



## Estimating the Reach of a Manifold

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

### ► To cite this version:

Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, et al.. Estimating the Reach of a Manifold. 2018. hal-01521955v2

**HAL Id: hal-01521955**

**<https://hal.science/hal-01521955v2>**

Preprint submitted on 31 Jan 2018 (v2), last revised 19 Mar 2019 (v3)

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

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

# Estimating the Reach of a Manifold

Eddie Aamari <sup>*†‡</sup>	Jisu Kim <sup>§¶</sup>
<i>University of California, San Diego</i>	<i>Carnegie Mellon University</i>
Frédéric Chazal <sup>*†</sup>	Bertrand Michel <sup>*†</sup>
<i>Inria Saclay</i>	<i>École Centrale Nantes</i>
Alessandro Rinaldo <sup>§</sup>	Larry Wasserman
<i>Carnegie Mellon University</i>	<i>Carnegie Mellon University</i>

## Abstract

Various problems in manifold estimation make use of a quantity called the *reach*, denoted by  $\tau_M$ , 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  $\tau_M$  is proposed in a 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.

## 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, [29, 24, 23, 10, 31, 8, 25].

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  $\tau_M$  of the manifold  $M$  is non-zero.

First introduced by Federer [21], the reach  $\tau_M$  of a set  $M \subset \mathbb{R}^D$  is the largest number such that any point at distance less than  $\tau_M$  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  $\tau_{min}$  around it [14]. 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.

---

<sup>\*</sup>Research supported by ANR project TopData ANR-13-BS01-0008

<sup>†</sup>Research supported by Advanced Grant of the European Research Council GUDHI

<sup>‡</sup>Supported by the Conseil régional d’Île-de-France program RDM-IdF

<sup>§</sup>Supported by Samsung Scholarship

<sup>¶</sup>Partially supported by NSF CAREER Grant DMS 1149677

Positive reach is the minimal regularity assumption on sets in geometric measure theory and integral geometry [22, 33]. 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 [14] and to derive theoretical guarantees. For example, in manifold reconstruction, the reach helps formalize minimax rates [24, 29]. The optimal manifold estimators of [1] implicitly use reach as a scale parameter in their construction. In homology inference [31, 7], the reach drives the minimal sample size required to consistently estimate topological invariants. It is used in [15] 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 [4] and manifold clustering [5]. Finally, the reach often plays the role of a scale parameter in dimension reduction techniques such as vector diffusions maps [32]. 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.

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 [18]. 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 [16, 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 [23]. The authors derived a consistent test procedure, but the results do not permit any inference bound on 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 study minimax rates 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 [19]. For every point  $p$  in

$M$ , the tangent space of  $M$  at  $p$  is denoted by  $T_p M$ : 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  $c : [a, b] \rightarrow M$ , the length of  $c$  is defined as  $Length(c) = \int_a^b \|\dot{c}(t)\| dt$ . One can show [19] that there exists a path  $\gamma_{p \rightarrow q}$  of minimal length joining  $p$  and  $q$ . Such an arc is called geodesic, and the geodesic distance between  $p$  and  $q$  is given by  $d_M(p, q) = Length(\gamma_{p \rightarrow q})$ . We let  $\mathcal{B}_M(p, s)$  denote the closed geodesic ball of center  $p \in M$  and of radius  $s$ . A geodesic  $\gamma$  such that  $\|\dot{\gamma}(t)\| = 1$  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  $v \in T_p M$ , we denote by  $\gamma_{p,v}$  the unique arc-length parametrized geodesic of  $M$  such that  $\gamma_{p,v}(0) = p$  and  $\dot{\gamma}_{p,v}(0) = v$ . The exponential map is defined as  $\exp_p(vt) = \gamma_{p,v}(t)$ . Note that from the compactness of  $M$ ,  $\exp_p : T_p M \rightarrow M$  is defined globally on  $T_p M$ . For any two nonzero vectors  $u, v \in \mathbb{R}^D$ , we let  $\angle(u, v) = d_{S^{D-1}}(\frac{u}{\|u\|}, \frac{v}{\|v\|})$  be the angle between  $u$  and  $v$ .

## 2.2 Reach

First introduced by Federer [21], 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  $A$ ,

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

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

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

Some authors refer to  $\tau_A^{-1}$  as the *condition number* [31, 32]. 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  $\pi_A$  is well defined on the  $\rho$ -offset  $\{x \in \mathbb{R}^D \mid d(x, A) < \rho\}$ . 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 [21]). *For all submanifolds  $M \subset \mathbb{R}^D$ ,*

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

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).

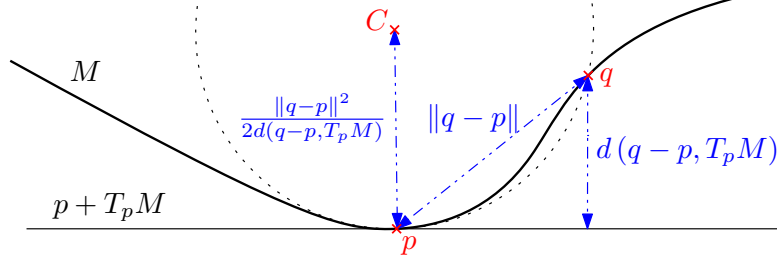


Figure 1: Geometric interpretation of quantities involved in (2.3).

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,  $\tau_M^{-1}$  is an upper bound on the curvature of  $M$ .

**Proposition 2.3** (Proposition 6.1 in [31]). *Let  $M \subset \mathbb{R}^D$  be a submanifold, and  $\gamma_{p,v}$  an arc-length parametrized geodesic of  $M$ . Then for all  $t$ ,*

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

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.

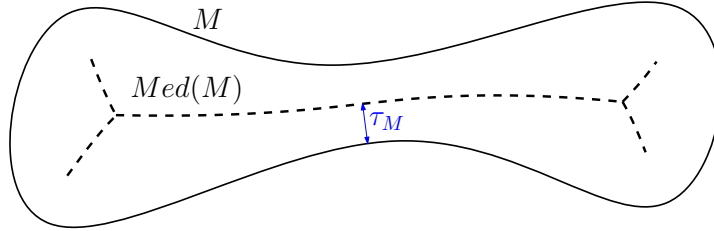


Figure 2: A narrow bottleneck structure yields a small reach  $\tau_M$ .

### 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  $\tau_M$  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.** We let  $\mathcal{M}_{\tau_{min}, L}^{d,D}$  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

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

It is important to note that any compact  $d$ -dimensional  $\mathcal{C}^3$ -submanifold  $M \subset \mathbb{R}^D$  belongs to such a class  $\mathcal{M}_{\tau_{\min}, L}^{d,D}$ , provided that  $\tau_{\min} \leq \tau_M$  and that  $L$  is large enough. Note also that since the third order condition  $\|\gamma_{p,v}'''(0)\| \leq L$  needs to hold for all  $(p, v)$ , we have in particular that  $\|\gamma_{p,v}'''(t)\| \leq L$  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  $\text{vol}_M$  from the  $d$ -dimensional Hausdorff measure  $\mathcal{H}^d$  on  $\mathbb{R}^D$  [22, p. 171]. We will consider distributions  $Q$  that have densities with respect to  $\text{vol}_M$  that are bounded away from zero.

**Definition 2.5.** We let  $\mathcal{Q}_{\tau_{\min}, L, f_{\min}}^{d,D}$  denote the set of distributions  $Q$  having support  $M \in \mathcal{M}_{\tau_{\min}, L}^{d,D}$  and with a Hausdorff density  $f = \frac{dQ}{d\text{vol}_M}$  satisfying  $\inf_{x \in M} f(x) \geq f_{\min} > 0$  on  $M$ .

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. We let  $\mathbb{G}^{d,D}$  denote the Grassmanian 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  $Q \in \mathcal{Q}_{\tau_{\min}, L, f_{\min}}^{d,D}$  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  $\mathcal{P}_{\tau_{\min}, L, f_{\min}}^{d,D}$  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  $M = \text{supp}(Q)$ . 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  $\tau_M$ . 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 in the Appendix, leading to a bound on the diameter of  $M$ , and on a topological argument (Lemma A.5 in the Appendix 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

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

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

*Remark 2.8.* For a distribution  $P \in \mathcal{P}_{\tau_{\min}, L, f_{\min}}^{d, D}$ , 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  $\tau_P$  is equivalent to the estimation of the condition number  $\tau_P^{-1}$ , 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 of the Appendix. Below,  $\sigma_d$  stands for the volume of the  $d$ -dimensional unit sphere  $\mathcal{S}^d$ .

**Proposition 2.9.** *Given  $\tau_{\min} > 0$ , provided that  $f_{\min} \leq (2^{d+1}\tau_{\min}^d\sigma_d)^{-1}$ , we have for all  $n \geq 1$ ,*

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

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

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 of the Appendix.

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  $\tau_M$  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  $\tau_M$ . 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  $z_0 \in \text{Med}(M)$  be their associated axis point, and write  $c_{z_0}(q_1, q_2)$  for the arc of the circle with center  $z_0$  and endpoints as  $q_1$  and  $q_2$ .*

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

In particular, in this “semi-local” situation, since  $\tau_M^{-1}$  is the norm of the second derivative of a geodesic of  $M$  (the exhibited arc of the circle of radius  $\tau_M$ ), 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  $\tau_M$  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  $z \in \text{Med}(M)$ ,  $d(z, M) > \tau_M$ . Then there exists  $q_0 \in M$  and a geodesic  $\gamma_0$  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:

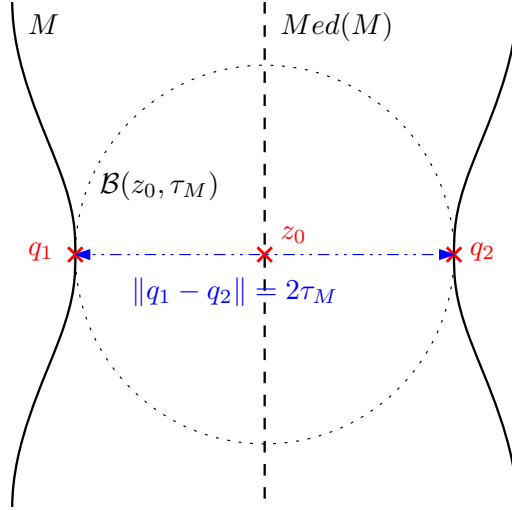
- (See Figure 3a)  $M$  has a bottleneck: by definition,  $\tau_M$  originates from a structure having scale  $2\tau_M$ .
- (See Figure 3b)  $M$  has a reach attaining pair but no bottleneck: then  $M$  contains an arc of a circle of radius  $\tau_M$  (Lemma 3.2), so that  $M$  actually contains a zone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

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.

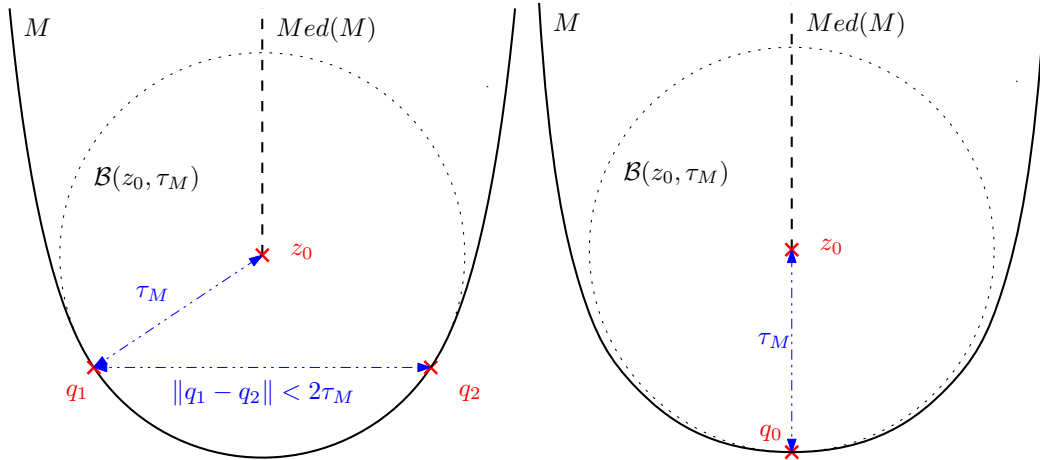
**Theorem 3.4.** *Let  $M \subset \mathbb{R}^D$  be a compact submanifold with reach  $\tau_M > 0$ . At least one of the following two assertions holds.*

- (Global Case)  $M$  has a bottleneck  $(q_1, q_2) \in M^2$ , that is, there exists  $z_0 \in \text{Med}(M)$  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  $\gamma_0$  such that  $\gamma_0(0) = q_0$  and  $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$ .





(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 (Lemma 3.2 and Lemma 3.3).

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 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  $\tau_M$  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  $\mathbb{X} \subset M$ ,

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

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  $\tau_M$ . 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  $\tau_M$  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 of the Appendix.

### 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  $\tau_M$ .

**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*

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

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

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

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

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  $\tau_M$  boils down to estimating the curvature of  $M$  at  $q_0$  in the direction  $v_0$ .

We first relate directional curvature to the increment  $\frac{\|y-x\|^2}{2d(y-x, T_x M)}$  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  $\gamma_{x \rightarrow y}$  denote the arc-length parametrized geodesic joining  $x$  and  $y$ , with the convention  $\gamma_{x \rightarrow y}(0) = x$ .

**Lemma 4.4.** *Let  $M \in \mathcal{M}_{\tau_{\min}, L}^{d, D}$  with reach  $\tau_M$  and  $\mathbb{X} \subset M$  be a subset. Let  $x, y \in \mathbb{X}$  with  $d_M(x, y) < \pi\tau_M$ . Then,*

$$0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x, y\})} \leq \frac{1}{\tau_M} - \|\gamma_{x \rightarrow y}''(0)\| + \frac{2}{3}Ld_M(x, y).$$

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

$$\kappa_p = \sup_{v \in \mathcal{B}_{T_p M}(0, 1)} \|\gamma_{p, v}''(0)\|.$$

**Lemma 4.5.** *Let  $M \in \mathcal{M}_{\tau_{\min}, L}^{d, D}$  with reach  $\tau_M$  and  $q_0, x, y \in M$  be such that  $x, y \in \mathcal{B}_M(q_0, \frac{\pi\tau_M}{2})$ . Let  $\gamma_0$  be a geodesic such that  $\gamma_0(0) = q_0$  and  $\|\gamma_0''(0)\| = \kappa_{q_0}$ . Write*

$$\theta_x := \angle(\gamma_0'(0), \gamma_{q_0 \rightarrow x}'(0)), \quad \theta_y := \angle(\gamma_0'(0), \gamma_{q_0 \rightarrow y}'(0)),$$

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

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

In particular, geodesics in a neighborhood of  $q_0$  with directions close to  $v_0$  have curvature close to  $\frac{1}{\tau_M}$ . A point cloud  $\mathbb{X}$  sampled densely enough in  $M$  would contain points in this neighborhood. Hence combining Lemma 4.4 and Lemma 4.5 yields the following deterministic bound in the local case.

**Proposition 4.6.** *Under the same conditions as Lemma 4.5,*

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

In other words, since the reach boils down to directional curvature in the local case,  $\hat{\tau}$  performs well if it is given as input a pair of points  $x, y$  which are close to the point  $q_0$  realizing the reach, and almost aligned with the direction of interest  $v_0$ . Note that the error bound in the local case (Proposition 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  $\tau_M = \|\gamma''_{q_0, v_0}(0)\|$  (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  $\mathbb{X}_n = \{X_1, \dots, X_n\}$  is random.

**Proposition 4.7.** *Let  $P \in \mathcal{P}_{\tau_{\min}, L, f_{\min}}^{d, D}$  and  $M = \text{supp}(P)$ . Suppose there exists  $q_0 \in M$  and a geodesic  $\gamma_0$  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  $\mathcal{P}_{\tau_{\min}, L, f_{\min}}^{d, D}$ , that is

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

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

The rate of convergence of the plugin estimator  $\hat{\tau}(\mathbb{X}_n)$  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 [34]. In what follows, let

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

denote the total variation distance between  $P$  and  $P'$ , where  $dP, dP'$  denote the respective densities of  $P, P'$  with respect to any dominating measure. Since  $|x - z|^p + |z - y|^p \geq 2^{1-p}|x - y|^p$ , the following version of Le Cam's lemma results from Lemma 1 in [34] and  $(1 - TV(P^n, P'^n)) \geq (1 - TV(P, P'))^n$ .

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

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

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

**Lemma 5.3.** *In accordance with the notation of Definition 2.6, let  $Q, Q' \in \mathcal{Q}_{\tau_{\min}, L, f_{\min}}^{d, D}$  be distributions on  $\mathbb{R}^D$  with associated distributions  $P, P' \in \mathcal{P}_{\tau_{\min}, L, f_{\min}}^{d, D}$  on  $\mathbb{R}^D \times \mathbb{G}^{d, D}$ . Then,*

$$TV(P, P') = TV(Q, Q').$$

In order to construct hypotheses in  $\mathcal{Q}_{\tau_{\min}, L, f_{\min}}^{d,D}$  we take advantage of the fact that the class  $\mathcal{M}_{\tau_{\min}, L}^{d,D}$  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  $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^D$ , we denote by  $d_x^i \Phi$  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 of the Appendix.

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

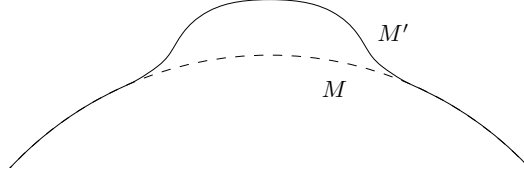


Figure 4: Hypotheses of Proposition 5.5.

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  $\Phi$  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 of the Appendix.

**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  $Q, Q' \in \mathcal{Q}_{\tau_{\min}, L, f_{\min}}^{d,D}$  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,*

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

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

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

From Proposition 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

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

In particular,  $\hat{\tau}(\mathbb{X}) = \hat{\tau}(\mathbb{X}, T_{\mathbb{X}}M)$ , where  $T_{\mathbb{X}}M = \{T_x M\}_{x \in \mathbb{X}}$ . 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 of the Appendix. 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  $\delta$ -sparse,  $T$  and  $\tilde{T}$  to be  $\theta$ -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  $|\hat{\tau}(\mathbb{X})^{-1} - \tau_M^{-1}|$  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 rates of tangent space estimation derived in  $\mathcal{C}^3$ -like models can be found in [13, 32, 2], yielding bounds on  $\sin \theta$ . 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 [27] 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.

- The minimax upper and lower bounds given in Theorem 5.1 and Theorem 5.6 do not match. They are yet to be sharpened.
- In practice, since large reach ensures regularity, one may be interested with having a lower bound on the reach  $\tau_M$ . Giving the limiting distribution of the statistic  $\hat{\tau}(\mathbb{X}_n)$  would allow to derive asymptotic confidence intervals for  $\tau_M$ .
- Other regularity parameters such as local feature size [10] and  $\lambda$ -reach [12] could be relevant to estimate, as they are used as tuning parameters in computational geometry techniques.

## Acknowledgements

This collaboration was made possible by the associated team CATS (Computations And Topological Statistics) between DataShape and Carnegie Mellon University.



## Table of Notation

$C_{\alpha,\beta}, c_{\alpha,\beta}, C'_{\alpha,\beta}, c'_{\alpha,\beta}$	Constant depending on the parameters $\alpha, \beta$
$D$	Ambient dimension
$d$	Manifold dimension
$T_x M$	Tangent space of $M$ at $x$
$\tau_M, \tau_Q, \tau_P$	Reach of $M = \text{Supp}(Q) = \text{Supp}(P)$
$\tau_{min}$	Prescribed lower bound on $\tau_M$
$L$	Prescribed third order derivative bound on geodesics
$\mathcal{H}^d$	$d$ -dimensional Hausdorff measure on $\mathbb{R}^D$
$\text{vol}_M$	Volume measure of $M \subset \mathbb{R}^D$
$f$	Density with respect to the volume measure
$f_{min}$	Prescribed lower bound on $f$
$\lambda_M$	Uniform probability distribution on $M$ , i.e. $f = \mathcal{H}^d(M)^{-1}$
$\mathcal{M}^{d,D}_{\tau_{min},L}$	Geometric model
$\mathcal{Q}^{d,D}_{\tau_{min},L,f_{min}}$	Statistical Model with unknown tangent spaces
$Q, Q'$	Element of $\mathcal{Q}^{d,D}_{\tau_{min},L,f_{min}}$
$\mathcal{P}^{d,D}_{\tau_{min},L,f_{min}}$	Statistical Model with known tangent spaces
$P, P'$	Element of $\mathcal{P}^{d,D}_{\tau_{min},L,f_{min}}$
$\text{Supp}(\cdot)$	Support of a distribution
$\ \cdot\ $	Euclidean norm
$\langle \cdot, \cdot \rangle$	Euclidean inner product
$\mathcal{B}(x, r)$	Closed Euclidean ball
$\gamma_{p,v}$	Geodesic passing through $p$ with direction $v$
$\gamma_{x \rightarrow y}$	Geodesic joining $x$ to $y$ with $\gamma_{x \rightarrow y}(0) = x$
$d_M(\cdot, \cdot)$	Geodesic distance
$\mathcal{B}_M(p, s)$	Closed Geodesic ball
$d(\cdot, A)$	Distance to a subset $A \subset \mathbb{R}^D$
$\pi_M$	Projection map onto $M$
$\text{Med}(\cdot)$	Medial axis
$d_x \Phi \cdot h$	Differential at $x$ in the direction $h$
$d_x^2 \Phi, d_x^3 \Phi$	Higher order differentials
$\ \cdot\ _{op}$	Operator norm

## A Some Technical Results on the Model

### A.1 Metric Properties

This section gathers geometric lemmas on embedded manifolds in the Euclidean space that are related to the reach, and that will be used several times in the proofs.

**Proposition A.1.** *Let  $M \subset \mathbb{R}^D$  be a submanifold with reach  $\tau_M > 0$ .*

- (i) *For all  $p \in M$ , we let  $II_p$  denote the second fundamental form of  $M$  at  $x$ . Then for all unit vector  $v \in T_p M$ ,  $\|II_p(v, v)\| \leq \frac{1}{\tau_M}$ .*
- (ii) *The injectivity radius of  $M$  is at least  $\pi\tau_M$ .*
- (iii) *The sectional curvatures  $\kappa$  of  $M$  satisfy  $-\frac{2}{\tau_M^2} \leq \kappa \leq \frac{1}{\tau_M^2}$ .*

(iv) For all  $p \in M$ , the map  $\exp_p : \mathring{\mathcal{B}}_{T_p M}(0, \pi\tau_M) \rightarrow \mathring{\mathcal{B}}_M(0, \pi\tau_M)$  is a diffeomorphism. Moreover, for all  $\|v\| < \frac{\pi\tau_M}{2\sqrt{2}}$  and  $w \in T_p M$ ,

$$\left(1 - \frac{\|v\|^2}{6\tau_M^2}\right) \|w\| \leq \|d_v \exp_p \cdot w\| \leq \left(1 + \frac{\|v\|^2}{\tau_M^2}\right) \|w\|$$

(v) For all  $p \in M$  and  $r \leq \frac{\pi\tau_M}{2\sqrt{2}}$ , given any Borel set  $A \subset \mathcal{B}_{T_p M}(0, r) \subset T_p M$  we have

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

(vi) Let  $\gamma$  be a geodesic at  $p \in M$ , and  $P_t$  the parallel transport operator along  $\gamma$ . Then for all  $t < \pi\tau_M$  and  $v \in T_p M$ ,

$$\angle(P_t(v), v) \leq \frac{t}{\tau_M}.$$

*Proof of Proposition A.1.* (i) is stated in Proposition 2.1 in [31], yielding (ii) from Corollary 1.4 in [3]. (iii) follows using (i) again and the Gauss equation [19, p. 130]. (iv) is derived from (iii) by a direct application of Lemma 8 in [20]. (v) follows from (iv) and Lemma 6 in [5]. All that remain to be showed is (vi).

For this, assume without loss of generality that  $\|v\| = 1$ . Let  $g : [0, t] \rightarrow \mathcal{S}^{d-1}$  be defined by  $g(s) = P_s(v)$ . Let  $u \in \mathbb{R}^D$  be a unit vector and denoting by  $\bar{\nabla}$  the ambient derivative. We may write

$$\langle g'(s), u \rangle = \langle \bar{\nabla}_{\gamma'(s)} P_s(w), u \rangle = \langle II(\gamma'(s), P_s(w)), u \rangle.$$

Hence  $\|g'(s)\| \leq \frac{1}{\tau_M}$  for all  $s \in [0, t]$ . Since  $g$  is a curve on  $\mathcal{S}^{d-1}$ , this implies

$$\angle(P_t(v), v) = d_{\mathcal{S}^{d-1}}(\gamma(t), \gamma(0)) \leq \int_0^t \|g'(s)\| ds \leq \frac{t}{\tau_M}.$$

□

## 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.

**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,*

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

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,*

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

*Proof of Proposition A.3.* Follows from a straightforward combination of Lemma A.4 and Lemma A.5.  $\square$

We recall that for two compact subsets  $A, B \subset \mathbb{R}^D$ , the Hausdorff distance [11, 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  $\text{conv}(\cdot)$  the closed convex hull of a set.

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

*Proof of Lemma A.4.* It is a straightforward corollary of Jung's Theorem 2.10.41 in [22], which states that  $K$  is contained in a (unique) closed ball with (minimal) radius at most  $\sqrt{\frac{D}{2(D+1)}} \text{diam}(K)$ .  $\square$

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

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

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

Therefore, the map  $\pi_K : \text{conv}(K) \rightarrow K$  is well defined and continuous, so that  $K$  is a retract of  $\text{conv}(K)$  (see Chapter 0 in [26]). Therefore,  $K$  is homotopy equivalent to a point, since the convex set  $\text{conv}(K)$  is.  $\square$

We are now in position to prove Proposition 2.7.

*Proof of Proposition 2.7.* From Theorem 3.26 in [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  $\tau_M \leq \text{diam}(M)$ , and we conclude by applying the bound  $\text{diam}(M) \leq C_d/(\tau_M^{d-1} f_{\min})$  given by Lemma A.2.  $\square$

## B Geometry of the Reach

For  $M \subset \mathbb{R}^D$ ,  $a \in M$ , and  $v \in \mathbb{R}^D$  a non-zero vector, we define the *local directional reach* by

$$\tau_M(a, v) = \inf \left\{ d(x, M) \mid x \in \overline{\text{Med}(M)} \text{ with } x = a + tv \text{ for some } t \geq 0 \right\},$$

with the convention  $\tau_M(a, v) = \infty$  if  $\overline{\text{Med}(M)} \cap \{a + tv \mid t \geq 0\} = \emptyset$ .

**Lemma B.1.** *(i) For  $x \notin \text{Med}(M) \cup M$ , writing  $a = \pi_M(x)$ , we have  $\tau_M(a, x - a) > 0$ , and for all  $b \in M$ ,*

$$\langle x - a, a - b \rangle \geq -\frac{\|a - b\|^2 \|x - a\|}{2\tau_M(a, x - a)}.$$

(ii) Let  $0 < r < q < \infty$  be fixed. Let  $x, y \notin \text{Med}(M) \cup M$  be such that  $d(x, M) \vee d(y, M) \leq r$  and

$$\tau_M(\pi_M(x), x - \pi_M(x)) \wedge \tau_M(\pi_M(y), y - \pi_M(y)) \geq q.$$

Then,

$$\|\pi_M(x) - \pi_M(y)\| \leq \frac{q}{q-r} \|x - y\|.$$

*Proof of Lemma B.1.* The proof of (i) follows that of Theorem 4.8 (7) in [21]. Let  $v = \frac{x-a}{\|x-a\|}$  and  $S = \{t | \pi_M(a + tv) = a\}$ . As  $\|x - a\| > 0$  belongs to  $S$ ,  $\sup S > 0$  and from [21, Theorem 4.8 (6)] we get  $\sup S \geq \tau_M(a, v)$ . Moreover, for  $0 < t \in S$ ,

$$\|a + tv - b\| \geq d(a + tv, M) = t.$$

Developing and rearranging the square of previous inequality yields

$$\begin{aligned} \|a - b\|^2 + 2t \langle v, a - b \rangle + t^2 &\geq t^2, \\ 2t \langle v, a - b \rangle &\geq -\|a - b\|^2, \\ \langle x - a, a - b \rangle &\geq -\frac{\|a - b\|^2 \|x - a\|}{2t}. \end{aligned}$$

On the other hand, the proof of (ii) follows that of Theorem 4.8 (8) in [21]. Writing  $a = \pi_M(x)$  and  $b = \pi_M(y)$ , the previous point yields

$$\langle x - a, a - b \rangle \geq -\frac{\|a - b\|^2 r}{2q} \quad \text{and} \quad \langle y - b, b - a \rangle \geq \frac{\|a - b\|^2 r}{2q}.$$

As a consequence,

$$\begin{aligned} \|x - y\| \|a - b\| &\geq \langle x - y, a - b \rangle \\ &= \langle (x - a) + (a - b) + (b - y), a - b \rangle \\ &\geq \|a - b\|^2 \left(1 - \frac{r}{q}\right), \end{aligned}$$

hence the result.  $\square$

**Lemma B.2.** Let  $M \subset \mathbb{R}^D$  be a submanifold with reach  $\tau_M > 0$  having a reach attaining pair  $(q_1, q_2) \in M^2$  such that  $\|q_1 - q_2\| < 2\tau_M$ . Write  $z_0 \in \text{Med}(M)$  for the associated axis point. Then there exists a sequence of curves  $\{\gamma_n\}_{n \in \mathbb{N}}$  of  $M$  joining  $q_1$  and  $q_2$  with

$$\lim_{n \rightarrow \infty} \text{Length}(\gamma_n) = \tau_M \angle(q_1 - z_0, q_2 - z_0).$$

*Proof of Lemma B.2.* Without loss of generality, assume that  $z_0$  coincides with the origin. Let  $c_{z_0}(q_1, q_2)$  be the circle arc of center  $z_0$  with endpoints  $q_1$  and  $q_2$ , and let  $\gamma : [-t_0, t_0] \rightarrow c_{z_0}(q_1, q_2)$  be its arc length parametrization with  $\gamma(-t_0) = q_1$  and  $\gamma(t_0) = q_2$ . Let  $\theta := \angle(q_1 - z_0, q_2 - z_0)$ . Since  $\|q_1 - z_0\| = \|q_2 - z_0\| = \tau_M$ , we have  $t_0 = \frac{1}{2}\tau_M\theta$ . For all  $t \in [-t_0, t_0]$ , let  $r_t := \sqrt{\tau_M^2 - \frac{\|q_1 - q_2\|^2}{4}} / \cos\left(\frac{t}{\tau_M}\right)$ ,

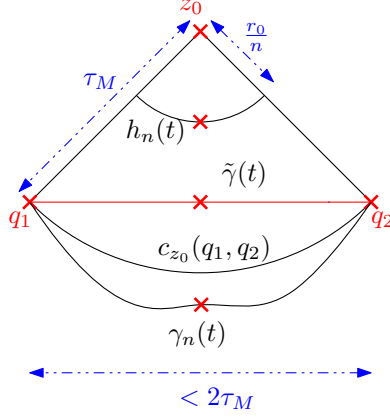


Figure 5: Layout of the proof of Lemma B.2.

and let  $\tilde{\gamma} : [-t_0, t_0] \rightarrow \mathbb{R}^D$  be  $\tilde{\gamma}(t) = \frac{r_t}{\tau_M} \gamma(t)$ . Let us show that for all  $r \in (0, r_0]$  and  $t \in [-t_0, t_0]$ , the following holds:

$$\mathring{\mathcal{B}}\left(\frac{r}{\tau_M} \gamma(t), r\right) \subset \mathring{\mathcal{B}}(\tilde{\gamma}(t), r_t) \subset \mathring{\mathcal{B}}(q_1, \tau_M) \cup \mathring{\mathcal{B}}(q_2, \tau_M). \quad (\text{B.1})$$

The left-hand side inclusion of (B.1) being straightforward, we turn to the second inclusion. First, note that by definition,

$$\tilde{\gamma}(t) = \left(\frac{1}{2} - \frac{\tan\left(\frac{t}{\tau_M}\right)}{2 \tan\left(\frac{t_0}{\tau_M}\right)}\right) q_1 + \left(\frac{1}{2} + \frac{\tan\left(\frac{t}{\tau_M}\right)}{2 \tan\left(\frac{t_0}{\tau_M}\right)}\right) q_2$$

for all  $t \in [-t_0, t_0]$ . Hence,

$$\tilde{\gamma}(t) - \tilde{\gamma}(0) = \frac{\tan\left(\frac{t}{\tau_M}\right)}{2 \tan\left(\frac{t_0}{\tau_M}\right)} (q_2 - q_1), \quad (\text{B.2})$$

and from  $\tan\left(\frac{t_0}{\tau_M}\right) = \frac{\|q_1 - q_2\|}{2r_0}$ , we get  $\|\tilde{\gamma}(t) - \tilde{\gamma}(0)\| = r_0 \tan\left(\frac{t}{\tau_M}\right)$ . Now suppose that  $x \in \mathring{\mathcal{B}}(\tilde{\gamma}(t), r_t)$ , then

$$\|x - \tilde{\gamma}(t)\|^2 < r_t^2. \quad (\text{B.3})$$

Then,

$$\|x - \tilde{\gamma}(t)\|^2 = \|x - \tilde{\gamma}(0)\|^2 - 2 \langle x - \tilde{\gamma}(0), \tilde{\gamma}(t) - \tilde{\gamma}(0) \rangle + \|\tilde{\gamma}(t) - \tilde{\gamma}(0)\|^2,$$

and  $r_t^2 = r_0^2 + r_0^2 \tan^2\left(\frac{t}{\tau_M}\right) = r_0^2 + \|\tilde{\gamma}(t) - \tilde{\gamma}(0)\|^2$ , hence applying these and (B.2) to (B.3) implies

$$\|x - \tilde{\gamma}(0)\|^2 - \frac{\tan\left(\frac{t}{\tau_M}\right)}{\tan\left(\frac{t_0}{\tau_M}\right)} \langle x - \tilde{\gamma}(0), q_2 - q_1 \rangle < r_0^2. \quad (\text{B.4})$$

Now applying  $\tilde{\gamma}(-t_0) = q_1$  to (B.2) gives  $q_1 - \tilde{\gamma}(0) = -\frac{1}{2}(q_2 - q_1)$ , so

$$\begin{aligned}\|x - q_1\|^2 &= \|x - \tilde{\gamma}(0)\|^2 + 2\langle x - \tilde{\gamma}(0), q_1 - \tilde{\gamma}(0) \rangle + \|q_1 - \tilde{\gamma}(0)\|^2 \\ &= \|x - \tilde{\gamma}(0)\|^2 - \langle x - \tilde{\gamma}(0), q_2 - q_1 \rangle + \frac{1}{4}\|q_1 - q_2\|^2.\end{aligned}$$

Similarly,

$$\|x - q_2\|^2 = \|x - \tilde{\gamma}(0)\|^2 + \langle x - \tilde{\gamma}(0), q_2 - q_1 \rangle + \frac{1}{4}\|q_1 - q_2\|^2,$$

and hence

$$\begin{aligned}\min\left\{\|x - q_1\|^2, \|x - q_2\|^2\right\} \\ = \|x - \tilde{\gamma}(0)\|^2 - |\langle x - \tilde{\gamma}(0), q_2 - q_1 \rangle| + \frac{1}{4}\|q_1 - q_2\|^2.\end{aligned}\tag{B.5}$$

Since  $\left|\tan\left(\frac{t_0}{\tau_M}\right)\right| \geq \left|\tan\left(\frac{t}{\tau_M}\right)\right|$ , applying (B.4) to (B.5) gives

$$\begin{aligned}\min\left\{\|x - q_1\|^2, \|x - q_2\|^2\right\} \\ \leq \|x - \tilde{\gamma}(0)\|^2 - \frac{\tan\left(\frac{t}{\tau_M}\right)}{\tan\left(\frac{t_0}{\tau_M}\right)} \langle x - \tilde{\gamma}(0), q_2 - q_1 \rangle + \frac{1}{4}\|q_1 - q_2\|^2 \\ < r_0^2 + \frac{1}{4}\|q_1 - q_2\|^2 = \tau_M^2,\end{aligned}$$

which asserts the second inclusion in (B.1).

Now, by definition of the reach in (2.2),  $\left(\mathring{\mathcal{B}}(q_1, \tau_M) \cup \mathring{\mathcal{B}}(q_2, \tau_M)\right) \cap \text{Med}(M) = \emptyset$ , hence (B.1) implies

$$\mathring{\mathcal{B}}\left(\frac{r}{\tau_M}\gamma(t), r\right) \cap \text{Med}(M) = \emptyset.$$

For all  $n \in \mathbb{N}$ , let us now define  $h_n, \gamma_n : [-t_0, t_0] \rightarrow M$  by (See Figure 5),

$$h_n(t) = \frac{r_0}{n\tau_M}\gamma(t) \quad \text{and} \quad \gamma_n(t) = \pi_M(h_n(t)).$$

Then for any fixed  $n \in \mathbb{N}$  and  $t_1, t_2 \in [-t_0, t_0]$  such that  $|t_1 - t_2| < \tau_M$ , from  $\mathring{\mathcal{B}}(h_n(t_i), \frac{r_0}{n}) \cap \text{Med}(M) = \emptyset$ , we get

$$\begin{aligned}\tau_M(\gamma_n(t_i), h_n(t_i) - \gamma_n(t_i)) &\geq d(h_n(t_i), M) + \frac{r_0}{n} \\ &\geq d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n},\end{aligned}$$

and since  $d(h_n(t_i), M) \leq d(h_n(t_1), M) \vee d(h_n(t_2), M)$ , Lemma B.1 (ii) yields

$$\begin{aligned}\|\gamma_n(t_1) - \gamma_n(t_2)\| &= \|\pi_M(h_n(t_1)) - \pi_M(h_n(t_2))\| \\ &\leq \frac{(d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n}) \|h_n(t_1) - h_n(t_2)\|}{d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n} - d(h_n(t_1), M) \vee d(h_n(t_2), M))} \\ &= \frac{d(h_n(t_1), M) \wedge d(h_n(t_2), M) + \frac{r_0}{n}}{\frac{r_0}{n} - |d(h_n(t_1), M) - d(h_n(t_2), M)|} \|h_n(t_1) - h_n(t_2)\|.\end{aligned}$$

Noticing furthermore that

$$|d(h_n(t_1), M) - d(h_n(t_2), M)| \leq \|h_n(t_1) - h_n(t_2)\| \leq \frac{r_0}{n\tau_M} |t_1 - t_2|,$$

and

$$d(h_n(t_i), M) \leq d(z_0, M) + \|h_n(t_i) - z_0\| \leq \tau_M + \frac{r_0}{n},$$

we get

$$\begin{aligned} \|\gamma_n(t_1) - \gamma_n(t_2)\| &\leq \frac{\tau_M + 2\frac{r_0}{n}}{\frac{r_0}{n} - \frac{r_0}{n\tau_M} |t_1 - t_2|} \frac{r_0}{n\tau_M} |t_1 - t_2| \\ &= \frac{\tau_M + 2\frac{r_0}{n}}{\tau_M - |t_1 - t_2|} |t_1 - t_2|. \end{aligned}$$

For any fixed  $k$  and  $0 \leq j \leq k$ , set  $t_{k,j} = \frac{2j-k}{k}t_0$ . The inequality above yields,

$$\sum_{j=1}^k \|\gamma_n(t_{k,j}) - \gamma_n(t_{k,j-1})\| \leq \frac{\tau_M + 2\frac{r_0}{n}}{\tau_M - \frac{2t_0}{k}} 2t_0,$$

so

$$\text{Length}(\gamma_n) = \limsup_k \sum_{j=1}^k \|\gamma_n(t_{k,j}) - \gamma_n(t_{k,j-1})\| \leq \left(1 + \frac{2r_0}{\tau_M n}\right) 2t_0.$$

Moreover, the  $\gamma_n$ 's are curves joining  $q_1$  to  $q_2$  with images  $\gamma_n([-t_0, t_0]) \subset \mathbb{R}^D \setminus \mathring{B}(z_0, \tau_M)$ , so that their lengths are at most that of the arc of great circle  $c_{z_0}(q_1, q_2)$ , that is

$$\text{Length}(\gamma_n) \geq \text{Length}(c_{z_0}(q_1, q_2)) = 2t_0.$$

Hence,

$$\lim_{n \rightarrow \infty} \text{Length}(\gamma_n) = 2t_0 = \tau_M \theta.$$

□

**Lemma B.3.** *Let  $M$  be a compact manifold, and  $q_1, q_2 \in M$  with  $q_1 \neq q_2$ . Let  $(\gamma_n)_{n \in \mathbb{N}}$  be a sequence of curves on  $M$  joining  $q_1$  and  $q_2$  such that  $\sup_n \text{Length}(\gamma_n) < \infty$ . Then there exists a curve  $\gamma$  on  $M$  joining  $q_1$  and  $q_2$  such that*

$$\liminf_{n \rightarrow \infty} \text{Length}(\gamma_n) \leq \text{Length}(\gamma) \leq \limsup_{n \rightarrow \infty} \text{Length}(\gamma_n).$$

*Proof of Lemma B.3.* Without loss of generality, we take the  $\gamma_n$ 's to be arc length parametrized. For all  $n \in \mathbb{N}$ , we let  $g_n : [0, 1] \rightarrow M$  be the reparametrization  $g_n(t) = \gamma_n(\text{Length}(\gamma_n)t)$ . Notice that for all  $t \in [0, 1]$ , the set  $(g_n(t))_{n \in \mathbb{N}}$  is contained in the compact set  $M$ , so that it is bounded uniformly in  $t$ . Moreover, writing  $K = \sup_n \text{Length}(\gamma_n) < \infty$ , we have that for all  $t_1, t_2 \in [0, 1]$ ,

$$\begin{aligned} \|g_n(t_1) - g_n(t_2)\| &= \|\gamma_n(\text{Length}(\gamma_n)t_1) - \gamma_n(\text{Length}(\gamma_n)t_2)\| \\ &\leq \text{Length}(\gamma_n) |t_1 - t_2| \\ &\leq K |t_1 - t_2|. \end{aligned}$$

Hence, the sequence  $(g_n)_{n \in \mathbb{N}}$  is pointwise bounded and equicontinuous. From Arzelà-Ascoli theorem [30, Theorem 45.4], there exists a curve  $\gamma : [0, 1] \rightarrow M$  and subsequence  $(g_{n_i})_{i \in \mathbb{N}}$  converging uniformly to  $\gamma$ .

For any fixed  $k$  and  $1 \leq j \leq k$ , set  $t_{k,j} = j/k$ . The (pointwise) convergence of  $(g_{n_i})_i$  towards  $\gamma$  ensures that

$$\sum_{j=0}^{k-1} \|\gamma(t_{k,j+1}) - \gamma(t_{k,j})\| = \lim_{i \rightarrow \infty} \sum_{j=0}^{k-1} \|g_{n_i}(t_{k,j+1}) - g_{n_i}(t_{k,j})\|.$$

Furthermore, from the uniform convergence of  $(g_{n_i})_i$  towards  $\gamma$  on the compact set  $[0, 1]$ ,

$$\begin{aligned} \text{Length}(\gamma) &= \lim_{k \rightarrow \infty} \sum_{j=0}^{k-1} \|\gamma(t_{k,j+1}) - \gamma(t_{k,j})\| \\ &= \lim_{k \rightarrow \infty} \lim_{i \rightarrow \infty} \sum_{j=0}^{k-1} \|g_{n_i}(t_{k,j+1}) - g_{n_i}(t_{k,j})\| \\ &= \lim_{i \rightarrow \infty} \text{Length}(g_{n_i}) = \lim_{i \rightarrow \infty} \text{Length}(\gamma_{n_i}), \end{aligned}$$

hence the result.  $\square$

*Proof of Lemma 3.2.* Combining Lemma B.2 and Lemma B.3 provides the existence of a curve  $\gamma \subset M$  joining  $q_1$  and  $q_2$  such that  $\text{Length}(\gamma) = \text{Length}(c_{z_0}(q_1, q_2))$ . But  $M \subset \mathbb{R}^D \setminus \mathring{\mathcal{B}}(z_0, \tau_M)$ , and since  $\|q_1 - q_2\| < 2\tau_M$ ,  $c_{z_0}(q_1, q_2)$  is the unique minimizing geodesic of  $\partial\mathcal{B}(z_0, \tau_M) \subset \mathbb{R}^D \setminus \mathring{\mathcal{B}}(z_0, \tau_M)$  joining  $q_1$  and  $q_2$ . Therefore,  $\gamma = c_{z_0}(q_1, q_2) \subset M$ , hence the result.  $\square$

**Lemma B.4.** *Let  $M \in \mathcal{M}_{\tau_{\min}, L}^{d,D}$  be a submanifold with reach  $\tau_M$ . For all  $p \in M$ , let us denote*

$$L_p := \sup_{\substack{q \in \mathcal{B}_M(p, \tau_M/2) \\ v \in \mathcal{B}_{T_q M}(0, 1)}} \|\gamma_{q,v}'''(0)\|.$$

*Then for all  $r \leq \tau_M/2$ ,*

$$\left| \sup_{v \in T_p M, \|v\|=1} \|\gamma_{p,v}''(0)\| - \sup_{q \in \mathcal{B}(p, r) \cap M} \frac{2d(q-p, T_p M)}{\|q-p\|^2} \right| \leq 3 \left( \frac{1}{\tau_M^2} + L_p \right) r.$$

To prove Lemma B.4 we need the following straightforward result.

**Lemma B.5.** *Let  $U$  be a linear space and  $u \in U$ ,  $n \in U^\perp$ . If  $v = u + n + e$ , then*

$$|d(v, U) - \|v - u\|| \leq \|e\|.$$

*Proof of Lemma B.4.* First note that for all unit vector  $v \in T_p M$ ,  $\gamma_{p,v}(r)$  belongs to  $\mathcal{B}(p, r) \cap M$  and, whenever  $0 < r \leq \frac{\tau_M}{2}$ , Proposition A.1 (ii) ensures that  $\gamma_{p,v}(r) \neq p$ . Therefore, it suffices to show that for all  $q \in \mathcal{B}(p, r) \cap M$ , there exists a unit tangent vector  $v \in T_p M$  such that

$$\left| \|\gamma_{p,v}''(0)\| - \frac{2d(q-p, T_p M)}{\|q-p\|^2} \right| \leq 3 \left( \frac{1}{\tau_M^2} + L_p \right) r.$$



Let  $q \in \mathcal{B}(p, r) \cap M$  be different from  $p$ . Denoting  $t = d_M(p, q) > 0$ , we let  $\gamma = \gamma_{p \rightarrow q}$  be the arc-length parametrized geodesic of minimal length such that  $\gamma(0) = p$  and  $\gamma(t) = q$ .  $\gamma$  exists from Proposition A.1 (ii) since  $r \leq \frac{\tau_M}{2} < \pi\tau_M$ . We will show that  $v = \gamma'(0)$  provides the desired bound.

First, a Taylor expansion at zero of  $\gamma$  yields,

$$\left\| \frac{q-p}{t} - \gamma'(0) - \frac{t}{2} \gamma''(0) \right\| \leq L_p \frac{t^2}{6}.$$

Since  $\gamma''(0) \in T_p M^\perp$ , Lemma B.5 shows that

$$\left| d\left(\frac{q-p}{t}, T_p M\right) - \left\| \frac{q-p}{t} - \gamma'(0) \right\| \right| \leq L_p \frac{t^2}{6}.$$

Therefore,

$$\begin{aligned} & \left| \frac{2}{t} d\left(\frac{q-p}{t}, T_p M\right) - \|\gamma''(0)\| \right| \\ & \leq \frac{2}{t} \left( \left| d\left(\frac{q-p}{t}, T_p M\right) - \left\| \frac{q-p}{t} - \gamma'(0) \right\| \right| + \left\| \frac{q-p}{t} - \gamma'(0) - \frac{t}{2} \gamma''(0) \right\| \right) \\ & \leq \frac{2}{3} L_p t. \end{aligned}$$

This yields,

$$\left| \frac{2d(q-p, T_p M)}{\|q-p\|^2} - \|\gamma''(0)\| \right| \leq 2d(q-p, T_p M) \left| \frac{1}{d_M(p, q)^2} - \frac{1}{\|q-p\|^2} \right| + \frac{2}{3} L_p t.$$

Moreover, from  $\|q-p\| \leq d_M(p, q)$  and Proposition 6.3 in [31], we derive

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

where the last two inequalities follow from elementary real analysis arguments.

Therefore, we get  $t \leq 2\|q-p\|$  and

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

Finally, using (2.3) we derive,

$$\begin{aligned} \left| \|\gamma''(0)\| - \frac{2d(q-p, T_p M)}{\|q-p\|^2} \right| & \leq 2d(q-p, T_p M) \frac{3}{\tau_M \|q-p\|} + \frac{4}{3} L_p \|q-p\| \\ & \leq \frac{3}{\tau_M^2} \|q-p\| + \frac{4}{3} L_p \|q-p\| \\ & \leq 3 \left( \frac{1}{\tau_M^2} + L_p \right) r. \end{aligned}$$

□

*Proof of Lemma 3.3.* For  $r > 0$ , let  $\Delta_r := \{(p, q) \in M^2 \mid \|p - q\| < r\}$ , and  $\bar{\Delta} = \bigcap_{r>0} \Delta_r$  denote the diagonal of  $M^2$ . Consider the map  $\varphi : M^2 \setminus \bar{\Delta} \rightarrow \mathbb{R}$  defined by  $\varphi(p, q) = 2d(q - p, T_p M) / \|q - p\|^2$ . From (2.3), if there exists  $p \neq q \in M$  such that  $\varphi(p, q) = \tau_M^{-1}$ , then there exists  $z \in \text{Med}(M)$  with  $d(z, M) = \tau_M$ . Hence, for all  $p \neq q \in T_p M$ ,  $\varphi(p, q) < \tau_M^{-1}$ , and by compactness of  $M^2 \setminus \Delta_r$ , we have  $\sup_{M^2 \setminus \Delta_r} \varphi < \tau_M^{-1}$ . Since we have the decomposition

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

we get  $\sup_{\Delta_r \setminus \bar{\Delta}} \varphi = \tau_M^{-1}$ . Moreover, Lemma B.4 implies that

$$\left| \sup_{\substack{p \in M \\ v \in T_p M, \|v\|=1}} \|\gamma''_{p,v}(0)\| - \sup_{(p,q) \in \Delta_r \setminus \bar{\Delta}} \varphi(p, q) \right| \leq 3 \left( \frac{1}{\tau_M^2} + L \right) r$$

for  $r > 0$  small enough. Letting  $r$  go to zero yields

$$\sup_{\substack{p \in M \\ v \in T_p M, \|v\|=1}} \|\gamma''_{p,v}(0)\| = \frac{1}{\tau_M}.$$

Finally, the unit tangent bundle  $T^{(1)} M = \{(p, v), p \in M, v \in T_p M, \|v\| = 1\}$  being compact, there exists  $(q_0, v_0) \in T^{(1)} M$  such that  $\gamma_0 = \gamma_{p_0, v_0}$  satisfies  $\|\gamma''_0(0)\| = \tau_M^{-1}$ , which concludes the proof.  $\square$

## C Analysis of the Estimator

### C.1 Global Case

To show Proposition 4.2, we show a stronger result (Proposition C.1) that applies to a reach attaining pair with any size  $2\lambda$  (see Definition 3.1), meaning that it is not necessarily a bottleneck.

*Proof of Proposition 4.2.* Follows by applying Proposition C.1 with  $\lambda = \tau_M$ .  $\square$

**Proposition C.1.** *Let  $M \subset \mathbb{R}^D$  be a submanifold, and  $0 < \lambda \leq \tau_M$ . Assume that  $M$  has a reach attaining pair  $(q_1, q_2) \in M^2$  (see Definition 3.1) with  $\|q_1 - q_2\| \geq 2\lambda$ . Let  $\mathbb{X} \subset M$ . If there exists  $x, y \in \mathbb{X}$  with  $\|q_1 - x\| < \lambda$  and  $\|q_2 - y\| < \lambda$ , then*

$$0 \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x, y\})} \leq C_{\tau_M, \lambda} \max \{d_M(q_1, x), d_M(q_2, y)\},$$

where  $C_{\tau_M, \lambda} = \frac{2\tau_M^2 + 6\tau_M\lambda + \lambda^2}{2\tau_M^2\lambda^2}$  depends only on the parameters  $\tau_M, \lambda$ , and is a decreasing function of  $\tau_M$  and  $\lambda$  when the other parameter is fixed.

*Proof of Proposition C.1.* The two left hand inequalities are a direct consequence of Corollary 4.1, let us then focus on the third one.

Without loss of generality, assume that  $\|q_1 - q_2\| = 2\lambda$ . We set  $t$  to be equal to  $\max \{d_M(q_1, x), d_M(q_2, y)\}$ , and  $z_1 := x + (q_2 - q_1)$ . We have  $\|z_1 - x\| =$

$\|q_2 - q_1\| = 2\lambda$  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

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

Let now  $\theta := \angle(q_2 - q_1, T_{q_1} M) = \min_{v \in T_{q_1} M} \angle(q_2 - q_1, v)$ . Since  $z_0 \in \text{Med}(M)$ , with  $q_1, q_2 \in \mathcal{B}(z_0, \tau_M)$  and  $\|q_1 - q_2\| = 2\lambda$ , for any  $v'$  such that  $v' \perp z_0 - q_1$ , we have  $\angle(q_2 - q_1, v') \geq \frac{\pi}{2} - \angle(q_2 - q_1, z_0 - q_1)$ . Hence,  $\sin \theta \geq \frac{\lambda}{\tau_M}$  and  $\cos \theta \leq \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M}$ . Let  $v_1 \in T_{q_1} M$  be any point in  $T_{q_1} M$  realizing this angle, in the sense that  $\angle(q_2 - q_1, v_1) = \angle(q_2 - q_1, T_{q_1} M)$ . Then we have

$$\angle(z_1 - x, v_1) = \angle(q_2 - q_1, v_1) = \theta.$$

Let  $\bar{v}_1 \in T_x M$  be the parallel transport of  $v_1$  along the geodesic between  $q_1$  and  $x$ . Since  $M$  has reach  $\tau_M$ , Proposition A.1 (vi) gives

$$\angle(v_1, \bar{v}_1) \leq \frac{d_M(x, q_1)}{\tau_M} \leq \frac{t}{\tau_M}.$$

Hence the angle  $\angle(z_1 - x, T_x M)$  can be lower bounded as

$$\begin{aligned} \angle(z_1 - x, T_x M) &\geq \angle(z_1 - x, \bar{v}_1) \\ &\geq \angle(z_1 - x, v) - \angle(v, \bar{v}_1) \\ &\geq \theta - \frac{t}{\tau_M}. \end{aligned}$$

And  $0 \leq \frac{\lambda}{\tau_M} - \frac{t}{\tau_M} \leq \theta - \frac{t}{\tau_M} \leq \angle(z_1 - x, T_x M) \leq \frac{\pi}{2}$ , so the inequality is preserved by the sine function, i.e.

$$\begin{aligned} d(z_1 - x, T_x M) &= \|z_1 - x\| \sin(\angle(z_1 - x, T_x M)) \\ &\geq 2\lambda \sin\left(\theta - \frac{t}{\tau_M}\right) \\ &= 2\lambda \left( \sin \theta \cos \frac{t}{\tau_M} - \cos \theta \sin \frac{t}{\tau_M} \right) \\ &= \frac{2\lambda^2}{\tau_M} \cos \frac{t}{\tau_M} - \frac{2\lambda \sqrt{\tau_M^2 - \lambda^2}}{\tau_M} \sin \frac{t}{\tau_M}. \end{aligned}$$

Combining the previous bounds yields,

$$\begin{aligned} \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\{x, y\})} &\leq \frac{1}{\tau_M} - \frac{d(z_1 - x, T_x M) - 2t}{2(\lambda + t)^2} \\ &\leq \frac{1}{\tau_M} - \frac{\frac{1}{\tau_M} \cos \frac{t}{\tau_M} - \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M \lambda} \sin \frac{t}{\tau_M} - \frac{t}{\lambda^2}}{\left(1 + \frac{t}{\lambda}\right)^2}. \end{aligned}$$

Using again that  $t < \lambda \leq \tau_M$ , the latter right-hand side term is itself upper bounded by,

$$\begin{aligned}
& \frac{1}{\tau_M} - \left( \frac{1}{\tau_M} \left( 1 - \frac{t^2}{2\tau_M^2} \right) - \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M \lambda} \frac{t}{\tau_M} - \frac{t}{\lambda^2} \right) \left( 1 - \frac{2t}{\lambda} \right) \\
& \leq \left( \frac{\lambda}{2\tau_M^3} + \frac{\sqrt{\tau_M^2 - \lambda^2}}{\tau_M^2 \lambda} + \frac{1}{\lambda^2} + \frac{2}{\lambda \tau_M} \right) t \\
& = \frac{2\tau_M^3 + 2\lambda \tau_M \sqrt{\tau_M^2 - \lambda^2} + 4\tau_M^2 \lambda + \lambda^3}{2\tau_M^3 \lambda^2} t \\
& \leq \frac{2\tau_M^2 + 6\tau_M \lambda + \lambda^2}{2\tau_M^2 \lambda^2} t := C_{\tau_M, \lambda} t,
\end{aligned}$$

which is the announced result.  $\square$

As for Proposition 4.2, we tackle the proof of Proposition 4.3 by showing the following stronger one, Proposition C.2 that contains an extra parameter  $0 < \lambda \leq \tau_M$ .

*Proof of Proposition 4.3.* Follows by applying Proposition C.2 with  $\lambda = \tau_M$ .  $\square$

**Proposition C.2.** *Let  $P \in \mathcal{P}_{\tau_{\min}, L, f_{\min}}^{d, D}$ ,  $M = \text{supp}(P)$  and  $0 < \lambda \leq \tau_M$ . Assume that  $M$  has a reach attaining pair  $(q_1, q_2) \in M^2$  (see Definition 3.1) with  $\|q_1 - q_2\| \geq 2\lambda$ . Then*

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

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

*Proof of Proposition C.2.* Let  $Q$  be the distribution on  $\mathbb{R}^D$  associated to  $P$ . Let  $s < \frac{1}{\tau_M}$ ,  $C_{\tau_M, \lambda} = \frac{2\tau_M^2 + 6\tau_M \lambda + \lambda^2}{2\tau_M^2 \lambda^2}$ , and  $t = \frac{1}{C_{\tau_M, \lambda}} s \leq 2\tau_M/9$ . Let  $\omega_d := \mathcal{H}^d(\mathcal{B}_{\mathbb{R}^d}(0, 1))$  be the volume of the  $d$ -dimensional unit ball. Then note that from Proposition A.1 (v), for all  $q \in M$ ,

$$\begin{aligned}
Q(\mathcal{B}_M(p, t)) & \geq f_{\min} \mathcal{H}^d(\mathcal{B}_M(p, t)) \\
& \geq \omega_d f_{\min} \left( 1 - \left( \frac{t}{6\tau_M} \right)^2 \right)^d t^d \\
& \geq \omega_d f_{\min} \left( \frac{728}{729} \right)^d t^d.
\end{aligned}$$

Moreover, Proposition 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,

$$\begin{aligned} \mathbb{P} \left( \left| \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X}_n)} \right| > s \right) &\leq \mathbb{P}(\mathcal{B}_M(q_1, t) \cap \mathbb{X}_n = \emptyset) + \mathbb{P}(\mathcal{B}_M(q_2, t) \cap \mathbb{X}_n = \emptyset) \\ &\leq 2 \left( 1 - \omega_d f_{\min} \left( \frac{728}{729} \right)^d t^d \right)^n \\ &\leq 2 \exp \left( -n \omega_d f_{\min} \left( \frac{728}{729} \right)^d C_{\tau_M, \lambda}^{-d} s^d \right). \end{aligned}$$

The integration of the above bound gives

$$\begin{aligned} \mathbb{E}_{P^n} \left[ \left| \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X}_n)} \right|^p \right] &= \int_0^{\frac{1}{\tau_M^p}} \mathbb{P} \left( \left| \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X}_n)} \right|^p > s \right) ds \\ &\leq 2 \int_0^\infty \exp \left( -n \omega_d f_{\min} \left( \frac{728}{729} \right)^d C_{\tau_M, \lambda}^{-d} s^{\frac{d}{p}} \right) ds \\ &= \frac{2 \left( \frac{729}{728} \right)^{\frac{p}{d}} C_{\tau_M, \lambda}^p}{(n \omega_d f_{\min})^{\frac{p}{d}}} \int_0^\infty x^{\frac{p}{d}-1} e^{-x} dx \\ &:= C_{\tau_M, \lambda, f_{\min}, d, p} n^{-\frac{p}{d}}. \end{aligned}$$

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

## C.2 Local Case

**Lemma C.3.** *Let  $M$  be a submanifold and  $p \in M$ . Let  $v_0, v_1 \in T_p M$  be a unit tangent vector, and let  $\theta = \angle(v_0, v_1)$ . Let  $\gamma_{p,v}$  be the arc length parametrized geodesic starting from  $p$  with velocity  $v$ , and write  $\gamma_i = \gamma_{p,v_i}$  for  $i = 0, 1$ . Let  $\kappa_p = \max_{v \in \mathcal{B}_{T_p M}(0,1)} \|\gamma''_{p,v}(0)\|$ . Then,*

$$\|\gamma''_1(0)\| \geq \|\gamma''_0(0)\| - \frac{\sqrt{2}}{\sqrt{2}-1} \sin^2 \theta (\kappa_p + \|\gamma''_0(0)\|) - \frac{1}{\sqrt{2}-1} (\kappa_p - \|\gamma''_0(0)\|), \quad (\text{C.1})$$

and

$$\begin{aligned} \|\gamma''_1(0)\| &\geq \|\gamma''_0(0)\| - \sin^2 \theta (\kappa_p + \|\gamma''_0(0)\|) \\ &\quad - \frac{|\cos \theta \sin \theta| \kappa_p \sqrt{\kappa_p - \|\gamma''_0(0)\|}}{(\sqrt{2}-1) \|\gamma''_0(0)\|} \left( \frac{2\kappa_p}{\|\gamma''_0(0)\|} + 1 \right). \end{aligned} \quad (\text{C.2})$$

*Proof of Lemma C.3.* Let  $w \in T_p M$  be a unit vector satisfying  $w \perp v_0$  and  $v_1 = \cos \theta v_0 + \sin \theta w$ . For  $t \in \mathbb{R}$ , let  $v(t) := (\cos t)v_0 + (\sin t)w \in T_p M$ , so that  $v_1 = v(\theta)$ . Then

$$\begin{aligned} \|d_0^2 \exp_p(v(t), v(t))\| &= \|\cos^2 t d_0^2 \exp_p(v_0, v_0) + 2 \cos t \sin t d_0^2 \exp_p(v_0, w) \\ &\quad + \sin^2 t d_0^2 \exp_p(w, w)\| \\ &\geq |\cos t| \|\cos t d_0^2 \exp_p(v_0, v_0) + 2 \sin t d_0^2 \exp_p(v_0, w) \\ &\quad - \sin^2 t d_0^2 \exp_p(w, w)\|. \end{aligned} \quad (\text{C.3})$$

Now, note that when  $x \in [-1, 1]$ ,  $\sqrt{1+x} \geq 1+f(x)$ , where  $f(x) = \min\{x, (\sqrt{2}-1)x\}$ . Hence for any  $v', v'' \in T_p M$ ,

$$\begin{aligned} \|v' + v''\| &= \sqrt{\|v'\|^2 + \|v''\|^2} \sqrt{1 + \frac{2\langle v', v'' \rangle}{\|v'\|^2 + \|v''\|^2}} \\ &\geq \sqrt{\|v'\|^2 + \|v''\|^2} \left( 1 + f\left(\frac{2\langle v', v'' \rangle}{\|v'\|^2 + \|v''\|^2}\right) \right) \\ &\geq \|v'\| + f\left(\frac{2\langle v', v'' \rangle}{\sqrt{\|v'\|^2 + \|v''\|^2}}\right). \end{aligned}$$

Applying the latter inequality to (C.3) and using  $d_0^2 \exp_p(v_0, v_0) = \gamma_0''(0)$  together with  $\|d_0^2 \exp_p(w, w)\| \leq \kappa_p$  gives

$$\begin{aligned} &\|d_0^2 \exp_p(v(t), v(t))\| \\ &\geq \cos^2 t \|d_0^2 \exp_p(v_0, v_0)\| - \sin^2 t \|d_0^2 \exp_p(w, w)\| \\ &\quad + |\cos t| f\left(\frac{4 \cos t \sin t \langle d_0 \exp_p(v_0, v_0), d_0 \exp_p(v_0, w) \rangle}{\sqrt{\cos^2 t \|d_0^2 \exp_p(v_0, v_0)\|^2 + 4 \sin^2 t \|d_0^2 \exp_p(v_0, w)\|^2}}\right) \\ &\geq \cos^2 t \|\gamma_0''(0)\| - \kappa_p \sin^2 t \\ &\quad + |\cos t| f\left(\frac{4 \cos t \sin t \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle}{\sqrt{\cos^2 t \|\gamma_0''(0)\|^2 + 4 \sin^2 t \|d_0^2 \exp_p(v_0, w)\|^2}}\right). \end{aligned}$$

Now, note that  $f(x) \geq -|x|$  for  $x \in [-1, 1]$ , so applying this with  $t = \theta$  gives

$$\begin{aligned} \|\gamma_1''(0)\| &= \|d_0^2 \exp_p(v_1, v_1)\| \\ &\geq \cos^2 \theta \|\gamma_0''(0)\| - \sin^2 \theta \kappa_p \\ &\quad - \frac{4 |\cos^2 \theta \sin \theta \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle|}{\sqrt{\cos^2 \theta \|\gamma_0''(0)\|^2 + 4 \sin^2 \theta \|d_0^2 \exp_p(v_0, w)\|^2}}. \end{aligned} \quad (\text{C.4})$$

We now focus on the third term of the right-hand side. For this, note that either

$$t \sin t \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle \geq 0,$$

or

$$\cos(-t) \sin(-t) \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle \geq 0,$$

so that

$$\begin{aligned} \kappa_p &\geq \max\{\|d_0^2 \exp_p(v(-t), v(-t))\|, \|d_0^2 \exp_p(v(t), v(t))\|\} \\ &\geq \cos^2 t \|\gamma_0''(0)\| + \frac{4(\sqrt{2}-1) |\cos^2 t \sin t \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle|}{\sqrt{\cos^2 t \|\gamma_0''(0)\|^2 + 4 \sin^2 t \|d_0^2 \exp_p(v_0, w)\|^2}} \\ &\quad - \sin^2 t \kappa_p. \end{aligned}$$

As a consequence,

$$\begin{aligned}
& \frac{|\cos^2 t \sin t \langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle|}{\sqrt{\cos^2 t \|\gamma_0''(0)\|^2 + 4 \sin^2 t \|d_0^2 \exp_p(v_0, w)\|^2}} \\
& \leq \frac{1}{4(\sqrt{2}-1)} ((1 + \sin^2 t) \kappa_p - \cos^2 t \|\gamma_0''(0)\|) \\
& = \frac{1}{4(\sqrt{2}-1)} (\cos^2 t (\kappa_p - \|\gamma_0''(0)\|) + 2 \sin^2 t \kappa_p).
\end{aligned}$$

First, setting  $t = \theta$ , we derive

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

Furthermore, let  $t_0$  be defined by  $\sin^2 t_0 = 1 - \frac{\|\gamma_0''(0)\|}{\kappa_p} + \epsilon$  for  $\epsilon > 0$  small enough.

Then  $\sqrt{\cos^2 t_0 \|\gamma_0''(0)\|^2 + 4 \sin^2 t_0 \|d_0^2 \exp_p(v_0, w)\|^2} \leq \kappa_p$ , yielding

$$\begin{aligned}
& |\langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle| \\
& \leq \frac{\sqrt{\kappa_p}}{4(\sqrt{2}-1) \cos^2 t_0 |\sin t_0|} (\cos^2 t_0 (\kappa_p - \|\gamma_0''(0)\|) + 2 \sin^2 t_0 \kappa_p) \\
& = \frac{\kappa_p^{\frac{3}{2}}}{4(\sqrt{2}-1)} \left( \frac{1 - \frac{\|\gamma_0''(0)\|}{\kappa_p}}{\sqrt{1 - \frac{\|\gamma_0''(0)\|}{\kappa_p} + \epsilon}} + \frac{2\sqrt{1 - \frac{\|\gamma_0''(0)\|}{\kappa_p} + \epsilon}}{\frac{\|\gamma_0''(0)\|}{\kappa_p} - \epsilon} \right).
\end{aligned}$$

Sending  $\epsilon \rightarrow 0$ , we obtain

$$|\langle \gamma_0''(0), d_0 \exp_p(v_0, w) \rangle| \leq \frac{\kappa_p \sqrt{\kappa_p - \|\gamma_0''(0)\|}}{4(\sqrt{2}-1)} \left( \frac{2\kappa_p}{\|\gamma_0''(0)\|} + 1 \right).$$

Using the previous bound together with

$$\cos^2 \theta \|\gamma_0''(0)\|^2 + 4 \sin^2 \theta \|d_0^2 \exp_p(v_0, w)\|^2 \geq |\cos \theta| \|\gamma_0''(0)\|,$$

we finally obtain

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

□

*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  $\gamma_{x \rightarrow y}$ . The two left hand inequalities are a direct consequence of Corollary 4.1. Let us then focus on the

third one. Let  $t_0 := d_M(x, y)$ , and write  $\gamma = \gamma_{x \rightarrow y}$  for short. By definition of  $\hat{\tau}$  in (4.1),

$$\frac{1}{\hat{\tau}(\{x, y\})} \geq \frac{2d(y-x, T_x M)}{\|y-x\|^2} \geq \frac{2d(y-x, T_x M)}{t_0^2}. \quad (\text{C.5})$$

Let  $H_{\gamma''(0)} := \{x + u \in \mathbb{R}^D \mid \langle u, \gamma''_{x \rightarrow y}(0) \rangle = 0\}$  denote the affine hyperplane with normal vector  $\gamma''(0)$  that contain  $x$ . Since  $\gamma''(0) \in T_x M^\perp$ ,  $T_x M \subset H_{\gamma''(0)}$ . As a consequence,

$$d(y-x, T_x M) \geq d(y-x, H_{\gamma''(0)}) = \frac{|\langle \gamma''(0), y-x \rangle|}{\|\gamma''(0)\|}. \quad (\text{C.6})$$

Using the Taylor expansion of  $\gamma$  at order two, we get

$$y-x = \gamma(t_0) - \gamma(0) = t_0 \gamma'(0) + \int_0^{t_0} \int_0^t \gamma''(s) ds dt. \quad (\text{C.7})$$

Since  $\gamma$  is parametrized by arc length,  $\langle \gamma'(t), \gamma'(t) \rangle = 1$ . Differentiating this identity at 0 yields  $\langle \gamma''(0), \gamma'(0) \rangle = 0$ . In addition, by definition of  $\mathcal{M}_{\tau_{\min}, L}^{d,D} \ni M$  (Definition 2.4), the geodesic  $\gamma$  satisfies  $\|\gamma''(s) - \gamma''(0)\| \leq L|s|$ . Therefore,

$$\begin{aligned} |\langle \gamma''(0), \gamma''(s) \rangle| &= |\langle \gamma''(0), \gamma''(0) \rangle - \langle \gamma''(0), \gamma''(s) - \gamma''(0) \rangle| \\ &\geq \|\gamma''(0)\|^2 - L\|\gamma''(0)\||s|. \end{aligned}$$

Combining the above bound together with (C.5), (C.6) and (C.7), we derive

$$\frac{1}{\hat{\tau}(\{x, y\})} \geq \|\gamma''(0)\| - \frac{2}{3} L t_0,$$

which is the announced inequality.  $\square$

*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  $\theta := \angle(\gamma'_{x \rightarrow y}(0), \gamma'_{x \rightarrow q_0}(0)) = \pi - \angle(\gamma'_{x \rightarrow y}(0), \gamma'_{q_0 \rightarrow x}(t_x))$  (see Figure 6). From (C.1) in Lemma C.3,

$$\begin{aligned} \|\gamma''_{x \rightarrow y}(0)\| &\geq \|\gamma''_{q_0 \rightarrow x}(t_x)\| - \frac{\sqrt{2}}{\sqrt{2}-1} \sin^2 \theta (\kappa_x + \|\gamma''_{q_0 \rightarrow x}(t_x)\|) \\ &\quad - \frac{1}{\sqrt{2}-1} (\kappa_x - \|\gamma''_{q_0 \rightarrow x}(t_x)\|) \\ &= \frac{\sqrt{2}}{\sqrt{2}-1} \cos^2 \theta \|\gamma''_{q_0 \rightarrow x}(t_x)\| - \left( \frac{1}{\sqrt{2}-1} + \frac{\sqrt{2}}{\sqrt{2}-1} \sin^2 \theta \right) \kappa_x. \end{aligned} \quad (\text{C.8})$$

We now focus on the term  $\|\gamma''_{q_0 \rightarrow x}(t_x)\|$ . Since  $\theta_x = \angle(\gamma'_0(0), \gamma'_{q_0 \rightarrow x}(0))$ , applying (C.2) in Lemma C.3 yields

$$\|\gamma''_{q_0 \rightarrow x}(0)\| \geq (1 - 2 \sin^2 \theta_x) \kappa_{q_0},$$

and since  $\gamma''_{q_0 \rightarrow x}$  is  $L$ -Lipschitz,

$$\begin{aligned} \|\gamma''_{q_0 \rightarrow x}(t_x)\| &\geq \|\gamma''_{q_0 \rightarrow x}(0)\| - \|\gamma''_{q_0 \rightarrow x}(t_x) - \gamma''_{q_0 \rightarrow x}(0)\| \\ &\geq (1 - 2 \sin^2 \theta_x) \kappa_{q_0} - L t_x. \end{aligned} \quad (\text{C.9})$$



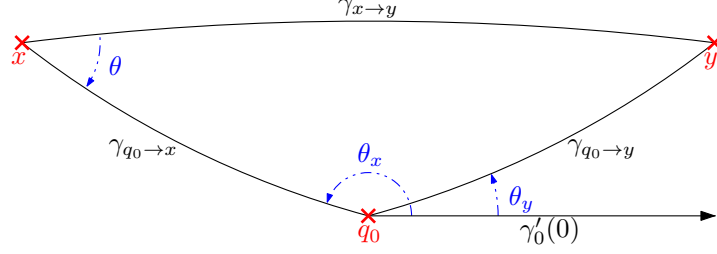


Figure 6: Layout of Lemma 4.5.

Now we focus on bounding the terms  $\sin^2 \theta$  and  $\cos^2 \theta$ . Let  $\mathcal{S}_{\tau_M}^2$  be a  $d$ -dimensional sphere of radius  $\tau_M$ . In what follows, for short,  $\angle abc$  stands for  $\angle(\gamma'_{b \rightarrow a}(0), \gamma'_{b \rightarrow c}(0))$ . First, let  $\tilde{q}_0, \tilde{x}, \tilde{y} \in \mathcal{S}_{\tau_M}^2$  be such that  $d_{\mathcal{S}_{\tau_M}^2}(\tilde{q}_0, \tilde{x}) = d_M(q_0, x)$ ,  $d_{\mathcal{S}_{\tau_M}^2}(\tilde{q}_0, \tilde{y}) = d_M(q_0, y)$ , and  $\angle \tilde{x} \tilde{q}_0 \tilde{y} = \angle x q_0 y$ . Then from Toponogov's comparison theorem (see [28]), we have  $d_{\mathcal{S}_{\tau_M}^2}(\tilde{x}, \tilde{y}) \leq d_M(x, y)$ . Moreover, the spherical law of cosines [9, Proposition 18.6.8] yields

$$\cos\left(\frac{d_{\mathcal{S}_{\tau_M}^2}(\tilde{x}, \tilde{y})}{\tau_M}\right) = \cos\left(\frac{t_x}{\tau_M}\right) \cos\left(\frac{t_y}{\tau_M}\right) + \sin\left(\frac{t_x}{\tau_M}\right) \sin\left(\frac{t_y}{\tau_M}\right) \cos(\angle \tilde{x} \tilde{q}_0 \tilde{y}),$$

and since  $t_x, t_y \leq \frac{\pi}{2}$  and  $\cos(\cdot)$  is decreasing on  $[0, \pi]$ , we get

$$t_y \leq d_{\mathcal{S}_{\tau_M}^2}(\tilde{x}, \tilde{y}) \leq d_M(x, y).$$

Now, let  $\bar{q}_0, \bar{x}, \bar{y} \in \mathcal{S}_{\tau_M}^2$  be such that  $d_{\mathcal{S}_{\tau_M}^2}(\bar{q}_0, \bar{x}) = d_M(q_0, x)$ ,  $d_{\mathcal{S}_{\tau_M}^2}(\bar{q}_0, \bar{y}) = d_M(q_0, y)$ , and  $d_{\mathcal{S}_{\tau_M}^2}(\bar{x}, \bar{y}) = d_M(x, y)$ . Applying Toponogov's comparison theorem (see [28]), we have  $\angle q_0 x y \leq \angle \bar{q}_0 \bar{x} \bar{y}$  and  $\angle x q_0 y \leq \angle \bar{x} \bar{q}_0 \bar{y}$ , and from the spherical law of cosines [9, Proposition 18.6.8],

$$\cos(\angle \bar{q}_0 \bar{x} \bar{y}) = \frac{\cos\left(\frac{t_y}{\tau_M}\right) - \cos\left(\frac{t_x}{\tau_M}\right) \cos\left(\frac{d_M(x, y)}{\tau_M}\right)}{\sin\left(\frac{t_x}{\tau_M}\right) \sin\left(\frac{d_M(x, y)}{\tau_M}\right)} \geq 0,$$

so that  $\angle q_0 x y \leq \angle \bar{q}_0 \bar{x} \bar{y} \leq \frac{\pi}{2}$ . Also,  $\angle x q_0 y \geq |\theta_x - \theta_y| \geq \frac{\pi}{2}$  yields  $\frac{\pi}{2} \leq \angle x q_0 y \leq \angle \bar{x} \bar{q}_0 \bar{y}$ , and  $\theta = \angle(\gamma'_{x \rightarrow y}(0), \gamma'_{q_0 \rightarrow x}(t_x)) = \pi - \angle q_0 x y$ . Hence applying the spherical law of sines and cosines [9, Proposition 18.6.8] yields

$$\begin{aligned} \sin \theta &= \sin(\angle q_0 x y) \leq \sin(\angle \bar{q}_0 \bar{x} \bar{y}) \\ &= \frac{\sin\left(\frac{t_y}{\tau_M}\right) \sin(\angle \bar{x} \bar{q}_0 \bar{y})}{\sqrt{1 - \left(\cos\left(\frac{t_x}{\tau_M}\right) \cos\left(\frac{t_y}{\tau_M}\right) + \sin\left(\frac{t_x}{\tau_M}\right) \sin\left(\frac{t_y}{\tau_M}\right) \cos(\angle \bar{x} \bar{q}_0 \bar{y})\right)^2}} \\ &\leq \frac{\sin\left(\frac{t_y}{\tau_M}\right) \sin(\angle \bar{x} \bar{q}_0 \bar{y})}{\sqrt{1 - \cos^2\left(\frac{t_x}{\tau_M}\right) \cos^2\left(\frac{t_y}{\tau_M}\right)}} \\ &= \frac{\sin\left(\frac{t_y}{\tau_M}\right) \sin(\angle \bar{x} \bar{q}_0 \bar{y})}{\sqrt{\sin^2\left(\frac{t_y}{\tau_M}\right) + \sin^2\left(\frac{t_x}{\tau_M}\right) \cos^2\left(\frac{t_y}{\tau_M}\right)}} \\ &\leq \sin(\angle \bar{x} \bar{q}_0 \bar{y}) \leq \sin(\angle x q_0 y) \leq \sin(|\theta_x - \theta_y|). \end{aligned} \tag{C.10}$$

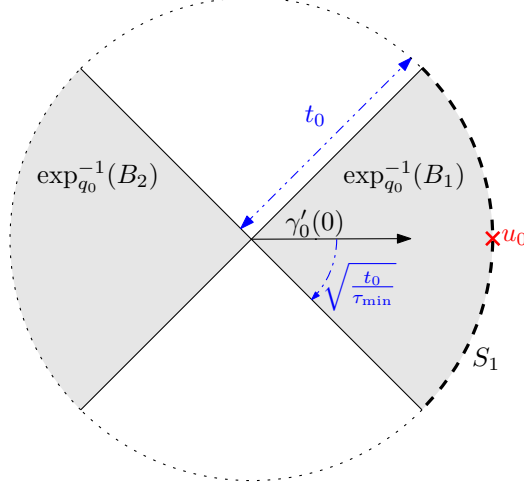


Figure 7: Layout of the proof of Proposition 4.7.

And accordingly,

$$|\cos \theta| = \sqrt{1 - \sin^2 \theta} \geq \sqrt{1 - \sin^2(|\theta_x - \theta_y|)} = |\cos(|\theta_x - \theta_y|)|. \quad (\text{C.11})$$

Hence applying (C.9), (C.10), and (C.11) to (C.8) gives

$$\begin{aligned} & \|\gamma''_{x \rightarrow y}(0)\| \\ & \geq \frac{\sqrt{2}}{\sqrt{2}-1} \cos^2(|\theta_x - \theta_y|) ((1 - 2 \sin^2 \theta_x) \kappa_{q_0} - Lt_x) - \left( \frac{1}{\sqrt{2}-1} + \frac{\sqrt{2}}{\sqrt{2}-1} \sin^2(|\theta_x - \theta_y|) \right) \kappa_x \\ & = \frac{(\sqrt{2} \kappa_{q_0} - \kappa_x)}{\sqrt{2}-1} - \frac{\sqrt{2}}{\sqrt{2}-1} ((\kappa_{q_0} + \kappa_x) \sin^2(|\theta_x - \theta_y|) + 2 \kappa_{q_0} \sin^2 \theta_x \cos^2(|\theta_x - \theta_y|)) \\ & \quad - \frac{\sqrt{2}}{\sqrt{2}-1} Lt_x \cos^2(\theta_x + \theta_y) \\ & \geq \kappa_{q_0} - \frac{1}{\sqrt{2}-1} \left( \kappa_x - \kappa_{q_0} + \sqrt{2}(3 \kappa_{q_0} + \kappa_x) \sin^2(|\theta_x - \theta_y|) + \sqrt{2} Lt_x \right). \end{aligned}$$

□

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

$$\begin{aligned} B_1 &:= \exp_{q_0} \left( \left\{ v \in T_{q_0} M : \|v\| \leq t_0, \angle(\gamma'_0(0), v) \leq \sqrt{\frac{t_0}{\tau_{\min}}} \right\} \right), \\ B_2 &:= \exp_{q_0} \left( \left\{ v \in T_{q_0} M : \|v\| \leq t_0, \angle(\gamma'_0(0), v) \geq \pi - \sqrt{\frac{t_0}{\tau_{\min}}} \right\} \right), \end{aligned}$$

and  $B_0 := B_1 \cup B_2$  (see Figure 7). Let  $\mathbb{X} \subset M$ , and  $x, y \in \mathbb{X}$  be such that  $x \in B_1$ ,  $y \in B_2$ . Writing  $\theta_x := \angle(\gamma'_0(0), \gamma'_{q_0 \rightarrow x}(0))$  and  $\theta_y := \angle(\gamma'_0(0), \gamma'_{q_0 \rightarrow y}(0))$ , 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 that

$$\begin{aligned}
0 &\leq \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \\
&\leq \frac{4\sqrt{2}\sin^2(|\theta_x - \theta_y|)}{(\sqrt{2}-1)\tau_M} + L \left( \frac{2}{3}d_M(x, y) + \frac{\sqrt{2}}{\sqrt{2}-1}d_M(q_0, x) \right) \\
&\leq \left( \frac{16\sqrt{2}}{(\sqrt{2}-1)\tau_{\min}\tau_M} + \frac{(7\sqrt{2}-4)L}{3(\sqrt{2}-1)} \right) t_0.
\end{aligned}$$

A symmetric argument also applies when  $x \in B_2$  and  $y \in B_1$ . Now, for any  $s < \frac{1}{\tau_M}$ , let  $t_0(s) := \left( \frac{16\sqrt{2}}{(\sqrt{2}-1)\tau_{\min}^2} + \frac{(7\sqrt{2}-4)L}{3(\sqrt{2}-1)} \right)^{-1} s < \frac{\tau_{\min}}{10}$ . The above argument implies that if  $\left| \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X})} \right| > s$ , then for any  $x, y \in \mathbb{X} \cap B_0$ , one has either  $x, y \in B_1$  or  $x, y \in B_2$ . Hence,

$$\begin{aligned}
&\mathbb{P} \left( \left| \frac{1}{\tau_M} - \frac{1}{\hat{\tau}(\mathbb{X}_n)} \right| > s \right) \\
&\leq \sum_{m=0}^n \binom{n}{m} \left\{ \mathbb{P}(X_1, \dots, X_m \in M \setminus B_0, X_{m+1}, \dots, X_n \in B_1) \right. \\
&\quad \left. + \mathbb{P}(X_1, \dots, X_m \in M \setminus B_0, X_{m+1}, \dots, X_n \in B_2) \right\} \\
&= \sum_{m=0}^n \binom{n}{m} \left\{ (1 - Q(B_0))^m Q(B_1)^{n-m} + (1 - Q(B_0))^m Q(B_2)^{n-m} \right\} \\
&\leq (1 - Q(B_2))^n + (1 - Q(B_1))^n. \tag{C.12}
\end{aligned}$$

Let us derive lower bounds for  $Q(B_1)$  and  $Q(B_2)$ . For this purpose, let  $S_1 := \exp_{q_0}^{-1}(B_1) \cap \partial \mathcal{B}_{T_{q_0}M}(0, t_0)$  (see Figure 7). Then  $\exp_{q_0}^{-1}(B_1) \subset \mathcal{B}_{T_{q_0}M}(0, t_0)$  is a cone satisfying

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

Let  $\omega_d := \mathcal{H}^d(\mathcal{B}_{\mathbb{R}^d}(0, 1))$  and  $\sigma_d := \mathcal{H}^d(\partial \mathcal{B}_{\mathbb{R}^{d+1}}(0, 1))$  be the volumes of the  $d$ -dimensional unit ball and the unit sphere respectively. Then by homogeneity,  $\mathcal{H}^d(\mathcal{B}_{T_{q_0}M}(0, t_0)) = \omega_d t_0^d$  and  $\mathcal{H}^{d-1}(\partial \mathcal{B}_{T_{q_0}M}(0, t_0)) = \sigma_{d-1} t_0^{d-1}$ . To derive a lower bound on  $\mathcal{H}^{d-1}(S_1)$ , consider  $u_0 := t_0 \gamma'_0(0) \in S_1$ . Since  $\tau_{S_1} = t_0$  and  $\exp_{u_0}^{-1}(S_1) \subset \mathcal{B}_{T_{u_0}S_1}\left(0, \tau_{\min}^{-\frac{1}{2}} t_0^{\frac{3}{2}}\right)$ , applying Proposition A.1 (v) yields

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

and hence

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

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

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

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.12) gives

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

As a consequence, by integration,

$$\begin{aligned} \mathbb{E}_{P^n} \left[ \left| \frac{1}{\hat{\tau}(\mathbb{X}_n)} - \frac{1}{\tau_M} \right|^p \right] &= \int_0^{\frac{1}{\tau_M^p}} \mathbb{P}\left(\left|\frac{1}{\hat{\tau}(\mathbb{X}_n)} - \frac{1}{\tau_M}\right|^p > s\right) ds \\ &\leq 2 \int_0^\infty \exp\left(-C_{\tau_{\min}, d, L, f_{\min}} n s^{\frac{3d-1}{2p}}\right) ds \\ &= 2 (C_{\tau_{\min}, d, L, f_{\min}} n)^{-\frac{2p}{3d-1}} \int_0^\infty x^{\frac{2p}{3d-1}} e^{-x} dx \\ &:= C_{\tau_{\min}, d, L, f_{\min}, p} n^{-\frac{2p}{3d-1}}. \end{aligned}$$

□

## D Minimax Lower Bounds

### D.1 Stability of the model with respect to diffeomorphism

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 [21]). *Let  $A \subset \mathbb{R}^D$  with  $\tau_A \geq \tau_{\min} > 0$  and  $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^D$  is a  $\mathcal{C}^1$ -diffeomorphism such that  $\Phi, \Phi^{-1}$ , 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  $\Phi$ . Since  $\Phi$  is a global diffeomorphism,  $M'$  is a closed submanifold of dimension one. Moreover,  $\Phi$  is  $\|d\Phi\|_{op} \leq (1 + \|d\Phi - I_D\|_{op})$ -Lipschitz,  $\Phi^{-1}$  is  $\|d\Phi^{-1}\|_{op} \leq (1 - \|d\Phi - I_D\|_{op})^{-1}$ -Lipschitz, and  $d\Phi$  is  $\|d^2\Phi\|_{op}$ -Lipschitz. From Lemma 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  $\|d^2\Phi\|_{op} \tau_{min} \leq 1/2$  and  $\|d\Phi - I_D\|_{op} \leq 0.1$ . All that remains to be proved now is the bound on the third order derivative of the geodesics of  $M'$ . We denote by  $\gamma$  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  $M \in \mathcal{M}_{\tau_{min}, L}^{d, D}$  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  $\Psi_p : \mathcal{B}_{\mathbb{R}^d}(0, \varepsilon) \rightarrow M$  with  $\Psi_p(0) = p$ . For a smooth curve  $\gamma$  on  $M$  nearby  $p$ , we let  $c = (c_1, \dots, c_d)^t$  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 [19] p.62). That is, the second order ordinary differential equation

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

where  $\Gamma^\ell = (\Gamma_{i,j}^\ell)_{1 \leq i,j \leq d}$  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  $\mathbf{y} = \Psi_{p'}^{-1} = \Psi_p^{-1} \circ \Phi^{-1}$  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  $\tilde{c} = \Psi_{p'}^{-1}(\tilde{\gamma})$  satisfies (D.1) with  $\Gamma^\ell$  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  $\Gamma$ ,  $d\Phi$ , 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  $\|I_D - d\Phi\|_{op}$ ,  $\|d^2\Phi\|_{op}$  and  $\|d^3\Phi\|_{op}$  small enough,  $\|c'''(0) - \tilde{c}'''(0)\|$  can be made arbitrarily small. In particular,  $\tilde{\gamma}'''(0)$  gets arbitrarily close to  $\gamma'''(0)$ , so that  $\|\tilde{\gamma}'''(0)\| \leq \|\gamma'''(0)\| + L \leq 2L$ , which concludes the proof.  $\square$

## D.2 Some Lemmas on the Total Variation Distance

Prior to any actual construction, we show this straightforward lemma bounding the total variation between uniform distribution on manifolds that are perturbations of each other. For  $M \subset \mathbb{R}^D$ , write  $\lambda_M = \mathbb{1}_M / \mathcal{H}^d(M)$  for the uniform probability distribution on  $M$ .

**Lemma D.2.** *Let  $M \subset \mathbb{R}^D$  be a  $d$ -dimensional submanifold and  $B \subset \mathbb{R}^D$  be a Borel set. Let  $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^D$  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  $TV(\lambda_M, \lambda_{\Phi(M)}) \leq 12\lambda_M(B)$ .*

*Proof of Lemma D.2.* Since  $\Phi$  is  $(1 + \|d\Phi - I_D\|_{op})$ -Lipschitz, Lemma 7 of [5] 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  $\Delta$  for the symmetric difference of sets, we have  $M \Delta \Phi(M) = (B \cap M) \Delta (B \cap \Phi(M)) \subset (B \cap M) \cup (B \cap \Phi(M))$ . Therefore, Lemma 7 in [5] yields,

$$\begin{aligned} TV(\lambda_M, \lambda_{\Phi(M)}) &\leq 4 \frac{\mathcal{H}^d(M \Delta \Phi(M))}{\mathcal{H}^d(M \cup \Phi(M))} \\ &\leq 4 \frac{\mathcal{H}^d(M \cap B) + \mathcal{H}^d(\Phi(M) \cap B)}{\mathcal{H}^d(M)} \\ &= 4 \frac{\mathcal{H}^d(M \cap B) + \mathcal{H}^d(\Phi(M \cap B))}{\mathcal{H}^d(M)} \\ &\leq 12 \frac{\mathcal{H}^d(M \cap B)}{\mathcal{H}^d(M)} = 12\lambda_M(B). \end{aligned}$$

□

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  $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$  be  $\mathcal{C}^1$  and  $x \in \mathbb{R}^d$  be such that  $g(x) = 0$  and  $d_x g \neq 0$ . Then there exists  $r > 0$  such that  $\mathcal{H}^d(g^{-1}(0) \cap \mathcal{B}(x, r)) = 0$ .*

*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  $g = (g_1, \dots, g_k)$ , assume without loss of generality that  $\partial_{x_1} g_1 \neq 0$ . Since  $g_1 : \mathbb{R}^d \rightarrow \mathbb{R}$  is nonsingular at  $x$ , the implicit function theorem asserts that  $g_1^{-1}(0)$  is a submanifold of dimension  $d - 1$  of  $\mathbb{R}^d$  in a neighborhood of  $x \in \mathbb{R}^d$ . Therefore, for  $r > 0$  small enough,  $g_1^{-1}(0) \cap \mathcal{B}(x, r)$  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  $g : \mathcal{B}_{\mathbb{R}^d}(0, r') \rightarrow \mathbb{R}^k$  for  $r' > 0$  small enough. The assumption  $T_x M \neq T_x M'$  translates to  $d_0 g \neq 0$ , and the previous transformation maps smoothly  $M \cap M' \cap \mathcal{B}(x, r'')$  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  $\mu = \llcorner_{M \cup M'} \mathcal{H}^d$ , with  $dQ(x) = f(x)d\mu(x)$  and  $dQ'(x) = f'(x)d\mu(x)$ , where  $f, f' : \mathbb{R}^D \rightarrow \mathbb{R}_+$  have support  $M$  and  $M'$  respectively. On the other hand,  $P$  and  $P'$  are dominated by  $\nu(dx dT) = \delta_{\{T_x M, T_x M'\}}(dT) \mu(dx)$  with respective densities  $\bar{f}(x, T) = \llcorner_{T=T_x M} f(x)$  and  $\bar{f}'(x, T) = \llcorner_{T=T_x M'} f'(x)$ , where we set arbitrarily  $T_x M = T_0$  for  $x \notin M$ , and  $T_x M' = T_0$  for  $x \notin M'$ . Recalling that  $f$  vanishes

outside  $M$  and  $f'$  outside  $M'$ ,

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

From Corollary D.4 and a straightforward compactness argument, we derive that

$$\mathcal{H}^d(M \cap M' \cap \{x | T_x M \neq T_x M'\}) = 0.$$

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.  $\square$

### 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  $\ell \leq \frac{R}{2} \wedge (2^{1/d} - 1)$  and  $\eta \leq \frac{\ell^2}{2R}$ . Then there exists a  $d$ -dimensional sphere of radius  $R$  that we call  $M$ , such that  $M \in \mathcal{M}_{R, \frac{1}{R^2}}^{d,D}$  and a global  $\mathcal{C}^\infty$ -diffeomorphism  $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^D$  such that,*

$$\|d\Phi - I_D\|_{op} \leq \frac{3\eta}{\ell}, \quad \|d^2\Phi\|_{op} \leq \frac{23\eta}{\ell^2}, \quad \|d^3\Phi\|_{op} \leq \frac{573\eta}{\ell^3},$$

and so that writing  $M' = \Phi(M)$ , we have  $\mathcal{H}^d(M') \leq 2\mathcal{H}^d(M) = 2\sigma_d R^d$ ,

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

*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  $M \in \mathcal{M}_{R, \frac{1}{R^2}}^{d,D}$ . Let  $\phi : \mathbb{R}^D \rightarrow \mathbb{R}_+$

be the map defined by  $\phi(x) = \exp(\frac{\|x\|^2}{\|x\|^2 - 1}) \mathbb{K}_{\|x\|^2 < 1}$ .  $\phi$  is a symmetric  $\mathcal{C}^\infty$  map with support equal to  $\mathcal{B}(0, 1)$  and elementary real analysis yields  $\phi(0) = 1$ ,  $\|d\phi\|_{op} \leq 3$ ,  $\|d^2\phi\|_{op} \leq 23$  and  $\|d^3\phi\|_{op} \leq 573$ . Let  $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^D$  be defined by

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

where  $v = (0, 1, 0, \dots, 0)$  is the unit vertical vector.  $\Phi$  is the identity map on  $\mathcal{B}(0, \ell)^c$ , and in  $\mathcal{B}(0, \ell)$ ,  $\Phi$  translates points on the vertical axis with a magnitude modulated by the weight function  $\phi(x/\ell)$ . From chain rule,  $\|d\Phi - I_D\|_{op} = \eta \|d\phi\|_{op} / \ell \leq 3\eta / \ell < 1$ . Therefore,  $d_x \Phi$  is invertible for all  $x \in \mathbb{R}^D$ , so that  $\Phi$  is a local  $\mathcal{C}^\infty$ -diffeomorphism according to the local inverse function theorem. Moreover,  $\|\Phi(x)\| \rightarrow \infty$  as  $\|x\| \rightarrow \infty$ , so that  $\Phi$  is a global  $\mathcal{C}^\infty$ -diffeomorphism

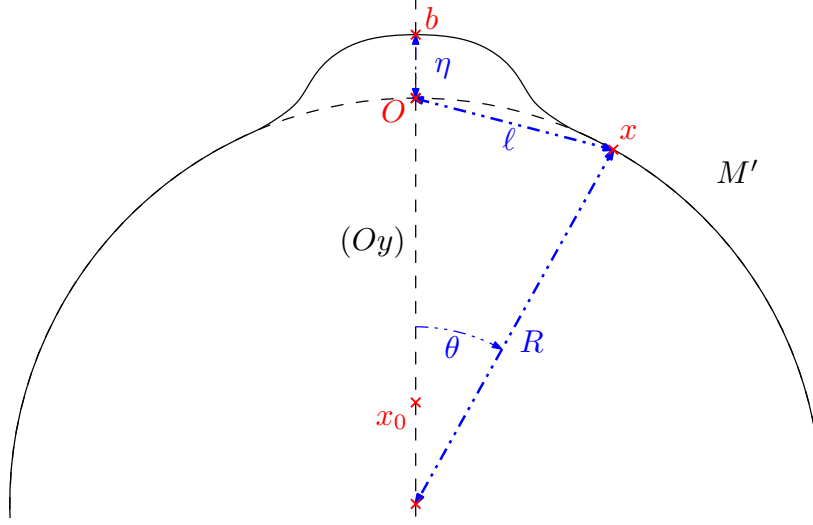


Figure 8: The bumped sphere  $M'$  of Lemma D.5.

by Hadamard-Cacciopoli theorem [17]. Similarly, from bounds on differentials of  $\phi$  we get

$$\|d^2\Phi\|_{op} \leq 23 \frac{\eta}{\ell^2} \quad \text{and} \quad \|d^3\Phi\|_{op} \leq 573 \frac{\eta}{\ell^3}.$$

Let us now write  $M' = \Phi(M)$  for the image of  $M$  by the map  $\Phi$  (see Figure 8). Denote by  $(Oy)$  the vertical axis  $\text{span}(v)$ , and notice that since  $\phi$  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  $\text{Med}(M')$  (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  $\text{Med}(M')$  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

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

which concludes the proof.  $\square$

*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  $\mathcal{M}_{2\tau_{min}, 1/(4\tau_{min}^2)}^{d,D}$ . Furthermore, taking  $\eta = c_d \ell^3 / \tau_{min}^2$  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  $\mathcal{M}_{\tau_{min}, 1/(2\tau_{min}^2)}^{d,D} \subset \mathcal{M}_{\tau_{min}, L}^{d,D}$ , since  $L \geq 1/(2\tau_{min}^2)$ . Moreover,

$$\mathcal{H}^d(M')^{-1} \wedge \mathcal{H}^d(M)^{-1} \geq (2^{d+1} \sigma_d \tau_{min}^d)^{-1} \geq f_{min},$$

so that  $\lambda_M, \lambda_{M'} \in \mathcal{Q}_{\tau_{min}, L, f_{min}}^{d,D}$ , which gives the result.  $\square$

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  $\|d^2\Phi\|_{op} \leq 23\eta/\ell^2 \leq 1/(2\tau_{min})$ . 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  $\mathcal{M}_{\tau_{min}, L=\infty}^{d,D}$ . Furthermore, since we have  $f_{min} \leq (2^{d+1} \tau_{min}^d \sigma_d)^{-1} \leq \mathcal{H}^d(M)^{-1} \wedge \mathcal{H}^d(M')^{-1}$ , we see that the uniform distributions  $\lambda_M, \lambda_{M'}$  belong to  $\mathcal{Q}_{\tau_{min}, L=\infty, f_{min}}^{d,D}$ . Let now  $P, P'$  denote the distributions of  $\mathcal{P}_{\tau_{min}, L=\infty, f_{min}}^{d,D}$  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  $\ell$  small enough,

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

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

## 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. For all  $x, y \in \mathbb{X}$  with  $x \neq y$ ,

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

□

## References

- [1] E. Aamari and C. Levrard. Stability and Minimax Optimality of Tangential Delaunay Complexes for Manifold Reconstruction. *ArXiv e-prints*, 2015.
- [2] E. Aamari and C. Levrard. Non-Asymptotic Rates for Manifold, Tangent Space, and Curvature Estimation. *ArXiv e-prints*, May 2017.
- [3] Stephanie B. Alexander and Richard L. Bishop. Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geom. Dedicata*, 117:65–84, 2006.
- [4] E. Arias-Castro, B. Pateiro-López, and A. Rodríguez-Casal. Minimax Estimation of the Volume of a Set with Smooth Boundary. *ArXiv e-prints*, 2016.
- [5] Ery Arias-Castro, Gilad Lerman, and Teng Zhang. Spectral clustering based on local PCA. *J. Mach. Learn. Res.*, 18:Paper No. 9, 57, 2017.
- [6] Dominique Attali, Jean-Daniel Boissonnat, and Herbert Edelsbrunner. 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., pages 109–125. Springer, Berlin, 2009.
- [7] Sivaraman Balakrishnan, Alessandro Rinaldo, Don Sheehy, Aarti Singh, and Larry A Wasserman. Minimax rates for homology inference. In *International Conference on Artificial Intelligence and Statistics*, pages 64–72, 2012.
- [8] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.*, 7:2399–2434, 2006.
- [9] Marcel Berger. *Geometry. II*. Universitext. Springer-Verlag, Berlin, 1987. Translated from the French by M. Cole and S. Levy.
- [10] Jean-Daniel Boissonnat and Arijit Ghosh. Manifold reconstruction using tangential Delaunay complexes. *Discrete Comput. Geom.*, 51(1):221–267, 2014.

- [11] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.
- [12] Frédéric Chazal and André Lieutier. The  $\lambda$ -medial axis. *J. Graphical Models*, 67:304–331, 2005.
- [13] Siu-Wing Cheng and Man-Kwun Chiu. Tangent estimation from point samples. *Discrete Comput. Geom.*, 56(3):505–557, 2016.
- [14] A. Cuevas, R. Fraiman, and B. Pateiro-López. On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.*, 44(2):311–329, 2012.
- [15] Antonio Cuevas, Ricardo Fraiman, and Alberto Rodríguez-Casal. A non-parametric approach to the estimation of lengths and surface areas. *Ann. Statist.*, 35(3):1031–1051, 2007.
- [16] Antonio Cuevas, Pamela Llop, and Beatriz Pateiro-López. On the estimation of the medial axis and inner parallel body. *J. Multivariate Anal.*, 129:171–185, 2014.
- [17] Giuseppe De Marco, Gianluca Gorni, and Gaetano Zampieri. Global inversion of functions: an introduction. *NoDEA Nonlinear Differential Equations Appl.*, 1(3):229–248, 1994.
- [18] Tamal K. Dey and Jian Sun. Normal and feature approximations from noisy point clouds. In *FSTTCS 2006: Foundations of software technology and theoretical computer science*, volume 4337 of *Lecture Notes in Comput. Sci.*, pages 21–32. Springer, Berlin, 2006.
- [19] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.
- [20] Ramsay Dyer, Gert Vegter, and Mathijs Wintraecken. Riemannian simplices and triangulations. *Geometriae Dedicata*, 179(1):91–138, 2015.
- [21] Herbert Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.
- [22] Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.
- [23] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *J. Amer. Math. Soc.*, 29(4):983–1049, 2016.
- [24] Christopher R. Genovese, Marco Perone-Pacífico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *J. Mach. Learn. Res.*, 13:1263–1291, 2012.

- [25] Evarist Giné and Vladimir Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 238–259. Inst. Math. Statist., Beachwood, OH, 2006.
- [26] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.
- [27] S. Kanagawa, Y. Mochizuki, and H. Tanaka. 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(1):121–131, 1992.
- [28] Hermann Karcher. Riemannian comparison constructions. In *Global differential geometry*, volume 27 of *MAA Stud. Math.*, pages 170–222. Math. Assoc. America, Washington, DC, 1989.
- [29] Arlene K. H. Kim and Harrison H. Zhou. Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.*, 9(1):1562–1582, 2015.
- [30] James R. Munkres. *Topology: a first course*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.
- [31] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008.
- [32] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.*, 65(8):1067–1144, 2012.
- [33] Christoph Thäle. 50 years sets with positive reach—a survey. *Surv. Math. Appl.*, 3:123–165, 2008.
- [34] Bin Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, New York, 1997.