

©Copyright 2019

Kitty Mohammed

Statistical Methods for Manifold Recovery and $C^{1,1}$ Regression on Manifolds

Kitty Mohammed

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Hariharan Narayanan, Chair

Marina Meilă

Yen-Chi Chen

Program Authorized to Offer Degree:
Department of Statistics

University of Washington

Abstract

Statistical Methods for Manifold Recovery and $C^{1,1}$ Regression on Manifolds

Kitty Mohammed

Chair of the Supervisory Committee:
Dr. Hariharan Narayanan
Department of Statistics

High-dimensional data sets often have lower-dimensional structure taking the form of a submanifold of a Euclidean space. It is challenging but necessary to develop statistical methods for these data sets that respect the manifold structure. We present research from two areas: manifold learning (i.e., support estimation) and smooth regression on manifolds.

First, we consider the problem of recovering a d -dimensional submanifold \mathcal{M} of \mathbb{R}^n when provided with N samples from \mathcal{M} . Ideally, the estimator of \mathcal{M} should be a manifold of a certain smoothness, and it should converge to \mathcal{M} in Hausdorff distance as N increases. Fefferman, Mitter, and Narayanan (2016) have developed an algorithm whose output is provably a manifold. The algorithm relies on the definition of an approximate squared-distance function (asdf) to \mathcal{M} . As long as the asdf meets certain regularity conditions (which can be difficult to verify), it can be used to define an estimator of \mathcal{M} that has the desired properties. We define two asdfs that can be calculated solely from the sample and show that they meet the required regularity conditions. The first asdf is based on kernel density estimation, and the second is based on the estimation of tangent spaces with local principal components analysis.

Second, we analyze a structural risk minimization-based algorithm for the regression of real-valued $C^{1,1}$ functions defined on a manifold. (These are differentiable functions whose derivative is Lipschitz.) We assume that we are provided with N sample points from \mathcal{M} , a

d -dimensional C^2 submanifold of \mathbb{R}^n , as well as noisy observations from a $C^{1,1}(\mathcal{M})$ function f^* . We first present results of independent interest on sampling a C^2 atlas of \mathcal{M} w.h.p. To do this, we 1) show that the sample contains, w.h.p., a fine-enough net of \mathcal{M} of bounded size, 2) derive uniform convergence rates for the empirical measure indexed by particular subsets of \mathcal{M} , and 3) prove that tangent spaces estimated with local PCA are close in angular distance to the true tangent spaces w.h.p. The estimated tangent spaces can be used to define charts of \mathcal{M} whose derivatives have a uniformly-bounded Lipschitz constant. After sampling an atlas, we use $C^{1,1}(\mathbb{R}^d)$ regression to locally estimate the pullbacks of f^* to the estimated tangent spaces and then use a partition of unity to define the global estimator of f^* . The $C^{1,1}(\mathbb{R}^d)$ regression algorithm was developed by our collaborators (Gustafson, Hirn, Mohammed, Narayanan, and Xu, 2018), and it is closely related to a $C^{1,1}(\mathbb{R}^d)$ interpolation algorithm of Herbert-Voss, Hirn, and McCollum (2017). We use tools from empirical processes to analyze the $C^{1,1}(\mathbb{R}^d)$ regression algorithm, and we combine these results with properties of the C^2 atlas to derive risk bounds and convergence rates for our $C^{1,1}(\mathcal{M})$ regression algorithm.

TABLE OF CONTENTS

	Page
Chapter 1: Introduction	1
1.1 Outline of the Dissertation	2
Chapter 2: Manifold Learning Using Kernel Density Estimation and Local Principal Components Analysis	5
2.1 Introduction	5
2.1.1 Related Work	6
2.2 Technical Background and Assumptions	7
2.2.1 Manifolds	7
2.2.2 Model	10
2.2.3 Approximate Squared-Distance Functions	10
2.2.4 Ridges and Gradient Descent	14
2.2.5 Empirical Processes	14
2.3 Kernel Density Estimation	15
2.3.1 Definition of the asdf	15
2.3.2 Selecting the Bandwidth σ	17
2.3.3 Bounding p_N in Expectation	17
2.3.4 Finite Sample Bounds for p_N and $\partial^\alpha(-\log p_N(x))$	24
2.3.5 $-\log p_N(x) + \log N_f$ is an asdf	31
2.4 Local Principal Components Analysis	35
2.4.1 Definition of the asdf and Selection of the Bandwidth $\bar{\tau}$	35
2.4.2 Cylinder Packets	36
2.4.3 Constructing an Admissible Cylinder Packet with Local PCA	39
2.4.4 $\{F^{\bar{o}}(z), C_p^{\widehat{Tan}}\}$ is an asdf	44
2.5 Simulations	45
2.6 Discussion	46

Chapter 3: $C^{1,1}(\mathcal{M})$ Regression Via Estimation of Charts and Local $C^{1,1}(\mathbb{R}^d)$ Regression	49
3.1 Introduction	49
3.1.1 Notation	50
3.2 $C^{1,1}(\mathbb{R}^d)$ Regression	52
3.2.1 Noiseless $C^{1,1}(\mathbb{R}^d)$ Interpolation	52
3.2.2 Estimating a $C^{1,1}(\mathbb{R}^d)$ Function from a Noisy Sample	56
3.2.3 Sample Complexity of $C^{1,1}(\mathbb{R}^d)$ Regression	59
3.3 Geometric Framework	69
3.3.1 Metric and Connection	70
3.3.2 Volume	73
3.3.3 Reach	73
3.3.4 Charts	75
3.4 Norm of $C^{1,1}(\mathcal{M})$ Functions	84
3.4.1 Pullback to Tangent Spaces	89
3.5 Tangent Space Estimation for Charts	91
3.5.1 Pullback to Estimated Tangent Spaces	103
3.6 Analysis of $C^{1,1}(\mathcal{M})$ Regression Algorithm	104
3.6.1 Sampling a C^2 Atlas for \mathcal{M}	106
3.6.2 Excess Risk and Convergence of \hat{f}	113
3.7 Discussion	118
Bibliography	120

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Hariharan Narayanan. I have learned so much from him over the course of the last few years. Apart from the actual statistical knowledge I've gained from him, he taught me how to select original research questions to work on and how to see the light at the end of the tunnel when your research is really abstract. The reason I was able to complete this dissertation is due to him being such a patient mentor, including weekly calls after he moved to TIFR to let me pick his brain. Even more than his academic mentorship, I would like to thank him for being a really kind and caring advisor. I had some personal things going on during my time in grad school, and he was very understanding and let me work at my own pace when I needed to. I would also like to thank Marina Meilă and Yen-Chi Chen for being on my committee, for giving me helpful ideas and comments about my research, and for being really inspiring and accomplished people. Finally, I couldn't have finished grad school without the people who care about me the most. Thank you to my family—my mom, my sister Maryam, my dad, and my brother. And thank you to my best friends, especially Mimi, Rozie, Navi, Nathan, and Jason. You've all taken 3 AM calls from me that neither of us remember ♡

DEDICATION

♡ To my mommy ♡

Chapter 1

INTRODUCTION

Often, high-dimensional data sets have lower-dimensional structure taking the form of a manifold. Estimating the structure of such a manifold is desirable because it effectively reduces the dimensionality of the data set. Of course, the end goal may be to perform some other statistical analysis, such as regression or classification, with this data set. In this case, it is necessary to develop statistical procedures that take the structure of the manifold into account, either implicitly or by learning it in a separate, pre-processing step.

In this dissertation, we are interested in two major questions:

- (Q1). Given N samples from \mathcal{M} , a d -dimensional C^2 submanifold of \mathbb{R}^n , devise a concrete procedure that estimates \mathcal{M} with a sufficiently-smooth submanifold \mathcal{M}_{put} that converges to \mathcal{M} in Hausdorff distance as N increases.
- (Q2). Assume that we have N samples from \mathcal{M} , a d -dimensional C^2 submanifold of \mathbb{R}^n , as well as corresponding (noisy) samples from a $C^{1,1}(\mathcal{M})$ function f^* . (A $C^{1,1}(\mathcal{M})$ function is a real-valued function defined on \mathcal{M} that is differentiable and has a Lipschitz Riemannian gradient.) Construct an algorithm that recovers f^* in risk and sup norm for increasing N .

(Q1) is a support estimation problem, and (Q2) is a smooth regression problem, with the covariates lying on a manifold. We choose to include both of these problems in this dissertation because our solutions demonstrate the interplay between support estimation and additional statistical analyses in the manifold setting. For example, in both cases we locally approximate the manifold using a particular class of charts that have a uniformly bounded Lipschitz constant. (This class of charts is useful for constructing various uniform bounds

and is discussed in Section 2.2.1 and, in more detail, in Section 3.3.4.) This construction can be used directly for support approximation, as a preliminary step for more sophisticated support estimation procedures, or as a preliminary step for tasks where the manifold is the domain. In general, we aim to provide concrete procedures that are also theoretically justified. Our solutions for both (Q1) and (Q2) produce estimators that can be directly calculated from the data set; we then use methods from differential geometry and empirical processes to analyze convergence and other desired theoretical properties.

1.1 *Outline of the Dissertation*

In Chapter 2, we present two solutions to (Q1). They use key results from the work of Fefferman et al. (2016), who introduce the concept of the approximate squared-distance function (asdf) and use it to define an estimating manifold. An asdf is a function that meets three specific regularity conditions relating to smoothness and curvature. If data is drawn from \mathcal{M} , a close estimate of \mathcal{M} (in Hausdorff distance) can be defined that is locally a sufficiently-smooth graph given implicitly in terms of the gradient and Hessian of the asdf. In order for this procedure to produce a concrete estimate of \mathcal{M} , it is required for the asdf to be a function only of the data. We define two such asdfs using kernel density estimation and local PCA and show that they meet the necessary conditions.

Section 2.2 contains the technical background required to read our main results and proofs. This section starts off with background material on submanifolds, including key definitions, conventions regarding coordinates and projection operators, and important geometric results. Section 2.2.2 contains the model assumptions. Section 2.2.3 summarizes the major theorems we use from Fefferman et al. (2016). In Section 2.2.4, we summarize the algorithm from Ozertem and Erdogmus (2011) that we use to actually compute the putative manifold. Section 2.2.5 lists a few key concepts from empirical process theory. We include these because a few of our proofs are simpler when we work in the continuous setting and then argue that a similar result holds for a finite sample from the manifold. Sections 2.3 and 2.4 contain the main results of this chapter. We provide the precise definition of the asdfs

and prove that they do indeed meet the conditions required to apply the theorems contained in Section 2.2.3.

We present a solution to (Q2) in Chapter 3 that is based on sampling a C^2 atlas of \mathcal{M} and using a partition of unity to patch together local estimates of the pullbacks of f^* to the estimated tangent spaces. The local estimates are produced by a $C^{1,1}(\mathbb{R}^d)$ regression algorithm created by our collaborators (Gustafson et al., 2018). Our contribution was to analyze the sample complexity of the algorithm. We include these results in Section 3.2.

In Section 3.3, we present most of the background material on manifolds required to understand this chapter, including definitions of the reach and the Hausdorff measure. In Section 3.3.4, we discuss a particular class of local parametrizations of submanifolds, defined as the preimages of orthogonal projections onto the tangent spaces, performed on regions whose diameters are sufficiently small. In Lemma 35 we show that this class consists of C^2 diffeomorphisms whose derivatives have operator norms and Lipschitz constants that are uniformly bounded above. This is important to know because our algorithm requires the projection of sample points onto estimated tangent spaces. Assuming for now that we know the actual tangent spaces (a requirement that will be removed later), the functions that must be analyzed are the compositions of f^* with a local parametrization; thus, we need analytic properties of these parametrizations. Although the results in this lemma are mostly known, we include the proof because we derive slightly better constants and because our proof technique is more easily followed by those familiar with the statistical literature.

In Section 3.4, we define a norm $\|\cdot\|_{C^{1,1}(\mathcal{M})}$ on the class $C^{1,1}(\mathcal{M})$, which is a generalization to nonlinear manifolds of the norm on the class $C^{1,1}(\mathbb{R}^d)$ that was defined in Section 3.2. We show in Lemma 40 that if $\|f^*\|_{C^{1,1}(\mathcal{M})} = M^* < \infty$, the $C^{1,1}(\mathbb{R}^d)$ norm of its composition with a local parametrization from Section 3.3.4 is finite, which is necessary to apply the $C^{1,1}(\mathbb{R}^d)$ regression algorithm as an intermediate step. We also show that these $C^{1,1}(\mathbb{R}^d)$ norms are uniformly upper bounded by a function of M^* , which allows us to apply sample complexity results for $C^{1,1}(\mathbb{R}^d)$ regression.

Of course, since we only sample points from \mathcal{M} and not tangent spaces, we need to

estimate the latter, which we do using local PCA. In Section 3.5, we analyze the class of local parametrizations defined as preimages of orthogonal projections onto the estimated tangent spaces. If local PCA is performed in a sufficiently small region and if the sample is large enough, the principal angles between an estimated tangent space and the true tangent space are very small with high probability. This can be shown using matrix perturbation and concentration theory. We use this result to prove Theorems 44 and 46, which are analogues of Lemmas 35 and 40, respectively.

In Section 3.6, we show that our algorithm works to solve the desired minimization problem, and we also give risk bounds. We do so by proving that w.h.p. the sample contains a net with certain desirable properties and then combine this with the results of Section 3.5 and the $C^{1,1}(\mathbb{R}^d)$ regression results.

Chapter 2

MANIFOLD LEARNING USING KERNEL DENSITY ESTIMATION AND LOCAL PRINCIPAL COMPONENTS ANALYSIS

2.1 Introduction

Manifold learning consists of algorithms that take a high-dimensional data set as input and output a fit of a manifold structure. Many of these algorithms (such as Isomap, Laplacian eigenmaps, locally linear embedding, etc.) are used in practice and have a theoretical literature supporting them. [Ma and Fu \(2011\)](#) give a concise overview of these methods.

A drawback of most manifold learning algorithms is that if we are given data from a manifold, their output is not an actual manifold that is close to the original manifold. [Fefferman, Mitter, and Narayanan \(2016\)](#) develop an algorithm whose output is provably a manifold of certain smoothness. They start by defining an approximate squared-distance function (asdf) from the data in a manner that uses exhaustive search, utilizing the data only indirectly. Thus, a very large number of potential asdfs are examined before an approximately optimal one is chosen. In this chapter, we do away with the exhaustive search, albeit in the specific case of noiseless data that is sampled uniformly from a manifold. [Fefferman et al. \(2016\)](#) prove a key theorem that states that as long as we are able to define an asdf meeting certain general conditions, their algorithm outputs a set that is a manifold with bounded smoothness and Hausdorff distance to the original manifold. We demonstrate two different methods of estimating the true manifold via asdfs that can be calculated from the data. The two asdfs in our work are based on 1) kernel density estimation, and 2) approximating the manifold using tangent planes which are in turn approximated with local principal components analysis (PCA).

Ozertem and Erdogmus (2011) learn manifolds by forming a kernel density estimator (KDE) from the data points and finding its d -dimensional ridges. We give a more precise definition later, but a ridge is essentially a higher-dimensional analog of the mode and is related to the output set from the algorithm of Fefferman et al. (2016). Ozertem and Erdogmus (2011) give a practical method for finding the ridges through a variant of gradient descent where the descent is constrained to the subspace spanned by the largest eigenvectors of the Hessian of the KDE. We state their algorithm in Section 2.5 and use it to produce simulation results. Although they only apply subspace-constrained gradient descent to find ridges of the KDE, the method is more general and can be used to find ridges of both of our asdfs.

2.1.1 Related Work

Manifold learning has existed as an area of statistics and machine learning since the early 2000s. Some classical manifold learning algorithms are Isomap (Tenenbaum, De Silva, and Langford, 2000), locally linear embedding (Roweis and Saul, 2000), and Laplacian eigenmaps (Belkin and Niyogi, 2003). Many of these early algorithms rely on spectral graph theory and start off by constructing a graph which is then used to produce a lower-dimensional embedding of the data set. The theoretical guarantees are centered around proving that asymptotically, certain values such as the geodesic distance can be approximated to arbitrary precision.

More recently, there have been quite a few papers combining ridge estimation with manifold learning (including the work of Ozertem and Erdogmus, 2011). Some early results on ridge estimation are due to Eberly (1996), Hall, Qian, and Titterington (1992), and Cheng, Hall, Hartigan, et al. (2004). Ridge sets can be constructed to estimate a probability density or an embedded submanifold. Theoretical guarantees in this setting have been given by Genovese, Perone-Pacifco, Verdinelli, Wasserman, et al. (2012b), Genovese, Perone-Pacifco, Verdinelli, and Wasserman (2012a), Genovese, Perone-Pacifco, Verdinelli, Wasserman, et al. (2014), and Chen, Genovese, Wasserman, et al. (2015). Of these, the most relevant results for

us are from [Genovese et al. \(2014\)](#). They prove that as the sample size goes to infinity, their ridge set gets arbitrarily close to an underlying manifold in Hausdorff distance. [Fefferman et al. \(2016\)](#) also define a procedure related to ridge estimation methods that can be used to estimate an underlying manifold. For our purposes, the major advances of their work are twofold. First, their method is general; as long as a function meets a few conditions, it can be used to define an estimator that can be made arbitrarily close to an underlying manifold in Hausdorff distance. Furthermore, they show that this estimator is itself a manifold with bounded reach (which measures how rough a submanifold can be). Second, their proofs rely on using the implicit function theorem concretely, allowing them to make quantitative statements about the bounds of interest.

2.2 Technical Background and Assumptions

We now provide the definitions and major theorems that we rely on in the rest of this chapter. The results we use the most often are Theorems [1](#), [2](#), [5](#), and [6](#); the rest of this section can be referred to as necessary.

2.2.1 Manifolds

This subsection is adapted from [Fefferman et al. \(2016\)](#). In this chapter, we use the terms manifold and submanifold interchangeably with compact imbedded d -manifold. A closed subset $\mathcal{M} \subset \mathbb{R}^n$ is a *compact imbedded d -manifold* if the following conditions hold. First, \mathcal{M} is compact. Next, there exists $r_1 > r_2 > 0$ such that for every $z \in \mathcal{M}$ there exists a d -dimensional subspace $T_z\mathcal{M}$ of \mathbb{R}^n such that $\mathcal{M} \cap B(z, r_2) = \Gamma \cap B(z, r_2)$ for a patch Γ over $T_z\mathcal{M}$ of radius r_1 , centered at z and tangent to $T_z\mathcal{M}$ at z . A *patch* of radius r over $T_z\mathcal{M}$ is a subset $\Gamma := \{x + \Psi(x) \mid x \in B_d(r) \subset T_z\mathcal{M}\}$ of \mathbb{R}^n where $\Psi(x) : B_d(r) \rightarrow T_z^\perp\mathcal{M}$ is a C^2 -function that is zero at the origin.

The tangent space can be defined in the usual way (corresponding to $T_z\mathcal{M}$) or by using the following definition which applies to arbitrary closed sets $A \subset \mathbb{R}^n$. At a point $a \in A$, $\text{Tan}^0(a, A)$ is the set of vectors v such that for all $\epsilon > 0$, there exists $b \in A$ such that

$0 < |a - b| < \epsilon$ and $\left|v/|v| - \frac{b-a}{|b-a|}\right| < \epsilon$. Let the *tangent space* $\text{Tan}(a, A)$ be the set of all x such that $x - a \in \text{Tan}^0(a, A)$.

The geometric quantities of a submanifold \mathcal{M} that we are most concerned with are the d -dimensional volume V and the *reach* τ . The reach is the largest number such that all points within τ of \mathcal{M} have a unique closest point on \mathcal{M} . Intuitively, the reach governs how “rough” an embedded submanifold is. For example, the reach of a line with a sharp cusp is zero, and the reach of a linear subspace is infinite.

The following theorem due to [Federer \(1959\)](#) is useful for bounding the distance from a point on a manifold to the tangent space at a nearby point.

Theorem 1 (Federer’s reach condition). *Let \mathcal{M} be an embedded submanifold of \mathbb{R}^n . Then*

$$\text{reach}(\mathcal{M})^{-1} = \sup\{2\|b - a\|^{-2}\|b - \Pi_a b\| \mid a, b \in \mathcal{M}, a \neq b\}.$$

In this chapter, we assume regularity conditions on the manifold we draw samples from. We assume it is in \mathcal{G} , where $\mathcal{G} = \mathcal{G}(d, V_{max}, \tau_{min})$ is the family of boundaryless C^2 -submanifolds of the unit ball of \mathbb{R}^n with dimension d , volume less than or equal to V_{max} , and reach at least τ_{min} .

Let the *tubular neighborhood* $\mathcal{M}_{\bar{\tau}}$ be the set of all points within a distance of $\bar{\tau}$ of \mathcal{M} . Now, for points $z \in \mathcal{M}$ and $y \in \mathcal{M}_{\bar{\tau}}$, denote the projection onto the tangent plane at z by

$$\Pi_z : \mathbb{R}^n \rightarrow T_z \mathcal{M}.$$

A number of our proofs rely on defining the following sets:

$$\begin{aligned} U_{\bar{\tau}}^z &:= \{y \mid \|y - \Pi_z(y)\| \leq \bar{\tau}\} \cap \{y \mid \|z - \Pi_z(y)\| \leq \bar{\tau}\} \\ \tilde{\mathcal{A}}_{z, \bar{\tau}} &:= U_{\bar{\tau}}^z \cap \mathcal{M} \\ \mathcal{A}_{z, \bar{\tau}} &:= U_{\bar{\tau}}^z \cap T_z \mathcal{M}. \end{aligned}$$

$U_{\bar{\tau}}^z$ is a cylinder centered at z , and $\tilde{\mathcal{A}}_{z, \bar{\tau}}$ and $\mathcal{A}_{z, \bar{\tau}}$ are nearby regions of the manifold and tangent space, respectively. $\mathcal{A}_{z, \bar{\tau}}$ can also be defined as the projection of the cylinder onto

the tangent space; i.e., as $\Pi_z(U_{\bar{\tau}}^z)$. These sets are especially useful because, as long as $\bar{\tau} < \tau$, they allow us to work with a local parametrization of the manifold. As mentioned earlier, manifolds can be defined locally as functions from the tangent space to the normal space. The functions we are working with are in the class $C^{1,1}$; i.e., they are once continuously differentiable and have a Lipschitz gradient. This is summarized in the next theorem.

Theorem 2. *Let $\mathcal{M} \in \mathcal{G}(d, V_{max}, \tau_{min})$. Let $z \in \mathcal{M}$ and $y \in \mathcal{M}_{\bar{\tau}}$. When $\bar{\tau}$ is sufficiently small, there exists a $C^{1,1}$ function*

$$F_{z, U_{\bar{\tau}}^z} : \mathcal{A}_{z, \bar{\tau}} \rightarrow \Pi_z^{-1}(\Pi_z(0))$$

such that

$$\{y + F_{z, U_{\bar{\tau}}^z}(y) \mid y \in \mathcal{A}_{z, \bar{\tau}}\} = \tilde{\mathcal{A}}_{z, \bar{\tau}}.$$

Additionally, there exists a constant C such that $\text{Lip}(\nabla F_{z, U_{\bar{\tau}}^z}) \leq C/\tau$.

The next theorem is from [Krantz and Parks \(2012\)](#). It states that \mathcal{M} has positive reach as long as it is embedded in a Euclidean space with strictly higher dimension.

Theorem 3. *Let \mathcal{M} be a d -dimensional C^2 -submanifold of \mathbb{R}^n . If $n > d$, then \mathcal{M} has positive reach.*

Now, suppose we want a discrete approximation of a manifold \mathcal{M} at a certain resolution. Let $Y \subset \mathcal{M}$ be an η -net for \mathcal{M} if for every $p \in \mathcal{M}$ there is a $y \in Y$ such that $\|p - y\| < \eta$. The following theorem states that the size of an η -net depends on the geometry of \mathcal{M} .

Theorem 4. *Let $\mathcal{M} \in \mathcal{G}(d, V_{max}, \tau_{min})$, and let \mathcal{M} be equipped with the Euclidean metric from \mathbb{R}^n . For any $\eta > 0$, there exists an η -net of \mathcal{M} consisting of at most $CV(1/\tau^d + 1/\eta^d)$ points, where C is a universal constant.*

How well a manifold approximates a point set $\{x_i\}_{i=1}^N$ can be quantified through the empirical loss, which is defined as

$$L_{emp}(\mathcal{M}) := \frac{1}{N} \sum_{i=1}^N d(x_i, \mathcal{M})^2,$$

where $d(x_i, \mathcal{M})$ is the length of the projection from x_i onto \mathcal{M} .

Given two subsets X and Y of Euclidean space, we can measure the distance between them using the *Hausdorff distance* $H(X, Y)$. This is defined as

$$H(X, Y) := \max \left\{ \sup_{x \in X} \inf_{y \in Y} \|x - y\|, \sup_{y \in Y} \inf_{x \in X} \|x - y\| \right\}.$$

It can be shown that, given adequate sampling density, two manifolds that are close in empirical risk to a given point set are also close in Hausdorff distance.

2.2.2 Model

We assume that we are provided with $\{y_i\}_1^N$ noiselessly sampled from the uniform distribution on $\mathcal{M} \in \mathcal{G}(d, V_{max}, \tau_{min})$. We take this approach to simplify calculations. The analysis would be similar if the sample came from a (potentially Lipschitz) density bounded away from zero.

2.2.3 Approximate Squared-Distance Functions

For our purposes the most important results from [Fefferman et al. \(2016\)](#) are Theorem 13 and Lemma 14. We reproduce them below as Theorems 5 and 6, and give an adapted proof of the latter. It is beyond the scope of this chapter to discuss the proof of Theorem 5. We merely note that it relies on the implicit function theorem, so there are concrete bounds on the constants c_2, \dots, c_7 and C that control the geometry of the putative manifold.

Theorem 5 states that an *approximate squared-distance function* can be used to recover a manifold with arbitrary precision (with increasing sample size) as long as F , a scaled version of the asdf, meets three conditions related to smoothness and curvature. The notation $\partial^\alpha F(x)$ means that given a set of vectors $\alpha := \{v_1, \dots, v_{|\alpha|}\}$, the partial derivative is computed successively in the directions v_i . The third condition is the reason for the term asdf: for a

small constant ρ , $F + \rho^2$ is bounded both above and below by a multiple of $|y|^2 + \rho^2$, the approximate squared distance to the manifold.

Note that the function F always has as its domain the unit ball (or a ball whose radius is not dependent on sample size). F is not the asdf itself, but a related function applied locally after the coordinate system has been scaled up by a constant. This constant is usually a kind of bandwidth parameter that we decrease in order to get a more precise estimate of the manifold. For example, in Section 2.3, we have a scheme to decrease the bandwidth σ of the kernel density estimator, and F is the KDE applied to coordinates scaled up by $1/\sigma$.

The output set from Theorem 5 is locally a smooth graph $(x, \Psi(x))$ that lies within a tubular neighborhood of the manifold. Theorem 6 uses bounds on the smoothness of $(x, \Psi(x))$ to show that it lies away from the boundary of the tubular neighborhood, and so it is itself a manifold. We show that it is in fact very close to the original manifold, giving a bound on the Hausdorff distance in terms of a constant that can be made as small as desired.

Theorem 5. *Suppose the following conditions hold for a function F :*

1. $F : B_n(0, 1) \rightarrow \mathbb{R}$ is C^k -smooth.
2. $\partial_{x,y}^\alpha F(x, y) \leq C_0$, where $(x, y) \in B_n(0, 1)$ and $|\alpha| \leq k$.
3. For $x \in \mathbb{R}^d$, $y \in \mathbb{R}^{n-d}$ and $(x, y) \in B_n(0, 1)$

$$c_1(|y|^2 + \rho^2) \leq F(x, y) + \rho^2 \leq C_1(|y|^2 + \rho^2),$$

for $0 < \rho < c$, where c is an arbitrarily small constant depending only on C_0, c_1, C_1, k , and n .

Then there are constants c_2, \dots, c_7 and C depending only on C_0, c_1, C_1, k , and n such that:

1. For $z \in B_n(0, c_2)$, let $N(z)$ be the subspace of \mathbb{R}^n spanned by the top $n - d$ eigenvectors of $\partial^2 F(z)$. Let $\Pi_{\text{hi}} : \mathbb{R}^n \rightarrow N(z)$ be the orthogonal projection from \mathbb{R}^n to $N(z)$. Then $|\partial^\alpha \Pi_{\text{hi}}(z)| \leq C$ for $z \in B_n(0, c_2)$ and $|\alpha| \leq k - 2$.

2. There is a map

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

such that $|\Psi(0)| \leq C\rho$ and $|\partial^\alpha \Psi| \leq C^{|\alpha|}$ for $1 \leq |\alpha| \leq k-2$. The set of all $z = (x, y) \in B_d(0, c_4) \times B_{n-d}(0, c_3)$ such that

$$\{z \mid \Pi_{hi}(z)\partial F(z) = 0\} = \{(x, \Psi(x)) \mid x \in B_d(0, c_4)\}$$

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

Theorem 6. Let c_1, C_1 , and C_0 be the constants appearing in Theorem 5. Assume that $C\rho$ is sufficiently small compared to r . Define the putative submanifold

$$\mathcal{M}_{\text{put}} = \{z \in \mathcal{M}_{\min(c_3, c_4)r} \mid \Pi_{hi}(z)\partial F(z) = 0\}.$$

Then, \mathcal{M}_{put} is a submanifold of \mathbb{R}^n which has a reach greater than cr , where c depends only on C_0, c_1, C_1, k, d , and n . Furthermore, the Hausdorff distance $H(\mathcal{M}, \mathcal{M}_{\text{put}})$ is bounded above by $(C^2 + C)\rho$.

The statement of this theorem assumes that we are provided with the output set from Theorem 5; that is, we are working in the scaled-up coordinates. In the original coordinate system, \mathcal{M}_{put} is contained in $\mathcal{M}_{\min(c_3, c_4)\sigma r}$, where σ is the bandwidth. In this case, the reach is bounded below by $c\sigma r$, and $H(\mathcal{M}, \mathcal{M}_{\text{put}})$ is bounded above by $(C^2 + C)\sigma\rho$.

Proof. \mathcal{M}_{put} is locally the graph of a C^{k-2} -smooth function Ψ . To prove that it is a manifold, it is sufficient to show that it does not intersect the boundary of the tubular neighborhood $\mathcal{M}_{\min(c_3, c_4)r}$. Since Theorem 5 gives bounds on $\|\partial\Psi\|$, we can show by contradiction of the mean value theorem that every point on \mathcal{M}_{put} is within $\min(c_3, c_4)r/2$ of \mathcal{M} .

Suppose there exists a point \hat{z} on \mathcal{M}_{put} which is at a distance greater than $\min(c_3, c_4)r/2$ from \mathcal{M} . Let $z := \Pi_{\mathcal{M}}\hat{z}$. By Theorem 5, there is a point $\tilde{z} \in \mathcal{M}_{\text{put}}$ such that $\|z - \tilde{z}\| < C\rho$. Let $\tilde{v} \in T_z\mathcal{M}$ be the vector $\Pi_z(\hat{z} - \tilde{z})$. Let $\tilde{\Psi} : [0, \|\tilde{v}\|] \rightarrow \mathbb{R}^{n-d}$ define a curve on \mathcal{M}_{put} whose endpoints are \hat{z} and \tilde{z} . The existence and smoothness of $\tilde{\Psi}$ are guaranteed by Ψ , the

C^{k-2} -smooth function that locally defines \mathcal{M}_{put} . The mean value theorem states that there exists a point $x \in [0, \|\tilde{v}\|]$ such that

$$\begin{aligned} \|\partial\tilde{\Psi}(x)\| &\geq \frac{1}{\|\tilde{v}\|} \|\tilde{\Psi}(z) - \tilde{\Psi}(z + \tilde{v})\| \\ &= \frac{\|\hat{z} - \tilde{z}\|}{\|\tilde{v}\|} \\ &\geq \frac{\|z - \hat{z}\| - C\rho}{C\rho} \\ &> \frac{\min(c_3, c_4)r/2 - C\rho}{C\rho}. \end{aligned}$$

Since $C\rho$ is sufficiently small compared to $\min(c_3, c_4)r/2$, $\|\partial\tilde{\Psi}(x)\|$ can be made as large as desired. This contradicts the bound $\|\partial\Psi(x)\| < C$ and shows that \mathcal{M}_{put} lies away from the boundary of $\mathcal{M}_{\min(c_3, c_4)r}$. In fact, the expression in the third line above must be less than C , which shows that $\|z - \hat{z}\| < (C^2 + C)\rho$. Theorem 5 states that every point on \mathcal{M} is within $C\rho$ of \mathcal{M}_{put} , so we have the desired bound on the Hausdorff distance.

By Theorem 1, the reach of \mathcal{M}_{put} is defined as follows:

$$\text{reach}(\mathcal{M}_{\text{put}}) = \inf_{\substack{x \neq y \\ x, y \in \mathcal{M}_{\text{put}}}} \frac{\|x - y\|^2}{2\|y - \Pi_x y\|}.$$

Let c' be a constant depending on C_0, c_1, C_1, k, d , and n . If $\|x - y\| \geq r/c'$, then

$$\frac{\|x - y\|^2}{2\|y - \Pi_x y\|} \geq \frac{(r/c')^2}{2(r/c')}.$$

Now, suppose $\|x - y\| < r/c'$. If x and y are close together, this quantity is controlled by the second derivative of the C^{k-2} function locally defining \mathcal{M}_{put} . That is, $\|y - \Pi_x y\|$ is on the order of $C^2\|x - y\|^2$, implying that

$$\inf_{\substack{x \neq y \\ x, y \in \mathcal{M}_{\text{put}} \\ \|x - y\| < r/c'}} \frac{\|x - y\|^2}{2\|x - \Pi_y x\|} \geq \frac{\|x - y\|^2}{2c''C^2\|x - y\|^2}$$

for some constant c'' . Therefore,

$$\text{reach}(\mathcal{M}_{\text{put}}) \geq \min\left(\frac{r}{2c'}, \frac{1}{2c''C^2}\right).$$

■

2.2.4 Ridges and Gradient Descent

To actually find a putative manifold using an approximate squared-distance function F , we can use a method introduced by [Ozertem and Erdogmus \(2011\)](#). Let ∂F and $\partial^2 F$ be the gradient and Hessian of F , respectively. At a point $x \in \mathbb{R}^n$, let $\{v_1, \dots, v_n\}$ be the eigenvectors of $\partial^2 F$ associated with the eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ (listed in decreasing order). Let $N(z)$ be the subspace of \mathbb{R}^n spanned by the top $n - d$ eigenvectors of $\partial^2 F(z)$. Recall that $\Pi_{hi} : \mathbb{R}^n \rightarrow N(z)$ is the orthogonal projection from \mathbb{R}^n to $N(z)$. Note that $\Pi_{hi} = VV^\top$, where V is a matrix whose columns are $[v_1 | \dots | v_{n-d}]$.

[Ozertem and Erdogmus \(2011\)](#) give an algorithm to compute the set

$$\{z \mid \Pi_{hi}(z)\partial F(z) = 0\},$$

which is termed the *d-dimensional ridge* of F . This is, of course, the local definition of \mathcal{M}_{put} from Theorems 5 and 6. In order to find a ridge, an initial set of points is chosen and then iteratively shifted in the direction $VV^\top \partial F$ until a tolerance condition is met. This is essentially a subspace-constrained variant of gradient descent.

2.2.5 Empirical Processes

In Section 2.3, we need to bound various quantities that are functions of the kernel density estimator. This is difficult to do because they are empirical averages over a finite number of samples. It is easier to bound the expectation of these quantities and then bound their difference using results from empirical processes, which we summarize here.

Let \mathcal{G} be a class of functions from $\mathbb{R}^n \rightarrow \mathbb{R}$. If \mathcal{G} consists of bounded functions, the *empirical Rademacher average* is given by

$$\mathcal{R}_N(\mathcal{G}) = \mathbb{E}_\sigma \frac{1}{N} \left[\sup_{g \in \mathcal{G}} \left(\sum_{i=1}^N \sigma_i g(x_i) \right) \right],$$

where $\{x_i\}_1^N$ is an i.i.d. sample from the distribution \mathcal{P} and $\sigma := \{\sigma_1, \dots, \sigma_N\}$ is a vector of Rademacher random variables. (Rademacher random variables take the values ± 1 with

equal probability). Letting \mathcal{P}_N denote the empirical distribution on $\{x_i\}_1^N$, the following holds for $0 < \delta < 1$:

$$\mathbb{P} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_{\mathcal{P}_N} g - \mathbb{E}_{\mathcal{P}} g| < 2\mathcal{R}_N(\mathcal{G}) + \sqrt{\frac{2 \log(2/\delta)}{N}} \right] > 1 - \delta.$$

It is usually difficult to calculate $\mathcal{R}_N(\mathcal{G})$ directly from the definition. However, the next theorem states an upper bound that is dependent on the size of \mathcal{G} , which is often easy to estimate. Let the *covering number* $N(\eta, \mathcal{G}, \|\cdot\|)$ be the minimum number of elements in an η -net of \mathcal{G} with respect to the norm $\|\cdot\|$. Let the *metric entropy* be defined as $\log N(\eta, \mathcal{G}, \|\cdot\|)$. The Rademacher complexity can be bounded using a modified form of Dudley's entropy integral (Sridharan and Srebro, 2010):

Theorem 7 (Modified Dudley's integral).

$$\mathcal{R}_N(\mathcal{G}) \leq \inf_{\gamma \geq 0} \left\{ 4\gamma + 12 \int_{\gamma}^{\sup_{g \in \mathcal{G}} \|g\|_{\infty}} \sqrt{\frac{\log N(\eta, \mathcal{G}, \|\cdot\|_{\mathcal{L}_2(\mathcal{P}_N)})}{N}} d\eta \right\}.$$

2.3 Kernel Density Estimation

Consider the kernel density estimator

$$\tilde{p}_N(x) := \frac{1}{N} \sum_{i=1}^N G_{\sigma}(x; y_i),$$

where $G_{\sigma}(x; y) := C_{\sigma} e^{-\|x-y\|^2/2\sigma^2}$, $C_{\sigma} := (2\pi\sigma^2)^{-d/2}$, and $x \in \mathcal{M}_{\sigma}$ (the tubular neighborhood of \mathcal{M} with width σ). The denominator of C_{σ} has $2\pi\sigma^2$ raised to the power $d/2$ and not $n/2$ because we are trying to estimate a d -dimensional surface. In Theorem 16, we show that a function based on \tilde{p}_N can recover a manifold \mathcal{M} when we are given noiseless samples from \mathcal{M} .

2.3.1 Definition of the asdf

Recall that Theorem 5 must actually be applied in a coordinate system scaled by a bandwidth parameter that becomes more precise with increasing sample size. For the kernel density

estimator, this parameter is, of course, σ . (If we do not scale by σ , it is clear that $|\partial^\alpha \tilde{p}_N|$ is bounded above by an increasing function of $1/\sigma$ instead of a universal constant C_0). To this purpose, make the following transformations:

$$\begin{aligned} x &\mapsto x/(\sigma\sqrt{2\pi}) \\ y_i &\mapsto y_i/(\sigma\sqrt{2\pi}) \\ \sigma &\mapsto 1/\sqrt{2\pi}. \end{aligned}$$

Note that the geometric properties of \mathcal{M} change in the obvious ways: the reach becomes $\tau/(\sigma\sqrt{2\pi})$ and the volume is $O(V/\sigma^d)$. In the transformed case, let $\hat{\tau}$, $\hat{\tau}$, and \hat{V} denote the analogs of the obvious quantities. For z the projection of x onto \mathcal{M} , let $\tilde{\mathcal{A}}_z := \tilde{\mathcal{A}}_{z, \hat{\tau}}$ and $\mathcal{A}_z := \mathcal{A}_{z, \hat{\tau}}$. Recall that these are regions of \mathcal{M} and $T_z\mathcal{M}$, respectively, which are near the point $z \in \mathcal{M}$. Define the normalizing factor

$$N_f := \text{Vol}(\tilde{\mathcal{A}}_z) / \left(\text{Vol}(\mathcal{A}_z) \times \hat{V} \right).$$

The appropriate estimator to analyze is any convenient function of p_N/N_f , where

$$p_N(x) := \frac{1}{N} \sum_{i=1}^N e^{-\pi\|x-y\|^2}.$$

We choose to work with $-\log p_N(x) + \log N_f$ as our potential asdf. The first condition from Theorem 5 follows immediately, as seen in the following lemma.

Lemma 8. $-\log p_N(x)$ is C^k -smooth.

Proof. $G_{1/\sqrt{2\pi}}$ is C^k -smooth, so by linearity, $p_N(x)$ is C^k -smooth. By the chain rule, $-\log p_N(x)$ is C^k -smooth. ■

In Lemmas 13 and 15 below, we show that $-\log p_N(x) + \log N_f$ also satisfies the second and third conditions from Theorem 5 with high probability. Before detailing the proofs, we briefly discuss our scheme for selecting σ and $\bar{\tau}$.

2.3.2 Selecting the Bandwidth σ

The procedure we assume is that a fixed value σ_1 of σ is chosen by the experimenter as well as a value of $\bar{\tau}$ that depends on σ . Without making any claims about optimality, we choose $\bar{\tau} := \sigma^{5/6}$. We prove that there exists a lower bound K_1 for $\mathbb{E}p_N(x)$ given σ_1 . We then use empirical processes to show that $p_N(x)$ concentrates around $\mathbb{E}p_N(x)$, and is within ε_1 of $\mathbb{E}p_N(x)$ with high probability. Since ε_1 is a decreasing function of the sample size N , we can increase N until $\varepsilon_1 < K_1$ giving us a lower bound for p_N . This allows us to derive an upper bound for $\partial^\alpha(-\log p_N(x))$. We also find an expression for ρ^2 in terms of σ_1 and use this to show that condition 3 of Theorem 5 holds. If ρ^2 is not small enough, we can repeat this procedure using a fixed value $\sigma_{i+1} := \frac{\sigma_i}{2}$ of σ in each subsequent iteration.

2.3.3 Bounding p_N in Expectation

To prove the second and third conditions, we need upper and lower bounds for p_N . It is more convenient to work initially in the continuous setting, which amounts to bounding $\mathbb{E}p_N$. This is the Gaussian kernel integrated against $\mu_{\mathcal{M}} := \frac{1}{V}d\text{Vol}(\mathcal{M})$, the measure that is uniform with respect to the volume form. Explicitly,

$$\mathbb{E}p_N(x) = \int_{\mathcal{M}} e^{-\pi\|x-y\|^2} d\mu_{\mathcal{M}}(y).$$

Points on \mathcal{M} that are far away from x do not contribute very much to the value of this integral. In fact, the value of $-\log \mathbb{E}p_N$ is very close to

$$\hat{F}_{\tilde{\mathcal{A}}_z} := -\log \int_{\tilde{\mathcal{A}}_z} e^{-\pi\|x-y\|^2} d\mu_{\mathcal{M}}(y),$$

where z is the projection of x onto \mathcal{M} . Define its approximation

$$\hat{F}_{\mathcal{A}_z} := -\log \int_{\mathcal{A}_z} e^{-\pi\|x-y\|^2} N_f d\mathcal{L}_d(y),$$

where \mathcal{L}_d is the d -dimensional Lebesgue measure on $T_z\mathcal{M}$.

Decreasing σ corresponds to estimating \mathcal{M} with greater precision. Even though this expands the unit ball, leading to $\hat{\tau} \rightarrow \infty$, the ratio $\hat{\tau}/\bar{\tau} \rightarrow 0$. This implies that \mathcal{A}_z and $\tilde{\mathcal{A}}_z$

are shrinking relatively closer and closer to z , and $\tilde{\mathcal{A}}_z$ is very close to an affine space. Thus, we expect $\hat{F}_{\tilde{\mathcal{A}}_z}$ and $\hat{F}_{\mathcal{A}_z}$ to grow closer together. To prove this, we first need to show that the pushforward of the uniform measure on the manifold has a density $p(y)$ that is close to the uniform density on $B_d(\hat{\tau})$. Of course, we don't want this to be a proper density on $B_d(\hat{\tau})$; we want it to have the same total measure as $\int_{\tilde{\mathcal{A}}_z} d\mu_{\mathcal{M}}(y')$. In the following lemma, we quantify how much $p(y)$ can deviate from N_f on \mathcal{A}_z .

Lemma 9. *Let $z \in \mathcal{M}$, and let $y \in T_z\mathcal{M}$. The pushforward of $\mu_{\mathcal{M}}$ to $T_z\mathcal{M}$ has density $p(y)$ on \mathcal{A}_z with respect to \mathcal{L}_d such that*

$$N_f \times \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2}\right)^{-d/2} \leq p(y) \leq \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2}\right)^{d/2} \times N_f.$$

Proof. Assume that z is the origin and the first d coordinates lie in $T_z\mathcal{M}$. \mathcal{M} is a submanifold of \mathbb{R}^n defined locally by the function $G_{z,U_{\hat{\tau}}^z}$ that maps $(x_1, \dots, x_d) \mapsto (x_1, \dots, x_d, F_{z,U_{\hat{\tau}}^z})$. Recall from Theorem 2 that $F_{z,U_{\hat{\tau}}^z} : \mathbb{R}^d \rightarrow \mathbb{R}^{n-d}$ is a $C^{1,1}$ function whose Jacobian $J \in \mathbb{R}^{(n-d) \times d}$ has a Lipschitz constant bounded above by $C/\hat{\tau}$. J evaluated at z is 0 since \mathbb{R}^d is tangent to \mathcal{M} at z ; this implies $\|J\|_F \leq (C\hat{\tau})/\hat{\tau}$ within a radius of $\hat{\tau}$, where $\|\cdot\|_F$ is the Frobenius norm. We can find the desired bound on $p(y)$ by finding the ratio of the volume elements of \mathcal{A}_z and $\tilde{\mathcal{A}}_z$, normalizing this so it integrates to one over \mathcal{A}_z , and multiplying by $\text{Vol}(\tilde{\mathcal{A}}_z) / (\hat{V})$. $G_{z,U_{\hat{\tau}}^z}$ has Jacobian $[I|J^\top]^\top$, allowing us to write

$$p(y) = \frac{\text{Vol}(\tilde{\mathcal{A}}_z)}{\hat{V}} \times \frac{\sqrt{\det(I + J^\top J)}}{\int_{\mathcal{A}_z} \sqrt{\det(I + J^\top J)} d\mathcal{L}_d(y)}.$$

Let λ_i be the eigenvalues of $J^\top J$. $J^\top J$ is positive semidefinite, so $\lambda_i \geq 0$. Then,

$$\begin{aligned} \sqrt{\det(I + J^\top J)} &= \left(\prod_{i=1}^d (1 + \lambda_i) \right)^{1/2} \\ &\leq \left(\prod_{i=1}^d (1 + \|J^\top J\|_F) \right)^{1/2} \\ &\leq \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2} \right)^{d/2}. \end{aligned}$$

Since the map from \mathcal{M} to \mathbb{R}^d is a contraction, $1 \leq \sqrt{\det(I + J^\top J)}$. Clearly, we also have

$$\text{Vol}(\mathcal{A}_z) \leq \int_{\mathcal{A}_z} \sqrt{\det(I + J^\top J)} d\mathcal{L}_d(y) \leq \text{Vol}(\mathcal{A}_z) \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2}\right)^{d/2},$$

which is enough to show the lemma. \blacksquare

We can use this bound on $p(y)$ to simplify the integration of functions over \mathcal{M} . As we mentioned earlier, $-\log \mathbb{E}p_N$ is a function of an integral whose major contribution comes from the region $\tilde{\mathcal{A}}_z$. (A crude bound suffices for the contribution from the region $\mathcal{M} \setminus \tilde{\mathcal{A}}_z$). In the following lemma, we show that $\hat{F}_{\tilde{\mathcal{A}}_z}$ and $\hat{F}_{\mathcal{A}_z}$ are very close together. By using the pushforward we can perform both integrals over \mathcal{A}_z using Lebesgue measure. To do so we need a bound on the ratio of $p(y)$ to N_f (which we have) as well as a bound on the ratio between the integrands. We find that $\hat{F}_{\tilde{\mathcal{A}}_z}$ and $\hat{F}_{\mathcal{A}_z}$ are within a constant C_f of each other. By decreasing σ , C_f can be made as small as desired.

Lemma 10. *Let $x \in \mathcal{M}_{\hat{\tau}}$, and let z be the projection of x onto \mathcal{M} . Then, $|\hat{F}_{\tilde{\mathcal{A}}_z} - \hat{F}_{\mathcal{A}_z}| \leq C_f$, where*

$$C_f := \frac{dC^2 \hat{\tau}^2}{2\hat{\tau}^2} + \left(\frac{\hat{\tau}^4}{\hat{\tau}^2} + \frac{2\sqrt{2}\hat{\tau}^3}{\hat{\tau}} \right) \pi.$$

Proof. Assume that z is the origin and $T_z \mathcal{M}$ is identified with the first d coordinates. The following chain of inequalities holds, where $y \in T_z \mathcal{M}$, $y' := y + F_{z, U_{\hat{\tau}}^z}$ is a point on the manifold, and J is the Jacobian of $F_{z, U_{\hat{\tau}}^z}$:

$$\begin{aligned} |\hat{F}_{\tilde{\mathcal{A}}_z} - \hat{F}_{\mathcal{A}_z}| &= \left| \log \frac{\int_{\mathcal{A}_z} e^{-\pi \|x-y\|^2} N_f d\mathcal{L}_d(y)}{\int_{\tilde{\mathcal{A}}_z} e^{-\pi \|x-y'\|^2} d\mu_{\mathcal{M}}(y')} \right| \\ &= \left| \log \frac{\int_{\mathcal{A}_z} e^{-\pi \|x-y\|^2} N_f d\mathcal{L}_d(y)}{\int_{\mathcal{A}_z} e^{-\pi \|x-y'\|^2} p(y) d\mathcal{L}_d(y)} \right| \\ &\leq \left| \sup_{y \in \mathcal{A}_z} \log \frac{e^{-\pi \|x-y\|^2} N_f}{e^{-\pi \|x-y'\|^2} p(y)} \right| \end{aligned}$$

$$\leq \sup_{y \in \mathcal{A}_z} \left| \log \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2} \right)^{d/2} \right| + \sup_{y \in \mathcal{A}_z} \left| -(x-y)^2 + (x-y')^2 \right| \pi.$$

The first term in the last line comes from Lemma 9. A Taylor expansion (valid for $|x| < 1$) shows that

$$\log(1+x)^{d/2} = \frac{dx}{2} - \frac{dx^2}{4} + \frac{dx^3}{6} + O(x^4).$$

Therefore,

$$\left| \log \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2} \right)^{d/2} \right| \leq \frac{dC^2 \hat{\tau}^2}{2\hat{\tau}^2}$$

as long as $\hat{\tau}/\hat{\tau}$ is smaller than a controlled constant. To bound the other term, we use the law of cosines in conjunction with Theorem 1, which shows that

$$\begin{aligned} \|y' - y\| &= \|y' - \Pi_z y'\| \leq \frac{\|y' - z\|^2}{2\hat{\tau}} \\ &\leq \frac{(\sqrt{2}\hat{\tau})^2}{2\hat{\tau}}. \end{aligned}$$

Let θ be the angle between $y - y'$ and $x - y$. Then, we have:

$$\begin{aligned} \left| \|y' - x\|^2 - \|y - x\|^2 \right| \pi &= \left| \|y - y'\|^2 - 2\|y - y'\| \|y - x\| \cos \theta \right| \pi \\ &\leq \left(\left(\frac{\hat{\tau}^2}{\hat{\tau}} \right)^2 + 2 \left(\frac{\hat{\tau}^2}{\hat{\tau}} \right) (\sqrt{2}\hat{\tau}) \right) \pi. \end{aligned}$$

Thus, $\left| \hat{F}_{\tilde{\mathcal{A}}_z} - \hat{F}_{\mathcal{A}_z} \right| \leq C_f$, where C_f is defined in the statement of the lemma. ■

To actually find the lower bound for p_N , we bound $\hat{F}_{\mathcal{A}_z}$ in the next lemma by using a d -dimensional Gaussian concentration inequality. The upper bound is much simpler to derive; we include it as well. These bounds are important in verifying the second and third conditions of Theorem 5. The third condition essentially says that our function is an approximate squared-distance function. That is, given a point $x \in \mathcal{M}_{1/\sqrt{2\pi}}$ and its projection $z \in \mathcal{M}$, we should have upper and lower bounds that are close to $\|x - z\|^2$. Since our putative

asdf is $-\log p_N$, we need bounds for p_N that are within a multiplicative factor of $e^{-\|x-z\|^2\pi}$. In the proof of the following lemma we find these pointwise bounds.

The second condition requires that we find an upper bound for $\partial^\alpha(-\log p_N)$. This derivative consists of terms that have powers of p_N in the denominator and combinations of powers of partial derivatives of p_N in the numerator. Thus, we need a uniform lower bound for p_N over $\mathcal{M}_{1/\sqrt{2\pi}}$; this follows by taking the infimum of the pointwise bound over the tubular neighborhood. We also need bounds for $|\mathbb{E}\partial^\alpha p_N|$, but we defer these to the proof of Lemma 13.

Lemma 11. $p_N(x)$ is bounded in expectation. More precisely, $\inf_{x \in \mathcal{M}_{1/\sqrt{2\pi}}} \mathbb{E}p_N(x) \geq K_1$, where

$$K_1 := N_f e^{-1/2} \left(1 - 2e^{-\left(\hat{r} - \sqrt{d/(2\pi)}\right)^2 \pi} \right) e^{-C_f};$$

furthermore, $\sup_{x \in \mathcal{M}_{1/\sqrt{2\pi}}} \mathbb{E}p_N(x) \leq K_2$, where

$$K_2 := e^{C_f} N_f + e^{-\hat{r}^2 \pi/2}.$$

Proof. Let z be the projection of x onto \mathcal{M} , and let $y \in T_z \mathcal{M}$.

$$\begin{aligned} \mathbb{E}p_N(x) &\geq \int_{\tilde{\mathcal{A}}_z} e^{-\pi\|x-y'\|^2} d\mu_{\mathcal{M}}(y') \\ &\geq \left(\int_{\mathcal{A}_z} e^{-\pi\|x-y\|^2} N_f d\mathcal{L}_d(y) \right) e^{-C_f}, \end{aligned}$$

where the second inequality is due to Lemma 10. By orthogonality, $(x-z)^\top(y-z) = 0$, so we rewrite the integral over \mathcal{A}_z as follows:

$$\begin{aligned} \int_{\mathcal{A}_z} e^{-\pi\|x-y\|^2} N_f d\mathcal{L}_d(y) &= N_f e^{-\|x-z\|^2\pi} \int_{\mathcal{A}_z} e^{-\pi\|z-y\|^2} d\mathcal{L}_d(y) \\ &= N_f e^{-\|x-z\|^2\pi} \mathbb{P}[\|y-z\| \leq \hat{r}], \end{aligned}$$

where the probability is with respect to a d -dimensional multivariate Gaussian with covariance $\frac{1}{2\pi}I$. Letting z be the origin for simplicity, we know from standard Gaussian concentration results (Boucheron, Lugosi, and Massart, 2013) that

$$\mathbb{P}[\|y\| - \mathbb{E}\|y\| \leq t] \geq 1 - 2e^{-t^2\pi}$$

for $t > 0$. We can calculate $\mathbb{E}[\|y\|^2]$ and then get a bound for $\mathbb{E}[\|y\|]$ by using Jensen's inequality.

Make the substitutions

$$\left\{ x_1 \mapsto r \cos \phi_1, x_{2 \leq i \leq d-1} \mapsto r \cos \phi_i \prod_{j=1}^{i-1} \sin \phi_j, x_d \mapsto r \sin \phi_{d-1} \prod_{j=1}^{d-2} \sin \phi_j \right\},$$

and let

$$dV := r^{d-1} \prod_{j=1}^{d-2} \sin^{d-j-1} \phi_j dr d\phi_1 \dots d\phi_{d-1}.$$

We have

$$\begin{aligned} \mathbb{E}[\|y\|^2] &= \int_{\mathbb{R}^d} \|y\|^2 e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \\ &= \int_0^\infty r^{d+1} e^{-r^2 \pi} dr \times \prod_{j=1}^{d-2} \int_0^\pi \sin^{d-j-1} \phi_j d\phi_j \\ &\quad \times \int_0^{2\pi} d\phi_{d-1} \\ &= \pi^{-(2+d)/2} \Gamma(1 + d/2) \times \prod_{j=1}^{d-2} \frac{\sqrt{\pi} \Gamma((d-j)/2)}{\Gamma(1 + (d-j-1)/2)} \times 2\pi \\ &= \frac{d}{2\pi}. \end{aligned}$$

The product in the third line telescopes to $\pi^{(d-2)/2} / (\Gamma(d/2))$; simplifying yields the fourth line.

It follows that $\mathbb{E}\|y\| \leq \sqrt{d/(2\pi)}$. Setting $t := \hat{\tau} - \sqrt{d/(2\pi)}$ (and assuming that σ is small enough so $t > 0$), we see that

$$\mathbb{P}[\|y\| \leq \hat{\tau}] \geq 1 - 2e^{-\left(\hat{\tau} - \sqrt{d/(2\pi)}\right)^2 \pi}.$$

Consequently,

$$\mathbb{E}p_N(x) \geq N_f e^{-\|x-z\|^2 \pi} \left(1 - 2e^{-\left(\hat{\tau} - \sqrt{d/(2\pi)}\right)^2 \pi} \right) e^{-C_f}.$$

The first part of the lemma follows by taking the infimum over $\mathcal{M}_{1/\sqrt{2\pi}}$.

To find an upper bound, first write the expectation as

$$\begin{aligned}\mathbb{E}p_N(x) &= \int_{\mathcal{M}} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') \\ &= \int_{\tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') + \int_{\mathcal{M}\setminus\tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y').\end{aligned}$$

The first term can be bounded as follows:

$$\begin{aligned}\int_{\tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') &\leq e^{C_f} \int_{\mathcal{A}_z} e^{-\|x-y\|^2\pi} N_f d\mathcal{L}_d(y) \\ &\leq e^{C_f} N_f e^{-\|x-z\|^2\pi} \int_{\mathcal{A}_z} e^{-\|z-y\|^2\pi} d\mathcal{L}_d(y) \\ &\leq e^{C_f} N_f e^{-\|x-z\|^2\pi}.\end{aligned}$$

Now consider $y' \in \mathcal{M}\setminus\tilde{\mathcal{A}}_z$. Since $\|z-y'\| \leq \|x-y'\| + \|x-z\|$ and $\|x-z\| \leq 1/\sqrt{2\pi} \leq \hat{\tau} \leq \|z-y'\|$, we have $(\|x-z\| - \hat{\tau})^2 \leq \|x-y'\|^2$. This gives us the following bound for the second term as long as $\hat{\tau}$ is large enough:

$$\begin{aligned}\int_{\mathcal{M}\setminus\tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') &\leq e^{-(\|x-z\| - \hat{\tau})^2\pi} \times \int_{\mathcal{M}\setminus\tilde{\mathcal{A}}_z} d\mu_{\mathcal{M}}(y') \\ &\leq e^{-\hat{\tau}^2\pi/2}.\end{aligned}$$

Thus, for $x \in \mathcal{M}_{1/\sqrt{2\pi}}$,

$$\mathbb{E}p_N(x) \leq e^{C_f} N_f + e^{-\hat{\tau}^2\pi/2}.$$

■

Note that the values we chose for σ and $\bar{\tau}$ are appropriate given our calculations in this section. For decreasing σ , we would like for $p(y)$ to grow closer to N_f in Lemma 9 and for C_f to tend to zero in Lemma 10; we also need $\left(1 - 2e^{-(\hat{\tau} - \sqrt{d/(2\pi)})^2\pi}\right)$, the Gaussian concentration probability, to grow closer to 1 in the previous lemma. Our choice of $\bar{\tau} := \sigma^{5/6}$ is appropriate given these constraints. For σ small enough, $\left(1 - 2e^{-(\hat{\tau} - \sqrt{d/(2\pi)})^2\pi}\right)e^{-C_f} \approx 1$ and $\text{Vol}(\tilde{\mathcal{A}}_z) / \left(\text{Vol}(\mathcal{A}_z) \times \hat{V}\right) \approx 1/\hat{V} \approx \sigma^d/V$. Thus, $K_1 \approx e^{-1/2}\sigma^d/V$ and $K_2 \approx \sigma^d/V$.

2.3.4 Finite Sample Bounds for p_N and $\partial^\alpha(-\log p_N(x))$

In Lemma 11, we proved a statement about $\inf_{x \in \mathcal{M}_{1/\sqrt{2\pi}}} \mathbb{E}p_N(x)$ whereas we really need a statement about $\inf_{x \in \mathcal{M}_{1/\sqrt{2\pi}}} p_N(x)$. We can use methods from empirical processes to relate these quantities. Let \mathcal{F} consist of functions $f : \mathcal{M} \rightarrow [0, 1]$ where each f has the form $e^{-\pi\|x-y'\|^2}$ with $y' \in \mathcal{M}$. Here, x is fixed, and each $x \in \mathcal{M}_{1/\sqrt{2\pi}}$ corresponds to a different f . Note that $p_N(x)$ is equivalent to $\mathbb{E}_N f$ and $\mathbb{E}p_N(x)$ is equivalent to $\mathbb{E}f$. We have the tools to prove that for $0 < \delta < 1$,

$$\mathbb{P} \left[\sup_{f \in \mathcal{F}} |\mathbb{E}_N f - \mathbb{E}f| \leq \varepsilon_1 \right] \geq 1 - \delta,$$

where ε_1 is a function of δ and N . We rewrite the form of the probability bound in part (a) of Lemma 12 so that it is in terms of p_N and $\mathbb{E}p_N$.

In part (b) of Lemma 12, we prove a similar concentration bound for particular derivatives of $e^{\|x-z\|^2} p_N$. Let $\mathcal{F}_{\beta,v}$ consist of functions $f : \mathcal{M} \rightarrow \mathbb{R}$ where each f is of the form $\partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi}$ with $x \in \mathcal{M}_{1/\sqrt{2\pi}}$, $y' \in \mathcal{M}$, and $v \in B_n(0, 1)$. These functions are involved in finding an upper bound for $\partial^\alpha(-\log(p_N(x)))$.

Lemma 12. *Let \mathcal{F} be the class of functions consisting of $e^{-\pi\|x-y'\|^2}$ indexed by $x \in \mathcal{M}_{1/\sqrt{2\pi}}$. For a given β and $v \in B_n(0, 1)$, let $\mathcal{F}_{\beta,v}$ be the class of functions consisting of $\partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi}$ indexed by $x \in \mathcal{M}_{1/\sqrt{2\pi}}$.*

(a) For $0 < \delta < 1$,

$$\mathbb{P} \left[\sup_{x \in \mathcal{M}_{1/\sqrt{2\pi}}} |p_N(x) - \mathbb{E}p_N(x)| \leq \varepsilon_1 \right] \geq 1 - \delta,$$

where

$$\varepsilon_1 := \frac{24}{\sqrt{N}} \left(\frac{\sqrt{\pi n}}{2} + \sqrt{\log C'} \right) + \sqrt{\frac{2 \log(2/\delta)}{N}}$$

and

$$C' := C\hat{V}100^d \left(2\sqrt{2\pi/e} \right)^n.$$

(b) For $0 < \delta < 1$,

$$\mathbb{P} \left[\sup_{x \in \mathcal{M}_{1/\sqrt{2\pi}}} \left| \partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi} - \mathbb{E} \left[\partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi} \right] \right| \leq \varepsilon_{1,\beta} \right] \geq 1 - \delta,$$

where

$$\varepsilon_{1,\beta} := \frac{24}{\sqrt{N}} \left(\frac{\sqrt{\pi n}}{2} + \sqrt{\log C'_\beta} \right) + \sqrt{\frac{2 \log(2/\delta)}{N}}$$

and

$$C'_\beta := C \hat{V} 100^d \left(2 \left(1 + \sqrt{5 + 4\beta} \right)^{\beta+1} e^{(-1-2\beta+\sqrt{5+4\beta})/4} \left(\frac{2}{\pi} \right)^{-(\beta+1)/2} \right)^n.$$

Proof. We can bound $\sup_x |p_N(x) - \mathbb{E}p_N(x)|$ through a method from empirical processes by first determining the covering number of \mathcal{F} and then using Dudley's integral. Since \mathcal{F} is a class of Lipschitz functions parametrized by points in $\mathcal{M}_{\hat{\pi}}$, we can relate its covering number to the covering number of this parameter space.

From empirical process theory, we know that

$$\mathbb{P} \left[\sup_{f \in \mathcal{F}} |p_N(x) - \mathbb{E}p_N(x)| \leq 2\mathcal{R}_N(\mathcal{F}) + \sqrt{\frac{2 \log(2/\delta)}{N}} \right] \geq 1 - \delta.$$

$\mathcal{R}_N(\mathcal{F})$ is the Rademacher complexity of \mathcal{F} , which can be bounded using Theorem 7. Let $\mathcal{N}(\eta, \mathcal{F}, \|\cdot\|)$ be the covering number at scale η with respect to norm $\|\cdot\|$. Then,

$$\begin{aligned} \mathcal{R}_N(\mathcal{F}) &\leq \inf_{\varepsilon' \geq 0} \left\{ 4\varepsilon' + 12 \int_{\varepsilon'/4}^{\sup_{f \in \mathcal{F}} \sqrt{\hat{\mathbb{E}}[f^2]}} \sqrt{\frac{\log \mathcal{N}(\eta, \mathcal{F}, \|\cdot\|_{\mathcal{L}_2(P_N)})}{N}} d\eta \right\} \\ &\leq \inf_{\varepsilon' \geq 0} \left\{ 4\varepsilon' + 12 \int_{\varepsilon'/4}^{\sup_{f \in \mathcal{F}'} \sqrt{\hat{\mathbb{E}}[f^2]}} \sqrt{\frac{\log \mathcal{N}(\eta, \mathcal{F}, \|\cdot\|_\infty)}{N}} d\eta \right\}. \end{aligned}$$

The second inequality is well-known. Each $f \in \mathcal{F}$ is parametrized by $x \in \mathcal{M}_{1/\sqrt{2\pi}}$ and is at most L -Lipschitz in this parameter. If we can calculate L , we can also bound the covering number of \mathcal{F} by relating it to the covering number of the tubular neighborhood.

That is,

$$\begin{aligned} \mathcal{N}(\eta, \mathcal{F}, \|\cdot\|_\infty) &\leq \mathcal{N}(\eta/L, \mathcal{M}_{1/\sqrt{2\pi}}, \|\cdot\|_2) \\ &\leq C\hat{V}100^d \left(\frac{1}{\eta/L} + 1 \right)^n, \end{aligned}$$

where the second line follows from taking a $1/100$ -net of \mathcal{M} , placing unit n -balls at each net point, and then finding an η/L -net of those.

Now we find L . For simplicity, assume x has coordinates (x_1, \dots, x_n) that have been centered around any point on the manifold. By the symmetry of $\|x\|$, we only need to consider one coordinate.

$$\begin{aligned} \text{Lip}(f) &= \sup_{f \in \mathcal{F}} \|\nabla f\| \leq \sup_{x \in \mathcal{M}} \left\| \nabla \left(e^{-\|x\|^2 \pi} \right) \right\| \\ &\leq \sup_{x \in \mathbb{R}^n} \left| \frac{\partial}{\partial x_1} e^{-\|x\|^2 \pi} \right| \\ &= 2e^{-x_1^2 \pi} x_1 \Big|_{x_1=1/\sqrt{2\pi}} \\ &= \sqrt{\frac{2\pi}{e}} \\ &=: L. \end{aligned}$$

Since η ranges between 0 and 1, $L/\eta > 1$. Define $C' := C\hat{V}100^d(2\sqrt{2\pi/e})^n$; then,

$$\mathcal{N}(\eta, \mathcal{F}, \|\cdot\|_\infty) \leq C' \eta^{-n}.$$

Using the monotonicity of log and the square root,

$$\begin{aligned} \mathcal{R}_N(\mathcal{F}) &\leq 12 \int_0^1 \left(\sqrt{\frac{\log C'}{N}} + \sqrt{\frac{-n \log \eta}{N}} \right) d\eta \\ &= \frac{12}{\sqrt{N}} \left(\frac{\sqrt{\pi n}}{2} + \sqrt{\log C'} \right). \end{aligned}$$

Thus, with high probability,

$$\sup_{f \in \mathcal{F}} |p_N(x) - \mathbb{E}p_N(x)| \leq \frac{24}{\sqrt{N}} \left(\frac{\sqrt{\pi n}}{2} + \sqrt{\log C'} \right) + \sqrt{\frac{2 \log(2/\delta)}{N}},$$

which proves (a).

The proof of part (b) is nearly the same, with the only difference being in the covering number of the parameter space. If each $f \in \mathcal{F}_{\beta, v}$ is at most L_β -Lipschitz, then

$$\mathcal{N}(\eta, \mathcal{F}_\beta, \|\cdot\|_\infty) \leq C'_\beta \eta^{-n},$$

where $C'_\beta := C\hat{V}100^d(2L_\beta)^n$. Since

$$\partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi} = e^{2\pi x^\top y'} (2\pi)^\beta (y'^\top v)^\beta e^{-\|y'\|^2 \pi},$$

we have

$$\begin{aligned} \text{Lip}(f) &= \sup_{f \in \mathcal{F}_{\beta, v}} \|\nabla f\| \\ &\leq \sup_{x \in \mathcal{M}_{1/\sqrt{2\pi}}, y' \in \mathbb{R}^n, v \in B_n(0,1)} \left\| \nabla \left(e^{2\pi x^\top y'} \right) (2\pi)^\beta (y'^\top v)^\beta e^{-\|y'\|^2 \pi} \right\| \\ &= \sup_{x \in \mathcal{M}_{1/\sqrt{2\pi}}, y' \in \mathbb{R}^n, v \in B_n(0,1)} (2\pi)^\beta |y'^\top v|^\beta e^{-\|y'\|^2 \pi} \sqrt{\sum_{i=1}^n (2\pi y'_i e^{2\pi x^\top y'})^2} \\ &= \sup_{x \in \mathcal{M}_{1/\sqrt{2\pi}}, y' \in \mathbb{R}^n, v \in B_n(0,1)} (2\pi)^\beta |y'^\top v|^\beta e^{-\|y'\|^2 \pi} 2\pi e^{2\pi x^\top y'} \|y'\| \\ &\leq \sup_{y' \in \mathbb{R}^n} (2\pi)^{\beta+1} \|y'\|^{\beta+1} e^{-\|y'\|^2 \pi + \sqrt{2\pi} \|y'\|}. \end{aligned}$$

In the third line, $\{y'_1, \dots, y'_n\}$ are the components of y' . The final line follows by the Cauchy-Schwarz inequality, which shows that $|y'^\top v| \leq \|y'\| \|v\| \leq \|y'\|$ and $2\pi x^\top y' \leq 2\pi \|x\| \|y'\| \leq \sqrt{2\pi} \|y'\|$. Differentiating with respect to $\|y'\|$ and setting equal to zero shows that the supremum is achieved at $\|y'\| = (1 + \sqrt{5 + 4\beta}) / (2\sqrt{2\pi})$. We can substitute this back in to set

$$L_\beta := \left(1 + \sqrt{5 + 4\beta}\right)^{\beta+1} e^{(-1-2\beta+\sqrt{5+4\beta})/4} \left(\frac{2}{\pi}\right)^{-(\beta+1)/2}.$$

■

It follows directly that $K_1 - \varepsilon_1 \leq p_N \leq K_2 + \varepsilon_1$ with high probability. For large enough N , $K_1/2 \leq p_N \leq 2K_2$. In the next lemma, we prove that a corresponding result holds for

$\partial^\alpha(-\log p_N(x))$ (which is exactly the second condition from Theorem 5). The derivation is more technical but the intuition is based on the arguments in Section 2.3.3.

Lemma 13. $\partial^\alpha(-\log p_N(x)) \leq C_0$ for $x \in B_n(0, 1/\sqrt{2\pi})$, $|\alpha| \leq k$, and C_0 depending only on n and k .

Proof. Start by defining

$$q_N(x) = \frac{1}{N} \sum_{i=1}^N e^{-\|z-y_i\|^2\pi} e^{-2(x-z)^\top(z-y_i)\pi},$$

where z is the projection of x onto \mathcal{M} . Then,

$$\partial^\alpha(-\log p_N(x)) = \partial^\alpha\left(-\log\left(e^{-\|x-z\|^2\pi} q_N(x)\right)\right).$$

A result due to Nemirovski (2004) shows that

$$\sup_{|v_i| \leq 1} \left| \partial_{v_1} \dots \partial_{v_{|\alpha|}} F(x) \right| \leq \sup_{|v| \leq 1} |\partial_v^{|\alpha|} F(x)|$$

for C^k -smooth F , implying that we do not need to bound mixed partials.

It is straightforward to calculate $C_{0,1} := \sup \left| \partial^\alpha \left(-\log \left(e^{-\|x-z\|^2\pi} \right) \right) \right|$. (The supremum is also over $|\alpha| \leq k$). To get an upper bound for $\partial^\alpha(-\log q_N(x))$, we first write it as an expression involving powers of $q_N(x)$ and partials of $q_N(x)$. For example, if $\alpha = \{x_1, x_1, x_1, x_1\}$,

$$\begin{aligned} \frac{\partial^4(-\log q_N(x))}{\partial x_1^4} &= 6 \frac{\left(\frac{\partial q_N}{\partial x_1}\right)^4}{q_N(x)^4} - 12 \frac{\left(\frac{\partial q_N}{\partial x_1}\right)^2 \left(\frac{\partial^2 q_N}{\partial x_1^2}\right)}{q_N(x)^3} + 3 \frac{\left(\frac{\partial^2 q_N}{\partial x_1^2}\right)^2}{q_N(x)^2} \\ &\quad + 4 \frac{\left(\frac{\partial q_N}{\partial x_1}\right) \left(\frac{\partial^3 q_N}{\partial x_1^3}\right)}{q_N(x)^2} - \frac{\left(\frac{\partial^4 q_N}{\partial x_1^4}\right)}{q_N(x)}. \end{aligned}$$

Faa di Bruno's formula is an explicit representation of this expression; the number of terms and the coefficients depend on $|\alpha|$. We can find a suitable C_0 if we can calculate a lower bound for q_N and an upper bound for $|\partial_v^\beta q_N|$ where $\beta \leq |\alpha|$. The first bound follows from two previous lemmas. Lemma 11 shows that $\mathbb{E}p_N(x) \geq K_1$, and Lemma 12 shows that p_N is within ε_1 of its expectation with high probability. Since $q_N(x) \geq p_N(x)$ and ε_1 can be

made smaller than $K_1/2$, $q_N \geq K_1/2$ with high probability for N sufficiently large. K_1 is a function of d , and ε_1 is a function of n .

To bound the partials of q_N , we start off by using the second part of Lemma 12, which shows

$$|\partial_v^\beta q_N| \leq |\mathbb{E} \partial_v^\beta q_N| + \varepsilon_{1,\beta}.$$

Let z be the origin and let the first d coordinates lie in $T_z \mathcal{M}$. We can write the expectation as

$$\begin{aligned} |\mathbb{E} \partial_v^\beta q_N| &= \left| \int_{\mathcal{M}} \partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi} d\mu_{\mathcal{M}}(y') \right| \\ &\leq \left| \int_{\tilde{\mathcal{A}}_z} e^{2\pi x^\top y'} (2\pi)^\beta (y'^\top v)^\beta e^{-\|y'\|^2 \pi} d\mu_{\mathcal{M}}(y') \right| \\ &\quad + \left| \int_{\mathcal{M} \setminus \tilde{\mathcal{A}}_z} e^{2\pi x^\top y'} (2\pi)^\beta (y'^\top v)^\beta e^{-\|y'\|^2 \pi} d\mu_{\mathcal{M}}(y') \right|. \end{aligned}$$

We first bound the integral over $\mathcal{M} \setminus \tilde{\mathcal{A}}_z$. For $\hat{\tau}$ large enough, the local extrema of $\partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi}$ with respect to y' lie within $U_{\hat{\tau}}^z$. Since

$$\lim_{\|y'\| \rightarrow \infty} \partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi} = 0,$$

$\left| \partial_v^\beta \left(e^{2\pi x^\top y'} \right) e^{-\|y'\|^2 \pi} \right|$ is decreasing with increasing $\|y'\|$ in $\mathcal{M} \setminus \tilde{\mathcal{A}}_z$. The following holds, where $y_0 \in \partial \tilde{\mathcal{A}}_z$:

$$\begin{aligned} &\left| \int_{\mathcal{M} \setminus \tilde{\mathcal{A}}_z} e^{2\pi x^\top y'} (2\pi)^\beta (y'^\top v)^\beta e^{-\|y'\|^2 \pi} d\mu_{\mathcal{M}}(y') \right| \\ &\leq e^{2\pi x^\top y_0} (2\pi)^\beta (y_0^\top v)^\beta e^{-\|y_0\|^2 \pi} \int_{\mathcal{M} \setminus \tilde{\mathcal{A}}_z} d\mu_{\mathcal{M}}(y') \\ &\leq e^{2\pi \|x\| \|y_0\|} (2\pi)^\beta \|y_0\|^\beta \|v\|^\beta e^{-\|y_0\|^2 \pi} \\ &\leq (2\sqrt{2\pi\hat{\tau}})^\beta e^{-2\pi\hat{\tau}^2 + 2\sqrt{\pi\hat{\tau}}}. \end{aligned}$$

The integral over $\tilde{\mathcal{A}}_z$ can be bounded by relating it to the corresponding integral over \mathcal{A}_z . Let y be the projection of $y' \in \tilde{\mathcal{A}}_z$ onto \mathcal{A}_z . Then,

$$\left| \int_{\tilde{\mathcal{A}}_z} e^{2\pi x^\top y'} (2\pi)^\beta (y'^\top v)^\beta e^{-\|y'\|^2 \pi} d\mu_{\mathcal{M}}(y') \right|$$

$$\begin{aligned}
&= \left| \int_{\mathcal{A}_z} e^{2\pi x^\top y'} (2\pi)^\beta (y'^\top v)^\beta e^{-\|y'\|^2 \pi} p(y) d\mathcal{L}_d(y) \right| \\
&\leq \left| \int_{\mathcal{A}_z} e^{2\pi x^\top y} (y'^\top v)^\beta e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \right| \times N_f \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2} \right)^{d/2} \\
&\quad \times \sup_{x, y'} \left| e^{2\pi x^\top (y' - y)} \right| \times \sup_{y'} \left| e^{(\|y'\|^2 - \|y\|^2) \pi} \right| \times (2\pi)^\beta \\
&\leq \left| \int_{\mathcal{A}_z} (y'^\top v)^\beta e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \right| \times N_f \left(1 + \frac{C^2 \hat{\tau}^2}{\hat{\tau}^2} \right)^{d/2} \\
&\quad \times e^{2\sqrt{2\pi} \hat{\tau}^2 / (2\hat{\tau})} \times e^{\pi \hat{\tau}^4 / \hat{\tau}^2} \times (2\pi)^\beta.
\end{aligned}$$

The third line comes from relating $p(y)$ and N_f (Lemma 9) and bounding the change in the integrand due to projecting y' onto \mathcal{A}_z . We do not project $(y'^\top v)^\beta$ because it can equal zero. The fourth line follows by noting that $x^\top y = 0$ by orthogonality and that $\|y' - y\| \leq \hat{\tau}^2 / \hat{\tau}$ by Federer's reach condition. The reach condition also shows that $(y'^\top v)^\beta$ is a polynomial whose terms either lie in \mathbb{R}^d or have arbitrarily small coefficients. This can be used to bound the integral. Starting off by applying the triangle inequality for integrals and then the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\left| \int_{\mathcal{A}_z} (y'^\top v)^\beta e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \right| &\leq \int_{\mathcal{A}_z} \|y'\|^\beta \|v\|^\beta e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \\
&\leq \int_{\mathcal{A}_z} (\|y\| + \|y' - y\|)^\beta e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \\
&\leq \int_{\mathcal{A}_z} \|y\|^\beta e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \\
&\quad + \sum_{i=1}^{\beta} \binom{\beta}{i} \left(\frac{\hat{\tau}^2}{\hat{\tau}} \right)^i \int_{\mathcal{A}_z} \|y\|^{\beta-i} e^{-\|y\|^2 \pi} d\mathcal{L}_d(y) \\
&\leq 2 \int_{\mathcal{A}_z} \|y\|^\beta e^{-\|y\|^2 \pi} d\mathcal{L}_d(y).
\end{aligned}$$

The second line holds because $\|v\| \leq 1$ and $\|y'\|$ can be bounded using the triangle inequality. The third line follows after expanding $(\|y\| + \|y' - y\|)^\beta$, substituting in the bound for $\|y' - y\|$, and rearranging. Each term in the summation can be made arbitrarily small, which gives the fourth line. This integral is a function of the moments (of order β or less) of a d -dimensional Gaussian with covariance $1/(2\pi)I$. We can calculate it using spherical

coordinates, following the calculation of $\mathbb{E}[\|y\|^2]$ in Lemma 11. We have

$$\begin{aligned}
2 \int_{\mathcal{A}_z} \|y\|^\beta e^{-\|y\|^2\pi} d\mathcal{L}_d(y) &\leq 2 \int_{\mathbb{R}^d} \|y\|^\beta e^{-\|y\|^2\pi} d\mathcal{L}_d(y) \\
&= 2 \int_0^\infty r^{\beta+d-1} e^{-r^2\pi} dr \times \prod_{j=1}^{d-2} \int_0^\pi \sin^{d-j-1} \phi_j d\phi_j \\
&\quad \times \int_0^{2\pi} d\phi_{d-1} \\
&= \pi^{-(\beta+d)/2} \Gamma\left(\frac{\beta+d}{2}\right) \times \prod_{j=1}^{d-2} \frac{\sqrt{\pi} \Gamma((d-j)/2)}{\Gamma(1+(d-j-1)/2)} \times 2\pi \\
&= 2\pi^{-\beta/2} \frac{\Gamma((\beta+d)/2)}{\Gamma(d/2)} \\
&=: C_\beta.
\end{aligned}$$

Therefore, for large enough $\hat{\tau}$,

$$\begin{aligned}
|\mathbb{E}\partial_v^\beta q_N| &\leq 2(2\pi)^\beta N_f C_\beta + (2\sqrt{2\pi\hat{\tau}})^\beta e^{-2\pi\hat{\tau}^2+2\sqrt{\pi\hat{\tau}}} \\
&\leq 3(2\pi)^\beta N_f C_\beta.
\end{aligned}$$

If N is large enough, $\varepsilon_{1,\beta}$ will be smaller than $(2\pi)^\beta N_f C_\beta$ with high probability, implying that $|\partial_v^\beta q_N| \leq 4(2\pi)^\beta N_f C_\beta$. Let $\beta' \leq |\alpha|$ be the value of β for which this is maximized. Each term of $|\partial^\alpha(-\log q_N(x))|$ is bounded above in absolute value by a multiple of $4(2\pi)^{\beta'} N_f C_{\beta'}/(K_1/2)$ raised to a power less than or equal to $|\alpha|$. Therefore, using the triangle inequality and letting $|\alpha| \leq k$, $|\partial^\alpha(-\log q_N(x))| \leq C_{0,2}$, a constant. The factors of N_f in K_1 and $|\partial_v^\beta q_N|$ cancel each other out, so $C_{0,2}$ is a function of n and k . Setting $C_0 := C_{0,1} + C_{0,2}$ yields the lemma. \blacksquare

2.3.5 $-\log p_N(x) + \log N_f$ is an asdf

In the next two lemmas, we prove that the third condition of Theorem 5 holds. Recall from the proof of Lemma 11 that $\mathbb{E}p_N$ can be bounded above and below to within a multiplicative factor of $N_f e^{-\|x-z\|^2\pi}$. By taking logarithms and defining a suitable constant $\tilde{\rho}$, we show in Lemma 14 that the third condition holds for $-\log \mathbb{E}p_N + \log N_f$. In Lemma 15, we show that

this condition also holds for $-\log p_N + \log N_f$ as long as we modify $\tilde{\rho}$ to take into account the concentration bound from Lemma 12.

Lemma 14. For $x \in \mathcal{M}_{1/\sqrt{2\pi}}$ and z the projection of x onto \mathcal{M} ,

$$c_1(\|x - z\|^2\pi + \tilde{\rho}^2) \leq -\log \mathbb{E}p_N + \log N_f + \tilde{\rho}^2 \leq C_1(\|x - z\|^2\pi + \tilde{\rho}^2)$$

for $0 < \tilde{\rho} < c$, with c depending on C_0, c_1, C_1, k, n .

Proof. Let $x \in \mathcal{M}_{1/\sqrt{2\pi}}$ and let z be its projection onto \mathcal{M} . Then we can bound $\mathbb{E}p_N(x)$ by calculating the expectation separately over $\tilde{\mathcal{A}}_z$ and $\mathcal{M} \setminus \tilde{\mathcal{A}}_z$:

$$\begin{aligned} \mathbb{E}p_N(x) &= \int_{\mathcal{M}} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') \\ &= \int_{\tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') + \int_{\mathcal{M} \setminus \tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y'). \end{aligned}$$

The first term can be bounded as follows:

$$\begin{aligned} \int_{\tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') &\leq e^{C_f} \int_{\mathcal{A}_z} e^{-\|x-y\|^2\pi} N_f d\mathcal{L}_d(y) \\ &\leq e^{C_f} N_f e^{-\|x-z\|^2\pi} \int_{\mathcal{A}_z} e^{-\|z-y\|^2\pi} d\mathcal{L}_d(y) \\ &\leq e^{C_f} N_f e^{-\|x-z\|^2\pi}. \end{aligned}$$

Now consider $y' \in \mathcal{M} \setminus \tilde{\mathcal{A}}_z$. Since $\|z - y'\| \leq \|x - y'\| + \|x - z\|$ and $\|x - z\| \leq 1/\sqrt{2\pi} \leq \hat{\tau} \leq \|z - y'\|$, we have $(\|x - z\| - \hat{\tau})^2 \leq \|x - y'\|^2$. This gives us the following bound for the second term:

$$\int_{\mathcal{M} \setminus \tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2\pi} d\mu_{\mathcal{M}}(y') \leq e^{-(\|x-z\| - \hat{\tau})^2\pi}.$$

Thus, letting

$$C_{neg} := \frac{e^{-\hat{\tau}^2\pi + 2\hat{\tau}\pi\|x-z\|}}{N_f e^{C_f}},$$

we have

$$\mathbb{E}p_N(x) \leq e^{C_f} N_f e^{-\|x-z\|^2\pi} (1 + C_{neg})$$

$$\leq e^{C_f} N_f e^{-\|x-z\|^2 \pi} \left(\frac{1}{1 - C_{neg}} \right).$$

The second inequality holds because C_{neg} is arbitrarily small for small σ , so we can use the Taylor expansion

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i.$$

Next, we find a lower bound for $\mathbb{E}p_N(x)$ as in Lemma 11.

$$\begin{aligned} \mathbb{E}p_N(x) &\geq (1 - C_{neg}) \int_{\tilde{\mathcal{A}}_z} e^{-\|x-y'\|^2 \pi} d\mu_{\mathcal{M}}(y') \\ &\geq (1 - C_{neg}) e^{-C_f} \int_{\mathcal{A}_z} e^{-\|x-y\|^2 \pi} N_f d\mathcal{L}_d(y) \\ &\geq (1 - C_{neg}) e^{-C_f} N_f e^{-\|x-z\|^2 \pi} \int_{\mathcal{A}_z} e^{-\|z-y\|^2 \pi} d\mathcal{L}_d(y) \\ &\geq (1 - C_{neg}) e^{-C_f} N_f e^{-\|x-z\|^2 \pi} \left(1 - 2e^{-\left(\hat{\tau} - \sqrt{d/(2\pi)}\right)^2 \pi} \right). \end{aligned}$$

Now, let $\alpha := -\log(1 - C_{neg})$ and $\beta := -\log\left(1 - 2e^{-\left(\hat{\tau} - \sqrt{d/(2\pi)}\right)^2 \pi}\right)$. We have shown the following:

$$-C_f + \|x-z\|^2 \pi - \alpha \leq -\log \mathbb{E}p_N(x) + \log N_f \leq \alpha + C_f + \|x-z\|^2 \pi + \beta.$$

Since $\alpha, \beta > 0$, we can add $2\alpha + 2C_f + \beta$ to the left-hand side, $2\alpha + 2\beta + 2C_f$ to the middle, and $2\alpha + 2\beta + 2C_f + \frac{1}{2}\|x-z\|^2 \pi$ to the right-hand side while preserving these inequalities. Let $\tilde{\rho} := \sqrt{2(\alpha + \beta + C_f)}$, $c_1 := \frac{1}{2}$, and $C_1 := \frac{3}{2}$. Then, we have

$$c_1(\|x-z\|^2 \pi + \tilde{\rho}^2) \leq -\log \mathbb{E}p_N + \log N_f + \tilde{\rho}^2 \leq C_1(\|x-z\|^2 \pi + \tilde{\rho}^2).$$

■

Lemma 15. *With high probability, for $x \in \mathcal{M}_{1/\sqrt{2\pi}}$ and z the projection of x onto \mathcal{M} ,*

$$c_1(\|x-z\|^2 \pi + \rho^2) \leq -\log p_N + \log N_f + \rho^2 \leq C_1(\|x-z\|^2 \pi + \rho^2)$$

for $0 < \rho < c$, with c depending on C_0, c_1, C_1, k, n .

Proof. From Lemmas 11 and 12, w.h.p. for large enough N , $\inf_x p_N(x) \geq K_1 - \varepsilon_1$, or $\sup_x [-\log p_N(x)] \leq -\log(K_1 - \varepsilon_1)$. By uniform continuity,

$$\mathbb{P} \left[\sup_{f \in \mathcal{F}} |-\log p_N(x) + \log \mathbb{E} p_N(x)| \leq \varepsilon' \right] \geq 1 - \delta,$$

where $\varepsilon' := \varepsilon_1 / (K_1 - \varepsilon_1)$. Thus, the statement of the lemma follows from Lemma 14 with $\rho := \sqrt{2(\alpha + \beta + C_f + \varepsilon')}$, $c_1 := \frac{1}{2}$, and $C_1 := \frac{3}{2}$. \blacksquare

We have proven all the conditions necessary in order to show that $-\log p_N(x) + \log N_f$ is an asdf. We summarize this in the next theorem, which is the major result of this section. We also prove that the constants we have defined are small enough to apply Theorem 6 and state that \mathcal{M}_{put} is a manifold with desirable properties.

Theorem 16. *$-\log p_N(x) + \log N_f$ is an approximate squared-distance function that meets the conditions in Theorem 5. Consider the output set*

$$\mathcal{M}_{\text{put}} = \left\{ z \in \mathcal{M}_{\min(c_3, c_4)\sigma/\sqrt{2\pi}} \mid \Pi_{hi}(z) \partial F(z) = 0 \right\}$$

in the original coordinate system (i.e., the coordinates not scaled by $1/\sigma$, where σ is the bandwidth of the KDE). By Theorem 6, \mathcal{M}_{put} is a manifold whose reach is bounded below by $c\sigma$, where c is a constant depending on C_0, c_1, C_1, k, d , and n . \mathcal{M}_{put} converges to \mathcal{M} in Hausdorff distance for increasing N ; more specifically, $H(\mathcal{M}, \mathcal{M}_{\text{put}}) = O(\sigma^{5/4})$.

Proof. Since $\log N_f$ is a constant, the first two conditions from Theorem 5 hold by Lemmas 8 and 13. The third condition holds by Lemma 15. Thus, $-\log p_N(x) + \log N_f$ is an asdf.

For \mathcal{M}_{put} to be a manifold, $\sigma\rho$ must be sufficiently small with respect to $\min(c_3, c_4)\sigma/\sqrt{2\pi}$. We will show that for a small enough σ and large enough N , $\sqrt{2(\alpha + \beta + C_f + \varepsilon')}$ (our choice of ρ) can be made as small as needed. Recall that $\hat{\tau} = \tau/\sigma$ and $\hat{\tau} = \sigma^{-1/6}$. This implies

$$C_f = \frac{dC^2\sigma^{5/3}}{2\tau^2} + \left(\frac{\sigma^{4/3}}{\tau^2} + \frac{2\sqrt{2}\sqrt{\sigma}}{\tau} \right) \pi,$$

which can be made as small as desired. α and β can be bounded by using the fact that $-\log(1-p) < 2p$ if p is sufficiently small. For a small enough σ ,

$$e^{dC^2\sigma^{5/3}/(4\tau^2)} \leq e^{C_f},$$

which gives the bound

$$\begin{aligned}\alpha &\leq 2C_{neg} \\ &\leq \frac{2e^{-\sigma^{-1/3}\pi/2}}{e^{dC^2\sigma^{5/3}/(4\tau^2)}\sigma^d/(2V)} \\ &\leq 4V\sigma^{-d}e^{-\sigma^{-1/3}\pi/4}.\end{aligned}$$

Similarly,

$$\beta \leq 4e^{-\sigma^{-1/3}\pi/2}.$$

As σ tends to zero, so do these quantities. Finally, for large enough N , ε_1 is sufficiently small such that $\varepsilon' = \varepsilon_1/(K_1 - \varepsilon_1)$ is as small as necessary. For a small enough σ and a large enough N , $\rho = O(C_f^{1/2}) = O(\sigma^{1/4})$. This is sufficient to apply Theorem 6, which implies that \mathcal{M}_{put} is a manifold with bounded reach that converges to \mathcal{M} in Hausdorff distance. The Hausdorff distance $H(\mathcal{M}, \mathcal{M}_{\text{put}})$ is $O(\sigma\rho)$, which is $O(\sigma^{5/4})$. ■

2.4 Local Principal Components Analysis

A manifold \mathcal{M} can be approximated by a finite collection of tangent spaces centered at a sufficiently dense set of points sampled from \mathcal{M} . Fefferman et al. (2016) use this as motivation to define the concept of a cylinder packet; they also define a function $F^{\bar{\sigma}}$ and show that it is an asdf when coupled with a suitably constructed cylinder packet. In this section we show that we can estimate tangent spaces directly from the data to create a cylinder packet; this leads to the construction of an approximate squared-distance function that satisfies Theorems 5 and 6 and produces a putative manifold.

2.4.1 Definition of the asdf and Selection of the Bandwidth $\bar{\tau}$

Let $C_p := \{\text{cyl}_i\}$ be a collection of cylinders with centers $\{x_i\}$. Each cylinder is isometric to $\text{cyl} := \bar{\tau}(B_d \times B_{n-d})$. We choose $\bar{\tau}$ so that it tends to zero but remains large compared to the distance between a sample point and its nearest neighbors. Since we are assuming uniform support on \mathcal{M} , for large N we can choose $\bar{\tau}$ on the order of $N^{-1/(d+\varepsilon)}$ for a small value of ε .

Let U_i be a proper rotation of cyl_i , Tr_i a translation, and o_i a composition of a proper rotation and translation that moves the origin to x_i and rotates the d -dimensional cross-section of cyl_i to \mathbb{R}^d . Define

$$F^{\bar{o}}(z) := \frac{\sum_{\text{cyl}_i \ni z} \phi_{\text{cyl}_i}(o_i^{-1}(z)) \theta(\Pi_d(o_i^{-1}(z)) / (2\bar{\tau}))}{\sum_{\text{cyl}_i \ni z} \theta(\Pi_d(o_i^{-1}(z)) / (2\bar{\tau}))},$$

where $\text{cyl}_i \in C_p$, $z \in \bigcup \text{cyl}_i$, $\phi_{\text{cyl}_i}(z)$ is the squared distance from z to the d -dimensional cross-section of cyl_i and $\theta : \mathbb{R}^d \rightarrow [0, 1]$ is a bump function such that

1. $\theta(y) = 0$ for $\|y\| \notin (-1, 1)$
2. $\partial^\alpha \theta(y) = 0$ for $|\alpha| \leq k$ and $y = 0$ or $\|y\| \notin (-1, 1)$
3. $|\partial^\alpha \theta(y)| < C$, a controlled constant for all y
4. $\theta(y) = 1$ for $\|y\| < 1/4$.

Note that whether or not $F^{\bar{o}}(z)$ satisfies Theorem 5 depends on our choice of C_p ; for convenience, we refer to the pair $\{F^{\bar{o}}, C_p\}$ as a putative asdf. $F^{\bar{o}}(z)$ measures the squared distances ϕ_{cyl_i} to the central cross-section of each cylinder containing a given point z , and averages them using the bump function θ . Let $\hat{F}_z : B_n(0, 1) \rightarrow \mathbb{R}$ be a related function defined by $\hat{F}_z(w) = F^{\bar{o}}(z + \bar{\tau}\Theta(w)) / \bar{\tau}^2$, where Θ is an isometry that fixes the origin at z and identifies the first d coordinates with $T_z\mathcal{M}$. \hat{F}_z is essentially $F^{\bar{o}}$ analyzed in a coordinate system scaled up by $1/\bar{\tau}$. This is analogous to our analysis of the kernel density estimator in the previous section, where we scaled the coordinate system by $1/\sigma$.

2.4.2 Cylinder Packets

In order for $\{F^{\bar{o}}, C_p\}$ to be an asdf, C_p needs to be a cylinder packet, which is a collection of cylinders that satisfies the geometric constraints given below in Definition 17. These conditions ensure that a cylinder packet doesn't contain pairs of cylinders that overlap too

much or intersect at too great of an angle. This is motivated by our desire to estimate a manifold with bounded reach.

Definition 17 (Cylinder packet). *Let C_p be a collection of cylinders as above. C_p is a cylinder packet if it satisfies the following conditions:*

1. *The number of cylinders is less than or equal to a constant factor times $\frac{V}{\bar{\tau}^d}$.*
2. *Consider the set $S_i := \{\text{cyl}_{i_1}, \dots, \text{cyl}_{i_{|S_i|}}\}$ of cylinders that intersect cyl_i and perform the rigid-body motion o_i . For each cyl_{i_j} , there exists a translation Tr_{i_j} and a proper rotation U_{i_j} fixing x_{i_j} so that*
 - (a) *For $1 \leq j \leq |S_i|$, $Tr_{i_j}U_{i_j}\text{cyl}_{i_j}$ is a translation of cyl_i by a vector with norm at least $\bar{\tau}/3$.*
 - (b) *$\{x_i\} \cup \{Tr_{i_j}U_{i_j}x_{i_j} \mid 1 \leq j \leq |S_i|\} \cap \text{cyl}_i$ forms a $\bar{\tau}/2$ -net of the d -dimensional cross-section of cyl_i .*
 - (c) *For $1 \leq j \leq |S_i|$ and $v \in \mathbb{R}^n$, $\|v - U_{i_j}v\| < 2\frac{\bar{\tau}}{\tau}\|v - x_{i_j}\|$.*
 - (d) *For $1 \leq j \leq |S_i|$, $\|Tr_{i_j}(0)\| < \frac{\bar{\tau}^2}{\tau}$.*

In Lemmas 16 and 17 due to [Fefferman et al. \(2016\)](#), it is shown that \hat{F}_z satisfies Theorem 5 when C_p is a cylinder packet, meaning that $\{F^{\bar{o}}, C_p\}$ is an asdf. We include this towards the end of this section as Theorem 23 and provide a sketch of the proof.

In the next lemma we construct a collection of cylinders C_p^{Tan} whose central cross sections are derived from the tangent planes of the manifold and show that it is indeed a cylinder packet. The putative manifold actually has reach $c\bar{\tau}$, so the right-hand sides of conditions 2(c) and (d) in Definition 17 can be within a constant factor of what is given above.

Lemma 18. *First, construct a set $\{x_i\}$ of centers. Assume the sample size is large enough to contain a $\bar{\tau}/2$ -net of \mathcal{M} such that no two net points are within $\bar{\tau}/2.9$ of each other. Let C_p^{Tan} be the collection of cylinders with centers $\{x_i\}$ and central cross sections contained in $T_{x_i}\mathcal{M}$. Then, C_p^{Tan} is a cylinder packet; we call it an ideal cylinder packet.*

Proof. We show that the conditions in Definition 17 hold. Fix an x_i and consider the set $S_i := \{\text{cyl}_{i_1}, \dots, \text{cyl}_{i_{|S_i|}}\}$ of cylinders that intersect cyl_i . Perform the rigid-body motion o_i so that we are working in a convenient coordinate system. For each $x_{i_j} \in S_i$, define U_{i_j} as a rotation fixing x_{i_j} and rotating the central cross section of cyl_{i_j} so that it is parallel to $T_{x_i}\mathcal{M}$. Also, define Tr_{i_j} as the translation that subsequently moves the central cross section so that it lies in $T_{x_i}\mathcal{M}$. Lemma 4 implies the first condition.

Let p be the projection of x_{i_j} onto $T_x\mathcal{M}$. Federer's reach condition implies that

$$\|x_{i_j} - p\| \leq \frac{\|x_{i_j} - x_i\|^2}{2\tau}.$$

Condition 2(d) holds since the right hand side must be less than $4\bar{\tau}^2/\tau$. Since $\|x_{i_j} - x_i\| \geq \bar{\tau}/2.9$, we also have

$$\begin{aligned} \|x_i - p\| &\geq \sqrt{\|x_{i_j} - x_i\|^2 - \frac{\|x_{i_j} - x_i\|^4}{4\tau^2}} \\ &\geq \sqrt{\frac{100\bar{\tau}^2}{841} - \frac{2500\bar{\tau}^4}{707281\tau^2}} \\ &\approx \frac{10\bar{\tau}}{29} - \frac{125\bar{\tau}^2}{24389\tau^2} + O(\bar{\tau}^4), \end{aligned}$$

where the last line follows by a Taylor expansion. Since this can be made arbitrarily close to $\bar{\tau}/2.9$, $\|x_i - p\| \geq \bar{\tau}/3$ and Condition 2(a) is satisfied. Condition 2(b) follows from the fact that we started off with a $\bar{\tau}/2$ -net of the manifold; projecting $\{x_{i_1}, \dots, x_{i_{|S_i|}}\}$ onto $T_{x_i}\mathcal{M}$ contracts interpoint distances so we end up with a $\bar{\tau}/2$ -net of the tangent space.

To show the bound in 2(c), we need an expression for the angle between two nearby tangent spaces (in this case $T_{x_i}\mathcal{M}$ and $T_{x_{i_j}}\mathcal{M}$). In Lemma B.3 from a paper by [Boissonnat, Dyer, and Ghosh \(2013\)](#), it is shown that the sine of the largest principal angle θ_1 between $T_{x_i}\mathcal{M}$ and $T_{x_{i_j}}\mathcal{M}$ is less than or equal to $6\|x_i - x_{i_j}\|/\tau$, which is $12\sqrt{2}\bar{\tau}/\tau$ in our setup. Now, translate the origin to x_{i_j} and translate $T_{x_i}\mathcal{M}$ so that it contains x_{i_j} . Without loss of generality, let $v \in T_{x_{i_j}}\mathcal{M}$. Let $\{e_i\}_1^d$ and $\{\hat{e}_i\}_1^d$ be orthonormal bases for $T_{x_{i_j}}$ and T_{x_i} , respectively, so that the angle between e_i and \hat{e}_i is the principal angle θ_i . Define U_{i_j} as the

rotation that maps $\{e_i\}_1^d$ onto $\{\hat{e}_i\}_1^d$. Let $\{v_i\}_1^d$ be the components of v and $U_{i_j}v$ with respect to the appropriate bases. Then we have the following:

$$\begin{aligned} \|v - U_{i_j}v\| &\leq \left\| \sum_{i=1}^d v_i e_i - \sum_{i=1}^d v_i \hat{e}_i \right\| \\ &\leq \sum_{i=1}^d \|v_i e_i - v_i \hat{e}_i\| \\ &\leq \sum_{i=1}^d |v_i| \|e_i - \hat{e}_i\| \\ &\leq \|e_1 - \hat{e}_1\| \|v\|_1 \\ &\leq \|e_1 - \hat{e}_1\| \sqrt{d} \|v\|. \end{aligned}$$

Using the law of cosines, $\|e_i - \hat{e}_i\| \leq \sqrt{2 - 2 \cos \theta_i}$. From the bound on $\sin \theta_1$ and a Taylor expansion of $\cos \arcsin(12\sqrt{2}\bar{\tau}/\tau)$, we can show

$$\begin{aligned} 2 - 2 \cos \theta_1 &\leq \frac{288\bar{\tau}^2}{\tau^2} + O(\bar{\tau}^4) \\ &\leq \frac{576\bar{\tau}^2}{\tau^2} \end{aligned}$$

for large enough N . Thus,

$$\|v - U_{i_j}v\| \leq \frac{24\sqrt{d}\bar{\tau}}{\tau} \|v\|,$$

which shows 2(c). ■

Corollary 19. *First, construct a set $\{x_i\}$ of centers. Assume the sample size is large enough to contain a $\bar{\tau}/2$ -net of \mathcal{M} such that no two net points are within $\bar{\tau}/2.9$ of each other. The collection of cylinders with centers $\{x_i\}$ and central cross sections within $O(\bar{\tau}/\tau)$ of $T_{x_i}\mathcal{M}$ in operator norm is a cylinder packet; we call it an admissible cylinder packet.*

2.4.3 Constructing an Admissible Cylinder Packet with Local PCA

Usually we only have access to points sampled from the manifold and not their associated tangent spaces. It is easy to see that we can also construct a cylinder packet if we can

estimate the tangent spaces accurately enough (as stated in Corollary 19). Let \widehat{C}_p^{Tan} be a collection of cylinders constructed by using the same net points as in Lemma 18 and performing local PCA to estimate the d -dimensional cross-sections. In this section, we show that \widehat{C}_p^{Tan} is an admissible cylinder packet. The d -dimensional cross-sections are estimated as follows. Given a sample point $z \in \mathcal{M}$, we construct the PCA matrix $N_z^{-1}YY^\top$, where Y has columns consisting of the N_z sample points lying within $\bar{\tau}(B_d \times B_{n-d})$. (We are using a coordinate system centered at z whose first d coordinates lie in $T_z\mathcal{M}$). Using the eigenvectors of $N_z^{-1}YY^\top$, we can get an estimate $\widehat{T}_z\mathcal{M}$ of the tangent space at z . We show that this is close to the true tangent space by using the Davis-Kahan $\sin \theta$ theorem. The version stated below is due to Yu, Wang, and Samworth (2015).

Let $\|\cdot\|_F$ denote the Frobenius norm of a matrix. Suppose $V, \widehat{V} \in \mathbb{R}^{n \times d}$ both have orthonormal columns. Theorem 20 gives an upper bound on $\left\|\sin \theta(\widehat{V}, V)\right\|_F$, where $\theta(\widehat{V}, V)$ is the $d \times d$ diagonal matrix whose diagonal consists of the principal angles between the column spaces of V and \widehat{V} and $\sin \theta(\widehat{V}, V)$ is defined entrywise. The principal angles are given by $\{\cos^{-1} \zeta_1, \dots, \cos^{-1} \zeta_d\}$, where $\{\zeta_1, \dots, \zeta_d\}$ are the singular values of $\widehat{V}^T V$.

Theorem 20 (Davis-Kahan **sin θ** Theorem). *Let $\Lambda, \widehat{\Lambda} \in \mathbb{R}^{n \times n}$ be symmetric, with eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$ and $\widehat{\lambda}_1 \geq \dots \geq \widehat{\lambda}_n$ respectively. Let $1 \leq d \leq n$ and assume $\lambda_d - \lambda_{d+1} > 0$. Let $V = (v_1, \dots, v_d) \in \mathbb{R}^{n \times d}$ and $\widehat{V} = (\widehat{v}_1, \dots, \widehat{v}_d) \in \mathbb{R}^{n \times d}$ have orthonormal columns satisfying $\Lambda v_j = \lambda_j v_j$ and $\widehat{\Lambda} \widehat{v}_j = \widehat{\lambda}_j \widehat{v}_j$ for $j = 1, \dots, d$. Then*

$$\left\|\sin \theta(\widehat{V}, V)\right\|_F \leq \frac{2\left\|\widehat{\Lambda} - \Lambda\right\|_F}{\lambda_d - \lambda_{d+1}}.$$

We also make use of the following concentration inequality due to Ahlswede and Winter (2002). Let $A \leq B$ mean that $A - B$ is positive semidefinite.

Theorem 21. *Let a_1, \dots, a_k be i.i.d. random positive semidefinite $d \times d$ matrices with expected value $\mathbb{E}[a_i] = M \geq \mu I$ and $a_i \leq I$. Then for all $\epsilon \in [0, 1/2]$,*

$$\mathbb{P}\left[\frac{1}{k} \sum_{i=1}^k a_i \notin [(1 - \epsilon)M, (1 + \epsilon)M]\right] \leq 2D \exp\left\{\frac{-\epsilon^2 \mu k}{2 \ln 2}\right\}.$$

We now prove the key result of this section.

Theorem 22. *Let z be a point sampled from \mathcal{M} . Translate z to the origin, and let the first d coordinates lie in $T_z\mathcal{M}$. Let Y be a matrix whose columns consist of the N_z sample points $\{y_i\}$ lying within $\bar{\tau}(B_d \times B_{n-d})$. Let X be a matrix whose columns are the projections $\{x_i\}$ of $\{y_i\}$ onto $T_z\mathcal{M}$. Construct the matrices $N_z^{-1}XX^\top$ and $N_z^{-1}YY^\top$, and let V and \hat{V} be their respective matrices of eigenvectors. Then, w.h.p.,*

$$\left\| \sin \theta(\hat{V}, V) \right\|_F \leq \frac{\left(\frac{2\bar{\tau}^3}{\tau} + \frac{2\bar{\tau}^4}{\tau^2} \right) (d+2)}{(1-\epsilon)\bar{\tau}^2 \left(1 + \frac{C^2\bar{\tau}^2}{\tau^2} \right)^{-d/2}},$$

where $\epsilon \in [0, 1/2]$.

Proof. Clearly, N_z increases with N . We can assume the matrices in the statement of the theorem can be defined. We apply Theorem 20 with $N_z^{-1}XX^\top$ and $N_z^{-1}YY^\top$ corresponding to Λ and $\hat{\Lambda}$, respectively.

We start off by bounding the numerator $\|N_z^{-1}(YY^\top - XX^\top)\|_F$. This is easiest if we consider Y as a perturbation of X by the matrix P since we can control P using Federer's reach condition. This gives:

$$\begin{aligned} YY^\top &= (X + P)(X + P)^\top \\ &= XX^\top + XP^\top + PX^\top + PP^\top. \end{aligned}$$

Therefore,

$$\begin{aligned} \|N_z^{-1}(YY^\top - XX^\top)\|_F &\leq N_z^{-1} \|XP^\top + PX^\top + PP^\top\|_F \\ &\leq N_z^{-1} (\|XP^\top\|_F + \|PX^\top\|_F + \|PP^\top\|_F) \\ &\leq N_z^{-1} (\|X\|_F \|P^\top\|_F + \|P\|_F \|X^\top\|_F + \|P\|_F \|P^\top\|_F). \end{aligned}$$

Because each column of X has norm less than or equal to $\bar{\tau}$, $\|X\| \leq \sqrt{N_z}\bar{\tau}$. By Federer's reach condition, we have

$$|y_i - x_i| \leq \frac{|z - y_i|^2}{2\tau}$$

$$\leq \frac{(\sqrt{2\bar{\tau}})^2}{2\tau},$$

which implies that $\|P\| \leq \sqrt{N_z \bar{\tau}^4 / (\tau^2)}$. Thus,

$$\|N_z^{-1}(YY^\top - XX^\top)\|_F \leq \frac{\bar{\tau}^3}{\tau} + \frac{\bar{\tau}^4}{\tau^2}.$$

Now we need to bound $\lambda_d - \lambda_{d+1}$. Let $\lambda_1 \geq \dots \geq \lambda_n$ be the eigenvalues of $N_z^{-1}XX^\top$, and let $\mu_1 \geq \dots \mu_n$ be the eigenvalues of $M := \mathbb{E}[N_z^{-1}XX^\top]$. We see that $\lambda_{d+1} = \dots = \lambda_n = \mu_{d+1} = \dots = \mu_n = 0$. So, we only need a lower bound for λ_d , which we can get by relating its value to μ_d through a concentration inequality. Assuming the first d coordinates are aligned with the eigenvectors of M , μ_d is the variance in the direction x_d . M is the population covariance matrix of the probability measure \mathcal{P} on $T_z\mathcal{M} \cap B_d(\bar{\tau})$ that is the pushforward of the uniform measure on $\mathcal{M} \cap \bar{\tau}(B_d \times B_{n-d})$. From Lemma 9, we know \mathcal{P} has a density $p(x)$ that is greater than or equal to $(1 + C^2\bar{\tau}^2/\tau^2)^{-d/2}$ multiplied by a normalizing factor, which in this case is just $\text{Vol}(B_d(\bar{\tau}))^{-1}$.

We have the following bound for μ_d :

$$\begin{aligned} \mu_d &= \int_{B_d(\bar{\tau})} x_d^2 d\mathcal{P}(x) \\ &\geq \int_{B_d(\bar{\tau})} x_d^2 \frac{\left(1 + \frac{C^2\bar{\tau}^2}{\tau^2}\right)^{-d/2}}{\text{Vol}(B_d(\bar{\tau}))} d\mathcal{L}_d(x) \\ &\geq \int_0^{\bar{\tau}} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \left(r \sin \phi_{d-1} \prod_{j=1}^{d-2} \sin \phi_j \right)^2 \frac{\left(1 + \frac{C^2\bar{\tau}^2}{\tau^2}\right)^{-d/2}}{\text{Vol}(B_d(\bar{\tau}))} dV \\ &= \frac{\Gamma(d/2 + 1) \left(1 + \frac{C^2\bar{\tau}^2}{\tau^2}\right)^{-d/2}}{\pi^{d/2} \bar{\tau}^d} \int_0^{\bar{\tau}} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} r^{d+1} \prod_{j=1}^{d-1} \sin^{d-j+1} \phi_j dr \prod_{j=1}^{d-1} d\phi_j. \end{aligned}$$

The third line follows by a change of coordinates. Substitute

$$\left\{ x_1 \mapsto r \cos \phi_1, x_{2 \leq i \leq d-1} \mapsto r \cos \phi_i \prod_{j=1}^{i-1} \sin \phi_j, x_d \mapsto r \sin \phi_{d-1} \prod_{j=1}^{d-2} \sin \phi_j \right\},$$

and let

$$dV := r^{d-1} \prod_{j=1}^{d-2} \sin^{d-j-1} \phi_j dr d\phi_1 \dots d\phi_{d-1}.$$

The integral in the fourth line can be evaluated by noting that $\int_0^{\bar{\tau}} r^{d+1} dr = \bar{\tau}^{d+2} / (d+2)$, $\int_0^{2\pi} \sin^2 \phi_{d-1} d\phi_{d-1} = \pi$, and $\int_0^\pi \sin^{d-j+1} \phi_j d\phi_j = \frac{\sqrt{\pi} \Gamma((d-j+2)/2)}{\Gamma(1+(d-j+1)/2)}$ for $1 \leq j \leq d-2$. We can simplify (by telescoping)

$$\prod_{j=1}^{d-2} \frac{\sqrt{\pi} \Gamma((d-j+2)/2)}{\Gamma(1+(d-j+1)/2)} = \frac{\pi^{(d-2)/2}}{\Gamma(1+d/2)}.$$

Therefore,

$$\mu_d \geq \frac{1}{d+2} \bar{\tau}^2 \left(1 + \frac{C^2 \bar{\tau}^2}{\tau^2}\right)^{-d/2}.$$

$N_z^{-1} X X^\top$ and M are zero outside the upper left $d \times d$ block. Call their nonzero blocks Ξ and $\tilde{\Xi}$, respectively; clearly these matrices have eigenvalues $\{\lambda_i\}_1^d$ and $\{\mu_i\}_1^d$. Ξ can be written as the empirical average $N_z^{-1} \sum_{i=1}^{N_z} x_{i,d} x_{i,d}^\top$, where the $\{x_{i,d}\}$ are the first d coordinates of the $\{x_i\}$. Note that $x_{i,d} x_{i,d}^\top \leq \|x_{i,d} x_{i,d}^\top\|_F I \leq \bar{\tau}^2 I$. For N_z large enough, this implies $x_{i,d} x_{i,d}^\top \leq I$. Additionally, since $\mu_d > 0$ is the smallest eigenvalue of $\tilde{\Xi}$, we have $\tilde{\Xi} \geq \mu_d I$. This is sufficient to apply Theorem 21. So, for all $\epsilon \in [0, 1/2]$,

$$\mathbb{P}\left[\Xi \notin \left[(1-\epsilon)\tilde{\Xi}, (1+\epsilon)\tilde{\Xi}\right]\right] \leq 2d \exp\left\{-\frac{\epsilon^2 \mu_d N_z}{2 \log 2}\right\}.$$

The matrix interval is in terms of the positive semidefinite ordering, so $\Xi \geq (1-\epsilon)\tilde{\Xi}$ w.h.p. This implies $\lambda_d \geq (1-\epsilon)\mu_d$.

Now, applying Theorem 20,

$$\begin{aligned} \left\|\sin \theta(\hat{V}, V)\right\|_F &\leq \frac{2\|N_z^{-1}(Y Y^\top - X X^\top)\|_F}{\lambda_d} \\ &\leq \frac{\left(\frac{2\bar{\tau}^3}{\tau} + \frac{2\bar{\tau}^4}{\tau^2}\right)(d+2)}{(1-\epsilon)\bar{\tau}^2\left(1 + \frac{C^2 \bar{\tau}^2}{\tau^2}\right)^{-d/2}} \\ &= O\left(\frac{\bar{\tau}}{\tau}\right). \end{aligned}$$

■

2.4.4 $\{F^{\bar{o}}(z), C_p^{\widehat{Tan}}\}$ is an asdf

In Theorem 23, we sketch a proof that $\{F^{\bar{o}}, C_p\}$ is an asdf for an arbitrary cylinder packet C_p ; we also show that Theorem 6 applies. Since we showed that $C_p^{\widehat{Tan}}$ is an admissible cylinder packet, it follows immediately that $\{F^{\bar{o}}(z), C_p^{\widehat{Tan}}\}$ is an asdf.

Theorem 23. *Assume that we are given a cylinder packet C_p . $\{F^{\bar{o}}, C_p\}$ is an approximate squared-distance function that meets the conditions in Theorem 5. Furthermore, by Theorem 6, the output set (in the original coordinate system)*

$$\mathcal{M}_{\text{put}} = \{z \in \mathcal{M}_{\min(c_3, c_4)\bar{\tau}} \mid \Pi_{hi}(z)\partial F(z) = 0\}.$$

is a manifold whose reach is bounded below by $c\bar{\tau}$, where c is a constant depending on C_0, c_1, C_1, k, d , and n . \mathcal{M}_{put} converges to \mathcal{M} in Hausdorff distance for increasing N ; more specifically, $H(\mathcal{M}, \mathcal{M}_{\text{put}}) = O(\bar{\tau}^2)$.

Proof. \widehat{F}_z is C^k -smooth by the chain rule and the smoothness of the projection, distance, and bump functions. $\partial^\alpha(\widehat{F}_z(w)) \leq C_0$ for $w \in B_n(0, 1)$, $|\alpha| \leq k$, and C_0 depending only on n and k . This is true by the chain rule since the bounds on the derivatives of the bump function and the distance function can be directly calculated. After rescaling by $\bar{\tau}$, these depend only on n and k .

The third condition is satisfied by setting ρ equal to $c_\rho\bar{\tau}/\tau$, where c_ρ is a constant depending on the geometry of C_p . Let $z' := z + \bar{\tau}\Theta(w)$. If C_p is a cylinder packet, the distance from $\Pi_{\mathcal{M}}z'$ to the central cross-section of any cylinder containing z' is on the order of $\bar{\tau}^2/\tau$. $\widehat{F}_z(w)$ is a rescaled convex combination of the squared distance between z' and the central cross section of the cylinders containing it. That is, $\widehat{F}_z(w)$ is essentially $\bar{\tau}^{-2} \sum b_i (\|\Pi_{\mathcal{M}}z' - z'\| + c_i\bar{\tau}^2/\tau)^2$, where $\sum b_i = 1$ and the c_i depend on C_p . Thus, setting ρ^2 to $c_\rho^2\bar{\tau}^2/\tau^2$ satisfies the third condition of Theorem 5 for appropriate values of c_1 and C_1 :

$$c_1 \left(\frac{\|\Pi_{\mathcal{M}}z' - z'\|^2}{\bar{\tau}^2} + \rho^2 \right) \leq \widehat{F}_z(w) + \rho^2 \leq C_1 \left(\frac{\|\Pi_{\mathcal{M}}z' - z'\|^2}{\bar{\tau}^2} + \rho^2 \right),$$

where $0 < \rho < c$, with c depending on C_0, c_1, C_1, k, n .

For Theorem 6 to apply, $\bar{\tau}\rho$ must be sufficiently small with respect to $\min(c_3, c_4)\bar{\tau}$. This is clearly true because $\bar{\tau}\rho/(\min(c_3, c_4)\bar{\tau}) = O(\bar{\tau}/\tau)$, which can be made as small as desired. The Hausdorff distance $H(\mathcal{M}, \mathcal{M}_{\text{put}})$ is $O(\bar{\tau}\rho)$, which is $O(\bar{\tau}^2)$. ■

Theorem 24. $\{F^{\bar{\sigma}}(z), C_p^{\widehat{Tan}}\}$, where $C_p^{\widehat{Tan}}$ is a cylinder packet constructed using local PCA is an approximate squared-distance function that meets the conditions in Theorem 5. Furthermore, by Theorem 6, the output set

$$\mathcal{M}_{\text{put}} = \{z \in \mathcal{M}_{\min(c_3, c_4)\bar{\tau}} \mid \Pi_{hi}(z)\partial F(z) = 0\}$$

is a manifold whose reach is bounded below by $c\bar{\tau}$, where c is a constant depending on C_0, c_1, C_1, k, d , and n . \mathcal{M}_{put} converges to \mathcal{M} in Hausdorff distance for increasing N ; more specifically, $H(\mathcal{M}, \mathcal{M}_{\text{put}}) = O(\bar{\tau}^2)$.

Proof. This is a direct consequence of Theorem 23, Corollary 19, and Theorem 22. ■

2.5 Simulations

In this section, we present simulation results showing that the two asdfs considered in this chapter can be used to find a discretized version of a putative manifold. All simulations were performed using the following gradient descent algorithm based on subspace-constrained mean shift (Ozertem and Erdogmus, 2011).

1. Initialize a mesh of points on which to perform gradient descent. They can be sample points with or without added noise.
2. Perform the following for each mesh point x :
 - (a) Calculate the gradient g and the Hessian H of the asdf f .
 - (b) Let V be a matrix whose columns are the eigenvectors corresponding to the largest $n - d$ eigenvalues of H .

- (c) Calculate $VV^\top g$ and take a step in this direction.
- (d) Go to step (a) until a tolerance condition is met.

We applied this algorithm to data points sampled from three different manifolds contained in the unit ball of a Euclidean space: a circle embedded in \mathbb{R}^2 , a closed curve embedded in \mathbb{R}^3 , and a sphere embedded in \mathbb{R}^3 . We sampled 1000 points from each manifold and used this data to construct asdfs based on KDE and local PCA. We then sampled 1000 additional points and added Gaussian noise with a standard deviation of 0.05; these were used as the starting mesh points. Finally, we ran the algorithm and took the final output to be points lying on the putative manifold. Figure 2.1 shows an example of each of the three manifolds for each asdf. To get a sense of the accuracy of this procedure, we found the RMS distance of each putative manifold to a 10000 point sample (i.e., an approximate net) derived from the original manifolds. The average RMS distance from 100 trials is given in Table 2.1.

	Circle $\subset \mathbb{R}^2$	Curve $\subset \mathbb{R}^3$	Sphere $\subset \mathbb{R}^3$
KDE	0.000433	0.000990	0.00221
Local PCA	0.000146	0.000453	0.000603

Table 2.1: Average RMS distance for subspace-constrained gradient descent on two asdfs

2.6 Discussion

In this chapter, we showed that if we are provided with data sampled from a manifold \mathcal{M} , we can use two different asdfs to construct an estimator of \mathcal{M} . The asdfs are based on kernel density estimation and local PCA, which are conceptually easy to understand and mainstays of nonparametric estimation. The estimator is a manifold itself, and there are concrete bounds on its geometry (for example, its reach). These bounds are derived from an application of the implicit function theorem and are given in a key theorem of [Fefferman](#),

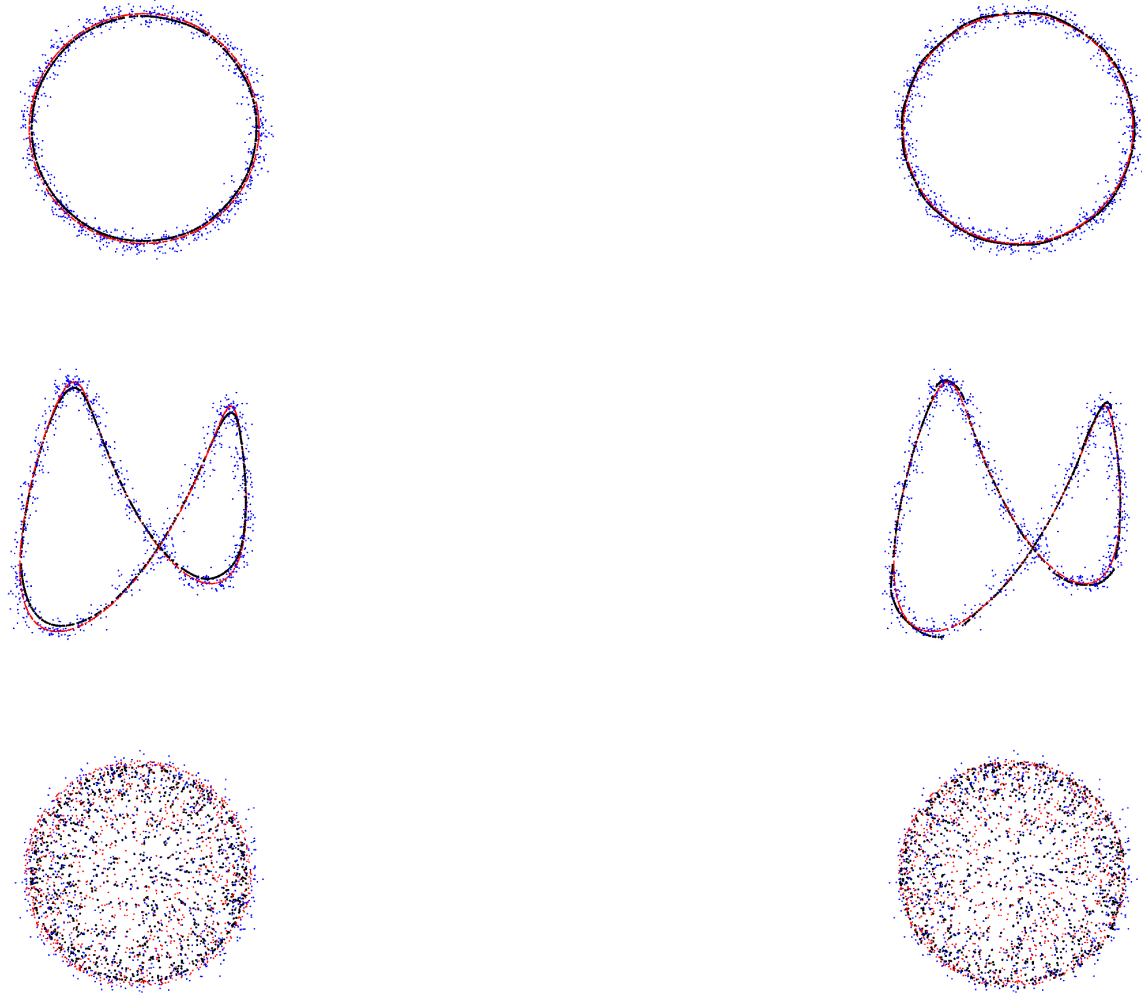


Figure 2.1: Circle $\subset \mathbb{R}^2$, curve $\subset \mathbb{R}^3$, and sphere $\subset \mathbb{R}^3$ examples (top to bottom). Red points are sampled from the manifold, blue points are sampled points with added Gaussian noise (sd = 0.05), and black points are the output points obtained by subspace-constrained gradient descent. The left column uses an asdf based on KDE, and the right column uses an asdf based on local PCA.

[Mitter, and Narayanan \(2016\)](#). Our contribution in this chapter was to create asdfs that can be calculated directly from the data as well as to give bounds on the reach and Hausdorff distance that depend on the sample size and properties of the asdfs. In the future, we aim to work on several natural extensions of our results. It remains to be seen what can be said about an estimator derived from a sample contaminated with noise (potentially bounded or sub-Gaussian). Additionally, it would be of theoretical interest to see how precise we can make the constants in our theorems, including the constants derived from the work of [Fefferman et al. \(2016\)](#).

Chapter 3

$C^{1,1}(\mathcal{M})$ REGRESSION VIA ESTIMATION OF CHARTS AND LOCAL $C^{1,1}(\mathbb{R}^d)$ REGRESSION

3.1 Introduction

In this chapter, we are concerned with the regression of functions from the class $C^{1,1}(\mathcal{M})$; these are real-valued functions defined on a submanifold $\mathcal{M} \subset \mathbb{R}^n$ that are differentiable and have a Lipschitz (Riemannian) gradient. This space of functions can be equipped with a norm $\|\cdot\|_{C^{1,1}(\mathcal{M})}$ defined as the maximum of the Lipschitz constant of the gradient and the supremum of the Euclidean norm of the gradient and the modulus of the function values. We make these notions more precise in Sections 3.3 and 3.4. When \mathcal{M} is a Euclidean space, this class is written as, for example, $C^{1,1}(\mathbb{R}^d)$, and \mathcal{M} may or may not be embedded in a higher-dimensional Euclidean space.

We place a few regularity conditions on \mathcal{M} itself. We assume that it is a closed and connected submanifold of the unit ball of \mathbb{R}^n that is of smoothness class C^2 and has dimension d , reach at least τ , and volume at most V . The smoothness class and dimension are defined at the beginning of Section 3.3. The reach is a quantity that measures the curvature of a submanifold, and it is defined in Section 3.3.3. The volume is measured through the Hausdorff measure $\mathcal{H}^d(\mathcal{M})$, which is a generalization of the Lebesgue measure to lower-dimensional subsets of a Euclidean space; we provide a more precise definition in Section 3.3.2.

As is standard in regression problems, we assume that we have access to a sample generated by an underlying function from the class of interest. Our model is as follows. We are provided with a sample $\mathcal{X} \times \mathcal{Y} \subset \mathcal{M} \times \mathbb{R}$ consisting of N points on \mathcal{M} and a real-valued observation associated with each point. We assume that \mathcal{X} is drawn from a probability

measure \mathcal{P} supported on \mathcal{M} , where \mathcal{P} is absolutely continuous with respect to $\mathcal{H}^d(\mathcal{M})$ and has density p bounded away from zero and bounded above. For each $x_i \in \mathcal{X}$, we also have $y_i \in \mathcal{Y}$ generated by a $C^{1,1}(\mathcal{M})$ function and observed with additive Gaussian noise. That is, $y_i = f^*(x_i) + \xi_i$, where $f^* \in C^{1,1}(\mathcal{M})$, $\|f^*\|_{C^{1,1}(\mathcal{M})} = M^*$, and $\xi_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.

Our objective is to construct an algorithm that finds an approximate empirical risk minimizer from a particular class of functions defined more technically in Section 3.6. This class consists of functions whose pullbacks to tangent spaces of \mathcal{M} indexed by a suitably fine net have $C^{1,1}(\mathbb{R}^d)$ norms bounded by an increasing function of N . We derive risk bounds for this procedure and show that it recovers f^* for $N \rightarrow \infty$.

In a paper due to [Gustafson et al. \(2018\)](#), co-authored by us, an algorithm for $C^{1,1}(\mathbb{R}^d)$ regression is constructed and analyzed. Here, we extend these results to the case where \mathcal{M} is not an open set in a Euclidean space, and we show that the sample complexity of the algorithm also relies on geometric properties of \mathcal{M} . A brief outline of our algorithm is as follows. We start by finding a net of \mathcal{M} consisting of sample points, and we use local principal components analysis to estimate the tangent spaces to \mathcal{M} at the net points. Then, for each net point, we find an associated local estimator for f^* by projecting the sample points onto the estimated tangent space and performing $C^{1,1}(\mathbb{R}^d)$ regression within a ball whose radius is chosen as a sufficiently small function of the reach. The final estimator is formed by patching together the local estimators using a partition of unity.

3.1.1 Notation

In this section, we collect notational conventions; those specifically pertaining to manifolds are deferred until Section 3.3.

Let $\mathcal{L}(\mathbb{R}^{m'}, \mathbb{R}^m)$ be the space of $m \times m'$ real-valued matrices. Let $M \in \mathcal{L}(\mathbb{R}^{m'}, \mathbb{R}^m)$. The element of M in the i^{th} row and j^{th} column is written as M_{ij} . The i^{th} column of M is written as M_i . We equip $\mathcal{L}(\mathbb{R}^{m'}, \mathbb{R}^m)$ with $\|\cdot\|_2$ and $\|\cdot\|_F$, the operator and Frobenius norms, respectively. They are defined as follows: for $v \in \mathbb{R}^{m'}$, $\|M\|_2 := \sup_{\|v\|=1} \|Mv\|$; additionally, $\|M\|_F := \sqrt{\text{Tr}(MM^\top)}$, where the trace $\text{Tr} : \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m) \rightarrow \mathbb{R}$ assigns $M' \mapsto \sum M'_{ii}$. We

write $\text{id} \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)$ as I or I_m . The singular-value decomposition allows us to write $M = UDV^\top$, where $U^\top U = I_m$, $V^\top V = I_{m'}$, and $D \in \mathcal{L}(\mathbb{R}^{m'}, \mathbb{R}^m)$ has nonnegative real numbers (called the singular values of M) on the diagonal and zeros everywhere else. The columns of U and V are called left and right singular vectors, respectively. Let $\sigma_i : \mathcal{L}(\mathbb{R}^{m'}, \mathbb{R}^m) \rightarrow \mathbb{R}$ be the i^{th} singular value, where $\sigma_1 \geq \dots \geq \sigma_{\min(m, m')}$. The singular values are the square roots of the eigenvalues of $M^\top M$. We define $\lambda_i : \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m) \rightarrow \mathbb{C}$ to be the i^{th} eigenvalue, where $|\lambda_1| \geq \dots \geq |\lambda_m|$.

Given a d -dimensional linear subspace $M \subset \mathbb{R}^n$, we write $V_M \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$ for a matrix whose columns form an orthonormal basis of M ; $V_M V_M^\top$ is the linear map that orthogonally projects vectors from \mathbb{R}^n onto M . We also denote the projection by Π_M . The orthogonal complement M^\perp is the $(n-d)$ -dimensional subspace of \mathbb{R}^n consisting of the vectors orthogonal to every vector in M . The projection onto M^\perp is written as Π_M^\perp .

We define the principal angles between two d -dimensional linear subspaces $M, N \subset \mathbb{R}^n$ as $\angle_i(M, N) := \arccos \sigma_i(V_M^\top V_N)$; we also write the largest principal angle as $\angle(M, N)$. Write $\Theta(M, N) \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)$ (interchangeably, $\Theta(V_M, V_N)$) for the diagonal matrix whose elements are $\Theta(M, N)_{ii} := \angle_i(M, N)$. They can also be defined recursively as

$$\cos \angle_i(M, N) := \sup_{u_i \in M} \sup_{v_i \in N} u_i^\top v_i,$$

where $\|u_i\| = 1$, $\|v_i\| = 1$, $u_i \perp u_j, j < i$, and $v_i \perp v_j, j < i$. Then, $\{u_i\}$ and $\{v_i\}$ are called principal vectors and form orthonormal bases for M and N , respectively. The nonzero principal angles of M^\perp and N^\perp are equal to $\angle_i(M, N)$.

We use C_d to denote any constant that is either absolute or depends solely on d . If we need to differentiate between multiple such constants, e.g., in the statement of a theorem, we will write them as $C_{d,1}, C_{d,2}$, and so on.

The indicator function of a set A is written as $\mathbb{1}_A$ or $\mathbb{1}\{A\}$. It assigns $x \mapsto 1$ if $x \in A$ and $x \mapsto 0$ otherwise.

3.2 $C^{1,1}(\mathbb{R}^d)$ Regression

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$ is in the class $C^{1,1}(\mathbb{R}^d, \mathbb{R}^n)$ if it is differentiable and has a Lipschitz gradient. We place the following norm on this class:

$$\|f\|_{C^{1,1}(\mathbb{R}^d, \mathbb{R}^n)} := \max \left\{ \sup_{x \in \mathbb{R}^d} \|f(x)\|, \sup_{x \in \mathbb{R}^d} \|Df|_x\|_F, \text{Lip}(Df) \right\}.$$

The maximum is over a norm and two seminorms, respectively, which can also be written as $\|f\|_{C^0(\mathbb{R}^d, \mathbb{R}^n)}$, $\|f\|_{\dot{C}^1(\mathbb{R}^d, \mathbb{R}^n)}$, and $\|f\|_{\dot{C}^{1,1}(\mathbb{R}^d, \mathbb{R}^n)}$. This notation is due to $C^0(\mathbb{R}^d, \mathbb{R}^n)$ being the class of continuous maps from \mathbb{R}^d to \mathbb{R}^n and $\dot{C}^1(\mathbb{R}^d, \mathbb{R}^n)$ the class of continuously differentiable maps. This norm is undefined for $f \notin C^{1,1}(\mathbb{R}^d, \mathbb{R}^n)$; in general, if we write $\|f\|_{C^{1,1}(\mathbb{R}^d, \mathbb{R}^n)}$, it is to be assumed that $f \in C^{1,1}(\mathbb{R}^d, \mathbb{R}^n)$. When f is real-valued, the codomain is not specified; e.g., we write $f \in C^{1,1}(\mathbb{R}^d)$.

[Gustafson et al. \(2018\)](#) consider the problem of recovering a $C^{1,1}(\mathbb{R}^d)$ function from noisy observations. They assume access to a sample $\mathcal{X}' \times \mathcal{Y}' \subset \mathbb{R}^d \times \mathbb{R}$, where \mathcal{X}' consists of N' points in $B_d(0, 1) \subset \mathbb{R}^d$ and \mathcal{Y}' of corresponding real-valued observations. \mathcal{X}' is drawn from a probability measure \mathcal{P}' , which is absolutely continuous with respect to \mathcal{L}^d and has density p' bounded from above and bounded away from zero. For each $x'_i \in \mathcal{X}'$, there is also an observation $y'_i \in \mathcal{Y}'$ such that $y'_i = f^{*'}(x'_i) + \xi'_i$, where $f^{*'} \in C^{1,1}(\mathbb{R}^d)$, $\|f^{*'}\|_{C^{1,1}(\mathbb{R}^d)} = M^{*'}$, and $\xi'_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$. A structural risk minimization approach is adopted to recover $f^{*'}$ —an estimator is chosen by minimizing the empirical risk (based on squared error loss) over the class of functions whose $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm is bounded above by a particular function of the sample size. Later in this section, we will describe this algorithm and its analysis; we will summarize our co-authors' contributions and include our own in more detail as a part of this dissertation. Before doing so, we first discuss the construction of a $C^{1,1}(\mathbb{R}^d)$ extension from noise-free observations, i.e., the interpolation problem.

3.2.1 Noiseless $C^{1,1}(\mathbb{R}^d)$ Interpolation

Here, we provide some background on the question of extending a function defined on a finite set (in this case, \mathcal{X}') to a $C^{1,1}(\mathbb{R}^d)$ function with minimal $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm. For

now, we assume that the function values are noiselessly observed; the extension discussed in this section passes through the observed function values instead of smoothing out presumed noise.

Let $f : \mathcal{X}' \rightarrow \mathbb{R}$ be the function that we wish to extend. Let $Df : \mathcal{X}' \rightarrow \mathbb{R}^d$ be gradient information; $Df(x'_i)$ is written as $Df|_{x'_i}$ to emphasize this fact. In most instances, gradients are not sampled, but an interpolant can still be constructed. Let $\mathcal{A}(\mathbb{R}^d, \mathbb{R})$ be the space of real-valued affine functions on \mathbb{R}^d . Let $P : \mathcal{X}' \rightarrow \mathcal{A}(\mathbb{R}^d, \mathbb{R})$ be the map assigning $x'_i \mapsto P_{x'_i}$, where

$$P_{x'_i}(x) = f(x'_i) + \left(Df|_{x'_i} \right)^\top (x - x'_i), \quad x'_i \in \mathcal{X}', x \in \mathbb{R}^d;$$

P is called a 1-field. Denote the first-order Taylor expansion of $\tilde{f} \in C^{1,1}(\mathbb{R}^d)$ at $x' \in \mathbb{R}^d$ by $J_{x'} \tilde{f}$. The classical Whitney's extension theorem applied to $C^{1,1}(\mathbb{R}^d)$ states that if $\mathcal{X}' \subset \mathbb{R}^d$ is closed and P satisfies particular conditions, there exists an extension $f_w \in C^{1,1}(\mathbb{R}^d)$ for which $J_{x'_i} f_w = P_{x'_i}$ for all $x'_i \in \mathcal{X}'$. The following must hold for P : for some $M < \infty$ and all $x'_i, x'_j \in \mathcal{X}'$, 1) $\left| P_{x'_i}(x'_i) - P_{x'_j}(x'_i) \right| \leq M(x'_i - x'_j)^2$ and 2) $\left\| \nabla P_{x'_i}|_{x'_i} - \nabla P_{x'_j}|_{x'_i} \right\|_\infty \leq M|x'_i - x'_j|$. Clearly, since \mathcal{X}' is not only closed but finite, Whitney's extension theorem applies. Moreover, f can be extended even when gradients are not sampled; any arbitrary Df can be used to define a 1-field for which an extension exists.

More care must be taken when a minimal-seminorm extension is desired—Whitney's extension theorem does not make any claims about optimality. Instead, several results of [Le Gruyer \(2009\)](#) are needed. The minimal $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm of an extension defines two closely-related seminorms on the spaces of 1-fields and functions with domain \mathcal{X}' . More explicitly,

$$\|P\|_{\dot{C}^{1,1}(\mathcal{X}')} := \inf \left\{ \text{Lip}(\nabla \tilde{f}) \mid \tilde{f} \in C^{1,1}(\mathbb{R}^d), J_{x'_i} \tilde{f} = P_{x'_i} \forall x'_i \in \mathcal{X}' \right\},$$

and, when gradients are unknown,

$$\|f\|_{\dot{C}^{1,1}(\mathcal{X}')} := \inf \left\{ \text{Lip}(\nabla \tilde{f}) \mid \tilde{f} \in C^{1,1}(\mathbb{R}^d), \tilde{f}(x'_i) = f(x'_i) \forall x'_i \in \mathcal{X}' \right\}.$$

Le Gruyer (2009) proves that $\|P\|_{\dot{C}^{1,1}(\mathcal{X}')}$ can be computed directly from the sample; $\|f\|_{\dot{C}^{1,1}(\mathcal{X}')}$ can be computed similarly after a preliminary convex optimization step to find an optimal 1-field. Thus, the lower bound on the $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm of an extension can be determined exactly. These results are summarized in the following theorem.

Theorem 25 (Le Gruyer, 2009). *Let $\mathcal{X}' \subset \mathbb{R}^d$ be a finite set. Let $f : \mathcal{X}' \rightarrow \mathbb{R}$, and let $Df : \mathcal{X}' \rightarrow \mathbb{R}^d$. Let $P : \mathcal{X}' \rightarrow \mathcal{A}(\mathbb{R}^d, \mathbb{R})$ be the 1-field corresponding to f and Df ; i.e., $P_{x'_i}(x'_i) = f(x'_i)$ and $\nabla P_{x'_i} = Df|_{x'_i}$ for all $x'_i \in \mathcal{X}'$. Define the functionals*

$$\Gamma^1(P; \mathcal{X}') := 2 \sup_{x \in \mathbb{R}^d} \max_{x'_i \neq x'_j \in \mathcal{X}'} \frac{P_{x'_i}(x) - P_{x'_j}(x)}{|x'_i - x|^2 + |x'_j - x|^2}$$

and

$$\Gamma^1(f; \mathcal{X}') := \inf \left\{ \Gamma^1(\tilde{P}; E) \mid \tilde{P}_{x'_i}(x'_i) = f(x'_i) \forall x'_i \in \mathcal{X}' \right\}.$$

Then, the following statements hold.

i) (Proposition 2.2) *The value of $\Gamma^1(P; \mathcal{X}')$ depends only on the sample. Specifically,*

$$\Gamma^1(P; \mathcal{X}') = \max_{x'_i \neq x'_j \in \mathcal{X}'} \sqrt{A(P; x'_i, x'_j)^2 + B(P; x'_i, x'_j)^2} + |A(P; x'_i, x'_j)|,$$

where

$$A(P; x'_i, x'_j) := \left(2(f(x'_i) - f(x'_j)) + \left(Df|_{x'_i} - Df|_{x'_j} \right)^\top (x'_j - x'_i) \right) / |x'_i - x'_j|^2$$

and

$$B(P; x'_i, x'_j) := \left| Df|_{x'_i} - Df|_{x'_j} \right| / |x'_i - x'_j|.$$

ii) (Theorem 2.6) *There exists $f_{l,P} \in C^{1,1}(\mathbb{R}^d)$ with $\text{Lip}(\nabla f_{l,P}) = \Gamma^1(P; \mathcal{X}')$ for which $J_{x'_i} f_{l,P} = P_{x'_i}$ for all $x'_i \in \mathcal{X}'$. Furthermore, this is a minimal extension; i.e.,*

$$\|P\|_{\dot{C}^{1,1}(\mathcal{X}')} = \Gamma^1(P; \mathcal{X}').$$

iii) (Theorem 3.2) In the case where only f is known, there exists $f_{l,f} \in C^{1,1}(\mathbb{R}^d)$ with $\text{Lip}(\nabla f_{l,f}) = \Gamma^1(f; \mathcal{X}')$ for which $f_{l,f}(x'_i) = f(x'_i)$ for all $x'_i \in \mathcal{X}'$. This is a minimal extension; i.e.,

$$\|f\|_{\dot{C}^{1,1}(\mathcal{X}')} = \Gamma^1(f; \mathcal{X}').$$

An interpolant can be explicitly defined by combining the work of [Le Gruyer \(2009\)](#) with a construction due to [Wells \(1973\)](#); this is done in a paper of [Herbert-Voss et al. \(2017\)](#), who also create and analyze efficient algorithms for implementation. The next theorem is key in constructing an interpolant.

Theorem 26 (Wells' Construction, Theorem 1, Wells, 1973). *Let $\mathcal{X}' \subset \mathbb{R}^d$ be a finite set. Let $f : \mathcal{X}' \rightarrow \mathbb{R}$, and let $Df : \mathcal{X}' \rightarrow \mathbb{R}^d$. Let $P : \mathcal{X}' \rightarrow \mathcal{A}(\mathbb{R}^d, \mathbb{R})$ be the 1-field corresponding to f and Df ; i.e., $P_{x'_i}(x'_i) = f(x'_i)$ and $\nabla P_{x'_i} = Df|_{x'_i}$ for all $x'_i \in \mathcal{X}'$. Let M be a constant satisfying*

$$f(x'_j) \leq f(x'_i) + \frac{1}{2} \left(Df|_{x'_i} + Df|_{x'_j} \right)^\top (x'_j - x'_i) + \frac{M}{4} |x'_j - x'_i|^2 - \frac{1}{4M} \left| Df|_{x'_i} - Df|_{x'_j} \right|^2$$

for all $x'_i, x'_j \in \mathcal{X}'$.

Then, there exists $F : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $F \in C^{1,1}(\mathbb{R}^d)$, $J_{x'_i} F = P_{x'_i}$ for all $x'_i \in \mathcal{X}'$, and $\text{Lip}(\nabla F) = M$.

The proof is constructive. The procedure is technical, but the basic idea is to define the interpolant piecewise on a partition of \mathbb{R}^d created from cell complexes whose geometry depends on \mathcal{X}' , P , and M . A few preliminary sets and functions must be defined before F can be stated. For each $x'_i \in \mathcal{X}'$, let $\tilde{x}'_i := x'_i - Df|_{x'_i}/M$, and define $d_{x'_i} : \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$x \mapsto f(x'_i) - \frac{\left| Df|_{x'_i} \right|^2}{2M} + \frac{M}{4} \left| x - \tilde{x}'_i \right|^2.$$

Given $S \subset \mathcal{X}'$, define $d_S : \mathbb{R}^d \rightarrow \mathbb{R}$ as the map assigning $x \mapsto \min \left\{ d_{x'_i}(x), x'_i \in S \right\}$, and let

$$\tilde{S} := \left\{ \tilde{x}'_i \mid x'_i \in S \right\}$$

$$\begin{aligned}
S_H &:= \text{the smallest affine space containing } \tilde{S} \\
\hat{S} &:= \text{the convex hull of } \tilde{S} \\
S_E &:= \left\{ x \in \mathbb{R}^d \mid d_{x'_i}(x) = d_{x'_j}(x) \forall x'_i, x'_j \in S \right\} \\
S_* &:= \left\{ x \in \mathbb{R}^d \mid d_{x'_i}(x) = d_{x'_j}(x) \leq d_{x'_k}(x) \forall x'_i, x'_j \in S, x'_k \in \mathcal{X}' \right\} \\
S_C &:= S_H \cap S_E.
\end{aligned}$$

For all $S \subset \mathcal{X}'$ for which there exists $x \in S_*$ such that $d_S(x) < d_{\mathcal{X}' \setminus S}(x)$, define

$$T_S := \left\{ \frac{1}{2}(v + w) \mid v \in \hat{S}, w \in S_* \right\}.$$

The collection $\{T_S\}$ partitions \mathbb{R}^d . The extension F is defined piecewise as $F(x) = F_S(x)$, where $F_S : T_S \rightarrow \mathbb{R}$ assigns

$$x \mapsto d_S(S_C) + \frac{M}{2} \inf_{w \in S_H} |x - w|^2 - \frac{M}{2} \inf_{z \in S_E} |x - z|^2.$$

The derivative of F is piecewise linear.

$\Gamma^1(P; \mathcal{X}')$ satisfies the inequality required for Theorem 26 to apply; thus, by Theorem 25, M can be taken to be $\|P\|_{\dot{C}^{1,1}(\mathcal{X}')}$ (or $\|f\|_{\dot{C}^{1,1}(\mathcal{X}')}$, as the case may be). This means that Wells' construction can be used to define a minimal interpolant. [Herbert-Voss et al. \(2017\)](#) compute $\|P\|_{\dot{C}^{1,1}(\mathcal{X}')}$ exactly, or, more efficiently when N' is large, to within a dimensionless constant. They use algorithms from computational geometry to make the calculation practical. Specifically, they rely on a decomposition of \mathcal{X}' known as the ε -well separated pairs decomposition, previously used in a similar context by [Fefferman and Klartag \(2009\)](#). When Df is not observed, they additionally use methods from convex optimization to compute $\|f\|_{\dot{C}^{1,1}(\mathcal{X}')}$ to within a dimensionless constant and an arbitrarily small additive error. They also efficiently compute the interpolant of Wells' construction in part by employing practical algorithms for the calculation of convex hulls.

3.2.2 Estimating a $C^{1,1}(\mathbb{R}^d)$ Function from a Noisy Sample

In the regression setting, the function values as well as the gradients must be estimated. Additionally, the sample is assumed to be generated by a particular function (in this case

$f^{*'})$ that should be recovered asymptotically by the estimation procedure. [Gustafson et al. \(2018\)](#) estimate $f^{*'})$ by a solution of the following constrained minimization problem:

$$\inf_{\tilde{f} \in C^{1,1}(\mathbb{R}^d)} \frac{1}{N'} \sum_{i=1}^{N'} \left(y'_i - \tilde{f}(x'_i) \right)^2, \quad \text{such that } \left\| \tilde{f} \right\|_{\dot{C}^{1,1}(\mathbb{R}^d)} \leq \tilde{M}',$$

where \tilde{M}' grows with N' . The precise rate of growth is chosen by appealing to properties of the function class $C^{1,1}(\mathbb{R}^d)$ and will be discussed later in this section alongside proofs of convergence. The objective function is the empirical expectation of squared error loss, i.e., the empirical risk. It will be denoted \hat{R} and considered as a functional that can be applied to either $C^{1,1}(\mathbb{R}^d)$ functions or 1-fields. The value of $\hat{R}(\tilde{f})$ is determined entirely by the values of \tilde{f} on \mathcal{X}' , and, by [Theorem 25](#) and the definition of the $\dot{C}^{1,1}(\mathcal{X}')$ seminorm, the constraint set is determined by the values of \tilde{f} and $D\tilde{f}$ on \mathcal{X}' . Thus, it is equivalent to first minimize \hat{R} over 1-fields \tilde{P} such that $\left\| \tilde{P} \right\|_{\dot{C}^{1,1}(\mathcal{X}')} \leq \tilde{M}'$ and then use the solution to construct an interpolant as discussed in [Section 3.2.1](#). Finding a minimizing 1-field is a convex optimization problem—the objective can be written as a real-valued convex function on $\mathbb{R}^{(d+1)N'}$, and the constraint is defined by a seminorm ball, which is a convex set.

The particular convex optimization algorithm used to solve this problem is due to [Vaidya \(1996\)](#). This is a cutting-plane method, which requires the derivation of separation oracles, i.e., hyperplanes separating any point not within a convex set of interest from all points within the set. Here, separation oracles are necessary for the sets $\left\{ \tilde{P} \mid \left\| \tilde{P} \right\|_{\dot{C}^{1,1}(\mathcal{X}')} \leq \tilde{M}' \right\}$ and $\left\{ \tilde{P} \mid \hat{R}(\tilde{P}) - \hat{R}(P^*) \leq \gamma \right\}$, where γ is a tolerance parameter and P^* is a minimizer of \hat{R} . The algorithm starts by finding a feasible point in the first set; then, fixing gradients, a set of function values are found that yield a feasible point in the second set. [Gustafson et al. \(2018\)](#) derive the desired separation oracles and prove that an empirical risk minimizer can be found to arbitrary precision.

Theorem 27 ([Gustafson et al., 2018](#)). *Let $x_i^{*'} \neq x_j^{*'} \in \mathcal{X}'$ and $x^* \in \mathbb{R}^d$ be values that*

maximize $\Gamma^1(P; \mathcal{X}')$. Let $\gamma > 0$ be a tolerance parameter. Let P^* be an optimal solution to

$$\inf_{\tilde{P}} \frac{1}{N'} \sum_{i=1}^{N'} \left(y'_i - \tilde{P}_{x'_i}(x'_i) \right)^2, \quad \text{such that } \left\| \tilde{P} \right\|_{\dot{C}^{1,1}(\mathcal{X}')} \leq \tilde{M}'.$$

Assume that $0 < r \leq |x'_i - x'_j| \leq R$ for all $x'_i \neq x'_j \in \mathcal{X}'$. Define the constants

$$\rho_1 := \sqrt{N'} \left(\frac{\tilde{M}' r^2}{8(1+r)} \right), \quad \rho_2 := \sqrt{N'} \left(\frac{\sum y_i'^2}{N'} + 4 \left(\frac{10(\sum y_i'^2)^{1/2}}{r} + \frac{5\tilde{M}'}{2} \right)^2 \right)^{1/2}.$$

Choose

$$L \geq \log_2 \left(\frac{4 \sum y_i'^2}{N' \gamma \rho_1} \right).$$

Then, the following statements hold.

i) Given $P_0 \notin \left\{ \tilde{P} \mid \left\| \tilde{P} \right\|_{\dot{C}^{1,1}(\mathcal{X}')} \leq \tilde{M}' \right\}$, a separating hyperplane is:

$$\frac{2 \left(\tilde{P}_{x_i^{*'}}(x^*) - \tilde{P}_{x_j^{*'}}(x^*) \right)}{|x_i^{*'} - x^*|^2 + |x_j^{*'} - x^*|^2} = \Gamma^1(P_0; \mathcal{X}').$$

Given $P_0 \notin \left\{ \tilde{P} \mid \hat{R}(\tilde{P}) - \hat{R}(P^*) \leq \gamma \right\}$, a separating hyperplane is:

$$\frac{2}{N'} \sum (P_{0,x'_i}(x'_i) - y'_i) \tilde{P}_{x'_i}(x'_i) = \frac{2}{N'} \sum (P_{0,x'_i}(x'_i) - y'_i) P_{0,x'_i}(x'_i).$$

ii) (Theorem 12) Using the separation oracles from part i) of this theorem, Vaidya's algorithm yields an approximate minimizer $\hat{P}_T \in \left\{ \tilde{P} \mid \hat{R}(\tilde{P}) - \hat{R}(P^*) \leq \gamma \right\}$ in $T = O((d+1)N'(L + \log \rho_2))$ steps.

The output of Vaidya's algorithm is a set of function values and gradients, i.e., a 1-field on \mathcal{X}' ; the $\dot{C}^{1,1}(\mathcal{X}')$ seminorm can be computed and then Wells' construction applied to create an estimator for $f^{*'}$. If a minimal interpolant is desired, the gradients can be discarded and the gradients of the minimal interpolant computed through the algorithms of [Herbert-Voss et al. \(2017\)](#); an application of Wells' construction gives the estimator of $f^{*'}$. Since the loss function only depends on the function values, either estimator is acceptable.

3.2.3 Sample Complexity of $C^{1,1}(\mathbb{R}^d)$ Regression

Gustafson et al. (2018) show the convergence of their estimator by using methods from empirical process theory to analyze the squared-error loss of the class $\mathcal{F}_{\widetilde{M}'} := \left\{ \tilde{f} \mid \tilde{f} \in C^{1,1}(\mathbb{R}^d), \left\| \tilde{f} \right\|_{C^{1,1}(\mathbb{R}^d)} \leq \widetilde{M}' \right\}$, where $\widetilde{M}' := N^{1/(2\tilde{d})}$ and $\tilde{d} := \max\{d, 5\}$. The supremum of the difference between the empirical risk and its expectation (i.e., the true risk) is bounded above by a function of the Rademacher complexity of $\mathcal{F}_{\widetilde{M}'}$, which is itself a function of the covering number of this class. The choice of \widetilde{M}' ensures that this supremum tends to zero with high probability. Theorem 31 gives a precise statement of this result, and also gives the rate of convergence of an empirical risk minimizer to $f^{*'}$ in sup norm. (Note that $\mathcal{F}_{\widetilde{M}'}$ is increasing in size; since $M^{*'} < \infty$, $f^{*'} \in \mathcal{F}_{\widetilde{M}'}$ eventually.) Its proof uses the empirical processes methods described in Section 2.2.5. Before stating this theorem, we state three preliminary results—McDiarmid’s inequality (Lemma 28), a concentration inequality for a loss class when given a function class of interest (Lemma 29), and an upper bound on the covering number of $\mathcal{F}_{\widetilde{M}'}$ with respect to the supremum norm (Lemma 30). We include the proofs of Lemma 29 and Theorem 31 since this was our major contribution to the paper of Gustafson et al. (2018).

Lemma 28 (McDiarmid’s inequality, McDiarmid, 1989). *Let $X_1, \dots, X_{N'}$ be independent random variables that take values in a set A . Suppose the function $f : A^{N'} \rightarrow \mathbb{R}$ satisfies*

$$\sup_{x_1, \dots, x_{N'}, x'_i \in A} |f(x_1, \dots, x_{N'}) - f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_{N'})| \leq c_i$$

for every $1 \leq i \leq N'$. Then, for $t > 0$,

$$\mathbb{P}[|f(X_1, \dots, X_{N'}) - \mathbb{E}f(X_1, \dots, X_{N'})| \geq t] \leq 2e^{-2t^2 / \sum_{i=1}^{N'} c_i^2}.$$

Lemma 29. *Let $\mathcal{F}_{\widetilde{M}'}$ be a class of functions $f : B_d(0, 1) \rightarrow \mathbb{R}$ with $\sup_{f \in \mathcal{F}_{\widetilde{M}'}} |f| \leq \widetilde{M}'$. Let $L : [-\widetilde{M}', \widetilde{M}'] \times Y \rightarrow \mathbb{R}$ be a bounded loss function with Lipschitz constant L_L and*

$0 \leq L \leq L_{max}$. Then, the following is true for $0 < \delta < 1$:

$$\mathbb{P} \left[\sup_{f \in \mathcal{F}_{\tilde{M}'}} |R(f) - \hat{R}(f)| < 4L_L \mathcal{R}_{N'}(\mathcal{F}_{\tilde{M}'}) + 7L_{max} \sqrt{\frac{\log(8/\delta)}{2N'}} \right] > 1 - \delta.$$

Proof. We provide a proof based on three results due to [Bartlett and Mendelson \(2003\)](#). We begin by adapting their Theorem 8 to find a bound on the risk that depends on a probabilistic term plus the expectation of the Rademacher average of the class of loss functions. We follow the proof of Lemma 4 of [Bousquet et al. \(2004\)](#) for guidance. We apply the two-sided form of McDiarmid's inequality as we want bounds on the absolute value of $R(f) - \hat{R}(f)$, and appeal to Theorems 11 and 12 of [Bartlett and Mendelson \(2003\)](#) to relate the expected Rademacher average of the loss class to the empirical Rademacher average of $\mathcal{F}_{\tilde{M}'}$.

Let $\tilde{L} \circ \mathcal{F}_{\tilde{M}'}$ be the class of functions consisting of $\{(x, y) \mapsto L(f(x), y) - L(0, y)\}$. If $h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}$, then $-L_{max} \leq h \leq L_{max}$. For any $f \in \mathcal{F}_{\tilde{M}'}$, the triangle inequality shows that

$$|R(f) - \hat{R}(f)| \leq \sup_{h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}} |\mathbb{E}h - \hat{\mathbb{E}}_{N'}h| + |\mathbb{E}L(0, y) - \hat{\mathbb{E}}_{N'}L(0, y)|.$$

McDiarmid's inequality yields more favorable expressions for both terms on the right-hand side as follows. The most that $\hat{\mathbb{E}}_{N'}L(0, y)$ can change by altering one sample is L_{max}/N' . Since $\mathbb{E}\hat{\mathbb{E}}_{N'}L(0, y) = \mathbb{E}L(0, y)$, we have, with probability $1 - \delta/4$,

$$|\mathbb{E}L(0, y) - \hat{\mathbb{E}}_{N'}L(0, y)| \leq \sqrt{\frac{L_{max}^2 \log 8/\delta}{2N'}}.$$

The most that $\sup_{h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}} |\mathbb{E}h - \hat{\mathbb{E}}_{N'}h|$ can change with an alteration of one sample is $2L_{max}/N'$. Therefore, with probability $1 - \delta/4$,

$$\left| \sup_{h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}} |\mathbb{E}h - \hat{\mathbb{E}}_{N'}h| - \mathbb{E} \sup_{h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}} |\mathbb{E}h - \hat{\mathbb{E}}_{N'}h| \right| \leq \sqrt{\frac{4L_{max}^2 \log 8/\delta}{2N'}}.$$

Now,

$$\mathbb{E} \sup_{h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}} |\mathbb{E}h - \hat{\mathbb{E}}_{N'}h| \leq \max \left\{ \mathbb{E} \sup_{h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}} (\mathbb{E}h - \hat{\mathbb{E}}_{N'}h), \mathbb{E} \sup_{h \in \tilde{L} \circ \mathcal{F}_{\tilde{M}'}} (\hat{\mathbb{E}}_{N'}h - \mathbb{E}h) \right\}.$$

Let $S' := \{(x''_1, y''_1), \dots, (x''_{N'}, y''_{N'})\}$ be a sample with the same distribution as $S := \mathcal{X}' \times \mathcal{Y}'$.

Conditioning on the original sample,

$$\begin{aligned}
\mathbb{E} \sup_{h \in \tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}} \left(\mathbb{E} h - \hat{\mathbb{E}}_{N'} h \right) &= \mathbb{E} \sup_{h \in \tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}} \mathbb{E} \left[\frac{1}{N'} \sum_{i=1}^{N'} h(x''_i, y''_i) - \hat{\mathbb{E}}_{N'} h \mid S \right] \\
&\leq \mathbb{E} \sup_{h \in \tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}} \left(\frac{1}{N'} \sum_{i=1}^{N'} h(x''_i, y''_i) - \hat{\mathbb{E}}_{N'} h \right) \\
&= \mathbb{E} \sup_{h \in \tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}} \frac{1}{N'} \sum_{i=1}^{N'} \sigma_i (h(x''_i, y''_i) - h(x'_i, y'_i)) \\
&\leq \mathbb{E} \sup_{h \in \tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}} \frac{1}{N'} \sum_{i=1}^{N'} \sigma_i h(x''_i, y''_i) + \mathbb{E} \sup_{h \in \tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}} \frac{1}{N'} \sum_{i=1}^{N'} -\sigma_i h(x'_i, y'_i) \\
&= 2\mathbb{E} \mathcal{R}_{N'}(\tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}).
\end{aligned}$$

The second line follows by applying Jensen's inequality to sup, which is convex. Note the preceding argument is symmetric in $\mathbb{E} h$ and $\hat{\mathbb{E}}_{N'} h$. Therefore, $\mathbb{E} \sup_{h \in \tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}} \left| \mathbb{E} h - \hat{\mathbb{E}}_{N'} h \right|$ has the same upper bound and, with probability $1 - \delta/2$,

$$\left| R(f) - \hat{R}(f) \right| \leq 2\mathbb{E} \mathcal{R}_{N'}(\tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}) + 3L_{max} \sqrt{\frac{\log(8/\delta)}{2N'}}.$$

Theorem 11 of [Bartlett and Mendelson \(2003\)](#) uses McDiarmid's inequality to bound the difference between the empirical and expected Rademacher averages, but assumes that we are interested in the Rademacher complexity of a class of functions mapping to $[-1, 1]$. Since $\mathcal{F}_{\tilde{M}'}$ maps to $[-\tilde{M}', \tilde{M}']$, we rederive the analogous result here. The most that one sample affects $\mathcal{R}_{N'}(\tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'})$ is $2L_{max}/N'$. We have

$$\mathbb{P} \left[\left| \mathcal{R}_{N'}(\tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}) - \mathbb{E} \mathcal{R}_{N'}(\tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}) \right| \geq t \right] \leq 2e^{-2N't^2/(4L_{max}^2)}.$$

Thus, with probability $1 - \delta/2$,

$$\begin{aligned}
2\mathbb{E} \mathcal{R}_{N'}(\tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}) &\leq 2\mathcal{R}_{N'}(\tilde{\mathcal{L}} \circ \mathcal{F}_{\tilde{M}'}) + 4L_{max} \sqrt{\frac{\log(4/\delta)}{2N'}} \\
&\leq 4L_L \mathcal{R}_{N'}(\mathcal{F}_{\tilde{M}'}) + 4L_{max} \sqrt{\frac{\log(8/\delta)}{2N'}}.
\end{aligned}$$

The second line follows from part 4 of Theorem 12, which states that, for $\tilde{L} : \mathbb{R} \rightarrow \mathbb{R}$ with Lipschitz constant $L_{\tilde{L}}$ and satisfying $\tilde{L}(0) = 0$, $\mathbb{E}\mathcal{R}_{N'}(\tilde{L} \circ \mathcal{F}_{\tilde{M}'}) \leq 2L_{\tilde{L}}\mathbb{E}\mathcal{R}_{N'}(\mathcal{F}_{\tilde{M}'})$. The reasoning from the proof also applies to the empirical Rademacher average, giving $\mathcal{R}_{N'}(\tilde{L} \circ \mathcal{F}_{\tilde{M}'}) \leq 2L_{\tilde{L}}\mathcal{R}_{N'}(\mathcal{F}_{\tilde{M}'})$. Since \tilde{L} has the same Lipschitz constant as L , we use the notation L_L .

Finally, with probability at least $1 - \delta$,

$$\left| R(f) - \hat{R}(f) \right| \leq 4L_L\mathcal{R}_{N'}(\mathcal{F}_{\tilde{M}'}) + 7L_{\max}\sqrt{\frac{\log(8/\delta)}{2N'}}.$$

■

Lemma 30 (adapted from Theorem 2.7.1 of [Van Der Vaart and Wellner, 1996](#)).

There exists a constant K depending only on d such that, for every $\eta > 0$,

$$\log N(\eta, \mathcal{F}_{\tilde{M}'}, \|\cdot\|_{\infty}) \leq K \left(\frac{\tilde{M}'}{\eta} \right)^{d/2}.$$

Theorem 31. Let $\tilde{M}' := N'^{1/(2\tilde{d})}$ and $\tilde{d} := \max\{d, 5\}$. Let $\mathcal{F}_{\tilde{M}'}$ be the class of functions whose $C^{1,1}(\mathbb{R}^d)$ norms are no greater than \tilde{M}' . Define $\tilde{L}'_{\max} := \left(\tilde{M}' + M^{*'} + \sigma\sqrt{2\log 2N'} + N'^{1/\tilde{d}} \right)^2$, $\tilde{L}'_L := 2\left(\tilde{M}' + M^{*'} + \sigma\sqrt{2\log 2N'} + N'^{1/\tilde{d}} \right)$, and

$$\tilde{R}' := \begin{cases} 4\left(81^{1/d}K^{2/d}\tilde{M}'N'^{-2/d}\right) + 12\frac{\sqrt{K}\tilde{M}'^{d/4}\left(4\tilde{M}'^{1-d/4} - 4\left(81^{1/d}K^{2/d}\tilde{M}'N'^{-2/d}\right)^{1-d/4}\right)}{(4-d)\sqrt{N'}} & : d \neq 4 \\ 4\left(3\sqrt{K}\tilde{M}'N'^{-1/2}\right) + 12\frac{\sqrt{K}\tilde{M}'}{\sqrt{N'}}\left(\log \tilde{M}' - \log\left(3\sqrt{K}\tilde{M}'N'^{-1/2}\right)\right) & : d = 4. \end{cases}$$

i) For $\delta \in (0, 1)$,

$$\mathbb{P}\left[\sup_{f \in \mathcal{F}_{\tilde{M}'}} \left| R(f) - \hat{R}(f) \right| < \varepsilon'\right] > 1 - \delta - e^{-N'^{2/\tilde{d}}/(2\sigma^2)},$$

where $\varepsilon' := 4\tilde{L}'_L\tilde{R}' + 7\tilde{L}'_{\max}\sqrt{\log(8/\delta)/2N'}$. ε' is a monotonically-decreasing function of N' for large enough N' and $\lim_{N' \rightarrow \infty} \varepsilon' = 0$. Furthermore, $\sup_{f \in \mathcal{F}_{\tilde{M}'}} \left| R(f) - \hat{R}(f) \right| \xrightarrow{a.s.} 0$.

ii) Now let $\widetilde{M}' := N^{1/(16\tilde{d}^2)}$. The conclusions of *i)* still hold with this choice of \widetilde{M}' . Assume that \mathcal{X}' is sampled i.i.d. from a probability measure \mathcal{P}' absolutely continuous with respect to \mathcal{L}^d with density p' such that $0 < p'_{\min} \leq p'(x)$ for all $x \in \mathbb{R}^d$. Let $\widehat{f}' \in \mathcal{F}_{\widetilde{M}'}$ be an empirical risk minimizer. Let $\tilde{r}' := \left(2\varepsilon'\alpha'^{-2}\Gamma\left(1 + \frac{d}{2}\right) / p'_{\min}\pi^{d/2}\right)^{1/d}$ and $\alpha' := N'^{-1/(10\tilde{d})}$. Then, for $\delta \in (0, 1)$,

$$\mathbb{P}\left[\sup_{x \in B_d(0,1)} \left|\widehat{f}'(x) - f^{*'}(x)\right| < \beta'\right] > 1 - \delta - e^{-N'^{2/\tilde{d}}/(2\sigma^2)},$$

where $\beta' := \widetilde{M}'\tilde{r}' + \alpha'$. β' is a monotonically-decreasing function of N' for large enough N' and $\lim_{N' \rightarrow \infty} \beta' = 0$. Furthermore, $\sup_{x \in B_d(0,1)} \left|\widehat{f}'(x) - f^{*'}(x)\right| \xrightarrow{a.s.} 0$.

Proof. *i)* $\mathcal{F}_{\widetilde{M}'}$ is a sequence of function classes with increasing $C^{1,1}(\mathbb{R}^d)$ norm. We set the rate $\widetilde{M}' := N'^{1/(2\tilde{d})}$, where $\tilde{d} := \max\{d, 5\}$, so that $f^{*'}$ is a candidate for large enough N' . We aim to use Lemma 29 to prove the desired probability statement, but our loss function is unbounded since the elements of \mathcal{Y}' can be arbitrarily large. To circumvent this, we also let the maximum value of $y'_i \in \mathcal{Y}'$ increase with N' ; samples violating this condition are part of the error probability. Write $y'_i = f^{*'}(x'_i) + \xi'_i$, where $\xi'_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. We condition on the event $\mathcal{H} := \left\{\max_{1 \leq i \leq N'} |\xi'_i| \leq \sigma\sqrt{2\log 2N'} + N'^{1/\tilde{d}}\right\}$. Theorem 7.1 of Ledoux (2005) gives the following bound for suprema of Gaussian processes:

$$\mathbb{P}\left[\max_{1 \leq i \leq N'} |\xi'_i| < \mathbb{E} \max_{1 \leq i \leq N'} |\xi'_i| + r\right] > 1 - e^{-r^2/2\sigma^2}.$$

The following is well-known (Boucheron, Lugosi, and Massart, 2013):

$$\mathbb{E} \max_{1 \leq i \leq N'} |\xi'_i| \leq \sigma\sqrt{2\log 2N'}.$$

Thus, $\mathbb{P}(\mathcal{H}) > 1 - e^{-N'^{2/\tilde{d}}/2\sigma^2}$.

Since the loss function is bounded after conditioning on \mathcal{H} , we can compute:

$$L_{max} < \sup_{x,y,f} (f(x) - y)^2$$

$$\begin{aligned}
&< \sup_{x,y,f} (|f(x)| + |y|)^2 \\
&< \left(\widetilde{M}' + M^{*'} + \sigma\sqrt{2\log 2N'} + N'^{1/\tilde{d}} \right)^2 \\
&= \left(N'^{1/(2\tilde{d})} + M^{*'} + \sigma\sqrt{2\log 2N'} + N'^{1/\tilde{d}} \right)^2 \\
&:= \widetilde{L}'_{max}.
\end{aligned}$$

We also find the Lipschitz constant as follows, where $f_1, f_2 \in \mathcal{F}_{\widetilde{M}'}$:

$$\begin{aligned}
\sup_{x,y,f_1,f_2} |(f_1(x) - y)^2 - (f_2(x) - y)^2| &= \sup_{x,y,f_1,f_2} |(-2y + f_1(x) + f_2(x))(f_1(x) - f_2(x))| \\
&\leq \sup_{x,y,f_1,f_2} |-2y + f_1(x) + f_2(x)| \|f_1 - f_2\|_\infty.
\end{aligned}$$

This implies:

$$\begin{aligned}
L_L &\leq \sup_{x,f_1,f_2} |f_1(x) + f_2(x)| + 2 \sup_y |y| \\
&< 2 \left(\widetilde{M}' + M^{*'} + \sigma\sqrt{2\log 2N'} + N'^{1/\tilde{d}} \right) \\
&= 2 \left(N'^{1/(2\tilde{d})} + M^{*'} + \sigma\sqrt{2\log 2N'} + N'^{1/\tilde{d}} \right) \\
&:= \widetilde{L}'_L.
\end{aligned}$$

Next, we bound the Rademacher complexity using the entropy integral:

$$\begin{aligned}
\mathcal{R}_{N'}(\mathcal{F}_{\widetilde{M}'}) &\leq \inf_{\gamma \geq 0} \left\{ 4\gamma + 12 \int_\gamma^{\widetilde{M}'} \sqrt{\frac{\log N(\eta, \mathcal{F}_{\widetilde{M}'}, \|\cdot\|_{\mathcal{L}_2(\mathcal{P}'_{N'})})}{N'}} d\eta \right\} \\
&\leq \inf_{\gamma \geq 0} \left\{ 4\gamma + 12 \int_\gamma^{\widetilde{M}'} \sqrt{\frac{\log N(\eta, \mathcal{F}_{\widetilde{M}'}, \|\cdot\|_\infty)}{N'}} d\eta \right\} \\
&\leq \inf_{\gamma \geq 0} \left\{ 4\gamma + 12 \int_\gamma^{\widetilde{M}'} \sqrt{\frac{K \widetilde{M}'^{d/2}}{N' \eta^{d/2}}} d\eta \right\}.
\end{aligned}$$

The second inequality is standard, and the third is from substituting the covering number bound from Lemma 30. The integral is different for $d \neq 4$ and $d = 4$. In the first case,

$$\mathcal{R}_{N'}(\mathcal{F}_{\widetilde{M}'}) \leq \inf_{\gamma \geq 0} \left\{ 4\gamma + 12 \frac{\sqrt{K} \widetilde{M}'^{d/4} \left(4\widetilde{M}'^{1-d/4} - 4\gamma^{1-d/4} \right)}{(4-d)\sqrt{N'}} \right\},$$

and the infimum is achieved at $\gamma = 81^{1/d} K^{2/d} \widetilde{M}' N'^{-2/d}$. When $d = 4$,

$$\mathcal{R}_{N'}(\mathcal{F}_{\widetilde{M}'}) \leq \inf_{\gamma \geq 0} \left\{ 4\gamma + 12 \frac{\sqrt{K} \widetilde{M}'}{\sqrt{N'}} (\log \widetilde{M}' - \log \gamma) \right\},$$

which is minimized at $\gamma = 3\sqrt{K} \widetilde{M}' N'^{-1/2}$. Substituting in γ and \widetilde{M}' gives us

$$\widetilde{R}' := \begin{cases} \frac{4N'^{1/(2\tilde{d})} \left(-\frac{12\sqrt{K}}{\sqrt{N'}} + 81^{1/d} d K^{2/d} N'^{-2/d} \right)}{d-4} & : d \neq 4 \\ \frac{6\sqrt{K} N'^{1/(2\tilde{d})} (2 + \log N' - \log 9 - \log K)}{\sqrt{N'}} & : d = 4, \end{cases}$$

so that $\mathcal{R}_{N'}(\mathcal{F}_{\widetilde{M}'}) \leq \widetilde{R}'$.

Set

$$\varepsilon' := 4\widetilde{L}'_L \widetilde{R}' + 7\widetilde{L}'_{max} \sqrt{\frac{\log(8/\delta)}{2N'}}.$$

Each term goes to zero, so $\lim_{N' \rightarrow \infty} \varepsilon' = 0$. Additionally, $\partial \varepsilon' / \partial N' = O\left(-N'^{-(2\tilde{d}+1)/(2\tilde{d})}\right)$. If N' is sufficiently large, $\partial \varepsilon' / \partial N' < 0$, and ε' is decreasing in N' . Finally, applying Lemma 29 yields the first part of the theorem.

To strengthen the result to almost-sure convergence, we appeal to the Borel-Cantelli lemma. It is enough to show that

$$\sum_{N'} \mathbb{P} \left[\sup_{f \in \mathcal{F}_{\widetilde{M}'}} |R(f) - \widehat{R}(f)| > \varepsilon'' \mid \mathcal{H} \right] + \sum_{N'} e^{-N'^{2/\tilde{d}}/2\sigma^2} < \infty,$$

where $\varepsilon'' > 0$ is an arbitrary, fixed value. The second series converges by comparison with the integral

$$\int_0^\infty e^{-N'^{2/\tilde{d}}/2\sigma^2} dN' = (2\sigma^2)^{\tilde{d}/2} \Gamma\left(1 + \tilde{d}/2\right).$$

Each term in the first series is bounded above by $\min\{1, \delta\}$, with δ satisfying $\varepsilon'' = 4L_L \mathcal{R}_{N'}(\mathcal{F}_{\widetilde{M}'}) + 7L_{max} \sqrt{\log(8/\delta)/2N'}$. For a given ε'' , a solution does not exist if N' is too

small. When N' is large enough, we have the following:

$$\begin{aligned} \frac{\delta}{8} &= \exp \left\{ -2N' \left(\frac{\varepsilon'' - 4L_L \mathcal{R}_{N'}(\mathcal{F}_{\tilde{M}'})}{7L_{max}} \right)^2 \right\} \\ &\leq \exp \left\{ -2N' \left(\frac{\varepsilon'' - 4\tilde{L}'_L \tilde{R}'}{7\tilde{L}'_{max}} \right)^2 \right\} \\ &:= \tilde{\delta} \end{aligned}$$

The second line follows because $\tilde{L}'_L \tilde{R}' \rightarrow 0$ and $\partial(\tilde{L}'_L \tilde{R}')/\partial N' < 0$. Eventually, $0 < \tilde{L}'_L \tilde{R}' < \varepsilon''$ and $(\varepsilon'' - 4\tilde{L}'_L \tilde{R}')^2 \leq (\varepsilon'' - 4L_L \mathcal{R}_{N'}(\mathcal{F}_{\tilde{M}'})^2$.

Asymptotically, $\log \tilde{\delta} = O(-N'^{1-4/\tilde{d}})$. Furthermore, its derivative is $O(-N'^{-4/\tilde{d}})$, so $\tilde{\delta}$ is decreasing for large N' . Since

$$\int_0^\infty e^{-N'^{1-4.1/\tilde{d}}} dN' = \left(1 - 4.1/\tilde{d}\right)^{-1} \Gamma\left\{\left(1 - 4.1/\tilde{d}\right)^{-1}\right\},$$

the integral test shows the tail of $\sum_{N'} \tilde{\delta}$ is finite, proving

$$\sup_{f \in \mathcal{F}_{\tilde{M}'}} \left| R(f) - \hat{R}(f) \right| \xrightarrow{a.s.} 0.$$

ii) We again condition on the event $\mathcal{H} := \left\{ \max_{1 \leq i \leq N'} |\xi'_i| \leq \sigma \sqrt{2 \log 2N'} + N'^{1/\tilde{d}} \right\}$. However, we now set $\tilde{M}' := N'^{1/(16\tilde{d}^2)}$. The conclusions of the first part of this theorem still hold with the appropriate modifications made to any constants depending on \tilde{M}' . To relate the uniform risk bound to the difference between \hat{f}' and $f^{*'}$, we start by decomposing the risk. With probability at least $(1 - \delta)(1 - e^{-N'^{2/\tilde{d}}/2\sigma^2})$ over the sample,

$$\begin{aligned} \mathbb{E} \left[\left(\hat{f}' - f^{*'} \right)^2 \right] &= R(\hat{f}') - R(f^{*'}) \\ &\leq \left| R(\hat{f}') - \hat{R}(\hat{f}') \right| + \left| \hat{R}(f^{*'}) - R(f^{*'}) \right| \\ &\leq 2 \sup_{f \in \mathcal{F}_{\tilde{M}'}} \left| R(f) - \hat{R}(f) \right| \\ &< 2\varepsilon'. \end{aligned}$$

Combining this with Chebyshev's inequality, we have

$$\begin{aligned} \mathbb{P}\left[|\hat{f}' - f^{*'}| > \alpha'\right] &< \mathbb{E}\left[\left(\hat{f}' - f^{*'}\right)^2\right] / \alpha'^2 \\ &< 2\varepsilon'\alpha'^{-2} \end{aligned}$$

for $\alpha' > 0$. In other words, \hat{f}' lies within a tube of radius α' , except on a set $A \subset B_d(0, 1)$ such that $P(A) < 2\varepsilon'\alpha'^{-2}$.

Let $h := \sup_{x \in A} |\hat{f}' - f^{*'}| - \alpha'$. Because \hat{f}' and $f^{*'}$ are Lipschitz, h is constrained by the inequality

$$\frac{(f^{*'}(x) + M^{*'}r + \alpha' + h) - (f^{*'}(x) + \alpha')}{r} \leq \widetilde{M}',$$

where x is on the boundary of A , and r is the inradius of A . This implies that $h \leq \widetilde{M}'r$. We can maximize this by taking A to be the d -dimensional ball of radius

$$\tilde{r}' := \left(\frac{2\varepsilon'\alpha'^{-2}\Gamma(1 + \frac{d}{2})}{p'_{min}\pi^{d/2}} \right)^{1/d},$$

where p'_{min} is a constant bounding the density p' away from zero. This shows

$$\sup_{x \in B^d} |\hat{f}' - f^{*'}| < \widetilde{M}'\tilde{r}' + \alpha'.$$

Set $\alpha' := N'^{-1/(10d)}$. Then $\widetilde{M}'\tilde{r}' = O(N'^{\rho})$ and $\partial(\widetilde{M}'\tilde{r}')/\partial N' = O(-N'^{\rho-1})$, where

$$\rho := \begin{cases} -3/(50d) + 1/400 & : d \leq 5 \\ (5 - 59d)/(80d^3) & : d > 5. \end{cases}$$

Now, defining $\beta' := \widetilde{M}'\tilde{r}' + \alpha'$ gives the first part of the theorem.

Almost-sure convergence follows from a similar argument as the last part of **i)**. It suffices to show that, for arbitrary $\beta' > 0$,

$$\sum_{N'} \mathbb{P}\left[\sup_{x \in B_d(0,1)} |\hat{f}' - f^{*'}| > \beta'' \mid \mathcal{H} \right] + \sum_{N'} e^{-N'^{2/\tilde{d}}/2\sigma^2} < \infty,$$

where we have already shown convergence of the second series. Let

$$\varepsilon'' := \left(\frac{\beta'' - \alpha'}{\widetilde{M}'} \right)^d \left(\frac{c\pi^{d/2}}{\Gamma(1 + d/2)} \right) \left(\frac{\alpha'^2}{2} \right)$$

be the solution when setting $\beta'' = \widetilde{M}'\tilde{r}' + \alpha'$ and solving for ε' . For fixed β'' and large N' , there is a corresponding $\varepsilon'' > 0$. Note that ε'' is not fixed, but decreasing in N' . In fact, $\varepsilon'' = O\left(N'^{-(5d+16\tilde{d})/(80\tilde{d}^2)}\right)$. Since $\tilde{L}'_L\tilde{R}' = O\left(N'^{-119/400}\right)$ for $d < 5$ and $O\left(N'^{-(16d-1)/(16d^2)}\right)$ for $d \geq 5$, eventually $0 < \tilde{L}'_L\tilde{R}' < \varepsilon''$. Therefore,

$$\tilde{\delta} := 8 \exp \left\{ -2N' \left(\frac{\varepsilon'' - 4\tilde{L}'_L\tilde{R}'}{7\tilde{L}'_{max}} \right)^2 \right\}$$

is an upper bound for the tail of the first series. Observe that $\log(\tilde{\delta}/8) = O\left(-N'^{1-d/(8\tilde{d}^2)} - 22/(5\tilde{d})\right)$ and $\partial(\log(\tilde{\delta}/8))/\partial N' = O\left(-N'^{-d/(8\tilde{d}^2)} - 22/(5\tilde{d})\right)$. Comparison with the integral

$$\begin{aligned} \int_0^\infty e^{-N'^{1-d/(8\tilde{d}^2)} - 22.1/(5\tilde{d})} dN' &= 1 / \left(1 - d/(8\tilde{d}^2) - 22.1/(5\tilde{d}) \right) \\ &\quad \times \Gamma \left\{ \left(1 - d/(8\tilde{d}^2) - 22.1/(5\tilde{d}) \right)^{-1} \right\} \end{aligned}$$

is enough to give almost-sure uniform convergence of the empirical risk minimizer. \blacksquare

Of course, the usefulness of the bounds in Theorem 31 depends on being able to find an empirical risk minimizer in the class $\mathcal{F}_{\widetilde{M}'}$. In Section 3.2.2, an estimator is found whose $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm is controlled. The following theorem shows that with our assumptions on \mathcal{P}' , for N' large enough, this estimator is in $\mathcal{F}_{\widetilde{M}'}$ and is the empirical risk minimizer of this class.

Theorem 32. *Assume that \mathcal{X}' is sampled i.i.d. from a probability measure \mathcal{P}' absolutely continuous with respect to \mathcal{L}^d with density p' such that $0 < p'_{min} \leq p'(x) \leq p'_{max}$ for all $x \in \mathbb{R}^d$. Let $\mathcal{K}' \subseteq L^2(\mathcal{X}')$ be the closed convex set of all functions $f : \mathcal{X}' \rightarrow \mathbb{R}$ such that*

$$\|f\|_{\dot{C}^{1,1}(\mathcal{X}')} \leq \widetilde{M}',$$

where $\widetilde{M}' = o(N'^{2/d})$. Let h be the projection of \mathcal{Y}' onto \mathcal{K}' with respect to the Hilbert space $L^2(\mathcal{X}')$.

i) (Theorem 13, [Gustafson et al., 2018](#)) Then when N' is sufficiently large, with probability at least $1 - \exp(-N'^{1/100})$,

$$\max\left(\|h\|_{C^0(\mathcal{X}')} , \|h\|_{\dot{C}^1(\mathcal{X}')} \right) < \widetilde{M}'/2.$$

ii) (Theorem 16, [Gustafson et al., 2018](#)) Let $\|h\|_{\dot{C}^{1,1}(\mathcal{X}')} \leq \widetilde{M}'$ and $\max(\|h\|_{C^0(\mathcal{X}')} , \|h\|_{\dot{C}^1(\mathcal{X}')}) \leq \widetilde{M}'/2$. Then, with probability at least $1 - \exp(-N'^{1/100})$, any minimal $\dot{C}^{1,1}(\mathbb{R}^d)$ -seminorm extension f of h to the unit ball satisfies

$$\max\left(\|f\|_{C^0(\mathbb{R}^d)} , \|f\|_{\dot{C}^1(\mathbb{R}^d)} \right) \leq \widetilde{M}'.$$

3.3 Geometric Framework

In this section, we aim to collect the geometric background we will need for the rest of this chapter. Most of the concepts below are fundamental to differential geometry and can be found in standard references ([Lee, 2018](#); [Lee, 2012](#); [Krantz and Parks, 2012](#)); we will provide specific sources for lesser-known results.

We use manifold and submanifold interchangeably. A C^2 submanifold of dimension d is a subset \mathcal{M} of \mathbb{R}^n with the following differentiable structure. For every point $x \in \mathcal{M}$, there is an open set $U \subset \mathbb{R}^n$ containing x , a convex open set $W \subset \mathbb{R}^d$, and C^2 functions $\phi : U \rightarrow W$ and $\psi : W \rightarrow U$ such that $\mathcal{M} \cap U = \psi(W)$ and $\phi \circ \psi = \text{id}$. (U, ϕ) is known as a chart and (W, ψ) as a local parametrization. (A C^2 function is a function whose second partial derivatives all exist and are continuous). This is equivalent to other standard definitions of a submanifold of \mathbb{R}^n ; e.g., \mathcal{M} can be regarded as a set that is locally the graph of a C^2 function or as a subset of \mathbb{R}^n such that at each point there exists a C^2 chart $\phi : \mathbb{R}^n \supset U \rightarrow \mathbb{R}^n$ of the ambient space such that $\phi(U \cap \mathcal{M}) \subset \mathbb{R}^d \times \{0\}$.

If we have a collection of charts that form an open cover of \mathcal{M} , we can define a partition of unity subordinate to this cover. This is a collection of real-valued functions $\{\eta_i\}$ with support lying in a single chart such that for each $x \in \mathcal{M}$, $\eta_i(x) \geq 0$, there are only a finite number of η_i that are nonzero, and $\sum_i \eta_i(x) = 1$. The $\{\eta_i\}$ can be chosen to be of the same

smoothness class as \mathcal{M} . Partitions of unity are useful for performing geometric calculations on regions of \mathcal{M} that span multiple charts.

At each point x of \mathcal{M} , there exists a d -dimensional affine subspace of \mathbb{R}^n tangent to \mathcal{M} ; this is called the tangent space and is denoted by $T_x\mathcal{M}$. Its orthogonal complement is known as the normal space and is denoted by $T_x^\perp\mathcal{M}$. The elements of $T_x\mathcal{M}$ are equivalence classes of curves passing through x that have a particular tangent vector; this is equivalent to considering them as elements of \mathbb{R}^d transformed to lie tangent to \mathcal{M} at x . When we consider $T_x\mathcal{M}$ as embedded in \mathbb{R}^n , the norm of an element of $T_x\mathcal{M}$ is given by the standard Euclidean norm. We denote an orthonormal basis of the tangent space $T_x\mathcal{M}$ as V_x and the projection either by $V_x V_x^\top$ or Π_x ; the projection onto the normal space $T_x^\perp\mathcal{M}$ is written as Π_x^\perp . We can also consider $T_x\mathcal{M}$ as an abstract subspace or a subspace induced by a parametrization of \mathcal{M} ; in this case, we denote the norm as $\|\cdot\|_{\mathcal{M}}$ and calculate it by using the metric, which we discuss below. The disjoint union of all the tangent spaces of \mathcal{M} is a $2d$ -dimensional C^1 manifold called the tangent bundle; it is denoted by $T\mathcal{M}$. A C^k vector field, also known as a section of the tangent bundle, is a C^k map from $\mathcal{M} \rightarrow T\mathcal{M}$ assigning $x \mapsto v_x, v_x \in T_x\mathcal{M}$.

We only consider Riemannian manifolds that satisfy particular geometric constraints; specifically, we require that \mathcal{M} be closed and connected and that its volume and curvature not be too large. We describe the Riemannian structure of \mathcal{M} in Section 3.3.1. The d -dimensional volume can be defined in terms of the Hausdorff measure, which is discussed in Section 3.3.2. The curvature of \mathcal{M} is measured through a quantity known as the reach; we define it and discuss its relationship to other geometric aspects of \mathcal{M} in Section 3.3.3. We end the background material on manifolds with Section 3.3.4, in which we discuss a special class of charts consisting of orthogonal projections from \mathcal{M} onto the tangent spaces; the projections are performed within a ball of suitably small radius.

3.3.1 Metric and Connection

\mathcal{M} inherits the Riemannian structure of \mathbb{R}^n and has an induced metric $g_{\mathcal{M}}(x) : T_x\mathcal{M} \times T_x\mathcal{M} \rightarrow \mathbb{R}$ varying C^1 smoothly with $x \in \mathcal{M}$. If (W, ψ) is a local parametrization of

\mathcal{M} , and \tilde{x} is such that $\psi(\tilde{x}) = x$, the metric at x is represented by the positive-definite matrix $(D\psi)^\top|_{\tilde{x}}D\psi|_{\tilde{x}}$. If $v, \tilde{v} \in T_x\mathcal{M}$, their inner product is given by $v^\top g_{\mathcal{M}}(x)\tilde{v}$, which allows us to calculate quantities that depend on the length of tangent vectors and the angle between two tangent vectors. Let $\|\cdot\|_{\mathcal{M}}$ be the norm defined through this inner product; i.e., $\|v\|_{\mathcal{M}} = (v^\top g_{\mathcal{M}}(x)v)^{1/2}$. For a piecewise C^1 curve $\gamma : [t_0, t_1] \rightarrow \mathcal{M}$, its length $L_{\mathcal{M}}(\gamma)$ is defined as follows:

$$L_{\mathcal{M}}(\gamma) := \int_{t_0}^{t_1} \|\gamma'(t)\|_{\mathcal{M}} dt.$$

If γ is not contained in the image of a single local parametrization, a partition of unity can be used to define $g_{\mathcal{M}}(x)$ in the desired region. \mathcal{M} can be given a metric space structure by considering the length of piecewise C^1 curves connecting two given points of \mathcal{M} . The distance $d_{\mathcal{M}}(x, y)$ between $x, y \in \mathcal{M}$ is defined as:

$$d_{\mathcal{M}}(x, y) := \inf\{L_{\mathcal{M}}(\gamma) \mid \gamma : [t_0, t_1] \rightarrow \mathcal{M} \text{ is piecewise } C^1, \gamma(t_0) = x, \gamma(t_1) = y\}.$$

Unless stated otherwise, we assume that any curve is arclength-parametrized, in which case its tangent vectors have unit length. Note that in the above we can also work directly with the embedding of $T_x\mathcal{M}$ in \mathbb{R}^n . In this case, the columns of $D\psi$ form a basis for $T_x\mathcal{M}$, and $(D\psi)v$ gives the coordinates of v with respect to the standard basis of \mathbb{R}^n . Computations can be performed using the standard inner product.

\mathcal{M} also has an induced connection $\nabla : C^1(T\mathcal{M}) \times C^1(T\mathcal{M}) \rightarrow C^0(T\mathcal{M})$, where $C^k(T\mathcal{M})$ is the space of C^k smooth vector fields on \mathcal{M} . A vector field is a function assigning an element of the tangent space to each point of \mathcal{M} ; smoothness is measured componentwise. The connection is represented by the Christoffel symbols $\Gamma_{ij}^k(x)$. If $\{x_1, \dots, x_d\}$ is the standard coordinate system on $\mathbb{R}^d \supset W$,

$$\Gamma_{ij}^k(x) = \frac{1}{2} \sum_{l=1}^d g^{kl}(x) \left(\frac{\partial g_{jl}(x)}{\partial x_i} + \frac{\partial g_{il}(x)}{\partial x_j} - \frac{\partial g_{ij}(x)}{\partial x_l} \right),$$

where the elements of $g_{\mathcal{M}}(x)$ are written with subscripts and those of $g_{\mathcal{M}}^{-1}(x)$ are written with superscripts. The connection can be used to find the covariant derivative of a vector

field with respect to another vector field. Let e_i be the basis vector in the direction of x_i . Let u with components $\{u_1, \dots, u_d\}$ and v with components $\{v_1, \dots, v_d\}$ be two vector fields. Then,

$$\nabla_v u = \left(\sum_{j=1}^d v_j \frac{\partial u_k}{\partial x_j} + \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d u_i v_j \Gamma_{ij}^k \right) e_k.$$

The connection is also used to define a procedure that allows the comparison of vectors from different tangent spaces. In Euclidean space, this can be done directly by translating two vectors to the same origin. However, this is not true in the general case: for two distinct points x and y on \mathcal{M} , $T_x \mathcal{M}$ and $T_y \mathcal{M}$ are different vector spaces. One way of comparing their elements is by using parallel transport. For $v \in T_y \mathcal{M}$, let $v_x^{\text{par}} : \gamma \rightarrow C^1(T\mathcal{M})$ be the function constructing a vector field along an arclength-parametrized curve $\gamma : [t_0, t_1] \rightarrow \mathcal{M}$ connecting x and y satisfying the following: $v_x^{\text{par}}(\gamma(t_0)) = v$, $\nabla_{\dot{\gamma}} v_x^{\text{par}}(\gamma(t)) = 0$ for all $t \in [t_0, t_1]$, and $v_x^{\text{par}} := v_x^{\text{par}}(\gamma(t_1)) \in T_x \mathcal{M}$. In other words, for a curve γ starting at y and ending at x , v_x^{par} translates v to $T_x \mathcal{M}$ along γ so that it stays parallel to itself.

Geodesics are a special class of curves that parallel transport their own tangent vectors. In other words, $\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = 0$. Given $x \in \mathcal{M}$ and $v \in T_x \mathcal{M}$, differential equations existence and uniqueness theorems show that there is a geodesic $\gamma_{x,v}$ with $\gamma_{x,v}(0) = x$ and $\dot{\gamma}_{x,v}(0) = v$. The exponential map $\exp_x : T_x \mathcal{M} \rightarrow \mathcal{M}$ is the map assigning $v \mapsto \gamma_{x,v}(1)$. If \exp_x is defined on all of $T_x \mathcal{M}$ for all $x \in \mathcal{M}$, \mathcal{M} is called geodesically complete. The Hopf-Rinow theorem states that \mathcal{M} is a complete metric space if and only if it is geodesically complete. (We assume \mathcal{M} is a compact subset of \mathbb{R}^n , so it is complete). This can be shown to imply that any two points $x, y \in \mathcal{M}$ can be joined by a geodesic whose length is $d_{\mathcal{M}}(x, y)$. In general, there can be multiple geodesics joining two points of \mathcal{M} , not all of which are length-minimizing. However, if the distance between two points is smaller than a quantity called the injectivity radius, there is a unique length-minimizing geodesic connecting them. The injectivity radius is the largest value such that the exponential map is a diffeomorphism on an open ball of this size at each point of \mathcal{M} . In Theorem 34 iii), we quote a result stating that the injectivity radius is at least as large as the reach, which means that parallel transport along geodesics

is well-defined within the charts that we are working in.

3.3.2 Volume

The d -dimensional volume of \mathcal{M} can be computed through the volume form induced by the metric. In this chapter, we do not use differential forms or explicitly define the volume form. Instead, we use the Hausdorff measure \mathcal{H}^d , which is equivalent for d -dimensional submanifolds of \mathbb{R}^n ; additionally, \mathcal{H}^0 is the counting measure and \mathcal{H}^n is equivalent to \mathcal{L}^n . The Hausdorff measure of a set S is defined in terms of its covering by a countable union of open sets. Let $S \subset \bigcup_i A_i$ and $\text{diam} A_i \leq \varepsilon$. Then,

$$\mathcal{H}^d(S) := \liminf_{\varepsilon \rightarrow 0} \frac{\omega_d}{2^d} \sum_i (\text{diam} A_i)^d,$$

where $\omega_d := \pi^{d/2} / \Gamma(d/2 + 1)$ is the volume of the d -dimensional unit ball. If S is a subset of a d -dimensional submanifold of a Euclidean space, we also write $\text{vol}_d(S)$ for $\mathcal{H}^d(S)$.

\mathcal{H}^d is a Borel measure, which means that Borel sets and continuous functions are \mathcal{H}^d -measurable. The standard measure-theoretic treatment of the integral suffices for the integration of real-valued functions with respect to the Hausdorff measure. The following theorem is useful when \mathcal{M} is parametrized by Lipschitz mappings.

Theorem 33 (Theorem 3.2.5, Federer, 1969). *Let $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^n$ be an injective Lipschitz mapping, and let $g : \mathbb{R}^n \rightarrow \mathbb{R}$. For $S \subset \mathbb{R}^n$, if its preimage $\psi^{-1}(S)$ is \mathcal{L}^d -measurable,*

$$\int_S g(x) d\mathcal{H}^d(x) = \int_{\psi^{-1}(S)} (g \circ \psi)(x') \sqrt{\det D\psi^\top D\psi} d\mathcal{L}^d(x').$$

3.3.3 Reach

Let $\Pi_{\mathcal{M}} : \mathbb{R}^n \rightarrow 2^{\mathcal{M}}$ be the function that assigns to $y \in \mathbb{R}^n$ its nearest point(s) as a subset of \mathcal{M} . If y has a unique nearest point, we also use $\Pi_{\mathcal{M}} : \mathbb{R}^n \rightarrow \mathcal{M}$ to denote the projection onto \mathcal{M} . Due to the compactness of \mathcal{M} and the continuity and boundedness of the Euclidean distance function, there exists at least one point in $\Pi_{\mathcal{M}}(y)$ for every $y \in \mathbb{R}^n$. Define the

tubular neighborhood of \mathcal{M} of radius $r \in [0, \infty)$ as $\mathcal{M}_r := \bigcup_{x \in \mathcal{M}} B_n(x, r)$. The reach is essentially defined as the size of the largest tubular neighborhood whose points all have unique projections onto \mathcal{M} ; that is,

$$\text{reach}(\mathcal{M}) := \sup\{r' \mid \#\Pi_{\mathcal{M}}(y) = 1 \forall y \in \mathcal{M}_{r'}\}.$$

The reach can arise from both global and local properties of \mathcal{M} . If $\text{reach}(\mathcal{M}) = \tau$, then one of the following is true: there exists $z \in \mathbb{R}^n$ that is the midpoint of the line connecting $x, y \in \Pi_{\mathcal{M}}(z)$ and $\|x - y\| = 2\tau$; or, there exists a geodesic γ on \mathcal{M} such that $\|\gamma''(t)\| = 1/\tau$ at a point $\gamma(t) \in \mathcal{M}$ (Aamari et al., 2017). The reach can also be used to bound other geometric quantities associated with \mathcal{M} , such as the difference between the geodesic and Euclidean distances and the principal angles between two nearby tangent spaces. The following theorem summarizes a few important observations.

Theorem 34. *Let \mathcal{M} be a d -dimensional C^2 submanifold of \mathbb{R}^n with reach τ . Then, the following statements hold.*

- i) (Federer's reach condition, Theorem 4.18, Federer, 1959) The distance from a point y on \mathcal{M} to the tangent space at a point $x \in \mathcal{M}$ can be bounded using the following characterization of the reach:*

$$\frac{1}{\tau} = \sup_{\substack{x, y \in \mathcal{M} \\ x \neq y}} 2 \frac{\|y - \Pi_x y\|}{\|x - y\|^2}.$$

- ii) Let $x \in \mathcal{M}$ and $\tilde{y} \in T_x \mathcal{M}$ such that $\|x - \tilde{y}\| < \tau$. Let $y \in \mathcal{M}$ be the preimage of \tilde{y} under Π_x . Then,*

$$\|y - \tilde{y}\| \leq \tau - \sqrt{\tau^2 - \|x - \tilde{y}\|^2}; \quad \|x - y\| \leq \sqrt{2} \sqrt{\tau^2 - \sqrt{\tau^4 - \tau^2 \|x - \tilde{y}\|^2}}.$$

- iii) (Proposition A.1 ii), Aamari et al., 2017) \mathcal{M} has injectivity radius $\rho \geq \pi\tau$.*

- iv) (Lemma 2.5, Boissonnat et al., 2017) Let $x, y \in \mathcal{M}$ such that $\|x - y\| < 2\tau$. Then, the geodesic distance between x and y is bounded as*

$$d_{\mathcal{M}}(x, y) \leq 2\tau \arcsin \frac{\|x - y\|}{2\tau}.$$

v) (Lemma 3.5, [Boissonnat et al., 2017](#)) Let $x, y \in \mathcal{M}$ such that $\|x - y\| < \tau$. The largest principal angle between the tangent spaces at x and y has the bound:

$$\angle_1(T_x\mathcal{M}, T_y\mathcal{M}) \leq \frac{d_{\mathcal{M}}(x, y)}{\tau}.$$

Proof. The proofs of **i)**, **iii)**, **iv)**, and **v)** are given in the cited sources. **ii)** is a consequence of **i)** as follows. By Federer's reach condition, $\|y - \tilde{y}\| \leq \|x - y\|^2/(2\tau)$. Then, since x , \tilde{y} , and y form a right triangle, $\|x - \tilde{y}\| \geq \sqrt{\|x - y\|^2 - \|x - y\|^4/(4\tau^2)}$. Solving this inequality gives the bound for $\|x - y\|$, and plugging back into the reach condition gives the bound for $\|y - \tilde{y}\|$. \blacksquare

3.3.4 Charts

The charts that we make the most use of are the projections onto the tangent spaces, which are performed on regions with sufficiently small diameter. Assume that we are working in a coordinate system where $x \in \mathcal{M}$ is the origin and $T_x\mathcal{M}$ lies in \mathbb{R}^d . Define the cylinder $U_{\tilde{\tau}}^x := B_d(x, \tilde{\tau}) \times B_{n-d}(x, \tilde{\tau})$, where $B_d(x, \tilde{\tau}) \subset T_x\mathcal{M}$ and $B_{n-d}(x, \tilde{\tau}) \subset T_x^\perp\mathcal{M}$. Let $\mathcal{A}_{x, \tilde{\tau}} := U_{\tilde{\tau}}^x \cap T_x\mathcal{M}$ and $\tilde{\mathcal{A}}_{x, \tilde{\tau}} := U_{\tilde{\tau}}^x \cap \mathcal{M}$. Let $\tilde{\tau} < \tau_0 := \tau/C_d$, where C_d is a constant greater than one. (Note that [Theorem 34 iii\)](#) implies that we can parallel transport vectors between the tangent spaces of any two points of $\tilde{\mathcal{A}}_{x, \tilde{\tau}}$. The collection of charts (i.e., the atlas) that we use is $\left\{G_{x, U_{\tilde{\tau}}^x}^{-1} \mid x \in \mathcal{M}, \tilde{\tau} < \tau_0\right\}$, where $G_{x, U_{\tilde{\tau}}^x}^{-1} : \tilde{\mathcal{A}}_{x, \tilde{\tau}} \rightarrow \mathcal{A}_{x, \tilde{\tau}}$ is the orthogonal projection from \mathcal{M} onto $T_x\mathcal{M}$ restricted to $\tilde{\mathcal{A}}_{x, \tilde{\tau}}$. There exists a C^2 function $F_{x, U_{\tilde{\tau}}^x} : \mathcal{A}_{x, \tilde{\tau}} \rightarrow T_x^\perp\mathcal{M}$ such that $G_{x, U_{\tilde{\tau}}^x} : \mathcal{A}_{x, \tilde{\tau}} \rightarrow \tilde{\mathcal{A}}_{x, \tilde{\tau}}$ assigning $y \mapsto (y, F_{x, U_{\tilde{\tau}}^x}(y))$ is a local parametrization of the manifold. The existence of $F_{x, U_{\tilde{\tau}}^x}$ and $G_{x, U_{\tilde{\tau}}^x}$ is guaranteed by the inverse function theorem; this is discussed in detail in [Claim 1 of Fefferman et al. \(2016\)](#), which states that as long as τ_0 is sufficiently small, the $\{F_{x, U_{\tilde{\tau}}^x}\}$ are C^2 functions with $C^{1,1}$ norms uniformly bounded above by C_d/τ (i.e., $\sup_{x \in \mathcal{M}, \tilde{\tau} < \tau_0} \text{Lip}(DF_{x, U_{\tilde{\tau}}^x}) \leq C_d/\tau$), where C_d depends at most on d . We give a proof of this claim in [Lemma 35 i\)](#) and [ii\)](#) and show that C_d can be taken to be $3\sqrt{d}$.

The subspace angle bound from [Theorem 34 v\)](#) can be used to show that $G_{x, U_{\tilde{\tau}}^x}$ is approximately a local isometry. In parts [iii\)](#) and [iv\)](#) of the next lemma and [Corollary 36](#), we

bound the geodesic distance between two points contained in the same chart from the collection $\left\{G_{x,U_{\tilde{\tau}}}^{-1} \mid x \in \mathcal{M}, \tilde{\tau} < \tau_0\right\}$ as well as the volume of a region contained in a single chart. Throughout this chapter, we will make repeated use of the fact that working in a subset of \mathcal{M} contained in $U_{\tilde{\tau}}^x$ is very close to working in $B_d(x, \tilde{\tau})$, thus justifying our treatment of $C^{1,1}(\mathcal{M})$ regression as locally equivalent to $C^{1,1}(\mathbb{R}^d)$ regression.

Lemma 35. *Let $x \in \mathcal{M}$, and assume $\tilde{\tau} < \tau/2$. Let $G_{x,U_{\tilde{\tau}}}^{-1} : \tilde{\mathcal{A}}_{x,\tilde{\tau}} \rightarrow \mathcal{A}_{x,\tilde{\tau}}$ be the orthogonal projection from $\mathcal{M} \cap U_{\tilde{\tau}}^x$ onto $T_x\mathcal{M}$. Then, the following statements hold.*

i) $G_{x,U_{\tilde{\tau}}}^{-1}$ is a bijection; in fact, it is a C^2 diffeomorphism. Its inverse, $G_{x,U_{\tilde{\tau}}} : \mathcal{A}_{x,\tilde{\tau}} \rightarrow \tilde{\mathcal{A}}_{x,\tilde{\tau}}$ is a local parametrization of \mathcal{M} that maps $y \in T_x\mathcal{M}$ to $(y, F_{x,U_{\tilde{\tau}}}(y)) \in \mathcal{M}$ for a function $F_{x,U_{\tilde{\tau}}} : \mathcal{A}_{x,\tilde{\tau}} \rightarrow T_x^\perp\mathcal{M}$. Let $O_{U_{\tilde{\tau}}} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a rigid motion translating x to the origin and rotating $T_x\mathcal{M}$ to lie in \mathbb{R}^d . Decompose \mathbb{R}^n as $\mathbb{R}^d \times \mathbb{R}^{n-d}$, and let $\Pi_d : \mathbb{R}^n \rightarrow \mathbb{R}^d$ and $\Pi_{n-d} : \mathbb{R}^n \rightarrow \mathbb{R}^{n-d}$ be the projections onto the first and second factors, respectively. Then, there exists a finite set $\Psi := \{\psi_i : \mathbb{R}^d \supset W_i \rightarrow U_i \subset \mathbb{R}^n\}$ of parametrizations of \mathcal{M} whose images cover $\mathcal{M} \cap U_{\tilde{\tau}}^x$ and where each domain W_i is an open set of \mathbb{R}^d where the inverse of $\Pi_d \circ O_{U_{\tilde{\tau}}} \circ G_{x,U_{\tilde{\tau}}}^{-1} \circ \psi_i$ exists. $G_{x,U_{\tilde{\tau}}}$ assigns $y \mapsto \left(\psi \circ \left(\Pi_d \circ O_{U_{\tilde{\tau}}} \circ G_{x,U_{\tilde{\tau}}}^{-1} \circ \psi\right)^{-1}\right)(y)$, where $\psi \in \Psi$ such that $G_{x,U_{\tilde{\tau}}}^{-1} \circ \psi$ contains y in its image. $F_{x,U_{\tilde{\tau}}}$ can be defined as the C^2 function $\Pi_{n-d} \circ O_{U_{\tilde{\tau}}} \circ G_{x,U_{\tilde{\tau}}}$.

ii) Let $x' \in \mathcal{A}_{x,\tilde{\tau}}$, and set $\tilde{x}' := G_{x,U_{\tilde{\tau}}}(x')$. Then, the singular values of $DG_{x,U_{\tilde{\tau}}}^{-1}|_{\tilde{x}'}$ are $\{\cos \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})\}$, and those of $DF_{x,U_{\tilde{\tau}}}|_{x'}$ and $DG_{x,U_{\tilde{\tau}}}|_{x'}$ are $\{\tan \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})\}$ and $\{\sec \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})\}$, respectively. This leads to the following bounds:

$$0 \leq \inf_{\substack{x \in \mathcal{M} \\ x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_d(DF_{x,U_{\tilde{\tau}}}|_{x'}) \leq \sup_{\substack{x \in \mathcal{M} \\ x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_1(DF_{x,U_{\tilde{\tau}}}|_{x'}) \leq \frac{7\tilde{\tau}}{6\tau};$$

and,

$$1 \leq \inf_{\substack{x \in \mathcal{M} \\ x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_d(DG_{x,U_{\tilde{\tau}}}|_{x'}) \leq \sup_{\substack{x \in \mathcal{M} \\ x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_1(DG_{x,U_{\tilde{\tau}}}|_{x'}) \leq \left(1 + \frac{49\tilde{\tau}^2}{36\tau^2}\right)^{1/2}.$$

Uniform bounds on the operator and Frobenius norms follow immediately. Furthermore, the Lipschitz constants of $DF_{x,U_{\tilde{\tau}}^x}$ and $DG_{x,U_{\tilde{\tau}}^x}$ are equal, and

$$\sup_{x \in \mathcal{M}} \text{Lip}(DF_{x,U_{\tilde{\tau}}^x}) \leq \frac{3\sqrt{d}}{\tau}.$$

iii) Let $x_1, x_2 \in \tilde{\mathcal{A}}_{x,\tilde{\tau}}$, and let $\gamma : [t_0, t_1] \rightarrow \mathcal{M}$ be a C^1 curve connecting x_1 and x_2 . Then,

$$L_{\mathcal{A}_{x,\tilde{\tau}}}(G_{x,U_{\tilde{\tau}}^x}^{-1} \circ \gamma) \leq L_{\mathcal{M}}(\gamma) \leq \left(1 + \frac{C_d \tilde{\tau}^2}{\tau^2}\right)^{1/2} L_{\mathcal{A}_{x,\tilde{\tau}}}(G_{x,U_{\tilde{\tau}}^x}^{-1} \circ \gamma).$$

iv) Let $S \subset \tilde{\mathcal{A}}_{x,\tilde{\tau}}$ such that $G_{x,U_{\tilde{\tau}}^x}^{-1}(S)$ is \mathcal{L}^d -measurable. For $g : \mathcal{M} \rightarrow [0, \infty)$,

$$\begin{aligned} \int_{G_{x,U_{\tilde{\tau}}^x}^{-1}(S)} (g \circ G_{x,U_{\tilde{\tau}}^x})(x') d\mathcal{L}^d(x') &\leq \int_S g(x) d\mathcal{H}^d(x) \\ &\leq \left(1 + \frac{C_d \tilde{\tau}^2}{\tau^2}\right)^{d/2} \int_{G_{x,U_{\tilde{\tau}}^x}^{-1}(S)} (g \circ G_{x,U_{\tilde{\tau}}^x})(x') d\mathcal{L}^d(x'). \end{aligned}$$

Proof. i) $G_{x,U_{\tilde{\tau}}^x}^{-1}$ is the restriction of the C^∞ function $V_x V_x^\top : \mathbb{R}^n \rightarrow T_x \mathcal{M}$ to \mathcal{M} ; by composition with a chart, $G_{x,U_{\tilde{\tau}}^x}^{-1}$ is seen to be C^2 . We actually apply the inverse function theorem to a slightly different function. \mathcal{M} is a C^2 submanifold of \mathbb{R}^n so for any $\tilde{y} \in \mathcal{M}$ we can choose a set $U_{\tilde{y}} \subset \mathbb{R}^n$ containing \tilde{y} , a set $W_{\tilde{y}} \subset \mathbb{R}^d$, and C^2 functions $\psi_{\tilde{y}} : W_{\tilde{y}} \rightarrow U_{\tilde{y}}$ and $\phi_{\tilde{y}} : U_{\tilde{y}} \rightarrow W_{\tilde{y}}$ such that $\mathcal{M} \cap U_{\tilde{y}} = \psi_{\tilde{y}}(W_{\tilde{y}})$ and $\phi_{\tilde{y}} \circ \psi_{\tilde{y}} = \text{id}$. For every $\tilde{y} \in \tilde{\mathcal{A}}_{x,\tilde{\tau}}$, $\Pi_d \circ O_{U_{\tilde{\tau}}^x} \circ G_{x,U_{\tilde{\tau}}^x}^{-1} \circ \psi_{\tilde{y}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a composition of C^2 functions and thus C^2 . Its derivative has the same rank as $DG_{x,U_{\tilde{\tau}}^x}^{-1}|_{\tilde{y}}$, which is the projection of the derivative of $V_x V_x^\top$ onto $T_{\tilde{y}} \mathcal{M}$. $V_x V_x^\top$ is an orthogonal projection onto a linear subspace, i.e., a linear function, so its derivative is $V_x V_x^\top$; therefore, $DG_{x,U_{\tilde{\tau}}^x}^{-1}|_{\tilde{y}} = V_{\tilde{y}} V_{\tilde{y}}^\top V_x V_x^\top$. Note that $V_{\tilde{y}} \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$, $V_x^\top \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^d)$, and $\text{rank}(V_{\tilde{y}}) = \text{rank}(V_x^\top) = d$, so by the properties of rank we have $\text{rank}(V_{\tilde{y}} V_{\tilde{y}}^\top V_x V_x^\top) = \text{rank}(V_{\tilde{y}}^\top V_x)$, which is equal to d if $\sigma_d(V_{\tilde{y}}^\top V_x) > 0$.

The singular value in question is the cosine of the largest principal angle between $V_{\tilde{y}}$ and V_x , implying that $DG_{x,U_{\tilde{\tau}}^x}^{-1}$ is of full rank if the principal angles between $T_x \mathcal{M}$ and the tangent space to \mathcal{M} at any $\tilde{y} \in \tilde{\mathcal{A}}_{x,\tilde{\tau}}$ are strictly less than $\pi/2$. An application of Theorem 34 ii), iv),

and **v)** shows that this must hold:

$$\begin{aligned}
\sup_{x \in \mathcal{M}} \sup_{\tilde{y} \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}} \angle_1(T_x \mathcal{M}, T_{\tilde{y}} \mathcal{M}) &\leq \sup_{x \in \mathcal{M}} \sup_{\tilde{y} \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}} \frac{d_{\mathcal{M}}(x, \tilde{y})}{\tau} \\
&\leq \sup_{\tilde{\tau} < \tau_0} 2 \arcsin \frac{\sqrt{1 - \sqrt{1 - \tilde{\tau}^2/\tau^2}}}{\sqrt{2}} \\
&< \frac{2}{3}.
\end{aligned}$$

With the choice of a $\psi_{\tilde{y}}$ for which $\psi_{\tilde{y}}(0) = \tilde{y}$ and $\phi_{\tilde{y}}(\tilde{y})$ is a fixed point, the hypotheses of the inverse function theorem hold. Theorem 43 shows that $\left(\Pi_d \circ O_{U_{\tilde{\tau}}^x} \circ G_{x, U_{\tilde{\tau}}^x}^{-1} \circ \psi_{\tilde{y}}\right)^{-1}$ exists within a neighborhood $\tilde{W}_{\tilde{y}} \subset W_{\tilde{y}}$ of $\phi_{\tilde{y}}(\tilde{y})$. Let $\tilde{U}_{\tilde{y}} := \psi_{\tilde{y}}(\tilde{W}_{\tilde{y}})$ so that $\{\tilde{U}_{\tilde{y}}\}$ is an open cover of $(U_{\tilde{\tau}}^x \cup \partial U_{\tilde{\tau}}^x) \cap \mathcal{M}$, which is a compact set. Let $\{\tilde{V}_{\tilde{y}}\}$ be the projections (which are open sets) of the $\{\tilde{U}_{\tilde{y}}\}$ onto $T_x \mathcal{M}$. We can choose a finite set $\{\tilde{y}_i\}$ such that $\{\tilde{U}_{\tilde{y}_i}\}$ covers $\tilde{\mathcal{A}}_{x, \tilde{\tau}}$ and $\{\tilde{V}_{\tilde{y}_i}\}$ covers $\mathcal{A}_{x, \tilde{\tau}}$. $G_{x, U_{\tilde{\tau}}^x}$ can be defined as a map with restrictions $G_{x, U_{\tilde{\tau}}^x}|_{\tilde{V}_{\tilde{y}_i}} := \psi_{\tilde{y}_i} \circ \left(\Pi_d \circ O_{U_{\tilde{\tau}}^x} \circ G_{x, U_{\tilde{\tau}}^x}^{-1} \circ \psi_{\tilde{y}_i}\right)^{-1}$. (This defines $G_{x, U_{\tilde{\tau}}^x}$ uniquely). $G_{x, U_{\tilde{\tau}}^x}$ is well defined: inverses are unique and it is well-known that $G_{x, U_{\tilde{\tau}}^x}^{-1}$ is bijective on all of $U_{\tilde{\tau}}^x$ (Niyogi et al., 2008). $G_{x, U_{\tilde{\tau}}^x}$ is of smoothness class C^2 because the local inverses are composed of C^2 functions, and they are defined on open sets and agree on the intersections of these open sets. We can also say that $\mathcal{M} \cap U_{\tilde{\tau}}^x$ is the union of a finite number of graphs $\left(y, F_{x, U_{\tilde{\tau}}^x}|_{\tilde{V}_{\tilde{y}_i}}(y)\right)$, where $F_{x, U_{\tilde{\tau}}^x}|_{\tilde{V}_{\tilde{y}_i}} : \mathbb{R}^d \rightarrow \mathbb{R}^{n-d}$ assigns $y \mapsto \Pi_{n-d} \circ O_{U_{\tilde{\tau}}^x} \circ \psi_{\tilde{y}_i} \circ \left(\Pi_d \circ O_{U_{\tilde{\tau}}^x} \circ G_{x, U_{\tilde{\tau}}^x}^{-1} \circ \psi_{\tilde{y}_i}\right)^{-1}(y)$. $F_{x, U_{\tilde{\tau}}^x}$ is defined as the unique function with these restrictions. Clearly, $F_{x, U_{\tilde{\tau}}^x}$ is C^2 . Let $\Psi := \{\psi_{\tilde{y}_i}|_{\tilde{W}_{\tilde{y}_i}}\}$. Our convention will be to write $F_{x, U_{\tilde{\tau}}^x}$ and $G_{x, U_{\tilde{\tau}}^x}$ without reference to the particular $\psi_{\tilde{y}_i}$; instead, we write ψ , which can be any element of Ψ such that $G_{x, U_{\tilde{\tau}}^x}^{-1} \circ \psi$ contains y in its image.

ii) $G_{x, U_{\tilde{\tau}}^x}$ and $G_{x, U_{\tilde{\tau}}^x}^{-1}$ can be viewed as maps between $\tilde{\mathcal{A}}_{x, \tilde{\tau}} \subset \mathcal{M}$ and $\mathcal{A}_{x, \tilde{\tau}} \subset T_x \mathcal{M}$ considered as submanifolds of \mathbb{R}^n . Let $\tilde{x}' := G_{x, U_{\tilde{\tau}}^x}(x')$. Then, $DG_{x, U_{\tilde{\tau}}^x}|_{x'}$ and $DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}'}$ are maps between $T_{\tilde{x}'} \tilde{\mathcal{A}}_{x, \tilde{\tau}}$ and $T_{x'} \mathcal{A}_{x, \tilde{\tau}}$. Note that $G_{x, U_{\tilde{\tau}}^x}^{-1}$ is the restriction to $\tilde{\mathcal{A}}_{x, \tilde{\tau}}$ of the projection from $\mathbb{R}^n \rightarrow \mathcal{A}_{x, \tilde{\tau}}$ that assigns $y \mapsto V_x V_x^\top y$, where V_x is a matrix whose columns form an orthonormal basis for $T_x \mathcal{M}$. Since this is a linear function, its derivative is $V_x V_x^\top$; its projection onto $T_{\tilde{x}'} \tilde{\mathcal{A}}_{x, \tilde{\tau}}$ (and therefore the derivative of $G_{x, U_{\tilde{\tau}}^x}^{-1}$ at \tilde{x}') is $V_{\tilde{x}'} V_{\tilde{x}'}^\top V_x V_x^\top$. Then, $DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}'} : T_{\tilde{x}'} \tilde{\mathcal{A}}_{x, \tilde{\tau}} \rightarrow$

$T_{x'}\mathcal{A}_{x,\tilde{\tau}}$ is the map assigning $\tilde{v} \mapsto (V_{\tilde{x}'}V_{\tilde{x}'}^\top V_x V_x^\top)^\top \tilde{v}$, and $DG_{x,U_{\tilde{\tau}}^x}|_{x'}$ is its inverse. $F_{x,U_{\tilde{\tau}}^x}$ is the projection of $G_{x,U_{\tilde{\tau}}^x}$ onto $T_x^\perp\mathcal{M}$. Therefore, $DF_{x,U_{\tilde{\tau}}^x}|_{x'}$ is a map from $T_{x'}\mathcal{A}_{x,\tilde{\tau}} \rightarrow T_{F_{x,U_{\tilde{\tau}}^x}(x')}T_x^\perp\mathcal{M}$.

We now show that the singular values of $DG_{x,U_{\tilde{\tau}}^x}|_{x'}$ are a function of the principal angles between $T_{\tilde{x}'}\mathcal{M}$ and $T_x\mathcal{M}$. By the singular-value decomposition, we can write $DG_{x,U_{\tilde{\tau}}^x}|_{x'} = V_{\tilde{x}'}D_x^{\tilde{x}'}V_x^\top$, where the columns of $V_{\tilde{x}'}$ and V_x form orthonormal bases for $T_{\tilde{x}'}\mathcal{M}$ and $T_x\mathcal{M}$, respectively, and $D_x^{\tilde{x}'}$ is a diagonal matrix whose diagonal consists of the singular values of $DG_{x,U_{\tilde{\tau}}^x}|_{x'}$ in decreasing order. Correspondingly, $DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'} = V_x D_x^{\tilde{x}'-1} V_{\tilde{x}'}^\top$. The singular values and vectors must satisfy $DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'} V_{\tilde{x}',i} = (D_x^{\tilde{x}'})_{ii}^{-1} V_{x,i}$. Since $DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'}$ is a projection, this implies that the columns of V_x are the projections onto $T_x\mathcal{M}$ of the columns of $V_{\tilde{x}'}$, normalized to be of unit length. The entries of $D_x^{\tilde{x}'-1}$ are the factors by which $DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'}$ contracts elements of the basis V_x , i.e., the cosines of the angles between $V_{\tilde{x}',i}$ and $V_{x,i}$. It is not hard to see that these are in fact the principal angles and vectors.

Recall the definition $\cos \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M}) := \sup_{v_i \in T_{\tilde{x}'}\mathcal{M}} \sup_{u_i \in T_x\mathcal{M}} u_i^\top v_i$, where we require that $\{u_i\}$ and $\{v_i\}$ be orthonormal bases. Write the inner product using the decomposition $v_i = DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'} v_i + \Pi_x^\perp v_i$. This shows that $u_i^\top v_i$ is maximized when u_i is the normalized projection of v_i onto $T_x\mathcal{M}$; that is,

$$\begin{aligned} DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'} v_i &= \left\| DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'} v_i \right\| u_i \\ &= \cos \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M}) u_i. \end{aligned}$$

Analogously, v_i also must be the normalized projection of u_i onto $T_{\tilde{x}'}\mathcal{M}$, which means $\left(DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'} \right)^\top u_i = V_{\tilde{x}'} V_{\tilde{x}'}^\top V_x V_x^\top u_i = \cos \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M}) v_i$. The previous two equations characterize singular values; since they are unique, $(D_x^{\tilde{x}'})_{ii}^{-1} = \cos \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})$. We can also (uniquely up to certain allowable transformations) take $V_{x,i}$ to be u_i and $V_{\tilde{x}',i}$ to be v_i . Clearly, this determines the singular values of $DG_{x,U_{\tilde{\tau}}^x}|_{x'}$ as well; most importantly, the operator norm $\|DG_{x,U_{\tilde{\tau}}^x}|_{x'}\|_2$ is equal to $\sec \angle_1(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})$. The singular values of $DF_{x,U_{\tilde{\tau}}^x}|_{x'}$ also have related expressions. They are $\{\|DF_{x,U_{\tilde{\tau}}^x}|_{x'} v\| \|v\|^{-1} \mid v \in V_x\}$, which are the tangents of the principal angles. To see this, write $\|DF_{x,U_{\tilde{\tau}}^x}|_{x'} v\|$ as $\|V_x V_x^\top \tilde{v} - \tilde{v}\|$ for $\tilde{v} := DG_{x,U_{\tilde{\tau}}^x}|_{x'} v$. This is equal to $\|\tilde{v}\| \sin \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})$, and since $\|v\| = \|\tilde{v}\| \cos \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})$, we have

$\sigma_i(DF_{x,U_{\tilde{\tau}}^x}|_{x'}) = \tan \angle_i(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})$. Of course, by the properties of sec and tan, each singular value of $DG_{x,U_{\tilde{\tau}}^x}|_{x'}$ satisfies $\sigma_i(DG_{x,U_{\tilde{\tau}}^x}|_{x'}) = \left(1 + \sigma_i(DF_{x,U_{\tilde{\tau}}^x}|_{x'})^2\right)^{1/2}$.

We will use the preceding discussion to bound the singular values of $DF_{x,U_{\tilde{\tau}}^x}|_{x'}$ and $DG_{x,U_{\tilde{\tau}}^x}|_{x'}$ from above. However, we first note the lower bound

$$\begin{aligned} \inf_{x \in \mathcal{M}} \inf_{x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_d(DG_{x,U_{\tilde{\tau}}^x}|_{x'}) &= \inf_{x \in \mathcal{M}} \inf_{\substack{x' \in \mathcal{A}_{x,\tilde{\tau}} \\ v \in T_{x'}\mathcal{A}_{x,\tilde{\tau}}}} \frac{\|DG_{x,U_{\tilde{\tau}}^x}|_{x'}v\|}{\|v\|} \\ &= \inf_{x \in \mathcal{M}} \inf_{\substack{x' \in \mathcal{A}_{x,\tilde{\tau}} \\ \tilde{v} \in T_{\tilde{x}'}\mathcal{M}}} \frac{\|\tilde{v}\|}{\|DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'}\tilde{v}\|} \\ &\geq 1, \end{aligned}$$

which holds because $DG_{x,U_{\tilde{\tau}}^x}|_{\tilde{x}'}$ is a composition of projections. The bound is achieved when $T_{x'}\mathcal{M} = T_x\mathcal{M}$, in which case $DF_{x,U_{\tilde{\tau}}^x}|_{x'} = 0$. This also gives $\inf_{x \in \mathcal{M}} \inf_{x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_d(DF_{x,U_{\tilde{\tau}}^x}|_{x'}) \geq 0$.

For the largest singular values of $DF_{x,U_{\tilde{\tau}}^x}|_{x'}$ we have, using the bound on $\angle_1(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M})$ from Theorem 34 v),

$$\begin{aligned} \sup_{x \in \mathcal{M}} \sup_{x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_1(DF_{x,U_{\tilde{\tau}}^x}|_{x'}) &= \sup_{x \in \mathcal{M}} \sup_{x' \in \mathcal{A}_{x,\tilde{\tau}}} \|DF_{x,U_{\tilde{\tau}}^x}|_{x'}\|_2 \\ &= \sup_{x \in \mathcal{M}} \sup_{\substack{x' \in \mathcal{A}_{x,\tilde{\tau}} \\ v \in T_{x'}\mathcal{A}_{x,\tilde{\tau}}}} \frac{\|DF_{x,U_{\tilde{\tau}}^x}|_{x'}v\|}{\|v\|} \\ &= \sup_{x \in \mathcal{M}} \sup_{x' \in \mathcal{A}_{x,\tilde{\tau}}} \tan \angle_1(T_{\tilde{x}'}\mathcal{M}, T_x\mathcal{M}) \\ &\leq \tan \left(2 \arcsin \frac{\sqrt{1 - \sqrt{1 - \tilde{\tau}^2/\tau^2}}}{\sqrt{2}} \right) \\ &\leq \frac{7\tilde{\tau}}{6\tau}. \end{aligned}$$

The second-to-last line is due to Theorem 34 ii), iv), and v); the last line holds because $\tilde{\tau}/\tau < 1/2$. Similarly, for $DG_{x,U_{\tilde{\tau}}^x}$,

$$\sup_{x \in \mathcal{M}} \sup_{x' \in \mathcal{A}_{x,\tilde{\tau}}} \sigma_1(DG_{x,U_{\tilde{\tau}}^x}|_{x'}) = \sup_{x \in \mathcal{M}} \sup_{x' \in \mathcal{A}_{x,\tilde{\tau}}} \|DG_{x,U_{\tilde{\tau}}^x}|_{x'}\|_2$$

$$\begin{aligned}
&= \sup_{x \in \mathcal{M}} \sup_{x' \in \mathcal{A}_{x, \tilde{\tau}}} \sec \angle_1(T_{\tilde{x}'} \mathcal{M}, T_x \mathcal{M}) \\
&= \sup_{x \in \mathcal{M}} \sup_{x' \in \mathcal{A}_{x, \tilde{\tau}}} \left(1 + \tan^2 \angle_1(T_{\tilde{x}'} \mathcal{M}, T_x \mathcal{M})\right)^{1/2} \\
&\leq \left(1 + \frac{49\tilde{\tau}^2}{36\tau^2}\right)^{1/2}.
\end{aligned}$$

Because the squared Frobenius norm is equal to the sum of the squared singular values, we have:

$$0 \leq \sup_{\substack{x \in \mathcal{M} \\ x' \in \mathcal{A}_{x, \tilde{\tau}}} \|DF_{x, U_{\tilde{\tau}}^x}|_{x'}\|_F \leq \frac{7\sqrt{d}\tilde{\tau}}{6\tau}; \quad \sqrt{d} \leq \sup_{\substack{x \in \mathcal{M} \\ x' \in \mathcal{A}_{x, \tilde{\tau}}} \|DG_{x, U_{\tilde{\tau}}^x}|_{x'}\|_F \leq \sqrt{d} \left(1 + \frac{49\tilde{\tau}^2}{36\tau^2}\right)^{1/2}.$$

The Lipschitz constant of $DG_{x, U_{\tilde{\tau}}^x}$ is defined as

$$\text{Lip}(DG_{x, U_{\tilde{\tau}}^x}) := \sup_{\substack{x_1, x_2 \in \mathcal{A}_{x, \tilde{\tau}} \\ x_1 \neq x_2}} \frac{\|DG_{x, U_{\tilde{\tau}}^x}|_{x_1} - DG_{x, U_{\tilde{\tau}}^x}|_{x_2}\|_F}{\|x_1 - x_2\|};$$

thus, it is necessary to find an upper bound for $\sigma_1(DG_{x, U_{\tilde{\tau}}^x}|_{x_1} - DG_{x, U_{\tilde{\tau}}^x}|_{x_2})$. (The lower bound is, of course, zero). By partitioning $DG_{x, U_{\tilde{\tau}}^x}$ as $\left[I|DF_{x, U_{\tilde{\tau}}^x}^\top\right]^\top$, it is clear that $\text{Lip}(DF_{x, U_{\tilde{\tau}}^x}) = \text{Lip}(DG_{x, U_{\tilde{\tau}}^x})$. Let $\tilde{x}_1 := G_{x, U_{\tilde{\tau}}^x}(x_1)$ and $\tilde{x}_2 := G_{x, U_{\tilde{\tau}}^x}(x_2)$. For $v \in T_x \mathcal{M}$, let $v_{\tilde{x}_1}$ and $v_{\tilde{x}_2}$ be such that $DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}_1} v_{\tilde{x}_1} = DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}_2} v_{\tilde{x}_2} = v$. Note that both $\|v_{\tilde{x}_1}\|$ and $\|v_{\tilde{x}_2}\|$ are no greater than $\|v\| \sup_{x \in \mathcal{M}, x' \in \mathcal{A}_{x, \tilde{\tau}}} \sigma_1(DG_{x, U_{\tilde{\tau}}^x}|_{x'})$. We find an upper bound for the supremum of $\|DG_{x, U_{\tilde{\tau}}^x}|_{x_1} v - DG_{x, U_{\tilde{\tau}}^x}|_{x_2} v\| = \|v_{\tilde{x}_1} - v_{\tilde{x}_2}\|$ by projecting $v_{\tilde{x}_2}$ onto $T_{\tilde{x}_1} \mathcal{M}$, bounding the distance of this point from $v_{\tilde{x}_1}$ and $v_{\tilde{x}_2}$, and using the triangle inequality. Let $\tilde{v}_{\tilde{x}_2} := V_{\tilde{x}_1} V_{\tilde{x}_1}^\top v_{\tilde{x}_2}$; $\|v_{\tilde{x}_2} - \tilde{v}_{\tilde{x}_2}\|$ is no greater than $\|V_{\tilde{x}_1} V_{\tilde{x}_1}^\top - V_{\tilde{x}_2} V_{\tilde{x}_2}^\top\|_2 \|v_{\tilde{x}_2}\|$. Note the following:

$$\begin{aligned}
\left\|DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}_1} \tilde{v}_{\tilde{x}_2} - DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}_1} v_{\tilde{x}_1}\right\| &= \left\|DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}_1} \tilde{v}_{\tilde{x}_2} - DG_{x, U_{\tilde{\tau}}^x}^{-1}|_{\tilde{x}_2} v_{\tilde{x}_2}\right\| \\
&= \|V_x V_x^\top (V_{\tilde{x}_1} V_{\tilde{x}_1}^\top - V_{\tilde{x}_2} V_{\tilde{x}_2}^\top) v_{\tilde{x}_2}\| \\
&\leq \|V_x V_x^\top\|_2 \|V_{\tilde{x}_1} V_{\tilde{x}_1}^\top - V_{\tilde{x}_2} V_{\tilde{x}_2}^\top\|_2 \|v_{\tilde{x}_2}\| \\
&\leq \|v_{\tilde{x}_2}\| \sin \angle_1(T_{\tilde{x}_1} \mathcal{M}, T_{\tilde{x}_2} \mathcal{M}) \\
&\leq \|v_{\tilde{x}_2}\| \sin \frac{d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)}{\tau}.
\end{aligned}$$

Then,

$$\begin{aligned} \|\tilde{v}_{\tilde{x}_2} - v_{\tilde{x}_1}\| &\leq \sigma_d \left(DG_{x, U_{\tilde{\tau}}}^{-1} \Big|_{\tilde{x}_1} \right)^{-1} \left\| DG_{x, U_{\tilde{\tau}}}^{-1} \Big|_{\tilde{x}_1} \tilde{v}_{\tilde{x}_2} - DG_{x, U_{\tilde{\tau}}}^{-1} \Big|_{\tilde{x}_1} v_{\tilde{x}_1} \right\| \\ &\leq \left(1 + \frac{49\tilde{\tau}^2}{36\tau^2} \right)^{1/2} \|v_{\tilde{x}_2}\| \sin \frac{d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)}{\tau}. \end{aligned}$$

Using the definition of the Lipschitz constant,

$$\begin{aligned} \sup_{x \in \mathcal{M}} \text{Lip}(DG_{x, U_{\tilde{\tau}}}) &\leq \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \mathcal{A}_{x, \tilde{\tau}}}} \sup_{\substack{v \in T_x \mathcal{M} \\ \|v\|=1}} \frac{\sqrt{d} \left\| DG_{x, U_{\tilde{\tau}}} \Big|_{x_1} v - DG_{x, U_{\tilde{\tau}}} \Big|_{x_2} v \right\|}{\|x_1 - x_2\|} \\ &\leq \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \mathcal{A}_{x, \tilde{\tau}}}} \sup_{\substack{v \in T_x \mathcal{M} \\ \|v\|=1}} \frac{\sqrt{d} (\|v_{\tilde{x}_1} - \tilde{v}_{\tilde{x}_2}\| + \|\tilde{v}_{\tilde{x}_2} - v_{\tilde{x}_2}\|)}{\|x_1 - x_2\|} \\ &\leq \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \mathcal{A}_{x, \tilde{\tau}}}} \sqrt{d} \left(1 + \frac{49\tilde{\tau}^2}{36\tau^2} + \left(1 + \frac{49\tilde{\tau}^2}{36\tau^2} \right)^{1/2} \right) \frac{\sin(d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)/\tau)}{\|x_1 - x_2\|} \\ &\leq \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \mathcal{A}_{x, \tilde{\tau}}}} \frac{5\sqrt{d} d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)}{2\tau \|x_1 - x_2\|}. \end{aligned}$$

The fourth line holds because $\tilde{\tau} < \tau/2$ and $\sin(d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)/\tau) \leq (d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)/\tau)$. In Corollary 36 i), we show that the ratio between the geodesic distance and the distance between the projections onto $T_x \mathcal{M}$ is bounded from above. We only use the fact that $G_{x, U_{\tilde{\tau}}}^{-1}$ has a C^2 inverse whose derivative has bounded operator norm, which we have already proven, so we apply the result here. This finally yields

$$\begin{aligned} \sup_{x \in \mathcal{M}} \text{Lip}(DG_{x, U_{\tilde{\tau}}}) &\leq \sup_{\tilde{\tau} < \tau_0} \frac{5\sqrt{d}}{2\tau} \left(1 + \frac{49\tilde{\tau}^2}{36\tau^2} \right)^{1/2} \\ &\leq \frac{3\sqrt{d}}{\tau}. \end{aligned}$$

iii) Let $\tilde{\gamma} : [\tilde{t}_0, \tilde{t}_1] \rightarrow \mathcal{A}_{x, \tilde{\tau}}$ be a C^1 curve connecting $G_{x, U_{\tilde{\tau}}}^{-1}(x_1)$ and $G_{x, U_{\tilde{\tau}}}^{-1}(x_2)$. Then, $G_{x, U_{\tilde{\tau}}} \circ \tilde{\gamma}$ is a C^1 curve on \mathcal{M} lying within $\tilde{\mathcal{A}}_{x, \tilde{\tau}}$, and

$$\begin{aligned} L_{\mathcal{M}}(G_{x, U_{\tilde{\tau}}} \circ \tilde{\gamma}) &= \int_{\tilde{t}_0}^{\tilde{t}_1} \left\| (G_{x, U_{\tilde{\tau}}} \circ \tilde{\gamma})'(t) \right\| dt \\ &= \int_{\tilde{t}_0}^{\tilde{t}_1} \left\| DG_{x, U_{\tilde{\tau}}} \Big|_{\tilde{\gamma}(t)} \tilde{\gamma}'(t) \right\| dt \end{aligned}$$

$$\begin{aligned}
&\leq \sup_{x \in \mathcal{M}, x' \in \mathcal{A}_{x, \tilde{\tau}}} \|DG_{x, U_{\tilde{\tau}}^x}|_{x'}\|_2 \int_{\tilde{t}_0}^{\tilde{t}_1} \|\tilde{\gamma}'(t)\| dt \\
&\leq \left(1 + \frac{C_d \tilde{\tau}^2}{\tau^2}\right)^{1/2} L_{\mathcal{A}_{x, \tilde{\tau}}}(\tilde{\gamma}).
\end{aligned}$$

Conversely,

$$\begin{aligned}
L_{\mathcal{M}}(G_{x, U_{\tilde{\tau}}^x} \circ \tilde{\gamma}) &\geq \inf_{x \in \mathcal{M}} \inf_{x' \in \mathcal{A}_{x, \tilde{\tau}}} \sigma_d(DG_{x, U_{\tilde{\tau}}^x}|_{x'}) \int_{\tilde{t}_0}^{\tilde{t}_1} \|\tilde{\gamma}'(t)\| dt \\
&\geq L_{\mathcal{A}_{x, \tilde{\tau}}}(\tilde{\gamma}).
\end{aligned}$$

iv) By Theorem 33, we have

$$\int_S g(x) d\mathcal{H}^d(x) = \int_{G_{x, U_{\tilde{\tau}}^x}^{-1}(S)} (g \circ G_{x, U_{\tilde{\tau}}^x})(x') \sqrt{\det(DG_{x, U_{\tilde{\tau}}^x}^\top|_{x'} DG_{x, U_{\tilde{\tau}}^x}|_{x'})} d\mathcal{L}^d(x').$$

The determinant in the above line is equal to the product of the squared singular values of $DG_{x, U_{\tilde{\tau}}^x}|_{x'}$. The result follows by applying part ii) of this lemma. \blacksquare

Corollary 36. *Let $x \in \mathcal{M}$, and let $\tilde{\tau} < \tau/2$. Then,*

i) *If $x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}$,*

$$\left\|G_{x, U_{\tilde{\tau}}^x}^{-1}(x_1) - G_{x, U_{\tilde{\tau}}^x}^{-1}(x_2)\right\| \leq d_{\mathcal{M}}(x_1, x_2) \leq \left(1 + \frac{C_d \tilde{\tau}^2}{\tau^2}\right)^{1/2} \left\|G_{x, U_{\tilde{\tau}}^x}^{-1}(x_1) - G_{x, U_{\tilde{\tau}}^x}^{-1}(x_2)\right\|.$$

ii) *Let ω_d be the volume of the d -dimensional unit ball. Then, the Hausdorff measure of $\tilde{\mathcal{A}}_{x, \tilde{\tau}}$ is bounded as*

$$\tilde{\tau}^d \omega_d \leq \mathcal{H}^d(\tilde{\mathcal{A}}_{x, \tilde{\tau}}) \leq \left(1 + \frac{C_d \tilde{\tau}^2}{\tau^2}\right)^{d/2} \tilde{\tau}^d \omega_d.$$

Proof. i) Since $\tilde{\tau} < \rho$ and \mathcal{M} is geodesically complete, the distance between x_1 and x_2 is realized by a unique geodesic γ^* . Take $\gamma_s \subset \mathcal{A}_{x, \tilde{\tau}}$ to be the straight line connecting $G_{x, U_{\tilde{\tau}}^x}^{-1}(x_1)$ and $G_{x, U_{\tilde{\tau}}^x}^{-1}(x_2)$. Then, using Theorem 35 iii),

$$L_{\mathcal{A}_{x, \tilde{\tau}}}(\gamma_s) = \inf_{\tilde{\gamma} \subset \mathcal{A}_{x, \tilde{\tau}}} L_{\mathcal{A}_{x, \tilde{\tau}}}(\tilde{\gamma}) \leq L_{\mathcal{A}_{x, \tilde{\tau}}}(G_{x, U_{\tilde{\tau}}^x}^{-1} \circ \gamma^*)$$

$$\begin{aligned}
&\leq L_{\mathcal{M}}(\gamma^*) = d_{\mathcal{M}}(x_1, x_2) \\
&\leq L_{\mathcal{M}}(G_{x, U_{\tilde{\tau}}} \circ \gamma_s) \leq \left(1 + \frac{C_d \tilde{\tau}^2}{\tau^2}\right)^{1/2} L_{\mathcal{A}_{x, \tilde{\tau}}}(\gamma_s).
\end{aligned}$$

ii) Take g to be the constant function 1 in Theorem 35 iv). ■

3.4 Norm of $C^{1,1}(\mathcal{M})$ Functions

The smoothness class of a function $f : \mathcal{M} \rightarrow \mathbb{R}$ is defined in terms of its composition with a chart. For example, f is in $C^{1,1}(\mathcal{M})$ if, for any $x \in \mathcal{M}$ and any chart (U, ϕ) in our atlas such that $x \in U$, the composition $f \circ \psi : \mathbb{R}^d \rightarrow \mathbb{R}$ is $C^{1,1}(\mathbb{R}^d)$ on $\phi(U)$. We write $\|\cdot\|_{C^{1,1}(\mathcal{M})}$ for the norm of a $C^{1,1}(\mathcal{M})$ function. As in the $C^{1,1}(\mathbb{R}^d)$ case, we start by writing $\|f\|_{C^{1,1}(\mathcal{M})}$ as the maximum of a norm and two seminorms:

$$\|f\|_{C^{1,1}(\mathcal{M})} := \max \left\{ \|f\|_{C^0(\mathcal{M})}, \|f\|_{\dot{C}^1(\mathcal{M})}, \|f\|_{\dot{C}^{1,1}(\mathcal{M})} \right\}.$$

In general, if we apply this norm to a function f , it is to be assumed that $f \in C^{1,1}(\mathcal{M})$. Of course, we define $\|f\|_{C^0(\mathcal{M})} := \sup_{x \in \mathcal{M}} |f|$. It is easy to see from first principles that $\|\cdot\|_{\dot{C}^1(\mathcal{M})}$ and $\|\cdot\|_{\dot{C}^{1,1}(\mathcal{M})}$ as defined below are seminorms and that $\|\cdot\|_{C^{1,1}(\mathcal{M})}$ is a norm. Alternatively, it follows by comparison with $C^{1,1}(\mathbb{R}^d)$ and noting that the calculations below are independent of the choice of chart.

We can characterize the $\dot{C}^1(\mathcal{M})$ seminorm of f as an upper bound on directional derivatives. Given a point x on \mathcal{M} and an arclength-parametrized geodesic $\gamma_{x,v}$ such that $\gamma_{x,v}(0) = x$ and $\gamma'_{x,v}(0) = v$, the directional derivative of f is defined as $(f \circ \gamma_{x,v})'(0)$. At each point on the manifold, this quantity is maximized in a particular direction; let $\text{grad}f$ be the vector field associating each $x \in \mathcal{M}$ with this direction scaled by the maximum rate of change. Then, we define:

$$\begin{aligned}
\|f\|_{\dot{C}^1(\mathcal{M})} &:= \sup_{x \in \mathcal{M}} \sup_{\substack{v \in T_x \mathcal{M} \\ \|v\|=1}} (f \circ \gamma_{x,v})'(0) \\
&= \sup_{x \in \mathcal{M}} \|\text{grad}f(x)\|.
\end{aligned}$$

Let ϕ be a chart whose domain contains x , and let $\tilde{x} := \phi(x)$. Then, $\text{grad}f(x) = g_{\mathcal{M}}^{-1}(x)\nabla(f \circ \psi)|_{\tilde{x}}$. Writing the metric in terms of the parametrization, we have the following :

$$\|f\|_{\dot{C}^1(\mathcal{M})} = \sup_{x \in \mathcal{M}} \left(\nabla(f \circ \psi)^\top|_{\tilde{x}} \left((D\psi)^\top|_{\tilde{x}} D\psi|_{\tilde{x}} \right)^{-\top} \nabla(f \circ \psi)|_{\tilde{x}} \right)^{1/2}.$$

Additionally, because \mathcal{M} is geodesically complete, we can express the Riemannian distance as the length of a geodesic and use the second fundamental theorem of calculus to show that the $\dot{C}^1(\mathcal{M})$ seminorm of f is also equal to the Lipschitz constant

$$\text{Lip}(f) := \sup_{x, y \in \mathcal{M}, x \neq y} \frac{|f(x) - f(y)|}{d_{\mathcal{M}}(x, y)},$$

which is analogous to the Euclidean case.

It is a slightly more involved process to define the $\dot{C}^{1,1}(\mathcal{M})$ seminorm of f . We want this to measure the Lipschitzness of $\text{grad}f$, essentially bounding the amount that directional derivatives of f can change from one point on the manifold to another. We can either use parallel transport or use the embedded version of the gradient. Let $\text{grad}f_x^{\text{par}}(y) : \gamma \rightarrow C^1(T\mathcal{M})$ be the parallel transport function constructing a vector field along an arclength-parametrized geodesic $\gamma : [t_0, t_1] \rightarrow \mathcal{M}$ connecting x and y satisfying the following: $\text{grad}f_x^{\text{par}}(y)(\gamma(t_0)) = \text{grad}f(y)$, $\nabla_{\dot{\gamma}} \text{grad}f_x^{\text{par}}(y)(\gamma(t)) = 0$ for all $t \in [t_0, t_1]$, and $\text{grad}f_x^{\text{par}}(y) := \text{grad}f_x^{\text{par}}(y)(\gamma(t_1)) \in T_x\mathcal{M}$. By Theorem 34 iii), this is a well-defined procedure within the charts that we are working in. Then, the $\dot{C}^{1,1}(\mathcal{M})$ seminorm can be defined as the maximum of the local (within the injectivity radius) Lipschitz constant

$$\text{Lip}_\rho(\text{grad}f) := \sup_{\substack{x \neq y \in \mathcal{M} \\ d_{\mathcal{M}}(x, y) \leq \rho}} \frac{\|\text{grad}f(x) - \text{grad}f_x^{\text{par}}(y)\|}{d_{\mathcal{M}}(x, y)}$$

and an upper bound on the Lipschitz constant for $x, y \in \mathcal{M}$ such that $d_{\mathcal{M}}(x, y) > \rho$; in the second case, the geodesic connecting x and y is not necessarily unique so the upper bound is crude and based on $\|f\|_{C^1(\mathcal{M})}$.

A simpler way to define the $\dot{C}^{1,1}(\mathcal{M})$ seminorm of f is with respect to the embedding in \mathbb{R}^n ; this is the approach we will take. Let $\widetilde{\text{grad}f}$ denote the embedded version of $\text{grad}f$.

Using the same local coordinates as above,

$$\widetilde{\text{grad}}f(x) = D\psi|_{\tilde{x}} \left((D\psi)^\top|_{\tilde{x}} D\psi|_{\tilde{x}} \right)^{-1} \nabla(f \circ \psi)|_{\tilde{x}}.$$

We define $\|f\|_{\dot{C}^{1,1}(\mathcal{M})} := \text{Lip}(\widetilde{\text{grad}}f)$, where

$$\text{Lip}(\widetilde{\text{grad}}f) := \sup_{x,y \in \mathcal{M}, x \neq y} \frac{\|\widetilde{\text{grad}}f(x) - \widetilde{\text{grad}}f(y)\|}{d_{\mathcal{M}}(x,y)}.$$

For our purposes, this is sufficient; to make this clearer, in the following lemma we bound the parallel transport of vectors from nearby tangent spaces and show that $\text{Lip}(\widetilde{\text{grad}}f)$ and the embedding of \mathcal{M} determine an upper bound for the local Lipschitz constant of $\text{grad}f$.

Lemma 37. *Let f be a function defined on \mathcal{M} such that $\|f\|_{\dot{C}^{1,1}(\mathcal{M})} \leq M^*$; i.e., $\sup_{x \in \mathcal{M}} |f| \leq M^*$, $\sup_{x \in \mathcal{M}} \|\text{grad}f(x)\| \leq M^*$, and $\text{Lip}(\widetilde{\text{grad}}f) \leq M^*$. For $x \in \mathcal{M}$ and $\tilde{\tau} < \tilde{\tau}_{max}$, where $\tilde{\tau}_{max} := \left(\max\left\{1/\tau_0, 8/\tau, \sup_{x \in \mathcal{M}} \text{Lip}\left(DF_{x, U_{\tau/2}^x}\right)\right\} \right)^{-1}$, define the local Lipschitz constant of $\text{grad}f$ over $\tilde{\mathcal{A}}_{x, \tilde{\tau}}$ as*

$$\text{Lip}_{\tilde{\mathcal{A}}_{x, \tilde{\tau}}}(\text{grad}f) := \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\|\text{grad}f(x_1) - \text{grad}f_{x_1}^{\text{par}}(x_2)\|}{d_{\mathcal{M}}(x_1, x_2)}.$$

Then,

$$\inf_{x \in \mathcal{M}} \text{Lip}_{\tilde{\mathcal{A}}_{x, \tilde{\tau}}}(\text{grad}f) \leq \left(1 + \frac{C_d}{\tau} + \frac{C_d \tilde{\tau}}{\tau^2}\right) M^*.$$

Proof. We work in the chart $G_{x_1, U_{2\tilde{\tau}}^{x_1}}^{-1}$; in this chart $DG_{x_1, U_{2\tilde{\tau}}^{x_1}}|_{x_1} = [I|0]^\top$, so, conveniently, the first d coordinates of $\widetilde{\text{grad}}f(x_1)$ are the d coordinates of $\text{grad}f(x_1)$ and the last $n-d$ coordinates are zero. We assume that $\tilde{\tau} < \tilde{\tau}_{max}$, where $\tilde{\tau}_{max} := \left(\max\left\{1/\tau_0, 8/\tau, \sup_{x \in \mathcal{M}} \text{Lip}\left(DF_{x, U_{\tau/2}^x}\right)\right\} \right)^{-1}$. This ensures that $\tilde{\tau}$ is small enough for x_2 to be in a valid, geodesically convex chart centered at x_1 and also for Lemma 35 to guarantee $d_{\mathcal{M}}(x_1, x_2) < (1 + C_d \tilde{\tau}^2 / \tau^2)^{1/2} (2\tilde{\tau}) \leq \rho$ so that the parallel transport of tangent vectors from $T_{x_2}\mathcal{M}$ to $T_{x_1}\mathcal{M}$ along the geodesic joining x_1 and x_2 is well-defined. We start by relating $\text{Lip}_{\tilde{\mathcal{A}}_{x, \tilde{\tau}}}(\text{grad}f)$ to an upper bound consisting of three components: the Lipschitz constant of $\widetilde{\text{grad}}f$, the influence of the parametrization, and the change in components due to parallel transport.

$$\text{Lip}_{\tilde{\mathcal{A}}_{x, \tilde{\tau}}}(\text{grad}f) = \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\|[I|0]^\top \text{grad}f(x_1) - [I|0]^\top \text{grad}f_{x_1}^{\text{par}}(x_2)\|}{d_{\mathcal{M}}(x_1, x_2)}$$

$$\begin{aligned}
&= \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} (d_{\mathcal{M}}(x_1, x_2))^{-1} \left\| \widetilde{\text{grad}}f(x_1) - \widetilde{\text{grad}}f(x_2) \right. \\
&\quad \left. + \widetilde{\text{grad}}f(x_2) - [I|0]^\top \text{grad}f(x_2) \right. \\
&\quad \left. + [I|0]^\top \text{grad}f(x_2) - [I|0]^\top \text{grad}f_{x_1}^{\text{par}}(x_2) \right\| \\
&\leq \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\left\| \widetilde{\text{grad}}f(x_1) - \widetilde{\text{grad}}f(x_2) \right\|}{d_{\mathcal{M}}(x_1, x_2)} \\
&\quad + \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\left\| \left(DG_{x_1, U_{2\tilde{\tau}}^{x_1}} \Big|_{G_{x_1, U_{2\tilde{\tau}}^{x_1}}^{-1}(x_2)} - [I|0]^\top \right) \text{grad}f(x_2) \right\|}{d_{\mathcal{M}}(x_1, x_2)} \\
&\quad + \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\left\| [I|0]^\top \text{grad}f(x_2) - [I|0]^\top \text{grad}f_{x_1}^{\text{par}}(x_2) \right\|}{d_{\mathcal{M}}(x_1, x_2)}.
\end{aligned}$$

The first term on the right-hand side of the inequality is less than or equal to $\text{Lip}(\widetilde{\text{grad}}f)$, which is bounded above by M^* by assumption. We deal with the second term as follows.

$$\begin{aligned}
&\sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\left\| \left(DG_{x_1, U_{2\tilde{\tau}}^{x_1}} \Big|_{G_{x_1, U_{2\tilde{\tau}}^{x_1}}^{-1}(x_2)} - [I|0]^\top \right) \text{grad}f(x_2) \right\|}{d_{\mathcal{M}}(x_1, x_2)} \\
&\leq \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\left\| \left(\left[I \Big| DF_{x_1, U_{2\tilde{\tau}}^{x_1}}^\top \Big|_{G_{x_1, U_{2\tilde{\tau}}^{x_1}}^{-1}(x_2)} \right] - [I|0]^\top \right) \text{grad}f(x_2) \right\|}{d_{\mathcal{M}}(x_1, x_2)} \\
&\leq \text{Lip} \left(DF_{x_1, U_{2\tilde{\tau}}^{x_1}}^\top \right) \times \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\|x_2 - x_1\|}{d_{\mathcal{M}}(x_1, x_2)} \times \sup_{x_2 \in \mathcal{M}} \|\text{grad}f(x_2)\| \\
&\leq \frac{C_d}{\tau} M^*.
\end{aligned}$$

It remains to bound the difference between $\text{grad}f(x_2)$ and its parallel transport to $T_{x_1}\mathcal{M}$ along the geodesic γ . Let v be the vector field that is the parallel transport of $\text{grad}f(x_2)$, and let (v^1, \dots, v^d) be the components of v with respect to the basis (e_1, \dots, e_d) defined by the columns of $DG_{x_1, U_{2\tilde{\tau}}^{x_1}} \Big|_{G_{x_1, U_{2\tilde{\tau}}^{x_1}}^{-1}(\gamma(t))}$. Assume that γ is parametrized by arclength, and let $\dot{\gamma}$ have components $(\dot{\gamma}^1, \dots, \dot{\gamma}^d)$. Recall that v must satisfy $\nabla_{\dot{\gamma}} v = 0$, where $\nabla_{\dot{\gamma}}$ denotes the

covariant derivative with respect to the tangent vector field of γ . This means the following must hold

$$\dot{v} - \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \dot{\gamma}^i v^j \Gamma_{ij}^k e_k = 0,$$

where Γ_{ij}^k are the Christoffel symbols of the connection ∇ . In local coordinates, in terms of the metric they are

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^d g^{kl} \left(\frac{\partial g_{jl}}{\partial x_i} + \frac{\partial g_{il}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_l} \right),$$

where the elements of $g_{\mathcal{M}}(\gamma(t))$ in matrix form are written with subscripts and those of $g_{\mathcal{M}}^{-1}(\gamma(t))$ are written with superscripts. Note that $\sup_{k,l} |g^{kl}| \leq \|g_{\mathcal{M}}^{-1}\|_2 \leq 1$. Additionally,

$$\sup_{i,j,l} \left| \frac{\partial g_{jl}}{\partial x_i} \right| \leq \left\| DF_{x_1, U_{2\tilde{\tau}}^{x_1}} \Big|_{G_{x_1, U_{2\tilde{\tau}}^{x_1}}^{-1}(\gamma(t))} \right\|_{\mathcal{F}} \text{Lip} \left(DF_{x_1, U_{2\tilde{\tau}}^{x_1}} \right),$$

implying that $\sup_{i,j,k} |\Gamma_{ij}^k| \leq C_d \tilde{\tau} / \tau^2$. Since γ is unit-speed, $|\dot{\gamma}^i| \leq 1$. So, $\|\dot{v}\| \leq C_d \sup_j |v^j| \sup_{i,j,k} |\Gamma_{ij}^k| \leq C_d \tilde{\tau} M^* / \tau^2$ for any x_2 in this chart. This implies that

$$\begin{aligned} \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\| [I|0]^\top \text{grad} f(x_2) - [I|0]^\top \text{grad} f_{x_1}^{\text{par}}(x_2) \|}{d_{\mathcal{M}}(x_1, x_2)} &\leq \sup_{x_1, x_2 \in \tilde{\mathcal{A}}_{x, \tilde{\tau}}, x_1 \neq x_2} \frac{\|\dot{v}\| d_{\mathcal{M}}(x_1, x_2)}{d_{\mathcal{M}}(x_1, x_2)} \\ &\leq \frac{C_d \tilde{\tau}}{\tau^2} M^*. \quad \blacksquare \end{aligned}$$

If $y \notin \tilde{\mathcal{A}}_{x, \tilde{\tau}_{max}}$, then $d_{\mathcal{M}}(x, y) \geq \tilde{\tau}_{max}$. Suppose $d_{\mathcal{M}}(x, y) < \rho$. Then,

$$\begin{aligned} \frac{\|\text{grad} f(x) - \text{grad} f_x^{\text{par}}(y)\|}{d_{\mathcal{M}}(x, y)} &\leq \frac{2 \sup_{x \in \mathcal{M}} \|\text{grad} f(x)\|}{d_{\mathcal{M}}(x, y)} \\ &\leq \frac{2}{\tilde{\tau}_{max}} M^*. \end{aligned}$$

If $d_{\mathcal{M}}(x, y) \geq \rho$, there may not be a unique geodesic connecting x and y , but the same bound holds regardless of the path because $\sup_{x \in \mathcal{M}} \|\text{grad} f\|$ is finite. Writing $\tilde{\tau}_{max}$ as $C_d \tau$, we have the following corollary.

Corollary 38. *Let f be a function defined on \mathcal{M} such that $\|f\|_{C^{1,1}(\mathcal{M})} \leq M^*$; then,*

$$\text{Lip}(\text{grad} f) \leq \max \left\{ \left(1 + \frac{C_{d,1}}{\tau} \right), \frac{C_{d,2}}{\tau} \right\} M^*.$$

If instead of assuming a bound on $\text{Lip}(\widetilde{\text{grad}}f)$ we had assumed that $\text{Lip}(\text{grad}f) \leq M^*$, we could have used identical arguments to those in Lemma 37 and Corollary 38 to show that $\text{Lip}(\widetilde{\text{grad}}f)$ is less than or equal to the same constant times M^* . This is summarized in the following theorem.

Theorem 39. *Let f be a function defined on \mathcal{M} such that*

$$\max \left\{ \sup_{x \in \mathcal{M}} |f(x)|, \sup_{x \in \mathcal{M}} \|\text{grad}f(x)\| \right\} \leq M^*.$$

Then, $\text{Lip}(\text{grad}f)$ has a finite upper bound if and only if $\text{Lip}(\widetilde{\text{grad}}f)$ does.

3.4.1 Pullback to Tangent Spaces

Our algorithm for $C^{1,1}(\mathcal{M})$ regression relies on the ability to locally perform $C^{1,1}(\mathbb{R}^d)$ regression on a real-valued function defined on subsets of specified tangent spaces. Given a function $f : \tilde{\mathcal{A}}_{x,\bar{\tau}} \rightarrow \mathbb{R}$, we can define its pullback to $\mathcal{A}_{x,\bar{\tau}}$ along any local diffeomorphism between $\tilde{\mathcal{A}}_{x,\bar{\tau}}$ and $\mathcal{A}_{x,\bar{\tau}}$. Projection onto the tangent spaces is one such diffeomorphism; denote the pullback of f to $\mathcal{A}_{x,\bar{\tau}}$ along $G_{x,U_{\bar{\tau}}^x}$ by $G_{x,U_{\bar{\tau}}^x}^*(f) := f \circ G_{x,U_{\bar{\tau}}^x}$. It is a requirement of the $C^{1,1}(\mathbb{R}^d)$ regression algorithm that the sample is generated by a function with finite $C^{1,1}(\mathbb{R}^d)$ norm. In the next lemma, we prove that $\|G_{x,U_{\bar{\tau}}^x}^*(f)\|_{C^{1,1}(\mathbb{R}^d)}$ is finite whenever $\|f\|_{C^{1,1}(\mathcal{M})}$ is.

Lemma 40. *Let $f^* : \mathcal{M} \rightarrow \mathbb{R}$ be a $C^{1,1}(\mathcal{M})$ function such that $\|f^*\|_{C^{1,1}(\mathcal{M})} \leq M^*$. Let $x \in \mathcal{M}$, and let $\bar{\tau} < \tau_0$. Let $g_{x,\bar{\tau}}^* : T_x\mathcal{M} \supset \mathcal{A}_{x,\bar{\tau}} \rightarrow \mathbb{R}$ be the function assigning $y \mapsto (f^* \circ G_{x,U_{\bar{\tau}}^x})(y)$. Then, the $C^{1,1}(\mathbb{R}^d)$ norms of $\{g_{x,\bar{\tau}}^* \mid x \in \mathcal{M}\}$ are uniformly bounded above by $M_{g,\bar{\tau}}^*$:*

$$\begin{aligned} \sup_{x \in \mathcal{M}} \|g_{x,\bar{\tau}}^*\|_{C^{1,1}(\mathbb{R}^d)} &\leq \left(1 + \frac{3\sqrt{d}}{\tau} + \frac{49\bar{\tau}^2}{36\tau^2} \right) M^* \\ &=: M_{g,\bar{\tau}}^* \end{aligned}$$

Proof. Clearly, $\|g_{x,\bar{\tau}}^*\|_{C^0(\mathbb{R}^d)} \leq \|f^*\|_{C^0(\mathcal{M})} \leq M^*$.

By the chain rule for maps between manifolds,

$$\begin{aligned} \sup_{y \in \mathcal{A}_{x, \tilde{\tau}}} \|\nabla g_{x, \tilde{\tau}}^*(y)\| &= \sup_{y \in \mathcal{A}_{x, \tilde{\tau}}} \left\| (DG_{x, U_{\tilde{\tau}}^x})^\top \Big|_y \nabla f^*(G_{x, U_{\tilde{\tau}}^x}(y)) \right\| \\ &\leq \sup_{y \in \mathcal{A}_{x, \tilde{\tau}}} \left\| (DG_{x, U_{\tilde{\tau}}^x})^\top \Big|_y \right\|_2 \times \sup_{x \in \mathcal{M}} \|\nabla f^*(x)\|. \end{aligned}$$

In the proof of Lemma 35, we showed that $\sup_{x \in \mathcal{M}, \tilde{y} \in \mathcal{A}_{x, \tilde{\tau}}} \left\| DG_{x, U_{\tilde{\tau}}^x} \Big|_{\tilde{y}} \right\|_2 \leq (1 + C_d^2 \tilde{\tau}^2 / \tau^2)^{1/2}$ for all $\tilde{\tau} < \tau_0$. Here, ∇f^* is the embedded version of the Riemannian gradient, i.e., $\widetilde{\text{grad} f^*}$. Therefore,

$$\|g_{x, \tilde{\tau}}^*\|_{C^1(\mathbb{R}^d)} \leq \left(1 + \frac{C_d^2 \tilde{\tau}^2}{\tau^2}\right)^{1/2} M^*.$$

The $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm of $g_{x, \tilde{\tau}}^*$ can be bounded as follows:

$$\begin{aligned} \text{Lip}(\nabla g_{x, \tilde{\tau}}^*) &= \sup_{y_1 \neq y_2 \in \mathcal{A}_{x, \tilde{\tau}}} \frac{\|\nabla g_{x, \tilde{\tau}}^*(y_1) - \nabla g_{x, \tilde{\tau}}^*(y_2)\|}{\|y_1 - y_2\|} \\ &\leq \sup_{y_1 \neq y_2 \in \mathcal{A}_{x, \tilde{\tau}}} \frac{\left\| (DG_{x, U_{\tilde{\tau}}^x})^\top \Big|_{y_1} \nabla f^*(G_{x, U_{\tilde{\tau}}^x}(y_1)) - (DG_{x, U_{\tilde{\tau}}^x})^\top \Big|_{y_1} \nabla f^*(G_{x, U_{\tilde{\tau}}^x}(y_2)) \right\|}{\|y_1 - y_2\|} \\ &\quad + \sup_{y_1 \neq y_2 \in \mathcal{A}_{x, \tilde{\tau}}} \frac{\left\| (DG_{x, U_{\tilde{\tau}}^x})^\top \Big|_{y_1} \nabla f^*(G_{x, U_{\tilde{\tau}}^x}(y_2)) - (DG_{x, U_{\tilde{\tau}}^x})^\top \Big|_{y_2} \nabla f^*(G_{x, U_{\tilde{\tau}}^x}(y_2)) \right\|}{\|y_1 - y_2\|}. \end{aligned}$$

Let $\tilde{y}_1 := G_{x, U_{\tilde{\tau}}^x}(y_1)$ and $\tilde{y}_2 := G_{x, U_{\tilde{\tau}}^x}(y_2)$. The first term on the right-hand side of the inequality is bounded above by

$$\left\| (DG_{x, U_{\tilde{\tau}}^x})^\top \Big|_{y_1} \right\|_2 \times \sup_{y_1 \neq y_2 \in \mathcal{A}_{x, \tilde{\tau}}} \frac{\left\| \widetilde{\text{grad} f^*}(G_{x, U_{\tilde{\tau}}^x}(y_1)) - \widetilde{\text{grad} f^*}(G_{x, U_{\tilde{\tau}}^x}(y_2)) \right\|}{d_{\mathcal{M}}(y_1, y_2)} \times \sup_{y_1 \neq y_2 \in \mathcal{A}_{x, \tilde{\tau}}} \frac{d_{\mathcal{M}}(\tilde{y}_1, \tilde{y}_2)}{\|y_1 - y_2\|},$$

which is no greater than $(1 + C_d \tilde{\tau}^2 / \tau^2) M^*$ by an application of Lemma 35 ii), Corollary 36 i), and the assumption that $\text{Lip}(\widetilde{\text{grad} f^*}) \leq M^*$. The second term is less than or equal to $\text{Lip}(DG_{x, U_{\tilde{\tau}}^x}^\top) \times \sup_{x \in \mathcal{M}} \|\nabla f^*(x)\|$, which is less than or equal to $C_d M^* / \tau$. Thus,

$$\|g_{x, \tilde{\tau}}^*\|_{\dot{C}^{1,1}(\mathbb{R}^d)} \leq \left(1 + \frac{C_d}{\tau} + \frac{C_d \tilde{\tau}^2}{\tau^2}\right) M^*.$$

Since our arguments were uniform in x , the result follows. ■

3.5 Tangent Space Estimation for Charts

In Lemma 40, we rely on having knowledge of the true tangent spaces of \mathcal{M} . More realistically, we have access only to a sample of points from \mathcal{M} and we must estimate other quantities, including the tangent spaces. The simplest way of doing so is by using local PCA. Start by forming the matrix $X_{x,\bar{\tau}}$ whose columns are the $N_{x,\bar{\tau}}$ sample points $B_n(x, \bar{\tau}) \cap \mathcal{X}$; then, the estimate of $T_x\mathcal{M}$ is the subspace $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ with orthonormal basis given by the eigenvectors of $N_{x,\bar{\tau}}^{-1}X_{x,\bar{\tau}}X_{x,\bar{\tau}}^\top$ corresponding to the d largest eigenvalues. This estimator is actually very close to $T_x\mathcal{M}$. For fixed $\bar{\tau}$, Theorem 22 states a finite sample bound on the sines of the principal angles between $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and $T_x\mathcal{M}$. The proof assumes that the probability measure on \mathcal{M} is uniform with respect to $\mathcal{H}^d(\mathcal{M})$. We restate this result in terms of the assumptions we place on \mathcal{P} . The only difference is in the derivation of a lower bound on the second moment of the pushforward of \mathcal{P} to $T_x\mathcal{M}$; in the present case, this needs to be adjusted by a factor of p_{min}/p_{max} .

Theorem 41. *Fix $x \in \mathcal{M}$ and $\bar{\tau} \leq \tau/2$. Let $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ be the subspace with orthonormal basis given by the eigenvectors corresponding to the d largest eigenvalues of $N_{x,\bar{\tau}}^{-1}X_{x,\bar{\tau}}X_{x,\bar{\tau}}^\top$, where $X_{x,\bar{\tau}}$ is the matrix whose columns are the $N_{x,\bar{\tau}}$ sample points $B_n(x, \bar{\tau}) \cap \mathcal{X}$. Let $\widehat{V_{x,\bar{\tau}}}$ and V_x be matrices whose columns are orthonormal bases for $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and $T_x\mathcal{M}$, respectively, and let $\sin \Theta(\widehat{V_{x,\bar{\tau}}}, V_x)$ be a diagonal matrix with entries given by $\sin \arccos \sigma_i(\widehat{V_{x,\bar{\tau}}}^\top V_x)$. Then,*

$$\mathbb{P} \left[\left\| \sin \Theta(\widehat{V_{x,\bar{\tau}}}, V_x) \right\|_F \leq 2 \left(\frac{\bar{\tau}}{\tau} + \frac{\bar{\tau}^2}{\tau^2} \right) \frac{(d+2)}{(1-\varepsilon_{p,1})} \left(1 + \frac{C_d \bar{\tau}^2}{\tau^2} \right)^{d/2} \frac{p_{max}}{p_{min}} \right] \geq 1 - 2d \exp \left(- \frac{\varepsilon_{p,1}^2 \bar{\tau}^2 N_{x,\bar{\tau}}}{2 \log 2 (d+2)} \left(1 + \frac{C_d \bar{\tau}^2}{\tau^2} \right)^{-d/2} \frac{p_{min}}{p_{max}} \right),$$

for all $\varepsilon_{p,1} \in [0, 1/2]$.

This probability bound is enough to imply convergence in probability when $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and $T_x\mathcal{M}$ are viewed as elements of the Grassmannian $\text{Grass}(d, n)$, the space of all d -dimensional linear subspaces of \mathbb{R}^n . $\text{Grass}(d, n)$ can be given a manifold structure where the geodesic

distance between $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and $T_x\mathcal{M}$ is $\left\| \Theta \left(\widehat{V_{x,\bar{\tau}}}, V_x \right) \right\|_F$. Note that $0 \leq \arccos \sigma_i \left(\widehat{V_{x,\bar{\tau}}}^\top V_x \right) \leq \pi/2$, so by the monotonicity of sine on this interval, $\widehat{T_{x,\bar{\tau}}\mathcal{M}} \xrightarrow{p} T_x\mathcal{M}$ for $\bar{\tau} \rightarrow 0$ and $N_{x,\bar{\tau}} \rightarrow \infty$.

We now define the requisite sets and functions for parametrizing the manifold using the estimated tangent space. Let the cylinder $\widehat{U_{\bar{\tau}}^x} := B_d(x, \bar{\tau}) \times B_{n-d}(x, \bar{\tau})$, where $B_d(x, \bar{\tau}) \subset \widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and $B_{n-d}(x, \bar{\tau}) \subset \widehat{T_{x,\bar{\tau}}^\perp\mathcal{M}}$. Let $\widehat{\mathcal{A}}_{x,\bar{\tau}} := \widehat{U_{\bar{\tau}}^x} \cap \widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and $\widetilde{\mathcal{A}}_{x,\bar{\tau}} := \widehat{U_{\bar{\tau}}^x} \cap \mathcal{M}$. Let $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1} : \widetilde{\mathcal{A}}_{x,\bar{\tau}} \rightarrow \widehat{\mathcal{A}}_{x,\bar{\tau}}$ be the orthogonal projection of \mathcal{M} onto $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$. If the inverse of $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1}$ is well-defined, let $F_{x,\widehat{U_{\bar{\tau}}^x}} : \widehat{\mathcal{A}}_{x,\bar{\tau}} \rightarrow \widehat{T_{x,\bar{\tau}}\mathcal{M}}$ be a function such that $G_{x,\widehat{U_{\bar{\tau}}^x}} : \widehat{\mathcal{A}}_{x,\bar{\tau}} \rightarrow \widetilde{\mathcal{A}}_{x,\bar{\tau}}$ assigning $y \mapsto (y, F_{x,\widehat{U_{\bar{\tau}}^x}}(y))$ is a local parametrization of the manifold. We want to use $\left\{ G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1} \mid x \in \mathcal{M}, \bar{\tau} < \bar{\tau}_{max} \right\}$ as a collection of approximately locally isometric charts for \mathcal{M} ; in Theorem 44, we prove that as long as $\bar{\tau} < \bar{\tau}_{max} := \left(\max \left\{ 1/\tau_0, 96(\sqrt{2})^{d/2}(d+2)/\tau, 5 \sup_{x \in \mathcal{M}} \text{Lip} \left(DF_{x,U_{\tau/2}^x} \right) \right\} \right)^{-1}$, there exist functions $F_{x,\widehat{U_{\bar{\tau}}^x}}$ and $G_{x,\widehat{U_{\bar{\tau}}^x}}$ with the properties needed to do so. In the remainder of the text, although our results are stated in terms of $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ defined through local PCA, they apply to any tangent space estimation technique for which the sine of the largest principal angle between $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and $T_x\mathcal{M}$ is bounded above by a constant ε_p (preferably decreasing in $\bar{\tau}/\tau$) with probability at least $1 - \delta_p$. Of course, for our purposes, we can take ε_p to be the error bound from Theorem 41 and δ_p the error probability. Before outlining the proof of Theorem 44, we include a preliminary lemma relating the geometry of $\widehat{U_{\bar{\tau}}^x}$ and $U_{\bar{\tau}}^x$ followed by a statement of the inverse function theorem.

Lemma 42. *Let $x \in \mathcal{M}$ and $\bar{\tau} < \bar{\tau}_{max}$. Let $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ be the subspace with orthonormal basis given by the eigenvectors corresponding to the d largest eigenvalues of $N_{x,\bar{\tau}}^{-1} X_{x,\bar{\tau}} X_{x,\bar{\tau}}^\top$, where $X_{x,\bar{\tau}}$ is the matrix whose columns are the $N_{x,\bar{\tau}}$ sample points $B_n(x, \bar{\tau}) \cap \{x_i\}$. Let ε_p be such that $\sin \angle_1 \left(\widehat{T_{x,\bar{\tau}}\mathcal{M}}, T_x\mathcal{M} \right) \leq \varepsilon_p$ with probability at least $1 - \delta_p$. Then, also with probability at least $1 - \delta_p$,*

- i) The preimage of $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1} \left(\widehat{\mathcal{A}}_{x,\bar{\tau}} \right)$ satisfies the following relationship with the preimage of the projection onto $T_x\mathcal{M}$:*

$$\widetilde{\mathcal{A}}_{x,\bar{\tau}} / (1 + \varepsilon_p \sqrt{2}) \subset \widetilde{\mathcal{A}}_{x,\bar{\tau}} \subset \widetilde{\mathcal{A}}_{x,\bar{\tau}} / (1 - \varepsilon_p \sqrt{2}).$$

ii) The largest principal angle between $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and the tangent space to \mathcal{M} at $y \in \widehat{\tilde{\mathcal{A}}_{x,\bar{\tau}}}$ is bounded as

$$\angle_1\left(\widehat{T_{x,\bar{\tau}}\mathcal{M}}, T_y\mathcal{M}\right) \leq \arcsin\left(\sin\frac{d_{\mathcal{M}}(x,y)}{\tau} + \varepsilon_p\right).$$

Proof. Assume that $\bar{\tau}$ and ε_p are small enough so that $\bar{\tau}/(1-\varepsilon_p\sqrt{2}) < \tau_0$ and $\sin(d_{\mathcal{M}}(x,y)/\tau) + \varepsilon_p < 1$.

i) Let $y \in \widehat{\tilde{\mathcal{A}}_{x,\bar{\tau}}}$, and let $\widehat{V_{x,\bar{\tau}}V_{x,\bar{\tau}}^\top} y$ be its projection onto $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$. Note that its difference from the projection of y onto $T_x\mathcal{M}$ is very small: $\left\|\widehat{V_{x,\bar{\tau}}V_{x,\bar{\tau}}^\top} y - V_x V_x^\top y\right\| \leq \left\|\widehat{V_{x,\bar{\tau}}V_{x,\bar{\tau}}^\top} - V_x V_x^\top\right\|_2 \|y\| \leq \left\|\sin\Theta\left(\widehat{V_{x,\bar{\tau}}}, V_x\right)\right\|_F \|y\| \leq \varepsilon_p \|y\|$. By the triangle inequality, $\|V_x V_x^\top y\| \leq \left\|\widehat{V_{x,\bar{\tau}}V_{x,\bar{\tau}}^\top} y\right\| + \left\|\widehat{V_{x,\bar{\tau}}V_{x,\bar{\tau}}^\top} y - V_x V_x^\top y\right\| \leq \bar{\tau} + \varepsilon_p \|y\|$. Since $\|y\| < \sqrt{2}\|V_x V_x^\top y\|$, we can solve the preceding inequality to show that $\|V_x V_x^\top y\| < \bar{\tau}/(1-\varepsilon_p\sqrt{2})$, implying that every point in $\widehat{\tilde{\mathcal{A}}_{x,\bar{\tau}}}$ also lies within $\tilde{\mathcal{A}}_{x,\bar{\tau}/(1-\varepsilon_p\sqrt{2})}$. To show the other side of the containment, let $y \in \tilde{\mathcal{A}}_{x,\bar{\tau}/(1+\varepsilon_p\sqrt{2})}$ and reverse the roles of $\|V_x V_x^\top y\|$ and $\left\|\widehat{V_{x,\bar{\tau}}V_{x,\bar{\tau}}^\top} y\right\|$ in the above argument.

ii) follows by applying the triangle inequality to the metric $(X, Y) \mapsto \|\sin\Theta(X, Y)\|_\infty$ defined on $\text{Grass}(d, n)$ and using Theorem 34 v) to bound $\sin\angle_1(T_x\mathcal{M}, T_y\mathcal{M})$. \blacksquare

Theorem 43 (Inverse Function Theorem, Theorem 3.3.2, Krantz and Parks, 2012). Let $\widehat{W} \subset \mathbb{R}^Q$ be an open set and let $G : \widehat{W} \rightarrow \mathbb{R}^Q$ be a mapping of class C^k , $k \geq 1$.

Let x^0 be a fixed point of \widehat{W} and assume that $\det DG(x^0) \neq 0$. Then there exists a neighborhood $W \subset \widehat{W}$ of x^0 such that

i) The restriction $G|_W$ is univalent.

ii) The set $V := G(W)$ is open.

iii) The inverse G^{-1} of $G|_W$ is of class C^k .

It is fairly easy to see that the inverse function theorem implies the invertibility of the projection onto $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ in a small (nonspecific) neighborhood of any point of \mathcal{M} where the

tangent spaces are not orthogonal to $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$. However, we want invertibility on an explicitly defined neighborhood; that is, we want to treat \mathcal{M} as the graph of a C^2 function from $\widehat{T_{x,\bar{\tau}}\mathcal{M}} \rightarrow \widehat{T_{x,\bar{\tau}}^\perp\mathcal{M}}$ on all of $\widehat{U_{\bar{\tau}}^x}$. For this reason, the proof of Theorem 44 is slightly technical in parts. In **i)**, we show the existence of the inverse of $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1}$. We demonstrate both injectivity and surjectivity by using the mean value theorem for vector-valued functions and showing that a failure of either condition leads to a contradiction with the bounds on $\|DG_{x,\widehat{U_{\bar{\tau}}^x}}\|_F$ and $\left\|\sin\Theta\left(\widehat{V_{x,\bar{\tau}}}, V_x\right)\right\|_F$. This shows that $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1}$ is bijective and thus has a unique inverse on $\widehat{U_{\bar{\tau}}^x}$.

In **ii)**, we use the inverse function theorem to show that $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1}$ is a C^2 diffeomorphism. By the definition of a C^2 submanifold, for every $\tilde{y} \in \widetilde{\mathcal{A}}_{x,\bar{\tau}}$ there exist sets $W \subset \mathbb{R}^d$ and $U \subset \mathbb{R}^n$, the latter containing \tilde{y} , and C^2 functions $\psi : W \rightarrow U$ and $\phi : U \rightarrow W$ such that ψ locally defines \mathcal{M} and $\phi \circ \psi$ is the identity. By composition with ψ , we see that $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1}$ is C^2 since it is the restriction of a smooth function on \mathbb{R}^n to \mathcal{M} . We show that its derivative has rank d because of the small angular distances between $T_x\mathcal{M}$ and $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ and between $T_x\mathcal{M}$ and $T_{\tilde{y}}\mathcal{M}$ for all $\tilde{y} \in \widetilde{\mathcal{A}}_{x,\bar{\tau}}$. This implies the local invertibility of $G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1}$ on a neighborhood $\tilde{U} \subset U$ of \tilde{y} . The neighborhoods \tilde{U} form an open cover of $\left(\widehat{U_{\bar{\tau}}^x} \cup \partial\widehat{U_{\bar{\tau}}^x}\right) \cap \mathcal{M}$, which is compact; they can be refined so that $G_{x,\widehat{U_{\bar{\tau}}^x}}$ is a union of a finite number of graphs.

Throughout most of this chapter, we do not strictly differentiate between \mathbb{R}^d and its inclusion into \mathbb{R}^n or its embedding as a tangent space to \mathcal{M} . Here we will be more precise. Let $O_{\widehat{U_{\bar{\tau}}^x}} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a rigid motion translating x to the origin and performing a rotation so that $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ lies in \mathbb{R}^d . If \mathbb{R}^n is decomposed as $\mathbb{R}^d \times \mathbb{R}^{n-d}$, let $\Pi_d : \mathbb{R}^n \rightarrow \mathbb{R}^d$ be the projection onto the first factor and $\Pi_{n-d} : \mathbb{R}^n \rightarrow \mathbb{R}^{n-d}$ the projection onto the second factor. At every $\tilde{y} \in \widetilde{\mathcal{A}}_{x,\bar{\tau}}$, we apply the inverse function theorem to the function $\Pi_d \circ O_{\widehat{U_{\bar{\tau}}^x}} \circ G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1} \circ \psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$. By the uniqueness of the inverse (up to coordinate changes on \mathbb{R}^d), we can define the C^2 function $F_{x,\widehat{U_{\bar{\tau}}^x}}$ on $\widehat{U_{\bar{\tau}}^x}$ as $\Pi_{n-d} \circ O_{\widehat{U_{\bar{\tau}}^x}} \circ \psi \circ \left(\Pi_d \circ O_{\widehat{U_{\bar{\tau}}^x}} \circ G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1} \circ \psi\right)^{-1}$; then, $G_{x,\widehat{U_{\bar{\tau}}^x}}$ is the C^2 function $\psi \circ \left(\Pi_d \circ O_{\widehat{U_{\bar{\tau}}^x}} \circ G_{x,\widehat{U_{\bar{\tau}}^x}}^{-1} \circ \psi\right)^{-1}$, which assigns $y \mapsto \left(y, F_{x,\widehat{U_{\bar{\tau}}^x}}(y)\right)$. The $C^1(\mathbb{R}^d, \mathbb{R}^{n-d})$ and $\dot{C}^{1,1}(\mathbb{R}^d, \mathbb{R}^{n-d})$ norms of $DF_{x,\widehat{U_{\bar{\tau}}^x}}$, and the corresponding norms of $DG_{x,\widehat{U_{\bar{\tau}}^x}}$, are not much

larger than the norms of $DF_{x, U_{\bar{\tau}}^x}$ and $DG_{x, U_{\bar{\tau}}^x}$, respectively, and we end by deriving upper bounds in **iii**).

Theorem 44. *Fix $x \in \mathcal{M}$ and $\bar{\tau} < \bar{\tau}_{max}$. Let $\widehat{T_{x, \bar{\tau}} \mathcal{M}}$ be the estimator of the tangent space to \mathcal{M} at x resulting from principal components analysis within the ball $B_n(x, \bar{\tau})$. Specifically, if $X_{x, \bar{\tau}}$ has columns consisting of the $N_{x, \bar{\tau}}$ sample points $B_n(x, \bar{\tau}) \cap \{x_i\}$, $\widehat{T_{x, \bar{\tau}} \mathcal{M}}$ has an orthonormal basis given by the eigenvectors of $N_{x, \bar{\tau}}^{-1} X_{x, \bar{\tau}} X_{x, \bar{\tau}}^T$ corresponding to the d largest eigenvalues. Let $\varepsilon_p < 1/12$ be such that, with probability at least $1 - \delta_p$, $\sin \angle_1(\widehat{T_{x, \bar{\tau}} \mathcal{M}}, T_x \mathcal{M}) \leq \varepsilon_p$. Let $G_{x, \widehat{U_{\bar{\tau}}^x}}^{-1} : \widehat{\mathcal{A}}_{x, \bar{\tau}} \rightarrow \widehat{\mathcal{A}}_{x, \bar{\tau}}$ be the orthogonal projection of \mathcal{M} onto $\widehat{T_{x, \bar{\tau}} \mathcal{M}}$. Then, the following hold on $\widehat{U_{\bar{\tau}}^x}$ with probability at least $1 - \delta_p$:*

i) $G_{x, \widehat{U_{\bar{\tau}}^x}}^{-1}$ is a one-to-one function. It possesses an inverse, $G_{x, \widehat{U_{\bar{\tau}}^x}} : \widehat{\mathcal{A}}_{x, \bar{\tau}} \rightarrow \widehat{\mathcal{A}}_{x, \bar{\tau}}$ that can be written as a local parametrization of \mathcal{M} ; specifically, $G_{x, \widehat{U_{\bar{\tau}}^x}}$ maps $y \in \widehat{T_{x, \bar{\tau}} \mathcal{M}}$ to $(y, F_{x, \widehat{U_{\bar{\tau}}^x}}(y)) \in \mathcal{M}$ for a function $F_{x, \widehat{U_{\bar{\tau}}^x}} : \widehat{\mathcal{A}}_{x, \bar{\tau}} \rightarrow \widehat{T_{x, \bar{\tau}}^{\perp} \mathcal{M}}$.

ii) Let $O_{\widehat{U_{\bar{\tau}}^x}} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a rigid motion translating x to the origin and rotating $\widehat{T_{x, \bar{\tau}} \mathcal{M}}$ to lie in \mathbb{R}^d . Decompose \mathbb{R}^n as $\mathbb{R}^d \times \mathbb{R}^{n-d}$, and let $\Pi_d : \mathbb{R}^n \rightarrow \mathbb{R}^d$ and $\Pi_{n-d} : \mathbb{R}^n \rightarrow \mathbb{R}^{n-d}$ be the projections onto the first and second factors, respectively. Then, there exists a finite set $\Psi := \{\psi_i : \mathbb{R}^d \supset W_i \rightarrow U_i \subset \mathbb{R}^n\}$ of parametrizations of \mathcal{M} whose images cover $\mathcal{M} \cap \widehat{U_{\bar{\tau}}^x}$ and where each domain W_i is an open set of \mathbb{R}^d where the inverse of $\Pi_d \circ O_{\widehat{U_{\bar{\tau}}^x}} \circ G_{x, \widehat{U_{\bar{\tau}}^x}}^{-1} \circ \psi_i$ exists. $G_{x, \widehat{U_{\bar{\tau}}^x}}$ is the C^2 diffeomorphism that assigns $y \mapsto \left(\psi \circ \left(\Pi_d \circ O_{\widehat{U_{\bar{\tau}}^x}} \circ G_{x, \widehat{U_{\bar{\tau}}^x}}^{-1} \circ \psi \right)^{-1} \right)(y)$, where $\psi \in \Psi$ such that $G_{x, \widehat{U_{\bar{\tau}}^x}}^{-1} \circ \psi$ contains y in its image. $F_{x, \widehat{U_{\bar{\tau}}^x}}$ can be defined as the C^2 function $\Pi_{n-d} \circ O_{\widehat{U_{\bar{\tau}}^x}} \circ G_{x, \widehat{U_{\bar{\tau}}^x}}$.

iii) The derivatives of $F_{x, \widehat{U_{\bar{\tau}}^x}}$ and $G_{x, \widehat{U_{\bar{\tau}}^x}}$ have operator norms bounded as follows:

$$\sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \left\| DF_{x, \widehat{U_{\bar{\tau}}^x}}|_{x'} \right\|_2 \leq \frac{8}{5} \left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p \right); \quad \sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \left\| DG_{x, \widehat{U_{\bar{\tau}}^x}}|_{x'} \right\|_2 \leq \left(1 + \frac{64}{25} \left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p \right)^2 \right)^{1/2}.$$

Furthermore, the Lipschitz constants of $DF_{x, \widehat{U_{\bar{\tau}}^x}}$ and $DG_{x, \widehat{U_{\bar{\tau}}^x}}$ are equal, and

$$\sup_{x \in \mathcal{M}} \text{Lip} \left(DF_{x, \widehat{U_{\bar{\tau}}^x}} \right) \leq \frac{7\sqrt{d}}{\tau}.$$

Proof. We start the proof of **i)** by assuming $G_{x, \widehat{U_{\bar{\tau}}}}^{-1}$ is not injective and show through the mean value theorem that this leads to a contradiction with the bound $\|DF_{x, U_{\bar{\tau}}}\|_F \leq C_d \bar{\tau} / \tau$. Let $x_1, x_2 \in \mathcal{M}$ be two distinct points in the preimage of $y \in \widehat{T_{x, \bar{\tau}} \mathcal{M}}$ under $G_{x, \widehat{U_{\bar{\tau}}}}^{-1}$; by Lemma 42 **i)**, they are both contained in $\tilde{\mathcal{A}}_{x, \bar{\tau} / (1 - \varepsilon_p \sqrt{2})}$. Note that $\bar{\tau} < \tau / (96(\sqrt{2})^{d/2}(d+2))$, and so $\varepsilon_p < (\bar{\tau} / \tau) (8(\sqrt{2})^{d/2}(d+2)) < 1/12$, which means $\tilde{\mathcal{A}}_{x, \bar{\tau} / (1 - \varepsilon_p \sqrt{2})} \subset \tilde{\mathcal{A}}_{x, 8\bar{\tau}/7}$. Let $\tilde{\gamma} : [\tilde{t}_0, \tilde{t}_1] \rightarrow T_x \mathcal{M}$ be the projection onto $\mathcal{A}_{x, \bar{\tau}}$ of the geodesic connecting x_1 and x_2 . Then, $\tilde{\gamma}$ is an arclength-parametrized C^1 curve connecting $G_{x, U_{\bar{\tau}}}^{-1}(x_1)$ and $G_{x, U_{\bar{\tau}}}^{-1}(x_2)$, which, since $\bar{\tau} / (1 - \varepsilon_p \sqrt{2}) < \tau_0$, are shown by Lemma 35 **i)** to be two distinct points of $T_x \mathcal{M}$. $G_{x, U_{\bar{\tau}}} \circ \tilde{\gamma}$ is a C^1 curve on \mathcal{M} lying within $\tilde{\mathcal{A}}_{x, 8\bar{\tau}/7}$. The mean value theorem states that there exists $\tilde{t} \in (\tilde{t}_0, \tilde{t}_1)$ such that

$$\|(G_{x, U_{\bar{\tau}}} \circ \tilde{\gamma})'(\tilde{t})\| \geq \frac{\|(G_{x, U_{\bar{\tau}}} \circ \tilde{\gamma})(\tilde{t}_0) - (G_{x, U_{\bar{\tau}}} \circ \tilde{\gamma})(\tilde{t}_1)\|}{|\tilde{t}_0 - \tilde{t}_1|}.$$

By Corollary 36 **i)** and the assumption that $\bar{\tau} < \tau / (5C_d)$, $\sup_{x' \in \tilde{\mathcal{A}}_{x, 8\bar{\tau}/7}} \|DG_{x, U_{\bar{\tau}}}|_{x'}\|_2 \leq \sqrt{20/19}$, showing $|\tilde{t}_0 - \tilde{t}_1| \leq \sqrt{20/19} \|G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2)\|$. On the left-hand side, we use the chain rule and the fact that $\|\tilde{\gamma}'(\tilde{t})\| = 1$ to see that $\|(G_{x, U_{\bar{\tau}}} \circ \tilde{\gamma})'(\tilde{t})\| \leq \|DG_{x, U_{\bar{\tau}}}|_{\tilde{\gamma}(\tilde{t})}\|_2 \|\tilde{\gamma}'(\tilde{t})\| \leq \sqrt{20/19}$. Thus,

$$\begin{aligned} \frac{20}{19} &\geq \frac{\|x_1 - x_2\|}{\|G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2)\|} \\ &= \sec \angle \left(x_1 - x_2, G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2) \right). \end{aligned}$$

Note that $G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2)$ is orthogonal to both $x_1 - G_{x, U_{\bar{\tau}}}^{-1}(x_1)$ and $x_2 - G_{x, U_{\bar{\tau}}}^{-1}(x_2)$; this means that $\angle \left(x_1 - x_2, G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2) \right)$ is in the interval $[0, \pi/2]$ and $\sec \angle \left(x_1 - x_2, G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2) \right) \geq 1$.

We now show that the bound on the principal angles between $\widehat{T_{x, \bar{\tau}} \mathcal{M}}$ and $T_x \mathcal{M}$ implies a lower bound on the secant that is greater than 20/19. Project $G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2)$ onto $\widehat{T_{x, \bar{\tau}} \mathcal{M}}$ through left-multiplication by $\widehat{V_{x, \bar{\tau}}} \widehat{V_{x, \bar{\tau}}}^\top$; this vector is orthogonal to $x_1 - x_2$ by construction. Using this together with the Cauchy-Schwarz inequality,

$$(x_1 - x_2)^\top \left(G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2) \right) = (x_1 - x_2)^\top \left(G_{x, U_{\bar{\tau}}}^{-1}(x_1) - G_{x, U_{\bar{\tau}}}^{-1}(x_2) \right)$$

$$\begin{aligned}
& -\widehat{V}_{x,\bar{\tau}}\widehat{V}_{x,\bar{\tau}}^\top \left(G_{x,U_{\bar{x}}}^{-1}(x_1) - G_{x,U_{\bar{x}}}^{-1}(x_2) \right) \\
& \leq \|x_1 - x_2\| \left\| \widehat{V}_{x,\bar{\tau}}\widehat{V}_{x,\bar{\tau}}^\top - V_x V_x^\top \right\|_2 \left\| G_{x,U_{\bar{x}}}^{-1}(x_1) - G_{x,U_{\bar{x}}}^{-1}(x_2) \right\|,
\end{aligned}$$

which yields

$$\begin{aligned}
\sec \angle \left(x_1 - x_2, G_{x,U_{\bar{x}}}^{-1}(x_1) - G_{x,U_{\bar{x}}}^{-1}(x_2) \right) &= \frac{\|x_1 - x_2\| \left\| G_{x,U_{\bar{x}}}^{-1}(x_1) - G_{x,U_{\bar{x}}}^{-1}(x_2) \right\|}{(x_1 - x_2)^\top \left(G_{x,U_{\bar{x}}}^{-1}(x_1) - G_{x,U_{\bar{x}}}^{-1}(x_2) \right)} \\
&\geq \left\| \widehat{V}_{x,\bar{\tau}}\widehat{V}_{x,\bar{\tau}}^\top - V_x V_x^\top \right\|_2^{-1} \\
&\geq \left\| \sin \Theta(\widehat{V}_{x,\bar{\tau}}, V_x) \right\|_F^{-1} > 12
\end{aligned}$$

with probability at least $1 - \delta_p$, a contradiction. Thus, there cannot exist points $x_1, x_2 \in \mathcal{M}$ lying in the same normal fiber of $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$, and $G_{x,U_{\bar{x}}}^{-1}$ is one-to-one.

We must also show that $G_{x,U_{\bar{x}}}^{-1}$ is surjective. Suppose there exists $y \in \widehat{T_{x,\bar{\tau}}\mathcal{M}}$ whose preimage under $G_{x,U_{\bar{x}}}^{-1}$ is empty, and assume that the closure and connectedness of \mathcal{M} are not violated. We will show that this contradicts the bounded curvature of \mathcal{M} . Note that $\{y\} \times B_{n-d}(x, r)$ intersects \mathcal{M} for some $r < \tau$. To see this, construct the cylinder $U_{\tilde{r}}^x$, where $\tilde{r} := 2\bar{\tau}/(\cos \arcsin \varepsilon_p) + 2\bar{\tau}\varepsilon_p < 3\bar{\tau}$; then, for every $v \in B_{n-d}(x, 1)$, $y + \alpha_v v \in (B_d(x, \tilde{r}) \times \partial B_{n-d}(x, \tilde{r}))$ for some α_v depending on v . Since the projection of $\mathcal{M} \cap U_{\tilde{r}}^x$ onto $T_x \mathcal{M}$ is the entire unit ball $B_d(x, \tilde{r})$, $(\{y\} \times B_{n-d}(x, r)) \cap \mathcal{M} \neq \emptyset$ for some $r \in [\bar{\tau}, \tilde{r} \sec \arcsin \varepsilon_p]$. Let $y'' \in (\{y\} \times B_{n-d}(x, r)) \cap \mathcal{M}$, $y' := (y'' - y) \cap (\{y\} \times \partial B_{n-d}(x, \bar{\tau}))$, and \tilde{y}' and \tilde{y}'' be the projections of y' and y'' , respectively, onto $T_x \mathcal{M}$. We see that

$$\begin{aligned}
\|y'' - \tilde{y}''\| &\geq \|y'' - y\| - \|\tilde{y}'' - y\| \\
&\geq \|y' - y\| - \left\| \widehat{V}_{x,\bar{\tau}}\widehat{V}_{x,\bar{\tau}}^\top - V_x^\top V_x \right\|_2 \|y''\| \\
&\geq \bar{\tau} \left(1 - \sqrt{11} \varepsilon_p \right).
\end{aligned}$$

Similarly, identifying x with the origin, $\|\tilde{y}'' - x\| \leq \|y - x\| + \|y - \tilde{y}''\| \leq \bar{\tau}(1 + \sqrt{11}\varepsilon_p)$. Let $\tilde{\gamma} : [\tilde{t}_0, \tilde{t}_1] \rightarrow T_x \mathcal{M}$ be the projection onto $T_x \mathcal{M}$ of the geodesic connecting x and y'' . Then, $\tilde{\gamma}$ is an arclength-parametrized C^1 curve connecting x and \tilde{y}'' . $G_{x,U_{\bar{x}}} \circ \tilde{\gamma}$ is a C^1 curve on \mathcal{M} .

Since $\varepsilon_p < 1/12$, $\bar{\tau}(1 + \sqrt{11}\varepsilon_p) < 4/3$, and $x, \tilde{y}'' \in \mathcal{A}_{x, 4\bar{\tau}/3}$. The mean value theorem states that there exists $\tilde{t} \in (\tilde{t}_0, \tilde{t}_1)$ such that

$$\begin{aligned} \left\| (G_{x, U_{\bar{\tau}}^x} \circ \tilde{\gamma})'(\tilde{t}) \right\| &\geq \frac{\|y'' - x\|}{\sqrt{14/13}\|\tilde{y}'' - x\|} \\ &\geq \frac{(\|y'' - \tilde{y}''\|^2 + \|x - \tilde{y}''\|^2)^{1/2}}{\sqrt{14/13}\|\tilde{y}'' - x\|} \\ &\geq \sqrt{\frac{13}{14}} \left(1 + \frac{(1 - \sqrt{11}\varepsilon_p)^2}{(1 + \sqrt{11}\varepsilon_p)^2} \right)^{1/2} \geq \frac{11}{10}. \end{aligned}$$

However, this contradicts $\left\| (G_{x, U_{\bar{\tau}}^x} \circ \tilde{\gamma})'(\tilde{t}) \right\| \leq \sup_{x' \in \tilde{\mathcal{A}}_{x, 4\bar{\tau}/3}} \|DG_{x, U_{\bar{\tau}}^x}|_{x'}\|_2 \leq \sqrt{14/13}$. This proves that every point in $\widehat{T_{x, \bar{\tau}}\mathcal{M}}$ has a point of \mathcal{M} in its preimage under $G_{x, U_{\bar{\tau}}^x}^{-1}$. Therefore, $G_{x, U_{\bar{\tau}}^x}^{-1}$ is a bijection, and its inverse $G_{x, U_{\bar{\tau}}^x}$ is well-defined; furthermore, by decomposing \mathbb{R}^n as $\widehat{T_{x, \bar{\tau}}\mathcal{M}} \times \widehat{T_{x, \bar{\tau}}^\perp\mathcal{M}}$, $G_{x, U_{\bar{\tau}}^x}(y)$ can be written in the form $(y, F_{x, U_{\bar{\tau}}^x}(y))$ for a function $F_{x, U_{\bar{\tau}}^x} : \widehat{\mathcal{A}}_{x, \bar{\tau}} \rightarrow \widehat{T_{x, \bar{\tau}}^\perp\mathcal{M}}$.

We now proceed to derive the smoothness properties stated in **ii**). $\widehat{V_{x, \bar{\tau}}V_{x, \bar{\tau}}^\top} : \mathbb{R}^n \rightarrow \widehat{T_{x, \bar{\tau}}\mathcal{M}}$ is an orthogonal projection onto a linear subspace, so, viewed as a function on the ambient space, it is C^∞ with derivative $\widehat{V_{x, \bar{\tau}}V_{x, \bar{\tau}}^\top}$. $G_{x, U_{\bar{\tau}}^x}^{-1}$ is the restriction of this function to \mathcal{M} , so it is at most of the same smoothness class as \mathcal{M} . Specifically, \mathcal{M} is a C^2 submanifold of \mathbb{R}^n so for any $\tilde{y} \in \mathcal{M}$ we can by definition choose a set $U_{\tilde{y}} \subset \mathbb{R}^n$ containing \tilde{y} , a set $W_{\tilde{y}} \subset \mathbb{R}^d$, and C^2 functions $\psi_{\tilde{y}} : W_{\tilde{y}} \rightarrow U_{\tilde{y}}$ and $\phi_{\tilde{y}} : U_{\tilde{y}} \rightarrow W_{\tilde{y}}$ such that $\mathcal{M} \cap U_{\tilde{y}} = \psi_{\tilde{y}}(W_{\tilde{y}})$ and $\phi_{\tilde{y}} \circ \psi_{\tilde{y}} = \text{id}$. For every $\tilde{y} \in \widehat{\mathcal{A}}_{x, \bar{\tau}}$, we can see that $\Pi_d \circ O_{U_{\bar{\tau}}^x} \circ G_{x, U_{\bar{\tau}}^x}^{-1} \circ \psi_{\tilde{y}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is C^2 because it is a composition of functions that are all at least C^2 . Its derivative has the same rank as $DG_{x, U_{\bar{\tau}}^x}^{-1}$, which at a point \tilde{y} is the projection of the derivative in the ambient space onto $T_{\tilde{y}}\mathcal{M}$, i.e., $V_{\tilde{y}}V_{\tilde{y}}^\top \widehat{V_{x, \bar{\tau}}V_{x, \bar{\tau}}^\top}$. Note that $V_{\tilde{y}} \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$, $\widehat{V_{x, \bar{\tau}}^\top} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^d)$, and $\text{rank}(V_{\tilde{y}}) = \text{rank}(\widehat{V_{x, \bar{\tau}}^\top}) = d$, so by the properties of rank we have $\text{rank}(V_{\tilde{y}}V_{\tilde{y}}^\top \widehat{V_{x, \bar{\tau}}V_{x, \bar{\tau}}^\top}) = \text{rank}(V_{\tilde{y}}^\top \widehat{V_{x, \bar{\tau}}})$. $V_{\tilde{y}}^\top \widehat{V_{x, \bar{\tau}}}$ is of full rank whenever $\sigma_d(V_{\tilde{y}}^\top \widehat{V_{x, \bar{\tau}}}) > 0$.

Recall that the cosines of the principal angles between $V_{\tilde{y}}$ and $\widehat{V_{x, \bar{\tau}}}$ are defined as the singular values of $V_{\tilde{y}}^\top \widehat{V_{x, \bar{\tau}}}$. Hence, $DG_{x, U_{\bar{\tau}}^x}^{-1}$ has full rank on $\widehat{U_{\bar{\tau}}^x}$ whenever the principal angles

between $\widehat{V}_{x,\bar{\tau}}$ and any tangent space from $\left\{T_{\tilde{y}}\mathcal{M} \mid \tilde{y} \in \widehat{\tilde{\mathcal{A}}}_{x,\bar{\tau}}\right\}$ are strictly less than $\pi/2$. Suppose this does not hold and there does exist $y' \in \widehat{\tilde{\mathcal{A}}}_{x,\bar{\tau}}$ such that $\sin \angle_1\left(T_{y'}\mathcal{M}, \widehat{T}_{x,\bar{\tau}}\mathcal{M}\right) = 1$. Applying Lemma 42 ii), we see that

$$\begin{aligned} \sin \angle_1\left(T_{y'}\mathcal{M}, \widehat{T}_{x,\bar{\tau}}\mathcal{M}\right) &\leq \sin \frac{d_{\mathcal{M}}(x, y')}{\tau} + \varepsilon_p \\ &\leq \sin \frac{8\left(1 + 64C_d\bar{\tau}^2/49\tau^2\right)^{1/2}\bar{\tau}}{7\tau} + \varepsilon_p \\ &\leq \frac{1}{250} + \frac{1}{12} < \frac{1}{11} \end{aligned}$$

with probability at least $1 - \delta_p$. Thus, the hypothesized y' cannot exist, and $\text{rank}\left(DG_{x,\widehat{U}_{\bar{\tau}}^x}^{-1}\right) = d$ on $\widehat{\tilde{\mathcal{A}}}_{x,\bar{\tau}}$ with the same probability. Although it appears that we are conditioning separately on the event that $\text{rank}\left(\widehat{V}_{x,\bar{\tau}}^\top\right) = d$, the probabilistic bound on $\left\|\sin \Theta\left(V_x, \widehat{V}_{x,\bar{\tau}}\right)\right\|_F$ from Theorem 22 assumes that the d^{th} eigenvalue of the projection of $N_{x,\bar{\tau}}^{-1}X_{x,\bar{\tau}}X_{x,\bar{\tau}}^\top$ onto $T_x\mathcal{M}$ is bounded away from zero; this is also a sufficient condition to ensure that $\text{rank}\left(\widehat{V}_{x,\bar{\tau}}^\top\right) = d$.

We have shown that the hypotheses of the inverse function theorem hold. (Without loss of generality we can choose $\psi_{\tilde{y}}$ such that $\psi_{\tilde{y}}(0) = \tilde{y}$ and $\phi_{\tilde{y}}(\tilde{y})$ is a fixed point). We see that $\left(\Pi_d \circ O_{\widehat{U}_{\bar{\tau}}^x} \circ G_{x,\widehat{U}_{\bar{\tau}}^x}^{-1} \circ \psi_{\tilde{y}}\right)^{-1}$ exists within a small neighborhood $\widetilde{W}_{\tilde{y}} \subset W_{\tilde{y}}$ of $\phi_{\tilde{y}}(\tilde{y})$. Let $\tilde{U}_{\tilde{y}} := \psi_{\tilde{y}}(\widetilde{W}_{\tilde{y}})$; then, $\{\tilde{U}_{\tilde{y}}\}$ is an open cover of $\left(\widehat{U}_{\bar{\tau}}^x \cup \partial\widehat{U}_{\bar{\tau}}^x\right) \cap \mathcal{M}$, a compact set. Let $\{\tilde{V}_{\tilde{y}}\}$ be the projections of the $\{\tilde{U}_{\tilde{y}}\}$ onto $\widehat{T}_{x,\bar{\tau}}\mathcal{M}$; these are clearly open sets as well. We can choose a finite set $\{\tilde{y}_i\}$ such that $\{\tilde{U}_{\tilde{y}_i}\}$ covers $\widehat{\tilde{\mathcal{A}}}_{x,\bar{\tau}}$ and $\{\tilde{V}_{\tilde{y}_i}\}$ covers $\widehat{\mathcal{A}}_{x,\bar{\tau}}$. Let $G_{x,\widehat{U}_{\bar{\tau}}^x}$ be the unique map with restrictions $G_{x,\widehat{U}_{\bar{\tau}}^x}|_{\tilde{V}_{\tilde{y}_i}} := \psi_{\tilde{y}_i} \circ \left(\Pi_d \circ O_{\widehat{U}_{\bar{\tau}}^x} \circ G_{x,\widehat{U}_{\bar{\tau}}^x}^{-1} \circ \psi_{\tilde{y}_i}\right)^{-1}$. Since $G_{x,\widehat{U}_{\bar{\tau}}^x}^{-1}$ is bijective on all of $\widehat{U}_{\bar{\tau}}^x$ and since inverses are unique, this is sufficient to show that $G_{x,\widehat{U}_{\bar{\tau}}^x}$ is well-defined. It is also of class C^2 ; the local inverses are C^2 by composition, and they are defined on open sets and agree on the intersections of their domains, which preserves differentiability. Note that this is equivalent to writing $\mathcal{M} \cap \widehat{U}_{\bar{\tau}}^x$ as the union of a finite number of graphs $\left(y, F_{x,\widehat{U}_{\bar{\tau}}^x}|_{\tilde{V}_{\tilde{y}_i}}(y)\right)$, where the C^2 function $F_{x,\widehat{U}_{\bar{\tau}}^x}|_{\tilde{V}_{\tilde{y}_i}} : \mathbb{R}^d \rightarrow \mathbb{R}^{n-d}$ assigns $y \mapsto \Pi_{n-d} \circ O_{\widehat{U}_{\bar{\tau}}^x} \circ \psi_{\tilde{y}_i} \circ \left(\Pi_d \circ O_{\widehat{U}_{\bar{\tau}}^x} \circ G_{x,\widehat{U}_{\bar{\tau}}^x}^{-1} \circ \psi_{\tilde{y}_i}\right)^{-1}(y)$. $F_{x,\widehat{U}_{\bar{\tau}}^x}$ is defined as the unique function with restrictions $F_{x,\widehat{U}_{\bar{\tau}}^x}|_{\tilde{V}_{\tilde{y}_i}}$. Let $\Psi := \left\{\psi_{\tilde{y}_i}|_{\widetilde{W}_{\tilde{y}_i}}\right\}$. In the remainder of the text, we write

$F_{x, \widehat{U}_{\bar{\tau}}}$ and $G_{x, \widehat{U}_{\bar{\tau}}}$ identically to their restrictions to $\{\widetilde{V}_{\tilde{y}_i}\}$, except we substitute ψ for $\psi_{\tilde{y}_i}$, understanding ψ to be any function from Ψ such that the image of $G_{x, \widehat{U}_{\bar{\tau}}}^{-1} \circ \psi$ includes the point of application.

It is clear that $F_{x, \widehat{U}_{\bar{\tau}}}$ and $G_{x, \widehat{U}_{\bar{\tau}}}$ have bounded derivatives. To prove **iii)** we use largely the same setup as Lemma **35 ii)**, which we summarize here, making the necessary changes for our setting. These statements are justified by the same reasoning as in the earlier lemma, so we do not restate the proofs here. Let $x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}$, and let $\tilde{x}' := G_{x, \widehat{U}_{\bar{\tau}}}(x')$. Recall that by Lemma **42 i)**, $\tilde{x}' \in \widehat{\tilde{\mathcal{A}}}_{x, \bar{\tau}} \subset \tilde{\mathcal{A}}_{x, \bar{\tau}} / (1 - \varepsilon_p \sqrt{2})$. $DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}$ and $DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}'}$ are maps between $T_{x'} \widehat{\mathcal{A}}_{x, \bar{\tau}}$ and $T_{\tilde{x}'} \widehat{\tilde{\mathcal{A}}}_{x, \bar{\tau}}$. Specifically, $DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}'} : T_{\tilde{x}'} \widehat{\tilde{\mathcal{A}}}_{x, \bar{\tau}} \rightarrow T_{x'} \widehat{\mathcal{A}}_{x, \bar{\tau}}$ assigns $\tilde{v} \mapsto \left(V_{\tilde{x}'} V_{\tilde{x}'}^\top \widehat{V}_{x, \bar{\tau}} \widehat{V}_{x, \bar{\tau}}^\top \right)^\top \tilde{v}$, and $DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}$ is the inverse map. Since $F_{x, \widehat{U}_{\bar{\tau}}}$ is the projection of $G_{x, \widehat{U}_{\bar{\tau}}}$ onto $\widehat{T}_{x, \bar{\tau}}^\perp \mathcal{M}$, $DF_{x, \widehat{U}_{\bar{\tau}}}|_{x'}$ is a map from $T_{x'} \widehat{\mathcal{A}}_{x, \bar{\tau}}$ to the tangent space of $\widehat{T}_{x, \bar{\tau}}^\perp \mathcal{M}$ at $F_{x, \widehat{U}_{\bar{\tau}}}(x')$. The singular values of $DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}'}$ are $\left\{ \cos \angle_i \left(T_{\tilde{x}'} \mathcal{M}, \widehat{T}_{x, \bar{\tau}} \mathcal{M} \right) \right\}$, those of $DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}$ are $\left\{ \sec \angle_i \left(T_{\tilde{x}'} \mathcal{M}, \widehat{T}_{x, \bar{\tau}} \mathcal{M} \right) \right\}$, and those of $DF_{x, \widehat{U}_{\bar{\tau}}}|_{x'}$ are $\left\{ \tan \angle_i \left(T_{\tilde{x}'} \mathcal{M}, \widehat{T}_{x, \bar{\tau}} \mathcal{M} \right) \right\}$. Because $DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}'}$ is a composition of projections, $\inf_{x \in \mathcal{M}} \inf_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \sigma_d \left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'} \right) \geq 1$, and $\inf_{x \in \mathcal{M}} \inf_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \sigma_d \left(DF_{x, \widehat{U}_{\bar{\tau}}}|_{x'} \right) \geq 0$.

The operator norm of $DF_{x, \widehat{U}_{\bar{\tau}}}|_{x'}$ is bounded above as

$$\begin{aligned} \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \sigma_1 \left(DF_{x, \widehat{U}_{\bar{\tau}}}|_{x'} \right) &= \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \left\| DF_{x, \widehat{U}_{\bar{\tau}}}|_{x'} \right\|_2 \\ &= \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \tan \angle_1 \left(T_{\tilde{x}'} \mathcal{M}, \widehat{T}_{x, \bar{\tau}} \mathcal{M} \right) \\ &\leq \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \tan \arcsin \left(\sin \frac{d_{\mathcal{M}}(x, \tilde{x}')}{\tau} + \varepsilon_p \right) \\ &\leq \tan \arcsin \left(\left(1 + \frac{16\bar{\tau}^2}{9\tau^2} \right)^{1/2} \frac{8\bar{\tau}}{7\tau} + \varepsilon_p \right) \\ &< \frac{8}{5} \left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p \right) \end{aligned}$$

for $\bar{\tau}/\tau < 1/2$ and $\varepsilon_p < 1/12$. This also implies

$$\begin{aligned} \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \sigma_1 \left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'} \right) &= \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \left\| DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'} \right\|_2 \\ &= \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \sec \angle_1 \left(T_{\tilde{x}'} \mathcal{M}, \widehat{T}_{x, \bar{\tau}} \mathcal{M} \right) \end{aligned}$$

$$\leq \left(1 + \frac{64}{25} \left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p\right)^2\right)^{1/2}.$$

We now bound the Lipschitz constants of $DF_{x, \widehat{U}_{\bar{\tau}}}$ and $DG_{x, \widehat{U}_{\bar{\tau}}}$, which are equal because $DG_{x, \widehat{U}_{\bar{\tau}}} = \left[I \left| DF_{x, \widehat{U}_{\bar{\tau}}}^\top \right|^\top\right]^\top$. Let $x_1, x_2 \in \widehat{\mathcal{A}}_{x, \bar{\tau}}$, and let $\tilde{x}_1 := G_{x, \widehat{U}_{\bar{\tau}}}(x_1)$ and $\tilde{x}_2 := G_{x, \widehat{U}_{\bar{\tau}}}(x_2)$. Let $v \in \widehat{T_{x, \bar{\tau}}\mathcal{M}}$. Then, there exist vectors $v_{\tilde{x}_1}, v_{\tilde{x}_2}$ with norms no greater than $\|v\| \sup_{x \in \mathcal{M}} \sup_{x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}\right)$ satisfying $DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}_1} v_{\tilde{x}_1} = DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}_2} v_{\tilde{x}_2} = v$. We have

$$\begin{aligned} \sup_{x \in \mathcal{M}} \text{Lip}\left(DG_{x, \widehat{U}_{\bar{\tau}}}\right) &\leq \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sup_{\substack{v \in \widehat{T_{x, \bar{\tau}}\mathcal{M}} \\ \|v\|=1}} \frac{\sqrt{d} \left\| DG_{x, \widehat{U}_{\bar{\tau}}}|_{x_1} v - DG_{x, \widehat{U}_{\bar{\tau}}}|_{x_2} v \right\|}{\|x_1 - x_2\|} \\ &= \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sup_{\substack{v \in \widehat{T_{x, \bar{\tau}}\mathcal{M}} \\ \|v\|=1}} \frac{\sqrt{d} \|v_{\tilde{x}_1} - v_{\tilde{x}_2}\|}{\|x_1 - x_2\|} \\ &\leq \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sup_{\substack{v \in \widehat{T_{x, \bar{\tau}}\mathcal{M}} \\ \|v\|=1}} \frac{\sqrt{d} (\|v_{\tilde{x}_1} - \tilde{v}_{\tilde{x}_2}\| + \|\tilde{v}_{\tilde{x}_2} - v_{\tilde{x}_2}\|)}{\|x_1 - x_2\|}, \end{aligned}$$

where $\tilde{v}_{\tilde{x}_2}$ is the projection of $v_{\tilde{x}_2}$ onto $T_{\tilde{x}_1}\mathcal{M}$. Using the principal angle bound from Lemma 34

v), $\|\tilde{v}_{\tilde{x}_2} - v_{\tilde{x}_2}\| \leq \|V_{\tilde{x}_1} V_{\tilde{x}_1}^\top - V_{\tilde{x}_2} V_{\tilde{x}_2}^\top\|_2 \|\tilde{v}_{\tilde{x}_2}\| \leq \|v_{\tilde{x}_2}\| \sin \angle_1(T_{\tilde{x}_1}\mathcal{M}, T_{\tilde{x}_2}\mathcal{M}) \leq \|v_{\tilde{x}_2}\| \sin(d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)/\tau)$.

We also have

$$\begin{aligned} \|\tilde{v}_{\tilde{x}_2} - v_{\tilde{x}_1}\| &\leq \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x_1}\right) \left\| DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}_1} \tilde{v}_{\tilde{x}_2} - DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}_1} v_{\tilde{x}_1} \right\| \\ &= \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x_1}\right) \left\| DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}_1} \tilde{v}_{\tilde{x}_2} - DG_{x, \widehat{U}_{\bar{\tau}}}^{-1}|_{\tilde{x}_2} v_{\tilde{x}_2} \right\| \\ &\leq \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x_1}\right) \left\| \widehat{V_{x, \bar{\tau}}} \widehat{V_{x, \bar{\tau}}}^\top \right\|_2 \|V_{\tilde{x}_1} V_{\tilde{x}_1}^\top - V_{\tilde{x}_2} V_{\tilde{x}_2}^\top\|_2 \|v_{\tilde{x}_2}\| \\ &\leq \|v_{\tilde{x}_2}\| \sin \frac{d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)}{\tau} \times \sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}\right). \end{aligned}$$

Thus,

$$\begin{aligned} \sup_{x \in \mathcal{M}} \text{Lip}\left(DG_{x, \widehat{U}_{\bar{\tau}}}\right) &\leq \sup_{\substack{x \in \mathcal{M} \\ x_1 \neq x_2 \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sup_{\substack{v \in \widehat{T_{x, \bar{\tau}}\mathcal{M}} \\ \|v\|=1}} \frac{\sqrt{d} d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)}{\tau \|x_1 - x_2\|} \|v_{\tilde{x}_2}\| \left(1 + \sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}\right)\right) \\ &\leq \frac{\sqrt{d}}{\tau} \times \sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}\right) \times \sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}}} \sigma_1\left(DG_{x, \widehat{U}_{\bar{\tau}}}|_{x'}\right) \end{aligned}$$

$$\begin{aligned} & \times \left(1 + \sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} \sigma_1 \left(DG_{x, \widehat{U}_{\bar{\tau}}} \Big|_{x'} \right) \right) \\ & \leq \frac{7\sqrt{d}}{\tau}. \end{aligned} \quad \blacksquare$$

Clearly, analogous statements to Lemma 35 iii) and iv) and Corollary 36 hold for the maps $\left\{ G_{x, \widehat{U}_{\bar{\tau}}}^{-1} \mid x \in \mathcal{M}, \bar{\tau}/(1 - \varepsilon_p \sqrt{2}) < \tau_0 \right\}$. In fact, a version of Corollary 36 i) was needed for the bound on $d_{\mathcal{M}}(\tilde{x}_1, \tilde{x}_2)/\|x_1 - x_2\|$ in the final part of the proof of Theorem 44 iii); this ratio depends on the operator norm of $DG_{x, \widehat{U}_{\bar{\tau}}}$ and not on its Lipschitz constant, so there are no circular dependencies. We now prove a related lemma concerning the densities of the pushforwards of \mathcal{P} to the estimated tangent spaces.

Lemma 45. *Let $\bar{\tau} < \tilde{\tau}_{max,2}$. Let $\mathcal{P}_{x, \bar{\tau}}$, for $x \in \mathcal{M}$, be the pushforward of \mathcal{P} to $\widehat{\mathcal{A}}_{x, \bar{\tau}}$ considered as a probability measure. Then, it has density $p_{x, \bar{\tau}}$ with respect to \mathcal{L}^d that is bounded as*

$$\begin{aligned} 0 &< \left(1 + \frac{64}{25} \left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p \right)^2 \right)^{-d/2} \frac{p_{min}}{p_{max}} \omega_d^{-1} \bar{\tau}^{-d} \leq \inf_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} p_{x, \bar{\tau}}(x') \\ \infty &> \left(1 + \frac{64}{25} \left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p \right)^2 \right)^{d/2} \frac{p_{max}}{p_{min}} \omega_d^{-1} \bar{\tau}^{-d} \geq \sup_{\substack{x \in \mathcal{M} \\ x' \in \widehat{\mathcal{A}}_{x, \bar{\tau}}} p_{x, \bar{\tau}}(x') \end{aligned}$$

Proof. Note that, by Theorem 33, $\mathcal{P}_{z_i, \bar{\tau}}$ has density

$$p_{z_i, \bar{\tau}}(x') = \left(p \circ G_{z_i, \widehat{U}_{\bar{\tau}}^{z_i}} \right)(x') \sqrt{\det \left(DG_{z_i, \widehat{U}_{\bar{\tau}}^{z_i}}^\top \Big|_{x'} DG_{z_i, \widehat{U}_{\bar{\tau}}^{z_i}} \Big|_{x'} \right)} \mathcal{P} \left(\widehat{\mathcal{A}}_{z_i, \bar{\tau}} \right)^{-1}$$

(after renormalizing) with respect to \mathcal{L}^d . The determinant is the product of the squared singular values of $DG_{z_i, \widehat{U}_{\bar{\tau}}^{z_i}} \Big|_{x'}$, which have a uniform upper bound of $\left(1 + 64/25(7\bar{\tau}/5\tau + \varepsilon_p)^2 \right)^{1/2}$ by Theorem 44 iii). By another application of these theorems, $0 < p_{min} \omega_d \bar{\tau}^d \leq \inf_{x \in \mathcal{M}} \mathcal{P} \left(\widehat{\mathcal{A}}_{x, \bar{\tau}} \right)$ and $\sup_{x \in \mathcal{M}} \mathcal{P} \left(\widehat{\mathcal{A}}_{x, \bar{\tau}} \right) \leq p_{max} \omega_d \bar{\tau}^d \left(1 + 64/25(7\bar{\tau}/5\tau + \varepsilon_p)^2 \right)^{d/2} < \infty$. This is enough to prove the statement. \blacksquare

3.5.1 Pullback to Estimated Tangent Spaces

In order to perform local $C^{1,1}(\mathbb{R}^d)$ regression on the estimated tangent spaces, we need a result showing that we are estimating a function with finite $C^{1,1}(\mathbb{R}^d)$ norm. In the following theorem, we show that for $f : \widehat{\mathcal{A}}_{x,\bar{\tau}} \rightarrow \mathbb{R}$ with finite $C^{1,1}(\mathcal{M})$ norm, the $C^{1,1}(\mathbb{R}^d)$ norm of the pullback $G_{x,\widehat{U}_{\bar{\tau}}}^*(f) := f \circ G_{x,\widehat{U}_{\bar{\tau}}}$ is bounded above by a multiple of $\|f\|_{C^{1,1}(\mathcal{M})}$. The proof is similar to that of Lemma 40, with the appropriate operator norms substituted for $\sup_{x \in \mathcal{M}, y \in \mathcal{A}_{x,\bar{\tau}}} \|DG_{x,U_{\bar{\tau}}}|_y\|_2$.

Theorem 46. *Let $x \in \mathcal{M}$, and let $\widehat{T_{x,\bar{\tau}}\mathcal{M}}$ be an estimate of $T_x\mathcal{M}$ such that $\sin \angle_1(\widehat{T_{x,\bar{\tau}}\mathcal{M}}, T_x\mathcal{M}) \leq \varepsilon_p$. Assume $\bar{\tau}/(1 - \varepsilon_p\sqrt{2}) < \tau_0$. Let $f^* : \mathcal{M} \rightarrow \mathbb{R}$ be a $C^{1,1}(\mathcal{M})$ function such that $\|f^*\|_{C^{1,1}(\mathcal{M})} \leq M^*$, and let $h_{x,\bar{\tau}}^* : \widehat{T_{x,\bar{\tau}}\mathcal{M}} \supset \widehat{\mathcal{A}}_{x,\bar{\tau}} \rightarrow \mathbb{R}$ be the function assigning $y \mapsto (f^* \circ G_{x,\widehat{U}_{\bar{\tau}}})(y)$. Then, the $C^{1,1}(\mathbb{R}^d)$ norm of $h_{x,\bar{\tau}}^*$ is bounded above by $M_{h,\bar{\tau}}^*$, a constant independent of x :*

$$\begin{aligned} \sup_{x \in \mathcal{M}} \|h_{x,\bar{\tau}}^*\|_{C^{1,1}(\mathbb{R}^d)} &\leq \left(1 + \frac{7\sqrt{d}}{\tau} + \frac{64}{25} \left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p \right)^2 \right) M^* \\ &=: M_{h,\bar{\tau}}^*. \end{aligned}$$

Proof. Clearly, $\|h_{x,\bar{\tau}}^*\|_{C^0(\mathbb{R}^d)} \leq M^*$, and

$$\begin{aligned} \sup_{\substack{x \in \mathcal{M} \\ y \in \widehat{\mathcal{A}}_{x,\bar{\tau}}}} \|\nabla h_{x,\bar{\tau}}^*(y)\| &\leq \sup_{\substack{x \in \mathcal{M} \\ y \in \widehat{\mathcal{A}}_{x,\bar{\tau}}}} \left\| \left(DG_{x,\widehat{U}_{\bar{\tau}}} \right)^\top \Big|_y \right\|_2 \times \sup_{x \in \mathcal{M}} \|\nabla f^*(x)\| \\ &\leq \left(1 + C_d \left(\frac{\bar{\tau}}{\tau} + \varepsilon_p \right)^2 \right)^{1/2} M^*, \end{aligned}$$

where the second line follows by Theorem 44 iii) and the fact that $\|f^*\|_{C^{1,1}(\mathcal{M})} \leq M^*$. As we did for $\text{Lip}(\nabla g_{x,\bar{\tau}}^*)$, we can use the triangle inequality to bound $\text{Lip}(\nabla h_{x,\bar{\tau}}^*)$ by the sum of a term depending on the norm of $DG_{x,\widehat{U}_{\bar{\tau}}}$ and the Lipschitz constant of ∇f^* and another term depending on the Lipschitz constant of $DG_{x,\widehat{U}_{\bar{\tau}}}$ and the norm of ∇f^* . This yields, for $\tilde{y}_1 := G_{x,\widehat{U}_{\bar{\tau}}}(y_1)$ and $\tilde{y}_2 := G_{x,\widehat{U}_{\bar{\tau}}}(y_2)$,

$$\sup_{x \in \mathcal{M}} \text{Lip}(\nabla h_{x,\bar{\tau}}^*) \leq \sup_{\substack{x \in \mathcal{M} \\ y \in \widehat{\mathcal{A}}_{x,\bar{\tau}}}} \left\| \left(DG_{x,\widehat{U}_{\bar{\tau}}} \right)^\top \Big|_y \right\|_2 \times \text{Lip}(\widetilde{\text{grad} f^*}) \times \sup_{\substack{x \in \mathcal{M} \\ y_1 \neq y_2 \in \widehat{\mathcal{A}}_{x,\bar{\tau}}}} \frac{d_{\mathcal{M}}(\tilde{y}_1, \tilde{y}_2)}{\|y_1 - y_2\|}$$

$$\begin{aligned}
& + \sup_{x \in \mathcal{M}} \text{Lip} \left(DG_{x, \widehat{U}_{\bar{\tau}}} \right) \times \sup_{x \in \mathcal{M}} \|\nabla f^*(x)\| \\
& \leq \left(1 + C_d \left(\frac{\bar{\tau}}{\tau} + \varepsilon_p \right)^2 \right) M^* + \frac{C_d M^*}{\tau}. \quad \blacksquare
\end{aligned}$$

3.6 Analysis of $C^{1,1}(\mathcal{M})$ Regression Algorithm

We begin by restating the regression problem. We are provided with N i.i.d. samples from \mathcal{M} , a d -dimensional C^2 submanifold of \mathbb{R}^n with reach τ ; we are also provided with samples from a real-valued function associated with each point. Let $\mathcal{X} \times \mathcal{Y} \subset \mathcal{M} \times \mathbb{R}$ denote the sample. We place minor restrictions on the probability measure \mathcal{P} that \mathcal{X} is drawn from — we assume that \mathcal{P} is absolutely continuous with respect to $\mathcal{H}^d(\mathcal{M})$ and has a density p such that $0 < p_{\min} \leq p(x) \leq p_{\max} < \infty$ for all $x \in \mathcal{M}$. \mathcal{Y} is drawn i.i.d from a Gaussian with mean $f^* \in C^{1,1}(\mathcal{M})$ and variance $\sigma^2 < \infty$. f^* is assumed to have bounded norm $\|f^*\|_{C^{1,1}(\mathcal{M})} \leq M^*$. The foregoing assumptions apply throughout this section, so we will not repeat them in the statement of each result.

We will now describe the estimation procedure that we use and then prove that it recovers f^* in risk and sup norm as $N \rightarrow \infty$. Start by fixing $\bar{\tau} < \tilde{\tau}_{\max,2}$, which serves as a bandwidth parameter. Then, proceed with the following steps to define the estimator \hat{f} :

1. Construct a minimal $\bar{\tau}/6$ -net \mathcal{Z} of \mathcal{X} . (A minimal η -net is both an η -covering and an η -packing.) We will locally estimate f^* around each point of \mathcal{Z} and then patch together these estimators using a partition of unity.
2. For each $z_i \in \mathcal{Z}$, estimate $T_{z_i} \mathcal{M}$ by performing local PCA within the region $B_n(z_i, \bar{\tau}) \cap \mathcal{X}$. Let $\widehat{T}_{z_i, \bar{\tau}} \mathcal{M}$ be the estimator, and let $\widehat{U}_{\bar{\tau}}^{z_i} := B_d(z_i, \bar{\tau}) \times B_{n-d}(z_i, \bar{\tau})$, where the coordinate system has origin z_i and $\widehat{T}_{z_i, \bar{\tau}} \mathcal{M}$ is identified with \mathbb{R}^d . Define $\widehat{\mathcal{A}}_{z_i, \bar{\tau}} := \widehat{U}_{\bar{\tau}}^{z_i} \cap \widehat{T}_{z_i, \bar{\tau}} \mathcal{M}$, and $\widetilde{\mathcal{A}}_{z_i, \bar{\tau}} := \widehat{U}_{\bar{\tau}}^{z_i} \cap \mathcal{M}$. Let $G_{z_i, \widehat{U}_{\bar{\tau}}^{z_i}}^{-1} : \widehat{\mathcal{A}}_{z_i, \bar{\tau}} \rightarrow \widetilde{\mathcal{A}}_{z_i, \bar{\tau}}$ be the orthogonal projection of \mathcal{M} onto $\widehat{T}_{z_i, \bar{\tau}} \mathcal{M}$, and let $G_{z_i, \widehat{U}_{\bar{\tau}}^{z_i}}$ be its inverse, which is well-defined by Theorem 44.

3. Estimate the pullback of f^* to each $\widehat{T_{z_i, \bar{\tau}} \mathcal{M}}$. By Theorem 46, the $\{h_{z_i, \bar{\tau}}^* := f^* \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}} \mid z_i \in \mathcal{Z}\}$ are $C^{1,1}(\mathbb{R}^d)$ and have uniformly bounded $C^{1,1}(\mathbb{R}^d)$ norms. For each $h_{z_i, \bar{\tau}}^*$, use the methods of Gustafson et al. (2018) outlined in Section 3.2 to perform $C^{1,1}(\mathbb{R}^d)$ regression on $\mathcal{X}'_{z_i} \times \mathcal{Y}_{z_i}$, where $\mathcal{X}'_{z_i} := G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}(\widehat{\tilde{\mathcal{A}}_{z_i, \bar{\tau}}} \cap \mathcal{X}) \subset \widehat{T_{z_i, \bar{\tau}} \mathcal{M}}$ and \mathcal{Y}_{z_i} is the corresponding subset of \mathcal{Y} . This involves finding a solution to

$$\arg \inf_{f \in \widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}} \frac{1}{|\mathcal{X}'_{z_i}|} \sum_{(x_i, y_i) \in \mathcal{X}'_{z_i} \times \mathcal{Y}_{z_i}} (y_i - f(x_i))^2,$$

where $\widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}} := \{f : \widehat{\mathcal{A}}_{z_i, \bar{\tau}} \rightarrow \mathbb{R} \mid \|f\|_{C^{1,1}(\mathbb{R}^d)} \leq \widetilde{M}\}$. \widetilde{M} is increasing with N , and is chosen via metric entropy arguments, as in Theorem 31. Denote the estimator by $\widehat{h}_{z_i, \bar{\tau}}$.

4. Define the function $\widehat{f} : \mathcal{M} \rightarrow \mathbb{R}$ through the assignment

$$x \mapsto \sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(\widehat{h}_{z_i, \bar{\tau}} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} \right)(x),$$

where $\{\alpha_{z_i} \mid z_i \in \mathcal{Z}\}$ is any suitable C^2 partition of unity subordinate to $\{\widehat{\tilde{\mathcal{A}}_{z_i, \bar{\tau}}} \mid z_i \in \mathcal{Z}\}$.

Before proving risk bounds and convergence properties of \widehat{f} , we need to show that \mathcal{Z} can be used to partition \mathcal{M} in a favorable manner, which we do in Section 3.6.1. We start by deriving bounds on the size of minimal η -nets of \mathcal{M} (where $\eta < \tau/2$) in terms of η and the volume of \mathcal{M} . Then, we derive uniform convergence rates for the empirical measure indexed by sets formed by intersecting \mathcal{M} with an n -dimensional ball centered on \mathcal{M} . We use these results to derive an upper bound on the size of a $\bar{\tau}/6$ -net \mathcal{Z} of \mathcal{X} and to prove that \mathcal{Z} is a $\bar{\tau}/2$ -net of \mathcal{M} w.h.p. The latter fact implies that if f^* can be locally estimated around each point of \mathcal{Z} , a partition of unity can be used to construct an estimator \widehat{f} whose domain is all of \mathcal{M} .

We use the convergence of the empirical measure to derive a uniform lower bound, increasing in N , on the number of points of \mathcal{X} within $\bar{\tau}$ of each point of \mathcal{Z} . This leads to uniform bounds for other quantities, such as the error probability of local PCA. Since we

have a bound on $|\mathcal{Z}|$, we can apply the union bound to get the probability of the estimated tangent spaces being close in angular distance to the true tangent spaces. From this, it follows that charts based on orthogonal projection onto the estimated tangent spaces are C^2 diffeomorphisms with uniformly bounded norms. One of the implications of this is that rates can be set uniformly when locally performing $C^{1,1}(\mathbb{R}^d)$ regression. We also prove that the preimages of the projections within a radius of $\bar{\tau}$ cover \mathcal{M} when taken together, allowing us to construct a C^2 atlas for \mathcal{M} .

3.6.1 Sampling a C^2 Atlas for \mathcal{M}

We will assume that minimal η -nets for the metric space (K, dist) (with distance function $\text{dist} : K \times K \rightarrow \mathbb{R}$) are constructed using the following procedure. Choose $k_1 \in K$. Choose subsequent $k_i \in K \setminus \bigcup_j \overline{B_K(k_j, \eta)}$, where $B_K(k_j, \eta)$ is the set of all points $k \in K$ such that $\text{dist}(k_j, k) < \eta$. Clearly this procedure terminates for compact K otherwise $\{k_i\}$ does not contain a convergent subsequence. An η -covering is guaranteed by the stopping condition. Note that $\{k_i\}$ is also an η -packing by construction.

Lemma 47. *$(\mathcal{M}, \|\cdot\|_2)$ is a metric space where $\|\cdot\|_2$ is the standard norm of \mathbb{R}^n . For $\eta \in (0, \tau/2)$, a minimal η -net S_η of \mathcal{M} has cardinality bounded as*

$$\left(\frac{5}{7}\right)^{d/2} \omega_d^{-1} \frac{V}{\eta^d} \leq |S_\eta| \leq 5^{d/2} \omega_d^{-1} \frac{V}{\eta^d}.$$

Proof. Let $S_\eta := \{s_{\eta,i}\}$ be a minimal η -net of \mathcal{M} ; at least one such net exists and can be constructed according to the procedure outlined earlier in this section. We prove the bounds on its size by noting that

$$\sum_i \mathcal{H}^d(B_n(s_{\eta,i}, \eta/2) \cap \mathcal{M}) \leq \mathcal{H}^d(\mathcal{M}) \leq \sum_i \mathcal{H}^d(B_n(s_{\eta,i}, \eta) \cap \mathcal{M}).$$

The first inequality holds because S_η is an η -packing and $\{B_n(s_{\eta,i}, \eta/2) \cap \mathcal{M}\}$ are disjoint, and the second inequality is true since \mathcal{M} is covered by $\{B_n(s_{\eta,i}, \eta) \cap \mathcal{M}\}$. From this we see that

$$|S_\eta| \inf_{x \in \mathcal{M}} \mathcal{H}^d(B_n(x, \eta/2) \cap \mathcal{M}) \leq V \leq |S_\eta| \sup_{x \in \mathcal{M}} \mathcal{H}^d(B_n(x, \eta) \cap \mathcal{M}).$$

Applying Lemma 35 iv) with g as the constant function 1 shows that

$$\begin{aligned} \sup_{x \in \mathcal{M}} \mathcal{H}^d(B_n(x, \eta) \cap \mathcal{M}) &\leq \left(1 + \frac{49\eta^2}{36\tau^2}\right)^{d/2} \sup_{x \in \mathcal{M}} \mathcal{L}^d(\Pi_x(B_n(x, \eta) \cap \mathcal{M})) \\ &\leq \left(\frac{7}{5}\right)^{d/2} \omega_d \eta^d, \end{aligned}$$

where the second line follows from the fact that $\Pi_x(B_n(x, \eta) \cap \mathcal{M}) \subset \mathcal{A}_{x, \eta}$. Now, note that Federer's reach condition implies that $\Pi_x(B_n(x, \eta/2) \cap \mathcal{M})$ contains a d -dimensional ball of radius at least $\eta/2\sqrt{1 - \eta^2/(16\tau^2)}$. Using the other side of Lemma 35 iv),

$$\begin{aligned} \inf_{x \in \mathcal{M}} \mathcal{H}^d(B_n(x, \eta/2) \cap \mathcal{M}) &\geq \inf_{x \in \mathcal{M}} \mathcal{L}^d(\Pi_x(B_n(x, \eta/2) \cap \mathcal{M})) \\ &\geq \omega_d \left(\frac{\eta}{2}\right)^d \left(1 - \frac{\eta^2}{16\tau^2}\right)^{d/2} \\ &\geq \left(\frac{1}{5}\right)^{d/2} \omega_d \eta^d. \end{aligned}$$

Rearranging yields the lemma. ■

Lemma 48. *Let $\mathcal{B}_0 := \{B_n(x, r) \mid x \in \mathcal{M}, 0 < r < \tau/2\}$. Then, for sets of the form $B \cap \mathcal{M}$, $B \in \mathcal{B}_0$, the following uniform bound holds for the convergence of the empirical measure to \mathcal{P} :*

$$\mathbb{P} \left[\sup_{\substack{B_n(x, r) \in \mathcal{B}_0 \\ N^{-1/(4d)} \leq r \leq \tau/2}} \left| \frac{1}{N} |B \cap \mathcal{X}| - \mathcal{P}(B \cap \mathcal{M}) \right| \leq \frac{1}{N^{1/4}} + \left(\frac{7}{5}\right)^{d/2} p_{\max} \omega_d \left(\frac{\tau}{2} + 1\right)^d \left(\frac{5}{N}\right) \right] \geq 1 - \frac{2(5^{d/2})V\tau N^{d+1}}{\omega_d} \exp\left(-N^{1/4} p_{\min} \omega_d \left(\frac{15}{16}\right)^{d/2} / 3\right).$$

With the same probability, $\sup_{B_n(x, r) \in \mathcal{B}_0, 0 < r < N^{-1/(4d)}} \left| \frac{1}{N} |B \cap \mathcal{X}| - \mathcal{P}(B \cap \mathcal{M}) \right| \leq N^{-1/4} + p_{\max} (7/5)^d \omega_d N^{-1/4}$.

Proof. Let $\mathcal{F}_{\mathcal{B}_0} := \{\mathbb{1}\{B \cap \mathcal{M}\} \mid B \in \mathcal{B}_0\}$. Then, the supremum in the probability statement becomes $\sup_{f \in \mathcal{F}_{\mathcal{B}_0}} \left| \frac{1}{N} \sum_{i=1}^N f(x_i) - \mathbb{E}f \right|$. Measurability is not an issue because $\mathcal{F}_{\mathcal{B}_0}$ contains a countable subset $\mathcal{F}_{\mathcal{B}_1}$ from which a pointwise convergent sequence of functions can be constructed for any $f \in \mathcal{F}_{\mathcal{B}_0}$; that is, $\mathcal{F}_{\mathcal{B}_0}$ is pointwise measurable (Van Der Vaart and

Wellner, 1996). $\mathcal{F}_{\mathcal{B}_1}$ can be taken as the indicators of the intersection of \mathcal{M} with balls of rational radii and whose centers have rational coordinates in charts centered at points of a finite net of \mathcal{M} .

We use a slightly different (finite) subset of $\mathcal{F}_{\mathcal{B}_0}$ to prove the finite sample bound stated in the lemma. Choose $\delta_1, \delta_2, \delta_3 > 0$ and sufficiently small with respect to τ . Construct a minimal δ_1 -net S_{δ_1} of \mathcal{M} . Let $\mathcal{B}_{2,N} := \{B_n(x, r) \mid x \in S_{\delta_1}, r \in (\delta_3, \tau) \cap \delta_2\mathbb{N}\}$. Let $\mathcal{F}_{\mathcal{B}_{2,N}}$ be the class of indicator functions of the intersection of \mathcal{M} with members of $\mathcal{B}_{2,N}$. For any $f \in \mathcal{F}_{\mathcal{B}_0}$, there exist $f_l, f_u \in \mathcal{F}_{\mathcal{B}_{2,N}}$ such that $f_l \leq f \leq f_u$ for all $x \in \mathcal{M}$.

Fix $f_0 := \mathbb{1}\{\mathcal{M} \cap B_n(x_0, r_0)\}$. Then, we can take $f_{0,l} := \mathbb{1}\{\mathcal{M} \cap B_n(x_{0,l}, r_{0,l})\}$ and $f_{0,u} := \mathbb{1}\{\mathcal{M} \cap B_n(x_{0,u}, r_{0,u})\}$, where $x_{0,l}, x_{0,u} \in S_{\delta_1}$ are within δ_1 of x_0 and $r_{0,l}, r_{0,u} \in (\delta_3, \tau) \cap \delta_2\mathbb{N}$ are within $\delta_1 + \delta_2$ of r_0 . An upper bound for $\mathbb{E}_N f_0 - \mathbb{E} f_0$ can be obtained by adding and subtracting $f_{0,u}$ within the expectations:

$$\begin{aligned} \mathbb{E}_N f_0 - \mathbb{E} f_0 &= \mathbb{E}_N [f_0 - f_{0,u}] - \mathbb{E} [f_0 - f_{0,u}] + \mathbb{E}_N f_{0,u} - \mathbb{E} f_{0,u} \\ &\leq \mathbb{E} [f_{0,u} - f_{0,l}] + |\mathbb{E}_N f_{0,u} - \mathbb{E} f_{0,u}|. \end{aligned}$$

Similarly, $\mathbb{E}_N f_0 - \mathbb{E} f_0 \geq -\mathbb{E} [f_{0,u} - f_{0,l}] - |\mathbb{E}_N f_{0,l} - \mathbb{E} f_{0,l}|$. Thus,

$$\begin{aligned} \sup_{f \in \mathcal{F}_{\mathcal{B}_0}} |\mathbb{E}_N f - \mathbb{E} f| &\leq \sup_{\substack{x \in \mathcal{M} \\ 0 < r < \tau/2}} \mathbb{E} [\mathbb{1}\{\mathcal{M} \cap B_n(x, r_0 + 2\delta_1 + 2\delta_2)\} - \mathbb{1}\{\mathcal{M} \cap B_n(x, r_0)\}] \\ &\quad + \sup_{f \in \mathcal{F}_{\mathcal{B}_{N,2}}} |\mathbb{E}_N f - \mathbb{E} f|. \end{aligned}$$

We start by bounding the first term on the right-hand side, rewriting it as a probability statement. For balls with radius less than δ_3 , we have

$$\begin{aligned} \sup_{\substack{x \in \mathcal{M} \\ 0 < r < \delta_3}} \mathcal{P}(\mathcal{M} \cap B_n(x, r_0 + \delta_1 + \delta_2)) &= \sup_{\substack{x \in \mathcal{M} \\ 0 < r < \delta_3}} \int_{\mathcal{M} \cap B_n(x, r_0 + \delta_1 + \delta_2)} p(x') d\mathcal{H}^d(x') \\ &\leq p_{max} \left(\frac{7}{5} \right)^{d/2} \sup_{\substack{x \in \mathcal{M} \\ 0 < r < \delta_3}} \mathcal{L}^d(\Pi_x(\mathcal{M} \cap B_n(x, r_0 + \delta_1 + \delta_2))) \\ &\leq p_{max} \left(\frac{7}{5} \right)^{d/2} \omega_d (\delta_3 + \delta_1 + \delta_2)^d. \end{aligned}$$

For balls with radius at least δ_3 the calculation is harder because we need a bound on the volume of the intersection of \mathcal{M} with an n -dimensional spherical shell. Letting $\tilde{\delta} := 2\delta_1 + 2\delta_2$,

$$\begin{aligned} \sup_{\substack{x \in \mathcal{M} \\ \delta_3 \leq r \leq \tau/2}} \mathcal{P}\left(\mathcal{M} \cap B_n(x, r_0 + \tilde{\delta}) \setminus B_n(x, r_0)\right) &= \sup_{\substack{x \in \mathcal{M} \\ \delta_3 \leq r \leq \tau/2}} \int_{\mathcal{M} \cap B_n(x, r_0 + \tilde{\delta}) \setminus B_n(x, r_0)} p(x') d\mathcal{H}^d(x') \\ &\leq p_{max} \left(\frac{7}{5}\right)^{d/2} \sup_{\substack{x \in \mathcal{M} \\ \delta_3 \leq r \leq \tau/2}} \left(\mathcal{L}^d\left(\Pi_x\left(\mathcal{M} \cap B_n(x, r_0 + \tilde{\delta})\right)\right) \right. \\ &\quad \left. - \mathcal{L}^d\left(\Pi_x\left(B_n(x, r_0)\right)\right) \right) \\ &= p_{max} \left(\frac{7}{5}\right)^{d/2} \sup_{\substack{x \in \mathcal{M} \\ \delta_3 \leq r \leq \tau/2}} \int_{\tilde{r}_0}^{\tilde{r}_1} r'^{d-1} V(\Theta) dr' d\Theta, \end{aligned}$$

where the last line is a spherical integral with Θ a vector of $d - 1$ angular variables and $V(\Theta)$ the volume element. The radii \tilde{r}_0 and \tilde{r}_1 may be functions of Θ . We need to bound the variation in the radial distances of the projections of $\mathcal{M} \cap B_n(x, r_0 + \tilde{\delta})$ and $\mathcal{M} \cap B_n(x, r_0)$, i.e., $\tilde{r}_1 - \tilde{r}_0$, and relate it to $\tilde{\delta}$. For a fixed set of angular values, consider a line starting at x . Denote its intersection with the boundary of $\Pi_x(\mathcal{M} \cap B_n(x, r_0))$ by x_1 and with the boundary of $\Pi_x(\mathcal{M} \cap B_n(x, r_0 + \tilde{\delta}))$ by x_2 . For a given x_1 , the distance $\|x_2 - x_1\|$ is maximized when x_2 is chosen such that $G_{x, U_{\tau/2}^x}(x_2) - G_{x, U_{\tau/2}^x}(x_1)$ lies parallel to $x_1 - x$ and $\|F_{x, U_{\tau/2}^x}(x_2)\| \leq \|F_{x, U_{\tau/2}^x}(x_1)\|$. In addition to the constraints $\|G_{x, U_{\tau/2}^x}(x_2) - x\| = r_0 + \tilde{\delta}$ and $\|G_{x, U_{\tau/2}^x}(x_1) - x\| = r_0$, we also require $\|G_{x, U_{\tau/2}^x}(x_2) - G_{x, U_{\tau/2}^x}(x_1)\| / \|x_2 - x_1\| \leq \sqrt{7/5}$ by Lemma 35 ii) and $\|F_{x, U_{\tau/2}^x}(x_1)\| \leq r_0^2 / (2\tau)$ by Federer's reach condition. Let $h := \|F_{x, U_{\tau/2}^x}(x_1)\|$, $t_2 := \|x_2 - x_1\|$, $t_1 := \|F_{x, U_{\tau/2}^x}(x_2) - F_{x, U_{\tau/2}^x}(x_1)\|$, and $c := t_1/t_2$. Then, we have $t_2 = \sqrt{(r_0 + \tilde{\delta})^2 - (h - ct_2)^2} - \sqrt{r_0^2 - h^2}$; solving this for t_2 and optimizing subject to the above constraints, we see that t_2 achieves its maximum when $h = r_0^2 / (2\tau)$ and $c = \sqrt{2/5}$. A series expansion shows $t_2 < 5/4\tilde{\delta}$. The desired integral can now be bounded as follows:

$$\begin{aligned} \sup_{\substack{x \in \mathcal{M} \\ \delta_3 \leq r \leq \tau/2}} \int_{\tilde{r}_0}^{\tilde{r}_0 + \frac{5}{4}\tilde{\delta}} r'^{d-1} V(\Theta) dr' d\Theta &= \sup_{\substack{x \in \mathcal{M} \\ \delta_3 \leq r \leq \tau/2}} \int \frac{1}{d} \left((\tilde{r}_0 + \frac{5}{4}\tilde{\delta})^d - \tilde{r}_0^d \right) V(\Theta) d\Theta \\ &\leq \omega_d \sum_{k=1}^d \binom{d}{k} \left(\frac{\tau}{2}\right)^{d-k} \left(\frac{5}{4}\tilde{\delta}\right)^k \end{aligned}$$

$$\leq \omega_d \left(\frac{\tau}{2} + 1 \right)^d \left(\frac{5}{4} \tilde{\delta} \right).$$

Finally, we show that $\sup_{f \in \mathcal{F}_{\mathcal{B}_{N,2}}} |\mathbb{E}_N f - \mathbb{E} f|$ is bounded with high probability. For a fixed $f \in \mathcal{F}_{\mathcal{B}_{N,2}}$ (i.e., a fixed $B \in \mathcal{B}_{N,2}$), a two-sided Chernoff bound gives

$$\mathbb{P}[|B \cap \mathcal{X}| - N\mathcal{P}(B \cap \mathcal{M})| \leq N^{3/4}\mathcal{P}(B \cap \mathcal{M})] \geq 1 - 2 \exp(-N^{1/2}\mathcal{P}(B \cap \mathcal{M})/3).$$

Note that $\inf_{B \in \mathcal{B}_{N,2}} \mathcal{P}(B \cap \mathcal{M})$ is bounded away from zero:

$$\begin{aligned} \inf_{B \in \mathcal{B}_{N,2}} \mathcal{P}(B \cap \mathcal{M}) &\geq p_{\min} \inf_{x \in \mathcal{M}} \mathcal{L}^d(\Pi_x(B_n(x, \delta_3) \cap \mathcal{M})) \\ &\geq p_{\min} \omega_d \delta_3^d \left(\frac{15}{16} \right)^{d/2}. \end{aligned}$$

This, along with the bound on $|S_{\delta_1}|$ from Lemma 47 allows us to apply the union bound:

$$\begin{aligned} \mathbb{P} \left[\sup_{B \in \mathcal{B}_{N,2}} \left| \frac{|B \cap \mathcal{X}|}{N} - \mathcal{P}(B \cap \mathcal{M}) \right| \leq \frac{1}{N^{1/4}} \right] &\geq 1 - \frac{2|S_{\delta_1}| \tau}{\delta_2} \exp \left(-N^{1/2} \inf_{B \in \mathcal{B}_{N,2}} \mathcal{P}(B \cap \mathcal{M})/3 \right) \\ &\geq 1 - \frac{2(5^{d/2})V\tau}{\omega_d \delta_1^d \delta_2} \exp \left(-N^{1/2} p_{\min} \omega_d \delta_3^d \left(\frac{15}{16} \right)^{d/2} / 3 \right). \end{aligned}$$

Taking $\delta_1 := N^{-1}$, $\delta_2 := N^{-1}$, and $\delta_3 := N^{-1/(4d)}$ gives the bounds in the statement. \blacksquare

Lemma 49. Fix $\bar{\tau} \leq \tilde{\tau}_{\max,2}$. Construct a minimal $\bar{\tau}/6$ -net \mathcal{Z} of \mathcal{X} . Then,

$$i) |\mathcal{Z}| \leq 5^{d/2} \omega_d^{-1} \frac{V}{(\bar{\tau}/12)^d}.$$

$$ii) \sup_{x \in \mathcal{M}} |\{z_i \in \mathcal{Z} \mid x \in B_n(z_i, \bar{\tau}) \cap \mathcal{M}\}| \leq 30^d \left(1 + \frac{49\bar{\tau}^2}{36\tau^2} \right)^{d/2}.$$

Furthermore, with probability at least $1 - 2(5^{d/2})V\tau\omega_d^{-1}N^{d+1} \exp(-N^{1/4}p_{\min}\omega_d(15/16)^{d/2}/3)$, for N large enough,

iii) \mathcal{Z} is a $\bar{\tau}/2$ -net of \mathcal{M} and

$$iv) \inf_{z_i \in \mathcal{Z}} |B_n(z_i, \bar{\tau}) \cap \mathcal{X}| \geq (15/16)^{d/2} N p_{\min} \omega_d \bar{\tau}^d / 2.$$

Proof. Start by constructing a minimal $\bar{\tau}/6$ -net $\mathcal{Z} := \{z_i\}$ of \mathcal{X} , where \mathcal{X} has a metric space structure from its embedding into \mathbb{R}^n . Note that \mathcal{Z} is a $\bar{\tau}/6$ -packing of \mathcal{X} and therefore of \mathcal{M} . A $\bar{\tau}/6$ -packing of a metric space has cardinality bounded above by that of any given $\bar{\tau}/12$ -net. Otherwise, there would be two elements of the packing within $\bar{\tau}/12$ of a member of the net, or within $\bar{\tau}/6$ of each other, which is a contradiction. In particular, using Lemma 47 to bound the size of a minimal $\bar{\tau}/12$ -net of \mathcal{M} , we have i):

$$|\mathcal{Z}| \leq 5^{d/2} \omega_d^{-1} \frac{V}{(\bar{\tau}/12)^d}.$$

Next, in ii), we prove an upper bound on the number of $z_i \in \mathcal{Z}$ that are within $\bar{\tau}$ of any point of \mathcal{M} . Fix $x \in \mathcal{M}$. Note that $\{B_n(z_i, \bar{\tau}/12) \cap \mathcal{M} \mid z_i \in B_n(x, \bar{\tau}) \cap \mathcal{Z}\}$ are pairwise disjoint, and by the injectivity of Π_x within a radius of τ , so are $\{\Pi_x(B_n(z_i, \bar{\tau}/12) \cap \mathcal{M}) \mid z_i \in B_n(x, \bar{\tau}) \cap \mathcal{Z}\}$. Note that $B_d(\Pi_x(z_i), (1 + 49\bar{\tau}^2/36\tau^2)^{-1/2}\bar{\tau}/12) \subset \Pi_x(B_n(z_i, \bar{\tau}/12) \cap \mathcal{M})$, so the maximum cardinality of a $(1 + 49\bar{\tau}^2/36\tau^2)^{-1/2}\bar{\tau}/6$ -packing of $B_d(x, 13\bar{\tau}/12)$ is the desired upper bound. By Lemma 47, this is no greater than $30^d(1 + 49\bar{\tau}^2/36\tau^2)^{d/2}$.

It remains to be shown that, w.h.p., $\mathcal{M} \subset \bigcup_i B_n(z_i, \bar{\tau}/2)$ and $|B_n(z_i, \bar{\tau}) \cap \mathcal{X}|$ has a uniform lower bound strictly greater than zero and increasing in N . (These are the statements from parts iii) and iv), respectively). The latter is a straightforward application of Lemma 48, in which we proved a uniform bound on the number of points contained in any region that is the intersection of \mathcal{M} with an n -dimensional ball centered on \mathcal{M} . In other words, with high probability, $\inf_{z_i \in \mathcal{Z}} |B_n(z_i, \bar{\tau}) \cap \mathcal{X}|$ is not much smaller than $\inf_{x \in \mathcal{M}} N\mathcal{P}(B_n(x, \bar{\tau}) \cap \mathcal{M})$, which has a lower bound of $Np_{min}\omega_d \left(\bar{\tau}\sqrt{15/16}\right)^d$ by an application of Lemma 35 iv). The error bound in the probability statement from Lemma 48 is much less than this for large enough N . (If an explicit bound is desired, we can set each term equal to 1/4 of the expected number of points and solve for N). Thus,

$$\begin{aligned} \mathbb{P} \left[\inf_{z_i \in \mathcal{Z}} |B_n(z_i, \bar{\tau}) \cap \mathcal{X}| \geq (15/16)^{d/2} Np_{min}\omega_d \bar{\tau}^d / 2 \right] &\geq 1 - 2(5^{d/2})V\tau\omega_d^{-1}N^{d+1} \\ &\quad \times \exp\left(-N^{1/4}p_{min}\omega_d(15/16)^{d/2}/3\right). \end{aligned}$$

We can demonstrate that $\{B_n(z_i, \bar{\tau}/2) \mid z_i \in \mathcal{Z}\}$ is a cover of \mathcal{M} by relating \mathcal{Z} to any

minimal $\bar{\tau}/6$ -net of \mathcal{M} . Consider one such net $S_{\bar{\tau}/6} := \{s_{\bar{\tau}/6,i}\}$. By another application of Lemma 48, we see that $\inf_{s_{\bar{\tau}/6,i} \in S_{\bar{\tau}/6}} |B_n(s_{\bar{\tau}/6,i}, \bar{\tau}/6) \cap \mathcal{X}| \geq 37^{-d/2} N p_{\min} \omega_d \bar{\tau}^d / 2$ with the same probability as above. This shows that \mathcal{Z} is a $\bar{\tau}/2$ -net of \mathcal{M} as follows. Any point of \mathcal{M} is within $\bar{\tau}/6$ of an element of $S_{\bar{\tau}/6}$, which is within $\bar{\tau}/6$ of a point of \mathcal{X} w.h.p. (since $\inf_{s_{\bar{\tau}/6,i} \in S_{\bar{\tau}/6}} |B_n(s_{\bar{\tau}/6,i}, \bar{\tau}/6) \cap \mathcal{X}| > 0$), which is in turn within $\bar{\tau}/6$ of some element of \mathcal{Z} . Thus, every point of \mathcal{M} is within $\bar{\tau}/2$ of a member of \mathcal{Z} . \blacksquare

Theorem 50. *Let \mathcal{Z} be a minimal $\bar{\tau}/6$ -net of \mathcal{X} , where $\bar{\tau} \leq \tilde{\tau}_{\max,2}$. For each $z_i \in \mathcal{Z}$, estimate $T_{z_i} \mathcal{M}$ with $\widehat{T_{z_i, \bar{\tau}} \mathcal{M}}$, the subspace with orthonormal basis given by the d largest eigenvectors of $N_{z_i, \bar{\tau}}^{-1} X_{z_i, \bar{\tau}} X_{z_i, \bar{\tau}}^\top$, where $X_{z_i, \bar{\tau}} := B_n(z_i, \bar{\tau}) \cap \mathcal{X}$ and $N_{z_i, \bar{\tau}} := |B_n(z_i, \bar{\tau}) \cap \mathcal{X}|$. Center the coordinate system at z_i and let the first d coordinates lie in $\widehat{T_{z_i, \bar{\tau}} \mathcal{M}}$. Define $\widehat{U_{\bar{\tau}}^{z_i}} := B_d(z_i, \bar{\tau}) \times B_{n-d}(z_i, \bar{\tau})$, $\widehat{\mathcal{A}_{z_i, \bar{\tau}}} := \widehat{U_{\bar{\tau}}^{z_i}} \cap \widehat{T_{z_i, \bar{\tau}} \mathcal{M}}$, and $\widetilde{\mathcal{A}_{z_i, \bar{\tau}}} := \widehat{U_{\bar{\tau}}^{z_i}} \cap \mathcal{M}$. Let $G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} : \widetilde{\mathcal{A}_{z_i, \bar{\tau}}} \rightarrow \widehat{\mathcal{A}_{z_i, \bar{\tau}}}$ be the orthogonal projection of \mathcal{M} onto $\widehat{T_{z_i, \bar{\tau}} \mathcal{M}}$. Then, with probability no less than $1 - 2(5^{d/2}) V \tau \omega_d^{-1} N^{d+1} \times \exp\left(-N^{1/4} p_{\min} \omega_d (15/16)^{d/2} / 3\right) - 5^{d/2} \omega_d^{-1} V / (\bar{\tau}/12)^d \times 2d \exp\left(- (2/3)^{d/2} \bar{\tau}^{d+2} N \omega_d p_{\min}^2 / 12(d+2) p_{\max}\right)$, the following statements hold.*

i) Each member of $\left\{G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} \mid z_i \in \mathcal{Z}\right\}$ is a well-defined local C^2 parametrization of \mathcal{M} possessing a derivative whose operator norm and Lipschitz constant are uniformly bounded above.

ii) The collection $\mathcal{C}_{\mathcal{Z}} := \left\{\left(\widehat{\mathcal{A}_{z_i, \bar{\tau}}}, G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}\right) \mid z_i \in \mathcal{Z}\right\}$ is a C^2 atlas for \mathcal{M} .

Proof. This proof is conditional on the statements of Lemma 49, which hold w.h.p. Recall from Theorem 41 that local PCA produces a subspace $\widehat{T_{z_i, \bar{\tau}} \mathcal{M}}$ such that $\left\|\sin \Theta\left(\widehat{T_{z_i, \bar{\tau}} \mathcal{M}}, T_{z_i} \mathcal{M}\right)\right\|_F$ is bounded above by ε_p , where ε_p is $O(\bar{\tau}/\tau)$ with probability at least $1 - 2d \exp\left(- (5/7)^{d/2} \bar{\tau}^2 N_{z_i, \bar{\tau}} p_{\min} / 6(d+2)\right)$. By Lemma 49 iv), $\inf_{z_i \in \mathcal{Z}} N_{z_i, \bar{\tau}} \geq (15/16)^{d/2} N p_{\min} \omega_d \bar{\tau}^d / 2$, implying that the error probability of local PCA is uniformly bounded for all charts of radius $\bar{\tau}$ (taken individually). The bound on $|\mathcal{Z}|$ from Lemma 49 i), together with the union bound, gives the probability that the domain of every chart in $\mathcal{C}_{\mathcal{Z}}$ is close in angular distance to its corresponding true tangent

space and thus meets the assumptions required for the conclusions of Lemma 42 and Theorem 44 to hold. Part i) of the present lemma follows immediately from the latter, which shows that the $\left\{G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} \mid z_i \in \mathcal{Z}\right\}$ are C^2 diffeomorphisms with derivatives that all have the stated uniform bounds on their operator norms and Lipschitz constants.

For ii), we need to show that $\mathcal{M} \subset \bigcup_{z_i \in \mathcal{Z}} \widehat{\tilde{\mathcal{A}}_{z_i, \bar{\tau}}}$ and that the charts in $\mathcal{C}_{\mathcal{Z}}$ are C^2 compatible with each other. The former is true because $\mathcal{M} \subset \bigcup_{z_i \in \mathcal{Z}} \tilde{\mathcal{A}}_{z_i, \bar{\tau}/2} \subset \bigcup_{z_i \in \mathcal{Z}} \widehat{\tilde{\mathcal{A}}_{z_i, c'\bar{\tau}}} \subset \bigcup_{z_i \in \mathcal{Z}} \widehat{\tilde{\mathcal{A}}_{z_i, c''\bar{\tau}}}$, where $c' := (1 + \varepsilon_p \sqrt{2})/2$ and $c'' := 9/16$. The first inclusion follows from Lemma 49 iii) and the fact that $\Pi_{z_i}(B_n(z_i, \bar{\tau}/2) \cap \mathcal{M}) \subset B_d(z_i, \bar{\tau}/2)$. The second and third are due to Lemma 42 i) and $\bar{\tau}$ being small enough such that $\varepsilon_p < 1/12$. Finally, compatibility of the charts follows from $\left\{G_{z_j, \widehat{U_{\bar{\tau}}^{z_j}}}^{-1} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}} \Big|_{G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}(\widehat{\tilde{\mathcal{A}}_{z_i, \bar{\tau}} \cap \widehat{\tilde{\mathcal{A}}_{z_j, \bar{\tau}}})} \mid z_i, z_j \in \mathcal{Z}\right\}$ being C^2 by composition. \blacksquare

3.6.2 Excess Risk and Convergence of \hat{f}

After sampling the net \mathcal{Z} and constructing $\mathcal{C}_{\mathcal{Z}}$, we carry out a structural risk minimization procedure separately on each $\widehat{U_{\bar{\tau}}^{z_i}}$, giving a collection of local estimators $\left\{\widehat{h_{z_i, \bar{\tau}}} \mid z_i \in \mathcal{Z}\right\}$. We then use a partition of unity $\left\{\alpha_{z_i} : \widehat{\tilde{\mathcal{A}}_{z_i, \bar{\tau}}} \rightarrow \mathbb{R} \mid z_i \in \mathcal{Z}\right\}$ to define the estimator of f^* as

$$\hat{f}(x) = \sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(\widehat{h_{z_i, \bar{\tau}}} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} \right)(x).$$

Besides having locally finite support, the partition of unity must also satisfy $0 \leq \alpha_{z_i}(x) \leq 1$ for all $z_i \in \mathcal{Z}, x \in \mathcal{M}$ and $\sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) = 1$ for all $x \in \mathcal{M}$. These conditions are satisfied when each α_{z_i} is defined through the assignment

$$x \mapsto \frac{\tilde{\alpha}_{z_i} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}(x)}{\sum_{z_i \in \mathcal{Z}} \tilde{\alpha}_{z_i} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}(x)},$$

where $\tilde{\alpha}_{z_i}$ is a recentered and rotated version of a C^2 bump function $\tilde{\alpha}$ on $B_d(0, \bar{\tau})$. We will not specify the exact form of $\tilde{\alpha}$, but we will note that it can be chosen so that $\|\tilde{\alpha}\|_{C^{1,1}(\mathbb{R}^d)} < \infty$ and $\inf_{x \in \mathcal{M}} \sum_{z_i \in \mathcal{Z}} \tilde{\alpha}_{z_i} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}(x) > 0$, implying that $\sup_{z_i \in \mathcal{Z}} \|\alpha_{z_i}\|_{C^{1,1}(\mathcal{M})} < \infty$.

Define the class of functions $\mathcal{F}_{z_i, \bar{\tau}, \tilde{M}} := \left\{ f : \mathcal{M} \rightarrow \mathbb{R} \mid \|f_{z_i, \bar{\tau}}\|_{C^{1,1}(\mathbb{R}^d)} \leq \tilde{M}, \text{ where } f(x) = \sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(f_{z_i, \bar{\tau}} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} \right)(x) \right\}$. \tilde{M} is chosen as an increasing function of N . Define $\dot{\mathcal{F}}_{z_i, \bar{\tau}, \tilde{M}}$ similarly, but with the constraint on the $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm. Note that checking if a particular f is in either of these classes requires finding a suitable decomposition $\{f_{z_i, \bar{\tau}}\}$, which may not be unique. Now define $\mathcal{F}_{\mathcal{Z}, \bar{\tau}, \tilde{M}} := \bigcap_{z_i \in \mathcal{Z}} \mathcal{F}_{z_i, \bar{\tau}, \tilde{M}}$, and $\dot{\mathcal{F}}_{\mathcal{Z}, \bar{\tau}, \tilde{M}} := \bigcap_{z_i \in \mathcal{Z}} \dot{\mathcal{F}}_{z_i, \bar{\tau}, \tilde{M}}$. Our algorithm finds $\hat{f} \in \dot{\mathcal{F}}_{\mathcal{Z}, \bar{\tau}, \tilde{M}}$. In the next lemma, we show that $\hat{f} \in \mathcal{F}_{\mathcal{Z}, \bar{\tau}, \tilde{M}}$ as well, w.h.p.; this is required for the derivation of risk bounds via the $C^{1,1}(\mathbb{R}^d)$ case. We also show that $f^* \in \mathcal{F}_{\mathcal{Z}, \bar{\tau}, \tilde{M}}$ for large enough N , meaning that $\{h_{z_i, \bar{\tau}}^*\}$ are eventually candidate functions for all z_i . The results in this section are conditional on Section 3.6.1. Specifically, we assume that \mathcal{Z} is a minimal $\bar{\tau}/6$ -net of \mathcal{X} that is also a $\bar{\tau}/2$ -net of \mathcal{M} satisfying the conclusions of Lemma 49. We also assume that $\mathcal{C}_{\mathcal{Z}}$ is a collection of charts forming a C^2 atlas for \mathcal{M} and satisfying the conclusions of Theorem 50.

Lemma 51. *Choose $\tilde{M} = o(N^{2/d})$ and increasing in N . Then, the following statements hold for large enough N .*

- i) $\mathbb{P}\left[\hat{f} \in \mathcal{F}_{\mathcal{Z}, \bar{\tau}, \tilde{M}}\right] \geq 1 - 2(5^{d/2})\omega_d^{-1} \frac{V}{(\bar{\tau}/12)^d} \exp\left(-\left((15/16)^{d/2} N p_{\min} \omega_d (8\bar{\tau}/9)^d / 2\right)^{1/100}\right)$
- ii) $f^* \in \mathcal{F}_{\mathcal{Z}, \bar{\tau}, \tilde{M}}$

Proof. i) By Lemma 45, the pushforwards $\{\mathcal{P}_{z_i, \bar{\tau}} \mid z_i \in \mathcal{Z}\}$ of \mathcal{P} to $\{\widehat{\mathcal{A}}_{z_i, \bar{\tau}} \mid z_i \in \mathcal{Z}\}$ have densities uniformly bounded away from zero and infinity. Thus, by Theorem 32, each $\widehat{h}_{z_i, \bar{\tau}}$ has $C^{1,1}(\mathbb{R}^d)$ norm no greater than \tilde{M} with probability at least $1 - 2 \exp\left(-\left(\inf_{z_i \in \mathcal{Z}} |\mathcal{X}'_{z_i}|\right)^{1/100}\right)$, where $\mathcal{X}'_{z_i} := G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}\left(\widehat{\mathcal{A}}_{z_i, \bar{\tau}} \cap \mathcal{X}\right)$, as long as $\tilde{M} = o\left(|\mathcal{X}'_{z_i}|^{2/d}\right)$. The setup of the present lemma assumes the uniform bound of Lemma 48, which we can employ after noting that $\bar{\tau} < \tilde{\tau}_{\max, 2}$ implies $\varepsilon_p < 1/12$. Then, $\tilde{\mathcal{A}}_{z_i, 8\bar{\tau}/9} \subset \widehat{\mathcal{A}}_{z_i, \bar{\tau}} \subset \tilde{\mathcal{A}}_{z_i, 8\bar{\tau}/7}$ by Lemma 42 i), which gives

$$\begin{aligned} \inf_{z_i \in \mathcal{Z}} |\mathcal{X}'_{z_i}| &\geq \inf_{x \in \mathcal{M}} N \mathcal{P}(B_n(x, 8\bar{\tau}/9) \cap \mathcal{M})/2 \\ &\geq (15/16)^{d/2} N p_{\min} \omega_d (8\bar{\tau}/9)^d / 2 \end{aligned}$$

and

$$\begin{aligned} \sup_{z_i \in \mathcal{Z}} |\mathcal{X}'_{z_i}| &\leq \sup_{x \in \mathcal{M}} 2N\mathcal{P}\left(B_n\left(x, \left(1 + \frac{49\bar{\tau}^2}{36\tau^2}\right)^{1/2} 8\bar{\tau}/7\right) \cap \mathcal{M}\right) \\ &\leq 2\left(1 + \frac{49\bar{\tau}^2}{36\tau^2}\right)^d Np_{max}\omega_d(8\bar{\tau}/7)^d \end{aligned}$$

for large enough N . That is, $C_{d,1}N \leq |\mathcal{X}'_{z_i}| \leq C_{d,2}N$ for every z_i , which is sufficient to show that \widetilde{M} can be chosen as $o(N^{2/d})$. Applying the union bound and then Lemma 49 i) yields the statement.

ii) Clearly, $f^*(x)$ can be written as $\sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(h_{z_i, \bar{\tau}}^* \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}\right)(x)$, and Theorem 46 shows that $\sup_{z_i \in \mathcal{Z}} \|h_{z_i, \bar{\tau}}^*\|_{C^{1,1}(\mathbb{R}^d)}$ is uniformly bounded above by $M_{h, \bar{\tau}}^* = o(\widetilde{M})$. ■

Theorem 52. *Let the function $\widehat{f} : \mathcal{M} \rightarrow \mathbb{R}$ defined as*

$$x \mapsto \sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(\widehat{h_{z_i, \bar{\tau}}} \circ G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}\right)(x)$$

be an estimator of f^* , where each $\widehat{h_{z_i, \bar{\tau}}}$ is a solution to the minimization problem

$$\arg \inf_{f \in \widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}} \frac{1}{|\mathcal{X}'_{z_i}|} \sum_{(x_i, y_i) \in \mathcal{X}'_{z_i} \times \mathcal{Y}_{z_i}} (y_i - f(x_i))^2$$

carried out on its respective $\widehat{T_{z_i, \bar{\tau}}\mathcal{M}}$. $\mathcal{X}'_{z_i} := G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1}\left(\widehat{\widetilde{\mathcal{A}}_{z_i, \bar{\tau}}} \cap \mathcal{X}\right)$, \mathcal{Y}_{z_i} is the corresponding subset of \mathcal{Y} , and $\widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}} := \left\{f : \widehat{\mathcal{A}}_{z_i, \bar{\tau}} \rightarrow \mathbb{R} \mid \|f\|_{C^{1,1}(\mathbb{R}^d)} \leq \widetilde{M}\right\}$. $\{\alpha_{z_i} \mid z_i \in \mathcal{Z}\}$ is a partition of unity subordinate to $\left\{\widehat{\widetilde{\mathcal{A}}_{z_i, \bar{\tau}}} \mid z_i \in \mathcal{Z}\right\}$ with $\sup_{z_i \in \mathcal{Z}} \|\alpha_{z_i}\|_{C^{1,1}(\mathcal{M})} < \infty$. Assume that $\widehat{f}, f^* \in \mathcal{F}_{\mathcal{Z}, \bar{\tau}, \widetilde{M}}$. Then, the following statements hold.

$$i) \left\|\widehat{f}\right\|_{C^{1,1}(\mathcal{M})} < \infty.$$

Let $N_{\mathcal{Z}, min} := (15/16)^{d/2} Np_{min}\omega_d(8\bar{\tau}/9)^d/2$. Let $\widetilde{d} := \max\{d, 5\}$. Define $\widetilde{L}_{max} := \left(\widetilde{M} +$

$M_{h,\bar{\tau}}^* + \sigma\sqrt{2\log 2N} + N^{1/\tilde{d}})^2$, $\tilde{L}_L := 2\left(\tilde{M} + M_{h,\bar{\tau}}^* + \sigma\sqrt{2\log 2N} + N^{1/\tilde{d}}\right)$, and

$$\tilde{R} := \begin{cases} 4\frac{81^{1/d}K^{2/d}\tilde{M}}{N_{\mathcal{Z},\min}^{2/d}} + 12\frac{\sqrt{K}\tilde{M}^{d/4}\left(4\tilde{M}^{1-d/4} - 4\left(81^{1/d}K^{2/d}\tilde{M}N_{\mathcal{Z},\min}^{-2/d}\right)^{1-d/4}\right)}{(4-d)\sqrt{N_{\mathcal{Z},\min}}} & : d \neq 4 \\ 4\left(3\sqrt{K}\tilde{M}N_{\mathcal{Z},\min}^{-1/2}\right) + 12\frac{\sqrt{K}\tilde{M}}{\sqrt{N_{\mathcal{Z},\min}}}\left(\log \tilde{M} - \log\left(3\sqrt{K}\tilde{M}N_{\mathcal{Z},\min}^{-1/2}\right)\right) & : d = 4. \end{cases}$$

ii) Set $\tilde{M} := N^{1/(2\tilde{d})}$. For $\delta \in (0, 1)$,

$$\mathbb{P}\left[R(\hat{f}) - R(f^*) \leq 2(27^d)Vp_{\max}\left(1 + \frac{64}{25}\left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p\right)^2\right)^{d/2} \varepsilon\right] \geq 1 - 5^{d/2}\omega_d^{-1}\frac{V}{(\bar{\tau}/12)^d}\left(\delta + e^{-N_{\mathcal{Z},\min}^{2/\tilde{d}}/(2\sigma^2)}\right),$$

where $\varepsilon := 4\tilde{L}_L\tilde{R} + 7\tilde{L}_{\max}\sqrt{\log(8/\delta)/2N_{\mathcal{Z},\min}}$. ε is a monotonically-decreasing function of N for large enough N and $\lim_{N \rightarrow \infty} \varepsilon = 0$.

iii) Set $\tilde{M} := N^{1/(16\tilde{d}^2)}$. The conclusions of **ii)** still hold. For $\delta \in (0, 1)$,

$$\mathbb{P}\left[\sup_{x \in \mathcal{M}} |\hat{f}(x) - f^*(x)| \leq \beta\right] \geq 1 - 5^{d/2}\omega_d^{-1}\frac{V}{(\bar{\tau}/12)^d}\left(\delta + e^{-N_{\mathcal{Z},\min}^{2/\tilde{d}}/(2\sigma^2)}\right),$$

where $\beta := \left(1 + \frac{64}{25}\left(\frac{7\bar{\tau}}{5\tau} + \varepsilon_p\right)^2\right)^{1/2} \tilde{M}\bar{\tau}\left(\frac{2\varepsilon N^{1/(5\tilde{d})}p_{\max}}{p_{\min}}\right)^{1/d} + N_{\mathcal{Z},\min}^{-1/(10\tilde{d})}$. β is a monotonically-decreasing function of N for large enough N and $\lim_{N \rightarrow \infty} \beta = 0$.

Proof. **i)** It holds that $\hat{f} \in C^{1,1}(\mathcal{M})$. Since $\alpha_{z_i} \in C^{1,1}(\mathcal{M})$, $\widehat{h_{z_i,\bar{\tau}}} \in C^{1,1}(\mathbb{R}^d)$, and $G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} \in C^{1,1}(\mathcal{M})$ for every $z_i \in \mathcal{Z}$, the linearity of the derivative and the product and chain rules show that \hat{f} is differentiable. Furthermore, $D\hat{f}$ is the sum of products of Lipschitz functions and is therefore Lipschitz itself. It also holds that $\|\hat{f}\|_{C^{1,1}(\mathcal{M})} < \infty$. Theorem 44 shows that $\sup_{z_i \in \mathcal{Z}} \left\| G_{z_i, \widehat{U_{\bar{\tau}}^{z_i}}}^{-1} \right\|_{C^{1,1}(\mathcal{M})} < \infty$, and by assumption, $\sup_{z_i \in \mathcal{Z}} \|\alpha_{z_i}\|_{C^{1,1}(\mathcal{M})} < \infty$. Additionally, we are conditioning on the event that $\hat{f} \in \mathcal{F}_{\mathcal{Z},\bar{\tau},\tilde{M}}$, which holds w.h.p. for large enough N and

suitable \widetilde{M} by Lemma 51 i), implying that $\sup_{z_i \in \mathcal{Z}} \left\| \widehat{h_{z_i, \bar{\tau}}} \right\|_{C^{1,1}(\mathbb{R}^d)} < \infty$. Since $|\mathcal{Z}| < \infty$, it follows that \widehat{f} has finite $C^{1,1}(\mathcal{M})$ norm.

ii) By a standard decomposition of the risk, $R(\widehat{f}) - R(f^*) = \mathbb{E} \left[\left(\widehat{f} - f^* \right)^2 \right]$. After substituting the definition of \widehat{f} and writing f^* in terms of its pullbacks to $\left\{ \widehat{T_{z_i, \bar{\tau}} \mathcal{M}} \mid z_i \in \mathcal{Z} \right\}$, we can apply Jensen's inequality pointwise to see that

$$\begin{aligned} \mathbb{E} \left[\left(\widehat{f} - f^* \right)^2 \right] &= \mathbb{E} \left[\left(\sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(\left(\widehat{h_{z_i, \bar{\tau}}} - h_{z_i, \bar{\tau}}^* \right) \circ G_{z_i, \widehat{U_{z_i, \bar{\tau}}}}^{-1} \right)(x) \right)^2 \right] \\ &\leq \mathbb{E} \left[\sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(\left(\left(\widehat{h_{z_i, \bar{\tau}}} - h_{z_i, \bar{\tau}}^* \right) \circ G_{z_i, \widehat{U_{z_i, \bar{\tau}}}}^{-1} \right)(x) \right)^2 \right] \\ &\leq \sum_{z_i \in \mathcal{Z}} \mathcal{P} \left(\widehat{\mathcal{A}_{z_i, \bar{\tau}}} \right) \int_{\widehat{\mathcal{A}_{z_i, \bar{\tau}}}} \left(\widehat{h_{z_i, \bar{\tau}}}(x') - h_{z_i, \bar{\tau}}^*(x') \right)^2 d\mathcal{P}_{z_i, \bar{\tau}}(x') \\ &\leq |\mathcal{Z}| \sup_{z_i \in \mathcal{Z}} \mathcal{P} \left(\widehat{\mathcal{A}_{z_i, \bar{\tau}}} \right) \sup_{z_i \in \mathcal{Z}} \left(R(\widehat{h_{z_i, \bar{\tau}}}) - R(h_{z_i, \bar{\tau}}^*) \right). \end{aligned}$$

We have $|\mathcal{Z}| \sup_{z_i \in \mathcal{Z}} \mathcal{P} \left(\widehat{\mathcal{A}_{z_i, \bar{\tau}}} \right) \leq 27^d V p_{max} \left(1 + 64/25 (7\bar{\tau}/5\tau + \varepsilon_p)^2 \right)^{d/2}$ by Lemma 49 i) and Theorem 33. $\sup_{z_i \in \mathcal{Z}} \left(R(\widehat{h_{z_i, \bar{\tau}}}) - R(h_{z_i, \bar{\tau}}^*) \right)$ can be bounded above w.h.p. Let $\widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}} := \left\{ f : \widehat{\mathcal{A}_{z_i, \bar{\tau}}} \rightarrow \mathbb{R} \mid \|f\|_{C^{1,1}(\mathbb{R}^d)} \leq \widetilde{M} \right\}$. By assumption, $\widehat{f}, f^* \in \mathcal{F}_{\mathcal{Z}, \bar{\tau}, \widetilde{M}}$, so $\widehat{h_{z_i, \bar{\tau}}}, h_{z_i, \bar{\tau}}^* \in \widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}$ for all z_i . For fixed z_i , $R(\widehat{h_{z_i, \bar{\tau}}}) - R(h_{z_i, \bar{\tau}}^*) \leq 2 \sup_{f \in \widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}} \left| R(f) - \widehat{R}(f) \right| \leq 2\varepsilon'$ with probability at least $1 - \delta - e^{-N_{\mathcal{Z}, min}^{2/d}/(2\sigma^2)}$, where $\delta \in (0, 1)$ and ε' is as in Theorem 31. Note that $\varepsilon' \leq 4\widetilde{L}_L \widetilde{R} + 7\widetilde{L}_{max} \sqrt{\log(8/\delta)/2N_{\mathcal{Z}, min}} =: \varepsilon$ by comparison of each term, where $\widetilde{L}_L, \widetilde{L}_{max}$, and \widetilde{R} are as defined in the statement of the present theorem. In Lemma 51 i), we proved as an intermediate step that $C_{d,1}N \leq |\mathcal{X}'_{z_i}| \leq C_{d,2}N$ for every z_i ; this implies that the asymptotics of ε' and ε are the same. The union bound over \mathcal{Z} gives the result.

The random quantity $\sup_{f \in \widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}} \left| R(f) - \widehat{R}(f) \right|$ is measurable because of the pointwise measurability of $\widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}$. This was implicitly assumed by Gustafson et al. (2018). Every $f \in \widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}$ is Lipschitz continuous and defined on $\widehat{\mathcal{A}_{z_i, \bar{\tau}}}$, a bounded subset of \mathbb{R}^d , so it can be extended continuously to $\partial \widehat{\mathcal{A}_{z_i, \bar{\tau}}}$. Let $\widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}^{ext}$ be the class consisting of these extensions. $\widetilde{\mathcal{F}}_{z_i, \bar{\tau}, \widetilde{M}}^{ext} \subset C^0 \left(\widehat{\mathcal{A}_{z_i, \bar{\tau}}} \right)$, the space of continuous functions on $\widehat{\mathcal{A}_{z_i, \bar{\tau}}}$. $C^0 \left(\widehat{\mathcal{A}_{z_i, \bar{\tau}}} \right)$ equipped with

the sup norm is separable by the Stone-Weierstrass Theorem, which can be used to show that the space of continuous functions on a compact metric space has a countable dense subset. As a subset of a separable metric space, $\tilde{\mathcal{F}}_{z_i, \bar{\tau}, \tilde{M}}^{ext}$ contains a countable dense subset itself, which we denote $\tilde{\mathcal{F}}_{z_i, \bar{\tau}, \tilde{M}, 0}^{ext}$. Clearly, $\left\{ f|_{\widehat{\mathcal{A}}_{z_i, \bar{\tau}}} \mid f \in \tilde{\mathcal{F}}_{z_i, \bar{\tau}, \tilde{M}, 0}^{ext} \right\} \subset \tilde{\mathcal{F}}_{z_i, \bar{\tau}, \tilde{M}}$ is countable and can be used to construct a pointwise convergent sequence of functions for any $f \in \tilde{\mathcal{F}}_{z_i, \bar{\tau}, \tilde{M}}$.

iii) The following holds because the partition of unity is a convex combination for each x : $\sup_{x \in \mathcal{M}} |\hat{f}(x) - f^*(x)| = \sup_{x \in \mathcal{M}} \left| \sum_{z_i \in \mathcal{Z}} \alpha_{z_i}(x) \left(\left(\widehat{h}_{z_i, \bar{\tau}} - h_{z_i, \bar{\tau}}^* \right) \circ G_{z_i, \widehat{U}_{\bar{\tau}}^{z_i}}^{-1} \right)(x) \right| \leq \sup_{z_i \in \mathcal{Z}} \left\| \widehat{h}_{z_i, \bar{\tau}} - h_{z_i, \bar{\tau}}^* \right\|_{\infty}$. For each $z_i \in \mathcal{Z}$, $\left\| \widehat{h}_{z_i, \bar{\tau}} - h_{z_i, \bar{\tau}}^* \right\|_{\infty} \leq \beta'$, defined in Theorem 31 ii), w.h.p. It holds that $\beta' \leq \tilde{M} \left(2\varepsilon N^{1/(5\tilde{d})} / (\tilde{p}_{min} \omega_d) \right)^{1/d} + N_{\mathcal{Z}, min}^{-1/(10\tilde{d})} =: \beta$. \tilde{p}_{min} is a uniform lower bound on the densities of the pushforward measures $\{\mathcal{P}_{z, \bar{\tau}} \mid z_i \in \mathcal{Z}\}$; it can be taken as the value given in Lemma 45. Substituting this for \tilde{p}_{min} and applying the union bound gives the result. The asymptotic behavior of β is the same as that of β' . ■

3.7 Discussion

In this chapter, we considered the problem of recovering a $C^{1,1}(\mathcal{M})$ function f^* from noisy observations. We described an approach based on local $C^{1,1}(\mathbb{R}^d)$ regression on estimated tangent spaces (indexed by a fine-enough net of the sample) and showed that it does recover f^* as the sample size increases.

The $C^{1,1}(\mathbb{R}^d)$ regression algorithm we used is due to our collaborators (Gustafson et al., 2018) and extends previous work on $C^{1,1}(\mathbb{R}^d)$ function interpolation (Wells, 1973; Le Gruyer, 2009; Herbert-Voss, Hirn, and McCollum, 2017). They solve a convex optimization problem to find the empirical risk minimizer within a $\dot{C}^{1,1}(\mathbb{R}^d)$ seminorm ball whose diameter is increasing with the sample size. Our contribution was to use empirical process methods to bound probabilistically the difference between the empirical risk and the true risk and to use this to derive the convergence rate of the estimator to the true function in sup norm. We included these results in this chapter.

In order to extend this to a $C^{1,1}(\mathcal{M})$ regression algorithm, we proved that the sample con-

tains enough information to define a suitable C^2 atlas of \mathcal{M} w.h.p. We showed that tangent spaces estimated with local PCA are sufficiently close to the actual tangent spaces w.h.p., allowing us to define charts for \mathcal{M} whose derivatives have uniformly bounded Lipschitz constants. We also proved several other facts about the sample; e.g., we found uniform bounds for the convergence of the empirical measure indexed by sets of interest, and we showed that the sample contains a suitably-fine net of \mathcal{M} with bounded cardinality. We estimated f^* by using a partition of unity to combine local estimators (obtained via $C^{1,1}(\mathbb{R}^d)$ regression) of the pullbacks of f^* to the estimated tangent spaces. We used properties of the C^2 atlas along with our $C^{1,1}(\mathbb{R}^d)$ regression sample complexity results to find risk bounds and convergence rates for the estimator of f^* .

BIBLIOGRAPHY

- Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *arXiv preprint arXiv:1705.04565*, 2017.
- Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.
- Peter L Bartlett and Shahar Mendelson. Rademacher and Gaussian complexities: Risk bounds and structural results. *The Journal of Machine Learning Research*, 3:463–482, 2003.
- Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.
- Jean-Daniel Boissonnat, Ramsay Dyer, and Arijit Ghosh. Constructing intrinsic Delaunay triangulations of submanifolds. *arXiv preprint arXiv:1303.6493*, 2013.
- Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. 2017.
- Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.
- Olivier Bousquet, Stéphane Boucheron, and Gábor Lugosi. Introduction to statistical learning theory. In *Advanced Lectures on Machine Learning*, pages 169–207. Springer, 2004.
- Yen-Chi Chen, Christopher R Genovese, Larry Wasserman, et al. Asymptotic theory for density ridges. *The Annals of Statistics*, 43(5):1896–1928, 2015.

- Ming-Yen Cheng, Peter Hall, John A Hartigan, et al. Estimating gradient trees. In *A Festschrift for Herman Rubin*, pages 237–249. Institute of Mathematical Statistics, 2004.
- David Eberly. *Ridges in Image and Data Analysis*, volume 7. Springer Science & Business Media, 1996.
- Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, pages 418–491, 1959.
- Herbert Federer. *Geometric Measure Theory*. Grundlehren der mathematischen Wissenschaften. Springer, 1969.
- Charles Fefferman and Bo’az Klartag. Fitting a C^m -smooth function to data I. *Annals of Mathematics*, pages 315–346, 2009.
- Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 2016.
- Christopher R Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *The Journal of Machine Learning Research*, 13(1):1263–1291, 2012a.
- Christopher R Genovese, Marco Perone-Pacifico, Isabella Verdinelli, Larry Wasserman, et al. Manifold estimation and singular deconvolution under Hausdorff loss. *The Annals of Statistics*, 40(2):941–963, 2012b.
- Christopher R Genovese, Marco Perone-Pacifico, Isabella Verdinelli, Larry Wasserman, et al. Nonparametric ridge estimation. *The Annals of Statistics*, 42(4):1511–1545, 2014.
- Adam Gustafson, Matthew Hirn, Kitty Mohammed, Hariharan Narayanan, and Jason Xu. Structural risk minimization for $C^{1,1}(\mathbb{R}^d)$ regression. *arXiv preprint arXiv:1803.10884*, 2018.

- Peter Hall, Wei Qian, and DM Titterton. Ridge finding from noisy data. *Journal of Computational and Graphical Statistics*, 1(3):197–211, 1992.
- Ariel Herbert-Voss, Matthew J Hirn, and Frederick McCollum. Computing minimal interpolants in $C^{1,1}(\mathbb{R}^d)$. *Revista Matemática Iberoamericana*, 33(1):29–66, 2017.
- S.G. Krantz and H.R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Modern Birkhäuser Classics. Springer New York, 2012.
- Erwan Le Gruyer. Minimal Lipschitz extensions to differentiable functions defined on a Hilbert space. *Geometric and Functional Analysis*, 19(4):1101–1118, 2009.
- Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2005.
- John Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 2nd edition, 2012.
- John Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer International Publishing, 2nd edition, 2018.
- Yunqian Ma and Yun Fu. *Manifold Learning Theory and Applications*. CRC press, 2011.
- Colin McDiarmid. On the method of bounded differences. *Surveys in Combinatorics*, 141(1):148–188, 1989.
- Arkadi Nemirovski. *Interior Point Polynomial Time Methods in Convex Programming*. http://www2.isye.gatech.edu/~nemirovs/Lect_IPM.pdf, 2004.
- Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

- Umut Ozertem and Deniz Erdogmus. Locally defined principal curves and surfaces. *The Journal of Machine Learning Research*, 12:1249–1286, 2011.
- Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- Karthik Sridharan and Nathan Srebro. Note on refined Dudley integral covering number bound, 2010.
- Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- Pravin M Vaidya. A new algorithm for minimizing convex functions over convex sets. *Mathematical Programming*, 73(3):291–341, 1996.
- Aad W Van Der Vaart and Jon A Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.
- John C Wells. Differentiable functions on Banach spaces with Lipschitz derivatives. *Journal of Differential Geometry*, 8(1):135–152, 1973.
- Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.