New Analysis of Manifold Embeddings and Signal Recovery from Compressive Measurements

Armin Eftekhari and Michael B. Wakin[*]

Department of Electrical Engineering and Computer Science, Colorado School of Mines

June 2013

Abstract

Compressive Sensing (CS) exploits the surprising fact that the information contained in a sparse signal can be preserved in a small number of compressive, often random linear measurements of that signal. Strong theoretical guarantees have been established concerning the embedding of a sparse signal family under a random measurement operator and on the accuracy to which sparse signals can be recovered from noisy compressive measurements. In this paper, we address similar questions in the context of a different modeling framework. Instead of sparse models, we focus on the broad class of manifold models, which can arise in both parametric and non-parametric signal families. Using tools from the theory of empirical processes, we improve upon previous results concerning the embedding of low-dimensional manifolds under random measurement operators. We also establish both deterministic and probabilistic instance-optimal bounds in ℓ_2 for manifold-based signal recovery and parameter estimation from noisy compressive measurements. In line with analogous results for sparsity-based CS, we conclude that much stronger bounds are possible in the probabilistic setting. Our work supports the growing evidence that manifold-based models can be used with high accuracy in compressive signal processing.

Keywords. Manifolds, Compressive Sensing, dimensionality reduction, random projections, manifold embeddings, signal recovery, parameter estimation.

AMS Subject Classification. 53A07, 57R40, 62H12, 68P30, 94A12, 94A29.

1 Introduction

1.1 Concise signal models

A significant byproduct of the Information Age has been an explosion in the sheer quantity of raw data demanded from sensing systems. From digital cameras to mobile devices, scientific computing to medical imaging, and remote surveillance to signals intelligence, the size (or dimension) N of a typical desired signal continues to increase. Naturally, the dimension N imposes a direct burden on the various stages of the data processing pipeline, from the data acquisition itself to the subsequent transmission, storage, and/or analysis.

Fortunately, in many cases, the information contained within a high-dimensional signal actually obeys some sort of concise, low-dimensional model. Such a signal may be described as having just

$K \ll N$ degrees of freedom for some K. Periodic signals bandlimited to a certain frequency are one example; they live along a fixed K-dimensional linear subspace of $\mathbb{R}^N$. Piecewise smooth signals are an example of *sparse signals*, which can be written as a succinct linear combination of just K elements from some basis such as a wavelet dictionary. Still other signals may live along K-dimensional submanifolds of the ambient signal space $\mathbb{R}^N$; examples include collections of signals observed from multiple viewpoints in a camera or sensor network. In general, the conciseness of these models suggests the possibility for efficient processing and compression of these signals.

1.2 Compressive measurements

Recently, the conciseness of certain signal models has led to the use of *compressive measurements* for simplifying the data acquisition process. Rather than designing a sensor to measure a signal $x \in \mathbb{R}^N$, for example, it often suffices to design a sensor that can measure a much shorter vector $y = \Phi x$, where Φ is a linear measurement operator represented as an $M \times N$ matrix, and where typically $M \ll N$. As we discuss below in the context of Compressive Sensing (CS), when Φ is properly designed, the requisite number of measurements M typically scales with the information level K of the signal, rather than with its ambient dimension N.

Surprisingly, the requirements on the measurement matrix Φ can often be met by choosing Φ randomly from an acceptable distribution. Most commonly, the entries of Φ are chosen to be independent and identically distributed (i.i.d.) Gaussian random variables, although the use of structured random matrices is on the rise [25, 37]. Physical architectures have been proposed for hardware that will enable the acquisition of signals using compressive measurements [10, 22, 30, 36]. The potential benefits for data acquisition are numerous. These systems can enable simple, low-cost acquisition of a signal directly in compressed form without requiring knowledge of the signal structure in advance. Some of the many possible applications include distributed source coding in sensor networks [23], medical imaging [40], and high-rate analog-to-digital conversion [10, 30, 56].

1.3 Signal understanding from compressive measurements

Having acquired a signal x in compressed form (in the form of a measurement vector y), there are many questions that may then be asked of the signal. These include:

Q1. *Recovery:* What was the original signal x?

Q2. *Sketching:* Supposing that x was sparse or nearly so, what were the K basis vectors used to generate x?

Q3. *Parameter estimation:* Supposing x was generated from a K-dimensional parametric model, what was the original K-dimensional parameter that generated x?

Given only the measurements y (possibly corrupted by noise), solving any of the above problems requires exploiting the concise, K-dimensional structure inherent in the signal.[1] CS addresses questions Q1 and Q2 under the assumption that the signal x is K-sparse (or approximately so) in some basis or dictionary; in Section 2.1 we outline some key theoretical bounds from CS regarding the accuracy to which these questions may be answered.

---

[1] Other problems, such as finding the nearest neighbor to x in a large database of signals [34], can also be solved using compressive measurements and do not require assumptions about the concise structure in x.
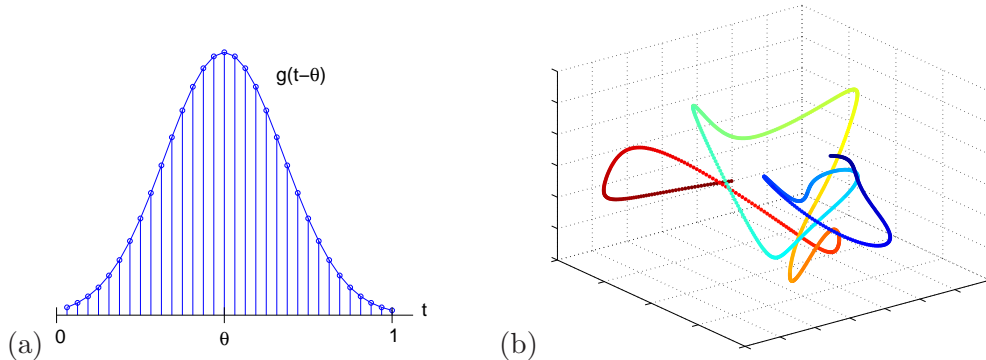
Figure 1: (a) The articulated signal $f_\theta(t) = g(t - \theta)$ is defined via shifts of a primitive function g, where g is a Gaussian pulse. Each signal is sampled at N points, and as θ changes, the resulting signals trace out a 1-D manifold in $\mathbb{R}^N$. (b) Projection of the manifold from $\mathbb{R}^N$ into $\mathbb{R}^3$ via a random $3 \times N$ matrix; the color/shading represents different values of $\theta \in [0, 1]$.

1.4 Manifold models for signal understanding

In this paper, we will address these questions in the context of a different modeling framework for concise signal structure. Instead of sparse models, we focus on the broad class of *manifold models*, which arise both in settings where a K-dimensional parameter θ controls the generation of the signal and also in non-parametric settings.

As a very simple illustration, consider the articulated signal in Figure 1(a). We let $g(t)$ be a fixed continuous-time Gaussian pulse centered at $t = 0$ and consider a shifted version of g denoted as the parametric signal $f_\theta(t) := g(t - \theta)$ with $t, \theta \in [0, 1]$. We then suppose the discrete-time signal $x = x_\theta \in \mathbb{R}^N$ arises by sampling the continuous-time signal $f_\theta(t)$ uniformly in time, i.e., $x_\theta(n) = f_\theta(n/N)$ for $n = 1, 2, \ldots, N$. As the parameter θ changes, the signals x_θ trace out a continuous one-dimensional (1-D) curve $\mathcal{M} = \{x_\theta : \theta \in [0, 1]\} \subset \mathbb{R}^N$. The conciseness of our model (in contrast with the potentially high dimension N of the signal space) is reflected in the low dimension of the path $\mathcal{M}$.

In the real world, manifold models may arise in a variety of settings. A K-dimensional parameter θ could reflect uncertainty about the 1-D timing of the arrival of a signal (as in Figure 1(a); see also [24]), the 2-D orientation and position of an edge in an image, the 2-D translation of an image under study [45], the multiple degrees of freedom in positioning a camera or sensor to measure a scene [15], the physical degrees of freedom in an articulated robotic or sensing system, or combinations of the above. Manifolds have also been proposed as approximate models for signal databases such as collections of images of human faces or of handwritten digits [4, 33, 57].

Consequently, the potential applications of manifold models are numerous in signal processing. In some applications, the signal x itself may be the object of interest, and the concise manifold model may facilitate the acquisition or compression of that signal. Alternatively, in parametric settings one may be interested in using a signal $x = x_\theta$ to infer the parameter θ that generated that signal. In an application known as manifold learning, one may be presented with a collection of data $\{x_{\theta_1}, x_{\theta_2}, \ldots, x_{\theta_n}\}$ sampled from a parametric manifold and wish to discover the underlying parameterization that generated that manifold. Multiple manifolds can also be considered simultaneously, for example in problems that require recognizing an object from one of n possible classes, where the viewpoint of the object is uncertain during the image capture process. In this case, we may wish to know which of n manifolds is closest to the observed image x.

While any of these questions may be answered with full knowledge of the high-dimensional signal $x \in \mathbb{R}^N$, there is growing theoretical and experimental support that they can also be answered from

only compressive measurements $y = \Phi x$. In past work [3], we have shown that given a sufficient number M of random measurements, one can ensure with high probability that a manifold $\mathcal{M} \subset \mathbb{R}^N$ has a stable embedding in the measurement space $\mathbb{R}^M$ under the operator Φ, such that pairwise Euclidean and geodesic distances are approximately preserved on its image $\Phi\mathcal{M}$. We will discuss this in more detail later in Section 3, but a key aspect is that the number of requisite measurements M is linearly proportional to the information level of the signal, i.e., the dimension K of the manifold. In that work, the number of measurements was also logarithmically dependent on the ambient dimension N, although this dependence was later removed in the asymptotic case in [12] using a different set of assumptions on the manifold.

The first contribution of this paper—presented in Section 3—is that we provide an improved lower bound on the number of random measurements to guarantee a stable embedding of a signal manifold. In particular, we make the same assumptions on the manifold as in our past work [3] but provide a measurement bound that is independent of the ambient dimension N. Our bound is non-asymptotic, and we provide explicit constants.

In order to do this, we use tools from the theory of empirical processes (namely, the idea of "generic chaining" [55]), which have recently been used to develop state-of-the-art RIP results for structured measurement matrices in CS [24, 37, 47, 49–52, 56]. More elementary arguments (e.g., involving simple concentration of measure inequalities) have previously been used in CS (see, e.g., [2]) for deriving RIP bounds for unstructured i.i.d. measurement matrices, and we also used such arguments in [3] to derive a manifold embedding guarantee. However, it appears that the stronger machinery of the empirical process approach is necessary to derive stronger bounds, both in RIP problems and in manifold embedding problems. A chaining argument was employed in [12], and in this paper we present a chaining argument that is suitable for studying the manifold embedding problem under our set of assumptions on the manifold. Because this chaining framework is fairly technical, we develop it entirely in the appendices so that the body of the paper will be as self-contained and expository as possible for someone seeking merely to understand the substance and context of our results.

As a very simple illustration of the embedding phenomenon, Figure 1(b) presents an experiment where just $M = 3$ compressive measurements are acquired from each point x_θ described in Figure 1(a). We let $N = 1024$ and construct a randomly generated $3 \times N$ matrix Φ whose entries are i.i.d. Gaussian random variables with zero mean and variance of $1/3$. Each point x_θ from the original manifold $\mathcal{M} \subset \mathbb{R}^{1024}$ maps to a unique point Φx_θ in $\mathbb{R}^3$; the manifold embeds in the low-dimensional measurement space. Given any $y = \Phi x_{\theta'}$ for θ' unknown, then, it is possible to infer the value θ' using only knowledge of the parametric model for $\mathcal{M}$ and the measurement operator Φ. Moreover, as the number M of compressive measurements increases, the manifold embedding becomes much more stable and remains highly self-avoiding.

Indeed, there is strong evidence that, as a consequence of this phenomenon, questions such as Q1 (signal recovery) and Q3 (parameter estimation) can be accurately solved using only compressive measurements of a signal x, and that these procedures are robust to noise and to deviations of the signal x away from the manifold $\mathcal{M}$ [14, 53, 60]. Additional theoretical and empirical justification has followed for the manifold learning [32] and multiclass recognition problems [14] described above. Consequently, many of the advantages of compressive measurements that are beneficial in sparsity-based CS (low-cost sensor design, reduced transmission requirements, reduced storage requirements, lack of need for advance knowledge of signal structure, simplified computation in the low-dimensional space $\mathbb{R}^M$, etc.) may also be enjoyed in settings where manifold models capture the concise signal structure. Moreover, the use of a manifold model can often capture the structure of a signal in many fewer degrees of freedom K than would be required in any sparse representation, and thus the measurement rate M can be greatly reduced compared to sparsity-based CS approaches.

In this paper, we will focus on questions Q1 (signal recovery) and Q3 (parameter estimation). *The second contribution of this paper—presented in Section 4—is that we establish theoretical bounds on the accuracy to which these questions may be answered.* To do this, we rely largely on the new analytical chaining framework described above. We consider both deterministic and probabilistic instance-optimal bounds, and we see strong similarities to analogous results that have been derived for sparsity-based CS. As with sparsity-based CS, we show for manifold-based CS that for any fixed Φ, uniform deterministic ℓ_2 recovery bounds for recovery of all x are necessarily poor. We then show that, as with sparsity-based CS, providing for any x a probabilistic bound that holds over most Φ is possible with the desired accuracy. We consider both noise-free and noisy measurement settings and compare our bounds with sparsity-based CS.

We feel that a third contribution of this paper comes in the form of the analytical tools we use to study the above problems. Our chaining argument allows us to study not only the embedding problem (as in [12]) but also Q1 and Q3. Moreover, in Appendix A, which we call the "Toolbox," we present a collection of implications of our assumption that the manifold has bounded *condition number* (see Section 2.2 for definition). This elementary property, also known as the *reach* of a manifold in the geometric measure theory literature [26, 43], has become somewhat popular in the analysis of manifold models for signal processing (e.g., see [3, 14, 15, 35, 43, 58, 62]). The seminal paper [43] contains a collection of implications of bounded condition number that have been used directly or indirectly in numerous works, including [3, 14, 15, 35, 58, 62]. We restate some of these implications in the Toolbox. Unfortunately, after very careful study we were unable to confirm for ourselves some of the original proofs appearing in [43]. Therefore, some of the statements and proofs in the Toolbox below differ slightly from their original counterparts in [43]. We hope that these results will provide a useful reference for the continued study of manifolds with bounded condition number.

1.5 Paper organization

Section 2 provides the necessary background on sparsity-based CS and on manifold models to place our work in the proper context. In Section 3, we state our improved bound regarding stable embeddings of manifolds. In Section 4, we then formalize our criteria for answering questions Q1 and Q3 in the context of manifold models. We first confront the task of deriving deterministic instance-optimal bounds in ℓ_2 and then consider probabilistic instance-optimal bounds in ℓ_2. We conclude in Section 5 with a final discussion. The Toolbox (Appendix A) establishes a collection of useful results in differential geometry that are frequently used throughout our technical proofs, which appear in the remaining appendices.

2 Background

2.1 Sparsity-Based Compressive Sensing

2.1.1 Sparse models

The concise modeling framework used in CS is *sparsity*. Consider a signal $x \in \mathbb{R}^N$ and suppose the $N \times N$ matrix $\Psi = [\psi_1 \ \psi_2 \ \cdots \ \psi_N]$ forms an orthonormal basis for $\mathbb{R}^N$. We say x is K-sparse in the basis Ψ if for $\alpha \in \mathbb{R}^N$ we can write $x = \Psi\alpha$, where $\|\alpha\|_0 = K < N$. (The ℓ_0-norm notation counts the number of nonzeros of the entries of α.) In a sparse representation, the actual information content of a signal is contained exclusively in the $K < N$ positions and values of its nonzero coefficients.

For those signals that are approximately sparse, we may measure their proximity to sparse signals as follows. We define $\alpha_K \in \mathbb{R}^N$ to be the vector containing only the largest K entries of α in magnitude, with the remaining entries set to zero. Similarly, we let $x_K = \Psi\alpha_K$. It is then common to measure the proximity to sparseness using either $\|\alpha - \alpha_K\|_1$ or $\|\alpha - \alpha_K\|$ (the latter of which equals $\|x - x_K\|$ because Ψ is orthonormal). Here and elsewhere in this paper, $\|\cdot\|$ stands for the ℓ_2 norm.

2.1.2 Stable embeddings of sparse signal families

CS uses the concept of sparsity to simplify the data acquisition process. Rather than designing a sensor to measure a signal $x \in \mathbb{R}^N$, for example, it often suffices to design a sensor that can measure a much shorter vector $y = \Phi x$, where Φ is a linear measurement operator represented as an $M \times N$ matrix, and typically $M \ll N$.

The measurement matrix Φ must have certain properties in order to be suitable for CS. One desirable property (which leads to the theoretical results we mention in Section 2.1.3) is known as the Restricted Isometry Property (RIP) [6–8]. We say a matrix Φ meets the *RIP of order K with respect to the basis Ψ* if for some $\delta_K > 0$,

$$(1 - \delta_K)\|\alpha\| \leq \|\Phi\Psi\alpha\| \leq (1 + \delta_K)\|\alpha\|$$

holds for all $\alpha \in \mathbb{R}^N$ with $\|\alpha\|_0 \leq K$. Intuitively, the RIP can be viewed as guaranteeing a *stable embedding* of the collection of K-sparse signals within the measurement space $\mathbb{R}^M$. In particular, supposing the RIP of order $2K$ is satisfied with respect to the basis Ψ, then for all pairs of K-sparse signals $x_1, x_2 \in \mathbb{R}^N$, we have

$$(1 - \delta_{2K})\|x_1 - x_2\| \leq \|\Phi x_1 - \Phi x_2\| \leq (1 + \delta_{2K})\|x_1 - x_2\|. \tag{1}$$

Although deterministic constructions of matrices meeting the RIP are still a work in progress, it is known that the RIP can often be met by choosing Φ randomly from an acceptable distribution. For example, let Ψ be a fixed orthonormal basis for $\mathbb{R}^N$ and suppose that

$$M \geq C_1 K \log(N/K) \tag{2}$$

for some constant C_1. Then supposing that the entries of the $M \times N$ matrix Φ are drawn as i.i.d. Gaussian random variables with mean 0 and variance $\frac{1}{M}$, it follows that with high probability Φ meets the RIP of order K with respect to the basis Ψ. Two aspects of this construction deserve special notice: first, the number M of measurements required is linearly proportional to the information level K (and logarithmic in the ambient dimension N), and second, neither the sparse basis Ψ nor the locations of the nonzero entries of α need be known when designing the measurement operator Φ. Other random distributions for Φ may also be used, all requiring approximately the same number of measurements [25, 37, 48].

2.1.3 Sparsity-based signal recovery and sketching

Although the sparse structure of a signal x need not be known when collecting measurements $y = \Phi x$, a hallmark of CS is the use of the sparse model in order to facilitate understanding from the compressive measurements. A variety of algorithms have been proposed to answer Q1 (signal recovery), where we seek to solve the apparently undercomplete set of M linear equations $y = \Phi x$ for N unknowns. The canonical method [5, 8, 18] is known as ℓ_1-*minimization* and is formulated as follows: first solve

$$\widehat{\alpha} = \operatorname{argmin}_{\alpha' \in \mathbb{R}^N} \|\alpha'\|_1 \text{ subject to } y = \Phi\Psi\alpha', \tag{3}$$

and then set $\widehat{x} = \Psi\widehat{\alpha}$. This recovery program can also be extended to account for measurement noise. The following bound is known.

Theorem 1. [9] *Suppose that Φ satisfies the RIP of order $2K$ with respect to Ψ and with constant $\delta_{2K} < \sqrt{2} - 1$. Let $x \in \mathbb{R}^N$, and suppose that $y = \Phi x + n$ where $\|n\| \leq \epsilon$. Then let*

$$\widehat{\alpha} = \arg \min_{\alpha' \in \mathbb{R}^N} \|\alpha'\|_1 \text{ subject to } \|y - \Phi\Psi\alpha'\| \leq \epsilon,$$

and set $\widehat{x} = \Psi\widehat{\alpha}$. Then

$$\|x - \widehat{x}\| = \|\alpha - \widehat{\alpha}\| \leq C_1 K^{-\frac{1}{2}} \|\alpha - \alpha_K\|_1 + C_2 \epsilon. \tag{4}$$

for constants C_1 and C_2.

This result is not unique to ℓ_1 minimization; similar bounds have been established for signal recovery using greedy iterative algorithms OMP [16], ROMP [42], and CoSAMP [41]. Bounds of this type are extremely encouraging for signal processing. From only M measurements, it is possible to recover x with quality that is comparable to its proximity to the nearest K-sparse signal, and if x itself is K-sparse and there is no measurement noise, then x can be recovered exactly. Moreover, despite the apparent ill-conditioning of the inverse problem, the measurement noise is not dramatically amplified in the recovery process.

These bounds are known as *deterministic, instance-optimal* bounds because they hold deterministically for any Φ that meets the RIP, and because for a given Φ they give a guarantee for recovery of any $x \in \mathbb{R}^N$ based on its proximity to the concise model.

The use of ℓ_1 as a measure for proximity to the concise model (on the right hand side of (4)) arises due to the difficulty in establishing ℓ_2 bounds on the right hand side. Indeed, it is known that deterministic ℓ_2 instance-optimal bounds cannot exist that are comparable to (4). In particular, for any Φ, to ensure that $\|x - \widehat{x}\| \leq C_3 \|x - x_K\|$ for all x, it is known [13] that this requires that $M \geq C_4 N$ regardless of K.

However, it is possible to obtain an instance-optimal ℓ_2 bound for sparse signal recovery in the noise-free setting by changing from a deterministic formulation to a *probabilistic* one [13, 17]. In particular, by considering any given $x \in \mathbb{R}^N$, it is possible to show that for *most* random Φ, letting the measurements $y = \Phi x$, and recovering $\widehat{x}$ via ℓ_1-minimization (3), it holds that

$$\|x - \widehat{x}\| \leq C_5 \|x - x_K\|. \tag{5}$$

While the proof of this statement [17] does not involve the RIP directly, it holds for many of the same random distributions that work for RIP matrices, and it requires the same number of measurements (2) up to a constant.

Similar bounds hold for the closely related problem of Q2 (sketching), where the goal is to use the compressive measurement vector y to identify and report only approximately K expansion coefficients that best describe the original signal, i.e., a sparse approximation to α_K. In the case where $\Psi = I$, an efficient randomized measurement process coupled with a customized recovery algorithm [29] provides signal sketches that meet a deterministic mixed-norm ℓ_2/ℓ_1 instance-optimal bound analogous to (4) in the noise-free setting. A desirable aspect of this construction is that the computational complexity scales with only $\log(N)$ (and is polynomial in K); this is possible because only approximately K pieces of information must be computed to describe the signal. Though at a higher computational cost, the aforementioned greedy algorithms (such as CoSaMP) for signal recovery can also be interpreted as sketching techniques in that they produce explicit sparse approximations to α_K. Finally, for signals that are sparse in the Fourier domain (Ψ consists of the DFT vectors), probabilistic ℓ_2/ℓ_2 instance-optimal bounds have been established for a specialized sketching algorithm [27, 28] that are analogous to (5).

2.2 Manifold models and properties

As we have discussed in Section 1.4, there are many possible modeling frameworks for capturing concise signal structure. Among these possibilities are the broad class of manifold models.

Manifold models arise, for example, in settings where the signals of interest vary continuously as a function of some K-dimensional parameter. Suppose, for instance, that there exists some parameter θ that controls the generation of the signal. We let $x_\theta \in \mathbb{R}^N$ denote the signal corresponding to the parameter θ, and we let Θ denote the K-dimensional parameter space from which θ is drawn. In general, Θ itself may be a K-dimensional manifold and need not be embedded in an ambient Euclidean space. For example, supposing θ describes the 1-D rotation parameter in a top-down satellite image, we have $\Theta = \mathbb{S}^1$.

Under certain conditions on the parameterization $\theta \mapsto x_\theta$, it follows that $\mathcal{M} := \{x_\theta : \theta \in \Theta\}$ forms a K-dimensional *submanifold* of $\mathbb{R}^N$. An appropriate visualization is that the set $\mathcal{M}$ forms a nonlinear K-dimensional "surface" within the high-dimensional ambient signal space $\mathbb{R}^N$. Depending on the circumstances, we may measure the distance between points two points x_{θ_1} and x_{θ_2} on the manifold $\mathcal{M}$ using either the ambient Euclidean distance $\|x_{\theta_1} - x_{\theta_2}\|$ or the geodesic distance along the manifold, which we denote as $d_\mathcal{M}(x_{\theta_1}, x_{\theta_2})$. In the case where the geodesic distance along $\mathcal{M}$ equals the native distance in parameter space, i.e., when

$$d_\mathcal{M}(x_{\theta_1}, x_{\theta_2}) = d_\Theta(\theta_1, \theta_2), \tag{6}$$

we say that $\mathcal{M}$ is *isometric* to Θ. The definition of the distance $d_\Theta(\theta_1, \theta_2)$ depends on the appropriate metric for the parameter space Θ; supposing Θ is a convex subset of Euclidean space, then we can let $d_\Theta(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|$.

While our discussion above concentrates on the case of manifolds $\mathcal{M}$ generated by underlying parameterizations, we stress that manifolds have also been proposed as approximate low-dimensional models within $\mathbb{R}^N$ for nonparametric signal classes such as images of human faces or handwritten digits [4, 33, 57]. These signal families may also be considered.

The results we present in this paper will make reference to certain characteristic properties of the manifold under study. These terms are originally defined in [3, 43] and are repeated here for completeness. First, our results will depend on a measure of regularity for the manifold. For this purpose, we adopt the notion of the *condition number* of a manifold, which is also known as the reach of a manifold in the geometric measure theory literature [26, 43].

Definition 1. [43] *Let $\mathcal{M}$ be a compact Riemannian submanifold of $\mathbb{R}^N$. The* condition number *is defined as $1/\tau$, where τ is the largest number having the following property: The open normal bundle about $\mathcal{M}$ of radius r is embedded in $\mathbb{R}^N$ for all $r < \tau$.*

The condition number $1/\tau$ controls both local properties and global properties of the manifold. Its role is summarized in two key relationships (see the Toolbox and [43] for more detail). First, the the curvature of any unit-speed geodesic path on $\mathcal{M}$ is bounded by $1/\tau$. Second, at long geodesic distances, the condition number controls how close the manifold may curve back upon itself. For example, supposing $x_1, x_2 \in \mathcal{M}$ with $d_\mathcal{M}(x_1, x_2) > \tau$, it must hold that $\|x_1 - x_2\| > \tau/2$.

We conclude with a brief but concrete example to illustrate specific values for these quantities. Let $N > 0$, $\kappa > 0$, $\Theta = \mathbb{R} \bmod 2\pi$, and suppose $x_\theta \in \mathbb{R}^N$ is given by

$$x_\theta = [\kappa\cos(\theta); \ \kappa\sin(\theta); \ 0; \ 0; \cdots 0]^T.$$

In this case, $\mathcal{M} = \{x_\theta : \theta \in \Theta\}$ forms a circle of radius κ in the $x(1), x(2)$ plane. The manifold dimension $K = 1$, and the condition number $1/\tau = 1/\kappa$. We also refer in our results to the K-dimensional volume of the $\mathcal{M}$, denoted by $V_\mathcal{M}$, which in this example corresponds to the circumference $2\pi\kappa$ of the circle.

3 Stable embeddings of manifolds

In cases where the signal class of interest $\mathcal{M}$ forms a low-dimensional submanifold of $\mathbb{R}^N$, we have theoretical justification that the information necessary to distinguish and recover signals $x \in \mathcal{M}$ can be well-preserved under a sufficient number of compressive measurements $y = \Phi x$. In particular, it was first shown in [3] that an RIP-like property holds for families of manifold-modeled signals. The result stated that, under a random projection onto $\mathbb{R}^M$, pairwise distances between the points on $\mathcal{M}$ are approximately preserved with high probability, provided that M, number of measurements, is large enough. Mainly, M should scale linearly with the dimension K of $\mathcal{M}$ and logarithmically with the ambient dimension N. The dependence on N was later removed in [12], which used a different set of assumptions on the manifold to help derive a sharper lower bound on the requisite number of random measurements. Unfortunately, the results given in [12] hold only as the isometry constant $\epsilon \to 0$ in (1), with asymptotic threshold and constants fixed but unspecified. The manifold properties assumed in [12] are arguably more complicated and less commonly used than the manifold volume and condition number which are at the heart of our results. (On the other hand, there do exist manifolds where the properties assumed in [12] clearly provide a stronger analysis.)

In this section, we establish an improved lower bound on M to ensure a stable embedding of a manifold. We make the same assumptions on the manifold as in our past work [3] but provide a measurement bound that is independent of the ambient dimension N. Our bound holds for every $\epsilon \leq 1/3$ and we provide explicit constants. The proof, presented in Appendix B, draws from the ideas in generic chaining [55], which have been recently used to develop state-of-the-art RIP results for structured measurement matrices in CS [24, 37, 47, 49–52, 56]. As in [12], we control the failure probability of the manifold embedding by forming a so-called *chain* on a sequence of increasingly finer covers on the index set of the random process [39, 55]. Aside from delivering an improved bound (and also allowing us to study Q1 and Q3 in Section 4), we hope that our exposition in this paper will encourage yet more researchers in the field of CS to use this powerful technique.

Theorem 2. *Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having condition number $1/\tau$ and volume $V_{\mathcal{M}}$. Conveniently assume that[2]*

$$\frac{V_{\mathcal{M}}}{\tau^K} \geq \left(\frac{21}{2\sqrt{K}} \right)^K . \tag{7}$$

Fix $0 < \epsilon \leq 1/3$ and $0 < \rho < 1$. Let Φ be a random $M \times N$ matrix populated with i.i.d. zero-mean Gaussian random variables with variance of $1/M$ with

$$M \geq 18\epsilon^{-2} \max \left(24K + 2K \log \left(\frac{\sqrt{K}}{\tau\epsilon^2} \right) + \log(2V_{\mathcal{M}}^2), \log \left(\frac{8}{\rho} \right) \right) . \tag{8}$$

Then with probability at least $1 - \rho$ the following statement holds: For every pair of points $x_1, x_2 \in \mathcal{M}$,

$$(1 - \epsilon) \|x_1 - x_2\| \leq \|\Phi x_1 - \Phi x_2\| \leq (1 + \epsilon) \|x_1 - x_2\| . \tag{9}$$

The proof of the above result can be found in Appendix B. In essence, manifolds with higher volume or with greater curvature have more complexity, which leads to an increased number of measurements (8). By comparing (1) with (9), we see a strong analogy to the RIP of order $2K$.

---

[2]Theorem 2 still holds, with worse constants, after relaxing this assumption.

This theorem establishes that, like the class of K-sparse signals, a collection of signals described by a K-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$ can have a stable embedding in an M-dimensional measurement space. Moreover, the requisite number of random measurements M is once again almost linearly proportional to the information level (or number of degrees of freedom) K. It is important to note that in (8), the combined dependence on the manifold dimension K, condition number $1/\tau$, and volume $V_\mathcal{M}$ cannot, generally speaking, be improved. In particular, according to (64), when $\mathcal{M}$ is a K-dimensional unit ball in $\mathbb{R}^N$, $V_\mathcal{M} = V_{\mathcal{B}_K}$ scales like $K^{-K/2}$ and the lower bound on M in (8) scales with K (since the $\log K$ terms cancel), which is clearly optimal.

As was the case with the RIP for sparse signal processing, this sort of result has a number of possible implications for manifold-based signal processing. First, individual signals obeying a manifold model can be acquired and stored efficiently using compressive measurements, and it is unnecessary to employ the manifold model itself as part of the compression process. Rather, the model needs only to be used for signal understanding from the compressive measurements. Second, problems such as Q1 (signal recovery) and Q3 (parameter estimation) can be addressed, as we discuss in Section 4. Aside from this theoretical analysis, we have reported promising experimental recovery/estimation results with various classes of parametric signals [14, 60]. Also, taking a different analytical perspective (a statistical one, assuming additive white Gaussian measurement noise), estimation-theoretic quantities such as the Cramer-Rao Lower Bound (for a specialized set of parametric problems) have been shown to be preserved in the compressive measurement space as a consequence of the stable embedding [46]. Third, the stable embedding results can also be extended to the case of multiple manifolds that are simultaneously embedded [14]; this allows both the classification of an observed object to one of several possible models (different manifolds) and the estimation of a parameter within that class (position on a manifold). Fourth, collections of signals obeying a manifold model (such as multiple images of a scene photographed from different perspectives) can be acquired using compressive measurements, and the resulting manifold structure will be preserved among the suite of measurement vectors in $\mathbb{R}^M$ [15, 45]. Fifth, we have provided empirical and theoretical support for the use of manifold learning in the reduced-dimensional space [32]; this can dramatically simplify the computational and storage demands on a system for processing large databases of signals.

4 Manifold-based signal recovery and parameter estimation

In this section, we establish theoretical bounds on the accuracy to which problems Q1 (signal recovery) and Q3 (parameter estimation) can be solved under a manifold model. To be specific, let us consider a length-N signal x that is not necessarily K-sparse, but rather that we assume lives on or near some known K-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$. From a collection of measurements

$$y = \Phi x + n,$$

where Φ is a random $M \times N$ matrix and $n \in \mathbb{R}^M$ is an additive noise vector, we would like to recover either x or a parameter θ that generates x.

For the signal recovery problem, we will consider the following as a method for estimating x:

$$\widehat{x} = \arg \min_{x' \in \mathcal{M}} \left\| y - \Phi x' \right\|, \tag{10}$$

supposing here and elsewhere that the minimum is uniquely defined. We also let x^* be the optimal "nearest neighbor" to x on $\mathcal{M}$, i.e.,

$$x^* = \arg \min_{x' \in \mathcal{M}} \left\| x - x' \right\|. \tag{11}$$

To consider signal recovery successful, we would like to guarantee that $\|x - \widehat{x}\|$ is not much larger than $\|x - x^*\|$.

For the parameter estimation problem, where we presume $x \approx x_\theta$ for some $\theta \in \Theta$, we propose a similar method for estimating θ from the compressive measurements:

$$\widehat{\theta} = \arg \min_{\theta' \in \Theta} \|y - \Phi x_{\theta'}\|. \tag{12}$$

Let θ^* be the "optimal estimate" that could be obtained using the full data $x \in \mathbb{R}^N$, i.e.,

$$\theta^* = \arg \min_{\theta' \in \Theta} \|x - x_{\theta'}\|. \tag{13}$$

(If $x = x_\theta$ exactly for some θ, then $\theta^* = \theta$; otherwise this formulation allows us to consider signals x that are not precisely on the manifold $\mathcal{M}$ in $\mathbb{R}^N$. This generalization has practical relevance; a local image block, for example, may only approximately resemble a straight edge, which has a simple parameterization.) To consider parameter estimation successful, we would like to guarantee that $d_\Theta(\widehat{\theta}, \theta^*)$ is small.

As we will see, bounds pertaining to accurate signal recovery can often be extended to imply accurate parameter estimation as well. However, the relationships between distance d_Θ in parameter space and distances $d_\mathcal{M}$ and $\| \cdot \|$ in the signal space can vary depending on the parametric signal model under study. Thus, for the parameter estimation problem, our ability to provide generic bounds on $d_\Theta(\widehat{\theta}, \theta^*)$ will be restricted. In this paper we focus primarily on the signal recovery problem and provide preliminary results for the parameter estimation problem that pertain most strongly to the case of isometric parameterizations.

In this paper, we do not confront in depth the question of how a recovery program such as (10) can be efficiently solved. Some efforts in this direction have recently appeared in [31,35,53]. In [53], the authors guarantee the success of a gradient-projection algorithm for recovering a signal that lives exactly on the manifold from noisy compressive measurements. The keys to the success of this method are a stable embedding of the manifold (as is guaranteed by [3] or our Theorem 2) and the knowledge of the projection operator onto the manifold within $\mathbb{R}^N$. In [35], the authors construct a geometric multi-resolution approximation of a manifold using a collection of affine subspaces. A major contribution of that work is a recovery algorithm that works by assigning a measured signal to the closest projected affine subspace in the compressed domain. Two recovery results are presented. In the first of these, the number of measurements is independent of the ambient dimension and the recovery error holds for any given signal in the ambient space. All of this is analogous to our Theorem 4 (a probabilistic instance-optimal bound in ℓ_2), but the recovery is guaranteed for a particular algorithm. Unlike that result, however, our Theorem 4 includes explicit constants, allows for the consideration of measurement noise, and falls nearly for free out of our novel analytical framework based on chaining. A second result appearing in [35] provides a special type of deterministic instance-optimal bound for signal recovery and involves embedding arguments that extend those in [3]. It would be interesting to see if our improved embedding guarantees in the present paper could now be used to remove the dependence on the ambient dimension appearing in that result. In [11], the authors provide a Bayesian treatment of the signal recovery problem using a mixture of (low-rank) Gaussians for approximating the manifold. Furthermore, some discussion of signal recovery is provided in [3], with application-specific examples provided in [14,60]. Unfortunately, it is difficult to propose a single general-purpose algorithm for solving (10) in $\mathbb{R}^M$, as even the problem (11) in $\mathbb{R}^N$ may be difficult to solve depending on certain nuances (such as topology) of the individual manifold. Additional complications arise when the manifold $\mathcal{M}$ is non-differentiable, as may happen when the signals x represent 2-D images. However, just as

a multiscale regularization can be incorporated into Newton's method for solving (11) (see [61]), an analogous regularization can be incorporated into a compressive measurement operator Φ to facilitate Newton's method for solving (10) (see [19,21,60]). For manifolds that lack differentiability, additional care must be taken when applying results such as Theorem 2. We therefore expect that the research on signal recovery and approximation based on low-dimensional manifold models will witness even more growth in the future.

It is also crucial to study the theoretical limits and guarantees in this problem; in what follows, we will consider both deterministic and probabilistic instance-optimal bounds for signal recovery and parameter estimation, and we will draw comparisons to the sparsity-based CS results of Section 2.1.3. Our bounds are formulated in terms of generic properties of the manifold (as mentioned in Section 2.2), which will vary from signal model to signal model. In some cases, calculating these may be possible, whereas in other cases it may not. Nonetheless, we feel the results in this paper highlight the relative importance of these properties in determining the requisite number of measurements.

4.1   A deterministic instance-optimal bound in ℓ_2

We begin by seeking an instance-optimal bound. That is, for a measurement matrix Φ that meets (9) for all $x_1, x_2 \in \mathcal{M}$, we seek an upper bound for the relative reconstruction error

$$\frac{\|x - \widehat{x}\|}{\|x - x^*\|}$$

that holds uniformly for all $x \in \mathbb{R}^N$. In this section we consider only the signal recovery problem; however, similar bounds would apply to parameter estimation. We have the following result for the noise-free case, the proof of which can be found in Appendix H.

Theorem 3. *Fix $0 < \epsilon \leq 1/3$ and $0 < \rho < 1$. Let $\mathcal{M}$ be as described in Theorem 2. Assume that Φ satisfies (9) for all pairs of points $x_1, x_2 \in \mathcal{M}$. Take $x \in \mathbb{R}^N$, let $y = \Phi x$, and let the recovered estimate $\widehat{x}$ and the optimal estimate x^* be as defined in (10) and (11). Then the following holds:*

$$\frac{\|x - \widehat{x}\|}{\|x - x^*\|} \leq \frac{1}{1 - \epsilon} \left(2\sigma_M(\Phi) + 1\right) \tag{14}$$

where $\sigma_M(\Phi)$ is the largest singular value of Φ.

In particular, it is interesting to consider the case where Φ is a random Gaussian matrix as described in Theorem 2. It is well-known, e.g., [59, Corollary 5.35], that the nonzero singular values of Φ satisfy the following:

$$P \left\{ \sigma_M(\Phi) > \sqrt{\frac{N}{M}} + 1 + t \right\} \leq e^{-t^2 M/2}, \tag{15}$$

$$P \left\{ \sigma_m(\Phi) < \sqrt{\frac{N}{M}} - 1 - t \right\} \leq e^{-t^2 M/2}, \tag{16}$$

for $t > 0$. Here, $\sigma_M(\Phi)$ and $\sigma_m(\Phi)$ are the largest and smallest (nonzero) singular values of Φ, respectively. Suppose that M satisfies (8) so that the promises of Theorem 2 hold except for a probability of at most ρ. Set $t = 1$ in (15). Now since $e^{-M/2} \leq \rho$, we have that

$$\sigma_M(\Phi) \leq \sqrt{\frac{N}{M}} + 2,$$

except with a probability of at most ρ. In combination with Theorem 3, it finally follows that, except with a probability of at most 2ρ, Φ satisfies (9) for every pair of points on the manifold and that

$$\frac{\|x - \widehat{x}\|}{\|x - x^*\|} \leq \frac{1}{1-\epsilon}\left(2\sqrt{\frac{N}{M}} + 5\right), \tag{17}$$

for every $x \in \mathbb{R}^N$. Here, $\widehat{x}$ and x^* are as defined in (10) and (11). As $\frac{M}{N} \to 0$, the bound on the right hand side of (17) grows as $\frac{2}{1-\epsilon}\sqrt{\frac{N}{M}}$. Unfortunately, this is not desirable for signal recovery. Supposing, for example, that we wish to ensure $\|x - \widehat{x}\| \leq C_6 \|x - x^*\|$ for all $x \in \mathbb{R}^N$, then using the bound (17) we would require that $M \geq C_7 N$ regardless of the dimension K of the manifold.

The weakness of this bound is a geometric necessity; indeed, the bound itself is quite tight in general, as the following simple example illustrates. The proof can be found in Appendix I.

Proposition 1. *Fix $0 < \epsilon \leq 1/3$. Let $\mathcal{M}$ denote the line segment in $\mathbb{R}^N$ joining the origin and $e_1 := [1, 0, 0, \dots, 0]^T$. Suppose that Φ satisfies (9) for all $x_1, x_2 \in \mathcal{M}$ and that $\sigma_m(\Phi) \geq 8/3$. Then, there exists a point $x \in \mathbb{R}^N$ such that*

$$\frac{\|x - \widehat{x}\|}{\|x - x^*\|} \geq \frac{1}{2(1+\epsilon)}\sigma_m(\Phi),$$

where $\widehat{x}$ and x^* are defined in (10) and (11).

In particular, consider the case where Φ is a random Gaussian matrix as described in Theorem 2. According to (45) and (16) (with $t = 1$), the following two statements are valid except with a probability of at most $2e^{-M\epsilon^2/6} + 2e^{-M/2} \leq 4e^{-M\epsilon^2/6}$. First, (9) holds for every $x_1, x_2 \in \mathcal{M}$. Second,

$$\sigma_m(\Phi) \geq \sqrt{\frac{N}{M}} - 2. \tag{18}$$

If we assume that $N/M \geq (14/3)^2$, we can conclude, using Proposition 1, that Φ satisfies (9) for every $x_1, x_2 \in \mathcal{M}$ and yet there exists $x \in \mathbb{R}^N$ such that

$$\frac{\|x - \widehat{x}\|}{\|x - x^*\|} \geq \frac{1}{4(1+\epsilon)}\sqrt{\frac{N}{M}},$$

except with a probability of at most $4e^{-M\epsilon^2/6}$ on the choice of Φ.

It is worth recalling that, as we discussed in Section 2.1.3, similar difficulties arise in sparsity-based CS when attempting to establish a deterministic ℓ_2 instance-optimal bound. In particular, to ensure that $\|x - \widehat{x}\| \leq C_3 \|x - x_K\|$ for all $x \in \mathbb{R}^N$, it is known [13] that this requires $M \geq C_4 N$ regardless of the sparsity level K.

In sparsity-based CS, there have been at least two types of alternative approaches. The first are the deterministic "mixed-norm" results of the type given in (4). These involve the use of an alternative norm such as the ℓ_1 norm to measure the distance from the coefficient vector α to its best K-term approximation α_K. While it may be possible to pursue similar directions for manifold-modeled signals, we feel this is undesirable as a general approach because when sparsity is no longer part of the modeling framework, the ℓ_1 norm has less of a natural meaning. Instead, we prefer to seek bounds using ℓ_2, as that is the most conventional norm used in signal processing to measure energy and error.

Thus, the second type of alternative bounds in sparsity-based CS have involved ℓ_2 bounds in probability, as we discussed in Section 2.1.3. Indeed, the performance of both sparsity-based and

manifold-based CS is often much better in practice than a deterministic ℓ_2 instance-optimal bound might indicate. The reason is that, for any Φ, such bounds consider the *worst case* signal over all possible $x \in \mathbb{R}^N$. Fortunately, this worst case is not typical. As a result, it is possible to derive much stronger results that consider any given signal $x \in \mathbb{R}^N$ and establish that for most random Φ, the recovery error of that signal x will be small.

4.2 Probabilistic instance-optimal bounds in ℓ_2

For a given measurement operator Φ, our bound in Theorem 3 applies uniformly to any signal in $\mathbb{R}^N$. However, a much sharper bound can be obtained by relaxing the deterministic requirement.

4.2.1 Signal recovery

Our first bound applies to the signal recovery problem, and we include the consideration of additive noise in the measurements. This result is proved in Appendix J and, like that of Theorem 2, involves a generic chaining argument.

Theorem 4. *Suppose $x \in \mathbb{R}^N$. Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having condition number $1/\tau$ and volume $V_\mathcal{M}$. Conveniently assume that*[3]

$$\frac{V_\mathcal{M}}{\tau^K} \geq \left(\frac{21}{\sqrt{K}} \right)^K .$$

Fix $0 < \epsilon \leq 1/3$ and $0 < \rho < 1$. Let Φ be an $M \times N$ random matrix populated with i.i.d. zero-mean Gaussian random variables with variance $1/M$, chosen independently of x, with

$$M \geq 18\epsilon^{-2} \max \left(11K + K \log \left(\frac{\sqrt{K}}{\tau\epsilon^2} + \right) + \log V_\mathcal{M} \, , \, \log \left(\frac{16}{\rho} \right) \right). \tag{19}$$

Let $n \in \mathbb{R}^M$, let $y = \Phi x + n$, and let the recovered estimate $\widehat{x}$ and the optimal estimate x^* be as defined in (10) and (11). Then with a probability of at least $1 - \rho$, the following statement holds:

$$\|x - \widehat{x}\| \leq (1 + 2\epsilon) \|x - x^*\| + (2 + \epsilon) \|n\| + \frac{\epsilon\tau}{40}. \tag{20}$$

Let us now compare and contrast our bound with the analogous results for sparsity-based CS. Like Theorem 1, we consider the problem of signal recovery in the presence of additive measurement noise. Both bounds relate the recovery error $\|x - \widehat{x}\|$ to the proximity of x to its nearest neighbor in the concise model class (either x_K or x^* depending on the model), and both bounds relate the recovery error $\|x - \widehat{x}\|$ to the amount $\|n\|$ of additive measurement noise. However, Theorem 1 is a deterministic bound whereas Theorem 4 is probabilistic, and our bound (20) measures proximity to the concise model in the ℓ_2 norm, whereas (4) uses the ℓ_1 norm.

Our bound can also be compared with (5), as both are instance-optimal bounds in probability, and both use the ℓ_2 norm to measure proximity to the concise model. However, we note that unlike (5), our bound (20) allows the consideration of measurement noise.

Finally, we note that there is an additional term $\frac{\epsilon\tau}{40}$ appearing on the right hand side of (20). This term becomes relevant only when both $\|x - x^*\|$ and $\|n\|$ are significantly smaller than the

---

[3]Theorem 4 still holds, with worse constants, after relaxing this assumption.

condition number τ, since $\epsilon \leq 1/3$. Indeed, in these regimes the signal recovery remains accurate (much smaller than τ), but the quantity $\|x - \widehat{x}\|$ may not remain strictly proportional to $\|x - x^*\|$ and $\|n\|$. We emphasize that this bound may also be sharpened by artificially assuming a condition number $1/\tau' > 1/\tau$ for the purpose of choosing a number of measurements M in (19). This will decrease the last term in (20) as $\frac{\epsilon \tau'}{40}$. In the case where $n = 0$, it is also possible to resort to the bound (17); this bound is inferior to (20) when $\|x - x^*\|$ is large but ensures that $\|x - \widehat{x}\| \to 0$ when $\|x - x^*\| \to 0$.

4.2.2   Parameter estimation

Above we have derived a bound for the signal recovery problem, with an error metric that measures the discrepancy between the recovered signal $\widehat{x}$ and the original signal x.

However, in some applications it may be the case that the original signal $x \approx x_{\theta^*}$, where $\theta^* \in \Theta$ is a parameter of interest. In this case we may be interested in using the compressive measurements $y = \Phi x + n$ to solve the problem (12) and recover an estimate $\widehat{\theta}$ of the underlying parameter.

Of course, these two problems are closely related. However, we should emphasize that guaranteeing $\|x - \widehat{x}\| \approx \|x - x^*\|$ does not automatically guarantee that $d_{\mathcal{M}}(x_{\widehat{\theta}}, x_{\theta^*})$ is small (and therefore does not ensure that $d_{\Theta}(\widehat{\theta}, \theta^*)$ is small). If the manifold is shaped like a horseshoe, for example, then it could be the case that x_{θ^*} sits at the end of one arm but $x_{\widehat{\theta}}$ sits at the end of the opposing arm. These two points would be much closer in a Euclidean metric than in a geodesic one.

Consequently, in order to establish bounds relevant for parameter estimation, our concern focuses on guaranteeing that the geodesic distance $d_{\mathcal{M}}(x_{\widehat{\theta}}, x_{\theta^*})$ is itself small. Our next result is proved in Appendix L.

Theorem 5. *Suppose $x \in \mathbb{R}^N$. Let $\mathcal{M}$ and Φ be as described in Theorem 2. Fix $0 < \epsilon \leq 1/3$ and $0 < \rho < 1$. Assume that M satisfies (19) and that the convenient assumption (4) holds. Let $n \in \mathbb{R}^M$, let $y = \Phi x + n$, and let the recovered estimate $\widehat{x}$ and the optimal estimate x^* be as defined in (10) and (11). If $\|x - x^*\| + \frac{7}{8}\|n\| \leq 0.184\tau$, then with probability at least $1 - \rho$ the following statement holds:*

$$d_{\mathcal{M}}(\widehat{x}, x^*) \leq (4 + 4\epsilon)\|x - x^*\| + (4 + 2\epsilon)\|n\| + \frac{\epsilon\tau}{20}. \qquad (21)$$

In several ways, this bound is similar to (20). Both bounds relate the recovery error to the proximity of x to its nearest neighbor x^* on the manifold and to the amount $\|n\|$ of additive measurement noise. Both bounds also have an additive term on the right hand side that is small in relation to the condition number τ.

In contrast, (21) guarantees that the recovered estimate $\widehat{x}$ is near to the optimal estimate x^* in terms of geodesic distance along the manifold. Establishing this condition required the additional assumption that $\|x - x^*\| + \frac{7}{8}\|n\| \leq 0.184\tau$. Because τ relates to the degree to which the manifold can curve back upon itself at long geodesic distances, this assumption prevents exactly the type of "horseshoe" problem that was mentioned above, where it may happen that $d_{\mathcal{M}}(\widehat{x}, x^*) \gg \|\widehat{x} - x^*\|$. Suppose, for example, it were to happen that $\|x - x^*\| \approx \tau$ and x was approximately equidistant from both ends of the horseshoe; a small distortion of distances under Φ could then lead to an estimate $\widehat{x}$ for which $\|x - \widehat{x}\| \approx \|x - x^*\|$ but $d_{\mathcal{M}}(\widehat{x}, x^*) \gg 0$. Similarly, additive noise could cause a similar problem of "crossing over" in the measurement space. Although our bound provides no guarantee in these situations, we stress that under these circumstances, accurate parameter estimation would be difficult (or perhaps even unimportant) in the original signal space $\mathbb{R}^N$.

Finally, we revisit the situation where the original signal $x \approx x_{\theta^*}$ for some $\theta^* \in \Theta$ (with θ^* satisfying (13)), where the measurements $y = \Phi x + n$, and where the recovered estimate $\widehat{\theta}$ satisfies (12). We consider the question of whether (21) can be translated into a bound on $d_\Theta(\widehat{\theta}, \theta^*)$. As described in Section 2.2, in signal models where $\mathcal{M}$ is isometric to Θ, this is automatic: we have simply that

$$d_\mathcal{M}(x_{\widehat{\theta}}, x_{\theta^*}) = d_\Theta(\widehat{\theta}, \theta^*).$$

Such signal models are not nonexistent. Work by Donoho and Grimes [20], for example, has characterized a variety of articulated image classes for which (6) holds or for which $d_\mathcal{M}(x_{\theta_1}, x_{\theta_2}) = C_8 d_\Theta(\theta_1, \theta_2)$ for some constant $C_8 > 0$. In other models it may hold that

$$C_9 d_\mathcal{M}(x_{\theta_1}, x_{\theta_2}) \leq d_\Theta(\theta_1, \theta_2) \leq C_{10} d_\mathcal{M}(x_{\theta_1}, x_{\theta_2})$$

for constants $C_9, C_{10} > 0$. Each of these relationships may be incorporated to the bound (21).

5 Conclusions

In this paper, we have provided an improved and non-asymptotic lower bound on the number of requisite measurements to ensure a stable embedding of a manifold under a random linear measurement operator. We have also considered the tasks of signal recovery and parameter estimation using compressive measurements of a manifold-modeled signal, and we have established theoretical bounds on the accuracy to which these questions may be answered. Although these problems differ substantially from the mechanics of sparsity-based CS, we have seen a number of similarities that arise due to the low-dimensional geometry of the each of the concise models. First, we have seen that a sufficient number of compressive measurements can guarantee a stable embedding of either type of signal family, and the requisite number of measurements scales essentially linearly with the information level of the signal. Second, we have seen that deterministic instance-optimal bounds in ℓ_2 are necessarily weak for both problems. Third, we have seen that probabilistic instance-optimal bounds in ℓ_2 can be derived that give the optimal scaling with respect to the signal proximity to the concise model and with respect to the amount of measurement noise. Thus, our work supports the growing evidence that manifold-based models can be used with high accuracy in compressive signal processing.

Most of our analysis in this paper rests on a new analytical framework for studying manifold embeddings that uses tools from the theory of empirical processes (namely, the idea of generic chaining). While such tools are becoming more widely adopted in the analysis of sparsity-based CS problems, we believe they are also very promising for studying the interactions of nonlinear signal families (such as manifolds) with random, compressive measurement operators. We hope that the chaining argument in this paper will be useful for future investigations along these lines.

Acknowledgements

A Toolbox

We begin by introducing some notation that will be used throughout the rest of the appendices.

In this paper, $\mathbb{N}$ stands for the set of nonnegative integers. The tangent space of $\mathcal{M}$ at $p \in \mathcal{M}$ is denoted $\mathcal{T}_p$. The orthogonal projection operator onto this linear subspace is denoted by $\downarrow_p$. We let $\angle[\cdot, \cdot]$ represent the angle between two vectors after being shifted to the same starting point. Throughout this paper, $d_{\mathcal{M}}(\cdot, \cdot)$ measures the geodesic distance between two points on $\mathcal{M}$. By r-*ball* we refer to a Euclidean (open) ball of radius $r > 0$. In addition, with $\mathcal{B}_N$ we denote the unit ball in $\mathbb{R}^N$ with volume $V_{\mathcal{B}_N}$ and we reserve $\mathcal{B}_N(p, r)$ to represent an N-dimensional r-ball centered at p in $\mathbb{R}^N$. For $r > 0$, let $\mathcal{A}_{\mathcal{M}}(p, r) := \mathcal{M} \cap \mathcal{B}_N(p, r) - p$ denote a (relatively) open neighborhood of p on $\mathcal{M}$ after being shifted to the origin. Here the subtraction is in the Minkowski's sense. A K-dimensional r-ball centered at the origin and along $\mathcal{T}_p$ will be denoted by $\mathcal{B}_{\mathcal{T}_p}(r)$. Unless otherwise stated, all distances are measured in the Euclidean metric.

A collection of N-dimensional r-balls that covers $\mathcal{M}$ is called an r-*cover* for $\mathcal{M}$, with their centers forming an r-*net* for $\mathcal{M}$. Notice that in general we do not require a net for $\mathcal{M}$ to be a subset of $\mathcal{M}$. However, we define the *covering number* of $\mathcal{M}$ at scale r, $\mathcal{N}_{\mathcal{M}}(r)$, to be the cardinality of a minimal r-net for $\mathcal{M}$ among all subsets of $\mathcal{M}$. (In other words, $\mathcal{N}_{\mathcal{M}}(r)$ is the smallest number of r-balls centered on $\mathcal{M}$ that it takes to cover $\mathcal{M}$.) A maximal r-separated subset of $\mathcal{M}$ is called an r-*packing* for $\mathcal{M}$. The *packing number* of $\mathcal{M}$ at scale $r > 0$, denoted by $\mathcal{P}_{\mathcal{M}}(r)$, is the cardinality of such a set. It can be easily verified that an r-packing for $\mathcal{M}$ is also an r-cover for $\mathcal{M}$, so

$$\mathcal{N}_{\mathcal{M}}(r) \leq \mathcal{P}_{\mathcal{M}}(r). \tag{22}$$

The concept of (principal) angle between subspaces will later come in handy. The (principal) angle between two linear subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined such that $\cos(\angle[\mathcal{S}_1, \mathcal{S}_2]) := \min_u \max_v |\langle u, v \rangle|$, where the unit vectors u and v belong to $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. It is known that

$$\|(\downarrow_{\mathcal{S}_1} - \downarrow_{\mathcal{S}_2})(\cdot)\|_{2,2} = \sin(\angle[\mathcal{S}_1, \mathcal{S}_2]), \tag{23}$$

where $\downarrow_{\mathcal{S}_1}$ and $\downarrow_{\mathcal{S}_2}$ are orthogonal projectors onto $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively [54, Theorem 2.5]. The norm above is the spectral norm.

We will also use the following conventions to clarify the exposition. For $x_1 \neq x_2 \in \mathbb{R}^N$, define

$$U(x_1, x_2) := \frac{x_2 - x_1}{\|x_2 - x_1\|}.$$

Additionally, we let $U(S_1, S_2)$ denote the set of directions of all the chords connecting two sets $S_1, S_2 \subseteq \mathbb{R}^N$, namely

$$U(S_1, S_2) := \{U(x_1, x_2) : x_1 \in S_1, x_2 \in S_2, x_1 \neq x_2\}.$$

Clearly, $U(S_1, S_2) \subseteq \mathbb{S}^{N-1}$, where $\mathbb{S}^{N-1}$ is the unit sphere in $\mathbb{R}^N$. Whenever possible, we also simplify our notation by using $U(S) := U(S, S)$.

Below we list a few useful results (mostly from differential geometry) which are used throughout the rest of the paper. We begin with a well-known bound on the covering number of Euclidean balls, e.g., [59, Lemma 5.2].[4]

Lemma 1. *A K-dimensional unit ball can be covered by at most $(3/r)^K$ r-balls with $r \leq 1$.*

---

[4]Lemma 5.2 in [59] concerns the unit sphere in $\mathbb{R}^K$, but the result still holds for the unit Euclidean ball using essentially the same argument.

We now recall several results from Sections 5 and 6 in [43]. Unfortunately we were unable to confirm for ourselves some of the original proofs appearing in [43]. Therefore, some of the statements and proofs below differ slightly from their original counterparts. The first result is closely related to Lemma 5.3 in [43].

Lemma 2. *Fix $p, q \in \mathcal{M}$, such that $\|q - p\| < 2\tau$. Then $\angle[q - p, \downarrow_p (q - p)] < \sin^{-1}(\|q - p\|/2\tau)$.*

Proof. Consider the unit vector v along $(q - p) - \downarrow_p (q - p) \perp \mathcal{T}_p$ and the point $z := p + \tau \cdot v$. By definition of the condition number of the manifold, $\|z - q\| > \tau$, since p is the nearest point on the manifold to z.[5] Now consider the triangle formed by the points p, q, z and the line l passing through z and perpendicular to $q - p$. Let z' denote the intersection of l with the line passing through p and q. It is clear that $\angle[q - p, z - p] \leq \pi/2$. Also since $\|z - q\| > \|z - p\| = \tau$, we have $\angle[z - q, p - q] \leq \angle[z - p, q - p] \leq \pi/2$. Therefore, z' is indeed between p and q. The angle between l and the line passing through p and z equals the angle between $q - p$ and $\downarrow_p (q - p)$. To obtain an upper bound for the latter angle, we again note that $\|z - q\| > \|z - p\|$ and therefore $\|z' - q\| > \|z' - p\|$, or $\|z' - p\| < \frac{1}{2}\|q - p\|$. So, this angle is bounded by $\sin^{-1}(\|q - p\|/2\tau)$. This completes the proof of Lemma 2. $\qquad\square$

Lemma 3. *[43, Lemma 5.4] For $p \in \mathcal{M}$, the derivative of $\downarrow_p$ is nonsingular on $\mathcal{A}_{\mathcal{M}}(p, \tau/2)$.*

Lemma 4. *[43, Proposition 6.1] Let $\gamma(\cdot)$ denote a smooth unit-speed geodesic curve on $\mathcal{M}$ defined on an interval $I \subset \mathbb{R}$. For every $t \in I$, the following holds.*

$$\|\gamma''(t)\| \leq 1/\tau.$$

Lemma 5. *[43, Proposition 6.2] Fix $p, q \in \mathcal{M}$. The angle between $\mathcal{T}_p$ and $\mathcal{T}_q$, $\angle[\mathcal{T}_p, \mathcal{T}_q]$, satisfies $\cos(\angle[\mathcal{T}_p, \mathcal{T}_q]) \geq 1 - d_{\mathcal{M}}(p, q)/\tau$.*

The next lemma guarantees that two points separated by a small Euclidean distance are also separated by a small geodesic distance, and so the manifold does not "curve back" upon itself.

Lemma 6. *[43, Proposition 6.3] For $p, q \in \mathcal{M}$ with $\|q - p\| \leq \tau/2$, we have*

$$d_{\mathcal{M}}(p, q) \leq \tau - \tau\sqrt{1 - \frac{2}{\tau}\|q - p\|}. \tag{24}$$

Proof. The first part of the proof of Proposition 6.3 in [43] establishes that for any $p, q \in \mathcal{M}$,

$$\|q - p\| \geq d_{\mathcal{M}}(p, q) - \frac{(d_{\mathcal{M}}(p, q))^2}{2\tau}, \tag{25}$$

which is satisfied only if (24) is satisfied or if

$$d_{\mathcal{M}}(p, q) \geq \tau + \tau\sqrt{1 - \frac{2}{\tau}\|q - p\|} \tag{26}$$

is satisfied. We provide the following argument to complete the proof.

---

[5]To see this, suppose, in contradiction, that $q' \neq p$ is a point on $\mathcal{M}$ that minimizes the distance to z so that $\|z - q'\| \leq \tau = \|z - p\|$. On the other hand, $z - q'$ is orthogonal to $\mathcal{T}_{q'}$ and thus $\|z - q'\| > \tau$ by the definition of condition number. Therefore we have reached a contradiction and p is indeed the unique nearest neighbor of z on $\mathcal{M}$.

For fixed $p \in \mathcal{M}$, let us consider

$$\widehat{q} := \arg \min_{q \in \mathcal{M}, d_{\mathcal{M}}(p,q) \geq \tau} \|q - p\|.$$

We know the minimizer $\widehat{q}$ exists because we are minimizing a continuous function over a compact set. We consider two cases. First, if $d_{\mathcal{M}}(p, \widehat{q}) = \tau$, then by (25), we will have $\|\widehat{q} - p\| \geq \tau/2$. Second, if $d_{\mathcal{M}}(p, \widehat{q}) > \tau$, then there must exist an open neighborhood of $\widehat{q}$ on $\mathcal{M}$ over which the distance to p is minimized at $\widehat{q}$. This implies that $p - \widehat{q}$ will be normal to $\mathcal{M}$ at $\widehat{q}$, which by the definition of condition number (and the fact that $p \in \mathcal{M}$) means that we must have $\|\widehat{q} - p\| \geq 2\tau$.

Now, for any $p, q \in \mathcal{M}$ such that $\|q - p\| < \tau/2$, (24) would imply that $d_{\mathcal{M}}(p, q) < \tau$ and (26) would imply that $d_{\mathcal{M}}(p, q) > \tau$. From the paragraph above, we see that if $d_{\mathcal{M}}(p, q) \geq \tau$, then $\|q - p\| \geq \|\widehat{q} - p\| \geq \tau/2$, and so we can rule out the possibility that (26) is true. Thus, (24) must hold for any $p, q \in \mathcal{M}$ with $\|q - p\| < \tau/2$.

For any $p, q \in \mathcal{M}$ such that $\|q - p\| = \tau/2$, (24) would imply that $d_{\mathcal{M}}(p, q) \leq \tau$ and (26) would imply that $d_{\mathcal{M}}(p, q) \geq \tau$. From the paragraph above involving $\widehat{q}$, we see that any point $q \in \mathcal{M}$ satisfying both $d_{\mathcal{M}}(p, q) \geq \tau$ and $\|q - p\| = \tau/2$ would have to be a local minimizer of $\|q - p\|$ on the convex set and in fact would have to fall into the first case, implying that $d_{\mathcal{M}}(p, q) = \tau$ exactly. It follows that (24) must hold for any $p, q \in \mathcal{M}$ with $\|q - p\| = \tau/2$. $\qquad\square$

The next lemma concerns the invertibility of $\downarrow_p$ within the neighborhood of p and is closely related to Lemma 5.3 in [43].

Lemma 7. *For $p \in \mathcal{M}$, $\downarrow_p$ is invertible on $\mathcal{A}_{\mathcal{M}}(p, \tau/4)$.*

Proof. Lemma 3 states that the derivative of $\downarrow_p$ is nonsingular on $\mathcal{A}_{\mathcal{M}}(p, \tau/2)$. The inverse function theorem then implies that there exists an $r > 0$ such that $\downarrow_p$ is invertible on $\mathcal{A}_{\mathcal{M}}(p, r\tau)$; without loss of generality assume that $r < 1/4$ (otherwise we are done). Now, suppose that there exists $c > 0$ and distinct points $q, z \in \mathcal{M}$ such that $\|q - p\| = c\tau$, $\|z - p\| \leq c\tau$, and $\downarrow_p (q - p) = \downarrow_p (z - p)$. In particular, this implies that

$$z - q \perp \mathcal{T}_p. \tag{27}$$

That is, for any unit vector $v \in \mathcal{T}_p$, we have

$$\langle z - q, v \rangle = 0. \tag{28}$$

Our goal is to show that $c > 1/4$. Suppose, in contradiction that indeed $c \leq 1/4$. Let $\gamma(\cdot)$ be the unit-speed geodesic curve connecting q to z, such that $\gamma(0) = q$ and $\gamma(d_{\mathcal{M}}(q, z)) = z$. By applying the fundamental theorem of calculus twice, we realize that

$$
\begin{aligned}
z - q &= \gamma(d_{\mathcal{M}}(q, z)) - \gamma(0) \\
&= \int_0^{d_{\mathcal{M}}(q,z)} \gamma'(\alpha) \, d\alpha \\
&= \int_0^{d_{\mathcal{M}}(q,z)} \left(\gamma'(0) + \int_0^\alpha \gamma''(\beta) \, d\beta \right) d\alpha \\
&= \gamma'(0) \cdot d_{\mathcal{M}}(q, z) + \int_0^{d_{\mathcal{M}}(q,z)} \int_0^\alpha \gamma''(\beta) \, d\beta d\alpha.
\end{aligned}
$$

Invoking Lemma 4, we can write that

$$
\begin{aligned}
\|(z - q) - \gamma'(0) \cdot d_{\mathcal{M}}(q, z)\| &\leq \int_0^{d_{\mathcal{M}}(q,z)} \int_0^{\alpha} \|\gamma''(\beta)\| \, d\beta d\alpha \\
&\leq \frac{1}{\tau} \int_0^{d_{\mathcal{M}}(q,z)} \int_0^{\alpha} d\beta d\alpha \\
&= \frac{d_{\mathcal{M}}^2(q, z)}{2\tau}.
\end{aligned}
\tag{29}
$$

Meanwhile, having $\|z - q\| \leq 2c\tau$ implies, via Lemma 6, that

$$
d_{\mathcal{M}}(q, z) \leq \tau - \tau\sqrt{1 - 4c},
\tag{30}
$$

which, after plugging back into (29), yields

$$
\left\| \frac{z - q}{d_{\mathcal{M}}(q, z)} - \gamma'(0) \right\| \leq \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4c}.
\tag{31}
$$

So, for any unit vector $v \in \mathcal{T}_p$, we have

$$
\begin{aligned}
|\langle \gamma'(0), v \rangle| &\leq \left| \left\langle \gamma'(0) - \frac{z - q}{d_M(q, z)}, v \right\rangle \right| + \left| \left\langle \frac{z - q}{d_M(q, z)}, v \right\rangle \right| \\
&= \left| \left\langle \gamma'(0) - \frac{z - q}{d_M(q, z)}, v \right\rangle \right| \\
&\leq \left\| \gamma'(0) - \frac{z - q}{d_M(q, z)} \right\| \\
&\leq \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4c},
\end{aligned}
\tag{32}
$$

where the first line follows from the triangle inequality, and the second line uses (28). The last line uses (31). To reiterate, (32) is valid for any unit vector $v \in \mathcal{T}_p$.

On the other hand, Lemma 5 implies that

$$
\begin{aligned}
\cos\left(\angle[\mathcal{T}_q, \mathcal{T}_p]\right) &\geq 1 - \frac{1}{\tau} d_{\mathcal{M}}(p, q) \\
&\geq \sqrt{1 - \frac{2}{\tau}\|q - p\|} \\
&= \sqrt{1 - 2c},
\end{aligned}
\tag{33}
$$

where the second line follows from Lemma 6, and the last line uses $\|q - p\| = c\tau$. By the definition of the angle between subspaces, (33) implies that there exists a unit vector $v_0 \in \mathcal{T}_p$ such that

$$
\left| \langle v_0, \gamma'(0) \rangle \right| \geq \sqrt{1 - 2c}
\tag{34}
$$

because $\gamma'(0) \in \mathcal{T}_q$. Combining this bound with (32) for $v = v_0$, we realize that

$$
\sqrt{1 - 2c} \leq \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4c}.
$$

This inequality is not met for any $c \leq 1/4$. Thus, indeed $c > 1/4$. In particular, this means that $\downarrow_p$ is invertible on $\mathcal{A}_{\mathcal{M}}(p, \tau/4)$. $\qquad\square$

The next three lemmas are of importance when approximating the long and short chords on $\mathcal{M}$ with, respectively, nearby long chords and vectors on the nearby tangent planes.

Lemma 8. [12, Implicit in Lemma 4.1] *Consider two pair of points a_1, a_2 and b_1, b_2, all in $\mathbb{R}^N$, such that $\|a_1 - b_1\|, \|a_2 - b_2\| \le r$, and that $\|a_1 - a_2\| \ge \kappa\sqrt{r}$, for some $r, \kappa > 0$. Then $\|U(a_1, a_2) - U(b_1, b_2)\| \le 4\kappa^{-1}\sqrt{r}$.*

Lemma 9. *For $a, b \in \mathcal{M}$ with $\|a - b\| \le l_1 < \tau/2$, it holds true that*

$$\| \downarrow_a v - \downarrow_b v\| \le \sqrt{\frac{2l_1}{\tau}},$$

for every unit vector $v \in \mathbb{R}^N$.

Proof. It follows from (23) that

$$\| \downarrow_a v - \downarrow_b v\| \le \|(\downarrow_a - \downarrow_b)(\cdot)\|_{2,2} = \sin(\angle[\mathcal{T}_a, \mathcal{T}_b]). \tag{35}$$

On the other hand, since $\|a - b\| \le l_1 < \tau/2$, Lemma 6 implies that

$$d_{\mathcal{M}}(a, b) \le \tau - \tau\sqrt{1 - \frac{2l_1}{\tau}},$$

and thus, using Lemma 5, we arrive at

$$\cos\left(\angle[\mathcal{T}_a, \mathcal{T}_b]\right) \ge \sqrt{1 - \frac{2l_1}{\tau}}.$$

Plugging back the estimate above into (35), we conclude that $\| \downarrow_a v - \downarrow_b v\| \le \sqrt{2l_1/\tau}$, as claimed. $\quad\square$

Lemma 10. *Fix $p \in \mathcal{M}$, and take two points $x_1, x_2 \in \mathcal{M}$ such that $\|x_1 - p\| \le l_1$ and $\|x_2 - x_1\| \le l_2$, $l_1, l_2 < \tau/2$. Then, we have that*

$$\|U(x_1, x_2) - \downarrow_p U(x_1, x_2)\| \le \sqrt{\frac{2l_1}{\tau}} + \frac{l_2}{2\tau}.$$

Proof. The triangle inequality implies that

$$\|U(x_1, x_2) - \downarrow_p U(x_1, x_2)\| \le \|U(x_1, x_2) - \downarrow_{x_1} U(x_1, x_2)\| + \| \downarrow_{x_1} U(x_1, x_2) - \downarrow_p U(x_1, x_2)\|. \tag{36}$$

Since $\|x_2 - x_1\| \le l_2 < 2\tau$, Lemma 2 is the right tool to bound the first term on the right hand side of (36):

$$\begin{aligned}
\|U(x_1, x_2) - \downarrow_{x_1} U(x_1, x_2)\| &= \sin\left(\angle[U(x_1, x_2), \downarrow_{x_1} U(x_1, x_2)]\right) \\
&= \sin\left(\angle[x_2 - x_1, \downarrow_{x_1}(x_2 - x_1)]\right) \\
&\le \frac{l_2}{2\tau}.
\end{aligned} \tag{37}$$

Since $\|x_1 - p\| \le l_1 < \tau/2$, a bound on the second term follows from an application of Lemma 9:

$$\| \downarrow_{x_1} U(x_1, x_2) - \downarrow_p U(x_1, x_2)\| \le \sqrt{\frac{2l_1}{\tau}}. \tag{38}$$

Combining (37) and (38) immediately proves our claim. $\quad\square$

B Proof of Theorem 2

It is easily verified that our objective is to find an upper bound for

$$\mathrm{P}\left\{\sup_{y\in U(\mathcal{M})}|\|\Phi y\|-1|>\epsilon\right\},$$

when $\epsilon \leq 1/3$.

The remainder of this section is divided to two parts. In the first part, we construct a sequence of increasingly finer nets for $\mathcal{M}$. This is in turn used to construct a sequence of covers for the set of all directions in $\mathcal{M}$, $U(\mathcal{M})$. In the second part, we apply a chaining argument that utilizes this later sequence of covers to prove Theorem 2.

B.1 Sequence of covers for $U(\mathcal{M})$

For $\eta > 0$, let $C_0(\eta) \subset \mathcal{M}$ denote a minimal η-net for $\mathcal{M}$ over all η-nets that are a subset of $\mathcal{M}$. Upper and lower bounds for $\#C_0(\eta) = \mathcal{N}_{\mathcal{M}}(\eta)$ are known for sufficiently small η [43], where $\#C_0(\eta)$ denotes the cardinality of $C_0(\eta)$. Since the claim below slightly differs from the one in [43], the proof is included in Appendix C.

Lemma 11. *When $\eta \leq \tau/2$, it holds that*

$$\#C_0(\eta) \leq \left(\frac{2}{\theta(\eta/4\tau)\eta}\right)^K \frac{V_{\mathcal{M}}}{V_{\mathcal{B}_K}} =: \mathrm{c}_0(\eta), \tag{39}$$

where $\theta(\alpha) := \sqrt{1-\alpha^2}$ for $|\alpha| \leq 1$.

By replacing η with $4^{-j}\eta$, we can construct a sequence of increasingly finer nets for $\mathcal{M}$, $\{C_j(\eta)\}$, such that $C_j(\eta) \subset \mathcal{M}$ is a $(4^{-j}\eta)$-net for $\mathcal{M}$, for every $j \in \mathbb{N}$. In light of Lemma 11, we have that

$$\#C_j(\eta) \leq 4^{jK} \cdot \mathrm{c}_0(\eta). \tag{40}$$

Construction of a sequence of covers for $U(\mathcal{M})$ demands the following setup. For $\eta' > 0$ and $j \in \mathbb{N}$, let $C^j(\eta')$ denote a minimal $(2^{-j}\eta')$-net for $\mathcal{B}_K$. For $p \in C_j(\eta)$, we can naturally map $C^j(\eta')$ to live in the K-dimensional unit ball along $\mathcal{T}_p$ (and anchored at the origin). We represent this set of vectors by $C^{j,p}(\eta')$ and define

$$C'_j(\eta,\eta') := \bigcup_{p\in C_j(\eta)} C^{j,p}(\eta'),$$

which forms a $(2^{-j}\eta')$-net for the unit balls along the tangent spaces at every point in $C_j(\eta) \subset \mathcal{M}$. For $\delta > 0$, let us specify η and η' as functions of δ. For $C_2, C_3 > 0$ to be set later, take $\eta = \eta(\delta) = C_2^2\tau\delta^2$ and $\eta' = \eta'(\delta) = C_3\sqrt{\eta/\tau} = C_2C_3\delta$. Now, for every $j \in \mathbb{N}$, simply set

$$T_j(\delta) := U(C_j(\eta)) \cup C'_j(\eta,\eta').$$

It turns out that $U(C_j(\eta))$, the set of all directions in $C_j(\eta)$, provides a net for the directions of long chords on $\mathcal{M}$. In contrast, $C'_j(\eta,\eta')$ forms a net for the directions in $U(\mathcal{M})$ that correspond to the short chords on $\mathcal{M}$. It is therefore not surprising that $\{T_j(\delta)\}$ proves to be a sequence of increasingly finer covers for $U(\mathcal{M})$. This discussion is formalized in the next lemma and proved in Appendix D. We remark that Lemma 12 holds more generally for all constants C_2, C_3 that satisfy the conditions listed in the proof.

Lemma 12. *Set $C_2 = 0.4$ and $C_3 = 1.7 - \sqrt{2}$. For every $j \in \mathbb{N}$, $T_j(\delta)$, as constructed above, is a $(2^{-j}\delta)$-net for $U(\mathcal{M})$, when $\delta \leq 1/2$. Under the mild assumption that*

$$\frac{V_{\mathcal{M}}}{\tau^K} \geq \left(\frac{21}{2\sqrt{K}}\right)^K, \tag{41}$$

it also holds that

$$\#T_j(\delta) \leq 2 \cdot 4^{2jK} \left(\frac{6.12\sqrt{K}}{\delta^2}\right)^{2K} \left(\frac{V_{\mathcal{M}}}{\tau^K}\right)^2 =: \mathsf{t}_j(\delta). \tag{42}$$

B.2 Applying the chaining argument

Every $y \in U(\mathcal{M})$ can be represented with a chain of points in $\{T_j(\delta)\}$. Let $\pi_j(y)$ be the nearest point to y in $T_j(\delta)$. This way we obtain a sequence $\{\pi_j(y)\}$ that represents y via an almost surely convergent telescoping sum, that is

$$y = \pi_0(y) + \sum_{j \in \mathbb{N}} (\pi_{j+1}(y) - \pi_j(y)). \tag{43}$$

Note that, for every $j \in \mathbb{N}$ and every $y \in \mathcal{M}$, the length of the chord connecting $\pi_j(y)$ to $\pi_{j+1}(y)$ is no longer than $2^{-j+1}\delta$. We are now ready to state a generic chaining argument, proved in Appendix E, that allows us to bound the failure probability of obtaining a stable embedding of $\mathcal{M}$ in terms of its geometrical properties. The interested reader is referred to [55] for more information about the generic chaining.

Lemma 13. *Fix $0 < \delta < \epsilon_1 < \epsilon \leq 1/3$, and $\epsilon_2 > 0$ such that $\epsilon_1 + \epsilon_2 = \epsilon$. Choose $C_4, C_5 > 0$ so that $\epsilon_1/\delta \geq \frac{1+C_4}{1-C_4}$ and $\epsilon_2/\delta \geq C_5$. Then, under* (41), *we have that*

$$\mathrm{P}\left\{\sup_{y \in U(\mathcal{M})} |\|\Phi y\| - 1| > \epsilon\right\} \leq 2\,\mathsf{t}_0(\delta) \cdot \max_{t_0 \in T_0(\delta)} \mathrm{P}\left\{|\|\Phi t_0\| - \|t_0\|| > C_4\epsilon_1\|t_0\|\right\}$$

$$+ 2\sum_{j \in \mathbb{N}} \mathsf{t}_{j+1}^2(\delta) \cdot \max_{(t_j, s_j) \in Q_j(\delta)} \mathrm{P}\left\{\|\Phi s_j - \Phi t_j\| > 8^{-1}C_5(j+1)\|s_j - t_j\|\right\}, \tag{44}$$

where $\{\mathsf{t}_j(\delta)\}$ were previously defined in Lemma 12. For $j \in \mathbb{N}$, $Q_j(\delta)$ is defined as

$$Q_j(\delta) := \{(t_j, s_j) : \pi_j(y) = t_j \text{ and } \pi_{j+1}(y) = s_j \text{ for some } y \in U(\mathcal{M})\}.$$

There are two type of probabilities involved in the upper bound above. One controls the large deviations of $\|\Phi t_0\|$ from its expectation, and the other corresponds to very large (one sided) deviations of $\|\Phi s_j - \Phi t_j\|$ from its expectation. As claimed in the next lemma and proved in Appendix F, both of these probabilities are exponentially small when M is large enough.

Lemma 14. *Fix $0 \leq \lambda \leq 1/3$ and $\lambda' \geq 1/5$. Then, for fixed $y_1, y_2 \in \mathbb{R}^N$, we have*

$$\mathrm{P}\left\{|\|\Phi y_1\| - \|y_1\|| > \lambda\|y_1\|\right\} \leq 2e^{-\frac{M\lambda^2}{6}} \tag{45}$$

$$\mathrm{P}\left\{\|\Phi y_2 - \Phi y_1\| > (1 + \lambda')\|y_2 - y_1\|\right\} \leq e^{-\frac{M\lambda'}{7}}. \tag{46}$$

Now fix $\epsilon \leq 1/3$ and set $\epsilon_1 = 9\epsilon/10$. Taking $C_4 = \sqrt{6/7}$, $C_5 = 16$, $\delta = \epsilon/160$ and finally assuming (41) guarantees that Lemma 12 is in force. Under this setup, note that an upper bound for the first term on the right hand side of (44) can be found by applying (45) (after plugging in for C_4):

$$2\,\mathsf{t}_0(\delta) \cdot \max_{t_0 \in T_0(\delta)} \mathrm{P}\left\{ |\|\Phi t_0\| - \|t_0\|| > \sqrt{\frac{6}{7}}\epsilon_1 \|t_0\| \right\} \leq 2\,\mathsf{t}_0(\delta) \cdot 2e^{-\frac{M\epsilon_1^2}{7}},$$

and, assuming that

$$M \geq 14\epsilon_1^{-2} \log N_0(\delta), \tag{47}$$

we arrive at

$$2\,\mathsf{t}_0(\delta) \cdot \max_{t_0 \in T_0(\delta)} \mathrm{P}\left\{ |\|\Phi t_0\| - \|t_0\|| > \sqrt{\frac{6}{7}}\epsilon_1 \|t_0\| \right\} \leq 4e^{-\frac{M\epsilon_1^2}{14}}. \tag{48}$$

In order to bound the second term on the right hand side of (44), we proceed as follows. After plugging in for C_5 and applying (46), we obtain that

$$\max_{(t_j,s_j) \in Q_j(\delta)} \mathrm{P}\left\{ \|\Phi s_j - \Phi t_j\| > 2(j+1)\|s_j - t_j\| \right\} \leq e^{-\frac{(2j+1)M}{7}}.$$

Using the estimate above and Lemma 12, we get an upper bound for the second term on the right hand side of (44) (after plugging in for C_5):

$$2\sum_{j \in \mathbb{N}} \mathsf{t}_{j+1}^2(\delta) \cdot \max_{(t_j,s_j) \in Q_j(\delta)} \mathrm{P}\left\{ \|\Phi s_j - \Phi t_j\| > 2(j+1)\|s_j - t_j\| \right\}$$
$$\leq 2\,\mathsf{t}_0^2(\delta)e^{-\frac{M}{7}} 4^{4K} \sum_{j \in \mathbb{N}} 4^{4jK} e^{-\frac{2}{7}jM}. \tag{49}$$

Assuming that

$$M \geq \max(32\log \mathsf{t}_0(\delta), 310K) \tag{50}$$

allows us to continue simplifying (49), therefore arriving at

$$2\sum \mathsf{t}_{j+1}^2(\delta) \cdot \max_{(t_j,s_j) \in Q_j(\delta)} \mathrm{P}\left\{ \|\Phi s_j - \Phi t_j\| > 2(j+1)\|s_j - t_j\| \right\}$$
$$\leq 4e^{-\frac{M}{17}}. \tag{51}$$

We can now combine (48) and (51) to obtain

$$\mathrm{P}\left\{ \sup_y |\|\Phi y\| - 1| > \epsilon \right\} \leq 4e^{-\frac{M\epsilon_1^2}{14}} + 4e^{-\frac{M}{17}} \leq 8e^{-\frac{M\epsilon_1^2}{14}},$$

where the second inequality follows since $\epsilon_1 \leq 1/3$ and thus $\epsilon_1^2/14 \leq 1/17$. In particular, to achieve a failure probability of at most $\rho \leq 1$, we need

$$M \geq 14\epsilon_1^{-2} \log\left(8/\rho\right). \tag{52}$$

Assuming that (41) holds and $\epsilon \leq 1/3$, (50) is absorbed into (47) and (52), which are in turn lumped into a single lower bound on M (after plugging in for δ), that is

$$
\begin{aligned}
M &\geq 18\epsilon^{-2} \max \left(\log(2V_{\mathcal{M}}^2) + 24K + 2K \log \left(\frac{\sqrt{K}}{\tau\epsilon^2} \right), \log \left(\frac{8}{\rho} \right) \right) \\
&\geq 18\epsilon^{-2} \max \left(\log \left(2V_{\mathcal{M}}^2 \left(\frac{6\sqrt{K}}{\tau\delta^2} \right)^{2K} \right), \log \left(\frac{8}{\rho} \right) \right) \\
&= 18\epsilon^{-2} \max \left(\log N_0(\delta), \log(8/\rho) \right) \\
&\geq 14\epsilon_1^{-2} \max \left(\log N_0(\delta), \log(8/\rho) \right).
\end{aligned}
\tag{53}
$$

So far, we proved that

$$
\mathrm{P} \left\{ \sup_{y \in U(\mathcal{M})} |\|\Phi y\| - 1| > \epsilon \right\} \leq \rho,
$$

provided that M satisfies (53). This completes the proof of Theorem 2.

C Proof of Lemma 11

To prove this bound, we first need the following result regarding the local properties of $\mathcal{M}$, which is closely related to Lemma 5.3 in [43]. The proof can be found in Appendix G.

Lemma 15. *Let $p \in \mathcal{M}$ and $r \leq \tau/4$. Then the following holds:*

$$
\mathrm{vol}_K(\mathcal{A}_{\mathcal{M}}(p, r)) \geq \left(1 - \frac{r^2}{4\tau^2} \right)^{\frac{K}{2}} r^K V_{\mathcal{B}_K},
$$

where $\mathrm{vol}_K(\cdot)$ measures the K-dimensional volume.

Now, using (22) and a simple volume comparison argument, we observe that

$$
\#C_0(\eta) = \mathcal{N}_{\mathcal{M}}(\eta) \leq \mathcal{P}_{\mathcal{M}}(\eta) \leq \frac{V_{\mathcal{M}}}{\inf_{p \in \mathcal{M}} \mathrm{vol}_K(\mathcal{A}_{\mathcal{M}}(p, \eta/2))}.
$$

Since $\eta/2 \leq \tau/4$, we can apply Lemma 15 (with $r = \eta/2$) and obtain that

$$
\begin{aligned}
\#C_0(\eta) &\leq \frac{V_{\mathcal{M}}}{\left(1 - \frac{\eta^2}{16\tau^2} \right)^{\frac{K}{2}} \left(\frac{\eta}{2} \right)^K V_{\mathcal{B}_K}} \\
&= \left(\frac{2}{\theta(\eta/4\tau)\eta} \right)^K \frac{V_{\mathcal{M}}}{V_{\mathcal{B}_K}},
\end{aligned}
$$

This completes the proof of Lemma 11.

D Proof of Lemma 12

Consider two arbitrary but distinct points $x_1, x_2 \in \mathcal{M}$. For $C_6 > 0$ to be set later in the proof, we separate the treatment of long and short chords, i.e., $\|x_2 - x_1\|/\tau > C_6\sqrt{\eta/\tau} =: \gamma$ and $\|x_2 - x_1\|/\tau \leq$

γ, and in this strategy we follow [3, 12]. Short chords are distinct in that, as we will see later, they have to be approximated with nearby tangent vectors. For convenience, let us also define

$$U_\gamma^l(\mathcal{M}) := \{U(z_1, z_2) : \|z_2 - z_1\| > \gamma\tau, \ z_1, z_2 \in \mathcal{M}\},$$

$$U_\gamma^s(\mathcal{M}) := \{U(z_1, z_2) : 0 < \|z_2 - z_1\| \leq \gamma\tau, \ z_1, z_2 \in \mathcal{M}\}.$$

Of course, $U_\gamma^l(\mathcal{M}) \cup U_\gamma^s(\mathcal{M}) = U(\mathcal{M})$, although their intersection might not be empty.

Suppose that $\|x_2 - x_1\|/\tau > \gamma = C_6\sqrt{\eta/\tau}$ so that $U(x_1, x_2) \in U_\gamma^l(\mathcal{M})$. Since $C_0(\eta)$ is an η-net for $\mathcal{M}$, there exist p_1 and p_2 in $C_0(\eta)$ such that $\|x_1 - p_1\|, \|x_2 - p_2\| \leq \eta$. It then follows from Lemma 8 (with $a_1 = x_1$, $a_2 = x_2$, $b_1 = p_1$, and $b_2 = p_2$) that

$$\|U(x_1, x_2) - U(p_1, p_2)\| \leq \frac{4}{C_6}\sqrt{\frac{\eta}{\tau}} = \frac{4C_2\delta}{C_6}. \tag{54}$$

Now, assuming that

$$4C_2 = C_6, \tag{55}$$

and leveraging the fact that the choice of $x_1, x_2 \in \mathcal{M}$ was arbitrary, we conclude that $U(C_0(\eta))$ is a δ-net for $U_\gamma^l(\mathcal{M})$.

On the other hand, suppose that $0 < \|x_2 - x_1\|/\tau \leq \gamma = C_6\sqrt{\eta/\tau} = 4C_2\sqrt{\eta/\tau}$ so that $U(x_1, x_2) \in U_\gamma^s(\mathcal{M})$. We assume that

$$\frac{\eta}{\tau} = C_2^2\delta^2 < \min\left(\frac{1}{64C_2^2}, \frac{1}{2}\right), \tag{56}$$

so that, in particular, $\|x_2 - x_1\| < \tau/2$. Since $C_0(\eta)$ is an η-net for $\mathcal{M}$, there exists a point $p \in C_0(\eta)$ such that $\|x_1 - p\| \leq \eta < \tau/2$. Lemma 10 (with $l_1 = \eta$ and $l_2 = 4C_2\sqrt{\tau\eta}$) then implies that the direction of the chord connecting x_1 to x_2 can be approximated with a tangent vector in $\mathcal{T}_p$, that is

$$\|U(x_1, x_2) - \downarrow_p U(x_1, x_2)\| \leq \sqrt{\frac{2\eta}{\tau}} + 2C_2\sqrt{\frac{\eta}{\tau}} = \left(\sqrt{2} + 2C_2\right)C_2\delta. \tag{57}$$

Recall that $C^{0,p}(\eta')$ is an η'-net for the unit ball centered at p and along $\mathcal{T}_p$. So, there also exists a vector $v \in C^{0,p}(\eta')$ such that $\|\downarrow_p U(x_1, x_2) - v\| \leq \eta' = C_2C_3\delta$. Using the triangle inequality, we therefore arrive at

$$\|U(x_1, x_2) - v\| \leq \left(\sqrt{2} + 2C_2 + C_3\right)C_2\delta. \tag{58}$$

Assuming that

$$\sqrt{2} + 2C_2 + C_3 = C_2^{-1}, \tag{59}$$

and leveraging the fact that the choice of $x_1, x_2 \in \mathcal{M}$ was arbitrary, we conclude that $C_0'(\eta, \eta')$ is a δ-net for $U_\gamma^s(\mathcal{M})$. Overall, under (55), (56), and (59), $T_0(\delta) = U(C_0(\eta)) \cup C_0'(\eta, \eta')$ is a δ-net for $U(\mathcal{M})$. By repeating the argument above (with $\eta, \delta, \eta', \gamma$ replaced with $\eta/4^j, \delta/2^j, \eta'/2^j, \gamma/2^j$) we observe that $T_j(\delta)$ is a $(2^{-j}\delta)$-net for $U(\mathcal{M})$, for every $j \in \mathbb{N}$. In particular, the choice of $C_2 = 0.4, C_3 = 1.7 - \sqrt{2}, C_6 = 1.6$ satisfies the conditions above for every $\delta \leq 1/2$ and completes the proof of the first statement in Lemma 12.

In order to bound the cardinality of $T_j(\delta)$, we begin with estimating $\#C_j'(\eta, \eta')$. According to Lemma 1, we can write that

$$\#C_j'\left(\eta, \eta'\right) \leq \left(\frac{3 \cdot 2^j}{\eta'}\right)^K \cdot \#C_j\left(\eta\right), \tag{60}$$

which holds assuming that $\eta' \leq 1$, i.e., $C_2 C_3 \delta \leq 1$. (Our choice of C_2, C_3 above satisfies this condition.) It is possible now to write that

$$
\begin{aligned}
\#T_j(\delta) &\leq (\#C_j(\eta))^2 + \#C_j'(\eta, \eta') \\
&\leq (\#C_j(\eta))^2 + \left(\frac{3 \cdot 2^j}{\eta'}\right)^K \cdot \#C_j(\eta) \\
&\leq 2\max\left(4^{jK} c_0(\eta), \left(\frac{3 \cdot 2^j}{\eta'}\right)^K\right) \cdot 4^{jK} c_0(\eta),
\end{aligned}
\tag{61}
$$

where we used (60) in the second line and (40) in the last line. To guarantee that the first term dominates the maximum in (61), it suffices (according to the definition of $c_0(\eta)$ in (39)) to enforce that

$$
\left(\frac{2}{\theta(\eta/4\tau)\eta}\right)^K \frac{V_\mathcal{M}}{V_{\mathcal{B}_K}} \geq \left(\frac{3}{\eta'}\right)^K,
$$

which, after plugging in for η and η' in terms of δ and using the hypothesis that $\delta \leq 1$, is satisfied under the mild assumption that

$$
\frac{V_\mathcal{M}}{\tau^K} \geq 2.5^K V_{\mathcal{B}_K} \geq \left(\frac{3C_2}{2C_3}\right)^K V_{\mathcal{B}_K}.
\tag{62}
$$

The assumption in (62) allows us to simplify (61) and obtain that

$$
\#T_j(\delta) \leq 2 \cdot 4^{2jK} c_0^2(\eta).
\tag{63}
$$

It follows from (63) and the definition of $c_0(\eta)$ in (39) that

$$
\#T_j(\delta) \leq 2 \cdot 4^{2jK} \left(\frac{2}{\theta(C_2^2/4)C_2^2 \tau \delta^2}\right)^{2K} \left(\frac{V_\mathcal{M}}{V_{\mathcal{B}_K}}\right)^2 \leq 2 \cdot 4^{2jK} \left(\frac{12.52}{\tau\delta^2}\right)^{2K} \left(\frac{V_\mathcal{M}}{V_{\mathcal{B}_K}}\right)^2,
$$

where we used the fact that $\delta \leq 1$. We remind the reader that

$$
\left(\frac{4\pi}{K+2}\right)^{K/2} \leq V_{\mathcal{B}_K} = \frac{\pi^{K/2}}{\Gamma\left(\frac{K}{2}+1\right)} \leq \left(\frac{2e\pi}{K+2}\right)^{K/2},
\tag{64}
$$

where the inequalities follow from the fact that $(K/e)^{K-1} \leq \Gamma(K) \leq (K/2)^{K-1}$ for $K \in \mathbb{N}$ [44]. Here $\Gamma(\cdot)$ denotes the Gamma function. The above inequality leads us to

$$
\#T_j(\delta) \leq 2 \cdot 4^{2jK} \left(\frac{6.12\sqrt{K}}{\delta^2}\right)^{2K} \left(\frac{V_\mathcal{M}}{\tau^K}\right)^2,
$$

which holds under the mild assumption that $V_\mathcal{M}/\tau^K \geq (21/\sqrt{K})^K$. Indeed, this assumption is obtained by plugging our choice of C_2, C_3 into (62). This completes the proof of Lemma 12.

E Proof of Lemma 13

For notational convenience, let us denote the infinite sum in (43) by $\Sigma(y)$. Then, using the triangle inequality, we observe that

$$
\begin{aligned}
\mathrm{P}\left\{\sup_{y \in U(\mathcal{M})}\|\Phi y\| > 1 + \epsilon\right\} &= \mathrm{P}\left\{\sup_y \|\Phi \pi_0(y) + \Phi\Sigma(y)\| > 1 + \epsilon_1 + \epsilon_2\right\} \\
&\leq \mathrm{P}\left\{\sup_y \|\Phi\pi_0(y)\| + \sup_y \|\Phi\Sigma(y)\| > 1 + \epsilon_1 + \epsilon_2\right\} \\
&\leq \mathrm{P}\left\{\sup_y \|\Phi\pi_0(y)\| - 1 > \epsilon_1\right\} + \mathrm{P}\left\{\sup_y \|\Phi\Sigma(y)\| > \epsilon_2\right\},
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
\mathrm{P}\left\{\inf_{y \in U(\mathcal{M})}\|\Phi y\| < 1 - \epsilon\right\} &= \mathrm{P}\left\{\inf_y \|\Phi \pi_0(y) + \Phi\Sigma(y)\| < 1 - \epsilon_1 - \epsilon_2\right\} \\
&\leq \mathrm{P}\left\{\inf_y \|\Phi\pi_0(y)\| - \sup_y \|\Phi\Sigma(y)\| < 1 - \epsilon_1 - \epsilon_2\right\} \\
&\leq \mathrm{P}\left\{\sup_y 1 - \|\Phi\pi_0(y)\| > \epsilon_1\right\} + \mathrm{P}\left\{\sup_y \|\Phi\Sigma(y)\| > \epsilon_2\right\}.
\end{aligned}
$$

We can therefore argue that

$$
\begin{aligned}
\mathrm{P}\left\{\sup_{y \in U(\mathcal{M})} |\|\Phi y\| - 1| > \epsilon\right\} &\leq \mathrm{P}\left\{\sup_y \|\Phi y\| > 1 + \epsilon\right\} + \mathrm{P}\left\{\inf_y \|\Phi y\| < 1 - \epsilon\right\} \\
&\leq 2\,\mathrm{P}\left\{\sup_y |\|\Phi\pi_0(y)\| - 1| > \epsilon_1\right\} + 2\,\mathrm{P}\left\{\sup_y \|\Phi\Sigma(y)\| > \epsilon_2\right\}. \quad (65)
\end{aligned}
$$

Consider the first probability on the last line of (65):

$$
\begin{aligned}
\mathrm{P}\left\{\sup_{y \in U(\mathcal{M})} |\|\Phi\pi_0(y)\| - 1| > \epsilon_1\right\} &\leq \mathrm{P}\left\{\sup_y |\|\Phi\pi_0(y)\| - \|\pi_0(y)\|| + \sup_y |\|\pi_0(y)\| - 1| > \epsilon_1\right\} \\
&\leq \mathrm{P}\left\{\sup_y |\|\Phi\pi_0(y)\| - \|\pi_0(y)\|| > \epsilon_1 - \delta\right\} \\
&\leq \mathrm{P}\left\{\sup_y \frac{|\|\Phi\pi_0(y)\| - \|\pi_0(y)\||}{\|\pi_0(y)\|} > \frac{\epsilon_1 - \delta}{1 + \delta}\right\} \\
&\leq \mathrm{P}\left\{\sup_y \frac{|\|\Phi\pi_0(y)\| - \|\pi_0(y)\||}{\|\pi_0(y)\|} > C_4\epsilon_1\right\} \\
&\leq \mathrm{P}\left\{\max_{t_0 \in T_0(\delta)} \frac{|\|\Phi t_0\| - \|t_0\||}{\|t_0\|} > C_4\epsilon_1\right\} \\
&\leq \#T_0(\delta) \cdot \max_{t_0 \in T_0(\delta)} \mathrm{P}\left\{|\|\Phi t_0\| - \|t_0\|| > C_4\epsilon_1\|t_0\|\right\}
\end{aligned}
$$

where the first line uses the triangle inequality. The second and third lines hold on account of $T_0(\delta)$ being a net for a subset of $\mathbb{S}^{N-1}$, namely $U(\mathcal{M})$. An application of the union bound gives the last line above.

Now consider the second probability on the last line of (65). By the definition of $\Sigma(y)$, we observe that

$$
\mathrm{P}\left\{\sup_{y \in U(\mathcal{M})} \|\Phi\Sigma(y)\| > \epsilon_2\right\}
$$

$$
= \mathrm{P}\left\{\sup_y \left\|\sum \Phi\pi_{j+1}(y) - \Phi\pi_j(y)\right\| > \epsilon_2\right\}
$$

$$
\leq \mathrm{P}\left\{\sum_j \max_{(t_j,s_j) \in Q_j(\delta)} \|\Phi s_j - \Phi t_j\| > C_5\delta\right\}
$$

$$
= \mathrm{P}\left\{\sum_j \max_{(t_j,s_j) \in Q_j(\delta)} \|\Phi s_j - \Phi t_j\| > C_5 \sum_j (j+1)2^{-j-2}\delta\right\}
$$

$$
\leq \sum_j \mathrm{P}\left\{\max_{(t_j,s_j) \in Q_j(\delta)} \|\Phi s_j - \Phi t_j\| > C_5(j+1)2^{-j-2}\delta\right\}
$$

$$
\leq \sum_j \mathrm{P}\left\{\max_{(t_j,s_j) \in Q_j(\delta)} \|\Phi s_j - \Phi t_j\| > 8^{-1}C_5(j+1)\|s_j - t_j\|\right\}
$$

$$
\leq \sum_j \#T_{j+1}^2(\delta) \max_{(t_j,s_j) \in Q_j(\delta)} \mathrm{P}\left\{\|\Phi s_j - \Phi t_j\| > 8^{-1}C_5(j+1)\|s_j - t_j\|\right\}.
$$

The third line above uses the triangle inequality and the assumption on ϵ_2, while the fifth and last lines use the union bound. It can be easily verified that the infinite sum on the right hand side of the inequality in the fourth line equals one. Having upper bounds for both terms on the last line of (65), we overall arrive at

$$
\mathrm{P}\left\{\sup_{y \in U(\mathcal{M})} |\|\Phi y\| - 1| > \epsilon\right\} \leq 2\#T_0(\delta) \cdot \max_{t_0 \in T_0(\delta)} \mathrm{P}\left\{|\|\Phi t_0\| - \|t_0\|| > C_4\epsilon_1\|t_0\|\right\}
$$

$$
+ 2\sum_j \#T_{j+1}^2(\delta) \max_{(t_j,s_j) \in Q_j(\delta)} \mathrm{P}\left\{\|\Phi s_j - \Phi t_j\| > 8^{-1}C_5(j+1)\|s_j - t_j\|\right\}.
$$

From Lemma 12, $\#T_j(\delta) \leq \mathsf{t}_j(\delta)$. This establishes Lemma 13.

F   Proof of Lemma 14

The proof is elementary. It is easily verified that $\mathbb{E}\|\Phi y_1\|^2 = \|y_1\|^2$, and we then note that

$$
\mathrm{P}\left\{|\|\Phi y_1\| - \|y_1\|| > \lambda\|y_1\|\right\} = \mathrm{P}\left\{\|\Phi y_1\| > (1+\lambda)\|y_1\|\right\} + \mathrm{P}\left\{\|\Phi y_1\| < (1-\lambda)\|y_1\|\right\}
$$

$$
\leq \mathrm{P}\left\{\|\Phi y_1\|^2 > (1+\lambda)\|y_1\|^2\right\} + \mathrm{P}\left\{\|\Phi y_1\|^2 < (1-\lambda)\|y_1\|^2\right\}
$$

$$
\leq 2e^{-\frac{M}{2}\left(\frac{\lambda^2}{2} - \frac{\lambda^3}{3}\right)}
$$

$$
\leq 2e^{-\frac{M\lambda^2}{2}\left(\frac{1}{2} - \frac{1}{9}\right)}
$$

$$
\leq 2e^{-\frac{M\lambda^2}{6}},
$$

where the third line uses a well-known concentration bound [1]. The fourth line holds because $\lambda \leq 1/3$. This establishes the first inequality in Lemma 14. For the second inequality, we begin by

observing that

$$\mathrm{P}\left\{\|\Phi\left(y_2 - y_1\right)\| > \left(1 + \lambda'\right)\|y_2 - y_1\|\right\} = \mathrm{P}\left\{\|\Phi \cdot U\left(y_1, y_2\right)\| > 1 + \lambda'\right\}$$
$$\leq \mathrm{P}\left\{\|\Phi \cdot U\left(y_1, y_2\right)\|^2 > 1 + 2\lambda'\right\}$$
$$= \mathrm{P}\left\{M^{-1}\sum_{i=1}^{M}n_i^2 - 1 > 2\lambda'\right\}$$
$$= \mathrm{P}\left\{\sum_{i=1}^{M}n_i^2 - M > 2\lambda'M\right\}, \tag{66}$$

where $n_1, n_2, \cdots, n_M$ are zero-mean and unit-variance Gaussian random variables. The third line above follows since the entries of the vector $\Phi \cdot U(y_1, y_2)$ are distributed as i.i.d. zero-mean Gaussians with variance of $1/M$. We now recall Lemma 1 in [38], which states that

$$\mathrm{P}\left\{\sum_{i=1}^{M}n_i^2 - M > 2\sqrt{M\alpha} + 2\alpha\right\} \leq e^{-\alpha}, \tag{67}$$

for $\alpha > 0$. Comparing the last line in (66) to the inequality above, we observe that taking

$$\alpha = \frac{M}{4}\left(\sqrt{1 + 4\lambda'} - 1\right)^2$$

allows us to continue simplifying (66) to obtain that

$$\mathrm{P}\left\{\|\Phi\left(y_2 - y_1\right)\| > \left(1 + \lambda'\right)\|y_2 - y_1\|\right\} \leq \mathrm{P}\left\{\sum_{i=1}^{M}n_i^2 - M > 2\sqrt{M\alpha} + 2\alpha\right\} \leq e^{-\alpha}. \tag{68}$$

It is easily verified that $\sqrt{1 + 4\lambda'} - 1 \geq (3 - \sqrt{5})\sqrt{\lambda'}$ when $\lambda' \geq 1/5$. It follows that

$$\alpha \geq \frac{M}{4} \cdot (3 - \sqrt{5})^2\lambda' \geq M\lambda'/7, \tag{69}$$

and consequently,

$$\mathrm{P}\left\{\|\Phi\left(y_2 - y_1\right)\| > \left(1 + \lambda'\right)\|y_2 - y_1\|\right\} \leq e^{-\frac{M\lambda'}{7}},$$

as claimed. This establishes the second inequality in Lemma 14 and completes the proof.

G   Proof of Lemma 15

As in the proof of Lemma 5.3 in [43], we will show that for some $r' > 0$ to be defined below,

$$B_{\mathcal{T}_p}(r') \subset \downarrow_p (A_{\mathcal{M}}(p, r)),$$

as our claim follows directly from the inclusion above. To show the above inclusion, we use the following argument. Let us denote the inverse of $\downarrow_p$ on $\mathcal{A}_{\mathcal{M}}(p, \tau/4)$ with $g(\cdot)$.

From Lemma 7, $\downarrow_p$ is invertible on $A_{\mathcal{M}}(p, r)$ and therefore $\downarrow_p (\mathcal{A}_{\mathcal{M}}(p, r))$ is an open set. Thus there exists $s > 0$ such that $B_{\mathcal{T}_p}(s) \subset \downarrow_p (\mathcal{A}_{\mathcal{M}}(p, r))$. We can keep increasing s until at $s = s^*$ we reach a point y on the boundary of the closure of $B_{\mathcal{T}_p}(s^*)$ such that $y \notin \downarrow_p (\mathcal{A}_{\mathcal{M}}(p, r))$. Consider a sequence $\{y_i\} \subset B_{\mathcal{T}_p}(s^*) \subset \downarrow_p (\mathcal{A}_{\mathcal{M}}(p, r))$ such that $y_i \to y$ when $i \to \infty$. Note that $\{g(y_i)\} \subset$

$\mathcal{A}_{\mathcal{M}}(p, r)$ and, because every sequence in a compact space contains a convergent subsequence, there exist a convergent sebsequence $\{g(y_{i_k})\}$ and x in the closure of $\mathcal{A}_{\mathcal{M}}(p, r)$ such that $g(y_{i_k}) \to x$. Since $\downarrow_p$ is continuous, $\downarrow_p x = y$. Therefore $y =\downarrow_p x \notin\downarrow_p (\mathcal{A}_{\mathcal{M}}(p, r))$, and $x \notin \mathcal{A}_{\mathcal{M}}(p, r)$ and thus x is on the boundary of the closure of $\mathcal{A}_{\mathcal{M}}(p, r)$ and $\|x\| = r$. Now we invoke Lemma 2 with $q = x+p$ to obtain that

$$\cos(\angle[x, y]) \geq \sqrt{1 - \frac{r^2}{4\tau^2}}.$$

It follows that

$$
\begin{aligned}
s^* &= \|y\| \\
&= \cos(\angle[x, y]) \cdot r \\
&\geq \sqrt{1 - \frac{r^2}{4\tau^2}} \cdot r =: r',
\end{aligned}
$$

and thus $B_{\mathcal{T}_p}(r') \subset\downarrow_p (\mathcal{A}_{\mathcal{M}}(p, r))$. This completes the proof of Lemma 15 since

$$\mathrm{vol}_K(\mathcal{A}_{\mathcal{M}}(p, r)) \geq \mathrm{vol}_K(\downarrow_p (\mathcal{A}_{\mathcal{M}}(p, r))) \geq \mathrm{vol}_K(\mathcal{B}_{\mathcal{T}_p}(r')) = (r')^K \mathrm{vol}(\mathcal{B}_K),$$

where the first inequality holds because projection onto a subspace is non-expansive.

H   Proof of Theorem 3

Fix $\alpha \in [1 - \epsilon, 1 + \epsilon]$. We consider any two points $w_a, w_b \in \mathcal{M}$ such that

$$\frac{\|\Phi w_a - \Phi w_b\|}{\|w_a - w_b\|} = \alpha,$$

and supposing that x is closer to w_a, i.e.,

$$\|x - w_a\| \leq \|x - w_b\|,$$

but Φx is closer to Φw_b, i.e.,

$$\|\Phi x - \Phi w_b\| \leq \|\Phi x - \Phi w_a\|,$$

we seek the maximum value that

$$\frac{\|x - w_b\|}{\|x - w_a\|}$$

may take. In other words, we wish to bound the worst possible "mistake" (according to our error criterion) between two candidate points on the manifold whose distance is scaled by the factor α.

This can be posed in the form of an optimization problem

$$
\max_{x \in \mathbb{R}^N, w_a, w_b \in \mathcal{M}} \frac{\|x - w_b\|}{\|x - w_a\|} \quad \text{s.t.} \quad \|x - w_a\| \leq \|x - w_b\|,
$$
$$
\|\Phi x - \Phi w_b\| \leq \|\Phi x - \Phi w_a\|,
$$
$$
\frac{\|\Phi w_a - \Phi w_b\|}{\|w_a - w_b\|} = \alpha.
$$

For simplicity, we may expand the constraint set to include all $w_a, w_b \in \mathbb{R}^N$; the solution to this larger problem is an upper bound for the solution to the case where $w_a, w_b \in \mathcal{M}$.

The constraints and objective function now are invariant to adding a constant to all three variables or to a constant rescaling of all three. Hence, without loss of generality, we set $w_a = \mathbf{0}$ and $\|x\| = 1$. This leaves

$$\max_{x, w_b \in \mathbb{R}^N} \|x - w_b\| \quad \text{s.t.} \quad \|x\| = 1,$$

$$\|x - w_b\| \geq 1,$$
$$\|\Phi x - \Phi w_b\| \leq \|\Phi x\|,$$
$$\frac{\|\Phi w_b\|}{\|w_b\|} = \alpha.$$

We may safely ignore the second constraint (because of its relation to the objective function). Under the constraints above, the objective satisfies

$$\|x - w_b\| \leq \|x\| + \|w_b\|$$
$$= 1 + \|\Phi w_b\|/\alpha$$
$$\leq 1 + 2\|\Phi x\|/\alpha$$
$$\leq 1 + \frac{2\sigma_M(\Phi)}{1 - \epsilon}$$
$$\leq \frac{1}{1 - \epsilon}\left(2\sigma_M(\Phi) + 1\right).$$

The first line follows from the triangle inequality. The identity above uses the first constraint. The third constraint (via the triangle inequality) implies that $\|\Phi w_b\| \leq 2\|\Phi x\|$ and the third line thus follows. The fourth line follows from the fact that $\|x\| = 1$ and after considering the possible range of α. Returning to the original optimization problem, this implies that

$$\frac{\|x - w_b\|}{\|x - w_a\|} \leq \frac{1}{1 - \epsilon}\left(2\sigma_M(\Phi) + 1\right)$$

for any observation x that could be mistakenly paired with w_b instead of w_a (under a Φ that scales the distance $\|w_a - w_b\|$ by α). This completes the proof of Theorem 3.

I   Proof of Proposition 1

Set

$$\delta := (1 + \epsilon)\left(\sigma_m(\Phi)\right)^{-1},$$

and let

$$x = e_1 + \delta u,$$

where $\|u\| \leq 1$ belongs to the row span of Φ and satisfies $\Phi x = \Phi(e_1 + \delta u) = 0$. Finding such u is possible because

$$\|\Phi e_1\| \leq (1 + \epsilon)\|e_1\| = 1 + \epsilon = \delta \cdot \sigma_m(\Phi) = \delta \cdot \sigma_m(\Phi)\|v\| \leq \delta\|\Phi v\|,$$

for every unit vector v in the row span of Φ. The first inequality holds because $e_1, \mathbf{0} \in \mathcal{M}$ and Φ stably embeds $\mathcal{M}$. The second equality holds by our choice of δ, and the last inequality holds because v belongs to the row span of Φ. With our choice of x above, we have $\Phi x = 0$ and therefore $\widehat{x} = 0$. On the other hand,

$$\|x - x^*\| \leq \|x - e_1\| \leq \delta.$$

It follows that

$$\frac{\|x - \widehat{x}\|}{\|x - x^*\|} \geq \frac{\|x - \widehat{x}\|}{\|x - e_1\|} \geq \frac{\|x\|}{\delta} \geq \frac{1 - \delta}{\delta} \geq \frac{1}{2\delta} = \frac{1}{2(1 + \epsilon)} \sigma_m(\Phi).$$

Indeed, one can verify that $\delta \leq 1/2$ because (by hypothesis) $\epsilon \leq 1/3$ and $\sigma_m(\Phi) \geq 8/3$. This immediately implies the second to last inequality above. This completes the proof of Proposition 1.

J   Proof of Theorem 4

Our success in stably embedding $\mathcal{M}$ via random linear measurements (and what distinguished Theorem 2 from embedding of a point cloud) relied on the smoothness of $\mathcal{M}$. This assumption enabled us to control the behavior of short chords on $\mathcal{M}$. However, x does not generally belong to $\mathcal{M}$ and hence, in general, we cannot control the direction of short chords connecting x to $\mathcal{M}$. To deal with this issue, we proceed as follows. For fixed $\gamma > 0$ to be specified later, define

$$\mathcal{M}_\gamma := \{z \in \mathcal{M} \, : \, \|z - x\| > \gamma\tau\},$$

and let $\mathcal{M}_\gamma^C := \mathcal{M} \backslash \mathcal{M}_\gamma$, i.e., the complement of $\mathcal{M}_\gamma$ in $\mathcal{M}$. Note that one of the two sets may be empty. Our first step towards a proof is to show that, for every $z \in \mathcal{M}_\gamma$ with an appropriately chosen γ, we have

$$(1 - \epsilon) \|z - x\| \leq \|\Phi z - \Phi x\| \leq (1 + \epsilon) \|z - x\|. \tag{70}$$

In other words, we first study the stable embedding of the directions of all the chords connecting x to $\mathcal{M}_\gamma$, namely $U(\mathcal{M}_\gamma, x)$, for an appropriate γ. This is addressed next and proved in Appendix K.

Lemma 16. *Choose $0 < \epsilon \leq 1/3$ and $0 < \rho < 1$. Conveniently assume that*

$$\frac{V_\mathcal{M}}{\tau^K} \geq \left(\frac{21}{\sqrt{K}}\right)^K.$$

If

$$M \geq 18\epsilon^{-2} \max\left(11K + K \log\left(\frac{\sqrt{K}}{\tau\epsilon^2}\right) + \log V_\mathcal{M} \, , \, \log\left(\frac{8}{\rho}\right)\right),$$

then, except with a probability of at most ρ, (70) holds for every $z \in \mathcal{M}_{\epsilon/40}$.

We now combine Lemma 16 and an elementary argument to complete the proof of Theorem 4. It is possible to recognize two different cases: when $\widehat{x} \in \mathcal{M}_{\epsilon/40}^C$ and when $\widehat{x} \in \mathcal{M}_{\epsilon/40}$. Clearly,

$$\|x - x^*\| \leq \|x - \widehat{x}\| \leq \frac{\epsilon\tau}{40}, \qquad \text{when } \widehat{x} \in \mathcal{M}_{\epsilon/40}^C. \tag{71}$$

If, however, $\widehat{x} \in \mathcal{M}_{\epsilon/40}$, then a more detailed analysis is required. An application on Lemma 14 implies that (70) holds for $z = x^*$, except with a probability of at most ρ and provided that $M \geq 6\epsilon^{-2} \log(2/\rho)$. Suppose the assumptions in Lemma 16 are met. Therefore, (70) holds for every $z \in \mathcal{M}_{\epsilon/40} \cup \{x^*\}$, except with a probability of at most 2ρ. Also, by the definition of x^* and $\widehat{x}$, it holds true that

$$\|x - x^*\| \leq \|x - \widehat{x}\|$$

$$\|(\Phi x + n) - \Phi\widehat{x}\| \leq \|(\Phi x + n) - \Phi x^*\|.$$

Now, combining all these bounds and using several applications of the triangle inequality we obtain that

$$
\begin{aligned}
\|x - \widehat{x}\| &\leq (1 - \epsilon)^{-1} \|\Phi x - \Phi \widehat{x}\| \\
&\leq (1 - \epsilon)^{-1} \|(\Phi x + n) - \Phi \widehat{x}\| + (1 - \epsilon)^{-1} \|n\| \\
&\leq (1 - \epsilon)^{-1} \|(\Phi x + n) - \Phi x^*\| + (1 - \epsilon)^{-1} \|n\| \\
&\leq (1 - \epsilon)^{-1} \|(\Phi x + n) - \Phi x^*\| + (1 - \epsilon)^{-1} \|n\| \\
&\leq (1 - \epsilon)^{-1} \|\Phi x - \Phi x^*\| + 2(1 - \epsilon)^{-1} \|n\| \\
&\leq \frac{1 + \epsilon}{1 - \epsilon} \|x - x^*\| + 2(1 - \epsilon)^{-1} \|n\|.
\end{aligned}
$$

Since $\epsilon \leq 1/3$, one can easily check that

$$
(1 - \epsilon)^{-1} \leq 1 + \epsilon/2,
$$

$$
\frac{1 + \epsilon}{1 - \epsilon} \leq 1 + 2\epsilon.
$$

Consequently, we obtain that

$$
\|x - \widehat{x}\| \leq (1 + 2\epsilon) \|x - x^*\| + (2 + \epsilon) \|n\|, \qquad \text{when } \widehat{x} \in M_{\epsilon/40}. \tag{72}
$$

Combining (71) and (72), we overall obtain that

$$
\|x - \widehat{x}\| \leq \max\left(\frac{\epsilon\tau}{40}, \, (1 + 2\epsilon) \|x - x^*\| + (2 + \epsilon) \|n\| \right),
$$

which, to emphasize, is valid under the assumptions of Lemma 16 and with a probability of at least $1 - 2\rho$. This completes the proof of Theorem 4.

K   Proof of Lemma 16

The proof strategy is identical to that in Section B. We will prove that (70) holds for every $z \in \mathcal{M}_{\epsilon/40}$, with high probability and provided that M is large enough. As before, this is achieved by finding an upper bound on

$$
P\left\{ \sup_{y \in U(\mathcal{M}_{\epsilon/40}, x)} |\|\Phi y\| - 1| > \epsilon \right\}, \tag{73}
$$

for $\epsilon \leq 1/3$.

We begin again by constructing a sequence of increasingly finer covers for $U(\mathcal{M}_\gamma, x)$, with γ to be set later. We denote this sequence by $\{L_j(\delta)\}$—each $L_j(\delta)$ is a $(2^{-j}\delta)$-net for $U(\mathcal{M}_\gamma, x)$. For $0 < \delta \leq 1/\sqrt{2}$, set $\eta = \delta^2 \tau$ and $\gamma = 4\delta$. We form $\{L_j(\delta)\}$ from $\{C_j(\eta)\}$, the sequence of covers for $\mathcal{M}$ constructed in Section B.1. Indeed, the same argument in that section proves that $U(C_j(\eta), x)$ is a $(2^{-j}\delta)$-cover for $U(\mathcal{M}_\gamma, x)$. It also holds that

$$
\#L_j(\delta) \leq \#C_j(\eta) \leq 4^{jK} \left(\frac{2}{\theta(\delta^2/4)\delta^2 \tau} \right)^K \frac{V_\mathcal{M}}{V_{\mathcal{B}_K}} \leq 4^{jK} \left(\frac{\sqrt{K}}{\delta^2 \tau} \right)^K V_\mathcal{M} =: l_j(\delta). \tag{74}
$$

As before, we can represent every $y \in U(\mathcal{M}_\gamma, x)$ with an infinite chain of points from the sequence of covers $\{L_j(\delta)\}$. After setting $\delta = \epsilon/160$, using the same argument as the one in Section B.2,

and exploiting the estimates above, one can verify that the failure probability in (73) is at most ρ, provided that

$$M \geq 18\epsilon^{-2} \max\left(11K + K\log\left(\frac{\sqrt{K}}{\tau\epsilon^2}\right) + \log V_{\mathcal{M}}, \log\left(\frac{8}{\rho}\right)\right).$$

L   Proof of Theorem 5

Using the triangle inequality and (20), we have

$$\|\widehat{x} - x^*\| \leq \|x - \widehat{x}\| + \|x - x^*\| \leq 2(1 + \epsilon)\|x - x^*\| + (2 + \epsilon)\|n\| + \frac{\epsilon\tau}{40}. \tag{75}$$

Now, since both $\widehat{x}$ and x^* belong to $\mathcal{M}$, we can invoke Lemma 6 from the Toolbox, which states that if $\|\widehat{x} - x^*\| \leq \tau/2$, then

$$d_{\mathcal{M}}(\widehat{x}, x^*) \leq \tau - \tau\sqrt{1 - 2\|\widehat{x} - x^*\|/\tau}. \tag{76}$$

To apply this lemma, it is sufficient to know that

$$2(1 + \epsilon)\|x - x^*\| + (2 + \epsilon)\|n\| + \frac{\epsilon\tau}{40} \leq \tau/2,$$

i.e., that

$$\|x - x^*\| + \frac{2 + \epsilon}{2(1 + \epsilon)}\|n\| \leq \frac{\tau}{4}\left(\frac{1 - \epsilon/20}{1 + \epsilon}\right).$$

For the sake of neatness, we may tighten this condition to $\|x - x^*\| + \frac{7}{8}\|n\| \leq 0.184\tau$, which implies the sufficient condition above (since $\epsilon \leq 1/3$). Thus, if $\|x - x^*\|$ and $\|n\|$ are sufficiently small (on the order of the condition number τ), then we may combine (75) and (76), giving

$$\begin{aligned}
d_{\mathcal{M}}(\widehat{x}, x^*) &\leq \tau - \tau\sqrt{1 - \frac{2}{\tau}\left(2(1 + \epsilon)\|x - x^*\| + (2 + \epsilon)\|n\| + \frac{\epsilon\tau}{40}\right)} \\
&= \tau - \tau\sqrt{1 - \left(\frac{4 + 4\epsilon}{\tau}\|x - x^*\| + \frac{4 + 2\epsilon}{\tau}\|n\| + \frac{\epsilon}{20}\right)}.
\end{aligned} \tag{77}$$

Under the assumption that $\|x - x^*\| + \frac{7}{8}\|n\| \leq 0.184\tau$, it follows that the term inside the parenthesis in the last line above does not exceed one and therefore,

$$d_{\mathcal{M}}(\widehat{x}, x^*) \leq (4 + 4\epsilon)\|x - x^*\| + (4 + 2\epsilon)\|n\| + \frac{\epsilon\tau}{20}.$$

This completes the proof of Theorem 5.

References

[1] D. Achlioptas. Database-friendly random projections. In *Proc. Symp. on Principles of Database Systems (PODS '01)*, pages 274–281. ACM, 2001.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the Restricted Isometry Property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[3] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Found. Comput. Math.*, 9(1):51–77, 2009.

[4] D. S. Broomhead and M. J. Kirby. The Whitney Reduction Network: A method for computing autoassociative graphs. *Neural Comput.*, 13(11):2595–2616, November 2001.

[5] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, February 2006.

[6] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, August 2006.

[7] E. Candès and T. Tao. Decoding via linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, December 2005.

[8] E. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, December 2006.

[9] E. J. Candès. The Restricted Isometry Property and its implications for compressed sensing. *Compte Rendus de l'Academie des Sciences, Paris*, 346:589–592, 2008.

[10] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Proc. Mag.*, 25(2):21–30, 2008.

[11] M. Chen, J. Silva, J. Paisley, C. Wang, D. Dunson, and L. Carin. Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds. *IEEE Trans. Signal Proc.*, 58(12):6140–6155, 2010.

[12] K. L. Clarkson. Tighter bounds for random projections of manifolds. In *Proc. Symp. Comput. Geom.*, pages 39–48. ACM, 2008.

[13] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best *k*-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.

[14] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk. The smashed filter for compressive classification and target recognition. In *Proc. Comput. Imaging V at SPIE Electronic Imaging*, January 2007.

[15] M. A. Davenport, C. Hegde, M. F. Duarte, and R. G. Baraniuk. Joint manifolds for data fusion. *IEEE Trans. Image Proc.*, 19(10):2580–2594, 2010.

[16] M. A. Davenport and M. B. Wakin. Analysis of orthogonal matching pursuit using the restricted isometry property. *IEEE Trans. Inform. Theory*, 56(9):4395–4401, 2010.

[17] R. DeVore, G. Petrova, and P. Wojtaszczyk. Instance-optimality in probability with an ℓ_1-minimization decoder. *Appl. Comput. Harmonic Anal.*, 27(3):275–288, 2009.

[18] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4), April 2006.

[19] D. Donoho and Y. Tsaig. Extensions of compressed sensing. *Signal Proc.*, 86(3):533–548, March 2006.

[20] D. L. Donoho and C. Grimes. Image manifolds which are isometric to Euclidean space. *J. Math. Imaging Comp. Vision*, 23(1):5–24, July 2005.

[21] M. Duarte, M. Davenport, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk. Multiscale random projections for compressive classification. In *Proc. IEEE Conf. Image Proc. (ICIP)*, September 2007.

[22] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Proc. Mag.*, 25(2):83–91, 2008.

[23] M. F. Duarte, M. B. Wakin, D. Baron, S. Sarvotham, and R. G. Baraniuk. Measurement bounds for sparse signal ensembles via graphical models. *IEEE Trans. Inform. Theory*, 2013. To appear.

[24] A. Eftekhari, J. Romberg, and M. B. Wakin. Matched filtering from limited frequency samples. *IEEE Trans. Inform. Theory*, 59(6):3475–3496, June 2013.

[25] A. Eftekhari, H. L. Yap, C. J. Rozell, and M. B. Wakin. The Restricted Isometry Property for random block diagonal matrices. *arXiv preprint arXiv:1210.3395*, 2012.

[26] H. Federer. Curvature measures. *Trans. Amer. Math. Soc*, 93(3):418–491, 1959.

[27] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss. Improved time bounds for near-optimal sparse Fourier representations. In *Proc. SPIE Wavelets XI*, 2005.

[28] A. C. Gilbert, M. J. Strauss, and J. A. Tropp. A tutorial on fast Fourier sampling. *IEEE Signal Proc. Mag.*, 25(2):57–66, 2008.

[29] A.C. Gilbert, M.J. Strauss, J.A. Tropp, and R. Vershynin. One sketch for all: Fast algorithms for compressed sensing. In *Proc. Symp. Theory Comput. (STOC)*, 2007.

[30] D. Healy and D. J. Brady. Compression at the physical interface. *IEEE Signal Proc. Mag.*, 25(2):67–71, 2008.

[31] C. Hegde and R. G. Baraniuk. Signal recovery on incoherent manifolds. *IEEE Trans. Inform. Theory*, 58(12):7204–7214, 2012.

[32] C. Hegde, M. B. Wakin, and R. G. Baraniuk. Random projections for manifold learning. In *Proc. Neural Inform. Proc. Sys. (NIPS)*, December 2007.

[33] G. E. Hinton, P. Dayan, and M. Revow. Modeling the manifolds of images of handwritten digits. *IEEE Trans. Neural Networks*, 8(1):65–74, January 1997.

[34] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimenstionality. In *Proc. Symp. Theory Comput. (STOC)*, 1998.

[35] M. A. Iwen and M. Maggioni. Approximation of points on low-dimensional manifolds via random linear projections. *Information and Inference*, 2013.

[36] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk. Analog-to-information conversion via random demodulation. In *Proc. IEEE Dallas Circuits Systems Workshop (DCAS)*, 2006.

[37] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the Restricted Isometry Property. *arXiv preprint arXiv:1207.0235*, 2012.

[38] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Stat.*, 28(5):1302–1338, 2000.

[39] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Ergebnisse der Mathematik Und Ihrer Grenzgebiete. Springer, 2011.

[40] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Proc. Mag.*, 25(2):72–82, 2008.

[41] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmonic Anal.*, 26(3):301–321, 2009.

[42] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J. Sel. Topics Signal Proc.*, 4(2):310–316, 2010.

[43] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1):419–441, 2008.

[44] F. W. J. Olver and National Institute of Standards and Technology (U.S.). *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

[45] J. Y. Park and M. B. Wakin. A geometric approach to multi-view compressive imaging. *EURASIP J. Advances Signal Proc.*, 2012(1):37, 2012.

[46] D. Ramasamy, S. Venkateswaran, and U. Madhow. Compressive estimation in AWGN: General observations and a case study. In *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2012.

[47] H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS'09*, Saint-Malo, France, 2009.

[48] H. Rauhut. Compressive sensing and structured random matrices. In M. Fornasier, editor, *Theoretical foundations and numerical methods for sparse recovery*, pages 1–92. De Gruyter, 2010.

[49] H. Rauhut, J. Romberg, and J. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmonic Anal.*, 32(2):242–254, 2012.

[50] J. Romberg. A uniform uncertainty principle for Gaussian circulant matrices. In *Proc. Int. Conf. Digital Signal Proc.*, 2009.

[51] J. Romberg and R. Neelamani. Sparse channel separation using random probes. *Inverse Problems*, 26(11):115015, 2010.

[52] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.

[53] P. Shah and V. Chandrasekaran. Iterative projections for signal identification on manifolds: Global recovery guarantees. In *Allerton Conf. Comm., Control, Comput. (Allerton)*, pages 760–767, 2011.

[54] G. W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev.*, 15(4):727–764, 1973.

[55] M. Talagrand. *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer Monographs in Mathematics. Springer, 2005.

[56] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk. Beyond Nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Trans. Inform. Theory*, 56(1):520–544, 2010.

[57] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cogn. Neurosci.*, 3(1):71–83, 1991.

[58] N. Verma. Distance preserving embeddings for general n-dimensional manifolds. 2012. Preprint.

[59] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y.C. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[60] M. B. Wakin. *The Geometry of Low-Dimensional Signal Models*. PhD thesis, Department of Electrical and Computer Engineering, Rice University, Houston, TX, 2006.

[61] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk. The multiscale structure of non-differentiable image manifolds. In *Proc. Wavelets XI at SPIE Optics and Photonics*, August 2005.

[62] H. L. Yap, M. B. Wakin, and C.J. Rozell. Stable manifold embeddings with structured random matrices. *IEEE J. Sel. Topics Signal Proc.*, 2013. To appear.