Estimating the Reach of a Manifold

Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, Larry Wasserman

Estimating the Reach of a Manifold

Eddie Aamari[1,*,†], **Jisu Kim**[2,‡,§], **Frédéric Chazal**[2,*,†], **Bertrand Michel**[4,*,†], **Alessandro Rinaldo**[3,§], **and Larry Wasserman**[3]

[1] *CNRS, LPSM*
Université Paris Diderot
e-mail: aamari@lpsm.paris *url:* lpsm.paris/ aamari/

[2] *Inria Saclay – Île-de-France*
e-mail: jisuk1@andrew.cmu.edu; frederic.chazal@inria.fr
url: stat.cmu.edu/ jisuk/; geometrica.saclay.inria.fr/team/Fred.Chazal/

[3] *Department of Statistics*
Carnegie Mellon University
e-mail: arinaldo@cmu.edu; larry@stat.cmu.edu
url: stat.cmu.edu/ arinaldo/; stat.cmu.edu/ larry/

[4] *Département Informatique et Mathématiques*
École Centrale de Nantes
e-mail: bertrand.michel@ec-nantes.fr *url:* bertrand.michel.perso.math.cnrs.fr/

Abstract: Various problems in manifold estimation make use of a quantity called the *reach*, denoted by τ_M, which is a measure of the regularity of the manifold. This paper is the first investigation into the problem of how to estimate the reach. First, we study the geometry of the reach through an approximation perspective. We derive new geometric results on the reach for submanifolds without boundary. An estimator $\hat{\tau}$ of τ_M is proposed in an oracle framework where tangent spaces are known, and bounds assessing its efficiency are derived. In the case of i.i.d. random point cloud $\mathbb{X}_n$, $\hat{\tau}(\mathbb{X}_n)$ is showed to achieve uniform expected loss bounds over a $\mathcal{C}^3$-like model. Finally, we obtain upper and lower bounds on the minimax rate for estimating the reach.

MSC 2010 subject classifications: Primary 62G05; secondary 62C20, 68U05.
Keywords and phrases: Geometric Inference, Reach, Minimax Risk.

1. Introduction

1.1. Background and Related Work

Manifold estimation has become an increasingly important problem in statistics and machine learning. There is now a large literature on methods and theory for estimating manifolds. See, for example, [31, 25, 24, 10, 33, 8, 26].

*Research supported by ANR project TopData ANR-13-BS01-0008
†Research supported by Advanced Grant of the European Research Council GUDHI
‡Supported by Samsung Scholarship
§Partially supported by NSF CAREER Grant DMS 1149677

Estimating a manifold, or functionals of a manifold, requires regularity conditions. In nonparametric function estimation, regularity conditions often take the form of smoothness constraints. In manifold estimation problems, a common assumption is that the reach τ_M of the manifold M is non-zero.

First introduced by Federer [22], the reach τ_M of a set $M \subset \mathbb{R}^D$ is the largest number such that any point at distance less than τ_M from M has a unique nearest point on M. If a set has its reach greater than $\tau_{min} > 0$, then one can roll freely a ball of radius τ_{min} around it [15]. The reach is affected by two factors: the curvature of the manifold and the width of the narrowest bottleneck-like structure of M, which quantifies how close M is from being self-intersecting.

Positive reach is the minimal regularity assumption on sets in geometric measure theory and integral geometry [23, 37]. Sets with positive reach exhibit a structure that is close to being differential — the so-called tangent and normal cones. The value of the reach itself quantifies the degree of regularity of a set, with larger values associated to more regular sets. The positive reach assumption is routinely imposed in the statistical analysis of geometric structures in order to ensure good statistical properties [15] and to derive theoretical guarantees. For example, in manifold reconstruction, the reach helps formalize minimax rates [25, 31]. The optimal manifold estimators of [1] implicitly use reach as a scale parameter in their construction. In homology inference [33, 7], the reach drives the minimal sample size required to consistently estimate topological invariants. It is used in [16] as a regularity parameter in the estimation of the Minkowski boundary lengths and surface areas. The reach has also been explicitly used as a regularity parameter in geometric inference, such as in volume estimation [5] and manifold clustering [4]. Finally, the reach often plays the role of a scale parameter in dimension reduction techniques such as vector diffusions maps [36]. Problems in computational geometry such as manifold reconstruction also rely on assumptions on the reach [10].

In this paper we study the problem of estimating reach. To do so, we first provide new geometric results on the reach. We also give the first bounds on the minimax rate for estimating reach. As a first attempt to study reach estimation in the literature, we will mainly work in a framework where a point cloud is observed jointly with tangent spaces, before relaxing this constraint in Section 6. Such an oracle framework has direct applications in digital imaging [32, 28], where a very high resolution image or 3D-scan, represented as a manifold, enables to determine precisely tangent spaces for arbitrary finite set of points [28].

There are very few papers on this problem. When the embedding dimension is 3, the estimation of the local feature size (a localized version of the reach) was tackled in a deterministic way in [19]. To some extent, the estimation of the medial axis (the set of points that have strictly more than one nearest point on M) and its generalizations [17, 6] can be viewed as an indirect way to estimate the reach. A test procedure designed to validate whether data actually comes from a smooth manifold satisfying a condition on the reach was developed in [24]. The authors derived a consistent test procedure, but the results do not permit any inference bound on the reach. When a sample is uniformly distributed over

a full-dimensional set, [35] proposes a selection procedure for the radius of r-convexity of the set, a quantity closely related to the reach.

1.2. Outline

In Section 2 we provide some differential geometric background and define the statistical problem at hand. New geometric properties of the reach are derived in Section 3, and their consequences for its inference follow in Section 4 in a setting where tangent spaces are known. We then derive minimax bounds in Section 5. An extension to a model where tangent spaces are unknown is discussed in Section 6, and we conclude with some open questions in Section 7. For sake of readability, the proofs are given in the Appendix.

2. Framework

2.1. Notions of Differential Geometry

In what follows, $D \geq 2$ and $\mathbb{R}^D$ is endowed with the Euclidean inner product $\langle \cdot, \cdot \rangle$ and the associated norm $\|\cdot\|$. The associated closed ball of radius r and center x is denoted by $\mathcal{B}(x, r)$. We will consider compact connected submanifolds M of $\mathbb{R}^D$ of fixed dimension $1 \leq d < D$ and without boundary [20]. For every point p in M, the tangent space of M at p is denoted by T_pM: it is the d-dimensional linear subspace of $\mathbb{R}^D$ composed of the directions that M spans in the neighborhood of p. Besides the Euclidean structure given by $\mathbb{R}^D \supset M$, a submanifold is endowed with an intrinsic distance induced by the ambient Euclidean one, and called the geodesic distance. Given a smooth path $c : [a, b] \to M$, the length of c is defined as $Length(c) = \int_a^b \|c'(t)\|\, dt$. One can show [20] that there exists a path $\gamma_{p\to q}$ of minimal length joining p and q. Such an arc is called geodesic, and the geodesic distance between p and q is given by $d_M(p, q) = Length(\gamma_{p\to q})$. We let $\mathcal{B}_M(p, s)$ denote the closed geodesic ball of center $p \in M$ and of radius s. A geodesic γ such that $\|\gamma'(t)\| = 1$ for all t is called arc-length parametrized. Unless stated otherwise, we always assume that geodesics are parametrized by arc-length. For all $p \in M$ and all unit vectors $v \in T_pM$, we denote by $\gamma_{p,v}$ the unique arc-length parametrized geodesic of M such that $\gamma_{p,v}(0) = p$ and $\gamma'_{p,v}(0) = v$. The exponential map is defined as $\exp_p(vt) = \gamma_{p,v}(t)$. Note that from the compactness of M, $\exp_p : T_pM \to M$ is defined globally on T_pM. For any two nonzero vectors $u, v \in \mathbb{R}^D$, we let $\angle(u, v) = d_{\mathcal{S}^{D-1}}(\frac{u}{\|u\|}, \frac{v}{\|v\|})$ be the angle between u and v.

2.2. Reach

First introduced by Federer [22], the reach regularity parameter is defined as follows. Given a closed subset $A \subset \mathbb{R}^D$, the medial axis $Med(A)$ of A is the

subset of $\mathbb{R}^D$ consisting of the points that have at least two nearest neighbors on A. Namely, denoting by $d(z, A) = \inf_{p\in A} \|p - z\|$ the distance function to A,

$$Med(A) = \left\{ z \in \mathbb{R}^D | \exists p \neq q \in A, \|p - z\| = \|q - z\| = d(z, A) \right\}. \tag{2.1}$$

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

Definition 2.1. The reach of a closed subset $A \subset \mathbb{R}^D$ is defined as

$$\tau_A = \inf_{p\in A} d\left(p, Med(A)\right) = \inf_{z\in Med(A)} d\left(z, A\right). \tag{2.2}$$

Some authors refer to τ_A^{-1} as the *condition number* [33, 36]. From the definition of the medial axis in (2.1), 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 $\left\{x \in \mathbb{R}^D | d(x, A) < \rho\right\}$. Hence, the reach condition can be seen as a generalization of convexity, since a set $A \subset \mathbb{R}^D$ is convex if and only if $\tau_A = \infty$. In the case of submanifolds, one can reformulate the definition of the reach in the following manner.

Theorem 2.2 (Theorem 4.18 in [22])**.**

$$\tau_M = \inf_{q\neq p\in M} \frac{\|q - p\|^2}{2d(q - p, T_pM)}. \tag{2.3}$$

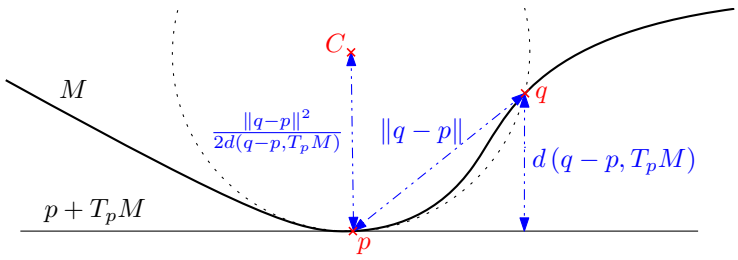


Figure 1: Geometric interpretation of quantities involved in (2.3).

This formulation has the advantage of involving only points on M and its tangent spaces, while (2.2) uses the distance to the medial axis $Med(M)$, which is a global quantity. The formula (2.3) will be the starting point of the estimator proposed in this paper (see Section 4).

The ratio appearing in (2.3) can be interpreted geometrically, as suggested in Figure 1. This ratio is the radius of an ambient ball, tangent to M at p and passing through q. 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 2.3 (Proposition 6.1 in [33])**.** *Let $M \subset \mathbb{R}^D$ 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 $\{M \subset \mathbb{R}^D | \tau_M \geq \tau_{min} > 0\}$ can be interpreted as the Hölder space $\mathcal{C}^2(1/\tau_{min})$. In addition, as illustrated in Figure 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.

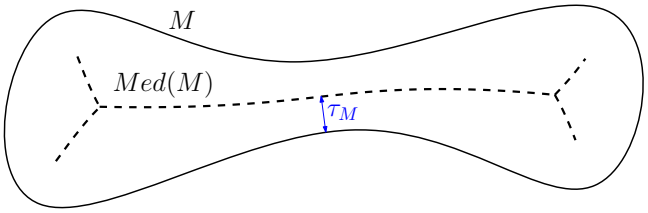


Figure 2: A narrow bottleneck structure yields a small reach τ_M.

2.3. 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 2.9 — to require an extra degree of smoothness. For example, by imposing an upper bound on the third order derivatives of geodesics.

Definition 2.4. Let $\mathcal{M}^{d,D}_{\tau_{min},L}$ denote the set of compact connected d-dimensional submanifolds $M \subset \mathbb{R}^D$ 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

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

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 $\hat{\tau}$ (4.1) does not require them in its construction.

It is important to note that any compact d-dimensional $\mathcal{C}^3$-submanifold $M \subset \mathbb{R}^D$ belongs to such a class $\mathcal{M}^{d,D}_{\tau_{min},L}$, 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 $M \subset \mathbb{R}^D$ of dimension d inherits a natural measure vol_M from the d-dimensional Hausdorff measure $\mathcal{H}^d$ on $\mathbb{R}^D$ [23, p. 171]. We will consider distributions Q that have densities with respect to vol_M that are bounded away from zero.

Definition 2.5. We let $\mathcal{Q}^{d,D}_{\tau_{\min},L,f_{\min}}$ denote the set of distributions Q having support $M \in \mathcal{M}^{d,D}_{\tau_{min},L}$ 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 $\hat{\tau}$ (4.1) 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 [32]. This assumption will eventually be relaxed in Section 6.

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

Definition 2.6. For any distribution $Q \in \mathcal{Q}^{d,D}_{\tau_{\min},L,f_{\min}}$ with support M we associate the distribution P of the random variable $(X, T_X M)$ on $\mathbb{R}^D \times \mathbb{G}^{d,D}$, where X has distribution Q. We let $\mathcal{P}^{d,D}_{\tau_{\min},L,f_{\min}}$ 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 $(X_1, T_1), \dots, (X_n, T_n) \in \mathbb{R}^D \times \mathbb{G}^{d,D}$, where $X_1, \dots, X_n$ 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 $\mathbb{R}^D \times \mathbb{G}^{d,D}$, we will write $\tau_P = \tau_Q = \tau_M$, with a slight abuse of notation.

Note that the model does not explicitly impose an upper bound on τ_M. Such an upper bound would be redundant, since the lower bound on f_{min} does impose such an upper bound, as we now state in the following result. The proof relies on a volume argument (Lemma A.2), leading to a bound on the diameter of M, and on a topological argument (Lemma A.3) to link the reach and the diameter.

Proposition 2.7. *Let $M \subset \mathbb{R}^D$ be a connected closed d-dimensional manifold, and let Q be a probability distribution with support M. Assume that Q has a density f with respect to the Hausdorff measure on M such that $\inf_{x \in M} f(x) \geq f_{min} > 0$. Then,*

$$\tau_M^d \leq \frac{C_d}{f_{min}},$$

for some constant $C_d > 0$ depending only on d.

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

$$\ell(\tau, \tau') = \left| \frac{1}{\tau} - \frac{1}{\tau'} \right|^p, \quad p \geq 1. \tag{2.4}$$

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

Remark 2.8. For a distribution $P \in \mathcal{P}^{d,D}_{\tau_{\min},L,f_{\min}}$, Proposition 2.7 asserts that $\tau_{min} \leq \tau_P \leq \tau_{max} := (C_d / f_{min})^{1/d}$. Therefore, in an inference set-up, we can always restrict to estimators $\hat{\tau}$ within the bounds $\tau_{min} \leq \hat{\tau} \leq \tau_{max}$. Consequently,

$$\frac{1}{\tau_{max}^{2p}} |\tau_P - \hat{\tau}|^p \leq \left| \frac{1}{\tau_P} - \frac{1}{\hat{\tau}} \right|^p \leq \frac{1}{\tau_{min}^{2p}} |\tau_P - \hat{\tau}|^p,$$

so that the estimation of the reach τ_P is equivalent to the estimation of the condition number τ_P^{-1}, up to constants.

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 2.9 demonstrates that relaxing this constraint — *i.e.* setting $L = \infty$ — renders the problem of reach estimation intractable. Its proof is to be found in Section D.3. Below, σ_d stands for the volume of the d-dimensional unit sphere $\mathcal{S}^d$.

Proposition 2.9. *There exists a universal constant $c > 1/100$ such that given $\tau_{min} > 0$, provided that $f_{min} \leq (2^{d+1}\tau_{min}^d \sigma_d)^{-1}$, we have for all $n \geq 1$,*

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

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. 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 3.4). 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.

Consider the formulation (2.2) of the reach as the infimum of the distance between M and its medial axis $Med(M)$. By definition of the medial axis (2.1), 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 3b). By definition, q_1 and q_2 belong to $\mathcal{B}(z_0, \tau_M)$ 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 $\mathcal{B}(z_0, \tau_M)$ (see Figure 3a). Note that the ball $\mathcal{B}(z_0, \tau_M)$ meets M only on its boundary $\partial\mathcal{B}(z_0, \tau_M)$.

Definition 3.1. Let $M \subset \mathbb{R}^D$ 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 $q_1, q_2 \in \mathcal{B}(z_0, \tau_M)$. 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 3.2, if a reach attaining pair is not a bottleneck — that is $\|q_1 - q_2\| < 2\tau_M$, as in Figure 3b —, 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 3.2. *Let $M \subset \mathbb{R}^D$ 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 shorter 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 shorter 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 (2.2) is not attained. In this case, the following Lemma 3.3 asserts that τ_M is created by curvature.

Lemma 3.3. *Let $M \subset \mathbb{R}^D$ 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:

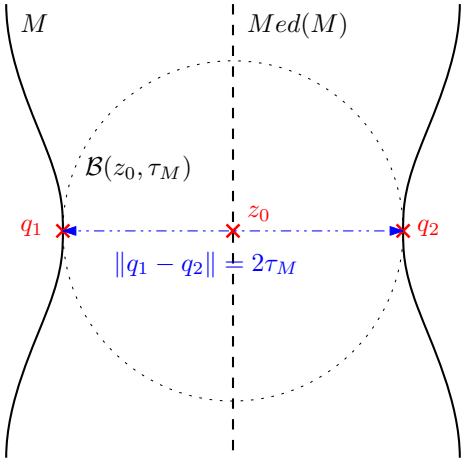
- (See Figure 3a) M has a bottleneck: by definition, τ_M originates from a structure having scale $2\tau_M$.
- (See Figure 3b) M has a reach attaining pair but no bottleneck: then M contains an arc of a circle of radius τ_M (Lemma 3.2), so that M actually contains a zone with radius of curvature τ_M.
- (See Figure 3c) M does not have a reach attaining pair: then τ_M comes from a curvature-attaining point (Lemma 3.3), 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 3.2 and Lemma 3.3.

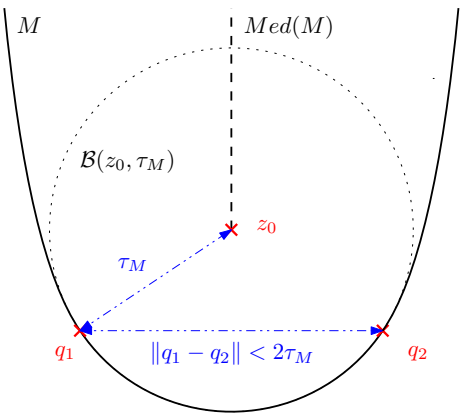
Theorem 3.4. *Let $M \subset \mathbb{R}^D$ be a compact submanifold with reach $\tau_M > 0$. At least one of the following two assertions holds.*

- (Global Case) *M has a bottleneck $(q_1, q_2) \in M^2$, that is, there exists $z_0 \in Med(M)$ such that $q_1, q_2 \in \partial\mathcal{B}(z_0, \tau_M)$ 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 $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$.*

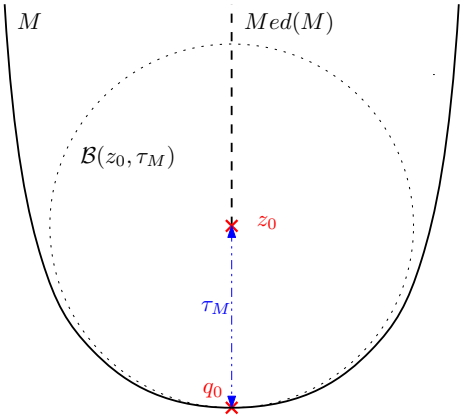
Let us emphasize the fact that the global case and the local case of Theorem 3.4 are not mutually exclusive. Theorem 3.4 provides a description of the reach as arising from global and local geometric structures that, to the best of



(a) A bottleneck.



(b) A non-bottleneck reach attaining pair.



(c) Curvature-attaining point.

Figure 3: The different ways for the reach to be attained, as described in Lemma 3.2 and Lemma 3.3.

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.

4. 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 (2.4). For the sake of clarity in the analysis, we assume the tangent spaces to be known at every sample point. This assumption will be relaxed in Section 6.

We rely on the formulation of the reach given in (2.3) (see also Figure 1), and define $\hat{\tau}$ as a plugin estimator as follows: given a point cloud $\mathbb{X} \subset M$,

$$\hat{\tau}(\mathbb{X}) = \inf_{x \neq y \in \mathbb{X}} \frac{\|y-x\|^2}{2d(y-x, T_x M)}. \tag{4.1}$$

In particular, we have $\hat{\tau}(M) = \tau_M$. Since the infimum (4.1) is taken over a set $\mathbb{X}$ smaller than M, $\hat{\tau}(\mathbb{X})$ always overestimates τ_M. In fact, $\hat{\tau}(\mathbb{X})$ is decreasing in the number of distinct points in $\mathbb{X}$, a useful property that we formalize in the following result, whose proof is immediate.

Corollary 4.1. *Let M be a submanifold with reach τ_M and $\mathbb{Y} \subset \mathbb{X} \subset M$ be two nested subsets. Then $\hat{\tau}(\mathbb{Y}) \geq \hat{\tau}(\mathbb{X}) \geq \tau_M$.*

We now derive the rate of convergence of $\hat{\tau}$. We analyze the global case (Section 4.1) and the local case (Section 4.2) 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 4.1 and Section 4.2 are to be found in Section C.1 and Section C.2.

4.1. Global Case

Consider the global case, that is, M has a bottleneck structure (Theorem 3.4). Then the infimum (2.3) is achieved at a bottleneck pair $(q_1, q_2) \in M^2$. When $\mathbb{X}$ 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 (2.3): that is, that $\hat{\tau}(\mathbb{X})$ should be close to τ_M.

Proposition 4.2. *Let $M \subset \mathbb{R}^D$ be a submanifold with reach $\tau_M > 0$ that has a bottleneck $(q_1, q_2) \in M^2$ (see Definition 3.1), and $\mathbb{X} \subset M$. If there exist $x, y \in \mathbb{X}$ with $\|q_1 - x\| < \tau_M$ and $\|q_2 - y\| < \tau_M$, then*

$$0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x,y\})} \leq \frac{4}{\tau_M^2} \max\left\{d_M(q_1,x), d_M(q_2,y)\right\}.$$

The error made by $\hat{\tau}(\mathbb{X})$ 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 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 4.2, we can now give an upper bound on the expected loss under the model $\mathcal{P}^{d,D}_{\tau_{\min},L,f_{\min}}$. We recall that, throughout the paper, $\mathbb{X}_n = \{X_1, \dots, X_n\}$ is an i.i.d. sample with common distribution Q associated to P (see Definition 2.6).

Proposition 4.3. *Let $P \in \mathcal{P}^{d,D}_{\tau_{\min},L,f_{\min}}$ and $M = supp(P)$. Assume that M has a bottleneck $(q_1, q_2) \in M^2$ (see Definition 3.1). Then,*

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

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

Proposition 4.3 follows straightforwardly from Proposition 4.2 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 $\mathbb{X}_n$.

4.2. 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 3.4). 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 $\hat{\tau}$ (4.1). Indeed, since the latter quantity is the radius of a sphere tangent at x and passing through y (Figure 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 4.4. *Let $M \in \mathcal{M}^{d,D}_{\tau_{\min},L}$ with reach τ_M and $\mathbb{X} \subset M$ be a subset. Let $x, y \in \mathbb{X}$ with $d_M(x, y) < \pi\tau_M$. Then,*

$$0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x,y\})} \leq \frac{1}{\tau_M} - \left\|\gamma''_{x\to y}(0)\right\| + \frac{1}{3}L d_M(x,y).$$

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,

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

Lemma 4.5. *Let $M \in \mathcal{M}^{d,D}_{\tau_{\min},L}$ with reach τ_M and $q_0, x, y \in M$ be such that $x, y \in \mathcal{B}_M\left(q_0, \frac{\pi\tau_M}{2}\right)$. 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 $|\theta_x - \theta_y| \geq \frac{\pi}{2}$. Then,

$$\left\|\gamma''_{x\to y}(0)\right\| \geq \kappa_{q_0} - (\kappa_x - \kappa_{q_0}) - 2Ld_M(q_0, x) - (2\kappa_x + 6\kappa_{q_0})\sin^2(|\theta_x - \theta_y|).$$

In particular, geodesics in a neighborhood of q_0 with directions close to v_0 have curvature close to $\frac{1}{\tau_M}$. Combining Lemma 4.4 and Lemma 4.5 yields the following deterministic bound in the local case.

Proposition 4.6. *Let $M \in \mathcal{M}^{d,D}_{\tau_{\min},L}$ be such that there exist $q_0 \in M$ and a geodesic γ_0 such that $\gamma_0(0) = q_0$ and $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$. Let $\mathbb{X} \subset M$ and $x, y \in \mathbb{X}$ be such that $x, y \in \mathcal{B}_M\left(q_0, \frac{\pi\tau_M}{2}\right)$. 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 $|\theta_x - \theta_y| \geq \frac{\pi}{2}$. Then,

$$\begin{aligned} 0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} &\leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x, y\})} \\ &\leq \frac{8\sin^2(|\theta_x - \theta_y|)}{\tau_M} + L\left(\frac{1}{3}d_M(x, y) + 2d_M(q_0, x)\right). \end{aligned}$$

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 4.6) is very similar to that of the global case (Proposition 4.2) 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 3.4). Hence, it is natural that the accuracy of $\hat{\tau}(\mathbb{X})$ depends on how precisely $\mathbb{X}$ 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 4.6 yields a bound on the risk of $\hat{\tau}(\mathbb{X}_n)$ when $\mathbb{X}_n = \{X_1, \ldots, X_n\}$ is random.

Proposition 4.7. *Let $P \in \mathcal{P}^{d,D}_{\tau_{\min},L,f_{\min}}$ 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,*

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

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 4.6 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 4.3 and 4.7 yield a convergence rate of $\hat{\tau}(\mathbb{X}_n)$ which is slower in the local case than in the global case. Recall that from Theorem 3.4, 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 4.3 and 4.7 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.

5. Minimax Estimates

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

$$R_n = \inf_{\hat{\tau}_n} \sup_{P \in \mathcal{P}^{d,D}_{\tau_{\min},L,f_{min}}} \mathbb{E}_{P^n} \left| \frac{1}{\tau_P} - \frac{1}{\hat{\tau}_n} \right|^p , \tag{5.1}$$

where the infimum ranges over all estimators $\hat{\tau}_n\big((X_1, T_{X_1}), \dots, (X_n, T_{X_n})\big)$ based on an i.i.d. sample of size n with the knowledge of the tangent spaces at sample points. The minimax risk R_n corresponds to the best expected risk that an estimator, based on n samples, can achieve uniformly over the model $\mathcal{P}^{d,D}_{\tau_{\min},L,f_{min}}$ without the knowledge of the underlying distribution P.

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

Theorem 5.1. *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 [38]. 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 [38, Lemma 1] and $(1 - TV(P^n, P'^n)) \geq (1 - TV(P, P'))^n$.

Lemma 5.2 (Le Cam's Lemma)**.** *Let $P, P' \in \mathcal{P}^{d,D}_{\tau_{\min},L,f_{\min}}$ with respective supports M and M'. Then for all $n \geq 1$,*

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

Lemma 5.2 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 $\mathcal{P}^{d,D}_{\tau_{min},L,f_{min}}$ requires the hypotheses to belong to the class. Luckily, in our problem it will be enough to construct hypotheses from the simpler class $\mathcal{Q}^{d,D}_{\tau_{\min},L,f_{\min}}$. Indeed, we have the following isometry result between $\mathcal{Q}^{d,D}_{\tau_{\min},L,f_{\min}}$ and $\mathcal{P}^{d,D}_{\tau_{min},L,f_{min}}$ for the total variation distance, as proved in Section D.2. We use here the notation of Definition 2.6

Lemma 5.3. *Let $Q, Q' \in \mathcal{Q}^{d,D}_{\tau_{\min},L,f_{\min}}$ be distributions on $\mathbb{R}^D$ with associated distributions $P, P' \in \mathcal{P}^{d,D}_{\tau_{\min},L,f_{\min},}$ on $\mathbb{R}^D \times \mathbb{G}^{d,D}$. Then,*

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

In order to construct hypotheses in $\mathcal{Q}^{d,D}_{\tau_{\min},L,f_{\min}}$ we take advantage of the fact that the class $\mathcal{M}^{d,D}_{\tau_{min},L}$ 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 $\mathbb{R}^D \supset M$. Given a smooth map $\Phi : \mathbb{R}^D \to \mathbb{R}^D$, 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 5.4 can be found in Section D.3.

Proposition 5.4. *Let $M \in \mathcal{M}^{d,D}_{\tau_{min}L}$ be fixed, and let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ be a global $\mathcal{C}^3$-diffeomorphism. If $\|I_D - d\Phi\|_{op}$, $\left\|d^2\Phi\right\|_{op}$ and $\left\|d^3\Phi\right\|_{op}$ are small enough, then $M' = \Phi(M) \in \mathcal{M}^{d,D}_{\frac{\tau_{min}}{2},2L}$.*

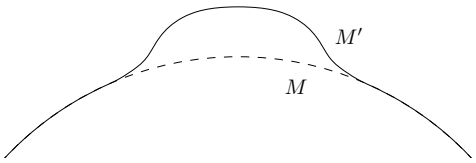


Figure 4: Hypotheses of Proposition 5.5.

Now we construct the two hypotheses Q, Q' as follows (see Figure 4). 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 5.5 is to be found in Section D.3.

Proposition 5.5. *Assume that* $L \geq (2\tau_{min}^2)^{-1}$ *and* $f_{min} \leq (2^{d+1}\tau_{min}^d \sigma_d)^{-1}$. *Then for* $\ell > 0$ *small enough, there exist* $Q, Q' \in \mathcal{Q}^{d,D}_{\tau_{\min}, L, f_{min}}$ *with respective supports* M *and* M' *such that*

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

Hence, applying Lemma 5.2 with the hypotheses P, P' associated to Q, Q' of Proposition 5.5, and taking $12\,(\ell/2\tau_{min})^d = 1/n$, together with Lemma 5.3, yields the following lower bound.

Proposition 5.6. *Assume that* $L \geq (2\tau_{min}^2)^{-1}$ *and* $f_{min} \leq (2^{d+1}\tau_{min}^d \sigma_d)^{-1}$. *Then for* n *large enough,*

$$R_n \geq \left(\frac{c_d}{\tau_{min}} \right)^p n^{-p/d},$$

where $c_d > 0$ *depends only on* d.

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 5.6, the plugin estimation $\hat{\tau}(\mathbb{X}_n)$ provably achieves the optimal rate in the global case (Theorem 4.3) up to numerical constants. In the local case (Theorem 4.7) the rate obtained presents a gap, yielding a gap in the overall rate. As explained above (Section 4.2), the slower rate in the local case is a consequence of the alignment required in order to estimate directional curvature. Though, let us note that in the one-dimensional case $d = 1$, the rate of Proposition 5.6 matches the convergence rate of $\hat{\tau}(\mathbb{X}_n)$ (Theorem 5.1). Indeed, for curves, the alignment requirement is always fulfilled. Hence, the rate is exactly n^{-p} for $d = 1$, and $\hat{\tau}(\mathbb{X}_n)$ is minimax optimal.

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

6. Towards Unknown Tangent Spaces

So far, in our analysis we have used the key assumption that both the point cloud and the tangent spaces were jointly observed. We now focus on the more realistic framework where only points are observed. We once again rely on the formulation of the reach given in Theorem 2.3 and consider a new plug-in estimator in which the true tangent spaces are replaced by estimated ones. Namely,

given a point cloud $\mathbb{X} \subset \mathbb{R}^D$ and a family $T = \{T_x\}_{x\in\mathbb{X}}$ of linear subspaces of $\mathbb{R}^D$ indexed by $\mathbb{X}$, the estimator is defined as

$$\hat{\tau}(\mathbb{X}, T) = \inf_{x\neq y\in\mathbb{X}} \frac{\|y-x\|^2}{2d(y-x, T_x)}. \tag{6.1}$$

In particular, $\hat{\tau}(\mathbb{X}) = \hat{\tau}(\mathbb{X}, T_{\mathbb{X}}M)$, where $T_{\mathbb{X}}M = \{T_xM\}_{x\in\mathbb{X}}$. Adding uncertainty on tangent spaces in (6.1) does not change drastically the estimator as the formula is stable with respect to T. We state this result quantitatively in the following Proposition 6.1, the proof of which can be found in Section E. In what follows, the distance between two linear subspaces $U, V \in \mathbb{G}^{d,D}$ is measured with their principal angle $\|\pi_U - \pi_V\|_{op}$.

Proposition 6.1. *Let $\mathbb{X} \subset \mathbb{R}^D$ and $T = \{T_x\}_{x\in\mathbb{X}}$, $\tilde{T} = \{\tilde{T}_x\}_{x\in\mathbb{X}}$ be two families of linear subspaces of $\mathbb{R}^D$ indexed by $\mathbb{X}$. Assume $\mathbb{X}$ to be δ-sparse, T and $\tilde{T}$ to be θ-close, in the sense that*

$$\inf_{x\neq y\in\mathbb{X}} \|y-x\| \geq \delta \quad \text{and} \quad \sup_{x\in\mathbb{X}} \|T_x - \tilde{T}_x\|_{op} \leq \sin\theta.$$

Then,

$$\left| \frac{1}{\hat{\tau}(\mathbb{X}, T)} - \frac{1}{\hat{\tau}(\mathbb{X}, \tilde{T})} \right| \leq \frac{2\sin\theta}{\delta}.$$

In other words, the map $T \mapsto \hat{\tau}(\mathbb{X}, T)^{-1}$ is smooth, provided that the basis point cloud $\mathbb{X}$ contains no zone of accumulation at a too small scale $\delta > 0$. As a consequence, under the assumptions of Proposition 6.1, the bounds on $\left|\hat{\tau}(\mathbb{X})^{-1} - {\tau_M}^{-1}\right|$ of Proposition 4.2 and Proposition 4.6 still hold with an extra error term $2\sin\theta/\delta$ if we replace $\hat{\tau}(\mathbb{X})$ by $\hat{\tau}(\mathbb{X}, T)$.

For an i.i.d. point cloud $\mathbb{X}_n$, asymptotic and nonasymptotic rates of tangent space estimation derived in $\mathcal{C}^3$-like models can be found in [2, 14, 36], yielding bounds on $\sin\theta$ of order $(\log n/n)^{1/d}$. In that case, the typical scale of minimum interpoint distance is $\delta \asymp n^{-2/d}$, as stated in the asymptotic result Theorem 2.1 in [29] for the flat case of $\mathbb{R}^d$. However, the typical covering scale of M used in the global case (Theorem 4.3) is $\varepsilon \asymp (1/n)^{1/d}$. It appears that we can sparsify the point cloud $\mathbb{X}_n$ — that is, removing accumulation points — while preserving the covering property at scale $\varepsilon = 2\delta \asymp (\log n/n)^{1/d}$. This can be performed using the *farthest point sampling algorithm* [1, Section 3.3]. Such a sparsification pre-processing allows to lessen the possible instability of $\hat{\tau}(\mathbb{X}_n, \cdot)^{-1}$. Though, whether the alignment property used in the local case (Theorem 4.7) is preserved under sparsification remains to be investigated.

7. Conclusion and Open Questions

In the present work, we gave new insights on the geometry of the reach. Inference results were derived in both deterministic and random frameworks. For i.i.d. samples, non-asymptotic minimax upper and lower bounds were derived under assumptions on the third order derivative of geodesic trajectories. Let us conclude with some open questions.

- Interestingly, the derivation of the minimax lower bound (Theorem 5.6) involves hypotheses that correspond to the local case, but yields the rate $n^{-p/d}$. But, on the upper bound side, this rate matches with that of the global case (Theorem 4.3), the local case being slower (Theorem 4.7). The minimax upper and lower bounds given in Theorem 5.1 and Theorem 5.6 do not match. They are yet to be sharpened. This results into minimax upper and lower bounds that do not match. They are yet to be sharpened.
- As mentioned earlier, Section 6 is only a first step towards a framework where tangent spaces are unknown. A minimax upper bound in this case is still an open question. Considering smoother $\mathcal{C}^k$ models ($k \geq 3$) such as those of [2], or data with additive noise would also be of interest.
- In practice, since large reach ensures regularity, one may be interested with having a lower bound on the reach τ_M. Studying the limiting distribution of the statistic $\hat{\tau}(\mathbb{X}_n)$ would allow to derive asymptotic confidence intervals for τ_M.
- Other regularity parameters such as local feature size [10] and λ-reach [13] could be relevant to estimate, as they are used as tuning parameters in computational geometry techniques.

Acknowledgments

This collaboration was made possible by the associated team CATS (Computations And Topological Statistics) between DataShape and Carnegie Mellon University. We thank warmly the anonymous reviewers for their insight, which led to various significant improvements of the paper.

Appendix A: Some Technical Results on the Model

A.1. Geometric Properties

The following Proposition A.1 garners geometric properties of submanifolds of the Euclidean space that are related to the reach. We will use them numerous times in the proofs.

Proposition A.1. *Let $M \subset \mathbb{R}^D$ be a closed 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 K of M satisfy $-\frac{2}{\tau_M^2} \leq K \leq \frac{1}{\tau_M^2}$.
(iv) For all $p \in M$, the map $\exp_p : \mathring{\mathcal{B}}_{T_pM}(0, \pi\tau_M) \to \mathring{\mathcal{B}}_M(0, \pi\tau_M)$ 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 $A \subset \mathcal{B}_{T_pM}(0,r) \subset T_pM$,

$$\left(1 - \frac{r^2}{6\tau_M^2}\right)^d \mathcal{H}^d(A) \leq \mathcal{H}^d(\exp_p(A)) \leq \left(1 + \frac{r^2}{\tau_M^2}\right)^d \mathcal{H}^d(A).$$

Proof of Proposition A.1. (i) is stated as in [33, Proposition 2.1], yielding (ii) from [3, Corollary 1.4]. (iii) follows using (i) again and the Gauss equation [20, p. 130]. (iv) is derived from (iii) by a direct application of [21, Lemma 8]. (v) follows from (iv) and [4, Lemma 6]. □

A.2. Comparing Reach and Diameter

Let us prove Proposition 2.7. For this aim, we first state the following analogous bound on the (Euclidean) diameter $diam(M) = \sup_{x,y \in M} \|x - y\|$.

Lemma A.2 (Lemma 2 in [1])**.** *Let $M \subset \mathbb{R}^D$ be a connected closed d-dimensional manifold, and let Q be a probability distribution having support M with a density $f \geq f_{min}$ with respect to the Hausdorff measure on M. Then,*

$$diam(M) \leq \frac{C_d}{\tau_M^{d-1} f_{min}},$$

for some constant $C_d > 0$ depending only on d.

Proposition A.3. *If $K \subset \mathbb{R}^D$ is not homotopy equivalent to a point,*

$$\tau_K \leq \sqrt{\frac{D}{2(D+1)}} diam(K).$$

Proof of Proposition A.3. Combine Lemma A.4 and Lemma A.5. □

Let us recall that for two compact subsets $A, B \subset \mathbb{R}^D$, the Hausdorff distance [12, p. 252] between them is defined by

$$d_H(A,B) = \max\left\{\sup_{a \in A} d(a,B), \sup_{b \in B} d(b,A)\right\}.$$

We denote by $conv(\cdot)$ the closed convex hull of a set.

Lemma A.4. *For all $K \subset \mathbb{R}^D$, $d_H(K, conv(K)) \leq \sqrt{\frac{D}{2(D+1)}} diam(K)$.*

Proof of Lemma A.4. It is a straightforward corollary of Jung's Theorem 2.10.41 in [23], which states that K is contained in a (unique) closed ball with (minimal) radius at most $\sqrt{\frac{D}{2(D+1)}} diam(K)$. □

Lemma A.5. *If $K \subset \mathbb{R}^D$ is not homotopy equivalent to a point, then $\tau_K \leq d_H(K, conv(K))$.*

Proof of Lemma A.5. Let us prove the contrapositive. For this, assume that $\tau_K > d_H(K, conv(K))$. Then,

$$conv(K) \subset \bigcup_{x\in K} \mathcal{B}(x, d_H(K, conv(K))) \subset \bigcup_{x\in K} \mathring{\mathcal{B}}(x, \tau_K) \subset Med(K)^c.$$

Therefore, the map $\pi_K : conv(K) \to K$ is well defined and continuous, so that K is a retract of $conv(K)$ (see [27, Chapter 0]). Therefore, K is homotopy equivalent to a point, since the convex set $conv(K)$ is. □

We are now in position to prove Proposition 2.7.

Proof of Proposition 2.7. From [27, Theorem 3.26], M has a non trivial homology group of dimension d over $\mathbb{Z}/2\mathbb{Z}$, so that it cannot be homotopy equivalent to a point. Therefore, Proposition A.3 yields $\tau_M \leq diam(M)$, and we conclude by applying the bound $diam(M) \leq C_d/(\tau_M^{d-1} f_{min})$ given by Lemma A.2. □

Appendix B: Geometry of the Reach

Lemma B.1. *Let $V \subset \mathbb{R}^D$ be a 2-dimensional affine space and $q_1, q_2, z, p \in V$ be such that $\|p - q_1\| = \|p - q_2\| = r_p$ and $\|z - q_1\| = \|z - q_2\| = r_z$. If $r_p < r_z$, then*

$$V \cap \partial\mathcal{B}_{\mathbb{R}^D}(z, r_z) \cap \mathcal{B}_{\mathbb{R}^D}(p, r_p) = c_z(q_1, q_2),$$

where $c_z(q_1, q_2)$ is the shorter arc of the circle with center z and endpoints as q_1 and q_2.

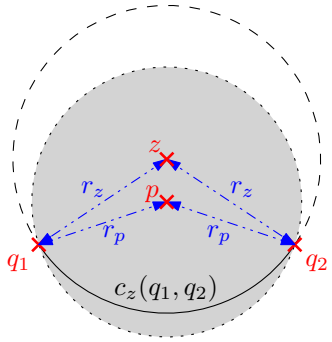


Figure 5: Layout of Lemma B.1.

Proof of Lemma B.1. Since everything is intersected with the 2-dimensional space V, we can assume that $D = 2$ without loss of generality. For short, we write $K = \partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \mathcal{B}_{\mathbb{R}^2}(p, r_p)$.

First note that $\{q_1, q_2\} \subset K$, so that $K \neq \emptyset$. Furthermore, for all $x \in K$, $d(x, Med(\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z)) = r_z > r_p$, so that $\tau_K > r_p$ from [34, Lemma 3.4 (i)].

Hence, applying [34, Lemma 3.4 (ii)], we get that K is contractible. In particular, K is connected.

Since K is a closed connected subset of the circle $\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z)$, K is an arc of a circle. Let c_1, c_2 denote its endpoints.

Let us now show that $\{c_1, c_2\} \subset \partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \partial\mathcal{B}_{\mathbb{R}^2}(p, r_p)$, or equivalently that $\|c_1 - p\| = \|c_2 - p\| = r_p$. Indeed, if $x \in K$ is such that $\|x - p\| < r_p$ then there exists $r_x > 0$ such that $\mathcal{B}_{\mathbb{R}^2}(x, r_x) \subset \mathcal{B}_{\mathbb{R}^2}(p, r_p)$. Then $\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \mathcal{B}_{\mathbb{R}^2}(x, r_x) \neq \emptyset$, so $\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \mathcal{B}_{\mathbb{R}^2}(x, r_x)$ is also an arc of a circle, and since $x \in \partial\mathcal{B}_{\mathbb{R}^2}(z, r_z)$, x cannot be an end point of the arc $\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \mathcal{B}_{\mathbb{R}^2}(x, r_x)$.

The two circles $\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z)$ and $\partial\mathcal{B}_{\mathbb{R}^2}(p, r_p)$ are different ($r_z > r_p$), so their intersection contains at most two points. Since $q_1 \neq q_2 \in K = \partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \mathcal{B}_{\mathbb{R}^2}(p, r_p)$, in fact $\{q_1, q_2\} = \partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \partial\mathcal{B}_{\mathbb{R}^2}(p, r_p)$. Consequently, $\{c_1, c_2\} = \{q_1, q_2\}$. That is, q_1 and q_2 are the endpoints of the arc K.

Note that there are two arcs of the circle $\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z)$ with endpoints q_1 and q_2. Since $K = \partial\mathcal{B}_{\mathbb{R}^2}(z, r_z) \cap \mathcal{B}_{\mathbb{R}^2}(p, r_p) \subset \mathcal{B}_{\mathbb{R}^2}(p, r_p)$ and $r_p < r_z$, K cannot contain two points at distance equal to $2r_z$. Hence, K is the shorter arc of the circle $\partial\mathcal{B}_{\mathbb{R}^2}(z, r_z)$ with endpoints q_1 and q_2, which is exactly $c_z(q_1, q_2)$.

□

Lemma B.2. *Let $V \subset \mathbb{R}^D$ be a 2-dimensional affine space and $q_1, q_2, z, x \in V$. Denote by L be the line passing q_1 and q_2. Assume that $x, z \notin L$, and that the segment joining z and x intersects L. Let $p \in \mathbb{R}^D$ be such that $\|p - q_1\| = \|z - q_1\|$ and $\|p - q_2\| = \|z - q_2\|$. Then $\|p - x\| \leq \|z - x\|$, and the equality holds if and only if $p = z$.*

Proof of Lemma B.2. Let y denote the intersection point of L and the line segment between z and x. Since $\|p - q_1\| = \|z - q_1\|$ and $\|p - q_2\| = \|z - q_2\|$,

$$\begin{aligned}\cos(\angle(p - q_1, q_2 - q_1)) &= \cos(\angle(z - q_1, q_2 - q_1)) \\ &= \frac{\|z - q_1\|^2 + \|q_2 - q_1\|^2 - \|z - q_2\|^2}{2\,\|z - q_1\|\,\|q_2 - q_1\|},\end{aligned}$$

from which we derive

$$\begin{aligned}\|p - y\|^2 &= \|p - q_1\|^2 + \|y - q_1\|^2 - 2\,\|p - q_1\|\,\|y - q_1\|\cos(\angle(p - q_1, q_2 - q_1)) \\ &= \|z - q_1\|^2 + \|y - q_1\|^2 - 2\,\|z - q_1\|\,\|y - q_1\|\cos(\angle(z - q_1, q_2 - q_1)) \\ &= \|z - y\|^2.\end{aligned}$$

Using the fact that y belongs to the segment joining x and z, we get

$$\begin{aligned}\|z - x\| &= \|z - y\| + \|y - x\| \\ &= \|p - y\| + \|y - x\| \\ &\geq \|p - x\|.\end{aligned}$$

Finally, note that since $x, z \notin L$ and $y \in L$, the equality holds if and only if $\angle(x - y, p - y) = \pi$. But $\|p - y\| = \|z - y\|$ and x, y, and z are colinear, so this is possible if and only if $p = z$.

□

The following lemma can be seen as an extension of [34, Lemma 3.4 (i)].

Lemma B.3. *Let $A \subset \mathbb{R}^D$ be a set with positive reach $\tau_A > 0$, and let $\{B_i\}_{i\in I}$ be a collection of balls indexed by I. Suppose $\bigcap_{i\in I} B_i \cap A$ is nonempty. Let r_i be the radius of B_i, and suppose $r_i < \tau_A$ for all $i \in I$.*

(i) If I is finite, then $\tau_{\bigcap_{i\in I} B_i\cap A} > \min_{i\in I} r_i$.
(ii) If I is countably infinite, then $\tau_{\bigcap_{i\in I} B_i\cap A} \geq \inf_{i\in I} r_i$.

Proof of Lemma B.3. (i) Since I is finite, we can assume that $I = \{1, \dots, k\}$ and that the sequence $(r_i)_{1\leq i\leq k}$ is nonincreasing. We use an induction on k:

- If $k = 1$, since for all $x \in A \cap B_1$, $d(x, Med(A)) \geq \tau_A > r_1$, [34, Lemma 3.4 (i)] gives that $\tau_{B_1\cap A} > r_1$.
- Suppose now that $\tau_{\bigcap_{i-1}^{j} B_i\cap A} > r_j$ for some $j < k$. Then for all $x \in \bigcap_{i=1}^{j+1} B_i \cap A = \left(\bigcap_{i=1}^{j} B_i \cap A\right) \cap B_{j+1}$, $d\left(x, Med\left(\bigcap_{i=1}^{j} B_i \cap A\right)\right) > r_j \geq r_{j+1}$. Applying again [34, Lemma 3.4 (i)] gives

$$\tau_{\bigcap_{i=1}^{j+1} B_i\cap A} = \tau_{(\bigcap_{i=1}^{j} B_i\cap A)\cap B_{j+1}} > r_{j+1} = \min_{1\leq i\leq j+1} r_i.$$

By induction on k, we get the result.

(ii) Note that if $\inf_{i\in I} r_i = 0$, there is nothing to prove. Hence we only consider the case where $\inf_{i\in I} r_i > 0$.
Since I is countable, we can assume that $I = \mathbb{N}$. For $k \in \mathbb{N}$, let $C_k := \bigcap_{i=1}^{k} B_i \cap A$. In particular, $C_\infty := \cap_{k=1}^{\infty} C_k = \cap_{i=1}^{\infty} B_i \cap A$. From the finite case (i),

$$\tau_{C_k} > \min_{1\leq i\leq k} r_i \geq \inf_{i\in\mathbb{N}} r_i.$$

Now, since $\{C_k\}_{k=1}^{\infty}$ is a decreasing sequence of sets, the distance functions $d(\cdot, C_k)$ converges to $d(\cdot, C_\infty)$. As the distance functions $d(\cdot, C_k)$ are continuous, there convergence is uniform on any compact subset of $\mathbb{R}^D$. Hence, [22, Theorem 5.9] yields $\tau_{C_\infty} \geq \inf_{i\in\mathbb{N}} r_i$, which concludes the proof.

□

Proof of Lemma 3.2. Let $p_0 := \frac{z_0+q_1+q_2}{3}$ and $\tau_0 = \|p_0 - q_1\| < \tau_M$. Consider the subset C_0 of the median hyperplane of q_1 and q_2 defined by

$$C_0 := \left\{p \in \mathbb{R}^D \,|\, \|p - q_1\| = \|p - q_2\| \in (\tau_0, \tau_M)\right\},$$

and let $\{p_i\}_{i\in\mathbb{N}} \subset C_0$ be its countable dense subset. Write $\tau_i := \|p_i - q_1\|$ and $B_i := \mathcal{B}_{\mathbb{R}^D}(p_i, \tau_i)$. Let $A_\infty := \bigcap_{k=0}^{\infty} B_k$. Note that $\{q_1, q_2\} \subset M \cap A_\infty$ which implies that $M \cap A_\infty$ is nonempty. Note also that by definition, $\tau_i \in (\tau_0, \tau_M)$ for all $i \in \mathbb{N} \cup \{0\}$. Hence from Lemma B.3 (ii), $\tau_{M\cap A_\infty} \geq \tau_0$. In addition, $M \subset \mathbb{R}^D \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M)$, so that

$$\{q_1, q_2\} \subset M \cap A_\infty \subset A_\infty \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M).$$

Note that it is sufficient to show that $A_\infty \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M) = c_{z_0}(q_1, q_2)$ to conclude the proof. Indeed, since $\tau_{M\cap A_\infty} \geq \tau_0 > \frac{\|q_2-q_1\|}{2}$ and $\emptyset \neq M \cap A_\infty \subset c_{z_0}(q_1, q_2) \subset \mathcal{B}_{\mathbb{R}^D}\left(\frac{q_1+q_2}{2}, \frac{\|q_2-q_1\|}{2}\right)$, [34, Lemma 3.4 (ii)] implies that $M \cap A_\infty$ is contractible. In other words, $M \cap A_\infty$ is a contractible subset of the shorter arc of a circle $c_{z_0}(q_1, q_2)$ containing its endpoints q_1 and q_2, and hence $M \cap A_\infty = c_{z_0}(q_1, q_2)$. Therefore, $c_{z_0}(q_1, q_2) \subset M$, which concludes the proof.

It is left to show that $A_\infty \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M) = c_{z_0}(q_1, q_2)$. To this aim, let us write $V := z_0 + span\{q_1 - z_0, q_2 - z_0\}$ for the 2-dimensional plane passing through q_1, q_2, and z_0. Then $\tau_0 = \|p_0 - q_1\| = \|p_0 - q_2\| < \|z_0 - q_1\| = \|z_0 - q_2\| = \tau_M$, and hence from Lemma B.1, $c_{z_0}(q_1, q_2)$ can be represented as

$$c_{z_0}(q_1, q_2) = V \cap \partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M) \cap \mathcal{B}_{\mathbb{R}^D}(p_0, \tau_0). \tag{B.1}$$

The proof will hence be complete as soon as we have showed the equality

$$A_\infty \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M) = V \cap \partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M) \cap \mathcal{B}_{\mathbb{R}^D}(p_0, \tau_0),$$

which we tackle by showing the two inclusions.

- *(Direct inclusion)* Let $x \in \mathbb{R}^D \setminus \mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M)$. Since $\|z_0 - q_1\| = \|z_0 - q_2\| = \tau_M$, their exists p_i satisfying $\|p_i - z_0\| < \frac{\|z_0-x\|-\tau_M}{2}$. Then,

$$\|p_i - x\| \geq \|z_0 - x\| - \|p_i - z_0\| \geq \frac{\|z_0 - x\| + \tau_M}{2} > \tau_M > \|p_i - q_1\|,$$

so that $x \notin B_i = \mathcal{B}_{\mathbb{R}^D}(p_i, \|p_i - q_1\|)$, and $x \notin A_\infty = \bigcap_{i=1}^\infty B_i$ as well. Hence this implies that

$$(\mathbb{R}^D \setminus \mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M)) \cap A_\infty = \emptyset. \tag{B.2}$$

Let now $x \in (\partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M)) \setminus V$. Since x, q_1, and q_2 are not colinear, we can find $p' \in V_x = x + span\{q_1 - x, q_2 - x\}$ such that $\|p' - q_1\| = \|p' - q_2\| = \tau_M$ and the line segment between p' and x intersects the line L passing by q_1 and q_2. Then q_1, q_2, x, and p' are lying on a 2-dimensional plane V_x, and $x \notin L$. Also, $x \notin V$, $q_1, q_2 \in V$, and $\|p' - q_1\| = \|p' - q_2\| = \tau_M > \frac{\|q_1-q_2\|}{2}$ implies that $p' \notin V$, and hence $p' \neq z_0$. Hence from Lemma B.2,

$$\|p' - x\| > \|z_0 - x\| = \tau_M.$$

Now, since $\|p' - q_1\| = \|p' - q_2\| = \tau_M$, there exists $p_{i'}$ be satisfying $\|p_{i'} - p'\| < \frac{\|p'-x\|-\tau_M}{2}$. Then

$$\|p_{i'} - x\| \geq \|p' - x\| - \|p_{i'} - p'\| \geq \frac{\|p' - x\| + \tau_M}{2} > \tau_M > \|p_{i'} - q_1\|,$$

and hence $x \notin B_{i'} = \mathcal{B}_{\mathbb{R}^D}(p_{i'}, \|p_{i'} - q_1\|)$, $x \notin A_\infty = \bigcap_{i=1}^\infty B_i$ as well. Hence this implies that

$$((\partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M)) \setminus V) \cap A_\infty = \emptyset. \tag{B.3}$$

Finally, by construction, $A_\infty \subset \mathcal{B}_{\mathbb{R}^D}(p_0, \tau_0)$. Combining this last inclusion with (B.2) and (B.3) yields the desired inclusion

$$A_\infty \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M) \subset V \cap \partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M) \cap \mathcal{B}_{\mathbb{R}^D}(p_0, \tau_0). \tag{B.4}$$

- *(Reverse inclusion)* Let $x \in V \cap \partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M) \cap \mathcal{B}_{\mathbb{R}^D}(p_0, \tau_0)$, and fix $B_i = \mathcal{B}_{\mathbb{R}^D}(p_i, \|p_i - q_1\|)$. Let $z_0' \in V$ be such that $\|z_0' - q_1\| = \|z_0' - q_2\| = \|p_i - q_1\|$ and the line segment between z_0' and x intersects the line passing q_1 and q_2. Then $q_1, q_2, x, z_0' \in V$, and x is not lying on the line passing q_1 and q_2. Hence from Lemma B.2,

$$\|p_i - x\| \leq \|z_0' - x\|. \tag{B.5}$$

Since $x \in c_{z_0}(q_1, q_2)$ and $\|z_0' - q_1\| = \|z_0' - q_2\| < \tau_M = \|z_0 - q_1\| = \|z_0 - q_2\|$, Lemma B.1 yields

$$\|z_0' - x\| \leq \|z_0' - q_1\|. \tag{B.6}$$

Hence (B.5) and (B.6) gives the upper bound on $\|p_i - x\|$ as

$$\|p_i - x\| \leq \|z_0' - x\| \leq \|z_0' - q_1\| = \|p_i - q_1\|.$$

Hence $x \in B_i$, and since choice of x and B_i were arbitrary, $V \cap \partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M) \cap \mathcal{B}_{\mathbb{R}^D}(p_0, \tau_0) \subset A_\infty$. But $\partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M) \cap \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M) = \emptyset$, so that we get the desired inclusion

$$V \cap \partial\mathcal{B}_{\mathbb{R}^D}(z_0, \tau_M) \cap \mathcal{B}_{\mathbb{R}^D}(p_0, \tau_0) \subset A_\infty \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M). \tag{B.7}$$

Putting together (B.1), (B.4), and (B.7) we get

$$A_\infty \setminus \mathring{\mathcal{B}}_{\mathbb{R}^D}(z_0, \tau_M) = c_{z_0}(q_1, q_2).$$

□

Lemma B.4. *Let $M \subset \mathbb{R}^D$ be a compact submanifold with reach $\tau_M > 0$. If there exist $p \neq q \in M$ such that $\tau_M = \frac{\|q-p\|^2}{2d(q-p, T_pM)}$, then there exists $z_0 \in Med(M)$ with $d(z_0, M) = \tau_M$.*

Proof of Lemma B.4. Write $z_0 := p + \tau_M \frac{\pi_{T_pM^\perp}(q-p)}{\left\|\pi_{T_pM^\perp}(q-p)\right\|}$. Clearly, $\|z_0 - p\| = \tau_M$, and $z_0 - p \in T_pM^\perp$, so [22, Theorem 4.8 (12)] implies that for all $\lambda \in (0,1)$, $\pi_M(p + \lambda(z_0 - p)) = p$, and hence $d(p + \lambda(z_0 - p), M) = \|\lambda(z_0 - p)\| = \lambda\tau_M$. Sending $\lambda \to 1$ yields that $d(z_0, M) = \tau_M$. Let us show that $\|z_0 - q\| = \tau_M$, which will imply that $\|z_0 - p\| = \|z_0 - q\| = d(z_0, M) = \tau_M$, and hence that $z_0 \in Med(M)$, which will conclude the proof.

Let $z_1 := p + \pi_{T_pM^\perp}(q - p)$ (see Figure 6). Note that $z_0 - z_1$ and $q - z_1$ are simplified as

$$\begin{aligned} z_0 - z_1 &= \left(\frac{\tau_M}{\left\|\pi_{T_pM^\perp}(q-p)\right\|} - 1 \right) \pi_{T_pM^\perp}(q-p), \\ q - z_1 &= (q-p) - \pi_{T_pM^\perp}(q-p) = \pi_{T_pM}(q-p). \end{aligned}$$

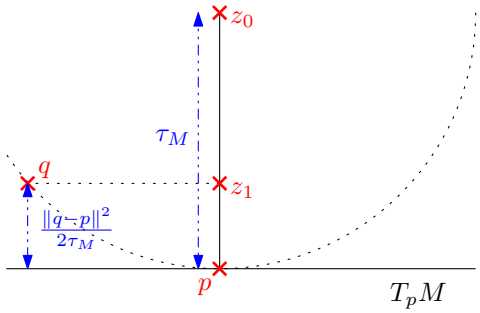


Figure 6: Layout of the proof of Lemma B.4.

In particular, $z_0 - z_1 \perp q - z_1$, which yields

$$\begin{aligned}\|z_0 - q\|^2 &= \|z_0 - z_1\|^2 + \|q - z_1\|^2 \\ &= \left(\tau_M - \left\|\pi_{T_pM^\perp}(q-p)\right\|\right)^2 + \left\|\pi_{T_pM}(q-p)\right\|^2.\end{aligned}$$

Noticing that

$$\left\|\pi_{T_pM^\perp}(q-p)\right\| = d(q-p, T_pM) = \frac{\|q-p\|^2}{2\tau_M},$$

and

$$\left\|\pi_{T_pM}(q-p)\right\|^2 = \|q-p\|^2 - \left\|\pi_{T_pM^\perp}(q-p)\right\|^2 = \|q-p\|^2\left(1 - \frac{\|q-p\|^2}{4\tau_M^2}\right),$$

we finally get

$$\begin{aligned}\|z_0 - q\|^2 &= \left(\tau_M - \frac{\|q-p\|^2}{2\tau_M}\right)^2 + \|q-p\|^2\left(1 - \frac{\|q-p\|^2}{4\tau_M^2}\right) \\ &= \tau_M^2.\end{aligned}$$

□

Lemma B.5. *Let $M \subset \mathbb{R}^D$ be a closed submanifold with reach $\tau_M > 0$. Then for all $p, q \in M$ with $t_0 := d_M(p,q) \le \tau_M/2$,*

$$\left\|\gamma''_{p\to q}(0)\right\| \le \frac{2d(q-p, T_pM)}{\|q-p\|^2} + \frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t (\gamma''_{p\to q}(s) - \gamma''_{p\to q}(0))dsdt\right\|,$$

and

$$\left\|\gamma''_{p\to q}(0)\right\| \ge \frac{2d(q-p, T_pM)}{\|q-p\|^2} - \frac{3\|q-p\|}{\tau_M^2} - \frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t (\gamma''_{p\to q}(s) - \gamma''_{p\to q}(0))dsdt\right\|.$$

In particular, when M is $\mathcal{C}^2$,

$$\sup_{\substack{p\in M\\ v\in T_pM,\|v\|=1}} \left\|\gamma''_{p,v}(0)\right\| = \limsup_{\substack{q\to p\\ q\in M}} \frac{2d(q-p,T_pM)}{\|q-p\|^2}.$$

To prove Lemma B.5 we need the following straightforward result.

Lemma B.6. *Let U be a linear space and $u\in U$, $n\in U^\perp$. If $v=u+n+e$, then*

$$|d(v,U)-\|n\|| \le \|e\|.$$

Proof of Lemma B.5. First note that from Proposition A.1 (ii), $d_M(x,y)<\pi\tau_M$ ensures the existence and uniqueness of the geodesic γ. For short, let us write $\gamma=\gamma_{p\to q}$.

The Taylor expansion of γ at order two yields

$$\begin{aligned} q-p=\gamma(t_0)-\gamma(0) &= t_0\gamma'(0)+\int_0^{t_0}\int_0^t \gamma''(s)dsdt \\ &= t_0\gamma'(0)+\frac{t_0^2}{2}\gamma''(0)+\int_0^{t_0}\int_0^t(\gamma''(s)-\gamma''(0))dsdt. \end{aligned} \tag{B.8}$$

Since $\gamma'(0)\in T_pM$ and $\gamma''(0)\in T_pM^\perp$, Lemma B.6 shows that

$$\left|\frac{2d(q-p,T_pM)}{t_0^2}-\|\gamma''(0)\|\right| \le \frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t(\gamma''(s)-\gamma''(0))dsdt\right\|.$$

Now, from $t_0=d_M(p,q)\ge\|q-p\|$, we derive the upper bound

$$\begin{aligned} \|\gamma''(0)\| &\le \frac{2d(q-p,T_pM)}{d_M(p,q)^2}+\frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t(\gamma''(s)-\gamma''(0))dsdt\right\| \\ &\le \frac{2d(q-p,T_pM)}{\|q-p\|^2}+\frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t(\gamma''(s)-\gamma''(0))dsdt\right\|. \end{aligned}$$

For the lower bound, we apply [33, Proposition 6.3] to get

$$\begin{aligned} d_M(p,q)^2 &\le \tau_M^2\left(1-\sqrt{1-\frac{2\|q-p\|}{\tau_M}}\right)^2 \\ &\le \tau_M^2\frac{\left(\frac{\|q-p\|}{\tau_M}\right)^2}{\left(1-\frac{2\|q-p\|}{\tau_M}\right)^{3/2}} \le \frac{\|q-p\|^2}{1-3\frac{\|q-p\|}{\tau_M}}, \end{aligned}$$

or equivalently,

$$\frac{1}{\|q-p\|^2}-\frac{1}{d_M(p,q)^2}\le\frac{3}{\tau_M\|q-p\|}.$$

As $d(q-p, T_pM) \leq \frac{\|q-p\|^2}{2\tau_M}$ (2.3), we finally derive

$$\begin{aligned}\|\gamma''(0)\| &\geq \frac{2d(q-p,T_pM)}{d_M(p,q)^2} - \frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t (\gamma''(s)-\gamma''(0))dsdt\right\| \\ &= \frac{2d(q-p,T_pM)}{\|q-p\|^2} - 2d(q-p,T_pM)\left(\frac{1}{\|q-p\|^2} - \frac{1}{d_M(p,q)^2}\right) \\ &\quad - \frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t (\gamma''(s)-\gamma''(0))dsdt\right\| \\ &\geq \frac{2d(q-p,T_pM)}{\|q-p\|^2} - \frac{3\|q-p\|}{\tau_M^2} - \frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t (\gamma''(s)-\gamma''(0))dsdt\right\|.\end{aligned}$$

□

Proof of Lemma 3.3. 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 $\varphi(p,q) = 2d(q-p,T_pM)/\|q-p\|^2$. By assumption, $d(z,M) > \tau_M$ for all $z \in Med(M)$. From Lemma B.4, this implies that for all $p \neq q \in M$, $\varphi(p,q) < \tau_M^{-1}$. By compactness of $M^2 \setminus \Delta_r$, this yields $\sup_{M^2\setminus\Delta_r}\varphi < \tau_M^{-1}$. Hence, from the decomposition of (2.3) as

$$\frac{1}{\tau_M} = \sup_{(p,q)\in M^2\setminus\bar{\Delta}} \varphi(p,q) = \max\left\{\sup_{(p,q)\in M^2\setminus\Delta_r}\varphi(p,q), \sup_{(p,q)\in\Delta_r\setminus\bar{\Delta}}\varphi(p,q)\right\},$$

we get $\sup_{\Delta_r\setminus\bar{\Delta}}\varphi = \tau_M^{-1}$. By letting $r>0$ go to zero and applying Lemma B.5, this yields

$$\sup_{\substack{p\in M \\ v\in T_pM, \|v\|=1}} \|\gamma''_{p,v}(0)\| = \lim_{r\to 0}\sup_{(p,q)\in\Delta_r\setminus\bar{\Delta}}\varphi(p,q) = \frac{1}{\tau_M}.$$

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 $\gamma_0 = \gamma_{q_0,v_0}$ attains the supremum, i.e. $\|\gamma_0''(0)\| = \tau_M^{-1}$, which concludes the proof. □

Appendix C: Analysis of the Estimator

C.1. Global Case

Proof of Proposition 4.2. The two left hand inequalities are direct consequences of Corollary 4.1, let us then focus on the third one.

We set t to be equal to $\max\{d_M(q_1,x), d_M(q_2,y)\}$, and $z_1 := x + (q_2-q_1)$. We have $\|z_1 - x\| = \|q_2-q_1\| = 2\tau_M$ and $\|y-q_2\|, \|q_1-x\| \leq t$. Therefore, from the definition of $\hat{\tau}$ in (4.1) and the fact that the distance function to a

linear space is 1-Lipschitz, we get

$$\begin{aligned}\frac{1}{\hat{\tau}(\{x,y\})} &\geq \frac{2d(y-x,T_xM)}{\|y-x\|^2} \\ &= \frac{2d\left((y-q_2)+(z_1-x)+(q_1-x),T_xM\right)}{\|(y-q_2)+(z_1-x)+(q_1-x)\|^2} \\ &\geq \frac{d(z_1-x,T_xM)-2t}{2(\tau_M+t)^2}.\end{aligned}$$

Since $q_1, q_2 \in \mathcal{B}(z_0,\tau_M)$ and $\|q_1-q_2\| = 2\tau_M$, $z_1 - x = q_2 - q_1 \in T_{q_1}M^\perp$. Furthermore, from [11, Lemma 11], $\sin\angle(T_xM, T_{q_1}M) \leq t/\tau_M$ and hence

$$\begin{aligned}d(z_1-x,T_xM) &\geq d(z_1-x,T_{q_1}M) - \|z_1-x\|\sin\angle(T_xM,T_{q_1}M) \\ &\geq d(q_2-q_1,T_{q_1}M) - \|q_2-q_1\|\frac{t}{\tau_M} \\ &= 2\tau_M\left(1-\frac{t}{\tau_M}\right).\end{aligned}$$

Combining the two previous bounds finally yields the announced result

$$\begin{aligned}\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x,y\})} &\leq \frac{1}{\tau_M} - \frac{d(z_1-x,T_xM)-2t}{2(\tau_M+t)^2} \\ &\leq \frac{1}{\tau_M}\left(1-\frac{1-2t/\tau_M}{(1+t/\tau_M)^2}\right) \\ &\leq \frac{4}{\tau_M^2}t,\end{aligned}$$

where the last inequality follows from the concavity of $[0,1] \ni u \mapsto 1 - \frac{1-2u}{(1+u)^2}$. □

Proof of Proposition 4.3. Let Q be the distribution on $\mathbb{R}^D$ associated to P. Let $s < \frac{1}{\tau_M}$ and $t = \frac{\tau_M^2}{4}s \leq \tau_M/4$. Write $\omega_d := \mathcal{H}^d(\mathcal{B}_{\mathbb{R}^d}(0,1))$ for the volume of the d-dimensional unit ball. From Proposition A.1 (v), for all $q \in M$,

$$\begin{aligned}Q\left(\mathcal{B}_M(q,t)\right) &\geq f_{min}\mathcal{H}^d\left(\mathcal{B}_M(q,t)\right) \\ &\geq \omega_d f_{min}\left(1-\left(\frac{t}{6\tau_M}\right)^2\right)^d t^d \\ &\geq \omega_d f_{min}\left(\frac{575}{576}\right)^d t^d.\end{aligned}$$

Moreover, Proposition 4.2 asserts that $\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X}_n)}\right| > s$ implies that either

$\mathcal{B}_M(q_1,t)\cap\mathbb{X}_n=\emptyset$ or $\mathcal{B}_M(q_2,t)\cap\mathbb{X}_n=\emptyset$. Hence,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\tau_M}-\frac{1}{\hat\tau(\mathbb{X}_n)}\right|>s\right)&\le\mathbb{P}\left(\mathcal{B}_M(q_1,t)\cap\mathbb{X}_n=\emptyset\right)+\mathbb{P}\left(\mathcal{B}_M(q_2,t)\cap\mathbb{X}_n=\emptyset\right)\\&\le 2\left(1-\omega_d f_{min}\left(\frac{575}{576}\right)^d t^d\right)^n\\&\le 2\exp\left(-n\omega_d f_{min}\left(\frac{575}{2304}\right)^d\tau_M^{2d}s^d\right).\end{aligned}$$

Integrating the above bound gives

$$\begin{aligned}\mathbb{E}_{P^n}\left[\left|\frac{1}{\tau_M}-\frac{1}{\hat\tau(\mathbb{X}_n)}\right|^p\right]&=\int_0^{\frac{1}{\tau_M}}\mathbb{P}\left(\left|\frac{1}{\tau_M}-\frac{1}{\hat\tau(\mathbb{X}_n)}\right|^p>s\right)ds\\&\le 2\int_0^\infty\exp\left(-n\omega_d f_{min}\left(\frac{575}{2304}\right)^d\tau_M^{2d}s^{\frac{d}{p}}\right)ds\\&=\frac{2\left(\frac{2304}{575}\right)^{\frac{p}{d}}}{(n\omega_d f_{min})^{\frac{p}{d}}\tau_M^{2p}}\int_0^\infty x^{\frac{p}{d}-1}e^{-x}dx\\&:=C_{\tau_M,f_{\min},d,p}n^{-\frac{p}{d}}.\end{aligned}$$

where $C_{\tau_M,f_{\min},d,p}$ depends only on τ_M, f_{min}, d, p, and is a decreasing function of τ_M when the other parameters are fixed. □

C.2. Local Case

Proof of Lemma 4.4. First note that from Proposition A.1 (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 direct consequences of Corollary 4.1. Let us then focus on the third one. We write $t_0=d_M(x,y)$ and $\gamma=\gamma_{x\to y}$ for short. By the definition (4.1) of $\hat\tau$,

$$\frac{1}{\hat\tau(\{x,y\})}\ge\frac{2d(y-x,T_xM)}{\|y-x\|^2}. \tag{C.1}$$

Furthermore, from Lemma B.5,

$$\frac{2d(y-x,T_xM)}{\|y-x\|^2}\ge\|\gamma''(0)\|-\frac{2}{t_0^2}\left\|\int_0^{t_0}\int_0^t(\gamma''(s)-\gamma''(0))dsdt\right\|. \tag{C.2}$$

But by definition of $\mathcal{M}^{d,D}_{\tau_{\min},L}\ni M$ (Definition 2.4), the geodesic γ satisfies $\|\gamma''(s)-\gamma''(0)\|\le L|s|$, so that

$$\frac{2}{t_0^2}\int_0^{t_0}\int_0^t\|\gamma''(s)-\gamma''(0)\|\,dsdt\le\frac{2}{t_0^2}\int_0^{t_0}\int_0^t L|s|dsdt=\frac{1}{3}Lt_0. \tag{C.3}$$

Combining (C.1), (C.2) and (C.3) gives the announced inequality. □

To prove Lemma 4.5, we will use the following lemma on bilinear maps.

Lemma C.1. *Let $(V, \langle\cdot,\cdot\rangle)$ and $(W, \langle\cdot,\cdot\rangle)$ be Hilbert spaces. Let $B : V \times V \to W$ be a continuous bilinear map, and write*

$$\lambda_{\max} := \sup_{\substack{v\in V\\ \|v\|=1}} \|B(v,v)\|.$$

Then for all unit vectors $v, w \in V$,

(i) $\left\|B(w,w) - 2\langle v,w\rangle^2 B(v,v)\right\| \le (3 - 2\langle v,w\rangle^2)\lambda_{\max}$.

(ii) If $v \in V$ satisfies that for all $\tilde{v} \perp v$, $\langle B(v,v), B(v,\tilde{v}) + B(\tilde{v},v)\rangle = 0$, then

$$\|B(w,w)\| \ge \langle v,w\rangle^2 \|B(v,v)\| - (1 - \langle v,w\rangle^2)\lambda_{\max}.$$

In particular, this holds whenever $\|v\| = 1$ with $\|B(v,v)\| = \lambda_{\max}$.

Proof of Lemma C.1. Let $\theta = \arccos(\langle v,w\rangle) \in [0,\pi]$, and write $w = \cos\theta v + \sin\theta v^\perp$ for some unit vector $v^\perp \in V$ with $v^\perp \perp v$. Then $B(w,w)$ can be expanded as

$$B(w,w) = \cos^2\theta B(v,v) + \cos\theta\sin\theta(B(v,v^\perp) + B(v^\perp,v)) + \sin^2\theta B(v^\perp,v^\perp). \tag{C.4}$$

(i) Consider $\bar{w} := -\cos\theta v + \sin\theta v^\perp \in V$. Then $\bar{w}$ is a unit vector, and $B(\bar{w},\bar{w})$ can be similarly expanded as

$$B(\bar{w},\bar{w}) = \cos^2\theta B(v,v) - \cos\theta\sin\theta(B(v,v^\perp) + B(v^\perp,v)) + \sin^2\theta B(v^\perp,v^\perp), \tag{C.5}$$

and hence summing up (C.4) and (C.5) gives

$$B(w,w) + B(\bar{w},\bar{w}) = 2\cos^2\theta B(v,v) + 2\sin^2\theta B(v^\perp,v^\perp).$$

As $\left\|B(v^\perp,v^\perp)\right\|$ and $\|B(\bar{w},\bar{w})\|$ are upper bounded by $\lambda_{\max}$, this yields

$$\begin{aligned}\left\|B(w,w) - 2\cos^2\theta B(v,v)\right\| &= \left\|2\sin^2\theta B(v^\perp,v^\perp) - B(\bar{w},\bar{w})\right\| \\ &\le (1 + 2\sin^2\theta)\lambda_{\max} \\ &= (3 - 2\cos^2\theta)\lambda_{\max},\end{aligned}$$

which is the announced bound.

(ii) From (C.4), $\|B(w,w)\|$ can be lower bounded as

$$\begin{aligned}&\|B(w,w)\| \\ &\ge \left\|\cos^2\theta B(v,v) + \cos\theta\sin\theta(B(v,v^\perp) + B(v^\perp,v))\right\| - \sin^2\theta\left\|B(v^\perp,v^\perp)\right\|.\end{aligned} \tag{C.6}$$

But since $\left\langle B(v,v), B(v,v^\perp) + B(v^\perp,v)\right\rangle = 0$, Pythagoras's theorem yields

$$\begin{aligned}&\left\|\cos^2\theta B(v,v) + \cos\theta\sin\theta(B(v,v^\perp) + B(v^\perp,v))\right\| \\ &= \sqrt{\cos^4\theta\left\|(B(v,v)\right\|^2 + \cos^2\theta\sin^2\theta\left\|(B(v,v^\perp) + B(v^\perp,v)\right\|^2} \\ &\ge \cos^2\theta\left\|(B(v,v)\right\|.\end{aligned}$$

Applying this and $\|B(v^\perp, v^\perp)\| \leq \lambda_{\max}$ to (C.6) gives the final bound

$$\|B(w,w)\| \geq \cos^2\theta\, \|B(v,v)\| - \sin^2\theta \lambda_{\max}.$$

We now show the last claim, namely that $\|v\| = 1$ and $\|B(v,v)\| = \lambda_{\max}$ are sufficient conditions for

$$\langle B(v,v), B(v,\tilde{v}) + B(\tilde{v},v)\rangle = 0 \text{ for all } \tilde{v} \perp v. \tag{C.7}$$

For this aim, we take such a $v \in V$ and we consider $h : V \to \mathbb{R}$ defined by $h(u) = \|B(u,u)\|^2$ and $g : V \to \mathbb{R}$ defined by $g(u) = \|u\|^2 - 1$. Then v is a solution of the optimization problem:

$$\begin{aligned} &\text{maximize } h(u) \\ &\text{subject to } g(u) = 0. \end{aligned}$$

Since h and g are continuously differentiable, the Lagrange multiplier theorem asserts that their Fréchet derivatives at v satisfy $\ker d_v g \subset \ker d_v h$. As $d_v h(u) = 2\langle B(v,v), B(v,u) + B(u,v)\rangle$ and $d_v g(u) = 2\langle v, u\rangle$, this rewrites exactly as the claim (C.7).

□

Corollary C.2. *Let $M \subset \mathbb{R}^D$ be a $\mathcal{C}^2$-submanifold and $p \in M$. Let $v_0, v_1 \in T_pM$ be unit tangent vectors, 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 $\kappa_p = \max_{v \in \mathcal{B}_{T_pM}(0,1)} \|\gamma''_{p,v}(0)\|$. Then,*

(i) $\|\gamma_1''(0)\| \geq 2\|\gamma_0''(0)\|\cos^2\theta - \kappa_p(1 + 2\sin^2\theta)$.

(ii) If v_0 is a direction of maximum directional curvature, i.e. $\|\gamma_0''(0)\| = \kappa_p$, then $\|\gamma_1''(0)\| \geq \|\gamma_0''(0)\| - 2\kappa_p \sin^2\theta$.

Proof of Corollary C.2. Consider the symmetric bilinear map $B : T_pM \times T_pM \to T_pM^\perp$ given by the hessian of the exponential map $B(v,w) := d_0^2 \exp_p(v,w)$. In particular, for all $v \in T_pM$, $\gamma''_{p,v}(0) = B(v,v)$ and $\sup_{v \in V, \|v\|=1} \|B(v,v)\| = \kappa_p$. This allows us to tackle the two points of the result.

(i) Applying Lemma C.1 (i) to B with $v = v_0$ and $w = v_1$ yields

$$\begin{aligned} \|\gamma_1''(0)\| &\geq \|2\cos^2\theta \gamma_0''(0)\| - \|2\cos^2\theta\gamma_0''(0) - \gamma_1''(0)\| \\ &\geq 2\|\gamma_0''(0)\|\cos^2\theta - \kappa_p(1 + 2\sin^2\theta) \end{aligned}$$

(ii) Since v_0 gives the maximal directional curvature, applying Lemma C.1 (ii) to B, $v = v_0$ and $w = v_1$ precisely yields $\|\gamma_1''(0)\| \geq \|\gamma_0''(0)\| - 2\kappa_p \sin^2\theta$.

□

For a triangle in a Euclidean space, the sum of any two angles is upper bounded by π. The same property holds for a geodesic triangle on a manifold if its side lengths are not too large compared to its reach, which is formalized in the following Lemma C.3.

Lemma C.3. *Let $M \subset \mathbb{R}^D$ be a closed submanifold with reach $\tau_M > 0$, and $x, y, z \in M$ be three distinct points. Consider the geodesic triangle with vertices x, y, z, that is, the triangle formed by $\gamma_{x\to y}$, $\gamma_{y\to z}$, $\gamma_{z\to x}$.*

If at least two of the side lengths of the triangle are strictly less than $\frac{\pi\tau_M}{2}$, then the sum of any two of its angles is less than or equal to π.

Proof of Lemma C.3. Without loss of generality, suppose $d_M(x,y)$ is the longest side length: $d_M(y,z), d_M(z,x) \le d_M(x,y)$. Then $d_M(y,z), d_M(z,x) \in \big(0, \frac{\pi\tau_M}{2}\big)$, so that $d_M(x,y) \in (0, \pi\tau_M)$ by triangle inequality.

Let $\mathcal{S}^2_{\tau_M}$ be a d-dimensional sphere of radius τ_M. In what follows, for short, $\angle abc$ stands for $\angle(\gamma'_{b\to a}(0), \gamma'_{b\to c}(0))$. Let $\bar{x}, \bar{y}, \bar{z} \in \mathcal{S}^2_{\tau_M}$ be such that $d_{\mathcal{S}^2_{\tau_M}}(\bar{x}, \bar{y}) = d_M(x,y)$, $d_{\mathcal{S}^2_{\tau_M}}(\bar{y}, \bar{z}) = d_M(y,z)$, and $d_{\mathcal{S}^2_{\tau_M}}(\bar{z}, \bar{x}) = d_M(z,x)$. From Proposition A.1 (ii) and the fact that $d_M(x,y) + d_M(y,z) + d_M(z,x) < 2\pi\tau_M$, Toponogov's comparison theorem [30, Section 4] yields $\angle xyz \le \angle \bar{x}\bar{y}\bar{z}$, $\angle yzx \le \angle \bar{y}\bar{z}\bar{x}$, and $\angle xzy \le \angle \bar{x}\bar{z}\bar{y}$. Furthermore, the spherical law of cosines [9, Proposition 18.6.8] together with $d_M(y,z), d_M(z,x) \in \big(0, \frac{\pi\tau_M}{2}\big)$, $d_M(x,y) \in (0, \pi\tau_M)$, and the fact that $\cos(\cdot)$ is decreasing on $[0,\pi]$ imply

$$\cos\left(\angle \bar{z}\bar{x}\bar{y}\right) = \frac{\cos\left(\frac{d_M(y,z)}{\tau_M}\right) - \cos\left(\frac{d_M(z,x)}{\tau_M}\right)\cos\left(\frac{d_M(x,y)}{\tau_M}\right)}{\sin\left(\frac{d_M(z,x)}{\tau_M}\right)\sin\left(\frac{d_M(x,y)}{\tau_M}\right)} \ge 0,$$

so that $\angle zxy \le \angle \bar{z}\bar{x}\bar{y} \le \frac{\pi}{2}$. Symmetrically, we also have $\angle xyz \le \frac{\pi}{2}$.

If $\angle yzx \le \frac{\pi}{2}$ also holds, then the final result is trivial, so from now on we will assume that $\angle yzx \ge \frac{\pi}{2}$.

Thus, $\sin(\angle \bar{y}\bar{z}\bar{x}) \le \sin(\angle yzx)$, so applying the spherical law of sines and cosines [9, Proposition 18.6.8], $d_M(y,z), d_M(z,x) \in \big(0, \frac{\pi\tau_M}{2}\big)$, and $\angle \bar{y}\bar{z}\bar{x} \in \big[\frac{\pi}{2}, \pi\big]$ yield

$$\begin{aligned}
&\sin(\angle zxy) \le \sin(\angle \bar{z}\bar{x}\bar{y}) \\
&= \frac{\sin\left(\frac{d_M(y,z)}{\tau_M}\right)\sin(\angle \bar{y}\bar{z}\bar{x})}{\sqrt{1 - \left(\cos\left(\frac{d_M(z,x)}{\tau_M}\right)\cos\left(\frac{d_M(y,z)}{\tau_M}\right) + \sin\left(\frac{d_M(z,x)}{\tau_M}\right)\sin\left(\frac{d_M(y,z)}{\tau_M}\right)\cos(\angle \bar{y}\bar{z}\bar{x})\right)^2}} \\
&\le \frac{\sin\left(\frac{d_M(y,z)}{\tau_M}\right)\sin(\angle \bar{y}\bar{z}\bar{x})}{\sqrt{1 - \cos^2\left(\frac{d_M(z,x)}{\tau_M}\right)\cos^2\left(\frac{d_M(y,z)}{\tau_M}\right)}} \le \sin(\angle \bar{y}\bar{z}\bar{x}) \le \sin(\angle yzx).
\end{aligned}$$

This last bound together with $\angle zxy \le \frac{\pi}{2} \le \angle yzx$ yield $\angle zxy + \angle yzx \le \pi$. Symmetrically, we also have $\angle xxz + \angle xzy \le \pi$. Hence, the sum of any two angles is less than or equal to π.

□

We are now in position to prove Lemma 4.5.

Proof of Lemma 4.5. 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 7). From Corollary C.2 (i),

$$\left\|\gamma''_{x\to y}(0)\right\| \geq (2-2\sin^2\theta)\left\|\gamma''_{q_0\to x}(t_x)\right\| - (1+2\sin^2\theta)\kappa_x. \tag{C.8}$$

We now focus on the term $\left\|\gamma''_{q_0\to x}(t_x)\right\|$. Since the direction $\gamma'_0(0)$ maximizes the directional curvature at q_0 and $\theta_x = \angle(\gamma'_0(0), \gamma'_{q_0\to x}(0))$, Corollary C.2 (ii) 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,

$$\begin{aligned}\left\|\gamma''_{q_0\to x}(t_x)\right\| &\geq \left\|\gamma''_{q_0\to x}(0)\right\| - \left\|\gamma''_{q_0\to x}(t_x) - \gamma''_{q_0\to x}(0)\right\| \\ &\geq (1-2\sin^2\theta_x)\kappa_{q_0} - Lt_x. \end{aligned}\tag{C.9}$$

Now, consider the geodesic triangle with vertices x, y, q_0, that is, the triangle

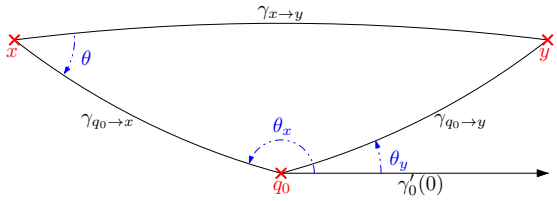


Figure 7: Layout of Lemma 4.5.

formed by $\gamma_{x\to y}$, $\gamma_{q_0\to x}$, $\gamma_{q_0\to y}$, as in Figure 7. Then Lemma C.3 implies that $\theta + |\theta_x - \theta_y| \leq \pi$. Combined with the assumption that $|\theta_x - \theta_y| \geq \frac{\pi}{2}$, this yields

$$\sin\theta \leq \sin(|\theta_x - \theta_y|). \tag{C.10}$$

Putting together (C.8), (C.9) and (C.10) gives the final bound

$$\begin{aligned}&\left\|\gamma''_{x\to y}(0)\right\| \\ &\geq 2(1-\sin^2(|\theta_x-\theta_y|))\left((1-2\sin^2\theta_x)\kappa_{q_0} - Lt_x\right) - (1+2\sin^2(|\theta_x-\theta_y|))\kappa_x \\ &= 2\kappa_{q_0} - \kappa_x(1+2\sin^2(|\theta_x-\theta_y|)) - 2Lt_x\cos^2(|\theta_x-\theta_y|) \\ &\quad - 2\kappa_{q_0}\left(\sin^2(|\theta_x-\theta_y|)) + 2\sin^2\theta_x - \sin^2\theta_x\sin^2(|\theta_x-\theta_y|)\right) \\ &\geq \kappa_{q_0} - (\kappa_x - \kappa_{q_0}) - 2Lt_x - (2\kappa_x + 6\kappa_{q_0})\sin^2(|\theta_x-\theta_y|).\end{aligned}$$

□

Proof of Proposition 4.7. In what follows, we let $t_0 \leq \frac{\tau_{\min}}{10}$,

$$B_1 := \exp_{q_0}\left(\left\{v \in T_{q_0}M : \|v\| \leq t_0,\ \angle(\gamma'_0(0), v) \leq \sqrt{\frac{t_0}{\tau_{\min}}}\right\}\right),$$

$$B_2 := \exp_{q_0}\left(\left\{v \in T_{q_0}M : \|v\| \leq t_0,\ \angle(\gamma'_0(0), v) \geq \pi - \sqrt{\frac{t_0}{\tau_{\min}}}\right\}\right),$$

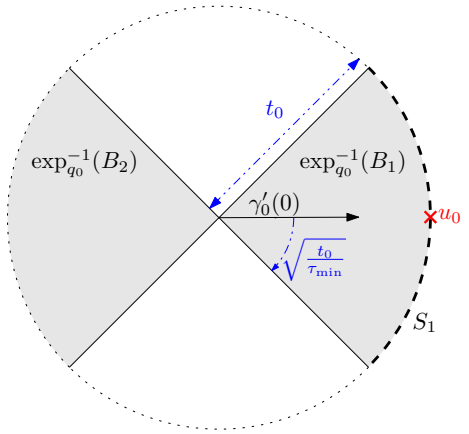


Figure 8: Layout of the proof of Proposition 4.7.

and $B_0 := B_1 \cup B_2$ (see Figure 8). Let $\mathbb{X} \subset M$, and $x, y \in \mathbb{X}$ be such that $x \in B_1$, $y \in B_2$. Writing $\theta_x := \angle(\gamma_0'(0), \gamma'_{q_0 \to x}(0))$ and $\theta_y := \angle(\gamma_0'(0), \gamma'_{q_0 \to y}(0))$, then $\theta_x \leq \sqrt{\frac{t_0}{\tau_{\min}}} \leq \frac{\pi}{4}$ and $\theta_y \geq \pi - \sqrt{\frac{t_0}{\tau_{\min}}} \geq \frac{3\pi}{4}$. Also, $d_M(q_0, x) \leq t_0$ and $d_M(x, y) \leq 2t_0$, so Proposition 4.6 rewrites as

$$\begin{aligned} 0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} &\leq \frac{8\sin^2(|\theta_x - \theta_y|)}{\tau_M} + L\left(\frac{1}{3} d_M(x,y) + 2 d_M(q_0, x)\right) \\ &\leq \left(\frac{16}{\tau_{\min}\tau_M} + \frac{8L}{3}\right) t_0. \end{aligned}$$

A symmetric argument also applies when $x \in B_2$ and $y \in B_1$. Now, for any $s < \frac{1}{\tau_M}$, let $t_0(s) := \left(\frac{16}{\tau_{\min}^2} + \frac{8L}{3}\right)^{-1} s < \frac{\tau_{\min}}{10}$. 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,

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X}_n)}\right| > s\right) \\ &\leq \sum_{m=0}^{n} \binom{n}{m} \Big\{ \mathbb{P}\left(X_1, \ldots, X_m \in M \backslash B_0, X_{m+1}, \ldots, X_n \in B_1\right) \\ &\qquad\qquad + \mathbb{P}\left(X_1, \ldots, X_m \in M \backslash B_0, X_{m+1}, \ldots, X_n \in B_2\right) \Big\} \\ &= \sum_{m=0}^{n} \binom{n}{m} \left\{ (1 - Q(B_0))^m Q(B_1)^{n-m} + (1 - Q(B_0))^m Q(B_2)^{n-m} \right\} \\ &\leq (1 - Q(B_2))^n + (1 - Q(B_1))^n. \end{aligned} \tag{C.11}$$

We now derive lower bounds for $Q(B_1)$ and $Q(B_2)$. For this purpose, let $S_1 := \exp_{q_0}^{-1}(B_1) \cap \partial\mathcal{B}_{T_{q_0}M}(0, t_0)$ (see Figure 8). Then $\exp_{q_0}^{-1}(B_1) \subset \mathcal{B}_{T_{q_0}M}(0, t_0)$ is a

cone satisfying

$$\frac{\mathcal{H}^d\left(\exp_{q_0}^{-1}(B_1)\right)}{\mathcal{H}^d\left(\mathcal{B}_{T_{q_0}M}(0,t_0)\right)} = \frac{\mathcal{H}^{d-1}\left(S_1\right)}{\mathcal{H}^{d-1}\left(\partial\mathcal{B}_{T_{q_0}M}(0,t_0)\right)}.$$

Let $\omega_d := \mathcal{H}^d(\mathcal{B}_{\mathbb{R}^d}(0,1))$ and $\sigma_d := \mathcal{H}^d(\partial\mathcal{B}_{\mathbb{R}^{d+1}}(0,1))$ denote the volumes of the d-dimensional unit ball and sphere respectively, so that $\mathcal{H}^d\left(\mathcal{B}_{T_{q_0}M}(0,t_0)\right) = \omega_d t_0^d$ and $\mathcal{H}^{d-1}\left(\partial\mathcal{B}_{T_{q_0}M}(0,t_0)\right) = \sigma_{d-1}t_0^{d-1}$. In view of deriving a lower bound on $\mathcal{H}^{d-1}\left(S_1\right)$, consider $u_0 := t_0\gamma_0'(0) \in S_1$. Since $\tau_{S_1} = t_0$ and $\exp_{u_0}^{-1}(S_1) \subset \mathcal{B}_{T_{u_0}S_1}\left(0, \tau_{\min}^{-\frac{1}{2}}t_0^{\frac{3}{2}}\right)$, applying Proposition A.1 (v) yields

$$\begin{aligned}\mathcal{H}^{d-1}\left(S_1\right) &\geq \left(1 - \frac{t_0}{6\tau_{\min}}\right)^{d-1} \mathcal{H}^{d-1}\left(\mathcal{B}_{T_{u_0}S_1}\left(0, \tau_{\min}^{-\frac{1}{2}}t_0^{\frac{3}{2}}\right)\right)\\ &\geq \left(\frac{59}{60}\right)^{d-1} \omega_{d-1}\tau_{\min}^{-\frac{d-1}{2}}t_0^{\frac{3d-3}{2}},\end{aligned}$$

and hence

$$\begin{aligned}\mathcal{H}^{d-1}\left(\exp_{q_0}^{-1}(B_1)\right) &= \frac{\mathcal{H}^d\left(\mathcal{B}_{T_{q_0}M}(0,t_0)\right)\mathcal{H}^{d-1}\left(S_1\right)}{\mathcal{H}^{d-1}\left(\partial\mathcal{B}_{T_{q_0}M}(0,t_0)\right)}\\ &\geq \left(\frac{59}{60}\right)^{d-1} \frac{\omega_{d-1}}{d}\tau_{\min}^{-\frac{d-1}{2}}t_0^{\frac{3d-1}{2}}.\end{aligned}$$

Finally, since $\exp_{q_0}^{-1}(B_1) \subset \mathcal{B}_{T_{q_0}M}(q_0, \frac{\tau_M}{10})$, Proposition A.1 (v) yields

$$\mathcal{H}^d\left(B_1\right) \geq \left(\frac{599}{600}\right)^d \mathcal{H}^d\left(\exp_{q_0}^{-1}(B_1)\right) \geq \left(\frac{35341}{36000}\right)^d \frac{1}{d}\tau_{\min}^{-\frac{d-1}{2}}t_0^{\frac{3d-1}{2}},$$

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 (C.11) gives

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X}_n)}\right| > s\right) &\leq 2\left(1 - C_{\tau_{\min},d,L,f_{\min}}s^{\frac{3d-1}{2}}\right)^n\\ &\leq 2\exp\left(-C_{\tau_{\min},d,L,f_{\min}}ns^{\frac{3d-1}{2}}\right).\end{aligned}$$

As for the proof of Proposition 4.3, the result then follows by integration. □

Appendix D: Minimax Lower Bounds

D.1. Stability of the Model With Respect to Diffeomorphisms

To prove Proposition 5.4, we will use the following result stating that the reach is a stable quantity with respect to $\mathcal{C}^2$-perturbations.

Lemma D.1 (Theorem 4.19 in [22]). *Let $A \subset \mathbb{R}^D$ with $\tau_A \geq \tau_{min} > 0$ and $\Phi : \mathbb{R}^D \longrightarrow \mathbb{R}^D$ 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 5.4. 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 $\left\|d\Phi^{-1}\right\|_{op} \leq (1 - \|d\Phi - I_D\|_{op})^{-1}$-Lipschitz, and $d\Phi$ is $\left\|d^2\Phi\right\|_{op}$-Lipschitz. From Lemma D.1,

$$\tau_{M'} \geq \frac{\tau_{min}(1 - \|d\Phi - I_D\|_{op})^2}{\|d^2\Phi\|_{op}\tau_{min} + (1 + \|d\Phi - I_D\|_{op})} \geq \tau_{min}/2,$$

where we used that $\left\|d^2\Phi\right\|_{op}\tau_{min} \leq 1/2$ and $\|d\Phi - I_D\|_{op} \leq 1/10$. 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 $M \in \mathcal{M}^{d,D}_{\tau_{min},L}$ 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 $\Psi_p : \mathcal{B}_{\mathbb{R}^d}(0,\varepsilon) \to M$ 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 [20, p.62]). That is, the second order ordinary differential equation

$$\begin{cases} c''_\ell(t) + \left\langle \Gamma^\ell\left(c(t)\right) \cdot c'(t), c'(t) \right\rangle = 0, & (1 \leq \ell \leq d) \\ c(0) = 0 \text{ and } c'(0) = d_p\mathbf{x}.v, \end{cases} \tag{D.1}$$

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 (D.1) 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 (D.1): since $c(0) = \tilde{c}(0) = 0$ and $c''(0) = \tilde{c}''(0)$, we get that for $\|I_D - d\Phi\|_{op}$, $\left\|d^2\Phi\right\|_{op}$ and $\left\|d^3\Phi\right\|_{op}$ 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. □

D.2. Lemmas on the Total Variation Distance

Prior to any actual construction, we show the following straightforward lemma bounding the total variation between uniform distribution on manifolds that

are perturbations of each other. For $M \subset \mathbb{R}^D$, write $\lambda_M = \mathbb{1}_M \mathcal{H}^d / \mathcal{H}^d(M)$ for the uniform probability distribution on M.

Lemma D.2. *Let $M \subset \mathbb{R}^D$ be a compact d-dimensional submanifold and $B \subset \mathbb{R}^D$ be a Borel set. Let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ 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 D.2. Since Φ is $(1 + \|d\Phi - I_D\|_{op})$-Lipschitz, [4, Lemma 7] 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, [4, Lemma 7] 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 5.3. For this, we will need the following elementary differential geometry results Lemma D.3 and Corollary D.4.

Lemma D.3. *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 $\mathcal{H}^d\left(g^{-1}(0) \cap \mathcal{B}(x, r)\right) = 0$.*

Proof of Lemma D.3. Let us prove that for $r > 0$ small enough, the intersection $g^{-1}(0) \cap \mathcal{B}(x, r)$ 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 non-singular 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, $g_1^{-1}(0) \cap \mathcal{B}(x, r)$ has d-dimensional Hausdorff measure zero. The result hence follows, noticing that $g^{-1}(0) \subset g_1^{-1}(0)$. □

Corollary D.4. *Let $M, M' \subset \mathbb{R}^D$ 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 $A = M \cap M' \cap \mathcal{B}(x, r)$ satisfies $\lambda_M(A) = \lambda_{M'}(A) = 0$.*

Proof of Corollary D.4. Writing $k = D - d$, 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 $g : \mathcal{B}_{\mathbb{R}^d}(0, r') \to \mathbb{R}^k$ for $r' > 0$ small enough. The assumption $T_xM \neq T_xM'$ translates to $d_0 g \neq 0$, and the previous transformation maps smoothly $M \cap M' \cap \mathcal{B}(x, r'')$ to $g^{-1}(0) \cap \mathcal{B}(0, r')$ for $r'' > 0$ small enough. We conclude by applying Lemma D.3. □

We are now in position to prove Lemma 5.3.

Proof of Lemma 5.3. 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 $f, f' : \mathbb{R}^D \to \mathbb{R}_+$ 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',

$$\begin{aligned} &TV(P, P') \\ &\quad = \frac{1}{2} \int_{\mathbb{R}^D \times \mathbb{G}^{d,D}} |\bar{f} - \bar{f}'| d\nu \\ &\quad = \frac{1}{2} \int_{\mathbb{R}^D} \mathbb{1}_{T_xM = T_xM'} |f(x) - f'(x)| + \mathbb{1}_{T_xM \neq T_xM'} (f(x) + f'(x)) \mathcal{H}^d(dx). \end{aligned}$$

From Corollary D.4 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

$$TV(P, P') = \frac{1}{2} \int_{\mathbb{R}^D} |f - f'| d\mathcal{H}^d = TV(Q, Q'),$$

which concludes the proof. □

D.3. Construction of the Hypotheses

This section is devoted to the construction of hypotheses that will be used in Le Cam's lemma (Lemma 5.2), to derive Proposition 2.9 and Theorem 5.6.

Lemma D.5. *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 $M \in \mathcal{M}^{d,D}_{R, \frac{1}{R^2}}$ and a global $\mathcal{C}^\infty$-diffeomorphism $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ 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 $\mathcal{H}^d(M') \leq 2\mathcal{H}^d(M) = 2\sigma_d R^d$,

$$\left| \frac{1}{\tau_M} - \frac{1}{\tau_{M'}} \right| \geq \frac{\eta}{\ell^2}, \quad \text{and} \quad TV(\lambda_M, \lambda_{M'}) \leq 12 \left(\frac{\ell}{R}\right)^d.$$

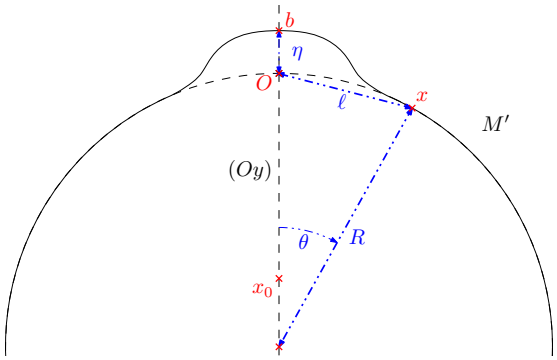


Figure 9: The bumped sphere M' of Lemma D.5.

Proof of Lemma D.5. Let $M \subset \mathbb{R}^{d+1} \times \{0\}^{D-d-1} \subset \mathbb{R}^D$ be the sphere of radius R with center $(0, -R, 0, \dots, 0)$. 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 $M \in \mathcal{M}^{d,D}_{R,\frac{1}{R^2}}$. Let $\phi : \mathbb{R}^D \to \mathbb{R}_+$ 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 $\mathcal{B}(0,1)$ and elementary real analysis yields $\phi(0) = 1$, $\|d\phi\|_{op} \leq 3$, $\left\|d^2\phi\right\|_{op} \leq 23$ and $\left\|d^3\phi\right\|_{op} \leq 573$. Let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ be defined by

$$\Phi(x) = x + \eta\phi\left(x/\ell\right) \cdot v,$$

where $v = (0, 1, 0, \dots, 0)$ is the unit vertical vector. Φ is the identity map on $\mathcal{B}\left(0,\ell\right)^c$, and in $\mathcal{B}\left(0,\ell\right)$, Φ translates points on the vertical axis with a magnitude modulated by the weight function $\phi(x/\ell)$. 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 $x \in \mathbb{R}^D$, 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 [18]. 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 $M' = \Phi\left(M\right)$ for the image of M by the map Φ (see Figure 9). 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 $x_0 = \left(0, \frac{R+\eta/2}{1+\frac{\ell^2}{2R\eta}}, 0, \dots, 0\right)$ belongs to its medial axis $Med(M')$ (see (2.1)). For this, write

$$b = (0, \eta, 0, \dots, 0), \quad b' = (0, -2R, 0, \dots, 0),$$

together with $\theta = \arccos(1 - \ell^2/(2R^2))$, and

$$x = (R\sin\theta, R\cos\theta - R, 0, \dots, 0).$$

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,

$$\begin{aligned}\frac{1}{\tau_{M'}} \geq \frac{1}{d(x_0, M')} &\geq \frac{1}{\|x_0 - x\|} \\ &\geq \frac{1}{R\left|1 - \frac{\ell^2}{2R^2} - \frac{1+\frac{\eta}{2R}}{1+\frac{\ell^2}{2R\eta}}\right|} \\ &\geq \frac{1}{R\left(1 - \frac{1+\frac{\eta}{2R}}{1+\frac{\ell^2}{2R\eta}}\right)} \geq \frac{1}{R}\left(1 + \frac{1+\frac{\eta}{2R}}{1+\frac{\ell^2}{2R\eta}}\right) \geq \frac{1}{R} + \frac{\eta}{\ell^2},\end{aligned}$$

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 $\Phi_{|\mathcal{B}(0,\ell)^c}$ coinciding with the identity map, Lemma D.2 yields $\mathcal{H}^d(M') \leq 2\mathcal{H}^d(M) = 2\sigma_d R^d$ and

$$\begin{aligned}TV(\lambda_M, \lambda_{M'}) &\leq 12\lambda_M(\mathcal{B}(0,\ell)) \\ &= 12\frac{\mathcal{H}^d\left(\mathcal{B}_{\mathcal{S}^d}\left(0, 2\arcsin\left(\frac{\ell}{2R}\right)\right)\right)}{\mathcal{H}^d(\mathcal{S}^d)} \\ &\leq 12\left(\frac{\ell}{R}\right)^d,\end{aligned}$$

which concludes the proof. □

Proof of Proposition 5.5. Apply Lemma D.5 with $R = 2\tau_{min}$. Then the sphere M of radius $2\tau_{min}$ belongs to $\mathcal{M}^{d,D}_{2\tau_{min},1/(4\tau^2_{min})}$. Furthermore, taking $\eta = c_d\ell^3/\tau^2_{min}$ for $c_d > 0$ and $\ell > 0$ small enough, Proposition 5.4 (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 $\mathcal{M}^{d,D}_{\tau_{min},1/(2\tau^2_{min})} \subset \mathcal{M}^{d,D}_{\tau_{min},L}$, 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 $\lambda_M, \lambda_{M'} \in \mathcal{Q}^{d,D}_{\tau_{min},L,f_{min}}$, 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 2.9. Let M and M' be given by Lemma D.5 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 $\left\|d^2\Phi\right\|_{op} \leq 23\eta/\ell^2 \leq 1/(2\tau_{min})$. Since $\tau_M \geq 2\tau_{min}$, Lemma D.1 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 $\mathcal{M}^{d,D}_{\tau_{min},L=\infty}$. 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 $\mathcal{Q}^{d,D}_{\tau_{min},L=\infty,f_{min}}$. Let now P, P' denote the distributions of $\mathcal{P}^{d,D}_{\tau_{min},L=\infty,f_{min}}$ associated to $\lambda_M, \lambda_{M'}$ (Definition 2.6). Lemma 5.3 asserts that $TV(P,P') = TV(\lambda_M, \lambda_{M'})$. Applying Lemma 5.2 to P, P', we get that for all $n \geq 1$, for ℓ small enough,

$$\begin{aligned}\inf_{\hat{\tau}_n}\sup_{P\in\mathcal{P}^{d,D}_{\tau_{min},L=\infty,f_{min}}} \mathbb{E}_{P^n}\left|\frac{1}{\tau_P} - \frac{1}{\hat{\tau}_n}\right|^p &\geq \frac{1}{2^p}\left|\frac{1}{\tau_M} - \frac{1}{\tau_{M'}}\right|^p (1-TV(P,P'))^n \\ &\geq \frac{1}{2^p}\left(\frac{\eta}{\ell^2}\right)^p\left(1-12\left(\frac{\ell}{2\tau_{min}}\right)^d\right)^n \\ &= \frac{1}{2^p}\left(\frac{1}{46\tau_{min}}\right)^p\left(1-12\left(\frac{\ell}{2\tau_{min}}\right)^d\right)^n.\end{aligned}$$

Sending $\ell \to 0$ with $n \geq 1$ fixed yields the announced result.

□

Appendix E: Stability with Respect to Tangent Spaces

Proof of Proposition 6.1. To get the bound on the difference of suprema, we show the (stronger) pointwise bound. Indeed, for all $x, y \in \mathbb{X}$ with $x \neq y$,

$$\begin{aligned}\left|\frac{2d(y-x,T_x)}{\|y-x\|^2} - \frac{2d(y-x,\tilde{T}_x)}{\|y-x\|^2}\right| &\leq \frac{2\|\pi_{T_x}(y-x) - \pi_{\tilde{T}_x}(y-x)\|}{\|y-x\|^2} \\ &\leq \frac{2\|\pi_{T_x} - \pi_{\tilde{T}_x}\|_{\mathrm{op}}}{\|y-x\|} \leq \frac{2\sin\theta}{\delta}.\end{aligned}$$

□

References

[1] Aamari, E. and Levrard, C. (2018). Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. *Discrete Comput. Geom.* **59** 923–971. MR3802310

[2] Aamari, E. and Levrard, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.* **47** 177–204. MR3909931

[3] Alexander, S. B. and Bishop, R. L. (2006). Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geom. Dedicata* **117** 65–84. MR2231159 (2007c:53110)
[4] Arias-Castro, E., Lerman, G. and Zhang, T. (2017). Spectral clustering based on local PCA. *J. Mach. Learn. Res.* **18** Paper No. 9, 57. MR3634876
[5] Arias-Castro, E., Pateiro-López, B. and Rodríguez-Casal, A. (2018). Minimax Estimation of the Volume of a Set Under the Rolling Ball Condition. *Journal of the American Statistical Association* **0** 1-12.
[6] Attali, D., Boissonnat, J.-D. and Edelsbrunner, H. (2009). Stability and computation of medial axes: a state-of-the-art report. In *Mathematical foundations of scientific visualization, computer graphics, and massive data exploration. Math. Vis.* 109–125. Springer, Berlin. MR2560510
[7] Balakrishnan, S., Rinaldo, A., Sheehy, D., Singh, A. and Wasserman, L. A. (2012). Minimax rates for homology inference. In *International Conference on Artificial Intelligence and Statistics* 64–72.
[8] Belkin, M., Niyogi, P. and Sindhwani, V. (2006). Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **7** 2399–2434. MR2274444
[9] Berger, M. (1987). *Geometry. II. Universitext*. Springer-Verlag, Berlin Translated from the French by M. Cole and S. Levy. MR882916
[10] Boissonnat, J.-D. and Ghosh, A. (2014). Manifold reconstruction using tangential Delaunay complexes. *Discrete Comput. Geom.* **51** 221–267. MR3148657
[11] Boissonnat, J.-D., Lieutier, A. and Wintraecken, M. (2018). The reach, metric distortion, geodesic convexity and the variation of tangent spaces. In *34th International Symposium on Computational Geometry. LIPIcs. Leibniz Int. Proc. Inform.* **99** Art. No. 10, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3824254
[12] Burago, D., Burago, Y. and Ivanov, S. (2001). *A course in metric geometry. Graduate Studies in Mathematics* **33**. American Mathematical Society, Providence, RI. MR1835418
[13] Chazal, F. and Lieutier, A. (2005). The λ-medial axis. *J. Graphical Models* **67** 304–331.
[14] Cheng, S.-W. and Chiu, M.-K. (2016). Tangent estimation from point samples. *Discrete Comput. Geom.* **56** 505–557. MR3544007
[15] Cuevas, A., Fraiman, R. and Pateiro-López, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.* **44** 311–329. MR2977397
[16] Cuevas, A., Fraiman, R. and Rodríguez-Casal, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.* **35** 1031–1051. MR2341697
[17] Cuevas, A., Llop, P. and Pateiro-López, B. (2014). On the estimation of the medial axis and inner parallel body. *J. Multivariate Anal.* **129** 171–185. MR3215988
[18] De Marco, G., Gorni, G. and Zampieri, G. (1994). Global inversion of

functions: an introduction. *NoDEA Nonlinear Differential Equations Appl.* **1** 229–248. MR1289855 (95h:58014)

[19] Dey, T. K. and Sun, J. (2006). Normal and feature approximations from noisy point clouds. In *FSTTCS 2006: Foundations of software technology and theoretical computer science. Lecture Notes in Comput. Sci.* **4337** 21–32. Springer, Berlin. MR2335319

[20] do Carmo, M. P. (1992). *Riemannian geometry. Mathematics: Theory & Applications*. Birkhäuser Boston, Inc., Boston, MA Translated from the second Portuguese edition by Francis Flaherty. MR1138207 (92i:53001)

[21] Dyer, R., Vegter, G. and Wintraecken, M. (2015). Riemannian simplices and triangulations. *Geometriae Dedicata* **179** 91–138.

[22] Federer, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. MR0110078

[23] Federer, H. (1969). *Geometric measure theory. Die Grundlehren der mathematischen Wissenschaften, Band 153*. Springer-Verlag New York Inc., New York. MR0257325 (41 ##1976)

[24] Fefferman, C., Mitter, S. and Narayanan, H. (2016). Testing the manifold hypothesis. *J. Amer. Math. Soc.* **29** 983–1049. MR3522608

[25] Genovese, C. R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012). Minimax manifold estimation. *J. Mach. Learn. Res.* **13** 1263–1291. MR2930639

[26] Giné, E. and Koltchinskii, V. (2006). Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In *High dimensional probability. IMS Lecture Notes Monogr. Ser.* **51** 238–259. Inst. Math. Statist., Beachwood, OH. MR2387773

[27] Hatcher, A. (2002). *Algebraic topology*. Cambridge University Press, Cambridge. MR1867354

[28] Hug, D., Kiderlen, M. and Svane, A. M. (2017). Voronoi-based estimation of Minkowski tensors from finite point samples. *Discrete Comput. Geom.* **57** 545–570. MR3614771

[29] Kanagawa, S., Mochizuki, Y. and Tanaka, H. (1992). Limit theorems for the minimum interpoint distance between any pair of i.i.d. random points in $\mathbf{R}^d$. *Ann. Inst. Statist. Math.* **44** 121–131. MR1165576

[30] Karcher, H. (1989). Riemannian comparison constructions. In *Global differential geometry. MAA Stud. Math.* **27** 170–222. Math. Assoc. America, Washington, DC. MR1013810

[31] Kim, A. K. H. and Zhou, H. H. (2015). Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.* **9** 1562–1582. MR3376117

[32] Klette, R. and Rosenfeld, A. (2004). *Digital geometry*. Morgan Kaufmann Publishers, San Francisco, CA; Elsevier Science B.V., Amsterdam Geometric methods for digital picture analysis. MR2095127

[33] Niyogi, P., Smale, S. and Weinberger, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441. MR2383768

[34] Rataj, J. and Zajíček, L. (2017). On the structure of sets with positive

reach. *Math. Nachr.* **290** 1806–1829. MR3683461

[35] Rodríguez-Casal, A. and Saavedra-Nieves, P. (2016). A fully data-driven method for estimating the shape of a point cloud. *ESAIM Probab. Stat.* **20** 332–348. MR3557598

[36] Singer, A. and Wu, H. T. (2012). Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.* **65** 1067–1144. MR2928092

[37] Thäle, C. (2008). 50 years sets with positive reach—a survey. *Surv. Math. Appl.* **3** 123–165. MR2443192

[38] Yu, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963 (99c:62137)