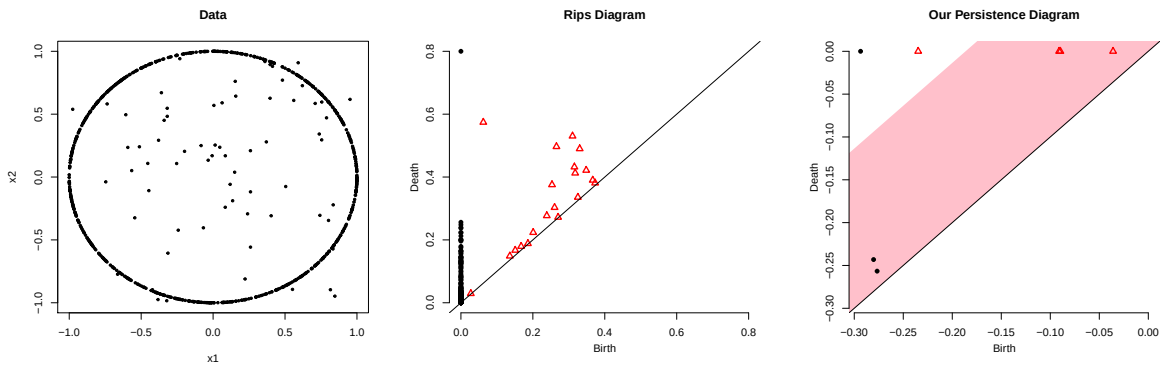


(a) (b) (c)

Figure 5.3: *One circle with additive noise example.* (a) 700 data points uniformly distributed over a circle of radius 1 with additive Gaussian noise $\mathcal{N}(0, ?)$. (b) The usual Rips persistence diagram of the filtration in (5.26). (c) Persistence diagram of KDE filtration ($h = 0.6$) on Rips complex as in Definition 70. The shaded area represents the confidence set as in (5.25).

5.4 Computation time comparison

In worst-case, the time complexity of persistent homology computation is known to be the order of $O(N^3)$ where N is the number of simplices in the underlying simplicial complex. Therefore, when the



(a) (b) (c)

Figure 5.4: *One circle with background noise example.* (a) 700 data points uniformly distributed over a circle of radius 1, and 70 outliers are added to the data set ($n = 770$). (b) the usual Rips persistence diagram of the filtration in (5.26). (c) Persistence diagram of KDE filtration ($h = 0.6$) on Rips complex as in Definition 70. The shaded area represents the confidence set as in (5.25).

ambient space has large dimension or topological features are heterogeneously distributed, in which case we need large size of grid points to approximate the ambient space precisely, our proposed estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ in Definition 70 could be computationally efficient to infer the topological features of the data generating distributions.

In this section, we demonstrate the computational advantage of our method in 2 series of synthetic data sets in which we expect the Rips complex based approach is computationally more efficient than the grid-based ones.

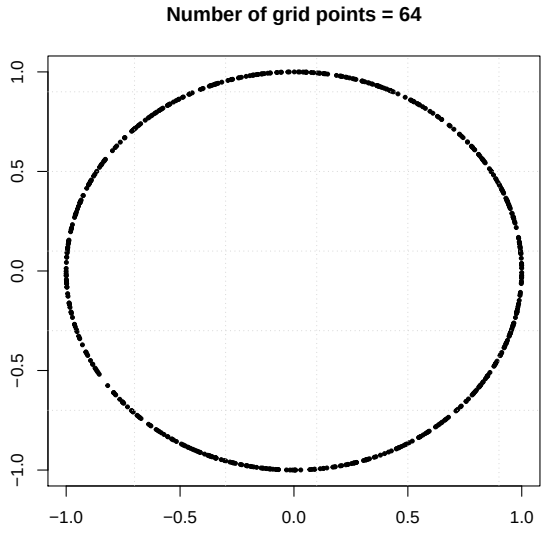
5.4.1 Large dimensional ambient space

We generate a set of 600 sample points uniformly distributed on a 2-dimensional circle of radius 1 (Figure 5.5a). Then, by using a fixed orthonormal matrix, we embed the 2-dimensional circular sample points in higher dimensional spaces ($d = 3, 4, 5$). Figure 5.5b shows the computation time of grid and Rips complex based persistent homology estimators in log scale. For both methods, a fixed bandwidth ($h = 0.2$) is used for all cases. The dashed lines in Figure 5.5a represent the grid used for the 2-dimensional sample points. Grids with the same resolution are used for higher dimensional cases. The parameter r in the Rips complex based estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ is chosen to minimize $\hat{c}_r \vee \hat{c}_{\sqrt{2}r}$ in the 2-dimensional case, and the same r is used for higher dimensional cases.

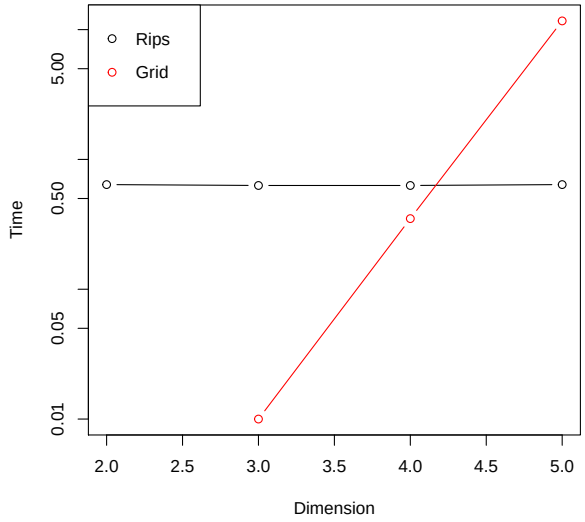
The time complexity of grid-based estimator increases exponentially as the dimension of the ambient space increases because the number of grid points required to approximate the space increase exponentially. In contrast, the Rips-complex based estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ in Definition 70 has constant time complexity because the computational time is dominated by the number of sample points which is constant in this experiment. A similar result is obtained for two circles case described in Figure 5.5c and 5.5d.

5.4.2 Heterogeneously distributed topological features

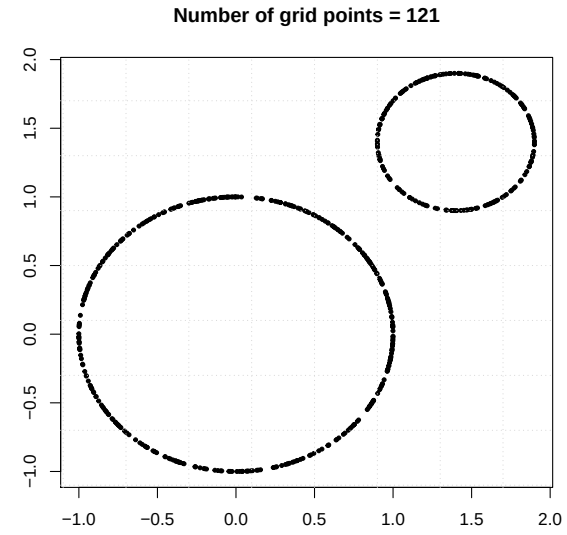
We generate two sets of sample points uniformly distributed on two circles in $\mathbb{R}^2$ (Figure 5.6a). Then we increase the distance between two circles from $2\sqrt{2}$ to $32\sqrt{2}$. Figure 5.6d shows the computation



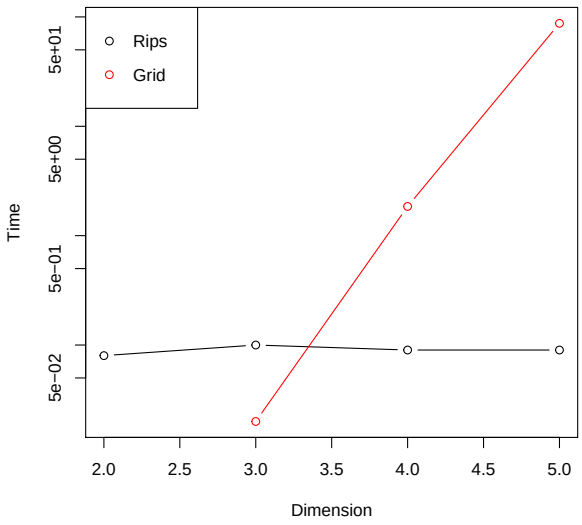
(a) Sample points on a circle ($n = 600$)



(b) Computation time vs Dimension of ambient space



(c) Sample points on two circles ($n = 600$)

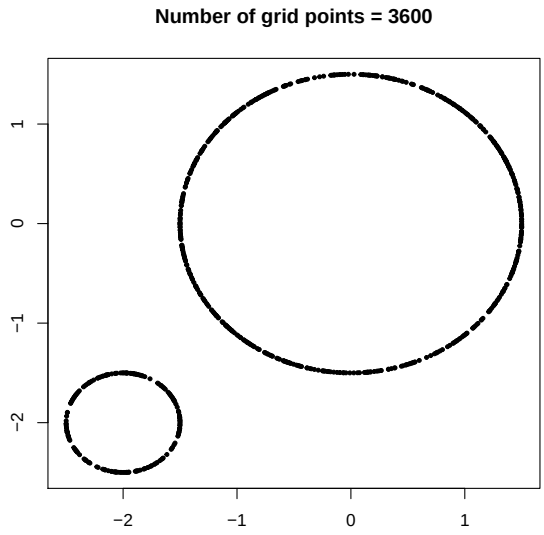


(d) Computation time vs Dimension of ambient space

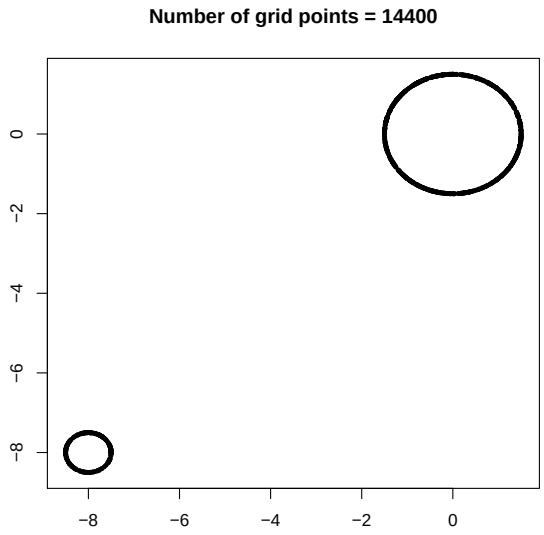
Figure 5.5: Time complexity comparison between grid and Rips complex based persistent homology estimator when the dimension of ambient space increases.

time of grid and Rips complex based persistent homology estimators in log scale. For both methods, a fixed bandwidth ($h = 0.2$) is used for all cases. Grids with the same resolution are used for all cases. The parameter r in $\mathrm{PH}_*^R(\hat{p}_h, r)$ is chosen to minimize $\hat{c}_r \vee \hat{c}_{\sqrt{2}r}$ in the 2-dimensional case, and the same r is used for all the other cases.

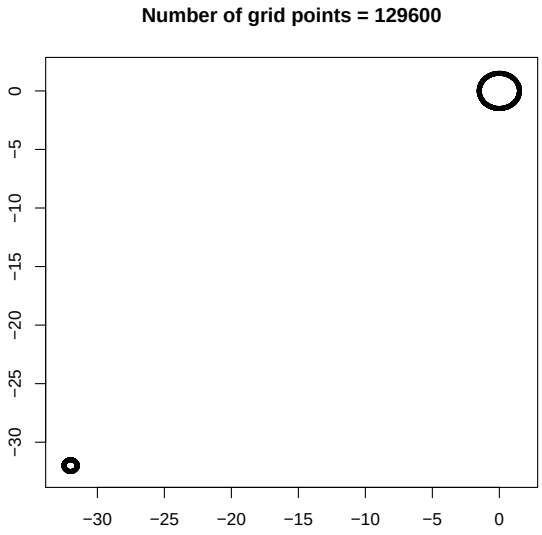
The time complexity of grid-based estimator increase as the distance between centers of two circles increase because a larger number of grid points are required to cover the larger ambient space. In contrast, the Rips-complex based estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ in Definition 70 has constant time complexity because the computational time is dominated by the number of sample points which is constant in this experiment.



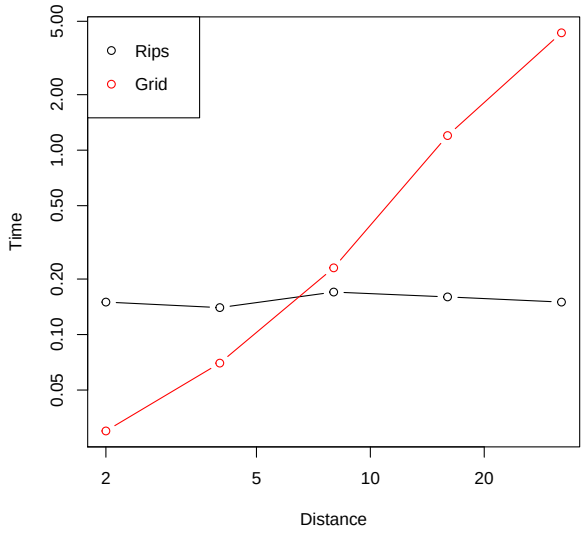
(a) Sample points on two circles (Distance = $2\sqrt{2}$)



(b) Sample points on two circles (Distance = $8\sqrt{2}$)



(c) Sample points on two circles (Distance = $32\sqrt{2}$)



(d) Computation time vs Distance between two circular points

Figure 5.6: Time complexity comparison between grid and Rips complex based persistent homology estimator when the distance between the centers of two circles increases.

Chapter 6

R Package TDA: Statistical Tools for Topological Data Analysis

This chapter presents the work in Fasy et al. [2014a].

This chapter is devoted to the presentation of the R package **TDA**, which provides a user-friendly interface for the efficient algorithms of the C++ libraries **GUDHI** [Maria, 2014], **Dionysus** [Morozov, 2007], and **PHAT** [Bauer et al., 2012].

In Section 6.1, we describe how to compute some widely studied functions that, starting from a point cloud, provide some topological information about the underlying space: the distance function (distFct), the distance to a measure function (dtm), the k Nearest Neighbor density estimator (knnDE), the kernel density estimator (kde), and the kernel distance (kernelDist). Section 6.2 is devoted to the computation of persistence diagrams: the function gridDiag can be used to compute persistent homology of sublevel sets (or superlevel sets) of functions evaluated over a grid of points; the function ripsDiag returns the persistence diagram of the Rips filtration built on top of a point cloud.

One of the key challenges in persistent homology is to find a way to isolate the points of the persistence diagram representing the topological noise. Statistical methods for persistent homology provide an alternative to its exact computation. Knowing with high confidence that an approximated persistence diagrams is close to the true–computationally infeasible–diagram is often enough for practical purposes. Fasy et al. [2014b], Chazal et al. [2014c], and Chazal et al. [2014a] propose several statistical methods to construct confidence sets for persistence diagrams and other summary functions that allow us to separate topological signal from topological noise. The methods are implemented in the **TDA** package and described in Section 6.2.

Finally, the **TDA** package provides the implementation of an algorithm for density clustering. This method allows us to identify and visualize the spatial organization of the data, without specific knowledge about the data generating mechanism and in particular without any a priori information about the number of clusters. In Section 6.3, we describe the function clusterTree, that, given a density estimator, encodes the hierarchy of the connected components of its superlevel sets into a dendrogram, the cluster tree [Kpotufe and von Luxburg, 2011, Kent, 2013].

6.1 Distance Functions and Density Estimators

As a first toy example to using the **TDA** package, we show how to compute distance functions and density estimators over a grid of points. The setting is the typical one in TDA: a set of points $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ has been sampled from some distribution P and we are interested in recovering

the topological features of the underlying space by studying some functions of the data. The following code generates a sample of 400 points from the unit circle and constructs a grid of points over which we will evaluate the functions.

```
library("TDA")
X <- circleUnif(400)

Xlim <- c(-1.6, 1.6);  Ylim <- c(-1.7, 1.7);  by <- 0.065

Xseq <- seq(Xlim[1], Xlim[2], by = by)
Yseq <- seq(Ylim[1], Ylim[2], by = by)
Grid <- expand.grid(Xseq, Yseq)
```

The **TDA** package provides implementations of the following functions:

- The distance function is defined for each $y \in \mathbb{R}^d$ as $\Delta(y) = \inf_{x \in X} \|x - y\|_2$ and is computed for each point of the Grid with the following code:

```
distance <- distFct(X = X, Grid = Grid)
```

- Given a probability measure P, the distance to measure (DTM) is defined for each $y \in \mathbb{R}^d$ as

$$d_{m0}(y) = \left(\frac{1}{m0} \int_0^{m0} (G_y^{-1}(u))^r du \right)^{1/r},$$

where $G_y(t) = P(\|X - y\| \leq t)$, and $m0 \in (0, 1)$ and $r \in [1, \infty)$ are tuning parameters. As m0 increases, DTM function becomes smoother, so m0 can be understood as a smoothing parameter. r affects less but also changes DTM function as well. The default value of r is 2. The DTM can be seen as a smoothed version of the distance function. See [Chazal et al., 2011a, Definition 3.2] and [Chazal et al., 2015, Equation (2)] for a formal definition of the "distance to measure" function.

Given $X = \{x_1, \ldots, x_n\}$, the empirical version of the DTM is

$$\hat{d}_{m0}(y) = \left(\frac{1}{k} \sum_{x_i \in N_k(y)} \|x_i - y\|^r \right)^{1/r},$$

where $k = \lceil m0 * n \rceil$ and $N_k(y)$ is the set containing the k nearest neighbors of y among $x_1, \ldots, x_n$.

For more details, see [Chazal et al., 2011a] and [Chazal et al., 2015].

The DTM is computed for each point of the Grid with the following code:

```
m0 <- 0.1
DTM <- dtm(X = X, Grid = Grid, m0 = m0)
```

- The k Nearest Neighbor density estimator, for each $y \in \mathbb{R}^d$, is defined as

$$\hat{\delta}_k(y) = \frac{k}{n \, v_d \, r_k^d(y)},$$

where v_n is the volume of the Euclidean d dimensional unit ball and $r_k^d(x)$ is the Euclidean distance form point x to its kth closest neighbor among the points of X. It is computed for each point of the Grid with the following code:

```
k <- 60
kNN <- knnDE(X = X, Grid = Grid, k = k)
```

- The Gaussian Kernel Density Estimator (KDE), for each $y \in \mathbb{R}^d$, is defined as

$$\hat{p}_h(y) = \frac{1}{n(\sqrt{2\pi}h)^d} \sum_{i=1}^{n} \exp\left(\frac{-\|y - x_i\|_2^2}{2h^2}\right).$$

  where h is a smoothing parameter. It is computed for each point of the Grid with the following code:

```
h <- 0.3
KDE <- kde(X = X, Grid = Grid, h = h)
```

- The Kernel distance estimator, for each $y \in \mathbb{R}^d$, is defined as

$$\hat{\kappa}_h(y) = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_h(x_i, x_j) + K_h(y, y) - 2\frac{1}{n}\sum_{i=1}^{n} K_h(y, x_i)},$$

  where $K_h(x, y) = \exp\left(\frac{-\|x-y\|_2^2}{2h^2}\right)$ is the Gaussian Kernel with smoothing parameter h. The Kernel distance is computed for each point of the Grid with the following code:

```
h <- 0.3
Kdist <- kernelDist(X = X, Grid = Grid, h = h)
```

For this 2 dimensional example, we can visualize the functions using persp form the **graphics** package. For example the following code produces the KDE plot in Figure 6.1:

```
persp(Xseq, Yseq,
    matrix(KDE, ncol = length(Yseq), nrow = length(Xseq)), xlab = "",
    ylab = "", zlab = "", theta = -20, phi = 35, ltheta = 50,
    col = 2, border = NA, main = "KDE", d = 0.5, scale = FALSE,
    expand = 3, shade = 0.9)
```

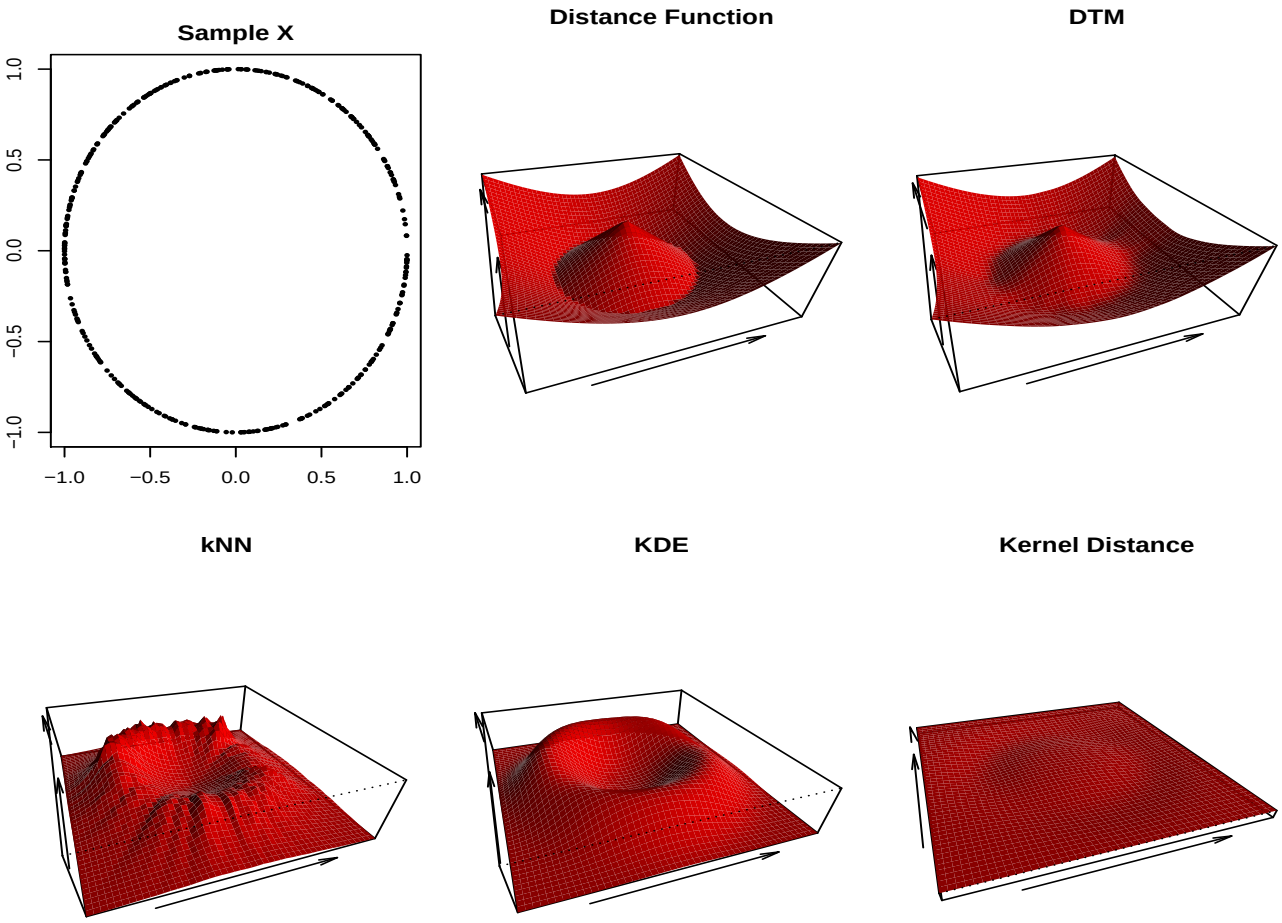


Figure 6.1: distance functions and density estimators evaluated over a grid of points.

6.1.1 Bootstrap Confidence Bands

We can construct a $(1-\alpha)$ confidence band for a function using the bootstrap algorithm, which we briefly describe using the kernel density estimator:

1. Given a sample $X = \{x_1, \ldots, x_n\}$, compute the kernel density estimator $\hat{p}_h$;
2. Draw $X^* = \{x_1^*, \ldots, x_n^*\}$ from $X = \{x_1, \ldots, x_n\}$ (with replacement), and compute $\theta^* = \sqrt{n}\|\hat{p}_h^*(x) - \hat{p}_h(x)\|_\infty$, where $\hat{p}_h^*$ is the density estimator computed using X^*;
3. Repeat the previous step B times to obtain $\theta_1^*, \ldots, \theta_B^*$;
4. Compute $q_\alpha = \inf\left\{q : \frac{1}{B}\sum_{j=1}^{B} I(\theta_j^* \geq q) \leq \alpha\right\}$;
5. The $(1-\alpha)$ confidence band for $\mathbb{E}[\hat{p}_h]$ is $\left[\hat{p}_h - \frac{q_\alpha}{\sqrt{n}}\,,\ \hat{p}_h + \frac{q_\alpha}{\sqrt{n}}\right]$.

Fasy et al. [2014b] and Chazal et al. [2014a] prove the validity of the bootstrap algorithm for kernel density estimators, distance to measure, and kernel distance, and use it in the framework of persistent homology. The bootstrap algorithm is implemented in the function bootstrapBand, which provides the option of parallelizing the algorithm (parallel = TRUE) using the package **parallel**. The following code computes a 90% confidence band for $\mathbb{E}[\hat{p}_h]$, showed in Figure 6.2.

```
band <- bootstrapBand(X = X, FUN = kde, Grid = Grid, B = 100,
                      parallel = FALSE, alpha = 0.1, h = h)
```

Figure 6.2: the 90% confidence band for $\mathbb{E}[\hat{p}_h]$ has the form $[\ell, u] = [\hat{p}_h - q_\alpha/\sqrt{n}\,,\, \hat{p}_h + q_\alpha/\sqrt{n}]$. The plot on the right shows a section of the functions: the red surface is the KDE $\hat{p}_h$; the pink surfaces are ℓ and u.

6.2 Persistent Homology

We provide an informal description of the implemented methods of persistent homology. We assume the reader is familiar with the basic concepts and, for a rigorous exposition, we refer to the textbook Edelsbrunner and Harer [2010].

6.2.1 Persistent Homology Over a Grid

In this section, we describe how to use the gridDiag function to compute the persistent homology of sublevel (and superlevel) sets of the functions described in Section 6.1. The function gridDiag evaluates a given real valued function over a triangulated grid, constructs a filtration of simplices using the values of the function, and computes the persistent homology of the filtration. From version 1.2, gridDiag works in arbitrary dimension. The core of the function is written in C++ and the user can choose to compute persistence diagrams using either the C++ library **GUDHI**, **Dionysus**, or **PHAT**.

The following code computes the persistent homology of the superlevel sets (sublevel = FALSE) of the kernel density estimator (FUN = kde, h = 0.3) using the point cloud stored in the matrix X from the previous example. The same code would work for the other functions defined in Section 6.1 (it is sufficient to replace kde and its smoothing parameter h with another function and the corresponding parameter). The function gridDiag returns an object of the class "diagram". The other inputs are the features of the grid over which the kde is evaluated (lim and by), the smoothing parameter h, and a logical variable that indicates whether a progress bar should be printed (printProgress).

```
DiagGrid <- gridDiag(
    X = X, FUN = kde, h = 0.3, lim = cbind(Xlim, Ylim), by = by,
    sublevel = FALSE, library = "Dionysus", location = TRUE,
    printProgress = FALSE)
```

We plot the data and the diagram, using the function plot, implemented as a standard S3 method for objects of the class "diagram". The following command produces the third plot in Figure 6.3.

```
plot(DiagGrid[["diagram"]], band = 2 * band[["width"]],
     main = "KDE Diagram")
```

The option (band = 2 * band[["width"]]) produces a pink confidence band for the persistence diagram, using the confidence band constructed for the corresponding kernel density estimator in the previous section. The features above the band can be interpreted as representing significant homological features, while points in the band are not significantly different from noise. The validity of the bootstrap confidence band for persistence diagrams of KDE, DTM, and Kernel Distance derive from the *Stability Theorem* [Chazal et al., 2012] and is discussed in detail in Fasy et al. [2014b] and Chazal et al. [2014a].

The function plot for the class "diagram" provide the options of rotating the diagram (rotated = TRUE), drawing the barcode in place of the diagram (barcode = TRUE), as well as other standard graphical options. See Figure 6.4.

6.2.2 Rips Diagrams

The *Vietoris-Rips complex* $R(X, \varepsilon)$ consists of simplices with vertices in $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ and diameter at most ε. In other words, a simplex σ is included in the

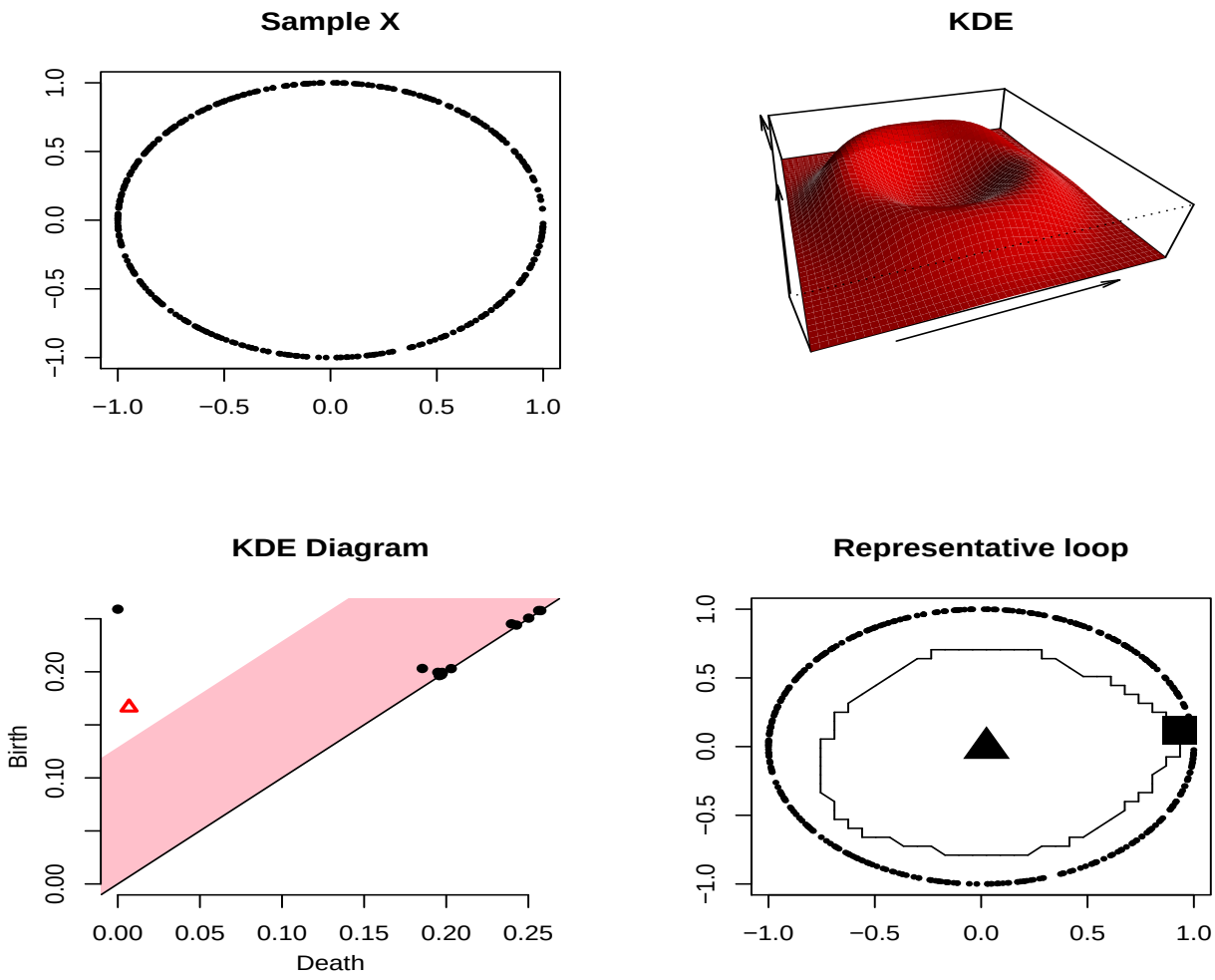


Figure 6.3: The plot on the right shows the persistence diagram of the superlevel sets of the KDE. Black points represent connected components and red triangles represent loops. The features are born at high levels of the density and die at lower levels. The pink 90% confidence band separates significant features from noise.

complex if each pair of vertices in σ is at most ε apart. The sequence of Rips complexes obtained by gradually increasing the radius ε creates a filtration.

The ripsDiag function computes the persistence diagram of the Rips filtration built on top of a point cloud. The user can choose to compute the Rips filtration using either the C++ library **GUDHI** or **Dionysus**. Then for computing the persistence diagram from the Rips filtration, the user can use either the C++ library **GUDHI**, **Dionysus**, or **PHAT**.
The following code generates 60 points from two circles:

```
Circle1 <- circleUnif(60)
Circle2 <- circleUnif(60, r = 2) + 3
Circles <- rbind(Circle1, Circle2)
```

We specify the limit of the Rips filtration and the max dimension of the homological features we are interested in (0 for components, 1 for loops, 2 for voids, etc.):

```
par(mfrow = c(1, 2), mai = c(0.8, 0.8, 0.3, 0.1))
plot(DiagGrid[["diagram"]], rotated = TRUE, band = band[["width"]],
     main = "Rotated Diagram")
plot(DiagGrid[["diagram"]], barcode = TRUE, main = "Barcode")
```

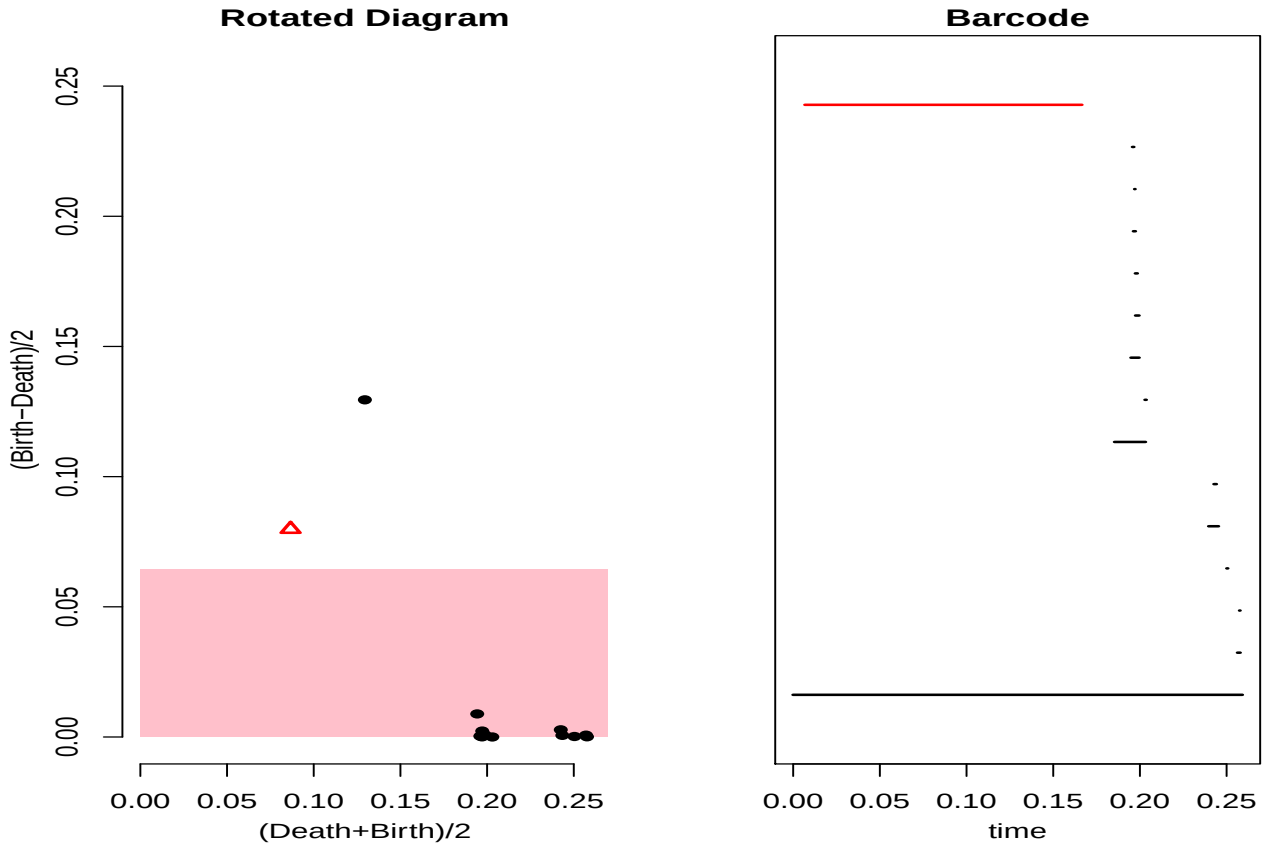


Figure 6.4: Rotated Persistence Diagram and Barcode

```
maxscale <- 5          # limit of the filtration
maxdimension <- 1      # components and loops
```

and we generate the persistence diagram:

```
DiagRips <- ripsDiag(X = Circles, maxdimension, maxscale,
    library = c("GUDHI", "Dionysus"), location = TRUE,
    printProgress = FALSE)
```

Alternatively, using the option (dist = ”arbitrary”) in ripsDiag(), the input X can be an $n \times n$ matrix of distances. This option is useful when the user wants to consider a Rips filtration constructed using an arbitrary distance and is currently only available for the option (library = ”Dionysus”).

Finally we plot the data and the diagram, as in Figure 6.5.:

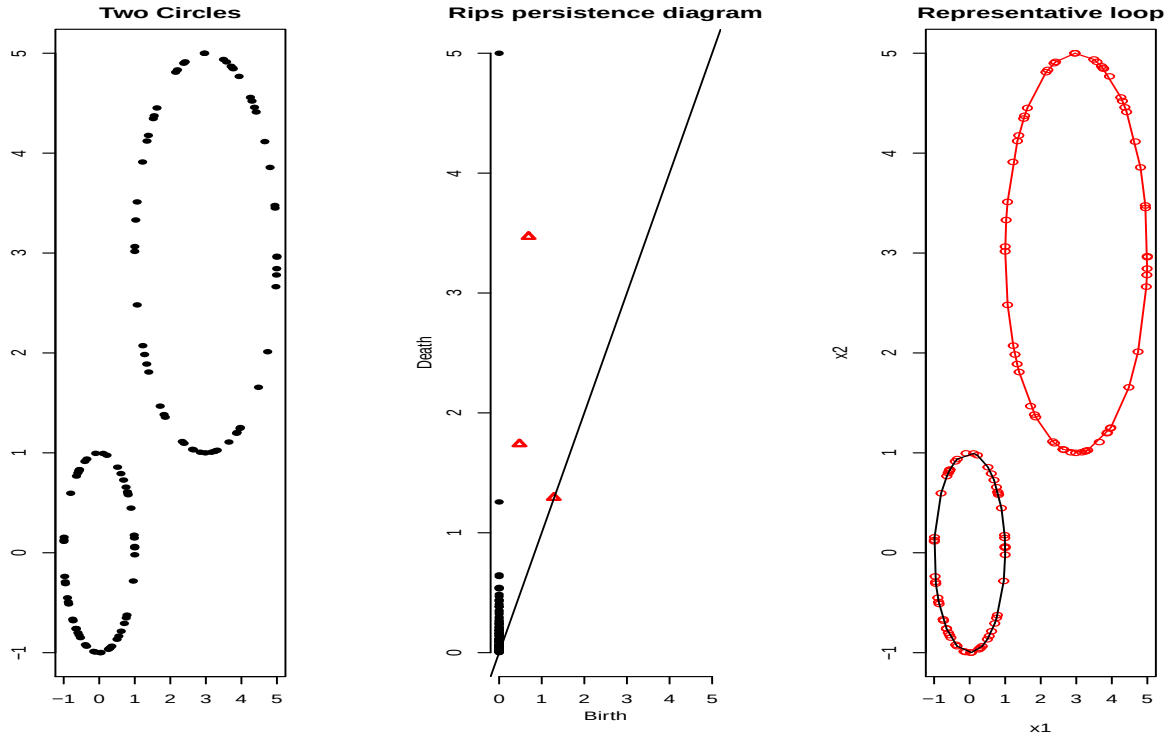


Figure 6.5: Rips persistence diagram. Black points represent connected components and red triangles represent loops.

6.2.3 Alpha Complex Persistence Diagram

For a finite set of points $X \subset \mathbb{R}^d$, the *Alpha complex* $Alpha(X, s)$ is a simplicial subcomplex of the Delaunay complex of X consisting of simplices of circumradius less than or equal to $\sqrt{s}$. For each $u \in X$, let V_u be its Voronoi cell, i.e. $V_u = \{x \in \mathbb{R}^d : d(x, u) \leq d(x, v) \text{ for all } v \in X\}$, and $B_u(r)$ be the closed ball with center u and radius r. Let $R_u(r)$ consists of be the intersection of earh ball of radius r with the voronoi cell of u, i.e. $R_u(r) = B_u(r) \cap V_u$. Then $Alpha(X, s)$ is defined as

$$Alpha(X, r) = \left\{ \sigma \subset X : \bigcap_{u \in \sigma} R_u(\sqrt{s}) \neq \emptyset \right\}.$$

See [Edelsbrunner and Harer, 2010, Section 3.4] and [Rouvreau, 2015]. The sequence of Alpha complexes obtained by gradually increasing the parameter s creates an Alpha complex filtration.

The alphaComplexDiag function computes the Alpha complex filtration built on top of a point cloud, using the C++ library **GUDHI**. Then for computing the persistence diagram from the Alpha complex filtration, the user can use either the C++ library **GUDHI**, **Dionysus**, or **PHAT**.

We first generate 30 points from a circle:

```
X <- circleUnif(n = 30)
```

and the following code compute the persistence diagram of the alpha complex filtration using the point cloud X, with printing its progress (printProgress = FALSE). The function alphaComplexDiag returns an object of the class "diagram".

```
# persistence diagram of alpha complex
DiagAlphaCmplx <- alphaComplexDiag(
    X = X, library = c("GUDHI", "Dionysus"), location = TRUE,
    printProgress = TRUE)
## # Generated complex of size: 115
##
## 0%   10   20   30   40   50   60   70   80   90   100%
## |----|----|----|----|----|----|----|----|----|----|
## **************************************************
## # Persistence timer: Elapsed time [ 0.000000 ] seconds
```

And we plot the diagram in Figure 6.6.

6.2.4 Persistence Diagram of Alpha Shape

The *Alpha shape complex* $S(X, \alpha)$ is the polytope with its boundary consisting of α-exposed simplices, where a simplex σ is α-exposed if there is an open ball b of radius α such that $b \cap X = \emptyset$ and $\partial b \cap X = \sigma$. Suppose $\mathbb{R}^d$ is filled with ice cream, then consider scooping out the ice cream with sphere-shaped spoon of radius α without touching the points X. $S(X, \alpha)$ is the remaining polytope with straightening round surfaces. See [Fischer, 2005] and [Edelsbrunner and Mücke, 1994]. The sequence of Alpha shape complexes obtained by gradually increasing the parameter α creates an Alpha shape complex filtration.

The alphaShapeDiag function computes the persistence diagram of the Alpha shape filtration built on top of a point cloud in 3 dimension, using the C++ library **GUDHI**. Then for computing the persistence diagram from the Alpha shape filtration, the user can use either the C++ library **GUDHI**, **Dionysus**, or **PHAT**. Currently the point data cloud should lie in 3 dimension.

We first generate 30 points from a cylinder:

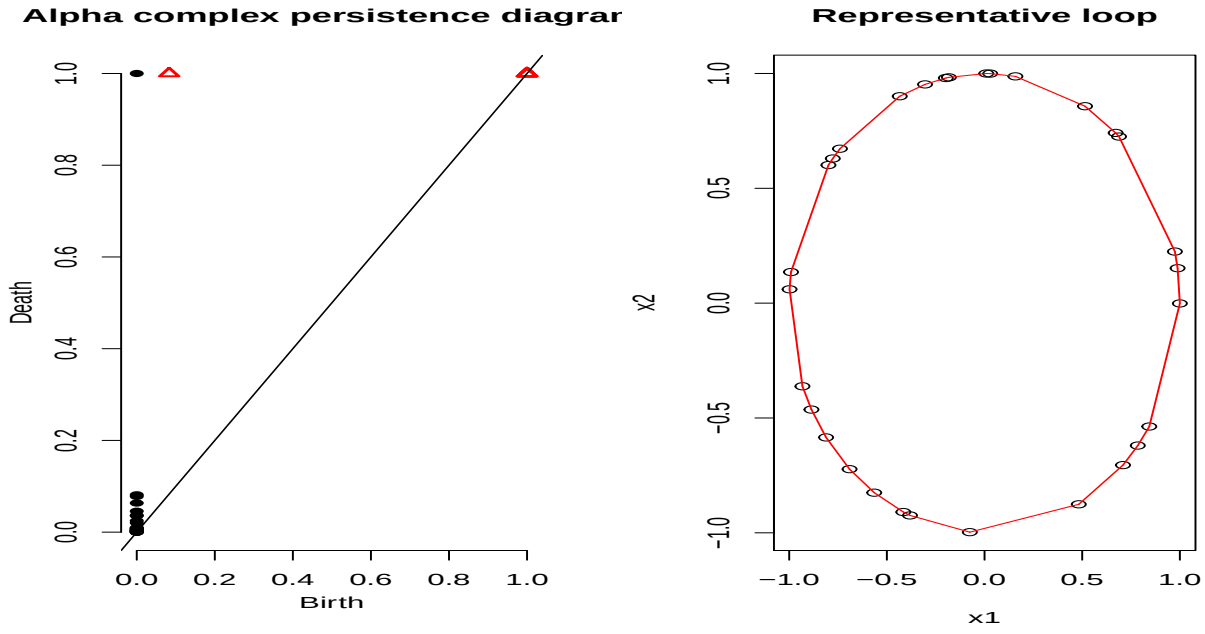
```
n <- 30
X <- cbind(circleUnif(n = n), runif(n = n, min = -0.1, max = 0.1))
```

and the following code compute the persistence diagram of the alpha shape filtration using the point cloud X, with printing its progress (printProgress = TRUE). The function alphaShapeDiag returns an object of the class "diagram".

```
DiagAlphaShape <- alphaShapeDiag(
    X = X, maxdimension = 1, library = c("GUDHI", "Dionysus"),
    location = TRUE, printProgress = TRUE)
## # Generated complex of size: 543
##
## 0%   10   20   30   40   50   60   70   80   90   100%
## |----|----|----|----|----|----|----|----|----|----|
## **************************************************
## # Persistence timer: Elapsed time [ 0.002000 ] seconds
```

And we plot the diagram and first two dimension of data in Figure 6.7.

```
# plot
par(mfrow = c(1, 2))
plot(DiagAlphaCmplx[["diagram"]],
    main = "Alpha complex persistence diagram")
one <- which(DiagAlphaCmplx[["diagram"]][, 1] == 1)
one <- one[which.max(DiagAlphaCmplx[["diagram"]][one, 3] -
    DiagAlphaCmplx[["diagram"]][one, 2])]
plot(X, col = 1, main = "Representative loop")
for (i in seq(along = one)) {
  for (j in
      seq_len(dim(DiagAlphaCmplx[["cycleLocation"]][[one[i]]])[1])) {
    lines(DiagAlphaCmplx[["cycleLocation"]][[one[i]]][j, , ],
        pch = 19, cex = 1, col = i + 1)
  }
}
```



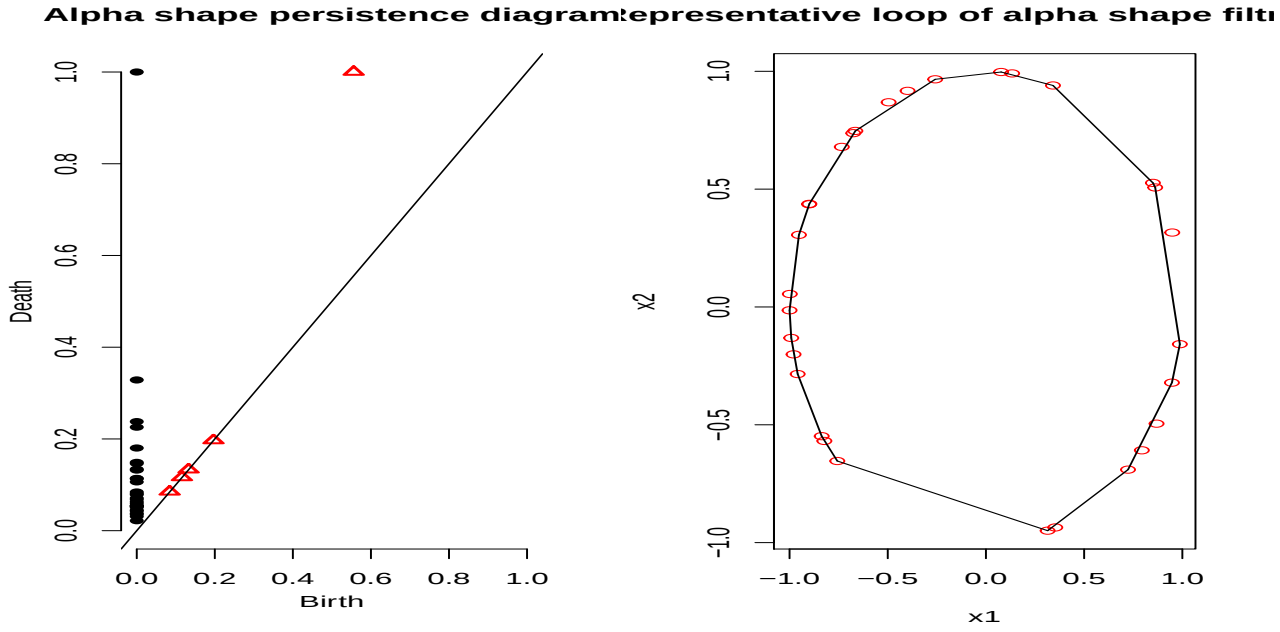

```
par(mfrow = c(1, 1))
```

Figure 6.6: Persistence diagram of Alpha complex. Black points represent connected components and red triangles represent loops.

6.2.5 Persistence Diagrams from Filtration

Rather than computing persistence diagrams from built-in function, it is also possible to compute persistence diagrams from a user-defined filtration. A filtration consists of simplicial complex and the filtration values on each simplex. The functions ripsDiag, alphaComplexDiag, alphaShapeDiag have their counterparts for computing corresponding filtrations instead of persistence diagrams: namely, ripsFiltration corresponds to the Rips filtration built on top of a point cloud, alphaComplexFiltration to the alpha complex filtration, and alphaShapeFiltration to the alpha shape filtration.

```
par(mfrow = c(1, 2))
plot(DiagAlphaShape[["diagram"]],
    main = "Alpha shape persistence diagram")
plot(X[, 1:2], col = 2,
    main = "Representative loop of alpha shape filtration")
one <- which(DiagAlphaShape[["diagram"]][, 1] == 1)
one <- one[which.max(DiagAlphaShape[["diagram"]][one, 3] -
    DiagAlphaShape[["diagram"]][one, 2])]
for (i in seq(along = one)) {
  for (j in
      seq_len(dim(DiagAlphaShape[["cycleLocation"]][[one[i]]])[1])) {
    lines(
        DiagAlphaShape[["cycleLocation"]][[one[i]]][j, , 1:2],
        pch = 19, cex = 1, col = i)
  }
}
```




```
par(mfrow = c(1, 1))
```

Figure 6.7: Persistence diagram of Alpha shape. Black points represent connected components and red triangles represent loops.

We first generate 100 points from a circle:

```
X <- circleUnif(n = 100)
```

Then, after specifying the limit of the Rips filtration and the max dimension of the homological features, the following code compute the Rips filtration using the point cloud X.

```
maxscale <- 0.4       # limit of the filtration
maxdimension <- 1     # components and loops
FltRips <- ripsFiltration(X = X, maxdimension = maxdimension,
    maxscale = maxscale, dist = "euclidean", library = "GUDHI",
        printProgress = TRUE)
## # Generated complex of size: 2730
```

One way of defining a user-defined filtration is to build a filtration from a simplicial complex and function values on the vertices. The function funFiltration takes function values (FUNvalues) and simplicial complex (cmplx) as input, and build a filtration, where a filtration value on a simplex is defined as the maximum of function values on the vertices of the simplex.
In the following example, the function funFiltration construct a filtration from a Rips complex and the DTM function values on data points.

```
m0 <- 0.1
dtmValues <- dtm(X = X, Grid = X, m0 = m0)
FltFun <- funFiltration(
    FUNvalues = dtmValues, cmplx = FltRips[["cmplx"]])
```

Once the filtration is computed, the function filtrationDiag computes the persistence diagram from the filtration. The user can choose to compute the persistence diagram using either the C++ library **GUDHI** or **Dionysus**.

```
DiagFltFun <- filtrationDiag(
    filtration = FltFun, maxdimension = maxdimension,
    library = "Dionysus", location = TRUE, printProgress = TRUE)
##
## 0%   10   20   30   40   50   60   70   80   90   100%
## |----|----|----|----|----|----|----|----|----|----|
## **************************************************
## # Persistence timer: Elapsed time [ 0.007000 ] seconds
```

Then we plot the data and the diagram in Figure 6.8.

6.2.6 Bottleneck and Wasserstein Distances

Standard metrics for measuring the distance between two persistence diagrams are the bottleneck distance and the pth Wasserstein distance [Edelsbrunner and Harer, 2010]. The **TDA** package includes the functions bottleneck and wasserstein, which are R wrappers of the functions "bottleneck_distance" and "wasserstein_distance" of the C++ library **Dionysus**.

We generate two persistence diagrams of the Rips filtrations built on top of the two (separate) circles of the previous example,

```
Diag1 <- ripsDiag(Circle1, maxdimension = 1, maxscale = 5)
Diag2 <- ripsDiag(Circle2, maxdimension = 1, maxscale = 5)
```

and we compute the bottleneck distance and the 2nd Wasserstein distance between the two diagrams. In the following code, the option dimension = 1 specifies that the distances between diagrams are computed using only one dimensional features (loops).

```
par(mfrow = c(1, 2), mai=c(0.8, 0.8, 0.3, 0.3))
plot(X, pch = 16, xlab = "",ylab = "")
plot(DiagFltFun[["diagram"]], diagLim = c(0, 1))
```

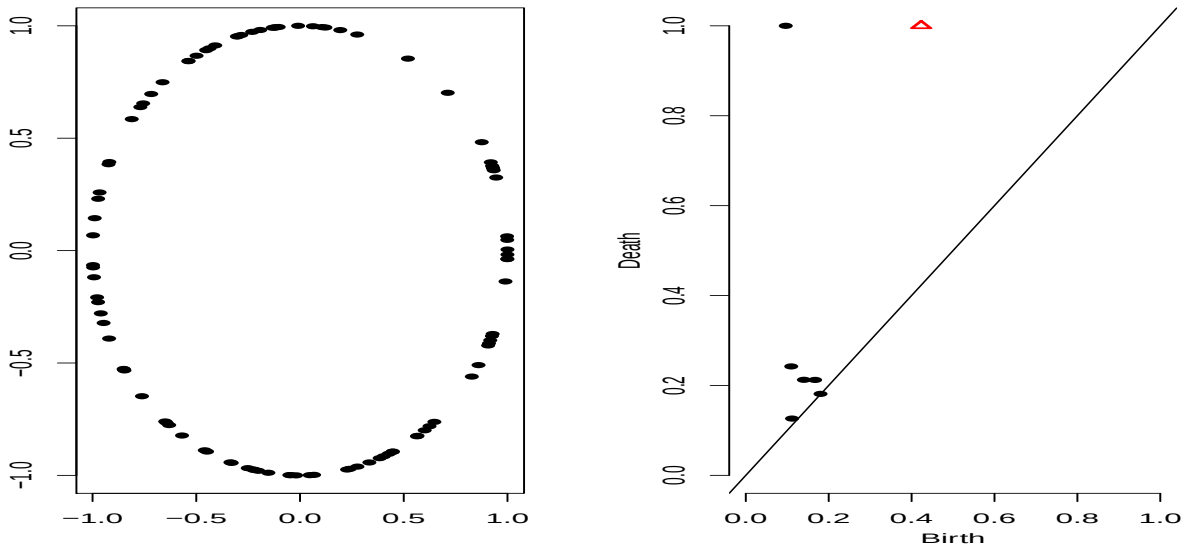


Figure 6.8: Persistence diagram from Rips filtration and DTM function values. Black points represent connected components and red triangles represent loops.

```
print(bottleneck(Diag1[["diagram"]], Diag2[["diagram"]],
                 dimension = 1))
## [1] 1.38913
print(wasserstein(Diag1[["diagram"]], Diag2[["diagram"]], p = 2,
                  dimension = 1))
## [1] 2.327802
```

6.2.7 Landscapes and Silhouettes

Persistence landscapes and silhouettes are real-valued functions that further summarize the information contained in a persistence diagram. They have been introduced and studied in Bubenik [2012], Chazal et al. [2014c], and Chazal et al. [2014b]. We briefly introduce the two functions.

Landscape. The persistence landscape is a collection of continuous, piecewise linear functions $\lambda\colon \mathbb{Z}^+ \times \mathbb{R} \to \mathbb{R}$ that summarizes a persistence diagram. To define the landscape, consider the set of functions created by tenting each each point $p = (x, y) = \left(\frac{b+d}{2}, \frac{d-b}{2}\right)$ representing a birth-death pair (b, d) in the persistence diagram D as follows:

$$\Lambda_p(t) = \begin{cases} t - x + y & t \in [x - y, x] \\ x + y - t & t \in (x, x + y] \\ 0 & \text{otherwise} \end{cases} = \begin{cases} t - b & t \in [b, \frac{b+d}{2}] \\ d - t & t \in (\frac{b+d}{2}, d] \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

We obtain an arrangement of piecewise linear curves by overlaying the graphs of the functions $\{\Lambda_p\}_p$; see Figure 6.9 (left). The persistence landscape of D is a summary of this arrangement. Formally, the

persistence landscape of D is the collection of functions

$$\lambda(k,t) = \underset{p}{k\text{max}}\, \Lambda_p(t), \quad t \in [0,T], k \in \mathbb{N}, \tag{6.2}$$

where $k$max is the kth largest value in the set; in particular, 1max is the usual maximum function. see Figure 6.9 (middle).

Silhouette. Consider a persistence diagram with N off diagonal points $\{(b_j, d_j)\}_{j=1}^N$. For every $0 < p < \infty$ we define the power-weighted silhouette

$$\phi^{(p)}(t) = \frac{\sum_{j=1}^N |d_j - b_j|^p \Lambda_j(t)}{\sum_{j=1}^N |d_j - b_j|^p}.$$

The value p can be thought of as a trade-off parameter between uniformly treating all pairs in the persistence diagram and considering only the most persistent pairs. Specifically, when p is small, $\phi^{(p)}(t)$ is dominated by the effect of low persistence features. Conversely, when p is large, $\phi^{(p)}(t)$ is dominated by the most persistent features; see Figure 6.9 (right).

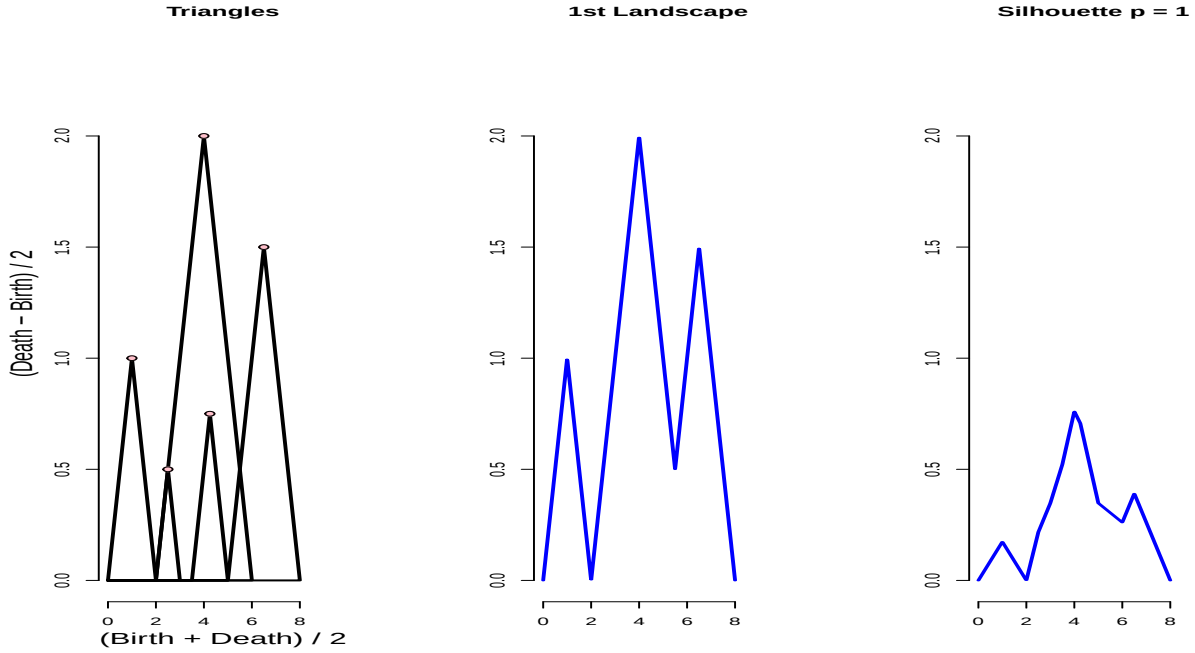


Figure 6.9: Left: we use the rotated axes to represent a persistence diagram D. A feature $(b,d) \in D$ is represented by the point $(\frac{b+d}{2}, \frac{d-b}{2})$ (pink). In words, the x-coordinate is the average parameter value over which the feature exists, and the y-coordinate is the half-life of the feature. Middle: the blue curve is the landscape $\lambda(1,\cdot)$. Right: the blue curve is the silhouette $\phi^{(1)}(\cdot)$.

The landscape and silhouette functions can be evaluated over a one-dimensional grid of points tseq using the functions landscape and silhouette. In the following code, we use the persistence diagram from Figure 6.5 to construct the corresponding landscape and silhouette for one-dimensional features (dimension = 1). The option (KK = 1) specifies that we are interested in the 1st landscape function, and (p = 1) is the power of the weights in the definition of the silhouette function.

```
maxscale <- 5
tseq <- seq(0, maxscale, length = 1000)    #domain
Land <- landscape(DiagRips[["diagram"]], dimension = 1, KK = 1, tseq)
Sil <- silhouette(DiagRips[["diagram"]], p = 1, dimension = 1, tseq)
```

The functions landscape and silhouette return real valued vectors, which can be simply plotted with plot(tseq, Land, type = "l"); plot(tseq, Sil, type = "l"). See Figure 6.10.

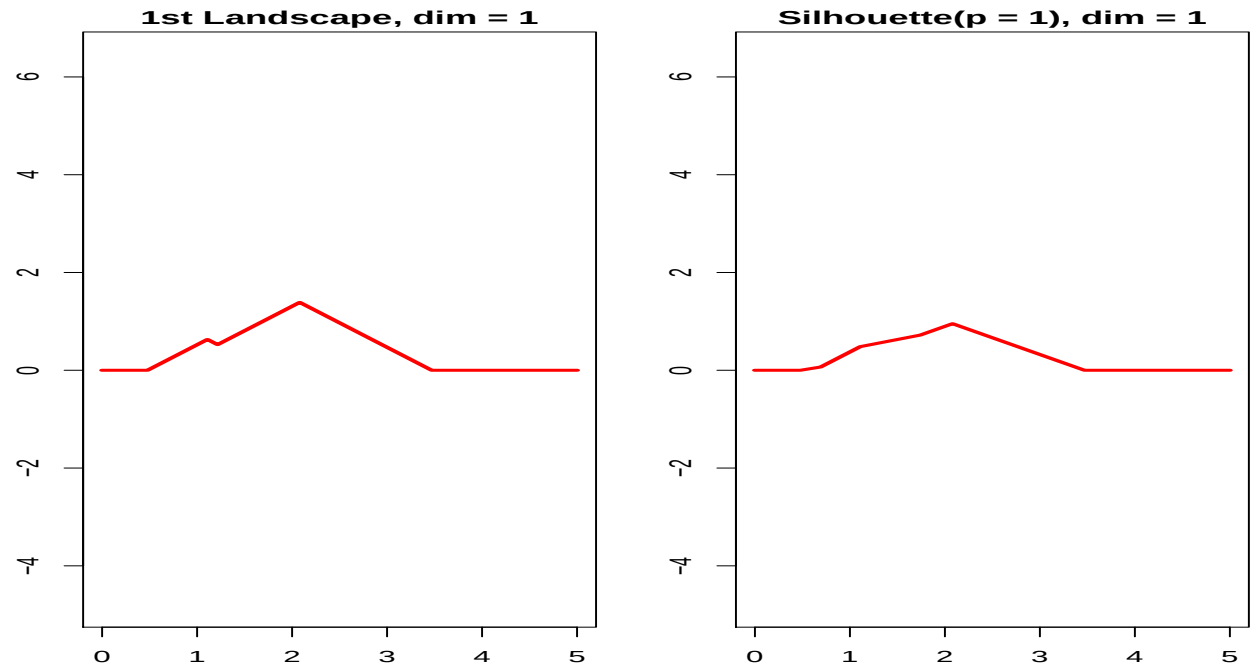


Figure 6.10: Landscape and Silhouette of the one-dimensional features of the diagram of Figure 6.5.

6.2.8 Confidence Bands for Landscapes and Silhouettes

Recent results in Chazal et al. [2014c] and Chazal et al. [2014b] show how to construct confidence bands for landscapes and silhouettes, using a bootstrap algorithm (multiplier bootstrap). This strategy is useful in the following scenario. We have a very large dataset with N points. There is a diagram D and landscape λ corresponding to some filtration built on the data. When N is large, computing D is prohibitive. Instead, we draw n subsamples, each of size m. We compute a diagram and a landscape for each subsample yielding landscapes $\lambda_1, \ldots, \lambda_n$. (Assuming m is much smaller than N, these subsamples are essentially independent and identically distributed.) Then we compute $\frac{1}{n}\sum_i \lambda_i$, an estimate of $\mathbb{E}(\lambda_i)$, which can be regarded as an approximation of λ. The function multipBootstrap uses the landscapes $\lambda_1, \ldots, \lambda_n$ to construct a confidence band for $\mathbb{E}(\lambda_i)$. The same strategy is valid for silhouette functions. We illustrate the method with a simple example.
First we sample N points from two circles:

```
N <- 4000
XX1 <- circleUnif(N / 2)
XX2 <- circleUnif(N / 2, r = 2) + 3
X <- rbind(XX1, XX2)
```

Then we specify the number of subsamples n, the subsample size m, and we create the objects that will store the n diagrams and landscapes:

```
m <- 80      # subsample size
n <- 10      # we will compute n landscapes using subsamples of size m
tseq <- seq(0, maxscale, length = 500)          #domain of landscapes

#here we store n Rips diags
Diags <- list()
#here we store n landscapes
Lands <- matrix(0, nrow = n, ncol = length(tseq))
```

For n times, we subsample from the large point cloud, compute n Rips diagrams and the corresponding 1st landscape functions (KK = 1), using 1 dimensional features (dimension = 1):

```
for (i in seq_len(n)) {
  subX <- X[sample(seq_len(N), m), ]
  Diags[[i]] <- ripsDiag(subX, maxdimension = 1, maxscale = 5)
  Lands[i, ] <- landscape(Diags[[i]][["diagram"]], dimension = 1,
                          KK = 1, tseq)
}
```

Finally we use the n landscapes to construct a 95% confidence band for the mean landscape

```
bootLand <- multipBootstrap(Lands, B = 100, alpha = 0.05,
                            parallel = FALSE)
```

which is plotted by the following code. See Figure 6.11.

```
plot(tseq, bootLand[["mean"]], main = "Mean Landscape with 95% band")
polygon(c(tseq, rev(tseq)),
        c(bootLand[["band"]][, 1], rev(bootLand[["band"]][, 2])),
        col = "pink")
lines(tseq, bootLand[["mean"]], lwd = 2, col = 2)
```

6.2.9 Selection of Smoothing Parameters

An unsolved problem in topological inference is how to choose the smoothing parameters, for example h for KDE and m0 for DTM.

Chazal et al. [2014a] suggest the following method, that we describe here for the kernel density estimator, but works also for the kernel distance and the distance to measure.

Let $\ell_1(h), \ell_2(h), \ldots,$ be the lifetimes of the features of a persistence diagram at scale h. Let $q_\alpha(h)/\sqrt{n}$ be the width of the confidence band for the kernel density estimator at scale h, as described in Section 6.1.1. We define two quantities that measure the amount of significant information at level h:

- The number of significant features, $N(h) = \#\left\{i : \ \ell(i) > 2\frac{q_\alpha(h)}{\sqrt{n}}\right\}$;
- The total significant persistence, $S(h) = \sum_i \left[\ell_i - 2\frac{q_\alpha(h)}{\sqrt{n}}\right]_+$.

These measures are small when h is small since $q_\alpha(h)$ is large. On the other hand, they are small when h is large since then all the features of the KDE are smoothed out. Thus we have a kind of topological

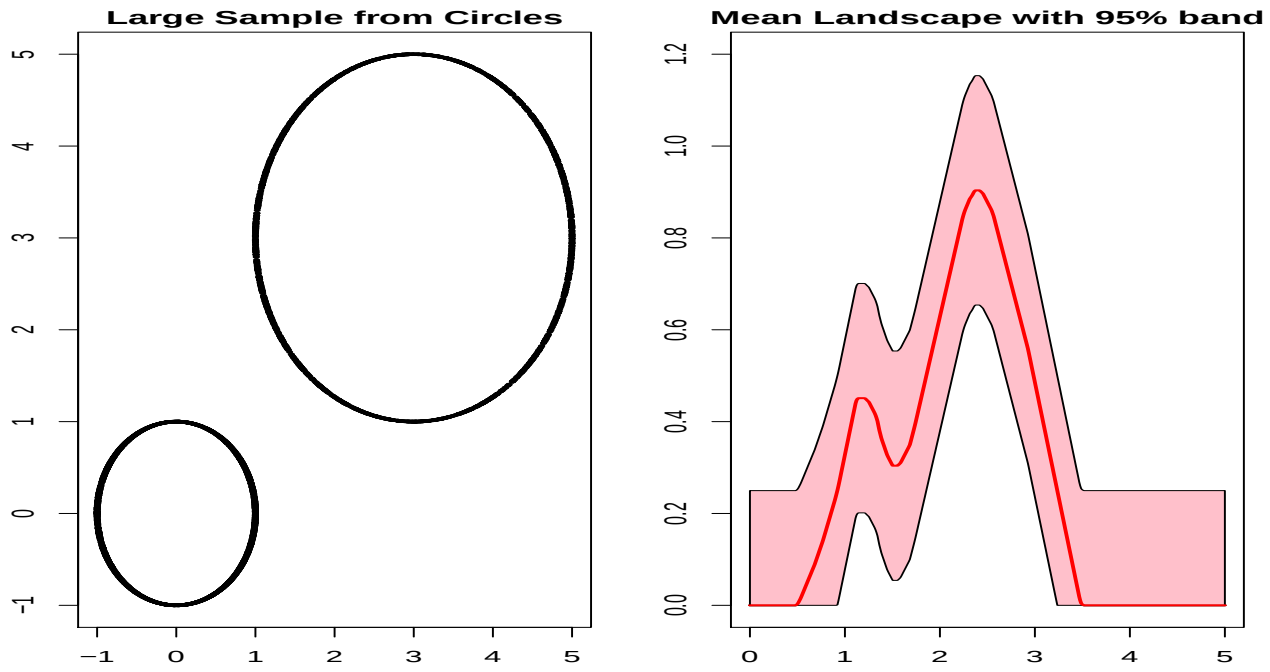


Figure 6.11: 95% confidence band for the mean landscape function.

bias-variance tradeoff. We choose h to maximize $N(h)$ or $S(h)$.

The method is implemented in the function maxPersistence, as shown in the following toy example. First, we sample 1600 point from two circles (plus some clutter noise) and we specify the limits of the grid over which the KDE is evaluated:

```
XX1 <- circleUnif(600)
XX2 <- circleUnif(1000, r = 1.5) + 2.5
noise <- cbind(runif(80, -2, 5), runif(80, -2, 5))
X <- rbind(XX1, XX2, noise)

# Grid limits
Xlim <- c(-2, 5)
Ylim <- c(-2, 5)
by <- 0.2
```

Then we specify a sequence of smoothing parameters among which we will select the optimal one, the number of bootstrap iterations and the level of the confidence bands to be computed:

```
parametersKDE <- seq(0.1, 0.6, by = 0.05)

B <- 50        # number of bootstrap iterations. Should be large.
alpha <- 0.1   # level of the confidence bands
```

The function maxPersistence can be parallelized (parallel = TRUE) and a progress bar can be printed (printProgress = TRUE):

```
maxKDE <- maxPersistence(kde, parametersKDE, X,
                lim = cbind(Xlim, Ylim), by = by, sublevel = FALSE,
```

```
              B = B, alpha = alpha, parallel = TRUE,
              printProgress = TRUE, bandFUN = "bootstrapBand")
## 0   10   20   30   40   50   60   70   80   90   100
## |----|----|----|----|----|----|----|----|----|----|
## **************************************************
```

The S3 methods summary and plot are implemented for the class "maxPersistence". We can display the values of the parameters that maximize the two criteria:

```
print(summary(maxKDE))
## Call:
## maxPersistence(FUN = kde, parameters = parametersKDE, X = X,
##     lim = cbind(Xlim, Ylim), by = by, sublevel = FALSE, B = B,
##     alpha = alpha, bandFUN = "bootstrapBand", parallel = TRUE,
##     printProgress = TRUE)
##
## The number of significant features is maximized by
## [1] 0.25 0.30 0.35
##
## The total significant persistence is maximized by
## [1] 0.15
```

and produce the summary plot of Figure 6.12.

6.3 Density Clustering

The last example of this vignette illustrates the use of the function clusterTree, which is an implementation of Algorithm 1 in Kent et al. [2013].

First, we briefly describe the task of density clustering; we defer the reader to Kent [2013] for a more rigorous and complete description. Let f be the density of the probability distribution P generating the observed sample $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$. For a threshold value $\lambda > 0$, the corresponding super level set of f is $L_f(\lambda) := \text{cl}(\{x \in \mathbb{R}^s : f(x) > \lambda\})$, and its d-dimensional subsets are called high-density regions. The high-density clusters of P are the maximal connected subsets of $L_f(\lambda)$. By considering all the level sets simultaneously (from $\lambda = 0$ to $\lambda = \infty$), we can record the evolution and the hierarchy of the high-density clusters of P. This naturally leads to the notion of the cluster density tree of P (see, e.g., Hartigan [1981]), defined as the collection of sets $T := \{L_f(\lambda), \lambda \geq 0\}$, which satisfies the tree property: $A, B \in T$ implies that $A \subset B$ or $B \subset A$ or $A \cap B = \emptyset$. We will refer to this construction as the λ-tree. Alternatively, Kent et al. [2013] introduced the α-tree and κ-tree, which facilitate the interpretation of the tree by precisely encoding the probability content of each tree branch rather than the density level. Cluster trees are particularly useful for high dimensional data, whose spatial organization is difficult to represent.

We illustrate the strategy with a simple example. First we generate a 2D point cloud from three (not so well) separated clusters (see top left plot of Figure 6.13):

```
X1 <- cbind(rnorm(300, 1, .8), rnorm(300, 5, 0.8))
X2 <- cbind(rnorm(300, 3.5, .8), rnorm(300, 5, 0.8))
```

```
par(mfrow = c(1, 2), mai = c(0.8, 0.8, 0.35, 0.3))
plot(X, pch = 16, cex = 0.5, main = "Two Circles")
plot(maxKDE, main = "Max Persistence - KDE")
```

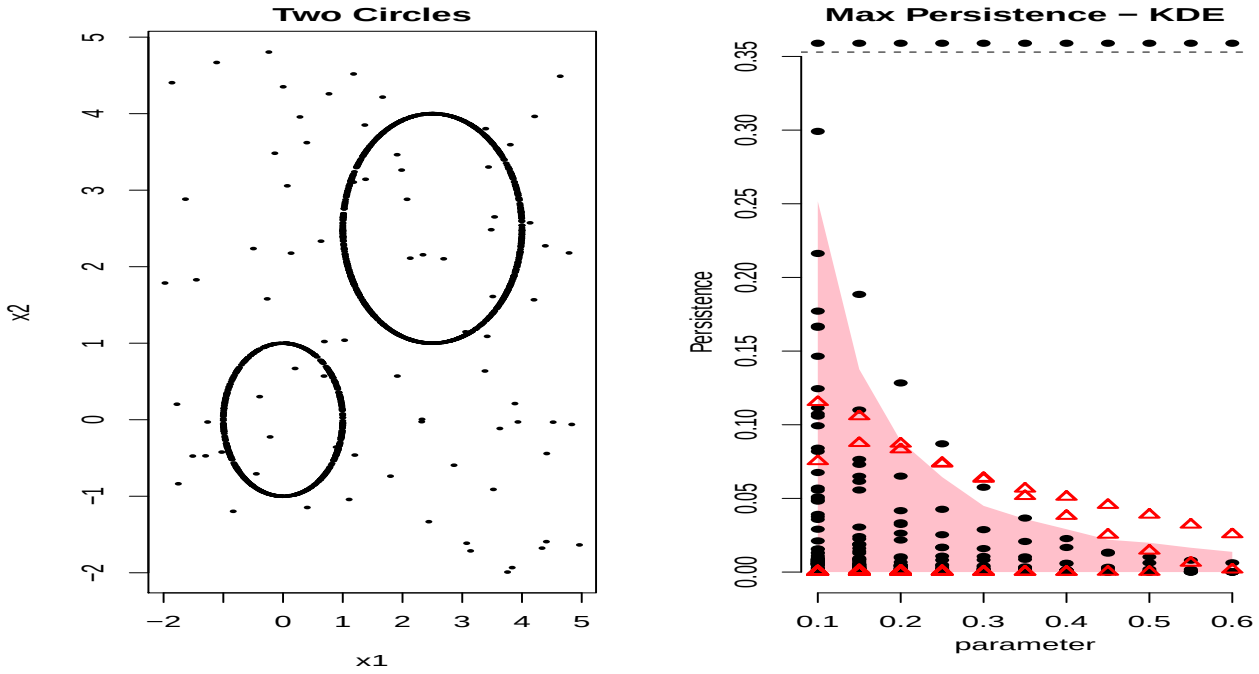


Figure 6.12: Max Persistence Method for the selection of smoothing parameters. For each value of the smoothing parameter we display the persistence of the corresponding homological features, along with a (pink) confidence band that separates the statistically significant features from the topological noise.

```
X3 <- cbind(rnorm(300, 6, 1), rnorm(300, 1, 1))
XX <- rbind(X1, X2, X3)
```

Then we use the function clusterTree to compute cluster trees using the k Nearest Neighbors density estimator (k = 100 nearest neighbors) and the Gaussian kernel density estimator, with smoothing parameter h.

```
Tree <- clusterTree(XX, k = 100, density = "knn",
                    printProgress = FALSE)
TreeKDE <- clusterTree(XX, k = 100, h = 0.3, density = "kde",
                       printProgress = FALSE)
```

Note that, even when kde is used to estimate the density, we have to provide the option (k = 100), so that the algorithm can compute the connected components at each level of the density using a k Nearest Neighbors graph.

The "clusterTree" objects Tree and TreeKDE contain information about the λ-tree, α-tree and κ-tree. The function plot for objects of the class "clusterTree" produces the plots in Figure 6.13.

```
plot(Tree, type = "lambda", main = "lambda Tree (knn)")
plot(Tree, type = "kappa", main = "kappa Tree (knn)")
plot(TreeKDE, type = "lambda", main = "lambda Tree (kde)")
plot(TreeKDE, type = "kappa", main = "kappa Tree (kde)")
```

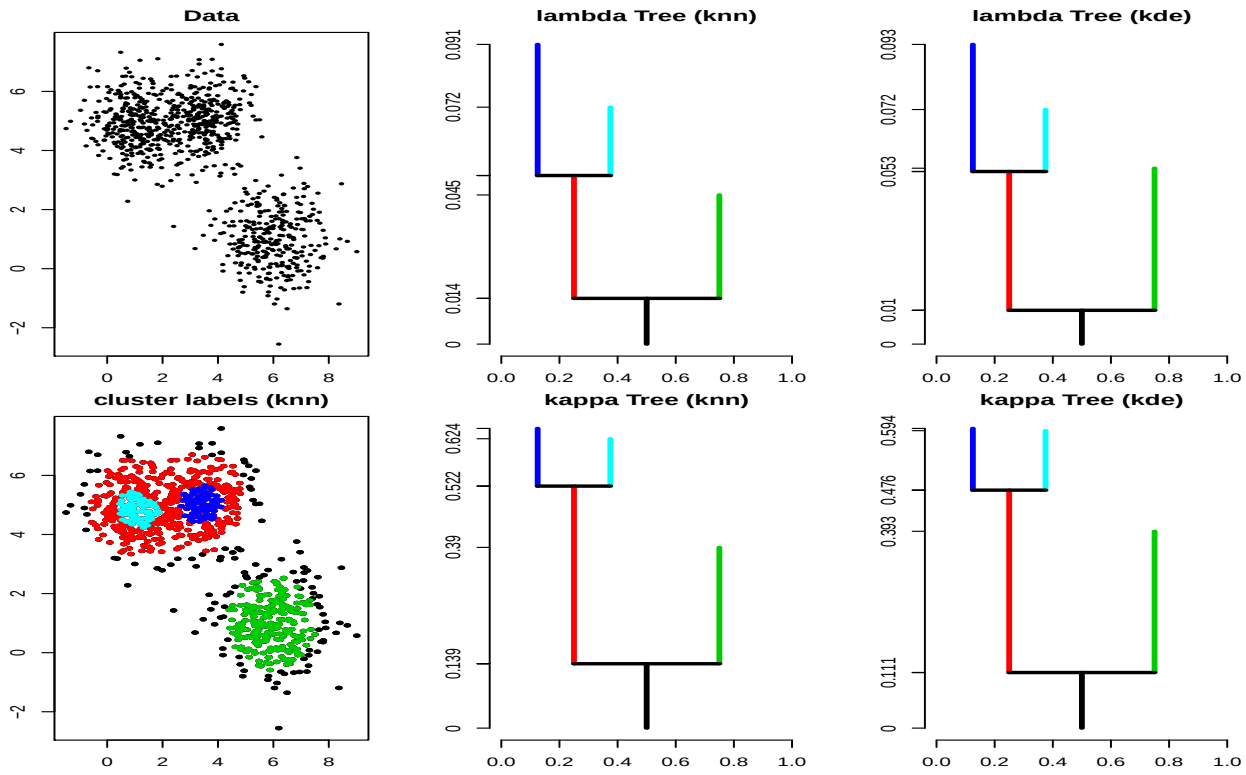


Figure 6.13: The lambda trees and kappa trees of the k Nearest Neighbor density estimator and the kernel density estimator.

Bibliography

E. Aamari and C. Levrard. Stability and Minimax Optimality of Tangential Delaunay Complexes for Manifold Reconstruction. *ArXiv e-prints*, December 2015. 3

Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the Reach of a Manifold. *ArXiv e-prints*, May 2017. 3, A.1, A.3

Stephanie B. Alexander and Richard L. Bishop. Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geom. Dedicata*, 117:65–84, 2006. ISSN 0046-5755. doi: 10.1007/s10711-005-9011-6. URL `http://dx.doi.org/10.1007/s10711-005-9011-6`. B.1.1

E. Arias-Castro, G. Lerman, and T. Zhang. Spectral Clustering Based on Local PCA. *ArXiv e-prints*, January 2013. 3, B.1.1, B.4.2

E. Arias-Castro, B. Pateiro-López, and A. Rodríguez-Casal. Minimax Estimation of the Volume of a Set with Smooth Boundary. *ArXiv e-prints*, May 2016. 3

S. Balakrishnan, S. Narayanan, A. Rinaldo, A. Singh, and L. Wasserman. Cluster Trees on Manifolds. *ArXiv e-prints*, July 2013a. 5

S. Balakrishnan, A. Rinaldo, A. Singh, and L. Wasserman. Tight Lower Bounds for Homology Inference. *ArXiv e-prints*, July 2013b. 3

Ulrich Bauer, Michael Kerber, and Jan Reininghaus. PHAT, a software library for persistent homology, 2012. `https://bitbucket.org/phat-code/phat`. 6

Ulrich Bauer, Elizabeth Munch, and Yusu Wang. Strong equivalence of the interleaving and functional distortion metrics for reeb graphs. In *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34, pages 461–475. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015. 4.2.1, 4.2.2

Richard E. Bellman. *Adaptive Control Processes - A Guided Tour*. Princeton Legacy Library. Princeton University Press, 1961. URL `https://books.google.com/books?id=POAmAAAAMAAJ`. 1, 2

Marcel Berger. *Geometry. II*. Universitext. Springer-Verlag, Berlin, 1987. ISBN 3-540-17015-4. URL `https://doi.org/10.1007/978-3-540-93816-3`. Translated from the French by M. Cole and S. Levy. B.3.2

A. Björner. Handbook of combinatorics (vol. 2). chapter Topological Methods, pages 1819–1872. MIT Press, Cambridge, MA, USA, 1995. D.3

O. Bobrowski, S. Mukherjee, and J. E. Taylor. Topological consistency via kernel estimation. *ArXiv e-prints*, July 2014. 12, 5

Jean-Daniel Boissonnat and Arijit Ghosh. Manifold reconstruction using tangential Delaunay com-

plexes. *Discrete Comput. Geom.*, 51(1):221–267, 2014. ISSN 0179-5376. doi: 10.1007/s00454-013-9557-2. URL http://dx.doi.org/10.1007/s00454-013-9557-2. 3

Martin R. Bridson and André Häfliger. *Metric Spaces of Non-Positive Curvature*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer-Verlag Berlin Heidelberg, 1999. ISBN 978-3-540-64324-1. doi: 10.1007/978-3-662-12494-9. URL https://books.google.com/books?id=3DjaqB08AwAC. A.1

Ryan Remy Brinkman, Maura Gasparetto, Shang-Jung Jessica Lee, Albert J Ribickas, Janelle Perkins, William Janssen, Renee Smiley, and Clay Smith. High-content flow cytometry and temporal data analysis for defining a cellular signature of graft-versus-host disease. *Biology of Blood and Marrow Transplantation*, 13(6):691–700, 2007. 4.4.2

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *arXiv preprint arXiv:1207.6437*, 2012. 6.2.7

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, January 2015. 5

Kevin. Buchin. 2. space-filling curves. In *Organizing Point Sets:Space-Filling Curves, Delaunay Tessellations of Random Point Sets, and Flow Complexes*, chapter 2, pages 5–29. Freien Universität Berlin, 2008. URL http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000003494. 2.2.1, A.2

Benoît Cadre. Kernel estimation of density level sets. *Journal of Multivariate Analysis*, 97(4):999–1023, 2006. 5

Claire Caillerie, Frederic Chazal, Jerome Dedecker, and Bertrand Michel. Deconvolution for the wasserstein metric and geometric inference. *Electron. J. Statist.*, 5:1394–1423, 2011. doi: 10.1214/11-EJS646. URL http://dx.doi.org/10.1214/11-EJS646. 1.4.3

Francesco Camastra and Antonino Staiano. Intrinsic dimension estimation: Advances and open problems. *Inf. Sci.*, 328:26–41, 2016. doi: 10.1016/j.ins.2015.08.029. URL http://dx.doi.org/10.1016/j.ins.2015.08.029. 2

Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for the cluster tree. In *Advances in Neural Information Processing Systems 23*, pages 343–351. 2010. 5

Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246. ACM, 2009. 12, 13, D.1, 111, 113

Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011a. 6.1

Frédéric Chazal, Leonidas J Guibas, Steve Y Oudot, and Primoz Skraba. Scalar field analysis over point cloud data. *Discrete & Computational Geometry*, 46(4):743–775, 2011b. 5

Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. *arXiv preprint arXiv:1207.3674*, 2012. 6.2.1

Frédéric Chazal, Leonidas J Guibas, Steve Y Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):41, 2013. 5

Frédéric Chazal, Brittany T Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance to a measure and kernel distance. *arXiv preprint arXiv:1412.7197*, 2014a. 5.2.1, 5.3, 6, 6.1.1, 6.2.1, 6.2.9, C.7.1

Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Subsampling methods for persistent homology. *arXiv preprint arXiv:1406.1901*, 2014b. 6.2.7, 6.2.8

Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Annual Symposium on Computational Geometry*, pages 474–483. ACM, 2014c. 6, 6.2.7, 6.2.8

Frederic Chazal, Brittany Therese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. On the bootstrap for persistence diagrams and landscapes. *Modeling and Analysis of Information Systems*, 20(6):111, 2014d. 5.2.2

Frédéric Chazal, Pascal Massart, and Bertrand Michel. Rates of convergence for robust geometric inference. *CoRR*, abs/1505.07602, 2015. URL http://arxiv.org/abs/1505.07602. 6.1

Yen-Chi Chen, Christopher R Genovese, and Larry Wasserman. Density level sets: Asymptotics, inference, and visualization. *arXiv:1504.05438*, 2015. 4.2.2, 4.3.1

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, 2016. 4.2.2, 4.3.1, 4.3.1, C.4.1

Moo K. Chung, Peter Bubenik, and Peter T. Kim. Persistence diagrams of cortical surface data. In *Information Processing in Medical Imaging, 21st International Conference, IPMI 2009, Williamsburg, VA, USA, July 5-10, 2009. Proceedings*, pages 386–397, 2009. 5

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. volume 37, pages 103–120, Jan 2007. doi: 10.1007/s00454-006-1276-5. URL https://doi.org/10.1007/s00454-006-1276-5. 13

A. Cuevas, R. Fraiman, and B. Pateiro-López. On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.*, 44(2):311–329, 2012. ISSN 0001-8678. doi: 10.1239/aap/1339878713. URL http://dx.doi.org/10.1239/aap/1339878713. 3

Antonio Cuevas. Set estimation: Another bridge between statistics and geometry. *Bol. Estad. Investig. Oper*, 25(2):71–85, 2009. 5.2.1

Antonio Cuevas and Alberto Rodríguez-Casal. On boundary estimation. *Advances in Applied Probability*, 36(2):340–354, 2004. 5.2.1

Antonio Cuevas, Ricardo Fraiman, and Alberto Rodríguez-Casal. A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.*, 35(3):1031–1051, 2007. ISSN 0090-5364. doi: 10.1214/009053606000001532. URL http://dx.doi.org/10.1214/009053606000001532. 3

Giuseppe De Marco, Gianluca Gorni, and Gaetano Zampieri. Global inversion of functions: an introduction. *NoDEA Nonlinear Differential Equations Appl.*, 1(3):229–248, 1994. ISSN 1021-9722. doi: 10.1007/BF01197748. URL http://dx.doi.org/10.1007/BF01197748. B.4.3

Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. ISBN 0-8176-3490-8. doi: 10.1007/978-1-4757-2201-7. URL http://dx.doi.org/10.1007/978-1-4757-2201-7. Translated from the second Portuguese edition by Francis Flaherty. 2.1, B.1.1, B.4.1

David Donoho. One-sided inference about functionals of a density. *The Annals of Statistics*, 16(4):1390–1420, 1988. 4

Ramsay Dyer, Gert Vegter, and Mathijs Wintraecken. Riemannian simplices and triangulations. *Ge-*

ometriae Dedicata, 179(1):91–138, 2015. ISSN 1572-9168. doi: 10.1007/s10711-015-0069-5. URL http://dx.doi.org/10.1007/s10711-015-0069-5. B.1.1

H. Edelsbrunner and J. Harer. Persistent homology -— a survey. In *Surveys on discrete and computational geometry*, volume 453, page 257. Amer Mathematical Society, 2008. 5

H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. Applied mathematics. American Mathematical Society, 2010. ISBN 9780821849255. URL http://books.google.com/books?id=MDXa6gFRZuIC. 5, 6.2, 6.2.3, 6.2.6

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. 6.2.4

Uwe Einmahl and David M Mason. Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, 33(3):1380–1403, 2005. C.4.1

Justin Eldridge, Mikhail Belkin, and Yusu Wang. Beyond hartigan consistency: Merge distortion metric for hierarchical clustering. In *Proceedings of The 28th Conference on Learning Theory*, volume 40, pages 588–606. PMLR, 2015a. 5

Justin Eldridge, Mikhail Belkin, and Yusu Wang. Beyond hartigan consistency: Merge distortion metric for hierarchical clustering. In *Proceedings of The 28th Conference on Learning Theory*, pages 588–606, 2015b. 4.1, 4.2.1, 4.2.2, C.6.1

Brittany T. Fasy, Jisu Kim, Fabrizio Lecci, Clément Maria, David L. Millman, and Vincent Rouvreau. Introduction to the R package TDA. *CoRR*, abs/1411.1830, 2014a. URL http://arxiv.org/abs/1411.1830. 6

Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *Ann. Statist.*, 42(6):2301–2339, 12 2014b. doi: 10.1214/14-AOS1252. URL http://dx.doi.org/10.1214/14-AOS1252. 1.4.3, 5, 5.2.1, 5.3, 6, 6.1.1, 6.2.1

Herbert Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959. ISSN 0002-9947. 1.3, 3, 3, 3.1, B.2, 95, D.2, D.2, D.2, D.2, D.2

Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969. 3

Stephen Few. Tapping the Power of Visual Perception. 2004. 1

Stephen Few. Data visualization for human perception. In *The Encyclopedia of Human-Computer Interaction*, chapter 35. 2013. 1

Kaspar Fischer. Introduction to alpha shapes, 2005. http://www.cs.uu.nl/docs/vakken/ddm/texts/Delaunay/alphashapes.pdf. 6.2.4

Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *J. Mach. Learn. Res.*, 13:1263–1291, 2012. ISSN 1532-4435. 3

Christopher R Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Nonparametric ridge estimation. *The Annals of Statistics*, 42(4):1511–1545, 2014. C.4.1

John A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition, 1975. 5

John A Hartigan. Consistency of single linkage for high-density clusters. *Journal of the American Statistical Association*, 76(374):388–394, 1981. 5, 6.3

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 14. unsupervised learning. In *The Elements of Statistical Learning*, chapter 14, pages 485–586. Springer-Verlag, 2009. URL http://statweb.stanford.edu/~tibs/ElemStatLearn/. 1, 2

Matthias Hein and Jean-Yves Audibert. Intrinsic dimensionality estimation of submanifolds in $\mathbb{R}^d$. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, pages 289–296. ACM, 2005. ISBN 1-59593-180-5. doi: 10.1145/1102351.1102388. URL http://doi.acm.org/10.1145/1102351.1102388. 2

Hermann Karcher. Riemannian comparison constructions. In *Global differential geometry*, volume 27 of *MAA Stud. Math.*, pages 170–222. Math. Assoc. America, Washington, DC, 1989. B.3.2

Balázs Kégl. Intrinsic dimension estimation using packing numbers, 2003. URL http://papers.nips.cc/paper/2290-intrinsic-dimension-estimation-using-packing-numbers. 2

Brian Kent. *Level Set Trees for Applied Statistics*. PhD thesis, Department of Statistics, Carnegie Mellon University, 2013. 6, 6.3

Brian P Kent, Alessandro Rinaldo, and Timothy Verstynen. Debacl: A python package for interactive density-based clustering. *arXiv preprint arXiv:1307.8136*, 2013. 6.3

Arlene K. H. Kim and Harrison H. Zhou. Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.*, 9(1):1562–1582, 2015. ISSN 1935-7524. doi: 10.1214/15-EJS1039. URL http://dx.doi.org/10.1214/15-EJS1039. 3

Jisu Kim, Yen-Chi Chen, Sivaraman Balakrishnan, Alessandro Rinaldo, and Larry Wasserman. Statistical inference for cluster trees. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 1839–1847. Curran Associates, Inc., 2016. URL http://papers.nips.cc/paper/6508-statistical-inference-for-cluster-trees.pdf. 4, 5

Jisu Kim, Alessandro Rinaldo, and Larry Wasserman. Minimax Rates for Estimating the Dimension of a Manifold. *ArXiv e-prints*, May 2016. 2

Reinhard Klette and Azriel Rosenfeld. *Digital geometry*. Morgan Kaufmann Publishers, San Francisco, CA; Elsevier Science B.V., Amsterdam, 2004. ISBN 1-55860-861-3. Geometric methods for digital picture analysis. 3.1

V. I. Koltchinskii. Empirical geometry of multivariate data: a deconvolution approach. *Ann. Statist.*, 28 (2):591–629, 04 2000. doi: 10.1214/aos/1016218232. URL http://dx.doi.org/10.1214/aos/1016218232. 2

S. Kpotufe and U. von Luxburg. Pruning nearest neighbor cluster trees. In *International Conference on Machine Learning (ICML)*, 2011. 6

John A. Lee and Michel. Verleysen. 1. high-dimensional data. In *Nonlinear Dimensionality Reduction*, chapter 1, pages 1–16. Springer New York, 2007a. URL https://books.google.com/books?id=o_TIoyeO7AsC&dq=isbn:038739351X&source=gbs_navlinks_s. 1, 2

John A. Lee and Michel. Verleysen. 3. estimation of the intrinsic dimension. In *Nonlinear Dimensionality Reduction*, chapter 3, pages 47–68. Springer New York, 2007b. URL https://books.google.com/books?id=o_TIoyeO7AsC&dq=isbn:038739351X&source=gbs_navlinks_s. 2

John Marshall Lee. *Introduction to Topological Manifolds*. Graduate texts in mathematics.

Springer, 2000. ISBN 978-0-3879-5026-6. URL https://books.google.com/books?id=5LqQgkS3--MC. 2.1

John Marshall Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2003. ISBN 978-0-3879-5448-6. URL https://books.google.com/books?id=eqfgZtjQceYC. 2.1

Elizaveta Levina, Peter J Bickel, Elizaveta Levina, and Peter J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems 17 (NIPS 2004)*, pages 777–784, 2004. URL http://papers.nips.cc/paper/2577-maximum-likelihood-estimation-of-intrinsic-dimension. 2

Anna V. Little, Yoon-Mo Jung, and Mauro Maggioni. Multiscale estimation of intrinsic dimensionality of data sets. In *AAAI Fall Symposium: Manifold Learning and Its Applications*, volume FS-09-04 of *AAAI Technical Report*. AAAI, 2009. URL http://aaai.org/ocs/index.php/FSS/FSS09/paper/view/950. 2

Anna V Little, Mauro Maggioni, and Lorenzo Rosasco. Multiscale geometric methods for estimating intrinsic dimension. *Proc. SampTA*, 2011. URL https://services.math.duke.edu/~mauro/Papers/IntrinsicDimensionality_SAMPTA2011.pdf. 2

Yunqian Ma and Yun Fu. *Manifold Learning Theory and Applications*. CRC Press, Inc., 1st edition, 2011. ISBN 978-1-4398-7109-6. URL https://books.google.de/books?id=LjeGZwEACAAJ. A.1

Clément Maria. GUDHI, simplicial complexes and persistent homology packages, 2014. https://project.inria.fr/gudhi/software/. 6

Dmitriy Morozov. Dionysus, a c++ library for computing persistent homology, 2007. http://www.mrzv.org/software/dionysus/. 6

Dmitriy Morozov, Kenes Beketayev, and Gunther Weber. Interleaving distance between merge trees. *Discrete and Computational Geometry*, 49:22–45, 2013. 4.2.1, 4.2.2

James R. Munkres. *Topology: a first course*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975. B.2

Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 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. 1.3, 4, 3, B.1.1, B.2

Peter Petersen. *Riemannian Geometry*. Graduate Texts in Mathematics. Springer New York, 2006. ISBN 978-0-3872-9246-5. doi: 10.1007/978-0-387-29403-2. URL https://books.google.com/books?id=9cekXdo52hEC. 2.1, A.1

Jeff M. Phillips, Bei Wang, and Yan Zheng. Geometric inference on kernel density estimates. *CoRR*, abs/1307.7760, 2013. URL http://arxiv.org/abs/1307.7760. 5

Maxim Raginsky and Svetlana Lazebnik. Estimation of intrinsic dimensionality using high-rate vector quantization. In *Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada]*, pages 1105–1112, 2005. URL http://papers.nips.cc/paper/2945-estimation-of-intrinsic-dimensionality-using-high-rate-vector-quantiza
2

Alessandro Rinaldo and Larry Wasserman. Generalized density clustering. *The Annals of Statistics*, 38

(5):2678–2722, 2010. 5

Vincent Rouvreau. Alpha complex. In *GUDHI User and Reference Manual*. GUDHI Editorial Board, 2015. URL http://gudhi.gforge.inria.fr/doc/latest/group__alpha__complex.html. 6.2.3

Alessandro Rozza, Gabriele Lombardi, Claudio Ceruti, Elena Casiraghi, and Paola Campadelli. Novel high intrinsic dimensionality estimators. *Machine learning*, 89(1-2):37–65, 2012. ISSN 1573-0565. doi: 10.1007/s10994-012-5294-7. URL http://dx.doi.org/10.1007/s10994-012-5294-7. 2

David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015. C.4.1, C.7.1

Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986. 4.4.1, 4.4.2

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. 1.3, 3

Kumar Sricharan, Raviv Raich, and Alfred O. Hero III. Optimized intrinsic dimension estimator using nearest neighbor graphs. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5418–5421. IEEE, 2010. ISBN 978-1-4244-4296-6. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5494931. 2

J. Michael Steele. 2. concentration of measure and the classical theorems. In *Probability Theory and Combinatorial Optimization*, chapter 2, pages 27–51. Society for Industrial and Applied Mathematics, 1997. doi: 10.1137/1.9781611970029.ch2. URL http://epubs.siam.org/doi/abs/10.1137/1.9781611970029.ch2. 2.2.1

Christoph Thäle. 50 years sets with positive reach—a survey. *Surv. Math. Appl.*, 3:123–165, 2008. ISSN 1843-7265. 3

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 1st edition, 2008. ISBN 978-0-3877-9051-0. URL https://books.google.com/books?id=mwB8rUBsbqoC. 1.1, 1.1

Larry Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006. C.4.1, C.7.1

Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media, 2010. ISBN 1441923225, 9781441923226. 4

Jon Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media, 2013. 4.2.2, 106, C.3

Bin Yu. Assouad, fano, and le cam. In David Pollard, Erik Torgersen, and GraceL. Yang, editors, *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pages 423–435. Springer New York, 1997. ISBN 978-1-4612-1880-7. doi: 10.1007/978-1-4612-1880-7_29. URL http://dx.doi.org/10.1007/978-1-4612-1880-7_29. 1.1, 3.3

Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33 (2):249–274, February 2005. 5

Appendix A

Appendix for Chapter 2

A.1 Proofs for Section 2.1

Lemma 16. *Fix $\tau_g,\ \tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, with $\tau_g \le \tau_\ell$. For $M \in \mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$ and $r \in (0,\tau_g)$, let $M_r := \{x \in \mathbb{R}^m :\ dist_{\mathbb{R}^m}(x,M) < r\}$ be a r-neighborhood of M in $\mathbb{R}^m$. Then, the volume of M is upper bounded as*

$$\begin{aligned} vol_M(M) &\le \frac{m!}{d!} r^{d-m} vol_{\mathbb{R}^m}(M_r) \\ &\le C^{(16)}_{K_I,d,m}\left(1+\tau_g^{d-m}\right), \end{aligned} \tag{A.1}$$

where $C^{(16)}_{K_I,d,m}$ is a constant depending only on K_I, d and m.

Proof of Lemma 16. Suppose $\{A_1, \ldots, A_l\}$ is a disjoint cover of M, i.e. measurable subsets of M such that $A_i \cap A_j = \emptyset$, $\bigcup_{i=1}^{l} A_i = M$, and each A_i is equipped with chart maps $\varphi^{(i)} : U_i \subset \mathbb{R}^d \to A_i$. Such a triangulation is always possible. For each A_i, define $M_r^{(i)} := \{x \in \mathbb{R}^m :\ \pi_M(x) \in A_i,\ dist_{\mathbb{R}^m,||\cdot||_1}(x,M) \le r\}$ so that each A_i is a projection of $M_r^{(i)}$ on M, as in Figure A.1. Then,

$$vol_{\mathbb{R}^m}(M_r) = \sum_{i=1}^{l} vol_{\mathbb{R}^m}(M_r^{(i)}). \tag{A.2}$$

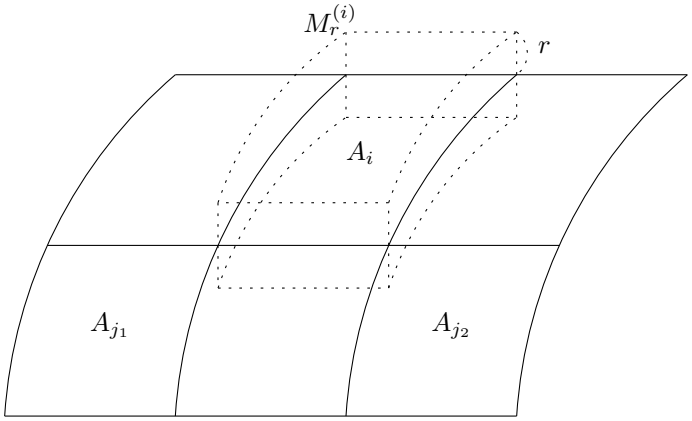


Figure A.1: $\{A_1, \ldots, A_l\}$ is a disjoint cover of M, and each A_i is a projection of $M_r^{(i)}$ on M.

Fix $i \in \{1, \ldots, l\}$. Then for each $u \in U_i$, there exists a linear isometry $R^{(i)}(u) : \mathbb{R}^{m-d} \to (T_{\varphi^{(i)}(u)}M)^{\perp}$, which can be identified as an $m \times (m-d)$ matrix with j^{th} column being $R^{(i,j)}(u)$, so that $M_r^{(i)}$ can be parametrized as $\psi^{(i)} : U_i \times \mathbb{B}_{\mathbb{R}^{m-d}, \|\cdot\|_1}(0, r) \to M_r^{(i)}$ with

$$\psi^{(i)}(u, t) = \varphi^{(i)}(u) + R^{(i)}(u)t = \varphi^{(i)}(u) + \sum_{j=1}^{m-d} t_j R^{(i,j)}(u). \tag{A.3}$$

Then, because $R^{(i)}$ is an isometry,

$$R^{(i)}(u)^{\top} R^{(i)}(u) = I_{m-d}. \tag{A.4}$$

Let $\psi_u^{(i)} = \frac{\partial \psi^{(i)}}{\partial u} = \left(\frac{\partial \psi^{(i)}}{\partial u_1}, \ldots, \frac{\partial \psi^{(i)}}{\partial u_d} \right) \in \mathbb{R}^{m \times d}$ be the partial derivative of $\psi^{(i)}$ with respect to u and let $\psi_t^{(i)} = \frac{\partial \psi^{(i)}}{\partial t}$ be the partial derivative of $\psi^{(i)}$ with respect to t. Define $\varphi_u^{(i)}$ and $R_u^{(i,j)}$ similarly. Then, since $R^{(i)}$ is an isometry, $image(R^{(i)}(u)) = (T_{\varphi^{(i)}(u)}M)^{\perp}$ holds, and hence

$$R^{(i)}(u)^{\top} \varphi_u^{(i)}(u) = 0. \tag{A.5}$$

Also by differentiating (A.4), for all j,

$$R_u^{(i,j)}(u)^{\top} R^{(i)}(u) = 0. \tag{A.6}$$

Also by differentiating (A.3), we get

$$\psi_u^{(i)}(u, t) = \varphi_u^{(i)}(u) + \sum_{j=1}^{m-d} t_j R_u^{(i,j)}(u), \tag{A.7}$$

and

$$\psi_t^{(i)}(u, t) = R^{(i)}(u). \tag{A.8}$$

Hence by multiplying (A.7) and (A.8), and by applying (A.4), (A.5), and (A.6), we get

$$\psi_t^{(i)}(u, t)^{\top} \psi_u^{(i)}(u, t) = R^{(i)}(u)^{\top} \varphi_u^{(i)}(u) + R^{(i)}(u)^{\top} R_u^{(i)}(u)t = 0, \tag{A.9}$$

and

$$\psi_t^{(i)}(u, t)^{\top} \psi_t^{(i)}(u, t) = R^{(i)}(u)^{\top} R^{(i)}(u) = I_{m-d}. \tag{A.10}$$

Now let's consider $\psi_u^{(i)}(u, t)^{\top} \psi_u^{(i)}(u, t)$. From (A.6) and $image(R^{(i)}(u)) = (T_{\varphi^{(i)}(u)}M)^{\perp}$, column space generated by $R_u^{(i,j)}(u)$ is contained in $T_{\varphi^{(i)}(u)}M$, i.e.

$$\left\langle R_u^{(i,j)}(u) \right\rangle \subset T_{\varphi^{(i)}(u)}(M) = span(\varphi_u^{(i)}(u)).$$

Therefore, there exists $\Lambda^{(i,j)}(u) : d \times d$ matrix such that

$$R_u^{(i,j)}(u) = \varphi_u^{(i)}(u) \Lambda^{(i,j)}(u).$$

Then by applying this to (A.7),

$$\psi_u^{(i)}(u, t) = \varphi_u^{(i)}(u) \left(I + \sum_{j=1}^{m-d} t_j \Lambda^{(i,j)}(u) \right). \tag{A.11}$$

Now M being of global reach $\geq \tau_g$ implies $\psi_u^{(i)}(u,t)$ is of full rank for all $t \in \mathbb{B}_{\mathbb{R}^{m-d},\|\cdot\|_1}(0,\tau_g)$. From (A.11), this implies $I + \sum_{j=1}^{m-d} t_j \Lambda^{(i,j)}(u)$ is invertible for all $t \in \mathbb{B}_{\mathbb{R}^{m-d},\|\cdot\|_1}(0,\tau_g)$, and this implies all singular values of $\Lambda^{(i,j)}(u)$ are bounded by $\frac{1}{\tau_g}$. Hence for all $v \in \mathbb{R}^d$,

$$\left|v^\top \Lambda^{(i,j)}(u) v\right| \leq \frac{\|v\|_2^2}{\tau_g},$$

and accordingly,

$$\begin{aligned}\left|v^\top \left(I + \sum_{j=1}^{m-d} t_j \Lambda^{(i,j)}(u)\right) v\right| &\geq \|v\|_2^2 - \sum_{j=1}^{m-d} |t_j| \left|v^\top \Lambda^{(i,j)}(u) v\right| \\ &\geq \left(1 - \frac{\|t\|_1}{\tau_g}\right) \|v\|_2^2.\end{aligned}$$

Hence any singular values σ of $I + \sum_{j=1}^{m-d} t_j \Lambda^{(i,j)}(u)$ satisfies $|\sigma| \geq 1 - \frac{\|t\|_1}{\tau_g}$. And since $\|t\|_1 \leq \tau_g$,

$$\left|I + \sum_{j=1}^{m-d} t_j \Lambda^{(i,j)}(u)\right| \geq \left(1 - \frac{\|t\|_1}{\tau_g}\right)^d.$$

By applying this result to (A.11), the determinant of $\psi_u^{(i)}(u,t)^\top \psi_u^{(i)}(u,t)$ is lower bounded as

$$\begin{aligned}\left|\psi_u^{(i)}(u,t)^\top \psi_u^{(i)}(u,t)\right| &= \left|I + \sum_{j=1}^{m-d} t_j \Lambda^{(i,j)}(u)\right|^2 \left|\varphi_u^{(i)}(u)^\top \varphi_u^{(i)}(u)\right| \\ &\geq \left(1 - \frac{\|t\|_1}{\tau_g}\right)^{2d} \left|\varphi_u^{(i)}(u)^\top \varphi_u^{(i)}(u)\right|. \end{aligned} \tag{A.12}$$

Now, let $g_{ij}^{(M_r)}$ be the Riemannian metric tensor of M_r, and $g_{ij}^{(M)}$ be the Riemannian metric tensor of M. Then from (A.9), (A.10), and (A.12), the determinant of Riemannian metric tensor $g_{ij}^{(M_r)}$ is lower bounded by

$$\begin{aligned}|\det(g_{ij}^{(M_r)})| &= \left|\left(\psi_u^{(i)}(u,t)\ \psi_t^{(i)}(u,t)\right)^\top \left(\psi_u^{(i)}(u,t)\ \psi_t^{(i)}(u,t)\right)\right| \\ &= \begin{vmatrix} \psi_u^{(i)}(u,t)^\top \psi_u^{(i)}(u,t) & \psi_u^{(i)}(u,t)^\top \psi_t^{(i)}(u,t) \\ \psi_u^{(i)}(u,t)^\top \psi_t^{(i)}(u,t) & \psi_t^{(i)}(u,t)^\top \psi_t^{(i)}(u,t) \end{vmatrix} \\ &= \left| \psi_u^{(i)}(u,t)^\top \psi_u^{(i)}(u,t) \right| \\ &\geq \left(1 - \frac{\|t\|_1}{\tau_g}\right)^{2d} \left|\varphi_u^{(i)}(u)^\top \varphi_u^{(i)}(u)\right| \\ &= \left(1 - \frac{\|t\|_1}{\tau_g}\right)^{2d} |\det(g_{ij}^{(M)})|.\end{aligned}$$

And from this, volume of $M_r^{(i)}$ is lower bounded as

$$\begin{aligned}
vol_{\mathbb{R}^m}(M_r^{(i)}) &= \int_{U_i \times \mathbb{B}_{\mathbb{R}^m, \|\cdot\|_1}(0,r)} \sqrt{|\det(g_{ij}^{(M_r)})|} du dt \\
&\geq \int_{U_i} \int_{\mathbb{B}_{\mathbb{R}^m, \|\cdot\|_1}(0,r)} (1 - \|t\|_1 \kappa_g)^d \sqrt{|\det(g_{ij}^{(M)})|} dt du \\
&= vol(U_i) \int_0^r \int_{t_1 + \cdots + t_{m-d-1} \leq s} \left(1 - \frac{s}{\tau_g}\right)^d dt_1 \cdots dt_{m-d-1} ds \\
&= \frac{1}{(m-d-1)!} vol(U_i) \int_0^r s^{m-d-1} \left(1 - \frac{s}{\tau_g}\right)^d ds \\
&= \frac{1}{(m-d-1)!} r^{m-d} vol(U_i) \int_0^1 u^{m-d-1} \left(1 - \frac{r}{\tau_g} u\right)^d du \\
&\geq \frac{1}{(m-d-1)!} r^{m-d} vol(U_i) \int_0^1 u^{m-d-1} (1-u)^d du \\
&= \frac{d!}{m!} r^{m-d} vol(U_i). \qquad \text{(A.13)}
\end{aligned}$$

By applying (A.13) to (A.2), we can lower bound volume of M_r as

$$\begin{aligned}
vol_{\mathbb{R}^m}(M_r) &\geq \frac{d!}{m!} r^{m-d} \sum_{i=1}^{l} vol(U_i) \\
&= \frac{d!}{m!} r^{m-d} vol_M(M). \qquad \text{(A.14)}
\end{aligned}$$

Also, with $r = \tau_g$, M_r is contained in τ_g-neighborhood of I, hence

$$vol_{\mathbb{R}^m}(M_r) \leq 2^m (K_I + \tau_g)^m. \qquad \text{(A.15)}$$

By combining (A.14) and (A.15), we get the desired upper bound of $vol_M(M)$ in (A.1) as

$$\begin{aligned}
vol_M(M) &\leq \frac{m!}{d!} r^{d-m} vol_{\mathbb{R}^m}(M_r) \\
&\leq C_{K_I,d,m}^{(16)} \left(1 + \tau_g^{d-m}\right),
\end{aligned}$$

where $C_{K_I,d,m}^{(16)} \in (0, \infty)$ is a constant depending only on K_I, d and m.

□

Lemma 17. ***Fix*** τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, ***with*** $\tau_g \leq \tau_\ell$. ***Let*** $M \in \mathcal{M}^d_{\tau_g, \tau_\ell, K_I, K_v}$ ***and*** $r \in (0, 2\sqrt{3}\tau_g]$. ***Then*** M ***can be covered by*** N ***radius*** r ***balls*** $\mathbb{B}_M(p_1, r)$, ..., $\mathbb{B}_M(p_N, r)$, ***with***

$$N \leq \left\lfloor \frac{2^d vol(M)}{K_v r^d \omega_d} \right\rfloor. \qquad \text{(A.16)}$$

Proof of Lemma 17. We follow the strategy in [Ma and Fu, 2011, 4.3.1. Lemma 3].
Consider a maximal family of disjoint balls $\left\{\mathbb{B}_M(p_1, \frac{r}{2}), \ldots, \mathbb{B}_M(p_N, \frac{r}{2})\right\}$, i.e. $\mathbb{B}_M(p_i, \frac{r}{2}) \cap \mathbb{B}_M(p_j, \frac{r}{2}) =$

$\emptyset$ for $i \neq j$ and for all $q \in M$, there exists $i \in [1, N]$ such that $\mathbb{B}_M(q, \frac{r}{2}) \cap \mathbb{B}_M(p_i, \frac{r}{2}) \neq \emptyset$. Then $\|q - p_i\|_2 < r$ holds, so $\{\mathbb{B}_M(p_1, r), \ldots, \mathbb{B}_M(p_N, r)\}$ covers M. Now, note that $\mathbb{B}_M(p_i, \frac{r}{2})$ are disjoint, and hence

$$\sum_{i=1}^{N} vol(\mathbb{B}_M(p_i, \frac{r}{2})) \leq vol(M). \tag{A.17}$$

Then since $\frac{r}{2} \leq \sqrt{3}\tau_g$, condition (4) in Definition 15 implies $vol(\mathbb{B}_M(p_i, \frac{r}{2})) \geq K_v 2^{-d} r^d \omega_d$ for all i, hence applying this to (A.17) yields

$$N \leq \frac{2^d vol(M)}{K_v r^d \omega_d},$$

hence M can be covered by N radius r balls with N satisfying (A.16). □

Lemma 81. *(Toponogov comparison theorem, 1959) Let (M, g) be a complete Riemannian manifold with sectional curvature$\geq$ κ, and let S_κ be a surface of constant Gaussian curvature κ. Given any geodesic triangle with vertices $p, q, r \in M$ forming an angle α at q, consider a (comparison) triangle with vertices $\bar{p}, \bar{q}, \bar{r} \in S_\kappa$ such that $dist_{S_\kappa}(\bar{p}, \bar{q}) = dist_M(p, q)$, $dist_{S_\kappa}(\bar{r}, \bar{q}) = dist_M(r, q)$, and $\angle \bar{p}\bar{q}\bar{r} = \angle pqr$. Then*

$$dist_M(\bar{p}, \bar{r}) \leq dist_{S_\kappa}(p, r).$$

Proof of Lemma 81. [See Petersen, 2006, Theorem 79, p.339]. Note that for a manifold with boundary, the complete Riemannian manifold condition can be relaxed to requiring the existence of a geodesic path joining p and q whose image lies on $intM$. □

Lemma 82. *(Hyperbolic law of cosines) Let H_κ be a hyperbolic plane whose Gaussian curvature is $-\kappa^2$. Then given a hyperbolic triangle ABC with angles α, β, γ, and side lengths $BC = a$, $CA = b$, and $AB = c$, the following holds:*

$$\cosh(\kappa a) = \cosh(\kappa b)\cosh(\kappa c) - \sinh(\kappa b)\sinh(\kappa c)\cos\alpha.$$

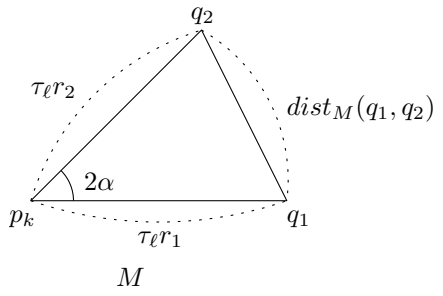
Proof of Lemma 82. [See Bridson and Häfliger, 1999, 2.13 The Law of Cosines in M^n_κ, p.24]. □

Claim 83. Let $\lambda \in [0, 1]$ and let $a, b \in [0, \infty)$ satisfy $a < b$. Then

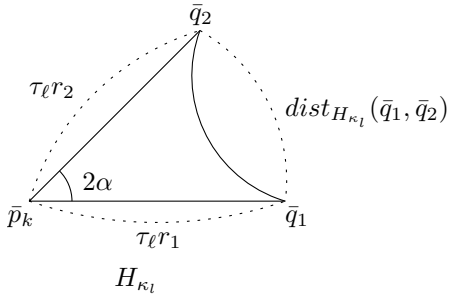
$$\frac{\cosh^{-1}\left((1-\lambda)\cosh a + \lambda \cosh b\right)}{\sqrt{(1-\lambda)a^2 + \lambda b^2}} \leq \frac{\sinh\left(\frac{b}{2}\right)}{b/2}. \tag{A.18}$$

Proof of Claim 83. Consider functions $F, G : [0, \infty)^2 \times [0, 1] \to \mathbb{R}$ defined as $F(a, b, \lambda) = f^{-1}((1 - \lambda)f(a) + \lambda f(b))$ and $G(a, b, \lambda) = g^{-1}((1 - \lambda)g(a) + \lambda g(b))$, for $0 \leq a < b$, $\lambda \in [0, 1]$, $f(t) = \cosh t$, and $g(t) = t^2$. Toponogov comparison theorem in Lemma 81 implies $F(a, b, \lambda) \geq G(a, b, \lambda)$, and f and g being strictly increasing function implies $a < G(a, b, \lambda) \leq F(a, b, \lambda) < b$. Also differentiating log fraction $\frac{\partial}{\partial a} \log \frac{F(a,b,\lambda)}{G(a,b,\lambda)}$ gives

$$\begin{aligned}
\frac{\partial}{\partial a} \log \frac{F(a, b, \lambda)}{G(a, b, \lambda)} &= \frac{(1-\lambda)f'(a)}{f'(F(a, b, \lambda))F(a, b, \lambda)} - \frac{(1-\lambda)g'(a)}{g'(G(a, b, \lambda))G(a, b, \lambda)} \\
&= \frac{1-\lambda}{F(a, b, \lambda)} \exp\left(-\int_a^{F(a,b,\lambda)} (\log f')'(t)dt\right) \\
&\quad - \frac{1-\lambda}{G(a, b, \lambda)} \exp\left(-\int_a^{G(a,b,\lambda)} (\log g')'(t)dt\right).
\end{aligned} \tag{A.19}$$



(a) triangle $\triangle p_k q_1 q_2$ in M



(b) comparison triangle $\triangle \bar{p}_k \bar{q}_1 \bar{q}_2$ in H_{κ_ℓ}

Figure A.2: a triangle $\triangle p_k q_1 q_2$ in M formed by p_k, q_1, q_2, and b its comparison triangle $\triangle \bar{p}_k \bar{q}_1 \bar{q}_2$ in H_{κ_ℓ}.

Then by applying $(\log f')'(t) = \coth t > \frac{1}{t} = (\log g')'(t)$ and $F(a, b, \lambda) \geq G(a, b, \lambda)$ to (A.19) implies

$$0 < \forall a < b, \ \frac{\partial}{\partial a} \log \frac{F(a, b, \lambda)}{G(a, b, \lambda)} < 0,$$

and hence

$$\frac{F(a, b, \lambda)}{G(a, b, \lambda)} \leq \frac{F(0, b, \lambda)}{G(0, b, \lambda)}.$$

By expanding F and G from this, we get

$$\begin{aligned} \frac{\cosh^{-1}\left((1-\lambda)\cosh a + \lambda \cosh b\right)}{\sqrt{(1-\lambda)a^2 + \lambda b^2}} &\leq \frac{\cosh^{-1}\left(\lambda \cosh b + (1-\lambda)\right)}{\sqrt{\lambda b^2}} \\ &= \frac{\cosh^{-1}\left(1 + 2\lambda \sinh^2\left(\frac{b}{2}\right)\right)}{b\sqrt{\lambda}} \\ &\leq \frac{2\sinh\left(\frac{b}{2}\right)}{b}, \end{aligned}$$

where last line is coming from $1 + x \leq \cosh\sqrt{2x} \Longrightarrow \cosh^{-1}(1 + x) \leq \sqrt{2x}$. Hence we get (A.18). □

Lemma 18. *Fix τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, with $\tau_g \leq \tau_\ell$. Let $M \in \mathcal{M}^d_{\tau_g, \tau_\ell, K_I, K_v}$ and let $\exp_{p_k} : \mathcal{E}_k \subset \mathbb{R}^m \to \mathcal{M}$ be an exponential map, where $\mathcal{E}_k$ is the domain of the exponential map $\exp_{p_k}$ and $T_{p_k}M$ is identified with $\mathbb{R}^d$. For all $v, w \in \mathcal{E}_k$, let $R_k := \max\{||v||, ||w||\}$. Then*

$$\| \exp_{p_k}(v) - \exp_{p_k}(w) \|_{\mathbb{R}^m} \leq \frac{\sinh(\sqrt{2}R_k/\tau_\ell)}{\sqrt{2}R_k/\tau_\ell} \|v - w\|_{\mathbb{R}^d}. \tag{A.20}$$

Proof of Lemma 18. Let $q_1 = \exp_{p_k}(v)$ and $q_2 = \exp_{p_k}(w)$. Let $dist_M(p_k, q_1) = \frac{\tau_\ell}{\sqrt{2}} r_1$, $dist_M(p_k, q_2) = \frac{\tau_\ell}{\sqrt{2}} r_2$, and $\angle q_1 p_k q_2 = 2\alpha$ with $0 \leq \alpha \leq \pi$, as in Figure A.2a. Then

$$\begin{aligned} \|v - w\|_{\mathbb{R}^d} &= \frac{\tau_\ell}{\sqrt{2}} \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos 2\alpha} \\ &= \frac{\tau_\ell}{\sqrt{2}} \sqrt{(r_1 + r_2)^2 \sin^2\alpha + (r_1 - r_2)^2 \cos^2\alpha}. \end{aligned} \tag{A.21}$$

Let $\kappa_\ell := \frac{1}{\tau_\ell}$, $H_{-2\kappa_\ell^2}$ be a surface of constant sectional curvature $-2\kappa_\ell^2$, and let $\bar{p}_k, \bar{q}_1, \bar{q}_2 \in H_{-2\kappa_\ell^2}$ be such that $dist_{H_{-2\kappa_\ell^2}}(\bar{p}_k, \bar{q}_1) = dist_M(p_k, q_1)$, $dist_{H_{-2\kappa_\ell^2}}(\bar{p}_k, \bar{q}_2) = dist_M(p_k, q_2)$, and $\angle \bar{q}_1 \bar{p}_k \bar{q}_2 = \angle q_1 p_k q_2$, so that $\triangle \bar{p}_k \bar{q}_1 \bar{q}_2$ becomes a comparison triangle of $p_k q_1 q_2$, as in Figure A.2b. Then since (sectional curvature of M) $\geq -2\kappa_\ell^2$ by [Aamari et al., 2017, Proposition A.1 (iii)], from the Toponogov comparison theorem in Lemma 81,

$$dist_M(q_1, q_2) \leq dist_{H_{-2\kappa_\ell^2}}(\bar{q}_1, \bar{q}_2). \tag{A.22}$$

Also, by applying the hyperbolic law of cosines in Lemma 82 to comparison triangle $\triangle \bar{p}_k \bar{q}_1 \bar{q}_2$ in Figure A.2a,

$$\begin{aligned}
&\cosh r_1 \cosh r_2 - \sinh r_1 \sinh r_2 \cos 2\alpha \\
&\qquad \cosh(\sqrt{2}\kappa_\ell dist_{H_{\kappa_\ell}}(\bar{q}_1, \bar{q}_2)) = \cosh r_1 \cosh r_2 - \sinh r_1 \sinh r_2 \cos 2\alpha \\
&\qquad\qquad = (\sin^2 \alpha) \cosh(r_1 + r_2) + (\cos^2 \alpha) \cosh(r_1 - r_2).
\end{aligned} \tag{A.23}$$

From (A.21) and (A.23), we can expand the fraction of distances $\frac{dist_{H_{-2\kappa_\ell^2}}(\bar{q}_1, \bar{q}_2)}{\|v - w\|_{\mathbb{R}^d}}$ as

$$\frac{dist_{H_{-2\kappa_\ell^2}}(\bar{q}_1, \bar{q}_2)}{\|v - w\|_{\mathbb{R}^d}} = \frac{\cosh^{-1}\left(\sin^2 \alpha \cosh(r_1 + r_2) + \cos^2 \alpha \cosh(r_1 - r_2)\right)}{\sqrt{(\sin^2 \alpha)(r_1 + r_2)^2 + (\cos^2 \alpha)(r_1 - r_2)^2}}. \tag{A.24}$$

Then we can upper bound the fraction of distances $\frac{dist_{H_{-2\kappa_\ell^2}}(\bar{q}_1, \bar{q}_2)}{\|v - w\|_{\mathbb{R}^d}}$ by plugging in $a = |r_1 - r_2|$, $b = r_1 + r_2$, $\lambda = \sin^2 \alpha$ to Claim 83 implies

$$\frac{\cosh^{-1}\left(\sin^2 \alpha \cosh(r_1 + r_2) + \cos^2 \alpha \cosh(r_1 - r_2)\right)}{\sqrt{(\sin^2 \alpha)(r_1 + r_2)^2 + (\cos^2 \alpha)(r_1 - r_2)^2}} \leq \frac{\sinh\left(\frac{r_1 + r_2}{2}\right)}{(r_1 + r_2)/2}. \tag{A.25}$$

Then since $t \mapsto \frac{\sinh t}{t}$ is an increasing function of t and $\frac{r_1 + r_2}{2} \leq \sqrt{2} R_k / \tau_\ell$, so

$$\frac{\sinh\left(\frac{r_1 + r_2}{2}\right)}{(r_1 + r_2)/2} \leq \frac{\sinh(\sqrt{2} R_k / \tau_\ell)}{\sqrt{2} R_k / \tau_\ell}. \tag{A.26}$$

Combining (A.24), (A.25), and (A.26), we have upper bound of the fraction of distances $\frac{dist_{H_{-2\kappa_\ell^2}}(\bar{q}_1, \bar{q}_2)}{\|v - w\|_{\mathbb{R}^d}}$ uniform over r_1, r_2 as

$$\frac{dist_{H_{-2\kappa_\ell^2}}(\bar{q}_1, \bar{q}_2)}{\|v - w\|_{\mathbb{R}^d}} \leq \frac{\sinh(\sqrt{2} R_k / \tau_\ell)}{\sqrt{2} R_k / \tau_\ell}. \tag{A.27}$$

And finally, combining (A.22) and (A.27), we get desired upper bound of $\| \exp_{p_k}(v) - \exp_{p_k}(w) \|_{\mathbb{R}^m}$ in (A.20) as

$$\begin{aligned}
\| \exp_{p_k}(v) - \exp_{p_k}(w) \|_{\mathbb{R}^m} &\leq dist_M(q_1, q_2) \\
&\leq dist_{H_{-2\kappa_\ell^2}}(\bar{q}_1, \bar{q}_2) \\
&\leq \frac{\sinh(\sqrt{2} R_k / \tau_\ell)}{\sqrt{2} R_k / \tau_\ell} \|v - w\|_{\mathbb{R}^d}.
\end{aligned}$$

□

A.2 Proofs for Section 2.2

Claim 84. Fix $\tau_g,\ \tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, $d_1,\ d_2 \in \mathbb{N}$, with $\tau_g \leq \tau_\ell$ and $1 \leq d_1 < d_2 \leq m$. Let $X_1, \ldots, X_n \sim P \in \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}$. Then for all $y \in [0,\infty)$,

$$P^{(n)}\left(||X_n - X_{n-1}||^{d_1}_{\mathbb{R}^m} \leq y | X_1, \ldots, X_{n-1}\right) \leq C^{(84)}_{K_I,K_p,d_2,m}\left(1 + \tau_g^{d_2-m}\right) y^{\frac{d_2}{d_1}}, \tag{A.28}$$

where $C^{(84)}_{K_I,K_p,d_2,m}$ is a constant depending only on K_I, K_p, d_2, m.

Proof of Claim 84. Let p_{X_n} be the pdf of X_n. Then conditional cdf of $||X_n - X_{n-1}||^{d_1}_{\mathbb{R}^m}$ given $X_1, \ldots, X_{n-1}$ is upper bounded by volume of a ball in the manifold M as

$$\begin{aligned}
&P^{(n)}\left(||X_n - X_{n-1}||^{d_1}_{\mathbb{R}^m} \leq y | X_1, \ldots, X_{n-1}\right) \\
&= P^{(n)}\left(X_n \in \mathbb{B}_{\mathbb{R}^m}\left(X_{n-1}, y^{\frac{1}{d_1}}\right) \mid X_1, \ldots, X_{n-1}\right) \\
&= \int_{M \cap \left(\mathbb{B}_{\mathbb{R}^m}\left(X_{n-1}, y^{\frac{1}{d_1}}\right)\right)} p_{X_n}(x_n)\, dvol_M(x_n) \\
&\leq K_p vol_M\left(M \cap B\left(X_{n-1}, y^{\frac{1}{d_1}}\right)\right),
\end{aligned} \tag{A.29}$$

where last inequality is coming from condition (6) in Definition 15. And by applying Lemma 16, $vol_M\left(M \cap B\left(X_{n-1}, y^{\frac{1}{d_1}}\right)\right)$ can be further bounded as

$$\begin{aligned}
&vol_M\left(M \cap B\left(X_{n-1}, y^{\frac{1}{d_1}}\right)\right) \\
&\leq \frac{m!}{d_2!} \min\left\{y^{\frac{1}{d_1}}, \tau_g\right\}^{d_2-m} vol_{\mathbb{R}^m}\left(B\left(X_{n-1},\, y^{\frac{1}{d_1}} + \min\left\{y^{\frac{1}{d_1}}, \tau_g\right\}\right)\right) \text{ (Lemma 16)} \\
&= \frac{m!}{d_2!}\omega_m \left(y^{\frac{d_2}{d_1}} 2^m 1(y^{\frac{1}{d_1}} \leq \tau_g) + y^{\frac{d_2}{d_1}} \left(\frac{\tau_g}{y^{\frac{1}{d_1}}}\right)^{d_2-m} \left(1 + \left(\frac{\tau_g}{y^{\frac{1}{d_1}}}\right)\right)^m 1(y^{\frac{1}{d_1}} > \tau_g) \right) \\
&\leq \frac{m!}{d_2!}\omega_m 2^m \left(y^{\frac{d_2}{d_1}} 1(y^{\frac{1}{d_1}} \leq \tau_g) + y^{\frac{d_2}{d_1}} \left(\frac{\tau_g}{2K_I\sqrt{m}}\right)^{d_2-m} 1(y^{\frac{1}{d_1}} > \tau_g) \right) \\
&\leq C^{(84,1)}_{K_I,d_2,m}\left(1 + \tau_g^{d_2-m}\right) y^{\frac{d_2}{d_1}},
\end{aligned} \tag{A.30}$$

where $C^{(84,1)}_{K_I,d_2,m} = \frac{m!}{d_2!}\omega_m 2^m \left(2K_I\sqrt{m}\right)^{m-d_2}$. By applying (A.29) and (A.30), we get the upper bound on conditional cdf of $||X_n - X_{n-1}||^{d_1}_{\mathbb{R}^m}$ given $X_1, \ldots, X_{n-1}$ in (A.28) as

$$\begin{aligned}
P^{(n)}\left(||X_n - X_{n-1}||^{d_1}_{\mathbb{R}^m} \leq y | X_1, \ldots, X_{n-1}\right) &\leq K_p C^{(84,1)}_{K_I,d_2,m}\left(1 + \tau_g^{d_2-m}\right) y^{\frac{d_2}{d_1}} \\
&\leq C^{(84)}_{K_I,K_p,d_2,m}\left(1 + \tau_g^{d_2-m}\right) y^{\frac{d_2}{d_1}},
\end{aligned} \tag{A.31}$$

where $C^{(84)}_{K_I,K_p,d_2,m} = K_p C^{(84,1)}_{K_I,d_2,m} = \frac{m!}{d_2!} K_p \omega_m 2^m \left(2K_I\sqrt{m}\right)^{m-d_2}$.

□

Lemma 19. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \le \tau_\ell$ and $1 \le d_1 < d_2 \le m$. Let $X_1, \dots, X_n \sim P \in \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}$. Then for all $L > 0$,*

$$P^{(n)}\left[\sum_{i=1}^{n-1}\|X_{i+1}-X_i\|^{d_1} \le L\right] \le \frac{\left(C^{(19)}_{K_I,K_p,d_1,d_2,m}\right)^{n-1} L^{\frac{d_2}{d_1}(n-1)}\left(1+\tau_g^{(d_2-m)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}(n-1)!}, \tag{A.32}$$

where $C^{(19)}_{K_I,K_p,d_1,d_2,m}$ is a constant depending only on K_I, K_p, d_1, d_2, m.

Proof of Lemma 19. Let $Y_i := \|X_{i+1}-X_i\|^{d_1}_{\mathbb{R}^m}$, $i = 1,\dots,n-1$, and let $P^{(n)}_{\sum_{i=1}^{n-2} Y_i}$ be the cumulative distribution function of $\sum_{i=1}^{n-2} Y_i$. Then from Claim 84, probability of the d_1-squared length of the path being bounded by L, $P^{(n)}\left(\sum_{i=1}^{n-1} Y_i \le L\right)$, is upper bounded as

$$\begin{aligned}
&P^{(n)}\left(\sum_{i=1}^{n-1} Y_i \le L\right)\\
&= \int_0^L P^{(n)}\left(Y_{n-1} \le y_{n-1} \middle| \sum_{i=1}^{n-2} Y_i = L - y_{n-1}\right) dP^{(n)}_{\sum_{i=1}^{n-2} Y_i}(L-y_{n-1})\\
&\le C^{(84)}_{K_I,K_p,d_2,m}\left(1+\tau_g^{d_2-m}\right)\int_0^L y_{n-1}^{\frac{d_2}{d_1}} dP^{(n)}_{\sum_{i=1}^{n-2} Y_i}(L-y_{n-1}) \text{ (Claim 84)}\\
&= C^{(84)}_{K_I,K_p,d_2,m}\left(1+\tau_g^{d_2-m}\right)\\
&\quad\times\left(\left[-y_{n-1}^{\frac{d_2}{d_1}} P\left(\sum_{i=1}^{n-2} Y_i \le L - y_{n-1}\right)\right]_0^L + \int_0^L P\left(\sum_{i=1}^{n-2} Y_i \le L - y_{n-1}\right) d\left(y_{n-1}^{\frac{d_2}{d_1}}\right)\right)\\
&= C^{(84)}_{K_I,K_p,d_2,m}\left(1+\tau_g^{d_2-m}\right)\int_0^L P\left(\sum_{i=1}^{n-2} Y_i \le L - y_{n-1}\right)\frac{d_2}{d_1} y_{n-1}^{\frac{d_2-d_1}{d_1}} dy_{n-1}.
\end{aligned}$$

By repeating this argument, we get upper bound of $P^{(n)}\left(\sum_{i=1}^{n-1} Y_i \le L\right)$ as

$$P^{(n)}\left(\sum_{i=1}^{n-1} Y_i \le L\right) \le \left(\frac{d_2}{d_1} C^{(84)}_{K_I,K_p,d_2,m}\left(1+\tau_g^{d_2-m}\right)\right)^{n-1} \int_{\sum_{i=1}^{n-1} y_i \le L} \prod_{i=1}^{n-1} y_i^{\frac{d_2-d_1}{d_1}} dy.$$

From further upper bounding this, we get upper bound of $P^{(n)}\left(\sum_{i=1}^{n-1}\|X_{i+1}-X_i\|^{d_1}_{\mathbb{R}^m} \le L\right)$ in (A.32)

as

$$
\begin{aligned}
&P^{(n)}\left(\sum_{i=1}^{n-1}\|X_{i+1}-X_i\|_{\mathbb{R}^m}^{d_1}\le L\right)\\
&\le\left(\frac{d_2}{d_1}C^{(84)}_{K_I,K_p,d_2,m}\left(1+\tau_g^{d_2-m}\right)\right)^{n-1}\int_{\sum_{i=1}^{n-1}y_i\le L}\prod_{i=1}^{n-1}y_i^{\frac{d_2-d_1}{d_1}}dy\\
&\le\left(\frac{2d_2}{d_1}C^{(84)}_{K_I,K_p,d_2,m}\right)^{n-1}L^{\frac{d_2}{d_1}(n-1)}\left(1+\tau_g^{(d_2-m)(n-1)}\right)\\
&\quad\times\int_{\sum_{i=1}^{n-1}y_i\le 1}\left(\frac{1}{n-1}\sum_{i=1}^{n-1}y_i\right)^{\frac{(d_2-d_1)(n-1)}{d_1}}dy_{n-1}\cdots dy_1\\
&=\frac{\left(C^{(19)}_{K_I,K_p,d_1,d_2,m}\right)^{n-1}L^{\frac{d_2}{d_1}(n-1)}\left(1+\tau_g^{(d_2-m)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}}\\
&\quad\times\int_0^1\int_{\sum_{i=1}^{n-2}y_i\le z}z^{\frac{(d_2-d_1)(n-1)}{d_1}}dy_{n-2}\cdots dy_1dz\\
&=\frac{\left(C^{(19)}_{K_I,K_p,d_1,d_2,m}\right)^{n-1}L^{\frac{d_2}{d_1}(n-1)}\left(1+\tau_g^{(d_2-m)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}(n-2)!}\int_0^1 z^{\frac{d_2(n-1)}{d_1}-1}dz\\
&\le\frac{\left(C^{(19)}_{K_I,K_p,d_1,d_2,m}\right)^{n-1}L^{\frac{d_2}{d_1}(n-1)}\left(1+\tau_g^{(d_2-m)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}(n-1)!},
\end{aligned}
$$

where $C^{(19)}_{K_I,K_p,d_1,d_2,m}=\frac{2d_2}{d_1}C^{(84)}_{K_I,K_p,d_2,m}$.

□

Lemma 85. (Space-filling curve) *There exists a surjective map $\psi_d:\ [0,1]\to[0,1]^d$ which is Hölder continuous of order $1/d$, i.e.*

$$0\le\forall s,t\le 1,\ \|\psi_d(s)-\psi_d(t)\|_{\mathbb{R}^d}\le 2\sqrt{d+3}|s-t|^{1/d}. \tag{A.33}$$

Such a map is called a space-filling curve.

Proof of Lemma 85. [See Buchin, 2008, Chapter 2.1.6]. □

Lemma 20. *Fix τ_g, $\tau_\ell\in(0,\infty]$, $K_I\in[1,\infty)$, $K_v\in(0,2^{-m}]$, $d_1\in\mathbb{N}$, with $\tau_g\le\tau_\ell$. Let $M\in\mathcal{M}^{d_1}_{\tau_g,\tau_\ell,K_p,K_v}$ and $X_1,\dots,X_n\in M$. Then*

$$\min_{\sigma\in S_n}\sum_{i=1}^{n-1}\|X_{\sigma(i+1)}-X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1}\le C^{(20)}_{K_I,K_v,d_1,m}\left(1+\tau_g^{d_1-m}\right), \tag{A.34}$$

where $C^{(20)}_{K_I,K_v,d_1,m}$ is a constant depending only on m, d_1, K_v, and K_I.

Proof of Lemma 20. When $d_1 = 1$, the length of TSP path is bounded by the length of the curve $vol_M(M)$ as in Figure 2.3, and from Lemma 16 we have $vol_M(M) \leq C^{(16)}_{K_I,d,m}\left(1+\tau_g^{1-m}\right)$, hence $C^{(20)}_{K_I,K_v,d_1,m}$ can be set as $C^{(20)}_{K_I,K_v,d_1,m} = C^{(16)}_{K_I,d,m}$, as described before.

Consider $d_1 > 1$. By scaling the space-filling curve in Lemma 85, there exists a surjective map $\psi_{d_1} : [0,1] \to [-r,r]^{d_1}$ and $\psi_m : [0,1] \to [-K_I, K_I]^m$ that satisfies

$$0 \leq \forall s,t \leq 1,\ \|\psi_{d_1}(s) - \psi_{d_1}(t)\|_{\mathbb{R}^{d_1}} \leq 4r\sqrt{d_1+3}|s-t|^{1/d_1} \tag{A.35}$$

$$0 \leq \forall s,t \leq 1,\ \|\psi_m(s) - \psi_m(t)\|_{\mathbb{R}^m} \leq 4K_I\sqrt{m+3}|s-t|^{1/m} \tag{A.36}$$

Let $r := 2\sqrt{3}\tau_g$. From Lemma 17, M can be covered by N balls of radius r, denoted by

$$\mathbb{B}_M(p_1, r),\ \ldots,\ \mathbb{B}_M(p_N, r), \tag{A.37}$$

with $N \leq \left\lfloor \frac{2^{d_1} vol_M(M)}{K_v r^{d_1} \omega_{d_1}} \right\rfloor$. Since $\psi_m : [0,1] \to [-K_I, K_I]^m$ in (A.36) is surjective, we can find a right inverse $\Psi_m : [-K_I, K_I]^m \to [0,1]$ that satisfies $\psi_m(\Psi_m(p)) = p$, i.e.

$$[0,1] \underset{\Psi_m}{\overset{\psi_m}{\rightleftarrows}} [-K_I, K_I]^m. \tag{A.38}$$

Reindex p_k with respect to Ψ_m so that

$$\Psi_m(p_1) < \cdots < \Psi_m(p_N). \tag{A.39}$$

Now fix k, and consider the ball $\mathbb{B}_M(p_k, r)$ in the covering in (A.37). Then for all $p \in \mathbb{B}_M(p_k, r)$, since $d_M(p_k, p) < r$, condition (3) in Definition 15 implies that we can find $\varphi_k(p) \in \mathbb{B}_{\mathbb{R}^{d_1}}(0, r)$ such that $\exp_{p_k}(\varphi_k(p)) = p$. So this shows

$$\mathbb{B}_M(p_k, r) \subset \exp_{p_k}\left(\mathbb{B}_{\mathbb{R}^{d_1}}(0, r)\right).$$

Now consider the composition of the exponential map $\exp_{p_k}$ and ψ_{d_1} in (A.35), $\exp_{p_k} \circ \psi_{d_1} : [0,1] \to M$. Then

$$\mathbb{B}_M(p_k, r) \subset \exp_{p_k}\left(\mathbb{B}_{\mathbb{R}^{d_1}}(0, r)\right) \subset \exp_{p_k}\left([-r,r]^{d_1}\right) = \exp_{p_k} \circ \psi_{d_1}\left([0,1]\right),$$

where last equality is from that ψ_{d_1} in (A.35) is surjective. So $\exp_{p_k} \circ \psi_{d_1} : [0,1] \to M$ is surjective on $\mathbb{B}_M(p, r)$, so we can find right inverse $\Psi_k : \mathbb{B}_M(p_k, r) \to [0,1]$ that satisfies $(\exp_{p_k} \circ \psi_{d_1})(\Psi_k(p)) = p$, i.e.

$$[0,1] \overset{\psi_{d_1}}{\rightleftarrows} [-r,r] \overset{\exp_{p_k}}{\underset{\Psi_k}{\rightleftarrows}} M \supset \mathbb{B}_M(p_k, r). \tag{A.40}$$

Then, reindex $X_1, \ldots, X_n$ with respect to Ψ_m and Ψ_k as $\{X_{k,j}\}_{1 \leq k \leq N,\ 1 \leq j \leq n_k}$, where $X_{k,1}, \ldots, X_{k,n_k} \in \mathbb{B}_M(p_k, r)$ and

$$\Psi_k(X_{k,1}) < \cdots < \Psi_k(X_{k,n_k}). \tag{A.41}$$

Let $\sigma \in S_n$ be corresponding order of index, so that the d_1-squared length of the path $\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|^{d_1}_{\mathbb{R}^m}$ is factorized as

$$\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|^{d_1}_{\mathbb{R}^m} = \sum_{k=1}^{N}\sum_{j=1}^{n_k-1} \|X_{k,j+1} - X_{k,j}\|^{d_1}_{\mathbb{R}^m} + \sum_{k=1}^{N-1} \|X_{k+1,1} - X_{k,n_k}\|^{d_1}_{\mathbb{R}^m}. \tag{A.42}$$

First, consider the first term $\sum_{k=1}^{N}\sum_{j=1}^{n_k-1}\|X_{k,j+1}-X_{k,j}\|_{\mathbb{R}^m}^{d_1}$ in (A.42). For all $1\le k\le N$, by applying Lemma 18, $\sum_{j=1}^{n_k-1}\|X_{k,j+1}-X_{k,j}\|_{\mathbb{R}^m}^{d_1}$ is upper bounded as

$$
\begin{aligned}
&\sum_{j=1}^{n_k-1}\|X_{k,j+1}-X_{k,j}\|_{\mathbb{R}^m}^{d_1}\\
&\le \sum_{j=1}^{n_k-1}\|(\exp_{p_k}\circ\psi_{d_1})(\Psi_k(X_{k,j+1}))-(\exp_{p_k}\circ\psi_{d_1})(\Psi_k(X_{k,j}))\|_{\mathbb{R}^m}^{d_1} \text{ (from (A.40))}\\
&\le \left(\frac{\sinh(\sqrt{2}r/\tau_\ell)}{\sqrt{2}r/\tau_\ell}\right)^{d_1}\sum_{j=1}^{n_k-1}\|\psi_{d_1}(\Psi_k(X_{k,j+1}))-\psi_{d_1}(\Psi_k(X_{k,j}))\|_{\mathbb{R}^{d_1}}^{d_1} \text{ (Lemma 18)}\\
&\le \left(\frac{2\sqrt{2(d_1+3)}\sinh(\sqrt{2}r/\tau_\ell)}{r/\tau_\ell}\right)^{d_1} r^{d_1}\sum_{j=1}^{n_k-1}|\Psi_k(X_{k,j+1})-\Psi_k(X_{k,j})| \text{ (from (A.35))}\\
&\le \left(\frac{2\sqrt{2(d_1+3)}\sinh(\sqrt{2}r/\tau_\ell)}{r/\tau_\ell}\right)^{d_1} r^{d_1} \text{ (from (A.41))}.
\end{aligned}
$$

Then, by applying the fact that $r=2\sqrt{3}\tau_g\le 2\sqrt{3}\tau_\ell$ and that $t\mapsto\frac{\sinh t}{t}$ is increasing function on $t\ge 0$ to this, we have upper bound of $\sum_{j=1}^{n_k-1}\|X_{k,j+1}-X_{k,j}\|_{\mathbb{R}^m}^{d_1}$ as

$$
\sum_{j=1}^{n_k-1}\|X_{k,j+1}-X_{k,j}\|_{\mathbb{R}^m}^{d_1}\le\left(\frac{\sqrt{2(d_1+3)}\sinh 2\sqrt{6}}{\sqrt{3}}\right)^{d_1} r^{d_1}. \tag{A.43}
$$

And then, the second term $\sum_{k=1}^{N-1}\|X_{k+1,1}-X_{k,n_k}\|_{\mathbb{R}^m}^{d_1}$ in (A.42) is upper bounded as

$$
\begin{aligned}
&\sum_{k=1}^{N-1}\|X_{k+1,1}-X_{k,n_k}\|_{\mathbb{R}^m}^{d_1}\\
&\le 3^{d_1-1}\sum_{k=1}^{N-1}\left(\|X_{k+1,1}-p_{k+1}\|_{\mathbb{R}^m}^{d_1}+\|p_{k+1}-p_k\|_{\mathbb{R}^m}^{d_1}+\|p_k-X_{k,n_k}\|_{\mathbb{R}^m}^{d_1}\right)\\
&\le 2\cdot 3^{d_1-1}(N-1)r^{d_1}+3^{d_1-1}\sum_{k=1}^{N-1}\|\psi_m(\Psi_m(p_{k+1}))-\psi_m(\Psi_m(p_k))\|_{\mathbb{R}^{d_1}}^{d_1} \text{ (from (A.38))}\\
&< 3^{d_1}(N-1)r^{d_1}+2\cdot 3^{d_1}\sqrt{m+3}K_I\sum_{k=1}^{N-1}|\Psi_m(p_{k+1})-\Psi_m(p_k)|^{\frac{d_1}{m}} \text{ (from (A.36))}\\
&\le 3^{d_1}(N-1)r^{d_1}+2\cdot 3^{d_1}\sqrt{m+3}K_I\left(\sum_{k=1}^{N-1}|\Psi_m(p_{k+1})-\Psi_m(p_k)|^{\frac{d_1}{m}\times\frac{m}{d_1}}\right)^{\frac{d_1}{m}}\left(\sum_{k=1}^{N-1}1^{\frac{m}{m-d_1}}\right)^{\frac{m-d_1}{m}}\\
&\quad\text{(using Hölder's inequality)}\\
&\le 3^{d_1}(N-1)r^{d_1}+2\cdot 3^{d_1}\sqrt{m+3}K_I(N-1)^{1-\frac{d_1}{m}} \text{ (from (A.39))}.
\end{aligned}
\tag{A.44}
$$

Hence, by plugging in (A.43) and (A.44) to (A.42), $\sum_{i=1}^{n-1}\|X_{\sigma(i+1)}-X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1}$ is upper bounded as

$$\begin{aligned}
&\sum_{i=1}^{n-1}\|X_{\sigma(i+1)}-X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1}\\
&< \left(\left(\frac{\sqrt{2(d_1+3)}\sinh 2\sqrt{6}}{\sqrt{3}}\right)^{d_1}+3^{d_1}\right) r^{d_1}N + 2\cdot 3^{d_1}\sqrt{m+3}K_I N^{1-\frac{d_1}{m}}\\
&< \frac{\left(2\sqrt{d_1+3}\sinh 2\sqrt{6}\right)^{d_1}+6^{d_1}}{K_v\omega_{d_1}} vol_M(M) + \frac{2\cdot 3^{\frac{d_1}{2}}\sqrt{m+3}K_I}{\left(K_v\omega_{d_1}\right)^{1-\frac{d_1}{m}}}\tau_g^{d_1\left(\frac{d_1}{m}-1\right)}\left(vol_M(M)\right)^{1-\frac{d_1}{m}}\\
&\le C^{(20)}_{K_I,K_v,d_1,m}\left(1+\tau_g^{d_1-m}\right),
\end{aligned}$$

by some $C^{(20)}_{K_I,K_v,d_1,m}$ which depends only on m, d_1, K_v, and K_I, where the last line comes from inequality in Lemma 16. Hence we have same upper bound for $\min_{\sigma\in S_n}\sum_{i=1}^{n-1}\|X_{\sigma(i+1)}-X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1}$ as well, as in (A.34). □

Proposition 21. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0,2^{-m}]$, $K_p \in [(2K_I)^m,\infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \le \tau_\ell$ and $1 \le d_1 < d_2 \le m$. Let $\hat{d}_n$ be in* (2.10)*. Then either for $d = d_1$ or $d = d_2$,*

$$\begin{aligned}
&\sup_{P\in\mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right]\\
&\le 1(d=d_2)\left(C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}\right)^n\left(1+\tau_g^{-\left(\frac{d_2}{d_1}m+m-2d_2\right)n}\right)n^{-\left(\frac{d_2}{d_1}-1\right)n},
\end{aligned} \tag{A.45}$$

where $C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m} \in (0,\infty)$ is a constant depending only on $K_I, K_p, K_v, d_1, d_2, m$.

Proof of Proposition 21. Consider first the case $d = d_1$. Then for all $P \in \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p}$ and $X_1,\dots,X_n \sim P$, by Lemma 20,

$$\min_{\sigma\in S_n}\left\{\sum_{i=1}^{n-1}\|X_{\sigma(i+1)}-X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1}\right\} \le C^{(20)}_{K_I,K_v,d_1,m}\left(1+\tau_g^{d_1-m}\right),$$

hence $\hat{d}_n$ in (2.10) always satisfies $\hat{d}_n(X) = d_1 = d(P)$, i.e. the risk of $\hat{d}_n$ satisfies

$$P^{(n)}\left[\hat{d}_n(X_1,\dots,X_n) = d_2\right] = 0. \tag{A.46}$$

For the case when $d = d_2$, for all $P \in \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}$, the risk of $\hat{d}_n$ in (2.10) is upper bounded as

$$
\begin{aligned}
&P^{(n)}\left[\hat{d}_n(X_1,\ldots,X_n) = d_1\right] \\
&= P\left[\bigcup_{\sigma\in S_n}\sum_{i=1}^{n-1}|X_{\sigma(i+1)} - X_{\sigma(i)}| \le C^{(20)}_{K_I,K_v,d_1,m}\left(1+\tau_g^{d_1-m}\right)\right] \\
&\le \sum_{\sigma\in S_n} P\left[\sum_{i=1}^{n-1}|X_{\sigma(i+1)} - X_{\sigma(i)}| \le C^{(20)}_{K_I,K_v,d_1,m}\left(1+\tau_g^{d_1-m}\right)\right] \\
&= n!P\left[\sum_{i=1}^{n-1}|X_{i+1} - X_i| \le C^{(20)}_{K_I,K_v,d_1,m}\left(1+\tau_g^{d_1-m}\right)\right] \\
&= \frac{n\left(C^{(2,2)}_{K_p,d_1,d_2,m}\right)^{n-1}\left(C^{(20)}_{K_I,K_v,d_1,m}\left(1+\tau_g^{d_1-m}\right)\right)^{\frac{d_2}{d_1}(n-1)}\left(1+\tau_g^{(d_2-m)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}},
\end{aligned}
\tag{A.47}
$$

where last line is implied by Lemma 19. Therefore, by combining (A.46) and (A.47), the risk is upper bounded as in (A.45), as

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \\
&\le 1(d=d_2)\frac{n2^{\frac{d_2}{d_1}(n-1)+1}\left(C^{(2,2)}_{K_p,d_1,d_2,m}\left(C^{(20)}_{K_I,K_v,d_1,m}\right)^{\frac{d_2}{d_1}}\right)^{n-1}\left(1+\tau_g^{-\left(\frac{d_2}{d_1}m+m-2d_2\right)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}} \\
&\le 1(d=d_2)\left(C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}\right)^n\left(1+\tau_g^{-\left(\frac{d_2}{d_1}m+m-2d_2\right)n}\right)n^{-\left(\frac{d_2}{d_1}-1\right)n},
\end{aligned}
$$

for some $C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}$ that depends only on $K_I, K_p, K_v, d_1, d_2, m$. □

Proposition 22. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \le \tau_\ell$ and $1 \le d_1 < d_2 \le m$. Then*

$$
\begin{aligned}
&\inf_{\hat{d}_n}\sup_{P\in\mathcal{P}_1\cup\mathcal{P}_2}\mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \\
&\le \left(C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}\right)^n\left(1+\tau_g^{-\left(\frac{d_2}{d_1}m+m-2d_2\right)n}\right)n^{-\left(\frac{d_2}{d_1}-1\right)n},
\end{aligned}
\tag{A.48}
$$

where $C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m}$ is from Proposition 21 and

$$
\mathcal{P}_1 = \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p}, \quad \mathcal{P}_2 = \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}.
$$

Proof of Proposition 22. Applying Proposition 21 to (**??**) yields

$$\begin{aligned}
&\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p} \cup \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{d}_n, d(P) \right) \right] \\
&\leq \sup_{P \in \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p} \cup \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{d}_n, d(P) \right) \right] \\
&\leq \left(C^{(21)}_{K_I,K_p,K_v,d_1,d_2,m} \right)^n \left(1 + \tau_g^{-\left(\frac{d_2}{d_1} m + m - 2d_2 \right) n} \right) n^{-\left(\frac{d_2}{d_1} - 1 \right) n}.
\end{aligned}$$

Hence the minimax rate R_n in (2.5) is upper bounded as in (A.48). □

A.3 Proofs for Section 2.3

Lemma 23. *Fix τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, d, $\Delta d \in \mathbb{N}$, with $\tau_g \leq \tau_\ell$ and $1 \leq d + \Delta d \leq m$. Let $M \in \mathcal{M}^d_{\tau_g,\tau_\ell,K_I,K_v}$ be a d-dimensional manifold of global reach $\geq \tau_g$, local reach $\geq \tau_\ell$, which is embedded in $\mathbb{R}^{m-\Delta d}$. Then*

$$M \times [-K_I, K_I]^{\Delta d} \in \mathcal{M}^{d+\Delta d}_{\tau_g,\tau_\ell,K_I,K_v}, \tag{A.49}$$

which is embedded in $\mathbb{R}^m$.

Proof of Lemma 23. For showing (A.49), we need to show 4 conditions in Definition 15. The other conditions are rather obvious and the critical condition is (2), i.e. global reach condition and local reach condition. Showing the local reach condition is almost identical to showing the global reach condition, so we will focus on the global reach condition. From the definition of reach in Definition 2, we need to show that for all $x \in \mathbb{R}^m$ with $dist_{\mathbb{R}^m}(x, M \times [-K_I, K_I]^{\Delta d}) < \tau_g$, x has unique closest point $\pi_{M \times [-K_I,K_I]^{\Delta d}}(x)$ on $M \times [-K_I, K_I]$.

Let $x \in \mathbb{R}^m$ be satisfying $dist_{\mathbb{R}^m}(x, M \times [-K_I, K_I]^{\Delta d}) < \tau_g$, and let $y \in M \times [-K_I, K_I]^{\Delta d}$. Then the distance between x and y can be factorized as their distance on first $m - \Delta d$ coordinates and last Δd coordinates,

$$\begin{aligned}
&dist_{\mathbb{R}^m}(x,\, y) \\
&= \sqrt{dist_{\mathbb{R}^{m-\Delta d}}(\Pi_{1:m-\Delta d}(x),\, \Pi_{1:m-\Delta d}(y))^2 + dist_{\mathbb{R}^{\Delta d}}\left(\Pi_{(m-\Delta d+1):m}(x),\, \Pi_{(m-\Delta d+1):m}(y)\right)^2}.
\end{aligned} \tag{A.50}$$

For the first term in (A.50), note that the projection map $\Pi_{1:m-\Delta d} : \mathbb{R}^m \to \mathbb{R}^{m-\Delta d}$ is a contraction, i.e. for all $x, y \in \mathbb{R}^m$, $dist_{\mathbb{R}^{m-\Delta d}}(\Pi_{1:m-\Delta d}(x),\, \Pi_{1:m-\Delta d}(y)) \leq dist_{\mathbb{R}^m}(x,\, y)$ holds, so $\Pi_{1:m-\Delta d}(x)$ is also within a τ_g-neighborhood of $M = \Pi_{1:m-\Delta d}(M \times [-K_I, K_I]^{\Delta d})$, i.e.

$$\begin{aligned}
dist_{\mathbb{R}^{m-\Delta d}}(\Pi_{1:m-\Delta d}(x),\, M) &= dist_{\mathbb{R}^{m-\Delta d}}\left(\Pi_{1:m-\Delta d}(x),\, \Pi_{1:m-\Delta d}(M \times [-K_I, K_I]^{\Delta d})\right) \\
&\leq dist_{\mathbb{R}^m}(x,\, M \times [-K_I, K_I]^{\Delta d}) < \tau_g.
\end{aligned}$$

Hence from the definition of the global reach in Definition 2, $\pi_M(\Pi_{1:m-\Delta d}(x)) \in M$ uniquely exists. And from $\Pi_{1:m-\Delta d}(y) \in M$, distance of $\Pi_{1:m-\Delta d}(x)$ and $\Pi_{1:m-\Delta d}(y)$ is lower bounded by the distance of $\Pi_{1:m-\Delta d}(x)$ and M, i.e.

$$\begin{aligned}
dist_{\mathbb{R}^{m-\Delta d}}(\Pi_{1:m-\Delta d}(x),\, \Pi_{1:m-\Delta d}(y)) &\geq dist_{\mathbb{R}^{m-\Delta d}}(\Pi_{1:m-\Delta d}(x),\, \pi_M(\Pi_{1:m-\Delta d}(x))) \\
&= dist_{\mathbb{R}^{m-\Delta d}}(\Pi_{1:m-\Delta d}(x),\, M),
\end{aligned} \tag{A.51}$$

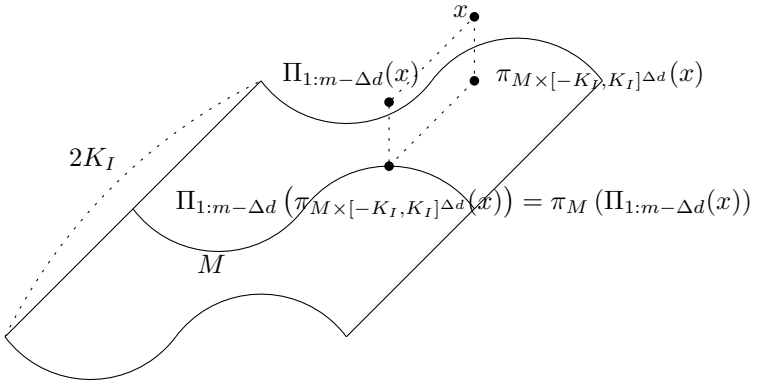


Figure A.3: $\pi_{M\times[-K_I,K_I]^{\Delta d}}(x)$ satisfies $\Pi_{1:m-\Delta d}\left(\pi_{M\times[-K_I,K_I]^{\Delta d}}(x)\right) = \pi_M\left(\Pi_{1:m-\Delta d}(x)\right)$.

and equality holds if and only if $\Pi_{1:m-\Delta d}(y) = \pi_M\left(\Pi_{1:m-\Delta d}(x)\right)$.

The second term in (A.50) is trivially lower bounded by 0, i.e.

$$dist_{\mathbb{R}^{\Delta d}}\left(\Pi_{(m-\Delta d+1):m}(x),\, \Pi_{(m-\Delta d+1):m}(y)\right) \geq 0, \tag{A.52}$$

and equality holds if and only if $\Pi_{(m-\Delta d+1):m}(x) = \Pi_{(m-\Delta d+1):m}(y)$.

Hence by applying (A.51) and (A.52) to (A.50), $dist_{\mathbb{R}^m}(x,\, y)$ is lower bounded by distance of $\Pi_{1:m-\Delta d}(x)$ and M, i.e.

$$\begin{aligned}
&dist_{\mathbb{R}^m}(x,\, y)\\
&= \sqrt{dist_{\mathbb{R}^{m-\Delta d}}\left(\Pi_{1:m-\Delta d}(x),\, \Pi_{1:m-\Delta d}(y)\right)^2 + dist_{\mathbb{R}^{\Delta d}}\left(\Pi_{(m-\Delta d+1):m}(x),\, \Pi_{(m-\Delta d+1):m}(y)\right)^2}\\
&\geq dist_{\mathbb{R}^{m-\Delta d}}\left(\Pi_{1:m-\Delta d}(x),\, M\right),
\end{aligned}$$

and equality holds if and only if $\Pi_{1:m-\Delta d}(y) = \pi_M\left(\Pi_{1:m-\Delta d}(x)\right)$ and $\Pi_{(m-\Delta d+1):m}(x) = \Pi_{(m-\Delta d+1):m}(y)$, i.e. when $y = \left(\pi_M\left(\Pi_{1:m-\Delta d}(x)\right),\, \Pi_{(m-\Delta d+1):m}(x)\right)$. Hence x has unique closest point $\pi_{M\times[-K_I,K_I]^{\Delta d}}(x)$ on $M\times[-K_I,K_I]$ as

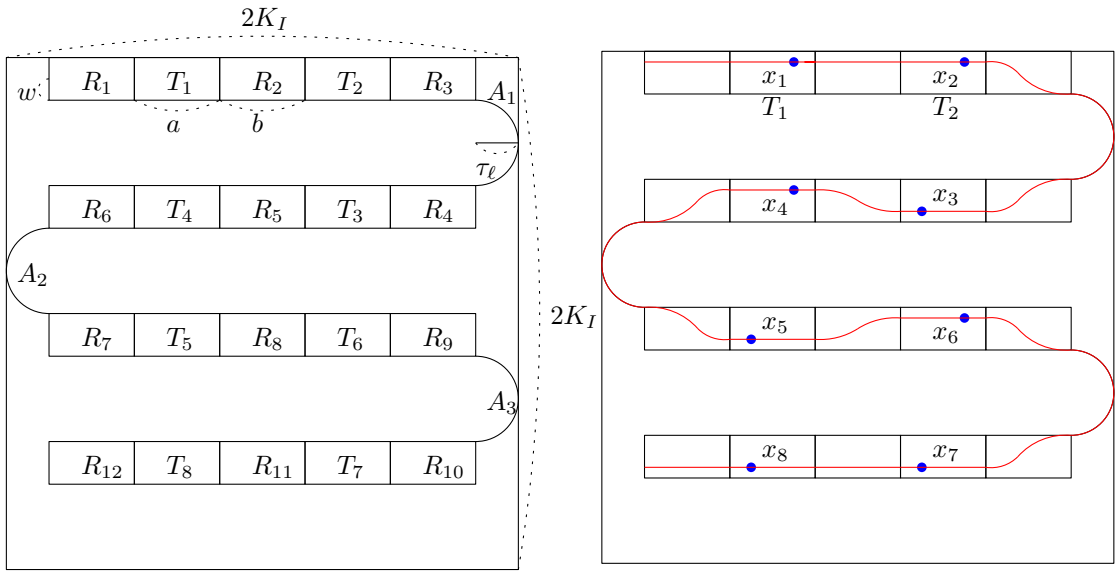
$$\pi_{M\times[-K_I,K_I]^{\Delta d}}(x) = \left(\pi_M\left(\Pi_{1:m-\Delta d}(x)\right),\, \Pi_{(m-\Delta d+1):m}(x)\right),$$

as in Figure A.3. □

Lemma 24. *Fix $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, d_1, $d_2 \in \mathbb{N}$, with $1 \leq d_1 \leq d_2$, and suppose $\tau_\ell < K_I$. Then there exist $T_1, \cdots, T_n \subset [-K_I, K_I]^{d_2}$ such that:*
(1) The T_i's are distinct.
(2) For each T_i, there exists an isometry Φ_i such that

$$T_i = \Phi_i\left([-K_I,K_I]^{d_1-1}\times[0,a]\times\mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0,w)\right), \tag{A.53}$$

where $c = \left\lceil \frac{K_I+\tau_\ell}{2\tau_\ell}\right\rceil$, $a = \frac{K_I-\tau_\ell}{\left(d_2-d_1+\frac{1}{2}\right)\left\lceil\frac{n}{c^{d_2-d_1}}\right\rceil}$, and $w = \min\left\{\tau_\ell,\ \frac{(d_2-d_1)^2(K_I-\tau_\ell)^2}{2\tau_\ell\left(d_2-d_1+\frac{1}{2}\right)^2\left(\left\lceil\frac{n}{c^{d_2-d_1}}\right\rceil+1\right)^2}\right\}$.
(3)There exists $\mathscr{M} : \left(\mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0,w)\right)^n \to \mathcal{M}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v}$ one-to-one such that for each $y_i \in \mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0,w)$, $1\leq i\leq n$, $\mathscr{M}(y_1,\ldots,y_n)\cap T_i = \Phi_i([-K_I,K_I]^{d_1-1}\times[0,a]\times\{y_i\})$. Hence for any $x_1\in T_1,\ldots,x_n\in T_n$, $\mathscr{M}(\{\Pi^{-1}_{(d_1+1):d_2}\Phi_i^{-1}(x_i)\}_{1\leq i\leq n})$ passes through $x_1,\ldots,x_n$.



(a) comparison triangle $\triangle \bar{p}_k \bar{q}_1 \bar{q}_2$ in H_{κ_ℓ}

(b) manifold passing through X_i's

Figure A.4: This figure illustrates the case where $d_1 = 1$ and $d_2 = 2$. shows how T_i, R_i, and A_i's are aligned in a zigzag. a shows for given $x_1 \in T_1, \ldots, x_n \in T_n$ (represented as blue points), how $\mathscr{M}(\{\Pi^{-1}_{(d_1+1):d_2} \Phi_i^{-1}(x_i)\}_{1 \le i \le n})$ (represented as a red curve) passes through $x_1, \ldots, x_n$
.

Proof of Lemma 24. By Lemma 23, we only need to show the case for $d_1 = 1$. This is since for $d_1 > 1$ case, we can build the set of manifolds in $\mathcal{M}^{d_1}_{\tau_g, \tau_\ell, K_I, K_v}$ by forming a Cartesian product of the manifold with the cube as in Lemma 23.

Let $b = \frac{2(d_2-d_1)(K_I-\tau_\ell)}{\left(d_2-d_1+\frac{1}{2}\right)\left(\left\lfloor \frac{n}{c^{d_2-d_1}} \right\rfloor + 1\right)}$, so that

$$b \ge 2\sqrt{2w\tau_\ell} \quad \text{and} \quad 2\tau_\ell + \left\lfloor \frac{n}{c^{d_2-d_1}} \right\rfloor a + \left(\left\lfloor \frac{n}{c^{d_2-d_1}} \right\rfloor + 1 \right) b = 2K_I.$$

With such values of a, b, and w, align T_i, R_i, and A_i in a zigzag way, as in Figure A.4.

Then from the definition of T_i, (1) the T_i's are distinct and (2) for each T_i, there exists an isometry Φ_i such that $T_i = \Phi_i\left([-K_I, K_I]^{d_1-1} \times [0, a] \times \mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0, w)\right)$. There exists isometry Ψ_i such that $R_i = \Psi_i\left([-K_I, K_I]^{d_1-1} \times [0, b] \times \mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0, w)\right)$ as well. Hence condition (1) and (2) are satisfied.

We are left to define $\mathscr{M}$ that satisfies condition (3). Now define a map from a set of points to a set of manifolds $\mathscr{M} : \left(\mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0, w)\right)^n \to \mathcal{M}^{d_1}_{\tau_g, \tau_\ell, K_I, K_v}$ as follows. For each $y_i \in \mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0, w)$, $1 \le i \le n$, $\bigcup_{i=1}^{4} A_i \subset \mathscr{M}(y_1, \ldots, y_n) \subset \left(\bigcup_{i=1}^{4} A_i\right) \bigcup \left(\bigcup_{i=1} T_i\right) \bigcup \left(\bigcup_{i=1} R_i\right)$. The intersection of $\mathscr{M}(y_1, \ldots, y_n)$ and T_i is a line segment $\Phi_i([-K_I, K_I]^{d_1-1} \times [0, a] \times \{y_i\})$, as in Figure A.4a. Our goal is to make $\mathscr{M}(y_1, \ldots, y_n)$ be C^1 and piecewise C^2.

See Figure A.5 for construction of intersection of $\mathscr{M}(y_1, \ldots, y_n)$ and R_i. Given that $\mathscr{M}(y_1, \ldots, y_n) \cap \left(\left(\bigcup_{i=1}^{4} A_i\right) \bigcup \left(\bigcup_{i=1} T_i\right)\right)$ is determined, two points on $\mathscr{M}(y_1, \ldots, y_n) \cap \partial R_i$ is already determined. By translation and rotation if necessary, for all p, q with $-w \le q \le p \le w$, we need to find C^2 curve with reach $\ge \tau_\ell$ that starts from $(0, p) \in \mathbb{R}^2$, ends at $(b, q) \in \mathbb{R}^2$, and velocity at each end points are both parallel to $(1, 0) \in \mathbb{R}^2$, as in Figure A.5.

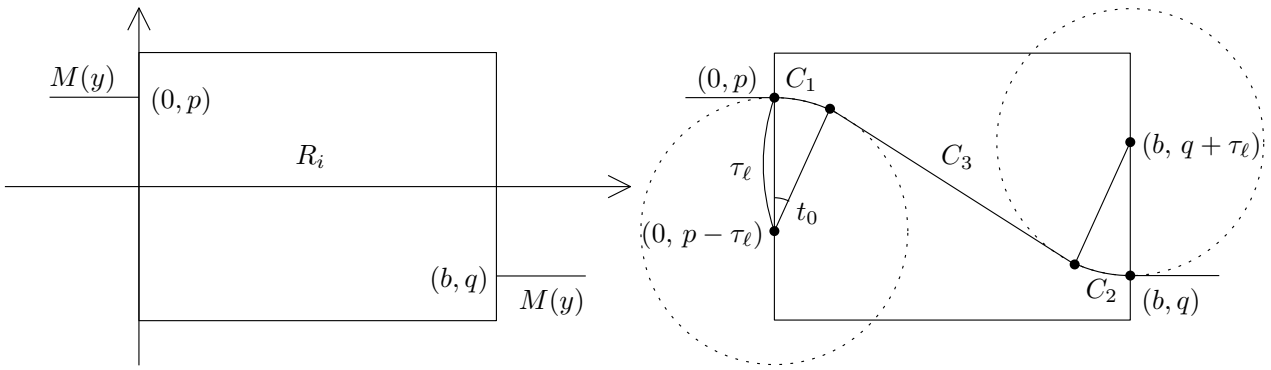


Figure A.5: We need to find C^2 curve with local reach $\geq \tau_\ell$ that starts from $(0,p) \in \mathbb{R}^2$, ends at (b,q), and velocity at each end points are both parallel to $(1,0)$. C_1 and C_2 are arcs of circles of radius R_l, and C_3 is the cotangent segment of two circles.

Let

$$t_0 = \cos^{-1}\left(\frac{2\tau_\ell\left(2\tau_\ell - (p-q)\right) + b\sqrt{b^2 - (p-q)\left(4\tau_\ell - (p-q)\right)}}{b^2 + \left(2\tau_\ell - (p-q)\right)^2}\right), \tag{A.54}$$

and let

$$C_1 = \left\{(0, p-\tau_\ell) + \tau_\ell\left(\sin t, \cos t\right) \mid 0 \leq t \leq t_0\right\}.$$

Then C_1 is an arc of circle of which center is $(0, p-\tau_\ell)$, and starts at $(0,p)$ when $t=0$ and ends at $(\tau_\ell \sin t_0,\ p - \tau_\ell(1-\cos t_0))$ when $t = t_0$. Also, the normalized velocities of C_1 at endpoints are

$$(1,\ 0) \text{ at } (0,\ p), \qquad (\cos t_0,\ -\sin t_0) \text{ at } \left(\tau_\ell \sin t_0,\ p - \tau_\ell(1-\cos t_0)\right). \tag{A.55}$$

Similarly, let

$$C_2 = \left\{(b, q+\tau_\ell) - \tau_\ell\left(\sin t, \cos t\right) \mid 0 \leq t \leq t_0\right\}.$$

Then C_2 is an arc of a circle of whose center is $(b, q+\tau_\ell)$, and starts at (b,q) when $t=0$ and ends at $(b - \tau_\ell \sin t_0,\ q + \tau_\ell\left(1-\cos t_0\right))$ when $t = t_0$. Also, the normalized velocities of C_2 at endpoints are

$$(-1,\ 0) \text{ at } (b,\ q), \qquad (-\cos t_0,\ \sin t_0) \text{ at } \left(b - \tau_\ell \sin t_0,\ q + \tau_\ell\left(1-\cos t_0\right)\right). \tag{A.56}$$

Let

$$C_3 = \Big\{(1-s)\left(\tau_\ell \sin t_0,\ p - \tau_\ell(1-\cos t_0)\right) + s\left(b - \tau_\ell \sin t_0,\ q + \tau_\ell\left(1-\cos t_0\right)\right) \\ \mid 0 \leq s \leq 1\Big\},$$

so that C_3 is a segment joining $(\tau_\ell \sin t_0,\ p - \tau_\ell(1-\cos t_0))$ (when $s=0$) and $(b - \tau_\ell \sin t_0,\ q + \tau_\ell(1-\cos t_0))$ (when $s=1$). Also, its velocity vector is

$$\left(b - \tau_\ell \sin t_0,\ q + \tau_\ell\left(1-\cos t_0\right)\right) \text{ for all } s \in [0,1]. \tag{A.57}$$

Then from definition of t_0 in (A.54),

$$\cos t_0\left(q - p + 2\tau_\ell\left(1-\cos t_0\right)\right) + \sin t_0\left(b - 2\tau_\ell \sin t_0\right) = 0,$$

and this implies that $(b - 2\tau_\ell \sin t_0,\ q - p + 2\tau_\ell (1 - \cos t_0))$ is parallel to $(\cos t_0, -\sin t_0)$. Hence the velocity vector of C_3 in (A.57) is parallel to the velocity vector of C_1 in (A.55) at $(\tau_\ell \sin t_0,\ p - \tau_\ell(1 - \cos t_0))$ and the velocity vector of C_2 in (A.56) at $(b - \tau_\ell \sin t_0,\ q + \tau_\ell(1 - \cos t_0))$, i.e. C_3 is cotangent to both C_1 and C_2. See Figure A.5.

Now we check whether is of global reach $\geq \tau_\ell$, which implies both global reach $\geq \tau_g$ and local reach $\geq \tau_\ell$ since $\tau_g \leq \tau_\ell$. From [Aamari et al., 2017, Theorem 3.4], the reach $\tau(M)$ of a manifold M is realized in either the global case or the local case, where the global case refers to that there exists two points $q_1, q_2 \in M$ with $\mathbb{B}(\frac{q_1+q_2}{2}, \tau(M)) \cap M = \emptyset$, and the local case refers to that there exists an arc-length parametrized geodesic γ such that $||\gamma''(0)||_2 = \frac{1}{\tau(M)}$. Now from the construction, any $q_1, q_2 \in \mathscr{M}(y_1, \ldots, y_n)$ with $\mathbb{B}(\frac{q_1+q_2}{2}, \tau) \cap \mathscr{M}(y_1, \ldots, y_n) = \emptyset$ can only happen when $\tau \geq \tau_\ell$, so it suffices to check whether any arc-length parametrized geodesics γ satisfies $||\gamma''(0)||_2 \leq \frac{1}{\tau_\ell}$. And this is satisfied since $\mathscr{M}(y_1, \ldots, y_n)$ is piecewise either a straight line segment or an arc of a circle of radius τ_ℓ. Hence $\mathscr{M}(y_1, \ldots, y_n)$ is of global reach $\geq \tau_\ell$. □

Claim 25. Let $T = S_n \prod_{i=1}^n T_i$ where the T_i's are from Lemma 24. Let Q_2 be the uniform distribution on $[-K_I, K_I]^{d_2}$, and let $\mathcal{P}_1^{d_1}$ be as in (2.13). Then there exists $Q_1 \in co(\mathcal{P}_1^{d_1})$ satisfying that for all $x \in intT$, there exists $r_x > 0$ such that for all $r < r_x$,

$$Q_1 \left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r) \right) \geq 2^{-n} Q_2 \left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r) \right). \tag{A.58}$$

Proof of Claim 25. Let Q_1 be from (A.63) in Proposition 26. By symmetry, we can assume that $x \in \prod_{i=1}^n T_i$, i.e. $x_1 \in T_1, \ldots, x_n \in T_n$. Choose r_x small enough so that $\mathbb{B}(x, r_x) \subset intT$. Then for all $r < r_x$, from the definition of Q_1 in (A.63),

$$\begin{aligned} Q_1 \left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r) \right) &= \int_{\mathcal{P}_1} P^{(n)} \left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r) \right) d\mu_1(P) \\ &= \int_{C^n} \Phi(y)^{(n)} \left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r) \right) \lambda_{C^n}(y) \\ &= \int_{C^n} \prod_{i=1}^n \lambda_{\mathscr{M}(y)} \left(\mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r) \right) \lambda_{C^n}(y). \end{aligned} \tag{A.59}$$

Then from condition (3) in Lemma 24, $\mathscr{M}(y) \cap T_i = \Phi_i \left([-K_I, K_I]^{d_1 - 1} \times [0, a] \times \{y_i\} \right)$ holds, hence

$$\mathscr{M}(y) \cap \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)$$
$$\begin{cases} = \Phi_i \left(\mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_1},\infty}} \left(\Pi_{1:d_1}(\Phi_i^{-1}(x_i)),\ r \right) \times \{y_i\} \right), & \text{if } \left\| y_i - \Pi_{(d_1+1):d_2}(\Phi_i^{-1}(x_i)) \right\|_{\mathbb{R}^{d_2 - d_1}} < r, \\ \supset \emptyset, & \text{otherwise.} \end{cases}$$

And hence the volume of $\mathscr{M}(y) \cap \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)$ can be lower bounded as

$$\lambda_{\mathscr{M}(y)} \left(\mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r) \right) \geq \frac{r^{d_1}}{2K_I^{d_1 - 1} an} I \left(\left\| y_i - \Pi_{(d_1+1):d_2}(\Phi_i^{-1}(x_i)) \right\|_{\mathbb{R}^{d_2 - d_1},\infty} < r \right).$$

By applying this to (A.59), $Q_1\left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i,r)\right)$ can be lower bounded as

$$\begin{aligned}
&Q_1\left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i,r)\right)\\
&\geq \int_{C^n}\prod_{i=1}^n \frac{r^{d_1}}{2K_I{}^{d_1-1}an} I\left(\left\|y_i-\Pi_{(d_1+1):d_2}(\Phi_i^{-1}(x_i))\right\|_{\mathbb{R}^{d_2-d_1},\infty}<r\right)\lambda_{C^n}(y)\\
&= \frac{r^{d_1 n}}{2^n K_I^{(d_1-1)n}(an)^n}\prod_{i=1}^n\int_C I\left(\left\|y_i-\Pi_{(d_1+1):d_2}(\Phi_i^{-1}(x_i))\right\|_{\mathbb{R}^{d_2-d_1},\infty}<r\right)\lambda_C(y_i)\\
&= \frac{r^{d_1 n}}{2^n K_I^{(d_1-1)n}(an)^n}\left(\frac{(2r)^{d_2-d_1}}{w^{d_2-d_1}\omega_{d_2-d_1}}\right)^n\\
&= \frac{2^{(d_2-d_1-1)n}r^{d_2 n}}{K_I^{(d_1-1)n}w^{(d_2-d_1)n}(an)^n\omega_{d_2-d_1}^n}\\
&\geq \frac{2^{(d_2-d_1-1)n}r^{d_2 n}}{K_I^{d_2 n}\omega_{d_2-d_1}^n},
\end{aligned}\tag{A.60}$$

where the last inequality uses $an \leq c^{d_2-d_1}K_I \leq \frac{K_I^{d_2-d_1+1}}{\tau_\ell^{d_2-d_1}}$ and $w \leq \tau_\ell$.

On the other hand, $Q_2\left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i,r)\right) = \left(\frac{2r}{2K_I}\right)^{d_2 n} = \frac{r^{d_2 n}}{K_I^{d_2 n}}$, so from this and (A.60), we get (A.58) as

$$\begin{aligned}
Q_1\left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i,r)\right) &\geq \frac{2^{(d_2-d_1-1)n}}{\omega_{d_2-d_1}^n} Q_2\left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i,r)\right)\\
&\geq 2^{-n} Q_2\left(\prod_{i=1}^n \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i,r)\right).
\end{aligned}$$

□

Proposition 26. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, d_1, $d_2 \in \mathbb{N}$, with $\tau_g \leq \tau_\ell$ and $1 \leq d_1 < d_2 \leq m$, and suppose that $\tau_\ell < K_I$. Then*

$$\begin{aligned}
&\inf_{\hat{d}_n}\sup_{P\in\mathcal{Q}}\mathbb{E}_{P^{(n)}}[\ell(\hat{d}_n, d(P))]\\
&\geq \left(C^{(26)}_{d_1,d_2,K_I}\right)^n \min\left\{\tau_\ell^{-2(d_2-d_1+1)}n^{-2}, 1\right\}^{(d_2-d_1)n},
\end{aligned}\tag{A.61}$$

where $C^{(26)}_{d_1,d_2,K_I} \in (0,\infty)$ is a constant depending only on d_1, d_2, and K_I and

$$\mathcal{Q} = \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p} \bigcup \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}.$$

Proof of Proposition 26. Let $J = [-K_I, K_I]^{d_2}$. Let S_n be the permutation group, and $S_n \curvearrowright J^n$ by coordinate change, i.e. $\sigma \in S_n$, $x \in J^n$, $\sigma x := (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. For any set $A \subset J^n$, let $S_n A := \{\sigma x \in J^n : \sigma \in S_n,\ x \in A\}$.

Let T_i be T_i's from Lemma 24. Let $T := S_n \prod_{i=1}^{n} T_i$, and $V := \bigcup_{i=1}^{n} T_i = \Pi_{1:d_2}(T)$. Intuitively, T is the set of points $x = (x_1, \ldots, x_n)$ where x_i lies on one of the T_j.

Let $C = \mathbb{B}_{\mathbb{R}^{d_2-d_1}}(0, w)$ where w is from Lemma 24, and precisely define a set of d_1-dimensional distribution $\mathcal{P}_1$ in (2.13) and a set of d_2-dimensional distribution $\mathcal{P}_2$ in (2.14) as

$$\begin{aligned}\mathcal{P}_1 &= \{P \in \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p} : \text{ there exists } M \in \mathscr{M}(C^n) \text{ such that } P \text{ is uniform on } M\},\\ \mathcal{P}_2 &= \{\lambda_J\} \subset \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}.\end{aligned} \tag{A.62}$$

Define a map $\Phi : C^n \to \mathcal{P}_1$ by $\Phi(y_1, \ldots, y_n) = \lambda_{\mathscr{M}(y_1,\ldots,y_n)}$, i.e. the uniform measure on $\mathscr{M}(y_1, \ldots, y_n)$. Impose a topology and probability measure structure on $\mathcal{P}_1$ by the pushforward topology and the uniform measure on C^n, i.e. $\mathcal{P}' \subset \mathcal{P}_1$ is open if and only if $\Phi^{-1}(\mathcal{P}')$ is open in C^n, $\mathcal{P}' \subset \mathcal{P}_1$ is measurable if and only if $\Phi^{-1}(\mathcal{P}') \in \mathcal{B}(C^n)$, and $\mu_1(\mathcal{P}') = \lambda_{C^n}(\Phi^{-1}(\mathcal{P}'))$.

Define a probability measure Q_1, Q_2 on $(J^n, \mathcal{B}(J^n))$ by

$$Q_1(A) := \int_{\mathcal{P}_1} P^{(n)}(A) d\mu_1(P) \qquad \text{and} \qquad Q_2 = \lambda_{J^n}. \tag{A.63}$$

Fix $P \in \mathcal{P}_1$, let $x = \Phi^{-1}(P)$. Then $P^{(n)}(A) = \lambda^{(n)}_{\mathscr{M}(x)}(A)$ is a measurable function of x and Φ is a homeomorphism. Hence, $p^{(n)}(A)$ is measurable function and $Q_1(A)$ is well defined. Define $\nu = Q_1 + \lambda_J$. Then $Q_1,\ Q_2 \ll \nu$, so there exist densities $q_1 = \frac{dQ_1}{d\nu},\ q_2 = \frac{dQ_2}{d\nu}$ with respect to ν.

Then by applying Le Cam's Lemma (Lemma 1) with $\theta(P) = d(P)$, $\mathcal{P}_1$ and $\mathcal{P}_2$ from (A.62), and Q_1 and Q_2 in (A.63), the minimax rate $\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} \mathbb{E}_P \left[\ell(\hat{d}_n,\, d(P))\right]$ can be lower bounded as

$$\begin{aligned}\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} \mathbb{E}_P \left[\ell(\hat{d}_n,\, d(P))\right] &\geq \frac{\ell(d_1,\, d_2)}{2} \int_{J^n} q_1(x) \wedge q_2(x) d\nu(x)\\ &= \frac{1}{2} \int_{J^n} q_1(x) \wedge q_2(x) d\nu(x).\end{aligned} \tag{A.64}$$

Then from Claim 25, for all $x \in intT$, there exists $r_x > 0$ s.t. for all $r < r_x$,

$$Q_1\left(\prod_{i=1}^{n} \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)\right) \geq 2^{-n} Q_2\left(\prod_{i=1}^{n} \mathbb{B}_{\|\cdot\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)\right).$$

Hence $q_1(x)$ is lower bounded by $q_2(x)$ whenever $x \in intT$ as

$$q_1(x) \geq 2^{-n} q_2(x) \text{ if } x \in intT,$$

and $q_1(x) \wedge q_2(x)$ is correspondingly lower bounded by $q_2(x)$ as

$$q_1(x) \wedge q_2(x) \geq 2^{-n} q_2(x) 1(x \in intT).$$

Hence the integration of $q_1(x) \wedge q_2(x)$ over T is lower bounded as

$$\frac{1}{2} \int_T q_1(x) \wedge q_2(x) d\nu(x) \geq 2^{-n-1} \lambda_{J^n}(T). \tag{A.65}$$

Then from $a = \frac{K_I - \tau_\ell}{\left(d_2 - d_1 + \frac{1}{2}\right)\left\lceil \frac{n}{c^{d_2-d_1}} \right\rceil}$ and $w = \min\left\{\tau_\ell,\ \frac{(d_2-d_1)^2(K_I-\tau_\ell)^2}{2\tau_\ell\left(d_2-d_1+\frac{1}{2}\right)^2\left(\left\lceil \frac{n}{c^{d_2-d_1}} \right\rceil + 1\right)^2}\right\}$, $\lambda_{J^n}(T)$ can be lower bounded as

$$\begin{aligned}
\lambda_{J^n}\left(S_n \prod_{i=1}^{n} T_i\right) &= n! \lambda_{J^1}(T_1)^n \\
&= n! \left(\frac{(2K_I)^{d_1-1}\omega_{d_2-d_1} a w^{d_2-d_1}}{(2K_I)^{d_2}}\right)^n \\
&\geq \left(C^{(26,1)}_{d_1,d_2,K_I}\right)^n \min\left\{\tau_\ell^{-2(d_2-d_1+1)} n^{-2}, 1\right\}^{(d_2-d_1)n}, \qquad \text{(A.66)}
\end{aligned}$$

for some constant $C^{(26,1)}_{d_1,d_2,K_I}$ that depends only on d_1, d_2, and K_I. Hence by combining (A.64), (A.65), and (A.66), the minimax rate $\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} \mathbb{E}_P\left[\ell(\hat{d}_n,\, d(P))\right]$ can be lower bounded as

$$\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} \mathbb{E}_P\left[\ell(\hat{d}_n,\, d(P))\right] \geq \left(C^{(26)}_{d_1,d_2,K_I}\right)^n \min\left\{\tau_\ell^{-2(d_2-d_1+1)} n^{-2}, 1\right\}^{(d_2-d_1)n},$$

for some constant $C^{(26)}_{d_1,d_2,K_I}$ that depends only on d_1, d_2, and K_I. Then since $\mathcal{P}_1 \subset \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p}$ and $\mathcal{P}_2 \subset \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}$, the minimax rate R_n in (2.5) can be lower bounded by the minimax rate $\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} \mathbb{E}_P\left[\ell(\hat{d}_n,\, d(P))\right]$, i.e.

$$\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p} \cup \mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_P[\ell(\hat{d}_n, d(P))] \geq \inf_{\hat{d}_n} \sup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} \mathbb{E}_P[\ell(\hat{d}_n, d(P))],$$

which completes the proof of showing (A.61). □

A.4 Proofs For Section 2.4

Proposition 27. *Fix τ_g, $\tau_\ell \in (0, \infty]$, $K_I \in [1, \infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, with $\tau_g \leq \tau_\ell$. Let $\hat{d}_n$ be in* (2.16)*. Then:*

$$\sup_{P \in \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \qquad \text{(A.67)}$$

$$\begin{cases} = 0, & d = 1, \\ \leq \left(C^{(27)}_{K_I,K_p,K_v,d,m}\right)^n \left(1 + \tau_g^{-(dm+m-2d)n}\right) n^{-\frac{1}{d-1}n}, & d > 1. \end{cases} \qquad \text{(A.68)}$$

where $C^{(27)}_{K_I,K_p,K_v,d,m} \in (0, \infty)$ is a constant depending only on K_I, K_p, K_v, d, m.

Proof of Proposition 27. Note that for all $P \in \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}$ and $X_1, \ldots, X_n \sim P$, by Lemma 20,

$$\min_{\sigma \in S_n}\left\{\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|^d_{\mathbb{R}^m}\right\} \leq C^{(20)}_{K_I,K_v,d,m}\left(1 + \tau_g^{d-m}\right),$$

hence $\hat{d}_n$ in (2.16) always satisfies

$$\hat{d}_n(X) \le d = d(P). \tag{A.69}$$

Hence when $d = 1$, the risk of $\hat{d}_n$ is 0. When $d > 1$, from (A.69) and Proposition 22, the risk of $\hat{d}_n$ in (2.16) is upper bounded as

$$\begin{aligned}
&P^{(n)}\left[\hat{d}_n(X_1, \cdots, X_n) \neq d\right] \\
&= P^{(n)}\Bigg[\max\left\{k \in [1,m] : \min_{\sigma \in S_n}\left\{\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^k\right\} \le C^{(20)}_{K_I,K_v,d,m}\left(1 + \tau_g^{k-m}\right)\right\} \\
&\qquad < d\Bigg] \text{(from (A.69))} \\
&\le \sum_{k=1}^{d-1} P^{(n)}\left[\min_{\sigma \in S_n}\left\{\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^k\right\} \le C^{(20)}_{K_I,K_v,k,m}\left(1 + \tau_g^{k-m}\right)\right] \\
&\le \sum_{k=1}^{d-1}\left(C^{(21)}_{K_I,K_p,K_v,k,d,m}\right)^n \left(1 + \tau_g^{-\left(\frac{d}{k}m+m-2d\right)n}\right) n^{-\left(\frac{d}{k}-1\right)n} \text{ (Proposition 22)} \\
&\le \left(C^{(27)}_{K_I,K_p,K_v,d,m}\right)^n \left(1 + \tau_g^{-(dm+m-2d)n}\right) n^{-\frac{1}{d-1}n},
\end{aligned}$$

for some $C^{(27)}_{K_I,K_p,K_v,d,m}$ that depends only on K_I, K_p, K_v, d, m. Therefore, the risk is upper bounded as in (A.67), as

$$\begin{aligned}
&\sup_{P \in \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \\
&\begin{cases} = 0, & d = 1, \\ \le \left(C^{(27)}_{K_I,K_p,K_v,d,m}\right)^n \left(1 + \tau_g^{-(dm+m-2d)n}\right) n^{-\frac{1}{d-1}n}, & d > 1. \end{cases}
\end{aligned}$$

□

Proposition 28. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, with $\tau_g \le \tau_\ell$. Then:*

$$\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \le \left(C^{(28)}_{K_I,K_p,K_v,m}\right)^n \left(1 + \tau_g^{-(m^2-m)n}\right) n^{-\frac{1}{m-1}n} \tag{A.70}$$

where $C^{(28)}_{K_I,K_p,K_v,m} \in (0,\infty)$ is a constant depending only on K_I, K_p, K_v, m.

Proof of Proposition 28. Note that (**??**) still holds when $\mathcal{P}$ is as in (2.7). Hence applying Proposition 27 to (**??**) yields

$$\begin{aligned}
&\inf_{\hat{d}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right] \\
&\le \max_{1 \le d \le n}\left\{\sup_{P \in \mathcal{P}^d_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}}\left[\ell\left(\hat{d}_n, d(P)\right)\right]\right\} \\
&\le \left(C^{(28)}_{K_I,K_p,K_v,m}\right)^n \left(1 + \tau_g^{-\left(m^2-m\right)n}\right) n^{-\frac{1}{m-1}n},
\end{aligned}$$

where $C^{(28)}_{K_I,K_p,K_v,m} = \max_{1\le d\le m} C^{(27)}_{K_I,K_p,K_v,d,m}$ depends only on K_I, K_p, K_v, m. Hence the minimax rate R_n in (2.5) is upper bounded as in (A.70). □

Proposition 29. *Fix τ_g, $\tau_\ell \in (0,\infty]$, $K_I \in [1,\infty)$, $K_v \in (0, 2^{-m}]$, $K_p \in [(2K_I)^m, \infty)$, with $\tau_g \le \tau_\ell$ and and suppose that $\tau_\ell < K_I$. Then,*

$$\inf_{\hat{d}_n}\sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}[\ell(\hat{d}_n, d(P))] \ge \left(C^{(29)}_{K_I}\right)^n \min\left\{\tau_\ell^{-4} n^{-2}, 1\right\}^n \tag{A.71}$$

where $C^{(29)}_{K_I} \in (0,\infty)$ is a constant depending only on K_I.

Proof of Proposition 29. For any d_1 and d_2, from Proposition 26,

$$\begin{aligned}
&\inf_{\hat{d}_n}\sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}[\ell(\hat{d}_n, d(P))] \\
&\ge \inf_{\hat{d}_n} \sup_{P\in\mathcal{P}^{d_1}_{\tau_g,\tau_\ell,K_I,K_v,K_p}\cup\mathcal{P}^{d_2}_{\tau_g,\tau_\ell,K_I,K_v,K_p}} \mathbb{E}_{P^{(n)}}[\ell(\hat{d}_n, d(P))] \\
&\ge \left(C^{(26)}_{d_1,d_2,K_I}\right)^n \min\left\{\tau_\ell^{-2(d_2-d_1+1)} n^{-2}, 1\right\}^{(d_2-d_1)n}
\end{aligned}$$

Hence by plugging in $d_1 = 1$ and $d_2 = 2$, the minimax rate R_n in (2.5) is lower bounded as in (A.70), as

$$\inf_{\hat{d}_n}\sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}[\ell(\hat{d}_n, d(P))] \ge \left(C^{(29)}_{K_I}\right)^n \min\left\{\tau_\ell^{-4} n^{-2}, 1\right\}^n$$

with $C^{(29)}_{K_I} = C^{(26)}_{d_1=1,d_2=2,K_I}$.

□

Appendix B

Appendix for Chapter 3

B.1 Some Technical Results on the Model

B.1.1 Metric Properties

This section garners geometric lemmas on embedded manifolds in the Euclidean space that are related to the reach, and that will be used several times in the proofs.

Proposition 86. *Let $M \subset \mathbb{R}^m$ be a submanifold with reach $\tau_M > 0$.*

(i) For all $p \in M$, we let II_p denote the second fundamental form of M at x. Then for all unit vector $v \in T_pM$, $\|II_p(v,v)\| \leq \frac{1}{\tau_M}$.

(ii) The injectivity radius of M is at least $\pi\tau_M$.

(iii) The sectional curvatures κ of M satisfy $-\frac{2}{\tau_M^2} \leq \kappa \leq \frac{1}{\tau_M^2}$.

(iv) For all $p \in M$, the map $\exp_p : \mathring{\mathbb{B}}_{T_pM}(0, \pi\tau_M) \to \mathring{\mathbb{B}}_M(0, \pi\tau_M)$ is a diffeomorphism. Moreover, for all $\|v\| < \frac{\pi\tau_M}{2\sqrt{2}}$ and $w \in T_pM$,

$$\left(1 - \frac{\|v\|^2}{6\tau_M^2}\right)\|w\| \leq \left\|d_v \exp_p \cdot w\right\| \leq \left(1 + \frac{\|v\|^2}{\tau_M^2}\right)\|w\|$$

(v) For all $p \in M$ and $r \leq \frac{\pi\tau_M}{2\sqrt{2}}$, given any Borel set $A \subset \mathbb{B}_{T_pM}(0,r) \subset T_pM$ we have

$$\left(1 - \frac{r^2}{6\tau_M^2}\right)^d \mathcal{H}^d(A) \leq \mathcal{H}^d(\exp_p(A)) \leq \left(1 + \frac{r^2}{\tau_M^2}\right)^d \mathcal{H}^d(A).$$

(vi) Let γ be a geodesic at $p \in M$, and P_t the parallel transport operator along γ. Then for all $t < \pi\tau_M$ and $v \in T_pM$,

$$\angle(P_t(v), v) \leq \frac{t}{\tau_M}.$$

Proof of Proposition 86. (i) is stated in Proposition 2.1 in Niyogi et al. [2008], yielding (ii) from Corollary 1.4 in Alexander and Bishop [2006]. (iii) follows using (i) again and the Gauss equation [do Carmo, 1992, p. 130]. (iv) is derived from (iii) by a direct application of Lemma 8 in Dyer et al. [2015]. (v) follows from (iv) and Lemma 6 in Arias-Castro et al. [2013]. All that remain to be showed is (vi).

For this, assume without loss of generality that $\|v\| = 1$. Let $g : [0,t] \to \mathcal{S}^{d-1}$ be defined by $g(s) = P_s(v)$. Let $u \in \mathbb{R}^m$ be a unit vector and denoting by $\bar{\nabla}$ the ambient derivative. We may write

$$\langle g'(s), u\rangle = \left\langle \bar{\nabla}_{\gamma'(s)} P_s(w), u\right\rangle = \langle II(\gamma'(s), P_s(w)), u\rangle .$$

Hence $\|g'(s)\| \leq \frac{1}{\tau_M}$ for all $s \in [0,t]$. Since g is a curve on $\mathcal{S}^{d-1}$, this implies

$$\angle(P_t(v), v) = d_{\mathcal{S}^{d-1}}(\gamma(t), \gamma(0)) \leq \int_0^t \|g'(s)\| \, ds \leq \frac{t}{\tau_M}.$$

□

B.2 Geometry of the Reach

For $M \subset \mathbb{R}^m$, $a \in M$, and $v \in \mathbb{R}^m$ a non-zero vector, we define the *local directional reach* by

$$\tau_M(a,v) = \inf \left\{ d(x,M) | x \in \overline{Med(M)} \text{ with } x = a + tv \text{ for some } t \geq 0 \right\},$$

with the convention $\tau_M(a,v) = \infty$ if $\overline{Med(M)} \cap \{a + tv | t \geq 0\} = \emptyset$.

Lemma 87. *(i) For $x \notin Med(M) \cup M$, writing $a = \pi_M(x)$, we have $\tau_M(a, x-a) > 0$, and for all $b \in M$,*

$$\langle x - a, a - b \rangle \geq -\frac{\|a-b\|^2 \|x-a\|}{2\tau_M(a, x-a)}.$$

(ii) Let $0 < r < q < \infty$ be fixed. Let $x, y \notin Med(M) \cup M$ be such that $d(x,M) \vee d(y,M) \leq r$ and

$$\tau_M(\pi_M(x), x - \pi_M(x)) \wedge \tau_M(\pi_M(y), y - \pi_M(y)) \geq q.$$

Then,

$$\|\pi_M(x) - \pi_M(y)\| \leq \frac{q}{q-r} \|x - y\|.$$

Proof of Lemma 87. The proof of *(i)* follows that of Theorem 4.8 (7) in Federer [1959]. Let $v = \frac{x-a}{\|x-a\|}$ and $S = \{t | \pi_M(a+tv) = a\}$. As $\|x-a\| > 0$ belongs to S, $\sup S > 0$ and from [Federer, 1959, Theorem 4.8 (6)] we get $\sup S \geq \tau_M(a,v)$. Moreover, for $0 < t \in S$,

$$\|a + tv - b\| \geq d(a + tv, M) = t.$$

Developing and rearranging the square of previous inequality yields

$$\begin{aligned} \|a-b\|^2 + 2t\langle v, a-b\rangle + t^2 &\geq t^2, \\ 2t\langle v, a-b\rangle &\geq -\|a-b\|^2, \\ \langle x-a, a-b\rangle &\geq -\frac{\|a-b\|^2 \|x-a\|}{2t}. \end{aligned}$$

On the other hand, the proof of *(ii)* follows that of Theorem 4.8 (8) in Federer [1959]. Writing $a = \pi_M(x)$ and $b = \pi_M(y)$, the previous point yields

$$\langle x-a, a-b\rangle \geq -\frac{\|a-b\|^2 r}{2q} \quad \text{and} \quad \langle y-b, b-a\rangle \geq \frac{\|a-b\|^2 r}{2q}.$$

As a consequence,

$$\begin{aligned}\|x-y\|\,\|a-b\| &\geq \langle x-y, a-b\rangle \\ &= \langle (x-a)+(a-b)+(b-y), a-b\rangle \\ &\geq \|a-b\|^2\left(1-\frac{r}{q}\right),\end{aligned}$$

hence the result.

□

Lemma 88. *Let $M \subset \mathbb{R}^m$ be a submanifold with reach $\tau_M > 0$ having a reach attaining pair $(q_1, q_2) \in M^2$ such that $\|q_1 - q_2\| < 2\tau_M$. Write $z_0 \in Med(M)$ for the associated axis point. Then there exists a sequence of curves $\{\gamma_n\}_{n\in\mathbb{N}}$ of M joining q_1 and q_2 with*

$$\lim_{n\to\infty} Length(\gamma_n) = \tau_M \angle(q_1 - z_0, q_2 - z_0).$$

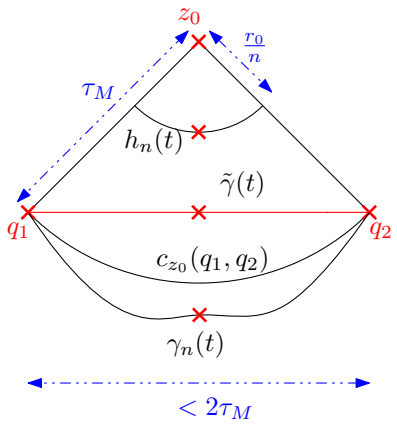


Figure B.1: Layout of the proof of Lemma 88.

Proof of Lemma 88. Without loss of generality, assume that z_0 coincides with the origin. Let $c_{z_0}(q_1, q_2)$ be the circle arc of center z_0 with endpoints q_1 and q_2, and let $\gamma : [-t_0, t_0] \to c_{z_0}(q_1, q_2)$ be its arc length parametrization with $\gamma(-t_0) = q_1$ and $\gamma(t_0) = q_2$. Let $\theta := \angle(q_1 - z_0, q_2 - z_0)$. Since $\|q_1 - z_0\| = \|q_2 - z_0\| = \tau_M$, we have $t_0 = \frac{1}{2}\tau_M\theta$. For all $t \in [-t_0, t_0]$, let $r_t := \sqrt{\tau_M^2 - \frac{\|q_1-q_2\|^2}{4}} / \cos\left(\frac{t}{\tau_M}\right)$, and let $\tilde{\gamma} : [-t_0, t_0] \to \mathbb{R}^m$ be $\tilde{\gamma}(t) = \frac{r_t}{\tau_M}\gamma(t)$. Let us show that for all $r \in (0, r_0]$ and $t \in [-t_0, t_0]$, the following holds:

$$\mathring{\mathbb{B}}\left(\frac{r}{\tau_M}\gamma(t), r\right) \subset \mathring{\mathbb{B}}\left(\tilde{\gamma}(t), r_t\right) \subset \mathring{\mathbb{B}}\left(q_1, \tau_M\right) \cup \mathring{\mathbb{B}}\left(q_2, \tau_M\right). \tag{B.1}$$

The left-hand side inclusion of (B.1) being straightforward, we turn to the second inclusion. First, note that by definition,

$$\tilde{\gamma}(t) = \left(\frac{1}{2} - \frac{\tan\left(\frac{t}{\tau_M}\right)}{2\tan\left(\frac{t_0}{\tau_M}\right)}\right) q_1 + \left(\frac{1}{2} + \frac{\tan\left(\frac{t}{\tau_M}\right)}{2\tan\left(\frac{t_0}{\tau_M}\right)}\right) q_2$$

for all $t \in [-t_0, t_0]$. Hence,

$$\tilde{\gamma}(t) - \tilde{\gamma}(0) = \frac{\tan\left(\frac{t}{\tau_M}\right)}{2\tan\left(\frac{t_0}{\tau_M}\right)}(q_2 - q_1), \tag{B.2}$$

and from $\tan\left(\frac{t_0}{\tau_M}\right) = \frac{\|q_1 - q_2\|}{2r_0}$, we get $\|\tilde{\gamma}(t) - \tilde{\gamma}(0)\| = r_0 \tan\left(\frac{t}{\tau_M}\right)$. Now suppose that $x \in \mathring{\mathbb{B}}\left(\tilde{\gamma}(t), r_t\right)$, then

$$\|x - \tilde{\gamma}(t)\|^2 < r_t^2. \tag{B.3}$$

Then,

$$\|x - \tilde{\gamma}(t)\|^2 = \|x - \tilde{\gamma}(0)\|^2 - 2\left\langle x - \tilde{\gamma}(0), \tilde{\gamma}(t) - \tilde{\gamma}(0)\right\rangle + \|\tilde{\gamma}(t) - \tilde{\gamma}(0)\|^2,$$

and $r_t^2 = r_0^2 + r_0^2 \tan^2\left(\frac{t}{\tau_M}\right) = r_0^2 + \|\tilde{\gamma}(t) - \tilde{\gamma}(0)\|^2$, hence applying these and (B.2) to (B.3) implies

$$\|x - \tilde{\gamma}(0)\|^2 - \frac{\tan\left(\frac{t}{\tau_M}\right)}{\tan\left(\frac{t_0}{\tau_M}\right)}\left\langle x - \tilde{\gamma}(0), q_2 - q_1\right\rangle < r_0^2. \tag{B.4}$$

Now applying $\tilde{\gamma}(-t_0) = q_1$ to (B.2) gives $q_1 - \tilde{\gamma}(0) = -\frac{1}{2}(q_2 - q_1)$, so

$$\begin{aligned}\|x - q_1\|^2 &= \|x - \tilde{\gamma}(0)\|^2 + 2\left\langle x - \tilde{\gamma}(0), q_1 - \tilde{\gamma}(0)\right\rangle + \|q_1 - \tilde{\gamma}(0)\|^2 \\ &= \|x - \tilde{\gamma}(0)\|^2 - \left\langle x - \tilde{\gamma}(0), q_2 - q_1\right\rangle + \frac{1}{4}\|q_1 - q_2\|^2.\end{aligned}$$

Similarly,

$$\|x - q_2\|^2 = \|x - \tilde{\gamma}(0)\|^2 + \left\langle x - \tilde{\gamma}(0), q_2 - q_1\right\rangle + \frac{1}{4}\|q_1 - q_2\|^2,$$

and hence

$$\begin{aligned}&\min\left\{\|x - q_1\|^2, \|x - q_2\|^2\right\} \\ &= \|x - \tilde{\gamma}(0)\|^2 - \left|\left\langle x - \tilde{\gamma}(0), q_2 - q_1\right\rangle\right| + \frac{1}{4}\|q_1 - q_2\|^2.\end{aligned} \tag{B.5}$$

Since $\left|\tan\left(\frac{t_0}{\tau_M}\right)\right| \geq \left|\tan\left(\frac{t}{\tau_M}\right)\right|$, applying (B.4) to (B.5) gives

$$\begin{aligned}&\min\left\{\|x - q_1\|^2, \|x - q_2\|^2\right\} \\ &\leq \|x - \tilde{\gamma}(0)\|^2 - \frac{\tan\left(\frac{t}{\tau_M}\right)}{\tan\left(\frac{t_0}{\tau_M}\right)}\left\langle x - \tilde{\gamma}(0), q_2 - q_1\right\rangle + \frac{1}{4}\|q_1 - q_2\|^2 \\ &< r_0^2 + \frac{1}{4}\|q_1 - q_2\|^2 = \tau_M^2,\end{aligned}$$

which asserts the second inclusion in (B.1).

Now, by definition of the reach in (1.6), $\left(\mathring{\mathbb{B}}(q_1, \tau_M) \cup \mathring{\mathbb{B}}(q_2, \tau_M)\right) \cap Med(M) = \emptyset$, hence (B.1) implies

$$\mathring{\mathbb{B}}\left(\frac{r}{\tau_M}\gamma(t), r\right) \cap Med(M) = \emptyset.$$

For all $n \in \mathbb{N}$, let us now define $h_n, \gamma_n : [-t_0, t_0] \to M$ by (See Figure B.1),

$$h_n(t) = \frac{r_0}{n\tau_M}\gamma(t) \quad \text{and} \quad \gamma_n(t) = \pi_M(h_n(t)).$$

Then for any fixed $n \in \mathbb{N}$ and $t_1, t_2 \in [-t_0, t_0]$ such that $|t_1 - t_2| < \tau_M$, from $\mathring{\mathbb{B}}\left(h_n(t_i), \frac{r_0}{n}\right) \cap Med(M) = \emptyset$, we get

$$\begin{aligned}\tau_M(\gamma_n(t_i), h_n(t_i) - \gamma_n(t_i)) &\geq d(h_n(t_i), M) + \frac{r_0}{n} \\ &\geq d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n},\end{aligned}$$

and since $d(h_n(t_i), M) \leq d(h_n(t_1), M) \vee d(h_n(t_2), M)$, Lemma 87 (ii) yields

$$\begin{aligned}&\|\gamma_n(t_1) - \gamma_n(t_2)\| = \|\pi_M(h_n(t_1)) - \pi_M(h_n(t_2))\| \\ &\leq \frac{\left(d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n}\right)\|h_n(t_1) - h_n(t_2)\|}{d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n} - d(h_n(t_1), M) \vee d(h_n(t_2), M)} \\ &= \frac{d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n}}{\frac{r_0}{n} - |d(h_n(t_1), M) - d(h_n(t_2), M)|}\|h_n(t_1) - h_n(t_2)\|.\end{aligned}$$

Noticing furthermore that

$$|d(h_n(t_1), M) - d(h_n(t_2), M)| \leq \|h_n(t_1) - h_n(t_2)\| \leq \frac{r_0}{n\tau_M}|t_1 - t_2|,$$

and

$$d(h_n(t_i), M) \leq d(z_0, M) + \|h_n(t_i) - z_0\| \leq \tau_M + \frac{r_0}{n},$$

we get

$$\begin{aligned}\|\gamma_n(t_1) - \gamma_n(t_2)\| &\leq \frac{\tau_M + 2\frac{r_0}{n}}{\frac{r_0}{n} - \frac{r_0}{n\tau_M}|t_1 - t_2|}\frac{r_0}{n\tau_M}|t_1 - t_2| \\ &= \frac{\tau_M + 2\frac{r_0}{n}}{\tau_M - |t_1 - t_2|}|t_1 - t_2|.\end{aligned}$$

For any fixed k and $0 \leq j \leq k$, set $t_{k,j} = \frac{2j-k}{k}t_0$. The inequality above yields,

$$\sum_{j=1}^{k}\|\gamma_n(t_{k,j}) - \gamma_n(t_{k,j-1})\| \leq \frac{\tau_M + 2\frac{r_0}{n}}{\tau_M - \frac{2t_0}{k}}2t_0,$$

so

$$Length(\gamma_n) = \limsup_k \sum_{j=1}^{k}\|\gamma_n(t_{k,j}) - \gamma_n(t_{k,j-1})\| \leq \left(1 + \frac{2r_0}{\tau_M n}\right)2t_0.$$

Moreover, the γ_n's are curves joining q_1 to q_2 with images $\gamma_n([-t_0, t_0]) \subset \mathbb{R}^m \setminus \mathring{\mathbb{B}}(z_0, \tau_M)$, so that their lengths are at most that of the arc of great circle $c_{z_0}(q_1, q_2)$, that is

$$Length(\gamma_n) \geq Length(c_{z_0}(q_1, q_2)) = 2t_0.$$

Hence,

$$\lim_{n\to\infty} Length(\gamma_n) = 2t_0 = \tau_M\theta.$$

□

Lemma 89. *Let M be a compact manifold, and $q_1, q_2 \in M$ with $q_1 \neq q_2$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of curves on M joining q_1 and q_2 such that $\sup_n Length(\gamma_n) < \infty$ Then there exists a curve γ on M joining q_1 and q_2 such that*

$$\liminf_{n\to\infty} Length(\gamma_n) \leq Length(\gamma) \leq \limsup_{n\to\infty} Length(\gamma_n).$$

Proof of Lemma 89. Without loss of generality, we take the γ_n's to be arc length parametrized. For all $n \in \mathbb{N}$, we let $g_n : [0,1] \to M$ be the reparametrization $g_n(t) = \gamma_n\left(Length(\gamma_n)t\right)$. Notice that for all $t \in [0,1]$, the set $(g_n(t))_{n\in\mathbb{N}}$ is contained in the compact set M, so that it is bounded uniformly in t. Moreover, writing $K = \sup_n Length(\gamma_n) < \infty$, we have that for all $t_1, t_2 \in [0,1]$,

$$\begin{aligned}\|g_n(t_1) - g_n(t_2)\| &= \|\gamma_n\left(Length(\gamma_n)t_1\right) - \gamma_n\left(Length(\gamma_n)t_2\right)\| \\ &\leq Length(\gamma_n)|t_1 - t_2| \\ &\leq K|t_1 - t_2|.\end{aligned}$$

Hence, the sequence $(g_n)_{n\in\mathbb{N}}$ is pointwise bounded and equicontinuous. From Arzelà-Ascoli theorem [Munkres, 1975, Theorem 45.4], there exists a curve $\gamma : [0,1] \to M$ and subsequence $(g_{n_i})_{i\in\mathbb{N}}$ converging uniformly to γ.

For any fixed k and $1 \leq j \leq k$, set $t_{k,j} = j/k$. The (pointwise) convergence of $(g_{n_i})_i$ towards γ ensures that

$$\sum_{j=0}^{k-1} \|\gamma(t_{k,j+1}) - \gamma(t_{k,j})\| = \lim_{i\to\infty} \sum_{j=0}^{k-1} \|g_{n_i}(t_{k,j+1}) - g_{n_i}(t_{k,j})\|.$$

Furthermore, from the uniform convergence of $(g_{n_i})_i$ towards γ on the compact set $[0,1]$,

$$\begin{aligned}Length(\gamma) &= \lim_{k\to\infty} \sum_{j=0}^{k-1} \|\gamma(t_{k,j+1}) - \gamma(t_{k,j})\| \\ &= \lim_{k\to\infty} \lim_{i\to\infty} \sum_{j=0}^{k-1} \|g_{n_i}(t_{k,j+1}) - g_{n_i}(t_{k,j})\| \\ &= \lim_{i\to\infty} Length(g_{n_i}) = \lim_{i\to\infty} Length(\gamma_{n_i}),\end{aligned}$$

hence the result. □

Proof of Lemma 35. Combining Lemma 88 and Lemma 89 provides the existence of a curve $\gamma \subset M$ joining q_1 and q_2 such that $Length(\gamma) = Length(c_{z_0}(q_1, q_2))$. But $M \subset \mathbb{R}^m \setminus \mathring{\mathbb{B}}(z_0, \tau_M)$, and since $\|q_1 - q_2\| < 2\tau_M$, $c_{z_0}(q_1, q_2)$ is the unique minimizing geodesic of $\partial\mathbb{B}(z_0, \tau_M) \subset \mathbb{R}^m \setminus \mathring{\mathbb{B}}(z_0, \tau_M)$ joining q_1 and q_2. Therefore, $\gamma = c_{z_0}(q_1, q_2) \subset M$, hence the result. □

Lemma 90. *Let $M \in \mathcal{M}^{d,m}_{\tau_{min},L}$ be a submanifold with reach τ_M. For all $p \in M$, let us denote*

$$L_p := \sup_{\substack{q\in\mathbb{B}_M(p,\tau_M/2) \\ v\in\mathbb{B}_{T_qM}(0,1)}} \left\|\gamma'''_{q,v}(0)\right\|.$$

Then for all $r \leq \tau_M/2$,

$$\left| \sup_{v\in T_pM, \|v\|=1} \left\|\gamma''_{p,v}(0)\right\| - \sup_{q\in\mathbb{B}(p,r)\cap M} \frac{2d(q-p, T_pM)}{\|q-p\|^2} \right| \leq 3\left(\frac{1}{\tau_M^2} + L_p\right) r.$$

To prove Lemma 90 we need the following straightforward result.

Lemma 91. *Let U be a linear space and $u \in U$, $n \in U^\perp$. If $v = u + n + e$, then*

$$|d(v, U) - \|v - u\|\,| \leq \|e\|\,.$$

Proof of Lemma 90. First note that for all unit vector $v \in T_pM$, $\gamma_{p,v}(r)$ belongs to $\mathbb{B}(p, r) \cap M$ and, whenever $0 < r \leq \frac{\tau_M}{2}$, Proposition 86 (ii) ensures that $\gamma_{p,v}(r) \neq p$. Therefore, it suffices to show that for all $q \in \mathbb{B}(p, r) \cap M$, there exists a unit tangent vector $v \in T_pM$ such that

$$\left| \left\|\gamma''_{p,v}(0)\right\| - \frac{2d(q-p, T_pM)}{\|q-p\|^2} \right| \leq 3\left(\frac{1}{\tau_M^2} + L_p\right) r.$$

Let $q \in \mathbb{B}(p, r) \cap M$ be different from p. Denoting $t = d_M(p, q) > 0$, we let $\gamma = \gamma_{p\to q}$ be the arc-length parametrized geodesic of minimal length such that $\gamma(0) = p$ and $\gamma(t) = q$. γ exists from Proposition 86 (ii) since $r \leq \frac{\tau_M}{2} < \pi\tau_M$. We will show that $v = \gamma'(0)$ provides the desired bound.

First, a Taylor expansion at zero of γ yields,

$$\left\| \frac{q-p}{t} - \gamma'(0) - \frac{t}{2}\gamma''(0) \right\| \leq L_p \frac{t^2}{6}.$$

Since $\gamma''(0) \in T_pM^\perp$, Lemma 91 shows that

$$\left| d\left(\frac{q-p}{t}, T_pM\right) - \left\| \frac{q-p}{t} - \gamma'(0) \right\| \right| \leq L_p \frac{t^2}{6}.$$

Therefore,

$$\begin{aligned}
&\left| \frac{2}{t} d\left(\frac{q-p}{t}, T_pM\right) - \|\gamma''(0)\| \right| \\
&\leq \frac{2}{t}\left(\left| d\left(\frac{q-p}{t}, T_pM\right) - \left\| \frac{q-p}{t} - \gamma'(0) \right\| \right| + \left\| \frac{q-p}{t} - \gamma'(0) - \frac{t}{2}\gamma''(0) \right\| \right) \\
&\leq \frac{2}{3} L_p t.
\end{aligned}$$

This yields,

$$\left| \frac{2d(q-p, T_pM)}{\|q-p\|^2} - \|\gamma''(0)\| \right| \leq 2d(q-p, T_pM) \left| \frac{1}{d_M(p,q)^2} - \frac{1}{\|q-p\|^2} \right| + \frac{2}{3} L_p t.$$

Moreover, from $\|q-p\| \leq d_M(p,q)$ and Proposition 6.3 in Niyogi et al. [2008], we derive

$$\begin{aligned}
\|q-p\|^2 \leq d_M(p,q)^2 &\leq \tau_M^2 \left(1 - \sqrt{1 - \frac{2\|q-p\|}{\tau_M}}\right)^2 \\
&\leq \tau_M^2 \frac{\left(\frac{\|q-p\|}{\tau_M}\right)^2}{\left(1 - \frac{2\|q-p\|}{\tau_M}\right)^{3/2}} \\
&\leq \frac{\|q-p\|^2}{1 - 3\frac{\|q-p\|}{\tau_M}},
\end{aligned}$$

where the last two inequalities follow from elementary real analysis arguments. Therefore, we get $t \leq 2\|q-p\|$ and

$$\left|\frac{1}{d_M(p,q)^2} - \frac{1}{\|q-p\|^2}\right| \leq \frac{3}{\tau_M \|q-p\|}.$$

Finally, using (1.7) we derive,

$$\begin{aligned}\left|\|\gamma''(0)\| - \frac{2d(q-p,T_pM)}{\|q-p\|^2}\right| &\leq 2d(q-p,T_pM)\frac{3}{\tau_M\|q-p\|} + \frac{4}{3}L_p\|q-p\| \\ &\leq \frac{3}{\tau_M^2}\|q-p\| + \frac{4}{3}L_p\|q-p\| \\ &\leq 3\left(\frac{1}{\tau_M^2} + L_p\right) r.\end{aligned}$$

□

Proof of Lemma 36. For $r > 0$, let $\Delta_r := \{(p,q) \in M^2 | \, \|p-q\| < r\}$, and $\bar{\Delta} = \cap_{r>0}\Delta_r$ denote the diagonal of M^2. Consider the map $\varphi : M^2 \setminus \bar{\Delta} \to \mathbb{R}$ defined by $\varphi(p,q) = 2d(q-p,T_pM)/\|q-p\|^2$. From (1.7), if there exists $p \neq q \in M$ such that $\varphi(p,q) = {\tau_M}^{-1}$, then there exists $z \in Med(M)$ with $d(z,M) = \tau_M$. Hence, for all $p \neq q \in T_pM$, $\varphi(p,q) < \tau_M^{-1}$, and by compactness of $M^2\setminus\Delta_r$, we have $\sup_{M^2\setminus\Delta_r}\varphi < \tau_M^{-1}$. Since we have the decomposition

$$\frac{1}{\tau_M} = \sup_{(p,q)\in M^2\setminus\bar{\Delta}} \varphi(p,q) = \max\left\{\sup_{(p,q)\in M^2\setminus\Delta_r}\varphi(p,q), \sup_{(p,q)\in\Delta_r\setminus\bar{\Delta}}\varphi(p,q)\right\},$$

we get $\sup_{\Delta_r\setminus\bar{\Delta}}\varphi = \tau_M^{-1}$. Moreover, Lemma 90 implies that

$$\left|\sup_{\substack{p\in M \\ v\in T_pM, \|v\|=1}} \left\|\gamma''_{p,v}(0)\right\| - \sup_{(p,q)\in\Delta_r\setminus\bar{\Delta}}\varphi(p,q)\right| \leq 3\left(\frac{1}{\tau_M^2} + L\right) r$$

for $r > 0$ small enough. Letting r go to zero yields

$$\sup_{\substack{p\in M \\ v\in T_pM, \|v\|=1}} \left\|\gamma''_{p,v}(0)\right\| = \frac{1}{\tau_M}.$$

Finally, the unit tangent bundle $T^{(1)}M = \{(p,v), p \in M, v \in T_pM, \|v\| = 1\}$ being compact, there exists $(q_0,v_0) \in T^{(1)}M$ such that $\gamma_0 = \gamma_{p_0,v_0}$ satisfies $\|\gamma_0''(0)\| = \tau_M^{-1}$, which concludes the proof. □

B.3 Analysis of the Estimator

B.3.1 Global Case

To show Proposition 39, we show a stronger result (Proposition 92) that applies to a reach attaining pair with any size 2λ (see Definition 34), meaning that it is not necessarily a bottleneck.

Proof of Proposition 39. Follows by applying Proposition 92 with $\lambda = \tau_M$. □

Proposition 92. *Let $M \subset \mathbb{R}^m$ be a submanifold, and $0 < \lambda \leq \tau_M$. Assume that M has a reach attaining pair $(q_1, q_2) \in M^2$ (see Definition 34) with $\|q_1 - q_2\| \geq 2\lambda$. Let $\mathbb{X} \subset M$. If there exists $x, y \in \mathbb{X}$ with $\|q_1 - x\| < \lambda$ and $\|q_2 - y\| < \lambda$, then*

$$0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x,y\})} \leq C_{\tau_M,\lambda} \max\left\{d_M(q_1,x), d_M(q_2,y)\right\},$$

where $C_{\tau_M,\lambda} = \frac{2\tau_M^2 + 6\tau_M\lambda + \lambda^2}{2\tau_M^2\lambda^2}$ depends only on the parameters τ_M, λ, and is a decreasing function of τ_M and λ when the other parameter is fixed.

Proof of Proposition 92. The two left hand inequalities are a direct consequence of Corollary 38, let us then focus on the third one.

Without loss of generality, assume that $\|q_1 - q_2\| = 2\lambda$. We set t to be equal to $\max\left\{d_M(q_1,x), d_M(q_2,y)\right\}$, and $z_1 := x + (q_2 - q_1)$. We have $\|z_1 - x\| = \|q_2 - q_1\| = 2\lambda$ and $\|y - q_2\|, \|q_1 - x\| \leq t$. Therefore, from the definition of $\hat{\tau}$ in (3.4) and the fact that the distance function to a linear space is 1-Lipschitz, we get

$$\begin{aligned}
\frac{1}{\hat{\tau}(\{x,y\})} &\geq \frac{2d(y-x, T_xM)}{\|y-x\|^2} \\
&= \frac{2d\left((y-q_2) + (z_1 - x) + (q_1 - x), T_xM\right)}{\|(y-q_2) + (z_1-x) + (q_1-x)\|^2} \\
&\geq \frac{d(z_1 - x, T_xM) - 2t}{2(\lambda + t)^2}.
\end{aligned}$$

Let now $\theta := \angle(q_2 - q_1, T_{q_1}M) = \min_{v \in T_{q_1}M} \angle(q_2 - q_1, v)$. Since $z_0 \in Med(M)$, with $q_1, q_2 \in \mathbb{B}(z_0, \tau_M)$ and $\|q_1 - q_2\| = 2\lambda$, for any v' such that $v' \perp z_0 - q_1$, we have $\angle(q_2 - q_1, v') \geq \frac{\pi}{2} - \angle(q_2 - q_1, z_0 - q_1)$. Hence, $\sin\theta \geq \frac{\lambda}{\tau_M}$ and $\cos\theta \leq \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M}$. Let $v_1 \in T_{q_1}M$ be any point in $T_{q_1}M$ realizing this angle, in the sense that $\angle(q_2 - q_1, v_1) = \angle(q_2 - q_1, T_{q_1}M)$. Then we have

$$\angle(z_1 - x, v_1) = \angle(q_2 - q_1, v_1) = \theta.$$

Let $\bar{v}_1 \in T_xM$ be the parallel transport of v_1 along the geodesic between q_1 and x. Since M has reach τ_M, Proposition 86 (vi) gives

$$\angle(v_1, \bar{v}_1) \leq \frac{d_M(x, q_1)}{\tau_M} \leq \frac{t}{\tau_M}.$$

Hence the angle $\angle(z_1 - x, T_xM)$ can be lower bounded as

$$\begin{aligned}
\angle(z_1 - x, T_xM) &\geq \angle(z_1 - x, \bar{v}_1) \\
&\geq \angle(z_1 - x, v) - \angle(v, \bar{v}_1) \\
&\geq \theta - \frac{t}{\tau_M}.
\end{aligned}$$

And $0 \leq \frac{\lambda}{\tau_M} - \frac{t}{\tau_M} \leq \theta - \frac{t}{\tau_M} \leq \angle(z_1 - x, T_xM) \leq \frac{\pi}{2}$, so the inequality is preserved by the sine function,

i.e.

$$\begin{aligned} d(z_1 - x, T_x M) &= \|z_1 - x\| \sin(\angle(z_1 - x, T_x M)) \\ &\geq 2\lambda \sin\left(\theta - \frac{t}{\tau_M}\right) \\ &= 2\lambda \left(\sin\theta \cos\frac{t}{\tau_M} - \cos\theta \sin\frac{t}{\tau_M}\right) \\ &= \frac{2\lambda^2}{\tau_M} \cos\frac{t}{\tau_M} - \frac{2\lambda\sqrt{\tau_M^2 - \lambda^2}}{\tau_M} \sin\frac{t}{\tau_M}. \end{aligned}$$

Combining the previous bounds yields,

$$\begin{aligned} \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x,y\})} &\leq \frac{1}{\tau_M} - \frac{d(z_1 - x, T_x M) - 2t}{2(\lambda + t)^2} \\ &\leq \frac{1}{\tau_M} - \frac{\frac{1}{\tau_M}\cos\frac{t}{\tau_M} - \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M \lambda}\sin\frac{t}{\tau_M} - \frac{t}{\lambda^2}}{\left(1 + \frac{t}{\lambda}\right)^2}. \end{aligned}$$

Using again that $t < \lambda \leq \tau_M$, the latter right-hand side term is itself upper bounded by,

$$\begin{aligned} &\frac{1}{\tau_M} - \left(\frac{1}{\tau_M}\left(1 - \frac{t^2}{2\tau_M^2}\right) - \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M \lambda}\frac{t}{\tau_M} - \frac{t}{\lambda^2}\right)\left(1 - \frac{2t}{\lambda}\right) \\ &\leq \left(\frac{\lambda}{2\tau_M^3} + \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M^2 \lambda} + \frac{1}{\lambda^2} + \frac{2}{\lambda \tau_M}\right) t \\ &= \frac{2\tau_M^3 + 2\lambda\tau_M\sqrt{\tau_M^2 - \lambda^2} + 4\tau_M^2\lambda + \lambda^3}{2\tau_M^3\lambda^2} t \\ &\leq \frac{2\tau_M^2 + 6\tau_M\lambda + \lambda^2}{2\tau_M^2\lambda^2} t := C_{\tau_M,\lambda} t, \end{aligned}$$

which is the announced result. □

As for Proposition 39, we tackle the proof of Proposition 40 by showing the following stronger one, Proposition 93 that contains an extra parameter $0 < \lambda \leq \tau_M$.

Proof of Proposition 40. Follows by applying Proposition 93 with $\lambda = \tau_M$. □

Proposition 93. *Let $P \in \mathcal{P}^{d,m}_{\tau_{\min}, L, f_{\min}}$, $M = supp(P)$ and $0 < \lambda \leq \tau_M$. Assume that M has a reach attaining pair $(q_1, q_2) \in M^2$ (see Definition 34) with $\|q_1 - q_2\| \geq 2\lambda$. Then*

$$\mathbb{E}_{P^n}\left[\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right|^p\right] \leq C_{\tau_M,\lambda,f_{min},d,p} n^{-\frac{p}{d}},$$

where $C_{\tau_M,\lambda,f_{min},d,p}$ depends only on τ_M, λ, f_{min} d, p, and is a decreasing function of τ_M and λ when other parameters are fixed.

Proof of Proposition 93. Let Q be the distribution on $\mathbb{R}^m$ associated to P. Let $s < \frac{1}{\tau_M}$, $C_{\tau_M,\lambda} = \frac{2\tau_M^2+6\tau_M\lambda+\lambda^2}{2\tau_M^2\lambda^2}$, and $t = \frac{1}{C_{\tau_M,\lambda}} s \leq 2\tau_M/9$. Let $\omega_d := \mathcal{H}^d(\mathbb{B}_{\mathbb{R}^d}(0,1))$ be the volume of the d-dimensional unit ball. Then note that from Proposition 86 (v), for all $q \in M$,

$$\begin{aligned} Q\left(\mathbb{B}_M(p,t)\right) &\geq f_{min}\mathcal{H}^d\left(\mathbb{B}_M(p,t)\right) \\ &\geq \omega_d f_{min}\left(1-\left(\frac{t}{6\tau_M}\right)^2\right)^d t^d \\ &\geq \omega_d f_{min}\left(\frac{728}{729}\right)^d t^d. \end{aligned}$$

Moreover, Proposition 39 asserts that $\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right| > s$ implies that either $\mathbb{B}_M(q_1,t) \cap \mathcal{X}_n = \emptyset$ or $\mathbb{B}_M(q_2,t) \cap \mathcal{X}_n = \emptyset$. Hence,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right| > s\right) &\leq \mathbb{P}\left(\mathbb{B}_M(q_1,t)\cap\mathcal{X}_n = \emptyset\right) + \mathbb{P}\left(\mathbb{B}_M(q_2,t)\cap\mathcal{X}_n = \emptyset\right) \\ &\leq 2\left(1-\omega_d f_{min}\left(\frac{728}{729}\right)^d t^d\right)^n \\ &\leq 2\exp\left(-n\omega_d f_{min}\left(\frac{728}{729}\right)^d C_{\tau_M,\lambda}^{-d} s^d\right). \end{aligned}$$

The integration of the above bound gives

$$\begin{aligned} \mathbb{E}_{P^n}\left[\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right|^p\right] &= \int_0^{\frac{1}{\tau_M^p}} \mathbb{P}\left(\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right|^p > s\right) ds \\ &\leq 2\int_0^\infty \exp\left(-n\omega_d f_{min}\left(\frac{728}{729}\right)^d C_{\tau_M,\lambda}^{-d} s^{\frac{d}{p}}\right) ds \\ &= \frac{2\left(\frac{729}{728}\right)^{\frac{p}{d}} C_{\tau_M,\lambda}^p}{\left(n\omega_d f_{min}\right)^{\frac{p}{d}}} \int_0^\infty x^{\frac{p}{d}-1} e^{-x} dx \\ &:= C_{\tau_M,\lambda,f_{\min},d,p} n^{-\frac{p}{d}}. \end{aligned}$$

where $C_{\tau_M,\lambda,f_{\min},d,p}$ depends only on τ_M, λ, f_{min}, d, p, and is a decreasing function of τ_M and λ when other parameters are fixed. □

B.3.2 Local Case

Lemma 94. *Let M be a submanifold and $p \in M$. Let $v_0, v_1 \in T_pM$ be a unit tangent vector, and let $\theta = \angle(v_0, v_1)$. Let $\gamma_{p,v}$ be the arc length parametrized geodesic starting from p with velocity v, and write $\gamma_i = \gamma_{p,v_i}$ for $i = 0, 1$. Let $\kappa_p = \max_{v\in\mathbb{B}_{T_pM}(0,1)} \left\|\gamma''_{p,v}(0)\right\|$. Then,*

$$\|\gamma_1''(0)\| \geq \|\gamma_0''(0)\| - \frac{\sqrt{2}}{\sqrt{2}-1}\sin^2\theta\left(\kappa_p + \|\gamma_0''(0)\|\right) - \frac{1}{\sqrt{2}-1}\left(\kappa_p - \|\gamma_0''(0)\|\right), \tag{B.6}$$

and

$$\|\gamma_1''(0)\| \geq \|\gamma_0''(0)\| - \sin^2\theta\left(\kappa_p + \|\gamma_0''(0)\|\right) - \frac{|\cos\theta\sin\theta|\,\kappa_p\sqrt{\kappa_p - \|\gamma_0''(0)\|}}{(\sqrt{2}-1)\,\|\gamma_0''(0)\|}\left(\frac{2\kappa_p}{\|\gamma_0''(0)\|}+1\right). \tag{B.7}$$

Proof of Lemma 94. Let $w \in T_pM$ be a unit vector satisfying $w \perp v_0$ and $v_1 = \cos\theta v_0 + \sin\theta w$. For $t \in \mathbb{R}$, let $v(t) := (\cos t)v_0 + (\sin t)w \in T_pM$, so that $v_1 = v(\theta)$. Then

$$\begin{aligned}\left\|d_0^2\exp_p(v(t),v(t))\right\| &= \left\|\cos^2 t\, d_0^2\exp_p(v_0,v_0) + 2\cos t\sin t\, d_0^2\exp_p(v_0,w)\right. \\ &\qquad \left. + \sin^2 t\, d_0^2\exp_p(w,w)\right\| \\ &\geq |\cos t|\left\|\cos t\, d_0^2\exp_p(v_0,v_0) + 2\sin t\, d_0^2\exp_p(v_0,w)\right\| \\ &\qquad - \sin^2 t\left\|d_0^2\exp_p(w,w)\right\|. \end{aligned} \tag{B.8}$$

Now, note that when $x \in [-1,1]$, $\sqrt{1+x} \geq 1 + f(x)$, where $f(x) = \min\{x, (\sqrt{2}-1)x\}$. Hence for any $v', v'' \in T_pM$,

$$\begin{aligned}\|v'+v''\| &= \sqrt{\|v'\|^2+\|v''\|^2}\sqrt{1+\frac{2\langle v', v''\rangle}{\|v'\|^2+\|v''\|^2}} \\ &\geq \sqrt{\|v'\|^2+\|v''\|^2}\left(1+f\left(\frac{2\langle v', v''\rangle}{\|v'\|^2+\|v''\|^2}\right)\right) \\ &\geq \|v'\| + f\left(\frac{2\langle v', v''\rangle}{\sqrt{\|v'\|^2+\|v''\|^2}}\right).\end{aligned}$$

Applying the latter inequality to (B.8) and using $d_0^2\exp_p(v_0,v_0) = \gamma_0''(0)$ together with $\left\|d_0^2\exp_p(w,w)\right\| \leq \kappa_p$ gives

$$\begin{aligned}&\left\|d_0^2\exp_p(v(t),v(t))\right\| \\ &\geq \cos^2 t\left\|d_0^2\exp_p(v_0,v_0)\right\| - \sin^2 t\left\|d_0^2\exp_p(w,w)\right\| \\ &\quad + |\cos t| f\left(\frac{4\cos t\sin t\left\langle d_0\exp_p(v_0,v_0), d_0\exp_p(v_0,w)\right\rangle}{\sqrt{\cos^2 t\left\|d_0^2\exp_p(v_0,v_0)\right\|^2 + 4\sin^2 t\left\|d_0^2\exp_p(v_0,w)\right\|^2}}\right) \\ &\geq \cos^2 t\,\|\gamma_0''(0)\| - \kappa_p\sin^2 t \\ &\quad + |\cos t| f\left(\frac{4\cos t\sin t\left\langle \gamma_0''(0), d_0\exp_p(v_0,w)\right\rangle}{\sqrt{\cos^2 t\,\|\gamma_0''(0)\|^2 + 4\sin^2 t\left\|d_0^2\exp_p(v_0,w)\right\|^2}}\right).\end{aligned}$$

Now, note that $f(x) \geq -|x|$ for $x \in [-1,1]$, so applying this with $t = \theta$ gives

$$\begin{aligned}\|\gamma_1''(0)\| &= \left\|d_0^2\exp_p(v_1,v_1)\right\| \\ &\geq \cos^2\theta\,\|\gamma_0''(0)\| - \sin^2\theta\kappa_p \\ &\qquad - \frac{4\left|\cos^2\theta\sin\theta\left\langle\gamma_0''(0), d_0\exp_p(v_0,w)\right\rangle\right|}{\sqrt{\cos^2\theta\,\|\gamma_0''(0)\|^2 + 4\sin^2\theta\left\|d_0^2\exp_p(v_0,w)\right\|^2}}.\end{aligned} \tag{B.9}$$

We now focus on the third term of the right-hand side. For this, note that either

$$t \sin t \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle \geq 0,$$

or

$$\cos(-t) \sin(-t) \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle \geq 0,$$

so that

$$\begin{aligned}
\kappa_p &\geq \max \left\{ \left\| d_0^2 \exp_p(v(-t), v(-t)) \right\|, \left\| d_0^2 \exp_p(v(t), v(t)) \right\| \right\} \\
&\geq \cos^2 t \left\| \gamma_0''(0) \right\| + \frac{4(\sqrt{2}-1) \left| \cos^2 t \sin t \left\langle \gamma_0''(0), d_0 \exp_p(v_0, w) \right\rangle \right|}{\sqrt{\cos^2 t \left\| \gamma_0''(0) \right\|^2 + 4 \sin^2 t \left\| d_0^2 \exp_p(v_0, w) \right\|^2}} \\
&\quad - \sin^2 t \kappa_p.
\end{aligned}$$

As a consequence,

$$\begin{aligned}
&\frac{\left| \cos^2 t \sin t \left\langle \gamma_0''(0), d_0 \exp_p(v_0, w) \right\rangle \right|}{\sqrt{\cos^2 t \left\| \gamma_0''(0) \right\|^2 + 4 \sin^2 t \left\| d_0^2 \exp_p(v_0, w) \right\|^2}} \\
&\leq \frac{1}{4(\sqrt{2}-1)} \left((1 + \sin^2 t) \kappa_p - \cos^2 t \left\| \gamma_0''(0) \right\| \right) \\
&= \frac{1}{4(\sqrt{2}-1)} \left(\cos^2 t \left(\kappa_p - \left\| \gamma_0''(0) \right\| \right) + 2 \sin^2 t \kappa_p \right).
\end{aligned}$$

First, setting $t = \theta$, we derive

$$\begin{aligned}
&\left\| \gamma_1''(0) \right\| \\
&\geq \cos^2 \theta \left\| \gamma_0''(0) \right\| - \left(1 + \frac{2}{\sqrt{2}-1} \right) \sin^2 \theta \kappa_p - \frac{1}{\sqrt{2}-1} \cos^2 \theta \left(\kappa_p - \left\| \gamma_0''(0) \right\| \right) \\
&= \left\| \gamma_0''(0) \right\| - \frac{\sqrt{2}}{\sqrt{2}-1} \sin^2 \theta \left(\kappa_p + \left\| \gamma_0''(0) \right\| \right) - \frac{1}{\sqrt{2}-1} \left(\kappa_p - \left\| \gamma_0''(0) \right\| \right).
\end{aligned}$$

Furthermore, let t_0 be defined by $\sin^2 t_0 = 1 - \frac{\left\| \gamma_0''(0) \right\|}{\kappa_p} + \epsilon$ for $\epsilon > 0$ small enough. Then

$$\sqrt{\cos^2 t_0 \left\| \gamma_0''(0) \right\|^2 + 4 \sin^2 t_0 \left\| d_0^2 \exp_p(v_0, w) \right\|^2} \leq \kappa_p,$$

yielding

$$\begin{aligned}
&\left| \left\langle \gamma_0''(0), d_0 \exp_p(v_0, w) \right\rangle \right| \\
&\leq \frac{\sqrt{\kappa_p}}{4(\sqrt{2}-1) \cos^2 t_0 |\sin t_0|} \left(\cos^2 t_0 \left(\kappa_p - \left\| \gamma_0''(0) \right\| \right) + 2 \sin^2 t_0 \kappa_p \right) \\
&= \frac{\kappa_p^{\frac{3}{2}}}{4(\sqrt{2}-1)} \left(\frac{1 - \frac{\left\| \gamma_0''(0) \right\|}{\kappa_p}}{\sqrt{1 - \frac{\left\| \gamma_0''(0) \right\|}{\kappa_p} + \epsilon}} + \frac{2\sqrt{1 - \frac{\left\| \gamma_0''(0) \right\|}{\kappa_p} + \epsilon}}{\frac{\left\| \gamma_0''(0) \right\|}{\kappa_p} - \epsilon} \right).
\end{aligned}$$

Sending $\epsilon \to 0$, we obtain

$$\left|\left\langle \gamma_0''(0), d_0 \exp_p(v_0, w)\right\rangle\right| \leq \frac{\kappa_p \sqrt{\kappa_p - \|\gamma_0''(0)\|}}{4(\sqrt{2}-1)} \left(\frac{2\kappa_p}{\|\gamma_0''(0)\|} + 1\right).$$

Using the previous bound together with

$$\cos^2\theta \left\|\gamma_0''(0)\right\|^2 + 4\sin^2\theta \left\|d_0^2 \exp_p(v_0, w)\right\|^2 \geq \left|\cos\theta\right| \left\|\gamma_0''(0)\right\|,$$

we finally obtain

$$\begin{aligned}\|\gamma_1''(0)\| \geq \|\gamma_0''(0)\| - \sin^2\theta \left(\kappa_p + \|\gamma_0''(0)\|\right) \\ - \frac{\left|\cos\theta \sin\theta\right| \kappa_p \sqrt{\kappa_p - \|\gamma_0''(0)\|}}{(\sqrt{2}-1)\|\gamma_0''(0)\|} \left(\frac{2\kappa_p}{\|\gamma_0''(0)\|} + 1\right).\end{aligned}$$

□

Proof of Lemma 41. First note that from Proposition 86 (ii), $d_M(x,y) < \pi\tau_M$ ensures the existence and uniqueness of the geodesic $\gamma_{x\to y}$. The two left hand inequalities are a direct consequence of Corollary 38. Let us then focus on the third one. Let $t_0 := d_M(x,y)$, and write $\gamma = \gamma_{x\to y}$ for short. By definition of $\hat{\tau}$ in (3.4),

$$\frac{1}{\hat{\tau}(\{x,y\})} \geq \frac{2d(y-x, T_xM)}{\|y-x\|^2} \geq \frac{2d(y-x, T_xM)}{t_0^2}. \tag{B.10}$$

Let $H_{\gamma''(0)} := \{x + u \in \mathbb{R}^m | \left\langle u, \gamma''_{x\to y}(0)\right\rangle = 0\}$ denote the affine hyperplane with normal vector $\gamma''(0)$ that contain x. Since $\gamma''(0) \in T_xM^\perp$, $T_xM \subset H_{\gamma''(0)}$. As a consequence,

$$d\left(y-x, T_xM\right) \geq d\left(y-x, H_{\gamma''(0)}\right) = \frac{\left|\langle \gamma''(0), y-x\rangle\right|}{\|\gamma''(0)\|}. \tag{B.11}$$

Using the Taylor expansion of γ at order two, we get

$$y - x = \gamma(t_0) - \gamma(0) = t_0\gamma'(0) + \int_0^{t_0}\int_0^t \gamma''(s)dsdt. \tag{B.12}$$

Since γ is parametrized by arc length, $\langle \gamma'(t), \gamma'(t)\rangle = 1$. Differentiating this identity at 0 yields $\langle \gamma''(0), \gamma'(0)\rangle = 0$. In addition, by definition of $\mathcal{M}^{d,m}_{\tau_{\min},L} \ni M$ (Definition 30), the geodesic γ satisfies $\|\gamma''(s) - \gamma''(0)\| \leq L|s|$. Therefore,

$$\begin{aligned}|\langle \gamma''(0), \gamma''(s)\rangle| &= |\langle \gamma''(0), \gamma''(0)\rangle - \langle \gamma''(0), \gamma''(s) - \gamma''(0)\rangle| \\ &\geq ||\gamma''(0)||^2 - L||\gamma''(0)|||s|.\end{aligned}$$

Combining the above bound together with (B.10), (B.11) and (B.12), we derive

$$\frac{1}{\hat{\tau}(\{x,y\})} \geq \|\gamma''(0)\| - \frac{2}{3}Lt_0,$$

which is the announced inequality. □

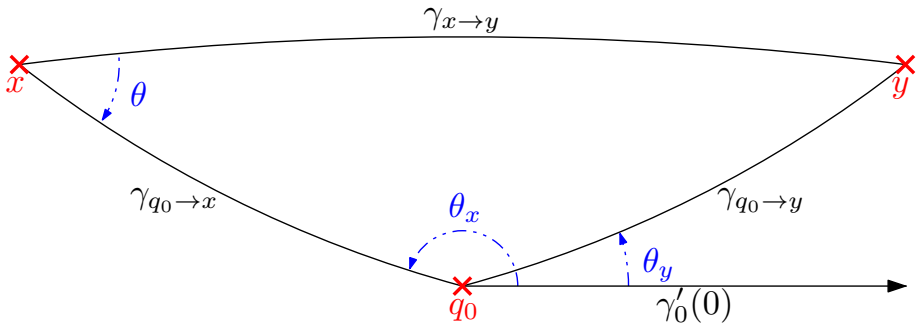


Figure B.2: Layout of Lemma 42.

Proof of Lemma 42. For short, in what follows, we let $t_x := d_M(q_0, x)$, $t_y := d_M(q_0, y)$, and $\theta := \angle(\gamma'_{x\to y}(0), \gamma'_{x\to q_0}(0)) = \pi - \angle(\gamma'_{x\to y}(0), \gamma'_{q_0\to x}(t_x))$ (see Figure B.2). From (B.6) in Lemma 94,

$$\begin{aligned}\left\|\gamma''_{x\to y}(0)\right\| &\geq \left\|\gamma''_{q_0\to x}(t_x)\right\| - \frac{\sqrt{2}}{\sqrt{2}-1}\sin^2\theta\left(\kappa_x + \left\|\gamma''_{q_0\to x}(t_x)\right\|\right) \\ &\quad - \frac{1}{\sqrt{2}-1}\left(\kappa_x - \left\|\gamma''_{q_0\to x}(t_x)\right\|\right) \\ &= \frac{\sqrt{2}}{\sqrt{2}-1}\cos^2\theta\left\|\gamma''_{q_0\to x}(t_x)\right\| - \left(\frac{1}{\sqrt{2}-1} + \frac{\sqrt{2}}{\sqrt{2}-1}\sin^2\theta\right)\kappa_x. \end{aligned} \tag{B.13}$$

We now focus on the term $\left\|\gamma''_{q_0\to x}(t_x)\right\|$. Since $\theta_x = \angle(\gamma'_0(0), \gamma'_{q_0\to x}(0))$, applying (B.7) in Lemma 94 yields

$$\left\|\gamma''_{q_0\to x}(0)\right\| \geq (1 - 2\sin^2\theta_x)\kappa_{q_0},$$

and since $\gamma''_{q_0\to x}$ is L-Lipschitz,

$$\begin{aligned}\left\|\gamma''_{q_0\to x}(t_x)\right\| &\geq \left\|\gamma''_{q_0\to x}(0)\right\| - \left\|\gamma''_{q_0\to x}(t_x) - \gamma''_{q_0\to x}(0)\right\| \\ &\geq (1 - 2\sin^2\theta_x)\kappa_{q_0} - Lt_x. \end{aligned} \tag{B.14}$$

Now we focus on bounding the terms $\sin^2\theta$ and $\cos^2\theta$. Let $\mathcal{S}^2_{\tau_M}$ be a d-dimensional sphere of radius τ_M. In what follows, for short, $\angle abc$ stands for $\angle(\gamma'_{b\to a}(0), \gamma'_{b\to c}(0))$. First, let $\tilde{q}_0, \tilde{x}, \tilde{y} \in \mathcal{S}^2_{\tau_M}$ be such that $d_{\mathcal{S}^2_{\tau_M}}(\tilde{q}_0, \tilde{x}) = d_M(q_0, x)$, $d_{\mathcal{S}^2_{\tau_M}}(\tilde{q}_0, \tilde{y}) = d_M(q_0, y)$, and $\angle\tilde{x}\tilde{q}_0\tilde{y} = \angle xq_0y$. Then from Toponogov's comparison theorem (see Karcher [1989]), we have $d_{\mathcal{S}^2_{\tau_M}}(\tilde{x}, \tilde{y}) \leq d_M(x, y)$. Moreover, the spherical law of cosines [Berger, 1987, Proposition 18.6.8] yields

$$\cos\left(\frac{d_{\mathcal{S}^2_{\tau_M}}(\tilde{x}, \tilde{y})}{\tau_M}\right) = \cos\left(\frac{t_x}{\tau_M}\right)\cos\left(\frac{t_y}{\tau_M}\right) + \sin\left(\frac{t_x}{\tau_M}\right)\sin\left(\frac{t_y}{\tau_M}\right)\cos\left(\angle\tilde{x}\tilde{q}_0\tilde{y}\right),$$

and since $t_x, t_y \leq \frac{\pi}{2}$ and $\cos(\cdot)$ is decreasing on $[0, \pi]$, we get

$$t_y \leq d_{\mathcal{S}^2_{\tau_M}}(\tilde{x}, \tilde{y}) \leq d_M(x, y).$$

Now, let $\bar{q}_0, \bar{x}, \bar{y} \in \mathcal{S}^2_{\tau_M}$ be such that $d_{\mathcal{S}^2_{\tau_M}}(\bar{q}_0, \bar{x}) = d_M(q_0, x)$, $d_{\mathcal{S}^2_{\tau_M}}(\bar{q}_0, \bar{y}) = d_M(q_0, y)$, and $d_{\mathcal{S}^2_{\tau_M}}(\bar{x}, \bar{y}) = d_M(x, y)$. Applying Toponogov's comparison theorem (see Karcher [1989]), we have $\angle q_0xy \leq \angle\bar{q}_0\bar{x}\bar{y}$ and $\angle xq_0y \leq \angle\bar{x}\bar{q}_0\bar{y}$, and from the spherical law of cosines [Berger, 1987, Proposition 18.6.8],

$$\cos\left(\angle\bar{q}_0\bar{x}\bar{y}\right) = \frac{\cos\left(\frac{t_y}{\tau_M}\right) - \cos\left(\frac{t_x}{\tau_M}\right)\cos\left(\frac{d_M(x,y)}{\tau_M}\right)}{\sin\left(\frac{t_x}{\tau_M}\right)\sin\left(\frac{d_M(x,y)}{\tau_M}\right)} \geq 0,$$

so that $\angle q_0xy \leq \angle \bar{q}_0\bar{x}\bar{y} \leq \frac{\pi}{2}$. Also, $\angle xq_0y \geq |\theta_x - \theta_y| \geq \frac{\pi}{2}$ yields $\frac{\pi}{2} \leq \angle xq_0y \leq \angle \bar{x}\bar{q}_0\bar{y}$, and $\theta = \angle(\gamma'_{x\to y}(0), \gamma'_{q_0\to x}(t_x)) = \pi - \angle q_0xy$. Hence, applying the spherical law of sines and cosines [Berger, 1987, Proposition 18.6.8] yields

$$\begin{aligned}
\sin\theta = \sin(\angle q_0xy) &\leq \sin(\angle \bar{q}_0\bar{x}\bar{y}) \\
&= \frac{\sin\left(\frac{t_y}{\tau_M}\right)\sin(\angle \bar{x}\bar{q}_0\bar{y})}{\sqrt{1-\left(\cos\left(\frac{t_x}{\tau_M}\right)\cos\left(\frac{t_y}{\tau_M}\right)+\sin\left(\frac{t_x}{\tau_M}\right)\sin\left(\frac{t_y}{\tau_M}\right)\cos(\angle \bar{x}\bar{q}_0\bar{y})\right)^2}} \\
&\leq \frac{\sin\left(\frac{t_y}{\tau_M}\right)\sin(\angle \bar{x}\bar{q}_0\bar{y})}{\sqrt{1-\cos^2\left(\frac{t_x}{\tau_M}\right)\cos^2\left(\frac{t_y}{\tau_M}\right)}} \\
&= \frac{\sin\left(\frac{t_y}{\tau_M}\right)\sin(\angle \bar{x}\bar{q}_0\bar{y})}{\sqrt{\sin^2\left(\frac{t_y}{\tau_M}\right)+\sin^2\left(\frac{t_x}{\tau_M}\right)\cos^2\left(\frac{t_y}{\tau_M}\right)}} \\
&\leq \sin(\angle \bar{x}\bar{q}_0\bar{y}) \leq \sin(\angle xq_0y) \leq \sin(|\theta_x - \theta_y|).
\end{aligned} \tag{B.15}$$

And accordingly,

$$|\cos\theta| = \sqrt{1-\sin^2\theta} \geq \sqrt{1-\sin^2(|\theta_x-\theta_y|)} = |\cos(|\theta_x-\theta_y|)|. \tag{B.16}$$

Hence, applying (B.14), (B.15), and (B.16) to (B.13) gives

$$\begin{aligned}
\left\|\gamma''_{x\to y}(0)\right\| &\geq \frac{\sqrt{2}}{\sqrt{2}-1}\cos^2(|\theta_x-\theta_y|)\left((1-2\sin^2\theta_x)\kappa_{q_0} - Lt_x\right) \\
&\quad - \left(\frac{1}{\sqrt{2}-1} + \frac{\sqrt{2}}{\sqrt{2}-1}\sin^2(|\theta_x-\theta_y|)\right)\kappa_x \\
&= \frac{(\sqrt{2}\kappa_{q_0}-\kappa_x)}{\sqrt{2}-1} - \frac{\sqrt{2}}{\sqrt{2}-1}Lt_x\cos^2(\theta_x+\theta_y) \\
&\quad - \frac{\sqrt{2}}{\sqrt{2}-1}\left((\kappa_{q_0}+\kappa_x)\sin^2(|\theta_x-\theta_y|) + 2\kappa_{q_0}\sin^2\theta_x\cos^2(|\theta_x-\theta_y|)\right) \\
&\geq \kappa_{q_0} - \frac{1}{\sqrt{2}-1}\left(\kappa_x - \kappa_{q_0} + \sqrt{2}(3\kappa_{q_0}+\kappa_x)\sin^2(|\theta_x-\theta_y|) + \sqrt{2}Lt_x\right).
\end{aligned}$$

□

Proof of Proposition 44. In what follows, we let $t_0 \leq \frac{\tau_{\min}}{10}$,

$$\begin{aligned}
B_1 &:= \exp_{q_0}\left(\left\{v \in T_{q_0}M : \|v\| \leq t_0,\, \angle(\gamma'_0(0), v) \leq \sqrt{\frac{t_0}{\tau_{\min}}}\right\}\right), \\
B_2 &:= \exp_{q_0}\left(\left\{v \in T_{q_0}M : \|v\| \leq t_0,\, \angle(\gamma'_0(0), v) \geq \pi - \sqrt{\frac{t_0}{\tau_{\min}}}\right\}\right),
\end{aligned}$$

and $B_0 := B_1 \cup B_2$ (see Figure B.3). Let $\mathbb{X} \subset M$, and $x, y \in \mathbb{X}$ be such that $x \in B_1$, $y \in B_2$. Writing

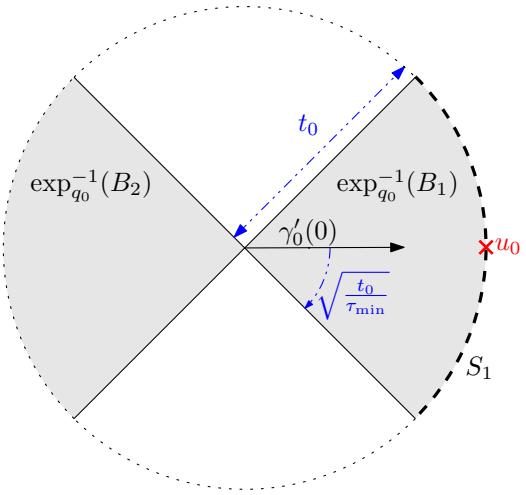


Figure B.3: Layout of the proof of Proposition 44.

$\theta_x := \angle(\gamma_0'(0), \gamma_{q_0\to x}'(0))$ and $\theta_y := \angle(\gamma_0'(0), \gamma_{q_0\to y}'(0))$, then $\theta_x \le \sqrt{\frac{t_0}{\tau_{\min}}} \le \frac{\pi}{4}$ and $\theta_y \ge \pi - \sqrt{\frac{t_0}{\tau_{\min}}} \ge \frac{3\pi}{4}$. Also, $d_M(q_0, x) \le t_0$ and $d_M(x, y) \le 2t_0$, so that

$$\begin{aligned}
0 &\le \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \\
&\le \frac{4\sqrt{2}\sin^2(|\theta_x - \theta_y|)}{(\sqrt{2}-1)\tau_M} + L\left(\frac{2}{3}d_M(x,y) + \frac{\sqrt{2}}{\sqrt{2}-1}d_M(q_0,x)\right) \\
&\le \left(\frac{16\sqrt{2}}{(\sqrt{2}-1)\tau_{\min}\tau_M} + \frac{(7\sqrt{2}-4)L}{3(\sqrt{2}-1)}\right) t_0.
\end{aligned}$$

A symmetric argument also applies when $x \in B_2$ and $y \in B_1$. Now, for any $s < \frac{1}{\tau_M}$, let $t_0(s) := \left(\frac{16\sqrt{2}}{(\sqrt{2}-1)\tau_{\min}^2} + \frac{(7\sqrt{2}-4)L}{3(\sqrt{2}-1)}\right)^{-1} s < \frac{\tau_{\min}}{10}$. The above argument implies that if $\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})}\right| > s$, then for any $x, y \in \mathbb{X} \cap B_0$, one has either $x, y \in B_1$ or $x, y \in B_2$. Hence,

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right| > s\right) \\
&\le \sum_{m=0}^{n} \binom{n}{m} \Big\{\mathbb{P}\left(X_1, \ldots, X_m \in M\backslash B_0, X_{m+1}, \ldots, X_n \in B_1\right) \\
&\qquad\qquad + \mathbb{P}\left(X_1, \ldots, X_m \in M\backslash B_0, X_{m+1}, \ldots, X_n \in B_2\right)\Big\} \\
&= \sum_{m=0}^{n} \binom{n}{m} \left\{(1-Q(B_0))^m Q(B_1)^{n-m} + (1-Q(B_0))^m Q(B_2)^{n-m}\right\} \\
&\le (1-Q(B_2))^n + (1-Q(B_1))^n. \qquad \text{(B.17)}
\end{aligned}$$

Let us derive lower bounds for $Q(B_1)$ and $Q(B_2)$. For this purpose, let $S_1 := \exp_{q_0}^{-1}(B_1) \cap \partial\mathbb{B}_{T_{q_0}M}(0, t_0)$ (see Figure B.3). Then $\exp_{q_0}^{-1}(B_1) \subset \mathbb{B}_{T_{q_0}M}(0, t_0)$ is a cone satisfying

$$\frac{\mathcal{H}^d\left(\exp_{q_0}^{-1}(B_1)\right)}{\mathcal{H}^d\left(\mathbb{B}_{T_{q_0}M}(0,t_0)\right)} = \frac{\mathcal{H}^{d-1}\left(S_1\right)}{\mathcal{H}^{d-1}\left(\partial\mathbb{B}_{T_{q_0}M}(0,t_0)\right)}.$$

Let $\omega_d := \mathcal{H}^d(\mathbb{B}_{\mathbb{R}^d}(0,1))$ and $\sigma_d := \mathcal{H}^d(\partial\mathbb{B}_{\mathbb{R}^{d+1}}(0,1))$ be the volumes of the d-dimensional unit ball and the unit sphere respectively. Then by homogeneity, $\mathcal{H}^d\left(\mathbb{B}_{T_{q_0}M}(0,t_0)\right) = \omega_d t_0^d$ and $\mathcal{H}^{d-1}\left(\partial\mathbb{B}_{T_{q_0}M}(0,t_0)\right) = \sigma_{d-1}t_0^{d-1}$. To derive a lower bound on $\mathcal{H}^{d-1}(S_1)$, consider $u_0 := t_0\gamma_0'(0) \in S_1$. Since $\tau_{S_1} = t_0$ and $\exp_{u_0}^{-1}(S_1) \subset \mathbb{B}_{T_{u_0}S_1}\left(0, \tau_{\min}^{-\frac{1}{2}} t_0^{\frac{3}{2}}\right)$, applying Proposition 86 (v) yields

$$\begin{aligned}\mathcal{H}^{d-1}(S_1) &\geq \left(1 - \frac{t_0}{6\tau_{\min}}\right)^{d-1} \mathcal{H}^{d-1}\left(\mathbb{B}_{T_{u_0}S_1}\left(0, \tau_{\min}^{-\frac{1}{2}} t_0^{\frac{3}{2}}\right)\right) \\ &\geq \left(\frac{59}{60}\right)^{d-1} \omega_{d-1}\tau_{\min}^{-\frac{d-1}{2}} t_0^{\frac{3d-3}{2}},\end{aligned}$$

and hence

$$\begin{aligned}\mathcal{H}^{d-1}\left(\exp_{q_0}^{-1}(B_1)\right) &= \frac{\mathcal{H}^d\left(\mathbb{B}_{T_{q_0}M}(0,t_0)\right)\mathcal{H}^{d-1}(S_1)}{\mathcal{H}^{d-1}\left(\partial\mathbb{B}_{T_{q_0}M}(0,t_0)\right)} \\ &\geq \left(\frac{59}{60}\right)^{d-1} \frac{\omega_{d-1}}{d}\tau_{\min}^{-\frac{d-1}{2}} t_0^{\frac{3d-1}{2}}.\end{aligned}$$

Finally, since $\exp_{q_0}^{-1}(B_1) \subset \mathbb{B}_{T_{q_0}M}(q_0, \frac{\tau_M}{10})$, Proposition 86 (v) yields

$$\mathcal{H}^d(B_1) \geq \left(\frac{599}{600}\right)^d \mathcal{H}^d\left(\exp_{q_0}^{-1}(B_1)\right) \geq \left(\frac{35341}{36000}\right)^d \frac{1}{d}\tau_{\min}^{-\frac{d-1}{2}} t_0^{\frac{3d-1}{2}},$$

and hence,

$$Q(B_1) \geq \left(\frac{35341}{36000}\right)^d \frac{f_{\min}}{d}\tau_{\min}^{-\frac{d-1}{2}} t_0^{\frac{3d-1}{2}} \geq C_{\tau_{\min},d,L,f_{\min}} s^{\frac{3d-1}{2}}.$$

By symmetry, the same bound holds for $Q(B_2)$. Applying these bounds to (B.17) gives

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathcal{X}_n)}\right| > s\right) &\leq 2\left(1 - C_{\tau_{\min},d,L,f_{\min}} s^{\frac{3d-1}{2}}\right)^n \\ &\leq 2\exp\left(-C_{\tau_{\min},d,L,f_{\min}} n s^{\frac{3d-1}{2}}\right).\end{aligned}$$

As a consequence, by integration,

$$\begin{aligned}\mathbb{E}_{P^n}\left[\left|\frac{1}{\hat{\tau}(\mathcal{X}_n)} - \frac{1}{\tau_M}\right|^p\right] &= \int_0^{\frac{1}{\tau_M^p}} \mathbb{P}\left(\left|\frac{1}{\hat{\tau}(\mathcal{X}_n)} - \frac{1}{\tau_M}\right|^p > s\right) ds \\ &\leq 2\int_0^\infty \exp\left(-C_{\tau_{\min},d,L,f_{\min}} n s^{\frac{3d-1}{2p}}\right) ds \\ &= 2\left(C_{\tau_{\min},d,L,f_{\min}} n\right)^{-\frac{2p}{3d-1}} \int_0^\infty x^{\frac{2p}{3d-1}} e^{-x} dx \\ &:= C_{\tau_{\min},d,L,f_{\min},p} n^{-\frac{2p}{3d-1}}.\end{aligned}$$

□

B.4 Minimax Lower Bounds

B.4.1 Stability of the Model With Respect to Diffeomorphisms

To prove Proposition 48, we will use the following result stating that the reach is a stable quantity with respect to $\mathcal{C}^2$-perturbations.

Lemma 95 (Theorem 4.19 in Federer [1959]). *Let $A \subset \mathbb{R}^m$ with $\tau_A \geq \tau_{min} > 0$ and $\Phi : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ is a $\mathcal{C}^1$-diffeomorphism such that Φ,Φ^{-1}, and $d\Phi$ are Lipschitz with Lipschitz constants K,N and R respectively, then*

$$\tau_{\Phi(A)} \geq \frac{\tau_{min}}{(K + R\tau_{min})N^2}.$$

Proof of Proposition 48. Let $M' = \Phi(M)$ be the image of M by the mapping Φ. Since Φ is a global diffeomorphism, M' is a closed submanifold of dimension one. Moreover, Φ is $\|d\Phi\|_{op} \leq (1 + \|d\Phi - I_D\|_{op})$-Lipschitz, Φ^{-1} is $\|d\Phi^{-1}\|_{op} \leq (1 - \|d\Phi - I_D\|_{op})^{-1}$-Lipschitz, and $d\Phi$ is $\|d^2\Phi\|_{op}$-Lipschitz. From Lemma 95,

$$\tau_{M'} \geq \frac{\tau_{min}(1 - \|d\Phi - I_D\|_{op})^2}{\|d^2\Phi\|_{op}\,\tau_{min} + (1 + \|d\Phi - I_D\|_{op})} \geq \tau_{min}/2,$$

where we used that $\|d^2\Phi\|_{op}\,\tau_{min} \leq 1/2$ and $\|d\Phi - I_D\|_{op} \leq 0.1$. All that remains to be proved now is the bound on the third order derivative of the geodesics of M'. We denote by γ and $\tilde{\gamma}$ the geodesics of M and M' respectively.

Let $p' = \Phi(p) \in M'$ and $v' = d_p\Phi.v \in T_{p'}M'$ be fixed. Since $M \in \mathcal{M}^{d,m}_{\tau_{min},L}$ is a compact $\mathcal{C}^3$-submanifold with geodesics $\|\gamma'''(0)\| \leq L$, M can be parametrized locally by a $\mathcal{C}^3$ bijective map $\Psi_p : \mathbb{B}_{\mathbb{R}^d}(0,\varepsilon) \to M$ with $\Psi_p(0) = p$. For a smooth curve γ on M nearby p, we let $c = (c_1, \ldots, c_d)^t$ denote its lift in the coordinates $\mathbf{x} = \Psi_p^{-1}$, that is $\gamma(t) = \Psi_p \circ c(t)$. $\gamma = \gamma_{p,v}$ is the geodesic of M with initial conditions p and v if and only if c satisfies the geodesic equations (see do Carmo [1992] p.62). That is, the second order ordinary differential equation

$$\begin{cases} c''_\ell(t) + \left\langle \Gamma^\ell\left(c(t)\right) \cdot c'(t), c'(t) \right\rangle = 0, & (1 \leq \ell \leq d) \\ c(0) = 0 \text{ and } c'(0) = d_p\mathbf{x}.v, \end{cases} \tag{B.18}$$

where $\Gamma^\ell = \left(\Gamma^\ell_{i,j}\right)_{1\leq i,j\leq d}$ are the Christoffel symbols of the $\mathcal{C}^3$ chart $\mathbf{x}$, which depends only on $\mathbf{x}$ and its differentials of order 1 and 2. By construction, M' is parametrized locally by $\Psi'_{p'} = \Phi \circ \Psi_p$ yielding local coordinates $\mathbf{y} = \Psi'^{-1}_{p'} = \Psi_p^{-1} \circ \Phi^{-1}$ nearby $p' \in M'$. Writing $\tilde{\Gamma}^\ell$ for the Christoffel's symbols of M', $\tilde{\gamma}$ is a geodesic of M' at p' if its lift $\tilde{c} = \Psi'^{-1}_{p'}(\tilde{\gamma})$ satisfies (B.18) with Γ^ℓ replaced by $\tilde{\Gamma}^\ell$, and initial conditions $\tilde{c}(0) = c$ and $\tilde{c}'(0) = d_{p'}\mathbf{y}.v' = d_p\mathbf{x}.v$. From chain rule, the $\tilde{\Gamma}^\ell$'s depend on Γ, $d\Phi$, and $d^2\Phi$.

Write $c'''(0) - \tilde{c}'''(0)$ by differentiating (B.18): since $c(0) = \tilde{c}(0) = 0$ and $c''(0) = \tilde{c}''(0)$, we get that for $\|I_D - d\Phi\|_{op}$, $\|d^2\Phi\|_{op}$ and $\|d^3\Phi\|_{op}$ small enough, $\|c'''(0) - \tilde{c}'''(0)\|$ can be made arbitrarily small. In particular, $\tilde{\gamma}'''(0)$ gets arbitrarily close to $\gamma'''(0)$, so that $\|\tilde{\gamma}'''(0)\| \leq \|\gamma'''(0)\| + L \leq 2L$, which concludes the proof. □

B.4.2 Lemmas on the Total Variation Distance

Prior to any actual construction, we show this straighforward lemma bounding the total variation between uniform distribution on manifolds that are perturbations of each other. For $M \subset \mathbb{R}^m$, write $\lambda_M = \mathbb{1}_M\mathcal{H}^d/\mathcal{H}^d(M)$ for the uniform probability distribution on M.

Lemma 96. *Let $M \subset \mathbb{R}^m$ be a d-dimensional submanifold and $B \subset \mathbb{R}^m$ be a Borel set. Let $\Phi : \mathbb{R}^m \to \mathbb{R}^m$ be a global diffeomorphism such that $\Phi_{|B^c}$ is the identity map and $\|d\Phi - I_D\|_{op} \leq 2^{1/d} - 1$. Then $\mathcal{H}^d(\Phi(M)) \leq 2\mathcal{H}^d(M)$ and $TV\left(\lambda_M, \lambda_{\Phi(M)}\right) \leq 12\lambda_M(B)$.*

Proof of Lemma 96. Since Φ is $(1 + \|d\Phi - I_D\|_{op})$-Lipschitz, Lemma 7 of Arias-Castro et al. [2013] asserts that

$$\mathcal{H}^d(\Phi(M \cap B)) \leq (1 + \|d\Phi - I_D\|_{op})^d \mathcal{H}^d(M \cap B) \leq 2\mathcal{H}^d(M \cap B).$$

Therefore,

$$\begin{aligned}\mathcal{H}^d(\Phi(M)) - \mathcal{H}^d(M) &= \mathcal{H}^d(\Phi(M \cap B)) - \mathcal{H}^d(M \cap B) \\ &\leq \mathcal{H}^d(M \cap B) \leq \mathcal{H}^d(M).\end{aligned}$$

Now, writing $\triangle$ for the symmetric difference of sets, we have $M \triangle \Phi(M) = (B \cap M) \triangle (B \cap \Phi(M)) \subset (B \cap M) \cup (B \cap \Phi(M))$. Therefore, Lemma 7 in Arias-Castro et al. [2013] yields,

$$\begin{aligned}TV\left(\lambda_M, \lambda_{\Phi(M)}\right) &\leq 4\frac{\mathcal{H}^d(M \triangle \Phi(M))}{\mathcal{H}^d(M \cup \Phi(M))} \\ &\leq 4\frac{\mathcal{H}^d(M \cap B) + \mathcal{H}^d(\Phi(M) \cap B)}{\mathcal{H}^d(M)} \\ &= 4\frac{\mathcal{H}^d(M \cap B) + \mathcal{H}^d(\Phi(M \cap B))}{\mathcal{H}^d(M)} \\ &\leq 12\frac{\mathcal{H}^d(M \cap B)}{\mathcal{H}^d(M)} = 12\lambda_M(B).\end{aligned}$$

□

Let us now tackle the proof of Lemma 47. For this, we will need the following elementary differential geometry results Lemma 97 and Corollary 98.

Lemma 97. *Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be $\mathcal{C}^1$ and $x \in \mathbb{R}^d$ be such that $g(x) = 0$ and $d_x g \neq 0$. Then there exists $r > 0$ such that $\mathcal{H}^d(g^{-1}(0) \cap \mathbb{B}(x, r)) = 0$.*

Proof of Lemma 97. Let us prove that for $r > 0$ small enough, the intersection $g^{-1}(0) \cap \mathbb{B}(x, r)$ is contained in a submanifold of codimension one of $\mathbb{R}^d$. Writing $g = (g_1, \ldots, g_k)$, assume without loss of generality that $\partial_{x_1} g_1 \neq 0$. Since $g_1 : \mathbb{R}^d \to \mathbb{R}$ is nonsingular at x, the implicit function theorem asserts that $g_1^{-1}(0)$ is a submanifold of dimension $d-1$ of $\mathbb{R}^d$ in a neighborhood of $x \in \mathbb{R}^d$. Therefore, for $r > 0$ small enough, $g_1^{-1}(0) \cap \mathbb{B}(x, r)$ has d-dimensional Hausdorff measure zero. The result hence follows, noticing that $g^{-1}(0) \subset g_1^{-1}(0)$. □

Corollary 98. *Let $M, M' \subset \mathbb{R}^m$ be two compact d-dimensional submanifolds, and $x \in M \cap M'$. If $T_x M \neq T_x M'$, there exists $r > 0$ such that $A = M \cap M' \cap \mathbb{B}(x, r)$ satisfies $\lambda_M(A) = \lambda_{M'}(A) = 0$.*

Proof of Corollary 98. Writing $k = m - d$, we see that up to ambient diffeomorphism — which preserves the nullity of measure — we can assume that locally around x, M' coincides with $\mathbb{R}^d \times \{0\}^k$ and that M is the graph of a $\mathcal{C}^\infty$ function $g : \mathbb{B}_{\mathbb{R}^d}(0, r') \to \mathbb{R}^k$ for $r' > 0$ small enough. The assumption $T_x M \neq T_x M'$ translates to $d_0 g \neq 0$, and the previous transformation maps smoothly $M \cap M' \cap \mathbb{B}(x, r'')$ to $g^{-1}(0) \cap \mathbb{B}(0, r')$ for $r'' > 0$ small enough. We conclude by applying Lemma 97. □

We are now in position to prove Lemma 47.

Proof of Lemma 47. Notice that Q and Q' are dominated by the measure $\mu = \mathbb{1}_{M\cup M'}\mathcal{H}^d$, with $dQ(x) = f(x)d\mu(x)$ and $dQ'(x) = f'(x)d\mu(x)$, where $f, f' : \mathbb{R}^m \to \mathbb{R}_+$ have support M and M' respectively. On the other hand, P and P' are dominated by $\nu(dx\,dT) = \delta_{\{T_xM, T_xM'\}}(dT)\,\mu(dx)$ with respective densities $\bar{f}(x,T) = \mathbb{1}_{T=T_xM}f(x)$ and $\bar{f}'(x,T) = \mathbb{1}_{T=T_xM'}f'(x)$, where we set arbitrarily $T_xM = T_0$ for $x \notin M$, and $T_xM' = T_0$ for $x \notin M'$. Recalling that f vanishes outside M and f' outside M',

$$\begin{aligned} TV(P,P') &= \frac{1}{2}\int_{\mathbb{R}^m\times\mathbb{G}^{d,m}} |\bar{f} - \bar{f}'|d\nu \\ &= \frac{1}{2}\int_{\mathbb{R}^m} \mathbb{1}_{T_xM=T_xM'}|f(x) - f'(x)| + \mathbb{1}_{T_xM\neq T_xM'}(f(x)+f'(x))\mathcal{H}^d(dx). \end{aligned}$$

From Corollary 98 and a straightforward compactness argument, we derive that

$$\mathcal{H}^d\left(M \cap M' \cap \{x | T_xM \neq T_xM'\}\right) = 0.$$

As a consequence, the above integral expression becomes

$$TV(P,P') = \frac{1}{2}\int_{\mathbb{R}^m} |f - f'|d\mathcal{H}^d = TV(Q,Q'),$$

which concludes the proof. □

B.4.3 Construction of the Hypotheses

This section is devoted to the construction of hypotheses that will be used in Le Cam's lemma (Lemma 46), to derive Proposition 33 and Theorem 50.

Lemma 99. *Let $R, \ell, \eta > 0$ be such that $\ell \leq \frac{R}{2} \wedge \left(2^{1/d} - 1\right)$ and $\eta \leq \frac{\ell^2}{2R}$. Then there exists a d-dimensional sphere of radius R that we call M, such that $M \in \mathcal{M}^{d,m}_{R,\frac{1}{R^2}}$ and a global $\mathcal{C}^\infty$-diffeomorphism $\Phi : \mathbb{R}^m \to \mathbb{R}^m$ such that,*

$$\|d\Phi - I_D\|_{op} \leq \frac{3\eta}{\ell}, \quad \left\|d^2\Phi\right\|_{op} \leq \frac{23\eta}{\ell^2}, \quad \left\|d^3\Phi\right\|_{op} \leq \frac{573\eta}{\ell^3},$$

and so that writing $M' = \Phi(M)$, we have $\mathcal{H}^d(M') \leq 2\mathcal{H}^d(M) = 2\sigma_d R^d$,

$$\left|\frac{1}{\tau_M} - \frac{1}{\tau_{M'}}\right| \geq \frac{\eta}{\ell^2}, \quad \textit{and} \quad TV(\lambda_M, \lambda_{M'}) \leq 12\left(\frac{\ell}{R}\right)^d.$$

Proof of Lemma 99. Let $M \subset \mathbb{R}^{d+1}\times\{0\}^{m-d-1} \subset \mathbb{R}^m$ be the sphere of radius R with center $(0, -R, 0, \ldots, 0)$. The reach of M is $\tau_M = R$, and its arc-length parametrized geodesics are arcs of great circles, which have third derivatives of constant norm $\|\gamma'''(t)\| = \frac{1}{R^2}$. Hence we see that $M \in \mathcal{M}^{d,m}_{R,\frac{1}{R^2}}$. Let $\phi : \mathbb{R}^m \to \mathbb{R}_+$ be the map defined by $\phi(x) = \exp\left(\frac{\|x\|^2}{\|x\|^2-1}\right)\mathbb{1}_{\|x\|^2<1}$. ϕ is a symmetric $\mathcal{C}^\infty$ map with support equal to $\mathbb{B}(0,1)$ and elementary real analysis yields $\phi(0) = 1$, $\|d\phi\|_{op} \leq 3$, $\|d^2\phi\|_{op} \leq 23$ and $\|d^3\phi\|_{op} \leq 573$. Let $\Phi : \mathbb{R}^m \to \mathbb{R}^m$ be defined by

$$\Phi(x) = x + \eta\phi\left(x/\ell\right)\cdot v,$$

where $v = (0, 1, 0, \ldots, 0)$ is the unit vertical vector. Φ is the identity map on $\mathbb{B}\left(0,\ell\right)^c$, and in $\mathbb{B}\left(0,\ell\right)$, Φ translates points on the vertical axis with a magnitude modulated by the weight function $\phi(x/\ell)$.

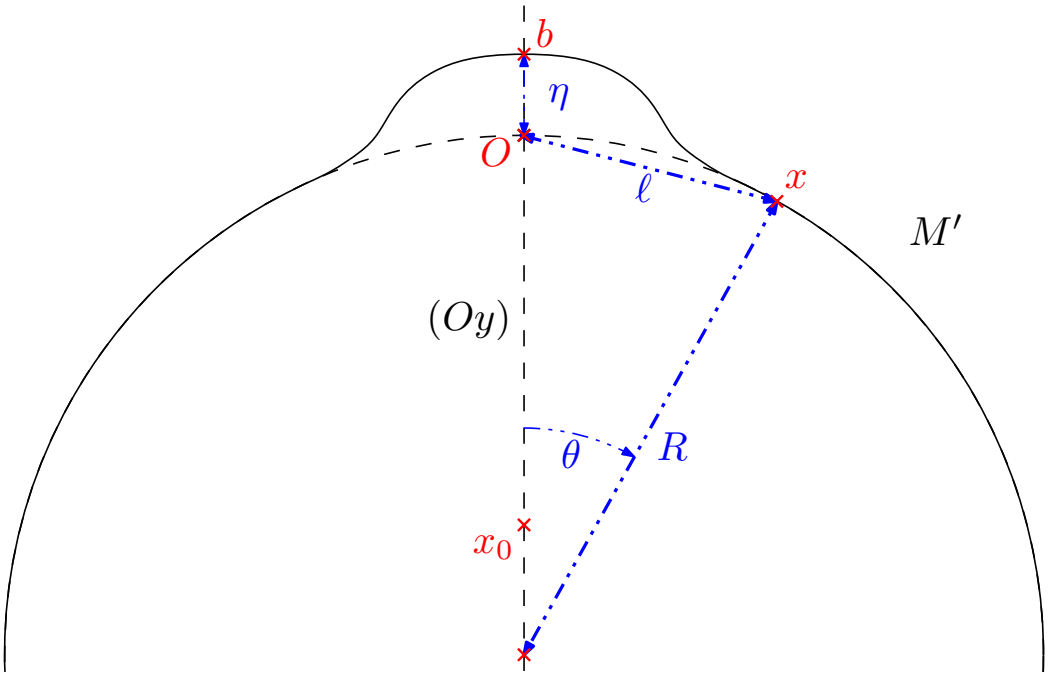


Figure B.4: The bumped sphere M' of Lemma 99.

From chain rule, $\|d\Phi - I_D\|_{op} = \eta \|d\phi\|_\infty / \ell \leq 3\eta/\ell < 1$. Therefore, $d_x\Phi$ is invertible for all $x \in \mathbb{R}^m$, so that Φ is a local $\mathcal{C}^\infty$-diffeomorphism according to the local inverse function theorem. Moreover, $\|\Phi(x)\| \to \infty$ as $\|x\| \to \infty$, so that Φ is a global $\mathcal{C}^\infty$-diffeomorphism by Hadamard-Cacciopoli theorem De Marco et al. [1994]. Similarly, from bounds on differentials of ϕ we get

$$\left\|d^2\Phi\right\|_{op} \leq 23\frac{\eta}{\ell^2} \quad \text{and} \quad \left\|d^3\Phi\right\|_{op} \leq 573\frac{\eta}{\ell^3}.$$

Let us now write $M' = \Phi(M)$ for the image of M by the map Φ (see Figure B.4). Denote by (Oy) the vertical axis $\mathrm{span}(v)$, and notice that since ϕ is symmetric, M' is symmetric with respect to the vertical axis (Oy). We now bound from above the reach $\tau_{M'}$ of M' by showing that the point $x_0 = \left(0, \frac{R+\eta/2}{1+\frac{\ell^2}{2R\eta}}, 0, \ldots, 0\right)$ belongs to its medial axis $Med(M')$ (see (1.5)). For this, write

$$b = (0, \eta, 0, \ldots, 0), \quad b' = (0, -2R, 0, \ldots, 0),$$

together with $\theta = \arccos(1 - \ell^2/(2R^2))$, and

$$x = (R\sin\theta, R\cos\theta - R, 0, \ldots, 0).$$

By construction, b, b' and x belong to M'. One easily checks that $\|x_0 - x\| < \|x_0 - b\|$ and $\|x_0 - x\| < \|x_0 - b'\|$, so that neither b nor b' is the nearest neighbor of x_0 on M'. But $x_0 \in (Oy)$ which is an axis of symmetry of M', and $(Oy) \cap M' = \{b, b'\}$. As a consequence, x_0 has strictly more than one nearest

neighbor on M'. That is, x_0 belongs to the medial axis $Med(M')$ of M'. Therefore,

$$\begin{aligned}\frac{1}{\tau_{M'}} \geq \frac{1}{d\left(x_0, M'\right)} &\geq \frac{1}{\|x_0 - x\|} \\ &\geq \frac{1}{R\left|1 - \frac{\ell^2}{2R^2} - \frac{1+\frac{\eta}{2R}}{1+\frac{\ell^2}{2R\eta}}\right|} \\ &\geq \frac{1}{R\left(1 - \frac{1+\frac{\eta}{2R}}{1+\frac{\ell^2}{2R\eta}}\right)} \geq \frac{1}{R}\left(1 + \frac{1+\frac{\eta}{2R}}{1+\frac{\ell^2}{2R\eta}}\right) \geq \frac{1}{R} + \frac{\eta}{\ell^2},\end{aligned}$$

which yields the bound $\left|\frac{1}{\tau_M} - \frac{1}{\tau_{M'}}\right| = \left|\frac{1}{R} - \frac{1}{\tau_{M'}}\right| \geq \frac{\eta}{\ell^2}$.

Finally, since $M' = \Phi(M)$ with $\|d\Phi - I_D\|_{op} \leq 2^{1/d} - 1$ with $\Phi_{|\mathbb{B}(0,\ell)^c}$ coinciding with the identity map, Lemma 96 yields $\mathcal{H}^d(M') \leq 2\mathcal{H}^d(M) = 2\sigma_d R^d$ and

$$\begin{aligned}TV\left(\lambda_M, \lambda_{M'}\right) &\leq 12\lambda_M\left(\mathbb{B}(0,\ell)\right) \\ &= 12\frac{\mathcal{H}^d\left(\mathbb{B}_{\mathcal{S}^d}\left(0, 2\arcsin\left(\frac{\ell}{2R}\right)\right)\right)}{\mathcal{H}^d\left(\mathcal{S}^d\right)} \\ &\leq 12\left(\frac{\ell}{R}\right)^d,\end{aligned}$$

which concludes the proof. □

Proof of Proposition 49. Apply Lemma 99 with $R = 2\tau_{min}$. Then the sphere M of radius $2\tau_{min}$ belongs to $\mathcal{M}^{d,m}_{2\tau_{min},1/(4\tau^2_{min})}$. Furthermore, taking $\eta = c_d\ell^3/\tau^2_{min}$ for $c_d > 0$ and $\ell > 0$ small enough, Proposition 48 (applied to the unit sphere, yielding c_d, and reasoning by homogeneity for the sphere of radius $2\tau_{min}$) asserts that $M' = \Phi(M)$ belongs to $\mathcal{M}^{d,m}_{\tau_{min},1/(2\tau^2_{min})} \subset \mathcal{M}^{d,m}_{\tau_{min},L}$, since $L \geq 1/(2\tau^2_{min})$. Moreover,

$$\mathcal{H}^d(M')^{-1} \wedge \mathcal{H}^d(M)^{-1} \geq \left(2^{d+1}\sigma_d\tau^d_{min}\right)^{-1} \geq f_{min},$$

so that $\lambda_M, \lambda_{M'} \in \mathcal{Q}^{d,m}_{\tau_{min},L,f_{min}}$, which gives the result. □

Let us now prove the minimax inconsistency of the reach estimation for $L = \infty$, using the same technique as above.

Proof of Proposition 33. Let M and M' be given by Lemma 99 with $\ell \leq \frac{R}{2} \wedge (2^{1/d} - 1)$, $\eta = \ell^2/(23R)$ and $R = 2\tau_{min}$. We have $\|d\Phi - I_D\|_{op} \leq 3\eta/\ell \leq 0.1$ and $\|d^2\Phi\|_{op} \leq 23\eta/\ell^2 \leq 1/(2\tau_{min})$. Since $\tau_M \geq 2\tau_{min}$, Lemma 95 yields

$$\tau_{M'} \geq \frac{\tau_M(1 - \|d\Phi - I_D\|_{op})^2}{\|d^2\Phi\|_{op}\tau_M + (1 + \|d\Phi - I_D\|_{op})} \geq \tau_{min}.$$

As a consequence, M and M' belong to $\mathcal{M}^{d,m}_{\tau_{min},L=\infty}$. Furthermore, since we have $f_{min} \leq \left(2^{d+1}\tau^d_{min}\sigma_d\right)^{-1} \leq \mathcal{H}^d(M)^{-1} \wedge \mathcal{H}^d(M')^{-1}$, we see that the uniform distributions $\lambda_M, \lambda_{M'}$ belong to $\mathcal{Q}^{d,m}_{\tau_{min},L=\infty,f_{min}}$. Let now P, P' denote the distributions of $\mathcal{P}^{d,m}_{\tau_{min},L=\infty,f_{min}}$ associated to $\lambda_M, \lambda_{M'}$ (Definition 32). Lemma 47

asserts that $TV(P, P') = TV(\lambda_M, \lambda_{M'})$. Applying Lemma 46 to P, P', we get that for all $n \geq 1$, for ℓ small enough,

$$\begin{aligned}
\inf_{\hat{\tau}_n} \sup_{P \in \mathcal{P}^{d,m}_{\tau_{min}, L=\infty, f_{min}}} \mathbb{E}_{P^n} \left| \frac{1}{\tau_P} - \frac{1}{\hat{\tau}_n} \right|^p &\geq \frac{1}{2^p} \left| \frac{1}{\tau_M} - \frac{1}{\tau_{M'}} \right|^p (1 - TV(P, P'))^n \\
&\geq \frac{1}{2^p} \left(\frac{\eta}{\ell^2} \right)^p \left(1 - 12 \left(\frac{\ell}{2\tau_{min}} \right)^d \right)^n \\
&= \frac{1}{2^p} \left(\frac{1}{46\tau_{min}} \right)^p \left(1 - 12 \left(\frac{\ell}{2\tau_{min}} \right)^d \right)^n .
\end{aligned}$$

Sending $\ell \to 0$ with $n \geq 1$ fixed yields the announced result. □

Appendix C

Appendix for Chapter 4

C.1 Topological Preliminaries

The goal of this section is to define an appropriate topology on the cluster tree T_f in Definition 51. Defining an appropriate topology for the cluster tree T_f is important in Chapter 4 for several reasons: (1) the topology gives geometric insight for the cluster tree, (2) homeomorphism (topological equivalence) is connected to equivalence in the partial order $\preceq$ in Definition 54, and (3) the topology gives a justification for using a fixed bandwidth h for constructing confidence set $\hat{C}_\alpha$ as in Lemma 56 to obtain faster rates of convergence.

We construct the topology of the cluster tree T_f by imposing a topology on the corresponding collection of connected components $\{T_f\}$ in Definition 51. For defining a topology on $\{T_f\}$, we define the tree distance function d_{T_f} in Definition 100, and impose the metric topology induced from the tree distance function. Using a distance function for topology not only eases formulating topology but also enables us to inherit all the good properties of the metric topology.

The desired tree distance function $d_{T_f} : \{T_f\} \times \{T_f\} \to [0, \infty)$ is based on the merge height function m_f in Definition 52. For later use in the proof, we define the tree distance function d_{T_f} on both $\mathbb{X}$ and $\{T_f\}$ as follows:

Definition 100. Let $f : \mathbb{X} \to [0, \infty)$ be a function, and T_f be its cluster tree in Definition 51. For any two points $x, y \in \mathbb{X}$, the tree distance function $d_{T_f} : \mathbb{X} \times \mathbb{X} \to [0, \infty)$ of T_f on $\mathbb{X}$ is defined as

$$d_{T_f}(x, y) = f(x) + f(y) - 2m_f(x, y).$$

Similarly, for any two clusters $C_1, C_2 \in \{T_f\}$, we first define $\lambda_1 = \sup\{\lambda : C_1 \in T_f(\lambda)\}$, and λ_2 analogously. We then define the tree distance function $d_{T_f} : \{T_f\} \times \{T_f\} \to [0, \infty)$ of T_f on $\mathbb{X}$ as:

$$d_{T_f}(C_1, C_2) = \lambda_1 + \lambda_2 - 2m_f(C_1, C_2).$$

The tree distance function d_{T_f} in Definition 52 is a pseudometric on $\mathbb{X}$ and is a metric on $\{T_f\}$ as desired, proven in Lemma 101. The proof is given later in Appendix C.5.

Lemma 101. *Let $f : \mathbb{X} \to [0, \infty)$ be a function, T_f be its cluster tree in Definition 51, and d_{T_f} be its tree distance function in Definition 100. Then d_{T_f} on $\mathbb{X}$ is a pseudometric and d_{T_f} on $\{T_f\}$ is a metric.*

From the metric d_{T_f} on $\{T_f\}$ in Definition 100, we impose the induced metric topology on $\{T_f\}$. We say T_f is homeomorphic to T_g, or $T_f \cong T_g$, when their corresponding collection of connected components are homeomorphic, i.e. $\{T_f\} \cong \{T_g\}$. (Two spaces are homeomorphic if there exists a bijective continuous function between them, with a continuous inverse.)

To get some geometric understanding of the cluster tree in Definition 51, we identify edges that constitute the cluster tree. Intuitively, edges correspond to either leaves or internal branches. An edge is roughly defined as a set of clusters whose inclusion relationship with respect to clusters outside an edge are equivalent, so that when the collection of connected components is divided into edges, we observe the same inclusion relationship between representative clusters whenever any cluster is selected as a representative for each edge.

For formally defining edges, we define an interval in the cluster tree and the equivalence relation in the cluster tree. For any two clusters $A, B \in \{T_f\}$, the interval $[A, B] \subset \{T_f\}$ is defined as a set clusters that contain A and are contained in B, i.e.

$$[A, B] := \{C \in \{T_f\} : A \subset C \subset B\},$$

The equivalence relation $\sim$ is defined as $A \sim B$ if and only if their inclusion relationship with respect to clusters outside $[A, B]$ and $[B, A]$, i.e.

$A \sim B$ if and only if
for all $C \in \{T_f\}$ such that $C \notin [A, B] \cup [B, A]$, $C \subset A$ iff $C \subset B$ and $A \subset C$ iff $B \subset C$.

Then it is easy to see that the relation $\sim$ is reflexive($A \sim A$), symmetric($A \sim B$ implies $B \sim A$), and transitive ($A \sim B$ and $B \sim C$ implies $A \sim C$). Hence the relation $\sim$ is indeed an equivalence relation, and we can consider the set of equivalence classes $\{T_f\}/_\sim$. We define the edge set $E(T_f)$ as $E(T_f) := \{T_f\}/_\sim$.

For later use, we define the partial order on the edge set $E(T_f)$) as follows: $[C_1] \leq [C_2]$ if and only if for all $A \in [C_1]$ and $B \in [C_2]$, $A \subset B$. We say that a tree T_f is finite if its edge $E(T_f)$ is a finite set.

C.2 The Partial Order

As discussed in Section 4.1, to see that the partial order $\preceq$ in Definition 54 is indeed a partial order, we need to check the reflexivity, the transitivity, and the antisymmetry. The reflexivity and the transitivity are easier to check, but to show antisymmetric, we need to show that if two trees T_f and T_g satisfies $T_f \preceq T_g$ and $T_g \preceq T_f$, then T_f and T_g are equivalent in some sense. And we give the equivalence relation as the topology on the cluster tree defined in Appendix C.1. The argument is formally stated in Lemma 102. The proof is done later in Appendix C.5.

Lemma 102. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. Then if f, g are continuous and T_f, T_g are finite, $T_f \preceq T_g$ and $T_g \preceq T_f$ implies that there exists a homeomorphism $\Phi : \{T_f\} \to \{T_g\}$ that preserves the root, i.e. $\Phi(\mathbb{X}) = \mathbb{X}$. Conversely, if there exists a homeomorphism $\Phi : \{T_f\} \to \{T_g\}$ that preserves the root, $T_f \preceq T_g$ and $T_g \preceq T_f$ hold.*

The partial order $\preceq$ in Definition 54 gives a formal definition of simplicity of trees, and it is used to justify pruning schemes in Section 4.3.2. Hence it is important to match the partial order $\preceq$ with the intuitive notions of the complexity of the tree. We provided three arguments in Section 4.1: (1) if $T_f \preceq T_g$ holds then it must be the case that (number of edges of T_f) $\leq$ (number of edges of T_g), (2) if T_g can be obtained from T_f by adding edges, then $T_f \preceq T_g$ holds, and (3) the existence of a topology preserving embedding from $\{T_f\}$ to $\{T_g\}$ implies the relationship $T_f \preceq T_g$. We formally state each item in Lemma 103, 104, and 105. Proofs of these lemmas are done later in Appendix C.5.

Lemma 103. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. Suppose $T_f \preceq T_g$ via $\Phi : \{T_f\} \to \{T_g\}$. Define $\bar{\Phi} : E(T_f) \to E(T_g)$ by for $[C] \in E(T_f)$*

choosing any $C \in [C]$ and defining as $\bar{\Phi}([C]) = [\Phi(C)]$. Then $\bar{\Phi}$ is injective, and as a consequence, $|E(T_f)| \leq |E(T_g)|$.

Lemma 104. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. If T_g can be obtained from T_f by adding edges, then $T_f \preceq T_g$ holds.*

Lemma 105. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. If there exists a one-to-one map $\Phi : \{T_f\} \to \{T_g\}$ that is a homeomorphism between $\{T_f\}$ and $\Phi(\{T_f\})$ and preserves the root, i.e. $\Phi(\mathbb{X}) = \mathbb{X}$, then $T_f \preceq T_g$ holds.*

C.3 Hadamard Differentiability

Definition 106 (see page 281 of Wellner [2013])**.** Let $\mathbb{D}$ and $\mathbb{E}$ be normed spaces and let $\phi : \mathbb{D}_\phi \to \mathbb{E}$ be a map defined on a subset $\mathbb{D}_\phi \subset \mathbb{D}$. Then ϕ is Hadamard differentiable at θ if there exists a continuous, linear map $\phi'_\theta : \mathbb{D} \to \mathbb{E}$ such that

$$\left\| \frac{\phi(\theta + tq_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\|_{\mathbb{E}} \to 0$$

as $t \to 0$, for every $q_t \to q$.

Hadamard differentiability is a key property for bootstrap inference since it is a sufficient condition for the delta method; for more details, see section 3.1 of Wellner [2013]. Recall that d_{MM} is based on the function $d_{T_p}(x, y) = p(x) + p(y) - 2m_p(x, y)$. The following theorem shows that the function d_{T_p} is not Hadamard differentiable for some pairs (x, y). In our case $\mathbb{D}$ is the set of continuous functions on the sample space, $\mathbb{E}$ is the real line, $\theta = p$, $\phi(p)$ is $d_{T_p}(x, y)$ and the norm on $\mathbb{E}$ is the usual Euclidean norm.

Theorem 107. *Let $B(x)$ be the smallest set $B \in T_p$ such that $x \in B$. $d_{T_p}(x, y)$ is not Hadamard differentiable for $x \neq y$ when one of the following two scenarios occurs:*

(i) $\min\{p(x), p(y)\} = p(c)$ *for some critical point c.*

(ii) $B(x) = B(y)$ *and* $p(x) = p(y)$.

The merge distortion metric d_{M} is also not Hadamard differentiable.

C.4 Confidence Sets Constructions

C.4.1 Regularity conditions on the kernel

To apply the results in Chernozhukov et al. [2016] which imply that the bootstrap confidence set is consistent, we consider the following two assumptions.

(K1) The kernel function K has the bounded second derivative and is symmetric, non-negative, and

$$\int x^2 K(x) dx < \infty, \qquad \int K(x)^2 dx < \infty.$$

(K2) The kernel function K satisfies

$$\mathcal{K} = \left\{ y \mapsto K\left(\frac{x - y}{h}\right) : x \in \mathbb{R}^d, h > 0 \right\}. \tag{C.1}$$

We require that $\mathcal{K}$ satisfies

$$\sup_P N\left(\mathcal{K}, L_2(P), \epsilon\|F\|_{L_2(P)}\right) \leq \left(\frac{A}{\epsilon}\right)^v \tag{C.2}$$

for some positive numbers A and v, where $N(T, d, \epsilon)$ denotes the ϵ-covering number of the metric space (T, d), F is the envelope function of $\mathcal{K}$, and the supremum is taken over the whole $\mathbb{R}^d$. The A and v are usually called the VC characteristics of $\mathcal{K}$. The norm $\|F\|^2_{L_2(P)} = \int |F(x)|^2 dP(x)$.

Assumption (K1) is to ensure that the variance of the KDE is bounded and p_h has the bounded second derivative. This assumption is very common in statistical literature, see e.g. Wasserman [2006], Scott [2015]. Assumption (K2) is to regularize the complexity of the kernel function so that the supremum norm for kernel functions and their derivatives can be bounded in probability. A similar assumption appears in Einmahl and Mason [2005] and Genovese et al. [2014]. The Gaussian kernel and most compactly supported kernels satisfy both assumptions.

C.4.2 Pruning

The goal of this section is to formally define the pruning scheme in Section 4.3.2. Note that when pruning leaves and internal branches, when the cumulative length is computed for each leaf and internal branch, then the pruning process can be done at once. We provide two pruning schemes in Section 4.3.2 in a unifying framework by defining an appropriate notion of lifetime for each edge, and deleting all insignificant edges with small lifetimes. To follow the pruning schemes in Section 4.3.2, we require that the lifetime of a child edge is shorter than the lifetime of a parent edge, so that we can delete edges from the top. We evaluate the lifetime of each edge by an appropriate nonnegative (possibly infinite) function life. We formally define the pruned tree $Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h)$ as follows:

Definition 108. Suppose the function life : $E(\hat{T}_h) \to [0, +\infty]$ satisfies that $[C_1] \leq [C_2] \implies \text{life}([C_1]) \subset \text{life}([C_2])$. We define the pruned tree $Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h) : \mathbb{R} \to 2^{\mathbb{X}}$ as

$$Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h)(\lambda) = \left\{C \in \hat{T}_h(\lambda - \hat{t}_\alpha) : \text{life}([C]) > \hat{t}_\alpha\right\}.$$

We suggest two life functions corresponding to two pruning schemes in Section 4.3.2. We first need several definitions. For any $[C] \in E(\hat{T}_h)$, define its level as

$$level([C]) := \left\{\lambda : \text{ there exists } A \in [C] \cap \hat{T}_h(\lambda)\right\},$$

and define its cumulative level as

$$cumlevel([C]) := \left\{\lambda : \text{ there exists } A \in \hat{T}_h(\lambda),\ B \in [C] \text{ such that } A \subset B\right\}.$$

Then life^{leaf} corresponds to first pruning scheme in Section 4.3.2, which is to prune out only insignificant leaves.

$$\text{life}^{leaf}([C]) = \begin{cases} \sup\{level([C])\} - \inf\{level([C])\} & \text{if } \inf\{level([C])\} \neq \inf\{cumlevel([C])\} \\ +\infty & \text{otherwise.} \end{cases}.$$

And life^{top} corresponds to second pruning scheme in Section 4.3.2, which is to prune out insignificant edges from the top.

$$\text{life}^{top}([C]) = \sup\{cumlevel([C])\} - \inf\{cumlevel([C])\}.$$

Note that life^{leaf} is lower bounded by life^{top}. In fact, for any life function that is lower bounded by life^{top}, the pruned tree $Pruned_{\text{life},\hat{t}_\alpha}$ is a valid tree in the confidence set $\hat{C}_\alpha$ that is simpler than the original estimate $\hat{T}_h$, so that the pruned tree is the desired tree as discussed in Section 4.3.2. We formally state as follows. The proof is given in Appendix C.7.

Lemma 109. *Suppose that the* life *function satisfies: for all* $[C] \in E(\hat{T}_h)$, $\text{life}^{top}([C]) \leq \text{life}([C])$. *Then*

(i) $Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h) \preceq T_{\hat{p}_h}$.

(ii) there exists a function $\tilde{p}$ *such that* $T_{\tilde{p}} = Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h)$.

(iii) $\tilde{p}$ *in (ii) satisfies* $\tilde{p} \in \hat{C}_\alpha$.

Remark: It can be shown that complete pruning — simultaneously removing all leaves and branches with length less than $2\hat{t}_\alpha$ — can in general yield a tree that is outside the confidence set. For example, see Figure 4.3. If we do complete pruning to this tree, we will get the trivial tree.

C.5 Proofs for Appendix C.1 and C.2

C.5.1 Proof of Lemma 101

Lemma 101. *Let* $f : \mathbb{X} \to [0, \infty)$ *be a function,* T_f *be its cluster tree in Definition 51, and* d_{T_f} *be its tree distance function in Definition 100. Then* d_{T_f} *on* $\mathbb{X}$ *is a pseudometric and* d_{T_f} *on* T_f *is a metric.*

Proof. First, we show that d_{T_f} on $\mathbb{X}$ is a pseudometric. To do this, we need to show non-negativity($d_{T_f}(x, y) \geq 0$), $x = y$ implying $d_{T_f}(x, y) = 0$, symmetry($d_{T_f}(x, y) = d_{T_f}(y, x)$), and subadditivity($d_{T_f}(x, y) + d_{T_f}(y, z) \leq d_{T_f}(x, z)$).

For non-negativity, note that for all $x, y \in \mathbb{X}$, $m_f(x, y) \leq \min\{f(x), f(y)\}$, so

$$d_{T_f}(x, y) = f(x) + f(y) - 2m_f(x, y) \geq 0. \tag{C.3}$$

For $x = y$ implying $d_{T_f}(x, y) = 0$, $x = y$ implies $m_f(x, y) = f(x) = f(y)$, so

$$x = y \Longrightarrow d_{T_f}(x, y) = 0. \tag{C.4}$$

For symmetry, since $m_f(x, y) = m_f(y, x)$,

$$d_{T_f}(x, y) = d_{T_f}(y, x). \tag{C.5}$$

For subadditivity, note first that $m_f(x, y) \leq f(y)$ and $m_f(y, z) \leq f(y)$ holds, so

$$\max\{m_f(x, y),\, m_f(y, z)\} \leq f(y). \tag{C.6}$$

And also note that there exists $C_{xy}, C_{yz} \in T_f$ $(\min\{m_f(x, y),\, m_f(y, z)\})$ that satisfies $x, y \in C_{xy}$ and $y, z \subset C_{yz}$. Then $y \in C_{xy} \cap C_{yz} \neq \emptyset$, so $x, z \in C_{xy} = C_{yz}$. Then from definition of $m_f(x, z)$, this implies that

$$\min\{m_f(x, y),\, m_f(y, z)\} \leq m_f(x, z). \tag{C.7}$$

And by applying (C.6) and (C.7), $d_{T_f}(x, y) + d_{T_f}(y, z)$ is upper bounded by $d_{T_f}(x, z)$ as

$$\begin{aligned}
&d_{T_f}(x, y) + d_{T_f}(y, z) \\
&= f(x) + f(y) - 2m_f(x, y) + f(y) + f(z) - 2m_f(y, z) \\
&= f(x) + f(z) - 2\left(\min\{m_f(x, y),\, m_f(y, z)\} + \max\{m_f(x, y),\, m_f(y, z)\} - f(y)\right) \\
&\geq f(x) + f(z) - 2m_f(x, z) \\
&= d_{T_f}(x, z).
\end{aligned} \tag{C.8}$$

Hence (C.3), (C.4), (C.5), and (C.8) implies that d_{T_f} on $\mathbb{X}$ is a pseudometric.

Second, we show that d_{T_f} on T_f is a metric. To do this, we need to show non-negativity($d_{T_f}(x, y) \geq 0$), identity of indiscernibles($x = y \iff d_{T_f}(x, y) = 0$), symmetry($d_{T_f}(x, y) = d_{T_f}(y, x)$), and subadditivity($d_{T_f}(x, y) + d_{T_f}(y, z) \leq d_{T_f}(x, z)$).

For nonnegativity, note that if $C_1 \in T_f(\lambda_1)$ and $C_2 \in T_f(\lambda_2)$, then $m_f(C_1, C_2) \leq \min\{\lambda_1, \lambda_2\}$, so

$$d_{T_f}(C_1, C_2) = \lambda_1 + \lambda_2 - 2m_f(C_1, C_2) \geq 0. \tag{C.9}$$

For identity of indiscernibles, $C_1 = C_2$ implies $m_f(C_1, C_2) = \lambda_1 = \lambda_2$, so

$$C_1 = C_2 \Longrightarrow d_{T_f}(C_1, C_2) = 0. \tag{C.10}$$

And conversely, $d_{T_f}(C_1, C_2) = 0$ implies $\lambda_1 = \lambda_2 = m_f(C_1, C_2)$, so there exists $C \in T_f(\lambda_1)$ such that $C_1 \subset C$ and $C_2 \subset C$. Then since $C_1, C_2, C \in T_f(\lambda_1)$, so $C_1 \cap C \neq \emptyset$ implies $C_1 = C$ and similarly $C_2 = C$, so

$$d_{T_f}(C_1, C_2) = 0 \Longrightarrow C_1 = C_2. \tag{C.11}$$

Hence (C.10) and (C.11) implies identity of indiscernibles as

$$C_1 = C_2 \iff d_{T_f}(C_1, C_2) = 0. \tag{C.12}$$

For symmetry, since $m_f(C_1, C_2) = m_f(C_2, C_1)$,

$$d_{T_f}(C_1, C_2) = d_{T_f}(C_2, C_1). \tag{C.13}$$

For subadditivity, note that $m_f(C_1, C_2) \leq \lambda_2$ and $m_f(C_2, C_3) \leq \lambda_2$ holds, so

$$\max\{m_f(C_1, C_2),\, m_f(C_2, C_3)\} \leq \lambda_2. \tag{C.14}$$

And also note that there exists $C_{12}, C_{23} \in T_f\left(\min\{m_f(C_1, C_2),\, m_f(C_2, C_3)\}\right)$ that satisfies $C_1, C_2 \subset C_{12}$ and $C_2, C_3 \subset C_{23}$. Then $C_2 \subset C_{12} \cap C_{23} \neq \emptyset$, so $C_1, C_3 \in C_{12} = C_{23}$. Then from definition of $m_f(C_1, C_3)$, this implies that

$$\min\{m_f(C_1, C_2),\, m_f(C_2, C_3)\} \leq m_f(C_1, C_3). \tag{C.15}$$

And by applying (C.14) and (C.15), $d_{T_f}(C_1, C_2) + d_{T_f}(C_2, C_3)$ is upper bounded by $d_{T_f}(C_1, C_3)$ as

$$\begin{aligned}
&d_{T_f}(C_1, C_2) + d_{T_f}(C_2, C_3) \\
&= \lambda_1 + \lambda_2 - 2m_f(C_1, C_2) + \lambda_2 + \lambda_3 - 2m_f(C_2, C_3) \\
&= \lambda_1 + \lambda_3 - 2\left(\min\{m_f(C_1, C_2),\, m_f(C_2, C_3)\} + \max\{m_f(C_1, C_2),\, m_f(C_2, C_3)\} - \lambda_2\right) \\
&\geq \lambda_1 + \lambda_3 - 2m_f(C_1, C_3) \\
&= d_{T_f}(C_1, C_3).
\end{aligned} \tag{C.16}$$

Hence (C.9), (C.12), (C.13), and (C.16) d_{T_f} on T_f is a metric.

□

C.5.2 Proof of Lemma 102

Lemma 102. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. Then if f, g are continuous and T_f, T_g are finite, $T_f \preceq T_g$ and $T_g \preceq T_f$ implies that there exists a homeomorphism $\Phi : T_f \to T_g$ that preserves the root, i.e. $\Phi(\mathbb{X}) = \mathbb{X}$. Conversely, if there exists a homeomorphism $\Phi : T_f \to T_g$ that preserves the root, $T_f \preceq T_g$ and $T_g \preceq T_f$ hold.*

Proof. First, we show that $T_f \preceq T_g$ and $T_g \preceq T_f$ implies homeomorphism. Let $\Phi : T_f \to T_g$ be the map that gives the partial order $T_f \preceq T_g$ in Definition 54. Then from Lemma 103, $\bar{\Phi} : E(T_f) \to E(T_g)$ is injective and $|E(T_f)| \leq |E(T_g)|$. With a similar argument, $|E(T_g)| \leq |E(T_f)|$ holds, so

$$|E(T_f)| = |E(T_g)|.$$

Since we assumed that T_f and T_g are finite, i.e. $|E(T_f)|$ and $|E(T_g)|$ are finite, $\bar{\Phi}$ becomes a bijection.

Now, let $[C_1]$ and $[C_2]$ be adjacent edges in $E(T_f)$, and without loss of generality, assume $C_1 \subset C_2$. We argue below that $\bar{\Phi}([C_1])$ and $\bar{\Phi}([C_2])$ are also adjacent edges. Then $\Phi(C_1) \subset \Phi(C_2)$ holds from Definition 54, and since $\bar{\Phi}$ is bijective, $[\Phi(C_1)] = \bar{\Phi}([C_1])$ and $[\Phi(C_2)] = \bar{\Phi}([C_2])$ holds. Suppose there exists $\tilde{C}_3 \in T_g$ such that $[\tilde{C}_3] \notin \{\bar{\Phi}([C_1]), \bar{\Phi}([C_2])\}$ and $\Phi(C_1) \subset \tilde{C}_3 \subset \Phi(C_2)$. Then since $\bar{\Phi}$ is bijective, there exists $C_3 \in T_f$ such that $[\Phi(C_3)] = [\tilde{C}_3]$. Then $\Phi(C_1) \subset \tilde{C}_3 \subset \Phi(C_2)$ implies that $C_1 \subset C_3 \subset C_2$, and $\bar{\Phi}$ being a bijection implies that $[C_3] \notin \{[C_1], [C_3]\}$. This is a contradiction since $[C_1]$ and $[C_2]$ are adjacent edges. Hence there is no such $\tilde{C}_3$, and $\bar{\Phi}([C_1])$ and $\bar{\Phi}([C_2])$ are adjacent edges. Therefore, $\bar{\Phi} : E(T_f) \to E(T_g)$ is a bijective map that sends adjacent edges to adjacent edges, and also sends root edge to root edge.

Then combining $\bar{\Phi} : E(T_f) \to E(T_g)$ being bijective sending adjacent edges to adjacent edges and root edge to root edge, and f, g being continuous, the map $\bar{\Phi} : E(T_f) \to E(T_g)$ can be extended to a homeomorphism $T_g \to T_f$ that preserves the root.

Second, the part that homeomorphism implies $T_f \preceq T_g$ and $T_g \preceq T_f$ follows by Lemma 105. □

C.5.3 Proof of Lemma 103

Lemma 103. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. Suppose $T_f \preceq T_g$ via $\Phi : T_f \to T_g$. Define $\bar{\Phi} : E(T_f) \to E(T_g)$ by for $[C] \in E(T_f)$ choosing any $C \in [C]$ and defining as $\bar{\Phi}([C]) = [\Phi(C)]$. Then $\bar{\Phi}$ is injective, and as a consequence, $|E(T_f)| \leq |E(T_g)|$.*

Proof. We will first show that equivalence relation on T_g implies equivalence relation on T_f, i.e.

$$\Phi(C_1) \sim \Phi(C_2) \Longrightarrow C_1 \sim C_2. \tag{C.17}$$

Suppose $\Phi(C_1) \sim \Phi(C_2)$ in T_g. Then from Definition 54 of Φ, for any $C \in T_f$ such that $C \notin [C_1, C_2] \cup [C_2, C_1]$, $\Phi(C) \notin [\Phi(C_1), \Phi(C_2)] \cup [\Phi(C_2), \Phi(C_1)]$ holds. Then from definition of $\Phi(C_1) \sim \Phi(C_2)$,

$$\Phi(C) \subset \Phi(C_1) \text{ iff } \Phi(C) \subset \Phi(C_2) \text{ and } \Phi(C_1) \subset \Phi(C) \text{ iff } \Phi(C_2) \subset \Phi(C).$$

Then again from Definition 54 of Φ, equivalence relation holds for C_1 and C_2 holds as well, i.e.

$$C \subset C_1 \text{ iff } C \subset C_2 \text{ and } C_1 \subset C \text{ iff } C_2 \subset C.$$

Hence (C.17) is shown, and this implies that

$$\begin{aligned}\bar{\Phi}([C_1]) = \bar{\Phi}([C_2]) &\Longrightarrow [\Phi(C_1)] = [\Phi(C_2)] \\ &\Longrightarrow \Phi(C_1) \sim \Phi(C_2) \\ &\Longrightarrow C_1 \sim C_2 \\ &\Longrightarrow [C_1] = [C_2],\end{aligned}$$

so $\bar{\Phi}$ is injective. □

C.5.4 Proof of Lemma 104

Lemma 104. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. If T_g can be obtained from T_f by adding edges, then $T_f \preceq T_g$ holds.*

Proof. Since T_g can be obtained from T_f by adding edges, there is a map $\Phi : T_f \to T_g$ which preserves order, i.e. $C_1 \subset C_2$ if and only if $\Phi(C_1) \subset \Phi(C_2)$. Hence $T_f \preceq T_g$ holds. □

C.5.5 Proof of Lemma 105

Lemma 105. *Let $f, g : \mathbb{X} \to [0, \infty)$ be functions, and T_f, T_g be their cluster trees in Definition 51. If there exists a one-to-one map $\Phi : T_f \to T_g$ that is a homeomorphism between T_f and $\Phi(T_f)$ and preserves root, i.e. $\Phi(\mathbb{X}) = \mathbb{X}$, then $T_f \preceq T_g$ holds.*

Proof. For any $C \in T_f$, note that $[C, \mathbb{X}] \subset T_f$ is homeomorphic to an interval, hence $\Phi([C, \mathbb{X}]) \subset T_g$ is also homeomorphic to an interval. Since T_g is topologically a tree, an interval in a tree with fixed boundary points is uniquely determined, i.e.

$$\Phi([C, \mathbb{X}]) = [\Phi(C), \Phi(\mathbb{X})] = [\Phi(C), \mathbb{X}]. \tag{C.18}$$

For showing $T_f \preceq T_g$, we need to argue that for all $C_1, C_2 \in T_f$, $C_1 \subset C_2$ holds if and only if $\Phi(C_1) \subset \Phi(C_2)$. For only if direction, suppose $C_1 \subset C_2$. Then $C_2 \in [C_1, \mathbb{X}]$, so Definition 54 and (C.18) implies

$$\Phi(C_2) \subset \Phi([C_1, \mathbb{X}]) = [\Phi(C_1), \mathbb{X}].$$

And this implies

$$\Phi(C_1) \subset \Phi(C_2). \tag{C.19}$$

For if direction, suppose $\Phi(C_1) \subset \Phi(C_2)$. Then since $\Phi^{-1} : \Phi(T_f) \to T_f$ is also an homeomorphism with $\Phi^{-1}(\mathbb{X}) = \mathbb{X}$, hence by repeating above argument, we have

$$C_1 = \Phi^{-1}(\Phi(C_1)) \subset \Phi^{-1}(\Phi(C_2)) = C_2. \tag{C.20}$$

Hence (C.19) and (C.20) implies $T_f \preceq T_g$. □

C.6 Proofs for Section 4.2 and Appendix C.3

C.6.1 Proof of Lemma 55 and extreme cases

Lemma 55. *For any densities p and q, the following relationships hold:*

(i) When p and q are continuous, then $d_\infty(T_p, T_q) = d_{\mathrm{M}}(T_p, T_q)$.

(ii) $d_{\mathrm{MM}}(T_p, T_q) \leq 4 d_\infty(T_p, T_q)$.

(iii) $d_{\mathrm{MM}}(T_p, T_q) \geq d_\infty(T_p, T_q) - a$, where a is defined as above. Additionally when $\mu(\mathbb{X}) = \infty$, then $d_{\mathrm{MM}}(T_p, T_q) \geq d_\infty(T_p, T_q)$.

Proof. (i)

First, we show $d_M(T_p,\, T_q) \leq d_\infty(T_p,\, T_q)$. Note that this part is implicitly shown in Eldridge et al. [2015b, Proof of Theorem 6]. For all $\epsilon > 0$ and for any $x, y \in \mathbb{X}$, let $C_0 \in T_p(m_p(x,y) - \epsilon)$ with $x, y \in C_0$. Then for all $z \in C_0$, $q(z)$ is lower bounded as

$$\begin{aligned} q(z) &> p(z) - d_\infty(T_p, T_q) \\ &\geq m_p(x,y) - \epsilon - d_\infty(T_p, T_q), \end{aligned}$$

so $C_0 \subset q^{-1}\left(m_p(x,y) - \epsilon - d_\infty(T_p, T_q),\, \infty\right)$ and C_0 is connected, so x and y are in the same connected component of $q^{-1}\left(m_p(x,y) - \epsilon - d_\infty(T_p, T_q),\, \infty\right)$, which implies

$$m_q(x,y) \leq m_p(x,y) - \epsilon - d_\infty(T_p, T_q). \tag{C.21}$$

A similar argument holds for other direction as

$$m_p(x,y) \leq m_q(x,y) - \epsilon - d_\infty(T_p, T_q), \tag{C.22}$$

so (C.21) and (C.22) being held for all $\epsilon > 0$ implies

$$|m_p(x,y) - m_q(x,y)| \leq d_\infty(T_p, T_q). \tag{C.23}$$

And taking $\sup$ over all $x, y \in \mathbb{X}$ in (C.23) $d_M(T_p,\, T_q)$ is upper bounded by $d_\infty(T_p,\, T_q)$, i.e.

$$d_M(T_p,\, T_q) \leq d_\infty(T_p,\, T_q). \tag{C.24}$$

Second, we show $d_M(T_p,\, T_q) \geq d_\infty(T_p,\, T_q)$. For all $\epsilon > 0$, Let x be such that $|p(x) - q(x)| > d_\infty(T_p, T_q) - \frac{\epsilon}{2}$. Then since p and q are continuous, there exists $\delta > 0$ such that

$$\mathbb{B}(x,\, \delta) \subset p^{-1}\left(p(x) - \frac{\epsilon}{2},\, \infty\right) \cap q^{-1}\left(q(x) - \frac{\epsilon}{2},\, \infty\right).$$

Then for any $y \in \mathbb{B}(x,\, \delta)$, since $\mathbb{B}(x,\, \delta)$ is connected, $p(x) - \frac{\epsilon}{2} \leq m_p(x,y) \leq p(x)$ holds and $q(x) - \frac{\epsilon}{2} \leq m_q(x,y) \leq q(x)$, so

$$\begin{aligned} |m_p(x,y) - m_q(x,y)| &\geq |p(x) - q(x)| - \frac{\epsilon}{2} \\ &> d_\infty(T_p, T_q) - \epsilon. \end{aligned}$$

Since this holds for any $\epsilon > 0$, $d_M(T_p,\, T_q)$ is lower bounded by $d_\infty(T_p,\, T_q)$, i.e.

$$d_M(T_p,\, T_q) \geq d_\infty(T_p,\, T_q). \tag{C.25}$$

(C.24) and (C.25) implies $d_\infty(T_p, T_q) = d_{\mathrm{M}}(T_p, T_q)$.

(ii)

We have already seen that for all $x, y \in \mathbb{X}$, $|m_p(x,y) - m_q(x,y)| \leq d_\infty(T_p, T_q)$ in (C.23). Hence for all $x, y \in \mathbb{X}$,

$$\begin{aligned}
&|[p(x) + p(y) - 2m_p(x,y)] - [q(x) + q(y) - 2m_q(x,y)]| \\
&\leq |p(x) - q(x)| + |p(y) - q(y)| + 2|m_p(x,y) - m_q(x,y)| \\
&\leq 4d_\infty(T_p, T_q).
\end{aligned}$$

Since this holds for all $x, y \in \mathbb{X}$, so

$$d_{\mathrm{MM}}(T_p, T_q) \leq 4d_\infty(T_p, T_q).$$

(iii)

For all $\epsilon > 0$, Let x be such that $|p(x) - q(x)| > d_\infty(T_p, T_q) - \frac{\epsilon}{2}$, and without loss of generality assume that $p(x) > q(x)$. Let y be such that $p(y) + q(y) < \inf_x (p(x) + q(x)) + \frac{\epsilon}{2}$. Then $m_p(x,y) \leq p(y)$ holds, and since $\mathbb{X}$ is connected, $q_{\inf} \leq m_q(x,y)$ holds. Hence

$$\begin{aligned}
&[p(x) + p(y) - 2m_p(x,y)] - [q(x) + q(y) - 2m_q(x,y)] \\
&\geq [p(x) + p(y) - 2p(y)] - [q(x) + q(y) - 2q_{\inf}] \\
&= p(x) - q(x) - (p(y) + q(y) - 2q_{\inf}) \\
&> d_\infty(T_p,\, T_q) - \left(\inf_x (p(x) + q(x)) - 2q_{\inf}\right) - \epsilon \\
&\geq d_\infty(T_p,\, T_q) - a - \epsilon,
\end{aligned}$$

where $a = \inf_{x \in \mathbb{X}} (p(x) + q(x)) - 2\min\{p_{\inf},\, q_{\inf}\}$. Since this holds for all $\epsilon > 0$, we have

$$d_{\mathrm{MM}}(T_p,\, T_q) \geq d_\infty(T_p,\, T_q) - a.$$

□

Hence $0 \leq d_{\mathrm{MM}}(T_p, T_q) \leq 4d_\infty(T_p, T_q)$ holds. And both extreme cases can happen, i.e. $d_{\mathrm{MM}}(T_p, T_q) = 4d_\infty(T_p, T_q) > 0$ and $d_{\mathrm{MM}}(T_p, T_q) = 0,\ d_\infty(T_p, T_q) > 0$ can happens.

Lemma 110. *There exists densities p, q for both $d_{\mathrm{MM}}(T_p, T_q) = 4d_\infty(T_p, T_q) > 0$ and $d_{\mathrm{MM}}(T_p, T_q) = 0,\ d_\infty(T_p, T_q) > 0$.*

Proof. Let $\mathbb{X} = \mathbb{R}$, $p(x) = I(x \in [0,1])$ and $q(x) = 2I\left(x \in \left[0,\, \frac{1}{4}\right]\right) + 2I\left(x \in \left[\frac{3}{4},\, 1\right]\right)$. Then $d_\infty(T_p, T_q) = 1$. And with $x = \frac{1}{8}$ and $y = \frac{7}{8}$,

$$\begin{aligned}
|[p(x) + p(y) - 2m_p(x,y)] - [q(x) + q(y) - 2m_q(x,y)]| &= |[1 + 1 - 2] - [2 + 2 - 0]| \\
&= 4,
\end{aligned}$$

hence $d_{\mathrm{MM}}(T_p, T_q) = 4d_\infty(T_p, T_q)$.

Let $\mathbb{X} = [0,1)$, $p(x) = 2I\left(x \in \left[0,\, \frac{1}{2}\right)\right)$ and $q(x) = 2I\left(x \in \left[\frac{1}{2},\, 1\right)\right)$. Then $d_\infty(T_p, T_q) = 2$. And for any $x \in \left[0,\, \frac{1}{2}\right)$ and $y \in \left[\frac{1}{2},\, 1\right)$,

$$\begin{aligned}
|[p(x) + p(y) - 2m_p(x,y)] - [q(x) + q(y) - 2m_q(x,y)]| &= |(2 + 0 - 0) + (0 + 2 - 0)| \\
&= 0.
\end{aligned}$$

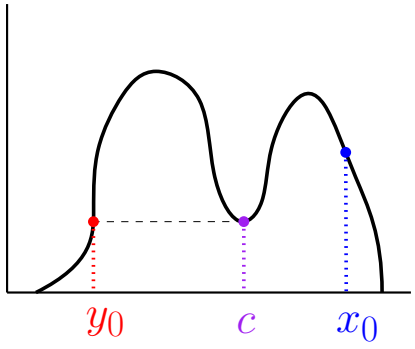


Figure C.1: The example used in the proof of Theorem 107.

A similar case holds for $x \in \left[\frac{1}{2}, 1\right)$ and $y \in \left[0, \frac{1}{2}\right)$. And for any $x, y \in \left[0, \frac{1}{2}\right)$,

$$\begin{aligned} |[p(x) + p(y) - 2m_p(x,y)] - [q(x) + q(y) - 2m_q(x,y)]| &= |(2+2-4) + (0+0-0)| \\ &= 0. \end{aligned}$$

and a similar case holds for $x, y \in \left[\frac{1}{2}, 1\right)$. Hence $d_{\text{MM}}(T_p, T_q) = 0$. □

C.6.2 Proof of Theorem 107

Theorem 107. *Let $B(x)$ be the smallest set $B \in T_p$ such that $x \in B$. $d_{T_p}(x,y)$ is not Hadamard differentiable for $x \neq y$ when one of the following two scenarios occurs:*

(i) $\min\{p(x), p(y)\} = p(c)$ *for some critical point c.*

(ii) $B(x) = B(y)$ *and* $p(x) = p(y)$.

Proof. For $x, y \in \mathbb{K}$, note that the merge height satisfies

$$m_p(x,y) = \min\{t : (x,y) \text{ are in the same connected component of } L(t)\}.$$

Recall that

$$d_{T_p}(x,y) = p(x) + p(y) - 2m_p(x,y).$$

Note that the modified merge distortion metric is $d_{\text{MM}}(p,q) = \sup_{x,y} |d_{T_p}(x,y) - d_{T_q}(x,y)|$.

A feature of the merge height is that

$$\begin{aligned} m_p(x,y) = p(x) &\Rightarrow B(y) \subset B(x) \\ m_p(x,y) = p(y) &\Rightarrow B(x) \subset B(y) \\ m_p(x,y) \neq p(y) \text{ or } p(x) &\Rightarrow \exists c(x,y) \in \mathcal{C} \;\; s.t. \;\; m_p(x,y) = p(c(x,y)). \end{aligned}$$

where $\mathcal{C}$ is the collection of all critical points. Thus, we have

$$d_{T_p}(x,y) = \begin{cases} p(x) - p(y) & \text{if } B(y) \subset B(x) \\ p(y) - p(x) & \text{if } B(x) \subset B(y) \\ p(x) + p(y) - 2p(c(x,y)) & \text{otherwise} \end{cases}.$$

Case 1:
We pick a pair of x_0, y_0 as in Figure C.1. Now we consider a smooth symmetric function $g(x) > 0$ such that it peaks at 0 and monotonically decay and has support $[-\delta, \delta]$ for some small $\delta > 0$. We pick δ small enough such that $p_\epsilon(x_0) = p(x_0), p_\epsilon(y_0) = p(y_0)$. For simplicity, let $g(0) = \max_x g(x) = 1$.

Now consider perturbing $p(x)$ along $g(x - c)$ with amount ϵ. Namely, we define

$$p_\epsilon(x) = p(x) + \epsilon \cdot g(x - c).$$

For notational convenience, define $\xi_{p,\epsilon} = d_{T_{p_\epsilon}}(x_0, y_0)$. When $|\epsilon|$ is sufficiently small, define

$$\xi_{p,\epsilon}(x_0, y_0) = d_{T_p}(x_0, y_0) \quad \text{if } \epsilon > 0,$$
$$\xi_{p,\epsilon}(x_0, y_0) = d_{T_p}(x_0, y_0) - 2\epsilon \quad \text{if } \epsilon < 0.$$

This is because when $\epsilon > 0$, the $p_\epsilon(c) > p(c)$, so the merge height for x_0, y_0 using p_ϵ is still the same as $p(y_0)$, which implies $\xi_{p,\epsilon}(x_0, y_0) = d_{T_p}(x_0, y_0)$. On the other hand, when $\epsilon < 0$, $p_\epsilon(c) < p(c)$, so the merge height is no longer $p(y_0)$ but $p_\epsilon(c)$. Then using the fact that $|\epsilon| = p(c) - p_\epsilon(c)$ we obtain the result.

Now we show that $d_{T_p}(x_0, y_0)$ is not Hadamard differentiable. In this case, $\phi(p) = \xi_p(x_0, y_0)$. First, we pick a sequence of ϵ_n such that $\epsilon_n \to 0$ and $\epsilon_n > 0$ if n is even and $\epsilon_n < 0$ if n is odd. Plugging $t \equiv \epsilon_n$ and $q_t = g$ into the definition of Hadamard differentiability, we have

$$\phi'(p) \equiv \frac{\xi_{p,\epsilon_n}(x_0, y_0) - d_{T_p}(x_0, y_0)}{\epsilon_n}$$

is alternating between 0 and 2, so it does not converge. This shows that the function $d_{T_p}(x, y)$ at such a pair of (x_0, y_0) is non-Hadamard differentiable.

Case 2:
The proof of this case uses the similar idea as the proof of case 1. We pick the pair (x_0, y_0) satisfying the desire conditions. We consider the same function g but now we perturb p by

$$p_\epsilon(x) = p(x) + \epsilon \cdot g(x - x_0),$$

and as long as δ is small, we will have $p_\epsilon(y_0) = p(y_0)$. Since $B(x_0) = B(y_0)$ and $p(x_0) = p(y_0)$, $d_{T_p}(x_0, y_0) = 0$. When $\epsilon > 0$, $\xi_{p,\epsilon}(x_0, y_0) = \epsilon$, and on the other hand, when $\epsilon < 0$, $\delta_\epsilon(x_0, y_0) = -\epsilon$.

In this case, again, $\phi(p) = \xi_p(x_0, y_0)$. Now we use the similar trick as case 1: picking a sequence of ϵ_n such that $\epsilon_n \to 0$ and $\epsilon_n > 0$ if n is even and $\epsilon_n < 0$ if n is odd. Under this sequence of ϵ_n, the 'derivative' along g

$$\phi'(p) \equiv \frac{\xi_{p,\epsilon_n}(x_0, y_0) - d_{T_p}(x_0, y_0)}{\epsilon_n}$$

is alternating between 1 and -1, so it does not converge. Thus, $d_{T_p}(x, y)$ at such a pair of (x_0, y_0) is non-Hadamard differentiable. □

C.7 Proofs for Section 4.3 and Appendix C.4

C.7.1 Proof of Lemma 56

Lemma 56. *Let $p_h = \mathbb{E}[\hat{p}_h]$ where $\hat{p}_h$ is the kernel estimator with bandwidth h. We assume that p is a Morse function supported on a compact set with finitely many, distinct, critical values. There exists $h_0 > 0$ such that for all $0 < h < h_0$, T_p and T_{p_h} have the same topology in Appendix C.1.*

Proof. Let S be the compact support of p. By the classical stability properties of the Morse function, there exists a constant $C_0 > 0$ such that for any other smooth function $q : S \to \mathbb{R}$ with $\|q-p\|_\infty, \|\nabla q - \nabla p\|_\infty, \|\nabla^2 q - \nabla^2 p\|_\infty < C_0$, q is a Morse function. Moreover, there exist two diffeomorphisms $h : \mathbb{R} \to \mathbb{R}$ and $\phi : S \to S$ such that $q = h \circ p \circ \phi$ See e.g., proof of [Chazal et al., 2014a, Lemma 16]. Further, h should be nondecreasing if C_0 is small enough. Hence for any $C \in T_p(\lambda)$, since $q \circ \phi^{-1}(C) = h \circ p(C)$, so $\phi^{-1}(C)$ is a connected component of $T_q(h(\lambda))$. Now define $\Phi : \{T_p\} \to \{T_q\}$ as $\Phi(C) = \phi^{-1}(C)$. Then since ϕ is a diffeomorphism, $C_1 \subset C_2$ if and only if $\Phi(C_1) = \phi^{-1}(C_1) \subset \phi^{-1}(C_2) = \Phi(C_2)$, hence $T_p \preceq T_q$ holds. And from $p \circ \phi = h^{-1} \circ q$, we can similarly show $T_q \preceq T_p$ as well. Hence from Lemma 102, two trees T_p and T_q are topologically equivalent according to the topology in Appendix C.1.

Now by the nonparametric theory (see e.g. page 144-145 of Scott [2015], and Wasserman [2006]), there is a constant $C_1 > 0$ such that $\|p_h - p\|_{2,\max} \le C_1 h^2$ when $h < 1$. Thus, when $0 \le h \le \sqrt{\frac{C_0}{C_1}}$, $T_h = T_{p_h}$ and $T = T_p$ have the same topology. □

C.7.2 Proof of Lemma 109

Lemma 109. *Suppose that the* life *function satisfies: for all* $[C] \in E(\hat{T}_h)$, $\text{life}^{top}([C]) \le \text{life}([C])$. *Then*

(i) $Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h) \preceq T_{\hat{p}_h}$.

(ii) there exists a function $\tilde{p}$ *such that* $T_{\tilde{p}} = Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h)$.

(iii) $\tilde{p}$ *in (ii) satisfies* $\tilde{p} \in \hat{C}_\alpha$.

Proof. (i)

This is implied by Lemma 104.

(ii)

Note that $Pruned_{\text{life},\hat{t}_\alpha}(\hat{T}_h)$ is generated by function $\tilde{p}$ defined as

$$\tilde{p}(x) = \sup\left\{\lambda : \text{ there exists } C \in \hat{T}_h(\lambda) \text{ such that } x \in C \text{ and } \text{life}([C]) > 2\hat{t}_\alpha\right\} + \hat{t}_\alpha.$$

(iii)

Let $C_0 := \bigcup\{C : \text{life}([C]) \le 2\hat{t}_\alpha\}$. Then note that

$$\hat{p}(x) = \sup\left\{\lambda : \text{ there exists } C \in \hat{T}_h(\lambda) \text{ such that } x \in C\right\},$$

so for all x, $\tilde{p}(x) \le \hat{p}(x) + \hat{t}_\alpha$, and if $x \notin C_0$, $\tilde{p}(x) = \hat{p}(x) + \hat{t}_\alpha$. Then note that

$$\begin{aligned}
&\left\{\lambda : \text{ there exists } C \in \hat{T}_h(\lambda) \text{ such that } x \in C\right\} \\
&\setminus \left\{\lambda : \text{ there exists } C \in \hat{T}_h(\lambda) \text{ such that } x \in C \text{ and } \text{life}([C]) > 2\hat{t}_\alpha\right\} \\
&\subset \left\{\lambda : \text{ there exists } C \in \hat{T}_h(\lambda) \text{ such that } x \in C \text{ and } \text{life}([C]) \le 2\hat{t}_\alpha\right\}
\end{aligned}$$

Let $e_x := \max\left\{e : x \in \cup e, \text{life}(e) \le 2\hat{t}_\alpha\right\}$. Then note that $x \in C$ and $\text{life}([C]) \le 2\hat{t}_\alpha$ implies that we can find some $B \in e_x$ such that $C \subset B$, so

$$\left\{\lambda : \text{ there exists } C \in \hat{T}_h(\lambda) \text{ such that } x \in C \text{ and } \text{life}([C]) \le 2\hat{t}_\alpha\right\} \subset cumlevel(e_x).$$

Hence

$$\begin{aligned}\hat{p}(x) + \hat{t}_\alpha - \tilde{p}(x) &\leq \sup\{cumlevel(e_x)\} - \inf\{cumlevel(e_x)\} \\ &= \text{life}^{top}(e_x) \\ &\leq \text{life}(e_x) \leq 2\hat{t}_\alpha,\end{aligned}$$

and hence

$$\hat{p}(x) - \hat{t}_\alpha \leq \tilde{p}(x) \leq \hat{p}(x) + \hat{t}_\alpha.$$

□

Appendix D

Appendix for Chapter 5

D.1 Stability Theorem for Persistence module

This section gives an introduction to the Stability Theorem on persistence module. We refer to Chazal et al. [2009] for more details.

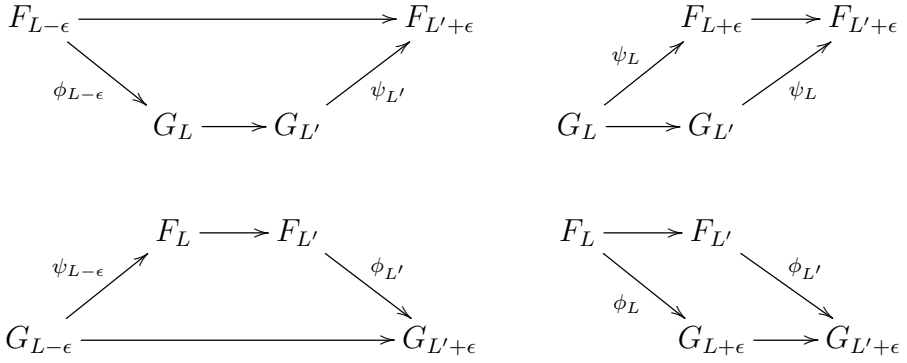
A persistence module is an algebraic abstraction of a persistent homology. Let $\mathcal{R}$ be a connected subset of $\mathbb{R}$.

Definition 111. [Chazal et al., 2009, Definition 2.1] A *persistence module* $\mathcal{F}$ is a family $\{F_L\}_{L\in\mathcal{R}}$ of $\mathbb{Z}_2$-vector spaces indexed by the elements of $\mathcal{R}$, together with a family $\{f_L^{L'} : F_L \to F_{L'}\}_{L\leq L'}$ of homomorphisms such that: $\forall L \leq L' \leq L''$, $f_L^{L''} = f_{L'}^{L''} \circ f_L^{L'}$ and $f_L^L = id_{F_L}$.

We say that $\mathcal{F}$ is tame if F_L is a finite dimensional vector spaces for all $L \in \mathcal{R}$.

For two functions $f, g : \mathbb{X} \to \mathbb{R}$ satisfying $\|f - g\|_\infty \leq \epsilon$, their sublevel sets filtrations are nested as follows: $\forall L \in \mathbb{R}$ with L, $L + \epsilon \in \mathcal{R}$, $\mathbb{X}_L^f \subset \mathbb{X}_{L+\epsilon}^g$ and $\mathbb{X}_L^g \subset \mathbb{X}_{L+\epsilon}^f$. By letting $F_L = H_k(\mathbb{X}_L^f)$ and $G_L = H_k(\mathbb{X}_L^g)$, this induces the homomorphisms induced by the inclusions as $F_L \to G_{L+\epsilon}$ and $G_L \to F_{L+\epsilon}$. Also, the canonical inclusions $\mathbb{X}_L^f \subset \mathbb{X}_{L'}^f$ and $\mathbb{X}_L^g \subset \mathbb{X}_{L'}^g$ for $L \leq L'$ induces homomorphisms as $F_L \to F_{L'}$ and $G_L \to G_{L'}$. This homomorphisms relations can be extended to persistence modules as follows:

Definition 112. Two persistence modules $\mathcal{F}$ and $\mathcal{G}$ are said to be strongly ϵ-interleaved if there exist two families of homomorphisms $\{\phi_L : F_L \to G_{L+\epsilon}\}_{L\in\mathcal{R}}$ and $\{\psi_L : G_L \to F_{L+\epsilon}\}_{L\in\mathcal{R}}$ such that the following diagrams commute for all $L \leq L'$:



(D.1)

If two persistence modules are strongly interleaved, then their bottleneck distance are close, which is the strong stability theorem.

Theorem 113 (Strong Stability Theorem)**.** *[Chazal et al., 2009, Theorem 4.4] Let $\mathcal{F}_\mathcal{R}$ and $\mathcal{G}_\mathcal{R}$ be two tame persistence modules. If $\mathcal{F}_\mathcal{R}$ and $\mathcal{G}_\mathcal{R}$ are strongly interleaved, then $d_B(\mathcal{F}_\mathcal{R}, \mathcal{G}_\mathcal{R}) \leq \epsilon$.*

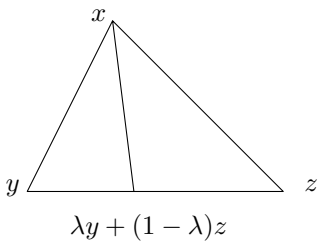


Figure D.1: The distance from one point x of a triangle to another point $\lambda y + (1-\lambda)z$ on the opposite side, as in Claim 114.

D.2 Geometry and Topology of a Set of Positive Reach

Nerve Theorem requires that any intersection of balls is contractible. This section analyzes the geometry and topology of a set of positive reach, and in particular, shows that the intersection of small enough balls is contractible. This contractibility will be used in our main theorem.

For a set A, let τ be its reach. For $u \in \mathbb{R}^m$ with $d(u, A) < \tau$, let $\pi_A(u) \in A$ be its projection on A.

We first start with simple calculation of the distance from one point of a triangle to another point lying on the opposite side, as in Claim 114.

Claim 114. Let $x, y, z \in \mathbb{R}^m$ and $\lambda \in [0, 1]$. Then

$$\|(\lambda y + (1-\lambda)z) - x\| = \sqrt{\lambda \|y - x\|^2 + (1-\lambda)\|z - x\|^2 - \lambda(1-\lambda)\|y - z\|^2}.$$

Proof of Claim 114. The distance from $\lambda y + (1-\lambda)z$ to x can be expanded as

$$\begin{aligned}
&\|(\lambda y + (1-\lambda)z) - x\|^2 \\
&= \|\lambda(y - x) + (1-\lambda)(z - x)\|^2 \\
&= \lambda^2 \|y - x\|^2 + (1-\lambda)^2 \|z - x\|^2 + 2\lambda(1-\lambda)\langle y - x, z - x\rangle .
\end{aligned}$$

Then applying $2\langle y - x, z - x\rangle = \|y - x\|^2 + \|z - x\|^2 - \|y - z\|^2$ to above gives

$$\|(\lambda y + (1-\lambda)z) - x\|^2 = \lambda \|y - x\|^2 + (1-\lambda)\|z - x\|^2 - \lambda(1-\lambda)\|y - z\|^2 ,$$

and the claim directly follows. □

Given a line segment whose end points are on A, Lemma 115 gives a bound on a distance from any point on that segment to its projection on A.

Lemma 115. *Let $A \subset \mathbb{R}^m$ be a set with reach $\tau > 0$, and let $y, z \in A$. Let $\lambda \in [0, 1]$, and let $u := \lambda y + (1-\lambda)z$ be satisfying $d(u, A) < \tau$. Then*

$$\|\pi_A(u) - u\| \leq \tau - \sqrt{\left(\tau^2 - \lambda(1-\lambda)\|y - z\|^2\right)_+}.$$

Proof of Lemma 115. 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 $\|w - \pi_A(u)\| = \tau$ holds. And $w - u = \left(\frac{\tau - \|\pi_A(u) - u\|}{\|\pi_A(u) - u\|}\right)(u - \pi_A(u))$ holds.

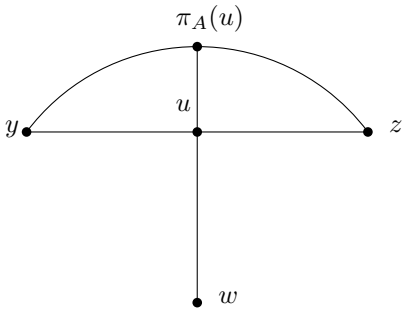


Figure D.2: Bound on the distance from any point on the segment to its projection on A, as in Lemma 115.

Since $\|u - \pi_A(u)\| < \tau$, $\langle w - u, u - \pi_A(u)\rangle = \|w - u\| \|u - \pi_A(u)\|$ and $\|u - \pi_A(u)\| + \|w - u\| = \|w - \pi_A(u)\|$ holds. Since Theorem 4.8 (2) and (6) in Federer [1959] implies that

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

for all $r < \tau$, hence $\mathbb{B}(w, \tau) \cap A = \emptyset$. Then $\|w - y\| \geq \tau$ and $\|w - z\| \geq \tau$ holds, so applying Claim 114 on $\|w - u\|$ implies

$$\begin{aligned}\|w - u\| &= \sqrt{\lambda \|w - y\|^2 + (1 - \lambda) \|w - z\|^2 - \lambda(1 - \lambda) \|y - z\|^2} \\ &\geq \sqrt{\left(\tau^2 - \lambda(1 - \lambda) \|y - z\|^2\right)_+}.\end{aligned}$$

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

$$\|u - \pi_A(u)\| \leq \tau - \sqrt{\left(\tau^2 - \lambda(1 - \lambda) \|y - z\|^2\right)_+}.$$

□

For showing the contractibility, it is sufficient to show that when two points are in a ball, then the projection of a path connecting them also lies on the ball as well. In particular, we will show that given two points in a ball, the projection of the internally dividing points to the set of positive reach is also in a ball in Claim 116 and 118. First, we consider the case when the radius of the ball is bounded by τ, where τ is the reach of the positive reach set, in Claim 116.

Claim 116. Let $A \subset \mathbb{R}^m$ be a set with reach $\tau > 0$. Let $y, z \in A$, $\lambda \in [0, 1]$, and let $u := \lambda y + (1 - \lambda)z$. Let $x \in \mathbb{R}^m$ with $\|x - y\|, \|x - z\| < \tau$. Then

$$\|x - \pi_A(u)\| \leq \sqrt{\lambda \|y - x\|^2 + (1 - \lambda) \|z - x\|^2}.$$

Proof of Claim 116. Let $r := \sqrt{\lambda \|y - x\|^2 + (1 - \lambda) \|z - x\|^2}$. Then from Claim 114,

$$\begin{aligned}\|x - u\| &= \sqrt{\lambda \|y - x\|^2 + (1 - \lambda) \|z - x\|^2 - \lambda(1 - \lambda) \|y - z\|^2} \\ &= \sqrt{r^2 - \lambda(1 - \lambda) \|y - z\|^2}.\end{aligned} \tag{D.2}$$

Also, since $\|u-y\|+\|u-z\| = \|y-z\| \leq \|x-y\|+\|x-z\| < 2\tau$ and $y, z \in A$,

$$d(u, A) \leq \min\{\|u-y\|, \|u-z\|\} < \tau,$$

and hence Lemma 115 and $\|y-z\| < 2\tau$ implies

$$\|u - \pi_A(u)\| \leq \tau - \sqrt{\tau^2 - \lambda(1-\lambda)\|y-z\|^2}. \tag{D.3}$$

Then (D.2), (D.3), and $r \leq \tau$ imply

$$\begin{aligned}
&\|x - \pi_A(u)\| \\
&\leq \|x-u\| + \|u - \pi_A(u)\| \\
&\leq \sqrt{r^2 - \lambda(1-\lambda)\|y-z\|^2} + \tau - \sqrt{\tau^2 - \lambda(1-\lambda)\|y-z\|^2} \\
&= r - \lambda(1-\lambda)\|y-z\|^2 \left(\frac{1}{r + \sqrt{r^2 - \lambda(1-\lambda)\|y-z\|^2}} - \frac{1}{\tau + \sqrt{\tau^2 - \lambda(1-\lambda)\|y-z\|^2}} \right) \\
&\leq r.
\end{aligned}$$

□

For the case when the center of the ball lies on the positive reach set, we need a slightly different version of Theorem 4.8 (8) in Federer [1959].

Lemma 117. *Let $A \subset \mathbb{R}^m$ be a set with reach $\tau > 0$, $x \in A$, and $u \in \mathbb{R}^m$ with $d(u, A) < \tau$. Then*

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

Proof of Lemma 117. From Theorem 4.8 (7) in Federer [1959],

$$\langle u - \pi_A(u),\, \pi_A(u) - x \rangle \geq -\frac{\|\pi_A(u) - x\|^2 \|u - \pi_A(u)\|}{2\tau}.$$

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

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

Rearranging this gives

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

□

Now we consider the case when center of the ball lies on the positive reach set and the radius of the ball is bounded by $\sqrt{2}\tau$, where τ is the reach, in Claim 118.

Claim 118. Let $A \subset \mathbb{R}^m$ be a set with reach $\tau > 0$. Let $y, z \in A$, $\lambda \in [0,1]$, and let $u := \lambda y + (1-\lambda)z$. Let $x \in A$ with $\|x-y\|, \|x-z\| < \sqrt{2}\tau$. Then

$$\|x - \pi_A(u)\| \leq \sqrt{\lambda \|y-x\|^2 + (1-\lambda)\|z-x\|^2}.$$

Proof of Claim 118. Let $r := \sqrt{\lambda \|y-x\|^2 + (1-\lambda)\|z-x\|^2}$, then $r < \sqrt{2}\tau$. Then from Claim 114,

$$\begin{aligned}\|x-u\| &= \sqrt{\lambda\|y-x\|^2 + (1-\lambda)\|z-x\|^2 - \lambda(1-\lambda)\|y-z\|^2} \\ &= \sqrt{r^2 - \lambda(1-\lambda)\|y-z\|^2}. \end{aligned} \tag{D.4}$$

Now, note that

$$\begin{aligned}\|u-x\|^2 + \|u-y\|\|u-z\| &< \left(r^2 - \lambda(1-\lambda)\|y-z\|^2\right) + \left((1-\lambda)\|y-z\|\right)\left(\lambda\|y-z\|\right) \\ &= r^2 < 2\tau^2,\end{aligned}$$

which implies that at least one of $\|u-x\|, \|u-y\|, \|u-z\|$ should be less than τ. And hence

$$d(u,A) \leq \min\{\|u-x\|, \|u-y\|, \|u-z\|\} < \tau.$$

Then Lemma 115 and $\|y-z\| < 2\tau$ implies

$$\|u - \pi_A(u)\| \leq \tau - \sqrt{\tau^2 - \lambda(1-\lambda)\|y-z\|^2}. \tag{D.5}$$

Now, Lemma 117 gives the upper bound of $\|x - \pi_A(u)\|$ as

$$\|x-\pi_A(u)\| \leq \sqrt{\frac{\tau\left(\|u-x\|^2 - \|u-\pi_A(u)\|^2\right)}{\tau - \|u-\pi_A(u)\|}}. \tag{D.6}$$

Consider first the case where $\lambda(1-\lambda)\|y-z\|^2 \geq \frac{1}{2}r^2$. Then applying $\|u-x\| \leq \frac{r}{\sqrt{2}}$ to (D.6) gives the bound for $\|x-\pi_A(u)\|^2$ as

$$\|x-\pi_A(u)\|^2 \leq \frac{\tau\left(\frac{r^2}{2} - \|u-\pi_A(u)\|^2\right)}{\tau - \|u-\pi_A(u)\|}.$$

Now, for further upper bounding RHS, consider a function f as

$$f(t) := \frac{\tau(\frac{r^2}{2} - t^2)}{\tau - t} \text{ for } t \in \left[0, \tau - \sqrt{\tau^2 - \lambda(1-\lambda)\|y-z\|^2}\right].$$

Then $f'(t) = \frac{\tau(t^2 - 2\tau t + \frac{r^2}{2})}{(\tau-t)^2} \leq 0$ if and only if $\tau - \sqrt{\tau^2 - \frac{r^2}{2}} \leq t \leq \tau + \sqrt{\tau^2 - \frac{r^2}{2}}$. Since $\tau - \sqrt{\tau^2 - \frac{r^2}{2}} \leq$

$\tau - \sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2} \leq \tau + \sqrt{\tau^2 - \frac{r^2}{2}}$, $f(t)$ is maximized at $t = \tau - \sqrt{\tau^2 - \frac{r^2}{2}}$, and hence

$$\begin{aligned}
\left\|x - \pi_A(u)\right\| &< \frac{\tau\left(\frac{r^2}{2} - \left\|u - \pi_A(u)\right\|^2\right)}{\tau - \left\|u - \pi_A(u)\right\|} \\
&\leq \frac{\tau\left(\frac{r^2}{2} - \left(2\tau^2 - \frac{r^2}{2} - 2\tau\sqrt{\tau^2 - \frac{r^2}{2}}\right)\right)}{\sqrt{\tau^2 - \frac{r^2}{2}}} \\
&= \frac{\tau\left(r^2 - 2\tau^2 + 2\tau\sqrt{\tau^2 - \frac{r^2}{2}}\right)}{\sqrt{\tau^2 - \frac{r^2}{2}}} \\
&= r^2 - (2\tau^2 - r^2)\left(\frac{\tau}{\sqrt{\tau^2 - \frac{r^2}{2}}} - 1\right) \\
&\leq r^2. \qquad \text{(D.7)}
\end{aligned}$$

Now, consider the case when $\lambda(1-\lambda)\left\|y-z\right\|^2 \leq \frac{1}{2}r^2$. Then applying (D.4) to (D.6) gives the bound for $\left\|x - \pi_A(u)\right\|^2$ as

$$\left\|x - \pi_A(u)\right\|^2 \leq \frac{\tau\left((r^2 - \lambda(1-\lambda)\left\|y-z\right\|^2) - \left\|u - \pi_A(u)\right\|^2\right)}{\tau - \left\|u - \pi_A(u)\right\|}.$$

Now,for further upper bounding RHS, let $\tilde{r} = \sqrt{r^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}$, and consider a function f as

$$f(t) := \frac{\tau(\tilde{r}^2 - t^2)}{\tau - t} \text{ for } t \in \left[0, \tau - \sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}\right].$$

Then $f'(t) = \frac{\tau(t^2 - 2\tau t + \tilde{r}^2)}{(\tau - t)^2} \leq 0$ if and only if $\tau - \sqrt{\tau^2 - \tilde{r}^2} \leq t \leq \tau + \sqrt{\tau^2 - \tilde{r}^2}$. Since $\tau - \sqrt{\tau^2 - \tilde{r}^2} = \tau - \sqrt{\tau^2 - (r^2 - \lambda(1-\lambda)\left\|y-z\right\|^2)} \geq \tau - \sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}$, $f(t)$ is maximized at $t =$

$\tau - \sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}$, and hence

$$
\begin{aligned}
&\left\|x - \pi_A(u)\right\| \\
&\leq \frac{\tau\left((r^2 - \lambda(1-\lambda)\left\|y-z\right\|^2) - \left\|u - \pi_A(u)\right\|^2\right)}{\tau - \left\|u - \pi_A(u)\right\|} \\
&\leq \frac{\tau\left((r^2 - \lambda(1-\lambda)\left\|y-z\right\|^2) - (2\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2 - 2\tau\sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2})\right)}{\sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}} \\
&= \frac{\tau\left(r^2 - 2\tau^2 + 2\tau\sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}\right)}{\sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}} \\
&= r^2 - (2\tau^2 - r^2)\left(\frac{\tau}{\sqrt{\tau^2 - \lambda(1-\lambda)\left\|y-z\right\|^2}} - 1\right) \\
&\leq r^2. \qquad \text{(D.8)}
\end{aligned}
$$

Hence for either cases, (D.7) and (D.8) give the desired upper bound for $\left\|x - \pi_A(u)\right\|$ as

$$\left\|x - \pi_A(u)\right\| \leq r = \sqrt{\lambda\left\|y-x\right\|^2 + (1-\lambda)\left\|z-x\right\|^2}.$$

□

Proposition 119 is the main statement of this section. Given a set A with its reach τ, it asserts that the intersection of any balls whose radius being bounded by τ is contractible.

Proposition 119. *Let $A \subset \mathbb{R}^m$ be a set with reach $\tau_A > 0$ and $\{B_\alpha\}_{\alpha\in I}$ be a collection of balls of the form $B_\alpha = \mathbb{B}_{\mathbb{R}^m}(x_\alpha, r_\alpha)$. Suppose for all $\alpha \in I$, either $x_\alpha \in A$ and $r_\alpha \leq \sqrt{2}\tau$ holds or $r_\alpha \leq \tau$ holds. Then $\bigcap_{\alpha\in I} B_\alpha \cap A$ is contractible.*

Hence, applying Proposition 119 to Nerve Theorem establish that the topology of $\mathrm{supp}(P)$ can be still well approximated by the Cech complex, as in Theorem 59.

Proof of Proposition 119. Fix $\alpha \in I$, and fix $y_1, y_2 \in B_\alpha \cap A$. Let $l : [0,1] \to B_\alpha$ with $l(t) = ty_1 + (1-t)y_2$ be the line segment from y_1 to y_2, and define a curve $\gamma_{y_1,y_2} : [0,1] \to A$ as $\gamma(t) = \pi_A(l(t))$. Thoerem 4.8 (4) in Federer [1959] implies that γ is continuous.

We will further argue that $\gamma_{y_1,y_2}(t) \in B_\alpha$ for $t \in [0,1]$. For notational convenience, let $\gamma = \gamma_{y_1,y_2}$ here. Then from Claim 116 or 118,

$$\left\|x_\alpha - \gamma(t)\right\| \leq \sqrt{\lambda\left\|x_\alpha - y_1\right\|^2 + (1-\lambda)\left\|x_\alpha - y_2\right\|^2} < r_\alpha.$$

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

Now, fix $y_0 \in \bigcap_{\alpha\in I} B_\alpha \cap A$, and define homotopic map $F : \left(\bigcap_{\alpha\in I} B_\alpha \cap A\right) \times [0,1] \to \left(\bigcap_{\alpha\in I} B_\alpha \cap A\right)$ as $F(y,t) = \gamma_{y_0,y}(t)$. Since $\gamma_{y_0,y}$ is continuous and above argument implies that $\gamma_{y_0,y}(t) \in \bigcap_{\alpha\in I} B_\alpha \cap A$, hence F is well-defined continuous map. And $F(y,0) = y$ and $F(y,y_0) = y_0$ for all $y \in \bigcap_{\alpha\in I} B_\alpha \cap A$, hence $\bigcap_{\alpha\in I} B_\alpha \cap A$ is contractible. □

Lemma 120. *Let $A \subset \mathbb{R}^m$ be a set with reach $\tau > 0$, and let $x, y \in \mathbb{R}^m$ with $\|x - \pi_A(x)\|, \|y - \pi_A(y)\| < \tau$. Then*

$$\|x - y\| \geq \|\pi_A(y) - \pi_A(x)\| \left(1 - \frac{\|x - \pi_A(x)\| + \|y - \pi_A(y)\|}{2\tau}\right).$$

Proof. From Theorem 4.8 (7) in Federer [1959],

$$\langle y - \pi_A(y),\, \pi_A(y) - \pi_A(x)\rangle \geq -\frac{\|\pi_A(y) - \pi_A(x)\|^2 \|y - \pi_A(y)\|}{2\tau},$$
$$\langle x - \pi_A(x),\, \pi_A(x) - \pi_A(y)\rangle \geq -\frac{\|\pi_A(x) - \pi_A(y)\|^2 \|x - \pi_A(x)\|}{2\tau}.$$

Then applying and gives

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

Hence,

$$\|x - y\| \geq \|\pi_A(y) - \pi_A(x)\| \left(1 - \frac{\|x - \pi_A(x)\| + \|y - \pi_A(y)\|}{2\tau}\right).$$

□

D.3 Proofs for Section 5.1

This section is for providing rigorous proofs for Section 5.1. Recall the setting in Section 5.1 that the upper level set filtration of f on $\mathbb{X}$ is defined by $\{D_L\}_{L>0}$ where

$$D_L := \{x \in \mathbb{X} : f(x) \geq L\}.$$

And the upper level set estimator $\hat{D}_L(r)$ is defined by

$$\hat{D}_L(r) := \bigcup_{\{X_i : \hat{f}(X_i) \geq L\}} \mathbb{B}_{\mathbb{X}}(X_i, r_i),$$

where

$$\mathbb{B}_{\mathbb{X}}(x, r) := \{y \in \mathbb{X} : d(x, y) < r\}, \quad r > 0.$$

From Strong stability Theorem (Theorem 113), Upper bounding the bottleneck distance by ϵ for Lemma 58, Theorem 62, and Theorem 64, is derived by showing ϵ-strongly interleaving of the corresponding persistence modules. Lemma 58, Theorem 62, and Theorem 64 are based on different interleaving relation, but they all use the interleaving between the upper level set filtration $\{D_L\}_{L>0}$ and the upper level set estimator filtration $\{\hat{D}_L(r)\}_{L>0}$, as in Lemma 121.

Lemma 121. *Suppose either f or $\hat{f}$ is M-Lipschitz continuous. For any given $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, suppose the samples form an r-covering of $\mathbb{X}$, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{D.9}$$

Then the following inclusion holds,

$$D_{L+\left\|\hat{f}-f\right\|_\infty + M\|r\|_\infty} \subset \hat{D}_L(r) \quad \textit{and} \quad \hat{D}_{L+\left\|\hat{f}-f\right\|_\infty + M\|r\|_\infty}(r) \subset D_L, \quad \forall L > 0. \tag{D.10}$$

Proof of Lemma 121. Fix $L > 0$. For the first inclusion of (D.10), suppose $x \in D_{L+\left\|\hat{f}-f\right\|_\infty + M\|r\|_\infty}$, which is equivalent to $f(x) \geq L + \left\|\hat{f} - f\right\|_\infty + M\|r\|_\infty$ and $x \in \mathbb{X}$. From (D.9), there exists some X_i such that $\|x - X_i\| \leq r_i$. If f is M-Lipschitz, $\hat{f}(X_i)$ can be lower bounded as

$$\hat{f}(X_i) \geq f(X_i) - \left\|\hat{f} - f\right\|_\infty \geq f(x) - M\|r\|_\infty - \left\|\hat{f} - f\right\|_\infty \geq L.$$

If $\hat{f}$ is M-Lipschitz, $\hat{f}(X_i)$ can be similarly lower bounded as

$$\hat{f}(X_i) \geq \hat{f}(x) - M\|r\|_\infty \geq f(x) - \left\|\hat{f} - f\right\|_\infty - M\|r\|_\infty \geq L.$$

Hence for either cases, $x \in \hat{D}_L(r)$, which implies

$$D_{L+\left\|\hat{f}-f\right\|_\infty + M\|r\|_\infty} \subset \hat{D}_L(r). \tag{D.11}$$

For the second inclusion of (D.10), suppose $x \in \hat{D}_{L+\left\|\hat{f}-f\right\|_\infty + M\|r\|_\infty}(r)$. Then $x \in \mathbb{X}$ and there exists X_i such that $\|x - X_i\| \leq r_i$ and $\hat{f}(X_i) \geq L + \left\|\hat{f} - f\right\|_\infty + M\|r\|_\infty$. If f is M-Lipschitz, $f(x)$ can be lower bounded as

$$f(x) \geq f(X_i) - M\|r\|_\infty \geq \hat{f}(X_i) - \left\|\hat{f} - f\right\|_\infty - M\|r\|_\infty \geq L.$$

If $\hat{f}$ is M-Lipschitz, $f(x)$ can be similarly lower bounded as

$$f(x) \geq \hat{f}(x) - \left\|\hat{f} - f\right\|_\infty \geq \hat{f}(X_i) - M\|r\|_\infty - \left\|\hat{f} - f\right\|_\infty \geq L.$$

Hence for either cases, $x \in D_L$, which implies

$$\hat{D}_{L+\left\|\hat{f}-f\right\|_\infty + M\|r\|_\infty}(r) \subset D_L. \tag{D.12}$$

Hence (D.11) and (D.12) imply (D.10). □

Then Lemma 58 is a direct consequence from Lemma 121 and Strong stability Theorem (Theorem 113).

Lemma 58. *Suppose either f or $\hat{f}$ is M-Lipschitz continuous. For any given $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, suppose the samples form an r-covering of $\mathbb{X}$, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{D.13}$$

Then the bottleneck distance between $\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r)$ and $\mathrm{PH}_*^{\mathbb{X}}(f)$ is upper bounded as

$$d_B\left(\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r), \mathrm{PH}_*^{\mathbb{X}}(f)\right) \leq \|\hat{f} - f\|_\infty + M\|r\|_\infty. \tag{D.14}$$

Proof of Lemma 58. From (D.13), Lemma 121 implies that $\{D_L\}_{L>0}$ and $\{\hat{D}_L(r)\}_{L>0}$ are strongly $\|\hat{f} - f\|_\infty + M\|r\|_\infty$-interleaved. Hence from Strong stability Theorem (Theorem 113), (D.14) is derived. □

In the following proofs of Claim 122 and Lemma 123, we refer to $\check{C}ech(\mathcal{X}_n, r)$ as $\check{C}(r)$ for notational convenience. Also, for $r, r' \in \mathbb{R}^n$, use the notation $r \leq r'$ as $r_i \leq r'_i$ for all i.

Claim 122. Let τ be the reach of $\mathbb{X}$. Fix $L > 0$ and $r = (r_1, \ldots, r_n) \in \left(0, \sqrt{2}\tau\right]^n$. Suppose $\mathbb{X}$ is triangulated so that $\hat{D}_L(r)$ and $\mathbb{B}(X_i, r_i)$ are subcomplices. Then there exist simplicial maps $\phi_L^r : sd\left(\hat{D}_L(r)\right) \to sd\left(\check{C}_L(r)\right)$ and $\psi_L^r : sd\left(\check{C}_L(r)\right) \to sd\left(\hat{D}_L(r)\right)$ that are homotopic equivalent to each other, i.e.

$$\psi_L^r \circ \phi_L^r \simeq id_{\hat{D}_L(r)} \text{ and } \phi_L^r \circ \psi_L^r \simeq id_{\check{C}_L(r)}. \tag{D.15}$$

Let $L, L' \in (0, \infty)$, $r, r' \in \left(0, \sqrt{2}\tau\right]^n$ and $\mathbb{X}$ is triangulated so that $\hat{D}_L(r)$, $\hat{D}_{L'}(r')$, $\mathbb{B}(X_i, r_i)$, $\mathbb{B}(X_i, r'_i)$ are subcomplices. Then ϕ_L^r and $\phi_{L'}^{r'}$ further satisfy that if $r \leq r'$ and $L' \leq L$,

$$(\phi_L^r)_* = (\phi_{L'}^{r'})_* \text{ on } H_*(sd(\hat{D}_{L'}(r')). \tag{D.16}$$

Also, ψ_L^r and $\psi_{L'}^{r'}$ further satisfy that

$$\psi_L^r = \psi_{L'}^{r'} \text{ on } sd\left(\check{C}_L(r)\right) \cap sd\left(\check{C}_{L'}(r')\right). \tag{D.17}$$

Proof of Claim 122. For showing (D.15), we consider two simplicial maps from Nerve Theorem [Björner, 1995, Theorem 10.6]. We define a simplicial map $\phi_L^r : sd\left(\hat{D}_L(r)\right) \to sd\left(\check{C}_L(r)\right)$ to be a barycentric map induced from $\sigma \mapsto \left\{X_i \in \mathcal{X}_{n,L}^{\hat{f}} : \sigma \in \mathbb{B}_{\mathbb{X}}(X_i, r_i)\right\}$ (where each $\mathbb{B}_{\mathbb{X}}(X_i, r_i)$ is understood as a simplicial subcomplex of $\mathbb{X}$). We also define a simplicial map $\psi_L^r : sd\left(\check{C}_L(r)\right) \to sd\left(\hat{D}_L(r)\right)$ to be a barycentric map induced from $\{X_{n_1}, \ldots, X_{n_k}\} \mapsto \frac{\sum_{j=1}^k r_j X_{n_j}}{\sum_{j=1}^k r_j}$. From $r_i \leq \sqrt{2}\tau$ for all i and Proposition 119, the proof of Björner [1995, Theorem 10.6] implies that ψ_L^r and ϕ_L^r gives the homotopy equivalence between $\hat{D}_L(r)$ and $\check{C}(r)$, i.e.

$$\psi_L^r \circ \phi_L^r \simeq id_{\hat{D}_L(r)} \text{ and } \phi_L^r \circ \psi_L^r \simeq id_{\check{C}_L(r)}.$$

For showing (D.16), suppose $r \leq r'$ and $L' \leq L$. For each $\sigma \in sd\left(\hat{D}_L(r)\right)$, since vertices of σ can be ordered by inclusion relation, we can define its minimal vertex $\min\sigma := \min\{v : v \in \sigma\}$. And let $\Delta_\sigma := \left\{X_i \in \mathcal{X}_{n,L'}^{L'} : \min\sigma \in \mathbb{B}_{\mathbb{X}}(X_i, r_i)\right\}$ be the set of vertices that is r'_i-close from $\min\sigma$. Then $\Delta_\sigma \subset \mathcal{X}_{n,L'}^{L'}$ and $\min\sigma \in \bigcap_{X_i \in \Delta_\sigma} \mathbb{B}_{\mathbb{X}}(X_i, r'_i) \neq \emptyset$ implies that Δ_σ is a subcomplex of $\check{C}_{L'}(r')$, i.e.

$$\Delta_\sigma \subset \check{C}_{L'}(r').$$

Also, $\|\phi_L^r(\sigma)\|, \left\|\phi_{L'}^{r'}(\sigma)\right\| \subset \|\Delta_\sigma\|$ holds from the definition of ϕ_L^r and Δ_σ. Hence for any $\gamma \in B_*\left(sd\left(\hat{D}_L(r)\right)\right)$, $\phi_L^r(\gamma)$ and $\phi_{L'}^{r'}(\gamma)$ are homotopic to each other in $sd\left(\check{C}_{L'}(r')\right)$, i.e., $\phi_L^r(\gamma) - \phi_{L'}^{r'}(\gamma) \in Z_*\left(sd\left(\check{C}_{L'}(r')\right)\right)$and hence in $H_*\left(sd\left(\check{C}_{L'}(r')\right)\right)$,

$$(\phi_L^r)_*[\gamma] = (\phi_{L'}^{r'})_*[\gamma].$$

Therefore (D.16) holds.

For showing (D.17), note that from the definition of ψ_L^r, if $\sigma \in sd\left(\check{C}_L(r)\right) \cap sd\left(\check{C}_{L'}(r')\right)$, then ψ_L^r and $\psi_{L'}^{r'}$ coincide on $sd\left(\check{C}_L(r)\right) \cap sd\left(\check{C}_{L'}(r')\right)$, i.e.

$$\psi_L^r(\sigma) = \psi_{L'}^{r'}(\sigma).$$

Hence (D.17) holds.

□

Lemma 123. *Let τ be the reach of $\mathbb{X}$ and $r', r'' \in \left(0, \sqrt{2}\tau\right]^n$ with $r' \leq r''$. Let $\epsilon > 0$ be satisfying*

$$D_{L+\epsilon} \subset \hat{D}_L(r'), \text{ and } \hat{D}_{L+\epsilon}(r'') \subset D_L, \text{ for all } L > 0.$$

Let $r \in \left(0, \sqrt{2}\tau\right]^n$ and let $\mathcal{S} = \{S_L(r)\}_{L\in(0,\infty)}$ be a filtration of simplicial complexes satisfying

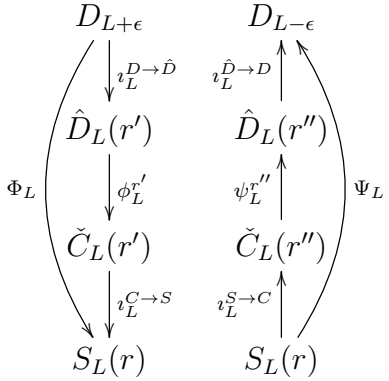
$$\check{C}ech_{\mathbb{X}}(\mathcal{X}_{n,L}^{\hat{f}}, r') \subset S_L(r) \subset \check{C}ech_{\mathbb{X}}(\mathcal{X}_{n,L}^{\hat{f}}, r'') \text{ for all } L > 0.$$

Then $\{H_*(D_L)\}_{L>0}$ and $\{H_*(S_L(r))\}_{L>0}$ are strongly ϵ-interleaved. In particular,

$$d_B\left(\mathrm{PH}_*(\mathcal{S}), \mathrm{PH}_*^{\mathbb{X}}(f)\right) \leq \epsilon. \tag{D.18}$$

Proof of Lemma 123. Our goal is to define simplicial maps $\Phi_L : D_{L+\epsilon} \to sd(S_L(r))$ and $\Psi_L : sd(S_L(r)) \to D_{L-\epsilon}$ so that $(\Phi_L)_* : H_*(D_{L+\epsilon}) \to H_*(S_L(r))$ and $(\Psi_L)_* : H_*(S_L(r)) \to H_*(D_{L-\epsilon})$ are homomorphisms satisfying strong ϵ-interleaving conditions in (D.1). Then Strong Stability Theorem (Theorem 113) implies (D.18).

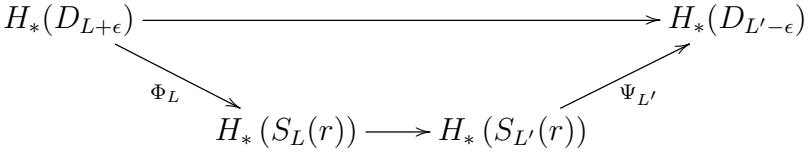
Now we construct Φ_L and Ψ_L. Let $\imath_L^{D\to\hat{D}} : D_{L+\epsilon} \to sd\left(\hat{D}_L(r')\right)$, $\imath_L^{C\to S} : sd\left(\check{C}_L(r')\right) \to sd(S_L(r))$, $\imath_L^{S\to C} : sd(S_L(r)) \to sd\left(\check{C}_L(r'')\right)$, $\imath_L^{\hat{D}\to D} : sd\left(\hat{D}_L(r'')\right) \to D_{L-\epsilon}$ be simplicial maps induced from the inclusion maps. And then we define $\Phi_L := \imath_L^{C\to S} \circ \phi_L^{r'} \circ \imath_L^{D\to\hat{D}} : D_{L+\epsilon} \to sd(S_L(r))$ and $\Psi_L := \imath_L^{\hat{D}\to D} \circ \psi_L^{r''} \circ \imath_L^{S\to C} : sd(S_L(r)) \to D_{L-\epsilon}$, as in (D.19).



(D.19)

For $L' \in (0,\infty)$ with $L' \leq L$, let $\imath_{L\to L'}^D : D_L \to D_{L'}$, $\imath_{L\to L'}^S : sd(S_L(r)) \to sd(S_L(r))$ be simplicial maps induced from the inclusion maps.

First we show that the diagram in (D.20) commutes,



(D.20)

i.e. compare $\Psi_{L'} \circ \imath^S_{L\to L'} \circ \Phi_L : D_{L+\epsilon} \to D_{L'-\epsilon}$ to inclusion map $\imath^D_{L+\epsilon\to L'-\epsilon} : D_{L+\epsilon} \to D_{L'-\epsilon}$. For $\gamma \in B_*\left(D_{L+\epsilon}\right)$, note that $\Phi_L = \imath^{C\to S}_L \circ \phi^{r'}_L \circ \imath^{D\to\hat{D}}_L$ and $\Psi_{L'} = \imath^{\hat{D}\to D}_{L'} \circ \psi^{r''}_{L'} \circ \imath^{S\to C}_{L'}$, hence $\Psi_{L'} \circ \imath^S_{L\to L'} \circ \Phi_L(\gamma)$ can be expanded as

$$\begin{aligned}\Psi_{L'} \circ \imath^S_{L\to L'} \circ \Phi_L(\gamma) &= (\imath^{\hat{D}\to D}_{L'} \circ \psi^{r''}_{L'} \circ \imath^{S\to C}_{L'}) \circ \imath^S_{L\to L'} \circ (\imath^{C\to S}_L \circ \phi^{r'}_L \circ \imath^{D\to\hat{D}}_L)(\gamma) \\ &= \psi^{r''}_{L'} \circ \phi^{r'}_L(\gamma). \end{aligned} \tag{D.21}$$

Now, note that from $L \geq L'$ and $r' \leq r''$, $\check{C}_L(r') \subset \check{C}_{L'}(r'')$ holds, and hence

$$\phi^r_L(\gamma) \in B_*\left(sd\left(\check{C}_L(r')\right)\right) = B_*\left(sd\left(\check{C}_L(r')\right) \cap sd\left(\check{C}_{L'}(r'')\right)\right).$$

Then (D.17) in Claim 122 implies that $\psi^{r''}_{L'} = \psi^{r'}_L$ on $sd\left(\check{C}_L(r')\right) \cap sd\left(\check{C}_{L'}(r'')\right)$, hence combined with above gives

$$\psi^{r''}_{L'} \circ \phi^{r'}_L(\gamma) = \psi^{r'}_L \circ \phi^{r'}_L(\gamma). \tag{D.22}$$

Then (D.15) in Claim 122 implies that $\psi^{r'}_L$ and $\phi^{r'}_L$ are homotopic inverses to each other in $H_*\left(\hat{D}_L(r')\right)$, i.e.

$$\left(\psi^{r'}_L \circ \phi^{r'}_L\right)_*[\gamma] = id_{\hat{D}_L(r')}[\gamma] = [\gamma] \text{ in } H_*\left(\hat{D}_L(r')\right). \tag{D.23}$$

Since $\hat{D}_L(r') \subset D_{L'-\epsilon}$, combining (D.21), (D.22), and (D.23) gives that in $H_*(D_{L'-\epsilon})$,

$$\begin{aligned}\left(\Psi_{L'} \circ \imath^S_{L\to L'} \circ \Phi_L\right)_*[\gamma] &= \left(\psi^{r''}_L \circ \phi^{r'}_L\right)_*[\gamma] \\ &= \left(\psi^{r'}_L \circ \phi^{r'}_L\right)_*[\gamma] \\ &= [\gamma] \\ &= \left(\imath^D_{L+\epsilon\to L'-\epsilon}\right)_*[\gamma],\end{aligned}$$

i.e. $\Psi_{L'} \circ \imath^S_{L\to L'} \circ \Phi_{L+\epsilon}$ and $\imath^D_{L+\epsilon\to L'-\epsilon}$ coincide on $H_*(D_{L'-\epsilon})$, and hence (D.20) is shown.

Second, we show that the diagram in (D.24) commutes,

$$\begin{array}{ccc} & & H_*\left(D_{L-\epsilon}\right) \longrightarrow H_*\left(D_{L'-\epsilon}\right) \\ & \overset{\Psi_L}{\nearrow} & \overset{\Psi_{L'}}{\nearrow} \\ H_*\left(S_L(r)\right) & \longrightarrow & H_*\left(S_{L'}(r)\right) \end{array} \tag{D.24}$$

i.e. compare $\Psi_{L'} \circ \imath^S_{L\to L'} : sd(S_L(r)) \to D_{L'-\epsilon}$ to $\imath^D_{L-\epsilon\to L'-\epsilon} \circ \Psi_L : sd(S_L(r)) \to D_{L'-\epsilon}$. For $\gamma \in B_*\left(sd(S_L(r))\right)$, note that $\Psi_{L'} = \imath^{\hat{D}\to D}_{L'} \circ \psi^{r''}_{L'} \circ \imath^{S\to C}_{L'}$ and $\Psi_L = \imath^{\hat{D}\to D}_L \circ \psi^{r''}_L \circ \imath^{S\to C}_L$, hence

$$\Psi_{L'} \circ \imath^S_{L\to L'}(\gamma) = (\imath^{\hat{D}\to D}_{L'} \circ \psi^{r''}_{L'} \circ \imath^{S\to C}_{L'}) \circ \imath^S_{L\to L'}(\gamma) = \psi^{r''}_{L'}(\gamma), \tag{D.25}$$

$$\imath^D_{L-\epsilon\to L'-\epsilon} \circ \Psi_L(\gamma) = \imath^D_{L-\epsilon\to L'-\epsilon} \circ (\imath^{\hat{D}\to D}_L \circ \psi^{r''}_L \circ \imath^{S\to C}_L)(\gamma) = \psi^{r''}_L(\gamma). \tag{D.26}$$

From $L \geq L'$, $\check{C}_L(r'') \subset \check{C}_{L'}(r'')$ holds, and hence

$$\gamma \in B_*\left(sd\left(\check{C}_L(r'')\right)\right) = B_*\left(sd\left(\check{C}_L(r'')\right) \cap sd\left(\check{C}_{L'}(r'')\right)\right).$$

Also, (D.17) in Claim 122 implies that $\psi^{r''}_{L'} = \psi^{r''}_L$ on $sd\left(\check{C}_L(r'')\right) \cap sd\left(\check{C}_{L'}(r'')\right)$, hence (D.25) and (D.26) indeed eqaul, i.e.

$$\Psi_{L'} \circ \imath^S_{L\to L'}(\gamma) = \psi^{r''}_{L'}(\gamma) = \psi^{r''}_L(\gamma) = \imath^D_{L-\epsilon\to L'-\epsilon} \circ \Psi_L(\gamma).$$

Hence they equal in $H_*(D_{L'-\epsilon})$ as well, i.e.

$$\left(\Psi_{L'} \circ \imath^S_{L\to L'}\right)_* [\gamma] = \left(\imath^D_{L-\epsilon\to L'-\epsilon} \circ \Psi_L\right)_* [\gamma] \text{ in } H_*(D_{L'-\epsilon}),$$

and hence (D.24) is shown.

Third, we show that the diagram in (D.27) commutes,

$$\begin{array}{ccc} & H_*\left(D_L\right) \longrightarrow H_*\left(D_{L'}\right) & \\ {\scriptstyle \Psi_{L+\epsilon}} \nearrow & & \searrow {\scriptstyle \Phi_{L'-\epsilon}} \\ H_*\left(S_{L+\epsilon}(r)\right) & \longrightarrow & H_*\left(S_{L'-\epsilon}(r)\right) \end{array} \tag{D.27}$$

i.e. compare $\Phi_{L'-\epsilon} \circ \imath^D_{L\to L'} \circ \Psi_{L+\epsilon} : sd\left(S_{L+\epsilon}(r)\right) \to sd\left(S_{L'-\epsilon}(r)\right)$ to inclusion map $\imath^S_{L+\epsilon\to L'-\epsilon} : sd\left(S_{L+\epsilon}(r)\right) \to sd\left(S_{L'-\epsilon}(r)\right)$. For $\gamma \in B_*\left(sd\left(S_{L+\epsilon}(r)\right)\right)$, note that $\Phi_{L'-\epsilon} = \imath^{C\to S}_{L'-\epsilon} \circ \phi^{r'}_{L'-\epsilon} \circ \imath^{D\to \hat{D}}_{L'-\epsilon}$ and $\Psi_{L+\epsilon} = \imath^{\hat{D}\to D}_{L+\epsilon} \circ \psi^{r''}_{L+\epsilon} \circ \imath^{S\to C}_{L+\epsilon}$, hence $\Phi_{L'-\epsilon} \circ \imath^D_{L\to L'} \circ \Psi_{L+\epsilon}(\gamma)$ can be expanded as

$$\begin{aligned} \Phi_{L'-\epsilon} \circ \imath^D_{L\to L'} \circ \Psi_{L+\epsilon}(\gamma) &= (\imath^{C\to S}_{L'-\epsilon} \circ \phi^{r'}_{L'-\epsilon} \circ \imath^{D\to\hat{D}}_{L'-\epsilon}) \circ \imath^D_{L\to L'} \circ (\imath^{\hat{D}\to D}_{L+\epsilon} \circ \psi^{r''}_{L+\epsilon} \circ \imath^{S\to C}_{L+\epsilon})(\gamma) \\ &= \phi^{r'}_{L'-\epsilon} \circ \psi^{r''}_{L+\epsilon}(\gamma). \end{aligned} \tag{D.28}$$

Now, note that $\left\|\check{C}_{L+\epsilon}(r'')\right\| = \hat{D}_{L+\epsilon}(r'') \subset D_L \subset D_{L'} \subset \hat{D}_{L'-\epsilon}(r') = \left\|\check{C}_{L'-\epsilon}(r')\right\|$, hence with subdivisions if necessasry,

$$\gamma \in B_*\left(sd\left(\check{C}_{L+\epsilon}(r'')\right)\right) = B_*\left(sd\left(\check{C}_{L+\epsilon}(r'')\right) \cap sd\left(\check{C}_{L'-\epsilon}(r')\right)\right).$$

Then (D.17) in Claim 122 implies that $\psi^{r''}_{L+\epsilon} = \psi^{r'}_{L'-\epsilon}$ on $sd\left(\check{C}_{L+\epsilon}(r'')\right) \cap sd\left(\check{C}_{L'-\epsilon}(r')\right)$, hence combined with above gives

$$\phi^{r'}_{L'-\epsilon} \circ \psi^{r''}_{L+\epsilon}(\gamma) = \phi^{r'}_{L'-\epsilon} \circ \psi^{r'}_{L'-\epsilon}(\gamma). \tag{D.29}$$

Then (D.15) in Claim 122 implies that $\psi^{r'}_{L'-\epsilon}$ and $\phi^{r'}_{L'-\epsilon}$ are homotopic inverses to each other in $H_*\left(sd\left(\check{C}_{L'-\epsilon}(r')\right)\right)$, i.e.

$$\left(\phi^{r'}_{L'-\epsilon} \circ \psi^{r'}_{L'-\epsilon}\right)_* [\gamma] = id_{sd\left(\check{C}_{L'-\epsilon}(r')\right)}[\gamma] = [\gamma] \text{ in } H_*\left(sd\left(\check{C}_{L'-\epsilon}(r')\right)\right). \tag{D.30}$$

Since $sd\left(\check{C}_{L'-\epsilon}(r')\right) \subset sd\left(S_{L'-\epsilon}(r)\right)$, combining (D.28), (D.29), and (D.30) gives that in $H_*\left(sd\left(S_{L'-\epsilon}(r)\right)\right) \cong H_*\left(S_{L'-\epsilon}(r)\right)$,

$$\begin{aligned} \left(\Phi_{L'-\epsilon} \circ \imath^D_{L\to L'} \circ \Psi_{L+\epsilon}\right)_* [\gamma] &= \left(\phi^{r'}_{L'-\epsilon} \circ \psi^{r''}_{L+\epsilon}\right)_* [\gamma] \\ &= \left(\phi^{r'}_{L'-\epsilon} \circ \psi^{r'}_{L'-\epsilon}\right)_* [\gamma] \\ &= [\gamma] \\ &= \left(\imath^S_{L+\epsilon\to L'-\epsilon}\right)_* [\gamma], \end{aligned}$$

i.e. $\Phi_{L'-\epsilon} \circ \imath^D_{L\to L'} \circ \Psi_{L+\epsilon}$ and $\imath^S_{L+\epsilon\to L'-\epsilon}$ coincide on $H_*\left(S_{L'+\epsilon}(r)\right)$, and hence (D.27) is shown.

Fourth, we show that the diagram in (D.31) commutes,

$$\begin{array}{ccc} H_*\left(D_L\right) & \longrightarrow & H_*\left(D_{L'}\right) \\ & {\scriptstyle \Phi_{L-\epsilon}} \searrow & \searrow {\scriptstyle \Phi_{L'-\epsilon}} \\ & & H_*\left(S_{L-\epsilon}(r)\right) \longrightarrow H_*\left(S_{L'-\epsilon}(r)\right) \end{array} \tag{D.31}$$

i.e. compare $\Phi_{L'-\epsilon} \circ \imath^D_{L\to L'} : D_L \to sd(S_{L'-\epsilon}(r))$ to $\imath^S_{L-\epsilon\to L'-\epsilon} \circ \Phi_{L-\epsilon} : D_L \to sd(S_{L'-\epsilon}(r))$. For $\gamma \in B_*(D_L)$, note that $\Phi_{L-\epsilon} = \imath^{C\to S}_{L-\epsilon} \circ \phi^{r'}_{L-\epsilon} \circ \imath^{D\to\hat{D}}_{L-\epsilon}$ and $\Phi_{L'-\epsilon} = \imath^{C\to S}_{L'-\epsilon} \circ \phi^{r'}_{L'-\epsilon} \circ \imath^{D\to\hat{D}}_{L'-\epsilon}$, hence

$$\Phi_{L'-\epsilon} \circ \imath^D_{L\to L'}(\gamma) = (\imath^{C\to S}_{L'-\epsilon} \circ \phi^{r'}_{L'-\epsilon} \circ \imath^{D\to\hat{D}}_{L'-\epsilon}) \circ \imath^D_{L\to L'}(\gamma) = \phi^{r'}_{L'-\epsilon}(\gamma), \tag{D.32}$$

$$\imath^S_{L-\epsilon\to L'-\epsilon} \circ \Phi_{L-\epsilon}(\gamma) = \imath^S_{L-\epsilon\to L'-\epsilon} \circ (\imath^{C\to S}_{L-\epsilon} \circ \phi^{r'}_{L-\epsilon} \circ \imath^{D\to\hat{D}}_{L-\epsilon})(\gamma) = \phi^{r'}_{L-\epsilon}(\gamma). \tag{D.33}$$

Then (D.16) in Claim 122 implies that $\phi^{r'}_{L'-\epsilon} = \phi^{r'}_{L-\epsilon}$ on $H_*\left(sd\left(\check{C}_{L'-\epsilon}(r)\right)\right)$, hence (D.32) and (D.33) are equal in $H_*\left(sd\left(\check{C}_{L'-\epsilon}(r)\right)\right)$, i.e.

$$\left(\Phi_{L'-\epsilon} \circ \imath^D_{L\to L'}\right)_*[\gamma] = \left(\phi^{r'}_{L'-\epsilon}\right)_*[\gamma] = \left(\phi^{r'}_{L-\epsilon}\right)_*[\gamma] = \left(\imath^S_{L-\epsilon\to L'-\epsilon} \circ \Phi_L\right)_*[\gamma] \text{ in } H_*\left(sd\left(\check{C}_{L'-\epsilon}(r)\right)\right).$$

Since $\check{C}_{L'-\epsilon}(r) \subset S_{L'-\epsilon}(r)$, the same relation holds in $H_*(sd(S_{L'-\epsilon}(r))) \cong H_*(S_{L'-\epsilon}(r))$ as well, and hence (D.24) is shown.

From (D.20), (D.24), (D.27), and (D.31), $\{H_*(D_L)\}_{L>0}$ and $\{H_*(S_L(r))\}_{L>0}$ are strongly ϵ-interleaved. Hence from Strong stability Theorem (Theorem 113), (D.18) is derived.

□

Theorem 62. *Let τ be the reach of $\mathbb{X}$. Suppose either f or $\hat{f}$ is M-Lipschitz continuous. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \tau/\sqrt{2}]^n$, suppose the samples form an r-covering of $\mathbb{X}$, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{D.34}$$

Then the bottleneck distance between $\mathrm{PH}^{\check{C}}_*(\hat{f}, r)$ and $\mathrm{PH}^{\mathbb{X}}_*(f)$ is upper bounded as

$$d_B\left(\mathrm{PH}^{\check{C}}_*(\hat{f}, r), \mathrm{PH}^{\mathbb{X}}_*(f)\right) \leq \|\hat{f} - f\|_\infty + 2M\|r\|_\infty \tag{D.35}$$

Proof of Theorem 62. From (D.34), Lemma 121 implies that for all $L > 0$,

$$D_L \subset \hat{D}_{L-\|\hat{f}-f\|_\infty - M\|r\|_\infty}(r) \subset \hat{D}_{L-\|\hat{f}-f\|_\infty - 2M\|r\|_\infty}(r),$$
$$\hat{D}_L(2r) \subset D_{L-\|\hat{f}-f\|_\infty - 2M\|r\|_\infty}.$$

And Čech complexes on $\mathbb{X}$ and Čech complexes on $\mathbb{R}^m$ have the following inclusion relation as

$$\check{C}ech_{\mathbb{X}}(\mathcal{X}^{\hat{f}}_{n,L}, r) \subset \check{C}ech_{\mathbb{R}^m}(\mathcal{X}^{\hat{f}}_{n,L}, r) \subset \check{C}ech_{\mathbb{X}}(\mathcal{X}^{\hat{f}}_{n,L}, 2r).$$

Hence from Lemma 123, $\{H_*(D_L)\}_{L\in\mathbb{R}}$ and $\left\{H_*\left(\check{C}ech_{\mathbb{R}^m}(\mathcal{X}^{\hat{f}}_{n,L}, r)\right)\right\}_{L\in\mathbb{R}}$ are strongly $\|\hat{f} - f\|_\infty + 2M\|r\|_\infty$-interleaved, and in particular, (D.35) is derived.

□

Theorem 64. *Let τ be the reach of $\mathbb{X}$. Suppose either f or $\hat{f}$ is M-Lipschitz continuous. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \tau/\sqrt{2}]^n$, suppose the samples form an r-covering of $\mathbb{X}$, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i). \tag{D.36}$$

Then the bottleneck distance between $\mathrm{PH}^R_*(\hat{f}, r)$ and $\mathrm{PH}^{\mathbb{X}}_*(f)$ is upper bounded as

$$d_B\left(\mathrm{PH}^R_*(\hat{f}, r), \mathrm{PH}^{\mathbb{X}}_*(f)\right) \leq \|\hat{f} - f\|_\infty + 2M\|r\|_\infty. \tag{D.37}$$

Proof of Theorem 64. From (D.36), Lemma 121 implies that for all $L \in \mathbb{R}$,

$$D_L \subset \hat{D}_{L-\|\hat{f}-f\|_\infty - M\|r\|_\infty}(r) \subset \hat{D}_{L-\|\hat{f}-f\|_\infty - 2M\|r\|_\infty}(r),$$
$$\hat{D}_L(2r) \subset D_{L-\|\hat{f}-f\|_\infty - 2M\|r\|_\infty}.$$

And Čech complexes on $\mathbb{X}$ and Rips complexes have the following inclusion relation as

$$\check{C}ech_{\mathbb{X}}(\mathcal{X}_{n,L}^{\hat{f}}, r) \subset R(\mathcal{X}_{n,L}^{\hat{f}}, r) \subset \check{C}ech_{\mathbb{X}}(\mathcal{X}_{n,L}^{\hat{f}}, 2r).$$

Hence from Lemma 123, $\{H_*(D_L)\}_{L\in\mathbb{R}}$ and $\left\{H_*\left(R(\mathcal{X}_{n,L}^{\hat{f}}, r)\right)\right\}_{L\in\mathbb{R}}$ are strongly $\|\hat{f}-f\|_\infty + 2M\|r\|_\infty$-interleaved, and in particular, (D.37) is derived.

□

D.4 Proofs for Section 5.2

Claim 124. Let P be a probability measure on $\mathbb{R}^m$ and K be a kernel function satisfying Assumption 66, 67, and 72. Let $C_K := \int_{\mathbb{R}^m} |x|K(x)dx$. Then,

$$\|p_h - p\|_\infty \leq C_K M_P h.$$

Proof of Claim 124. Note that $p_h(x)$ can be expanded as

$$p_h(x) = h^{-d}\int_{\mathbb{R}^m} K\left(\frac{x-z}{h}\right) dP(z).$$

Then under Assumption 72, $dP(z) = p(z)dz$, and hence the integral is further expanded as

$$p_h(x) = h^{-d}\int_{\mathbb{R}^m} K\left(\frac{x-z}{h}\right) dP(z) = \int_{\mathbb{R}^m} K(t)p(x-ht)dt.$$

Hence $p_h(x) - p(x)$ can be bounded as

$$\begin{aligned}
|p_h(x) - p(x)| &= \left|\int_{\mathbb{R}^m} K(t)p(x-ht)dt - p(x)\right| \\
&= \left|\int_{\mathbb{R}^m} K(t)(p(x-ht) - p(x))dt\right| \\
&\leq \int_{\mathbb{R}^m} K(t)\left|p(x-ht) - p(x)\right| dt \\
&\leq hM_P \int_{\mathbb{R}^m} |t|K(t)dt \\
&= C_K M_P h.
\end{aligned}$$

□

Proposition 68. *Let P be a probability measure on $\mathbb{R}^m$ and K be a kernel function satisfying Assumption 66 and 67. Let p be the Lebesgue density of P, and assume p is Lipschitz continuous. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, the following hold :*

(a) $d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*^{\mathbb{R}^d}(p_h)\right) \leq \sup_{x \notin supp(P)} |p_h(x)| \leq C_K M_P h$,

(b) $d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*(p)\right) \leq \sup_{x \in supp(P)} |p_h(x) - p(x)| \leq C_K M_P h$,

where $C_K = \int \|x\| K(x) \mathrm{d}x$ *and* $M_P > 0$ *is the Lipschitz constant of* p.

Proof of Proposition 68. We will first show that

$$\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) = \mathrm{PH}_*^{\mathbb{R}^m}(p_h I_{\mathrm{supp}(P)}), \tag{D.38}$$

where $I_{\mathrm{supp}(P)}(x) = I(x \in \mathrm{supp}(P))$ is an indicator function on a set $\mathrm{supp}(P)$. Note that the level set of (5.12) equals

$$D_L = \{x \in \mathrm{supp}(P) : p_h(x) \geq L\} = \{x \in \mathbb{R}^m : p_h(x) I_{\mathrm{supp}(P)}(x) \geq L\},$$

and hence D_L is a level set of $p_h I_{\mathrm{supp}(P)}$ at L. Hence, $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) = \mathrm{PH}_*^{\mathbb{R}^m}(p_h I_{\mathrm{supp}(P)})$ holds.

(a)

Applying (D.38) to Theorem 13 gives

$$\begin{aligned} d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*^{\mathbb{R}^m}(p_h)\right) &= d_B\left(\mathrm{PH}_*^{\mathbb{R}^m}(p_h I_{\mathrm{supp}(P)}), \mathrm{PH}_*^{\mathbb{R}^m}(p_h)\right) \\ &\leq \left\|p_h I_{\mathrm{supp}(P)} - p_h\right\|_\infty = \sup_{x \notin \mathrm{supp}(P)} |p_h(x)|. \end{aligned}$$

Note that $p(x) = 0$ on $\mathbb{R}^m \backslash \mathrm{supp}(P)$, and hence

$$\sup_{x \notin \mathrm{supp}(P)} |p_h(x)| = \sup_{x \notin \mathrm{supp}(P)} |p_h(x) - p(x)| \leq \|p_h - p\|_\infty.$$

Hence applying Claim 124 gives

$$d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*^{\mathbb{R}^m}(p_h)\right) \leq \sup_{x \notin \mathrm{supp}(P)} |p_h(x)| \leq C_K M_P h.$$

(b)

Similarly, applying (D.38) to Theorem 13 gives

$$d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*(p)\right) = d_B\left(\mathrm{PH}_*^{\mathbb{R}^m}(p_h I_{\mathrm{supp}(P)}), \mathrm{PH}_*^{\mathbb{R}^m}(p)\right) \leq \|p_h - p\|_\infty.$$

Hence applying Claim 124 gives

$$d_B\left(\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h), \mathrm{PH}_*(p)\right) \leq \|p_h - p\|_\infty \leq C_K M_P h.$$

□

Lemma 125. *Suppose the distribution* P *and the kernel function* K *satisfies Assumption 66 and 67. Then the following inequalities hold for any* $x, y \in \mathbb{R}^m$*:*

(a) If K *is* M_K*-Lipschitz, then*

$$|\hat{p}_h(x) - \hat{p}_h(y)| \leq \frac{M_K}{h^{d+1}} \|x - y\|.$$

(b) Under Assumption 71,

$$|p_h(x) - p_h(y)| \leq \frac{a_{\max} M_K}{h^{d+1-\nu_{\min}}} \|x - y\|.$$

(c) Under Assumption 72,

$$|p_h(x) - p_h(y)| \leq M_P \|x - y\|.$$

Proof of Lemma 125. (a)

The first inequality comes from the M_K-Lipschitz continuity of K.

$$\begin{aligned}
|\hat{p}_h(x) - \hat{p}_h(y)| &\leq \frac{1}{nh^d} \sum_{i=1}^{n} \left| K\left(\frac{x - X_i}{h}\right) - K\left(\frac{y - X_i}{h}\right) \right| \\
&\leq \frac{1}{nh^d} M_K \left\| \frac{x - y}{h} \right\| \\
&= \frac{M_K}{nh^{d+1}} \|x - y\|.
\end{aligned}$$

(b)

If we further suppose Assumption 71 holds, note that $p_h(x) - p_h(y)$ can be factorized as

$$\begin{aligned}
p_h(x) - p_h(y) &= \mathbb{E}_P \left[\frac{1}{h^d} \left(K\left(\frac{x - X}{h}\right) - K\left(\frac{y - X}{h}\right) \right) \right] \\
&= h^{-d} \int_{\mathbb{R}^m} \left(K\left(\frac{x - z}{h}\right) - K\left(\frac{y - z}{h}\right) \right) dP(z) \\
&= h^{-d} \int_{\mathbb{B}(x,h) \cup \mathbb{B}(y,h)} \left(K\left(\frac{x - z}{h}\right) - K\left(\frac{y - z}{h}\right) \right) dP(z) \\
&\quad + h^{-d} \int_{\mathbb{R}^m \setminus (\mathbb{B}(x,h) \cup \mathbb{B}(y,h))} \left(K\left(\frac{x - z}{h}\right) - K\left(\frac{y - z}{h}\right) \right) dP(z).
\end{aligned}$$

Then note that for $z \in \mathbb{R}^m \setminus (\mathbb{B}(x,h) \cup \mathbb{B}(y,h))$, $\left\| \frac{x-z}{h} \right\|, \left\| \frac{y-z}{h} \right\| \geq 1$ and hence $K\left(\frac{x-z}{h}\right) = K\left(\frac{y-z}{h}\right) = 0$ under Assumption 71. Hence the integral reduces to and is further bounded as

$$\begin{aligned}
|p_h(x) - p_h(y)| &= h^{-d} \left| \int_{\mathbb{B}(x,h) \cup \mathbb{B}(y,h)} \left(K\left(\frac{x - z}{h}\right) - K\left(\frac{y - z}{h}\right) \right) dP(z) \right| \\
&\leq h^{-d} \int_{\mathbb{B}(x,h) \cup \mathbb{B}(y,h)} \left| K\left(\frac{x - z}{h}\right) - K\left(\frac{y - z}{h}\right) \right| dP(z) \\
&\leq h^{-d} \int_{\mathbb{B}(x,h) \cup \mathbb{B}(y,h)} M_K \left\| \frac{x - y}{h} \right\| dP(z) \\
&= \frac{M_K}{h^{d+1}} \|x - y\| P(\mathbb{B}(x,h) \cup \mathbb{B}(y,h)) \\
&\leq \frac{M_K}{h^{d+1}} \|x - y\| (P(\mathbb{B}(x,h)) + P(\mathbb{B}(y,h))) \\
&\leq \frac{a_{\max} M_K}{h^{d+1-\nu_{\max}}} \|x - y\|.
\end{aligned}$$

(c)

Now, we suppose Assumption 72. Note that $p_h(x)$ can be expanded as

$$\begin{aligned} p_h(x) &= \mathbb{E}_P\left[\frac{1}{h^d}K\left(\frac{x-X}{h}\right)\right] \\ &= h^{-d}\int_{\mathbb{R}^m} K\left(\frac{x-z}{h}\right) dP(z). \end{aligned}$$

Then under Assumption 72, $dP(z) = p(z)dz$, and hence the integral is further expanded as

$$\begin{aligned} p_h(x) &= h^{-d}\int_{\mathbb{R}^m} K\left(\frac{x-z}{h}\right) p(z)dz \\ &= \int_{\mathbb{R}^m} K(t)p(x-ht)dt. \end{aligned}$$

And hence $p_h(x) - p_h(y)$ can be bounded as

$$\begin{aligned} |p_h(x) - p_h(y)| &= \left|\int_{\mathbb{R}^m} K(t)(p(x-ht) - p(y-ht))dt\right| \\ &\leq \int_{\mathbb{R}^m} K(t)\left|p(x-ht) - p(y-ht))\right| dt \\ &= \int_{\mathbb{R}^m} K(t)M_P\left\|x-y\right\| dt \\ &= M_P\left\|x-y\right\|. \end{aligned}$$

□

Proposition 73. *Let P be a probability measure on $\mathbb{R}^m$ and K be a kernel function satisfying Assumption 66 and 67. For any given $h > 0$, $r = (r_1, \ldots, r_n) \in (0,\infty)^n$ with $\sqrt{2}\|r\|_\infty \leq \tau$, suppose the samples form an r-covering of the support of P, that is,*

$$\mathbb{X} \subset \bigcup_i \mathbb{B}_{\mathbb{X}}(X_i, r_i).$$

Then the bottleneck distance between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ and its estimator $\mathrm{PH}_*^R(\hat{p}_h, r)$ is upper bounded as, under Assumption 71,

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq \|\hat{p}_h - p_h\|_\infty + \frac{2a_{\max}M_K\|r\|_\infty}{h^{d+1-\nu_{\min}}}, \tag{D.39}$$

while, under Assumption 72,

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq \|\hat{p}_h - p_h\|_\infty + 2M_P\|r\|_\infty. \tag{D.40}$$

Proof of Proposition 73. Under Assumption 71, Lemma 125 (b) imply that p_h is $\frac{a_{\max}M_K}{h^{d+1-\nu_{\min}}}$-Lipschitz. Hence Theorem 64 implies (D.39).

Similarly under Assumption 72, Lemma 125 (c) imply that p_h is M_P-Lipschitz. Hence Theorem 64 implies (D.40). □

Lemma 126. *Suppose Assumption 66 holds. Let $\{r_n = (r_{n,1}, \dots, r_{n,n})\}_{n\in\mathbb{N}}$ be a triangular array of positive numbers. Then the probability of the samples forming an r_n-covering of* $\mathrm{supp}(P)$ *is bounded as*

$$\mathbb{P}\left(\mathrm{supp}(P) \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i})\right) \geq 1 - a_{\min}^{-1} \exp\left(\nu_{\max} \log(\min_i r_{n,i})^{-1} - 2^{-\nu_{\max}} a_{\min} n (\min_i r_{n,i})^{\nu_{\max}}\right). \tag{D.41}$$

In particular, if $\min_i r_{n,i} \geq 2\left(\frac{\beta \log n}{a_{\min} n}\right)^{1/\nu_{\max}}$*, then*

$$\mathbb{P}\left(\mathrm{supp}(P) \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i})\right) \geq 1 - \frac{1}{2^{\nu_{\max}} n^{\beta-1} \log n}. \tag{D.42}$$

Proof of Lemma 126. Let $\epsilon := \frac{1}{2}\min_i r_{n,i}$. Under Assumption 66, there exists $x_1, \dots, x_N$ with $N \leq a_{\min}^{-1}\epsilon^{-\nu_{\max}}$ satisfying

$$\mathrm{supp}(P) \subset \bigcup_{j=1}^{N} \mathbb{B}_{\mathbb{R}^m}(x_j, \epsilon).$$

Let E' be the event that all $\mathbb{B}_{\mathbb{R}^m}(x_j, \epsilon)$ have intersections with $\{X_1, \dots, X_n\}$, i.e. for each $1 \leq j \leq N$, there exists $1 \leq i \leq n$ with $X_i \in \mathbb{B}_{\mathbb{R}^m}(x_j, \epsilon)$. Then note that $2\epsilon = \min_i r_{n,i} \leq r_{n,i}$, and hence $\mathbb{B}_{\mathbb{R}^m}(x_j, \epsilon) \subset \mathbb{B}_{\mathbb{R}^m}(X_i, 2\epsilon) \subset \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i})$. Hence under E',

$$\mathrm{supp}(P) \subset \bigcup_{j=1}^{N} \mathbb{B}_{\mathbb{R}^m}(x_j, \epsilon) \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i}),$$

and hence E' implies $\mathrm{supp}(P) \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i})$, i.e.

$$\mathbb{P}\left(\mathrm{supp}(P) \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i})\right) \geq \mathbb{P}(E'). \tag{D.43}$$

Then $P(E')$ can be expanded and lower bounded as

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

where the last line is from that $1-t \leq \exp(-t)$ for all $t \in \mathbb{R}$. Then from Assumption 66, $P(\mathbb{B}_{\mathbb{R}^m}(x_j, \epsilon)) \geq a_{\min}\epsilon^{\nu_{\max}}$ holds, and hence applying this and $N \leq a_{\min}^{-1}\epsilon^{-\nu_{\max}}$ gives

$$\begin{aligned} P(E') &\geq 1 - N\exp\left(-a_{\min}n\epsilon^{\nu_{\max}}\right) \\ &\geq 1 - a_{\min}^{-1}\exp\left(\nu_{\max}\log\epsilon^{-1} - a_{\min}n\epsilon^{\nu_{\max}}\right) \\ &\geq 1 - a_{\min}^{-1}\exp\left(\nu_{\max}\log(\min_i r_{n,i})^{-1} - 2^{-\nu_{\max}}a_{\min}n(\min_i r_{n,i})^{\nu_{\max}}\right). \end{aligned} \tag{D.44}$$

Hence applying (D.44) to (D.43) gives (D.41).

Now, suppose $\min_i r_{n,i} \geq 2\left(\frac{\beta\log n}{a_{\min}n}\right)^{1/\nu_{\max}}$. Note that RHS of (D.41) is an increasing function of $\min_i r_{n,i}$, and hence

$$\begin{aligned} \mathbb{P}\left(\mathrm{supp}(P) \subset \bigcup_{i=1}^n \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i})\right) &\geq 1 - a_{\min}^{-1}\exp\left(\log\left(\frac{a_{\min}n}{2^{\nu_{\max}}\log n}\right) - \beta\log n\right) \\ &= 1 - \frac{1}{2^{\nu_{\max}}n^{\beta-1}\log n}. \end{aligned}$$

Hence (D.42) is shown. □

Theorem 74. *Suppose Assumption 66 and 67 holds. Let $\{r_n = (r_{n,1}, \ldots, r_{n,n})\}_{n\in\mathbb{N}}$ be a triangular array of positive numbers such that*

$$\min_i r_{n,i} \geq C_P\left(\frac{\log n}{n}\right)^{1/\nu_{\max}}, \tag{D.45}$$

with a constant C_P depending only on $a_{\min}$. Let also assume $\sqrt{2}\|r_n\|_\infty \leq \tau$ for all sufficiently large n. Then, under Assumption 71, for a fixed $h > 0$, there exists a positive constant $C_{K,P}$ depends only on $\|K\|_\infty$, $\|K\|_2$, $\nu_{\min}$, $\nu_{\max}$, $a_{\min}$, $a_{\max}$ such that with probability at least $1-\delta$, the bottleneck distance between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)$ and its estimator $\mathrm{PH}_*^R(\hat{p}_h, r_n)$ is upper bounded as

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq C_{K,P}\left(\sqrt{\frac{\log(1/\delta)}{n}} + \|r_n\|_\infty\right), \tag{D.46}$$

for $\forall n$ with $\sqrt{2}\|r_n\|_\infty \leq \tau$.
Under Assumption 72, suppose $h_n \leq h_0$ for some fixed $h_0 \in (0,1)$ for sufficiently large n and $h_n^{-d}\log(1/h_n) \leq C_{h_0}n$ for some constant C_{h_0}. Then there exists a positive constant C_{K,P,h_0} depends only on $\|K\|_\infty$, $\|K\|_2$, d , $a_{\min}$, $\|p\|_\infty$, h_0 such that with probability at least $1-\delta$, the bottleneck distance between the persistent homology of the density filtration $\mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})$ and its estimator $\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n)$ is upper bounded as

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})\right) \leq C_{K,P,h_0}\left(\sqrt{\frac{\log(1/\delta)}{nh_n^d}} + \sqrt{\frac{\log(1/h_n)}{nh_n^d}} + \|r_n\|_\infty\right). \tag{D.47}$$

for $\forall n$ with $\sqrt{2}\|r_n\|_\infty \leq \tau$.

Proof of Theorem 74. Note first that, under Assumption 66 and (D.45), Lemma 126 implies that when $n^{\beta-1}\log n \geq \frac{1}{2^{\nu_{\max}-1}\delta}$,

$$\mathbb{P}\left(\mathrm{supp}(P) \subset \bigcup_{i=1}^{n} \mathbb{B}_{\mathbb{R}^m}(X_i, r_{n,i})\right) \geq 1 - \frac{1}{2^{\nu_{\max}} n^{\beta-1}\log n} \geq 1 - \frac{\delta}{2}, \tag{D.48}$$

i.e. the sample forms an r_n-covering of the support of P.

First, suppose the assumptions 66, 67, and 71. When the sample forms an r_n-covering of $supp(P)$, we have the following inequality from (5.14) in Proposition 73 as

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})\right) \leq \|\hat{p}_{h_n} - p_{h_n}\|_\infty + \frac{2a_{\max}M_K \|r_n\|_\infty}{h_n^{d+1-\nu_{\min}}}.$$

Then under the Assumption 71, with probability $1-\frac{\delta}{2}$, we have

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})\right) \leq C_{P,K,h_0}\sqrt{\frac{\log(1/h_n)+\log(2/\delta)}{nh_n^{2d-\nu_{\min}}}} + \frac{2a_{\max}M_K \|r_n\|_\infty}{h_n^{d+1-\nu_{\min}}}.$$

Hence when $h_n = h$ for all n, with probability $1-\delta$, we have

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_h, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h)\right) \leq C_{P,K,h,M_K}\left(\sqrt{\frac{\log(1/\delta)}{nh^{2d-\nu_{\min}}}} + \frac{\|r_n\|_\infty}{h^{d+1-\nu_{\min}}}\right),$$

where C_{P,K,h,M_K} depends only on $\|K\|_\infty$, $\|K\|_2$, $\nu_{\min}$, $\nu_{\max}$, $a_{\min}$, $a_{\max}$, h, M_K.

Second, suppose the assumptions 66, 67, and 72. When the sample forms an r_n-covering of $supp(P)$, we have the following inequality from (5.15) in Proposition 73 as

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})\right) \leq \|\hat{p}_{h_n} - p_{h_n}\|_\infty + 2M_P \|r_n\|_\infty.$$

Then under the Assumption 72, with probability $1-\frac{\delta}{2}$, we have

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})\right) \leq C_{P,K,h_0}\sqrt{\frac{\log(1/h_n)+\log(2/\delta)}{nh_n^{2d-\nu_{\min}}}} + 2M_P \|r_n\|_\infty.$$

And hence with probability $1-\delta$, we have

$$d_B\left(\mathrm{PH}_*^R(\hat{p}_{h_n}, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_{h_n})\right) \leq C_{P,K,h_0,M_P}\left(\sqrt{\frac{\log(1/h_n)}{nh_n^{2d-\nu_{\min}}}} + \sqrt{\frac{\log(1/\delta)}{nh_n^{2d-\nu_{\min}}}} + \|r_n\|_\infty\right),$$

where C_{P,K,h_0,M_P} depends only on $\|K\|_\infty$, $\|K\|_2$, $\nu_{\min}$, $\nu_{\max}$, $a_{\min}$, $a_{\max}$, h_0, M_P. □

We generalize the setting of Lemma 77. For any given $\hat{f}: \mathbb{R}^m \to \mathbb{R}$ and $r = (r_1, \ldots, r_n) \in (0,\infty)^n$, let $\mathcal{E}_r(\hat{f}) \subset \mathbb{R}$ be a version of (5.20) for $\hat{f}$, i.e.

$$\mathcal{E}_r(\hat{f}) := \left\{\epsilon \in \mathbb{R}_+ : \left\{x : \hat{f}(x) \geq \epsilon\right\} \subset \bigcup_i \mathbb{B}_{\mathbb{R}^d}(X_i, r_i)\right\}, \tag{D.49}$$

and let $\hat{c}_r(\hat{f})$ a version of (5.22) for $\hat{f}$, i.e.

$$\hat{c}_r(\hat{f}) := \inf\{\epsilon \in \mathcal{E}_r(\hat{f})\} \vee \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^d}(X_i, r_i)} |\hat{f}(X_i) - \hat{f}(x)|. \tag{D.50}$$

Claim 127. For any $\hat{f} : \mathbb{R}^m \to \mathbb{R}$ and $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, the following holds:

(a) $\left(\sup \hat{f} \vee 0, \infty\right) \subset \mathcal{E}_r(\hat{f})$.

(b) $\hat{f}^{-1}\left(\inf \mathcal{E}_r(\hat{f}), \infty\right) \subset \bigcup_i \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)$.

(c) $\hat{c}_r(\hat{f}) \in \left[\inf \mathcal{E}_r(\hat{f}),\ \sup \hat{f} - \inf \hat{f} \wedge 0\right]$.

(d) For $x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)$, $\left|\hat{f}(x) - \hat{f}(X_i)\right| \leq \hat{c}_r(\hat{f})$.

Proof of Claim 127. (a)

Note that for any $\epsilon > \sup \hat{f} \vee 0$, $\epsilon \in \mathbb{R}_+$ and $\left\{x : \hat{f}(x) \geq \epsilon\right\} = \emptyset \subset \bigcup_i \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)$, and hence

$$\left(\sup \hat{f} \vee 0, \infty\right) \subset \mathcal{E}_r(\hat{f}).$$

(b)

From the definition of $\mathcal{E}_r(\hat{f})$ in (D.49), $\hat{f}(x) > \inf \mathcal{E}_r(\hat{f})$ implies that $\hat{f}(x) \in \mathcal{E}_r(\hat{f})$, and hence

$$x \in \left\{y : \hat{f}(y) \geq \hat{f}(x)\right\} \subset \bigcup_i \mathbb{B}_{\mathbb{R}^m}(X_i, r_i).$$

(c)

$\hat{c}_r(\hat{f}) \geq \inf \mathcal{E}_r(\hat{f})$ is apparent from the definition in (D.50) as

$$\hat{c}_r(\hat{f}) = \inf \mathcal{E}_r(\hat{f}) \vee \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)} |\hat{f}(X_i) - \hat{f}(x)| \geq \inf \mathcal{E}_r(\hat{f}).$$

For $\hat{c}_r(\hat{f}) \leq \sup \hat{f} - \inf \hat{f}$, note that

$$\max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)} |\hat{f}(X_i) - \hat{f}(x)| \leq \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)} \sup \hat{f} - \inf \hat{f} \leq \sup \hat{f} - \inf \hat{f} \wedge 0. \tag{D.51}$$

Also from (a),

$$\inf \mathcal{E}_r(\hat{f}) \leq \sup \hat{f} \vee 0 \leq \sup \hat{f} - \inf \hat{f} \wedge 0. \tag{D.52}$$

Hence from (D.51) and (D.52), $\hat{c}_r(\hat{f})$ is upper bounded as

$$\begin{aligned}\hat{c}_r(\hat{f}) &= \inf \mathcal{E}_r(\hat{f}) \vee \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)} |\hat{f}(X_i) - \hat{f}(x)| \\ &\leq \sup \hat{f} - \inf \hat{f} \wedge 0.\end{aligned}$$

(d)

Let $x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)$. Then $\left|\hat{f}(x) - \hat{f}(X_i)\right|$ can be bounded as

$$\begin{aligned}\left|\hat{f}(x) - \hat{f}(X_i)\right| &\leq \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)} |\hat{f}(X_i) - \hat{f}(x)| \\ &\leq \inf \mathcal{E}_r(\hat{f}) \vee \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^m}(X_i, r_i)} |\hat{f}(X_i) - \hat{f}(x)| = \hat{c}_r(\hat{f}).\end{aligned}$$

□

Lemma 128. *For any bounded function $\hat{f} : \mathbb{R}^m \to \mathbb{R}$ and $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, the following inclusion holds:*

$$D_{L+\left\|\hat{f}-f\right\|_\infty + \hat{c}_r(\hat{f})} \subset \hat{D}_L(r) \quad \textit{and} \quad \hat{D}_{L+\left\|\hat{f}-f\right\|_\infty + \hat{c}_r(\hat{f})}(r) \subset D_L, \quad \forall L > 0, \tag{D.53}$$

where

$$\hat{D}_L(r) = \bigcup_{\{X_i : \hat{f}(X_i) \geq L\}} \mathbb{B}_{\mathbb{X}}(X_i, r_i),$$

and

$$D_L = \{x \in \mathbb{X} : \ f(x) \geq L\}.$$

Proof of Lemma 128. Fix $L > 0$. Note first that from Claim 127 (c) and $\hat{f}$ bounded,

$$\hat{c}_r(\hat{f}) \leq \sup \hat{f} - \inf \hat{f} \wedge 0 < \infty.$$

To prove the first inclusion of (D.53), suppose $x \in D_{L+\left\|\hat{f}-f\right\|_\infty + \hat{c}_r(\hat{f})}$, which is equivalent to $x \in \mathbb{X}$ and $f(x) \geq L + \left\|\hat{f} - f\right\|_\infty + \hat{c}_r(\hat{f})$. Then from $\hat{c}_r(\hat{f}) < \infty$,

$$\begin{aligned} \hat{f}(x) &\geq f(x) - \left\|\hat{f} - f\right\|_\infty \geq L + \hat{c}_r(\hat{f}) \\ &> \hat{c}_r(\hat{f}). \end{aligned} \tag{D.54}$$

Then from Claim 127 (c), $\hat{f}(x) > \inf \mathcal{E}_r(\hat{f})$, and hence from Claim 127 (b),

$$x \in \bigcup_i \mathbb{B}_{\mathbb{R}^m}(X_i, r_i),$$

i.e. there exists some X_i such that $\|x - X_i\| \leq r_i$. Then from Claim 127 (d) and (D.54),

$$\hat{f}(X_i) \geq \hat{f}(x) - \hat{c}_r(\hat{f}) \geq L,$$

Hence $x \in \hat{D}_L$, which implies that

$$D_{L+\left\|\hat{f}-f\right\|_\infty + \hat{c}_r(\hat{f})} \subset \hat{D}_L. \tag{D.55}$$

For the second inclusion of (D.53), suppose $x \in \hat{D}_{L+\left\|\hat{f}-f\right\|_\infty + \hat{c}_r(\hat{f})}(r)$. Then $x \in \mathbb{X}$ and there exists X_i such that $\|x - X_i\| \leq r_i$ and $\hat{f}(X_i) \geq L + \left\|\hat{f} - f\right\|_\infty + \hat{c}_r(\hat{f})$. Then from Claim 127 (d),

$$\hat{f}(x) \geq \hat{f}(X_i) - \hat{c}_r(\hat{f}) \geq L + \left\|\hat{f} - f\right\|_\infty.$$

Therefore,

$$f(x) \geq \hat{f}(x) - \left\|\hat{f} - f\right\|_\infty \geq L.$$

Hence $x \in D_L$, which implies that

$$\hat{D}_{L+\left\|\hat{f}-f\right\|_\infty + \hat{c}_r(\hat{f})}(r) \subset D_L. \tag{D.56}$$

Hence (D.55) and (D.56) imply (D.53). □

Lemma 129. *For any given $\hat{f} : \mathbb{R}^m \to \mathbb{R}$ bounded above and $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, set*

$$\mathcal{E}_r(\hat{f}) = \left\{ \epsilon \in \mathbb{R}_+ : \left\{ x : \hat{f}(x) \geq \epsilon \right\} \subset \bigcup_i \mathbb{B}_{\mathbb{R}^d}(X_i, r_i) \right\}.$$

Then,

$$d_B \left(\mathrm{PH}_*^{\mathbb{X}}(\hat{f}, r), \mathrm{PH}_*^{\mathbb{X}}(f) \right) \leq \|\hat{f} - f\|_\infty + \hat{c}_r, \tag{D.57}$$

where

$$\hat{c}_r(\hat{f}) := \inf\{\epsilon \in \mathcal{E}_r(\hat{f})\} \vee \max_i \sup_{x \in \mathbb{B}_{\mathbb{R}^d}(X_i, r_i)} |\hat{f}(X_i) - \hat{f}(x)|.$$

Proof of Lemma 129. Lemma 128 implies that $\{D_L\}_{L \in (0,\infty)}$ and $\{\hat{D}_L(r)\}_{L \in (0,\infty)}$ are strongly $\left\|\hat{f} - f\right\|_\infty + \hat{c}_r$-interleaved. Hence from Strong stability Theorem (Theorem 113), (D.57) is derived. □

Lemma 130. *For any given $\hat{f} : \mathbb{R}^m \to \mathbb{R}$ bounded above and $r = (r_1, \ldots, r_n) \in (0, \infty)^n$, the following relation holds:*

$$d_B \left(\mathrm{PH}_*^{R}(\hat{f}, r), \mathrm{PH}_*^{\mathbb{X}}(f) \right) \leq \|\hat{f} - f\|_\infty + \hat{c}_r \vee \hat{c}_{2r}. \tag{D.58}$$

Proof of Lemma 130. Lemma 128 implies that for all $L \in (0, \infty)$,

$$D_{L + \left\|\hat{f} - f\right\|_\infty + \hat{c}_r \vee \hat{c}_{2r}} \subset D_{L + \left\|\hat{f} - f\right\|_\infty + \hat{c}_r} \subset \hat{D}_L(r),$$
$$\hat{D}_{L + \left\|\hat{f} - f\right\|_\infty + \hat{c}_r \vee \hat{c}_{2r}}(2r) \subset \hat{D}_{L + \left\|\hat{f} - f\right\|_\infty + \hat{c}_{2r}}(2r) \subset D_L.$$

And Čech complexes on $\mathbb{X}$ and Rips complexes have the following inclusion relation as

$$\check{C}ech_{\mathbb{X}}(\mathcal{X}_{n,L}^{\hat{f}}, r) \subset R(\mathcal{X}_{n,L}^{\hat{f}}, r) \subset \check{C}ech_{\mathbb{X}}(\mathcal{X}_{n,L}^{\hat{f}}, 2r).$$

Hence from Lemma 123, $\{H_*(D_L)\}_{L \in (0,\infty)}$ and $\left\{ H_* \left(R(\mathcal{X}_{n,L}^{\hat{f}}, r) \right) \right\}_{L \in (0,\infty)}$ are strongly $\left\|\hat{f} - f\right\|_\infty + \hat{c}_r \vee \hat{c}_{2r}$-interleaved, and in particular, (D.58) is derived. □

Theorem 78. *Suppose Assumption 66 and 67 holds. Let $\{r_n = (r_{n,1}, \ldots, r_{n,n})\}_{n \in \mathbb{N}}$ be a triangular array of positive numbers such that $\sqrt{2}\|r_n\|_\infty \leq \tau$ for all sufficiently large n. Then, the confidence set $\hat{C}_\alpha^R$ in* (5.25) *is asymptotically valid and satisfies*

$$\mathbb{P}\left(d_B \left(\mathrm{PH}_*^{R}(\hat{p}_h, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) \right) \leq \frac{\hat{z}_\alpha}{\sqrt{nh^d}} + \hat{c}_{r_n} \vee \hat{c}_{2r_n} \right) \geq 1 - \alpha + O\left(\frac{1}{\sqrt{n}} \right).$$

Proof of Theorem 78. Applying Lemma 130 gives the lower bound for LHS of (D.59) as

$$\begin{aligned}
&\mathbb{P}\left(d_B \left(\mathrm{PH}_*^{R}(\hat{p}_h, r_n), \mathrm{PH}_*^{\mathrm{supp}(P)}(p_h) \right) \leq \frac{\hat{z}_\alpha}{\sqrt{nh^d}} + \hat{c}_{r_n} \vee \hat{c}_{2r_n} \right) \\
&\geq \mathbb{P}\left(\|\hat{p}_h - p_h\|_\infty + \hat{c}_{r_n} \vee \hat{c}_{2r_n} \leq \frac{\hat{z}_\alpha}{\sqrt{nh^d}} + \hat{c}_{r_n} \vee \hat{c}_{2r_n} \right) \\
&= \mathbb{P}\left(\sqrt{nh^d} \|\hat{p}_h - p_h\|_\infty \leq \hat{z}_\alpha \right).
\end{aligned} \tag{D.59}$$

Then from the $1-\alpha$ asymptotic confidence set for $\|\hat{p}_h - p_h\|_\infty$ with fixed $h > 0$ in (5.24), we have

$$\mathbb{P}\left(\sqrt{nh^d}\,\|\hat{p}_h - p_h\|_\infty \le \hat{z}_\alpha\right) = 1 - \alpha + O\left(\sqrt{\frac{1}{n}}\right). \tag{D.60}$$

Then combining (D.59) and (D.60) gives (D.59). □