Random Projections of Smooth Manifolds

Richard G. Baraniuk[*] and Michael B. Wakin[†]

October 2006; Revised June 2007

Abstract

We propose a new approach for *nonadaptive* dimensionality reduction of manifold-modeled data, demonstrating that a small number of *random linear projections* can preserve key information about a manifold-modeled signal. We center our analysis on the effect of a random linear projection operator $\Phi : \mathbb{R}^N \to \mathbb{R}^M$, $M < N$, on a smooth K-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^N$. As our main theoretical contribution, we establish a sufficient number M of random projections to guarantee that, with high probability, all pairwise Euclidean and geodesic distances between points on $\mathcal{M}$ are well-preserved under the mapping Φ.

Our results bear strong resemblance to the emerging theory of Compressed Sensing (CS), in which sparse signals can be recovered from small numbers of random linear measurements. As in CS, the random measurements we propose can be used to recover the original data in $\mathbb{R}^N$. Moreover, like the fundamental bound in CS, our requisite M is linear in the "information level" K and logarithmic in the ambient dimension N; we also identify a logarithmic dependence on the volume and curvature of the manifold. In addition to recovering faithful approximations to manifold-modeled signals, however, the random projections we propose can also be used to discern key properties about the manifold. We discuss connections and contrasts with existing techniques in manifold learning, a setting where dimensionality reducing mappings are typically nonlinear and constructed adaptively from a set of sampled training data.

Keywords. Manifolds, dimensionality reduction, random projections, Compressed Sensing, sparsity, manifold learning, Johnson-Lindenstrauss lemma.

AMS Subject Classification. 53A07, 57R40, 62H99, 65C99, 68P30, 68T05, 94A12, 94A29.

1 Introduction

1.1 Concise models for data and information

Effective techniques for processing and understanding data and information often rely on some sort of *model* that characterizes the *expected* behavior of the information. In many cases, the model conveys a notion of constrained structure or *conciseness* to the information; considering a data vector (or "signal") $x \in \mathbb{R}^N$, for example, one may believe that x has "few degrees of freedom" relative to its size N. Such constraints can arise, for example, in data associated with a physical system that has few parameters or in digital samples of a continuous-time function collected at an unnecessarily high sampling rate.

The notion of conciseness is a very powerful assumption, and it suggests the potential for developing highly accurate and efficient data processing algorithms that capture and exploit the inherent structure of the model. Such algorithms can often be inspired (or interpreted) by taking a geometric perspective. Letting $\mathcal{F} \subset \mathbb{R}^N$ denote the class of signals of interest under a given model, many concise models correspond to a *low-dimensional geometric structure* for $\mathcal{F}$, and many algorithms for efficiently processing signals from $\mathcal{F}$ can be loosely characterized as attempts at *dimensionality reduction* relative to the ambient dimension N.

Generally speaking, dimensionality reduction techniques aim to extract low-dimensional information about a signal or collection of signals from some high-dimensional ambient space. Common tasks include approximation and compression, in which the goal is to maintain a low-dimensional representation of a signal x from which a faithful approximation to x can be recovered. For other tasks in dimensionality reduction, however, the goal may not always be to maintain a faithful representation of each signal. Instead, the purpose may be to preserve some critical relationships among elements of a data set or to discover information about the structure $\mathcal{F}$ on which the data lives.

1.2 Low-dimensional geometry

Concise models can arise in a variety of settings and take a variety of forms; we now overview specifically three of the most common model classes and discuss some of the corresponding techniques for dimensionality reduction.

1.2.1 Linear models

Linear signal models arise in cases where a signal depends linearly on a low-dimensional set of parameters (its degrees of freedom), and as a result, the signal class $\mathcal{F}$ takes the form of a K-dimensional linear subspace of $\mathbb{R}^N$. Such models commonly arise in the context of bases and representations. For example, let $\Psi = \{\psi_1, \psi_2, \dots, \psi_N\}$ denote an orthonormal basis[1] for $\mathbb{R}^N$, in which we may represent a signal x in terms of its coefficient vector α,

$$x = \sum_{i=1}^{N} \alpha_i \psi_i.$$

A linear model can be obtained by choosing a particular subset $\Omega \subset \{1, 2, \dots, N\}$, $\#\Omega = K$, of elements from the basis to represent signals

$$x = \sum_{i \in \Omega} \alpha_i \psi_i. \tag{1}$$

In such a case, the resulting signal class $\mathcal{F} = \text{span}(\{\psi_i\}_{i \in \Omega})$ forms a K-dimensional linear subspace of $\mathbb{R}^N$; examples include low-frequency subspaces in the Fourier domain or scaling spaces in the wavelet domain for representing and approximating smooth signals [32].

The linear geometry of these signal classes leads to simple, *linear* algorithms for dimensionality reduction. An ℓ_2-nearest "linear approximation" to a signal $x \in \mathbb{R}^N$ can be computed via orthogonal projection onto the subspace $\mathcal{F}$ (setting α_i to zero for $i \notin \Omega$). Also, the best K-dimensional subspace to approximate a class of signals in $\mathbb{R}^N$ can be discovered using principal components analysis (PCA) (also known as the Karhunen-Loève transform) [42].

---

[1]Although we focus on the case of an orthonormal basis for simplicity, many of the basic ideas we discuss generalize to frames and arbitrary dictionaries in $\mathbb{R}^N$.

1.2.2 Sparse models

Sparse signal models also arise in the context of bases and representations and can be viewed as a generalization of linear models. In a sparse model, every signal from the class $\mathcal{F}$ can again be represented (either exactly or approximately) using a K-term representation from some basis Ψ, but unlike the linear case in (1), the relevant set Ω of basis elements may change from signal to signal. A prototypical example is the wavelet representation of piecewise smooth signals; discontinuities in each signal can be sparsely represented in the wavelet domain, but if the *locations* of the discontinuities can vary from signal to signal, then the required set of wavelets also varies from signal to signal.

Since few coefficients are required to represent any given signal, sparse representations lead to efficient algorithms for dimensionality reduction; naturally these algorithms must typically *adapt* to the changing locations of the significant coefficients. Best K-term "nonlinear approximations" can be computed simply by thresholding the expansion coefficients α in the basis Ψ (letting Ω contain the indices of the K largest coefficients and then setting α_i to zero for $i \notin \Omega$). Transform coding algorithms (which form the heart of many modern signal and image compression standards such as JPEG and JPEG-2000 [39]) also build upon this basic idea.

Again, the geometry of the sparse signal class plays an important role in the signal processing. Unlike the linear case described above, no single K-dimensional subspace suffices to represent all K-sparse signals; instead the set of all sparse signals in the basis Ψ forms a *nonlinear* union of $\binom{N}{K}$ distinct K-dimensional subspaces in $\mathbb{R}^N$,

$$\Sigma_K := \mathcal{F} = \bigcup_{\Omega \subset \{1,2,\ldots,N\},\ \#\Omega=K} \mathrm{span}(\{\psi_i\}_{i\in\Omega}).$$

Approximation by thresholding can be interpreted as the orthogonal projection of x onto the nearest subspace in Σ_K, a simple but *nonlinear* technique owing to the nonlinear geometry of Σ_K.

Despite the apparent need for adaptive, nonlinear methods for dimensionality reduction of sparse signals, a radically different technique known as *Compressed Sensing* (CS) [12, 18] has emerged that relies on a *nonadaptive, linear* method for dimensionality reduction. Like traditional approaches to approximation and compression, the goal of CS is to maintain a low-dimensional representation of a signal x from which a faithful approximation to x can be recovered. In CS, however, the encoder requires no a priori knowledge of the signal structure. Only the decoder uses the model (sparsity in the basis Ψ) to recover the signal.

The CS theory states that with high probability, every K-sparse signal $x \in \mathbb{R}^N$ can be recovered from just $M = O(K \log(N/K))$ *linear* measurements $y = \Phi x$, where Φ represents an $M \times N$ measurement (or encoding) matrix drawn *randomly* from an acceptable distribution [12, 18]. CS decoding involves recovering the signal $x \in \mathbb{R}^N$ from its measurements $y = \Phi x$, where $y \in \mathbb{R}^M$ and $M < N$ in general. Although such inverse problems are generally ill-posed, CS recovery algorithms exploit the additional assumption of *sparsity* in the basis Ψ to identify the correct signal x from an uncountable number of possibilities. CS has many promising applications in signal acquisition, compression, medical imaging, and sensor networks [3, 8, 10, 11, 22, 27, 38]; the random nature of the operator Φ makes it a particularly intriguing *universal* measurement scheme for settings in which the basis Ψ is unknown at the encoder or multi-signal settings in which distributed, collaborative compression can be difficult to coordinate across multiple sensors.

Nonlinear geometry plays a critical role in developing and understanding the CS theory. For example, CS recovery is possible because Φ embeds the sparse signal set Σ_K in $\mathbb{R}^M$ (no two sparse signals in $\mathbb{R}^N$ are mapped to the same point in $\mathbb{R}^M$); such an embedding is ensured with probability one if Φ has i.i.d. Gaussian entries and $M \geq 2K$ [3]. Stable recovery of sparse signals

(via efficient algorithms), however, requires a more "stable" embedding (where sparse signals remain well-separated in $\mathbb{R}^M$). One way of characterizing this stability is known as the *Restricted Isometry Property* (RIP) [9, 10, 12]; we say Φ has RIP of order K if for every $x \in \Sigma_K$,

$$(1 - \epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x\|_2}{\|x\|_2} \leq (1 + \epsilon)\sqrt{\frac{M}{N}}.$$

(Observe that the RIP of order $2K$ ensures that distinct K-sparse signals remain well-separated in $\mathbb{R}^M$.) A random orthoprojector[2] from $\mathbb{R}^N$ to $\mathbb{R}^M$ can be shown to meet the RIP of order K (with respect to any fixed sparsity basis Ψ) with high probability if $M = O(K \log(N/K))$. Hence, the cost for stability is a modest logarithmic "excess" dimensionality factor with respect to the sparsity level K. Various other geometric arguments surrounding CS have also been made involving n-widths of ℓ^p balls and the properties of randomly projected polytopes [11, 12, 17–21, 25, 34, 37].

In a recent paper [2], we have identified a fundamental connection between CS and the Johnson-Lindenstrauss (JL) lemma [1, 16], which concerns the stable embedding of a finite point cloud under a random dimensionality-reducing projection.

Lemma 1.1 [Johnson-Lindenstrauss] *Let Q be a finite collection of points in $\mathbb{R}^N$. Fix $0 < \epsilon < 1$ and $\beta > 0$. Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with*

$$M \geq \left(\frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3}\right) \ln(\#Q).$$

If $M \leq N$, then, with probability exceeding $1 - (\#Q)^{-\beta}$, the following statement holds: For every $x, y \in Q$,

$$(1 - \epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x - \Phi y\|_2}{\|x - y\|_2} \leq (1 + \epsilon)\sqrt{\frac{M}{N}}.$$

At first glance there are several apparent differences between CS (which deals with embedding an *uncountable* point set and correctly identifying a signal from its projections) and the JL lemma (which deals only with embedding a *finite* number of points and makes no mention of signal recovery). However, for the purpose of ensuring a stable CS embedding, Σ_K (because of its limited complexity) can be characterized in terms of a finite number point samples. By applying the JL lemma only to these points we can deduce the RIP for all of the remaining points on Σ_K, which in turn permits stable CS signal recovery. The required number M of random projections matches the known asymptotics for the previous RIP theory.

1.2.3 Manifold models

Manifold signal models generalize the notion of concise signal structure beyond the framework of bases and representations. These models arise in more broad cases where (i) a K-dimensional parameter θ can be identified that carries the relevant information about a signal and (ii) the signal $x_\theta \in \mathbb{R}^N$ changes as a continuous (typically nonlinear) function of these parameters. In general, this dependence may not be neatly reflected in a sparse set of transform coefficients. Some simple explicit examples include:

---

[2]By an orthoprojector, we mean an orthogonal projection from $\mathbb{R}^N$ to $\mathbb{R}^M$ that can be expressed as an $M \times N$ matrix Φ with orthonormal rows. A random orthoprojector may be constructed, for example, by running the Gram-Schmidt process on M random length-N vectors having i.i.d. Gaussian entries (assuming the vectors are linearly independent). We note also that other formulations of the CS problem pertain to random $M \times N$ matrices Φ with Gaussian entries, or to matrices renormalized by $\sqrt{N/M}$ to ensure that $\|\Phi x\|_2 \approx (1 \pm \epsilon) \|x\|_2$. However, we find it useful in this paper to work with random orthoprojectors and the resulting "compaction" by $\sqrt{M/N}$.

- time delay of a 1-D signal (parametrized by 1 variable for translation);

- amplitude, phase, and frequency of a pure sinusoid (3 parameters);

- starting and ending time and frequency of a linear radar chirp (4 parameters);

- local signal parameters such as the configuration of a straight edge in a small image segment (2 parameters: slope and offset);

- global parameters affecting signal generation such as the position of a camera or microphone recording a scene or the relative placement of objects/speakers in a scene; and

- parameters governing the output of some articulated physical system [23, 26, 45].

In these and many other cases, the geometry of the signal class forms a nonlinear K-dimensional *submanifold* of $\mathbb{R}^N$,

$$\mathcal{F} = \{x_\theta : \theta \in \Theta\},$$

where Θ is the K-dimensional parameter space.[3] (Note the dimension K can be interpreted as an "information level" of the signal, analogous to the sparsity level in Section 1.2.2 that we also denote by K.) Low-dimensional manifolds have also been proposed as approximate models for nonparametric signal classes such as images of human faces or handwritten digits [7, 28, 42].

Most algorithms for dimensionality reduction of manifold-modeled signals involve "learning" the manifold structure from a collection of data points, typically by constructing *nonlinear* mappings from $\mathbb{R}^N$ to $\mathbb{R}^M$ for some $M < N$ (ideally $M = K$) that are *adapted* to the training data and intended to preserve some characteristic property of the manifold. Example algorithms include ISOMAP [40], Hessian Eigenmaps (HLLE) [24], and Maximum Variance Unfolding (MVU) [46], which attempt to learn isometric embeddings of the manifold (preserving pairwise geodesic distances); Locally Linear Embedding (LLE) [36], which attempts to preserve local linear neighborhood structures among the embedded points; Local Tangent Space Alignment (LTSA) [47], which attempts to preserve local coordinates in each tangent space; and a method for charting a manifold [5] that attempts to preserve local neighborhood structures. These algorithms can be useful for learning the dimension and parametrizations of manifolds, for sorting data, for visualization and navigation through the data, and as preprocessing to make further analysis more tractable; common demonstrations include analysis of face images and classification of handwritten digits. A related technique, the Whitney Reduction Network [6, 7], uses a training data set to adaptively construct a *linear* mapping from $\mathbb{R}^N$ to $\mathbb{R}^M$ that attempts to preserve ambient pairwise distances on the manifold; this is particularly useful for processing the output of dynamical systems having low-dimensional attractors. Additional algorithms have also been proposed for characterizing manifolds from sampled data without constructing an explicit embedding in $\mathbb{R}^M$ [13, 15, 33] and for constructing functions on the point samples in $\mathbb{R}^N$ that reflect the intrinsic structure of the manifold [4, 14]. Naturally, the burden of storing the sampled data points and implementing any of these manifold learning algorithms increases with the native dimension N of the data.

1.3 Contributions: Universal dimensionality reduction for manifold models via random projections

In this paper, we propose a new approach for *nonadaptive, universal dimensionality reduction* of manifold-modeled data, demonstrating that small numbers of *random linear projections* can

---

[3]In general, Θ itself can be a K-dimensional manifold and need not be a subset of $\mathbb{R}^K$. We refer the reader to [35] for an excellent overview of several manifolds with relevance to signal processing, including the rotation group SO(3), which can be used for representing orientations of objects in 3-D space.

preserve key information about manifold-modeled signals. As we discuss, these projections can in turn be used either to recover faithful approximations to manifold-modeled signals or to discern key properties about the manifold.

We center our analysis on the effect of a random linear projection operator $\Phi : \mathbb{R}^N \to \mathbb{R}^M$ on a smooth K-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^N$. Our main theoretical contribution (Theorem 3.1) is an RIP/JL-like result that ensures a stable embedding of the manifold in $\mathbb{R}^M$. In particular, we establish a sufficient number M of random projections to guarantee that, with high probability, all pairwise distances between points on $\mathcal{M}$ are well-preserved under the mapping Φ. Like the fundamental bound in CS, our requisite M is linear in the "information level" K and logarithmic in the ambient dimension N; additionally we identify a logarithmic dependence on the volume and curvature of the manifold. Although the manifold itself consists of an uncountable number of points, we again exploit its low-dimensional structure to specify an effective sampling of points drawn from the manifold (plus its tangent spaces), employ the JL lemma to ensure these points are well embedded, and generalize the result to the remaining points on the manifold.

Our results suggest that, in contrast with most techniques in manifold learning, the essential information in many manifold-modeled signals can be captured via a dimensionality reducing mapping that is both *linear* and *nonadaptive*, requiring no training on sampled data and only rudimentary knowledge of $\mathcal{M}$ itself. Additionally, our results suggest that, for processing large volumes of data concentrated on manifolds, the number of requisite dimensions for a structure-preserving mapping should derive from the properties of the manifold itself, rather than the number of data points (in contrast to the JL lemma). As we discuss, these facts have promising implications in both (i) CS recovery, which can be extended beyond sparse signals to include manifold-modeled signals, and (ii) manifold learning, in which the computational and/or storage requirements can be lessened by identifying key manifold characteristics from lower-dimensional data.

The remainder of this paper is organized as follows. In Section 2, we set our notation and explain our regularity assumptions on the manifold. We state and prove our main results in Section 3. We then conclude in Section 4 by discussing possible applications of these results in CS and manifold learning.

2 Preliminaries

2.1 Notation

Let $d_{\mathcal{M}}(x, y)$ denote the geodesic distance between two points x and y on a K-dimensional manifold $\mathcal{M}$. Let Tan_x denote the K-dimensional tangent space to $\mathcal{M}$ at the point $x \in \mathcal{M}$. (We use the convention that Tan_x is shifted to have origin 0, rather than x.)

2.2 Condition number

To give ourselves a firm footing for analysis, we must assume a certain regularity of the manifold. For this purpose, we adopt the *condition number* defined recently by Niyogi et al. [33].

Definition 2.1 [33] *Let $\mathcal{M}$ be a compact Riemannian submanifold of $\mathbb{R}^N$. The* condition number *is defined as $1/\tau$, where τ is the largest number having the following property: The open normal bundle about $\mathcal{M}$ of radius r is embedded in $\mathbb{R}^N$ for all $r < \tau$.*

The embedding described above of the open normal bundle of radius r is a tubular neighborhood of $\mathcal{M}$ in $\mathbb{R}^N$: $\mathrm{Tub}_r = \{x + \eta \in \mathbb{R}^N : x \in \mathcal{M}, \eta \in \mathrm{Tan}_x^\perp, \|\eta\|_2 < r\}$, where $\mathrm{Tan}_x^\perp$ denotes the set

of vectors normal to Tan_x. Intuitively, supposing that $r < \tau$, this tubular neighborhood does not intersect itself. More precisely, every $p \in \text{Tub}_r$ corresponds to a unique $x \in \mathcal{M}$. (See also [33] for a discussion relating the condition number to the medial axis transform.)

The condition number $1/\tau$ thus controls both local properties of the manifold (such as curvature) and global properties (such as self-avoidance). The precise properties of the manifold that we require can each be derived as a consequence of the condition number. We list these explicitly below.

Lemma 2.1 [33] *(Curvature) Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two distinct points on $\mathcal{M}$, and let $\gamma(t)$ denote a unit speed parameterization of the geodesic path joining p and q. Then the curvature of γ is bounded by $1/\tau$.*

Lemma 2.2 [33] *(Twisting of tangent spaces) Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two distinct points on $\mathcal{M}$ with geodesic distance given by $d_{\mathcal{M}}(p, q)$. Let θ be the angle between the tangent spaces Tan_p and Tan_q defined by $\cos(\theta) = \min_{u \in \text{Tan}_p} \max_{v \in \text{Tan}_q} |\langle u, v \rangle|$. Then $\cos(\theta) > 1 - \frac{1}{\tau} d_{\mathcal{M}}(p, q)$.*

Lemma 2.3 [33] *(Self-avoidance) Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points such that $\|p - q\|_2 = d$. Then for all $d \le \tau/2$, the geodesic distance $d_{\mathcal{M}}(p, q)$ is bounded by $d_{\mathcal{M}}(p, q) \le \tau - \tau\sqrt{1 - 2d/\tau}$.*

From Lemma 2.3 we have an immediate corollary.

Corollary 2.1 *Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points such that $\|p - q\|_2 = d$. If $d \le \tau/2$, then $d \ge d_{\mathcal{M}}(p, q) - \frac{(d_{\mathcal{M}}(p,q))^2}{2\tau}$.*

2.3 Covering regularity

We also introduce a notion of "geodesic covering regularity" for a manifold.

Definition 2.2 *Let $\mathcal{M}$ be a compact Riemannian submanifold of $\mathbb{R}^N$. Given $T > 0$, the geodesic covering number $G(T)$ of $\mathcal{M}$ is defined as the smallest number such that there exists a set A of points on $\mathcal{M}$, $\#A = G(T)$, so that for all $x \in \mathcal{M}$,*

$$\min_{a \in A} d_{\mathcal{M}}(x, a) \le T.$$

Definition 2.3 *Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having volume V. We say that $\mathcal{M}$ has geodesic covering regularity R for resolutions $T \le T_0$ if*

$$G(T) \le \frac{R^K V K^{K/2}}{T^K} \tag{2}$$

for all $0 < T \le T_0$.

The bound described in (2) reflects a natural scaling relationship as balls of radius T are used to cover a K-dimensional surface; indeed the geodesic covering regularity of the manifold is closely related to its more traditional ambient-distance covering number $C(T)$. In fact, for a manifold with condition number $1/\tau$, we can make this connection explicit. Lemma 2.3 implies that for small d, $d_{\mathcal{M}}(p, q) \le \tau - \tau\sqrt{1 - 2d/\tau} \le \tau(1 - (1 - 2d/\tau)) = 2d$. This implies that $G(T) \le C(T/2)$ for small T. Lemmata 5.2 and 5.3 of [33] then provide a means to bound $C(T/2)$, using the bounded curvature

of the manifold to ensure that any ball centered on the manifold captures a certain amount of its volume. In particular it follows that

$$G(T) \leq C(T/2) \leq \frac{V}{\cos(\arcsin(\frac{T}{8\tau}))^K \mathrm{vol}(B_{T/4}^K)} \leq \frac{V \cdot \Gamma(K/2+1)}{(1-(\frac{T}{8\tau})^2)^{K/2}\pi^{K/2}(T/4)^K} \leq \frac{3^K V K^{K/2}}{T^K}$$

for T sufficiently small (e.g., $T < \tau$). Although we point out this connection between the geodesic covering regularity and the condition number, for future reference and flexibility we prefer to specify these as distinct properties in our main result below. For notational brevity in the remainder of this paper, we neglect the minor implicit dependence of the geodesic covering regularity R on the maximum resolution T_0.

3   Main Results

3.1   Theory

The following result establishes a sufficient number of random projections to ensure a satisfactory embedding of a well-conditioned manifold.

Theorem 3.1 *Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having condition number $1/\tau$, volume V, and geodesic covering regularity R. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with*

$$M = O\left(\frac{K \log(NVR\tau^{-1}\epsilon^{-1})\log(1/\rho)}{\epsilon^2}\right). \tag{3}$$

If $M \leq N$, then with probability at least $1 - \rho$ the following statement holds: For every pair of points $x, y \in \mathcal{M}$,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x - \Phi y\|_2}{\|x - y\|_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}}. \tag{4}$$

Proof: See Section 3.2.

Theorem 3.1 concerns the preservation of pairwise ambient distances on the manifold; this can be immediately extended to geodesic distances as well.

Corollary 3.1 *Let $\mathcal{M}$ and Φ be as in Theorem 3.1. Assuming (4) holds for all pairs of points on $\mathcal{M}$, then for every pair of points $x, y \in \mathcal{M}$,*

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{d_{\Phi\mathcal{M}}(\Phi x, \Phi y)}{d_{\mathcal{M}}(x,y)} \leq (1+\epsilon)\sqrt{\frac{M}{N}}, \tag{5}$$

where $d_{\Phi\mathcal{M}}(\Phi x, \Phi y)$ denotes the geodesic distance between the projected points on the image of $\mathcal{M}$.

Proof: See Section 3.3.

Sections 3.2 and 3.3 present proofs of Theorem 3.1 and Corollary 3.1, respectively, and Section 3.4 follows with some brief remarks.

3.2 Proof of Theorem 3.1

A quick sketch of the proof is as follows. After a few preliminary notes in Section 3.2.1, we first specify in Section 3.2.2 a high-resolution sampling of points on the manifold. At each of these points we consider the tangent space to the manifold and specify a sampling of points drawn from this space as well. We then employ the JL lemma to ensure an embedding with satisfactory preservation of all pairwise distances between these points.

Based on the preservation of these pairwise distances, in Section 3.2.3 we then ensure isometry for all tangents to the sampled points and then (using the bounded twisting of tangent spaces) ensure isometry for all tangents at all points on the manifold.

Finally, in Section 3.2.4, we establish pairwise distance preservation between all points on the manifold. For nearby points, our results rely on the tangent space isometries and the bounded curvature of the manifold. For distant points, we essentially rely on the original pairwise distance preservation between the sample points (plus their nearby tangent points). We conclude the proof in Section 3.2.5.

3.2.1 Preliminaries

Consider an orthoprojector $\Phi : \mathbb{R}^N \to \mathbb{R}^M$. For shorthand, we say a point $x \in \mathbb{R}^N$ has "isometry constant ϵ" if the following condition is met:

$$(1 - \epsilon)\sqrt{M/N}\,\|x\|_2 \leq \|\Phi x\|_2 \leq (1 + \epsilon)\sqrt{M/N}\,\|x\|_2\,.$$

We say a set has isometry constant ϵ if the above condition is met for every point in the set. We say a point $x \in \mathbb{R}^N$ has "squared isometry constant ϵ" if the following condition is met:

$$(1 - \epsilon)(M/N)\,\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \epsilon)(M/N)\,\|x\|_2^2\,.$$

These notions are very similar — an isometry constant of ϵ implies a squared isometry constant of 3ϵ, and a squared isometry constant of ϵ implies an isometry constant of ϵ.

We also note that Φ is a nonexpanding operator (by which we mean that $\|\Phi\|_2 \leq 1$, i.e., $\|\Phi x\|_2 \leq \|x\|_2$ for all $x \in \mathbb{R}^N$).

We will also find the following inequalities useful throughout:

$$\frac{1}{1-s} \leq (1 + 2s), \quad 0 \leq s \leq 1/2, \tag{6}$$

and

$$\frac{1}{1+s} \geq (1 - s), \quad s \geq 0. \tag{7}$$

3.2.2 Sampling the manifold

We begin by specifying a high-resolution (but finite) sampling of points on and near the manifold and examining the behavior of these points under the mapping Φ. Our construction will rely on the following variables, to which we will assign specific values in Section 3.2.5: $T > 0$ and $\delta > 0$ control the resolution of the sampling, while $\epsilon_1 \geq 2\delta$ affects the embedding of the points.

First, let A be a minimal set of points *on the manifold* such that, for every $x \in \mathcal{M}$,

$$\min_{a \in A} d_{\mathcal{M}}(x, a) \leq T. \tag{8}$$

We call A the set of *anchor points*. From (2) we have that $\#A \leq \frac{R^K V K^{K/2}}{T^K}$.

9

Second, let B be a set of points *on and near the manifold* constructed as follows. For each anchor point $a \in A$ we consider the tangent space Tan_a to $\mathcal{M}$ at a. We construct a covering of points $Q_1(a) \subset \text{Tan}_a$ such that $\|q\|_2 \leq 1$ for all $q \in Q_1(a)$ and such that for every $u \in \text{Tan}_a$ with $\|u\|_2 \leq 1$,

$$\min_{q \in Q_1(a)} \|u - q\|_2 \leq \delta.$$

This can be accomplished with $\#Q_1(a) \leq (3/\delta)^K$ (see e.g. Chapter 13 of [31]). We then define the renormalized set

$$Q_2(a) = \{Tq : \ q \in Q_1(a)\}$$

and note that $\|q\|_2 \leq T$ for all $q \in Q_2(a)$ and that for every $u \in \text{Tan}_a$ with $\|u\|_2 \leq T$,

$$\min_{q \in Q_2(a)} \|u - q\|_2 \leq T\delta. \tag{9}$$

We now define the set

$$B = \bigcup_{a \in A} \{a\} \cup (a + Q_2(a)),$$

where $a + Q_2(a)$ denotes the set of tangents anchored at the point a (rather than at 0).

Now let $\beta = -\ln(\rho)$, set

$$M \geq \left(\frac{4 + 2\beta}{\epsilon_1^2/2 - \epsilon_1^3/3} \right) \ln(\#B), \tag{10}$$

and let Φ be as specified in Theorem 3.1. According to Lemma 1.1 (Johnson-Lindenstrauss), with probability exceeding $1 - (\#B)^{-\beta} > 1 - \rho$, the following statement holds: For all $u, v \in B$, the difference vector $(u - v)$ has isometry constant ϵ_1. We assume this to hold and must now extend it to show (4) for every $x, y \in \mathcal{M}$.

3.2.3 Isometry constants of tangent planes

To proceed, we must first examine the behavior of tangents to the manifold under the mapping Φ. We immediately have that for every $a \in A$, every $q \in Q_2(a)$ has isometry constant ϵ_1, and because Φ is linear, every $q \in Q_1(a)$ also has isometry constant ϵ_1. Following the derivation in Lemma 5.1 of [2] (and recalling that we assume $\delta \leq \epsilon_1/2$), we have that for all $a \in A$, the tangent space Tan_a has isometry constant

$$\epsilon_2 := 2\epsilon_1,$$

where := denotes equality by definition. That is, for every $a \in A$, every $u \in \text{Tan}_a$ has isometry constant ϵ_2.

Using the bounded twisting of tangent spaces (Lemma 2.2), we can then generalize this isometry result from the anchored tangent spaces to arbitrary tangent spaces on the manifold.

Lemma 3.1 *Let Φ be as specified in Section 3.2.2, and suppose $T/\tau < 1/4$. Then for every $x \in \mathcal{M}$, every $u \in \text{Tan}_x$ has isometry constant*

$$\epsilon_3 := \epsilon_2 + \sqrt{\frac{3TN}{\tau M}}.$$

Proof: See Appendix A.

3.2.4 Differences between points on the manifold

We are now ready to consider the behavior of difference vectors connecting two points on the manifold. We divide our analysis into two cases: (i) for points nearby in terms of geodesic distance, we rely on the bounded curvature of geodesic paths on $\mathcal{M}$ (Lemma 2.1) and on the tangent space isometry constants (Lemma 3.1), and (ii) for points well-separated in terms of geodesic distance, we rely on the fact that such points are also well-separated in ambient distance (Lemma 2.3) and on the isometry constants of pairwise distances between points in the set B (Section 3.2.2). In the following lemmata, we use $C_1 T$ as a demarcation between "nearby" and "distant" geodesic distance; we assign a specific value for C_1 in Section 3.2.5.

Lemma 3.2 *Let Φ be as specified in Section 3.2.2, let $C_1 > 0$, and suppose $C_1 T/\tau < 1/2$. Let x_1 and x_2 be two points on $\mathcal{M}$ separated by geodesic distance $d_{\mathcal{M}}(x_1, x_2) \leq C_1 T$. Then the difference vector $x_1 - x_2$ has isometry constant*

$$\epsilon_4 := \epsilon_3 + C_1 T \sqrt{N/M}/\tau + 2C_1 T/\tau + 2\epsilon_3 C_1 T/\tau + 2C_1^2 T^2 \sqrt{N/M}/\tau^2.$$

Proof: See Appendix B.

Lemma 3.3 *Let Φ be as specified in Section 3.2.2, and suppose $C_1 \geq 10$, $T \leq \tau/C_1$, and $\delta \leq 1/4$. Let x_1 and x_2 be two points on $\mathcal{M}$ separated by geodesic distance $d_{\mathcal{M}}(x_1, x_2) > C_1 T$. Then the difference vector $x_1 - x_2$ has isometry constant*

$$\epsilon_6 := 9\epsilon_5 + \frac{36\epsilon_5 T}{\tau} + \frac{2T}{\tau} + \frac{8T^2}{\tau^2} + \frac{2T\sqrt{N/M}}{\tau} + \frac{8T^2\sqrt{N/M}}{\tau^2}$$

where

$$\epsilon_5 := \epsilon_1 + 4\delta + 4\delta\epsilon_1 + 2\delta\sqrt{N/M} + 8\delta^2\sqrt{N/M}.$$

Proof: See Appendix C.

3.2.5 Synthesis

Finally, we specify values for C_1, ϵ_1, T, and δ that will ensure an isometry constant of ϵ for all pairwise distances on the manifold. Thus far, we have established the following isometry constants. For nearby points (Lemma 3.2) we have isometry constant

$$
\begin{aligned}
\epsilon_4 &= \epsilon_3 + \frac{C_1 T}{\tau}\sqrt{\frac{N}{M}} + \frac{2C_1 T}{\tau} + \frac{2\epsilon_3 C_1 T}{\tau} + \frac{2C_1^2 T^2}{\tau^2}\sqrt{\frac{N}{M}} \\
&= 2\epsilon_1 + \frac{4\epsilon_1 C_1 T}{\tau} + \sqrt{\frac{3TN}{\tau M}} + \frac{C_1 T}{\tau}\sqrt{\frac{N}{M}} + \frac{2C_1 T}{\tau} + \frac{2C_1 T}{\tau}\sqrt{\frac{3TN}{\tau M}} + \frac{2C_1^2 T^2}{\tau^2}\sqrt{\frac{N}{M}}.
\end{aligned}
$$

For distant points (Lemma 3.3) we have isometry constant

$$\epsilon_6 \;=\; 9\epsilon_5 + \frac{36\epsilon_5 T}{\tau} + \frac{2T}{\tau} + \frac{8T^2}{\tau^2} + \frac{2T}{\tau}\sqrt{\frac{N}{M}} + \frac{8T^2}{\tau^2}\sqrt{\frac{N}{M}}$$

$$\begin{aligned}
=\;& 9\epsilon_1 + 36\delta + 36\delta\epsilon_1 + 18\delta\sqrt{N/M} + 72\delta^2\sqrt{N/M} \\
&+ \frac{36\epsilon_1 T}{\tau} + \frac{144\delta T}{\tau} + \frac{144\delta\epsilon_1 T}{\tau} + \frac{72\delta T}{\tau}\sqrt{\frac{N}{M}} + \frac{288\delta^2 T}{\tau}\sqrt{\frac{N}{M}} \\
&+ \frac{2T}{\tau} + \frac{8T^2}{\tau^2} + \frac{2T}{\tau}\sqrt{\frac{N}{M}} + \frac{8T^2}{\tau^2}\sqrt{\frac{N}{M}}.
\end{aligned}$$

We must ensure that both $\epsilon_4 \leq \epsilon$ and $\epsilon_6 \leq \epsilon$.

We first set $C_1 = 10$. For constants C_2, C_3, and C_4 (which we will soon specify), we let

$$\epsilon_1 = C_2\epsilon, \quad T = \frac{C_3\epsilon^2\tau}{N}, \quad \text{and} \quad \delta = \frac{C_4\epsilon}{\sqrt{N}}.$$

Plugging in to the above and using the fact that $\epsilon < 1$, we have

$$\begin{aligned}
\epsilon_4 \;=\;& 2C_2\epsilon + \frac{40C_2C_3\epsilon^3}{N} + \sqrt{\frac{3C_3\epsilon^2}{M}} + \frac{10C_3\epsilon^2}{N}\sqrt{\frac{N}{M}} + \frac{20C_3\epsilon^2}{N} \\
&+ \frac{20C_3\epsilon^2}{N}\sqrt{\frac{3C_3\epsilon^2}{M}} + \frac{200C_3^2\epsilon^4}{N^2}\sqrt{\frac{N}{M}} \\
\leq\;& \epsilon(2C_2 + 40C_2C_3 + \sqrt{3C_3} + 30C_3 + 20\sqrt{3}C_3\sqrt{C_3} + 200C_3^2)
\end{aligned}$$

and

$$\begin{aligned}
\epsilon_6 \;=\;& 9C_2\epsilon + \frac{36C_4\epsilon}{\sqrt{N}} + \frac{36C_2C_4\epsilon^2}{\sqrt{N}} + \frac{18C_4\epsilon}{\sqrt{M}} + \frac{72C_4^2\epsilon^2}{\sqrt{NM}} \\
&+ \frac{36C_2C_3\epsilon^3}{N} + \frac{144C_3C_4\epsilon^3}{N\sqrt{N}} + \frac{144C_2C_3C_4\epsilon^4}{N\sqrt{N}} + \frac{72C_3C_4\epsilon^3}{N\sqrt{M}} + \frac{288C_3C_4^2\epsilon^4}{N\sqrt{NM}} \\
&+ \frac{2C_3\epsilon^2}{N} + \frac{8C_3^2\epsilon^4}{N^2} + \frac{2C_3\epsilon^2}{\sqrt{NM}} + \frac{8C_3^2\epsilon^4}{N\sqrt{NM}} \\
\leq\;& \epsilon(9C_2 + 36C_4 + 36C_2C_4 + 18C_4 + 72C_4^2 + 36C_2C_3 + 144C_3C_4 \\
&+ 144C_2C_3C_4 + 72C_3C_4 + 288C_3C_4^2 + 2C_3 + 8C_3^2 + 2C_3 + 8C_3^2).
\end{aligned}$$

We now must set the constants C_2, C_3, and C_4 to ensure that $\epsilon_4 \leq \epsilon$ and $\epsilon_6 \leq \epsilon$. Due to the role of ϵ_1 in determining our ultimate bound on M, we wish to be most aggressive in setting the constant C_2. To ensure $\epsilon_6 \leq \epsilon$, we must set $C_2 < 1/9$; for neatness we choose $C_2 = 1/10$. For the remaining constants we may choose $C_3 = 1/1900$ and $C_4 = 1/633$ and confirm that both $\epsilon_4 \leq \epsilon$ and $\epsilon_6 \leq \epsilon$. One may also verify that, by using these constants, all of our assumptions in Section 3.2.2, Lemma 3.1, Lemma 3.2, and Lemma 3.3 are met (in particular, that $\epsilon_1 \geq 2\delta$, $T/\tau < 1/4$, $C_1T/\tau < 1/2$, $C_1 \geq 10$, $T \leq \tau/C_1$, and $\delta \leq 1/4$).

We can now determine the size of the set B, which according to (10) will dictate the required

number of measurements M. We have

$$
\begin{aligned}
\#B &\leq \sum_{a \in A}(1 + \#Q_2(a)) = \sum_{a \in A}(1 + \#Q_1(a)) \\
&\leq \left(\frac{R^K V K^{K/2}}{T^K}\right)(1 + (3/\delta)^K) \\
&\leq \left(\frac{R^K V K^{K/2}}{T^K}\right)\left(1 + \left(\frac{3 \cdot 633\sqrt{N}}{\epsilon}\right)^K\right) \\
&\leq \left(\frac{R^K V K^{K/2} 1900^K N^K}{\epsilon^{2K} \tau^K}\right)\left(\frac{(3 \cdot 633 + 1)^K N^{K/2}}{\epsilon^K}\right).
\end{aligned}
$$

Plugging in to (10), we require

$$
\begin{aligned}
M &\geq \left(\frac{4 + 2\beta}{\epsilon_1^2/2 - \epsilon_1^3/3}\right)\ln(\#B) \\
&\geq \left(\frac{4 - 2\ln(\rho)}{\epsilon^2/200 - \epsilon^3/3000}\right)\ln\left(\frac{1900^{2K} K^{K/2} N^{3K/2} R^K V}{\epsilon^{3K} \tau^K}\right). \tag{11}
\end{aligned}
$$

This completes the proof of Theorem 3.1. $\qquad\square$

3.3 Proof of Corollary 3.1

The corollary follows simply from the fact that length of a smooth curve on the manifold can be written as a limit sum of ambient distances between points on that curve and the observation that (4) can be applied to each of these distances.

So if we let $x, y \in \mathcal{M}$, define $\mu = d_{\mathcal{M}}(x, y)$, and let γ denote the unit speed geodesic path joining x and y on $\mathcal{M}$ in $\mathbb{R}^N$, then the length of the image of γ along $\Phi\mathcal{M}$ in $\mathbb{R}^M$ will be bounded above by $(1 + \epsilon)\sqrt{M/N}\mu$. Hence, $d_{\Phi\mathcal{M}}(\Phi x, \Phi y) \leq (1 + \epsilon)\sqrt{M/N}d_{\mathcal{M}}(x, y)$.

Similarly, if we let $x, y \in \mathcal{M}$, define $\mu_\Phi = d_{\Phi\mathcal{M}}(\Phi x, \Phi y)$, and let γ_Φ denote the unit speed geodesic path joining Φx and Φy on the image of $\mathcal{M}$ in $\mathbb{R}^M$, then the length of the preimage of γ_Φ is bounded above by $\frac{1}{1-\epsilon}\sqrt{N/M}\mu_\Phi$. Hence,

$$
d_{\mathcal{M}}(x, y) \leq \frac{1}{1 - \epsilon}\sqrt{N/M}\mu_\Phi,
$$

which implies that

$$
d_{\Phi\mathcal{M}}(\Phi x, \Phi y) \geq (1 - \epsilon)\sqrt{M/N}d_{\mathcal{M}}(x, y).
$$

$\qquad\square$

3.4 Remarks

We offer some brief remarks on these results.

1. Like the fundamental bound in Compressed Sensing, the requisite number of random projections M to ensure a stable embedding of the manifold is linear in the "information level" K and logarithmic in the ambient dimension N; additionally we identify a logarithmic dependence on the volume and curvature of the manifold.

2. The proof of Theorem 3.1 in fact establishes the bound (3) up to actual constants; see (11) for the complete result.

3. The $\ln(1/\rho)$ factor in the numerator of (3) and (11) can be immediately sharpened to

$$\frac{\ln(1/\rho)}{\ln\left(\frac{1900^{2K}K^{K/2}N^{3K/2}R^{K}V}{\epsilon^{3K}\tau^{K}}\right)}$$

to dramatically reduce the dependence on the failure probability ρ. (This follows simply from Lemma 1.1 and a more careful accounting in Section 3.2.2 of the proof.)

4. The constant 200 appearing in (11) can likely be improved by increasing C_1 and by sharpening Lemma 3.3.

5. Our results can also be extended to include other random distributions for Φ, such as the case where Φ is a random $M \times N$ matrix with i.i.d. $\mathcal{N}(0,\sigma^2)$ entries, where $\sigma^2 = 1/N$. (See [2] for other possibilities.) In order to adapt the proof, one must account for the fact that Φ may no longer be nonexpanding; however with high probability the norm $\|\Phi\|_2$ can be bounded by a small constant.

4 Stylized Applications

We conclude with a brief discussion of possible applications of these results in signal acquisition and processing.

4.1 Compressed Sensing for Manifolds

First, we consider a generalization of Compressed Sensing (CS) that takes advantage of the stable embedding of the manifold into $\mathbb{R}^M$. In the traditional CS problem, recovery of a signal $x \in \mathbb{R}^N$ from its measurement $y = \Phi x \in \mathbb{R}^M$ (where $M < N$) is possible thanks to a model that characterizes *where* in $\mathbb{R}^N$ the signal x might be expected to reside. In particular, this model comes as an assumption of sparsity in the dictionary Ψ, and the sparsity level K dictates the requisite number of measurements M to ensure stable recovery of sparse signals x. As we discussed in Section 1.2.2, recovery of sparse signals is possible because the signal set Σ_K has a stable embedding under Φ.

We have also established in this paper, however, that a signal manifold $\mathcal{M}$ can be stably embedded under a random projection from $\mathbb{R}^N$ to $\mathbb{R}^M$ for some M. This suggests that *signals obeying manifold models can also be recovered from CS measurements*, simply by replacing the traditional CS model of sparsity with a manifold model for x. Moreover, the requisite number M of measurements should depend only on the properties of the manifold $\mathcal{M}$ on which the signal x is hypothesized to live, and not on the sparsity level of x in any particular dictionary Ψ. To make this more concrete, suppose that x lives exactly on a compact submanifold $\mathcal{M} \subset \mathbb{R}^N$ with a particular dimension and condition number. Theorem 3.1 then suggests a sufficient number M of measurements so that (with high probability),[4] the operator Φ embeds $\mathcal{M}$ into $\mathbb{R}^M$. This implies also that x will be uniquely identifiable from its projection $y = \Phi x$.

---

[4]Whitney's Embedding Theorem [29] actually suggests for certain K-dimensional manifolds that the number of measurements need be no larger than $2K + 1$ to ensure an embedding. However, as it offers no guarantee of stability, the practical recovery of a signal on the manifold could be complicated. Interestingly, the number $2K$ also arises in CS as the number of random measurements required to embed the set of all K-sparse signals (but again with no guarantee on the conditioning of the recovery problem); see [3].

Beyond simple invertibility, however, Theorem 3.1 also indicates a certain "stability" to the embedding of $\mathcal{M}$ in $\mathbb{R}^M$ (specifically, in the preservation of ambient distances). This stability can be useful for guaranteeing approximate recovery of signals that live *near* the manifold $\mathcal{M}$, rather than exactly on it. In particular, suppose $x \notin \mathcal{M}$, and let x^* be the "nearest neighbor" to x on $\mathcal{M}$, i.e.,

$$x^* = \arg \min_{x' \in \mathcal{M}} \left\| x - x' \right\|_2, \tag{12}$$

supposing that this point is uniquely defined. Let $\widehat{x}$ be the "approximate nearest neighbor" as estimated in the projected space $\mathbb{R}^M$, i.e.,

$$\widehat{x} = \arg \min_{x' \in \mathcal{M}} \left\| \Phi x - \Phi x' \right\|_2. \tag{13}$$

This point $\widehat{x}$ could be thought of as a "recovered" CS approximation to x. To consider this recovery successful, we would like to guarantee that $\|x - \widehat{x}\|_2$ is not much larger than $\|x - x^*\|_2$. Such a guarantee comes again from the JL lemma. Assuming that the random orthoprojector Φ is statistically independent of the signal x, then we may recall Section 3.2.2 of the proof and consider the embedding of the set $\{x\} \cup B$ under Φ. With high probability,[5] each pairwise distance in this set will have isometry constant ϵ_1. Hence, the distance from x to each anchor point will be well-preserved, and since every manifold point is no more than distance T from an anchor point, then (assuming $\|x - x^*\|_2$ is sufficiently larger than T) the distance from x to every point on $\mathcal{M}$ will be well-preserved. This guarantees a satisfactory recovery $\widehat{x}$ in the approximate nearest neighbor problem. (By examining, for example, the tangent spaces, this can all be made more precise and extended to consider the case where $\|x - x^*\|_2$ is small.)

The question remains of how a recovery program such as (13) can be efficiently solved. Unfortunately it is difficult to provide a single general-purpose algorithm, as certain nuances (such as topology) of the individual manifold can cause problem-specific challenges. (This is true even when solving (12) with the full data $x \in \mathbb{R}^N$.) However, given a suitable initial guess for the solution $\widehat{x}$ to (13), local navigation of the manifold, driven for example by an iterative Newton's method [45] to minimize $\|\Phi x - \Phi x'\|_2$, could be used to converge to the true solution. An initial guess for $\widehat{x}$ could be obtained, for example, by solving (13) using only a coarse sampling of points from $\Phi \mathcal{M}$ or by employing prior information discovered through other means. Such prior information is not unreasonable in multi-signal settings or when recovering a single signal in a multiscale fashion [22]. In general, because many key properties of the manifold are preserved, one would expect the challenge of solving (13) in the measurement space $\mathbb{R}^M$ to roughly reflect the difficulty of solving (12) in the initial ambient space $\mathbb{R}^N$.

We refer the reader to [43, 44] for additional discussion of recovery algorithms in more specific contexts such as image processing. These papers also include a series of simple but promising experiments that support manifold-based CS on a more empirical level. Additional research and experiments are ongoing, including an investigation into the relationship between the condition number and certain signal properties such as differentiability [45].

4.2 Manifold learning

Recall from Section 1.2.3 that the basic problem of manifold learning is to discover some information about a manifold based on a collection of data sampled from that manifold. In standard problems,

---

[5]By the addition of an extra point to the embedding, there is a nominal increase in the required number of measurements. This increase becomes much more relevant in the case where a large number of signals x would need to be embedded well with respect to the manifold.

this data is presented in $\mathbb{R}^N$ (the natural ambient signal space). For several reasons it may be desirable to reduce the data dimension N. First of all, the process of acquiring and storing a large number of manifold samples may be difficult when the dimension N is large. Second, the dimension N can also impose a significant computational burden on manifold learning algorithms.

Fortunately, Theorem 3.1 and Corollary 3.1 imply that many of the properties of a manifold $\mathcal{M}$ one may wish to discover from sampled data in $\mathbb{R}^N$ are approximately preserved in its image $\Phi\mathcal{M}$ under a random projection to $\mathbb{R}^M$. Among the properties preserved are:[6]

- ambient and geodesic distances between pairs of points;

- dimension of the manifold;

- topology, local neighborhoods, and local angles;

- lengths and curvature of paths on the manifold; and

- volume of the manifold.

These are some of the basic properties sought by the manifold learning algorithms such as those listed in Section 1.2.3, and so one can simply apply such algorithms to random projections of the original data and get an approximation to the true answer. The quality of the approximation could be weighed against the potential savings in data acquisition, storage, and computation.

To give a simple example, the first step of the canonical ISOMAP algorithm [40] computes pairwise Euclidean distances between some or all of the training points Q. Superficially, this step requires $O(N(\#Q)^2)$ arithmetic operations, though the actual distances are required only for the k nearest neighbors of each point, which potentially lowers the computational burden as far as $O(Nk(\#Q))$. (From these Euclidean distances, geodesic distances are estimated and then used in constructing the embedding. Hence the complexity of all subsequent steps depends only on $\#Q$ and not on N.) However, Theorem 3.1 (and Corollary 3.1) suggest that each Euclidean (geodesic) distance can be accurately estimated in the projected space, which further reduces the computational burden of the first step of ISOMAP to $O(Mk(\#Q))$. Moreover, the requisite number of dimensions M depends only on the properties of the manifold and not the number of training points $\#Q$ (in contrast to the JL lemma).

Of course, to properly measure the true computational savings of manifold learning in $\mathbb{R}^M$ instead of $\mathbb{R}^N$, one would need to account for the complexity[7] of applying the operator Φ. If the data is initially provided in $\mathbb{R}^N$, then the future computational savings may not always warrant the cost of applying Φ. However, in some cases the reduced storage requirement could justify the computational burden. Moreover, to support the developing CS theory, physical "analog-to-information" hardware devices [30, 41] and imaging systems [38] have been proposed that *directly* acquire random projections of signals into $\mathbb{R}^M$ without first sampling the signals in $\mathbb{R}^N$; this eliminates the computational burden of applying the operator Φ. Our results suggest that such devices could be a useful and efficient means for acquiring and storing manifold-modeled data while preserving the relevant structural features within the data. Again, we refer the reader to [43] for promising preliminary experiments.

---

[6]Some of these follow directly from Theorem 3.1 and Corollary 3.1; others depend on the near-isometry of the projected tangent spaces as discussed in Section 3.2.3.

[7]The computational complexity of applying an arbitrary $M \times N$ operator Φ is $O(MN)$ per data point, though this complexity can be reduced through a careful, pseudorandom design of Φ [41].

Acknowledgements

A   Proof of Lemma 3.1

Let x be an arbitrary point on the manifold and let a be its nearest anchor point (in geodesic distance), recalling from (8) that $d_{\mathcal{M}}(x, a) \leq T$. Let $v \in \mathrm{Tan}_x$ with $\|v\|_2 = 1$. From Lemma 2.2 it follows that there exists $u \in \mathrm{Tan}_a$ such that $\|u\|_2 = 1$ and $\cos(\mathrm{angle}(u, v)) > 1 - T/\tau$.

Because $\|u\|_2 = \|v\|_2 = 1$, it follows that $\|u - v\|_2 \leq \mathrm{angle}(u, v)$. Define $\theta := \mathrm{angle}(u, v)$; our bound above specifies that $\cos(\theta) > 1 - T/\tau$. Using a Taylor expansion we have that $\cos(\theta) < 1 - \theta^2/2 + \theta^4/24 = 1 - \theta^2/2(1 - \theta^2/12)$, and because we assume $T/\tau < 1/4$, then $\theta < 2$, which implies $\cos(\theta) < 1 - \theta^2/3$. Combining, we have $1 - \theta^2/3 > \cos(\theta) > 1 - T/\tau$, which implies that $T/\tau > \theta^2/3$, and so $\|u - v\|_2 \leq \theta < \sqrt{3T/\tau}$.

Since $u \in \mathrm{Tan}_a$ with $a \in A$, we recall that u has isometry constant ϵ_2. We aim to determine the isometry constant for v. Using the triangle inequality and the fact that Φ is nonexpanding, we have $\|\Phi v\|_2 \leq \|\Phi u\|_2 + \|\Phi(u - v)\|_2 \leq (1 + \epsilon_2)\sqrt{M/N} + \sqrt{3T/\tau}$. Similarly, $\|\Phi v\|_2 \geq \|\Phi u\|_2 - \|\Phi(u - v)\|_2 \geq (1 - \epsilon_2)\sqrt{M/N} - \sqrt{3T/\tau}$. Since $\|v\|_2 = 1$, this implies that v has isometry constant

$$\epsilon_3 := \epsilon_2 + \sqrt{\frac{3TN}{\tau M}}.$$

Because the choices of x and v were arbitrary, it follows that all tangents to the manifold have isometry constant ϵ_3. $\qquad\square$

B   Proof of Lemma 3.2

Define $\mu := d_{\mathcal{M}}(x_1, x_2) \leq C_1 T$. Let $\gamma(t)$ denote a unit speed parameterization of the geodesic path connecting x_1 and x_2, with $\gamma(0) = x_1$ and $\gamma(\mu) = x_2$. From Lemma 2.1 we have that the curvature of γ is bounded by $1/\tau$. From Taylor's theorem we then have that

$$\gamma(\mu) - \gamma(0) = \mu\gamma'(0) + R_1 \tag{14}$$

where $\gamma'(0)$ denotes the tangent to the curve γ at 0, and where the norm of the remainder obeys $\|R_1\|_2 \leq \mu^2/\tau$. Using the triangle inequality and the fact that $\|\gamma'(0)\|_2 = 1$, we have that

$$(1 - \mu/\tau)\mu \leq \|\gamma(\mu) - \gamma(0)\|_2 \leq (1 + \mu/\tau)\mu, \tag{15}$$

and combining (14) with the isometry constant ϵ_3 of $\gamma'(0)$ and the fact that Φ is nonexpanding we have

$$(1 - (\epsilon_3 + \mu\sqrt{N/M}/\tau))\mu\sqrt{M/N} \leq \|\Phi\gamma(\mu) - \Phi\gamma(0)\|_2 \leq (1 + (\epsilon_3 + \mu\sqrt{N/M}/\tau))\mu\sqrt{M/N}. \tag{16}$$

Combining (15) and (16), the ratio

$$
\frac{\|\Phi\gamma(\mu) - \Phi\gamma(0)\|_2}{\|\gamma(\mu) - \gamma(0)\|_2} \leq \frac{(1 + \epsilon_3 + \mu\sqrt{N/M}/\tau)\mu\sqrt{M/N}}{(1 - \mu/\tau)\mu}
$$

$$
\leq \frac{(1 + \epsilon_3 + C_1 T\sqrt{N/M}/\tau)}{(1 - C_1 T/\tau)}\sqrt{M/N}
$$

$$
\leq (1 + \epsilon_3 + C_1 T\sqrt{N/M}/\tau + 2C_1 T/\tau + 2\epsilon_3 C_1 T/\tau + 2C_1^2 T^2\sqrt{N/M}/\tau^2)\sqrt{\frac{M}{N}}.
$$

In the third step above we have employed (6) and the fact that $C_1 T/\tau < 1/2$. Similarly, the ratio

$$
\frac{\|\Phi\gamma(\mu) - \Phi\gamma(0)\|_2}{\|\gamma(\mu) - \gamma(0)\|_2} \geq \frac{(1 - \epsilon_3 - \mu\sqrt{N/M}/\tau)\mu\sqrt{M/N}}{(1 + \mu/\tau)\mu}
$$

$$
\geq \frac{(1 - \epsilon_3 - C_1 T\sqrt{N/M}/\tau)}{(1 + C_1 T/\tau)}\sqrt{M/N}
$$

$$
\geq (1 - \epsilon_3 - C_1 T\sqrt{N/M}/\tau - C_1 T/\tau + \epsilon_3 C_1 T/\tau + C_1^2 T^2\sqrt{N/M}/\tau^2)\sqrt{M/N}
$$

$$
\geq (1 - \epsilon_3 - C_1 T\sqrt{N/M}/\tau - C_1 T/\tau)\sqrt{M/N}.
$$

Here the third step uses (7). Of the bounds we have now derived, the upper bound is the looser of the two, and so it follows that the difference vector $\gamma(\mu) - \gamma(0) = x_2 - x_1$ has isometry constant

$$
\epsilon_4 := \epsilon_3 + C_1 T\sqrt{N/M}/\tau + 2C_1 T/\tau + 2\epsilon_3 C_1 T/\tau + 2C_1^2 T^2\sqrt{N/M}/\tau^2.
$$

This isometry constant ϵ_4 will hold for any two points on the manifold separated by geodesic distance $\leq C_1 T$. $\qquad\square$

C   Proof of Lemma 3.3

Let a_1 and a_2 be the nearest (in terms of geodesic distance) anchor points to x_1 and x_2, respectively. We consider the geodesic path from a_1 to x_1 and let $u_1 \in \text{Tan}_{a_1}$ denote the tangent to this path at a_1. (For convenience we scale u_1 to have norm $\|u_1\|_2 = T$.) Similarly, we let $u_2 \in \text{Tan}_{a_2}$ denote the tangent at the start of the geodesic path from a_2 to x_2 (choosing $\|u_2\|_2 = T$).

We recall from (9) that there exists $q_1 \in Q_2(a_1)$ such that $\|u_1 - q_1\|_2 \leq T\delta$ and there exists $q_2 \in Q_2(a_2)$ such that $\|u_2 - q_2\|_2 \leq T\delta$. Additionally, the points $a_1 + q_1$ and $a_2 + q_2$ belong to the set B, and so the difference $(a_1 + q_1) - (a_2 + q_2)$ has isometry constant ϵ_1.

Recalling the assumption that $T \leq \tau/C_1$, we consider the ambient distance between x_1 and x_2. We have either that $\|x_1 - x_2\|_2 > \tau/2 \geq C_1 T/2$ or that $\|x_1 - x_2\|_2 \leq \tau/2$, which by Corollary 2.1 would then imply that $\|x_1 - x_2\|_2 \geq d_{\mathcal{M}}(x_1, x_2) - \frac{(d_{\mathcal{M}}(x_1, x_2))^2}{2\tau}$ with $d_{\mathcal{M}}(x_1, x_2) > C_1 T$ by assumption and $d_{\mathcal{M}}(x_1, x_2) \leq \tau - \tau\sqrt{1 - 2\|x_1 - x_2\|_2/\tau} \leq \tau(1 - (1 - 2\|x_1 - x_2\|_2/\tau)) = 2\|x_1 - x_2\|_2 \leq \tau$ by Lemma 2.3. In this range $C_1 T < d_{\mathcal{M}}(x_1, x_2) \leq \tau$, it follows that $\|x_1 - x_2\|_2 \geq d_{\mathcal{M}}(x_1, x_2) - \frac{(d_{\mathcal{M}}(x_1, x_2))^2}{2\tau} > C_1 T/2$. Since we assume $C_1 \geq 10$, then $\|x_1 - x_2\|_2 > 5T$. Using the triangle inequality, $\|a_1 - a_2\|_2 > 3T$ and $\|(a_1 + q_1) - (a_2 + q_2)\|_2 > T$.

Now we consider the isometry constant of $(a_1 + u_1) - (a_2 + u_2)$. Using the triangle inequality and the fact that Φ is nonexpanding, we have

$$
\begin{aligned}
\frac{\|\Phi(a_1 + u_1) - \Phi(a_2 + u_2)\|_2}{\|(a_1 + u_1) - (a_2 + u_2)\|_2} &\leq \frac{\|\Phi(a_1 + q_1) - \Phi(a_2 + q_2)\|_2 + 2T\delta}{\|(a_1 + q_1) - (a_2 + q_2)\|_2 - 2T\delta} \\[2mm]
&\leq \frac{(1 + \epsilon_1)\,\|(a_1 + q_1) - (a_2 + q_2)\|_2\,\sqrt{M/N} + 2T\delta}{\|(a_1 + q_1) - (a_2 + q_2)\|_2 - 2T\delta} \\[2mm]
&= \frac{(1 + \epsilon_1)\sqrt{M/N} + 2T\delta/\,\|(a_1 + q_1) - (a_2 + q_2)\|_2}{1 - 2T\delta/\,\|(a_1 + q_1) - (a_2 + q_2)\|_2} \\[2mm]
&< \frac{(1 + \epsilon_1)\sqrt{M/N} + 2\delta}{1 - 2\delta} \\[2mm]
&\leq (1 + \epsilon_1 + 4\delta + 4\delta\epsilon_1 + 2\delta\sqrt{N/M} + 8\delta^2\sqrt{N/M})\sqrt{M/N}.
\end{aligned}
$$

The fifth step above uses (6) and assumes $\delta \leq 1/4$. Similarly,

$$
\begin{aligned}
\frac{\|\Phi(a_1 + u_1) - \Phi(a_2 + u_2)\|_2}{\|(a_1 + u_1) - (a_2 + u_2)\|_2} &\geq \frac{\|\Phi(a_1 + q_1) - \Phi(a_2 + q_2)\|_2 - 2T\delta}{\|(a_1 + q_1) - (a_2 + q_2)\|_2 + 2T\delta} \\[2mm]
&\geq \frac{(1 - \epsilon_1)\,\|(a_1 + q_1) - (a_2 + q_2)\|_2\,\sqrt{M/N} - 2T\delta}{\|(a_1 + q_1) - (a_2 + q_2)\|_2 + 2T\delta} \\[2mm]
&= \frac{(1 - \epsilon_1)\sqrt{M/N} - 2T\delta/\,\|(a_1 + q_1) - (a_2 + q_2)\|_2}{1 + 2T\delta/\,\|(a_1 + q_1) - (a_2 + q_2)\|_2} \\[2mm]
&> \frac{(1 - \epsilon_1)\sqrt{M/N} - 2\delta}{1 + 2\delta} \\[2mm]
&\geq (1 - \epsilon_1 - 2\delta + 2\delta\epsilon_1 - 2\delta\sqrt{N/M} + 4\delta^2\sqrt{N/M})\sqrt{M/N} \\[2mm]
&> (1 - \epsilon_1 - 2\delta - 2\delta\sqrt{N/M})\sqrt{M/N}.
\end{aligned}
$$

Here the fifth step uses (7). Of the bounds we have now derived, the upper bound is the looser of the two, and so the difference vector $(a_1 + u_1) - (a_2 + u_2)$ has isometry constant

$$
\epsilon_5 := \epsilon_1 + 4\delta + 4\delta\epsilon_1 + 2\delta\sqrt{N/M} + 8\delta^2\sqrt{N/M}.
$$

Using very similar arguments one can show that the difference vectors $a_1 - (a_2 + u_2)$ and $(a_1 + u_1) - a_2$ also have isometry constant ϵ_5.

Define $b_i = a_i + u_i$, $\mu_i = d_{\mathcal{M}}(a_i, x_i)$, and $c_i = a_i + (\mu_i/T)u_i$ for $i = 1, 2$. The points c_i represent traversals of length μ_i along the tangent path rather than the geodesic path from a_i to x_i; they can also be expressed as the linear combination

$$
c_i = (1 - \mu_i/T)a_i + (\mu_i/T)b_i, \quad i = 1, 2. \tag{17}
$$

We have established above that all pairwise differences of vectors from the set $\{a_1, a_2, b_1, b_2\}$ have isometry constant ϵ_5. As we recall from Section 3.2.1, this implies squared isometry constant $3\epsilon_5$ for each of these difference vectors. We now use this fact to establish a similar bound for the

difference $c_1 - c_2$. First, we can express the distance $\|c_1 - c_2\|_2^2$ in terms of the distances between the a_i's and b_i's. Define

$$
\begin{aligned}
d_{\text{cross}} &= (\mu_1/T)(\mu_2/T)\|b_1 - b_2\|_2^2 + (1 - \mu_1/T)(\mu_2/T)\|a_1 - b_2\|_2^2 \\
&\quad + (\mu_1/T)(1 - \mu_2/T)\|b_1 - a_2\|_2^2 + (1 - \mu_1/T)(1 - \mu_2/T)\|a_1 - a_2\|_2^2
\end{aligned}
$$

and

$$
d_{\text{local}} = (\mu_1/T)(1 - \mu_1/T)\|a_1 - b_1\|_2^2 + (\mu_2/T)(1 - \mu_2/T)\|a_2 - b_2\|_2^2.
$$

Then we can use (17) to show that

$$
\|c_1 - c_2\|_2^2 = d_{\text{cross}} - d_{\text{local}}.
$$

Noting that $\|a_1 - b_1\|_2^2 = \|a_2 - b_2\|_2^2 = T^2$, we have that $d_{\text{local}} \le T^2/2$. Because $\|x_1 - x_2\|_2 > 5T$, a_1 and b_1 are at least distance T from each of a_2 and b_2, which implies that $d_{\text{cross}} > T^2 \ge 2d_{\text{local}}$. We will use this fact below. We can also express

$$
\Phi c_i = (1 - \tau_i/T)\Phi a_i + (\tau_i/T)\Phi b_i, \quad i = 1, 2,
$$

define

$$
\begin{aligned}
\widehat{d_{\text{cross}}} &= (\mu_1/T)(\mu_2/T)\|\Phi b_1 - \Phi b_2\|_2^2 + (1 - \mu_1/T)(\mu_2/T)\|\Phi a_1 - \Phi b_2\|_2^2 \\
&\quad + (\mu_1/T)(1 - \mu_2/T)\|\Phi b_1 - \Phi a_2\|_2^2 + (1 - \mu_1/T)(1 - \mu_2/T)\|\Phi a_1 - \Phi a_2\|_2^2, \text{ and}
\end{aligned}
$$

$$
\widehat{d_{\text{local}}} = (\mu_1/T)(1 - \mu_1/T)\|\Phi a_1 - \Phi b_1\|_2^2 + (\mu_2/T)(1 - \mu_2/T)\|\Phi a_2 - \Phi b_2\|_2^2,
$$

and establish that

$$
\|\Phi c_1 - \Phi c_2\|_2^2 = \widehat{d_{\text{cross}}} - \widehat{d_{\text{local}}}.
$$

Using the squared isometry constant of all pairwise differences of a_1, a_2, b_1, and b_2, we have that

$$
\begin{aligned}
\|\Phi c_1 - \Phi c_2\|_2^2 &= \widehat{d_{\text{cross}}} - \widehat{d_{\text{local}}} \\
&\le (1 + 3\epsilon_5)(M/N)d_{\text{cross}} - (1 - 3\epsilon_5)(M/N)d_{\text{local}} \\
&= \left(1 + 3\epsilon_5 + 6\epsilon_5\left(\frac{d_{\text{local}}}{d_{\text{cross}} - d_{\text{local}}}\right)\right)(M/N)(d_{\text{cross}} - d_{\text{local}}) \\
&< (1 + 9\epsilon_5)(M/N)\|c_1 - c_2\|_2^2.
\end{aligned}
$$

For the last inequality we used the fact that $d_{\text{cross}} > 2d_{\text{local}}$. Similarly, we have that

$$
\|\Phi c_1 - \Phi c_2\|_2^2 > (1 - 9\epsilon_5)(M/N)\|c_1 - c_2\|_2^2.
$$

Combining, these imply squared isometry constant $9\epsilon_5$ for the vector $c_1 - c_2$, which also implies isometry constant $9\epsilon_5$ for $c_1 - c_2$.

Finally, we are ready to compute the isometry constant for the vector $x_1 - x_2$. Using Taylor's theorem anchored at the points a_i, we have $\|x_i - c_i\|_2 \le \mu_i^2/\tau \le T^2/\tau$, $i = 1, 2$. Using the triangle inequality we also have that $\|c_1 - c_2\|_2 > T$. Thus

$$
\begin{aligned}
\frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} &\le \frac{(1 + 9\epsilon_5)\sqrt{M/N}\|c_1 - c_2\|_2 + 2T^2/\tau}{\|c_1 - c_2\|_2 - 2T^2/\tau} \\
&= \left(1 + \frac{9\epsilon_5 + 2T^2/(\tau\|c_1 - c_2\|_2) + 2T^2\sqrt{N/M}/(\tau\|c_1 - c_2\|_2)}{1 - 2T^2/(\tau\|c_1 - c_2\|_2)}\right)\sqrt{M/N} \\
&\le \left(1 + \frac{9\epsilon_5 + 2T/\tau + 2T\sqrt{N/M}/\tau}{1 - 2T/\tau}\right)\sqrt{M/N}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} & \geq \frac{(1 - 9\epsilon_5)\sqrt{M/N}\,\|c_1 - c_2\|_2 - 2T^2/\tau}{\|c_1 - c_2\|_2 + 2T^2/\tau} \\
& = \left(1 - \frac{9\epsilon_5 + 2T^2/(\tau\,\|c_1 - c_2\|_2) + 2T^2\sqrt{N/M}/(\tau\,\|c_1 - c_2\|_2)}{1 + 2T^2/(\tau\,\|c_1 - c_2\|_2)}\right)\sqrt{M/N} \\
& \geq \left(1 - (9\epsilon_5 + 2T/\tau + 2T\sqrt{N/M}/\tau)\right)\sqrt{M/N}.
\end{aligned}
$$

Considering both bounds, we have

$$
\begin{aligned}
9\epsilon_5 + 2T/\tau + 2T\sqrt{N/M}/\tau & \leq \frac{9\epsilon_5 + 2T/\tau + 2T\sqrt{N/M}/\tau}{1 - 2T/\tau} \\
& \leq (9\epsilon_5 + 2T/\tau + 2T\sqrt{N/M}/\tau)(1 + 4T/\tau).
\end{aligned}
$$

(For the second inequality, we use the assumption that $T/\tau < 1/4$.) Hence, $x_1 - x_2$ has isometry constant

$$
\epsilon_6 := 9\epsilon_5 + \frac{36\epsilon_5 T}{\tau} + \frac{2T}{\tau} + \frac{8T^2}{\tau^2} + \frac{2T\sqrt{N/M}}{\tau} + \frac{8T^2\sqrt{N/M}}{\tau^2}.
$$

$\square$

References

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

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 2007. To appear.

[3] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk. Distributed compressed sensing. 2005. Preprint.

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

[5] M. Brand. Charting a manifold. In *Advances in Neural Information Processing Systems (NIPS)*, volume 15, pages 985–992. MIT Press, 2003.

[6] D. S. Broomhead and M. Kirby. A new approach for dimensionality reduction: Theory and algorithms. *SIAM J. Applied Math.*, 60(6):2114–2142, 2000.

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

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

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

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

[11] E. Candès and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.*, 2006. To appear.

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

[13] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. In *Proc. Symp. on Geometry processing (SGP '04)*, pages 124–135. ACM Press, 2004.

[14] R. R. Coifman and M. Maggioni. Diffusion wavelets. *Appl. Comput. Harmon. Anal.*, 21(1):53–94, June 2006.

[15] J. A. Costa and A. O. Hero. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. Signal Processing*, 52(8):2210–2221, August 2004.

[16] S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. Technical Report TR-99-006, Berkeley, CA, 1999.

[17] D. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. 2005. Preprint.

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

[19] D. Donoho. For most large underdetermined systems of linear equations, the minimal L1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59(6), June 2006.

[20] D. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Disc. Comput. Geometry*, 35(4):617–652, 2006.

[21] D. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102(27):9452–9457, 2005.

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

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

[24] D. L. Donoho and C. E. Grimes. Hessian Eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA*, 100(10):5591–5596, May 2003.

[25] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when then projection radically lowers dimension. Technical Report 2006-11, Stanford University Department of Statistics, 2006.

[26] C. Grimes. *New methods in nonlinear dimensionality reduction*. PhD thesis, Department of Statistics, Stanford University, 2003.

[27] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Trans. Inform. Theory*, 52(9):4036–4048, 2006.

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

[29] M. W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*. Springer, 1976.

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

[31] G. G. Lorentz, M. von Golitschek, and Y. Makovoz. *Constructive approximation: Advanced problems*, volume 304. Springer Grundlehren, Berlin, 1996.

[32] S. Mallat. *A wavelet tour of signal processing.* Academic Press, San Diego, CA, USA, 1999.

[33] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with confidence from random samples. *Disc. Comput. Geometry*, 2006. To appear.

[34] A. Pinkus. *n*-widths and optimal recovery. In C. de Boor, editor, *Proc. Symp. Applied Math.*, volume 36, pages 51–66. AMS, 1986.

[35] I. Ur Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schroeder. Multiscale representations for manifold-valued data. *SIAM J. Multiscale Model. Simul.*, 4(4):1201–1232, 2005.

[36] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[37] M. Rudelson and R. Vershynin. Geometric approach to error correcting codes and reconstruction of signals. *Int. Math. Res. Not.*, 64:4019–4041, 2005.

[38] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, K. F. Kelly, and R. G. Baraniuk. A compressed sensing camera: New theory and an implementation using digital micromirrors. In *Proc. Comp. Imaging IV at SPIE Electronic Imaging*, San Jose, California, January 2006.

[39] D. S. Taubman and M. W. Marcellin. JPEG 2000: Image compression fundamentals, standards and practice. *Kluwer*, 2001.

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

[41] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk. Random filters for compressive sampling and reconstruction. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*. IEEE, May 2006.

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

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

[44] M. B. Wakin and R. G. Baraniuk. Random projections of signal manifolds. In *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*. IEEE, May 2006.

[45] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk. The multiscale structure of nondifferentiable image manifolds. In *Proc. Wavelets XI at SPIE Optics and Photonics*, San Diego, California, August 2005.

[46] K. Q. Weinberger and L. K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *Int. J. Comp. Vision – Special Issue: Comp. Vision Pattern Recog. (CVPR 2004)*, 70(1):77–90, 2006.

[47] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimension reduction via tangent space alignment. *SIAM J. Sci. Comput.*, 26(1):313–338, 2005.