JMLR: Workshop and Conference Proceedings vol 23 (2012) 32.1–32.28 25th Annual Conference on Learning Theory

Distance Preserving Embeddings for General n-Dimensional Manifolds

Nakul Verma NAVERMA@CS.UCSD.EDU
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093-0404, USA

Editor: Shie Mannor, Nathan Srebro, Robert C. Williamson

Abstract

Low dimensional embeddings of manifold data have gained popularity in the last decade. However, a systematic finite sample analysis of manifold embedding algorithms largely eludes researchers. Here we present two algorithms that, given access to just the samples, embed the underlying n-dimensional manifold into $\mathbb{R}^d$ (where d only depends on some key manifold properties such as its intrinsic dimension, volume and curvature) and *guarantee* to approximately preserve all interpoint geodesic distances.

Keywords: Manifold Learning, Isometric Embeddings, Nash's Embedding Theorem

1. Introduction

Finding low dimensional representations of manifold data have gained popularity over the years. One typically assumes that points are sampled from an n-dimensional manifold residing in some high-dimensional ambient space $\mathbb{R}^D$ and analyze to what extent their low dimensional embedding maintains some important manifold property, say, interpoint geodesic distances. Despite an abundance of manifold embedding algorithms, only a few provide any kind of distance preserving guarantee. Isomap (Tenebaum et al., 2000), for instance, provides an asymptotic guarantee that as one increases the amount of data sampled from an underlying manifold, one can approximate the geodesic distances between the sample points well (Bernstein et al., 2000). Then, under a very restricted class of n-dimensional manifolds, one can show that the n-dimensional embedding returned by Isomap is approximately distance preserving on the input samples.

Unfortunately any kind of systematic finite sample analysis of manifold embedding algorithms—especially for general classes of manifolds—still largely eludes the manifold learning community. Part of the difficulty is due to the restriction of finding an embedding in exactly n dimensions. It turns out that many simple manifolds (such as a closed loop, a cylinder, a section of a sphere) cannot be isometrically embedded in $\mathbb{R}^n$, where n is the manifold's intrinsic dimension. If these manifolds reside in some high dimensional ambient space, we would at least like to embed them in a lower dimensional space (possibly slightly larger than n) while still preserving interpoint geodesic distances.

Here we are interested in investigating low-dimensional distance-preserving manifold embeddings more formally. Given a sample X from an underlying n-dimensional manifold $M \subset \mathbb{R}^D$, and an embedding procedure $\mathcal{A} : M \to \mathbb{R}^d$ that (uses X in training and) maps points from M into some low dimensional space $\mathbb{R}^d$, we define the quality of the embedding $\mathcal{A}$ as $(1 \pm \epsilon)$-isometric if for all $p, q \in M$:

$$(1-\epsilon)D_G(p,q) \leq D_G(\mathcal{A}(p),\mathcal{A}(q)) \leq (1+\epsilon)D_G(p,q),$$

where D_G denotes the geodesic distance. We would like to know i) can one come up with an embedding algorithm $\mathcal{A}$ that achieves $(1 \pm \epsilon)$-isometry for *all* points in M? ii) how much can one reduce the target dimension d and still have $(1 \pm \epsilon)$-isometry? and, iii) what kinds of restrictions (if any) does one need on M and X?

Since $\mathcal{A}$ only gets to access a finite size sample X from the underlying non-linear manifold M, it is essential to assume certain amount of curvature regularity on M. Niyogi et al. (2006) provide a nice characterization of manifold curvature via a notion of manifold condition number that will be useful throughout the text (details later).

Perhaps the first algorithmic result for embedding a general n-dimensional manifold is due to Baraniuk and Wakin (2007). They show that an orthogonal linear projection of a well-conditioned n-dimensional manifold $M \subset \mathbb{R}^D$ into a sufficiently high dimensional *random* subspace is enough to approximately preserve *all* pairwise geodesic distances. To get $(1 \pm \epsilon)$-isometry, they show that a target dimension d of size about $O\left(\frac{n}{\epsilon^2}\log\frac{VD}{\tau}\right)$ is sufficient, where V is the n-dimensional volume of the manifold and τ is the manifold's curvature condition number. This result was sharpened by Clarkson (2007) and Verma (2011) by completely removing the dependence on ambient dimension D and partially substituting the curvature-condition τ with more average-case manifold properties. In either case, the $1/\epsilon^2$ dependence is troublesome: if we want an embedding with all distances within 99% of the original (i.e., $\epsilon = 0.01$), the bounds require the target dimension to be at least 10,000!

Our Contributions. In this work, we give two algorithms that achieve $(1 \pm \epsilon)$-isometry where the dimension of the target space is *independent* of the isometry constant ϵ. As one expects, this dependency shows up in the sampling density (i.e. the size of X) required to compute the embedding. The first algorithm we propose is simple and easy to implement but embeds the given n-dimensional manifold in $\tilde{O}(2^{cn})$ dimensions[1] (where c is an absolute constant). The second algorithm, a variation on the first, focuses on minimizing the target dimension. It is computationally more involved and serves a more theoretical purpose: it shows that one can embed the manifold in just $\tilde{O}(n)$ dimensions.

We would like to highlight that both our proposed algorithms work for a very general class of well-conditioned manifolds. There is no requirement that the underlying manifold is connected, or is globally isometric (or even globally diffeomorphic) to some subset of $\mathbb{R}^n$ as is frequently assumed by several manifold embedding algorithms. In addition, unlike spectrum-based embedding algorithms in the literature, our algorithms yield an explicit embedding that cleanly embeds out-of-sample data points, and provide isometry guarantees over the entire manifold (not just the input samples).

2. Isometrically Embedding n-Dimensional Manifolds: Intuition

Given an underlying n-dimensional manifold $M \subset \mathbb{R}^D$, we shall use ideas from Nash's embedding (Nash, 1954) to develop our algorithms. To ease the burden of finding a $(1 \pm \epsilon)$-isometric embedding directly, our proposed algorithm will be divided in two stages. The first stage will embed

1. $\tilde{O}(\cdot)$ notation suppresses the logarithmic dependence on volume and curvature-condition terms.

M in a lower dimensional space without having to worry about preserving any distances. Since interpoint distances will potentially be distorted by the first stage, the second stage will focus on adjusting these distances by applying a series of corrections. The combined effect of both stages is a distance preserving embedding of M in lower dimensions. We now describe the stages in detail.

Embedding Stage. We shall use the random projection result by Clarkson (2007) (with ϵ set to a constant) to embed M into $d = \tilde{O}(n)$ dimensions. This gives an easy one-to-one low-dimensional embedding that doesn't collapse interpoint distances. Note that a projection does contract interpoint distances; by appropriately scaling the random projection, we can make sure that the distances are contracted by at most a constant amount, with high probability.

Correction Stage. Since the random projection can contract different parts of the manifold by different amounts, we will apply several corrections—each corresponding to a different local region—to stretch-out and restore the local distances.

To understand a single correction better, we can consider its effect on a small section of the contracted manifold. Since manifolds are locally linear, the section effectively looks like a contracted n-dimensional affine space. Our correction map needs to restore distances over this n-flat.

For simplicity, let's temporarily assume $n = 1$ (this corresponds to a 1-dimensional manifold), and let $t \in [0, 1]$ parameterize a unit-length segment of the contracted 1-flat. Suppose we want to stretch the segment by a factor of $L \geq 1$ to restore the contracted distances. How can we accomplish this?

Perhaps the simplest thing to do is apply a linear correction mapping $\Psi : t \mapsto Lt$. While this mapping works well for individual local corrections, it turns out that this mapping makes it difficult to control interference caused by different local corrections with overlapping localities.

We instead use extra coordinates and apply a *non-linear* map $\Psi : t \mapsto (t, \sin(Ct), \cos(Ct))$, where C controls the stretch-size. Note that such a spiral map stretches the length of the tangent vectors by a factor of $\sqrt{1 + C^2}$, since $\|\Psi'\| = \|d\Psi/dt\| = \|(1, C\cos(Ct), -C\sin(Ct))\| = \sqrt{1 + C^2}$. Now since the geodesic distance between any two points p and q on a manifold is given by the expression $\int \|\gamma'(s)\| ds$, where γ is a parameterization of the geodesic curve between points p and q (that is, length of a curve is infinitesimal sum of the length of tangent vectors along its path), Ψ stretches the interpoint geodesic distances by a factor of $\sqrt{1 + C^2}$ on the resultant surface as well. Thus, to stretch the distances by a factor of L, we can set $C := \sqrt{L^2 - 1}$.

Now generalizing this to a local region for an arbitrary n-dimensional manifold, let $U := [u^1, \ldots, u^n]$ be a $d \times n$ matrix whose columns form an orthonormal basis for the (local) contracted n-flat in the embedded space $\mathbb{R}^d$ and let $\sigma^1, \ldots, \sigma^n$ be the corresponding shrinkages along the n orthogonal directions. Then one can consider applying an n-dimensional analog of the spiral mapping: $\Psi : t \mapsto (t, \Psi_{\sin}(t), \Psi_{\cos}(t))$, where $t \in \mathbb{R}^d$

$$\Psi_{\sin}(t) := (\sin((Ct)_1), \ldots, \sin((Ct)_n)), \text{ and}$$
$$\Psi_{\cos}(t) := (\cos((Ct)_1), \ldots, \cos((Ct)_n)).$$

Here C is an $n \times d$ "correction" matrix that encodes how much of the surface needs to stretch in the various orthogonal directions. It turns out that if one sets C to be the matrix SU^{T}, where S is a diagonal matrix with entry $S_{ii} := \sqrt{(1/\sigma^i)^2 - 1}$ (recall that σ^i was the shrinkage along direction u^i), then the correction Ψ precisely restores the shrinkages along the n orthonormal directions on the resultant surface.

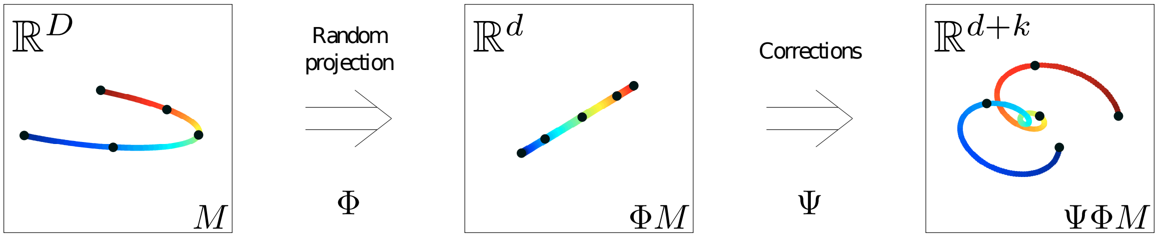

Figure 1: A simple example demonstrating our embedding technique on a 1-dimensional manifold. Left: The original 1-dimensional manifold in some high dimensional space. Middle: A low dimensional mapping of the original manifold via, say, a linear projection onto the vertical plane. Different parts of the manifold are contracted by different amounts – distances at the tail-ends are contracted more than the distances in the middle. Right: Final embedding after applying a series of spiralling corrections. Small size spirals are applied to regions with small distortion (middle), large spirals are applied to regions with large distortions (tail-ends). Resulting embedding is isometric (i.e., geodesic distance preserving) to the original manifold.

This takes care of the local regions individually. Now, globally, since different parts of the contracted manifold need to be stretched by different amounts, we localize the effect of the individual Ψ's to a small enough neighborhood by applying a specific kind of kernel function known as the "bump" function in the analysis literature, given by

$$\lambda_x(t) := \mathbf{1}_{\{\|t-x\|<\rho\}} \cdot e^{-1/(1-(\|t-x\|/\rho)^2)}.$$

Applying different Ψ's at different parts of the manifold has an aggregate effect of creating an approximate isometric embedding.

We now have a basic outline of our algorithm. Let M be an n-dimensional manifold in $\mathbb{R}^D$. We first find a contraction of M in $d = \tilde{O}(n)$ dimensions via a random projection. This embeds the manifold in low dimensions but distorts the interpoint geodesic distances. We estimate the distortion at different regions of the projected manifold by comparing a sample from M (i.e. X) with its projection. We then perform a series of corrections—each applied locally—to adjust the lengths in the local neighborhoods. We will conclude that restoring the lengths in all neighborhoods yields a globally consistent approximately isometric embedding of M. See also Figure 1.

As briefly mentioned earlier, a key issue in preserving geodesic distances across points in different neighborhoods is reconciling the interference between different corrections with overlapping localities. Based on exactly *how* we apply these different local Ψ's gives rise to our two algorithms. For the first algorithm, we shall allocate a fresh set of coordinates for each correction Ψ so that the different corrections don't interfere with each other. Since a local region of an n-dimensional manifold can potentially have up to $O(2^{cn})$ overlapping regions, we shall require $O(2^{cn})$ additional coordinates to apply the corrections, making the final embedding dimension of $\tilde{O}(2^{cn})$ (where c is an absolute constant). For the second algorithm, we will follow Nash's technique (Nash, 1954) more closely and apply Ψ maps iteratively in the same embedding space without the use of extra coordinates. At each iteration we need to compute a pair of vectors *normal* to the embedded manifold. Since locally the manifold spreads across its tangent space, these normals indicate the locally empty regions in the embedded space. Applying the local Ψ correction in the direction of these normals gives a way to mitigate the interference between different Ψ's. Since we don't use extra coordinates, the final embedding dimension remains $\tilde{O}(n)$.

3. Preliminaries

Let M be a smooth, n-dimensional compact Riemannian submanifold of $\mathbb{R}^D$. We will frequently refer to such a manifold as an n-manifold. Since we will be working with samples from M, we need to ensure certain amount of curvature regularity. Here we borrow the notation from Niyogi et al. (2006) about the condition number of M.

Definition 1 (condition number (Niyogi et al., 2006)) *Let $M \subset \mathbb{R}^D$ be a compact Riemannian manifold. The condition number of M is $\frac{1}{\tau}$, if τ is the largest number such that the normals of length $r < \tau$ at any two distinct points $p, q \in M$ don't intersect.*

Throughout the text we shall assume that our given M has condition number $1/\tau$. We will use $D_G(p,q)$ to indicate the geodesic distance between points p and q where the underlying manifold is understood from the context, and $\|p-q\|$ to indicate the Euclidean distance between points p and q where the ambient space is understood from the context.

To correctly estimate the distortion induced by the initial contraction mapping, we additionally require a high-resolution covering of our manifold.

Definition 2 (bounded manifold cover) *Let $M \subset \mathbb{R}^D$ be a Riemannian n-manifold. We call $X \subset M$ an α-bounded (ρ, δ)-cover of M if for all $p \in M$ and ρ-neighborhood $X_p := \{x \in X : \|x-p\| < \rho\}$ around p, we have*

- *exist $x_0, \ldots, x_n \in X_p$ such that $\left|\frac{x_i - x_0}{\|x_i - x_0\|} \cdot \frac{x_j - x_0}{\|x_j - x_0\|}\right| \leq \frac{1}{2n}$, for $i \neq j$. (covering criterion)*
- *$|X_p| \leq \alpha$. (local boundedness criterion)*
- *exists point $x \in X_p$ such that $\|x - p\| \leq \rho/2$. (point representation criterion)*
- *for any $n+1$ points in X_p satisfying the covering criterion, let $\hat{T}_p$ denote the n-dimensional affine space passing through them. Then, for any unit vector $\hat{v}$ in $\hat{T}_p$, we have $\left|\hat{v} \cdot \frac{v}{\|v\|}\right| \geq 1 - \delta$, where v is the projection of $\hat{v}$ onto the tangent space of M at p. (tangent space approximation criterion)*

Remark 3 *Given an n-manifold M with condition number $1/\tau$, and some $0 < \delta \leq 1$, if $\rho \leq \tau\delta/3\sqrt{2}n$, then there exists a 2^{10n+1}-bounded (ρ, δ)-cover of M (see Appendix B).*

4. Algorithms

Inputs. We assume the following quantities are given. (i) n: the intrinsic dimension of M, (ii) $1/\tau$: the condition number of M, (iii) X: an α-bounded (ρ, δ)-cover of M, (iv) ρ: the ρ parameter of the cover.

Notation. Let ϕ be a random orthogonal projection map that maps points from $\mathbb{R}^D$ into a random subspace of dimension d ($n \leq d \leq D$). We will have d to be about $\tilde{O}(n)$. Set $\Phi := (2/3)(\sqrt{D/d})\phi$

as a scaled version of ϕ. Since Φ is linear, Φ can also be represented as a $d \times D$ matrix. In our discussion below we will use the function notation and the matrix notation interchangeably, that is, for any $p \in \mathbb{R}^D$, we will use the notation $\Phi(p)$ (applying function Φ to p) and the notation Φp (matrix-vector multiplication) interchangeably.

For any $x \in X$, let $x_0, \ldots, x_n$ be $n+1$ points from the set $\{x' \in X : \|x - x'\| < \rho\}$ such that $\left| \frac{x_i - x_0}{\|x_i - x_0\|} \cdot \frac{x_j - x_0}{\|x_j - x_0\|} \right| \leq 1/2n$, for $i \neq j$ (cf. Definition 2). Let F_x be the $D \times n$ matrix whose column vectors form some orthonormal basis of the n-dimensional subspace spanned by the vectors $\{x_i - x_0\}_{i \in [n]}$. Note that F_x serves as a good approximation to the tangent spaces at different points in the neighborhood of $x \in M \subset \mathbb{R}^D$.

Estimating local contractions. We estimate the contraction caused by Φ at a small enough neighborhood of M containing the point $x \in X$, by computing the "thin" Singular Value Decomposition (SVD) $U_x \Sigma_x V_x^{\mathsf{T}}$ of the $d \times n$ matrix ΦF_x and representing the singular values in the conventional descending order. That is, $\Phi F_x = U_x \Sigma_x V_x^{\mathsf{T}}$, and since ΦF_x is a tall matrix ($n \leq d$), we know that the bottom $d - n$ singular values are zero. Thus, we only consider the top n (of d) left singular vectors in the SVD (so, U_x is $d \times n$, Σ_x is $n \times n$, and V_x is $n \times n$) and $\sigma_x^1 \geq \sigma_x^2 \geq \ldots \geq \sigma_x^n$ where σ_x^i is the i^{th} largest singular value.

Observe that the singular values $\sigma_x^1, \ldots, \sigma_x^n$ are precisely the distortion amounts in the directions $u_x^1, \ldots, u_x^n$ at $\Phi(x) \in \mathbb{R}^d$ ($[u_x^1, \ldots, u_x^n] = U_x$) when we apply Φ. To see this, consider the direction $w^i := F_x v_x^i$ in the column-span of F_x ($[v_x^1, \ldots, v_x^n] = V_x$). Then $\Phi w^i = (\Phi F_x) v_x^i = \sigma_x^i u_x^i$, which can be interpreted as: Φ maps the vector w^i in the subspace F_x (in $\mathbb{R}^D$) to the vector u_x^i (in $\mathbb{R}^d$) with the scaling of σ_x^i.

Note that if $0 < \sigma_x^i \leq 1$ (for all $x \in X$ and $1 \leq i \leq n$), we can define an $n \times d$ correction matrix (corresponding to each $x \in X$) $C^x := S_x U_x^{\mathsf{T}}$, where S_x is a diagonal matrix with $(S_x)_{ii} := \sqrt{(1/\sigma_x^i)^2 - 1}$. We can also write S_x as $(\Sigma_x^{-2} - I)^{1/2}$. The correction matrix C^x will have an effect of stretching the direction u_x^i by the amount $(S_x)_{ii}$ and killing any direction v that is orthogonal to (column-span of) U_x.

Algorithm 1 Compute Corrections C^x's

for $x \in X$ (in any order) **do**
 Let $x_0, \ldots, x_n \in \{x' \in X : \|x' - x\| < \rho\}$ be such that $\left| \frac{x_i - x_0}{\|x_i - x_0\|} \cdot \frac{x_j - x_0}{\|x_j - x_0\|} \right| \leq 1/2n$ (for $i \neq j$).
 Let F_x be a $D \times n$ matrix whose columns form an orthonormal basis of the n-dimensional span of the vectors $\{x_i - x_0\}_{i \in [n]}$.
 Let $U_x \Sigma_x V_x^{\mathsf{T}}$ be the "thin" SVD of ΦF_x.
 Set $C^x := (\Sigma_x^{-2} - I)^{1/2} U_x^{\mathsf{T}}$.
end for

Algorithm 2 Embedding Technique I

Preprocessing: Partition the given covering X into disjoint subsets such that no subset contains points that are too close to each other. Let $x_1, \ldots, x_{|X|}$ be the points in X in some arbitrary but fixed order. We partition as follows:

1: Initialize $X^{(1)}, \ldots, X^{(K)}$ as empty sets.
2: **for** $x_i \in X$ (in any fixed order) **do**
3: Let j be the smallest positive integer such that x_i is not within distance 2ρ of any element in $X^{(j)}$. That is, the smallest j such that for all $x \in X^{(j)}$, $\|x - x_i\| \geq 2\rho$.
4: $X^{(j)} \leftarrow X^{(j)} \cup \{x_i\}$.
5: **end for**

The Embedding: For *any* $p \in M \subset \mathbb{R}^D$, embed it in $\mathbb{R}^{d+2nK}$ as follows:

1: Let $t = \Phi(p)$. [Φ is the scaled random projection from $\mathbb{R}^D$ to R^d]
2: Define $\Psi(t) := (t, \Psi_{1,\text{sin}}(t), \Psi_{1,\text{cos}}(t), \ldots, \Psi_{K,\text{sin}}(t), \Psi_{K,\text{cos}}(t))$ where

$$\Psi_{j,\text{sin}}(t) := (\psi^1_{j,\text{sin}}(t), \ldots, \psi^n_{j,\text{sin}}(t)),$$
$$\Psi_{j,\text{cos}}(t) := (\psi^1_{j,\text{cos}}(t), \ldots, \psi^n_{j,\text{cos}}(t)).$$

The individual terms are given by

$$\begin{array}{ll} \psi^i_{j,\text{sin}}(t) := \sum_{x \in X^{(j)}} \left(\sqrt{\Lambda_{\Phi(x)}(t)}/\omega\right) \sin(\omega (C^x t)_i) & i = 1, \ldots, n; \\ \psi^i_{j,\text{cos}}(t) := \sum_{x \in X^{(j)}} \left(\sqrt{\Lambda_{\Phi(x)}(t)}/\omega\right) \cos(\omega (C^x t)_i) & j = 1, \ldots, K \end{array}$$

where $\Lambda_a(b) = \frac{\lambda_a(b)}{\sum_{q \in X} \lambda_{\Phi(q)}(b)}$.
3: **return** $\Psi(t)$ as the embedding of p in $\mathbb{R}^{d+2nK}$.

A few remarks are in order.

Remark 4 *The goal of the Preprocessing step is to identify samples from X that can have overlapping (ρ-size) local neighborhoods. The partitioning procedure described above ensures that corrections associated with nearby neighborhoods are applied in separate coordinates to minimize interference.*

Remark 5 *If $\rho \leq \tau/4$, the number of subsets (i.e. K) produced by the Preprocessing is at most $\alpha 2^{cn}$ for an α-bounded (ρ, δ) cover X of M (where $c \leq 4$). See Appendix C for details.*

Remark 6 *The function Λ acts as a (normalized) localizing kernel that helps in localizing the effects of the spiralling corrections (discussed in detail in Section 5.2).*

Remark 7 *$\omega > 0$ is a free parameter that controls the interference due to overlapping local corrections.*

Algorithm 3 Embedding Technique II

The Embedding: Let $x_1, \ldots, x_{|X|}$ be the points in X in some arbitrary but fixed order. For *any* point $p \in M \subset \mathbb{R}^D$, we embed it in $\mathbb{R}^{2d+3}$ by:

1: Let $t = \Phi(p)$. [Φ is the scaled random projection from $\mathbb{R}^D$ to R^d]
2: Define $\Psi_{0,n}(t) := (t, \underbrace{0, \ldots, 0}_{d+3})$. [Extension needed to efficiently find the normal vectors]
3: **for** $i = 1, \ldots, |X|$ **do**
4: Define $\Psi_{i,0} := \Psi_{i-1,n}$.
5: **for** $j = 1, \ldots, n$ **do**
6: Let $\eta_{i,j}(t)$ and $\nu_{i,j}(t)$ be two mutually orthogonal unit vectors normal to $\Psi_{i,j-1}(M)$ at $\Psi_{i,j-1}(t)$.
7: Define

$$\Psi_{i,j}(t) := \Psi_{i,j-1}(t) + \eta_{i,j}(t)\left(\frac{\sqrt{\Lambda_{\Phi(x_i)}(t)}}{\omega_{i,j}}\right)\sin(\omega_{i,j}(C^{x_i}t)_j) + \nu_{i,j}(t)\left(\frac{\sqrt{\Lambda_{\Phi(x_i)}(t)}}{\omega_{i,j}}\right)\cos(\omega_{i,j}(C^{x_i}t)_j),$$

where $\Lambda_a(b) = \frac{\lambda_a(b)}{\sum_{q \in X} \lambda_{\Phi(q)}(b)}$.
8: **end for**
9: **end for**
10: **return** $\Psi_{|X|,n}(t)$ as the embedding of p into $\mathbb{R}^{2d+3}$.

Remark 8 *The function Λ, and the free parameters $\omega_{i,j}$ (one for each i, j iteration) have roles similar to those in Embedding I.*

Remark 9 *The success of Embedding II depends upon finding a pair of normal unit vectors η and ν in each iteration (Step 6); we discuss how to approximate these in Appendix E.*

We shall see that for appropriate choice of d, ρ, δ and ω (or $\omega_{i,j}$), our algorithms yield an approximate isometric embedding of M.

4.1. Main Result

Theorem 10 *Let $M \subset \mathbb{R}^D$ be a compact n-manifold with volume V and condition number $1/\tau$ (as above). Let $d = \Omega\left(n + \ln(V/\tau^n)\right)$ be the target dimension of the initial random projection mapping such that $d \leq D$. For any $0 < \epsilon \leq 1$, let $\rho \leq (\tau d/D)(\epsilon/350)^2$, $\delta \leq (d/D)(\epsilon/250)^2$, and let $X \subset M$ be an α-bounded (ρ, δ)-cover of M. Now, let*

i. $N_{\mathrm{I}} \subset \mathbb{R}^{d+2\alpha n 2^{cn}}$ *be the embedding of M returned by Algorithm* I *(where $c \leq 4$),*

ii. $N_{\mathrm{II}} \subset \mathbb{R}^{2d+3}$ *be the embedding of M returned by Algorithm* II.

Then, with probability at least $1 - 1/poly(n)$ over the choice of the initial random projection, for all $p, q \in M$ and their corresponding mappings $p_{\mathrm{I}}, q_{\mathrm{I}} \in N_{\mathrm{I}}$ and $p_{\mathrm{II}}, q_{\mathrm{II}} \in N_{\mathrm{II}}$, we have

i. $(1-\epsilon)D_G(p,q) \leq D_G(p_{\mathrm{I}}, q_{\mathrm{I}}) \leq (1+\epsilon)D_G(p,q)$,

ii. $(1-\epsilon)D_G(p,q) \leq D_G(p_{\mathrm{II}}, q_{\mathrm{II}}) \leq (1+\epsilon)D_G(p,q)$.

5. Proof

Our goal is to show that the two proposed embeddings approximately preserve the lengths of all geodesic curves. Now, since the length of any given curve $\gamma : [a, b] \to M$ is given by $\int_a^b \|\gamma'(s)\| ds$, it suffices to ensure that our embeddings distort the length of the tangent vectors at any point $p \in M$ by at most a factor of $1 \pm \epsilon$.

In what follows, we will fix an arbitrary point $p \in M$ and a tangent vector $v \in T_pM$ and analyze how the various steps of the algorithm modify the length of v. Let Φ be the initial (scaled) random projection map (from $\mathbb{R}^D$ to $\mathbb{R}^d$) that may contract distances on M by various amounts, and let Ψ be the subsequent correction map that attempts to restore these distances (as defined in Step 2 for Embedding I or as a sequence of maps in Step 7 for Embedding II). To get a firm footing for our analysis, we need to study how Φ and Ψ modify the tangent vector v. It is well known from differential geometry (see e.g. Chapter 0 of do Carmo, 1992) that for any smooth map $F : M \to N$ that maps a manifold $M \subset \mathbb{R}^k$ to a manifold $N \subset \mathbb{R}^{k'}$, there exists a linear map $(DF)_p : T_pM \to T_{F(p)}N$, known as the derivative map or the pushforward (at p), that maps tangent vectors incident at p in M to tangent vectors incident at $F(p)$ in N. We shall thus study how the respective pushforwards ($(D\Phi)_p$ and $(D\Psi)_{\Phi(p)}$) affect the length of v. Observe that since pushforward maps are linear, without loss of generality we can assume that v has unit length. Figure 2 provides a quick sketch of our two stage mapping with various quantities of interest.

A quick roadmap for the proof. In the next three sections, we take a brief detour to study the effects of applying Φ, applying Ψ for Algorithm I, and applying Ψ for Algorithm II separately. This will give us the necessary tools to analyze the combined effect of applying $\Psi \circ \Phi$ on v (Section 5.4). We defer the proofs of all the supporting lemmas to Appendix D.

5.1. Effects of Applying Φ

Lemma 11 *Let $M \subset \mathbb{R}^D$ be a smooth n-manifold (as defined above) with volume V and condition number $1/\tau$. Let R be a random projection matrix that maps points from $\mathbb{R}^D$ into a random subspace of dimension d ($d \leq D$). Define $\Phi := (2/3)(\sqrt{D/d})R$ as a scaled projection mapping. If $d = \Omega(n + \ln(V/\tau^n))$, then with probability at least $1 - 1/poly(n)$ over the choice of the random projection matrix, we have*

(a) For all $p \in M$ and all tangent vectors $v \in T_pM$, $(1/2)\|v\| \leq \|(D\Phi)_p(v)\| \leq (5/6)\|v\|$.

(b) For all $p, q \in M$, $(1/2)\|p - q\| \leq \|\Phi p - \Phi q\| \leq (5/6)\|p - q\|$.

(c) For all $x \in \mathbb{R}^D$, $\|\Phi x\| \leq (2/3)(\sqrt{D/d})\|x\|$.

In what follows, we assume that Φ is such a scaled random projection map. Then, a bound on the length of tangent vectors also gives us a bound on the spectrum of ΦF_x (recall the definition of F_x from Section 4).

Corollary 12 *Let Φ, F_x and n be as described above (recall that $x \in X$ that forms a bounded (ρ, δ)-cover of M). Let σ_x^i represent the i^{th} largest singular value of the matrix ΦF_x. Then, for $\delta \leq d/32D$, we have $1/4 \leq \sigma_x^n \leq \sigma_x^1 \leq 1$ (for all $x \in X$).*

We will be using these facts in our discussion below in Section 5.4.

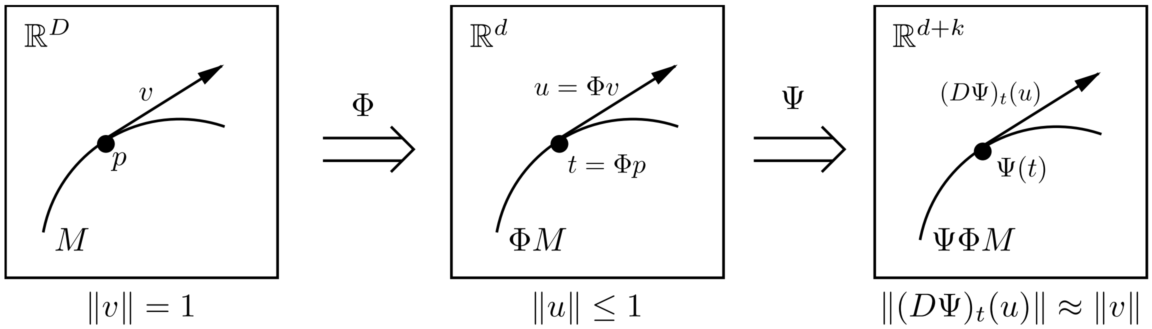

Figure 2: Two stage mapping of our embedding technique. Left: Underlying manifold $M \subset \mathbb{R}^D$ with the quantities of interest – a fixed point p and a fixed unit-vector v tangent to M at p. Center: A (scaled) linear projection of M into a random subspace of d dimensions. The point p maps to Φp and the tangent vector v maps to $u := (D\Phi)_p(v) = \Phi v$. The length of v contracts to $\|u\|$. Right: Correction of ΦM via a non-linear mapping Ψ into $\mathbb{R}^{d+k}$. We have $k = O(\alpha 2^{cn})$ for correction technique I, and $k = d + 3$ for correction technique II (see also Section 4). Our goal is to show that Ψ stretches length of contracted v (i.e. u) back to approximately its original length.

5.2. Effects of Applying Ψ (Algorithm I)

Embedding Technique I restores the contraction induced by Φ by applying a non-linear map $\Psi(t) := (t, \Psi_{1,\text{sin}}(t), \Psi_{1,\text{cos}}(t), \ldots, \Psi_{K,\text{sin}}(t), \Psi_{K,\text{cos}}(t))$ (recall that K is the number of subsets we decompose X into, as described in the Preprocessing step of Embedding I in Section 4), with $\Psi_{j,\text{sin}}(t) := (\psi^1_{j,\text{sin}}(t), \ldots, \psi^n_{j,\text{sin}}(t))$ and $\Psi_{j,\text{cos}}(t) := (\psi^1_{j,\text{cos}}(t), \ldots, \psi^n_{j,\text{cos}}(t))$. The individual terms are given as

$$\begin{array}{ll} \psi^i_{j,\text{sin}}(t) := \sum_{x \in X^{(j)}} \left(\sqrt{\Lambda_{\Phi(x)}(t)}/\omega\right) \sin(\omega(C^x t)_i) & \\ \psi^i_{j,\text{cos}}(t) := \sum_{x \in X^{(j)}} \left(\sqrt{\Lambda_{\Phi(x)}(t)}/\omega\right) \cos(\omega(C^x t)_i) & \end{array} \quad i = 1, \ldots, n; j = 1, \ldots, K,$$

where C^x's are the correction amounts for different locations x on the manifold, $\omega > 0$ controls the frequency (cf. Section 4), and $\Lambda_{\Phi(x)}(t)$ is defined to be $\lambda_{\Phi(x)}(t) / \sum_{q \in X} \lambda_{\Phi(q)}(t)$, with

$$\lambda_{\Phi(x)}(t) := \begin{cases} \exp(-1/(1 - \|t - \Phi(x)\|^2/\rho^2)) & \text{if } \|t - \Phi(x)\| < \rho. \\ 0 & \text{otherwise.} \end{cases}$$

λ is a classic example of a *bump function*. It is a smooth function with compact support. Its applicability arises from the fact that it can be made "to specifications". That is, it can be made to vanish outside any interval of our choice. Here we exploit this property to localize the effect of our corrections. The normalization of λ (the function Λ) creates the so-called smooth partition of unity that helps to vary smoothly between the corrections applied at different regions of M.

Since any tangent vector in $\mathbb{R}^d$ can be expressed in terms of the basis vectors, it suffices to study how $D\Psi$ acts on the standard basis $\{e^i\}$. Note that

$$(D\Psi)_t(e^i) = \left(\frac{dt}{dt^i}, \frac{d\Psi_{1,\text{sin}}(t)}{dt^i}, \frac{d\Psi_{1,\text{cos}}(t)}{dt^i}, \ldots, \frac{d\Psi_{K,\text{sin}}(t)}{dt^i}, \frac{d\Psi_{K,\text{cos}}(t)}{dt^i} \right) \Bigg|_t,$$

where

$$\begin{array}{ll} \frac{d\psi^k_{j,\sin}(t)}{dt^i} = \sum_{x\in X^{(j)}} \frac{1}{\omega}\Big(\sin(\omega(C^x t)_k)\frac{d\Lambda^{1/2}_{\Phi(x)}(t)}{dt^i}\Big) + \sqrt{\Lambda_{\Phi(x)}(t)}\cos(\omega(C^x t)_k)C^x_{k,i} & k\in[n];\ i\in[d] \\ \frac{d\psi^k_{j,\cos}(t)}{dt^i} = \sum_{x\in X^{(j)}} \frac{1}{\omega}\Big(\cos(\omega(C^x t)_k)\frac{d\Lambda^{1/2}_{\Phi(x)}(t)}{dt^i}\Big) - \sqrt{\Lambda_{\Phi(x)}(t)}\sin(\omega(C^x t)_k)C^x_{k,i} & j\in[K] \end{array}.$$

One can observe the advantage of having the term ω. By picking ω sufficiently large, we can make the first part of the expression sufficiently small. In particular,

Lemma 13 *Let t be any point in $\Phi(M)$ and u be any vector tagent to $\Phi(M)$ at t such that $\|u\| \leq 1$. Let ϵ be the isometry parameter chosen in Theorem 10. Pick $\omega \geq \Omega(n\alpha^2 9^n\sqrt{d}/\rho\epsilon)$, then for some ζ, such that $|\zeta| \leq \epsilon/2$,*

$$\|(D\Psi)_t(u)\|^2 \quad = \quad \|u\|^2 + \sum_{x\in X}\Lambda_{\Phi(x)}(t)\sum_{k=1}^{n}(C^x u)^2_k + \zeta. \tag{1}$$

We use this derivation to study the combined effect of $\Psi \circ \Phi$ on M in Section 5.4.

5.3. Effects of Applying Ψ (Algorithm II)

The goal of the second algorithm is to apply the spiralling corrections while using the coordinates more economically. We achieve this goal by applying them sequentially in the same embedding space (rather than simultaneously by making use of extra $2nK$ coordinates as done in Algorithm I), see also Nash (1954). Since all the corrections share the same coordinate space, one needs to keep track of a pair of normal vectors to prevent interference among the different local corrections. More specifically, $\Psi : \mathbb{R}^d \to \mathbb{R}^{2d+3}$ (in Algorithm II) is defined recursively as $\Psi := \Psi_{|X|,n}$ such that (see also Embedding II in Section 4)

$$\Psi_{i,j}(t) := \Psi_{i,j-1}(t) + \eta_{i,j}(t)\frac{\sqrt{\Lambda_{\Phi(x_i)}(t)}}{\omega_{i,j}}\sin(\omega_{i,j}(C^{x_i}t)_j) + \nu_{i,j}(t)\frac{\sqrt{\Lambda_{\Phi(x_i)}(t)}}{\omega_{i,j}}\cos(\omega_{i,j}(C^{x_i}t)_j),$$

where $\Psi_{i,0}(t) := \Psi_{i-1,n}(t)$, and the base function $\Psi_{0,n}(t)$ is given as $t \mapsto (t, \overbrace{0,\dots,0}^{d+3})$. $\eta_{i,j}(t)$ and $\nu_{i,j}(t)$ are mutually orthogonal unit vectors that are approximately normal to $\Psi_{i,j-1}(\Phi M)$ at $\Psi_{i,j-1}(t)$. We assume that the normals η and ν have the following properties:

- $|\eta_{i,j}(t)\cdot v| \leq \epsilon_0$ and $|\nu_{i,j}(t)\cdot v| \leq \epsilon_0$ for all unit-length v tangent to $\Psi_{i,j-1}(\Phi M)$ at $\Psi_{i,j-1}(t)$. (quality of normal approximation)
- For all $1 \leq l \leq d$, we have $\|d\eta_{i,j}(t)/dt^l\| \leq K_{i,j}$ and $\|d\nu_{i,j}(t)/dt^l\| \leq K_{i,j}$. (bounded directional derivatives)

We refer the reader to Section E for details on how to estimate such normals. Now, as before, by picking $\omega_{i,j}$ sufficiently large, we have the following.

Lemma 14 *Let t be any point in $\Phi(M)$ and u be any vector tangent to $\Phi(M)$ at t such that $\|u\| \leq 1$. Let ϵ be the isometry parameter chosen in Theorem 10. Pick $\omega_{i,j} \geq \Omega\big((K_{i,j} +$*

$(\alpha 9^n/\rho))(nd|X|)^2/\epsilon\big)$ *(recall that $K_{i,j}$ is the bound on the directional derivate of η and ν). If $\epsilon_0 \leq O\big(\epsilon/d(n|X|)^2\big)$ (recall that ϵ_0 is the quality of approximation of the normals η and ν), then we have for some ζ, such that $|\zeta| \leq \epsilon/2$,*

$$\|(D\Psi)_t(u)\|^2 = \|(D\Psi_{|X|,n})_t(u)\|^2 = \|u\|^2 + \sum_{i=1}^{|X|} \Lambda_{\Phi(x_i)}(t) \sum_{j=1}^{n} (C^{x_i}u)_j^2 + \zeta. \tag{2}$$

5.4. Combined Effect of $\Psi(\Phi(M))$

We can now analyze the aggregate effect of both our embeddings on the length of an arbitrary unit vector v tangent to M at p. Let $u := (D\Phi)_p(v) = \Phi v$ be the pushforward of v. Then $\|u\| \leq 1$ (cf. Lemma 11). See also Figure 2. Now, recalling that $D(\Psi \circ \Phi) = D\Psi \circ D\Phi$, and noting that pushforward maps are linear, we have $\|(D(\Psi \circ \Phi))_p(v)\|^2 = \big\|(D\Psi)_{\Phi(p)}(u)\big\|^2$. Thus, representing u as $\sum_i u_i e^i$ in ambient coordinates of $\mathbb{R}^d$, and using Eq. (1) (for Algorithm I) or Eq. (2) (for Algorithm II), we get

$$\big\|(D(\Psi \circ \Phi))_p(v)\big\|^2 \;=\; \big\|(D\Psi)_{\Phi(p)}(u)\big\|^2 = \|u\|^2 + \sum_{x \in X} \Lambda_{\Phi(x)}(\Phi(p)) \|C^x u\|^2 + \zeta,$$

where $|\zeta| \leq \epsilon/2$. We can give simple lower and upper bounds for the above expression by noting that $\Lambda_{\Phi(x)}$ is a localization function. Define $N_p := \{x \in X : \|\Phi(x) - \Phi(p)\| < \rho\}$ as the neighborhood around p (ρ as per the theorem statement). Then only the points in N_p contribute to above equation, since $\Lambda_{\Phi(x)}(\Phi(p)) = d\Lambda_{\Phi(x)}(\Phi(p))/dt^i = 0$ for $\|\Phi(x) - \Phi(p)\| \geq \rho$. Also note that for all $x \in N_p$, $\|x - p\| < 2\rho$ (cf. Lemma 11).

Let $x_M := \arg\max_{x \in N_p} \|C^x u\|^2$ and $x_m := \arg\min_{x \in N_p} \|C^x u\|^2$ are quantities that attain the maximum and the minimum respectively, then:

$$\|u\|^2 + \|C^{x_m}u\|^2 - \epsilon/2 \leq \|(D(\Psi \circ \Phi))_p(v)\|^2 \leq \|u\|^2 + \|C^{x_M}u\|^2 + \epsilon/2. \tag{3}$$

Notice that ideally we would like to have the correction factor "$C^p u$" in Eq. (3) since that would give the perfect stretch around the point p. But what about correction $C^x u$ for closeby x's? The following lemma helps us continue in this situation.

Lemma 15 *Let p, v, u be as above. For any $x \in N_p \subset X$, let C^x and F_x also be as discussed above (recall that $\|p - x\| < 2\rho$, and $X \subset M$ forms a bounded (ρ, δ)-cover of the fixed underlying manifold M with condition number $1/\tau$). Define $\xi := (4\rho/\tau) + \delta + 4\sqrt{\rho\delta/\tau}$. If $\rho \leq \tau/4$ and $\delta \leq d/32D$, then*

$$1 - \|u\|^2 - 40 \cdot \max\big\{\sqrt{\xi D/d}, \xi D/d\big\} \leq \|C^x u\|^2 \leq 1 - \|u\|^2 + 51 \cdot \max\big\{\sqrt{\xi D/d}, \xi D/d\big\}.$$

Note that we chose $\rho \leq (\tau d/D)(\epsilon/350)^2$ and $\delta \leq (d/D)(\epsilon/250)^2$ (cf. theorem statement). Thus, combining Eq. (3) and Lemma 15, we get (recall $\|v\| = 1$)

$$(1 - \epsilon)\|v\|^2 \leq \|(D(\Psi \circ \Phi))_p(v)\|^2 \leq (1 + \epsilon)\|v\|^2.$$

Since the choice of the vector v and the point p was arbitrary, our embedding approximately preserves the tangent vector lengths throughout the embedded manifold uniformly.

Acknowledgements. The author thanks Sanjoy Dasgupta for his guidance throughout the project. This research was supported in part by NSF Grant CNS-0932403.

References

R. Baraniuk and M. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 2007.

M. Bernstein, V. de Silva, J. Langford, and J. Tenebaum. Graph approximations to geodesics on embedded manifolds. *Techical Report*, 2000.

K. Clarkson. Tighter bounds for random projections of manifolds. *Comp. Geometry*, 2007.

S. Dasgupta and Y. Freund. Random projection trees and low dimensional manifolds. *ACM Symposium on Theory of Computing*, 2008.

M. do Carmo. *Riemannian Geometry*. Birkhauser, 1992.

Q. Han and J. Hong. *Isometric embedding of Riemannian manifolds in Euclidean spaces*. American Mathematical Society, 2006.

W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Conf. in Modern Analysis and Probability*, pages 189–206, 1984.

J. Milnor. *Topology from the differential viewpoint*. Univ. of Virginia Press, 1972.

J. Nash. C^1 isometric imbeddings. *Annals of Mathematics*, 60(3):383–396, 1954.

P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Disc. Computational Geometry*, 2006.

J. Tenebaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2000.

N. Verma. A note on random projections for preserving paths on a manifold. *UC San Diego, Tech. Report CS2011-0971*, 2011.

H. Whitney. Differentiable manifolds. *Annals of Mathematics*, 37:645–680, 1936.

Appendix A. Properties of a Well-conditioned Manifold

Throughout this section we will assume that M is a compact submanifold of $\mathbb{R}^D$ of dimension n, and condition number $1/\tau$. The following are some properties of such a manifold that would be useful throughout the text.

Lemma 16 (relating closeby tangent vectors – implicit in the proof of Proposition 6.2 Niyogi et al. (2006)) *Pick any two (path-connected) points $p, q \in M$. Let $u \in T_pM$ be a unit length tangent vector and $v \in T_qM$ be its parallel transport along the (shortest) geodesic path to q. Then*[2]*, i)* $u \cdot v \geq 1 - D_G(p,q)/\tau$, *ii)* $\|u - v\| \leq \sqrt{2D_G(p,q)/\tau}$.

2. Technically, it is not possible to directly compare two vectors that reside in different tangent spaces. However, since we only deal with manifolds that are immersed in some ambient space, we can treat the tangent spaces as n-dimensional affine subspaces. We can thus parallel translate the vectors to the origin of the ambient space, and do the necessary comparison (such as take the dot product, etc.). We will make a similar abuse of notation for any calculation that uses vectors from different affine subspaces to mean to first translate the vectors and then perform the necessary calculation.

Lemma 17 (relating geodesic distances to ambient distances – Proposition 6.3 of Niyogi et al. (2006)) *If $p, q \in M$ such that $\|p - q\| \leq \tau/2$, then $D_G(p, q) \leq \tau(1 - \sqrt{1 - 2\|p - q\|/\tau}) \leq 2\|p - q\|$.*

Lemma 18 (projection of a section of a manifold onto the tangent space) *Pick any $p \in M$ and define $M_{p,r} := \{q \in M : \|q - p\| \leq r\}$. Let f denote the orthogonal linear projection of $M_{p,r}$ onto the tangent space T_pM. Then, for any $r \leq \tau/2$*

(i) the map $f : M_{p,r} \to T_pM$ is one-to-one. (see Lemma 5.4 of Niyogi et al. (2006))

(ii) for any $x, y \in M_{p,r}$, $\|f(x) - f(y)\|^2 \geq (1 - (r/\tau)^2) \cdot \|x - y\|^2$. (implicit in the proof of Lemma 5.3 of Niyogi et al. (2006))

Lemma 19 (coverings of a section of a manifold) *Pick any $p \in M$ and define $M_{p,r} := \{q \in M : \|q - p\| \leq r\}$. If $r \leq \tau/2$, then there exists $C \subset M_{p,r}$ of size at most 9^n with the property: for any $p' \in M_{p,r}$, exists $c \in C$ such that $\|p' - c\| \leq r/2$.*

Proof The proof closely follows the arguments presented in the proof of Theorem 22 of Dasgupta and Freund (2008).

For $r \leq \tau/2$, note that $M_{p,r} \subset \mathbb{R}^D$ is (path-)connected. Let f denote the projection of $M_{p,r}$ onto $T_pM \cong \mathbb{R}^n$. Quickly note that f is one-to-one (see Lemma 18(i)). Then, $f(M_{p,r}) \subset \mathbb{R}^n$ is contained in an n-dimensional ball of radius r. By standard volume arguments, $f(M_{p,r})$ can be covered by at most 9^n balls of radius $r/4$. WLOG we can assume that the centers of these covering balls are in $f(M_{p,r})$. Now, noting that the inverse image of each of these covering balls (in $\mathbb{R}^n$) is contained in a D-dimensional ball of radius $r/2$ (see Lemma 18(ii)) finishes the proof. ■

Lemma 20 (relating closeby manifold points to tangent vectors) *Pick any point $p \in M$ and let $q \in M$ (distinct from p) be such that $D_G(p, q) \leq \tau$. Let $v \in T_pM$ be the projection of the vector $q - p$ onto T_pM. Then, i) $\left|\frac{v}{\|v\|} \cdot \frac{q-p}{\|q-p\|}\right| \geq 1 - (D_G(p, q)/2\tau)^2$, ii) $\left\|\frac{v}{\|v\|} - \frac{q-p}{\|q-p\|}\right\| \leq D_G(p, q)/\tau\sqrt{2}$.*

Proof If vectors v and $q - p$ are in the same direction, we are done. Otherwise, consider the plane spanned by vectors v and $q - p$. Then since M has condition number $1/\tau$, we know that the point q cannot lie within any τ-ball tangent to M at p (see Figure 3). Consider such a τ-ball (with center c) whose center is closest to q and let q' be the point on the surface of the ball which subtends the same angle ($\angle pcq'$) as the angle formed by q ($\angle pcq$). Let this angle be called θ. Then using cosine rule, we have $\cos\theta = 1 - \|q' - p\|^2/2\tau^2$.

Define α as the angle subtended by vectors v and $q - p$, and α' the angle subtended by vectors v and $q' - p$. WLOG we can assume that the angles α and α' are less than π. Then, $\cos\alpha \geq \cos\alpha' = \cos\theta/2$. Using the trig identity $\cos\theta = 2\cos^2\left(\frac{\theta}{2}\right) - 1$, and noting $\|q - p\|^2 \geq \|q' - p\|^2$, we have

$$\left|\frac{v}{\|v\|} \cdot \frac{q-p}{\|q-p\|}\right| = \cos\alpha \geq \cos\frac{\theta}{2} \geq \sqrt{1 - \|q-p\|^2/4\tau^2} \geq 1 - (D_G(p,q)/2\tau)^2.$$

Now, by applying the cosine rule, we have $\left\|\frac{v}{\|v\|} - \frac{q-p}{\|q-p\|}\right\|^2 = 2(1 - \cos\alpha)$. The lemma follows. ■

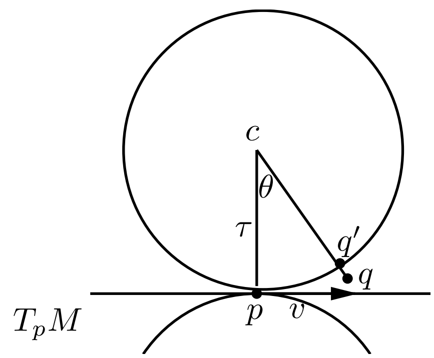

Figure 3: Plane spanned by vectors $q-p$ and $v \in T_pM$ (where v is the projection of $q-p$ onto T_pM), with τ-balls tangent to p. Note that q' is the point on the ball such that $\angle pcq = \angle pcq' = \theta$.

Lemma 21 (approximating tangent space by closeby samples) *Let $0 < \delta \leq 1$. Pick any point $p_0 \in M$ and let $p_1, \ldots, p_n \in M$ be n points distinct from p_0 such that (for all $1 \leq i \leq n$)*

(i) $D_G(p_0, p_i) \leq \tau\delta/\sqrt{n}$,

(ii) $\left|\frac{p_i - p_0}{\|p_i - p_0\|} \cdot \frac{p_j - p_0}{\|p_j - p_0\|}\right| \leq 1/2n$ *(for $i \neq j$).*

Let $\hat{T}$ be the n dimensional subspace spanned by vectors $\{p_i - p_0\}_{i\in[n]}$. For any unit vector $\hat{u} \in \hat{T}$, let u be the projection of $\hat{u}$ onto $T_{p_0}M$. Then, $\left|\hat{u} \cdot \frac{u}{\|u\|}\right| \geq 1 - \delta$.

Proof Define the vectors $\hat{v}_i := \frac{p_i - p_0}{\|p_i - p_0\|}$ (for $1 \leq i \leq n$). Observe that $\{\hat{v}_i\}_{i\in[n]}$ forms a basis of $\hat{T}$. For $1 \leq i \leq n$, define v_i as the projection of vector $\hat{v}_i$ onto $T_{p_0}M$. Also note that by applying Lemma 20, we have that for all $1 \leq i \leq n$, $\|\hat{v}_i - v_i\|^2 \leq \delta^2/2n$.

Let $V = [\hat{v}_1, \ldots, \hat{v}_n]$ be the $D \times n$ matrix. We represent the unit vector $\hat{u}$ as $V\alpha = \sum_i \alpha_i \hat{v}_i$. Also, since u is the projection of $\hat{u}$, we have $u = \sum_i \alpha_i v_i$. Then, $\|\alpha\|^2 \leq 2$. To see this, we first identify $\hat{T}$ with $\mathbb{R}^n$ via an isometry S (a linear map that preserves the lengths and angles of all vectors in $\hat{T}$). Note that S can be represented as an $n \times D$ matrix, and since V forms a basis for $\hat{T}$, SV is an $n \times n$ invertible matrix. Then, since $S\hat{u} = SV\alpha$, we have $\alpha = (SV)^{-1}S\hat{u}$. Thus, (recall $\|S\hat{u}\| = 1$)

$$
\begin{aligned}
\|\alpha\|^2 &\leq \max_{x\in S^{n-1}} \|(SV)^{-1}x\|^2 = \lambda_{\max}((SV)^{-\mathsf{T}}(SV)^{-1}) \\
&= \lambda_{\max}((SV)^{-1}(SV)^{-\mathsf{T}}) = \lambda_{\max}((V^{\mathsf{T}}V)^{-1}) = 1/\lambda_{\min}(V^{\mathsf{T}}V) \\
&\leq 1/1 - ((n-1)/2n) \leq 2,
\end{aligned}
$$

where i) $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the largest and smallest eigenvalues of a square symmetric matrix A respectively, and ii) the second inequality is by noting that $V^{\mathsf{T}}V$ is an $n \times n$ matrix with 1's on the diagonal and at most $1/2n$ on the off-diagonal elements, and applying the Gershgorin circle theorem.

Now we can bound the quantity of interest. Note that

$$\begin{aligned}\left|\hat{u}\cdot\frac{u}{\|u\|}\right| &\geq |\hat{u}^{\mathsf{T}}(\hat{u}-(\hat{u}-u))| \geq 1-\|\hat{u}-u\| = 1-\Big\|\sum_i \alpha_i(\hat{v}_i - v_i)\Big\| \\ &\geq 1-\sum_i |\alpha_i|\|\hat{v}_i - v_i\| \geq 1-(\delta/\sqrt{2n})\sum_i|\alpha_i| \geq 1-\delta,\end{aligned}$$

where the last inequality is by noting $\|\alpha\|_1 \leq \sqrt{2n}$. ■

Appendix B. On Constructing a Bounded Manifold Cover

Given a compact n-manifold $M \subset \mathbb{R}^D$ with condition number $1/\tau$, and some $0 < \delta \leq 1$. We can construct an α-bounded (ρ, δ) cover X of M (with $\alpha \leq 2^{10n+1}$ and $\rho \leq \tau\delta/3\sqrt{2}n$) as follows.

Set $\rho \leq \tau\delta/3\sqrt{2}n$ and pick a $(\rho/2)$-net C of M (that is $C \subset M$ such that, i. for $c, c' \in C$ such that $c \neq c'$, $\|c - c'\| \geq \rho/2$, ii. for all $p \in M$, exists $c \in C$ such that $\|c - p\| < \rho/2$). WLOG we shall assume that all points of C are in the interior of M. Then, for each $c \in C$, define $M_{c,\rho/2} := \{p \in M : \|p - c\| \leq \rho/2\}$, and the orthogonal projection map $f_c : M_{c,\rho/2} \to T_cM$ that projects $M_{c,\rho/2}$ onto T_cM (note that, cf. Lemma 18(i), f_c is one-to-one). Note that T_cM can be identified with $\mathbb{R}^n$ with the c as the origin. We will denote the origin as $x_0^{(c)}$, that is, $x_0^{(c)} = f_c(c)$.

Now, let B_c be any n-dimensional closed ball centered at the origin $x_0^{(c)} \in T_cM$ of radius $r > 0$ that is completely contained in $f_c(M_{c,\rho/2})$ (that is, $B_c \subset f_c(M_{c,\rho/2})$). Pick a set of n points $x_1^{(c)}, \ldots, x_n^{(c)}$ on the surface of the ball B_c such that $(x_i^{(c)} - x_0^{(c)}) \cdot (x_j^{(c)} - x_0^{(c)}) = 0$ for $i \neq j$. Define the bounded manifold cover as

$$X := \bigcup_{c\in C, i=0,\ldots,n} f_c^{-1}(x_i^{(c)}). \tag{4}$$

Lemma 22 *Let $0 < \delta \leq 1$ and $\rho \leq \tau\delta/3\sqrt{2}n$. Let C be a $(\rho/2)$-net of M as described above, and X be as in Eq. (4). Then X forms a 2^{10n+1}-bounded (ρ, δ) cover of M.*

Proof Pick any point $p \in M$ and define $X_p := \{x \in X : \|x - p\| < \rho\}$. Let $c \in C$ be such that $\|p - c\| < \rho/2$. Then X_p has the following properties.

Covering criterion: For $0 \leq i \leq n$, since $\|f_c^{-1}(x_i^{(c)}) - c\| \leq \rho/2$ (by construction), we have $\|f_c^{-1}(x_i^{(c)}) - p\| < \rho$. Thus, $f_c^{-1}(x_i^{(c)}) \in X_p$ (for $0 \leq i \leq n$). Now, for $1 \leq i \leq n$, noting that $D_G(f_c^{-1}(x_i^{(c)}), f_c^{-1}(x_0^{(c)})) \leq 2\|f_c^{-1}(x_i^{(c)}) - f_c^{-1}(x_0^{(c)})\| \leq \rho$ (cf. Lemma 17), we have that for the vector $\hat{v}_i^{(c)} := \frac{f_c^{-1}(x_i^{(c)}) - f_c^{-1}(x_0^{(c)})}{\|f_c^{-1}(x_i^{(c)}) - f_c^{-1}(x_0^{(c)})\|}$ and its (normalized) projection $v_i^{(c)} := \frac{x_i^{(c)} - x_0^{(c)}}{\|x_i^{(c)} - x_0^{(c)}\|}$ onto T_cM, $\|\hat{v}_i^{(c)} - v_i^{(c)}\| \leq \rho/\sqrt{2}\tau$ (cf. Lemma 20). Thus, for $i \neq j$, we have (recall, by construction, we have

$v_i^{(c)} \cdot v_j^{(c)} = 0$)

$$\begin{aligned} |\hat{v}_i^{(c)} \cdot \hat{v}_j^{(c)}| &= |(\hat{v}_i^{(c)} - v_i^{(c)} + v_i^{(c)}) \cdot (\hat{v}_j^{(c)} - v_j^{(c)} + v_j^{(c)})| \\ &= |(\hat{v}_i^{(c)} - v_i^{(c)}) \cdot (\hat{v}_j^{(c)} - v_j^{(c)}) + v_i^{(c)} \cdot (\hat{v}_j^{(c)} - v_j^{(c)}) + (\hat{v}_i^{(c)} - v_i^{(c)}) \cdot v_j^{(c)}| \\ &\leq \|(\hat{v}_i^{(c)} - v_i^{(c)})\|\|(\hat{v}_j^{(c)} - v_j^{(c)})\| + \|\hat{v}_i^{(c)} - v_i^{(c)}\| + \|\hat{v}_j^{(c)} - v_j^{(c)}\| \\ &\leq 3\rho/\sqrt{2}\tau \leq 1/2n. \end{aligned}$$

Point representation criterion: There exists $x \in X_p$, namely $f_c^{-1}(x_0^{(c)})$ $(= c)$, such that $\|p - x\| \leq \rho/2$.

Local boundedness criterion: Define $M_{p,3\rho/2} := \{q \in M : \|q - p\| < 3\rho/2\}$. Note that $X_p \subset \{f_c^{-1}(x_i^{(c)}) : c \in C \cap M_{p,3\rho/2}, 0 \leq i \leq n\}$. Now, using Lemma 19 we have that there exists a cover $N \subset M_{p,3\rho/2}$ of size at most 9^{3n} such that for any point $q \in M_{p,3\rho/2}$, there exists $n' \in N$ such that $\|q - n'\| < \rho/4$. Note that, by construction of C, there cannot be an $n' \in N$ such that it is within distance $\rho/4$ of two (or more) distinct $c, c' \in C$ (since otherwise the distance $\|c - c'\|$ will be less than $\rho/2$, contradicting the packing of C). Thus, $|C \cap M_{p,3\rho/2}| \leq 9^{3n}$. It follows that $|X_p| \leq (n+1)9^{3n} \leq 2^{10n+1}$.

Tangent space approximation criterion: Let $\hat{T}_p$ be the n-dimensional span of $\{\hat{v}_i^{(c)}\}_{i \in [n]}$ (note that $\hat{T}_p$ may not necessarily pass through p). Then, for any unit vector $\hat{u} \in \hat{T}_p$, we need to show that its projection u_p onto T_pM has the property $|\hat{u} \cdot \frac{u_p}{\|u_p\|}| \geq 1 - \delta$. Let θ be the angle between vectors $\hat{u}$ and u_p. Let u_c be the projection of $\hat{u}$ onto T_cM, and θ_1 be the angle between vectors $\hat{u}$ and u_c, and let θ_2 be the angle between vectors u_c (at c) and its parallel transport[3] along the geodesic path to p using the manifold connection. WLOG we can assume that θ_1 and θ_2 are at most $\pi/2$. Then, $\theta \leq \theta_1 + \theta_2 \leq \pi$. We get the bound on the individual angles as follows. By applying Lemma 21, $\cos(\theta_1) \geq 1 - \delta/4$, and by applying Lemma 16, $\cos(\theta_2) \geq 1 - \delta/4$. Finally, by using Lemma 23, we have $\left|\hat{u} \cdot \frac{u_p}{\|u_p\|}\right| = \cos(\theta) \geq \cos(\theta_1 + \theta_2) \geq 1 - \delta$. ■

Lemma 23 *Let* $0 \leq \epsilon_1, \epsilon_2 \leq 1$. *If* $\cos\alpha \geq 1 - \epsilon_1$ *and* $\cos\beta \geq 1 - \epsilon_2$, *then* $\cos(\alpha + \beta) \geq 1 - \epsilon_1 - \epsilon_2 - 2\sqrt{\epsilon_1\epsilon_2}$.

Proof Applying the identity $\sin\theta = \sqrt{1 - \cos^2\theta}$ immediately yields $\sin\alpha \leq \sqrt{2\epsilon_1}$ and $\sin\beta \leq \sqrt{2\epsilon_2}$. Now, $\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \geq (1 - \epsilon_1)(1 - \epsilon_2) - 2\sqrt{\epsilon_1\epsilon_2} \geq 1 - \epsilon_1 - \epsilon_2 - 2\sqrt{\epsilon_1\epsilon_2}$. ■

Remark 24 *A dense enough sample from* M *constitutes as a bounded cover. One can selectively prune the dense sampling to control the total number of points in each neighborhood, while still maintaining the cover properties.*

3. Parallel transport is the translation of a (tangent) vector from one point to another while remaining tangent to the manifold. As the vector is transported infinitesimally along a path, it is also required to be parallel. See e.g. Chapter 2 of do Carmo (1992) for details.

Appendix C. Bounding the number of subsets K in Embedding I

By construction (see the Preprocessing step of Embedding I), $K = \max_{x \in X} |X \cap B(x, 2\rho)|$ (where $B(x, r)$ denotes a Euclidean ball centered at x of radius r). That is, K is the largest number of x's ($\in X$) that are within a 2ρ ball of some $x \in X$.

Now, pick any $x \in X$ and consider the set $M_x := M \cap B(x, 2\rho)$. Then, if $\rho \leq \tau/4$, M_x can be covered by 2^{cn} balls of radius ρ (see Lemma 19). By recalling that X forms an α-bounded (ρ, δ)-cover, we have $|X \cap B(x, 2\rho)| = |X \cap M_x| \leq \alpha 2^{cn}$ (where $c \leq 4$).

Appendix D. Various Proofs

D.1. Proof of Lemma 11

Since R is a random orthoprojector from $\mathbb{R}^D$ to $\mathbb{R}^d$, it follows that

Lemma 25 (random projection of n-manifolds – adapted from Theorem 1.5 of Clarkson (2007)) *Let M be a smooth compact n-manifold with volume V and condition number $1/\tau$. Let $\bar{R} := \sqrt{D/d}R$ be a scaling of R. Pick any $0 < \epsilon \leq 1$ and $0 < \delta \leq 1$. If $d = \Omega\big(\epsilon^{-2} \log(V/\tau^n) + \epsilon^{-2} n \log(1/\epsilon) + \ln(1/\delta)\big)$, then with probability at least $1 - \delta$, for all $p, q \in M$*

$$(1 - \epsilon)\|p - q\| \leq \|\bar{R}p - \bar{R}q\| \leq (1 + \epsilon)\|p - q\|.$$

We apply this result with $\epsilon = 1/4$. Then, for $d = \Omega(\log(V/\tau^n) + n)$, with probability at least $1 - 1/\mathrm{poly}(n)$, $(3/4)\|p - q\| \leq \|\bar{R}p - \bar{R}q\| \leq (5/4)\|p - q\|$. Now let $\Phi : \mathbb{R}^D \to \mathbb{R}^d$ be defined as $\Phi x := (2/3)\bar{R}x = (2/3)(\sqrt{D/d})x$ (as per the lemma statement). Then we immediately get $(1/2)\|p - q\| \leq \|\Phi p - \Phi q\| \leq (5/6)\|p - q\|$.

Also note that for any $x \in \mathbb{R}^D$, we have $\|\Phi x\| = (2/3)(\sqrt{D/d})\|Rx\| \leq (2/3)(\sqrt{D/d})\|x\|$ (since R is an orthoprojector).

Finally, for any point $p \in M$, a unit vector u tangent to M at p can be approximated arbitrarily well by considering a sequence $\{p_i\}_i$ of points (in M) converging to p (in M) such that $(p_i - p)/\|p_i - p\|$ converges to u. Since for all points p_i, $(1/2) \leq \|\Phi p_i - \Phi p\|/\|p_i - p\| \leq (5/6)$ (with high probability), it follows that $(1/2) \leq \|(D\Phi)_p(u)\| \leq (5/6)$.

D.2. Proof of Corollary 12

Let v_x^1 and v_x^n ($\in \mathbb{R}^n$) be the right singular vectors corresponding to singular values σ_x^1 and σ_x^n respectively of the matrix ΦF_x. Then, quickly note that $\sigma_x^1 = \|\Phi F_x v^1\|$, and $\sigma_x^n = \|\Phi F_x v^n\|$. Note that since F_x is orthonormal, we have that $\|F_x v^1\| = \|F_x v^n\| = 1$. Now, since $F_x v^n$ is in the span of column vectors of F_x, by the sampling condition (cf. Definition 2), there exists a unit length vector $\bar{v}_x^n$ tangent to M (at x) such that $|F_x v_x^n \cdot \bar{v}_x^n| \geq 1 - \delta$. Thus, decomposing $F_x v_x^n$ into two vectors a_x^n and b_x^n such that $a_x^n \perp b_x^n$ and $a_x^n := (F_x v_x^n \cdot \bar{v}_x^n)\bar{v}_x^n$, we have

$$\begin{aligned}
\sigma_x^n &= \|\Phi(F_x v^n)\| = \|\Phi((F_x v_x^n \cdot \bar{v}_x^n)\bar{v}_x^n) + \Phi b_x^n\| \\
&\geq (1 - \delta)\,\|\Phi \bar{v}_x^n\| - \|\Phi b_x^n\| \\
&\geq (1 - \delta)(1/2) - (2/3)\sqrt{2\delta D/d},
\end{aligned}$$

since $\|b_x^n\|^2 = \|F_x v_x^n\|^2 - \|a_x^n\|^2 \leq 1 - (1 - \delta)^2 \leq 2\delta$ and $\|\Phi b_x^n\| \leq (2/3)(\sqrt{D/d})\|b_x^n\| \leq (2/3)\sqrt{2\delta D/d}$. Similarly decomposing $F_x v_x^1$ into two vectors a_x^1 and b_x^1 such that $a_x^1 \perp b_x^1$ and

$a_x^1 := (F_x v_x^1 \cdot \bar{v}_x^1)\bar{v}_x^1$, we have

$$\begin{aligned} \sigma_x^1 &= \|\Phi(F_x v_x^1)\| = \|\Phi((F_x v_x^1 \cdot \bar{v}_x^1)\bar{v}_x^1) + \Phi b_x^1\| \\ &\le \left\|\Phi \bar{v}_x^1\right\| + \|\Phi b_x^1\| \\ &\le (5/6) + (2/3)\sqrt{2\delta D/d}, \end{aligned}$$

where the last inequality is by noting $\|\Phi b_x^1\| \le (2/3)\sqrt{2\delta D/d}$. Now, by our choice of δ ($\le d/32D$), and by noting that $d \le D$, the corollary follows.

D.3. Proof of Lemma 13

For any tangent vector $u = \sum_i u_i e^i$ such that $\|u\| \le 1$, we have (by algebra)

$$\begin{aligned} \|(D\Psi)_t(u)\|^2 &= \left\|\sum u_i (D\Psi)_t(e^i)\right\|^2 \\ &= \sum_{k=1}^d u_k^2 + \sum_{k=1}^n \sum_{j=1}^K \Big[\sum_{x \in X^{(j)}} \Big(\frac{A_{\text{sin}}^{k,x}(t)}{\omega}\Big) + \sqrt{\Lambda_{\Phi(x)}(t)} \cos(\omega (C^x t)_k)(C^x u)_k \Big]^2 \\ &\qquad + \Big[\sum_{x \in X^{(j)}} \Big(\frac{A_{\text{cos}}^{k,x}(t)}{\omega}\Big) - \sqrt{\Lambda_{\Phi(x)}(t)} \sin(\omega (C^x t)_k)(C^x u)_k \Big]^2, \end{aligned} \tag{5}$$

where $A_{\text{sin}}^{k,x}(t) := \sum_i u_i \sin(\omega(C^x t)_k)(d\Lambda_{\Phi(x)}^{1/2}(t)/dt^i)$ and $A_{\text{cos}}^{k,x}(t) := \sum_i u_i \cos(\omega(C^x t)_k)(d\Lambda_{\Phi(x)}^{1/2}(t)/dt^i)$.

We can simplify Eq. (5) by recalling how the subsets $X^{(j)}$ were constructed (see preprocessing stage of Embedding I). Note that for any fixed t, at most one term in the set $\{\Lambda_{\Phi(x)}(t)\}_{x \in X^{(j)}}$ is non-zero. Thus,

$$\begin{aligned} \|(D\Psi)_t(u)\|^2 &= \sum_{k=1}^d u_k^2 + \sum_{k=1}^n \sum_{x \in X} \Lambda_{\Phi(x)}(t) \cos^2(\omega(C^x t)_k)(C^x u)_k^2 + \Lambda_{\Phi(x)}(t) \sin^2(\omega(C^x t)_k)(C^x u)_k^2 \\ &\qquad + \frac{1}{\omega} \Big[\underbrace{\big(\big(A_{\text{sin}}^{k,x}(t)\big)^2 + \big(A_{\text{cos}}^{k,x}(t)\big)^2\big)/\omega}_{\zeta_1} + \underbrace{2A_{\text{sin}}^{k,x}(t)\sqrt{\Lambda_{\Phi(x)}(t)} \cos(\omega(C^x t)_k)(C^x u)_k}_{\zeta_2} \\ &\qquad \underbrace{-2A_{\text{cos}}^{k,x}(t)\sqrt{\Lambda_{\Phi(x)}(t)} \sin(\omega(C^x t)_k)(C^x u)_k}_{\zeta_3} \Big] \\ &= \|u\|^2 + \sum_{x \in X} \Lambda_{\Phi(x)}(t) \sum_{k=1}^n (C^x u)_k^2 + \zeta, \end{aligned}$$

where $\zeta := (\zeta_1 + \zeta_2 + \zeta_3)/\omega$. Noting that i) the terms $|A_{\text{sin}}^{k,x}(t)|$ and $|A_{\text{cos}}^{k,x}(t)|$ are at most $O(\alpha 9^n \sqrt{d}/\rho)$ (see Lemma 26), ii) $|(C^x u)_k| \le 4$, and iii) $\sqrt{\Lambda_{\Phi(x)}(t)} \le 1$, we can pick ω sufficiently large (say, $\omega \ge \Omega(n\alpha^2 9^n \sqrt{d}/\rho\epsilon)$ such that $|\zeta| \le \epsilon/2$ (where ϵ is the isometry constant from our main theorem).

Lemma 26 *For all k, x and t, the terms $|A_{\text{sin}}^{k,x}(t)|$ and $|A_{\text{cos}}^{k,x}(t)|$ are at most $O(\alpha 9^n \sqrt{d}/\rho)$.*

Proof We shall focus on bounding $|A^{k,x}_{\sin}(t)|$ (the steps for bounding $|A^{k,x}_{\cos}(t)|$ are similar). Note that

$$|A^{k,x}_{\sin}(t)| \;=\; \Big|\sum_{i=1}^{d} u_i \sin(\omega(C^x t)_k)\frac{d\Lambda^{1/2}_{\Phi(x)}(t)}{dt^i}\Big| \leq \sum_{i=1}^{d}|u_i|\cdot\Big|\frac{d\Lambda^{1/2}_{\Phi(x)}(t)}{dt^i}\Big| \leq \sqrt{\sum_{i=1}^{d}\Big|\frac{d\Lambda^{1/2}_{\Phi(x)}(t)}{dt^i}\Big|^2},$$

since $\|u\| \leq 1$. Thus, we can bound $|A^{k,x}_{\sin}(t)|$ by $O(\alpha 9^n \sqrt{d}/\rho)$ by noting the following lemma. ■

Lemma 27 *For all i, x and t, $|d\Lambda^{1/2}_{\Phi(x)}(t)/dt^i| \leq O(\alpha 9^n/\rho)$.*

Proof Pick any $t \in \Phi(M)$, and let $p_0 \in M$ be (the unique element) such that $\Phi(p_0) = t$. Define $N_{p_0} := \{x \in X : \|\Phi(x) - \Phi(p_0)\| < \rho\}$ as the neighborhood around p_0. Fix an arbitrary $x_0 \in N_{p_0} \subset X$ (since if $x_0 \notin N_{p_0}$ then $d\Lambda^{1/2}_{\Phi(x_0)}(t)/dt^i = 0$), and consider the function

$$\Lambda^{1/2}_{\Phi(x_0)}(t) = \left(\frac{\lambda_{\Phi(x_0)}(t)}{\sum_{x\in N_{p_0}}\lambda_{\Phi(x)}(t)}\right)^{1/2} = \left(\frac{e^{-1/(1-(\|t-\Phi(x_0)\|^2/\rho^2))}}{\sum_{x\in N_{p_0}} e^{-1/(1-(\|t-\Phi(x)\|^2/\rho^2))}}\right)^{1/2}.$$

Pick an arbitrary coordinate $i_0 \in \{1,\ldots,d\}$ and consider the (directional) derivative of this function.

$$\begin{aligned}
\frac{d\Lambda^{1/2}_{\Phi(x_0)}(t)}{dt^{i_0}} &= \frac{1}{2}\big(\Lambda^{-1/2}_{\Phi(x_0)}(t)\big)\Big(\frac{d\Lambda_{\Phi(x_0)}(t)}{dt^{i_0}}\Big)\\
&= \frac{\Big(\sum_{x\in N_{p_0}} e^{-A_t(x)}\Big)^{1/2}}{2\Big(e^{-A_t(x_0)}\Big)^{1/2}}\left[\frac{\Big(\sum_{x\in N_{p_0}} e^{-A_t(x)}\Big)\Big(\frac{-2(t_{i_0}-\Phi(x_0)_{i_0})}{\rho^2}(A_t(x_0))^2\Big)\Big(e^{-A_t(x_0)}\Big)}{\Big(\sum_{x\in N_{p_0}} e^{-A_t(x)}\Big)^2}\right.\\
&\qquad\left. - \frac{\Big(e^{-A_t(x_0)}\Big)\Big(\sum_{x\in N_{p_0}}\frac{-2(t_{i_0}-\Phi(x)_{i_0})}{\rho^2}(A_t(x))^2 e^{-A_t(x)}\Big)}{\Big(\sum_{x\in N_{p_0}} e^{-A_t(x)}\Big)^2}\right]\\
&= \frac{\Big(\sum_{x\in N_{p_0}} e^{-A_t(x)}\Big)\Big(\frac{-2(t_{i_0}-\Phi(x_0)_{i_0})}{\rho^2}(A_t(x_0))^2\Big)\Big(e^{-A_t(x_0)}\Big)^{1/2}}{2\Big(\sum_{x\in N_{p_0}} e^{-A_t(x)}\Big)^{1.5}}\\
&\qquad - \frac{\Big(e^{-A_t(x_0)}\Big)^{1/2}\Big(\sum_{x\in N_{p_0}}\frac{-2(t_{i_0}-\Phi(x)_{i_0})}{\rho^2}(A_t(x))^2 e^{-A_t(x)}\Big)}{2\Big(\sum_{x\in N_{p_0}} e^{-A_t(x)}\Big)^{1.5}},
\end{aligned}$$

where $A_t(x) := 1/(1-(\|t-\Phi(x)\|^2/\rho^2))$. Observe that the domain of A_t is $\{x \in X : \|t-\Phi(x)\| < \rho\}$ and the range is $[1, \infty)$. Recalling that for any $\beta \geq 1$, $|\beta^2 e^{-\beta}| \leq 1$ and $|\beta^2 e^{-\beta/2}| \leq 3$, we have that $|A_t(\cdot)^2 e^{-A_t(\cdot)}| \leq 1$ and $|A_t(\cdot)^2 e^{-A_t(\cdot)/2}| \leq 3$. Thus,

$$\begin{aligned}
\left|\frac{d\Lambda^{1/2}_{\Phi(x_0)}(t)}{dt^{i_0}}\right| &\leq \frac{3 \cdot \left|\sum_{x \in N_{p_0}} e^{-A_t(x)}\right| \cdot \left|\frac{2(t_{i_0} - \Phi(x_0)_{i_0})}{\rho^2}\right| + \left|e^{-A_t(x_0)/2}\right| \cdot \left|\sum_{x \in N_{p_0}} \frac{2(t_{i_0} - \Phi(x)_{i_0})}{\rho^2}\right|}{2\Big(\sum_{x \in N_{p_0}} e^{-A_t(x)}\Big)^{1.5}} \\
&\leq \frac{(3)(2/\rho)\Big|\sum_{x \in N_{p_0}} e^{-A_t(x)}\Big| + \Big|e^{-A_t(x_0)/2}\Big| \sum_{x \in N_{p_0}} (2/\rho)}{2\Big(\sum_{x \in N_{p_0}} e^{-A_t(x)}\Big)^{1.5}} \\
&\leq O(\alpha 9^n/\rho),
\end{aligned}$$

where the last inequality is by noting: i) $|N_{p_0}| \leq \alpha 9^n$ (since for all $x \in N_{p_0}$, $\|x - p_0\| \leq 2\rho$ – cf. Lemma 11, X is an α-bounded cover, and by noting that for $\rho \leq \tau/4$, a ball of radius 2ρ can be covered by 9^n balls of radius ρ on the given n-manifold – cf. Lemma 19), ii) $|e^{-A_t(x)}| \leq |e^{-A_t(x)/2}| \leq 1$ (for all x), and iii) $\sum_{x \in N_{p_0}} e^{-A_t(x)} \geq \Omega(1)$ (since our cover X ensures that for any p_0, there exists $x \in N_{p_0} \subset X$ such that $\|p_0 - x\| \leq \rho/2$ – see also Remark 3, and hence $e^{-A_t(x)}$ is non-negligible for some $x \in N_{p_0}$). ■

D.4. Proof of Lemma 14

Representing a tangent vector $u = \sum_l u_l e^l$ (such that $\|u\|^2 \leq 1$) in terms of its basis vectors, it suffices to study how $D\Psi$ acts on basis vectors. Observe that $(D\Psi_{i,j})_t(e^l) = \left(\frac{d\Psi_{i,j}(t)}{dt^l}\right)_{k=1}^{2d+3}\Big|_t$, with the k^{th} component given as

$$\begin{aligned}
\left(\frac{d\Psi_{i,j-1}(t)}{dt^l}\right)_k &+ (\eta_{i,j}(t))_k \sqrt{\Lambda_{\Phi(x_i)}(t)} C^{x_i}_{j,l} B^{i,j}_{\text{cos}}(t) - (\nu_{i,j}(t))_k \sqrt{\Lambda_{\Phi(x_i)}(t)} C^{x_i}_{j,l} B^{i,j}_{\text{sin}}(t) \\
&+ \frac{1}{\omega_{i,j}} \Big[\Big(\frac{d\eta_{i,j}(t)}{dt^l}\Big)_k \sqrt{\Lambda_{\Phi(x_i)}(t)} B^{i,j}_{\text{sin}}(t) + \Big(\frac{d\nu_{i,j}(t)}{dt^l}\Big)_k \sqrt{\Lambda_{\Phi(x_i)}(t)} B^{i,j}_{\text{cos}}(t) \\
&\qquad + (\eta_{i,j}(t))_k \frac{d\Lambda^{1/2}_{\Phi(x_i)}(t)}{dt^l} B^{i,j}_{\text{sin}}(t) + (\nu_{i,j}(t))_k \frac{d\Lambda^{1/2}_{\Phi(x_i)}(t)}{dt^l} B^{i,j}_{\text{cos}}(t)\Big],
\end{aligned}$$

where $B^{i,j}_{\text{cos}}(t) := \cos(\omega_{i,j}(C^{x_i}t)_j)$ and $B^{i,j}_{\text{sin}}(t) := \sin(\omega_{i,j}(C^{x_i}t)_j)$. For ease of notation, let $R^{k,l}_{i,j}$ be the terms in the bracket (being multiplied to $1/\omega_{i,j}$) in the above expression. Then, we have (for any i, j)

$$
\begin{aligned}
\|(D\Psi_{i,j})_t(u)\|^2 &= \Big\|\sum_l u_l (D\Psi_{i,j})_t(e^l)\Big\|^2 \\
&= \sum_{k=1}^{2d+3}\Big[\underbrace{\sum_l u_l \left(\frac{d\Psi_{i,j-1}(t)}{dt^l}\right)_k}_{\zeta_{i,j}^{k,1}} + \underbrace{(\eta_{i,j}(t))_k\sqrt{\Lambda_{\Phi(x_i)}(t)}\cos(\omega_{i,j}(C^{x_i}t)_j)\sum_l C^{x_i}_{j,l}u_l}_{\zeta_{i,j}^{k,2}} \\
&\qquad \underbrace{-(\nu_{i,j}(t))_k\sqrt{\Lambda_{\Phi(x_i)}(t)}\sin(\omega_{i,j}(C^{x_i}t)_j)\sum_l C^{x_i}_{j,l}u_l}_{\zeta_{i,j}^{k,3}} + (1/\omega_{i,j})\underbrace{\sum_l u_l R^{k,l}_{i,j}}_{\zeta_{i,j}^{k,4}}\Big]^2 \\
&= \underbrace{\|(D\Psi_{i,j-1})_t(u)\|^2}_{=\sum_k \left(\zeta_{i,j}^{k,1}\right)^2} + \underbrace{\Lambda_{\Phi(x_i)}(t)(C^{x_i}u)_j^2}_{=\sum_k \left(\zeta_{i,j}^{k,2}\right)^2 + \left(\zeta_{i,j}^{k,3}\right)^2} \\
&\quad + \underbrace{\sum_k \Big[\big(\zeta_{i,j}^{k,4}/\omega_{i,j}\big)^2 + \big(2\zeta_{i,j}^{k,4}/\omega_{i,j}\big)\big(\zeta_{i,j}^{k,1}+\zeta_{i,j}^{k,2}+\zeta_{i,j}^{k,3}\big) + 2\big(\zeta_{i,j}^{k,1}\zeta_{i,j}^{k,2} + \zeta_{i,j}^{k,1}\zeta_{i,j}^{k,3}\big)\Big]}_{Z_{i,j}}, \qquad (6)
\end{aligned}
$$

where the last equality is by expanding the square and by noting that $\sum_k \zeta_{i,j}^{k,2}\zeta_{i,j}^{k,3} = 0$ since η and ν are orthogonal to each other. The base case $\|(D\Psi_{0,n})_t(u)\|^2$ equals $\|u\|^2$.

Note that the cross terms $\sum_k(\zeta_{i,j}^{k,1}\zeta_{i,j}^{k,2})$ and $\sum_k(\zeta_{i,j}^{k,1}\zeta_{i,j}^{k,3})$ are very close to zero since η and ν are approximately normal to the tangent vector.

By definition, $\|(D\Psi)_t(u)\|^2 = \|(D\Psi_{|X|,n})_t(u)\|^2$. Thus, using Eq. (6) and expanding the recursion, we have

$$
\begin{aligned}
\|(D\Psi)_t(u)\|^2 &= \|(D\Psi_{|X|,n})_t(u)\|^2 \\
&= \|(D\Psi_{|X|,n-1})_t(u)\|^2 + \Lambda_{\Phi(x_{|X|})}(t)(C^{x_{|X|}}u)_n^2 + Z_{|X|,n} \\
&\;\;\vdots \\
&= \|(D\Psi_{0,n})_t(u)\|^2 + \Big[\sum_{i=1}^{|X|}\Lambda_{\Phi(x_i)}(t)\sum_{j=1}^{n}(C^{x_i}u)_j^2\Big] + \sum_{i,j} Z_{i,j}.
\end{aligned}
$$

Note that $(D\Psi_{i,0})_t(u) := (D\Psi_{i-1,n})_t(u)$. Now recalling that $\|(D\Psi_{0,n})_t(u)\|^2 = \|u\|^2$ (the base case of the recursion), all we need to show is that $|\sum_{i,j} Z_{i,j}| \le \epsilon/2$. This follows directly from the lemma below.

Lemma 28 *Let $\epsilon_0 \le O\big(\epsilon/d(n|X|)^2\big)$, and for any i, j, let $\omega_{i,j} \ge \Omega\big((K_{i,j} + (\alpha 9^n/\rho))(nd|X|)^2/\epsilon\big)$ (as per the statement of Lemma 14). Then, for any i, j, $|Z_{i,j}| \le \epsilon/2n|X|$.*

Proof Recall that (cf. Eq. (6))

$$
Z_{i,j} = \underbrace{\frac{1}{\omega_{i,j}^2}\sum_k \big(\zeta_{i,j}^{k,4}\big)^2}_{(a)} + \underbrace{2\sum_k \frac{\zeta_{i,j}^{k,4}}{\omega_{i,j}}\big(\zeta_{i,j}^{k,1}+\zeta_{i,j}^{k,2}+\zeta_{i,j}^{k,3}\big)}_{(b)} + \underbrace{2\sum_k \zeta_{i,j}^{k,1}\zeta_{i,j}^{k,2}}_{(c)} + \underbrace{2\sum_k \zeta_{i,j}^{k,1}\zeta_{i,j}^{k,3}}_{(d)}.
$$

Term (a): Note that $|\sum_k (\zeta_{i,j}^{k,4})^2| \leq O\big(d^3(K_{i,j} + (\alpha 9^n/\rho))^2\big)$ (cf. Lemma 29 (iv)). By our choice of $\omega_{i,j}$, we have term (a) at most $O(\epsilon/n|X|)$.

Term (b): Note that $\big|\zeta_{i,j}^{k,1} + \zeta_{i,j}^{k,2} + \zeta_{i,j}^{k,3}\big| \leq O(n|X| + (\epsilon/dn|X|))$ (by noting Lemma 29 (i)-(iii), recalling the choice of $\omega_{i,j}$, and summing over all i', j'). Thus, $\big|\sum_k \zeta_{i,j}^{k,4}(\zeta_{i,j}^{k,1} + \zeta_{i,j}^{k,2} + \zeta_{i,j}^{k,3})\big| \leq O\big(\big(d^2(K_{i,j} + (\alpha 9^n/\rho))\big)\big(n|X| + (\epsilon/dn|X|)\big)\big)$. Again, by our choice of $\omega_{i,j}$, term (b) is at most $O(\epsilon/n|X|)$.

Terms (c) and (d): We focus on bounding term (c) (the steps for bounding term (d) are same). Note that $|\sum_k \zeta_{i,j}^{k,1}\zeta_{i,j}^{k,2}| \leq 4|\sum_k \zeta_{i,j}^{k,1}(\eta_{i,j}(t))_k|$. Now, observe that $\big(\zeta_{i,j}^{k,1}\big)_{k=1,\ldots,2d+3}$ is a tangent vector with length at most $O(dn|X| + (\epsilon/dn|X|))$ (cf. Lemma 29 (i)). Thus, by noting that $\eta_{i,j}$ is almost normal (with quality of approximation ϵ_0), we have term (c) at most $O(\epsilon/n|X|)$.

By choosing the constants in the order terms appropriately, we can get the lemma. ■

Lemma 29 *Let $\zeta_{i,j}^{k,1}$, $\zeta_{i,j}^{k,2}$, $\zeta_{i,j}^{k,3}$, and $\zeta_{i,j}^{k,4}$ be as defined in Eq. (6). Then for all $1 \leq i \leq |X|$ and $1 \leq j \leq n$, we have*

(i) $|\zeta_{i,j}^{k,1}| \leq 1 + 8n|X| + \sum_{i'=1}^{i}\sum_{j'=1}^{j-1} O(d(K_{i',j'} + (\alpha 9^n/\rho))/\omega_{i',j'})$,

(ii) $|\zeta_{i,j}^{k,2}| \leq 4$,

(iii) $|\zeta_{i,j}^{k,3}| \leq 4$,

(iv) $|\zeta_{i,j}^{k,4}| \leq O(d(K_{i,j} + (\alpha 9^n/\rho)))$.

Proof First note for any $\|u\| \leq 1$ and for any $x_i \in X$, $1 \leq j \leq n$ and $1 \leq l \leq d$, we have $|\sum_l C_{j,l}^{x_i} u_l| = |(C^{x_i}u)_j| \leq 4$ (cf. Lemma 31 (b) and Corollary 12).

Noting that for all i and j, $\|\eta_{i,j}\| = \|\nu_{i,j}\| = 1$, we have $|\zeta_{i,j}^{2,k}| \leq 4$ and $|\zeta_{i,j}^{3,k}| \leq 4$.

Observe that $\zeta_{i,j}^{k,4} = \sum_l u_l R_{i,j}^{k,l}$. For all i, j, k and l, note that i) $\|d\eta_{i,j}(t)/dt^l\| \leq K_{i,j}$ and $\|d\nu_{i,j}(t)/dt^l\| \leq K_{i,j}$ and ii) $|d\lambda_{\Phi(x_i)}^{1/2}(t)/dt^l| \leq O(\alpha 9^n/\rho)$ (cf. Lemma 27). Thus we have $|\zeta_{i,j}^{k,4}| \leq O(d(K_{i,j} + (\alpha 9^n/\rho)))$.

Now for any i, j, note that $\zeta_{i,j}^{k,1} = \sum_l u_l d\Psi_{i,j-1}(t)/dt^l$. Thus by recursively expanding, $|\zeta_{i,j}^{k,1}| \leq 1 + 8n|X| + \sum_{i'=1}^{i}\sum_{j'=1}^{j-1} O(d(K_{i',j'} + (\alpha 9^n/\rho))/\omega_{i',j'})$. ■

D.5. Proof of Lemma 15

We start by stating the following useful observations:

Lemma 30 *Let A be a linear operator such that $\max_{\|x\|=1} \|Ax\| \leq \delta_{\max}$. Let u be a unit-length vector. If $\|Au\| \geq \delta_{\min} > 0$, then for any unit-length vector v such that $|u \cdot v| \geq 1 - \epsilon$, we have*

$$1 - \frac{\delta_{\max}\sqrt{2\epsilon}}{\delta_{\min}} \leq \frac{\|Av\|}{\|Au\|} \leq 1 + \frac{\delta_{\max}\sqrt{2\epsilon}}{\delta_{\min}}.$$

Proof Let $v' = v$ if $u \cdot v > 0$, otherwise let $v' = -v$. Quickly note that $\|u - v'\|^2 = \|u\|^2 + \|v'\|^2 - 2u \cdot v' = 2(1 - u \cdot v') \leq 2\epsilon$. Thus, we have,

i. $\|Av\| = \|Av'\| \leq \|Au\| + \|A(u - v')\| \leq \|Au\| + \delta_{\max}\sqrt{2\epsilon}$,

ii. $\|Av\| = \|Av'\| \geq \|Au\| - \|A(u - v')\| \geq \|Au\| - \delta_{\max}\sqrt{2\epsilon}$.

Noting that $\|Au\| \geq \delta_{\min}$ yields the result. ■

Lemma 31 *Let $x_1, \ldots, x_n \in \mathbb{R}^D$ be a set of orthonormal vectors, $F := [x_1, \ldots, x_n]$ be a $D \times n$ matrix and let Φ be a linear map from $\mathbb{R}^D$ to $\mathbb{R}^d$ ($n \leq d \leq D$) such that for all non-zero $a \in$ span(F) we have $0 < \|\Phi a\| \leq \|a\|$. Let $U\Sigma V^{\mathsf{T}}$ be the thin SVD of ΦF. Define $C = (\Sigma^{-2} - I)^{1/2} U^{\mathsf{T}}$. Then,*

(a) $\|C(\Phi a)\|^2 = \|a\|^2 - \|\Phi a\|^2$, for any $a \in$ span(F),

(b) $\|C\|^2 \leq (1/\sigma^n)^2$, where $\|\cdot\|$ denotes the spectral norm of a matrix and σ^n is the n^{th} largest singular value of ΦF.

Proof Note that FV forms an orthonormal basis for the subspace spanned by columns of F that maps to $U\Sigma$ via the mapping Φ. Thus, since $a \in \text{span}(F)$, let y be such that $a = FVy$. Note that i) $\|a\|^2 = \|y\|^2$, ii) $\|\Phi a\|^2 = \|U\Sigma y\|^2 = y^{\mathsf{T}}\Sigma^2 y$. Now,

$$\begin{aligned}
\|C\Phi a\|^2 &= \|((\Sigma^{-2} - I)^{1/2} U^{\mathsf{T}})\Phi F V y\|^2 \\
&= \|(\Sigma^{-2} - I)^{1/2} U^{\mathsf{T}} U \Sigma V^{\mathsf{T}} V y\|^2 \\
&= \|(\Sigma^{-2} - I)^{1/2} \Sigma y\|^2 \\
&= y^{\mathsf{T}} y - y^{\mathsf{T}} \Sigma^2 y \\
&= \|a\|^2 - \|\Phi a\|^2.
\end{aligned}$$

Now, consider $\|C\|^2$.

$$\begin{aligned}
\|C\|^2 &\leq \|(\Sigma^{-2} - I)^{1/2}\|^2 \|U^{\mathsf{T}}\|^2 \\
&\leq \max_{\|x\|=1} \|(\Sigma^{-2} - I)^{1/2} x\|^2 \\
&\leq \max_{\|x\|=1} x^{\mathsf{T}} \Sigma^{-2} x \\
&= \max_{\|x\|=1} \sum_i x_i^2/(\sigma^i)^2 \\
&\leq (1/\sigma^n)^2,
\end{aligned}$$

where σ^i are the (top n) singular values forming the diagonal matrix Σ. ■

Lemma 32 *Let $M \subset \mathbb{R}^D$ be a compact Riemannian n-manifold with condition number $1/\tau$. Pick any $x \in M$ and let F_x be any n-dimensional affine space with the property: for any unit vector v_x tangent to M at x, and its projection v_{xF} onto F_x, $\left|v_x \cdot \frac{v_{xF}}{\|v_{xF}\|}\right| \geq 1-\delta$. Then for any $p \in M$ such that $\|x-p\| \leq \rho \leq \tau/2$, and any unit vector v tangent to M at p, ($\xi := (2\rho/\tau) + \delta + 2\sqrt{2\rho\delta/\tau}$)*

i. $\left|v \cdot \frac{v_F}{\|v_F\|}\right| \geq 1-\xi$,

ii. $\|v_F\|^2 \geq 1-2\xi$,

iii. $\|v_r\|^2 \leq 2\xi$,

where v_F is the projection of v onto F_x and v_r is the residual (i.e. $v = v_F + v_r$ and $v_F \perp v_r$).

Proof Let γ be the angle between v_F and v. We will bound this angle.

Let v_x (at x) be the parallel transport of v (at p) via the (shortest) geodesic path via the manifold connection. Let the angle between vectors v and v_x be α. Let v_{xF} be the projection of v_x onto the subspace F_x, and let the angle between v_x and v_{xF} be β. WLOG, we can assume that the angles α and β are acute. Then, since $\gamma \leq \alpha + \beta \leq \pi$, we have that $\left|v \cdot \frac{v_F}{\|v_F\|}\right| = \cos\gamma \geq \cos(\alpha+\beta)$. We bound the individual terms $\cos\alpha$ and $\cos\beta$ as follows.

Now, since $\|p-x\| \leq \rho$, using Lemmas 16 and 17, we have $\cos(\alpha) = |v \cdot v_x| \geq 1 - 2\rho/\tau$. We also have $\cos(\beta) = \left|v_x \cdot \frac{v_{xF}}{\|v_{xF}\|}\right| \geq 1-\delta$. Then, using Lemma 23, we finally get $\left|v \cdot \frac{v_F}{\|v_F\|}\right| = |\cos(\gamma)| \geq 1 - 2\rho/\tau - \delta - 2\sqrt{2\rho\delta/\tau} = 1-\xi$.

Also note since $1 = \|v\|^2 = (v \cdot \frac{v_F}{\|v_F\|})^2 \left\|\frac{v_F}{\|v_F\|}\right\|^2 + \|v_r\|^2$, we have $\|v_r\|^2 = 1 - \left(v \cdot \frac{v_F}{\|v_F\|}\right)^2 \leq 2\xi$, and $\|v_F\|^2 = 1 - \|v_r\|^2 \geq 1-2\xi$. ■

Now we are in a position to prove Lemma 15. Let v_F be the projection of the unit vector v (at p) onto the subspace spanned by (the columns of) F_x and v_r be the residual (i.e. $v = v_F + v_r$ and $v_F \perp v_r$). Then, noting that p, x, v and F_x satisfy the conditions of Lemma 32 (with ρ in the Lemma 32 replaced with 2ρ from the statement of Lemma 15), we have ($\xi := (4\rho/\tau) + \delta + 4\sqrt{\rho\delta/\tau}$)

a) $\left|v \cdot \frac{v_F}{\|v_F\|}\right| \geq 1-\xi$,

b) $\|v_F\|^2 \geq 1-2\xi$,

c) $\|v_r\|^2 \leq 2\xi$.

We can now bound the required quantity $\|C^x u\|^2$. Note that

$$\begin{aligned}
\|C^x u\|^2 &= \|C^x \Phi v\|^2 = \|C^x \Phi(v_F + v_r)\|^2 \\
&= \|C^x \Phi v_F\|^2 + \|C^x \Phi v_r\|^2 + 2C^x \Phi v_F \cdot C^x \Phi v_r \\
&= \underbrace{\|v_F\|^2 - \|\Phi v_F\|^2}_{(a)} + \underbrace{\|C^x \Phi v_r\|^2}_{(b)} + \underbrace{2C^x \Phi v_F \cdot C^x \Phi v_r}_{(c)}
\end{aligned}$$

where the last equality is by observing v_F is in the span of F_x and applying Lemma 31 (a). We now bound the terms (a),(b), and (c) individually.

Term (a)**:** Note that $1 - 2\xi \leq \|v_F\|^2 \leq 1$ and observing that Φ satisfies the conditions of Lemma 30 with $\delta_{\text{max}} = (2/3)\sqrt{D/d}$, $\delta_{\text{min}} = (1/2) \leq \|\Phi v\|$ (cf. Lemma 11) and $\left|v \cdot \frac{v_F}{\|v_F\|}\right| \geq 1 - \xi$, we have (recall $\|\Phi v\| = \|u\| \leq 1$)

$$\begin{aligned}
\|v_F\|^2 - \|\Phi v_F\|^2 &\leq 1 - \|v_F\|^2 \left\|\Phi \frac{v_F}{\|v_F\|}\right\|^2 \\
&\leq 1 - (1 - 2\xi) \left\|\Phi \frac{v_F}{\|v_F\|}\right\|^2 \\
&\leq 1 + 2\xi - \left\|\Phi \frac{v_F}{\|v_F\|}\right\|^2 \\
&\leq 1 + 2\xi - \left(1 - (4/3)\sqrt{2\xi D/d}\right)^2 \|\Phi v\|^2 \\
&\leq 1 - \|u\|^2 + \left(2\xi + (8/3)\sqrt{2\xi D/d}\right),
\end{aligned} \tag{7}$$

where the fourth inequality is by using Lemma 30. Similarly, in the other direction

$$\begin{aligned}
\|v_F\|^2 - \|\Phi v_F\|^2 &\geq 1 - 2\xi - \|v_F\|^2 \left\|\Phi \frac{v_F}{\|v_F\|}\right\|^2 \\
&\geq 1 - 2\xi - \left\|\Phi \frac{v_F}{\|v_F\|}\right\|^2 \\
&\geq 1 - 2\xi - \left(1 + (4/3)\sqrt{2\xi D/d}\right)^2 \|\Phi v\|^2 \\
&\geq 1 - \|u\|^2 - \left(2\xi + (32/9)\xi(D/d) + (8/3)\sqrt{2\xi D/d}\right).
\end{aligned} \tag{8}$$

Term (b)**:** Note that for any x, $\|\Phi x\| \leq (2/3)(\sqrt{D/d})\|x\|$. We can apply Lemma 31 (b) with $\sigma_x^n \geq 1/4$ (cf. Corollary 12) and noting that $\|v_r\|^2 \leq 2\xi$, we immediately get

$$0 \leq \|C^x \Phi v_r\|^2 \leq 4^2 \cdot (4/9)(D/d)\|v_r\|^2 \leq (128/9)(D/d)\xi. \tag{9}$$

Term (c)**:** Recall that for any x, $\|\Phi x\| \leq (2/3)(\sqrt{D/d})\|x\|$, and using Lemma 31 (b) we have that $\|C^x\|^2 \leq 16$ (since $\sigma_x^n \geq 1/4$ – cf. Corollary 12).

Now let $a := C^x \Phi v_F$ and $b := C^x \Phi v_r$. Then $\|a\| = \|C^x \Phi v_F\| \leq \|C^x\| \|\Phi v_F\| \leq 4$, and $\|b\| = \|C^x \Phi v_r\| \leq (8/3)\sqrt{2\xi D/d}$ (see Eq. (9)).

Thus, $|2a \cdot b| \leq 2\|a\|\|b\| \leq 2 \cdot 4 \cdot (8/3)\sqrt{2\xi D/d} = (64/3)\sqrt{2\xi D/d}$. Equivalently,

$$-(64/3)\sqrt{2\xi D/d} \leq 2C^x \Phi v_F \cdot C^x \Phi v_r \leq (64/3)\sqrt{2\xi D/d}. \tag{10}$$

Combining (7)-(10), and noting $d \leq D$, yields the lemma.

Appendix E. Computing the Normal Vectors

The success of the second embedding technique crucially depends upon finding (at each iteration step) a pair of mutually orthogonal unit vectors that are normal to the embedding of manifold M

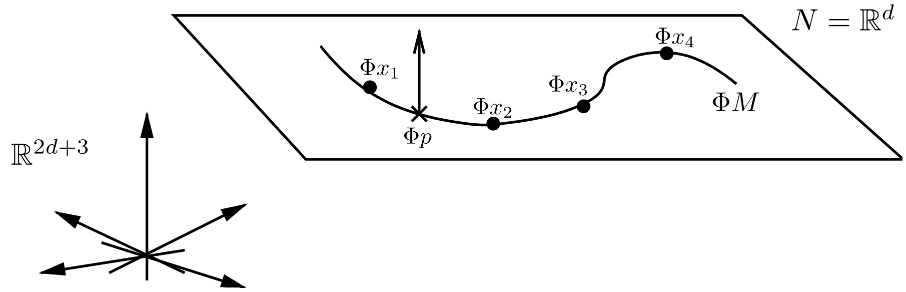

Figure 4: Basic setup for computing the normals to the underlying n-manifold ΦM at the point of interest Φp. Observe that even though it is difficult to find vectors normal to ΦM at Φp within the containing space $\mathbb{R}^d$ (because we only have a finite-size sample from ΦM, viz. Φx_1, Φx_2, etc.), we can treat the point Φp as part of the bigger ambient manifold N (= $\mathbb{R}^d$, that contains ΦM) and compute the desired normals in a space that contains N itself. Now, for each i, j iteration of Algorithm II, $\Psi_{i,j}$ acts on the entire N, and since we have complete knowledge about N, we can compute the desired normals.

(from the previous iteration step) at a given point p. At a first glance finding such normal vectors seems infeasible since we only have access to a finite size sample X from M. The saving grace comes from noting that the corrections are applied to the n-dimensional manifold $\Phi(M)$ that is actually a *submanifold* of d-dimensional space $\mathbb{R}^d$. Let's denote this space $\mathbb{R}^d$ as a flat d-manifold N (containing our manifold of interest $\Phi(M)$). Note that even though we only have partial information about $\Phi(M)$ (since we only have samples from it), we have full information about N (since it is the entire space $\mathbb{R}^d$). What it means is that given some point of interest $\Phi p \in \Phi(M) \subset N$, finding a vector normal to N (at Φp) automatically is a vector normal to $\Phi(M)$ (at Φp). Of course, to find two mutually orthogonal normals to a d-manifold N, N itself needs to be embedded in a larger dimensional Euclidean space (although embedding into $d + 2$ should suffice, for computational reasons we will embed N into Euclidean space of dimension $2d + 3$). This is precisely the first thing we do before applying any corrections (cf. Step 2 of Embedding II in Section 4). See Figure 4 for an illustration of the setup before finding any normals.

Now for every iteration of the algorithm, note that we have complete knowledge of N and exactly what function (namely $\Psi_{i,j}$ for iteration i, j) is being applied to N. Thus with additional computation effort, one can compute the necessary normal vectors.

More specifically, We can estimate a pair of mutually orthogonal unit vectors that are normal to $\Psi_{i,j}(N)$ at Φp (for any step i, j) as follows.

Algorithm 4 Compute Normal Vectors

Preprocessing Stage:

1: Let $\eta_{i,j}^{\text{rand}}$ and $\nu_{i,j}^{\text{rand}}$ be vectors in $\mathbb{R}^{2d+3}$ drawn independently at random from the surface of the unit-sphere (for $1 \leq i \leq |X|$, $1 \leq j \leq n$).

Compute Normals: For any point of interest $p \in M$, let $t := \Phi p$ denote its projection into $\mathbb{R}^d$. Now, for any iteration i, j (where $1 \leq i \leq |X|$, and $1 \leq j \leq n$), we shall assume that vectors η and ν upto iterations $i, j-1$ are already given. Then we can compute the (approximated) normals $\eta_{i,j}(t)$ and $\nu_{i,j}(t)$ for the iteration i, j as follows.

1: Let $\Delta > 0$ be the quality of approximation.
2: **for** $k = 1, \ldots, d$ **do**
3: Approximate the k^{th} tangent vector as

$$T^k := \frac{\Psi_{i,j-1}(t + \Delta e^k) - \Psi_{i,j-1}(t)}{\Delta},$$

where $\Psi_{i,j-1}$ is as defined in Section 5.3, and e^k is the k^{th} standard vector.
4: **end for**
5: Let $\eta = \eta_{i,j}^{\text{rand}}$, and $\nu = \nu_{i,j}^{\text{rand}}$.
6: Use Gram-Schmidt orthogonalization process to extract $\hat{\eta}$ (from η) that is orthogonal to vectors $\{T^1, \ldots, T^d\}$.
7: Use Gram-Schmidt orthogonalization process to extract $\hat{\nu}$ (from ν) that is orthogonal to vectors $\{T^1, \ldots, T^d, \hat{\eta}\}$.
8: **return** $\hat{\eta}/\|\hat{\eta}\|$ and $\hat{\nu}/\|\hat{\nu}\|$ as mutually orthogonal unit vectors that are approximately normal to $\Psi_{i,j-1}(\Phi M)$ at $\Psi_{i,j-1}(t)$.

A few remarks are in order.

Remark 33 *The choice of target dimension of size $2d+3$ (instead of $d+2$) ensures that a pair of random unit-vectors η and ν are not parallel to any vector in the tangent bundle of $\Psi_{i,j-1}(N)$ with probability 1. This follows from Sard's theorem (see e.g. Milnor, 1972), and is the key observation in reducing the embedding size in Whitney's embedding (Whitney, 1936). This also ensures that our orthogonalization process (Steps 6 and 7) will not result in a null vector.*

Remark 34 *By picking Δ sufficiently small, we can approximate the normals η and ν arbitrarily well by approximating the tangents $T^1, \ldots, T^d$ well.*

Remark 35 *For each iteration i, j, the vectors $\hat{\eta}/\|\hat{\eta}\|$ and $\hat{\nu}/\|\hat{\nu}\|$ that are returned (in Step 8) are a smooth modification to the starting vectors $\eta_{i,j}^{\text{rand}}$ and $\nu_{i,j}^{\text{rand}}$ respectively. Now, since we use the same starting vectors $\eta_{i,j}^{\text{rand}}$ and $\nu_{i,j}^{\text{rand}}$ regardless of the point of application ($t = \Phi p$), it follows that the respective directional derivates of the returned vectors are bounded as well.*

By noting Remarks 34 and 35, the approximate normals we return satisfy the conditions needed for Embedding II (see our discussion in Section 5.3).