arXiv:1910.05084v3 [math.ST] 15 Apr 2021

Fitting a manifold of large reach to noisy data

Charles Fefferman[*] Sergei Ivanov[†] Matti Lassas[‡]
Hariharan Narayanan[§]

April 16, 2021

To the memory of Yaroslav Kurylev.

Abstract

Let $\mathcal{M} \subset \mathbb{R}^n$ be a C^2-smooth compact submanifold of dimension d. Assume that the volume of $\mathcal{M}$ is at most V and the reach (i.e. the normal injectivity radius) of $\mathcal{M}$ is greater than τ. Moreover, let μ be a probability measure on $\mathcal{M}$ which density on $\mathcal{M}$ is a strictly positive Lipschitz-smooth function. Let $x_j \in \mathcal{M}$, $j = 1, 2, \dots, N$ be N independent random samples from distribution μ. Also, let ξ_j, $j = 1, 2, \dots, N$ be independent random samples from a Gaussian random variable in $\mathbb{R}^n$ having covariance $\sigma^2 I$, where σ is less than a certain specified function of d, V and τ. We assume that we are given the data points $y_j = x_j + \xi_j$, $j = 1, 2, \dots, N$, modelling random points of $\mathcal{M}$ with measurement noise. We develop an algorithm which produces from these data, with high probability, a d dimensional submanifold $\mathcal{M}_o \subset \mathbb{R}^n$ whose Hausdorff distance to $\mathcal{M}$ is less than Δ for $\Delta > Cd\sigma^2/\tau$ and whose reach is greater than $c\tau/d^6$ with universal constants $C, c > 0$. The number N of random samples required depends almost linearly on n, polynomially on Δ^{-1} and exponentially on d.

Contents

[*]Princeton University, Mathematics Department, Fine Hall, Washington Road, Princeton NJ, 08544-1000, USA.
[†]St. Petersburg Department of Steklov Institute of Mathematics, Russian Academy of Sciences, 27 Fontanka, 191023 St. Petersburg, Russia.
[‡]University of Helsinki, Department of Mathematics and Statistics, P.O. Box 68, 00014, Helsinki, Finland.
[§]School of Technology and Computer Science, Tata Institute for Fundamental Research, Mumbai 400005, India.

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. 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 a small 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?

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 general category. Cryo-EM images are X-ray images of three-dimensional macromolecules, e.g. viruses, 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 [51]). 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 A note on constants

In the following sections, we will denote positive absolute constants by $c, C, C_1, C_2, \bar{c}_1$ etc. These constants are universal and positive, but their precise value may differ from occurrence to occurrence.

1.2 Model

Let $\mathcal{M}$ be a d dimensional $\mathcal{C}^2$ submanifold of $\mathbb{R}^n$. We assume $\mathcal{M}$ has volume ($d-$dimensional Hausdorff measure) less or 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 $d\mu/d\lambda_{\mathcal{M}}$ is bounded above and below by ρ_{max}/V and ρ_{min}/V respectively, where ρ_{max} and ρ_{min} lie in $[c, C]$ and $\ln\left(d\mu/d\lambda_{\mathcal{M}}\right)$ is $C/\tau-$Lipschitz (as specified in (2)). Thus, we assume that

$$c < \rho_{min} < V d\mu/d\lambda_{\mathcal{M}} < \rho_{max} < C, \tag{1}$$

and also for all $x, y \in \mathcal{M}$,

$$\frac{\left|\ln\left(d\mu/d\lambda_{\mathcal{M}}(x)\right) - \ln\left(d\mu/d\lambda_{\mathcal{M}}(y)\right)\right|}{|x-y|} \le \frac{C}{\tau}. \tag{2}$$

Definition 1.1. *Let $G_\sigma^{(n)}$ denote the Gaussian distribution supported on $\mathbb{R}^n$ whose density (Radon-Nikodym derivative with respect to the Lebesgue measure) at x is*

$$\pi_{G_\sigma^{(n)}}(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right). \tag{3}$$

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 $G_\sigma^{(n)}$. We observe

$$y_i = x_i + \zeta_i, \quad \text{for } i = 1, 2, \dots, N,$$

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 $G_\sigma^{(n)}$. This is denoted by $\mu * G_\sigma^{(n)}$. Let ω_d be the volume of a d dimensional unit Euclidean ball. Suppose that

$$\sigma < r_c D^{-1/2}, \quad \text{where } r_c := cd^{-C}\tau,\ D = \min\left(n, \frac{V}{c^d \omega_d \beta^d}\right),\ \beta = \tau\sqrt{\frac{c^d\omega_d\tau^d}{V}}. \tag{4}$$

and $\Delta \geq \frac{Cd\sigma^2}{\tau}$. We observe $y_1, y_2, \dots, y_N$ and for $k \geq 3$, will produce a description of a $\mathcal{C}^k-$manifold $\mathcal{M}_o$ such that the Hausdorff distance between $\mathcal{M}_o$ and $\mathcal{M}$ is at most Δ and $\mathcal{M}_o$ has reach that is bounded below by $\frac{c\tau}{d^6}$ with probability at least $1 - \eta$. Note that the required upper bound on σ does not degrade to 0 with as the ambient dimension $n \to \infty$, but can be controlled by the intrinsic parameters d, V and τ. The following is our main theorem.

Theorem 1.1. *Let $\mathcal{M}$ be a d dimensional $\mathcal{C}^2$ submanifold of $\mathbb{R}^n$. We assume $\mathcal{M}$ has volume ($d-$dimensional Hausdorff measure) less or equal to V, reach (i.e. normal injectivity radius) greater or equal to τ, and that $\mathcal{M}$ has no boundary.*

Let μ be a probability measure on $\mathcal{M}$ which density with respect to Hausdorff measure of $\mathcal{M}$ satisfies (1) and (2) with bounds $\rho_{max}, \rho_{min} \in [c, C]$. Let $k \in [3, C]$ be a fixed integer. Let $x_j \in \mathcal{M}$, $j = 1, 2, \dots, N$ be N independent random samples from distribution μ. Also, let ξ_j, $j = 1, 2, \dots, N$ be independent random samples from a Gaussian random variable in $\mathbb{R}^n$ having covariance $\sigma^2 I$, see (3). Suppose that $\sigma < r_c D^{-1/2}$, where r_c and D are given in (4) and $\Delta \geq \frac{Cd\sigma^2}{\tau}$. Let $0 < \eta < 1$.

$$N_0 = \frac{CV}{\omega_d r_c^d} \ln\left(\frac{CV}{\omega_d r_c^d}\right),$$
$$N = \left(\frac{n\tau\Delta}{d} + \tau^2\right)\left(\frac{d}{\Delta}\right)^2 \left(\frac{r_c\sqrt{d}}{\sqrt{\tau\Delta}}\right)^d N_0 (\log N_0)^3 \log(\eta^{-1}).$$

Suppose we observe the data points

$$y_i = x_i + \xi_i \quad \text{for } i = 1, 2, \dots, N.$$

Then with probability at least $1 - \eta$, using these data we can construct a $\mathcal{C}^k-$manifold $\mathcal{M}_o \subset \mathbb{R}^n$ such that the Hausdorff distance between $\mathcal{M}_o$ and $\mathcal{M}$ is at most Δ and $\mathcal{M}_o$ has reach that is bounded below by $c\tau/d^6$.

The above theorem may seem counterintuitive, in that the smaller σ is, the larger N is. However note that the Hausdorff distance that we are achieving is $O(\sigma^2)$, which itself decreases quadratically as σ tends to zero. This is the reason for the anomaly. We believe that going below a Hausdorff distance of $O(\sigma^2)$ in the case of $\mathcal{C}^2$ manfolds would take different techniques and a significantly larger number of samples. Indeed that this is the case for sufficiently small Hausdorff distances was shown in [53]. In Proposition 7.1, we obtain an explicit bound on the magnitude of the third derivatives of $\mathcal{M}_0$. We emphasize that in Theorem 1.1 the Hausdorff distance of the constructed manifold $\mathcal{M}_o$ and the original manifold $\mathcal{M}$ as well as the reach of $\mathcal{M}_o$ do not depend on the dimension n of the ambient space. To prove Theorem 1.1 we develop an algorithm, using a number of analytic tools, which ensures that the degradation of the reach is polynomial and not exponential in the dimension of the manifold, d. We believe that this is the first time this have been achieved. Secondly the number of samples required depends almost linearly on the ambient dimension n. This is the second novel feature of our algorithm. A detailed comparison to earlier results is given in Subsection 1.3.

1.3 A survey of related work

Let $f : K \to \mathbb{R}$ be a function defined on a given (arbitrary) set $K \subset \mathbb{R}^n$, and let $m \geq 1$ be a given integer. The classical Whitney problem is the question whether f extends to a function $F \in C^m(\mathbb{R}^n)$ and if such an F exists, what is the optimal C^m norm of the extension. Furthermore, one is interested in the questions if the derivatives of F, up to order m, at a given point can be estimated, or if one can construct extension F so that it depends linearly on f.

These questions go back to the work of H. Whitney [96, 97, 98] in 1934. In the decades since Whitney's seminal work, fundamental progress was made by G. Glaeser [56], Y. Brudnyi and P. Shvartsman [18, 19, 20, 21, 22, 23] and [86, 87, 88], and E. Bierstone-P. Milman-W. Pawluski [11]. (See also N. Zobin [103, 104] for the solution of a closely related problem.)

The above questions have been answered in the last few years, thanks to work of E. Bierstone, Y. Brudnyi, C. Fefferman, P. Milman, W. Pawluski, P. Shvartsman and others, (see [11, 17, 18, 20, 21, 23, 41, 42, 43, 44, 45].) Along the way, the analogous problems with $C^m(\mathbb{R}^n)$ replaced by $C^{m,\omega}(\mathbb{R}^n)$, the space of functions whose m^{th} derivatives have a given modulus of continuity ω, (see [44, 45]), were also solved.

The solution of Whitney's problems has led to a new algorithm for interpolation of data, due to C. Fefferman and B. Klartag [46, 47], where the authors show how to compute efficiently an interpolant $F(x)$, whose C^m norm lies within a factor C of least possible, where C is a constant depending only on m and n.

In traditional manifold learning, for instance, by using the ISOMAP algorithm introduced in the seminal paper [91], one often aims to map points X_j to points $Y_j = F(X_j)$ in an Euclidean space $\mathbb{R}^m$, where $m \geq n$ is as small as possible so that the Euclidean distances $\|Y_j - Y_k\|_{\mathbb{R}^m}$ are close to the intrinsic distances $d_M(X_j, X_k)$ and find a submanifold $\tilde{M} \subset \mathbb{R}^m$ that is close to the points Y_j. This method has turned out to be very useful, in particular in finding the topological manifold structure of the manifold (M, g). It has been shown that when the original manifold (M, g) has a vanishing Riemann curvature and satisfies certain convexity conditions, the manifold reconstructed by the ISOMAP approaches the original manifold as the number of the sample points tends to infinity (see the results in [25, 36, 37] for ISOMAP and [101] for the continuum version of ISOMAP). We note that for a general Riemannian manifold, the construction of a map $F : M \to \mathbb{R}^m$, for which the intrinsic metric of the embedded manifold $F(M) = \tilde{M} \subset \mathbb{R}^m$ is isometric to (M, g) is a very difficult task numerically as it means finding a map, the existence of which is proved by the Nash embedding theorem (see [71, 72] and [92] on numerical techniques based on the Nash embedding theorem). We emphasize that the construction of an isometric embedding $f : M \to \mathbb{R}^n$ is outside of the context of the paper.

One can overcome the difficulties related to the construction of the Nash embedding by formulating the problem in a coordinate invariant way: Given the geodesic distances of points sampled from a Riemannian manifold (M, g), construct a manifold M^* with an intrinsic metric tensor g^* so that the Lipschitz distance of (M^*, g^*) to the original manifold (M, g) is small. The construction of abstract manifolds from the distances of sampled data points has also been considered by Coifman and Lafon [32] and Coifman et al. [30, 31] using "Diffusion Maps", and by Belkin and Niyogi [6] using "EigenMaps", where the data points are mapped to the values of the approximate eigenfunctions or diffusion kernels at the sample points. These methods construct a non-isometric embedding of the manifold M into $\mathbb{R}^m$ with a sufficiently large m. This construction is continued in [65] by computing an approximation the metric tensor g by using finite differences to find the Laplacian of the products of the local coordinate functions. In [50], we extend the results of [49] that deals with the question how a smooth manifold, that approximates a manifold (M, g), can be constructed, when one is given the distances of the points of in a discrete subset X of M with small deterministic errors. In this paper we extend these results to two directions. First, the discrete set is randomly sampled and the distances have (possibly large) random errors. Second, we consider the case when some distance information is missing.

The question of fitting a manifold to data is of interest to data analysts and statisticians [1, 27, 58, 53, 54, 64, 89]. We will focus our attention on 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 [55], building over [74] provides an upper bound on the Hausdorff distance between the output manifold and the true manifold equal to $O((\frac{\log N}{N})^{\frac{2}{n+8}}) + O(\sigma^2 \log(\sigma^{-1}))$. Note that in order to obtain a Hausdorff distance of $c\epsilon$, one needs more than $\epsilon^{-n/2}$ samples, where n is the ambient dimension. This bound is exponential in n and thus differs significantly from our results.

The results of the present work guarantee (for σ satisfying (4)) that the Hausdorff distance between the output manifold and the true manifold $\mathcal{M}$ is less than

$$\frac{C\sigma^2 d}{\tau} = O(\sigma^2)$$

with probability at least $1 - \xi$ (with less than N samples). Thus our bound on the Hausdorff distance is $O(\sigma^2)$ which is an improvement over $O(\sigma^2 \log(\sigma^{-1}))$, and also, the number of samples needed to get there depends exponentially on the intrinsic dimension d, but linearly on n. The upper bound on the number of samples depends polynomially on σ^{-1}, the exponent being $d + 4$. Moreover, if the ambient dimension n increases while σ decreases, in such a manner as to have $n\sigma^2 > c\tau^2$, we have the exponent of σ^{-1} to be $d + 2$.

The results of [48] guarantee (for sufficiently small σ) a Hausdorff distance of

$$Cd^7(\sigma\sqrt{n})$$

with less than

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

samples, where d is the dimension of the submanifold, V is in upper bound in the d dimensional volume, and σ is the standard deviation of the noise projected in one dimension. A preliminary version of our results, under stronger conditions, and only a part of our algorithm to construct $\mathcal{M}_o$, appeared in the proceedings of a conference [48]. However, the present work improves the results of [48] in the following two ways. Firstly, the upper bound on the standard deviation σ of the permissible noise is independent of the ambient dimension, while in [48] this upper bound depended inversely on the square root of the ambient dimension. Secondly, the bound on the Hausdorff distance between the output manifold and the true manifold is less than $\frac{Cd\sigma^2}{\tau}$ rather than $Cd^7\sigma\sqrt{n}$ as was the case in [48], which for permissible values of σ is significantly smaller.

As shown in [53] the question of manifold estimation with additive noise, in certain cases can be viewed as a question of regression with errors in variables [38]. The asymptotic rates that can be achieved in the latter question are extremely slow. The results of [53] imply among other things the following. Suppose that in a manifold that is the graph of $y = \sin(x + \phi)$ we wish to identify with constant probability, the phase ϕ to within an additive error of at most ϵ, from samples of the form $(x + \eta_1, y + \eta_2)$ where x, η_1 and η_2 are standard Gaussians. Then the number of samples needed is at least $\exp(C/\epsilon)$.

Definition 1.2. *Given two subsets X and Y of a metric space (M, d_M), we denote by $dist(X, Y)$, the one-sided distance from X to Y which equals $\sup_{x\in X} \inf_{y\in Y} d_M(x, y)$. We denote the Hausdorff distance between X and Y, which equals $\max(dist(X, Y), dist(Y, X))$ by $\mathbf{d}_{\mathtt{haus}}(X, Y)$. When X is a singleton $\{x\}$, we abbreviate $dist(\{x\}, Y)$ to $dist(x, Y)$.*

It follows that it is not possible to provide sample complexity bounds with a inverse polynomial dependence on $\mathbf{d}_{\mathtt{haus}}(\mathcal{M}_o, \mathcal{M})$, where the Hausdorff distance is arbitrarily small.

Finally, we mention that there is an interesting body of literature [4, 28] in computational geometry that deals with fitting piecewise linear manifolds (as opposed to $C^2 - smooth$ manifolds) to data. The paper [28] 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 $\mathcal{M}$.

1.4 Overview of sections of this paper.

1. In Section 2 we discuss preliminaries needed for working with $\mathcal{C}^2$ submanifolds of positive reach, as well as record some analytic and probabilistic facts needed subsequently.

2. In Section 3, we start with projecting the raw data points on to a D dimensional linear subspace S which is such that it minimizes the sum of the squares of the distances to the points. The span of the eigenvectors corresponding to the top D eigenvalues of the covariance matrix of the points, is such a linear subspace (the probability of ties in the eigenvalues is 0). Projection on to this subspace reduces the variance of the noise by a multiplicative factor of $\frac{D}{n}$, while the reach of the projected manifold is almost as large as the reach of the original manifold (see Lemma 3.3).

3. In Section 4, we find a family of d dimensional putative discs D_i of radius roughly $\sqrt{\sigma\sqrt{D}\tau}$ that approximates the set of projected points to within a Hausdorff distance of order roughly $\sigma\sqrt{D}$. We then view the projected data as points in a fiber bundle over a base space. The base space is the disjoint union of the discs in the family mentioned in (2). Each fiber is a disc of dimension $D - d$ and radius $\sigma\sqrt{D}$ centered at its basepoint.

4. In Section 5, in each disc D_i, we consider a set of lattice points and for each lattice point $x \in D_i$ we consider the Voronoi region in D_i of points closer to x than to the other lattice points in D_i (see Figure 5). This region is a $d-$dimensional cube unless x is close to the boundary of D_i. We compute the average of all points in fibers that have as their basepoint, a point in the Voronoi region just mentioned. The distance of individual points in the set Rnet of averages so obtained is within roughly $\frac{d\sigma^2}{\tau}$ of the original manifold, but the points are now contained in $\mathbb{R}^n$ and not necessarily S. Using this "refined net" Rnet, we now, for a second time, find a family of discs $\{D'_i\}_{i\in[N_3]}$ that approximate the new data in Hausdorff distance. However this time, the discs have radius roughly $\sqrt{d}\sigma$. and the Hausdorff distance between the new data set and the union of the discs is of the order of $\frac{d\sigma^2}{\tau}$. Further we prove that the Hausdorff distance between the union of the discs and the manifold $\mathcal{M}$ is also of the order of $\frac{d\sigma^2}{\tau}$ (see Lemma 5.6).

5. In Section 6, we design a set of weights $\{\alpha_i\}_{i\in[N_3]}$ associated with the set of discs $\{D'_i\}_{i\in[N_3]}$ to be used in the next step. These weights play a crucial role in obtained a lower bound on the reach of the output manifold $\mathcal{M}_o$ that differs from the reach of $\mathcal{M}$ by a factor that is polynomially bounded in d. Since a point on the submanifold $\mathcal{M}$ can be contained inside exponential in d many Euclidean balls of dimension n and radius less than $c\tau$ in any cover of $\mathcal{M}$ with such balls, if one does not take special care their interaction leads to the reach being potentially smaller by a multiplicative factor that is exponential in d.

6. In Section 7, we consider n dimensional balls $U_i \subseteq S$ containing the respective discs D'_i and having the same radius. We use the union of these balls to construct the output manifold as follows. We construct a vector bundle in which the base space is $\bigcup_i U_i$, and the fiber at a point x is a $n - d$ dimensional affine subspace that is roughly orthogonal to the affine span of the disc D'_i closest to x. This step is uses partitions of unity for defining a subspace as a kind of weighted average of subspaces (See Definition 7.1), with specially designed weights discussed in Section 6. For each disc D'_i, we consider a bump function supported on U_i corresponding to a partition of unity for $\bigcup_{i\in[N_3]} U_i$, which we use to generate a function F that approximates the squared distance to the manifold from the individual squared distances

to the discs. Finally, the output manifold is defined to be the set of all points x at which F is stationary on the fiber at x (see Definition 7.2). This manifold is $\mathcal{C}^k$ and a lower bound on the reach of this manifold is obtained that has a polynomial rather than exponential dependence on d. A concrete upper bound on the third derivatives of the manifold viewed as the graph of a function is obtained in Proposition 7.1.

2 Geometric Preliminaries

We need the following form of the Gaussian concentration inequality, which may be found in Proposition 1.5.7 of [90]. Recall from Definition 1.1 that $G_\sigma^{(m)}$, is the centered Gaussian distribution with variance σ^2 supported on $\mathbb{R}^m$ whose density at x is given by $\rho_{G_\sigma}(x) := \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$.

Lemma 2.1 (Gaussian concentration)**.** *Let $g : \mathbb{R}^m \to \mathbb{R}$ be a $1-$Lipschitz function and $a = \mathbb{E}g(X)$ where X is a random variable having distribution respect to $G_\sigma^{(m)}$. Then, for $t > 0$,*

$$G_\sigma^{(m)}\{x : |g(x) - a| \geq t\sigma\} \leq C \exp\left(-ct^2\right). \tag{5}$$

for some absolute constants c, C.

Definition 2.1. *For a closed $A \subseteq \mathbb{R}^m$, and $a \in A$, let the tangent space (in the sense of Federer [39]) $Tan^0(a, A)$ denote the set of all vectors $v \in \mathbb{R}^D$ 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 $X \subseteq \mathbb{R}^m$ and a point $p \in \mathbb{R}^m$, 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 2.2 (Reach)**.** *The reach of a closed set $A \subseteq \mathbb{R}^m$, 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, [39]), gives an alternate characterization of the reach.

Proposition 2.1 (Federer's reach criterion)**.** *Let A be a closed subset of $\mathbb{R}^m$. Then,*

$$\text{reach}(A)^{-1} = \sup\left\{2|b - a|^{-2} dist(b, Tan(a)) \,\middle|\, a, b \in A, a \neq b\right\}. \tag{6}$$

Corollary 2.2. *Suppose $a, b \in A$ and $|a - b| < reach(A) := \tau$. Let $\Pi_a b$ denote the unique nearest point to b in $Tan(a)$. Then,*

$$\frac{dist(b, Tan(a))}{\tau} \le \left(\frac{|a - \Pi_a b|}{\tau}\right)^2. \tag{7}$$

Proof. We have

$$\frac{dist(b, Tan(a))}{\tau} \le \frac{|a-b|^2}{2\tau^2} = \frac{|a - \Pi_a b|^2 + dist(b, Tan(a))|^2}{2\tau^2}. \tag{8}$$

Solving the quadratic inequality in $dist(b, Tan(a))$, we get

$$\frac{dist(b, Tan(a))}{\tau} \quad \le \quad 1 - \sqrt{1 - \left(\frac{|a - \Pi_a b|}{\tau}\right)^2} \le \left(\frac{|a - \Pi_a b|}{\tau}\right)^2. \tag{9}$$

□

Definition 2.3. *We say that $\mathcal{M}$ is a compact d dimensional C^2 submanifold of $\mathbb{R}^m$ if $\mathcal{M}$ is compact and the following is true. Firstly, for any $p \in \mathcal{M}$, the tangent space at p is a d dimensional affine subspace of $\mathbb{R}^m$, and therefore, by an orthogonal transformation, one can choose Euclidean coordinates in $\mathbb{R}^m$ so that the tangent space at p has the form $Tan(p) = \{(z_1, z_2) \in \mathbb{R}^d \oplus \mathbb{R}^{m-d} | Az_1 + b = z_2\}$ for some matrix $A = A(p)$ and vector $b = b(p)$. Secondly, there exists a neighborhood $V \subseteq \mathbb{R}^m$ of p, an open set $U \subset \mathbb{R}^d$ and a C^2-smooth function $F : U \to \mathbb{R}^{m-d}$ such that in the above coordinates*

$$\mathcal{M} \cap V = \{(u, F(u)) \big| u \in U \cap \mathbb{R}^d\}. \tag{10}$$

Let $\mathcal{G}(d, m, V, \tau)$ be the set of all d dimensional, compact $\mathcal{C}^2$ manifolds embedded in $\mathbb{R}^m$ and having reach at least τ and d dimensional Hausdorff measure less or equal to V. Let $\mathcal{M} \in \mathcal{G}(d, m, V, \tau)$. 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.

3 Projecting the manifold on to a $D-$dimensional subspace

We shall be assuming that σ is known exactly, however, all the arguments that we use go through if σ^2 is merely an upper bound on the true variance. This assumption, and how it can be made practical will be discussed at the end of this section. Further, for the purposes of the proof, we may assume that $\Delta = \frac{Cd\sigma^2}{\tau}$, because if Δ is larger by more than C than this, we can simply add i.i.d Gaussian noise of standard deviation $\sqrt{(\sigma')^2 - \sigma^2}$ to each sample, and then assume that we have samples where the standard deviation of the noise is σ' rather than σ. Here, σ' is chosen so that $\Delta = C(\sigma')^2/(d\tau)$.

This section describes the effect of Principal Component Analysis (PCA), with sufficiently many components on the hidden manifold. It is a preprocessing step involving projecting data sampled from μ on to a suitable D dimensional linear subspace. After this step, one may assume that the data is D dimensional rather than n dimensional, where D is an integer that depends only of d, V and τ as given in (4).

Suppose that we are in the following setting: there is a manifold in the class $\mathcal{G}(d, n, V, \tau)$ and a probability measure μ supported on this manifold that has a density with respect to the uniform measure on the manifold the logarithm of which is $\widetilde{L}$ Lipschitz. Let

$$\tilde{\mu} := \mu * G_\sigma^{(n)}.$$

Definition 3.1. *If $x \in \mathbb{R}^m$ and $S \subseteq \mathbb{R}^m$, we define*

$$dist(x, S) := \inf_{y \in S} |x - y|.$$

Let S be an affine subspace of $\mathbb{R}^n$. Let Π_S denote orthogonal projection onto S. Let the span of the first d canonical basis vectors be denoted $\mathbb{R}^d$ and the span of the last $n-d$ canonical basis vectors be denoted $\mathbb{R}^{n-d}$. Let ω_d be the d dimensional Lebesgue measure of the unit Euclidean ball in $\mathbb{R}^d$. Given $\alpha \in (0, 1)$, let

$$\beta := \beta(\alpha) = \sqrt{(1/10)\left(\frac{\alpha^2\tau}{2}\right)^2\left(\frac{\alpha^2\tau}{4}\right)^d\left(\frac{\omega_d\rho_{min}}{V}\right)}. \tag{11}$$

Let

$$R \geq C\sigma\sqrt{n} + \sigma \max\left(C\sqrt{\log CN/\delta}, C\sqrt{\log(Cn\sigma^2/\epsilon)}\right), \tag{12}$$

where C is a sufficiently large universal constant. Let

$$D := \left\lfloor \frac{V}{\omega_d \beta^d} \right\rfloor + 1. \tag{13}$$

Let

$$N_0 \geq C(R^2 D/\epsilon^2)\sqrt{\log(C/\delta)}. \tag{14}$$

Note that due to the slow growth of the the $\sqrt{\log N_0}$ term in (12), it is possible to set N_0 in (14) in a way that is consistant with the definition of R in (12). Let $\epsilon < \beta^2/2$. Below, $\delta > 0$ will be a small parameter that gives a bound on the probability that the conclusion $\sup_{x \in \mathcal{M}} dist(x, S) < \alpha^2\tau$ in Proposition 3.1 fails.

In fact, we may choose

$$R = C\sigma\sqrt{n} + C\sigma\sqrt{\log(Cn\sigma^2/(\epsilon\delta))}, \tag{15}$$

and

$$N_0 = \lfloor C(n\sigma^2 + \sigma^2\log(Cn\sigma^2/(\epsilon\delta)))\sqrt{\log(C/\delta)}(D/\epsilon^2)\rfloor, \tag{16}$$

where C is a sufficiently large universal constant, and in doing so simultaneously satisfy (12) and (14).

Proposition 3.1. *Given N_0 data points $\{x_1, \dots, x_{N_0}\}$ drawn i.i.d from $\tilde{\mu}$, let S be a D dimensional affine subspace that minimizes*

$$\sum_{i=1}^{N_0} dist(x_i, \tilde{S})^2, \tag{17}$$

subject to the condition that $\tilde{S}$ is an affine subspace of dimension D, and $\beta < c\tau$, where β is given by (11).

Then,

$$\mathbb{P}[\sup_{x \in \mathcal{M}} dist(x, S) < \alpha^2 \tau] > 1 - \delta. \tag{18}$$

In order to prove Proposition 3.1, we need some tools, which we proceed to develop. We will present the proof of the above proposition after presenting the proof of Lemma 3.2. We will need the following form of Hoeffding's inequality.

Lemma 3.1 (Hoeffding's Inequality). *Let $X_1, \dots, X_{N_0}$ be i.i.d copies of the random variable X whose range is $[0, 1]$. Then,*

$$\mathbb{P}\left[\left|\frac{1}{s}\left(\sum_{i=1}^{N_0} X_i\right) - \mathbb{E}[X]\right| \leq \epsilon\right] \geq 1 - 2\exp(-2N_0\epsilon^2). \tag{19}$$

Let $\mathcal{P}$ be a probability distribution supported on $B := \{x \in \mathbb{R}^n \big| \|x\| \leq 1\}$. Let $\mathbb{H} := \mathbb{H}_D$ be the set whose elements are affine subspaces $S \subseteq \mathbb{R}^n$ of dimension D, each of which intersects B. Let $\mathbb{H}^0 = \mathbb{H}_D^0$ be the set of linear subspaces of dimension D. Let $\mathcal{F}_D$ be the set of all loss functions $F(x) = dist(x, H)^2$ for some $H \in \mathbb{H}$ (where $dist(x, S) := \inf_{y \in S} \|x - y\|$). Let $\mathcal{F}_D^0$ be the set of all loss functions $F(x) = dist(x, H)^2$ for some $H \in \mathbb{H}^0$. We wish to obtain a probabilistic upper bound on

$$\sup_{F \in \mathcal{F}_d} \left|\frac{\sum_{i=1}^{N_0} F(x_i)}{N_0} - \mathbb{E}_{\mathcal{P}} F(x)\right|, \tag{20}$$

where $\{x_i\}_1^s$ is the training set and $\mathbb{E}_{\mathcal{P}} F(x)$ is the expected value of F with respect to $\mathcal{P}$. In our situation, (20) is measurable and hence a random variable because $\mathcal{F}$ is a family of bounded piecewise quadratic functions, continuously parameterized by $\mathbb{H}_d$, which has a countable dense subset, for example, the subset of elements specified using rational data. We obtain a probabilistic upper bound on (20) that is independent of n, the ambient dimension.

Lemma 3.2. *Let $x_1, \dots, x_{N_0}$ be i.i.d samples from $\mathcal{P}$, a distribution supported on the ball of radius 1 in $\mathbb{R}^m$.*

Then, firstly,

$$\mathbb{P}\left[\left\|\frac{\sum_{i=1}^{N_0} x_i}{N_0} - \mathbb{E}_{\mathcal{P}} x\right\| \leq 2\left(\sqrt{\frac{1}{N_0}}\right)\left(1 + \sqrt{2\ln(4/\delta)}\right)\right] > 1 - \delta.$$

Secondly,

$$\mathbb{P}\left[\sup_{F\in\mathcal{F}_D}\left|\frac{\sum_{i=1}^{N_0}F(x_i)}{N_0}-\mathbb{E}_{\mathcal{P}}F(x)\right|\leq 2\left(\frac{\sqrt{D}+2}{\sqrt{N_0}}\right)\left(1+\sqrt{2\ln(4/\delta)}\right)\right]>1-2\delta.$$

Proof. Any $F\in\mathcal{F}_D$ can be expressed as $F(x)=dist(x,H)^2$ where H is an affine subspace of dimension equal to D that intersects the unit ball. We see that $dist(x,H)^2$ can be expressed as

$$\left(\|x\|^2-x^\dagger A^\dagger Ax\right),$$

where A is the orthogonal projection onto the linear subspace H. Thus, F is defined using $H\in\mathbb{H}^0$, where

$$F(x):=\left(\|x\|^2-x^\dagger A^\dagger Ax\right).$$

Now, define vector valued maps Φ and Ψ whose respective domains are the space of D dimensional affine subspaces and B respectively.

$$\Phi(H):=\left(\frac{1}{\sqrt{d}}\right)A^\dagger A$$

and

$$\Psi(x):=xx^\dagger,$$

where $A^\dagger A$ and $xx^\dagger$ are interpreted as rows of n^2 real entries.

Thus,

$$\begin{aligned}F(x) &= \left(\|x\|^2-x^\dagger A^\dagger Ax\right) &(21)\\ &= \|x\|^2+\sqrt{d}\Phi(H)\cdot\Psi(x), &(22)\end{aligned}$$

where the dot product is the inner product corresponding to Frobenius norm. We see that since $\|x\|\leq 1$, the Frobenius norm (which equals the operator norm in this case) of $\Psi(x)$ is $\|\Psi(x)\|\leq 1$. The Frobenius norm $\|A^\dagger A\|_F^2$ is equal to $Tr(AA^\dagger AA^\dagger)$, which is the rank of A_i since A_i is a projection. Therefore,

$$d\|\Phi(H)\|^2\leq\|A^\dagger A\|^2\leq D$$

and

$$\mathbb{P}\left[\sup_{F\in\mathcal{F}_D^0}\left|\frac{\sum_{i=1}^{N_0}F(x_i)}{N_0}-\mathbb{E}_{\mathcal{P}}F(x)\right|>\epsilon\right]\leq p^{(1)}+p^{(2)}$$

where

$$p^{(1)}=\mathbb{P}\left[\left|\frac{\sum_{i=1}^{N_0}\|x_i\|^2}{N_0}-\mathbb{E}_{\mathcal{P}}\|x\|^2\right|>\epsilon/2\right] \tag{23}$$

and

$$p^{(2)} = \mathbb{P}\left[\sup_{H\in\mathbb{H}_D^0}\left|\frac{\sum_{i=1}^{N_0}\Phi(H)\cdot\Psi(x_i)}{N_0} - \mathbb{E}_{\mathcal{P}}\Phi(H)\cdot\Psi(x)\right| > \frac{\epsilon}{2\sqrt{D}}\right]. \tag{24}$$

The first term, namely $p^{(1)}$ can be bounded above using Hoeffding's inequality as follows

$$\mathbb{P}\left[\left|\frac{\sum_{i=1}^{N_0}\|x_i\|^2}{N_0} - \mathbb{E}_{\mathcal{P}}\|x\|^2\right| \geq \epsilon/2\right] \leq 2\exp(-N_0\epsilon^2/2). \tag{25}$$

In order to bound $p^{(2)}$, we will use the notion of Rademacher complexity described below.

Definition 3.2 (Rademacher Complexity)**.** *Given a class $\mathcal{F}$ of functions $f: X \to \mathbb{R}$ a measure μ supported on X, and a natural number $s \in \mathbb{N}$, and an $s-$tuple of points $(x_1, \dots, x_s)$, where each $x_i \in X$ we define the empirical Rademacher complexity $R_s(\mathcal{F}, x)$ as follows. Let $\sigma = (\sigma_1, \dots, \sigma_s)$ be a vector of s independent Rademacher random variables (which take values 1 and -1 with equal probability. Then,*

$$R_s(\mathcal{F}, x) := \mathbb{E}_\sigma\left(\frac{1}{s}\right)\left[\sup_{f\in\mathcal{F}}\left(\sum_{i=1}^{s}\sigma_i f(x_i)\right)\right].$$

We will use Rademacher complexities to bound the sample complexity from above. Let $X = B_n$ be the unit ball in $\mathbb{R}^n$. Let μ be a measure supported on X. Let $\mathcal{F}$ be a class of functions $f: X \to \mathbb{R}$. In our context, the functions f are indexed by elements H in H_d^0 and $f(x) = \Phi(H)\cdot\Psi(x)$ for any $x \in X$. Let μ_{N_0} denote the uniform counting probability measure on $\{x_1, \dots, x_{N_0}\}$, where $x_1, \dots, x_{N_0}$ are N_0 i.i.d draws from μ. Thus $\mathbb{E}_{\mu_{N_0}} f$ is shorthand for $(1/N_0)\sum_i f(x_i)$. We know (see Theorem 3.2, [5]) that for all $\delta > 0$,

$$\mathbb{P}_{(x_i)\sim\mathcal{P}^{N_0}}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}_\mu f - \mathbb{E}_{\mu_{N_0}} f\right| \leq 2R_{N_0}(\mathcal{F}, x) + \sqrt{\frac{2\log(2/\delta)}{N_0}}\right] \geq 1-\delta. \tag{26}$$

Applying this inequality to the term in (24) we see that

$$\mathbb{P}\left[\sup_{H\in\mathbb{H}_d^0}\left|\frac{\sum_{i=1}^{N_0}\Phi(H)\cdot\Psi(x_i)}{N_0} - \mathbb{E}_{\mathcal{P}}\Phi(H)\cdot\Psi(x)\right| < \frac{\epsilon}{2\sqrt{D}}\right] > 1-\delta, \tag{27}$$

where

$$\frac{\epsilon}{2\sqrt{d}} > \mathbb{E}_\sigma\frac{1}{N_0}\left[\sup_{H\in\mathbb{H}_D^0}\left(\sum_{i=1}^{N_0}\sigma_i\Phi(H)\cdot\Psi(x_i)\right)\right] + \sqrt{\frac{2\log(2/\delta)}{N_0}}. \tag{28}$$

In order for the last statement to be useful, we need a concrete upper bound on

$$\mathbb{E}_\sigma\frac{1}{N_0}\left[\sup_{H\in\mathbb{H}_D^0}\left(\sum_{i=1}^{N_0}\sigma_i\Phi(H)\cdot\Psi(x_i)\right)\right],$$

which we proceed to obtain. We have

$$
\begin{aligned}
\mathbb{E}_\sigma \frac{1}{N_0}\left[\sup_{H\in\mathbb{H}_D^0}\left(\sum_{i=1}^{N_0}\sigma_i\Phi(H)\cdot\Psi(x_i)\right)\right] &\leq \mathbb{E}_\sigma \frac{1}{N_0}\left[\left\|\sum_{i=1}^{N_0}\sigma_i\Psi(x_i)\right\|\right] &(29)\\
&\leq \mathbb{E}_\sigma \frac{1}{N_0}\left[\left\|\sum_{i=1}^{N_0}\sigma_i\Psi(x_i)\right\|^2\right]^{\frac{1}{2}} &(30)\\
&= \frac{1}{N_0}\left[\sum_i \|\Psi(x_i)\|^2\right]^{\frac{1}{2}} &(31)\\
&\leq \frac{1}{\sqrt{N_0}}. &(32)
\end{aligned}
$$

Plugging this into (28) we see that for

$$\epsilon > 2\left(\sqrt{\frac{d}{N_0}}\right)\left(1+\sqrt{2\ln(4/\delta)}\right)$$

and

$$\mathbb{P}\left[\sup_{F\in\mathcal{F}_d}\left|\frac{\sum_{i=1}^{s}F(x_i)}{N_0}-\mathbb{E}_{\mathcal{P}}F(x)\right|<\epsilon\right] > 1-\delta.$$

The first claim of the lemma similarly follows from (26). Let $\mathcal{P}_{N_0}$ be the uniform measure on $\{x_1,\dots,x_{N_0}\}$. A direct calculation shows that if H^0 is the translate of H containing the origin, and for any x, the foot of the perpendicular from x to H is q_x and the foot of the perpendicular from x to H^0 is q_x^0, then

$$\mathbb{E}_{\mathcal{P}_{N_0}} dist(x,H^0)^2 - \mathbb{E}_{\mathcal{P}} dist(x,H^0)^2 - (\mathbb{E}_{\mathcal{P}_{N_0}} dist(x,H)^2 - \mathbb{E}_{\mathcal{P}} dist(x,H)^2)$$

can be expressed as

$$
\begin{aligned}
(\mathbb{E}_{\mathcal{P}_{N_0}}-\mathbb{E}_{\mathcal{P}})(dist(x,H^0)^2-dist(x,H)^2) &= (\mathbb{E}_{\mathcal{P}_{N_0}}-\mathbb{E}_{\mathcal{P}})(|x-q_x^0|^2-|x-q|^2)\\
&= (\mathbb{E}_{\mathcal{P}_{N_0}}-\mathbb{E}_{\mathcal{P}})(-2\langle x, q_x^0-q\rangle+|q_x^0|^2-|q|^2).
\end{aligned}
$$

This is in magnitude less than $|2(\mathbb{E}_{\mathcal{P}_{N_0}}-\mathbb{E}_{\mathcal{P}})(x)|$ which by the first claim of the lemma is bounded by

$$4\left(\sqrt{\frac{1}{N_0}}\right)\left(1+\sqrt{2\ln(4/\delta)}\right)$$

with probability greater than $1-\delta$. The second claim of the Lemma follows.

□

Proof of Proposition 3.1. Let $x_i = y_i + z_i$ where for each $i \in [N_0]$, y_i is a random draw from μ supported on $\mathcal{M}$ and z_i is an independent Gaussian sampled from $G(0,\sigma^2)$, and the collection $\{(y_i,z_i)\}$ is independent, i. e. comes from the appropriate product distribution $(\mu\times G(0,\sigma^2))^{\times n}$.

For each i, let $\widehat{z}_i$ equal z_i if $|z_i| < R$ and let $\widehat{z}_i = 0$ otherwise. Let the distribution of $\widehat{z}_i$ be denoted $\widehat{G}$.

We shall first establish the following claim.

Claim 3.1. *If*

$$\mathbb{E}_{y\sim\mu} dist(y,S)^2 < \left(\frac{\alpha^2\tau}{2}\right)^2 \left(\frac{\alpha^2\tau}{4}\right)^d \omega_d \rho_{min}.$$

then

$$\sup_{x\in\mathcal{M}} dist(x,S) < \alpha^2\tau.$$

Proof. If

$$\sup_{x\in\mathcal{M}} dist(x,S) \geq \alpha^2\tau,$$

since $\mathcal{M}$ is compact, the supremum is achieved at some point x_0. Thus, any point within $B_{\frac{\alpha^2\tau}{2}}(x_0)$ is at a Euclidean distance of at least $\alpha^2\tau/2$ from S. Observe that, for any $x \in \mathcal{M}$, $B_{\frac{\alpha^2\tau}{2}}(x) \cap \mathcal{M}$ is the graph of a function over the orthogonal projection of $B_{\frac{\alpha^2\tau}{2}}(x) \cap \mathcal{M}$ onto $Tan(x)$, which, by Federer's reach criterion, contains a d dimensional ball of radius at least $\frac{\alpha^2\tau}{4}$. Consequently,

$$\begin{aligned}
\mathbb{E}_{y\sim\mu} dist(y,S)^2 &= \int_{\mathcal{M}} dist(y,S)^2 \mu(dy) && (33)\\
&\geq \left(\frac{\alpha^2\tau}{2}\right)^2 \inf_{\widehat{y}\in\mathcal{M}} \mu\left(B_{\frac{\alpha^2\tau}{2}}(\widehat{y}) \cap \mathcal{M}\right) && (34)\\
&\geq \left(\frac{\alpha^2\tau}{2}\right)^2 \left(\frac{\alpha^2\tau}{4}\right)^d \omega_d \rho_{min}. && (35)
\end{aligned}$$

□

Definition 3.3 ($\epsilon-$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 $dist(x, x_1) < \epsilon$.*

Note that $\mathcal{M}$ can be provided with $3\beta/2$ net with respect to Euclidean distance of size D because the volume of the intersection of an n dimensional ball of radius $3\beta/2$ centered at a point in $\mathcal{M}$ with $\mathcal{M}$ is greater than $\omega_d\beta^n$ (by Federer's reach criterion), and $\omega_d\beta^n$ is greater than $\frac{V}{D}$ by (13). Next, let $\widehat{S} \subset \mathbb{R}^n$ be the linear span of a minimal 3β net of $\mathcal{M}$. Then,

$$\mathbb{E}_{y\sim\mu} dist(y,\widehat{S})^2 \leq 9\beta^2. \tag{36}$$

Let $\epsilon < \beta^2/2$. By the definition of S,

$$\sum_{i=1}^{N} dist(x_i,S)^2 \leq \sum_{i=1}^{N} dist(x_i,\widehat{S})^2. \tag{37}$$

Claim 3.2. *By our choice of R and N_0, with probability greater than $1-\delta/2$, for all $i \in [N_0]$, $x_i = y_i + \widehat{z}_i$.*

Proof. It suffices to show that

$$I_R := \int_{|x|>R} (2\pi\sigma^2)^{n/2} \exp(-|x|^2/(2\sigma^2))dx < 1 - (1-\delta/2)^{1/N_0}. \tag{38}$$

The left hand side I_R can be bounded above as follows.

$$\begin{aligned} I_R \exp(R^2/(4\sigma^2)) &\leq \int_{\mathbb{R}^n} (2\pi\sigma^2)^{-n/2} \exp(-|x|^2/(2\sigma^2)) \exp(|x|^2/(4\sigma^2))dx \\ &= 2^{n/2}. \end{aligned}$$

From (12),

$$R \geq C\sigma\sqrt{n} + C\sigma\sqrt{\log(CN/\delta)},$$

and so

$$\begin{aligned} I_R \leq 2^{n/2}\exp(-R^2/(4\sigma^2)) &\leq 2^{n/2}\exp\left(-Cn - C\log(CN_0/\delta)\right) \qquad (39) \\ &\leq \frac{\delta}{CN_0} \qquad (40) \\ &\leq 1-(1-\delta/2)^{1/N_0}. \qquad (41) \end{aligned}$$

□

By Lemma 3.2 with probability greater than $1-\delta/2$, we have

$$\sup_{\widetilde{S}} |(1/N_0)\sum_{i=1}^{N_0} dist(y_i + \widehat{z}_i, \widetilde{S})^2 - \mathbb{E}_{(y,\widehat{z})\sim\mu\times\widehat{G}} dist(y+\widehat{z}, \widetilde{S})^2| < \epsilon/2. \tag{42}$$

Claim 3.3. *By our choice of R,*

$$\sup_{\widetilde{S}} |\mathbb{E}_{x\sim\widetilde{\mu}} dist(x, \widetilde{S})^2 - \mathbb{E}_{(y,\widehat{z})\sim\mu\times\widehat{G}} dist(y+\widehat{z}, \widetilde{S})^2| < \epsilon/2. \tag{43}$$

Proof. For any fixed $\widetilde{S}$ and $x \in \mathbb{R}^n$,

$$-\mathbb{E}_{x\sim\widetilde{\mu}} dist(x,\widetilde{S})^2 + \mathbb{E}_{(x,\widehat{z})\sim\mu\times\widehat{G}} dist(x+\widehat{z},\widetilde{S})^2 = \mathbb{E}_{\widehat{z}\sim\widehat{G}} dist(\widehat{z},\widetilde{S})^2,$$

because for the vector valued random variable $p = x - \Pi_{\widetilde{S}}x$, $|p| = dist(x,\widetilde{S})$ and we have

$$\mathbb{E}|p|^2 = \mathbb{E}|p-\mathbb{E}p|^2 + |\mathbb{E}p|^2.$$

Therefore,

$$\begin{aligned} & \sup_{\widetilde{S}} |\mathbb{E}_{x\sim\widetilde{\mu}} dist(x,\widetilde{S})^2 - \mathbb{E}_{(y,\widehat{z})\sim\mu\times\widehat{G}} dist(y+\widehat{z},\widetilde{S})^2| \\ =\; & \sup_{\widetilde{S}} \mathbb{E}_{\widehat{z}\sim\widehat{G}} dist(\widehat{z},\widetilde{S})^2 \\ =\; & \int_{|(x_{D+1},\dots,x_n)|>R} |x|^2 (2\pi\sigma^2)^{n/2}\exp(-|x|^2/(2\sigma^2))dx \\ \leq\; & \int_{|x|>R} |x|^2(2\pi\sigma^2)^{n/2}\exp(-|x|^2/(2\sigma^2))dx \\ =:\; & J_R. \end{aligned}$$

Proceeding as with the preceeding claim,

$$\begin{aligned} J_R \exp(R^2/(4\sigma^2)) &\leq \int_{\mathbb{R}^n} |x|^2 (2\pi\sigma^2)^{-n/2} \exp(-|x|^2/(2\sigma^2)) \exp(|x|^2/(4\sigma^2)) dx \\ &= 2n\sigma^2 2^{n/2}. \end{aligned}$$

Since by (12), $R \geq C\sigma\sqrt{n} + C\sigma\sqrt{\log(Cn\sigma^2/\epsilon)}$, we have $J_R \leq \epsilon/2$ and the claim follows. □

Thus, with probability greater than $1-\delta$,

$$\sup_{\widetilde{S}} |(1/N_0) \sum_{i=1}^{N_0} dist(x_i, \widetilde{S})^2 - \mathbb{E}_{x\sim\widetilde{\mu}} dist(x_i, \widetilde{S})^2| < \epsilon. \tag{44}$$

Therefore, by (37) and the above, with probability greater than $1-\delta$,

$$\mathbb{E}_{x\sim\widetilde{\mu}} dist(x, S)^2 \leq \mathbb{E}_{x\sim\widetilde{\mu}} dist(x, \widehat{S})^2 + 2\epsilon. \tag{45}$$

Therefore, expanding $x = y + z$, we have

$$\begin{aligned} \mathbb{E}_{x\sim\widetilde{\mu}} dist(y, S)^2 &\leq \mathbb{E}_{x\sim\widetilde{\mu}} dist(y, \widehat{S})^2 + 2\epsilon & (46) \\ &\leq 9\beta^2 + 2\epsilon \qquad\qquad (by\ (36)) & (47) \\ &< \left(\frac{\alpha^2\tau}{2}\right)^2 \left(\frac{\alpha^2\tau}{4}\right)^d \left(\frac{\omega_d \rho_{min}}{V}\right). & (48) \end{aligned}$$

Therefore, by Claim 3.1, with probability greater than $1-\delta$,

$$\sup_{x\in\mathcal{M}} dist(x, S) < \alpha^2\tau.$$

This proves Proposition 3.1. □

Lemma 3.3. *Suppose $\mathcal{M}$ is a C^2 submanifold of $\mathbb{R}^n$ having reach τ and S is a D dimensional linear subspace such that $\sup_{x\in\mathcal{M}} dist(x, S) < \alpha^2\tau$ where $\alpha < \frac{1}{4}$. Then $\Pi_S(\mathcal{M})$ is a submanifold of $\mathbb{R}^n$ having reach at least $(1-4\alpha^2)\tau$.*

Proof. Without loss of generality, set $\tau = 1$. Since $\mathcal{M}$ is compact, and Π_S is continuous, $\Pi_S(\mathcal{M})$ is compact. We will first show that the reach of $\Pi_S(\mathcal{M})$ is greater than $1 - 4\alpha^2$. Note that for any $x, y \in \mathcal{M}$ such that $|x - y| > \alpha$,

$$\frac{|\Pi_S(x-y)|}{|x-y|} = \sqrt{1 - \frac{|\Pi_{S^\perp}(x-y)|^2}{|x-y|^2}} \geq \sqrt{1-4\alpha^2}. \tag{49}$$

Let $|x - y| \leq \alpha$. Let

$$\widehat{U} := \{y \in \mathbb{R}^n \big| |y - \Pi_x y| \leq 1/4\} \cap \{y \in \mathbb{R}^n \big| |x - \Pi_x y| \leq 1/4\}.$$

As $\mathcal{M}$ is a C^2 submanifold of $\mathbb{R}^n$, by Lemma A.2, there exists a C^2 function $F_{x,\widehat{U}}$ from $\Pi_x(\widehat{U})$ to $\Pi_x^{-1}(\Pi_x(0))$ such that

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

By Corollary 2.2,

$$|F_{x,\widehat{U}}(y)| \;\leq\; |x - \Pi_x y|^2 \leq \alpha^2. \tag{50}$$

Therefore, the Hausdorff distance between $\mathcal{M} \cap \widehat{U}$ and the disc $Tan(x) \cap \widehat{U}$ is less or equal to α^2. Consequently,

$$\sup_{x \in Tan(x) \cap \widehat{U}} dist(x, S) < \alpha^2 + \alpha^2 = 2\alpha^2. \tag{51}$$

In particular, this implies that the dimension of S is at least d. We observe that

$$\frac{|\Pi_S(x-y)|}{|x-y|} = \frac{|\Pi_S((x + \frac{\alpha(y-x)}{|y-x|}) - (x + \frac{\alpha(x-y)}{|x-y|}))|}{|(x + \frac{\alpha(y-x)}{|y-x|}) - (x + \frac{\alpha(x-y)}{|x-y|})|} \tag{52}$$

$$= \sqrt{1 - \left(\frac{|\Pi_{S^\perp}((x + \frac{\alpha(y-x)}{|y-x|}) - (x + \frac{\alpha(x-y)}{|x-y|}))|}{|(x + \frac{\alpha(y-x)}{|y-x|}) - (x + \frac{\alpha(x-y)}{|x-y|})|}\right)^2} \tag{53}$$

$$\geq \sqrt{1 - \left(\frac{4\alpha^2}{2\alpha}\right)^2} = \sqrt{1 - 4\alpha^2}. \tag{54}$$

For $\epsilon > 0$ let

$$U_\epsilon := \{y \in \mathbb{R}^n \big| |y - \Pi_x y| \leq 1/4\} \cap \{y \in \mathbb{R}^n \big| |x - \Pi_x y| \leq \epsilon\}. \tag{55}$$

Then,

$$U_\epsilon \cap \{x \in \mathbb{R}^n | dist(x, S) < \alpha^2\} = \{y \in \mathbb{R}^n \big| |x - \Pi_x y| \leq \epsilon\} \cap \{x \in \mathbb{R}^n | dist(x, S) < \alpha^2\}.$$

Therefore, by (51) and the Lipschitz nature of the gradient (by (227)) of $F_{x,\widehat{U}}$ if we choose a frame where the origin is $\Pi_S(x)$ and $\mathbb{R}^d$ is $\Pi_S(Tan(x))$, we see that for sufficiently small ϵ, $\Pi_S(U_\epsilon \cap \mathcal{M})$ is the graph of a C^2 function. Since $x \in \mathcal{M}$ was arbitrary, this proves that $\Pi_S \mathcal{M}$ is a submanifold of $\mathbb{R}^n$. Finally, we note that the reach of $\Pi_S \mathcal{M}$ is given by

$$reach(\Pi_S \mathcal{M}) = \inf_{\mathcal{M} \ni x \neq y \in \mathcal{M}} \frac{|\Pi_S(x-y)|^2}{2dist(\Pi_S x, \Pi_S Tan(y))} \tag{56}$$

$$\geq \inf_{\mathcal{M} \ni x \neq y \in \mathcal{M}} \frac{(1-4\alpha^2)|x-y|^2}{2dist(\Pi_S x, \Pi_S Tan(y))} \tag{57}$$

$$\geq (1-4\alpha^2) \inf_{\mathcal{M} \ni x \neq y \in \mathcal{M}} \frac{|x-y|^2}{2dist(x, Tan(y))} \tag{58}$$

$$= (1-4\alpha^2) reach(\mathcal{M}). \tag{59}$$

□

3.1 Estimating σ.

As we mentioned in the beginning of this section, we have assumed knowledge of σ. However in an interesting regime, namely the regime where the values of n, D, τ and σ satisfy

$$n - D \geq \left(\frac{\tau^2 D}{\sigma^2}\right),$$

we shall now show how an upper bound on σ can be obtained which is good enough for our purposes. Let S be the D dimensional subspace defined in the statement of Proposition 3.1. Let

$$\hat{\sigma} := \sqrt{\mathbb{E}_{z \sim \tilde{\mu}} \left(\frac{dist(z, S)^2}{n - D}\right)}.$$

It then follows from Proposition 3.1 that with probability at least $1 - \delta$,

$$(n - D)\sigma^2 \leq (n - D)\hat{\sigma}^2 \leq \alpha^4 \tau^2 + (n - D)\sigma^2.$$

This implies that

$$\sigma^2 \leq \hat{\sigma}^2 \leq (\alpha^4 + 1)\sigma^2.$$

The Monte-Carlo method can be used to estimate $\hat{\sigma}^2$ to within a prescribed error with high probability.

4 Learning discs that approximate the data

4.1 Properties of $\Pi_D X_0$.

In what follows, we shall, without loss of generality, identify S constructed in Proposition 3.1 with $\mathbb{R}^D \subseteq \mathbb{R}^n$. We will denote the orthogonal projection $\Pi_{\mathbb{R}^D}$ of $\mathbb{R}^n$ on to $\mathbb{R}^D$ by Π_D.

Let

$$r_p \in \left[\sqrt{\sigma\tau} D^{1/4}, \frac{\tau}{C d^C}\right]. \tag{60}$$

Here r_p is a preliminary radius that will be subsequently be replaced by a smaller radius r_c. Let N_0 be chosen to be an integer such that

$$N_0 / \ln(N_0) > \frac{CV}{\rho_{min} \omega_d (r_p^2/\tau)^d}, \tag{61}$$

where ω_d is the volume of a Euclidean unit ball in $\mathbb{R}^d$. We will assume that D is large enough to that we can choose N_0 such that

$$N_0 \leq e^D. \tag{62}$$

Lemma 4.1. *Let $\widetilde{X}_0$ be a set of N_0 i.i.d random samples from the distribution μ. Let X_0 be a set of N_0 i.i.d samples from $\mu * G_\sigma^{(n)}$, obtained by adding i.i.d noise sampled from $G_\sigma^{(n)}$ to the points in $\widetilde{X}_0$. With probability $1 - N_0^{-C}$, $\Pi_D X_0$ will be $Cr_p^2/\tau-$close to $\Pi_D \mathcal{M}$ in Hausdorff distance.*

Proof. By the coupon collector problem applied to the Voronoi cells corresponding to a $6r_p^2/\tau$ net of $\mathcal{M}$ that is also $r_p^2/2\tau$ separated (such a net always exists, and can be constructed by a greedy procedure), we see that if we examine the set $\Pi_D\widetilde{X_0}$ of N_0 i.i.d random samples from the push forward of μ under Π_D, with probability at least $1 - N_0^{-C}$, every Voronoi cell has at least one random sample. Therefore the Hausdorff distance of $\Pi_D\widetilde{X_0}$ to $\Pi_D\mathcal{M}$ is less than $12r_p^2/\tau$. Due to Gaussian concentration (see Lemma 2.1), the maximum distance of a point $\Pi_D y_i$ of $\Pi_D X_0$ to the corresponding $\Pi_D x_i$ in $\Pi_D\widetilde{X_0}$ is bounded above by

$$\sigma\left(\sqrt{D} + \sqrt{\ln(N_0^C)}\right) < Cr_p^2/\tau,$$

with probability at least $1 - N_0^{-C}$. This is an upper bound on the Hausdorff distance between $\Pi_D X_0$ and $\Pi_D\widetilde{X_0}$. Therefore, we have proved the lemma. $\square$

4.2 Putative discs

Let X be a finite set of points in $E = \mathbb{R}^D$ and $X \cap B_1(x) := \{x, \widetilde{x}_1, \dots, \widetilde{x}_s\}$ 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 (i.e. a ball in a $d-$ dimensional isometrically embedded Euclidean space) centered at x within a Hausdorff distance $Cd\delta$ of $\widetilde{X}_0 := X \cap B_1(x)$, 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 [49] but 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 $\widetilde{A}_x$ be the affine d-dimensional subspace containing $x, x_1, \dots, x_n$, and the unit d-disc $\widetilde{D}_1(x)$ be $\widetilde{A}_x \cap B_1(x)$. Recall that for two subsets A, B of $\mathbb{R}^D$, $d_H(A, B)$ represents the Hausdorff distance between the sets. We will denote large absolute constants by C and small absolute constants by c.

Lemma 4.2. *Suppose there exists a d-dimensional affine subspace A_x containing x such that $D_1(x) = A_x \cap B_1(x)$ satisfies $d_H(X_0, D_1(x)) \le \delta$. Suppose $0 < \delta < \frac{1}{2d}$. Then $d_H(X_0, \widetilde{D}_1(x)) \le Cd\delta$, 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 $m \le d - 1$,

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

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

$$\max\left(|1 - d(x, y)|, \left|\left\langle \frac{(x_1)}{|x_1|}, y\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, y\right\rangle\right|\right) \tag{63}$$

is 0, because the dimension of $D_1(x)$ is d and there are only $m \le d - 1$ 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 (63) 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

$$\max\left(|1 - d(x, y_{m+1})|, \left|\left\langle \frac{(x_1)}{|x_1|}, y_{m+1}\right\rangle\right|, \dots, \left|\left\langle \frac{(x_m)}{|x_m|}, y_{m+1}\right\rangle\right|\right) \le \delta. \tag{64}$$

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 (64), we see that for each $m + 1 \le n$,

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

Let $\widetilde{V}$ 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,

$$\|\widetilde{V}^t\widetilde{V} - I\| \le C_1 d\delta.$$

Therefore, the singular values of $\widetilde{V}$ 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 $i \le n$, 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 $i \le n$. Let $\widehat{V}$ be an $D \times d$ matrix whose i^{th} column is x'_i. Since for any distinct i, j we have $|\langle x'_i, x'_j\rangle| < 3\delta + \delta^2$, and for any i, $|\langle x'_i, x'_i\rangle - 1| < 4\delta$,

$$\|\widehat{V}^t\widehat{V} - I\| \le Cd\delta.$$

This means that the singular values of $\widehat{V}$ 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 $\widetilde{D}_1(x)$. Recall that the unit d-disc $\widetilde{D}_1(x)$ is $\widetilde{A}_x \cap B_1(x)$. 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 $\widetilde{D}_1(x)$ is less than $Cd\delta$.

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

Let $\alpha \in \mathbb{R}^d$ be such that $\widehat{V}\alpha = x'$. By the bound on the singular values of $\widehat{V}$, we have $|\alpha| < \exp(Cd\delta)$. Let $y' = \widetilde{V}\alpha$. Then, by the bound on the singular values of $\widetilde{V}$, 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 $\widetilde{D}_1(x)$. We next obtain an upper bound on $|x' - z'|$

$$
\begin{aligned}
|x' - z'| &\leq |x' - y'| \qquad (65)\\
&\quad + |y' - z'|. \qquad (66)
\end{aligned}
$$

We examine the term in (65)

$$
\begin{aligned}
|x' - y'| &= |\widehat{V}\alpha - \widetilde{V}\alpha| \\
&\leq \sup_i |x_i - x_i'| (\sum_{j=1}^{d} |\alpha_j|) \\
&\leq \delta d \exp(Cd\delta).
\end{aligned}
$$

We next bound the term in (66).

$$
\begin{aligned}
|y' - z'| &\leq |y'|(1 - \exp(-Cd\delta)) \\
&\leq Cd\delta.
\end{aligned}
$$

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

$$
\begin{aligned}
|p' - z''| &\leq |q' - z''| + |p' - q'| \\
&\leq |\widetilde{V}\beta - V\beta| + |1 - \widetilde{V}\beta| \\
&\leq \sup_i |x_i - x_i'| (\sum_{j=1}^{d} |\beta_j|) + C\delta d \\
&\leq \delta d \exp(Cd\delta) + C\delta d \leq C\delta d.
\end{aligned}
$$

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

4.3 Fine-tuning the discs

In this subsection, we will assume that the discs are centered at the origin and have a radius 1. We also assume that that we have constructed a disc $\widetilde{D}_1$ such that $d_H(\widetilde{D}_1, B_1 \cap X) < c$ and we know that there exists some disc D_1 (which we have not constructed) such that

1. $d_H(D_1, B_1 \cap X) = \delta_1 < c$ is the Hausdorff distance between a unit disc D_1 and $B_1 \cap X$.

2. $\delta_2 := \sup_{z \in B_1 \cap X} dist(z, D_1)$.

We will describe an algorithm that produces a disc $\widehat{D}$ such that

1. $\sup_{z \in B_1 \cap X} dist(z, \widehat{D}) \leq 2\delta_2$.

2. $d_H(\widehat{D}, D_1) \leq 12\delta_2$.

This algorithm will be applied to $X = \Pi_D X_0$, from Lemma 4.1. Note that in this context, $\delta_2 < \frac{Cr^2}{\tau}$, while we only need $\delta_1 < cr$. Consider the family $\bar{\mathfrak{D}}$ of all unit $d-$dimensional discs centered at the origin defined by

$$\bar{\mathfrak{D}} := \{D | (\sup_{z \in B_1 \cap X} dist(z, D) \leq 2\delta_2) \text{ and } (d_H(B_1 \cap X, D) \leq 2 d_H(\widetilde{D}_1, B_1 \cap X))\}.$$

We will obtain an estimate $\overline{\delta}$ of the diameter of this family in the Hausdorff metric. By the triangle inequality applied to the Hausdorff metric, there exist $D_2 \in \bar{\mathfrak{D}}$ and $x \in \partial D_1$ such that

$$dist(x, D_2) = d_H(D_1, D_2) \in [\overline{\delta}/2, \overline{\delta}].$$

Let $\phi : Tan(0, D_1) \to Nor(0, D_1)$ be a linear map from the tangent space of D_1 at the origin to the normal space at D_1 at the origin such that $D_2 \subseteq \{(z, \phi(z)) | z \in D_1\}$. Then, ϕ is $\overline{\delta}/\sqrt{1 - \overline{\delta}^2}-$Lipschitz (measured with respect to the Euclidean metric on the range and domain of ϕ).

Let $y \in X \cap B_1$ be such that $|y - x| \leq 2 d_H(\widetilde{D}_1, B_1 \cap X)$. Since $2 d_H(\widetilde{D}_1, B_1 \cap X) < c$, we have

$$|x - \Pi_{D_1} y| < 2 d_H(\widetilde{D}_1, B_1 \cap X) < c,$$

implying that

$$dist(\Pi_{D_1} y, D_2) \geq \overline{\delta}/4.$$

Also $|y - \Pi_{D_1} y| \leq \delta_2$. Therefore,

$$2\delta_2 \geq dist(y, D_2) \geq \overline{\delta}/4 - \delta_2.$$

Therefore, $\overline{\delta}/4 \leq 3\delta_2$. Thus, $\overline{\delta} \leq 12\delta_2 < c$.

Therefore, by the preceding bound on the Hausdorff diameter and $D_1 \in \bar{\mathfrak{D}}$, it suffices to find any one element of $\bar{\mathfrak{D}}$ to obtain an additive approximation $\widehat{D}$ to D_1 to within $12\delta_2$ in the Hausdorff metric which also satisfies $\sup_{z \in B_1 \cap X} dist(z, \widehat{D}) \leq 2\delta_2.$. This is done as follows. Let the linear span of $\widetilde{D}_1$ be identified with $\mathbb{R}^d$ with the canonical Euclidean metric. Then, we consider the set $\widetilde{\bar{\mathfrak{D}}}$ of discs of the form

$$\widetilde{D}_A = B_1 \cap \{(z_1', z_2') | A z_1' - I z_2' = 0\}, \tag{67}$$

where A is a $d \times (D-d)$ matrix, such that the operator norm of A is bounded above as follows:

$$\|A\| \leq 2d_H(\widetilde{D}_1, B_1 \cap X).$$

By (4.3), $\widetilde{\widetilde{\mathcal{D}}} \supseteq \bar{\mathcal{D}}$. It suffices to find one element of $\bar{\mathcal{D}}$ in $\widetilde{\widetilde{\mathcal{D}}}$. We do this by solving the following convex program using Vaidya's algorthm [93]:
Find A such that

1. $\|A\|_2 \leq 2d_H(\widetilde{D}_1, B_1 \cap X)$, and
2. for all $x = (x_1, x_2) \in B_1 \cap X$, where $x_1 \in \mathbb{R}^d$ and $x_2 \in \mathbb{R}^{D-d}$,

$$\|Ax_1 - Ix_2\|_2 \leq 2\delta_2.$$

Once such an A is found, the corresponding $\widehat{D}$ given by (67) satisfies the following. For any $(x_1, x_2) = x \in B_1 \cap X$,

$$dist(x, \widehat{D}) \;=\; \left\| (I + AA^T)^{-1/2}(Ax_1 - Ix_2) \right\|_2 \leq 2\delta_2. \tag{68}$$

5 Obtaining a refined net of $\mathcal{M}$

We recall the setup. $\mathcal{M}$ is a submanifold of $\mathbb{R}^n$. $\mathbb{R}^D \subseteq \mathbb{R}^n$ is a coordinate subspace such that the orthogonal projection Π_D of $\mathbb{R}^n$ onto $\mathbb{R}^D$ satisfies

$$\sup_{x \in \mathcal{M}} |x - \Pi_D x| \leq \alpha^2 \tau,$$

where we choose α^2 to be less than $\frac{1}{Cd}$. Let the set of all points within a distance of $2\sqrt{D}\sigma$ of $\Pi_D\mathcal{M}$, be denoted $(\Pi_D\mathcal{M})_{2\sqrt{D}\sigma}$. Let $x \in (\Pi_D\mathcal{M})_{2\sqrt{D}\sigma}$. Recall that X_0 is a set of N_0 independent, randomly sampled points from $\mu * G_\sigma^{(n)}$. Let r_c satisfy

$$\frac{r_c^2}{\tau} = \frac{\tau}{Cd^C} > 4\sigma D^{\frac{1}{2}}. \tag{69}$$

Let $B_D(x, r_c)$ denote the Euclidean $D-$dimensional ball of radius r_c, centered at x, contained in $\mathbb{R}^D$. Let x be an arbitrary point in $(\Pi_D\mathcal{M})_{2\sqrt{D}\sigma}$. Let D_0 be a $d-$dimensional disc centered at x, having radius r_c and having a Hausdorff distance to $\Pi_D(X_0) \cap B_D(x, r_c)$ that is less than $\delta = \frac{Cr_c^2}{\tau}$, that is

$$\mathbf{d}_{\mathtt{haus}}(D_0, \Pi_D(X_0) \cap B_D(x, r_c)) < \delta = \frac{Cr_c^2}{\tau}. \tag{70}$$

Disc D_0 can be obtained using the algorithm in the Section 4.

Lemma 5.1.

$$\mathbb{P}\left[\mathbf{d}_{\mathtt{haus}}(D_0, B_D(x, r_c) \cap \Pi_D(\mathcal{M})) \leq 2\delta\right] \geq 1 - cN_0^{-C}.$$

Proof. Since $\sigma(\sqrt{D} + \sqrt{\ln(N_0^C)})$ is less than $\delta = Cr_c^2/\tau < \frac{\tau}{Cd^C}$, by Gaussian concentration (see Lemma 2.1), with probability at least $1 - \frac{N_0^{-C}}{C}$, every point $y_i = x_i + \zeta_i$ satisfies

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

By replacing r_p by r_c in Lemma 4.1, we see that

$$\mathbb{P}\left[\mathbf{d}_{\mathtt{haus}}(\Pi_D X_0, \Pi_D(\mathcal{M})) \le \delta/2\right] \ge 1 - N_0^{-C}.$$

Therefore, in particular,

$$\mathbb{P}\left[\mathbf{d}_{\mathtt{haus}}(\Pi_D X_0 \cap B_D(x, r_c), \Pi_D(\mathcal{M})) \le \delta\right] \ge 1 - N_0^{-C}.$$

By the triangle inequality with respect to $\mathbf{d}_{\mathtt{haus}}$,

$$\begin{aligned}\mathbf{d}_{\mathtt{haus}}(D_0, B_D(x, r_c) \cap \Pi_D(\mathcal{M})) &\le \mathbf{d}_{\mathtt{haus}}(D_0, \Pi_D(X_0) \cap B_D(x, r_c)) \\ &+ \mathbf{d}_{\mathtt{haus}}(\Pi_D(X_0) \cap B_D(x, r_c), B_D(x, r_c) \cap \Pi_D(\mathcal{M})).\end{aligned}$$

The Lemma now follows from the last two inequalities and (70). □

We now translate the origin to x and denote by $\mathbb{R}^d$, the affine span of D_0. Let $\Pi_d : \mathbb{R}^n \to \mathbb{R}^d$ denote the map that projects a point in $\mathbb{R}^n$ orthogonally on to $\mathbb{R}^d$. Let S_0 denote the cylindrical set given by

$$S_0 := B_d(0, \frac{r_c}{2}) \times B_{D-d}(0, \frac{r_c}{2}) \subset \mathbb{R}^D, \tag{71}$$

where $B_d(0, \frac{r_c}{2})$ is the ball of radius $\frac{r_c}{2}$ contained in $\mathbb{R}^d$ centered at the origin, and $B_{D-d}(0, \frac{r_c}{2})$ is the ball of radius $\frac{r_c}{2}$ contained in $\mathbb{R}^{D-d}$ centered at the origin. Here $\mathbb{R}^{D-d}$ is the orthogonal compliment of $\mathbb{R}^d$ inside $\mathbb{R}^D$. Note that

$$\Pi_D^{-1} S_0 \subseteq \mathbb{R}^n.$$

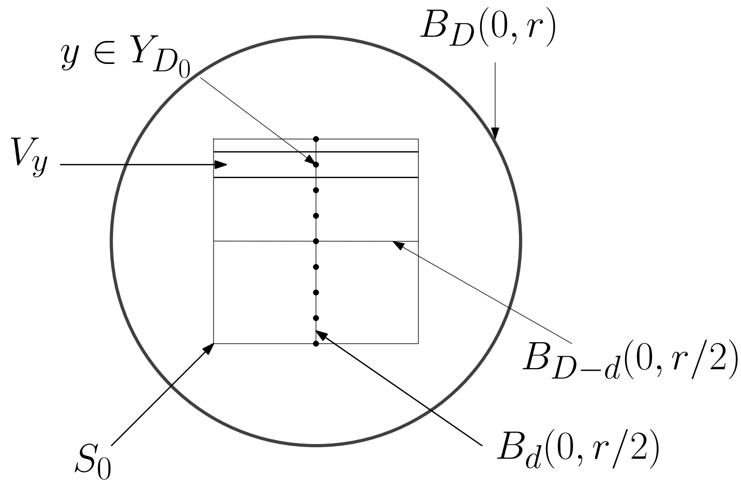
We will obtain a refined net of $\mathcal{M} \cap \Pi_D^{-1} S_0$. Let $Y_{D_0} = B_d(0, \frac{r_c}{2}) \cap 10\sigma\mathbb{Z}^d$. Due to volumetric considerations,

$$|Y_{D_i}| \le \left(\frac{r_c}{\sigma}\right)^d. \tag{72}$$

For $y \in Y_{D_0}$, let V_y denote the "truncated Voronoi cell" defined by

$$V_y = \{y' \in S_0 | \forall y'' \in Y_{D_0}, |y - y'| \le |y' - y''|.\} \tag{73}$$

Thus, V_y is the set of all points in S_0 that are at least as close to y as they are to any other member of Y_{D_0}.



The points of X_0 have the form $q_i = p_i + \gamma_i$, where p_i is a sample from μ supported on $\mathcal{M}$ and the γ_i are i.i.d samples from the measure whose density is $G_\sigma^{(n)}$. Define the net

$$Z_{S_0} := \{z_y | y \in Y_{D_0}\},$$

where z_y is defined to be the expectation of a random point q_i conditional on it belonging to $\Pi_D^{-1} V_y$. Let us first focus on $\Pi_D z_i$, and its distance to $\Pi_D \mathcal{M}$. The push-forward of $G_\sigma^{(n)}$ under Π_D is a Gaussian density $G_\sigma^{(D)}$ restricted to $\mathbb{R}^D$, the measure corresponding to which is $N_D(0, \sigma^2)$. Further $\Pi_D \mathcal{M}$ has already been shown to be of reach at least $(1 - 4\alpha^2)\tau$ in Lemma 3.3. Secondly, by (50), Π_D applied to a unit tangent vector v at a point on $\mathcal{M}$ satisfies

$$|\|\Pi_D v\| - 1| \leq \|(\Pi_D v) - v\| \leq 1 - \sqrt{1 - 4\alpha^2} \leq 4\alpha^2 \leq \frac{1}{4d}.$$

Due to the consequences of this on the Jacobian of the restriction of Π_D on the tangent space at a point on $\mathcal{M}$, the push forward of μ under Π_D has a Radon-Nikodym derivative with respect to the Hausdorff measure on $\Pi_D \mathcal{M}$ that takes values in $[\frac{\rho_{min}}{2}, 2\rho_{max}]$.

In this section, we will use the fact that that the logarithm of the Radon-Nikodym derivative of μ with respect to the Hausdorff measure on $\mathcal{M}$ is $\frac{1}{\tau}-$Lipschitz.

As a consequence of Lemma A.2, $\Pi_D(\mathcal{M}) \cap \frac{\tau S_0}{10 r_c}$ is the graph of a C^2-function f from $B_d(0, \frac{\tau}{20})$ to the orthogonal compliment of $\mathbb{R}^d$ in $\mathbb{R}^D$, (which we henceforth denote by $\mathbb{R}^{D-d}$). This function f, when restricted to $B_d(0, \frac{r_c}{2})$ has a C^0-norm of at most $C\delta$ with probability $1 - N_0^{-C}$ by Lemma 5.1. Thus, we henceforth assume

$$\forall x \in B_d(0, r), |f(x)| \leq C\delta. \tag{74}$$

The bound on the reach together with Lemma A.2 implies the following for all $x \in B_d(0, \frac{\tau}{20})$. Firstly,

$$\forall_{v \in \mathbb{R}^d} \forall_{w \in \mathbb{R}^{n-d}} \langle \partial_v^2 f(x), w \rangle \leq \frac{C|v|^2|w|}{\tau}.$$

Secondly,

$$\forall v \in \mathbb{R}^d, |\partial_v f(x)| \le \frac{C\delta|v|}{r_c} + \frac{C|v||x|}{\tau}. \tag{75}$$

This implies that $\forall x, y \in B_d(0, \frac{\tau}{20})$, and $\forall w \in B_{n-d}(0,1)$, denoting $x - y$ by v,

$$|\langle f(x) - f(y) - \partial_v f(y), w\rangle| \le \frac{C|v|^2}{\tau}. \tag{76}$$

Let $X_0^{(2)}$ be a set of

$$\begin{aligned} N_2 &:= C^d n \left(\frac{\sigma}{\sigma^2/\tau}\right)^2 N_0 \max_i |Y_{D_i}| \log N_0 \log \frac{N_0 r_c^d}{\sigma^d} \\ &\le C^d n \left(\frac{\sigma}{\sigma^2/\tau}\right)^2 N_0 \left(\frac{r_c}{\sigma}\right)^d \log N_0 \log \frac{N_0 r_c^d}{\sigma^d} \end{aligned}$$

independent random samples from μ (which are hence independent to X_0). Let

$$\#_y := |V_y \cap \Pi_D X_0^{(2)}|.$$

We will be interested in the case where

$$\#_y > 100\left(\frac{n\tau^2}{\sigma^2}\right)\log N_0,$$

since the complimentary event will be absorbed in the error probability.

Lemma 5.2.

$$\mathbb{P}\left[\left(\min_{D_i}\min_{y\in Y_{D_i}} \#_y\right) \le 100\left(\frac{n\tau^2}{\sigma^2}\right)\log N_0\right] < N_0^{-100}. \tag{77}$$

Proof. By (75), $\mu(\Pi_D^{-1}(V_y)\cap\mathcal{M}) > \frac{\rho_{min} V}{N_0 \max_i |Y_{D_i}|} > \frac{c\sigma^d}{r_c^d N_0}$. If $x \in \Pi_D^{-1}(V_y)\cap\mathcal{M}$, and ζ is a random sample from the measure associated with $G_\sigma^{(n)}$, then

$$\mathbb{P}[x + \zeta \in V_y] \ge c^d,$$

since $V_y \cap \mathbb{R}^d$ is a cube of side length more than $c\sigma$. Therefore, if x is a random sample from μ and ζ is a random sample from the measure associated with $G_\sigma^{(n)}$ and $z = x + \zeta$, then $\mathbb{P}[z \in \Pi_D^{-1} V_y] > \frac{c^d \sigma^d}{r_c^d N_0}$. The lemma follows from the solution to the coupon collector problem with $N_0 \max_i |Y_{D_i}| < \frac{N_0 r_c^d}{\sigma^d}$ bins, corresponding to the different V_y. □

Recall from (71) that $S_0 = B_d(0, \frac{r_c}{2}) \times B_{D-d}(0, \frac{r_c}{2})$. We split $V_y \cap \Pi_D X_0^{(2)}$ into two sets. Let S_1 consist of those points $\Pi_D(p_i + \gamma_i)$ such that $\Pi_D(p_i + \gamma_i) \in V_y \cap \Pi_D X_0$ and $\Pi_D(p_i) \in \frac{\tau S_0}{10 r_c}$, where we recall that $p_i \in \mathcal{M}$. If $\#_y \geq 1$, let

$$\widehat{z}_{y,1} := \frac{1}{\#_y} \sum_{z \in S_1} z.$$

If $\#_y = 0$, we define $\widehat{z}_{y,1} = 0$. Let

$$z_{y,1} := \mathbb{E}\left[\widehat{z}_{y,1}\right].$$

Let S_2 consist of those points $\Pi_D(p_i+\gamma_i)$ such that $\Pi_D(p_i+\gamma_i) \in V_y$ and $\Pi_D(p_i) \notin \frac{\tau S_0}{10 r_c}$. If $\#_y \geq 1$, let

$$\widehat{z}_{y,2} := \frac{1}{\#_y} \sum_{z \in S_2} z.$$

If $\#_y = 0$, we define $\widehat{z}_{y,2} = 0$. Let

$$z_{y,2} := \mathbb{E}\left[\widehat{z}_{y,2}\right].$$

Let

$$\widehat{z}_y := \widehat{z}_{y,1} + \widehat{z}_{y,2}.$$

Let μ' be the pushforward under Π_d of the restriction of μ to $\Pi_D^{-1}\left(\frac{\tau S_0}{10 r_c}\right)$, normalized to be a probability measure. Thus μ' is the push forward of a measure (derived from μ) supported in $\mathcal{M}$. We will also need to work with an analogous measure derived from $\mu * G_\sigma^{(n)}$, which we denote by μ'' below.

Let $\mathbb{P}[x \in A] = \mu'(A)$ for every Borel set $A \subseteq B_d(0, \frac{\tau}{20})$. Let Π_{D-d} denote the orthogonal projection of $\mathbb{R}^n$ onto $\mathbb{R}^{D-d}$, which we define to be the orthocomplement of $\mathbb{R}^d$ in $\mathbb{R}^D$. Let μ'' be the probability measure supported on $B_d(0, \frac{\tau}{20})$ given by

$$\mu''(A) \quad := \quad \mathbb{P}[x \in A | (x' + x + f(x) \in V_y)],$$

where $x + f(x)$ is a sample from the pushforward of μ via Π_D and x' is the image via Π_D of an independent sample from $G_\sigma^{(n)}$. Thus

$$\begin{aligned} \mu''(A) &= \frac{\mathbb{P}[(x \in A) \wedge (x' + x + f(x) \in V_y)]}{\mathbb{P}[(x' + x + f(x) \in V_y)]} \\ &= \frac{\mathbb{P}[(x \in A) \wedge (\Pi_d x' + x \in \Pi_d V_y) \wedge (\Pi_{D-d} x' + f(x) \in \Pi_{D-d} V_y)]}{\mathbb{P}[(\Pi_d x' + x \in \Pi_d V_y) \wedge (\Pi_{D-d} x' + f(x) \in \Pi_{D-d} V_y)]}. \end{aligned}$$

Let γ_d and γ_{D-d} denote Gaussian measures in $\mathbb{R}^d$ and $\mathbb{R}^{D-d}$ having covariances $\sigma^2 I_d$ and $\sigma^2 I_{D-d}$ respectively. Note that the denominator in the above expression

$$\mathbb{P}[(\Pi_d x' + x \in \Pi_d V_y) \wedge (\Pi_{D-d} x' + f(x) \in \Pi_{D-d} V_y)]$$

equals

$$\int_{\mathbb{R}^d} \gamma_d(\Pi_d V_y - x)\gamma_{D-d}(\Pi_{D-d} V_y - f(x))\mu'(dx) =: \Gamma. \tag{78}$$

Let $B_d^\infty(x, R)$ denote ℓ_∞ ball in $\mathbb{R}^d$ whose center is x and side length is $2R$. Then, the Radon-Nikodym derivative $\frac{d\mu''}{d\mu'}$ at $x \in B_d(0, \frac{\tau}{20})$ is given by

$$\Gamma^{-1}\gamma_d(\Pi_d V_y - x)\gamma_{D-d}(\Pi_{D-d} V_y - f(x)). \tag{79}$$

Moreover,

$$\gamma_d(\Pi_d V_y - x) = \int_{B_d^\infty(y, 5\sqrt{d}\sigma)} (\sqrt{2\pi}\sigma)^{-d} \exp(-\frac{(|x-z|)^2}{2\sigma^2})\lambda_d(dz). \tag{80}$$

Assuming the $D - d$ is larger than a sufficiently large universal constant C, we see by Gaussian concentration (Lemma 2.1), (74) and the fact that $\Pi_{D-d} V_y \subseteq B_{D-d}(0, \frac{r}{2})$ that if $x \in B_d(y, \frac{r_c}{2})$, then $f(x) < C\delta < \frac{r_c}{6}$, implying that

$$\gamma_{D-d}(\Pi_{D-d} V_y - f(x)) > 1 - \frac{\sigma^2}{\tau}. \tag{81}$$

It follows that $\Gamma^{-1} \leq C\lambda_d(B_d^\infty(y, 5\sqrt{d}\sigma))$. We observe that due to the symmetries of the lattice $10\sigma\mathbb{Z}^d$, provided that $B_d^\infty(y, 5\sqrt{d}\sigma) \subseteq S_0$,

$$\gamma_d(\Pi_d V_y - x) = \gamma_d(\Pi_d V_y - (2y - x)). \tag{82}$$

Let us first analyze $|z_{y,2}|$. Since V_y is contained in S_0, this is clearly in the interval $[-\frac{r_c}{2}, \frac{r_c}{2}]$. We see that

$$\left|z_{y,2}\right| \leq \frac{r_c}{2}\left((\sqrt{2\pi}\sigma)^{-D} \exp(-\frac{c\tau^2}{2\sigma^2})\rho_{max}\lambda_D(V_y)\right) \leq \left(\frac{\sigma^2}{\tau}\right),$$

by Gaussian concentration (Lemma 2.1), since $\frac{\tau}{Cd^C} > \sigma\sqrt{D}$.

We will now fix one $y \in Y_{D_0} = B_d(0, \frac{r_c}{2}) \cap 10\sigma\mathbb{Z}^d$*, and translate the origin to this point.* Next, let us analyze

$$e_y := \Pi_{D-d}(z_{y,1} + z_{y,2}).$$

5.1 Controlling the distance of e_y to $\Pi_D\mathcal{M}$

Let $\widehat{G}^x_{D-d}$ be a random variable, whose distribution is that of a a mean 0 Gaussian of covariance $\sigma^2 I_{D-d}$, conditioned to be in $\Pi_{D-d} V_y - f(x)$. Then,

$$\Pi_{D-d}(z_{y,1}) = \int_{x \in B_d(0, \frac{\tau}{20})} \mathbb{E}(f(x) + \widehat{G}^x_{D-d})\mu''(dx),$$

and so

$$e_y = \Pi_{D-d} z_{y,2} + \int_{x \in B_d(0, \frac{\tau}{20})} \mathbb{E}(f(x) + \widehat{G}^x_{D-d}) \mu''(dx). \tag{83}$$

By (74), for $x \in B_d(0,r)$, we have $|f(x)| \leq C\delta$. Recall that

$$\frac{r_c^2}{\tau} = \frac{\delta}{C} > 4\sigma\sqrt{D}.$$

Let $g : B_d(0, \frac{\tau}{20}) \to \mathbb{R}^{D-d}$ be given by

$$g(x) := \mathbb{E}(\widehat{G}^x_{D-d}).$$

Lemma 5.3. *If $x \in B_d(0, r_c)$, then $|g(x)| < \frac{\sigma^2}{\tau}$.*

Proof. Let all lengths be rescaled so that $\sqrt{2\pi}\sigma = 1$. Let $\widehat{x} = \frac{f(x)}{|f(x)|}$, and denote $|f(x)|$ by T. By (74), for $x \in B_d(0, r_c)$, $T \leq C\delta$. Then,

$$g(x) = \frac{\left(\int\limits_{B_{D-d}(\frac{r_c}{2})} \langle z - T\widehat{x}, \widehat{x} \rangle \exp\left(-\pi \|z - T\widehat{x}\|^2\right) \lambda_{D-d}(dz) \right) \widehat{x}}{\int\limits_{B_{D-d}(\frac{r_c}{2})} \exp(-\pi\|z - T\widehat{x}\|^2) \lambda_{D-d}(dz)}.$$

Fot $t \in \mathbb{R}$, let

$$h(t) := \int\limits_{B_{D-d}(\frac{r_c}{2})} \langle z - t\widehat{x}, \widehat{x} \rangle \exp\left(-\pi\|z - t\widehat{x}\|^2\right) \lambda_{D-d}(dz).$$

Then,

$$\begin{aligned} \partial_t h(t) &= 2\pi \int\limits_{B_{D-d}(\frac{r_c}{2})} \langle z - t\widehat{x}, \widehat{x} \rangle \langle z - t\widehat{x}, \widehat{x} \rangle \exp\left(-\pi\|z - t\widehat{x}\|^2\right) \lambda_{D-d}(dz) \\ &\quad - \int\limits_{B_{D-d}(\frac{r_c}{2})} \exp\left(-\pi\|z - t\widehat{x}\|^2\right) \lambda_{D-d}(dz). \end{aligned}$$

Let

$$J_r := 2\pi \int\limits_{\mathbb{R}^{D-d} \setminus B_{D-d}(r_c)} \langle z - t\widehat{x}, \widehat{x} \rangle \langle z - t\widehat{x}, \widehat{x} \rangle \exp\left(-\pi\|z - t\widehat{x}\|^2\right) \lambda_{D-d}(dz)$$

Let $r' = \frac{r_c}{3} > \left(\frac{2}{3}\right) \sqrt{\sigma\tau D^{\frac{1}{2}}}$. Note that

$$B_{D-d}(0, \frac{r_c}{3}) \subseteq B_{D-d}(T\widehat{x}, \frac{r_c}{2}).$$

Then,

$$J_{r'} \exp(\pi r'^2/2) \quad \leq \quad 2\pi \int_{\mathbb{R}^{D-d}} \frac{|x|^2}{D-d} \exp(-\pi|x|^2) \exp(\pi|x|^2/2) \lambda_{D-d}(dx) = 2^{\frac{D-d}{2}}.$$

Therefore

$$J_{r'} \leq 2^{\frac{D-d}{2}} \exp(-\frac{\pi r'^2}{2}).$$

Let

$$I_{r'} := \int_{|x|>r'} \exp(-\pi|x|^2) \lambda_{D-d}(dx). \tag{84}$$

The left hand side $I_{r'}$ can be bounded above as follows.

$$\begin{aligned} I_{r'} \exp(\pi r'^2/2) \quad &\leq \quad \int_{\mathbb{R}^{D-d}} \exp(-\pi|x|^2) \exp(\pi|x|^2/2) \lambda_{D-d}(dx) \\ &= \quad 2^{\frac{D-d}{2}}. \end{aligned}$$

Therefore

$$I_{r'} \leq 2^{\frac{D-d}{2}} \exp(-\pi r'^2/2) \leq 2^{\frac{D-d}{2}} \exp(-\frac{\pi r'^2}{2}).$$

We see that $J_{r'}$ and $I_{r'}$ are trivially non-negative. It follows that for $t \in [0, T]$,

$$|\partial_t h(t)| \leq 2^{\frac{D-d}{2}} \exp(-\frac{\pi r'^2}{2}). \tag{85}$$

Since $h(0) = 0$, we see that $|h(T)| \leq 2^{\frac{D-d}{2}} \exp(-\frac{\pi r'^2}{2})T$. Therefore, $\forall x \in B_d(0, r)$,

$$|g(x)| \quad \leq \quad \frac{2^{\frac{D-d}{2}} \exp(-\frac{\pi r'^2}{2})T}{1 - 2^{\frac{D-d}{2}} \exp(-\frac{\pi r'^2}{2})} \tag{86}$$

$$\leq \quad 2^{2+\frac{D-d}{2}} \exp(-\frac{\pi r'^2}{4\pi\sigma^2})T \leq \frac{\sigma^2}{\tau}. \tag{87}$$

□

The following lemma shows that $dist(e_y, \Pi_D \mathcal{M}) < \frac{Cd\sigma^2}{\tau}$.

Lemma 5.4.

$$\left|\Pi_{D-d} e_y - f(y)\right| < \frac{Cd\sigma^2}{\tau}.$$

Proof.

$$\begin{aligned}\left|\Pi_{D-d}(e_y - z_{y,2}) - f(y)\right| &= \left|\int_{x\in B_d(0,\frac{\tau}{20})} \mathbb{E}(f(x) - f(y) + \widehat{G}^x_{D-d})\mu''(dx)\right| \\ &\leq \left|\int_{x\in B_d(y,\frac{r_c}{2})} \mathbb{E}(f(x) - f(y) + \widehat{G}^x_{D-d})\mu''(dx)\right| \quad (88)\\ &+ \left|\int_{x\in B_d(0,\frac{\tau}{20})\setminus B_d(y,\frac{\tau}{2})} \mathbb{E}(f(x) - f(y) + \widehat{G}^x_{D-d})\mu''(dx)\right|. \quad (89)\end{aligned}$$

Observe that the signed version of (88) can be rewritten as follows.

$$\int_{x\in B_d(y,\frac{r_c}{2})} \mathbb{E}(f(x) - f(y) + \widehat{G}^x_{D-d})\mu''(dx) \quad (90)$$

$$= \left(\frac{1}{2}\right)\left(\int_{x\in B_d(y,\frac{r_c}{2})} \mathbb{E}(f(x) - f(y) - \partial_{x-y}f(y) + \widehat{G}^x_{D-d})\mu''(dx)\right) \quad (91)$$

$$+ \left(\frac{1}{2}\right)\left(\int_{x\in B_d(y,\frac{r_c}{2})} \mathbb{E}(f(2y-x) - f(y) - \partial_{y-x}f(y) + \widehat{G}^{2y-x}_{D-d})\mu''(d(2y-x))\right) \quad (92)$$

$$+ \left(\frac{1}{2}\right)\left(\int_{x\in B_d(y,\frac{r}{2})} \partial_{x-y}f(y)(\mu''(dx) - \mu''(d(2y-x))\right). \quad (93)$$

We observe that twice the magnitude of (91) satisfies,

$$\left|\int_{x\in B_d(y,\frac{r_c}{2})} \mathbb{E}(f(x) - f(y) - \partial_{x-y}f(y) + \widehat{G}^x_{D-d})\mu''(dx)\right| \leq$$

$$\left|\int_{x\in B_d(y,\frac{r_c}{2})} (f(x) - f(y) - \partial_{x-y}f(y))\mu''(dx)\right| + \frac{\sigma^2}{\tau} \leq \quad (94)$$

$$\left|\int_{x\in B_d(y,\frac{r_c}{2})} \left(\frac{C|x-y|^2}{\tau}\right)\Gamma^{-1}\gamma_d(\Pi_d V_y - x)\gamma_{D-d}(\Pi_{D-d}V_y - f(x))\mu'(dx)\right| + \frac{\sigma^2}{\tau} \leq \frac{Cd\sigma^2}{\tau}.$$

By symmetry, The same bound applies to twice the next term. Now we need to bound (93). To do so, observe that

$$\begin{aligned}&\left|\int_{x\in B_d(y,\frac{r_c}{2})} \partial_{x-y}f(y)(\mu''(dx) - \mu''(d(2y-x))\right| \\ \leq\ & \int_{x\in B_d(y,\frac{r_c}{2})} |\partial_{x-y}f(y)|\cdot|(\mu''(dx) - \mu''(d(2y-x))| \\ \leq\ & \int_{x\in B_d(y,\frac{r_c}{2})} \left(\frac{C\delta|x-y|}{r_c} + \frac{C|x-y|^2}{\tau}\right)\Gamma^{-1}\gamma_d(\Pi_d V_y - x)\left(\frac{|x-y|}{\tau}\right)\mu'(dx) \\ \leq\ & \left(\frac{Cd\sigma^2 r_c}{\tau^2} + \frac{C\sigma^3 d^{\frac{3}{2}}}{\tau^2}\right) < \frac{C\sigma^2}{\tau}.\end{aligned}$$

In the last step, we used the fact that $r_c < \frac{\tau}{Cd}$. Lastly, from the tail decay of $\frac{d\mu''}{d\mu'}$ as expressed by (79) and (80) and the fact that $\frac{r_c}{2} \geq \left|\mathbb{E}(f(x) - f(y) + \widehat{G}^x_{D-d})\right|$, it follows that the term (89) is bounded above by $\frac{\sigma^2}{\tau}$. □

We define the "refined net" of $\Pi_D\mathcal{M} \cap B_D(0, \frac{r_c}{2})$ to be $\{e_y | y \in Y_{D_0}\}$. To extend this to a net of $\Pi_D\mathcal{M}$, we take the union over all such refined nets corresponding to balls of radius r_c with centers in X_1.

5.1.1 The ψ_2 norm

A random variable Z in $\mathbb{R}$ that satisfies for some positive real K,

$$\mathbb{E}[\exp(|Z|^2/K^2)] \leq 2$$

is called subgaussian.

Definition 5.1 (ψ_2 norm)**.** *We define*

$$\|Z\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}[\exp(|Z|^2/t^2)] \leq 2\}.$$

That this corresponds to a norm is known, see for example, Exercise 2.5.7 of [94]. For a subgaussian random variable, by Proposition 2.5.2 of [94],

$$\mathbb{P}[Z \geq t] \leq 2\exp\left(\frac{-ct^2}{\|Z\|^2_{\psi_2}}\right), \tag{95}$$

for all $t \geq 0$.

We appeal to Theorem 3.1.1 from [94], which implies the following.

Proposition 5.1. *Let $Z = (Z_1, \dots, Z_n) \in \mathbb{R}^n$ be a random vector with independent Gaussian coordinates Z_i that satisfy $\mathbb{E}Z_i^2 = \sigma^2$. Then*

$$\left\| |Z| - \sqrt{n}\sigma \right\|_{\psi_2} \leq C\sigma,$$

where C is an absolute constant.

5.2 Controlling the distance of a point in the refined net to $\mathcal{M}$

We now proceed to obtain a refined net of $\mathcal{M}$. We split $(V_y \oplus \mathbb{R}^{n-D}) \cap X_0^{(2)}$ into two sets.

Let S_1^n consist of those points $p_i + \gamma_i$ such that $p_i + \gamma_i \in (V_y \oplus \mathbb{R}^{n-D}) \subset \mathbb{R}^n$ and $\Pi_D p_i \in \frac{\tau S_0}{10r}$. Let $\widehat{z}^n_{y,1}$ be $\frac{1}{\#_y}\sum\limits_{z \in S_1^n} z$ if $\#_y \geq 1$ and be 0 if $\#_y = 0$ and $z^n_{y,1} := \mathbb{E}\left[\widehat{z}^n_{y,1}\right]$. Let S_2^n consist of those points $p_i + \gamma_i$ such that $\Pi_D(p_i + \gamma_i) \in V_y$ and $\Pi_D(p_i) \notin \frac{\tau S_0}{10r_c}$. Let $\widehat{z}^n_{y,2}$ be $\frac{1}{\#_y}\sum\limits_{z \in S_2^n} z$ if $\#_y \geq 1$ and be 0 if $\#_y = 0$. Let $z^n_{y,2} := \mathbb{E}\left[\widehat{z}^n_{y,2}\right]$. Let

$$\widehat{z}^n_y := \widehat{z}^n_{y,1} + \widehat{z}^n_{y,2}.$$

Let us first analyze $|z^n_{y,2}|$. We see that there is a distribution γ''_τ supported on $[c\tau, \infty)$, such that

$$\left|z^n_{y,2}\right| \;\; \leq \;\; \int \tau \left((\sqrt{2\pi}\sigma)^{-D} \exp(-\frac{cx^2}{2\sigma^2}) \lambda_D(S_0) \right) \gamma''_\tau(dx) \leq \left(\frac{\sigma^2}{\tau}\right). \tag{96}$$

Let us next analyze

$$e^n_y := e_y + \Pi_{n-D}(z^n_{y,1} + z^n_{y,2}).$$

Definition 5.2. *Let $\widehat{G}^x_{n-D}$ denote a mean 0 Gaussian of covariance $\sigma^2 I_{n-D}$. Then,*

$$\Pi_{n-D}(e^n_y - z^n_{y,2}) = \int_{x \in B_d(0, \frac{\tau}{20})} \mathbb{E}(f(x) + \widehat{G}^x_{n-D})\mu''(dx) = \int_{x \in B_d(0, \frac{\tau}{20})} f^{n-D}(x)\mu''(dx).$$

As a consequence of Proposition 3.1, $\mathcal{M} \cap \Pi_D^{-1}\left(\frac{\tau S_0}{10 r_c}\right)$ is the graph of a C^2-function f^{n-D} from $B_d(0, \frac{\tau}{20})$ to the orthogonal compliment of $\mathbb{R}^D$ in $\mathbb{R}^n$, (which we henceforth denote by $\mathbb{R}^{n-D}$). In order to show that

$$dist(e^n_y, \mathcal{M}) < \frac{Cd\sigma^2}{\tau}, \tag{97}$$

it suffices to prove the following lemma. Note that the difference between Lemma 5.4 and the following Lemma 5.5 is that in the latter we have the projection Π_{n-D} from $\mathbb{R}^n$ to $\mathbb{R}^D$, while the former involves the projection Π_{D-d} from $\mathbb{R}^D$ to $\mathbb{R}^d$.

Lemma 5.5.

$$\left|\Pi_{n-D} e^n_y - f^{n-D}(y)\right| < \frac{Cd\sigma^2}{\tau}.$$

Proof of Lemma 5.5. We see that

$$\begin{aligned} \left|\Pi_{n-D}(e^n_y - z^n_{y,2}) - f^{n-D}(y)\right| \;\; &= \;\; \left|\int_{x \in B_d(0, \frac{\tau}{20})} \mathbb{E}(f^{n-D}(x) - f^{n-D}(y))\mu''(dx)\right| \\ &\leq \left|\int_{x \in B_d(y, \frac{r_c}{2})} (f^{n-D}(x) - f^{n-D}(y))\mu''(dx)\right| \qquad (98) \\ &+ \left|\int_{x \in B_d(0, \frac{\tau}{20}) \setminus B_d(y, \frac{r_c}{2})} \mathbb{E}(f^{n-D}(x) - f^{n-D}(y))\mu''(dx)\right|. \qquad (99) \end{aligned}$$

Observe that the signed version of (98) can be rewritten as

$$\int_{x \in B_d(y, \frac{r_c}{2})} (f^{n-D}(x) - f^{n-D}(y))\mu''(dx) = \tag{100}$$

$$\left(\frac{1}{2}\right) \left(\int_{x \in B_d(y, \frac{r_c}{2})} (f^{n-D}(x) - f^{n-D}(y) - \partial_{x-y} f^{n-D}(y))\mu''(dx) \right) + \tag{101}$$

$$\left(\frac{1}{2}\right)\left(\int_{x\in B_d(y,\frac{r_c}{2})}(f^{n-D}(2y-x)-f^{n-D}(y)-\partial_{y-x}f^{n-D}(y))\mu''(d(2y-x))\right)+$$

$$\left(\frac{1}{2}\right)\left(\int_{x\in B_d(y,\frac{r_c}{2})}\partial_{x-y}f^{n-D}(y)(\mu''(dx)-\mu''(d(2y-x))\right). \tag{102}$$

We observe that twice the magnitude of (101) satisfies,

$$\begin{aligned}\Big|\int_{x\in B_d(y,\frac{r_c}{2})}\mathbb{E}(f^{n-D}(x)-f^{n-D}(y)-\partial_{x-y}f^{n-D}(y)+\widehat{G}^x_{n-D})\mu''(dx)\Big| &= \\ \Big|\int_{x\in B_d(y,\frac{r_c}{2})}(f^{n-D}(x)-f^{n-D}(y)-\partial_{x-y}f^{n-D}(y))\mu''(dx)\Big| &\leq \\ \Big|\int_{x\in B_d(y,\frac{r_c}{2})}\left(\frac{C|x-y|^2}{\tau}\right)\Gamma^{-1}\gamma_d(\Pi_dV_y-x)\gamma_{D-d}(\Pi_{D-d}V_y-f^{n-D}(x))\mu'(dx)\Big| &\leq \frac{Cd\sigma^2}{\tau}.\end{aligned}$$

By symmetry, the same bound applies to twice the next term. Now we need to bound (102). To do so, observe that

$$\begin{aligned}&\Big|\int_{x\in B_d(y,\frac{r_c}{2})}\partial_{x-y}f^{n-D}(y)(\mu''(dx)-\mu''(d(2y-x))\Big| \\ \leq\ &\int_{x\in B_d(y,\frac{r_c}{2})}|\partial_{x-y}f^{n-D}(y)|\cdot|(\mu''(dx)-\mu''(d(2y-x))| \\ \leq\ &\int_{x\in B_d(y,\frac{r_c}{2})}\left(C|x-y|+\frac{C|x-y|^2}{\tau}\right)\Gamma^{-1}\gamma_d(\Pi_dV_y-x)\left(\frac{|x-y|}{\tau}\right)\mu'(dx) \\ \leq\ &\left(\frac{Cd\sigma^2}{\tau}+\frac{C\sigma^3d^{\frac{3}{2}}}{\tau^2}\right)<\frac{C\sigma^2}{\tau}.\end{aligned}$$

Lastly, as we show below, from the tail decay of $\frac{d\mu''}{d\mu'}$, proceeding in a way that is analogous to (96), it follows that the term (99) is bounded above by $\frac{\sigma^2}{\tau}$. Thus,

$$\begin{aligned}
&\left|\int_{x\in B_d(y,\frac{\tau}{20})\setminus B_d(y,\frac{r_c}{2})} (f^{n-D}(x) - f^{n-D}(y))\mu''(dx)\right| \\
= &\left(\frac{1}{2}\right)\left|\int_{x\in B_d(y,\frac{\tau}{20})\setminus B_d(y,\frac{r_c}{2})} (f^{n-D}(x) - f^{n-D}(y) - \partial_{x-y}f^{n-D}(y))\mu''(dx)\right| \\
&+\left(\frac{1}{2}\right)\left|\int_{x\in B_d(y,\frac{\tau}{20})\setminus B_d(y,\frac{r_c}{2})} (f^{n-D}(2y-x) - f^{n-D}(y) - \partial_{y-x}f^{n-D}(y))\mu''(d(2y-x))\right| \\
&+\left(\frac{1}{2}\right)\left|\int_{x\in B_d(y,\frac{\tau}{20})\setminus B_d(y,\frac{r_c}{2})} \partial_{x-y}f^{n-D}(y)(\mu''(dx) - \mu''(d(2y-x))\right| \\
\leq &\int_{x\in B_d(y,\frac{\tau}{20})\setminus B_d(y,\frac{r_c}{2})} \left(\frac{C|x-y|^2}{\tau}\right)\mu''(dx) \\
&+\int_{x\in B_d(y,\frac{\tau}{20})\setminus B_d(y,\frac{r_c}{2})} \left(\frac{C|x-y|^2}{\tau}\right)\mu''(d(2y-x)) \\
&+\int_{x\in B_d(y,\frac{\tau}{20})\setminus B_d(y,\frac{r_c}{2})} \left(C|x-y| + \frac{C|x-y|^2}{\tau}\right)\mu''(dx) \leq \frac{\sigma^2}{\tau}.
\end{aligned}$$

□

Definition 5.3. *We define the "refined-net"* Rnet_0 *of* $\mathcal{M} \cap (S_0 + \mathbb{R}^{n-D})$ *by*

$$\mathrm{Rnet}_0 := \{y + \widehat{z}_y + \widehat{z}_y^n | y \in Y_{D_0}\}.$$

Note that $\mathbb{E}[y + \widehat{z}_y + \widehat{z}_y^n] = e_y^n$.

Definition 5.4. *We extend this to a net* Rnet *of* $\mathcal{M}$*, by taking the union over all refined nets* Rnet_i *corresponding to balls of the form* $B_D(x, r)$ *where the center* $x \in X_0$*.*

We observe that $\widehat{z}_y^n$ can be expressed as the sum of two random variables, one that is a sample $\zeta_{y,1}^n$ from the push-forward of μ via the orthogonal projection onto $\mathbb{R}^{n-D}$ and another that is an independent Gaussian $\zeta_{y,2}^n$ belonging to $\mathbb{R}^{n-D}$ having the distribution $N_{n-D}(0, \sigma^2)$. Due to the bound on the distance of any point on $\mathcal{M}$ to $\mathbb{R}^D$, we see that conditional on the correctness of the Principal Component Analysis step in Proposition 3.1, $|\zeta_{y,1}^n| < \frac{\tau}{Cd}$.

Recall from Definition 1.2 that for two subsets X and Y in a metric space $\mathbb{M}$, we define $dist(X, Y)$ to be

$$\sup_{x\in X} \inf_{y\in Y} d_{\mathbb{M}}(x, y).$$

Thus $\mathbf{d}_{\mathsf{haus}}(X, Y) = \max(dist(X, Y), dist(Y, X))$.

Lemma 5.6.

$$\mathbb{P}\left[(dist(\mathrm{Rnet}, \mathcal{M}) < \frac{Cd\sigma^2}{\tau})\ and\ (dist(\mathcal{M}, \mathrm{Rnet}) < C\sigma)\right] > 1 - N_0^{-50}.$$

Proof. We know that assuming the Principle Component Analysis step does not produce an erroneous output, $\mathbf{d}_{\mathtt{haus}}(\mathcal{M}, \Pi_D\mathcal{M}) < \alpha^2\tau$, from Proposition 3.1 and hence the function f^{n-D} (which was introduced in discussion immediately following Definition 5.2,)when restricted to $B_d(0, \frac{r}{2})$ has a C^0-norm of at most $\frac{\tau}{Cd}$. This, together with Lemma 5.5 implies that

$$\mathbb{P}[dist(\bigcup_i \{e_y | y \in Y_{D_i}\}, \mathcal{M}) < \frac{Cd\sigma^2}{\tau}] > 1 - N_0^{-75}. \tag{103}$$

It also implies that

$$\mathbb{P}[dist(\mathcal{M}, \bigcup_i \{e_y | y \in Y_{D_i}\}) < C\sigma] > 1 - N_0^{-75}. \tag{104}$$

We next show that with probability at least $1 - N_0^{-75}$, $\mathbf{d}_{\mathtt{haus}}(\mathrm{Rnet}, \cup_i\{e_y|y \in Y_{D_i}\}) < \frac{C\sigma^2}{\tau}$. This follows from Lemma 5.2, (95) and Proposition 5.1 applied separately to the random variables $\widehat{z}_y$ and $\widehat{z}_y^n$. Indeed $\widehat{z}_y$ is the average of independent samples contained inside $\Pi_D V_y$ and hence $\|\widehat{z}_y\|_{\psi_2}$ is less than $\frac{C\tau}{\sqrt{\#y}}$. On the other hand $\widehat{z}_y^n$ is the average of $\#_y$ random points, each of which is the sum of two independent random variables, one that is a sample from $N_{n-D}(0, \sigma^2)$, and the other that is absolutely continuous with respect to the push forward of μ under Π_{n-D}. Since $\Pi_{n-D}\mathcal{M}$ is contained in a ball of radius $\frac{\tau}{Cd}$ if the Principal Component Analysis step in Proposition 3.1 executes correctly (which is a high probability event), this implies $\|\widehat{z}_y^n\|_{\psi_2}$ is less than $\frac{\tau + \sigma\sqrt{n}}{\sqrt{\#y}}$. We conclude that with probability at least $1 - N_0^{-75}$, $\mathbf{d}_{\mathtt{haus}}(\mathrm{Rnet}, \cup_i\{e_y|y \in Y_{D_i}\}) < \frac{C\sigma^2}{\tau}$. The lemma follows from (103) and (104). □

5.3 Boosting the probability of correctness of Rnet to $1 - \xi$.

Now consider G to be metric space whose elements are finite subsets of $\mathbb{R}^n$, and the metric is the Hausdorff distance. Let $p = \cup_i\{e_y|y \in Y_{D_i}\}$, and $\epsilon = \frac{C\sigma^2}{\tau}$. We have a procedure (see Definition 5.4) by which we can produce independent $p_1, p_2, \ldots$ in G such that for each i, $\mathbb{P}[dist(p_i, p) < \epsilon] > \frac{2}{3}$. Then, we may take $C\log(\xi^{-1})$ points p_i and search for an index j such that at least a $\frac{4}{7}$ fraction of all the points p_i lie within a 2ϵ ball of p_j. If no such p_j exists we declare failure, but if such a p_j is found, as will happen with probability at least $1 - \frac{\xi}{10}$, this p_j has the property that it is within 3ϵ of p with probability at least $1 - \frac{\xi}{10}$.

In what follows we set r to a much smaller value than r_c, namely

$$r := C\sqrt{d}\sigma, \tag{105}$$

and apply the algorithm in Section 4 to the refined net obtained in Section 5.

5.4 Using discs to approximate $\mathcal{M}$

Let $X_3 = \{p_i\}$ be a minimal $cr/d-$net of Rnet. 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_3 be denoted N_3.

We introduce a family of n dimensional balls of radius r, $\{U_i\}_{i\in[N_3]}$ where the center of U_i is p_i and a family of $d-$dimensional embedded discs of radius r, denoted $\{D_i\}_{i\in[N_3]}$, $D_i \subseteq U_i$ 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 4.2. We will need the following properties of (D_i, p_i), which hold with probability at least $1 - N_0^{-C}$:

1. The Hausdorff distance between $\cup_i D_i$ and $\mathcal{M}$ is less than $\frac{Cr^2}{\tau} = \delta$.
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 $|z - p_i| < 3\inf_{i\neq j}, |p_i - p_j|$.

6 Computing weights used to define output manifold $\mathcal{M}_o$

For $v \in \mathbb{R}^n$, let $\theta(v) = (1 - |v|^2)^{d+k}$ for $|v| \le 1$, and $\theta(v) = 0$ for $|v| > 1$. Consider the bump function $\tilde{\alpha}_i : \mathbb{R}^n \to \mathbb{R}$ given by

$$\tilde{\alpha}_i(p_i + rv) = c_i\theta(v).$$

Here k is some fixed integer greater or equal to 3. Let

$$\tilde{\alpha}(x) := \sum_{i\in[N_3]} \tilde{\alpha}_i(x), \quad \text{where } \alpha_i(x) = \frac{\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)}, \text{ for each } i \in [N_3].$$

Lemma 6.1. *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.

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

Claim 6.1. *Let $\lambda_{\mathcal{M}}$ denote the Hausdorff measure supported on $\mathcal{M}$ and let λ denote the Lebesgue measure on $\mathbb{R}^D$. There exists $\kappa \in \mathbb{R}$ such that the following is true. For all z in a $\frac{r}{4d}$ neighborhood of $\mathcal{M}$,*

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

Proof. We make the following claim.

Claim 6.2. *If* $|v| < \frac{1}{\sqrt{2}}$*, then* $\exp(-2(d+k)|v|^2) < \theta(v)$*. Also*

$$\forall v \in \mathbb{R}^D,\ \exp(-(d+k)|v|^2) > \theta(v). \tag{106}$$

Proof. To see the first inequality, note that

$$\begin{aligned}
|v| &< \frac{1}{\sqrt{2}} && (107)\\
\implies (-2)(1-|v|^2) &< -1 && (108)\\
\implies (-2)(d+k)|v|^2 &< (d+k)\left(-\frac{|v|^2}{1-|v|^2}\right) && (109)\\
\implies (-2)(d+k)|v|^2 &< (d+k)\left(\ln(1-|v|^2)\right) && (110)\\
\implies \exp((-2)(d+k)|v|^2) &< \left(1-|v|^2\right)^{d+k} = \theta(v). && (111)
\end{aligned}$$

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

$$-(d+k)|v|^2 \ > \ (d+k)\left(\ln(1-|v|^2)\right). \tag{112}$$

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

□

We now provide the proof of Lemma 6.1.

We will need the following Proposition which follows from Theorem 3.2.3 in [40].

Proposition 6.1. *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 an injective* C^2 *function with* $m \leq n$ *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*

$$\int_A u(x) J_m f(x) \mathcal{L}^m(dx) = \int_{f(A)} u(f^{-1}(y)) \mathcal{H}^m(dy). \tag{113}$$

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

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

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

$$\begin{aligned}
\int_{\mathbb{R}^d} \theta(v)\lambda(dv) &= \int_{B_d} \theta(v)\lambda(dv) && (114)\\
&\leq \int_{B_d} \exp(-(d+k)v)\lambda(dv) \leq (\frac{\pi}{d+k})^{d/2}. && (115)
\end{aligned}$$

Also,

$$\int_{B_d} \theta(v)\lambda(dv), \ \geq \ \int_{\frac{B_d}{\sqrt{2}}} \exp(-2(d+k)v)\lambda(dv) \geq c(\frac{\pi}{2(d+k)})^{d/2}. \tag{116}$$

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

Next consider a unit disk $B_d \subseteq \mathbb{R}^n$ equipped with the measure $c_\theta\lambda$. 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

$$\frac{((c_\theta\lambda 1_{B_d}) * \theta)(q)}{((c_\theta\lambda 1_{B_d}) * \theta)(0)} = \frac{\int_{B_d}\theta(q-v)(c_\theta\lambda(dv))}{\int_{B_d}\theta(-v)(c_\theta\lambda(dv))}, \tag{117}$$

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

$$\frac{((c_\theta\lambda 1_{B_d}) * \theta)(q)}{((c_\theta\lambda 1_{B_d}) * \theta)(0)} \leq 1. \tag{118}$$

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

$$\Delta^2 \leq \left(\frac{1-|v|^2}{4d}\right). \tag{119}$$

Therefore,

$$\begin{aligned}
\int_{B_d}\beta(q-v)(c_\theta\lambda(dv)) &= \int_{B_d}(1-|v|^2-\Delta^2)^{d+k}1_{\{v||v|^2\leq 1-\Delta^2\}}(c_\theta\lambda(dv)) \\
&\geq \int_{\sqrt{1-\frac{1}{2d}}B_d}((1-|v|^2)(1-\frac{1}{4d}))^{d+k}(c_\theta\lambda(dv)) \\
&\geq \int_{\sqrt{1-\frac{1}{2d}}B_d}c(1-|v|^2)^{d+k}(c_\theta\lambda(dv)) \qquad (120)\\
&\geq c\int_{B_d}(1-|v|^2)^{d+k}(c_\theta\lambda(dv)). \qquad (121)
\end{aligned}$$

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 $\theta(v_1) \geq \theta(v_2)$ if $|v_1| < |v_2|$.

We thus have

$$c \leq \frac{((c_\theta\lambda 1_{B_d}) * \theta)(q)}{((c_\theta\lambda 1_{B_d}) * \theta)(0)} = \frac{\int_{B_d}\theta(q-v)(c_\theta\lambda(dv))}{\int_{B_d}\theta(-v)(c_\theta\lambda(dv))} \leq 1, \tag{122}$$

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

$$\frac{((c_\theta\mathcal{H}^d_{\mathcal{M}}1_{B_m}) * \theta)(q)}{((c_\theta\lambda 1_{B_d}) * \theta)(0)} = \frac{\int_{\mathcal{M}\cap B_D}\theta(-v)(c_\theta\mathcal{H}^d_{\mathcal{M}}(dv))}{\int_{B_d}\theta(-v)(c_\theta\lambda(dv))}, \tag{123}$$

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

$$\sup_{x \in \mathcal{M} \cap B_n} |x - \Pi_d x| = \Delta. \tag{124}$$

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 $n-d$ 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, $v = w + f(w)$. Then,

$$\begin{aligned} \int_{\mathcal{M} \cap B_n} \theta(-v)(c_\theta \mathcal{H}^d_{\mathcal{M}}(dv)) &= \int_{\Pi_d(\mathcal{M} \cap B_n)} \beta(-(w + f(w)))(c_\beta \mathcal{H}^d_{\mathcal{M}}(dv)) \\ &\leq \int_{\Pi_d(\mathcal{M} \cap B_n)} \theta(-w)(c_\theta \mathcal{H}^d_{\mathcal{M}}(dv)) \qquad (125) \\ &\leq \int_{\Pi_d(\mathcal{M} \cap B_n)} \theta(-w)(c_\theta J(w)\lambda(dw)). \qquad (126) \end{aligned}$$

Since $\|Df\|$ is of the order of $\frac{1}{Cd^C}$ by Lemma A.2 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, in view of Proposition 6.1, implies that

$$\int_{\Pi_d(\mathcal{M} \cap B_n)} \theta(-w)(c_\theta J(w)\lambda(dw)) \leq C \int_{B_d} \theta(-v)(c_\theta \lambda(dv)). \tag{127}$$

This in turn implies that

$$c^{-1} > \frac{\int_{\mathcal{M} \cap B_n} \theta(-v)(c_\theta \mathcal{H}^d_{\mathcal{M}}(dv))}{\int_{B_d} \theta(-v)(c_\beta \lambda(dv))}. \tag{128}$$

for an appropriately small universal constant c.

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

$$\begin{aligned} \int_{\mathcal{M} \cap B_n} \theta(-v)(c_\theta \mathcal{H}^d_{\mathcal{M}}(dv)) &= \int_{\Pi_d(\mathcal{M} \cap B_n)} \theta(-(w + f(w)))(c_\theta \mathcal{H}^d_{\mathcal{M}}(dv)) \\ &\geq \int_{B_d(1-1/d)} \theta(-(w + f(w)))(c_\theta \mathcal{H}^d_{\mathcal{M}}(dv)) \qquad (129) \\ &\geq \int_{B_d(1-1/d)} (1 - |w|^2 - \Delta^2)^{d+k}(c_\theta J(w)\lambda(dw)) \\ &\geq \int_{B_d(1-1/d)} ((1 - |w|^2)(1 - 1/d))^{d+k}(c_t \beta \lambda(dw)) \\ &\geq c \int_{B_d(1-1/d)} (1 - |w|^2)^{d+k}(c_\theta \lambda(dw)) \qquad (130) \\ &\geq c^2 \int_{B_d} (1 - |w|^2)^{d+k}(c_\theta \lambda(dw)). \qquad (131) \end{aligned}$$

The last step comes from dilating the disk $(1 - \frac{1}{d})B_d$ to B_d and observing that $\theta(v_1) \geq \theta(v_2)$ 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

$$\frac{\int_{\mathcal{M}\cap B_n} \theta(-v)(c_\theta \mathcal{H}^d_{\mathcal{M}}(dv))}{\int_{B_d} \theta(-v)(c_\theta \lambda(dv))} > c. \tag{132}$$

□

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.

Claim 6.3. *There exists a measure μ_P supported on $\{p_i\}$ such that*

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

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

Proof. For any $\epsilon \in (0,1)$, let $\theta_\epsilon(\epsilon r v) = c_{\epsilon,\theta}(1 - \|v\|^2)^{d+k}$ if $|v| \leq 1$, and $\theta_\epsilon(\epsilon r v) = 0$ if $|v| > 1$. Here $c_{\epsilon,\theta}$ is chosen so that θ_ϵ integrates to 1 over $\mathbb{R}^n$.

Definition 6.1. *For $i \in [N_3]$, 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*

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

We note $\frac{d(c_\theta \mathcal{H}^d_{\mathcal{M}} * \theta)}{d\lambda}(z)$ is a $\frac{d}{cr}-$Lipschitz function of z, and so is also the function

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

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

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

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

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

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

$$\text{diam(supp}(c_\theta \mathcal{H}^d_{\mathcal{M}} * \theta_\epsilon) \cap \text{Vor}_i) < \frac{cr}{d}.$$

Let π denote the map defined on $supp(c_\theta \mathcal{H}^d_{\mathcal{M}} * \theta_\epsilon)$ from Vor$_i$ to p_i. Then,

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

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

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

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

□

Let λ^i_d denote the $d-$dimensional Lebesgue measure restricted to the disc D_i.
Recall that from Definition 6.1 that

$$\mu_P(p_i) = (c_\theta \mathcal{H}^d_{\mathcal{M}} * \theta_\epsilon)(\text{Vor}_i),$$

where the $\text{Vor}_i \subset \mathbb{R}^n$. Let

$$\widetilde{\mu_P}(p_i) = (c_\theta \lambda^i_d)(\text{Vor}_i \cap D_i).$$

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,

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

for a suitable universal constant c. We see that $(c_\theta \lambda^i_d)(\text{Vor}_i \cap D_i)$ is the volume of the polytope $\text{Vor}_i \cap D_i$ multiplied by c_θ, and membership of a point in $\text{Vor}_i \cap D_i$ can be answered in time $(Cd)^d$. Thus by placing a sufficiently fine grid, and counting the lattice points in $\text{Vor}_i \cap D_i$, $\widetilde{\mu_P}(p_i)$ can be computed using $(Cd)^{2d}$ deterministic steps. Even faster randomized algorithms exist for the task, which we choose not to delve into here. This concludes the proof of Lemma 6.1. □

7 The output manifold

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

Let Π^i be the orthogonal projection of $\mathbb{R}^n$ onto the $n-d-$dimensional subspace containing the origin that is orthogonal to the affine span of D_i. Recall that the p_i are the centers of the discs D_i as i ranges over $[N_3]$. 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

$$F(x) = \sum_{i \in [N_3]} \alpha_i(x) F_i(x). \tag{133}$$

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.

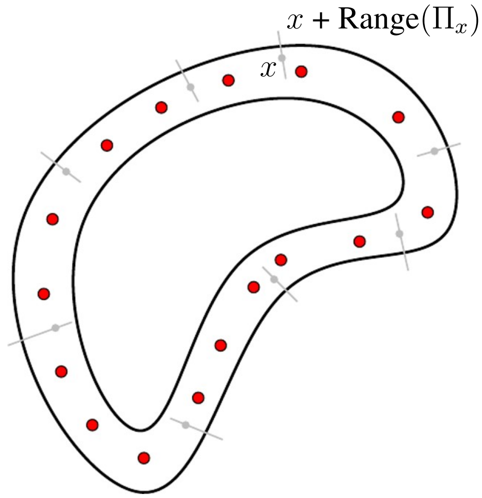


Figure 1: A "vector bundle" over a neighborhood of the data, used to produce the output manifold. In the figure, x is the gray point and x + Range(Π_x) is the gray line segment containing x.

Definition 7.1. *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 $\widetilde{U}_i$ be defined as the $\frac{cr}{d}$−Eucidean neighborhood of D_i intersected with U_i. Given a matrix X, its Frobenius norm $\|X\|_F$ is defined as the square root of the sum of the squares of all the entries of X. This norm is unchanged when X is premultiplied or postmultiplied by orthogonal matrices (of the appropriate order). Note that Π_x is $\mathcal{C}^2$ when restricted to $\bigcup_i \widetilde{U}_i$, because the $\alpha_i(x)$ are C^2 and when x is in this set, $c < \sum_i \widetilde{\alpha}_i(x) < c^{-1}$, 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 7.2. *The output manifold $\mathcal{M}_o$ is the set of all points $x \in \bigcup_i \widetilde{U}_i$ such that $\Pi_x F(x) = 0$.*

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

$$\Pi_{hi}(\sum_{i\in[N_3]} \alpha_i(x)\Pi^i)(\sum_{i\in[N_3]} \alpha_i(x)\Pi^i(x - p_i)) = 0. \tag{134}$$

We see that

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

using diagonalization and Cauchy's integral formula, and so

$$\frac{1}{2\pi i}\left[\oint_\gamma (zI - (\sum_i \alpha_i(x)\Pi^i))^{-1} dz\right]\left(\sum_i \alpha_i(x)\Pi^i(x - p_i)\right) = 0 \tag{135}$$

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

Let

$$\sum \alpha_i(x)\Pi^i = M(x), \tag{136}$$

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 (135) can be written as

$$\oint_\gamma \frac{dz}{2\pi i}\left(\sum_i \alpha_i(x)(zI - M(x))^{-1}F_i(x)\right). \tag{137}$$

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

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

Then,

$$\partial_v \oint_\gamma \frac{dz}{2\pi i}\left(\sum_i \alpha_i(x)(zI - M(x))^{-1}F_i(x)\right) = \sum_i \alpha_i(x)\Pi_x(\partial_v F_i(x)) \tag{138}$$

$$+ \sum_i \alpha_i(x)(\partial_v \Pi_x)F_i(x) \tag{139}$$

$$+ \sum_i (\partial_v \alpha_i(x))\Pi_x F_i(x). \tag{140}$$

Let $v \in \mathbb{R}^n$ be such that $\|v\| = 1$. 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 7.1. *The reach of* $\mathcal{M}_\Pi^d \subset \mathbb{R}^{n \times n}$ *is greater or equal to* $1/2$.

Proof. Let

$$\mathcal{M}_\Pi := \bigcup_{\widehat{d}=0}^{n} \mathcal{M}_\Pi^{\widehat{d}}.$$

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,

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

□

In what follows, we will make repeated use of Hölder's inequality for ℓ_p norms and ℓ_q norms: 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 $w \in \mathbb{R}^{N_3}$ that is supported on the set of all j such that $U_i \cap U_j \neq \emptyset$ will satisfy

$$\|w\|_{d+k} \leq Cd\|w\|_\infty, \quad \|w\|_{\frac{d+k}{2}} \leq Cd^2\|w\|_\infty, \quad \|w\|_{\frac{d+k}{3}} \leq Cd^3\|w\|_\infty.$$

Thirdly, we will use 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. Recall that N_3 is the total number of balls U_i.

Lemma 7.1. *We have for any $v \in \mathbb{R}^D$ such that $|v| = 1$, and any $x \in \mathbb{R}^D$ such that $dist(x, \mathcal{M}) \leq \frac{cr}{d}$,*

$$\|(\partial_v \alpha_i(x))_{i \in [N_3]}\|_{\frac{d+k}{d+k-1}} \leq Cd^2. \tag{141}$$

Proof. We have

$$\begin{aligned}
\|(\partial_v \alpha_i(x))_{i \in [N_3]}\|_{\frac{d+k}{d+k-1}} &= \|(\partial_v \frac{\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)})_{i \in [N_3]}\|_{\frac{d+k}{d+k-1}} \\
&\leq \frac{\|(\partial_v \tilde{\alpha}_i(x))_{i \in [N_3]}\|_{\frac{d+k}{d+k-1}}}{\tilde{\alpha}} + \frac{\|((\partial_v \tilde{\alpha}(x))\tilde{\alpha}_i(x))_{i \in [N_3]}\|_{\frac{d+k}{d+k-1}}}{\tilde{\alpha}^2} \\
&\leq (c^{-1})\|Cd(\tilde{\alpha}_i(x))_{i \in [N_3]}\|_1^{\frac{d+k-1}{d+k}} + (c^{-2})\|(\tilde{\alpha}_i(x))_{i \in [N_3]}\|_{\frac{d+k}{d+k-1}} |\partial_v \tilde{\alpha}| \\
&\leq Cd + C|\partial_v \tilde{\alpha}| \\
&\leq Cd + C\|(\partial_v \tilde{\alpha}_i)_{i \in [N_3]}\|_{\frac{d+2}{d+1}} \|(1)_{i \in [N_3]}\|_{d+2} \\
&\leq Cd^2.
\end{aligned}$$

□

Recall that as the F_i are affine maps, $\partial F_i(x) = \Pi^i$. We first look at the right hand side of (138). This can be rewritten as

$$\sum_{i \in [N_3]} \alpha_i(x) \Pi_x \Pi_i v = \Pi_x v + \Pi_x \left(M(x) - \Pi_x\right) v. \tag{142}$$

It follows from properties of the Frobenius norm that

$$\Pi_{hi}(A_x) = \arg \min_{\Pi \in \mathcal{M}_\Pi^s} \|A_x - \Pi\|_F.$$

Thus, recalling from (136) that

$$\sum \alpha_i(x) \Pi^i = M(x),$$

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

We look at (139) next. Observe that

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

Lemma 7.2. *We have for any $v \in \mathbb{R}^n$ such that $|v| = 1$, and any $x \in \mathbb{R}^n$ such that $dist(x, \mathcal{M}) \leq \frac{cr}{d}$,*

$$\|(\partial_v^2 \alpha_i(x))_{i \in [N_3]}\|_{\frac{d+k}{d+k-2}} \leq C d^4. \tag{143}$$

Proof. We have

$$\|(\partial_v^2 \alpha_i(x))_{i \in [N_3]}\|_{\frac{d+k}{d+k-2}} = \|(\partial_v^2 \frac{\tilde\alpha_i(x)}{\tilde\alpha(x)})_{i \in [N_3]}\|_{\frac{d+k}{d+k-2}}$$

$$= \|(\frac{\partial_v^2 \tilde\alpha_i(x)}{\tilde\alpha(x)} + \frac{(-2)(\partial_v \tilde\alpha_i(x))(\partial_v \tilde\alpha(x))}{\tilde\alpha(x)^2} + \frac{\tilde\alpha_i(x)}{\tilde\alpha(x)^3}(2(\partial_v \tilde\alpha)^2 - \partial_v^2 \tilde\alpha(x)(\tilde\alpha(x)))_{i \in [N_3]}\|_{\frac{d+k}{d+k-2}}.$$

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 7.2. *We have*

$$\|(\frac{\partial_v^2 \tilde\alpha_i(x)}{\tilde\alpha(x)})_{i \in [N_3]}\|_{\frac{d+k}{d+k-2}} \leq C d^2, \tag{144}$$

Proof. This follows from $c < \tilde{\alpha} < C$, and the discussion below. Suppose x belongs to the unit ball in $\mathbb{R}^D$. Then,

$$\begin{aligned}
\partial_v^2(1-\|x\|^2)^{k+d} &= \partial_v((k+d)(1-\|x\|^2)^{k+d-1}(2\langle x, v\rangle)) && (145)\\
&= (k+d)(k+d-1)(1-\|x\|^2)^{k+d-2}(4\langle x, v\rangle)^2) && (146)\\
&+ (k+d)(1-\|x\|^2)^{k+d-1}(2\langle v, v\rangle)). && (147)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|(\partial_v^2\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}} &\leq C\left(d^2\tilde{\alpha} + d\|(\partial_v\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}}\right) && (148)\\
&\leq C\left(d^2 + d\|(\partial_v\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-1}}\right) \leq Cd^2. && (149)
\end{aligned}$$

□

Claim 7.3. *We have*

$$\|(\frac{(-2)(\partial_v\tilde{\alpha}_i(x))(\partial_v\tilde{\alpha}(x))}{\tilde{\alpha}(x)^2})_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}} \leq Cd^3. \quad (150)$$

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

$$\begin{aligned}
\|(\frac{(-2)(\partial_v\tilde{\alpha}_i(x))(\partial_v\tilde{\alpha}(x))}{\tilde{\alpha}(x)^2})_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}} &< Cd^2\|((\partial_v\tilde{\alpha}_i(x)))_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}}\\
&\leq Cd^2\|((\partial_v\tilde{\alpha}_i(x)))_{i\in[N_3]}\|_{\frac{d+k}{d+k-1}} && (151)\\
&\leq Cd^3. && (152)
\end{aligned}$$

□

Claim 7.4. *We have*

$$\|(\frac{\tilde{\alpha}_i(x)}{\tilde{\alpha}(x)^3}(2(\partial_v\tilde{\alpha})^2 - \partial_v^2\tilde{\alpha}(x)(\tilde{\alpha}(x)))_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}} \leq Cd^4. \quad (153)$$

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

$$\begin{aligned}
C|\partial_v^2\tilde{\alpha}| &\leq C\|(\partial_v^2\tilde{\alpha}_i)_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}}\|(1)_{i\in[N_3]}\|_{(d+k)/2} && (154)\\
&\leq Cd^4. && (155)
\end{aligned}$$

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

This proves Lemma 7.2. □

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

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

We proceed to obtain an upper bound on $\|\partial_v \Pi_x\|$, for $x \in \bigcup_i \tilde{U}_i$. Recall that this implies that $c < \tilde{\alpha}(x) < C$. Recall that the radius of the circle γ is $\frac{1}{2}$. Thus,

$$\begin{aligned}
\|\partial_v \Pi_x\| &\leq \left(\frac{1}{2}\right) \left\|\partial_v((zI - M(x))^{-1})\right\| &(156)\\
&= \left(\frac{1}{2}\right) \|(zI - M(x))^{-1} \partial_v M(x)(zI - M)^{-1}\| &(157)\\
&\leq \left(\frac{1}{2}\right) \|(zI - M(x))\|^{-2} \|\partial_v M(x)\| &(158)\\
&\leq 8\|\partial_v M(x)\| &(159)\\
&= 8\|\sum_i \partial_v \alpha_i(x)(\Pi^i - \Pi^1) + \partial_v \sum_i \alpha_i(x)\Pi_1\| &(160)\\
&\leq 8\sum_i |\partial_v \alpha_i(x)|\delta + 0 &(161)\\
&\leq 8\|(\partial_v \alpha_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-1}} \|(\delta)_{i\in[N_3]}\|_{d+k} &(162)\\
&\leq Cd^3\delta, &(163)
\end{aligned}$$

where C is an absolute constant.

Therefore,

$$\left\| \oint_\gamma \frac{dz}{2\pi i} \left(\sum_i \alpha_i(x) \left(\partial_v((zI - M(x))^{-1})\right) F_i(x) \right) \right\| \leq Cd^3\delta. \quad (164)$$

Finally, we bound (140) from above,

$$\begin{aligned}
&\left\| \oint_\gamma \frac{dz}{2\pi i} \left(\sum_i \left(\partial_v \alpha_i(x)\right) (zI - M(x))^{-1} F_i(x) \right) \right\| \\
&\leq \left\| \Pi_x \left(\sum_i (\partial_v \alpha_i(x))(F_i(x) - F_1(x)) \right) \right\| + \left\| \left(\sum_i \partial_v \alpha_i(x) \right) F_1(x) \right\| \\
&\leq \|\Pi_x\| \sum_i |\partial_v \alpha_i(x)| \|F_i(x) - F_1(x)\| + 0 \\
&\leq \|(\partial_v \alpha_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-1}} \|(F_i(x) - F_1(x))_{i\in[N_3]}\|_{d+k} \\
&\leq Cd^3\delta.
\end{aligned}$$

Therefore,

$$\|\partial_v (\Pi_x F(x)) - \Pi_x v\| \leq Cd^3\delta. \quad (165)$$

Note also by (156)-(163) that

$$\|\partial_v \Pi_x\| \leq Cd^3\delta. \quad (166)$$

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

We now proceed to obtain an upper bound on $\|\partial_v^2 (\Pi_x F(x))\|$. To this end, we use that

$$\begin{aligned} \|\partial_v^2 (\Pi_x F(x))\| &\leq \|(\partial_v^2 \Pi_x) F(x)\| && (167)\\ &+ \|2(\partial_v \Pi_x)\partial_v F(x)\| && (168)\\ &+ \|\Pi_x \partial_v^2 F(x)\|. && (169) \end{aligned}$$

We first bound from above the right side of (167). To this end, we observe that

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

Therefore,

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

Next, we bound (168) from above. Note that

$$\begin{aligned} \|\partial_v F(x)\| &\leq \|(\sum_i (\partial_v \alpha_i(x)(F_i(x) - F_1(x))) + \Pi_x v\| && (170)\\ &\leq 1 + Cd^3\delta && (171) \end{aligned}$$

and

$$\begin{aligned} \|(\partial_v \Pi_x)\partial_v F(x)\| &\leq \|(\partial_v \Pi_x)\|\|\partial_v F(x)\| && (172)\\ &\leq (Cd^3\delta)(1 + d^3\delta) && (173)\\ &= Cd^3\delta + Cd^6\delta^2. && (174) \end{aligned}$$

Finally, we bound (169) from above by observing that

$$\begin{aligned}
\|\Pi_x \partial_v^2 F(x)\| &\leq \|\partial_v^2 F(x)\| && (175)\\
&\leq \|\partial_v^2 (F(x) - F_1(x))\| && (176)\\
&\leq \sum_i |\partial_v^2 \alpha_i(x)| \|F_i(x) - F_1(x)\| && (177)\\
&+ \sum_i 2|\partial_v \alpha_i(x)| \|\partial_v F_i(x) - \partial_v F_1(x)\| && (178)\\
&+ \sum_i |\alpha_i(x)| \|\partial_v^2 F_i(x)\|. && (179)
\end{aligned}$$

We first bound (177) from above.

$$\begin{aligned}
&\sum_i |\partial_v^2 \alpha_i(x)| \|F_i(x) - F_1(x)\| \\
&\leq \|(|\partial_v^2 \alpha_i(x)|)_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}} \|(\|F_i(x) - F_1(x)\|)_{i\in[N_3]}\|_{\frac{d+k}{2}} && (180)\\
&\leq (Cd^4)(d^2\delta) && (181)\\
&= Cd^6\delta. && (182)
\end{aligned}$$

Next we bound (178) from above.

$$\begin{aligned}
&\sum_i |\partial_v \alpha_i(x)| \|\partial_v F_i(x) - \partial_v F_1(x)\| \\
&\leq \|(|\partial_v \alpha_i(x)|)_{i\in[N_3]}\|_{\frac{d+k}{d+k-1}} \|(\|\partial_v F_i(x) - \partial_v F_1(x)\|)_{i\in[N_3]}\|_{d+k} \\
&\leq Cd^2(d\delta) \\
&= Cd^3\delta.
\end{aligned}$$

Observe, that the term (179) is equal to 0. Therefore,

$$\left\|\partial_v^2 \left(\Pi_x F(x)\right)\right\| \leq Cd^6\delta. \tag{183}$$

7.3 A bound on the third derivative of $\Pi_x F(x)$

We now proceed to obtain an upper bound on $\|\partial_v^3 \left(\Pi_x F(x)\right)\|$.

Claim 7.5.

$$\|(\partial_v^3 \tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}} \leq Cd^3. \tag{184}$$

Proof. This follows from $c < \tilde{\alpha} < C$, and the discussion below. Suppose x belongs to the unit ball in $\mathbb{R}^D$. Then,

$$\begin{aligned}
\partial_v^3(1-\|x\|^2)^{k+d} &= \partial_v^2((k+d)(1-\|x\|^2)^{k+d-1}(2\langle x,v\rangle)) \\
&= (k+d)\partial_v\Big((k+d-1)(1-\|x\|^2)^{k+d-2}(4\langle x,v\rangle)^2 \\
&\qquad +(1-\|x\|^2)^{k+d-1}(2\langle v,v\rangle)\Big) \\
&= (k+d)((k+d-1)(\partial_v((1-\|x\|^2)^{k+d-2})(4\langle x,v\rangle^2)) \\
&\quad +(k+d)(\partial_v((1-\|x\|^2)^{k+d-1}))(2\langle v,v\rangle).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\|(\partial_v^3\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}} \\
&\leq C\left(d^3\tilde{\alpha} + d^2\|(\partial_v\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}} + d\|(\partial_v^2\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}}\right) \\
&\leq C\left(d^3\tilde{\alpha} + d^2\|(\partial_v\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-1}} + d\|(\partial_v^2\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-2}}\right) \\
&\leq C\left(d^3\right).
\end{aligned}$$

□

As a consequence, we see the following.

$$|(\partial_v^3\tilde{\alpha}(x))| \leq \|(\partial_v^3\tilde{\alpha}_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}}\|(1)_{i\in[N_3]}\|_{\frac{d+k}{3}} \tag{185}$$

$$\leq Cd^6. \tag{186}$$

Lemma 7.3. *We have for any $v\in\mathbb{R}^n$ such that $|v|=1$, and any $x\in\mathbb{R}^n$ such that $dist(x,\mathcal{M})\leq\frac{cr}{d}$,*

$$\|(\partial_v^3\alpha_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}} \leq Cd^6. \tag{187}$$

Proof. We see that

$$\partial_v^3(\alpha_i\tilde{\alpha}) = (\partial_v^3\alpha_i)\tilde{\alpha} + (3\partial_v^2\alpha_i)(\partial_v\tilde{\alpha}) + (3\partial_v\alpha_i)(\partial_v^2\tilde{\alpha}) + (\alpha_i)\partial_v^3\tilde{\alpha}. \tag{188}$$

Therefore,

$$\begin{aligned}
&\tilde{\alpha}(x)\|(\partial_v^3\alpha_i(x))_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}} \\
=\ &\|\left(-\partial_v^3(\tilde{\alpha}_i) + (3\partial_v^2\alpha_i)(\partial_v\tilde{\alpha}) + (3\partial_v\alpha_i)(\partial_v^2\tilde{\alpha}) + (\alpha_i)\partial_v^3\tilde{\alpha}\right)_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}}
\end{aligned}$$

The right hand side above can be bounded above by

$$Cd^3 + \|((3\partial_v^2\alpha_i)\partial_v\tilde{\alpha})_i\|_{\frac{d+k}{d+k-3}} + \|((3\partial_v\alpha_i)\partial_v^2\tilde{\alpha})_i\|_{\frac{d+k}{d+k-3}} + \|((\alpha_i)\partial_v^3\tilde{\alpha})_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}}.$$

This is bounded above by

$$Cd^3 + |\partial_v\tilde{\alpha}|\|((3\partial_v^2\alpha_i))_i\|_{\frac{d+k}{d+k-2}} + |\partial_v^2\tilde{\alpha}|\|((3\partial_v\alpha_i))_i\|_{\frac{d+k}{d+k-1}} + |\partial_v^3\tilde{\alpha}|\|((\alpha_i))_{i\in[N_3]}\|_1,$$

which is in turn bounded above by

$$Cd^3 + (Cd^2)(Cd^2) + (Cd^4)(Cd) + (Cd^6),$$

in which the dominant term is Cd^6.

□

Lemma 7.4.

$$\|\partial_v^3 \left(\Pi_x F(x)\right)\| \leq Cd^9\delta. \tag{189}$$

Proof. We have

$$\begin{aligned}
\|\partial_v^3 \left(\Pi_x F(x)\right)\| &\leq \|(\partial_v^3 \Pi_x) F(x)\| && (190)\\
&+ \|3(\partial_v^2 \Pi_x)\partial_v F(x)\| && (191)\\
&+ \|3(\partial_v \Pi_x)\partial_v^2 F(x)\| && (192)\\
&+ \|\Pi_x \partial_v^3 F(x)\|. && (193)
\end{aligned}$$

We first bound from above the right side of (190). Let $A := zI - M(x)$ and $B := (zI - M(x))^{-1}$. Then,

$$\begin{aligned}
0 &= \partial_v^3(AB)\\
&= (\partial_v^3 A)B + 3(\partial_v^2 A)(\partial_v B) + 3(\partial_v A)(\partial_v^2 B) + A(\partial_v^3 B).
\end{aligned}$$

Thus,

$$-A(\partial_v^3 B) = (\partial_v^3 A)B + 3(\partial_v^2 A)(\partial_v B) + 3(\partial_v A)(\partial_v^2 B),$$

and so,

$$\partial_v^3 B = -B(\partial_v^3 A)B - 3B(\partial_v^2 A)(\partial_v B) - 3B(\partial_v A)(\partial_v^2 B),$$

Thus,

$$\|\partial_v^3 B\| \leq C\left[\|\partial_v^3 A\| + \|\partial_v^2 A\|\|\partial_v B\| + \|\partial_v A\|\|\partial_v^2 B\|\right]$$

and

$$\begin{aligned}
\|\partial_v^3 A\| &= \|\partial_v^3 \sum_i \alpha_i \Pi_i\|\\
&= \sum_i |\partial_v^3 \alpha_i (\Pi_i - \Pi_1)|\\
&\leq \|(\partial_v^3 \alpha_i)_{i\in[N_3]}\|_{\frac{d+k}{d+k-3}} \|(\delta)_{i\in[N_3]}\|_{\frac{d+k}{3}}\\
&\leq Cd^9\delta.
\end{aligned}$$

We already know by (166) that

$$\|\partial_v \Pi_x\| \le C d^3 \delta$$

and

$$\|\partial_v^2 \Pi_x\| \le C d^6 \delta.$$

We have shown that

$$\|\partial_v F(x)\| \le 1 + C d^3 \delta.$$

We have also already shown in (179) that

$$\|\partial_v^2 F(x)\| \le C d^6 \delta.$$

We proceed to get an upper bound on $\|\Pi_x \partial_v^3 F(x)\|$,

$$\begin{aligned}
\|\Pi_x \partial_v^3 F(x)\| &\le \|\partial_v^3 F(x)\| && (194)\\
&\le \|\partial_v^3 (F(x) - F_1(x))\| && (195)\\
&\le \sum_i |\partial_v^3 \alpha_i(x)| \|F_i(x) - F_1(x)\| && (196)\\
&\quad + \sum_i 3|\partial_v^2 \alpha_i(x)| \|\partial_v F_i(x) - \partial_v F_1(x)\| && (197)\\
&\quad + \sum_i 3|\partial_v \alpha_i(x)| \|\partial_v^2 F_i(x)\| && (198)\\
&\quad + \sum_i |\alpha_i(x)| \|\partial_v^3 F_i(x)\|. && (199)
\end{aligned}$$

For each i, $\partial_v^2 F_i$ is 0, and so, the above expression reduces to

$$\sum_i |\partial_v^3 \alpha_i(x)| \|F_i(x) - F_1(x)\| + \sum_i 3|\partial_v^2 \alpha_i(x)| \|\partial_v F_i(x) - \partial_v F_1(x)\|.$$

Here, we have

$$\begin{aligned}
\sum_i |\partial_v^3 \alpha_i(x)| \|F_i(x) - F_1(x)\| &\le \|(|\partial_v^3 \alpha_i(x)|)_{i \in [N_3]}\|_{\frac{d+k}{d+k-3}} \|(\|F_i(x) - F_1(x)\|)_{i \in [N_3]}\|_{\frac{d+k}{3}}\\
&\le (C d^6)(d^3 \delta) = C d^9 \delta, && (200)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_i |\partial_v^2 \alpha_i(x)| \|\partial_v F_i(x) - \partial_v F_1(x)\|\\
&\le \|(|\partial_v^2 \alpha_i(x)|)_{i \in [N_3]}\|_{\frac{d+k}{d+k-2}} \|(\|\partial_v F_i(x) - \partial_v F_1(x)\|)_{i \in [N_3]}\|_{\frac{d+k}{2}}\\
&\le C d^4 (d^2 \delta) = C d^6 \delta.
\end{aligned}$$

□

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

$$\Pi_{hi}(\sum_i \alpha_i(x)\Pi^i)(\sum_i \alpha_i(x)\Pi^i(x-p_i)) = 0. \tag{201}$$

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 $\mathbb{R}^d$), thus $z = (x, y) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$. We define $g(x, y) = (x, h(x, y))$. This function is then substituted into the quantitative inverse function theorem of Subsection B.

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

For this subsection, we choose a new length scale so that $\tau = 1$. Let $r = C\sqrt{d}\sigma$. Suppose that $dist(X_3, \mathcal{M}) < \delta := Cr^2$, and $dist(\mathcal{M}, X_3) < cr$. These are the parameters for the refined net.

Theorem 7.5. *The Hausdorff distance between $\mathcal{M}_o$ and $\mathcal{M}$ is less than $\frac{Cd\sigma^2}{\tau}$.*

Proof. Since $dist(X_3, \mathcal{M}) < \delta := Cr^2$, and $dist(\mathcal{M}, X_3) < cr$ the Hausdorff distance between $\bigcup_i D_i$ and $\mathcal{M}$ is less than $\delta = Cr^2$ by Subsection 4.3. The Hausdorff distance between $\bigcup_i D_i$ and $\mathcal{M}_o$ is less than $C\delta$ by the quantitative implicit function theorem (Subsection B), applying Taylor's theorem together with (251) and (254). Thus, by the triangle inequality, the Hausdorff distance between $\mathcal{M}_o$ and $\mathcal{M}$ is less than Cr^2. □

Next, we address the reach of $\mathcal{M}_o$. By Federer's criterion (Proposition 2.1) we know that

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

Let $a, b \in \mathcal{M}_o, a \neq b$.

If $|a-b| > \frac{1}{Cd^6}$, then $|b-a|^2(2dist(b, Tan(a)))^{-1} > \frac{1}{Cd^6}$, because

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

Therefore, we may suppose that $|a-b| \leq \frac{1}{Cd^6}$. By the bound on the Hausdorff distance between $\mathcal{M}$ and $\mathcal{M}_o$, the distances of a and b to their projections onto $\mathcal{M}$, which we denote a' and b' respectively, are less than $C\delta$. By the quantitative implicit function theorem (Subsection 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 (165), (251) and (254) the Hausdorff distance between the two unit discs $Tan(a) \cap B(a, r)$ and $(Tan_{\mathcal{M}}(a') \cap B(a', r)) + (a - a')$ which are centered at a, is bounded above by $Cd^3\delta$. Therefore, 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 $\frac{Cd^3\delta}{r}$.

Then, $\mathcal{M}_o$ and $\mathcal{M}$ are δ close in Hausdorff distance and $(\mathcal{M}_o \cap B(a, 2|a-b|))$ and $(\mathcal{M} \cap B(a', 2|a-b|))$ are $C(d^3\delta/r)$ close in C^1 as graphs of functions over $(Tan(a) \cap B(a, 3|a-b|/2))$. Let these functions be respectively $\widehat{f}_o$ and $\widehat{f}$. Note that the range is $Nor(a)$, the fiber of the normal bundle at a;

please see Lemma A.2 in Section 2. We know that the C^1 norm of $\widehat{f}$ on $(Tan(a) \cap B(a, 3|a-b|/2))$ is at most $\frac{Cd^3\delta}{r} + C|a-b|$. Therefore, the C^1 norm of $\widehat{f_o}$ on $(Tan(a) \cap B(a, 3|a-b|/2))$ is at most $\frac{Cd^3\delta}{r} + C|a-b|$. But using this and the Hessian bound of Cd^6 from (183), we also know that the Hessian of $\widehat{f_o}$ is bounded above by Cd^6. But now, by Taylor's theorem, $dist(b, Tan(a)) \leq \sup \|Hess\widehat{f_o}\||a-b|^2/2$, where the supremum is taken over $(Tan(a) \cap B(a, 3/2|a-b|))$. This, we know is bounded above by $Cd^6|a-b|^2$. Substituting this into Federer's criterion for the reach, we see that $reach(\mathcal{M}_o) \geq \frac{1}{Cd^6}$. Thus, we have just proved the following.

Theorem 7.6. *The reach of $\mathcal{M}_o$ is at least $\frac{1}{Cd^6}$.*

Finally, we provide an estimate on the third derivatives of $\mathcal{M}_o$. Since our guarantee about the true manifold is only that it is $\mathcal{C}^2$, it is inevitable that as the Hausdorff distance between the output manifold and the true manifold tends to zero, the guarantees on the third derivatives of $\mathcal{M}_o$ viewed as the graph of a function tend to infinity. The inverse dependence on σ in the following lemma reflects that fact.

Proposition 7.1. *Let $a \in \mathcal{M}_o$ and $\widehat{f_o}$ be a function from $(Tan(a) - a)$ to $Nor(a)$ such that $(\mathcal{M}_o - a)$ agrees with the graph of $\widehat{f_o}$ in a $\frac{1}{20}-$neighborhood U of a. Then, for any unit vector v in the domain, and any unit vector w in the range, the third derivative $\langle \partial_v^3 \widehat{f_o}, w\rangle$ at $x \in U$ satisfies*

$$\left|\langle \partial_v^3 \widehat{f_o}(x), w\rangle\right| \leq \frac{Cd^9}{r} \leq \frac{Cd^{8.5}}{\sigma}.$$

Proof. This follows from Lemma 7.4 and the quantitative implicit function theorem from Subsection B.3. □

Concluding remarks

We have studied the problem of reconstructing a compact embedded d dimensional C^2 submanifold $\mathcal{M}$ of $\mathbb{R}^n$ from random samples. These random samples are obtained from sampling the manifold independently and identically at random from some density μ and adding Gaussian noise having a spherically symmetric distribution where the standard deviation of any component is σ. In the present paper, we developed an algorithm that uses $O(\sigma^{-d-4})$ samples and produces a manifold $\mathcal{M}_o$ whose reach is no more than Cd^6 times the reach of $\mathcal{M}$ and whose Hausdorff distance to $\mathcal{M}$ is at most $\frac{Cd\sigma^2}{\tau}$.

Acknowledgements

Ch.F. was partly supported by the US-Israel Binational Science Foundation, grant number 2014055, AFOSR, grant DMS-1265524, and NSF, grant FA9550- 12-1-0425. S.I. was partly supported RFBR, grant 20-01-00070, 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

[1] Aamari, E., and Levrard, C. Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist. 47*, 1 (02 2019), 177–204.

[2] M. Anderson, *Convergence and rigidity of manifolds under Ricci curvature bounds*, Invent. Math. **102** (1990), 429–445.

[3] M. Anderson, A. Katsuda, Y. Kurylev, M. Lassas, M. Taylor, *Boundary regularity for the Ricci equation, geometric convergence, and Gel'fand's inverse boundary problem*, Invent. Math. **158** (2004), 261–321.

[4] Boissonnat, J., Guibas, L. J., and Oudot, S. Manifold reconstruction in arbitrary dimensions using witness complexes. *Discrete & Computational Geometry 42*, 1 (2009), 37–70.

[5] Boucheron, S., Bousquet, O. and Lugosi, G., Theory of classification : a survey of some recent advances, *ESAIM: Probability and Statistics*, 9 (2005), 323–375.

[6] M. Belkin, P. Niyogi, *Laplacian eigenmaps and spectral techniques for embedding and clustering*, Adv. in Neural Inform. Process. Systems, **14** (2001), 586–691.

[7] M. Belkin, P. Niyogi, *Semi-Supervised Learning on Riemannian Manifolds*, Machine Learning, **56** (2004), 209–239.

[8] M. Belkin, P. Niyogi, *Convergence of Laplacian eigenmaps*, Adv. in Neural Inform. Process. Systems **19** (2007), 129–136.

[9] V. Berestovskij, I. Nikolaev, *Multidimensional generalized Riemannian spaces*, In: Geometry IV, Encyclopaedia Math. Sci. **70**, Springer, 1993, pp. 165–243.

[10] E. Beretta, M. de Hoop, L. Qiu, *Lipschitz Stability of an Inverse Boundary Value Problem for a Schrödinger-Type Equation* SIAM J. Math. Anal. **45** (2012), 679-699.

[11] E. Bierstone, P. Milman, W. Paulucki, *Differentiable functions defined on closed sets. A problem of Whitney*, Invent. Math., **151** (2003), 329–352.

[12] J. Boissonnat, L. Guibas, S. Oudot, *Manifold reconstruction in arbitrary dimensions using witness complexes*, Discrete & Computational Geometry 42 (2009), 37–70.

[13] L. Borcea, V. Druskin, F. Guevara Vasquez, *Electrical impedance tomography with resistor networks* Inverse Problems **24** (2008), 035013.

[14] L. Borcea, V. Druskin, L. Knizhnerman, *On the continuum limit of a discrete inverse spectral problem on optimal finite difference grids*, Comm. Pure Appl. Math. **58** (2005), 1231-1279.

[15] M. Brand, *Charting a manifold, NIPS 15* (2002), 985–992.

[16] M. Bridson, A. Haefliger, *Metric spaces of non-positive curvature*, Springer-Verlag, 1999.

[17] S. Bromberg, *An extension in the class C^1,* Bol. Soc. Mat. Mex. II, Ser. **27**, (1982), 35–44.

[18] Y. Brudnyi, *On an extension theorem,* Funk. Anal. i Prilzhen. 4 (1970), 97–98; English transl. in Func. Anal. Appl. **4** (1970), 252–253.

[19] Y. Brudnyi, P. Shvartsman, *The traces of differentiable functions to closed subsets of $\mathbb{R}^n$,* in Function Spaces (1989), Teubner-Texte Math. **120**, 206–210.

[20] Y. Brudnyi, P. Shvartsman, *A linear extension operator for a space of smooth functions defined on closed subsets of $\mathbb{R}^n$,* Dokl. Akad. Nauk SSSR **280** (1985), 268–270. English transl. in Soviet Math. Dokl. **31**, No. 1 (1985), 48–51.

[21] Y. Brudnyi, P. Shvartsman, *Generalizations of Whitney's extension theorem,* Int. Math. Research Notices **3** (1994), 129–139.

[22] Y. Brudnyi, P. Shvartsman, *The traces of differentiable functions to closed subsets of $\mathbb{R}^n$*, Dokl. Akad. Nauk SSSR **289** (1985), 268–270.

[23] Y. Brudnyi, P. Shvartsman, *The Whitney problem of existence of a linear extension operator*, J. Geom. Anal. **7**(1997), 515–574.

[24] Y. Brudnyi, P. Shvartsman, *Whitney's extension problem for multivariate $C^{1,\omega}$ functions,* Trans. Amer. Math. Soc. **353** No. 6 (2001), 2487–2512.

[25] M. Bernstien, V. de Silva, J. Langford, J. Tenenbaum, *Graph approximations to geodesics on embedded manifolds*. Technical Report, Stanford University, 2000.

[26] D. Burago, S. Ivanov, Y. Kurylev, *A graph discretisation of the Laplace-Beltrami operator*, J. Spectr. Theory **4** (2014), 675–714.

[27] Chen, Y.-C., Genovese, C. R., and Wasserman, L. Asymptotic theory for density ridges. *Ann. Statist. 43*, 5 (10 2015), 1896–1928.

[28] Cheng, S., Dey, T. K., and Ramos, E. A. 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* (2005), pp. 1018–1027.

[29] D. Chigirev and W. Bialek, *Optimal Manifold Representation of Data: An Information Theoretic Approach*. In: Advances in Neural Information Processing Systems **16,** Ed. S. Thrun et al, The MIT press, 2004, pp. 164–168.

[30] R. Coifman, et al. *Geometric diffusions as a tool for harmonic analysis and structure definition of data* Part II: Multiscale methods. Proc. of Nat. Acad. Sci. **102** (2005), 7432–7438.

[31] R. Coifman, et al. R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker Geometric diffusions as a tool for harmonic analysis and structure definition of data Part II: Multiscale methods. *Proc. of Nat. Acad. Sci.* 102 (2005), 7432–7438.

[32] R. Coifman, S. Lafon, Diffusion maps. *Appl. Comp. Harm. Anal.* 21 (2006), 5-30.

[33] T. Cox, M. and Cox *Multidimensional Scaling.* Chapman & Hall, London, (1994).

[34] S. Dasgupta, Y. Freund, *Random projection trees and low dimensional manifolds.* In Proc. the 40th ACM symposium on Theory of computing (2008), STOC '08 537–546.

[35] D. Donoho, D. Grimes, *Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data,* Proceedings of the National Academy of Sciences, **100**, 5591–5596.

[36] D. Donoho, C. Grimes, When does geodesic distance recover the true hidden parametrization of families of articulated images? *Proceedings of ESANN 2002*, Bruges, Belgium, 2002.

[37] D. Donoho, C. Grimes, Image Manifolds which are Isometric to Euclidean Space. *J. Math. Im. Vis* 23 (2005), 5.

[38] Fan, J., and Truong, Y. K. Nonparametric regression with errors in variables. *Ann. Statist. 21*, 4 (12 1993), 1900–1925.

[39] Federer, H. Curvature measures. *Transactions of the American Mathematical Society 93* (1959).

[40] Federer, H. *Geometric measure theory*. Springer, 2014.

[41] Ch. Fefferman, *A sharp form of Whitney's extension theorem*, Ann. of Math. **161** (2005), 509–577.

[42] Ch. Fefferman, *Whitney's extension problem for C^m*, Ann. of Math. **164** (2006), 313–359.

[43] Ch. Fefferman, *C^m-extension by linear operators*, Ann. of Math. **166** (2007), 779–835.

[44] Ch. Fefferman, *A generalized sharp Whitney theorem for jets, Rev. Mat. Iberoam.* **21,** No. 2 (2005), 577–688.

[45] Ch. Fefferman, Extension of $C^{m,\omega}$ *smooth functions by linear operators,* Rev. Mat. Iberoam. **25,** No. 1 (2009), 1–48.

[46] Ch. Fefferman, B. Klartag, *Fitting C^m-smooth function to data I*, Ann. of Math, **169** (2009), 315–346.

[47] Ch. Fefferman, B. Klartag, *Fitting C^m-smooth function to data II*, Rev. Mat. Iberoam. **25** (2009), 49–273.

[48] Fefferman, C., Ivanov, S., Kurylev, Y., Lassas, M., and Narayanan, H. Fitting a putative manifold to noisy data. In *Proceedings of the 31st Conference On Learning Theory* (06–09 Jul 2018), S. Bubeck, V. Perchet, and P. Rigollet, Eds., vol. 75 of *Proceedings of Machine Learning Research*, PMLR, pp. 688–720.

[49] Fefferman, C., Ivanov, S., Kurylev, Y., Lassas, M., and Narayanan, H. Reconstruction and interpolation of manifolds I: The geometric whitney problem. *Foundations of Computational Mathematics*. Preprint arXiv:1508.00674.

[50] Fefferman, C., Ivanov, S., Lassas, M., and Narayanan, H. Reconstruction of a Riemannian Manifold from Noisy Intrinsic Distances SIAM Journal on Mathematics of Data Science 2020 2:3, 770-808

[51] Fefferman, C., Mitter, S., and Narayanan, H. Testing the manifold hypothesis. *Journal of the American Mathematical Society 29*, 4 (2016), 983–1049.

[52] K. Fukaya, *A boundary of the set of the Riemannian manifolds with bounded curvatures and diameters* J. Differential Geom. **28** (1988), No. 1, 1–21.

[53] Genovese, C. R., Perone-Pacifico, M., Verdinelli, I., and Wasserman, L. Manifold estimation and singular deconvolution under hausdorff loss. *Annals of Statistics 40*, 2 (2012).

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

[55] Genovese, C. R., Perone-Pacifico, M., Verdinelli, I., and Wasserman, L. Nonparametric ridge estimation. *Ann. Statist. 42*, 4 (2014), 1511–1545.

[56] G. Glaeser, *Etudes de quelques algebres Tayloriennes*, J. d'Analyse **6** (1958), 1–124.

[57] M. Gromov with appendices by M. Katz, P. Pansu, and S. Semmes, *Metric Structures for Riemannian and Non-Riemanian Spaces.* Birkhauser (1999).

[58] Hein, M., Maier, M. Manifold denoising. *In Advances in neural information processing systems* (pp. 561-568).

[59] H. Hotelling, Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24** (1933), 417–441, 498–520.

[60] E. Iversen, M. Tygel, B. Ursin, and M. V. de Hoop, *Kinematic time migration and demigration of re ections in pre-stack seismic data*, Geophys. J. Int. **189** (2012), 1635–1666.

[61] P. Jones, M., Maggioni, R. Schul, *Universal local parametrizations via heat kernels and eigenfunctions of the laplacian,* Ann. Acad. Scient. Fen. **35** (2010), 1–44.

[62] A. Katchalov, Y. Kurylev, M. Lassas: *Inverse Boundary Spectral Problems*, Monographs and Surveys in Pure and Applied Mathematics **123**, CRC-press, 2001, xi+290 pp.

[63] A. Katsuda, Y. Kurylev, M. Lassas, *Stability and Reconstruction in Gel'fand Inverse Boundary Spectral Problem*, in: New analytic and geometric methods in inverse problems. (Ed. K. Bingham, Y. Kurylev, and E. Somersalo), 309–320, Springer-Verlag, 2003.

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

[65] D. Perraul-Joncas, M. Meila, Non-linear dimensionality reduction: Riemannian metric estimation and the problem of geometric discovery, arXiv:1305-7255, 2013.

[66] R. Kress, *Numerical analysis*. Springer-Verlag, 1998. xii+326 pp.

[67] M. Lassas, G. Uhlmann, *Determining Riemannian manifold from boundary measurements,* Ann. Sci. École Norm. Sup. **34** (2001), 771–787.

[68] J. Lee, G. Uhlmann, *Determining anisotropic real-analytic conductivities by boundary measurements,* Comm. Pure Appl. Math. **42** (1989), 1097–1112.

[69] L. Ma, M. Crawford, J. W. Tian, *Generalised supervised local tangent space alignment for hyperspectral image classification*, Electronics Letters **46** (2010), 497.

[70] J. Mueller, S. Siltanen, *Linear and nonlinear inverse problems with practical applications*. SIAM, Philadelphia, 2012. xiv+351 pp.

[71] J. Nash, C^1*-isometric imbeddings*, Ann. of Math. **60** (1954), 383–396.

[72] J. Nash, *The imbedding problem for Riemannian manifolds*, Ann. of Math. **63** (1956), 20–63.

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

[74] Ozertem, U., and Erdogmus, D. Locally defined principal curves and surfaces. *Journal of Machine Learning Research 12* (2011), 1249–1286.

[75] G. Paternain, M. Salo, G. Uhlmann *Tensor Tomography on Simple Surfaces,* Inventiones Math. **193** (2013), 229–247.

[76] L. Pestov, G. Uhlmann, *Two Dimensional Compact Simple Riemannian manifolds are Boundary Distance Rigid,* Ann. of Math. **161** (2005), 1089–1106.

[77] K. Pearson, *On lines and planes of closest fit to systems of points in space,* Philosophical Magazine **2** (1901), 559–572.

[78] S. Peters, *Cheeger's finiteness theorem for diffeomorphism classes of Riemannian manifolds*, J. Reine Angew. Math. **349** (1984), 77–82.

[79] P. Petersen, *Riemannian geometry*. 2nd Ed. Springer, (2006), xvi+401 pp.

[80] G. Rosman, M. M. Bronstein, A. M. Bronstein, R. Kimmel, *Nonlinear Dimensionality Reduction by Topologically Constrained Isometric Embedding*, International Journal of Computer Vision, **89** (2010), 56–68.

[81] S. Roweis, L. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, Science, **290** (2000), 2323–326.

[82] S. Roweis, L. Saul, G. Hinton, *Global coordination of local linear models,* Advances in Neural Information Processing Systems **14** (2001) 889–896.

[83] V. Ryaben'kii, S. Tsynkov, *A Theoretical Introduction to Numerical Analysis*, CRC Press, 2006, 537 pp.

[84] T. Sakai, *Riemannian geometry.* AMS, (1996), xiv+358 pp.

[85] J. Shawe-Taylor, N. Christianini, *Kernel Methods for Pattern Analysis,* Cambridge University Press, (2004).

[86] P. Shvartsman, *Lipschitz selections of multivalued mappings and traces of the Zygmund class of functions to an arbitrary compact,* Dokl. Acad. Nauk SSSR **276** (1984), 559–562; English transl. in Soviet Math. Dokl. **29** (1984), 565–568.

[87] P. Shvartsman, *On traces of functions of Zygmund classes,* Sibirskyi Mathem. J. **28** (1987), 203–215; English transl. in Siberian Math. J. **28** (1987), 853–863.

[88] P. Shvartsman, *Lipschitz selections of set-valued functions and Helly's theorem,* J. Geom. Anal. **12** (2002), 289–324.

[89] Sober B. and Levin, D., Manifold Approximation by Moving Least-Squares Projection (MMLS), *Constructive Approximation*, 2019

[90] Tao, T. An epsilon of room, ii: pages from year three of a mathematical blog.

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

[92] N. Verma, Distance preserving embeddings for general n-dimensional manifolds. (aka An algorithmic realization of Nash's embedding theorem), *Journal of Machine Learning Research* 23 (2012) 32.1–32.28.

[93] Vaidya, P. M. A new algorithm for minimizing convex functions over convex sets. *Mathematical Programming 73*, 3 (1996), 291–341.

[94] Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[95] K. Weinberger, L. Saul, *Unsupervised learning of image manifolds by semidefinite programming,* Int. J. Comput. Vision **70**, 1 (2006), 77–90.

[96] H. Whitney, *Analytic extensions of differentiable functions defined on closed sets*, Trans. Amer. Math. Soc., **36** (1934), 63–89.

[97] H. Whitney, *Differentiable functions defined in closed sets I,* Trans. Amer. Math. Soc. **36** (1934), 369–389.

[98] H. Whitney, *Functions differentiable on the boundaries of regions,* Ann. of Math. **35** (1934), 482–485.

[99] H. Whitney, *Differentiable manifolds*, Ann. of Math. **37** (1936), 645–680.

[100] H. Whitney, J. Eells, D. Toledo, eds., *The collected papers of Hassler Whitney*. Volumes I-II., Contemporary Mathematicians, Birkhäuser, (1992).

[101] H. Zha and Z. Zhang, Continuum Isomap for manifold learnings. Comp. Stat. Data Anal. 52 (2007), 184-200.

[102] Z. Zhang and H. Zha, *Principal manifolds and nonlinear dimension reduction via local tangent space alignment*, SIAM J. Sci. Computing, **26** (2005), 313–338.

[103] N. Zobin, *Whitney's problem on extendability of functions and an intrinsic metric*, Advances in Math. **133** (1998), 96–132.

[104] N. Zobin, *Extension of smooth functions from finitely connected planar domains*, J. Geom. Anal. **9** (1999), 489–509.

A Some basic lemmas

Lemma A.1. *Suppose that* $\mathcal{M} \in \mathcal{G}(d, m, V, \tau)$. *Let*

$$U := \{y \in \mathbb{R}^m \big| |y - \Pi_x y| \le \tau/4\} \cap \{y \in \mathbb{R}^m \big| |x - \Pi_x y| \le \tau/4\}.$$

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 closed unit ball in $\mathbb{R}^d$. Suppose otherwise, then let $\emptyset \ne Y := B_d \setminus \Pi_0(\mathcal{N})$. 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 $B_d \setminus \Pi_0(\mathcal{N})$. In order to prove this lemma, it suffices to show

$$|x_1| \ge \frac{1}{2}. \tag{202}$$

Since $Y \ne \emptyset$ and $B_d \setminus \Pi_0(\mathcal{N})$ is open relative to B_d,

$$|x_1| < 1. \tag{203}$$

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, (i. e. Corollary 2.2) $\forall y_1 \in \Pi_0^{-1}(x_1) \cap \mathcal{N}$,

$$dist(y_1, Tan(0)) \leq \frac{\|y_1\|^2}{4}. \tag{204}$$

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

$$dist(y_1, \mathbb{R}^d) \leq \frac{dist(y_1, \mathbb{R}^d)^2 + |x_1|^2}{4}. \tag{205}$$

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

$$|y_1 - x_1|/2 \ \ \leq \ \ 1 - \sqrt{1 - \left(\frac{|x_1|}{2}\right)^2} \leq \left(\frac{|x_1|}{2}\right)^2. \tag{206}$$

This implies that

$$|y_1 - x_1| \leq \frac{1}{8} < \frac{\tau}{8}. \tag{207}$$

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

$$|z - \Pi_{y_1}(z)|/2 \ \ \leq \ \ 1 - \sqrt{1 - \left(\frac{|y_1 - \Pi_{y_1} z|}{2}\right)^2} \leq \left(\frac{|y_1 - \Pi_{y_1} z|}{2}\right)^2. \tag{208}$$

By (203), there exists no neighborhood $V \subseteq \mathbb{R}^m$ of y_1, such that there exists an open set $U_0 \subseteq \mathbb{R}^d$ containing x_1 and a $\mathcal{C}^2$ function $F : U_0 \to \mathbb{R}^{m-d}$ with $DF(u)$ of rank d for all $u \in U_0$ such that

$$\mathcal{N} \cap V = \{(u, F(u))\big|u \in U \cap \mathbb{R}^d\}. \tag{209}$$

Therefore, we have the following.

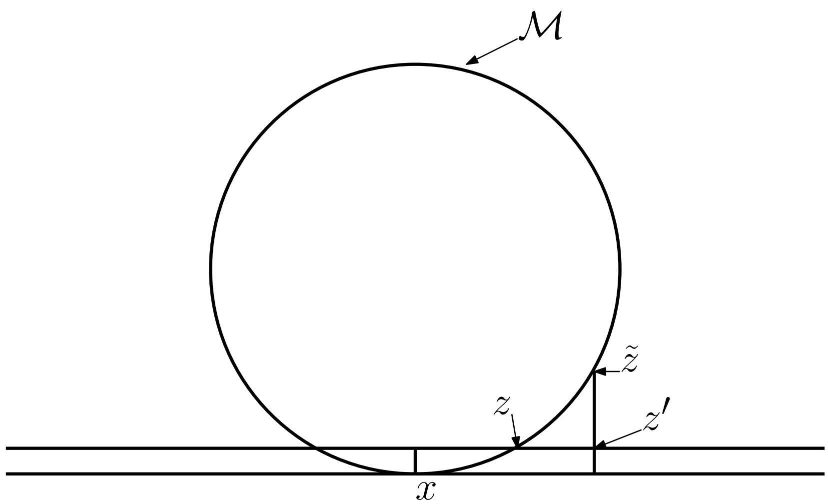
Claim A.1. *Let* $y_1 \in \Pi_0^{-1}(x_1) \cap \mathcal{N}$. *Then there exists* $v \in \partial B_d$ *such that if* $y_1' \in Tan(y_1)$ *then* $\langle y_1' - y_1, v\rangle = 0$.

The cause behind the preceding claim is that the only way $\mathcal{N}$ is locally not the graph of a function over a neighborhood contained in the interior of B_d, is if such a vector v exists.

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 A.1, $\Pi_\ell(Tan(y_1))$ 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 $y_2 \in \Pi_0^{-1}(x_2) \cap \mathcal{N}$. Note that $(y_2 - x_2)$ and $(x_1 - y_1)$ are both



orthogonal to $(x_2 - x_1)$, which will be used to obtain (213) below. Then,

$$
\begin{aligned}
|x_1| &\le |\Pi_\ell(y_1) - x_2| && (210)\\
&\le dist(y_2, Tan(y_1)) && (211)\\
&\le \frac{|y_2 - y_1|^2}{4} && (212)\\
&\le \frac{2|y_2 - x_2|^2 + |x_1 - x_2|^2 + 2|y_1 - x_1|^2}{4} && (213)\\
&\le \frac{2\left(\frac{|x_2|^2}{2}\right)^2 + 4|x_1|^2 + 2\left(\frac{|x_1|^2}{2}\right)^2}{4}. && (214)
\end{aligned}
$$

Therefore,

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

Therefore, $1 \le \alpha^3/4 + \alpha$. This implies that $|x_1| > \frac{1}{2}$, which proves Lemma A.1 (see (202)).

□

Lemma A.2. *Suppose that* $\mathcal{M} \in \mathcal{G}(d, m, V, \tau)$. *Let* $x \in \mathcal{M}$ *and*

$$\widehat{U} := \{y \in \mathbb{R}^m \big| |y - \Pi_x y| \le \tau/8\} \cap \{y \in \mathbb{R}^m \big| |x - \Pi_x y| \le \tau/8\}.$$

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

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

Secondly, for $\delta \le \tau/8$, *let* $z \in \mathcal{M} \cap \widehat{U}$ *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* $m - d$ *canonical basis vectors be denoted* $\mathbb{R}^{m-d}$. *In this coordinate frame,* *let a point $z' \in \mathbb{R}^m$ be represented as (z'_1, z'_2), where $z'_1 \in \mathbb{R}^d$ and $z'_2 \in \mathbb{R}^{m-d}$. By Lemma A.1, there exists an $(m-d) \times d$ matrix A_z such that*

$$Tan(z) = \{(z'_1, z'_2)|A_z z'_1 - I z'_2 = 0\} \tag{215}$$

where the identity matrix is $(m-d)\times(m-d)$. Let $z \in \mathcal{M} \cap \{z\big||z - \Pi_x z| \le \delta\} \cap \{z\big||x - \Pi_x z| \le \delta\}$. Then $\|A_z\|_2 \le 15\delta/\tau$. Lastly, the following upper bound on the second derivative of $F_{x,\widehat{U}}$ holds for $y \in \Pi_x(\widehat{U})$.

$$\forall_{v\in\mathbb{R}^d}\forall_{w\in\mathbb{R}^{m-d}} \;\; \langle \partial_v^2 F_{x,\widehat{U}}(y), w\rangle \le \frac{C|v|^2|w|}{\tau}.$$

Proof. We will first show that there exists a function $F_{x,\widehat{U}}$ that satisfies the given conditions and then show that it is C^2. Let $z \in \mathcal{M} \cap \widehat{U}$ 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 $m-d$ canonical basis vectors be denoted $\mathbb{R}^{m-d}$. In this coordinate frame, let a point $z' \in \mathbb{R}^m$ be represented as (z'_1, z'_2), where $z'_1 \in \mathbb{R}^d$ and $z'_2 \in \mathbb{R}^{m-d}$. By Lemma A.1, there exists a matrix A such that

$$Tan(z) = \{(z'_1, z'_2)|A z'_1 - I z'_2 = 0\}. \tag{216}$$

Further, a linear algebraic calculation shows that

$$dist(z', Tan(z)) = \left|(I + AA^T)^{-1/2}(Az'_1 - Iz'_2)\right|. \tag{217}$$

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

$$|\widetilde{z} - \Pi_x \widetilde{z}| \le \frac{|x - \Pi_x \widetilde{z}|^2}{\tau} \le \frac{4\delta^2}{\tau}. \tag{218}$$

The last inequality holds because

$$|x - \Pi_x \widetilde{z}| \le |x - \Pi_x z| + |\Pi_x z - \Pi_x \widetilde{z}| = 2\delta.$$

Therefore, denoting x by (x_1, x_2), where $x_1 \in \mathbb{R}^d$ and $x_2 \in \mathbb{R}^{m-d}$, we have $|x_2| \le \frac{|x_1|^2}{\tau} = \frac{\delta^2}{\tau}$, and so

$$|\widetilde{z} - z'| = |\widetilde{z} - ((\Pi_x \widetilde{z}) - x_2)| \le \frac{4\delta^2}{\tau} + \frac{\delta^2}{\tau} = \frac{5\delta^2}{\tau}. \tag{219}$$

Therefore,

$$\begin{aligned} dist(z', Tan(z)) &\le dist(\widetilde{z}, Tan(z)) + |\widetilde{z} - z'| && (220)\\ &\le \frac{|z - \widetilde{z}|^2}{\tau} + \frac{5\delta^2}{\tau} && (221)\\ &= \frac{|z - z'|^2 + |z' - \widetilde{z}|^2}{\tau} + \frac{5\delta^2}{\tau} && (222)\\ &\le \frac{\delta^2 + (5\delta^2/\tau)^2}{\tau} + \frac{5\delta^2}{\tau}. && (223) \end{aligned}$$

Therefore, for any $z'_1 \in S_\delta^{d-1}$,

$$\left|(I + AA^T)^{-1/2}(Az'_1)\right| \leq \frac{\delta^2}{\tau}\left(6 + \frac{25\delta^2}{\tau^2}\right) \leq \frac{\delta^2}{\tau}\left(6 + \frac{64\delta^2}{\tau^2}\right). \tag{224}$$

Thus,

$$\left\|(I + AA^T)^{-1/2}A\right\|_2 \leq \frac{\delta}{\tau}\left(6 + \frac{64\delta^2}{\tau^2}\right) =: \delta'. \tag{225}$$

Therefore, we see that

$$\left\|(I + AA^T)^{-1/2}AA^T(I + AA^T)^{-1/2}\right\|_2 \leq \delta'^2. \tag{226}$$

Let $\|A\|_2 = \lambda$. We then see that λ^2 is an eigenvalue of AA^T. Therefore, $\frac{\lambda^2}{1+\lambda^2} \leq \delta'^2$. This gives us $\lambda^2 \leq \frac{\delta'^2}{1-\delta'^2}$, which implies that

$$\lambda \leq \frac{\delta'}{\sqrt{1-\delta'^2}}. \tag{227}$$

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 $\widehat{z} \neq z$ that also belongs to $\Pi_x^{-1}(\Pi_x z) \cap \mathcal{M} \cap U$. Then,

$$dist(\widehat{z}, Tan(z)) \leq |\widehat{z} - z| \leq \frac{|\widehat{z}_2|^2}{\tau}, \tag{228}$$

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

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

Therefore,

$$1 \leq |\widehat{z}_2|(1/\sqrt{(1-\delta'^2)})/\tau \leq \frac{2\delta^2}{\tau^2\sqrt{1-\delta'^2}}. \tag{229}$$

Hence,

$$1 - \delta'^2 \leq \frac{2\delta^2}{\tau^2}, \tag{230}$$

and so $\delta' \geq 1 - 2\delta^2/\tau^2$. Assuming $\delta \leq \frac{\tau}{8}$ we infer from (225) that

$$\delta' \leq \frac{7\delta}{\tau}. \tag{231}$$

Therefore

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

This implies that $\delta/\tau > 1/8$. This is a contradiction. This proves that $\Pi_x^{-1}(\Pi_x z) \cap \mathcal{M} \cap U$ contains the single point z. Further, (227) and (231) together imply that $\|A_z\|_2 \leq \frac{56\delta}{\sqrt{15}\tau} < \frac{15\delta}{\tau}$. Since $\mathcal{M}$ is a C^2 submanifold of $\mathbb{R}^n$, by (216), $F_{x,\widehat{U}}$ is C^2. For $y \in \Pi_x(\widehat{U})$, we shall obtain the following upper bound on the second derivative

$$\forall_{v\in\mathbb{R}^d}\forall_{w\in\mathbb{R}^{n-d}} \ \langle \partial_v^2 F_{x,\widehat{U}}(y), w\rangle \leq \frac{C|v|^2|w|}{\tau}$$

below. Given $v \in \mathbb{R}^d$, let $z = y + F_{x,\widehat{U}}(y)$.

Let $\widetilde{z} = (y + \epsilon v, F_{x,\widehat{U}}(y+\epsilon v))$. Then,

$$\begin{aligned}
&\left|F_{x,\widehat{U}}(y+\epsilon v) - F_{x,\widehat{U}}(y) - \epsilon A_z v\right|\epsilon^{-2} \\
&\quad= \left|(y+\epsilon v, F_{x,\widehat{U}}(y+\epsilon v)) - (y+\epsilon v, F_{x,\widehat{U}}(y) + \epsilon A_z v)\right|\epsilon^{-2} \\
&\quad= \|I + A_z A_z^T\|_2^{1/2} |dist(\widetilde{z}, Tan(z))|\epsilon^{-2} \\
&\quad\leq \|I + A_z A_z^T\|_2^{1/2} \left(\frac{|z-\widetilde{z}|^2}{2\tau}\right)\epsilon^{-2} \\
&\quad\leq \|I + A_z A_z^T\|_2^{1/2} \left(\frac{\|A_z\|^2|v|^2}{2\tau}\right) \leq \frac{C|v|^2}{\tau}.
\end{aligned}$$

This yields

$$\partial_v^2 F_{x,\widehat{U}}(y) = \lim_{\epsilon\to 0}\left(\frac{F_{x,\widehat{U}}(y+\epsilon v) - F_{x,\widehat{U}}(y) - \epsilon A_z v}{\epsilon^2}\right). \tag{232}$$

Therefore,

$$\forall_{v\in\mathbb{R}^d} \left|\partial_v^2 F_{x,\widehat{U}}(y)\right| \leq \frac{C|v|^2}{\tau},$$

implying by Cauchy-Schwartz that

$$\forall_{v\in\mathbb{R}^d}\forall_{w\in\mathbb{R}^{m-d}} \ \langle \partial_v^2 F_{x,\widehat{U}}(y), w\rangle \leq \frac{C|v|^2|w|}{\tau}.$$

□

B Quantitative implicit and inverse function theorems

In this subsection, we provide for the reader's convenience, versions of the implicit and inverse function theorems with quantitative bounds on the derivatives that do not depend on the dimensions involved. We think it is very likely that such theorems exist in the literature, but are not aware of a specific reference.

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 $x \in B_p(0,1)$,

$$\|Jac_g - I\| \leq \epsilon_1/4 \tag{233}$$

for some $\epsilon_1 \in [0,1]$.

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

$$\left\|\frac{\partial^2 g(x)}{\partial v^2}\right\| \leq \left(\frac{\epsilon_2}{4}\right)|v|^2. \tag{234}$$

By (233), for any $x \neq x'$, both belonging to $B_p(0,1)$,

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

which implies that $g(x) \neq g(x')$. Applying the Inverse Function Theorem ([73]), 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,

$$\left(\left(\frac{df_i}{dg_j}\right)_{i,j\in[p]}\right) Jac_g = I, \tag{235}$$

implying by (233) that

$$\left\|\left(\left(\frac{df_i}{dg_j}\right)_{i,j\in[p]}\right)\right\| \leq (1 - \epsilon_1/4)^{-1}. \tag{236}$$

The second derivative of a linear function is 0 and so

$$0 = \frac{\partial^2 \hat{F}(g)}{\partial v^2}(x) = \sum_{i,j} \frac{d^2\hat{F}}{dg_i dg_j}\left(\frac{dg_i}{dv}\right)\left(\frac{dg_j}{dv}\right) + \sum_j \frac{d\hat{F}}{dg_j}\left(\frac{d^2 g_j}{dv^2}\right). \tag{237}$$

Therefore,

$$\sum_{i,j} \frac{d^2\hat{F}}{dg_i dg_j}\left(\frac{dg_i}{dv}\right)\left(\frac{dg_j}{dv}\right) = (-1)\sum_j \frac{d\hat{F}}{dg_j}\left(\frac{d^2 g_j}{dv^2}\right), \tag{238}$$

and so by Cauchy-Schwartz inequality,

$$\left|\sum_{i,j} \frac{d^2\hat{F}}{dg_i dg_j}\left(\frac{dg_i}{dv}\right)\left(\frac{dg_j}{dv}\right)\right| \leq \left\|\left(\left(\frac{d\hat{F}}{dg_j}\right)_{j\in[p]}\right)\right\| \left\|\left(\left(\frac{d^2 g_j}{dv^2}\right)_{j\in[p]}\right)\right\|. \tag{239}$$

By (233) there exists a unit vector $\widetilde{v}$ such that

$$\left|\sum_{i,j} \frac{d^2\hat{F}}{dg_i dg_j}\left(\frac{dg_i}{d\widetilde{v}}\right)\left(\frac{dg_j}{d\widetilde{v}}\right)\right| = \left\|Hess\,\widehat{F}\right\|\left\|\frac{dg}{d\widetilde{v}}\right\|^2 \tag{240}$$

$$\geq \left\|Hess\,\widehat{F}\right\| \inf_{\|v\|=1}\left\|\frac{dg}{dv}\right\|^2. \tag{241}$$

Together (234), (236), (239) and (240) imply that

$$\left\|Hess\,\widehat{F}\right\| \inf_{\|v\|=1}\left\|\frac{dg}{dv}\right\|^2 \leq \left\|\left(\left(\frac{df_i}{dg_j}\right)_{i,j\in[p]}\right) w\right\| \sup_{\|v\|=1}\left(\frac{\epsilon_2}{4}\right)\|v\|^2 \leq \left(\frac{\epsilon_2}{4-\epsilon_1}\right)\|w\|.$$

It follows that

$$\left\|Hess\,\widehat{F}\right\| \leq \left(\frac{\epsilon_2}{4-\epsilon_1}\right)\|w\| \sup_{\|v\|=1}\left\|\frac{dg}{dv}\right\|^{-2} \leq \left(\frac{16\epsilon_2}{(4-\epsilon_1)^3}\right)\|w\|. \tag{242}$$

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

$$\|Jac_g - I\| < \epsilon_1/4$$

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

$$\left\|\frac{\partial^2 g(x)}{\partial v^2}\right\| \leq \left(\frac{\epsilon_1}{4}\right)\|v\|^2$$

where $\epsilon_0, \epsilon_1, \epsilon_2 \in [0,1]$. Suppose also that $\|g(0)\| < \frac{\epsilon_0}{20}$.

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

$$\left\|Hess\,\widehat{F}\right\| \leq \frac{16\epsilon_2}{(4-\epsilon_1)^3}. \tag{243}$$

Lemma B.1. *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 $\|x\| \leq \frac{\eta}{2}$, where $\eta \in [0, 1]$,

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

and

$$\|(x, e(x,0))\| \leq \frac{8}{5}\left(\frac{\epsilon_0}{20} + \frac{\eta}{2}\right).$$

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

$$\|Hess(e \cdot w)\| \leq \frac{16\epsilon_2}{(4 - \epsilon_1)^3}. \tag{244}$$

Proof. It suffices to prove that if $z = (x, y) \in \mathbb{R}^p$ and $\|z\| \leq \eta/2$, where $\eta \in [0, 1]$, then there exists a point $\widehat{z}$, where $\|\widehat{z}\| \leq \frac{8}{5}\left(\frac{\epsilon_0}{20} + \frac{\eta}{2}\right)$, such that $g(\widehat{z}) = z$. We will achieve this by analysing Newton's method for finding a sequence $\widehat{z}_0, \dots, \widehat{z}_k, \dots$ converging to a point $\widehat{z}$ that satisfies $g(\widehat{z}) = z$. We will start with $\widehat{z}_0 = 0$.

The iterations of Newton's method proceed as follows.

For $i \geq 0$,

$$\widehat{z}_{i+1} = \widehat{z}_i - J_g^{-1}(\widehat{z}_i)(g(\widehat{z}_i) - z). \tag{245}$$

Claim B.1. *For any $i \geq 0$, $\|\widehat{z}_i\| \leq \frac{8}{5}\left(\frac{\epsilon_0}{20} + \frac{\eta}{2}\right)$.*

Proof. Observe that

$$\|\widehat{z}_{i+1} - \widehat{z}_i\| = \|J_g^{-1}(\widehat{z}_i)(g(\widehat{z}_i) - z)\|. \tag{246}$$

For $i = 0$,

$$\|g(\widehat{z}_i) - z\| \leq \frac{\epsilon_0}{20} + \frac{\eta}{2}. \tag{247}$$

and since $\|J_g^{-1}(\widehat{z}_i)\| \leq \frac{1}{1-\epsilon_1/4} \leq 4/3$, therefore

$$\|\widehat{z}_{i+1} - \widehat{z}_i\| \leq \left(\frac{4}{3}\right)\left(\frac{\epsilon_0}{20} + \frac{\eta}{2}\right). \tag{248}$$

Suppose $i \geq 1$.

$$g(\widehat{z}_i) - z = g\left(\widehat{z}_{i-1} - J_g^{-1}(\widehat{z}_{i-1})(g(\widehat{z}_{i-1}) - z)\right) - z. \tag{249}$$

Using the integral form of the remainder in Taylor's theorem, the right hand side of (249) equals

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

which simplifies to Λ, where

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

The norm of Λ is bounded above as follows. Note that by the induction hypothesis, $\|\widehat{z}_i\| \leq \frac{8}{5}\left(\frac{\epsilon_0}{20}+\frac{\eta}{2}\right)$, and $\|\widehat{z}_{i-1}\| \leq \frac{8}{5}\left(\frac{\epsilon_0}{20}+\frac{\eta}{2}\right)$, which places both $\widehat{z}_i$ and $\widehat{z}_{i-1}$ within the unit ball. Therefore $\|(\widehat{z}_i - \widehat{z}_{i-1})^T Hess_g(\widehat{z}_{i-1} + t(\widehat{z}_i - \widehat{z}_{i-1}))(\widehat{z}_i - \widehat{z}_{i-1})\| \leq (\epsilon_2/4)\|\widehat{z}_i - \widehat{z}_{i-1}\|^2$ for any $t \in [0,1]$. Moreover,

$$\|\Lambda\| \leq \int_0^1 (1-t)\|(\widehat{z}_i - \widehat{z}_{i-1})\|^2(\epsilon_2/4)dt = \left(\frac{\epsilon_2}{8}\right)\|(\widehat{z}_i - \widehat{z}_{i-1})\|^2.$$

Therefore

$$\|\widehat{z}_{i+1} - \widehat{z}_i\| = \|J_g^{-1}(\widehat{z}_i)(g(\widehat{z}_i) - z)\| \leq$$

$$\left(\frac{4}{3}\right)\left(\frac{\epsilon_2}{8}\right)\|(\widehat{z}_i - \widehat{z}_{i-1})\|^2 = \left(\frac{\epsilon_2}{6}\right)\|(\widehat{z}_i - \widehat{z}_{i-1})\|^2.$$

By recursion,

$$\|\widehat{z}_{i+1} - \widehat{z}_i\| \leq \left(\frac{\epsilon_2^{2i}}{6^i}\right)\|\widehat{z}_1 - \widehat{z}_0\|^{2^i}. \tag{250}$$

Therefore,

$$\|\widehat{z}_{i+1}\| = \|\widehat{z}_{i+1} - \widehat{z}_0\| \leq \sum_{j=1}^{i}\|\widehat{z}_{j+1} - \widehat{z}_j\| \leq$$

$$\frac{\|\widehat{z}_1 - \widehat{z}_0\|}{1-\frac{\epsilon_2}{6}} \leq \left(\frac{4}{3}\left(\frac{\epsilon_0}{20}+\frac{\eta}{2}\right)\right)\left(\frac{6}{5}\right) = \frac{8}{5}\left(\frac{\epsilon_0}{20}+\frac{\eta}{2}\right).$$

□

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 (248) and (250) $(\widehat{z}_0, \dots, \widehat{z}_i, \dots)$ is a Cauchy sequence, and therefore has a unique limit point. By the preceding Claim, this limit $\widehat{z}$ satisfies $\|\widehat{z}\| \leq \frac{22}{25} < 1$. 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,

$$\|f((x,0))\| \leq \frac{8}{5}\left(\frac{\epsilon_0}{20}+\frac{\eta}{2}\right).$$

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

$$\|f((0,0))\| \leq \frac{2\epsilon_0}{25}. \tag{251}$$

By (236) the function e satisfies, for $\|x\| \leq 1/2$,

$$\begin{aligned} \|D_x e\|^2 &= \|D_x f\|^2 - 1 && (252)\\ &\leq (1-\epsilon_1/4)^{-2} - 1 && (253)\\ &\leq \epsilon_1. && (254) \end{aligned}$$

By (243) the function e satisfies, for any $w \in \mathbb{R}^n$ such that $\|w\| = 1$,

$$\|Hess(e \cdot w)\| \leq \frac{16\epsilon_2}{(4-\epsilon_1)^3}. \tag{255}$$

□

We next obtain bounds for the $m^{th}-$order derivatives. Our focus will be in the case of $m \geq 2$.

In the remainder of this section, all norms on Euclidean spaces $\mathbb{R}^N, \mathbb{R}^M, \mathbb{R}^D, \dots$ are Euclidean norms.

$$|(v_1, \dots, v_N)| = \left(\sum_1^N v_i^2\right)^{\frac{1}{2}}.$$

Note that the usual implicit function theorem gives the function $\psi(x,z)$ that solves the equation $F(x, \psi(x,z)) = z$ at the end of this section. The purpose of this section is to derive bounds for the derivatives of ψ in terms of the derivatives of F.

B.1 Differentiating composed maps

Let $y = (y_1, \dots, y_M) = \Phi(x_1, \dots, x_N)$.

$$z = G(y) = G \circ \Phi(x).$$

Let $v_1, \dots, v_m$ be vectors in $\mathbb{R}^N$. Let ∂_v denote the directional derivative in the direction v. Then, $\partial_{v_1} \dots \partial_{v_m}(G \circ \Phi)(x)$ is a sum of terms

$$\sum_{p_1,\dots,p_{\nu_{max}}} \prod_{\nu=1}^{\nu_{max}} (\partial_{w_{1,\nu}} \dots \partial_{w_{s_\nu,\nu}} y_{p_\nu})(x) \cdot (\partial_{y_{p_1}} \dots \partial_{y_{p_{\nu_{max}}}} G(y))\big|_{y=\Phi(x)},$$

where each $s_\nu \geq 1$, and the list

$$w_{1,1}, \dots w_{s_1,1}, w_{1,2}, \dots, w_{s_2,2}, \dots, w_{1,\nu_{max}}, \dots, w_{s_{\nu_{max}},\nu_{max}},$$

may be permuted into the list $v_1, \dots, v_m$. This follows by induction on m.

So

$$\partial_{v_1} \dots \partial_{v_m}(G \circ \Phi),$$

is a sum of terms

(A) $\partial_{\zeta_1} \dots \partial_{\zeta_{\nu_{max}}} G(y)\big|_{y=\Phi(x)}$, where $\zeta_\nu \in \mathbb{R}^M$ is the vector whose $p^{th}-$coordinate is $\zeta_{\nu,p} = \partial_{w_{1,\nu}} \dots \partial_{w_{s_\nu,\nu}} y_p$. That is $\zeta_\nu = \partial_{w_{1,\nu}} \dots \partial_{w_{s_\nu,\nu}} y$.

where, as before, the concatenated list of all the $w's$ may be permuted into the list of $v's$.
The only term of the form (A) in which y is differentiated m times is

$$\partial_{[\partial_{v_1}\dots\partial_{v_m} y]} G(y) = \sum_p (\partial_{v_1} \dots \partial_{v_m} y_p)(\frac{\partial G}{\partial y_p}(y))\big|_{y=\Phi(x)}.$$

Suppose we know that

(*1) $|\partial_{w_1} \dots \partial_{w_s} y| \leq C_s$ for $s < m$, whenever $|w_1|, \dots, |w_s| \leq 1$,

(*2) $|\partial_{\zeta_1} \dots \partial_{\zeta_{\nu_{max}}} G| \leq C_m^*$ for $\nu_{max} \leq m$, all $|\zeta_\nu| \leq 1$.

Then,

$$|\partial_{\zeta_1} \dots \partial_{\zeta_{\nu_{max}}} G| \leq C_m^* |\zeta_1| \dots |\zeta_{\nu_{max}}|$$

for any $\zeta's$ provided $\nu_{max} \leq m$.

Then $(*1)$, $(*2)$ and our discussion of $\partial_{v_1} \dots \partial_{v_m}(G \circ \Phi)$ (see (A)) together imply that (for $|v_1|, \dots, |v_m| \leq 1$) together imply that (for $|v_1|, \dots, |v_m| \leq 1$)

$$\partial_{v_1} \dots \partial_{v_m}(G \circ \Phi)(x) = \sum_p (\partial_{v_1} \dots \partial_{v_m} y_p) \left(\frac{\partial G}{\partial y_p}\right)\bigg|_{y=\Phi(x)} + \delta_0,$$

where $|\delta_0|$ is less or equal to a constant determined by the C_s and C_m^*. We write $\overline{C}_m$ to denote any such constant. Write

$$G(x) = (G_1(x), \dots, G_M(x)).$$

Then,

$$\partial_{v_1} \dots \partial_{v_m}(G_q \circ \Phi)(x) = \sum_p (\partial_{v_1} \dots \partial_{v_m} y_p) \left(\frac{\partial G_q}{\partial y_p}\right)\bigg|_{y=\Phi(x)} + |\overline{\delta}_q|,$$

where $(\sum_q |\delta_q|^2)^{\frac{1}{2}} \leq \overline{C}_m$.

Let Ω_q^r be the $M \times M$ matrix that inverts the matrix $\frac{\partial G_q}{\partial y_p}\big|_{y=\Phi(x)}$. Assume that (Ω_q^r) has norm ≤ 10 (say) as a linear map from $\mathbb{R}^M$ to $\mathbb{R}^M$. Then, we find that

(*3) $\sum_q \Omega_q^r \partial_{v_1} \dots \partial v_m (G_q \circ \Phi)(x) = \sum_{p,q} (\partial_{v_1} \dots \partial_{v_m} y_p) \cdot \Omega_q^r \left(\frac{\partial G_q}{\partial y_p}\right)\bigg|_{y=\Phi(x)} + \widetilde{\delta}_r,$

which is equal to $\partial_{v_1} \dots \partial_{v_m} y_r(x) + \widetilde{\delta}_r$, where $(\sum_r |\widetilde{\delta}_r|^2)^{\frac{1}{2}} \leq \overline{C}_m$. Here is what that means:

Let $\Phi : \mathbb{R}^N \to \mathbb{R}^N$, and let G be the inverse function of Φ in some neighborhood.

Suppose we have bounds on

(!1) $|\partial_{v_1} \dots \partial_{v_k} G|$ for $k \leq m$ and $|v_1|, \dots, |v_k| \leq 1$.

(!2) $|\partial_{v_1} \dots \partial_{v_k} \Phi|$ for $k \leq m-1$ and $|v_1|, \dots, |v_k| \leq 1$.

Suppose the inverse of the Jacobian $\nabla_x G$ has norm ≤ 10 as a matrix (i. e. as a bounded linear operator on $\mathbb{R}^N$).

Then, we obtain bounds on

(!3) $|\partial_{v_1} \dots \partial_{v_k} \Phi|$ for $k = m$ and $|v_1|, \dots, |v_k| \leq 1$.

This holds for $m \geq 2$. Our bounds for (!3) depend only on m, and our bounds for (!1), (!2).

Starting from $m = 2$, we may now use induction on m to obtain the following result.

B.2 Quantitative Inverse Function Theorem.

Let $m \geq 2$, and let G, Φ be inverse images of each other in a neighborhood of a point in $\mathbb{R}^N$.

Suppose

- $|(\nabla_x G)^{-1}(v)| \leq 10|v|$ for all values $v \in \mathbb{R}^N$.
- $|\partial_{v_1} \dots \partial_{v_k} G| \leq C$ for $k \leq m$, $|v_1|, \dots, |v_k| \leq 1$.

Then, $|\partial_{v_1} \dots \partial_{v_k} \Phi| \leq C'$ for $k \leq m$, $|v_1|, \dots, |v_k| \leq 1$, where C' depends only on C and m. In particular, C' does not depend on N (unless C does).

Now for $x \in \mathbb{R}^N, y \in \mathbb{R}^D$, let $G(x,y)$ take values in $\mathbb{R}^D$.

We want to solve the equation

$$G(x,y) = z,$$

for the unknown y. Say the solution is $y = \Psi(x,z)$. So

$$G(x, \Psi(x,z)) = z.$$

Then, the following maps from $\mathbb{R}^N \times \mathbb{R}^D$ to itself are inverses of each other.

$$(x,y) \mapsto (x, G(x,y)),$$

$$(x,z) \mapsto (x, \Psi(x,z)).$$

Applying the quantitative inverse function theorem to these two maps, we obtain the following.

B.3 Quantitative Implicit Function Theorem

Let $m \geq 2$. Let

$$G(x,y) = (G_1(x,y), \dots, G_D(x,y)),$$

for $x = (x_1, \dots, x_N), y = (y_1, \dots, y_D)$.

Suppose $y = \Psi(x,z)$ solves the equation $G(x,y) = z$. Assume that

$$|\partial_{v_1} \dots \partial_{v_k} G| \leq C$$

for $k \leq m$, $v_1, \ldots, v_k \in \mathbb{R}^{N+D}$ of length ≤ 1. Assume that the inverse of the matrix

$$\left(\frac{\partial G_p}{\partial y_q} \right),$$

has norm at most 10 as a linear map from $\mathbb{R}^D$ to itself. Then also

$$|\partial_{v_1} \ldots \partial_{v_k} \Psi| \leq \overline{C},$$

for $k \leq m$, and $v_1, \ldots, v_k \in \mathbb{R}^{N+D}$, of length ≤ 1, where $\overline{C}$ is determined by C and m.