SUPPLEMENTARY MATERIAL FOR: NON-ASYMPTOTIC RATES FOR MANIFOLD, TANGENT SPACE AND CURVATURE ESTIMATION

By Eddie Aamari[§,*,†,‡] and Clément Levrard[¶,*,†]

U.C. San Diego[§] , *Université Paris-Diderot*[¶]

Appendix A: Properties and Stability of the Models

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

Here we show the following Lemma 1, reproduced as Lemma A.1.

Lemma A.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 $v \in \mathcal{B}_{T_pM}(0, \tau_{min}/4)$,

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

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

Proof of Lemma A.1. Proposition 6.1 in [13] states that for all $x \in M$, $\left\|II^M_x\right\|_{op} \leq 1/\tau_{min}$. In particular, Gauss equation ([8, 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 [3], 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 $d_v\mathbf{N}_p =$

*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

$d_v \exp_p - Id_{T_pM}$. 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 [9],

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

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,

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

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 [10] 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 $\mathcal{C}^k$

Lemma A.2. *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

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

where L_i' *depends on* $d, k, \tau_{min}, L_\perp, \ldots, L_i$, *and* C *on* d, k, τ_{min}, $L_\perp$, $\ldots$, 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*

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

PROOF OF LEMMA A.2. (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 A.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 [4], 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 [4].

(iv) Notice that Lemma A.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 A.3, 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 $T_2^* = II_x^M$. 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 [8, 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 [13]. It yields the bound on sectional curvature, using the Gauss equation [8, Proposition 3.1 (a), p.135].

□

In the proof of Lemma A.2 (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 A.3. *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 $2 \leq i \leq k$,*

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

Proof of Lemma A.3. 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 $\{1, \ldots, i\}$ 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 $h_1, \ldots, h_i \in \mathbb{R}^D$,

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

Isolating the term for $j = 1$ entails

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

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. Proof of Proposition 1

This section is devoted to prove Proposition 1 (reproduced below as Proposition A.4), that asserts the stability of the model with respect to ambient diffeomorphisms.

PROPOSITION A.4. *Let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ be a global $\mathcal{C}^k$-diffeomorphism. If $\|d\Phi - I_D\|_{op}$, $\left\|d^2\Phi\right\|_{op}$, ... , $\left\|d^k\Phi\right\|_{op}$ 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 $\mathbf{L}_{(\lambda)} = (L_\perp/\lambda, L_3/\lambda^2, \ldots, L_k/\lambda^{k-1})$.

Proof of Proposition A.4. The second part is 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.

The first part follows combining Proposition A.5 and Lemma A.6. $\square$

Proposition A.5 asserts the stability of the geometric model, that is, the reach bound and the existence of a smooth parametrization when a submanifold is perturbed.

Proposition A.5. *Let $\Phi : \mathbb{R}^D \to \mathbb{R}^D$ be a global $\mathcal{C}^k$-diffeomorphism. If $\|d\Phi - I_D\|_{op}$, $\|d^2\Phi\|_{op}$, ..., $\|d^k\Phi\|_{op}$ are small enough, then for all M in $\mathcal{C}^k_{\tau_{min},\mathbf{L}}$, the image $M' = \Phi(M)$ belongs to $\mathcal{C}^k_{\tau_{min}/2, 2L_\perp, 2L_3, \ldots, 2L_k}$.*

Proof of Proposition A.5. 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 [10],

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

for $\|I_D - d\Phi\|_{op}$ and $\|d^2\Phi\|_{op}$ 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 $v' \in T_{p'}M' = d_p\Phi(T_pM)$, 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 $\mathbf{N}'_{p'}(v') = \left\{\Phi\left(\Psi_p\left(d_{p'}\Phi^{-1}.v'\right)\right) - p' - v'\right\}$.

$$\begin{tikzcd} M \arrow[r, "\Phi"] & M' \\ T_pM \arrow[u, "\Psi_p"] \arrow[r, "d_p\Phi"'] & T_{p'}M' \arrow[u, "\Psi'_{p'}"'] \end{tikzcd}$$

The maps $\Psi'_{p'}(v')$ and $\mathbf{N}'_{p'}(v')$ are well defined whenever $\left\|d_{p'}\Phi^{-1}.v'\right\| \leq \frac{1}{4L_\perp}$, so in particular if $\|v'\| \leq \frac{1}{4(2L_\perp)} \leq \frac{1-\|I_D - d\Phi\|_{op}}{4L_\perp}$ 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 $w' \in T_{p'}M'$,

$$
\begin{aligned}
\left\|d^2_{v'}\mathbf{N}'_{p'}(w'^{\otimes 2})\right\| &= \Big\| d^2_{c(v')}\Phi\left(\left\{d_{d_{p'}\Phi^{-1}.v'}\Psi_p \circ d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right) \\
&\quad + d_{c(v')}\Phi \circ d^2_{d_{p'}\Phi^{-1}.v'}\Psi_p\left(\left\{d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right)\Big\| \\
&= \Big\| d^2_{c(v')}\Phi\left(\left\{d_{d_{p'}\Phi^{-1}.v'}\Psi_p \circ d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right) \\
&\quad + \left(d_{c(v')}\Phi - Id\right) \circ d^2_{d_{p'}\Phi^{-1}.v'}\Psi_p\left(\left\{d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right) \\
&\quad + d^2_{d_{p'}\Phi^{-1}.v'}\Psi_p\left(\left\{d_{p'}\Phi^{-1}.w'\right\}^{\otimes 2}\right)\Big\| \\
&\leq \left\|d^2\Phi\right\|_{op}\left(1 + L_\perp\left\|d_{p'}\Phi^{-1}.v'\right\|\right)^2\left\|d_{p'}\Phi^{-1}.w'\right\|^2 \\
&\quad + \|I_D - d\Phi\|_{op} L_\perp \left\|d_{p'}\Phi^{-1}.w'\right\|^2 \\
&\quad + L_\perp\left\|d_{p'}\Phi^{-1}.w'\right\|^2 \\
&\leq \left\|d^2\Phi\right\|_{op}(1 + 1/4)^2\left\|d_{p'}\Phi^{-1}\right\|^2_{op} \\
&\quad + \|I_D - d\Phi\|_{op} L_\perp\left\|d\Phi^{-1}\right\|^2_{op} \\
&\quad + L_\perp\left\|d_{p'}\Phi^{-1}\right\|^2_{op}.
\end{aligned}
$$

Writing further $\left\|d\Phi^{-1}\right\|_{op} \leq (1 - \|I_D - d\Phi\|_{op})^{-1} \leq 1 + 2\|I_D - \Phi\|_{op}$ for $\|I_D - d\Phi\|_{op}$ small enough depending only on $L_\perp$, it is clear that the right-hand side of the latter inequality goes below $2L_\perp$ for $\|I_D - d\Phi\|_{op}$ and $\left\|d^2\Phi\right\|_{op}$ small enough. Hence, for $\|I_D - d\Phi\|_{op}$ 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 $\|v'\| \leq \frac{1}{4(2L_\perp)}$. From the chain rule, the same argument applies for the order $3 \leq i \leq k$ differential of $\mathbf{N}'_{p'}$. $\square$

Lemma A.6 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.

Lemma A.6 (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 $\|I_D - d\Phi\|_{\mathrm{op}} < 1/3$. 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 $z \in \Phi(M)$,*

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

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

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

Proof of Lemma A.6. 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 ([11, Theorem 3.2.5]) states that if h is one-to-one,

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

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 $h = \Phi \circ \Psi_p$,

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

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

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

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

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

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

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

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

and

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

which yields the result. □

Appendix 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 B.7. *Let $P_0 \in \mathcal{P}^k$ 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 $p_x(r) = P_0\big(\mathcal{B}(x, r)\big)$.

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 $h \leq \tau_{min}/4$, with probability at least $1 - \left(\frac{1}{n}\right)^{k/d}$,

$$d_H(M, \mathbb{Y}_n) \leq h/2.$$

PROOF OF LEMMA B.7. Denoting by $\mathcal{B}_M(x,r)$ the geodesic ball of radius r centered at x, Proposition 25 of [1] 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 [1] 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 [7] entails, for $s \leq \tau_{min}/8$,

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

Hence, taking $s = h/2$, and $h = \left(\frac{C'_d k}{f_{min}} \frac{\log n}{n}\right)^{1/d}$ with C'_d so that $C'_d \geq \frac{8^d}{c_d k} \vee \frac{2^d(1+k/d)}{c_d k}$ yields the result. Since $k \geq 1$, taking $C'_d = \frac{8^d}{c_d}$ is sufficient. □

B.2. Concentration Bounds for Local Polynomials

This section is devoted to the proof of the following proposition.

PROPOSITION B.8. *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 $3h/2 \leq h_0 \leq \tau_{min}/4$, then with probability at least $1 - \left(\frac{1}{n}\right)^{\frac{k}{d}+1}$, we have*

$$\begin{array}{rcl} P_{0,n-1}[S^2(\pi^*(x)) \mathbb{1}_{\mathcal{B}(h/2)}(x)] & \geq & c_{k,d} h^d f_{min} \|S_h\|_2^2, \\ N(3h/2) & \leq & C_d f_{max} (n-1) h^d, \end{array}$$

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

A first step is to ensure that empirical expectations of order k polynomials are close to their deterministic counterparts.

Proposition B.9. *Let $b \leq \tau_{min}/8$. For any $y_0 \in M$, we have*

$$\mathbb{P}\left[\sup_{u_1,\ldots,u_k,\varepsilon \in \{0,1\}^k} \left|(P_0 - P_{0,n-1}) \prod_{j=1}^{p} \left(\frac{\langle u_j, y\rangle}{b}\right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(y_0,b)}(y)\right| \right.$$
$$\left. \geq p_{y_0}(b)\left(\frac{4k\sqrt{2\pi}}{\sqrt{(n-1)p_{y_0}(b)}} + \sqrt{\frac{2t}{(n-1)p_{y_0}(b)}} + \frac{2}{3(n-1)p_{y_0}(b)}\right)\right] \leq e^{-t},$$

where $P_{0,n-1}$ denotes the empirical distribution of $n-1$ i.i.d. random variables Y_i drawn from P_0.

Proof of Proposition B.9. Without loss of generality we choose $y_0 = 0$ and shorten notation to $\mathcal{B}(b)$ and $p(b)$. Let $\mathcal{Z}$ denote the empirical process on the left-hand side of Proposition B.9. Denote also by $f_{u,\varepsilon}$ the map $\prod_{j=1}^{k} \left(\frac{\langle u_j,y\rangle}{b}\right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(b)}(y)$, 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 $Pf_{u,\varepsilon}^2 \leq p(b)$, the Talagrand-Bousquet inequality ([6, Theorem 2.3]) yields

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

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

Lemma B.10. *We may write*

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

Proof of Lemma B.10. 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

$$
\begin{aligned}
\mathbb{E}\mathcal{Z} &= \mathbb{E}_Y \sup_{u,\varepsilon} \left| (P_0 - P_{0,n-1}) \prod_{j=1}^{k} \left(\frac{\langle u_j, y \rangle}{b} \right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(b)}(y) \right| \\
&\leq \frac{2}{n-1} \mathbb{E}_Y \mathbb{E}_\sigma \sup_{u,\varepsilon} \sum_{i=1}^{n-1} \sigma_i \prod_{j=1}^{k} \left(\frac{\langle u_j, Y_i \rangle}{b} \right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(b)}(Y_i) \\
&\leq \frac{\sqrt{2\pi}}{n-1} \mathbb{E}_Y \mathbb{E}_g \sup_{u,\varepsilon} \sum_{i=1}^{n-1} g_i \prod_{j=1}^{k} \left(\frac{\langle u_j, Y_i \rangle}{b} \right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(b)}(Y_i).
\end{aligned}
$$

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

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

we deduce that

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

where $\Theta_{u,\varepsilon} = \sqrt{k} \sum_{i=1}^{n-1} \sum_{r=1}^{k} g_{i,r} \frac{\langle \varepsilon_r u_r, Y_i \rangle}{b} \mathbb{1}_{\mathcal{B}(b)}(Y_i)$. According to Slepian's

Lemma [5, Theorem 13.3], it follows that

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

We deduce that

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

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

Combining Lemma B.10 with Talagrand-Bousquet's inequality gives the result of Proposition B.9. □

We are now in position to prove Proposition B.8.

Proof of Proposition B.8. If $h/2 \leq \tau_{min}/4$, then, according to Lemma B.7, $p(h/2) \geq c_d f_{min} h^d$, hence, if $h = \left(K \frac{\log(n)}{n-1}\right)^{\frac{1}{d}}$, $(n-1)p(h/2) \geq K c_d f_{min} \log(n)$. Choosing $b = h/2$ and $t = (k/d+1)\log(n) + \log(2)$ in Proposition B.9 and $K = K'/f_{min}$, with $K' > 1$ leads to

$$
\begin{aligned}
\mathbb{P}\Bigg[\sup_{u_1,\ldots,u_k,\varepsilon \in \{0,1\}^k} &\left| (P_0 - P_{0,n-1}) \prod_{j=1}^{k} \left(2\frac{\langle u_j, y \rangle}{h} \right)^{\varepsilon_j} \mathbb{1}_{\mathcal{B}(y_0,h/2)}(y) \right| \\
&\geq \frac{c_{d,k} f_{max}}{\sqrt{K'}} h^d \Bigg] \leq \frac{1}{2} \left(\frac{1}{n}\right)^{\frac{k}{d}+1}.
\end{aligned}
$$

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

$$
\begin{aligned}
(P_{0,n-1} - P_0) S^2(y_{1:d}) \mathbb{1}_{\mathcal{B}(h/2)}(y) &\geq - \sum_{\alpha,\beta} \frac{c_{d,k} f_{max}}{\sqrt{K'}} |a_\alpha a_\beta| h^{d+|\alpha|+|\beta|} \\
&\geq - \frac{c_{d,k} f_{max}}{\sqrt{K'}} h^d \|S_h\|_2^2.
\end{aligned}
$$

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

$$
\int_{\mathcal{B}(0,r)} S^2(y_{1:d}) dy_1 \dots dy_d \geq C_{d,k} r^d \|S_r\|_2^2,
$$

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

$$
P_0 S^2(y_{1:d}) \mathbb{1}_{\mathcal{B}(h/2)}(y) \geq P_0 S^2(y_{1:d}) \mathbb{1}_{B(7h/16)}(y_{1:d}) \geq c_{k,d} h^d f_{min} \|S_h\|_2^2,
$$

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

$$
P_{0,n-1} S^2(x_{1:d}) \mathbb{1}_{\mathcal{B}(h/2)}(y) \geq c_{k,d} f_{min} h^d \|S_h\|_2^2.
$$

The second inequality of Proposition B.8 is derived the same way from Proposition B.9, choosing $\varepsilon = (0, \dots, 0)$, $b = 3h/2$ and $h \leq \tau_{min}/8$ so that $b \leq \tau_{min}/4$. □

Appendix C: Minimax Lower Bounds

C.1. Conditional Assouad's Lemma

This section is dedicated to the proof of Lemma 7, reproduced below as Lemma C.11.

Lemma C.11 (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

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

Then,

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

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

Proof of Lemma C.11. The proof follows that of Lemma 2 in [14]. 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,

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

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 $1 \le k \le m$. 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. Therefore,

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

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 above bound $|\{1, \ldots, m\} \times \{0,1\}^m| = m2^m$ times. □

C.2. Construction of Generic Hypotheses

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 in $\mathbb{R}^{d+1} \times \{0\}^{D-d-1}$. 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, and from Proposition A.4, we have

$$
\tau_{M_0} \geq 2\tau_{min}, \qquad \mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(0, \tau_{min}) \subset M_0, \qquad Vol(M_0) = C_d \tau_{min}^d,
$$

and the uniform probability distribution P_0 on M_0 satisfies

$$
P_0 \in \mathcal{P}^k_{2\tau_{min}, \mathbf{L}/2, 2f_{min}, f_{max}/2},
$$

whenever $L_\perp/2 \geq L_\perp^{(0)}/(2\tau_{min})$, $\ldots$, $L_k/2 \geq L_k^{(0)}/(2\tau_{min})^{k-1}$, and provided that $2f_{min} \leq \left((2\tau_{min})^d V_0^{(0)}\right)^{-1} \leq f_{max}/2$. 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 $0 < \delta \leq \tau_{min}/4$, let $x_1, \ldots, x_m \in M_0 \cap \mathcal{B}(0, \tau_{min}/4)$ be a family of points such that

$$\text{for} \quad 1 \leq k \neq k' \leq m, \quad \|x_k - x_{k'}\| \geq \delta.$$

For instance, considering the family $\left\{ \left(l_1\delta, \ldots, l_d\delta, 0, \ldots, 0\right) \right\}_{l_i \in \mathbb{Z}, |l_i| \leq \lfloor \tau_{min}/(4\delta) \rfloor}$,

$$m \geq c_d \left(\frac{\tau_{min}}{\delta} \right)^d,$$

for some $c_d > 0$.

We let $e \in \mathbb{R}^D$ denote the $(d+1)$th vector of the canonical basis. In particular, we have the orthogonal decomposition of the ambient space

$$\mathbb{R}^D = \left(\mathbb{R}^d \times \{0\}^{D-d} \right) + span(e) + \left(\{0\}^{d+1} \times \mathbb{R}^{D-d-1} \right).$$

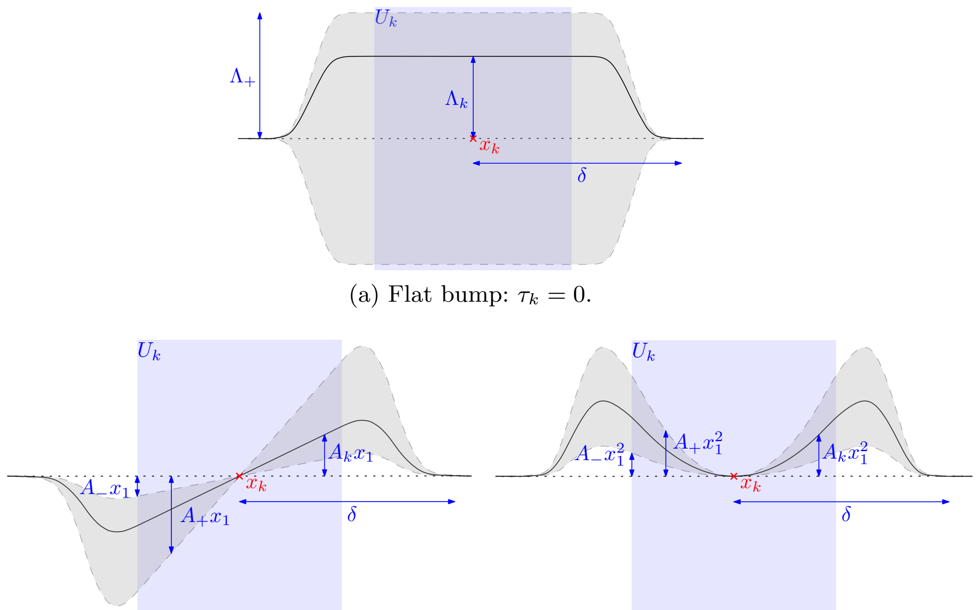
Let $\phi : \mathbb{R}^D \to [0,1]$ be a smooth scalar map such that $\phi|_{\mathcal{B}\left(0, \frac{1}{2}\right)} = 1$ and $\phi|_{\mathcal{B}(0,1)^c} = 0$.

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

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

An analogous deformation map was considered in [1]. 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 support. Roughly speaking, $M_\tau^{\mathbf{\Lambda}, \mathbf{A}, i}$ consists of m bumps at the x_k's having different shapes (Figure 1). 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 C.12 gives differential bounds and geometric properties of $\Phi_\tau^{\mathbf{\Lambda}, \mathbf{A}, i}$.

Lemma C.12. *There exists $c_{\phi,i} < 1$ such that if $A_+ \leq c_{\phi,i}\delta^{i-1}$ and $\Lambda_+ \leq c_{\phi,i}\delta$, then $\Phi_\tau^{\mathbf{\Lambda}, \mathbf{A}, i}$ is a global $\mathcal{C}^\infty$-diffeomorphism of $\mathbb{R}^D$ such that for*



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

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

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

Figure 1: The three shapes of the bump map $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}$ around x_k.

all $1 \leq k \leq m$, $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}(\mathcal{B}(x_k,\delta)) = \mathcal{B}(x_k,\delta)$. *Moreover,*

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

and for $j \geq 2$,

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

Proof of Lemma C.12. Follows straightforwardly from chain rule, similarly to Lemma 11 in [1]. □

Lemma C.13. *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 provided that* $\Lambda_+ \vee A_+ \delta^i \leq c_{k,d,\tau_{min}} \delta^k$*, for all* $\tau \in \{0,1\}^m$, $P_\tau^{\mathbf{\Lambda},\mathbf{A},i} \in \mathcal{P}^k_{\tau_{min},\mathbf{L},f_{min},f_{max}}$

Proof of Lemma C.13. Follows using the stability of the model Lemma A.4 applied to the distribution $P_0 \in \mathcal{P}^k_{2\tau_{min},\mathbf{L}/2,2f_{min},f_{max}/2}$ and the map $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}$, of which differential bounds are asserted by Lemma C.12. □

C.3. Hypotheses for Tangent Space and Curvature

C.3.1. Proof of Lemma 8

This section is devoted to the proof of Lemma 8, for which we first derive two slightly more general results, with parameters to be tuned later. The proof is split into two intermediate results Lemma C.14 and Lemma C.15.

Let us write $\bar{Q}^{(i)}_{\tau,n}$ for the mixture distribution on $(\mathbb{R}^D)^n$ defined by

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

Although the probability distribution $\bar{Q}^{(i)}_{\tau,n}$ depends on A_-, A_+ and Λ_+, we omit this dependency for the sake of compactness. Another way to define $\bar{Q}^{(i)}_{\tau,n}$ 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 $Y_1,\ldots,Y_n$ is an i.i.d. n-sample with common distribution P_0 on M_0. Then $(Z_1,\ldots,Z_n)$ has distribution $\bar{Q}^{(i)}_{\tau,n}$.

Lemma C.14. *Assume that the conditions of Lemma C.12 hold, and let*

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

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

Moreover, if $(X_1,\ldots,X_n)$ 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* $i = 2$*:* $\left\| II_{X_1}^{M_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}} \circ \pi_{T_{X_1} M_\tau^{\mathbf{\Lambda},\mathbf{A},(2)}} \right\|_{op} \geq A_-/2.$

PROOF OF LEMMA C.14. It is clear from the definition (2) that $\bar{Q}_{\tau,n}^{(i)} \in \overline{Conv}\big((\mathcal{P}_\tau^{(i)})^{\otimes n}\big)$. By construction of the $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},i}$'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 $z = \Phi_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}(x_0) \in U_k \cap M_\tau^{\mathbf{\Lambda},\mathbf{A},(i)}$. 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, $\Phi_\tau^{\mathbf{\Lambda},\mathbf{A},(1)}$ 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 $II_{x_0}^{M_0} = 0$, we have

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

- 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,

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

□

Lemma C.15. *Assume that the conditions of Lemma C.12 and Lemma C.14 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,*

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

and

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

Proof of Lemma C.15. First note that all the involved distributions 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 (2), on the event $\{Z \in U_k\}$, a random variable Z having distribution $\bar{Q}^{(i)}_{\tau,1}$ can be represented by $Z = X + \phi\left(\frac{X - x_k}{\delta}\right)\Lambda_k e = X + \Lambda_k e$ 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, $\bar{Q}^{(i)}_{\tau,1}$ has a density with respect to the Lebesgue measure λ_{d+1} on $\mathbb{R}^d \times span(e)$ that can be written as

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

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

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

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

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

It follows that

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

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,

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

hence the result. □

Proof of Lemma 8. The properties of $\left\{\bar{Q}^{(i)}_{\tau,n}\right\}_\tau$ and $\left\{U_k \times U'_k\right\}_k$ given by Lemma C.14 and Lemma C.15 yield the result, setting $\Lambda_+ = A_+ \delta^i/4$, $A_+ = 2A_- = \varepsilon \delta^{k-i}$ for $\varepsilon = \varepsilon_{k,d,\tau_{min}}$, and δ such that $c'_d \left(\frac{\delta}{\tau_{min}}\right)^d = \frac{1}{n-1}$. □

C.3.2. Proof of Lemma 9

This section details the construction leading to Lemma 9 that we restate in Lemma C.16.

LEMMA C.16. *Assume that* $\tau_{min}L_\perp,\ldots,\tau_{min}^{k-1}L_k,(\tau_{min}^d f_{min})^{-1}$, $\tau_{min}^d f_{max}$ *are large enough (depending only on* d *and* k*), and* $\sigma \geq C_{k,d,\tau_{min}}\left(1/(n-1)\right)^{k/d}$ *for* $C_{k,d,\tau_{min}} > 0$ *large enough. Given* $i \in \{1,2\}$*, there exists a collection of* 2^m *distributions* $\left\{\mathbf{P}_\tau^{(i),\sigma}\right\}_{\tau\in\{0,1\}^m} \subset \mathcal{P}^k(\sigma)$ *with associated submanifolds* $\left\{M_\tau^{(i),\sigma}\right\}_{\tau\in\{0,1\}^m}$*, together with pairwise disjoint subsets* $\{U_k^\sigma\}_{1\leq k\leq m}$ *of* $\mathbb{R}^D$ *such that the following holds for all* $\tau \in \{0,1\}^m$ *and* $1 \leq k \leq m$.

If $x \in U_k^\sigma$ *and* $y = \pi_{M_\tau^{(i),\sigma}}(x)$*, we have*

- *if* $\tau_k = 0$,

$$T_y M_\tau^{(i),\sigma} = \mathbb{R}^d \times \{0\}^{D-d} \quad , \quad \left\| II_y^{M_\tau^{(i),\sigma}} \circ \pi_{T_y M_\tau^{(i),\sigma}} \right\|_{op} = 0,$$

- *if* $\tau_k = 1$,
  - *for* $i = 1$*:* $\angle\left(T_y M_\tau^{(1),\sigma}, \mathbb{R}^d \times \{0\}^{D-d}\right) \geq c_{k,d,\tau_{min}}\left(\dfrac{\sigma}{n-1}\right)^{\frac{k-1}{k+d}}$,
  - *for* $i = 2$*:* $\left\| II_y^{M_\tau^{(2),\sigma}} \circ \pi_{T_y M_\tau^{(2),\sigma}} \right\|_{op} \geq c'_{k,d,\tau_{min}}\left(\dfrac{\sigma}{n-1}\right)^{\frac{k-2}{k+d}}$.

Furthermore,

$$\int_{(\mathbb{R}^D)^{n-1}} \left(\mathbf{P}_\tau^{(i),\sigma}\right)^{\otimes n-1} \wedge \left(\mathbf{P}_{\tau^k}^{(i),\sigma}\right)^{\otimes n-1} \geq c_0, \text{ and } \quad m \cdot \int_{U_k^\sigma} \mathbf{P}_\tau^{(i),\sigma} \wedge \mathbf{P}_{\tau^k}^{(i),\sigma} \geq c_d.$$

PROOF OF LEMMA C.16. Following the notation of Section C.2, for $i \in \{1,2\}$, $\tau \in \{0,1\}^m$, $\delta \leq \tau_{min}/4$ and $A > 0$, consider

$$\Phi_\tau^{A,i}(x) = x + \sum_{k=1}^m \phi\left(\frac{x-x_k}{\delta}\right)\left\{\tau_k A (x-x_k)_1^i\right\} e. \tag{3}$$

Note that (3) is a particular case of (1). Clearly from the definition, $\Phi_\tau^{A,i}$ and $\Phi_{\tau^k}^{A,i}$ coincide outside $\mathcal{B}(x_k,\delta)$, $(\Phi(x)-x) \in span(e)$ for all $x \in \mathbb{R}^D$, and $\|I_D - \Phi\|_\infty \leq A\delta^i$. Let us define $M_\tau^{A,i} = \Phi_\tau^{A,i}(M_0)$. From Lemma C.13, we have $M_\tau^{A,i} \in \mathcal{C}^k_{\tau_{min},\mathbf{L}}$ provided that $\tau_{min}L_\perp,\ldots,\tau_{min}^{k-1}L_k$ are large enough, and that $\delta \leq \tau_{min}/2$, with $A/\delta^{k-i} \leq \varepsilon$ for $\varepsilon = \varepsilon_{k,d,\tau_{min},i}$ small enough.

Furthermore, let us write

$$U_k^\sigma = \mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}\left(x_k, \delta/2\right) + \mathcal{B}_{\{0\}^d \times \mathbb{R}^{D-d}}\left(x_k, \sigma/2\right).$$

Then the family $\{U_k^\sigma\}_{1 \le k \le m}$ is pairwise disjoint. Also, since $\tau_k = 0$ implies that $M_\tau^{A,i}$ coincides with M_0 on $\mathcal{B}(x_k, \delta)$, we get that if $x \in U_k^\sigma$ and $y = \pi_{M_\tau^{A,i}}(x)$,

$$T_y M_\tau^{A,i} = \mathbb{R}^d \times \{0\}^{D-d} \quad , \quad \left\| II_y^{M_\tau^{A,i}} \circ \pi_{T_y M_\tau^{A,i}} \right\|_{op} = 0.$$

Furthermore, by construction of the bump function $\Phi_\tau^{A,i}$, if $x \in U_k^\sigma$ and $\tau_k = 1$, then

$$\angle\left(T_y M_\tau^{A,i}, \mathbb{R}^d \times \{0\}^{D-d}\right) \geq \frac{A}{2},$$

and

$$\left\| II_y^{M_\tau^{A,i}} \circ \pi_{T_y M_\tau^{A,i}} \right\|_{op} \geq \frac{A}{2}.$$

Now, let us write

$$\mathcal{O}_\tau^{A,i} = \left\{ y + \xi \,\middle|\, y \in M_\tau^{A,i}, \xi \in \left(T_y M_\tau^{A,i}\right)^\perp, \|\xi\| \leq \sigma/2 \right\}$$

for the offset of $M_\tau^{\Lambda,A,i}$ of radius $\sigma/2$. The sets $\left\{\mathcal{O}_\tau^{A,i}\right\}_\tau$ are closed subsets of $\mathbb{R}^D$ with non-empty interiors. Let $\mathbf{P}_\tau^{A,i}$ denote the uniform distribution on $\mathcal{O}_\tau^{A,i}$. Finally, let us denote by $P_\tau^{A,i} = \left(\pi_{M_\tau^{A,i}}\right)_* \mathbf{P}_\tau^{A,i}$ the pushforward distributions of $\mathbf{P}_\tau^{A,i}$ by the projection maps $\pi_{M_\tau^{A,i}}$. From Lemma 19 in [12], $P_\tau^{A,i}$ has a density $f_\tau^{A,i}$ with respect to the volume measure on $M_\tau^{A,i}$, and this density satisfies

$$Vol\left(M_\tau^{A,i}\right) f_\tau^{A,i} \leq \left(\frac{\tau_{min} + \sigma/2}{\tau_{min} - \sigma/2}\right)^d \leq \left(\frac{5}{3}\right)^d,$$

and

$$Vol\left(M_\tau^{A,i}\right) f_\tau^{A,i} \geq \left(\frac{\tau_{min} - \sigma/2}{\tau_{min} + \sigma/2}\right)^d \geq \left(\frac{3}{5}\right)^d.$$

Since, by construction, $Vol(M_0) = c_d \tau_{min}^d$, and $c'_d \leq Vol\left(M_\tau^{\Lambda,A,i}\right)/Vol(M_0) \leq C'_d$ whenever $A/\delta^{i-1} \leq \varepsilon'_{d,\tau_{min},i}$, we get that $P_\tau^{A,i}$ belongs to the model $\mathcal{P}^k$ provided that $(\tau_{min}^d f_{min})^{-1}$ and $\tau_{min}^d f_{max}$ are large enough. This proves that under these conditions, the family $\left\{\mathbf{P}_\tau^{A,i}\right\}_{\tau \in \{0,1\}^m}$ is included in the model $\mathcal{P}^k(\sigma)$.

Let us now focus on the bounds on the L^1 test affinities. Let $\tau \in \{0,1\}^m$ and $1 \leq k \leq m$ be fixed, and assume, without loss of generality, that $\tau_k = 0$ (if not, flip the role of τ and τ^k). First, note that

$$\int_{(\mathbb{R}^D)^{n-1}} \left(\mathbf{P}_\tau^{A,i}\right)^{\otimes n-1} \wedge \left(\mathbf{P}_{\tau^k}^{A,i}\right)^{\otimes n-1} \geq \left(\int_{\mathbb{R}^D} \mathbf{P}_\tau^{A,i} \wedge \mathbf{P}_{\tau^k}^{A,i}\right)^{n-1}.$$

Furthermore, since $\mathbf{P}_\tau^{A,i}$ and $\mathbf{P}_{\tau^k}^{A,i}$ are the uniform distributions on $\mathcal{O}_\tau^{A,i}$ and $\mathcal{O}_\tau^{A,i}$,

$$\begin{aligned}\int_{\mathbb{R}^D} \mathbf{P}_\tau^{A,i} \wedge \mathbf{P}_{\tau^k}^{A,i} &= 1 - \frac{1}{2}\int_{\mathbb{R}^D} \left|\mathbf{P}_\tau^{A,i} - \mathbf{P}_{\tau^k}^{A,i}\right| \\ &= 1 - \frac{1}{2}\int_{\mathbb{R}^D} \left| \frac{\mathbb{1}_{\mathcal{O}_\tau^{A,i}}(a)}{Vol\left(\mathcal{O}_\tau^{A,i}\right)} - \frac{\mathbb{1}_{\mathcal{O}_\tau^{A,i}}(a)}{Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)} \right| d\mathcal{H}^D(a).\end{aligned}$$

Furthermore,

$$\begin{aligned}&\frac{1}{2}\int_{\mathbb{R}^D} \left| \frac{\mathbb{1}_{\mathcal{O}_\tau^{A,i}}(a)}{Vol\left(\mathcal{O}_\tau^{A,i}\right)} - \frac{\mathbb{1}_{\mathcal{O}_{\tau^k}^{A,i}}(a)}{Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)} \right| d\mathcal{H}^D(a) \\ &\qquad = \frac{1}{2} Vol\left(\mathcal{O}_\tau^{A,i} \cap \mathcal{O}_{\tau^k}^{A,i}\right) \left| \frac{1}{Vol\left(\mathcal{O}_\tau^{A,i}\right)} - \frac{1}{Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)} \right| \\ &\qquad\quad + \frac{1}{2}\left(\frac{Vol\left(\mathcal{O}_\tau^{A,i} \setminus \mathcal{O}_{\tau^k}^{A,i}\right)}{Vol\left(\mathcal{O}_\tau^{A,i}\right)} + \frac{Vol\left(\mathcal{O}_{\tau^k}^{A,i} \setminus \mathcal{O}_\tau^{A,i}\right)}{Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)} \right) \\ &\qquad \leq \frac{3}{2} \frac{Vol\left(\mathcal{O}_\tau^{A,i} \setminus \mathcal{O}_{\tau^k}^{A,i}\right) \vee Vol\left(\mathcal{O}_{\tau^k}^{A,i} \setminus \mathcal{O}_\tau^{A,i}\right)}{Vol\left(\mathcal{O}_\tau^{A,i}\right) \wedge Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)}.\end{aligned}$$

To get a lower bound on the denominator, note that for $\delta \leq \tau_{min}/2$, $M_\tau^{A,i}$ and $M_{\tau^k}^{A,i}$ both contain

$$\mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(0, \tau_{min}) \setminus \mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(0, \tau_{min}/4),$$

so that $\mathcal{O}_\tau^{A,i}$ and $\mathcal{O}_{\tau^k}^{A,i}$ both contain

$$\left(\mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(0, \tau_{min}) \setminus \mathcal{B}_{\mathbb{R}^d \times \{0\}^{D-d}}(0, \tau_{min}/4)\right) + \mathcal{B}_{\{0\}^d \times \mathbb{R}^{D-d}}(0, \sigma/2).$$

As a consequence, $Vol\left(\mathcal{O}_\tau^{A,i}\right) \wedge Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right) \geq c_d \omega_d \tau_{min}^d \omega_{D-d} (\sigma/2)^{D-d}$, where ω_ℓ denote the volume of a ℓ-dimensional unit Euclidean ball.

We now derive an upper bound on $Vol\big(\mathcal{O}_\tau^{A,i} \setminus \mathcal{O}_{\tau^k}^{A,i}\big)$. To this aim, let us consider $a_0 = y + \xi \in \mathcal{O}_\tau^{A,i} \setminus \mathcal{O}_{\tau^k}^{A,i}$, with $y \in M_\tau^{A,i}$ and $\xi \in \big(T_y M_\tau^{A,i}\big)^\perp$. Since $\Phi_\tau^{A,i}$ and $\Phi_{\tau^k}^{A,i}$ coincide outside $\mathcal{B}(x_k,\delta)$, so do $M_\tau^{A,i}$ and $M_{\tau^k}^{A,i}$. Hence, one necessarily has $y \in \mathcal{B}(x_k,\delta)$. Thus, $\big(T_y M_\tau^{A,i}\big)^\perp = T_y M_0^\perp = span(e) + \{0\}^{d+1} \times \mathbb{R}^{D-d-1}$, so we can write $\xi = se + z$ with $s \in \mathbb{R}$ and $z \in \{0\}^{d+1} \times \mathbb{R}^{D-d-1}$. By definition of $\mathcal{O}_\tau^{A,i}$, $\|\xi\| = \sqrt{s^2 + \|z\|^2} \leq \sigma/2$, which yields $\|z\| \leq \sigma/2$ and $|s| \leq \sqrt{(\sigma/2)^2 - \|z\|^2}$. Furthermore, y_0 does not belong to $\mathcal{O}_{\tau^k}^{A,i}$, which translates to

$$\begin{aligned}\sigma/2 < d\big(a_0, M_{\tau^k}^{A,i}\big) &\leq \left\| y_0 + se + z - \Phi_{\tau^k}^{A,i}(y_0) \right\| \\ &= \sqrt{\left| s + \left\langle e, y_0 - \Phi_{\tau^k}^{A,i}(y_0) \right\rangle \right|^2 + \|z\|^2},\end{aligned}$$

from what we get $|s| \geq \sqrt{(\sigma/2)^2 - \|z\|^2} - \left\| I_D - \Phi_{\tau^k}^{A,i} \right\|_\infty$. We just proved that $\mathcal{O}_\tau^{A,i} \setminus \mathcal{O}_{\tau^k}^{A,i}$ is a subset of

$$\begin{aligned}\mathcal{B}_d(x_k,\delta) + \Bigg\{ se + z \Bigg| \ &(s,z) \in \mathbb{R} \times \mathbb{R}^{D-d-1}, \|z\| \leq \sigma/2 \text{ and} \\ &\sqrt{(\sigma/2)^2 - \|z\|^2} - \left\| I_D - \Phi_{\tau^k}^{A,i} \right\|_\infty \leq |s| \leq \sqrt{(\sigma/2)^2 - \|z\|^2} \Bigg\}.\end{aligned}$$

Hence,

$$Vol\left(\mathcal{O}_\tau^{A,i} \setminus \mathcal{O}_{\tau^k}^{A,i}\right) \leq \omega_d \delta^d \times 2 \left\| I_D - \Phi_{\tau^k}^{A,i} \right\|_\infty \times \omega_{D-d-1} (\sigma/2)^{D-d-1}. \tag{4}$$

Similar arguments lead to

$$Vol\left(\mathcal{O}_{\tau^k}^{A,i} \setminus \mathcal{O}_\tau^{A,i}\right) \leq \omega_d \delta^d \times 2 \left\| I_D - \Phi_\tau^{A,i} \right\|_\infty \times \omega_{D-d-1} (\sigma/2)^{D-d-1}. \tag{5}$$

Since $\left\| I_D - \Phi_\tau^{A,i} \right\|_\infty \vee \left\| I_D - \Phi_{\tau^k}^{A,i} \right\|_\infty \leq A\delta^i$, summing up bounds (4) and (5) yields

$$\begin{aligned}\int_{\mathbb{R}^D} \mathbf{P}_\tau^{A,i} \wedge \mathbf{P}_{\tau^k}^{A,i} &\geq 1 - 3 \frac{\omega_d \omega_{D-d-1} A\delta^i \cdot \delta^d (\sigma/2)^{D-d-1}}{\omega_d \tau_{min}^d \omega_{D-d} (\sigma/2)^{D-d}} \\ &\geq 1 - 3 \frac{A\delta^i}{\sigma} \left(\frac{\delta}{\tau_{min}} \right)^d.\end{aligned}$$

To derive the last bound, we notice that since $U_k^\sigma \subset \mathcal{O}_\tau^{A,i} = Supp\big(\mathbf{P}_\tau^{A,i}\big)$, we have

$$\begin{aligned}
\int_{U_k^\sigma} \mathbf{P}_\tau^{A,i} \wedge \mathbf{P}_{\tau^k}^{A,i} &\geq \frac{Vol\left(U_k^\sigma \cap \mathcal{O}_{\tau^k}^{A,i}\right)}{Vol\left(\mathcal{O}_\tau^{A,i}\right) \wedge Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)} \\
&\geq \frac{Vol\left(U_k^\sigma\right) - Vol\left(U_k^\sigma \setminus \mathcal{O}_{\tau^k}^{A,i}\right)}{Vol\left(\mathcal{O}_\tau^{A,i}\right) \wedge Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)} \\
&\geq \frac{Vol\left(U_k^\sigma\right) - Vol\left(\mathcal{O}_\tau^{A,i} \setminus \mathcal{O}_{\tau^k}^{A,i}\right)}{Vol\left(\mathcal{O}_\tau^{A,i}\right) \wedge Vol\left(\mathcal{O}_{\tau^k}^{A,i}\right)} \\
&\geq \frac{\omega_d(\delta/2)^d \omega_{D-d}(\sigma/2)^{D-d} - \omega_d \delta^d A \delta^i \omega_{D-d-1}(\sigma/2)^{D-d-1}}{\omega_d \tau_{min}^d \omega_{D-d}(\sigma/2)^{D-d}}.
\end{aligned}$$

Hence, whenever $A\delta^i \leq c_d \sigma$ for c_d small enough, we get

$$\int_{U_k^\sigma} \mathbf{P}_\tau^{A,i} \wedge \mathbf{P}_{\tau^k}^{A,i} \geq c'_d \left(\frac{\delta}{\tau_{min}}\right)^d.$$

Since m can be chosen such that $m \geq c_d(\tau_{min}/\delta)^d$, we get the last bound.

Eventually, writting $\mathbf{P}_\tau^{(i),\sigma} = \mathbf{P}_\tau^{A,i}$ for the particular parameters $A = \varepsilon \delta^{k-i}$, for $\varepsilon = \varepsilon_{k,d,\tau_{min}}$ small enough, and δ such that $\frac{3A\delta^i}{\sigma}\left(\frac{\delta}{\tau_{min}}\right)^d = \frac{1}{n-1}$ yields the result. Such a choice of parameter δ does meet the condition $A\delta^i = \varepsilon \delta^k \leq c_d \sigma$, provided that $\sigma \geq \frac{c_d}{\varepsilon}\left(\frac{1}{n-1}\right)^{k/d}$. □

C.4. Hypotheses for Manifold Estimation

C.4.1. Proof of Lemma 5

Let us prove Lemma 5, stated here as Lemma C.17.

Lemma C.17. *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), there exist $P_0, P_1 \in \mathcal{P}^k$ with associated submanifolds M_0, M_1 such that*

$$d_H(M_0, M_1) \geq c_{k,d,\tau_{min}} \left(\frac{1}{n}\right)^{\frac{k}{d}}, \text{ and } \quad \|P_0 \wedge P_1\|_1^n \geq c_0.$$

PROOF OF LEMMA C.17. Following the notation of Section C.2, for $\delta \leq \tau_{min}/4$ and $\Lambda > 0$, consider

$$\Phi_\tau^\Lambda(x) = x + \phi\left(\frac{x}{\delta}\right)\Lambda \cdot e,$$

which is a particular case of (1). Define $M^\Lambda = \Phi^\Lambda(M_0)$, and $P^\Lambda = \Phi_*^\Lambda P_0$. Under the conditions of Lemma C.13, P_0 and P^Λ belong to $\mathcal{P}^k$, and by construction, $d_H(M_0, M^\Lambda) = \Lambda$. In addition, since P_0 and P^Λ coincide outside $\mathcal{B}(0, \delta)$,

$$\int_{\mathbb{R}^D} dP_0 \wedge dP^\Lambda = P_0\big(\mathcal{B}(0,\delta)\big) = \omega_d \left(\frac{\delta}{\tau_{min}}\right)^d.$$

Setting $P_1 = P^\Lambda$ with $\omega_d \left(\frac{\delta}{\tau_{min}}\right)^d = \frac{1}{n}$ and $\Lambda = c_{k,d,\tau_{min}}\delta^k$ for $c_{k,d,\tau_{min}} > 0$ small enough yields the result. □

C.4.2. Proof of Lemma 6

Here comes the proof of Lemma 6, stated here as Lemma C.17.

LEMMA C.18. *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), there exist $P_0^\sigma, P_1^\sigma \in \mathcal{P}^k(\sigma)$ with associated submanifolds M_0^σ, M_1^σ such that*

$$d_H(M_0^\sigma, M_1^\sigma) \geq c_{k,d,\tau_{min}} \left(\frac{\sigma}{n}\right)^{\frac{k}{d+k}}, \quad \text{and} \quad \|P_0^\sigma \wedge P_1^\sigma\|_1^n \geq c_0.$$

PROOF OF LEMMA C.18. The proof follows the lines of that of Lemma C.16. Indeed, with the notation of Section C.2, for $\delta \leq \tau_{min}/4$ and $0 < \Lambda \leq c_{k,d,\tau_{min}}\delta^k$ for $c_{k,d,\tau_{min}} > 0$ small enough, consider

$$\Phi_\tau^\Lambda(x) = x + \phi\left(\frac{x}{\delta}\right)\Lambda \cdot e.$$

Define $M^\Lambda = \Phi^\Lambda(M_0)$. Write $\mathcal{O}_0$, $\mathcal{O}^\Lambda$ for the offsets of radii $\sigma/2$ of M_0, M^Λ, and and $\mathbf{P}_0, \mathbf{P}^\Lambda$ for the uniform distributions on these sets.

By construction, we have $d_H(M_0, M^\Lambda) = \Lambda$, and as in the proof of Lemma C.16, we get

$$\int_{\mathbb{R}^D} \mathbf{P}_0 \wedge \mathbf{P}^\Lambda \geq 1 - 3\frac{\Lambda}{\sigma}\left(\frac{\delta}{\tau_{min}}\right)^d.$$

Denoting $P_0^\sigma = \mathbf{P}_0$ and $P_1^\sigma = \mathbf{P}^\Lambda$ with $\Lambda = \varepsilon_{k,d,\tau_{min}}\delta^k$ and δ such that $3\frac{\Lambda}{\sigma}\left(\frac{\delta}{\tau_{min}}\right)^d$ yields the result.

□

C.5. Minimax Inconsistency Results

This section is devoted to the proof of Theorem 1, reproduced here as Theorem C.19.

THEOREM C.19. *Assume that $\tau_{min} = 0$. If $D \geq d+3$, then, for all $k \geq 2$ 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 $\hat{T} = \hat{T}\big(X_1, \ldots, X_n\big)$.

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 $n \geq 1$,

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

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

We will make use of Le Cam's Lemma, which we recall here.

THEOREM C.20 (Le Cam's Lemma [14]). *For all pairs P, P' in $\mathcal{P}$,*

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

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

PROOF OF THEOREM C.19. For $\delta \geq \Lambda > 0$, let $\mathcal{C}, \mathcal{C}' \subset \mathbb{R}^3$ be closed curves of the Euclidean space as in Figure 2, 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 (1) and 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 $M_1^{(0)}, M'^{(0)}_1$, similarly to Section 5.3.2 in [2], yielding, from Proposition A.4, that P_1, P_1' belong to $\mathcal{P}^k_{(x)}$ 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

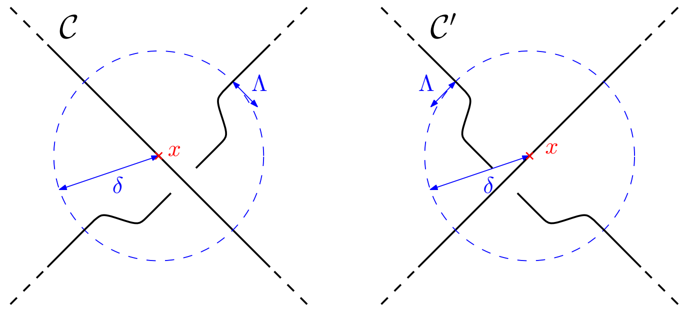


Figure 2: Hypotheses for minimax lower bound on tangent space estimation with $\tau_{min} = 0$.

only on d and k), and that Λ, δ and Λ^k/δ are small enough. From Le Cam's Lemma C.20, we have for all $n \geq 1$,

$$\inf_{\hat{T}} \sup_{P \in \mathcal{P}^k_{(x)}} \mathbb{E}_{P^{\otimes n}} \angle\big(T_x M, \hat{T}\big) \geq \frac{1}{2} \angle\big(T_x M_1, T_x M_1'\big) \left\| P_1 \wedge P_1' \right\|_1^n .$$

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

$$\begin{aligned} \left\| P_1 \wedge P_1' \right\|_1 &= 1 - Vol\left((\mathcal{B}_{\mathbb{R}^3}(0,\delta) \cap \mathcal{C}) \times \mathcal{S}^{d-1} \right) / Vol\left(\mathcal{C} \times \mathcal{S}^{d-1} \right) \\ &= 1 - Length\left(\mathcal{B}_{\mathbb{R}^3}(0,\delta) \cap \mathcal{C} \right) / Length(\mathcal{C}) \\ &\geq 1 - c_{L_\perp} \delta. \end{aligned}$$

Hence, at fixed $n \geq 1$, letting Λ, δ go to 0 with Λ^k/δ small enough, we get the announced bound.

We now tackle the lower bound on curvature estimation with the same strategy. Let $M_2, M_2' \subset \mathbb{R}^D$ be d-dimensional submanifolds as in Figure 3: 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 (1), centered at x. Outside $\mathcal{B}(x, \delta)$, M_2 and 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

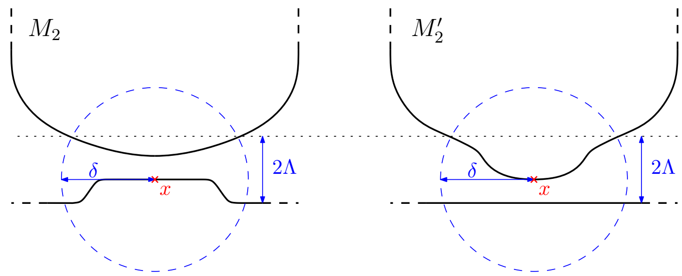


Figure 3: Hypotheses for minimax lower bound on curvature estimation with $\tau_{min} = 0$.

A.4. Hence from Le Cam's Lemma C.20 we deduce

$$\inf_{\widehat{II}} \sup_{P \in \mathcal{P}^k_{(x)}} \mathbb{E}_{P^{\otimes n}} \left\| II^M_x \circ \pi_{T_x M} - \widehat{II} \right\|_{op} \geq \frac{1}{2} \left\| II^{M_2}_x \circ \pi_{T_x M_2} - II^{M_2'}_x \circ \pi_{T_x M_2'} \right\|_{op} \left\| P_2 \wedge P_2' \right\|_1^n .$$

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

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

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

$$\left\| P_2 \wedge P_2' \right\|_1 = 1 - P_2(\mathcal{B}(x,\delta)) \geq 1 - c_{d,L_\perp} \delta^d.$$

At $n \geq 1$ fixed, letting Λ, δ go to 0 with Λ^k/δ small enough, we get the desired result.

□

REFERENCES

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

[2] Aamari, E. and Levrard, C. (2017). Non-asymptotic rates for manifold, tangent space and curvature estimation.

[3] Alexander, S. B. and Bishop, R. L. (2006). Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geom. Dedicata* **117** 65–84. MR2231159 (2007c:53110)

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

[5] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities*. Oxford University Press, Oxford A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR3185193

[6] Bousquet, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** 495–500. MR1890640 (2003f:60039)

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

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

[9] Dyer, R., Vegter, G. and Wintraecken, M. (2015). Riemannian simplices and triangulations. *Geom. Dedicata* **179** 91–138. MR3424659

[10] Federer, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. MR0110078 (22 ##961)

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

[12] Maggioni, M., Minsker, S. and Strawn, N. (2016). Multiscale dictionary learning: non-asymptotic bounds and robustness. *J. Mach. Learn. Res.* **17** Paper No. 2, 51. MR3482922

[13] Niyogi, P., Smale, S. and Weinberger, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441. MR2383768 (2009b:60038)

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

Department of Mathematics
University of California San Diego
9500 Gilman Dr. La Jolla
CA 92093
United States
E-mail: eaamari@ucsd.edu
URL: http://www.math.ucsd.edu/ eaamari/

Laboratoire de probabilités et modèles aléatoires
Bâtiment Sophie Germain
Université Paris-Diderot
75013 Paris
Paris, France
E-mail: levrard@math.univ-paris-diderot.fr
URL: http://www.normalesup.org/ levrard/