Minimax Manifold Estimation

Christopher R. Genovese   GENOVESE@STAT.CMU.EDU
Department of Statistics
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213, USA

Marco Perone-Pacifico   MARCO.PERONEPACIFICO@UNIROMA1.IT
Dipartimento di Scienze Statistiche
Sapienza University of Rome
Piazzale Aldo Moro 5 – 00185 Roma, Italy

Isabella Verdinelli[*]   ISABELLA@STAT.CMU.EDU
Larry Wasserman[†]   LARRY@STAT.CMU.EDU
Department of Statistics
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213, USA

Editor: Ulrike von Luxburg

Abstract

We find the minimax rate of convergence in Hausdorff distance for estimating a manifold M of dimension d embedded in $\mathbb{R}^D$ given a noisy sample from the manifold. Under certain conditions, we show that the optimal rate of convergence is $n^{-2/(2+d)}$. Thus, the minimax rate depends only on the dimension of the manifold, not on the dimension of the space in which M is embedded.

Keywords: manifold learning, minimax estimation

1. Introduction

We consider the problem of estimating a manifold M given noisy observations near the manifold. The observed data are a random sample $Y_1, \ldots, Y_n$ where $Y_i \in \mathbb{R}^D$. The model for the data is

$$Y_i = \xi_i + Z_i$$

where $\xi_1, \ldots, \xi_n$ are unobserved variables drawn from a distribution supported on a manifold M with dimension $d < D$. The noise variables $Z_1, \ldots, Z_n$ are drawn from a distribution F. Our main assumption is that M is a compact, d-dimensional, smooth Riemannian submanifold in $\mathbb{R}^D$; the precise conditions on M are given in Section 2.1.

A manifold M and a distribution for (ξ, Z) induce a distribution $Q \equiv Q_M$ for Y. In Section 2.2, we define a class of such distributions

$$Q = \left\{ Q_M : M \in \mathcal{M} \right\}$$

---

*. Also in the Department of Statistical Sciences, Sapienza University of Rome, Italy.

†. Also in the Machine Learning Department, Carnegie Mellon University.

where $\mathcal{M}$ is a set of manifolds. Given two sets A and B, the Hausdorff distance between A and B is

$$H(A,B) = \inf\Big\{ \varepsilon : \ A \subset B \oplus \varepsilon \ \ \text{and} \ \ B \subset A \oplus \varepsilon \Big\}$$

where

$$A \oplus \varepsilon = \bigcup_{x \in A} B_D(x, \varepsilon)$$

and $B_D(x, \varepsilon)$ is an open ball in $\mathbb{R}^D$ centered at x with radius ε. We are interested in the minimax risk

$$R_n(Q) = \inf_{\widehat{M}} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[H(\widehat{M}, M)]$$

where the infimum is over all estimators $\widehat{M}$. By an estimator $\widehat{M}$ we mean a measurable function of $Y_1, \ldots, Y_n$ taking values in the set of all manifolds. Our first main result is the following minimax lower bound which is proved in Section 3.

Theorem 1 *Under assumptions (A1)-(A4) given in Section 2, there is a constant $C_1 > 0$ such that, for all large n,*

$$\inf_{\widehat{M}} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q\left[H(\widehat{M}, M)\right] \geq C_1 \left(\frac{1}{n}\right)^{\frac{2}{2+d}}$$

where the infimum is over all estimators $\widehat{M}$.

Thus, no method of estimating M can have an expected Hausdorff distance smaller than the stated bound. Note that the rate depends on d but not on D even though the support of the distribution Q for Y has dimension D. Our second result is the following upper bound which is proved in Section 4.2.

Theorem 2 *Under assumptions (A1)-(A4) given in Section 2, there exists an estimator $\widehat{M}$ such that, for all large n,*

$$\sup_{Q \in \mathcal{Q}} \mathbb{E}_Q\left[H(\widehat{M}, M)\right] \leq C_2 \left(\frac{\log n}{n}\right)^{\frac{2}{2+d}}$$

for some $C_2 > 0$.

Thus the rate is tight, up to logarithmic factors. The estimator in Theorem 2 is of theoretical interest because it establishes that the lower bound is tight. But, the estimator constructed in the proof of that theorem is not practical and so in Section 5, we construct a very simple estimator $\widehat{M}$ such that

$$\sup_{Q \in \mathcal{Q}} \mathbb{E}_Q\left[H(\widehat{M}, M)\right] \leq \left(\frac{C \log n}{n}\right)^{1/D}.$$

This is slower than the minimax rate, but the estimator is computationally very simple and requires no knowledge of d or the smoothness of M.

1.1 Related Work

There is a vast literature on manifold estimation. Much of the literature deals with using manifolds for the purpose of dimension reduction. See, for example, Baraniuk and Wakin (2007) and references therein. We are interested instead in actually estimating the manifold itself. There is a large literature on this problem in the field of computational geometry; see, for example, Dey (2006), Dey and Goswami (2004), Chazal and Lieutier (2008) Cheng and Dey (2005) and Boissonnat and Ghosh (2010). However, very few papers allow for noise in the statistical sense, by which we mean observations drawn randomly from a distribution. In the literature on computational geometry, observations are called noisy if they depart from the underlying manifold in a very specific way: the observations have to be close to the manifold but not too close to each other. This notion of noise is quite different from random sampling from a distribution. An exception is Niyogi et al. (2008) who constructed the following estimator. Let $I = \{i : \widehat{p}(Y_i) > \lambda\}$ where $\widehat{p}$ is a density estimator. They define $\widehat{M} = \bigcup_{i \in I} B_D(Y_i, \varepsilon)$ and they show that if λ and ε are chosen properly, then $\widehat{M}$ is homologous to M. (This means that M and $\widehat{M}$ share certain topological properties.) However, the result does not guarantee closeness in Hausdorff distance. Note that $\bigcup_{i=1}^{n} B_D(Y_i, \varepsilon)$ is precisely the Devroye-Wise estimator for the support of a distribution (Devroye and Wise, 1980).

1.2 Notation

Given a set S, we denote its boundary by ∂S. We let $B_D(x, r)$ denote a D-dimensional open ball centered at x with radius r. If A is a set and x is a point then we write $d(x, A) = \inf_{y \in A} ||x - y||$ where $|| \cdot ||$ is the Euclidean norm. Let

$$A \circ B = (A \cap B^c) \bigcup (A^c \cap B)$$

denote symmetric set difference between sets A and B.

The uniform measure on a manifold M is denoted by μ_M. Lebesgue measure on $\mathbb{R}^k$ is denoted by ν_k. In case $k = D$, we sometimes write V instead of ν_D; in other words $V(A)$ is simply the volume of A. Any integral of the form $\int f$ is understood to be the integral with respect to Lebesgue measure on $\mathbb{R}^D$. If P and Q are two probability measures on $\mathbb{R}^D$ with densities p and q then the *Hellinger distance* between P and Q is

$$h(P, Q) \equiv h(p, q) = \sqrt{\int (\sqrt{p} - \sqrt{q})^2} = \sqrt{2 \left(1 - \int \sqrt{pq} \right)}$$

where the integrals are with respect to ν_D. Recall that

$$\ell_1(p, q) \leq h(p, q) \leq \sqrt{\ell_1(p, q)} \tag{1}$$

where $\ell_1(p, q) = \int |p - q|$. Let $p(x) \wedge q(x) = \min\{p(x), q(x)\}$. The *affinity* between P and Q is

$$||P \wedge Q|| = \int p \wedge q = 1 - \frac{1}{2} \int |p - q|.$$

Let P^n denote the n-fold product measure based on n independent observations from P. In the appendix Section 7.1 we show that

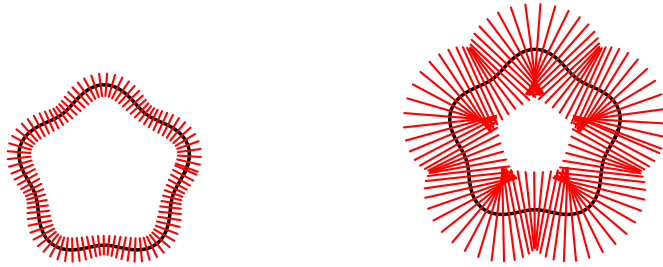$$||P^n \wedge Q^n|| \geq \frac{1}{2} \left(1 - \frac{1}{2} \int |p - q| \right)^{2n}. \tag{2}$$

Figure 1: The condition number $\Delta(M)$ of a manifold is the largest number κ such that the normals to the manifold do not cross as long as they are not extended beyond κ. The plot on the left shows a one-dimensional manifold (a curve) and some normals of length $r < \kappa$. The plot on the right shows the same manifold and some normals of length $r > \kappa$.

We write $X_n = O_P(a_n)$ to mean that, for every $\varepsilon > 0$ there exists $C > 0$ such that $\mathbb{P}(||X_n||/a_n > C) \le \varepsilon$ for all large n. Throughout, we use symbols like $C, C_0, C_1, c, c_0, c_1 \ldots$ to denote generic positive constants whose value may be different in different expressions.

2. Model Assumptions

In this section we describe all the assumptions on the manifold and on the underlying distributions.

2.1 Manifold Conditions

We shall be concerned with d-dimensional compact Riemannian submanifolds without boundary embedded in $\mathbb{R}^D$ with $d < D$. (Informally, this means that M looks like $\mathbb{R}^d$ in a small neighborhood around any point in M.) We assume that M is contained in some compact set $\mathcal{K} \subset \mathbb{R}^D$.

At each $u \in M$ let $T_u M$ denote the tangent space to M and let $T_u^\perp M$ be the normal space. We can regard $T_u M$ as a d-dimensional hyperplane in $\mathbb{R}^D$ and we can regard $T_u^\perp M$ as the $D - d$ dimensional hyperplane perpendicular to $T_u M$. Define the *fiber of size a at u* to be $L_a(u) \equiv L_a(u, M) = T_u^\perp M \bigcap B_D(u, a)$.

Let $\Delta(M)$ be the largest r such that each point in $M \oplus r$ has a unique projection onto M. The quantity $\Delta(M)$ will be small if either M highly curved or if M is close to being self-intersecting. Let $\mathcal{M} \equiv \mathcal{M}(\kappa)$ denote all d-dimensional manifolds embedded in $\mathcal{K}$ such that $\Delta(M) \ge \kappa$. Throughout this paper, κ is a fixed positive constant. The quantity $\Delta(M)$ has been rediscovered many times. It is called the *condition number* in Niyogi et al. (2006), the *thickness* in Gonzalez and Maddocks (1999) and the *reach* in Federer (1959).

An equivalent definition of $\Delta(M)$ is the following: $\Delta(M)$ is the largest number r such that the fibers $L_r(u)$ never intersect. See Figure 1. Note that if M is a sphere then $\Delta(M)$ is just the radius of the sphere and if M is a linear space then $\Delta(M) = \infty$. Also, if $\sigma < \Delta(M)$ then $M \oplus \sigma$ is the disjoint union of its fibers:

$$M \oplus \sigma = \bigcup_{u \in M} L_\sigma(u). \tag{3}$$

Define $\mathrm{tube}(M, a) = \bigcup_{u \in M} L_a(u)$. Thus, if $\sigma < \Delta(M)$ then $M \oplus \sigma = \mathrm{tube}(M, \sigma)$.

Let $p, q \in M$. The angle between two tangent spaces T_p and T_q is defined to be

$$\mathsf{angle}(T_p, T_q) = \cos^{-1}\left(\min_{u \in T_p} \max_{v \in T_q} |\langle u - p, v - q \rangle|\right)$$

where $\langle u, v \rangle$ is the usual inner product in $\mathbb{R}^D$. Let $d_M(p, q)$ denote the geodesic distance between $p, q \in M$.

We now summarize some useful results from Niyogi et al. (2006).

Lemma 3 *Let $M \subset \mathcal{K}$ be a manifold and suppose that $\Delta(M) = \kappa > 0$. Let $p, q \in M$.*

1. *Let γ be a geodesic connecting p and q with unit speed parameterization. Then the curvature of γ is bounded above by $1/\kappa$.*

2. *$\cos(\mathsf{angle}(T_p, T_q)) > 1 - d_M(p, q)/\kappa$. Thus, $\mathsf{angle}(T_p, T_q) \le \sqrt{2 d_M(p,q)/\kappa} + o(\sqrt{d_M(p,q)/\kappa})$.*

3. *If $a = ||p - q|| \le \kappa/2$ then $d_M(p, q) \le \kappa - \kappa\sqrt{1 - (2a)/\kappa} = a + o(a)$.*

4. *If $a = ||p - q|| \le \kappa/2$ then $a \ge d_M(p, q) - (d_M(p, q))^2/(2\kappa)$.*

5. *If $||q - p|| > \varepsilon$ and $v \in B_D(q, \varepsilon) \cap T_p^{\perp} M \cap B_D(p, \kappa)$ then $||v - p|| < \varepsilon^2/\kappa$.*

6. *Fix any $\delta > 0$. There exists points $x_1, \dots, x_N \in M$ such that $M \subset \bigcup_{j=1}^{N} B_D(x_j, \delta)$ and such that $N \le (c/\delta)^d$.*

For further information about manifolds, see Lee (2002).

2.2 Distributional Assumptions

The distribution of Y is induced by the distribution of ξ and Z. We will assume that ξ is drawn uniformly on the manifold. Then we assume that Z is drawn uniformly on the normal to M. More precisely, given ξ, we draw Z uniformly on $L_\sigma(\xi)$. In other words, the noise is perpendicular to the manifold. The result is that, if $\sigma < \kappa$, then the distribution $Q = Q_M$ of Y has support equal to $M \oplus \sigma$.

The distributional assumption on ξ is not critical. Any smooth density bounded away from 0 on the manifold will lead to similar results. However, the assumption on the noise Z is critical. We have chosen the simplest noise distribution here. (Perpendicular noise is also assumed in Niyogi et al., 2008.) In current work, we are deriving the rates for more complicated noise distributions. The rates are quite different and the proofs are more complex. Those results will be reported elsewhere.

The set of distributions we consider is as follows. Let κ and σ be fixed positive numbers such that $0 < \sigma < \kappa$. Let

$$Q \equiv Q(\kappa, \sigma) = \left\{ Q_M : M \in \mathcal{M}(\kappa) \right\}.$$

For any $M \in \mathcal{M}(\kappa)$ consider the corresponding distribution Q_M, supported on $S_M = M \oplus \sigma$. Let q_M be the density of Q_M with respect to Lebesgue measure. We now show that q_M is bounded above and below by a uniform density.

Recall that the essential supremum and essential infimum of q_M are defined by

$$\operatorname*{ess\,sup}_{y \in A} q_M = \inf\left\{ a \in \mathbb{R} : \nu_D(\{y : q_M(y) > a\} \cap A) = 0 \right\}$$

and

$$\underset{y \in A}{\mathrm{ess\,inf}}\, q_M = \sup\Big\{a \in \mathbb{R}: \; \nu_D(\{y: \; q_M(y) < a\} \cap A) = 0\Big\}.$$

Also recall that, by the Lebesgue density theorem, $q_M(y) = \lim_{\varepsilon \to 0} Q_M(B_D(y,\varepsilon))/V(B_D(y,\varepsilon))$ for almost all y. Let U_M be the uniform distribution on $M \oplus \sigma$ and let $u_M = 1/V(M \oplus \sigma)$ be the density of U_M. Note that, for $A \subset M \oplus \sigma$, $U_M(A) = V(A)/V(M \oplus \sigma)$.

Lemma 4 *There exist constants $0 < C_* \leq C^* < \infty$, depending only on κ and d, such that*

$$C_* \leq \inf_{M \in \mathcal{M}} \mathrm{ess\,inf}_{y \in S_M} \frac{q_M(y)}{u_M(y)} \leq \sup_{M \in \mathcal{M}} \mathrm{ess\,sup}_{y \in S_M} \frac{q_M(y)}{u_M(y)} \leq C^*.$$

Proof Choose any $M \in \mathcal{M}(\kappa)$. Let x by any point in the interior of S_M. Let $B = B_D(x,\varepsilon)$ where $\varepsilon > 0$ is small enough so that $B \subset S_M = M \oplus \sigma$. Let y be the projection of x onto M. We want to upper and lower bound $Q(B)/V(B)$. Then we will take the limit as $\varepsilon \to 0$. Consider the two spheres of radius κ tangent to M at y in the direction of the line between x and y. (See Figure 2.) Note that $Q(B)$ is maximized by taking M to be equal to the upper sphere and $Q(B)$ is minimized by taking M to be equal to the lower sphere. Let us consider first the case where M is equal to the upper sphere. Let

$$U = \Big\{u \in M: \; L_\sigma(u) \cap B \neq \emptyset\Big\}$$

be the projection of B onto M. By simple geometry, $U = M \cap B_D(y, r\varepsilon)$ where

$$\Big(1 + \frac{\sigma}{\kappa}\Big)^{-1} \leq r \leq \Big(1 + \frac{\sigma}{\kappa}\Big).$$

Let Vol denote d-dimensional volume on M. Then $\mathrm{Vol}(B_D(y, r\varepsilon) \cap M) \leq c_1 r^d \varepsilon^d \omega_d$ where ω_d is the volume of a unit d-ball and c_1 depends only on κ and d. To see this, note that because M is a manifold and $\Delta(M) \geq \kappa$, it follows that near y, M may be locally parameterized as a smooth function $f = (f_1, \ldots, f_{D-d})$ over $B \cap T_y M$. The surface area of the graph of f over $B \cap T_y M$ is bounded by $\int_{B_D(y, r\varepsilon) \cap T_y M} \sqrt{1 + \|\nabla f_i\|^2}$, which is bounded by a constant c_1 uniformly over $\mathcal{M}$. Hence, $\mathrm{Vol}(B_D(y, r\varepsilon) \cap M) \leq c_1 \mathrm{Vol}(B_D(y, r\varepsilon) \cap T_y M) = c_1 r^d \varepsilon^d \omega_d$.

Let Λ_M be the uniform distribution on M and let Γ_u denote the uniform measure on $L_\sigma(u)$. Note that, for $u \in U$, $L_\sigma(u) \cap B$ is a $(D-d)$-ball whose radius is at most ε. Hence,

$$\Gamma_u(L_\sigma(u) \cap B) \leq \frac{\varepsilon^{D-d} \omega_{D-d}}{\sigma^{D-d} \omega_{D-d}} = \Big(\frac{\varepsilon}{\sigma}\Big)^{D-d}.$$

Thus,

$$\begin{aligned}
Q_M(B) &= \int_M \Gamma_u(B \cap L_\sigma(u)) d\Lambda_M(u) = \int_U \Gamma_u(B \cap L_\sigma(u)) d\Lambda_M(u) \\
&\leq \Big(\frac{\varepsilon}{\sigma}\Big)^{D-d} \Lambda(U) = \Big(\frac{\varepsilon}{\sigma}\Big)^{D-d} \frac{\mathrm{Vol}(B_D(y,r) \cap M)}{\mathrm{Vol}(M)} \\
&\leq \Big(\frac{\varepsilon}{\sigma}\Big)^{D-d} \frac{\varepsilon^d r^d \omega_d}{\mathrm{Vol}(M)} \leq \Big(\frac{\varepsilon}{\sigma}\Big)^{D-d} \frac{\varepsilon^d (1 + \sigma/\kappa)^d \omega_d}{\mathrm{Vol}(M)}.
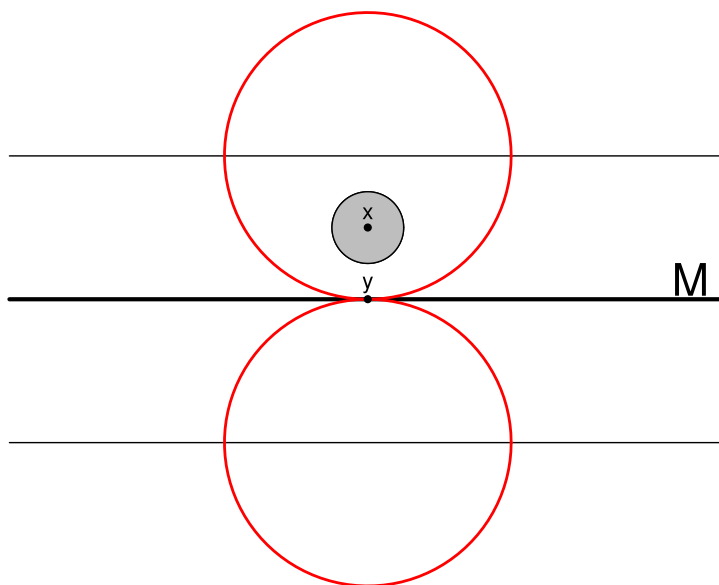\end{aligned}$$

Figure 2: Figure for proof of Lemma 4. x is a point in the support $M \oplus \sigma$. y is the projection of x onto M. The two spheres are tangent to M at y and have radius κ.

Now, $U_M(B) = V(B)/V(M \oplus \sigma) = \varepsilon^D \omega_D/(\sigma^{D-d} \text{Vol}(M))$. Hence,

$$\frac{Q_M(B)}{U_M(B)} \leq \left(1 + \frac{\sigma}{\kappa}\right)^d \omega_d.$$

Taking limits as $\varepsilon \to 0$ we have that $q_M(y) \leq C^* u_M(y)$ for almost all y.

The proof of the lower bound is similar to the upper bound except for the following changes: let U_0 denote all $u \in U$ such that the radius of $B \cap L_\sigma(u)$ is at least $\varepsilon/2$. Then $\Lambda(U_0) \geq \Lambda(U)(1 - O(\varepsilon))$ and the projection of U_0 onto M is again of the form $B_D(y, r\varepsilon) \cap M$. By Lemma 5.3 of Niyogi et al. (2006),

$$\text{Vol}(B_D(y,r) \cap M) \geq \left(1 - \frac{r^2 \varepsilon^2}{4\kappa^2}\right)^{d/2} r^d \varepsilon^d \omega_d$$

and the latter is larger than $2^{-d/2} r^d \varepsilon^d \omega_d$ for all small ε. Also, $\Gamma_u(L_\sigma(u) \cap B) \geq (\varepsilon/(2\sigma))^{D-d}$ for all $u \in U_0$. ∎

Of course, an immediate consequence of the above lemma is that, for every $M \in \mathcal{M}(\kappa)$ and every measurable set A, $C_* U_M(A) \leq Q_M(A) \leq C^* U_M(A)$. We conclude this section by recording all the assumptions in Theorems 1 and 2:

(A1) The manifold M is d-dimensional and is contained in a compact set $\mathcal{K} \subset \mathbb{R}^D$ with $d < D$.

(A2) The manifold M satisfies $\Delta(M) \geq \kappa > 0$.

(A3) The observed data $Y_1, \ldots, Y_n$ are iid observations with $Y_i = X_i + \xi_i$. Here, $\xi_1, \ldots, \xi_n$ are drawn uniformly on M. X_i given ξ_i is drawn uniformly on $L_\sigma(\xi_i) = T_{\xi_i}^{\perp} \cap B_D(\xi_i, \sigma)$.

(A4) The noise level σ satisfies $0 < \sigma < \kappa$.

Remark: As noted by a referee, the assumptions are very specific and the results do depend critically on the assumptions especially the assumption that d is known.

Remark: A referee has pointed out that another reasonable model is to assume that the Y_i have a uniform distribution on the tube of size σ around the manifold. To the best of our knowledge, this does not correspond to our model except in the special case where $\Delta(M) = \infty$. However, all the results of our paper still apply in this case as long as $\sigma < \kappa$.

3. Minimax Lower Bound

In this section we derive a lower bound on the minimax rate of convergence for this problem. We will make use of the following result due to LeCam (1973). The following version is from Lemma 1 of Yu (1997).

Lemma 5 (Le Cam 1973) *Let Q be a set of distributions. Let $\theta(Q)$ take values in a metric space with metric ρ. Let $Q_0, Q_1 \in Q$ be any pair of distributions in Q. Let $Y_1, \ldots, Y_n$ be drawn iid from some $Q \in Q$ and denote the corresponding product measure by Q^n. Let $\widehat{\theta}(Y_1, \ldots, Y_n)$ be any estimator. Then*

$$\sup_{Q \in Q} \mathbb{E}_{Q^n} \left[\rho(\widehat{\theta}(Y_1, \ldots, Y_n), \theta(Q)) \right] \geq \rho\left(\theta(Q_0), \theta(Q_1)\right) \, ||Q_0^n \wedge Q_1^n||.$$

To get a useful bound from Le Cam's lemma, we need to construct an appropriate pair Q_0 and Q_1. This is the topic of the next subsection.

3.1 A Geometric Construction

In this section, we construct a pair of manifolds $M_0, M_1 \in \mathcal{M}(\kappa)$ and corresponding distributions Q_0, Q_1 for use in Le Cam's lemma. An informal description is as follows. Roughly speaking, M_0 and M_1 minimize the Hellinger distance $h(Q_0, Q_1)$ subject to their Hausdorff distance $H(M_0, M_1)$ being equal to a given value γ.

Let

$$M_0 = \left\{ (u_1, \ldots, u_d, 0, \ldots, 0) : -1 \leq u_j \leq 1, \, 1 \leq j \leq d \right\}$$

be a d-dimensional hyperplane in $\mathbb{R}^D$. Hence $\Delta(M_0) = \infty$. Place a hypersphere of radius κ below M_0. Push the sphere upwards into M_0 causing a bump of height γ at the origin. This creates a new manifold M_0' such that $H(M_0, M_0') = \gamma$. However, M_0' is not smooth. We will roll a sphere of radius κ around M_0' to get a smooth manifold M_1 as in Figure 3. We re-iterate that this is only an informal description and the reader should see Section 7.2 for the formal details.

Theorem 6 *Let γ be a small positive number. Let M_0 and M_1 be as defined in Section 7.2. Let Q_i be the corresponding distributions on $M_i \oplus \sigma$ for $i = 0, 1$. Then:*

1. *$\Delta(M_i) \geq \kappa$, $i = 0, 1$.*

2. *$H(M_0, M_1) = \gamma$.*

3. *$\int |q_0 - q_1| = O(\gamma^{(d+2)/2})$.*
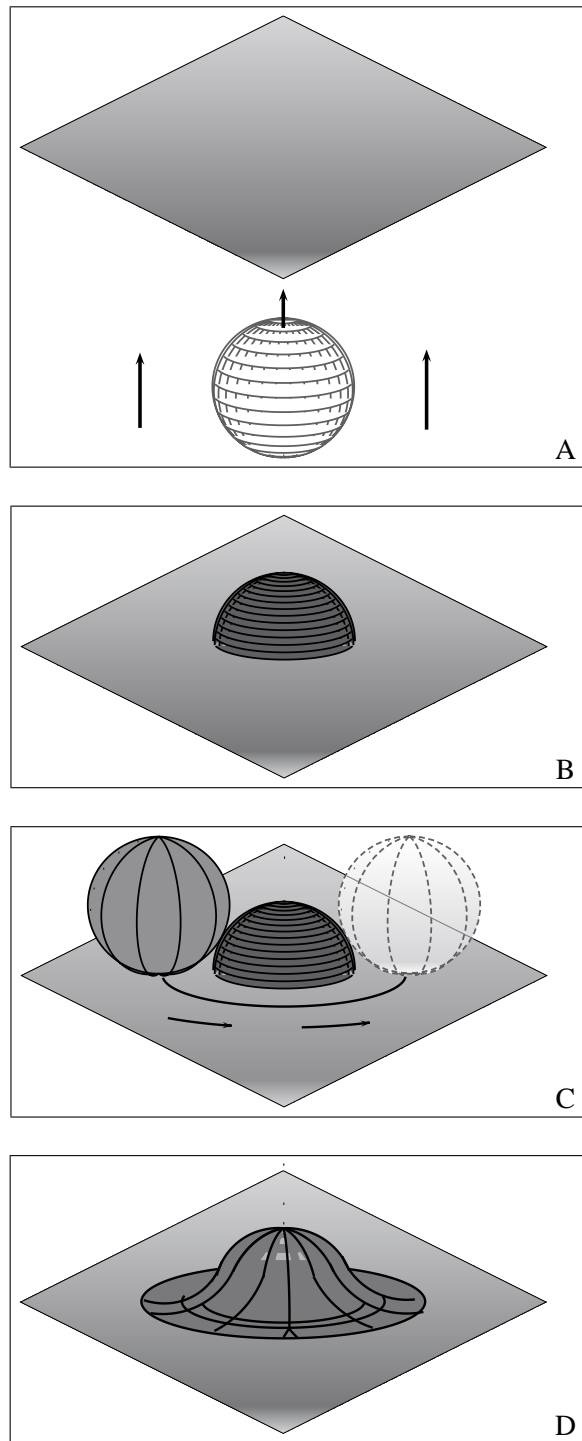
Proof See Section 7.2. ∎

Figure 3: A sphere of radius κ is pushed upwards into the plane M_0 (panel A). The resulting manifold M_0' is not smooth (panel B). A sphere is then rolled around the manifold (panel C) to produce a smooth manifold M_1 (panel D).

3.2 Proof of the Lower Bound

Now we are in a position to prove the first theorem. Let us first restate the theorem.

Theorem 1. *Under assumptions (A1)-(A4), there is a constant $C > 0$ such that, for all large n,*

$$\inf_{\widehat{M}} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q \left[H(\widehat{M}, M) \right] \geq C n^{-\frac{2}{2+d}}$$

where the infimum is over all estimators $\widehat{M}$.

Proof of Theorem 1. Let M_0 and M_1 be as defined in Section 3.1. Let Q_i be the uniform distribution on $M_i \oplus \sigma$, $i = 0, 1$. Let q_i be the density of Q_i with respect to Lebesgue measure ν_D, $i = 0, 1$. Then, from Theorem 6, $H(M_0, M_1) = \gamma$ and $\int |q_0 - q_1| = O(\gamma^{(d+2)/2})$. Le Cam's lemma then gives, for any $\widehat{M}$,

$$\sup_{Q \in \mathcal{Q}} \mathbb{E}_{Q^n}[H(M, \widehat{M})] \geq H(M_0, M_1) \, ||Q_0^n \wedge Q_1^n|| \geq \frac{\gamma}{2}(1 - c\gamma^{(d+2)/2})^{2n}$$

where we used Equation (2). Setting $\gamma = n^{-2/(d+2)}$ yields the result. ∎

4. Upper bound

To establish the upper bound, we will construct an estimator that achieves the appropriate rate. The estimator is intended only for the theoretical purpose of establishing the rate. (A simpler but non-optimal method is discussed in Section 5.) Recall that $\mathcal{M} = \mathcal{M}(\kappa)$ is the set of all d-dimensional submanifolds M contained in $\mathcal{K}$ such that $\Delta(M) \geq \kappa > 0$. Before proceeding, we need to discuss sieve maximum likelihood.

4.1 Sieve Maximum Likelihood

Let $\mathcal{P}$ be any set of distributions such that each $P \in \mathcal{P}$ has a density p with respect to Lebesgue measure ν_D. Recall that h denotes Hellinger distance. A set of pairs of functions $\mathcal{B} = \{(\ell_1, u_1), \dots, (\ell_N, u_N)\}$ is an ε-Hellinger bracketing for $\mathcal{P}$ if, (i) for each $p \in \mathcal{P}$ there is a $(\ell, u) \in \mathcal{B}$ such that $\ell(y) \leq p(y) \leq u(y)$ for all y and (ii) $h(\ell, u) \leq \varepsilon$. The logarithm of the size of the smallest ε-bracketing is called the *bracketing entropy* and is denoted by $\mathcal{H}_{[]}(\varepsilon, \mathcal{P}, h)$.

We will make use of the following result which is Example 4 of Shen and Wong (1995).

Theorem 7 (Shen and Wong, 1995) *Let ε_n solve the equation $\mathcal{H}_{[]}(\varepsilon_n, \mathcal{P}, h) = n\varepsilon_n^2$. Let $(\ell_1, u_1), \dots, (\ell_N, u_N)$ be an ε_n bracketing where $N = \mathcal{H}_{[]}(\varepsilon_n, \mathcal{P}, h)$. Define the set of densities $S_n^* = \{p_1^*, \dots, p_N^*\}$ where $p_t^* = u_t / \int u_t$. Let $\widehat{p}^*$ maximize the likelihood $\prod_{i=1}^n p_t^*(Y_i)$ over the set S_n^*. Then*

$$\sup_{P \in \mathcal{P}} P^n \left(\{ h(p, \widehat{p}^*) \geq \varepsilon_n \} \right) \leq c_1 e^{-c_2 n \varepsilon_n^2}.$$

The sequence $\{S_n^*\}$ in Theorem 7 is called a *sieve* and the estimator $\widehat{p}^*$ is called a *sieve-maximum likelihood estimator*. The estimator $\widehat{p}^*$ need not be in $\mathcal{P}$. We will actually need an estimator that is contained in $\mathcal{P}$. We may construct one as follows. Let $\widehat{p}^*$ be the sieve mle corresponding to S_n^*. Then $\widehat{p}^* = p_t^*$ for some t. Let $(\widehat{\ell}, \widehat{u}) \equiv (\ell_t, u_t)$ be the corresponding bracket.

Lemma 8 *Assume the conditions in Theorem 7. Let $\widehat{p}$ be any density in $\mathcal{P}$ such that $\widehat{\ell} \leq \widehat{p} \leq \widehat{u}$. If $\varepsilon_n \leq 1$ then*

$$\sup_{P \in \mathcal{P}} P^n \left(\{ h(p, \widehat{p}) \geq c\varepsilon_n \} \right) \leq c_1 e^{-c_2 n \varepsilon_n^2}.$$

Proof By the triangle inequality, $h(p,\widehat{p}) \le h(p,\widehat{p}^*) + h(\widehat{p},\widehat{p}^*) = h(p,\widehat{p}^*) + h(\widehat{p},u_t/\int u_t)$ where $\widehat{p}^* = u_t/\int u_t$ for some t. From Theorem 7, $h(p,\widehat{p}^*) \le \varepsilon_n$ with high probability. Thus we need to show that $h(\widehat{p},u_t/\int u_t) \le C\varepsilon_n$. It suffices to show that, in general, $h(p,u/\int u) \le C h(\ell,u)$ whenever $\ell \le p \le u$.

Let (ℓ,u) be a bracket and let $\delta^2 = h^2(\ell,u) \le 1$. Let $\ell \le p \le u$. We claim that $h^2(p,u/\int u) \le 4\delta^2$. (Taking $\delta = \varepsilon_n$ then proves the result.) Let $c^2 = \int u$. Then $1 \le c^2 = \int u = \int p + \int (u-p) = 1 + \int (u-p) = 1 + \ell_1(u,p) \le 1 + 2h(u,\ell) = 1 + 2\delta$. Now,

$$
\begin{aligned}
h^2\left(p,\frac{u}{\int u}\right) &= \int (\sqrt{u}/c - \sqrt{p})^2 = \frac{1}{c^2}\int (\sqrt{u} - c\sqrt{p})^2 \le \int (\sqrt{u} - c\sqrt{p})^2 \\
&= \int ((\sqrt{u} - \sqrt{p}) + (c-1)\sqrt{p})^2 \le 2\int (\sqrt{u} - \sqrt{p})^2 + 2(c-1)^2 \\
&\le 2\delta^2 + 2(\sqrt{1+2\delta} - 1)^2 \le 2\delta^2 + 2\delta^2 = 4\delta^2
\end{aligned}
$$

where the last inequality used the fact that $\delta \le 1$. ∎

In light of the above result, we define modified maximum likelihood sieve estimator $\widehat{p}$ to be any $p \in \mathcal{P}$ such that $\widehat{\ell} \le \widehat{p} \le \widehat{u}$. For simplicity, in the rest of the paper, we refer to the modified sieve estimator $\widehat{p}$, simply as the maximum likelihood estimator (mle).

4.2 Outline of Proof

We are now ready to find an estimator $\widehat{M}$ that converges at the optimal rate (up to logarithmic terms.) Our strategy for estimating M has the following steps:

Step 1. We split the data into two halves.

Step 2. Let $\widetilde{Q}$ be the maximum likelihood estimator using the first half of the data. Define $\widetilde{M}$ to be the corresponding manifold. We call $\widetilde{M}$, the pilot estimator. We show that $\widetilde{M}$ is a consistent estimator of M that converges at a sub-optimal rate $a_n = n^{-\frac{2}{D(d+2)}}$. To show this we:

  a. Compute the Hellinger bracketing entropy of Q. (Theorem 9, Lemmas 10 and 11).
  b. Establish the rate of convergence of the mle in Hellinger distance, using the bracketing entropy and Theorem 7.
  c. Relate the Hausdorff distance to the Hellinger distance and hence establish the rate of convergence a_n of the mle in Hausdorff distance. (Lemma 13).
  d. Conclude that the true manifold is contained, with high probability, in $\mathcal{M}_n = \{M \in \mathcal{M}(\kappa) : H(M,\widetilde{M}) \le a_n\}$ (Lemma 14). Hence, we can now restrict attention to $\mathcal{M}_n$.

Step 3. To improve the pilot estimator, we need to control the relationship between Hellinger and Hausdorff distance and thus need to work over small sets on which the manifold cannot vary too greatly. Hence, we cover the pilot estimator with long, thin slabs $R_1,\ldots,R_N$. We do this by first covering $\widetilde{M}$ with spheres $\beth_1,\ldots,\beth_N$ of radius $\delta_n = O((\log n/n)^{1/(2+d)})$. We define a slab R_j to be the union of fibers of size $b = \sigma + a_n$ within one of the spheres: $R_j = \cup_{x \in \beth_j} L_b(x,\widetilde{M})$. We then show that:

  a. The set of fibers on $\widetilde{M}$ cover each $M \in \mathcal{M}_n$ in a nice way. In particular, if $M \in \mathcal{M}_n$ then each fiber from $\widetilde{M}$ is nearly normal to M. (Lemma 15).

    b. As M cuts through a slab, it stays nearly parallel to $\widetilde{M}$. Roughly speaking, M behaves like a smooth, nearly linear function within each slab. (Lemma 16).

Step 4. Using the second half of the data, we apply maximum likelihood within each slab. This defines estimators $\widehat{M}_j$, for $1 \leq j \leq N$. We show that:

    a. The entropy of the set of distributions within a slab is very small. (Lemma 18).

    b. Because the entropy is small, the maximum likelihood estimator within a slab converges fairly quickly in Hellinger distance. The rate is $\varepsilon_n = (\log n/n)^{1/(2+d)}$. (Lemma 19).

    c. Within a slab, there is a tight relationship between Hellinger distance and Hausdorff distance. Specifically, $H(M_1, M_2) \leq c\, h^2(Q_1, Q_2)$. (Lemma 20).

    d. Steps (4b) and (4c) imply that $H(M \cap R_j, \widehat{M}_j) = O_P(\varepsilon_n^2) = O_P((\log n/n)^{2/(d+2)})$.

Step 5. Finally we define $\widehat{M} = \bigcup_{j=1}^{N} \widehat{M}_j$ and show that $\widehat{M}$ converges at the optimal rate because each $\widehat{M}_j$ does within its own slab.

    The reason for getting a preliminary estimator and then covering the estimator with thin slabs is that, within a slab, there is a tight relationship between Hellinger distance and Hausdorff distance. This is not true globally but only in thin slabs. Maximum likelihood is optimal with respect to Hellinger distance. Within a slab, this allows us to get optimal rates in Hausdorff distance.

4.3 Step 1: Data Splitting

For simplicity assume the sample size is even and denote it by $2n$. We split the data into two halves which we denote by $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$.

4.4 Step 2: Pilot Estimator

Let $\widetilde{q}$ be the maximum likelihood estimator over Q. Let $\widetilde{M}$ be the corresponding manifold. To study the properties of $\widetilde{M}$ requires two steps: computing the bracketing entropy of Q and relating $H(M, \widetilde{M})$ to $h(q, \widetilde{q})$. The former allows us to apply Theorem 7 to bound $h(q, \widetilde{q})$, and the latter allows us to control the Hausdorff distance.

4.5 Step 2a: Computing the Entropy of Q

To compute the entropy of Q we start by constructing a finite net of manifolds to cover $\mathcal{M}(\kappa)$. A finite set of d-manifolds $\mathbb{M}_\gamma = \{M_1, \ldots, M_N\}$ is a γ-net (or a γ-cover) if, for each $M \in \mathcal{M}$ there exists $M_j \in \mathbb{M}_\gamma$ such that $H(M, M_j) \leq \gamma$. Let $N(\gamma) = N(\gamma, \mathcal{M}, H)$ be the size of the smallest covering set, called the (Hausdorff) covering number of $\mathcal{M}$.

Theorem 9 *The Hausdorff covering number of $\mathcal{M}$ satisfies the following:*

$$N(\gamma) \equiv N(\gamma, \mathcal{M}, H) \leq c_1 \kappa_2(\kappa, d, D) \exp\left(\kappa_3(\kappa, d, D)\gamma^{-d/2}\right) \equiv c \exp\left(c' \gamma^{-d/2}\right)$$

where $\kappa_2(\kappa, d, D) = \binom{D}{d}^{(c_2/\kappa)^D}$ and $\kappa_3(\kappa, d, D) = 2^{d/2}(D-d)(c_2/\kappa)^D$, for a constant c_2 that depends only on κ and d.

Proof Recall that the manifolds in $\mathcal{M}$ all lie within $\mathcal{K}$. Consider any hypercube containing $\mathcal{K}$. Divide this cube into a grid of $J = (2c/\kappa)^D$ sub-cubes $\{C_1, \ldots, C_J\}$ of side length κ/c, where $c \geq 4$ is a positive constant chosen to be sufficiently large. Our strategy is to show that within each of these cubes, the manifold is the graph of a smooth function. We then only need count the number of such smooth functions.

In thinking about the manifold as (locally) the graph of a smooth function, it helps to be able to translate easily between the natural coordinates in $\mathcal{K}$ and the domain-range coordinates of the function. To that end, within each subcube C_j for $j \in \{1, \ldots, J\}$, we define $K = \binom{D}{d}$ coordinate frames, F_{jk} for $k \in \{1, \ldots, K\}$, in which d out of D coordinates are labeled as "domain" and the remaining $D - d$ coordinates are labeled as "range."

Each frame is associated with a relabeling of the coordinates so that the d "domain" coordinates are listed first and $D - d$ "range" coordinates last. That is, F_{jk} is defined by a one-to-one correspondence between $x \in C_j$ and $(u, v) \in \pi_{jk}(x)$ where $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^{D-d}$ and $\pi_{jk}(x_1, \ldots, x_D) = (x_{i_1}, \ldots, x_{i_d}, x_{j_1}, \ldots, x_{j_{D-d}})$ for domain coordinate indices $i_1 < \ldots < i_d$ and range coordinate indices $j_1 < \ldots < j_{D-d}$.

We define $\text{domain}(F_{jk}) = \{u \in \mathbb{R}^d : \exists v \in \mathbb{R}^{D-d} \text{ such that } (u, v) \in F_{jk}\}$, and let $\mathcal{G}_{jk}$ denote the class of functions defined on $\text{domain}(F_{jk})$ whose second derivative (i.e., second fundamental form) is bounded above by a constant $C(\kappa)$ that depends only on κ. To say that a set $R \subset C_j$ is the graph of a function on a d-dimensional subset of the coordinates in C_j is equivalent to saying that for some frame F_{jk} and some set $A \subset \text{domain}(F_{jk})$, $R = \pi_{jk}^{-1}\{(u, f(u)) : u \in A\}$.

We will prove the theorem by establishing the following claims.

Claim 1. Let $M \in \mathcal{M}$ and C_j be a subcube that intersects M. Then: (i) for at least one $k \in \{1, \ldots, K\}$, the set $M \cap C_j$ is the graph of a function (i.e., single-valued mapping) defined on a set $\mathcal{A} \subset \text{domain}(F_{jk})$, of the form $(u_1, \ldots, u_d) \mapsto \pi_{jk}^{-1}((u, f(u)))$ for some function f on $\mathcal{A}$, and (ii) this function lies in $\mathcal{G}_{jk}$.

Claim 2. $\mathcal{M}$ is in one-to-one correspondence with a subset of $\mathcal{G} = \prod_{j=1}^{J} \bigcup_{k=1}^{K} \mathcal{G}_{jk}$.

Claim 3. The L^∞ covering number of $\mathcal{G}$ satisfies

$$N(\gamma, \mathcal{G}, L^\infty) \leq c_1 \binom{D}{d}^{(2c/\kappa)^D} \exp\left((D-d)(2c/\kappa)^D \gamma^{-d/2}\right).$$

Claim 4. There is a one-to-one correspondence between an $\gamma/2$ L^∞-cover of $\mathcal{G}$ and an γ Hausdorff-cover of $\mathcal{M}$.

Taken together, the claims imply that

$$N(\gamma, \mathcal{M}, H) \leq c_1 \binom{D}{d}^{(2c/\kappa)^D} \exp((D-d)(2c/\kappa)^D 2^{d/2} \gamma^{-d/2}).$$

Taking $c_2 = 2c$ proves the theorem.

Proof of Claim 1. We begin by showing that (i) implies (ii). By part 1 of Lemma 3, each $M \in \mathcal{M}$ has curvature (second fundamental form) bounded above by $1/\kappa$. This implies that the function identified in (i) has uniformly bounded second derivative and thus lies in the corresponding $\mathcal{G}_{jk}$.

We prove (i) by contradiction. Suppose that there is an $M \in \mathcal{M}$ such that for every j with $M \cap C_j \neq \emptyset$, the set $M \cap C_j$ is not the graph of a single-valued mapping for any of the K coordinate frames.

Fix $j \in \{1, \ldots, J\}$. Then in each domain(F_{jk}), there is a point u such that $C_j \cap \pi_{jk}^{-1}(u \times \mathbb{R}^{D-d})$ intersects M in at least two points, call them a_k and b_k. By construction $\|a_k - b_k\| \leq \sqrt{D-d} \cdot \kappa/c$, and hence by choosing c large enough (making the cubes small), part 3 of Lemma 3 tells us that $d_M(a_k, b_k) \leq 2\sqrt{D-d}\kappa/c$. Then we argue as follows:

1. By parts 2 and 3 of Lemma 3 and the fact that C_j has diameter $\sqrt{D}\kappa/c$ and

$$\max_{p,q \in C_j \cap M} \cos(\mathsf{angle}(T_p M, T_q M)) \geq 1 - \frac{2\sqrt{D}}{c}.$$

   For large enough c, the maximum angle between tangent vectors can be made smaller than $\pi/3$.

2. By part 2 of Lemma 3, any point z along a geodesic between a_k and b_k,

$$\cos(\mathsf{angle}(T_{a_k} M, T_z M)) \geq 1 - \frac{2\sqrt{D-d}}{c}.$$

   It follows that there is a point in $C_j \cap M$ and a tangent vector v_k at that point such that $\mathsf{angle}(v_k, b_k - a_k) = O(1/\sqrt{c})$.

3. We have for each of $K = \binom{D}{d}$ coordinate frames and associated tangent vectors $v_1, \ldots, v_K$ that are each nearly orthogonal to at least d of the others. Consequently, there are $\geq d+1$ nearly orthogonal tangent vectors of M within C_j. This contradicts point 1 and proves the claim.

Proof of Claim 2. We construct the correspondence as follows. For each cube C_j, let k_j^* be the smallest k such that $M \cap C_j$ is the graph of a function $\phi_{jk} \in \mathcal{G}_{jk}$ as in Claim 1. Map M to $\varphi = (\phi_{1k_1^*}, \ldots, \phi_{Jk_J^*})$, and let $\mathcal{F} \subset \mathcal{G}$ be the image of this map. If $M \neq M' \in \mathcal{M}$, then the corresponding φ and φ' must be distinct. If not, then $M \cap C_j = M' \cap C_j$ for all j, contradicting $M \neq M'$. The correspondence from $\mathcal{M}$ to $\mathcal{F}$ is thus a one-to-one correspondence.

Proof of Claim 3. From the results in Birman and Solomjak (1967), the set of functions defined on a pre-compact d-dimensional set that take values in a fixed dimension space $\mathbb{R}^m$ with uniformly bounded second derivative has L^∞ covering number bounded above by $c_1 e^{m(1/\gamma)^{d/2}}$ for some c_1. Part 1 of Lemma 3 shows that each $M \in \mathcal{M}$ has curvature (second fundamental form) bounded above by $1/\kappa$, so each $\mathcal{G}_{jk}$ satisfies Birman and Solomjak's conditions. Hence, $N(\gamma, \mathcal{G}_{jk}, L^\infty) \leq c_1 e^{(D-d)(1/\gamma)^{d/2}}$. Because all the $\mathcal{G}_{jk}$'s are disjoint, simple counting arguments show that $N(\gamma, \mathcal{G}, L^\infty) = \left(\binom{D}{d} N(\gamma, \mathcal{G}_{jk}, L^\infty) \right)^J$, where J is the number of cubes defined above. The claim follows. (Note that the functions in Claim 1 are defined on a subset of domain(F_{jk}). But because all such functions have an extension in $\mathcal{G}_{jk}$, a covering of $\mathcal{G}_{jk}$ also covers these functions defined on restricted domains.)

Proof of Claim 4. First, note that if two functions are less than γ distant in L^∞, their graphs are less than γ distant in Hausdorff distance, and vice versa. This implies that a γ L^∞-cover of a set of functions corresponds directly to an γ Hausdorff-cover of the set of the functions' graphs. Hence, in the argument that follows, we can work with functions or graphs interchangeably.

For $k \in \{1, \ldots, K\}$, let $\mathcal{G}_{jk}^\gamma$ be a minimal L^∞ cover of $\mathcal{G}_{jk}$ by $\gamma/2$ balls; specifically, we assume that $\mathcal{G}_{jk}^\gamma$ is the set of centers of these balls. For each $g_{jk} \in \mathcal{G}_{jk}^\gamma$, define $f_{jk}(u) = \pi_{jk}^{-1}(u, g_{jk}(u))$. For every j, choose one such f_{jk}, and define a set $M' = \bigcup_j (C_j \cap \mathsf{range}(f_{jk_j}))$, which is a union of manifolds with boundary that have curvature bounded by $1/\kappa$. That is, such an M' is piecewise

smooth (smooth within each cube) but may fail to satisfy $\Delta(M') \geq \kappa$ globally. Let $\mathcal{A}$ be the collection of M' constructed this way. There are $N(\gamma/2, \mathcal{G}, L^\infty)$ elements in this collection.

By construction and Claim 2, for each $M \in \mathcal{M}$, there exists an $M' \in \mathcal{A}$ such that $H(M, M') \leq \gamma/2$. In other words, the set of $\gamma/2$ Hausdorff balls around the manifolds in $\mathcal{A}$ covers $\mathcal{M}$ but the elements of $\mathcal{A}$ are not themselves necessarily in $\mathcal{M}$. Let $B_H(A, \gamma/2)$ denote the set of all d-manifolds $M \in \mathcal{M}$ such that $H(A, M) \leq \gamma/2$. Let

$$\mathcal{A}_0 = \left\{ A \in \mathcal{A} : \ B_H(A, \gamma/2) \cap \mathcal{M} \neq \emptyset \right\}.$$

For each $A \in \mathcal{A}_0$, choose some $\widetilde{A} \in B_H(A, \gamma/2) \cap \mathcal{M}$. By the triangle inequality, the set $\{\widetilde{A} : A \in \mathcal{A}_0\}$ forms an γ Hausdorff-net for $\mathcal{M}$. This proves the claim. ∎

We are almost ready to compute the entropy. We will need the following lemma.

Lemma 10 *Let $0 < \gamma < \kappa - \sigma$. There exists a constant $K > 0$ (depending only on $\mathcal{K}, \kappa$ and σ) such that, for any $M_1, M_2 \in \mathcal{M}(\kappa)$, $H(M_1, M_2) \leq \gamma$ implies that $|V(M_1 \oplus \sigma) - V(M_2 \oplus \sigma)| \leq K\gamma$. Also, for any $M \in \mathcal{M}(\kappa)$, $|V(M \oplus (\sigma + \gamma)) - V(M \oplus \sigma)| \leq K\gamma$.*

Proof Let $S_j = M_j \oplus \sigma$, $j = 1, 2$. Then, using (3),

$$S_2 \subset M_1 \oplus (\sigma + \gamma) = \bigcup_{u \in M_1} L_{\sigma + \gamma}(u).$$

Hence, uniformly over $\mathcal{M}$,

$$V(S_2) \leq \int_{M_1} \nu_{D-d}(L_{\sigma+\gamma}(u)) d\mu_{M_1} \leq \int_{M_1} \nu_{D-d}(L_\sigma(u)) d\mu_{M_1} + K\gamma = V(S_1) + K\gamma$$

since $\nu_{D-d}(B(u, \sigma + \gamma)) \leq \nu_{D-d}(B(u, \sigma)) + K\gamma$ for some $K > 0$ not depending on M_1 or M_2. By a symmetric argument, $V(S_1) \leq V(S_2) + K\gamma$. Hence, $|V(M_1 \oplus \sigma) - V(M_2 \oplus \sigma)| \leq K\gamma$. The second statement is proved in a similar way. ∎

Now we construct a Hellinger bracketing. Let $\gamma = \varepsilon^2$. Let $\mathbb{M}_\gamma = \{M_1, \ldots, M_N\}$ be a γ-Hausdorff net of manifolds. Thus, by Theorem 9, $N = N(\varepsilon^2, \mathcal{M}, H) \leq c_1 e^{c_2(1/\varepsilon)^d}$. Let ω denote the volume of a sphere of radius σ. Let q_j be the density corresponding to M_j. Define

$$u_j(y) = \left(q_j(y) + \frac{2\varepsilon^2}{V(M_j \oplus (\sigma + \varepsilon^2))} \right) I(y \in M_j \oplus (\sigma + \varepsilon^2))$$

and

$$\ell_j(y) = \left(q_j(y) - \frac{2\varepsilon^2}{V(M_j \oplus (\sigma - \varepsilon^2))} \right) I(y \in M_j \oplus (\sigma - \varepsilon^2)).$$

Let $\mathcal{B} = \{(\ell_1, u_1), \ldots, (\ell_N, u_N)\}$.

Lemma 11 *$\mathcal{B}$ is an ε-Hellinger bracketing of Q. Hence, $\mathcal{H}_{[]}(\varepsilon, Q, h) \leq C(1/\varepsilon)^d$.*

Proof Let $M \in \mathcal{M}(\kappa)$ and let $Q = Q_M$ be the corresponding distribution. Let q be the density of Q. Q is supported on $S = M \oplus \sigma$. There exists $M_j \in \mathbb{M}_\gamma$ such that $H(M, M_j) \leq \varepsilon^2$. Let y be in S. Then there is a $x \in M$ such that $||y - x|| \leq \sigma$. There is a $x' \in M_j$ such that $||x - x'|| \leq \varepsilon^2$. Hence, $d(y, M_j) \leq \sigma + \varepsilon^2$ and thus y is in the support of u_j. Now, for $y \in S$, $u_j(y) - q(y) = 2\varepsilon^2 / V(M_j \oplus (\sigma + \varepsilon^2)) \geq 0$. Hence, $q(y) \leq u_j(y)$. By a similar argument, $\ell_j(y) \leq q(y)$. Thus $\mathcal{B}$ is a bracketing. Now

$$\ell_1(\ell_j, u_j) \;=\; \int u_j - \int \ell_j = \left(1 + \frac{2K\varepsilon^2}{\omega}\right) - \left(1 - \frac{2K\varepsilon^2}{\omega}\right) = \frac{4K\varepsilon^2}{\omega}.$$

Finally, by (1), $h(u_j, \ell_j) \leq \sqrt{\ell_1(\ell_j, u_j)} = C\varepsilon$. Thus $\mathcal{B}$ is a $C\varepsilon$-Hellinger bracketing. ∎

4.6 Step 2b. Hellinger Rate

Lemma 12 *Let $\widetilde{Q}$ be the mle. Then*

$$\sup_{Q \in \mathcal{Q}} Q^n \left(\left\{h(Q, \widetilde{Q}) > C_0 n^{-\frac{1}{d+2}}\right\}\right) \leq \exp\left\{-Cn^{\frac{d}{2+d}}\right\}.$$

Proof We have shown (Lemma 11) that $\mathcal{H}_{[]}(\varepsilon, Q, h) \leq C(1/\varepsilon)^d$. Solving the equation $H_{[]}(\varepsilon_n, Q, h) = n\varepsilon_n^2$ from Theorem 7 we get $\varepsilon_n = (1/n)^{1/(d+2)}$. From Lemma 8, for all Q

$$Q^n \left(\left\{h(Q, \widetilde{Q}) > C_0 n^{-\frac{1}{d+2}}\right\}\right) \leq c_1 e^{-c_2 n\varepsilon_n^2} = \exp\left\{-Cn^{\frac{d}{2+d}}\right\}.$$

∎

4.7 Step 2c. Relating Hellinger Distance and Hausdorff Distance

Lemma 13 *Let $c = (\kappa - \sigma)\sqrt{\pi}C_* / (2\Gamma(D/2 + 1))$. If $M_1, M_2 \in \mathcal{M}(\kappa)$ and $h(Q_1, Q_2) < c$ then*

$$H(M_2, M_2) \leq \left[\frac{2}{\sqrt{\pi}}\left(\frac{\Gamma(D/2 + 1)}{C_*}\right)^{1/D}\right] h^{\frac{1}{D}}(Q_1, Q_2)$$

Proof Let $b = H(M_1, M_2)$ and $\gamma = \min\{\kappa - \sigma, b\}$. Let S_1, S_2 be the supports of Q_1 and Q_2. Because $H(M_1, M_2) = b$, we can find points $x \in M_1$ and $y \in M_2$ such that $||y - x|| = b$. Note that $T_x M_1$ and $T_y M_2$ are parallel, otherwise we could move x or y and increase $||y - x||$. It follows that the line segment $[x, y]$ is along a common normal vector of the two manifolds and we can write $y = x \pm bu$ for some $u \in L_\sigma(u, M)$. Without loss of generality, assume that $y = x + bu$. Let $x' = x + \sigma u$ and $y' = y + \sigma u$. Hence, $x' \in \partial S_1$, $y' \in \partial S_2$ and $||x' - y'|| = b$. Note that ∂S_1 and ∂S_2 are themselves smooth D-manifolds with $\Delta(\partial S_i) \geq \kappa - \sigma > 0$.

We now make the following three claims:

1. $y' \in S_2 - S_1$.

2. $(x', y'] \subset S_2 - S_1$

3. interior $B\left(\frac{x' + y'}{2}, \frac{\gamma}{2}\right) \subset S_2 - S_1$

First, note that y' differs from y along a fiber of M_2 by exactly σ, therefore $[x', y'] \subset S_2$. Second, because $x' \in \partial S_1$, there is a neighborhood of x' in $[x', y']$ that is not contained in S_1. Hence, if there is a point in $S_1 \cap [x', y']$ there must be a point $z' \in \partial S_1 \cap [x', y']$, with $z' \neq x'$. This implies the existence of two distinct points whose fibers of length less than $\kappa - \sigma$ cross, which contradicts the fact that $\Delta(\partial S_1) \geq \kappa - \sigma$. Claims 1 and 2 follows.

Let $B = B\left(\frac{x'+y'}{2}, \frac{\gamma}{2}\right)$. By construction, B is tangent to ∂S_1 at x' and tangent to ∂S_2 at y', and B contains $[x', y']$. The ball has radius $\gamma/2 = (1/2)\min\{\kappa - \sigma, b\} < \kappa - \sigma$. Because B intersects $S_2 - S_1$, the interior of B cannot intersect either ∂S_1 or ∂S_2. Claim 3 follows by a similar argument as in the proof of Claim 2. (In particular, if there were a point in the interior of B that is either in S_1 or outside S_2, a line segment from $(x'+y')/2$ to that point would have to intersect the corresponding boundary, which cannot happen.)

Now $V(B) = (\gamma/2)^D \pi^{D/2} / \Gamma(D/2+1)$. So

$$
\begin{aligned}
h(Q_1, Q_2) &\geq \ell_1(Q_1, Q_2) = \int |q_1 - q_2| \geq \int_{S_1 \cap S_2^c} |q_1 - q_2| \\
&= \int_{S_1 \cap S_2^c} q_1 = Q_1(S_1 \cap S_2^c) \geq C_* V(S_1 \cap S_2^c) = C_*(\gamma/2)^D \pi^{D/2} / \Gamma(D/2+1).
\end{aligned}
$$

Hence,

$$
\gamma = \min\{\kappa - \sigma, b\} \leq \left[\frac{2}{\sqrt{\pi}} \left(\frac{\Gamma(D/2+1)}{C_*} \right)^{1/D} \right] h^{1/D}(Q_1, Q_2).
$$

If $\kappa - \sigma \leq b$ this implies that $h(Q_1, Q_2) > c$ which contradicts the assumption that $h(Q_1, Q_2) < c$. Therefore, $\gamma = b$ and the conclusion follows. ∎

4.8 Step 2d. Computing The Hausdorff Rate of the Pilot

Lemma 14 *Let $a_n = \left(\frac{C_0}{n}\right)^{\frac{2}{D(d+2)}}$. For all large n,*

$$
\sup_{Q \in \mathcal{Q}} Q^n \left(\{H(M, \widetilde{M}) > a_n\} \right) \leq \exp\left\{ -Cn^{\frac{d}{2+d}} \right\}.
$$

Proof Follows by combining Lemma 12 and Lemma 13. ∎

We conclude that, with high probability, the true manifold M is contained in the set $\mathcal{M}_n = \Big\{ M \in \mathcal{M}(\kappa) : H(\widetilde{M}, M) \leq a_n \Big\}$.

4.9 Step 3: Cover With Slabs

Now we cover the pilot estimator $\widetilde{M}$ with (possibly overlapping) slabs. Let $\delta_n = \left(\frac{C \log n}{n}\right)^{\frac{1}{2+d}}$. It follows from part 6 of Lemma 3 that there exists a collection of points $F = \{x_1, \ldots, x_N\} \subset \widetilde{M}$, such that $N = (c\delta_n)^{-d} = (Cn/\log n)^{d/(2+d)}$ and such that $\widetilde{M} \subset \bigcup_{j=1}^N B_D(x_j, c\delta)$.
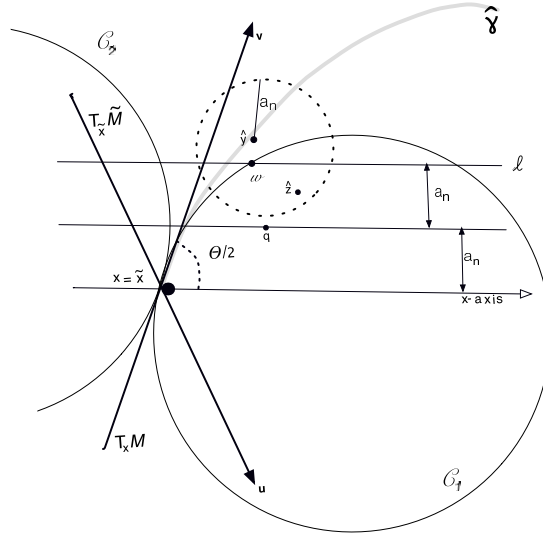
Figure 4: Figure for the proof of part 1 of Lemma 15.

4.10 Step 3a. The Fibers of $\widetilde{M}$ Cover M Nicely

Lemma 15 *Let* $b = \sigma + a_n$. *For* $\widetilde{x} \in \widetilde{M}$, *let* $L_b(\widetilde{x}) = T_{\widetilde{x}}^{\perp}\widetilde{M} \cap B_D(\widetilde{x},b)$ *be a fiber at* $\widetilde{x}$ *of size b. Let* $M \in \mathcal{M}_n$. *Then:*

1. *If* $\widetilde{x} \in \widetilde{M}$ *and* $x \in M$ *are such that* $\|x - \widetilde{x}\| \leq a_n$, *then* $\mathsf{angle}(T_xM, T_{\widetilde{x}}\widetilde{M}) < \pi/4$.

2. $L_b(\widetilde{x}) \cap M \neq \emptyset$.

3. *If* $x \in L_b(\widetilde{x}) \cap M$, *then* $\|x - \widetilde{x}\| \leq 2a_n$.

4. *For any* $\widetilde{x} \in \widetilde{M}$, $\#\{L_b(\widetilde{x}) \cap M\} = 1$.

5. *We have* $M \subset \bigcup_{\widetilde{x} \in \widetilde{M}} L_b(\widetilde{x})$.

Proof *1.* Let x and $\widetilde{x}$ be as given in the statement of the lemma and let $\theta = \mathsf{angle}(T_xM, T_{\widetilde{x}}\widetilde{M})$. Suppose that $\theta \geq \pi/4$. There exists unit vectors $u \in T_{\widetilde{x}}\widetilde{M}$ and $v \in T_xM$ such that $\mathsf{angle}(u,v) = \theta$. Without loss of generality, we can assume that $x = \widetilde{x}$. (The extension to the case $x \neq \widetilde{x}$ is straightforward.)

Consider the plane defined by u and v as in Figure 4. We assume, without loss of generality, that $(u+v)/2$ generates the x-axis in this plane and that v lies above the x-axis and u lies below the x axis. Let ℓ denote the horizontal line, parallel to the x-axis and lying $2a_n$ units above the horizontal axis. Hence, u and v each make an angle greater than $\pi/8$ with respect to the x-axis.

Consider the two circles C_1 and C_2 tangent to M at x with radius κ where C_1 lies below v and C_2 lies above v. Let w be the point at which C_1 intersects ℓ. The arclength of C_1 from x to w is Ca_n for some $C > 1$. Let γ be the geodesic on M through x with gradient v. The projection $\widehat{\gamma}$ of γ into the plane must fall between C_1 and C_2. Let $y = \gamma(Ca_n)$ and $\widehat{y}$ be the projection of y into the plane.

Now $\|y - \widetilde{x}\| \geq \|\widehat{y} - \widetilde{x}\| \geq \|w - \widetilde{x}\| \geq 2a_n > a_n$. There exists $\widetilde{z} \in \widetilde{M}$ such that $\|\widetilde{z} - y\| \leq a_n$. Hence, $\|\widehat{z} - \widehat{y}\| \leq a_n$ where $\widehat{z}$ is the projection of $\widetilde{z}$ into the plane. Let q be the point on the plane

with coordinates $(a_n\sqrt{C^2-1}, a_n)$. Thus, $||q-\widetilde{x}|| = Ca_n$. Note that $\mathsf{angle}(\widehat{z}-\widetilde{x}, u)$ is larger than the angle between $q-\widetilde{x}$ and the x-axis which is $\arctan\left(\frac{1}{\sqrt{C^2-1}}\right) \equiv \alpha > 0$. Hence,

$$\mathsf{angle}(\widetilde{z}-\widetilde{x}, u) \geq \mathsf{angle}(\widehat{z}-\widetilde{x}, u) \geq \alpha.$$

Let $\widetilde{\gamma}$ be a geodesic on $\widetilde{M}$, parameterized by arclength connecting $\widetilde{x}$ and $\widetilde{z}$. Thus $\widetilde{\gamma}(0) = \widetilde{x}$ and $\widetilde{\gamma}(T) = \widetilde{z}$ for some T. There exists some $0 \leq t \leq T$ such that $\gamma'(t) \propto \widetilde{z}-\widetilde{x}$. So

$$\mathsf{angle}(\gamma'(t), \gamma'(0)) = \alpha > 0.$$

However, $||\widetilde{z}-\widetilde{x}|| \leq (C+1)a_n$ which implies, by part 2 of Lemma 3, that $\mathsf{angle}(\gamma'(t), \gamma'(0)) = O(\sqrt{a_n}) < \alpha$ which is a contradiction.

2. For any $\widetilde{x} \in \widetilde{M}$, the closest point $x \in M$ must satisfy $||x-\widetilde{x}|| \leq a_n$. Let y be the projection of x onto $T_{\widetilde{x}}\widetilde{M}$. Let $U = T_{\widetilde{x}}\widetilde{M} \cap B_d(y, a_n)$. Let $\mathrm{Cyl} = \bigcup_{u \in U} B_D(u, 3a_n) \cap \left(T_{\widetilde{x}}\widetilde{M}\right)^{\perp}$. Cyl is a small hyper-cylinder containing y and $\widetilde{x}$, with the former in the center. M cannot intersect the top or bottom faces of the cylinder. Otherwise, we can find a point $p \in M$ such that $\mathsf{angle}(T_{\widetilde{x}}\widetilde{M}, T_p M) > \arctan(1) = \pi/4$ contradicting 1. Thus, any path through x on M must intersect the sides of Cyl. Hence, $L_b(\widetilde{x}) \cap M \neq \emptyset$.

3. Let $x \in M \cap L_b(\widetilde{x})$. Suppose that $||x-\widetilde{x}|| > 2a_n$. There exists $q \in \widetilde{M}$ such that $||q-x|| \leq a_n$. Note that $||q-\widetilde{x}|| > a_n$. Now we apply part 5 Lemma 3 with $p = \widetilde{x}$ and $v = x$. This implies that $||v-p|| = ||x-\widetilde{x}|| < a_n^2/\kappa$ which contradicts the assumption that $||x-\widetilde{x}|| > 2a_n$.

4. Suppose that more than one point of M were in $L_b(\widetilde{x})$. Pick two and call them x_1 and x_2. By 3, $||x_i - \widetilde{x}|| \leq 2a_n$. It follows that $||x_1 - x_2|| \leq 4a_n$ and thus they are $O(a_n)$ close in geodesic distance by part 3 of Lemma 3. Hence, there is a geodesic on M connecting x_1 and x_2 that is contained strictly within the Ca_n ball. Because $x_2 - x_1$ lies in $L_b(\widetilde{x})$ and is consequently orthogonal to $T_{\widetilde{x}}\widetilde{M}$, there must exist a point on the geodesic whose angle with $T_{\widetilde{x}}\widetilde{M}$ equals $\pi/2$, contradicting part 1.

5. Because $H(\widetilde{M}, M) \leq a_n$, we have that $M \subset \mathrm{tube}(\widetilde{M}, a_n)$. Because $a_n < \kappa$, the fibers $L_b(\widetilde{x})$ partition $\mathrm{tube}(\widetilde{M}, a_n)$. Hence, each $x \in M$ must lie on one (and only one) $L_b(\widetilde{x})$. ∎

4.11 Step 3b. Construct Slabs that Cover M Nicely

Let $\beth_j = B_D(x_j, \delta_n) \cap \widetilde{M}$. Define the slab

$$R_j = \bigcup_{x \in \beth_j} L_b(x, \widetilde{M}).$$

Lemma 16 *The collection of slabs $R_1, \ldots, R_N$ has the following properties. Let $M \in \mathcal{M}_n$.*

1. *$M \subset \bigcup_{j=1}^{N} R_j$.*

2. *$M \cap R_j$ is function-like over R_j. That is, there exists a function $g_j : \beth_j \to \mathbb{R}^{D-d}$ such that $M \cap R_j = \{g_j(x) : x \in \beth_j\}$.*

3. *For each $x \in \beth_j$, $L_b(x) \cap M \neq \emptyset$.*

4. *There exists a linear function $\ell_j : \beth_j \to \mathbb{R}^{D-d}$ such that $\sup_{x \in \beth_j} ||g_j(x) - \ell_j(x)|| \leq C\delta_n^2$.*

5. $\sup_{M \in \mathcal{M}_n} \mathrm{diam}(M \cap R_j) \leq C\delta_n$.

Thus the slabs cover M and M cuts across R_j is a function-like way. Moreover, $M \cap R_j$ is nearly linear.

Proof The first three claims follow immediately from Lemma 15. In particular, g_j in claim *2* is defined by $g_j(x) = \{M \cap L_b(x)\}$. Now we show *4*. We can write $g_j(x) = g_j(x_j) + (x - x_j)^T \nabla g + \frac{1}{2}(x - x_j)^T \mathrm{Hess}\,(x - x_j)$ where Hess is the Hessian matrix of g_j evaluated at some point between x and x_j. By part 1 of Lemma 3, the largest eigenvalue of Hess is bounded above by $1/\kappa$. Since $\|x - x_j\| \leq c\delta_n^2$, the claim follows. Part *5* follows easily. ∎

4.12 Step 4: Local Conditional Likelihood

Recall that $\mathcal{M}_n = \{M \in \mathcal{M}(\kappa) : H(\widetilde{M}, M) \leq a_n\}$. Let

$$Q_n = \{Q_M : M \in \mathcal{M}_n\}.$$

Consider a slab R_j. For each $Q \in Q_n$ define $Q_j \equiv Q(\cdot | R_j)$ by $Q_j(A) = Q(A \cap R_j)/Q(R_j)$. Note that Q_j is supported over $\mathrm{tube}(M, \sigma) \cap R_j$. Let $Q_{n,j} = \{Q_j : Q \in Q_n\}$. Before we proceed we need to establish the following.

Lemma 17 *Let* $I_j(M) = \mathrm{tube}(M, \sigma) \cap R_j$. *Then there exists* $c_0 > 0$ *such that*

$$\inf_{M \in \mathcal{M}_n} V(I_j(M)) \geq c_0 \delta_n^d.$$

Proof By Lemma 16, $M \cap R_j$ lies in a slab of size a_n orthogonal to $\beth_j$. Because the angle between the two manifolds on this set must be no more than $\pi/4$ and because $a_n > \delta_n$, the manifold M cannot intersect both the "top" and "bottom" surfaces of the slab. Hence, for large enough $C > 0$, $\mathcal{I}_j = \bigcup_{x \in \beth_j} B_D(x, \sigma/C) \subset I_j$. By construction, $V(I_j) \geq V(\mathcal{I}_j) \geq c\delta_n^d$. ∎

4.13 Step 4a. The Entropy of $Q_{n,j}$

Lemma 18 $\mathcal{H}_{[]}(\varepsilon, Q_{n,j}, h) \leq c_1 \log(c_2/\varepsilon)$.

Proof We begin by creating a γ Hausdorff net for $Q_{n,j}$. To do this, we will parameterize the support of these distributions. Each $Q \in Q_{n,j}$ has support in the collection $\mathcal{S}_{n,j} = \{(M \oplus \sigma) \cap R_j : M \in \mathcal{M}_n\}$. We will construct a γ-Hausdorff net for $\mathcal{S}_{n,j}$.

Let $\widetilde{x} \in \widetilde{M}$ be the center of $\beth_j$. Let $y_1, \ldots, y_r$ be a $c_1\gamma$-net of $L_b(\widetilde{x})$, and let $\theta_1 < \theta_2 < \cdots < \theta_s < \pi/2 - \eta$ for a small, fixed $\eta > 0$ where $\theta_j - \theta_{j-1} \leq c_2\gamma$. Note that $r = O(\gamma^{-(D-d)})$ and $s = O(1/\gamma)$. For every pair y_i and θ_j, let M_{ij} be a $M \in \mathcal{M}_n$ that crosses through y_i with $\mathrm{angle}(T_{y_i}M, T_{\widetilde{x}}\widetilde{M}) = \theta_j$. These manifolds comprise a collection of size $O((1/\gamma)^{D-d-1})$ which we will denote by $\mathrm{Net}(\gamma)$.

Let $M \in \mathcal{M}_n$. Let y be the point where M crosses $L_b(\widetilde{x})$. Let y_i be the closest point in the net to y and let θ_j be the closest angle in the net to $\mathrm{angle}(T_y M, T_{\widetilde{x}}\widetilde{M})$. Because the angle between M and M_{ij} is strictly less than $\pi/4$ (part 1 of Lemma 15) and the slab R_j has radius δ_n, it follows that $H(M, M_{ij}) \leq C_1\gamma + \delta_n C_2\gamma \leq C\gamma$. Hence, $\mathrm{Net}(\gamma)$ is a γ-Hausdorff net.

Now consider $\mathsf{Net}(\gamma)$ with $\gamma = \varepsilon^2$. For each $M_{ij} \in \mathsf{Net}(\gamma)$ let q_{ij} be the corresponding density and define u_{ij} and ℓ_{ij} by

$$u_{ij}(y) = \left(q_{ij}(y) + \frac{C\varepsilon^2}{V(M_{ij} \oplus (\sigma + \varepsilon^2))} \right) I(y \in M_{ij} \oplus (\sigma + \varepsilon^2))$$

and

$$\ell_{ij}(y) = \left(q_{ij}(y) - \frac{C\varepsilon^2}{V(M_{ij} \oplus (\sigma - \varepsilon^2))} \right) I(y \in M_j \oplus (\sigma - \varepsilon^2)).$$

Let $\mathcal{B} = \{(\ell_{ij}, u_{ij})\}$.

Let $M \in \mathcal{M}_n$ and let M_{ij} be the element of the net closest to M. It follows easily that $u_{ij} \geq q_M \geq \ell_{ij}$. Thus $\mathcal{B}$ is a bracketing. Now,

$$\int u_{ij} - \ell_{ij} = 1 + C\varepsilon^2 - (1 - C\varepsilon^2) = 2C\varepsilon^2.$$

Hence, $h(u_{ij}, \ell_{ij}) \leq \sqrt{\int |u_{ij} - \ell_{ij}|} = \sqrt{2C}\varepsilon$. Hence, $\mathcal{B}$ is an $\sqrt{2C} - \varepsilon$-bracketing. So,

$$\mathcal{H}_{[]}(\varepsilon, Q_{n,j}, h) \leq (D - d - 1)\log(c/\varepsilon),$$

which proves the lemma. ∎

4.14 Step 4b. Hellinger Rate of the Conditional MLE

Let $\widehat{q}$ be the mle over $Q_{n,j}$ using the Y_i's in R_j. Let $\widehat{M}$ be the manifold corresponding to $\widehat{q}$ and let $\widehat{M}_j = \widehat{M} \cap R_j$.

Lemma 19 *For all Q, all A > 0 and all large n,*

$$Q^n \left(\left\{ h(Q, \widehat{Q}) > \left(\frac{C_0 \log n}{n} \right)^{\frac{1}{2+d}} \right\} \right) \leq n^{-A}.$$

Proof Let N_j be the number of observations from the second half of the data that are in R_j. Let $\mu_j = \mathbb{E}(N_j)$ and define $m_n = n^{\frac{2}{2+d}}$. First, we claim that $N_j \geq \mu_j/2 = O(m_n)$ for all j, except on a set of probability $e^{-cn^{2/(2+d)}}$. Let $\pi_j = Q(R_j)$. By Lemma 17 and Lemma 4, $\pi_j \geq c\delta_n^d$ for some $c > 0$. Hence, $\mu_j \geq m_n$. Note that $\sigma^2 \equiv \mathsf{Var}(N_j)/n = \pi_j(1 - \pi_j) \leq \pi_j$. Let $t = \mu_j/2$. By Bernstein's inequality,

$$\mathbb{P}(N_j \leq \mu_j/2) = \mathbb{P}(N_j - \mu_j \leq -\mu_j/2) \leq \exp\left\{ -\frac{t^2}{2n\sigma^2 + 2t/3} \right\} \leq \exp\left\{ -cn^{2/(2+d)} \right\}.$$

Hence, by the union bound,

$$\mathbb{P}(N_j \leq \mu_j/2 \text{ for some } j) \leq \frac{1}{N} \exp\left\{ -cn^{2/(2+d)} \right\} \leq \exp\left\{ -c'n^{2/(2+d)} \right\}$$

since there are $N = O(1/\delta_n)$ slabs. Thus we can assume that there are at least order m_n observations in each R_j.

Since $\mathcal{H}_{[]}(\varepsilon, Q_{v,j}, h) \leq \log(C(1/\varepsilon))$, solving the equation $\mathcal{H}_{[]}(\varepsilon, Q_{v,j}, h) = m_n \varepsilon^2$ we get $\varepsilon_m \geq \sqrt{C \log m_n / m_n} = (\log n / n)^{2/(2(2+d))} = \delta_n$. From Lemma 8, we have, for all $Q \in Q_{n,j}$,

$$Q^n\left(\left\{h(Q,\widehat{Q}) > \delta_n\right\}\right) = Q^n\left(\left\{h(Q,\widehat{Q}) > \varepsilon_m\right\}\right) \leq c_1 e^{-c_2 m_n \varepsilon_m^2} \leq n^{-A}.$$

■

4.15 Step 4c. Relating Hausdorff Distance to Hellinger Distance Within a Slab

Lemma 20 *For each $M_1, M_2 \in \mathcal{M}_n$, $H(M_1 \cap R_j, M_2 \cap R_j) \leq C h^2(Q_{j1}, Q_{j2})$.*

Proof Let g_1 and g_2 be defined as in Lemma 16. There exists $x \in \beth_j$ such that $g_1(x) \in M_1$, $g_2(x) \in M_2$ and $\|g_1(x) - g_2(x)\| = \gamma$. We claim there exists $\beth' \subset \beth_j$ such that $\inf_{x \in \beth'} \|g_1(x) - g_2(x)\| \geq \gamma/2$ and such that $V(\beth') \geq c\delta_n^d$. This follows since g_1 and g_2 are smooth, they both lie in a slab of size a_n around $\beth_j$ and the angle between the tangent of $g_j(x)$ and $\beth_j$ is bounded by $\pi/4$.

Create a modified manifold M_2' such that M_2' differs from M_1 over $\beth'$ by a $\gamma/2$ shift orthogonal to $\beth_j$ and such that M_2' is otherwise equal to M_1. It follows that $\ell_1(M_1, M_2) \geq \ell_1(M_1, M_2')$ and $h(Q_1, Q_2) \geq h(Q_1, Q_2')$.

Every point in the support of the conditioned distributions can be written as an ordered pair (x, y) where $x \in \beth_j$ and y lies in a d' ball of radius σ. M_2' is shifted a distance of $\gamma/2$ in the direction orthogonal to $\beth_j$. As a result, the ℓ_1 distance between M_1 and M_2' equals the integral over C' of the volume difference between two d' balls of the same radius that are shifted by $\gamma/2$ relative to each other. This volume $\delta_n^d \gamma$. Hence, $V(M_1 \cap \beth_j) \circ (M_2 \cap \beth_j) \geq \gamma \delta_n^d$. Let $A = \{x \in \beth_j : q_1 > 0, q_2 = 0\}$, $B = \{x \in \beth_j : q_1 > 0, q_2 > 0\}$, $C = \{x \in \beth_j : q_1 = 0, q_2 > 0\}$. At least one of A or B has volume at least $\gamma \delta_n^d / 2$. Without loss of generality, assume that it is A. Then

$$
\begin{aligned}
h^2(q_1, q_2) &= \int (\sqrt{q_1} - \sqrt{q_2})^2 \geq \int_A (\sqrt{q_1} - \sqrt{q_2})^2 = \int_A q_1 \\
&\geq \frac{C_* c \delta_n^d \gamma}{\delta_n^d} = c C_* \gamma = c C_* H(M_1, M_2).
\end{aligned}
$$

■

4.16 Step 4d. The Hausdorff Rate

Lemma 21 *For any $A > 0$ there exists C_0 such that*

$$Q^n\left(\left\{H(M \cap R_j, \widehat{M}_j) > \left(\frac{C_0 \log n}{n}\right)^{\frac{2}{2+d}}\right\}\right) \leq \frac{1}{n^A}.$$

Proof This follows by combining Lemma 20 and Lemma 19.

■

4.17 Step 5: Final Estimator

Now we can combine the estimators from the difference slabs. Let $\widehat{M} = \bigcup_{j=1}^{N} \widehat{M}_j$. Recall that the number of slabs is $N = (c\delta_n)^{-d} = (Cn/\log n)^{d/(2+d)}$.

Proof of Theorem 2. Choose an $A > 2/(2+d)$. We have:

$$
\begin{aligned}
Q^n\left(\left\{H(\widehat{M},M) > \left(\frac{C_0 \log n}{n}\right)^{\frac{2}{2+d}}\right\}\right) &\leq \sum_j Q^n\left(\left\{H(\widehat{M}_j, M \cap R_j) > \left(\frac{C_0 \log n}{n}\right)^{\frac{2}{2+d}}\right\}\right) \\
&\leq \frac{N}{n^A} \\
&= \left(\frac{n}{C \log n}\right)^{\frac{1}{2+d}} \times \frac{1}{n^A} \leq \frac{c}{n^A}.
\end{aligned}
$$

Let $r_n = \left(\frac{C_0 \log n}{n}\right)^{2/(2+d)}$. Since M and $\widehat{M}$ are contained in a compact set, $H(\widehat{M},M)$ is uniformly bounded above by a constant K_0. Hence,

$$
\begin{aligned}
\mathbb{E}_Q H(\widehat{M},M) &= \mathbb{E}_Q[H(\widehat{M},M)I(H(\widehat{M},M) > r_n)] + \mathbb{E}_Q[H(\widehat{M},M)I(H(\widehat{M},M) \leq r_n)] \\
&\leq K_0\, Q^n(H(\widehat{M},M) > r_n) + r_n \\
&\leq \frac{c}{n^A} + r_n = O\left(\left(\frac{\log n}{n}\right)^{2/(2+d)}\right).
\end{aligned}
$$

∎

5. A Simple, Consistent Estimator

Here we give a practical, consistent estimator, one that does not converge at the optimal rate. It is a generalization of the estimator in Genovese et al. (2010) and is similar to the estimator in Niyogi et al. (2006). Let

$$
\widehat{S} = \bigcup_{i=1}^{n} B_D(Y_i, \varepsilon)
$$

and define $\widehat{\partial S} = \partial(\widehat{S})$, $\widehat{\sigma} = \max_{y \in \widehat{S}} d(y, \widehat{\partial S})$ and

$$
\widehat{M} = \left\{y \in \widehat{S}: \; d(y, \widehat{\partial S}) \geq \widehat{\sigma} - 2\varepsilon\right\}.
$$

Lemma 22 *Let $\varepsilon_n = C(\log n/n)^{1/D}$ in the estimator $\widehat{M}$. Then*

$$
H(M, \widehat{M}) = O\left(\frac{\log n}{n}\right)^{1/D}
$$

almost surely for all large n.

Before proving the lemma we need a few definitions. Following Cuevas and Rodríguez-Casal (2004), we say that a set S is (χ, λ)-*standard* if there exist positive numbers χ and λ such that

$$
\nu_D(B_D(y, \varepsilon) \cap S) \geq \chi\, \nu_D(B(y, \varepsilon)) \quad \text{for all } y \in S, \; 0 < \varepsilon \leq \lambda.
$$

We say that S is *partly expandable* if there exist $r > 0$ and $R \geq 1$ such that $H(\partial S, \partial(S \oplus \varepsilon)) \leq R\varepsilon$ for all $0 \leq \varepsilon < r$. A standard set has no sharp peaks while a partly expandable set has not deep inlets.

Lemma 23 *If $\sigma < \Delta(M)$ then $S = M \oplus \sigma$ is standard with $\chi = 2^{-D}$ and $\lambda = \sigma$ and partly expandable with $r = \Delta(M) - \sigma$ and $R = 1$.*

Proof Let $\chi = 2^{-D}$. Let y be a point in S and let $\Lambda(y) \leq \sigma$ be its distance from the boundary ∂S. If $\Lambda(y) \geq \varepsilon$ then $B_D(y, \varepsilon) \cap S = B_D(y, \varepsilon)$ so that $\nu_D(B_D(y, \varepsilon) \cap S) = \nu_D(B_D(y, \varepsilon)) \geq \chi \nu_D(B_D(y, \varepsilon))$.

Suppose that $\Lambda(y) < \varepsilon$. Let v be a point on the manifold closest to y and let y^* be the point on the segment joining y to v such that $||y - y^*|| = \varepsilon/2$. The ball $A = B_D(y^*, \varepsilon/2)$ is contained in both $B_D(y, \varepsilon)$ and S. Hence, $\nu_D(B_D(y, \varepsilon) \cap S) \geq \nu_D(A) \geq \chi \nu_D(B_D(y, \varepsilon))$. This is true for all $\varepsilon \leq \sigma$, hence S is (χ, λ)-standard for $\chi = 1/2^D$ and $\lambda = \sigma$.

Now we show that S is partly expandable. By Proposition 1 in Cuevas and Rodríguez-Casal (2004) it suffices to show that a ball of radius r rolls freely outside S for some r, meaning that, for each $y \in \partial S$, there is an a such that $y \in B(a, r) \subset \overline{S^c}$, where S^c is the complement of S. Let O_y be the ball of radius $\Delta - \sigma$ tangent to y such that $O_y \subset S^c$. Such a ball exists by virtue of the fact that $\sigma < \Delta(M)$. ∎

Theorem 24 (Cuevas and Rodríguez-Casal, 2004) *Let $Y_1, \ldots, Y_n$ be a random sample from a distribution with support S. Let S be compact, (λ, χ)-standard and partly expandable. Let*

$$\widehat{S} = \bigcup_{i=1}^{n} B(Y_i, \varepsilon_n)$$

and let $\widehat{\partial S}$ be the boundary of $\widehat{S}$. Let $\varepsilon_n = C(\log n/n)^{1/D}$ with $C > (2/(\chi \, \omega_D))^{1/D}$ where $\omega_D = V(B_D(0,1))$. Then, with probability one,

$$H(S, \widehat{S}) \leq C \left(\frac{\log n}{n} \right)^{1/D} \quad \text{and} \quad H(\partial S, \widehat{\partial S}) \leq C \left(\frac{\log n}{n} \right)^{1/D}$$

for all large n. Also, $S \subset \widehat{S}$ almost surely for all large n.

Proof of Lemma 22. Theorem 24 and Lemma 23 imply that $H(S, \widehat{S}) \leq C(\log n/n)^{1/D}$ and $H(\partial S, \widehat{\partial S}) \leq C(\log n/n)^{1/D}$. It follows that $\widehat{\sigma} \geq \sigma - \varepsilon$. First we show that $y \in \widehat{M}$ implies that $d(y, M) \leq 4\varepsilon$. Let $y \in \widehat{M}$. Then $d(y, \partial S) \geq d(y, \widehat{\partial S}) - \varepsilon \geq \widehat{\sigma} - 2\varepsilon - \varepsilon \geq \sigma - \varepsilon - 2\varepsilon - \varepsilon = \sigma - 4\varepsilon$. So $d(y, M) = \sigma - d(y, \partial S) \leq \sigma - \sigma + 4\varepsilon = 4\varepsilon$. Now we show that $M \subset \widehat{M}$. Suppose that $y \in M$. Then,

$$d(y, \widehat{\partial S}) \geq d(y, \partial S) - \varepsilon = \sigma - \varepsilon \geq \widehat{\sigma} - 2\varepsilon$$

so that $y \in \widehat{M}$. ∎

6. Conclusion and Open Questions

We have established that the optimal rate for estimating a smooth manifold in Hausdorff distance is $n^{-\frac{2}{2+d}}$. We conclude with some comments and open questions.

1. We have assumed that the noise is perpendicular to the manifold. In current work we are deriving the minimax rate under the more general assumption that ε is drawn from a general,

spherically symmetric distribution. We also allow the distribution along the manifold to be any smooth density bounded away from 0. The rates are quite different and the methods for proving the rates are substantially more involved. Moreover, the rates depends on the behavior of the noise density near the boundary of its support. We will report on this elsewhere.

2. Perhaps the most important open question is to find a computationally tractable estimator that achieves the optimal rate. It is possible that combining the estimator in Section 5 with one of the estimators in the computational geometry literature (Dey, 2006) could work. However, it appears that some modification of such an estimator is needed. This is a difficult question which we hope to address in the future.

3. It is interesting to note that Niyogi et al. (2006) have a Gaussian noise distribution. While it is possible to infer the homology of M with Gaussian noise it is not possible to infer M itself with any accuracy. The reason is that manifold estimation is similar to (and in fact, more difficult than) nonparametric regression with measurement error. In that case, it is well known that the fastest possible rates under Gaussian noise are logarithmic. This highlights an important distinction between estimating the topological structure of M versus estimating M in Hausdorff distance.

4. The current results take $\Delta(M)$, d and σ as known (or at least bounded by known constants). In practice these must be estimated. We do not know whether there exist minimax estimators that are adaptive over $d, \Delta(M)$ and σ.

Acknowledgments

7. Appendix

This appendix contains proofs of some technical results used earlier in the paper.

7.1 Proof of Equation 2

We will use the following two results (see Section 2.4 of Tsybakov, 2008):

$$h^2(P^n, Q^n) = 2\left(1 - \left[1 - \frac{h^2(P,Q)}{2}\right]^n\right)$$

and

$$\|P \wedge Q\| \geq \frac{1}{2}\left(1 - \frac{h^2(P,Q)}{2}\right)^2.$$

We have

$$
\begin{aligned}
\|P^n \wedge Q^n\| &\geq \frac{1}{2}\left(1 - \frac{h^2(P^n, Q^n)}{2}\right)^2 = \frac{1}{2}\left(1 - \frac{h^2(P,Q)}{2}\right)^{2n} \\
&\geq \frac{1}{2}\left(1 - \frac{\ell_1(P,Q)}{2}\right)^{2n}
\end{aligned}
$$

since $h^2(P,Q) \le \ell_1(P,Q)$.

7.2 Proof of Theorem 6

We define two manifolds M_0 and M_1 with corresponding distributions Q_0 and Q_1 such that (i) $\Delta(M_i) \ge \kappa$ $i = 0,1$, (ii) $H(M_0,M_1) = \gamma$ and (iii) such that the volume of $S_0 \circ S_1$ is of order $\gamma^{\frac{d}{2}+1}$, where S_i is the support of Q_i.

We write a generic D-dimensional vector as $y = (u,v,z)$, with $u \in \mathbb{R}^d$, $v \in \mathbb{R}$, $z \in \mathbb{R}^{D-d-1}$. For each $u \in \mathbb{R}^d$ with $||u|| \le 1$, define the disk in $\mathbb{R}^{d+1}$

$$D_0 = \left\{ (u,0) \in \mathbb{R}^{d+1} : u \in B_d(0,1) \right\}$$

and let

$$F_0 = \partial \left(\bigcup_{(u,v) \in D_0} B_{d+1}((u,v),\kappa) \right).$$

Now define the following d-dimensional manifold in $\mathbb{R}^D$

$$
\begin{aligned}
M_0 &= \left\{ (u,v,0_{D-d-1}) : (u,v) \in F_0 \right\} \\
&= \left\{ (u,a(u),0_{D-d-1}) : u \in B_d(0,1+\kappa) \right\} \cup \left\{ (u,-a(u),0_{D-d-1}) : u \in B_d(0,1+\kappa) \right\}
\end{aligned}
$$

where

$$a(u) = \begin{cases} \kappa & \text{if } ||u|| \le 1 \\ \sqrt{\kappa^2 - (||u|| - 1)^2} & \text{if } 1 < ||u|| \le 1+\kappa. \end{cases}$$

The manifold M_0 has no boundary and, by construction, $\Delta(M_0) \ge \kappa$.

Now define a second manifold that coincides with M_0 but has a small perturbation. Let $\gamma \in (0,4\kappa)$ and define

$$M_1 = \left\{ (u,b(u),0_{D-d-1}) : u \in B_d(0,1+\kappa) \right\} \cup \left\{ (u,-a(u),0_{D-d-1}) : u \in B_d(0,1+\kappa) \right\}$$

where

$$b(u) = \begin{cases} \gamma + \sqrt{\kappa^2 - ||u||^2} & \text{if } ||u|| \le \frac{1}{2}\sqrt{4\gamma\kappa - \gamma^2} \\ 2\kappa - \sqrt{\kappa^2 - (||u|| - \sqrt{4\gamma\kappa - \gamma^2})^2} & \text{if } \frac{1}{2}\sqrt{4\gamma\kappa - \gamma^2} < ||u|| \le \sqrt{4\gamma\kappa - \gamma^2} \\ a(u) & \text{if } \sqrt{4\gamma\kappa - \gamma^2} < ||u|| \le \sqrt{4\gamma\kappa - \gamma^2} + \kappa. \end{cases}$$

Note that $\Delta(M_1) \ge \kappa$ since the perturbation is obtained using portions of spheres of radius κ. In fact

- for $||u|| \le \frac{1}{2}\sqrt{4\gamma\kappa - \gamma^2}$, $b(u)$ is the $d+1$-th coordinate of the "upper" portion of the $(d+1)$-dimensional sphere with radius κ centered at $(0,\cdots,0,\gamma)$, hence $b(u)$ satisfies

$$||u||^2 + (b(u) - \gamma)^2 = \kappa^2 \qquad \text{with } b(u) \ge \gamma;$$

- for $\frac{1}{2}\sqrt{4\gamma\kappa - \gamma^2} < ||u|| \le \sqrt{4\gamma\kappa - \gamma^2}$, $b(u)$ is the $(d+1)$-th coordinate of the "lower" portion of the $(d+1)$-dimensional sphere with radius κ centered at $(u \cdot \sqrt{4\gamma\kappa - \gamma^2}/||u||, 2\kappa)$ (note that the center of the sphere differs according to the direction of u), hence $b(u)$ satisfies

$$\left|\left| u - \frac{u}{||u||}\sqrt{4\gamma\kappa - \gamma^2} \right|\right|^2 + (b(u) - 2\kappa)^2 = \kappa^2 \qquad \text{with } b(u) \le 2\kappa.$$

To summarize, M_0 and M_1 are both manifolds with no boundary, $\Delta(M_0) \geq \kappa$ and $\Delta(M_1) \geq \kappa$. See Figure 5. Now

$$
\begin{aligned}
E_0 &= M_0 - M_1 = \left\{ (u, a(u), 0_{D-d-1}) : u \in B_d(0, \sqrt{4\gamma\kappa - \gamma^2}) \right\} \\
E_1 &= M_1 - M_0 = \left\{ (u, b(u), 0_{D-d-1}) : u \in B_d(0, \sqrt{4\gamma\kappa - \gamma^2}) \right\}.
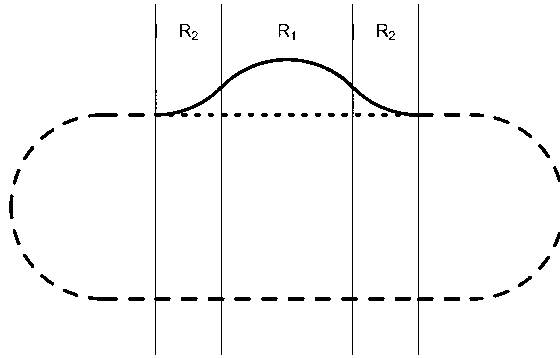\end{aligned}
$$

Figure 5: One section of manifolds M_0 and M_1. The common part is dashed, E_0 is dotted and E_1 solid. R_1 and R_2 denote the regions where the different definitions of the perturbation apply: R_1 is $||u|| \leq \frac{1}{2}\sqrt{4\gamma\kappa - \gamma^2}$ while R_2 denotes $\frac{1}{2}\sqrt{4\gamma\kappa - \gamma^2} < ||u|| \leq \sqrt{4\gamma\kappa - \gamma^2}$.

Note that for each point $y \in E_0$ there exists $y' \in E_1$ such that $||y - y'|| \leq |a(u) - b(u)| \leq \gamma$. Also, $y_0 = (0, a(0), 0) \in M_0$ has as its closest M_1 point $y_1 = (0, b(0), 0)$, so that $||y_0 - y_1|| = \gamma$. Hence $H(M_0, M_1) = H(E_0, E_1) = \gamma$.

To find an upper bound for $V(S_0 \circ S_1)$, we show that each $y = (u, v, z) \in S_1 - S_0$ satisfies the following conditions:

(i) $u \in B_d(0, \sqrt{4\gamma\kappa - \gamma^2})$;

(ii) $z \in B_{D-d-1}(0, \sigma)$;

(iii) $\kappa + \sigma - ||z|| < v \leq \kappa + \gamma + \sigma - ||z||$.

If $y = (u, v, z)$ belongs to S_1 and has $||u|| > \sqrt{4\gamma\kappa - \gamma^2}$, then there is a point of $M_0 \cap M_1$ within distance σ, hence $y \notin S_1 - S_0$. This proves (i). Before proving (ii) and (iii), note that if $u \in B_d(0, \sqrt{4\gamma\kappa - \gamma^2})$ then

$$\kappa = a(u) \leq b(u) \leq \kappa + \gamma.$$

Now, let $y' = (u', b(u'), 0) \in E_1$ be the point in S_1 closest to y. We have

$$d(y, S_1) = ||y - y'|| \leq ||u - u'|| + |v - b(u')| + ||z|| \leq \sigma.$$

This gives condition (ii) above $||z|| \leq \sigma$ and also

$$|v - b(u')| \leq \sigma - ||z||. \tag{4}$$

Since $b(u') \leq \kappa + \gamma$, we obtain

$$v \leq b(u') + \sigma - ||z|| \leq \kappa + \gamma + \sigma - ||z||$$

which is the right inequality in (iii). Finally,

$$\sigma < d(y, M_0) \leq ||y - (u, a(u), 0)|| \leq |v - a(u)| + ||z||$$

which implies either $v < a(u) - (\sigma - ||z||)$ or $v > a(u) + (\sigma - ||z||)$. The former inequality would imply

$$v < a(u) - (\sigma - ||z||) = \kappa - (\sigma - ||z||) \leq \inf_{u'} b(u') - (\sigma - ||z||)$$

so that $|v - b(u')| > \sigma - ||z||$ for all u', which is in contradiction with (4). Hence we have $v > a(u) + (\sigma - ||z||) = \kappa + (\sigma - ||z||)$ that is the left inequality in (iii).

As a consequence,

$$S_1 - S_0 \subset B_d(0, \sqrt{4\gamma\kappa - \gamma^2}) \times \left\{ (v, z) \in \mathbb{R}^{D-d} : \kappa - \gamma + \sigma - ||z|| < v \leq \kappa + \gamma + \sigma - ||z||, z \in B_{D-d-1}(0, \sigma) \right\}$$

and

$$V(S_0 - S_1) \leq C \cdot (\sqrt{4\gamma\kappa - \gamma^2})^d \cdot \gamma \cdot \sigma^{D-d-1}.$$

Hence, $V(S_0 - S_1) = O(\gamma^{\frac{d}{2}+1})$.

With similar arguments one can show that $V(S_1 - S_0) = O(\gamma^{\frac{d}{2}+1})$ so that

$$V(S_0 \circ S_1) = O(\gamma^{\frac{d}{2}+1}).$$

It then follows that $\int |q_0 - q_1| = O(\gamma^{(d+2)/2})$.

References

R. Baraniuk and M. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9:51–77, 2007.

M. Birman and M. Solomjak. Piecewise-polynomial approximation of functions of the classes w_p. *Mathematics of USSR Sbornik*, 73:295–317, 1967.

J. Boissonnat and A. Ghosh. Manifold reconstruction using tangential delaunay complexes. In *Proceedings of the 2010 Annual Symposium on Computational Geometry*, pages 324–333. ACM, 2010.

F. Chazal and A. Lieutier. Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Computational Geometry*, 40:156–170, 2008.

S. Cheng and T. Dey. Manifold reconstruction from point samples. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1018–1027. SIAM, 2005.

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

L. Devroye and G. Wise. Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal on Applied Mathematics*, 38:480–488, 1980.

T. Dey. *Curve and Surface Reconstruction: Algorithms with Mathematical Analysis*. Cambridge University Press, 2006.

T. Dey and S. Goswami. Provable surface reconstruction from noisy samples. In *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, pages 330–339. ACM, 2004.

H. Federer. Curvature measures. *Transactions of the American Statistical Society*, 93:418–491, 1959.

C. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Nonparametric filament estimation. *arXiv:1003.5536*, 2010.

O. Gonzalez and J. Maddocks. Global curvature, thickness, and the ideal shapes of knots. *Proceedings of the National Academy of Sciences*, 96(9):4769–4773, 1999.

L. LeCam. Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, pages 38–53, 1973.

J.M. Lee. *Introduction to Smooth Manifolds*. Springer, 2002.

P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 39:419–441, 2006.

P. Niyogi, S. Smale, and S. Weinberger. A topological view of unsupervised learning from noisy data. *Unpublished technical report, University of Chicago*, 2008.

X. Shen and W. Wong. Probability inequalities for likelihood ratios and convergence rates of sieve mles. *The Annals of Statistics*, 23:339–362, 1995.

A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.

B. Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*. Springer, 1997.