Proceedings of Machine Learning Research vol 75:1–32, 2018 31st Annual Conference on Learning Theory

Fitting a putative manifold to noisy data

Charles Fefferman CF@MATH.PRINCETON.EDU
Princeton University, Mathematics Department, Fine Hall, Washington Road, Princeton NJ, 08544-1000, USA.

Sergei Ivanov SVIVANOV@PDMI.RAS.RU
Steklov Institute of Mathematics, Russian Academy of Sciences, 27 Fontanka, 191023 St.Petersburg, Russia.

Yaroslav Kurylev Y.KURYLEV@UCL.AC.UK
University College London, Department of Mathematics, Gower Street WC1E 6BT, London, UK.

Matti Lassas MATTI.LASSAS@HELSINKI.FI
University of Helsinki, Department of Mathematics and Statistics, P.O. Box 68, 00014, Helsinki, Finland.

Hariharan Narayanan HARIHARAN.NARAYANAN@TIFR.RES.IN
School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, INDIA

Editors: Sebastien Bubeck, Vianney Perchet and Philippe Rigollet

Abstract

In the present work, we give a solution to the following question from manifold learning. Suppose data belonging to a high dimensional Euclidean space is drawn independently, identically distributed from a measure supported on a low dimensional twice differentiable embedded manifold $\mathcal{M}$, and corrupted by a small amount of gaussian noise. How can we produce a manifold $\mathcal{M}_o$ whose Hausdorff distance to $\mathcal{M}$ is small and whose reach is not much smaller than the reach of $\mathcal{M}$?

Keywords: Manifold learning, Hausdorff distance, reach

1. Introduction

One of the main challenges in high dimensional data analysis is dealing with the exponential growth of the computational and sample complexity of generic inference tasks as a function of dimension, a phenomenon termed "the curse of dimensionality". One intuition that has been put forward to lessen or even obviate the impact of this curse is that high dimensional data tend to lie near a low dimensional submanifold of the ambient space. Algorithms and analyses that are based on this hypotheses constitute the subfield of learning theory known as manifold learning; papers from this subfield include Belkin and Niyogi (2003); Carlsson (2009); Dasgupta and Freund (2008); Donoho and Grimes (2003); Fefferman et al. (2015, 2016); Genovese et al. (2012, 2014); Kégl et al. (2000); Narayanan and Niyogi (2009); Niyogi et al. (2008); Roweis and Saul (2000); Smola and Williamson (2001); Tenenbaum et al. (2000); Weinberger and Saul (2006). In the present work, we give a solution to the following question from manifold learning. Suppose data is drawn independently, identically distributed (i.i.d) from a measure supported on a low dimensional twice differentiable (C^2) manifold $\mathcal{M}$ whose reach is $\geq \tau$, and corrupted by amall amount of (i.i.d) gaussian noise. How can can we produce a manifold $\mathcal{M}_o$ whose Hausdorff distance to $\mathcal{M}$ is small and whose reach is not much smaller than τ?

© 2018 C. Fefferman, S. Ivanov, Y. Kurylev, M. Lassas & H. Narayanan.

This question is an instantiation of the problem of understanding the geometry of data. To give a specific real-world example, the issue of denoising noisy Cryo-electron microscopy (Cryo-EM) images falls into this category. Cryo-EM images are X-ray images of three-dimensional macromolecules, possessing an arbitrary orientation. The space of orientations is in correspondence with the Lie group $SO_3(\mathbb{R})$, which is only three dimensional. However, the ambient space of greyscale images on $[0,1]^2$ can be identified with an infinite dimensional subspace of $\mathcal{L}^2([0,1]^2)$, which gets projected down to a finite but very high dimensional subspace through the process of dividing $[0,1]^2$ into pixels. Thus noisy Cryo-EM X-ray images lie approximately on an embedding of a compact $3-$dimensional manifold in a very high dimensional space. If the errors are modelled as being gaussian, then fitting a manifold to the data can subsequently allow us to project the data onto this output manifold. Due to the large codimension and small dimension of the true manifold, the noise vectors are almost perpendicular to the true manifold and the projection would effectively denoise the data. The immediate rationale behind having a good lower bound on the reach is that this implies good generalization error bounds with respect to squared loss (See Theorem 1 in Fefferman et al. (2016)). Another reason why this is desirable is that the projection map onto such a manifold is Lipschitz within a tube of the manifold of radius equal to c times the reach for any c less than 1.

LiDAR (Light Detection and Ranging) also produces point cloud data for which the methods of this paper could be applied.

1.1. Model

Let $\mathcal{M}$ be a d dimensional $\mathcal{C}^2$ submanifold of $\mathbb{R}^D$. We assume $\mathcal{M}$ has volume ($d-$dimensional Hausdorff measure) equal to V, reach (i.e. normal injectivity radius) greater or equal to τ, and that $\mathcal{M}$ has no boundary. Let $x_1, \dots, x_N$ be a sequence of points chosen i.i.d at random from a measure μ absolutely continuous with respect to the d-dimensional Hausdorff measure $\mathcal{H}^d_{\mathcal{M}} = \lambda_{\mathcal{M}}$ on $\mathcal{M}$. More precisely, the Radon-Nikodym derivative $d\mu/d\lambda_{\mathcal{M}}$ satisfies

$$0 < \rho_{min} < d\mu/d\lambda_{\mathcal{M}} < \rho_{max} < \infty. \tag{1}$$

Let G_σ denote the Gaussian distribution supported on $\mathbb{R}^D$ whose density (Radon-Nikodym derivative with respect to the Lebesgue measure) at x is

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{D}{2}} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right).$$

Let $\zeta_1, \dots, \zeta_N$ be a sequence of i.i.d random variables independent of $x_1, \dots, x_N$ having the distribution G_σ. We observe $y_i = x_i + \zeta_i$ for $i = 1, 2, \dots$ and wish to construct a manifold $\mathcal{M}_o$ close to $\mathcal{M}$ in Hausdorff distance but at the same time having a reach not much less than τ. Note that the distribution of y_i (for each i), is the convolution of μ and G_σ. This is denoted by $\mu * G_\sigma$.

We observe $y_1, y_2, \dots, y_N$ and will produce a description of a manifold $\mathcal{M}_o$ such that for σ satisfying a certain upper bound, the Hausdorff distance between $\mathcal{M}$ and $\mathcal{M}_o$ is at most $O(\sigma)$ and $\mathcal{M}_o$ has reach that is bounded below by $\frac{c\tau}{d^7}$.

1.2. Prior work

The question of fitting a manifold to data is of interest to data analysts and statisticians. While there are several results dealing exclusively with sample complexity such as Genovese et al. (2012);

Narayanan and Mitter (2010), we will restrict our attention to results that provide an algorithm for describing a manifold to fit the data together with upper bounds on the sample complexity.

A work in this direction, Genovese et al. (2014), building over Ozertem and Erdogmus (2011) provides an upper bound on the Hausdorff distance between the output manifold and the true manifold equal to $O((\frac{\log N}{N})^{\frac{2}{D+8}}) + O(\sigma^2 \log(\sigma^{-1}))$. Note that in order to obtain a Hausdorff distance of $c\epsilon$, one needs more than $\epsilon^{-D/2}$ samples, where D is the ambient dimension. The results of this paper guarantee (for sufficiently small σ,) a Hausdorff distance of

$$Cd^7(\sigma\sqrt{D}) = O(\sigma)$$

with less than

$$\frac{CV}{\omega_d(\sigma\sqrt{D})^d} = O(\sigma^{-d})$$

samples. Thus while the asymptotic bound on the Hausdorff distance is $O(\sigma)$ which is worse than $\tilde{O}(\sigma^2)$, the number of samples needed to get there depends exponentially on the intrinsic dimension d. We note here that using Principal Component analysis, it is possible to replace D by a quantity that is at most exponential in d^2, however we will present these details elsewhere.

The question of fitting a manifold $\mathcal{M}_o$ to data with control both on the reach τ_o and mean squared distance of the data to the manifold was considered in Fefferman et al. (2016). However, Fefferman et al. (2016) did not assume a generative model for the data, and had to use an exhaustive search over the space of candidate manifolds whose time complexity was doubly exponential in the intrinsic dimension d of $\mathcal{M}_o$. In the present paper the construction of $\mathcal{M}_o$ has a sample complexity that is singly exponential in d, made possible by the generative model. The time complexity of our construction is governed by the complexity of constructing a set of weights as described in Lemma 6, which is bounded above by $N(Cd)^{2d}$, N being the sample compexity. Also, Fefferman et al. (2016) did not specify the bound on τ_o, beyond stating that the multiplicative degradation $\frac{\tau}{\tau_o}$ in the reach depended on the intrinsic dimension alone. In this paper, we pin down this degradation to within $(0, Cd^7]$, where C is an absolute constant and d is the dimension of $\mathcal{M}$.

The paper Mohammed and Narayanan (2017) assumes a generative model with no noise and proves that two algorithms related to the algorithm in Fefferman et al. (2016), can fit manifolds whose Hausdorff distance is arbitrarily close to 0 to the data but the guaranteed lower bound on the reach degrades to 0 as the Hausdorff distance tends to 0.

Finally, we mention that there is an interesting body of literature (Boissonnat et al. (2009); Cheng et al. (2005)) in computational geometry that deals with fitting piecewise linear manifolds (as opposed to C^2-manifolds) to data . Cheng et al. (2005) presented the first algorithm for arbitrary d, that takes samples from a smooth $d-$dimensional manifold $\mathcal{M}$ embedded in an Euclidean space and outputs a simplicial manifold that is homeomorphic and close in Hausdorff distance to M.

2. Some basic estimates and definitions.

2.1. A note on constants

In this section, and the following sections, we will make frequent use of constants $c, C, C_1, C_2, \bar{c}_1$ etc. These constants are absolute and positive, and may change in their value from line to line. Also, the value of a constant can depend on the values of constants defined before it, but not those defined after it. This convention eliminates the possibility of circularity. We will use upper case letters to denote constants greater than 1 and lower case to denote constants less than 1.

We need the following form of the Gaussian concentration inequality. Let G_σ, as stated earlier, be the distribution supported on $\mathbb{R}^D$ whose density at x is given by $\left(\frac{1}{2\pi}\right)^{\frac{D}{2}} \exp\left(-\frac{\|x\|^2}{2}\right)$. Let $g : \mathbb{R}^D \to \mathbb{R}$ be a $1-$Lipschitz function and $a = \mathbb{E}g(x)$ be the average value of g with respect to G_σ. Then,

$$G_\sigma\{x : |g(x) - a| \geq t\sigma\} \leq C \exp\left(-ct^2\right). \tag{2}$$

for some absolute constants c, C.

Definition 1 (Tangent and Normal Space)
For a closed $A \subseteq \mathbb{R}^D$, and $a \in A$, let the tangent space (in the sense of Federer) $Tan^0(a, A)$ denote the set of all vectors v such that for all $\epsilon > 0$, there exists $b \in A$ such that $0 < |a - b| < \epsilon$ and $\left|v/|v| - \frac{b-a}{|b-a|}\right| < \epsilon$. Let the normal space $Nor^0(a, A)$ denote the set of all v such that for all $w \in Tan^0(a, A)$, we have $\langle v, w\rangle = 0$. Let $Tan(a, A)$ (or $Tan(a)$ when A is clear from context) denote the set of all x such that $x - a \in Tan^0(a, A)$. For a set $X \subseteq \mathbb{R}^D$ and a point $p \in \mathbb{R}^D$, let $dist(p, X)$ denote the Euclidean distance of the nearest point in X to p. Let $Nor(a, A)$ (or $Nor(a)$ when A is clear from context) denote the set of all x such that $x - a \in Nor^0(a, A)$.

Definition 2 (Reach) *The reach of a closed set $A \subseteq \mathbb{R}^D$, denoted $reach(A)$, is the supremum of all r satisfying the following property. If $dist(p, A) \leq r$, then there exists a unique $q \in A$ such that $|p - q| = dist(p, A)$.*

For a smooth submanifold, the reach is the size of the largest neighborhood where the tubular coordinates near the submanifold are defined.

The following result of Federer (Theorem 4.18, Federer (1959)), gives an alternate characterization of the reach.

Proposition 3 *Let A be a closed subset of $\mathbb{R}^D$. Then,*

$$\mathrm{reach}(A)^{-1} = \sup\left\{2|b - a|^{-2} dist(b, Tan(a)) \,\middle|\, a, b \in A, a \neq b\right\}. \tag{3}$$

Definition 4 *We say $\mathcal{M}$ is a $d-$dimensional C^2-submanifold of $\mathbb{R}^D$ if $\mathcal{M}$ is compact and for every point $p \in \mathcal{M}$ there is a neigborhood $U \subset \mathbb{R}^D$ of p, a convex open $W \in \mathbb{R}^d$ and $\mathcal{C}^2$ functions $\phi : U \to W$, $\psi : W \to U$ such that $\mathcal{M} \cap U = \psi(W)$ and $\phi \circ \psi$ is the identity map on W.*

3. Preliminary structures

Definition 5 ($\epsilon-$net) *Let $(X, dist)$ be a metric space. We say that X_1 is an $\epsilon-$net of X, if $X_1 \subseteq X$ and for every $x \in X$, there is an $x_1 \in X_1$ such that $dist(x, x_1) < \epsilon$.*

Let

$$r \in \left[\sqrt{\sigma\tau}D^{1/4}, \frac{\tau}{Cd^C}\right]. \tag{4}$$

Let N_0 be chosen to be an integer such that

$$N_0/\ln(N_0) > \frac{CV}{\rho_{min}\omega_d(r^2/\tau)^d}, \tag{5}$$

where ω_d is the volume of a Euclidean unit ball in $\mathbb{R}^d$. We will assume that V and σ are sufficiently small that we can choose r and N_0 such that

$$N_0 \leq e^D. \tag{6}$$

By the coupon collector applied to the Voronoi cells corresponding to a $6r^2/\delta-$net of $\mathcal{M}$ that is also $r^2/2\delta-$separated (such a net always exists, and can be constructed by a greedy procedure), we see that if one chooses a set of N_0 i.i.d random samples $\tilde{X_0}$ from μ, with probability at least $1 - N_0^{-C}$, every Voronoi cell has at least one random sample. Therefore the Hausdorff distance of $\tilde{X_0}$ to $\mathcal{M}$ is less than $12r^2/\delta$. The maximum distance of a point y_i of X_0 to the corresponding x_i in $\tilde{X_0}$ is bounded above with probability at least $1 - N_0^{-C}$ by

$$\sigma(\sqrt{D} + \sqrt{\ln(N_0^C)}) < Cr^2/\tau.$$

This is due to Gaussian concentration. This is an upper bound on the Hausdorff distance between X_0 and $\tilde{X_0}$. Therefore, with probability $1 - N_0^{-C}$, if one chooses a set X_0 of N_0 i.i.d random samples from $\mu * G_\sigma$, X_0 will be $Cr^2/\tau-$close to $\mathcal{M}$ in Hausdorff distance.

Then, let $X_1 = \{p_i\}$ be a minimal $cr/d-$net of X_0. Such a net can be chosen greedily, ensuring at every step that no element included in the net thus far is within $\frac{cr}{2d}$ of the point currently chosen. The process continues while progress is possible. Let the size of X_1 be denoted $\bar{N}$.

We introduce a family of D dimensional balls of radius r, $\{U_i\}_{i\in[\bar{N}]}$ where the center of U_i is p_i and a family of $d-$dimensional embedded discs of radius r, denoted $\{D_i\}_{i\in[\bar{N}]}$, $D_i \subseteq U_i$ where D_i is centered at p_i. The D_i are chosen by fitting a disc that approximately minimizes among all discs of radius r centered at p_i the Hausdorff distance to $U_i \cap X_0$ by a procedure described in Subsection C.1. We will need the following properties of (D_i, p_i), which hold with probability $1 - N_0^{-C}$:

1. The Hausdorff distance between $\cup_i D_i$ and $\mathcal{M}$ is less than $\frac{Cdr^2}{\tau} = \delta$.
2. For any $i \neq j$, $|p_i - p_j| > \frac{cr}{d}$.
3. For every $z \in \mathcal{M}$, there exists a point p_i such that $|z - p_i| < 3\inf_{i\neq j}, |p_i - p_j|$.

Consider the bump function $\tilde{\alpha}_i : \mathbb{R}^D \to \mathbb{R}$ given by

$$\tilde{\alpha}_i(p_i + rv) = c_i(1 - \|v\|^2)^{d+2}$$

for any $v \in B_D$ and 0 otherwise. Let

$$\tilde{\alpha}(x) := \sum_i \tilde{\alpha}_i(x).$$

Let

$$\alpha_i(x) = \frac{\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)},$$

for each i.

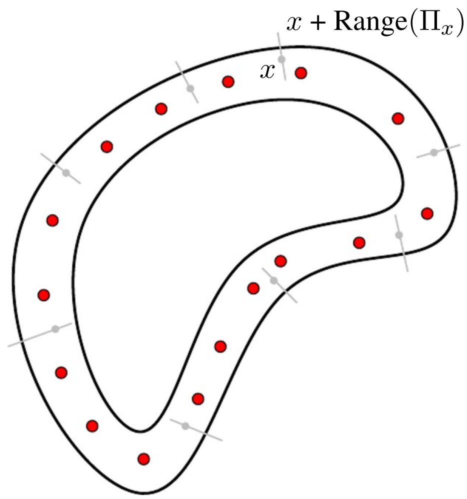

Figure 1: A vector bundle over a neighborhood of the data, used to produce the output manifold. In the figure, x is the gray point and x + Range(Π_x) is the gray line segment containing x.

Lemma 6 *It is possible to choose c_i such that for any z in a $\frac{r}{4d}$ neighborhood of $\mathcal{M}$,*

$$c^{-1} > \tilde{\alpha}(z) > c,$$

where c is a small universal constant. Further, such c_i can be computed using no more than $N_0(Cd)^{2d}$ operations involving vectors of dimension D.

Proof The details of the proof appear in the appendix in section D. ■

4. The output manifold

For the course of this section, we consider the scaled setting where $r = 1$. Thus, in the new Euclidean metric, $\tau \geq Cd^C$.

Let Π^i be the orthogonal projection onto the $n - d-$dimensional subspace containing the origin that is orthogonal to the affine span of D_i.

We define the function $F_i : U_i \to \mathbb{R}^n$ by $F_i(x) = \Pi^i(x - p_i)$. Let $\cup_i U_i = U$. We define

$$F : U \to \mathbb{R}^n$$

by $F(x) = \sum_i \alpha_i(x) F_i(x)$.

Given a symmetric matrix A such that A has $n - d$ eigenvalues in $(1/2, 3/2)$ and d eigenvalues in $(-1/2, 1/2)$, let $\Pi_{hi}(A)$ denote the projection onto the span of the eigenvectors corresponding to the largest $n - d$ eigenvalues.

For $x \in \cup_i U_i$, we define $\Pi_x = \Pi_{hi}(A_x)$ where $A_x = \sum_i \alpha_i(x)\Pi^i$. Let $\tilde{U}_i$ be defined as the $\frac{cr}{d}-$Eucidean neighborhood of D_i intersected with U_i. Note that Π_x is $\mathcal{C}^2$ when restricted to $\cup_i \tilde{U}_i$, because the $\alpha_i(x)$ are C^2 and when x is in this set, $c < \sum_i \tilde{\alpha}_i(x) < c^{-1}$, and for any i, j such that $\alpha_i(x) \neq 0 \neq \alpha_j(x)$, we have $\|\Pi^i - \Pi^j\|_F < Cd\delta$.

Definition 7 *The output manifold $\mathcal{M}_o$ is the set of all points $x \in \cup_i \tilde{U}_i$ such that $\Pi_x F(x) = 0$.*

As stated above, $\mathcal{M}_o$ is the set of points $x \in \cup_i \tilde{U}_i$ such that

$$\Pi_{hi}(\sum_i \alpha_i(x)\Pi^i)(\sum_i \alpha_i(x)\Pi^i(x-p_i)) = 0. \tag{7}$$

Using diagonalization and Cauchy's integral formula, we have

$$\frac{1}{2\pi i}\left[\oint_\gamma (zI - (\sum_i \alpha_i(x)\Pi^i))^{-1} dz\right]\left(\sum_i \alpha_i(x)\Pi^i(x-p_i)\right) = 0 \tag{8}$$

where γ is the circle of radius $1/2$ centered at 1.

Let $\sum \alpha_i(x)\Pi^i = M(x)$, and as stated earlier, $\Pi^i(x-p_i) = F_i(x)$. Let $\Pi_{hi}(M(x))$ be denoted Π_x.

Then the left hand side of (8) can be written as

$$\oint_\gamma \frac{dz}{2\pi i}\left(\sum_i \alpha_i(x)(zI - M(x))^{-1}F_i(x)\right). \tag{9}$$

for any $v \in \mathbb{R}^{\hat{n}}$ and and $f : \mathbb{R}^{\hat{n}} \to \mathbb{R}^{\tilde{n}}$ where $\hat{n}, \tilde{n} \in \mathbb{N}_+$ let

$$\partial_v f(x) := \lim_{\alpha \to 0} \frac{f(x+\alpha v) - f(x)}{\alpha}.$$

$$\begin{aligned}
\partial_v \oint_\gamma \frac{dz}{2\pi i}\left(\sum_i \alpha_i(x)(zI - M(x))^{-1}F_i(x)\right) &= \sum_i \alpha_i(x)\Pi_x(\partial_v F_i(x)) && (10)\\
&+ \sum_i \alpha_i(x)(\partial_v \Pi_x)F_i(x) && (11)\\
&+ \sum_i (\partial_v \alpha_i(x))\Pi_x F_i(x). && (12)
\end{aligned}$$

Let $\|v\| = 1$. Let $\mathcal{M}_\Pi^d$ denote the set of all projection matrices of rank d. This is an analytic submanifold of the space of $n \times n$ matrices.

Claim 1 *The reach of $\mathcal{M}_\Pi^d$ is greater or equal to* $1/2$.

Proof This proof appears in Section E of the appendix. ■

We first look at the right hand side of (10). This can be rewritten as

$$\sum_i \alpha_i(x)\Pi_x\Pi_i v = \Pi_x v + \Pi_x (M(x) - \Pi_x) v. \tag{13}$$

$$
\begin{aligned}
\|M(x) - \Pi_x\|_F &= dist(M(x), Tan(\Pi_x, \mathcal{M}^d_\Pi)) && (14)\\
&\leq \sup_i dist(\Pi_i, Tan(\Pi_x, \mathcal{M}^d_\Pi)) && (15)\\
&\leq \sup_i \|\Pi_i - \Pi_x\|_F^2/(2\, reach(\mathcal{M}^d_\Pi)) && (16)\\
&\leq 4 \sup_{i,j} \|\Pi_i - \Pi_j\|_F^2 && (17)\\
&\leq 8d\delta^2. && (18)
\end{aligned}
$$

We look at (11) next. Observe that

$$
\left\| \oint_\gamma \frac{dz}{2\pi i} \left(\sum_i \alpha_i(x) \left(\partial_v((zI - M(x))^{-1})\right) F_i(x) \right) \right\| \quad \leq \quad \|\partial_v \Pi_x\| \left\| \sum_i \alpha_i(x) F_i(x) \right\|.
$$

In what follows, we will make repeated use of Hölder's inequality: Let $p, q \in \mathbb{R}$ and $\frac{1}{p} + \frac{1}{q} = 1$, then,

$$
\forall x, y \in \mathbb{R}^n, \langle x, y\rangle \leq \|x\|_p \|y\|_q.
$$

Secondly, we will use the fact that for any ball U_i, the number of j such that $U_i \cap U_j$ is nonempty is bounded above by $(Cd)^d$ because of the lower bound of $\frac{cr}{d}$ on the spacing between the p_i and p_j for any two distinct i and j. A consequence of this is that any vector $v \in \mathbb{R}^{\bar{N}}$ that is supported on the set of all j such that $U_i \cap U_j \neq \emptyset$ will satisfy

$$
\|v\|_{d+2} \leq Cd\|v\|_\infty, \tag{19}
$$

and

$$
\|v\|_{\frac{d+2}{2}} \leq Cd^2 \|v\|_\infty. \tag{20}
$$

Thirdly, we will use the following bounds on the derivatives of the bump functions at points x that are within a distance of cr/d of $\mathcal{M}$. Recall that $\sum_i \tilde{\alpha}_i(x)$ is denoted $\tilde{\alpha}(x)$. Then we know that $c < \tilde{\alpha}(x) < C$ if the distance of x from $\mathcal{M}$ is less than cr/d.

Lemma 8 *We have for any $v \in \mathbb{R}^D$ such that $|v| = 1$, and any $x \in \mathbb{R}^D$ such that $dist(x, \mathcal{M}) \leq \frac{cr}{d}$,*

$$
\|(\partial_v \alpha_i(x))_{i \in [\bar{N}]}\|_{\frac{d+2}{d+1}} \leq Cd^2. \tag{21}
$$

Proof This proof appears in section F in the Appendix. ■

Lemma 9 *We have for any $v \in \mathbb{R}^D$ such that $|v| = 1$, and any $x \in \mathbb{R}^D$ such that $dist(x, \mathcal{M}) \leq \frac{cr}{d}$,*

$$
\|(\partial_v^2 \alpha_i(x))_{i \in [\bar{N}]}\|_{\frac{d+2}{d}} \leq Cd^4. \tag{22}
$$

Proof This proof appears in section F in the Appendix. ■

Recall that $F(x) = \sum \alpha_i(x) F_i(x)$.

4.1. A bound on the first derivative of $\Pi_x F(x)$

We proceed to obtain an upper bound on $\|\partial_v \Pi_x\|$.

$$
\begin{aligned}
\|\partial_v \Pi_x\| &\leq (radius(\gamma)) \left\|\partial_v((zI - M(x))^{-1})\right\| && (23)\\
&= \left(\frac{1}{2}\right) \|(zI - M(x))^{-1}\partial_v M(x)(zI - M)^{-1}\| && (24)\\
&\leq \left(\frac{1}{2}\right) \|(zI - M(x))\|^{-2}\|\partial_v M(x)\| && (25)\\
&\leq 8\|\partial_v M(x)\| && (26)\\
&= 8\| \sum_i \partial_v \alpha_i(x)(\Pi^i - \Pi^1) + \partial_v \sum_i \alpha_i(x)\Pi_1\| && (27)\\
&\leq 8\sum_i |\partial_v \alpha_i(x)|\delta + 0 && (28)\\
&\leq 8\|(\partial_v \alpha_i(x))_{i\in[\bar{N}]}\|_{\frac{d+2}{d+1}}\|(\delta)_{i\in[\bar{N}]}\|_{d+2} && (29)\\
&\leq Cd^3\delta. && (30)
\end{aligned}
$$

where C is an absolute constant.

Therefore,

$$
\left\|\oint_\gamma \frac{dz}{2\pi i}\left(\sum_i \alpha_i(x)\left(\partial_v((zI - M(x))^{-1})\right) F_i(x)\right)\right\| \leq Cd^3\delta. \qquad (31)
$$

Finally, we bound (12) from above.

$$
\begin{aligned}
\left\|\oint_\gamma \frac{dz}{2\pi i}\left(\sum_i (\partial_v \alpha_i(x))\,(zI - M(x))^{-1}F_i(x)\right)\right\| &\leq \left\|\Pi_x\left(\sum_i (\partial_v \alpha_i(x))(F_i(x) - F_1(x))\right)\right\|\\
&+ \left\|\left(\sum_i \partial_v \alpha_i(x)\right) F_1(x)\right\| && (32)\\
&\leq \|\Pi_x\| \sum_i |\partial_v \alpha_i(x)|\|F_i(x) - F_1(x)\| + 0 && (33)\\
&\leq \|(\partial_v \alpha_i(x))_{i\in[\bar{N}]}\|_{\frac{d+2}{d+1}}\|(F_i(x) - F_1(x))_{i\in[\bar{N}]}\|_{d+2}\\
&\leq Cd^3\delta. && (34)
\end{aligned}
$$

Therefore

$$
\|\partial_v\left(\Pi_x F(x)\right) - \Pi_x v\| \leq Cd^3\delta. \qquad (35)
$$

Note also by (23)-(30) that

$$
\|\partial_v \Pi_x\| \leq Cd^3\delta. \qquad (36)
$$

4.2. A bound on the second derivative of $\Pi_x F(x)$

We now proceed to obtain an upper bound on $\|\partial_v^2 \left(\Pi_x F(x)\right)\|$.

$$
\begin{aligned}
\|\partial_v^2 \left(\Pi_x F(x)\right)\| &\leq \|(\partial_v^2 \Pi_x) F(x)\| && (37)\\
&+ \|2(\partial_v \Pi_x)\partial_v F(x)\| && (38)\\
&+ \|\Pi_x \partial_v^2 F(x)\|. && (39)
\end{aligned}
$$

We first bound from above the right side of (37).

$$
\begin{aligned}
(\partial_v^2 \Pi_x) &= \partial_v^2 \left[\frac{1}{2\pi i}\oint [zI - M(x)]^{-1} dz\right] && (40)\\
&= \partial_v \left[\frac{1}{2\pi i}\oint (zI - M(x))^{-1}\partial_v M(x)(zI - M(x))^{-1} dz\right] && (41)\\
&= \frac{1}{2\pi i}\oint 2(zI - M(x))^{-1}\partial_v M(x)(zI - M(x))^{-1}\partial_v M(x)(zI - M(x))^{-1} dz\\
&+ \oint (zI - M(x))^{-1}\partial_v^2 M(x)(zI - M(x))^{-1} dz. && (42)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\|(\partial_v^2 \Pi_x) F(x)\| &\leq \sup_{z\in\gamma}\left(C\|(zI - M(x)\|^{-3}\|(\partial_v M(x))^2\| + C\|(zI - M(x)\|^{-2}\|(\partial_v^2 M(x))\|\right)\\
&\leq C(\|\partial_v M(x)\|^2 + \|\partial_v^2 M(x)\|) && (43)\\
&\leq Cd^6\delta^2 + \|\partial_v^2 M(x)\| && (44)\\
&= Cd^6\delta^2 + \|\partial_v^2 \sum_i \alpha_i(x)\Pi_i\| && (45)\\
&\leq Cd^6\delta^2 + \sum_i |\partial_v^2 \alpha_i(x)(\Pi_i - \Pi_1)| && (46)\\
&\leq Cd^6\delta^2 + \|(\partial_v^2 \alpha_i(x))_i\|_{\frac{d+2}{d}}\|(\delta)_i\|_{\frac{d+2}{2}} && (47)\\
&\leq Cd^6\delta^2 + Cd^6\delta. && (48)
\end{aligned}
$$

Next, we bound (38) from above. Note that

$$
\begin{aligned}
\|\partial_v F(x)\| &\leq \|(\sum_i \partial_v \alpha_i(x))(F_i(x) - F_1(x)) + \Pi_x v\| && (49)\\
&\leq 1 + Cd^3\delta. && (50)
\end{aligned}
$$

$$
\begin{aligned}
\|(\partial_v \Pi_x)\partial_v F(x)\| &\leq \|(\partial_v \Pi_x)\|\|\partial_v F(x)\| && (51)\\
&\leq (Cd^3\delta)(1 + d^3\delta) && (52)\\
&= Cd^3\delta + Cd^6\delta^2. && (53)
\end{aligned}
$$

Finally, we bound (39) from above.

$$
\begin{aligned}
\|\Pi_x \partial_v^2 F(x)\| &\leq \|\partial_v^2 F(x)\| && (54)\\
&\leq \|\partial_v^2 (F(x) - F_1(x))\| && (55)\\
&\leq \sum_i |\partial_v^2 \alpha_i(x)| \|F_i(x) - F_1(x)\| && (56)\\
&+ \sum_i 2|\partial_v \alpha_i(x)| \|\partial_v F_i(x) - \partial_v F_1(x)\| && (57)\\
&+ \sum_i |\alpha_i(x)| \|\partial_v^2 F_i(x)\|. && (58)
\end{aligned}
$$

We first bound (56) from above.

$$
\begin{aligned}
\sum_i |\partial_v^2 \alpha_i(x)| \|F_i(x) - F_1(x)\| &\leq \|(|\partial_v^2 \alpha_i(x)|)_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} \|(\|F_i(x) - F_1(x)\|)_{i\in[\bar{N}]}\|_{\frac{d+2}{2}}\\
&\leq (Cd^4)(d^2\delta) && (59)\\
&= Cd^6\delta. && (60)
\end{aligned}
$$

Next we bound (57) from above.

$$
\begin{aligned}
\sum_i |\partial_v \alpha_i(x)| \|\partial_v F_i(x) - \partial_v F_1(x)\| &\leq \|(|\partial_v \alpha_i(x)|)_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} \|(\|\partial_v F_i(x) - \partial_v F_1(x)\|)_{i\in[\bar{N}]}\|_{\frac{d+2}{2}}\\
&\leq Cd^4(d^2\delta) && (61)\\
&= Cd^6\delta. && (62)
\end{aligned}
$$

The term (58) is 0.
Therefore

$$
\left\|\partial_v^2 \left(\Pi_x F(x)\right)\right\| \leq Cd^6\delta. \tag{63}
$$

Recall that $\mathcal{M}_o$ is the set of points $x \in \cup_i \tilde{U}_i$ such that

$$
\Pi_{hi}(\sum_i \alpha_i(x)\Pi^i)(\sum_i \alpha_i(x)\Pi^i(x - p_i)) = 0. \tag{64}
$$

In particular, $x \in \mathcal{M}_o \cap U_i$ if and only if $h(z) = \Pi_i \Pi_{hi}(\sum_i \alpha_i(z)\Pi^i)(\sum_i \alpha_i(z)\Pi^i(z - p_i)) = 0$, where Π^i is the orthogonal projection onto the subspace orthogonal to D_i, containing the center of D_i. We take U_i to be the unit ball and the center of U_i to be the origin and take the linear span of D_i to be $\mathbb{R}^d$. We split z into its x component (projection onto $\mathbb{R}^d$) and y component (projection orthogonal to $\mathbb{R}^d$). and define $g(x, y) = (x, h(x, y))$. This function is then substituted into the quantitative inverse function theorem of section B in the appendix.

5. Hausdorff distance of $\mathcal{M}_o$ to $\mathcal{M}$ and the reach of $\mathcal{M}_o$.

Theorem 10 *With probability at least $1 - N_0^{-C}$, the following is true. The Hausdorff distance between $\mathcal{M}_o$ and $\mathcal{M}$ is less than $Cd^3\sqrt{r\delta} < \frac{C\tau}{d^C}$ and the reach of $\mathcal{M}_o$ is at least $\frac{\tau}{Cd^7}$.*

Proof Since $\sigma(\sqrt{D} + \sqrt{\ln(N_0^C)})$ is less than $Cr^2/\tau < \frac{\tau}{Cd^C}$, by Gaussian concentration, with probability at least $1 - N_0^{-C}$, every point $y_i = x_i + \zeta_i$ satisfies

$$|y_i - x_i| = |\zeta_i| < \sigma(\sqrt{D} + \sqrt{\ln(N_0^C)}) < Cr^2/\tau.$$

All the statements for the rest of this proof will hold with probability at least $1-N_0^{-C}$. Since X_0 has a Hausdorff distance $\frac{Cr^2}{\tau}$ to $\mathcal{M}$, and X_1 has a Hausdorff distance $\frac{cr}{d}$ to $\mathcal{M}$, as a consequence, the Hausdorff distance between $\cup_i D_i$ and $\mathcal{M}$ is less than $\delta = \frac{Cdr^2}{\tau}$ by subsection C.1. The Hausdorff distance between $\cup_i D_i$ and $\mathcal{M}_o$ is less than

$$Cd^6\delta = Cd^7\frac{r^2}{\tau} < \frac{C\tau}{d^C} \tag{65}$$

by the quantitative implicit function theorem (appendix, section B), applying Taylor's theorem together with (122) and (125). Thus, by the triangle inequality, the Hausdorff distance between $\mathcal{M}_o$ and $\mathcal{M}$ is less than $Cd^7\frac{r^2}{\tau} < \frac{C\tau}{d^C}$. Next, we address the reach of $\mathcal{M}_o$. By Federer's criterion (Proposition 3) we know that

$$reach(\mathcal{M}_o) = \inf\left\{|b-a|^2(2dist(b, Tan(a)))^{-1}\middle|\, a, b \in \mathcal{M}_o, a \neq b\right\}. \tag{66}$$

We wish to prove that $reach(\mathcal{M}_o) > \frac{\tau}{Cd^7}$. Let $a, b \in \mathcal{M}_o, a \neq b$.

If $|a-b| > \frac{\tau}{Cd^7}$, then $|b-a|^2(2dist(b, Tan(a)))^{-1} > \frac{\tau}{Cd^7}$, because

$$|b-a| \geq dist(b, Tan(a)).$$

Therefore, we may suppose that $|a-b| \leq \frac{\tau}{Cd^7}$. By the bound on the Hausdorff distance between $\mathcal{M}$ and $\mathcal{M}_o$, the respective distances of a and b to the images of their projections onto $\mathcal{M}$, which we denote a' and b' respectively, are less than $Cd^3\sqrt{r\delta}$. By the quantitative implicit function theorem (appendix, section B) and the covering property of $\{U_i\}$, $\mathcal{M}_o$ is a C^2-submanifold of $\mathbb{R}^n$. Therefore $Tan(a)$ is a $d-$dimensional affine subspace. By (122) and (125) the Hausdorff distance between the two unit discs $Tan(a) \cap B(a, 1)$ and $(Tan_{\mathcal{M}}(a') \cap B(a', 1)) + (a - a')$ which are centered at a, is bounded above by

$$Cd^3\sqrt{\delta/r} = Cd^{7/2}\sqrt{r/\tau} = \frac{1}{Cd^C}. \tag{67}$$

Then, $\mathcal{M}$ and $\mathcal{M}_o$ are $Cd^3\sqrt{r\delta}$ close in Hausdorff distance and $(\mathcal{M}_o \cap B(a, 2|a-b|))$ and $(\mathcal{M} \cap B(a', 2|a-b|))$ are $Cd^3\sqrt{\delta/r}|a-b| = \frac{1}{Cd^C}$ close in C^1 over the maximal subset U^a of $Tan(a)$ on which both are defined as graphs of functions (respectively $\hat{f}_o$ and $\hat{f}$) whose range is $Nor(a)$, the fiber of the normal bundle at a by Lemma 13; please see section A in the appendix. This subset contains $(Tan(a) \cap B(a, 3|a-b|/2))$ because we further know by (67) and the reach of $\mathcal{M}$ being at least τ, that the C^1 norm of $\hat{f}$ on $(Tan(a) \cap B(a, 3|a-b|/2))$ is at most $(Cd^7)^{-1}$. Therefore, the C^1 norm of $\hat{f}_o$ on $(Tan(a) \cap B(a, 3|a-b|/2))$ is at most $(Cd^7)^{-1}$. But using this and the Hessian bound of $Cd^6\delta$ from (63), we also know that the Hessian of $\hat{f}_o$ is bounded above by Cd^7/τ. But now, by Taylor's theorem, $dist(b, Tan(a)) \leq \sup \|Hess\hat{f}_o\||a-b|^2/2$, where the supremum is taken over $(Tan(a) \cap B(a, 3/2|a-b|))$. This, we know is bounded above by $Cd^7|a-b|^2/\tau$. Substituting this into Federer's criterion for the reach, we see that

$$reach(\mathcal{M}_o) \geq \frac{\tau}{Cd^7}.$$

■

Acknowledgments

Ch.F. was partly supported AFOSR, grant DMS-1265524, and NSF, grant FA9550- 12-1-0425. S.I. was partly supported RFBR, grant 14-01-00062, Y.K. was partly supported by EPSRC and the AXA professorship, M.L. was supported by Academy of Finland, grants 273979 and 284715, and H.N. was partly supported by NSF grant DMS-1620102 and a Ramanujan Fellowship.

References

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, 2003. ISSN 0899-7667. doi: http://dx.doi.org/10.1162/089976603321780317.

Jean-Daniel Boissonnat, Leonidas J. Guibas, and Steve Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. *Discrete & Computational Geometry*, 42(1): 37–70, 2009. doi: 10.1007/s00454-009-9175-1. URL http://dx.doi.org/10.1007/s00454-009-9175-1.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, January 2009. doi: 10.1090/S0273-0979-09-01249-X. URL http://www.ams.org/bull/2009-46-02/S0273-0979-09-01249-X/home.html.

Siu-Wing Cheng, Tamal K. Dey, and Edgar A. Ramos. Manifold reconstruction from point samples. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 1018–1027, 2005. URL http://dl.acm.org/citation.cfm?id=1070432.1070579.

Sanjoy Dasgupta and Yoav Freund. Random projection trees and low dimensional manifolds. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 537–546, 2008.

David L. Donoho and Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591–5596, May 2003. doi: 10.1073/pnas.1031596100. URL http://dx.doi.org/10.1073/pnas.1031596100.

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93, 1959.

Herbert Federer. *Geometric Measure Theory*. Springer New York, 1969.

Charles Fefferman, Sergei Ivanov, Yaroslav Kurylev, Matti Lassas, and Hariharan Narayanan. Reconstruction and interpolation of manifolds i: the geometric whitney problem. *CoRR*, abs/1508.00674, 2015. URL http://arxiv.org/abs/1508.00674.

Charles L. Fefferman, Sanjoy K. Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the Americal Mathematical Society*, 2016.

Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Manifold estimation and singular deconvolution under hausdorff loss. *Annals of Statistics*, 40(2), 2012.

Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Nonparametric ridge estimation. *Ann. Statist.*, 42(4):1511–1545, 2014. doi: 10.1214/14-AOS1218. URL `http://dx.doi.org/10.1214/14-AOS1218`.

Balázs Kégl, Adam Krzyzak, Tamás Linder, and Kenneth Zeger. Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:281–297, 2000. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/34.841759.

Kitty Mohammed and Hariharan Narayanan. Manifold learning using kernel density estimation and local principal components analysis. *CoRR*, abs/1709.03615., 2017. URL `http://arxiv.org/abs/1709.03615.`

R. Narasimhan. *Lectures on Topics in Analysis*. Tata Institute of Fundamental Research, Bombay, 1965.

Hariharan Narayanan and Sanjoy Mitter. On the sample complexity of testing the manifold hypothesis. In *NIPS*, 2010.

Hariharan Narayanan and Partha Niyogi. On the sample complexity of learning smooth cuts on a manifold. In *Proc. of the 22nd Annual Conference on Learning Theory (COLT)*, June 2009.

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

Umut Ozertem and Deniz Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12:1249–1286, 2011.

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

Alexander J. Smola and Robert C. Williamson. Regularized principal manifolds. In *In Computational Learning Theory: 4th European Conference*, pages 214–229. Springer, 2001.

J. B. Tenenbaum, V. Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000.

Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *Int. J. Comput. Vision*, 70(1):77–90, 2006.

Appendix A. Geometric Preliminaries

Let $\mathcal{M} \in \mathcal{G}(d, V, \tau)$. In the remainder of this section, for $x \in \mathcal{M}$ denote the orthogonal projection from $\mathbb{R}^n$ to the affine subspace tangent to $\mathcal{M}$ at x, $Tan(x)$ by Π_x.

Lemma 11 *For each $p \in \mathcal{M}$, such that $Tan(p) = \{(z_1, z_2) \in \mathbb{R}^d \oplus \mathbb{R}^{n-d} | Az_1 + b = z_2\}$ for some matrix $A(p)$ and vector $b(p)$, there exists a neighborhood $V \subseteq \mathbb{R}^n$ of p, an open set $U \in \mathbb{R}^d$ and a $\mathcal{C}^2$ function $F : U \to \mathbb{R}^{n-d}$ with $DF(u)$ of rank d for all $u \in U$ such that*

$$\mathcal{M} \cap V = ((u, F(u)), u \in U \cap \mathbb{R}^d). \tag{68}$$

Proof Let U, W, ϕ, ψ be as in the definition of a C^2 manifold. Since $\phi \circ \psi = id$, $D\psi$ has rank d at each point of W, in particular $q = \phi(p)$. Therefore,

$$\det\left(\frac{\partial}{\partial w_1}\Pi_{\mathbb{R}^d} \circ \psi, \dots, \frac{\partial}{\partial w_d}\Pi_{\mathbb{R}^d} \circ \psi\right) \neq 0 \tag{69}$$

holds at q. By (69), we can apply the inverse function theorem to $\Pi_{\mathbb{R}^d} \circ \psi$ at $u_0 = \Pi_{\mathbb{R}^d} \circ \psi(q)$ to see that $f = (\Pi_d \circ \psi)^{-1} : U' \to \mathbb{R}^d$ exists, is $\mathcal{C}^2$ and Df has rank d at each $u \in U'$ for some neighborhood U' (possibly smaller than U) of $\Pi_{\mathbb{R}^d} q$. Setting $F = \Pi_{\mathbb{R}^{n-d}} \circ \psi \circ f$, we obtain (68). ■

Lemma 12 *Suppose that $\mathcal{M} \in \mathcal{G}(d, V, \tau)$. Let*

$$U := \{y \big| |y - \Pi_x y| \leq \tau/4\} \cap \{y \big| |x - \Pi_x y| \leq \tau/4\}.$$

Then,

$$\Pi_x(U \cap \mathcal{M}) = \Pi_x(U).$$

Proof Without loss of generality, we will assume $\tau/2 = 1$, and $x = \{0\}$, and $Tan(x) = \mathbb{R}^d$. Let $\mathcal{N} = U \cap \mathcal{M}$. We will first show that $\Pi_0(\mathcal{N}) = B_d$, where B_d is the unit ball in $\mathbb{R}^d$. Suppose otherwise, then let $\emptyset \neq Y := B_d \setminus \Pi_0(\mathcal{N})$. Note that $\mathcal{N}$ is closed and bounded and is therefore compact. The image of a compact set under a continuous map is compact, therefore $\Pi_0(\mathcal{N})$ is compact. Therefore $\mathbb{R}^d \setminus \Pi_0(\mathcal{N})$ is open. Let x_1 be a point of minimal distance from $0 = \Pi_0(0) \subseteq \Pi_0(\mathcal{N})$ among all points in the closure Z of $\mathbb{R}^d \setminus \Pi_0(\mathcal{N})$. Since $\emptyset \neq Y$ and $\mathbb{R}^d \setminus \Pi_0(\mathcal{N})$ is open, $|x_1| < 1$. Since $Tan(0) = \mathbb{R}^d$ and $\mathcal{M}$ is a closed imbedded C^2-submanifold, 0 does not belong to Z. Therefore $x_1 \neq 0$. By Federer's criterion for the reach, $\forall y_1 \in \Pi_0^{-1}(x_1) \cap \mathcal{N}$,

$$dist(y_1, Tan(0)) \leq \frac{\|y_1\|^2}{4}. \tag{70}$$

Therefore, $\forall y_1 \in \Pi_0^{-1}(x_1) \cap \mathcal{N}$,

$$dist(y_1, \mathbb{R}^d) \leq \frac{dist(y_1, \mathbb{R}^d)^2 + |x_1|^2}{4}. \tag{71}$$

Noting that $2 \geq 1 \geq dist(y_1, \mathbb{R}^d)$ and solving the above quadratic inequality, we see that

$$|y_1 - x_1|/2 \leq 1 - \sqrt{1 - \left(\frac{|x_1|}{2}\right)^2} \tag{72}$$

$$\leq \left(\frac{|x_1|}{2}\right)^2. \tag{73}$$

This implies that

$$|y_1 - x_1| \leq \frac{1}{8} < \frac{\tau}{8}. \tag{74}$$

Again by Federer's criterion, for any $z \in \Pi_0^{-1}(|x_1|B_d) \cap \mathcal{N}$,

$$
\begin{aligned}
|z - \Pi_{y_1}(z)|/2 &\leq 1 - \sqrt{1 - \left(\frac{|y_1 - \Pi_{y_1} z|}{2}\right)^2} && (75)\\
&\leq \left(\frac{|y_1 - \Pi_{y_1} z|}{2}\right)^2. && (76)
\end{aligned}
$$

By Lemma 11 and (74), we have the following.

Claim 2 *Let $y_1 \in \Pi_0^{-1}(x_1) \cap \mathcal{N}$. Then there exists $v \in \partial B_d$ such that if $y_1' \in Tan(y_1)$ then $\langle y_1' - y_1, v\rangle = 0$.*

Let $\ell = \{\lambda v | \lambda \in \mathbb{R}\}$ and let Π_ℓ denote the orthogonal projection on to ℓ. Then,

$$\Pi_\ell(|x_1|B_d) = \{\lambda v | \lambda \in [-|x_1|, |x_1|]\}.$$

By Claim 2, $\Pi_\ell(Tan(y_1))$ is the single point $\Pi_\ell(y_1)$. Let $\Pi_\ell(y_1) = \lambda_0 v$. Let $x_2 = |x_1|v$ if $\lambda_0 \leq 0$ and $x_2 = -|x_1|v$ if $\lambda_0 > 0$. Let $y_2 \in \Pi_0^{-1}(x_2) \cap \mathcal{N}$. Then,

$$
\begin{aligned}
|x_1| &\leq |\Pi_\ell(y_1) - x_2| && (77)\\
&\leq dist(y_2, Tan(y_1)) && (78)\\
&\leq \frac{|y_2 - y_1|^2}{4} && (79)\\
&\leq \frac{2|y_2 - x_2|^2 + |x_1 - x_2|^2 + 2|y_1 - x_1|^2}{4} && (80)\\
&\leq \frac{2\left(\frac{|x_2|^2}{2}\right)^2 + 4|x_1|^2 + 2\left(\frac{|x_1|^2}{2}\right)^2}{4} && (81)
\end{aligned}
$$

Therefore,

$$\alpha := |x_1| \leq |x_1|^4/4 + |x_1|^2.$$

Therefore, $1 \leq \alpha^3/4 + \alpha$. This implies that

$$|x_1| > \frac{1}{2}. \quad (82)$$

■

Lemma 13

Suppose that $\mathcal{M} \in \mathcal{G}(d, V, \tau)$. Let

$$\hat{U} := \{y \big| |y - \Pi_x y| \leq \tau/8\} \cap \{y \big| |x - \Pi_x y| \leq \tau/8\}.$$

There exists a C^2 function $F_{x,\hat{U}}$ from $\Pi_x(\hat{U})$ to $\Pi_x^{-1}(\Pi_x(0))$ such that

$$\{y + F_{x,\hat{U}}(y) \big| y \in \Pi_x(\hat{U})\} = \mathcal{M} \cap \hat{U}.$$

Secondly, let $z \in \mathcal{M} \cap \hat{U}$ satisfy $|\Pi_x(z) - x| = \delta$. Let z be taken to be the origin and let the span of the first d canonical basis vectors be denoted $\mathbb{R}^d$ and let $\mathbb{R}^d$ be a translate of $Tan(x)$. Let the span of the last $n-d$ canonical basis vectors be denoted $\mathbb{R}^{n-d}$. In this coordinate frame, let a point $z' \in \mathbb{R}^n$ be represented as (z'_1, z'_2), where $z'_1 \in \mathbb{R}^d$ and $z'_2 \in \mathbb{R}^{n-d}$. By the preceding lemma, there exists an $(n-d) \times d$ matrix A_z such that

$$Tan(z) = \{(z'_1, z'_2) | A_z z'_1 - I z'_2 = 0\} \tag{83}$$

where the identity matrix is $(n-d) \times (n-d)$. For $\delta < \tau/8$, let $z \in \mathcal{M} \cap \{z \big| |z - \Pi_x z| \leq \delta\} \cap \{z \big| |x - \Pi_x z| \leq \delta\}$. Then $\|A_z\|_2 \leq 20\delta/\tau$.

Proof Let

$$U := \{y \big| |y - \Pi_x y| \leq \tau/4\} \cap \{y \big| |x - \Pi_x y| \leq \tau/4\}.$$

We will first show that there exists a function $F_{x,\hat{U}}$ that satisfies the given conditions and then show that it is C^2. Let $z \in \mathcal{M} \cap \hat{U}$ satisfy $|\Pi_x(z) - x| = \delta$. Let z be taken to be the origin and let the span of the first d canonical basis vectors be denoted $\mathbb{R}^d$ and let $\mathbb{R}^d$ be a translate of $Tan(x)$. Let the span of the last $n-d$ canonical basis vectors be denoted $\mathbb{R}^{n-d}$. In this coordinate frame, let a point $z' \in \mathbb{R}^n$ be represented as (z'_1, z'_2), where $z'_1 \in \mathbb{R}^d$ and $z'_2 \in \mathbb{R}^{n-d}$. By the preceding lemma, there exists a matrix A such that

$$Tan(z) = \{(z'_1, z'_2) | A z'_1 - I z'_2 = 0\}. \tag{84}$$

Further, a linear algebraic calculation shows that

$$dist(z', Tan(z)) = \left\| (I + AA^T)^{-1/2} (A z'_1 - I z'_2) \right\|_2. \tag{85}$$

Let S_δ^{d-1} denote the $d-1-$dimensional sphere of radius δ centered at the origin contained in $\mathbb{R}^d$. By the preceding lemma, there is a point $\tilde{z} \in \mathcal{M}$ for every $z' \in S_\delta^{d-1}$ such that $\tilde{z} \in U$, $\Pi_x \tilde{z} = \Pi_x z'$ and

$$\|\tilde{z} - \Pi_x \tilde{z}\| \leq \frac{\|x - \Pi_x \tilde{z}\|^2}{\tau} = \frac{4\delta^2}{\tau}. \tag{86}$$

Therefore,

$$\left\|\tilde{z} - z'\right\| = \|\tilde{z} - ((\Pi_x \tilde{z}) - x_2)\| \leq \frac{4\delta^2}{\tau} + \frac{\delta^2}{\tau} = \frac{5\delta^2}{\tau}. \tag{87}$$

Therefore,

$$\begin{aligned} dist(z', Tan(z)) &\leq dist(\tilde{z}, Tan(z)) + \left\|\tilde{z} - z'\right\| && (88)\\ &\leq \frac{|z - \tilde{z}|^2}{\tau} + \frac{5\delta^2}{\tau} && (89)\\ &= \frac{|z - z'|^2 + |z' - \tilde{z}|^2}{\tau} + \frac{5\delta^2}{\tau} && (90)\\ &\leq \frac{\delta^2 + (8\delta^2/\tau)^2}{\tau} + \frac{5\delta^2}{\tau}. && (91) \end{aligned}$$

Therefore, for any $z_1' \in S_\delta^{d-1}$,

$$\left\|(I+AA^T)^{-1/2}(Az_1')\right\|_2 \leq \frac{\delta^2}{\tau}\left(6+\frac{64\delta^2}{\tau^2}\right). \tag{92}$$

Thus,

$$\left\|(I+AA^T)^{-1/2}(A)\right\|_2 \leq \frac{\delta}{\tau}\left(6+\frac{64\delta^2}{\tau^2}\right) =: \delta'. \tag{93}$$

Therefore, we see that

$$\left\|(I+AA^T)^{-1/2}(AA^T)(I+AA^T)^{-1/2}\right\|_2 \leq \delta'^2. \tag{94}$$

Let $\|A\|_2 = \lambda$. We then see that λ^2 is an eigenvalue of AA^T. Therefore,

$$\frac{\lambda^2}{1+\lambda^2} \leq \delta'^2. \tag{95}$$

This gives us $\lambda^2 \leq \frac{\delta'^2}{1-\delta'^2}$, which implies that

$$\lambda \leq \frac{\delta'}{\sqrt{1-\delta'^2}}. \tag{96}$$

We will use this to show that , $\Pi_x^{-1}(\Pi_x z) \cap \mathcal{M} \cap U$ contains the single point z. Suppose to the contrary, there is a point $\hat{z} \neq z$ that also belongs to $\Pi_x^{-1}(\Pi_x z) \cap \mathcal{M} \cap U$. Then,

$$dist(\hat{z}, Tan(z)) \leq \frac{\|\hat{z}_2\|^2}{\tau}, \tag{97}$$

where $\|\hat{z}_2\| \leq 2\delta^2/\tau$. Thus,

$$\|\hat{z}_2\| \leq \frac{\|\hat{z}_2\|^2\|I+AA^T\|^{1/2}}{\tau}.$$

Therefore,

$$1 \leq \|\hat{z}_2\|(1/\sqrt{(1-\delta'^2)})/\tau \leq \frac{2\delta^2}{\tau^2\sqrt{1-\delta'^2}}. \tag{98}$$

Therefore

$$1-\delta'^2 \leq \frac{2\delta^2}{\tau^2}, \tag{99}$$

and so

$$\delta' \geq 1 - 2\delta^2/\tau^2. \tag{100}$$

assuming $\delta \leq \frac{\tau}{8}$,

$$\delta' \leq \frac{7\delta}{\tau}. \tag{101}$$

Therefore

$$2\delta^2/\tau^2 + 7\delta/\tau \geq 1.$$

This implies that $\delta/\tau > 1/8$. This is a contradiction.

By (84) and Lemma 11, $F_{x,U}$ is C^2. ∎

Appendix B. Quantitative inverse and implicit function theorems

We now prove versions of the implicit and inverse function theorems with quantitative bounds on the second derivatives that do not depend on the dimensions involved.

We begin with the inverse function theorem.

Let $g : \mathbb{R}^p \to \mathbb{R}^p$ be a $\mathcal{C}^2$ function on whose derivatives the following bounds hold.

At any point $x \in B_p(0, 1)$,

$$\|Jac_g - I\| \le \epsilon^2/4 \tag{102}$$

for some $\epsilon \in [0, 1]$

For any non-zero vector v and x as before,

$$\left\| \frac{\partial^2 g(x)}{\partial v^2} \right\| \le \left(\frac{\epsilon^2}{4} \right) |v|^2 \tag{103}$$

for the same ϵ.

By (102), for any $x \neq x'$, both belonging to $B_p(0, 1)$,

$$|g(x) - g(x') - (x - x')| \le |x - x'|(1/4),$$

which implies that $g(x) \neq g(x')$. Applying the Inverse Function Theorem (Narasimhan (1965)), there exists a function $f : g(B_p(0, 1)) \to B_p(0, 1)$ such that $f(g(x)) = x$, for all $x \in B(0, 1)$. Let $\hat{F} = w \cdot f$ for some fixed non-zero vector w. Let $g = (g_1, \dots, g_p)$, where each g_i is a real-valued function. The Jacobian of the identity function is I. Therefore, by the chain rule,

$$\left(\left(\frac{df_i}{dg_j} \right)_{i,j \in [p]} \right) Jac_g = I, \tag{104}$$

implying by (102) that

$$\left\| \left(\left(\frac{df_i}{dg_j} \right)_{i,j \in [p]} \right) \right\| \le (1 - \epsilon^2/4)^{-1} \tag{105}$$

The second derivative of a linear function is 0 and so

$$0 = \frac{\partial^2 \hat{F}(g)}{\partial v^2}(x) = \sum_{i,j} \frac{d^2 \hat{F}}{dg_i dg_j} \left(\frac{dg_i}{dv} \right) \left(\frac{dg_j}{dv} \right) + \sum_j \frac{d\hat{F}}{dg_j} \left(\frac{d^2 g_j}{dv^2} \right). \tag{106}$$

Therefore,

$$\sum_{i,j} \frac{d^2 \hat{F}}{dg_i dg_j} \left(\frac{dg_i}{dv} \right) \left(\frac{dg_j}{dv} \right) = (-1) \sum_j \frac{d\hat{F}}{dg_j} \left(\frac{d^2 g_j}{dv^2} \right). \tag{107}$$

and so by Cauchy-Schwartz,

$$\left|\sum_{i,j} \frac{d^2\hat{F}}{dg_i dg_j}\left(\frac{dg_i}{dv}\right)\left(\frac{dg_j}{dv}\right)\right| \leq \left\|\left(\left(\frac{d\hat{F}}{dg_j}\right)_{j\in[p]}\right)\right\| \left\|\left(\left(\frac{d^2 g_j}{dv^2}\right)_{j\in[p]}\right)\right\|. \tag{108}$$

By (102) there exists a unit vector $\tilde{v}$ such that

$$\left|\sum_{i,j} \frac{d^2\hat{F}}{dg_i dg_j}\left(\frac{dg_i}{d\tilde{v}}\right)\left(\frac{dg_j}{d\tilde{v}}\right)\right| = \left\|Hess\,\hat{F}\right\| \left\|\frac{dg}{d\tilde{v}}\right\|^2 \geq \left\|Hess\,\hat{F}\right\| \inf_{\|v\|=1} \left\|\frac{dg}{dv}\right\|^2. \tag{109}$$

Together (103), (105), (108) and (109) imply that

$$\left\|Hess\,\hat{F}\right\| \inf_{\|v\|=1} \left\|\frac{dg}{dv}\right\|^2 \leq \left\|\left(\left(\frac{df_i}{dg_j}\right)_{i,j\in[p]}\right) w\right\| \sup_{\|v\|=1} \left(\frac{\epsilon^2}{4}\right)\|v\|^2 \leq \left(\frac{\epsilon^2}{4-\epsilon}\right)\|w\|. \tag{110}$$

It follows that

$$\left\|Hess\,\hat{F}\right\| \leq \left(\frac{\epsilon^2}{4-\epsilon}\right)\|w\| \sup_{\|v\|=1} \left\|\frac{dg}{dv}\right\|^{-2} \leq \left(\frac{16\epsilon^2}{(4-\epsilon)^3}\right)\|w\|. \tag{111}$$

Next, consider the setting of the Implicit Function Theorem. Let $h : \mathbb{R}^{m+n} \to \mathbb{R}^n$ be a $\mathcal{C}^2-$function,

$$h : (x, y) \mapsto h(x, y).$$

Let $g : B_{m+n} \to \mathbb{R}^{m+n}$ be defined by

$$g : (x, y) \mapsto (x, h(x, y)).$$

Suppose the Jacobian of g, Jac_g satisfies

$$\|Jac_g - I\| < \epsilon^2/4$$

on B_{m+n} and that for any vector $v \in \mathbb{R}^{m+n}$,

$$\left\|\frac{\partial^2 g(x)}{\partial v^2}\right\| \leq \left(\frac{\epsilon^2}{4}\right)\|v\|^2$$

where $\epsilon \in [0, 1]$. Suppose also that $\|g(0)\| < \frac{\epsilon^2}{20}$ for the same ϵ.

Let $p = m + n$. Then, applying the inverse function theorem, we see that defining f and $\hat{F}$ as before, and choosing $\|w\| = 1$,

$$\left\|Hess\,\hat{F}\right\| \leq \frac{16\epsilon^2}{(4-\epsilon)^3}. \tag{112}$$

Lemma 14 *On the domain of definition of f, i. e. $g(B_{m+n})$*

$$f((x, y)) = (x, e(x, y))$$

for an appropriate e and in particular, for $\|x\| \le \frac{\eta}{2}$, where $\eta \in [0, 1]$,

$$f((x, 0)) = (x, e(x, 0))$$

and

$$\|(x, e(x, 0))\| \le \frac{8}{5}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right).$$

Finally, for any $w \in \mathbb{R}^n$ such that $\|w\| = 1$,

$$\|Hess(e \cdot w)\| \le \frac{16\epsilon^2}{(4-\epsilon)^3}. \tag{113}$$

Proof It suffices to prove that if $z = (x, y) \in \mathbb{R}^p$ and $\|z\| \le \eta/2$, where $\eta \in [0, 1]$, then there exists a point $\hat{z}$, where $\|\hat{z}\| \le \frac{8}{5}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right)$, such that $g(\hat{z}) = z$. We will achieve this by analysing Newton's method for finding a sequence $\hat{z}_0, \dots, \hat{z}_k, \dots$ converging to a point $\hat{z}$ that satisfies $g(\hat{z}) = z$. We will start with $\hat{z}_0 = 0$.

The iterations of Newton's method proceed as follows.

For $i \ge 0$,

$$\hat{z}_{i+1} = \hat{z}_i - J_g^{-1}(\hat{z}_i)(g(\hat{z}_i) - z). \tag{114}$$

Claim 3

We shall first show that for any $i \ge 0$, $\|\hat{z}_i\| \le \frac{8}{5}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right)$.

Proof Observe that

$$\|\hat{z}_{i+1} - \hat{z}_i\| = \|J_g^{-1}(\hat{z}_i)(g(\hat{z}_i) - z)\|. \tag{115}$$

For $i = 0$,

$$\|g(\hat{z}_i) - z\| \le \frac{\epsilon^2}{20} + \frac{\eta}{2}. \tag{116}$$

and since $\|J_g^{-1}(\hat{z}_i)\| \le \frac{1}{1-\epsilon/4} \le 4/3$, therefore

$$\|\hat{z}_{i+1} - \hat{z}_i\| \le \left(\frac{4}{3}\right)\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right). \tag{117}$$

Suppose $i \ge 1$.

$$g(\hat{z}_i) - z = g\left(\hat{z}_{i-1} - J_g^{-1}(\hat{z}_{i-1})(g(\hat{z}_{i-1}) - z)\right) - z. \tag{118}$$

Using the integral form of the remainder in Taylor's theorem, the right hand side of (118) equals

$$g(\hat{z}_{i-1}) + J_g(\hat{z}_{i-1})\left(-J_g^{-1}(\hat{z}_{i-1})(g(\hat{z}_{i-1}) - z)\right) + \Lambda - z,$$

which simplifies to Λ, where

$$\Lambda = \int_0^1 (1-t)(\hat{z}_i - \hat{z}_{i-1})^T Hess_g(\hat{z}_{i-1} + t(\hat{z}_i - \hat{z}_{i-1}))(\hat{z}_i - \hat{z}_{i-1})dt.$$

The norm of Λ is bounded above as follows. Note that by the induction hypothesis, $\|\hat{z}_i\| \leq \frac{8}{5}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right)$, and $\|\hat{z}_{i-1}\| \leq \frac{8}{5}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right)$, which places both $\hat{z}_i$ and $\hat{z}_{i-1}$ within the unit ball. Therefore $\|(\hat{z}_i - \hat{z}_{i-1})^T Hess_g(\hat{z}_{i-1} + t(\hat{z}_i - \hat{z}_{i-1}))(\hat{z}_i - \hat{z}_{i-1})\| \leq (\epsilon^2/4)\|\hat{z}_i - \hat{z}_{i-1}\|^2$ for any $t \in [0, 1]$.

$$\|\Lambda\| \leq \int_0^1 (1-t)\|(\hat{z}_i - \hat{z}_{i-1})\|^2(\epsilon^2/4)dt = \left(\frac{\epsilon^2}{8}\right)\|(\hat{z}_i - \hat{z}_{i-1})\|^2.$$

Therefore

$$\|\hat{z}_{i+1} - \hat{z}_i\| = \|J_g^{-1}(\hat{z}_i)(g(\hat{z}_i) - z)\| \leq \left(\frac{4}{3}\right)\left(\frac{\epsilon^2}{8}\right)\|(\hat{z}_i - \hat{z}_{i-1})\|^2 = \left(\frac{\epsilon^2}{6}\right)\|(\hat{z}_i - \hat{z}_{i-1})\|^2 \tag{119}$$

By recursion,

$$\|\hat{z}_{i+1} - \hat{z}_i\| \leq \left(\frac{\epsilon^{2i}}{6^i}\right)\|\hat{z}_1 - \hat{z}_0\|^{2^i}. \tag{120}$$

Therefore,

$$\|\hat{z}_{i+1}\| = \|\hat{z}_{i+1} - \hat{z}_0\| \leq \sum_{j=1}^{i}\|\hat{z}_{j+1} - \hat{z}_j\| \leq \frac{\|\hat{z}_1 - \hat{z}_0\|}{1 - \frac{\epsilon^2}{6}} \leq \left(\frac{4}{3}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right)\right)\left(\frac{6}{5}\right) = \frac{8}{5}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right) \tag{121}$$

■

Recall that $g : B_{m+n} \to \mathbb{R}^{m+n}$ is given by

$$g : (x, y) \mapsto (x, h(x, y)).$$

Since g is injective, it follows that on the domain of definition of f, i. e. $g(B_{m+n})$

$$f((x, y)) = (x, e(x, y))$$

for an appropriate e. By (117) and (120) $(\hat{z}_0, \dots, \hat{z}_i, \dots)$ is a Cauchy sequence, and therefore has a unique limit point. By the preceding Claim, this limit $\hat{z}$ satisfies $\|\hat{z}\| \leq \frac{22}{25} < 1$. Therefore any point in $B_m \times B_n$ of the form $(x, 0)$ where $\|x\| = \frac{\eta}{2} \leq \frac{1}{2}$ belongs to $g(B_{m+n})$. Further,

$$\|f((x, 0))\| \leq \frac{8}{5}\left(\frac{\epsilon^2}{20} + \frac{\eta}{2}\right).$$

In particular, setting $\eta = 0$, we have

$$\|f((0, 0))\| \leq \frac{2\epsilon^2}{25}. \tag{122}$$

By (105) the function e satisfies, for $\|x\| \le 1/2$,

$$\begin{aligned} \|D_x e\|^2 &= \|D_x f\|^2 - 1 && (123)\\ &\le (1-\epsilon^2/4)^{-2} - 1 && (124)\\ &\le \epsilon^2. && (125) \end{aligned}$$

By (112) the function e satisfies, for any $w \in \mathbb{R}^n$ such that $\|w\| = 1$,

$$\|Hess(e \cdot w)\| \le \frac{16\epsilon^2}{(4-\epsilon)^3}. \tag{126}$$

■

Appendix C. Finding good discs

Let G_σ^z be the spherical gaussian measure whose variance (of the marginal along any fixed direction) is σ^2 and center z. When z is the origin, we will drop the superscript. Let $\hat{\mu} = \mu * G_\sigma$ be the distribution from which the y_i are drawn. We will need the following theorem which follows from Theorem 3.2.3 in Federer (1969).

Theorem 15 *Let $\mathcal{L}^m$ denote the $m-$dimensional Lebesgue measure and $\mathcal{H}^m$ denote the $m-$dimensional Hausdorff measure. Suppose $f : A \to \mathbb{R}^n$ be a $1 \to 1$ C^2 function with $m \le n$ where A is a $\mathcal{L}^m-$measurable subset of $\mathbb{R}^m$ and $J_m f$ is the Jacobian of f. If u is a $\mathcal{L}^m$ integrable function, then*

$$\int_{\mathbb{R}^m} u(x) J_m f(x) \mathcal{L}^m(dx) = \int_{\mathbb{R}^n} u(f^{-1}(y)) \mathcal{H}^m(dy). \tag{127}$$

C.1. Learning unit discs that approximate the data locally

Let X be a finite set of points in $E = \mathbb{R}^D$ and $X \cap B_1(x) := \{x, \tilde{x}_1, \dots, \tilde{x}_s\}$ be a set of points within a Hausdorff distance δ of some (unknown) unit d-dimensional disc $D_1(x)$ centered at x. Here $B_1(x)$ is the set of points in $\mathbb{R}^D$ whose distance from x is less or equal to 1. We give below a simple algorithm that finds a unit d-disc centered at x within a Hausdorff distance $Cd\delta$ of $X_0 := X \cap B_1(x)$, where C is an absolute constant.

The basic idea is to choose a near orthonormal basis from X_0 where x is taken to be the origin and let the span of this basis intersected with $B_1(x)$ be the desired disc. This algorithm appeared previously in Fefferman et al. (2015), but has been included in the interest of readability.

Algorithm FindDisc:

1. Let x_1 be a point that minimizes $|1 - |x - x'||$ over all $x' \in X_0$.
2. Given $x_1, \dots x_m$ for $m \le d-1$, choose x_{m+1} such that

$$\max(|1 - |x - x'||, |\langle x_1/|x_1|, x'\rangle|, \dots, |\langle x_m/|x_m|, x'\rangle|)$$

is minimized among all $x' \in X_0$ for $x' = x_{m+1}$.

Let $\tilde{A}_x$ be the affine d-dimensional subspace containing $x, x_1, \dots, x_n$, and the unit d-disc $\tilde{D}_1(x)$ be $\tilde{A}_x \cap B_1(x)$. Recall that for two subsets A, B of $\mathbb{R}^D$, $d_H(A, B)$ represents the Hausdorff distance between the sets. The same letter C can be used to denote different constants, even within one formula.

Lemma 16 *Suppose there exists a d-dimensional affine subspace A_x containing x such that $D_1(x) = A_x \cap B_1(x)$ satisfies $d_H(X_0, D_1(x)) \leq \delta$. Suppose $0 < \delta < \frac{1}{2d}$. Then $d_H(X_0, \tilde{D}_1(x)) \leq Cd\delta$, where C is an absolute constant.*

Proof Without loss of generality, let x be the origin. Let $d(x, y)$ be used to denote $|x - y|$. We will first show that for all $m \leq d - 1$,

$$\max\left(|1 - d(x, x_{m+1})|, \left|\left\langle \frac{x_1}{|x_1|}, x_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{x_m}{|x_m|}, x_{m+1}\right\rangle\right|\right) < \delta.$$

To this end, we observe that the minimum over $D_1(x)$ of

$$\max\left(|1 - d(x, y)|, \left|\left\langle \frac{(x_1)}{|x_1|}, y\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, y\right\rangle\right|\right) \tag{128}$$

is 0, because the dimension of $D_1(x)$ is d and there are only $m \leq d - 1$ linear equality constraints. Also, the radius of $D_1(x)$ is 1, so $|1 - d(x, z_{m+1})|$ has a value of 0 where a minimum of (128) occurs at $y = z_{m+1}$. Since the Hausdorff distance between $D_1(x)$ and X_0 is less than δ there exists a point $y_{m+1} \in X_0$ whose distance from z_{m+1} is less than δ. For this point y_{m+1}, we have δ greater than

$$\max\left(|1 - d(x, y_{m+1})|, \left|\left\langle \frac{(x_1)}{|x_1|}, y_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, y_{m+1}\right\rangle\right|\right). \tag{129}$$

Since

$$\max\left(|1 - d(x, x_{m+1})|, \left|\left\langle \frac{(x_1)}{|x_1|}, x_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, x_{m+1}\right\rangle\right|\right)$$

is no more than the corresponding quantity in (129), we see that for each $m + 1 \leq n$,

$$\max\left(|1 - d(x, x_{m+1})|, \left|\left\langle \frac{(x_1)}{|x_1|}, x_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, x_{m+1}\right\rangle\right|\right) < \delta.$$

Let $\tilde{V}$ be an $D \times d$ matrix whose i^{th} column is the column x_i. Let the operator 2-norm of a matrix Z be denoted $\|Z\|$. For any distinct i, j we have $|\langle x_i, x_j\rangle| < \delta$, and for any i, $|\langle x_i, x_i\rangle - 1| < 2\delta$, because $0 < 1 - \delta < |x_i| < 1$. Therefore,

$$\|\tilde{V}^t\tilde{V} - I\| \leq C_1 d\delta.$$

Therefore, the singular values of $\tilde{V}$ lie in the interval

$$I_C = (\exp(-Cd\delta), \exp(Cd\delta)) \supseteq (1 - C_1 d\delta, 1 + C_1 d\delta).$$

For each $i \leq n$, let x_i' be the nearest point on $D_1(x)$ to the point x_i. Since the Hausdorff distance of X_0 to $D_1(x)$ is less than δ, this implies that $|x_i' - x_i| < \delta$ for all $i \leq n$. Let $\hat{V}$ be an $D \times d$

matrix whose i^{th} column is x'_i. Since for any distinct i, j $|\langle x'_i, x'_j\rangle| < 3\delta + \delta^2$, and for any i, $|\langle x'_i, x'_i\rangle - 1| < 4\delta$,

$$\|\hat{V}^t\hat{V} - I\| \leq Cd\delta.$$

This means that the singular values of $\hat{V}$ lie in the interval I_C.

We shall now proceed to obtain an upper bound of $Cd\delta$ on the Hausdorff distance between X_0 and $\tilde{D}_1(x)$. Recall that the unit d-disc $\tilde{D}_1(x)$ is $\tilde{A}_x \cap B_1(x)$. By the triangle inequality, since the Hausdorff distance of X_0 to $D_1(x)$ is less than δ, it suffices to show that the Hausdorff distance between $D_1(x)$ and $\tilde{D}_1(x)$ is less than $Cd\delta$.

Let x' denote a point on $D_1(x)$. We will show that there exists a point $z' \in \tilde{D}_1(x)$ such that $|x' - z'| < Cd\delta$.

Let $\hat{V}\alpha = x'$. By the bound on the singular values of $\hat{V}$, we have $|\alpha| < \exp(Cd\delta)$. Let $y' = \tilde{V}\alpha$. Then, by the bound on the singular values of $\tilde{V}$, we have $|y'| \leq \exp(Cd\delta)$. Let $z' = z' = \min(1-\delta, |y'|)|y'|^{-1}y'$. By the preceding two lines, z' belongs to $\tilde{D}_1(x)$. We next obtain an upper bound on $|x' - z'|$

$$\begin{aligned} |x' - z'| &\leq |x' - y'| && (130)\\ &\quad +|y' - z'|. && (131)\end{aligned}$$

We examine the term in (130)

$$|x' - y'| = |\hat{V}\alpha - \tilde{V}\alpha| \leq \sup_i |x_i - x'_i|(\sum_i |\alpha_i|) \leq \delta d \exp(Cn\delta).$$

We next bound the term in (131).

$$|y' - z'| \leq |y'|(1 - \exp(-Cd\delta)) \leq Cn\delta.$$

Together, these calculations show that

$$|x' - z'| < Cd\delta.$$

A similar argument shows that if z'' belongs to $\tilde{D}_1(x)$ then there is a point $p' \in D_1(x)$ such that $|p' - z''| < Cd\delta$; the details follow. Let $\hat{V}\beta = z''$. From the bound on the singular values of $\hat{V}$, $|\beta| < \exp(Cd\delta)$. Let $q' := \tilde{V}\beta$. Let $p' := \min(1-\delta, |q'|)|q'|^{-1}q'$.

$$\begin{aligned} |p' - z''| &\leq |q' - z''| + |p' - q'| \\ &\leq |\tilde{V}\beta - V\beta| + |1 - \tilde{V}\beta| \\ &\leq \sup_i |x_i - x'_i|(\sum_i |\beta_i|) + C\delta d \\ &\leq \delta d \exp(Cd\delta) + C\delta d \\ &\leq C\delta d.\end{aligned}$$

This proves that the Hausdorff distance between X_0 and $\tilde{D}_1(x)$ is bounded above by $Cd\delta$ where C is a universal constant.

■

Appendix D. Proof of Lemma 6

Proof

Claim 4 *There exists $\kappa \in \mathbb{R}$ independent of z such that*

$$c\kappa^{-1} < \frac{d(\lambda_{\mathcal{M}} * \beta)}{d\lambda}(z) < c^{-1}\kappa^{-1}.$$

Proof After appropriate scaling, we will assume that $r = 1$. We make the following claim.

Claim 5 *If $|v| < \frac{1}{\sqrt{2}}$,*

$$\exp(-2(d+2)|v|^2) < \beta(v). \tag{132}$$

Also

$$\forall v \in \mathbb{R}^D,\ \exp(-(d+2)|v|^2) > \beta(v). \tag{133}$$

Proof To see the first inequality, note that

$$\begin{aligned} |v| &< \frac{1}{\sqrt{2}} && (134)\\ \implies (-2)(1-|v|^2) &< -1 && (135)\\ \implies (-2)(d+2)|v|^2 &< (d+2)\left(-\frac{|v|^2}{1-|v|^2}\right) && (136)\\ \implies (-2)(d+2)|v|^2 &< (d+2)\left(\ln(1-|v|^2)\right) && (137)\\ \implies \exp((-2)(d+2)|v|^2) &< \left(1-|v|^2\right)^{d+2} = \beta(v). && (138) \end{aligned}$$

To see the second inequality, exponentiate the following inequality for $|v| < 1$:

$$-(d+2)|v|^2 \quad > \quad (d+2)\left(\ln(1-|v|^2)\right). \tag{139}$$

When $|v| \geq 1$, $\beta(v) = 0$, so the inequality holds.

∎

We will now use the preceding claim to get upper and lower bounds on

$$\int_{\mathbb{R}^d} \beta(v)\lambda(dv) =: c_\beta^{-1},$$

where λ corresponds to the $d-$dimensional Lebesgue measure.

$$\begin{aligned} \int_{\mathbb{R}^d} \beta(v)\lambda(dv) &= \int_{B_d} \beta(v)\lambda(dv) && (140)\\ &\leq \int_{B_d} \exp(-(d+2)v)\lambda(dv) && (141)\\ &\leq (\frac{\pi}{d+2})^{d/2}. && (142) \end{aligned}$$

Also,

$$\int_{B_d} \beta(v)\lambda(dv), \geq \int_{\frac{B_d}{\sqrt{2}}} \exp(-2(d+2)v)\lambda(dv) \tag{143}$$

$$\geq (\frac{\pi}{2(d+2)})^{d/2}(1-\exp(-\left(\frac{d+4}{2d+4}\right)^2\frac{d}{4})) \tag{144}$$

$$\geq (\frac{\pi}{2(d+2)})^{d/2}(1-\exp(-d/16)) \tag{145}$$

$$\geq c(\frac{\pi}{2(d+2)})^{d/2}. \tag{146}$$

Using numerical integration the value of c_β can be estimated to within a multiplicative factor of 2 using $(Cd)^d$ operations on real numbers.

Next consider a unit disk $B_d \subseteq \mathbb{R}^D$ equipped with the measure $c_\beta\lambda$. We consider a point q at a distance Δ from the projection of q onto B_d, which we assume is the origin. As a warm-up, we will be interested in

$$\frac{((c_\beta\lambda 1_{B_d}) * \beta)(q)}{((c_\beta\lambda 1_{B_d}) * \beta)(0)} = \frac{\int_{B_d}\beta(q-v)(c_\beta\lambda(dv))}{\int_{B_d}\beta(-v)(c_\beta\lambda(dv))}, \tag{147}$$

as a function of Δ. We observe that $v \in B_d \implies \beta(-v) \geq \beta(q-v)$, and so

$$\frac{((c_\beta\lambda 1_{B_d}) * \beta)(q)}{((c_\beta\lambda 1_{B_d}) * \beta)(0)} \leq 1. \tag{148}$$

Let $\Delta^2 \leq \frac{1}{8d^2}$. Suppose $|v|^2 < 1 - \frac{1}{2d}$, then

$$\Delta^2 \leq \left(\frac{1-|v|^2}{4d}\right). \tag{149}$$

Therefore,

$$\int_{B_d}\beta(q-v)(c_\beta\lambda(dv)) = \int_{B_d}(1-|v|^2-\Delta^2)^{d+2}1_{\{v||v|^2\leq 1-\Delta^2\}}(c_\beta\lambda(dv))$$

$$\geq \int_{\sqrt{1-\frac{1}{2d}}B_d}((1-|v|^2)(1-\frac{1}{4d}))^{d+2}(c_\beta\lambda(dv))$$

$$\geq \int_{\sqrt{1-\frac{1}{2d}}B_d} e^{-1}((1-|v|^2))^{d+2}(c_\beta\lambda(dv)) \tag{150}$$

$$\geq c\int_{B_d}((1-|v|^2))^{d+2}(c_\beta\lambda(dv)). \tag{151}$$

In the above sequence of inequalities the last step comes from dilating the disk $\sqrt{1-\frac{1}{2d}}B_d$ to B_d and observing that $\beta(v_1) \geq \beta(v_2)$ if $|v_1| < |v_2|$.

We thus have

$$c \leq \frac{((c_\beta\lambda 1_{B_d}) * \beta)(q)}{((c_\beta\lambda 1_{B_d}) * \beta)(0)} = \frac{\int_{B_d}\beta(q-v)(c_\beta\lambda(dv))}{\int_{B_d}\beta(-v)(c_\beta\lambda(dv))} \leq 1, \tag{152}$$

for some absolute constant $c > 0$ provided $\Delta^2 \leq \frac{1}{8d^2}$.

Next consider a point q at a distance $\leq 1/(2d)$ from $\mathcal{M}$. We let q be the origin. Consider a unit disk $B_d \subseteq \mathbb{R}^D$ that is parallel to the tangent plane to $\mathcal{M}$ at the point nearest to q. We will be interested in

$$\frac{((c_\beta \mathcal{H}^d_\mathcal{M} 1_{B_m}) * \beta)(q)}{((c_\beta \lambda 1_{B_d}) * \beta)(0)} = \frac{\int_{\mathcal{M}\cap B_n} \beta(-v)(c_\beta \mathcal{H}^d_\mathcal{M}(dv))}{\int_{B_d} \beta(-v)(c_\beta \lambda(dv))}, \tag{153}$$

as a function of Δ. Let Π_d denote the projection onto B_d. Let

$$\sup_{x \in \mathcal{M}\cap B_n} |x - \Pi_d x| = \Delta. \tag{154}$$

Then, by Federer's criterion for the reach, $\Delta < 1/d$. Also, $\mathcal{M} \cap B_n$ is the graph of a function $f(x)$ from $\Pi_d(\mathcal{M} \cap B_n)$ to the $D-d$ dimensional normal space to B_d. For $v \in \mathcal{M} \cap B_n$, let $w = \Pi_d v$, and by the definition of f, $v = w + f(w)$.

$$\begin{aligned}
\int_{\mathcal{M}\cap B_n} \beta(-v)(c_\beta \mathcal{H}^d_\mathcal{M}(dv)) &= \int_{\Pi_d(\mathcal{M}\cap B_n)} \beta(-(w + f(w)))(c_\beta \mathcal{H}^d_\mathcal{M}(dv)) \\
&\leq \int_{\Pi_d(\mathcal{M}\cap B_n)} \beta(-w)(c_\beta \mathcal{H}^d_\mathcal{M}(dv)) \qquad (155)\\
&\leq \int_{\Pi_d(\mathcal{M}\cap B_n)} \beta(-w)(c_\beta J(w)\lambda(dw)). \qquad (156)
\end{aligned}$$

Since $\|Df\|$ is of the order of $\frac{1}{Cd^C}$ by lemma 13 and the upper bound on r, the Jacobian

$$J(w) = \sqrt{\det(I + (Df(w))(Df(w))^T)}$$

is less or equal to an absolute constant C. This implies that

$$\int_{\Pi_d(\mathcal{M}\cap B_n)} \beta(-w)(c_\beta J(w)\lambda(dw)) \leq C \int_{B_d} \beta(-v)(c_\beta \lambda(dv)). \tag{157}$$

This in turn implies that

$$c^{-1} > \frac{\int_{\mathcal{M}\cap B_n} \beta(-v)(c_\beta \mathcal{H}^d_\mathcal{M}(dv))}{\int_{B_d} \beta(-v)(c_\beta \lambda(dv))}. \tag{158}$$

for an appropriately small universal constant c.

We now proceed to the lower bound. As noted above, $\Delta < 1/d$.

$$\int_{\mathcal{M}\cap B_n} \beta(-v)(c_\beta \mathcal{H}^d_{\mathcal{M}}(dv)) = \int_{\Pi_d(\mathcal{M}\cap B_n)} \beta(-(w+f(w)))(c_\beta \mathcal{H}^d_{\mathcal{M}}(dv)) \tag{159}$$

$$\geq \int_{B_d(1-1/d)} \beta(-(w+f(w)))(c_\beta \mathcal{H}^d_{\mathcal{M}}(dv)) \tag{160}$$

$$\geq \int_{B_d(1-1/d)} (1-|w|^2-\Delta^2)^{d+2}(c_\beta J(w)\lambda(dw))$$

$$\geq \int_{B_d(1-1/d)} ((1-|w|^2)(1-1/d))^{d+2}(c_\beta \lambda(dw))$$

$$\geq c\int_{B_d(1-1/d)} ((1-|w|^2))^{d+2}(c_\beta \lambda(dw)) \tag{161}$$

$$\geq c^2\int_{B_d} ((1-|w|^2))^{d+2}(c_\beta \lambda(dw)). \tag{162}$$

The last step comes from dilating the disk $(1-\frac{1}{d})B_d$ to B_d and observing that $\beta(v_1) \geq \beta(v_2)$ if $|v_1| < |v_2|$. In dropping $J(w)$, we used the fact that $J(w) \geq 1$.

Relabelling c^2 by c, the above sequence of inequalities shows that

$$\frac{\int_{\mathcal{M}\cap B_n} \beta(-v)(c_\beta \mathcal{H}^d_{\mathcal{M}}(dv))}{\int_{B_d} \beta(-v)(c_\beta \lambda(dv))} > c. \tag{163}$$

■

We next, using the fact that the Hausdorff distance of the set $\{p_i\}$ to $\mathcal{M}$ is less than $\frac{cr}{d}$ show the following.

Lemma 17 *There exists a measure μ_P supported on $\{p_i\}$ such that*

$$c < \frac{d(\mu_P * \beta)}{d\lambda}(z) < c^{-1},$$

for all z in a $\frac{r}{4d}$ neighborhood of $\mathcal{M}$.

Proof For any $\epsilon \in (0,1)$, let $\beta_\epsilon(\epsilon r v) = c^\epsilon_\beta(1-\|v\|^2)^{d+2}$ if $|v| \leq 1$, and $\beta_\epsilon(\epsilon r v) = 0$ if $|v| > 1$. Here c^ϵ_β is chosen so that β_ϵ integrates to 1 over $\mathbb{R}^n$. Let Vor_i denote the open set of all points $p \in \mathbb{R}^n$ such that for all $j \neq i$, $|p-p_i| < |p-p_j|$. Let

$$\mu_P(p_i) = (c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta_\epsilon)(\mathrm{Vor}_i).$$

We note

$$\frac{d(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta)}{d\lambda}(z)$$

is a $\frac{d}{cr}-$Lipschitz function of z, and this continues to be true for

$$\frac{d(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta * \beta_\epsilon)}{d\lambda}(z),$$

for any $\epsilon \in (0,1)$. Further, there exists an $\epsilon_0 \in (0,1)$ such that

$$\forall \epsilon \in (0,\epsilon_0), \left\| \frac{d(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta * \beta_\epsilon)}{d\lambda} - \frac{d(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta)}{d\lambda} \right\|_{\mathcal{L}^\infty(\mathbb{R}^n)} < c(\epsilon),$$

where $\lim_{\epsilon \to 0} \frac{c(\epsilon)}{\epsilon}$ exists and is finite. It thus suffices to prove that for all $\epsilon \in (0,\epsilon_0)$,

$$\left\| \frac{d(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta * \beta_\epsilon)}{d\lambda} - \frac{d(\mu_P * \beta)}{d\lambda} \right\|_{\mathcal{L}^\infty(\mathbb{R}^d)} < \frac{c}{2} - c(\epsilon)$$

for all z in a $\frac{r}{4d}-$neighborhood of $\mathcal{M}$. For any i, the diameter of

$$supp(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta_\epsilon) \cap \mathrm{Vor}_i$$

is less than $\frac{cr}{d}$. Let π denote the map defined on $supp(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta_\epsilon)$ from Vor_i to p_i. Then,

$$\left| \frac{d(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta * \beta_\epsilon)}{d\lambda}(z) - \frac{d(\mu_P * \beta)}{d\lambda}(z) \right| \quad = \quad \left| \frac{d(((c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta_\epsilon) - \mu_P) * \beta)}{d\lambda}(z) \right|.$$

For any $w \in supp(c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta_\epsilon) \cap \mathrm{Vor}_i$, $|\pi(w) - w| < \frac{cr}{d}$. Let $c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta * \beta_\epsilon$ be denoted ν. Then,

$$\begin{aligned} \frac{(\nu - \mu_P) * \beta}{d\lambda}(z) &= \int_{z+supp(\beta)} \nu(dx)\beta(z-x) - \int_{z+supp(\beta)} \mu_P(dy)\beta(z-y) \\ &= \int_{z+supp(\beta)} \nu(dx)\beta(z-x) - \int_{z+supp(\beta)} \nu(dx)\beta(z-\pi(x)). \end{aligned}$$

The Lemma follows noting that β is $\frac{d}{cr}-$Lipschitz.

■

Let λ^i_d denote the $d-$dimensional Lebesgue measure restricted to the disc D_i.

Recall that

$$\mu_P(p_i) = (c_\beta \mathcal{H}^d_{\mathcal{M}} * \beta_\epsilon)(\mathrm{Vor}_i).$$

Let

$$\tilde{\mu_P}(p_i) = (c_\beta \lambda^i_d)(\mathrm{Vor}_i \cap D_i).$$

By making $\frac{r}{\tau} < \frac{1}{Cd^C}$ for a sufficiently large universal constant C, and ϵ a sufficiently small quantity, we see that for each i,

$$c \le \frac{\tilde{\mu_P}(p_i)}{\mu_P(p_i)} \le c^{-1}.$$

for a suitable universal constant c. We see that $(c_\beta \lambda^i_d)(\mathrm{Vor}_i \cap D_i)$ is the volume of the polytope $\mathrm{Vor}_i \cap D_i$ multiplied by c_β, and $\mathrm{Vor}_i \cap D_i$ is given by a membership oracle, whose answer to any query takes time $(Cd)^d$. Thus by placing a sufficiently fine grid, and counting the lattice points in $\mathrm{Vor}_i \cap D_i$, $\tilde{\mu_P}(p_i)$ can be computed using $(Cd)^{2d}$ deterministic steps. Even faster randomized algorithms exist for the task, which we choose not to delve into here. ■

Appendix E. Proof of Claim 1

Proof Let

$$\mathcal{M}_{\Pi} := \cup_{\hat{d}=0}^{n} \mathcal{M}_{\Pi}^{\hat{d}}.$$

The various connected components of $\mathcal{M}_{\Pi}$ are the different $\mathcal{M}_{\Pi}^{d}$ (whose dimensions are respectively $(n-d)d$), and by evaluating Frobenius norms, we see that the distance between any two points on distinct connected components is at least 1. Since it suffices to show that a normal disc bundle of radius less than $1/2$ injectively embeds into the ambient space (which is $\mathbb{R}^{n(n-1)/2}$,) it suffices to show that

$$reach(\mathcal{M}_{\Pi}) = 1/2.$$

Let $x \in \mathcal{M}_{\Pi}^{d}$. Let z belong to the normal fiber at x and let $\|x-z\|_F < 1/2$. Without loss of generality we may (after diagonalization if necessary) take $x = diag(1,\dots,1,0,\dots,0)$ where the number of $1s$ is d and the number of $0s$ is $n-d$. Further, (using block diagonalization if necessary), we may assume that z is a diagonal matrix as well. All the eigenvalues of z lie in $(1/2, 3/2)$ and further the span of the corresponding eigenvectors is the space of eigenvectors of x corresponding to the eigenvalue 1. Therefore $\Pi_{hi}(z)$ is well defined through Cauchy's integral formula and equals x. Thus the normal discs of radius $< 1/2$ do not intersect, and so $reach(\mathcal{M}_{\Pi}^{d}) \geq 1/2$. Conversely, $\mathcal{M}_{\Pi}^{0}$ is the origin and $\mathcal{M}_{\Pi}^{1}$ contains the point $diag(1,0,\dots,0)$. We see that $diag(1/2,0,\dots,0)$ is equidistant from $\mathcal{M}_{\Pi}^{0}$ and $\mathcal{M}_{\Pi}^{1}$ and the distance is $1/2$. Therefore $reach(\mathcal{M}_{\Pi}^{d}) \leq 1/2$. Therefore,

$$reach(\mathcal{M}_{\Pi}^{d}) \geq reach(\mathcal{M}_{\Pi}) = 1/2.$$

■

Appendix F. Proofs of Lemma 8 and Lemma 9

Proof [Proof of Lemma 8]

$$\begin{aligned}
\|(\partial_v \alpha_i(x))_{i\in[\bar{N}]}\|_{\frac{d+2}{d+1}} &\leq \frac{\|(\partial_v \tilde{\alpha}_i(x))_i\|_{\frac{d+2}{d+1}}}{\tilde{\alpha}} + \frac{\|((\partial_v \tilde{\alpha}(x))\tilde{\alpha}_i(x))_i\|_{\frac{d+2}{d+1}}}{\tilde{\alpha}^2} && (164)\\
&\leq (c^{-1})\|Cd(\tilde{\alpha}_i(x))_i\|_1^{\frac{d+1}{d+2}} + (c^{-2})\|(\tilde{\alpha}_i(x))_i\|_{\frac{d+2}{d+1}}|\partial_v\tilde{\alpha}| && (165)\\
&\leq Cd + C|\partial_v \tilde{\alpha}| && (166)\\
&\leq Cd + C\|(\partial_v \tilde{\alpha}_i)_i \in [\bar{N}]\|_{\frac{d+2}{d+1}}\|(1)_i\|_{d+2} && (167)\\
&\leq Cd^2. && (168)
\end{aligned}$$

■

Proof [Proof of Lemma 9]

$$\begin{aligned}
\|(\partial_v^2 \alpha_i(x))_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} &= \|(\partial_v^2 \frac{\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)})_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} && (169)\\
&= \|(\frac{\partial_v^2 \tilde{\alpha}_i(x)}{\tilde{\alpha}(x)} + \frac{(-2)(\partial_v \tilde{\alpha}_i(x))(\partial_v \tilde{\alpha}(x))}{\tilde{\alpha}(x)^2} + \frac{\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)^3}(2(\partial_v\tilde{\alpha})^2 - \partial_v^2\tilde{\alpha}(x)(\tilde{\alpha}(x)))_{i\in[\bar{N}]}\|_{\frac{d+2}{d}}.
\end{aligned}$$

We use the triangle inequality on the above expression, and reduce the task of obtaining an upper bound to that of separately obtaining the following bounds.

Claim 6 *We have*

$$\|(\frac{\partial_v^2\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)})_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} \leq Cd^2, \tag{170}$$

Proof This follows from $c < \tilde{\alpha} < C$. ■

Claim 7 *We have*

$$\|(\frac{(-2)(\partial_v\tilde{\alpha}_i(x))(\partial_v\tilde{\alpha}(x))}{\tilde{\alpha}(x)^2})_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} \leq Cd^3, \tag{171}$$

Proof We have seen that $|\partial_v\tilde{\alpha}(x)| < Cd^2$. Therefore,

$$\begin{aligned}
\|(\frac{(-2)(\partial_v\tilde{\alpha}_i(x))(\partial_v\tilde{\alpha}(x))}{\tilde{\alpha}(x)^2})_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} &< Cd^2\|((\partial_v\tilde{\alpha}_i(x)))_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} && (172)\\
&\leq Cd^2\|((\partial_v\tilde{\alpha}_i(x)))_{i\in[\bar{N}]}\|_{\frac{d+2}{d+1}} && (173)\\
&\leq Cd^3. && (174)
\end{aligned}$$

■

Claim 8

$$\|(\frac{\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)^3}(2(\partial_v\tilde{\alpha})^2 - \partial_v^2\tilde{\alpha}(x)(\tilde{\alpha}(x)))_{i\in[\bar{N}]}\|_{\frac{d+2}{d}} \leq Cd^4. \tag{175}$$

Proof The only term that we have not already bounded is $|\partial_v^2\tilde{\alpha}(x)|$. To bound this, we observe that

$$\begin{aligned}
C|\partial_v^2\tilde{\alpha}| &\leq C\|(\partial_v^2\tilde{\alpha}_i)_i\|_{\frac{d+2}{d}}\|(1)_i\|_{(d+2)/2} && (176)\\
&\leq Cd^4. && (177)
\end{aligned}$$

Therefore, the entire expression gets bounded by Cd^4 as well. ■

■