# Embedding Metric Spaces in their Intrinsic Dimension

Ittai Abraham* Yair Bartal† Ofer Neiman‡

**Abstract**

A fundamental question of metric embedding is whether the *metric dimension* of a metric space is related to its *intrinsic dimension*. That is whether the dimension in which it can be embedded in some real normed space is implied by the intrinsic dimension which is reflected by the inherent geometry of the space. The existence of such an embedding was conjectured by Assouad and was later posed as an open problem by others. This question is tightly related to a major goal of many practical application fields: developing tools to represent intrinsically low dimensional metric data sets in a succinct manner. In this paper we give the first algorithmic technique with formal guarantees for finding faithful and low dimensional representations of data lying in high dimensional space. Our main theorem states that every finite metric space $X$ embeds into Euclidean space with dimension $O(\dim(X)/\epsilon)$ and distortion $O(\log^{1+\varepsilon} n)$, where $\dim(X)$ is the *doubling* dimension of the space $X$. Moreover, we show that $X$ can be embedded into dimension $\tilde{O}(\dim(X))$ with *constant* average distortion and $\ell_q$-distortion for any $q < \infty$. Our technique also provides a dimension-distortion tradeoff and an extension of Assouad's theorem, providing distance oracles that improve known construction when $\dim(X) = o(\log |X|)$.

## 1 Introduction

One of the main goals of the theory of metric embedding is to understand how well do finite metric spaces embed into normed spaces. Two measures are of particular importance, the *dimension* of the target normed space and the *distortion*, the extent to which the metrics disagree. Metric embedding has important applications in many practical fields. Finding compact and faithful representations of large and complex data sets is a major goal in fields like data mining, information retrieval and learning. Many real world measurements are of intrinsically low dimensional data that lie in extremely high dimensional space. For example, consider a camera capturing a simple object moving from one side to another. Given two images we would like to know their temporal relationship, how close they are to each other. Assuming each gray scale $m \times m$ image is captured by $O(m^2)$ brightness pixel measurements, this video stream lies in a $\mathbb{R}^{O(m^2)}$ space. However since the object's movement is intrinsically low dimensional its natural representation should be in a very low dimensional space. To the best of our knowledge we provide the first formal grantees for algorithmically finding faithful and low dimensional representations of data lying in high dimensional space.

A celebrated theorem of Bourgain [10] states that any $n$ point metric space $X$ embeds into Euclidean space with distortion $O(\log n)$. Together with the dimension reduction result of Johnson and Lindenstrauss [19], such an embedding requires only $O(\log n)$ dimensions. In [4] it was shown that such an embedding exists into $L_p$ for any $p$. In general, this result is tight, since metric spaces induced by expander graphs require $\Omega(\log n)$ distortion [23] and $\Omega(\log n)$ dimension as well [4].

These lower bounds on the dimension are associated with metrics that have high *intrinsic dimension*. The intrinsic dimension of a metric space $X$ is naturally measured by the doubling constant of the space: the minimum $\lambda$ such that every ball can be covered by $\lambda$ balls of half the radius. The *doubling dimension* of $X$ is defined as $\dim(X) = \log_2 \lambda$.

The doubling dimension of a metric space provides an inherent lower bound on the dimension in which a metric space can be embedded into a normed space with small distortion. Specifically, the dimension for embedding a metric space $X$ with distortion $\alpha$ is $\Omega(\dim(X)/\log \alpha)$, by a simple volume argument. The dimension in which a metric space can be embedded in some real normed space with low distortion is referred to as the *metric dimension* of the space. This paper addresses the fundamental problem of whether the intrinsic dimension of a metric space determines its metric dimension.

Variants of this question were posed by Assouad

*School of Engineering and Computer Science, Hebrew University, Israel. Email: ittaia@cs.huji.ac.il.

†School of Engineering and Computer Science, Hebrew University, Israel and Center of the Mathematics of Information, Caltech, CA, USA. Email: yair@cs.huji.ac.il. Supported in part by a grant from the Israeli Science Foundation (195/02).

‡School of Engineering and Computer Science, Hebrew University, Israel. Email: neiman@cs.huji.ac.il. Supported in part by a grant from the Israeli Science Foundation (195/02).

[6] as well as by Linial, London and Rabinovich [23], Gupta, Krauthgamer and Lee [17], and mentioned in [25]. Assouad [6] proved that for any $\gamma < 1$ there exists numbers $D = D(\lambda, \gamma)$ and $C = C(\lambda, \gamma)$ such that for any metric space $(X, d)$ with $\dim(X) = \log \lambda$, its "snowflake" version $(X, d^\gamma)$ can be embedded into a $D$-dimensional Euclidean space with distortion at most $C$. Assouad conjectured that similar results are possible for $\gamma = 1$, however this conjecture was disproved by Semmes [29]. Gupta, Krathgamer and Lee [17] initiated a comprehensive study of embeddings of doubling metrics. They analyzed the Euclidean distortion of the Laakso graph, which has constant doubling dimension, and show a lower bound of $\Omega(\sqrt{\log n})$ on the distortion. They also show a matching upper bound on the distortion of embedding doubling metrics, more generally the distortion is $O(\log^{1/p} n)$ for embedding into $L_p$. The best dependency on $\dim(X)$ of the distortion for embedding doubling metrics is given by Krauthgamer *et. al.* [21]. They show an embedding into $L_p$ with distortion $O((\dim(X))^{1-1/p}(\log n)^{1/p})$, and dimension $O(\log^2 n)$.

However, all known embeddings for general spaces (e.g. [10, 23, 4]), and even those that were tailored specifically for bounded doubling dimension spaces [17, 21] require $\Omega(\log n)$ dimensions. In this paper we give the first general low-distortion embeddings into a normed space whose dimension depends only on $\dim(X)$.

**1.1 Our results** Given two metric spaces $(X, d_X)$ and $(Y, d_Y)$ an *injective* mapping $f : X \to Y$ is called an *embedding* of $X$ into $Y$. An embedding is *non-contractive* if for any $u \neq v \in X$: $d_Y(f(u), f(v)) \geq d_X(u, v)$. For a non-contractive embedding let the distortion of the pair $\{u, v\}$ be $\mathrm{dist}_f(u, v) = \frac{d_Y(f(u),f(v))}{d_X(u,v)}$. The *distortion* of the embedding is defined as $\mathrm{dist}(f) = \max_{u,v\in X} \mathrm{dist}_f(u, v)$.

Our main theorem is the following:

THEOREM 1.1. *For every $n$-point metric space $X$ and $\theta \in (0, 1]$, there exists an embedding of $X$ into $L_p$ with distortion $O(\log^{1+\theta} n)$ and dimension $O(\frac{\dim(X)}{\theta})$.*

In this extended abstract we focus on the proof on this main theorem while the proofs of the rest of the theorems described in sequel appear in the full version of the paper.

The proof uses the framework for metric embedding defined in [4] and several new techniques: A new probabilistic partition that has local properties, a new technique for reducing the dimension by obtaining contributions over multiple scales, and a non-trivial use of the Lovász Local Lemma.

All the embeddings we present in this paper can be transformed into polynomial time algorithms. This requires a non-trivial use of an algorithmic version of the Local Lemma.

We extend our main theorem and provide embeddings with constant average distortion and constant $\ell_q$-distortion for any fixed $q < \infty$. Recall the definition from [4]:

DEFINITION 1.1. ($\ell_q$-DISTORTION) *For $1 \leq q \leq \infty$, define the* $\ell_q$-distortion *of an embedding $f$ as:*

$$\mathrm{dist}_q(f) = \|\mathrm{dist}_f(u, v)\|_q^{(\mathcal{U})} = \mathbb{E}[\mathrm{dist}_f(u, v)^q]^{1/q},$$

*where the expectation is taken according to the uniform distribution $\mathcal{U}$ over $\binom{X}{2}$. The classic notion of* distortion *is expressed by the $\ell_\infty$-distortion and the* average distortion *is expressed by the $\ell_1$-distortion.*

To bound the $\ell_q$-distortion we provide embeddings with scaling distortion. An embedding has scaling distortion $D(\cdot)$ if for every $\epsilon \in (0, 1]$ there is a $(1 - \epsilon)$ fraction of the pairs whose distortion is at most $D(\epsilon)$. The notion of scaling distortion [20, 2, 4] is tightly related to the notion of $\ell_q$-distortion.

We prove the following theorem effectively showing that Assouad's conjecture is in fact true in the following practical sense: data of low intrinsic dimension embeds into *constant* dimensional space with *constant* average distortion:

THEOREM 1.2. *For every finite metric space $X$, there exists an embedding $f$ from $X$ into $L_p$ with scaling distortion $O(\log^C(\frac{1}{\epsilon}))$ in dimension $O(\dim(X) \log \dim(X))$. In particular, for any fixed $q$, $\mathrm{dist}_q(f) = O(q^C)$ where $C$ is a universal constant.*

Obtaining bounds on the scaling distortion in a dimension which depends only on $\dim(X)$ is considerably more demanding.

We also study "snowflake" embeddings, embeddings of the metric $d^\alpha$ for $0 < \alpha < 1$ first considered by [6].

THEOREM 1.3. *For any $n$ point metric space $(X, d)$, any $p \geq 1$, any $0 < \alpha < 1$, any $\theta \leq 1$ and any $2^{C/\theta} \leq k \leq \dim(X)$ for a universal constant $C$, there exists an embedding of $(X, d^\alpha)$ into $L_p$ with distortion $O\left(\frac{k^{1+\theta} 2^{\dim(X)/(pk)}}{1-\alpha}\right)$ and dimension $D = O\left(\frac{2^{\dim(X)/k}\dim(X)}{\alpha\cdot\theta}\left(1 - \frac{\log(1-\alpha)}{\log k}\right)\right)$.*

Taking $\alpha = 1/2$, $k = \dim(X)$ yields for any $p \geq 1$ distortion $O(\dim(X)^{1+\theta})$ and dimension

$O((\dim(X))/\theta)$. The special case when $k = \dim(X)$ and $\theta = 1/\log \dim(X)$ was shown by [17][1]

A known tradeoff for *distance oracles* yields stretch $O(k)$ with memory $O(kn^{1/k})$ [33]. Taking $p = \infty$, $\alpha = 1 - 1/\log k$ and $\theta = 1/\log k$ in Theorem 1.3 yields a distance oracle with $\tilde{O}(k)$ stretch and $O(2^{\dim(X)/k}\dim(X)\log k)$ memory. This distance oracle improves known constructions when $\dim(X) = o(\log n)$.

In addition, we have a dimension/distortion tradeoff result, which improves the theorem of [21]:

Proposition 1.1. *Any metric space $(X, d)$ on $n$ points embeds into $L_p^D$ space with distortion $\tilde{O}\left(\log n \left(\frac{\dim(X)}{D}\right)^{1-1/p}\right)$ where $\dim(X)\log^2 \dim(X) \log\log n \le D \le \log n \log^2 \dim(X)$.*

Following our work, a similar tradeoff result was recently obtained by Chan, Gupta and Talwar [12].

**1.2 Applications** One of main goals of several fields: machine learning, artificial intelligence, data mining, information retrieval and pattern recognition is to develop tools to represent complex data sets in a succinct manner. In many important application areas, an intrinsically low dimensionally data set lies in a very high dimensional space. Finding compact and faithful representation of such data is an important problem. Belkin and Niyogi [9] use Laplacian methods to obtain heuristic approximations to low dimensional manifolds. In this paper we give the first formal proof that low dimensional data can indeed be embedded in its intrinsic dimension in a faithful manner.

The results of Kleinberg, Slivkins, and Wexler [20] and Abraham et. al. [3, 4] draw a striking connection between the theoretical notions of partial and scaling embeddings and the recent research in the network community on performing *passive distance estimation* (see e.g. [16, 26, 22, 14, 30, 13]) using *network embedding*. They suggest that the low intrinsic dimensionality of the Internet latency matrix may contribute to the ability to obtain practical network embeddings. While all previous theoretical results required $\Omega(\log n)$ dimensions, most practical schemes manage to embed internet latencies into *constant*-dimensional Euclidean space [14]. Our results provide theoretical support for this observed phenomena. We prove that such low distortion embeddings into constant-dimensional Euclidean space exist given that the underlying communication network has low intrinsic dimensionality.

Our results can be viewed in the context of *locality in distributed computing* [27] as providing *distance oracles* for doubling metrics. Theorem 1.2 implies distance oracle which can naturally be distributed into a *distance labeling* scheme where each label requires only $\tilde{O}(\dim(X))$ space and provides *constant* average stretch[2].

Another tradeoff suggested in Theorem 1.3 yields stretch $O(k \log k)$ labels of size $O(2^{\dim(X)/k}\dim(X)\log k)$ improving the tradeoff of [33] when $\dim(X) = o(\log n)$. Several previous results provide $1 + \epsilon$ stretch distance labels with $O((1/\epsilon)^{\dim(X)})$ space per label ([32, 31, 18]). However theses results require super polylogarithmic space per label when $\dim(X) = \omega(\log\log n)$. Our construction provides the first distance labeling scheme whose memory is proportional to the intrinsic dimensionality of the metric, even for metrics whose doubling dimension is not constant.

**1.3 Preliminaries** Consider a finite metric space $(X, d)$ and let $n = |X|$. 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 $\mathrm{diam}(X) = \max_{x,y\in X} d(x, y)$. For a point $x \in X$ and $r \ge 0$, the ball at radius $r$ around $x$ is defined as $B_X(x, r) = \{z \in X | d(x, z) \le r\}$. We omit the subscript $X$ when it is clear form the context. A metric space $(X, d)$ is **$\lambda$-doubling** if for any $x \in X$, $r \in \mathbb{R}$ the ball $B(x, r)$ can be covered by $\lambda$ balls of radius $r/2$. The **doubling dimension** of $X$ is $\dim(X) = \log \lambda$. An $r$-net for a metric space $(X, d)$ is a set $S \subseteq X$ such that for any $x, y \in S$, $d_X(x, y) \ge r$ and $X \subseteq \bigcup_{x\in S} B(x, r)$.

## 2 Embedding Doubling Metrics into Euclidean Space

In this section we show our main result.

**Theorem 1.1.** *For any $n$-point metric space $(X, d)$ with $\dim(X) = \log \lambda$ and any $c_1/\log\log n < \theta \le 1$, there exists an embedding $f : X \to L_p^D$ with distortion $O\left(\log^{1+3\theta} n\right)$ where $D = O\left(\frac{\log \lambda}{\theta}\right)$ and $c_1$ is a universal constant.*

### 2.1 Proof Overview

**Uniformly Padded Local Probabilistic Partitioning for Doubling Metrics.** Our embeddings are based on probabilistic partitions [7] which have become

[1] The proofs in [17] as well as the ones presented here require a local probabilistic partition for doubling metrics. The description in [17] was based on the partitions of [11, 15], however these cannot be made local. Instead one needs to base the construction of such partitions on the probabilistic partitions of [7, 4] as shown in this paper.

[2] We measure the space of a distance labeling scheme by the number of distance values needed to describe each label.

a standard tool in embeddings into normed spaces (e.g. [28, 21, 4]). A main tool in all our results is a new probabilistic partition lemma with "local" properties. It provides improved properties that depend on the doubling dimension and whose padding probability depends only on local events. Both these properties are crucial for our results. Many known probabilistic partitions do not posses the "local" property [15, 11, 4][3]. Our new probabilistic partition lemma is based on the uniform partition lemma of [4]. However some subtle changes to the algorithm and analysis are required to obtain the desired "local" property.

Given a metric space $(X, d)$, in Lemma 2.7 we construct a distribution on partitions and functions $\eta_P : X \to [0, 1]$. Each partition is composed of clusters with a bounded diameter $\Delta$. The distribution provides *uniform padding* with respect to $\eta_P$: All points $x$ that belong to the same cluster $C$ in $P$ have the *same* padding parameter $\eta_P(x)$. The lemma guarantees the probability that $B(x, \eta_P(x)\Delta)$ is fully contained in a cluster of $P$ is at least $1/2$ (notice that $\eta_P(x)$ is also a random variable). Moreover, this probability of $x$ being padded is *local*, it depends only on events that occur at distance at most $4\Delta$ away. The lemma ensures that the value of $\eta_P(x)$ is simultaneously lower bounded by two factors: one depends on the doubling dimension, $\eta_P(x) > \Omega(1/\log\lambda)$ and the other is the local growth rate, $\eta_P(x) > \Omega(|B(x, \Delta/64)|/|B(x, O(\log^\theta n\Delta))|)$. Note that the local growth rate has a non-standard factor of $O(\log^\theta n)$. Indeed, our embedding is parameterized by $\theta$. Finally we note that the lemma provides the desired properties (denoted by $\xi_P(x) = 1$) only when the local growth rate $\approx |B(x, O(\log^\theta n\Delta))|/|B(x, \Delta/64)|$ is larger than a constant.

**The Embedding.** The embedding follows the general approach of [4]. While the basic scheme of [4] uses $O(\log n)$ dimensions, we use only $D \approx (\log\lambda)/\theta$ dimensions. For each dimension, probabilistic partitions are chosen for each scale $\Delta_i$ and each cluster of each partition is randomly colored 0 or 1. The value of $x$ at dimension $t$ is the sum $\sum f_i^{(t)}(x)$ over all scales $i$. At each scale $i$ the value $f_i^{(t)}(x)$ is 0 if the cluster $P(x)$ containing $x$ at scale $i$ is colored 0 or there is no growth rate ($\xi_P(x) = 0$); and otherwise it is the distance between $x$ and $X \setminus P(x)$ times the inverse of the uniform padding parameter $\eta_P(x)$ obtained from the probabilistic partition lemma. Also the value of $f_i^{(t)}(x)$ is truncated by $\Delta_i$. Lemma 2.10 shows that the embedding does not expand distances too much. The proof is similar to the one in [4].

[3] We note that [7, 24] do have this property, but do not obtain the same padding parameters needed in this paper.

**The Lower Bound.** In all previous proofs of partitions based embeddings, the lower bound is constructed by examining a *single* scale for each pair of points. Here, we defer from this approach and obtain a lower bound on distortion that follows from examining *several* scales. Our result is obtained by a subtle balancing between distortion and dimension obtained by this examination of several scales.

The high level idea is to apply Lovász Local Lemma to prove a small dimension suffices to obey all pairwise distortions with some constant probability. We start by choosing a set of $r$-nets for each scale. Since the doubling dimension bounds the density of $r$-nets, the Local Lemma can be applied and the number of dependencies becomes a function of $\dim(X)$ instead of $\log|X|$. For some of the scales $i$, $N_i$ is a $\approx \frac{\Delta_i}{\log^3 n}$-net of $X$.

Consider net points $u, v \in N_j$ such that $d(u, v) \approx \Delta_j$. Using a standard technique of examining the $j$-th term we obtain a basic result of $O(\log n)$ distortion and $O(\log\lambda \log\log n)$ dimensions. Obtaining optimal dimension, independent of $n$, requires to examine several scales. We iteratively examine scales from $j$ to $\approx j + \theta\log\log n$. For each dimension $t$ there is a constant probability to succeed in any one of the $\theta\log\log n$ scales. The success probability of each scale $\ell$ is constant no matter what are the values of the previous scales $j, \ldots, \ell-1$. Upon the first success in a given dimension, the contribution from this scale is at least $\Omega(\Delta_j/\log^\theta n)$. Obtaining this property requires to define the local growth rate as $\approx |B(x, O(\log^\theta n\Delta_j))|/|B(x, \Delta_j/64)|$ so as to ensure that at least one of the points in the pair will obtain the desired $\xi = 1$ property from the probabilistic partition lemma for each of the examined scales. Therefore with very low probability all scales will fail for a given dimension. Hence two properties hold: the nets are sparse enough and the probability for such a failure of all scales is small enough. Lemma 2.11 uses these two properties and applies a variant of the Local Lemma to ensure that there exists an event in which all required net points have the desired lower bound, using only $D$ dimensions. As our events associated with a particular scale may depend on other events associated with higher scales, we require here an application of a special variant of the LLL which allows such directional dependencies. Finally, since all pairs of net points have small distortion, Lemma 2.12 shows that all pairs have the desired distortion as well. This follows using the triangle inequality, observing that the distance of a point from its nearest net point is at most $O(\frac{\Delta_i}{\log^3 n})$, while the distortion of the embedding is at most $O(\log^{1+\theta} n)$.

**Algorithmic version.** In the full version we prove that it is possible to derandomize our LLL variant using the approach of [8, 5] tailored to the specifics of our problem.

### 2.2 A Local Uniform Padding Lemma for Doubling Metrics

Definition 2.1. *The local growth rate of $x \in X$ at radius $r > 0$ for given scales $\gamma_1, \gamma_2 > 0$ is defined as*

$$\rho(x, r, \gamma_1, \gamma_2) = |B(x, r\gamma_1)|/|B(x, r\gamma_2)|.$$

*Given a subspace $Z \subseteq X$, the minimum local growth rate of $Z$ at radius $r > 0$ and scales $\gamma_1, \gamma_2 > 0$ is defined as $\rho(Z, r, \gamma_1, \gamma_2) = \min_{x \in Z} \rho(x, r, \gamma_1, \gamma_2)$. The minimum local growth rate of $x \in X$ at radius $r > 0$ and scales $\gamma_1, \gamma_2 > 0$ is defined as $\bar{\rho}(x, r, \gamma_1, \gamma_2) = \rho(B(x, r), r, \gamma_1, \gamma_2)$.*

The following claim was shown in [4]

Claim 2.2. *Let $x, y \in X$, let $\gamma_1, \gamma_2 > 0$ and let $r$ be such that for any $w \in X$, $2(1 + \gamma_2)r < d(x, y) \leq (\gamma_1 - \gamma_2 - 2)r$, then*

$$\max\{\bar{\rho}(x, r, \gamma_1, \gamma_2), \bar{\rho}(y, r, \gamma_1, \gamma_2)\} \geq 2.$$

Definition 2.3. (Partition) *A partition $P$ of a finite metric space $(X, d)$ is a collection of disjoint sets $\mathcal{C}(P) = \{C_1, C_2, \ldots, C_t\}$ such that $X = \cup_j C_j$. The sets $C_j \subseteq X$ are called clusters. For $x \in X$ denote by $P(x)$ the cluster containing $x$. Given $\Delta > 0$, a partition is* $\Delta$-bounded *if for all $1 \leq j \leq t$, $\mathrm{diam}(C_j) \leq \Delta$. For $Z \subseteq X$ we denote by $P_{|Z}$ the restriction of $P$ to points in $Z$.*

Definition 2.4. (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 $\Delta$-bounded if each $P \in \mathcal{P}$ is $\Delta$-bounded.*

*Let $supp(\hat{\mathcal{P}}) \subseteq \mathcal{P}$ be the set of partitions with non-zero probability under $\hat{\mathcal{P}}$.*

Definition 2.5. (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 2.6. (Uniformly Padded Local PP) *Given $\Delta > 0$ and $0 < \delta \leq 1$, let $\hat{\mathcal{P}}$ be a $\Delta$-bounded probabilistic partition of $(X, d)$. Given collection of functions $\eta = \{\eta_P : X \to [0, 1] | P \in \mathcal{P}\}$ such that $\eta_P$ is a uniform function with respect to $P$. We say that $\hat{\mathcal{P}}$ is a $(\eta, \delta)$-*uniformly padded local probabilistic partition *if the event $B(x, \eta_P(x)\Delta) \subseteq P(x)$ occurs with probability at least $\delta$ 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 $C$,*

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

Lemma 2.7. (Local Uniform Padding Lemma) *Let $(X, d)$ be a $\lambda$-doubling finite metric space. Let $0 < \Delta \leq \mathrm{diam}(X)$. Let $\hat{\delta} \in (\lambda^{-2}, 1/2]$, and let $\gamma_1 : X \to \mathbb{R}_+$ such that $\gamma_1(x) \geq 1$ for all $x \in X$ and let $\gamma_2 \leq 1/64$. There exists a $\Delta$-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: $2^{-7}/\ln \rho(x, 8\Delta, \gamma_1(x), \gamma_2) \leq \eta_P(x) \leq 2^{-7}/\ln(1/\hat{\delta})$.*
- *If $\xi_P(x) = 0$ then: $\eta_P(x) = 2^{-7}/\ln(1/\hat{\delta})$ and $\bar{\rho}(x, 8\Delta, \gamma_1(\cdot), \gamma_2) < 1/\hat{\delta}$.*

The proof of Lemma 2.7 is deferred to the full version.

**2.3 The Embedding** Let $\Delta_0 = \mathrm{diam}(X)$, $I = \{i \in \mathbb{Z} \mid 1 \leq i \leq (\log \Delta_0 + \theta \log\log n)/3\}$. For $i \in I$ let $\Delta_i = \Delta_0/8^i$. Let $c$ be a constant to be determined later and set $D = \frac{c \log \lambda}{\theta}$.

In the analysis we will use Claim 2.2. To guarantee that there will be growth rate greater than 2 in scales $\log\log^\theta n$ lower than some "critical scale", we define $\gamma_1 = 2^8 \log^\theta n$, $\gamma_2 = 1/64$. Notice that this will inflict loss of $O(\log\log^\theta n)$ term in the distortion, but it is consumed by the $O(\log^\theta n)$ factor.

We shall define the embedding $f$ by defining for each $1 \leq t \leq D$, a function $f^{(t)} : X \to \mathbb{R}^+$ and let $f = D^{-1/p} \bigoplus_{1 \leq t \leq D} f^{(t)}$. Fix $t$, $1 \leq t \leq D$. In what follows we define $f^{(t)}$. For each $0 < i \in I$ construct a $\Delta_i$-bounded $(\eta_i, 1/2)$-padded probabilistic partition $\hat{\mathcal{P}}_i$, as in Lemma 2.7 with parameters $\gamma_1, \gamma_2$ and $\hat{\delta} = 1/2$. Fix some $P_i \in \mathcal{P}_i$ for all $i \in I$.

Define for $x \in X$, $0 < i \in I$, $\phi_i^{(t)} : X \to \mathbb{R}^+$, by $\phi_i^{(t)}(x) = \xi_{P,i}(x)\eta_{P,i}(x)^{-1}$. Lemma 2.7 ensures that $\xi_i$ and $\eta_i$ are uniform functions with respect to $\mathcal{P}_i$. Hence $\phi_i$ is also uniform with respect to $\mathcal{P}_i$.

Claim 2.8. *For any $x \in X$, $1 \leq t \leq D$ we have $\sum_{j \in I} \phi_j^{(t)}(x) \leq 2^9 \log^{1+\theta}(n)$.*

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

$P_i, 0 < i \in I\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. The embedding is defined as follows: for each $x \in X$:

- For each $0 < i \in I$, let $f_i^{(t)}(x) = \sigma_i^{(t)}(P_i(x)) \cdot \min\{\phi_i^{(t)}(x) \cdot d(x, X \setminus P_i(x)), \Delta_i\}$.

CLAIM 2.9. *For any $0 < i \in I$ and $x, y \in X$: $f_i^{(t)}(x) - f_i^{(t)}(y) \le \min\{\phi_i^{(t)}(x) \cdot d(x,y), \Delta_i\}$.*

LEMMA 2.10. *For any $(x,y) \in X$: $|f^{(t)}(x) - f^{(t)}(y)| \le 2^9 \log^{1+\theta}(n) \cdot d(x,y)$.*

The upper bound is similar to [4], the proofs for this section are deferred to the full version.

**2.4 Lower Bound Analysis** The lower bound analysis uses a set of nets. First we define a set of scales in which we hope to succeed with high probability. Let $k = \lceil (\theta/3) \log\log n \rceil$, $K = \{i \in I : k | i\}$. For any $0 < i \in K$ let $N_i$ be a $\frac{\Delta_i}{2^{13} \log^3 n}$-net of $X$. Let $M = \{(i,u,v) \mid i \in K,\ u,v \in N_i, 7\Delta_{i-1} \le d(u,v) \le 65\Delta_{i-k-1}\}$. Given an embedding $f$ define a function $T : M \to 2^D$ such that for $t \in [D]$ :

$$t \in T(i,u,v) \Leftrightarrow \left| f^{(t)}(u) - f^{(t)}(v) \right| \ge \frac{\Delta_i}{4 \log^\theta n}.$$

For all $(i,u,v) \in M$, let $\mathcal{E}_{(i,u,v)}$ be the event $|T(i,u,v)| \ge D/2$. Roughly speaking, $\mathcal{E}_{(i,u,v)}$ is the event that $u, v$, which belong to the net $N_i$ and whose distance is about $\approx \Delta_i, \ldots, \Delta_i \log n$ have the property that for at least half the coordinates of the embedding, the distance between $f^{(t)}(u)$ and $f^{(t)}(v)$ is at least $\frac{\Delta_i}{4 \log^\theta n}$.

Define the event $\mathcal{E} = \bigcap_{(i,u,v) \in M} \mathcal{E}_{(i,u,v)}$ 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 2.11. $\Pr[\mathcal{E}] > 0$.

We defer the proof for later, and now show that if the event $\mathcal{E}$ took place, then we can show the lower bound. Let $x, y \in X$, and let $0 < i' \in I$ be such that $8\Delta_{i'-1} \le d(x,y) < 64\Delta_{i'-1}$. Let $i \in K$ be the minimal such that $i \ge i'$, note that $\Delta_i \ge \frac{\Delta_{i'}}{\log^\theta n}$. Consider $u, v \in N_i$ satisfying $d(x,u) = d(x, N_i)$ and $d(y,v) = d(y, N_i)$, as $d(u,v) \le d(u,x) + d(x,y) + d(y,v) \le 64\Delta_{i'-1} + \Delta_i \le 65\Delta_{i-k-1}$, by the definition of $M$ follows that $(i,u,v) \in M$. 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 2.12. *Let $x, y \in X$, let $i'$ such that $8\Delta_{i'-1} \le d(x,y) \le 64\Delta_{i'-1}$, let $i \in K$ be the minimal such that $i \ge i'$, let $u, v \in N_i$ satisfying $d(x,u) = d(x,N_i)$ and $d(y,v) = d(y,N_i)$.*

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

$$\left| f^{(t)}(x) - f^{(t)}(y) \right| \ge \frac{\Delta_i}{8 \log^\theta n}$$

*Proof.* Since $N_i$ is $\frac{\Delta_i}{2^{13}\log^3 n}$-net, then $d(x,u) \le \frac{\Delta_i}{2^{13}\log^3 n}$. By Lemma 2.10 $|f^{(t)}(x) - f^{(t)}(u)| \le 2^9 \log^{1+\theta}(n) \cdot d(x,u) \le \frac{\Delta_i}{2^4 \log n}$, and similarly $|f^{(t)}(y) - f^{(t)}(v)| \le \frac{\Delta_i}{2^4 \log n}$. Then

$$\begin{aligned}
&|f^{(t)}(x) - f^{(t)}(y)| \\
&= |f^{(t)}(x) - f^{(t)}(u) + f^{(t)}(u) - f^{(t)}(v) + f^{(t)}(v) - f^{(t)}(y)| \\
&\ge |f^{(t)}(u) - f^{(t)}(v)| - |f^{(t)}(x) - f^{(t)}(u)| - |f^{(t)}(y) - f^{(t)}(v)| \\
&\ge \frac{\Delta_i}{4 \log^\theta n} - \frac{2\Delta_i}{16 \log n} \ge \frac{\Delta_i}{8 \log^\theta n}.
\end{aligned}$$

This Lemma and Lemma 2.11 implies the following:

LEMMA 2.13. *There exists a universal constant $C_2 > 0$ and an embedding $f$ such that for any $x, y \in X$*

$$\|f(x) - f(y)\|_p \ge C_2 \frac{d(x,y)}{\log^{2\theta} n}.$$

*Proof.* Let $f$ be an embedding such that event $\mathcal{E}$ took place. Let $i' \in I$ such that $\Delta_{i'-2} \le d(x,y) < \Delta_{i'-3}$, $i \in K$ the minimal such that $i \ge i'$ and $u, v$ be the nearest points to $x, y$ respectively in the net $N_i$. Noticing that $\Delta_i \ge \frac{d(x,y)}{512 \log^\theta n}$ and that $|T(i,u,v)| \ge D/2$ we get from Lemma 2.12 that

$$\begin{aligned}
\|f(x) - f(y)\|_p^p &= D^{-1} \sum_{t \in [D]} |f^{(t)}(x) - f^{(t)}(y)|^p \\
&\ge D^{-1} \sum_{t \in T(i,u,v)} \left( \frac{\Delta_i}{8 \log^\theta n} \right)^p \\
&\ge D^{-1} |T(i,u,v)| \left( \frac{d(x,y)}{2^{12} \log^{2\theta} n} \right)^p \\
&\ge \left( \frac{d(x,y)}{2^{13} \log^{2\theta} n} \right)^p
\end{aligned}$$

**2.4.1 Proof of Lemma 2.11** Recall that each coordinate $t \in \{1, 2, \ldots, D\}$ of an embedding of a point $x$ is defined as a sum of terms over all possible scales $\sum_j f_j^{(t)}(x)$. Also recall that each triplet $(i,u,v) \in M$ represents two net points $u, v$ for which we would like to the embedding to have $|f^{(t)}(u) - f^{(t)}(v)| \ge \frac{\Delta_i}{4 \log^\theta n}$ for at least $D/2$ (half) of the coordinates. For each coordinate

$t$, we will try to obtain $|f^{(t)}(u)-f^{(t)}(v)| \geq \frac{\Delta_i}{4\log^\theta n}$ by iteratively checking scales $\ell \in \{i,\dots,i+k-1\}$ (recall that $k = \lceil(\theta/3)\log\log n\rceil$). For each $\ell$ in increasing order we will check if $|\sum_{j\leq\ell} f_j^{(t)}(u) - \sum_{j\leq\ell} f_j^{(t)}(v)| \geq \Delta_\ell/2$. Since scales decrease by a factor of 8, once such a successful event occurs, no matter what the remaining scales are, we will have $|f^{(t)}(u)-f^{(t)}(v)| \geq \frac{\Delta_\ell}{4} \geq \frac{\Delta_i}{4\log^\theta n}$ as required.

Formally we define for every $(i,u,v) \in M$, $i \leq \ell < i+k$ and $t \in [D]$ the event $\mathcal{F}_{(i,u,v,t,\ell)}$ as

$$\Big|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\Big| \leq \frac{\Delta_\ell}{2} \Rightarrow \big|f_\ell^{(t)}(u) - f_\ell^{(t)}(v)\big| \geq \Delta_\ell$$

$$\Big|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\Big| > \frac{\Delta_\ell}{2} \Rightarrow f_\ell^{(t)}(u) = f_\ell^{(t)}(v) = 0$$

In words, event $\mathcal{F}_{(i,u,v,t,\ell)}$ means that the pair $u,v$ is successful, it has a difference of at least $\Delta_\ell/2$ from the sum of their first $\ell$ terms of their $t$-th coordinate. Now define event $\hat{\mathcal{E}}_{(i,u,v)}$ as

$$\exists S \subseteq [D], |S| \geq D/2, \forall t \in S, \exists i \leq \ell < i+k \ : \ \mathcal{F}_{(i,u,v,t,\ell)}$$

In words $\hat{\mathcal{E}}_{(i,u,v)}$ is the event that at least half of the coordinates are successful. Note that each successful coordinate may have a different $\ell$ value.

Claim 2.14. *For all $(i,u,v) \in M$, $\hat{\mathcal{E}}_{(i,u,v)}$ implies $\mathcal{E}_{(i,u,v)}$.*

*Proof.* Let $S \subseteq [D]$ from the definition of $\hat{\mathcal{E}}_{(i,u,v)}$. For any $t \in S$, let $i \leq \ell(t) < i+k$ be such that $\mathcal{F}_{(i,u,v,t,\ell(t))}$ holds. Then for such $t \in S$: $\big|\sum_{j\leq\ell(t)} f_j^{(t)}(u) - f_j^{(t)}(v)\big| \geq \frac{\Delta_{\ell(t)}}{2}$, from Claim 2.9 it follows that $\big|\sum_{j>\ell(t)} f_j^{(t)}(u) - f_j^{(t)}(v)\big| \leq \sum_{j>\ell(t)} \Delta_j \leq \frac{\Delta_{\ell(t)}}{4}$, which implies that $\big|\sum_{j\in I} f_j^{(t)}(u) - f_j^{(t)}(v)\big| \geq \frac{\Delta_{\ell(t)}}{4} \geq \frac{\Delta_i}{4\log^\theta n}$ as required.

We shall make use the following variation of the Local Lemma due to Erdős and Lovász. See the full version for a proof.

Lemma 2.15. (Local Lemma) *Let $\mathcal{A}_1, \mathcal{A}_2, \dots \mathcal{A}_n$ be events in some probability space. Let $G(V,E)$ be a directed graph on $n$ vertices with out-degree at most $d$, each vertex corresponding 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 $c(\mathcal{A}_i) \leq c(\mathcal{A}_j)$. Assume that for any $i = 1,\dots,n$*

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j\in Q} \neg\mathcal{A}_j\right] \leq p$$

*for all $Q \subseteq \{j : (\mathcal{A}_i,\mathcal{A}_j) \notin E \wedge c(\mathcal{A}_i) \geq c(\mathcal{A}_j\}$. If $ep(d+1) \leq 1$, then*

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

Define a graph $G = (V,E)$, where $V = \{\hat{\mathcal{E}}_{(i,u,v)} \mid (i,u,v) \in M\}$, and the rating of a vertex $c(\hat{\mathcal{E}}_{(i,u,v)}) = i$. Define that $(\hat{\mathcal{E}}_{(i,u,v)}, \hat{\mathcal{E}}_{(i',u',v')}) \in E$ iff $d(\{u,v\},\{u',v'\}) \leq 4\Delta_i$, and $i = i'$.

Claim 2.16. *The out-degree of $G$ is bounded by $\lambda^{10\log\log n}$*

*Proof.* Fix $\hat{\mathcal{E}}_{(i,u,v)} \in V$, we bound the number of pairs pairs $u',v' \in N_i$ such that $(\hat{\mathcal{E}}_{(i,u,v)}, \hat{\mathcal{E}}_{(i,u',v')}) \in E$.

Assume w.l.o.g $d(u,u') \leq 4\Delta_i$, since $d(u,v), d(u',v') \leq 65\Delta_{i-k-1}$ follows $u,v,u',v' \in B = B(u,\Delta_{i-k-4})$. The number of pairs can be bounded by $|N_i \cap B|^2$. Since $(X,d)$ is $\lambda$-doubling, the ball $B$ can be covered by $\lambda^{26+4\log\log n}$ balls of radius $\frac{\Delta_i}{2^{14}\log^3 n}$, each of these contains at most one point in the net $N_i$. Assuming $\log\log n > 26$ it follows that $|N_i \cap B|^2 \leq \lambda^{10\log\log n}$.

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. Notice that events $\hat{\mathcal{E}}_{(i,u,v)}$ do not depend on the choice of partitions for scales greater than $i+k$.

Lemma 2.17.

$$\Pr\left[\neg\hat{\mathcal{E}}_{(i,u,v)} \mid \bigwedge_{(i',u',v')\in Q} \hat{\mathcal{E}}_{(i',u',v')}\right] \leq \lambda^{-11\log\log n},$$

*for all* $Q \subseteq \Big\{(i',u',v') \mid i \geq i' \wedge \Big(\hat{\mathcal{E}}_{(i,u,v)}, \hat{\mathcal{E}}_{(i',u',v')}\Big) \notin E\Big\}$.

Before we prove this lemma, let us see that it implies Lemma 2.11.

Apply Lemma 2.15 to the graph $G$ we defined, by Claim 2.16 let $d = \lambda^{10\log\log n}$ and by Lemma 2.17 we can let $p = \lambda^{-11\log\log n}$ satisfying the first condition of Lemma 2.15. It is easy to see that the second condition also holds (since $\lambda \geq 2$ and assuming $\log\log n \geq 2$). Therefore $\Pr[\bigwedge_{(i,u,v)\in M} \hat{\mathcal{E}}_{(i,u,v)}] > 0$, and Claim 2.14 concludes the proof of Lemma 2.11 by $\Pr[\mathcal{E}] = \Pr[\bigwedge_{(i,u,v)\in M} \mathcal{E}_{(i,u,v)}] > 0$.

#### 2.4.2 Proof of Lemma 2.17

In order to prove this lemma, we use a variation of a claim shown in [4].

CLAIM 2.18. *Let* $(i,u,v) \in M$, $t \in [D]$ *and* $i \le \ell < i+k$ *then* $\Pr\left[\mathcal{F}_{(i,u,v,t,\ell)}\right] \ge 1/8$.

*Proof.* Let $i \le \ell < i+k$. Since $\gamma_1 \Delta_{\ell-1} \ge 2^8 \log^\theta n \cdot \Delta_{\ell-1} \ge 2^{11}\Delta_i \ge 2d(u,v)$, by Claim 2.2 we have $\max\{\bar\rho(u,\Delta_{\ell-1},\gamma_1,\gamma_2), \bar\rho(v,\Delta_{\ell-1},\gamma_1,\gamma_2)\} \ge 2$. Assume w.l.o.g that $\bar\rho(u,\Delta_{\ell-1},\gamma_1,\gamma_2) \ge 2$. It follows from Lemma 2.7 that $\xi_{P,\ell}(u) = 1$ which implies that $\phi_\ell^{(t)}(u) = \eta_{P,\ell}(u)^{-1}$. We now consider the two cases in $\mathcal{F}_{(i,u,v,t,\ell)}$:

If it is the case that $\left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \le \frac{\Delta_\ell}{2}$ then it is enough for the following three events occur: $B(u,\eta_{P,\ell}(u)\Delta_\ell) \subseteq P_\ell(u)$, $\sigma_\ell^{(t)}(P_\ell(u)) = 1$ and $\sigma_\ell^{(t)}(P_\ell(v)) = 0$. Each event happens independently with probability $\ge 1/2$, the first since $P_\ell$ is $(\eta_\ell, 1/2)$-padded and the other two follow from $d(u,v) \ge 3\Delta_\ell \Rightarrow P_\ell(u) \ne P_\ell(v)$. If all these events occur then $|f_\ell^{(t)}(u) - f_\ell^{(t)}(v)| \ge \min\{\eta_{P,\ell}^{-1}(u) \cdot d(u, X \setminus P_\ell(u)), \Delta_\ell\} \ge \Delta_\ell$.

Similarly, if it is the case that $\left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| > \frac{\Delta_\ell}{2}$ then it is enough that $\sigma_\ell^{(t)}(P_\ell(u)) = \sigma_\ell^{(t)}(P_\ell(v)) = 0$. This occurs with probability $\ge 1/4$. So event $\mathcal{F}_{(i,u,v,t,\ell)}$ occurs with probability $\ge 1/8$.

CLAIM 2.19. *Let* $(i,u,v) \in M$, $t \in [D]$ *and* $i \le \ell < i+k$. *Then*

$$\Pr\left[\neg\mathcal{F}_{(i,u,v,t,\ell)} \mid \bigwedge_{(i',u',v')\in Q} \hat{\mathcal{E}}_{(i',u',v')}\right] \le 7/8,$$

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

*Proof.* If $i' < i$, then event $\hat{\mathcal{E}}_{(i',u',v')}$ depend on events $\mathcal{F}_{(i',u',v',t',\ell')}$, where by definition $\ell' < i'+k \le i$ (recall that $K$ contains only integers that divide by $k$), and these events depend only on the choice of partition for scales at most $\ell'$. Hence the padding probability for $u,v$ in scale $\ell$ and the choice of $\sigma_\ell$ is independent of these events.

Otherwise, if $i' = i$, let $(i,u',v') \in M$ such that $\left(\hat{\mathcal{E}}_{(i,u,v)}, \hat{\mathcal{E}}_{(i,u',v')}\right) \notin E$. We know by the construction of $G$ that $u',v' \notin B(u,4\Delta_i)$ and $u',v' \notin B(v,4\Delta_i)$. Hence $u',v'$ are far from $u,v$ and they fall into different clusters in every possible partition of scale $\ell$. From Lemma 2.7, the padding of $u,v$ in any scale $\ell \in [i,i+k)$ depends only on the local neighborhoods, $B(u,2\Delta_\ell) \cup B(v,2\Delta_\ell)$, which are disjoint from those of $u',v'$.

By Claim 2.18 there is probability $\ge 1/8$ to succeed, no matter what happened in scales $\ell' < \ell$ or "far away" in scale $\ell$.

We now prove the Lemma. For every coordinate $t \in [D]$, we have $k = \lceil(\theta/3)\log\log n\rceil$ possible values of $\ell$. In each scale $\ell$, by Claim 2.19 there is probability at most $(7/8)$ to fail, this probability bound is unaffected by events for scales $\ell' < \ell$. Let $\mathcal{Y}_\ell$ be the indicator event for $\neg\mathcal{F}_{(i,u,v,t,\ell)}$. The probability that all scales $\ell \in [i,i+k)$ failed is bounded by:

$$\begin{aligned}\Pr\left[\bigwedge_{\ell=i}^{i+k-1} \mathcal{Y}_\ell\right] &= \prod_{\ell=i}^{i+k-1}\left(\Pr\left[\mathcal{Y}_\ell \mid \bigwedge_{j=i}^{\ell-1} \mathcal{Y}_j\right]\right)\\ &\le (7/8)^{\lceil(\theta/3)\log\log n\rceil}\\ &\le \frac{1}{\log^{\theta/18} n} = z.\end{aligned}$$

Let $Z_t$ be indicator that the $t$-th coordinate failed (*i.e.* , $\mathcal{F}_{(i,u,v,t,\ell)}$ does not hold for all $\ell \in [i,i+k)$), so $\Pr[Z_t] \le z$. Let $Z = \sum_{t\in D} Z_t$, since $\mathbb{E}[Z] \le zD$, let $\alpha \ge 1$ be such that $\mathbb{E}[Z] = \frac{zD}{\alpha}$. Using Chernoff's bound,

$$\begin{aligned}\Pr[Z > D/2] &= \Pr\left[Z > \left(\frac{\alpha}{2z}\right)\mathbb{E}[Z]\right]\\ &\le \left(\frac{e^{\alpha/(2z)-1}}{(\alpha/(2z))^{\alpha/(2z)}}\right)^{zD/\alpha}\\ &\le (2ez)^{D/2} \le \lambda^{\frac{3c}{\theta} - \frac{c}{18}\log\log n}\end{aligned}$$

setting $c_1 = 60$ implies that $\Pr[Z > D/2] \le \lambda^{-(c/200)\log\log n} = \lambda^{-11\log\log n}$ for large enough $c$, as required.

## 3 Scaling Embedding of Doubling Metrics

In this section an extension of the previous result to embedding with scaling distortion is sketched, the proof appears in the full version.

For any $x \in X$, let $r_\epsilon(x)$ be the minimal radius such that $|B(x,r_\epsilon(x))| \ge \epsilon n$.
Let $\hat{G}(\epsilon) = \left\{(x,y) \mid d_X(x,y) > \max\{r_{\epsilon/2}(x), r_{\epsilon/2}(y)\}\right\}$.

DEFINITION 3.1. *For metric spaces* $(X,d),(Y,\rho)$ *a non-contracting embedding* $f : X \to Y$ *has coarse scaling distortion* $D(\epsilon)$ *if for any* $\epsilon > 0$, *any* $x,y \in \hat{G}(\epsilon)$ *satisfy*

$$d(x,y) \le \rho(f(x),f(y)) \le D(\epsilon)\cdot d(x,y)$$

THEOREM 3.1. *For any metric space* $(X,d)$ *with* $\dim(X) = \log\lambda$ *there exists an embedding* $f : X \to L_p^D$ *with coarse scaling distortion* $O\left(\log^{26}(\frac{1}{\epsilon})\right)$ *where* $D = O(\log\lambda\log\log\lambda)$.

### 3.1 Proof overview

Note the the proof shown in the previous section, in conjunction of the techniques of [4] yields scaling distortion $O(\log^\theta n)$ in dimension $O((\log\lambda)/\theta)$.

In this section we highlight the differers between the proof of Theorem 1.1 and Theorem 1.2. We assume the reader is familiar with the proof of Theorem 1.1.

1. Every pair $x, y$ has some $\epsilon$ value which corresponds to the maximal $\epsilon$ such that $\max\{r_{\epsilon/2}(x), r_{\epsilon/2}(y)\} < d(x, y)$. We partition the possible $\epsilon \in (0, 1]$ values into $\approx \log\log\log n$ buckets . For each scale $\Delta_i$ and each of the $\approx \log\log\log n$ possible values of $\epsilon$ we build a $\approx \Delta_i/\text{polylog}(\lambda, 1/\epsilon)$-net.

   A naive approach would be to assign separate coordinates for each possible $\epsilon$ value, and increase the dimension and hence the distortion by a factor of $\log\log\log n$. To avoid paying this $\log\log\log n$ factor we sieve the nets in a subtle manner

2. The local growth rate of each point is defined with respect to some $\epsilon$ value in non standard manner - this is done so that for sufficiently many levels (as a function of $\epsilon$) there will be a density change.

3. A pair with distance $\approx \Delta_i$ and epsilon that falls into bucket $k$ (hence $k \approx \log\log(1/\epsilon)$) "looks" for a contribution in the levels $i + k/2, \ldots, i + k$ . This is necessary in order to avoid collisions between contributing scales of pairs with different $\epsilon$ values.

4. Showing independence of lower bound successes between two pairs is technical and relies on the sieving process. For a pair $u, v$ related to a net the scales examined are $\approx i + k/2, \ldots i + k$. We show that examining only these scales ensures that $u, v$ are independent of a pair $u', v'$ if one of the following occurs (1) $u', v'$ belong to a different scale than that of $u, v$; (2) $u', v'$ are far enough from $u, v$ in the metric space; (3) $u', v'$ has a different $\epsilon$ value from that of $u, v$.

5. Proving that all pairs have the desired scaling distortion given that the sieved net points have this property is more involved now since it depends on the $\epsilon$, hence an argument that uses the the fact that the padding parameter is lower bounded by the inverse doubling dimension must be used.

6. The application of the local lemma is complicated due to two issues - (1) we use the general case (2) we do not proceed simply from scale $i$ to scale $i+1$, but rather use the ranking function in a non-trivial manner.

## 4 Snowflake Results

In this section we show the main ideas of the proof of Theorem 1.3.

### 4.1 Proof overview

Let $\dim(X) = \log\lambda$. The high level approach is similar to that of Theorem 1.1. However, there are several differences

- In order to obtain an upper bound $O(k)$ where $k$ can be independent of the number of points in the metric *and* of the doubling dimension, our partition has a large padding parameter $\tau = \Omega(1/k)$, this will cause the probability of the padding event to be as small as $\lambda^{-1/k}$.

- In each term of the embedding for scale $i$ (i.e. $f_i(x)$) we introduce a factor of $\Delta_i^{\alpha-1}$, this ensures that for a given $x, y$, the sum over all scales larger than $d(x, y)$ behaves like a geometric progression, hence the upper bound for $|f(x) - f(y)|$ is independent of the number of scales.

- In the lower bound analysis, the density of the nets is now only a function of $k$ (and $\alpha$), hence the number of scales needed to be examined for a contribution is also only a function of $k$, which yields the desired lower bound with small probability (the padding probaility). The fact that this probability is $\approx \lambda^{-1/k}$ induces the inverse of this factor in the dimension and a factor of $\lambda^{1/(pk)}$ in the distortion.

## 5 Conclusion

The natural remaining open question is to understand the best possible distortion of embeddings a metric space $X$ into Euclidean space in dimension $O(\dim(X))$. Specifically it seems interesting to understand the best possible distortion of embeddings into Euclidean space with constat dimension for the metrics arising form the Laakso graph and from the Heisenberg group. Another intriguing question is whether a version of the Johnson-Lindenstrauss lemma [19] exists which relates the embedding dimension to the intrinsic dimension, specifically given a subset $X$ of Euclidean space, does it embed into $O(\dim(X))$ dimensional Euclidean space with constant distortion ?

## References


[1] *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*. IEEE Computer Society, 2005.

[2] Ittai Abraham, Yair Bartal, Hubert T.-H. Chan, Kedar Dhamdhere, Anupam Gupta, Jon M. Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *FOCS* [1], pages 83–100.

[3] Ittai Abraham, Yair Bartal, Hubert T.-H. Chan, Kedar Dhamdhere, Anupam Gupta, Jon M. Kleinberg, Ofer

Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *FOCS* [1], pages 83–100.

[4] 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.

[5] Noga Alon. A parallel algorithmic version of the local lemma. In *Proceedings of the 32nd annual symposium on Foundations of computer science*, pages 586–593, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.

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

[7] 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.

[8] J. Beck. An algorithmic approach to the lovasz local lemma. *Random Struct. Algorithms*, 2:343–365, 1991.

[9] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, 2003.

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

[11] Gruia Calinescu, Howard J. Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Symposium on Discrete Algorithms*, pages 8–16, 2001.

[12] H. Chan, A. Gupta, and K. Talwar, 2007. Personal communication.

[13] Manuel Costa, Miguel Castro, Antony I. T. Rowstron, and Peter B. Key. Pic: Practical internet coordinates for distance estimation. In *24th International Conference on Distributed Computing Systems*, pages 178–187, 2004.

[14] Russ Cox, Frank Dabek, M. Frans Kaashoek, Jinyang Li, and Robert Morris. Practical, distributed network coordinates. *ACM SIGCOMM Computer Communication Review*, 34(1):113–118, 2004.

[15] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 448–455. ACM Press, 2003.

[16] Paul Francis, Sugih Jamin, Cheng Jin, Yixin Jin, Danny Raz, Yuval Shavitt, and Lixia Zhang. Idmaps: a global internet host distance estimation service. *IEEE/ACM Trans. Netw.*, 9(5):525–540, 2001.

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

[18] Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM J. Comput*, 35(5):1148–1184, 2006.

[19] 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.

[20] Jon M. Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. In *FOCS*, pages 444–453, 2004.

[21] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 434–443. IEEE, October 2004.

[22] Hyuk Lim, Jennifer C. Hou, and Chong-Ho Choi. Constructing internet coordinate system based on delay measurement. In *3rd ACM SIGCOMM Conference on Internet Measurement*, pages 129–142, 2003.

[23] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[24] Nathan Linial and Michael Saks. Decomposing graphs into regions of small diameter. In *Proceedings of the second annual ACM-SIAM symposium on Discrete algorithms*, pages 320–330, Philadelphia, PA, USA, 1991. Society for Industrial and Applied Mathematics.

[25] Jiri Matoušek. Open problems on low-distortion embeddings of finite metric spaces, 2005. Avaliable in http://kam.mff.cuni.cz/ matousek/metrop.ps.gz.

[26] T. S. Eugene Ng and Hui Zhang. Predicting internet network distance with coordinates-based approaches. In *21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pages 178–187, 2002.

[27] David Peleg. *Distributed Computing: A Locality-Sensitive Approach*. SIAM Monographs on Discrete Mathematics and Applications, 2000.

[28] 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.

[29] S. Semmes. On the nonexistence of bilipschitz parameterizations and geometric problems about $a_\infty$ weights. *Revista Matemática Iberoamericana*, 12:337–410, 1996.

[30] Yuval Shavitt and Tomer Tankel. Big-bang simulation for embedding network distances in euclidean space. *IEEE/ACM Trans. Netw.*, 12(6):993–1006, 2004.

[31] Aleksandrs Slivkins. Distance estimation and object location via rings of neighbors. In $24^{th}$ *Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 41–50. ACM Press, July 2005. Appears earlier as Cornell CIS technical report TR2005-1977.

[32] Kunal Talwar. Bypassing the embedding: algorithms for low dimensional metrics. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 281–290. ACM Press, 2004.

[33] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24, 2005.