Optimal Reach Estimation and Metric Learning

Eddie Aamari, Clément Berenfeld, Clément Levrard

Optimal Reach Estimation and Metric Learning

Eddie Aamari* Clément Berenfeld† Clément Levrard‡

Abstract

We study the estimation of the *reach*, an ubiquitous regularity parameter in manifold estimation and geometric data analysis. Given an i.i.d. sample over an unknown d-dimensional $\mathcal{C}^k$-smooth submanifold of $\mathbb{R}^D$, we provide optimal nonasymptotic bounds for the estimation of its reach. We build upon a formulation of the reach in terms of maximal curvature on one hand, and geodesic metric distortion on the other hand. The derived rates are adaptive, with rates depending on whether the reach of M arises from curvature or from a bottleneck structure. In the process, we derive optimal geodesic metric estimation bounds.

1 Introduction

1.1 Geometric Inference

Topological data analysis and geometric methods now constitute a standard toolbox in statistics and machine learning [44, 20]. In this family of methods, data $\mathbb{X}_n := \{X_1, \dots, X_n\}$ are usually seen as point clouds in high dimension, for which complex structural correlations give rise to an underlying structure that is neither full-dimensional, nor even linear. Dealing with non-linearity is very well understood through the prism of *non-parametric regression*. However, in absence of distinguished "covariate" and "response" variables (i.e. coordinates), regression does not make sense anymore. Hence, one needs to adopt a more global and coordinate-free approach: data are naturally viewed as lying on a submanifold $M \subset \mathbb{R}^D$ of dimension $d \ll D$, where d corresponds to its *true* number of degrees of freedom.

This approach opens the way to the estimation of numerous geometric and topological quantities to describe data. Central to it is the manifold itself [30, 29, 35, 28, 25, 4, 40], where error is most commonly measured in Hausdorff distance. Among many others, let us also mention the homology [9], persistent homology [18], differential quantities [3], intrinsic metric [5] and regularity [2].

1.2 Reach and Regularity

Similarly to functional estimation, the theoretical study of nonparametric geometric problems naturally comes with regularity conditions. By far, the most ubiquitous regularity and scale parameter in this context is the *reach*. First introduced by H. Federer's seminal paper [26] on geometric measure theory, the reach $\mathrm{rch}(K) \in \mathbb{R}_+$ of a set $K \subset \mathbb{R}^D$ measures how far K is from being convex [8]. It hence provides a typical scale at which it shares most of the properties of a convex set. These properties include – among others – uniqueness of the projection map, contractibility of balls, and

*CNRS & U. Paris & Sorbonne U., Paris, France (https://www.lpsm.paris/~aamari/)
†CNRS & U. Paris-Dauphine & PSL, Paris, France (https://www.ceremade.dauphine.fr/~berenfeld/)
‡CNRS & U. Paris & Sorbonne U., Paris, France (http://www.normalesup.org/~levrard/)

explicit formulas for the volume of thickenings (see [26]). When $K = M$ is a submanifold, the reach also assesses quantitatively how it deviates from its tangent spaces. Therefore, the reach also provides an upper bound on curvature (that is, a bound in $\mathcal{C}^2$) and a minimal scale of possible quasi self-intersections [3].

For all these reasons, the reach practically appears in *all* geometric inference methods as a natural scale parameter, which either drives a bandwidth used in a localization method [30, 2], a minimal regularity scale in a minimax study [35], or a *signal* part in a signal-to-noise ratio [29, 28, 4]. See [2] for more examples of its use. On the estimation side, the reach has already been studied under several angles.

- The formulation of $\mathrm{rch}(M)$ in terms of deviation to tangent spaces from [26, Theorem 4.18] has been put to use through a plugin in [3]. The authors derived non-matching upper and lower bounds for the estimation of $\mathrm{rch}(M)$ over $\mathcal{C}^3$ submanifolds. In addition to being suboptimal, the method of [3] requires the knowledge of tangent spaces, and is very sensitive to uncertainty on them (see [3, Section 6]).

- Extending the minimax study of [3], [10] took advantage of the so-called *convexity defect function* introduced by [8] to propose another plugin strategy, with rates obtained over more general $\mathcal{C}^k$-smooth manifold classes. Despite still deriving non-matching upper and lower bounds, [10] managed to exhibit two different estimation rates, depending on whether the reach testifies of a high curvature zone (the so-called local case, with slow rates) or of a narrow bottleneck structure (global case, with faster rates). In this work, the derived rates are only suboptimal when the reach is achieved by curvature.

- More recently, [12, Theorem 1] gave a new formulation of the reach in terms of geodesic distortion. Informally, they showed that $\mathrm{rch}(K)$ is the largest radius $r \geqslant 0$ for which the geodesic distance d_K is smaller than the geodesic distance $\mathrm{d}_{\mathcal{S}(r)}$ on a Euclidean ball of radius r. Based on this purely metric statement, [22] proposed to plug-in a nearest-neighbor graph distance of the data in this formulation. This method provides a consistent estimator under very weak assumptions. Unfortunately, it fails to take advantage of high order regularity, when the reach is achieved by curvature (again).

With this analysis of possible estimation flaws in mind, this article proposes a two-step method. In short, we decouple the estimation of the local and global reaches [2], and estimate them separately via max-curvature estimation and geodesic distance estimation respectively.

1.3 Metric Learning

In the data analysis area, *metric learning* refers to the problem of finding a distance $\widehat{\mathrm{d}}$ over the space of observations $\mathbb{X}_n \times \mathbb{X}_n$ that is relevant for a given task at stake [46, 41]. For instance, in a supervised framework where one is provided with tuples of allegedly *similar* or *dissimilar* observations, the goal is to find a distance that is small on the similar tuples and large on the dissimilar ones. There is a wide range of existing methods in the literature, ranging from parametric (LSI [45], MCML [31], LDML [32] among others) to nonparametric (DMLMJ [37], kernel methods [36, 15], to cite a few).

In an unsupervised setting, metric learning aims at finding a metric that takes into account the underlying geometry of the data. That is, it amounts to estimating of shortest path (or *geodesic*) distance. Often, this is done via a dimension reduction technique: any low-dimensional embedding of the data gives rise to a new distance over the data in the embedded space. Existing algorithms include PCA, t-SNE [34], MDS [24], Isomap [42], or MVU [7]. See [41] for a thorough overview of the field.

Astonishingly, despite the variety of existing methods, we are not aware of any general minimax study of geodesic metric learning. Though, two major theoretical references seem to stand out:

- In [43], the authors use a neighborhood graph to estimate distances and derive convergence rates in the $\mathcal{C}^2$ case, but only for nearby points.
- In [5], estimation rates of geodesic distances are derived in the $\mathcal{C}^2$ case using a reconstructing mesh. Lower bounds are also obtained, but in a fixed-design setting only.

We propose a simple plugin method, and show that estimating the geodesic metric is no harder than estimating the manifold itself in Hausdorff distance. This general strategy is also supported by a matching minimax lower bound.

1.4 Contribution and Outline

This article deals with the framework where data lies on an unknown d-dimensional $\mathcal{C}^k$-submanifold of $\mathbb{R}^D$ (Section 2). The main contribution consists of nearly-tight minimax bounds for reach estimation (Section 6). Along the way, three major building blocks, interesting in their own rights, are developed thoroughly:

(Section 3) We propose a general plug-in strategy for estimating the reach of a manifold. It is based on curvature estimation on one hand, and on the estimation of an intermediate scale (framed between the reach and the weak feature size) on the other hand.

(Section 4) We define the so-called *spherical distortion radius* at scale $\delta > 0$ and study its estimation. From the metric characterization of the reach from [12], we notice that this purely metric quantity can be used to play the role of an intermediate scale for reach estimation. We show that its stability properties make it well-suited to play the role of the intermediate scale of Section 3.

(Section 5) We propose a general plugin strategy for metric learning, and derive optimal geodesic metric estimation upper and lower bounds.

The proofs and the most technical points are deferred to the Appendix.

1.5 General Notation

In what follows, $\mathbb{R}^D$ ($D \geqslant 2$) is endowed with the Euclidean norm $\|\cdot\|$. The closed ball of radius $r \geqslant 0$ centered at $x \in \mathbb{R}^D$ is denoted by $\mathsf{B}(x,r)$. If $x \in T \subset \mathbb{R}^D$ is a linear subspace, we write $\mathsf{B}_T(x,r) := T \cap \mathsf{B}(x,r)$ for the same ball in T. Throughout, $c_\square, c'_\square, C_\square, C'_\square \geqslant 0$ denote generic constants that depend on $\square$, and that shall change from line to line to shorten notation. Similarly, universal constants shall generically be denoted by $c, c', C, C' \geqslant 0$.

2 Geometric and Statistical Model

Let us first present the models in which we will work throughout. As will be defined and discussed at length in Section 3.1, we let $\mathrm{rch}(K)$ denote the *reach* of a subset $K \subset \mathbb{R}^D$ of the Euclidean space.

Building upon the standard regression setup, the following class is a good analog of Hlder classes of order $k \geqslant 2$, that is well adapted to submanifolds for stability reasons (see [3, Proposition 1]). Here, the analogy is to be understood as T_pM being the (local) covariate space, and Ψ_p being the regression function.

Definition 2.1 ([2, Definition 1]). Let $k \geqslant 2$, $\mathrm{rch}_{\min} > 0$, and $\mathbf{L} = (L_2, L_3, \ldots, L_k)$. We let $\mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$ denote the set of d-dimensional compact connected submanifolds M of $\mathbb{R}^D$ with $\mathrm{rch}(M) \geqslant \mathrm{rch}_{\min}$, such that for all $p \in M$, there exists a local one-to-one parametrization Ψ_p of the form:

$$\begin{aligned} \Psi_p \colon \mathsf{B}_{T_pM}(0,r) &\longrightarrow M \\ v &\longmapsto p + v + \mathbf{N}_p(v) \end{aligned}$$

for some $r \geqslant \frac{1}{4L_2}$, with $\mathbf{N}_p \in \mathcal{C}^k\left(\mathsf{B}_{T_pM}(0,r), \mathbb{R}^D\right)$ such that for all $\|v\| \leqslant \frac{1}{4L_2}$,

$$\mathbf{N}_p(0) = 0, \quad \mathrm{d}_0\mathbf{N}_p = 0, \quad \text{and} \quad \left\|\mathrm{d}^j_v\mathbf{N}_p\right\|_{\mathrm{op}} \leqslant L_j \text{ for all } j \in \{2, \ldots, k\},$$

where $\mathrm{d}^j_v\mathbf{N}_p$ stands for the jth differential of $\mathbf{N}_p$ at v, and $\|\cdot\|_{\mathrm{op}}$ for the Euclidean operator norm over tensors.

As explained in [2, Section 2.2], radii $1/(4L_2)$ in local parametrizations have only been chosen for convenience. For $k = 2$, the existence of parametrizations Ψ_p is always guaranteed as soon as $\mathrm{rch}_{\min} > 0$ and $L_2 \geqslant 2/\mathrm{rch}_{\min}$ (see [2, Lemma 1]).

Definition 2.2. We let $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$ denote the set of Borel probability distributions P on $\mathbb{R}^D$ satisfying:

- Its support $M := \mathrm{Support}(P)$ belongs to $\mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$;
- It has a density f with respect to the volume measure on M, such that

$$f_{\min} \leqslant f(x) \leqslant f_{\max} \quad \text{for all} \quad x \in M.$$

On the estimation side, the uniform smoothness of the parametrizations in Definition 2.1 allows for estimation of the manifold via local polynomial fitting around sample points in the models $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$. Recall that the *Hausdorff distance* between two compact subsets $K, K' \subset \mathbb{R}^D$ is defined by

$$\mathrm{d_H}(K, K') := \max\left\{\sup_{x\in K} \mathrm{d}(x, K'),\ \sup_{x'\in K'} \mathrm{d}(x', K)\right\}, \tag{1}$$

where for all $u \in \mathbb{R}^D$,

$$\mathrm{d}(u, K) := \min_{x\in K} \|x - u\| \tag{2}$$

stands for the *distance function* to K. The estimation rates over the model $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$ have been studied in [2]. A key result that we will use is the following.

Theorem 2.3 ([2, Theorem 6])**.** *There exists an estimator $\widehat{M}$ such that for n large enough,*

$$\sup_{P\in\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} \mathbb{E}_{P^n}[\mathrm{d_H}(\widehat{M}, M)] \leqslant C_{d,k,\mathrm{rch}_{\min},\mathbf{L},f_{\min},f_{\max}} \left(\frac{\log n}{n}\right)^{k/d},$$

where in the supremum, M stands for $\mathrm{Support}(P)$.

This rate is minimax optimal up to $\log n$ factors [2, Theorem 7]. It can be achieved by a local polynomial patch estimator $\widehat{M}$ (see (7) below) that we will use as a preliminary step towards reach estimation. Let us also mention here that these fitted local polynomials also allow for estimation of differential quantities of M, such as tangent spaces and curvature at sample points, with (minimax) convergence rates of order $O(n^{-(k-1)/d})$ and $O(n^{-(k-2)/d})$ respectively (see [2, Theorems 2 to 5]). This fact will be of key importance in Section 3.3, where estimating the maximal curvature of M will allow to estimate the so-called "local reach".

3 Reach and Related Quantities

3.1 Characterizations and Relaxations of the Reach

Let K be a compact subset of $\mathbb{R}^D$. Following the original definition of [26], the *reach* of K, denoted by $\mathrm{rch}(K)$, may be thought of as the largest radius of a neighborhood of K onto which the projection map π_K onto K is well-defined. More formally, define the *medial axis* of K by

$$\mathrm{Med}(K) := \left\{u \in \mathbb{R}^D \mid \exists x_1 \neq x_2 \in K \quad \|u - x_1\| = \|u - x_2\| = \mathrm{d}(u, K)\right\}.$$

The reach of K is then defined as the smallest distance between K and $\mathrm{Med}(K)$.

Definition 3.1. For all closed $K \subset \mathbb{R}^D$, the *reach* of K is defined by

$$\mathrm{rch}(K) := \min_{x \in K} \mathrm{d}(x, \mathrm{Med}(K)) = \inf_{u \in \mathrm{Med}(K)} \mathrm{d}(u, K).$$

Note that in full generality, the medial axis might not be a closed set, so that the infimum in Definition 3.1 may not be attained (for instance in the case where K is one-dimensional with a sharp edge). From a topological viewpoint, a key property of sets with positive reach is that the projection onto K induces continuous retractions from the *offset* $K^r := \left\{u \in \mathbb{R}^D \mid \mathrm{d}(u, K) \leqslant r\right\}$ onto K, whenever $r < \mathrm{rch}(K)$ [26, Theorem 4.8]. This property is at the core of topologically consistent reconstruction procedures such as that of [11].

Sets with positive reach can also been thought of as generalizations of convex sets, characterized by the smoothness of their distance function. Indeed, based on the remark that $x \mapsto \mathrm{d}(x, K)$ is $\mathcal{C}^1$ on $\mathbb{R}^D \backslash K$ whenever K is convex, [23] define *r-proximally-smooth* sets as the sets K such that $\mathrm{d}(\cdot, K)$ is $\mathcal{C}^1$ over $\left\{u \in \mathbb{R}^D \mid 0 < \mathrm{d}(u, K) < r\right\}$. Interestingly, for subsets of $\mathbb{R}^D$, r-proximally smooth sets are exactly sets with reach $\mathrm{rch}(K) \geqslant r$ [39], so that the reach may be alternatively defined in terms of gradients of the distance function. To this aim, following [19], a generalized gradient function can be defined over $\mathbb{R}^D \setminus K$. For all $x \in \mathbb{R}^D \setminus K$, we write

$$\nabla \mathrm{d}(x, K) := \frac{x - c_K(x)}{\mathrm{d}(x, K)}, \tag{3}$$

where $c_K(x)$ is the center of the smallest enclosing ball of the set $\pi_K(\{x\})$ of nearest neighbors of x on K. Since $c_K(x) = \pi_K(x)$ whenever $x \notin \mathrm{Med}(K)$, the medial axis can actually be characterized as

$$\mathrm{Med}(K) = \left\{x \in \mathbb{R}^D \setminus K \mid \|\nabla \mathrm{d}(x, K)\| < 1\right\},$$

and the reach as

$$\mathrm{rch}(K) = \sup\left\{r > 0 \mid 0 < \mathrm{d}(x, K) < r \Rightarrow \|\nabla \mathrm{d}(x, K)\| = 1\right\}.$$

This characterization of the reach allows for a straightforward relaxation. Namely, for a parameter $\mu \in [0, 1]$, the seminal paper [16] introduces the so-called *μ-medial axis* as being

$$\mathrm{Med}_\mu(K) := \left\{x \in \mathbb{R}^D \setminus K \mid \|\nabla \mathrm{d}(x, K)\| \leqslant \mu\right\},$$

and the μ-reach as

$$\mathrm{rch}_\mu(K) := \inf_{u \in \mathrm{Med}_\mu(K)} \mathrm{d}(u, K). \tag{4}$$

It is clear that for all $\mu < 1$, $\mathrm{rch}(K) \leqslant \mathrm{rch}_\mu(K)$, with $\mathrm{rch}(K)$ corresponding to the limit $\mathrm{rch}_{1^-}(K)$. Furthermore, this relaxation of the reach still yields enough regularity guarantees that the offsets $K^r = \left\{u \in \mathbb{R}^D \mid \mathrm{d}(u, K) \leqslant r\right\}$ are isotopic for all $r \in (0, \mathrm{rch}_\mu(K))$ [16, Lemma 2.1]. Hence, the condition that $\mathrm{rch}_\mu(K) > 0$ conveys enough regularity properties for many topological estimators to work [17].

Through this lens, the largest radius that ensures the topological stability of the offsets is the 0-reach, also called *weak-feature size*,

$$\mathrm{wfs}(K) := \inf_{u \in \mathrm{Med}_0(K)} \mathrm{d}(u, K), \tag{5}$$

that is the distance from K to the set of critical points of $\mathrm{d}(\cdot, K)$. As detailed in the following section, the weak-feature size plays a special role in the case where K is a manifold. Here come a few elementary properties of the weak feature size that we will use later on.

Proposition 3.2. *Let $K \subset \mathbb{R}^D$ be compact.*

(i) If K is a closed submanifold of $\mathbb{R}^D$, then $\mathrm{wfs}(K) < +\infty$;

(ii) If $\mathrm{wfs}(K) < +\infty$, then for all $\mu \in [0, 1)$,

$$\mathrm{rch}(K) \leqslant \mathrm{rch}_\mu(K) \leqslant \mathrm{wfs}(K) \leqslant \sqrt{\frac{D}{2(D+1)}}\,\mathrm{diam}(K).$$

A proof is given in Section A.1. Proposition 3.2 thus ensures that $\mathrm{wfs}(M)$ is uniformly bounded over the classes $\mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$ introduced in Section 2. Since $\mathrm{wfs}(K)$ and $\mathrm{rch}_\mu(K)$ both measure a typical scale for topological stability, estimating them from sample could be of practical interest for topological inference. Unfortunately, the following negative result shows that this estimation problem is intractable, even over a well-behaved model of closed $\mathcal{C}^k$-submanifolds such as $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$.

Theorem 3.3. *Assume that $f_{\min} \leqslant c_{d,k}/\,\mathrm{rch}^d_{\min}$ and $f_{\max} \geqslant C_{d,k}/\,\mathrm{rch}^d_{\min}$, and $L_j \geqslant C_{d,k}/\,\mathrm{rch}^{j-1}_{\min}$ for all $j \in \{2, \ldots, k\}$. Then there exists $\tilde{c}_{d,k} > 0$ such that for all $n \geqslant 1$ and $\mu \in [0, 1)$,*

$$\inf_{\widehat{r}_\mu} \sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}}\left[|\widehat{r}_\mu - \mathrm{rch}_\mu(M)|\right] \geqslant \tilde{c}_{d,k}\,\mathrm{rch}_{\min} > 0,$$

where $\widehat{r}_\mu$ ranges among all the possible estimators based on n samples.

An intuition behind Theorem 3.3 is that for all $\mu < 1$, the μ-medial axis is an unstable structure. For certain manifolds $M_0 \in \mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$, one can find arbitrarily small perturbations of M_0 whose μ-medial axes remain at a fixed Hausdorff distance from $\mathrm{Med}_\mu(M_0)$. See the proof of Theorem 3.3 in Section A.2 for a precise statement of this intuition.

Despite the fact that $\mathrm{rch}(K) = \mathrm{rch}_{1^-}(K)$, this negative result indicates that we cannot leverage μ-reach estimation to obtain quantitative bounds for reach estimation. We shall hence turn towards other reach-related quantities. In fact, the particular case where $K = M$ is a manifold offers us several other characterizations of the reach, which suggest other estimation strategies.

3.2 Reach of Submanifolds

In what follows, M stands for a d-dimensional closed submanifold of $\mathbb{R}^D$. Note that [26, Remarks 4.20 and 4.21] and [12] assert that a closed submanifold with positive reach is at least of

regularity $\mathcal{C}^{1,1}$, so that geodesics and tangent spaces are always defined in the usual differential sense. For the manifold case, the intuition of $\mathrm{rch}(M)$ as a generalized convexity parameter is further backed by [26, Theorem 4.8]. Indeed, the inequality $\langle x - \pi_C(x), \pi_C(x) - c\rangle \geqslant 0$ valid for all $c \in C$ and $x \in \mathbb{R}^D$ whenever C is convex, translates to $\langle x - \pi_M(x), \pi_M(x) - y\rangle \geqslant -\|\pi_M(x) - y\|^2 \|x - \pi_M(x)\|/(2\,\mathrm{rch}(M))$ being valid for all $y \in \mathbb{R}^D$ and $x \in \mathbb{R}^D$ such that $\mathrm{d}(x, M) < \mathrm{rch}(M)$. This leads to the following characterization of the reach, in the manifold case.

Theorem 3.4 ([26, Theorem 4.18])**.** *For a submanifold $M \subset \mathbb{R}^D$ without boundary,*

$$\mathrm{rch}(M) = \inf_{p \neq q \in M} \frac{\|p - q\|^2}{2\mathrm{d}(q - p, T_pM)},$$

where T_pM denotes the tangent space of M at p.

This result provides a natural plugin estimator, proposed by [2], which consists in replacing M and T_pM by suitable estimators of them. A key result from [2] is a description of how the infimum in Theorem 3.4 is achieved, possibly asymptotically.

Theorem 3.5 ([2, Theorem 3.4])**.** *Let $M \subset \mathbb{R}^D$ be a compact $\mathcal{C}^2$ submanifold without boundary. Then,*

$$\mathrm{rch}(M) = \mathrm{wfs}(M) \wedge R_\ell(M),$$

where denoting by $\mathrm{I\!I}_p : T_pM \times T_pM \to T_pM^\perp$ the second fundamental form *of M at $p \in M$,*

$$R_\ell(M) := \min_{p \in M} \| \mathrm{I\!I}_p \|_{\mathrm{op}}^{-1}$$

stands for the minimal curvature radius of M.

This result conveys the following intuition in the manifold case: the infimum in the right-hand side of Theorem 3.4 may be attained:

(Local case) Asymptotically, for pairs of points (p, q) converging to a maximal curvature point in some direction, so that $\mathrm{rch}(M) = R_\ell(M)$.

(Global case) For a pair of points (p, q) belonging to parallel areas of M, forming a bottleneck zone, so that $\mathrm{rch}(M) = \mathrm{wfs}(M)$.

This local/global dichotomy of the reach may also be retrieved in the recent characterization given by [12] in terms of metric distortion.

Theorem 3.6 ([12, Theorem 1])**.** *Let $K \subset \mathbb{R}^D$ be a closed subset. Then*

$$\mathrm{rch}(K) = \sup \left\{ r > 0 \mid \forall p, q \in K, \|p - q\| < 2r \Rightarrow \mathrm{d}_K(p, q) \leqslant 2r \arcsin\left(\frac{\|p - q\|}{2r}\right) \right\},$$

where $\mathrm{d}_K : K \times K \to \bar{\mathbb{R}}_+$ stands for the shortest-path *(or* geodesic*) distance on K.*

Recall that, for all $p, q \in K$, the distance $\mathrm{d}_K(p, q)$ is the infimum of the length of all the continuous path in K between p and q. As will be detailed in Section 4, the above result allows to characterize the reach in terms of metric distortion with respect to metrics on spheres of radii r. In the same spirit as Theorem 3.5, when $K = M$ is a submanifold, the configurations of (p, q, r) in the supremum of Theorem 3.6 are limited by the same two local and global layouts:

(Local case) When p and q tend to a maximal curvature point in some direction, the geodesic distance d_K behaves like that of a sphere of radius $R_\ell(M)$ at this point in this direction.

(Global case) When p and q are in parallel areas, their geodesic distance must be larger than the spherical distance of radius $\|p - q\|/2$.

3.3 Plug-in Methods for Reach Estimation

The characterizations of the reach given in Section 3.2 all lead to their associate plug-in estimators:

- Studying a $\mathcal{C}^3$ model similar to $\mathcal{P}^3_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$, [2] took advantage of the characterization with tangent spaces (Theorem 3.4) to conceive a reach estimator that converges at rate $O(n^{-2/(3d-1)})$ in the local case ($\mathrm{rch}(M) = R_\ell(M)$), and $O(n^{-1/d})$ in the global case ($\mathrm{rch}(M) = \mathrm{wfs}(M)$).
- Based on the metric distortion characterization of Theorem 3.6, [22] propose a reach estimator that is consistent whenever M has positive reach.

In light of Theorem 3.5, differences of convergence rates between the local and global case are to be expected. To quantify this intuition, [10] introduces subclasses of the model $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$, parametrized by the gap between $R_\ell(M)$ and $\mathrm{wfs}(M)$. They obtain the following lower bounds.

Theorem 3.7 ([10, Theorem 7.1] and [2, Proposition 2.9])**.** *Let $\alpha \in \mathbb{R}$, $k \geqslant 2$, and write*

$$\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L},\alpha}(f_{\min}, f_{\max}) := \left\{ P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max}) \mid R_\ell(M) \geqslant \mathrm{wfs}(M) + \alpha \right\},$$

where M denotes $\mathrm{Support}(P)$. Then, for all $\mathrm{rch}_{\min} > 0$ there exists small enough $f_{\min}$ and large enough $f_{\max}$, $\mathbf{L}$ such that

$$\inf_{\widehat{\mathrm{rch}}} \sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L},\alpha}(f_{\min},f_{\max})} \mathbb{E}|\widehat{\mathrm{rch}} - \mathrm{rch}(M)| \geqslant c_{\mathrm{rch}_{\min},d,k} \left(\frac{1}{n}\right)^{(k-2)/d}, \quad \text{if } \alpha \leqslant 0,$$

$$\inf_{\widehat{\mathrm{rch}}} \sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L},\alpha}(f_{\min},f_{\max})} \mathbb{E}|\widehat{\mathrm{rch}} - \mathrm{rch}(M)| \geqslant c_{\mathrm{rch}_{\min},d,k,\alpha} \left(\frac{1}{n}\right)^{k/d}, \quad \text{if } \alpha > 0.$$

These bounds indicate that estimating the reach is at least as hard as estimating the curvature in the local case ($\mathrm{rch}(M) = R_\ell(M)$), and at least as hard as estimating the manifold in the global case ($\mathrm{rch}(M) = \mathrm{wfs}(M)$). We will prove in Section 6 that these rates are in fact minimax optimal up to $\log n$ factors. This means that reducing reach estimation to curvature *and* manifold estimation is a good way to go, as it leads to optimal rates. To do so, following the idea behind Theorem 3.5, estimating $R_\ell(M)$ – or some notion of local reach – and $\mathrm{wfs}(M)$ – or some notion of global reach – separately seems a sensible approach.

3.3.1 Local Reach Estimation

For (max-)curvature estimation, the strategy that we adopt follows from the polynomial patches estimator proposed in [3]. Given a localization bandwidth $h > 0$, and a parameter $t > 0$, for all $i \in \{1, \dots, n\}$, we let $\hat{\pi}_i : \mathbb{R}^D \to \mathbb{R}^D$ be an orthogonal projector of rank d and $\hat{\mathbb{T}}^{(j)}_i : \left(\mathbb{R}^D\right)^{\otimes j} \to \mathbb{R}^D$ be symmetric tensors solutions of the least squares problem

$$\min_{\substack{\pi \\ \max_{2\leqslant j\leqslant k-1} \|\mathbb{T}^{(j)}\|^{\frac{1}{j-1}} \leqslant t}} P^{(i)}_{n-1} \left[\left\| x - \pi(x) - \sum_{j=2}^{k-1} \mathbb{T}^{(j)}(\pi(x)^{\otimes j}) \right\|^2 \mathbb{1}_{\mathsf{B}(0,h)}(x) \right], \tag{6}$$

where $P^{(i)}_{n-1} := \frac{1}{n-1} \sum_{p \neq i} \delta_{X_p - X_i}$ denotes the empirical measure centered at point X_i. Following [3, Section 3], if h is taken to be of order $\Theta\left((\log n/n)^{1/d}\right)$, that t is chosen such that $t^k h \leqslant 1$, and

that $\hat{T}_i := \mathrm{Im}(\hat{\pi}_i)$ denotes the image of $\hat{\pi}_i$ – which is a d-dimensional vector space by construction –, then the local patches

$$\begin{aligned} \widehat{\Psi}_i \colon \ \mathsf{B}_{\hat{T}_i}(0, 7h/8) &\longrightarrow \mathbb{R}^D \\ v &\longmapsto X_i + v + \sum_{j=2}^{k-1} \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j}) \end{aligned} \tag{7}$$

are local $O(h^k)$ approximations of M whenever n is large enough. Furthermore, for $v \in \mathsf{B}_{\hat{T}_i}(0, h/4)$, we can estimate the curvature tensor at $\pi_M(\widehat{\Psi}_i(v))$ via the second derivative of $\widehat{\Psi}_i$ at v, expressed in local coordinates around $\widehat{\Psi}_i(v)$ given by a basis of $\mathrm{Im}(\mathrm{d}_v\widehat{\Psi}_i)$. To summarize, for all $v \in \mathsf{B}_{\hat{T}_i}(0, h/4)$, (7) provides a d-dimensional space $\hat{T}_{i,v} := \mathrm{Im}(\mathrm{d}_v\widehat{\Psi}_i)$, as well as a symmetric bilinear map

$$\hat{\mathbb{T}}_{i,v}^{(2)} : \hat{T}_{i,v} \times \hat{T}_{i,v} \to \hat{T}_{i,v}^{\perp},$$

that is provably close to $\mathrm{I\!I}_{\pi_M(\widehat{\Psi}_i(v))}$. The precise definition of $\hat{\mathbb{T}}_{i,v}^{(2)}$ is given in Section A.3. A minimal curvature radius (i.e. maximal curvature) estimator may then be computed as the minimal curvature radius of all the polynomial patches around sample points, that is

$$\widehat{R}_\ell := \min_{1 \leqslant i \leqslant n} \min_{v \in \mathsf{B}_{\hat{T}_i}(0, h/4)} \|\hat{\mathbb{T}}_{i,v}^{(2)}\|_{\mathrm{op}}^{-1}. \tag{8}$$

Provided M is uniformly well approximated by $\bigcup_{i=1}^n \widehat{\Psi}_i(\mathsf{B}_{\hat{T}_i}(0, h/4))$, the convergence rate of $\widehat{R}_\ell$ towards $R_\ell(M)$ will follow from uniform curvature bounds, similar to the pointwise ones from [3, Theorem 4]. We are able to prove the following.

Theorem 3.8. *Let $k \geqslant 3$ and $P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$. Write $h = \left(C_{d,k} \frac{f_{\max}^2 \log n}{f_{\min}^3 n}\right)^{1/d}$. Then for n large enough, with probability larger than $1 - 2n^{-k/d}$, we have*

$$\left|\widehat{R}_\ell - R_\ell(M)\right| \leqslant C_{d,k,\mathbf{L},\mathrm{rch}_{\min}} R_\ell^2(M) \sqrt{\frac{f_{\max}}{f_{\min}}} h^{k-2}.$$

We refer to Section A.3 for a proof of this result. In particular, the estimator $\widehat{R}_\ell$ achieves the rate of the lower bound from Theorem 3.7 in the case where $\mathrm{rch}(M) = R_\ell(M)$ (i.e. $\alpha \leqslant 0$), up to $\log n$ factors.

3.3.2 Global Reach Estimation

To complete the construction of an estimator of $\mathrm{rch}(M)$, building an estimator of $\mathrm{wfs}(M)$ could be a possibility. However, Theorem 3.3 shows that building an estimator of the weak feature size with a uniform convergence rates over $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$ is hopeless. Nonetheless, it is important to note that a uniform estimation rate of $\mathrm{wfs}(M)$ over $\mathcal{P}^k$ is not necessary to obtain uniform convergence rate for $\mathrm{rch}(M)$. Indeed, an estimator $\widehat{\mathrm{wfs}}$ of $\mathrm{wfs}(M)$ that exhibits an optimal uniform convergence rate whenever $\mathrm{wfs}(M) \leqslant R_\ell(M)$, and that is provably larger than $R_\ell(M)$ otherwise, is enough to build an optimal reach estimator when combined with $\widehat{R}_\ell$. This is the case, for instance, of the weak feature size estimator of [10] based on the so-called *convexity defect function*.

Based on this remark, we adopt a more general strategy, by seeking for an intermediate geometric scale $\theta(M)$ (or *feature size*) such that for all $M \in \mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$,

$$\mathrm{rch}(M) \leqslant \theta(M) \leqslant \mathrm{wfs}(M).$$

In such a case, Theorem 3.5 extends trivially, with $\mathrm{wfs}(M)$ replaced by $\theta(M)$.

Proposition 3.9. *Assume that* $\theta : \mathcal{C}^2_{\mathrm{rch}_{\min}} \to \mathbb{R}_+$ *is such that* $\mathrm{rch}(M) \leqslant \theta(M) \leqslant \mathrm{wfs}(M)$ *for all* $M \in \mathcal{C}^2_{\mathrm{rch}_{\min}}$. *Then,*

$$\mathrm{rch}(M) = \theta(M) \wedge R_\ell(M).$$

Given such an intermediate scale parameter of interest $\theta(M)$, and assuming that a consistent estimator $\widehat{\theta}$ of $\theta(M)$ is available, one can naturally consider the plugin $\widehat{\mathrm{rch}} := \widehat{R}_\ell \wedge \widehat{\theta}$. For free, Proposition 3.9 yields that $\theta(M)\mathbb{1}_{R_\ell(M)>\mathrm{rch}(M)} = \mathrm{rch}(M)\mathbb{1}_{R_\ell(M)>\mathrm{rch}(M)}$, so that

$$|\,\mathrm{rch}(M) - \widehat{\mathrm{rch}}| \leqslant |\widehat{R}_\ell - R_\ell(M)|\mathbb{1}_{R_\ell(M)\leqslant\mathrm{rch}(M)} + |\widehat{\theta} - \theta(M)|\mathbb{1}_{R_\ell(M)>\mathrm{rch}(M)}, \tag{9}$$

as soon as $|R_\ell(M) - \widehat{R}_\ell| + |\theta(M) - \widehat{\theta}| \leqslant |R_\ell(M) - \theta(M)|$. In addition, such a quantity would provide a local scale that is of interest for further topological inference, as exposed in Section 3.1.

According to Theorem 3.3, taking $\theta(M)$ to be related to the medial axis characterization of the reach – such as the μ-reach, or the λ-reach defined in [19]) – is likely to lead to an unsolvable statistical problem, because of the inherent instability of the medial axis. Hence, we rather build upon the metric distortion characterization of the reach given by Theorem 3.6, and provide a better-behaved intermediate scale $\theta(M)$: the *spherical distortion radius*.

4 Spherical Distortion Radius

4.1 Motivation and Definition

Based on Theorem 3.6, we now build a geometrically stable feature size that measures the maximum radius (or scale) at which the geodesic distance can be compared to the corresponding spherical distance. To be more precise, for $x, y \in \mathbb{R}^D$ and $r > 0$, we define the *spherical distance* $\mathrm{d}_{\mathcal{S}(r)}(x, y)$ – or *great-circle distance* – as the distance between x and y when seen as both lying on a sphere of radius r. That is,

$$\mathrm{d}_{\mathcal{S}(r)}(x, y) := \begin{cases} 2r \arcsin\left(\frac{\|x-y\|}{2r}\right) & \text{if } \|x - y\| \leqslant 2r, \\ +\infty & \text{otherwise} \end{cases}$$

Note that the map $r \mapsto \mathrm{d}_{\mathcal{S}(r)}(x, y)$ is decreasing on $[\|x - y\|/2, \infty)$ and that

$$\mathrm{d}_{\mathcal{S}(r)}(x, y) = \frac{1}{2}\pi\|x - y\| \quad \text{for} \quad r = \frac{\|x - y\|}{2} \quad \text{and} \quad \mathrm{d}_{\mathcal{S}(r)}(x, y) \xrightarrow[r\to\infty]{} \|x - y\|.$$

Then, Theorem 3.6 can be rewritten as

$$\mathrm{rch}(K) = \sup\left\{r > 0 \mid \forall x, y \in K, \|x - y\| < 2r \Rightarrow \mathrm{d}_K(x, y) \leqslant \mathrm{d}_{\mathcal{S}(r)}(x, y)\right\}.$$

It should be noted that $\mathrm{d}_{\mathcal{S}(r)}$ is not formally a distance on K (unless K is a subset of a sphere of radius r), but this is of little importance in what follows.

Based on the same idea that motivates the introduction of the μ-reach, we intend to discard curvature effects to obtain some notion of global reach. In the metric characterization of the reach from Theorem 3.6, this can be done by the supremum restricting to points that are not too close.

Definition 4.1. Let K be a compact subset of $\mathbb{R}^D$, d a distance on K and $\delta > 0$. The *spherical distortion radius* of the metric space (K, d) at scale δ is defined by

$$\mathrm{sdr}_\delta(K, \mathrm{d}) := \sup \left\{ r > 0 \;\middle|\; \forall x, y \in K,\ \delta \leqslant \|x - y\| < 2r \ \Rightarrow\ \mathrm{d}(x, y) \leqslant \mathrm{d}_{\mathcal{S}(r)}(x, y) \right\}.$$

In words, the spherical distortion radius at scale $\delta > 0$ is the largest radius r for which the distance d is bounded above by the spherical distance at radius r, when restricted to points that are at least δ-apart for the Euclidean distance.

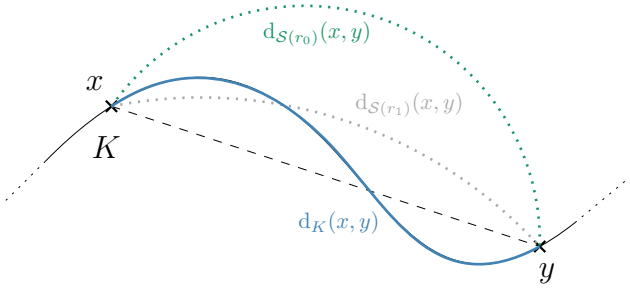


Figure 1: A curve K in the plane. In blue is the shortest path between two points x and y, whose length is $\mathrm{d}_K(x, y)$. In green (resp. grey) is the circle portion of radius r_0 (resp. r_1) going through x and y. The layout is chosen so that $r_0 \leqslant r_1$ and $\mathrm{d}_{\mathcal{S}(r_1)}(x, y) \leqslant \mathrm{d}_K(x, y) \leqslant \mathrm{d}_{\mathcal{S}(r_0)}(x, y)$.

By construction, $\mathrm{sdr}_\delta(K, d) \geqslant \delta/2$ for all $\delta > 0$. Furthermore, whenever δ is strictly greater than $\operatorname{diam} K$, then no pairs of points in $x, y \in K$ satisfies $\|x - y\| \geqslant \delta$ so that $\mathrm{sdr}_\delta(K) = +\infty$. On the other hand, if $\delta = 0$, then the spherical distortion radius of (K, d_K), coincides with the reach of K (Theorem 3.6). In fact, Proposition 4.2 below confirms that the spherical distortion radius interpolates between the reach and the weak feature size.

Proposition 4.2. *For all closed $K \subset \mathbb{R}^D$ and all metric d on K, the map $\delta \mapsto \mathrm{sdr}_\delta(K, \mathrm{d})$ is non-decreasing. Furthermore, for $\mathrm{d} = \mathrm{d}_K$,*

$$\mathrm{rch}(K) \leqslant \mathrm{sdr}_\delta(K, \mathrm{d}_K) \leqslant \mathrm{wfs}(K) \quad \textit{for all} \quad 0 \leqslant \delta \leqslant \sqrt{\frac{2(D+1)}{D}}\, \mathrm{wfs}(K).$$

A proof of Proposition 4.2 is given in Appendix B.1.

Example 4.3. As a toy example, let us study the spherical distortion radius of the wedge shape $K_\alpha = \mathcal{L}_1 \cup \mathcal{L}_2$ where $\mathcal{L}_1$ and $\mathcal{L}_2$ are two half-line originated from a common point $z \in \mathbb{R}^D$ (see Figure 2). We let $\alpha \in (0, \pi)$ be the angle between these two lines. In this context, we have $\mathrm{rch}(K_\alpha) = 0$, and it is easy to see that $\mathrm{wfs}(K_\alpha) = \infty$. Furthermore, the usual interpolations between the reach and the weak feature size exhibit a very degenerate behavior in the presence of an angular configuration such as this one, with for instance

$$\mathrm{rch}_\mu(K_\alpha) = \begin{cases} 0 & \text{if } \ \mu \geqslant \sin(\alpha/2), \\ \infty & \text{if } \ \mu < \sin(\alpha/2). \end{cases}$$

On the contrary, we show hereafter that the spherical distortion radius interpolates non-trivially between $\mathrm{rch}(K_\alpha)$ and $\mathrm{wfs}(K_\alpha)$ in this case, giving rise to a new family of relevant characteristic scales even for non-smooth subsets K_α.

To see this, take $x \in \mathcal{L}_1$ and $y \in \mathcal{L}_2$, and denote by $a := \|x - z\|$ and $b := \|y - z\|$. The intrinsic distance $\mathrm{d}_{K_\alpha}(x, y)$ is given by $a + b$ while $\|x - y\|^2 = a^2 + b^2 - 2ab\cos(\alpha)$. Now the solution of the minimization problem

$$\min\left\{a^2 + b^2 - 2ab\cos(\alpha) \mid a + b = \mathrm{d}_{K_\alpha}(x, y)\right\}$$

is given by $a = b = \mathrm{d}_{K_\alpha}(x, y)/2$ and equals $\mathrm{d}^2_{K_\alpha}(x, y)\sin^2(\alpha/2)$. The spherical distortion radius of K_α at scale δ is thus the largest r such that

$$\frac{\delta}{\sin(\alpha/2)} \leqslant 2r \arcsin\left(\frac{\delta}{2r}\right). \tag{10}$$

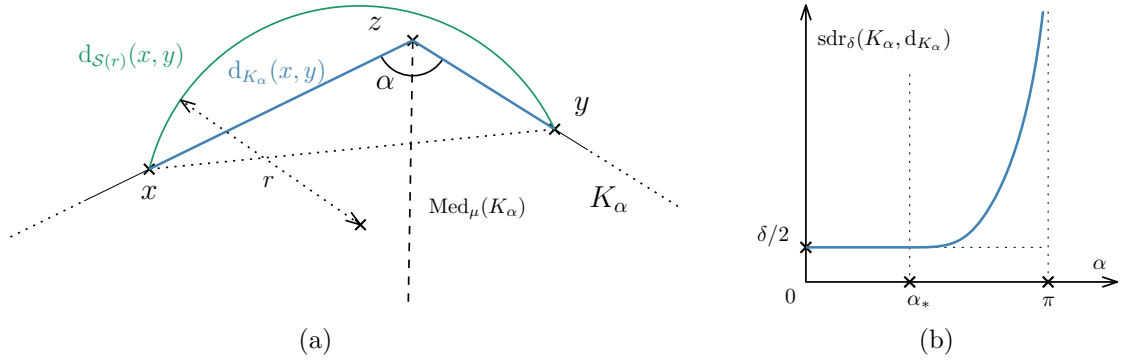
Since the right-hand side above ranges between δ and $\delta\pi/2$, we distinguish two cases:

- If $\sin(\alpha/2) < 2/\pi$, then no r can fulfill (10). Hence, $\mathrm{sdr}_\delta(K_\alpha, \mathrm{d}_{K_\alpha}) = \delta/2$.
- Otherwise $\sin(\alpha/2) \geqslant 2/\pi$, in which case the largest r is given by the equality $\varphi(2r/\delta) = 1/\sin(\alpha/2)$, where $\varphi(u) := u\arcsin(1/u)$ is a bijection between $[1, \infty)$ and $(1, \pi/2]$.

All in all, it holds

$$\mathrm{sdr}_\delta(K_\alpha, \mathrm{d}_{K_\alpha}) = \begin{cases} \delta/2 & \text{if} \quad \alpha < \alpha_* \\ (\delta/2)\varphi^{-1}(1/\sin(\alpha/2)) & \text{if} \quad \alpha \geqslant \alpha_* \end{cases}$$

where $\alpha_* = 2\arcsin(2/\pi) < \pi/2$. Note that compared to $\mathrm{rch}_\mu(K_\alpha)$, there is no discontinuity in $\mathrm{sdr}_\delta(K_\alpha, \mathrm{d}_{K_\alpha})$ as α varies.



(a) (b)

Figure 2: (a) Diagram of $K_\alpha = \mathcal{L}_1 \cup \mathcal{L}_2$ with an angle α between the two half-lines. The shortest path between x and y is drawn in blue. In dashed the μ-medial axis for $\mu > \sin(\alpha/2)$, showing in particular that $\mathrm{rch}_\mu(K_\alpha) = 0$ in this case. (b) Plot of the function $\alpha \mapsto \mathrm{sdr}_\delta(K_\alpha, \mathrm{d}_{K_\alpha})$, which operates a smooth interpolation between $\delta/2$ and ∞.

Example 4.3 above carries the intuition that the spherical distortion radius seems somehow stable with respect to Hausdorff perturbations, contrary to the μ-reach. We quantify this intuition in the following section.

4.2 Stability Properties

In this section, we will be comparing different metric spaces on subsets of $\mathbb{R}^D$. Let K and K' be two subsets of $\mathbb{R}^D$, endowed with distances d and d' respectively. We intend to prove that $\mathrm{sdr}_\delta(K, \mathrm{d})$ and $\mathrm{sdr}_\delta(K', \mathrm{d}')$ are close whenever (K, d) and (K', d') are close, and that (K, d) has good properties. The notion of proximity between K and K' will be measured in Hausdorff distance (see (1)). It remains to define a notion of proximity between d and d', which is called the *mutual distortion*.

Definition 4.4. Let (K, d) and (K', d') be two metric subspaces of $\mathbb{R}^D$. The *metric distortion* of d' relative to d at scale $\delta > 0$ is

$$\mathrm{D}_\delta(\mathrm{d}'|\mathrm{d}) := \sup_{\substack{x',y' \in K' \\ \|x'-y'\| \geqslant \delta}} \frac{\mathrm{d}'(x', y')}{\mathrm{d}(\pi_K(\{x'\}), \pi_K(\{y'\}))}.$$

where π_K is the (possibly multivalued) closest-point projection onto K for the ambient Euclidean distance, and where

$$\mathrm{d}(\pi_K(\{x'\}), \pi_K(\{y'\})) := \inf \left\{ \mathrm{d}(x, y) \mid x \in \pi_K(\{x'\}),\ y \in \pi_K(\{y'\}) \right\}.$$

We adopt the convention $\mathrm{D}_\delta(\mathrm{d}'|d) = 0$ if $\delta > \mathrm{diam}(K')$. The *mutual distortion* of d and d' is then defined as

$$\mathrm{D}_\delta(\mathrm{d}, \mathrm{d}') := \max \left\{ \mathrm{D}_\delta(\mathrm{d}'|\mathrm{d}), \mathrm{D}_\delta(\mathrm{d}|\mathrm{d}') \right\}.$$

The mutual distortion defined above allows to compare distances on different spaces, while taking into account their respective embeddings in $\mathbb{R}^D$. A small distortion $\mathrm{D}_\delta(\mathrm{d}, \mathrm{d}')$ means that, if $a, b \in K$ and $x, y \in K'$ are two couples of points that are δ-separated and such that x and a, and y and b are respectively close to each other, then $\mathrm{d}(a, b)$ and $\mathrm{d}'(x, y)$ should be close as well. This definition of *mutual distortion* between metric subspaces of $\mathbb{R}^D$ is related to the existing notion *metric distortion* of an embedding. See for instance [14] or more recently [21] which deals with distortion measures in a statistical framework. It is nonetheless significantly different, in particular because the usual notion of distortion is invariant through re-scaling of either d or d'. In our framework, invariance with respect to scaling is an undesirable property, since we want to estimate the reach, which is itself a scale factor (or feature size).

Remark 4.5. When $K = K'$, the mutual distortion can be seen as the bi-Lipschitz coefficient of $\mathrm{Id} : (K, \mathrm{d}) \to (K, \mathrm{d}')$ at *scale* δ, meaning that for all $x, y \in K$

$$\|x - y\| \geqslant \delta \ \Rightarrow\ \frac{1}{L}\mathrm{d}'(x, y) \leqslant \mathrm{d}(x, y) \leqslant L\mathrm{d}'(x, y),$$

where $L = \mathrm{D}_\delta(\mathrm{d}, \mathrm{d}')$. In particular, a mutual distortion that is close to 1 means that (K, d) is quasi-isometric to (K, d'), at scale δ.

If the two subspaces K and K' are too far apart, then it makes no sense to compare two distances d and d' defined on them, and one could expect the mutual distortion to explode. This is will typically the case when $\mathrm{d_H}(K, K') \geqslant \delta$.

It is clear from the definition that using the notion of relative metric distortion defined above, the spherical distortion radius of K may be expressed as

$$\mathrm{sdr}_\delta(K, \mathrm{d}) = \sup \left\{ r > 0 \ \middle|\ \mathrm{D}_\delta(\mathrm{d}|\mathrm{d}_{\mathcal{S}(r)}) \leqslant 1 \right\}.$$

This point supports the idea that the relative metric distortion we defined is a suitable notion of proximity to assess stability of the spherical distortion radius, as exposed by the following proposition.

Proposition 4.6. *Let $\delta_0 > 0$ and $\varepsilon, \nu > 0$. Assume that both $\mathrm{d_H}(K', K) \leqslant \varepsilon$ and $\mathrm{D}_{\delta_0}(\mathrm{d}'|\mathrm{d}) \leqslant 1 + \nu$. Define*

$$\xi(r) := 384(1 + \pi)\frac{r^4}{\delta_0^4} \text{ for all } r \geqslant 0.$$

Then, for all $\delta \geqslant \delta_0$, letting $\Upsilon := (\delta\nu) \vee \varepsilon$ and $\mathrm{r}_1 := \mathrm{sdr}_{\delta+2\varepsilon}(K', \mathrm{d}')$, if $\xi(\mathrm{r}_1)\Upsilon < \mathrm{r}_1$, then

$$\mathrm{sdr}_\delta(K, \mathrm{d}) \leqslant \mathrm{sdr}_{\delta+2\varepsilon}(K', \mathrm{d}') + \xi(\mathrm{r}_1)\Upsilon.$$

A proof of Proposition 4.6 is given in Appendix B.2. Note that the condition $\mathrm{d_H}(K', K) \leqslant \varepsilon$ may be relaxed via $\mathrm{d_H}(K'|K) \leqslant \varepsilon$, where $\mathrm{d_H}(K'|K) := \sup_{x \in K'} \mathrm{d}(x, K)$. Also, under the assumptions of Proposition 4.6, let us remark that if $\mathrm{sdr}_{\delta+2\varepsilon}(K', \mathrm{d}')$ is finite, then so is $\mathrm{sdr}_\delta(K, \mathrm{d})$ with $\mathrm{sdr}_\delta(K, \mathrm{d}) \leqslant 2\,\mathrm{sdr}_{\delta+2\varepsilon}(K', \mathrm{d}')$.

Proposition 4.6 can be symmetrized to get the following two-sided control.

Corollary 4.7. *Let $0 < \delta_0 < \delta_1$ and $\varepsilon, \nu > 0$. Assume that both $\mathrm{d_H}(K', K) \leqslant \varepsilon$ and $\mathrm{D}_{\delta_0}(\mathrm{d}', \mathrm{d}) \leqslant 1 + \nu$. Then, for any $\delta \in (\delta_0 + 2\varepsilon, \delta_1 - 2\varepsilon)$, it holds*

$$\mathrm{sdr}_{\delta-2\varepsilon}(K, \mathrm{d}) - \xi_0 \Upsilon \leqslant \mathrm{sdr}_\delta(K', \mathrm{d}') \leqslant \mathrm{sdr}_{\delta+2\varepsilon}(K, \mathrm{d}) + \xi_0 \Upsilon$$

with $\xi_0 := \xi(2\,\mathrm{sdr}_{\delta_1}(K, \mathrm{d}))$ and $\Upsilon := (\nu\delta) \vee \varepsilon$, provided that $\xi_0 \Upsilon \leqslant 2\,\mathrm{sdr}_{\delta_1}(K, \mathrm{d})$.

Corollary 4.7 is proven in Appendix B.2. It ensures that the spherical distortion radius enjoys an interleaving property. That is the SDR of (K, d) at scale δ may be framed by the SDR of an approximation (K', d') at scales $\delta \pm \varepsilon$. This interleaving property is a common thread with the μ-reach (see, e.g., [16, Theorem 3.4]) and the λ-reach ([19, Theorem 3]), that is not enough to ensure consistent estimation. In fact, for the two aforementioned quantities, consistency may be proved with the additional assumption of $\mu \mapsto \mathrm{rch}_\mu(K)$ (resp. $\lambda \mapsto \lambda$-reach) are continuous at the targeted μ (resp. λ).

As opposed to the μ-reach the λ-reach, the SDR is also stable with respect to its the scale parameter δ. Next, we prove that $\delta \mapsto \mathrm{sdr}_\delta(K, \mathrm{d})$ is continuous over a fixed range $(0, \Delta^*)$ under mild structural assumptions on (K, d). These assumptions will be easily checked in the model $\mathcal{C}^k_{\mathrm{rch}_{\min}, \mathbf{L}}$, hence ensuring consistency of the subsequent reach estimator.

Assumption A1. We say that $K \subset \mathbb{R}^D$ is *spreadable* if there exist $\Delta_0 > 0$, $\varepsilon_0 > 0$, and $C_0 > 0$ such that for all $x, y \in K$ such that $\|x - y\| \leqslant \Delta_0$ and all $\varepsilon \leqslant \varepsilon_0$, there exists a point $a \in K$ such that either

- $\|a - y\| \leqslant \varepsilon$ and $\|x - a\| \geqslant \|x - y\| + C_0\varepsilon$, or
- $\|a - x\| \leqslant \varepsilon$ and $\|y - a\| \geqslant \|x - y\| + C_0\varepsilon$.

Assumption A1 requires that every point y of K may be locally pushed away from any (close enough) point $x \in K$. In particular, this means that K is nowhere discrete. In the manifold case, this pushing may be carried out using the exponential map (see Proposition 6.2).

Assumption A2. We say that (K, d) is *sub-Euclidean* if there exist $C_1 > 0$ and $\Delta_1 > 0$ such that for all $x, y \in K$ such that $\|x - y\| \leqslant \Delta_1$, we have $\mathrm{d}(x, y) \leqslant C_1\|x - y\|$.

Assumption A2 requires that the distance locally compares with the ambient Euclidean distance. This essentially means that the identity map $(K, \mathrm{d}) \to (K, \|\cdot\|)$ is locally Lipschitz. Such an assumption is automatically fulfilled whenever K has positive reach and $\mathrm{d} = \mathrm{d}_K$ (see [26]), with explicit constants in the manifold case (see Proposition 6.2) Whenever these two conditions are met, the spherical distortion radius of (K, d) can be proved to be locally Lipschitz in δ.

Theorem 4.8. *Assume that the metric space (K, d) fulfills Assumptions A1 and A2. Then $\delta \mapsto \mathrm{sdr}_\delta(K, \mathrm{d})$ is locally Lipschitz on $(0, \Delta^*)$ where*

$$\Delta^* := \min\left\{\Delta_0, \Delta_1, \sup\left\{\delta \geqslant 0 \mid \mathrm{sdr}_\delta(K, \mathrm{d}) < \infty\right\}\right\}.$$

More precisely, for all $0 < \delta_0 < \delta_1 < \Delta^*$, the map $\delta \mapsto \mathrm{sdr}_\delta(K, \mathrm{d})$ is L_0-Lipschitz on $[\delta_0, \delta_1]$ with

$$L_0 := \frac{192\mathrm{r}_1^3}{C_0\delta_0^3}\left(C_1 + \pi\frac{\mathrm{r}_1}{\delta_0}\right),$$

where $\mathrm{r}_1 := \mathrm{sdr}_{\delta_1}(K, \mathrm{d})$.

A proof of Theorem 4.8 can be found in Appendix B.2. Not only does it ensure that the spherical distortion radius at scale δ is continuous with respect to δ, that is enough to guarantee consistency, but it also allows to control its variation via an explicit local Lipschitz constant. Combined with Corollary 4.7, this allows to convert a bound between (K, d) and (K', d') in terms of Hausdorff distance and metric distortion into a bound on the SDR's at scale δ.

Theorem 4.9. *Let (K, d) fulfill Assumptions A1 and A2, and let (K', d') be such that $\mathrm{d_H}(K, K') \leqslant \varepsilon$ and $\mathrm{D}_{\delta_0}(\mathrm{d}, \mathrm{d}') \leqslant 1 + \nu$ for some $\delta_0 < \Delta^*$. Then, for all $\delta_1 \in (\delta_0, \Delta^*)$ and $\delta \in (\delta_0 + 2\varepsilon, \delta_1 - 2\varepsilon)$, provided that $\xi_0 \Upsilon \leqslant 2\,\mathrm{sdr}_{\delta_1}(K, \mathrm{d})$, we have*

$$\left|\mathrm{sdr}_\delta(K, \mathrm{d}) - \mathrm{sdr}_\delta(K', \mathrm{d}')\right| \leqslant \zeta_0 \Upsilon,$$

with $\Upsilon = (\delta\nu) \vee \epsilon$ and $\zeta_0 = \xi_0 + 2L_0$, where ξ_0 is defined in Corollary 4.7, L_0 is defined in Theorem 4.8.

We refer to Appendix B.2 for a proof of this result and to Figure 3 for a diagram of the scales at play. Note that the constant ζ_0 only depends on δ_0 and features of (K, d), that the assumptions are required on (K, d) only, and that the constraint on ε depends only on (K, d) as well.

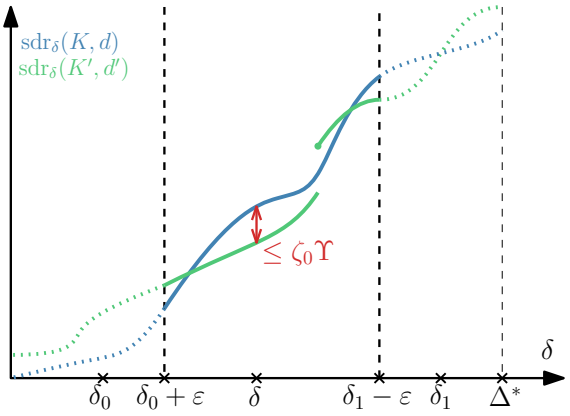


Figure 3: Plot of $\delta \mapsto \mathrm{sdr}_\delta$ for (K, d) and (K', d') in the context of Theorem 4.9. On the interval $(\delta_0 + \varepsilon, \delta_1 - \varepsilon)$, the two functions do not differ of more than $\zeta_0 \Upsilon$. Even though (K', d') might not be well-behaved, the regularity of $\delta \mapsto \mathrm{sdr}_\delta(K, d)$ (Theorem 4.8) is sufficient to insure stability.

The estimation of K is a now well-understood in the manifold case (see [2]). To obtain guarantees on the estimation of $\mathrm{sdr}_\delta(K, \mathrm{d}_K)$, it hence remains to investigate the estimation of d_K. This is the aim of the following section.

5 Optimal Metric Learning

5.1 Unsupervised Distance Metric Learning

As explained in the introduction, various learning tasks lead to the problem of estimation the shortest-pat distance d_K, via an estimator $\hat{d}$ on a sample of $K \subset \mathbb{R}^D$. Though, there is no canonical choice of loss for measuring the proximity of $\widehat{\mathrm{d}}$ to d_K. One could consider for instance the empirical sup-loss

$$\ell_n(\widehat{\mathrm{d}}|\mathrm{d}_K) := \sup_{x \neq y \in \mathbb{X}_n} \left| 1 - \frac{\widehat{\mathrm{d}}(x, y)}{\mathrm{d}_K(x, y)} \right|,$$

or the global sup-loss

$$\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_K) := \sup_{x\neq y\in K}\left|1-\frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_K(x,y)}\right|.$$

It might seem counter-intuitive to ask an estimator $\widehat{\mathrm{d}}$ of $\mathrm{d}_K : K\times K\to\mathbb{R}_+$ to be defined on the whole set $K\times K$, while the this domain is unknown. It actually is easy to extend any metric estimator to the whole space $\mathbb{R}^D\times\mathbb{R}^D$. Indeed, given such a metric estimation procedure $\widehat{\mathrm{d}}_n : \mathbb{X}_n\times\mathbb{X}_n\to\mathbb{R}_+$ that outputs a distance $\widehat{\mathrm{d}}_n[\mathbb{X}_n](x,y)$ between any pair of points of $\mathbb{X}_n$, we can define $\widetilde{\mathrm{d}}_n(x,y) := \widehat{\mathrm{d}}_{n+2}[\mathbb{X}_n,x,y](x,y)$ for all $(x,y)\in\mathbb{R}^D\times\mathbb{R}^D$. Informally this means that one can treat any given tuple of points (x,y) as actual data points in the estimation process, and that we are only interested in the behavior of the later when x and y are in fact from K.

The losses ℓ_n and ℓ_∞ are naturally multiplicative, in particular because the usual notions of distortions are multiplicative by nature (see Section 4). Indeed, the sup-loss $\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_K)$ being smaller than ν means that

$$\forall x,y\in K,\quad (1-\nu)\mathrm{d}_K(x,y)\leqslant\widehat{\mathrm{d}}(x,y)\leqslant(1+\nu)\mathrm{d}_K(x,y),$$

which is the usual way to quantify if the intrinsic metric is well-estimated. See for instance [42, 5]. When ν is small, it yields that $(K,\widehat{\mathrm{d}})$ is quasi-isometric to (K,d_K).

Remark 5.1. We emphasize the fact that the global sup-loss ℓ_∞ and the mutual metric distortion D_δ from Definition 4.4 are different in essence. Indeed, while the mutual metric distortion D_δ allows to compare different metrics on *different* subsets of $\mathbb{R}^D$, the sup-loss ℓ_∞ compares two distances defined on the *same* subset.

However, the global sup-loss and the mutual distortion metric may be related as follows. Consider K endowed with either $\widehat{\mathrm{d}}$ or d_K. Denote by $\mathrm{D}_{0^+}(\mathrm{d}_K,\widehat{\mathrm{d}}) := \lim_{\delta\to0}\mathrm{D}_\delta(\mathrm{d}_K,\widehat{\mathrm{d}})$. Then, straightforward computation entails

$$\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_K)+1\leqslant\mathrm{D}_{0^+}(\mathrm{d}_K,\widehat{\mathrm{d}})\leqslant(1-\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_K))_+^{-1}.$$

Hence, the global sup-loss $\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_K)$ is somehow an additive counterpart to the mutual distortion $\mathrm{D}_{0^+}(\widehat{\mathrm{d}},\mathrm{d}_K)$ in the case where $K=K'$. That is, when the support of the two metrics coincide in Definition 4.4, as already noticed in Remark 4.5.

When $K=M$ is a $\mathcal{C}^2$ submanifold of $\mathbb{R}^D$ of dimension d with reach bounded below, methods using neighborhood graphs such as Isomap provably estimate d_M at rate $O(n^{-2/3d})$ [6]. As we will show in Theorem 5.5, this rate is far from being optimal. To date, the best minimax lower bound in this setting is due to [5], who obtain a rate of order $\Omega(n^{-2/d})$ in the particular case of a deterministic design on $\mathcal{C}^2$ submanifolds. Actually, we can extend the result of [5] to our random design setting, and to general $\mathcal{C}^k$ submanifolds with $k\geqslant2$.

Theorem 5.2. *Assume that $f_{\min}\leqslant c_{d,k}/\operatorname{rch}_{\min}^d$ and $f_{\max}\geqslant C_{d,k}/\operatorname{rch}_{\min}^d$, and $L_j\geqslant C_{d,k}/\operatorname{rch}_{\min}^{j-1}$ for all $j\in\{2,\dots,k\}$. Then for n large enough,*

$$\inf_{\widehat{\mathrm{d}}}\sup_{P\in\mathcal{P}^k_{\operatorname{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})}\mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_M)]\geqslant\tilde{c}_{d,k,\operatorname{rch}_{\min}}\left(\frac{1}{n}\right)^{k/d},$$

where the infimum is taken over all measurable estimator $\widehat{\mathrm{d}}$ of d_M based on n samples.

This theorem is proved in Appendix C.1. As we shall prove shortly in Section 5.2, this lower-bound can be provided with a matching upper-bound up to $\log n$ factors (Theorem 5.4), and is thus optimal.

5.2 An optimal Approach of Metric Estimation

The existing unsupervised methods for metric learning are known to either have no theoretical guarantees, or to have a sub-optimal rate for estimating the intrinsic metric. As stated before, Isomap reaches a rate of $n^{-2/3d}$, which is very far from the theoretical lower-bound $n^{-k/d}$ shown in Theorem 5.2. Other methods, such as taking the shortest path distance over a Delaunay triangulation [5], are shown to attain a precision of $n^{-2/d}$ which is optimal for $\mathcal{C}^2$-model but not for $k \geqslant 3$. We propose here a fairly general approach that can output a family of minimax-optimal metric estimators. It relies on the following bound.

Proposition 5.3. *Let $K \subset \mathbb{R}^D$ be a set of positive reach $\mathrm{rch}(K) > 0$, and $K' \subset \mathbb{R}^D$ be any set such that $\mathrm{d_H}(K', K) < \varepsilon \leqslant \mathrm{rch}(K)/2$. Then,*

$$\ell_\infty(\mathrm{d}_{(K')^\varepsilon} | \mathrm{d}_K) \leqslant \frac{2\varepsilon}{\mathrm{rch}(K)},$$

where we recall that $(K')^\varepsilon = \left\{u \in \mathbb{R}^D \mid \mathrm{d}(u, K') \leqslant \varepsilon\right\}$, so that $K \subset (K')^\varepsilon$.

Proposition 5.3 is proved in Appendix C.2. It asserts that estimating geodesic distances of sets of positive reach is never harder than estimating the sets themselves in Hausdorff distance. Beyond the framework of closed manifold developed here, note that for the convex case $\mathrm{rch}(K) = \infty$, d_K coincides with the Euclidean metric, so that estimating d_K becomes trivial.

A significant consequence of Proposition 5.3 is that we can derive a consistent estimator of the intrinsic distance from any consistent estimator of the support, and with the same rate of convergence. In what follows, we write

$$\mathrm{d}_{\max} := \frac{5^d}{\omega_d f_{\min} \mathrm{rch}_{\min}^{d-1}}, \tag{11}$$

where ω_d is the volume of the d-dimensional unit ball. In Lemma C.2, the length $\mathrm{d}_{\max}$ is proved to be an upper bound on the geodesic diameter of the supports of any distribution in the model $\mathcal{P}^k_{\mathrm{rch}_{\min}, \mathbf{L}}(f_{\min}, f_{\max})$.

Theorem 5.4. *Let $k \geqslant 2$ and let $\widehat{M}$ be an estimator satisfying*

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min}, \mathbf{L}}(f_{\min}, f_{\max})} P^{\otimes n}(\mathrm{d_H}(\widehat{M}, M) \geqslant \varepsilon_n) \leqslant \eta_n,$$

for some positive sequences ε_n and η_n converging to 0. Then the metric estimator

$$\widehat{\mathrm{d}}(x, y) := \mathrm{d}_{\max} \wedge \mathrm{d}_{(\widehat{M}_{x,y})^{\varepsilon_n}}(x, y) \quad \textit{with} \quad \widehat{M}_{x,y} := \widehat{M} \cup \{x, y\},$$

which is defined for all $x, y \in \mathbb{R}^D$, satisfies

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min}, \mathbf{L}}(f_{\min}, f_{\max})} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}} | \mathrm{d}_M)] \leqslant \frac{2}{\mathrm{rch}_{\min}} \varepsilon_n + \left(1 + \frac{\mathrm{d}_{\max}}{\varepsilon_n}\right) \eta_n.$$

Theorem 5.4 is proved in Appendix C.2. A particular advantage of this result is that it does not require the estimator $\widehat{M}$ to have any geometric structure, nor to be regular in any sense. This contrasts sharply with [5], which extensively uses the structural properties of the intermediate estimator $\widehat{M}$. Theorem 5.4 is much more versatile, since here, $\widehat{M}$ could just as easily be anything

as a point cloud, a metric graph, a triangulation, or a union of polynomial patches. For instance, taking $\widehat{M} = \{X_1, \dots, X_n\}$ to be the observed data, we can take $\varepsilon_n = C(\log n/n)^{1/d}$ for C large enough yields $\eta_n \leqslant \varepsilon_n^2$ so that

$$\sup_{P \in \mathcal{P}^2_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_M)] \leqslant C_{\mathrm{rch}_{\min},d,f_{\min}} \left(\frac{\log n}{n}\right)^{1/d},$$

which is faster than the known rate of order $O(n^{-2/3d})$ for Isomap (see for instance [5, Eq (1.2)]). Now, taking $\widehat{M}$ to be a minimax optimal estimator of M for the Hausdorff loss — as that of [3], for instance — and $\varepsilon_n = C(\log n/n)^{k/d}$ for some large constant $C > 0$ yields $\eta_n \leqslant \varepsilon_n^2$ (see Lemma A.4), and a metric estimator $\widehat{\mathrm{d}}$ that achieves the following rate.

Theorem 5.5. *Let $\widehat{\mathrm{d}}$ be the estimator described in Theorem 5.4 built on top of $\widehat{M}$ described in Lemma A.4. Then for n large enough,*

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_M)] \leqslant C_{\mathrm{rch}_{\min},d,f_{\max},f_{\min},\mathbf{L},k} \left(\frac{\log n}{n}\right)^{k/d}.$$

In virtue of Theorem 5.2, this rate is minimax optimal up to $\log n$ factors.

6 Optimal Reach Estimation

6.1 Optimal Spherical Distortion Radius Estimation

Interesting as it is in its own right, we now investigate the estimation rates of the spherical distortion radius at scale $\delta > 0$. To obtain a minimax lower bound, we simply note that $\mathrm{sdr}_\delta(M, \mathrm{d}_M)$ coincides with $\mathrm{rch}(M)$ whenever $\mathrm{rch}(M) = \mathrm{wfs}(M)$ (Proposition 4.2). Hence, any lower bound for the estimation of $\mathrm{rch}(M)$ on a model over which $\mathrm{rch}(M) = \mathrm{wfs}(M)$ yields a lower bound for the estimation of $\mathrm{sdr}_\delta(M, \mathrm{d}_M)$. In application of Theorem 3.7 with $\alpha \geqslant 0$, this immediately gives the following lower bound.

Theorem 6.1. *Assume that $f_{\min} \leqslant c_{d,k}/\operatorname{rch}^d_{\min}$ and $f_{\max} \geqslant C_{d,k}/\operatorname{rch}^d_{\min}$, and $L_j \geqslant C_{d,k}/\operatorname{rch}^{j-1}_{\min}$ for all $j \in \{2, \dots, k\}$. Then for n large enough, for all $\delta \in (0, \mathrm{rch}_{\min})$,*

$$\inf_{\widehat{\mathrm{sdr}}_\delta} \sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} \mathbb{E}_{P^{\otimes}}[|\widehat{\mathrm{sdr}}_\delta - \mathrm{sdr}_\delta(M, \mathrm{d}_M)|] \geqslant \tilde{c}_{\mathrm{rch}_{\min},d,k} n^{-k/d}.$$

where the infimum is taken over all measurable estimators $\widehat{\mathrm{sdr}}_\delta$ of $\mathrm{sdr}_\delta(M, \mathrm{d}_M)$ based on n samples.

It turns out that this bound is optimal. To exhibit an estimator that achieves this rate, we take advantage of the Hausdorff and metric stability of the spherical distortion radius shown in Theorem 4.9. In order to apply it, we first need to check that Assumptions A1 and A2 are fulfilled for every manifolds in our models $\mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$.

Proposition 6.2. *Let $M \subset \mathbb{R}^D$ be a submanifold with bounded reach $\mathrm{rch}(M) > 0$. Then M satisfies Assumptions A1 and A2 with parameters*

$$\varepsilon_0 = \mathrm{rch}(M)/4, \quad \Delta_0 = \mathrm{rch}(M), \quad C_0 = 3/16, \quad \Delta_1 = \mathrm{rch}(M)/2 \quad \text{and} \quad C_1 = 2.$$

Proposition 6.2 is proven in Appendix D. In the vein of Theorem 5.4, and using the stability of the spherical distortion radius with respect to the pair (K, d), we can now build an estimator of $\mathrm{sdr}_\delta(M, \mathrm{d}_M)$ in a plug-in fashion over $\mathcal{C}^k$ submanifolds. Recall that when M is in $\mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$, and $\delta \in (0, \sqrt{2(D+1)/D}\,\mathrm{wfs}(M))$, then according to Propositions 4.2 and 3.2, and to Lemma C.2,

$$0 < \mathrm{rch}_{\min} \leqslant \mathrm{rch}(M) \leqslant \mathrm{sdr}_\delta(M, \mathrm{d}_M) \leqslant \mathrm{wfs}(M) \leqslant \sqrt{\frac{D}{2(D+1)}}\,\mathrm{diam}(M) \leqslant \mathrm{s}_{\max} < \infty,$$

where $\mathrm{s}_{\max} := \sqrt{D/(2(D+1))}\mathrm{d}_{\max}$, with $\mathrm{d}_{\max}$ being the constant introduced in (11).

Theorem 6.3. *Given $k \geqslant 2$, let $\widehat{M}$ be an estimator satisfying*

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} P^{\otimes n}(\mathrm{d_H}(M, \widehat{M}) \geqslant \varepsilon_n) \leqslant \eta_n$$

for some positive sequences ε_n, η_n converging to 0. Then, for any $\delta \in (0, \mathrm{rch}_{\min})$, the estimator $\widehat{\mathrm{sdr}}_\delta := \mathrm{sdr}_\delta(\widehat{M}, \widehat{\mathrm{d}}) \wedge \mathrm{s}_{\max}$, where $\widehat{\mathrm{d}}$ is defined in Theorem 5.4, satisfies

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\mathrm{sdr}}_\delta - \mathrm{sdr}_\delta(M, \mathrm{d}_M)| \leqslant C \left(\frac{\mathrm{s}^4_{\max}}{\delta^4} \varepsilon_n + \mathrm{s}_{\max} \eta_n \right).$$

We refer to Appendix D for a proof of this result.

Remark 6.4. In place of $\widehat{\mathrm{d}} = \mathrm{d}_{\widehat{M}^{\varepsilon_n}}$, one could actually plug any estimator $\widehat{\mathrm{d}}$ of the metric intoTheorem 6.3. In light of the stability result of Theorem 4.9, as long as $\widehat{\mathrm{d}}$ satisfies

$$\sup_{P \in \mathcal{P}^k} P^{\otimes n} \left(\mathrm{D}_\delta(\widehat{\mathrm{d}}, \mathrm{d}_M) \geqslant 1 + \frac{\varepsilon_n}{\delta} \right) \leqslant \eta_n,$$

the conclusion of Theorem 6.3 would still hold. This comes in handy, especially if one wants to input a computationally efficient distance estimator, such as shortest-path distance on a neigbhorhood graph [42] or on Delaunay triangulations [5].

Again, taking $\widehat{M}$ to be a minimax optimal estimator for the Hausdorff loss [3] outputs an estimator $\widehat{\mathrm{sdr}}_\delta$ of the spherical distortion radius satisfying

Theorem 6.5. *For all $\delta \in (0, \mathrm{rch}_{\min})$, with the construction of $\widehat{\mathrm{sdr}}_\delta$ above, we have that for n large enough,*

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\mathrm{sdr}}_\delta - \mathrm{sdr}_\delta(M, \mathrm{d}_M)| \leqslant C_{\mathrm{rch}_{\min},d,f_{\max},f_{\min},\mathbf{L},k} \frac{1}{\delta^4} \left(\frac{\log n}{n} \right)^{k/d},$$

and this rate is optimal in regard of Theorem 6.1.

Note the presence of the factor $1/\delta^4$ in the bound, which makes the rate diverge as $\delta \to 0$. This blowup is to be expected for the following reason. As δ goes to 0, the spherical distortion radius goes to the reach $\mathrm{rch}(M)$ (Proposition 4.2). Since the estimation of $\mathrm{rch}(M)$ cannot be faster than $n^{-(k-2)/d}$ (Theorem 3.7), the estimation rate of $\mathrm{sdr}_\delta(M, \mathrm{d}_M)$ must deteriorate in some way as $\delta \to 0$.

6.2 Optimal Reach Estimation

In light of Proposition 3.9 and (9), it only remains to combine the maximal curvature estimator and the spherical distortion radius estimator to obtain an estimator of the reach. Naely, we let $\widehat{M}$ be the minimax-Hausdorff estimator of Lemma A.4. According to the very same Lemma A.4, there exists $c_{\mathrm{rch}_{\min},d,f_{\max},f_{\min},\mathbf{L},k} > 0$ such that denoting by

$$\varepsilon_n := c_{\mathrm{rch}_{\min},d,f_{\max},f_{\min},\mathbf{L},k} \left(\frac{\log n}{n}\right)^{k/d}, \tag{12}$$

there holds

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} P^{\otimes n}(\mathrm{d_H}(\widehat{M}, M) \geqslant \varepsilon_n) \leqslant \varepsilon_n^2. \tag{13}$$

We also let $\widehat{\mathrm{d}}$ be the estimator of the intrinsic distance of Theorem 5.4 from $\widehat{M}$ and ε_n. We let $\widehat{\mathrm{sdr}}_\delta := \mathrm{sdr}_\delta(\widehat{M}, \widehat{\mathrm{d}}) \wedge \mathrm{s}_{\max}$ for some $\delta \in (0, \mathrm{rch}_{\min})$ as in Theorem 6.3. Finally, we write

$$\widehat{\mathrm{rch}} := \widehat{R}_\ell \wedge \widehat{\mathrm{sdr}}_\delta.$$

The following Theorem 6.6 is a straightforward consequence of Theorems 3.8 and 6.3, inserted in the plugin strategy of Proposition 3.9 and (9).

Theorem 6.6. *The estimator* $\widehat{\mathrm{rch}}$ *described above with* $\delta = \mathrm{rch}_{\min}/2$ *satisfies*

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\mathrm{rch}} - \mathrm{rch}(M)| \leqslant C_{\mathrm{rch}_{\min},d,f_{\max},f_{\min},\mathbf{L},k} \left(\frac{\log n}{n}\right)^{(k-2)/d},$$

and, for all $\alpha > 0$,

$$\sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L},\alpha}(f_{\min},f_{\max})} \mathbb{E}_{P^{\otimes n}} |\widehat{\mathrm{rch}} - \mathrm{rch}(M)| \leqslant C_{\mathrm{rch}_{\min},d,f_{\max},f_{\min},\mathbf{L},k,\alpha} \left(\frac{\log n}{n}\right)^{k/d}.$$

As a conclusion, Theorems 3.7 and 6.6 assert that $\widehat{\mathrm{rch}}$ is minimax optimal, and that its rate of convergence adapts to whether $\mathrm{rch}(M)$ is attained by curvature (yielding the slower rate $O(n^{-(k-2)/d})$) or by a bottleneck (yielding the faster rate rate $O(n^{-k/d})$).

The computation of $\widehat{\mathrm{rch}}$ depends explicitly on the parameters of the models at two levels. First, in tuning the value of ε_n as in (12). Second, in choosing $\delta \in (0, \mathrm{rch}_{\min})$. These two dependencies may be circumvented by picking

$$\widetilde{\varepsilon}_n = \log n \left(\frac{\log n}{n}\right)^{k/d},$$

and $\delta_n = 1/\log n$. Then, for n large enough, both (13) and $\delta_n \in (0, \mathrm{rch}_{\min})$ will be fulfilled. The price to pay for this way-around to calibration of constants limits to multiplicative $\log n$ factors in the upper-bound of Theorem 6.6.

7 Conclusion and Further Prospects

We developed a general strategy for estimating the reach of a manifold M. It relies on two independent plugins, accountable for the estimation of the minimal curvature radius $R_\ell(M)$ and any

another set-defined feature size $\theta(M)$ that lie between the reach and the weak feature size. We then introduced and studied the spherical distortion radius, the estimation of which reduces to geodesic distance estimation, itself reducing to set estimation in Hausdorff distance. All the derived results are minimax optimal, as testified by associated matching lower bounds up to $\log n$ factors.

Geometrically, one should note that this overall method relies heavily on the local/global dichotomy of the reach for closed submanifolds [2]. Hence, it still remains unclear how to extend it to manifolds with boundary, even though their curvature and spherical distortion radius are likely to be estimated in a similar way [1].

On the statistical side, a major extension of the results would consist in allowing for additive noise. Recent works obtained Hausdorff estimation rates for the support [28, 4, 40] in such a noisy setting, so that the estimation of the spherical distortion radius inherits the same rates straightforwardly. In the same spirit as the iterated local polynomial fitting of [4], we expect that the same method could likewise lead to maximal curvature estimation.

Finally, since the main goal of this work was of minimax nature, we did not focus on the algorithmic properties of our estimators. As they stand, $\widehat{R}_\ell$ and $\widehat{\mathrm{sdr}}$ both require to compute a supremum over the union of continuous patches $\widehat{M}$, which is computationally prohibitive. Actually, one can easily show that taking the same supremum over a discretization of $\widehat{M}$ at scale $O\big(n^{-\beta/d}\big)$ – i.e. $O(n^\beta)$ points in total – yields estimation rates of order $O\big(n^{-(\beta\wedge(k-2))/d}\big)$ for $R_\ell(M)$, and $O\big(n^{-(\beta\wedge k)/d}\big)$ for $\mathrm{sdr}_\delta(M, \mathrm{d}_M)$. This suggests a possible estimation-computation tradeoff which one could take advantage of. Yet, this is not a fully satisfactory solution, as sdr_δ still requires to compute costly geodesic distances on a high-dimensional set. More globally, the quest for computationally efficient – yet optimal – geometric estimators in high dimensions is still in its infancy.

Acknowledgments

The authors would like to thank heartily *Chez Adel* for its unconditional warmth and creative atmosphere, and Vincent Divol for helpful discussions.

References

[1] Aamari, E., C. Aaron, and C. Levrard (2021, August). Minimax Boundary Estimation and Estimation with Boundary. *arXiv e-prints*, arXiv:2108.03135.

[2] Aamari, E., J. Kim, F. Chazal, B. Michel, A. Rinaldo, and L. Wasserman (2019). Estimating the reach of a manifold. *Electron. J. Stat. 13*(1), 1359–1399.

[3] Aamari, E. and C. Levrard (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics 47*(1), 177–204.

[4] Aizenbud, Y. and B. Sober (2021, May). Non-Parametric Estimation of Manifolds from Noisy Data. *arXiv e-prints*, arXiv:2105.04754.

[5] Arias-Castro, E. and P. A. Chau (2020, November). Minimax Estimation of Distances on a Surface and Minimax Manifold Learning in the Isometric-to-Convex Setting. *arXiv e-prints*, arXiv:2011.12478.

[6] Arias-Castro, E. and T. Le Gouic (2019). Unconstrained and curvature-constrained shortest-path distances and their approximation. *Discrete & Computational Geometry 62*(1), 1–28.

[7] Arias-Castro, E. and B. Pelletier (2013). On the convergence of maximum variance unfolding. *Journal of Machine Learning Research 14*(18), 1747–1770.

[8] Attali, D., A. Lieutier, and D. Salinas (2013). Vietorisrips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry 46*(4), 448–465. 27th Annual Symposium on Computational Geometry (SoCG 2011).

[9] Balakrishnan, S., A. Rinaldo, D. Sheehy, A. Singh, and L. Wasserman (2012, 21–23 Apr). Minimax rates for homology inference. In N. D. Lawrence and M. Girolami (Eds.), *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, Volume 22 of *Proceedings of Machine Learning Research*, La Palma, Canary Islands, pp. 64–72. PMLR.

[10] Berenfeld, C., J. Harvey, M. Hoffmann, and K. Shankar (2022). Estimating the reach of a manifold via its convexity defect function. *Discrete Comput. Geom. 67*(2), 403–438.

[11] Boissonnat, J.-D. and A. Ghosh (2014). Manifold reconstruction using tangential Delaunay complexes. *Discrete Comput. Geom. 51*(1), 221–267.

[12] Boissonnat, J.-D., A. Lieutier, and M. Wintraecken (2019a). The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *J. Appl. Comput. Topol. 3*(1-2), 29–58.

[13] Boissonnat, J.-D., A. Lieutier, and M. Wintraecken (2019b). The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of applied and computational topology 3*(1), 29–58.

[14] Bourgain, J. (1985). On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics 52*(1), 46–52.

[15] Chatpatanasiri, R., T. Korsrilabutr, P. Tangchanachaianan, and B. Kijsirikul (2010). A new kernelization framework for mahalanobis distance learning algorithms. *Neurocomputing 73*(10-12), 1570–1579.

[16] Chazal, F., D. Cohen-Steiner, and A. Lieutier (2006). A sampling theory for compact sets in Euclidean space. In *Computational geometry (SCG'06)*, pp. 319–326. ACM, New York.

[17] Chazal, F., D. Cohen-Steiner, and Q. Mérigot (2011). Geometric inference for probability measures. *Found. Comput. Math. 11*(6), 733–751.

[18] Chazal, F., M. Glisse, C. Labrure, and B. Michel (2014, 22–24 Jun). Convergence rates for persistence diagram estimation in topological data analysis. In E. P. Xing and T. Jebara (Eds.), *Proceedings of the 31st International Conference on Machine Learning*, Volume 32 of *Proceedings of Machine Learning Research*, Bejing, China, pp. 163–171. PMLR.

[19] Chazal, F. and A. Lieutier (2005, jul). The "λ-medial axis". *Graph. Models 67*(4), 304331.

[20] Chazal, F. and B. Michel (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence 4*.

[21] Chennuru Vankadara, L. and U. von Luxburg (2018). Measures of distortion for machine learning. *Advances in Neural Information Processing Systems 31*.

[22] Cholaquidis, A., R. Fraiman, and L. Moreno (2021, oct). Universally consistent estimation of the reach. *arXiv e-prints*, arXiv:2110.12208.

[23] Clarke, F. H., R. J. Stern, and P. R. Wolenski (1995). Proximal smoothness and the lower-C^2 property. *J. Convex Anal. 2*(1-2), 117–144.

[24] Cox, M. A. and T. F. Cox (2008). Multidimensional scaling. In *Handbook of data visualization*, pp. 315–347. Springer.

[25] Divol, V. (2021). Minimax adaptive estimation in manifold inference. *Electronic Journal of Statistics 15*(2), 5888 – 5932.

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

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

[28] Fefferman, C., S. Ivanov, M. Lassas, and H. Narayanan (2019, October). Fitting a manifold of large reach to noisy data. *arXiv e-prints*, arXiv:1910.05084.

[29] Genovese, C. R., M. Perone-Pacifico, I. Verdinelli, and L. Wasserman (2012a). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist. 40*(2), 941–963.

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

[31] Globerson, A. and S. Roweis (2005). Metric learning by collapsing classes. *Advances in neural information processing systems 18*.

[32] Guillaumin, M., J. Verbeek, and C. Schmid (2009). Is that you? metric learning approaches for face identification. In *2009 IEEE 12th international conference on computer vision*, pp. 498–505. IEEE.

[33] Hatcher, A. (2002). *Algebraic Topology*. Cambridge University Press.

[34] Hinton, G. E. and S. Roweis (2002). Stochastic neighbor embedding. *Advances in neural information processing systems 15*.

[35] Kim, A. K. H. and H. H. Zhou (2015). Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat. 9*(1), 1562–1582.

[36] Kwok, J. T. and I. W. Tsang (2003). Learning with idealized kernels. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 400–407.

[37] Nguyen, B., C. Morell, and B. De Baets (2017). Supervised distance metric learning through maximization of the jeffrey divergence. *Pattern Recognition 64*, 215–225.

[38] Niyogi, P., S. Smale, and S. Weinberger (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry 39*(1), 419–441.

[39] Poliquin, R. A., R. T. Rockafellar, and L. Thibault (2000). Local differentiability of distance functions. *Trans. Amer. Math. Soc. 352*(11), 5231–5249.

[40] Puchkin, N. and V. Spokoiny (2022). Structure-adaptive manifold estimation. *Journal of Machine Learning Research 23*(40), 1–62.

[41] Suárez, J. L., S. García, and F. Herrera (2021). A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges. *Neurocomputing 425*, 300–322.

[42] Tenenbaum, J. B., V. de Silva, and J. C. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science 290*(5500), 2319–2323.

[43] Trillos, N. G., D. Sanz-Alonso, and R. Yang (2019). Local regularization of noisy point clouds: Improved global geometric estimates and data analysis. *Journal of Machine Learning Research 20*(136), 1–37.

[44] Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application 5*(1), 501–532.

[45] Xing, E., M. Jordan, S. J. Russell, and A. Ng (2002). Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems 15*.

[46] Yang, L. and R. Jin (2006). Distance metric learning: A comprehensive survey. *Michigan State Universiy 2*(2), 4.

[47] Yu, B. (1997). Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pp. 423–435. Springer.

A Proofs of Section 3

A.1 Comparing Reaches, Weak Feature Size and Diameter

This Section is devoted to the Proof of Proposition 3.2, which goes as follows.

Proof of Proposition 3.2. For (i), recall that no closed compact submanifold can be contractible [33, Theorem 3.26]. Furthermore, [26, Theorem 4.8] and [16, Lemma 2.1] combined together yield that K^r is isotopic to K for all $r < \mathrm{wfs}(K)$. On the other hand whenever $r > \mathrm{Rad}(K)$ where $\mathrm{Rad}(K)$ is the radius of the smallest ball enclosing K, K^r is star-shaped with respect to any point of the non-empty intersection $\cap_{x\in M}\, \mathsf{B}(x,r)$. We conclude that $\mathrm{wfs}(K) \leqslant \mathrm{Rad}(K)$, Since $\mathrm{Rad}(K) < \infty$ because K is compact, we obtain $\mathrm{wfs}(K) < \infty$.

For (ii), the first two inequalities come from the definition of $\mathrm{rch}_\mu(K)$ (see (4)). The rightmost comes Jung's Theorem [27, Theorem 2.10.41], which asserts that $\mathrm{Rad}(K) \leqslant \sqrt{\frac{D}{2(D+1)}}\,\mathrm{diam}(K)$, and the fact that $\mathrm{wfs}(K) \leqslant \mathrm{Rad}(K)$ whenever $\mathrm{wfs}(K)$ is finite (same argument as for (i)). □

A.2 Minimax Lower Bound for μ-Reach Estimation

This Section is devoted to the proof of Theorem 3.3. It builds upon the possible discontinuities of the map $M \mapsto \mathrm{Med}_\mu(M)$ in Hausdorff distance. The exhibition of such a discontinuity can be done in dimension $d = 1$ and $D = 2$, and can then be generalized to arbitrary $1 \leqslant d < D$ by using symmetry and rotation arguments.

The building block of the construction is the following arc of curve. For all $\alpha \in (0, \pi/4]$, write $R_\alpha := 1/\sin(\alpha)$. Let also $\mathsf{C}_\alpha : [0,1] \to \mathbb{R}_+$ be defined as $\mathsf{C}_\alpha(t) := R_\alpha - \sqrt{R_\alpha^2 - t^2}$, which graph is an arc of circle of radius R_α and aperture α (see Figure 4). To be able to glue up smoothly α-turns like C_α with straight lines, we smooth it as follows.

Lemma A.1. *There exists $G_\alpha : [0,1] \to \mathbb{R}_+$ infinitely differentiable such that:*

1. $G_\alpha^{(\ell)}(0) = 0$ *for all* $\ell \geqslant 0$;
2. $G_\alpha(1) = \mathsf{C}_\alpha(1)$, $G'_\alpha(1) = \mathsf{C}'_\alpha(1)$ *and* $G_\alpha^{(\ell)}(1) = 0$ *for all* $\ell \geqslant 2$;
3. $\|G_\alpha^{(\ell)}\|_\infty \leqslant C_\ell / R_\alpha$ *for all* $\ell \geqslant 1$;
4. $G_\alpha(t) < \mathsf{C}_\alpha(t)$ *for all* $t \in (0,1)$;
5. G_α *is convex.*

See Figure 4 for a diagram of such a G_α. Let us first comment on the requirements on G_α. Items 1 and 2 say that G_α is a $\mathcal{C}^k$ interpolation between the two tangent lines of two points of C_α who are α-apart in term of polar coordinate. Item 3 says that the graph of G_α, once rescaled by $1/R_\alpha$, will be bounded in $\mathcal{C}^k$-norm for all k. Items 4 and 5 ensure well-behavior of the medial axes of our future construct (see Figure 5).

Proof of Lemma A.1. The following construction applies to general convex functions, although we restrict it to G_α for simplicity. Consider the piecewise linear map A_α given by the tangent lines of C_α at $t = 0$ and $t = 1$. That is, define $A_\alpha(t)$ for all $t \in \mathbb{R}$ by

$$\begin{aligned} A_\alpha(t) &:= \max\left\{\mathsf{C}_\alpha(0) + \mathsf{C}'_\alpha(0)t, \mathsf{C}_\alpha(1) + (t-1)\mathsf{C}'_\alpha(1)\right\} \\ &= \max\left\{0, \mathsf{C}_\alpha(1) + (t-1)\mathsf{C}'_\alpha(1)\right\}. \end{aligned}$$

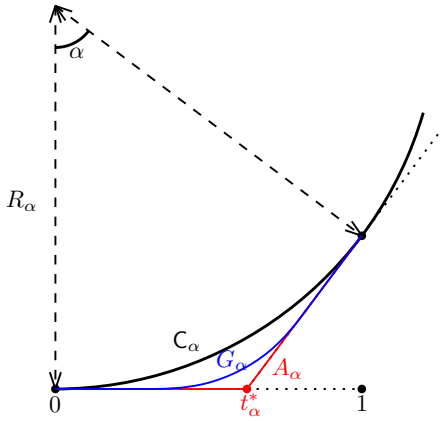


Figure 4: Construction for Lemma A.1: curves associated to C_α, A_α, and G_α.

As C_α is strictly convex, $A_\alpha < \mathsf{C}_\alpha$ on $\mathbb{R} \setminus \{0,1\}$. We also denote by t^*_α the (unique) point of non-differentiability of A_α, that is

$$t^*_\alpha := 1 - \frac{\mathsf{C}_\alpha(1)}{\mathsf{C}'_\alpha(1)} = R_\alpha \tan(\alpha/2).$$

Note by now that for all $\alpha \in (0, \pi/4)$, $1/2 \leqslant t^*_\alpha \leqslant 2 - \sqrt{2} \leqslant 6/10$. Given $h > 0$ to be chosen later, write $K_h(t) := h^{-1}K(t/h)$, where $K(t) := c_0 \exp(-1/(1-t^2))\mathbb{1}_{|t|<1}$ is a non-negative $\mathcal{C}^\infty$ kernel, and c_0 is chosen so that $\int_{\mathbb{R}} K = 1$. Finally, consider the convolution

$$G_\alpha(t) := \int_{\mathbb{R}} K_h(x) A_\alpha(t-x)\,\mathrm{d}x.$$

By smoothness of K_h and non-negativity of both K_h and A_α, $G_\alpha = K_h * A_\alpha$ is infinitely differentiable and non-negative. Also, since A_α is convex and K_h non-negative, G_α is convex (Item 5). Furthermore, one easily checks that outside the interval $[t^*_\alpha - h, t^*_\alpha + h]$, G_α coincides with A_α. Hence, if $h \leqslant 1/4$, we have $[t^*_\alpha - h, t^*_\alpha + h] \subset [1/2 - 1/4, 6/10 + 1/4] = [1/4, 17/20]$, so that Items 1 and 2 holds directly.

To check that $G_\alpha < \mathsf{C}_\alpha$ on $(0,1)$, fix $t \in (0,1)$. If $t \notin [t^*_\alpha - h, t^*_\alpha + h]$, $G_\alpha(t) = A_\alpha(t) < \mathsf{C}_\alpha(t)$ by construction. If $t \in [t^*_\alpha - h, t^*_\alpha + h]$, we have $G_\alpha(t) \leqslant G_\alpha(t^*_\alpha + h) = h\mathsf{C}'_\alpha(1)$. But on the other hand, $\mathsf{C}_\alpha(t) \geqslant \mathsf{C}_\alpha(t^*_* - h) > \mathsf{C}_\alpha(1/4)$. Hence, we do have $G_\alpha(t) < \mathsf{C}_\alpha(t)$ as soon as $h \leqslant 1/100$, since $\mathsf{C}_\alpha(1/4)/\mathsf{C}'_\alpha(1) > 1/100$ for all $\alpha \in (0, \pi/4)$. This yields Item 4.

Finally, letting $h = h_0 = 1/100$, we obtain for all $\ell \geqslant 1$ and $t \in [0,1]$,

$$|G^{(\ell)}_\alpha(t)| = \left|K^{(\ell)}_h * \mathsf{C}_\alpha(t)\right| \leqslant \left\|K^{(\ell)}_h\right\|_\infty \left\|\mathsf{C}_\alpha\right\|_\infty \leqslant C_\ell \mathsf{C}_\alpha(1) \leqslant C_\ell / R_\alpha,$$

which yields Item 3 and concludes the proof. □

Given $R > 0$, we now let $G_{\alpha,R}$ be the curve obtained by dilating homogeneously the graph of G_α by a scale factor R/R_α. We extend the construction of these smooth α-turns for $\alpha \in (\pi/4, \pi]$: for this, we glue two $G_{\alpha/2,R}$ or four $G_{\alpha/4,R}$ to define $G_{\alpha,R}$.

Proposition A.2. *Assume that for all $j \in \{2, \dots, k\}$, $L_j \geqslant C_{d,k}/\operatorname{rch}_{\min}^{j-1}$ for $C_{d,k} > 0$ large enough. Then for all $\mu \in [0,1)$ and $\varepsilon > 0$ small enough, there exist $M, M' \in \mathcal{C}^k_{\operatorname{rch}_{\min},\mathbf{L}}$ such that:*

– $|\operatorname{rch}_\mu(M) - \operatorname{rch}_\mu(M')| \geqslant c_{d,k} \operatorname{rch}_{\min}$;

– $c'_{d,k}\,\mathrm{rch}^d_{\min} \leqslant \mathrm{vol}_d(M) \wedge \mathrm{vol}_d(M') \leqslant \mathrm{vol}_d(M) \vee \mathrm{vol}_d(M') \leqslant C''_{d,k}\,\mathrm{rch}^d_{\min}$;

– $\mathrm{vol}_d(M \triangle M') \leqslant C'''_{d,k}\,\mathrm{rch}^d_{\min}\,\varepsilon$.

Proof of Proposition A.2. For small enough (and arbitrarily small) $\varepsilon > 0$, we let $\alpha \in [0, \pi]$ be such that $\sin\big((\alpha+\varepsilon)/2\big)^2 = 1 - \mu^2$. Such an α always exists since $\mu^2 < 1$. Given $\Delta, R_0, R_1 > 0$ to be chosen later, we glue smooth turns from Lemma A.1 with straight lines to create a $\mathcal{C}^k$ closed curve in $\mathbb{R}^2$, as shown in Figure 5. Then, we obtain a $\mathcal{C}^k$ closed d-dimensional submanifold M_α of $\mathbb{R}^{d+1}$, with a symmetry of revolution with respect to the horizontal axis of Figure 5.

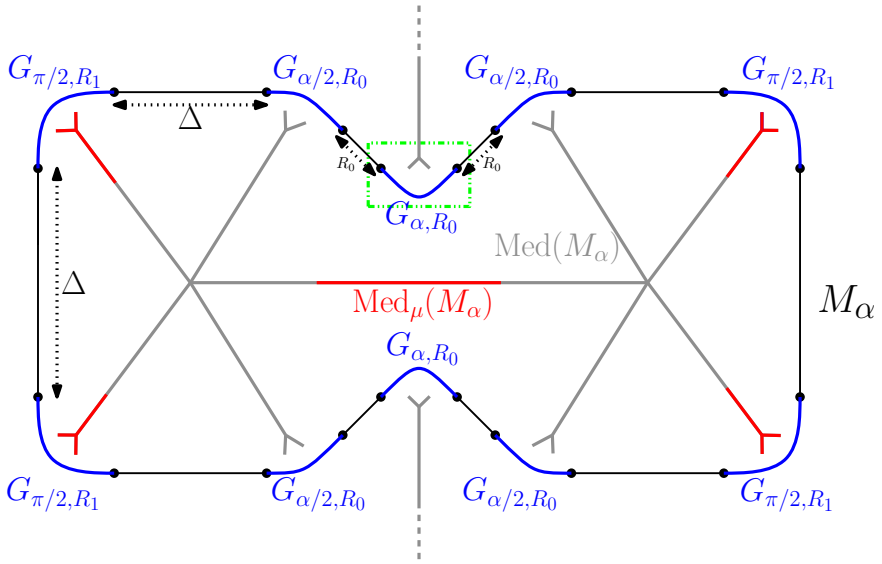


Figure 5: Construction of M_α in the proof of Proposition A.2.

By construction, if $\Delta \geqslant 8R_0$, then M_α has local parametrizations on top of its tangent spaces (see Definition 2.1) with $L_j \leqslant C_{d,k}/(\Delta \wedge R_0 \wedge R_1)^{j-1}$ for all $j \geqslant 2$, and has volume $\mathrm{vol}_d(M_\alpha) \leqslant C_{d,k}(\Delta \vee R_0 \vee R_1)^d$ and $\mathrm{vol}_d(M_\alpha) \geqslant c_{d,k}(\Delta \wedge R_0 \wedge R_1)^d$.

We now examine the structure of the medial axis and the reach of M_α. If $u \in \mathrm{Med}(M_\alpha)$ is a point on the medial axis, rotational symmetry yields that two of its projections points must lie either:

– In a plane containing its horizontal axis of symmetry (i.e. Figure 5). As a result, its distance to M_α cannot be smaller than the smallest reach of each of its parts $G_{\pi/2,R_1}, G_{\alpha/2,R_0}$ and G_{α,R_0}, so that $\mathrm{d}(u, M_\alpha) \geqslant c_{d,k}R_0 \wedge R_1$.

– In a d-plane orthogonal to the horizontal axis. By rotational invariance, this forces u to be on this axis of symmetry. As a result, $\mathrm{d}(u, M_\alpha) \geqslant \Delta/2 - 3R_0 \geqslant c_{d,k}\Delta$ since $\Delta \geqslant 8R_0$.

In all, we get $\mathrm{rch}(M_\alpha) \geqslant c_{d,k}(\Delta \wedge R_0 \wedge R_1)$.

We now examine the μ-reach of M_α. By definition, if $u \in \mathrm{Med}_\mu(M_\alpha)$ has two nearest neighbors $x, y \in M_\alpha$, the angle between $(u-x)$ and $(u-y)$ must be at most $2\arcsin(\sqrt{1-\mu^2})$. As a result, a single branch of M_α between the two arcs of $G_{\alpha/2,R_0}$ cannot not generate any point of the μ-medial axis, since α has been chosen so that $\alpha < 2\arcsin(\sqrt{1-\mu^2})$. Hence, for Δ, R_1 large enough compared to R_0, we have $\mathrm{rch}_\mu(M_\alpha) \geqslant c'_{d,k}(\Delta \wedge R_1)$.

Finally, we build M'_α from M_α by bumping the curve near G_{α,R_0} as shown in Figure 6 (while still preserving the radial symmetry as before). The manifold M'_α satisfies the same regularity

conditions at M_α. Furthermore, M_α and M'_α only differ on a set of volume $\mathrm{vol}_d(M_\alpha \triangle M'_\alpha) \leqslant C_{d,k}(\Delta \vee R_1)^{d-1}(R_0\varepsilon)$.

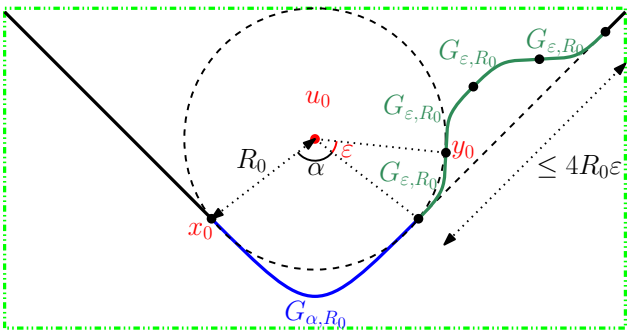


Figure 6: Local bump of M'_α for Proposition A.2, in the boxed area of Figure 5.

With this extra bump, we create a point $u_0 \in \mathrm{Med}(M'_\alpha)$ that has two nearest neighbors $x_0, y_0 \in M'_\alpha$ at distance R_0, with angle between $(u_0 - x_0)$ and $(u_0 - y_0)$ equal to $\alpha' = \alpha + \varepsilon$, which satisfies $\sin(\alpha'/2)^2 = 1 - \mu^2$. As a result, $u_0 \in \mathrm{Med}_\mu(M'_\alpha)$, so that $\mathrm{rch}_\mu(M'_\alpha) \leqslant \|u_0 - y_0\| = R_0$. In particular, we have

$$|\,\mathrm{rch}_\mu(M_\alpha) - \mathrm{rch}_\mu(M'_\alpha)| \geqslant c'_{d,k}(\Delta \wedge R_1) - R_0.$$

The proof is hence complete by setting $M = M_\alpha$ and $M' = M'_\alpha$, with $R_1 = \Delta = R_0/c'_{d,k}$ and $R_0 = \mathrm{rch}_{\min}/c_{d,k}$ for small enough $c_{d,k}, c'_{d,k} > 0$. □

Proof of Theorem 3.3. From Proposition A.2, for $\varepsilon > 0$ small enough, take $M, M' \in \mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$ such that $|\,\mathrm{rch}_\mu(M) - \mathrm{rch}_\mu(M')| \geqslant c_{d,k}\,\mathrm{rch}_{\min}$, $c'_{d,k}\,\mathrm{rch}^d_{\min} \leqslant \mathrm{vol}_d(M), \mathrm{vol}_d(M') \leqslant C_{d,k}\,\mathrm{rch}^d_{\min}$, and $\mathrm{vol}_d(M \triangle M') \leqslant C'_{d,k}\,\mathrm{rch}^d_{\min}\,\varepsilon$. Let us denote by P and P' the uniform distributions over M and M' respectively. Elementary calculations directly yield that

$$\mathrm{TV}(P, P') \leqslant \frac{\mathrm{vol}_d(M \triangle M')}{\mathrm{vol}_d(M) \vee \mathrm{vol}_d(M')} \leqslant C''_{d,k}\varepsilon.$$

Furthermore, since $c'_{d,k}\,\mathrm{rch}^d_{\min} \leqslant \mathrm{vol}_d(M) \wedge \mathrm{vol}_d(M') \leqslant \mathrm{vol}_d(M) \vee \mathrm{vol}_d(M') \leqslant C_{d,k}\,\mathrm{rch}^d_{\min}$, we obtain that $P, P' \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$ as soon as $f_{\min} \leqslant 1/(C_{d,k}\,\mathrm{rch}^d_{\min})$ and $f_{\max} \geqslant 1/(c'_{d,k}\,\mathrm{rch}^d_{\min})$. As a result, for all $n \geqslant 1$, Le Cam's Lemma [47] yields

$$\begin{aligned}
\inf_{\widehat{r}_\mu} \sup_{P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min},f_{\max})} & \mathbb{E}_{P^{\otimes n}}\left[|\widehat{r}_\mu - \mathrm{rch}_\mu(M)|\right] \\
&\geqslant \frac{1}{2}|\,\mathrm{rch}_\mu(M) - \mathrm{rch}_\mu(M')|\big(1 - \mathrm{TV}(P,P')\big)^n \\
&\geqslant c_{d,k}\,\mathrm{rch}_{\min}(1-\varepsilon)^n.
\end{aligned}$$

As this construction is valid for all $\varepsilon > 0$ small enough, we obtain the result by letting ε tend to zero. □

A.3 Maximal Curvature Estimation

This section is devoted to the proof of Theorem 3.8. It is based on a careful investigation of the local polynomial fitting procedure described in [2]. First, recall that from [2, Lemma 2], if $M \in \mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$,

$y \in M$ and $y' \in \mathsf{B}\left(y, \frac{L_2 \wedge \mathrm{rch}_{\min}}{4}\right) \cap M$, we may write

$$y' - y = \pi_y^*(y'-y) + \mathbb{T}_y^{(2),*}(\pi_y^*(y'-y)^{\otimes 2}) + \ldots + \mathbb{T}_y^{(k-1),*}(\pi_y^*(y'-y)^{\otimes k-1}) + R_y^{(k)}(y'-y), \tag{14}$$

where $\pi_y^* := \pi_{T_yM}$, $\mathbb{T}_y^{(j),*}$ are j-multilinear maps from T_yM to $\mathbb{R}^D$, and $R_y^{(k)}$ satisfies

$$\left\| R_y^{(k)}(y'-y) \right\| \leqslant C t_*^{k-1} \left\| y' - y \right\|^k,$$

where $t_* = \max_{2 \leqslant j \leqslant k, y \in M} \|T_y^{(j),*}\|_{\mathrm{op}}^{\frac{1}{j-1}} \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}}$. As assessed by [2, Lemma 2], the polynomial decomposition expressed in (14) allows to recover the curvature tensor via $\mathbb{I}_y M = \mathbb{T}_y^{(2),*}$. Following [2], we estimate this curvature tensor via the second term of the polynomial decomposition provided by local fit to data points (6). To this aim, a slight adaptation of [2, Lemma 3] is needed, that allows to translate quality of approximation in terms of Hausdorff distance to guarantees on the monomial terms.

Lemma A.3. *Set $h_0 = (\tau_{min} \wedge L_2^{-1})/8$ and $h \leqslant h_0$. Let $M \in \mathcal{C}^k_{\tau_{min},\mathbf{L}}$, $x_0 = y_0 + z_0$, with $y_0 \in M$ and $\|z_0\| \leqslant \sigma \leqslant h/4$. Denote by $\pi^*_{y_0}$ the orthogonal projection onto $T_{y_0}M$, and by $\mathbb{T}^{(2),*}_{y_0}, \ldots, \mathbb{T}^{(k-1),*}_{y_0}$ the multilinear maps given by (14).*

Let $x = y + z$ be such that $y \in M$, $\|z\| \leqslant \sigma \leqslant h/4$ and $x \in \mathsf{B}(x_0, h)$. We also let π be an orthogonal projection, and $\mathbb{T}^{(2)}, \ldots, \mathbb{T}^{(k-1)}$ be multilinear maps that satisfy

$$\left(\max_{2 \leqslant j \leqslant k-1} \|\mathbb{T}^{(j)}\|_{\mathrm{op}}^{\frac{1}{j-1}} \right) \vee t^* \leqslant t,$$
$$th \leqslant \frac{1}{4},$$

for some $t \geqslant 0$. Then it holds

$$x - x_0 - \pi(x - x_0) - \sum_{j=2}^{k-1} \mathbb{T}^{(j)}(\pi(x-x_0)^{\otimes j}) = \sum_{j=1}^{k} \mathbb{T}^{(j),\prime}_{y_0}(\pi^*_{y_0}(y - y_0)^{\otimes j}) + R^{(k)}_{y_0}(x - x_0),$$

where $\mathbb{T}^{(j),\prime}_{y_0}$ are j-linear maps, and $\|R^{(k)}_{y_0}(x-x_0)\| \leqslant C\left(\sigma + h^k(t_*^{k-1} + t^k h)\right)$, where C depends on d, k, $\mathrm{rch}_{\min}$, L_2,..., L_k. Moreover, we have

$$\begin{aligned} \mathbb{T}^{(1),\prime}_{y_0} &= (\pi^*_{y_0} - \pi), \\ \mathbb{T}^{(2),\prime}_{y_0} &= (\pi^*_{y_0} - \pi) \circ \mathbb{T}^{(2),*}_{y_0} + (\mathbb{T}^{(2),*}_{y_0} \circ \pi^*_{y_0} - \mathbb{T}^{(2)} \circ \pi), \end{aligned}$$

and, if $\pi = \pi^*_{y_0}$ and $\mathbb{T}^{(j)} = \mathbb{T}^{(j),*}_{y_0}$ for all $j \in \{2, \ldots, k-1\}$, then $\mathbb{T}^{(j),\prime}_{y_0} = 0$ for all $j \in \{1, \ldots, k\}$.

The proof of Lemma A.3 is deferred to Section A.4. To ensure that our local curvature estimators allow to approximate the maximal curvature of M, we have to ensure that the sample covers M well enough. That is the aim of the following Lemma.

Lemma A.4 ([2, Appendix, Lemma B.7 & Section 5.1.4])**.** *Let $P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$. Write $\mathbb{X}_n$ for an i.i.d. n-sample drawn from P. Let $h = \left(C_{d,k} \frac{f_{\max}^2}{f_{min}^3} \frac{\log n}{n}\right)^{1/d}$, for $C_{d,k}$ large enough. Then,*

for n large enough so that $h \leqslant \mathrm{rch}_{\min}/4$, with probability at least $1 - 2\left(\frac{1}{n}\right)^{2k/d}$, it holds

$$\begin{aligned} &\mathrm{d_H}\left(M, \mathbb{X}_n\right) \leqslant h/4, \\ &\mathrm{d_H}(M, \hat{M}) \leqslant C_{d,k,\mathrm{rch}_{\min},\mathbf{L}}(t^*)^{k-1} \left(\frac{f_{\max}^{2+\frac{d}{2k}} \log n}{f_{\min}^{3+\frac{d}{2k}} n} \right)^{k/d}, \end{aligned}$$

where $\hat{M}$ denotes the union of local polynomial patches

$$\hat{M} := \bigcup_{i=1}^{n} \hat{\Psi}_i \left(\mathsf{B}_{\hat{T}_i}(0, 7h/8) \right)$$

defined by (6) and (7), and $t_* = \max_{y \in M, 2 \leqslant j \leqslant k} \|\mathbb{T}_y^{(j),*}\|_{\mathrm{op}}^{\frac{1}{j-1}} \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}}$ as in Lemma A.3.

Equipped with these two lemmas, we are in position to prove Theorem 3.8.

Proof of Theorem 3.8. Based on Lemma A.4, for $h = \left(C_{d,k} \frac{f_{\max}^2 \log n}{f_{\min}^3 n} \right)^{\frac{1}{d}}$, given $i \in \{1, \dots, n\}$, we denote by $\hat{\Psi}_i$ the polynomial estimator around X_i defined by

$$\hat{\Psi}_i(v) := X_i + v + \sum_{j=2}^{k-1} \hat{\mathbb{T}}_i^{(j)}(v^{\otimes j}),$$

for all $v \in \hat{T}_i$. Setting

$$\hat{M} := \bigcup_{i=1}^{n} \hat{\Psi}_i \left(\mathsf{B}_{\hat{T}_i}(0, 7h/8) \right),$$

we have that with probability larger than $1 - 2\left(\frac{1}{n}\right)^{\frac{2k}{d}}$,

$$\mathrm{d_H}(\hat{M}, M) \leqslant C_{d,k,\mathrm{rch}_{\min},\mathbf{L}}(t_*)^{k-1} \left(\frac{f_{\max}^{2+\frac{d}{2k}} \log n}{f_{\min}^{3+\frac{d}{2k}} n} \right)^{\frac{k}{d}} := \varepsilon_1, \tag{15}$$

for n large enough, according to Lemma A.4. In what follows we settle on the probability event of Lemma A.4. In particular, denoting by

$$\hat{t} = \max_{1 \leqslant i \leqslant n} \max_{2 \leqslant j \leqslant k-1} \|\hat{\mathbb{T}}_i^{(j)}\|_{\mathrm{op}}^{\frac{1}{j-1}},$$

note that [2, Section 5.1.2] ensures that $\hat{t} \vee t_* \leqslant t \leqslant 1/(4h)$, for some fixed t, provided n is large enough.

We let $i \in \{1, \dots, n\}$, $v \in \mathsf{B}_{\hat{T}_i}(0, h/4)$, and intend to approximate $\mathrm{I\!I}_{\pi_M(\hat{\Psi}_i(v))}$. To do so, we consider the following polynomial expansion centered at v: for $u \in \mathsf{B}_{\hat{T}_i}(0, h/4)$,

$$\hat{\Psi}_i(v+u) - \hat{\Psi}_i(v) = u + \sum_{j=2}^{k-1} j \hat{\mathbb{T}}_i^{(j)} \left(v^{\otimes j-1} \otimes u \right) + \sum_{j=2}^{k-1} \sum_{r=j}^{k-1} \binom{r}{j} \hat{\mathbb{T}}_i^{(r)} \left(v^{\otimes r-j} \otimes u^{\otimes j} \right). \tag{16}$$

First we deduce from (16) an estimate for the tangent space at $\pi_M(X_i + v)$, as well as a coordinate system. Namely, we let

$$\begin{aligned}\hat{J}_{i,v} \colon \hat{T}_i &\longrightarrow \hat{J}_{i,v}(\hat{T}_i) \\ u &\longmapsto u + \sum_{j=2}^{k-1} j \hat{\mathbb{T}}_i^{(j)} \left(v^{\otimes j-1} \otimes u \right).\end{aligned}$$

Note that since $th \leqslant 1/4$, we have

$$\begin{aligned}\|\hat{J}_{i,v}(u) - u\| &\leqslant \sum_{j=2}^{k-1} j \left(\frac{th}{4} \right)^{j-1} \|u\| \\ &\leqslant \left(\sum_{j=1}^{\infty} j \left(\frac{th}{4} \right)^{j-1} - 1 \right) \|u\| \\ &\leqslant \left(\left(\frac{1}{1 - \frac{th}{4}} \right)^2 - 1 \right) \|u\| \leqslant \frac{\|u\|}{2},\end{aligned}$$

so that $\hat{J}_{i,v}$ is full-rank. In what follows we write $\hat{T}_{i,v} := \mathrm{Im}(\hat{J}_{i,v})$ and $\hat{\pi}_{i,v} := \pi_{\hat{T}_{i,v}}$. We now may express (16) in terms of the coordinate system given by $\hat{T}_{i,v}$:

$$\hat{\Psi}_i(v + u) - \hat{\Psi}_i(v) = \hat{J}_{i,v}(u) + \sum_{j=2}^{k-1} \widetilde{\mathbb{T}}_{i,v}^{(j)}(\hat{J}_{i,v}(u)^{\otimes j}), \tag{17}$$

where the symmetric tensor of order j centered at v, $\widetilde{\mathbb{T}}_{i,v}^{(j)}$, is defined by

$$\widetilde{\mathbb{T}}_{i,v}^{(j)}(w^{\otimes j}) := \sum_{r=j}^{k-1} \binom{r}{j} \hat{\mathbb{T}}_i^{(r)} \left(v^{\otimes r-j} \otimes \hat{J}_{i,v}^{-1}(w)^{\otimes j} \right),$$

for $w \in \hat{T}_{i,v}$. As well, since $th \leqslant \frac{1}{4}$, we may write

$$\begin{aligned}\left\| \widetilde{\mathbb{T}}_{i,v}^{(j)} \right\|_{\mathrm{op}} &\leqslant \sum_{r=j}^{k-1} \binom{r}{j} (3/2)^j t^{r-1} \left(\frac{h}{4} \right)^{r-j} \\ &\leqslant \left(\sum_{r=j}^{\infty} \binom{r}{j} \left(\frac{th}{4} \right)^{r-j} \right) (3/2)^j t^{j-1} \\ &\leqslant \left(\frac{1}{1 - \frac{th}{4}} \right)^j (3/2)^j t^{j-1} \leqslant (3/2)^{2j} t^{j-1},\end{aligned}$$

so that $\max_{2 \leqslant j \leqslant k-1} \left\| \widetilde{\mathbb{T}}_{i,v}^{(j)} \right\|_{\mathrm{op}}^{\frac{1}{j-1}} \leqslant \widetilde{t} \leqslant \left(\frac{3}{2} \right)^4 t$.

In particular, the bilinear form $\widetilde{\mathbb{T}}_{i,v}^{(2)} : \hat{T}_{i,v} \times \hat{T}_{i,v} \to \mathbb{R}^D$ may be expressed by

$$\widetilde{\mathbb{T}}_{i,v}^{(2)}(w^{\otimes 2}) := \sum_{j=2}^{k-1} \binom{j}{2} \hat{\mathbb{T}}_i^{(j)} \left(v^{\otimes j-2} \otimes \hat{J}_{i,v}^{-1}(w)^{\otimes 2} \right)$$

for all $w \in \hat{T}_{i,v}$. Our second fundamental form estimator at $\pi_M(\hat{\Psi}_i(v))$ is then defined by

$$\hat{\mathbb{T}}^{(2)}_{i,v} := \widetilde{\mathbb{T}}^{(2)}_{i,v} \circ \hat{\pi}_{i,v} - \hat{\pi}_{i,v} \circ \widetilde{\mathbb{T}}^{(2)}_{i,v} \circ \hat{\pi}_{i,v},$$

where with a slight abuse of notation, $\mathbb{T} \circ \pi(u) := \mathbb{T}\big(\pi(u)^{\otimes 2}\big)$. Note that composition with $\hat{\pi}_{i,v}$ is performed to ensure that $\hat{\mathbb{T}}^{(2)}_{i,v}$ ranges into $\hat{T}^{\perp}_{i,v}$.

Our final max-curvature estimator can now be defined as

$$\hat{R}_\ell^{-1} := \max_{1 \leqslant i \leqslant n} \max_{v \in \mathsf{B}_{\hat{T}_i}(h/4)} \left\| \hat{\mathbb{T}}^{(2)}_{i,v} \right\|_{\mathrm{op}}.$$

First, we intend to show that, for a given $v \in \mathsf{B}_{\hat{T}_i}(h/4)$, $\hat{\mathbb{T}}^{(2)}_{i,v}$ is close to $\mathrm{I\!I}_{y_0}$, for some $y_0 \in M$. To do so, we let $u \in \mathsf{B}_{\hat{T}_i}(0, h/4)$, $x := \hat{\Psi}_i(v+u), x_0 := \hat{\Psi}_i(v)$, and $\widetilde{P}^{(r:k-1)}_{i,v} := \sum_{j=r}^{k-1} \widetilde{\mathbb{T}}^{(j)}_{i,v}$. Then, we have the decomposition

$$\begin{aligned}
\hat{J}_{i,v}(u) &= \hat{\pi}_{i,v}(x - x_0) - \sum_{j=2}^{k-1} \hat{\pi}_{i,v} \circ \widetilde{\mathbb{T}}^{(j)}_{i,v}(\hat{J}_{i,v}(u)^{\otimes j}) \\
&= \hat{\pi}_{i,v}(x - x_0) - \sum_{j=2}^{k-1} \hat{\pi}_{i,v} \circ \widetilde{\mathbb{T}}^{(j)}_{i,v} \left[\left(\hat{\pi}_{i,v}(x - x_0) - \hat{\pi}_{i,v} \circ \widetilde{P}^{(2:k-1)}_{i,v}(\hat{J}_{i,v}(u)) \right)^{\otimes j} \right] \\
&= \hat{\pi}_{i,v}(x - x_0) + \sum_{j=2}^{k} \mathbb{T}^{(j),''}_{i,v}(\hat{\pi}_{i,v}(x - x_0)^{\otimes j}) + R^{(k)}_{i,v}(x - x_0),
\end{aligned}$$

with $\mathbb{T}^{(2),''}_{i,v} = -\hat{\pi}_{i,v} \circ \widetilde{\mathbb{T}}^{(2)}_{i,v}$, higher order tensors satisfying $\left\| \mathbb{T}^{(j),''}_{i,v} \right\|_{\mathrm{op}} \leqslant C_k \widetilde{t}^{j-1} \leqslant C_k t^{j-1}$, and remainder term $\|R^{(k)}_{i,v}\| \leqslant C_k t^k h^{k+1}$. Plugging the above inequalities into (17) yields

$$x - x_0 = \hat{\pi}_{i,v}(x - x_0) + \mathbb{T}^{(2)}_{i,v}\left(\hat{\pi}_{i,v}(x - x_0)^{\otimes 2}\right) + \sum_{j=3}^{k} \mathbb{T}^{(j)}_{i,v}(\hat{\pi}_{i,v}(x - x_0)^{\otimes j}) + R^{(k),'}_{i,v}(x - x_0), \tag{18}$$

with $\mathbb{T}^{(2)}_{i,v} = \widetilde{\mathbb{T}}^{(2)}_{i,v} - \hat{\pi}_{i,v} \circ \widetilde{\mathbb{T}}^{(2)}_{i,v}$, $\left\| \mathbb{T}^{(j)}_{i,v} \right\|_{\mathrm{op}} \leqslant C_k t^{j-1}$, and $\left\| R^{(k),'}_{i,v}(x - x_0) \right\| \leqslant C_k t^k h^{k+1}$.

Then, according to Lemma A.4, there exists $y_0 \in \mathsf{B}(X_i, \frac{8}{7 \times 4} h) \cap M$ such that $\|y_0 - x_0\| \leqslant \varepsilon_1$, where ε_1 is defined by (15). We further have

$$\begin{aligned}
\|v - \hat{\pi}_i(y_0 - X_i)\| &\leqslant \varepsilon_1 + \left\| \hat{\Psi}_i(v) - (X_i + v) \right\| \\
&\leqslant \varepsilon_1 + \left\| \sum_{j=2}^{k-1} \hat{\mathbb{T}}^{(j)}_i(v^{\otimes j}) \right\| \\
&\leqslant \varepsilon_1 + h/16 \\
&\leqslant h/8,
\end{aligned}$$

since $\hat{t} h \leqslant \frac{1}{4}$, provided that $\varepsilon_1 \leqslant h/16$ (satisfied for n large enough). Next, if $z \in \mathsf{B}\left(y_0, \frac{h}{8}\right) \cap M$, we have

$$\begin{aligned}
\|\hat{\pi}_i(z - X_i) - v\| &\leqslant \|\hat{\pi}_i(z - y_0)\| + \|v - \hat{\pi}_i(y_0 - X_i)\| \\
&\leqslant h/4,
\end{aligned}$$

so that, writing $x_z := \hat{\Psi}_i(\hat{\pi}_i(z - X_i))$, it holds $\|z - x_z\| \leqslant \varepsilon_1$ and (18) applies. Next, provided $C_k t h < 1/4$ and $C_k t \geqslant t_*$ (satisfied whenever n is large enough), Lemma A.3 yields that

$$x_z - x_0 - \left(\hat{\pi}_{i,v}(x_z - x_0) + \mathbb{T}^{(2)}_{i,v}(\hat{\pi}_{i,v}(x_z - x_0)^{\otimes 2}) + \sum_{j=3}^{k} \mathbb{T}^{(j)}_{i,v}(\hat{\pi}_{i,v}(x_z - x_0)^{\otimes j}) \right)$$
$$= \sum_{j=1}^{k} \mathbb{T}^{(j),\prime}_{i,v}(\pi^*_y(z - y_0)^{\otimes j}) + R^{(k)}_{y_0}(x_z - x_0),$$

so that

$$\left\| \sum_{j=1}^{k} \mathbb{T}^{(j),\prime}_{i,v}(\pi^*_y(z - y_0)^{\otimes j}) \right\| = \left\| R^{(k),\prime}_{i,v}(x_z - x_0) - R^{(k)}_{y_0}(x_z - x_0) \right\| \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1,$$

according to (18) and Lemma A.3, since $t^k h \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}}$. Using the development (16) and the inclusion $\mathsf{B}_{T_{y_0}M}(0, h/16) \subset \pi^*_{y_0}\left(\mathsf{B}(y_0, h/8) \cap M - y_0\right)$ from [2, Lemma 2] then entails

$$\left\| \sum_{j=1}^{k} \mathbb{T}^{(j),\prime}_{i,v}(\pi^*_{y_0}(w)^{\otimes j}) \right\| \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1,$$

for all $w \in \mathsf{B}_{T_{y_0}M}(0, h/16)$. Proceeding as in [2, Proof of Theorem 2], we get

$$\left\| \mathbb{T}^{(1),\prime}_{i,v} \right\|_{\mathrm{op}} \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-1},$$
$$\text{and } \left\| \mathbb{T}^{(2),\prime}_{i,v} \right\|_{\mathrm{op}} \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2}.$$

In turn, following [2, Proof of Theorem 4] entails

$$\left\| \hat{\mathbb{T}}^{(2)}_{i,v} \circ \hat{\pi}_{i,v} - \mathbb{T}^{(2),*}_{y_0} \circ \pi^*_{y_0} \right\|_{\mathrm{op}} \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2}.$$

Since $\mathbb{I}_{y_0} = \mathbb{T}^{(2),*}_{y_0}$ ([2, Lemma 2]), we deduce that

$$\max_{1 \leqslant i \leqslant n} \max_{v \in \mathsf{B}_{\hat{T}_j}(0,h/4)} \left\| \hat{\mathbb{T}}^{(2)}_{i,v} \circ \hat{\pi}_{i,v} \right\|_{\mathrm{op}} \leqslant \max_{y \in M} \|\mathbb{I}_y\|_{\mathrm{op}} + C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2}. \tag{19}$$

Conversely, since $X_1, \ldots, X_n$ is a $(h/4)$-covering of M onto the probability event described in Lemma A.4, we deduce that for all $y \in M$, there exists $i_0 \in \{1, \ldots, n\}$ such that $\|X_{i_0} - y\| \leqslant h/4$. In particular, we have

$$v := \hat{\pi}_{i_0,v}(y - X_{i_0}) \in \mathsf{B}_{\hat{T}_{i_0}}(0, h/4).$$

Proceeding as above similarly leads to

$$\left\| \hat{\mathbb{T}}^{(2)}_{i_0,v} \circ \hat{\pi}_{i_0,v} - \mathbb{I}_y \circ \pi^*_y \right\|_{\mathrm{op}} \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2},$$

so that

$$\max_{y \in M} \|\mathbb{I}_y\|_{\mathrm{op}} \leqslant \max_{1 \leqslant i \leqslant n} \max_{v \in \mathsf{B}_{\hat{T}_i}(0,h/4)} \left\| \hat{\mathbb{T}}^{(2)}_{i,v} \circ \hat{\pi}_{i,v} \right\|_{\mathrm{op}} + C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2}. \tag{20}$$

Combining (19) and (20) yields that for n large enough,

$$\left| \hat{R}_\ell - R_\ell(M) \right| \leqslant R_\ell(M)^2 C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} \varepsilon_1 h^{-2},$$

which concludes the proof. □

A.4 Proof of Lemma A.3

Proof of Lemma A.3. We follow the proof of [2, Lemma 3]. Without loss of generality we take $y_0 = 0$, so that $\|y\| \leqslant 3h/2$. Let $z' = z - z_0$, so that $\|z'\| \leqslant h/2$. We write

$$\begin{aligned}
&x - x_0 - \pi(x - x_0) - \sum_{j=2}^{k} \mathbb{T}^{(j)}(\pi(x - x_0)^{\otimes j}) \\
&\quad = y + z' - \pi(y + z') - \sum_{j=2}^{k} \mathbb{T}^{(j)}((\pi(y) + \pi(z'))^{\otimes j}) \\
&\quad = y + z' - \pi(y + z') - \sum_{j=2}^{k} \left[\mathbb{T}^{(j)}(\pi(y)^{\otimes j}) + \sum_{r=0}^{j-1} \binom{j}{r} \mathbb{T}^{(j)} \left(\pi(y)^{\otimes r} \otimes \pi(z')^{\otimes j-r} \right) \right].
\end{aligned}$$

Since, for any $j \geqslant 2$ and $r \in \{0, \dots, j-1\}$,

$$\begin{aligned}
\left\| \mathbb{T}^{(j)} \left(\pi(y)^{\otimes r} \otimes \pi(z')^{\otimes j-r} \right) \right\| &\leqslant t^{j-1} (3h/2)^r (2\sigma)^{j-r} \\
&\leqslant C_k \sigma t^{j-1} h^{j-1} \leqslant C_k \sigma,
\end{aligned}$$

we may write

$$\begin{aligned}
&x - x_0 - \pi(x - x_0) - \sum_{j=2}^{k} \mathbb{T}^{(j)}(\pi(x - x_0)^{\otimes j}) \\
&\qquad = y - \pi(y) - \sum_{j=2}^{k} \mathbb{T}^{(j)}(\pi(y)^{\otimes j}) + R^{(k),\prime}(x - x_0), \quad (21)
\end{aligned}$$

where $\left\| R^{(k),\prime}(x - x_0) \right\| \leqslant C_k \sigma$. Next, (14) entails

$$\begin{aligned}
y = \pi^*_{y_0}(y) + \mathbb{T}^{(2),*}_{y_0}(\pi^*_{y_0}(y)^{\otimes 2}) + \ldots + \mathbb{T}^{(k-1),*}_{y_0}(\pi^*_{y_0}(y)^{\otimes k-1}) \\
+ R^{(k),\prime\prime}_{y_0}(y),
\end{aligned}$$

with $\|R^{(k),\prime\prime}_{y_0}(y)\| \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} t_*^{k-1} h^k$. Denoting by

$$P^{*,(1:k-1)}_{y_0}(\pi^*_{y_0}(y)) := \pi^*_{y_0}(y) + \sum_{r=2}^{k-1} \mathbb{T}^{(r),*}_{y_0}(\pi^*_{y_0}(y)^{\otimes r}),$$

we deduce that

$$\begin{aligned}
&y - \pi(y) - \sum_{j=2}^{k} \mathbb{T}^{(j)}(\pi(y)^{\otimes j}) = \\
&\qquad P^{*,(1:k-1)}_{y_0}(\pi^*_{y_0}(y)) + R^{(k),\prime\prime}_{y_0}(y) - \pi \left(P^{*,(1:k-1)}_{y_0}(\pi^*_{y_0}(y)) + R^{(k),\prime\prime}_{y_0}(y) \right) \\
&\qquad\qquad - \sum_{j=2}^{k} \mathbb{T}^{(j)} \left(\pi \left[\left(P^{*,(1:k-1)}_{y_0}(\pi^*_{y_0}(y)) + R^{(k),\prime\prime}_{y_0}(y) \right) \right]^{\otimes j} \right).
\end{aligned}$$

Note that

$$\left\|\pi(R_{y_0}^{(k),''}(y))\right\| \leqslant \left\|R_{y_0}^{(k),''}(y)\right\| \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} t_*^{k-1} h^k.$$

Next, since $\|y\| \leqslant 3h/2$, it holds

$$\begin{aligned}\left\|P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right\| &\leqslant \sum_{r=1}^{k-1} t_*^{r-1} \left(\frac{3h}{2}\right)^r \\ &\leqslant \frac{3h}{2} \frac{1}{1-\frac{3t_*h}{2}} \leqslant 3h,\end{aligned}$$

so that, for all $j \in \{2,\dots,k\}$,

$$\begin{aligned}\left\|\mathbb{T}^{(j)}\left(\pi\left[\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) + R_{y_0}^{(k),''}(y)\right)\right]^{\otimes j}\right) - \mathbb{T}^{(j)}\left(\pi\left[\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right)\right]^{\otimes j}\right)\right\| \\ \leqslant t^{j-1}\sum_{r=1}^{j}\binom{j}{r}\left\|R_{y_0}^{(k),''}\right\|^r (3h)^{j-r} \\ \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} t^{j-1}h^j \max_{1\leqslant r\leqslant j} t_*^{(k-1)r} h^{(k-1)r} \\ \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} t^k h^{k+1}.\end{aligned}$$

Thus, we may write

$$\begin{aligned}y - \pi(y) - \sum_{j=2}^{k}\mathbb{T}^{(j)}(\pi(y)^{\otimes j}) = P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y)) - \pi\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right) \\ - \sum_{j=2}^{k}\mathbb{T}^{(j)}\left(\pi\left[\left(P_{y_0}^{*,(1:k-1)}(\pi_{y_0}^*(y))\right)\right]^{\otimes j}\right) + R_{y_0}^{(k),'''}(y),\end{aligned}$$

where $\left\|R_{y_0}^{(k),'''}(y)\right\| \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} h^k(t_*^{k-1} + t^k h)$. At last, for $j \in \{2,\dots,k\}$, and $r_1,\dots,r_j \in \{1,\dots,k-1\}$ such that $\sum_{s=1}^{j} r_s \geqslant k+1$, we have

$$\begin{aligned}\left\|\mathbb{T}^{(j)}\left(\bigotimes_{s=1}^{j}\pi\left(\mathbb{T}_{y_0}^{(r_s),*}\left(\pi_{y_0}^*(y)^{\otimes r_s}\right)\right)\right)\right\| &\leqslant t^{j-1}\prod_{s=1}^{j} t_*^{r_s-1}h^{r_s} \\ &\leqslant (th)^{(\sum_{s=1}^{j} r_s)-1} h \leqslant t^k h^{k+1},\end{aligned}$$

where $\mathbb{T}_{y_0}^{(1),*} = \pi_{y_0}^*$, with a slight abuse of notation. Hence, it holds

$$\begin{aligned}y - \pi(y) - \sum_{j=2}^{k}\mathbb{T}^{(j)}(\pi(y)^{\otimes j}) = (\pi_{y_0}^* - \pi\circ\pi_{y_0}^*)(y) + \mathbb{T}_{y_0}^{(2),*}(\pi_{y_0}^*(y)^{\otimes 2}) - \pi\left(\mathbb{T}_{y_0}^{(2),*}(\pi^*(y)^{\otimes 2})\right) \\ - \mathbb{T}^{(2)}\left(\left(\pi\circ\pi_{y_0}^*(y)\right)^{\otimes 2}\right) + \sum_{j=3}^{k}\mathbb{T}_{y_0}^{(j),'}\left(\pi_{y_0}^*(y)^{\otimes j}\right) + R_{y_0}^{(k),''''}(y),\end{aligned}$$

where $\left\|R_{y_0}^{(k),''''}(y)\right\| \leqslant C_{k,d,\mathrm{rch}_{\min},\mathbf{L}} h^k(t_*^{k-1} + t^k h)$. Plugging the above equation into (21) gives the result. □

B Proofs of Section 4

B.1 Comparing Reach, Weak Feature Size and Spherical Distortion Radius

Let us prove Proposition 4.2.

Proof of Proposition 4.2. The monotonicity follows trivially from the definition, and since by [13, Theorem 1], $\mathrm{sdr}_0(K, \mathrm{d}_K) = \mathrm{rch}(K, \mathrm{d}_K)$, there holds immediately that $\mathrm{sdr}_\delta(K, \mathrm{d}_K) \geqslant \mathrm{rch}(K)$ for any $\delta \geqslant 0$. Now take $\delta \leqslant \sqrt{2(D+1)/D}\,\mathrm{wfs}(K)$, and take z a critical point of K, so that $z \in \operatorname{conv}\Gamma$ where $\Gamma := \{x \in K \mid \|x - z\| = \mathrm{d}(z, K)\}$. Using Jung's theorem [27, Theorem 2.10.41], there holds

$$\mathrm{diam}(\Gamma) \geqslant \sqrt{\frac{2(D+1)}{D}}\,\mathrm{Rad}(\Gamma) = \sqrt{\frac{2(D+1)}{D}}\mathrm{d}(z, K) \geqslant \sqrt{\frac{2(D+1)}{D}}\,\mathrm{wfs}(K) \geqslant \delta$$

so that there exists two points $x, y \in \Gamma$ such that $\|x - y\| \geqslant \delta$. Furthermore, since the interior of $\mathsf{B}(z, \mathrm{wfs}(K))$ contains no point of K, there holds

$$\mathrm{d}_K(x, y) \geqslant \mathrm{d}_{\mathcal{S}(\mathrm{wfs}(K))}(x, y) > \mathrm{d}_{\mathcal{S}(r)}(x, y), \text{ for all } r > \mathrm{wfs}(K),$$

so that indeed $\mathrm{sdr}_\delta(K, \mathrm{d}_K) \leqslant \mathrm{wfs}(K)$. □

B.2 Stability Properties of the Spherical Distortion Radius

We now move to the proofs of the stability properties of the SDR. As a first step, we will need the following lemma on geodesic distances over spheres.

Lemma B.1. *Let $r, \varepsilon > 0$ and take $x, y, a, b \in K$ such that $\|x - y\| < 2r$ and*

$$\|a - b\| \leqslant \left(1 + \frac{A\varepsilon}{r}\right)\|x - y\|$$

for some $A > 0$. For all $\lambda > 0$, define

$$\zeta_\lambda = \max\left\{\frac{192r^3}{\|a - b\|^3}(\lambda + A\pi), 4A\right\}.$$

Then, for all $\zeta \geqslant \zeta_\lambda$ such that $\zeta\varepsilon \leqslant r$, there holds

$$\mathrm{d}_{\mathcal{S}(r+\zeta\varepsilon)}(a, b) \leqslant \mathrm{d}_{\mathcal{S}(r)}(x, y) - \lambda\varepsilon.$$

Proof of Lemma B.1. Notice that, denoting by $\rho = \|x - y\|$,

$$\mathrm{d}_{\mathcal{S}(r)}(x, y) = 2r \arcsin\left(\frac{\rho}{2r}\right) = \rho \times \varphi(2r/\rho) \quad \text{with} \quad \varphi(u) := u \arcsin(1/u).$$

The map φ is decreasing on $[1, \infty)$ and, using the development of

$$\arcsin(u) = \sum_{n=0}^{\infty}(2n)!u^{2n+1}/(2^{2n}n!^2(2n+1)),$$

we find that

$$\varphi'(u) = -\sum_{n=1}^{\infty}\frac{(2n)! \times 2n}{2^{2n}n!^2(2n+1)}\frac{1}{u^{2n+1}} \leqslant -\frac{1}{3u^3}.$$

Notice furthermore that, by assumption

$$\begin{aligned}\frac{2(r+\zeta\varepsilon)}{\|a-b\|} &\geqslant \frac{2(r+\zeta\varepsilon)}{(1+A\varepsilon/r)\|x-y\|} = \frac{1+\zeta\varepsilon/r}{1+A\varepsilon/r}\frac{2r}{\|x-y\|}\\ &\geqslant \left(1+\frac{\zeta\varepsilon}{2r}\right)\frac{2r}{\|x-y\|}\end{aligned}$$

where we used that $A \leqslant \zeta/4$, $\zeta\varepsilon \leqslant r$, and that $(1+u)/(1+u/4) \geqslant 1+u/2$ for $|u| \leqslant 2$. Now, as $\varphi \leqslant \pi/2$ and that $|\varphi'|$ is decreasing, we can write

$$\begin{aligned}\mathrm{d}_{\mathcal{S}(r+\zeta\varepsilon)}(a,b) &\leqslant \frac{A\varepsilon}{r}\|x-y\|\varphi\left(2(r+\zeta\varepsilon)/\|a-b\|\right) + \|x-y\|\varphi\left(2(r+\zeta\varepsilon)/\|a-b\|\right)\\ &\leqslant A\pi\varepsilon + \mathrm{d}_{\mathcal{S}(r)}(x,y)\\ &\qquad - \|x-y\| \times |\varphi'|\left(\frac{2(r+\zeta\varepsilon)}{\|a-b\|}\right) \times \left(\frac{2(r+\zeta\varepsilon)}{\|a-b\|} - \frac{2r}{\|x-y\|}\right)\\ &\leqslant \mathrm{d}_{\mathcal{S}(r)}(x,y) + A\pi\varepsilon - \frac{\|a-b\|^3}{3(2(r+\zeta\varepsilon))^3}\zeta\varepsilon\\ &\leqslant \mathrm{d}_{\mathcal{S}(r)}(x,y) + \left(A\pi - \frac{\|a-b\|^3}{192r^3}\zeta\right)\varepsilon,\end{aligned}$$

and using $\zeta \geqslant \zeta_\lambda$ ends the proof. □

We are now in position to prove Proposition 4.6 and Theorem 4.8.

Proof of Proposition 4.6. If $\mathrm{r}_1 = \infty$ there is nothing to show. Otherwise, notice that because $\mathrm{r}_1 \geqslant \delta_0/2$ by definition, there holds that

$$\xi(R) \geqslant \max\left\{\frac{192R^3}{\delta_0^3}\left(1+\pi\frac{2R}{\delta_0}\right), \frac{8R}{\delta_0}\right\} \geqslant 1$$

for all $R > \mathrm{r}_1$. Now, since $\xi(\mathrm{r}_1)\Upsilon < \mathrm{r}_1$, one can find $R > \mathrm{r}_1$ such that $\xi(R)\Upsilon < R$. By definition of r_1, there exist $x, y \in K'$ such that $\delta + 2\varepsilon \leqslant \|x-y\| < 2R$ and $\mathrm{d}_{\mathcal{S}(R)}(x,y) < \mathrm{d}'(x,y)$. Now, let $a, b \in K$ be two closest points (in Euclidean distance) from x and y such that $\mathrm{d}(a,b) = \mathrm{d}\left(\pi_K(\{x\}), \pi_K(\{y\})\right.$. Then

$$\delta \leqslant \|a-b\| \leqslant \|x-y\| + 2\varepsilon < 2R + 2\Upsilon \leqslant 2(R+\xi(R)\Upsilon)$$

and

$$\|a-b\| \leqslant \|x-y\| + 2\varepsilon \leqslant \left(1+\frac{2\Upsilon R}{\delta_0 R}\right)\|x-y\|.$$

We now can apply Lemma B.1 with $A = 2R/\delta_0$ and $\lambda = 1$ to find that

$$\begin{aligned}\mathrm{d}(a,b) &\geqslant \frac{1}{1+\nu}\mathrm{d}'(x,y)\\ &> \frac{1}{1+\nu}\mathrm{d}_{\mathcal{S}(R)}(x,y)\\ &\geqslant \frac{1}{1+\nu}\left(\mathrm{d}_{\mathcal{S}(R+\xi(R)\Upsilon)}(a,b) + \Upsilon\right)\\ &\geqslant \frac{1+\Upsilon/\delta}{1+\nu}\mathrm{d}_{\mathcal{S}(R+\xi(R)\Upsilon)}(a,b),\end{aligned}$$

where the last inequality uses that $\mathrm{d}_{\mathcal{S}(R+\xi(R)\Upsilon)}(a,b) \geqslant \|a-b\| \geqslant \delta$. At the end of the day, since $\Upsilon \geqslant \delta\nu$, we have $\mathrm{d}(a,b) > \mathrm{d}_{\mathcal{S}(R+\xi(R)\Upsilon)}(a,b)$, so that $\mathrm{sdr}_\delta(K,\delta) < R + \xi(R)\Upsilon$. Taking R to r_1 yields the result. □

Proof of Theorem 4.8. We take $\varepsilon > 0$ such that

$$\varepsilon < C_0\varepsilon_0, \quad \varepsilon < (\delta_1 - \delta_0)/2, \quad \text{and} \quad \varepsilon < r_0/L_0,$$

and take $\delta \in [\delta_0, \delta_1 - \varepsilon)$. We write $r_\delta := \mathrm{sdr}_\delta(K, \mathrm{d})$ and $r_{\delta+\varepsilon} := \mathrm{sdr}_{\delta+\varepsilon}(K, \mathrm{d})$ for short. Recall that $r_\delta \leqslant r_{\delta+\varepsilon}$. Now take $r \leqslant r_{\delta+\varepsilon} - L_0\varepsilon$, and two points $x, y \in K$ such that $\delta \leqslant \|x - y\| < 2r$ (if there are none, then $r \leqslant r_\delta$ automatically). If $\|x - y\| \geqslant \delta + \varepsilon$, then $\mathrm{d}(x, y) \leqslant \mathrm{d}_{\mathcal{S}(r)}(x, y)$ because $r \leqslant r_{\delta+\varepsilon}$. If now $\|x - y\| < \delta + \varepsilon$, since $\|x - y\| \leqslant \Delta_0$, we can use Assumption A1 and find a point $a \in K$ such that $\|a - y\| \leqslant \varepsilon/C_0$ and $\|x - a\| \geqslant \|x - y\| + \varepsilon \geqslant \delta + \varepsilon$. Now, since $r + L_0\varepsilon \leqslant r_{\delta+\varepsilon}$, it holds $\mathrm{d}(x, a) \leqslant \mathrm{d}_{\mathcal{S}(r+L_0\varepsilon)}(x, a)$. Furthermore, notice that

$$\|x - a\| \leqslant \|x - y\| + \frac{1}{C_0}\varepsilon \leqslant \left(1 + \frac{r_1\varepsilon}{C_0\delta_0 r}\right)\|x - y\|.$$

Using Assumption A2 and Lemma B.1 with $A = r_1/(C_0\delta_0)$ and $\lambda = C_1/C_0$, we find

$$\mathrm{d}(x, y) \leqslant \mathrm{d}(x, a) + \mathrm{d}(a, y) \leqslant \mathrm{d}_{\mathcal{S}(r+L_0\varepsilon)}(x, a) + \frac{C_1}{C_0}\varepsilon \leqslant \mathrm{d}_{\mathcal{S}(r)}(x, y),$$

so that in the end $r \leqslant r_\delta$. Taking r to $r_{\delta+\varepsilon} - L_0\varepsilon$ yields that $r_{\delta+\varepsilon} \leqslant r_\delta + L_0\varepsilon$, ending the proof. □

Finally, Corollary 4.7 follows as a direct corollary of Proposition 4.6.

Proof of Corollary 4.7. Since $\xi_0\varepsilon \leqslant 2\,\mathrm{sdr}_{\delta_1}(K, \mathrm{d})$, the radius $\mathrm{sdr}_{\delta_1}(K, \mathrm{d})$ is in particular finite so that, according to Proposition 4.6, $\mathrm{sdr}_\delta(K', \mathrm{d}') \leqslant 2\,\mathrm{sdr}_{\delta_1}(K, \mathrm{d})$ and, consequently, $\xi_1\Upsilon \leqslant \mathrm{sdr}_\delta(K', \mathrm{d}')$ and $\xi_2\Upsilon \leqslant \mathrm{sdr}_{\delta+2\varepsilon}(K, \mathrm{d})$, where $\xi_1 = \xi(\mathrm{sdr}_\delta(K', \mathrm{d}'))$ and $\xi_2 = \xi(\mathrm{sdr}_{\delta+2\varepsilon}(K, \mathrm{d}))$. Applying Proposition 4.6 twice – which is possible, since $\Upsilon \geqslant ((\delta - 2\varepsilon)\nu) \vee \varepsilon)$ –, we thus find

$$\mathrm{sdr}_{\delta-2\varepsilon}(K, \mathrm{d}) - \xi_1\Upsilon \leqslant \mathrm{sdr}_\delta(K', \mathrm{d}') \leqslant \mathrm{sdr}_{\delta+2\varepsilon}(K, \mathrm{d}) + \xi_2\Upsilon,$$

and we conclude by noticing that both ξ_1 and ξ_2 are less than ξ_0. □

Proof of Theorem 4.9. Using Corollary 4.7 and Theorem 4.8, one find that

$$\begin{aligned}\mathrm{sdr}_\delta(K', \mathrm{d}') &\leqslant \mathrm{sdr}_{\delta+2\varepsilon}(K, \mathrm{d}) + \xi_0\Upsilon \\ &\leqslant \mathrm{sdr}_\delta(K, \mathrm{d}) + 2L_0\varepsilon + \xi_0\Upsilon \\ &\leqslant \mathrm{sdr}_\delta(K, \mathrm{d}) + \zeta_0\Upsilon,\end{aligned}$$

and likewise for the lower bound. □

C Proofs of Section 5

C.1 Minimax Lower Bound for Metric Learning

We now turn towards the proof of Theorem 5.2. It relies on an adaptation of the classical Le Cam's argument [47] to the asymmetric loss ℓ_∞.

Lemma C.1. *Let $x, y \in \mathbb{R}^D$ and let M_0 and M_1 be two submanifolds of $\mathbb{R}^D$ such that $x, y \in M_0 \cap M_1$ and the uniform distribution P_0 (resp. P_1) on M_0 (resp. M_1) is in $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$. Then if $\mathrm{d}_{M_0}(x, y) \leqslant \mathrm{d}_{M_1}(x, y)$,*

$$\inf_{\widehat{\mathrm{d}}} \sup_{P \in \mathcal{P}^k} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_M)] \geqslant \frac{1}{2} \times \left|1 - \frac{\mathrm{d}_{M_0}(x, y)}{\mathrm{d}_{M_1}(x, y)}\right| \times (1 - \mathrm{TV}(P_0^{\otimes n}, P_1^{\otimes n})), \tag{22}$$

Proof of Lemma C.1. For brevity, we write $\mathcal{R}_n$ be the minimax risk appearing in the left-hand side of (22). First, we write

$$\begin{aligned}
\mathcal{R}_n &\geqslant \inf_{\widehat{\mathrm{d}}} \sup_{P\in\{P_0,P_1\}} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_M)] \\
&\geqslant \inf_{\widehat{\mathrm{d}}} \sup_{P\in\{P_0,P_1\}} \mathbb{E}_{P^{\otimes n}}\left[\left|1-\frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_M(x,y)}\right|\right] \\
&\geqslant \frac{1}{2}\inf_{\widehat{\mathrm{d}}}\left\{\mathbb{E}_{P_0^{\otimes n}}\left[\left|1-\frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_{M_0}(x,y)}\right|\right]+\mathbb{E}_{P_1^{\otimes n}}\left[\left|1-\frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_{M_1}(x,y)}\right|\right]\right\} \\
&\geqslant \frac{1}{2}\inf_{\widehat{\mathrm{d}}}\mathbb{E}_{P_0^{\otimes n}}\left[\left(\left|1-\frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_{M_0}(x,y)}\right|+\left|1-\frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_{M_1}(x,y)}\right|\right)\times\left(1\wedge\frac{\mathrm{d}P_1^{\otimes n}}{\mathrm{d}P_0^{\otimes n}}\right)\right].
\end{aligned}$$

But now, using that $\mathrm{d}_{M_0}(x,y)\leqslant\mathrm{d}_{M_1}(x,y)$, a simple computation shows that the functional

$$\delta\mapsto\left|1-\frac{\delta}{\mathrm{d}_{M_0}(x,y)}\right|+\left|1-\frac{\delta}{\mathrm{d}_{M_1}(x,y)}\right|$$

is minimal for $\delta=\mathrm{d}_{M_0}(x,y)$ so that

$$\begin{aligned}
\mathcal{R}_n &\geqslant \frac{1}{2}\mathbb{E}_{P_0^{\otimes n}}\left[\left|1-\frac{\mathrm{d}_{M_0}(x,y)}{\mathrm{d}_{M_1}(x,y)}\right|\times\left(1\wedge\frac{\mathrm{d}P_1^{\otimes n}}{\mathrm{d}P_0^{\otimes n}}\right)\right] \\
&= \frac{1}{2}\times\left|1-\frac{\mathrm{d}_{M_0}(x,y)}{\mathrm{d}_{M_1}(x,y)}\right|\times(1-\mathrm{TV}(P_0^{\otimes n},P_1^{\otimes n})),
\end{aligned}$$

which ends the proof. □

Proof of Theorem 5.2. Without loss of generality, we set the analysis in $\mathbb{R}^{d+1}\simeq\mathbb{R}^{d+1}\times\{0\}^{D-(d+1)}\subset\mathbb{R}^D$.

C.1.1 Submanifolds Construction

We let $M_0\subset\mathbb{R}^{d+1}$ be a submanifold of $\mathcal{C}^k_{2\,\mathrm{rch}_{\min},\mathbf{L}/2}$ such that it contains the cylinder

$$\left\{(s,z)\in\mathbb{R}^2\times\mathbb{R}^{d-1}\mid\|s\|=R\text{ and }\|z\|\leqslant 3R\right\}.$$

Such a manifold always exists as soon as $R\geqslant 2\,\mathrm{rch}_{\min}$ and L_j is large enough compared to $1/R^{j-1}$. For instance, one can design M_0 as a hypersurface of revolution obtained based on patches the interpolating curves of Lemma A.1.

In what follows, we denote any $x\in\mathbb{R}^{d+1}=\mathbb{R}^d\times\mathbb{R}$ as $x=(w,h)\in\mathbb{R}^d\times\mathbb{R}$. With this notation, we define, for $\varepsilon>0$ and $c>0$ to be chosen later,

$$\Phi_\varepsilon(x):=x+c\varepsilon^k K(w/\varepsilon)e_{d+1}\quad\text{where}\quad e_{d+1}=(0,\dots,0,1)\in\mathbb{R}^{d+1},$$

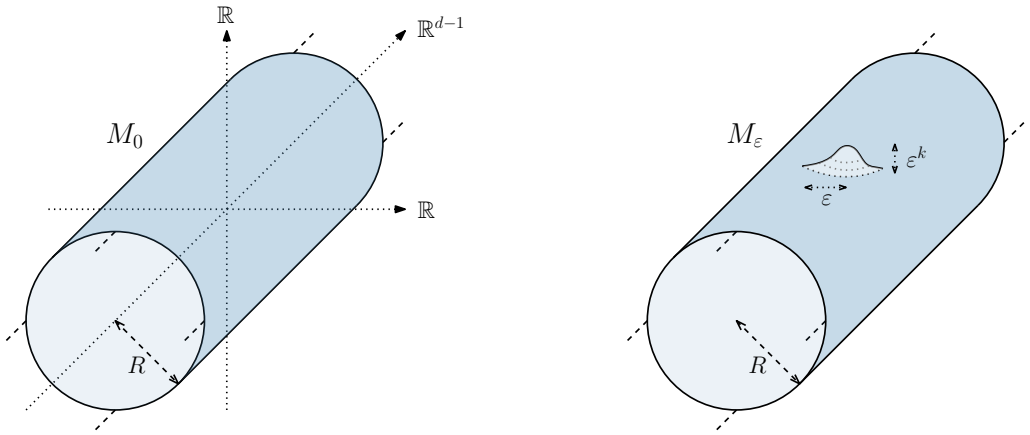
where $K(w)$ equals $\exp(-1/(1-\|w\|^2)_+)$ for $\|w\|<1$ and 0 otherwise.

For $\varepsilon\leqslant 1$ and c small enough, Φ_ε is a diffeomorphism of $\mathbb{R}^{d+1}$ with derivative bounded up to the order k. Using [3, Proposition A.4], we get that $M_\varepsilon:=\Phi_\varepsilon(M)$, the image of M_0 by Φ_ε, belongs to $\mathcal{C}^k_{\mathrm{rch}_{\min},\mathbf{L}}$ provided that c is small enough (depending on R) and $\varepsilon\leqslant cR$.

Locally around the apex $(0, R) \in \mathbb{R}^{d+1}$, M_0 can be seen as the graph of $\Psi_0(w) := \sqrt{R^2 - w_1^2}$, defined on $(-R, R) \times \mathsf{B}_{\mathbb{R}^{d-1}}(0, 3R)$, while M_ε is the graph of

$$\Psi_\varepsilon(w) := \Psi_0(w) + c\varepsilon^k K(w/\varepsilon).$$

Finally, we let $\bar{\Psi}_\varepsilon(w) := (w, \Psi_\varepsilon(w))$ and similarly define $\bar{\Psi}_0$. We refer to Figure 7 for a diagram of the situation.



(a) (b)

Figure 7: (a) The cylindrical section of M_0 used in the proof of Theorem 5.2, and (b) the perturbed submanifold M_ε.

C.1.2 Shortest-Path Properties

In this section, we seek to derive a lower bound on $|1 - \mathrm{d}_{M_0}(x, y)/\mathrm{d}_{M_\varepsilon}(x, y)|$, so as to apply Lemma C.1. For this, we will consider well-chosen $x, y \in M_0 \cap M_\varepsilon$ and derive a lower bound on $\mathrm{d}_{M_\varepsilon}(x, y) - \mathrm{d}_{M_0}(x, y)$.

We let $\ell < R$, and we pick $x := \bar{\Psi}_0(-\ell e_1)$ and $y := \bar{\Psi}_0(\ell e_1)$ where $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$. By construction, x and y belong to M_0. Furthermore, provided that $\ell \geqslant \varepsilon$, there holds that $x = \bar{\Psi}_\varepsilon(-\ell e_1)$ and $y = \bar{\Psi}_\varepsilon(\ell e_1)$ so that x and y are also in M_ε. We let $\gamma_\varepsilon : [-1, 1] \to M_\varepsilon$ be a shortest path in M_ε between x and y, parametrized at constant speed. We denote paths

$$w_\varepsilon := a_\varepsilon e_1 + b_\varepsilon := \pi_{\mathbb{R}^d \times \{0\}}(\gamma_\varepsilon),$$

where $b_\varepsilon \in \{0\} \times \mathbb{R}^{d-1}$. We refer to Figure 8 for a diagram of the situation. Several observations are in order.

- Since $w_\varepsilon(\pm 1) = \pm \ell e_1$, we have $a_\varepsilon(\pm 1) = \pm \ell$ and $b_\varepsilon(\pm 1) = 0$. Also, because γ_ε is a minimizing path, a_ε is nondecreasing, and $\|b_\varepsilon\|_\infty \leqslant \varepsilon$ (see Figure 8).
- Because γ_ε has constant speed on $[-1, 1]$, there holds

$$\|\gamma_\varepsilon'(t)\| = \frac{1}{2}\mathrm{d}_{M_\varepsilon}(x, y) \in [A_1 \ell, A_2 \ell], \quad \text{for all} \quad t \in [-1, 1], \tag{23}$$

  with A_1, A_2 depending on R only, uniformly on small ε.

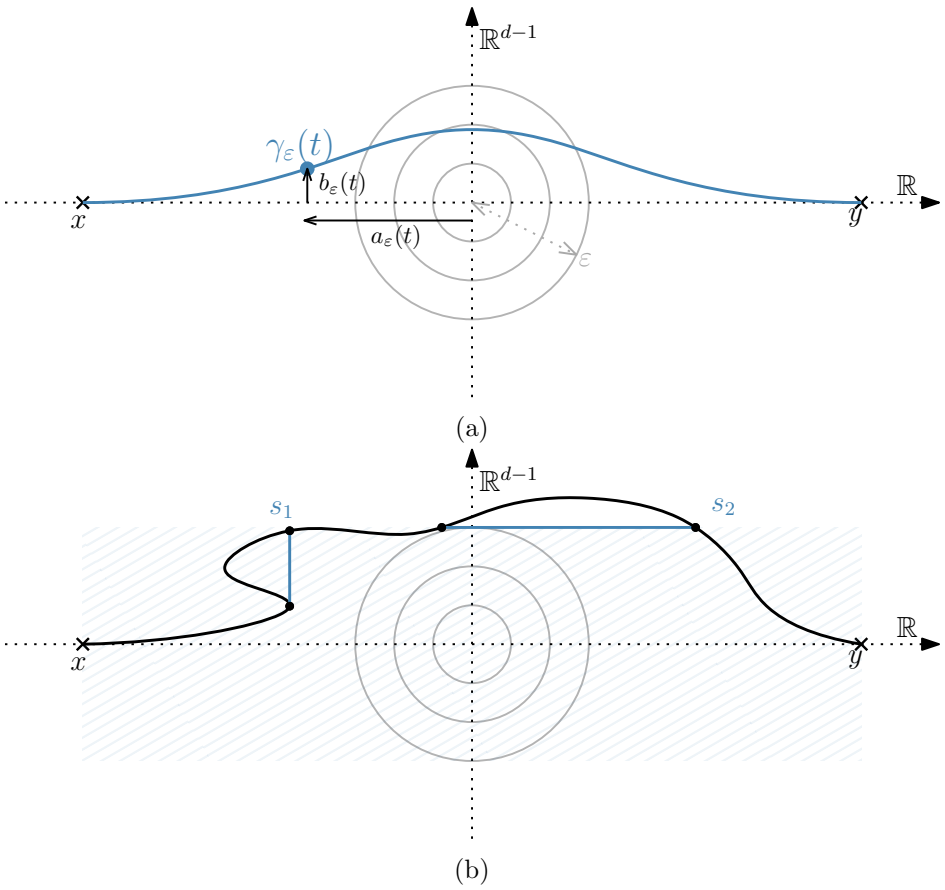


Figure 8: (a) Top view of M_ε and of one of the shortest path between x and y, in blue. In light grey is represented the bump of size ε. (b) Same view of M_ε as (a), illustrating the fact that any shortest path must go from left to right (otherwise one can construct a shorter path, through s_1 in the figure) and cannot go outside the shaded area (otherwise one can construct a shorter path, through s_2 in the figure).

- a_ε and b_ε are smooth and $\gamma_\varepsilon = \bar{\Psi}_\varepsilon(w_\varepsilon)$.
- Since M_ε is symmetric with respect to $\{0\} \times \mathbb{R}^d$, so should be the shortest path between x and y. This entails in particular that b_ε is even and that a_ε is odd;
- As γ_ε has constant speed and has a curvature bounded from above (as a shortest path in a bounded-curvature space), the ratio $\|\gamma_\varepsilon''\|/\|\gamma_\varepsilon'\|^2$ is bounded in sup-norm by a constant depending on R only. Therefore, there exists a constant $B > 0$ depending on R only such that, uniformly on ε small enough,

$$\max\left\{\|a_\varepsilon'\|_\infty/\ell, \|a_\varepsilon''\|_\infty/\ell^2, \|b_\varepsilon'\|_\infty/\ell, \|b_\varepsilon''\|_\infty/\ell^2\right\} \leqslant B. \tag{24}$$

- By symmetry also, γ_ε crosses the hyperplane $\{0\} \times \mathbb{R}^d$ orthogonally. As a consequence $\langle \gamma_\varepsilon'(0), e_{d+1}\rangle = 0$, $b_\varepsilon'(0) = 0$ and

$$a_\varepsilon'(0) = \|w_\varepsilon'(0)\| = \|\gamma_\varepsilon'(0)\| \in [A_1\ell, A_2\ell]\,,$$

where A_1 and A_2 were introduced in (23).

– Finally, using (24), we deduce that there exists $C > 0$ depending on R only such that for all $t \in [-1, 1]$,

$$\begin{cases} |a_\varepsilon(t) - a'_\varepsilon(0)t| & \leqslant C\ell^2 t^2, \\ |a'_\varepsilon(t) - a'_\varepsilon(0)| & \leqslant C\ell^2 t, \\ |a'_\varepsilon(t)a_\varepsilon(t) - a'_\varepsilon(0)^2 t| & \leqslant C\ell^3 t^2, \\ |b_\varepsilon(t) - b_\varepsilon(0)| & \leqslant C\ell t. \end{cases} \tag{25}$$

C.1.3 Perturbative Expansion of the Geodesic Length

We let $\gamma_0(t) := \bar{\Psi}_0(a_\varepsilon(t)e_1)$. Although not constant-speed, monotonicity of a_ε implies that γ_0 is the shortest path in M_0 between x and y, and we get, using (24) and (25), that for some constant A_3 depending on R,

$$\frac{1}{2}A_1\ell \leqslant \|\gamma'_0(t)\| \leqslant A_3\ell \quad \text{if} \quad \ell \leqslant \frac{A_1}{2C}, \tag{26}$$

which we will assume henceforth. Furthermore, the velocity of γ_ε writes

$$\begin{aligned} \gamma'_\varepsilon &= \mathrm{d}\bar{\Psi}_\varepsilon(w_\varepsilon)[w'_\varepsilon] = \mathrm{d}\bar{\Psi}_0(w_\varepsilon)[w'_\varepsilon] + c\varepsilon^{k-1}\langle \nabla K(w_\varepsilon/\varepsilon), w'_\varepsilon\rangle e_{d+1} \\ &= w'_\varepsilon + \langle \nabla\Psi_0(w_\varepsilon), w'_\varepsilon\rangle e_{d+1} + c\varepsilon^{k-1}\langle \nabla K(w_\varepsilon/\varepsilon), w'_\varepsilon\rangle e_{d+1} \\ &= a'_\varepsilon e_1 + b'_\varepsilon + \big(\underbrace{\langle \nabla\Psi_0(a_\varepsilon), a'_\varepsilon\rangle}_{:=\nabla_0} + \underbrace{c\varepsilon^{k-1}\langle \nabla K(w_\varepsilon/\varepsilon), w'_\varepsilon\rangle}_{:=\nabla_1}\big)e_{d+1}, \end{aligned}$$

where we used the fact that Ψ_0 depends only on its first variable. We write the last term as $(\nabla_0 + \nabla_1)e_{d+1}$. Using that each three terms in the preceding development are orthogonal, we obtain

$$\|\gamma'_\varepsilon\|^2 = a'^2_\varepsilon + \|b'_\varepsilon\|^2 + (\nabla_0 + \nabla_1)^2 = \underbrace{a'^2_\varepsilon + \nabla_0^2}_{=\|\gamma'_0\|^2} + \underbrace{\|b'_\varepsilon\|^2 + 2\nabla_0\nabla_1 + \nabla_1^2}_{:=Q_\varepsilon}, \tag{27}$$

and it only remains to study the last three terms, denoted by Q_ε. First, notice that using (24), one can find two constants D_0 depending on R such that $Q_\varepsilon \geqslant -D_0\varepsilon^2\ell^2$. Together with (26), this yields that $Q_\varepsilon/\|\gamma'_0\|^2 \geqslant -1$ for ε small enough (depending on R). Likewise, we can show that $Q_\varepsilon \leqslant D_1(\ell^2 + \ell^2\varepsilon^2 + \varepsilon^4)$, for some constant D_1 depending on R. This again yields

$$\frac{Q_\varepsilon}{\|\gamma'_0\|^2} \leqslant D_2 \quad \text{if} \quad \varepsilon \leqslant D_3\ell, \tag{28}$$

for some constants D_2 and D_3 depending on R only. All in all, we have that $Q_\varepsilon/\|\gamma'_0\| \in [-1, D_2]$. Using that

$$\sqrt{1+z} \geqslant \begin{cases} 1+z & \text{if} \quad z \in [-1, 0], \\ 1 + D_4 z & \text{if} \quad z \in [0, D_2], \ \text{with} \ \ D_4 = \frac{1}{D_2}(\sqrt{1+D_2} - 1), \end{cases}$$

we can finally derive from (27) and (28) the following bound

$$\|\gamma'_\varepsilon\| = \|\gamma'_0\|\sqrt{1 + \frac{Q_\varepsilon}{\|\gamma'_0\|^2}} \geqslant \|\gamma'_0\| + \tau(Q_\varepsilon)Q_\varepsilon, \tag{29}$$

where

$$\tau(z) := \frac{2}{A_1\ell}\mathbb{1}_{z<0} + \frac{D_4}{A_3\ell}\mathbb{1}_{z\geqslant 0},$$

and where we also used (26) to bound $1/\|\gamma_0'\|$. In particular, integrating (29) over $[-1,1]$ yields that

$$\mathrm{d}_{M_\varepsilon}(x,y) \geqslant \mathrm{d}_{M_0}(x,y) + \int_{-1}^{1} \tau(Q_\varepsilon) Q_\varepsilon.$$

To obtain a more explicit bound, let us now study Q_ε. For this, first rewrite ∇_0 and ∇_1 more explicitly as

$$\nabla_0 = -\frac{a_\varepsilon a_\varepsilon'}{\sqrt{R^2 - a_\varepsilon^2}} \quad \text{and} \quad \nabla_1 = -2c\varepsilon^{k-2} \frac{K(w_\varepsilon/\varepsilon)}{(1-\|w_\varepsilon/\varepsilon\|^2)^2} \langle w_\varepsilon, w_\varepsilon' \rangle.$$

Hence, noticing that $\langle w_\varepsilon, w_\varepsilon' \rangle = a_\varepsilon a_\varepsilon' + \langle b_\varepsilon, b_\varepsilon' \rangle$, one can write $2\nabla_0\nabla_1$ as $P_0 + P_1$ with

$$\begin{cases} P_0 &= \varepsilon^{k-2} (a_\varepsilon a_\varepsilon')^2 T_\varepsilon \\ P_1 &= \varepsilon^{k-2} T_\varepsilon a_\varepsilon a_\varepsilon' \langle b_\varepsilon, b_\varepsilon' \rangle \end{cases} \quad \text{with} \quad T_\varepsilon := \frac{4cK(w_\varepsilon/\varepsilon)}{\sqrt{R^2 - a_\varepsilon^2}(1-\|w_\varepsilon/\varepsilon\|^2)^2}.$$

For $\ell \leqslant A_1/4C$, condition (25) together with $a_\varepsilon'(0) \geqslant A_1\ell/2$ imply that

$$\|w_\varepsilon(t)\| \geqslant |a_\varepsilon(t)| \geqslant \varepsilon \quad \text{for all } |t| \geqslant t_\varepsilon \ \text{ with } \ t_\varepsilon := \frac{4\varepsilon}{A_1\ell},$$

so that in particular, $T_\varepsilon(t) = 0$ for $|t| \geqslant t_\varepsilon$. Furthermore, notice that, provided that ℓ is small before R, T_ε is bounded by some constant $E > 0$ depending on R only. Using again (25), we find that for $\ell \leqslant A_1^2/8C$, there holds

$$\begin{aligned} &(a_\varepsilon' a_\varepsilon)^2(t) \geqslant \frac{1}{2} a_\varepsilon'(0)^4 t^2 - C^2\ell^6 t^4 \geqslant \frac{1}{32} A_1^4 \ell^4 t^2, \qquad (30) \\ \text{and} \qquad &|a_\varepsilon' a_\varepsilon|(t) \leqslant a_\varepsilon'(0)^2 |t| + C\ell^3 t^2 \leqslant 5A_1^2\ell^2|t| \qquad \text{for all } \ t \in [-1,1]. \end{aligned}$$

In particular, we find that

$$\begin{aligned} \int_{-1}^{1} |P_1(t)| \,\mathrm{d}t &\leqslant 5\varepsilon^{k-2} E A_1^2 \ell^2 \|b_\varepsilon\|_\infty \|b_\varepsilon'\|_\infty \int_{-t_\varepsilon}^{t_\varepsilon} |t| \,\mathrm{d}t \\ &= 5\varepsilon^{k-2} E A_1^2 \ell^2 \|b_\varepsilon\|_\infty \|b_\varepsilon'\|_\infty t_\varepsilon^2 \\ &\leqslant 80 BE \|b_\varepsilon\|_\infty \ell^2 \varepsilon^k, \end{aligned}$$

where we used (24) in the last inequality. On the other hand, letting $t_0 \in (-1,1)$ be a time at which $\|b_\varepsilon(t_0)\| = \|b_\varepsilon\|_\infty$, notice that

$$\begin{aligned} \int_{-1}^{1} \|b_\varepsilon'\|^2 &= \int_{-1}^{t_0} \|b_\varepsilon'\|^2 + \int_{t_0}^{1} \|b_\varepsilon'\|^2 \\ &\geqslant \frac{1}{1+t_0} \left\| \int_{-1}^{t_0} b_\varepsilon' \right\|^2 + \frac{1}{1-t_0} \left\| \int_{t_0}^{1} b_\varepsilon' \right\|^2 \\ &= \left(\frac{1}{1+t_0} + \frac{1}{1-t_0} \right) \|b_\varepsilon(t_0)\|^2 \geqslant 2\|b_\varepsilon\|_\infty^2. \end{aligned}$$

Integrating (27) and using that $\nabla_1^2 \geqslant 0$ thus yields

$$\int_{-1}^{1} \tau(Q_\varepsilon) Q_\varepsilon \geqslant 2\|b_\varepsilon\|_\infty \left(\tau_1 \|b_\varepsilon\|_\infty - 40\tau_2 BE\ell^2\varepsilon^k \right) + \tau_1 \varepsilon^{k-2} \int_{-1}^{1} (a_\varepsilon a_\varepsilon')^2 T_\varepsilon. \qquad (31)$$

where τ_1 is the smallest value of τ, and τ_2 its greatest value. Now we distinguish on the value of $\|b_\varepsilon(0)\|$:

– If $\|b_\varepsilon(0)\| \geqslant \varepsilon/2$, then $\|b_\varepsilon\|_\infty \geqslant \varepsilon/2$ and for ε small enough, we get, noticing that the last term in (31) is non-negative,

$$\int_{-1}^{1} \tau(Q_\varepsilon) Q_\varepsilon \geqslant c_R \varepsilon(\varepsilon/2 - \varepsilon/4)/\ell \geqslant c_R \varepsilon^2/\ell.$$

– Otherwise, if $\|b_\varepsilon(0)\| \leqslant \varepsilon/2$, then, using (25), we find that

$$\begin{cases} \|b_\varepsilon(t)\| & \leqslant 3\varepsilon/4, \\ |a_\varepsilon(t)| & \leqslant \varepsilon/2, \end{cases} \quad \text{for all } |t| \leqslant t_\varepsilon^* \quad \text{with} \quad t_\varepsilon^* := \min\left\{ \frac{\varepsilon}{4C\ell}, \frac{\varepsilon}{8A_2\ell}, \frac{2A_2}{C\ell} \right\}.$$

For ε small before R, t_ε^* is of the form $t_\varepsilon^* = G\varepsilon/\ell$ with G depending on R only. Furthermore, notice that for $|t| \leqslant t_\varepsilon^*$, there holds $\|w_\varepsilon(t)\|^2 = |a_\varepsilon(t)|^2 + \|b_\varepsilon(t)\|^2 \leqslant 13\varepsilon^2/16$. In particular, T_ε is lower-bounded on $[-t_\varepsilon^*, t_\varepsilon^*]$ by a constant H depending on R only. Noticing that $t_\varepsilon^* \leqslant t_\varepsilon$, we can use the inequality in (30) to obtain

$$\int_{-1}^{1} (a_\varepsilon a_\varepsilon')^2 T_\varepsilon \geqslant \frac{1}{32} A_1^4 \ell^4 H \int_{-t_\varepsilon^*}^{t_\varepsilon^*} t^2 \,\mathrm{d}t = \frac{1}{48} A_1^4 H G^3 \ell \varepsilon^3.$$

Finally, since $z \mapsto z(z-\nu)$ is minimal on $\mathbb{R}_+$ at $z = \nu/2$ with minimal value $-\nu^2/4$, we find the bound

$$\int_{-1}^{1} \tau(Q_\varepsilon) Q_\varepsilon \geqslant c_R \varepsilon^{k+1} - c_R' \ell^3 \varepsilon^{2k} \geqslant c_R \varepsilon^{k+1},$$

provided that ε is small enough before R.

In both cases, we find that $\int_{-1}^{1} \tau(Q_\varepsilon) Q_\varepsilon \geqslant c_R \varepsilon^{k+1}$. Now integrating (29) gives

$$\mathrm{d}_{M_\varepsilon}(x,y) \geqslant \mathrm{d}_{M_0}(x,y) + c_R' \varepsilon^{k+1} > \mathrm{d}_{M_0}(x,y).$$

Finally, (23) yields $\mathrm{d}_{M_\varepsilon}(x,y) \geqslant 2A_2\ell$ and letting $\ell := (1 \vee D_3^{-1})\varepsilon$, which we can from (28), finally gives

$$\left| 1 - \frac{\mathrm{d}_{M_0}(x,y)}{\mathrm{d}_{M_\varepsilon}(x,y)} \right| \geqslant c_R \varepsilon^k. \tag{32}$$

C.1.4 Concluding with Le Cam's lemma

We apply Lemma C.1 with M_0 and $M_1 := M_\varepsilon$ for ε properly chosen. Their volumes are bounded from above and below by something depending on R and d only, so that the uniform distribution on M_0 and M_ε are in $\mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$ provided that $f_{\min}$ and $f_{\max}$ are respectively small enough and large enough compared to $1/R^d$. Finally, we set $R = 2\,\mathrm{rch}_{\min}$ and $\varepsilon = (C_{\mathrm{rch}_{\min},d} n)^{-1/d}$. For n large enough so that all previous controls are verified, Lemma C.1 finally yields

$$\inf_{\widehat{\mathrm{d}}} \sup_{P \in \mathcal{P}^k} \mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_M)] \geqslant \frac{1}{2} c_{\mathrm{rch}_{\min}} \varepsilon^k (1 - C_{\mathrm{rch}_{\min},d} n \varepsilon^d) \geqslant c_{\mathrm{rch}_{\min},d,k} n^{-k/d},$$

where the total variation was bounded using [10, Lemma 7]. □

C.2 Plug-in Estimation for Metric Learning

We start by giving the proof of Proposition 5.3.

Proof of Proposition 5.3. Let $x, y \in K$. Notice that, since $K \subset (K')^\varepsilon$, there holds trivially that $\mathrm{d}_{(K')^\varepsilon}(x,y) \leqslant d_K(x,y)$. For the converse inequality, let $\gamma : [0,1] \to \mathbb{R}^D$ be a continuous path in $(K')^\varepsilon$ between x and y. Since $\varepsilon < \mathrm{rch}(K)/2$ the closest-point projection on K is well-defined on $(K')^\varepsilon \subset K^{2\varepsilon}$ and we can consider $\gamma_0 = \pi_K \circ \gamma$, which is a continuous path in K. For any subdivision $0 = t_0 < t_1 < \cdots < t_k = 1$, there holds

$$\sum_{i=0}^{k-1} \|\gamma_0(t_{i+1}) - \gamma_0(t_i)\| \leqslant \frac{\mathrm{rch}(K)}{\mathrm{rch}(K) - 2\varepsilon} \sum_{i=0}^{k-1} \|\gamma(t_{i+1}) - \gamma(t_i)\|$$

where we used the fact that π_K is $\mathrm{rch}(K)/(\mathrm{rch}(K) - 2\varepsilon)$-Lipschitz on $K^{2\varepsilon}$ [26, Theorem 4.8 (8)]. Taking the supremum over all subdivision yields

$$\mathrm{d}_K(x,y) \leqslant L(\gamma_0) \leqslant \frac{\mathrm{rch}(K)}{\mathrm{rch}(K) - 2\varepsilon} L(\gamma)$$

and then taking the infimum on all continuous path γ finally gives

$$\mathrm{d}_{(K')^\varepsilon}(x,y) \geqslant \left(1 - \frac{2\varepsilon}{\mathrm{rch}(K)}\right) \mathrm{d}_K(x,y)$$

ending the proof. □

To prove Theorem 5.4, an intermediate result that bounds the intrinsic diameters of the supports in our statistical model is needed.

Lemma C.2. *For any $P \in \mathcal{P}^k_{\mathrm{rch}_{\min},\mathbf{L}}(f_{\min}, f_{\max})$, if $M = \mathrm{Support}(P)$, then*

$$\sup_{x,y \in M} \mathrm{d}_M(x,y) \leqslant \mathrm{d}_{\max}.$$

where $\mathrm{d}_{\max}$ is defined in Theorem 5.4.

Proof of Lemma C.2. We let $x_1, \ldots, x_N$ be a $\mathrm{rch}_{\min}/4$-packing of M. We let $x, y \in M$, and G be the neighborhood graph built on top of $x, y, x_1, \ldots, x_N$ with connectivity radius $\mathrm{rch}_{\min}/2$. Using [38, Theorem 6.3], denoting $z_0 = x, z_1, \ldots, z_k = y$ the shortest path between x and y in G, there holds

$$\mathrm{d}_M(x,y) \leqslant \sum_{i=0}^{k-1} \mathrm{d}_M(z_i, z_{i+1}) \leqslant \sum_{i=0}^{k-1} 2\|z_i - z_{i+1}\| \leqslant k\, \mathrm{rch}_{\min}\,.$$

But now $k \leqslant N - 1$ and

$$N \leqslant \frac{\mathrm{vol}_d(M)}{\min_{x \in M} \mathrm{vol}_d(M \cap \mathsf{B}(x, \mathrm{rch}_{\min}/4))} \leqslant \frac{\mathrm{vol}_d(M)}{(1 - 1/8^2)^{d/2}\, \omega_d\, \mathrm{rch}^d_{\min}/4^d},$$

where we used [38, Lemma 5.3]. Noticing that $\mathrm{vol}_d(M) \leqslant 1/f_{\min}$, we easily conclude. □

We are now in position to prove Theorem 5.4.

Proof of Theorem 5.4. We let $\mathcal{A}_n := \left\{ \mathrm{d_H}(\widehat{M}, M) \leqslant \varepsilon_n \right\}$ denote the event where $\widehat{M}$ is ε_n-precise in Hausdorff distance, and we take $x, y \in M$. On the event $\mathcal{A}_n$, for n large enough such that $\varepsilon_n \leqslant \mathrm{rch}_{\min}/2$, Proposition 5.3 applies to $K' = \widehat{M} \cup \{x, y\}$ and, together with Lemma C.2, yields

$$\left| 1 - \frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_M(x,y)} \right| \leqslant \frac{2\varepsilon_n}{\mathrm{rch}_{\min}}.$$

On $\mathcal{A}_n^c$, we distinguish whether $\|x - y\| \leqslant \varepsilon_n$ or not. If so, then $\widehat{\mathrm{d}}(x,y) = \|x-y\| \leqslant \mathrm{d}_M(x,y)$. In the other case, $\mathrm{d}_M(x,y) \geqslant \|x-y\| \geqslant \varepsilon_n$ and $\mathrm{d}(x,y) \leqslant \mathrm{d}_{\max}$ so that, in any case

$$\left| 1 - \frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_M(x,y)} \right| \leqslant 1 + \frac{\widehat{\mathrm{d}}(x,y)}{\mathrm{d}_M(x,y)} \leqslant 1 + \frac{\mathrm{d}_{\max}}{\varepsilon_n},$$

for n large enough such that $\varepsilon_n \leqslant \mathrm{d}_{\max}$. Patching these two bounds together yields

$$\mathbb{E}_{P^{\otimes n}}[\ell_\infty(\widehat{\mathrm{d}}|\mathrm{d}_M)] \leqslant \frac{2\varepsilon_n}{\mathrm{rch}_{\min}} P^{\otimes n}(\mathcal{A}_n) + \left(1 + \frac{\mathrm{d}_{\max}}{\varepsilon_n}\right) P^{\otimes n}(\mathcal{A}_n^c),$$

ending the proof. □

D Proofs of Section 6

We first prove that submanifolds of the model do fulfill Assumption A1 and Assumption A2.

Proof of Proposition 6.2. Assumption A2 is a simple consequence of [38, Proposition 6.3] which yields fulfillment for $\Delta_1 = \mathrm{rch}(M)/2$ and $C_1 = 2$. For Assumption A1, take $x, y \in M$ such that $\|x - y\| \leqslant \mathrm{rch}(M)$ and take $\varepsilon < \mathrm{rch}(M)/4$. We consider $a = \exp_y(v)$, where

$$v = -\varepsilon \frac{\pi_{T_yM}(x-y)}{\|\pi_{T_yM}(x-y)\|}.$$

Thanks to [26, Theorem 4.8 (7)], there holds

$$\begin{aligned} \|\pi_{T_yM}(x-y)\|^2 &= \|x-y\|^2 - \mathrm{d}^2(x-y, T_yM) \\ &\geqslant \|x-y\|^2 - \frac{\|x-y\|^4}{4\,\mathrm{rch}^2(M)} \\ &\geqslant \frac{3}{4}\|x-y\|^2, \end{aligned}$$

and

$$\langle v, y - x \rangle = \varepsilon \frac{\langle x - y, \pi_{T_yM}(x-y) \rangle}{\|\pi_{T_yM}(x-y)\|} = \varepsilon \|\pi_{T_yM}(x-y)\| \geqslant \frac{1}{2}\varepsilon \|x-y\|,$$

so that

$$\|x - y - v\|^2 \geqslant \|x-y\|^2 + \varepsilon\|x-y\| + \varepsilon^2 \geqslant \left(\|x-y\| + \frac{1}{2}\varepsilon\right)^2,$$

and thus $\|x - y - v\| \geqslant \|x-y\| + \varepsilon/2$. But now $\|x - a\| \geqslant \|x - y - v\| - \|a - y - v\|$ and $\|a - y - v\| \leqslant 5\varepsilon^2/4\,\mathrm{rch}(M)$ according to [3, Lemma 1]. All in all, we get that

$$\|x - a\| \geqslant \|x - y\| + \frac{1}{2}\varepsilon - \frac{5}{4\,\mathrm{rch}(K)}\varepsilon^2 \geqslant \|x-y\| + \frac{3}{16}\varepsilon,$$

ending the proof. □

To prove Theorem 6.3, a bound on the metric distortion between our distance estimator and d_M is needed, that easily follows from Proposition 5.3.

Proposition D.1. *In the context of Proposition 5.3, we have that for all $\delta > 4\varepsilon$,*

$$\mathrm{D}_\delta(\mathrm{d}_K, \mathrm{d}_{(K')^\varepsilon}) \leqslant 1 + \frac{4\varepsilon}{(\delta - 4\varepsilon) \wedge \mathrm{rch}(K)}.$$

Proof of Proposition D.1. Proposition 5.3 already gives that $\mathrm{D}_\delta(\mathrm{d}_K|\mathrm{d}_{(K')^\varepsilon}) \leqslant 1 + 2\varepsilon/\mathrm{rch}(K)$. For the other control, notice that for any two $x, y \in (K')^\varepsilon$ that are δ-apart for the Euclidean distance, there holds denoting $x_0 = \pi_K(x)$ and $y_0 = \pi_K(y)$,

$$\mathrm{d}_{(K')^\varepsilon}(x, y) \leqslant 4\varepsilon + \mathrm{d}_K(x_0, y_0)$$

because the piecewise-defined path consisting of the segment $[x, x_0]$ of the (or a near-minimizing) shortest-path between x_0 and y_0 in K, and of the segment $[y_0, y]$, is a continuous path in $(K')^\varepsilon$ between x and y of length the RHS of the display above. Now notice that

$$\mathrm{d}_K(x_0, y_0) \geqslant \|x_0 - y_0\| \geqslant \delta - 4\varepsilon,$$

which immediately yields $\mathrm{D}_\delta(\mathrm{d}_{(K')^\varepsilon}|\mathrm{d}_K) \leqslant 1 + \frac{4\varepsilon}{\delta - 4\varepsilon}$. □

The rate of the plug-in SDR estimator follows straightforwardly.

Proof of Theorem 6.3. Let $\mathcal{A}_n := \left\{\mathrm{d}_{\mathrm{H}}(M, \widehat{M}) \leqslant \varepsilon_n\right\}$. On this event, we have $\mathrm{D}_\delta(\widehat{\mathrm{d}}, \mathrm{d}_M) \leqslant 1 + 8\varepsilon_n/\delta$ according to Proposition D.1, so that applying Theorem 4.9 with $\delta_0 = \delta/2$, $\varepsilon = \varepsilon_n$ and $\nu = 8\varepsilon_n/\delta$ yields $|\widehat{\mathrm{sdr}}_\delta - \mathrm{sdr}_\delta(M, \mathrm{d}_M)| \leqslant \zeta_0 \varepsilon_n$ with $\zeta_0 \leqslant C\mathrm{s}_{\max}^4/\delta^4$. We conclude that

$$\mathbb{E}_{P^{\otimes n}}|\widehat{\mathrm{sdr}}_\delta - \mathrm{sdr}_\delta(M, \mathrm{d}_M)| \leqslant \zeta_0 \varepsilon_n P^{\otimes n}(\mathcal{A}_n) + 2\mathrm{s}_{\max} P^{\otimes n}(\mathcal{A}_n^c),$$

which ends the proof. □