On Fast Johnson-Lindenstrauss Embeddings of Compact Submanifolds of $\mathbb{R}^N$ with Boundary

Mark A. Iwen*, Benjamin Schmidt†, Arman Tavakoli‡

May 24, 2022

Abstract

Let $\mathcal{M}$ be a smooth d-dimensional submanifold of $\mathbb{R}^N$ with boundary that's equipped with the Euclidean (chordal) metric, and choose $m \leq N$. In this paper we consider the probability that a random matrix $A \in \mathbb{R}^{m\times N}$ will serve as a bi-Lipschitz function $A : \mathcal{M} \to \mathbb{R}^m$ with bi-Lipschitz constants close to one for three different types of distributions on the $m \times N$ matrices A, including two whose realizations are guaranteed to have fast matrix-vector multiplies. In doing so we generalize prior randomized metric space embedding results of this type for submanifolds of $\mathbb{R}^N$ by allowing for the presence of boundary while also retaining, and in some cases improving, prior lower bounds on the achievable embedding dimensions m for which one can expect small distortion with high probability. In particular, motivated by recent modewise embedding constructions for tensor data, herein we present a new class of highly structured distributions on matrices which outperform prior structured matrix distributions for embedding sufficiently low-dimensional submanifolds of $\mathbb{R}^N$ (with $d \lesssim \sqrt{N}$) with respect to both achievable embedding dimension, and computationally efficient realizations. As a consequence we are able to present, for example, a general new class of Johnson-Lindenstrauss embedding matrices for $\mathcal{O}(\log^c N)$-dimensional submanifolds of $\mathbb{R}^N$ which enjoy $\mathcal{O}(N \log(\log N))$-time matrix vector multiplications.

Keywords Randomized manifold embeddings, Johnson-Lindenstrauss lemma, Manifolds with boundary, Fast dimension reduction

Mathematics Subject Classification 53C40, 53Z99, 68P30 , 65D99

1 Introduction

Given a subset S of $\mathbb{R}^N$, $m \leq N$, and $\epsilon \in (0, 1)$, we will consider random matrices $A \in \mathbb{R}^{m\times N}$ satisfying

*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 and by NSF DMS 2106472.

†Michigan State University, 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.

$$(1-\epsilon)\|\mathbf{x}-\mathbf{y}\|_2^2 \leq \|A\mathbf{x}-A\mathbf{y}\|_2^2 \leq (1+\epsilon)\|\mathbf{x}-\mathbf{y}\|_2^2 \qquad (\dagger)$$

for all $\mathbf{x}, \mathbf{y} \in S$ simultaneously with high probability, where $\|\cdot\|_2$ denotes the ℓ_2-norm. Herein we will refer to any successful realization $A \in \mathbb{R}^{m\times N}$ satisfying (†) as an **ϵ-JL embedding of S into $\mathbb{R}^m$** in keeping with the extensive literature (see, e.g., [16, 2, 3, 35, 24, 7]) related to the many applications, extensions, and modifications of the celebrated Johnson-Lindenstrauss (JL) Lemma [39]. More specifically, this paper is principally concerned with the case where S is a low-dimensional compact submanifold of $\mathbb{R}^N$. In such cases the primary goal then becomes to bound the minimum embedding dimension m achievable by any ϵ-JL embedding of the submanifold in terms of its geometric characteristics, including, e.g, its dimension, volume, and reach [22]. Of course, the sufficient minimum achievable embedding dimension of a given submanifold generally depends on the distributions of the random matrices A considered above. As a result, there is a large body of work bounding the minimal embedding dimension of submanifolds achievable by various classes of random matrices [27, 9, 15, 49, 20, 18, 36] including, e.g., matrices with independent sub-gaussian rows [20, 18] as well as more structured random matrices which support faster matrix-vector multiplies [49]. In this paper we prove three new embedding theorems of this type which apply to submanifolds of $\mathbb{R}^N$ *both with and without boundary*, including results which provide both improved embedding dimension and runtime bounds for ϵ-JL embeddings of sufficiently low-dimensional manifold data.

The Importance of Boundaries: The applications of random low-distortion embeddings of type (†) are wide-ranging due to their ability to provide dimensionality reduction of incoming data prior to the user having any detailed knowledge of the data's characteristics beyond some rough measures of its likely complexity (e.g., in terms of an upper bound on its Gaussian width [46, Section 7.5], etc.). This has lead to ϵ-JL embeddings being proposed as a means to reduce measurement costs for many applications involving data conforming to a manifold model. Such applications include compressive sensing with manifold models [14, 33, 31, 32, 19], manifold learning and parameter estimation from compressive measurements [27, 9, 20, 21], and target recognition and classification via manifold models [17]. In addition, low-distortion manifold embeddings have recently been used to, e.g., help explain successful medical imaging from subsampled data via deep learning techniques [28]. In most of these applications the manifold models one considers often have boundary, and often for natural reasons. Consider, e.g., the standard "Swiss-roll" manifold one commonly encounters in the manifold learning literature (see, e.g., [44]) which has a boundary. More pertinently, however, one might also consider applications such as the aforementioned work on target recognition and classification [17] where one encounters image manifolds whose parameters include, e.g., the direction of view between an overflying aircraft collecting data and the object one wishes to classify. In such settings the physical limitations of the data collection (e.g., the pilot's understandable desire for an above-ground flight path which limits viewing directions to at most half of $\mathbb{S}^2$) will generally necessitate the presence of a boundary in the collectable manifold data. For such reasons we believe a careful analysis of boundary effects on ϵ-JL embeddings of submanifolds of $\mathbb{R}^N$ to be of fundamental importance in the context of all of the applications mentioned above.

Mathematically, the presence of a boundary in a given manifold $\mathcal{M}$ makes formulating covering number bounds for $\mathcal{M}$ more difficult by complicating the estimation of the volume of the portion of the manifold contained within a given Euclidean ball whose center lies too close to its boundary. As a consequence, the types of uniform volume estimates present in prior ϵ-JL embedding proofs for manifolds without boundary do not apply near $\partial\mathcal{M}$. A further complication is the assumption in prior work for manifolds without boundary that geodesics have a well defined external acceleration.

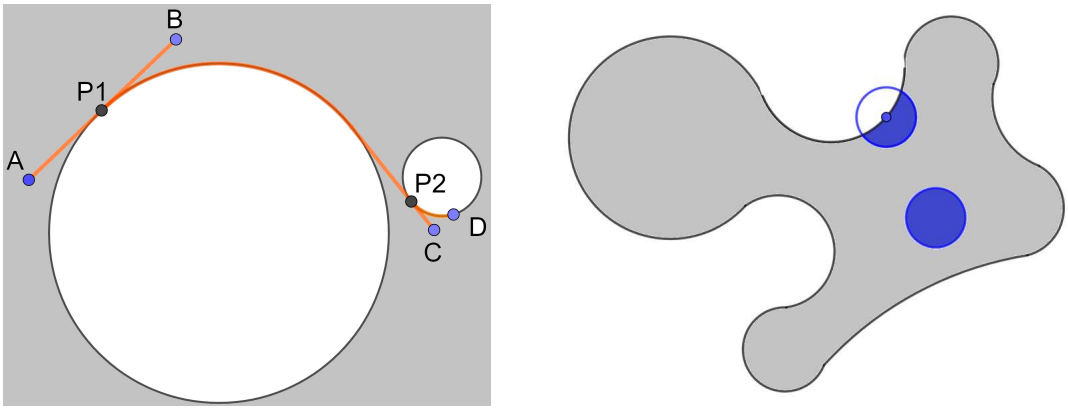

Figure 1: The presence of boundary can cause geodesics to bifurcate. In the left figure, the geodesic connecting A to B does not have a unique continuation as it can bifurcate at P_1 to reach C. This process can repeat, as it further bifurcates at P_2 to reach D. Such unit speed geodesics are $\mathcal{C}^1$ but not $\mathcal{C}^2$. In the right figure, we have two Euclidean balls overlapping with a manifold with boundary. If the center of the ball is close to the boundary, the ball will cover less of the manifold. This situation is further amplified in higher dimensions as the volume of the collar of the boundary typically grows exponentially [46, Remark 5.1.10]. In theorem 4.3 we address this issue by treating the collar of the boundary and the interior regions separately.

This is not the case in manifolds with boundary as a geodesic may not be $\mathcal{C}^2$, and may not have a unique continuation even if the underlying manifold is smooth (see Figure 1). In this paper we address these difficulties in order to extend prior results to the case of manifolds with boundary by carefully treating boundary and interior regions separately. The end result of this work is a general bound on the Gaussian width of the unit secants of a given submanifold of $\mathbb{R}^N$, potentially with boundary, in terms of its dimension, volume, and reach properties. With these bounds in hand we are then able to apply embedding results for general infinite sets with bounded Gaussian width to prove several new manifold embedding results. *To the best of our knowledge the resulting ϵ-JL embedding theorems proven herein are the first to apply to manifolds with boundary, and as such greatly generalize the class of manifold models for which such embedding techniques can be theoretically proven to work.*

Improved ϵ-JL Embedding Dimensions and Runtimes for Low-Dimensional Manifolds: In addition to allowing for the presence of boundary, we also provide improved ϵ-JL embedding results for submanifolds of $\mathbb{R}^N$ via highly structured random matrices which admit fast matrix-vector multiplies. Perhaps the most widely considered structured random matrices of this type are Subsampled Orthonormal with Random Signs (SORS) matrices of the form $A = \sqrt{\frac{N}{m}} RUD \in \mathbb{R}^{m \times N}$, where $R \in \{0,1\}^{m \times N}$ contains m rows independently sampled uniformly at random from the $N \times N$ identity matrix, $U \in \mathbb{R}^{N \times N}$ is a unitary matrix, and $D \in \{-1,0,1\}^{N \times N}$ is a diagonal matrix with independently and identically distributed (i.i.d.) Rademacher random variables on its diagonal. Note that such SORS matrices A will have fast matrix vector multiplies if, e.g., the orthonormal basis U is chosen to be related to a Discrete Fourier Transform (DFT) matrix with an $\mathcal{O}(N \log N)$

time matrix-vector multiply.[1] Herein we generalize existing results concerning SORS embeddings of submanifolds [49] to accommodate for the presence of boundary, while simultaneously removing a few logarithmic factors from prior lower bounds by appealing to recent concentration inequalities.

More interestingly, though, we also propose a new class of structured random matrices for embedding manifold data motivated by recent developments in the construction of fast modewise JL-embeddings for tensor data (see, e.g., [34, 7]). This new class of structured linear JL maps has several advantages over more commonly considered random embedding matrices including (i) lower-storage costs, (ii) trivially parallelizable data evaluations, (iii) the use of fewer random bits, and (iv) faster serial matrix-vector multiplies for structured data. The many useful computational characteristics of these embeddings for tensor data motivate the following naive question: Is it possible to effectively reshape vector data into tensor data, apply one of these low-cost linear maps, and obtain a new embedding that out-competes, e.g., SORS matrices on a rich class of vector data? Herein we answer this question to the affirmative using a vectorized form of a two-stage modewise tensor embedding matrix constructed along the lines of those proposed in [34]. *In particular, we show herein that a general class of random matrices exists which outperforms SORS embeddings on sufficiently low-dimensional manifold data with respect to both their provably achievable embedding dimensions and matrix-vector multiplication runtimes, all while maintaining similar embedding quality.* We consider this to be an exciting demonstration of the power of such modewise maps, and hope it helps to spur additional analysis of such JL embedding maps for tensor data going forward.

1.1 The Proposed Construction and A Motivating Experiment

We now present the proposed matrix construction aimed at combining the benefits of (i) fast JL-embeddings using matrices with a fast matrix-vector multiply and low memory requirements, with (ii) subgaussian matrices that have no simplifying structure but that offer optimal reduction in the embedding dimension of the given data. In particular, we will focus on an approach where we divide the data in blocks, apply a fast JL-map to each block, recombine the outputs, and then feed them to a sub-gaussian JL-embedding for additional compression. See Figure 2 for a graphical illustration. By designing each step carefully in this way we will see that one can retain the fast matrix-vector multiplication property of the first map along with the near-optimal dimension reduction of the second.

More specifically, the proposed matrices $E \in \mathbb{R}^{m_2 \times N}$ are constructed from two other matrices $B \in \mathbb{R}^{m_2 \times N/m_1}$ and $A \in \mathbb{R}^{m_1 \times m_1^2}$ where, for ease of notation, m_1^2 divides N. Given A and B as above, we let $C := \begin{pmatrix} A & & \\ & \ddots & \\ & & A \end{pmatrix} \in \mathbb{R}^{N/m_1 \times N}$ be the block diagonal matrix formed using N/m_1^2 copies of A, and then set

$$E := BC \in \mathbb{R}^{m_2 \times N}. \tag{1}$$

One can now see that this construction is analogous to reshaping the vector data one wishes to compress into a matrix, applying A to each column of the matrix, and then reshaping the resulting matrix back into a vector before applying B. As such, it is a specific example of a modewise

[1] Common choices for U include discrete cosine transform and Hadamard matrices. In addition, one can also see that choosing U to be a complex-valued DFT matrix outright will also work as a consequence of Euler's formula.

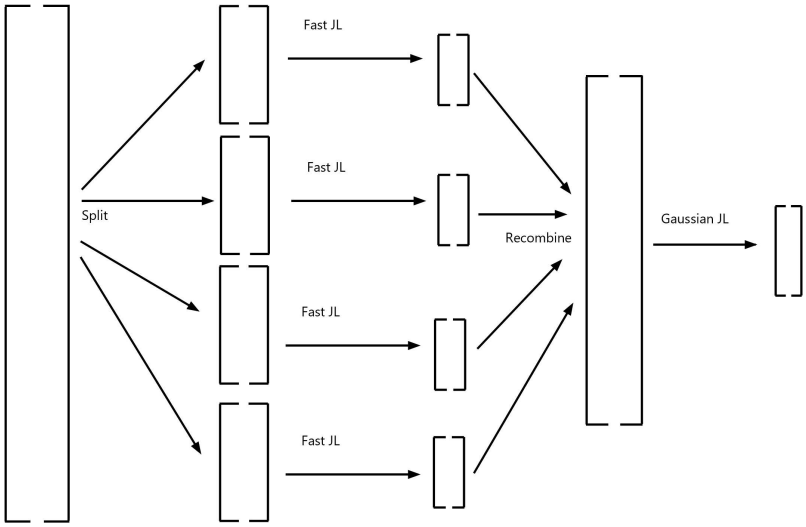

Figure 2: A Schematic diagram for an example matrix E of type (1). In this approach we split a vector into pieces, process each part with a fast JL map, and then recombine the outputs and feed it as a vector to a Gaussian JL map for optimal secondary dimensionality reduction. Note that this compression scheme is intrinsically parallel in nature and so should also be easily implementable in distributed settings.

JL-operator being applied to a vector after reshaping it into (in this case) a 2-mode tensor. When A above is chosen to be a matrix with a fast matrix-vector multiply (e.g., either a Partial Random Circulant (PRC) matrix [49, Corollary III.4], or a SORS matrix), and B is chosen to be a Gaussian random matrix, we obtain a matrix E corresponding to Figure 2.

The following lemma describes the properties of the matrices A and B that guarantee E in (1) will have a fast matrix vector multiply. We emphasize again that this lemma is compatible with choosing A as, e.g., either a PRC or SORS matrix, and B as a Gaussian matrix as per Figure 2.

Lemma 1.1. *Let $A \in \mathbb{R}^{m_1 \times m_1^2}$, $B \in \mathbb{R}^{m_2 \times N/m_1}$, $C \in \mathbb{R}^{N/m_1 \times N}$, and $E \in \mathbb{R}^{m_2 \times N}$ be as above in (1) with $m_1 \geq m_2$. Furthermore, suppose that $A \in \mathbb{R}^{m_1 \times m_1^2}$ has an $m_1^2 \cdot f(m_1)$ time matrix-vector multiplication algorithm. Then $E = BC \in \mathbb{R}^{m_2 \times N}$ will also have an $\mathcal{O}(N \cdot f(m_1))$-time matrix-vector multiply.*

Proof. The number of required operations for multiplying E against a vector is

$$\frac{N}{m_1^2}(m_1^2 \cdot f(m_1)) + \mathcal{O}\left(m_2 \frac{N}{m_1}\right) = \mathcal{O}(N \cdot f(m_1)).$$

Here, the first term comes from the $\dfrac{N}{m_1^2}$ multiplications of the matrix A that must be performed during a multiplication of a vector $\in \mathbb{R}^N$ by C. The second term results from a naive multiplication of a vector in the range of C by B, together with the assumption that $m_1 \geq m_2$. □

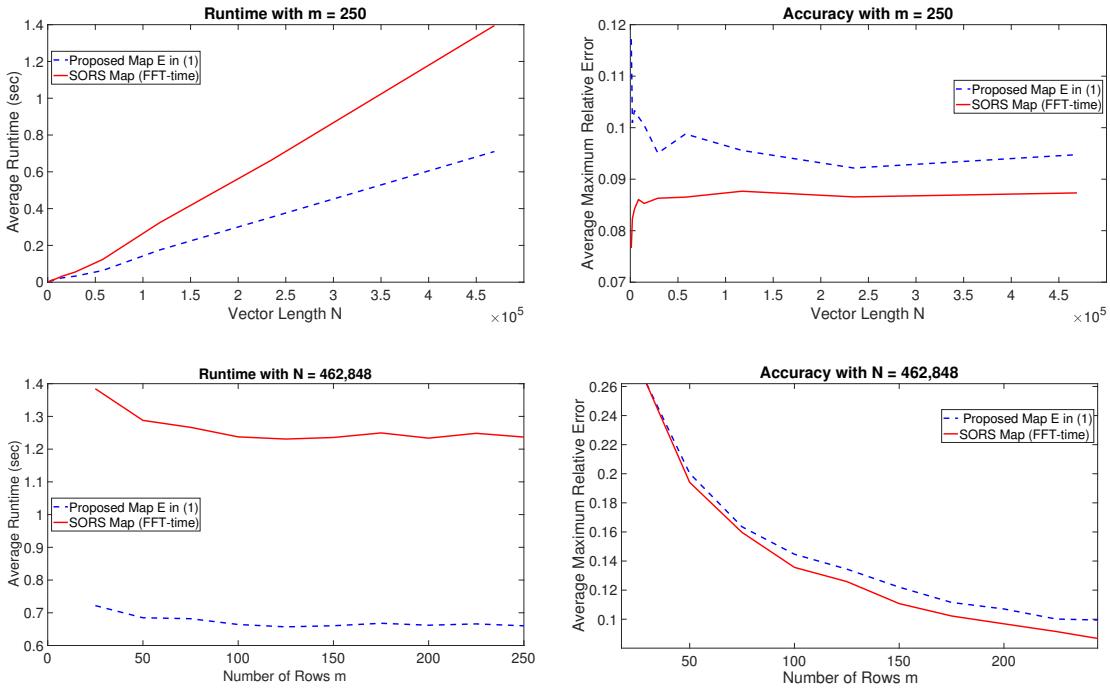

Figure 3: Proposed embedding matrices E of the form (1) where B is a Gaussian random matrix and A is a Discrete Fourier Transform (DFT)-based SORS matrix (in dotted blue), versus a standard $m \times N$ DFT-based SORS matrix with an $\mathcal{O}(N \log N)$-time matrix-vector multiply (in solid red). The reported runtimes are the average time in seconds needed to map randomly generated subsets of 100 vectors in $\mathbb{R}^N$ into $\mathbb{C}^m$, averaged over 100 randomly generated subsets. The errors reported in the accuracy plots are obtained by averaging the maximum relative errors $\max_{\mathbf{x} \in S} \left| \|E\mathbf{x}\|_2 - \|\mathbf{x}\|_2 \right| / \|\mathbf{x}\|_2$ over the 100 randomly generated subsets $S \subset \mathbb{R}^N$ for each given matrix $E \in \mathbb{C}^{m \times N}$. Here we briefly note that though our theory is developed for real-valued SORS matrices, it is relatively straightforward to extend all the results herein to the setting of complex-valued SORS matrices built using complex unitary matrices $U \in \mathbb{C}^{N \times N}$. See, e.g., the homework exercises in [29, Chapter 4.4] for more details.

Note that if A is chosen to be, e.g., a SORS matrix, $A = \sqrt{m_1} RUD$, where U is, e.g., an $m_1^2 \times m_1^2$ Discrete Cosine Transform (DCT) matrix, then $f(m_1)$ above will be $\mathcal{O}(\log m_1)$. As a result, the matrix E guaranteed by Lemma 1.1 will have an $\mathcal{O}(N \log m_1)$-time serial matrix-vector multiply in this setting, and can also easily benefit from parallel evaluation of C in (1) in a blockwise fashion. Thus, for example, we can see that such matrices E of the form (1) will have $o(N \log N)$-time matrix vector multiplies whenever m_1 can be chosen to be sufficiently small while still maintaining the desired level of embedding accuracy. But, how do they perform in practice? See Figure 3 for an example comparison between SORS and the proposed (1) random matrices when embedding finite point sets.

Looking at Figure 3 we can see that the proposed matrices E in (1) retain similar accuracy to standard SORS embeddings (i.e., their maximum relative errors generally differ by less than 1%)

while simultaneously being twice as fast or more for sufficiently large values of N. Taking such results as motivation, we will now turn our focus to proving theoretically that the proposed matrices E in (1) can also accurately embed submanifolds of $\mathbb{R}^N$ into much lower dimensional Euclidean space. In the process we will carefully compare the developed theory for the proposed matrices to similar embedding results via both standard SORS and sub-gaussian matrices. Our main results along these lines follow below.

1.2 Main Results and Discussion

We will begin by proving bounds for the embedding dimension of submanifolds of $\mathbb{R}^N$ with boundary using sub-gaussian random matrices for the purposes of later comparison. The reach of a submanifold used as a parameter below is provided in definition 4.1.

Theorem 1.1 (Embedding a Submanifold of $\mathbb{R}^N$ with Boundary via Sub-gaussian Random Matrices)**.** *Fix $\epsilon, p \in (0,1)$ and let A be a $m \times N$ sub-gaussian random matrix. Then, there exists a constant c' depending only on the distribution of the rows of A such that the following holds. Let $\mathcal{M} \hookrightarrow \mathbb{R}^N$ be a compact d-dimensional submanifold of $\mathbb{R}^N$ with $d \geq 2$, boundary $\partial\mathcal{M}$, finite reach $\tau_{\mathcal{M}}$, and volume $V_{\mathcal{M}}$.*[2] *Enumerate the connected components of $\partial\mathcal{M}$ and let τ_i be the reach of the i^{th} connected component of $\partial\mathcal{M}$ as a submanifold of $\mathbb{R}^N$. Set $\tau := \min_i\{\tau_{\mathcal{M}}, \tau_i\}$, let $V_{\partial\mathcal{M}}$ be the volume of $\partial\mathcal{M}$, and denote the volume of the d-dimensional Euclidean ball of radius 1 by ω_d. Finally, define*

$$\begin{aligned} \alpha_{\mathcal{M}} &:= \frac{V_{\mathcal{M}}}{\omega_d}\left(\frac{41}{\tau}\right)^d + \frac{V_{\partial\mathcal{M}}}{\omega_{d-1}}\left(\frac{81}{\tau}\right)^{d-1} \quad \text{and} \\ \beta_{\mathcal{M}} &:= \left(\alpha_{\mathcal{M}}^2 + 3^d\alpha_{\mathcal{M}}\right), \end{aligned} \tag{2}$$

and suppose that

$$m \geq \frac{c'\left(\sqrt{\ln(\beta_{\mathcal{M}})} + \sqrt{\ln(2/p)}\right)^2}{\epsilon^2}.$$

Then, $\frac{1}{\sqrt{m}}A$ will be an ϵ-JL embedding of $\mathcal{M}$ into $\mathbb{R}^m$ with probability at least $1-p$.

Proof. Apply Corollary 2.1 together with Theorem 4.5. □

Considering the sufficient lower bound on the embedding dimension m, one can analyze the dependence of m on d while keeping the other variables fixed. [3] If one puts $m = \frac{c'\left(\sqrt{\ln(\beta_{\mathcal{M}})} + \sqrt{\ln(2/p)}\right)^2}{\epsilon^2}$ as the least sufficient value of m, then m depends on d with order $\mathcal{O}(d\ln d)$. To see this we note that $\frac{1}{\omega_d} = \frac{\Gamma(\frac{d}{2}+1)}{\pi^{\frac{d}{2}}} = \mathcal{O}(d^d)$ so that $\frac{V_{\mathcal{M}}}{\omega_d}\left(\frac{41}{\tau}\right)^d + \frac{V_{\partial\mathcal{M}}}{\omega_{d-1}}\left(\frac{81}{\tau}\right)^{d-1} = \mathcal{O}(d^d)$ and $\beta_{\mathcal{M}} = \left(\frac{\alpha^2}{2} + 3^d\alpha\right) = \mathcal{O}(d^{2d})$. Comparing Theorem 1.1 to the state-of-the-art work in [20, Theorem 2], we have removed

[2] Note that one can prove similar results for one dimensional manifolds and for manifolds with infinite reach using the results herein. However, they require different definitions of $\alpha_{\mathcal{M}}$ and $\beta_{\mathcal{M}}$ below. See Theorem 4.4 and Proposition 4.3 for details on these special cases.

[3] One can show that $\beta_{\mathcal{M}}$ is guaranteed to be $> (d-1)\cdot 41^{2d-3}$ in this setting so that $m \geq c''d$ always holds in keeping with our intuition. See, e.g., Proposition 4.2 and (36) – (37) below for additional related discussion.

the mild geometric condition on reach $\frac{V_{\mathcal{M}}}{\tau^d} \geq \left(\frac{21}{2\sqrt{d}}\right)^d$ therein and can also accommodate the presence of a boundary while still having the embedding dimension, m, scale like $\mathcal{O}(d\log(d))$.

Most interestingly, we emphasize that the lower bound on the embedding dimension m for sub-gaussian matrices given by Theorem 1.1 has no dependence on the ambient dimension N whatsoever. However, sub-gaussian matrices are generally unstructured which means that they can not benefit from, e.g., fast specialized matrix vector multiplication methods such as Fast Fourier Transform (FFT) techniques. SORS matrices, on the other hand, do allow for such fast $\mathcal{O}(N\log N)$-time matrix-vector multiplies. The following result considers manifold embeddings by such fast-to-multiply structured matrices. SORS matrices and their constant K are introduced in definition 2.2.

Theorem 1.2 (Embedding a Submanifold of $\mathbb{R}^N$ with Boundary via SORS Matrices)**.** *Fix $\epsilon, p \in (0,1)$ and let $A = \sqrt{N/m}RUD$ be a $m \times N$ random SORS matrix with constant K. Then, there exist absolute constants c_0, c_1 such that such that the following holds. Let $\mathcal{M} \hookrightarrow \mathbb{R}^N$ be a compact d-dimensional submanifold of $\mathbb{R}^N$ with boundary $\partial\mathcal{M}$, define $\beta_{\mathcal{M}}$ as per (2) in Theorem 1.1, and suppose that*

$$m \geq \frac{c_0}{\epsilon^2}K^2\ln(\beta_{\mathcal{M}})\ln^2\left(\frac{c_1\ln(\beta_{\mathcal{M}})\ln(2/p)K^2}{\epsilon^2}\right)\ln(2/p)\ln(2eN/p).$$

Then, A will be an ϵ-JL embedding of $\mathcal{M}$ into $\mathbb{R}^m$ with probability at least $1-p$.

Proof. Apply Theorem 2.6 together with Theorem 4.5. □

Comparing Theorem 1.2 to Theorem 1.1 we can see that the the lower bound on the embedding dimension m provided by SORS matrices via Theorem 1.2 now does exhibit logarithmic N dependence[4], though these matrices can also benefit from fast matrix-vector multiplication techniques in practice. Comparing to prior state-of-the-art manifold embedding bounds for similar matrices [49, Corollary III.2] we see that they provide a sufficient embedding dimension lower bound of

$$m \geq \frac{c_0}{\epsilon^2}\left(d\ln\left(\frac{N}{\tau_{\mathcal{M}}\epsilon}\right) + \ln(V_{\mathcal{M}}/p)\right)\ln^4(N)\ln(1/p) \tag{3}$$

via SORS matrices. Again noting that $\ln(\beta_{\mathcal{M}})$ has no N dependence, we see that Theorem 1.2 improves the logarithmic dependence on N in (3) while again also allowing for the presence of a manifold boundary.

We are now prepared to prove our main result concerning the embedding of submanifolds of $\mathbb{R}^N$ that possibly have boundary via matrices which are structured along the lines of (1). More specifically, the matrices we propose for submanifolds of $\mathbb{R}^N$ herein (as well as for more general infinite sets with sufficiently small Gaussian width) will have the form

$$E := \sqrt{\frac{m_1}{m_2}}B\begin{pmatrix} RU & & \\ & \ddots & \\ & & RU \end{pmatrix}D \in \mathbb{R}^{m_2\times N} \tag{4}$$

[4] We believe that this is at least partially an artifact of the proof technique which ultimately depends on establishing the Restricted Isometry Property for a subsampled orthonormal basis system.

where $B \in \mathbb{R}^{m_2 \times N/m_1}$ has i.i.d. mean 0 and variance 1 sub-gaussian entries, $R \in \{0,1\}^{m_1 \times m_1^2}$ contains m_1 rows independently selected uniformly at random from the $m_1^2 \times m_1^2$ identity matrix, $U \in \mathbb{R}^{m_1^2 \times m_1^2}$ is a unitary matrix with $\max_{i,j} |u_{i,j}| \leq K/m_1$ for a constant K, and $D \in \{0,-1,1\}^{N \times N}$ is a random diagonal sign matrix with i.i.d. Rademacher random variables on its diagonal. We further assume that the matrix U has a $\mathcal{O}(m_1^2 \log(m_1))$-time matrix vector multiply (as will be the case if it is, e.g., a Hadamard or DCT matrix). We have the following manifold embedding result for this type of matrix.

Theorem 1.3 (Embedding a Submanifold of $\mathbb{R}^N$ with a Matrix of Type (4))**.** *There exist absolute constants $c_1, c_2, c_3, c_4 \in \mathbb{R}^+$ such that following holds for a given compact d-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^N$ with boundary $\partial\mathcal{M}$ and $\beta_{\mathcal{M}}$ defined as per (2) in Theorem 1.1. Suppose that $N \geq 50$, $\epsilon \in (0,1)$, $p \in \left(e^{-c_1 N}, 1/3\right)$, $\ln(\beta_{\mathcal{M}}) \leq c_2 \epsilon^2 \sqrt{N}/\ln^6(c_3 N/\epsilon p)$ and that $m_2 \in \mathbb{Z}^+$ satisfies*

$$m_2 \geq c_4 \ln(\beta_{\mathcal{M}}) \frac{\ln(N/\epsilon p)\ln(1/p)}{\epsilon^2}.$$

Then, one may randomly select an $m_2 \times N$ matrix E of the form in (4) such that E will be an ϵ-JL embedding of $\mathcal{M}$ into $\mathbb{R}^{m_2}$ with probability at least $1-p$. Furthermore, E will always have an $\mathcal{O}\left(N \cdot \left(\log\left(\sqrt{\ln(\beta_{\mathcal{M}})}/\epsilon\right) + \log\log(N/\epsilon p)\right)\right)$ run-time matrix-vector multiply.

Proof. Apply Theorem 3.6 in light of Remark 3.2, together with Theorem 4.5. □

First, we note that the mild restrictions on p and N in Theorem 1.3 are somewhat artificial and were made mainly to allow for greater simplification of the other derived bounds on m_2 and $d \leq \ln(\beta_{\mathcal{M}})$. They can be removed without real consequences beyond cosmetics. The restriction that $\ln(\beta_{\mathcal{M}}) \leq c_2 \epsilon^2 \sqrt{N}/\ln^6(c_3 N/\epsilon p)$ can also be made less severe at the cost of becoming less interpretable. However, it can not be discarded entirely and is ultimately required to allow for a valid choice of the intermediate matrix dimension $m_1 \leq \sqrt{N}$ to be made in the proposed construction (4). Ignoring log factors and considering ϵ and p to be constant, this restriction will ultimately always force the submanifolds we seek to embed via Theorem 1.3 to have dimension $d \lesssim \sqrt{N}$. Removing this restriction on d while preserving the nice lower bound on m_2 is of great interest, but appears to be difficult.

Similarly, retaining the restriction on d and obtaining a better lower bound on m_2 which is entirely independent of N similar to the one provided by Theorem 1.1 for sub-gaussian matrices would also be of great interest. This in fact appears possible if one can rigorously argue that the Gaussian width of $S_{\mathcal{M}} := U\left(U(\mathcal{M}-\mathcal{M}) - U(\mathcal{M}-\mathcal{M})\right)$ is always independent of N for a d-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^N$, where U here denotes normalization $\mathbf{x} \to \mathbf{x}/\|\mathbf{x}\|_2$, and $\mathcal{M}-\mathcal{M} := \left\{\mathbf{x}-\mathbf{y} \;\middle|\; \mathbf{x},\mathbf{y} \in \mathcal{M}, \mathbf{x} \neq \mathbf{y}\right\}$. Though this statement seems intuitively plausible, quantifying a concrete upper bound on the Gaussian width of $S_{\mathcal{M}}$ in terms of the original manifold parameters appears to be a non-trivial task. Another path toward removing the logarithmic N dependence in the lower bound for m_2 might be to carry out a modified chaining argument using, e.g., a result along the lines of Corollary 3.1 below at each level. Though this idea appears potentially promising in the abstract, the restrictions (8) that need to be satisfied in order to apply embedding results such as Corollary 3.1 to each cover involved complicate the standard approach.

Focusing now on the positive aspects of Theorem 1.3 we note that the lower bound on the embedding dimension m_2 it provides removes additional log factors from the embedding dimension lower bound for structured (SORS) matrices given by Theorem 1.2. In fact, ignoring constants and the logarithmic dependencies on ϵ and p, we believe that the lower bound provided by Theorem 1.3 for m_2 is the best one can ever hope to achieve in this setting via embedding arguments that require the embedding matrices to have the RIP. In addition, the structure of the proposed embedding matrices (4) endow them with $\mathcal{O}(N \log\log N)$-time matrix-vector multiplies whenever, e.g., $\ln(\beta_{\mathcal{M}}) \leq \ln^{c}(N)$ holds for fixed ϵ, p. Using the earlier estimates on $\beta_{\mathcal{M}}$, it is sufficient to have $d^{d/c} \leq N$. Finally, we again emphasize that these results hold for a general class of submanifolds of $\mathbb{R}^N$ both with and without boundary.

1.3 Paper Outline and Comments on Proof Elements

The proofs of all of Theorems 1.1, 1.2, and 1.3 are split into two independent parts: A general embedding result for infinite subsets $S \subset \mathbb{R}^N$ via a particular type of random matrix in terms of the subsets' Gaussian widths (i.e., Corollary 2.1, Theorem 2.6, and Theorem 3.6), combined with a Gaussian width bound for submanifolds of $\mathbb{R}^N$ which may (or may not) have boundary (i.e., Theorem 4.5). These component results are proven in three different sections below.

First, Corollary 2.1 and Theorem 2.6 are proven in Section 2, and are largely the result of updating existing compressive sensing and high dimensional probability bounds using some recent results by, e.g., Brugiapaglia, Dirksen, Jung, and Rauhut [13]. As a result, Section 2 is written in the form of a review of relevant prior work from these areas which makes some minor but useful (for our purposes later) modifications of existing theory along the way. The reader who is well familiar with these areas can safely skip to Section 3 and refer back as needed. To the less initiated reader, however, we recommend a more careful look and hope that the section may serve as a crash course to some current state-of-the-art results, techniques, and tools.

Next, Theorem 3.6 is proven in Section 3 in three phases. First, fast embedding results are proven for finite point sets with cardinalities bounded by $e^{\mathcal{O}(\sqrt{N})}$ using matrices of the form (1). We note that these results can be considered a simplification and generalization of a prior and more specialized JL-construction by Ailon and Liberty [4]. Next, these finite embedding results are then used together with a modified covering argument to prove that matrices of the form (1) also have the RIP for sufficiently small sparsities $s \lesssim \sqrt{N}$. In fact, for this range of sparsities, these structured RIP matrices have both an optimal number of rows (up to constant factors) and an $\mathcal{O}(N(\log(s) + \log\log(N))$-time matrix-vector multiply, a result of potential independent interest. Finally, Theorem 3.6 is then proven by using these new RIP matrices together with results by Oymak, Recht, and Soltanolkotabi [42].

To finish, Theorem 4.5 which bounds the Gaussian width of the closure of the unit secants of a submanifold $\mathcal{M}$ of $\mathbb{R}^N$ (potentially with boundary), i.e. $w\left(\overline{U(\mathcal{M}-\mathcal{M})}\right)$, is proven in Section 4. The proof begins by established covering number bounds for manifolds (possibly with boundary) by applying Günther's volume comparison theorem from Riemannian geometry. Next, covering number estimates for the unit secants of submanifolds of $\mathbb{R}^N$ (possibly with boundary) are then proven by modifying arguments motivated by the work of Eftekhari and Wakin for manifolds without boundary [20]. Once finished, these covering number estimates are then used in combination

with Dudley's inequality to prove Theorem 4.5.

In the next somewhat long section we will set terminology and review some relevant work from the compressive sensing and high dimensional probability literature.

2 Definitions, Notation, and Preliminaries

A matrix $A \in \mathbb{R}^{m \times N}$ is an ϵ-*JL* ***map*** *of a set* $T \subset \mathbb{R}^N$ *into* $\mathbb{R}^m$ if

$$(1-\epsilon)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1+\epsilon)\|\mathbf{x}\|_2^2$$

holds for all $\mathbf{x} \in T$. Note that this is equivalent to $A \in \mathbb{R}^{m \times N}$ having the property that

$$\sup_{\mathbf{x} \in T \setminus \{\mathbf{0}\}} \left| \|A(\mathbf{x}/\|\mathbf{x}\|_2)\|_2^2 - 1 \right| = \sup_{\mathbf{x} \in U(T)} \left| \|A\mathbf{x}\|_2^2 - 1 \right| \leq \epsilon,$$

where $U(T) \subset \mathbb{R}^N$ is the normalized version of $T \subset \mathbb{R}^N$ defined by

$$U(T) := \left\{ \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \;\middle|\; \mathbf{x} \in T \setminus \{\mathbf{0}\} \right\}.$$

We will say that a matrix $A \in \mathbb{R}^{m \times n}$ is an ϵ-*JL* ***embedding*** *of a set* $T \subset \mathbb{R}^n$ *into* $\mathbb{R}^m$ if A is an ϵ-JL map of

$$T - T := \{\mathbf{x} - \mathbf{y} \mid \mathbf{x}, \mathbf{y} \in T\}$$

into $\mathbb{R}^m$. Here we will be working with random matrices which will embed any fixed set T of bounded size measured in an appropriate way with high probability. Such matrix (distributions) are often called **oblivious** and discussed in the absence of any particular set T since they are independent of any properties of T beyond its size.

Of course, the discussion above now requires us to define what we actually mean by the "size" of an arbitrary and potentially infinite set $T \subset \mathbb{R}^N$. The following notions of the size of a set T will be useful and utilized heavily throughout. We will denote the cardinality of a finite set T by $|T|$. For a (potentially infinite) set $T \subset \mathbb{R}^N$ we then define its **radius** and **diameter** to be

$$\text{rad}(T) := \sup_{\mathbf{x} \in T} \|\mathbf{x}\|_2$$

and

$$\text{diam}(T) := \text{rad}(T - T) = \sup_{\mathbf{x}, \mathbf{y} \in T} \|\mathbf{x} - \mathbf{y}\|_2,$$

respectively. Given a value $\delta \in \mathbb{R}^+$ a δ**-cover of** T (also sometimes called a δ**-net of** T) will be a subset $S \subset T$ such that the following holds

$$\forall \mathbf{x} \in T \; \exists \mathbf{y} \in S \text{ so that } \|\mathbf{x} - \mathbf{y}\|_2 \leq \delta.$$

The δ**-covering number of** T, denoted by $\mathcal{N}(T, \delta) \in \mathbb{N}$, is then the smallest achievable cardinality of a δ-cover of T. Finally, the **Gaussian width** of a set T is defined as follows.

Definition 2.1. *[46, 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.

For more detail about the properties of the Gaussian width see [46, Proposition 7.5.2].

For simplicity we will focus on two general types of random matrices in this paper: sub-gaussian random matrices with independent, isotropic, and sub-gaussian rows (referred to simply as **sub-gaussian random matrices** below), and Krahmer-Ward Subsampled Orthonormal with Random Signs (SORS) matrices [35]. We will discuss each of these classes of random matrices in more detail next.

2.1 Sub-gaussian Random Matrices as Oblivious ϵ-JL maps

Sub-gaussian random matrices include, e.g., matrices with i.i.d. mean 0 and variance 1 Gaussian or Rademacher entries as special cases. We refer the reader to, e.g., [46, Section 2.5, Chapter 3, and Chapter 4] and/or [24, Chapters 7 and 9] for details regarding this rich class of random matrices. The following results demonstrate the use of these matrices as oblivious ϵ-JL maps of arbitrary sets.

Theorem 2.1 (See Theorem 9.1.1 and Exercise 9.1.8 in [46])**.** *Let A be $m \times N$ matrix whose rows are independent, isotropic, and sub-gaussain random vectors in $\mathbb{R}^N$. Let $p \in (0,1)$ and $T \subset \mathbb{R}^N$. Then there exists a constant c depending only on the distribution of the rows of A such that*

$$\sup_{\mathbf{x} \in T} \left| \|A\mathbf{x}\|_2 - \sqrt{m}\|\mathbf{x}\|_2 \right| \leq c \left[w(T) + \sqrt{\ln(2/p)} \cdot \mathrm{rad}(T) \right]$$

holds with probability at least $1 - p$.

Remark 2.1. *The constant c's dependence on the distributions of the rows of A can be bounded explicitly via their sub-guassian norms (see [46, Definition 3.4.1]). For simplicity we will neglect these more exact expressions and simply note here that once a distribution for A is fixed this constant will be completely independent of T and all its attributes. In particular, if the rows of A are all distributed identically as is common in practice then c will be an absolute constant with no dependence on any other quantities or entities whatsoever.*

The following simple corollary of Theorem 2.1 demonstrates how sub-gaussian matrices may be used to produce ϵ-JL maps of arbitrary subsets into lower dimensional Euclidean space with high probability.

Corollary 2.1 (Sub-gaussian Matrices Embed Infinite Sets)**.** *Let $S \subset \mathbb{R}^N$ and $\epsilon, p \in (0,1)$. Let A be a $m \times N$ sub-gaussian random matrix. Then, there exists a constant c' depending only on the distribution of the rows of A such that $\frac{1}{\sqrt{m}} A$ will be an ϵ-JL map of S into $\mathbb{R}^m$ with probability at least $1 - p$ provided that*

$$m \geq \frac{c' \left(w\left(U(S)\right) + \sqrt{\ln\left(2/p\right)} \right)^2}{\epsilon^2}.$$

Proof. Let $T = U(S) \subset \mathbb{S}^{N-1} := U(\mathbb{R}^N) = \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x}\|_2 = 1\}$. Since $T = U(S) \subset \mathbb{S}^{N-1}$, $\operatorname{rad}(T) = 1$ and $\|\mathbf{x}\|_2 = 1$ for all $\mathbf{x} \in T$. Furthermore, for all $u \in \mathbb{R}$ with $|u-1| \leq \epsilon/3$ one has that

$$|u^2 - 1| = |u+1||u-1| \leq (2 + \epsilon/3)(\epsilon/3) < \epsilon.$$

Hence, we may apply Theorem 2.1 to $T = U(S)$ with $m \geq \dfrac{9c\left(w(T) + \sqrt{\ln(2/p)}\right)^2}{\epsilon^2}$ to see that

$$\sup_{\mathbf{x} \in U(S)} \left| \left\| \frac{1}{\sqrt{m}} A\mathbf{x} \right\|_2^2 - 1 \right| \leq \epsilon$$

holds. □

The following simplification of Corollary 2.1 to finite sets S will be useful later.

Corollary 2.2 (Sub-gaussian Matrices Embed Finite Sets)**.** *Let $S \subset \mathbb{R}^N$ be finite and $\epsilon, p \in (0,1)$. Let A be a $m \times N$ sub-gaussian random matrix. Then, there exists a constant c'' depending only on the distribution of the rows of A such that $\dfrac{1}{\sqrt{m}}A$ will be an ϵ-JL map of S into $\mathbb{R}^m$ with probability at least $1 - p$ provided that*

$$m \geq \frac{c'' \ln(2|S|/p)}{\epsilon^2}.$$

Proof. Note that $T = U(S) \subset \mathbb{S}^{N-1}$ will also be finite with $|T| \leq |S|$, and with $\operatorname{diam}(T) \leq 2$. Hence, [46, Exercise 7.5.10] implies that $w(T) \leq c\sqrt{\ln|S|}$ for an absolute constant $c \in (1, \infty)$. As a consequence it suffices to take

$$\begin{aligned} m \geq c'' \frac{\ln(2|S|/p)}{\epsilon^2} &\geq c' \frac{2(c^2 \ln(|S|) + \ln(2/p))}{\epsilon^2} \\ &\geq c' \frac{\left(c\sqrt{\ln(|S|)} + \sqrt{\ln(2/p)}\right)^2}{\epsilon^2} \geq \frac{c'\left(w\left(U(S)\right) + \sqrt{\ln\left(2/p\right)}\right)^2}{\epsilon^2} \end{aligned}$$

for $c'' \in \mathbb{R}^+$ sufficiently large when applying Corollary 2.1. We also used $2(a^2 + b^2) \geq (a+b)^2$ in the line above. □

It can be shown that sub-gaussian random matrices are near-optimal with respect to the embedding dimension m they provide for ϵ-JL embeddings [30]. However, they are generally unstructured matrices which do not benefit from having, e.g., fast specialized algorithms for computing matrix-vector multiplies quickly. We will discuss more structured classes of random matrices that do have such algorithms next.

2.2 Oblivious ϵ-JL maps for Finite Sets from SORS Matrices

SORS matrices are derived from orthonormal bases and so can benefit from their inherent structure. They are defined as follows.

Definition 2.2. *[**SORS Matrices**] Let $U \in \mathbb{R}^{N\times N}$ be an orthogonal matrix obeying*

$$U^*U = I \qquad \text{and} \qquad \max_{i,j} |u_{i,j}| \leq \frac{K}{\sqrt{N}}$$

where I is the $N \times N$ identity matrix. Let $R \in \mathbb{R}^{m\times N}$ be a random matrix created by independently selecting m rows of I uniformly at random with replacement. Let $D \in \mathbb{R}^{N\times N}$ be a random diagonal matrix with i.i.d Rademacher random variables on its diagonal. Then $A = \sqrt{\frac{N}{m}}RUD$ is a Subsampled Orthogonal with Random Sign (SORS) matrix with constant $K \geq 1$.

The analysis of SORS matrices as ϵ-JL maps depends on the Restricted Isometry Constants (RICs) of the **Subsampled Orthonormal Basis (SOB)** matrices $\sqrt{N/m}RU$ with constant K defined above as a part of the SORS matrix definition. These constants are also closely associated with the Restricted Isometry Property (RIP) from compressive sensing [24].

Definition 2.3 (RICs). *[24, Definition 6.1] The s^{th} Restricted Isometry Constant (RIC) ϵ_s of a matrix $A \in \mathbb{R}^{m\times N}$ is the smallest $\epsilon \geq 0$ such that all at most s-sparse $x \in \mathbb{R}^N$ satisfy*

$$(1-\epsilon)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1+\epsilon)\|\mathbf{x}\|_2^2.$$

Definition 2.4 (RIP). *If a given value $\epsilon \in (0,1)$ is larger than the s^{th} RIC of A so that $\epsilon_s \leq \epsilon$ we say that A has the Restricted Isometry Property (RIP) of order (s,ϵ).*

As we shall see, the following theorem by Brugiapaglia, Dirksen, Jung, and Rauhut allows one to prove that a general class of random matrices have the RIP.

Theorem 2.2. *[13, Theorem 1.1] There exist absolute constants $\kappa > 0$ and $c_0, c_1 > 1$ such that the following holds. Let $X_1, ..., X_m$ be independent copies of a random vector $X \in \mathbb{C}^N$ with bounded coordinates, i.e. $\max_{1\leq i\leq N} |\langle X, \mathbf{e}_i\rangle| \leq K$ for some $K > 0$, where $\{\mathbf{e}_i\}_{i=1}^N$ is the standard basis of $\mathbb{C}^N$. Let $T \subseteq \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{x}\|_1 \leq \sqrt{s}\}, \epsilon \in (0,\kappa)$, and assume that*

$$m \geq c_0 K^2 \epsilon^{-2} s \ln(eN) \ln^2\left(sK^2/\epsilon\right).$$

Then, with probability exceeding $1 - 2\exp\left(-\epsilon^2 m/(sK^2)\right)$,

$$\sup_{\mathbf{y}\in T} \left| \frac{1}{m}\sum_{i=1}^m |\langle \mathbf{y}, X_i\rangle|^2 - \mathbb{E}|\langle \mathbf{y}, X\rangle|^2 \right| \leq c_1\epsilon\left(1 + \sup_{\mathbf{y}\in T} \mathbb{E}|\langle \mathbf{y}, X\rangle|^2\right).$$

Specializing Theorem 2.2 to the case of SOB matrices we arrive at the following corollary which upper bounds their RICs, thereby proving they have the RIP.

Corollary 2.3 (SOB Matrices have the RIP for Small ϵ)**.** *There exists absolute constants, $a_0, a_1 > 1$ and $a_2 > 0$ such that the following holds for any $\epsilon \in (0, a_2]$. Assume A is a $m \times N$ SOB matrix with*

$$m \geq a_0 K^2 \frac{s}{\epsilon^2} \ln(eN) \ln^2\left(\frac{a_1 sK^2}{\epsilon}\right).$$

Then A will have RIP of order (s,ϵ) with probability at least $1 - 2\exp(-\epsilon^2 m/(a_1^2 sK^2))$.

Proof. Using Theorem 2.2, we consider the set of unit length vectors $T := \{\mathbf{x} \in \mathbb{C}^N \text{ with } \|\mathbf{x}\|_2 = 1 \text{ and } \|\mathbf{x}\|_1 \leq \sqrt{s}\}$, which includes all unit length s-sparse vectors by Cauchy–Schwarz. Let X_j be the uniform selection of a row of $\sqrt{N}U$ for a unitary matrix $U \in \mathbb{C}^{N\times N}$. We then have $\mathbb{E}|\langle \mathbf{x}, X_j\rangle|^2 = \|\mathbf{x}\|_2^2 = 1$. Thus, if the rows of the SOB matrix A are selected uniformly at random we get that

$$\sup_{\mathbf{x}\in T}\left|\left\|\frac{1}{\sqrt{m}}A\mathbf{x}\right\|_2^2 - \|\mathbf{x}\|_2^2\right| \leq 2c_1\epsilon.$$

Changing constants to account for the extra $2c_1$-factor accompanying the ϵ above gives the stated bounds on the probability and m. □

The following additional variant of Corollary 2.3 provides an explicit probability variable, and will be more convenient to apply in some settings.

Corollary 2.4 (SOB Matrices have the RIP)**.** *There exist absolute constants $a_0', a_1' > 1$ such that the following holds. Let $\epsilon, p \in (0,1)$. Any SOB matrix $A \in \mathbb{R}^{m\times N}$ with constant K that has*

$$m \geq \frac{a_0'}{\epsilon^2}K^2 s\left(\ln(eN)\ln^2\left(\frac{a_1' sK^2}{\epsilon}\right) + \ln(e/p)\right)$$

will have the RIP of order (s,ϵ) with probability at least $1-p$.

Proof. If $\epsilon \leq a_2$ then

$$m \geq \max\{a_1^2, a_0\}K^2\frac{s}{\epsilon^2}\left(\ln(eN)\ln^2\left(\frac{a_1 sK^2}{\epsilon}\right) + \ln(e/p)\right) \geq \frac{a_1^2 sK^2\ln(e/p)}{\epsilon^2}. \tag{5}$$

Now Corollary 2.3 tells us that we will have the RIP of order (s,ϵ) with probability at least

$$1 - 2\exp(-\epsilon^2 m/(a_1^2 sK^2)) \geq 1 - 2p/e \geq 1 - p.$$

If $\epsilon \geq a_2$ Corollary 2.3 tells us that when

$$m \geq \max\{a_1^2, a_0\}K^2\frac{s}{\min^2\{a_2,1\}}\left(\ln(eN)\ln^2\left(\frac{a_1 sK^2}{\min\{a_2,1\}}\right) + \ln(e/p)\right) \geq \frac{a_1^2 sK^2\ln(e/p)}{a_2^2} \tag{6}$$

we will again have the RIP of order (s,ϵ) with probability at least

$$1 - 2\exp(-a_2^2 m/(a_1^2 sK^2)) \geq 1 - 2p/e \geq 1 - p.$$

Combining (5) and (6) we can, e.g., set $a_0' = \max\{a_1^2, a_0\}/\min\{a_2^2, 1\}$ and $a_1' = a_1/\min\{a_2, 1\}$. □

With Corollary 2.4 in hand we can now make a minor improvement to the embedding dimension m provided by the Krahmer-Ward theorem in the case of SORS matrices [35, Section 4].

Corollary 2.5 (SORS Matrices Embed Finite Sets)**.** *Let $S \subset \mathbb{R}^N$ be finite and $\epsilon, p \in (0,1)$. Let A be a $m \times N$ random SORS matrix with constant K. Then, there exist absolute constants c_0', c_1', c_2' such that A will be an ϵ-JL map of S into $\mathbb{R}^m$ with probability at least $1-p$ provided that*

$$m \geq c_0'\frac{K^2}{\epsilon^2}\ln(c_1'|S|/p)\cdot\left(\ln^2\left(\frac{\ln(c_2'|S|/p)K^2}{\epsilon}\right)\ln(eN) + \ln(2e/p)\right).$$

Proof. There are two steps: establishing an RIP bound and obtaining a JL map from the RIP bound. In both steps there is a failure probability which we control via the union bound. Let $s = 16 \ln(8|S|/p)$. For this choice of s [24, Theorem 9.36] guarantees that A will be an ϵ-JL map of S into $\mathbb{R}^m$ with probability at least $1 - p/2$ provided that $\sqrt{N/m}RU$ has the RIP of order $(2s, \epsilon/4)$. This RIP condition is provided by Corollary 2.4 with probability at least $1 - p/2$. Applying the union bound and adjusting the absolute constants now yields the desired result. □

Looking at Corollary 2.5 we can see that the embedding dimension m provided there is about a factor of $\mathcal{O}\left(\ln^2(\ln|S|)\log N\right)$ worse than that provided by Corollary 2.2 for sub-gaussian random matrices (holding ϵ, p, and K constant). One the other hand, if the unitary matrix U used to build the SORS matrix has an efficient matrix-vector multiply, then the SORS matrix will also have one. To try to get the best of both of these worlds (i.e., a near optimal embedding dimension together with a fast matrix-vector multiply) we will use the proposed construction (1). However, in order to demonstrate that this construction can in fact embed arbitrary (and potentially infinite) sets we will need a few more tools. These will be discussed in the next section.

2.3 Oblivious ϵ-JL maps for Infinite Sets via Structured Matrices

Referring back to Corollary 2.1, we can see that sub-gaussian random matrices can embed arbitrary infinite sets into lower dimensional Euclidean space. Note that we have not seen such a result for SORS matrices yet (note that, e.g., Corollary 2.5 only applies to finite sets). This is due to the proofs of such embedding results for infinite sets using structured matrices (such as SORS matrices) being significantly more involved in general. In this section we will outline a general approach for proving such results by Oymak, Recht, and Soltanolkotabi [42] which will require, among other things, the use of a couple of modified RIP definitions. The first one is essentially identical to the original RIP.

Definition 2.5 (Extended Restricted Isometry Property (ERIP) [42]). *Let $s \in [N]$ and $\epsilon \in \mathbb{R}^+$. A matrix $A \in \mathbb{R}^{m \times N}$ satisfies the extended RIP of order (s, ϵ) if*

$$\left| \|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \leq \max\{\epsilon, \epsilon^2\} \|\mathbf{x}\|_2^2$$

holds for all at most s-sparse $\mathbf{x} \in \mathbb{R}^N$.

Remark 2.2. *Note that the above definition only differs from the RIP in Definition 2.4 when $\epsilon \geq 1$.*

One can use results about the RICs of matrices to see that RIP results can be used to imply the ERIP for $\epsilon \geq 1$. In particular, the following facts are useful for this purpose.

Proposition 2.1. *[24, Proposition 6.6] For a matrix $A \in \mathbb{R}^{m \times N}$, let ϵ_s be the s^{th} restricted isometry constant of A. Then for integers $1 \leq s \leq t$,*

$$\epsilon_t \leq \frac{t-d}{s}\epsilon_{2s} + \frac{d}{s}\epsilon_s, \qquad d = gcd(s,t).$$

In particular since $\epsilon_s \leq \epsilon_{2s}$ we have that

$$\epsilon_t \leq \frac{t}{s}\epsilon_{2s}.$$

Proposition 2.2. *Let $s \in \mathbb{N}$, and $k \geq 1$ be a real number. Then*

$$\epsilon_s \leq k\epsilon_{\left(2\lceil s/k \rceil\right)}.$$

Proof. From proposition 2.1, for $1 \leq s \leq t$, we have $\epsilon_t \leq \frac{t}{s}\epsilon_{2s}$. Since $k \geq 1$ and s is an integer, $1 \leq \lceil s/k \rceil \leq s$, and hence $\epsilon_s \leq \frac{s}{\lceil \frac{s}{k} \rceil}\epsilon_{\left(2\lceil \frac{s}{k} \rceil\right)} \leq k\epsilon_{\left(2\lceil \frac{s}{k} \rceil\right)}$. □

As noted above, the RIP and ERIP coincide for $\epsilon < 1$. With the following two propositions we can now further see that the ERIP of order (s, ϵ) with $\epsilon > 1$ follows from the RIP of order, e.g., $(2\lceil s/\epsilon^2 \rceil, 0.9)$.

Lemma 2.1. *Let $b, a \in (0, 1]$ with $a < b$, $s \in [N]$, $\epsilon \in [b, \infty)$, and suppose that $A \in \mathbb{R}^{m \times N}$ has the RIP of order $(2\lceil sb^2/\epsilon^2 \rceil, ab)$. Then, A will also have the ERIP of order (s, ϵ).*

Proof. Note that A will have the RIC $\epsilon_{\left(2\lceil sb^2/\epsilon^2 \rceil\right)} \leq ab$ by assumption. Applying Proposition 2.2 with $k = \epsilon^2/b^2 \geq 1$ we can then see that $\epsilon_s \leq \frac{\epsilon^2}{b^2}\epsilon_{\left(2\lceil sb^2/\epsilon^2 \rceil\right)} \leq \epsilon^2 \frac{a}{b} < \epsilon^2 \leq \max\{\epsilon, \epsilon^2\}$. □

We are now prepared to define the central RIP variant of this section.

Definition 2.6. (Multiresolution Restricted Isometry Property (MRIP)) *[42, Definition 2.2]. A matrix $A \in \mathbb{R}^{m \times N}$ satisfies the MRIP of order (s, ϵ) if it possesses the extended RIP of order $(2^l s, 2^{l/2}\epsilon)$ for all integers l with $0 \leq l \leq \lceil \log_2(N/s) \rceil$.*

The following theorem can be used to convert RIP guarantees into MRIP guarantees.

Theorem 2.3 (RIP implies MRIP)**.** *Let $a \in (0, 1]$, $A \in \mathbb{R}^{m \times N}$ be a random matrix, and $f_N : [N] \times (0, a) \times (0, 1) \to \mathbb{R}^+$ have the property that*

$$m \geq f_N(s', \epsilon', p') \implies A \text{ has the RIP of order } (s', \epsilon') \text{ with probability at least } 1 - p'.$$

Fix $\epsilon, p \in (0, 1)$ and $s \in [N]$. Then, A will have the MRIP of order (s, ϵ) with probability at least $1 - p$ provided that

$$m \geq \max\left\{ f_N\left(2\left\lceil \frac{a^2 s}{\epsilon^2} \right\rceil, a^2/2, \frac{p}{\lceil \log_2(N/s) \rceil + 1}\right), \max_{0 \leq l < L} f_N\left(2^l s, 2^{l/2}\epsilon, \frac{p}{\lceil \log_2(N/s) \rceil + 1}\right)\right\} \quad (7)$$

where $L := \min\{2\log_2(a/\epsilon), \lceil \log_2(N/s) \rceil + 1\}$.

Proof. We need to establish that A has the ERIP of order $(2^l s, 2^{l/2}\epsilon)$ for all integers l with $0 \leq l \leq \lceil \log_2(N/s) \rceil$. To do so we will consider two separate ranges of the integers l:

(a) The L integers $l < L \leq 2\log_2(a/\epsilon)$ for which $2^{l/2}\epsilon < a$ holds, and

(b) The remaining $\lceil \log_2(N/s) \rceil + 1 - L$ integers l for which $2^{l/2}\epsilon \geq a$ holds.

For integers l in range (a) the ERIP of order $(2^l s, 2^{l/2}\epsilon)$ is equivalent to the RIP of order $(2^l s, 2^{l/2}\epsilon)$, and so choosing m as in (7) immediately provides each of these L ERIP conditions with probability at least $1 - p/(\lceil \log_2(N/s) \rceil + 1)$. For each of the integers l in range (b) the assumed RIP of order $\left(2\left\lceil 2^l s a^2/(2^{l/2}\epsilon)^2 \right\rceil, a^2/2\right)$ together with an application of Lemma 2.1 with $b \leftarrow a$ and $a \leftarrow a/2$ yields the desired result, where $x \leftarrow y$ means substitute y for x. Again, one can see that choosing m as in (7) therefore provides each of these $\lceil \log_2(N/s) \rceil + 1 - L$ ERIP conditions with probability at least $1 - p/(\lceil \log_2(N/s) \rceil + 1)$. An application of the union bound now establishes that A will therefore satisfy all of the $\lceil \log_2(N/s) \rceil + 1$ required ERIP conditions with probability at least $1 - p$ as claimed. □

Using Theorem 2.3 with $a = 1$ together with Corollary 2.4 we can now see that SOB matrices have the MRIP.

Theorem 2.4 (SOB Matrices have the MRIP)**.** *There exist absolute constants $c_0', c_1' > 1$ such that the following holds. Let $\epsilon, p \in (0,1)$. Any SOB matrix $A \in \mathbb{R}^{m \times N}$ with constant K that has*

$$m \geq \frac{c_0'}{\epsilon^2} K^2 s \ln(eN/p) \ln^2\left(\frac{c_1' s K^2}{\epsilon^2}\right)$$

will have the MRIP of order (s, ϵ) with probability at least $1 - p$.

Having defined and discussed the MRIP condition we can now state the main theorem of [42] which will ultimately allow us to construct ϵ-JL maps for arbitrary infinite sets using their Gaussian width via structured matrices (including, e.g., SORS matrices).

Theorem 2.5 (MRIP implies Embedding of Infinite Sets [42])**.** *Fix $p, \epsilon \in (0,1)$. Let $T \subset \mathbb{R}^N$ and suppose that $E \in \mathbb{R}^{m \times N}$ has the MRIP of order (s, ϵ') with*

$$s = 200(1 + \ln(1/p)) \qquad \text{and} \qquad \epsilon' \leq \frac{\epsilon \cdot \mathrm{rad}(T)}{c \cdot \max\{\mathrm{rad}(T), w(T)\}}$$

where $c > 0$ is an absolute constant. Let $D \in \mathbb{R}^{N \times N}$ be a random diagonal matrix with i.i.d Rademacher random variables on its diagonal. Then, the matrix $A = ED$ will obey

$$\sup_{\mathbf{x} \in T} \left| \|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \leq \epsilon \cdot \mathrm{rad}^2(T)$$

with probability at least $1 - p$.

We can now use Theorems 2.4 and 2.5 to prove a generalized version of Corollary 2.5 that still holds when S is an infinite set.

Theorem 2.6 (SORS Matrices Embed Infinite Sets)**.** *Let $S \subset \mathbb{R}^N$ and $\epsilon, p \in (0,1)$. Let $A = \sqrt{N/m} RU'D$ be a $m \times N$ random SORS matrix with constant K. Then, there exist absolute constants c_0, c_1 such that A will be an ϵ-JL map of S into $\mathbb{R}^m$ with probability at least $1 - p$ provided that*

$$m \geq \frac{c_0}{\epsilon^2} K^2 w^2(U(S)) \ln^2\left(\frac{c_1 w^2(U(S)) \ln(2/p) K^2}{\epsilon^2}\right) \ln(2/p) \ln(2eN/p).$$

Note: We use the U' notation for the unitary matrix in the theorem 2.6 above to avoid confusion with the notation $U(S)$ for the set of unit vectors corresponding to set S.

Proof. We will apply Theorem 2.5 with $T = U(S)$ and $p \leftarrow p/2$ noting that $\text{rad}(T) = 1$. Hence, by the union bound it suffices to invoke Theorem 2.4 for $\sqrt{N/m}RU' \in \mathbb{R}^{m \times N}$ with $s = 200(1+\ln(2/p))$ and $\epsilon' = \epsilon/(c + c \cdot w(T))$ again with $p \leftarrow p/2$. Doing so we learn that we will obtain the desired result as long as

$$m \geq \frac{c_0''}{\epsilon^2} K^2 (1 + w(T))^2 (1 + \ln(2/p)) \ln(2eN/p) \ln^2 \left(\frac{c_1''(1 + w(T))^2 (1 + \ln(2/p)) K^2}{\epsilon^2} \right).$$

Simplifying and combining absolute constants now leads to our final bound. □

Looking at Theorem 3.3 in [42] we can see that Theorem 2.6 improves the bound on m provided there while retaining the fast $\mathcal{O}(N \log N)$-time matrix-vector multiplies provided by SORS matrices. It is important to remember, however, that unstructured sub-gaussian matrices provide the smallest bounds on m (recall Corollary 2.1) in the setting where fast matrix-vector multiplies are of secondary importance. We now have all the tools necessary to prove that our proposed construction (1) can serve as an oblivious ϵ-JL map for infinite sets.

3 New Fast Embeddings for Infinite Sets

This section is devoted to showing that the variant of the proposed construction (1) corresponding to Figure 2 can indeed embed arbitrary infinite subsets of $\mathbb{R}^N$ into $\mathbb{R}^m$ with m near optimal. We will do this in four steps. First, we will establish that the proposed construction (1) can indeed embed finite point sets near-optimally provided that their cardinality is not too large. As mentioned above, this result can be considered a simplification and generalization of a prior embedding result due to Ailon and Liberty [4]. Once we have the embedding result for finite point sets, we will then show that, in fact, it also means that our proposed matrices (1) have the RIP for sufficiently small sparsities. Next, having established the RIP we will then prove the MRIP for the proposed matrices by applying Theorem 2.3. Finally, Theorem 2.5 ([42, Theorem 3.1]) can then be used to prove the desired oblivious embedding result for arbitrary infinite sets. We are now prepared to begin.

3.1 The Case of Finite Point Sets

The following lemma shows that a very general set of choices for both $A \in \mathbb{R}^{m_1 \times m_1^2}$ and $B \in \mathbb{R}^{m_2 \times N/m_1}$ in the proposed construction (1) lead to a matrix $E \in \mathbb{R}^{m_2 \times N}$ which will embed arbitrary finite subsets of $\mathbb{R}^N$. Before stating the result, however, we need some additional notation that will be useful later. Let $P_j : \mathbb{R}^N \mapsto \mathbb{R}^{m_1^2}$ for $j \in [N/m_1^2]_0 := \{0, \ldots, \lceil N/m_1^2 \rceil - 1\}$ be the orthogonal projection defined by $(P_j(\mathbf{x}))_\ell := x_{jm_1^2+\ell}$ for all $\ell \in [m_1^2] := \{1, \ldots, m_1^2\}$. For notational simplicity we will generally assume that m_1^2 divides N below. If not, $P_{\lceil N/m_1^2 \rceil - 1}$ can still map into m_1^2 by padding its output with zeros as needed. All instances of N/m_1 can then also be replaced by $m_1 \lceil N/m_1^2 \rceil$ in such cases without harm. We have the following result.

Lemma 3.1. *Let $\epsilon \in (0, 1)$, $S \subset \mathbb{R}^N$ be finite, and $A \in \mathbb{R}^{m_1 \times m_1^2}$, $B \in \mathbb{R}^{m_2 \times N/m_1}$, $C \in \mathbb{R}^{N/m_1 \times N}$, and $E = BC \in \mathbb{R}^{m_2 \times N}$ be as above in (1). Furthermore, suppose that*

(a) A is an ϵ-JL map of P_jS into $\mathbb{R}^{m_1}$ for all $j \in [N/m_1^2]_0$, and that

(b) B is an ϵ-JL map of CS into $\mathbb{R}^{m_2}$.

Then, E will be a 3ϵ-JL map of S into $\mathbb{R}^{m_2}$.

Proof. To begin we note that C will be an ϵ-JL map of S into $\mathbb{R}^{N/m_1}$ since

$$\begin{aligned}\big|\|C\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\big| &= \left|\sum_{j\in[N/m_1^2]_0} \|AP_j\mathbf{x}\|_2^2 - \|P_j\mathbf{x}\|_2^2\right| \leq \sum_{j\in[N/m_1^2]_0} \big|\|AP_j\mathbf{x}\|_2^2 - \|P_j\mathbf{x}\|_2^2\big| \\ &\leq \epsilon \sum_{j\in[N/m_1^2]_0} \|P_j\mathbf{x}\|_2^2 = \epsilon\|\mathbf{x}\|_2^2\end{aligned}$$

holds for all $\mathbf{x} \in S$ by assumption (a) about A. As a result, we can further see that E will be a 3ϵ-JL map of S into $\mathbb{R}^{m_2}$ since

$$\begin{aligned}(1-2\epsilon)\|\mathbf{x}\|_2^2 \leq (1-\epsilon)^2\|\mathbf{x}\|_2^2 \leq (1-\epsilon)\|C\mathbf{x}\|_2^2 \leq \|E\mathbf{x}\|_2^2 \\ \leq (1+\epsilon)\|C\mathbf{x}\|_2^2 \leq (1+\epsilon)^2\|\mathbf{x}\|_2^2 \leq (1+3\epsilon)\|\mathbf{x}\|_2^2\end{aligned}$$

will hold for all $\mathbf{x} \in S$. Here we have used assumption (b) about B to obtain the third and fourth inequalities just above. □

We can now use Lemma 3.1 to prove the promised fast ϵ-JL mapping result for finite sets.

Theorem 3.1 (Fast Embedding of Finite Sets by General Setup)**.** *Let $\epsilon, p \in (0,1)$, $S \subset \mathbb{R}^N$ be finite, $A = \sqrt{m_1}RUD$ be an $m_1 \times m_1^2$ random SORS matrix with constant K, and $B \in \mathbb{R}^{m_2 \times N/m_1}$ have i.i.d. mean zero, sub-gaussian entries. Furthermore, suppose that $m_1, m_2 \in \mathbb{Z}^+$ satisfy*

$$\begin{aligned}\sqrt{N} \;\geq\; m_1 \;&\geq c_0 \frac{K^2}{\epsilon^2} \ln(c_1 N|S|/m_1^2 p) \cdot \ln^2\left(\frac{\ln(c_2 N|S|/m_1^2 p)K^2}{\epsilon}\right) \ln(4eN/p) \quad \text{and} \\ m_2 \;&\geq \frac{c_3 \ln(4|S|/p)}{\epsilon^2}\end{aligned}$$

where $c_0, c_1, c_2, c_3 \in \mathbb{R}^+$ are absolute constants. Then, $E = \dfrac{1}{\sqrt{m_2}}BC \in \mathbb{R}^{m_2 \times N}$ as in (1) will be an ϵ-JL map of S into $\mathbb{R}^{m_2}$ with probability at least $1-p$. Furthermore, if $U \in \mathbb{R}^{m_1^2 \times m_1^2}$ has an $m_1^2 \cdot f(m_1)$ time matrix-vector multiplication algorithm, then E will have an $\mathcal{O}(N \cdot f(m_1))$-time matrix-vector multiply.

Proof. Note the stated result follows from the union bound together with Lemma 3.1 provided that its assumptions (a) and (b) both hold with probability at least $1 - p/2$ for $\epsilon \leftarrow \epsilon/3$. Hence, we seek to establish that both of these assumptions will hold with probability at least $1 - p/2$ for our choices of A and B above. This can be done by applying Corollaries 2.5 and 2.2, respectively, utilizing the union bound and adjusting constants as necessary. Finally, the runtime result for E follows from the structure of A combined with Lemma 1.1 after noting that c_0 and c_1 can easily be increased, if necessary, to ensure that $m_1 \geq m_2$ always holds. □

Note that an application of Theorem 3.1 requires a valid choice of m_1 to be made. This will effectively limit the sizes of the sets S which we can embed quickly below. In order to make the discussion of this limitation a bit easier below we can further simplify the lower bound for m_1 by noting that for a fixed and nonempty $S \subset \mathbb{R}^N$ with, e.g., $N \geq 4e$ we will have

$$\ln(c_1 N|S|/m_1^2 p) \cdot \ln^2\left(\frac{\ln(c_2 N|S|/m_1^2 p)K^2}{\epsilon}\right) \ln(4eN/p) \leq c \ln(N|S|/p) \cdot \ln^3\left(\frac{NK^2}{\epsilon p}\right)$$

for an absolute constant $c \in \mathbb{R}^+$, provided that $|S| \leq p m_1^2 e^N/N$. As a consequence, we may weaken the lower bound for m_1 and instead focus on the smaller interval

$$\sqrt{N} \geq m_1 \geq c' \frac{K^2}{\epsilon^2} \ln(N|S|/p) \cdot \ln^3\left(\frac{NK^2}{\epsilon p}\right)$$

for simplicity. Further assuming that K is upper bounded by a universal constant below (as it will be in all subsequent applications) we can see that our smaller range for m_1 will be nonempty whenever

$$1 \leq |S| \leq \frac{p}{N} e^{c'' \epsilon^2 \sqrt{N}/\ln^3\left(\frac{N}{\epsilon p}\right)} \leq p m_1^2 e^N/N \tag{8}$$

holds for another sufficiently small and absolute constant $c'' \in \mathbb{R}^+$. We will use (8) below to limit the sizes of the sets that we embed so that Theorem 3.1 can always be applied with a valid minimal choice of $m_1 \leq c''' \frac{K^2}{\epsilon^2} \ln(N|S|/p) \ln^3\left(\frac{NK^2}{\epsilon p}\right) \leq \sqrt{N}$ below. The following corollary of Theorem 3.1 is based on making more explicit choices for both A and B.

Corollary 3.1 (Fast Embedding of Finite Sets by SORS and Sub-gaussian Matrices)**.** *There exist absolute constants $c, c' \in \mathbb{R}^+$ such that the following holds. Let $\epsilon, p \in (0,1)$ and $S \subset \mathbb{R}^N$ with $N \geq 4e$ be finite with cardinality satisfying (8). Then, one may randomly select an $m \times N$ matrix E of the form (1) such that E will be an ϵ-JL map of S into R^m with probability at least $1-p$ provided that*

$$m \geq c\epsilon^{-2} \ln(|S|/p).$$

Furthermore, E will always have an $\mathcal{O}(N \log N)$ run-time matrix-vector multiply, and will in fact have, e.g., an $\mathcal{O}(N \log(\log N))$-time matrix-vector multiply for all S, p, ϵ with $|S|/p \leq N^{c'}$ and $\epsilon \geq 1/N^{c'}$.

Proof. We will let B have, e.g., i.i.d. Rademacher entries and will choose $U \in \mathbb{R}^{m_1^2 \times m_1^2}$ to be, e.g., a Hadamard or DCT matrix (see, e.g., [24, Section 12.1].) Making either choice for U will endow A with an $\mathcal{O}(m_1^2 \log(m_1))$-time matrix vector multiply via FFT-techniques, and will also ensure that $K = \sqrt{2}$ always suffices. As a result, we note that $f(m_1) = \mathcal{O}(\log(m_1))$ in Theorem 3.1. Combining this with the runtime guarantee of Theorem 3.1 gives the runtime bound when using the minimal choice of m_1. The lower bound for m results from the m_2 lower bound in Theorem 3.1. □

Looking at Corollary 3.1 we can see that the resulting matrices E achieve near-optimal embedding dimensions while simultaneously having $o(N \log N)$-time matrix vector multiplies for sufficiently small finite sets. Comparing Corollary 3.1 to Corollary 2.5 we can see that our proposed matrices of the form (1) have matrix-vector multiplies which are always at least as fast as SORS matrices

while simultaneously improving on their current best embedding dimension, m, bounds by a multiplicative factor of size roughly $\Theta\left(\ln^2\left(\frac{\ln(c|S|/p)}{\epsilon}\right)\ln(eN)+\ln(2e/p)\right)$. Of course, it must also be remembered that Corollary 3.1 only applies to finite sets S whose cardinality satisfies (8) whereas Corollary 2.5 applies more generally to larger sets.

3.2 New Fast Oblivious Subspace Embeddings and RIP Matrices

Let $S \subset \mathbb{R}^N$ be a d-dimensional subspace. The following fact will be useful.

Lemma 3.2 (See, e.g., Corollary 4.2.13 in [46]). *Let $S^{d-1} \subset S \subset \mathbb{R}^N$ be the $d-1$-dimensional unit Euclidean sphere in S. Then $\mathcal{N}(S^{d-1},\delta) \leq \left(\frac{3}{\delta}\right)^d$ for all $\delta > 0$.*

We are now prepared to apply Corollary 3.1 in order to produce an oblivious ϵ-JL map of S into $\mathbb{R}^m$ with m near-optimal.

Theorem 3.2 (Fast Oblivious Subspace Embedding). *There exist absolute constants $c, c' \in \mathbb{R}^+$ such that following holds for d-dimensional subspaces of $\mathbb{R}^N$. Let $\epsilon, p \in (0,1)$ and $S \subset \mathbb{R}^N$ with $N \geq 50$ be a d-dimensional subspace with $d \leq c\epsilon^2\sqrt{N}/\ln^4(N/\epsilon p) - 1$. Furthermore, suppose that $m \in \mathbb{Z}^+$ satisfies*

$$m \geq c' d\epsilon^{-2}\ln\left(1/\epsilon\sqrt[d]{p}\right).$$

Then, one may randomly select an $m \times N$ matrix E of the form in (1) such that E will be an ϵ-JL embedding of S into $\mathbb{R}^m$ with probability at least $1-p$. Furthermore, E will always have an $\mathcal{O}\left(N\cdot\left(\log(d/\epsilon^2)+\log\log\left(N/\epsilon p\right)\right)\right)$-time matrix-vector multiply.

Proof. Note that, e.g., Lemma 3 in [34] implies the desired embedding result if E embeds an $\epsilon/16$-net of S^{d-1} = the $d-1$-dimensional unit Euclidean sphere in S. Applying Corollary 3.1 to such a minimal net whose size is bounded by Lemma 3.2 then finishes the proof. □

With Theorem 3.2 in hand we can now easily consider RIP matrices of order (s,ϵ) of the form in (1). The approach proposed in [8], for example, would be to simply apply Theorem 3.2 to all $\binom{N}{s}$ subspaces of $\mathbb{R}^N$ spanned by s canonical basis vectors, and then to use the union bound. The following bound on $\binom{N}{s}$ is useful for such a strategy.

Lemma 3.3. *[24, Lemma C.5] For integers $N \geq s > 0$,*

$$\left(\frac{N}{s}\right)^s \leq \binom{N}{s} \leq \left(\frac{eN}{s}\right)^s.$$

Pursuing the simple strategy above yields the following RIP result which we will not use going forward due to its highly strict requirements on the size of s. Nonetheless, we state it here for the purposes of comparison.

Theorem 3.3. *There exist absolute constants $c, c', c'' \in \mathbb{R}^+$ such that following holds for $N \geq 50$. Let $\epsilon, p \in (0,1)$ and $s^5 \leq c\epsilon^2\sqrt{N}/\ln^4(c'N/s\epsilon p)$.*[5] *Furthermore, suppose that $m \in \mathbb{Z}^+$ satisfies*

$$m \geq c'' s\epsilon^{-2} \ln\left(eN/s\epsilon\sqrt[s]{p}\right).$$

Then, one may randomly select an $m \times N$ matrix E of the form in (1) such that E will have the RIP of order (s,ϵ) with probability at least $1-p$. Furthermore, E will always have an $\mathcal{O}\left(N \cdot \left(\log(s/\epsilon^2) + \log\log\left(N/\epsilon p\right)\right)\right)$-time matrix-vector multiply.

Remark 3.1. *Note that Theorem 3.3 requires $s^5 \leq c\epsilon^2\sqrt{N}/\ln^4(c'N/s\epsilon p)$ to hold. In fact, simply being more careful about the p dependence in the derivation of (8) can improve the exponent 5 on s above, even within this simple proof framework. However, achieving linear scaling on s appears to require us to use a different argument that avoids aggressive use of the union bound at this stage.*

The following alternate and improved RIP result achieves better scaling on the allowable size of s. It is proven using a covering argument over all unit length s-sparse vectors as opposed to the simpler approach outlined above. Effectively this alternate argument allows us to scale $|S|$ in an expression analogous to (8) by a factor of $\binom{N}{s}$ while leaving p fixed, instead of forcing us to apply (8) with $p \to p/\binom{N}{s}$.

Theorem 3.4 (Fast RIP Matrices)**.** *There exist absolute constants $c, c', c'' \in \mathbb{R}^+$ such that following holds for $N \geq 50$. Let $\epsilon \in \left(0, \frac{1}{3}\right)$, and $p \in \left(e^{-N}, \frac{1}{3}\right)$, and $s \leq c\epsilon^2\sqrt{N}/\ln^5(c'N/\epsilon p)$. Furthermore, suppose that $m \in \mathbb{Z}^+$ satisfies*

$$m \geq c'' s\epsilon^{-2} \ln\left(N/\epsilon\sqrt[s]{p}\right).$$

Then, one may randomly select an $m \times N$ matrix E of the form in (1) such that E will have the RIP of order (s,ϵ) with probability at least $1-p$. Furthermore, E will always have an $\mathcal{O}\left(N \cdot \left(\log(s/\epsilon^2) + \log\log\left(N/\epsilon p\right)\right)\right)$-time matrix-vector multiply.

Proof. See Appendix A. □

Comparing Theorem 3.4 to Theorem 3.3, we can see that Theorem 3.4 applies to a much larger range of sparsities s. Nonetheless, both theorems achieve a near-optimal scaling of the embedding dimension m and have fast matrix-vector multiplies. Comparing Theorem 3.4 to Corollary 2.4 we can see that our proposed matrices of the form (1) have matrix-vector multiplies which are always at least as fast as SOB RIP matrices while simultaneously improving on the current best bounds for their embedding dimension, m, by a multiplicative factor of size roughly $\Theta\left(\ln^2 s\right)$, having fixed ϵ and p. Of course, it must also be remembered that Theorem 3.4 applies to a smaller range of sparsities than Corollary 2.4 does. We are now equipped with the tools necessary to prove our main oblivious embedding result for infinite sets.

[5] The condition on s here is highly pessimistic. See Remark 3.1 for additional discussion about other admissible upper bounds which scale better in s.

3.3 Fast Embeddings of Infinite Sets with Small Gaussian Width

Having proven the RIP for matrices of the form (1) we can now establish the MRIP for such matrices using Theorem 2.3 with $a = 1/3$. Doing so while carefully considering the domain of the function f_N corresponding to Theorem 3.4 produces the following result. As usual, the absolute constants have been adjusted and simplified as needed.

Theorem 3.5 (Fast MRIP Matrices)**.** *There exist absolute constants $c_1, c_2, c_3 \in \mathbb{R}^+$ such that following holds for $N \geq 50$. Let $\epsilon \in (0,1)$, and $p \in \left(e^{-N\left(\lceil \log_2(N/s) \rceil + 1\right)}, \frac{1}{3}\right)$, and $s \leq c_1 \epsilon^2 \sqrt{N} / \ln^5(c_2 N/\epsilon p)$. Furthermore, suppose that $m \in \mathbb{Z}^+$ satisfies*

$$m \geq c_3 s \epsilon^{-2} \ln\left(N/\epsilon \sqrt[s]{p}\right).$$

Then, one may randomly select an $m \times N$ matrix E of the form in (1) such that E will have the MRIP of order (s, ϵ) with probability at least $1 - p$. Furthermore, E will always have an $\mathcal{O}\left(N \cdot \left(\log(s/\epsilon^2) + \log\log\left(N/\epsilon p\right)\right)\right)$-time matrix-vector multiply.

Finally, we may now apply Theorem 2.5 in light of Theorem 3.5 in order to obtain the main result of this section.

Theorem 3.6 (Fast Embedding of Infinite Sets)**.** *There exist absolute constants $c_1', c_2', c_3', c_4' \in \mathbb{R}^+$ such that following holds. Let $S \subset \mathbb{R}^N$ be nonempty for $N \geq 50$, $\epsilon \in (0,1)$, $p \in \left(e^{-c_1' N}, 1/3\right)$, and $D' \in \mathbb{R}^{N \times N}$ be a random diagonal matrix with i.i.d Rademacher random variables on its diagonal. Furthermore, suppose that $w^2(U(S)) \leq c_2' \epsilon^2 \sqrt{N} / \ln^6(c_3' N/\epsilon p)$ and that $m \in \mathbb{Z}^+$ satisfies*

$$m \geq c_4' w^2(U(S)) \frac{\ln\left(N/\epsilon p\right) \ln(1/p)}{\epsilon^2}.$$

Then, one may randomly select an $m \times N$ matrix E of the form in (1) such that ED' will be an ϵ-JL map of S into $\mathbb{R}^m$ with probability at least $1 - p$. Furthermore, ED' will always have an $\mathcal{O}\left(N \cdot \left(\log\left(w(U(S))/\epsilon\right) + \log\log\left(N/\epsilon p\right)\right)\right)$ run-time matrix-vector multiply.

Proof. We will apply Theorem 2.5 with $T = U(S)$ and $p \leftarrow p/2$ noting that $\mathrm{rad}(T) = 1$. Hence, by the union bound it suffices to invoke Theorem 3.5 with $s = 200(1 + \ln(2/p))$, $\epsilon' = \epsilon/(c + c \cdot w(T))$, and $p \leftarrow p/2$. Simplifying using the fact that $c' \leq w(U(S)) \leq w(U(\mathbb{R}^N)) \leq \sqrt{N} + c''$ and combining absolute constants now leads to our final bounds. □

Comparing Theorem 3.6 to Theorem 2.6 one can see that our proposed matrices (1) have matrix-vector multiplies which are always at least as fast as SORS matrices while simultaneously improving on the current best bounds for their embedding dimension, m, by a multiplicative factor of size roughly $\Theta\left(\ln^2\left(\frac{cw^2(U(S))\ln(2/p)}{\epsilon^2}\right)\right)$, with ϵ fixed. Of course, it must also be remembered that Theorem 3.6 applies to a smaller range of Gaussian widths than Theorem 2.6 does.

Remark 3.2. *Note that there is some redundancy in the final form of the embedding matrices constructed by Theorem 3.6. In particular, they look like*

$$ED' = \sqrt{\frac{1}{m_2}} BCD' = \sqrt{\frac{1}{m_2}} B \begin{pmatrix} A & & \\ & \ddots & \\ & & A \end{pmatrix} D' = \sqrt{\frac{m_1}{m_2}} B \begin{pmatrix} RUD & & \\ & \ddots & \\ & & RUD \end{pmatrix} D'$$

where $B \in \mathbb{R}^{m_2 \times N/m_1}$ *has i.i.d. mean* 0 *and variance* 1 *sub-gaussian entries,* $R \in \{0,1\}^{m_1 \times m_1^2}$ *contains* m_1 *rows independently selected uniformly at random from the* $m_1^2 \times m_1^2$ *identity matrix,* $U \in \mathbb{R}^{m_1^2 \times m_1^2}$ *is a unitary matrix with a bounded SOB constant,* $D \in \{0,-1,1\}^{m_1^2 \times m_1^2}$ *is a diagonal matrix with i.i.d.* ± 1 *Rademacher random variables on its diagonal, and* $D' \in \{0,-1,1\}^{N \times N}$ *is a random diagonal matrix with i.i.d. Rademacher random variables on its diagonal. Now one can see, for example, that the same embedding result will hold without having to use the smaller diagonal matrix* D *since* $\begin{pmatrix} D & & \\ & \ddots & \\ & & D \end{pmatrix} D' = D' \begin{pmatrix} D & & \\ & \ddots & \\ & & D \end{pmatrix}$ *and* D' *are identically distributed.*

Of course, before we can apply Theorem 3.6 to, e.g., submanifolds of $\mathbb{R}^N$ we will need covering bounds for their normalized secants. We derive such bounds in the next section.

4 Generalized Covering Bounds for Compact Smooth Submanifolds of $\mathbb{R}^N$ with Respect to Reach

In this section we prove four main theorems. In Theorem 4.2 we give upper bounds for the covering numbers of compact and smooth submanifolds of Euclidean spaces with empty boundary. The method of proof is based on Günther's volume comparison theorem [25, page 169, Theorem 3.101, part ii]. In Theorem 4.3 we use Theorem 4.2 to give upper bounds for the covering numbers of a compact and smooth submanifold with nonempty boundary. We do so by first covering the boundary as an independent manifold. This covers a collar of the boundary, after which we cover the interior. In Theorem 4.4 we utilize our bounds for the covering numbers of submanifolds to bound above the covering numbers of their unit secant sets.
Finally, Theorem 4.4 is applied in Theorem 4.5 to bound the Gaussian widths of the unit secant sets of submanifolds of $\mathbb{R}^N$ with boundary. These Gaussian width bounds can then be employed together with the general embedding results from Sections 2 and 3 to produce our main theorems in Section 1.2.

4.1 Reach and its Basic Properties for Submanifolds of $\mathbb{R}^N$

Here we review the definition and basic properties of the *reach* of a subset of Euclidean space. We specialize to the case when the subset is a compact and smooth submanifold and review the relationship of reach to intrinsic Riemannian geometric features of the submanifold. We include the case when the submanifold has nonempty boundary as is often the case for a manifold modeling real world data.

Reach is an extrinsic parameter of a subset S of Euclidean space defined based on how far away points can lie from S while having a unique closest point in S. Reach has been used extensively as a regularity parameter for S since 1959 when it was defined by Federer in [22]. A historical viewpoint of its development can be found in [45]. Its applications can be found in [1], [10], and [20].
Here, our focus will be on the case when S is a smooth submanifold of Euclidean space. In this case, the inner-product on the ambient $\mathbb{R}^N$ restricts to a Riemannian metric g_S on S. The Riemannian metric g_S equips S with the structure of a geodesic metric space

$$d_S : S \times S \to \mathbb{R}$$

described below. While reach is defined extrinsically, it bounds some intrinsic properties of the metrics g_S such as its sectional curvatures and the injectivity radii of its points. With these bounds in place, we employ Riemannian geometric techniques to obtain lower bounds on the intrinsic volumes of metric balls in S having sufficiently small radii, and in turn, upper bounds on the covering numbers of compact and smooth submanifolds.

We begin by recalling the definition of reach and will then review some of its basic properties.

Definition 4.1. *(**Reach** [22], Definition 4.1) For a subset S of Euclidean space $S \subset \mathbb{R}^N$, the reach τ_S is defined as*

$$\tau_S = \sup\left\{t \geq 0 \;\middle|\; \forall \mathbf{x} \in \mathbb{R}^n \text{ such that } d(\mathbf{x}, S) < t, \mathbf{x} \text{ has a unique closest point in } S\right\}.$$

Open subsets of Euclidean space have zero reach. Closed subsets can also have zero reach. For example, the closed subset $\{(x, |x|) \,|\, x \in \mathbb{R}\}$ of $\mathbb{R}^2$ has zero reach because of the singular point $(0, 0)$. However, sufficiently regular closed subsets have nonzero reach. In particular, compact smooth submanifolds of Euclidean spaces, the subsets under consideration herein, have positive reach [22]. The reach of a closed subset can also be infinite; the closed convex subsets of Euclidean space are precisely the closed subsets having infinite reach. We include a proof of this well-known characterization of convexity as it will be used below.

Lemma 4.1. *A closed subset of $\mathbb{R}^N$ is convex if and only if it has infinite reach.*

Proof. First assume S is a closed and convex subset of $\mathbb{R}^N$. Seeking a contradiction, assume S has finite reach. Then there exists a point $\mathbf{x} \in \mathbb{R}^n$ and distinct points $\mathbf{p}, \mathbf{q} \in S$ such that $d(\mathbf{x}, S) = \|\mathbf{x} - \mathbf{p}\|_2 = \|\mathbf{x} - \mathbf{q}\|_2$. Let $\mathbf{z}$ be the midpoint of the line segment $\mathbf{pq}$. As S is convex, $\mathbf{z} \in S$. The Pythagorean Theorem implies $\|\mathbf{x} - \mathbf{z}\|_2 < \|\mathbf{x} - \mathbf{p}\|_2$, a contradiction.
Now suppose that S is a closed subset of $\mathbb{R}^N$ with infinite reach. As S has infinite reach, the nearest point projection map $P : \mathbb{R}^N \to S$ is well-defined. By [22, Theorem 4.8(8)], P is 1-Lipshitz. Let $\mathbf{p}, \mathbf{q} \in S$ be distinct points, and seeking a contradiction, suppose the line segment $\mathbf{pq}$ does not lie entirely in S. Then since $P(\mathbf{pq})$ is a continuous path joining $\mathbf{p}$ to $\mathbf{q}$ lying entirely in S, it cannot lie entirely in $\mathbf{pq}$. Therefore, there exists a point $\mathbf{z} \in \mathbf{pq}$ with $P(\mathbf{z}) \notin \mathbf{pq}$.
We now have

$$\begin{aligned}\|\mathbf{p} - \mathbf{q}\|_2 &< \|\mathbf{p} - P(\mathbf{z})\|_2 + \|P(\mathbf{z}) - \mathbf{q}\|_2 = \|P(\mathbf{p}) - P(\mathbf{z})\|_2 + \|P(\mathbf{z}) - P(\mathbf{q})\|_2 \\ &\leq \|\mathbf{p} - \mathbf{z}\|_2 + \|\mathbf{z} - \mathbf{q}\|_2 = \|\mathbf{p} - \mathbf{q}\|_2,\end{aligned}$$

a contradiction. Here the last inequality uses that P is 1-Lipshitz. □

We now restrict to the case of compact smooth d-dimensional submanifolds of $\mathbb{R}^N$ ($d \leq N$). Throughout we denote such a manifold by $\mathcal{M}$ to emphasize the manifold setting. We quickly review the definition of these spaces.

By a slight abuse of notation, below we let

$$\mathbb{R}^d = \{\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N \,|\, x_{d+1} = \cdots = x_N = 0\} \quad \text{and} \quad \mathbb{H}^d = \{\mathbf{x} \in \mathbb{R}^d \,|\, x_d \geq 0\}.$$

By definition, $\mathcal{M}$ is a compact subset of $\mathbb{R}^N$ having the property that for each $\mathbf{x} \in \mathcal{M}$ there exists open subsets U and V of $\mathbb{R}^N$ with $\mathbf{x} \in U$ and $\mathbf{0} \in V$ and a smooth diffeomorphism $\phi : U \to V$ with $\phi(\mathbf{x}) = \mathbf{0}$ and such that

1. $\phi(U \cap \mathcal{M}) = V \cap \mathbb{R}^d$, or
2. $\phi(U \cap \mathcal{M}) = V \cap \mathbb{H}^d$.

Precisely one of 1 or 2 holds for each $\mathbf{x} \in \mathcal{M}$. The interior of $\mathcal{M}$, denoted $int\mathcal{M}$ is the union of points for which 1 holds. The boundary of $\mathcal{M}$, denoted $\partial\mathcal{M}$ is the union of points for which 2 holds. The following Lemma is readily deduced from the definition; we omit its standard proof.

Lemma 4.2. *Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$.*

1. *$int\mathcal{M}$ is nonempty.*
2. *$\partial\mathcal{M}$ has finitely many connected components.*
3. *If $\mathcal{C}$ is a nonempty connected component of $\partial\mathcal{M}$, then $\mathcal{C}$ is a compact smooth $(d-1)$-dimensional submanifold of $\mathbb{R}^N$ with $\partial\mathcal{C} = \emptyset$.*

Lemma 4.3. *Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$ with infinite reach. Then*

1. *There exists a d-dimensional affine subspace V of $\mathbb{R}^N$ such that $\mathcal{M} \subset V$.*
2. *The boundary $\partial\mathcal{M}$ is homeomorphic to the $(d-1)$-dimensional sphere.*

Proof. By Lemma 4.1, $\mathcal{M}$ is convex. Let $\mathbf{x} \in int\mathcal{M}$. For each $\mathbf{p} \in \mathcal{M}$ the line segment $\mathbf{xp}$ lies entirely in $\mathcal{M}$ and so also in the tangent space, concluding the proof of 1. The manifold $\mathcal{M}$ is a compact convex subset of the d-dimensional affine space $T_{\mathbf{x}}\mathcal{M}$ with nonempty interior and so has boundary homeomorphic to a $(d-1)$-dimensional sphere, concluding 2. □

We will now briefly review relevant facts from Riemannian geometry used to establish our covering number bounds. Let $\mathcal{M}$ be a connected, compact, and smooth d-dimensional submanifold of $\mathbb{R}^N$. The Euclidean inner-product on $\mathbb{R}^N$ induces a Riemannian metric $g_{\mathcal{M}}$ on $\mathcal{M}$ defined by restricting, for each $\mathbf{x} \in \mathcal{M}$, the Euclidean inner-product to the tangent space $T_{\mathbf{x}}\mathcal{M}$. Each sufficiently regular curve in $\mathcal{M}$ has a well defined $g_{\mathcal{M}}$-length: If $I \subset \mathbb{R}$ is an interval and $\gamma : I \to \mathcal{M}$ is a piecewise C^1-regular curve in $\mathcal{M}$, then γ has $g_{\mathcal{M}}$-length

$$L(\gamma) = \int_I \sqrt{g_{\mathcal{M}}(\gamma'(t), \gamma'(t))}dt.$$

Define

$$d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathbb{R},$$

by setting $d_{\mathcal{M}}(\mathbf{p}, \mathbf{q})$ equal to the infimum of the $g_{\mathcal{M}}$-lengths of piecewise C^1-regular curves joining $\mathbf{p}$ to $\mathbf{q}$ for each $(\mathbf{p}, \mathbf{q}) \in \mathcal{M} \times \mathcal{M}$. It is routine to check that $(\mathcal{M}, d_{\mathcal{M}})$ is a complete metric space majorizing the Euclidean (chordal) metric on $\mathcal{M}$: For all $(\mathbf{p}, \mathbf{q}) \in \mathcal{M} \times \mathcal{M}$,

$$\|\mathbf{p} - \mathbf{q}\|_2 \leq d_{\mathcal{M}}(\mathbf{p}, \mathbf{q}).$$

Given $\mathbf{x} \in \mathcal{M}$ and $r \in (0, \infty)$, let

$$B_{\mathcal{M}}(\mathbf{x}, r) = \{\mathbf{y} \in \mathcal{M} \,|\, d_{\mathcal{M}}(\mathbf{x}, \mathbf{y}) < r\} \quad \text{and} \quad B(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^N \,|\, \|\mathbf{x} - \mathbf{y}\|_2 < r\}$$

and note that

$$B_{\mathcal{M}}(\mathbf{x}, r) \subset \mathcal{M} \cap B(\mathbf{x}, r). \tag{9}$$

When the manifold $\mathcal{M}$ has empty boundary, *geodesics* in $\mathcal{M}$ are classically defined as the smooth curves $\gamma : I \to \mathcal{M}$ with zero internal acceleration: For each $t \in I$, $\gamma''(t)$ is normal to the tangent space $T_{\gamma(t)}\mathcal{M}$. It is standard, albeit nontrivial, to argue that geodesics are equivalently defined as the locally distance minimizing curves in $\mathcal{M}$. This metric-geometry approach is the starting point for defining geodesics in a Riemannian manifold with possibly nonempty boundary.
Herein, a *geodesic* in $\mathcal{M}$ is defined to be a locally distance minimizing constant speed parameterized path: A continuous path $\gamma : I \to \mathcal{M}$ such that there exists $\nu \geq 0$ having the property for each $t \in I$ there exists a subinterval $J \subset I$ with $t \in J$ and such that for each $t_1, t_2 \in J$,

$$d_{\mathcal{M}}(\gamma(t_1), \gamma(t_2)) = \nu |t_1 - t_2|.$$

A geodesic can be reparameterized so that its speed $\nu = 1$. In this case, it is said to be parameterized by arclength. A geodesic $\gamma : I \to \mathcal{M}$ is *minimizing* if the above equality holds for all $t_1, t_2 \in I$.
Geodesics in Riemannian manifolds with boundary are C^1-regular [5] and have one sided second derivative [6]. In particular, since $\mathcal{M}$ is compact, for each $(\mathbf{p}, \mathbf{q}) \in \mathcal{M}$, there exists a C^1-regular minimizing geodesic γ joining $\mathbf{p}$ to $\mathbf{q}$. We define an *interior geodesic* to be a geodesic $\gamma : I \to \mathcal{M}$ with image disjoint from $\partial\mathcal{M}$, or equivalently, with image in $int\mathcal{M}$. An interior geodesics has C^∞-regularity, is characterized by having zero internal acceleration as above, and is uniquely determined by a pair $(\gamma(t), \gamma'(t))$ for any $t \in I$. In contrast, when $\partial\mathcal{M} \neq \emptyset$, geodesics need not be C^∞-regular and may no longer be determined by an initial position and velocity due to possible bifurcations at $\partial\mathcal{M}$ (see Figure 1). Our analysis below avoids these complications by working with interior geodesics. To this end, we adopt the convention that if $\partial\mathcal{M} = \emptyset$, then for each $\mathbf{x} \in \mathcal{M}$, $d_{\mathcal{M}}(\mathbf{x}, \partial\mathcal{M}) = \infty$. With this convention, if $\mathbf{x} \in int\mathcal{M}$, $0 < r < d_{\mathcal{M}}(\mathbf{x}, \partial\mathcal{M})$, and if $B_r \subset T_{\mathbf{x}}\mathcal{M}$ denotes the ball of radius r in $T_{\mathbf{x}}\mathcal{M}$ centered at the origin, then for each $\mathbf{v} \in B_r$ there is a unique interior geodesic

$$\gamma_{\mathbf{v}} : [0, 1] \to int\mathcal{M}$$

with $\gamma'_{\mathbf{v}}(0) = \mathbf{v}$. This gives rise to the exponential map

$$\exp_{\mathbf{x},r} : B_r \to int\mathcal{M},$$

defined by for each $\mathbf{v} \in B_r$, $\exp_{\mathbf{x},r}(\mathbf{v}) = \gamma_{\mathbf{v}}(1)$. Note that

$$\exp_{\mathbf{x},r}(B_r) = B_{\mathcal{M}}(\mathbf{x}, r).$$

The derivative of $\exp_{\mathbf{x},r}$ at the origin in B_r is the identity map of $T_{\mathbf{x}}\mathcal{M}$. It follows from the inverse function theorem that if r is sufficiently small, then $\exp_{\mathbf{x},r}$ is a diffeomorphism onto its image. In Lemma 4.5 below, we estimate, in terms of the reach of $\mathcal{M}$, how large r can be while retaining this property. As a preliminary step, we first consider how large r can be while having the property that $\exp_{\mathbf{x},r}$ is a local diffeomorphism. To do this, we will apply the well known Rauch comparison theorem stated below. This theorem bounds the range of r for which $\exp_{\mathbf{x},r}$ is a local diffeomorphism from below in terms of an upper bound on the *sectional curvatures* of $int\mathcal{M}$.
Intuitively, the sectional curvatures of $\mathcal{M}$ quantify how much $\mathcal{M}$ bends in $\mathbb{R}^N$ along each two-dimensional tangent plane. A precise definition can be found in any Riemannian geometry textbook. Here we are concerned with their relationship to the reach parameter. As an example, let S be a Euclidean sphere in $\mathbb{R}^N$ of radius r. Then S has constant sectional curvatures equal to r^{-2} and

reach $\tau_S = r$. For more general submanifolds, the reach bounds sectional curvatures above as in the following well-known lemma.

Lemma 4.4. *Suppose $\mathcal{M}$ has reach τ.*

- *If $\tau < \infty$, then $int\mathcal{M}$ has sectional curvatures bounded above by τ^{-2}.*
- *If $\tau = \infty$, then $int\mathcal{M}$ has sectional curvatures equal to zero.*

Proof. By Gauss' equation [41, page 100, Theorem 5], it suffices to argue that the norms of the second fundamental forms at points in $int\mathcal{M}$ are bounded by τ^{-1}. See [10, Lemma 4] or [40, Proposition 6.1] for a proof of this bound. □

Having bounded the sectional curvatures above, we may now bound the local diffeomorphism range of r below using the Rauch comparison theorem.

Theorem 4.1 (Rauch comparison)**.** *Let $\mathbf{x} \in \mathcal{M}$ and $0 < r < d_{\mathcal{M}}(\mathbf{x}, \partial\mathcal{M})$, and assume $int\mathcal{M}$ has sectional curvatures bounded above by $K \in [0, \infty)$.*

1. *If $K > 0$,then $\exp_{\mathbf{x},r}$ is a local diffeomorphism provided $r < \frac{\pi}{\sqrt{K}}$.*
2. *If $K = 0$, then $\exp_{\mathbf{x},r}$ is a local diffeomorphism.*

We now apply the local diffeomorphism bound to obtain a bound for the range of r for which $\exp_{\mathbf{x},r}$ is a diffeomorphism.

Lemma 4.5. *Let $\mathcal{M}$ be a smooth and compact d-dimensional submanifold of $\mathbb{R}^N$ with reach $\tau > 0$. Further assume that $\mathbf{x} \in \mathcal{M}$ and $r \in \mathbb{R}$ satisfy*

$$0 < r < \pi\tau \quad \text{and} \quad d_{\mathcal{M}}(\mathbf{x}, \partial\mathcal{M}) > r.$$

Then the exponential map $exp_{\mathbf{x},r} : B_r \to int\mathcal{M}$ is a diffeomorphism onto its image $B_{\mathcal{M}}(\mathbf{x}, r)$.

Proof. Since $r < \pi\tau$, [6, Theorem 3] implies that $\exp_{\mathbf{x},r}$ is one-to-one and so a bijection between B_r and $B_{\mathcal{M}}(\mathbf{x}, r)$. Applying Lemma 4.4 and Theorem 4.1, $\exp_{\mathbf{x},r}$ is a local diffeomorphism. This concludes the proof since a bijective local diffeomorphism is a diffeomorphism. □

Let $\mathcal{H}^d$ denote the d-dimensional Hausdorff measure on $\mathbb{R}^N$. The Riemannian volume of a measurable subset of a compact smooth d-dimensional submanifold $\mathcal{M}$ coincides with its $\mathcal{H}^d$-measure. We adopt the following notional conventions. Given $d \in \mathbb{N}$ and $s > 0$,

- Let $\mathbb{D}^d_s$ denote the closed Euclidean ball in $\mathbb{R}^d$ with center $\mathbf{0}$ and radius s. Furthermore we let $\mathbb{D}^d = \mathbb{D}^d_1$.
- Let $\mathbb{S}^d_s = \partial\mathbb{D}^{d+1}_s$ denote the d-dimensional sphere in $\mathbb{R}^{d+1}$ with center $\mathbf{0}$ and radius s, and let $\mathbb{S}^d = \mathbb{S}^d_1$.
- Let $\omega_d = \mathcal{H}^d(\mathbb{D}^d)$.
- If $0 < r \leq \pi s$, let $V(d, s, r)$ denote the $\mathcal{H}^d$-measure of one (hence any) intrinsic metric open r-ball in $\mathbb{S}^d_s$. Note that πs equals $\text{diam}(\mathbb{S}^d_s)$ with respect to the Riemannian metric on the sphere.

- If $\mathcal{M}$ is a compact smooth d-dimensional submanifold, let $V_{\mathcal{M}} = \mathcal{H}^d(\mathcal{M})$.

Given $\mathbf{x} \in \mathcal{M}$ and r as in Lemma 4.5, we obtain a lower bound on $\mathcal{H}^d(B_{\mathcal{M}}(\mathbf{x}, r))$ as described in the next Proposition.

Proposition 4.1 (Günther's Volume Comparison)**.** *Let $\mathcal{M}$ be a smooth and compact d-dimensional submanifold of $\mathbb{R}^N$ with reach $\tau > 0$ and $d \geq 2$.*

1. *If $\mathbf{x} \in int\mathcal{M}$ and $r \in \mathbb{R}$ satisfy $0 < r < \pi\tau$ and $d_{\mathcal{M}}(\mathbf{x}, \partial\mathcal{M}) > r$, then*

$$\mathcal{H}^d(B_{\mathcal{M}}(\mathbf{x}, r)) \geq V(d, \tau, r).$$

2. *If $r < \sqrt{6}\tau$, then*

$$V(d, \tau, r) \geq \omega_d \left(1 - \frac{r^2}{6\tau^2}\right)^{d-1} r^d.$$

Proof. If the reach of $\mathcal{M}$ is infinite, then by Lemma 4.3, $\mathcal{M}$ is a convex subset of a d-dimensional affine space. Since $\mathbf{x}$ is at least r away from the boundary, we have $\mathcal{H}^d(B_{\mathcal{M}}(\mathbf{x}, r)) = \omega_d r^d$, concluding the proof in this case. In the remainder of the proof, we assume τ is finite.
Let $\mathbf{x} \in int\mathcal{M}$ and $r > 0$ be as in the statement of 1. By Lemma 4.5, $\exp_{\mathbf{x},r}$ is a diffeomorphism onto its image, the geodesic ball $B_{\mathcal{M}}(\mathbf{x}, r)$. By Lemma 4.4 and Günther's volume comparison theorem [25, page 169, Theorem 3.101, part ii], $\mathcal{H}^d(B_{\mathcal{M}}(\mathbf{x}, r))$ is bounded below by the volume of a metric r-ball in the sphere $\mathbb{S}^d_\tau$ having constant sectional curvatures τ^{-2}, concluding the proof of 1.
To prove 2, let $f(x) = \dfrac{\sin(x)}{x}$. We use a formula derived from [38]:

$$V(d, \tau, r) = d\omega_d \left(\int_0^r \left(x f\left(\frac{x}{\tau}\right)\right) dx\right)^{d-1}.$$

As $f(x)$ is positive and decreasing on $(0, \pi)$ and $r < \sqrt{6}\tau < \pi\tau$, it follows

$$V(d, \tau, r) \geq d\omega_d \left(\int_0^r \left(x f\left(\frac{r}{\tau}\right)\right) dx\right)^{d-1} = \omega_d f\left(\frac{r}{\tau}\right)^{d-1} r^d.$$

Now using $0 < 1 - \dfrac{x^2}{6} < f(x)$ on $(0, \sqrt{6})$, we obtain

$$V(d, \tau, r) \geq \omega_d \left(1 - \frac{r^2}{6\tau^2}\right)^{d-1} r^d,$$

concluding the proof. □

Given a compact and smooth d-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^N$ with reach $\tau_{\mathcal{M}}$, the ratio

$$\frac{V_{\mathcal{M}}}{\tau_{\mathcal{M}}^d}$$

is invariant under a rescaling of the ambient $\mathbb{R}^N$. The preceding Proposition 4.1 applies to show this ratio is uniformly bounded below for compact smooth d-dimensional submanifolds with $\partial\mathcal{M} = \emptyset$ as in the next proposition.

Proposition 4.2. *Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$ with $d \geq 1$. If $\partial\mathcal{M} = \emptyset$, then*

$$\frac{V_{\mathcal{M}}}{\tau_{\mathcal{M}}^d} \geq \mathcal{H}^d(\mathbb{S}^d).$$

Proof. By Lemma 4.3, the reach $\tau_{\mathcal{M}}$ is finite. Since $\frac{V_{\mathcal{M}}}{\tau_{\mathcal{M}}^d}$ is scale invariant, it is sufficient to consider the case where $\tau_{\mathcal{M}} = 1$. The case of $d = 1$ is classical [11, 23]. When $d \geq 2$, let $\mathbf{x} \in \mathcal{M}$ and apply Proposition 4.1 to deduce

$$V_{\mathcal{M}} \geq \mathcal{H}^d(B_{\mathcal{M}}(\mathbf{x}, \pi)) \geq V(d, 1, \pi) = \mathcal{H}^d(\mathbb{S}^d).$$

□

With the volume comparison Proposition 4.1 in place, we are now prepared to prove covering number bounds for compact smooth submanifolds of $\mathbb{R}^N$ in terms of reach.

4.2 Upper Bounds for the Covering Numbers of Compact Smooth Submanifolds of $\mathbb{R}^N$

We begin by reviewing the related covering and packing numbers of a subset of a metric space.

Definition 4.2. *Let (X, d) be a metric space, E a subset of X, and $r > 0$.*

1. *The packing number $N_r^{pack}(E)$ is the largest number of points $x_1, \ldots, x_n \in E$ such that the metric balls $B(x_1, r), \ldots, B(x_n, r)$ are pairwise disjoint.*
2. *The covering number $N_r^{cover}(E)$ is the fewest number of points $x_1, \ldots, x_n \in E$ such that E lies in the union of the metric balls $\overline{B(x_1, r)}, \ldots, \overline{B(x_n, r)}$.*

This section presents upper bounds for the covering numbers of compact smooth submanifolds of Euclidean spaces. The method employed is to give upper bounds for the packing numbers of these submanifolds and to apply the following well known lemma [46, lemma 4.2.8].

Lemma 4.6. *For each $r > 0$, $N_r^{cover}(E) \leq N_{r/2}^{pack}(E)$.*

Theorem 4.2 (Covering a Compact Smooth Submanifold with Empty Boundary)**.** *Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$ with $\partial\mathcal{M} = \emptyset$. Let $\tau_{\mathcal{M}}, V_{\mathcal{M}} \in (0, \infty)$ denote the reach and volume of $\mathcal{M}$, respectively.*

1. *If $d = 0$ and $\epsilon > 0$, then $N_\epsilon^{cover}(\mathcal{M}) \leq V_{\mathcal{M}}$.*
2. *If $d > 0$ and if $0 < \epsilon < 2\sqrt{6}\tau_{\mathcal{M}}$, then $N_\epsilon^{cover}(\mathcal{M}) \leq \dfrac{V_{\mathcal{M}}}{\omega_d \left(1 - \frac{\epsilon^2}{24\tau_{\mathcal{M}}^2}\right)^{d-1} \left(\frac{\epsilon}{2}\right)^d}$.*

Proof. First assume $d = 0$. Then $V_{\mathcal{M}} \in \mathbb{N}$ and there exists a set of $V_{\mathcal{M}}$ points

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_{V_{\mathcal{M}}}\} \subset \mathbb{R}^N$$

such that $\mathcal{M} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{V_\mathcal{M}}\}$. The Euclidean balls $B(\mathbf{x}_i, \epsilon)$ with $i = 1, \ldots, V_\mathcal{M}$ cover $\mathcal{M}$, concluding the proof in this case. Next assume $d > 0$. By Lemma 4.6, it suffices to establish the inequality

$$N_{\epsilon/2}^{pack}(\mathcal{M}) \leq \frac{V_\mathcal{M}}{\omega_d \left(1 - \frac{\epsilon^2}{24\tau_\mathcal{M}^2}\right)^{d-1} \left(\frac{\epsilon}{2}\right)^d}.$$

Recall from (9) that for each $\mathbf{x} \in \mathcal{M}$ and $r > 0$,

$$B_\mathcal{M}(\mathbf{x}, r) \subset B(\mathbf{x}, r) \cap \mathcal{M}.$$

Conclude that if $\{\mathbf{x}_1, \ldots \mathbf{x}_p\}$ are p-points in $\mathcal{M}$ such that the Euclidean balls $B(\mathbf{x}_i, r)$ are pairwise disjoint, then the intrinsic metric balls $B_\mathcal{M}(\mathbf{x}_i, r)$ are also pairwise disjoint. In this case, since $\mathcal{H}^d$ is additive,

$$V_\mathcal{M} \geq p \min\{\mathcal{H}^d(B_\mathcal{M}(\mathbf{x}_i, r)) \,|\, i = 1, \ldots, p\}.$$

With this in mind, applying Proposition 4.1 with $r = \frac{\epsilon}{2}$ yields the desired conclusion. □

We will now apply our theorem to $\mathbb{S}^d$ to judge its tightness. A standard estimate for covering $\mathbb{S}^d$ with balls of radius ϵ centered on the sphere is $\left(\frac{3}{\epsilon}\right)^d$, see [46, corollary 4.2.13]. Theorem 4.2 yields an upper bound of comparable quality.

Corollary 4.1. *For $0 < \epsilon < 1$, $\mathbb{S}^d \subset \mathbb{R}^N$ can be covered with at most $(3.4\sqrt{d})\frac{2.1^d}{\epsilon^d}$ Euclidean N-dimensional balls of radius ϵ centered in $\mathbb{S}^d$.*

Proof. By Theorem 4.2, we need at most $\frac{V}{\omega_d \left(1 - \frac{\epsilon^2}{24\tau^2}\right)^{d-1} \left(\frac{\epsilon}{2}\right)^d}$ balls. We have $\tau = 1$, $V = \frac{2\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}$, $\omega_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$, and $\frac{V}{\omega_d} = 2\sqrt{\pi}\frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d+1}{2})} < 2\sqrt{\pi d}$. This leads to an upper bound of $2\sqrt{\pi d}\left(\frac{24}{23}\right)^{d-1}\frac{2^d}{\epsilon^d} \leq 3.4\sqrt{d}\left(\frac{2.1}{\epsilon}\right)^d$. □

While there are tighter bounds on the order of $\mathcal{O}\left(\frac{d^{1.5}\ln(d)}{\epsilon^d}\right)$ ([12, Theorem 6.8.1]), they only apply to $\mathbb{S}^d$, whereas Theorem 4.2 has the advantage of applying to a general submanifold.
Using Theorem 4.2 and Lemma 4.2 we now present a covering estimate for a manifold with nonempty boundary. We first introduce some notation. Given a compact smooth d-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^N$ with $\partial M \neq \emptyset$,

- Let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ $(k \geq 1)$ denote the nonempty connected components of $\partial\mathcal{M}$.
- Let τ_M denote the reach of $\mathcal{M}$.
- For each $i \in \{1, \ldots, k\}$ let $\tau_{\mathcal{C}_i}$ denote the reach of $\mathcal{C}_i$.
- Let $\mu_{\partial\mathcal{M}} = \min\{\tau_{\mathcal{C}_i} \mid i \in \{1, \ldots, k\}\}$

- Let $V_{\mathcal{M}} = \mathcal{H}^d(\mathcal{M})$ and $V_{\partial\mathcal{M}} = \mathcal{H}^{d-1}(\partial\mathcal{M})$.
- For each $i \in \{1, \ldots, k\}$, let $V_{\mathcal{C}_i} = \mathcal{H}^{d-1}(\mathcal{C}_i)$.

Note that

$$V_{\partial\mathcal{M}} = \sum_{i=1}^{k} V_{\mathcal{C}_i}.$$

Theorem 4.3 (Covering a Compact Smooth Submanifold with Nonempty Boundary)**.** *Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$ with $d \geq 1$ and $\partial\mathcal{M} \neq \emptyset$. Further assume $\epsilon \in (0, \min\{4\sqrt{6}\mu_{\partial\mathcal{M}}, 2\sqrt{6}\tau_{\mathcal{M}}\}]$.*

1. *If $d = 1$ then*

$$N_\epsilon^{cover}(\mathcal{M}) \leq \frac{V_{\mathcal{M}}}{\epsilon} + V_{\partial\mathcal{M}}.$$

2. *If $d \geq 2$, then*

$$N_\epsilon^{cover}(\mathcal{M}) \leq \frac{V_{\mathcal{M}}}{\omega_d \left(1 - \frac{\epsilon^2}{24\tau_{\mathcal{M}}^2}\right)^{d-1} \left(\frac{\epsilon}{2}\right)^d} + \frac{V_{\partial\mathcal{M}}}{\omega_{d-1} \left(1 - \frac{\epsilon^2}{96\mu_{\partial\mathcal{M}}^2}\right)^{d-2} \left(\frac{\epsilon}{4}\right)^{d-1}}.$$

Proof. First consider the case when $d = 1$. Each connected component of $\mathcal{M}$ is either an embedded circle or an embedded closed interval. Let $\mathcal{D}$ denote a connected component and let $V_{\mathcal{D}}$ denote its length. We claim

1. If $\partial\mathcal{D} = \emptyset$, then $N_\epsilon^{cover}(\mathcal{D}) \leq \frac{V_{\mathcal{D}}}{\epsilon}$, and
2. If $\partial\mathcal{D} \neq \emptyset$, then $N_\epsilon^{cover}(\mathcal{D}) \leq \frac{V_{\mathcal{D}}}{\epsilon} + 2$.

Assuming the claim, the desired upper bound follows from summing the above upper bounds over the connected components of $\mathcal{M}$, noting those components $\mathcal{D}$ with $\partial\mathcal{D} \neq \emptyset$ have $V_{\partial\mathcal{D}} = 2$ and those with $\partial\mathcal{D} = \emptyset$ have $V_{\partial\mathcal{D}} = 0$.
We now establish the claim. We apply Lemma 4.6, and instead give an upper bound for $N_{\frac{\epsilon}{2}}^{pack}(\mathcal{D})$. In case 1, by (9), each Euclidean ball of radius $\epsilon/2$ centered in $\mathcal{D}$ contains a geodesic ball of the same radius. The length of this ball is $2(\epsilon/2)$. Therefore, $N_{\frac{\epsilon}{2}}^{pack}(\mathcal{D}) \leq \frac{V_{\mathcal{D}}}{\epsilon}$, where we have used that $\frac{V_{\mathcal{D}}}{\epsilon} \geq \frac{V_{\mathcal{D}}}{2\sqrt{6}\tau_{\mathcal{M}}} \geq \frac{V_{\mathcal{D}}}{2\sqrt{6}\tau_{\mathcal{D}}} \geq \frac{2\pi}{2\sqrt{6}} > 1$ by Proposition 4.2. By the same reasoning, in case 2, all but at most two of the Euclidean balls in an $\epsilon/2$ packing of $\mathcal{D}$ will meet $\mathcal{D}$ in a geodesic interval of length at least ϵ. The two potentially exceptional balls are those centered at points nearest to the two boundary points. Now 2 follows, concluding the proof of the theorem when $d = 1$.

Now assume $d \geq 2$. We will cover the following two subsets of $\mathcal{M}$ separately:

$$S_1 = \left\{\mathbf{x} \in \mathcal{M} \;\middle|\; d_{\mathcal{M}}(\mathbf{x}, \partial\mathcal{M}) < \frac{\epsilon}{2}\right\} \quad \text{and} \quad S_2 = \mathcal{M} \backslash S_1.$$

We begin by obtaining a covering number bound for S_1. We first claim

$$N_\epsilon^{cover}(S_1) \leq N_{\frac{\epsilon}{2}}^{cover}(\partial\mathcal{M}). \tag{10}$$

Indeed, assume that $\partial\mathcal{M}$ has been covered by a finite set of Euclidean $\frac{\epsilon}{2}$-balls and let $C \subset \partial\mathcal{M}$ denote the set of centers of these balls. Given $\mathbf{x} \in S_1$, there exists $\mathbf{y} \in \partial\mathcal{M}$ and $\mathbf{c} \in C$ such that

$$d_{\mathcal{M}}(\mathbf{x}, \mathbf{y}) < \frac{\epsilon}{2} \quad \text{and} \quad \|\mathbf{y} - \mathbf{c}\|_2 < \epsilon/2.$$

Then,

$$\|\mathbf{x} - \mathbf{c}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2 + \|\mathbf{y} - \mathbf{c}\|_2 \leq d_{\mathcal{M}}(\mathbf{x}, \mathbf{y}) + \epsilon/2 < \epsilon,$$

demonstrating that C is the central set for an ϵ covering of S_1 and establishing (10).
By Lemma 4.2, the boundary $\partial\mathcal{M}$ is a compact smooth $(d-1)$-dimensional submanifold with empty boundary and with finitely many connected components $\mathcal{C}_1, \ldots, \mathcal{C}_k$. As $\epsilon \leq 4\sqrt{6}\mu_{\partial\mathcal{M}}$ and $\mu_{\partial\mathcal{M}} = \min\{\tau_{\mathcal{C}_i} \,|\, i \in \{1, \ldots, k\}\}$, we may apply Theorem 4.2 to each component $\mathcal{C}_i$ to deduce

$$N_{\frac{\epsilon}{2}}^{cover}(\mathcal{C}_i) \leq \frac{V_{\mathcal{C}_i}}{\omega_{d-1}(1 - \frac{\epsilon^2}{96\tau_{\mathcal{C}_i}^2})^{d-2}(\frac{\epsilon}{4})^{d-1}} \leq \frac{V_{\mathcal{C}_i}}{\omega_{d-1}(1 - \frac{\epsilon^2}{96\mu_{\partial\mathcal{M}}^2})^{d-2}(\frac{\epsilon}{4})^{d-1}}. \tag{11}$$

Combining (10), (11), and the obvious inequality $N_{\frac{\epsilon}{2}}^{cover}(\partial\mathcal{M}) \leq \sum_{i=1}^{k} N_{\frac{\epsilon}{2}}^{cover}(\mathcal{C}_i)$ we have

$$N_\epsilon^{cover}(S_1) \leq \sum_{i=1}^{k} N_{\frac{\epsilon}{2}}^{cover}(\mathcal{C}_i) \leq \frac{\sum_{i=1}^{k} V_{\mathcal{C}_i}}{\omega_{d-1}(1 - \frac{\epsilon^2}{96\mu_{\partial\mathcal{M}}^2})^{d-2}(\frac{\epsilon}{4})^{d-1}} = \frac{V_{\partial\mathcal{M}}}{\omega_{d-1}(1 - \frac{\epsilon^2}{96\mu_{\partial\mathcal{M}}^2})^{d-2}(\frac{\epsilon}{4})^{d-1}}. \tag{12}$$

We next obtain a covering bound for S_2 using the method employed in Theorem 4.3. By Lemma 4.6,

$$N_\epsilon^{cover}(S_2) \leq N_{\frac{\epsilon}{2}}^{pack}(S_2).$$

If $C \subset S_2$ is the set of centers of a packing by Euclidean $\frac{\epsilon}{2}$-balls, then the $d_{\mathcal{M}}$ metric $\frac{\epsilon}{2}$-balls are pairwise disjoint in $\mathcal{M}$. As $\epsilon < 2\sqrt{6}\tau_{\mathcal{M}}$, Proposition 4.1 applies with $r = \frac{\epsilon}{2}$ to show each such ball has $\mathcal{H}^d$-measure at least $\omega_d \left(1 - \frac{\epsilon^2}{24\tau_{\mathcal{M}}}\right)^{d-1} \left(\frac{\epsilon}{2}\right)^d$. It follows that

$$N_\epsilon^{cover}(S_2) \leq \frac{V_{\mathcal{M}}}{\omega_d(1 - \frac{\epsilon^2}{24\tau_{\mathcal{M}}})^{d-1}(\frac{\epsilon}{2})^d}. \tag{13}$$

The claimed upper bound for $N_\epsilon^{cover}(\mathcal{M})$ now follows from (12), (13), and the obvious inequality $N_\epsilon^{cover}(\mathcal{M}) \leq N_\epsilon^{cover}(S_1) + N_\epsilon^{cover}(S_1)$, concluding the proof of the theorem. □

To illustrate Theorem 4.3 we will now apply our estimate to the standard closed d-dimensional unit ball $\mathbb{D}^d$ (e.g., the closed unit disk for $d = 2$) as a manifold with boundary.

Corollary 4.2. *Consider $\mathbb{D}^d \subset \mathbb{R}^N$ with $d \leq N$ and $N \geq 2$. If $\epsilon \in (0,1)$, then*

$$N_\epsilon^{cover}(\mathbb{D}^d) \leq \left(\frac{2}{\epsilon}\right)^d + 2\pi\left(\frac{4.05}{\epsilon}\right)^{d-1}.$$

Proof. From Theorem 4.3, we need at most $\dfrac{V_{\mathbb{D}^d}}{\omega_d\left(1-\frac{\epsilon^2}{24\tau^2_{\mathbb{D}^d}}\right)^{d-1}\left(\frac{\epsilon}{2}\right)^d} + \dfrac{V_{\mathbb{S}^{d-1}}}{\omega_{d-1}\left(1-\frac{\epsilon^2}{96\tau^2_{\mathbb{S}^{d-1}}}\right)^{d-2}\left(\frac{\epsilon}{4}\right)^{d-1}}$ balls. The relevant parameters are

$\mathcal{M}$	$V_{\mathbb{D}^d}$	$\tau_{\mathbb{D}^d}$	$\partial\mathbb{D}^d$	$V_{\mathbb{S}^{d-1}}$	$\tau_{\mathbb{S}^{d-1}}$	ω_d	ω_{d-1}
$\mathbb{D}^d$	$\dfrac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$	∞	$\mathbb{S}^{d-1}$	$2\dfrac{\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}$	1	$V_{\mathbb{D}^d}$	$V_{\mathbb{D}^{d-1}}$

Using $0 < \epsilon < 1$ we get, $\left(\frac{2}{\epsilon}\right)^d + 2\pi\frac{95}{96}\left(\frac{96}{95}\frac{4}{\epsilon}\right)^{d-1} < \left(\frac{2}{\epsilon}\right)^d + 2\pi\left(\frac{4.05}{\epsilon}\right)^{d-1}$. □

4.3 Covering Estimate for the Unit Secants of a Submanifold from Above

Recall from Section 2, the unit rescaling map

$$U : \mathbb{R}^N \setminus \{0\} \to \mathbb{S}^{N-1}$$

defined by $U(\mathbf{v}) = \dfrac{\mathbf{v}}{\|\mathbf{v}\|_2}$ and that for a subset S of $\mathbb{R}^N$,

$$S - S = \left\{\mathbf{p} - \mathbf{q} \;\middle|\; \mathbf{p} \neq \mathbf{q},\ \mathbf{p}, \mathbf{q} \in S\right\}.$$

Elements in $S - S$ are the *secants* generated by S and elements in $U(S-S)$ are the *unit secants* generated by S. In this section we provide an upper bound for the covering number of the closure of the unit secant set generated by a compact smooth d-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^N$. Such an object, denoted herein by

$$\overline{U(\mathcal{M}-\mathcal{M})},$$

has been studied previously in [37, Section 3], [43, Page 1323], [47, Section 1], and [48, Section 3]. We first consider two special simple cases.

Proposition 4.3. *Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$. Let V_M denote the volume of $\mathcal{M}$ and $\tau_{\mathcal{M}}$ denote its reach. Let $0 < \epsilon < 1$.*

1. *If $d = 0$, then*
$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})}) \leq V_{\mathcal{M}}^2.$$
2. *If $d \geq 1$ and $\tau_{\mathcal{M}} = \infty$, then*
$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})}) \leq \left(1+\frac{2}{\epsilon}\right)^d.$$

Proof. 1. As $d = 0$, $\mathcal{M}$ consists of $V_{\mathcal{M}}$ points in $\mathbb{R}^N$. From its definition, $\overline{U(\mathcal{M}-\mathcal{M})}$ has cardinality at most the cardinality of $\mathcal{M}-\mathcal{M}$. The latter is bounded above by the cardinality of $\mathcal{M}\times\mathcal{M}$.

2. By Lemma 4.3, there is a d-dimensional affine subspace V of $\mathbb{R}^N$ such that $\mathcal{M}$ is a compact smooth convex body in V. It follows that $\overline{U(\mathcal{M}-\mathcal{M})}$ is congruent to the $(d-1)$-sphere $\mathbb{S}^{d-1} \subset \mathbb{S}^{N-1}$. This sphere has the standard covering bound $\left(1+\frac{2}{\epsilon}\right)^d$ [46, Corollary 4.2.13].

□

We now move to the general case of a compact smooth d-dimensional submanifold $\mathcal{M}$, with $d \geq 1$ and $\tau_{\mathcal{M}} < \infty$. We allow the possibility that $\partial\mathcal{M} \neq \emptyset$ and adopt the notation preceding Theorem 4.3. Further, we let

$$\tau = \min\{\tau_M, \mu_{\partial\mathcal{M}}\},$$

where we set $\mu_{\partial\mathcal{M}} = \infty$ when $\partial\mathcal{M} = \emptyset$.
Given a sufficiently small number $\epsilon > 0$, we will estimate the covering number $N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})})$ following related arguments for manifolds without boundary presented in [15], [20], and [36]. The strategy is to separate the secants $\mathcal{M}-\mathcal{M}$ into long and short secants and to cover their images in $U(\mathcal{M}-\mathcal{M})$ separately. Before proceeding, we record three lemmas that will be useful in the course of the proof. For long secants, we will use the following lemma.

Lemma 4.7. *[15, Lemma 4.1] Let $\mathbf{p}, \mathbf{p}^*, \mathbf{q}$ and $\mathbf{q}^*$ be 4 points in $\mathbb{R}^N$. Let $0 < l := \|\mathbf{p}-\mathbf{q}\|_2$ and $\|\mathbf{p}-\mathbf{p}^*\|_2, \|\mathbf{q}-\mathbf{q}^*\|_2 < d \in \mathbb{R}^+$. Let $0 < \epsilon < 1$ and assume $\frac{4d}{l} \leq \epsilon$. Then,*

$$\|U(\mathbf{p}-\mathbf{q}) - U(\mathbf{p}^*-\mathbf{q}^*)\|_2 \leq \epsilon.$$

For short secants, we will use the following two lemmas.

Lemma 4.8. *Assume $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}$ satisfy $0 < \|\boldsymbol{p}-\boldsymbol{q}\| \leq \frac{\tau}{2}$. Given a unit tangent vector $\mathbf{w} \in T_{\boldsymbol{p}}\mathcal{M}$, let $\mathbf{w}^* \in T_{\boldsymbol{q}}\mathcal{M}$ be a unit tangent vector obtained by parallel translating $\mathbf{w}$ along a minimizing geodesic joining $\boldsymbol{p}$ to $\boldsymbol{q}$. Then*

1. *$d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{q}) \leq \tau$,*
2. *$d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{q}) \leq \|\boldsymbol{p}-\boldsymbol{q}\|_2 \left(1 + \frac{2\|\boldsymbol{p}-\boldsymbol{q}\|_2}{\tau}\right)$, and*
3. *if θ is the angle between $\mathbf{w}$ and $\mathbf{w}^*$, then $\theta \leq \frac{d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{q})}{\tau}$.*

Proof. By [40, Lemma 6.3] and [20, Lemma 7],

$$d_{\mathcal{M}}(\mathbf{p}, \mathbf{q}) \leq \tau - \tau\sqrt{1 - \frac{2\|\mathbf{p}-\mathbf{q}\|_2}{\tau}}, \tag{14}$$

implying the first inequality in the Lemma. Note that for each $x \in [0, 1]$,

$$1 - \sqrt{1-x} \leq \frac{x + x^2}{2}. \tag{15}$$

Using this with $x = \dfrac{2\|\mathbf{p}-\mathbf{q}\|_2}{\tau}$ in (14) implies the second inequality in the Lemma. See [10, Lemma 6] for the third inequality in the lemma. □

Lemma 4.9. *Assume $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}$ satisfy $0 < \|\boldsymbol{p}-\boldsymbol{q}\|_2 \leq \dfrac{\tau}{2}$. Let $\boldsymbol{u} \in T_{\boldsymbol{p}}\mathcal{M}$ be the initial unit length velocity vector of a minimizing unit speed geodesic in $\mathcal{M}$ joining $\boldsymbol{p}$ to $\boldsymbol{q}$. Let ϕ denote the angle between $\mathbf{u}$ and $U(\boldsymbol{q}-\boldsymbol{p})$. Then*

$$\sin(\phi) \leq \frac{\|\boldsymbol{p}-\boldsymbol{q}\|_2}{2\tau}\left(1+\frac{2\|\boldsymbol{p}-\boldsymbol{q}\|_2}{\tau}\right)^2.$$

Proof. See Figure 4. Let h be the distance between $\mathbf{q}$ and the line through $\mathbf{p}$ with direction $\mathbf{u}$, and let $d = \|\mathbf{p}-\mathbf{q}\|_2$. We claim

$$h \leq \frac{d^2}{2\tau}\left(1+\frac{2d}{\tau}\right)^2. \tag{16}$$

Assuming (16), we have

$$\sin(\phi) = \frac{h}{d} \leq \frac{\frac{d^2}{2\tau}\left(1+\frac{2d}{\tau}\right)^2}{d} = \frac{\|\mathbf{p}-\mathbf{q}\|_2}{2\tau}\left(1+\frac{2\|\mathbf{p}-\mathbf{q}\|_2}{\tau}\right)^2$$

as stated in the lemma. We conclude by establishing (16).

To this end, let $\gamma : [0, d_{\mathcal{M}}(\mathbf{p},\mathbf{q})] \to \mathcal{M}$ be a unit speed minimizing geodesic joining $\mathbf{p}$ to $\mathbf{q}$. For each $s \in [0, d_{\mathcal{M}}(\mathbf{p},\mathbf{q})]$, let $V_s = \dot{\gamma}(s)$ and let $\theta(s)$ denote the angle between $\mathbf{u} = V_0$ and V_s. By Lemma 4.8,

$$d_{\mathcal{M}}(\mathbf{p},\mathbf{q}) \leq \tau \tag{17}$$

and

$$\theta(s) \leq \frac{s}{\tau}. \tag{18}$$

By (17) and (18), for each $s \in [0, d_{\mathcal{M}(\mathbf{p},\mathbf{q})}]$,

$$\theta(s) \leq \frac{s}{\tau} \leq \frac{d_{\mathcal{M}}(\mathbf{p},\mathbf{q})}{\tau} \leq 1 < \frac{\pi}{2}.$$

As $\sin(t)$ is increasing on $\left[0, \dfrac{\pi}{2}\right]$,

$$\sin(\theta(s)) \leq \sin\left(\frac{s}{\tau}\right).$$

The incremental gain of $\gamma(s)$ in the direction h is at most $\sin(\theta(s))$. Therefore

$$h \leq \int_0^{d_{\mathcal{M}}(\mathbf{p},\mathbf{q})} \sin(\theta(s))\, ds \leq \int_0^{d_{\mathcal{M}}(\mathbf{p},\mathbf{q})} \sin\left(\frac{s}{\tau}\right)\, ds = \tau\left(1-\cos\left(\frac{d_{\mathcal{M}}(\mathbf{p},\mathbf{q})}{\tau}\right)\right). \tag{19}$$

Use (19) and the fact that

$$1-\cos(t) < \frac{t^2}{2}$$

for nonzero t to conclude

$$h \leq \tau \left(\frac{\left(\frac{d_{\mathcal{M}}(\mathbf{p},\mathbf{q})}{\tau}\right)^2}{2} \right) = \frac{d_{\mathcal{M}}(\mathbf{p},\mathbf{q})^2}{2\tau}. \tag{20}$$

Combining (20) with inequality 2 in Lemma 4.8 establishes (16), completing the proof.

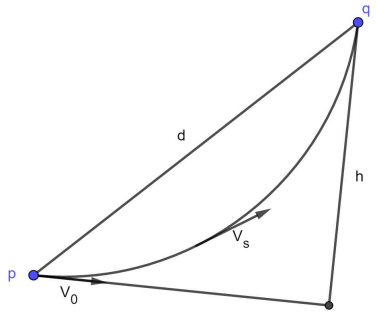

Figure 4: Two points with the geodesic and secant line between them. We bound the angle between $U(\mathbf{q}-\mathbf{p})$ and $V_0 = \mathbf{u}$ using τ.

□

Theorem 4.4 (Covering the Unit Secants)**.** *Let $d \geq 1$ and let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$ with $\tau_M < \infty$.*[6] *Let $\epsilon \in (0,1)$.*

1. *If $d = 1$, define $\alpha := \dfrac{20V_{\mathcal{M}}}{\tau} + V_{\partial\mathcal{M}}$. Then*

$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})}) \leq \left(\alpha^2 + 2\alpha\right)\frac{1}{\epsilon^4}.$$

2. *If $d \geq 2$, define $\alpha := \dfrac{V_{\mathcal{M}}}{\omega_d}\left(\dfrac{41}{\tau}\right)^d + \dfrac{V_{\partial\mathcal{M}}}{\omega_{d-1}}\left(\dfrac{81}{\tau}\right)^{d-1}$. Then*

$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})}) \leq \left(\alpha^2 + 3^d\alpha\right)\frac{1}{\epsilon^{4d}}.$$

Proof. Note that $\overline{U(\mathcal{M}-\mathcal{M})} \subset \mathbb{R}^N$ includes all the unit tangent vectors to $\mathcal{M}$. Fix a $\left(\dfrac{\tau\epsilon^2}{20}\right)$-net C in $\mathcal{M}$ having least cardinality and for each $\mathbf{c} \in C$, let $V_{\mathbf{c}}$ be an $\dfrac{\epsilon}{3}$-net in the unit tangent sphere $S_{\mathbf{c}}\mathcal{M}$ having least cardinality. Let

$$V = \bigcup_{\mathbf{c}\in C} V_{\mathbf{c}} \quad \text{and} \quad D = U(C-C) \cup V.$$

[6] For infinite reach, see Proposition 4.3.

To prove the Theorem, we will prove that D is an ϵ-net for $\overline{U(\mathcal{M}-\mathcal{M})}$ and then give an upper bound for its cardinality $|D|$.

Proving D is an ϵ-net for $\overline{U(\mathcal{M}-\mathcal{M})}$.

To this end, subdivide $\mathcal{M}-\mathcal{M}$ into disjoint subsets consisting of *long* and *short* secants:

$$S_1 = \left\{ \mathbf{v} \in \mathcal{M}-\mathcal{M} \,\middle|\, \|\mathbf{v}\|_2 > \frac{\tau\epsilon}{5} \right\} \quad \text{and} \quad S_2 = \left\{ \mathbf{v} \in \mathcal{M}-\mathcal{M} \,\middle|\, \|\mathbf{v}\|_2 \leq \frac{\tau\epsilon}{5} \right\}.$$

Note that the elements of $\overline{U(\mathcal{M}-\mathcal{M})}$ that do not lie in $U(\mathcal{M}-\mathcal{M})$ are the unit tangent vectors to $\mathcal{M}$ obtained as limits of elements in $U(S_2)$. Hence

$$\overline{U(\mathcal{M}-\mathcal{M})} = U(S_1) \cup \overline{U(S_2)}.$$

We argue in two steps, first showing that $U(C-C)$ is an ϵ-net for the unit-rescaled long secants $U(S_1)$, and then showing that V is an ϵ-net for the closure of the unit-rescaled short secants $\overline{U(S_2)}$.

Step 1: Proving $U(C-C)$ is an ϵ-net for $U(S_1)$.

Given $\mathbf{v} = \mathbf{p}-\mathbf{q} \in S_1$, let $\mathbf{p}^*$ and $\mathbf{q}^*$ be closest points to them in C. The triangle inequality implies

$$\|\mathbf{p}^*-\mathbf{q}^*\| \geq \|\mathbf{p}-\mathbf{q}\| - \|\mathbf{p}-\mathbf{p}^*\| - \|\mathbf{q}-\mathbf{q}^*\| \geq \frac{\tau\epsilon}{5} - 2\frac{\tau\epsilon^2}{20} > 0.$$

In particular, $\mathbf{p}^* \neq \mathbf{q}^*$, and so

$$\mathbf{v}^* := \mathbf{p}^*-\mathbf{q}^* \in C-C.$$

Applying Lemma 4.7 with $l = \|\mathbf{p}-\mathbf{q}\|_2$ and $d = \dfrac{\tau\epsilon^2}{20}$,

$$\|U(\mathbf{v}) - U(\mathbf{v}^*)\|_2 \leq \frac{4(\frac{\tau\epsilon^2}{20})}{\|\mathbf{p}-\mathbf{q}\|_2} < \frac{(\frac{\tau\epsilon^2}{5})}{(\frac{\tau\epsilon}{5})} \leq \epsilon,$$

concluding the proof that $U(C-C)$ is an ϵ-net for $U(S_1)$.

Step 2: Proving V is an ϵ-net for $\overline{U(S_2)}$.

The proof is based on the following two claims.

Claim 1. *If $\boldsymbol{v} \in \overline{U(S_2)}$, then there exists a unit tangent vector $\boldsymbol{w}$ to $\mathcal{M}$ such that*

$$\|\boldsymbol{v}-\boldsymbol{w}\|_2 < \frac{\epsilon}{3}. \tag{21}$$

Claim 2. *If $\boldsymbol{w}$ is a unit tangent vector to $\mathcal{M}$, then there exists $\boldsymbol{c} \in C$ and a unit tangent vector $\boldsymbol{w}^* \in T_{\boldsymbol{c}}\mathcal{M}$ such that*

$$\|\boldsymbol{w}-\boldsymbol{w}^*\|_2 < \frac{\epsilon}{3}. \tag{22}$$

We now prove that V is an ϵ-net for $\overline{U(S_2)}$ assuming the validity of these claims. Given $\mathbf{v} \in \overline{U(S_2)}$, let $\mathbf{w}$, $\mathbf{c}$, and $\mathbf{w}^*$ be as in the statements of the two Claims. By the definition of $V_{\mathbf{c}}$, there exists $\mathbf{v}^* \in V_{\mathbf{c}} \subset V$ such that

$$\|\mathbf{w}^* - \mathbf{v}^*\|_2 < \frac{\epsilon}{3}. \tag{23}$$

The triangle inequality and (21)-(23) now imply

$$\|\mathbf{v} - \mathbf{v}^*\| < \epsilon,$$

concluding the conditional proof. It remains to prove the claims.

Proof of Claim 1: Let $\mathbf{v} \in \overline{U(S_2)}$. Without loss of generality, $\mathbf{v}$ is not a unit tangent vector to $\mathcal{M}$ and so there exist $\mathbf{p}, \mathbf{q} \in \mathcal{M}$ with

$$0 < \|\mathbf{p} - \mathbf{q}\|_2 \leq \frac{\tau\epsilon}{5} \tag{24}$$

such that $\mathbf{v} = U(\mathbf{p} - \mathbf{q})$. There exists a unit speed minimizing geodesic $\gamma : [0, d_{\mathcal{M}}(\mathbf{p}, \mathbf{q})] \to \mathcal{M}$ joining $\mathbf{p}$ to $\mathbf{q}$. Consider the unit tangent vector $\mathbf{u} = \gamma'(0)$ and let ϕ denote the angle between $\mathbf{u}$ and $-\mathbf{v} = U(\mathbf{q} - \mathbf{p})$. Using Lemma 4.9, (24), and the hypothesis $0 < \epsilon < 1$, we have

$$\sin(\phi) \leq \frac{\|\mathbf{p} - \mathbf{q}\|_2}{2\tau}\left(1 + \frac{2\|\mathbf{p} - \mathbf{q}\|_2}{\tau}\right)^2 \leq \frac{\epsilon(5 + 2\epsilon)^2}{250} < \frac{49\epsilon}{250}. \tag{25}$$

By (25), $\sin(\phi) < \frac{1}{2}$ and so $0 \leq \phi < \frac{\pi}{6}$. It follows

$$2\sin(\phi/2) \leq 3/2\sin(\phi). \tag{26}$$

Let $\mathbf{w} = -\mathbf{u}$ and use (25)-(26) to conclude

$$\|\mathbf{v} - \mathbf{w}\|_2 = \|-\mathbf{v} - \mathbf{u}\|_2 = 2\sin(\frac{\phi}{2}) \leq \frac{3}{2}\sin(\phi) < \frac{147\epsilon}{500} < \frac{\epsilon}{3},$$

concluding the proof of Claim 1.

Proof of Claim 2: Let $\mathbf{x} \in \mathcal{M}$ and let $\mathbf{w} \in T_{\mathbf{x}}\mathcal{M}$ be a unit length tangent vector. Let $\mathbf{c} \in C$ be a closest net point to $\mathbf{x}$ so that

$$\|\mathbf{x} - \mathbf{c}\|_2 < \frac{\tau\epsilon^2}{20}. \tag{27}$$

Consider a unit speed minimizing geodesic

$$\gamma : [0, d_{\mathcal{M}}(\mathbf{x}, \mathbf{c})] \to \mathcal{M}$$

joining $\mathbf{x}$ to $\mathbf{c}$ and let $\mathbf{w}^* \in T_{\mathbf{c}}\mathcal{M}$ be the unit-tangent vector obtained by parallel translating $\mathbf{w}$ along the geodesic γ. In addition, let θ denote the angle between $\mathbf{w}$ and $\mathbf{w}^*$. By Lemma 4.8,

$$\theta \leq \frac{d_{\mathcal{M}}(\mathbf{x}, \mathbf{c})}{\tau} \tag{28}$$

and

$$d_{\mathcal{M}}(\mathbf{x},\mathbf{c}) \leq \|\mathbf{x}-\mathbf{c}\|_2 \left(1+\frac{2\|\mathbf{x}-\mathbf{c}\|_2}{\tau}\right). \tag{29}$$

Using (27)-(29) and the hypothesis $0<\epsilon<1$, we now have

$$\|\mathbf{w}-\mathbf{w}^*\|_2 = 2\sin\left(\frac{\theta}{2}\right) \leq \theta \leq \frac{\|\mathbf{x}-\mathbf{c}\|_2}{\tau}\left(1+\frac{2\|\mathbf{x}-\mathbf{c}\|_2}{\tau}\right) < \frac{\epsilon^2}{20}\left(1+\frac{\epsilon^2}{10}\right) < \frac{\epsilon}{3},$$

concluding the proof of Claim 2.

As we have now established that D is an ϵ-net for $\overline{U(\mathcal{M}-\mathcal{M})}$, we have

$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})}) \leq |D|.$$

It remains to bound $|D|$ from above.

Bounding $|D|$ from above.

We first consider the case when $d \geq 2$. As C is an $\frac{\tau\epsilon^2}{20}$-net for $\mathcal{M}$ of minimal cardinality, Theorem 4.3 implies

$$|C| \leq N\left(\frac{\tau\epsilon^2}{20}\right), \tag{30}$$

where

$$N(x) := \frac{V_{\mathcal{M}}}{\omega_d\left(1-\frac{x^2}{24\tau_{\mathcal{M}}^2}\right)^{d-1}\left(\frac{x}{2}\right)^d} + \frac{V_{\partial\mathcal{M}}}{\omega_{d-1}\left(1-\frac{x^2}{96\tau_{\partial\mathcal{M}}^2}\right)^{d-2}\left(\frac{x}{4}\right)^{d-1}}. \tag{31}$$

As $|U(C-C)| \leq |C\times C|$, (30) implies that

$$|U(C-C)| \leq \left(N\left(\frac{\tau\epsilon^2}{20}\right)\right)^2. \tag{32}$$

Next, we estimate $|V|$. For each $\mathbf{c}\in C$, $V_{\mathbf{c}}$ is a minimal $\frac{\epsilon}{3}$-net in the unit tangent sphere $S_{\mathbf{c}}\mathcal{M}$. This sphere is isometric to the unit sphere $\mathbb{S}^{d-1}$. By [46, Corollary 4.2.13], $|V_{\mathbf{c}}| \leq \left(\frac{3}{\epsilon}\right)^d$. As $V = \bigcup_{\mathbf{c}\in C} V_{\mathbf{c}}$, we have

$$|V| \leq |C|\left(\frac{3}{\epsilon}\right)^d \leq N\left(\frac{\tau\epsilon^2}{20}\right)\left(\frac{3}{\epsilon}\right)^d. \tag{33}$$

Finally, since $D = U(C-C)\cup V$,

$$|D| \leq |U(C-C)| + |V| \leq \left(N\left(\frac{\tau\epsilon^2}{20}\right)\right)^2 + N\left(\frac{\tau\epsilon^2}{20}\right)\left(\frac{3}{\epsilon}\right)^d. \tag{34}$$

This right hand side of (34) is rather inconvenient so we will simplify it.

One has

$$N\left(\frac{\tau\epsilon^2}{20}\right) = \frac{V_{\mathcal{M}}}{\omega_d\left(1-\frac{(\frac{\tau\epsilon^2}{20})^2}{24\tau_{\mathcal{M}}^2}\right)^{d-1}\left(\frac{(\frac{\tau\epsilon^2}{20})}{2}\right)^d} + \frac{V_{\partial\mathcal{M}}}{\omega_{d-1}\left(1-\frac{(\frac{\tau\epsilon^2}{20})^2}{96\mu_{\partial\mathcal{M}}^2}\right)^{d-2}\left(\frac{(\frac{\tau\epsilon^2}{20})}{4}\right)^{d-1}}.$$

We note $0<\epsilon<1$, $\tau=\min\{\tau_{\mathcal{M}},\mu_{\partial\mathcal{M}}\}$, and so

$$\frac{(2\times 20)^d}{\left(1-\frac{\epsilon^4}{(24)(20^2)}\left(\frac{\tau}{\tau_{\mathcal{M}}}\right)^2\right)^{d-1}} < 40.005^d < 41^d$$

$$\frac{(4\times 20)^{d-1}}{\left(1-\frac{\epsilon^4}{(96)(20^2)}\left(\frac{\tau}{\mu_{\partial\mathcal{M}}}\right)^2\right)^{d-2}} < 80.003^d < 81^d$$

Define $\alpha = \frac{V_{\mathcal{M}}}{\omega_d}\left(\frac{41}{\tau}\right)^d + \frac{V_{\partial\mathcal{M}}}{\omega_{d-1}}\left(\frac{81}{\tau}\right)^{d-1}$. Then

$$N\left(\frac{\tau\epsilon^2}{20}\right) < \frac{\alpha}{\epsilon^{2d}}. \tag{35}$$

Finally, substitute (35) in (34) and use $0<\epsilon<1$ to obtain

$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})}) \le |D| < \left(\alpha^2+\alpha 3^d\right)\frac{1}{\epsilon^{4d}}$$

as in the statement of the Theorem.
We conclude with the case when $d=1$. By Theorem 4.3,

$$|U(C-C)| \le |C\times C| = |C|^2 \le \left(\frac{20V_{\mathcal{M}}}{\tau\epsilon^2}+V_{\partial\mathcal{M}}\right)^2.$$

For each $\mathbf{c}\in C$, there are precisely two unit length tangent vectors in $T_{\mathbf{c}}\mathcal{M}$ and so $|V_{\mathbf{c}}|\le 2$ and

$$|V| \le 2|C| \le 2\left(\frac{20V_{\mathcal{M}}}{\tau\epsilon^2}+V_{\partial\mathcal{M}}\right).$$

Therefore,

$$|D| \le |U(C-C)| + |V| \le \left(\frac{20V_{\mathcal{M}}}{\tau\epsilon^2}+V_{\partial\mathcal{M}}\right)^2 + 2\left(\frac{20V_{\mathcal{M}}}{\tau\epsilon^2}+V_{\partial\mathcal{M}}\right).$$

Let $\alpha = \frac{20V_{\mathcal{M}}}{\tau}+V_{\partial\mathcal{M}}$. Then

$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})} \le |D| < \left(\alpha^2+2\alpha\right)\frac{1}{\epsilon^4},$$

as in the statement of the Theorem. □

One might be concerned about the existence of a sequence of d-dimensional submanifolds of $\mathbb{R}^N$ with $\alpha \to 0$ as such a sequence would invalidate Theorem 4.4. In fact, no such sequence of manifolds exists. Indeed, if $d \geq 2$ and $\partial\mathcal{M} = \emptyset$, one can apply Proposition 4.2 to $\mathcal{M}$ directly to obtain

$$\alpha \geq \frac{41^d}{\omega_d}\mathcal{H}^d(\mathbb{S}^d), \tag{36}$$

and if $\partial\mathcal{M} \neq \emptyset$, one can apply Proposition 4.2 to a boundary component, a $d-1$-manifold without boundary to obtain

$$\alpha \geq \frac{81^{d-1}}{\omega_{d-1}}\mathcal{H}^{d-1}(\mathbb{S}^{d-1}). \tag{37}$$

Similarly, if $d = 1$ and $\partial\mathcal{M} = \emptyset$, Proposition 4.2 implies

$$\alpha \geq 20\mathcal{H}^1(\mathbb{S}^1), \tag{38}$$

and if $\partial\mathcal{M} \neq \emptyset$, then $V_{\partial\mathcal{M}} \geq 2$ whence

$$\alpha \geq 2. \tag{39}$$

As a final example we apply our covering estimate again with $\mathcal{M} = \mathbb{S}^d$ in Corollary 4.3 below, noting that in this case, $\overline{U(\mathcal{M} - \mathcal{M})} = \mathbb{S}^d$. There is considerable redundancy in this example as many pairs of secants project under U to the same unit secant. It is an interesting geometry question to find submanifolds such that their secants avoid being parallel. In this direction there is the work on totally skew embeddings [26]. Such submanifolds would be great candidates for benchmarking JL maps as their unit secants are expected to be large and "worst case" in terms of size. Here we lose constant factors in the exponent compared to prior bounds for $\mathbb{S}^d$ we have stated due to the high level of redundancy present in the unit secants of $\mathbb{S}^d$ which our general argument over counts.

Corollary 4.3. *Let $d \geq 2$ and consider $\mathbb{S}^d$ as a submanifold of $\mathbb{R}^N$. If $\epsilon \in (0,1)$, then*

$$N_\epsilon^{cover}(\overline{U(\mathbb{S}^d - \mathbb{S}^d)}) \leq \frac{20d41^{2d}}{\epsilon^{4d}}.$$

Proof. Using Theorem 4.4, it suffices to show

$$\alpha^2 + \alpha 3^d \leq 20d41^{2d}.$$

We have

$$\alpha = \frac{V_{\mathbf{S}^d}}{\omega_d}\left(\frac{41}{\tau}\right)^d + \frac{V_{\partial\mathbf{S}^{d-1}}}{\omega_{d-1}}\left(\frac{81}{\tau}\right)^{d-1} = \frac{V_{\mathbf{S}^d}}{\omega_d}\left(\frac{41}{\tau}\right)^d$$

since $\partial\mathbb{S}^d = \emptyset$. We have $\tau = 1$, $V_{\mathbb{S}^d} = 2\frac{\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}$ and $V_{\mathbb{B}^d} = \omega_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$. Thus,

$$\frac{V_{\mathbb{S}^d}}{\omega_d} = 2\sqrt{\pi}\frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d+1}{2})} \leq 2\sqrt{\pi d}$$

. Therefore $\alpha \leq 2\sqrt{\pi d}41^d$, and

$$\alpha^2 + \alpha 3^d \leq 4\pi d41^{2d} + 2\sqrt{\pi d}41^d 3^d \leq 16d41^{2d} + 4d41^{2d}.$$

□

Having established covering number bounds for the unit secants of general compact smooth submanifolds of $\mathbb{R}^N$, we are now able to bound the Gaussian Widths of these sets. After doing so we will then be able to use the established bounds together with results from Sections 2 and 3 to produce a variety of new embedding results for submanifolds.

4.4 A Gaussian Width Bound for Unit Secants from Above

Theorem 4.4 will now be used to bound the Gaussian width of the closure of the unit secant set for a compact smooth submanifold of $\mathbb{R}^N$.

Theorem 4.5 (The Gaussian Width of the Unit Secants of a Submanifold of $\mathbb{R}^N$ with Boundary)**.** *Let $\mathcal{M}$ be a compact smooth d-dimensional submanifold of $\mathbb{R}^N$ with $d \geq 2$ and with $\tau_{\mathcal{M}} < \infty$. Let α and τ be as in Theorem 4.4 and let $c = \alpha^2 + \alpha 3^d$. Then the Gaussian width of $\overline{U(\mathcal{M}-\mathcal{M})}$ satisfies*

$$\omega(\overline{U(\mathcal{M}-\mathcal{M})}) \leq 8\sqrt{2}\sqrt{\ln(c)+4d}.$$

Proof. Note that by (36)-(39), $c > 1$. We use the covering number bounds in Theorem 4.4 and Dudley's inequality (see, e.g., Theorem 8.23 in [24]):

$$\omega\left(\overline{U(\mathcal{M}-\mathcal{M})}\right) \leq 4\sqrt{2}\int_0^\infty \sqrt{\ln\left(N_\epsilon^{cover}\left(\overline{U(\mathcal{M}-\mathcal{M})}\right)\right)}\, d\epsilon. \tag{40}$$

By Theorem 4.4, for each $\epsilon \in (0,1)$,

$$N_\epsilon^{cover}(\overline{U(\mathcal{M}-\mathcal{M})}) \leq \frac{c}{\epsilon^{4d}}.$$

As the covering numbers $N_\epsilon^{cover}\left(\overline{U(\mathcal{M}-\mathcal{M})}\right)$ are non-increasing in ϵ, for each $\epsilon \geq 1$,

$$N_\epsilon^{cover}\left(\overline{U(\mathcal{M}-\mathcal{M})}\right) \leq c.$$

As $U(\mathcal{M}-\mathcal{M}) \subset \mathbb{S}^{N-1}$, for each $\epsilon > 2$,

$$N_\epsilon^{cover}\left(\overline{U(\mathcal{M}-\mathcal{M})}\right) = 1.$$

Therefore

$$\begin{aligned}
\omega\left(\overline{U(\mathcal{M}-\mathcal{M})}\right) &\leq 4\sqrt{2}\left(\int_0^1 \sqrt{\ln\left(\frac{c}{\epsilon^{4d}}\right)}\, d\epsilon \;+\; \int_1^2 \sqrt{\ln(c)}\, d\epsilon\right) \\
&= 4\sqrt{2}\int_0^1 \sqrt{\ln(c) + 4d\ln\left(\frac{1}{\epsilon}\right)}\, d\epsilon \;+\; 4\sqrt{2}\sqrt{\ln(c)} \\
&\leq 4\sqrt{2}\sqrt{\ln(c)}\left(1 + \int_0^1 \sqrt{1 + \frac{4d}{\ln(c)}\ln\left(\frac{1}{\epsilon}\right)}\, d\epsilon\right) \\
&\leq 4\sqrt{2}\sqrt{\ln(c)}\left(1 + \sqrt{\int_0^1 1 + \frac{4d}{\ln(c)}\ln\left(\frac{1}{\epsilon}\right)\, d\epsilon}\right),
\end{aligned}$$

where the last inequality follows from Cauchy-Schwartz. Note that for $k > 0$

$$\begin{aligned}\int_0^1 1 + k \ln\left(\frac{1}{x}\right)\, dx &= 1 + \lim_{a \to 0} -k(x\ln(x) - x)\Big|_a^1 \\ &= 1 + k.\end{aligned}$$

Hence,

$$\begin{aligned}\omega\left(\overline{U(\mathcal{M} - \mathcal{M})}\right) &\leq 4\sqrt{2}\sqrt{\ln(c)}\left(1 + \sqrt{1 + \frac{4d}{\ln(c)}}\right) \\ &\leq 8\sqrt{2}\sqrt{\ln(c) + 4d}\end{aligned}$$

as claimed.

□

We have now established all the results needed to prove our main theorems.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

[1] Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electronic Journal of Statistics*, 13(1):1359–1399, 2019. 25

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

[3] Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 557–563, 2006. 2

[4] Nir Ailon and Edo Liberty. Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes. *Discrete & Computational Geometry*, 42(4):615–630, December 2009. 10, 19

[5] Ralph Alexander and Stephanie Alexander. Geodesics in Riemannian manifolds-with-boundary. *Indiana University Mathematics Journal*, 30(4):481–488, 1981. Publisher: JSTOR. 28

[6] Stephanie B. Alexander, I. David Berg, and Richard L. Bishop. The Riemannian obstacle problem. *Illinois Journal of Mathematics*, 31(1):167–184, March 1987. 28, 29

[7] Stefan Bamberger, Felix Krahmer, and Rachel Ward. Johnson-lindenstrauss embeddings with kronecker structure, 2021. 2, 4

[8] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008. 22

[9] Richard G Baraniuk and Michael B Wakin. Random projections of smooth manifolds. *Foundations of computational mathematics*, 9(1):51–77, 2009. 2

[10] Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of Applied and Computational Topology*, 3(1-2):29–58, June 2019. 25, 29, 37

[11] Karol Borsuk. Sur la courbure totale des courbes fermées. In *Annales de la Société Polonaise de Mathématique*. [sn], 1948. 31

[12] Károly Böröczky, Jr. *Finite Packing and Covering*. Cambridge Tracts in Mathematics. Cambridge University Press, 2004. 32

[13] Simone Brugiapaglia, Sjoerd Dirksen, Hans Christian Jung, and Holger Rauhut. Sparse recovery in bounded Riesz systems with applications to numerical methods for PDEs. *Applied and Computational Harmonic Analysis*, 53:231–269, July 2021. 10, 14

[14] Minhua Chen, Jorge Silva, John Paisley, Chunping Wang, David Dunson, and Lawrence Carin. Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds. *IEEE Transactions on Signal Processing*, 58(12):6140–6155, 2010. 2

[15] Kenneth L. Clarkson. Tighter bounds for random projections of manifolds. In *Proceedings of the twenty-fourth annual symposium on Computational geometry - SCG '08*, page 39, College Park, MD, USA, 2008. ACM Press. 2, 36

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

[17] Mark A Davenport, Marco F Duarte, Michael B Wakin, Jason N Laska, Dharmpal Takhar, Kevin F Kelly, and Richard G Baraniuk. The smashed filter for compressive classification and target recognition. In *Computational Imaging V*, volume 6498, page 64980H. International Society for Optics and Photonics, 2007. 2

[18] Sjoerd Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Foundations of Computational Mathematics*, 16(5):1367–1396, 2016. 2

[19] Sjoerd Dirksen, Mark Iwen, Sara Krause-Solberg, and Johannes Maly. Robust one-bit compressed sensing with manifold data. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pages 1–5. IEEE, 2019. 2

[20] Armin Eftekhari and Michael B. Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, 39(1):67–109, July 2015. 2, 7, 10, 25, 36

[21] Armin Eftekhari and Michael B Wakin. What happens to a manifold under a bi-lipschitz map? *Discrete & Computational Geometry*, 57(3):641–673, 2017. 2

[22] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–418, March 1959. 2, 25, 26

[23] Werner Fenchel. On the differential geometry of closed space curves. *Bulletin of the American Mathematical Society*, 57(1):44–55, January 1951. 31

[24] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Springer New York, New York, NY, 2013. 2, 12, 14, 16, 21, 22, 44, 50, 52, 54

[25] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry*, volume 2. Springer, 1990. 25, 30

[26] Mohammad Ghomi and Serge Tabachnikov. Totally skew embeddings of manifolds. *Mathematische Zeitschrift*, 258(3):499–512, 2008. 43

[27] Chinmay Hegde, Michael Wakin, and Richard Baraniuk. Random projections for manifold learning. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, pages 641–648. Curran Associates, Inc., 2008. 2

[28] Chang Min Hyun, Seong Hyeon Baek, Mingyu Lee, Sung Min Lee, and Jin Keun Seo. Deep learning-based solvability of underdetermined inverse problems in medical imaging. *arXiv preprint arXiv:2001.01432*, 2020. 2

[29] Mark Iwen. A mathematical introduction to fast and memory efficient algorithms for big data. *Publicly Available Course Notes*, 2020. 6

[30] Mark Iwen, Arman Tavakoli, and Benjamin Schmidt. Lower bounds on the low-distortion embedding dimension of submanifolds of $\mathbb{R}^N$. *arXiv preprint arXiv:2105.13512*, 2021. 13

[31] Mark A Iwen, Felix Krahmer, Sara Krause-Solberg, and Johannes Maly. On recovery guarantees for one-bit compressed sensing on manifolds. *Discrete & Computational Geometry*, 65(4):953–998, 2021. 2

[32] Mark A Iwen, Eric Lybrand, Aaron A Nelson, and Rayan Saab. New algorithms and improved guarantees for one-bit compressed sensing on manifolds. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pages 1–4. IEEE, 2019. 2

[33] Mark A Iwen and Mauro Maggioni. Approximation of points on low-dimensional manifolds via random linear projections. *Information and Inference: A Journal of the IMA*, 2(1):1–31, 2013. 2

[34] Mark A. Iwen, Deanna Needell, Elizaveta Rebrova, and Ali Zare. Lower memory oblivious (tensor) subspace embeddings with fewer random bits: Modewise methods for least squares. *SIAM Journal on Matrix Analysis and Applications*, 42(1):376–416, 2021. 4, 22

[35] Felix Krahmer and Rachel Ward. New and improved johnson–lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3):1269–1281, 2011. 2, 12, 15

[36] Subhaneil Lahiri, Peiran Gao, and Surya Ganguli. Random projections of random manifolds. *arXiv preprint arXiv:1607.04331*, 2016. 2, 36

[37] R. Lashof and S. Smale. On the Immersion of Manifolds in Euclidean Space. *The Annals of Mathematics*, 68(3):562, November 1958. 35

[38] S. Li. Concise Formulas for the Area and Volume of a Hyperspherical Cap. *Asian Journal of Mathematics & Statistics*, 4(1):66–70, December 2010. 30

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

[40] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete & Computational Geometry*, 39(1-3):419–441, March 2008. 29, 36

[41] Barrett O'Neill. *Semi-Riemannian geometry: with applications to relativity*. Number 103 in Pure and applied mathematics. Academic Press, New York, 1983. 29

[42] Samet Oymak, Benjamin Recht, and Mahdi Soltanolkotabi. Isometric sketching of any set via the Restricted Isometry Property. *Information and Inference: A Journal of the IMA*, 7(4):707–726, December 2018. 10, 16, 17, 18, 19

[43] William F. Pohl. Some Integral Formulas for Space Curves and Their Generalization. *American Journal of Mathematics*, 90(4):1321, October 1968. 35

[44] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000. 2

[45] Christoph Thäle. 50 years sets with positive reach–a survey. *Surveys in Mathematics and its Applications*, 3:123–165, 2008. Publisher: University Constantin Brancusi. 25

[46] Roman 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. 2, 3, 12, 13, 22, 31, 32, 36, 41, 51, 53

[47] James H. White. Self-Linking and the Gauss Integral in Higher Dimensions. *American Journal of Mathematics*, 91(3):693, July 1969. 35

[48] James H. White. Self-linking and the directed secant span of a differentiable manifold. *Journal of Differential Geometry*, 5(3-4):357–369, 1971. 35

[49] Han Lun Yap, Michael B Wakin, and Christopher J Rozell. Stable manifold embeddings with structured random matrices. *IEEE Journal of Selected Topics in Signal Processing*, 7(4):720–730, 2013. 2, 4, 5, 8

A The proof of Theorem 3.4

The following lemma describes properties that the matrices A and B can have in order to guarantee that E in (1) will approximately preserve the norms of all elements of an arbitrary bounded set

$S \subset \mathbb{R}^N$. Though we will present the proof in general, we will be primarily interested in the case where S contains all unit norm s-sparse vectors so that

$$S = U\left(\bigcup_{S' \subset [N],\ |S'|=s} \operatorname{span}\left(\{\mathbf{e}_j\}_{j \in S'}\right)\right). \tag{41}$$

Lemma A.1. *Let $\epsilon \in \left(0, \frac{1}{3}\right)$, $S \subset \mathbb{R}^N$, and $A \in \mathbb{R}^{m_1 \times m_1^2}$, $B \in \mathbb{R}^{m_2 \times N/m_1}$, $C \in \mathbb{R}^{N/m_1 \times N}$, and $E \in \mathbb{R}^{m_2 \times N}$ be as above in (1) with $m_1 \geq m_2$. Furthermore, let $a_E := \max\left\{\sup_{\mathbf{x} \in U(S-S)} \|E\mathbf{x}\|_2,\ 1\right\}$, $\mathcal{C}_\delta \subset S$ be a finite δ-cover of S for $\delta \leq \epsilon / a_E$, and suppose that*

(a) A is an ϵ-JL map of $P_j \mathcal{C}_\delta$ into $\mathbb{R}^{m_1}$ for all $j \in [N/m_1^2]_0$, and that

(b) $\frac{1}{\sqrt{m_2}} B$ is an ϵ-JL map of $C\mathcal{C}_\delta$ into $\mathbb{R}^{m_2}$.

Then,

$$(1-2\epsilon)\|\mathbf{x}\|_2 - \epsilon(1+\sqrt{5/3}) \leq \|E\mathbf{x}\|_2 \leq (1+3\epsilon/2)\|\mathbf{x}\|_2 + \epsilon(1+\sqrt{2}) \tag{42}$$

will hold for all $\mathbf{x} \in S$. If, in addition, S is a subset of the unit sphere so that $\|\mathbf{x}\|_2 = 1$ for all $\mathbf{x} \in S$, then $E = \frac{1}{\sqrt{m_2}} BC \in \mathbb{R}^{m_2 \times N}$ will also be a 14ϵ-JL map of S into $\mathbb{R}^{m_2}$.

Proof. By Lemma 3.1 we see that E will be a 3ϵ-JL map of $\mathcal{C}_\delta$ into $\mathbb{R}^{m_2}$ since

$$\begin{aligned}(1-2\epsilon)\|\mathbf{x}\|_2^2 \leq (1-\epsilon)^2\|\mathbf{x}\|_2^2 \leq (1-\epsilon)\|C\mathbf{x}\|_2^2 \leq \|E\mathbf{x}\|_2^2 \\ \leq (1+\epsilon)\|C\mathbf{x}\|_2^2 \leq (1+\epsilon)^2\|\mathbf{x}\|_2^2 \leq (1+3\epsilon)\|\mathbf{x}\|_2^2\end{aligned} \tag{43}$$

will hold for all $\mathbf{x} \in \mathcal{C}_\delta$. Continuing, let $\mathbf{y} \in S$ and choose $\mathbf{x} \in \mathcal{C}_\delta \subset S$ be such that $\|\mathbf{y} - \mathbf{x}\|_2 \leq \delta$. Using (43) we have that

$$\begin{aligned}\|E\mathbf{y}\|_2 &\leq \|E\mathbf{x}\|_2 + \|E(\mathbf{y}-\mathbf{x})\|_2 \leq \sqrt{1+3\epsilon}\|\mathbf{x}\|_2 + a_E\|\mathbf{y}-\mathbf{x}\|_2 \\ &\leq \sqrt{1+3\epsilon}\|\mathbf{y}\|_2 + \delta\sqrt{1+3\epsilon} + a_E\delta \leq (1+3\epsilon/2)\|\mathbf{y}\|_2 + \epsilon(1+\sqrt{2}).\end{aligned} \tag{44}$$

Similarly, we will also have that

$$\begin{aligned}\|E\mathbf{y}\|_2 &\geq \|E\mathbf{x}\|_2 - \|E(\mathbf{y}-\mathbf{x})\|_2 \geq \sqrt{1-2\epsilon}\|\mathbf{x}\|_2 - a_E\|\mathbf{y}-\mathbf{x}\|_2 \\ &\geq \sqrt{1-2\epsilon}\|\mathbf{y}\|_2 - \delta\sqrt{1+2\epsilon} - a_E\delta \geq (1-2\epsilon)\|\mathbf{y}\|_2 - \epsilon(1+\sqrt{5/3}).\end{aligned} \tag{45}$$

Combining (44) and (45) gives us (42). Finally, if all the elements of S are unit norm, then we can see from (44) and (45) that

$$(1-5\epsilon)\|\mathbf{y}\|_2 \leq \|E\mathbf{y}\|_2 \leq (1+4\epsilon)\|\mathbf{y}\|_2$$

will hold for all $\mathbf{y} \in S$. Squaring throughout now proves the remaining claim. □

Note that Lemma A.1 requires the matrix $B/\sqrt{m_2}$ to be an ϵ-JL map of a finite subset S_C of $C(S)$ (see assumption (b)). In addition, we need to have some way of bounding $a_E = \sup_{\mathbf{x}\in U(S-S)} \|E\mathbf{x}\|_2$ from above in order to safely upper bound the cardinality of $S_C = C(C_\delta)$ in the first place. The next lemma addresses both of these needs for sub-gaussian matrices B.

Lemma A.2. *Let $\epsilon, p \in \left(0, \frac{1}{3}\right)$, $S \subset \mathbb{R}^N$, and $S_C \subset C(S) \subset \mathbb{R}^{N/m_1}$ be finite. Furthermore, suppose that $B \in \mathbb{R}^{m_2 \times N/m_1}$ in the definition of E in (1) has $m_2 \geq c_1 \epsilon^{-2} \ln(|S_C|/p)$ independent, isotropic and sub-gaussian rows. Then, all of*

(1) $\frac{1}{\sqrt{m_2}} B$ *will be an ϵ-JL map of S_C into $\mathbb{R}^{m_2}$, and*

(2) $$\sup_{\mathbf{x}\in U(S-S)} \|E\mathbf{x}\|_2 \leq c_2 \, \|A\| \left(\frac{w\left(U(S-S)\right) + \sqrt{\ln(\frac{1}{p})}}{\sqrt{m_2}} + 1 \right) \leq c_3 \left(\frac{\|A\|}{\sqrt{m_2}} \right) \left(\sqrt{N} + \sqrt{\ln\left(\frac{1}{p}\right)} \right)$$

will hold simultaneously with probability at least $1-p$. Here $c_1, c_2, c_3 \in \mathbb{R}^+$ are constants that only depend on the distributions of the rows of B (i.e., they are absolute constants once distributions for the rows of B are fixed).

Proof. To prove that both conclusions (1) and (2) above hold simultaneously with probability at least $1-p$, we will prove that each one holds separately with probability at least $1-p/2$. The desired result will then follow from the union bound.

To establish conclusion (1) above with probability at least $1-p/2$ one may simply appeal, e.g., to the proof of [24, Lemma 9.35].

Toward establishing conclusion (2) above we first note that

$$\begin{aligned} \|C\| \;&=\; \sup_{\mathbf{x}\in U(\mathbb{R}^N)} \|C\mathbf{x}\|_2 \leq \sqrt{\sup_{\mathbf{x}\in U(\mathbb{R}^N)} \|C\mathbf{x}\|_2^2} \;=\; \sqrt{\sup_{\mathbf{x}\in U(\mathbb{R}^N)} \sum_{j\in[N/m_1^2]_0} \|AP_j\mathbf{x}\|_2^2} \\ &\leq \sqrt{\sup_{\mathbf{x}\in U(\mathbb{R}^N)} \sum_{j\in[N/m_1^2]_0} \|A\|^2 \|P_j\mathbf{x}\|_2^2} \;=\; \|A\|. \end{aligned} \tag{46}$$

Continuing, we have

$$\begin{aligned} \sup_{\mathbf{x}\in U(S-S)} \|E\mathbf{x}\|_2 \;&=\; \frac{1}{\sqrt{m_2}} \sup_{\mathbf{x}\in C(U(S-S))} \|B\mathbf{x}\|_2 \;\leq\; \frac{1}{\sqrt{m_2}} \sup_{\mathbf{x}\in C(U(S-S))} \big| \|B\mathbf{x}\|_2 - \sqrt{m_2}\|\mathbf{x}\|_2 \big| + \sqrt{m_2}\|\mathbf{x}\|_2 \\ &\leq \frac{1}{\sqrt{m_2}} \left(\sup_{\mathbf{x}\in C(U(S-S))} \big| \|B\mathbf{x}\|_2 - \sqrt{m_2}\|\mathbf{x}\|_2 \big| \right) + \sup_{\mathbf{x}\in U(S-S)} \|C\mathbf{x}\|_2. \end{aligned}$$

Now appealing to Theorem 2.1 and (46) we can see that

$$\begin{aligned}\sup_{\mathbf{x}\in U(S-S)} \|E\mathbf{x}\|_2 \leq & \frac{\tilde{c}\left(w\left(C(U(S-S))\right) + \sqrt{\ln(\frac{4}{p})}\cdot \sup_{\mathbf{x}\in C(U(S-S))}\|\mathbf{x}\|_2\right)}{\sqrt{m_2}} + \|C\| \\ \leq & \frac{\tilde{c}\left(w\left(C(U(S-S))\right) + \sqrt{\ln(\frac{4}{p})}\cdot \|A\|\right)}{\sqrt{m_2}} + \|A\|.\end{aligned}$$

will hold with probability at least $1-p/2$, where $\tilde{c}\in\mathbb{R}^+$ is a constant that only depends on the distributions of the rows of B. Finally, using properties of Gaussian width (see, e.g., [46, Exercise 7.5.4]) the last inequality can be simplified further to

$$\sup_{\mathbf{x}\in U(S-S)} \|D\mathbf{x}\|_2 \leq \frac{\tilde{c}\left(\|C\|\ w\left(U(S-S)\right) + \sqrt{\ln(\frac{4}{p})}\cdot \|A\|\right)}{\sqrt{m_2}} + \|A\|.$$

Using (46) one last time and simplifying using that $\ln(1/p)\geq 1$ now yields the first inequality in (2) above.

To obtain a different version of the second inequality in (2) one might be tempted to use, e.g., [46, Theorem 4.4.5] and then repeat analogous simplifications to those just performed above. Indeed, doing so provides a slight better bound on $\sup_{\mathbf{x}\in U(S-S)} \|E\mathbf{x}\|_2$ than the second inequality in (2) does in the end. However, for our purposes the second inequality in (2) suffices and also follows automatically from what we have already proven given that $w(U\left(S-S\right)) \leq w\left(U(\mathbb{R}^N)\right) \leq \sqrt{N}+c''$ for an absolute constant c'' (see, e.g., [46, Example 7.5.7]). Simplifying using that $N/m_2\geq 1$ finishes the job. □

Lemma A.2 proposes that the matrix B in the definition of E in (1) be chosen as a sub-gaussian random matrix. Indeed, it demonstrates that doing so will at least partially fulfill the requirements of Lemma A.1 with high probability. Our next lemma proposes an auspicious choice for the remaining matrix $A\in\mathbb{R}^{m_1\times m_1^2}$.

Lemma A.3. *Fix $p,\epsilon\in(0,1/3)$, a finite set $\tilde{S}\subset\mathbb{R}^N$, $K\in\left[1,N^{\frac{1}{4}}\right)$, and suppose that $m_1\in\mathbb{Z}^+$ satisfies*

$$\sqrt{N}\ \geq\ m_1\ \geq\ cK^2\frac{\ln\left(N|\tilde{S}|/p\right)}{\epsilon^2}\ln(N/p)\ln^2\left(\frac{\ln\left(N|\tilde{S}|/p\right)K^2}{\epsilon}\right),$$

where $c\in\mathbb{R}^+$ is an absolute constant. Next, let $U\in\mathbb{R}^{m_1^2\times m_1^2}$ be a unitary matrix with BOS constant $m_1\cdot\max_{k,t}|u_{t,k}|\leq K$, $D\in\{0,-1,1\}^{m_1^2\times m_1^2}$ be a diagonal matrix with i.i.d. ±1 Rademacher random variables on its diagonal, and $R\in\{0,1\}^{m_1\times m_1^2}$ be m_1 rows independently selected uniformly at random from the $m_1^2\times m_1^2$ identity matrix. Set $A:=\sqrt{m_1}RUD$. Then, $\|A\|\leq m_1$ always. Furthermore, A will be an ϵ-JL map of $P_j\tilde{S}$ into $\mathbb{R}^{m_1}$ for all $j\in[N/m_1^2]_0$ with probability at least $1-p$.

Proof. Due to the unitary nature of both U and all admissible D we have

$$\|A\| \leq \sqrt{m_1}\|R\|\|U\|\|D\| \;=\; \sqrt{m_1}\|R\| = \sqrt{m_1} \sup_{\mathbf{x}\in U(\mathbb{R}^{m_1^2})} \sqrt{\sum_{j=1}^{m_1^2}\left(\sum_{\ell=1}^{m_1} R_{\ell,j}\right)|x_j|^2} \leq \sqrt{m_1\|R\|_1},$$

where $\|R\|_1 := \max\limits_{1\leq j\leq m_1^2}\sum\limits_{\ell=1}^{m_1}|R_{\ell,j}| \leq m_1$ for all admissible R. This proves the claim regarding $\|A\|$. Now fix $j \in [N/m_1^2]_0$ and let $s := 16\ln\left(8eN|\tilde{S}|/p\right) \geq \max\limits_{j\in[N/m_1^2]_0} 16\ln\left(8N|P_j\tilde{S}|/m_1^2p\right)$. For this choice of s [24, Theorem 9.36] guarantees that A will be an ϵ-JL map of $P_j\tilde{S}$ into $\mathbb{R}^{m_1}$ with probability at least $1 - p/(2N/m_1^2)$ provided that $\sqrt{m_1}RU$ has the RIP of order $(2s, \epsilon/4)$. The union bound then guarantees that A will be an ϵ-JL map of $P_j\tilde{S}$ into $\mathbb{R}^{m_1}$ for all $j \in [N/m_1^2]_0$ with probability at least $1 - p/2$. To finish, by a final application of the union bound it suffices to prove that $\sqrt{m_1}RU$ will indeed have the RIP of order $(2s, \epsilon/4)$ with probability at least $1 - p/2$. This RIP condition is provided by Corollary 2.4 for any BOS matrix $\frac{m_1}{\sqrt{m'}}R'U \in \mathbb{C}^{m'\times m_1^2}$ with at least

$$m' \geq m_{\min} := \left\lceil cK^2\frac{\ln\left(N|\tilde{S}|/p\right)}{\epsilon^2}\ln(m_1/p)\ln^2\left(\frac{\ln\left(N|\tilde{S}|/p\right)K^2}{\epsilon}\right)\right\rceil$$

rows, where $c \in \mathbb{R}^+$ is a sufficiently large absolute constant. Note that our assumed bounds on m_1 guarantee that $m_1 \geq m_{\min}$. Furthermore, if $m_1 > m_{\min}$ we can simply increase m' to m_1 without losing the desired RIP condition. □

We are now prepared to prove the main result of this section.

Theorem A.1. *Let $S \subset U(\mathbb{R}^N)$, $K \in \left[1, N^{\frac{1}{4}}\right)$, $\epsilon \in \left(0, \frac{1}{3}\right)$, $p \in \left(e^{-N}, \frac{1}{3}\right)$, and fix a sequence $X = X_1, \ldots$ of i.i.d. mean zero, sub-gaussian random variables from which to draw the entries of B in (1). Next, suppose that $m_1 \in \mathbb{Z}^+$ satisfies*

$$\sqrt{N} \;\geq\; m_1 \;\geq\; c_2K^2\frac{\ln\left(N\mathcal{N}(S, \frac{\epsilon}{c_1N})/p\right)}{\epsilon^2}\ln(N/p)\ln^2\left(\frac{\ln\left(N\mathcal{N}(S, \frac{\epsilon}{c_1N})/p\right)K^2}{\epsilon}\right),$$

and that $m_2 \in \mathbb{Z}^+$ satisfies

$$m_1 \geq m_2 \geq c_3\epsilon^{-2}\ln\left(\mathcal{N}(S,\delta)/p\right)$$

for $\delta := c_4\epsilon/\left(m_1\left(w\left(U(S-S)\right) + \sqrt{\ln(1/p)}\right)\right)$, where $c_1, c_2, c_3, c_4 \in \mathbb{R}^+$ are absolute constants. Finally, choose $A \in \mathbb{R}^{m_1\times m_1^2}$ and $B \in \mathbb{R}^{m_2\times N/m_1}$ in (1) so that:

1. *$A := \sqrt{m_1}RUD$ where $U \in \mathbb{R}^{m_1^2\times m_1^2}$ be a unitary matrix with BOS constant $m_1 \cdot \max\limits_{k,t}|u_{t,k}| \leq K$, $D \in \{0,-1,1\}^{m_1^2\times m_1^2}$ be a diagonal matrix with i.i.d. ± 1 Rademacher random variables on its diagonal, and $R \in \{0,1\}^{m_1\times m_1^2}$ be m_1 rows independently selected uniformly at random from the $m_1^2 \times m_1^2$ identity matrix.*

2. *B has i.i.d. mean zero, sub-gaussian entries drawn according to the first $m_2 N/m_1$ random variables in X.*

Then, $E = \frac{1}{\sqrt{m_2}} BC \in \mathbb{R}^{m_2 \times N}$ will be an ϵ-JL map of S into $\mathbb{R}_2^m$ with probability at least $1-p$. Furthermore, if $A \in \mathbb{R}^{m_1 \times m_1^2}$ has an $m_1^2 \cdot f(m_1)$ time matrix-vector multiplication algorithm, then E will have an $\mathcal{O}(N \cdot f(m_1))$-time matrix-vector multiply.

Proof. Note the stated result follows from Lemma A.1 provided that its assumptions (a) and (b) both hold with $\epsilon \leftarrow \epsilon/14$ and δ sufficiently small. Hence, we seek to establish that both of these assumptions will simultaneously hold with probability at least $1-p$ for our choices of A and B above. We will use Lemmas A.2 and A.3 to accomplish this objective below, thereby proving the theorem.

To begin, we will apply Lemma A.2 with $S \leftarrow S$, $p \leftarrow p/3$, $\epsilon \leftarrow \epsilon/14$, and $S_C \leftarrow C\mathcal{C}_\delta$ where $\mathcal{C}_\delta \subset S$ is a minimal δ-cover of S. In doing so we note that c_1, c_2, c_3 in Lemma A.2 will be absolute constants given X.[7] As a consequence we learn that both the event $\mathcal{E}_{(c)} := \left\{ \frac{1}{\sqrt{m_2}} B \text{ is an } \epsilon/14\text{-JL map of } C\mathcal{C}_\delta \text{ into } \mathbb{R}^{m_2} \right\}$, and that

$$\begin{aligned} a_E &\leq \sup_{\mathbf{x} \in U(S-S)} \|E\mathbf{x}\|_2 + 1 \leq c\, m_1 \left(\frac{w\left(U(S-S)\right) + \sqrt{\ln(\frac{2}{p})}}{\sqrt{m_2}} + 1 \right) \\ &\leq c' m_1 \left(w\left(U(S-S)\right) + \sqrt{\ln(1/p)} \right) =: a'_E \\ &\leq c_1 N, \end{aligned} \tag{47}$$

will simultaneously hold with probability at least $1 - p/3$, where c, c', c_1 are absolute constants. In (47) we have used the assumptions that $m_1 \leq \sqrt{N}$ and $(1/p) \leq e^N$ as well as the fact that $\|A\| \leq m_1$ always (see Lemma A.3), and that $w\left(U(S-S)\right) \leq \sqrt{N} + c''$ for an absolute constant c'' (see, e.g., [46, Example 7.5.7]). Furthermore, we note that (47) implies that $\frac{\epsilon}{c_1 N} \leq \delta = \frac{c' c_4 \epsilon}{a'_E} \leq \frac{\epsilon}{a_E}$ holds provided that, e.g., c_4 is chosen to be $1/c'$.

Next, Lemma A.3 with $p \leftarrow p/3$, $\epsilon \leftarrow \epsilon/14$, $\tilde{S} \leftarrow \mathcal{C}_\delta$, $K \leftarrow K$ reveals that $\mathcal{E}_{(b)} := \left\{ A \text{ is an } \epsilon/14\text{-JL map of } P_j \mathcal{C}_\delta \text{ into } \mathbb{R}^{m_1} \text{ for all } j \in [N/m_1^2]_0 \right\}$ will also hold with probability at least $1 - p/3$ provided that (47) holds. Here we have used the fact that $\mathcal{N}\left(S, \frac{\epsilon}{c_1 N}\right) \geq \mathcal{N}(S, \delta) = |\mathcal{C}_\delta|$ when $\delta \geq \epsilon / c_1 N$. As a consequence we can finally see that both assumptions (a) and (b) of Lemma A.1 with $\epsilon \leftarrow \epsilon/14$

[7] Note that c_1, c_2, c_3 depend on the subgaussian norms of the rows of B in Lemma A.2 through an application of Theorem 2.1. However, after the distributions of B's entries, X, are fixed these norms are also both fixed and independent of the final length of the rows (see, e.g., [46, Lemma 3.4.2 and Theorem 9.1.1] in connection with the use of Theorem 2.1 in the proof of Lemma A.2).

will hold with probability at least $1-p$ since

$$\begin{aligned}\mathbb{P}\left[\mathcal{E}_{(c)}\cap(47)\cap\mathcal{E}_{(b)}\right] &\geq 1-\mathbb{P}\left[\overline{(47)\cap\mathcal{E}_{(b)}}\right]-p/3=\mathbb{P}\left[(47)\cap\mathcal{E}_{(b)}\right]-p/3\\ &=\mathbb{P}\left[\mathcal{E}_{(b)}\mid(47)\right]\mathbb{P}\left[(47)\right]-p/3\geq(1-p/3)\mathbb{P}\left[(47)\right]-p/3\\ &\geq(1-p/3)^2-p/3>1-p.\end{aligned}$$

Lemma A.1 now finishes the proof. The runtime result follows from Lemma 1.1. □

Note that an application of Theorem A.1 requires a valid choice of m_1 to be made. This will effectively limit the sizes of the sets which we can embed below. In order to make the discussion of this limitation a bit easier below we can further simplify the lower bound for m_1 by noting that for all fixed $S\subset U(\mathbb{R}^N)$, $K\in[1,N^{\frac{1}{4}})$, $\epsilon\in(0,1/3)$ and $p\in\left(e^{-N},\frac{1}{3}\right)$ we will have

$$\ln\left(\frac{N}{p}\right)\ln^2\left(\frac{\ln\left(N\mathcal{N}(S,\frac{\epsilon}{c_1N})/p\right)K^2}{\epsilon}\right)\leq\ln^3\left(\frac{NK^2}{\epsilon p}\right)\leq c\ln^3\left(\frac{N}{\epsilon p}\right)$$

for an absolute constant $c\in\mathbb{R}^+$, provided that $\mathcal{N}\left(S,\frac{\epsilon}{c_1N}\right)\leq pe^N/N$. As a consequence, we may weaken the lower bound for m_1 and instead focus on the smaller interval

$$\sqrt{N}\geq m_1\geq c_2''K^2\frac{\ln\left(\mathcal{N}(S,\frac{\epsilon}{c_1N})/p\right)}{\epsilon^2}\ln^4\left(\frac{N}{\epsilon p}\right)\geq c_2'K^2\frac{\ln\left(N\mathcal{N}(S,\frac{\epsilon}{c_1N})/p\right)}{\epsilon^2}\ln^3\left(\frac{N}{\epsilon p}\right)$$

for simplicity. Further assuming that K is upper bounded by a universal constant below (as it will be in all subsequent applications) we can see that our smaller range for m_1 will be nonempty when

$$\mathcal{N}\left(S,\delta\right)\leq\mathcal{N}\left(S,\frac{\epsilon}{c_1N}\right)\leq pe^{c\epsilon^2\sqrt{N}/\ln^4\left(\frac{N}{\epsilon p}\right)}\tag{48}$$

for a sufficiently small absolute constant $c\in\mathbb{R}^+$. We will use (48) in place of (8) below to limit the sizes of the sets that we embed so that Theorem A.1 can always be applied with a valid minimal choice of $m_1\leq c_2'''K^2\frac{\ln\left(\mathcal{N}(S,\frac{\epsilon}{c_1N})/p\right)}{\epsilon^2}\ln^4\left(\frac{N}{\epsilon p}\right)\leq\sqrt{N}$ below.

A.1 Theorem 3.4 as a Corollary of Theorem A.1

As above, we let B have, e.g., i.i.d. Rademacher entries and will choose $U\in\mathbb{R}^{m_1^2\times m_1^2}$ to be, e.g., a Hadamard or DCT matrix (see, e.g., [24, Section 12.1].) Making either choice for U will endow A with an $\mathcal{O}(m_1^2\log(m_1))$-time matrix vector multiply via FFT-techniques, and will also ensure that $K=\sqrt{2}$ always suffices. As a result, we note that $f(m_1)=\mathcal{O}(\log(m_1))$ in Theorem A.1.

We may therefore apply Theorem A.1 with S as in (41). To upper bound the final embedding dimension we note that we may safely choose

$$m_2\geq c_3\epsilon^{-2}\ln\left(\mathcal{N}\left(S,\frac{\epsilon}{c_1N}\right)/p\right)\geq c_3\epsilon^{-2}\ln\left(\mathcal{N}(S,\delta)/p\right).$$

Furthermore, applying Lemmas 3.2 and 3.3 we can further see that

$$\left(\frac{eN}{s}\right)^s \left(\frac{3c_1 N}{\epsilon}\right)^s \geq \mathcal{N}\left(S, \frac{\epsilon}{c_1 N}\right).$$

The stated lower bound on m now follows after adjusting and simplifying constants. Finally, and most crucially, the condition it suffices for s to satisfy now also follows from (48).