A Nash-type Dimensionality Reduction for Discrete Subsets of L_2

Yair Bartal* Ben Recht† Leonard J. Schulman‡

Abstract

Pursuing a line of work begun by Whitney, Nash showed that every C^1 manifold of dimension d can be embedded in $\mathbb{R}^{2d+2}$ in such a manner that the local structure at each point is preserved isometrically. We provide an analog of this result for discrete subsets of Euclidean space. For perfect preservation of infinitesimal neighborhoods we substitute near-isometric embedding of neighborhoods of bounded cardinality.

There are a variety of empirical situations, abstracted to metric space tasks, in which small distances are meaningful and reliable, but larger ones are not. Such situations arise in source coding, image processing, computational biology, and other applications, and are the motivation for widely-used heuristics such as Isomap and Locally Linear Embedding.

In such situations we offer the possibility of dimension reductions unobtainable with global methods, because the dimension of our locally $(1+\epsilon)$-distorting embedding is proportional to $\epsilon^{-2}\log k$, where k is the cardinality of the neighborhoods where distances are preserved and is independent of the number of points in the metric space.

One may view our work as a local version of the widely-used Johnson-Lindenstrauss lemma. We use a device of Nash, together with more recent metric embedding methods, to compose local dimensionality reducing embeddings within a global immersion (that preserves short distances), and with mild additional assumptions, a global embedding that also keeps distant points well-separated. We provide an efficient randomized algorithm for the embeddings.

*yair@cs.huji.ac.il. School of Engineering and Computer Science, Hebrew University, Israel and Center for the Mathematics of Information, Caltech, CA, USA. Supported in part by a grant from the Israeli Science Foundation (195/02) and in part by a grant from the National Science Foundation (NSF CCF-065253).

†brecht@ist.caltech.edu. Center for the Mathematics of Information, Caltech, Pasadena, CA 91125.

‡schulman@caltech.edu. Caltech, Pasadena, CA 91125. Supported in part by NSA H98230-06-1-0074 and NSF CCF-0515342.

1 Introduction

In many large-scale data processing applications, local distances convey more useful information than large distances and are sufficient for uncovering low-dimensional structure. Such situations would arise if the large distances are inaccurate or do not reflect the intrinsic geometry of the application. Moreover, there are a variety of situations that rely only on local distances, including nearest-neighbor search, the computation of vector quantization rate-distortion curves [21], and popular data-segmentation and clustering algorithms [35]. In all of these cases, it is often desirable to reduce the dimension of the data set for reductions of storage requirements or algorithm running times. If the long distances are unimportant, we may be able to reduce the dimensionality only preserving the local information, and such reduction can be into a far lower dimension than what is possible when attempting to preserve distances between all pairs of points.

Two overwhelmingly influential papers posited that if a high-dimensional data set lies on the embedding of a low-dimensional Riemannian manifold, the intrinsic dimensionality could then be found by examining only the nearest neighbor distances of the graph. The first algorithm, known as Isomap, uses Dijkstra's algorithm on the nearest neighbors graph to compute the global distances and then applies multi-dimensional scaling to the computed distances to find a low dimensional embedding of the data [37]. The second, Local Linear Embedding, computes the best linear approximation of each set of neighbors, and then stitches the neighborhoods together by solving an eigenvalue problem constraining the mappings of overlapping neighborhoods [33]. Based on these initial results and their accompanying empirical examples, these two papers gave rise to an active field, commonly referred to as *manifold learning*, and the ensuing years have seen a multitude of applications of these algorithms in areas as diverse as protein folding [16], motion planning in robotics [25], data-mining microarray assays [29], and face recognition [23]. All of these applications use the L_2 distance, even if it is not perfectly justified, because of its tractability and empirical power. Moreover, there have been a variety of alternative algorithms proposed to reduce dimensionality nearest neighbor distances problems, employing kernel methods [13], generative probabilistic models [15], semidefinite programming [39] or neural networks [24].

Despite their wide appeal, all of these algorithms assume some sort of manifold model underlies the data, and make implicit assumptions about intrinsic curvature, Riemmanian metrics, or volume. More importantly, not one of these manifold learning algorithms come with any provable guarantees for discrete data sets, and many authors have pointed out that the geometric assumptions of these algorithms are not reasonable in practice. For example, the algorithms are quite sensitive to the determination of neighborhood structure [9], have problems recovering non-convex domains or manifolds with nontrivial homology [18], and cannot recover manifold structures that require more than one coordinate chart [30].

From a more theoretical perspective, the concept of a "local embedding" was first introduced in the context of metric space embedding in [2]. Local embeddings share the same objective as manifold learning: to find a mapping of a metric space into a low-dimensional metric space where distances of close neighbors are preserved more faithfully than those of distant neighbors. The field of metric embedding has been an active field of research both in mathematics and computer science and has emerged as a powerful tool in many algorithmic application areas. Two cornerstone theorems in this field are a theorem of Bourgain [14] stating that that any n-point metric space embeds in L_2 with $O(\log n)$ distortion, and the Johnson-Lindenstrauss [26] dimension reduction lemma which says that any n-point subset of L_2 embeds with $(1+\epsilon)$-distortion in ℓ_2^D where $D = O(\epsilon^{-2}\log n)$. Both these theorems have many algorithmic consequences. Randomized constructions of a variety of these embeddings are given in [5], fast implementations of these embeddings are provided in [6], and deterministic constructions have been proposed in [19, 36].

Abraham, Bartal and Neiman [2] show that many of the known classic embedding results can be extended to the context of local embeddings. In particular, generalizing Bourgain's result (and [1]) they provide local embeddings requiring only $O(\log k)$ dimensions to achieve distortion $O(\log k)$ on the neighborhoods with at most k-points, assuming the the metric obeys a certain *weak growth rate* condition. This number k could have no relation to n, and in practice could be arbitrarily smaller than n. It should be emphasized that this type of embedding is an *immersion*, that is it preserves well the short distances but may arbitrarily distort the long ones. This is reasonable, for instance, if we desire a compact *distance oracle* [38] for close neighbors.

In this paper, we provide a local version of the Johnson-Lindenstrauss Lemma. Such a construction is challenging to achieve because all of the previously discussed algorithms based on the this Lemma require a globally consistent choice of random variables. For this reason, results extending the Johnson-Lindenstrauss Lemma to the projection of smooth manifolds end up depending on the dimension where the manifold is embedded and both the volume and curvature of the manifold [10]. We show that for any $\epsilon > 0$, only $O(\epsilon^{-2}\log(k))$ dimensions are required embedding with distortion $1+\epsilon$ on the neighborhoods with at most k-points, assuming the the metric obeys the weak growth rate condition defined by Abraham, Bartal and Neiman. Furthermore, as been recently addressed by Schechtman and Shraibman [34], the weak growth rate assumption is necessary. For the special case when our set X is isometric to an ultrametric [3] show an embedding in dimension $O(\epsilon^{-3}\log k)$. Our result both improves on this result and extends it to more general metric spaces.

For general metrics, this embedding is an immersion, but under the assumption that the metric has low dimensionality we can transform our immersion into a *global embedding* such that distances between far points can be bounded below so they don't destroy the local structure. This extension to a global embedding is useful in many applications of dimension reduction where it is necessary to maintain the local neighborhoods such as nearest neighbor search. Unlike the results in manifold learning, we make no assumptions that our data lie on some compact manifold, and further assume nothing about the volume or cardinality of our data set.

As an example application where our algorithm ideally suited, the principal computational problem in vector quantization is formally one of clustering (with ℓ_2^2 costs), but the parameters are different than in the classification literature: primarily, one studies here the limit that k (the number of clusters) tends to ∞, while the distortion (the average distance to a codeword) tends to 0. Known algorithms for construction of near-optimal clusterings are exponential in at least one of k or the dimension of the space. Our embedding is well-suited to taking advantage of dimensionality reduction for this problem, since our target dimension depends only on the size of the small regions in which the L_2 distance needs to be preserved.

Our methods owe a substantial debt to seminal papers in several areas of mathematics. Pursuing a line of work begun by Whitney [40, 41], Nash showed that every Riemannian manifold of dimension D could be embedded in $\mathbb{R}^{2D+2}$ by a C^1 mapping such that the metric at each point is preserved isometrically [28]. Nash achieves this embedding using a device which locally perturbs a non-distance preserving embedding provided by Whitney. The randomized trigonometric embedding of Section 3.1 is adapted from Nash's deterministic embedding procedure, and we give a probabilistic analysis showing that with high probability this yields an embedding of the local distances in each neighborhood. As observed in [31] in the context of fast algorithms for pattern recognition, our random trigonometric functions form an embedding into a Euclidean space where the inner product approximates a positive definite shift-invariant kernel function. In our case, we sample frequencies from a Gaussian distribution and use the smoothness properties of the gaussian kernel $k(x,y) = \exp(-\gamma\|x-y\|^2)$ to ensure the quality of our randomized Nash device.

To assemble these embedded neighborhoods into a global immersion (i.e., a mapping which preserves local structure, but may not be injective), we employ probabilistic partitioning [11] of our metric space in

Section 2. These partitions, developed in [1, 2, 3], decompose the metric space into clusters of bounded diameter and allow the coordinates of the embedding to smoothly transition between neighborhoods. With the additional assumption that the metric space has low doubling dimension [8, 22] we ensure in section 3.3 that our mapping is not only injective but also keeps distant points distant by combining recent techniques on embedding metrics in their intrinsic dimension [4] and local scaling distortion embeddings [2]. Finally, we guarantee the existence of our embedding using Lovász's powerful Local Lemma[20], and rely on algorithmic implementations of the LLL by Beck, and Molloy and Reed to provide a randomized algorithm to generate our local embedding [7, 12, 27].

Summary of Results. Given a discrete set of points X of cardinality n in U-dimensional Euclidean space we construct a low dimension local embedding, one that preserves distances to close neighbors with a $1+\epsilon$ multiplicative error. The main result of this paper is summarized by the following theorem.

Let $k \in \mathbb{N}$. For a point $x \in X$ and $r \geq 0$, the ball at radius r around x is defined as $B(x,r) = \{z \in X \mid \|x-z\| \leq r\}$. For a point $x \in X$ let $\Delta_k(x)$ be the largest radius r such that $|B(x,r)| \leq k$. For a pair $x, y \in X$, define: $\Delta_k(x,y) = \max\{\Delta_k(x), \Delta_k(y)\}$.

Theorem 1. *Let $k \in \mathbb{N}$. Given X a discrete subset of $\mathbb{R}^U$, then for any $\epsilon > 0$ there exists an embedding $\hat{\Phi} : X \to \mathbb{R}^D$, where $D = O(\log k/\epsilon^2)$ with the following properties:*

a. $\|\hat{\Phi}(x) - \hat{\Phi}(y)\| \leq (1+\epsilon)\|x-y\|$ *for all* $x,y \in X$

b. For all $x, y \in X$, let $\Delta_k^*(x,y) = c_1\epsilon\Delta_k(x,y)/\log k$, then:

$$\|\hat{\Phi}(x) - \hat{\Phi}(y)\| \geq \begin{cases} (1+\epsilon)^{-1}\|x-y\| & \text{if } \|x-y\| \leq \sqrt{\epsilon}\Delta_k^*(x,y) \\ (1+\epsilon')^{-1}\|x-y\| & \text{if } \|x-y\| = \sqrt{\epsilon'}\Delta_k^*(x,y) \text{ s.t. } \epsilon < \epsilon' \leq 1 \\ \frac{1}{2}\Delta_k^*(x,y) & \text{if } \Delta_k^*(x,y) < \|x-y\| \leq \frac{1}{2}\Delta_k(x,y), \end{cases} \tag{1}$$

where $c_1 < 1$ is a universal constant.

In the case that that X satisfies a *weak growth rate* condition (i.e., as in [2], there exists constants $\alpha > \beta \geq 1$ such that for every $x \in X$ and $r > 0$, $|B(x, \alpha r)| \leq |B(x,r)|^\beta$), then Theorem 1 provides $(1+\epsilon)$-distortion for all k-neighborhoods in X. In order to maintain separation among pairs of points at distances greater than $\frac{1}{2}\Delta_k(x,y)$, we need to make use of additional information about the metric space X. Our embedding dimension will depend on the doubling dimension $\dim(X)$ of X, and on a weak growth rate assumption about X, but not on the cardinality of X.

Theorem 2. *Let $k \in \mathbb{N}$, and X a discrete subset of $\mathbb{R}^U$. Suppose that X satisfies a weak growth rate assumption then for any $\epsilon, \theta > 0$ there exists an embedding $\bar{\Phi} : X \to \mathbb{R}^D$, where $D = O(\log k/\epsilon^2 + \dim(X)/\theta)$ such that Theorem 1 holds, and additionally if $\|x-y\| \geq \frac{1}{2}\Delta_k(x,y)$ then:*

$$\|\bar{\Phi}(x) - \bar{\Phi}(y)\| \geq \Delta_k(x,y) \cdot c_2\theta\sqrt{\epsilon}/\log^{1+\theta} k, \tag{2}$$

for some universal constant c_2.

2 Preliminaries

One of the tools we use are local probabilistic partitions. In particular, the following constructions are generalizations of the local probabilistic partitions of [2], and their analysis appears in [3]:

For any point $x \in X$ and a subset $S \subseteq X$ let $d(x, S) = \min_{s \in S} d(x, s)$. The *diameter* of X is denoted $\mathrm{diam}(X) = \max_{x,y \in X} d(x, y)$.

Definition 1 (Probabilistic Partition)**.** A *partition* P of X is a collection of disjoint set of *clusters* $\mathcal{C}(P) = \{C_1, C_2, \dots, C_t\}$ such that $X = \cup_j C_j$. A partition is called Δ-*bounded* where $\Delta : P \to \mathbb{R}^+$ if for all j, $\mathrm{diam}(C_j) \leq \Delta(C_j)$. For $x \in X$ we denote by $P(x)$ the cluster containing x. A *probabilistic partition* $\hat{\mathcal{P}}$ of a finite metric space (X, d) is a distribution over a set $\mathcal{P}$ of partitions of X. Such a partition is Δ-bounded if it is Δ-bounded for every $P \in \hat{\mathcal{P}}$.

Definition 2 (Locally Padded Probabilistic Partition)**.** Let $\hat{\mathcal{P}}$ be a Δ-bounded probabilistic partition of (X, d). Let $\mathcal{L}(x)$ denote the event that $B(x, \eta \cdot \Delta(P(x))) \subseteq P(x)$. For $\delta \in (0, 1]$, $\hat{\mathcal{P}}$ is called (η, δ)-*locally padded* if

$$\Pr[\mathcal{L}(x) | \bigwedge_{z \in Z} \mathcal{L}(z)] \geq \delta.$$

holds for any $x \in X$ and $Z \subseteq X \setminus B(x, 16\Delta(P(x)))$.

Lemma 3 (Locally Padded Cardinality-Based Probabilistic Partitions)**.** *Let (X, d) be a finite metric space. Let $k \in \mathbb{N}$. There exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of (X, d) with the following properties:*

- *For any $P \in \mathcal{P}$ and any $x \in X$: $|P(x)| \leq k$.*
- *For any $P \in \mathcal{P}$ is and any $x \in X$: $2^{-6} \leq \Delta(P(x))/\Delta_k(x) \leq 2^{-4}$.*
- *$\hat{\mathcal{P}}$ is $(\eta^{(\delta)}, \delta)$-locally padded for $\eta^{(\delta)} = 2^{-11} \ln k \cdot \ln(1/\delta)$, where $\delta \in (1/k, 1]$.*

Lemma 4. *Let (X, d) be a finite metric space. Let $k \in \mathbb{N}$ and $\xi > 0$. Let $\{\hat{\mathcal{P}}^{(t)}\}_{t \in T}$ be a collection of size $|T| \geq 8 \log k/\xi$ of independent Δ-bounded probabilistic partitions of (X, d) as in Lemma 3. Let $\delta = 1 - \xi$ and $\mathcal{L}_t^{(\delta)}(x)$ denote the event that $B(x, \eta^{(\delta)} \cdot \Delta(P^{(t)}(x))) \subseteq P^{(t)}(x)$, where $\eta^{(\delta)} = 2^{-11} \ln k \cdot \ln(1/\delta)$. Then with positive probability for every $x \in X$ there exists a set $T^{(\delta)}(x) \subseteq T$ of size $|T^{(\delta)}(x)| \geq (1 - 2\xi)|T|$ such that $\mathcal{L}_t^{(\delta)}(x)$ occurs for all $t \in T$.*

3 Local Dimensionality Reduction

In this section we describe the embedding and analysis to prove Theorem 1[1]. The main ingredients are a set of probabilistic partitions described in Section 2, and a compact embedding, based on a randomization of a device of Nash, provided in Section 3.1. Our construction proceeds in three steps. We prove the existence of an embedding Φ satisfying all of the properties in Theorem 1 for all $x, y \in X$ which are "close neighbors" in the sense that $\|x - y\| \leq \Delta_k^*(x, y)$. For farther neighbors, we use an simple additional construction in Section 3.3.

[1]We note that the constants may differ but a rescaling of the parameter ϵ would yield this formulation of the theorem.

3.1 The Randomized Nash Device

For any $\omega \in \mathbb{R}^U$ and $\sigma > 0$, we define the function $\varphi : \mathbb{R}^U \to \mathbb{R}^2$ as

$$\varphi(x; \sigma, \omega) = \frac{1}{\sigma} \left[\begin{array}{c} \cos(\sigma\omega' x) \\ \sin(\sigma\omega' x) \end{array} \right] \tag{3}$$

where $\omega' x$ denotes the inner product between ω and x. $\varphi(x; \sigma, \omega)$ maps onto a circle with radius σ^{-1} in $\mathbb{R}^2$. These functions were used by Nash in his construction of C^1-isometric embeddings of Riemannian manifolds [28], with the parameters chosen to correct errors in the metric. Note that as the parameter σ grows, the frequencies of the embedding function grow, but the amplitude becomes increasingly small. Using this procedure, Nash was able to correct the metric at every point in a manifold.

In this section we present a sequence of *random* parameter settings for these functions φ, first studied in [31], that with high probability approximate small distances in discrete metrics and bound large distances away from zero. Fix $\sigma > 0$ and let ω be a sample from a U-dimensional Gaussian $\mathcal{N}(0, I_U)$. For this choice of parameters, one may interpret Equation (3) as a random projection wrapped onto the circle. Using the intuition provided by the Johnson-Lindenstrauss lemma, one would expect nearby points x and y to be mapped to nearby points on the circle since the sine and cosine are Lipshitz. This intuition can be further reinforced by considering the expected distance between two points.

Claim 5. *For any x and y in $\mathbb{R}^U$, $|\varphi(x; \sigma, \omega) - \varphi(y; \sigma, \omega)|^2 = 2\sigma_t^{-2}(1 - \cos(\sigma\omega'(x-y)))$ and $\mathbb{E}[|\varphi(x; \sigma, \omega) - \varphi(y; \sigma, \omega)|^2] = 2\sigma^{-2}(1 - \exp(-\frac{1}{2}\sigma^2\|x-y\|^2)$.*

The first part of this claim can be verified using standard trigonometric identities, and the expectation follows because, for any $z \in \mathbb{R}^U$, $\mathbb{E}[\cos(\omega' z)] = \exp(-\frac{1}{2}\|z\|^2)$. For small distances relative to σ^{-1}, a Taylor expansion reveals that $2\sigma^{-2}(1 - \exp(-\frac{1}{2}\sigma^2\|x-y\|^2)$ is nearly the squared distance between x and y. For large distances, this function is equal to $2\sigma^{-2}$ and is bounded away from zero. The main result of this section is to note that these random variables are very well concentrated about their expected value and hence inherent their distance preserving property from this Gaussian kernel function. Hence, a concatenation of several φ corresponding to different samples of ω will provide a low-dimensional embedding.

Let $\sigma_1, \ldots, \sigma_D > 0$ be given real numbers bounded above by $\sigma_{\mathbf{m}}$, and let $\omega_1, \ldots, \omega_D$ be D samples from a U-dimensional Gaussian $\mathcal{N}(0, I_U)$. Let $\varphi^{(t)}(x) := \varphi(x; \sigma_t, \omega_t)$ and, for x and $y \in \mathbb{R}^U$, let $\Theta : X \to \mathbb{R}^{2D}$ denote the mapping $\Theta = \frac{1}{\sqrt{D}} \bigoplus_{1 \le t \le D} \varphi^{(t)}$. The main result of this section is the following Lemma.

Lemma 6. *Let $\frac{1}{2} > \epsilon > 0$ and x and $y \in \mathbb{R}^U$.*

a. $\|\Theta(x) - \Theta(y)\|^2 \le (1+\epsilon)\|x-y\|^2$ *with probability exceeding* $1 - \exp(-\frac{D}{2}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}))$.

b. If $\|x-y\| \le \frac{\sqrt{\epsilon}}{\sigma_{\mathbf{m}}}$, $\|\Theta(x) - \Theta(y)\|^2 \ge (1-\epsilon)\|x-y\|^2$ *with probability exceeding* $1 - \exp(-\frac{3D\epsilon^2}{32})$.

c. If $\|x-y\| \ge \frac{1}{\sqrt{2}\sigma_{\mathbf{m}}}$, $\|\Theta(x) - \Theta(y)\|^2 \ge \frac{1}{4\sigma_{\mathbf{m}}^2}$ *with probability exceeding* $1 - \exp(-\frac{D}{32})$.

The randomized embedding Θ maps onto a product of circles of varying radii, a compact set diffeomorphic to the D-torus. The different values of σ will be necessary in the following sections to stitch together regions of the metric space with differing densities, but the important point is all of the concentration results are only a function of the largest value of the σ_t. Intuitively, one can interpret this as saying the high frequency information is the dominant source of error in the approximation.

To prove part (a) of Lemma 6, note that $1 - \cos(\alpha) \leq \alpha^2/2$ for all α. Let $\ell = \|x - y\|$. $\tau_i := \omega_i'(x - y)$ is distributed as a one-dimensional Gaussian distribution $\mathcal{N}(0, \ell^2)$ and $\tau_1, \ldots, \tau_D$ are independent and we have

$$\|\Theta(x) - \Theta(y)\|^2 = \frac{1}{D}\sum_{t=1}^{D} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2 = \frac{1}{D}\sum_{t=1}^{D} \frac{2}{\sigma_t^2}\left(1 - \cos\left(\sigma_t \tau_t\right)\right) \leq \frac{1}{D}\sum_{t=1}^{D} \tau_t^2 \,. \tag{4}$$

It therefore follows that

$$\mathbf{Pr}\left[\|\Theta(x) - \Theta(y)\|^2 \geq (1+\epsilon)\ell^2\right] \leq \mathbf{Pr}\left[\frac{1}{D}\sum_{t=1}^{D} \tau_t^2 \geq (1+\epsilon)\ell^2\right] \leq e^{-\frac{D}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)} \tag{5}$$

where the second inequality is a well known known concentration inequality a χ-squared random variable (see, e.g., [17]).

Parts (b) and (c) require a more detailed verification, but follow from a standard Chernoff Bound analysis. We will sketch the argument here and not include the full proof. We explicitly bound the moment generating function of the everywhere non-positive process $\cos(\sigma\omega'(x - y)) - 1$ by using the upper bound $\exp(\alpha) \leq 1 + \alpha + \alpha^2$ for all $\alpha \leq 0$. Using this upper bound allows us to bound $\mathbb{E}_\omega[s(\cos(\sigma\omega'(x-y)) - 1)]$ by employing Claim 5 above. From this bound, one can then use some moderately tedious algebra and properties of the exponential function to yield the desired concentration inequalities.

The analysis in this section proves the following "compactified" version of the Johnson-Lindenstrauss Lemma where the entire metric space is mapped to a compact region of space with radius at most w. Indeed, this should not be surprising given our discussion above of how the randomized Nash device wraps random projections around the circle. In the limit that w goes to infinity, the following statement is exactly equivalent to the Johnson-Lindenstrauss Lemma (though weaker in the constants).

Corollary 7. *For any $0 < \epsilon < \frac{1}{2}$, $w > 0$, and any integer k, let D be an even integer greater than $22(\epsilon^2/2 - \epsilon^3/3)^{-1}\log(k)$. Then for any set V of k points in $\mathbb{R}^U$, there is a map Φ into R^D such that $\|\Phi(x)\| \leq w$ for all $x \in V$, and, for all x,$y \in V$, $\|\Phi(x) - \Phi(y)\|^2 \leq (1+\epsilon)\|x - y\|^2$ and*

$$\|\Phi(x) - \Phi(y)\|^2 \geq \begin{cases} (1-\epsilon)\|x-y\|^2 & \|x - y\| \leq \sqrt{\epsilon}w \\ (1 - \frac{1}{w^2}\|x-y\|^2)\|x-y\|^2 & \sqrt{\epsilon}w \leq \|x-y\| \leq \frac{1}{\sqrt{2}}w \\ \frac{1}{4}w^2 & \|x-y\| \geq \frac{1}{\sqrt{2}}w \end{cases} \tag{6}$$

Furthermore, this map can be found in randomized polynomial time.

This corollary is proven by a union bound argument, noting that for $D > 22(\epsilon^2/2 - \epsilon^3/3)^{-1}\log(k)$, the probability of all of the constraints being violated is at most $1/k$.

3.2 Embedding Close Neighbors

We now turn to a recipe for combining multiple instances of these trigonometric embeddings into a global map that preserves local distances using the probabilistic partitions discussed in 2. Specifically we concern ourselves with the "close neighbors," pairs x and y satisfying $\|x - y\| \leq \Delta_k^*(x, y)$. Let $D = C'\lceil\log k/\epsilon^2\rceil$, where C' is some universal constant to be determined later. We construct a locally padded cardinality-based probabilistic partition $\bar{\mathcal{P}}^{(t)}$ as in Lemma 3. Now fix a partition $P^{(t)} \in \mathcal{P}^{(t)}$. We define a trigonometric embedding for *every cluster* $C \in P^{(t)}$.

Let $\sigma_C = 2^{12} \ln k/\epsilon \cdot \Delta(C)^{-1}$, and let $\{\omega_C | C \in P^{(t)}, 1 \leq t \leq D\}$ be i.i.d. samples from a U-dimensional Gaussian $\mathcal{N}(0, I_U)$. For $x \in C$ define $\sigma^{(t)}(x) = \sigma_C$, $\omega^{(t)}(x) = \omega_C$, and $A^{(t)}(x) = \min\left\{d(x, X \setminus C), \sigma^{(t)}(x)^{-1}\right\}$, and let

$$\Phi^{(t)}(x) = A^{(t)}(x) \begin{bmatrix} \cos(\sigma^{(t)}(x)\omega^{(t)}(x)'x) \\ \sin(\sigma^{(t)}(x)\omega^{(t)}(x)'x) \end{bmatrix}.$$

The function $A^{(t)}$ serves as the amplitude of the embedding. For padded x, this number is equal to the amplitude defined in Section 3.1, and the amplitude rolls off to zero near the boundary of each cluster. In each cluster, we have a different trigonometric embedding, and continuity is maintained because the amplitude is zero at the boundaries of the clusters.

We define our embedding $\Phi : X \to l_2^{2D}$ by concatenating D instances of $\Phi^{(t)}$: $\Phi = \frac{1}{\sqrt{D}} \bigoplus_{1 \leq t \leq D} \Phi^{(t)}$.

3.2.1 Embedding Analysis

We first analyze the relationship between the embeddings $\Phi^{(t)}$ and the Nash-type embeddings of Section 3.1 and show that they have similar distortion. We then employ the Lovász Local Lemma to show that the distance constraints are satisfied everywhere with positive probability.

For each x, let $\varphi^{(t)}(x)$ denote $\varphi(x; \sigma^{(t)}(x), \omega^{(t)}(x))$, the trigonometric embedding function with parameters given by the σ and ω of the cluster $P^{(t)}(x)$. We have the following lemma:

Lemma 8. *Let $x, y \in X$. Then,*

1. *If $P^{(t)}(x) \neq P^{(t)}(y)$, $\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\| \leq 2\|x - y\|$.*
2. *If $C := P^{(t)}(x) = P^{(t)}(y)$, $\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\|^2 \leq \|x - y\|^2 + \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2$.*
3. *If $C := P^{(t)}(x) = P^{(t)}(y)$, $\sigma_C^{-1} \leq d(x, X \setminus P^{(t)}(x))$ and $\sigma_C^{-1} \leq d(y, X \setminus P^{(t)}(y))$, then $\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\| = \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|$.*

Proof. First, we observe that for all x and y

$$\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\|^2 = (A^{(t)}(x) - A^{(t)}(y))^2 + 2A^{(t)}(x)A^{(t)}(y)\left(1 - \cos\left(\sigma^{(t)}(x)\omega^{(t)}(x)x - \sigma^{(t)}(y)\omega^{(t)}(y)y\right)\right)$$

Here we used the trigonometric identities that $\sin^2(\theta) + \cos^2(\theta) = 1$ and $\cos(\theta_1 - \theta_2) = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2)$. We now proceed case by case.

For (1), note that since the cosine bounded below by -1, we have

$$\begin{aligned} \|\Phi^{(t)}(x) - \Phi^{(t)}(y)\|^2 &\leq (A^{(t)}(x) - A^{(t)}(y))^2 + 4A^{(t)}(x)A^{(t)}(y) \\ &= (A^{(t)}(x) + A^{(t)}(y))^2 \\ &\leq (d(x, X \setminus P^{(t)}(x)) + d(y, X \setminus P^{(t)}(y)))^2 \end{aligned}$$

but if x and y fall in different clusters, $\|x - y\| \geq d(y, X \setminus P^{(t)}(y))$ and $\|x - y\| \geq d(x, X \setminus P^{(t)}(x))$, and the assertion follows.

If $C = P^{(t)}(x) = P^{(t)}(y)$, recall that $\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\|^2 = \sigma_C^{-2}\left(1 - \cos\left(\sigma_C \omega_C'(x - y)\right)\right)$. We have that $A^{(t)}(x)A^{(t)}(y) \leq \sigma_C^{-2}$. We also need to show that $|A^{(t)}(x) - A^{(t)}(y)| \leq \|x - y\|$ for all $x, y \in P^{(t)}(x)$. We show that $A^{(t)}(x) - A^{(t)}(y) \leq \|x - y\|$ and the claim holds by reversing the roles of x and y. There are two cases: if $A^{(t)}(y) = \sigma^{-1}$ then $A^{(t)}(x) \leq \sigma^{-1}$ and $A^{(t)}(x) - A^{(t)}(y) \leq 0$. Otherwise $A^{(t)}(y) = d(y, X \setminus P^{(t)}(y))$ and $A^{(t)}(x) \leq d(x, X \setminus P^{(t)}(x))$ implying $A^{(t)}(x) - A^{(t)}(y) \leq d(x, X \setminus P^{(t)}(x)) - d(y, X \setminus P^{(t)}(y)) \leq \|x - y\|$ since $P^{(t)}(x) = P^{(t)}(y)$.

Finally, for (3), we only need use the fact that $A^{(t)}(x) = A^{(t)}(y) = \sigma_C^{-1}$. □

For $x, y \in X$, let us now classify the different coordinates t according to the cases of Lemma 8. Define the sets

$$\begin{aligned} T_{\neq}(x,y) &= \{t | P^{(t)}(x) \neq P^{(t)}(y)\} \\ T_{=}(x,y) &= \{t | P^{(t)}(x) = P^{(t)}(y)\} \\ T_{\circ}(x,y) &= \{t | d(x, X \setminus P^{(t)}(x)) \geq \sigma^{(t)}(x)^{-1} \vee d(y, X \setminus P^{(t)}(y)) \geq \sigma^{(t)}(y)^{-1}\} \end{aligned} \tag{7}$$

so that we have the upper and lower bounds for our embedded distances

$$\|\Phi(x) - \Phi(y)\|^2 \geq \tfrac{1}{D} \sum_{t \in T_{=}(x,y) \cap T_{\circ}(x,y)} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2, \text{ and} \tag{8}$$

$$\|\Phi(x) - \Phi(y)\|^2 \leq \frac{1}{D} \left[\sum_{t \in T_{=}(x,y)} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2 + \sum_{t \in T_{\neq}(x,y) \cup (T_{=}(x,y) \setminus T_{\circ}(x,y))} 2\|x - y\|^2 \right]. \tag{9}$$

We now turn to the part of the analysis that relies on the Lovász Local Lemma, to prove the existence of a probabilistic partition where for every pair of close neighbors (x, y), $|T_{=}(x,y) \cap T_{\circ}(x,y)| \geq (1 - 2\epsilon)D$. We then deduce from the analysis in Section 3.1, via a second application of the the Local Lemma, that the properties of the embedding hold with positive probability.

Apply Lemma 4 with $\xi = \epsilon$. Consider pairs x, y that are close neighbors, that is: $\|x - y\| \leq \Delta_k^*(x, y)$ where $\Delta_k^*(x, y) = c_1\sqrt{\epsilon}/\ln k \cdot \Delta_k(x, y)$, and $c_1 = 2^{-19}$. Note that c_1 is chosen so that $\frac{1}{8}\sigma^{(t)}(x, y)^{-1} \leq \Delta_k^*(x, y) \leq \frac{1}{2}\sigma^{(t)}(x, y)^{-1}$, where $\sigma^{(t)}(x, y) = \min\{\sigma^{(t)}(x), \sigma^{(t)}(y)\}$, for all $1 \leq t \leq D$ (this follows from Lemma 3). Assume w.l.o.g that $\sigma^{(t)}(x, y) = \sigma^{(t)}(x)$ (otherwise switch the roles of x and y). Consider $t \in T^{(\delta)}(x)$ then $B(x, \eta^{(\delta)} \cdot \Delta(P^{(t)}(x)) \subseteq P^{(t)}(x)$. It follows that $d(x, X \setminus P(x)) \geq \eta^{(\delta)} \cdot \Delta(P^{(t)}(x)) \geq 2\sigma^{(t)}(x)^{-1}$, by definition. Now consider $y \in X$ such that $\|x - y\| \leq \sigma^{(t)}(x)^{-1}$ then $P^{(t)}(y) = P^{(t)}(x)$ and $d(y, X \setminus P^{(t)}(y)) \geq d(y, X \setminus P^{(t)}(x)) - \|x - y\| \geq \sigma^{(t)}(x)^{-1}$. Hence, $t \in T_{=}(x, y) \cap T_{\circ}(x, y)$. Thus, we conclude from the lemma that $|T_{=}(x, y) \cap T_{\circ}(x, y)| \geq (1 - 2\epsilon)D$. Plugging this bound into 8 and 9 yields:

$$\|\Phi(x) - \Phi(y)\|^2 \quad \geq \quad (1 - 2\epsilon) \cdot \frac{\sum_{t \in T_{=}(x,y) \cap T_{\circ}(x,y)} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2}{|T_{=}(x, y) \cap T_{\circ}(x, y)|}, \text{ and} \tag{10}$$

$$\|\Phi(x) - \Phi(y)\|^2 \quad \leq \quad \frac{\sum_{t \in T_{=}(x,y)} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2}{|T_{=}(x, y)|} + 4\epsilon\|x - y\|^2. \tag{11}$$

We will next apply the Local Lemma again over events related to the Nash-type embeddings for the different clusters. Define:

$$L(x, y) = \frac{\sum_{t \in T_{=}(x,y) \cap T_{\circ}(x,y)} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2}{|T_{=}(x, y) \cap T_{\circ}(x, y)|} \quad \text{and} \quad U(x, y) = \frac{\sum_{t \in T_{=}(x,y)} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2}{|T_{=}(x, y)|}$$

Consider the pairs x, y that are close neighbors, that is: $\|x - y\| \leq \Delta_k^*(x, y)$. Let $\epsilon'(x, y) = \max\{\epsilon, \Delta_k^*(x, y)^{-2}\|x - y\|^2\}$. We define the following events for each such pair. Let $A_U(x, y)$ be the event that $U(x, y) > (1 + \epsilon)\|x - y\|^2$ and let $A_L(x, y)$ be the event that $L(x, y) < (1 - \epsilon'(x, y))$. Let $A(x, y) = A_L(x, y) \vee A_U(x, y)$.

We create a dependency graph G_A whose vertices are the events $A(x,y)$. Note that the event $A(x,y)$ depends only on the random variables associated with clusters $C \in P^{(t)}$ where $P^{(t)}(x) = P^{(t)}(y)$. We place an edge between two events $A(x,y)$ and $A(x',y')$ if $P^{(t)}(x) = P^{(t)}(x')$ for some $t \in T_=(x,y) \cap T_=(x',y')$. Note that if there is no edge between the two events then they are independent. On the other hand assume if there is an edge then for some t, $P^{(t)}(x) = P^{(t)}(y) = P^{(t)}(x') = P^{(t)}(y')$. Then $\max\{\|x-x'\|, \|x-y'\|\} \le \Delta(P^{(t)}(x)) \le \Delta_k(x)/16$, by Lemma 3, and $x', y' \in B(x, \Delta_k(x))$. This implies that the number of such pairs is at most $\binom{k}{2}$ which bounds the maximum degree of the dependency graph G_A.

The fact the x,y are close neighbors implies that $\|x-y\| \le \Delta_k^*(x,y) \le \frac{1}{\sqrt{2}}\sigma^{(t)}(x,y)^{-1}$. Now, by part (a) of Lemma 6 the probability that $U(x,y) > (1+\epsilon)\|x-y\|^2$ is at most $e^{-D(\epsilon^2/4+\epsilon^3/6)} \le k^{-2}/4$. Similarly by Lemma 6(b) the probability that $L(x,y) < (1 - \max\{\epsilon, \sigma_{\mathbf{m}}^2\|x-y\|^2\})\|x-y\|^2$ is at most $e^{-3D\epsilon^2/32} \le k^{-2}/4$, where $\sigma_{\mathbf{m}} \le \max_{t \in T} \sigma^{(t)}(x,y) \le \Delta_k^*(x,y)^{-1}/2$. Hence the probability the event $A(x,y)$ occurs is at most $k^{-2}/2 < 1/(e(\binom{k}{2}+1))$. Since the dependency graph has degree at most $\binom{k}{2}$, the probability of an event $A(x,y)$ occurring satisfies the conditions of the Local Lemma, implying that there is positive probability that none of the events $A(x,y)$ occur. Therefore for any close neighbors x,y such that $\|x-y\| \le \Delta_k^*(x,y)$ we have:

$$\begin{aligned}
\|\Phi(x) - \Phi(y)\|^2 &\ge (1-2\epsilon)L(x,y) \ge (1 - 2\epsilon - \max\{\epsilon, \Delta_k^*(x,y)^{-2}\|x-y\|^2\})\|x-y\|^2 \\
&\ge (1 - 3\max\{\epsilon, \epsilon'(x,y)\})\|x-y\|^2, \ \text{ and} \qquad (12) \\
\|\Phi(x) - \Phi(y)\|^2 &\le U(x,y) + 4\epsilon\|x-y\|^2 \le (1+5\epsilon)\|x-y\|. \qquad (13)
\end{aligned}$$

completing the analysis of the embedding of close neighbors.

3.3 Embedding farther Neighbors

In this section, we extend the embedding to cover all pairs such that $\|x-y\| \le \frac{1}{2}\Delta_k(x,y)$. To this end, we add another component to the embedding $\Psi : X \to \mathbb{R}^D$. The embedding Ψ is based on ideas similar to those of [32, 1]. For each $1 \le t \le D$, define a function $\Psi^{(t)} : X \to \mathbb{R}^2$ and let $\{\nu^{(t)}(C) | C \in P^{(t)}, t \in T\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. The embedding is defined for each $x \in X$ as $\Psi(x) = \frac{1}{\sqrt{D}} \bigoplus_{1 \le t \le D} \Psi^{(t)}(x)$ with

$$\Psi^{(t)} = \sqrt{\epsilon} \cdot \nu^{(t)}(P(x)) \cdot d(x, X \setminus P^{(t)}(x)) \,.$$

Our final embedding will be $\hat{\Phi} = \Phi \oplus \Psi$.

3.3.1 Embedding Analysis

First observe that the upper bound on the distance in the embedding is maintained with only small loss. This follows since $\|\Psi(x) - \Psi(y)\| \le \sqrt{\epsilon}\|x-y\|$, as follows by a standard argument (see, e.g., [1]), and we have

$$\|\hat{\Phi}(x) - \hat{\Phi}(y)\|^2 = \|\Phi(x) - \Phi(y)\|^2 + \|\Psi(x) - \Psi(y)\|^2 \le (1+5\epsilon)\|x-y\|^2 + \epsilon\|x-y\|^2 = (1+6\epsilon)\|x-y\|^2.$$

We now turn to show that the embedding provides a lower bound on the distance between images of neighbors which are not "close". We can partition the pairs x,y such that $\Delta_k^*(x,y) \le \|x-y\| \le \frac{1}{2}\Delta_k(x,y)$ into two sets as follows: $W_= = \{\{x,y\} \mid |T_=(x,y)| \ge D/2\}$ and $W_{\neq} = \{\{x,y\} \mid |T_{\neq}(x,y)| > D/2\}$.

Consider first a pair in $W_=$. For such pairs we show that the Φ component of the embedding gives a good lower bound on the distance. Recall that

$$\|\Phi(x)-\Phi(y)\|^2 \quad \geq \quad \frac{\sum_{t\in T_=(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2}{D} \geq \frac{1}{2}\cdot\frac{\sum_{t\in T_=(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2}{|T_=(x,y)|}. \quad (14)$$

Let $L_B(x,y) = \frac{\sum_{t\in T_=(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(x)\|^2}{|T_=(x,y)|}$ and define the event $B(x,y)$ that $L_B(x,y) < 2^{-5}\Delta_k^*(x,y)^2$. As before we create a dependency graph G_B whose vertices are these events and place an edge between two events $B(x,y)$ and $B(x',y')$ if $P^{(t)}(x) = P^{(t)}(x')$ for some $t \in T_=(x,y)\cap T_=(x',y')$. Note that if there is no edge between the two events then they are independent. By the same argument made before we can bound the degree of G_B by $\binom{k}{2}$.

We have that $\|x-y\| \geq \Delta_k^*(x,y) \geq \frac{1}{8}(\max_{t\in T}\sigma^{(t)}(x,y))^{-1} \geq \frac{1}{8}\sigma_{\mathbf{m}}^{-1}$. Now, by Lemma 6, the probability that $L_B(x,y) < 2^{-7}\sigma_{\mathbf{m}}^{-2}$ is at most $e^{-D\epsilon^2/32} < k^{-2}/2$, where $\sigma_{\mathbf{m}} \leq \max_{t\in T}\sigma^{(t)}(x,y) \leq \Delta_k^*(x,y)^{-1}/2$. Hence, the probability that event $B(x,y)$ occurs is at most $k^{-2}/2 < 1/(e(\binom{k}{2}+1))$, which satisfies the conditions of the Local Lemma, implying that there is positive probability that none of these event occur. We conclude that for every pair x,y in $W_=$,

$$\|\hat{\Phi}(x)-\hat{\Phi}(y)\|^2 \geq \|\Phi(x)-\Phi(y)\|^2 \geq \frac{1}{2}L_B(x,y) \geq 2^{-6}\Delta_k^*(x,y)^2, \quad (15)$$

that is: $\|\hat{\Phi}(x)-\hat{\Phi}(y)\| \geq \frac{1}{8}\Delta_k^*(x,y)$.

Next we deal with pairs in $W_{\neq}$. Here we will make use of the Ψ component of the embedding. By applying Lemma 4 with $\xi = 1/4$ we infer that with positive probability for every $x \in X$ there exists a set $T'(x) = T^{(3/4)}(x)$ such that $|T'(x)| \geq D/2$ and for each $t \in T'(x)$, $B(x,\eta^{(3/4)}\Delta(P^{(t)}(x))) \subseteq P^{(t)}(x)$, and therefore $\epsilon\cdot d(x, X\setminus P^{(t)}(x)) \geq \sigma^{(t)}(x)^{-1}/2$. We note that this event is positively correlated with the former application of the lemma and so this assertion holds in conjunction with our analysis of Φ. Assume w.l.o.g that $\sigma^{(t)}(x,y) = \sigma^{(t)}(x)$ (otherwise switch the roles of x and y), then we have that: $\epsilon\cdot d(x, X\setminus P^{(t)}(x)) \geq \Delta_k^*(x,y)$.

For such a pair x,y define $B'(x,y)$ to be the event that $\|\Psi(x)-\Psi(y)\| < \frac{1}{16}\Delta_k^*(x,y)$. Define a dependency graph $G_{B'}$ whose vertices are these events. We place an edge between two events $B'(x,y)$ and $B'(x',y')$ if one of $\{x,y\}$ is in the same cluster as $\{x',y'\}$ for some $t \in T$. Note that if there is no edge between two events then they are independent. On the other hand assume there exists $t \in T$ such that $P^{(t)}(x) = P^{(t)}(x')$. Therefore there are at most k such points x'. Now consider all such pairs including x'. Denote the other points in these pairs $y'_1,\ldots,y'_s$. Let z be the point which maximizes $\Delta_k(z)$ over all y'_js and x'. Since $\|x'-y'_j\| \leq \frac{1}{2}\Delta_k(x',y'_j) = \frac{1}{2}\max\{\Delta_k(x'),\Delta_k(y'_j)\} \leq \frac{1}{2}\Delta_k(z)$. We conclude that $\|z-y'_j\| \leq \|x'-z\|+\|z-y'_j\| \leq \Delta_k(z)$ and therefore all y'_js are in the ball around z containing at most k points so that $s \leq k$. We conclude that there are at most k^2 such pairs, which provides an upper bound on the degree of $G_{B'}$.

Note that for each $t \in T'(x)$ with probability at least $1/4$, $\nu(P^{(t)}(x)) = 1$ and $\nu(P^{(t)}(y)) = 0$, as $P^{(t)}(x) \neq P^{(t)}(y)$. Hence using Chernoff bounds we have that the probability that there are less than $1/8$ fraction of the coordinates $t \in T'(x)$ such $|\Psi^{(t)}(x)-\Psi^{(t)}(y)| \geq \Delta_k^*(x,y)$ is at most $e^{-D/16}$. Hence the probability that event $B'(x,y)$ occurs is at most $e^{-D/16} \leq k^{-2}/4 < 1/(e(k^2+1))$, satisfying the condition for the Local Lemma. We can therefore conclude that there is positive probability that non of the events $B'(x,y)$ occur.

Therefore for every $x,y \in W_{\neq}$, $\|\hat{\Phi}(x)-\hat{\Phi}(y)\| \geq \|\Psi(x)-\Psi(y)\| \geq \frac{1}{16}\Delta_k^*(x,y)$, thus completing the proof of Theorem 1.

4 Maintaining Separation of Distant Pairs

In many applications it is desirable that not only our distortion for neighbors is small but also that the distant pairs (non-neighbors) will not become too close in the embedding so that the local structure is preserved. To obtain this type of property we can use any non-expansive embedding $\Upsilon : X \to \ell_2^D$ that provides guarantees for the distortion of the distant pairs via a similar trick to the one in Section 3.3, i.e., add a component $\sqrt{\epsilon}\Upsilon$ to the embedding $\hat{\Phi}$. Let $\bar{\Phi} = \hat{\Phi} \oplus (\sqrt{\epsilon}\Upsilon)$ then:

$$\|\bar{\Phi}(x)-\bar{\Phi}(y)\|^2 = \|\hat{\Phi}(x)-\hat{\Phi}(y)\|^2+\epsilon\|\Upsilon(x)-\Upsilon(y)\|^2 \leq (1+\epsilon)\|x-y\|^2+\epsilon\|x-y\|^2 = (1+2\epsilon)\|x-y\|^2,$$

whereas the lower bound for neighbors given by $\hat{\Phi}$ still holds and the lower bound for far neighbors is given by Υ with just an additional $\sqrt{\epsilon}$ factor loss.

If we assume nothing about the metric space X there is no such low dimensional embedding that will give good guarantees. However, in this section we show that under reasonable assumptions on the local growth structure of the space there exists an embedding that provides reasonable bounds and in particular guarantees that the local structure of the space would be preserved. This is made possible by combining newly developed techniques of Abraham, Bartal and Neiman for local embeddings [2] and low-dimensional embedding doubling metrics [4].

Given a metric space (X, d) we say that X has *doubling constant* λ if for any $x \in X$, $r \in \mathbb{R}^+$ the ball $B(x, r)$ can be covered by at most λ balls of radius $r/2$. The *doubling dimension* of X is $\dim(X) = \log \lambda$.

In recent work of Abraham, Bartal and Neiman [4] it is shown that every metric space embeds in ℓ_2^D where $D = O(\dim(X)/\theta)$ with distortion $O(\log^{1+\theta} n)$ (in fact, their embedding can be shown to have constant average distortion [1, 4]). Hence one choice for the component Υ could be this embedding, and combining it with $\hat{\Phi}$ as described above, we obtain a global embedding in dimension $O(\epsilon^{-2}\log k+\theta^{-1}\dim(X))$ that guarantees that the distance distant pairs does not shrink below $\Delta_k(x, y)/\log^{1+\theta} n$. However, as this bound depends on the global size of the set this still does not promise full preservation of the local structure. To overcome this we give an improvement of this embedding using ideas from [2].

Following [2, 4] we obtain the following:

Theorem 3. *Given a metric space (X, d), $1 \leq p \leq \infty$, and $\theta > 0$, there exists an embedding of X into ℓ_p^D in dimension $D = O(\dim(X)/\theta)$ and scaling distortion where the distortion for pairs $x, y \in X$ and $\tilde{k}$ s.t. $d(x, y) \leq \Delta_{\tilde{k}}(x)$ is $O(\log^{1+\theta} \tilde{k}/\theta)$.*

The proof of Theorem 3 will appear in the full version of the paper. We use the embedding in Theorem 3 for the component Υ as explained above. This provides stronger distortion bounds for distant pairs. Moreover, under very weak condition on the subset of points X we can assure that the local structure of the space is maintained. Namely we say X satisfies a *weak growth rate* (a similar assumption appears in [2])condition $\mathrm{WGR}(\alpha, \beta)$ for some constants $\alpha > \beta \geq 1$ if for every $x \in X$ and $r > 0$, $|B(x, \alpha r)| \leq |B(x, r)|^\beta$. In fact, this implies that the lower bound on the distortion guarantees by Theorem 3 is a monotonic function of the distance from any particular point. The following is a corollary of Theorem 3:

Corollary 9. *Given a metric space (X, d) satisfying $\mathrm{WGR}(\alpha, \beta)$. For any $1 \leq p \leq \infty$, and $0 < \theta \leq \log_\beta(\alpha/\beta)$, there exists an embedding f of X into ℓ_p^D in dimension $D = O(\dim(X)/\theta)$ such that for any $x, y \in X$ and $\tilde{k}$ s.t. $d(x, y) \geq \Delta_{\tilde{k}}(x)$ then $\|f(x) - f(y)\|^p \geq \Delta_{\tilde{k}}(x) \cdot \Omega(\alpha^{-1}\theta/\log^{1+\theta} \tilde{k}(x)))$.*

References

[1] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. In *Proceedings of the thirty-eighth annual ACM Symposium on Theory of Computing*, pages 271–286, New York, NY, USA, 2006. ACM Press.

[2] Ittai Abraham, Yair Bartal, and Ofer Neiman. Local embedding of metric spaces. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing*, pages 631–640, 2007.

[3] Ittai Abraham, Yair Bartal, and Ofer Neiman. Local embeddings, dimension reduction and distance oracles, 2007. Tech Report.

[4] Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metric spaces in their intrinsic dimension. In *Proceedings of the 19th ACM-SIAM Symp. on Discrete Algorithms*, 2008.

[5] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and Systems Science*, 66(4):671–687, 2003. Special issue of invited papers from PODS'01.

[6] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast johnson-lindenstrauss transform. In *STOC '06: Proceedings of the thiry-eigth annual ACM Symposium on Theory of Computing*, pages 557–563, 2006.

[7] Noga Alon. A parallel algorithmic version of the local lemma. *Random Structures and Algorithms*, 2(4):367–378, 1991.

[8] P. Assouad. Plongements lipschitziens dans $\mathbb{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[9] M. Balasubramanian and E. L. Schwartz. The Isomap algorithm and topological stability. *Science*, 295(5552), 2002.

[10] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*. To appear. Preprint available at http://www.eecs.umich.edu/~wakin/publications.html.

[11] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 184–193. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[12] J. Beck. An algorithmic approach to the Lovász Local Lemma, I. *Random Structures and Algorithms*, 4(1991):343–365, 2.

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

[14] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[15] M. Brand. Charting a manifold. In *Neural Information Processing Systems (NIPS)*, 2002.

[16] Payel Das, Mark Moll, Hernán Stamati, Lydia E. Kavraki, and Cecilia Clementi. Low-dimensional, free-energy landscapes of protein folding reactions by nonlinear dimensionality reduction. *Proceedings of the National Academy of Science*, 103(26):9885–9890, 2006.

[17] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.

[18] D.L. Donoho and C. Grimes. Hessian eigenmaps: new locally linear embedding techniques for high dimensional data. Technical report, TR2003-08, Dept. of Statistics, Stanford University, 2003.

[19] Lars Engebretsen, Piotr Indyk, and Ryan O'Donnell. Derandomized dimensionality reduction with applications. In *SODA '02: Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 705–712, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics.

[20] P. Erdös and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In A. Hajnal et. al, editor, *Infinite and Finite Sets*, volume 11. Colloquia Mathematica Societas János Bolyai, North Holland, Amsterdam, 1975.

[21] Allen Gersho and Robert M. Gray. *Vector Quantization and Signal Compression*, volume 159 of *The Springer International Series in Engineering and Computer Science*. Springer, 1992.

[22] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS '03: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, page 534, Washington, DC, USA, 2003. IEEE Computer Society.

[23] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using Laplacianfaces. *IEEE. Trans. Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

[24] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[25] O. C. Jenkins and M. J. Mataric. Deriving action and behavior primitives from human motion data. In *IEEE/RSK International Conference on Intelligent Robots and Systems*, 2002.

[26] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven,Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[27] M. Molloy and B. Reed. Further algorithmic aspects of the Local Lemma. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 524–529, 1998.

[28] John Nash. C^1 isometric embeddings. *The Annals of Mathematics*, 60(3):383–396, 1954.

[29] J. Nilsson, T. Fioretos, M. Hoglund, and M. Fontes. Approximate geodesic distances reveal biologically relevant structures in microarray data. *Bioinformatics*, 20(6):874–880, 2004.

[30] Ali Rahimi and Ben Recht. Estimating observation functions in dynamical systems using unsupervised regression. In *Neural Infomration Processing Systems*, 2006.

[31] Ali Rahimi and Ben Recht. Random features for large-scale kernel machines. In *Neural Infomration Processing Systems*, 2007.

[32] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.

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

[34] Gideon Schechtman and Adi Shraibman. Lower bounds for local versions of dimension reductions, 2007. Tech Report.

[35] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[36] D. Sivakumar. Algorithmic derandomization via complexity theory. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 619–626, New York, NY, USA, 2002. ACM.

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

[38] M. Thorup and U. Zwick. Approximate distance oracles. In 33^{rd} *Annual ACM Symposium on Theory of Computing*, pages 183–192, Hersonissos, Crete, Greece, July 2001.

[39] K. Q. Weinberger and L. K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.

[40] H. Whitney. Differentiable manifolds. *Annals of Mathematics*, 31:645–680, 1936.

[41] H. Whitney. Self-intersection of a smooth n-manifold in $2n$-space. *Annals of Mathematics*, 45:220–246, 1944.