Lower Bounds on the Low-Distortion Embedding Dimension of Submanifolds of $\mathbb{R}^n$

Mark Iwen*, Benjamin Schmidt†, Arman Tavakoli‡

May 31, 2021

Abstract

Let $\mathcal{M}$ be a smooth submanifold of $\mathbb{R}^n$ equipped with the Euclidean (chordal) metric. This note considers the smallest dimension m for which there exists a bi-Lipschitz function $f : \mathcal{M} \mapsto \mathbb{R}^m$ with bi-Lipschitz constants close to one. The main result bounds the embedding dimension m below in terms of the bi-Lipschitz constants of f and the reach, volume, diameter, and dimension of $\mathcal{M}$. This new lower bound is applied to show that prior upper bounds by Eftekhari and Wakin [5] on the minimal low-distortion embedding dimension of such manifolds using random matrices achieve near-optimal dependence on both reach and volume. This supports random linear maps as being nearly as efficient as the best possible nonlinear maps at reducing the ambient dimension for manifold data. In the process of proving our main result, we also prove similar results concerning the impossibility of achieving better nonlinear measurement maps with the Restricted Isometry Property (RIP) in compressive sensing applications.

1 Introduction

Given a set $T \subset \mathbb{R}^n$ and $\epsilon \in (0, 1)$ we consider functions $f : T \mapsto \mathbb{R}^m$ satisfying

$$(1 - \epsilon)\|\mathbf{x} - \mathbf{y}\|_2 \leq \|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leq (1 + \epsilon)\|\mathbf{x} - \mathbf{y}\|_2$$

for all $\mathbf{x}, \mathbf{y} \in T$, where $\|\cdot\|_2$ denotes the ℓ_2-norm. Such a function f is said to be an ϵ-distortion (Euclidean) embedding of T into $\mathbb{R}^m$ with embedding dimension m. The use of such embeddings has exploded in growth during the last two

*Michigan State University, Department of Mathematics, and the Department of Computational Mathematics, Science and Engineering (CMSE), markiwen@math.msu.edu. Supported in part by NSF DMS 1912706.

†Michigan State Univeristy, Department of Mathematics, schmidt@math.msu.edu. Supported in part by a Simons Collaboration Grant.

‡Michigan State University, Department of Mathematics, tavakol4@msu.edu. Supported in part by NSF DMS 1912706, and by an MSU College of Natural Science Dissertation Completion Fellowship.

decades in computer science (where T is often a finite point set), numerical linear algebra (where T is often a low-dimensional subspace), and signal processing (where T is often a manifold, or all s-sparse vectors in $\mathbb{R}^n$). This is due in large part to results of Johnson and Lindenstrauss (JL) [11] and their later simplifications (see, e.g., [4, 2]), as well as subsequent developments in compressive sensing [7] which showed that random matrices could easily provide such embeddings. Questions concerning how sub-optimal such linear embeddings might be at reducing the low-distortion embedding dimension m in comparison to the best possible nonlinear f have remained less well explored, however see, e.g., [10] for a rare result of this kind concerning finite sets T.

In this note we are principally interested in the optimality of random matrices as ϵ-distortion embeddings of submanifolds of $\mathbb{R}^n$ into $\mathbb{R}^m$. More specifically, we will consider the optimality of the following result with respect to bounding the low-distortion embedding dimension m.

Theorem 1 (Theorem 2 in [5])**.** *Let $\mathcal{M}$ be a compact d-dimensional Riemannian submanifold of $\mathbb{R}^n$ having reach τ and volume V. Conveniently assume that*

$$\frac{V}{\tau^d} \geq \left(\frac{21}{2\sqrt{d}}\right)^d. \tag{1}$$

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

$$m \geq 18\epsilon^{-2} \max\left(24d + 2d\log\left(\frac{\sqrt{d}}{\tau\epsilon^2}\right) + \log(2V^2),\ \log\left(\frac{8}{\rho}\right)\right).$$

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

$$(1-\epsilon)\|\mathbf{p}-\mathbf{q}\|_2 \leq \|\Phi\mathbf{p} - \Phi\mathbf{q}\|_2 \leq (1+\epsilon)\|\mathbf{p}-\mathbf{q}\|_2.$$

The main result of this paper is a lower bound which demonstrates the near-optimality of Theorem 1. In particular, we will see that it is generally necessary for m to exhibit a logarithmic dependence on both the volume V and the inverse of reach $1/\tau$ when (1) holds.

1.1 The Main Result and Some Examples

In this paper we prove the following lower bound for the low-distortion embedding dimension of a smooth submanifold of $\mathbb{R}^n$.

Theorem 2 (Main Result)**.** *Let $\mathcal{M}$ be a d-dimensional smooth submanifold of $\mathbb{R}^n$, possibly with boundary, that has volume V and reach τ.*[1] *Furthermore, let $\epsilon \in (0,1)$, and suppose that $f : \mathcal{M} \mapsto \mathbb{R}^m$ satisfies*

$$(1-\epsilon)\|\mathbf{p}-\mathbf{q}\|_2 \leq \|f(\mathbf{p}) - f(\mathbf{q})\|_2 \leq (1+\epsilon)\|\mathbf{p}-\mathbf{q}\|_2$$

[1] See Definition 11 for the formal definition of reach.

for all $\mathbf{p}, \mathbf{q} \in \mathcal{M}$. *If* $\frac{1}{4\sqrt{e}}(\frac{V}{\omega_d})^{\frac{1}{d}} \leq \tau$ *then*

$$m \geq C_1 \left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \frac{d}{\operatorname{diam}^2(\mathcal{M})} \left(\frac{V}{\omega_d}\right)^{\frac{2}{d}}.$$

Else, if $\tau \leq \frac{1}{4\sqrt{e}}(\frac{V}{\omega_d})^{\frac{1}{d}}$ *then*

$$m \geq C_2 \left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \frac{d\tau^2}{\operatorname{diam}^2(\mathcal{M})} \log\left(\frac{V^{\frac{1}{d}}}{(\omega_d)^{\frac{1}{d}}(4\tau)}\right).$$

Here $C_1, C_2 > 0$ *are universal constants that are independent of all other quantities,* $\omega_d := \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ *is the volume of the* d*-dimensional Euclidean unit ball, and* $\operatorname{diam}(\mathcal{M}) := \sup_{\mathbf{p},\mathbf{q}\in\mathcal{M}} \|\mathbf{p}-\mathbf{q}\|_2$.

To help illustrate the use of Theorem 2 we will now apply it to the standard examples of the d-dimensional sphere of radius r, rS^d and ℓ_2-ball rB_2^d as submanifolds of $\mathbb{R}^n$. In the process we will see that it gives near-optimal (up to constants) lower bounds on m for these simple examples when ϵ is small.

Example 3 (The Low-Distortion Embedding Dimension of Euclidean ball of radius r, rB_2^d)**.** *The pertinent geometric parameters of* $rB_2^d \subset \mathbb{R}^n$ *as a submanifold of* $\mathbb{R}^n$ *are as follows: The volume of* rB_2^d *is* $r^d\omega_d$*, its reach is* $\tau(rB_2^d) = \infty$*, and its diameter is* $\operatorname{diam}(rB_2^d) = 2r$*. Applying Theorem 2 to* $\mathcal{M} = rB_2^d$ *(i.e., a ball of radius* r *in an* d*-dimensional subspace of* $\mathbb{R}^n$*) we therefore learn that*

$$m \geq C_1 \left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \frac{d}{(2r)^2} \left(\frac{r^d\omega_d}{\omega_d}\right)^{\frac{2}{d}} = \frac{C_1}{4}\left(\frac{1-\epsilon}{1+\epsilon}\right)^2 d.$$

This result agrees with our intuition based on the fact that for a no distortion $\epsilon = 0$ *embedding of* rB^d*, a target embedding dimension of* d *is both necessary and sufficient for all choices of the radius* r*.*

Example 4 (The Low-Distortion Embedding Dimension of a Sphere rS^d)**.** *The pertinent geometric parameters of* $rS^d \subset \mathbb{R}^n$ *as a submanifold of* $\mathbb{R}^n$ *are as follows: The volume of* rS^d *is* $2r^d\frac{\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}$*, its reach is* $\tau(rS^d) = r$*, and its diameter is* $\operatorname{diam}(rS^d) = 2r$*. Check the relation between reach and the other parameters to find that*

$$\frac{1}{4\sqrt{e}}\left(\frac{V}{\omega_d}\right)^{\frac{1}{d}} = \frac{r}{4\sqrt{e}}\left(2\sqrt{\pi}\frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d+1}{2})}\right)^{\frac{1}{d}}$$

Using that $\lim_{d\to\infty}\left(2\sqrt{\pi}\frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d+1}{2})}\right)^{\frac{1}{d}} = 1$*, for large enough* d *we have* $\frac{1}{4\sqrt{e}}\left(\frac{V}{\omega_d}\right)^{\frac{1}{d}} < r = \tau$*. Applying Theorem 2 we now learn that*

$$m \geq \frac{C_1}{4}\left(\frac{1-\epsilon}{1+\epsilon}\right)^2 d\left(2\sqrt{\pi}\frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d+1}{2})}\right)^{\frac{2}{d}}$$

holds for all r and sufficiently large d. This again agrees which our intuition that to embed rS^d without distortion requires at least d dimensions.

The attentive reader may have noticed that both Examples 3 and 4 end up utilizing the first $\frac{1}{4\sqrt{e}}(\frac{V}{\omega_d})^{\frac{1}{d}} \leq \tau$ case of Theorem 2. Somewhat crucially, the following discussion will largely hinge on the second case.

1.2 A Discussion of Theorems 1 and 2

Considering the second case of Theorem 2 with $\tau \leq \frac{1}{4\sqrt{e}}(\frac{V}{\omega_d})^{\frac{1}{d}}$ we note that the ratio $\frac{V}{\tau^d}$ must be at least $\omega_d(4\sqrt{e})^d = \frac{(4\sqrt{e\pi})^d}{\Gamma(1+d/2)}$, where Γ denotes the gamma function. Recalling Stirling's approximation one can show that $\left(\frac{d}{2e}\right)^{d/2} \leq \Gamma\left(1+d/2\right)$. As a result, we see, e.g., that assuming $\tau \leq \frac{1}{4\sqrt{e}}((\frac{d}{2e\pi})^{d/2}V)^{1/d} \leq \frac{1}{4\sqrt{e}}\left(\frac{\Gamma(1+d/2)}{\pi^{d/2}}V\right)^{1/d} = \frac{1}{4\sqrt{e}}(\frac{V}{\omega_d})^{\frac{1}{d}}$, or equivalently that $\left(\frac{32e^2\pi}{d}\right)^{d/2} = \frac{(4\sqrt{e})^d}{(\frac{d}{2e\pi})^{d/2}} \leq \frac{V}{\tau^d}$, implies the inequality $\tau \leq \frac{1}{4\sqrt{e}}(\frac{V}{\omega_d})^{\frac{1}{d}}$ in Theorem 2. We further note that this range of τ is also compatible with assumption (1) in Theorem 1 that $\frac{V}{\tau^d} \geq (\frac{21}{2\sqrt{d}})^d$.

Assuming then for the sake of comparison that $\frac{V}{\tau^d} \geq \max\left\{(\frac{21}{2\sqrt{d}})^d, \omega_d(4\sqrt{e})^d\right\}$ and setting $\epsilon = \rho = 1/3$ (for example) in Theorem 1 one can see that the minimum achievable, e.g., $\epsilon = 1/3$ embedding dimension m of $\mathcal{M}$ must satisfy

$$\frac{C_2' d\tau^2}{\mathrm{diam}^2(M)} \log\left(\frac{V^{\frac{1}{d}}}{(\omega_d)^{\frac{1}{d}}(4\tau)}\right) \underset{Thm\ 2}{\leq} m \underset{Thm\ 1}{\leq} \tilde{C} d\left(1+\log\left(\frac{d}{\tau^2}V^{\frac{2}{d}}\right)\right)$$

for all $d \geq 1$. Simplifying the expression further under the assumption that, e.g., $\frac{V}{\tau^d} = (\frac{C'}{\sqrt{d}})^d$ we have that

$$\frac{C_2'\tau^2}{\mathrm{diam}^2(M)} \log\left(\frac{C'}{4\sqrt{d}\sqrt[d]{\omega_d}}\right) \leq \frac{m}{d} \leq \tilde{C}\left(1+\log\left(2C'^2\right)\right)$$

In this case we can see that under the assumptions above that the upper and lower bounds on m differ by a factor proportional to $\frac{\tau^2}{\mathrm{diam}^2(M)}$ for d not too large (e.g., $d < 100$).

Taken all together we can see that Theorem 2 supports Theorem 1 as being optimal up to (at worst) a factor proportional to $\frac{\tau^2}{\mathrm{diam}^2(M)}$ for d not too large, and $\frac{V}{\tau^d} = (\frac{C'}{\sqrt{d}})^d$ for appropriate ranges of $C' \in \mathbb{R}^+$. This is perhaps most interesting due to the fact that Theorem 2 implies that it is not possible to improve Theorem 1 beyond this factor in this regime, even by utilizing *nonlinear* functions f in the place of the matrix Φ. Indeed, a similar fact is established as a consequence of the proof of Theorem 2 for matrices with the Restricted Isometry Property (RIP) (see Example 9 below), and an existing result of this kind is also reproduced for JL embeddings of finite sets (see Example 8).

1.3 An Outline of the Proof of Theorem 2

The remainder of the paper is organized as follows. In Section 2 we prove Theorem 7, a general method for bounding the low-distortion embedding dimension of an arbitrary subset $T \subset \mathbb{R}^n$ in terms of its covering numbers. Given this result the proof of Theorem 2 is then reduced to lower bounding the covering numbers of submanifolds of $\mathbb{R}^n$ as subsets of $\mathbb{R}^n$. This is done in Section 3 and results in Theorem 13. Finally, Theorem 2 is then proven in Section 4.

2 General Lower Bounds for Embedding Dimension via High Dimensional Probability

In this section we present a relatively simple lower bound for the achievable low-distortion Euclidean embedding dimension, m, of an arbitrary bounded subset $T \subset \mathbb{R}^n$ into $\mathbb{R}^m$ by any function. As we shall see, when coupled with lower bounds for the covering numbers of T, Theorem 7 below allows one to quickly reproduce more specialized results concerning the necessary low-distortion embedding dimension of finite point sets [3, 9, 10] in certain parameter regimes. Similarly, the result can also be used to quickly lower bound the number of rows that any matrix with the RIP must have, reproducing existing lower bounds of interest in compressive sensing (see, e.g., [7, Corollary 10.8]). In this second example one can further learn that existing RIP matrices are optimal in the sense that they match the necessary embedding dimension of even nonlinear embeddings of all s-sparse vectors in $\mathbb{R}^n$ (up to constants in certain parameter regimes).

To prove this general lower bound result we need the notion of the *Gaussian Width* of a set $T \subset \mathbb{R}^n$.

Definition 5. *[13, Definition 7.5.1] The Gaussian Width of a set $T \subset \mathbb{R}^n$ is*

$$w(T) := \mathbb{E} \sup_{\mathbf{x} \in T} \langle \mathbf{g}, \mathbf{x} \rangle$$

where $\mathbf{g}$ is a random vector with n independent and identically distributed (i.i.d.) mean 0 and variance 1 Gaussian entries.

The following lemma bounds the Gaussian width of the image of a low-distortion linear map of $T \subset \mathbb{R}^n$ in terms of the Gaussian width of T. It is largely a consequence of the Sudakov-Fernique comparison theorem for mean 0 Gaussian processes (see, e.g., [13, Theorem 7.2.11] for the statement used here).

Lemma 6. *Let $a, b \in \mathbb{R}^+$, $T \subset \mathbb{R}^n$, and $f : T \to \mathbb{R}^m$ be a function such that $a\|\mathbf{x} - \mathbf{y}\|_2 \leq \|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leq b\|\mathbf{x} - \mathbf{y}\|_2$ holds for all $\mathbf{x}, \mathbf{y} \in T$. Then, $a\, w(T) \leq w\left(f(T)\right) \leq b\, w(T)$ also holds.*

Proof. Define the mean 0 Gaussian processes via $X_{\mathbf{x}} := a\langle \mathbf{g}, \mathbf{x} \rangle$, $Y_{\mathbf{x}} := \langle \mathbf{g}', f(\mathbf{x}) \rangle$, and $Z_{\mathbf{x}} = b\,\langle \mathbf{g}, \mathbf{x} \rangle$ for all $\mathbf{x} \in T$, where $\mathbf{g}, \mathbf{g}'$ are Gaussian random vectors with

n and m i.i.d. mean 0 and variance 1 Gaussian entries, respectively. For each $\mathbf{x}, \mathbf{y} \in T$, we then have that their increments satisfy

$$\begin{aligned}
&\mathbb{E}\left(X_{\mathbf{x}} - X_{\mathbf{y}}\right)^2 = a^2 \mathbb{E}\langle \mathbf{g}, \mathbf{x} - \mathbf{y}\rangle^2 = a^2 \|\mathbf{x} - \mathbf{y}\|_2^2 \leq \\
&\mathbb{E}(Y_{\mathbf{x}} - Y_{\mathbf{y}})^2 = \mathbb{E}\langle \mathbf{g}', f(\mathbf{x}) - f(\mathbf{y})\rangle^2 = \|f(\mathbf{x}) - f(\mathbf{y})\|_2^2 \leq \\
&\mathbb{E}(Z_{\mathbf{x}} - Z_{\mathbf{y}})^2 = b^2 \mathbb{E}\langle \mathbf{g}, \mathbf{x} - \mathbf{y}\rangle^2 = b^2 \|\mathbf{x} - \mathbf{y}\|_2^2.
\end{aligned}$$

The stated result now follows from the Sudakov-Fernique comparison theorem together with the fact that, e.g., $\mathbb{E}\sup_{\mathbf{x}\in T} X_{\mathbf{x}} = a\,\mathbb{E}\sup_{\mathbf{x}\in T}\langle \mathbf{g}, \mathbf{x}\rangle = a\,w(T)$. □

With Lemma 6 in hand we are now able to prove the main theorem of this section. The proof also utilizes basic properties of Gaussian widths [13, Proposition 7.5.2], the Gaussian width of the open Euclidean unit ball in m dimensions, $B_2^m \subset \mathbb{R}^m$, being $w(B_2^m) \leq C'\sqrt{m}$ for an absolute constant $C' \in \mathbb{R}^+$ [13, Example 7.5.7], and Sudakov's minoration inequality [13, Theorem 7.4.1].

Theorem 7. *Let $a, b, \delta \in \mathbb{R}^+$, $T \subset \mathbb{R}^n$, and $f : T \to \mathbb{R}^m$ be a function such that $a\|\mathbf{x} - \mathbf{y}\|_2 \leq \|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leq b\|\mathbf{x} - \mathbf{y}\|_2$ holds for all $\mathbf{x}, \mathbf{y} \in T$. Then,*

$$\frac{Ca^2\delta^2}{b^2}\,\frac{\log N(T,\delta)}{\operatorname{diam}^2(T)} \leq \left(\frac{2a}{C'b}\,\frac{w(T)}{\operatorname{diam}(T)}\right)^2 \leq m$$

holds, where $\operatorname{diam}(T) := \sup_{\mathbf{x},\mathbf{y}\in T}\|\mathbf{x}-\mathbf{y}\|_2$, and where $N(T,\delta)$ is the δ-covering number of T by Euclidean balls of radius δ centered on T. Here $C, C' > 0$ are universal constants that are independent of all other quantities.

Proof. Consider the set $f(T) - f(T) := \{f(\mathbf{x}) - f(\mathbf{y}) \mid \mathbf{x}, \mathbf{y} \in T\}$. By Lemma 6 and the properties of Gaussian widths, $a\,w(T) \leq w\left(f(T)\right) = \frac{1}{2}w\left(f(T) - f(T)\right)$. Furthermore, $f(T) - f(T)$ is contained in a ball in $\mathbb{R}^m$ with radius $b\operatorname{diam}(T)$. Since $w(B_2^m) \leq C'\sqrt{m}$, monotonicity of the Gaussian width and scaling properties now imply that $w(f(T) - f(T)) \leq C'\,b\operatorname{diam}(T)\sqrt{m}$. Therefore,

$$a\,w(T) \leq \frac{1}{2}w\left(f(T) - f(T)\right) \leq \frac{C'\,b}{2}\operatorname{diam}(T)\,\sqrt{m}.$$

The rightmost inequality follows. The leftmost inequality is now a consequence of Sudakov's minoration inequality. □

We will now use Theorem 7 to reproduce the promised lower bounds concerning RIP matrices and the low-distortion embedding dimension of finite point sets.

Example 8 (The Necessary Low-Distortion Embedding Dimension of Finite Sets)**.** *Let $\epsilon \in (0,1)$, and $T = \{\mathbf{x}_1, \ldots \mathbf{x}_N\} \subset \mathbb{R}^n$ be a finite set of points whose minimal distance between any two elements is $\delta := \min_{j\neq k}\|\mathbf{x}_j - \mathbf{x}_k\|_2$. Note that the $\delta/2$-covering number of T is $N(T, \delta/2) = N$. As a result, Theorem 7 implies that any function $f : \mathbb{R}^n \to \mathbb{R}^m$ satisfying $\sqrt{1-\epsilon}\|\mathbf{x}_j - \mathbf{x}_k\|_2 \leq \|f(\mathbf{x}_j) - f(\mathbf{x}_k)\|_2 \leq \sqrt{1+\epsilon}\|\mathbf{x}_j - \mathbf{x}_k\|_2$ for all $\mathbf{x}_j, \mathbf{x}_k \in T$ must have $m \geq$*

$C''\frac{(1-\epsilon)\delta^2}{(1+\epsilon)}\frac{\log N}{\operatorname{diam}^2(T)}$ *for an absolute constant* $C'' \in \mathbb{R}^+$.

Now consider ϵ *not too small (e.g.,* $\epsilon > 0.01$*). Furthermore, let* T *to be a subset of, e.g., a* 0.01*-packing of the closed unit ball* $\overline{B_2^n} \subset \mathbb{R}^n$ *so that* $\delta > 0.01$ *and* $1.98 \leq \operatorname{diam}(T) \leq 2$ *both hold. For this range of* ϵ *and such* T *coming from a packing of the unit ball one can see that the lower bound on* m *from Theorem 7 matches (up to constants) the most recent worst case bounds of the form* $m = \Omega\left(\log(N)/\epsilon^2\right)$ *proven by Larson and Nelson in [10] for finite point sets. Furthermore, this example application of Theorem 7 helps to cement our intuition that (subsets of) packings of the unit ball are among the most difficult finite sets to embed into a lower dimensional Euclidean space.*

Example 9 (The Number of Rows Required by an RIP matrix, and the Non-Existence of Better Nonlinear Low-Distortion Embeddings for Sparse Vectors)**.** *Let* $S \subseteq [n] := \{1, \ldots, n\} \subset \mathbb{N}$*, and denote the* j^{th}*-column of the* $n \times n$ *identity matrix by* $\mathbf{e}_j$ *for all* $j \in [n]$*. Next, define* $C_S := \operatorname{span}\{\mathbf{e}_j \mid j \in S\}$ *and let* $U_s := \left(\bigcup_{S\subset[n],\ |S|\leq s} C_S\right) \bigcap \overline{B_2^n} \subset \mathbb{R}^n$ *be the set of all* s*-sparse vectors with* ℓ_2*-norm at most one. A matrix* $A \in \mathbb{R}^{m'\times n}$ *will have the RIP of order* (s, ϵ) *for* $s \in [n]$ *even and* $\epsilon \in (0, 1)$ *if and only if* $\sqrt{1-\epsilon}\|\mathbf{x}-\mathbf{y}\|_2 \leq \|A\mathbf{x} - A\mathbf{y}\|_2 \leq \sqrt{1+\epsilon}\|\mathbf{x}-\mathbf{y}\|_2$ *holds for all* $\mathbf{x}, \mathbf{y} \in U_{s/2}$*. As in the previous example we will consider* ϵ *not too small (e.g.,* $\epsilon > 0.01$*). In this regime it is known that a matrix* $A \in \mathbb{R}^{m'\times n}$ *with both the RIP of order* (s, ϵ) *and as few rows* m' *as possible will/must have* $m' = \Theta\left(s \log(N/s)\right)$*. See, e.g., [7, Theorem 9.2] and [7, Corollary 10.8] for specific supporting theorems.*

One natural question concerns whether allowing the matrix $A \in \mathbb{R}^{m'\times n}$ *to be replaced by an arbitrary nonlinear function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *might permit* $m \ll m' = \Theta\left(s \log(N/s)\right)$ *to hold while still approximately preserving the* ℓ_2*-norms of all* s*-sparse vectors up to the same* $\sqrt{1 \pm \epsilon}$*-distortion factors. Toward answering this question one can apply Theorem 7 to see that any function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *satisfying* $\sqrt{1-\epsilon}\|\mathbf{x}-\mathbf{y}\|_2 \leq \|f(\mathbf{x})-f(\mathbf{y})\|_2 \leq \sqrt{1+\epsilon}\|\mathbf{x}-\mathbf{y}\|_2$ *for all* $\mathbf{x}, \mathbf{y} \in U_{s/2}$ *must have* $m \geq C\frac{(1-\epsilon)\delta^2}{(1+\epsilon)}\frac{\log N(U_{s/2},\delta)}{\operatorname{diam}^2(U_{s/2})} = C\frac{(1-\epsilon)\delta^2}{4(1+\epsilon)}\log N(U_{s/2}, \delta)$ *for our choice of* $\delta \in \mathbb{R}^+$*. In order to make a good choice for* δ *in this bound we will use the following lemma.*

Lemma 10 (Lemma 10.12 in [7])**.** *Given integers* $s < n$*, there exist* $N \geq \left(\frac{n}{2s}\right)^{s/4}$ *subsets* $S_1, \ldots, S_N$ *of* $[n]$ *such that* $|S_j| = s/2$ *and* $|S_j \cap S_k| < \frac{s}{4}$ *whenever* $j \neq k$*.*

We can now use Lemma 10 to define $\mathbf{x}_1, \ldots, \mathbf{x}_N \in U_{s/2}$ *by letting* $\mathbf{x}_j$ *be* $\sqrt{2/s}$ *for each index in* S_j*, and* 0 *elsewhere. Note that* $\|\mathbf{x}_j - \mathbf{x}_k\|_2 > 1$ *will hold whenever* $j \neq k$*. As a consequence of the existence of these* N *points we can see that, e.g., setting* $\delta < 1/2$ *forces* $N(U_{s/2}, \delta) \geq N \geq \left(\frac{n}{2s}\right)^{s/4}$*. Hence, choosing any* $\epsilon > 0.01$ *and, e.g.,* $\delta = 1/3$ *in the bound on* m *from Theorem 7 above will yield the bound*

$m \geq C'''s\log(n/2s)$ for a universal constant $C''' \in \mathbb{R}^+$. As a consequence, we can see that linear measurements achieve the RIP with a near optimal level of compression even when compared to arbitrary nonlinear measurement functions, at least for larger values of ϵ. Allowing f to be nonlinear doesn't help here.

Having digressed with the two previous examples, we hope to have established a clear proof strategy for the manifold case. If we can lower bound the covering numbers of smooth submanifolds of $\mathbb{R}^n$, then we can use those lower bounds together with Theorem 7 to establish lower bounds of the achievable low-distortion embedding dimension m of any given submanifold $\mathcal{M}$ of $\mathbb{R}^n$ into $\mathbb{R}^m$. In persuance of that strategy we will develop lower bounds for the covering numbers of submanifolds of $\mathbb{R}^n$ in the next section.

3 Lower Bounds for the Covering Numbers of Smooth Submanifolds of Euclidean Space

Our lower covering number bounds for a smooth submanifold $\mathcal{M} \subset \mathbb{R}^n$ will in terms of its dimension d, d-dimensional volume V, and reach τ.

Definition 11 (**Reach:** Definition 4.1 in [6])**.** *For a subset of Euclidean space $T \subset \mathbb{R}^n$, the reach τ is defined as*

$$\tau(T) = \sup\{t \geq 0 \,\big|\, \forall \mathbf{x} \in \mathbb{R}^n \text{ with } d(\mathbf{x}, T) < t,\ \mathbf{x} \text{ has a unique closest point in } T\}.$$

The following lemma concerns the control of the sectional curvatures of $\mathcal{M}$ by τ, and also demonstrates how τ can be used to help relate geodesic distances between points on $\mathcal{M}$ to their Euclidean distances.

Lemma 12. *Let $\mathcal{M}$ be a compact, smooth submanifold of $\mathbb{R}^n$ possibly with boundary. Furthermore, let τ be the reach of $\mathcal{M}$, $\mathbf{p}, \mathbf{q} \in \mathcal{M}$, $\mathbf{x} \in \mathbb{R}^n$, and d and l be the Euclidean and geodesic distances between $\mathbf{p}$ and $\mathbf{q}$, respectively. Then:*

1. *In the interior of $\mathcal{M}$, any sectional curvature k satisfies $\frac{-2}{\tau^2} \leq k \leq \frac{1}{\tau^2}$.*
2. *$l - \frac{l^2}{2\tau} \leq d$.*
3. *When restricted to $d \leq \frac{\tau}{2}$, we further have $l \leq d + \frac{2d^2}{\tau}$.*

Proof. 1. See [1, proposition A.1]. 2. See [12, lemma 6.3]. 3. See [12, lemma 6.3] and [5, lemma 7] which gives $l \leq \tau - \tau\sqrt{1 - \frac{2d}{\tau}}$. Using $1 - \sqrt{1-x} \leq \frac{x+x^2}{2}$ we get the claimed $l \leq d + \frac{2d^2}{\tau}$. □

Lemma 12 in combination with Bishop's volume comparison theorem [8, Theorem 3.101, part i] now allows us to prove the main result of this section.

Theorem 13. *Let $\mathcal{M}$ be a d-dimensional smooth submanifold of $\mathbb{R}^n$ possibly with boundary, and with volume V and reach τ. Let $0 < \delta \leq \frac{\tau}{2}$, and $r := \delta(1 + \frac{2\delta}{\tau})$. Let $\omega_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ be the volume of the d-dimensional Euclidean unit ball B_2^d. Then the covering number of $\mathcal{M}$ with Euclidean balls centered on M of radius δ, $N(\mathcal{M}, \delta)$, satisfies*

$$\frac{V}{\omega_d(8\delta)^d} \leq \frac{V}{\omega_d(1 + \frac{2\sqrt{2}r}{\tau})^{d-1}r^d} \leq N(\mathcal{M}, \delta) \ .$$

Proof. We claim that the volume of the intersection of $\mathcal{M}$ with a Euclidean ball having center in $\mathcal{M}$ and radius δ is no greater than $\omega_d(1 + \frac{2\sqrt{2}r}{\tau})^{d-1}r^d$. This claim immediately implies the rightmost inequality.

Toward establishing this claim, we note that for two points $\mathbf{p}, \mathbf{q} \in \mathcal{M}$ with Euclidean distance d, geodesic distance l, and having $d \leq \frac{\tau}{2}$, one has $l \leq d(1 + \frac{2d}{\tau})$ by Lemma 12. Hence, the intersection of $\mathcal{M}$ with a Euclidean ball of radius δ centered on $\mathcal{M}$ is contained in a geodesic ball of radius $r = \delta(1 + \frac{2\delta}{\tau})$ with the same center provided that $\delta \leq \frac{\tau}{2}$. Thus, it suffices to bound the volume of such a geodesic ball from above by $w_d(1 + \frac{2\sqrt{2}r}{\tau})^{d-1}r^d$ in order to obtain the rightmost inequality.

By Lemma 12 the sectional curves of $\mathcal{M}$ are bounded below by $\frac{-2}{\tau^2}$. Hence by Bishop's theorem the volume of a geodesic ball of radius r in $\mathcal{M}$ is bounded above by the volume of the geodesic ball of radius r in the hyperbolic space with constant sectional curvature $\frac{-2}{\tau^2}$. Denote this volume by $V_{\frac{-2}{\tau^2}}(r)$; it is given by the formula

$$V_{\frac{-2}{\tau^2}}(r) = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \int_0^r S_{\frac{-2}{\tau^2}}(x)^{d-1} \, dx$$

where $S_k(x) = \frac{1}{\sqrt{-k}} \sinh(\sqrt{-k}x)$. Since $\frac{\sinh(x)}{x}$ is increasing on $0 < x$,

$$\begin{aligned} V_{\frac{-2}{\tau^2}}(r) = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \int_0^r S_{\frac{-2}{\tau^2}}(x)^{d-1} \, dx &\leq \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \int_0^r \left(\frac{\sinh(\frac{\sqrt{2}x}{\tau})}{\frac{\sqrt{2}x}{\tau}} \right)^{d-1} x^{d-1} \, dx \\ &\leq \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \left(\frac{\sinh(\frac{\sqrt{2}r}{\tau})}{\frac{\sqrt{2}r}{\tau}} \right)^{d-1} \frac{r^d}{d}. \end{aligned}$$

Using that $\frac{\sinh(x)}{x} \leq 1 + 2x$ for $0 < x < \sqrt{2}\pi$, we finally obtain that

$$V_{\frac{-2}{\tau^2}}(r) \leq \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \left(1 + \frac{2\sqrt{2}r}{\tau} \right)^{d-1} \frac{r^d}{d} = \omega_d \left(1 + \frac{2\sqrt{2}r}{\tau} \right)^{d-1} r^d.$$

This proves the rightmost inequality. To obtain the leftmost inequality we simplify $r = \delta(1 + \frac{2\delta}{\tau})$ via $r \leq 2\delta \leq \tau$ using that $0 < \delta \leq \frac{\tau}{2}$. We then get the factor of 8 from the fact that $(1 + 2\sqrt{2})2 < 8$. □

In light of Theorem 13, the following technical lemma will be useful in choosing the parameter δ in Theorem 7.

Lemma 14. *Let $d, C' \in \mathbb{R}^+$. If $\frac{2C'^{\frac{1}{d}}}{\sqrt{e}} \leq \tau$ then*

$$\frac{d}{2e} C'^{\frac{2}{d}} \leq \sup_{\delta \in (0, \tau/2]} \delta^2 \log\left(\frac{C'}{\delta^d}\right).$$

If $\tau \leq \frac{2C'^{\frac{1}{d}}}{\sqrt{e}}$ then

$$\frac{d\tau^2}{4} \log\left(\frac{2C'^{\frac{1}{d}}}{\tau}\right) \leq \sup_{\delta \in (0, \tau/2]} \delta^2 \log\left(\frac{C'}{\delta^d}\right).$$

Proof. One can see that $\delta^2 \log\left(\frac{C'}{\delta^d}\right)$ is maximized at $\delta^* := C'^{\frac{1}{d}}/\sqrt{e} > 0$. If $\delta^* \notin (0, \tau/2]$ we instead evaluate the function at $\tau/2$. □

We are now prepared to prove the main theorem of this paper.

4 The Proof of Theorem 2

Applying Theorem 7 to $\mathcal{M} \subset \mathbb{R}^n$ we learn that

$$m \geq C \left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \frac{\delta^2 \log N(\mathcal{M}, \delta)}{\operatorname{diam}^2(\mathcal{M})},$$

where we may choose $\delta \in \mathbb{R}^+$ as we wish. Restricting ourselves to $\delta \in (0, \tau/2]$ and applying Theorem 13 we further learn that

$$m \geq C \left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \frac{\delta^2 \log\left(\frac{V}{\omega_d (8\delta)^d}\right)}{\operatorname{diam}^2(\mathcal{M})}.$$

To maximize this lower bound as a function of δ we now apply Lemma 14 with $C' = V/\omega_d 8^d$. The stated result now follows.

References

[1] E. Aamari, J. Kim, F. Chazal, B. Michel, A. Rinaldo, and L. Wasserman. Estimating the reach of a manifold. *Electronic Journal of Statistics*, 13(1):1359–1399, 2019. 8

[2] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003. 2

[3] N. Alon. Problems and results in extremal combinatorics—i. *Discrete Mathematics*, 273(1-3):31–53, 2003. 5

[4] S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. *International Computer Science Institute, Technical Report*, 22(1):1–5, 1999. 2

[5] A. Eftekhari and M. B. Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, 39(1):67–109, July 2015. 1, 2, 8

[6] H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–418, Mar. 1959. 8

[7] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Springer New York, New York, NY, 2013. 2, 5, 7

[8] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian geometry*, volume 2. Springer, 1990. 8

[9] K. G. Larsen and J. Nelson. The Johnson-Lindenstrauss Lemma Is Optimal for Linear Dimensionality Reduction. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, pages 82:1–82:11. 5

[10] K. G. Larsen and J. Nelson. Optimality of the Johnson-Lindenstrauss lemma. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 633–638. IEEE, 2017. 2, 5, 7

[11] W. J. J. Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemp. Math*, 26:189–206, 1984. 2

[12] P. Niyogi, S. Smale, and S. Weinberger. Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete & Computational Geometry*, 39(1-3):419–441, Mar. 2008. 8

[13] R. Vershynin. *High-dimensional probability: an introduction with applications in data science*. Number 47 in Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge ; New York, NY, 2018. 5, 6