JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 29, Number 4, October 2016, Pages 983–1049
http://dx.doi.org/10.1090/jams/852
Article electronically published on February 9, 2016

TESTING THE MANIFOLD HYPOTHESIS

CHARLES FEFFERMAN, SANJOY MITTER, AND HARIHARAN NARAYANAN

Contents

Received by the editors March 9, 2014 and, in revised form, February 3, 2015 and August 9, 2015.

2010 *Mathematics Subject Classification.* Primary 62G08, 62H15; Secondary 55R10, 57R40.

The first author was supported by NSF grant DMS 1265524, AFOSR grant FA9550-12-1-0425 and U.S.-Israel Binational Science Foundation grant 2014055.

The second author was supported by NSF grant EECS-1135843.

1. Introduction

We are increasingly confronted with very high dimensional data from speech, images, genomes, and other sources. A collection of methodologies for analyzing high dimensional data based on the hypothesis that data tend to lie near a low dimensional manifold is now called "manifold learning" (see Figure 1). We refer to the underlying hypothesis as the "manifold hypothesis." Manifold learning, in particular, fitting low dimensional nonlinear manifolds to sampled data points in high dimensional spaces, has been an area of intense activity over the past two decades. These problems have been viewed as optimization problems generalizing the projection theorem in Hilbert space. We refer the interested reader to a limited set of papers associated with this field; see [3, 8, 9, 11, 16, 20, 26, 31, 32, 41, 43, 47, 50, 52, 55] and the references therein. Section 2 contains a brief review of manifold learning.

The goal of this paper is to develop an algorithm that tests the manifold hypothesis.

Examples of low dimensional manifolds embedded in high dimensional spaces include the following: image vectors representing three dimensional (3D) objects under different illumination conditions, and camera views and phonemes in speech signals. The low dimensional structure typically arises due to constraints arising from physical laws. A recent empirical study [9] of a large number of 3×3 images represented as points in $\mathbb{R}^9$ revealed that they approximately lie on a two dimensional manifold knows as the Klein bottle.

One of the characteristics of high dimensional data of the type mentioned earlier is that the number of dimensions is comparable to, or larger than, the number of samples. This has the consequence that the sample complexity of function approximation can grow exponentially. On the positive side, the data exhibit the phenomenon of "concentration of measure" [19, 33], and asymptotic analysis of statistical techniques is possible. Standard dimension reduction techniques, such as principal component analysis and factor analysis, work well when the data lie near a linear subspace of high dimensional space. They do not work well when the data lie near a nonlinear manifold embedded in the high dimensional space.

In this paper, we take a "worst case" viewpoint of the manifold learning problem. Let $\mathcal{H}$ be a separable Hilbert space, and let $\mathcal{P}$ be a probability measure supported on the unit ball $B_{\mathcal{H}}$ of $\mathcal{H}$. Let $|\cdot|$ denote the Hilbert space norm of $\mathcal{H}$, and for any $x, y \in \mathcal{H}$ let $d(x, y) = |x - y|$. For any $x \in B_{\mathcal{H}}$ and any $\mathcal{M} \subset B_{\mathcal{H}}$, a closed subset, let $d(x,\mathcal{M}) = \inf_{y\in\mathcal{M}} |x-y|$ and $\mathcal{L}(\mathcal{M},\mathcal{P}) = \int d(x,\mathcal{M})^2 d\mathcal{P}(x)$. We assume that i.i.d. data are generated from sampling $\mathcal{P}$, which is fixed but unknown. This is a worst-case view in the sense that no prior information about the data generating mechanism is assumed to be available or used for the subsequent development. This is the viewpoint of the modern statistical learning theory [54].

In order to state the problem more precisely, we need to describe the class of manifolds within which we will search for the existence of a manifold which satisfies the manifold hypothesis.

Let $\mathcal{M}$ be a submanifold of $\mathcal{H}$. The reach $\tau > 0$ of $\mathcal{M}$ is the largest number such that for any $0 < r < \tau$, any point at a distance r of $\mathcal{M}$ has a unique nearest point on $\mathcal{M}$.

Let $\mathcal{G} = \mathcal{G}(d,V,\tau)$ be the family of d dimensional $\mathcal{C}^2$-submanifolds of the unit ball in $\mathcal{H}$ with volume $\leq V$ and reach $\geq \tau$. We will assume that $\tau < 1$. We consider a bound on the reach to be a natural constraint since if data lie within a distance less than the reach of the manifold, it can be denoised by mapping data points to the nearest point on the manifold.

Let $\mathcal{P}$ be an unknown probability distribution supported on the unit ball of a separable (possibly infinite dimensional) Hilbert space and let $(x_1, x_2, \ldots)$ be i.i.d. random samples sampled from $\mathcal{P}$.

Let $\mathcal{B}$ be a black-box function which when given the labels $\ell(v), \ell(w)$ of two vectors $v, w \in \mathcal{H}$ outputs the inner product

$$\mathcal{B}(\ell(u), \ell(v)) = \langle v, w\rangle.$$

Note that while we permit the Hilbert space to be infinite dimensional, we require the labels to be finite dimensional for the finiteness of the algorithm.

The test for the manifold hypothesis answers the following affirmatively:

Given error ε, dimension d, volume V, reach τ, and confidence $1-\delta$, is there an algorithm that takes a number of samples depending on these parameters and with probability $1-\delta$ distinguishes between the following two cases (at least one must hold):

(a) whether there is a

$$\mathcal{M} \in \mathcal{G} = \mathcal{G}(d, CV, \tau/C)$$

such that

$$\int d(M,x)^2 dP(x) < C\varepsilon,$$

(b) whether there is no manifold

$$\mathcal{M} \in \mathcal{G}(d, V/C, C\tau)$$

such that

$$\int d(M,x)^2 dP(x) < \varepsilon/C\ ?$$

Here $d(M,x)$ is the distance from a random point x to the manifold $\mathcal{M}$, C is a constant depending only on d.

The basic statistical question is the following:

What is the number of samples needed for testing the hypothesis that data lie near a low dimensional manifold?

The desired result is that the sample complexity of the task depends only on the "intrinsic" dimension, volume, and reach, but not the "ambient" dimension.

We approach this by considering the empirical risk minimization problem.

Let

$$\mathcal{L}(M, P) = \int d(x, M)^2 dP(x) \ ,$$

and define the empirical loss

$$L_{\text{emp}}(M) = \frac{1}{s} \sum_{i=1}^{s} d(x_i, M)^2,$$

where $(x_1, \ldots, x_s)$ are the data points. The sample complexity is defined to be the smallest s such that there exists a rule $\mathcal{A}$ which assigns to given $(x_1, \ldots, x_s)$ a manifold $\mathcal{M}_{\mathcal{A}}$ with the property that if $x_1, \ldots, x_s$ are generated i.i.d. from $\mathcal{P}$, then

$$\mathbb{P}\left[\mathcal{L}(\mathcal{M}_A, \mathcal{P}) - \inf_{\mathcal{M} \in \mathcal{G}} \mathcal{L}(M, \mathcal{P}) > \varepsilon\right] < \delta.$$

We need to determine how large s needs to be so that

$$\mathbb{P}\left[\sup_{\mathcal{G}} \left|\frac{1}{s} \sum_{i=1}^{s} d(x_i, \mathcal{M})^2 - \mathcal{L}(\mathcal{M}, \mathcal{P})\right| < \varepsilon\right] > 1 - \delta.$$

The answer to this question is given by Theorem 1 in the paper.

The proof of the theorem proceeds by approximating manifolds using point clouds and then using uniform bounds for k-means (Lemma 6 of the paper).

The uniform bounds for k-means are proven by getting an upper bound on the fat shattering dimension of a certain function class and then using an integral related to Dudley's entropy integral. The bound on the fat shattering dimension is obtained using a random projection (along with the Johnson Lindenstrauss lemma) and the Sauer-Shelah lemma. The use of random projections in this context appears in Chapter 4 of [35] and in [40]. However, due to the absence of chaining, the bounds derived there are weaker. The Johnson-Lindenstrauss lemma has been used previously in the context of manifolds in [2, 13, 28], where random projections of low dimensional submanifolds of a high dimensional space are shown (after a suitable dilation) to be nearly isometric to the original manifold with high probability.

The algorithmic question can be stated as follows:

Given N points $x_1, \ldots, x_N$ in the unit ball in $\mathbb{R}^n$, distinguish between the following two cases (at least one must be true):

(a) whether there is a manifold $\mathcal{M} \in \mathcal{G} = \mathcal{G}(d, CV, C^{-1}\tau)$ such that

$$\frac{1}{N} \sum_{i=1}^{N} d(x_i, M)^2 \leq C\varepsilon,$$

where C is some constant depending only on d.

(b) whether there is no manifold $\mathcal{M} \in \mathcal{G} = \mathcal{G}(d, V/C, C\tau)$ such that

$$\frac{1}{N} \sum_{i=1}^{N} d(x_i, M)^2 \leq \varepsilon/C,$$

where C is some constant greater than 1 depending only on d.

The key step to solving this problem is to translate the question of optimizing the squared loss over a family of manifolds to that of optimizing over sections of a disc bundle. The former involves an optimization over a non-parameterized infinite dimensional space, while the latter involves an optimization over a parameterized (albeit infinite dimensional) set.

The proof of correctness of our algorithm requires showing two things:

(1) If a "good manifold" exists, then our algorithm is guaranteed to find "good local sections" close to pieces of this manifold. Therefore, if the algorithm is unable to find good local sections, then there is no good manifold.
(2) If good local sections are found by our algorithm, then these local sections can be patched together to form a manifold in the class of interest, such that each local section is close to a piece of this manifold.

These good local sections are local sections of a disc bundle. We introduce the notion of a cylinder packet in order to define a disc bundle. A cylinder packet is a finite collection of cylinders satisfying certain alignment constraints. An example of a cylinder packet corresponding to a d-manifold $\mathcal{M}$ of reach τ (see Definition 1) in $\mathbb{R}^n$ is obtained by taking a net (see Definition 6) of the manifold and for every point p in the net, throwing in a cylinder centered at p isometric to $2\bar{\tau}(B_d \times B_{n-d})$ whose d dimensional central cross section is tangent to $\mathcal{M}$. In general, p need not be at the center of the cylinder, but would lie inside the cylinder. Here $\bar{\tau} = c\tau$ for some appropriate constant c depending only on d, while B_d and B_{n-d} are d dimensional and $(n-d)$ dimensional balls, respectively.

On every cylinder $\mathtt{cyl}_i$ in the packet, we define a function f_i that is the squared distance to the d dimensional central cross section of $\mathtt{cyl}_i$. These functions are put together using a partition of unity defined on $\cup_i \mathtt{cyl}_i$. The resulting function f is an "approximate-squared-distance function" (see Definition 15). The base manifold is the set of points x at which the gradient ∂f is orthogonal to the eigenvectors corresponding to the top $n-d$ eigenvalues of the Hessian $\mathtt{Hess}\, f(x)$. The fiber of the disc bundle at a point x on the base manifold is defined to be the $(n-d)$ dimensional Euclidean ball centered at x contained in the span of the aforementioned eigenvectors of the Hessian. The base manifold and its fibers together define the disc bundle. The base manifold is a temporary approximation to the manifold that we are searching for.

We next perform an optimization over sections of the disc bundle in order to certify the existence of the desired manifold, if it exists. This optimization proceeds as follows. We fix a cylinder $\mathtt{cyl}_i$ of the cylinder packet. We optimize the squared loss over local sections corresponding to jets whose C^2-form is bounded above by $\frac{c_1}{\bar{\tau}}$, where c_1 is a controlled constant. The corresponding graphs are each contained inside $\mathtt{cyl}_i$. The optimization over local sections is performed by minimizing squared loss over a space of C^2-jets (see Definition 22) constrained by inequalities developed in [24]. The resulting local sections corresponding to various i are then patched together using the disc bundle and a partition of unity supported on the base manifold, to yield the actual manifold. The last step is performed implicitly, since we do not actually need to produce a manifold, but only need to certify the existence or non-existence of a manifold possessing certain properties.

The optimizations are performed over a large ensemble of cylinder packets. Indeed the size of this ensemble is the chief contribution in the complexity bound.

The results of this paper together with those of [24] lead to an algorithm for fitting a manifold to the data as well; the main additional step is to construct local sections from jets, rather than settling for the existence of good local sections as we do here.

1.1. Definitions.

Definition 1 (reach). *Let $\mathcal{M}$ be a subset of $\mathcal{H}$. The reach of $\mathcal{M}$ is the largest number τ to have the property that any point at a distance $r < \tau$ from $\mathcal{M}$ has a unique nearest point in $\mathcal{M}$.*

Definition 2 (tangent space). *Let $\mathcal{H}$ be a separable Hilbert space. For a closed $A \subseteq \mathcal{H}$, and $a \in A$, let the "tangent space" $Tan^0(a, A)$ denote the set of all vectors v such that for all $\epsilon > 0$, there exists $b \in A$ such that $0 < |a-b| < \epsilon$ and $\left|v/|v| - \frac{b-a}{|b-a|}\right| < \epsilon$. Let $Tan(a, A)$ denote the set of all x such that $x-a \in Tan^0(a, A)$. For a set $X \subseteq \mathcal{H}$ and a point $p \in \mathcal{H}$, let $\mathbf{d}(p, X)$ denote the Euclidean distance of the nearest point in X to p.*

The following result of Federer (Theorem 4.18 of [22]) gives an alternate characterization of the reach.

Proposition 1. *Let A be a closed subset of $\mathbb{R}^n$. Then*

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

1.2. Constants.

d is a fixed integer. Constants c, C, C', etc., depend only on d. These symbols may denote different constants in different occurrences, but d always stays fixed.

1.3. d-planes.

$\mathcal{H}$ denotes a fixed Hilbert space, possibly infinite dimensional, but in any case of dimension $> d$. A d-plane is a d dimensional vector subspace of $\mathcal{H}$. We write Π to denote a d-plane, and we write dPL to denote the space of all d-planes. If $\Pi, \Pi' \in dPL$, then we write $dist(\Pi, \Pi')$ to denote the infimum of $\|T - I\|$ over all orthogonal linear transformations $T : \mathcal{H} \to \mathcal{H}$ that carry Π to Π'. Here, the norm $\|A\|$ of a linear map $A : \mathcal{H} \to \mathcal{H}$ is defined as

$$\sup_{v \in \mathcal{H} \setminus \{0\}} \frac{\|Av\|_{\mathcal{H}}}{\|v\|_{\mathcal{H}}}.$$

One checks easily that $(dPL, dist)$ is a metric space. We write $\Pi^\perp$ to denote the orthocomplement of Π in $\mathcal{H}$.

1.4. Patches.

Suppose $B_\Pi(0, r)$ is the ball of radius r about the origin in a d-plane Π, and suppose

$$\Psi : B_\Pi(0, r) \to \Pi^\perp$$

is a C^2-map and hence a $C^{1,1}$-map, with $\Psi(0) = 0$. Then we call

$$\Gamma = \{x + \Psi(x) : x \in B_\Pi(0, r)\} \subset \mathcal{H}$$

a patch of radius r over Π centered at 0. We define

$$\|\Gamma\|_{\dot{C}^{1,1}(B_\Pi(0,r))} := \sup_{\text{distinct}\, x, y \in B_\Pi(0,r)} \frac{\|\partial\Psi(x) - \partial\Psi(y)\|}{\|x - y\|}.$$

Here,

$$\partial\Psi(x) : \Pi \to \Pi^{\perp}$$

is a linear map, and for linear maps $A : \Pi \to \Pi^{\perp}$, we define $\|A\|$ as

$$\sup_{v \in \Pi \setminus \{0\}} \frac{\|Av\|}{\|v\|}.$$

If also

$$\partial\Psi(0) = 0,$$

then we call Γ a patch of radius r tangent to Π at its center 0. If Γ_0 is a patch of radius r over Π centered at 0 and if $z \in \mathcal{H}$, then we call the translate $\Gamma = \Gamma_0 + z \subset \mathcal{H}$ a patch of radius r over Π, centered at z. If Γ_0 is tangent to Π at its center 0, then we say that Γ is tangent to Π at its center z.

1.5. **Imbedded manifolds.**

Definition 3. *Let $\mathcal{M} \subset \mathcal{H}$ be a "compact imbedded d-manifold" (for short, just a "manifold") if the following hold:*

- *$\mathcal{M}$ is compact.*
- *There exists an $r_1 > r_2 > 0$ such that for every $z \in \mathcal{M}$, there exists $T_z\mathcal{M} \in dPL$ such that $\mathcal{M} \cap B_{\mathcal{H}}(z, r_2) = \Gamma \cap B_{\mathcal{H}}(z, r_2)$ for some patch Γ over $T_z(\mathcal{M})$ of radius r_1, centered at z and tangent to $T_z(\mathcal{M})$ at z. We call $T_z(\mathcal{M})$ the tangent space to $\mathcal{M}$ at z.*

We say that $\mathcal{M}$ has infinitesimal reach $\geq \rho$ if for every $\rho' < \rho$, there is a choice of $r_1 > r_2 > 0$ such that for every $z \in \mathcal{M}$ there is a patch Γ over $T_z(\mathcal{M})$ of radius r_1, centered at z and tangent to $T_z(\mathcal{M})$ at z which has $C^{1,1}$-norm at most $\frac{1}{\rho'}$.

Definition 4 (a class of imbedded C^2 d-manifolds)**.** *Let $B_{\mathcal{H}}$ be the unit ball in $\mathcal{H}$. Let $\mathcal{G} = \mathcal{G}(d, V, \tau)$ be the family of imbedded C^2 d-submanifolds of $B_{\mathcal{H}}$ having volume less than or equal to V and reach greater than or equal to τ. We assume as mentioned before that $\tau < 1$.*

Let $\mathcal{H}$ be a separable Hilbert space and $\mathcal{P}$ be a probability distribution supported on its unit ball $B_{\mathcal{H}}$. Let $|\cdot|$ denote the Hilbert space norm on $\mathcal{H}$. For $x, y \in \mathcal{H}$, let $\mathbf{d}(x, y) := |x - y|$. For any $x \in B_{\mathcal{H}}$ and any $\mathcal{M} \subseteq B_{\mathcal{H}}$, let $\mathbf{d}(x, \mathcal{M}) := \inf_{y \in \mathcal{M}} |x - y|$, and

$$\mathcal{L}(\mathcal{M}, \mathcal{P}) := \int \mathbf{d}(x, \mathcal{M})^2 d\mathcal{P}(x).$$

Let $\mathcal{B}$ be a black-box function which when given the labels $\ell(v), \ell(w)$ of two vectors $v, w \in \mathcal{H}$ outputs the inner product

$$\mathcal{B}(\ell(u), \ell(v)) = \langle v, w \rangle.$$

We develop an algorithm which for given $\delta, \epsilon \in (0, 1)$, $V > 0$, integer d, and $\tau > 0$ takes i.i.d. random samples from $\mathcal{P}$ as input and determines which of the following two is true (at least one must be):

(1) there exists $\mathcal{M} \in \mathcal{G}(d, CV, \frac{\tau}{C})$ such that $\mathcal{L}(\mathcal{M}, \mathcal{P}) \leq C\epsilon$,
(2) there exists no $\mathcal{M} \in \mathcal{G}(d, V/C, C\tau)$ such that $\mathcal{L}(\mathcal{M}, \mathcal{P}) \leq \frac{\epsilon}{C}$.

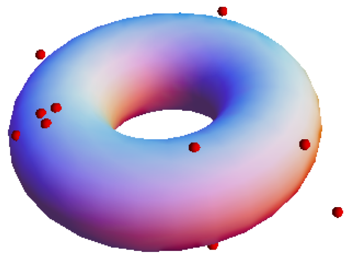

FIGURE 1. Data lying in the vicinity of a two dimensional torus.

The answer is correct with probability at least $1-\delta$. Here, C depends on d alone and is greater than 1.

The number of data points required (Theorem 1) is of the order of

$$n := \frac{N_p \ln^4\left(\frac{N_p}{\epsilon}\right) + \ln(\delta^{-1})}{\epsilon^2},$$

where

$$N_p := V\left(\frac{1}{\tau^d} + \frac{1}{\epsilon^{d/2}\tau^{d/2}}\right),$$

and the number of arithmetic operations is

$$\exp\left(C\left(\frac{V}{\tau^d}\right) n \ln(\tau^{-1})\right).$$

(Corollary 2 shows that N_p is an upper bound on the size of a $\sqrt{\epsilon\tau}$-net of $\mathcal{M}$.) The number of calls made to $\mathcal{B}$ is $O(n^2)$.

We say that such an algorithm *tests the manifold hypothesis.*

If one wishes to ascertain the mean of a bounded random variable to within ϵ, it requires $1/\epsilon^2$ samples. However, for more complicated questions such as estimating a manifold to within ϵ in Hausdorff distance, there are upper and lower bounds of $O(\epsilon^{\frac{-(2+d)}{2}})$ [12]. Thus our upper bound of $\epsilon^{-d/2-2}$ is not far from this bound.

The outline of the paper is as follows.

Section 2 is a brief survey of the literature on manifold learning.

Section 3 introduces sample complexity and has the statement of Theorem 1 which is our main result on sample complexity of testing the manifold hypothesis. This theorem is about the number of samples needed to fit a manifold of certain reach, volume, and dimension to an arbitrary probability distribution supported on the unit ball.

Section 4 contains the proof of Theorem 1. We reduce the problem to a uniform bound over a space of manifolds relating the empirical risk (or loss) to the true risk (i.e., expected squared distance between a random point to the manifold). Covering numbers at small scales play an important role. Here, by a covering number we mean the minimal size of a finite subset (net) of a manifold $\mathcal{M}$ such that every point of $\mathcal{M}$ is within ϵ of some point of the net. Primary tools include the Johnson-Lindenstrauss lemma, the Vapnik-Chervonenkis theory (Lemma 3 and Lemma 4), and tools from empirical processes (Lemma 5 and Lemma 6).

In Lemma 9 of Section 5, a uniform bound over the space of k-tuples of affine subspaces is obtained relating the empirical risk to the true risk.

In Section 6, we perform a dimension reduction that maps the manifold into a subspace spanned by a net of the manifold. This dimension reduction maps a manifold onto another with a similar reach, volume, and equal dimension. Further, the Hausdorff distance between the two manifolds is small. Since we are not assuming that the dimension on the ambient space is finite, such a dimension reduction is essential to obtain a finite algorithm. Our results are stated for separable Hilbert spaces. Once the dimension reduction is done, without loss of generality, we assume that the ambient space is $\mathbb{R}^n$.

In Section 7, we provide an overview of the algorithm for testing the manifold hypothesis.

In Section 8, we give the formal definitions of the disc bundles we use in our algorithm.

Section 9 contains the key technical result of the paper—Theorem 13.

In this theorem we consider, as a function ϕ of an open subset of the ambient space, the gradient of an approximate-squared-distance-function $F^{\bar{o}}$. For each point x in the domain of ϕ, we project $\phi(x)$ to the subspace spanned by the eigenvectors of the Hessian of $F^{\bar{o}}(x)$ corresponding to large eigenvalues, and we use the implicit function theorem on the zeros of that set. Specifically, we consider the set of points where ϕ is orthogonal to the span of these eigenvectors. We construct a disc bundle with a manifold (the "putative manifold") as the base space, with the fiber at a base point being given by the span of the eigenvectors corresponding to the large eigenvalues of the Hessian of $F^{\bar{o}}$ intersected with a ball of radius $\bar{\tau}$. Every point in the disc bundle can be expressed uniquely as a base point on the putative manifold plus a vector in the fiber corresponding to that base point. This unique decomposition is used later to patch together local sections to form a global section of the disc bundle. A key component is a lower bound on the gap between the top $(n-d)$ eigenvalues and bottom d eigenvalues of the Hessian of $F^{\bar{o}}$ that is given by a controlled constant. This gap affects both the reach of the manifold and the radius $\bar{\tau}$ of the fibers.

At this point our goal is to perform an optimization over the space of manifolds $\mathcal{G}(d, V, \tau)$ in order to certify the existence or non-existence of a manifold having a certain least squares error with respect to the data. Unfortunately this space is not equipped with a vector space structure and is difficult to optimize on. Our approach to handling this difficulty is to express it as the union of classes of manifolds, each class consisting of those manifolds that are near a given putative manifold. Each manifold in a fixed class can be associated with a section of a disc bundle over the relevant putative manifold. These sections enjoy a convex structure. Since the squared loss function is a convex function, we can use convex optimization techniques over the manifolds in a given class. It remains to describe how we come up with an exhaustive collection of disc bundles such that every manifold in $\mathcal{G}(d, V, \tau)$ corresponds to a section of some disc bundle.

In Section 10, it is shown how to construct cylinder packets consisting of cylinders isometric to $\bar{\tau}(B_d \times B_{n-d})$ that satisfy certain alignment constraints.

In Section 11, an approximate-squared-distance function (asdf) is defined, and it is shown how to construct a disc bundle from such a function. It is further shown that if an asdf has certain properties with respect to a manifold, then the manifold is the graph of a section of the corresponding disc bundle.

In Section 12 it is shown how to construct an asdf using cylinder packets. Each such function defines a disc bundle over a base putative manifold. A subset of the manifolds in $\mathcal{G}$ correspond to sections of this disc bundle. It is further shown that if the cylinder packet is "admissible" with respect to a manifold, then the corresponding disc bundle has a section of which this manifold is the graph.

In Section 13 results on Whitney interpolation are used to give a polyhedral description of the collection of jets that correspond to local sections with the appropriate bound on the C^2 norm. Vaidya's algorithm [53] is then used to optimize over the polytope thus constructed to estimate the optimal mean-squared error with respect to the data. The complexity of testing the existence of good local sections for a given cylinder packet is polynomial in the size of the data.

In Section 14 we describe how local sections are patched together to give global sections, using a partition of unity supported on a putative manifold.

In Section 15 we show that the reach of the manifold constructed in the previous step is of the order of τ.

In Section 16, we show that the mean-square error in approximating the data is within a controlled constant C of the optimal.

In Section 17, we provide bounds on the number of arithmetic operations required by the algorithm.

The Appendix contains proofs of some of the results in the main text.

1.6. **A note on controlled constants.** In this section, and the following sections, we will make frequent use of constants $c, C, C_1, C_2, \overline{c}_1, \dots, \overline{c}_{11}$ and c_{12}, etc. These constants are "controlled constants" in the sense that their value is entirely determined by the dimension d unless explicitly specified otherwise (as for example in Theorem 13). Also, the value of a constant can depend on the values of constants defined before it, but not those defined after it. This convention clearly eliminates the possibility of loops.

2. Literature on manifold learning

At present there are available a number of methods which aim to transform data lying near a d dimensional manifold in an N dimensional space into a set of points in a low dimensional space close to a d dimensional manifold. A comprehensive review of manifold learning can be found in a recent book [35]. The most basic method is "principal component analysis" (PCA) [27, 45], where data points are projected on to the span of the eigenvectors corresponding to the top d eigenvalues of the $(N \times N)$ covariance matrix of the data points. A variation is the kernel PCA [49] where one works in the "feature space" rather than the original ambient space.

In the case of "multi-dimensional scaling"(MDS) [15], only pairwise distances between points are attempted to be preserved when projecting to a lower dimensional space.

"Isomap" [52] attempts to improve on MDS by trying to capture geodesic distances between points while projecting. For each data point a "neighborhood graph" is constructed using its k neighbors (k could be varied based on other criteria), the edges carrying the length between points. Now the shortest distance between points is computed in the resulting global graph containing all the neighborhood graphs using a standard graph theoretic algorithm such as Dijkstra's. It is this "geodesic" distance which the method tries to preserve when projecting to a lower dimensional space.

"Maximum variance unfolding" [55] also constructs the neighborhood graph as in the case of Isomap but tries to maximize the distance between projected points keeping distance between the nearest points unchanged after projection.

In "diffusion maps" [14], a complete graph on the N data points is built. Each edge is assigned a weight based on a gaussian. The matrix is normalized to make it into a transition matrix of a Markov chain. The d nontrivial λ_i and their eigenvectors v_i of P^t are computed, the d eigenvectors form the rows of the $d \times N$ matrix, and the columns of this matrix constitute the lower dimensional representation of the data points.

"Local linear embedding" (LLE) [47] preserves solely local properties of the data once again using the neighborhood graph of each data point.

In the case of the "Laplacian eigenmap" [3, 30] again, a nearest neighbor graph is formed. Either this could be an undirected k-nearest neighbor graph or there could be a parameter ϵ that determines neighborhoods based on points that are within a Euclidean distance of ϵ. Weights are assigned to the edges as indicated below, a Laplacian matrix is computed, and a certain quadratic function is based on the Laplacian minimized through the solution of a generalized eigenvalue problem. The top d-eigenvectors constitute a representation of the data.

Hessian LLE (also called Hessian eigenmaps) [20] and "local tangent space alignment" (LTSA) [56] attempt to improve on LLE by also taking into consideration the curvature of the higher dimensional manifold while preserving the local pairwise distances.

The alignment of local coordinate mappings also underlies some other methods such as "local linear coordinates" [46] and "manifold charting" [7].

Methods which map higher dimensional data points to lower dimensional constructs (principal sets) more general than manifolds are described in [44] and studied more formally in [25]. Another line of research starts with principal curves/surfaces [26] and topology preserving networks [36]. Manifolds of probability distributions and connections to the work of Amari [1] have been studied in the work of Newton [42]. Uniform rectifiability offers an alternative to reach as a way of distinguishing complicated sets from simple ones (see [17]).

Some of the algorithms are known to perform correctly under the hypothesis that data lie on a manifold of a specific kind. In Isomap and LLE, the manifold has to be an isometric embedding of a convex subset of Euclidean space. In the case of [4, 10], the manifold is a simplicial complex and witness complex, respectively. In the limit as the number of data points tends to infinity, when the data approximate a manifold, then one can recover the geometry of this manifold by computing an approximation of the Laplace-Beltrami operator. Laplacian eigenmaps and diffusion maps rest on this idea. LTSA works for parameterized manifolds and detailed error analysis is available for it.

3. Sample complexity of manifold fitting

In this section, we show that if we randomly sample sufficiently many points as in the above mentioned algorithm and then find the least squares fit manifold to these data, we obtain an almost optimal manifold.

Definition 5 (sample complexity)**.** *Given error parameters ϵ, δ, a space Y, and a set of functions (henceforth function class) $\mathcal{F}$ of functions $f : Y \to \mathbb{R}$, we define the sample complexity $s = s(\epsilon, \delta, \mathcal{F})$ to be the least number such that the following is true. There exists a function $\mathcal{A} : Y^s \to \mathcal{F}$ such that, for any probability distribution $\mathcal{P}$ supported on Y, if $(x_1, \dots, x_s) \in Y^s$ is a sequence of i.i.d. draws from $\mathcal{P}$, then $f_{out} := \mathcal{A}((x_1, \dots, x_s))$ satisfies*

$$\mathbb{P}\left[\mathbb{E}_{x \dashv \mathcal{P}} f_{out}(x) < (\inf_{f \in \mathcal{F}} \mathbb{E}_{x \dashv \mathcal{P}} f(x)) + \epsilon\right] > 1 - \delta.$$

We state below, a sample complexity bound when the mean-squared error is minimized over $\mathcal{G}(d, V, \tau)$. Thus the function class $\mathcal{F}$ consists of functions $f_{\mathcal{M}}(x) := \mathbf{d}(x, \mathcal{M})^2$ indexed by $\mathcal{M}$. The manifold that minimizes the empirical risk will be denoted $\mathcal{M}_{erm}(X)$, *erm* standing for empirical risk minimization. This manifold is a function of $X = (x_1, \dots, x_s)$, a sequence of i.i.d. points from $\mathcal{P}$. The minimization involved is of a quantity $\mathcal{L}(\mathcal{M}, \mathcal{P}_X)$. The theorem as stated is true only if s, the number of data points, is greater than or equal to $s_{\mathcal{G}}(\epsilon, \delta)$. This theorem says that instead of optimizing $\mathcal{L}(\mathcal{M}, \mathcal{P})$ over manifolds $\mathcal{M}$, if s is sufficiently large, we might as well optimize $\mathcal{L}(\mathcal{M}, \mathcal{P}_X)$ over manifolds $\mathcal{M}$, where $\mathcal{P}_X$ is the empirical measure equally distributed over the data set $x_1, \dots, x_s$. The constant $C > 1$ in the definition of $U_{\mathcal{G}}(1/\epsilon)$ depends on the volume of a ball in d dimensional Euclidean space. The constant $C' > 1$ in $s_{\mathcal{G}}(\epsilon, \delta)$ is a universal constant.

Theorem 1. *For $r > 0$, let*

$$U_{\mathcal{G}}(1/r) = \frac{CV}{\tau^d} + \frac{CV}{(\tau r)^{d/2}}.$$

Let

$$s_{\mathcal{G}}(\epsilon, \delta) := C' \left(\frac{U_{\mathcal{G}}(1/\epsilon)}{\epsilon^2} \left(\log^4 \left(\frac{U_{\mathcal{G}}(1/\epsilon)}{\epsilon} \right) \right) + \frac{1}{\epsilon^2} \log \frac{1}{\delta} \right).$$

Let $s \geq s_{\mathcal{G}}(\epsilon, \delta)$ and $X = (x_1, \dots, x_s)$ be a sequence of i.i.d. points from $\mathcal{P}$ and $\mathcal{P}_X$ be the uniform probability measure over X. Let $\mathcal{M}_{erm}(X)$ denote a manifold in $\mathcal{G}(d, V, \tau)$ that approximately minimizes the quantity

$$\mathcal{L}(\mathcal{M}, \mathcal{P}_X) = s^{-1} \sum_{i=1}^{s} \mathbf{d}(x_i, \mathcal{M})^2$$

in the sense that

$$\mathcal{L}(\mathcal{M}_{erm}(X), \mathcal{P}_X) - \inf_{\mathcal{M} \in \mathcal{G}(d,V,\tau)} \mathcal{L}(\mathcal{M}, \mathcal{P}_X) < \frac{\epsilon}{2}.$$

Then,

$$\mathbb{P}\left[\mathcal{L}(\mathcal{M}_{erm}(X), \mathcal{P}) - \inf_{\mathcal{M} \in \mathcal{G}(d,V,\tau)} \mathcal{L}(\mathcal{M}, \mathcal{P}) < \epsilon\right] > 1 - \delta.$$

3.1. **Sketch of the proof of Theorem 1.** The first step involves obtaining new dimension independent bounds for the sample complexity of k-means, or in other words the problem of fitting k points to a probability distribution supported on the unit ball in a Hilbert space. This is essentially done in Lemma 6. Recall that given a data set $\{x_1, \dots, x_s\}$, k-means is the problem of producing k-centers $\mathbf{c} = \{c_1, \dots, c_k\}$ with the property that for any other set of k centers $\mathbf{c}'$,

$$\sum_{i \le s} \min_{j \le k} \mathbf{d}(x_i, c_j)^2 \le \sum_{i \le s} \min_{j \le k} \mathbf{d}(x_i, c'_j)^2.$$

(For the best previously known bound of $O\left(\frac{k^2}{\epsilon^2}\right)$ for the sample complexity of k-means in a ball in a Hilbert space, see [38].)

The second step involves upper bounding (see Lemma 7), the sample complexity of fitting the best manifold in $\mathcal{G}(d, V, \tau)$ to a probability distribution supported on the unit ball, by the sample complexity of fitting $U_{\mathcal{G}}(1/\epsilon)$ points in a least squares sense to the same probability distribution. This argument involves approximating manifolds in $\mathcal{G}(d, V, \tau)$ to within ϵ using point sets with respect to Hausdorff distance. This is done in Claim 1 and Corollary 2.

4. Proof of Theorem 1

Let $\mathcal{M} \in \mathcal{G}(d, V, \tau)$. For $x \in \mathcal{M}$ denote the orthogonal projection from $\mathcal{H}$ to the affine subspace $Tan(x, \mathcal{M})$ by Π_x. We will need the following claim to prove Theorem 1.

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

$$U := \{y \big| |y - \Pi_x y| \le \tau/C\} \cap \{y \big| |x - \Pi_x y| \le \tau/C\},$$

for a sufficiently large controlled constant C. There exists a $C^{1,1}$ function $F_{x,U}$ from $\Pi_x(U)$ to $\Pi_x^{-1}(\Pi_x(0))$ such that

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

and further such that the Lipschitz constant of the gradient of $F_{x,U}$ is bounded above by $\frac{C_1}{\tau}$.

The above claim is proved in the Appendix.

4.1. A bound on the size of an ϵ-net.

Definition 6. *Let $(X, \mathbf{d})$ be a metric space, and $r > 0$. We say that Y is an r-net of X if $Y \subseteq X$ and for every $x \in X$, there is a point $y \in Y$ such that $\mathbf{d}(x, y) < r$.*

Corollary 2. *Let*

$$U_{\mathcal{G}} : \mathbb{R}^+ \to \mathbb{R}$$

be given by

$$U_{\mathcal{G}}(1/r) = CV\left(\frac{1}{\tau^d} + \frac{1}{(\tau r)^{d/2}}\right).$$

Let $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ and $\mathcal{M}$ be equipped with the metric $\mathbf{d}_{\mathcal{H}}$ of the Hilbert space $\mathcal{H}$. Then, for any $r > 0$, there is a $\sqrt{\tau r}$-net of $\mathcal{M}$ consisting of no more than $U_{\mathcal{G}}(1/r)$ points.

Proof. It suffices to prove that for any $r \le \tau$, there is an r-net of $\mathcal{M}$ consisting of no more than $CV\left(\frac{1}{\tau^d} + \frac{1}{r^d}\right)$, since if $r > \tau$, a τ-net is also an r-net. Suppose $Y = \{y_1, y_2, \dots\}$ is constructed by the following greedy procedure. Let $y_1 \in \mathcal{M}$ be chosen arbitrarily. Suppose $y_1, \dots, y_k$ have been chosen. If the set of all y such

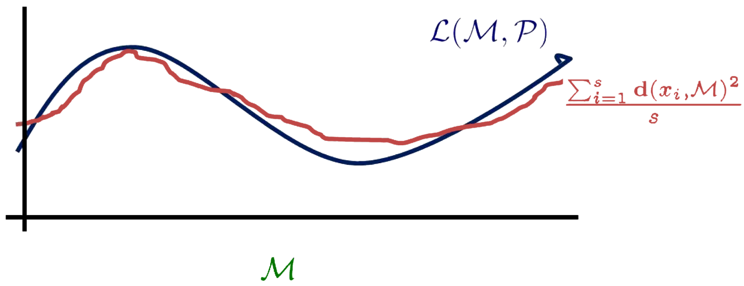

FIGURE 2. A uniform bound (over $\mathcal{G}$) on the difference between the empirical and true loss.

that $\min_{1\le i\le k} |y - y_i| \ge r$ is non-empty, let y_{k+1} be an arbitrary member of this set. Else declare the construction of Y to be complete.

We see that Y is an r-net of $\mathcal{M}$. Second, we see that the distance between any two distinct points $y_i, y_j \in Y$ is greater than or equal to r. Therefore the two balls $\mathcal{M} \cap B_{\mathcal{H}}(y_i, r/2)$ and $\mathcal{M} \cap B_{\mathcal{H}}(y_j, r/2)$ do not intersect.

By Claim 1, and the fact that the reach of $\mathcal{M}$ is greater than or equal to τ, it follows that for each $y \in Y$, there are controlled constants $0 < c < 1/2$ and $0 < c'$ such that for any $r \in (0, \tau]$, the volume of $\mathcal{M} \cap B_{\mathcal{H}}(y, cr)$ is greater than $c' r^d$. See footnote.[1] (In invoking Claim 1, the Lipschitz property of the gradient is not needed.)

Since the volume of

$$\{z \in \mathcal{M} | \mathbf{d}(z, Y) \le r/2\}$$

is less than or equal to V, the cardinality of Y is less than or equal to $\frac{V}{c' r^d}$ for all $r \in (0, \tau]$. The corollary follows. ☐

4.2. **Tools from empirical processes.** In this subsection, unless otherwise stated, data $(x_1, \dots, x_s)$ will be a sequence of i.i.d. draws from a probability measure $\mathcal{P}$ (or μ) supported on the unit ball $B_{\mathcal{H}}$ of a Hilbert space $\mathcal{H}$. In order to prove a uniform bound of the form

$$\mathbb{P}\left[\sup_{F\in\mathcal{F}} \left| \frac{\sum_{i=1}^{s} F(x_i)}{s} - \mathbb{E}_{\mathcal{P}} F(x) \right| < \epsilon \right] > 1 - \delta, \tag{2}$$

it suffices to bound a measure of the complexity of $\mathcal{F}$ known as the fat shattering dimension of the function class $\mathcal{F}$. The metric entropy (defined below) of $\mathcal{F}$ can be bounded using the fat shattering dimension, leading to a uniform bound of the form of (2), see Figure 2.

Definition 7 (metric entropy). *Given a metric space (Y, ρ), we call $Z \subseteq Y$ an η-net of Y if for every $y \in Y$, there is a $z \in Z$ such that $\rho(y, z) < \eta$. Let $\mathcal{P}$ be a measure supported on a metric space X, and $\mathcal{F}$ a class of functions from X to $\mathbb{R}$. Let $N(\eta, \mathcal{F}, \mathcal{L}_2(\mathcal{P}))$ denote the minimum number of elements that an η-net of $\mathcal{F}$ could have, with respect to the metric imposed by the Hilbert space $\mathcal{L}_2(\mathcal{P})$. Here,*

[1]This is because the area of a surface is no less than the area of a projection of it onto a subspace.

the distance between $f_1 : X \to \mathbb{R}$ *and* $f_2 : X \to \mathbb{R}$ *is*

$$\|f_1 - f_2\|_{\mathcal{L}_2(\mathcal{P})} = \sqrt{\int (f_1(x) - f_2(x))^2 d\mathcal{P}}.$$

We call $\ln N(\eta, \mathcal{F}, \mathcal{L}_2(\mathcal{P}))$ *the metric entropy of* $\mathcal{F}$ *at scale* η *with respect to* $\mathcal{L}_2(\mathcal{P})$*.*

Definition 8 (fat shattering dimension)**.** *Let* $\mathcal{F}$ *be a set of real valued functions. We say that a set of points* $x_1, \dots, x_k$ *is* γ*-shattered by* $\mathcal{F}$ *if there is a vector of real numbers* $t = (t_1, \dots, t_k)$ *such that for all binary vectors* $\mathbf{b} = (b_1, \dots, b_k)$*, there is a function* $f_{\mathbf{b},t}$ *satisfying,*

$$f_{\mathbf{b},t}(x_i) = \begin{cases} > t_i + \gamma, & \text{if } b_i = 1; \\ < t_i - \gamma, & \text{if } b_i = 0. \end{cases} \tag{3}$$

For each $\gamma > 0$*, the fat shattering dimension* $\mathrm{fat}_\gamma(\mathcal{F})$ *of the set* $\mathcal{F}$ *is defined to be the size of the largest* γ*-shattered set if this is finite; otherwise* $\mathrm{fat}_\gamma(\mathcal{F})$ *is declared to be infinite.*

The supremum taken over $(t_1, \dots, t_k)$ *of the number of binary vectors* $\mathbf{b}$*, for which there is a function* $f_{\mathbf{b},t} \in \mathcal{F}$ *which satisfies (3), is called the* γ*-shatter coefficient of* $(x_1, \dots, x_k)$*. (Thus the* γ*-shatter coefficient of a* k*-element set that is* γ*-shattered is* 2^k*.)*

We will also need to use the notion of VC dimension and some of its properties. These appear below.

Definition 9. *Let* Λ *be a collection of measurable subsets of* $\mathbb{R}^m$*. For* $x_1, \dots, x_k \in \mathbb{R}^m$*, let the number of different sets in* $\{\{x_1, \dots, x_k\} \cap H; H \in \Lambda\}$ *be denoted the shatter coefficient* $N_\Lambda(x_1, \dots, x_k)$ *of* $(x_1, \dots, x_k)$*. The VC dimension* VC_Λ *of* Λ *is the largest integer* k *such that there exist* $x_1, \dots, x_k$ *such that* $N_\Lambda(x_1, \dots, x_k) = 2^k$*.*

The following result concerning the VC dimension of half-spaces is well known (Corollary 13.1 [18]).

Lemma 3. *Let* Λ *be the class of half-spaces in* $\mathbb{R}^g$*. Then* $VC_\Lambda = g + 1$*.*

We state the Sauer-Shelah lemma below.

Lemma 4 (Theorem 13.2 [18])**.** *Let* Λ *be a collection of measurable subsets of* $\mathbb{R}^g$*. For any* $x_1, \dots, x_k \in \mathbb{R}^g$*,* $N_\Lambda(x_1, \dots, x_k) \le \sum_{i=0}^{VC_\Lambda} \binom{k}{i}$*. (Note that* $\binom{k}{i} = 0$ *if* $i > k$*.)*

For $VC_\Lambda > 2$, $\sum_{i=0}^{VC_\Lambda} \binom{k}{i} \le k^{VC_\Lambda}$.

The lemma below follows from existing results from the theory of empirical processes in a straightforward manner, but does not seem to have appeared in print before. We have provided a proof in the Appendix.

Lemma 5. *Let* μ *be a probability measure supported on* X *and* $\mathcal{F}$ *be a class of functions* $f : X \to \mathbb{R}$*. Let* $x_1, \dots, x_s$ *be independent random variables drawn from* μ *and* μ_s *be the uniform measure on* $x := \{x_1, \dots, x_s\}$*. If*

$$s \ge \frac{C}{\epsilon^2}\left(\left(\int_{c\epsilon}^{\infty} \sqrt{\mathrm{fat}_\gamma(\mathcal{F})} d\gamma\right)^2 + \log 1/\delta\right),$$

then

$$\mathbb{P}\left[\sup_{f \in \mathcal{F}} \left|\mathbb{E}_{\mu_s} f(x_i) - \mathbb{E}_\mu f\right| \ge \epsilon\right] \le \delta.$$

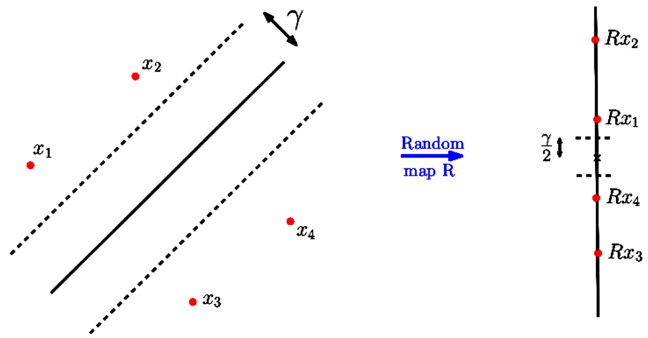

FIGURE 3. Random projections are likely to preserve linear separations.

A key component in the proof of the uniform bound in Theorem 1 is an upper bound on the fat shattering dimension of functions given by the maximum of a set of minima of collections of linear functions on a ball in $\mathcal{H}$. We will use the Johnson-Lindenstrauss lemma [29] in its proof.

Let J be a finite dimensional vector space of dimension greater than or equal to g. In what follows, by "uniformly random g dimensional subspace in J," we mean a random variable taking values in the set of g dimensional subspaces of J, possessing the property that its distribution is invariant under the action of the orthogonal group acting on J.

Johnson-Lindenstrauss lemma: Let $y_1, \dots, y_{\bar{\ell}}$ be points in the unit ball in $\mathbb{R}^m$ for some finite m. Let R be an orthogonal projection (see Figure 3) onto a random g dimensional subspace (where $g = C\frac{\log \bar{\ell}}{\gamma^2}$ for some $\gamma > 0$, and an absolute constant $C > 1$). Then,

$$\mathbb{P}\left[\sup_{i,j\in\{1,\dots,\bar{\ell}\}}\left|\left(\frac{m}{g}\right)\langle Ry_i, Ry_j\rangle - \langle y_i, y_j\rangle\right| > \frac{\gamma}{2}\right] < \frac{1}{2}.$$

Lemma 6. *Let μ be a probability distribution supported on $B_{\mathcal{H}} := \{x \in \mathcal{H} : \|x\| \leq 1\}$. Let $x_1, \dots, x_s$ be independent random variables drawn from μ and μ_s be the uniform measure on $x := \{x_1, \dots, x_s\}$. Let $\mathcal{F}_{k,\ell}$ be the set of all functions f from $\mathcal{H}$ to $\mathbb{R}$, such that for some $k\ell$ vectors $v_{11}, \dots, v_{k\ell} \in B_{\mathcal{H}}$,*

$$f(x) = \max_j \min_i (v_{ij} \cdot x).$$

Then,

(1) $\mathrm{fat}_\gamma(\mathcal{F}_{k,\ell}) \leq \frac{Ck\ell}{\gamma^2}\log^2\frac{Ck\ell}{\gamma^2}$.
(2) *If* $s \geq \frac{C}{\epsilon^2}\left(k\ell\ln^4(k\ell/\epsilon^2) + \ln 1/\delta\right)$, *then* $\mathbb{P}[\sup_{f\in\mathcal{F}_{k,\ell}}\left|\mathbb{E}_{\mu_s}f(x_i) - \mathbb{E}_\mu f\right| \geq \epsilon] \leq \delta$.

The proof of this lemma has been shifted to the Appendix.

In order to prove Theorem 1, we relate the empirical squared loss $s^{-1}\sum_{i=1}^s \mathbf{d}(x_i, \mathcal{M})^2$ and the expected squared loss over a class of manifolds whose covering numbers at a scale ϵ have a specified upper bound. Let $U : \mathbb{R}^+ \to \mathbb{Z}^+$ be a real-valued function. Let $\tilde{\mathcal{G}}$ be any family of subsets of the unit ball $B_{\mathcal{H}}$ in a Hilbert space $\mathcal{H}$ such that for all $r > 0$ every element $\mathcal{M} \in \tilde{\mathcal{G}}$ can be covered using $U(\frac{1}{r})$ open Euclidean balls.

A priori it is unclear if

$$\sup_{\mathcal{M}\in\tilde{\mathcal{G}}}\left|\frac{\sum_{i=1}^{s}\mathbf{d}(x_i,\mathcal{M})^2}{s}-\mathbb{E}_{\mathcal{P}}\mathbf{d}(x,\mathcal{M})^2\right| \tag{4}$$

is a random variable, since the supremum of a set of random variables is not always a random variable (although if the set is countable, this is true). Let $\mathbf{d}_{\mathtt{haus}}$ represent the Hausdorff distance. For each $n \geq 1$, $\tilde{\mathcal{G}}_n$ is a countable set of finite subsets of $\mathcal{H}$, such that for each $\mathcal{M} \in \tilde{\mathcal{G}}$, there exists $\mathcal{M}' \in \tilde{\mathcal{G}}_n$ such that $\mathbf{d}_{\mathtt{haus}}(\mathcal{M}, \mathcal{M}') \leq 1/n$, and for each $\mathcal{M}' \in \tilde{\mathcal{G}}_n$, there is an $\mathcal{M}'' \in \tilde{\mathcal{G}}$ such that $\mathbf{d}_{\mathtt{haus}}(\mathcal{M}'', \mathcal{M}') \leq 1/n$. For each n, such a $\tilde{\mathcal{G}}_n$ exists because $\mathcal{H}$ is separable. Now (4) is equal to

$$\lim_{n\to\infty}\sup_{\mathcal{M}'\in\tilde{\mathcal{G}}_n}\left|\frac{\sum_{i=1}^{s}\mathbf{d}(x_i,\mathcal{M}_n)^2}{s}-\mathbb{E}_{\mathcal{P}}\mathbf{d}(x,M_n)^2\right|,$$

and for each n, the supremum in the limits is over a countable set; thus, for a fixed n, the quantity in the limits is a random variable. Since the pointwise limit of a sequence of measurable functions (random variables) is a measurable function (random variable), this proves that

$$\sup_{\mathcal{M}\in\tilde{\mathcal{G}}}\left|\frac{\sum_{i=1}^{s}\mathbf{d}(x_i,\mathcal{M})^2}{s}-\mathbb{E}_{\mathcal{P}}\mathbf{d}(x,\mathcal{M})^2\right|$$

is a random variable.

Lemma 7. *Let ϵ and δ be error parameters. Let $U_{\mathcal{G}} : \mathbb{R}^+ \to \mathbb{R}^+$ be a function taking values in the positive reals. Suppose every $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ can be covered by the union of some $U_{\mathcal{G}}(\frac{1}{r})$ open Euclidean balls of radius $\frac{\sqrt{r\tau}}{16}$, for every $r > 0$. If*

$$s \geq C\left(\frac{U_{\mathcal{G}}(1/\epsilon)}{\epsilon^2}\left(\log^4\left(\frac{U_{\mathcal{G}}(1/\epsilon)}{\epsilon}\right)\right)+\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right),$$

then

$$\mathbb{P}\left[\sup_{\mathcal{M}\in\mathcal{G}(d,V,\tau)}\left|\frac{\sum_{i=1}^{s}\mathbf{d}(x_i,\mathcal{M})^2}{s}-\mathbb{E}_{\mathcal{P}}\mathbf{d}(x,\mathcal{M})^2\right|<\epsilon\right]>1-\delta.$$

Proof. Given a collection $\mathbf{c} := \{c_1, \ldots, c_k\}$ of points in $\mathcal{H}$, let

$$f_{\mathbf{c}}(x) := \min_{c_j\in\mathbf{c}}|x-c_j|^2. \tag{5}$$

Let $\mathcal{F}_k$ denote the set of all such functions for

$$\mathbf{c} = \{c_1, \ldots, c_k\} \subseteq B_{\mathcal{H}},$$

$B_{\mathcal{H}}$ being the unit ball in the Hilbert space.

Consider $\mathcal{M} \in \mathcal{G} := \mathcal{G}(d, V, \tau)$. Let $\mathbf{c}(\mathcal{M}, \epsilon) = \{\hat{c}_1, \ldots, \hat{c}_{\hat{k}}\}$ be a set of $\hat{k} := U_{\mathcal{G}}(1/\epsilon)$ points in $\mathcal{M}$, such that $\mathcal{M}$ is contained in the union of Euclidean balls of radius $\sqrt{\tau\epsilon}/16$ centered at these points. Suppose $x \in B_{\mathcal{H}}$. Since $c(\mathcal{M}, \epsilon) \subseteq \mathcal{M}$, we have $\mathbf{d}(x, \mathcal{M}) \leq \mathbf{d}(x, c(\mathcal{M}, \epsilon))$. To obtain a bound in the reverse direction, let $y \in \mathcal{M}$ be a point such that $|x - y| = \mathbf{d}(x, \mathcal{M})$, and let $z \in \mathbf{c}(\mathcal{M}, \epsilon)$ be a point such that $|y - z| < \sqrt{\tau\epsilon}/16$. Let z' be the point on $Tan(y, \mathcal{M})$ that is closest to z. By the reach condition, and Proposition 1,

$$\begin{aligned} |z - z'| &= \mathbf{d}(z, Tan(y, \mathcal{M})) \\ &\leq \frac{|y-z|^2}{2\tau} \\ &\leq \frac{\epsilon}{512}. \end{aligned}$$

Therefore,

$$\begin{aligned} 2\langle y - z, x - y\rangle &= 2\langle y - z' + z' - z, x - y\rangle \\ &= 2\langle z' - z, x - y\rangle \\ &\leq 2|z - z'||x - y| \\ &\leq \frac{\epsilon}{128}. \end{aligned}$$

In the last line above, we use the fact that both x and $\mathcal{M} \ni y$ belong to the unit ball and hence $|x - y| \leq 2$.

Thus

$$\begin{aligned} \mathbf{d}(x, \mathbf{c}(\mathcal{M}, \epsilon))^2 &\leq |x - z|^2 \\ &\leq |x - y|^2 + 2\langle y - z, x - y\rangle + |y - z|^2 \\ &\leq \mathbf{d}(x, \mathcal{M})^2 + \tfrac{\epsilon}{128} + \tfrac{\epsilon\tau}{256}. \end{aligned}$$

Since $\tau < 1$, this shows that

$$\mathbf{d}(x, \mathcal{M})^2 \leq \mathbf{d}(x, \mathbf{c}(\mathcal{M}, \epsilon))^2 \leq \mathbf{d}(x, \mathcal{M})^2 + \frac{\epsilon}{64}.$$

Note that

$$\mathbf{d}(x, \mathbf{c}(\mathcal{M}, \epsilon))^2 = f_{\mathbf{c}(\mathcal{M},\epsilon)}(x).$$

Therefore,

$$\begin{aligned} &\mathbb{P}\left[\sup_{\mathcal{M} \in \mathcal{G}} \left| \frac{\sum_{i=1}^{s} \mathbf{d}(x_i, \mathcal{M})^2}{s} - \mathbb{E}_{\mathcal{P}} \mathbf{d}(x, \mathcal{M})^2 \right| < \epsilon \right] \\ &\geq \mathbb{P}\left[\sup_{f_{\mathbf{c}} \in \mathcal{F}_{\hat{k}}} \left| \frac{\sum_{i=1}^{s} f_{\mathbf{c}}(x_i)}{s} - \mathbb{E}_{\mathcal{P}} f_{\mathbf{c}}(x_i) \right| < \frac{\epsilon}{3} \right]. \end{aligned} \tag{6}$$

Inequality (6) reduces the problem of deriving uniform bounds over a space of manifolds to a problem of deriving uniform bounds for k-means.

Let

$$\Phi : \mathbf{x} \mapsto 2^{-1/2}(\mathbf{x}, 1)$$

map a point $x \in \mathcal{H}$ to one in $\mathcal{H} \oplus \mathbb{R}$, which we equip with the natural Hilbert space structure. For each i, let

$$\tilde{c}_i := 2^{-1/2}(-c_i, \frac{\|c_i\|^2}{2}). \tag{7}$$

The factor of $2^{-1/2}$ (which could have been replaced by a slightly larger constant) is present because we want $\tilde{c}_i$ to belong to the unit ball. Then,

$$f_{\mathbf{c}}(x) = |x|^2 + 4\min(\langle \Phi(x), \tilde{c}_1\rangle, \ldots, \langle \Phi(x), \tilde{c}_k\rangle).$$

Let $\mathcal{F}_\Phi$ be the set of functions that map $\mathcal{H}$ to $\mathbb{R}$ having the form $4\min_{i=1}^k \Phi(x)\cdot\tilde{c}_i$ where $\tilde{c}_i$ is given by (7) and

$$\mathbf{c} = \{c_1, \dots, c_k\} \subseteq B_{\mathcal{H}}.$$

The metric entropy of the function class obtained by translating $\mathcal{F}_\Phi$ by adding $|x|^2$ to every function in it is the same as the metric entropy of $\mathcal{F}_\Phi$. However, this translated function class has the unit ball of a separable Hilbert space as its domain as well.

Therefore the integral of the square root of the metric entropy of functions of the form (5) in $\mathcal{F}_k$ can be bounded above, and by Lemma 6, if

$$s \geq C\left(\frac{k}{\epsilon^2}\left(\log^4\left(\frac{k}{\epsilon}\right)\right) + \frac{1}{\epsilon^2}\log\frac{1}{\delta}\right),$$

then

$$\mathbb{P}\left[\sup_{\mathcal{M}\in\mathcal{G}}\left|\frac{\sum_{i=1}^s \mathbf{d}(x_i,\mathcal{M})^2}{s} - \mathbb{E}_{\mathcal{P}}\mathbf{d}(x,\mathcal{M})^2\right| < \epsilon\right] > 1-\delta. \quad \square$$

Proof of Theorem 1. This follows immediately from Corollary 2 and Lemma 7. $\square$

5. Fitting k affine subspaces of dimension d

A natural generalization of k-means was proposed in [6] wherein one fits k d-dimensional planes to data in a manner that minimizes the average squared distance of a data point to the nearest d dimensional plane. For more recent results on this kind of model, with the average pth powers rather than squares, see [34]. We can view k-means as a zero dimensional special case of k-planes.

In this section, we derive an upper bound for the generalization error of fitting k-planes. Unlike the earlier bounds for fitting manifolds, the bounds here are linear in the dimension d rather than exponential in it. The dependence on k is linear up to logarithmic factors, as before. In the section, we assume for notation convenience that the dimension m of the Hilbert space is finite, though the results can be proved for any separable Hilbert space.

Let $\mathcal{P}$ be a probability distribution supported on $B := \{x \in \mathbb{R}^m \big| \, \|x\| \leq 1\}$. Let $\mathbb{H} := \mathbb{H}_{k,d}$ be the set whose elements are unions of not more than k affine subspaces of dimension $\leq d$, each of which intersects B. Let $\mathcal{F}_{k,d}$ be the set of all loss functions $F(x) = \mathbf{d}(x,H)^2$ for some $H \in \mathbb{H}$ (where $\mathbf{d}(x,S) := \inf_{y\in S}\|x-y\|$).

We wish to obtain a probabilistic upper bound on

$$\sup_{F\in\mathcal{F}_{k,d}}\left|\frac{\sum_{i=1}^s F(x_i)}{s} - \mathbb{E}_{\mathcal{P}}F(x)\right|, \tag{8}$$

where $\{x_i\}_1^s$ is the train set and $\mathbb{E}_{\mathcal{P}}F(x)$ is the expected value of F with respect to $\mathcal{P}$. Due to issues of measurability, (8) need not be a random variable for arbitrary $\mathcal{F}$. However, in our situation, this is the case because $\mathcal{F}$ is a family of bounded piecewise quadratic functions, continuously parameterized by $\mathbb{H}_{k,d}$, which has a countable dense subset, for example, the subset of elements specified using rational data. We obtain a bound that is independent of m, the ambient dimension. We will need the following form of Hoeffding's inequality.

Lemma 8 (Hoeffding's inequality). *Let $X_1, \dots, X_s$ be i.i.d. copies of the random variable X whose range is $[0,1]$. Then,*

$$\mathbb{P}\left[\left|\frac{1}{s}\left(\sum_{i=1}^{s} X_i\right) - \mathbb{E}[X]\right| \leq \epsilon\right] \geq 1 - 2\exp(-2s\epsilon^2). \tag{9}$$

Lemma 9. *Let $x_1, \dots, x_s$ be i.i.d. samples from $\mathcal{P}$, a distribution supported on the ball of radius 1 in $\mathbb{R}^m$. If*

$$s \geq C\left(\frac{dk}{\epsilon^2}\log^4\left(\frac{dk}{\epsilon}\right) + \frac{d}{\epsilon^2}\log\frac{1}{\delta}\right),$$

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

Proof. Any $F \in \mathcal{F}_{k,d}$ can be expressed as $F(x) = \min_{1\leq i\leq k} \mathbf{d}(x, H_i)^2$ where each H_i is an affine subspace of dimension less than or equal to d that intersects the unit ball. In turn, $\min_{1\leq i\leq k} \mathbf{d}(x, H_i)^2$ can be expressed as

$$\min_{1\leq i\leq k}\left(\|x - c_i\|^2 - (x-c_i)^\dagger A_i^\dagger A_i(x-c_i)\right),$$

where A_i is defined by the condition that for any vector z, $(z - (A_i z))^\dagger$ and $A_i z$ are the components of z parallel and perpendicular to H_i, and c_i is the point on H_i that is the nearest to the origin (it could have been any point on H_i). Thus

$$F(x) := \min_i\left(\|x\|^2 - 2c_i^\dagger x + \|c_i\|^2 - x^\dagger A_i^\dagger A_i x + 2c_i^\dagger A_i^\dagger A_i x - c_i^\dagger A_i^\dagger A_i c_i\right).$$

Now, define vector valued maps Φ and Ψ whose respective domains are the space of d dimensional affine subspaces and $\mathcal{H}$, respectively,

$$\Phi(H_i) := \left(\frac{1}{\sqrt{d+5}}\right)\left(\|c_i\|^2, A_i^\dagger A_i, (2A_i^\dagger A_i c_i - 2c_i)^\dagger\right)$$

and

$$\Psi(x) := \left(\frac{1}{\sqrt{3}}\right)(1, xx^\dagger, x^\dagger),$$

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

Thus,

$$\min_i\left(\|x\|^2 - 2c_i^\dagger x + \|c_i\|^2 - x^\dagger A_i^\dagger A_i x + 2c_i^\dagger A_i^\dagger A_i x - c_i^\dagger A_i^\dagger A_i c_i\right)$$

is equal to

$$\|x\|^2 + \sqrt{3(d+5)}\min_i \Phi(H_i)\cdot\Psi(x).$$

We see that since $\|x\| \leq 1$, $\|\Psi(x)\| \leq 1$. The Frobenius norm $\|A_i^\dagger A_i\|^2$ is equal to $Tr(A_i A_i^\dagger A_i A_i^\dagger)$, which is the rank of A_i since A_i is a projection. Therefore,

$$(d+5)\|\Phi(H_i)\|^2 \leq \|c_i\|^4 + \|A_i^\dagger A_i\|^2 + \|(2(I - A_i^\dagger A_i)c_i\|^2,$$

which is less than or equal to $d+5$.

Uniform bounds for classes of functions of the form $\min_i \Phi(H_i)\cdot\Psi(x)$ follow from Lemma 6. We infer from Lemma 6 and the Hoeffding's inequality that if

$$s \geq C\left(\frac{k}{\epsilon^2}\log^4\left(\frac{k}{\epsilon}\right) + \frac{1}{\epsilon^2}\log\frac{1}{\delta}\right),$$

then $\mathbb{P}\left[\sup_{F\in\mathcal{F}_{k,d}}\left|\frac{\sum_{i=1}^{s}F(x_i)}{s} - \mathbb{E}_{\mathcal{P}}F(x)\right| < \sqrt{3(d+5)}\epsilon\right] > 1-\delta$. The last statement can be rephrased as follows. If

$$s \geq C\left(\frac{dk}{\epsilon^2}\log^4\left(\frac{dk}{\epsilon}\right) + \frac{d}{\epsilon^2}\log\frac{1}{\delta}\right),$$

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

6. Dimension reduction

Suppose that $X = \{x_1, \dots, x_s\}$ is a set of i.i.d. random points drawn from $\mathcal{P}$, a probability measure supported on the unit ball $B_{\mathcal{H}}$ of a separable Hilbert space $\mathcal{H}$. Let $\mathcal{M}_{erm}(X)$ denote a manifold in $\mathcal{G}(d, V, \tau)$ that (approximately) minimizes

$$\sum_{i=1}^{s}\mathbf{d}(x_i, \mathcal{M})^2$$

over all $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ and denote by $\mathcal{P}_X$ the probability distribution on X that assigns a probability of $1/s$ to each point. More precisely, we know from Theorem 1 that there is some function $s_{\mathcal{G}}(\epsilon, \delta)$ of ϵ, δ, d, V, and τ such that if

$$s \geq s_{\mathcal{G}}(\epsilon, \delta),$$

then

$$\mathbb{P}\left[\mathcal{L}(\mathcal{M}_{erm}(X), \mathcal{P}_X) - \inf_{\mathcal{M}\in\mathcal{G}}\mathcal{L}(\mathcal{M}, \mathcal{P}) < \epsilon\right] > 1-\delta. \tag{10}$$

Lemma 10. *Suppose $\epsilon < c\tau$. Let W denote an arbitrary $2s_{\mathcal{G}}(\epsilon, \delta)$ dimensional linear subspace of $\mathcal{H}$ containing X. Then*

$$\inf_{\mathcal{G}(d,V,\tau(1-c))\ni\mathcal{M}\subseteq W}\mathcal{L}(\mathcal{M}, \mathcal{P}_X) \leq C\epsilon + \inf_{\mathcal{M}\in\mathcal{G}(d,V,\tau)}\mathcal{L}(\mathcal{M}, \mathcal{P}_X). \tag{11}$$

Proof. Let $\mathcal{M}_2 \in \mathcal{G} := \mathcal{G}(d, V, \tau)$ achieve

$$\mathcal{L}(\mathcal{M}_2, \mathcal{P}_X) \leq \inf_{\mathcal{M}\subseteq\mathcal{G}}\mathcal{L}(\mathcal{M}, \mathcal{P}_X) + \epsilon. \tag{12}$$

Let N_ϵ denote a set of no more than $s_G(\epsilon, \delta)$ points contained in $\mathcal{M}_2$ that is an ϵ-net of $\mathcal{M}_2$. Thus for every $x \in \mathcal{M}_2$, there is $y \in N_\epsilon$ such that $|x - y| < \epsilon$. Let O denote a unitary transformation from $\mathcal{H}$ to $\mathcal{H}$ that fixes each point in X and maps every point in N_ϵ to some point in W. Let Π_W denote the map from $\mathcal{H}$ to W that maps x to the point in W nearest to x. Let $\mathcal{M}_3 := O\mathcal{M}_2$. Since O is an isometry that fixes X,

$$\mathcal{L}(\mathcal{M}_3, \mathcal{P}_X) = \mathcal{L}(\mathcal{M}_2, \mathcal{P}_X) \leq \inf_{\mathcal{M}\subseteq\mathcal{G}}\mathcal{L}(\mathcal{M}, \mathcal{P}_X) + \epsilon. \tag{13}$$

Since $\mathcal{P}_X$ is supported on the unit ball and the Hausdorff distance between $\Pi_W\mathcal{M}_3$ and $\mathcal{M}_3$ is at most ϵ,

$$\begin{aligned}\left|\mathcal{L}(\Pi_W\mathcal{M}_3, \mathcal{P}_X) - \mathcal{L}(\mathcal{M}_3, \mathcal{P}_X)\right| &\leq \mathbb{E}_{x\dashv\mathcal{P}_X}\left|\mathbf{d}(x, \Pi_W\mathcal{M}_3)^2 - \mathbf{d}(x, \mathcal{M}_3)^2\right| \\ &\leq \mathbb{E}_{x\dashv\mathcal{P}_X}4\left|\mathbf{d}(x, \Pi_W\mathcal{M}_3) - \mathbf{d}(x, \mathcal{M}_3)\right| \\ &\leq 4\epsilon.\end{aligned}$$

By Lemma 11, we see that $\Pi_W\mathcal{M}_3$ belongs to $\mathcal{G}(d, V, \tau(1-c))$, thus proving the lemma. □

By Lemma 10, it suffices to find a manifold $\mathcal{G}(d, V, \tau) \ni \tilde{M}_{erm}(X) \subseteq V$ such that

$$\mathcal{L}(\tilde{M}_{erm}(X), \mathcal{P}_X) \leq C\epsilon + \inf_{V \supseteq \mathcal{M} \in \mathcal{G}(d,V,\tau)} \mathcal{L}(\mathcal{M}, \mathcal{P}_X).$$

Lemma 11. *Let $\mathcal{M} \in \mathcal{G}(d, V, \tau)$, and let Π be a map that projects $\mathcal{H}$ orthogonally onto a subspace containing the linear span of a $c\epsilon\tau$-net $\bar{S}$ of $\mathcal{M}$. Then, the image of $\mathcal{M}$, is a d dimensional submanifold of $\mathcal{H}$ and*

$$\Pi(\mathcal{M}) \in \mathcal{G}(d, V, \tau(1 - C\sqrt{\epsilon})).$$

Proof. The volume of $\Pi(\mathcal{M})$ is no more than the volume of $\mathcal{M}$ because Π is a contraction. Since $\mathcal{M}$ is contained in the unit ball, $\Pi(\mathcal{M})$ is contained in the unit ball.

Claim 2. *For any $x, y \in \mathcal{M}$,*

$$|\Pi(x - y)| \geq (1 - C\sqrt{\epsilon})|x - y|.$$

Proof. First suppose that $|x - y| < \sqrt{\epsilon}\tau$. Choose $\tilde{x} \in \bar{S}$ that satisfies

$$|\tilde{x} - x| < C_1\epsilon\tau.$$

Let $z := x + \frac{(y-x)\sqrt{\epsilon}\tau}{|y-x|}$. By linearity and Proposition 1,

$$\mathbf{d}(z, Tan(x, \mathcal{M})) = \mathbf{d}(y, Tan(x, \mathcal{M}))\left(\frac{\sqrt{\epsilon}\tau}{|y - x|}\right) \tag{14}$$

$$\leq \frac{|x - y|^2}{2\tau}\left(\frac{\sqrt{\epsilon}\tau}{|y - x|}\right) \tag{15}$$

$$\leq \frac{\epsilon\tau}{2}. \tag{16}$$

Therefore, there is a point $\hat{y} \in Tan(x, \mathcal{M})$ such that

$$\left|\hat{y} - \left(\tilde{x} + \frac{(y - x)\sqrt{\epsilon}\tau}{|y - x|}\right)\right| \leq C_2\epsilon\tau.$$

By Claim 1, there is a point $\bar{y} \in \mathcal{M}$ such that

$$\left|\bar{y} - \hat{y}\right| \leq C_3\epsilon\tau.$$

Let $\tilde{y} \in \bar{S}$ satisfy

$$|\tilde{y} - \bar{y}| < c\epsilon\tau.$$

Then,

$$\left|\tilde{y} - \left(\tilde{x} + \frac{(y - x)\sqrt{\epsilon}\tau}{|y - x|}\right)\right| \leq C_4\epsilon\tau,$$

i.e.,

$$\left|\left(\frac{\tilde{y} - \tilde{x}}{\sqrt{\epsilon}\tau}\right) - \frac{(y - x)}{|y - x|}\right| \leq C_4\sqrt{\epsilon}.$$

Consequently,

$$\left|\left(\frac{\tilde{y} - \tilde{x}}{\sqrt{\epsilon}\tau}\right)\right| - 1 \leq C_4\sqrt{\epsilon}. \tag{17}$$

We now have

$$\left\langle \frac{y-x}{|y-x|}, \frac{\tilde{y}-\tilde{x}}{\sqrt{\epsilon}\tau} \right\rangle = \left\langle \frac{y-x}{|y-x|}, \frac{y-x}{|y-x|} \right\rangle + \left\langle \frac{y-x}{|y-x|}, \left(\frac{\tilde{y}-\tilde{x}}{\sqrt{\epsilon}\tau} - \frac{y-x}{|y-x|} \right) \right\rangle \tag{18}$$

$$= 1 + \left\langle \frac{y-x}{|y-x|}, \left(\frac{\tilde{y}-\tilde{x}}{\sqrt{\epsilon}\tau} - \frac{y-x}{|y-x|} \right) \right\rangle \tag{19}$$

$$\geq 1 - C_4\sqrt{\epsilon}. \tag{20}$$

Since $\tilde{x}$ and $\tilde{y}$ belong to the range of Π, it follows from (17) and (20) that

$$|\Pi(x-y)| \geq (1 - C\sqrt{\epsilon})|x-y|.$$

Next, suppose that $|x-y| \geq \sqrt{\epsilon}\tau$. Choose $\tilde{x}, \tilde{y} \in \bar{S}$ such that $|x-\tilde{x}| + |y-\tilde{y}| < 2c\epsilon\tau$. Then,

$$\begin{aligned} \left\langle \frac{x-y}{|x-y|}, \frac{\tilde{x}-\tilde{y}}{|\tilde{x}-\tilde{y}|} \right\rangle &= \left\langle \frac{x-y}{|x-y|}, \frac{x-y}{|\tilde{x}-\tilde{y}|} \right\rangle + \left(|\tilde{x}-\tilde{y}|^{-1}\right) \\ &\quad \left\langle \frac{x-y}{|x-y|}, (\tilde{x}-x) - (\tilde{y}-y) \right\rangle \\ &\geq 1 - C\sqrt{\epsilon}, \end{aligned}$$

and the claim follows since $\tilde{x}$ and $\tilde{y}$ belong to the range of Π. □

By Claim 2, we see that

$$\forall\, x \in \mathcal{M},\ Tan^0(x, \mathcal{M}) \cap \ker(\Pi) = \{0\}. \tag{21}$$

Moreover, by Claim 2, we see that if $x, y \in \mathcal{M}$ and $\Pi(x)$ is close to $\Pi(y)$ then x is close to y. Therefore, to examine all $\Pi(x)$ in a neighborhood of $\Pi(y)$, it is enough to examine all x in a neighborhood of y. So by Definition 3, it follows that $\Pi(\mathcal{M})$ is a submanifold of $\mathcal{H}$. Finally, in view of Claim 2 and the fact that Π is a contraction, we see that

$$\operatorname{reach}(\Pi(\mathcal{M})) = \sup_{x,y\in\mathcal{M}} \frac{|\Pi(x)-\Pi(y)|^2}{2\mathbf{d}(\Pi(x), Tan(\Pi(y), \Pi(\mathcal{M})))} \tag{22}$$

$$\geq (1 - C\sqrt{\epsilon}) \sup_{x,y\in\mathcal{M}} \frac{|x-y|^2}{2\mathbf{d}(x, Tan(y, \mathcal{M}))} \tag{23}$$

$$= (1 - C\sqrt{\epsilon}) \operatorname{reach}(\mathcal{M}), \tag{24}$$

and the lemma follows. □

7. Overview of the algorithm for testing the manifold hypothesis

Given a set $X := \{x_1, \dots, x_s\}$ of points in $\mathbb{R}^n$, we give an overview of the algorithm that finds a nearly optimal interpolating manifold.

Definition 10. *Let $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ be called an ϵ-optimal interpolant if*

$$\sum_{i=1}^{s} \mathbf{d}(x_i, \mathcal{M})^2 \leq s\epsilon + \inf_{\mathcal{M}' \in \mathcal{G}(d, V/C, C\tau)} \sum_{i=1}^{s} \mathbf{d}(x_i, \mathcal{M}')^2, \tag{25}$$

where C is some constant depending only on d.

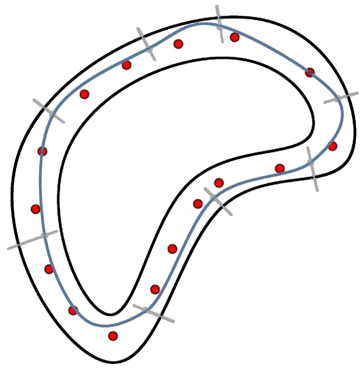

FIGURE 4. A disc bundle $D^{\mathtt{norm}} \in \bar{\mathcal{D}}^{\mathtt{norm}}$.

Given d, τ, V, ϵ, and δ, our goal is to output an implicit representation of a manifold $\mathcal{M}$ and an estimated error $\bar{\epsilon} \geq 0$ such that

(1) With probability greater than $1-\delta$, $\mathcal{M}$ is an ϵ-optimal interpolant, and
(2)
$$s\bar{\epsilon} \leq \sum_{x \in X} \mathbf{d}(x, \mathcal{M})^2 \leq s\left(\frac{\epsilon}{2} + \bar{\epsilon}\right).$$

Thus, we are required to perform an optimization over the set of manifolds $\mathcal{G} = \mathcal{G}(d, \tau, V)$. This set $\mathcal{G}$ can be viewed as a metric space $(G, \mathbf{d}_{\mathtt{haus}})$ by defining the distance between two manifolds $\mathcal{M}, \mathcal{M}'$ in $\mathcal{G}$ to be the Hausdorff distance between $\mathcal{M}$ and $\mathcal{M}'$. Our strategy for producing an approximately optimal manifold will be to execute the following steps. First identify a $O(\tau)$-net $S_{\mathcal{G}}$ of $(\mathcal{G}, \mathbf{d}_{\mathtt{haus}})$. Next, for each $\mathcal{M}' \in S_{\mathcal{G}}$, construct a disc bundle D' that approximates its normal bundle. The fiber of D' at a point $z \in \mathcal{M}'$ is a $n-d$ dimensional disc of radius $O(\tau)$ that is roughly orthogonal to $Tan(z, \mathcal{M}')$ (this is formalized in Definitions 11 and 12, see Figure 4). Suppose that $\mathcal{M}$ is a manifold in $\mathcal{G}$ such that

$$\mathbf{d}_{\mathtt{haus}}(\mathcal{M}, \mathcal{M}') < O(\tau). \tag{26}$$

As a consequence of (26) and the lower bounds on the reaches of $\mathcal{M}$ and $\mathcal{M}'$, it follows (as will be shown in Lemma 17) that $\mathcal{M}$ must be the graph of a section of D'. In other words, $\mathcal{M}$ intersects each fiber of D' in a unique point. We use convex optimization to find good local sections, and patch them up to find a good global section. Thus, our algorithm involves two main phases:

(1) Construct a set $\bar{\mathcal{D}}^{\mathtt{norm}}$ of disc bundles over manifolds in $\mathcal{G}(d, CV, \tau/C)$ which is rich enough that every ϵ-optimal interpolant is a section of some member of $\bar{\mathcal{D}}^{\mathtt{norm}}$.
(2) Given $D^{\mathtt{norm}} \in \bar{\mathcal{D}}^{\mathtt{norm}}$, use convex optimization to find a minimal $\hat{\epsilon}$ such that $D^{\mathtt{norm}}$ has a section (i.e., a small transverse perturbation of the base manifold of $D^{\mathtt{norm}}$) which is a $\hat{\epsilon}$-optimal interpolant. This is achieved by finding the right manifold in the vicinity of the base manifold of $D^{\mathtt{norm}}$ by finding good local sections (using results from [23, 24]) and then patching these up using a gentle partition of unity supported on the base manifold of $D^{\mathtt{norm}}$.

8. Disc bundles

The following definition specifies the kind of bundles we will be interested in. The constants have been named so as to be consistent with their appearance in (79) and Observation 3. Recall the parameter r from Definition 3.

Definition 11. *Let D be an open subset of $\mathbb{R}^n$ and $\mathcal{M}$ be a submanifold of D that belongs to $\mathcal{G}(d,\tau,V)$ for some choice of parameters d,τ,V. Let π be a $\mathcal{C}^4$ map $\pi : D \to \mathcal{M}$ such that for any $z \in \mathcal{M}$, $\pi(z) = z$ and $\pi^{-1}(z)$ is isometric to a Euclidean disc of dimension $n-d$, of some radius independent of z. We then say $D \xrightarrow{\pi} \mathcal{M}$ is a disc bundle. When $\mathcal{M}$ is clear from context, we will simply refer to the bundle as D. We refer to $D_z := \pi^{-1}(z)$ as the fiber of D at z. We call $s : \mathcal{M} \to D$ a section of D if for any $z \in \mathcal{M}$, $s(z) \in D_z$ and for some $\hat{\tau}, \hat{V} > 0$, $s(\mathcal{M}) \in \mathcal{G}(d,\hat{\tau},\hat{V})$. Let U be an open subset of $\mathcal{M}$. We call a given C^2-map $s_{loc} : U \to D$ a local section of D if for any $z \in U$, $s(z) \in D_z$ and $\{(z, s_{loc}(z)) | z \in U\}$ can locally be expressed as the graph of a $\mathcal{C}^2$-function.*

Definition 12. *For reals $\hat{\tau}, \hat{V} > 0$, let $\bar{\mathcal{D}}(d,\hat{\tau},\hat{V})$ denote the set of all disc bundles $D^{norm} \xrightarrow{\pi} \mathcal{M}$ with the following properties:*

1. *D^{norm} is a disc bundle over the manifold $\mathcal{M} \in \mathcal{G}(d,\hat{\tau},\hat{V})$.*
2. *Let $z_0 \in \mathcal{M}$. For $z_0 \in \mathcal{M}$, let $D^{norm}_{z_0} := \pi^{-1}(z_0)$ denote the fiber over z_0, and Π_{z_0} denote the projection of $\mathbb{R}^n$ onto the affine span of $D^{norm}_{z_0}$. Without loss of generality assume after rotation (if necessary) that $Tan(z_0,\mathcal{M}) = \mathbb{R}^d \oplus \{0\}$ and $Nor_{z_0,\mathcal{M}} = \{0\} \oplus \mathbb{R}^{n-d}$. Then, $D^{norm} \cap B(z_0, \bar{c}_{11}\hat{\tau})$ is a bundle over a graph $\{(z,\Psi(z))\}_{z\in\Omega_{z_0}}$ where the domain Ω_{z_0} is an open subset of $Tan(z_0,\mathcal{M})$.*
3. *Any $z \in B_n(z_0,\bar{c}_{11}\hat{\tau})$ may be expressed uniquely in the form $(x,\Psi(x)) + v$ with $x \in B_d(z_0,\bar{c}_{10}\hat{\tau}), v \in \Pi_{(x,\Psi(x))}B_{n-d}(x, \frac{\bar{c}_{10}\hat{\tau}}{2})$. Moreover, x and v here are $\mathcal{C}^{k-2}$-smooth functions of $z \in B_n(x,\bar{c}_{11}\hat{\tau})$, with derivatives up to order $k-2$ bounded by C in absolute value.*
4. *Let $x \in B_d(z_0,\bar{c}_{10}\hat{\tau})$, and let $v \in \Pi_{(x,\Psi(x))}\mathbb{R}^n$. Then, we can express v in the form*

$$v = \Pi_{(x,\Psi(x))} v^{\#}, \tag{27}$$

where $v^{\#} \in \{0\} \oplus \mathbb{R}^{n-d}$ and $|v^{\#}| \le 2|v|$.

Definition 13. *For any $D^{norm} \to \mathcal{M}_{base} \in \bar{\mathcal{D}}(d,\hat{\tau},\hat{V})$, and $\alpha \in (0,1)$, let $\alpha\bar{\mathcal{D}}(d,\hat{\tau},\hat{V})$ denote a bundle over $\mathcal{M}_{base}$, whose every fiber is a scaling by α of the corresponding fiber of D^{norm}.*

9. A key result

Given a function with prescribed smoothness, the following key result allows us to construct a bundle satisfying certain conditions, as well as assert that the base manifold has controlled reach. We decompose $\mathbb{R}^n$ as $\mathbb{R}^d \oplus \mathbb{R}^{n-d}$. The theorem roughly says that given a sufficiently smooth function F with bounded derivatives that resembles the squared distance to the intersection of the ball with a d dimensional subspace, we can construct from it a map Ψ that maps a neighborhood of the origin in $\mathbb{R}^d$ to a neighborhood of the origin in $\mathbb{R}^{n-d}$, whose graph is the intersection of a smooth manifold $\mathcal{M}$ with a neighborhood of the origin in $\mathbb{R}^n$. Further one can also construct a disc bundle over the manifold $\mathcal{M}$ using the large eigenspace of the Hessian of the function F. When we write $(x, y) \in \mathbb{R}^n$, we mean $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^{n-d}$. We will need Taylor's theorem (see [39]), which we state below.

Theorem 12 (Taylor's theorem). *Let Ω be open in $\mathbb{R}^n$, and $f \in \mathcal{C}^k(\Omega)$. Then, if $x, y \in \Omega$ and the closed line segment $[x, y]$ joining x to y is also contained in Ω, we have*

$$f(x) = \sum_{|\alpha| \le k-1} \frac{D^\alpha f(y)}{\alpha!}(x-y)^\alpha + \sum_{|\alpha|=k} \frac{D^\alpha f(\zeta)}{\alpha!}(x-y)^\alpha,$$

where ζ is a point of $[x, y]$.

Theorem 13. *Let the following conditions hold:*

(1) *Suppose $F : B_n(0,1) \to \mathbb{R}$ is $\mathcal{C}^k$-smooth.*

(2)

$$\partial^\alpha_{x,y} F(x, y) \le C_0 \tag{28}$$

for $(x, y) \in B_n(0, 1)$ and $|\alpha| \le k$.

(3) *For $x \in \mathbb{R}^d$, $y \in \mathbb{R}^{n-d}$, and $(x, y) \in B_n(0, 1)$, suppose also that*

$$c_1[|y|^2 + \rho^2] \le [F(x, y) + \rho^2] \le C_1[|y|^2 + \rho^2], \tag{29}$$

where

$$0 < \rho < c, \tag{30}$$

where c is a small enough constant determined by C_0, c_1, C_1, k, n.

Then there exist constants $c_2, \dots, c_7$ and C determined by C_0, c_1, C_1, k, n, such that the following hold:

(1) *For $z \in B_n(0, c_2)$, let $N(z)$ be the subspace of $\mathbb{R}^n$ spanned by the eigenvectors of the Hessian $\partial^2 F(z)$ corresponding to the $(n-d)$ largest eigenvalues. Let $\Pi_{hi}(z) : \mathbb{R}^n \to N(z)$ be the orthogonal projection from $\mathbb{R}^n$ onto $N(z)$. Then $|\partial^\alpha \Pi_{hi}(z)| \le C$ for $z \in B_n(0, c_2), |\alpha| \le k-2$. Thus, $N(z)$ depends $\mathcal{C}^{k-2}$-smoothly on z.*

(2) *There is a map*

$$\Psi : B_d(0, c_4) \to B_{n-d}(0, c_3), \tag{31}$$

with the following properties:

$$|\Psi(0)| \le C\rho; |\partial^\alpha \Psi| \le C^{|\alpha|} \tag{32}$$

on $B_d(0, c_4)$, for $1 \le |\alpha| \le k-2$. The set of all $z = (x, y) \in B_d(0, c_4) \times B_{n-d}(0, c_3)$, such that

$$\left\{z \middle| \Pi_{hi}(z)\partial F(z) = 0\right\} =: \left\{(x, \Psi(x)) \middle| x \in B_d(0, c_4)\right\}$$

is a $\mathcal{C}^{k-2}$-smooth graph.

Proof. We first study the gradient and Hessian of F. Taking $(x, y) = (0, 0)$ in (29), we see that

$$c_1\rho^2 \le F(0, 0) + \rho^2 \le C_1\rho^2. \tag{33}$$

A standard lemma in analysis asserts that non-negative F satisfying (28) must also satisfy

$$\left|\partial F(z)\right| \le C\left(F(z)\right)^{\frac{1}{2}}.$$

In particular, applying this result to the function $F + \rho^2$, we find that

$$\big|\partial F(0,0)\big| \le C\rho. \tag{34}$$

Next, we apply Taylor's theorem: For $(|x|^2+|y|^2)^{\frac{1}{2}} \le \rho^{\frac{2}{3}}$, for $z=(z_1,\dots,z_n)=(x,y)$, estimates (28) and (33) and Taylor's theorem yield

$$\Big|F(x,y)+F(-x,-y)-\sum_{i,j=1}^{n}\partial_{ij}^2 F(0,0)z_iz_j\Big| \le C\rho^2.$$

Hence, (29) implies that

$$c|y|^2 - C\rho^2 \le \sum_{i,j=1}^{n}\partial_{ij}^2 F(0,0)z_iz_j \le C(|y|^2+\rho^2).$$

Therefore,

$$c|y|^2 - C\rho^{2/3}|z|^2 \le \sum_{i,j=1}^{n}\partial_{ij}^2 F(0,0)z_iz_j \le C\left(|y|^2+\rho^{2/3}|z|^2\right)$$

for $|z|=\rho^{2/3}$, and hence for all $z\in\mathbb{R}^n$. Thus, the Hessian matrix $\left(\partial_{ij}^2F(0)\right)$ satisfies

$$\left(\begin{array}{c|c} -C\rho^{2/3}I & 0 \\ \hline 0 & cI\end{array}\right) \preceq \left(\partial_{ij}^2F(0,0)\right) \preceq \left(\begin{array}{c|c} +C\rho^{2/3}I & 0 \\ \hline 0 & CI\end{array}\right). \tag{35}$$

That is, the matrices

$$\left(\partial_{ij}^2F(0,0)-\left[-C\rho^{2/3}\delta_{ij}+c\delta_{ij}1_{i,j>d}\right]\right)$$

and

$$\left(C\left[\rho^{2/3}\delta_{ij}+\delta_{ij}1_{i,j>d}\right]-\partial_{ij}^2F(0,0)\right)$$

are positive definite, real, and symmetric. If (A_{ij}) is positive definite, real, and symmetric, then

$$\big|A_{ij}\big|^2 < A_{ii}A_{jj}$$

for $i\ne j$, since the two-by-two submatrix

$$\begin{pmatrix} A_{ii} & A_{ij} \\ A_{ji} & A_{jj}\end{pmatrix}$$

must also be positive definite and thus has a positive determinant. It follows from (35) that

$$\big|\partial_{ii}^2F(0,0)\big| \le C\rho^{2/3},$$

if $i\le d$, and

$$\big|\partial_{jj}^2F(0,0)\big| \le C$$

for any j. Therefore, if $i\le d$ and $j>d$, then

$$\big|\partial_{ij}^2F(0,0)\big|^2 \le \big|\partial_{ii}^2F(0,0)+C\rho^{2/3}\big|\cdot\big|\partial_{jj}^2F(0,0)-c\big| \le C\rho^{2/3}.$$

Thus,

$$\big|\partial_{ij}^2F(0,0)\big| \le C\rho^{1/3} \tag{36}$$

if $1\le i\le d$ and $d+1\le j\le n$. Without loss of generality, we can rotate the last $n-d$ coordinate axes in $\mathbb{R}^n$, so that the matrix

$$\left(\partial_{ij}^2F(0,0)\right)_{i,j=d+1,\dots,n}$$

is diagonal, say,

$$\left(\partial^2_{ij}F(0,0)\right)_{i,j=d+1,\dots,n} = \begin{pmatrix} \lambda_{d+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}.$$

For an $n \times n$ matrix $A = (a_{ij})$, let

$$\|A\|_\infty := \sup_{(i,j)\in[n]\times[n]} |a_{ij}|.$$

Then (35) and (36) show that

$$\left\| \left(\partial^2_{ij}F(0,0)\right)_{i,j=1,\dots,n} - \left(\begin{array}{c|ccc} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times 1} & \cdots & \mathbf{0}_{d\times 1} \\ \hline \mathbf{0}_{1\times d} & \lambda_{d+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times d} & 0 & \cdots & \lambda_n \end{array}\right) \right\|_\infty \leq C\rho^{1/3} \tag{37}$$

and

$$c \leq \lambda_j \leq C \tag{38}$$

for each $j = d+1, \dots, n$.

For λ_j satisfying (38), let $c^\#$ be a sufficiently small controlled constant. Let Ω be the set of all real symmetric $n \times n$ matrices A such that

$$\left\| A - \left(\begin{array}{c|ccc} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times 1} & \cdots & \mathbf{0}_{d\times 1} \\ \hline \mathbf{0}_{1\times d} & \lambda_{d+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times d} & 0 & \cdots & \lambda_n \end{array}\right) \right\|_\infty < c^\#. \tag{39}$$

We can pick controlled constants so that (37), (38) and (28), (30) imply that

$$\left(\partial^2_{ij}F(z)\right)_{i,j=1,\dots,n} \in \Omega \tag{40}$$

for $|z| < \overline{c}_4$. Here $\mathbf{0}_{d\times d}$, $\mathbf{0}_{1\times d}$, and $\mathbf{0}_{d\times 1}$ denote all-zero $d \times d, 1 \times d$, and $d \times 1$ matrices, respectively.

Definition 14. *If $A \in \Omega$, let $\Pi_{hi}(A) : \mathbb{R}^n \to \mathbb{R}^n$ be the orthogonal projection from $\mathbb{R}^n$ to the span of the eigenspaces of A that correspond to eigenvalues in $[\overline{c}_2, \overline{C}_3]$, and let $\Pi_{lo} : \mathbb{R}^n \to \mathbb{R}^n$ be the orthogonal projection from $\mathbb{R}^n$ onto the span of the eigenspaces of A that correspond to eigenvalues in $[-\overline{c}_1, \overline{c}_1]$.*

Then, $A \mapsto \Pi_{hi}(A)$ and $A \mapsto \Pi_{lo}(A)$ are smooth maps from the compact set Ω into the space of all real symmetric $n \times n$ matrices. For a matrix A, let $|A|$ denote its spectral norm, i.e.,

$$|A| := \sup_{\|u\|=1} \|Au\|.$$

Then, in particular,

$$\left|\Pi_{hi}(A) - \Pi_{hi}(A')\right| + \left|\Pi_{lo}(A) - \Pi_{lo}(A')\right| \leq C\left|A - A'\right| \tag{41}$$

for $A, A' \in \Omega$, and

$$\left|\partial^\alpha_A \Pi_{hi}(A)\right| + \left|\partial^\alpha_A \Pi_{lo}(A)\right| \leq C \tag{42}$$

for $A \in \Omega, |\alpha| \leq k$. Let

$$\Pi_{hi}(z) = \Pi_{hi}\left(\partial^2 F(z)\right) \tag{43}$$

and

$$\Pi_{lo}(z) = \Pi_{lo}\left(\partial^2 F(z)\right), \tag{44}$$

for $z < \overline{c}_4$, which make sense, thanks to the comment following (39). Also, we define projections $\Pi_d : \mathbb{R}^n \to \mathbb{R}^n$ and $\Pi_{n-d} : \mathbb{R}^n \to \mathbb{R}^n$ by setting

$$\Pi_d : (z_1, \dots, z_n) \mapsto (z_1, \dots, z_d, 0, \dots, 0) \tag{45}$$

and

$$\Pi_{n-d} : (z_1, \dots, z_n) \mapsto (0, \dots, 0, z_{d+1}, \dots, z_n). \tag{46}$$

From (37) and (41) we see that

$$\left|\Pi_{hi}(0) - \Pi_{n-d}\right| \leq C\rho^{1/3}. \tag{47}$$

Also, (28) and (42) together give

$$\left|\partial_z^\alpha \Pi_{hi}(z)\right| \leq C \tag{48}$$

for $|z| < \overline{c}_4, |\alpha| \leq k-2$. From (47), (48), and (30), we have

$$|\Pi_{hi}(z) - \Pi_{n-d}| \leq C\rho^{1/3} \tag{49}$$

for $|z| \leq \rho^{1/3}$. Note that $\Pi_{hi}(z)$ is the orthogonal projection from $\mathbb{R}^n$ onto the span of the eigenvectors of $\partial^2 F(z)$ with $(n-d)$ highest eigenvalues; this holds for $|z| < \overline{c}_4$. Now set

$$\zeta(z) = \Pi_{n-d}\Pi_{hi}\partial F(z) \tag{50}$$

for $|z| < \overline{c}_4$. Thus

$$\zeta(z) = (\zeta_{d+1}(z), \dots, \zeta_n(z)) \in \mathbb{R}^{n-d}, \tag{51}$$

where

$$\zeta_i(z) = \sum_{j=1}^{n} [\Pi_{hi}(z)]_{ij} \partial_{z_j} F(z) \tag{52}$$

for $i = d+1, \dots, n, |z| < \overline{c}_4$. Here, $[\Pi_{hi}(z)]_{ij}$ is the ij entry of the matrix $\Pi_{hi}(z)$. From (48) and (28) we see that

$$|\partial^\alpha \zeta(z)| \leq C \tag{53}$$

for $|z| < \overline{c}_4, |\alpha| \leq k-2$. Also, since Π_{n-d} and $\Pi_{hi}(z)$ are orthogonal projections from $\mathbb{R}^n$ to subspaces of $\mathbb{R}^n$, (34) and (50) yield

$$|\zeta(0)| \leq c\rho. \tag{54}$$

From (52), we have

$$\frac{\partial \zeta_i}{\partial z_\ell}(z) = \sum_{j=1}^{n} \frac{\partial}{\partial z_\ell}[\Pi_{hi}(z)]_{ij} \frac{\partial}{\partial z_j} F(z) + \sum_{j=1}^{n} [\Pi_{hi}(z)]_{ij} \frac{\partial^2 F(z)}{\partial z_\ell \partial z_j} \tag{55}$$

for $|z| < \overline{c}_4$ and $i = d+1, \dots, n$, $\ell = 1, \dots, n$. We take $z = 0$ in (55). From (34) and (48), we have

$$\left|\frac{\partial}{\partial z_\ell}[\Pi_{hi}(z)]_{ij}\right| \leq C$$

and

$$\left|\frac{\partial}{\partial z_j} F(z)\right| \leq C\rho$$

for $z = 0$. Also, from (47) and (37), we see that

$$\left|[\Pi_{hi}(z)]_{ij} - \delta_{ij}\right| \leq C\rho^{\frac{1}{3}}$$

for $z = 0, i = d+1, \dots, n, j = d+1, \dots, n$;

$$\left|[\Pi_{hi}(z)]_{ij}\right| \leq C\rho^{1/3}$$

for $z = 0$, $i = d+1, \dots, n$ and $j = 1, \dots, d$; and

$$\left|\frac{\partial^2 F}{\partial z_j \partial z_\ell}(z) - \delta_{j\ell}\lambda_\ell\right| \leq C\rho^{\frac{1}{3}}$$

for $z = 0$, $j = 1, \dots, n$, and $\ell = d+1, \dots, n$.

In view of the above remarks, (55) shows that

$$\left|\frac{\partial \zeta_i}{\partial z_\ell}(0) - \lambda_\ell \delta_{i\ell}\right| \leq C\rho^{1/3} \tag{56}$$

for $i, \ell = d+1, \dots, n$. Let $B_d(0, r), B_{n-d}(0, r)$, and $B_n(0, r)$ denote the open balls about 0 with radius r in $\mathbb{R}^d, \mathbb{R}^{n-d}$, and $\mathbb{R}^n$, respectively. Thanks to (30), (38), (53), (54), (56), and the implicit function theorem (see Section 3 of [39]), there exist controlled constants $\overline{c}_6 < \overline{c}_5 < \frac{1}{2}\overline{c}_4$ and a $\mathcal{C}^{k-2}$-map

$$\Psi : B_d(0, \overline{c}_6) \to B_{n-d}(0, \overline{c}_5), \tag{57}$$

with the following properties:

$$|\partial^\alpha \Psi| \leq C \tag{58}$$

on $B_d(0, \overline{c}_6)$, for $|\alpha| \leq k-2$;

$$|\Psi(0)| \leq C\rho. \tag{59}$$

Let $z = (x, y) \in B_d(0, \overline{c}_6) \times B_{n-d}(0, \overline{c}_5)$. Then

$$\zeta(z) = 0 \text{ if and only if } y = \Psi(x). \tag{60}$$

According to (47) and (48), the following holds for a small enough controlled constant $\overline{c}_7$. Let $z \in B_n(0, \overline{c}_7)$. Then $\Pi_{hi}(z)$ and $\Pi_{n-d}\Pi_{hi}(z)$ have the same null space. Therefore by (50), we have the following. Let $z \in B_n(0, \overline{c}_7)$. Then $\zeta(z) = 0$ if and only if $\Pi_{hi}(z)\partial F(z) = 0$. Consequently, after replacing $\overline{c}_5$ and $\overline{c}_6$ in (57), (58), (59), (60) by smaller controlled constants $\overline{c}_9 < \overline{c}_8 < \frac{1}{2}\overline{c}_7$, we obtain the following results:

$$\Psi : B_d(0, \overline{c}_9) \to B_{n-d}(0, \overline{c}_8) \tag{61}$$

is a $\mathcal{C}^{k-2}$-smooth map;

$$|\partial^\alpha \Psi| \leq C \tag{62}$$

on $B_d(0, \overline{c}_9)$ for $|\alpha| \leq k-2$;

$$|\Psi(0)| \leq C\rho. \tag{63}$$

Let

$$z = (x, y) \in B_d(0, \overline{c}_9) \times B_{n-d}(0, \overline{c}_8).$$

Then,

$$\Pi_{hi}(z)\partial F(z) = 0 \tag{64}$$

if and only if $y = \Psi(x)$. Thus we have understood the set $\{\Pi_{hi}(z)\partial F(z) = 0\}$ in the neighborhood of 0 in $\mathbb{R}^n$. Next, we study the bundle over $\{\Pi_{hi}(z)\partial F(z) = 0\}$ whose fiber at z is the image of $\Pi_{hi}(z)$. For $x \in B_d(0, \bar{c}_9)$ and $v = (0, \dots, 0, v_{d+1}, \dots, v_n) \in \{0\} \oplus \mathbb{R}^{n-d}$, we define

$$E(x, v) = (x, \Psi(x)) + [\Pi_{hi}(x, \Psi(x))]v \in \mathbb{R}^n. \tag{65}$$

From (48) and (58), we have

$$\left|\partial^\alpha_{x,v} E(x, v)\right| \le C \tag{66}$$

for $x \in B_d(0, \bar{c}_9), v \in B_{n-d}(0, \bar{c}_8), |\alpha| \le k - 2$. Here and below, we abuse notation by failing to distinguish between $\mathbb{R}^d$ and $\mathbb{R}^d \oplus \{0\} \in \mathbb{R}^n$. Let $E(x, v) = (E_1(x, v), \dots, E_n(x, v)) \in \mathbb{R}^n$. For $i = 1, \dots, d$, (65) gives

$$E_i(x, v) = x_i + \sum_{i=1}^{n} [\Pi_{hi}(x, \Psi(x))]_{ij} v_j. \tag{67}$$

For $i = d + 1, \dots, n$, (65) gives

$$E_i(x, v) = \Psi_i(x) + \sum_{i=1}^{n} [\Pi_{hi}(x, \Psi(x))]_{ij} v_j, \tag{68}$$

where we write $\Psi(x) = (\Psi_{d+1}(x), \dots, \Psi_n(x)) \in \mathbb{R}^{n-d}$. We study the first partials of $E_i(x, v)$ at $(x, v) = (0, 0)$. From (67), we find that

$$\frac{\partial E_i}{\partial x_j}(x, v) = \delta_{ij} \tag{69}$$

at $(x, v) = (0, 0)$, for $i, j = 1, \dots, d$. Also, (63) shows that $|(0, \Psi(0))| \le c\rho$; hence, (49) gives

$$\left|\Pi_{hi}(0, \Psi(0)) - \Pi_{n-d}\right| \le C\rho^{1/3}, \tag{70}$$

for $i \in \{1, \dots, d\}$ and $j \in \{1, \dots, n\}$. Therefore, another application of (67) yields

$$\left|\frac{\partial E_i}{\partial v_j}(x, v)\right| \le C\rho^{1/3} \tag{71}$$

for $i \in [d], j \in \{d + 1, \dots, n\}$, and $(x, v) = (0, 0)$. Similarly, from (70) we obtain

$$\left|[\Pi_{hi}(0, \Psi(0))]_{ij} - \delta_{ij}\right| \le C\rho^{1/3}$$

for $i = d + 1, \dots, n$ and $j = d + 1, \dots, n$. Therefore, from (68), we have

$$\left|\frac{\partial E_i}{\partial v_j}(x, v) - \delta_{ij}\right| \le C\rho^{1/3} \tag{72}$$

for $i, j = d + 1, \dots, n$, $(x, v) = (0, 0)$. In view of (66), (69), (71), (72), the Jacobian matrix of the map $(x_1, \dots, x_d, v_{d+1}, \dots, v_n) \mapsto E(x, v)$ at the origin is given by

$$\left(\begin{array}{c|c} I_d & O(\rho^{1/3}) \\ \hline O(1) & I_{n-d} + O(\rho^{1/3}) \end{array}\right), \tag{73}$$

where I_d and I_{n-d} denote (respectively) the $d \times d$ and $(n - d) \times (n - d)$ identity matrices, $O(\rho^{1/3})$ denotes a matrix whose entries have absolute values at most $C\rho^{1/3}$; and $O(1)$ denotes a matrix whose entries have absolute values at most C.

A matrix of the form (73) is invertible, and its inverse matrix has the norm at most C. (Here, we use (30).) Note also that $|E(0, 0)| = |(0, \Psi(0))| \le C\rho$.

Consequently, the inverse function theorem (see Section 3 of [39]) and (66) imply the following.

There exist controlled constants $\overline{c}_{10}$ and $\overline{c}_{11}$ with the following properties:

(74) The map $E(x,v)$ is one-to-one when restricted to $B_d(0,\overline{c}_{10}) \times B_{n-d}(0,\overline{c}_{10})$.

(75)

$$\text{The image of } E(x,v) : B_d(0,\overline{c}_{10}) \times B_{n-d}(0,\frac{\overline{c}_{10}}{2}) \to \mathbb{R}^n \text{contains a ball } B_n(0,\overline{c}_{11}).$$

(76) In view of (74), (75), the map

$$E^{-1} : B_n(0,\overline{c}_{11}) \to B_d(0,\overline{c}_{10}) \times B_{n-d}(0,\frac{\overline{c}_{10}}{2})$$

is well-defined.

(77) The derivatives of E^{-1} of order $\leq k-2$ have absolute value at most C.

Moreover, we may pick $\overline{c}_{10}$ in (74) small enough that the following holds.

Observation 1.

(78) *Let $x \in B_d(0,\overline{c}_{10})$, and let $v \in \Pi_{hi}(x,\Psi(x))\mathbb{R}^n$.*

(79) *Then, we can express v in the form $v = \Pi_{hi}(x,\psi(x))v^{\#}$ where $v^{\#} \in \{0\} \oplus \mathbb{R}^{n-d}$ and $|v^{\#}| \leq 2|v|$.*

Indeed, if $x \in B_d(0,\overline{c}_{10})$ for small enough $\overline{c}_{10}$, then by (30), (62), (63), we have $|(x,\Psi(x))| < c$ for small c; consequently, (79) follows from (47), (48). Thus (74), (75), (76), (77), and (79) hold for suitable controlled constants $\overline{c}_{10}, \overline{c}_{11}$. From (75), (76), (79), we learn the following.

Observation 2. *Let $x, \tilde{x} \in B_d(0,\overline{c}_{10})$, and let $v, \tilde{v} \in B_{n-d}(0,\frac{1}{2}\overline{c}_{10})$. Assume that $v \in \Pi_{hi}(x,\Psi(x))\mathbb{R}^n$ and $\tilde{v} \in \Pi_{hi}(\tilde{x},\Psi(\tilde{x}))\mathbb{R}^n$. If $(x,\Psi(x)) + v = (\tilde{x},\Psi(\tilde{x})) + \tilde{v}$, then $x = \tilde{x}$ and $v = \tilde{v}$.*

Observation 3. *Any $z \in B_n(0,\overline{c}_{11})$ may be expressed uniquely in the form $(x,\Psi(x)) + v$ with $x \in B_d(0,\overline{c}_{10}), v \in \Pi_{hi}(x,\Psi(x))\mathbb{R}^n \cap B_n(0,\frac{\overline{c}_{10}}{2})$. Moreover, x and v here are $\mathcal{C}^{k-2}$-smooth functions of $z \in B_n(0,\overline{c}_{11})$, with derivatives up to order $k-2$ bounded by C in absolute value.*

This concludes the proof of the lemma. □

10. Constructing cylinder packets

Let $\mathbb{R}^d$ and $\mathbb{R}^{n-d}$, respectively, denote the spans of the first d vectors and the last $n-d$ vectors of the canonical basis of $\mathbb{R}^n$. Let B_d and B_{n-d}, respectively, denote the unit Euclidean balls in $\mathbb{R}^d$ and $\mathbb{R}^{n-d}$. Let

$$\bar{\tau} := \overline{c}_{12}\tau. \tag{80}$$

Let Π_d be the map given by the orthogonal projection from $\mathbb{R}^n$ onto $\mathbb{R}^d$. Let $\mathtt{cyl} := \bar{\tau}(B_d \times B_{n-d})$, and $\mathtt{cyl}^2 = 2\bar{\tau}(B_d \times B_{n-d})$. Suppose that for any $x \in 2\bar{\tau}B_d$ and $y \in 2\bar{\tau}B_{n-d}$, $\phi_{\mathtt{cyl}^2}$ is given by

$$\phi_{\mathtt{cyl}^2}(x,y) = |y|^2,$$

and for any $z \notin \mathtt{cyl}^2$,

$$\phi_{\mathtt{cyl}^2}(z) = 0.$$

Suppose for each $i \in [\bar{N}] := \{1, \dots, \bar{N}\}$, $x_i \in B_n(0,1)$ and o_i is a proper rigid body motion, i.e., the composition of a proper rotation and translation of $\mathbb{R}^n$ and that $o_i(0) = x_i$.

For each $i \in [\bar{N}]$, let $\mathtt{cyl}_i := o_i(\mathtt{cyl})$, and $\mathtt{cyl}_i^2 := o_i(\mathtt{cyl}^2)$. Note that x_i is the center of $\mathtt{cyl}_i$.

We say that a set of cylinders $C_p := \{\mathtt{cyl}_1^2, \dots, \mathtt{cyl}_{\bar{N}}^2\}$ (where each $\mathtt{cyl}_i^2$ is isometric to $\mathtt{cyl}^2$) is a cylinder packet in $\mathcal{C}^{\bar{\tau}}(d, V, \tau)$ (or simply a cylinder packet) if the following conditions hold true:

(1) The number of cylinders $\bar{N}$ is less than or equal to $\frac{V}{\tau^d}$.

(2) Let $S_i := \{\mathtt{cyl}_{i_1}^2, \dots, \mathtt{cyl}_{i_{|S_i|}}^2\}$ be the set of cylinders that intersect $\mathtt{cyl}_i^2$. Translate the origin to the center of $\mathtt{cyl}_i^2$ (i.e., x_i) and perform a proper Euclidean transformation that puts the d dimensional central cross section of $\mathtt{cyl}_i^2$ in $\mathbb{R}^d$.

There exist proper rotations $U_{i_1}, \dots, U_{i_{|S_i|}}$, respectively, of the cylinders $\mathtt{cyl}_{i_1}^2, \dots, \mathtt{cyl}_{i_{|S_i|}}^2$ in S_i such that U_{i_j} fixes the center x_{i_j} of $\mathtt{cyl}_{i_j}^2$ and translations $Tr_{i_1}, \dots, Tr_{i_{|S_i|}}$ such that

(a) For each $j \in [|S_i|]$, $Tr_{i_j} U_{i_j} \mathtt{cyl}_{i_j}^2$ is a translation of $\mathtt{cyl}_i^2$ by a vector contained in $\mathbb{R}^d$ whose Euclidean norm is at least $\frac{\bar{\tau}}{3}$.

(b) The set $\{Tr_{i_j} U_{i_j} x_{i_j} | j \in [|S_i|]\} \cap \mathtt{cyl}_i^2$ is a $\frac{\bar{\tau}}{2}$ net of $\mathbb{R}^d \cap \mathtt{cyl}_i^2$.

(c) $\left| \left(Id - U_{i_j}\right) v \right| < 2\left(\frac{\bar{\tau}}{\tau}\right) |v - x_{i_j}|$, for each j in $\{1, \dots, |S_i|\}$, and any vector v.

(d) $|Tr_{i_j}(0)| < \frac{\bar{\tau}^2}{\tau}$ for each j in $\{1, \dots, |S_i|\}$.

Observation 4. *Any point in $B_d(0, (5/2)\bar{\tau})$ is within $\bar{\tau}/2$ of a point in $B_d(0, 2\bar{\tau})$, which in turn is within $\bar{\tau}/2$ of some $Tr_j x_{i_j}$. It therefore follows that*

$$\bigcup_j (Tr_{i_j} U_{i_j} (o_{i_j}(\mathbb{R}^d) \cap \mathtt{cyl}_{i_j})) \supseteq B_d(0, (5/2)\bar{\tau}).$$

We call $\{o_1, \dots, o_{\bar{N}}\}$ a packet of rigid body motions or simply a packet if $\{o_1(\mathtt{cyl}), \dots, o_{\bar{N}}(\mathtt{cyl})\}$ is a cylinder packet.

11. Constructing a disc bundle possessing the desired characteristics

11.1. Approximate squared distance functions.

Suppose that $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ is a submanifold of $\mathbb{R}^n$. For $\tilde{\tau} > 0$, let

$$\mathcal{M}_{\tilde{\tau}} := \{z \in \mathbb{R}^n | \inf_{\bar{z} \in \mathcal{M}} |z - \bar{z}| < \tilde{\tau}\}.$$

Note that $\mathcal{M}_{\tilde{\tau}}$ is a tubular neighborhood of the manifold $\mathcal{M}$. Let $\tilde{d}$ be a suitable large constant depending only on d, and which is a monotonically increasing function of d. Let

$$\bar{d} := \min(n, \tilde{d}). \tag{81}$$

We use a basis for $\mathbb{R}^n$ that is such that $\mathbb{R}^{\bar{d}}$ is the span of the first $\bar{d}$ basis vectors, and $\mathbb{R}^d$ is the span of the first d basis vectors. We denote by $\Pi_{\bar{d}}$, the corresponding projection of $\mathbb{R}^n$ onto $\mathbb{R}^{\bar{d}}$. Recall that

$$\bar{\tau} := \bar{c}_{12}\tau.$$

Definition 15. *Let* $asdf^{\bar\tau}_{\mathcal{M}}$ *denote the set of all functions* $\bar F : \mathcal{M}_{\bar\tau} \to \mathbb{R}$ *such that the following is true. For every* $z \in \mathcal{M}$*, there exists an isometry* Θ_z *of* $\mathbb{R}^n$ *that fixes the origin and maps* $\mathbb{R}^d$ *to a subspace parallel to the tangent plane at* z *such that* $\hat F_z : B_n(0,1) \to \mathbb{R}$ *given by*

$$\hat F_z(w) = \frac{\bar F(z + \bar\tau \Theta_z(w))}{\bar\tau^2} \tag{82}$$

satisfies the following:

<u>ASDF-1</u> $\hat F_z$ *satisfies the hypotheses of Theorem* 13 *for a sufficiently small controlled constant* ρ *which will be specified in Equation (*85*) in the proof of Lemma* 14*. The value of* k *equals* $r+2$*, where* $r = 2$ *is the degree of smoothness of the manifolds in Definition* 3*.*

<u>ASDF-2</u> *There is a function* $F_z : \mathbb{R}^{\bar d} \to \mathbb{R}$ *such that for any* $w \in B_n(0,1)$*,*

$$\hat F_z(w) = F_z\left(\Pi_{\bar d}(w)\right) + |w - \Pi_{\bar d}(w)|^2, \tag{83}$$

where $\mathbb{R}^d \subseteq \mathbb{R}^{\bar d} \subseteq \mathbb{R}^n$*, and* $\bar d$ *is a function of* d *alone.*

Let

$$\Gamma_z = \{w \,|\, \Pi^z_{hi}(w)\partial \hat F_z(w) = 0\}, \tag{84}$$

where Π^z_{hi} is as in Theorem 13 applied to the function $\hat F_z$.

Lemma 14. *Suppose that* $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ *is a submanifold of* $\mathbb{R}^n$*. Let* $\bar F$ *be in* $asdf^{\bar\tau}_{\mathcal{M}}$*, and let* Γ_z *and* Θ_z *be as in Definition* 15*.*

(1) *The graph* Γ_z *is contained in* $\mathbb{R}^{\bar d}$*.*
(2) *Let* c_4 *and* c_5 *be the constants appearing in (*31*) in Theorem* 13*, once we fix* C_0 *in (*28*) to be* 10*, and the constants* c_1 *and* C_1 *in (*29*) to* 1/10 *and* 10*, respectively. The set*

$$\mathcal{M}_{put} := \left\{z \in \mathcal{M}_{\min(c_4,c_5)\bar\tau} \middle| \Pi_{hi}(z)\partial \bar F(z) = 0\right\}$$

is a submanifold of $\mathbb{R}^n$ *and has a reach greater than* $c\tau$*, where* c *is a constant depending only on* d*.*

Here $\Pi_{hi}(z)$ *is the orthogonal projection onto the eigenspace corresponding to eigenvalues in the interval* $[\overline{c}_2, \overline{C}_2]$ *that is specified in Definition* 14*.*

Proof. To see the first part of the lemma, note that because of (83), for any $w \in B_n(0,1)$, the span of the eigenvectors corresponding to the eigenvalues of the Hessian of $F = \hat F_z$ that lie in $(\overline{c}_2, \overline{C}_3)$ contains the orthogonal complement of $\mathbb{R}^{\bar d}$ in $\mathbb{R}^n$ (henceforth referred to as $\mathbb{R}^{n-\bar d}$). Further, if $w \not\in \mathbb{R}^{\bar d}$, there is a vector in $\mathbb{R}^{n-\bar d}$ that is not orthogonal to the gradient $\partial \hat F_z(w)$. Therefore

$$\Gamma_z \subseteq \mathbb{R}^{\bar d}.$$

We proceed to the second part of the Lemma. We choose $\overline{c}_{12}$ to be a small enough monotonically decreasing function of $\bar d$ (by (81) and the assumed monotonicity of $\tilde d$, $\overline{c}_{12}$ is consequently a monotonically decreasing function of d) such that for every point $z \in \mathcal{M}$, F_z given by (83) satisfies the hypotheses of Theorem 13 with $\rho < \frac{\tilde c \bar\tau}{C}$ where C is the constant in Equation (32) and where $\tilde c$ is a sufficiently small controlled constant. Suppose, for the purpose of reaching a contradiction, that there is a point $\hat z$ in $\mathcal{M}_{put}$ such that $\mathbf{d}(\hat z, \mathcal{M})$ is greater than $\frac{min(c_4,c_5)\bar\tau}{2}$, where c_4 and c_5 are the constants in (31). Let z be the unique point on $\mathcal{M}$ nearest to $\hat z$.

We apply Theorem 13 to F_z. By Equation (32) in Theorem 13, there is a point $\tilde{z} \in \mathcal{M}_{put}$ such that

$$|z - \tilde{z}| < C\rho < c_{lem}\bar{\tau}. \tag{85}$$

The constant c_{lem} is controlled by $\tilde{c}$ and can be made as small as needed provided it is ultimately controlled by d alone. We have an upper bound of C on the first-order derivatives of Ψ in Equation (32), which is a function whose graph corresponds via Θ_z to $\mathcal{M}_{put}$ in a $\frac{\bar{\tau}}{2}$-neighborhood of z. Any unit vector $v \in Tan^0(z)$ is nearly orthogonal to $\tilde{z} - \hat{z}$ in that

$$\frac{\big|\langle \tilde{z} - \hat{z}, v\rangle\big|}{\big|\tilde{z} - \hat{z}\big|} < \frac{2c_{lem}}{\min(c_4, c_5)}. \tag{86}$$

We can choose c_{lem} small enough that (86) contradicts the mean value theorem applied to Ψ because of the upper bound of C on $|\partial\Psi|$ from Equation (32).

This shows that for every $\hat{z} \in \mathcal{M}_{put}$ its distance to $\mathcal{M}$ satisfies

$$\mathbf{d}(\hat{z}, \mathcal{M}) \le \frac{\min(c_4, c_5)\bar{\tau}}{2}. \tag{87}$$

Recall that

$$\mathcal{M}_{put} := \left\{z \in \mathcal{M}_{\min(c_4,c_5)\bar{\tau}} \middle| \Pi_{hi}(z)\partial\bar{F}(z) = 0\right\}.$$

Therefore, for every point $\hat{z}$ in $\mathcal{M}_{put}$, there is a point $z \in \mathcal{M}$ such that

$$B_n\left(\hat{z}, \frac{\min(c_4, c_5)\bar{\tau}}{2}\right) \subseteq \Theta_z\left(B_d(0, c_4\bar{\tau}) \times B_{n-d}(0, c_5\bar{\tau})\right). \tag{88}$$

We have now shown that $\mathcal{M}_{put}$ lies not only in $\mathcal{M}_{\min(c_4,c_5)\bar{\tau}}$ but also in $\mathcal{M}_{\frac{\min(c_4,c_5)\bar{\tau}}{2}}$. Recall that $\bar{\tau} = \bar{c}_{12}\tau$ by (80). This fact, in conjunction with (32) and Proposition 1 implies that $\mathcal{M}_{put}$ is a manifold with reach greater than $c\tau$. □

Let

$$\bar{D}_{\bar{F}}^{\tilde{\tau}} \to \mathcal{M}_{put} \tag{89}$$

be the bundle over $\mathcal{M}_{put}$ wherein the fiber at a point $\hat{z} \in \mathcal{M}_{put}$ consists of all points z such that

(1) $|\hat{z} - z| \le \tilde{\tau}$, and
(2) $z - \hat{z}$ lies in the span of the top $n - d$ eigenvectors of the Hessian of $\bar{F}$ evaluated at $\hat{z}$.

Observation 5. *By Theorem* 13*,* $\mathcal{M}$ *is a* $\mathcal{C}^r$*-smooth section of* $\bar{D}_{\bar{F}}^{\bar{c}_{11}\bar{\tau}}$ *and the controlled constants* $c_1, \dots, c_7$ *and* C *and depend only on* $c_1, C_1, C_0, k,$ *and* n *(these constants are identical to those in Theorem* 13*). By* (83)*, we conclude that the dependence on* n *can be replaced by a dependence on* $\bar{d}$*.*

11.2. **The disc bundles constructed from approximate-squared-distance functions are good.** In this subsection, we prove that any approximate-squared-distance function defined on a cylinder packet corresponds to a putative manifold and a disc bundle having good properties. By Lemma 17 of the next section and Observation 5, it will follow that every manifold in $\mathcal{G}(d, V, \tau)$ is achieved as the graph of a section of a disc bundle constructed from some element of $\mathcal{C}^{\bar{\tau}}(d, V, \tau)$.

Suppose that $\mathcal{C} \in \mathcal{C}^{\bar{\tau}}(d, V, \tau)$ is a cylinder packet corresponding to $\bar{o} = \{o_1, \dots, o_{\bar{N}}\}$. Let $\mathcal{A}$ be the union over i of the discs $\mathcal{A}_i := o_i(\mathbb{R}^d \cap \mathtt{cyl}^2)$. For $\tilde{\tau} > 0$, let

$$\mathcal{A}_{\tilde{\tau}} := \{z \in \mathbb{R}^n | \inf_{\bar{z} \in \mathcal{A}} |z - \bar{z}| < \tilde{\tau}\}.$$

Note that $\mathcal{A}_{\tilde{\tau}}$ is a neighborhood of the set $\mathcal{M}$. As before, let $\tilde{d}$ be a suitable large constant depending only on d, and which is a monotonically increasing function of d. Let

$$\bar{d} := \min(n, \tilde{d}). \tag{90}$$

As was the case earlier, we use a basis for $\mathbb{R}^n$ that is such that $\mathbb{R}^d$ is the span of the first d basis vectors, and $\mathbb{R}^{\bar{d}}$ is the span of the first $\bar{d}$ basis vectors. We denote by $\Pi_{\bar{d}}$, the corresponding projection of $\mathbb{R}^n$ onto $\mathbb{R}^{\bar{d}}$.

Definition 16. *Let $asdf_{\mathcal{A}}^{\bar{\tau}}$ denote the set of all functions $\bar{F} : \mathcal{A}_{\bar{\tau}} \to \mathbb{R}$ such that the following is true. For every $i \in \bar{N}$ and $z \in o_i(\mathbb{R}^d) \cap \mathtt{cyl}_i^2$, there exists an isometry Θ_z of $\mathbb{R}^n$ that fixes the origin and maps $\mathbb{R}^d$ to a subspace parallel to $o_i(\mathbb{R}^d)$ containing z such that $\hat{F}_z : B_n(0,1) \to \mathbb{R}$ given by*

$$\hat{F}_z(w) = \frac{\bar{F}(z + \bar{\tau}\Theta_z(w))}{\bar{\tau}^2} \tag{91}$$

satisfies the following.

<u>$ASDF_A - 1$</u> *$\hat{F}_z$ satisfies the hypotheses of Theorem 13 for a sufficiently small controlled constant ρ which will be specified in Equation (94) in the proof of Lemma 15. The value of k equals $r + 2$, where $r = 2$ is the degree of smoothness of the manifolds in Definition 3.*

<u>$ASDF_A - 2$</u> *There is a function $F_z : \mathbb{R}^{\bar{d}} \to \mathbb{R}$ such that for any $w \in B_n(0,1)$,*

$$\hat{F}_z(w) = F_z\left(\Pi_{\bar{d}}(w)\right) + |w - \Pi_{\bar{d}}(w)|^2, \tag{92}$$

where $\mathbb{R}^d \subseteq \mathbb{R}^{\bar{d}} \subseteq \mathbb{R}^n$, and $\bar{d}$ is a function of d alone.

Let

$$\Gamma_z = \{w \,|\, \Pi_{hi}^z(w)\partial\hat{F}_z(w) = 0\}, \tag{93}$$

where Π_{hi}^z is as in Theorem 13 applied to the function $\hat{F}_z$.

Lemma 15. *Let $\bar{F}$ be in $asdf_{\mathcal{A}}^{\bar{\tau}}$, and let Γ_z and Θ_z be as in Definition 16.*

1. *The graph Γ_z is contained in $\mathbb{R}^{\bar{d}}$.*
2. *Let c_4 and c_5 be the constants appearing in (31) in Theorem 13, once we fix C_0 in (28) to be 10, and the constants c_1 and C_1 in (29) to 1/10 and 10, respectively. The set*

$$\mathcal{M}_{put} := \left\{z \in \mathcal{A}_{\min(c_4,c_5)\bar{\tau}} \middle| \Pi_{hi}(z)\partial\bar{F}(z) = 0\right\}$$

is a submanifold of $\mathbb{R}^n$ and has a reach greater than $c\tau$, where c is a constant depending only on d.

Here $\Pi_{hi}(z)$ is the orthogonal projection onto the span of eigenvectors corresponding to eigenvalues in the interval $[\overline{c}_2, \overline{C}_2]$ that is specified in Definition 14.

Proof. The proof of this lemma closely follows the proof of Lemma 14.

To see the first part of the lemma, note that because of (92), for any $w \in B_n(0,1)$, the span of the eigenvectors corresponding to the eigenvalues of the Hessian of $F = \hat{F}_z$ that lie in $(\overline{c}_2, \overline{C}_3)$ contains the orthogonal complement of $\mathbb{R}^{\bar{d}}$ in $\mathbb{R}^n$ (henceforth referred to as $\mathbb{R}^{n-\bar{d}}$). Further, if $w \not\in \mathbb{R}^{\bar{d}}$, there is a vector in $\mathbb{R}^{n-\bar{d}}$ that is not orthogonal to the gradient $\partial\hat{F}_z(w)$. Therefore

$$\Gamma_z \subseteq \mathbb{R}^{\bar{d}}.$$

We proceed to the second part of the Lemma. We choose $\overline{c}_{12}$ to be a small enough monotonically decreasing function of $\bar{d}$ (by (81) and the assumed monotonicity of $\tilde{d}$, $\overline{c}_{12}$ is consequently a monotonically decreasing function of d) such that for every point $z \in \mathcal{A}$, F_z given by (92) satisfies the hypotheses of Theorem 13 with $\rho < \frac{\tilde{c}\bar{\tau}}{C}$ where C is the constant in Equation (32) and where $\tilde{c}$ is a sufficiently small controlled constant. Suppose, for the purpose of reaching a contradiction, that there is a point $\hat{z}$ in $\mathcal{M}_{put}$ such that $\mathbf{d}(\hat{z}, \mathcal{A})$ is greater than $\frac{min(c_4,c_5)\bar{\tau}}{2}$, where c_4 and c_5 are the constants in (31). Since $\hat{z}$ belongs to $\mathcal{A}_{\min(c_4,c_5)\bar{\tau}} \subseteq \bigcup_i \texttt{cyl}_i$, by Observation 4 and (b) of Section 10 it is possible to choose a point z on $o_i(\mathbb{R}^d)\cap\texttt{cyl}_i$ for some i such that $2\min(c_4,c_5) > \mathbf{d}(z,\hat{z}) > \frac{\min(c_4,c_5)\bar{\tau}}{2}$ and $z-\hat{z}$ is orthogonal to every vector in $(z - o_i(\mathbb{R}^d))$. We apply Theorem 13 to F_z. By Equation (32) in Theorem 13, there is a point $\tilde{z} \in \mathcal{M}_{put}$ such that

$$|z - \tilde{z}| < C\rho < c_{lem}\bar{\tau}. \tag{94}$$

The constant c_{lem} is controlled by $\tilde{c}$ and can be made as small as needed provided it is ultimately controlled by d alone. We have an upper bound of C on the first-order derivatives of Ψ in Equation (32), which is a function whose graph corresponds via Θ_z to $\mathcal{M}_{put}$ in a $\frac{\bar{\tau}}{2}$-neighborhood of z. Any unit vector $v \in Tan^0(z)$ is nearly orthogonal to $\tilde{z} - \hat{z}$ in that

$$\big|\langle \tilde{z} - \hat{z}, v\rangle\big| < \frac{2c_{lem}\big|\tilde{z} - \hat{z}\big|}{\min(c_4,c_5)}. \tag{95}$$

We can choose c_{lem} small enough that (95) contradicts the mean value theorem applied to Ψ because of the upper bound of C on $|\partial\Psi|$ from Equation (32).

This shows that for every $\hat{z} \in \mathcal{M}_{put}$ its distance to $\mathcal{A}$ satisfies

$$\mathbf{d}(\hat{z}, \mathcal{A}) \leq \frac{\min(c_4,c_5)\bar{\tau}}{2}. \tag{96}$$

Recall that

$$\mathcal{M}_{put} := \left\{z \in \mathcal{A}_{\min(c_4,c_5)\bar{\tau}} \middle| \Pi_{hi}(z)\partial\bar{F}(z) = 0\right\}.$$

Therefore, for every point $\hat{z}$ in $\mathcal{M}_{put}$, there is a point $z \in \mathcal{A}$ such that

$$B_n\left(\hat{z}, \frac{\min(c_4,c_5)\bar{\tau}}{2}\right) \subseteq \Theta_z\left(B_d(0, c_4\bar{\tau}) \times B_{n-d}(0, c_5\bar{\tau})\right). \tag{97}$$

We have now shown that $\mathcal{M}_{put}$ lies not only in $\mathcal{A}_{\min(c_4,c_5)\bar{\tau}}$ but also in $\mathcal{A}_{\frac{\min(c_4,c_5)\bar{\tau}}{2}}$. Recall that $\bar{\tau} = \bar{c}_{12}\tau$ by (80). This fact, in conjunction with (32) and Proposition 1, implies that $\mathcal{M}_{put}$ is a manifold with reach greater than $c\tau$. $\square$

Observation 6. *By Theorem* 13*, the graph of any function* $f : o_i(\mathbb{R}^d) \cap \texttt{cyl}_i \to o_i(\mathbb{R}^{n-d}) \cap \texttt{cyl}_i$ *such that* $\hat{f} : x \to (1/\bar{\tau})f(x/\bar{\tau})$ *has* $\mathcal{C}^2$ *norm less than a sufficiently* *small controlled constant (see Definition 22) corresponds to a $\mathcal{C}^2$-smooth local section of the disc bundle $\bar{D}_{\bar{F}}^{\bar{c}_{11}\bar{\tau}}$ (see (89)) and the controlled constants $c_1, \dots, c_7$ and C and depend only on $c_1, C_1, C_0, k,$ and n (these constants are identical to those in Theorem 13). By (83), we conclude that the dependence on n can be replaced by a dependence on $\bar{d}$.*

12. Constructing an exhaustive family of disc bundles

We wish to construct a family of functions $\bar{\mathcal{F}}$ defined on open subsets of $B_n(0,2)$ such that for every $\mathcal{M} \in \mathcal{G}(d, V, \tau)$ such that $\mathcal{M} \subseteq B_n(0,1)$, there is some $\hat{F} \in \bar{\mathcal{F}}$ such that the domain of $\hat{F}$ contains $\mathcal{M}_{\bar{\tau}}$ and the restriction of $\hat{F}$ to $\mathcal{M}_{\bar{\tau}}$ is contained in $asdf_{\mathcal{M}}^{\bar{\tau}}$.

We now show how to construct a set $\bar{D}$ of disc bundles rich enough that any manifold $\mathcal{M} \in \mathcal{G}(d, \tau, V)$ corresponds to a section of at least one disc bundle in $\bar{D}$. The constituent disc bundles in $\bar{D}$ will be obtained from cylinder packets.

Define

$$\theta : \mathbb{R}^d \to [0,1] \tag{98}$$

to be a bump function that has the following properties for any fixed k for a controlled constant C:

(1) For all α such that $0 < |\alpha| \le k$, for all $x \in \{0\} \cup (-\infty, -1] \cup [1, \infty)$

$$\partial^\alpha \theta(x) = 0,$$

and for all $x \in (-\infty, -1] \cup [1, \infty)$

$$\theta(x) = 0.$$

(2) for all x,

$$\left|\partial^\alpha \theta(x)\right| < C,$$

and for $|x| < \frac{1}{4}$,

$$\theta(x) = 1.$$

Definition 17. *Given a packet $\bar{o} := \{o_1, \dots, o_{\bar{N}}\}$, define $F^{\bar{o}} : \bigcup_i \boldsymbol{cyl}_i \to \mathbb{R}$ by*

$$F^{\bar{o}}(z) = \frac{\sum\limits_{\boldsymbol{cyl}_i^2 \ni z} \phi_{\boldsymbol{cyl}^2}(o_i^{-1}(z))\theta\left(\Pi_d(o_i^{-1}(z))/(2\bar{\tau})\right)}{\sum\limits_{\boldsymbol{cyl}_i^2 \ni z} \theta\left(\Pi_d(o_i^{-1}(z))/(2\bar{\tau})\right)}. \tag{99}$$

Definition 18. *Let A_1 and A_2 be two d dimensional affine subspaces of $\mathbb{R}^n$ for some $n \ge 1$ that, respectively, contain points x_1 and x_2. We define $\sphericalangle(A_1, A_2)$, the "angle between A_1 and A_2," by*

$$\sphericalangle(A_1, A_2) := \sup_{x_1 + v_1 \in A_1 \setminus x_1} \left(\inf_{x_2 + v_2 \in A_2 \setminus x_2} \arccos\left(\frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|} \right) \right).$$

Lemma 16. *Let* $\{cyl_1, \dots, cyl_{\bar N}\}$ *be a cylinder packet.*
Then,

$$F^{\bar o} \in asdf^{\bar\tau}_{\mathcal{A}}.$$

Proof. Recall that $asdf^{\bar\tau}_{\mathcal{A}}$ denotes the set of all $\bar F : \mathcal{A}_{\bar\tau} \to \mathbb{R}$ (where $\bar\tau = \bar c_{12}\tau$ and $\mathcal{M}_{\bar\tau}$ is a $\bar\tau$-neighborhood of $\mathcal{M}$) for which the following is true:

- For every $z \in \mathcal{A}_i$, there exists an isometry Θ of $\mathcal{H}$ that fixes the origin and maps $\mathbb{R}^d$ to a subspace parallel to $\mathcal{A}_i$ satisfying the conditions below.
  Let $\hat F_z : B_n(0,1) \to \mathbb{R}$ be given by

  $$\hat F_z(w) = \frac{\bar F(z + \bar\tau\Theta(w))}{\bar\tau^2}.$$

  Then, $\hat F_z$
  (1) satisfies the hypotheses of Theorem 13 with $k = r + 2 = 4$.
  (2) For any $w \in B_n$,

$$\hat F_z(w) = F_z\left(\Pi_{\bar d}(w)\right) + |w - \Pi_{\bar d}(w)|^2, \tag{100}$$

  where $\mathbb{R}^n \supseteq \mathbb{R}^{\bar d} \supseteq \mathbb{R}^d$, and $\Pi_{\bar d}$ is the projection of $\mathbb{R}^n$ onto $\mathbb{R}^{\bar d}$.

For any fixed $z \in \mathcal{A}_i$, it suffices to check that there exists an isometry Θ of $\mathcal{H}$ which satisfies as follows:

(A) The hypotheses of Theorem 13 are satisfied by

$$\hat F^{\bar o}_z(w) := \frac{F^{\bar o}(z + \bar\tau\Theta(w))}{\bar\tau^2}, \tag{101}$$

and

(B)

$$\hat F^{\bar o}_z(w) = \hat F^{\bar o}_z\left(\Pi_{\bar d}(w)\right) + |w - \Pi_{\bar d}(w)|^2,$$

where $\mathbb{R}^n \supseteq \mathbb{R}^{\bar d} \supseteq \mathbb{R}^d$, and $\Pi_{\bar d}$ is the projection of $\mathbb{R}^n$ onto $\mathbb{R}^{\bar d}$.

We begin by checking the condition (A). It is clear that $\hat F^{\bar o}_z : B_n(0,1) \to \mathbb{R}$ is $\mathcal{C}^k$-smooth.

Thus, to check condition (A), it suffices to establish the following claim.

Claim 3. *There is a constant* C_0 *depending only on* d *and* k *such that*

`C4.1A` $\partial^\alpha_{x,y}\hat F^{\bar o}_z(x,y) \le C_0$ *for* $(x,y) \in B_n(0,1)$ *and* $1 \le |\alpha| \le k$.
`C4.2A` *For* $(x,y) \in B_n(0,1)$,

$$c_1[|y|^2 + \rho^2] \le [\hat F^{\bar o}_z(x,y) + \rho^2] \le C_1[|y|^2 + \rho^2],$$

where, by making $\bar c_{12}$ *sufficiently small, we can ensure that* $\rho > 0$ *is less than any constant determined by* C_0, c_1, C_1, k, d.

Proof. That the first part of the claim, i.e., (`C4.1A`), is true follows from the chain rule and the definition of $\hat F^{\bar o}_z(x,y)$ after rescaling by $\bar\tau$. We proceed to show (`C4.2A`). For any $i' \in [\bar N]$ and any vector v in $\mathbb{R}^d$, for ρ taken to be the value from Theorem 13, for $(x,y) \in B_n(0,1)$, the corresponding $z' \in B_n(z,\bar\tau)$ belongs to some $\texttt{cyl}^2_i$. Suppose without loss of generality that $i = 1$ and the other cylinders that contain z' are $2, \dots, t$. Then, $F^{\bar o}$ is a convex combination of $\mathbf{d}(z', \mathcal{A}_1)^2, \dots, \mathbf{d}(z', \mathcal{A}_t)^2$. Note that $y = \mathbf{d}(z', \mathcal{A}_1)/(2\bar\tau)$. Thus, it suffices to prove that for each $j > 1$,

$|\mathbf{d}(z', \mathcal{A}_j) - \mathbf{d}(z', \mathcal{A}_1)| < \bar{\tau}\rho^2/8$. It follows from (c) and (d) of Section 10 that there is a rigid body motion of $\mathtt{cyl}_j^2$ that maps it to an isometric image such that no point of $\mathtt{cyl}_j^2$ is moved by more than $16\bar{\tau}^2/\tau$, such that the image of $\mathcal{A}_j$ is contained inside $o_1(\mathbb{R}^d)$. It follows that $|\mathbf{d}(z', \mathcal{A}_j) - \mathbf{d}(z', \mathcal{A}_1)| < 16\bar{\tau}^2/\tau$, which in turn by a proper choice of $\bar{c}_{12}$ can be made less than $\bar{\tau}\rho^2/8$ as we desire. This ends the proof of Claim 3. □

We proceed to check condition (B). This holds because for every point z in $\mathcal{A}$, the number of i such that the cylinder $\mathtt{cyl}_i$ has a non-empty intersection with a ball of radius $2\sqrt{2}(\bar{\tau})$ centered at z is bounded above by a controlled constant (i.e., a quantity that depends only on d). It follows from (a) of Section 10 that we can choose Θ so that $\Theta(\Pi_{\bar{d}}(w))$ contains the linear span of the d dimensional cross sections of all the cylinders containing z. This, together with the fact that $\mathcal{H}$ is a Hilbert space, is sufficient to yield condition (B). The lemma now follows. □

Let $\mathcal{M}$ belong to $\mathcal{G}(d, V, \tau)$. Let $Y := \{y_1, \dots, y_{\bar{N}}\}$ be a maximal subset of $\mathcal{M}$ with the property that no two distinct points are at a distance of less than $\frac{\bar{\tau}}{2}$ from each other. We construct an *ideal* cylinder packet $\{\mathtt{cyl}_1^2, \dots, \mathtt{cyl}_{\bar{N}}^2\}$ by fixing the center of $\mathtt{cyl}_i^2$ to be y_i, and fixing their orientations by the condition that for each cylinder $\mathtt{cyl}_i^2$, the d dimensional central cross section is a tangent disc to the manifold at y_i. Given an ideal cylinder packet, an *admissible* cylinder packet corresponding to $\mathcal{M}$ is obtained by perturbing the center of each cylinder by less than $\frac{\bar{c}_{12}}{10}\bar{\tau}$ and applying arbitrary unitary transformations to these cylinders whose difference with the identity has an operator norm less than $\frac{\bar{\tau}}{10\tau}$. It is not difficult to check that an admissible cylinder packet is a cylinder packet as per the definition in Section 10.

Lemma 17. *Let $\mathcal{M}$ belong to $\mathcal{G}(d, V, \tau)$, and let $\{\mathtt{cyl}_1, \dots, \mathtt{cyl}_{\bar{N}}\}$ be an admissible packet corresponding to $\mathcal{M}$.*

Then,

$$F^{\bar{o}} \in asdf_{\mathcal{M}}^{\bar{\tau}}.$$

Proof. Recall that $asdf_{\mathcal{M}}^{\bar{\tau}}$ denotes the set of all $\bar{F} : \mathcal{M}_{\bar{\tau}} \to \mathbb{R}$ (where $\bar{\tau} = \bar{c}_{12}\tau$ and $\mathcal{M}_{\bar{\tau}}$ is a $\bar{\tau}$-neighborhood of $\mathcal{M}$) for which the following is true:

- For every $z \in \mathcal{M}$, there exists an isometry Θ of $\mathcal{H}$ that fixes the origin and maps $\mathbb{R}^d$ to a subspace parallel to the tangent plane at z satisfying the conditions below.

  Let $\hat{F}_z : B_n(0,1) \to \mathbb{R}$ be given by

$$\hat{F}_z(w) = \frac{\bar{F}(z + \bar{\tau}\Theta(w))}{\bar{\tau}^2}.$$

  Then, $\hat{F}_z$

  (1) satisfies the hypotheses of Theorem 13 with $k = r + 2 = 4$.

  (2) For any $w \in B_n$,

$$\hat{F}_z(w) = F_z\left(\Pi_{\bar{d}}(w)\right) + |w - \Pi_{\bar{d}}(w)|^2, \tag{102}$$

  where $\mathbb{R}^n \supseteq \mathbb{R}^{\bar{d}} \supseteq \mathbb{R}^d$, and $\Pi_{\bar{d}}$ is the projection of $\mathbb{R}^n$ onto $\mathbb{R}^{\bar{d}}$.

For any fixed $z \in \mathcal{M}$, it suffices to check that there exists a proper isometry Θ of $\mathcal{H}$ such that

(A) The hypotheses of Theorem 13 are satisfied by

$$\hat{F}_z^{\bar{o}}(w) := \frac{F^{\bar{o}}(z + \bar{\tau}\Theta(w))}{\bar{\tau}^2} \tag{103}$$

and

(B)

$$\hat{F}_z^{\bar{o}}(w) = \hat{F}_z^{\bar{o}}(\Pi_{\bar{d}}(w)) + |w - \Pi_{\bar{d}}(w)|^2,$$

where $\mathbb{R}^n \supseteq \mathbb{R}^{\bar{d}} \supseteq \mathbb{R}^d$, and $\Pi_{\bar{d}}$ is the projection of $\mathbb{R}^n$ onto $\mathbb{R}^{\bar{d}}$.

We begin by checking the condition (A). It is clear that $\hat{F}_z^{\bar{o}} : B_n(0,1) \to \mathbb{R}$ is $\mathcal{C}^k$-smooth.

Thus, to check condition (A), it suffices to establish the following claim.

Claim 4. *There is a constant C_0 depending only on d and k such that*

`C4.1` $\partial^\alpha_{x,y}\hat{F}_z^{\bar{o}}(x,y) \leq C_0$ *for* $(x,y) \in B_n(0,1)$ *and* $1 \leq |\alpha| \leq k$.

`C4.2` *For* $(x,y) \in B_n(0,1)$,

$$c_1[|y|^2 + \rho^2] \leq [\hat{F}_z^{\bar{o}}(x,y) + \rho^2] \leq C_1[|y|^2 + \rho^2],$$

where, by making $\bar{c}_{12}$ sufficiently small, we can ensure that $\rho > 0$ is less than any constant determined by C_0, c_1, C_1, k, d.

Proof. That the first part of the claim, i.e., (`C4.1`), is true follows from the chain rule and the definition of $\hat{F}_z^{\bar{o}}(x,y)$ after rescaling by $\bar{\tau}$. We proceed to show (`C4.2`). For any $i \in [\bar{N}]$ and any vector v in $\mathbb{R}^d$, for ρ taken to be the value from Theorem 13, we see that for a sufficiently small value of $\bar{c}_{12} = \frac{\bar{\tau}}{\tau}$ (controlled by d alone), (104) and (105) follow because $\mathcal{M}$ is a manifold of reach greater than or equal to τ, and consequently Proposition 1 holds true,

$$|x_i - \Pi_{\mathcal{M}} x_i| < \frac{\rho^2 \bar{\tau}}{100}, \tag{104}$$

$$\sphericalangle \left(o_i(\mathbb{R}^d), Tan(\Pi_{\mathcal{M}}(x_i), \mathcal{M})\right) \leq \frac{\rho^2}{100}. \tag{105}$$

Making use of Proposition 1 and Claim 1, we see that for any x_i, x_j such that $|x_i - x_j| < 3\bar{\tau}$,

$$\sphericalangle (Tan(\Pi_{\mathcal{M}}(x_i), \mathcal{M}), Tan(\Pi_{\mathcal{M}}(x_j), \mathcal{M})) \leq \frac{3\rho^2}{100}. \tag{106}$$

The inequalities (104), (105), and (106) imply (`C4.2`), completing the proof of the claim. □

We proceed to check condition (B). This holds because for every point z in $\mathcal{M}$, the number of i such that the cylinder $\texttt{cyl}_i$ has a non-empty intersection with a ball of radius $2\sqrt{2}(\bar{\tau})$ centered at z is bounded above by a controlled constant (i.e., a quantity that depends only on d). This, in turn, is because $\mathcal{M}$ has a reach of τ and no two distinct y_i, y_j are at a distance less than $\frac{\bar{\tau}}{2}$ from each other. Therefore, we can choose Θ so that $\Theta(\Pi_{\bar{d}}(w))$ contains the linear span of the d dimensional cross sections of all the cylinders containing z. This, together with the fact that $\mathcal{H}$ is a Hilbert space, is sufficient to yield condition (B). The lemma now follows. □

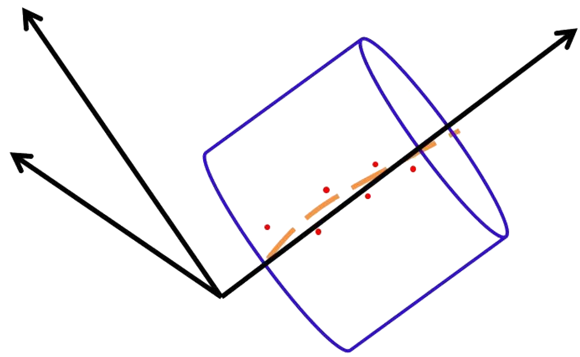

FIGURE 5. Optimizing over local sections.

Definition 19. *Let $\bar{\mathcal{F}}$ be a set of all functions $F^{\bar{o}}$ obtained as $\{\mathtt{cyl}_i^2\}_{i\in[\bar{N}]}$ ranges over all cylinder packets centered on points of a lattice whose spacing is a controlled constant multiplied by τ and the orientations are chosen arbitrarily from a net of the Grassmannian manifold Gr_d^n (with the usual Riemannian metric) of scale that is a sufficiently small controlled constant.*

By Lemma 17 $\bar{\mathcal{F}}$ has the following property.

Corollary 18. *For every $\mathcal{M} \in \mathcal{G}$ that is a $\mathcal{C}^r$-submanifold, there is some $\hat{F} \in \bar{\mathcal{F}}$ that is an approximate-squared-distance function for $\mathcal{M}$; i.e., the restriction of $\hat{F}$ to $\mathcal{M}_{\bar{\tau}}$ is contained in $asdf_{\mathcal{M}}^{\bar{\tau}}$.*

13. Finding good local sections

Definition 20. *Let $(x_1, y_1), \ldots, (x_N, y_N)$ be ordered tuples belonging to $B_d \times B_{n-d}$, and let $r \in \mathbb{N}$. Recall that by Definition 3, the value of r is 2. However, in the interest of clarity, we will use the symbol r to denote the number of derivatives. We say that a function*

$$f : B_d \to B_{n-d}$$

is an ϵ-optimal interpolant if the $\mathcal{C}^r$-norm of f (see Definition 22) satisfies

$$\|f\|_{\mathcal{C}^r} \leq c,$$

and

$$\sum_{i=1}^{N} |f(x_i) - y_i|^2 \leq CN\epsilon + \inf_{\{\check{f}:\|\check{f}\|_{\mathcal{C}^r}\leq C^{-1}c\}} \sum_{i=1}^{N} |\check{f}(x_i) - y_i|^2, \tag{107}$$

where c and $C > 1$ are some constants depending only on d (see Figure 5).

13.1. **Basic convex sets.** We will denote the codimension $n-d$ by $\bar{n}$. It will be convenient to introduce the following notation. For some $i \in \mathbb{N}$, an "i-Whitney field" is a family $\vec{P} = \{P^x\}_{x\in E}$ of i dimensional vectors of real-valued polynomials P_x indexed by the points x in a finite set $E \subseteq \mathbb{R}^d$. We say that $\vec{P} = (P_x)_{x\in E}$ is a Whitney field "on E," and we write $\mathtt{Wh}_r^{\bar{n}}(E)$ for the vector space of all $\bar{n}$-Whitney fields on E of degree at most r.

Definition 21. *Let $\mathcal{C}^r(\mathbb{R}^d)$ denote the space of all real functions on $\mathbb{R}^d$ that are r-times continuously differentiable and*

$$\sup_{|\alpha|\le r}\sup_{x\in\mathbb{R}^d} |\partial^\alpha f\big|_x| < \infty.$$

For a closed subset $U \in \mathbb{R}^d$ such that U is the closure of its interior U^o, we define the $\mathcal{C}^r$-norm of a function $f : U \to \mathbb{R}$ by

$$\|f\|_{\mathcal{C}^r(U))} := \sup_{|\alpha|\le r}\sup_{x\in U^o} |\partial^\alpha f\big|_x|. \tag{108}$$

When U is clear from context, we will abbreviate $\|f\|_{\mathcal{C}^r(U)}$ to $\|f\|_{\mathcal{C}^r}$.

Definition 22. *We define $\mathcal{C}^r(B_d, B_{\bar n})$ to consist of all $f : B_d \to B_{\bar n}$ such that $f(x) = (f^1(x), \dots, f^{\bar n}(x))$ and for each $i \in \bar n$, $f_i : B_d \to \mathbb{R}$ belongs to $\mathcal{C}^r(B_d)$. We define the $\mathcal{C}^r$-norm of $f(x) := (f^1(x), \dots, f^{\bar n}(x))$ by*

$$\|f\|_{\mathcal{C}^r(B_d,B_{\bar n})} := \sup_{|\alpha|\le r}\sup_{v\in B_{\bar n}}\sup_{x\in B_d} |\partial^\alpha(\langle f, v\rangle)\big|_x|.$$

Suppose $F \in \mathcal{C}^r(B_d)$, and $x \in B_d$, we denote by $J_x(F)$ the polynomial that is the rth *order Taylor approximation to F at x, and call it the "jet of F at x."*

If $\vec P = \{P_x\}_{x\in E}$ is an $\bar n$-Whitney field, and $F \in \mathcal{C}^r(B_d, B_{\bar n})$, then we say that "$F$ agrees with $\vec P$," or "F is an extending function for $\vec P$," provided $J_x(F) = P_x$ for each $x \in E$. If $E^+ \supset E$, and $(P^+_x)_{x\in E^+}$ is an $\bar n$-Whitney field on E^+, we say that $\vec P^+$ "agrees with $\vec P$ on E" if for all $x \in E$, $P_x = P^+_x$. We define a C^r-norm on $\bar n$-Whitney fields as follows. If $\vec P \in \mathtt{Wh}^{\bar n}_r(E)$, we define

$$\|\vec P\|_{\mathcal{C}^r(E)} = \inf_F \|F\|_{\mathcal{C}^r(B_d,B_{\bar n})}, \tag{109}$$

where the infimum is taken over all $F \in \mathcal{C}^r(B_d, B_{\bar n})$ such that F agrees with $\vec P$.

We are interested in the set of all $f \in \mathcal{C}^r(B_d, B_{\bar n})$ such that $\|f\|_{\mathcal{C}^r(B_d,B_{\bar n})} \le 1$. By results of Fefferman (see page 19 in [24]) we have the following.

Theorem 19. *Given $\epsilon > 0$, a positive integer r and a finite set $E \subset \mathbb{R}^d$, it is possible to construct in time and space bounded by $\exp(C/\epsilon)|E|$ (where C is controlled by d and r) a set E^+ and a convex set K having the following properties:*

- *Here K is the intersection of $\bar m \le \exp(C/\epsilon)|E|$ sets $\{x|(\alpha_i(x))^2 \le \beta_i\}$, where $\alpha_i(x)$ is a real valued linear function such that $\alpha(0) = 0$ and $\beta_i > 0$. Thus*

$$K := \{x|\forall i \in [\bar m],\, (\alpha_i(x))^2 \le \beta_i\} \subset \mathtt{Wh}^1_r(E^+).$$

- *If $\vec P \in \mathtt{Wh}^1_r(E^+)$ such that $\|\vec P\|_{\mathcal{C}^r(E)} \le 1-\epsilon$, then there exists a Whitney field $\vec P^+ \in K$ that agrees with $\vec P$ on E.*
- *Conversely, if there exists a Whitney field $\vec P^+ \in K$ that agrees with $\vec P$ on E, then $\|\vec P\|_{\mathcal{C}^r(E)} \le 1+\epsilon$.*

For our purposes, it would suffice to set the above ϵ to any controlled constant. To be specific, we set ϵ to $\frac{1}{2}$. By Theorem 1 of [23] we know the following.

Theorem 20. *There exists a linear map T from $\mathcal{C}^r(E)$ to $\mathcal{C}^r(\mathbb{R}^d)$ and a controlled constant C such that $Tf\big|_E = f$ and $\|Tf\|_{\mathcal{C}^r(\mathbb{R}^d)} \le C\|f\|_{\mathcal{C}^r(E)}$.*

Definition 23. *For $\{\alpha_i\}$ as in Theorem 19, let $\bar{K} \subset \bigoplus_{i=1}^{\bar{n}} \mathtt{Wh}_r^1(E^+)$ be the set of all $(x_1, \dots, x_{\bar{n}}) \in \bigoplus_{i=1}^{\bar{n}} \mathtt{Wh}_r^1(E^+)$ (where each $x_i \in \mathtt{Wh}_r^1(E^+)$) such that for each $i \in [\bar{m}]$*

$$\sum_{j=1}^{\bar{n}} (\alpha_i(x_j))^2 \leq \beta_i.$$

Thus, $\bar{K}$ is an intersection of $\bar{m}$ convex sets, one for each linear constraint α_i. We identify $\bigoplus_{i=1}^{\bar{n}} \mathtt{Wh}_r^1(E^+)$ with $\mathtt{Wh}_r^{\bar{n}}(E^+)$ via the natural isomorphism. Then, from Theorem 19 and Theorem 20 we obtain the following.

Corollary 21. *There is a controlled constant C depending on r and d such that*

- *If $\vec{P}$ is a $\bar{n}$-Whitney field on E such that $\|\vec{P}\|_{\mathcal{C}^r(E,\mathbb{R}^{\bar{n}})} \leq C^{-1}$, then there exists a $\bar{n}$-Whitney field $\vec{P}^+ \in \bar{K}$ that agrees with $\vec{P}$ on E.*
- *Conversely, if there exists a $\bar{n}$-Whitney field $\vec{P}^+ \in \bar{K}$ that agrees with $\vec{P}$ on E, then $\|\vec{P}\|_{\mathcal{C}^r(E,\mathbb{R}^{\bar{n}})} \leq C$.*

13.2. **Preprocessing.** Let $\bar{\epsilon} > 0$ be an error parameter.

Notation 1. *For $n \in \mathbb{N}$, we denote the set $\{1, \dots, n\}$ by $[n]$. Let $\{x_1, \dots, x_N\} \subseteq \mathbb{R}^d$.*

Suppose $x_1, \dots, x_N$ is a set of data points in $\mathbb{R}^{\bar{d}}$ and $y_1, \dots, y_N$ are corresponding values in $\mathbb{R}^{\bar{n}}$. The following procedure constructs a function $p : [N] \to [N]$ such that $\{x_{p(i)}\}_{i \in [N]}$ is an $\bar{\epsilon}$-net of $\{x_1, \dots, x_N\}$. For $i = 1$ to N, we sequentially define sets S_i and construct p.

Let $S_1 := \{1\}$ and $p(1) := 1$. For any $i > 1$,

(1) if $\{j : j \in S_{i-1} \text{ and } |x_j - x_i| < \bar{\epsilon}\} \neq \emptyset$, set $p(i)$ to be an arbitrary element of $\{j : j \in S_{i-1} \text{ and } |x_j - x_i| < \bar{\epsilon}\}$, and set $S_i := S_{i-1}$,

(2) and otherwise set $p(i) := i$, and set $S_i := S_{i-1} \cup \{i\}$.

Finally, set $S := S_N$, $\hat{N} = |S|$, and for each i, let

$$h(i) := \{j : p(j) = i\}.$$

For $i \in S$, let $\mu_i := N^{-1}\big|h(i)\big|$, and let

$$\bar{y}_i := \left(\frac{1}{|h(i)|}\right) \sum_{j \in h(i)} y_j. \tag{110}$$

It is clear from the construction that for each $i \in [N]$, $|x_{p(i)} - x_i| \leq \bar{\epsilon}$. The construction of S ensures that the distance between any two points in S is at least $\bar{\epsilon}$. The motivation for sketching the data in this manner was that now, the extension problem involving $E = \{x_i | i \in S\}$ that we will have to deal with will be better conditioned in a sense explained in the following subsection.

13.3. **Convex program.** Let the indices in $[N]$ be permuted so that $S = [\hat{N}]$. For any f such that $\|f\|_{\mathcal{C}^2} \leq C^{-1}c$, and $|x - y| < \bar{\epsilon}$, we have $|f(x) - f(y)| < \bar{\epsilon}$ (and so the grouping and averaging described in the previous section do not affect the quality of our solution); therefore we see that in order to find a $\bar{\epsilon}$-optimal interpolant, it suffices to minimize the objective function

$$\zeta := \sum_{i=1}^{\hat{N}} \mu_i |\bar{y}_i - P_{x_i}(x_i)|^2,$$

over all $\vec{P} \in \bar{K} \subseteq \mathtt{Wh}_r^{\bar{n}}(E^+)$, to within an additive error of $\bar{\epsilon}$, and to find the corresponding point in $\bar{K}$. We note that ζ is a convex function over $\bar{K}$.

Lemma 22. *Suppose that the distance between any two points in E is at least $\bar{\epsilon}$. Suppose $\vec{P} \in \mathtt{Wh}_r^1(E^+)$ has the property that for each $x \in E$, every coefficient of P_x is bounded above by $c'\bar{\epsilon}^2$. Then, if c' is less than some controlled constant depending on d,*

$$\|\vec{P}\|_{\mathcal{C}^2(E)} \leq 1.$$

Proof. Let

$$f(x) = \sum_{z \in E} \theta\left(\frac{10(x-z)}{\bar{\epsilon}}\right) P_z(x).$$

By the properties of θ listed in Section 12, we see that f agrees with $\vec{P}$ and that $\|f\|_{\mathcal{C}^2(\mathbb{R}^d)} \leq 1$ if c' is bounded above by a sufficiently small controlled constant. □

Let $z_{opt} \in \bar{K}$ be any point such that

$$\zeta(z_{opt}) = \inf_{z' \in \bar{K}} \zeta(z').$$

Observation 7. *By Lemma 22 we see that the set K contains a Euclidean ball of radius $c'\bar{\epsilon}^2$ centered at the origin, where c' is a controlled constant depending on d.*

It follows that $\bar{K}$ contains a Euclidean ball of the same radius $c'\bar{\epsilon}^2$ centered at the origin. Due to the fact that the magnitudes of the first m derivatives at any point in E^+ are bounded by C, every point in $\bar{K}$ is at a Euclidean distance of at most $C\hat{N}$ from the origin. We can bound $\hat{N}$ from above as follows:

$$\hat{N} \leq \frac{C}{\bar{\epsilon}^d}.$$

Thanks to Observation 7 and facts from computer science, we will see in a few paragraphs that the relevant optimization problems are tractable.

13.4. **Complexity.** Since we have an explicit description of $\bar{K}$ as in intersection of cylinders, we can construct a "separation oracle," which, when fed with z, does the following:

- If $z \in \bar{K}$, then the separation oracle outputs "Yes."
- If $z \notin \bar{K}$, then the separation oracle outputs "No" and in addition outputs a real affine function $a : \mathtt{Wh}_r^{\bar{n}}(E^+) \to \mathbb{R}$ such that $a(z) < 0$ and $\forall z' \in \bar{K}$ $a(z') > 0$.

To implement this separation oracle for $\bar{K}$, we need to do the following. Suppose we are presented with a point $x = (x_1, \dots, x_{\bar{n}}) \in \mathtt{Wh}_r^{\bar{n}}(E^+)$, where each $x_j \in \mathtt{Wh}_r^1(E^+)$.

(1) If, for each $i \in [\bar{m}]$,

$$\sum_{j=1}^{\bar{n}} (\alpha_i(x_j))^2 \leq \beta_i$$

holds, then declare that $x \in \bar{K}$.

(2) Else, let there be some $i_0 \in [\bar{m}]$ such that

$$\sum_{j=1}^{\bar{n}} (\alpha_{i_0}(x_j))^2 > \beta_{i_0}.$$

Output the following separating half-space:

$$\{(y_1, \dots, y_{\bar{n}}) : \sum_{j=1}^{\bar{n}} \alpha_{i_0}(x_j)\alpha_{i_0}(y_j - x_j) \leq 0\}.$$

The complexity A_0 of answering the above query is the complexity of evaluating $\alpha_i(x_j)$ for each $i \in [\bar{m}]$ and each $j \in [\bar{n}]$. Thus

$$A_0 \leq \bar{n}\bar{m}(dim(K)) \leq Cn\hat{N}^2. \tag{111}$$

Claim 5. *For some $a \in \bar{K}$,*

$$B(a, 2^{-L}) \subseteq \{z \in \bar{K} | \zeta(z) - \zeta(z_{opt}) < \bar{\epsilon}\} \subseteq B(0, 2^L),$$

where L can be chosen so that $L \leq C(1 + |\log(\bar{\epsilon})|)$.

Proof. By Observation 7, we see that the diameter of $\bar{K}$ is at most $C\bar{\epsilon}^{-d}$ and $\bar{K}$ contains a ball B_L of radius 2^{-L}. Let the convex hull of B_L and the point z_{opt} be K_h. Then,

$$\{z \in K_h | \zeta(z) - \zeta(z_{opt}) < \bar{\epsilon}\} \subseteq \{z \in \bar{K} | \zeta(z) - \zeta(z_{opt}) < \bar{\epsilon}\}$$

because $\bar{K}$ is convex. Let the set of all $\vec{P} \in \mathtt{Wh}_r^{\bar{n}}(E^+)$ at which

$$\zeta := \sum_{i=1}^{\hat{N}} \mu_i |\bar{y}_i - P_{x_i}(x_i)|^2 = 0$$

be the affine subspace H. Let $f : \mathtt{Wh}_r^{\bar{n}}(E^+) \to \mathbb{R}$ given by

$$f(x) = \mathbf{d}(x, z_{opt}) := |x - z_{opt}|,$$

where $|\cdot|$ denotes the Euclidean norm. We see that the magnitude of the gradient of ζ identity. Therefore,

$$\{z \in K_h | \zeta(z) - \zeta(z_{opt}) < \bar{\epsilon}\} \supseteq \{z \in K_h \big| 2C\hat{N}(f(z)) < \bar{\epsilon}\}.$$

We note that

$$\{z \in K_h \big| 2C\hat{N}(f(z)) < \bar{\epsilon}\} = K_h \cap B\left(z_{opt}, \frac{\bar{\epsilon}}{2C\hat{N}}\right),$$

where the right hand side denotes the intersection of K_h with a Euclidean ball of radius $\frac{\bar{\epsilon}}{2C\hat{N}}$ and center z_{opt}. By the definition of K_h, $K_h \cap B\left(z_{opt}, \frac{\bar{\epsilon}}{2C\hat{N}}\right)$ contains a ball of radius 2^{-2L}. This proves the claim. □

Given a separation oracle for $\bar{K} \in \mathbb{R}^{\bar{n}(\dim(K))}$ and the guarantee that for some $a \in \bar{K}$,

$$B(a, 2^{-L}) \subseteq \{z \in \bar{K} | \zeta(z) - \zeta(z_{opt}) < \bar{\epsilon}\} \subseteq B(0, 2^L), \tag{112}$$

if $\epsilon > \bar{\epsilon} + \zeta(z_{opt})$, Vaidya's algorithm (see [53]) finds a point in $\bar{K} \cap \{z | \zeta(z) < \epsilon\}$ using

$$O(\dim(\bar{K})A_0 L' + \dim(\bar{K})^{3.38} L')$$

arithmetic steps, where $L' \leq C(L + |\log(\bar{\epsilon}))|)$. Here A_0 is the number of arithmetic operations required to answer a query to the separation oracle.

Let ϵ_{va} denote the smallest real number such that

(1)

$$\epsilon_{va} > \bar{\epsilon}.$$

(2) For any $\epsilon > \epsilon_{va}$, Vaidya's algorithm finds a point in $\bar{K} \cap \{z|\zeta(z) < \epsilon\}$ using

$$O(\dim(\bar{K})A_0 L' + \dim(\bar{K})^{3.38} L')$$

arithmetic steps, where $L' \leq C(1 + |\log(\bar{\epsilon}))|)$.

A consequence of (112) is that $\epsilon_{va} \in [2^{-L}, 2^{L+1}]$. It is therefore clear that ϵ_{va} can be computed to within an additive error of $\bar{\epsilon}$ using binary search and $C(L + |\ln \bar{\epsilon}|)$ calls to Vaidya's algorithm.

The total number of arithmetic operations is therefore

$$O(\dim(\bar{K})A_0 L^2 + \dim(\bar{K})^{3.38} L^2)$$

where $L \leq C(1 + |\log(\bar{\epsilon})|)$.

14. Patching local sections together

This section starts by defining local sections (to be called $\mathcal{M}^i_{fin}$ later) and concludes with the definition of the final manifold $\mathcal{M}_{fin}$, which is obtained by patching local sections together.

For any $i \in [\bar{N}]$, recall the cylinders $\mathtt{cyl}_i$ and Euclidean motions o_i from Section 10.

Let $\mathtt{base}(\mathtt{cyl}_i) := o_i(\mathtt{cyl} \cap \mathbb{R}^d)$ and $\mathtt{stalk}(\mathtt{cyl}_i) := o_i(\mathtt{cyl} \cap \mathbb{R}^{n-d})$. Let $\check{f}_i : B_d \to B_{n-d}$ be an arbitrary C^2 function such that

$$\|\check{f}_i\|_{\mathcal{C}^2} \leq \frac{2\bar{\tau}}{\tau}. \tag{113}$$

Let $f_i : \mathtt{base}(\mathtt{cyl}) \to \mathtt{stalk}(\mathtt{cyl})$ be given by

$$f_i(x) = \bar{\tau}\check{f}_i\left(\frac{x}{\bar{\tau}}\right). \tag{114}$$

Now, fix an $i \in [\bar{N}]$. Without loss of generality, we will drop the subscript i (having fixed this i) and assume that $o_i := id$, by changing the frame of reference using a proper rigid body motion. Recall that $\hat{F}^{\bar{o}}$ was defined by (103), i.e.,

$$\hat{F}^{\bar{o}}(w) := \frac{F^{\bar{o}}(\bar{\tau}w)}{\bar{\tau}^2}$$

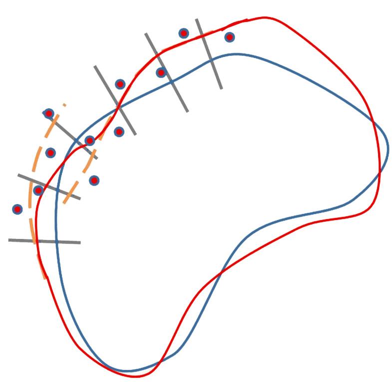

Figure 6. Patching local sections together: base manifold in blue, final manifold in red, and local sections in yellow.

(now 0 and $o_i = id$ play the role that z and Θ played in (103)). Let $N(z)$ be the linear subspace spanned by the top $n-d$ eigenvectors of the Hessian of $\hat{F}^{\bar{o}}$ at a variable point z. Let the intersection of

$$B_d(0,1) \times B_{n-d}(0,1)$$

with

$$\{\tilde{z} \big| \langle \partial \hat{F}^{\bar{o}}\big|_{\tilde{z}}, v\rangle = 0 \, \text{for all} \, v \in \Pi_{hi}(\tilde{z})(\mathbb{R}^n)\}$$

be locally expressed as the graph of a function g_i, where

$$g_i : B_d(0,1) \to \mathbb{R}^{n-d}. \tag{115}$$

For this fixed i, we drop the subscript and let $g : B_d(0,1) \to \mathbb{R}^{n-d}$ be given by

$$g := g_i. \tag{116}$$

As in (84), we see that

$$\Gamma = \{w \,|\, \Pi_{hi}(w)\partial F^{\bar{o}}(w) = 0\}$$

lies in $\mathbb{R}^{\bar{d}}$, and the manifold $\mathcal{M}_{put}$ obtained by patching up all such manifolds for $i \in [\bar{N}]$ is, as a consequence of Proposition 1 and Theorem 13, a submanifold, whose reach is at least $c\tau$. Let

$$\bar{D}^{\mathtt{norm}}_{\bar{F}^{\bar{o}}} \to \mathcal{M}_{put}$$

be the bundle over $\mathcal{M}_{put}$ defined by (89).

Let s_i be the local section of $\bar{D}^{\mathtt{norm}} := \bar{D}^{\mathtt{norm}}_{\bar{F}^{\bar{o}}}$ defined by

$$\{z + s_i(z) | z \in U_i\} := o_i\left(\{x + f_i(x)\}_{x \in \mathtt{base(cyl)}}\right), \tag{117}$$

where $U := U_i \subseteq \mathcal{M}_{put}$ is an open set fixed by (117). The choice of $\frac{\bar{\tau}}{\tau}$ in (113) is small enough to ensure that there is a unique open set U and a unique s_i such that (117) holds (by Observations 1, 2, and 3). We define U_j for any $j \in [\bar{N}]$ analogously. Next, we construct a partition of unity on $\mathcal{M}_{put}$. For each $j \in [\bar{N}]$, let $\tilde{\theta}_j : \mathcal{M}_{put} \to [0,1]$ be an element of a partition of unity defined as follows. For $x \in \mathtt{cyl}_j$,

$$\tilde{\theta}_j(x) := \begin{cases} \theta\left(\frac{\Pi_d(o_j^{-1}x)}{\bar{\tau}}\right), & \text{if } x \in \mathtt{cyl}_j; \\ 0 & \text{otherwise,} \end{cases}$$

where θ is defined by (98). Let

$$\theta_j(z) := \frac{\tilde{\theta}_j(z)}{\sum_{j' \in [\bar{N}]} \tilde{\theta}_{j'}(z)}. \tag{118}$$

We use the local sections $\{s_j | j \in [\bar{N}]\}$, defined separately for each j by (117) and the partition of unity $\{\theta_i\}_{i \in \bar{N}}$, to obtain a global section s of $D^{\mathtt{norm}}_{\bar{o}}$ defined as follows for $x \in U_i$ (see Figure 6):

$$s(x) := \sum_{j \in [\bar{N}]} \theta_j(x) s_j(x). \tag{119}$$

We also define $f : V_i \to B_{n-d}$ by

$$\{z + s(z) | z \in U_i\} := \{x + \bar{\tau} f(x/\bar{\tau})\}_{x \in V_i}. \tag{120}$$

The above equation fixes an open set V_i in $\mathbb{R}^d$. The graph of s, i.e.,

$$\big\{(x+s(x))\big|x \in \mathcal{M}_{put}\big\} =: \mathcal{M}_{fin}, \tag{121}$$

is the output manifold. We see that (121) defines a manifold $\mathcal{M}_{fin}$, by checking this locally. We will obtain a lower bound on the reach of $\mathcal{M}_{fin}$ in Section 15.

15. The reach of the final manifold $\mathcal{M}_{fin}$

Recall that $\hat{F}^{\bar{o}}$ was defined by (103), i.e.,

$$\hat{F}^{\bar{o}}(w) := \frac{F^{\bar{o}}(\bar{\tau}w)}{\bar{\tau}^2}$$

(now 0 and $o_i = id$ play the role that z and Θ played in (103)). We place ourselves in the context of Observation 3. By construction, $F^{\bar{o}} : B_n \to \mathbb{R}$ satisfies the conditions of Theorem 13, and therefore there exists a map

$$\Phi : B_n(0, \bar{c}_{11}) \to B_d(0, \bar{c}_{10}) \times B_{n-d}\left(0, \frac{\bar{c}_{10}}{2}\right),$$

satisfying the following condition:

$$\Phi(z) = (x, \Pi_{n-d} v), \tag{122}$$

where

$$z = x + g(x) + v$$

and

$$v \in N(x + g(x)).$$

Also, x and v are C^r-smooth functions of $z \in B_n(0, \bar{c}_{11})$ with derivatives of order up to r bounded above by C. Let

$$\check{\Phi} : B_n(0, \bar{c}_{11}\bar{\tau}) \to B_d(0, \bar{c}_{10}\bar{\tau}) \times B_{n-d}\left(0, \frac{\bar{c}_{10}\bar{\tau}}{2}\right) \tag{123}$$

be given by

$$\check{\Phi}(x) = \bar{\tau}\Phi(x/\bar{\tau}).$$

Let D_g be the disc bundle over the graph of g, whose fiber at $x+g(x)$ is the disc

$$B_n\left(x + g(x), \frac{\bar{c}_{10}}{2}\right) \bigcap N(x + g(x)).$$

By Lemma 23 below, we can ensure, by setting $\bar{c}_{12} \le \bar{c}$ for a sufficiently small controlled constant $\bar{c}$, that the derivatives of $\Phi - id$ of order less than or equal to $r = k - 2$ are bounded above by a prescribed controlled constant c.

Lemma 23. *For any controlled constant c, there is a controlled constant $\bar{c}$ such that if $\bar{c}_{12} \le \bar{c}$, then for each $i \in [\bar{N}]$ and each $|\alpha| \le 2$ the functions Φ and g, respectively, defined in* (122) *and* (116) *satisfy*

$$|\partial^\alpha(\Phi - id)| \le c,$$

$$|\partial^\alpha g| \le c.$$

Proof of Lemma 23. We would like to apply Theorem 13 here, but its conclusion would not directly help us, since it would give a bound of the form

$$|\partial^\alpha \Phi| \leq C,$$

where C is some controlled constant. To remedy this, we are going to use a simple scaling argument. We first provide an outline of the argument. We change scale by "zooming out," then apply Theorem 13, and thus obtain a bound of the desired form

$$|\partial^\alpha (\Phi - id)| \leq c.$$

We replace each cylinder $\mathtt{cyl}_j = o_j(\mathtt{cyl})$ by $\check{\mathtt{cyl}}_j := o_j(\bar{\tau}(B_d \times (\check{C} B_{n-d})))$. Since the guarantees provided by Theorem 13 have an unspecified dependence on $\bar{d}$ (which appears in (102)), we require an upper bound on the "effective dimension" that depends only on d and is independent of $\check{C}$. If we were only to "zoom out," this unspecified dependence on $\bar{d}$ renders the bound useless. To mitigate this, we need to modify the cylinders that are far away from the point of interest. More precisely, we consider a point $x \in \check{cyl}_i$ and replace each $\mathtt{cyl}_j$ that does not contribute to $\Phi(x)$ by $\check{\mathtt{cyl}}_j$, a suitable translation of

$$\bar{\tau}(B_d \times (\check{C} B_{n-d})).$$

This ensures that the dimension of

$$\{\sum_j \lambda_j v_j |\, \lambda_j \in \mathbb{R}, v_j \in \check{o}_j(\mathbb{R}^d)\}$$

is bounded above by a controlled constant depending only on d. We then apply Theorem 13 to the function $\check{F}^{\check{o}}(w)$ defined in (125). This concludes the outline; we now proceed with the details.

Recall that we have fixed our attention to $\check{cyl}_i$. Let

$$\check{\mathtt{cyl}} := \bar{\tau}(B_d \times (\check{C} B_{n-d})) = \check{\mathtt{cyl}}_i,$$

where $\check{C}$ is an appropriate (large) controlled constant, whose value will be specified later.

Let

$$\check{\mathtt{cyl}}^2 := 2\bar{\tau}(B_d \times (\check{C} B_{n-d})) = \check{\mathtt{cyl}}_i^2.$$

Given a packet $\bar{o} := \{o_1, \dots, o_{\bar{N}}\}$, define a collection of cylinders

$$\{\check{\mathtt{cyl}}_j | j \in [\check{N}]\}$$

in the following manner. Let

$$\check{S} := \left\{ j \in [\bar{N}] \big| |o_j(0)| < 6\bar{\tau} \right\}.$$

Let

$$\check{T} := \left\{ j \in [\bar{N}] \big| |\Pi_d(o_j(0))| < \check{C}\bar{\tau} \text{ and } |o_j(0)| < 4\sqrt{2}\hat{C}\bar{\tau} \right\},$$

and assume without loss of generality that $\check{T} = [\check{N}]$ for some integer $[\check{N}]$. Here $4\sqrt{2}\hat{C}$ is a constant chosen to ensure that for any $j \in [\bar{N}] \setminus [\check{N}]$, $\check{\mathtt{cyl}}^2{}_j \cap \check{\mathtt{cyl}}^2 = \emptyset$. For $v \in \mathbb{R}^n$, let $Tr_v : \mathbb{R}^n \to \mathbb{R}^n$ denote the map that takes x to $x + v$. For any $j \in \check{T} \setminus \check{S}$, let

$$v_j := \Pi_d o_j(0).$$

Next, for any $j \in \check{T}$, let

$$\check{o}_j := \begin{cases} o_j, & \text{if } \check{S}; \\ Tr_{v_j}, & \text{if } j \in \check{S} \setminus \check{T}. \end{cases}$$

For each $j \in \check{T}$, let $\check{\mathtt{cyl}}_j := \check{o}_j(\check{\mathtt{cyl}})$. Define $F^{\check{o}} : \bigcup_{j \in \check{T}} \check{\mathtt{cyl}}_j \to \mathbb{R}$ by

$$F^{\check{o}}(z) = \frac{\sum_{\check{\mathtt{cyl}}^2{}_j \ni z} \left|\Pi_{n-d}(\check{o}_j^{-1}(z))\right|^2 \theta\left(\frac{\Pi_d(\check{o}_j^{-1}(z))}{2\bar{\tau}}\right)}{\sum_{\check{\mathtt{cyl}}^2{}_j \ni z} \theta\left(\frac{\Pi_d(\check{o}_j^{-1}(z))}{2\bar{\tau}}\right)}. \tag{124}$$

Taking $\bar{c}_{12}$ to be a sufficiently small controlled constant depending on $\check{C}$, we see that

$$\check{F}^{\check{o}}(w) := \frac{F^{\check{o}}(\check{C}\bar{\tau}w)}{\check{C}^2\bar{\tau}^2}, \tag{125}$$

restricted to B_n, satisfies the requirements of Theorem 13. Choosing $\check{C}$ to be sufficiently large, for each $|\alpha| \in [2, k]$, the function Φ defined in (122) satisfies

$$|\partial^\alpha \Phi| \leq c, \tag{126}$$

and for each $|\alpha| \in [0, k-2]$, the function g defined in (122) satisfies

$$|\partial^\alpha g| \leq c. \tag{127}$$

Observe that we can choose $j \in [\bar{N}] \setminus [\check{N}]$ such that $|\check{o}_j(0)| < 10\tau$, for this j, $\check{\mathtt{cyl}}_j \cap \check{\mathtt{cyl}} = \emptyset$, and so

$$\partial\Phi\Big|_{(\bar{\tau}^{-1})\check{o}_j(0)} = id. \tag{128}$$

The lemma follows from Taylor's theorem, in conjunction with (126), (127), and (128).

Observation 8. *By choosing $\check{C} \geq 2/\bar{c}_{11}$ we find that the domains of both Φ and Φ^{-1} may be extended to contain the cylinder $\left(\frac{3}{2}\right) B_d \times B_{n-d}$, while satisfying (122).* □

Since $|\partial^\alpha(\Phi - Id)(x)| \leq c$ for $|\alpha| \leq r$ and $x \in \left(\frac{3}{2}\right) B_d \times B_{n-d}$, we have $|\partial^\alpha(\Phi^{-1} - Id)(w)| \leq c$ for $|\alpha| \leq r$ and $w \in B_d \times B_{n-d}$. For the remainder of this section, we will assume a scale where $\bar{\tau} = 1$.

For $u \in U_i$, we have the following equality which we restate from (119) for convenience:

$$s(u) = \sum_{j \in [\bar{N}]} \theta_j(u) s_j(u).$$

Let Π_{pseud} (for "pseudonormal bundle") be the map from a point x in $\mathtt{cyl}$ to the base point belonging to $\mathcal{M}_{put}$ of the corresponding fiber. The following relation exists between Π_{pseud} and Φ:

$$\Pi_{pseud} = \Phi^{-1}\Pi_d\Phi.$$

We define the C^{k-2} norm of a local section s_j over $U \subseteq U_j \cap U_i$ by

$$\|s_j\|_{C^{k-2}(U)} := \|s_j \circ \Phi^{-1}\|_{C^{k-2}(\Pi_d(U))}.$$

Recall that $k-2=r=2$. Suppose for a specific x and t,

$$x + f_j(x) = t + s_j(t),$$

where t belongs to $U_j \cap U_i$. Applying Π_{pseud} to both sides,

$$\Pi_{pseud}(x + f_j(x)) = t.$$

Let

$$\Pi_{pseud}(x + f_j(x)) =: \phi_j(x).$$

Substituting back, we have

$$x + f_j(x) = \phi_j(x) + s_j(\phi_j(x)). \tag{129}$$

By definition 20, we have the bound $\|f_j\|_{C^{k-2}(\phi_j^{-1}(U_i \cap U_j))} \leq c$. We have

$$\Pi_{pseud}(x + f_j(x)) = (\Pi_{pseud} - \Pi_d)(x + f_j(x)) + x,$$

which gives the bound

$$\|\phi_j - Id\|_{C^{k-2}(\phi_j^{-1}(U_i \cap U_j))} \leq c.$$

Therefore, from (129),

$$\|s_j \circ \phi_j\|_{C^{k-2}(\phi_j^{-1}(U_i \cap U_j))} \leq c. \tag{130}$$

Also,

$$\|\phi_j^{-1} \circ \Phi^{-1} - Id\|_{C^{k-2}(\Pi_d(U_i \cap U_j))} \leq c. \tag{131}$$

From the preceding two equations, it follows that

$$\|s_j\|_{C^{k-2}(U_i \cap U_j)} \leq c. \tag{132}$$

The cutoff functions θ_j satisfy

$$\|\theta_j\|_{C^{k-2}(U_i \cap U_j)} \leq C. \tag{133}$$

Therefore, by (119),

$$\|s\|_{C^{k-2}(U_i \cap U_j)} \leq Cc, \tag{134}$$

which we rewrite as

$$\|s\|_{C^{k-2}(U_i \cap U_j)} \leq c_1. \tag{135}$$

Recall the statements surrounding (120) for a definition of V_i. We will now show that

$$\|f\|_{C^{k-2}(V_i)} \leq c.$$

By (120) in view of $\bar{\tau} = 1$, for $u \in U_i$, there is an $x \in V_i$ such that

$$u + s(u) = x + f(x).$$

This gives us

$$\Pi_d(u + s(u)) = x.$$

Substituting back, we have

$$\Pi_d(u + s(u)) + f(\Pi_d(u + s(u))) = u + s(u).$$

Let

$$\psi(u) := \Pi_d(u + s(u)).$$

This gives us

(136) $$f(\psi(u)) = (u - \psi(u)) + s(u).$$

By (135) and the fact that $|\partial^\alpha(\Phi - Id)(x)| \leq c$ for $|\alpha| \leq r$, we see that

(137) $$\|\psi - Id\|_{C^{k-2}(U_i)} \leq c.$$

By (136), (137), and (135), we have $\|f \circ \psi\|_{C^{k-2}(U_i)} \leq c$.

By (137), we have

$$\|\psi^{-1} - Id\|_{C^{k-2}(V_i)} \leq c.$$

Therefore,

(138) $$\|f\|_{C^{k-2}(V_i)} \leq c.$$

For any point $u \in \mathcal{M}_{put}$, there is by Lemma 14 for some $j \in [\bar{N}]$, a U_j such that $\mathcal{M}_{put} \cap B(u, 1/10) \subseteq U_j$ (recall that $\bar{\tau} = 1$). Therefore, suppose a, b are two points on $\mathcal{M}_{fin}$ such that $|a - b| < 1/20$, then $|\Pi_{pseud}(a) - \Pi_{pseud}(b)| < 1/10$, and so both $\Pi_{pseud}(a)$ and $\Pi_{pseud}(b)$ belong to U_j for some j. Without loss of generality, let this j be i. This implies that a, b are points on the graph of f over V_i. Then, by (138) and Proposition 1, $\mathcal{M}_{fin}$ is a manifold whose reach is at least $c\tau$.

16. The mean-squared distance to the final manifold $\mathcal{M}_{fin}$ from a random data point

Let $\mathcal{M}_{opt}$ be an approximately optimal manifold in that

$$\text{reach}(\mathcal{M}_{opt}) > C\tau,$$

$$\text{vol}(\mathcal{M}_{opt}) < V/C,$$

and

$$\mathbb{E}_{\mathcal{P}}\mathbf{d}(x, \mathcal{M}_{opt})^2 \leq \inf_{\mathcal{M} \in \mathcal{G}(d, C\tau, cV)} \mathbb{E}_{\mathcal{P}}\mathbf{d}(x, \mathcal{M})^2 + \epsilon.$$

Suppose that $\bar{o}$ is the packet from the previous section and that the corresponding function $F^{\bar{o}}$ belongs to $asdf(\mathcal{M}_{opt})$. We need to show that the $\mathcal{M}_{fin}$ constructed using $\bar{o}$ serves the purpose it was designed for, namely, that the following Lemma holds.

Lemma 24.

$$\mathbb{E}_{x \dashv \mathcal{P}}\mathbf{d}(x, \mathcal{M}_{fin})^2 \leq C_0 \left(\mathbb{E}_{x \dashv \mathcal{P}}\mathbf{d}(x, \mathcal{M}_{opt})^2 + \epsilon\right).$$

Proof. Let us examine the manifold $\mathcal{M}_{fin}$. Recall that $\mathcal{M}_{fin}$ was constructed from a collection of local sections $\{s_i\}_{i \in \bar{N}}$, one for each i such that $o_i \in \bar{o}$. These local sections were obtained from functions $f_i : \texttt{base}(\texttt{cyl}_i) \to \texttt{stalk}(\texttt{cyl}_i)$. The s_i were patched together using a partition of unity supported on $\mathcal{M}_{put}$.

Let $\mathcal{P}_{in}$ be the measure obtained by restricting $\mathcal{P}$ to $\cup_{i \in [\bar{N}]} \texttt{cyl}_i$. Let $\mathcal{P}_{out}$ be the measure obtained by restricting $\mathcal{P}$ to $\left(\cup_{i \in [\bar{N}]} \texttt{cyl}_i\right)^c$. Thus,

$$\mathcal{P} = \mathcal{P}_{out} + \mathcal{P}_{in}.$$

For any $\mathcal{M} \in \mathcal{G}$,

(139) $$\mathbb{E}_{\mathcal{P}}\mathbf{d}(x, \mathcal{M})^2 = \mathbb{E}_{\mathcal{P}_{out}}\mathbf{d}(x, \mathcal{M})^2 + \mathbb{E}_{\mathcal{P}_{in}}\mathbf{d}(x, \mathcal{M})^2.$$

We will separately analyze the two terms on the right when $\mathcal{M}$ is $\mathcal{M}_{fin}$. We begin with $\mathbb{E}_{\mathcal{P}_{out}}\mathbf{d}(x, \mathcal{M}_{fin})^2$. We make two observations:

(1) By (113), the function $\check{f}_i$ satisfies

$$\|\check{f}_i\|_{L^\infty} \leq \frac{\bar{\tau}}{\tau}.$$

(2) By Lemma 23, the fibers of the disc bundle $D^{\mathtt{norm}}$ over $\mathcal{M}_{put} \cap \mathtt{cyl}_i$ are nearly orthogonal to $\mathtt{base}(\mathtt{cyl}_i)$.

Therefore, no point outside the union of the $\mathtt{cyl}_i$ is at a distance less than $\bar{\tau}(1 - \frac{2\bar{\tau}}{\tau})$ to $\mathcal{M}_{fin}$.

Since $F^{\bar{o}}$ belongs to $asdf(\mathcal{M}_{opt})$, we see that no point outside the union of the $\mathtt{cyl}_i$ is at a distance less than $\bar{\tau}(1 - C\bar{c}_{12})$ to $\mathcal{M}_{opt}$. Here C is a controlled constant.

For any given controlled constant c, by choosing $\bar{c}_{12}$ (i.e., $\frac{\bar{\tau}}{\tau}$) appropriately, we can arrange for

$$\mathbb{E}_{\mathcal{P}_{out}}[\mathbf{d}(x, \mathcal{M}_{fin})^2] \leq (1 + c)\mathbb{E}_{\mathcal{P}_{out}}[\mathbf{d}(x, \mathcal{M}_{opt})^2] \tag{140}$$

to hold.

Consider terms involving $\mathcal{P}_{in}$ now. We assume without loss of generality that $\mathcal{P}$ possesses a density, since we can always find an arbitrarily small perturbation of $\mathcal{P}$ (in the ℓ^2-Wasserstein metric) that is supported on a ball and also possesses a density. Let

$$\Pi_{put} : \cup_{i \in \bar{N}} \mathtt{cyl}_i \to \mathcal{M}_{put}$$

be the projection which maps a point in $\cup_{i \in \bar{N}} \mathtt{cyl}_i$ to the unique nearest point on $\mathcal{M}_{put}$. Let μ_{put} denote the d dimensional volume measure on $\mathcal{M}_{put}$.

Let $\{\mathcal{P}_{in}^z\}_{z \in \mathcal{M}_{put}}$ denote the natural measure induced on the fiber of the normal disc bundle of radius $2\bar{\tau}$ over z.

Then,

$$\mathbb{E}_{\mathcal{P}_{in}}[\mathbf{d}(x, \mathcal{M}_{fin})^2] = \int_{\mathcal{M}_{put}} \mathbb{E}_{\mathcal{P}_{in}^z}[\mathbf{d}(x, \mathcal{M}_{fin})^2] d\mu_{put}(z). \tag{141}$$

Using the partition of unity $\{\theta_j\}_{j \in [\bar{N}]}$ supported on $\mathcal{M}_{put}$, defined in (118), we split the right hand side of (141). We will soon use pieces of $\mathcal{M}_{fin}$ which we call $\mathcal{M}_{fin}^i$:

$$\int_{\mathcal{M}_{put}} \mathbb{E}_{\mathcal{P}_{in}^z}[\mathbf{d}(x, \mathcal{M}_{fin})^2] d\mu_{put}(z) = \sum_{i \in \bar{N}} \int_{\mathcal{M}_{put}} \theta_i(z) \mathbb{E}_{\mathcal{P}_{in}^z}[\mathbf{d}(x, \mathcal{M}_{fin})^2] d\mu_{put}(z). \tag{142}$$

For $x \in \mathtt{cyl}_i$, let $\mathbb{N}_x$ denote the unique fiber of $D^{\mathtt{norm}}$ that x belongs to. Observe that $\mathcal{M}_{fin} \cap \mathbb{N}_x$ consists of a single point. Define $\tilde{\mathbf{d}}(x, \mathcal{M}_{fin})$ to be the distance of x to this point, i.e.,

$$\tilde{\mathbf{d}}(x, \mathcal{M}_{fin}) := \mathbf{d}(x, \mathcal{M}_{fin} \cap \mathbb{N}_x).$$

We proceed to examine the right hand side in (142).

By (144)

$$\sum_i \int_{\mathcal{M}_{put}} \theta_i(z) \mathbb{E}_{\mathcal{P}_{in}^z}[\mathbf{d}(x, \mathcal{M}_{fin})^2] d\mu_{put}(z) \leq \sum_i \int_{\mathcal{M}_{put}} \theta_i(z) \mathbb{E}_{\mathcal{P}_{in}^z}[\tilde{\mathbf{d}}(x, \mathcal{M}_{fin})^2] d\mu_{put}(z).$$

For each $i \in [\bar{N}]$, let $\mathcal{M}^i_{fin}$ denote the manifold with boundary corresponding to the graph of f_i, i.e., let

$$\mathcal{M}^i_{fin} := \{x + f_i(x)\}_{x \in \mathtt{base}(\mathtt{cyl})}. \tag{143}$$

Since the quadratic function is convex, the average squared "distance" (where "distance" refers to $\tilde{\mathbf{d}}$) to $\mathcal{M}_{fin}$ is less than or equal to the average of the squared distances to the local sections in the following sense:

$$\sum_i \int_{\mathcal{M}_{put}} \theta_i(z)\mathbb{E}_{\mathcal{P}^z_{in}}[\tilde{\mathbf{d}}(x, \mathcal{M}_{fin})^2]d\mu_{put}(z) \le \sum_i \int_{\mathcal{M}_{put}} \theta_i(z)\mathbb{E}_{\mathcal{P}^z_{in}}[\tilde{\mathbf{d}}(x, \mathcal{M}^i_{fin})^2]d\mu_{put}(z).$$

Next, we will look at the summands of the right hand side. Lemma 23 tells us that $\mathbb{N}_x$ is almost orthogonal to $o_i(\mathbb{R}^d)$. By Lemma 23, and the fact that each f_i satisfies (138), we see that

$$\mathbf{d}(x, \mathcal{M}^i_{fin}) \le \tilde{\mathbf{d}}(x, \mathcal{M}^i_{fin}) \le (1 + c_0)\mathbf{d}(x, \mathcal{M}^i_{fin}). \tag{144}$$

Therefore,

$$\sum_i \int_{\mathcal{M}_{put}} \theta_i(z)\mathbb{E}_{\mathcal{P}^z_{in}}[\tilde{\mathbf{d}}(x, \mathcal{M}^i_{fin})^2]d\mu_{put}(z)$$
$$\le (1 + c_0) \sum_i \int_{\mathcal{M}_{put}} \theta_i(z)\mathbb{E}_{\mathcal{P}^z_{in}}[\mathbf{d}(x, \mathcal{M}^i_{fin})^2]d\mu_{put}(z).$$

We now fix $i \in [\bar{N}]$. Let $\mathcal{P}^i$ be the measure which is obtained, by the translation via o_i^{-1} of the restriction of $\mathcal{P}$ to $\mathtt{cyl}_i$. In particular, $\mathcal{P}^i$ is supported on $\mathtt{cyl}$.

Let μ^i_{base} be the push-forward of $\mathcal{P}^i$ onto $\mathtt{base}(\mathtt{cyl})$ under Π_d. For any $x \in \mathtt{cyl}_i$, let $v(x) \in \mathcal{M}^i_{fin}$ be the unique point such that $x - v(x)$ lies in $o_i(\mathbb{R}^{n-d})$. In particular,

$$v(x) = \Pi_d x + f_i(\Pi_d x).$$

By Lemma 23, we see that

$$\int_{\mathcal{M}_{put}} \theta_i(z)\mathbb{E}_{\mathcal{P}^z_{in}}[\tilde{\mathbf{d}}(x, \mathcal{M}^i_{fin})^2]d\mu_{put}(z) \le C_0 \mathbb{E}_{\mathcal{P}^i}|x - v(x)|^2.$$

Recall that $\mathcal{M}^i_{fin}$ is the graph of a function $f_i : \mathtt{base}(\mathtt{cyl}) \to \mathtt{stalk}(\mathtt{cyl})$. In Section 13, we have shown how to construct f_i so that it satisfies (113) and (145), where $\hat{\epsilon} = \frac{c\epsilon}{\bar{N}}$, for some sufficiently small controlled constant c,

$$\mathbb{E}_{\mathcal{P}^i}|f_i(\Pi_d x) - \Pi_{n-d}x|^2 \le \hat{\epsilon} + \inf_{f:\|f\|_{C^r} \le c\bar{\tau}^{-2}} \mathbb{E}_{\mathcal{P}^i}|f(\Pi_d x) - \Pi_{n-d}x|^2. \tag{145}$$

Let $f^{opt}_i : \mathtt{base}(\mathtt{cyl}) \to \mathtt{stalk}(\mathtt{cyl})$ denote the function (which exists because of the bound on the reach of $\mathcal{M}_{opt}$) with the property that

$$\mathcal{M}_{opt} \cap \mathtt{cyl}_i = o_i\left(\{x, f^{opt}_i(x)\}_{x \in \mathtt{base}(\mathtt{cyl})}\right).$$

By (145), we see that

$$\mathbb{E}_{\mathcal{P}^i}|f_i(\Pi_d x) - \Pi_{n-d}x|^2 \le \hat{\epsilon} + \mathbb{E}_{\mathcal{P}^i}|f^{opt}_i(\Pi_d x) - \Pi_{n-d}x|^2. \tag{146}$$

Lemma 23 and the fact that each f_i satisfies (138) and (145) show that

(147) $$\mathbb{E}_{\mathcal{P}_{in}}[\mathbf{d}(x, \mathcal{M}_{fin})^2] \leq C_0 \mathbb{E}_{\mathcal{P}_{in}}[\mathbf{d}(x, \mathcal{M}_{opt})^2] + C_0 \hat{\epsilon}.$$

The proof follows from (139), (140), and (147). □

17. Number of arithmetic operations

After the dimension reduction of Section 6, the ambient dimension is reduced to

$$n := O\left(\frac{N_p \ln^4\left(\frac{N_p}{\epsilon}\right) + \log \delta^{-1}}{\epsilon^2}\right),$$

where

$$N_p := V\left(\tau^{-d} + (\epsilon\tau)^{\frac{-d}{2}}\right).$$

The number of times that local sections are computed is bounded above by the product of the maximum number of cylinders in a cylinder packet (i.e., $\bar{N}$, which is less than or equal to $\frac{CV}{\tau^d}$) and the total number of cylinder packets whose centers are contained inside $B_n \cap (\bar{c}_{13}\tau)\mathbb{Z}_n$. The latter number is bounded above by $(\bar{c}_{13}\tau)^{-n\bar{N}}$. Each optimization for computing a local section requires only a polynomial number of computations as discussed in Section 13.4. Therefore, the total number of arithmetic operations required is bounded above by

$$\exp\left(C\left(\frac{V}{\tau^d}\right) n \ln \tau^{-1}\right).$$

18. Conclusion and future work

We developed an algorithm for testing if data drawn from a distribution supported on a separable Hilbert space have an expected squared distance of $O(\epsilon)$ to a submanifold (of the unit ball) of dimension d and volume at most V and reach at least τ. The number of data points required is of the order of

$$n := \frac{N_p \ln^4\left(\frac{N_p}{\epsilon}\right) + \ln \delta^{-1}}{\epsilon^2},$$

where

$$N_p := V\left(\frac{1}{\tau^d} + \frac{1}{\tau^{d/2}\epsilon^{d/2}}\right),$$

and the number of arithmetic operations and calls to the black box that evaluates inner products in the ambient Hilbert space is

$$\exp\left(C\left(\frac{V}{\tau^d}\right) n \ln \tau^{-1}\right).$$

An interesting question is to fit a manifold to data drawn i.i.d. from the uniform distribution on a manifold in $\mathcal{G}(d, V, \tau)$. In this case an exhaustive search for an appropriate disc bundle is unnecessary. Instead, one can use local principal component analysis to approximately learn the tangent spaces of the manifold from which data are being drawn. These tangent spaces can be used to produce a cylinder packet, which in turn can be used to construct a disc bundle that has as a section the manifold underlying the data. This manifold can be reconstructed by patching together local sections obtained using interpolation.

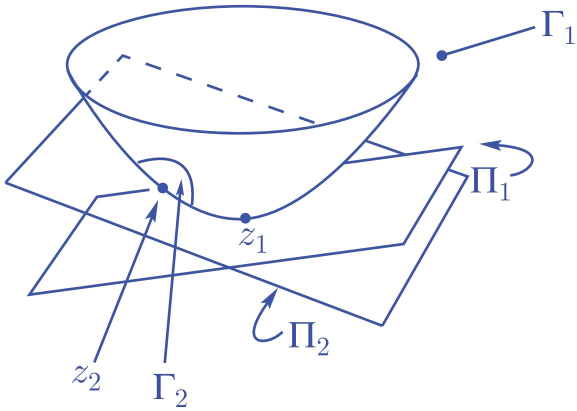

FIGURE 7. A patch.

APPENDIX A. PROOF OF CLAIM 1

The following is an easy consequence of the implicit function theorem in fixed dimension (d or $2d$).

Lemma 25. *Let Γ_1 be a patch of radius r_1 over Π_1 centered at z_1 and tangent to Π_1 at z_1. Let z_2 belong to Γ_1 and suppose $\|z_2 - z_1\| < c_0 r_1$. Assume*

$$\|\Gamma_1\|_{\dot{C}^{1,1}(B_\Pi(z_1,r_1))} \leq \frac{c_0}{r_1}.$$

Let $\Pi_2 \in dPL$ with $dist(\Pi_2, \Pi_1) < c_0$. Then there exists a patch Γ_2 of radius $c_1 r_1$ over Π_2 centered at z_2 with

$$\|\Gamma_2\|_{\dot{C}^{1,1}(B_\Pi(0,c_1r_1))} \leq \frac{200c_0}{r_1},$$

and

$$\Gamma_2 \cap B_{\mathcal{H}}\left(z_2, \frac{c_1r_1}{2}\right) = \Gamma_1 \cap B_{\mathcal{H}}\left(z_2, \frac{c_1r_1}{2}\right).$$

Here c_0 and c_1 are small constants depending only on d, and by rescaling, we may assume without loss of generality that $r_1 = 1$ when we prove Lemma 25.

The meaning of Lemma 25 is that if Γ is the graph of a map

$$\Psi : B_{\Pi_1}(0,1) \to \Pi_1^\perp$$

with $\Psi(0) = 0$ and $\partial\Psi(0) = 0$ and the $C^{1,1}$-norm of Ψ is small, then at any point $z_2 \in \Gamma$ is close to 0, and for any d-plane Π_2 close to Π_1, we may regard Γ near z_2 as the graph Γ_2 of a map

$$\tilde{\Psi} : B_{\Pi_2}(0,c) \to \Pi_2^\perp;$$

here Γ_2 is centered at z_2 and the $C^{1,1}$-norm of $\tilde{\psi}$ is not much bigger than that of Ψ, see Figure 7.

A.0.1. *Growing a patch.*

Lemma 26 ("growing patch")**.** *Let $\mathcal{M}$ be a manifold, and let r_1, r_2 be as in the definition of a manifold. Suppose $\mathcal{M}$ has infinitesimal reach ≥ 1. Let $\Gamma \subset \mathcal{M}$ be a patch of radius r centered at 0, over $T_0\mathcal{M}$. Suppose r is less than a small enough* *constant $\hat{c}$ determined by d. Then there exists a patch Γ^+ of radius $r + cr_2$ over $T_0\mathcal{M}$, centered at 0 such that $\Gamma \subset \Gamma^+ \subset \mathcal{M}$.*

Corollary 27. *Let $\mathcal{M}$ be a manifold with infinitesimal reach ≥ 1 and suppose $0 \in \mathcal{M}$. Then there exists a patch Γ of radius $\hat{c}$ over $T_0\mathcal{M}$ such that $\Gamma \subset \mathcal{M}$.*

Lemma 26 implies Corollary 27. Indeed, we can start with a tiny patch Γ (centered at 0) over $T_0\mathcal{M}$, with $\Gamma \subset \mathcal{M}$. Such Γ exists because $\mathcal{M}$ is a manifold. By repeatedly applying the lemma, we can repeatedly increase the radius of our patch by a fixed amount cr_2; we can continue doing so until we arrive at a patch of radius $\geq \hat{c}$.

Proof of Lemma 26. Without loss of generality, we can take $\mathcal{H} = \mathbb{R}^d \oplus \mathcal{H}'$ for a Hilbert space $\mathcal{H}'$; and we may assume that

$$T_0\mathcal{M} = \mathbb{R}^d \times \{0\} \subset \mathbb{R}^d \oplus \mathcal{H}'.$$

Our patch Γ is then a graph

$$\Gamma = \{(x, \Psi(x)) : x \in B_{\mathbb{R}^d}(0, r)\} \subseteq \mathbb{R}^d \oplus \mathcal{H}'$$

for a $C^{1,1}$ map

$$\Psi : B_{\mathbb{R}^d}(0, r) \to \mathcal{H}',$$

with $\Psi(0) = 0$, $\partial\Psi(0) = 0$, and

$$\|\Psi\|_{\dot{C}^{1,1}(B_{\mathbb{R}^d}(0,r))} \leq C_0.$$

For $y \in B_{\mathbb{R}^d}(0, r)$, we therefore have $|\partial\psi(y)| \leq C_0$. If r is less than a small enough $\hat{c}$, then Lemma 25 together with the fact that $\mathcal{M}$ agrees with a patch of radius r_1 in $B_{\mathbb{R}^d \oplus \mathcal{H}'}((y, \Psi(y)), r_2)$ (because $\mathcal{M}$ is a manifold) tells us that there exists a $C^{1,1}$ map

$$\Psi_y : B_{\mathbb{R}^d}(y, c'r_2) \to \mathcal{H}'$$

such that

$$\begin{aligned}&\mathcal{M} \cap B_{\mathbb{R}^d \oplus \mathcal{H}'}((y, \Psi(y)), c''r_2)\\ &= \{(z, \Psi_y(z)) : z \in B_{\mathbb{R}^d}(y, c'r_2)\} \cap B_{\mathbb{R}^d \oplus \mathcal{H}'}((y, \Psi(y)), c''r_2).\end{aligned}$$

Also, we have a priori bounds on $\|\partial_z \Psi_y(z)\|$ and on $\|\Psi_y\|_{\dot{C}^{1,1}}$. It follows that whenever $y_1, y_2 \in B_{\mathbb{R}^d}(0, r)$ and $z \in B_{\mathbb{R}^d}(y_1, c'''r_2) \cap B_{\mathbb{R}^d}(y_2, c'''r_2)$, we have $\Psi_{y_1}(z) = \Psi_{y_2}(z)$.

This allows us to define a global $C^{1,1}$ function

$$\Psi^+ : B_{\mathbb{R}^d}(0, r + c'''r_2) \to \mathcal{H}';$$

the graph of Ψ^+ is simply the union of the graphs of

$$\Psi_y|_{B_{\mathbb{R}^d}(y, c'''r_2)}$$

as y varies over $B_{\mathbb{R}^d}(0, r)$. Since the graph of each $\Psi_y|_{B_{\mathbb{R}^d}(y,c'''r_2)}$ is contained in $\mathcal{M}$, it follows that the graph of Ψ^+ is contained in $\mathcal{M}$. Also, by definition, Ψ^+ agrees on $B_{\mathbb{R}^d}(y, c'''r_2)$ with a $C^{1,1}$ function, for each $y \in B_{\mathbb{R}^d}(0, r)$. It follows that

$$\|\Psi^+\|_{\dot{C}^{1,1}(0, rc'''r_2)} \leq C.$$

Also, for each $y \in B_{\mathbb{R}^d}(0, r)$, the point $(y, \Psi(y))$ belongs to

$$\mathcal{M} \cap B_{\mathbb{R}^d \oplus \mathcal{H}'}((y, \Psi(y)), \frac{c'''r_2}{1000});$$

hence it belongs to the graph of $\Psi_y|_{B_{\mathbb{R}^d}(y,c'''r_2)}$, and therefore it belongs to the graph of Ψ^+. Thus $\Gamma^+ =$ graph of Ψ^+ satisfies $\Gamma \subset \Gamma^+ \subset \mathcal{M}$, and Γ^+ is a patch of radius $r + c'''r_2$ over $T_0\mathcal{M}$ centered at 0. That proves the lemma. □

A.0.2. *Global reach.* For a real number $\tau > 0$, a manifold $\mathcal{M}$ has reach $\geq \tau$ if and only if every $x \in \mathcal{H}$ such that $\mathbf{d}(x, \mathcal{M}) < \tau$ has a unique closest point of $\mathcal{M}$. By Federer's characterization of the reach in Proposition 1, if the reach is greater than one, the infinitesimal reach is greater than 1 as well.

Lemma 28. *Let $\mathcal{M}$ be a manifold of reach ≥ 1, with $0 \in \mathcal{M}$. Then, there exists a patch Γ of radius $\hat{c}$ over $T_0\mathcal{M}$ centered at 0, such that*

$$\Gamma \cap B_{\mathcal{H}}(0, \check{c}) = \mathcal{M} \cap B_{\mathcal{H}}(0, \check{c}).$$

Proof. There is a patch Γ of radius $\hat{c}$ over $T_0\mathcal{M}$ centered at 0 such that

$$\Gamma \cap B_{\mathcal{H}}(0, c^\sharp) \subseteq \mathcal{M} \cap B_{\mathcal{H}}(0, c^\sharp)$$

(see Lemma 26). For any $x \in \Gamma \cap B_{\mathcal{H}}(0, c^\sharp)$, there exists a tiny ball B_x (in $\mathcal{H}$) centered at x such that $\Gamma \cap B_x = \mathcal{M} \cap B_x$; that is because $\mathcal{M}$ is a manifold.

It follows that the distance from

$$\Gamma_{yes} := \Gamma \cap B_{\mathcal{H}}(0, \frac{c^\sharp}{2})$$

to

$$\Gamma_{no} := \left[\mathcal{M} \cap B_{\mathcal{H}}(0, \frac{c^\sharp}{2})\right] \setminus \left[\Gamma \cap B_{\mathcal{H}}(0, \frac{c^\sharp}{2})\right]$$

is strictly positive.

Suppose Γ_{no} intersects $B_H(0, \frac{c^\sharp}{100})$; say, $y_{no} \in B_{\mathcal{H}}(0, \frac{c^\sharp}{100}) \cap \Gamma_{no}$. Also, $0 \in B_{\mathcal{H}}(0, \frac{c^\sharp}{100}) \cap \Gamma_{yes}$.

The continuous function $B_{\mathcal{H}}(0, \frac{c^\sharp}{100}) \ni y \mapsto \mathbf{d}(y, \Gamma_{no}) - \mathbf{d}(y, \Gamma_{yes})$ is positive at $y = 0$ and negative at $y = y_{no}$. Hence at some point,

$$y_{Ham} \in B_{\mathcal{H}}(0, \frac{c^\sharp}{100}),$$

we have

$$\mathbf{d}(y_{Ham}, \Gamma_{yes}) = \mathbf{d}(y_{Ham}, \Gamma_{no}).$$

It follows that y_{Ham} has two distinct closest points in $\mathcal{M}$, and yet

$$\mathbf{d}(y_{Ham}, \mathcal{M}) \leq \frac{c^\sharp}{100}$$

since $0 \in \mathcal{M}$ and $y_{Ham} \in B_{\mathcal{H}}(0, \frac{c^\sharp}{100})$. That contradicts our assumption that $\mathcal{M}$ has reach ≥ 1. Hence our assumption that Γ_{no} intersects $B_{\mathcal{H}}(0, \frac{c^\sharp}{100})$ must be false. Therefore, by definition of Γ_{no} we have

$$\mathcal{M} \cap B_{\mathcal{H}}(0, \frac{c^\sharp}{100}) \subset \Gamma \cap B_{\mathcal{H}}(0, \frac{c^\sharp}{100}).$$

Since also

$$\Gamma \cap B_{\mathcal{H}}(0, c^\sharp) \subset \mathcal{M} \cap B_{\mathcal{H}}(0, c^\sharp),$$

it follows that

$$\Gamma \cap B_{\mathcal{H}}(0, \frac{c^\sharp}{100}) = \mathcal{M} \cap B_{\mathcal{H}}(0, \frac{c^\sharp}{100}),$$

proving the lemma. □

This completes the proof of Claim 1.

Appendix B. Proof of Lemma 5

Definition 24 (Rademacher complexity)**.** *Given a class $\mathcal{F}$ of functions $f : X \to \mathbb{R}$ a measure μ supported on X, a natural number $n \in \mathbb{N}$, and an n-tuple of points $(x_1, \dots, x_n)$, where each $x_i \in X$, we define the empirical Rademacher complexity $R_n(\mathcal{F}, x)$ as follows. Let $\sigma = (\sigma_1, \dots, \sigma_n)$ be a vector of n independent Rademacher (i.e., unbiased $\{-1, 1\}$-valued) random variables. Then,*

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

Proof. We will use Rademacher complexities to bound the sample complexity from above. We know (see Theorem 3.2 [5]) that for all $\delta > 0$

$$\mathbb{P}\left[\sup_{f \in \mathcal{F}} \left| \mathbb{E}_\mu f - \mathbb{E}_{\mu_s} f \right| \le 2R_s(\mathcal{F}, x) + \sqrt{\frac{2\log(2/\delta)}{s}} \right] \ge 1 - \delta. \tag{148}$$

Using a "chaining argument" the following claim is proved.

Claim 6.

$$R_s(\mathcal{F}, x) \le \epsilon + 12 \int_{\frac{\epsilon}{4}}^\infty \sqrt{\frac{\ln N(\eta, \mathcal{F}, \mathcal{L}_2(\mu_s))}{s}} d\eta. \tag{149}$$

When ϵ is taken to equal 0, the above is known as Dudley's entropy integral [21].

A result of Rudelson and Vershynin (Theorem 6.1, page 35 [48]) tells us that the integral in (149) can be bounded from above using an integral involving the square root of the fat shattering dimension (or in their terminology, combinatorial dimension). The precise relation that they prove is

$$\int_\epsilon^\infty \sqrt{\ln N(\eta, \mathcal{F}, \mathcal{L}_2(\mu_s))} d\eta \le C \int_\epsilon^\infty \sqrt{\mathrm{fat}_{c\eta}(\mathcal{F})} d\eta, \tag{150}$$

for universal constants c and C.

From Equations (148), (149), and (150), we see that if

$$s \ge \frac{C}{\epsilon^2} \left(\left(\int_{c\epsilon}^\infty \sqrt{\mathrm{fat}_\gamma(\mathcal{F})} d\gamma \right)^2 + \log 1/\delta \right),$$

then

$$\mathbb{P}\left[\sup_{f \in \mathcal{F}} \left| \mathbb{E}_{\mu_s} f(x_i) - \mathbb{E}_\mu f \right| \ge \epsilon \right] \le \delta. \quad \square$$

Appendix C. Proof of Claim 6

We begin by stating the finite class lemma of Massart ([37], Lemma 5.2).

Lemma 29. *Let X be a finite subset of $B(0,r) \subseteq \mathbb{R}^n$, and let $\sigma_1, \dots, \sigma_n$ be i.i.d. unbiased $\{-1,1\}$ random variables. Then, we have*

$$\mathbb{E}_\sigma \left[\sup_{x \in X} \frac{1}{n} \sum_{i=1}^n \sigma_i x_i \right] \leq \frac{r\sqrt{2 \ln |X|}}{n}.$$

We now move on to prove Claim 6. This claim is closely related to Dudley's integral formula, but appears to have been stated for the first time by Sridharan-Srebro [51]. We have furnished a proof following Sridharan-Srebro [51]. For a function class $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ and points $x_1, \dots, x_s \in \mathcal{X}$,

$$R_s(\mathcal{F}, x) \leq \epsilon + 12 \int_{\frac{\epsilon}{4}}^{\infty} \sqrt{\frac{\ln N(\eta, \mathcal{F}, \mathcal{L}_2(\mu_s))}{s}} d\eta. \tag{151}$$

Proof. Without loss of generality, we assume that $0 \in \mathcal{F}$; if not, we choose some function $f \in \mathcal{F}$ and translate $\mathcal{F}$ by $-f$. Let $M = \sup_{f \in \mathcal{F}} \|f\|_{L_2(P_n)}$, which we assume is finite. For $i \geq 1$, choose $\alpha_i = M2^{-i}$, and let T_i be a α_i-net of $\mathcal{F}$ with respect to the metric derived from $L_2(\mu_s)$. Here μ_s is the probability measure that is uniformly distributed on the s points $x_1, \dots, x_s$. For each $f \in \mathcal{F}$, and i, pick an $\hat{f}_i \in T_i$ such that f_i is an α_i-approximation of f, i.e., $\|f - f_i\|_{L_2(\mu_s)} \leq \alpha_i$. We use chaining to write

$$f = f - \hat{f}_N + \sum_{j=1}^N (\hat{f}_j - \hat{f}_{j-1}), \tag{152}$$

where $\hat{f}_0 = 0$. Now, choose N such that $\frac{\epsilon}{2} \leq M2^{-N} < \epsilon$,

(153)

$$\hat{R}_s(\mathcal{F}) = \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{s} \sum_{i=1}^s \sigma_i \left(f(x_i) - \hat{f}_N(x_i) + \sum_{j=1}^N (\hat{f}_j(x_i) - \hat{f}_{j-1}(x_i)) \right) \right]$$

$$\leq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{s} \sum_{i=1}^s \sigma_i (f(x_i) - \hat{f}_N(x_i)) \right] + \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{s} \sum_{i=1}^s \sigma_i (\hat{f}_j(x_i) - \hat{f}_{j-1}(x_i)) \right] \tag{154}$$

$$\leq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \langle \sigma, f - \hat{f}_N \rangle_{L_2(\mu_s)}) \right] + \sum_{j=1}^N \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{s} \sum_{i=1}^s \sigma_i (\hat{f}_j(x_i) - \hat{f}_{j-1}(x_i)) \right]. \tag{155}$$

We use Cauchy-Schwartz on the first term to give

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} \langle \sigma, f - \hat{f}_N \rangle_{L_2(\mu_s)}) \right] \leq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \|\sigma\|_{L_2(\mu_s)} \|f - \hat{f}_N\|_{L_2(\mu_s)}) \right] \tag{156}$$

$$\leq \epsilon. \tag{157}$$

Note that

$$\|\hat{f}_j - \hat{f}_{j-1}\|_{L_2(\mu_s)} \leq \|\hat{f}_j - f - (\hat{f}_{j-1} - f)\|_{L_2(\mu_s)} \leq \alpha_j + \alpha_{j-1} \tag{158}$$

$$\leq 3\alpha_j. \tag{159}$$

We use Massart's lemma to bound the second term,

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\frac{1}{s}\sum_{i=1}^{s}\sigma_i(\hat{f}_j(x_i)-\hat{f}_{j-1}(x_i))\right] = \mathbb{E}\left[\sup_{f\in\mathcal{F}}\langle\sigma,(\hat{f}_j-\hat{f}_{j-1})\rangle_{L_2(\mu_s)}\right] \tag{160}$$

$$\leq \frac{3\alpha_j\sqrt{2\ln(|T_j|\cdot|T_{j-1}|)}}{s} \tag{161}$$

$$\leq \frac{6\alpha_j\sqrt{\ln(|T_j|)}}{s}. \tag{162}$$

Now, from Equations (155), (157), and (162),

$$\hat{R}_s(\mathcal{F}) \leq \epsilon + 6\sum_{j=1}^{N}\alpha_j\sqrt{\frac{\ln N(\alpha_j,\mathcal{F},L_2(\mu))}{s}} \tag{163}$$

$$\leq \epsilon + 12\sum_{j=1}^{N}(\alpha_j-\alpha_{j+1})\sqrt{\frac{\ln N(\alpha_j,\mathcal{F},L_2(\mu_s))}{s}} \tag{164}$$

$$\leq \epsilon + 12\int_{\alpha_{N+1}}^{\alpha_0}\sqrt{\frac{\ln N(\alpha,\mathcal{F},L_2(\mu_s))}{s}}d\alpha \tag{165}$$

$$\leq \epsilon + 12\int_{\frac{\epsilon}{4}}^{\infty}\sqrt{\frac{\ln N(\alpha,\mathcal{F},L_2(\mu_s))}{s}}d\alpha. \quad \square \tag{166}$$

Appendix D. Proof of Lemma 6

Proof. We proceed to obtain an upper bound on the fat shattering dimension $\mathrm{fat}_\gamma(\mathcal{F}_{k,\ell})$. Let $x_1,\dots,x_s$ be s points such that

$$\forall\mathcal{A}\subseteq X := \{x_1,\dots,x_s\},$$

and there exists $V := \{v_{11},\dots,v_{k\ell}\}\subseteq B$ and $f\in\mathcal{F}_{k,\ell}$ where $f(x)=\max_j\min_i v_{ij}\cdot x$ such that for some $\mathbf{t}=(t_1,\dots,t_s)$, for all

$$x_r\in\mathcal{A},\forall\, j\in[\ell],\ \text{there exists}\ i\in[k]\quad v_{ij}\cdot x_r < t_r-\gamma \tag{167}$$

and

$$\forall x_r\notin\mathcal{A},\exists\, j\in[\ell],\forall i\in[k]\quad v_{ij}\cdot x_r > t_r+\gamma. \tag{168}$$

We will obtain an upper bound on s. Let $g := C_1\left(\gamma^{-2}\log(s+k\ell)\right)$ for a sufficiently large universal constant C_1. Consider a particular $\mathcal{A}\in X$ and $f(x) := \max_j\min_i v_{ij}\cdot x$ that satisfies (167) and (168).

Let R be an orthogonal projection onto a uniformly random g dimensional subspace of $span(X\cup V)$; we denote the family of all such linear maps $\Re$. Let RX denote the set $\{Rx_1,\dots,Rx_s\}$ and, likewise, RV denote the set $\{Rv_{11},\dots,Rv_{kl}\}$. Since all vectors in $X\cup V$ belong to the unit ball $B_{\mathcal{H}}$, by the Johnson-Lindenstrauss lemma, with probability greater than $1/2$, the inner product of every pair of vectors in $RX\cup RV$ multiplied by $\frac{m}{g}$ is within γ of the inner product of the corresponding vectors in $X\cup V$.

Therefore, we have the following.

Observation 9. *With probability at least $\frac{1}{2}$ the following statements are true:*

$$\forall x_r \in \mathcal{A}, \forall j \in [\ell], \exists i \in [k] \quad \left(\frac{m}{g}\right) Rv_{ij} \cdot Rx_r < t_r \tag{169}$$

and

$$\forall x_r \notin \mathcal{A}, \exists j \in [\ell], \forall i \in [k] \quad \left(\frac{m}{g}\right) Rv_{ij} \cdot Rx_r > t_r. \tag{170}$$

Let $R \in \Re$ be a projection onto a uniformly random g dimensional subspace in *span*$(X \cup V)$. Let $J := \text{span}(RX)$, and let $t^J : J \to \mathbb{R}$ be the function given by

$$t^J(y) := \begin{cases} t_i, & \text{if } y = Rx_i \text{ for some } i \in [s]; \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathcal{F}_{J,k,\ell}$ be the concept class consisting of all subsets of J of the form

$$\left\{ z : \max_j \min_i \begin{pmatrix} w_{ij} \\ 1 \end{pmatrix} \cdot \begin{pmatrix} z \\ -t^J(z) \end{pmatrix} \leq 0 \right\},$$

where $w_{11}, \dots, w_{k\ell}$ are arbitrary vectors in J.

Claim 7. *Let $y_1, \dots, y_s \in J$. Then, the number $W(s, \mathcal{F}_{J,k,\ell})$ of distinct sets $\{y_1, \dots, y_s\} \cap \imath$, $\imath \in \mathcal{F}_{J,k,\ell}$ is less than or equal to $s^{O((g+2)k\ell)}$.*

Proof of Claim 7. Classical VC theory (Lemma 3) tells us that the VC dimension of half-spaces in the span of all vectors of the form $(z; -t^J(z))$ is at most $g + 2$. Therefore, by the Sauer-Shelah lemma (Lemma 4), the number $W(s, \mathcal{F}_{J,1,1})$ of distinct sets $\{y_1, \dots, y_s\} \cap \jmath$, $\jmath \in \mathcal{F}_{J,1,1}$ is less than or equal to $\sum_{i=0}^{g+2} \binom{s}{i}$, which is less than or equal to s^{g+2}. Every set of the form $\{y_1, \dots, y_s\} \cap \imath$, $\imath \in \mathcal{F}_{J,k,\ell}$ can be expressed as an intersection of a union of sets of the form $\{y_1, \dots, y_s\} \cap \jmath$, $\jmath \in \mathcal{F}_{J,1,1}$, in which the total number of sets participating is $k\ell$. Therefore, the number $W(s, \mathcal{F}_{J,k,\ell})$ of distinct sets $\{y_1, \dots, y_s\} \cap \imath$, $\imath \in \mathcal{F}_{J,1,1}$ is less than or equal to $W(s, \mathcal{F}_{J,1,1})^{k\ell}$, which is in turn less than or equal to $s^{(g+2)k\ell}$. □

By Observation 9, for a random $R \in \Re$, the expected number of sets of the form $RX \cap \imath$, $\imath \in \mathcal{F}_{J,k,\ell}$ is greater than or equal to 2^{s-1}. Therefore, there exists an $R \in \Re$ such that the number of sets of the form $RX \cap \imath$, $\imath \in \mathcal{F}_{J,k,\ell}$ is greater than or equal to 2^{s-1}. Fix such an R and set $J := \text{span}(RX)$. By Claim 7,

$$2^{s-1} \leq s^{k\ell(g+2)}. \tag{171}$$

Therefore $s - 1 \leq k\ell(g+2)\log s$. Assuming without loss of generality that $s \geq k\ell$, and substituting $C_1\left(\gamma^{-2}\log(s + k\ell)\right)$ for g, we see that

$$s \leq O\left(k\ell\gamma^{-2}\log^2 s\right),$$

and hence

$$\frac{s}{\log^2(s)} \leq O\left(\frac{k\ell}{\gamma^2}\right),$$

implying that

$$s \leq O\left(\left(\frac{k\ell}{\gamma^2}\right)\log^2\left(\frac{k\ell}{\gamma}\right)\right).$$

Thus, the fat shattering dimension $\mathrm{fat}_\gamma(\mathcal{F}_{k,\ell})$ is $O\left(\left(\frac{k\ell}{\gamma^2}\right)\log^2\left(\frac{k\ell}{\gamma}\right)\right)$. We independently know that $\mathrm{fat}_\gamma(\mathcal{F}_{k,\ell})$ is 0 for $\gamma > 2$.

Therefore by Lemma 5, if

$$s \geq \frac{C}{\epsilon^2}\left(\left(\int_{c\epsilon}^{2}\frac{\sqrt{k\ell\log^2(k\ell/\gamma^2)}}{\gamma}d\gamma\right)^2 + \log 1/\delta\right), \tag{172}$$

then

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}_{\mu_s}f(x_i) - \mathbb{E}_\mu f\right| \geq \epsilon\right] \leq \delta.$$

Let $t = \ln\left(\frac{\sqrt{k\ell}}{\gamma}\right)$. Then the integral in (172) equals

$$\sqrt{k\ell}\int_{\ln(Ck\ell/\epsilon^2)}^{\ln(\sqrt{kl}/2)} -t dt < C\sqrt{kl}\left(\ln(Ck\ell/\epsilon^2)\right)^2,$$

and so if

$$s \geq \frac{C}{\epsilon^2}\left(k\ell\ln^4\left(k\ell/\epsilon^2\right) + \log 1/\delta\right),$$

then

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}}\left|\mathbb{E}_{\mu_s}f(x_i) - \mathbb{E}_\mu f\right| \geq \epsilon\right] \leq 1 - \delta. \quad \square$$

Acknowledgments

The authors are grateful to Alexander Rakhlin and Ambuj Tiwari for valuable feedback and for directing them to material related to Dudley's entropy integral. The authors thank Stephen Boyd, Thomas Duchamp, Sanjeev Kulkarni, Marina Miela, Thomas Richardson, Werner Stuetzle, and Martin Wainwright for valuable discussions.

References

[1] Shun-ichi Amari and Hiroshi Nagaoka, *Methods of information geometry*, Translations of Mathematical Monographs, vol. 191, American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000. Translated from the 1993 Japanese original by Daishi Harada. MR1800071 (2001j:62023)

[2] Richard G. Baraniuk and Michael B. Wakin, *Random projections of smooth manifolds*, Found. Comput. Math. **9** (2009), no. 1, 51–77, DOI 10.1007/s10208-007-9011-z. MR2472287 (2010g:94021)

[3] Mikhail Belkin and Partha Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Comput. **15** (2003), no. 6, 1373–1396.

[4] Jean-Daniel Boissonnat, Leonidas J. Guibas, and Steve Y. Oudot, *Manifold reconstruction in arbitrary dimensions using witness complexes*, Discrete Comput. Geom. **42** (2009), no. 1, 37–70, DOI 10.1007/s00454-009-9175-1. MR2506737 (2010h:68183)

[5] Olivier Bousquet, Stéphane Boucheron, and Gábor Lugosi, *Introduction to statistical learning theory*, Advanced Lectures on Machine Learning, 2004, pp. 169–207.

[6] P. S. Bradley and O. L. Mangasarian, *k-plane clustering*, J. Global Optim. **16** (2000), no. 1, 23–32, DOI 10.1023/A:1008324625522. MR1770524 (2001g:90095)

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

[8] Guillermo D. Cañas, Tomaso Poggio, and Lorenzo Rosasco, *Learning manifolds with k-means and k-flats*, 26th Annual Conference on Neural Information Processing Systems, December 3–6, 2012, Lake Tahoe, Nevada, Proceedings of Advances in Neural Information Processing Systems 25, Curran Associates, Red Hook, NY, 2013, pp. 2474–2482.

[9] Gunnar Carlsson, *Topology and data*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 2, 255–308, DOI 10.1090/S0273-0979-09-01249-X. MR2476414 (2010d:55001)

[10] Siu-Wing Cheng, Tamal K. Dey, and Edgar A. Ramos, *Manifold reconstruction from point samples*, Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, 2005, pp. 1018–1027. MR2298361

[11] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman, *Manifold estimation and singular deconvolution under Hausdorff loss*, Ann. Statist. **40** (2012), no. 2, 941–963, DOI 10.1214/12-AOS994. MR2985939

[12] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman, *Minimax manifold estimation*, J. Mach. Learn. Res. **13** (2012), no. 1, 1263–1291.

[13] Kenneth L. Clarkson, *Tighter bounds for random projections of manifolds*, Computational geometry (SCG'08), ACM, New York, 2008, pp. 39–48, DOI 10.1145/1377676.1377685. MR2504269

[14] Ronald R. Coifman, Stephane Lafon, Ann Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part II: Multiscale methods*, Proc. Natl. Acad. Sci. USA **102** (2005), no. 21, 7432–7438.

[15] Trevor F. Cox and Michael A. A. Cox, *Multidimensional scaling*, Monographs on Statistics and Applied Probability, vol. 59, Chapman & Hall, London, 1994. With 1 IBM-PC floppy disk (3.5 inch, HD). MR1335449 (96g:62118)

[16] Sanjoy Dasgupta and Yoav Freund, *Random projection trees and low dimensional manifolds*, STOC'08, ACM, New York, 2008, pp. 537–546, DOI 10.1145/1374376.1374452. MR2582911 (2011b:68036)

[17] Guy David and Stephen Semmes, *Analysis of and on uniformly rectifiable sets*, Mathematical Surveys and Monographs, vol. 38, American Mathematical Society, Providence, RI, 1993. MR1251061 (94i:28003)

[18] Luc Devroye, Lászlo Györfi, and Gábor Lugosi, *A probabilistic theory of pattern recognition*, Springer, New York, 1996.

[19] David L. Donoho, *Aide-memoire. High-dimensional data analysis: The curses and blessings of dimensionality*, http://statweb.stanford.edu/∼donoho/Lectures/CBMS/Curses.pdf (2000).

[20] David L. Donoho and Carrie Grimes, *Hessian eigenmaps: locally linear embedding techniques for high-dimensional data*, Proc. Natl. Acad. Sci. USA **100** (2003), no. 10, 5591–5596 (electronic), DOI 10.1073/pnas.1031596100. MR1981019 (2004e:62108)

[21] R. M. Dudley, *Uniform central limit theorems*, Cambridge Studies in Advanced Mathematics, vol. 63, Cambridge University Press, Cambridge, 1999. MR1720712 (2000k:60050)

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

[23] Charles Fefferman, *Interpolation and extrapolation of smooth functions by linear operators*, Rev. Mat. Iberoam. **21** (2005), no. 1, 313–348, DOI 10.4171/RMI/424. MR2155023 (2006h:58009)

[24] Charles Fefferman, *The C^m norm of a function with prescribed jets. II*, Rev. Mat. Iberoam. **25** (2009), no. 1, 275–421, DOI 10.4171/RMI/570. MR2514339 (2011d:46050)

[25] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman, *Nonparametric ridge estimation*, Ann. Statist. **42** (2014), no. 4, 1511–1545, DOI 10.1214/14-AOS1218. MR3262459

[26] Trevor Hastie and Werner Stuetzle, *Principal curves*, J. Amer. Statist. Assoc. **84** (1989), no. 406, 502–516. MR1010339 (90i:62073)

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

[28] Mark A. Iwen and Mauro Maggioni, *Approximation of points on low-dimensional manifolds via random linear projections*, Inf. Inference **2** (2013), no. 1, 1–31, DOI 10.1093/imaiai/iat001. MR3311442

[29] William B. Johnson and Joram Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in modern analysis and probability (New Haven, Conn., 1982), Contemp. Math., vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 189–206, DOI 10.1090/conm/026/737400. MR737400 (86a:46018)

[30] Peter W. Jones, Mauro Maggioni, and Raanan Schul, *Universal local parametrizations via heat kernels and eigenfunctions of the Laplacian*, Ann. Acad. Sci. Fenn. Math. **35** (2010), no. 1, 131–174, DOI 10.5186/aasfm.2010.3508. MR2643401 (2011g:58038)

[31] A. Kambhatla and Todd K. Leen, *Fast non-linear dimension reduction*, IEEE International Conference on Neural Networks, IEEE, New York, 1993, pp. 1213–1218.

[32] Balázs Kégl, Adam Krzyzak, Tamás Linder, and Kenneth Zeger, *Learning and design of principal curves*, IEEE Trans. Pattern Analysis and Machine Intelligence **22** (2000), 281–297.

[33] Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR1849347 (2003k:28019)

[34] Gilad Lerman and Teng Zhang, *Robust recovery of multiple subspaces by geometric l_p minimization*, Ann. Statist. **39** (2011), no. 5, 2686–2715, DOI 10.1214/11-AOS914. MR2906883

[35] *Manifold learning theory and applications*, CRC Press, Boca Raton, FL, 2012. Edited by Yunqian Ma and Yun Fu. MR2964193

[36] Thomas Martinetz and Klaus Schulten, *Topology representing networks*, Neural Netw. **7** (1994), no. 3, 507–522.

[37] Pascal Massart, *Some applications of concentration inequalities to statistics* (English, with English and French summaries), Ann. Fac. Sci. Toulouse Math. (6) **9** (2000), no. 2, 245–303. Probability theory. MR1813803 (2001m:60012)

[38] Andreas Maurer and Massimiliano Pontil, *K-dimensional coding schemes in Hilbert spaces*, IEEE Trans. Inform. Theory **56** (2010), no. 11, 5839–5846, DOI 10.1109/TIT.2010.2069250. MR2808936 (2012a:94107)

[39] Raghavan Narasimhan, *Lectures on topics in analysis*, Notes by M. S. Rajwade. Tata Institute of Fundamental Research Lectures on Mathematics, No. 34, Tata Institute of Fundamental Research, Bombay, 1965. MR0212837 (35 #3702)

[40] Hariharan Narayanan and Sanjoy Mitter, *On the sample complexity of testing the manifold hypothesis*, NIPS (2010), 1786–1794.

[41] Hariharan Narayanan and Partha Niyogi, *On the sample complexity of learning smooth cuts on a manifol*, Proc. of the 22nd Annual Conference on Learning Theory (COLT), June 2009, pp. 197–204.

[42] Nigel J. Newton, *Infinite-dimensional manifolds of finite-entropy probability measures*, Geometric science of information, Lecture Notes in Comput. Sci., vol. 8085, Springer, Heidelberg, 2013, pp. 713–720, DOI 10.1007/978-3-642-40020-9_79. MR3126105

[43] Partha Niyogi, Stephen Smale, and Shmuel Weinberger, *Finding the homology of submanifolds with high confidence from random samples*, Discrete Comput. Geom. **39** (2008), no. 1-3, 419–441, DOI 10.1007/s00454-008-9053-2. MR2383768 (2009b:60038)

[44] Umut Ozertem and Deniz Erdogmus, *Locally defined principal curves and surfaces*, J. Mach. Learn. Res. **12** (2011), 1249–1286. MR2804600 (2012e:62196)

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

[46] S. T. Roweis, L. Saul, and G. Hinton, *Global coordination of local linear models*, Adv. Neural Inf. Process. Syst. **14** (2002), 889–896.

[47] Sam T. Roweis and Lawrence K. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, Science **290** (2000), 2323–2326.

[48] M. Rudelson and R. Vershynin, *Combinatorics of random processes and sections of convex bodies*, Ann. of Math. (2) **164** (2006), no. 2, 603–648, DOI 10.4007/annals.2006.164.603. MR2247969 (2007j:60016)

[49] J. Shawe-Taylor and N. Christianini, *Kernel methods for pattern analysis*, Cambridge University Press, Cambridge, UK, 2004.

[50] Alex J. Smola, Robert C. Williamson, Sebastian Mika, and Bernhard Schölkopf, *Regularized principal manifolds*, Computational learning theory (Nordkirchen, 1999), Lecture Notes in Comput. Sci., vol. 1572, Springer, Berlin, 1999, pp. 214–229, DOI 10.1007/3-540-49097-3_17. MR1724991

[51] Karthik Sridharan and Nathan Srebro, *Note on refined Dudley integral covering number bound*. Unpublished.

[52] J. B. Tenenbaum, V. Silva, and J. C. Langford, *A Global Geometric Framework for Nonlinear Dimensionality Reduction*, Science **290** (2000), no. 2000, 2319–2323.

[53] Pravin M. Vaidya, *A new algorithm for minimizing convex functions over convex sets*, Math. Programming **73** (1996), no. 3, Ser. A, 291–341, DOI 10.1016/0025-5610(92)00021-S. MR1393123 (97e:90060)

[54] Vladimir Vapnik, *Estimation of dependences based on empirical data*, Springer Series in Statistics, Springer-Verlag, New York-Berlin, 1982. Translated from the Russian by Samuel Kotz. MR672244 (84a:62043)

[55] Kilian Q. Weinberger and Lawrence K. Saul, *Unsupervised learning of image manifolds by semidefinite programming*, Int. J. Comput. Vision **70** (2006), no. 1, 77–90.

[56] Zhenyue Zhang and Hongyuan Zha, *Principal manifolds and nonlinear dimensionality reduction via tangent space alignment*, SIAM J. Sci. Comput. **26** (2004), no. 1, 313–338 (electronic), DOI 10.1137/S1064827502419154. MR2114346 (2005k:65024)

Department of Mathematics, Princeton University, Princeton, New Jersey 08544
E-mail address: cf@math.princeton.edu

Laboratory for Information and Decision Systems, MIT, Cambridge, Massachusetts 02139
E-mail address: mitter@mit.edu

Department of Statistics and Department of Mathematics, University of Washington, Seattle, Washington 98195
E-mail address: harin@uw.edu