Dimensionality Reduction: beyond the Johnson-Lindenstrauss bound*

Yair Bartal† Ben Recht‡ Leonard J. Schulman§

Abstract

Dimension reduction of metric data has become a useful technique with numerous applications. The celebrated Johnson-Lindenstrauss lemma states that any n-point subset of Euclidean space can be embedded in $O(\epsilon^{-2} \log n)$ dimension with $1+\epsilon$ distortion. This bound is known to be nearly tight.

In many applications the demand that all distances would be nearly observed is too strong. In this paper we show that indeed under natural relaxations of the goal of the embedding, an improved dimension reduction is possible where the target dimension is independent of n. Our main result can be viewed as a *local dimension reduction*. There are a variety of empirical situations 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.

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. Our work is an analog of Nash's for discrete subsets of Euclidean space. For perfect preservation of infinitesimal neighborhoods we substitute near-isometric embedding of neighborhoods of bounded cardinality.

We provide a local $(1+\epsilon)$-distortion embedding (preserving short distances) for any finite subset of Euclidean space in dimension $O(\epsilon^{-2} \log k)$, where k is the cardinality of the neighborhoods within which short distances are preserved. We also show that with some additional assumptions, a global embedding that also keeps distant points well-separated may be obtained.

As an application of our result we obtain an (Assouad-style) dimension reduction for finite subsets of Euclidean space where the metric is raised to some fractional power (the resulting metrics are known as snowflakes). We show that any such metric can be embedded in dimension $\tilde{O}(\epsilon^{-3} \dim(X))$ with $1+\epsilon$ distortion, where $\dim(X)$ is the doubling dimension, a measure of the intrinsic dimension of the set. This result improves recent work by Gottlieb and Krauthgamer [20] to a nearly tight bound.

The new dimension reduction results are useful for applications such as clustering and distance labeling.

---

*A previous version of this paper was posted under the title: "A Nash-type Dimensionality Reduction for Discrete Subsets of L_2" [11]. The work of the first and second authors was performed in part while at the Center for the Mathematics of Information, Caltech.

†yair@cs.huji.ac.il. School of Engineering and Computer Science, Hebrew University, Israel 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@cs.wisc.edu. Computer Science Department, U Wisconsin, Madison.

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

1 Introduction

Dimension reduction for high dimensional metric data has been an extremely important paradigm in many application areas. In particular, the celebrated Johnson-Lindenstrauss Lemma [25] has played a central role in a plethora of applications. The lemma states that every n-point subset of Euclidean space can be embedded in dimension $O(\epsilon^{-2} \log n)$ with $1 + \epsilon$ distortion. This bound is known to be nearly tight [5]. However, in many practical instances it is often the case that the high-dimensional data is inherently low dimensional and it is therefore desirable to reduce its dimension close to its inherent dimensionality, which is independent of the size of the data set. In this paper we offer a first theoretical study of such dimension reduction methods.

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 [19], and popular data-segmentation and clustering algorithms [39]. 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.

Our main result is a *local dimension reduction* lemma which replaces the dependency in the global size of the data n in the Johnson-Lindenstrauss bound with a local parameter.

We then apply our lemma to provide dimension reduction for data with low "intrinsic dimension", often measured by the doubling dimension [6, 21] of the data set. We show that the snowflake version of the data, where distances are raised to some fixed fractional power, can be embedded in dimension close to the doubling dimension. This result provides a nearly tight bound to this problem, a variant of Assouad's problem [6], recently raised and studied by Gottlieb and Krauthgamer [20].

1.1 Local Dimension Reduction

Two 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 [40], 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. The second, Local Linear Embedding [36], 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. 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 [15], motion planning in robotics [24], data-mining microarray assays [32], and face recognition [22]. 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 [12], generative probabilistic models [14], semidefinite programming [42] or neural networks [23].

Despite their wide appeal, all of these algorithms assume some sort of manifold model underlies the data, and make implicit assumptions about intrinsic curvature, Riemannian 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 [7], have problems recovering non-convex domains or manifolds with nontrivial homology [17], and cannot recover manifold structures that require more than one coordinate chart [33].

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 the theorem of Bourgain [13] stating that that any n-point metric space embeds in L_2 with $O(\log n)$ distortion, and the Johnson-Lindenstrauss [25] dimension reduction lemma. Both these theorems have many algorithmic consequences.

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 theorem (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, and [4] remove this assumption at the cost of increasing the dimension to $O(\log^2 k)$. 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* [41] 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 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 [8]. Here, we present an embedding of dimension that has no dependence on the volume. 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 *et al.* [2]. Another way to state our result is that the $1 + \epsilon$ distortion is preserved inside a core neighborhood of diameter at least $\Omega(\epsilon^{1.5}/\log k)$ factor of the diameter of the k-neighborhood. Some assumption of this form is necessary, as follows from a lower bound by Schechtman and Shraibman [37] showing that there are worst case examples where no near-isometric local dimension reduction method can beat the Johnson-Lindenstrauss bound. Prior to our work the only case where such a result was known is when the input set is isometric to an ultrametric [4].

For general metrics, this embedding is an immersion, but under the assumption that the metric has low intrinsic dimensionality (i.e., small doubling dimension) we can transform our immersion into a *global embedding* such that distances between far points can be bounded below so they don't intrude on the local structure. This extension to a global embedding can be useful in 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 that our embedding is suited to, the principal computational problem in vector quantization [19] is formally one of clustering (with ℓ_2^2 costs), but the parameters are different than in the clustering literature: primarily, one studies here the limit that the number of clusters, s, tends to ∞, while the distortion (the average distance to a codeword) tends to 0. This means that only the small distances between data points are germane to the problem. Known algorithms for construction of near-optimal clusterings are exponential in either s or the dimension of the space. Our embedding is well-suited to taking advantage of dimensionality reduction for vector quantization, since our target dimension depends only on the size of the small regions in which the L_2 distance needs to be preserved. Using our embedding, the vector quantization algorithm can be run in a low-dimensional space, and the clustering ("codebook") can then be lifted back to the original space.

Our approach for local dimension reduction combines several metric embedding techniques. We first employ probabilistic partitioning [9] of our metric space (Section 2). These partitions, developed in [1, 2, 4], decompose the metric space into clusters of bounded diameter and allow the coordinates of the embedding to smoothly transition between neighborhoods. As opposed to the standard decompositions where cluster diameters are similar, the partitions of [4] allow varying diameters to capture neighborhoods of similar cardinality. The idea is to apply for each of the clusters of the partition separately a dimension reduction method on the points within the cluster and then assemble these embedded neighborhoods into a global immersion.

While this idea sounds simple it in fact fails if we attempt to directly apply the Johnson-Lindenstrauss embedding method in each of the clusters. The reason is that the values the embedding takes may be as large as the diameter of the cluster and that may temper the Lipschitz condition between points in separate clusters (that is the ratio of the embedded distance to the original distance may be unbounded). To avoid that we need to combine the dimension reduction method with a truncation mechanism. While there are several ways in which this may be done we introduce a natural and elegant mechanism for this aim which we call the randomized Nash device. To ensure the Lipschitz condition we finally apply a smoothing operator.

Our methods owe a substantial debt to seminal papers in several areas of mathematics. Pursuing a line of work begun by Whitney [43, 44], 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 [31]. 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 [34] 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. Our Nash device can also be viewed as a discretized version the the continuous truncation technique of Schoenberg [38] which has appeared in the embedding literature (e.g. [29, 28, 20]). (These methods, combined with the Johnson-Lindenstrauss dimension reduction, could have replaced the Nash device, but the latter is itself elegant, computationally efficient and simple to use, and may be of independent interest).

The existence of our embedding is guaranteed using the Lovász Local Lemma[18], and we rely on algorithmic implementation of the LLL by Moser and Tardos [30] to provide a randomized algorithm to generate our embeddings.[1].

Our main contribution is in the combination of these various ingredients to allow local dimension reduction. Following our work, this methodology has been applied in [20] in additional cases of dimensionality reduction. We mainly focus on applying these tools to obtain a *near optimal* local dimension reduction. Most notably, obtaining the near optimal bound requires a delicate probabilistic argument. The embedding must compose the coordinates associated with the probabilistic partitions and those associated with the Nash-type dimension reduction in an interlacing manner. The analysis follows with carefully balancing the

[1]We note that the application of the LLL together with probabilistic partitions was first applied in [26].

contributions of the different components through the dependencies of the relevant probabilistic events.

In some applications it may be important that the dimension reduction procedure will keep the embedded distant pairs away from the local neighborhoods. In general, this is impossible if no further assumptions are made. However, under the additional assumption that the metric space has low doubling dimension [6, 21] we ensure that our mapping has this property.

1.2 Dimension Reduction for Snowflakes

Let X be a subset of Euclidean space. The doubling constant of X is the minimum λ such that every ball can be covered by λ balls of half the radius. The *doubling dimension* of X is defined as $\dim(X) = \log_2 \lambda$. The question of whether the dimension bound in the Johnson-Lindensrauss lemma can be reduced to $O(\epsilon^{-2}\dim(X))$ has been posed by several researchers [27, 21, 3]. While this question remains open, it has been recently asked by Gottlieb and Krauthgamer [20] if a result along this line is possible for the "snowflake" version of the metric, i.e, if the distance function $d(x, y) = \|x-y\|$ is replaced with $d^\alpha(x, y) = \|x-y\|^\alpha$ for some $0 < \alpha < 1$. Such an embedding may suffice for certain applications. From a mathematical standpoint, this problem is motivated by Assouad's theorem [6] which states that the snowflake version of any metric space can be embedded in Euclidean space with dimension and distortion depending solely on the doubling dimension. Gottlieb and Krauthgamer [20] use a similar approach to ours to prove that such a dimension reduction is possible where the target dimension is $\tilde{O}((1-\alpha)^{-3}\epsilon^{-4}(\dim(X))^2)$. We observe that the main ingredient needed in the solution for this problem is a local dimension reduction theorem. Using a variant (in fact a simplified version) of our main local dimension reduction theorem (Theorem 1) we improve their result to a nearly tight bound: $\tilde{O}((1-\alpha)^{-2}\epsilon^{-3}\dim(X))$.

This theorem has applications for distance labeling schemes, problems such as nearest neighbor search where only relative relation between distances need to be preserved, and optimization problems where the objective function is composed of powers of distances, e.g., clustering problems.

1.3 Structure of the Paper

In Section 2 we provide the and background on the probabilistic partitions that we use. Theorem 1 is proved in Section 3. The local Nash-device is described in Section 3.1. We first give the main component of the embedding in Section 3.2 which provides the guarantee for "close" pairs. Then in Section 3.3 we provide the complete definition of the embedding which now deals with farther pairs that are still within the range of application of our main theorem (Theorem 1). In Section 4 we show how to extend the embedding to deal with all pairs and maintain separation of local and distant pairs (Theorem 2). Finally, in Section 5 we prove the dimension reduction for snowflakes (Theorem 3).

2 Preliminaries

We start with some basic definitions: 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 smallest radius r such that $|B(x, r)| \geq k$. For a pair $x, y \in X$, define: $\Delta_k(x, y) = \max\{\Delta_k(x), \Delta_k(y)\}$

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 $\mathrm{diam}(X) = \max_{x,y \in X} d(x, y)$.

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 [4]:

Definition 1 (Probabilistic Partition). A *partition* P of X is a collection of disjoint set of *clusters* $\mathcal{C}(P) = \{C_1, C_2, \ldots, C_t\}$ such that $X = \cup_j C_j$. A partition is called Δ-*bounded* where $\Delta : P \to \mathbb{R}^+$ if for all j, $\mathrm{diam}(C_j) \le \Delta(C_j)$. 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 for any $x \in X$ and $Z \subseteq X \setminus B(x, 16\Delta(P(x)))$: $\Pr[\mathcal{L}(x) | \bigwedge_{z \in Z} \mathcal{L}(z)] \ge \delta$.

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

Lemma 3 is a reformulation of Lemma 5 from [4]. A simple application of the Lovász Local Lemma implies:

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 $|T| \ge 8 \log k/\xi$ 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 $|T^{(\delta)}(x)| \ge (1-2\xi)|T|$ such that $\mathcal{L}_t^{(\delta)}(x)$ occurs for all $t \in T$.*

3 Local Dimension Reduction

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 $k \in \mathbb{N}$. Recall that for a point $x \in X$, $\Delta_k(x)$ denotes the smallest radius r such that $|B(x, r)| \ge k$, and for a pair $x, y \in X$: $\Delta_k(x, y) = \max\{\Delta_k(x), \Delta_k(y)\}$. Let $\Delta_k^*(x) = c_1 \epsilon \Delta_k(x)/\log k$, where $c_1 < 1$ is a universal constant, and $\Delta_k^*(x, y) = \max\{\Delta_k^*(x), \Delta_k^*(y)\}$.

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. For all x,$y \in X$, $\|\hat{\Phi}(x) - \hat{\Phi}(y)\| \le (1+\epsilon)\|x - y\|$

b. For all $x, y \in X$:

$$\|\hat{\Phi}(x) - \hat{\Phi}(y)\| \ge \begin{cases} (1+\epsilon)^{-1}\|x-y\| & \textit{if } \|x-y\| \le \sqrt{\epsilon}\Delta_k^*(x,y) \\ (1+\epsilon')^{-1}\|x-y\| & \textit{if } \|x-y\| = \sqrt{\epsilon'}\Delta_k^*(x,y) \text{ s.t. } \epsilon < \epsilon' \le 1 \\ \frac{1}{8}\Delta_k^*(x,y) & \textit{if } \Delta_k^*(x,y) < \|x-y\| \le \frac{1}{2}\Delta_k(x,y) \end{cases} \quad (1)$$

c. For all $x \in X$, $\|\hat{\Phi}(x)\| \le \Delta_k^*(x)$

We comment that property (c) is not needed in general but is useful for the application in Section 5.

We note that although Theorem 1 maintains $(1+\epsilon)$-distortion only in a core neighborhood within the k-neighborhood of a point, this implies $(1+\epsilon)$-distortion for *all* pairs within the entire k-neighborhood[2] if we demand that X satisfies a *weak growth rate* condition[3] (defined by [2]), where there exists a constant $\gamma < 1$ if for every $x \in X$ and $r_1, r_2 > 0$, $|B(x, r_2)| \le |B(x, r_1)|^{(r_2/r_1)^\gamma}$, and further assume $\gamma < 0.2$.

In the rest of this section we describe the embedding and analysis to prove Theorem 1[4]. 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. The core of the construction is presented in Section 3.2 where 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 $\|x - y\| \le \Delta_k^*(x, y)$, as well as the upper bound for all pairs. For farther neighbors, we use a simple additional construction in Section 3.3.

3.1 The Randomized Nash Device

In this section we introduce a new construct we call 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

$$\varphi(x; \sigma, \omega) = \frac{1}{\sigma} \begin{bmatrix} \cos(\sigma\omega' x) \\ \sin(\sigma\omega' x) \end{bmatrix} \tag{2}$$

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 [31], 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.

In this section we present a sequence of *random* parameter settings for these functions φ, first studied in [34], 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 (2) 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 Lipschitz. This intuition can be further reinforced by considering the expected distance between two points.

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

[2] The dimension can be bounded by: $O(\epsilon^{-2} \log k \lceil (\epsilon^{-3/2} \log k)^{\log_\alpha \beta} \rceil)$ and so for instance if $\beta = 2$ and $\alpha = \epsilon^{-3/2} \log k$ we get dimension $O(\epsilon^{-2} \log k)$. This bound is similar in flavor to bounds given in [2].

[3] The reason this condition is called weak is that it does not exclude rapidly expanding metrics.

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

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* $1 - \exp(-\frac{3D\epsilon^2}{128})$.

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* $1 - \exp(-\frac{D}{128})$.

The randomized embedding Θ maps onto a product of circles of varying radii, a subset of the $2D$-sphere. 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. The analysis of Lemma 6 appears in Appendix A.

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 Section 2. Specifically, we concern ourselves with the "close neighbors," pairs x and y satisfying $\|x-y\| \le \Delta_k^*(x,y)$ (for the lower bound, while the upper bound is proved for all pairs). 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 4, where $T = [D]$ and $\xi = \epsilon$. 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 $\{\omega_C | C \in P^{(t)}, 1 \le t \le D\}$ 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

$$\Phi^{(t)}(x) = A^{(t)}(x)\hat{\varphi}^{(t)}(x)$$

where,

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

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 $\Phi^{(t)}$: $\Phi = \frac{1}{\sqrt{D}} \bigoplus_{1 \le t \le D} \Phi^{(t)}$.

Analysis Overview: Our goal is to show that the embeddings Φ and the Nash-device based embeddings of Section 3.1 have similar distortion guarantees. The purpose of the padded probabilistic partitions and the smoothing amplitude function is to allow a smooth transition between the different local embeddings in different clusters. For a close pair the padded probabilistic partition guarantees that in $\approx 1-\epsilon$ of the coordinates they fall in the same cluster and therefore their distortion is governed by the local Nash-device based embedding, which still maintains its distortion guarantees over the random set of successful coordinates. With probability $\approx \epsilon$ that this fails we rely on the Lipschitz property (that the smoothing amplitude function provides) to make sure the distortion only deviates slightly and the overall distortion remains $1 + O(\epsilon)$. To enable this probabilistic argument our proof utilizes the Lovász Local Lemma, showing that the necessary constraints are satisfied everywhere with positive probability. The rest of this section is devoted to carrying out this proof strategy.

Embedding Analysis. We start with the following lemma which will be useful to bound the distance between embedded points:

Lemma 7. *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 $P^{(t)}(x) \neq P^{(t)}(y)$, $d(x, X \setminus P^{(t)}(x)) \geq 2\sigma^{(t)}(x)^{-1}$ and $d(y, X \setminus P^{(t)}(y)) \geq 2\sigma^{(t)}(y)^{-1}$, then $\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\| \leq \|x - y\|$.*
3. *If $P^{(t)}(x) = P^{(t)}(y)$, $\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\|^2 \leq \|x - y\|^2 + \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2$.*
4. *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

$$\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\| = \|A^{(t)}(x)\hat{\varphi}^{(t)}(x) - A^{(t)}(y)\hat{\varphi}^{(t)}(y)\|$$

We now proceed case by case.

For (1), note that since $\|\varphi^{(t)}(u)\| = 1$, we have

$$\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\| \leq A^{(t)}(x)\|\hat{\varphi}^{(t)}(x)\| + A^{(t)}(y)\|\hat{\varphi}^{(t)}(y)\| \leq A^{(t)}(x) + A^{(t)}(y)$$

For claim (2) we have that $A^{(t)}(x) + A^{(t)}(y) \leq d(x, X \setminus P^{(t)}(x)) + d(y, X \setminus P^{(t)}(y))$. Now 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. Claim (3) follows as $A^{(t)}(x) + A^{(t)}(y) \leq \sigma^{(t)}(x)^{-1} + \sigma^{(t)}(y)^{-1} \leq 2\max\{\sigma^{(t)}(x)^{-1}, \sigma^{(t)}(y)^{-1}\} \leq \max\{d(x, X \setminus P^{(t)}(x)), d(y, X \setminus P^{(t)}(y))\} \leq \|x - y\|$.

We now turn to claims (4). Assume $C := P^{(t)}(x) = P^{(t)}(y)$. Then

$$\|\Phi^{(t)}(x) - \Phi^{(t)}(y)\|^2 = (A^{(t)}(x) - A^{(t)}(y))^2 + A^{(t)}(x)A^{(t)}(y)\|\hat{\varphi}^{(t)}(x) - \hat{\varphi}^{(t)}(y)\|^2,$$

using $\|\varphi^{(t)}(u)\| = 1$. In this case 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 $A^{(t)}(y) = \sigma_C^{-1}$ then $A^{(t)}(x) \leq \sigma_C^{-1}$ 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 claim (5), we only need use the fact that $A^{(t)}(x) = A^{(t)}(y) = \sigma_C^{-1}$. □

We now proceed to proving Theorem 1. For $x, y \in X$, let us now classify the different coordinates t according to the cases of Lemma 7. Define the sets

$$\begin{aligned} T_{\neq}(x,y) &= \{t | P^{(t)}(x) \neq P^{(t)}(y)\}\ ,\ T_{=}(x,y) = \{t | P^{(t)}(x) = P^{(t)}(y)\} \\ T_{\circ}(x,y) &= \{t | d(x, X \setminus P^{(t)}(x)) \geq 2\sigma^{(t)}(x)^{-1} \wedge d(y, X \setminus P^{(t)}(y)) \geq 2\sigma^{(t)}(y)^{-1}\} \end{aligned} \tag{3}$$

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

$$\|\Phi(x) - \Phi(y)\|^2 \geq \frac{1}{D} \sum_{t \in T_{=}(x,y) \cap T_{\circ}(x,y)} \|\varphi^{(t)}(x) - \varphi^{(t)}(y)\|^2, \text{ and} \tag{4}$$

$$\|\Phi(x)-\Phi(y)\|^2 \le \frac{1}{D}\left[\sum_{t\in T_=(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2 + \sum_{t\in T_{\neq}(x,y)}\|x-y\|^2 + \sum_{t\in T\setminus T_\circ(x,y)}\|x-y\|^2\right]. \quad (5)$$

We now turn to show that the properties of the embedding hold with positive probability. For $t\in T$, let $\sigma^{(t)}(x,y)=\min\{\sigma^{(t)}(x),\sigma^{(t)}(y)\}$. Recall that we have applied Lemma 4 with $\xi=\epsilon$, so that $\delta=1-\epsilon$.

Consider $t\in T^{(\delta)}(x)$ then $B(x,\eta^{(\delta)}\cdot\Delta(P^{(t)}(x))\subseteq P^{(t)}(x)$, where $\eta^{(\delta)}=2^{-11}\epsilon/\ln k$. It follows that $d(x,X\setminus P^{(t)}(x))\ge \eta^{(\delta)}\cdot\Delta(P^{(t)}(x))\ge 2\sigma^{(t)}(x)^{-1}$, by definition. Similarly, if $t\in T^{(\delta)}(y)$ then $d(y,X\setminus P^{(t)}(y))\ge 2\sigma^{(t)}(y)^{-1}$. Hence, $T^{(\delta)}(x)\cap T^{(\delta)}(y)\subseteq T_\circ(x,y)$, implying that $|T\setminus T_\circ(x,y)|\le |T\setminus(T^{(\delta)}(x)\cap T^{(\delta)}(y))|\le |T\setminus T^{(\delta)}(x)|+|T\setminus T^{(\delta)}(y)|\le 4\epsilon D$, by Lemma 4. Plugging this bound into (5) we conclude that:

$$\|\Phi(x)-\Phi(y)\|^2 \le \frac{1}{D}\left[|T_=(x,y)|\cdot\frac{\sum_{t\in T_=(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2}{|T_=(x,y)|} + |T_{\neq}(x,y)|\cdot\|x-y\|^2\right] + 4\epsilon\|x-y\|^2. \quad (6)$$

Now consider pairs x,y that are close neighbors, that is: $\|x-y\|\le\Delta_k^*(x,y)$ 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 $\frac{1}{8}\sigma^{(t)}(x,y)^{-1}\le\Delta_k^*(x,y)\le\frac{1}{2}\sigma^{(t)}(x,y)^{-1}$ (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 we've seen that $d(x,X\setminus P(x))\ge 2\sigma^{(t)}(x)^{-1}$. Now consider $y\in X$ such that $\|x-y\|\le\Delta_k^*(x,y)\le\frac{1}{2}\sigma^{(t)}(x)^{-1}$ then $P^{(t)}(y)=P^{(t)}(x)$, implying that $T^{(\delta)}(x)\cap T^{(\delta)}(y)\subseteq T_=(x,y)\cap T_\circ(x,y)$ implying that $|T_=(x,y)\cap T_\circ(x,y)|\ge|T^{(\delta)}(x)\cap T^{(\delta)}(y)|\ge(1-4\epsilon)D$. Plugging this bound into 4 yields:

$$\|\Phi(x)-\Phi(y)\|^2 \;\ge\; (1-4\epsilon)\cdot\frac{\sum_{t\in T_=(x,y)\cap T_\circ(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2}{|T_=(x,y)\cap T_\circ(x,y)|},\ \text{and} \quad (7)$$

We will next apply the Local Lemma again over events related to the Nash-type embeddings in Section 3.1 for the different clusters. Define:

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

We define the following events for pairs. Let $A_U(x,y)$ be the event that $U(x,y)>(1+\epsilon)\|x-y\|^2$. For pairs x,y that are close neighbors, that is: $\|x-y\|\le\Delta_k^*(x,y)$. Let $\epsilon'(x,y)=\max\{\epsilon,\Delta_k^*(x,y)^{-2}\|x-y\|^2\}$, and define $A_L(x,y)$ be the event that $L(x,y)<(1-\epsilon'(x,y))\|x-y\|^2$. Let $A(x,y)=A_L(x,y)\vee A_U(x,y)$. If x,y are not close neighbors then $A(x,y)=A_U(x,y)$. The rest of the argument utilizes the Lovász Local Lemma to prove that there is positive probability that none of the events $A(x,y)$ occurs; the details that complete this argument can be found in Appendix B.

For property (c) of Theorem 1 note that it follows directly from the definition of Φ and Lemma 3.

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 [35, 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 $\hat{\Phi}=\Phi\oplus\Psi$. The analysis appears in Appendix C.

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. 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.

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:

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

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.

In recent work [3] it is shown that every metric space embeds in ℓ_2^D where $D = O(\dim(X)/\theta)$ with distortion $O(\log^{1+\theta} n)$, where $\dim(X)$ is the doubling dimension of X. Hence a possible 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 do 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 a refinement of this embedding using ideas from [2].[5]

In Appendix F we give a local scaling embedding for doubling metrics satisfying the weak growth rate condition[6]. By using this embedding for the component Υ as explained above we obtain the following theorem:

Theorem 2. *Let $k \in \mathbb{N}$, and X a discrete subset of $\mathbb{R}^U$. Suppose that X satisfies a weak growth rate condition then for any $0 < \epsilon, \theta \le 1$ 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 $\|x-y\| \ge \frac{1}{2}\Delta_k(x,y)$ then:*

$$\|\bar{\Phi}(x) - \bar{\Phi}(y)\| \ge \Delta_k(x,y) \cdot c_2\theta\sqrt{\epsilon}/\log^{1+\theta} k, \tag{8}$$

for some universal constant c_2.

5 Dimension Reduction for Euclidean Snowflakes

In this section we provide a dimension reduction for snowflakes of finite subsets of Euclidean space.

Theorem 3. *Given a subset X of Euclidean space, for every $0 < \alpha < 1$ there exists an embedding $\Phi : X \to \mathbb{R}^D$, where $D = O(\frac{\log(1/\epsilon)}{1-\alpha}\epsilon^{-3}\dim(X)(\log(\dim(X)) + \frac{\log(1/\epsilon)}{1-\alpha}))$ such that for all $x, y \in X$:*

$$(1+\epsilon)^{-1}\|x-y\|^\alpha \le \|\Phi(x)-\Phi(y)\| \le (1+\epsilon)\|x-y\|^\alpha$$

[5] Note that an alternate choice for Υ could be our snowflake embedding of Section 5, which would provide lower bound on the contraction of distant pairs which is a function of their distance. However, we prefer a bound as a function of k.

[6] X satisfies a *weak growth rate* condition [2] if for some constants $\alpha > \beta \ge 1$ if for every $x \in X$ and $r > 0$, $|B(x,\alpha r)| \le |B(x,r)|^\beta$, and further assume $\log_\alpha \beta < 0.2$.

The proof proceeds in two steps: we first use Theorem 1 to give an embedding of pairs of points whose distances fall in a single scale in dimension $\tilde{O}(\epsilon^{-2}\dim(X))$ and then use it to obtain the embedding in Theorem 3 that preserves small distortion for snowflakes in all scales simultaneously.

We apply a variant of Theorem 1 (in fact we only use a special case of it where all k neighborhoods are bounded below by a fixed parameter). We observe that the function $\Delta_k(x)$ can be replaced by any Lipschitz function[7] $\bar{\Delta}_k(x)$ bounded above by $\Delta_k(x)$, applying the same proof[8]. In particular, for our application we need to introduce a parameter $\Delta > 0$, and define: $\bar{\Delta}_k(x) = \min\{\Delta_k(x), \Delta\}$ and let $\Delta_k^*(x) = c_1\epsilon\bar{\Delta}_k(x)/\log k$. This provides the one scale embedding:

Lemma 8. *Given a subset X of Euclidean space, for every $r > 0$ and $\epsilon, \delta > 0$, there exists an embedding $\bar{\Phi} : X \to \mathbb{R}^D$, where $D = O(\epsilon^{-2}\dim(X)(\log(\dim(X)) + \log((\epsilon\delta)^{-1})))$, with the following properties:*

1. $\|\bar{\Phi}(x) - \bar{\Phi}(y)\| \le \|x - y\|$
2. *For all $x, y \in X$ such that $\delta r \le \|x - y\| \le r$: $\|\bar{\Phi}(x) - \bar{\Phi}(y)\| \ge (1+\epsilon)^{-1}\|x - y\|$*
3. *For all $x \in X$, $\|\bar{\Phi}(x)\| \le r/\sqrt{\epsilon}$*

Proof. Let $\hat{X}$ be an $\epsilon\delta r$-net of X. We show the theorem holds for $\hat{X}$. As in [20] claim (1) of the theorem can be easily obtained by using Kirszbraun's extension theorem[9], and observing that if $x, y \in X$ are such that $\delta r \le \|x - y\| \le r$ then there exist $x', y' \in \hat{X}$ such that $\delta(1-2\epsilon)r \le \|x' - y'\| \le r(1+2\epsilon)$ and a small adaptation of the parameters provides the statement in the theorem.

Let $k = 2^{c'\dim(X)(\log(\dim(X))+\log((\epsilon\delta)^{-1}))}$, where c' is an appropriate constant to be determined, and let $\Delta = \log k/(c_1\epsilon) \cdot r/\sqrt{\epsilon}$. Let x be an arbitrary point $x \in \hat{X}$ then $|B_{\hat{X}}(x, \Delta)| \le 2^{\dim(X)\log(\Delta/(\epsilon\delta r))} \le 2^{\dim(X)\log(\log k/(c_1\epsilon^3\delta))} < k$ (for an appropriate choice of c') and therefore for all $x \in \hat{X}$, $\Delta_k(x) > \Delta$ and so $\Delta_k^*(x) = r/\sqrt{\epsilon}$. The lemma now follows from the variant of Theorem 1 described above. □

Theorem 3 follows from a delicate application of Assouad's technique [6] (a similar somewhat more involved argument was used in [20] and it may also be seen as inspired by [10]). The proof is deferred to Appendix D.

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. Embedding metric spaces in their intrinsic dimension. In *Proceedings of the 19th ACM-SIAM Symp. on Discrete Algorithms*, 2008.

[4] Ittai Abraham, Yair Bartal, and Ofer Neiman. On low dimensional local embeddings. In *Proceedings of the 20th ACM-SIAM Symp. on Discrete Algorithms*, pages 875–884, 2009.

[7] Note that $\Delta_k(x)$ is itself Lipschitz.

[8] In fact the main thing that needs to be verified is that Lemma 3 holds with this definition. Moreover since we only use the theorem in the case that $\Delta_k(x) > \Delta$ for all x – a simpler lemma can be used to this aim where all clusters have the same parameter Δ and whose proof is simpler (this forms a special case of Theorem 1).

[9] Kirszbraun's theorem is not really necessary as an extension property can be shown to hold directly for the embedding in Theorem 1.

[5] Noga Alon. Perturbed identity matrices have high rank: proof and applications. *Combinatorics, Probability and Computing*, 18:3–15, 2009.

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

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

[8] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 2007. To appear. Preprint available at `http://www.eecs.umich.edu/~wakin/publications.html`.

[9] 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.

[10] Yair Bartal and Manor Mendel. Dimension reduction for ultrametrics. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 664–665. Society for Industrial and Applied Mathematics, 2004.

[11] Yair Bartal, Ben Recht, and Leonard Schulman. A nash-type dimensionality reduction for discrete subsets of l_2. In *http://www.ist.caltech.edu/ brecht/papers/07.BRS.local-dim.pdf*, 2007.

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

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

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

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] Lee-Ad Gottlieb and Robert Krauthgamer. A nonlinear approach to dimension reduction. In *arXiv:0907.5477v1 (also submitted to FOCS '10)*, 2009.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] Robert Krauthgamer and James R. Lee. The intrinsic dimensionality of graphs. In *Annual ACM Symposium on Theory of Computing*, pages 438–447, 2003.

[27] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.

[28] James R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *SODA '05: Proceedings of the 18th annual ACM-SIAM symposium on Discrete algorithms*, pages 92–101, 2005.

[29] Manor Mendel and Assaf Naor. Euclidean quotients of finite metric spaces. *Advances in Mathematics*, 189(2):451–494, 2004.

[30] Robin A. Moser and Gabor Tardos. A constructive proof of the general lov?sz local lemma. *Journal of the ACM*, 57(2), 2010.

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

[32] 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.

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

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

[35] 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.

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

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

[38] I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3), 1938.

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

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

[41] 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.

[42] 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.

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

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

A Randomized Nash Device Analysis

In section we prove Lemma 6

To prove part (a) of the lemma 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

$$\|\Theta(x)-\Theta(y)\|^2 = \frac{1}{D}\sum_{t=1}^{D}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2 = \frac{1}{D}\sum_{t=1}^{D}\frac{2}{\sigma_t^2}\left(1-\cos\left(\sigma_t\tau_t\right)\right) \leq \frac{1}{D}\sum_{t=1}^{D}\tau_t^2\,. \qquad (9)$$

It therefore follows that

$$\mathbf{Pr}\left[\|\Theta(x)-\Theta(y)\|^2 \geq (1+\epsilon)\ell^2\right] \leq \mathbf{Pr}\left[\frac{1}{D}\sum_{t=1}^{D}\tau_t^2 \geq (1+\epsilon)\ell^2\right] \leq e^{-\frac{D}{2}\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)} \qquad (10)$$

where the second inequality is a well known known concentration inequality a χ-squared random variable (see, e.g., [16]).

Parts (b) and (c) require a more detailed verification, but follow from a Chernoff Bound type analysis. 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.

Using the identity $\|\Theta(x)-\Theta(y)\|^2 = \frac{1}{D}\sum_{t=1}^{D}\frac{2}{\sigma_t^2}\left(1-\cos\left(\sigma_t\tau_t\right)\right)$ we have for any $u>0$

$$\begin{aligned}
&\mathbf{P}\left[\|\Theta(x)-\Theta(y)\|^2 \leq u\right] && (11)\\
=&\mathbf{P}\left[\frac{1}{D}\sum_{t=1}^{D}\frac{2}{\sigma_t^2}\left(1-\cos\left(\sigma_t\tau_t\right)\right) \leq u\right] && (12)\\
=&\mathbf{P}\left[\sum_{t=1}^{D}\frac{2}{\sigma_t^2}\left(\cos\left(\sigma_t\tau_t\right)-1\right)+uD \geq 0\right] && (13)\\
=&\mathbf{P}\left[\exp\left(s\sum_{t=1}^{D}\frac{2}{\sigma_t^2}\left(\cos\left(\sigma_t\tau_t\right)-1\right)+uDs\right) \geq 1\right] \qquad \forall s>0 && (14)\\
\leq&\mathbb{E}\left[\exp\left(s\sum_{t=1}^{D}\frac{2}{\sigma_t^2}\left(\cos\left(\sigma_t\tau_t\right)-1\right)+uDs\right)\right] \qquad \text{(by Markov's Inequality)} && (15)\\
=&\exp\left(uDs\right)\mathbb{E}\left[\prod_{t=1}^{D}\exp\left(s\frac{2}{\sigma_t^2}\left(\cos\left(\sigma_t\tau_t\right)-1\right)\right)\right] && (16)\\
=&\exp\left(uDs\right)\prod_{t=1}^{D}\mathbb{E}_{\tau_t}\left[\exp\left(s\frac{2}{\sigma_t^2}\left(\cos\left(\sigma_t\tau_t\right)-1\right)\right)\right]\,. && (17)
\end{aligned}$$

We first bound the expectations with respect to τ_t. Let τ be a zero-mean Gaussian random variable with variance ℓ^2. Since $\exp(t) \leq 1+t+t^2/2$ for all $t \leq 0$, we have, for all $s,\sigma>0$,

$$\begin{aligned}
\exp\left(s\tfrac{2}{\sigma^2}\left(\cos\left(\sigma\tau\right)-1\right)\right) &\leq 1+\tfrac{2}{\sigma^2}(\cos(\sigma\tau)-1)s+\tfrac{2}{\sigma^4}[\cos(\sigma\tau)-1]^2s^2 && (18)\\
&= 1+\tfrac{2}{\sigma^2}(\cos(\sigma\tau)-1)s+\tfrac{2}{\sigma^4}[1-2\cos(\sigma\tau)+\cos^2(\sigma\tau)]s^2 && (19)\\
&= 1+\tfrac{2}{\sigma^2}(\cos(\sigma\tau)-1)s+\tfrac{1}{\sigma^4}[3-4\cos(\sigma\tau)+\cos(2\sigma\tau)]s^2\,. && (20)
\end{aligned}$$

Using the fact that $\mathbb{E}[\cos(z\tau)] = \exp(-\ell^2 z^2/2)$ for all $z \in \mathbb{R}$, we can compute the expectation of (20)

$$\mathbb{E}\left[\exp(s\tfrac{2}{\sigma^2}[\cos(\sigma\tau)-1])\right] \le \mathbb{E}\left[1+\tfrac{2}{\sigma^2}(\cos(\sigma\tau)-1)s+\tfrac{1}{\sigma^4}[3-4\cos(\sigma\tau)+\cos(2\sigma\tau)]s^2\right] \tag{21}$$

$$= 1+\tfrac{2}{\sigma^2}\left(\exp(-\tfrac{1}{2}\sigma^2\ell^2)-1\right)s \tag{22}$$
$$+\tfrac{1}{\sigma^4}\left(3-4\exp(-\tfrac{1}{2}\sigma^2\ell^2)+\exp(-2\sigma^2\ell^2)\right)s^2\,.$$

The negative of the term linear in s is equal

$$b(\sigma) := \tfrac{2}{\sigma^2}\left(1-\exp(-\tfrac{1}{2}\sigma^2\ell^2)\right) \tag{23}$$

and that the term quadratic in s is equal to

$$a(\sigma) := \tfrac{1}{4}b(\sigma)^2\left((1+\exp(-\tfrac{1}{2}\sigma^2\ell^2))^2+2\right)\,. \tag{24}$$

Both $b(\sigma)$ and $a(\sigma)$ are positive decreasing functions of $\sigma > 0$.

To complete the proof, suppose we can find an $s_0 > 0$ such that

$$b(\sigma_t)s_0 - a(\sigma_t)s_0^2 < 1 \qquad \text{for all } 1 \le t \le D \tag{25}$$
$$b(\sigma_t)s_0 - a(\sigma_t)s_0^2 \ge \gamma + us_0 \qquad \text{for all } 1 \le t \le D\,. \tag{26}$$

for some constant $\gamma > 0$. Then, using the inequality $\log(1-t) \le -t$ for all $t < 1$ and the preceding analysis, we would have the probability of $\|\Theta(x)-\Theta(y)\|^2 < u$ being at most

$$\exp(uDs_0)\prod_{t=1}^{D}(1-b(\sigma_t)s_0+a(\sigma_t)s_0^2) = \exp\left(uDs_0+\sum_{t=1}^{D}\log(1-b(\sigma_t)s_0+a(\sigma_t)s_0^2)\right) \tag{27}$$
$$\le \exp\left(\sum_{t=1}^{D} us_0 - b(\sigma_t)s_0 + a(\sigma_t)s_0^2\right) \tag{28}$$
$$\le \exp(-\gamma D)\,. \tag{29}$$

Part (b) would be proven if we find an s_0 for which (25) and (26) hold when $\ell < \sigma_{\mathbf{m}}{}^{-1}\sqrt{\epsilon}$ with $u = (1-\epsilon)\ell^2$ and $\gamma = \frac{3}{128}\epsilon^2$. For part (c), we need to find an s_0 to show that when $\ell > (\sqrt{2}\sigma)^{-1}$, (25) and (26) hold with $u = (4\sigma^2)^{-1}$ and $\gamma = \frac{1}{128}$.

The strategy for both parts (b) and (c)is the same. We show that choosing s_0 such that the equality is attained in (26) when $\sigma = \sigma_{\mathbf{m}}$ suffices. That is, we set

$$s_0 = \frac{b(\sigma_{\mathbf{m}}) - u - \sqrt{(b(\sigma_{\mathbf{m}})-u)^2 - 4a(\sigma_{\mathbf{m}})\gamma}}{2a(\sigma_{\mathbf{m}})}\,. \tag{30}$$

If this choice of s_0 is positive, then (25) and (26) are automatically satisfied. For (25), note that for all $\sigma > 0$,

$$b(\sigma)s_0 - a(\sigma)s_0^2 = b(\sigma)s_0(1-\tfrac{1}{4}\left((1+\exp(-\tfrac{1}{2}\sigma^2\ell^2))^2+2\right)b(\sigma)s_0) \le b(\sigma)s_0(1-\tfrac{3}{4}b(\sigma)s_0) \le \tfrac{1}{3}\,. \tag{31}$$

For (26), a and b are both decreasing functions of σ so we have

$$(b(\sigma_t)-u)s_0 - a(\sigma_t)s_0^2 \ge (b(\sigma_{\mathbf{m}})-u)s_0 - a(\sigma_t)s_0^2 = \gamma + a(\sigma_{\mathbf{m}})s_0^2 - a(\sigma_t)s_0^2 \ge \gamma \tag{32}$$

All that remains is to verify that s_0 is positive for the values of u and γ in parts (b) and (c) respectively. Note that s_0 is positive if $b(\sigma_{\mathbf{m}}) > u$ and $b(\sigma_{\mathbf{m}}) - u \ge 2\sqrt{\gamma a(\sigma_{\mathbf{m}})}$. Certainly, if the latter inequality is strict, it implies the first, so we focus on the latter in the remainder of the argument.

For Lemma 6 (b), we set $u = (1-\epsilon)\ell^2$ and $\gamma = \frac{3}{128}\epsilon^2$. Rearranging terms, we must show $b(\sigma_{\mathbf{m}}) - \epsilon\sqrt{\frac{3}{128}a(\sigma_{\mathbf{m}})} \geq (1-\epsilon)\ell^2$ whenever $\ell \leq \sigma_{\mathbf{m}}{}^{-1}\sqrt{\epsilon}$. That is, plugging in our definitions for $a(\sigma_{\mathbf{m}})$, and $b(\sigma_{\mathbf{m}})$, we must show

$$(1-\exp(-\tfrac{1}{2}\sigma_{\mathbf{m}}^2\ell^2))\left(1-2\epsilon\sqrt{\tfrac{3}{128}}\sqrt{(1+\exp(-\tfrac{1}{2}\sigma_{\mathbf{m}}^2\ell^2))^2+2}\right) \geq (1-\epsilon)\frac{\sigma_{\mathbf{m}}^2\ell^2}{2} \tag{33}$$

for all $\ell \leq \sigma_{\mathbf{m}}{}^{-1}\sqrt{\epsilon}$. Using the bounds $1-\exp(-t) \geq -t+t^2/2$ for $t \geq 0$ and $(1+\exp(-\frac{1}{2}\sigma_{\mathbf{m}}^2\ell^2))^2+2 \leq 6$, we can compute

$$\begin{aligned}
&\left(1-\exp\left(-\frac{\sigma_{\mathbf{m}}^2\ell^2}{2}\right)\right)\left(1-2\epsilon\sqrt{\tfrac{3}{128}}\sqrt{\left(1+\exp\left(-\frac{\sigma_{\mathbf{m}}^2\ell^2}{2}\right)\right)^2+2}\right) && (34)\\
&\geq \left(\frac{\sigma_{\mathbf{m}}^2\ell^2}{2}-\frac{\sigma_{\mathbf{m}}^4\ell^4}{8}\right)\left(1-\tfrac{3}{4}\epsilon\right) && (35)\\
&= \left(1-\frac{\sigma_{\mathbf{m}}^2\ell^2}{4}-\tfrac{3}{4}\epsilon\right)\frac{\sigma_{\mathbf{m}}^2\ell^2}{2}+\frac{3\sigma_{\mathbf{m}}^4\ell^4}{32} && (36)\\
&\geq (1-\epsilon)\frac{\sigma_{\mathbf{m}}^2\ell^2}{2}+\frac{3\sigma_{\mathbf{m}}^4\ell^4}{32} && (37)\\
&\geq (1-\epsilon)\frac{\sigma_{\mathbf{m}}^2\ell^2}{2} && (38)
\end{aligned}$$

Where (37) used the fact that $\ell \leq \sqrt{\epsilon}/\sigma_{\mathbf{m}}$.

The argument for part (c) is more or less the same, now with $u = (4\sigma^2)^{-1}$ and $\gamma = \frac{1}{128}$. We must show

$$(1-\exp(-\tfrac{1}{2}\sigma_{\mathbf{m}}^2\ell^2))\left(1-2\sqrt{\tfrac{1}{128}}\sqrt{(1+\exp(-\tfrac{1}{2}\sigma_{\mathbf{m}}^2\ell^2))^2+2}\right) \geq \frac{1}{8} \tag{39}$$

for all $\ell \geq (\sqrt{2}\sigma_{\mathbf{m}})^{-1}$. Since $\sigma_{\mathbf{m}}^2\ell^2 > 2$, it follows that

$$\begin{aligned}
&(1-\exp(-\tfrac{1}{2}\sigma_{\mathbf{m}}^2\ell^2))\left(1-2\sqrt{\tfrac{1}{128}}\sqrt{(1+\exp(-\tfrac{1}{2}\sigma_{\mathbf{m}}^2\ell^2))^2+2}\right)\\
&\qquad \geq \left(1-\exp(-\tfrac{1}{4})\right)\left(1-\sqrt{\tfrac{3}{16}}\right) \approx 0.1254 > \frac{1}{8}\,. && (40)
\end{aligned}$$

B Close Neighbors Analysis

In this section we provide the Local Lemma argument which complete the proof of Theorem 1 in Section 3.2.

We create a dependency graph G_A whose vertices are the events $A(x,y)$. Let d_{G_A} denote its maximum degree. 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 $\max\{\|x-x'\|, \|x-y'\|\} \leq \Delta(P^{(t)}(x)) \leq \Delta_k(x)/16$, by Lemma 3, and hence $x', y' \in B(x, \Delta_k(x))$. This implies that the number of such pairs is bounded by $d_{G_A} \leq \binom{k}{2}$.

Now, by part (a) of Lemma 6 the probability that $U(x,y) > (1+\epsilon)\|x-y\|^2$ is at most $e^{-D(\epsilon^2/4+\epsilon^3/6)} \leq k^{-2}/4$. For pairs x, y that are not close neighbors this implies that the probability that event $A(x,y)$ occurs is at most $1/(e(\binom{k}{2}+1)) \leq 1/(e \cdot d_{G_A}+1)$.

For pair x, y that are close neighbors we have that $\|x-y\| \leq \Delta_k^*(x,y) \leq \frac{1}{2}\sigma^{(t)}(x,y)^{-1}$, we have by Lemma 6(b) that the probability that $L(x,y) < (1-\max\{\epsilon, \sigma_{\mathbf{m}}^2\|x-y\|^2\})\|x-y\|^2$ is at most $e^{-3D\epsilon^2/128} \leq k^{-2}/4$, where $\sigma_{\mathbf{m}} \leq \max_{t\in T}\sigma^{(t)}(x,y) \leq \Delta_k^*(x,y)^{-1}/2$. Hence the probability the event $A(x,y)$ occurs is at most $k^{-2}/2 < 1/(e \cdot d_{G_A}+1)$. This complete the proof that the events $A(x,y)$ satisfy the conditions of the Local Lemma, implying that there is positive probability that none of these events occur. Therefore we have for any pair $x, y \in X$:

$$\|\Phi(x)-\Phi(y)\|^2 \leq \frac{|T_=(x,y)|}{D}\cdot U(x,y) + \frac{|T_{\neq}(x,y)|}{D}\cdot\|x-y\|^2 + 4\epsilon\|x-y\|^2 \leq (1+5\epsilon)\|x-y\|^2, \quad (41)$$

and for all close neighbors x, y such that $\|x-y\| \leq \Delta_k^*(x,y)$ we have:

$$\begin{aligned}\|\Phi(x)-\Phi(y)\|^2 &\geq (1-4\epsilon)L(x,y) \geq (1-4\epsilon-\max\{\epsilon, \Delta_k^*(x,y)^{-2}\|x-y\|^2\})\|x-y\|^2 \\ &\geq (1-5\max\{\epsilon, \epsilon'(x,y)\})\|x-y\|^2. \quad (42)\end{aligned}$$

C Embedding Farther Pairs Analysis

In this section we complete the proof of the embedding from Section 3.3.

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

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

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 $\Delta_k^*(x,y) \leq \|x-y\| \leq \frac{1}{2}\Delta_k(x,y)$ into two sets as follows: $W_= = \{\{x,y\} \mid |T_=(x,y)| \geq D/2\}$ and $W_{\neq} = \{\{x,y\} \mid |T_{\neq}(x,y)| > D/2\}$. For pairs in $W_=$ we show that the Φ component of the embedding gives a good lower bound on the distance, whereas for pairs in $W_{\neq}$ such a contribution is obtain from the Ψ component of the embedding.

Consider first a pair in $W_=$. Recall that

$$\|\Phi(x)-\Phi(y)\|^2 \geq \frac{\sum_{t\in T_=(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2}{D} \geq \frac{1}{2}\cdot\frac{\sum_{t\in T_=(x,y)}\|\varphi^{(t)}(x)-\varphi^{(t)}(y)\|^2}{|T_=(x,y)|}. \quad (43)$$

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 as $d_{G_B} \leq \binom{k}{2}$.

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 $e^{-D\epsilon^2/128} < k^{-2}/2$, 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 $k^{-2}/2 < 1/(e(\binom{k}{2}+1)) \leq 1/(e(d_{G_B}+1))$, 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 $W_=$,

$$\|\hat{\Phi}(x)-\hat{\Phi}(y)\|^2 \geq \|\Phi(x)-\Phi(y)\|^2 \geq \frac{1}{2}L_B(x,y) \geq 2^{-6}\Delta_k^*(x,y)^2, \quad (44)$$

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 $T'(x) = T^{(7/8)}(x)$ such that $|T'(x)| \geq (1 - \frac{2}{8})D = \frac{3}{4}D$ and for each $t \in T'(x)$, $B(x, \eta^{(3/4)}\Delta(P^{(t)}(x))) \subseteq P^{(t)}(x)$, and therefore $d(x, X \setminus P^{(t)}(x)) \geq \sigma^{(t)}(x)^{-1}/(4\epsilon)$, by definition. 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 $\|\Psi(x) - \Psi(y)\| < \frac{1}{8}\Delta_k^*(x, y)$. 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 $P^{(t)}(x) = P^{(t)}(x')$. As before we have that $\|x - x'\| \leq \Delta(P^{(t)}(x)) \leq \Delta_k(x)/16$, by Lemma 3, and hence $x' \in B(x, \Delta_k(x))$ and 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 $\|x' - y'_j\| < \frac{1}{2}\Delta_k(x', y'_j) = \frac{1}{2}\max\{\Delta_k(x'), \Delta_k(y'_j)\} \leq \frac{1}{2}\Delta_k(z)$. We conclude that $\|z - y'_j\| \leq \|z - x'\| + \|x' - y'_j\| < \Delta_k(z)$ and therefore all y'_js are in a ball around z containing fewer than k points so that $s < k$. We conclude that there are at most k^2 such pairs. The same calculation can be for the case that $P^{(t)}(y) = P^{(t)}(x')$, giving a total bound of $2k^2$ pairs, which provides an upper bound on the degree $d_{G_{B'}}$ of the dependency graph $G_{B'}$.

Now, let $T''(x) = T'(x) \cap W_{\neq}$ then $|T''(x)| \geq D/4$. Then for each $t \in T''(x)$ with probability at least $1/4$, $\nu(P^{(t)}(x)) = 1$ and $\nu(P^{(t)}(y)) = 0$, as $P^{(t)}(x) \neq P^{(t)}(y)$. Applying a Chernoff bound we have that the probability that there are less than $1/8$ fraction of the coordinates $t \in T''(x)$ such that $|\Psi^{(t)}(x) - \Psi^{(t)}(y)| \geq \sqrt{\epsilon} \cdot d(x, X \setminus P^{(t)}(x)) \geq \Delta_k^*(x, y)$ is at most $e^{-D/16}$. But this means that with probability $1 - e^{-D/16}$, $\|\Psi(x) - \Psi(y)\| \geq \frac{1}{\sqrt{8 \cdot 4}}\Delta_k^*(x, y) > \frac{1}{8}\Delta_k^*(x, y)$. Therefore the probability that event $B'(x, y)$ occurs is at most $e^{-D/16} \leq k^{-2}/4 < 1/(e(k^2 + 1)) \leq 1/(e(d_{G_{B'}} + 1)$, satisfying the condition for the Local Lemma. We can therefore conclude that with positive probability none of the events $B'(x, y)$ occur. Therefore for every $x, y \in W_{\neq}$ we have: $\|\hat{\Phi}(x) - \hat{\Phi}(y)\| \geq \|\Psi(x) - \Psi(y)\| \geq \frac{1}{8}\Delta_k^*(x, y)$, completing the proof of Theorem 1.

D Proof of Dimension Reduction for Snowflakes

of Theorem 3. We may assume $\alpha \geq 1/2$, otherwise we can apply the embedding for this case to imply the conclusion for smaller α as well. Let $p_A = \lceil \log_{1+\epsilon} \epsilon^{-\frac{4}{1-\alpha}} \rceil$, $p_B = \lfloor \log_{1+\epsilon} \epsilon^{-2} \rfloor$ and $p = 1 + p_A + p_B$. Define $\Delta_i = \mathrm{diam}(X)(1 + \epsilon)^{-i}$, where $i \in I$, $I = \{i \in \mathbb{Z} | -p_A \leq i \leq \log_{1+\epsilon}(\mathrm{diam}(X)) + 1 + p_B\}$. Let $\bar{\Phi}_i$ be the embedding of Lemma 8 with $r = \epsilon^{-2}\Delta_i$ and $\delta = (1 + \epsilon)^{-1}\epsilon^{\frac{4}{1-\alpha}+2}$. Let $\Psi_i = \bar{\Phi}_i/\Delta_i^{1-\alpha}$.

For $j \in [p]$ let $\Phi_j = \sum_{i \in I; i \equiv_p j} \Psi_i$ and $\Phi = \bigoplus_{j \in [p]} \Phi_j$. The final embedding is $\Phi/M : X \to l_2^{pD'}$, where M is a parameter to be determined later and D' is the dimension of the embedding of Lemma 8.

Fix some pair $x, y \in X$. Let i^* be such that $(1 + \epsilon)^{-1}\Delta_{i^*} \leq \|x - y\| \leq \Delta_{i^*}$.

Let $A = \{i^* - p_A, \ldots, i^*, \ldots, i^* + p_B\}$, then for each $i \in A$: $\epsilon^{\frac{4}{1-\alpha}} \leq \Delta_{i^*}/\Delta_i \leq \epsilon^{-2}$ and therefore $(1 + \epsilon)^{-1}\epsilon^{\frac{4}{1-\alpha}}\Delta_i \leq \|x - y\| \leq \epsilon^{-2}\Delta_i$. Then it follows from Lemma 8 that for $i \in A$:

$$(1 + \epsilon)^{-1}\frac{\|x - y\|}{\Delta_i^{1-\alpha}} \leq \|\Psi_i(x) - \Psi_i(y)\| \leq \frac{\|x - y\|}{\Delta_i^{1-\alpha}}$$

We also have

$$\sum_{i'<i,i'\equiv_p i} \|\Psi_{i'}(x)-\Psi_{i'}(y)\| \sum_{i'<i,i'\equiv_p i} \frac{\|x-y\|}{\Delta_{i'}^{1-\alpha}} \leq \frac{\|x-y\|}{\Delta_i^{1-\alpha}} \sum_{j=1}^{\infty} (1+\epsilon)^{-(1-\alpha)p\cdot j} \leq (1+\epsilon)\epsilon^4 \frac{\|x-y\|}{\Delta_i^{1-\alpha}}$$

Using the bound $\|\bar{\Phi}_i(z)\| \leq \epsilon^{-2}\Delta_i/\sqrt{\epsilon} = \epsilon^{-2.5}\Delta_i$ from Lemma 8 and assuming $\alpha \geq 1/2$ we have

$$\begin{aligned}
\sum_{i'>i,i'\equiv_p i} \|\Psi_{i'}(x)-\Psi_{i'}(y)\| &\leq \sum_{i'>i,i'\equiv_p i} \frac{2\epsilon^{-2.5}\Delta_{i'}}{\Delta_{i'}^{1-\alpha}} \leq 2\epsilon^{-2.5} \sum_{i'>i,i'\equiv_p i} \Delta_{i'}^{\alpha} \\
&\leq 2\epsilon^{-2.5}\Delta_i^{\alpha} \sum_{j=1}^{\infty} (1+\epsilon)^{-\alpha p\cdot j} \leq 2(1+\epsilon)^2\epsilon^{1.5} \frac{\|x-y\|}{\Delta_i^{1-\alpha}}
\end{aligned}$$

Hence we get for $\epsilon < 1/8$:

$$\|\sum_{i'\equiv_p i} (\Psi_{i'}(x)-\Psi_{i'}(y))\|^2 \leq \left(\|\Psi_i(x)-\Psi_i(y)\| + \sum_{i'\neq i,i'\equiv_p i} \|\Psi_{i'}(x)-\Psi_{i'}(y)\| \right)^2 \leq (1+\epsilon)^2 \left(\frac{\|x-y\|}{\Delta_i^{1-\alpha}}\right)^2$$

$$\|\sum_{i'\equiv_p i} (\Psi_{i'}(x)-\Psi_{i'}(y))\|^2 \geq \left(\|\Psi_i(x)-\Psi_i(y)\| - \sum_{i'\neq i,i'\equiv_p i} \|\Psi_{i'}(x)-\Psi_{i'}(y)\| \right)^2 \geq (1-\epsilon)^2 \left(\frac{\|x-y\|}{\Delta_i^{1-\alpha}}\right)^2$$

Summing over all $i \in A$ we get

$$(1-\epsilon)^2 \sum_{i\in A} \left(\frac{\|x-y\|}{\Delta_i^{1-\alpha}}\right)^2 \leq \|\Phi(x)-\Phi(y)\|^2 \leq (1+\epsilon)^2 \sum_{i\in A} \left(\frac{\|x-y\|}{\Delta_i^{1-\alpha}}\right)^2$$

Finally,

$$\sum_{i\in A} \left(\frac{\|x-y\|}{\Delta_i^{1-\alpha}}\right)^2 = \left(\frac{\|x-y\|}{\Delta_{i^*}^{1-\alpha}}\right)^2 \sum_{i\in A} (1+\epsilon)^{2(1-\alpha)(i^*-i)} = \left(\frac{\|x-y\|}{\Delta_{i^*}^{1-\alpha}}\right)^2 \sum_{j=-p_A}^{p_B} (1+\epsilon)^{2(1-\alpha)j}$$

Choosing $M = \sum_{j=-p_A}^{p_B} (1+\epsilon)^{(1-\alpha)j}$ and observing that $\|x-y\|^{\alpha} \leq \frac{\|x-y\|}{\Delta_{i^*}^{1-\alpha}} \leq (1+\epsilon)^{\alpha}\|x-y\|^{\alpha}$ completes the proof. □

E Probabilistic Partitions Preliminaries

E.1 Preliminaries

Consider a finite metric space (X,d) and let $n = |X|$. The *diameter* of X is denoted $\mathrm{diam}(X) = \max_{x,y\in X} d(x,y)$. For a point x and $r \geq 0$, the ball at radius r around x is defined as $B_X(x,r) = \{z \in X | d(x,z) \leq r\}$. We omit the subscript X when it is clear form the context.

The following definitions are used in the context of partition-based embeddings into L_p:

Definition 9. The local growth rate of $x \in X$ at radius $r > 0$ for a given scale $\gamma > 0$ is defined as

$$\rho(x,r,\gamma) = |B(x,r\gamma)|/|B(x,r/\gamma)|.$$

Given a subspace $Z \subseteq X$, the minimum local growth rate of Z at radius $r > 0$ and scale $\gamma > 0$ is defined as $\rho(Z,r,\gamma) = \min_{x\in Z} \rho(x,r,\gamma)$. The minimum local growth rate at radius $r > 0$ and scale $\gamma > 0$ is defined as $\bar{\rho}(x,r,\gamma) = \rho(B(x,r),r,\gamma)$.

The following simple fact about minimum local growth rate is useful:

Claim 10. *Let $x, y \in X$, let $\gamma > 0$ and let r be such that $2(1 + 1/\gamma)r < d(x, y) \leq (\gamma - 2 - 1/\gamma)r$, then*

$$\max\{\bar{\rho}(x, r, \gamma), \bar{\rho}(y, r, \gamma)\} \geq 2.$$

E.2 Uniformly Padded Probabilistic Partitions

We start with describing the basic definition that captures the properties needed for the application for embeddings:

Definition 11 (Partition). Let (X, d) be a finite metric space. A partition P of X is a collection of disjoint set of clusters $\mathcal{C}(P) = \{C_1, C_2, \ldots, C_t\}$ such that $X = \cup_j C_j$. The sets C_j are called clusters. For $x \in X$ we denote by $P(x)$ the cluster containing x. Given $\Delta > 0$, a partition is Δ-*bounded* if for all $1 \leq j \leq t$, $\mathrm{diam}(C_j) \leq \Delta$.

Definition 12 (Uniform Function). Given a partition P of a metric space (X, d), a function f defined on X is called *uniform* with respect to P if for any $x, y \in X$ such that $P(x) = P(y)$ we have $f(x) = f(y)$.

Definition 13 (Probabilistic Partition). 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. Given $\Delta > 0$, $\hat{\mathcal{P}}$ is Δ-bounded if each $P \in \mathcal{P}$ is Δ-bounded.

Definition 14 (Uniformly Padded Local PP). Given $\Delta > 0$ and $0 < \delta \leq 1$, let $\hat{\mathcal{P}}$ be a Δ-bounded probabilistic partition of (X, d). Given collection of functions $\eta = \{\eta_P : X \to [0, 1] | P \in \mathcal{P}\}$ such that η_P is a uniform function with respect to P. We say that $\hat{\mathcal{P}}$ is a (η, δ)-*uniformly padded local probabilistic partition* if the event $B(x, \eta_P(x)\Delta) \subseteq P(x)$ occurs with probability at least δ and is independent of the structure of the partition outside $B(x, 2\Delta)$.

Formally for all $C \subseteq X \setminus B(x, 2\Delta)$ and all partitions P' of C,

$$\Pr[B(x, \eta_P(x)\Delta) \subseteq P(x) \mid P_{|C} = P'] \geq \delta$$

E.3 Local Uniform Padding Lemma for Doubling Metrics

Lemma 15 (Local Uniform Padding Lemma). *Let (X, d) be a λ-doubling finite metric space. Let $0 < \Delta \leq \mathrm{diam}(X)$. Let $\hat{\delta} \in (\lambda^{-2}, 1/2]$, and let $\Gamma = 64$. There exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of (X, d) and a collection of uniform functions $\{\xi_P : X \to \{0, 1\} \mid P \in \mathcal{P}\}$ and $\{\eta_P : X \to (0, 1/\ln(1/\hat{\delta})] \mid P \in \mathcal{P}\}$ such that for any $\hat{\delta} \leq \delta \leq 1$, and $\eta^{(\delta)}$ defined by $\eta_P^{(\delta)}(x) = \eta_P(x)\ln(1/\delta)$, the probabilistic partition $\hat{\mathcal{P}}$ is a $(\eta^{(\delta)}, \delta)$-uniformly padded local probabilistic partition; and the following conditions hold for any $P \in \mathcal{P}$ and any $x \in X$:*

- $\eta_P(x) \geq 2^{-9}/(\ln \lambda)$.
- *If $\xi_P(x) = 1$ then:* $2^{-7}/\ln \rho(x, 4\Delta, \Gamma) \leq \eta_P(x) \leq 2^{-7}/\ln(1/\hat{\delta})$.
- *If $\xi_P(x) = 0$ then:* $\eta_P(x) = 2^{-7}/\ln(1/\hat{\delta})$ *and* $\bar{\rho}(x, 4\Delta, \Gamma) < 1/\hat{\delta}$.

F Embedding Distant Pairs

Theorem 2 follows from the following theorem on local scaling embedding for doubling metrics.

Recall that X satisfies a *weak growth rate* condition (cf. [2]): WGR(γ) for some constants $\gamma < 1$ if for every $x \in X$ and $r_1, r_2 > 0$, $|B(x, r_2)| \leq |B(x, r_1)|^{(r_2/r_1)^\gamma}$, and further assume $\gamma < 0.2$.

Theorem 4. *Given a metric space (X, d) satisfying* WGR(γ)*. For any $1 \leq p \leq \infty$, and $0 < \theta \leq 1$, 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 $\hat{k}$ s.t. $d(x, y) \leq \Delta_{\hat{k}}(x)$ is $O(\log^{1+\theta} \hat{k}/\theta)$.*

The lower bound on the distortion guaranteed by Theorem 4 is a monotonic function of the distance from any particular point. This is stated in the following corollary:

Corollary 16. *Given a metric space (X, d) satisfying* WGR(γ)*. For any $1 \leq p \leq \infty$, and $0 < \theta \leq 1$, 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 $\hat{k}$ s.t. $d(x, y) \geq \Delta_{\hat{k}}(x)$ then $\|f(x) - f(y)\|^p \geq \Delta_{\hat{k}}(x) \cdot \Omega(\theta/\log^{1+\theta} \hat{k})$.*

In the rest of this section we prove Theorem 4.

F.1 Proof of Theorem 4

The Embedding.

Let $\theta > 0$. Let $D = \lceil \frac{c \log \lambda}{\theta} \rceil$, where c is a constant to be determined later. We will define an embedding $f : X \to l_p^D$ with scaling distortion where the distortion for pairs $x, y \in X$ and $\hat{k}$ s.t. $d(x, y) \leq \Delta_{\hat{k}}(x)$ is $O(\log^{1+\theta} \hat{k}/\theta)$. We define f by defining for each $1 \leq t \leq D$, a function $f^{(t)} : X \to \mathbb{R}^+$ and let $f = D^{-1/p} \bigoplus_{1 \leq t \leq D} f^{(t)}$.

In what follows we define the functions $f^{(t)}$. Let $\Delta_0 = \operatorname{diam}(X)$, $I = \{i \in \mathbb{Z} \mid 1 \leq i \leq \log \Delta_0\}$. For $i \in Z$ let $\Delta_i = \Delta_0/4^i$. For each $0 < i \in I$ construct a Δ_i-bounded uniformly padded probabilistic partition $\hat{\mathcal{P}}_i$, as in Lemma 15 with parameter $\Gamma = 64$, $\hat{\delta} = 1/2$. Fix some $P_i \in \mathcal{P}_i$ for all $i \in I$. In the usual embedding via partitions scheme we obtain a lower bound for every pair $x, y \in X$ from only one "critical" scale (which is approximately $d(x, y)$). Here, we use the same idea, but since the cluster in the critical scale may contain too many points, we get contribution from two scales lower than the critical one, which is guaranteed to be small enough. For this reason we define a new function $\bar{\xi}$ as follows, for each $i \in I$, $P \in \mathcal{H}$:

$$\bar{\xi}_{P,i}(x) = \begin{cases} 1 & \rho(\upsilon(P_i(x)), 4\Delta_i, \Gamma^4) \geq 2 \\ \xi_{P,i}(x) & \text{otherwise} \end{cases}$$

where $\upsilon(C)$ is the center of cluster $C \in \mathcal{P}_i$. It can be seen that the function $\bar{\xi}$ is uniform as well.

Let $\varepsilon(\bar{k}) = \ln^{-\theta} \bar{k}$, $\delta(\bar{k}) = 1 - \varepsilon(\bar{k})$, and let $\zeta(\bar{k}) = \ln^{1+2\theta} \bar{k}$. We define the embedding by defining the coordinates for each $x \in X$. Define for $x \in X$, $0 < i \in I$, $\hat{k}_i(x) = |B(\upsilon(P_i(x)), (4\Gamma + 1)\Delta_i)|$. Define $\phi_i^{(t)} : X \to \mathbb{R}^+$, as:

$$\phi_i^{(t)}(x) = \frac{\bar{\xi}_{P_i}(x)}{\eta_{P_i}^{(\delta(\hat{k}_i(x)))}(x) \cdot \zeta(\hat{k}_i(x))}.$$

Let $\{\sigma_i^{(t)}(C) | C \in P_i, 0 < i \in I\}$ be i.i.d random variables uniformly distributed in $[0, 1]$.

For each $0 < i \in I$ we define a function $f_i^{(t)} : X \to \mathbb{R}^+$ and for $x \in X$, let $f^{(t)}(x) = \sum_{i \in I} f_i^{(t)}(x)$.

The embedding is defined as follows: for each $x \in X$:

- For each $0 < i \in I$, let $f_i^{(t)}(x) = \sigma_i^{(t)}(P_i^{(t)}(x)) \cdot g_i^{(t)}(x)$, where $g_i^{(t)} : X \to \mathbb{R}^+$ is defined as: $g_i^{(t)}(x) = \min\{\phi_i^{(t)}(x) \cdot d(x, X \setminus P_i^{(t)}(x)), \Delta_i\}$.

We have the following claims:

Claim 17. *For any $x, y \in X$ and $i \in I$ if $P_i(x) = P_i(y)$ then $\phi_i^{(t)}(x) = \phi_i^{(t)}(y)$.*

Claim 18. *There exists universal constant C_1 such that for any $x \in X$, $1 \le t \le D$ we have $\sum_{j \in I} \phi_j^{(t)}(x) \le C_1/\theta$.*

Proof. Let $b_i = \lfloor \ln |B(x, 4\Delta_i)| \rfloor$. As $d(\upsilon(P_i(x)), x) \le \Delta_i$ we have that $\log \hat{k}_i(x) = \log |B(\upsilon(P_i(x)), (4\Gamma + 1)\Delta_i)| \ge \log |B(x, 4\Gamma\Delta_i)| \ge b_{i-3}$.

$$\begin{aligned}
\sum_{j \in I} \phi_j(x) &= \sum_{j \in I : \bar{\xi}_j(x) = 1} \frac{\eta_j^{(\delta(\hat{k}_j(x)))}(x)^{-1}}{\zeta(\hat{k}_j(x))} \\
&\le \sum_{j \in I : \xi_j(x) = 1} \frac{2^7 \ln \rho(x, 4\Delta_j, \Gamma)}{\zeta(\hat{k}_j(x)) \cdot \ln(\frac{1}{1 - \varepsilon(\hat{k}_j(x))})} + \sum_{j \in I : \bar{\xi}_j(x) = 1, \xi_j(x) = 0} \frac{2^7}{\zeta(\hat{k}_j(x)) \cdot \ln(\frac{1}{1 - \varepsilon(\hat{k}_j(x))})} \\
&\le 2^8 \sum_{j \in I : \bar{\xi}_j(x) = 1} \frac{\rho(x, 4\Delta_j, \Gamma)}{\ln^{1+\theta} \hat{k}_j(x)} + 2^7 \sum_{j \in I} \frac{1}{\ln^{1+\theta} \hat{k}_j(x)} \le 2^9 \sum_{j \in I : \bar{\xi}_j(x) = 1} \frac{b_{j-3} - b_{j+2}}{(b_{j-3})^{1+\theta}} + 2^7 \sum_{h=1}^{\infty} \frac{1}{h^{1+\theta}} \\
&\le 2^9 \sum_{j \in I} \sum_{h = b_{j+2}}^{b_{j-3}} \frac{1}{h^{1+\theta}} + O(1/\theta) \le 2^{12} \sum_{h=1}^{\infty} \frac{1}{h^{1+\theta}} + O(1/\theta) = O(1/\theta).
\end{aligned}$$

□

Define $\bar{g}_i^{(t)} : X \times X \to \mathbb{R}^+$ as follows: $\bar{g}_i^{(t)}(x, y) = \min\{\phi_i^{(t)}(x) \cdot d(x, y), \Delta_i\}$. We have the following claim:

Claim 19. *For any $0 < i \in I$ and $x, y \in X$: $f_i^{(t)}(x) - f_i^{(t)}(y) \le \bar{g}_i^{(t)}(x, y)$.*

Lemma 20. *There exists a universal constant $C_1 > 0$ such that for any $x, y \in X$:*

$$\|f(x) - f(y)\|_p \le (C_1/\theta) \cdot d(x, y).$$

Proof. From Claim 19 and Claim 18 we get

$$\begin{aligned}
\sum_{0 < i \in I} (f_i^{(t)}(x) - f_i^{(t)}(y)) &\le \sum_{0 < i \in I} \bar{g}_i^{(t)}(x, y) \le \sum_{0 < i \in I} \phi_i^{(t)}(x) \cdot d(x, y) \\
&\le (C_1/\theta) \cdot d(x, y).
\end{aligned}$$

It follows that $|f^{(t)}(x) - f^{(t)}(y)| = |\sum_{0 < i \in I} (f_i^{(t)}(x) - f_i^{(t)}(y))| \le (C_1/\theta) \cdot d(x, y)$,and therefore

$$\|f(x) - f(y)\|_p^p = D^{-1} \sum_{1 \le t \le D} |f^{(t)}(x) - f^{(t)}(y)|^p \le (C_1/\theta)^p \, d(x, y)^p.$$

□

Lemma 21. *There exists a universal constant $C_2 > 0$ such that with constant probability for any $x, y \in X$ s.t. $d(x, y) \le \Delta_{\hat{k}}(x)$:*

$$\|f(x) - f(y)\|_p \ge C_2 \ln^{-1-3\theta} \hat{k} \cdot d(x, y).$$

Proof. We will prove that with constant probability for every $x, y \in X$ s.t. $d(x, y) \le \Delta_{\hat{k}}(x)$, there exists a set $T(x, y) \subseteq \{1, \ldots, D\}$ of size at least $D/2$ such that for any $t \in T(x, y)$:

$$|f^{(t)}(x) - f^{(t)}(y)| \quad \ge \quad 2^{-6} \ln^{-1-3\theta} \hat{k} \cdot d(x, y). \tag{45}$$

The theorem follows directly:

$$\begin{aligned}\|f(x) - f(y)\|_p^p &= D^{-1} \sum_{1 \le t \le D} |f^{(t)}(x) - f^{(t)}(y)|^p \ge D^{-1} \sum_{t \in T(x,y)} |f^{(t)}(x) - f^{(t)}(y)|^p \\ &\ge D^{-1} |T(x, y)| \cdot \left(2^{-6} \ln^{-1-3\theta} \hat{k} \cdot d(x, y)\right)^p \ge \frac{1}{2} \left(2^{-6} \ln^{-1-3\theta} \hat{k} \cdot d(x, y)\right)^p.\end{aligned}$$

□

The proof of (45) uses a set of nets of the space. For any $0 < i \in I$, and $1 \le k = 2^j \le n$, let N_i^k be a $\frac{\theta \cdot \varepsilon(k) \Delta_i}{16 C_1 \zeta(4k)}$-net of X. Let

$$M = \left\{(i, k, u, v) \mid i \in I,\ u, v \in N_i^k, 3\Delta_{i-4} \le d(u, v) \le 17\Delta_{i-4}, k \le \min\{\hat{k}_i(u), \hat{k}_i(v)\} < 2k\right\}.$$

Given an embedding f define a function $T : M \to 2^{[D]}$ such that for $t \in [D]$:

$$t \in T(i, k, u, v) \Leftrightarrow \left|f^{(t)}(u) - f^{(t)}(v)\right| \ge \frac{1}{2} \frac{\varepsilon(k)}{\zeta(4k)} \Delta_i.$$

For all $(i, k, u, v) \in M$, let $\mathcal{E}_{(i,k,u,v)}$ be the event $|T(i, k, u, v)| \ge D/2$.

Define the event $\mathcal{E} = \bigcap_{(i,k,u,v) \in M} \mathcal{E}_{(i,k,u,v)}$ that captures the case that all triplets in M have the desired property. The main technical lemma is that $\mathcal{E}$ occurs with non-zero probability:

Lemma 22. $\Pr[\mathcal{E}] > 0$.

Let us first show that if the event $\mathcal{E}$ took place, then the lower bound follows. Let $x, y \in X$, and let $0 < i \in I$ be such that $4\Delta_{i-4} \le d(x, y) < 16\Delta_{i-4}$.

Consider $u, v \in N_i$ satisfying $d(x, u) = d(x, N_i^k)$ and $d(y, v) = d(y, N_i^k)$, then $d(u, v) \le d(x, y) + d(u, x) + d(y, v) \le 16\Delta_{i-4} + 2\frac{\Delta_i}{C_1} \le 17\Delta_{i-4}$ and $d(u, v) \ge d(x, y) - d(x, u) - d(y, v) \ge 4\Delta_{i-4} - 2\frac{\Delta_i}{C_1} \ge 3\Delta_{i-4}$.

Let k be such that $k \le \min\{\hat{k}_i(u), \hat{k}_i(v)\} < 2k$. By the definition of M it follows that $(i, k, u, v) \in M$. It also holds that $k \le |B(v(P_i(u)), (4\Gamma + 1)\Delta_i)| \le |B(x, 4\Delta_{i-4})| \le |B(x, d(x, y))| \le \hat{k}$.

The next lemma shows that since x, y are very close to u, v respectively, then by the triangle inequality the embedding f of x, y cannot differ by much from that of u, v (respectively).

Lemma 23. *Let $x, y \in X$, let i be such that $4\Delta_{i-4} \le d(x, y) \le 16\Delta_{i-4}$, and $u, v \in N_i^k$ satisfying $d(x, u) = d(x, N_i^k)$ and $d(y, v) = d(y, N_i^k)$.*

Given $\mathcal{E}$, for any $t \in T(i, k, u, v)$:

$$\left|f^{(t)}(x) - f^{(t)}(y)\right| \ge \frac{1}{4} \frac{\varepsilon(\hat{k})}{\zeta(4\hat{k})} \Delta_i.$$

Proof. Since N_i^k is $\frac{\theta\cdot\varepsilon(k)\Delta_i}{16C_1\zeta(4k)}$-net, then $d(x,u) \le \frac{\theta\cdot\varepsilon(k)\Delta_i}{16C_1\zeta(4k)}$. By Lemma 20 $|f^{(t)}(x) - f^{(t)}(u)| \le (C_1/\theta)\cdot d(x,u) \le \frac{1}{16}\frac{\varepsilon(k)}{\zeta(4k)}\Delta_i$, and similarly $|f^{(t)}(y) - f^{(t)}(v)| \le \frac{1}{16}\frac{\varepsilon(k)}{\zeta(4k)}\Delta_i$. Then

$$\begin{aligned}
&|f^{(t)}(x) - f^{(t)}(y)| \\
&\quad = |f^{(t)}(x) - f^{(t)}(u) + f^{(t)}(u) - f^{(t)}(v) + f^{(t)}(v) - f^{(t)}(y)| \\
&\quad \ge |f^{(t)}(u) - f^{(t)}(v)| - |f^{(t)}(x) - f^{(t)}(u)| - |f^{(t)}(y) - f^{(t)}(v)| \\
&\quad \ge \frac{1}{2}\frac{\varepsilon(k)}{\zeta(4k)}\Delta_i - 2\frac{1}{16}\frac{\varepsilon(k)}{\zeta(4k)}\Delta_i \ge \frac{1}{4}\frac{\varepsilon(k)}{\zeta(4k)}\Delta_i \ge \frac{1}{4}\frac{\varepsilon(\hat{k})}{\zeta(\hat{k})}\Delta_i.
\end{aligned}$$

□

Let $\kappa(k) = \lceil \log\log(4k) \rceil$. Let $(i,k,u,v) \in M$ and $t \in [D]$. Define $\mathcal{F}_{(i,k,u,v,t)}$ be the event that:

$$|\sum_{0<j\le i+\kappa(k)} (f_j^{(t)}(u) - f_j^{(t)}(v))| \ge \frac{\varepsilon(k)}{\zeta(4k)}\Delta_i.$$

Let $\hat{\mathcal{E}}_{(i,k,u,v)}$ be the event that $|\{t|\mathcal{F}_{(i,k,u,v,t)}\}| \ge D/2$.

Claim 24. *For all $(i,k,u,v) \in M$, $\hat{\mathcal{E}}_{(i,k,u,v)}$ implies $\mathcal{E}_{(i,u,v)}$.*

Proof. Let $S = \{t|\mathcal{F}_{(i,k,u,v,t)}\}$. Then for $t \in S$: $\left|\sum_{0<j\le i+\kappa(k)} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \ge \frac{\varepsilon(k)}{\zeta(4k)}\Delta_i$, from Claim 19 it follows that $\left|\sum_{j>i+\kappa(k)} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \le \sum_{j>i+\kappa(k)} \Delta_j \le \frac{1}{2}\frac{\varepsilon(k)}{\zeta(4k)}\Delta_i$, which implies that $|f^{(t)}(u) - f^{(t)}(v)| = \left|\sum_{j\in I} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \ge \frac{1}{2}\frac{\varepsilon(k)}{\zeta(4k)}\Delta_i$. □

Lemma 25 (Lovasz Local Lemma - General Case). *Let $\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_n$ be events in some probability space. Let $G(V,E)$ be a directed graph on n vertices, each vertex corresponds to an event. Let $c : V \to [m]$ be a rating function of events, such that if $(\mathcal{A}_i, \mathcal{A}_j) \in E$ then $c(\mathcal{A}_i) \le c(\mathcal{A}_j)$. Assume that for all $i = 1,\ldots,n$ there exists $x_i \in [0,1)$ such that*

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j\in Q} \neg\mathcal{A}_j\right] \le x_i \prod_{j:(i,j)\in E} (1 - x_j),$$

for all $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin E \wedge c(\mathcal{A}_i) \ge c(\mathcal{A}_j)\}$, then

$$\Pr\left[\bigwedge_{i=1}^{n} \neg\mathcal{A}_i\right] > 0$$

Define a graph $G = (V,E)$, where $V = \{\hat{\mathcal{E}}_{(i,k,u,v)} \mid (i,k,u,v) \in M\}$, and the rating of a vertex $c(\hat{\mathcal{E}}_{(i,k,u,v)}) = i$. Let $x_{(i,k,u,v)} = \lambda^{-60\ln(\frac{2\ln k}{\theta})}$.

Define that $(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}) \in E$ iff $d(\{u,v\},\{u',v'\}) \le 4\Delta_i$, and $i' \le i + \kappa(k)$, and $\frac{1}{3} \le \frac{\log\log(4k')}{\log\log(4k)} \le 3$.

Claim 26. *Let $\hat{\mathcal{E}}_{(i,k,u,v)} \in V$, then the number of edges $(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}) \in E$ is at most $\lambda^{20\ln(\frac{2\ln k}{\theta})}$.*

Proof. We bound the number of pairs $u', v' \in N_{i'}^{k}$ such that $(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}) \in E$ for $i \le i' \le i+\kappa(k)$ and $\frac{1}{3} \le \frac{\log\log(4k')}{\log\log(4k)} \le 3$.

Assume w.l.o.g $d(u,u') \le 4\Delta_i$, since $d(u',v') \le 17\Delta_{i-4}$ we have $u', v' \in B = B(u, 40\Delta_{i-4})$. The number of pairs can be bounded by $|N_{i'}^{k} \cap B|^2$. There is at most point from the net $N_{i'}^{k'}$ in every ball of radius $r = \frac{\theta \cdot \varepsilon(k)^3}{16C_1(\zeta(4k))^3}\Delta_{i+\kappa(k)}$. Since (X,d) is λ-doubling, the ball B can be covered by $\lambda^{\log(40\Delta_{i-4}/r)}$ balls of radius r. Now, $\log(40\Delta_{i-4}/r) \le 8\ln\ln k + 18 + \log(1/\theta)$. It conclude that the number of possible pairs is bounded above by $\lambda^{20\ln(\frac{2\ln k}{\theta})}$. $\square$

The construction of the graph is based on the proposition that vertices that do not have an edge are either farther than $\approx \Delta_i$ apart or have different scales and hence do not change each other's bound on their success probability.

Lemma 27.

$$\Pr\left[\neg\hat{\mathcal{E}}_{(i,k,u,v)} \mid \bigwedge_{(i',k',u',v')\in Q} \hat{\mathcal{E}}_{(i',k',u',v')}\right] \le \lambda^{-61\ln(\frac{2\ln k}{\theta})},$$

for all $Q \subseteq \left\{(i',k',u',v') \mid i \ge i' \wedge \left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right) \notin E\right\}$.

Before we prove this lemma, let us see that it implies Lemma 22.

Apply Lemma 25 to the graph G we defined. Using Claim 26 we can bound the number of edges $\left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right) \in E$ is at most $d = \lambda^{20\ln(\frac{2\ln k}{\theta})}$. Recall that $x_{(i,k,u,v)} = \lambda^{-60\ln(\frac{2\ln k}{\theta})}$. Also it follows that $x_{(i',k',u',v')} = \lambda^{-60\ln(\frac{2\ln k'}{\theta})} \le \lambda^{-20\ln(\frac{2\ln k}{\theta})}$. Therefore the probability bound in Lemma 27 satisfies the first condition of Lemma 25 $\lambda^{-61\ln(\frac{2\ln k}{\theta})} \le \lambda^{-60\ln(\frac{2\ln k}{\theta})}(1-\lambda^{-20\ln(\frac{2\ln k}{\theta})})^d$. Therefore $\Pr[\mathcal{E}] = \Pr[\bigwedge_{(i,k,u,v)\in M} \hat{\mathcal{E}}_{(i,k,u,v)}] > 0$, which concludes the proof of Lemma 22.

F.1.1 Proof of Lemma 27

In what follows we use of the following simple technical claim.

Claim 28. *Let $A, B \in \mathbb{R}^+$ and let α, β be i.i.d random variables uniformly distributed in $[0,1]$. Then for any $C \in \mathbb{R}$ and $\varepsilon > 0$:*

$$\Pr[|C + A\alpha - B\beta| < \varepsilon \cdot \max\{A,B\}] < 2\varepsilon.$$

Proof. Assume wlog $A \ge B$. Consider the condition $|C + A\alpha - B\beta| < \varepsilon \cdot \max\{A,B\} = \varepsilon A$. If $C - B\beta \ge 0$ then it implies $\alpha < \varepsilon$. Otherwise $|\alpha - \frac{B\beta - C}{A}| < \varepsilon$. $\square$

Claim 29. *Let $(i,k,u,v) \in M$, $t \in [D]$, then* $\Pr\left[\mathcal{F}_{(i,k,u,v,t)}\right] \ge 1 - 3\varepsilon(k)$.

Proof. Set $\varepsilon = \varepsilon(k)$ and $\delta = 1 - \varepsilon$. Consider some $(i,k,u,v) \in M$. Then $3\Delta_{i-4} \le d(u,v) \le 17\Delta_{i-4}$. By Claim 10 we have that $\max\{\bar{\rho}(u, \Delta_{i-4}, \Gamma), \bar{\rho}(v, \Delta_{i-4}, \Gamma)\} \ge 2$. Assume w.l.o.g that $\bar{\rho}(u, \Delta_{i-4}, \Gamma) \ge 2$. It follows that also $\rho(\upsilon(P_i(u)), 4\Delta_i, \Gamma^4) \ge 2$ from Lemma 15 that $\bar{\xi}_{P^{(t)},i}(u) = 1$ which implies that $\phi_i^{(t)}(u) = \frac{\eta_{P^{(t)},i}^{(\delta(k_i(u)))}(u)^{-1}}{\zeta(\hat{k}_i(u))}$. As $k_i(u) \ge k$ we have that $\phi_i^{(t)}(u) \ge \frac{\eta_{P^{(t)},i}^{(\delta)}(u)^{-1}}{\zeta(\hat{k}_i(u))}$. As $\hat{\mathcal{H}}^{(t)}$ is $(\eta^{(\delta)}, 1-\varepsilon)$-padded we have the following bound

$$\Pr[B(u, \eta_{P^{(t)},i}^{(\delta)}(u)\Delta_i) \subseteq P_i^{(t)}(u)] \ge 1 - \varepsilon.$$

Therefore with probability at least $1-\varepsilon$:

$$g_i^{(t)}(u) \quad \geq \quad \phi_i^{(t)}(u) \cdot d(u, X \setminus P_i^{(t)}(u)) \geq \frac{\Delta_i}{\zeta(\hat{k}_i(u))}. \tag{46}$$

If $\hat{k}_i(u) \leq 4k$ then $g_i^{(t)}(u) \geq \frac{\Delta_i}{\zeta(4k)}$. Otherwise it must be the case that $\hat{k}_i(v) \leq 2k$. It follows that $\rho(\upsilon(P_i(u)), 4\Delta_i, \Gamma^4) \geq 2$ and thus $\xi_{P^{(t)},i}(v) = 1$, and hence by analogues argument to the one above we get that $g_i^{(t)}(v) \geq \frac{\Delta_i}{\zeta(4k)}$. We conclude that $\max\{g_i^{(t)}(u), g_i^{(t)}(v)\} \geq \frac{\Delta_i}{\zeta(4k)}$.

Let $\mathcal{A}$ denote the event that (46) occurs.

Recall that we are interested in the expression: $|\sum_{0<j\leq i+\kappa(k)}(f_j^{(t)}(u) - f_j^{(t)}(v))|$ and

$$f_i^{(t)}(u) - f_i^{(t)}(v) = \sigma_i^{(t)}(P_i^{(t)}(u)) \cdot g_i^{(t)}(u) - \sigma_i^{(t)}(P_i^{(t)}(v)) \cdot g_i^{(t)}(v).$$

Define $A = g_i^{(t)}(u)$, $B = g_i^{(t)}(u)$, $\alpha = \sigma_i^{(t)}(P_i^{(t)}(u))$, $\beta = \sigma_i^{(t)}(P_i^{(t)}(v))$ and $C = \sum_{i\neq j\leq i+\kappa(k)}(f_j^{(t)}(u) - f_j^{(t)}(v))$. Since $\operatorname{diam}(P_i^{(t)}(u)) \leq \Delta_i < d(u,v)$ we have that $P_i^{(t)}(v) \neq P_i^{(t)}(u)$. Thus α and β are independent random variables uniformly distributed in $[0,1]$, hence we can apply claim 28 and using (46) we have:

$$\Pr[|\sum_{0<j\leq i+\kappa(k)}(f_j^{(t)}(u) - f_j^{(t)}(v))| < \varepsilon\frac{\Delta_i}{\zeta(4k)}|\mathcal{A}] = \Pr[|C + A\alpha - B\beta|] < \varepsilon \cdot \max\{A,B\}|\mathcal{A}] < 2\varepsilon.$$

Therefore with probability at least $1-3\varepsilon(k)$:

$$|f^{(t)}(u) - f^{(t)}(v)| \geq \frac{\varepsilon(k)}{\zeta(4k)}\Delta_i. \tag{47}$$

□

Claim 30. *Let $(i,k,u,v) \in M$, $t \in [D]$, then*

$$\Pr\left[\neg\mathcal{F}_{(i,k,u,v,t)} \mid \bigwedge_{(i',k',u',v')\in Q} \hat{\mathcal{E}}_{(i',k',u',v')}\right] \leq 3\varepsilon(k),$$

for all $Q \subseteq \{(i',k',u',v') \in M \mid i \geq i' \wedge \left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right) \notin E\}$.

Proof. If $i' + \kappa(k') < i$, then event $\hat{\mathcal{E}}_{(i',k',u',v')}$ depend on events $\mathcal{F}_{(i',k',u',v',t')}$, and these events depend only on the choice of partition for scales at most i. Hence the padding probability for u,v in scale i and the choice of σ_i is independent of these events.

Otherwise, if $i - \kappa(k') \leq i' \leq i$, let $(i',k',u',v') \in M$ such that $\left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right) \notin E$. By the construction of G there are two cases. If $u',v' \notin B(u, 4\Delta_{i'})$ and $u',v' \notin B(v, 4\Delta_{i'})$ then u',v' are far from u,v and they fall into different clusters in every possible partition of scale i. From Lemma 15, the padding of u,v in scale i depends only on the local neighborhoods, $B(u,2\Delta_i) \cup B(v,2\Delta_i)$, which are disjoint from those of u',v'. The second case is that $d(\{u,v\},\{u',v'\}) \leq 4\Delta_i$. Recall that $k' \leq k_{i'}(u') = |B(\upsilon(P_{i'}(u')), (4\Gamma+1)\Delta_{i'})|$ and $k \geq \frac{1}{2}k_i(u) = \frac{1}{2}|B(\upsilon(P_i(u)), (4\Gamma+1)\Delta_i)|$. We have $d(\upsilon(P_{i'}(u'), \upsilon(P_i(u))) \leq d(\upsilon(P_{i'}(u'), u') + d(u',u) + d(u, \upsilon(P_i(u)) \leq \Delta_{i'} + 4\Delta_i + \Delta_i \leq 6\Delta_{i'}$ and therefore $k' \leq |B(v(P_i(u)), 2(4\Gamma+1)\Delta_{i'})|$. It follows from the WGR(γ) assumption that $k' \leq 2k^{4^{\gamma\kappa(k')}}$ implying $\log\log(4k') \leq \log\log(4k) + 2\gamma\kappa(k') \leq \log\log(4k) + 3\gamma\log\log(4k')$, and therefore $\frac{\log\log(4k')}{\log\log(4k)} \leq 1/(1-3\gamma) \leq 3$ assuming $\gamma < 0.2$. A similar bound can be derived in the reverse direction which yields a contradiction.

By Claim 29 there is probability $\geq 1 - 3\varepsilon(k)$ to succeed, no matter what happened in scales $\neq i$ or "far away" in scale i. □

We now prove Lemma 27. By Claim 30the probability a single coordinate t fails is at most $3\varepsilon(k)$. It follows from Chernoff bounds that the probability that more than $D/2$ coordinates fail is bounded above by:

$$\Pr\left[\neg\hat{\mathcal{E}}_{(i,k,u,v)} \mid \bigwedge_{(i',k',u',v')\in Q} \hat{\mathcal{E}}_{(i',k',u',v')}\right] \leq (6e(3\varepsilon(k)))^{D/2} \leq \lambda^{-\frac{c}{8}\ln(\frac{2\ln k}{\theta})}. \tag{48}$$

Setting c large enough implies that (48) is at most $\lambda^{-61\ln(\frac{2\ln k}{\theta})}$, as required.