Non-Asymptotic Rates for Manifold, Tangent Space, and Curvature Estimation

Eddie Aamari, Clément Levrard

Non-asymptotic Rates for Manifold, Tangent Space and Curvature Estimation

Eddie Aamari*†‡
Laboratoire de Mathématiques d'Orsay, INRIA Saclay

Clément Levrard*†
Université Paris Diderot

Abstract

Given an n-sample drawn on a submanifold $M \subset \mathbb{R}^D$, we derive optimal rates for the estimation of tangent spaces $T_X M$, the second fundamental form II_X^M, and the submanifold M. After motivating their study, we introduce a quantitative class of $\mathcal{C}^k$-submanifolds in analogy with Hölder classes. The proposed estimators are based on local polynomials and allow to deal simultaneously with the three problems at stake. Minimax lower bounds are derived using a conditional version of Assouad's lemma when the base point X is random.

1 Introduction

A wide variety of data can be thought of as being generated on a shape of low dimensionality compared to possibly high ambient dimension. This point of view led to the developpement of the so-called topological data analysis, which proved fruitful for instance when dealing with physical parameters subject to constraints, biomolecule conformations, or natural images [Was16]. This field intends to associate geometric quantities to data without regard of any specific coordinate system or parametrization. If the underlying structure is sufficiently smooth, one can model a point cloud $\mathbb{X} = \{X_1, \dots, X_n\}$ as being sampled on a d-dimensional submanifold $M \subset \mathbb{R}^D$. In such a case, geometric and topological intrinsic quantities include (but are not limited to) homology groups [NSW08], persistent homology [FLR$^+$14], volume [APR16], differential quantities [CP05] or the submanifold itself [GPPVW12, AL15].

The present paper focuses on optimal rates for estimation of quantities up to order two: (0) the submanifold itself, (1) tangent spaces, and (2) second fundamental forms.

*Research supported by ANR project TopData ANR-13-BS01-0008
†Research supported by Advanced Grant of the European Research Council GUDHI
‡Supported by the Conseil régional d'Île-de-France program RDM-IdF

Among these three questions, a special attention has been paid to the estimation of the submanifold. In particular, it is a central problem in manifold learning. Indeed, there exists a wide bunch of algorithms intended to reconstruct submanifolds from point clouds (Isomap [TdSL00], LLE [RS00], and restricted Delaunay Complexes [BG14, CDR05] for instance), but a few come with theoretical guarantees [GPPVW12, AL15]. Up to our knowledge, a minimax lower bound has proved optimality of a reconstruction scheme in only one case [GPPVW12]. Some of these reconstruction procedures are based on tangent space estimation [BG14, AL15, CDR05]. Tangent space estimation itself also yields interesting applications in manifold clustering [GM11, ALZ13]. Estimation of curvature-related quantities naturally arises in shape reconstruction, since curvature can drive the size of a meshing. As a consequence, most of the associated results deal with the case $d = 2$ and $D = 3$, though some of them may be extended to higher dimensions [MOG11, GWM01]. Several algorithms have been proposed in that case [Rus04, CP05, MOG11, GWM01], but with no analysis of their performances from a statistical point of view.

To assess the quality of such a geometric estimator, the class of submanifolds over which the procedure is evaluated has to be specified. Up to now, the most commonly used model for submanifolds relied on the reach τ_M, a generalized convexity parameter. Assuming $\tau_M \geq \tau_{min} > 0$ involves both local regularity — a bound on curvature — and global regularity — no arbitrarily pinched area. This $\mathcal{C}^2$-like assumption has been extensively used in the computational geometry and geometric inference fields [AL15, NSW08, FLR$^+$14, APR16, GPPVW12]. One attempt of a specific investigation for higher orders of regularity $k \geq 3$ has been proposed in [CP05].

However, many works suggest that the regularity of the submanifold has an important impact on convergence rates. This is pretty clear for tangent space estimation, where convergence rates of PCA-based estimators range from $(1/n)^{1/d}$ in the $\mathcal{C}^2$ case [AL15] to $(1/n)^{\alpha}$ with $1/d < \alpha < 2/d$ in more regular settings [SW12, TVF13]. In addition, it seems that PCA-based estimators are outperformed by estimators taking into account higher orders of smoothness [CC16, CP05], for regularities at least $\mathcal{C}^3$. For instance fitting quadratic terms lead to a convergence rate of order $(1/n)^{2/d}$ in [CC16]. These remarks naturally led us to investigate the properties of local polynomial approximation for regular submanifolds, where "regular" has to be properly defined. Local polynomial fitting for geometric inference was studied in several frameworks such as [CP05]. In some sense, a part of our work extends these results, by investigating the dependency of convergence rates on the sample size n, but also on the order of regularity k and the ambient and intrinsic dimensions d and D.

1.1 Contribution

In this paper, we build a collection of models for $\mathcal{C}^k$-submanifolds ($k \geq 3$) that naturally generalize the commonly used one for $k = 2$ (Section 2). We emphasize the necessity of both local and global constraints for estimation. On these models, we study the non-asymptotic rates of estimation for tangent

space, second fundamental form, and submanifold estimation (Section 3). These results shed light on the influence of k, d, D and n on these estimation problems, showing for instance that the ambient dimension D plays no role. The estimators proposed all rely on the analysis of local polynomials, and allow to deal with the three estimation problems in a unified way (Section 5.1). Minimax lower bounds are derived using standard Bayesian techniques, although a new version of Assouad's Lemma is used for tangent spaces and second fundamental forms when the base point is random (Section 5.2). For the sake of completeness, geometric background and proofs of technical lemmas are given in the Appendix.

2 $\mathcal{C}^k$ Models for Submanifolds

2.1 Notation

Throughout the paper, we consider d-dimensional compact submanifolds $M \subset \mathbb{R}^D$ without boundary. The submanifolds will always be assumed to be at least $\mathcal{C}^2$. For all $p \in M$, T_pM stands for the tangent space of M at p [dC92, Chapter 0]. We let $II_p^M : T_pM \times T_pM \to T_pM^\perp$ denote the second fundamental form of M at p [dC92, p. 125]. II_p^M characterizes the curvature of M at p. The standard inner product in $\mathbb{R}^D$ is denoted by $\langle \cdot, \cdot \rangle$ and the Euclidean distance by $\|\cdot\|$. Given a linear subspace $T \subset \mathbb{R}^D$, write $T^\perp$ for its orthogonal space. We write $\mathcal{B}(p, r)$ for the closed Euclidean ball of radius $r > 0$ centered at $p \in \mathbb{R}^D$, and for short $\mathcal{B}_T(p, r) = \mathcal{B}(p, r) \cap T$. For a smooth function $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ and $i \geq 1$, we let $d_x^i\Phi$ denote the ith order differential of Φ at $x \in \mathbb{R}^D$. For a linear map A defined on $T \subset \mathbb{R}^D$, $\|A\|_{\text{op}} = \sup_{v \in T} \frac{\|Av\|}{\|v\|}$ stands for the operator norm. We adopt the same notation $\|\cdot\|_{op}$ for tensors, i.e. multilinear maps. Similarly, if $\{A_x\}_{x \in T'}$ is a family of linear maps, for short, its L^∞ operator norm is denoted by $\|A\|_{op} = \sup_{x \in T'} \|A_x\|_{op}$. When it is well defined, we will write $\pi_B(z)$ for the projection of $z \in \mathbb{R}^D$ onto the closed subset $B \subset \mathbb{R}^D$, that is the nearest neighbor of z in B. The distance between two linear subspaces $U, V \subset \mathbb{R}^D$ of the same dimension is measured by the principal angle $\angle(U, V) = \|\pi_U - \pi_V\|_{\text{op}}$. The Hausdorff distance [GPPVW12] in $\mathbb{R}^D$ is denoted by d_H. For a probability distribution P, $\mathbb{E}_P$ stands for the expectation with respect to P. We write $P^{\otimes n}$ for the n-times tensor product of P.

Throughout this paper, C_α will denote a generic constant depending on the parameter α. For clarity's sake, C'_α, c_α, or c'_α may also be used when several constants are involved.

2.2 Reach and Regularity of Submanifolds

As introduced in [Fed59], the reach τ_M of a subset $M \subset \mathbb{R}^D$ is the maximal neighborhood radius for which the projection π_M onto M is well defined. More precisely, denoting by $d(\cdot, M)$ the distance to M, the medial axis of M is defined

to be the set of points which have at least two nearest neighbors on M, that is

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

The reach is then defined by

$$\tau_M = \inf_{p \in M} d\left(p, Med(M)\right) = \inf_{z \in Med(M)} d\left(z, M\right).$$

It gives a minimal scale of geometric and topological features of M. As a generalized convexity parameter, τ_M is a key parameter in reconstruction [AL15, GPPVW12] and in topological inference [NSW08]. Having $\tau_M \geq \tau_{min} > 0$ prevents M from almost auto-intersecting, and bounds its curvature in the sense that $\left\| II_p^M \right\|_{op} \leq \tau_M^{-1} \leq \tau_{min}^{-1}$ for all $p \in M$ [NSW08, Proposition 6.1].

For $\tau_{min} > 0$, we let $\mathcal{C}^2_{\tau_{min}}$ denote the set of d-dimensional compact connected submanifolds M of $\mathbb{R}^D$ such that $\tau_M \geq \tau_{min} > 0$. A key property of submanifolds $M \in \mathcal{C}^2_{\tau_{min}}$ is the existence of a parametrization closely related to the projection onto tangent spaces. We let $\exp_p : T_pM \to M$ denote the geodesic map [dC92, Chapter 3], that is defined by $\exp_p(v) = \gamma_{p,v}(1)$, where $\gamma_{p,v}$ is the unique constant speed geodesic path of M with initial value p and velocity v.

Lemma 1. *If $M \in \mathcal{C}^2_{\tau_{min}}$, $\exp_p : \mathcal{B}_{T_pM}(0, \tau_{min}/4) \to M$ is one-to-one. Moreover, it can be written as*

$$\begin{aligned} \exp_p \colon \mathcal{B}_{T_pM}(0, \tau_{min}/4) &\longrightarrow M \\ v &\longmapsto p + v + \mathbf{N}_p(v) \end{aligned}$$

with $\mathbf{N}_p$ such that for all $v \in \mathcal{B}_{T_pM}(0, \tau_{min}/4)$,

$$\mathbf{N}_p(0) = 0, \quad d_0\mathbf{N}_p = 0, \quad \left\| d_v\mathbf{N}_p \right\|_{op} \leq L_\perp \|v\|,$$

where $L_\perp = 5/(4\tau_{min})$. Furthermore, for all $p, y \in M$,

$$y - p = \pi_{T_pM}(y-p) + R_2(y-p),$$

where $\|R_2(y-p)\| \leq \frac{\|y-p\|^2}{2\tau_{min}}$.

In other words, elements of $\mathcal{C}^2_{\tau_{min}}$ have local parametrizations on top of their tangent spaces that are defined on neighborhoods with a minimal radius, and these parametrizations differ from the identity map by at most a quadratic term. In addition, the reach condition provides an order 2 Taylor expansion of the submanifold on top of its tangent spaces. A natural extension to $\mathcal{C}^k$-submanifolds should ensure that such an expansion exists at order k and satisfies some regularity constraints. To this aim, we introduce the following class of regularity $\mathcal{C}^k_{\tau_{min},\mathbf{L}}$.

Definition 2. *For $k \geq 3$, $\tau_{min} > 0$, and $\mathbf{L} = (L_\perp, L_3, \ldots, L_k)$, we let $\mathcal{C}^k_{\tau_{min},\mathbf{L}}$ denote the set of d-dimensional compact connected submanifolds M of $\mathbb{R}^D$ with $\tau_M \geq \tau_{min}$ and such that, for all $p \in M$, there exists a local one-to-one parametrization Ψ_p of the form:*

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

for some $r \geq \frac{1}{8L_\perp}$, with $\mathbf{N}_p \in \mathcal{C}^k\left(\mathcal{B}_{T_pM}(0,r), \mathbb{R}^D\right)$ such that

$$\mathbf{N}_p(0) = 0, \quad d_0\mathbf{N}_p = 0, \quad \left\|d^2_v\mathbf{N}_p\right\|_{op} \leq L_\perp,$$

for all $\|v\| \leq \frac{1}{8L_\perp}$. Furthermore, we require that

$$\left\|d^i_v\mathbf{N}_p\right\|_{op} \leq L_i \text{ for all } 3 \leq i \leq k.$$

Let us precise that the radius $1/(8L_\perp)$ has been chosen for convenience. Other smaller scales would do and we could even parametrize this constant, but without substantial benefits in the results.

The Ψ_p's can be seen as unit parametrizations of M. The conditions on $\mathbf{N}_p(0)$, $d_0\mathbf{N}_p$, and $d^2_v\mathbf{N}_p$ ensure that Ψ_p^{-1} is close to the projection π_{T_pM}. The bounds on $d^i_v\mathbf{N}_p$ ($3 \leq i \leq k$) allow to control the coefficients of the polynomial expansion we seek. Indeed, whenever $M \in \mathcal{C}^k_{\tau_{min},\mathbf{L}}$, Lemma 12 shows that for every p in M, and y in $\mathcal{B}\big(p, \frac{\tau_{min} \wedge L_\perp^{-1}}{4}\big) \cap M$,

$$y - p = \pi^*(y-p) + \sum_{i=2}^{k-1} T^*_i(\pi^*(y-p)^{\otimes i}) + R_k(y-p), \tag{1}$$

where π^* denotes the orthogonal projection onto T_pM, the T^*_i are i-linear maps from T_pM to $\mathbb{R}^D$ with $\|T^*_i\|_{op} \leq L'_i$ and R_k satisfies $\|R_k(y-p)\| \leq C\|y-p\|^k$, where the constants C and the L'_i's depend on the parameters τ_{min}, d, k, $L_\perp, \ldots, L_k$.

Such Ψ_p's exist for any compact $\mathcal{C}^k$-submanifold, if one allows τ^{-1}_{min}, $L_\perp$, L_3,...,L_k to be large enough. Note that for $k \geq 3$ the exponential map can happen to be only $\mathcal{C}^{k-2}$ for a $\mathcal{C}^k$-submanifold [Har51]. Hence, it may not be a good choice of Ψ_p. However, for $k = 2$, taking $\Psi_p = \exp_p$ is sufficient for our purpose. For ease of notation, we may write $\mathcal{C}^2_{\tau_{min},\mathbf{L}}$ although the specification of $\mathbf{L}$ is useless. In this case, we implicitly set by default $\Psi_p = \exp_p$ and $L_\perp = 5/(4\tau_{min})$.

As will be shown in Theorem 5, the global assumption $\tau_M \geq \tau_{min} > 0$ cannot be dropped, even when higher order regularity bounds L_i's are fixed.

Let us now describe the statistical model. Every d-dimensional submanifold $M \subset \mathbb{R}^D$ inherits a natural uniform volume measure by restriction of the ambient d-dimensional Hausdorff measure $\mathcal{H}^d$. In what follows, we will consider probability distributions that are almost uniform on some M in $\mathcal{C}^k_{\tau_{min},\mathbf{L}}$, as stated below.

Definition 3. *For $k \geq 2$, $\tau_{min} > 0$, $\mathbf{L} = (L_\perp, L_3, \ldots, L_k)$ and $f_{min} \leq f_{max}$, we let $\mathcal{P}^k_{\tau_{min},\mathbf{L}}(f_{min}, f_{max})$ denote the set of distributions P with support $M \in \mathcal{C}^k_{\tau_{min},\mathbf{L}}$ that have a density f with respect to the volume measure on M, and such that for all $x \in M$,*

$$0 < f_{min} \leq f(x) \leq f_{max} < \infty.$$

For short, we write $\mathcal{P}^k$ when there is no ambiguity. We denote by $\mathbb{X}_n$ an i.i.d. n-sample $\{X_1, \ldots, X_n\}$, that is, a sample with distribution $P^{\otimes n}$ for some $P \in \mathcal{P}^k$. In what follows, though M is unknown, all the parameters of the model will be assumed to be known, including the intrinsic dimension d and the order of regularity k. We will also denote by $\mathcal{P}^k_{(x)}$ the subset of elements in $\mathcal{P}^k$ whose support contains a prescribed $x \in \mathbb{R}^D$.

In view of our minimax study on $\mathcal{P}^k$, it is important to ensure by now that $\mathcal{P}^k$ is stable with respect to deformations and dilations. Here, since we deal with submanifolds, a natural way to model deformations is through ambient diffeomorphisms.

Proposition 4. *Let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ be a global $\mathcal{C}^k$-diffeomorphism. If $\|d\Phi - I_D\|_{op}$, $\left\|d^2\Phi\right\|_{op}, \ldots, \left\|d^k\Phi\right\|_{op}$ are small enough, then for all P in $\mathcal{P}^k_{\tau_{min},\mathbf{L}}(f_{min}, f_{max})$, the pushforward distribution $P' = \Phi_* P$ belongs to $\mathcal{P}^k_{\tau_{min}/2,2\mathbf{L}}(f_{min}/2, 2f_{max})$.*

Moreover, if $\Phi = \lambda I_D$ ($\lambda > 0$) is an homogeneous dilation, then $P' \in \mathcal{P}^k_{\lambda\tau_{min},\mathbf{L}_{(\lambda)}}(f_{min}/\lambda^d, f_{max}/\lambda^d)$, where $\mathbf{L}_{(\lambda)} = (L_\perp/\lambda, L_3/\lambda^2, \ldots, L_k/\lambda^{k-1})$.

Proposition 4 follows from a geometric reparametrization argument (Proposition 21) and a change of variable result for the Hausdorff measure (Lemma 22).

2.3 Necessity of a Global Assumption

In the previous Section 2.2, we generalized $\mathcal{C}^2$-like models — stated in terms of reach — to $\mathcal{C}^k$ for $k \geq 3$ by imposing higher order differentiability bounds on parametrizations Ψ_p's. Though, we did not drop the global assumption $\tau_M \geq \tau_{min} > 0$. Indeed, it appears that such an assumption is necessary for estimation purpose.

Theorem 5. *Assume that $\tau_{min} = 0$. If $D \geq d + 3$, then for all $k \geq 3$ and $L_\perp > 0$, provided that $L_3/L_\perp^2, \ldots, L_k/L_\perp^{k-1}, L_\perp^d/f_{min}$ and $f_{max}/L_\perp^d$ are large enough (depending only on d and k), for all $n \geq 1$,*

$$\inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{(x)}} \mathbb{E}_{P^{\otimes n}} \angle\big(T_x M, \hat{T}\big) \geq \frac{1}{2} > 0,$$

where the infimum is taken over all the estimators $\hat{T} = \hat{T}\big(X_1, \ldots, X_n\big)$.

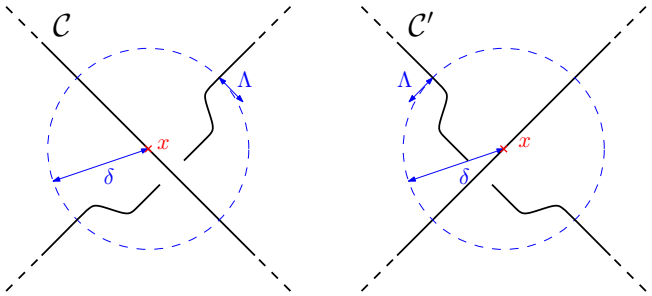


Figure 1: Inconsistency of tangent space estimation for $\tau_{min} = 0$.

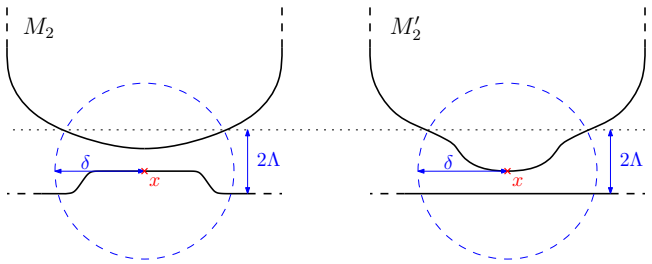


Figure 2: Inconsistency of curvature estimation for $\tau_{min} = 0$.

Moreover, for any $D \geq d+1$, provided that $L_3/L_\perp^2, \ldots, L_k/L_\perp^{k-1}, L_\perp^d/f_{min}$ and $f_{max}/L_\perp^d$ are large enough (depending only on d and k), for all $n \geq 1$,

$$\inf_{\widehat{II}} \sup_{P \in \mathcal{P}^k_{(x)}} \mathbb{E}_{P^{\otimes n}} \left\| II_x^M \circ \pi_{T_xM} - \widehat{II} \right\|_{op} \geq \frac{L_\perp}{4} > 0,$$

where the infimum is taken over all the estimators $\widehat{II} = \widehat{II}(X_1, \ldots, X_n)$.

In other words, if the class of submanifolds is allowed to have arbitrarily small reach, no estimator can perform uniformly well to estimate neither T_xM nor II_x^M. And this, even though each of the underlying submanifolds have arbitrarily smooth parametrizations. Indeed, if two parts of M can nearly intersect around x at an arbitrarily small scale $\Lambda \to 0$, no estimator can decide whether the direction (resp. curvature) of M at x is that of the first part or the second part (see Figures 1 and 2).

3 Main Results

Let us now move to the description of the main results, that consist of minimax upper and lower bounds for each object of interest. Given an i.i.d. n-sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$ with unknown common distribution $P \in \mathcal{P}^k$ having support M, we detail non-asymptotic rates for the estimation of tangent spaces $T_{X_j}M$, second fundamental forms $II_{X_j}^M$, and M itself.

For this, we need one more piece of notation. For $1 \leq j \leq n$, $P_n^{(j)}$ denotes integration with respect to $1/(n-1)\sum_{i\neq j}\delta_{(X_i - X_j)}$, and $y^{\otimes i}$ denotes the $D \times i$-dimensional vector $(y, \dots, y)$. For a constant $t > 0$ and a bandwidth $h > 0$ to be chosen later, we define the local polynomial estimator $(\hat{\pi}_j, \hat{T}_{2,j}, \dots, \hat{T}_{k-1,j})$ at X_j to be any element of

$$\underset{\pi, \sup_{2\leq i\leq k}\|T_i\|_{op}\leq t}{\arg\min} P_n^{(j)} \left[\left\| x - \pi(x) - \sum_{i=2}^{k-1} T_i(\pi(x)^{\otimes i}) \right\|^2 \mathbb{1}_{\mathcal{B}(0,h)}(x) \right], \qquad (2)$$

where π ranges among all the orthogonal projectors on d-dimensional subspaces, and $T_i : \left(\mathbb{R}^D\right)^i \to \mathbb{R}^D$ among the symmetric tensors of order i such that $\|T_i\|_{op} \leq t$. For $k = 2$, the sum over the tensors T_i is empty, and the integrated term reduces to $\|x - \pi(x)\|^2 \mathbb{1}_{\mathcal{B}(0,h)}(x)$. By compactness of the domain of minimization, such a minimizer exists almost surely. In what follows, we will work with a maximum scale $h \leq h_0$, with

$$h_0 = \frac{\tau_{min} \wedge L_\perp^{-1}}{8}.$$

Note that the set of d-dimensional orthogonal projectors is not convex, leading to a more involved optimization problem than usual least squares. In practice, this problem may be solved using tools from optimization on Grassman manifolds [UM14], or adopting a two-stage procedure such as in [CP05]: from local PCA, a first d-dimensional space is estimated at each sample point, along with an orthonormal basis of it. Then, the optimization problem (2) is expressed as a minimization problem in terms of the coefficients of $(\pi_j, T_{2,j}, \dots, T_{k,j})$ in this basis under orthogonality constraints. It is worth mentioning that a similar problem is explicitly solved in [CC16], leading to an optimal tangent space estimation procedure in the case $k = 3$.

The constraint $\|T_i\|_{op} \leq t$ involves a parameter t to be calibrated. As will be shown in the following section, it is enough to choose t roughly smaller than $1/h$, but still larger than the unknown norm of the optimal tensors $\|T_i^*\|_{op}$. Hence, for $h \to 0$, the choice $t = h^{-1}$ works to guarantee optimal convergence rates. Such a constraint on the higher order tensors might have been stated under the form of a $\|.\|_{op}$-penalized least squares minimization — as in ridge regression — leading to the same results.

3.1 Tangent Spaces

By definition, the tangent space $T_{X_j}M$ is the best linear approximation of M nearby X_j. Therefore, it is natural to take the range of the first order term minimizing (2) and write $\hat{T}_j = \operatorname{im}\hat{\pi}_j$. The $\hat{T}_j$'s approximate simultaneously the $T_{X_j}M$'s with high probability, as stated below.

Theorem 6. *Assume that* $t \geq C_{k,d,\tau_{min},\mathbf{L}} \geq \sup_{2\leq i\leq k}\|T_i^*\|_{op}$. *Set* $h = \left(C_{d,k}\frac{\log(n) f_{max}^2}{(n-1)f_{min}^3}\right)^{1/d}$, *for* $C_{d,k}$ *large enough. If* n *is large enough so that* $h \leq h_0$, *then with probability*

at least $1 - \left(\frac{1}{n}\right)^{k/d}$,

$$\max_{1\leq j\leq n} \angle\big(T_{X_j}M, \hat{T}_j\big) \leq C_{d,k,\tau_{min},\mathbf{L}} \sqrt{\frac{f_{max}}{f_{min}}} h^{k-1}(1+th).$$

As a consequence, taking $t = h^{-1}$*, for* n *large enough,*

$$\sup_{P\in\mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} \max_{1\leq j\leq n} \angle\big(T_{X_j}M, \hat{T}_j\big) \leq C \left(\frac{\log(n)}{n-1}\right)^{\frac{k-1}{d}},$$

where $C = C_{d,k,\tau_{min},\mathbf{L},f_{min},f_{max}}$.

The same bound holds for the estimation of T_xM at a prescribed $x \in M$. For that, simply take $P_n^{(x)} = 1/n \sum_i \delta_{(X_i - x)}$ as integration in (2).

This result is in line with those of [CP05] in terms of the sample size dependency $(1/n)^{(k-1)/d}$. Besides, it shows that the convergence rate of our estimator does not depend on the ambient dimension D, even in codimension greater than 2. When $k = 2$, we recover the same rate as [AL15], where we used local PCA for estimation, that is a reformulation of (2). When $k \geq 3$, this procedure outperforms PCA-based estimators of [SW12] and [TVF13], where convergence rates of the form $(1/n)^\alpha$ is obtained for $1/d < \alpha < 2/d$. This bound also recovers the result of [CC16] in the case $k = 3$, where a similar procedure is used. Moreover, Theorem 6 nearly matches the following lower bound.

Theorem 7. *If* $\tau_{min}L_\perp, \ldots, \tau_{min}^{k-1}L_k, (\tau_{min}^d f_{min})^{-1}$ *and* $\tau_{min}^d f_{max}$ *are large enough (depending only on* d *and* k*), then*

$$\inf_{\hat{T}} \sup_{P\in\mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} \angle\big(T_{X_1}M, \hat{T}\big) \geq c_{d,k,\tau_{min}} \left(\frac{1}{n-1}\right)^{\frac{k-1}{d}},$$

where the infimum is taken over all the estimators $\hat{T} = \hat{T}(X_1, \ldots, X_n)$.

Hence, up to a $\log n$ factor, the rate $n^{-(k-1)/d}$ is optimal for tangent space estimation on the model $\mathcal{P}^k$. The rate $(\log n/n)^{1/d}$ obtained in [AL15] for $k = 2$ is therefore optimal, as well as the rate $(\log n/n)^{2/d}$ given in [CC16] for $k = 3$. The rate $n^{-(k-1)/d}$ naturally appears on the the model $\mathcal{P}^k$, since it consists of $\mathcal{C}^k$-submanifolds, and tangent spaces are differential objects of order 1, yielding the shift $k-1$. Again, the same lower bound holds for the estimation of T_xM at a fixed point x in the model $\mathcal{P}^k_{(x)}$. Interestingly, the tools used to derive the lower bound for T_xM (x fixed) is much less involved than for $T_{X_1}M$ (X_1 random and depending on the distribution P). In the latter case, a conditional Assouad's Lemma (Lemma 16) is used. We will detail these differences in Section 5.2.

3.2 Curvature

The second fundamental form $II^M_{X_j} : T_{X_j}M \times T_{X_j}M \to T_{X_j}M^\perp \subset \mathbb{R}^D$ is a symmetric bilinear map that encodes completely the curvature of M at X_j

[dC92, Chap. 6, Proposition 3.1]. Estimating it only from a point cloud $\mathbb{X}_n$ does not trivially make sense, since $II^M_{X_j}$ has domain $T_{X_j}M$ which is unknown. To bypass this issue we extend $II^M_{X_j}$ to $\mathbb{R}^D$. That is, we consider the estimation of $II^M_{X_j} \circ \pi_{T_{X_j}M}$ which has full domain $\mathbb{R}^D$. Following the same ideas as in the previous Section 3.1, we use the second order tensor $\hat{T}_{2,j} \circ \hat{\pi}_j$ obtained in (2) to estimate $II^M_{X_j} \circ \pi_{T_{X_j}M}$.

Theorem 8. *Let $k \geq 3$. Take h as in Theorem 6 and $t = 1/h$. If n is large enough so that $h \leq h_0$ and $h^{-1} \geq C^{-1}_{k,d,\tau_{min},\mathbf{L}} \geq (\sup_{2\leq i\leq k} \|T_i^*\|_{op})^{-1}$, then with probability at least $1 - \left(\frac{1}{n}\right)^{k/d}$,*

$$\max_{1\leq j\leq n} \left\| II^M_{X_j} \circ \pi_{T_{X_j}M} - \hat{T}_{2,j} \circ \hat{\pi}_j \right\|_{op} \leq C_{d,k,\tau_{min},\mathbf{L}} \sqrt{\frac{f_{max}}{f_{min}}} h^{k-2}.$$

In particular, for n large enough,

$$\sup_{P\in\mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} \max_{1\leq j\leq n} \left\| II^M_{X_j} \circ \pi_{T_{X_j}M} - \hat{T}_{2,j} \circ \hat{\pi}_j \right\|_{op} \leq C \left(\frac{\log(n)}{n-1} \right)^{\frac{k-2}{d}},$$

where $C = C_{d,k,\tau_{min},\mathbf{L},f_{min},f_{max}}$.

Interestingly, Theorems 6 and 8 are enough to provide estimators of various notions of curvature. For instance, consider the scalar curvature [dC92, Section 4.4] at a point X_j, defined by

$$Sc^M_{X_j} = \frac{1}{d(d-1)} \sum_{r\neq s} \left[\left\langle II^M_{X_j}(e_r, e_r), II^M_{X_j}(e_s, e_s) \right\rangle - \|II^M_{X_j}(e_r, e_s)\|^2 \right],$$

where $(e_r)_{1\leq r\leq d}$ is an orthonormal basis of $T_{X_j}M$. A plugin estimator of $Sc^M_{X_j}$ is

$$\widehat{Sc}_j = \frac{1}{d(d-1)} \sum_{r\neq s} \left[\left\langle \hat{T}_{2,j}(\hat{e}_r, \hat{e}_r), \hat{T}_{2,j}(\hat{e}_s, \hat{e}_s) \right\rangle - \|\hat{T}_{2,j}(\hat{e}_r, \hat{e}_s)\|^2 \right],$$

where $(\hat{e}_r)_{1\leq r\leq d}$ is an orthonormal basis of $\hat{T}_{X_j}M$. Theorems 6 and 8 yield

$$\mathbb{E}_{P^{\otimes n}} \max_{1\leq j\leq n} \left| \widehat{Sc}_j - Sc^M_{X_j} \right| \leq C_{d,k,\tau_{min},\mathbf{L},f_{min},f_{max}} \left(\frac{\log(n)}{n-1} \right)^{\frac{k-2}{d}}.$$

The (near-)optimality of the bound stated in Theorem 8 is assessed by the following lower bound.

Theorem 9. *If $\tau_{min}L_\perp, \ldots, \tau^{k-1}_{min}L_k, (\tau^d_{min}f_{min})^{-1}$ and $\tau^d_{min}f_{max}$ are large enough (depending only on d and k), then*

$$\inf_{\widehat{II}} \sup_{P\in\mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} \left\| II^M_{X_1} \circ \pi_{T_{X_1}M} - \widehat{II} \right\|_{op} \geq c_{d,k,\tau_{min}} \left(\frac{1}{n-1} \right)^{\frac{k-2}{d}},$$

where the infimum is taken over all the estimators $\widehat{II} = \widehat{II}(X_1, \ldots, X_n)$.

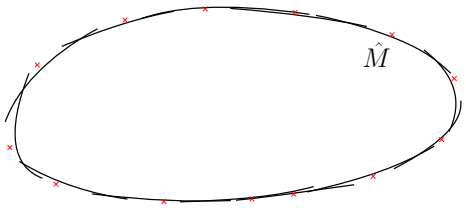


Figure 3: $\hat{M}$ is a union of polynomial patches at sample points.

The same remarks as in Section 3.1 hold. If the estimation problem consists in approximating II_x^M at a fixed point x known to belong to M beforehand, we obtain the same rate. The ambient dimension D still plays no role. The shift $k-2$ in the rate of convergence on a $\mathcal{C}^k$-model can be interpreted as the order of derivation of the object of interest, that is 2 for curvature.

Notice that the lower bound (Theorem 9) does not require $k \geq 3$. Hence, we get that for $k = 2$, curvature cannot be estimated uniformly consistently on the $\mathcal{C}^2$-model $\mathcal{P}^2$. This seems natural, since the estimation of a second order quantity should require an additional degree of smoothness.

3.3 Support Estimation

For each $1 \leq j \leq n$, the minimization (2) outputs a series of tensors $(\hat{\pi}_j, \hat{T}_{2,j}, \ldots, \hat{T}_{k-1,j})$. This collection of multidimensional monomials can be further exploited as follows. By construction, they fit M at scale h around X_j, so that

$$\widehat{\Psi}_j(v) = X_j + v + \sum_{i=2}^{k-1} \hat{T}_{i,j}\left(v^{\otimes i}\right)$$

is a good candidate for an approximate parametrization in a neighborhood of X_j. We do not know the domain $T_{X_j}M$ of the initial parametrization, though we have at hand an approximation $\hat{T}_j = \operatorname{im}\hat{\pi}_j$ which was proved to be consistent in Section 3.1. As a consequence, we let the support estimator based on local polynomials $\hat{M}$ be

$$\hat{M} = \bigcup_{j=1}^{n} \widehat{\Psi}_j\left(\mathcal{B}_{\hat{T}_j}(0, 7h/8)\right).$$

The set $\hat{M}$ has no reason to be globally smooth, since it consists of a union of polynomial patches that are not linked together (Figure 3). However, $\hat{M}$ is provably close to M for the Hausdorff distance.

Theorem 10. *With the same assumptions as Theorem 8, with probability at least* $1 - 2\left(\frac{1}{n}\right)^{\frac{k}{d}}$*, we have*

$$d_H\left(M, \hat{M}\right) \leq C_{d,k,\tau_{min},\mathbf{L},f_{min},f_{max}} h^k.$$

In particular, for n large enough,

$$\sup_{P\in\mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} d_H\big(M,\hat{M}\big) \leq C\left(\frac{\log(n)}{n-1}\right)^{\frac{k}{d}},$$

where $C = C_{d,k,\tau_{min},\mathbf{L},f_{min},f_{max}}$.

For $k=2$, we recover the rate $(\log n/n)^{2/d}$ obtained in [AL15, GPPVW12, KZ15]. However, our estimator $\hat{M}$ is an unstructured union of d-dimensional balls in $\mathbb{R}^D$. Consequently, $\hat{M}$ does not recover the topology of M as the estimator of [AL15] does.

When $k \geq 3$, $\hat{M}$ outperforms reconstruction procedures based on a somewhat piecewise linear interpolation [AL15, GPPVW12], and achieves the faster rate $(\log n/n)^{k/d}$ for the Hausdorff loss. This seems quite natural, since our procedure fits higher order terms. This is done at the price of a probably worse dependency on the dimension d than in [AL15, GPPVW12]. Theorem 10 is now proved to be (almost) minimax optimal.

Theorem 11. *If $\tau_{min}L_\perp, \ldots, \tau_{min}^{k-1}L_k, (\tau_{min}^d f_{min})^{-1}$ and $\tau_{min}^d f_{max}$ are large enough (depending only on d and k), then for n large enough,*

$$\inf_{\hat{M}} \sup_{P\in\mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} d_H\big(M,\hat{M}\big) \geq c_{d,k,\tau_{min}}\left(\frac{1}{n}\right)^{\frac{k}{d}},$$

where the infimum is taken over all the estimators $\hat{M} = \hat{M}(X_1,\ldots,X_n)$.

Theorem 11 is obtained from Le Cam's Lemma (Theorem 15). Let us note that it is likely for the extra $\log n$ term appearing in Theorem 10 to actually be present in the minimax rate. Roughly, it is due to the fact that the Hausdorff distance d_H is similar to a L^∞ loss. The $\log n$ term may be obtained in Theorem 11 with the same combinatorial analysis as in [KZ15] for $k=2$.

4 Conclusion, Prospects

In this article, we derived non-asymptotic bounds for inference of geometric objects associated with smooth submanifolds $M \subset \mathbb{R}^D$. We focused on tangent spaces, second fundamental forms, and the submanifold itself. We introduced new regularity classes $\mathcal{C}^k_{\tau_{min},\mathbf{L}}$ for submanifolds that extend naturally the case $k=2$. For each object of interest, the proposed estimator relies on local polynomials that can be computed through a least square minimization. Minimax lower bounds were presented, matching the upper bounds up to $\log n$ factors.

The implementation of (2) needs to be investigated. The non-convexity of the criterion comes from that we minimize over the space of orthogonal projectors, which is non-convex. However, that space is pretty well understood, and it seems possible to implement gradient descents on it [UM14]. Another way to improve our procedure could be to fit orthogonal polynomials instead

of monomials. Such a modification may also lead to improved dependency on the dimension d and the regularity k in the bounds for both tangent space and support estimation.

As a first attempt to a minimax study over models of higher order regularity $\mathcal{C}^k$ ($k \geq 3$) for submanifolds, we chose not to include noise. This is a limitation of the model $\mathcal{P}^k$, and one could argue that the methods described are not robust. However, with outliers in the model $\mathcal{C}^2$, [AL15] proposes an iterative denoising procedure based on tangent space estimation. It exploits the fact that tangent space estimation allows to remove a part of outliers, and removing outliers enhances tangent space estimation. An interesting question would be to study how this method can apply with local polynomials.

Another open question is that of exact topology recovering with fast rates for $k \geq 3$. Indeed, $\hat{M}$ converges at rate $(\log n/n)^{k/d}$ but is unstructured. It would be nice to glue the patches of $\hat{M}$ together, for example using interpolation techniques, following the ideas of [FIK+15].

5 Proofs

5.1 Local Polynomials

We now turn to the proof of the upper bounds of Section 3. First, to relate the existence of parametrizations Ψ_p's to a local polynomial decomposition, the following lemma is needed.

Lemma 12. *For any $M \in \mathcal{C}^k_{\tau_{min},\mathbf{L}}$ and $x \in M$, the following holds.*

(i) For all $v_1, v_2 \in \mathcal{B}_{T_xM}\left(0, \frac{1}{4L_\perp}\right)$,

$$\frac{3}{4}\|v_2 - v_1\| \leq \|\Psi_x(v_2) - \Psi_x(v_1)\| \leq \frac{5}{4}\|v_2 - v_1\|.$$

(ii) For all $h \leq \frac{1}{4L_\perp} \wedge \frac{2\tau_{min}}{5}$,

$$M \cap \mathcal{B}\left(x, \frac{3h}{5}\right) \subset \Psi_x\left(\mathcal{B}_{T_xM}(x,h)\right) \subset M \cap \mathcal{B}\left(x, \frac{5h}{4}\right).$$

(iii) For all $h \leq \frac{\tau_{min}}{2}$,

$$\mathcal{B}_{T_xM}\left(0, \frac{7h}{8}\right) \subset \pi_{T_xM}\left(\mathcal{B}(x,h) \cap M\right).$$

(iv) Denoting by $\pi^* = \pi_{T_xM}$ the orthogonal projection onto T_xM, for all $x \in M$, there exist multilinear maps $T_2^*, \ldots, T_{k-1}^*$ from T_xM to $\mathbb{R}^D$, and R_k such that for all $y \in \mathcal{B}\left(x, \frac{\tau_{min} \wedge L_\perp^{-1}}{4}\right) \cap M$,

$$\begin{aligned} y - x = \pi^*(y-x) + T_2^*(\pi^*(y-x)^{\otimes 2}) + \ldots + T_{k-1}^*(\pi^*(y-x)^{\otimes k-1}) \\ + R_k(y-x), \end{aligned}$$

with

$$\|R_k(y-x)\| \leq C\|y-x\|^k \qquad \text{and} \qquad \|T_i^*\|_{op} \leq L_i' \text{ for } 2 \leq i \leq k-1,$$

where L_i' *depends on* $d, k, \tau_{min}, L_\perp, \ldots, L_i$, *and* C *on* d, k, τ_{min}, $L_\perp$,..., L_k. *Moreover, for* $k \geq 3$, $T_2^* = II_x^M$.

(v) *For all* $x \in M$, $\left\|II_x^M\right\|_{op} \leq 1/\tau_{min}$. *In particular, the sectional curvatures of* M *satisfy*

$$\frac{-2}{\tau_{min}^2} \leq \kappa \leq \frac{1}{\tau_{min}^2}.$$

The proof of Lemma 12 can be found in Section A.2.

We are now in position to analyze local polynomial estimators. For clarity's sake, the bounds are given for $j = 1$, where we denote by $\hat{\pi}$, $\hat{T}_i$ $(2 \leq i \leq k-1)$ the fitted polynomials of (2), and $P_{n-1} = P_{n-1}^{(1)}$. The results of Theorems 6, 8, and 10 then follow from a straightforward union bound. We also set $k \geq 3$, the case $k = 2$ proceeding from the same derivation, omitting the higher order tensors. Without loss of generality, we can assume that $X_1 = 0$ and that T_0M is spanned by the first d vectors of the canonical basis, so that $\pi^*(x) = (x_1, \ldots, x_d, 0, \ldots, 0) = (x_{1:d}, 0, \ldots, 0)$.

Recall that $h_0 = (\tau_{min} \wedge L_\perp^{-1})/8$. According to Lemma 12, if $M \in \mathcal{C}^k_{\tau_{min},\mathbf{L}}$, for any $x \in M$ such that $\|x\| \leq h_0$, we may write

$$x = \pi^*(x) + T_2^*(\pi^*(x)^{\otimes 2}) + \ldots + T_{k-1}^*(\pi^*(x)^{\otimes k-1}) + R_k(x),$$

where $\|R_k(x)\| \leq C_{\tau_{min},\mathbf{L}}\|x\|^k$. Every coordinate of $(\hat{T}_i - T_i^*)(\pi^*(x))$ may be thought of as a polynomial map in the variable $x_{1:d}$. Thus, proximity between $\hat{T}_i$ and T_i^* will be first stated in terms of polynomial norm.

Let $\mathbb{R}^k[x_{1:d}]$ denote the set of real-valued polynomial functions in d variables with degree less than k. For $Q \in \mathbb{R}^k[x_{1:d}]$, we denote by $\|Q\|_2$ the Euclidean norm of its coefficients, and by Q_h the polynomial defined by $Q_h(x_{1:d}) = Q(hx_{1:d})$. The following result relates the $L^2(P_{n-1})$ norm involved in (2) to polynomial norms.

Proposition 13. *Set* $h = \left(K\frac{\log(n)}{n-1}\right)^{\frac{1}{d}}$. *There exist constants* $\kappa_{k,d}$, $c_{k,d}$ *and* C_d *such that, if* $K \geq (\kappa_{k,d} f_{max}^2/f_{min}^3)$ *and* n *is large enough so that* $h \leq h_0 \leq \tau_{min}/4$, *then with probability at least* $1 - \left(\frac{1}{n}\right)^{\frac{k}{d}+1}$, *we have*

$$\begin{array}{rcl} P_{n-1}[Q^2(\pi^*(x))\mathbb{1}_{\mathcal{B}(h)}(x)] & \geq & c_{k,d}h^d f_{min}\|Q_h\|_2^2, \\ N(h) & \leq & C_d f_{max}(n-1)h^d, \end{array}$$

for every $Q \in \mathbb{R}^k[x_{1:d}]$, *where* $N(h) = \sum_{j=2}^n \mathbb{1}_{\mathcal{B}(0,h)}(X_j)$.

The proof of Proposition 13 is deferred to Section B.2. From now on we assume that the probability event defined in Proposition 13 occurs. For short, with a slight abuse of notation, we denote by $T_{p:q}(x)$ the sum $T_p(x^{\otimes p}) + \ldots + T_q(x^{\otimes q})$, and by $\mathcal{R}_{n-1}(\pi, T_2, \ldots, T_{k-1})$ the empirical criterion defined by (2). Since for $t \geq \max_{i=2,\ldots,k-1} \|T_i^*\|_{op}$,

$$\mathcal{R}_{n-1}(\hat{\pi}, \hat{T}_1, \ldots, \hat{T}_{k-1}) \leq \mathcal{R}_{n-1}(\pi^*, T_2^*, \ldots, T_{k-1}^*) \leq C_{\tau_{min},\mathbf{L}} h^{2k} N(h)/(n-1)$$

according to (1), we may write

$$\begin{aligned} C_{\tau_{min},\mathbf{L}} h^{2k} \frac{N(h)}{n-1} &\geq \mathcal{R}_{n-1}(\hat{\pi}, \hat{T}_2, \ldots, \hat{T}_{k-1}) \\ &= P_{n-1}\left(\left\| (\pi^* - \hat{\pi})(x) + (T_{2:k-1}^* \circ \pi^* - \hat{T}_{2:k-1} \circ \hat{\pi})(x) \right. \right. \\ &\qquad \left. \left. + R_k(x) \right\|^2 \mathbb{1}_{\mathcal{B}(0,h)}(x) \right), \end{aligned}$$

with $\|R_k(x)\| \leq C_{\tau_{min},\mathbf{L}} h^{2k}$. It follows that

$$\begin{aligned} P_{n-1}\left(\left\| (\pi^* - \hat{\pi})(x) + (T_{2:k-1}^* \circ \pi^* - \hat{T}_{2:k-1} \circ \hat{\pi})(x) \right\|^2 \mathbb{1}_{\mathcal{B}(0,h)}(x) \right) & \\ \leq C_{\tau_{min},\mathbf{L}} h^{2k} \frac{N(h)}{n-1} & \\ \leq C_{\tau_{min},\mathbf{L},d} f_{max} h^{d+2k}. & \end{aligned}$$

On the other hand, using (1) again yields, for $x \in \mathcal{B}(0,h) \cap M$,

$$\begin{aligned} (\pi^* - \hat{\pi})(x) + (T_{2:k-1}^* \circ \pi^* - \hat{T}_{2:k-1} \circ \hat{\pi})(x) & \\ = T_1'(\pi^*(x)) + T_2'(\pi^*(x)^{\otimes 2}) + T_{3:k}'(\pi^*(x)) + \hat{\pi}(R_k(x)) + R_k'(x), & \end{aligned}$$

with $\|R_k(x)\| \leq C_{\tau_{min},\mathbf{L}} h^k$, $\|R_k'(x)\| \leq t C_{\tau_{min},k,\mathbf{L}} h^{k+1}$ since only tensors of order greater than 2 are involved in R_k', and

$$\begin{aligned} T_1'(\pi^*(x)) &= (\pi^* - \hat{\pi})\pi^*(x) \\ T_2'(\pi^*(x)^{\otimes 2}) &= (\pi^* - \hat{\pi})(T_2^*(\pi^*(x)^{\otimes 2})) + (T_2^* \circ \pi^* - \hat{T}_2 \circ \hat{\pi})\big(\pi^*(x)^{\otimes 2}\big). \end{aligned}$$

Hence,

$$\begin{aligned} P_{n-1}\left(\left\| T_1'(\pi^*(x)) + T_2'(\pi^*(x)^{\otimes 2}) + T_{3:k}'(\pi^*(x)) \right\|^2 \mathbb{1}_{\mathcal{B}(0,h)}(x) \right) & \\ \leq C_{\tau_{min},\mathbf{L},d} f_{max} h^{d+2k} (1 + ht). & \quad (3) \end{aligned}$$

The left-hand side of (3) may be decomposed coordinate-wise as

$$\begin{aligned} & P_{n-1}\left(\left\| T_1'(\pi^*(x)) + T_2'(\pi^*(x)^{\otimes 2}) + T_{3:k}'(\pi^*(x)) \right\|^2 \mathbb{1}_{\mathcal{B}(0,h)}(x) \right) \\ &= \sum_{j=1}^{D} P_{n-1}\left(\left(T_1'^{(j)}(\pi^* x) + T_2'^{(j)}(\pi^*(x)^{\otimes 2}) + T_{3:k}'^{(j)}(\pi^*(x)) \right)^2 \mathbb{1}_{\mathcal{B}(0,h)}(x) \right), \end{aligned}$$

where for any tensor T, $T^{(j)}$ denotes the j-th coordinate of T and is considered as a real valued j-order polynomial. Then, for every j, Proposition 13 leads to

$$\begin{aligned}
&P_{n-1}\left(\left(T'^{(j)}_1(\pi^*(x)) + T'^{(j)}_2(\pi^*(x)^{\otimes 2}) + T'^{(j)}_{3:k}(\pi^*(x))\right)^2 \mathbb{1}_{\mathcal{B}(0,h)}(x)\right) \\
&\qquad \geq c_{d,k} f_{min} h^d \left\| \left(T'^{(j)}_1(\pi^*(x)) + T'^{(j)}_2(\pi^*(x)^{\otimes 2}) + T'^{(j)}_{3:k}(\pi^*(x))\right)_h \right\|_2^2 \\
&\qquad = c_{d,k} f_{min} h^d \sum_{i=1}^{k} \left\| \left(T'^{(j)}_i(\pi^*(x)^{\otimes i})\right)_h \right\|_2^2.
\end{aligned}$$

Summing all contributions leads to

$$c_{d,k} f_{min} \sum_{j=1}^{D} \sum_{i=1}^{k} \left\| \left(T'^{(j)}_i(\pi^*(x)^{\otimes i})\right)_h \right\|_2^2 \leq C_{k,\mathbf{L},d,\tau_{min}} f_{max} h^{2k}(1 + t^2h^2).$$

This entails

$$\|T'_i\|_{op}^2 \leq C_{d,k,\mathbf{L},\tau_{min}} \frac{f_{max}}{f_{min}} h^{2(k-i)}(1 + t^2h^2), \tag{4}$$

for $1 \leq i \leq k$, as well as

$$\left\|(\pi^* - \hat{\pi})(x) + (T^*_{2:k-1} \circ \pi^* - \hat{T}_{2:k-1} \circ \hat{\pi})(x)\right\| \leq C_{d,k,\mathbf{L},\tau_{min}} \sqrt{\frac{f_{max}}{f_{min}}} h^k(1 + th), \tag{5}$$

for $x \in \mathcal{B}(0,h) \cap M$, according to (1).

5.1.1 Bounds for Tangent Space Estimation

Noting that

$$\|T'_1\|_{op} = \|(\pi^* - \hat{\pi})\pi^*\|_{op} = \|\pi_{\hat{T}_1^\perp} \circ \pi^*\| = \angle(T_0M, \hat{T}_1)$$

from [GVL96, Section 2.6.2], and using (4) for $i = 1$ yields Theorem 6.

5.1.2 Bounds for Curvature Estimation

In accordance with assumptions of Theorem 8, we assume that $\max_{2\leq i\leq k} \|T^*_i\|_{op} \leq t \leq 1/h$. Since

$$T'_2(\pi^*(x)^{\otimes 2}) = (\pi^* - \hat{\pi})(T^*_2(\pi^*(x)^{\otimes 2})) + (T^*_2 \circ \pi^* - \hat{T}_2 \circ \hat{\pi})(\pi^*(x)^{\otimes 2}),$$

we deduce that

$$\|T^*_2 \circ \pi^* - \hat{T}_2 \circ \hat{\pi}\|_{op} \leq \|T'_2\|_{op} + \|\hat{\pi} - \pi^*\|_{op} + \|\hat{T}_2 \circ \hat{\pi} \circ \pi^* - \hat{T}_2 \circ \hat{\pi} \circ \hat{\pi}\|_{op}.$$

Using (4) with $i = 1, 2$ and $th \leq 1$ leads to

$$\|T^*_2 \circ \pi^* - \hat{T}_2 \circ \hat{\pi}\|_{op} \leq C_{d,k,\mathbf{L},\tau_{min}} \sqrt{\frac{f_{max}}{f_{min}}} h^{k-2}.$$

Finally, Lemma 12 states that $II^M_{X_1} = T^*_2$, hence Theorem 8 is proved.

5.1.3 Bounds for Reconstruction

Let $v \in \mathcal{B}_{\hat{T}_0 M}(0, 7h/8)$ be fixed. Notice that $\pi^*(v) \in \mathcal{B}_{T_0 M}(0, 7h/8)$. Hence, according to Lemma 12, there exists $x \in \mathcal{B}(0,h) \cap M$ such that $\pi^*(v) = \pi^*(x)$. We may write

$$\widehat{\Psi}(v) = v + \sum_{i=2}^{k-1} \hat{T}_i(v^{\otimes i}) = \pi^*(v) + \sum_{i=2}^{k-1} \hat{T}_i(\pi^*(v)^{\otimes i}) + R_k(v),$$

where, since $\|\hat{T}_i\|_{op} \leq 1/h$, $\|R_k(v)\| \leq C_{k,d,\tau_{min},\mathbf{L}} \sqrt{f_{max}/f_{min}} h^k$ according to (4). Then, according to (5),

$$\begin{aligned}\pi^*(v) + \sum_{i=2}^{k-1} \hat{T}_i(\pi^*(v)^{\otimes i}) &= \pi^*(v) + \sum_{i=2}^{k-1} T_i^*(\pi^*(v)^{\otimes i}) + R'(\pi^*(v)) \\ &= \pi^*(x) + \sum_{i=2}^{k-1} T_i^*(\pi^*(x)^{\otimes i}) + R'(\pi^*(x)),\end{aligned}$$

where $\|R'(\pi^*(x))\| \leq C_{k,d,\tau_{min},L} \sqrt{f_{max}/f_{min}} h^{k+1}$. According to Lemma 12, we deduce that $\|\widehat{\Psi}(v) - x\| \leq C_{k,d,\tau_{min},\mathbf{L}} \sqrt{f_{max}/f_{min}} h^k$, hence

$$\sup_{u \in \hat{M}} d(u, M) \leq C_{k,d,\tau_{min},\mathbf{L}} \sqrt{\frac{f_{max}}{f_{min}}} h^k. \tag{6}$$

Now we focus on $\sup_{x \in M} d(x, \hat{M})$. For this, we need a lemma ensuring that $\mathbb{X}_n = \{X_1, \ldots, X_n\}$ covers M with high probability.

Lemma 14. *Let $h = \left(\frac{C'_d k}{f_{min}} \frac{\log n}{n}\right)^{1/d}$ with C'_d large enough. Then for n large enough so that $h \leq \tau_{min}/8$, with probability at least $1 - \left(\frac{1}{n}\right)^{k/d}$,*

$$d_H(M, \mathbb{X}_n) \leq h.$$

The proof of Lemma 14 is given in Section B.1. Now we choose h satisfying the conditions of Proposition 13 and Lemma 14. Let x be in M and assume that $\|x - X_{j_0}\| \leq h$. According to (5) and (1), we deduce that $\|\widehat{\Psi}_{j_0}(\hat{\pi}_{j_0}(x)) - x\| \leq C_{k,d,\tau_{min},\mathbf{L}} \sqrt{f_{max}/f_{min}} h^k$. Hence, from Lemma 14,

$$\sup_{x \in M} d(x, \hat{M}) \leq C_{k,d,\tau_M,L} \sqrt{\frac{f_{max}}{f_{min}}} h^k \tag{7}$$

with probability at least $1 - 2\left(\frac{1}{n}\right)^{k/d}$. Combining (6) and (7) gives Theorem 10.

5.2 Minimax Lower Bounds

This section is devoted to describe the main ideas of the proofs of the minimax lower bounds, Theorems 7, 9 and 11. The methods we use rely on hypothesis comparison [Yu97]. We recall that for two distributions Q and Q' defined on the same space, the total variation distance $TV(Q, Q')$ and the L^1 test affinity $\|Q \wedge Q'\|_1$ are given by

$$TV(Q, Q') = \frac{1}{2} \int |dQ - dQ'|, \qquad \|Q \wedge Q'\|_1 = \int dQ \wedge dQ',$$

where dQ and dQ' denote densities of Q and Q' with respect to any dominating measure.

5.2.1 Le Cam's Lemma and Consequences

The first technique we use, involving only two hypotheses, is usually referred to as Le Cam's Lemma. Let $\mathcal{P}$ be a model and $\theta(P)$ be the parameter of interest. Assume that $\theta(P)$ belongs to a pseudo-metric space $(\mathcal{D}, d)$, that is $d(\cdot, \cdot)$ is symmetric and satisfies the triangle inequality. Le Cam's Lemma can be adapted to our framework as follows.

Theorem 15 (Le Cam's Lemma [Yu97])**.** *For all P, P' in the model $\mathcal{P}$,*

$$\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{\otimes n}} d(\theta(P), \hat{\theta}) \geq \frac{1}{2} d\left(\theta(P), \theta(P')\right) \left\| P^{\otimes n} \wedge P'^{\otimes n} \right\|_1 ,$$

where the infimum is taken over all the estimators $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$.

Moreover, $\|P^{\otimes n} \wedge P'^{\otimes n}\|_1 \geq \|P \wedge P'\|_1^n = (1 - TV(P, P'))^n$.

We derive Theorem 11, as well as Theorems 7 and 9 with fixed base point x, $\theta(P)$ being $supp(P) = M$, $T_x M$ and $II_x^M \circ \pi_{T_x M}$ respectively. The hypotheses P, P' are built in Section 5.2.3. Such constructions are not substantially new in minimax geometric inference [GPPVW12]. Therefore, we do not detail it further.

5.2.2 Conditional Assouad's Lemma

Now, consider the estimation of the differential quantities $T_{X_1} M$ and $II_{X_1}^M$ with random base point X_1. In both cases, the loss can be cast as

$$\begin{aligned} E_{P^{\otimes n}}\ d(\theta_{X_1}(P), \hat{\theta}) &= \mathbb{E}_{P^{\otimes n-1}} \left[E_P\ d(\theta_{X_1}(P), \hat{\theta}) \right] \\ &= \mathbb{E}_{P^{\otimes n-1}} \left[\left\| d\big(\theta_\cdot(P), \hat{\theta}\big) \right\|_{L^1(P)} \right], \end{aligned}$$

where $\hat{\theta} = \hat{\theta}(X, X')$, with $X = X_1$ driving the parameter of interest, and $X' = (X_2, \ldots, X_n) = X_{2:n}$. Since $\|.\|_{L^1(P)}$ obviously depends on P, the technique exposed in the previous section does not apply anymore. However, a slight

adaptation of Assouad's Lemma [Yu97] with an extra conditioning on $X = X_1$ carries out for our purpose. Let us now detail a general framework where the method applies.

We let $\mathcal{X}, \mathcal{X}'$ denote measured spaces. For a probability distribution Q on $\mathcal{X} \times \mathcal{X}'$, we let (X, X') be a random variable with distribution Q. The marginals of Q on $\mathcal{X}$ and $\mathcal{X}'$ are denoted by μ and ν respectively. Let $(\mathcal{D}, d)$ be a pseudo-metric space. For $Q \in \mathcal{Q}$, we let $\theta_\cdot(Q) : \mathcal{X} \to \mathcal{D}$ be defined μ-almost surely, where μ is the marginal distribution of Q on $\mathcal{X}$. The parameter of interest is $\theta_X(Q)$, and the associated minimax risk over $\mathcal{Q}$ is

$$\inf_{\hat{\theta}} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q \left[d\big(\theta_X(Q), \hat{\theta}(X, X')\big) \right], \tag{8}$$

where the infimum is taken over all the estimators $\hat{\theta} : \mathcal{X} \times \mathcal{X}' \to \mathcal{D}$.

Given a set of probability distributions $\mathcal{Q}$ on $\mathcal{X} \times \mathcal{X}'$, write $\overline{Conv}(\mathcal{Q})$ for the set of mixture probability distributions with components in $\mathcal{Q}$. For all $\tau = (\tau_1, \dots, \tau_m) \in \{0,1\}^m$, τ^k denotes the m-tuple that differs from τ only at the kth position. We are now in position to state the conditional version of Assouad's Lemma that allows to lower bound the minimax risk (8).

Lemma 16 (Conditional Assouad)**.** *Let $m \geq 1$ be an integer and let $\{\mathcal{Q}_\tau\}_{\tau \in \{0,1\}^m}$ be a family of 2^m submodels $\mathcal{Q}_\tau \subset \mathcal{Q}$. Let $\{U_k \times U_k'\}_{1 \leq k \leq m}$ be a family of pairwise disjoint subsets of $\mathcal{X} \times \mathcal{X}'$, and $\mathcal{D}_{\tau,k}$ be subsets of $\mathcal{D}$. Assume that for all $\tau \in \{0,1\}^m$ and $1 \leq k \leq m$,*

- *for all $Q_\tau \in \mathcal{Q}_\tau$, $\theta_X(Q_\tau) \in \mathcal{D}_{\tau,k}$ on the event $\{X \in U_k\}$;*
- *for all $\theta \in \mathcal{D}_{\tau,k}$ and $\theta' \in \mathcal{D}_{\tau^k,k}$, $d(\theta, \theta') \geq \Delta$.*

For all $\tau \in \{0,1\}^m$, let $\overline{Q}_\tau \in \overline{Conv}(\mathcal{Q}_\tau)$, and write $\bar{\mu}_\tau$ and $\bar{\nu}_\tau$ for the marginal distributions of $\overline{Q}_\tau$ on $\mathcal{X}$ and $\mathcal{X}'$ respectively. Assume that if (X, X') has distribution $\overline{Q}_\tau$, X and X' are independent conditionally on the event $\{(X, X') \in U_k \times U_k'\}$, and that

$$\min_{\substack{\tau \in \{0,1\}^m \\ 1 \leq k \leq m}} \left\{ \left(\int_{U_k} d\bar{\mu}_\tau \wedge d\bar{\mu}_{\tau^k} \right) \left(\int_{U_k'} d\bar{\nu}_\tau \wedge d\bar{\nu}_{\tau^k} \right) \right\} \geq 1 - \alpha.$$

Then,

$$\inf_{\hat{\theta}} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q \left[d\big(\theta_X(Q), \hat{\theta}(X, X')\big) \right] \geq m \frac{\Delta}{2} (1 - \alpha),$$

where the infimum is taken over all the estimators $\hat{\theta} : \mathcal{X} \times \mathcal{X}' \to \mathcal{D}$.

Notice that for a model of the form $\mathcal{Q} = \{\delta_{x_0} \otimes P, P \in \mathcal{P}\}$ with fixed $x_0 \in \mathcal{X}$, one recovers the classical Assouad's Lemma [Yu97] taking $U_k = \mathcal{X}$ and $U_k' = \mathcal{X}'$. Indeed, when $X = x$ a.s, the parameter of interest $\theta_X(Q) = \theta(Q)$ can be seen as non-random.

5.2.3 Construction of Hypotheses

In order to apply Le Cam's Lemma (Theorem 15) or the conditional Assouad's Lemma (Lemma 16), we describe in this section the construction of the hypotheses involved in the different contexts of estimation. For this, the strategy consists in building distributions that are stochastically close — i.e. with a large test affinity — for which the associated parameters of interest are as different as possible. Before continuing to the precise construction, let us make two remarks about the lower bounds with random point X_1. First, the associated minimax risks (Theorems 7 and Theorem 9) involve the integration with respect to X_1. Hence, as for regression with L^p loss, multiple locations of bumps are required to yield the right rate. Second, building manifolds with different tangent spaces (resp curvature) would lead to locally singular distributions. Therefore it is natural to consider mixture distributions to get non-trivial bounds.

Let $M_0^{(0)}$ be a d-dimensional $\mathcal{C}^\infty$-submanifold of $\mathbb{R}^D$ with reach greater than 1 and such that it contains $\mathcal{B}_{\mathbb{R}^d\times\{0\}^{D-d}}(0,1/2)$. $M_0^{(0)}$ can be built for example by flattening smoothly a unit d-sphere. Since $M_0^{(0)}$ is $\mathcal{C}^\infty$, the uniform probability distribution $P_0^{(0)}$ on $M_0^{(0)}$ belongs to $\mathcal{P}^k_{1,\mathbf{L}^{(0)}}(1/V_0^{(0)}, 1/V_0^{(0)})$, for some $\mathbf{L}^{(0)}$ and $V_0^{(0)} = Vol(M_0^{(0)})$.

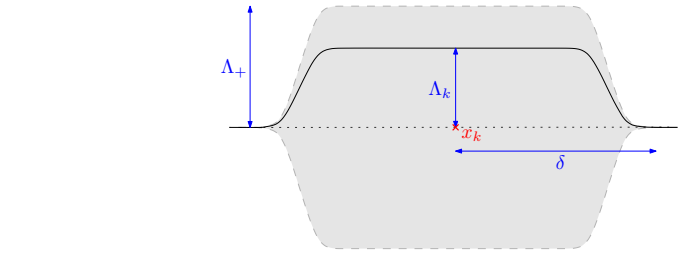
Let now $M_0 = (2\tau_{min})M_0^{(0)}$ be the submanifold obtained from $M_0^{(0)}$ by homothecy. By construction, from Proposition 4, we have $\tau_{M_0} \geq 2\tau_{min}$, $\mathcal{B}_{\mathbb{R}^d\times\{0\}^{D-d}}(0,\tau_{min}) \subset M_0$, and the uniform probability distribution P_0 on M_0 belongs to the model $\mathcal{P}^k_{2\tau_{min},\mathbf{L}}(f_{min}, f_{max})$ whenever $L_\perp \geq L_\perp^{(0)}/(2\tau_{min})$, ..., $L_k \geq L_k^{(0)}/(2\tau_{min})^{k-1}$, and provided that $f_{min} \leq \left((2\tau_{min})^d V_0^{(0)}\right)^{-1} \leq f_{max}$. Note that $L_\perp^{(0)}, \ldots, L_k^{(0)}$, $Vol(M_0^{(0)})$ depend only on d and k. For this reason, all the lower bounds will be valid for $\tau_{min}L_\perp, \ldots, \tau_{min}^{k-1}L_k, (\tau_{min}^d f_{min})^{-1}$ and $\tau_{min}^d f_{max}$ large enough to exceed the thresholds $L_\perp^{(0)}/2, \ldots, L_k^{(0)}/2^{k-1}$, $2^d V_0^{(0)}$ and $(2^d V_0^{(0)})^{-1}$ respectively.

For $0 < \delta \leq \tau_{min}/2$, let $x_1, \ldots, x_m \in M_0 \cap \mathcal{B}(0,\tau_{min})$ be such that for all $k \neq k'$, $\|x_k - x_{k'}\| \geq \delta$. A classical packing argument (see [Mas07] p. 71) shows that one can take up to $m = \lceil c_d/\delta^d \rceil$ for some $c_d > 0$. We let $e \in \mathbb{R}^D$ denote any unit vector orthogonal to $\mathbb{R}^d \times \{0\}^{D-d}$.

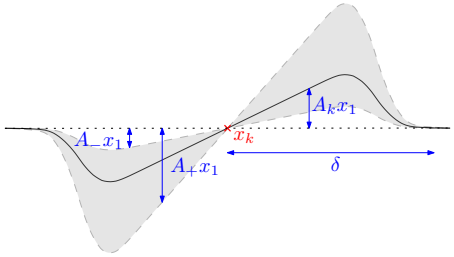
Let $\phi : \mathbb{R}^D \to [0,1]$ be a smooth scalar map such that $\phi|_{\mathcal{B}\left(0,\frac{1}{2}\right)} = 1$ and $\phi|_{\mathcal{B}(0,1)^c} = 0$. Let $\Lambda_+ > 0$ and $1 \geq A_+ > A_- > 0$ be real numbers to be chosen later. Let $\mathbf{\Lambda} = (\Lambda_1, \ldots, \Lambda_m)$ with entries $-\Lambda_+ \leq \Lambda_k \leq \Lambda_+$, and $\mathbf{A} = (A_1, \ldots, A_m)$ with entries $A_- \leq A_k \leq A_+$. For $z \in \mathbb{R}^D$, we write $z = (z_1, \ldots, z_D)$ for its coordinates in the canonical basis. For all $\tau = (\tau_1, \ldots, \tau_m) \in \{0,1\}^m$, define the bump map as

$$\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}(x) = x + \sum_{k=1}^m \phi\left(\frac{x - x_k}{\delta}\right)\left\{\tau_k A_k (x - x_k)_1^i + (1-\tau_k)\Lambda_k\right\} e. \qquad (9)$$

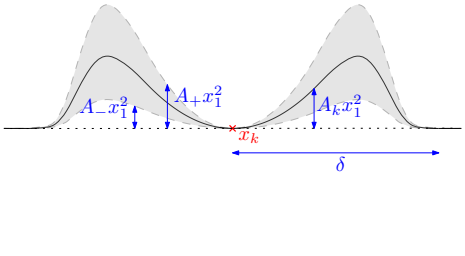
An analogous deformation map was considered in [AL15]. We let $P_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}$ denote the pushforward distribution of P_0 by $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}$, and write $M_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}$ for its



(a) Flat bump: $\tau_k = 0$.



(b) Linear bump: $\tau_k = 1$, $i = 1$.



(c) Quadratic bump: $\tau_k = 1$, $i = 2$.

Figure 4: The three shapes of the bump map $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}$ around a x_k.

support. Roughly speaking, $M_\tau^{\mathbf{\Lambda},\mathbf{A},i}$ consists of m bumps at the x_k's having different shapes (Figure 4). If $\tau_k = 0$, the bump at x_k is a symmetric plateau function and has height Λ_k. If $\tau_k = 1$, it fits the graph of the polynomial $A_k(x - x_k)_1^i$ locally. The following Lemma 17 gives differential bounds and geometric properties of $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}$. It follows straightforwardly from chain rule, similarly to Lemma 11 in [AL15].

Lemma 17. *There exists $c_{\phi,i} < 1$ such that if $A_+ \le c_{\phi,i}\delta^{i-1}$ and $\Lambda_+ \le c_{\phi,i}\delta$, then $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}$ is a global $\mathcal{C}^\infty$-diffeomorphism of $\mathbb{R}^D$ such that for all $1 \le k \le m$, $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}(\mathcal{B}(x_k,\delta)) = \mathcal{B}(x_k,\delta)$. Moreover,*

$$\left\| I_D - d\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i} \right\|_{op} \le C_{\phi,i} \left\{ \frac{A_+}{\delta^{1-i}} \right\} \vee \left\{ \frac{\Lambda_+}{\delta} \right\},$$

and for $j \ge 2$,

$$\left\| d^j \Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i} \right\|_{op} \le C_{\phi,i,j} \left\{ \frac{A_+}{\delta^{j-i}} \right\} \vee \left\{ \frac{\Lambda_+}{\delta^j} \right\}.$$

Finally, we define the mixture distribution $\bar{Q}_{\tau,n}^{(i)}$ on $(\mathbb{R}^D)^n$ by

$$\bar{Q}_{\tau,n}^{(i)} = \int_{[-\Lambda_+,\Lambda_+]^m} \int_{[A_-,A_+]^m} \left(P_\tau^{\mathbf{\Lambda},\mathbf{A},(i)} \right)^{\otimes n} \frac{d\mathbf{A}}{(A_+ - A_-)^m} \frac{d\mathbf{\Lambda}}{(2\Lambda_+)^m}. \tag{10}$$

Although the probability distribution $\bar{Q}_{\tau,n}^{(i)}$ depends on A_-, A_+ and Λ_+, we omit this dependency for the sake of compactness. Another way to define $\bar{Q}_{\tau,n}^{(i)}$ is the following: draw uniformly $\mathbf{\Lambda}$ in $[-\Lambda_+, \Lambda_+]^m$ and $\mathbf{A}$ in $[A_-, A_+]^m$, and given $(\mathbf{\Lambda}, \mathbf{A})$, take $Z_i = \Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}(Y_i)$, where $Y_1, \dots, Y_n$ is an i.i.d. n-sample with common distribution P_0 on M_0. Then $(Z_1, \dots, Z_n)$ has distribution $\bar{Q}_{\tau,n}^{(i)}$.

We now state useful probabilistic and geometric properties of $\bar{Q}_{\tau,n}^{(i)}$, in view of using Theorem 16. For this, let us denote by $\mathcal{P}_\tau^{(i)}$ the set composed of all the distributions $P_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}$ for $A_- \leq A_1, \dots, A_m \leq A_+$ and $-\Lambda_+ \leq \Lambda_1, \dots, \Lambda_m \leq \Lambda_+$. Again, we omit the dependency on A_-, A_+ and Λ_+.

Lemma 18. *Assume that the conditions of Lemma 17 hold, and let*

$$U_k = \mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(x_k, \delta/2) + \mathcal{B}_{span(e)}(0, \tau_{min}/2),$$

where for $B, B' \subset \mathbb{R}^D$, $B + B'$ denotes their Minkovski sum, and

$$U_k' = \left(\mathbb{R}^D \setminus \left\{\mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(x_k, \delta) + \mathcal{B}_{span(e)}(0, \tau_{min}/2)\right\}\right)^{n-1}.$$

Then the sets $U_k \times U_k'$ are pairwise disjoint, $\bar{Q}_{\tau,n}^{(i)} \in \overline{Conv}\left(\left(\mathcal{P}_\tau^{(i)}\right)^{\otimes n}\right)$, and if $(Z_1, \dots, Z_n) = (Z_1, Z_{2:n})$ has distribution $\bar{Q}_{\tau,n}^{(i)}$, Z_1 and $Z_{2:n}$ are independent conditionally on the event $\{(Z_1, Z_{2:n}) \in U_k \times U_k'\}$.

Moreover, if $(X_1, \dots, X_n)$ has distribution $\left(P_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}\right)^{\otimes n}$ (with fixed $\mathbf{A}$ and $\mathbf{\Lambda}$), then on the event $\{X_1 \in U_k\}$, we have:

- *if $\tau_k = 0$,*

$$T_{X_1} M_\tau^{\mathbf{\Lambda},\mathbf{A},(i)} = \mathbb{R}^d \times \{0\}^{D-d} \quad , \quad \left\| II_{X_1}^{M_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}} \circ \pi_{T_{X_1} M_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}} \right\|_{op} = 0,$$

  and $d_H\left(M_0, M_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}\right) \geq |\Lambda_k|$.

- *if $\tau_k = 1$,*
  - *for $i = 1$: $\angle\left(T_{X_1} M_\tau^{\mathbf{\Lambda},\mathbf{A},(1)}, \mathbb{R}^d \times \{0\}^{D-d}\right) \geq A_-/2$;*
  - *for $i = 2$: $\left\| II_{X_1}^{M_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}} \circ \pi_{T_{X_1} M_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}} \right\|_{op} \geq A_-/2$.*

To apply Theorem 16 to the $\bar{Q}_{\tau,n}^{(i)}$'s with $\mathcal{X} = \mathbb{R}^D$, $\mathcal{X}' = \left(\mathbb{R}^D\right)^{n-1}$, it remains to bound the test affinities between their marginals on $\mathcal{X}$ and $\mathcal{X}'$. By construction (10), these are respectively $\bar{Q}_{\tau,1}^{(i)}$ and $\bar{Q}_{\tau,n-1}^{(i)}$.

Lemma 19. *Assume that the conditions of Lemma 17 and Lemma 18 hold. If in addition, $cA_+(\delta/4)^i \leq \Lambda_+ \leq CA_+(\delta/4)^i$ for some absolute constants $C \geq c > 3/4$, and $A_- = A_+/2$, then,*

$$\int_{U_k} d\bar{Q}_{\tau,1}^{(i)} \wedge d\bar{Q}_{\tau^k,1}^{(i)} \geq \frac{c_{d,i}}{C} \left(\frac{\delta}{\tau_{min}}\right)^d,$$

and

$$\int_{U_k'} d\bar{Q}^{(i)}_{\tau,n-1} \wedge d\bar{Q}^{(i)}_{\tau^k,n-1} = \left(1 - c_d' \left(\frac{\delta}{\tau_{min}}\right)^d\right)^{n-1}.$$

Now, to derive Theorem 7, set $i = 1$, take $A_+ = 2A_- = \varepsilon\delta^{k-1}$, and $\Lambda_+ = \delta A_+/4$ for $\varepsilon = \varepsilon_{\phi,k,d,\tau_{min}}$ small enough so that $\mathcal{P}_\tau^{(1)} \subset \mathcal{P}^k_{\tau_{min},\mathbf{L}}(f_{min}, f_{max})$, according to Lemma 17 and Proposition 4. Hence, applying Lemma 16 together with Lemma 18 and Lemma 19, recalling that m can be taken of order c_d/δ^d, we get, for all estimators $\hat{T}$,

$$\begin{aligned}\sup_{P\in\mathcal{P}^k} \mathbb{E}_{P^{\otimes n}} \angle\big(T_{X_1} M, \hat{T}\big) &\geq c_{d,k}\varepsilon m \frac{A_-}{4}\left(\frac{\delta}{\tau_{min}}\right)^d \left(1 - c_d'\left(\frac{\delta}{\tau_{min}}\right)^d\right)^{n-1} \\ &\geq c'_{d,k,\tau_{min}} \frac{\delta^{k-1}}{\delta^d}\left(\frac{\delta}{\tau_{min}}\right)^d \left(1 - c_d'\left(\frac{\delta}{\tau_{min}}\right)^d\right)^{n-1}.\end{aligned}$$

Taking $(\delta/\tau_{min})^d = 1/(n-1)$ yields the result.

Similarly, to derive Theorem 9, set $i = 2$, take $A_+ = 2A_- = \varepsilon'\delta^{k-2}$, and $\Lambda_+ = \delta^2 A_+/4^2$ with $\varepsilon' = \varepsilon'_{\phi,k,d,\tau_{min}}$ small enough so that $\mathcal{P}_\tau^{(2)} \subset \mathcal{P}^k_{\tau_{min},\mathbf{L}}(f_{min}, f_{max})$. With $(\delta/\tau_{min})^d = 1/(n-1)$, the same derivation as above leads to the result.

Finally, for Theorem 11, simply take $m = 1$, $\tau = 0$ and $\Lambda_1 = \delta A_1/4 = \varepsilon\delta^k$ for $\varepsilon = \varepsilon_{\phi,k,d,\tau_{min}}$ as above. We may conclude using Theorem 15 with P_0 and $P_0^{\Lambda_1,A_1,(i)}$. Indeed, using $d_H\big(M_0, M_0^{\Lambda_1,A_1,(i)}\big) \geq |\Lambda_1| = \varepsilon\delta^k$ from Lemma 18, and noticing that the total variation distance between the two distributions is $P_0(\mathcal{B}(x_1,\delta)) = c_d(\delta/\tau_{min})^d$, since they differ only outside $\mathcal{B}(x_1,\delta)$, we get the result.

Acknowledgements

We would like to thank Frédéric Chazal and Pascal Massart for their constant encouragements, suggestions and stimulating discussions.

A Properties and Stability of the Models

A.1 Property of the Exponential Map in $\mathcal{C}^2_{\tau_{min}}$

Here we show the Lemma 1. Proposition 6.1 in [NSW08] states that for all $x \in M$, $\left\|II^M_x\right\|_{op} \leq 1/\tau_{min}$. In particular, Gauss equation ([dC92, Proposition 3.1 (a), p.135]) yields that the sectional curvatures of M satisfy $-2/\tau^2_{min} \leq \kappa \leq 1/\tau^2_{min}$. Using Corollary 1.4 of [AB06], we get that the injectivity radius of M is at least $\pi\tau_{min} \geq \tau_{min}/4$. Therefore, $\exp_p : \mathcal{B}_{T_pM}(0, \tau_{min}/4) \to M$ is one-to-one.

Let us write $\mathbf{N}_p(v) = \exp_p(v) - p - v$. We clearly have $\mathbf{N}_p(0) = 0$ and $d_0\mathbf{N}_p = 0$. Let now $v \in \mathcal{B}_{T_pM}(0, \tau_{min}/4)$ be fixed. We have $d_v\mathbf{N}_p = d_v \exp_p - Id_{T_pM}$. For $0 \leq t \leq \|v\|$, we write $\gamma(t) = \exp_p(tv/\|v\|)$ for the arc-length parametrized geodesic from p to $\exp_p(v)$, and P_t for the parallel translation along γ. From Lemma 18 of [DVW15],

$$\left\|d_{t\frac{v}{\|v\|}} \exp_p - P_t\right\|_{op} \leq \frac{2}{\tau^2_{min}}\frac{t^2}{2} \leq \frac{t}{4\tau_{min}}.$$

We now derive an upper bound for $\left\|P_t - Id_{T_pM}\right\|_{op}$. For this, fix two unit vectors $u \in \mathbb{R}^D$ and $w \in T_pM$, and write $g(t) = \langle P_t(w) - w, u\rangle$. Letting $\bar{\nabla}$ denote the ambient derivative in $\mathbb{R}^D$, by definition of parallel translation,

$$\begin{aligned}
|g'(t)| &= \left|\langle \bar{\nabla}_{\gamma'(t)} P_t(w) - w, u\rangle\right| \\
&= \left|\langle II^M_{\gamma(t)}\big(\gamma'(t), P_t(w)\big), u\rangle\right| \\
&\leq 1/\tau_{min}.
\end{aligned}$$

Since $g(0) = 0$, we get $\left\|P_t - Id_{T_pM}\right\|_{op} \leq t/\tau_{min}$. Finally, the triangle inequality leads to

$$\begin{aligned}
\|d_v\mathbf{N}_p\|_{op} &= \left\|d_v \exp - Id_{T_pM}\right\|_{op} \\
&\leq \left\|d_v \exp - P_{\|v\|}\right\|_{op} + \left\|P_{\|v\|} - Id_{T_pM}\right\|_{op} \\
&\leq \frac{5\,\|v\|}{4\tau_{min}}.
\end{aligned}$$

We conclude with the property of the projection $\pi^* = \pi_{T_pM}$. Indeed, defining $R_2(y-p) = (y-p) - \pi^*(y-p)$, Lemma 4.7 in [Fed59] gives

$$\begin{aligned}
\|R_2(y-p)\| &= d(y-p, T_pM) \\
&\leq \frac{\|y-p\|^2}{2\tau_{min}}.
\end{aligned}$$

A.2 Geometric Properties of the Models $\mathcal{C}^k$

We now more to the proof of Lemma 12.

Proof of Lemma 12. (i) Simply notice that from the reverse triangle inequality,

$$\left| \frac{\|\Psi_x(v_2) - \Psi_x(v_1)\|}{\|v_2 - v_1\|} - 1 \right| \leq \frac{\|N_x(v_2) - N_x(v_1)\|}{\|v_2 - v_1\|} \leq L_\perp(\|v_1\| \vee \|v_2\|) \leq \frac{1}{4}.$$

(ii) The right-hand side inclusion follows straightforwardly from (i). Let us focus on the left-hand side inclusion. For this, consider the map defined by $G = \pi_{T_xM} \circ \Psi_x$ on the domain $\mathcal{B}_{T_xM}(0,h)$. For all $v \in \mathcal{B}_{T_xM}(0,h)$, we have

$$\|d_vG - Id_{T_xM}\|_{op} = \|\pi_{T_xM} \circ d_v\mathbf{N}_x\|_{op} \leq \|d_v\mathbf{N}_x\|_{op} \leq L_\perp \|v\| \leq \frac{1}{4} < 1.$$

Hence, G is a diffeomorphism onto its image and it satisfies $\|G(v)\| \geq 3\|v\|/4$. It follows that

$$\mathcal{B}_{T_xM}\left(0, \frac{3h}{4}\right) \subset G\left(\mathcal{B}_{T_xM}(0,h)\right) = \pi_{T_xM}\left(\Psi_x\left(\mathcal{B}_{T_xM}(0,h)\right)\right).$$

Now, according to Lemma 1, for all $y \in \mathcal{B}\left(x, \frac{3h}{5}\right) \cap M$,

$$\|\pi_{T_xM}(y-x)\| \leq \|y-x\| + \frac{\|y-x\|^2}{2\tau_{min}} \leq \left(1 + \frac{1}{4}\right)\|y-x\| \leq \frac{3h}{4},$$

from what we deduce $\pi_{T_xM}\left(\mathcal{B}\left(x, \frac{3h}{5}\right) \cap M\right) \subset \mathcal{B}_{T_xM}\left(0, \frac{3h}{4}\right)$. As a consequence,

$$\pi_{T_xM}\left(\mathcal{B}\left(x, \frac{3h}{5}\right) \cap M\right) \subset \pi_{T_xM}\left(\Psi_x\left(\mathcal{B}_{T_xM}(0,h)\right)\right),$$

which yields the announced inclusion since π_{T_xM} is one to one on $\mathcal{B}\left(x, \frac{5h}{4}\right) \cap M$ from Lemma 3 in [ALZ13], and

$$\left(\mathcal{B}\left(x, \frac{3h}{5}\right) \cap M\right) \subset \Psi_x\left(\mathcal{B}_{T_xM}(0,h)\right) \subset \mathcal{B}\left(x, \frac{5h}{4}\right) \cap M.$$

(iii) Straightforward application of Lemma 3 in [ALZ13].

(iv) Notice that Lemma 1 gives the existence of such an expansion for $k = 2$. Hence, we can assume $k \geq 3$. Taking $h = \frac{\tau_{min} \wedge L_\perp^{-1}}{4}$, we showed in the proof of (ii) that the map G is a diffeomorphism onto its image, with $\|d_vG - Id_{T_xM}\|_{op} \leq \frac{1}{4} < 1$. Additionally, the chain rule yields $\left\|d_v^iG\right\|_{op} \leq \left\|d_v^i\Psi_x\right\|_{op} \leq L_i$ for all $2 \leq i \leq k$. Therefore, from Lemma 20, the differentials of G^{-1} up to order k are uniformly bounded. As a consequence, we get the announced expansion writing

$$y - x = \Psi_x \circ G^{-1}\left(\pi^*(y-x)\right),$$

and using the Taylor expansions of order k of Ψ_x and G^{-1}.

Let us now check that $T_2^* = II_x^M$. First, since by construction, T_2^* is the second order term of the Taylor expansion of $\Psi_x \circ G^{-1}$ at zero, a straightforward computation yields

$$\begin{aligned} T_2^* &= (I_D - \pi_{T_xM}) \circ d_0^2 \Psi_x \\ &= \pi_{T_xM^\perp} \circ d_0^2 \Psi_x. \end{aligned}$$

Let $v \in T_xM$ be fixed. Letting $\gamma(t) = \Psi_x(tv)$ for $|t|$ small enough, it is clear that $\gamma''(0) = d_0^2\Psi(v^{\otimes 2})$. Moreover, by definition of the second fundamental form [dC92, Proposition 2.1, p.127], since $\gamma(0) = x$ and $\gamma'(0) = v$, we have

$$II_x^M(v^{\otimes 2}) = \pi_{T_xM^\perp}(\gamma''(0)).$$

Hence

$$\begin{aligned} T_2^*(v^{\otimes 2}) &= \pi_{T_xM^\perp} \circ d_0^2\Psi_x(v^{\otimes 2}) \\ &= \pi_{T_xM^\perp}(\gamma''(0)) \\ &= II_x^M(v^{\otimes 2}), \end{aligned}$$

which concludes the proof.

(v) The first statement is a rephrasing of Proposition 6.1 in [NSW08]. It yields the bound on sectional curvature, using the Gauss equation [dC92, Proposition 3.1 (a), p.135].

□

In the proof of Lemma 12 (iv), we used a technical lemma of differential calculus that we now prove. It states quantitatively that if G is $\mathcal{C}^k$-close to the identity map, then it is a diffeomorphism onto its image and the differentials of its inverse G^{-1} are controlled.

Lemma 20. *Let $k \geq 2$ and U be an open subset of $\mathbb{R}^d$. Let $G : U \to \mathbb{R}^d$ be $\mathcal{C}^k$. Assume that $\|I_d - dG\|_{op} \leq \varepsilon < 1$, and that for all $2 \leq i \leq k$, $\left\|d^iG\right\|_{op} \leq L_i$ for some $L_i > 0$. Then G is a $\mathcal{C}^k$-diffeomorphism onto its image, and for all $2 \leq i \leq k$,*

$$\left\|I_d - dG^{-1}\right\|_{op} \leq \frac{\varepsilon}{1-\varepsilon} \quad \text{and} \quad \left\|d^iG^{-1}\right\|_{op} \leq L'_{i,\varepsilon,L_2,\ldots,L_i} < \infty \text{ for } 2 \leq i \leq k.$$

Proof of Lemma 20. For all $x \in U$, $\|d_xG - I_d\|_{op} < 1$, so G is one to one, and for all $y = G(x) \in G(U)$,

$$\begin{aligned} \left\|I_d - d_yG^{-1}\right\|_{op} &= \left\|I_d - (d_xG)^{-1}\right\|_{op} \\ &\leq \left\|(d_xG)^{-1}\right\|_{op} \|I_d - d_xG\|_{op} \\ &\leq \frac{\|I_d - d_xG\|_{op}}{1 - \|I_d - d_xG\|_{op}} \\ &\leq \frac{\varepsilon}{1-\varepsilon}. \end{aligned}$$

For $2 \leq i \leq k$ and $1 \leq j \leq i$, write $\Pi_i^{(j)}$ for the set of partitions of $\{1, \dots, i\}$ with j blocks. Differentiating i times the identity $G \circ G^{-1} = Id_{G(U)}$, Faa di Bruno's formula yields that, for all $y = G(x) \in G(U)$ and all unit vectors $h_1, \dots, h_i \in \mathbb{R}^D$,

$$0 = d_y \left(G \circ G^{-1}\right).(h_\alpha)_{1 \leq \alpha \leq i} = \sum_{j=1}^{i} \sum_{\pi \in \Pi_i^{(j)}} d_x^j G. \left(\left(d_y^{|I|} G^{-1}. (h_\alpha)_{\alpha \in I} \right)_{I \in \pi} \right).$$

Isolating the term for $j = 1$ entails

$$\begin{aligned}
&\left\| d_x \Phi. \left(d_y^i G^{-1}. (h_\alpha)_{1 \leq \alpha \leq i} \right) \right\|_{op} \\
&= \left\| - \sum_{j=2}^{i} \sum_{\pi \in \Pi_i^{(j)}} d_x^j G. \left(\left(d_y^{|I|} G^{-1}. (h_\alpha)_{\alpha \in I} \right)_{I \in \pi} \right) \right\|_{op} \\
&\leq \sum_{j=2}^{i} \sum_{\pi \in \Pi_i^{(j)}} \left\| d^j G \right\|_{op} \prod_{I \in \pi} \left\| d^{|I|} G^{-1} \right\|_{op}.
\end{aligned}$$

Using the first order Lipschitz bound on G^{-1}, we get

$$\left\| d^i G^{-1} \right\|_{op} \leq \frac{1+\varepsilon}{1-\varepsilon} \sum_{j=2}^{i} L_j \sum_{\pi \in \Pi_i^{(j)}} \prod_{I \in \pi} \left\| d^{|I|} G^{-1} \right\|_{op}.$$

The result follows by induction on i. □

A.3 Stability of the Model

This section is devoted to prove the stability of the model with respect to ambient diffeomorphisms (Proposition 4).

The second part is pretty straightforward since the dilation λM has reach $\tau_{\lambda M} = \lambda \tau_M$, and can be parametrized locally by $\tilde{\Psi}_{\lambda p}(v) = \lambda \Psi_p(v/\lambda) = \lambda p + v + \lambda \mathbf{N}_p(v/\lambda)$, yielding the differential bounds $\mathbf{L}_{(\lambda)}$. Bounds on the density follow from homogeneity of the d-dimensional Hausdorff measure.

For the first part, we split the proof into two intermediate results. Proposition 21 deals with the stability of the geometric model, that is, the reach bound and the existence of a smooth parametrization when a submanifold is perturbed. Lemma 22 deals with the condition on the density in the models $\mathcal{P}^k$. It gives a change of variable formula for pushforward of measure on submanifolds, ensuring a control on densities with respect to intrinsic volume measure.

Proposition 21. *Let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ be a global $\mathcal{C}^k$-diffeomorphism. If $\|d\Phi - I_D\|_{op}$, $\left\|d^2\Phi\right\|_{op}$, …, $\left\|d^k\Phi\right\|_{op}$ are small enough, then for all M in $\mathcal{C}^k_{\tau_{min},\mathbf{L}}$, the image $M' = \Phi(M)$ belongs to $\mathcal{C}^k_{\tau_{min}/2, 2L_\perp, 2L_3, \dots, 2L_k}$.*

Proof of Proposition 21. To bound $\tau_{M'}$ from below, we use the stability of the reach with respect to $\mathcal{C}^2$ diffeomorphisms. Namely, from Theorem 4.19 in [Fed59],

$$\begin{aligned}\tau_{M'} = \tau_{\Phi(M)} &\geq \frac{(1 - \|I_D - d\Phi\|_{op})^2}{\frac{1+\|I_D - d\Phi\|_{op}}{\tau_M} + \|d^2\Phi\|_{op}} \\ &\geq \tau_{min} \frac{(1 - \|I_D - d\Phi\|_{op})^2}{1 + \|I_D - d\Phi\|_{op} + \tau_{min} \|d^2\Phi\|_{op}} \geq \frac{\tau_{min}}{2}\end{aligned}$$

for $\|I_D - d\Phi\|_{op}$ and $\left\|d^2\Phi\right\|_{op}$ small enough. This shows the stability for $k = 2$, as well as that of the reach assumption for $k \geq 3$.

By now, take $k \geq 3$. We focus on the existence of a good parametrization of M' around a fixed point $p' = \Phi(p) \in M'$. For $v' \in T_{p'}M' = d_p\Phi\left(T_pM\right)$, let us define

$$\begin{aligned}\Psi'_{p'}(v') &= \Phi\left(\Psi_p\left(d_{p'}\Phi^{-1}.v'\right)\right) \\ &= p' + v' + \mathbf{N}'_{p'}(v'),\end{aligned}$$

where $\mathbf{N}'_{p'}(v') = \left\{\Phi\left(\Psi_p\left(d_{p'}\Phi^{-1}.v'\right)\right) - p' - v'\right\}$.

$$\begin{array}{ccc} M & \xrightarrow{\quad\Phi\quad} & M' \\ {\scriptstyle \Psi_p}\uparrow & & \uparrow{\scriptstyle \Psi'_{p'}} \\ T_pM & \xrightarrow[d_p\Phi]{} & T_{p'}M' \end{array}$$

The maps $\Psi'_{p'}(v')$ and $\mathbf{N}'_{p'}(v')$ are well defined whenever $\left\|d_{p'}\Phi^{-1}.v'\right\| \leq \frac{1}{8L_\perp}$, so in particular if $\|v'\| \leq \frac{1}{8(2L_\perp)} \leq \frac{1-\|I_D - d\Phi\|_{op}}{8L_\perp}$ and $\|I_D - d\Phi\|_{op} \leq \frac{1}{2}$. One easily checks that $\mathbf{N}'_{p'}(0) = 0$, $d_0\mathbf{N}'_{p'} = 0$ and writing $c(v') = p + d_{p'}\Phi^{-1}.v' + \mathbf{N}_{p'}\left(d_{p'}\Phi^{-1}.v'\right)$, for all unit vector $w' \in T_{p'}M'$,

$$
\begin{aligned}
\left\|d^2_{v'}\mathbf{N}'_{p'}(w'^{\otimes 2})\right\| &= \left\|d^2_{c(v')}\Phi\left(\left\{d_{d_{p'}\Phi^{-1}.v'}\Psi_p \circ d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right)\right. \\
&\quad \left. + d_{c(v')}\Phi \circ d^2_{d_{p'}\Phi^{-1}.v'}\Psi_p\left(\left\{d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right)\right\| \\
&= \left\|d^2_{c(v')}\Phi\left(\left\{d_{d_{p'}\Phi^{-1}.v'}\Psi_p \circ d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right)\right. \\
&\quad + \left(d_{c(v')}\Phi - Id\right) \circ d^2_{d_{p'}\Phi^{-1}.v'}\Psi_p\left(\left\{d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right) \\
&\quad \left. + d^2_{d_{p'}\Phi^{-1}.v'}\Psi_p\left(\left\{d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right)\right\| \\
&\leq \left\|d^2\Phi\right\|_{op}\left(1 + L_\perp\left\|d_{p'}\Phi^{-1}.v'\right\|\right)^2\left\|d_{p'}\Phi^{-1}.w'\right\|^2 \\
&\quad + \left\|I_D - d\Phi\right\|_{op} L_\perp\left\|d_{p'}\Phi^{-1}.w'\right\|^2 \\
&\quad + L_\perp\left\|d_{p'}\Phi^{-1}.w'\right\|^2 \\
&\leq \left\|d^2\Phi\right\|_{op}(1+1/8)^2\left\|d_{p'}\Phi^{-1}\right\|^2_{op} \\
&\quad + \left\|I_D - d\Phi\right\|_{op} L_\perp\left\|d\Phi^{-1}\right\|^2_{op} \\
&\quad + L_\perp\left\|d_{p'}\Phi^{-1}\right\|^2_{op}.
\end{aligned}
$$

Writing further $\left\|d\Phi^{-1}\right\|_{op} \leq (1 - \left\|I_D - d\Phi\right\|_{op})^{-1} \leq 1 + 2\left\|I_D - \Phi\right\|_{op}$ for $\left\|I_D - d\Phi\right\|_{op}$ small enough depending only on $L_\perp$, it is clear that the right-hand side of the latter inequality goes below $2L_\perp$ for $\left\|I_D - d\Phi\right\|_{op}$ and $\left\|d^2\Phi\right\|_{op}$ small enough. Hence, for $\left\|I_D - d\Phi\right\|_{op}$ and $\left\|d^2\Phi\right\|_{op}$ small enough depending only on $L_\perp$, $\|d^2_{v'}\mathbf{N}'_{p'}\|_{op} \leq 2L_\perp$ for all $\|v'\| \leq \frac{1}{8(2L_\perp)}$. From the chain rule, the same argument applies for the order $3 \leq i \leq k$ differential of $\mathbf{N}'_{p'}$. □

Lemma 22 (Change of variable for the Hausdorff measure)**.** *Let P be a probability distribution on $M \subset \mathbb{R}^D$ with density f with respect to the d-dimensional Hausdorff measure $\mathcal{H}^d$. Let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ be a global diffeomorphism such that $\left\|I_D - d\Phi\right\|_{\mathrm{op}} < 1/3$. Let $P' = \Phi_* P$ be the pushforward of P by Φ. Then P' has a density g with respect to $\mathcal{H}^d$. This density can be chosen to be, for all $z \in \Phi(M)$,*

$$
g(z) = \frac{f\left(\Phi^{-1}(z)\right)}{\sqrt{\det\left(\pi_{T_{\Phi^{-1}(z)}M} \circ d_{\Phi^{-1}(z)}\Phi^T \circ d_{\Phi^{-1}(z)}\Phi_{\left|T_{\Phi^{-1}(z)}M\right.}\right)}}.
$$

In particular, if $f_{min} \leq f \leq f_{max}$ on M, then for all $z \in \Phi(M)$,

$$
\left(1 - 3d/2\left\|I_D - d\Phi\right\|_{\mathrm{op}}\right) f_{min} \leq g(z) \leq f_{max}\left(1 + 3(2^{d/2} - 1)\left\|I_D - d\Phi\right\|_{\mathrm{op}}\right).
$$

Proof of Lemma 22. Let $p \in M$ be fixed and $A \subset \mathcal{B}(p, r) \cap M$ for r small enough. For a differentiable map $h : \mathbb{R}^d \to \mathbb{R}^D$ and for all $x \in \mathbb{R}^d$, we let $J_h(x)$ denote the

d-dimensional Jacobian $J_h(x) = \sqrt{\det\left(d_x h^T \, d_x h\right)}$. The area formula ([Fed69, Theorem 3.2.5]) states that if h is one-to-one,

$$\int_A u\left(h(x)\right) J_h(x) \lambda^d(dx) = \int_{h(A)} u(y) \mathcal{H}^d(dy),$$

whenever $u : \mathbb{R}^D \to \mathbb{R}$ is Borel, where λ^d is the Lebesgue measure on $\mathbb{R}^d$. By definition of the pushforward, and since $dP = f d\mathcal{H}^d$,

$$\int_{\Phi(A)} dP'(z) = \int_A f(y) \mathcal{H}^d(dy).$$

Writing $\Psi_p = \exp_p : T_pM \to \mathbb{R}^D$ for the exponential map of M at p, we have

$$\int_A f(y) \mathcal{H}^d(dy) = \int_{{\Psi_p}^{-1}(A)} f(\Psi_p(x)) J_{\Psi_p}(x) \lambda^d(dx).$$

Rewriting the right hand term, we apply the area formula again with $h = \Phi \circ \Psi_p$,

$$\begin{aligned}
&\int_{{\Psi_p}^{-1}(A)} f(\Psi_p(x)) J_{\Psi_p}(x) \lambda^d(dy) \\
&\qquad = \int_{{\Psi_p}^{-1}(A)} f\left(\Phi^{-1}\left(h(x)\right)\right) \frac{J_{\Psi_p}(h^{-1}\left(h(x)\right))}{J_{\Phi \circ \Psi_p}(h^{-1}\left(h(x)\right))} J_{\Phi \circ \Psi_p}(x) \lambda^d(dx) \\
&\qquad = \int_{\Phi(A)} f\left(\Phi^{-1}\left(z\right)\right) \frac{J_{\Psi_p}(h^{-1}\left(z\right))}{J_{\Phi \circ \Psi_p}(h^{-1}\left(z\right))} \mathcal{H}^d(dz).
\end{aligned}$$

Since this is true for all $A \subset \mathcal{B}(p, r) \cap M$, P' has a density g with respect to $\mathcal{H}^d$, with

$$g(z) = f\left(\Phi^{-1}\left(z\right)\right) \frac{J_{\Psi_{\Phi^{-1}(z)}}(\Psi^{-1}_{\Phi^{-1}(z)} \circ \Phi^{-1}\left(z\right))}{J_{\Phi \circ \Psi_{\Phi^{-1}(z)}}(\Psi^{-1}_{\Phi^{-1}(z)} \circ \Phi^{-1}\left(z\right))}.$$

Writing $p = \Phi^{-1}(z)$, it is clear that $\Psi^{-1}_{\Phi^{-1}(z)} \circ \Phi^{-1}\left(z\right) = \Psi_p^{-1}(p) = 0 \in T_pM$. Since $d_0 \exp_p : T_pM \to \mathbb{R}^D$ is the inclusion map, we get the first statement.

We now let B and π_T denote $d_p\Phi$ and π_{T_pM} respectively. For any unit vector $v \in T_pM$,

$$\begin{aligned}
\left|\left\|\pi_T B^T B v\right\| - \|v\|\right| &\leq \left\|\pi_T \left(B^T B - I_D\right) v\right\| \\
&\leq \left\|B^T B - I_D\right\|_{\text{op}} \\
&\leq \left(2 + \|I_D - B\|_{\text{op}}\right) \|I_D - B\|_{\text{op}} \\
&\leq 3 \|I_D - B\|_{\text{op}}.
\end{aligned}$$

Therefore, $1 - 3\|I_D - B\|_{\text{op}} \leq \left\|\pi_T B^T B_{|T_pM}\right\|_{\text{op}} \leq 1 + 3\|I_D - B\|_{\text{op}}$. Hence,

$$\sqrt{\det\left(\pi_T B^T B_{|T_pM}\right)} \leq \left(1 + 3\|I_D - B\|_{\text{op}}\right)^{d/2} \leq \frac{1}{1 - \frac{3d}{2}\|I_D - B\|_{\text{op}}},$$

and

$$\sqrt{\det\left(\pi_T B^T B_{|T_pM}\right)} \geq \left(1 - 3\left\|I_D - B\right\|_{\mathrm{op}}\right)^{d/2} \geq \frac{1}{1 + 3(2^{d/2} - 1)\left\|I_D - B\right\|_{\mathrm{op}}},$$

which yields the result. □

B Some Probabilistic Tools

B.1 Volume and Covering Rate

The first lemma of this section gives some details about the covering rate of a manifold with bounded reach.

Lemma 23. *Let $P \in \mathcal{P}^k$ have support $M \subset \mathbb{R}^D$. Then for all $r \leq \tau_{min}/4$ and x in M,*

$$c_d f_{min} r^d \leq p_x(r) \leq C_d f_{max} r^d,$$

for some $c_d, C_d > 0$, with $p_x(r) = P\big(\mathcal{B}(x,r)\big)$.

Moreover, letting $h = \left(\frac{C'_d k}{f_{min}} \frac{\log n}{n}\right)^{1/d}$ with C'_d large enough, the following holds. For n large enough so that $h \leq \tau_{min}/8$, with probability at least $1 - \left(\frac{1}{n}\right)^{k/d}$,

$$d_H\left(M, \mathbb{X}_n\right) \leq h.$$

Lemma 23 includes Lemma 14, hence we only prove Lemma 23.

Proof of Lemma 23. Denoting by $\mathcal{B}_M(x,r)$ the geodesic ball of radius r centered at x, Proposition 25 of [AL15] yields

$$\mathcal{B}_M(x,r) \subset \mathcal{B}(x,r) \cap M \subset \mathcal{B}_M(x, 6r/5).$$

Hence, the bounds on the Jacobian of the exponential map given by Proposition 27 of [AL15] yield

$$c_d r^d \leq Vol\big(\mathcal{B}(x,r) \cap M\big) \leq C_d r^d,$$

for some $c_d, C_d > 0$. Now, since P has a density $f_{min} \leq f \leq f_{max}$ with respect to the volume measure of M, we get the first result.

Now we notice that since $p_x(r) \geq c_d f_{min} r^d$, Theorem 3.3 in [CGLM13] entails, for $h \leq \tau_{min}/8$,

$$\mathbb{P}\left(d_H\big(M, \mathbb{X}_n\big) \geq h\right) \leq \frac{4^d}{c_d f_{min} h^d} \exp\left(-\frac{c_d f_{min}}{2^d} n h^d\right).$$

Hence, taking $h = \left(\frac{C'_d k}{f_{min}} \frac{\log n}{n}\right)^{1/d}$ with C'_d so that $C'_d \geq \frac{4^d}{c_d k} \vee \frac{(1+k/d)}{c_d k}$ yields the result. Since $k \geq 1$, taking $C'_d = \frac{4^d}{c_d}$ is sufficient. □

B.2 Concentration Bounds for Local Polynomials

This section is devoted to the proof of Proposition 13. A first step is to ensure that empirical expectations order k polynomials are close to their deterministic counterparts.

Proposition 24. *For any $x \in M$, we have*

$$\mathbb{P}\left[\sup_{u_1,\dots,u_k,\varepsilon\in\{0,1\}^k}\left|(P-P_{n-1})\prod_{j=1}^{p}\left(\frac{\langle u_j,y\rangle}{h}\right)^{\varepsilon_j}\mathbb{1}_{\mathcal{B}(x,h)}(y)\right|\right.$$
$$\left.\geq p_x(h)\left(\frac{4k\sqrt{2\pi}}{\sqrt{(n-1)p_x(h)}}+\sqrt{\frac{2t}{(n-1)p_x(h)}}+\frac{2}{3(n-1)p_x(h)}\right)\right]\leq e^{-t},$$

where P_{n-1} denotes the empirical distribution of $n-1$ i.i.d. random variables X_i drawn from P.

Proof of Proposition 24. Without loss of generality we choose $x=0$ and shorten notation to $\mathcal{B}(h)$ and $p(h)$. Let Z denote the empirical process on the left-hand side of Proposition 24. Denote also by $f_{u,\varepsilon}$ the map $\prod_{j=1}^{k}\left(\frac{\langle u_j,y\rangle}{h}\right)^{\varepsilon_j}\mathbb{1}_{\mathcal{B}(h)}(y)$, and let $\mathcal{F}$ denote the set of such maps, for u_j in $\mathcal{B}(1)$ and ε in $\{0,1\}^k$.

Since $\|f_{u,\varepsilon}\|_\infty \leq 1$ and $Pf_{u,\varepsilon}^2 \leq p(h)$, the Talagrand-Bousquet inequality ([Bou02, Theorem 2.3]) yields

$$Z \leq 4\mathbb{E}Z + \sqrt{\frac{2p(h)t}{n-1}} + \frac{2t}{3(n-1)},$$

with probability larger than $1-e^{-t}$. It remains to bound $\mathbb{E}Z$ from above.

Lemma 25. *We may write*

$$\mathbb{E}Z \leq \frac{\sqrt{2\pi p(h)}}{\sqrt{n-1}}k.$$

Proof of Lemma 25. Let σ_i and g_i denote some independent Rademacher and Gaussian variables. For convenience, we denote by $\mathbb{E}_A$ the expectation with respect to the random variable A. Using symmetrization inequalities we may write

$$\begin{aligned}\mathbb{E}Z &= \mathbb{E}_X \sup_{u,\varepsilon}\left|(P-P_{n-1})\prod_{j=1}^{k}\left(\frac{\langle u_j,y\rangle}{h}\right)^{\varepsilon_j}\mathbb{1}_{\mathcal{B}(h)}(y)\right| \\ &\leq \frac{2}{n-1}\mathbb{E}_X\mathbb{E}_\sigma \sup_{u,\varepsilon}\sum_{i=1}^{n-1}\sigma_i\prod_{j=1}^{k}\left(\frac{\langle u_j,X_i\rangle}{h}\right)^{\varepsilon_j}\mathbb{1}_{\mathcal{B}(h)}(X_i) \\ &\leq \frac{\sqrt{2\pi}}{n-1}\mathbb{E}_X\mathbb{E}_g \sup_{u,\varepsilon}\sum_{i=1}^{n-1}g_i\prod_{j=1}^{k}\left(\frac{\langle u_j,X_i\rangle}{h}\right)^{\varepsilon_j}\mathbb{1}_{\mathcal{B}(h)}(X_i).\end{aligned}$$

Now let Y_g denote the Gaussian process $\sum_{i=1}^{n-1} g_i \prod_{j=1}^{k} \left(\frac{\langle u_j, X_i\rangle}{h}\right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(h)}(X_i)$. Since, for any x in $\mathcal{B}(h)$, u,v in $\mathcal{B}(1)^k$, and ε, ε' in $\{0,1\}^k$, we have

$$
\begin{aligned}
&\left|\prod_{j=1}^{k}\left(\frac{\langle x,u_j\rangle}{h}\right)^{\varepsilon_j} - \prod_{j=1}^{k}\left(\frac{\langle x,v_j\rangle}{h}\right)^{\varepsilon'_j}\right| \\
&\quad\le \left|\sum_{r=1}^{k}\left(\prod_{j=1}^{k+1-r}\left(\frac{\langle x,u_j\rangle}{h}\right)^{\varepsilon_j}\prod_{j=k+2-r}^{k}\left(\frac{\langle x,v_j\rangle}{h}\right)^{\varepsilon'_j}\right.\right. \\
&\qquad\qquad \left.\left. - \prod_{j=1}^{k-r}\left(\frac{\langle x,u_j\rangle}{h}\right)^{\varepsilon_j}\prod_{j=k+1-r}^{k}\left(\frac{\langle x,v_j\rangle}{h}\right)^{\varepsilon'_j}\right)\right| \\
&\quad\le \sum_{r=1}^{k}\left|\prod_{j=1}^{k-r}\left(\frac{\langle x,u_j\rangle}{h}\right)^{\varepsilon_j}\prod_{j=k+2-r}^{k}\left(\frac{\langle x,v_j\rangle}{h}\right)^{\varepsilon'_j}\left[\left(\frac{\langle u_{k+1-r},x\rangle}{h}\right)^{\varepsilon_{k+1-r}}\right.\right. \\
&\qquad\qquad \left.\left. - \left(\frac{\langle v_{k+1-r},x\rangle}{h}\right)^{\varepsilon'_{k+1-r}}\right]\right| \\
&\quad\le \sum_{r=1}^{k}\left|\frac{\langle \varepsilon_r u_r - \varepsilon'_r v_r, x\rangle}{h}\right|.
\end{aligned}
$$

We deduce that

$$
\begin{aligned}
\mathbb{E}_g(Y_{u,\varepsilon} - Y_{v,\varepsilon'})^2 &\le k\sum_{i=1}^{n-1}\sum_{r=1}^{k}\left(\frac{\langle \varepsilon_r u_r, X_i\rangle}{h} - \frac{\langle \varepsilon'_r v_r, X_i\rangle}{h}\right)^2 \mathbb{1}_{\mathcal{B}(h)}(X_i) \\
&\le \mathbb{E}_g(\Theta_{u,\varepsilon} - \Theta_{v,\varepsilon'})^2,
\end{aligned}
$$

where $\Theta_{u,\varepsilon} = \sqrt{k}\sum_{i=1}^{n-1}\sum_{r=1}^{k} g_{i,r}\frac{\langle \varepsilon_r u_r, X_i\rangle}{h}\mathbb{1}_{\mathcal{B}(h)}(X_i)$. According to Slepian's Lemma [BLM13, Theorem 13.3], it follows that

$$
\begin{aligned}
\mathbb{E}_g \sup_{u,\varepsilon} Y_g &\le \mathbb{E}_g \sup_{u,\varepsilon} \Theta_g \\
&\le \sqrt{k}\mathbb{E}_g \sup_{u,\varepsilon}\sum_{r=1}^{k}\frac{\left\langle \varepsilon_r u_r, \sum_{i=1}^{n-1} g_{i,r}\mathbb{1}_{\mathcal{B}(h)}(X_i)X_i\right\rangle}{h} \\
&\le \sqrt{k}\mathbb{E}_g \sup_{u,\varepsilon}\sqrt{k\sum_{r=1}^{k}\frac{\left\langle \varepsilon_r u_r, \sum_{i=1}^{n-1} g_{i,r}\mathbb{1}_{\mathcal{B}(h)}(X_i)X_i\right\rangle^2}{h^2}}.
\end{aligned}
$$

We deduce that

$$\begin{aligned}\mathbb{E}_g \sup_{u,\varepsilon} Y_g &\leq \mathbb{E}_g \sup_{u,\varepsilon} \Theta_g \\ &\leq k\sqrt{\mathbb{E}_g \sup_{\|u\|=1,\varepsilon\in\{0,1\}} \frac{\left\langle \varepsilon u, \sum_{i=1}^{n-1} g_i \mathbb{1}_{\mathcal{B}(h)}(X_i) X_i \right\rangle^2}{h^2}} \\ &\leq k\sqrt{\mathbb{E}_g \left\| \sum_{i=1}^{n-1} \frac{g_i X_i}{h} \mathbb{1}_{\mathcal{B}(h)}(X_i) \right\|^2} \\ &\leq k\sqrt{N(h)}.\end{aligned}$$

Then we can deduce that $\mathbb{E}_X \mathbb{E}_g \sup_{u,\varepsilon} Y_g \leq k\sqrt{p(h)}$, hence the result. □

Combining Lemma 25 with Talagrand-Bousquet's inequality gives the result of Proposition 24. □

We are now in position to prove Proposition 13.

Proof of Proposition 13. If $h \leq \tau_{min}/8$, then, according to Lemma 23, $p(h) \geq c_d f_{min} h^d$, hence, if $h = \left(K\frac{\log(n)}{n-1}\right)^{\frac{1}{d}}$, $(n-1)p(h) \geq K c_d f_{min} \log(n)$. Choosing $t = (k/d+1)\log(n)$ in Proposition 24 and $K = K'/f_{min}$, with $K' > 1$ leads to

$$\mathbb{P}\left[\sup_{u_1,\ldots,u_k,\varepsilon\in\{0,1\}^k} \left| (P - P_{n-1}) \prod_{j=1}^{k} \left(\frac{\langle u_j, y\rangle}{h}\right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(x,h)}(y) \right| \geq \frac{c_{d,k} f_{max}}{\sqrt{K'}} h^d \right] \leq \left(\frac{1}{n}\right)^{\frac{k}{d}+1}.$$

On the complement of the probability event mentioned just above, for a polynomial $Q = \sum_{\alpha\in[0,k]^d || |\alpha|\leq k} a_\alpha x_{1:d}^\alpha$, we have

$$\begin{aligned}(P_{n-1} - P)Q^2(x_{1:d})\mathbb{1}_{\mathcal{B}(h)}(x) &\geq -\sum_{\alpha,\beta} \frac{c_{d,k} f_{max}}{\sqrt{K'}} |a_\alpha a_\beta| h^{d+|\alpha|+|\beta|} \\ &\geq -\frac{c_{d,k} f_{max}}{\sqrt{K'}} h^d \|Q_h\|_2^2.\end{aligned}$$

On the other hand, we may write, for all $r > 0$,

$$\int_{\mathcal{B}(0,r)} Q^2(x_{1:d}) dx_1 \ldots dx_d \geq C_{d,k} r^d \|Q_r\|_2^2,$$

for some constant $C_{d,k}$. It follows that

$$PQ^2(x_{1:d})\mathbb{1}_{\mathcal{B}(h)}(x) \geq PQ^2(x_{1:d})\mathbb{1}_{B(7h/8)}(x_{1:d}) \geq c_{k,d} h^d f_{min} \|Q_h\|_2^2,$$

according to Lemma 12. Then we may choose $K' = \kappa_{k,d}(f_{max}/f_{min})^2$, with $\kappa_{k,d}$ large enough so that

$$P_{n-1}Q^2(x_{1:d})\mathbb{1}_{\mathcal{B}(h)}(x) \geq c_{k,d}f_{min}h^d\|Q_h\|_2^2.$$

□

C Minimax Lower Bounds

C.1 Proof of the Conditional Assouad's Lemma

This section is dedicated to the proof of Lemma 16. The proof follows that of Lemma 2 in [Yu97]. Let $\hat{\theta} = \hat{\theta}(X, X')$ be fixed. For any family of 2^m distributions $\{Q_\tau\}_\tau \in \{\mathcal{Q}_\tau\}_\tau$, since the $U_k \times U_k'$'s are pairwise disjoint,

$$\begin{aligned}
&\sup_{Q\in\mathcal{Q}} \mathbb{E}_Q\left[d\big(\theta_X(Q), \hat{\theta}(X, X')\big)\right] \\
&\qquad \geq \max_\tau \mathbb{E}_{Q_\tau} d(\hat{\theta}, \theta_X(Q_\tau)) \\
&\qquad \geq \max_\tau \mathbb{E}_{Q_\tau} \sum_{k=1}^m d\big(\hat{\theta}, \theta_X(Q_\tau)\big)\mathbb{1}_{U_k\times U_k'}(X, X') \\
&\qquad \geq 2^{-m}\sum_\tau \sum_{k=1}^m \mathbb{E}_{Q_\tau} d\big(\hat{\theta}, \theta_X(Q_\tau)\big)\mathbb{1}_{U_k\times U_k'}(X, X') \\
&\qquad \geq 2^{-m}\sum_\tau \sum_{k=1}^m \mathbb{E}_{Q_\tau} d\big(\hat{\theta}, \mathcal{D}_{\tau,k}\big)\mathbb{1}_{U_k\times U_k'}(X, X') \\
&\qquad = \sum_{k=1}^m 2^{-(m+1)} \sum_\tau \Big(\mathbb{E}_{Q_\tau} d\big(\hat{\theta}, \mathcal{D}_{\tau,k}\big)\mathbb{1}_{U_k\times U_k'}(X, X') \\
&\qquad\qquad\qquad + \mathbb{E}_{Q_{\tau^k}} d\big(\hat{\theta}, \mathcal{D}_{\tau^k,k}\big)\mathbb{1}_{U_k\times U_k'}(X, X')\Big).
\end{aligned}$$

Since the previous inequality holds for all $Q_\tau \in \mathcal{Q}_\tau$, it extends to $\overline{Q}_\tau \in \overline{Conv}(\mathcal{Q}_\tau)$ by linearity. Let us now lower bound each of the terms of the sum for fixed $\tau \in \{0,1\}^m$ and $1 \leq k \leq m$. By assumption, if (X, X') has distribution $\overline{Q}_\tau$, then conditionally on $\{(X, X') \in U_k \times U_k'\}$, X and X' are independent. There-

fore,

$$\begin{aligned}
&\mathbb{E}_{\overline{Q}_\tau} d\big(\hat{\theta}, \mathcal{D}_{\tau,k}\big)\mathbb{1}_{U_k\times U_k'}(X,X') + \mathbb{E}_{\overline{Q}_{\tau^k}} d\big(\hat{\theta}, \mathcal{D}_{\tau^k,k}\big)\mathbb{1}_{U_k\times U_k'}(X,X') \\
&\quad \geq \mathbb{E}_{\overline{Q}_\tau} d\big(\hat{\theta}, \mathcal{D}_{\tau,k}\big)\mathbb{1}_{U_k}(X)\mathbb{1}_{U_k'}(X') + \mathbb{E}_{\overline{Q}_{\tau^k}} d\big(\hat{\theta}, \mathcal{D}_{\tau^k,k}\big)\mathbb{1}_{U_k}(X)\mathbb{1}_{U_k'}(X') \\
&\quad = \mathbb{E}_{\bar{\nu}_\tau}\left[\mathbb{E}_{\bar{\mu}_\tau}\left(d\big(\hat{\theta}, \mathcal{D}_{\tau,k}\big)\mathbb{1}_{U_k}(X)\right)\mathbb{1}_{U_k'}(X')\right] \\
&\qquad\quad + \mathbb{E}_{\bar{\nu}_{\tau^k}}\left[\mathbb{E}_{\bar{\mu}_{\tau^k}}\left(d\big(\hat{\theta}, \mathcal{D}_{\tau^k,k}\big)\mathbb{1}_{U_k}(X)\right)\mathbb{1}_{U_k'}(X')\right] \\
&\quad = \int_{U_k}\int_{U_k'} d(\hat{\theta}, \mathcal{D}_{\tau,k}) d\bar{\mu}_\tau(x) d\bar{\nu}_\tau(x') + \int_{U_k}\int_{U_k'} d(\hat{\theta}, \mathcal{D}_{\tau^k,k}) d\bar{\mu}_{\tau^k}(x) d\bar{\nu}_{\tau^k}(x') \\
&\quad \geq \int_{U_k}\int_{U_k'} \Big(d(\hat{\theta}, \mathcal{D}_{\tau,k}) + d(\hat{\theta}, \mathcal{D}_{\tau^k,k})\Big) d\bar{\mu}_\tau \wedge d\bar{\mu}_{\tau^k}(x) d\bar{\nu}_\tau \wedge d\bar{\nu}_{\tau^k}(x') \\
&\quad \geq \Delta\left(\int_{U_k} d\bar{\mu}_\tau \wedge d\bar{\mu}_{\tau^k}\right)\left(\int_{U_k'} d\bar{\nu}_\tau \wedge d\bar{\nu}_{\tau^k}\right) \\
&\quad \geq \Delta(1-\alpha),
\end{aligned}$$

where we used that $d(\hat{\theta}, \mathcal{D}_{\tau,k}) + d(\hat{\theta}, \mathcal{D}_{\tau^k,k}) \geq \Delta$. The result follows by summing the bound above $|\{1,\dots,m\}\times\{0,1\}^m| = m2^m$ times.

C.2 Construction of Generic Hypotheses

In this section we prove Lemma 18 and Lemma 19.

Proof of Lemma 18. It is clear from the definition (10) that $\bar{Q}^{(i)}_{\tau,n} \in \overline{Conv}\big(\big(\mathcal{P}^{(i)}_\tau\big)^{\otimes n}\big)$. By construction of the $\Phi^{\mathbf{\Lambda},\mathbf{A},i}_\tau$'s, these maps leave the sets

$$\mathcal{B}_{\mathbb{R}^d\times\{0\}^{D-d}}(x_k,\delta) + \mathcal{B}_{span(e)}(0,\tau_{min}/2)$$

unchanged for all $\mathbf{\Lambda},\mathbf{L}$. Therefore, on the event $\{(Z_1,Z_{2:n})\in U_k\times U_k'\}$, one can write Z_1 only as a function of X_1,Λ_k,A_k, and $Z_{2:n}$ as a function of the rest of the X_j's,Λ_k's and A_k's. Therefore, Z_1 and $Z_{2:n}$ are independent.

We now focus on the geometric statements. For this, we fix a deterministic point $z = \Phi^{\mathbf{\Lambda},\mathbf{A},(i)}_\tau(x_0) \in U_k \cap M^{\mathbf{\Lambda},\mathbf{A},(i)}_\tau$. By construction, one necessarily has $x_0 \in M_0 \cap \mathcal{B}(x_k,\delta/2)$.

- If $\tau_k = 0$, locally around x_0, $\Phi^{\mathbf{\Lambda},\mathbf{A},(1)}_\tau$ is the translation of vector $\Lambda_k e$. Therefore, since M_0 satisfies $T_{x_0}M_0 = \mathbb{R}^d\times\{0\}^{D-d}$ and $II^{M_0}_{x_0} = 0$, we have

  $$T_z M^{\mathbf{\Lambda},\mathbf{A},(i)}_\tau = \mathbb{R}^d\times\{0\}^{D-d} \quad \text{and} \quad \left\| II^{M^{\mathbf{\Lambda},\mathbf{A},(i)}_\tau}_z \circ \pi_{T_z M^{\mathbf{\Lambda},\mathbf{A},(i)}_\tau}\right\|_{op} = 0.$$

  Furthermore, by construction, $z_k = x_k + \Lambda_k e$ belongs to $M^{\mathbf{\Lambda},\mathbf{A},(i)}_\tau$. Since e is orthogonal to M_0, $d(z_0,M_0) \geq |\Lambda_k|$. Thus

  $$d_H\big(M_0, M^{\mathbf{\Lambda},\mathbf{A},(i)}_\tau\big) \geq |\Lambda_k|.$$

- if $\tau_k = 1$,
  - for $i = 1$: locally around x_0, $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},(1)}$ can be written as $x \mapsto x + A_k(x - x_k)_1 e$. Hence, $T_z M_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}$ contains the direction $(1, A_k)$ in the plane $span(e_1, e)$ spanned by the first vector of the canonical basis and e. As a consequence, since e is orthogonal to $\mathbb{R}^d \times \{0\}^{D-d}$,

$$\angle\left(T_z M_\tau^{\mathbf{\Lambda},\mathbf{A},(1)}, \mathbb{R}^d \times \{0\}^{D-d}\right) \geq \left(1 + 1/A_k^2\right)^{-1/2} \geq A_k/2 \geq A_-/2.$$

  - for $i = 2$: locally around x_0, $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}$ can be written as $x \mapsto x + A_k(x - x_k)_1^2 e$. Hence, $M_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}$ contains an arc of parabola of equation $y = A_k(x - x_k)_1^2$ in the plane $span(e_1, e)$. As a consequence,

$$\left\| II_z^{M_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}} \circ \pi_{T_z M_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}} \right\|_{op} \geq A_k/2 \geq A_-/2.$$

□

Proof of Lemma 19. First note that all the distributions involved have support in $\mathbb{R}^d \times span(e) \times \{0\}^{D-(d+1)}$. Therefore, we use the canonical coordinate system of $\mathbb{R}^d \times span(e)$, centered at x_k, and we denote the components by $(x_1, x_2, \ldots, x_d, y) = (x_1, x_{2:d}, y)$. Without loss of generality, assume that $\tau_k = 0$ (if not, flip τ and τ^k). Recall that ϕ has been chosen to be constant and equal to 1 on the ball $\mathcal{B}(0, 1/2)$.

By definition (10), on the event $\{Z \in U_k\}$, a random variable Z having distribution $\bar{Q}_{\tau,1}^{(i)}$ can be represented as $Z = X + \phi\left(\frac{X - x_k}{\delta}\right)\Lambda_k e = X + \Lambda_k e$ where X and Λ_k are independent and have respective distributions P_0 (the uniform distribution on M_0) and the uniform distribution on $[-\Lambda_+, \Lambda_+]$. Therefore, on U_k, $\bar{Q}_{\tau,1}^{(i)}$ has a density with respect to the Lebesgue measure λ_{d+1} on $\mathbb{R}^d \times span(e)$ that can be written as

$$\bar{q}_{\tau,1}^{(i)}(x_1, x_{2:d}, y) = \frac{\mathbb{1}_{[-\Lambda_+, \Lambda_+]}(y)}{2 Vol(M_0)\Lambda_+}.$$

Analogously, nearby x_k a random variable Z having distribution $\bar{Q}_{\tau^k,1}^{(i)}$ can be represented as $Z = X + A_k(X - x_k)_1^i e$ where A_k has uniform distribution on $[A_-, A_+]$. Therefore, a straightforward change of variable yields the density

$$\bar{q}_{\tau^k,1}^{(i)}(x_1, x_{2:d}, y) = \frac{\mathbb{1}_{[A_- x_1^i, A_+ x_1^i]}(y)}{Vol(M_0)\,(A_+ - A_-)\,x_1^i}.$$

We recall that $Vol(M_0) = (2\tau_{min})^d Vol\left(M_0^{(0)}\right) = c_d' \tau_{min}^d$. Let us now tackle the right-hand side inequality, writing

$$
\begin{aligned}
&\int_{U_k} d\bar{Q}^{(i)}_{\tau,1} \wedge d\bar{Q}^{(i)}_{\tau^k,1} \\
&\quad = \int_{\mathcal{B}(x_k,\delta/2)} \left(\frac{\mathbb{1}_{[-\Lambda_+,\Lambda_+]}(y)}{2Vol(M_0)\Lambda_+} \right) \wedge \left(\frac{\mathbb{1}_{[A_- x_1^i, A_+ x_1^i]}(y)}{Vol(M_0)\,(A_+ - A_-)\,x_1^i} \right) dy dx_1 dx_{2:d} \\
&\quad \geq \int_{\mathcal{B}_{\mathbb{R}^{d-1}}(0,\frac{\delta}{4})} \int_{-\delta/4}^{\delta/4} \int_{\mathbb{R}} \left(\frac{\mathbb{1}_{[-\Lambda_+,\Lambda_+]}(y)}{2\Lambda_+} \right) \wedge \left(\frac{\mathbb{1}_{[A_- x_1^i, A_+ x_1^i]}(y)}{A_+ x_1^i/2} \right) \frac{dy dx_1 dx_{2:d}}{Vol(M_0)}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\int_{U_k} d\bar{Q}^{(i)}_{\tau,1} \wedge d\bar{Q}^{(i)}_{\tau^k,1} \\
&\quad \geq \frac{c_d}{\tau_{min}^d} \delta^{d-1} \int_0^{\delta/4} \int_{A_+ x_1^i/2}^{\Lambda_+ \wedge (A_+ x_1^i)} \frac{1}{2\Lambda_+} \wedge \frac{2}{A_+ x_1^i} dy dx_1 \\
&\quad \geq \frac{c_d}{\tau_{min}^d} \delta^{d-1} \int_0^{\delta/4} \int_{A_+ x_1^i/2}^{(c \wedge 1)(A_+ x_1^i)} \frac{(2c \wedge 1/2)}{2\Lambda_+} dy dx_1 \\
&\quad = \frac{c_d}{\tau_{min}^d} \delta^{d-1} (2c \wedge 1/2)\,(c \wedge 1 - 1/2)\,\frac{A_+}{\Lambda_+} \frac{(\delta/4)^{i+1}}{i+1} \\
&\quad \geq \frac{c_{d,i}}{C} \left(\frac{\delta}{\tau_{min}} \right)^d.
\end{aligned}
$$

For the integral on U_k', notice that by definition, $\bar{Q}^{(i)}_{\tau,n-1}$ and $\bar{Q}^{(i)}_{\tau^k,n-1}$ coincide on U_k' since they are respectively the image distributions of P_0 by functions that are equal on that set. Moreover, these two functions leave $\mathbb{R}^D \setminus \left\{ \mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(x_k, \delta) + \mathcal{B}_{span(e)}(0, \tau_{min}/2) \right\}$ unchanged. Therefore,

$$
\begin{aligned}
&\int_{U_k'} d\bar{Q}^{(i)}_{\tau,n-1} \wedge d\bar{Q}^{(i)}_{\tau^k,n-1} \\
&\qquad = P_0^{\otimes n-1}(U_k') \\
&\qquad = \left(1 - P_0\left(\mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(x_k, \delta) + \mathcal{B}_{span(e)}(0, \tau_{min}/2)\right)\right)^{n-1} \\
&\qquad = \left(1 - \omega_d \delta^d / Vol(M_0)\right)^{n-1},
\end{aligned}
$$

hence the result. □

C.3 Minimax Inconsistency Results

This section is devoted to the proof of lower bound for tangent space estimation (Theorem 5): we build hypotheses P, P' and apply Theorem 15. For $\delta \geq \Lambda > 0$, let $\mathcal{C}', \mathcal{C}' \subset \mathbb{R}^3$ be closed curves of the Euclidean space as in Figure 1, and such that outside the figure, $\mathcal{C}'$ and $\mathcal{C}'$ coincide and are $\mathcal{C}^\infty$. The bumped parts are

obtained with a smooth diffeomorphism similar to (9), centered at x. Here, δ and Λ can be chosen arbitrarily small.

Let $\mathcal{S}^{d-1} \subset \mathbb{R}^d$ be a $d-1$-sphere of radius $1/L_\perp$. Consider the Cartesian products $M_1 = \mathcal{C} \times \mathcal{S}^{d-1}$ and $M_1' = \mathcal{C}' \times \mathcal{S}^{d-1}$. M_1 and M_1' are subsets of $\mathbb{R}^{d+3} \subset \mathbb{R}^D$. Finally, let P_1 and P_1' denote the uniform distributions on M and M'. Note that M, M' can be built by homothecy of ratio $\lambda = 1/L_\perp$ from some unitary scaled $M_1^{(0)}, {M'}_1^{(0)}$, similarly to Section 5.2.3, yielding, from Proposition 4, that P_1, P_1' belong to $\mathcal{P}^k_{(x)}$ provided that $L_3/L_\perp^2, \dots, L_k/L_\perp^{k-1}, L_\perp^d/f_{min}$ and $f_{max}/L_\perp^d$ are large enough (depending only on d and k), and that Λ, δ and Λ^k/δ are small enough. From Le Cam's Lemma 15, we have for all $n \geq 1$,

$$\inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{(x)}} \mathbb{E}_{P^{\otimes n}} \angle\big(T_x M, \hat{T}\big) \geq \frac{1}{2} \angle\big(T_x M_1, T_x M_1'\big) \left(1 - TV(P_1, P_1')\right)^n .$$

By construction, $\angle\big(T_x M_1, T_x M_1'\big) = 1$, and since $\mathcal{C}$ and $\mathcal{C}'$ coincide outside $\mathcal{B}_{\mathbb{R}^3}(0, \delta)$,

$$\begin{aligned} TV\big(P_1, P_1'\big) &= Vol\left(\left(\mathcal{B}_{\mathbb{R}^3}(0,\delta) \cap \mathcal{C}\right) \times \mathcal{S}^{d-1}\right) / Vol\left(\mathcal{C} \times \mathcal{S}^{d-1}\right) \\ &= Length\left(\mathcal{B}_{\mathbb{R}^3}(0,\delta) \cap \mathcal{C}\right) / Length(\mathcal{C}) \\ &\leq c_{L_\perp} \delta. \end{aligned}$$

Hence, letting Λ, δ go to 0 with Λ^k/δ small enough, we get the announced bound.

We now tackle the lower bound on second fundamental form estimation with the same strategy. Let $M_2, M_2' \subset \mathbb{R}^D$ be d-dimensional submanifolds as in Figure 2: they both contain x, the part on the top of M_2 is a half d-sphere of radius $2/L_\perp$, the bottom part of M_2' is a piece of a d-plane, and the bumped parts are obtained with a smooth diffeomorphism similar to (9) centered at x. Outside $\mathcal{B}(x, \delta)$, M_2, M_2' coincide and connect smoothly the upper and lower parts. Let P_2, P_2' be the probability distributions obtained by the pushforward given by the bump maps. Under the same conditions on the parameters as previously, P_2 and P_2' belong to $\mathcal{P}^k_{(x)}$ according to Proposition 4. From Le Cam's Lemma 15 we deduce

$$\begin{aligned} \inf_{\widehat{II}} \sup_{P \in \mathcal{P}^k_{(x)}} \mathbb{E}_{P^{\otimes n}} &\left\| II_x^M \circ \pi_{T_x M} - \widehat{II} \right\|_{op} \\ &\geq \frac{1}{2} \left\| II_x^{M_2} \circ \pi_{T_x M_2} - II_x^{M_2'} \circ \pi_{T_x M_2'} \right\|_{op} \left(1 - TV(P_2, P_2')\right)^n . \end{aligned}$$

By construction, $\left\| II_x^{M_2} \circ \pi_{T_x M_2} \right\|_{op} = 0$, and since M_2' is a part of a sphere of radius $2/\text{L}_\perp$ nearby x, $\left\| II_x^{M_2'} \circ \pi_{T_x M_2'} \right\|_{op} = L_\perp/2$. Hence,

$$\left\| II_x^{M_2} \circ \pi_{T_x M_2} - II_x^{M_2'} \circ \pi_{T_x M_2'} \right\|_{op} \geq L_\perp/2.$$

Moreover, since P_2 and P_2' coincide on $\mathbb{R}^D \setminus \mathcal{B}(x, \delta)$,

$$TV(P_2, P_2') = Pc_2(\mathcal{B}(x, \delta)) \leq c_{d, L_\perp} \delta^d.$$

Letting Λ, δ go to 0 with Λ^k/δ small enough, we have the desired result.

References

[AB06] Stephanie B. Alexander and Richard L. Bishop. Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geom. Dedicata*, 117:65–84, 2006.

[AL15] E. Aamari and C. Levrard. Stability and Minimax Optimality of Tangential Delaunay Complexes for Manifold Reconstruction. *ArXiv e-prints*, December 2015.

[ALZ13] E. Arias-Castro, G. Lerman, and T. Zhang. Spectral Clustering Based on Local PCA. *ArXiv e-prints*, January 2013.

[APR16] E. Arias-Castro, B. Pateiro-López, and A. Rodríguez-Casal. Minimax Estimation of the Volume of a Set with Smooth Boundary. *ArXiv e-prints*, May 2016.

[BG14] Jean-Daniel Boissonnat and Arijit Ghosh. Manifold reconstruction using tangential Delaunay complexes. *Discrete Comput. Geom.*, 51(1):221–267, 2014.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[Bou02] Olivier Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris*, 334(6):495–500, 2002.

[CC16] Siu-Wing Cheng and Man-Kwun Chiu. Tangent estimation from point samples. *Discrete Comput. Geom.*, 56(3):505–557, 2016.

[CDR05] Siu-Wing Cheng, Tamal K. Dey, and Edgar A. Ramos. Manifold reconstruction from point samples. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1018–1027. ACM, New York, 2005.

[CGLM13] F. Chazal, M. Glisse, C. Labruère, and B. Michel. Optimal rates of convergence for persistence diagrams in Topological Data Analysis. *ArXiv e-prints*, May 2013.

[CP05] F. Cazals and M. Pouget. Estimating differential quantities using polynomial fitting of osculating jets. *Comput. Aided Geom. Design*, 22(2):121–146, 2005.

[dC92] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[DVW15] Ramsay Dyer, Gert Vegter, and Mathijs Wintraecken. Riemannian simplices and triangulations. *Geom. Dedicata*, 179:91–138, 2015.

[Fed59] Herbert Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[Fed69] Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[FIK+15] C. Fefferman, S. Ivanov, Y. Kurylev, M. Lassas, and H. Narayanan. Reconstruction and interpolation of manifolds I: The geometric Whitney problem. *ArXiv e-prints*, August 2015.

[FLR+14] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *Ann. Statist.*, 42(6):2301–2339, 2014.

[GM11] Michael S. Gashler and Tony Martinez. Tangent space guided intelligent neighbor finding. In *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN'11*, pages 2617–2624. IEEE Press, 2011.

[GPPVW12] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.*, 40(2):941–963, 2012.

[GVL96] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[GWM01] Stefan Gumhold, Xinlong Wang, and Rob MacLeod. Feature Extraction from Point Clouds. In *10th International Meshing Roundtable*, pages 293–305. Sandia National Laboratories,, 2001.

[Har51] Philip Hartman. On geodesic coordinates. *Amer. J. Math.*, 73:949–954, 1951.

[KZ15] Arlene K. H. Kim and Harrison H. Zhou. Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.*, 9(1):1562–1582, 2015.

[Mas07] Pascal Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

[MOG11] Q. Merigot, M. Ovsjanikov, and L. J. Guibas. Voronoi-based curvature and feature estimation from point clouds. *IEEE Transactions on Visualization and Computer Graphics*, 17(6):743–756, June 2011.

[NSW08] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008.

[RS00] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE*, 290:2323–2326, 2000.

[Rus04] Szymon Rusinkiewicz. Estimating curvatures and their derivatives on triangle meshes. In *Proceedings of the 3D Data Processing, Visualization, and Transmission, 2Nd International Symposium*, 3DPVT '04, pages 486–493, Washington, DC, USA, 2004. IEEE Computer Society.

[SW12] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.*, 65(8):1067–1144, 2012.

[TdSL00] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319, 2000.

[TVF13] Hemant Tyagi, Elı f Vural, and Pascal Frossard. Tangent space estimation for smooth embeddings of Riemannian manifolds. *Inf. Inference*, 2(1):69–114, 2013.

[UM14] Konstantin Usevich and Ivan Markovsky. Optimization on a grassmann manifold with application to system identification. *Automatica*, 50(6):1656 – 1662, 2014.

[Was16] L. Wasserman. Topological Data Analysis. *ArXiv e-prints*, September 2016.

[Yu97] Bin Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.