Tighter Bounds for Random Projections of Manifolds

Kenneth L. Clarkson*

July 22, 2008

Abstract

The Johnson-Lindenstrauss random projection lemma gives a simple way to reduce the dimensionality of a set of points while approximately preserving their pairwise distances. The most direct application of the lemma applies to a finite set of points, but recent work has extended the technique to affine subspaces, curves, and general smooth manifolds. Here the case of random projection of smooth manifolds is considered, and a previous analysis is sharpened, reducing the dependence on such properties as the manifold's maximum curvature.

1 Introduction

The difficulty of dealing with high-dimensional data is a problem of long standing and continuing effort. One approach to it is to reduce the dimension: to map the original dataset in high dimension to another dataset in lower dimension, while preserving important properties as much as possible. While the singular value decomposition (SVD) gives such a mapping, and is optimal in some ways, an alternative approach simply rotates the data randomly, and then drops all but a small number of coordinates. Equivalently, it picks a random subspace of a given appropriate dimension, and projects the data to that subspace. Surprisingly, this approach can work well: as shown by Johnson and Lindenstrauss [JL84], it approximately preserves the distances between all pairs of points, with high probability. Its simplicity, and its obliviousness in some respects to the original data, allow it to be used in settings where the SVD is not appropriate.

An active area of research in recent years has been to extend this approach in various ways: finding similar transformations that also work well, finding additional properties that are preserved by random projection [Mag02], finding faster methods of implementing it [AC06], and determining the most general conditions on the dataset under which it can be applied.

*IBM Almaden Research Center, San Jose, CA.

This paper is a contribution to the latter line of work. Results for random projections were first proved for a single vector and so immediately for finite sets of vectors; recently affine subspaces [Sar06], sets with bounded doubling dimension [IN07], and curves and surfaces[AHPY07, BW06] have been considered. Here the results for curves and surfaces are extended, by giving somewhat weaker conditions under which the random projection technique can be applied.

These results give insight into a related question: what could we regard as the "complexity" of a geometric data set? While the (intrinsic) dimension of the set is a crucial property, including for algorithms, it is inadequate for telling the difference between a circle $\mathcal{C}$ and a curve $\mathcal{C}'$ that is approximately space-filling. We know heuristically that solving algorithmic problems for points that are samples of $\mathcal{C}$ should be easier than for samples of $\mathcal{C}'$, but why? This paper continues work that gives a partial but quantitative answer: in the analysis here, a dimension that is large enough for a random projection to preserve preserve distances among points of $\mathcal{C}$ is smaller than that needed for $\mathcal{C}'$.

The remainder of this introduction will introduce the topic in more technical detail, describe some previous work, and outline the results and give some pointers to the rest of the paper.

We are interested in linear mappings from $\mathbb{R}^m$ to its linear subspaces of some smaller dimension k. In particular, the maps will be chosen at random from a distribution constructed as follows: pick a random k-dimensional subspace F of $\mathbb{R}^m$, uniform under Haar measure. Call the orthogonal projection onto F, scaled by $\sqrt{m/k}$, a *k-map* for short. An equivalent k-map construction is: pick a random rotation of $\mathbb{R}^m$, apply it, project to $\mathbb{R}^k$, and then scale by $\sqrt{m/k}$. (The scaling by $\sqrt{m/k}$ serves to make the expected length of the projection of a vector equal to its length.)

For $S \subset \mathbb{R}^m$, say that the linear map $\mathbf{P} : \mathbb{R}^m \to \mathbb{R}^k$ is an *ϵ-isometry* on S, or *ϵ-isometrizes* S, if for all $x \in S$,

$$(1-\epsilon)\|x\| \leq \|\mathbf{P}x\| \leq (1+\epsilon)\|x\|.$$

Here, and throughout the paper, $\|x\|$ is the Euclidean norm of x. (The p-norm for $p \neq 2$ will have the explicitly subscripted form $\|x\|_p$.)

Note that if $\mathbf{P}$ ϵ-isometrizes point $\{x\}$, it also ϵ-isometrizes any scalar multiple of x, since $\mathbf{P}$ is linear. So we might assume for example that points of S have unit norm: if we define $\mathbf{U}(S) := \{x/\|x\| \mid x \in S, x \neq 0\}$, a subset of the unit sphere $\mathbb{S}^{m-1}$, then $\mathbf{P}$ ϵ-isometrizes S if and only if it ϵ-isometrizes $\mathbf{U}(S)$.

Say that $\mathbf{P}$ *ϵ-embeds* S if for all $a, b \in S$,

$$(1-\epsilon)\|a-b\| \leq \|\mathbf{P}a - \mathbf{P}b\| \leq (1+\epsilon)\|a-b\|.$$

Comparing these definitions, and using the linearity of $\mathbf{P}$, clearly $\mathbf{P}$ ϵ-embeds S if and only if it ϵ-isometrizes $\mathbf{U}(S-S)$, where for $A, B \subset \mathbb{R}^m$, the Minkowski difference $A - B := \{x - y \mid x \in A, y \in B\}$. The vectors of $S - S$ will be called the *chords* of S.

Johnson and Lindenstrauss showed that for large enough k, for a single point x, a k-map is an ϵ-isometry of $\{x\}$, with high probability.

Lemma 1.1. (single-point JL, [JL84]) *There is a constant c_{JL} so that for a given $x \in \mathbb{R}^m$ and $\epsilon, \delta > 0$, with probability at least $1-\delta$ a $(c_{\mathsf{JL}} \log(1/\delta)/\epsilon^2)$-map ϵ-isometrizes $\{x\}$. Put another way, with failure probability at most*

$$\exp(-k\epsilon^2/c_{\mathsf{JL}}),$$

a k-map ϵ-isometrizes $\{x\}$.

By the union bound, for sets S with more than one element, the failure probability for a k-map can simply be multiplied by the size $|S|$ of the set. Applying this to obtain an ϵ-isometry for S, and to $\mathbf{U}(S-S)$ to imply embedding of S, the following is obtained.

Lemma 1.2. (finite-set JL, [JL84]) *There is a constant c_{JL} so that for a given $S \subset \mathbb{R}^m$ of n points and $\epsilon, \delta > 0$, with probability at least $1-\delta$ a $(c_{\mathsf{JL}} \log(n/\delta)/\epsilon^2)$-map ϵ-isometrizes S. Put another way, with a probability of failure bounded by $n\exp(-k\epsilon^2/c_{\mathsf{JL}})$, a k-map ϵ-isometrizes S. Similarly, with probability at least $1-\delta$, a k-map with $k = c_{\mathsf{JL}} \log(n^2/\delta)/\epsilon^2$ ϵ-embeds S, or equivalently, with a probability of failure bounded by $n^2\exp(-k\epsilon^2/c_{\mathsf{JL}})$, a k-map ϵ-embeds S.*

Random projection as in the JL Lemma has found many applications, as surveyed by [Vem04, Lin02, Ind00, IM04]. It can be even extended beyond a finite number of points: as shown by Sarlós [Sar06], for example, a d-flat can be ϵ-embedded by an $O(d/\epsilon^2)$-map. (See [Sar06] also for previous and related results for embedding flats. A *d-flat* is a d-dimensional affine subspace, so a 1-flat is a line and an $(m-1)$-flat in $\mathbb{R}^m$ is a hyperplane.)

Theorem 1.3. (subspace JL, [Sar06]) *There is a constant c_{JLs} so that for a given d-flat $F \subset \mathbb{R}^m$ and $\epsilon, \delta > 0$, with probability at least $1-\delta$ a $(c_{\mathsf{JL}}(c_{\mathsf{JLs}}d + \log(1/\delta))/\epsilon^2)$-map ϵ-embeds F. Put another way, with a probability of failure bounded by $\exp(c_{\mathsf{JLs}}d - k\epsilon^2/c_{\mathsf{JL}})$, a k-map ϵ-embeds F.*

Proof. (Sketch) It is shown that F is ϵ-isometrized if all the points in an ϵ_0-cover of the unit ball of F are ϵ-isometrized, for a small enough constant ϵ_0. The number of points in such a cover is $1/\epsilon_0^d = \exp(c_{\mathsf{JLs}}d)$. Since $\mathbf{U}(F-F) \subset F$, an ϵ-isometry of F is also an ϵ-embedding. $\square$

For any of these lemmas, call the value that is the failure probability upper bound for ϵ-isometry, times $\exp(k\epsilon^2/c_{\mathsf{JL}})$, the *failure multiplier*, so that the failure multiplier for a single point is one, for a finite set is n, and for a d-dimensional linear subspace is $\exp(c_{\mathsf{JLs}}d)$.

Random projection results have been extended even beyond linear subspaces. Indyk and Naor showed [IN07] that $S \subset \mathbb{R}^m$ with bounded doubling dimension, or more generally with bounded $E[\sup\{x \cdot y \mid x \in S, y \sim N(0,1)\}]$, can be isometrized, or *additively* embedded, where the latter means that distance errors can bounded by an additive term. A key idea is to use finite *ϵ-nets* of S, as discussed in Section 2 on page 8, to approximate the points of S, and extend isometry and additive embedding results for those ϵ-nets to all of S.

A difficulty in extending results from additive embedding to relative embedding, in the general setting of bounded doubling dimension, is accounting for the embedding of "short" chords, vectors $a - b$ for $a, b \in S$ that are very close together. When S is smooth, however, short chords converge to tangent vectors, and it is possible to express the embedding complexity of short chords in terms of the overall complexity of the collection of tangent vectors. One measure of this complexity is the total absolute curvature of S, denoted here by $\mu_{\mathrm{III}}(S)$. As discussed in §4.2 on page 18, the tangent vectors can be well-approximated by a set of cardinality $O(\mu_{\mathrm{III}}(\mathcal{M})/\epsilon^d)$; in §4.3, the properties needed to define the threshold $\tau(\mathcal{M}, \epsilon)$ of shortness are given, that is, the threshold at which chords are short enough that they are adequately approximated by the manifold's tangent vectors.

One prior result involving μ_{III}, for the case of curves, is that of Agarwal *et al.* [AHPY07], who showed ϵ-isometry results for a k-map to a projection dimension that is proportional to the logarithm of curve total curvature. They also showed that manifolds that are contained in low-dimensional flats can be embedded, and that curves and sets with bounded doubling dimension can be embedded additively. The proof by Agarwal *et al.* of their Theorem 4.1 uses in part the ϵ-net approach of Indyk and Naor. Lemma 3.1 on page 14 in turn follows their proof, and extends it slightly in a few ways; for example, it is observed that the proof requires only a bounded box dimension of the associated metric space, and not the stronger condition of bounded doubling dimension.

The work most directly antecedent to this paper is that of Baraniuk and Wakin, who showed [BW06] that a smooth manifold $\mathcal{M}$ can be embedded by random projections, with a projection dimension that is $O(\epsilon^{-2}(d\log(1/\epsilon) + \log(1/\delta)))$ as $\epsilon \to 0$, with lower-order terms that depend on the manifold. In addition to the parameters ϵ and δ as above, their results depend on the ambient dimension m, and on several properties of the manifold: its (d-dimensional) surface area $\mu_{\mathrm{I}}(\mathcal{M})$, its *reach* ρ (defined in §4.3.1 on page 19), and some quantities that can be bounded in terms of the reach: the maximum curvature, the tangent space "twisting," the relation of the Euclidean distance between points of $\mathcal{M}$ to their geodesic distance, and the geodesic covering regularity. (For the precise definitions of these terms, please see [BW06]; they refer to $1/\rho$ as the *condition number* of the manifold.)

Their main result, in the notation here, is the following.

Theorem 1.4. (manifold JL, [BW06]) *Let $\mathcal{M}$ be a compact d-submanifold of $\mathbb{R}^m$ having d-volume $\mu_{\mathrm{I}}(\mathcal{M})$, reach ρ, and geodesic covering regularity R. Then a k-map ϵ-embeds $\mathcal{M}$ with probability at least $1 - \delta$, where*

$$k = O(\mathcal{L}(d, \epsilon, \delta)) + O(\epsilon^{-2}(d\log(mR\mu_I(\mathcal{M})/\rho)))$$

as $\epsilon \to 0$, where $\mathcal{L}(d, \epsilon, \delta) := \epsilon^{-2}(d\log(1/\epsilon) + \log(1/\delta))$.

The term $\mathcal{L}(d, \epsilon, \delta)$ is separated out for emphasis as the leading term. The geodesic covering regularity R in the lemma statement is motivated by considerations that led here to the use of the quantity $\psi(\mathcal{M}, \epsilon)$, as discussed in §2.4 on page 12 and below.

This work is an attempt to tighten this result; for example, the dependence on the ambient dimension m is removed entirely. More generally, this work reduces the dependence on worst-case properties, and puts the bounds more in terms of average properties: the dependence on the maximum curvature (implicitly bounded above by $1/\rho$), for example, is (partly) replaced by a dependence on the total absolute curvature. For connected manifolds, the dependence on reach and maximum curvature is generally replaced by a dependence on *low-torsion connectivity* as discussed in §4.3.1 on page 19. Also, improved bounds are given for the setting where only the preservation of geodesic distances is needed.

Another motivation is to continue the study of the complexity of a d-dimensional manifold $\mathcal{M}$ as a function of its d-dimensional measure (surface area), its total absolute curvature, and other integral curvature measures. The surface area, denoted here by $\mu_{\rm I}(\mathcal{M})$, together with the dimension d, has been seen to bound several quantities: the expected total length of an extremal graph of random points on $\mathcal{M}$; the size of ϵ-nets for $\mathcal{M}$ with respect to geodesic distance; the expected error of vector quantization; and the expected number of intersections of $\mathcal{M}$ with a random line. The total absolute curvature, denoted here by $\mu_{\rm III}(\mathcal{M})$, has a well-known relation to the expected number of maxima, minima, and saddle points of $\mathcal{M}$ with respect to a random orientation (height function), as well as a fundamental relation to topological invariants of $\mathcal{M}$.

The relevance of total absolute curvature to random projection bounds has already been mentioned; another integral measure of $\mathcal{M}$ of interest is the *total root curvature* of $\mathcal{M}$, denoted $\mu_{\rm II}(\mathcal{M})$, which essentially determines the difficulty of approximating $\mathcal{M}$ by a simplicial mesh. This was shown by Gruber for the case where $\mathcal{M}$ is the boundary of a convex set [Gru93], and for more general manifolds of codimension one in [Cla06]. (Recall that the codimension here is $m - d$.) Similar approximation relations are discussed in this paper, but the approximation is by tangent flats, and not by simplices, and in only one direction: $\mathcal{M}$ is approximated by a collection of flats, a subset of its tangent flats, in the sense that for each point of $\mathcal{M}$, there is a nearby point on a flat in the collection. This kind of approximation is called here a *generalized cover*, or just *cover* for short.) (Here the "tangent flat" at $a \in \mathcal{M}$ is $a + T_a$, the tangent subspace of $\mathcal{M}$ at a, translated to contain a.) The metric $D_{\rm II}$ is discussed in detail in §2.3 on page 10. A cover by flats can have considerably fewer members than a cover by a set of discrete points.

Moreover, consider the *long chords* $a - b$ in $\mathbf{U}(\mathcal{M} - \mathcal{M})$, those due to points $a, b \in \mathcal{M}$ that are far apart. A cover of $\mathcal{M}$ by flats can be used to build a cover for the long chords, as shown in Lemma 4.1 on page 17. The number of flats needed is no more than proportional to $(\mu_{\rm II}(\mathcal{M})/\epsilon^d)^2$, to approximate within distance ϵ; with this dependency, and the above dependence on $\mu_{\rm III}$ for handling short chords, a bound is given here on the projection dimension k sufficient for ϵ-embedding a smooth compact connected manifold; this is stated in the theorems below. A rough version of the new bound uses the quantity $\tau(\mathcal{M}, \epsilon)$, which is a threshold for shortness: it is the largest Euclidean distance τ so that any $a, b \in \mathcal{M}$ with $\|a - b\| \leq \tau$ are connected by either a low-curvature or a low-torsion path. The latter condition is motivated by Lemma 4.8 on page 21: the pairs $a, b \in \mathcal{M}$ that are connected by such paths can be adequately approximated by tangent vectors.

The rough version of the new bound is as follows.

Theorem 1.5. *A connected, compact, orientable, differentiable manifold $\mathcal{M}$ is ϵ-embedded with probability at least $1-\delta$ by a k-map with*

$$k = O(\mathcal{L}(d,\epsilon,\delta)) + O(\epsilon^{-2}(\log(\mu_{\mathrm{II}}/\tau^d + \mu_{\mathrm{III}})))$$

as $\epsilon \to 0$, with an asymptotic threshold depending on $\mathcal{M}$, omitting dependencies of $\mu_X(\mathcal{M},\epsilon)$ and $\tau(\mathcal{M},\epsilon)$ on $\mathcal{M}$ and ϵ. Here as before

$$\mathcal{L}(d,\epsilon,\delta) := \epsilon^{-2}(d\log(1/\epsilon) + \log(1/\delta)).$$

Analogous statements hold with "I" replacing "II" in the above.

The condition of orientability can be removed, by appealing to a double cover, and the differentiable (C^∞) condition is stronger than strictly necessary: it is enough to be C^k for large enough k. Also, $\mathcal{M}$ may have a boundary, but the constants here, and in the main theorem below, will then depend on the complexity of the boundary.

For comparison: the added term in the prior bound Theorem 1.4 on page 4 is

$$O(\epsilon^{-2}(d\log(mR\mu_I(\mathcal{M})/\rho)).$$

The "roughness" of this theorem is that the asymptotic threshold depends on $\mathcal{M}$. This can hide complexity; moreover, with this dependence, the same statement could be made, but omitting the lower-order term in the bound: with that omission, the unstated asymptotic threshold be different, but that is the only difference. To make the lower-order term meaningful, therefore, we need to be more precise, which unfortunately means more complicated, with the definition of a few more technical quantities:

- The value $\omega(\mathcal{M})$ is a threshold for the curvature-based distance D_{II}, related to $\mu_{\mathrm{II}}(\mathcal{M})$: when $D_{\mathrm{II}}(a,b) \leq \omega(\mathcal{M})$, then $\min_{y\in a+T_a}\|b-y\| \leq \beta D_{\mathrm{II}}(a,b)^2$ for some constant $\beta > 0$ (§2.3 on page 10). The analogous threshold for D_{I}, such that $D_E(a,b) = \|a-b\| \leq D_{\mathrm{I}}(a,b)$, is not needed, since the inequality always holds;
- The values $\psi_X(\mathcal{M},\epsilon)$, for $X \in \{\mathrm{I},\mathrm{II},\mathrm{III}\}$, which are $O(1)$ as $\epsilon \to 0$, relate the size $C_X(\mathcal{M},\epsilon)$ of an ϵ-net in the metric D_X to an asymptotic expression $\mu_X(\mathcal{M})/\epsilon^d$ for that $C_X(\mathcal{M},\epsilon)$ (§2.4 on page 12).

With these quantities defined, it is possible to state the bounds using *absolute asymptotic constants*: that is, the threshold and constant factor of the asymptotic $O()$ notation are absolute constants, and so in particular are independent of $\mathcal{M}$.

Theorem 1.6. *A connected, compact, orientable, differentiable manifold $\mathcal{M}$ is ϵ-embedded with probability at least $1-\delta$ by a k-map with*

$$k = O(\mathcal{L}(d,\epsilon,\delta)) + O(\epsilon^{-2}(\log(\mu_{\mathrm{II}}\psi_{\mathrm{II}}^d(1/\tau^d + 1/\omega^d) + \mu_{\mathrm{III}}\psi_{\mathrm{III}}^d)))$$

as $\epsilon \to 0$, with absolute asymptotic constants, under the assumption that $\psi_X(\mathcal{M},\epsilon)$ is increasing in ϵ for $X \in \{\mathrm{II}, \mathrm{III}\}$. The dependence of some quantities on $\mathcal{M}$ and ϵ is omitted for brevity. Here as before

$$\mathcal{L}(d,\epsilon,\delta) := \epsilon^{-2}(d\log(1/\epsilon) + \log(1/\delta)).$$

If τ is not needed, that is, $\tau(\mathcal{M},\epsilon) = \mathrm{diam}(\mathcal{M})$, then the term depending on $\mu_{\mathrm{II}}(\mathcal{M})$ is not needed. If only geodesic distances on $\mathcal{M}$ need be approximately preserved, a k-map with

$$k = O(\epsilon^{-2}(\log(1/\delta) + \log(\mu_{\mathrm{II}}\psi_{\mathrm{II}}^d(1/\tau^d + 1/\omega^d) + \mu_{\mathrm{III}}\psi_{\mathrm{III}}^d/\epsilon^d))$$

suffices. Analogous statements hold with "I" replacing "II" in the above, where $\omega(\mathcal{M})$ can be omitted.

This theorem is proved in §4.4 on page 22; it implies Theorem 1.5 on the previous page. As mentioned, a key lemma is that the existence of families of small generalized ϵ-covers (as defined in Section 2 on the following page) implies ϵ-isometry results for k-mappings. Section 2 also discusses the measures μ_X, for $X \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$, and the associated distance metrics D_X. The ϵ-isometry results can be applied to $\mathbf{U}(\mathcal{M}-\mathcal{M})$, and as mentioned, the short-enough chords can be approximated by tangent vectors. (As a reminder: the short/long distinction comes from the length of vector in $\mathcal{M}-\mathcal{M}$, before it is normalized to be in $\mathbf{U}(\mathcal{M}-\mathcal{M})$.) The long chords can be handled more directly: Lemma 4.1 on page 17 says that for an ϵ-cover N of $\mathcal{M}$, the set $\mathbf{U}(N-N)$ is an ϵ-cover of long chords of $\mathbf{U}(M-M)$. As mentioned, in §4.3, the properties needed to show that the threshold $\tau(\mathcal{M},\epsilon)$ can be useful are discussed.

While the use of μ_{II} gives the strongest results, it may seem unnatural. This is particularly so when it is extended to submanifolds of dimension or codimension not equal to one; then the corresponding metric D_{II} is not Riemannian in general (as defined here; see §2.3 on page 10). If μ_{II} is not palatable, the similar, but weaker result can be obtained using $\mu_{\mathrm{I}}(\mathcal{M})$, and analogous related quantities, instead of $\mu_{\mathrm{II}}(\mathcal{M})$. (This possibility may also be more general, as discussed in Section 5 on page 23.)

There are at least a few simple examples where the results here improve on Baraniuk and Wakin [BW06], beyond the removal of m from the bounds. For example, for a cylinder $\mathcal{M}$ of some given radius and length, their bound involves the length, since that quantity figures in the volume $\mu_{\mathrm{I}}(\mathcal{M})$. (We assume that the "caps" of the cylinder are closed off in a reasonable way.) The results here, in contrast, give a bound that is independent of the length; the bounds use the existence of a good approximation for $\mathcal{M}$ involving tangent flats that are close to $\mathcal{M}$ over a long distance; the analog for simplices would be approximation by very long and skinny simplices. This also shows why using μ_{II} in the main theorem is stronger than using μ_{I}. (This example involves a slight falsehood: the bound is dependent on an arbitrarily small, but fixed, multiple of the length, as implied by the perturbation discussed in §2.3 on page 10.) For a manifold that is the graph of a quadratic function, the results here show that a sufficient projection dimension depends only on the total absolute curvature $\mu_{\mathrm{III}}(\mathcal{M})$, and not on the surface area of $\mathcal{M}$, in contrast to prior work.

Some recent work using Baraniuk and Wakin's bound [BW06] for learning tasks relating to manifolds is correspondingly improved by the results here [HWB07].

As observed by Baraniuk and Wakin [BW06], various properties of a manifold are also preserved approximately if its pairwise Euclidean and geodesic distances are preserved: its dimension, topology, local neighborhoods, local angles, path curvatures, and surface area, for example.

A different direction of generalization of the JL Lemma is to show similar embedding results for mappings that are not projections to a random linear subspace, for example, the mapping that is multiplication by a $d \times m$ matrix of i.i.d. normal random variables. (Such variations are reviewed by Sarlós [Sar06], for example.) Since the theorem above ultimately relies only on embedding results for families of finite sets of points, and on the linearity of the embeddings, analogous results can be proved for those variant embedding techniques.

1.1 Approximation

For values x and y, and $\beta > 0$, let $x \leq_\beta y$ denote the condition $x \leq (1+\beta)y$, and let $x \approx_\beta y$ indicate that $x \leq_\beta y$ and $y \leq_\beta x$ both hold. When $x \leq_\beta y \leq_\beta z$, for some value z, it follows that $x \leq (1+\beta)^2 z$, and thus $x \leq_{3\beta} z$ for $\beta < 1$. So a version of the relation holds transitively, up to a constant factor. The β subscript may be sometimes dropped from $\approx_\beta$, and such book-keeping understood.

2 Covers and Manifold Metrics

Some basic facts about metric spaces, ϵ-covers, ϵ-packings, ϵ-nets, and associated measures will be needed. Given a metric space $(\mathbb{U}, D)$ and a subset $S \subset \mathbb{U}$, recall that an ϵ-cover of S is a subset $C \subset \mathbb{U}$ such that for any point $a \in S$, there is some $a' \in C$ with $D(a, a') \leq \epsilon$. Let $C(S, D, \epsilon)$ denote the minimum size of of an ϵ-cover for S. An ϵ-packing is a subset $N \subset \mathbb{U}$ such that no two points $a', a'' \in N$ have $D(a', a'') < \epsilon$. An ϵ-net for S is both an ϵ-cover and an ϵ-packing. It is a nearly optimal ϵ-cover, in the sense that if the size of an ϵ-net is no larger than the size of any $(\epsilon/2)$-cover. The space (S, D) has box dimension d when $C(S, D, \epsilon) = 1/\epsilon^{d+o(1)}$ as $\epsilon \to 0$, and the Hausdorff measure associated with D is approximately $C(S, D, \epsilon)/\epsilon^d$, as $\epsilon \to 0$, as discussed further in §2.4 on page 12.

It will be helpful to generalize the concept of an ϵ-cover from sets of points to sets of regions: that is, a *generalized ϵ-cover* of S is here a collection of subsets $N := \{F \mid F \subset \mathbb{U}\}$ such that for any $a \in S$, there is $a' \in F \in N$

such that $D(a, a') \le \epsilon$. Given a collection N of subsets of $\mathbb{R}^m$, let $N - N := \{F - F' \mid F, F' \in N\}$, where as before $F - F' := \{x - x' \mid x \in F, x' \in F'\}$. Also for N a collection of subsets, let $\mathbf{U}(N) := \{\mathbf{U}(F) \mid F \in N\}$.

As sketched above, two such generalized ϵ-covers that will be used for manifolds are: a collection of tangent flats that approximate the manifold, bounded using μ_{II}, and a collection of tangent subspaces that approximate all the tangent subspaces of the manifold, bounded using μ_{III}.

Both kinds of generalized cover are derived from ϵ-nets of $\mathcal{M}$, with respect to two different distance measures on $\mathcal{M}$, called D_{II} and to D_{III}. These distance measures, and the measure D_{I}, are discussed in the next subsection.

The distances D_{I}, D_{II}, and D_{III} on manifolds are central to the constructions here. In contrast to these distances, the Euclidean distance between two points $x, y \in \mathbb{R}^m$ will be denoted $\|x-y\|$, while the minimum distance between a point p and a set, such as a tangent space T, will be denoted $D_E(p, T)$.

The notation D_X, for $X \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$, derives from the relation of these distances to the corresponding *fundamental forms*, which generally correspond to the metric tensors of these distances.

2.1 Distance D_{I}

Given a manifold $\mathcal{M} \subset \mathbb{R}^m$ and $a, b \in \mathcal{M}$, the geodesic distance between a and b is denoted here by $D_{\mathrm{I}}(a, b)$; this is the length of the shortest path on $\mathcal{M}$ between a and b, where the metric tensor on vector v is simply v^2; that is, the length of a very short path is the Euclidean distance in $\mathbb{R}^m$ between its endpoints. Note that always $D_{\mathrm{I}}(a, b) \le \|a - b\|$.

2.2 Distance D_{III}, Grassmann manifolds, principal angles

Another metric on $\mathcal{M}$ that will be needed is the curvature-based distance $D_{\mathrm{III}}(a, b)$, which is the length of the shortest curve between a and b in $\mathcal{M}$, where the measure of the curve $\mathcal{C}$ is the length of the *Gauss map* image $\mathbf{N}(\mathcal{C})$ of $\mathcal{C}$. The Gauss map $\mathbf{N}$ takes a point $a \in \mathcal{M}$ to its tangent subspace T_a, which can also be regarded as a point in $\mathcal{G}_{d,m}$, the *Grassmann manifold* of d-dimensional linear subspaces of $\mathbb{R}^m$. (More precisely, the tangent subspace and the Grassmann manifold are in general oriented, but here the assumption of manifold orientability, which is nearly without loss of generality, using the *double cover* construction, allows this to be glossed over.) That is, for $a, b \in \mathcal{M}$ that are very close together, the distance $D_{\mathrm{III}}(a, b) \approx D_{\mathrm{I}}(T_a, T_b)$, where $T_a, T_b \in \mathcal{G}_{d,m}$. For example, when $\mathcal{M}$ is a curve, the tangent subspaces are lines through the origin, and so (the oriented version of) $cG_{1,m}$ is identifiable with $\mathbb{S}^{m-1}$. Similarly, when $\mathcal{M}$ has codimension one, the Grassmann manifold can be identified with $\mathbb{S}^{m-1}$, by mapping from an $(m-1)$-dimensional tangent subspace to its unit normal.

More generally, the distance between two nearby $T_a, T_b \in \mathcal{G}_{d,m}$ is a function of the *principal angle* vector between T_a and T_b. The principal angle vector

$\theta \in \mathbb{R}^d$ can be described as follows: $\cos\theta$ is the vector of singular values of $Y_a^T Y_b$, where Y_a and Y_b are an $m \times d$ orthonormal bases for T_a and T_b. (Here $\cos\theta$ means the vector whose i'th coordinate is $\cos\theta_i$.) The arclength between T_a and T_b in $\mathcal{G}_{d,m}$ is $\|\theta\|$. The matrix $Y_a Y_a^T$ projects points of $\mathbb{R}^m$ onto T_a, and similarly for $Y_b Y_b^T$. The *projection 2-norm* $\|Y_a Y_a^T - Y_b Y_b^T\|_2$ is a measure of distance between T_a and T_b, related to the principal angles by $\|Y_a Y_a^T - Y_b Y_b^T\| = \|\sin\theta\|_\infty$. (The matrix norm $\|\|_2$ here is $\|M\|_2 := \sup_{x \in \mathbb{R}^d} \|Mx\|/\|x\|$.) Note that

$$\|\sin\theta\|_\infty \leq \|\sin\theta\| \leq \|\theta\|, \tag{1}$$

so the projection 2-norm $\|Y_a Y_a^T - Y_b Y_b^T\|_2$ is always bounded above by the Grassmannian arc length between T_a and T_b, which is bounded above by $D_{\mathrm{III}}(a, b)$. The relations among metrics on Grassmann manifolds are discussed in detail by Edelman *et al.* [EAS99].

A key property of D_{III} needed here follows from these relations. It implies that an ϵ-cover for D_{III} can be used to obtain a generalized ϵ-cover for all unit tangent vectors.

Lemma 2.1. *For $a, b \in \mathcal{M}$ and unit vector $v \in T_a$, there is a unit vector $v' \in T_b$ such that $\|v - v'\| \leq D_{\mathrm{III}}(a, b)$.*

Proof. Using the notation and discussion just above, since $\|Y_a Y_a^T - Y_b Y_b^T\|_2 \leq D_{\mathrm{III}}(a, b)$, it follows by definition of the matrix norm and by (1) that

$$\begin{aligned} \|v - Y_b Y_b^T v\| &= \|(Y_a Y_a^T - Y_b Y_b^T)v\| \leq D_{\mathrm{III}}(a, b)\|v\| \\ &= D_{\mathrm{III}}(a, b), \end{aligned}$$

and so the lemma follows, with $v' = Y_b Y_b^T v$. ☐

Section B on page 27 gives an explicit description of D_{III} and its metric tensor, in terms of the local Taylor expansions of charts of the manifold.

2.3 Distance D_{II}

The distance metric $D_{\mathrm{II}}(a, b)$ is the length of the shortest path between a and b in $\mathcal{M}$, where the length of a very short path is roughly the square root of the absolute value of the curvature at a in the direction of b. For manifolds whose dimension and codimension are greater than one, the metric D_{II} as defined here is not Riemannian, and so it has no metric tensor as such. To define this metric and understand its properties, some preliminaries are needed, and discussed next.

Coordinates oriented to $a \in \mathcal{M}$ Given a point a on a d-manifold $\mathcal{M}$ in $\mathbb{R}^m$, translate the coordinate system so that a is at the origin, and rotate the coordinate system so that the tangent d-flat $a + T_a$ to $\mathcal{M}$ at a is $T_a := \{x \in \mathbb{R}^m | x_{d+1} = x_{d+2} = \ldots x_m = 0\}$. With this change, $a + T_a$ is naturally identified with $\mathbb{R}^d$. Let $\phi : \mathbb{R}^d \to \mathbb{R}^m$ be a coordinate chart of $\mathcal{M}$ at a, so that we can equivalently take the domain of ϕ to be a neighborhood of the origin in $T_a = \mathbb{R}^d$. To expand ϕ in Taylor series about zero, we can consider each coordinate of $\phi(x)$ as a separate function $\phi_i : \mathbb{R}^d \to \mathbb{R}$. Such a function has Taylor expansion

$$\phi_i(x) = \phi_i(0) + \nabla\phi_i(0)^T x + \frac{1}{2} x^T \nabla^2 \phi_i(0) x + O(\|x\|^3).$$

Stacking the $\nabla\phi_i(0)^T$ as rows of an $n \times d$ matrix G, and letting $H_i := \nabla^2\phi_i(0)$, we obtain

$$\phi(x) = \phi(0) + Gx + \frac{1}{2} \begin{bmatrix} x^T H_1 x \\ x^T H_2 x \\ \vdots \\ x^T H_m x \end{bmatrix} + O(\|x\|^3). \tag{2}$$

By the translation, $\phi(0) = 0$, and by a further rotation in $\mathbb{R}^d$, $G = \left[\begin{smallmatrix} D \\ 0 \end{smallmatrix}\right]$, where D is a $d \times d$ diagonal matrix. By an appropriate change of variables for ϕ, we can assume that D is in fact the identity matrix I. (For $d = 1$, this says that the tangent vector is a unit vector, that is, the parameterization is "unit speed.")

D_{II}, and Covering Manifolds by Tangent Flats We will define the D_{II} distance near a using the expansion (2): let A_i denote the matrix whose eigenvalues are the square root of the absolute values of the eigenvalues of H_i. let $q_{\mathrm{II}}(x; a)$ be defined by

$$q_{\mathrm{II}}(x; a) := \left[\sum_{d \le i \le m} \|A_i x\|^4 \right]^{1/2}. \tag{3}$$

Then the D_{II}-length of a curve $\mathcal{C}$ on $\mathcal{M}$ can be defined as $\int_{\mathcal{M}} \sqrt{q_{\mathrm{II}}(f'(t); f(t))}dt$, where f is a unit-speed parameterization of $\mathcal{C}$, and $D_{\mathrm{II}}(a, b)$ is then defined as the infinum of the D_{II}-lengths of curves connecting a and b.

The motivation for this definition is that for a given point $\phi(x) \in \mathcal{M}$ close to a, for a chart ϕ oriented to a as above, its distance to $a + T_a$ satisfies

$$\begin{aligned} D_E(\phi(x), a + T_a)^2 &= \sum_{d \le i \le m} \phi_i(x)^2 \\ &\approx \sum_{d \le i \le m} (x^T H_i x)^2. \end{aligned} \tag{4}$$

By choosing the A_i as described, we have that $D_E(\phi(x), a+T_a)$ is dominated by right hand side above; also, $\sqrt{q_{\mathrm{II}}(x; a)}$ is a seminorm on x, via the seminorms $\|A_i x\|$, $i = d \ldots m$, and the ℓ_4-norm on the vector of those $\|A_i x\|$ values. (Here a seminorm satisfies the properties of a norm, except that some nonzero vectors may have seminorm zero.) This implies that D_{II}, as defined, is a *pseudometric*, that is, a function that satisfies the conditions of a metric, except that some distinct points may be at D_{II}-distance zero from each other.

For the approximate condition of (4) on the previous page to hold, strictly speaking it's necessary to augment q_{II} so that the ϕ_i terms are dominated even when all $x^T H_i x$ are zero. This can be done by replacing the $\|A_i x\|^4$ terms in (3) on the preceding page by $\|A_i x\|^4 + \eta\|x\|^2$, where η is an arbitrarily small, but fixed, value greater than zero. The result is that for $\|x\|$ small enough, the approximation of (4) holds. The result of this perturbation is that $\mu_{\mathrm{II}}(\mathcal{M})$ will have an arbitrarily small, but fixed, additive term behaving like $\mu_{\mathrm{I}}(\mathcal{M})$. A positive side effect is that D_{II} becomes a metric.

With the given definition of D_{II}, we have the following lemma.

Lemma 2.2. *There is a value $\omega(\mathcal{M}) > 0$ and absolute constant $\beta > 0$ so that for any $a, a' \in \mathcal{M}$, if $D_{\mathrm{II}}(a, a') \leq \omega(\mathcal{M})$, then $D_E(a', a + T_a) \leq_\beta D_{\mathrm{II}}(a, a')^2$.*

Proof. By construction, the asserted inequality holds for a' in a sufficiently small neighborhood of a given point a. Moreover, there is a neighborhood V of a that is the image of a single chart ϕ. For a point $a' \in V$ close enough to a, the projection $\mathbf{P}_a x'$ of a tangent vector $x' \in T_{a'}$ onto T_a yields a value for the metric tensor $q_{\mathrm{II}}(\mathbf{P}_a x'; a)$. For all points a' close enough, by the smoothness of $\mathcal{M}$ this value is within a constant factor of $q_{\mathrm{II}}(x'; a')$. Thus $D_{\mathrm{II}}(a, a')^2$ is within a constant factor of of $q_{\mathrm{II}}(\phi^{-1}a'; a)$, which for close enough a' is within a constant factor of the distance of a' to T_a.

Since for every $a \in \mathcal{M}$, there is a neighborhood V' of a within which the asserted inequality $D_E(a', a+T_a) \leq_2 D_{\mathrm{II}}(a, a')^2$ holds, it follows by the Lebesgue Number Lemma that there is a value ω such that this condition holds for *any* a, a' with $D_{\mathrm{II}}(a, a') \leq \omega(\mathcal{M})$. □

Note that while H_i may not be positive semidefinite, A_i^2 always is, so that for any x, $\|A_i x\|^2 = x^T A_i^2 x \geq x^T H_i x$, so $D_{\mathrm{II}}(a, b)$ may be much larger than $D_E(b, a + T_a)$.

The D_{II} metric is related to curvature in several ways: when $d = 1$ and $D_{\mathrm{II}}(q, a)$ is small, it is approximately the integral of the square root of the curvature from a to q. When $d = m - 1$ and $D_{\mathrm{II}}(q, a)$ is small, it is proportional to the square root of the normal curvature at q in the direction $a - q$. In general, the D_{II}-length of a curve from $a \in \mathcal{M}$ to $\phi(\alpha x)$, for $x \in T_a$ and $\alpha = 0 \ldots \delta$, is approximately equal to δ times the square root of the directional curvature at a in the direction x.

In the extreme cases of dimension or codimension equal to one, the function (3) on the previous page is a quadratic form, and so D_{II} is a Riemannian metric. Otherwise, D_{II} is only a *Finsler* metric.

2.4 Measures, Net Sizes, and $\psi_X(\mathcal{M}, \epsilon)$

Each distance measure D_X, for $X \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$, has an associated Hausdorff measure μ_X, where $\mu_{\mathrm{I}}(\mathcal{M})$ is the d-dimensional measure (surface area) of $\mathcal{M}$, and $\mu_{\mathrm{III}}(\mathcal{M})$ is its total absolute curvature. The corresponding measure $\mu_{\mathrm{II}}(\mathcal{M})$ is the integral over $\mathcal{M}$ of the square root of the absolute value of the Gaussian curvature, for $\mathcal{M}$ with dimension or codimension one. For a manifold of codimension one, there is a triangulation comprising d-simplices whose Hausdorff distance to $\mathcal{M}$ is $O(\mu_{\mathrm{II}}(\mathcal{M})/\epsilon^{2/d})$ as $\epsilon \to 0$ [Cla06], and this is asymptotically tight for a large class of manifolds, under some mild restrictions. The triangulation vertices are a $\sqrt{\epsilon}$-net of $\mathcal{M}$ with respect to D_{II}. Here that ϵ-net construction and approximation of the manifold is extended to manifolds of higher codimension, but the approximation is not with respect to Hausdorff distance, but a weaker one, by a generalized cover comprising tangent flats. The existence of such a covering is then used to prove bounds for random projections.

We may write D_X for $X \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$, and correspondingly μ_X, and so on, or drop the suffix when the ambiguity is harmless. Rather than write $C(\mathcal{M}, D_X, \epsilon)$, we will generally write $C_X(\mathcal{M}, \epsilon)$, or when the intention is clear, omit the subscript.

For small enough ϵ, $\mu(\mathcal{M})$ (referring to any of the three) is within some $2^{\Theta(d)}$ of $C(\mathcal{M}, \epsilon)\epsilon^d$, but this need not be true for large ϵ: for example, for a curve $\mathcal{C}$ in $\mathbb{R}^d$, consider the boundary $\mathcal{C}' := \partial(\mathcal{C} + B_r)$of the Minkowski sum $\mathcal{C} + B_r$, where B_r is a d-dimensional ball of radius r centered at the origin, and $r > 0$. If $r \ll \epsilon$, then the points of an ϵ-net of $\mathcal{C}'$ will be scattered along the length of $\mathcal{C}$, and the length $\mu_{\mathrm{I}}(\mathcal{C}) \approx \epsilon C_{\mathrm{I}}(\mathcal{C}', \epsilon)$, but the surface area $\mu_{\mathrm{I}}(\mathcal{C}')$ is proportional to $\mu_{\mathrm{I}}(\mathcal{C})r^{d-1}$. (Note that the measures μ_{I} are d-dimensional for the d-manifold $\mathcal{C}'$, but 1-dimensional for the 1-manifold $\mathcal{C}$.) A related fact here is that for a point $a \in \mathcal{C}'$ and the set $B(a, \epsilon)$ of points $p \in \mathcal{C}'$ with $D_{\mathrm{I}}(p, a) \le \epsilon$, the surface area $\mu_{\mathrm{I}}(B(a, \epsilon)) \approx r^{d-1}\epsilon \ll \epsilon^d$: a ball of such a radius ϵ in $\mathcal{C}'$) has much smaller measure than a ball in $\mathbb{R}^d$ of the same radius.

To avoid the need for reference to "small enough ϵ," which could also be viewed as not counting lower-order asymptotic terms that depend on $\mathcal{M}$, while keeping to the philosophy of accounting for all dependence on $\mathcal{M}$, let $\psi_X(\mathcal{M}, \epsilon)$ be the "fudge factor" relating $C(\mathcal{M}, \epsilon)$ and $\mu(\mathcal{M})/\epsilon^d$:

$$\psi(\mathcal{M}, \epsilon) := \left[\frac{C(\mathcal{M}, \epsilon)}{\mu(\mathcal{M})/\epsilon^d} \right]^{1/d}.$$

Thus $\psi(\mathcal{M}, \epsilon) = O(1)$ as $\epsilon \to 0$. Results will be expressed in terms of $\mu(\mathcal{M})$, ϵ, and $\psi(\mathcal{M}, \epsilon)$, instead of simply in terms of $C(\mathcal{M}, \epsilon)$, so that the asymptotic dependence will be clearer, but the non-asymptotic, large ϵ situation will still be accounted for.

For convenience of calculation, we will often assume that $\psi(\mathcal{M}, \epsilon)$ is nonincreasing as $\epsilon \to 0$. (Alternatively, the quantity $\hat{\psi}(\mathcal{M}, \epsilon) := \sup_{\epsilon' \le \epsilon} \psi(\mathcal{M}, \epsilon)$ could be used, which by construction is nonincreasing as $\epsilon \to 0$.)

As defined, this factor simply translates from the net size $C(\mathcal{M}, \epsilon)$ to the asymptotically equivalent $\mu(\mathcal{M})/\epsilon^d$, but it is possible to relate $C(\mathcal{M}, \epsilon)$ to some known quantities nonasymptotically, under mild assumptions. Such a relation for $C(\mathcal{M}, \epsilon)$ then implies a relation for $\psi(\mathcal{M}, \epsilon)$.

Roughly, the size of an ϵ-net depends on what fraction of the manifold an average point will cover, and the latter is the *correlation integral* of the manifold. More formally, suppose for an ϵ-net $N \subset \mathcal{M}$ that the Voronoi regions $V_a \subset \mathcal{M}$

of $a \in N$ all have about the same measure $\mu(V_a) \approx \mu(\mathcal{M})/|N|$. Suppose also that balls of radius ϵ centered at points in V_a have about the same measure on average as $B(a,\epsilon)$, that is, $\mu(B(a,\epsilon)) \approx \int_{V_a} \mu(B(b,\epsilon))d\mu(b)/\mu(V_a)$. The latter condition is implied by *Ahlfors regularity*, that balls of the same size have about the same measure, but plainly is weaker. Under these assumptions,

$$\begin{aligned}
\mu(\mathcal{M}) &\leq \sum_{a\in N} \mu(B(a,\epsilon)) \\
&\approx \sum_{a\in N} \mu(B(a,\epsilon))\mu(V_a)|N|/\mu(\mathcal{M}) \\
&\approx |N| \sum_{a\in N} \int_{V_a} \mu(B(b,\epsilon))d\mu(b)/\mu(\mathcal{M}) \\
&= \frac{|N|}{\mu(\mathcal{M})} \int_{\mathcal{M}} \mu(B(b,\epsilon))d\mu(b),
\end{aligned}$$

that is, for some $\beta \geq 0$,

$$\frac{1}{|N|} \leq_\beta \frac{1}{\mu(\mathcal{M})^2} \int_{\mathcal{M}} \mu(B(b,\epsilon))d\mu(b).$$

A similar relation with $\int_{\mathcal{M}} \mu(B(b,\epsilon/2))d\mu(b)$ holds in the other direction, using $\mu(\mathcal{M}) \geq \sum_{a\in N} \mu(B(a,\epsilon/2))$.

That is, under some reasonable "smoothness" conditions, $C(\mathcal{M},\epsilon)$ can be naturally expressed within $2^{\Theta(d)}$ of $\int_{\mathcal{M}} \mu(B(b,\gamma))d\mu(b)/\mu(\mathcal{M})^2$, the *correlation integral*, which is the probability that two points randomly chosen from $\mathcal{M}$ according to μ will be closer than γ. (The normalization by $\mu(\mathcal{M})^2$ is needed to obtain a probability.) Here $\gamma = \epsilon/2$ and $\gamma = \epsilon$ provide bounds for $C(\mathcal{M},\epsilon)$ and so for $\psi(\mathcal{M},\epsilon)$. The correlation integral thus encodes the situation of manifolds that are d-dimensional but are roughly lower-dimensional, as in the above example. Such manifolds have a small injectivity radius, and the correlation integral reflects that. Bounds could also be expressed in terms of the injectivity radius itself, but that may be unduly pessimistic.

The *correlation dimension* characterizes the asymptotic behavior of the correlation integral as $\epsilon \to 0$; the random projection technique has been proposed as a method for speeding up the estimation of the the correlation dimension [HWB07].

3 Sets of Bounded Box Dimension

The lemma below considers approximate isometries of subsets of the unit sphere that have bounded box dimension, or more generally subsets that have generalized ϵ-covers of size $1/\epsilon^{d+o(1)}$ as $\epsilon \to 0$.

Lemma 3.1. *Let $S \subset \mathbb{S}^m$ have a family $N(S,\epsilon)$ of generalized ϵ-covers of size $C(S,\epsilon)$, with respect to the Euclidean metric. Suppose for convenience that*

$$C(S,\epsilon') \leq C(S,\epsilon)(\epsilon/\epsilon')^d$$

for $\epsilon' \leq \epsilon$. Then S is ϵ-isometrized with probability at least $1-\delta$ by a k-map where

$$k = O(\epsilon^{-2} c_{\text{JL}} \log((\exp(d) + f_N) C(S,\epsilon)/\delta)),$$

as $\epsilon \to 0$, with absolute asymptotic constants. Here f_N must be an upper bound on the failure multiplier for isometrizing any member of any $N(S,\epsilon)$, and also of the square root of the failure multiplier needed for isometrizing $F - F'$, for $F, F' \in N(S,\epsilon)$.

For finite sets, this result is uninteresting: for small enough ϵ, the promised k-map has $k \geq |S|$, and so the result is trivially true.

The lemma is more interesting for an infinite set S: then for finite $S' \subset S$ with $d/\log|S'|$ sufficiently small, the results here are better than those for JL as applied directly to S'.

When the family $N(S,\epsilon)$ is a family of (nongeneralized) ϵ-covers, then the condition $C(S,\epsilon) = 1/\epsilon^{d+o(1)}$ is the definition of S having box dimension d. Here the monotonicity assumption $C(S,\epsilon') \leq C(S,\epsilon)(\epsilon/\epsilon')^d$ for $\epsilon' \leq \epsilon$ is simply a strong form of the box dimension condition, assumed in the setting of generalized ϵ-covers. The monotonicity assumption could be removed by phrasing the results in terms of $\hat{C}(S,\epsilon) := \sup_{\epsilon' \leq \epsilon} C(S,\epsilon')(\epsilon')^d/\epsilon^d$, which has $\hat{C}(S,\epsilon)\epsilon^d$ monotone nonincreasing as $\epsilon \to 0$.

In the applications here of this lemma, the generalized covers are always collections of d-flats or $2d$-flats, so f_N is always implied by subspace JL, Lemma 1.3 on page 3.

As mentioned in the introduction, this proof is a slight extension of one by Agarwal *et al.* [AHPY07], which in turn is inspired by ideas of Indyk and Naor [IN07].

Proof. Let N_i be a sequence of ϵ_i-covers of S, for $i = 0, \ldots, \infty$, where $\epsilon_i \to 0$ as $i \to \infty$, but otherwise to be determined. For a given point $a \in S$, let $u_i(a)$ denote the closest point in any member of N_i to a, so that $u_i(a) \to a$ as $i \to \infty$. Since

$$a = u_0(a) + \sum_{i \geq 0} u_{i+1}(a) - u_i(a),$$

we have, for linear projection $\mathbf{P}$,

$$\mathbf{P}a = \mathbf{P}u_0(a) + \sum_{i \geq 0} \mathbf{P}u_{i+1}(a) - \mathbf{P}u_i(a),$$

and so

$$\|\mathbf{P}a\| \leq \|\mathbf{P}u_0(a)\| + \sum_{i \geq 0} \|\mathbf{P}(u_{i+1}(a) - u_i(a))\|, \tag{5}$$

remarking that this includes the trivial case where the sum diverges to infinity. We will consider the probabilities δ' and δ_i that for some β and β_i to be chosen, a

k-map $\mathbf{P}$ β-isometrizes N_0, and also $(\beta_i - 1)$-isometrizes the sets in the collection $N_i - N_{i+1}$, which are

$$N_i - N_{i+1} = \{F - F' \mid F \in N_i, F' \in N_{i+1}\},$$

as this was defined in Section 2 on page 8. Under the assumption of these isometries,

$$\begin{aligned}\|\mathbf{P}(u_{i+1}(a) - u_i(a))\| &\le \beta_i \|u_{i+1}(a) - u_i(a)\| \\ &\le \beta_i(\|u_{i+1}(a) - a\| + \|u_i(a) - a\|) \\ &\le \beta_i(\epsilon_{i+1} + \epsilon_i),\end{aligned}$$

which implies by (5) on the preceding page

$$\|\mathbf{P}a\| \le 1 + \beta + \sum_{i\ge 0} \beta_i(\epsilon_{i+1} + \epsilon_i). \tag{6}$$

Now choose $\beta := \epsilon/2$, $\beta_i := 4^{i+1}$, and $\epsilon_i := \epsilon/8^i/18$ for $i \ge 0$. Assuming that k can be chosen so that a k-map $\mathbf{P}$ satisfies these conditions,

$$\begin{aligned}\|\mathbf{P}a\| &\le 1 + \epsilon/2 + \sum_{i\ge 0} \frac{4^{i+1}}{18}(\epsilon/8^{i+1} + \epsilon/8^i) \\ &= 1 + \epsilon/2 + \epsilon\frac{1}{18}\sum_{i\ge 0}(1/2^{i+1} + 4/2^i) \\ &= 1 + \epsilon(1/2 + \frac{1}{18}(1+8)) \\ &= 1 + \epsilon.\end{aligned}$$

It remains to bound the probability of failure. Suppose

$$k \ge K \log(\hat{f}_N C(S,\epsilon)/\delta)c_{\text{JL}}/\epsilon^2$$

for a value K to be determined, and where $\hat{f}_N := \exp(d) + f_N$. Then from the definition of "failure multiplier," and the union bound, the probability δ', that the k-map $\mathbf{P}$ fails to β-isometrize N_0, satisfies

$$\begin{aligned}\delta' &\le C(S,\epsilon_0)\hat{f}_N \exp(-k(\epsilon/2)^2/c_{\text{JL}}) \\ &\le C(S,\epsilon)18^d \hat{f}_N(\delta/\hat{f}_N C(S,\epsilon))^{K/4}, \\ &\le 18^d \delta^{K/4}(\hat{f}_N C(S,\epsilon))^{1-K/4},\end{aligned}$$

which for $K > 4(1 + \ln 18)$, $\hat{f}_N C(S,\epsilon) > 2\exp(d)$, and $\delta < 1/2$ is less than $\delta/2$. Also

$$\begin{aligned}\delta_i &\le C(S,\epsilon_i)C(S,\epsilon_{i+1})\hat{f}_N^2 \exp(-k(4^{i+1}-1)^2/c_{\text{JL}}) \\ &\le C(S,\epsilon_i)C(S,\epsilon_{i+1})\hat{f}_N^2(\hat{f}_N C(S,\epsilon)/\delta)^{-K((4^{i+1}-1)/\epsilon)^2} \\ &\le 18^2 \delta^{K((4^{i+1}-1)/\epsilon)^2} 8^{d(2i+1)}(\hat{f}_N C(S,\epsilon))^{2-K((4^{i+1}-1)/\epsilon)^2}\end{aligned}$$

which satisfies $\delta/2 \geq \sum_i \delta_i$ under the same conditions on K, $\hat{f}_N C(S,\epsilon)$, and δ, and with also $\epsilon \leq 2$. (Regarding ϵ, recall that S is a collection of unit vectors, so any given point is a 2-net.) Thus a k-map $\mathbf{P}$ fails the needed conditions with probability at most $\delta' + \delta/2 \leq \delta$.

A lower bound on $\|\mathbf{P}a\|$, with similar failure probability, can be proved in a similar way, and the lemma then follows with some adjustment of constants. $\square$

4 Manifolds

For a given manifold $\mathcal{M}$, consider the following subset of $\mathbf{U}(\mathcal{M}-\mathcal{M})$:

$$\mathbf{U}_\lambda(\mathcal{M}-\mathcal{M}) := \left\{ \frac{a-b}{\|a-b\|} \;\middle|\; a, b \in \mathcal{M}, \|a-b\| > \lambda \right\},$$

the chords of $\mathcal{M}$ that are longer than λ, scaled to have unit length. As discussed above, if a linear map is an approximate isometry for $\mathbf{U}(\mathcal{M}-\mathcal{M})$, then it embeds $\mathcal{M}$. If $\mathbf{U}(\mathcal{M}-\mathcal{M})$ has bounded box dimension, then the above theorem implies an embedding of $\mathcal{M}$. Thus, we would like to know the sizes of (generalized) ϵ-covers of $\mathbf{U}(\mathcal{M}-\mathcal{M})$. An ϵ^2-cover for $\mathcal{M}$ implies an ϵ-cover for $\mathbf{U}_{4\epsilon}(\mathcal{M}-\mathcal{M})$, as shown in Lemma 4.1. When points $a, a' \in \mathcal{M}$ are close together, on the other hand, $\frac{a'-a}{\|a'-a\|}$ approaches a unit tangent vector to $\mathcal{M}$ at a, and so for sufficiently small λ, an ϵ-cover for $\mathbf{U}(\mathcal{M}-\mathcal{M}) \setminus \mathbf{U}_\lambda(\mathcal{M}-\mathcal{M})$ can be obtained from a cover for the tangent spaces of $\mathcal{M}$. Together, these facts will be used to show that Lemma 3.1 on page 14 can be applied to $\mathbf{U}(\mathcal{M}-\mathcal{M})$ to show that k-maps are embeddings, with high probability.

4.1 Long Chords

Lemma 4.1. *For a set $S \subset \mathbb{R}^m$, and given $\epsilon, \gamma > 0$, if N is a generalized $\epsilon\gamma$-cover of S with respect to Euclidean distance, then $\mathbf{U}(N-N)$ is a generalized ϵ-cover for $\mathbf{U}_{4\gamma}(S-S)$.*

If the number of members of N is bounded by $C(S,\epsilon\gamma)$, then the number of members of $\mathbf{U}(N-N)$ is bounded by $C(S,\epsilon\gamma)^2$.

Proof. Suppose N is a generalized $\epsilon\gamma$-cover of S, and $a, a' \in S$ have $b \in F_a \in N$ and $b' \in F_{a'} \in N$, respectively, closest in N. Then since $\|a-b\| \leq \epsilon\gamma$, and similarly $\|a'-b'\| \leq \epsilon\gamma$, we have $\|(a-a')-(b-b')\| \leq 2\epsilon\gamma$, and so also

$\|b-b'\|-\|a-a'\|$ is smaller than $2\epsilon\gamma$ in magnitude. Using also that $\|a-a'\| > 4\gamma$,

$$\begin{aligned}
&\left\| \frac{a-a'}{\|a-a'\|} - \frac{b-b'}{\|b-b'\|} \right\| \\
&= \left\| \frac{(a-a')-(b-b')}{\|a-a'\|} + \frac{(b-b')(\|b-b'\|-\|a-a'\|)}{\|a-a'\|\|b-b'\|} \right\| \\
&\le \left\| \frac{(a-a')-(b-b')}{\|a-a'\|} \right\| + \left\| \frac{(b-b')(\|b-b'\|-\|a-a'\|)}{\|a-a'\|\|b-b'\|} \right\| \\
&\le \frac{2\epsilon\gamma}{4\gamma} + \frac{2\epsilon\gamma}{4\gamma} \\
&\le \epsilon,
\end{aligned}$$

and the lemma follows, since $\frac{b-b'}{\|b-b'\|} \in \mathbf{U}(F_a - F_{a'}) \in \mathbf{U}(N-N)$. □

This lemma can be applied to a generalized covering of $\mathcal{M}$ by a collection of tangent flats of $\mathcal{M}$, using Lemma 4.1 on the previous page and Lemma 2.2 on page 12.

Lemma 4.2. *Let $\mathcal{M}$ be a compact differentiable manifold, and let $\gamma, \epsilon > 0$, and $\hat{\epsilon} := \min\{\omega(\mathcal{M}), \sqrt{\epsilon\gamma}/2\}$. Then for an absolute constant $\beta > 0$, there is a generalized $\beta\epsilon$-cover for $\mathbb{U}_\gamma(\mathcal{M}-\mathcal{M})$, comprising at most $C_{\mathrm{II}}(\mathcal{M},\hat{\epsilon})^2$ $(2d)$-flats.*

Proof. A $\hat{\epsilon}$-cover N of $\mathcal{M}$ with respect to the D_{II} metric has size $C_{\mathrm{II}}(\mathcal{M},\hat{\epsilon})$, and by definition, for any $b \in \mathcal{M}$, there is $a \in N$ such that $D_{\mathrm{II}}(a,b) \le \hat{\epsilon}$. Since $\hat{\epsilon} \le \omega(\mathcal{M})$, it follows by Lemma 2.2 on page 12 that for such a and b, $D_E(b, a+T_a) \le_\beta \hat{\epsilon}^2$, for some absolute constant β; that is, $N' := \{a+T_a \mid a \in N\}$ is a generalized $\beta\epsilon\gamma/4$-cover of $\mathcal{M}$. By Lemma 4.1 on the preceding page, $\mathbf{U}(N'-N')$ is a generalized $\beta\epsilon$-cover of $\mathbb{U}_\gamma(\mathcal{M}-\mathcal{M})$. □

4.2 Very Short Chords

As the distance $\|a-b\|$ between $a, b \in \mathcal{M}$ goes to zero, the vector $a-b$ looks more and more like a tangent vector to $\mathcal{M}$ at a (or b). This suggests that to approximate short chords $a-b$, it is necessary, at the minimum, to approximate the unit tangent vectors to $\mathcal{M}$. The size of an ϵ-net for these vectors can be bounded using the measure of the image of $\mathcal{M}$ under the *Gauss map*, which maps a point $a \in \mathcal{M}$ to its d-dimensional tangent subspace T_a, a member of the Grassmann manifold $\mathcal{G}_{d,m}$.

Lemma 4.3. *There is a generalized ϵ-cover for the set of unit tangent vectors to the smooth manifold $\mathcal{M} \subset \mathbb{R}^m$, of size $C_{\mathrm{III}}(\mathcal{M},\epsilon)$. Each member of the cover is a d-dimensional linear subspace of $\mathbb{R}^m$.*

Proof. Let N be an ϵ-cover of $\mathcal{M}$, with respect to the metric D_{III}. Then by Lemma 2.1 on page 10, for any $b \in \mathcal{M}$, there is some $a \in N$ so that for any unit $v' \in T_b$, there is unit $v \in T_a$ such that $\|v - v'\| \le D_{\mathrm{III}}(a,b) \le \epsilon$. Thus $N' := \{T_a \mid a \in N\}$ is a generalized ϵ-cover for all unit tangent vectors of $\mathcal{M}$. □

4.3 Embedding All Chords

The above has shown that all long chords and all very short chords can be approximated (and hence embedded) economically. It remains to consider chords of intermediate length. The main question is, which chords are close to tangent vectors? The local curvature gives one answer.

Lemma 4.4. *For $a, b \in \mathcal{M}$, if there is a curve $\mathcal{C} \subset \mathcal{M}$ between a and b whose total curvature $\mu_{\mathrm{III}}(\mathcal{C})$ is no more than ϵ, then the tangent to $\mathcal{C}$ at b has angle no more than ϵ with $a - b$.*

Proof. From the bound on the total curvature of $\mathcal{C}$, the tangent v_a to $\mathcal{C}$ at a has angle at most ϵ with the tangent v_b to $\mathcal{C}$ at b. The angle of $a - b$ to v_b is maximized if the turning from v_b occurs near to b; that is, if $a - b$ is close to parallel to v_a. So the angle of $a - b$ is at most the angle of v_a to v_b, which is at most ϵ. □

For any given target total curvature $\kappa > 0$, there is a distance threshold L so that that for $a, b \in \mathcal{M}$, if $\|a - b\| < L$, then there is a curve $\mathcal{C}$ connecting a and b with $\mu_{\mathrm{III}}(\mathcal{C}) \leq \kappa$. The threshold L thus gives one way to separate "short enough" chords, that can be approximated using tangent vectors, from "long enough" chords, that can be approximated using $\mathbf{U}(N - N)$ for a γ-cover N.

4.3.1 Planar and Low-Torsion Connectivity

The threshold value L is partly a function of the *reach* $\rho(\mathcal{M})$ of $\mathcal{M}$, the shortest distance such that some point of $\mathbb{R}^m$ has two distinct points of $\mathcal{M}$ as closest points in $\mathcal{M}$. The reach is also related to the maximum curvature of $\mathcal{M}$, which could be used to bound L. However, it is also possible to show that there is a tangent vector parallel to $a - b$, under conditions that do not depend on maximum curvature of $\mathcal{M}$, or its reach. The following lemma is an example; it involves a *planar curve in* $\mathcal{M}$, defined here as a curve that is contained in $\mathcal{M}$ and also in a plane (a 2-flat).

Lemma 4.5. *Suppose points $a, b \in \mathcal{M}$ are connected by a planar curve $\mathcal{C} \subset \mathcal{M}$. Then there is some point $c \in \mathcal{C}$ such that the tangent space T_c at c contains a vector parallel to $a - b$.*

Proof. Orient the plane containing the given curve $\mathcal{C}$ so that $a - b$ is contained in the horizontal axis, and consider the maximum and minimum vertical coordinates of $\mathcal{C}$. One of these must be nonzero, by the smoothness of $\mathcal{M}$ (and hence of the curve $\mathcal{C}$), and so there must be a horizontal tangent vector to $\mathcal{C}$, at some point $c \in \mathcal{M}$. This tangent vector is in T_c, where $c \in \mathcal{M}$, and is parallel to $a - b$, and so is the promised vector. □

Thus if there is a distance λ such that all chords of $\mathcal{M}$ shorter than λ are connected by a planar curve, then an ϵ-net for the tangent vectors of $\mathcal{M}$ is also an ϵ-net for all chords of $\mathcal{M}$ shorter than λ.

Such a planar curve connectivity condition holds when the neighborhood of a is a (pure) quadric surface. This condition does not rely on the shortness of $a-b$, except for b to be in that quadric neighborhood, nor does it rely on the distance to other parts of $\mathcal{M}$, or on the local flatness of $\mathcal{M}$.

Lemma 4.6. *Suppose d-manifold $\mathcal{M}$ has the form*

$$\mathcal{M} = \left\{ a \in \mathbb{R}^m \;\middle|\; a = \phi(x), x \in \mathbb{R}^d \right\},$$

for a mapping $\phi : \mathbb{R}^d \to \mathbb{R}^m$, where for all $i = 1 \ldots m$, the i'th coordinate $\phi_i(x)$ is a quadratic function

$$\phi_i(x) = z_i + x^T g_i + \frac{1}{2} x^T H_i x,$$

for $z_i \in \mathbb{R}$, $g_i \in \mathbb{R}^d$, and $d \times d$ symmetric matrix H_i. Then every chord $a-b$ for $a, b \in \mathcal{M}$ is parallel to a vector v tangent to $\mathcal{M}$.

Proof. Given $a, b \in \mathcal{M}$, let $x := \phi^{-1}(a)$ and $y := \phi^{-1}(b)$, the curve $\mathcal{C} := \{\phi(\alpha x + (1-\alpha)y) \mid \alpha \in [0,1]\}$ has coordinates

$$\begin{aligned}
\phi_i(y + \alpha(x-y)) &= [z_i + y^T g_i + y^T H_i y/2] \\
&\qquad + \alpha(x-y)^T(g_i + H_i y) + \alpha^2 (x-y)^T H_i (x-y)/2 \\
&=: \hat{z}_i + \alpha \hat{g}_i + \alpha^2 h_i
\end{aligned}$$

for suitable values $\hat{z}_i$, $\hat{g}_i$, and h_i, that are coordinates of $\hat{z}, \hat{g}, h \in \mathbb{R}^m$. That is, the curve $\mathcal{C}$ can be expressed as $\left\{ \hat{z} + \alpha \hat{g} + \alpha^2 h \;\middle|\; \alpha \in [0,1] \right\}$. This curve is contained in the plane that contains $\hat{z}$, $\hat{g}$, and h, and so is planar. Thus there is a point $c \in \mathcal{C}$ so that the tangent vector to $\mathcal{C}$ at c is parallel to $a-b$. (In fact, that point c is $\phi((x+y)/2)$.) $\square$

As noted in Theorem 1.6 on page 7, bounds for embedding can be simplified for a pure quadric surface: there is a bound for the projection dimension k that depends only on the total absolute curvature $\mu_{\mathrm{III}}(\mathcal{M})$, since all chords can be approximated tangent vectors. It's also worth mentioning that a pure quadric surface sits in a flat of dimension $O(d^2)$, and so a projection dimension $k = O(d^2/\epsilon^2)$ suffices to ϵ-embed it, with high probability [Sar06, AHPY07].

Turning to a more approximate setting: if there is a curve connecting a and b that is nearly planar, then the curve has a tangent vector nearly parallel to $a-b$. To prove this, a characterization of "nearly planar" is needed, starting with this lemma relating the angle between the osculating planes at two nearby points on a curve with the torsion and curvature at those points.

Lemma 4.7. *Given a curve $\mathcal{C} = \{\phi(x) \mid x \in [0 \ldots \mu_{\mathrm{I}}(\mathcal{C})]\}$ with a unit-speed parameterization ϕ, the Grassmannian arc length between the osculating planes at two points $\phi(x)$ and $\phi(x+\delta)$ is at most $\delta\tau/\kappa(1+O(\delta))$ as $\delta \to 0$, where τ is the torsion τ at $\phi(x)$ and κ is the curvature at $\phi(x)$.*

(In the lemma and below, $\mu_{\mathrm{I}}(\mathcal{C}))$ denotes the length of $\mathcal{C}$, that is, its 1-dimensional measure and not its d-dimensional measure.)

Proof. Please see Section A on page 26 $\square$

Motivated by this lemma, for a curve $\mathcal{C}$ define the *total torsion ratio* $\hat{\tau}(\mathcal{C})$ to be the integral over $\mathcal{C}$ of ratio of the torsion to the curvature. From the above lemma, the Grassmannian arc length between the osculating planes at points $w, w' \in \mathcal{C}$ is at most $\hat{\tau}(\mathcal{C})$.

Lemma 4.8. *For a d-manifold $\mathcal{M}$, suppose $a, b \in \mathcal{M}$ are connected by a curve $\mathcal{C} \subset \mathcal{M}$ such that the total torsion ratio $\hat{\tau}(\mathcal{C}) \leq \epsilon^2 \|a-b\|/\mu_{\mathrm{I}}(\mathcal{C})$. Then there is a tangent vector t to $\mathcal{C}$ that is within angle 3ϵ of $a-b$, for ϵ smaller than an absolute constant.*

Proof. For a point $w \in \mathcal{C}$, let $h(w)$ be the osculating plane (through the origin) that, by definition contains the tangent vector to $\mathcal{C}$ at w, and also contains the acceleration vector to $\mathcal{C}$ at w. Here, as above, a unit-speed parametrization of $\mathcal{C}$ is assumed, so that the tangent and acceleration vectors at each point of $\mathcal{C}$ are orthogonal. Let $\mathbf{P}_w : \mathbb{R}^m \to h(c)$ be the projection onto $h(w)$.

We will need a fact about the projections $\mathbf{P}_w$: by hypothesis, $\hat{\tau}(\mathcal{C}) \leq \epsilon^2\|a-b\|/\mu_{\mathrm{I}}(\mathcal{C}) \leq \epsilon^2$, and so by Lemma 4.7 on the preceding page, the Grassmannian distance from $h(w)$ to $h(b)$ is at most ϵ^2. As discussed in Section 2 on page 8, that Grassmannian distance bounds the *projection 2-norm* distance between $\mathbf{P}_w$ and $\mathbf{P}_b$; as applied to any vector y, this implies that

$$\|\mathbf{P}_w y - \mathbf{P}_b y\| \leq \epsilon^2. \tag{7}$$

The idea of the proof is to apply Lemma 4.5 on page 19 to $\mathbf{P}_b\mathcal{C}$, implying that there is some unit tangent vector $\hat{v}$ of $\mathbf{P}_b\mathcal{C}$ at some $\hat{c} \in \mathbf{P}_b\mathcal{C}$ that is parallel to $\mathbf{P}_b(a-b)$, and to show (i) that the angle of $\mathbf{P}_b(a-b)$ to $a-b$ is no more than 2ϵ, and also (ii) that the angle of $\hat{v}$ to the tangent vector v of $\mathcal{C}$ at $c := \mathbf{P}_b^{-1}(\hat{c})$ is $O(\epsilon^2)$. The combination of these two claims implies the result, for small enough ϵ. (The choice of $\mathbf{P}_b$ as a projection was somewhat arbitrary: any $\mathbf{P}_w$ for $w \in \mathcal{C}$ would do.)

For claim (ii): we apply (7) to point $c \in \mathcal{C}$ and its unit tangent vector v; this implies that $\|\mathbf{P}_c v - \mathbf{P}_b v\| \leq \epsilon^2$. Since v is the tangent vector to $\mathcal{C}$ at c, $\mathbf{P}_c v = v$, and so $\|v - \mathbf{P}_b v\| \leq \epsilon^2$. Also $\mathbf{P}_b v$ is parallel to the tangent vector $\hat{v}$ to $\mathbf{P}_b\mathcal{C}$ at $\hat{c}$, by the linearity of the projection, and its length is of $1 - O(\epsilon^2)$. So the angle of v to $\hat{v}$ is $O(\epsilon^2)$.

For claim (i), let $n := a - b - \mathbf{P}_b(a-b)$ be the component of $a-b$ normal to $h(b)$. Since n is normal to $h(b)$, and the tangent to $\mathcal{C}$ at b is in $h(b)$, motion along $\mathcal{C}$ at b has no component in the direction of n. Moreover, applying (7) to vector n, we have that at point $b \in \mathcal{C}$, and at any point $w \in \mathcal{C}$, $\|P_w n - P_b n\| = \|P_w n\| \leq \|n\|\hat{\tau}(\mathcal{C})$, or $\|P_w n/\|n\|\| \leq \hat{\tau}(\mathcal{C})$. Thus, in walking along $\mathcal{C}$ from b to a, the travel in the direction of n is at most $\hat{\tau}(\mathcal{C})$, per unit distance traveled. The length of $\mathcal{C}$ is $\mu_{\mathrm{I}}(\mathcal{C})$, and $\mathcal{C}$ travels $\|n\|$ in the direction of n in going from b to a. It follows, using the lemma hypothesis, that

$$\|n\| \leq \hat{\tau}(\mathcal{C})\mu_{\mathrm{I}}(\mathcal{C}) \leq \frac{\epsilon^2\|a-b\|}{\mu_{\mathrm{I}}(\mathcal{C})}\mu_{\mathrm{I}}(\mathcal{C}) = \epsilon^2\|a-b\|.$$

Letting θ denote the angle between $a-b$ and its projection $\mathbf{P}_b(a-b)$ onto $h(b)$, it follows that

$$\cos\theta = \frac{(a-b)\cdot(a-b-n)}{\|a-b\|\|a-b-n\|} \geq \frac{(a-b)^2 - n\cdot(a-b)}{(a-b)^2} \geq 1-\epsilon^2,$$

and so $\theta \leq 2\epsilon$ for ϵ small enough.

This completes the proof of claim (i), and so with (ii) and the argument above, the lemma follows. □

4.4 The Main Result

It is now possible to prove the main result of this paper, tying together the bounds for isometrizing long chords and short chords. To do this, we need the quantity $\tau(\mathcal{M},\epsilon)$; as defined just before Theorem 1.5, this is the largest Euclidean distance τ such that $\|a-b\| \leq \tau$ implies that there is a curve $\mathcal{C} \subset \mathcal{M}$ connecting a and b so that either the total curvature of $\mathcal{C}$ is less than ϵ, or the total torsion ratio $\hat{\tau}(\mathcal{C}) \leq \epsilon^2\|a-b\|/\mu_{\mathrm{I}}(C)$. The motivation for this definition, and its immediate consequence, is that for $a, b \in \mathcal{M}$ with $\|a-b\| \leq \tau$, the vector $a-b$ must be within angle $O(\epsilon)$ of a tangent vector to a; this holds either via Lemma 4.4 on page 19 or Lemma 4.8 on the preceding page.

Theorem 1.6, restated. *A connected, compact, orientable, differentiable manifold $\mathcal{M}$ is ϵ-embedded with probability at least $1-\delta$ by a k-map with*

$$k = O(\mathcal{L}(d,\epsilon,\delta)) + O(\epsilon^{-2}(\log(\mu_{\mathrm{II}}\psi_{\mathrm{II}}^d(1/\tau^d + 1/\omega^d) + \mu_{\mathrm{III}}\psi_{\mathrm{III}}^d))$$

as $\epsilon \to 0$, with absolute asymptotic constants, under the assumption that $\psi_X(\mathcal{M},\epsilon)$ is increasing in ϵ for $X \in \{\mathrm{II}, \mathrm{III}\}$. The dependence of some quantities on $\mathcal{M}$ and ϵ is omitted for brevity. Here as before

$$\mathcal{L}(d,\epsilon,\delta) := \epsilon^{-2}(d\log(1/\epsilon) + \log(1/\delta)).$$

If τ is not needed, that is, $\tau(\mathcal{M},\epsilon) = \mathrm{diam}(\mathcal{M})$, then the term depending on $\mu_{\mathrm{II}}(\mathcal{M})$ is not needed. If only geodesic distances on $\mathcal{M}$ need be approximately preserved, a k-map with

$$k = O(\epsilon^{-2}(\log(1/\delta) + \log(\mu_{\mathrm{II}}\psi_{\mathrm{II}}^d(1/\tau^d + 1/\omega^d) + \mu_{\mathrm{III}}\psi_{\mathrm{III}}^d/\epsilon^d))$$

suffices. Analogous statements hold with "I" replacing "II" in the above, where $\omega(\mathcal{M})$ can be omitted.

Proof. We show that $\mathbf{U}(\mathcal{M}-\mathcal{M})$ has generalized ϵ-covers of an appropriate size, so that Lemma 3.1 on page 14 can be usefully applied. To construct such covers, we consider covers for the long chords, in $\mathbf{U}_{\tau(\mathcal{M},\epsilon)}(\mathcal{M}-\mathcal{M})$, and also for the remaining short chords. (As in the theorem statement, dependencies on $\mathcal{M}$ and ϵ may be omitted.)

Lemma 4.3 on page 18 says that there is a generalized ϵ-cover for the unit tangent vectors of $\mathcal{M}$, of size $C_{\mathrm{III}}(\mathcal{M},\epsilon)$. For $a,b\in\mathcal{M}$, if $\|a-b\|\le\tau$, then as discussed just before the theorem statement, there is some unit tangent vector v of angle $O(\epsilon)$ with $a-b$, or in other words $\|v-\frac{a-b}{\|a-b\|}\|=O(\epsilon)$. Thus there is a generalized $O(\epsilon)$-cover of the short chords, of size $C_{\mathrm{III}}(\mathcal{M},\epsilon)$, with absolute asymptotic constants, and the cover comprises d-dimensional linear subspaces.

Applying Lemma 4.2 on page 18 with $\gamma=\tau$, there is a generalized $O(\epsilon)$-cover of size $C_{\mathrm{II}}(\mathcal{M},\hat\epsilon)^2$ for the long chords, where $\hat\epsilon:=\min\{\omega(\mathcal{M}),\sqrt{\epsilon\tau}/2\}$, and the members of the cover are $2d$-flats.

So $\mathbf{U}(\mathcal{M}-\mathcal{M})$ has a generalized $O(\epsilon)$-cover of size

$$\begin{aligned}(8)\qquad & O(C_{\mathrm{II}}(\mathcal{M},\hat\epsilon)^2)+O(C_{\mathrm{III}}(\mathcal{M},\epsilon))\\ &\qquad =O(\mu_{\mathrm{II}}\psi_{\mathrm{II}}^d/\hat\epsilon^d)^2+\mu_{\mathrm{III}}\psi_{\mathrm{III}}^d/\epsilon^d),\end{aligned}$$

plugging in the definitions of the ψ_X functions.

This argument applies for any $\epsilon>0$, and by the theorem hypothesis, $\psi_X(\mathcal{M},\epsilon)$ is decreasing as $\epsilon\to 0$, for $X\in\{\mathrm{II},\mathrm{III}\}$, implying the monotonicity assumption of Lemma 3.1 on page 14 for $C(S,\epsilon')$ and $\epsilon'\le\epsilon$, where $S=\mathbf{U}(\mathcal{M}-\mathcal{M})$. Hence Lemma 3.1 on page 14 can be applied, with $C(S,\epsilon)$ bounded by (8), and with f_N taken as $\exp(4c_{\mathsf{JLs}}d)$.

We have

$$\begin{aligned}(9)\qquad k&=O(\epsilon^{-2}(\log((\exp(d)+f_N)(O(\mu_{\mathrm{II}}\psi_{\mathrm{II}}^d/\hat\epsilon^d)^2\\ &\qquad +\mu_{\mathrm{III}}\psi_{\mathrm{III}}^d/\epsilon^d)/\delta)/c_{\mathsf{JL}})\\ &=O(\mathcal{L}(d,\epsilon,\delta))\\ &\qquad +O(\epsilon^{-2}\log(\mu_{\mathrm{II}}\psi_{\mathrm{II}}^d(1/\tau^d+1/\omega^d)+\mu_{\mathrm{III}}\psi_{\mathrm{III}}^d)),\end{aligned}$$

as $\epsilon\to 0$, with absolute asymptotic constants, as claimed. Neglecting the $\psi_X=O(1)$ terms, this is

$$O(\mathcal{L}(d,\epsilon,\delta))+O(\epsilon^{-2}\log(\mu_{\mathrm{II}}(1/\tau^d+1/\omega^d)+\mu_{\mathrm{III}})).$$

For preservation of geodesic distances, it is enough to ϵ-embed short chords, and also to (1/2)-embed long chords so that the projected version of the manifold is not self-intersecting. This implies the same cover with respect to D_{III}, but a smaller cover with respect to D_{II}, the latter having no dependence on ϵ. The appropriate simplification of (9) yields the claim in the theorem statement.

For replacement of D_{II} by D_{I}, an $\hat\epsilon$-cover of $\mathcal{M}$ with respect to D_{I} can be used instead of the generalized $\hat\epsilon$-cover by D_{II}, to obtain the cover of long chords. □

5 Concluding Remarks

A natural question here is the extension of these results to polyhedral manifolds, part of which might involve the generalization of the D_X distance measures.

While D_{III} and D_{I} have natural non-smooth analogs, this doesn't seem to be true for D_{II}. Perhaps ϵ-nets in the hybrid metric $D_{\mathrm{I+III}}$ could be used; such a construction would generalize Dudley's construction and *isophotic* triangulations [Cla06, Dud74, PSH+04].

Another natural question regards lower bounds: do these results give a sharp characterization of the projection dimension? Unfortunately not: a curve of arbitrarily high curvature has a small projection dimension if it is planar. More generally, the bounds here are not monotone as a function of set inclusion: the bounds for a helix may be large, while those for a cylinder containing the helix are small. However, these situations are "special," and non-generic, at least heuristically.

Acknowledgment I am grateful to the SOCG referees for their careful comments.

References

[AC06] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *Proc. 38th STOC*, pages 557–563, 2006. 1

[AHPY07] P. K. Agarwal, S. Har-Peled, and H. Yu. Embeddings of surfaces, curves, and moving points in Euclidean space. In *SOCG '07: Proc. Symp. on Computational Geometry*, 2007. 1, 1, 3, 4.3.1

[BW06] R. Baraniuk and M. Wakin. Random projections of smooth manifolds. to appear, *Foundations of Computational Mathematics*, 2006. 1, 1, 1.4, 1

[Cla06] K. L. Clarkson. Building triangulations using ϵ-nets. In *STOC 2006: Proceedings of the Thirty-eighth Annual SIGACT Symposium*, 2006. 1, 2.4, 5

[Dud74] R. M. Dudley. Metric entropy of some classes of sets with differentiable boundaries. *J. Approximation Theory*, 10:227–236, 1974. 5

[EAS99] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1999. 2.2, A, B, B

[Gru93] P. M. Gruber. Asymptotic estimates for best and stepwise approximation of convex bodies I. *Forum Math.*, 5:281–297, 1993. 1

[HWB07] C. Hegde, M. B. Wakin, and R. G. Baraniuk. Random projections for manifold learning. In *Neural Information Processing Systems (NIPS)*, December 2007. 1, 2.4

[IM04] P. Indyk and J. Matoušek. *Low-distortion embeddings of finite metric spaces*, pages 177–196. CRC Press, 2004. 1

[IN07] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. *ACM Trans. Algorithms*, 3(3), 2007. 1, 1, 3

[Ind00] P. Indyk. Dimensionality reduction techniques for proximity problems. In *Symposium on Discrete Algorithms*, pages 371–378, 2000. 1

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mapping into Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984. 1, 1.1, 1.2

[Lin02] N. Linial. Finite metric spaces: combinatorics, geometry, and algorithms. In *International Congress of Mathematicians III*, pages 573–586, 2002. 1

[Mag02] A. Magen. Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications. To appear. Preliminary version in RANDOM 2002, 2002. 1

[PSH+04] H. Pottmann, T. Steiner, M. Hofer, C. Haider, and A. Hanbury. The isophotic metric and its application to feature sensitive morphology on surfaces. In T. Pajdla and J. Matas, editors, *Computer Vision — ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 560–572. Springer, 2004. 5

[Sar06] T. Sarlós. Improved approximation algorithms for large matrices via random projections. In *FOCS2006: Proc. 47th Annual IEEE Symp. on Foundations of Computer Science*, pages 143–152, 2006. 1, 1, 1.3, 1, 4.3.1

[Ste91] G. W. Stewart. Perturbation theory for the singular value decomposition. In R. Vaccaro, editor, *SVD and Signal Processing, II: Algorithms, Analysis and Applications*, pages 99–109. Elsevier, 1991. A, B

[Vem04] S. Vempala. *The Random Projection Method*, volume 65 of *Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, 2004. 1

[Wey12] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf der Theorie der Hohlraumstrahlung). *Mathematische Annalen*, 71:441–479, 1912. A, B

A Grassmannian Arc Length Between Osculating Planes

Before a proof of Lemma 4.7 on page 20, here for convenience is a restatement of it.

Lemma 4.7 restatement *Given a curve $\mathcal{C} = \{\phi(x) \mid x \in [0 \ldots \mu_{\mathrm{I}}(\mathcal{C})]\}$ with a unit-speed parameterization ϕ, the Grassmannian arc length between the osculating planes at two points $\phi(x)$ and $\phi(x+\delta)$ is at most $\delta\tau/\kappa(1+O(\delta))$ as $\delta \to 0$, where τ is the torsion τ at $\phi(x)$ and κ is the curvature at $\phi(x)$.*

Proof. The tangent vectors $\phi'(x)$ to a unit speed curve have $\|\phi'(x)\| = 1$ for all relevant x; this implies that the acceleration vector $\phi''(x)$ is orthogonal to $\phi'(x)$. The osculating plane at x is the linear span of $\phi'(x)$ and $\phi''(x)$, and so is the column space of the matrix $Y_1 := [\, \phi'(x)\ \phi''(x) \,]$. Without loss of generality, via Euclidean rotation we can assume $\phi'(x) = [1, 0, \ldots, 0]^T$, $\phi''(x) = [0, \kappa, 0, \ldots, 0]^T$, and so

$$Y_1 = \begin{bmatrix} 1 & 0 \\ 0 & \kappa \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix},$$

where $\kappa = \|\phi''(x)\|$, the curvature at x. Equivalently, in these coordinates the osculating plane at $\phi(x)$ is the column space of the matrix $\hat{Y}_1 := [\, I\ 0 \ldots 0 \,]^T$ with orthonormal columns, where I is here the 2×2 identity matrix. Turning to the osculating plane at $\phi(x+\delta)$, for small $\delta > 0$: the torsion at $\phi(x)$ is $\tau := \|\phi'''(x) - g(x)\|$, where $g(x)$ is the projection of $\phi'''(x)$ onto the osculating plane. Without loss of generality, assume that $\phi'''(x) - g(x) = [0, 0, \tau, 0, \ldots, 0]^T$. Letting κ_1 and κ_2 denote the components of $\phi'''(x)$ in the osculating plane at $\phi(x)$, we have that the osculating plane at $\phi(x+\delta)$ is the column space of

$$Y_2 := \begin{bmatrix} \phi'(x+\delta) & \phi''(x+\delta) \end{bmatrix} = \begin{bmatrix} 1 & \delta\kappa_1 \\ \delta\kappa & \kappa + \delta\kappa_2 \\ 0 & \delta\tau \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} (1 + O(\delta)).$$

(The $O(\delta)$ is as $\delta \to 0$, here and in the remainder of the proof.) As discussed by Edelman *et al.*[EAS99], and in Section 2 on page 8, the cosines of the principal angles between the planes generated by Y_1 and by Y_2 are the singular values of $\hat{Y}_1^T \hat{Y}_2$, where $\hat{Y}_1$ is given above and $\hat{Y}_2$ is a matrix with orthonormal columns and the the same column space as Y_2.

To produce $\hat{Y}_2$, multiply the second column of Y_2 by δ and subtract it from the first column, then multiply the first column by $\delta\kappa_1$ and subtract it from the second, and then normalize each column. The result is that

$$X := \hat{Y}_1^T \hat{Y}_2 = \begin{bmatrix} 1 & O(\delta^2) \\ O(\delta^2) & \frac{(\kappa+\delta\kappa_2)}{\sqrt{(\kappa+\delta\kappa_2)^2+(\delta\tau)^2}} \end{bmatrix} (1+O(\delta)).$$

Weyl [Wey12, Ste91] gave a bound on the perturbation of the singular values of a matrix due to the perturbation of its entries. Weyl's bound implies that the singular values of X are within $\|X - D\|_F = O(\delta^2)$ of the singular values of the diagonal matrix D with the same diagonal entries as X. That is, the singular values of X are $1 + O(\delta^2)$ and

$$\begin{aligned} \frac{(\kappa + \delta\kappa_2))}{\sqrt{(\kappa+\delta\kappa_2)^2 + (\delta\tau)^2}}(1+O(\delta)) &= \frac{1+O(\delta)}{\sqrt{1+(\delta\tau)^2/(\kappa+\delta\kappa_2)^2}} \\ &= (1 - \frac{(\delta\tau)^2}{2\kappa^2})(1+O(\delta)). \end{aligned}$$

These are the cosines of the principal angles between the osculating planes, so those angles are $O(\delta^2)$ and $\delta\tau/\kappa(1+O(\delta))$. The Euclidean norm of the vector comprising these angles is $\delta\tau/\kappa(1+O(\delta))$, and that norm is the Grassmannian arc length between the planes. □

B Explicit Description of D_{III}

The fundamental forms, and other aspects of manifolds, are often discussed in an *intrinsic*, coordinate-free, terminology, which is motivated by the fact that many of these concepts need not depend on a particular choice of coordinate system, and can be defined for manifolds that are not embedded submanifolds of $\mathbb{R}^m$. However, this paper is concerned only with embedded submanifolds, and so as in [EAS99], the following discussion uses the *extrinsic* coordinates of $\mathbb{R}^m$. An advantage of this approach is that little background is needed beyond linear algebra and Taylor's theorem.

Recalling the discussion of Section 2 on page 8 and in particular the Taylor expansion (2) on page 11 of a coordinate chart $\phi(x)$ at $a \in \mathcal{M}$: if the Taylor expansion is differentiated, we obtain an expression for the derivative matrix $G(x)$ of ϕ at x near the origin:

$$G(x) \approx \begin{bmatrix} I \\ 0 \end{bmatrix} + \begin{bmatrix} H_1 x \\ H_2 x \\ \vdots \\ H_m x \end{bmatrix} = G + H(x),$$

where this defines the matrix $H(x)$.

Given the d-subspace T_0 that is the linear span of the columns of G, and the d-subspace $T_{\phi(x)}$ that is the linear span of the columns of $G + H(x)$, as shown below, the squared arc length in the Grassmann manifold $\mathcal{G}_{d,m}$ between T_0 and $T_{\phi(x)}$ is

$$
\begin{aligned}
(1{+}O(\|x\|)) \sum_{d<i\leq n} x^T H_i^2 x \\
&= (1 + O(\|x\|) x^T \left[\sum_{d<i\leq n} H_i^2 \right] x \\
&= (1 + O(\|x\|) q_{\mathrm{III}}(x; a),
\end{aligned}
$$

where $q_{\mathrm{III}}(x;a) := x^T \left[\sum_{d<i\leq n} H_i^2\right] x$, the *third fundamental form* at a, evaluated at the tangent vector x. (Here $H_i^2 := H_i^T H_i = H_i H_i$, remembering that each H_i is a Hessian matrix and hence symmetric.) For example, if $d = 1$, then x and each H_i are scalars, and the vector $\left[\begin{smallmatrix} 0 & H_3 & \cdots & H_m \end{smallmatrix}\right]^T$ is the normal component of the acceleration, whose norm is the normal curvature. For a hypersurface, that is, where the codimension is one, the function $x^T H_m^2 x$ is the third fundamental form at a, evaluated at x. In general, the third fundamental form gives a measure of the turning of the tangent space from a to $T_{\phi(x)}$, and the Riemannian metric D_{III} induced by this form gives a measure of winding angle.

Claim B.1. *The squared Grassmannian arc length between the subspace spanned by the columns of G and the subspace spanned by the columns of $G + H(x)$ is $(1 + \|x\|_\infty)\|\bar{H}(x)\|_F^2$, where $\bar{H}(x)$ is the $(m-d) \times d$ matrix that is the lower $m - d$ rows of $H(x)$.*

Proof. As discussed by, for example, Edelman *et al.* [EAS99], if the principal angles between two subspaces generated by the columns of orthonormal matrices Y_1 and Y_2 are θ_i, $i = 1 \ldots d$, then the values $\cos(\theta_i)$ are the singular values of $Y_1^T Y_2$. (Here *orthonormal* means that the columns are orthogonal to each other, and all have unit Euclidean norm.) Since G^T is $\begin{bmatrix} I \\ 0 \end{bmatrix}$, it is orthonormal, and multiplication by it simply selects the upper $d \times d$ submatrix of $G + H(x)$. Since the entries of $H(x)$ are $O(\|x\|_\infty)$ as $\|x\|_\infty \to 0$, the columns of $G + H(x)$ are nearly orthonormal, but not quite. We can do elementary column operations and scaling on those columns, to obtain an orthonormal matrix, and not affect the subspace the columns generate. We will do such operations, and at the same time obtain a matrix whose upper square submatrix has very small off-diagonal elements. As a result, the singular values of that upper square submatrix, selected by G^T, will be provably close to those diagonal entries.

Let $B := G + H(x)$, and let B' denote the upper square submatrix of B. For $i = 1 \ldots d$, use column operations to zero out all but the diagonal entries of the upper square submatrix B' of B. Then scale the columns so that the diagonal entries of B' are equal to one. Then, use Gram-Schmidt orthogonalization to make the columns of B orthonormal: that is, for $i = 1 \ldots d$, subtract from column b_i the vector $b_j(b_j^T b_i)$, for all $j < i$, and then normalize b_i.

The effects of these operations on the entries of B can be bounded. Let $\alpha := \|x\|_\infty$. The entries of $G+H(x)$ are all $O(\alpha)$, as $\alpha \to 0$, except the diagonal entries (due to G), which are $1 + O(\alpha)$. The elimination operations involve multiplication by a term that is $O(\alpha)$, and so entries that are not eventually zero are perturbed by $O(\alpha^2)$. A similar claim holds for the Gram-Schmidt orthogonalization.

During the orthogonalization step, the off-diagonal entries of B' become nonzero. However, because they were zero before the step, the dot products used for the orthogonalization are $O(\alpha^2)$, and so the off-diagonal entries of B' are $O(\alpha^2)$, as are the perturbations to the diagonal entries.

The above operations are repeated. Since the off-diagonal entries of B' are $O(\alpha^2)$, the Gram-Schmidt orthogonalizations are done with vectors that are orthogonal, up to a perturbation of $O(\alpha^2)$ due to the elimination step. Hence their dot products are $O(\alpha^4)$, and the nonzero entries of B' are $O(\alpha^4)$.

The diagonal entries of B' are $1/\sqrt{1 + w_i^2(1 + O(\alpha))}$, where w_i is the $(m-d)$-vector comprising the lower $m - d$ entries of $G + H(x)$. Now we can appeal to the bound of Weyl [Wey12, Ste91] on the perturbation of the singular values due to a matrix perturbation. Weyl's bound implies that the singular values of D, the diagonal matrix whose entries are the diagonal entries of B', are within $\|B' - D\|_F = O(\alpha^4)$ of those of B'. Thus the singular values of $G^T B$ are

$$1/\sqrt{1 + w_i^2(1 + O(\alpha))} + O(\alpha^4) = 1/\sqrt{1 + w_i^2} + O(\alpha^3).$$

As discussed above, these are the values of $\cos\theta_i$, where the θ_i are the principal angles between the subspaces generated by the columns of G and of $G + H(x)$.

Since $\cos\theta_i = 1/\sqrt{1 + w_i^2} + O(\alpha^3)$, $\sin^2\theta_i = w_i^2 + O(\alpha^3)$, and so $\sum_i \sin^2\theta_i = \sum_i w_i^2 = \|\bar{H}(x)\|_F^2$, which is the squared "Projection F-norm" between two points in the Grassmannian, and asymptotically equal to the other definitions of squared distance on the Grassmannian, including arc length, as discussed by Edelman *et al.* [EAS99]. The claim follows. □