APPROXIMATION OF POINTS ON LOW-DIMENSIONAL MANIFOLDS VIA RANDOM LINEAR PROJECTIONS

MARK A. IWEN MAURO MAGGIONI
DUKE UNIVERSITY, BOX 90320
DURHAM, NC 27708-0320

ABSTRACT. This paper considers the approximate reconstruction of points, $\vec{x} \in \mathbb{R}^D$, which are close to a given compact d-dimensional submanifold, $\mathcal{M}$, of $\mathbb{R}^D$ using a small number of linear measurements of $\vec{x}$. In particular, it is shown that a number of measurements of $\vec{x}$ which is independent of the extrinsic dimension, D, suffices for highly accurate reconstruction of $\vec{x}$. Furthermore, it is also proven that all vectors, $\vec{x}$, which are sufficiently close to $\mathcal{M}$ can be reconstructed with uniform approximation guarantees when the number of linear measurements of $\vec{x}$ depends logarithmically on D. Finally, the proofs of these facts are constructive: A practical algorithm for manifold-based signal recovery is presented in the process of proving the two main results mentioned above.

1. INTRODUCTION

In this paper we present a simple reconstruction technique which facilitates compressive sensing for general classes of high-dimensional signals with low intrinsic dimension. Two types of models are often considered: sparse models and low-dimensional/manifold models. The former type of model assumes that each data point has a sparse representation in terms of a (typically known) dictionary Φ, which geometrically means that data points lie on unions of a small number of planes spanned by the elements of the dictionary [27]. The latter type of model assumes that data possesses an intrinsically low-dimensional geometrical structure, for example that of a manifold (see e.g. [46, 7, 22, 26], among many others) or a union of planes (see e.g. [48, 18, 36, 31]), motivated by many applications, for example in image processing [32], computer vision [45], and pattern recognition [36].

Given the low-intrinsic dimension of these models, it is natural to ask whether a small number of linear projections ("measurements") of a data point, together with knowledge of the low-dimensional model, suffices to encode and reconstruct a data point. In the setting of sparsity, compressed sensing [24, 42] not only says that, under suitable assumptions [27], this is indeed possible, but a convex optimization problem leads to the stable recovery of the original data point. In the setting where data lies on a low-dimensional manifold, the work of Wakin et al. [6, 49] on manifold-based signal recovery shows that low-dimensional (random) projections provide small distortion embeddings for manifolds, but leave open the question of reconstructing a data point.

Standard compressed sensing [24, 42] deals with the approximation of vectors, $\vec{x} \in \mathbb{R}^D$, which can be sparsely represented in terms of a given $D \times n$ dictionary matrix, Φ. Note that such Φ-sparse vectors can be compactly stored in a compressed form which is easy to transmit and store. Moreover, they can be recovered from their compressed representations when necessary. This compression/recovery problem has been well studied when $\vec{x}$ is available in its entirety before compression (see, e.g., [37]). However, in situations where $\vec{x}$ is costly to observe one may only have the ability to collect a very small set of measurements of $\vec{x}$ to begin with, thus making standard compression techniques inapplicable (see, e.g., [2, 1] and references therein). This is the *compressed sensing regime*, where lossless compression must occur *before* one determines which vector components or transform coefficients are actually important. Hence, the goal of standard compressed sensing becomes to design an $m \times D$ measurement matrix M, with m as small as absolutely possible, subject to the constraint that a computationally efficient reconstruction algorithm, $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^D$, exists such that $\mathcal{A}(M\vec{x}) \approx \vec{x}$ anytime $\vec{x} \in \mathbb{R}^D$ is sufficiently compressible with respect to a given dictionary matrix Φ.

More precisely, given an integer $d \ll n$ suppose that

$$\vec{x} = \Phi\left(\vec{f_d} + \vec{\epsilon}\right)$$

This research was partially supported by ONR N00014-07-1-0625, The Alfred P. Sloan Foundation, DARPA N66001-11-1-4002, NSF DMS-1045153, and NSF DMS-0847388.

where $\vec{\epsilon} \in \mathbb{R}^n$ is in the row space of $\Phi \in \mathbb{R}^{D\times n}$ and

$$\vec{f_d} \;=\; \underset{\vec{y}\in\mathbb{R}^n \text{ with } \|\vec{y}\|_0\leq d}{\arg\min} \left\|\vec{x} - \Phi\vec{y}\right\|.$$

The goal of a compressed sensing method is to approximate $\vec{x}$ as well as possible by approximating the at most d nonzero elements of $\vec{f_d} \in \mathbb{R}^n$. Furthermore, compressed sensing techniques aim to accomplish this task using as few linear measurements of $\vec{x} \in \mathbb{R}^D$,

$$\left\langle \vec{m}_j, \vec{x} \right\rangle = \left\langle \vec{m}_j, \Phi\vec{f_d} \right\rangle + \left\langle \vec{m}_j, \Phi\vec{\epsilon} \right\rangle, \tag{1}$$

as absolutely possible.

Let $M \in \mathbb{R}^{m\times D}$ be the matrix whose j^{th}-row is the measurement vector $\vec{m}_j \in \mathbb{R}^D$ from Equation 1 above. A compressed sensing method consists of both a choice of $M \in \mathbb{R}^{m\times D}$, and a recovery algorithm, $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^D$, such that

$$\left\|\vec{x} - \mathcal{A}\left(M\vec{x}\right)\right\|_p \;=\; \left\|\vec{x} - \mathcal{A}\left(M\Phi\vec{f_d} + M\Phi\vec{\epsilon}\right)\right\|_p \;\leq\; C_{p,q}\cdot d^{\frac{1}{p}-\frac{1}{q}} \left\|\,\vec{\epsilon}\,\right\|_q \tag{2}$$

in fixed ℓ_p,ℓ_q norms, $1 \leq q \leq p \leq 2$, for an absolute constant $C_{p,q} \in \mathbb{R}$. Note that $M \in \mathbb{R}^{m\times D}$ forms a compressed representation of $\vec{x} \in \mathbb{R}^D$ whenever $m < D$, which is then stably inverted by $\mathcal{A}$. Many recovery algorithms, $\mathcal{A}$, have been developed for solving this problem when Φ is a square $D \times (n = D)$ orthonormal matrix, and $M\Phi$ has either restricted isometry [14] or incoherence [25, 30] properties (e.g., see [13, 11, 12, 47, 39, 40, 38, 9]). Perhaps the best such results are achieved by $(m = O(d\log(D/d))) \times D$ measurement matrices, M, whose entries are independent and identically distributed standard Gaussian random variables. These Gaussian matrices allow for near optimal compression (i.e., a near minimal size for m) while still allowing for the existence of recovery algorithms, $\mathcal{A}$, which achieve Equation 2 for an arbitrarily given square orthonormal matrix Φ. Furthermore, if M is Gaussian then Φ need not be known when the measurements, $M\vec{x}$, are computed: It suffices to know Φ only during reconstruction with $\mathcal{A}$.

One strand of work in compressed sensing has dealt with extending the results mentioned above concerning square orthonormal matrices to include settings where Φ is a more general (i.e., rectangular) $D \times n$ matrix. The first of these results extended compressed sensing to include $D\times n$ dictionaries, Φ, whose columns are all nearly pairwise orthogonal [43]. This work shares all of the advantages of the aforementioned results concerning compressed sensing when Φ is square orthonormal matrix (e.g., nearly orthogonal Φ also do not need to be known until reconstruction via $\mathcal{A}$) when M is a random matrix exhibiting concentration of measure properties (e.g., if M is Gaussian as above). These results were later generalized further to allow recovery along the lines of Equation 2 when Φ has columns with less limited forms of coherence and redundancy [15] (e.g., if Φ is a tight frame). In this paper we consider a geometric variant of standard compressed sensing results for signals which are sparsely representable with respect to a square orthonormal matrix, Φ, by focussing instead on signals which are well represented by manifold models. More specifically, herein the $D \times n$ dictionary matrix Φ utilized in standard compressed sensing models will be replaced by a piecewise linear approximation to a given submanifold of $\mathbb{R}^D$.

The work herein utilizes ideas introduced by Baranuik and Wakin which demonstrate the existence of simple linear operators capable of (nearly) isometrically embedding a given compact d-dimensional submanifold of $\mathbb{R}^D$ into $\mathbb{R}^{O(d\log D)}$ without utilizing detailed knowledge regarding the submanifold's structure [6]. In some sense, this work immediately yields measurement matrices, $M \in \mathbb{R}^{m\times D}$, for manifold-based compressed sensing. However, a complete compressed sensing strategy also requires an associated reconstruction algorithm, $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^D$, capable of accurately approximating points near the given manifold in a computationally efficient fashion. Algorithms of this kind were first considered by Wakin in [49]. Therein, Wakin showed that approximating a given point, $\vec{x}$, near a compact d-dimensional submanifold of $\mathbb{R}^D$ via an $O(d\log D)$ linear measurements (i.e., see Equation 1) was possible with high probability if the measurements were randomly regenerated for each new $\vec{x}$. Furthermore, [49] concluded that achieving strong reconstruction guarantees using one fixed set of linear measurements for all possible points, $\vec{x}$, near a given compact submanifold of $\mathbb{R}^D$ was difficult. However, it is important to mention that the results presented in [49] were derived independently of any particular numerical reconstruction algorithm, $\mathcal{A}$. As a consequence, this line of work did not result an implementable recovery algorithm with accompanying approximation guarantees.

In this paper we propose a computationally efficient reconstruction algorithm for manifold-based compressed sensing and prove accompanying approximation guarantees. In the process, we prove that a given point, $\vec{x}$, near a compact d-dimensional submanifold of $\mathbb{R}^D$ can be accurately approximated using $O(d\log d)$ linear measurements with high probability when the measurements are randomly regenerated for each new $\vec{x}$. This improves on previous results [49]

by removing all dependence on the extrinsic dimension of the submanifold, D, from the number of linear measurements required for accurate approximation. Furthermore, we provide stability guarantees for the algorithm when one fixed set of $O(d \log D)$ linear measurements are used for all possible points, $\vec{x}$, near a given compact submanifold of $\mathbb{R}^D$. Finally, an empirical evaluation of our method indicates that it also works well in practice.

Before moving on to discuss our methods and results in more detail we hasten to add that other techniques have also been proposed for manifold-based compressed sensing since the initial work of Baranuik and Wakin. Perhaps most notable among these are the statistical methods proposed by Chen et al. [19]. Chen et al. use training data from a compact d-dimensional submanifold of $\mathbb{R}^D$ in order to estimate the manifold data's distribution via a Gaussian mixture model composed of Gaussians whose covariance matrices are all rank $O(d)$. They then use the probability density resulting from their low-rank Gaussian mixture model to approximate points on the manifold, $\vec{x}$, with a maximum likelihood estimator when given only linear measurements, $M\vec{x} \in \mathbb{R}^m$. In contrast, we utilize geometric and analytic techniques herein and make no attempt to estimate the statistical properties of any observed manifold data.

Other related manifold-based compressed sensing methods include those recently proposed by Shah et al. [44]. The method in [44] assumes that a black-box manifold projector, which takes an input point and then outputs its nearest neighbor on a given submanifold of $\mathbb{R}^D$, is given. Given access to such a projector, the authors present an iterative projection method for approximately recovering a point, $\vec{x}$, on the given manifold using compressive linear measurements $M\vec{x} \in \mathbb{R}^m$. Similar to the recovery algorithm presented herein, the recovery algorithm in [44] is shown to be accurate whenever (*i*) the given black-box manifold projector is accurate, and (*ii*) the measurement matrix M (nearly) isometrically embeds the given manifold into $\mathbb{R}^m$. However, unlike [44], we analyze (approximate) manifold projectors that can be explicitly and efficiently computed using noisy training data sampled from a general manifold of interest (e.g., the manifold projectors resulting from Geometric Multi-Resolution Analysis [3]). Thus, we obtain implementable recovery algorithms that are applicable to a general class of smooth and compact submanifolds of $\mathbb{R}^D$. Furthermore, we are able to improve the computational efficiency of our recovery algorithm by calculating our manifold projections using computations that are performed in the smaller embedding space, $\mathbb{R}^m$, whenever possible.

1.1. **Methods and Results.** As discussed above, the standard compressed sensing setup assumes that the signal to be approximated has a compressible representation with respect to an orthonormal basis (or frame [15], or incoherent dictionary [43]). Although this is certainly a useful setting, there are many applications where signals might be better approximated via more geometrical considerations. For example, consider the setting where the class of potential input signals varies continuously as a function of a small number of parameters (e.g., see [50, 6, 49]). In this case it makes more sense to consider the approximate reconstruction of signals, $\vec{x} \in \mathbb{R}^D$, which are close to a given compact d-dimensional submanifold, $\mathcal{M}$, of $\mathbb{R}^D$. The optimal approximation for $\vec{x} \in \mathbb{R}^D$ is then defined to be

$$\vec{x}_{\text{opt}} = \arg\min_{\vec{y} \in \mathcal{M}} \left\| \vec{x} - \vec{y} \right\|_2 .$$

In effect, $\vec{x}_{\text{opt}}$ is the best approximation to $\vec{x}$ on $\mathcal{M}$. Our objective is to approximate $\vec{x}_{\text{opt}} \in \mathcal{M} \subset \mathbb{R}^D$ given only a small number of linear measurements, $M\vec{x} \in \mathbb{R}^m$, where M is an $m \times D$ measurement matrix as above. Hence, in this paper we seek to design a measurement matrix $M \in \mathbb{R}^{m \times D}$ with m as small as absolutely possible, together with a computationally efficient reconstruction algorithm $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^D$, so that $\mathcal{A}(M\vec{x}) \approx \vec{x}$ whenever $\vec{x} \in \mathbb{R}^D$ is sufficiently close to a given compact d-dimensional submanifold of $\mathbb{R}^D$, $\mathcal{M}$.[1]

Note that a manifold, $\mathcal{M}$, is now taking the place of the dictionary matrix, $\Phi \in \mathbb{R}^{D \times n}$, in the standard compressed sensing setup discussed above. Of course, it is unreasonable to expect that we can always have an exact representation of the signal manifold at our disposal. Instead, we assume that we have a set of locally linear approximations to the given manifold which capture the local geometric structure of the manifold's tangent spaces. In fact, such piecewise linear manifold representations are exactly the type of approximations produced by existing manifold learning algorithms like LTSA [52] and Geometric Multi-Resolution Analysis [3]. Thus, we assume that the signal manifold, $\mathcal{M}$, is approximated by such a method at some point. However, as in standard compressed sensing methods, the manifold-based compressed sensing strategies developed below do not require that these piecewise linear manifold representations are known when the compressed measurements, $M\vec{x} \in \mathbb{R}^m$, are collected. Approximation of the signal manifold can be put off until later when signal reconstruction takes place (i.e., one does not need a piecewise linear manifold approximation until $\mathcal{A}(M\vec{x})$ is actually computed).

[1] Put another way, we require that $\mathcal{A}(M\vec{x}) \approx \vec{x}_{\text{opt}}$ which implies that $\mathcal{A}(M\vec{x}) \approx \vec{x}$ whenever $\vec{x} \approx \vec{x}_{\text{opt}}$.

Although the manifold-based compressed sensing methods developed herein will work with any locally linear approximation to the given signal manifold, $\mathcal{M}$, we will focus on *multiscale* piecewise linear manifold approximations to $\mathcal{M}$ in particular. As opposed to fixed-scale locally linear approximations, multiscale representations better approximate non-smooth manifolds, and manifolds contaminated with noise [17, 3]. For example, multiscale locally linear approximation is particularly beneficial for signal processing tasks involving image manifolds, which tend to be non-differentiable in many realistic settings [50]. Hence, we formulate our compressed sensing methods below with respect to general multiscale piecewise linear manifold approximations of the type produced by Geometric Multi-Resolution Analysis (GMRA) [3].

As mentioned above, the manifold embeddings of Baraniuk and Wakin [6, 49] can be considered as manifold-based compressed sensing matrices, for which however no associated recovery algorithms were explicitly defined. Indeed, the measurement matrices, $M \in \mathbb{R}^{m \times D}$, used in the manifold-based compressed sensing methods developed below are modifications of their embedding matrices. However, unlike the embedding matrices considered in [6], the measurement matrices considered herein (nearly) isometrically embed *both* the underlying signal manifold, $\mathcal{M}$, *and* the multiscale piecewise linear approximation to $\mathcal{M}$ into $\mathbb{R}^m$ in a way which preserves the fidelity of the embedded multiscale locally linear approximation to the embedded image of $\mathcal{M}$. Accomplishing this requires us to reengineer the arguments from [6] using Johnson-Lindenstrauss embedding [34] techniques similar to those utilized in [5]. The resulting measurement matrices, M, ultimately justify this complication by allowing us to develop reconstruction algorithms which work exclusively with locally linear approximations to $\mathcal{M}$ while still preserving approximation accuracy with respect to the true manifold, $\mathcal{M}$.

The reconstruction algorithm, $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^D$, proposed below consists of two well-studied computational subroutines: a method for solving approximate nearest neighbor problems (e.g., [33, 8, 4]) in a space of dimension comparable to the intrinsic dimension of the data, and a method for solving an overdetermined least squares problem (e.g., via the singular value decomposition of the associated matrix). The algorithm works by first using the compressed measurements, $M\vec{x}$, of $\vec{x}$ to locate the best local linear approximation to $\mathcal{M}$ at $\vec{x}$. This is accomplished by running a nearest neighbor algorithm on a set of "center points" from near the manifold, each of which represents a particular linear approximation to $\mathcal{M}$ in a neighborhood of the center point. Because $\mathcal{M}$ has low intrinsic dimension, and the center points are arranged in a multiscale hierarchy as per [3], this search can be carried out relatively quickly. To finish, the algorithm then approximates $\vec{x}_{\text{opt}}$, the best approximation to $\vec{x}$ on $\mathcal{M}$, by solving an overdetermined least squares problem using the linear approximation to the manifold located in the first step.

In this paper we prove two compressed sensing results for the proposed reconstruction algorithm, each of which utilizes randomly generated measurement matrices, $M \in \mathbb{R}^{m \times D}$, satisfying a different set of properties. Roughly speaking, the first result indicates that $m = O(d \log(d/\delta))$ linear measurements of a given $\vec{x} \in \mathbb{R}^D$ suffice to create a compact representation, $M\vec{x} \in \mathbb{R}^m$, from which the reconstruction algorithm, $\mathcal{A}$, discussed above will recover an approximation to $\vec{x}_{\text{opt}} \in \mathcal{M}$ satisfying

$$\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\| \;<\; C\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \delta.$$

Here $C \in \mathbb{R}^+$ represents a fixed universal constant, $\delta \in \mathbb{R}^+$ can be freely chosen, and $\mathcal{M}$ is the given d-dimensional submanifold of $\mathbb{R}^D$. This result provides what is commonly referred to as a *nonuniform recovery result*, by which we mean that the upper bound on $\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\|$ holds with high probability *for each* $\vec{x} \in \mathbb{R}^D$ over the choice of random measurement matrix.

The second theorem proven below provides a type of *uniform recovery result* which holds with high probability *for all* vectors, $\vec{x} \in \mathbb{R}^D$, of a particular class. Simply put, it asserts the existence of a D-dimensional tube around the given manifold, $T \supset \mathcal{M}$, within which accurate approximation will always take place with high probability over the choice of random measurement matrix $M \in \mathbb{R}^{m \times D}$. More specifically, the second theorem says that $m = O(d \log(D/\delta))$ linear measurements of any $\vec{x} \in T \subset \mathbb{R}^D$ suffice to create a compact representation, $M\vec{x} \in \mathbb{R}^m$, from which the reconstruction algorithm discussed above, $\mathcal{A}$, will recover an approximation to $\vec{x}_{\text{opt}} \in \mathcal{M}$ satisfying

$$\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\| \;<\; C\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|_2 + \frac{C}{\sqrt{d}}\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|_1 + \delta.$$

Here, as above, $C \in \mathbb{R}^+$ represents a fixed universal constant and $\delta \in \mathbb{R}^+$ can be freely chosen.

The reminder of this paper is organized as follows: In the next section we begin by fixing terminology and reviewing relevant definitions and theorems. Having established the necessary notation, we then give precise statements of the two main results proven in this paper in Section 2.4. Finally, in Section 2.5, we conclude Section 2 with a discussion

of the different types of measurement matrices, $M \in \mathbb{R}^{m\times D}$, associated with each of our two main results. In Section 3 the recovery algorithm, $\mathcal{A}$, is presented and analyzed. In particular, the approximation error of $\mathcal{A}$ for a given $\vec{x}$, $\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\|$, is bounded for each of the two possible types of measurement matrices, M, considered herein. The runtime complexity of $\mathcal{A}$ is also determined. Next, in Section 4, the number of rows, m, required for each type of measurement matrix defined in Section 2.5 is upper bounded. This formally establishes the amount of compression possible in our manifold-based compressed sensing schemes. To finish, the compressed sensing methods developed herein are evaluated empirically in Section 5.

2. Notation and Setup

Given $n \in \mathbb{N}$ we will define $[n]$ to be the set $\{0, 1, 2, \ldots, n\} \subset \mathbb{Z}$. All norms, $\|\cdot\|$, will refer to the standard Euclidean norm unless otherwise stated. We will denote an open ball of radius $\delta \in \mathbb{R}^+$ centered at $\vec{y} \in \mathbb{R}^D$ by $\mathcal{B}_\delta(\vec{y})$. Our real valued $m \times D$ measurement matrix will always be denoted by M. Furthermore, M will always be linear Johnson-Lindenstrauss embedding [34, 29, 23, 35] of a finite set $S \subset \mathbb{R}^D$ into $\mathbb{R}^m$.

Definition 1. *Let $\epsilon \in (0, 1/2)$, and $S \subset \mathbb{R}^D$ be finite. An $m\times D$ matrix M is a linear Johnson-Lindenstrauss embedding of S into $\mathbb{R}^m$ if*

$$(1-\epsilon)\|\vec{u} - \vec{v}\|^2 \leq \|M\vec{u} - M\vec{v}\|^2 \leq (1+\epsilon)\|\vec{u} - \vec{v}\|^2$$

for all $\vec{u}, \vec{v} \in S$. In this case we will say that M embeds S into $\mathbb{R}^m$ with ϵ-distortion.

The following theorem is proven by showing that an $m\times D$ matrix with randomized entries will satisfy Definition 1 for a given set $S \subset \mathbb{R}^D$ with high probability whenever m is sufficiently large (e.g., see [23]).

Theorem 1. ***(See** [34, 23]**.)** Let $\epsilon \in (0, 1/2)$, and $S \subset \mathbb{R}^D$ be finite. Let $m = O(\epsilon^{-2}\log|S|)$ be a natural number. Then, there exists an $m \times D$ linear Johnson-Lindenstrauss embedding of S into $\mathbb{R}^m$ with ϵ-distortion.*

For the remainder of this paper $\mathcal{M}$ will denote a compact d-dimensional submanifold of $\mathbb{R}^D$ with d-dimensional volume V. We will characterize results concerning any such manifold $\mathcal{M}$ via its reach [28], denoted $\text{reach}(\mathcal{M})$, which is defined as follows: Let

$$D(\mathcal{M}) = \left\{\vec{x} \in \mathbb{R}^D \;\middle|\; \exists \text{ a unique } \vec{y} \in \mathcal{M} \text{ with } \|\vec{x} - \vec{y}\| = d(\vec{x}, \mathcal{M})\right\}$$

and

$$\text{tube}_r(\mathcal{M}) = \left\{\vec{x} \in \mathbb{R}^D \;\middle|\; d(\vec{x}, \mathcal{M}) < r\right\},$$

where $d(\vec{x}, \mathcal{M})$ is the standard Hausdorff distance. We then define

$$\text{reach}(\mathcal{M}) = \sup\{r \geq 0 \mid \text{tube}_r(\mathcal{M}) \subset D(\mathcal{M})\}. \tag{3}$$

Intuitively, $\text{reach}(\mathcal{M})$ is the radius of the largest possible non-self-intersecting tube around $\mathcal{M}$. For example, if $\mathcal{M}$ is a d-sphere of radius r, then $\text{reach}(\mathcal{M}) = r$. More generally, any compact and non-self-intersecting smooth submanifold of $\mathbb{R}^D$ will have nonzero reach. The reach of a manifold is particularly useful because it allows the development of concise bounds for many manifold properties of interest (e.g., curvature, self-avoidance, packing numbers, etcetera). See [28, 41, 6, 20] for more details.

Given a compact set $\mathcal{S} \subset \mathbb{R}^D$ we define a *δ-cover of $\mathcal{S}$* to be any finite set $S \subset \mathbb{R}^D$ with the following property:

$$\forall \vec{x} \in \mathcal{S},\ \exists \vec{y} \in S \text{ such that } \vec{x} \in \mathcal{B}_\delta(\vec{y}).$$

We will refer to a δ-cover of $\mathcal{S}$, S, as *minimal* if $|S| \leq |\tilde{S}|$ for all other δ-covers of $\mathcal{S}$, $\tilde{S}$. Hereafter, $C_\delta(\mathcal{S})$ will denote a minimal δ-cover of a given compact set $\mathcal{S}$ in $\mathbb{R}^D$. The following lemma, easily proven using results from [41], bounds $|C_\delta(\mathcal{M})|$ for any compact d-dimensional Riemannian manifold, $\mathcal{M}$, in terms of δ and $\text{reach}(\mathcal{M})$.

Lemma 1. ***(See** [41]**.)** Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact d-dimensional Riemannian manifold with d-dimensional volume V, and suppose that $\delta \in \mathbb{R}^+$ is less than $\text{reach}(\mathcal{M})$. Then, any minimal δ-cover of $\mathcal{M}$, $C_\delta(\mathcal{M})$, will have*

$$\left|C_\delta(\mathcal{M})\right| < \frac{V\left(\frac{d}{2}+1\right)^{\frac{d}{2}+1}}{2^{\frac{d}{2}}\delta^d}.$$

We will now briefly summarize one means of constructing the type of piecewise linear manifold approximations required by our manifold-based compressed sensing algorithm.

2.1. **Geometric Multi-Resolution Analysis (GMRA).** The manifold-based compressed sensing strategies developed herein will utilize manifold-approximations of the type produced by the GMRA framework [3]. GMRA is a dictionary learning technique that aims to concisely summarize data sets in $\mathbb{R}^D$ in a multiscale fashion. When a given data set lies near a low dimensional smooth and compact submanifold of $\mathbb{R}^D$, GMRA produces small dictionaries with accompanying guarantees on the dictionary size/approximation error trade-off. GMRA also has computationally efficient algorithms for both sparsifying, and subsequently reconstructing, data points via its multiscale dictionary.

The construction of a GMRA dictionary proceeds as follows: First, GMRA produces a set of $J+1$ partitions of the input data set, $X \subset \mathbb{R}^D$, organized in a tree structure. Denote these $J+1$ partitions by

$$C_j := \left\{ C_{j,k} \subset \mathbb{R}^D \;\middle|\; k \in [K_j] \right\},\ j \in [J]$$

where $K_0 = 0$ and $K_1, \ldots, K_J \in \mathbb{N}$. Here j represents the scale of the partition (i.e, there exists a constant C for all $j \in [J]$ and $k \in [K_j]$ such that $C_{j,k} \subset \mathcal{B}_{C\cdot 2^{-j}}\left(\vec{y}_{j,k}\right)$ for some $\vec{y}_{j,k} \in C_{j,k}$). Furthermore, C_{j+1} is a refinement of C_j for each $j \in [J-1]$. Thus, we can organize the partitions in a tree structure in a natural fashion (i.e., each subset $C_{j,k} \subset \mathbb{R}^D$ corresponds to a node in the tree at depth j, the root of the tree corresponding to the coarsest scale $j = 0$). Every $C_{j,k}$-node has a unique *parent node* corresponding to $C_{j-1,k} \supset C_{j,k}$, and, conversely, $C_{j,k} \subset C_{j-1,k}$ implies that the node associated with $C_{j,k}$ is a *child node* of the node associated with $C_{j-1,k}$. Finally, a node with no children is called a *leaf node*. Note that the covering numbers bounded by Lemma 1 above can be used to bound the number of nodes at depth j of our tree when the input data, X, is sampled from a compact d-dimensional submanifold of $\mathbb{R}^D$.

Next, GMRA processes the data points contained in each $C_{j,k}$-subset to produce a local approximation for $X \cap C_{j,k}$. The local approximation associated with each non-root $C_{j,k}$-node of the partition tree consists of:

(1) The mean of the points in $X \cap C_{j,k}$, denoted $c_{j,k}$.
(2) An orthogonal matrix, $\Phi_{j,k}$, whose rows are the top principal components of the data points in $X \cap C_{j,k}$. If, for example, X is sampled from a smooth submanifold of $\mathbb{R}^D$, $\mathcal{M}$, then the rows of $\Phi_{j,k}$ will form a basis for an approximate tangent space to $\mathcal{M}$ at $c_{j,k}$.
(3) A wavelet constant,
$$w_{j,k} = \left(I - \Phi_{j-1,k'}^T \Phi_{j-1,k'}\right) \cdot \left(c_{j,k} - c_{j-1,k'}\right),$$
where the $C_{j-1,k'}$-node is the parent of the $C_{j,k}$-node.
(4) An orthogonal matrix, $\Psi_{j,k}$, whose rows form a basis for the projection of row span$\{\Phi_{j,k}\}$ onto
$$\text{column span}\left\{I - \Phi_{j-1,k'}^T \Phi_{j-1,k'}\right\},$$
where the $C_{j-1,k'}$-node is the parent of the $C_{j,k}$-node. Note that
$$\text{row span}\left\{\Phi_{j,k}\right\} \subseteq \text{column span}\left\{\left[\Phi_{j-1,k'}^T, \Psi_{j,k}^T\right]\right\}.$$

The rows of the orthogonal matrices $\Phi_{j,k}, \Psi_{j,k}$ are called the geometric scaling and wavelet bases, respectively. The information above collectively defines the GMRA dictionary for X. Note that the root node associated with $C_{0,0}$ stores only items (1) and (2) above.

For $j > 0$ we can approximate any $x \in \left(C_{j,k} \cap X\right) \subset C_{j-1,k'}$ at scale j by

$$x_j = \Phi_{j,k}^T \Phi_{j,k}(x - c_{j,k}) + c_{j,k}.$$

Note that the difference between the approximations to x at scales $j-1$ and j can be compactly represented using the geometric wavelet basis/constant. In particular,

$$x_j - x_{j-1} = \left[\Phi_{j-1,k'}^T, \Psi_{j,k}^T\right] \begin{pmatrix} e_{j,k} \\ q_{j,k} \end{pmatrix} + w_{j,k}, \tag{4}$$

where $e_{j,k} = \Phi_{j-1,k'}(x - x_j)$ and $q_{j,k} = \Psi_{j,k}(x_j - c_{j,k})$. Iterating Equation 4 yields a multiscale transform of x in terms of the differences $x_j - x_{j-1}$, each of which belongs to a low-dimensional subspace stored in the GMRA dictionary.

In the next section we will list several properties which GMRA dictionaries are guaranteed to have whenever they are built using a sufficiently dense set of points, X, sampled from a compact and smooth submanifold of $\mathbb{R}^D$. As we shall see, these properties will be critical to the development of our manifold-based compressed sensing scheme. However, it is important to note that *it is only these properties which matter*. That is, the use of *any* piecewise linear manifold approximation satisfying the properties listed in Section 2.2 will suffice for the purposes of proving our main results. Furthermore, it is worth noting that the multiscale structure of the manifold approximations we consider

herein is not strictly necessary (i.e., a sufficiently fine fixed scale approximation will also satisfy Propositions 1 and 2 below). In fact, only the finest scale, J, matters in our main results' approximation guarantees. In principle, all other scales can be ignored. That is, any fixed-scale piecewise linear approximation of a given manifold that satisfies Properties 1, 2b, 3a, and 3b from Section 2.2 (e.g., with $j = J > j_0$) will also satisfy both Propositions 1 and 2 below. Hence, in principle, GMRA can be replaced herein by any other method that (implicitly) computes a piecewise linear approximation to a given manifold (e.g., [10, 52, 16, 19]). As long as the manifold approximation we utilize has the properties we consider in the next section, our main results will still hold as stated.

2.2. **The Manifold Approximation.** In order to help us develop a practical recovery algorithm we will assume we have a multiscale piecewise linear approximation of $\mathcal{M}$ of the type yielded by GMRA. Let $J \in \mathbb{N}$ and $K_0, K_1, \ldots, K_J \in \mathbb{N}$. For each $j \in [J]$ we assume that we have a set of affine projectors,

$$\mathbb{P}_j = \left\{ \mathbb{P}_{j,k} : \mathbb{R}^D \to \mathbb{R}^D \;\middle|\; k \in [K_j] \right\},$$

which approximate $\mathcal{M}$ at scale j. More precisely, these affine projectors will collectively satisfy the three following properties:

(1) **Affine Projections**: Every $\mathbb{P}_{j,k}$ has both an associated vector, $\vec{c}_{j,k} \in \mathbb{R}^D$, and an associated orthogonal $d \times D$ matrix, $\Phi_{j,k}$, so that

$$\mathbb{P}_{j,k}(\vec{x}) = \Phi_{j,k}^{\mathrm{T}} \Phi_{j,k} \left(\vec{x} - \vec{c}_{j,k} \right) + \vec{c}_{j,k}.$$

(2) **Dyadic Structure**: There exist two universal constants, $C_1 \in \mathbb{R}^+$ and $C_2 \in (0, 1]$, so that the following conditions are satisfied:
   (a) $K_j \leq K_{j+1}$ for all $j \in [J-1]$.
   (b) $\|\vec{c}_{j,k_1} - \vec{c}_{j,k_2}\| > C_1 \cdot 2^{-j}$ for all $j \in [J]$ and $k_1, k_2 \in [K_j]$ with $k_1 \neq k_2$. In other words, the $\vec{c}_{j,k}$-vectors at each scale $j \in [J]$ are well separated from one another.
   (c) For each $j \in [J] - \{0\}$ there is exactly one well defined parent function, $p_j : [K_j] \to [K_{j-1}]$, with the property that

$$\left\| \vec{c}_{j,k} - \vec{c}_{j-1,p_j(k)} \right\| < C_2 \min_{k' \in [K_{j-1}] - \{p_j(k)\}} \left\| \vec{c}_{j,k} - \vec{c}_{j-1,k'} \right\|.$$

   Together these J parent functions collectively define a tree structure on the $\vec{c}_{j,k}$-vectors. In particular, each $\vec{c}_{0,k}$ with $k \in [K_0]$ is a root node while each $\vec{c}_{J,k}$ with $k \in [K_J]$ is a leaf.

(3) **Multiscale Approximation**: When $\mathcal{M}$ is sufficiently smooth the affine projectors at each scale $j \in [J]$, $\left\{ \mathbb{P}_{j,k} \;\middle|\; k \in [K_j] \right\}$, approximate $\mathcal{M}$ pointwise with error $O\left(2^{-2j}\right)$.
   (a) There exists a constant $j_0 \in [J-1]$ so that $\vec{c}_{j,k} \in \mathrm{tube}_{C_1 \cdot 2^{-j-2}}(\mathcal{M})$ for all $j \in [J] - [j_0]$ and $k \in [K_j]$. Note that j_0 is a function of the constant C_1 from Property 2b. We will generally assume that a $j_0 \in [J-1]$ satisfying this condition exists when C_1 is chosen to be as large as possible above.
   (b) For each $j \in [J]$ and $\vec{x} \in \mathbb{R}^D$ let $k_j(\vec{x}) \in [K_j]$ be such that $\vec{c}_{j,k_j(\vec{x})}$ is one of the nearest neighbors of $\vec{x}$ in the set $\left\{ \vec{c}_{j',k} \;\middle|\; j' = j,\ k \in [K_j] \right\}$. That is, for each $j \in [J]$, let

$$k_j(\vec{x}) = \arg\min_{k \in [K_j]} \|\vec{x} - \vec{c}_{j,k}\|.$$

   Then, for each $\vec{x} \in \mathcal{M}$ there exists a constant $C \in \mathbb{R}^+$ such that

$$\left\| \vec{x} - \mathbb{P}_{j,k_j(\vec{x})}(\vec{x}) \right\| \leq C \cdot 2^{-2j}$$

   for all $j \in [J]$. In addition, affine projectors associated with $\vec{c}_{j,k}$-vectors that are nearly as close to any $\vec{x} \in \mathcal{M}$ as $\vec{c}_{j,k_j(\vec{x})}$ can also accurately represent $\vec{x}$. In particular, their exists a constant $\tilde{C} \in \mathbb{R}^+$ such that

$$\left\| \vec{x} - \mathbb{P}_{j,k'}(\vec{x}) \right\| \leq \tilde{C} \cdot 2^{-j}$$

   for all $\vec{x} \in \mathcal{M}$, $j \in [J]$, and $k' \in [K_j]$ satisfying

$$\left\| \vec{x} - \vec{c}_{j,k'} \right\| \leq 16 \cdot \max \left\{ \left\| \vec{x} - \vec{c}_{j,k_j(\vec{x})} \right\|,\ C_1 \cdot 2^{-j-1} \right\}. \tag{5}$$

[2]Note that $\tilde{C}$ is a uniform constant over $\mathcal{M}$ for all scales j. To see why GMRA will produce a piecewise linear approximation with this property given sufficiently representative training data from $\mathcal{M}$, recall that GMRA builds a sequence of refined partitions of the dataset it approximates (one

Note that the affine projectors approximate $\mathcal{M}$ more accurately as the scale $j \in [J]$ increases. The finest scale resolution is obtained when $j = J$. See [3] for details.

The remainder of this paper is devoted to analyzing the number of measurements required in order to approximately reconstruct an arbitrary point $\vec{x} \in \mathbb{R}^D$ which is nearly on a compact d-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^D$. In order to yield substantive progress we must first assume some knowledge of $\mathcal{M}$ (i.e., our manifold-based signal dictionary). Thus, we will assume below that we have a set of affine projectors, $\left\{\mathbb{P}_{j,k} \;\middle|\; j \in [J], k \in [K_j]\right\}$, for $\mathcal{M}$ as discussed above, and will primarily focus our analysis on bounding the number of measurements, m, sufficient to accurately compute $\mathbb{P}_{j,k_j(\vec{x})}(\vec{x})$ for any given input vector $\vec{x} \in \mathbb{R}^D$ and scale $j \in [J]$.

2.3. **The Goal: Approximating Manifold Data via Compressive Measurements.** Let $\mathbb{P} = \left\{\mathbb{P}_j \;\middle|\; j \in [J]\right\}$ be a multiscale piecewise linear approximation to $\mathcal{M}$ as discussed above. Given such a $\mathbb{P}$ we can accurately approximate any $\vec{x} \in \mathcal{M} \subset \mathbb{R}^D$ (e.g., see Property 3b). However, herein we are primarily interested in approximating arbitrary vectors, $\vec{x} \in \mathbb{R}^D \setminus \mathcal{M}$, as well as they can be approximated by a nearest neighbor on the manifold, $\vec{x}_{\text{opt}} \in \mathcal{M}$. As we shall see, $\mathbb{P}$ can be utilized for this task. The following lemma demonstrates that $\mathbb{P}_{j,k_j(\vec{x})}(\vec{x})$ approximates any vector $\vec{x} \in \mathbb{R}^D$ nearly as well as $\vec{x}_{\text{opt}} \in \mathcal{M}$ does.

Lemma 2. *Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact d-dimensional Riemannian submanifold of $\mathbb{R}^D$, and $\vec{x} \in \mathbb{R}^D$. Furthermore, let $\mathbb{P}_j = \left\{\mathbb{P}_{j,k} \mid k \in [K_j]\right\}$ be a scale $j \in [J]$ GMRA approximation to $\mathcal{M}$. Then,*

$$\left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\| \;\leq\; 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + O\left(2^{-j}\right)$$

for all $k' \in [K_j]$ satisfying

$$\left\|\vec{x} - \vec{c}_{j,k'}\right\| \leq 8 \cdot \max\left\{\left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\|,\; C_1 \cdot 2^{-j-1}\right\}.$$

Proof: Let $\delta = \max\left\{\left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\|,\; C_1 \cdot 2^{-j-1}\right\}$, where $C_1 \in \mathbb{R}$ is defined as in Property 2b above. Furthermore, let $k' \in [K_j]$ be such that $\left\|\vec{x} - \vec{c}_{j,k'}\right\| \leq 8\delta$. To begin, suppose that $\left\|\vec{x} - \vec{c}_{j,k'}\right\| \;\leq\; 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|$. In this case we are essentially finished since

$$\left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\| \;=\; \left\|\left[I - \Phi_{j,k'}^{\mathrm{T}}\Phi_{j,k'}\right]\left(\vec{x} - \vec{c}_{j,k'}\right)\right\| \;\leq\; \left\|\vec{x} - \vec{c}_{j,k'}\right\| \;\leq\; 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|.$$

Thus, we will hereafter assume that $\left\|\vec{x} - \vec{c}_{j,k'}\right\| \;>\; 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|$ without loss of generality.

Repeatedly applying the triangle inequality we see that $\left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\|$ is bounded above by

$$\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \left\|\vec{x}_{\text{opt}} - \mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right)\right\| + \left\|\mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right) - \mathbb{P}_{j,k'}(\vec{x})\right\|.$$

The third term in the sum immediately above can be bounded by

$$\left\|\mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right) - \mathbb{P}_{j,k'}(\vec{x})\right\| \;=\; \left\|\Phi_{j,k'}^{T}\Phi_{j,k'}\left(\vec{x} - \vec{x}_{\text{opt}}\right)\right\| \;\leq\; \left\|\vec{x} - \vec{x}_{\text{opt}}\right\|.$$

To bound the second term we note that $\left\|\vec{x} - \vec{c}_{j,k'}\right\| \;>\; 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|$ implies that $\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k_j\left(\vec{x}_{\text{opt}}\right)}\right\| \;>\; 9\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|/8$. Therefore,

$$\begin{aligned}\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k'}\right\| \;&\leq\; \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \left\|\vec{x} - \vec{c}_{j,k'}\right\| \;\leq\; \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + 8 \cdot \max\left\{\left\|\vec{x} - \vec{c}_{j,k_j\left(\vec{x}_{\text{opt}}\right)}\right\|,\; C_1 \cdot 2^{-j-1}\right\} \\ &\leq\; 9\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + 8 \cdot \max\left\{\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k_j\left(\vec{x}_{\text{opt}}\right)}\right\|,\; C_1 \cdot 2^{-j-1}\right\} \;<\; 16 \cdot \max\left\{\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k_j\left(\vec{x}_{\text{opt}}\right)}\right\|,\; C_1 \cdot 2^{-j-1}\right\}.\end{aligned}$$

Property 3b now guarantees that $\left\|\vec{x}_{\text{opt}} - \mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right)\right\| \;\leq\; \tilde{C} \cdot 2^{-j}$. Hence, we now have

$$\left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\| \;\leq\; 2\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \tilde{C} \cdot 2^{-j}.$$

partition per scale). The diameters of these partitions' elements decrease uniformly with scale, and are therefore bounded above by $C \cdot 2^{-j}$ for some universal $C \in \mathbb{R}^+$. Hence, if $\vec{x} \in \mathcal{M}$ satisfies Equation 5, we know that $\left\|\vec{x} - \vec{c}_{j,k'}\right\| \leq 16 \cdot C' 2^{-j}$ for some universal $C' \in \mathbb{R}^+$. Noting that $\left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\| \;\leq\; \left\|\vec{x} - \vec{c}_{j,k'}\right\|$ establishes the desired uniformity of $\tilde{C}$ for full (i.e., undecimated) GMRA manifold approximations.

The result follows. □

In this paper we are primarily concerned with achieving approximation results akin to Lemma 2 utilizing *compressive measurements*. This will allow us to extend the successful sparse approximation techniques and results of compressive sensing to the recovery of signals which belong to low dimensional submanifolds of $\mathbb{R}^D$. In order to accomplish this goal we must first propose and then subsequently analyze both a measurement operator and an associated recovery algorithm. Furthermore, in order for it to be of practical value, we must demonstrate that the proposed recovery algorithm is computationally efficient, easy to implement, and provably accurate. We begin this process by considering our measurement matrices in Section 2.5. We then develop a practical reconstruction algorithm in Section 3. Before we begin, however, we will first state the main results proven herein.

2.4. **Main Results.** In the statements of the two propositions below, $C \in \mathbb{R}^+$ is an absolute universal constant which is independent of $\vec{x}$, $\mathcal{M}$, $\mathcal{M}$'s GMRA approximation, etcetera. Note that the upper bounds provided for this constant in Section 3 are almost surely quite loose. We state our first result.

Proposition 1. *Fix precision parameter $\delta \in \mathbb{R}^+$ and let $\vec{x} \in \mathbb{R}^D$. In addition, let $\mathbb{P}_J$, $J = O\left(\log\left[1/(\delta \operatorname{reach}(\mathcal{M}))\right]\right)$, be a GMRA approximation to a given compact d-dimensional Riemannian manifold, $\mathcal{M} \subset \mathbb{R}^D$, with volume V. Finally, let*

$$m = O\left(d \log\left(\frac{d}{\delta \operatorname{reach}(\mathcal{M})}\right) + \log V\right)$$

be a natural number[3], and define $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^D$ to be Algorithm 1 from Section 3 below. Then, there exists an $m \times D$ matrix, M, such that

$$\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\| \;<\; C \cdot \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \delta.$$

Furthermore, $\mathcal{A}(M\vec{x})$ can be evaluated in $\left(m^{O(1)} + O(dD)\right)$-time.

Proof: The result follows from Theorem 3, the first part of Theorem 2, and the discussion in Section 3.1. □

Proposition 1 provides a nonuniform recovery guarantee for each given $\vec{x} \in \mathbb{R}^D$. The measurement matrices, M, referred to by the proposition can be any standard Johnson-Lindenstrauss embedding matrix (e.g., a Gaussian random matrix [23]). Hence, they are well understood.[4] The worst case theoretical runtime complexity of the recovery algorithm is polynomial in m. We refer the reader to Section 5 for an empirical evaluation of the recovery algorithm's computational efficiently in practice. Finally, we note that the number of required measurements, m, is entirely independent of the extrinsic dimension, D. Next, we state a uniform approximation guarantee for Algorithm 1.

Proposition 2. *Fix precision parameter $\delta \in \mathbb{R}^+$. In addition, let $\mathbb{P}_J$, $J = O\left(\log\left[1/(\delta \operatorname{reach}(\mathcal{M}))\right]\right)$, be a GMRA approximation to a given compact d-dimensional Riemannian manifold, $\mathcal{M} \subset \mathbb{R}^D$, with volume V. Finally, let*

$$m = O\left(d \log\left(\frac{D}{\delta \operatorname{reach}(\mathcal{M})}\right) + \log V\right)$$

be a natural number, and define $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^D$ to be Algorithm 1 from Section 3 below. Then, there exists an $m \times D$ matrix, M, such that

$$\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\| \;<\; C\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|_2 + \frac{C}{\sqrt{d}}\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|_1 + \delta$$

for all $\vec{x} \in \mathbb{R}^D$ with

$$2\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|_2 \;+\; \frac{6}{5\sqrt{d}}\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|_1 \;\leq\; \max\left\{\left\|\vec{x} - \vec{c}_{J,k_J}(\vec{x})\right\|,\; \delta\right\}.$$

Furthermore, $\mathcal{A}(M\vec{x})$ can be evaluated in worst case $\left(2^{O(d)} \log V + O\left(md^2 + dD\right)\right)$-time.

[3]Note that the upper bounds on m stated both here and in Proposition 2 assume that both δ and $\operatorname{reach}(\mathcal{M})$ are small, i.e., $O(1)$.

[4]Most importantly, we note that many variants of both Propositions 1 and 2 can be easily obtained by using any of the embedding results from [34, 29, 23, 5, 35] in order to define different measurement matrices, M.

Proof: The result follows from Theorem 4, the second part of Theorem 2, and the discussion in Section 3.1. □

Proposition 2 is best interpreted as a general stability result. It guarantees that Algorithm 1 will uniformly approximate all points which are sufficiently close to the manifold $\mathcal{M}$ (i.e., the points need not be exactly on $\mathcal{M}$). Thus, Algorithm 1 has some limited robustness to arbitrary additive input noise. The examples in the experimental section suggest that the constants involved in both the runtime and embedding dimension, m, are very mild in practice.

2.5. **The Measurement Matrix.** In the process of developing an algorithm to approximate $\mathbb{P}_{j,k_j(\vec{x})}(\vec{x})$, and subsequently demonstrating its accuracy, we will require some knowledge regarding our $m \times D$ measurement matrix M. We shall consider two sets of assumptions regarding M's interaction with both the manifold $\mathcal{M}$ and our given set of affine projectors for $\mathcal{M}$ at each scale $j \in [J]$. Each set of assumptions will ultimately result in both different approximation guarantees for our reconstruction algorithm, and different measurements bounds (i.e., sufficient upper bounds on m) for M. We will postpone discussion of how to create M and how to bound the number of rows it must have in order to satisfy each set of assumptions below until Section 4. In Section 3 below we will begin by presenting our reconstruction algorithm together with approximation error bounds under each set of assumptions regarding M.

Let $\vec{x} \in \mathbb{R}^D$ and $\mathbb{P} = \left\{\mathbb{P}_j \;\middle|\; j \in [J]\right\}$ be a fixed set of affine projectors for $\mathcal{M}$ for each scale $j \in [J]$. Fix $\epsilon \in \left(0, \frac{1}{2}\right)$. In Sections 3 and 4 we will assume that our $m \times D$ measurement matrix M satisfies each of these sets of assumptions in turn.

(1) **Assumption Set 1: Required for Nonuniform Recovery of a Given $\vec{x} \in \mathbb{R}^D$ (see Proposition 1)**

(a) Let $S_1 \subset \mathbb{R}^D$ be

$$S_1 \;=\; \left\{\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\left(\vec{x} - \vec{c}_{j,k}\right) \;\middle|\; j \in [J],\; k \in [K_j]\right\} \bigcup \left\{\vec{x} - \vec{c}_{j,k} \;\middle|\; j \in [J],\; k \in [K_j]\right\} \bigcup \left\{\vec{0}\right\}.$$

We will assume that

$$(1-\epsilon)\left\|\vec{y} - \vec{z}\right\|^2 \;\leq\; \left\|M\vec{y} - M\vec{z}\right\|^2 \;\leq\; (1+\epsilon)\left\|\vec{y} - \vec{z}\right\|^2$$

for all $\vec{y}, \vec{z} \in S_1$.

(b) Furthermore, we will assume that

$$(1-\epsilon)\left\|\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\| \;\leq\; \left\|M\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\| \;\leq\; (1+\epsilon)\left\|\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\|$$

for all $j \in [J]$, $k \in [K_j]$, and $\vec{y} \in \mathbb{R}^D$.

(2) **Assumption Set 2: Required for General Stability (see Proposition 2)**

(a) Let $S_2 = \mathcal{M} \bigcup \left\{\vec{c}_{j,k} \;\middle|\; j \in [J],\; k \in [K_j]\right\} \subset \mathbb{R}^D$. We will assume that

$$(1-\epsilon)\left\|\vec{y} - \vec{z}\right\|^2 \;\leq\; \left\|M\vec{y} - M\vec{z}\right\|^2 \;\leq\; (1+\epsilon)\left\|\vec{y} - \vec{z}\right\|^2$$

for all $\vec{y}, \vec{z} \in S_2$.

(b) Furthermore, we will assume that $\left\|M\vec{y}\right\|$ is bounded above by $E_M(\vec{y})$ for all $\vec{y} \in \mathbb{R}^D$, where $E_M : \mathbb{R}^D \to \mathbb{R}^+$ is a continuous function with $E_M\left(\vec{0}\right) = 0$. E_M is discussed in detail in Section 4.2.

(c) As before, we will assume that

$$(1-\epsilon)\left\|\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\| \;\leq\; \left\|M\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\| \;\leq\; (1+\epsilon)\left\|\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\|$$

for all $j \in [J]$, $k \in [K_j]$, and $\vec{y} \in \mathbb{R}^D$.

(d) Finally, we will also assume that

$$(1-\epsilon)\left\|\vec{y} - \mathbb{P}_{j,k}(\vec{y})\right\| - 2^{-J} \;\leq\; \left\|M\vec{y} - M\mathbb{P}_{j,k}(\vec{y})\right\| \;\leq\; (1+\epsilon)\left\|\vec{y} - \mathbb{P}_{j,k}(\vec{y})\right\| + 2^{-J}$$

for all $j \in [J]$, $k \in [K_j]$, and $\vec{y} \in \mathcal{M}$.

Note that the critical difference between the two sets of assumptions above concerns the treatment of $\vec{x} \in \mathbb{R}^D$ and $\vec{x}_{\mathrm{opt}} \in \mathcal{M} \subset \mathbb{R}^D$. If possible we would like to obtain measurement bounds which are independent of the ambient dimension, D. Since an arbitrary vector $\vec{x} \in \mathbb{R}^D \setminus \mathcal{M}$ may contain a substantial portion of its energy in the subspace orthogonal the tangent space to $\mathcal{M}$ at $\vec{x}_{\mathrm{opt}}$, results which are entirely independent of D generally appear to be unattainable unless our measurement matrix happens to successfully preserve information in the direction of $\vec{x} - \vec{x}_{\mathrm{opt}}$. We assume that M preserves lengths of vectors in the general direction of $\vec{x} - \vec{x}_{\mathrm{opt}}$ as part of our first set of assumptions. In

Algorithm 1 APPROXIMATE $\mathbb{P}_{j,k_j(\vec{x})}(\vec{x})$

1: **Input: Measurements** $M\vec{x} \in \mathbb{R}^m$**, Scale** $j \in [J]$**, Approximation** $\mathbb{P}_j = \left\{\mathbb{P}_{j,k} \,\middle|\, k \in [K_j]\right\}$ **to manifold** $\mathcal{M} \subset \mathbb{R}^D$

2: **Output:** $\mathcal{A}(M\vec{x})$**, an approximation to** $\mathbb{P}_{j,k_j(\vec{x})}(\vec{x}) \approx \vec{x}$

3: $k' \longleftarrow \arg\min_{k \in [K_j]} \left\| M\vec{x} - M\vec{c}_{j,k} \right\|$

4: $\vec{u}\,' \longleftarrow \arg\min_{\vec{u} \in \mathbb{R}^d} \left\| M\Phi_{j,k'}^{\mathrm{T}}\vec{u} - M\vec{x} + M\vec{c}_{j,k'} \right\|$

5: $\mathcal{A}(M\vec{x}) \longleftarrow \Phi_{j,k'}^{\mathrm{T}}\vec{u}\,' + \vec{c}_{j,k'}$

6: Output $\mathcal{A}(M\vec{x})$

the second set of assumptions we do not. It is primarily this difference which leads to different measurement bounds and error guarantees in each case.

With respect to reducing the number of required measurements, note that the first set of assumptions allows us to ignore the original manifold and focus on embedding piecewise linear approximations to the manifold instead. Embedding these sets of approximating subspaces to $\mathcal{M}$ instead of directly embedding $\mathcal{M}$ itself is ultimately what allows us to use a number of linear measurements, m, that is independent of the ambient dimension, D. In the second set of assumptions we must embed $\mathcal{M}$ itself. Hence, the number of linear measurements, m, obtained in this case depends logarithmically on the ambient dimension, D.

3. THE RECONSTRUCTION ALGORITHM

We will ultimately upper bound the number of measurements required in order to approximate a given $\vec{x} \in \mathbb{R}^D$ which is close to $\mathcal{M} \subset \mathbb{R}^D$ via the simple reconstruction technique presented in this section. In doing so we will require that the reconstruction algorithm approximates $\vec{x}$ nearly as well as the vector on $\mathcal{M}$ closest to $\vec{x}$,

$$\vec{x}_{\mathrm{opt}} = \arg\min_{\vec{y} \in \mathcal{M}} \|\vec{x} - \vec{y}\|,$$

approximates $\vec{x}$. Our first order of business, therefore, will be to derive explicit error guarantees for the reconstruction technique considered herein which demonstrate that it is indeed "near-optimal" in the sense discussed in Section 1 above. Let $\mathcal{A}(M\vec{x}) \in \mathbb{R}^D$ denote the output of our reconstruction procedure for a given input $\vec{x} \in \mathbb{R}^D$. We wish to bound the approximation error

$$\left\| \vec{x} - \mathcal{A}(M\vec{x}) \right\|$$

in terms of the optimal approximation error, $\|\vec{x} - \vec{x}_{\mathrm{opt}}\|$, and an additive error term of size $O\left(2^{-j}\right)$ whenever possible. Before this task can be accomplished, however, we must first describe the recovery algorithm we will use to calculate $\mathcal{A}(M\vec{x})$.

Our reconstruction procedure uses compressive measurements of $\vec{x}$ in order to approximate $\mathbb{P}_{j,k_j(\vec{x})}(\vec{x})$ in two steps (see Algorithm 1 above). First, the compressive measurements of $\vec{x}$ are used to determine a "center" vector, $\vec{c}_{j,k'}$, which is nearly as close to $\vec{x}$ as its nearest neighboring center, $\vec{c}_{j,k_j(\vec{x})}$, is. This step is guaranteed to work well as long as our measurement matrix, M, preserves appropriate distances between $\vec{x}$ and all the center vectors at scale j. Next, an accurate projection of $\vec{x} - \vec{c}_{j,k'}$ onto the d-dimensional subspace associated with $\vec{c}_{j,k'}$ is found by solving an overdetermined least squares problem. This step will also work well as long as our measurement matrix M is well conditioned on all of the d-dimensional subspaces associated with the scale j center vectors. As we demonstrate below, the two sets of assumptions for M in Section 2.5 are sufficient to guarantee that both steps work well.

The following lemma guarantees that the center found in line 3 of Algorithm 1 is nearly as close to $\vec{x}$ as $\vec{x}$'s true nearest center is.

Lemma 3. *Fix $\epsilon \in \left(0, \frac{1}{2}\right)$. Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact d-dimensional Riemannian submanifold of $\mathbb{R}^D$, and $\vec{x} \in \mathbb{R}^D$. Furthermore, let $\mathbb{P}_j = \left\{\mathbb{P}_{j,k} \mid k \in [K_j]\right\}$ be a scale $j \in [J]$ GMRA approximation to $\mathcal{M}$. Then, if our $m \times D$ measurement*

matrix M satisfies Assumption Set 1 in Section 2.5 above, line 3 of Algorithm 1 will select a $k' \in [K_j]$ which has

$$\left\|\vec{x} - \vec{c}_{j,k'}\right\| \;\leq\; \sqrt{\frac{1+\epsilon}{1-\epsilon}} \cdot \left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\|.$$

If our $m \times D$ measurement matrix M satisfies Assumption Set 2 in Section 2.5 above, then line 3 of Algorithm 1 will select a $k' \in [K_j]$ which has

$$\left\|\vec{x} - \vec{c}_{j,k'}\right\| \;\leq\; \sqrt{\frac{1+\epsilon}{1-\epsilon}} \cdot \left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\| + \left(1 + \sqrt{\frac{1+\epsilon}{1-\epsilon}}\right)\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \sqrt{\frac{4}{1-\epsilon}} \cdot E_M\left(\vec{x} - \vec{x}_{\text{opt}}\right). \tag{6}$$

Proof: Using the first set of assumptions for M together with the definition of $k' \in [K_j]$ from Algorithm 1 we can see that

$$\left\|\vec{x} - \vec{c}_{j,k'}\right\| \;\leq\; \sqrt{\frac{1}{1-\epsilon}} \cdot \left\|M\vec{x} - M\vec{c}_{j,k'}\right\| \;\leq\; \sqrt{\frac{1}{1-\epsilon}} \cdot \left\|M\vec{x} - M\vec{c}_{j,k_j(\vec{x})}\right\| \;\leq\; \sqrt{\frac{1+\epsilon}{1-\epsilon}} \cdot \left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\|.$$

We now turn our attention to the case where M satisfies the second set of assumptions. We have that

$$\begin{aligned}
\left\|\vec{x} - \vec{c}_{j,k'}\right\| \;&\leq\; \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k'}\right\| \;\leq\; \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \sqrt{\frac{1}{1-\epsilon}} \cdot \left\|M\vec{x}_{\text{opt}} - M\vec{c}_{j,k'}\right\| \\
&\leq\; \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \sqrt{\frac{1}{1-\epsilon}} \left(\left\|M\left(\vec{x} - \vec{x}_{\text{opt}}\right)\right\| + \left\|M\vec{x} - M\vec{c}_{j,k_j(\vec{x})}\right\|\right) \\
&\leq\; \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \sqrt{\frac{4}{1-\epsilon}} \cdot \left\|M\left(\vec{x} - \vec{x}_{\text{opt}}\right)\right\| + \sqrt{\frac{1+\epsilon}{1-\epsilon}} \cdot \left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k_j(\vec{x})}\right\|.
\end{aligned}$$

Focusing on the first and third terms in the line immediately above, we note that

$$\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \sqrt{\frac{1+\epsilon}{1-\epsilon}} \cdot \left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k_j(\vec{x})}\right\| \;\leq\; \sqrt{\frac{1+\epsilon}{1-\epsilon}} \cdot \left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\| + \left(1 + \sqrt{\frac{1+\epsilon}{1-\epsilon}}\right)\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|.$$

The result follows. □

Next, we prove a lemma which guarantees the accuracy of the solution of the overdetermined least squares problem produced by line 4 of Algorithm 1.

Lemma 4. *Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact d-dimensional Riemannian submanifold of $\mathbb{R}^D$, and $\vec{x} \in \mathbb{R}^D$. Furthermore, let $\mathbb{P}_j = \left\{\mathbb{P}_{j,k} \mid k \in [K_j]\right\}$ be a scale $j \in [J]$ GMRA approximation to $\mathcal{M}$, and $k' \in [K_j]$ be the value computed by line 3 of Algorithm 1. Then, if our $m \times D$ measurement matrix M satisfies either set of assumptions in Section 2.5 above, line 5 of Algorithm 1 will produce an $\mathcal{A}(M\vec{x}) \in \mathbb{R}^D$ which has*

$$\left\|\mathbb{P}_{j,k'}(\vec{x}) - \mathcal{A}(M\vec{x})\right\| \;\leq\; \frac{2}{1-\epsilon} \cdot \left\|M\left[\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right]\right\|.$$

Proof: Let $\vec{u}\,' \in \mathbb{R}^d$ be as defined in line 4 of Algorithm 1. Given either set of assumptions for M we will have

$$\left\|\Phi_{j,k'}^{\mathrm{T}}\vec{u}\,' - \Phi_{j,k'}^{\mathrm{T}}\Phi_{j,k'}\left(\vec{x} - \vec{c}_{j,k'}\right)\right\| \;\leq\; \frac{1}{1-\epsilon}\left(\left\|M\Phi_{j,k'}^{\mathrm{T}}\vec{u}\,' - M\left(\vec{x} - \vec{c}_{j,k'}\right)\right\| + \left\|M\left[I - \Phi_{j,k'}^{\mathrm{T}}\Phi_{j,k'}\right]\left(\vec{x} - \vec{c}_{j,k'}\right)\right\|\right),$$

where I is the $D \times D$ identity matrix. By the definition of $\vec{u}\,'$ in Algorithm 1 we can now see that

$$\left\|\Phi_{j,k'}^{\mathrm{T}}\vec{u}\,' - \Phi_{j,k'}^{\mathrm{T}}\Phi_{j,k'}\left(\vec{x} - \vec{c}_{j,k'}\right)\right\| \;\leq\; \frac{2}{1-\epsilon} \cdot \left\|M\left[I - \Phi_{j,k'}^{\mathrm{T}}\Phi_{j,k'}\right]\left(\vec{x} - \vec{c}_{j,k'}\right)\right\|.$$

The stated result follows. □

Finally, we demonstrate the accuracy of the output of Algorithm 1 as an approximation to $\vec{x}$.

Theorem 2. *Fix $\epsilon \in \left(0, \frac{1}{2}\right)$. Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact d-dimensional Riemannian submanifold of $\mathbb{R}^D$, and $\vec{x} \in \mathbb{R}^D$. Furthermore, let $\mathbb{P}_j = \left\{\mathbb{P}_{j,k} \mid k \in [K_j]\right\}$ be a scale $j \in [J]$ GMRA approximation to $\mathcal{M}$. Then, if our $m \times D$ measurement matrix M satisfies Assumption Set 1 in Section 2.5 above, Algorithm 1 will output a point, $\mathcal{A}(M\vec{x}) \in \mathbb{R}^D$, which satisfies*

$$\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\| \ < \ 100.3 \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + O\left(2^{-j}\right).$$

Now suppose that our $m \times D$ measurement matrix M satisfies Assumption Set 2 in Section 2.5 above, and that $\mathbb{P}_j$ is a scale j GMRA approximation to $\mathcal{M}$ for some $j > j_0$ (revisit Properties 3a and 3b in Section 2.2 for the definitions of the constants j_0, C_1, and $\tilde{C}$). Furthermore, suppose that $\vec{x} \in \mathbb{R}^D - \mathcal{M}$ has

$$2 \cdot E_M\left(\vec{x} - \vec{x}_{\text{opt}}\right) \leq \left(8\sqrt{1-\epsilon} - \sqrt{1+\epsilon}\right)\left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\| - \left(\sqrt{1-\epsilon} + \sqrt{1+\epsilon}\right)\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|. \tag{7}$$

Then, Algorithm 1 will output a point, $\mathcal{A}(M\vec{x}) \in \mathbb{R}^D$, which satisfies

$$\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\| \ < \ 220 \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + 4 \cdot E_M\left(\vec{x} - \vec{x}_{\text{opt}}\right) + O\left(2^{-j}\right). \tag{8}$$

Proof: To begin we note that

$$\left\|\vec{x} - \mathcal{A}(M\vec{x})\right\| \ \leq \ \left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\| + \left\|\mathbb{P}_{j,k'}(\vec{x}) - \mathcal{A}(M\vec{x})\right\| \tag{9}$$

where $k' \in [K_j]$ is defined as in line 3 of Algorithm 1. The first set of assumptions for M together with Lemmas 2 and 3 tells us that

$$\left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\| \ \leq \ 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + O\left(2^{-j}\right)$$

since $\epsilon \in \left(0, \frac{1}{2}\right)$. Furthermore, the first set of assumptions for M together with Lemma 4 indicates that

$$\left\|\mathbb{P}_{j,k'}(\vec{x}) - \mathcal{A}(M\vec{x})\right\| \ \leq \ \frac{2}{1-\epsilon} \cdot \left\|M\left[\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right]\right\| \ \leq \ 2\frac{\sqrt{1+\epsilon}}{1-\epsilon} \cdot \left\|\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right\|.$$

Hence, we obtain the stated bound in the first case.

Now assume that M satisfies Assumption Set 2 in Section 2.5. We will begin by bounding the $\left\|\mathbb{P}_{j,k'}(\vec{x}) - \mathcal{A}(M\vec{x})\right\|$ term in Equation 9. Applying Lemma 4 and then utilizing our second set of assumptions regarding M we can see that

$$\begin{aligned}
\left\|\mathbb{P}_{j,k'}(\vec{x}) - \mathcal{A}(M\vec{x})\right\| \ &\leq \ \frac{2}{1-\epsilon} \cdot \left\|M\left[\vec{x} - \mathbb{P}_{j,k'}(\vec{x})\right]\right\| \ \leq \ \frac{2}{1-\epsilon}\left(\left\|Mx - M\vec{x}_{\text{opt}}\right\| + \left\|M\vec{x}_{\text{opt}} - M\mathbb{P}_{j,k'}(\vec{x})\right\|\right) \\
&\leq \ \frac{2}{1-\epsilon}\left(\left\|Mx - M\vec{x}_{\text{opt}}\right\| + \left\|M\vec{x}_{\text{opt}} - M\mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right)\right\| + \left\|M\mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right) - M\mathbb{P}_{j,k'}(\vec{x})\right\|\right) \\
&\leq \ 2 \cdot \frac{1+\epsilon}{1-\epsilon} \cdot \left\|\vec{x} - \vec{x}_{\text{opt}}\right\| + \frac{2}{1-\epsilon}\left(\left\|Mx - M\vec{x}_{\text{opt}}\right\| + \left\|M\vec{x}_{\text{opt}} - M\mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right)\right\|\right). \qquad (10)
\end{aligned}$$

In order to bound the last term in Equation 10 above, we note that $\left\|\vec{x} - \vec{c}_{j,k'}\right\| \ \leq \ 8\left\|\vec{x} - \vec{c}_{j,k_j(\vec{x})}\right\|$ whenever $E_M\left(\vec{x} - \vec{x}_{\text{opt}}\right)$ satisfies Equation 7. Therefore, we will have $\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k'}\right\| \ < \ 16\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k_j\left(\vec{x}_{\text{opt}}\right)}\right\|$ whenever $\left\|\vec{x} - \vec{c}_{j,k'}\right\| > 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|$ by an argument identical to that presented in the second paragraph of the proof of Lemma 2. Hence, Property 3b in Section 2 guarantees that $\left\|\vec{x}_{\text{opt}} - \mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right)\right\| \ \leq \ \tilde{C} \cdot 2^{-j}$ whenever $\left\|\vec{x} - \vec{c}_{j,k'}\right\| > 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|$. Item (d) of Assumption Set 2 in Section 2.5 now guarantees that $\left\|M\vec{x}_{\text{opt}} - M\mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right)\right\|$ will also be $O\left(2^{-j}\right)$ whenever $\left\|\vec{x} - \vec{c}_{j,k'}\right\| > 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|$.

To finish, suppose that $\left\|\vec{x} - \vec{c}_{j,k'}\right\| \leq 17\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|$. Continuing to bound the last term of Equation 10 in this case we obtain

$$\begin{aligned}
\frac{2}{1-\epsilon}\left\|M\vec{x}_{\text{opt}} - M\mathbb{P}_{j,k'}\left(\vec{x}_{\text{opt}}\right)\right\| \ &\leq \ \frac{2}{1-\epsilon}\left(\left\|M\vec{x}_{\text{opt}} - M\vec{c}_{j,k'}\right\| + \left\|M\Phi_{j,k'}^{\mathrm{T}}\Phi_{j,k'}\left(\vec{x}_{\text{opt}} - \vec{c}_{j,k'}\right)\right\|\right) \\
&\leq \ 2 \cdot \frac{\sqrt{1+\epsilon}}{1-\epsilon}\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k'}\right\| + 2 \cdot \frac{1+\epsilon}{1-\epsilon}\left\|\vec{x}_{\text{opt}} - \vec{c}_{j,k'}\right\| \\
&\leq \ 36 \cdot \frac{\sqrt{1+\epsilon}}{1-\epsilon}\left(1 + \sqrt{1+\epsilon}\right)\left\|\vec{x} - \vec{x}_{\text{opt}}\right\|.
\end{aligned}$$

Combining this bound with the previous paragraph concludes the proof. □

Theorem 2 demonstrates that Algorithm 1 can stably approximate vectors $\vec{x} \in \mathbb{R}^D \setminus \mathcal{M}$ as long as the measurement matrix, M, satisfies one of the two sets of assumptions detailed in Section 2.5. However, the strength of the approximation guarantee depends on which set of assumptions M satisfies. When M possess the attributes listed in Assumption Set 1 (most notably, attribute (a)) the vector returned by Algorithm 1 will always provide an approximation to $\vec{x}$ whose error is a within a constant multiple of the optimal approximation error. When M satisfies Assumption Set 2, on the other hand, Algorithm 1 is only guaranteed to provide near optimal approximations for vectors, $\vec{x}$, which are relatively close to the manifold $\mathcal{M}$.

3.1. **Practical Implementation of Algorithm 1.** In line 3 of Algorithm 1 we want to locate the nearest neighbor of $M\vec{x} \in \mathbb{R}^m$ from the set $\left\{M\vec{c}_{j,k} \,\middle|\, k \in [K_j]\right\} \subset \mathbb{R}^m$. This can be accomplished naively in $O(mK_j)$-time. However, K_j is potentially large in the worst case (see Lemma 6 below). Therefore, it is important to note that the runtime's dependence on K_j can be greatly reduced in practice with the aid of standard space partitioning techniques (e.g., by building a k-d tree to solve the nearest neighbor problem). Alternatively, other fast nearest neighbor methods could also be utilized (e.g., see [33, 8, 4] and the references therein). Due to the dyadic structure of our $\vec{c}_{j,k}$-vectors, the worst case theoretical runtime complexity of line 3 can be improved slightly to $\left(2^{O(d)} \log V\right)$-time by using cover trees [8].[5] Alternatively, when it suffices to find a $(1+\delta)$-nearest neighbor of $M\vec{x}$ with high probability, we can utilize even faster algorithms which run in $m^{O(1)}$-time (see Proposition 3 in [33] together with the bound for m in Theorem 3 below).

Line 4 of Algorithm 1 requires the solution of an overdetermined least squares problem. This can be accomplished in $O(md^2)$-time via the singular value decomposition of $M\Phi_{j,k'}^{\mathrm{T}}$. Furthermore, the solution can be computed accurately since both sets of assumptions in Section 2.5 guarantee that $M\Phi_{j,k'}^{\mathrm{T}}$ is well conditioned. Finally, explicitly forming $\mathcal{A}(M\vec{x})$ in line 5 of Algorithm 1 can be accomplished in $O(Dd)$-time. The total runtime of Algorithm 1 will therefore be $O\left(d(md + D) + T_{\mathrm{NN}}\right)$, where T_{NN} bounds the runtime of the nearest neighbor algorithm used in line 3.

4. Upper Bounds on the Number of Required Measurements

In this section we will bound the number of rows, m, needed in order for our $m \times D$ measurement matrix, M, to satisfy each set of assumptions discussed in Section 2.5. In order to do so, it will suffice to let M be a linear Johnson-Lindenstrauss embedding of a well chosen set of points in $\mathbb{R}^D$ into $\mathbb{R}^m$. Of course, this set of points will vary depending on which set of assumptions from Section 2.5 we want M to satisfy. Below we consider each set of assumptions separately. However, we will first establish two lemmas which will be useful in both cases.

Lemma 5. *Let $\epsilon \in \left(0, \frac{1}{2}\right)$. Furthermore, let $j \in [J]$ and $k \in [K_j]$ denote an affine projector $\mathbb{P}_{j,k}$ (see Property 1 in Section 2). Then, there exists a finite set of vectors, $Q_{j,k} \subset X_{j,k} = \left\{\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y} \,\middle|\, \vec{y} \in \mathbb{R}^D\right\}$ with $\left|Q_{j,k}\right| \leq (12/\epsilon)^d + 1$, such that*

$$(1-\epsilon)\left\|\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\| \leq \left\|M\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\| \leq (1+\epsilon)\left\|\Phi_{j,k}^{\mathrm{T}}\Phi_{j,k}\vec{y}\right\|$$

for all $\vec{y} \in \mathbb{R}^D$ whenever M embeds $Q_{j,k}$ into $\mathbb{R}^m$ with $\epsilon/2$-distortion.

Proof: We let $Q'_{j,k}$ be a minimal $\epsilon/4$-cover of the d-dimensional unit ball in $X_{j,k}$ centered at $\vec{0} \in X_{j,k}$. Now set $Q_{j,k} = Q'_{j,k} \bigcup \left\{\vec{0}\right\}$. The stated upper bound of $\left|Q_{j,k}\right|$ follows from existing covering results (see [5] for references). Furthermore, if M embeds $Q_{j,k}$ into $\mathbb{R}^m$ with $\epsilon/2$-distortion it is easy to see that

$$(1-\epsilon/2)\|q\| \leq \|Mq\| \leq (1+\epsilon/2)\|q\|$$

for all $q \in Q_{j,k}$. The remainder of the proof now directly parallels the proof of Lemma 5.1 in [5]. □

Lemma 6. *Fix $J \in \mathbb{N}$ and let $\mathbb{P}_j$, $j \in [J]$, be a GMRA approximation to a given compact d-dimensional Riemannian manifold, $\mathcal{M} \subset \mathbb{R}^D$, with d-dimensional volume V. Furthermore, suppose that $j' \in [J] - \left[\max\left\{j_0, \log_2\left(\frac{C_1}{\mathrm{reach}(\mathcal{M})}\right) - 2\right\}\right]$,*

[5] Here V is the volume of the d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^D$.

where j_0 and C_1 are defined as in Property 3a of Section 2. Then, the number of affine projectors at scale j', $K_{j'}$, is bounded above by $\frac{2^{d(j'+1.5)}}{C_1^d} \cdot V \cdot \left(\frac{d}{2}+1\right)^{\frac{d}{2}+1}$.

Proof: We know that $\mathcal{B}_{C_1 \cdot 2^{-j'-2}}\left(\vec{c}_{j',k}\right) \cap \mathcal{M}$ is nonempty for all $k \in [K_{j'}]$ since $j' > j_0$. Now consider a minimal $C_1 \cdot 2^{-j'-2}$-cover of $\mathcal{M}$, $C_{C_1 \cdot 2^{-j'-2}}(\mathcal{M})$. It is not difficult to see that every $\vec{c}_{j',k}$ will be contained in $\mathcal{B}_{C_1 \cdot 2^{-j'-1}}(\vec{y})$ for some $\vec{y} \in C_{C_1 \cdot 2^{-j'-2}}(\mathcal{M})$. Furthermore, there can be no $\vec{y} \in C_{C_1 \cdot 2^{-j'-2}}(\mathcal{M})$ such that two distinct $\vec{c}_{j',k}$ are contained in the same ball, $\mathcal{B}_{C_1 \cdot 2^{-j'-1}}(\vec{y})$, by Property 2b in Section 2. Hence, $K_{j'} \leq \left|C_{C_1 \cdot 2^{-j'-2}}(\mathcal{M})\right|$. Applying Lemma 1 concludes the proof. □

We are now prepared to upper bound the number of rows required by our $m \times N$ measurement matrix, M, in order to satisfy each set of assumptions listed in Section 2.5.

4.1. **Bounding the Number of Rows Required to Satisfy Assumption Set 1.**

Theorem 3. *Fix $\epsilon \in \left(0, \frac{1}{2}\right)$, $\vec{x} \in \mathbb{R}^D$, and $J \in \mathbb{N}$ sufficiently large. Furthermore, let $\mathbb{P}_j$, $j \in [J]$, be a GMRA approximation to a given compact d-dimensional Riemannian manifold, $\mathcal{M} \subset \mathbb{R}^D$, with volume V. Then, there exists an $m \times D$ matrix, M, which satisfies Assumption Set 1 in Section 2.5 with $m = O\left(d\epsilon^{-2}(J + \log(d/\epsilon)) + \epsilon^{-2}\log V\right)$.*

Proof: The set $S_1 \subset \mathbb{R}^D$ defined in item (a) of Assumption Set 1 has $|S_1| \leq 2(J+1)K_J + 1$. Furthermore, applying Lemma 5 to all at most $(J+1)K_J$ affine projectors yields a set of size at most $(J+1)K_J\left((12/\epsilon)^d + 1\right)$ for item (b) of Assumption Set 1. Lemma 6 together with Theorem 1 now finishes the proof. □

It is important to recall that Theorem 1 is proven by showing that a random matrix will (nearly) isometrically embed a given subset of $\mathbb{R}^D$ into $\mathbb{R}^m$ with high probability. In the proof of Theorem 3 above, Theorem 1 is applied to embed a set which depends on the given $\vec{x} \in \mathbb{R}^D$ we are ultimately interested in approximating (i.e., the set S_1 defined in Section 2.5 depends on $\vec{x}$). Thus, Theorem 3 provides us with a high probability recovery guarantee for each separate $\vec{x} \in \mathbb{R}^D$ on which we apply Algorithm 1.

4.2. **Bounding the Number of Rows Required to Satisfy Assumption Set 2.** We will begin this section by considering item (b) of Assumption Set 2. Among other things, this will allow us to finally define the function $E_M : \mathbb{R}^D \to \mathbb{R}^+$. However, we must first define the Restricted Isometry Property [13] on which the subsequent discussion relies.

Definition 2. *Let $D, d \in \mathbb{N}$, and $\epsilon \in (0,1)$. An $m \times D$ matrix M' has the Restricted Isometry Property, RIP(D,d,ϵ), if*

$$(1-\epsilon)\left\|\vec{x}\right\|^2 \leq \left\|M'\vec{x}\right\|^2 \leq (1+\epsilon)\left\|\vec{x}\right\|^2 \tag{11}$$

for all $\vec{x} \in \mathbb{R}^D$ containing at most d nonzero coordinates.

We have the following lemma.

Lemma 7. *Let $\epsilon \in \left(0, \frac{1}{2}\right)$. There exists a finite set of vectors, $Q \subset X = \left\{\vec{y} \,\middle|\, \vec{y} \in \mathbb{R}^D \text{ contains } d \text{ nonzero coordinates}\right\}$ with $\left|Q\right| \leq \binom{D}{d}\left((12/\epsilon)^d + 1\right)$, such that an $m \times D$ matrix M' has the RIP(D,d,ϵ) whenever it embeds Q into $\mathbb{R}^m$ with $\epsilon/2$-distortion. Furthermore, any such matrix M' will have $\left\|M'\vec{y}\right\|_2$ bounded above by*

$$E_{M'}(\vec{y}) = \sqrt{1+\epsilon} \cdot \left[\left\|\vec{y}\right\|_2 + \frac{1}{\sqrt{d}}\left\|\vec{y}\right\|_1\right]$$

for all $\vec{y} \in \mathbb{R}^D$.

Proof: To prove that M' has the RIP(D,d,ϵ) we employ an argument similar to the proof of Theorem 5.2 in [5]. To begin, we define $\vec{e}_j$, $j \in [D] - \{0\}$, to be the j^{th} row of the $D \times D$ identity matrix. Then, for each d-element subset $S = \{j_1, \ldots, j_d\} \subset [D] - \{0\}$, we define X_S to be the d-dimensional subspace spanned by $\vec{e}_{j_1}, \ldots, \vec{e}_{j_d}$. Next, we let Q'_S be a minimal $\epsilon/4$-cover of the d-dimensional unit ball in X_S centered at $\vec{0}$, and define $Q_S = Q'_S \bigcup \left\{\vec{0}\right\}$ as per Lemma 5. Finally, we let

$$Q := \bigcup_{S \subset [D]-\{0\},\ |S|=d} Q_S.$$

The upper bound on $|Q|$ follows immediately.

Now suppose that M' embeds Q into $\mathbb{R}^m$ with $\epsilon/2$-distortion. Every $\vec{x} \in \mathbb{R}^D$ containing at most d nonzero coordinates belongs to some subspace, X_S, whose associated set, $Q_S \subset Q$, is also embedded into $\mathbb{R}^m$ with $\epsilon/2$-distortion by M'. Hence, a trivial variant of Lemma 5 guarantees that every such $\vec{x}$ will satisfy Equation 11. Therefore, M' will have the RIP(D,d,ϵ) as claimed. The equation for $E_{M'}$ now follows from Proposition 3.5 in [38]. □

We are now sufficiently equipped to consider item (a) of Assumption Set 2 in Section 2.5. We have the following lemma.

Lemma 8. *Fix $\epsilon \in \left(0, \frac{1}{2}\right)$ and $J \in \mathbb{N} - \left[\max\left\{j_0, \log_2\left(\frac{C_1}{\text{reach}(\mathcal{M})}\right) - 2\right\}\right]$, where j_0 and C_1 are defined as in Property 3a of Section 2. In addition, let $\mathbb{P}_j$, $j \in [J]$, be a GMRA approximation to a given compact d-dimensional Riemannian manifold, $\mathcal{M} \subset \mathbb{R}^D$, with d-dimensional volume V. Then, there exist absolute universal constants, $C_3, C_4 \in \mathbb{R}^+$, which are independent of both $\mathcal{M}$ and its GMRA approximation, together with a finite set of vectors, $\tilde{B} \subset \mathbb{R}^D$, so that any $m \times D$ matrix M' which embeds $\tilde{B}$ into $\mathbb{R}^m$ with $(C_3 \cdot \epsilon)$-distortion will satisfy*

$$(1-\epsilon)\left\|\vec{y}-\vec{z}\right\|^2 \;\leq\; \left\|M'\vec{y}-M'\vec{z}\right\|^2 \;\leq\; (1+\epsilon)\left\|\vec{y}-\vec{z}\right\|^2$$

for all $\vec{y}, \vec{z} \in \mathcal{M} \bigcup \left\{\vec{c}_{j,k} \;\middle|\; j \in [J],\ k \in [K_j]\right\} \subset \mathbb{R}^D$. Furthermore, $\tilde{B} \subset \mathbb{R}^D$ will have

$$\left|\tilde{B}\right| = O\left(2^{C_4 J \cdot d} V^2 \left(\frac{D}{\epsilon \cdot \min\{1, \text{reach}(\mathcal{M})\} \cdot \min\{1, C_1\}}\right)^{C_4 d}\right).$$

Proof: See Appendix A. □

Furthermore, a modification of the proof of Lemma 8 yeilds our final lemma concerning Assumption Set 2 in Section 2.5. We have the following result regarding item (d) of Assumption Set 2.

Lemma 9. *Fix $\epsilon \in \left(0, \frac{1}{2}\right)$ and $J \in \mathbb{N} - \left[\max\left\{j_0, \log_2\left(\frac{C_1}{\text{reach}(\mathcal{M})}\right) - 2\right\}\right]$, where j_0 and C_1 are defined as in Property 3a of Section 2. In addition, let $\mathbb{P}_j$, $j \in [J]$, be a GMRA approximation to a given compact d-dimensional Riemannian manifold, $\mathcal{M} \subset \mathbb{R}^D$, with d-dimensional volume V. Then, there exist absolute universal constants, $C_5, C_6 \in \mathbb{R}^+$, which are independent of both $\mathcal{M}$ and its GMRA approximation, together with a finite set of vectors, $B' \subset \mathbb{R}^D$, so that any $m \times D$ matrix M' which embeds B' into $\mathbb{R}^m$ with $(C_5 \cdot \epsilon)$-distortion will satisfy*

$$(1-\epsilon)\left\|\vec{y}-\mathbb{P}_{j,k}(\vec{y})\right\| - 2^{-J} \;\leq\; \left\|M'\vec{y}-M'\mathbb{P}_{j,k}(\vec{y})\right\| \;\leq\; (1+\epsilon)\left\|\vec{y}-\mathbb{P}_{j,k}(\vec{y})\right\| + 2^{-J}$$

for all $j \in [J]$, $k \in [K_j]$, and $\vec{y} \in \mathcal{M}$. Furthermore, $B' \subset \mathbb{R}^D$ will have

$$|B'| = O\left(2^{C_6 J \cdot d} V^2 \left(\frac{D}{\epsilon \cdot \min\{1, \text{reach}(\mathcal{M})\} \cdot \min\{1, C_1\}}\right)^{C_6 d}\right).$$

Proof: See Appendix B. □

We are finally ready to provide a useful upper bound for the number of rows required in any measurement matrix satisfying Assumption Set 2 in Section 2.5. We have the following theorem.

Theorem 4. *Fix $\epsilon \in \left(0, \frac{1}{2}\right)$ and $J \in \mathbb{N} - \left[\max\left\{j_0, \log_2\left(\frac{C_1}{\text{reach}(\mathcal{M})}\right) - 2\right\}\right]$, where j_0 and C_1 are defined as in Property 3a of Section 2. In addition, let $\mathbb{P}_j$, $j \in [J]$, be a GMRA approximation to a given compact d-dimensional Riemannian manifold, $\mathcal{M} \subset \mathbb{R}^D$, with d-dimensional volume V. Then, there exists an $m \times D$ matrix, M, which satisfies Assumption Set 2 in Section 2.5 with*

$$m = O\left(d\epsilon^{-2}\log\left(\frac{D}{\epsilon \cdot \text{reach}(\mathcal{M})}\right) + d\epsilon^{-2}J + \epsilon^{-2}\log V\right)$$

and

$$E_M(\vec{y}) = \sqrt{1+\epsilon} \cdot \left[\left\|\vec{y}\right\|_2 + \frac{1}{\sqrt{d}}\left\|\vec{y}\right\|_1\right].$$

Proof: Any $m \times D$ matrix which embeds $\tilde{B} \subset \mathbb{R}^D$ from Lemma 8 into $\mathbb{R}^m$ with $(C_3 \cdot \epsilon)$-distortion will satisfy both items (a) and (b) of Assumption Set 2 in Section 2.5 (see Lemmas 7 and 8). Similarly, any given $m \times D$ matrix which embeds $B' \subset \mathbb{R}^D$ from Lemma 9 into $\mathbb{R}^m$ with $(C_5 \cdot \epsilon)$-distortion will satisfy item (d) of Assumption Set 2. Finally, just as in the proof of Theorem 3 above, Lemma 5 applied to all at most $(J+1)K_J$ affine projectors yields a subset of $\mathbb{R}^D$ of size at most $(J+1)K_J\left((12/\epsilon)^d + 1\right)$ for item (c) of Assumption Set 2. Theorem 1 applied to the union of this subset with $\tilde{B} \cup B'$ guarantees the existence of

$$O\left(\epsilon^{-2}\log\left(\left|\tilde{B}\right| + |B'| + (J+1)K_J\left((12/\epsilon)^d + 1\right)\right)\right) \times D$$

Johnson-Lindenstrauss embedding matrices which satisfy Assumption Set 2 with high probability. Applying Lemmas 6, 8, and 9 to bound K_J, $\left|\tilde{B}\right|$, and $|B'|$, respectively, now finishes the proof. □

In the proof of Theorem 4 above, Theorem 1 is applied to embed a set which only depends on the given manifold, $\mathcal{M}$, and its GMRA approximation. More specifically, no knowledge was assumed regarding any point $\vec{x} \in \mathbb{R}^D \setminus \mathcal{M}$ which we might be interested in approximating via Algorithm 1. Thus, Theorem 4 provides us with a uniform approximation guarantee for all $\vec{x} \in \mathbb{R}^D$ on which we might apply Algorithm 1. However, we pay several penalties for this uniformity. First, the number of rows in our measurement matrix, m, now depends on the *extrinsic dimensionality*, D, of the given manifold. Second, the resulting uniform error bounds are only nontrivial for input points, $\vec{x}$, which are close to the given manifold. Hence, although Theorem 4 implies that Algorithm 1 enjoys a limited form of stability, it does not provide very robust uniform error guarantees in practice.

5. Empirical Evaluation

We implemented Algorithm 1 and present an empirical evaluation of the algorithm in this section.[6] We consider the following examples:

(i) $\mathcal{M}_1$: $20,000$ points sampled from a "swiss roll", a 2-dimensional manifold $\mathcal{S}$ embedded into $\mathbb{R}^{100}$;
(ii) $\mathcal{M}_2$: $40,000$ points sampled from a unit 9-dimensional sphere $\mathbb{S}^9$ embedded into $\mathbb{R}^{100}$;
(iii) $\mathcal{M}_3$: $5,000$ pictures of the digit '1' from the MNIST data base of images, 28×28 pixels, of handwritten digits[7], with each picture having pixel intensity normalized to have unit L^2 norm.
(iii) $\mathcal{M}_4$: $15,000$ points from the MNIST data base, with $5,000$ points sampled from each of the digits $1, 3, 5$, with each picture having pixel intensity normalized to have unit L^2 norm.
(iv) $\mathcal{M}_5$: the Science News text document data set, which comprises 1163 text documents, modeled as vectors in 1153 dimensions, whose i-th entry is the frequency of the i-th word in a dictionary (see [21] for detailed information about this data set), normalized so that every document vector has unit Euclidean norm.

We construct the GMRA on these data sets in order to obtain the linear approximations, $\mathbb{P}_j$ for each scale j considered below in the noiseless setting. For the noisy experiments we add Gaussian noise, $\mathcal{N}(0, \frac{\sigma^2}{D} I_D)$ where D is the (ambient) dimension of the data, to each data point for $\sigma \in (0, 0.05, 0.1)$. We then use the noisy data to compute the GMRA approximations of the noisy data, as well as the random projections utilized by the proposed reconstruction algorithm $\mathcal{A}$. We consider the following measures of approximation:

$$\text{relMSE}(\mathcal{A}, \mathrm{M}, \mathrm{j})^2 := \frac{1}{n}\sum_{i=1}^{n} \frac{||\vec{x}_i - \mathcal{A}(M\vec{x}_i)||^2}{||\vec{x}_i||^2} \quad , \quad \text{relMSE}_{\mathrm{j}}{}^2 := \frac{1}{n}\sum_{i=1}^{n} \frac{||\vec{x}_i - \mathbb{P}_j(\vec{x}_i)||^2}{||\vec{x}_i||^2} \tag{12}$$

where $\{\vec{x}_i\}_{i=1}^n$ are the data points, j is the level in the GMRA, ranging from 0 to J (dependent on the data set), $\mathcal{A}$ is Algorithm 1, and M is a fixed random (with respect to Haar measure) orthogonal projection at each scale j with range of dimension $(d_j \cdot m) \wedge D$, where the "oversampling factor" $m = 1, 2, 4, 16$, and the "intrinsic dimension" $d_j = \max_k \dim(\text{range}(\mathbb{P}_{j,k}))$. Therefore, d_j is the dimension of the manifold (2 and 9, respectively) for $\mathcal{M}_1$ and $\mathcal{M}_2$. The dimension parameter, d_j, is adaptively chosen in a scale-dependent way for $\mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5$ as described in [3], with actual values used in these examples reported in Figure 1.

[6] All GMRA and Algorithm 1 code is freely available at `http://www.math.duke.edu/~mauro/code.html` as part of the GMRA package. An example script for manifold-based compressed sensing, `script_IGWT_CS_2.m`, has been included here.

[7] Available at `http://yann.lecun.com/exdb/mnist/`.

In Figure 1 we also run SpaRSA [51], one of the leading algorithms, among many, for sparse reconstructions. SpaRSA provides additional benchmarking by solving the basis pursuit denoising problem,

$$\vec{y}_i = \underset{\vec{z}\in\mathbb{R}^{O\left((d_j+1)\cdot(K_j+1)\right)}}{\arg\min} \frac{1}{2}\left\|M\vec{x}_i - MW_j\vec{z}\right\|_2^2 + \tau\left\|\vec{z}\right\|_1, \tag{13}$$

for each data point $\vec{x}_i \in \mathbb{R}^D$ at scale j. Here $\tau \in \mathbb{R}^+$ was hand tuned as recommend in [51], and $W_j \in \mathbb{R}^{D\times O\left((d_j+1)\cdot(K_j+1)\right)}$ was the GWRA-based dictionary

$$W_j = \left[\Phi_{j,0}^T\ \vec{c}_{j,0}\ \Phi_{j,1}^T\ \vec{c}_{j,1}\ \cdots\ \Phi_{j,K_j}^T\ \vec{c}_{j,K_j}\right]$$

for each scale $j \in [J]$. We then output $W_j\vec{y}_i \in \mathbb{R}^D$ as SpaRSA's approximation to each $\vec{x}_i$ at scale j, where $\vec{y}_i$ is defined as in Equation 13.

The dashed black line in each Figure 1 plot reports the relMSE_j-error of the GMRA approximation computed with no compression (see Equation 12). Looking at the Figure 1 plots we note that that our manifold-based compressed sensing error, $\text{relMSE}(\mathcal{A}, M, j)$, is close to the uncompressed projector's approximation error, relMSE_j, at each scale j. Hence, we appear able to achieve our stated goals from Section 2.3 computationally as well as theoretically. We also note that: (a) for general real world data sets SpaRSA achieves comparable precision to our algorithm, suggesting that the GMRA dictionaries, W_j, may be used in concert with standard sparse approximation techniques for compressive sensing; (b) for low-dimensional manifold synthetic data sets, which do not curve in many dimensions, SpaRSA achieves higher accuracy, since the directions of a few tangent planes are sufficient to span a subspace containing the entire manifold.

Finally, in Figure 2 we report running times, for the same data sets as in Figure 1, for both our algorithm $\mathcal{A}$ and SpaRSA. These graphs suggest that our algorithm can perform several orders of magnitude faster than SpaRSA. In the examples shown it took a few seconds to run Algorithm 1 on all the points, with SpaRSA taking a significant fraction of a second to run on a single point. In some sense this is entirely expected since Algorithm 1 was designed to take advantage of the GMRA dictionary's structure in order to be faster, while SpaRSA is using the GMRA dictionary generically, without any knowledge of its structure.

6. Conclusion

In this paper we discussed the ability of random projections to embed an intrinsically low d-dimensional submanifold of $\mathbb{R}^D$, together with a piecewise linear approximation to the submanifold, into $\mathbb{R}^{O(d\log d)}$ in a way which (approximately) preserves the fidelity of the embedded piecewise linear approximation. Although any collection of approximating affine spaces suffice, we focussed on the type of multi-scale linear approximations provided by GMRA [3]. It is worth mentioning that the entire Geometric Wavelet Transform (GWT) [3] of a point near a given manifold can also be preserved by the type of random projections discussed herein.

Note that the GWT of a point on a given manifold will always be approximated by the sum of at most Jd vectors (where J is the number of scales in the GWT). So, pessimistically, a random projection needs to preserve all distances in a number of $O(Jd)$-dimensional subspaces which is bounded above by Lemma 6 in order to approximately preserve the entire geometric wavelet transform of each point on the manifold. Thus, the GWT of each point on a given manifold should be preserved in compressed form by a random linear projection onto a subspace whose dimension, m, satisfies a variant of Theorem 3 with d replaced everywhere by Jd.

References

[1] Bootstrap methods in signal processing. *IEEE Signal Proc. Mag.*, 24(4), 2007.

[2] Sensing, sampling, and compression. *IEEE Signal Proc. Mag.*, 25(2), 2008.

[3] W. Allard, G. Chen, and M. Maggioni. Multiscale geometric methods for data sets II: Geometric wavelets. *Appl. Comp. Harm. Anal., accepted*, May 2011.

[4] A. Andoni and P. Indyk. Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions. *Communications of the ACM*, 51(1):419 – 441, 2008.

[5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[6] R. Baraniuk and M. Wakin. Random Projections of Smooth Manifolds. *Foundations of Computational Mathematics*, 9(1):51 – 77, 2009.

[7] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[8] A. Beygelzimer, S. Kakade, and J. Langford. Cover trees for nearest neighbor. In *Proceedings of the 23rd international conference on Machine learning*, pages 97–104. ACM, 2006.

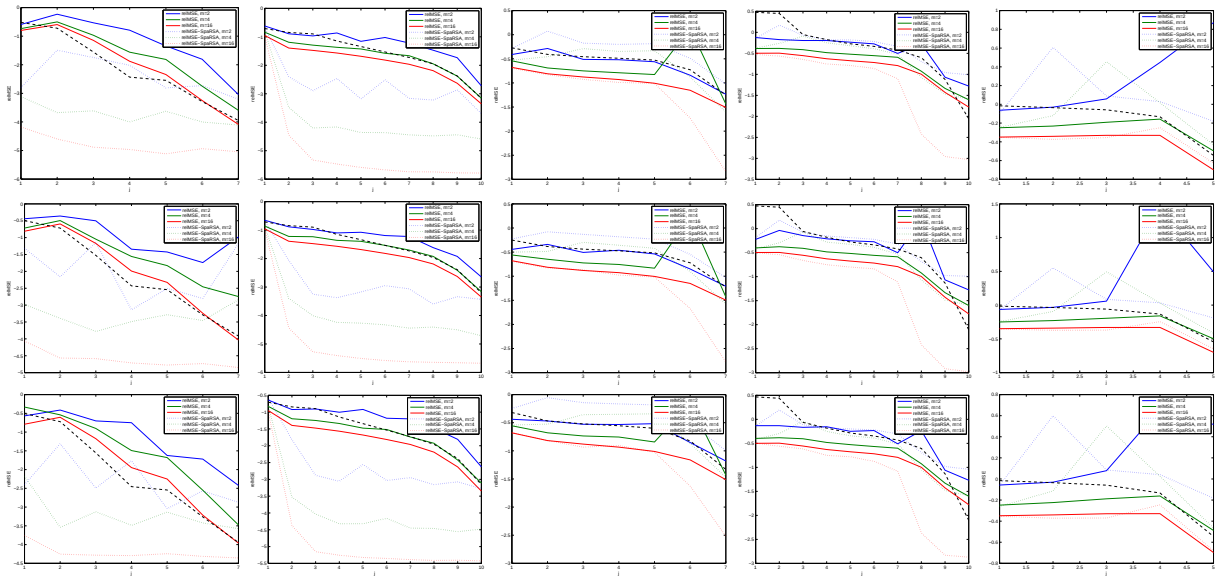

FIGURE 1. In this matrix of Figures, entry (i,k) refers to $\mathcal{M}_k + \mathcal{N}(0, \frac{\sigma_i^2}{D} I_D)$, with $\sigma = (0, 0.05, 1)$ as defined at the beginning of Section 5. Here the horizontal axis representing the scale j, the vertical axis the relative mean square errors relMSE$(\mathcal{A}, M, j)$ defined in (12), and we vary the oversampling parameter $m = 2, 4, 16$ (For $\mathcal{M}_3$ we have $(d_j)_j = (3,3,3,3,9,37,45,45)$, and for $\mathcal{M}_4$ we have $d_j = (9,7,7,7,6,6,9,36,61,66,66)$, and for $\mathcal{M}_5$ and $d_j = (83,51,33,21,43,50)$.). We also plot (dashed black) the relMSE$_J$ defined again (12). We report the average result of 5 draws of the random matrix M (we do not show the negligible standard deviation). Note: the maximum square error (not shown) is about 10 times larger than the mean square error in all cases (not shown). We also run SpaRSA [51] (see comments in the text).

[9] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265 – 274, 2009.

[10] M. Brand. Charting a Manifold. In *Advances in Neural Information Processing Systems*, 2003.

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

[12] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[13] E. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[14] E. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Information Theory*, 2006.

[15] E. J. Cands, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2010.

[16] G. Chen, G. Lerman. Foundations of a multi-way spectral clustering framework for hybrid linear modeling. *Foundations of Computational Mathematics*, 9(5):517–558, 2009.

[17] G. Chen, A. Little, M. Maggioni, and L. Rosasco. Some recent advances in the geometric analysis of point clouds in high dimensions. *Wavelets and Multiscale Analysis: Theory and Applications*, Springer, March 2011.

[18] G. Chen and M. Maggioni. Multiscale geometric and spectral analysis of plane arrangements. In *Proc. CVPR*, 2011. to appear.

[19] 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 Processing*, pages 6140–6155, Dec. 2010.

[20] K. L. Clarkson. Tighter bounds for random projections of manifolds. *In SoCG '08: Proceedings of the Twenty-Fourth Annual Symposium on Computational Geometry*, 2008.

[21] R. Coifman and M. Maggioni. Multiscale data analysis with diffusion wavelets. *Proc. SIAM Bioinf. Workshop, Minneapolis*, April 2007.

[22] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, 102(21):7426–7431, 2005.

[23] S. Dasgupta and A. Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[24] D. Donoho. Compressed Sensing. *IEEE Trans. on Information Theory*, 52:1289–1306, 2006.

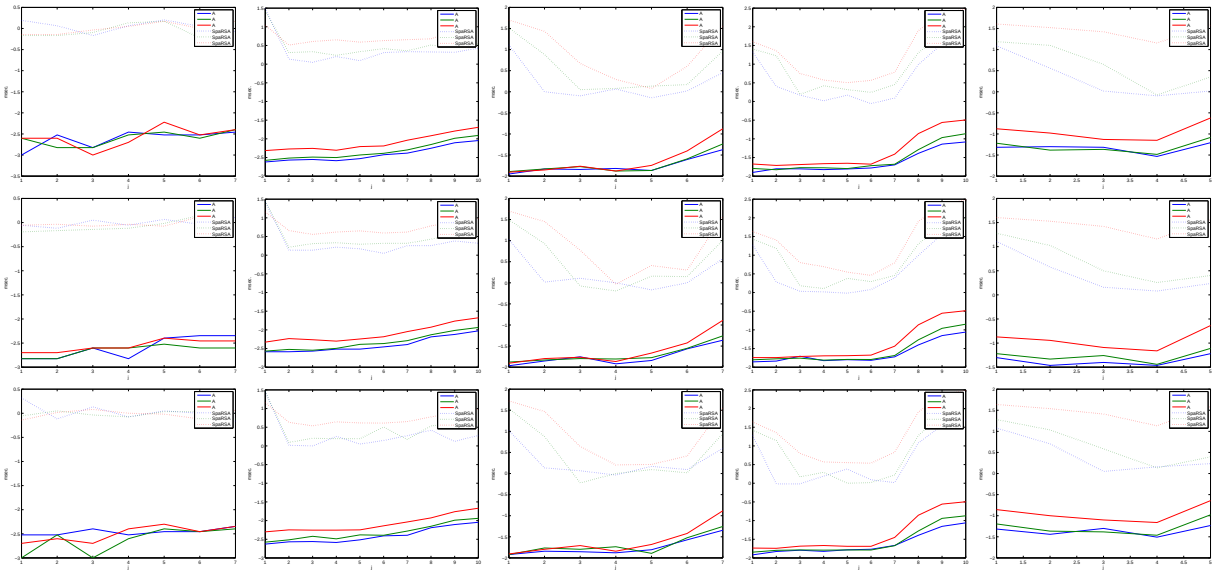

FIGURE 2. WIth the grid of sub-figures referring to the same examples as in Figure 1, we plot as a function of the horizontal axis representing the scale j, the time (in milliseconds in the vertical axis) per point x needed to compute either $\mathcal{A}(x)$ or SpaRSA(x) [51] (with the matrices involved in the latter algorithm, and their transposed, precomputed). Because of the long waiting time involved, we ran SpaRSA only on 50 randomly chosen points rather than all the points.

[25] D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[26] D. L. Donoho and C. Grimes. Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data. *Proc. Nat. Acad. Sciences*, pages 5591–5596, March 2003.

[27] D. N. E. J. Cands, Y. Eldar and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comp. Harm. Anal.*, 31(1):59–73.

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

[29] P. Frankl and H. Maehara. The johnson-lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory, Series B*, 44(3):355–362, 1988.

[30] A. C. Gilbert, M. Muthukrishnan, and M. J. Strauss. Approximation of functions over redundant dictionaries using coherence. *In Proc. of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, Jan. 2003.

[31] G. Haro, G. Randall, and G. Sapiro. Translated Poisson mixture model for stratification learning. *Int. J. Comput. Vision*, 80(3):358–374, 2008.

[32] W. Hong, J. Wright, K. Huang, and Y. Ma. A multi-scale hybrid linear model for lossy image representation. In *ICCV*, pages 764–771, 2005.

[33] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 604–613. ACM, 1998.

[34] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemp. Math*, 26:189–206, 1984.

[35] F. Krahmer and R. Ward. New and improved johnson-lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[36] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 50(3):413–458, 2008.

[37] S. Mallat. A wavelet tour of signal processing. *Academic Press*, 2003.

[38] D. Needell and J. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[39] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9:317–334, 2009.

[40] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):310–316, 2010.

[41] P. Niyogi, S. Smale, and S. Weinberger. Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete Comput Geom*, 39:419 – 441, 2008.

[42] H. Rauhut. Compressive sensing and structured random matrices. *Theoretical Foundations and Numerical Methods for Sparse Recovery*, 9:1–92, 2010.

[43] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Trans. Inform. Theory*, 54(5):22102219, 2008.

[44] P. Shah and V. Chandrasekaran. Iterative projections for signal identification on manifolds: Global recovery guarantees. In *49th Annual Allerton Conference on Communication, Control, and Computing*, pages 760 -767. Sept. 2011.
[45] Y. Sugaya and K. Kanatani. Multi-stage unsupervised learning for multi-body motion segmentation. *IEICE Transactions on Information and Systems*, E87-D(7):1935–1942, 2004.
[46] J. B. Tenenbaum, V. D. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
[47] J. Tropp and A. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. *IEEE Trans. Info. Theory*, 53(12):4655–4666, Dec. 2007.
[48] R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011.
[49] M. B. Wakin. Manifold-Based Signal Recovery and Parameter Estimation from Compressive Measurements. *Preprint*, 2008.
[50] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk. The multiscale structure of non-differentiable image manifolds. *Proc. Wavelets XI*, SPIE Optics and Photonics, San Diego, California, August, 2005.
[51] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.
[52] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal of Scientific Computing*, 26:313–338, 2002.

Appendix A. Proof of Lemma 8

To prove this lemma we will modify the proof of Theorem 3.1 in [6]. The proof of Theorem 3.1 proceeds in two steps. First, a finite set, $B \subset \mathbb{R}^D$, of points on/near the given manifold $\mathcal{M}$ is defined. The main body of the proof then consists of demonstrating that any $m \times D$ matrix, M', which embeds B into $\mathbb{R}^m$ with $\Theta(\epsilon)$-distortion will also satisfy

$$(1-\epsilon)\left\|\vec{x}-\vec{y}\right\| \;\leq\; \left\|M'\vec{x}-M'\vec{y}\right\| \;\leq\; (1+\epsilon)\left\|\vec{x}-\vec{y}\right\|$$

for all $\vec{x}, \vec{y} \in \mathcal{M}$. Our proof will proceed along a similar path. We will begin by first defining a modified version of the set, B, considered in [6]. We will call this set $\tilde{B}$. Then, we will prove that any $m \times D$ matrix which which embeds $\tilde{B}$ into $\mathbb{R}^m$ with $\Theta(\epsilon)$-distortion will also satisfy item (a) of Assumption Set 2 in Section 2.5.

Let $d_{\mathcal{M}}(\vec{x}, \vec{y})$ denote the geodesic distance between $\vec{x}, \vec{y} \in \mathcal{M}$. Furthermore, let $\mathbf{Tan}_{\vec{x}}$ denote the d-dimensional tangent space to $\mathcal{M}$ at each $\vec{x} \in \mathcal{M}$. Finally, let

$$\mathcal{B}_{\mathcal{M},\delta}(\vec{x}) = \left\{\vec{y} \in \mathcal{M} \;\middle|\; d_{\mathcal{M}}(\vec{x}, \vec{y}) \;\leq\; \delta\right\}$$

for each $\delta \in \mathbb{R}^+$ and $\vec{x} \in \mathcal{M}$.

We are now ready to construct $B \subset \mathbb{R}^D$ as per [6] as follows: Set $T = O\left(\frac{\epsilon^2}{D} \cdot \min\{1, \mathrm{reach}(\mathcal{M})\}\right)$ and, for each $\vec{x} \in \mathcal{M}$, let $Q_2(\vec{x}) \subset \mathbf{Tan}_{\vec{x}}$ denote a minimal $\Theta\left(\epsilon \cdot T/\sqrt{D}\right)$-cover of the d-dimensional Euclidean ball of radius T centered at $\vec{0} \in \mathbf{Tan}_{\vec{x}}$. Next, choose $A \subset \mathcal{M}$ to be a minimal finite cover of $\mathcal{M}$ satisfying

$$\min_{\vec{a} \in A} d_{\mathcal{M}}(\vec{a}, \vec{x}) \;\leq\; T,$$

for all $\vec{x} \in \mathcal{M}$. Then,

$$B := \bigcup_{\vec{a} \in A} \{\vec{a}\} \cup (\vec{a} + Q_2(\vec{a})).$$

In the next paragraph we will define our modified set, $\tilde{B} \subset \mathbb{R}^D$, which is a superset of the set B defined above.

Fix $j \in [J]$ and $k \in [K_j]$. For each $\vec{a} \in A$ above, let $\vec{a}_{j,k} \in \mathcal{B}_{\mathcal{M},T}(\vec{a})$ be such that

$$\left\|\vec{a}_{j,k} - \vec{c}_{j,k}\right\| \;\leq\; \left\|\vec{y} - \vec{c}_{j,k}\right\| \;\forall \vec{y} \in \mathcal{B}_{\mathcal{M},T}(\vec{a}).$$

Let $A_{j,k} = \left\{\vec{a}_{j,k} \;\middle|\; \vec{a} \in A\right\}$. Furthermore, denote the $(d+1)$-dimensional vector space spanned by $\mathbf{Tan}_{\vec{a}_{j,k}} \bigcup \left\{\vec{c}_{j,k} - \vec{a}_{j,k}\right\}$ by $\mathbf{Tan}_{\vec{a},j,k}$, and then let $Q_{j,k}(\vec{a}) \subset \mathbf{Tan}_{\vec{a},j,k}$ be a minimal $\Theta\left(\epsilon \cdot T/\sqrt{D}\right)$-cover of the $(d+1)$-dimensional Euclidean ball of radius T centered at $\vec{0}$. To finish, define

$$B_{j,k} := \bigcup_{\vec{a} \in A} \left\{\vec{a}_{j,k}\right\} \cup \left(\vec{a}_{j,k} + Q_{j,k}(\vec{a})\right)$$

and then set

$$\tilde{B} := \left(\bigcup_{j \in [J],\; k \in [K_j]} B_{j,k} \cup \left\{\vec{c}_{j,k}\right\}\right) \cup B \cup Q,$$

where $Q \subset \mathbb{R}^D$ is as defined in Lemma 7.

Note that $\left|\tilde{B}\right|$ will be bounded above by

$$(J+1)\cdot K_J\left(1+\max_{j\in[J],\ k\in[K_j]}\left|B_{j,k}\right|\right)+|B|+|Q|\,.$$

Applying Lemma 7 to bound $|Q|$, Lemma 6 to bound K_J, and appealing to Section 3.2.5 of [6] to bound $|B|$, the previous line reveals that

$$\left|\tilde{B}\right| \ll 2^{O(J\cdot d)}\cdot V\cdot\left(\frac{d}{\min\{1,C_1\}}\right)^{O(d)}\left(\max_{j\in[J],\ k\in[K_j]}\left|B_{j,k}\right|\right)+V\left(\frac{D}{\epsilon\cdot\min\{1,\text{reach}\,(\mathcal{M})\}}\right)^{O(d)}. \tag{14}$$

We now finish bounding the cardinality of $\tilde{B}$ by noting that $\left|B_{j,k}\right|$ will always be bounded above by the upper bounds for $|B|$ in Section 3.2.5 of [6] after every occurrence of $K=d$ is replaced with $d+1$.[8] The stated upper bound on $\left|\tilde{B}\right|$ follows.

We will now complete the second portion of our proof by demonstrating that a sufficiently precise linear embedding of $\tilde{B}$ will satisfy item (a) of Assumption Set 2. First, since $B\subset\tilde{B}$, Theorem 3.1 in [6] guarantees that a low-distortion embedding of $\tilde{B}$ will preserve all pairwise distances between points on the manifold $\mathcal{M}$. Furthermore, any embedding of $\tilde{B}$ will also embed all $\vec{c}_{j,k}$-vectors since they form a proper subset of $\tilde{B}$. Hence, if suffices for us to show that a sufficiently precise linear embedding of $\tilde{B}$ will (approximately) preserve the distance from each $\vec{c}_{j,k}$-vector to all points on the manifold $\mathcal{M}$.

Fix $j\in[J]$, $k\in[K_j]$, and $\vec{x}\in\mathcal{M}$. Let $\vec{a}\,'\in A$ be the closest element of A to $\vec{x}$,

$$\vec{a}\,' \;=\; \arg\min_{\vec{a}\in A}\; d_{\mathcal{M}}(\vec{a},\vec{x})\,.$$

Finally, let $\vec{x}\,'_{j,k}$ denote the projection of $\vec{x}$ onto the $(d+1)$-dimensional affine subspace $\vec{a}\,'_{j,k}+\mathbf{Tan}_{\vec{a}\,',j,k}$. By considering the Taylor series expansion of the unit speed parameterization of the geodesic path from $\vec{a}\,'_{j,k}$ to $\vec{x}$ on $\mathcal{M}$, we find that

$$\vec{x} \;=\; \vec{x}\,'_{j,k}+\vec{r},\ \text{where}\ \left\|\vec{r}\right\| = O\left(\frac{d^2_{\mathcal{M}}\left(\vec{x},\vec{a}\,'_{j,k}\right)}{\text{reach}\,(\mathcal{M})}\right).$$

In fact, the magnitude of the remainder, $\vec{r}$, is also $O\left(\left\|\vec{x}-\vec{a}\,'_{j,k}\right\|^2\right)$ since $T<\text{reach}\,(\mathcal{M})/2$ (see Corollary 2.1 in [6]). Furthermore, the definition of $\vec{a}\,'_{j,k}\in\mathcal{M}$ implies that $\left\|\vec{x}-\vec{a}\,'_{j,k}\right\| = O\left(\left\|\vec{x}-\vec{c}_{j,k}\right\|\right)$.

Continuing with the proof, suppose that an $m\times D$ matrix, M', embeds $\tilde{B}$ into $\mathbb{R}^m$ with $\Theta(\epsilon)$-distortion. A trivial variant of Lemma 5 then implies that

$$\begin{aligned}\left\|M'\vec{x}-M'\vec{c}_{j,k}\right\| \;&\le\; \left\|M'\vec{x}-M'\vec{x}\,'_{j,k}\right\|+\left\|M'\vec{x}\,'_{j,k}-M'\vec{c}_{j,k}\right\| \;\le\; \left\|M'\vec{r}\right\|+(1+\Theta(\epsilon))\left\|\vec{x}\,'_{j,k}-\vec{c}_{j,k}\right\| \\ &\le\; (1+\Theta(\epsilon))\left(\left\|\vec{x}-\vec{c}_{j,k}\right\|+\left\|\vec{r}\right\|\right)+\left\|M'\vec{r}\right\| \;\le\; (1+\Theta(\epsilon))\left\|\vec{x}-\vec{c}_{j,k}\right\|+\left\|M'\vec{r}\right\|+O\left(\left\|\vec{x}-\vec{a}\,'_{j,k}\right\|^2\right)\end{aligned}$$

since $Q_{j,k}(\vec{a}\,')\subset\mathbf{Tan}_{\vec{a}\,',j,k}$ is a proper subset of $\tilde{B}$, and $\left(\vec{x}\,'_{j,k}-\vec{c}_{j,k}\right)\in\mathbf{Tan}_{\vec{a}\,',j,k}$. In addition, the fact that $Q\subset\tilde{B}$ together with Lemma 7 guarantees that M' will have the RIP(D,d,$\Theta(\epsilon)$). This fact combined with the Hölder inequality finally reveals that

$$\begin{aligned}\left\|M'\vec{x}-M'\vec{c}_{j,k}\right\| \;&\le\; (1+\Theta(\epsilon))\left\|\vec{x}-\vec{c}_{j,k}\right\|+O\left(\sqrt{\frac{D}{d}}\cdot\left\|\vec{x}-\vec{a}\,'_{j,k}\right\|^2\right) \;\le\; \left(1+\Theta(\epsilon)+O\left(\sqrt{\frac{D}{d}}\cdot T\right)\right)\left\|\vec{x}-\vec{c}_{j,k}\right\| \\ &\le\; (1+O(\epsilon))\left\|\vec{x}-\vec{c}_{j,k}\right\|.\end{aligned}$$

The lower bound for $\left\|M'\vec{x}-M'\vec{c}_{j,k}\right\|$ is established in an analogous fashion. We have the stated theorem.

[8] Intuitively, we are increasing the effective intrinsic dimensionality of $\mathcal{M}$ from d to $d+1$ in the process of creating our $B_{j,k}$-subsets.

Appendix B. Proof of Lemma 9

The proof of this Lemma borrows heavilly from the proof of Lemma 8. Set $T = O\left(\frac{2^{-J}\epsilon^2}{D} \cdot \min\{1, \text{reach}\,(\mathcal{M})\}\right)$. We will begin by defining the set $B' \subset \mathbb{R}^D$. Let $A \subset \mathcal{M}$, $B \subset \mathbb{R}^D$, and $A_{j,k} = \left\{\vec{a}_{j,k} \,\middle|\, \vec{a} \in A\right\} \subset \mathcal{M}$ for each $j \in [J], k \in [K_j]$ be defined as in Appendix A above (except now using the smaller value of T from the second sentence of this appendix). Let $\widetilde{\mathbf{Tan}}_{\vec{a},j,k}$ denote the $(2d+1)$-dimensional vector space spanned by

$$\mathbf{Tan}_{\vec{a}_{j,k}} \bigcup \left\{\vec{c}_{j,k} - \vec{a}_{j,k}\right\} \bigcup \left\{\Phi_{j,k}^{\mathrm{T}} \Phi_{j,k} \vec{y} \,\middle|\, \vec{y} \in \mathbb{R}^D\right\}$$

for each $\vec{a}_{j,k} \in A_{j,k}$. Furthermore, for each $\vec{a}_{j,k} \in A_{j,k}$, let $Q'_{j,k}(\vec{a}) \subset \widetilde{\mathbf{Tan}}_{\vec{a},j,k}$ be a minimal $\Theta\left(\epsilon \cdot T / \sqrt{D}\right)$-cover of the $(2d+1)$-dimensional Euclidean ball of radius T centered at $\vec{0} \in \widetilde{\mathbf{Tan}}_{\vec{a},j,k}$. To finish, define

$$B'_{j,k} := \bigcup_{\vec{a} \in A} \left\{\vec{a}_{j,k}\right\} \cup \left(\vec{a}_{j,k} + Q'_{j,k}(\vec{a})\right)$$

for each $j \in [J], k \in [K_j]$, and then set

$$B' := \left(\bigcup_{j \in [J],\ k \in [K_j]} B'_{j,k} \cup \left\{\vec{c}_{j,k}\right\} \right) \cup B \cup Q,$$

where $Q \subset \mathbb{R}^D$ is as defined in Lemma 7. It is not difficult to see that $|B'|$ will be bounded above as per Equation 14 after ϵ is replaced everywhere by $2^{-J}\epsilon$. Simplifying yields the stated upper bound.

We will now complete our proof by demonstrating that a sufficiently precise linear embedding of B' will satisfy item (d) of Assumption Set 2. Fix $j \in [J]$, $k \in [K_j]$, and $\vec{x} \in \mathcal{M}$. Let $\vec{a}\,' \in A$ be the closest element of A to $\vec{x}$,

$$\vec{a}\,' \;=\; \arg\min_{\vec{a} \in A} \; d_{\mathcal{M}}(\vec{a}, \vec{x}).$$

Finally, let $\vec{x}\,'_{j,k}$ denote the projection of $\vec{x}$ onto the $(2d+1)$-dimensional affine subspace $\vec{a}\,'_{j,k} + \widetilde{\mathbf{Tan}}_{\vec{a},j,k}$. By considering the Taylor series expansion of the unit speed parameterization of the geodesic path from $\vec{a}\,'_{j,k}$ to $\vec{x}$ on $\mathcal{M}$, we find that

$$\vec{x} \;=\; \vec{x}\,'_{j,k} + \vec{r}, \text{ where } \left\|\vec{r}\right\| = O\left(\frac{d^2_{\mathcal{M}}\left(\vec{x}, \vec{a}\,'_{j,k}\right)}{\text{reach}\,(\mathcal{M})}\right).$$

Furthermore, we recall that the magnitude of the remainder, $\vec{r}$, is also $O\left(\left\|\vec{x} - \vec{a}\,'_{j,k}\right\|^2\right)$ since T is sufficiently small.

To finish, suppose that an $m \times D$ matrix, M', embeds B' into $\mathbb{R}^m$ with $\Theta(\epsilon)$-distortion. A trivial variant of Lemma 5 implies that

$$\begin{aligned}
\left\|M'\vec{x} - M'\mathbb{P}_{j,k}(\vec{x})\right\| \;&\leq\; \left\|M'\vec{x} - M'\vec{x}\,'_{j,k}\right\| + \left\|M'\vec{x}\,'_{j,k} - M'\mathbb{P}_{j,k}(\vec{x})\right\| \;\leq\; \left\|M'\vec{r}\right\| + (1+\Theta(\epsilon))\left\|\vec{x}\,'_{j,k} - \mathbb{P}_{j,k}(\vec{x})\right\| \\
&\leq\; (1+\Theta(\epsilon))\left(\left\|\vec{x} - \mathbb{P}_{j,k}(\vec{x})\right\| + \left\|\vec{r}\right\|\right) + \left\|M'\vec{r}\right\| \\
&\leq\; (1+\Theta(\epsilon))\left\|\vec{x} - \mathbb{P}_{j,k}(\vec{x})\right\| + \left\|M'\vec{r}\right\| + O\left(\left\|\vec{x} - \vec{a}\,'_{j,k}\right\|^2\right)
\end{aligned}$$

since $Q'_{j,k}(\vec{a}\,') \subset \widetilde{\mathbf{Tan}}_{\vec{a},j,k}$ is a subset of B', and $\left(\vec{x}\,'_{j,k} - \mathbb{P}_{j,k}(\vec{x})\right) \in \widetilde{\mathbf{Tan}}_{\vec{a},j,k}$. In addition, the fact that $Q \subset B'$ together with Lemma 7 guarantees that M' will have the RIP(D,d,$\Theta(\epsilon)$). This fact combined with the Hölder inequality reveals that

$$\begin{aligned}
\left\|M'\vec{x} - M'\mathbb{P}_{j,k}(\vec{x})\right\| \;&\leq\; (1+\Theta(\epsilon))\left\|\vec{x} - \mathbb{P}_{j,k}(\vec{x})\right\| + O\left(\sqrt{\frac{D}{d}} \cdot \left\|\vec{x} - \vec{a}\,'_{j,k}\right\|^2\right) \;\leq\; (1+\Theta(\epsilon))\left\|\vec{x} - \mathbb{P}_{j,k}(\vec{x})\right\| + O\left(\sqrt{\frac{D}{d}} \cdot T^2\right) \\
&\leq\; (1+\Theta(\epsilon))\left\|\vec{x} - \mathbb{P}_{j,k}(\vec{x})\right\| + 2^{-J}
\end{aligned}$$

whenever T is weighted by a sufficiently small (universal) constant. The lower bound for $\left\|M'\vec{x} - M'\mathbb{P}_{j,k}(\vec{x})\right\|$ is established in an analogous fashion.