Ultra-Low-Dimensional Embeddings for Doubling Metrics

T-H. Hubert Chan* Anupam Gupta† Kunal Talwar‡

Abstract

We consider the problem of embedding a metric into low-dimensional Euclidean space. The classical theorems of Bourgain and of Johnson and Lindenstrauss imply that *any* metric on n points embeds into an $O(\log n)$-dimensional Euclidean space with $O(\log n)$ distortion. Moreover, a simple "volume" argument shows that this bound is nearly tight: the uniform metric on n points requires $\Omega(\log n/\log\log n)$ dimensions to embed with logarithmic distortion. It is natural to ask whether such a volume restriction is the only hurdle to low-dimensional low-distortion embeddings. Do *doubling* metrics, which do not have large uniform submetrics, embed in low dimensional Euclidean spaces with small distortion? In this paper, we answer the question positively and show that any *doubling metric* embeds into $O(\log\log n)$ dimensions with $o(\log n)$ distortion. In fact, we give a suite of embeddings with a smooth trade-off between distortion and dimension: given an n-point metric (V, d) with doubling dimension $\dim_D$, and any target dimension T in the range $\Omega(\dim_D \log\log n) \le T \le O(\log n)$, we embed the metric into Euclidean space $\mathbb{R}^T$ with $O(\log n\sqrt{\dim_D/T})$ distortion.

1 Introduction

We consider the problem of representing a metric (V, d) using a small number of dimensions. Often, applications represent their data as points in a Euclidean space with thousands of dimensions, and this high-dimensionality poses a computational challenge: algorithms tend to have an exponential dependence on the dimension. Hence we constantly seek ways to combat this "curse of dimensionality," by finding low-dimensional yet faithful representations of the data. In this work, we attempt to maintain all pairwise distances, i.e. we seek to minimize the *distortion* of an embedding, and ask the following question: *given a metric space* (which may or may not be Euclidean to begin with), *what is the least number of dimensions in which it can be represented with "reasonable" distortion?*

Dimension reduction (in Euclidean spaces) has been studied extensively. The celebrated "flattening" lemma of Johnson and Lindenstrauss [29] states that the dimension of any Euclidean metric on n points can be reduced to $O(\frac{\log n}{\varepsilon^2})$ with $(1+\varepsilon)$ distortion, and moreover, this can be done via a random linear map. This result is existentially tight: a simple packing argument shows that any distortion-D embedding of a uniform metric on n points into Euclidean space requires at least $\Omega(\log_D n)$ dimensions (and a lower bound pinning down the dependence on ε appears in [6]). Hence we do need the $\Omega(\log n)$ dimensions for the uniform metric, and even allowing $O(\log n)$ distortion cannot reduce the number of dimensions below $\Omega(\log n/\log\log n)$.

It is natural to ask if such "volume" restrictions form the only bottleneck to low-dimensional embeddings. In other words, can metrics that do not have large uniform sub-metrics be embedded into low-dimensional spaces with small distortion? The notion of *doubling dimension* [7] makes this idea concrete: roughly speaking, a metric has doubling dimension $\dim_D = k$ if and only if it has (nearly-)uniform submetrics of size about 2^k, but no larger. A metric (or more strictly, a family of metrics) is simply called *doubling* if the doubling dimension is bounded by a universal constant. (See section 1.2 for a more precise definition).

A packing lower bound shows that any metric requires $\Omega(\dim_D)$ dimensions for a constant-distortion embedding into Euclidean space: is this lower bound tight? We now know the existence of n-point metrics with $\dim_D = O(1)$ that require $\Omega(\sqrt{\log n})$-distortion into Euclidean space of *any* dimension [24], but can we actually achieve this distortion with $o(\log n)$-dimensions? What if we give up a bit in the distortion? Bourgain's embedding combined with the JL-lemma shows that all metrics embed into Euclidean space of $O(\log n)$ dimensions and $O(\log n)$ distortion [38], but we do not know if doubling metrics embed into $O(\log^{1-\varepsilon} n)$ dimensions with $O(\log^{1-\varepsilon} n)$ distortion. Moreover, we

*Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. This research was partly supported by the NSF grant CCR-0122581 (The ALADDIN project), the NSF CAREER award CCF-0448095, and by an Alfred P. Sloan Fellowship.

†Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. This research was partly supported by the NSF CAREER award CCF-0448095, and by an Alfred P. Sloan Fellowship.

‡Microsoft Research, Silicon Valley Campus, 1065 La Avenida, Mountain View, CA 94043.

do not know whether this is true even for *Euclidean* doubling metrics: while it is conceivable that all Euclidean metrics with $\dim_D = O(1)$ embed into $O(1)$ dimensional Euclidean space with $O(1)$ distortion, the best result known is the JL-Lemma (which is completely oblivious to the doubling dimension, and is known to be tight even for doubling metrics).

We make progress on the problem of embedding doubling metrics into Euclidean space with small dimension and distortion. (Our results hold for *all* doubling metrics, not just Euclidean ones.)

THEOREM 1.1. (ULTRA-LOW-DIMENSION EMBEDDING) *Any metric space with doubling dimension* $\dim_D \leq O(\frac{\log n}{\log\log n})$ *embeds into* $O(\dim_D \log\log n)$ *dimensions with* $O(\frac{\log n}{\sqrt{\log\log n}})$ *distortion.*

Hence we can embed any such metric into few Euclidean dimensions (i.e., $\tilde{O}(\dim_D)$, where the notation $\tilde{O}(\cdot)$ suppresses a multiplicative factor polynomial in $\log\log n$), and achieve a distortion comparable to Bourgain's embedding. Note that to achieve distortion $O(\log n)$, any metric with doubling dimension $\dim_D$ requires at least $\Omega(\frac{\dim_D}{\log\log n})$ Euclidean dimensions, and hence we are within an $O(\log\log n)^2$ factor to the *optimal dimension* for this value of distortion. Theorem 1.1 is a special case of our general trade-off theorem:

THEOREM 1.2. (MAIN THEOREM) *Suppose* (V, d) *is a metric space with doubling dimension* $\dim_D$. *For any integer* T *such that* $\Omega(\dim_D \log\log n) \leq T \leq \ln n$, *there exists* $F : V \to \mathbb{R}^T$ *into* T*-dimensional space such that for all* $x, y \in V$, $d(x,y) \leq \|F(x) - F(y)\|_2 \leq O\left(\sqrt{\frac{\dim_D}{T}}\ \log n\right) \cdot d(x,y)$.

Varying the target dimension T, we can get some interesting tradeoffs between the distortion and dimension. For instance, we can balance the two quantities and get $O(\log^{2/3} n)$ dimensions and $O(\log^{2/3} n)$ distortion for doubling metrics, as desired. On the other hand, for large target dimension $T = \ln n$, we get distortion $O(\sqrt{\dim_D \log n})$, which matches the best known result from [33]. In the interests of clarity of presentation, we only show the *existence* of such embeddings. Standard techniques (e.g,. [10, 5, 41]) can be used to give algorithmic versions of our results. Moreover, due to space constraints, some of the proofs have been deferred to the full version of the paper.

Techniques. Our embedding can be thought of as an extension of Rao's embedding [42]: there are $O(\log n)$ copies of coordinates for each distance scale, hence leading to $O(\log n \log \Delta)$ dimensions. As observed in [2], it is possible to sum up the coordinates over different distance scales to form one coordinate, and in expectation the contraction is bounded. Using bounded doubling dimension, we show that there is limited dependency between pairs of points (using the Lovasz Local Lemma), and hence we only need much less than $O(\log n)$ coordinates to ensure that the contraction for all points are bounded.

For the tradeoff between the target dimension and the distortion, we apply a random sign (± 1) to the contribution for each distance scale before summing them up to form a coordinate. This process is analogous to the random projection in JL-type embeddings. Indeed, we use analysis similar to that in [3] to obtain a tradeoff between the target dimension and the expansion, although in our case the original metric needs not be Euclidean.

We give two embeddings: the first one uses a simple decomposition scheme [24, 44, 17] and illustrates the above ideas in bounding both the contraction and the expansion. The resulting embedding has distortion $O(\dim_D / \sqrt{T} \cdot \log n)$ with T dimensions. In order to reduce the dependence on the doubling dimension to $\sqrt{\dim_D}$, we use *uniform* padded decomposition schemes based on [2].

Bibliographic Note. In another paper appearing in this conference, Abraham, Bartal, and Neiman, also present results giving embeddings achieving a tradeoff between distortion and dimension as a function of the doubling dimension $\dim_D$ and the number of points n. Though of a similar nature, the two papers use somewhat different techniques, and the results are not strictly comparable. For instance, they can achieve $O(\dim_D)$-dimensional embeddings—smaller than ours by an $O(\log\log n)$ factor—though with slightly superlogarithmic distortion. On the other hand, our trade-off at the higher end of dimension is slightly better. They also present results on gracefully degrading distortion and average distortion (in the sense defined in [1, 2]).

1.1 Related Work

Dimension reduction for Euclidean space was first studied by Johnson and Lindenstrauss [29], using random projections. The results and techniques have since been sharpened and simplified in [23, 27, 20, 3, 4]. The embeddings have been derandomized, see [21, 43]. Moreover, Matousek [40] has obtained an almost tight tradeoff between the dimension of the target space and the distortion of the embedding. On the other hand, dimension reduction for L_1 space has been shown to be much harder in [13, 37].

The notion of doubling dimension was introduced by Larman [36] and Assouad [7], and first used in algorithm design by Clarkson [18]. The properties of doubling metrics and their algorithmic applications have since been studied extensively, a few examples of which appear in [24, 34, 35, 44, 25, 11, 19, 28, 32, 31].

There is extensive work on metric embeddings, see [26]. Bourgain [12] gave an embedding whose coordinates are formed by distances from random subsets. Low diameter decomposition is a useful tool and was studied by Awerbuch [8], and Linial and Saks [39]. Randomized decompositions for general metrics are given in [9, 16, 22]. Klein et al. [30] gave decomposition schemes for minor-excluding graphs, which were used by Rao [42] to obtain embeddings for planar graphs into Euclidean space. These ideas were developed further in [33, 1, 2]. Finally, approximation algorithms for embeddings into constant dimensional spaces have also been investigated, both for general metrics [15] and special classes of metrics, for instance ultra-metrics [14].

1.2 Notation and Preliminaries We denote a finite metric space by (V, d), its size by $n = |V|$, and its doubling dimension $\dim_D$ by k. For any positive integer A, we denote $[A] := \{0, 1, 2, \ldots, A-1\}$. We assume that the minimum distance between two points is 1, and hence its diameter Δ is also the aspect ratio of the metric. A *ball* $B(x, r)$ is the set $\{y \in V \mid d(x, y) \leq r\}$. Recall that for $r > 0$, an r-net N for (V, d) is a subset of V such that (i) for all $x \in V$, there exists $y \in N$ such that $d(x, y) \leq r$; and (ii) for alll $x, y \in N$ such that $x \neq y$, $d(x, y) > r$.

Definition 1.1. (Doubling Dimension $\dim_D$) *The* doubling dimension *of a metric (V, d) is at most k if for all $x \in V$, for all $r > 0$, every ball $B(x, 2r)$ can be covered by the union of at most 2^k balls of the form $B(z, r)$, where $z \in V$.*

Definition 1.2. (Padded Decompostion) *Given a finite metric space (V, d), a positive parameter $D > 0$ and $\alpha > 1$, a D-bounded α-*padded decomposition *is a distribution Π over partitions of V such that the following conditions hold.*

(a) For each partition P in the support of Π, the diameter of every cluster in P is at most D.

(b) If P is sampled from Π, then $Pr[B(x, \frac{D}{\alpha}) \subseteq P(x)] \geq \frac{1}{2}$, where $P(x)$ is the cluster in P containing x.

2 The Basic Embedding

We give two embeddings in this paper: this section we present the basic embedding, which achieves the following trade-off between dimension and distortion.

Theorem 2.1. (The Basic Embedding) *Given a metric space (V, d) with doubling dimension $\dim_D$, and a target dimension T in the range $\Omega(\dim_D \log\log n) \leq T \leq \ln n$, there exists a mapping $f : V \to \mathbb{R}^T$ such that for all $x, y \in V$, $\Omega\left(\frac{\sqrt{T}}{\dim_D}\right) \cdot d(x, y) \leq ||f(x) - f(y)||_2 \leq O(\log n) \cdot d(x, y)$. Hence, the distortion is $O(\frac{\dim_D \log n}{\sqrt{T}})$.*

Note that this trade-off is slightly worse than than the one claimed in Theorem 1.2 in terms of its dependence on the doubling dimension; however, the advantage is that this embedding is easier to state and prove. We will then improve on this embedding in the next section.

2.1 Basic Embedding: Defining The Embedding The embedding $f : (V, d) \to \mathbb{R}^T$ we describe is of the form $f := \oplus_{t \in [T]} \Phi^{(t)}$, where the symbol $\oplus$ is used to denote the concatenation of the various coordinates. Each $\Phi^{(t)} : V \to \mathbb{R}$ is a single coordinate generated independently of the other coordinates according to a probability distribution described as follows. To simplify notation, we drop the superscript t and describe how a random map $\Phi : V \to \mathbb{R}$ is constructed, and f is just the concatenation of T such coordinates.

Let $D_i := H^i$, for some constant $H \geq 2$. (Later we see that H is set large enough to bound the contraction.) Suppose all distances in the metric space are at least 1, and I is the largest integer such that $D_{I-1} < \Delta$. The mapping $\Phi : V \to \mathbb{R}$ is of the form $\Phi := \sum_{i \in [I]} \varphi_i$. We describe how $\varphi_i : V \to \mathbb{R}$ is constructed, for each $i \in [I]$.

Fix $i \in [I]$. We view the metric (V, d) as a weighted complete graph, and contract all edges with lengths at most $D_i/2n$. The points that are contracted together in this process would obtain the same value under φ_i. Let the resulting metric be (V, d_i). Here are a few properties of the metric (V, d_i).

Proposition 2.1. *Suppose for each $i \in [I]$, the metric (V, d_i) is defined as above. Then, for all $x, y \in V$, the following results hold.*

(a) For all $i \in [I]$, $d_i(x, y) \leq d(x, y) \leq d_i(x, y) + \frac{D_i}{2}$.

(b) For $j \geq i$, $d_j(x, y) \leq d_i(x, y)$.

Observe that Proposition 2.1 implies that the metric (V, d_i) gives good approximations of the distances in (V, d) of scales above D_i. In particular, (V, d_i) admits an $O(k)$-padded D_i-bounded stochastic decomposition.

Proposition 2.2. (Padded Decomposition for Doubling Metrics [24, 44, 17]) *Suppose the metric (V, d) has doubling dimension k. Then, there is an α-padded D_i-bounded stochastic decomposition Π_i for the metric (V, d_i), where $\alpha = O(k)$. Moreover, the event $\{B_i(x, D_i/\alpha) \subseteq P_i(x)\}$ is independent of all the events $\{B_i(z, D_i/\alpha) \subseteq P_i(z) : z \notin B_i(x, 3D_i/2)\}$, where $B_i(u, r) := \{v \in V : d_i(u, v) \leq r\}$.*

Suppose P_i is a random partition of (V, d_i) sampled from the padded decomposition Π_i of Proposition 2.2. Let $\{\sigma_i(C) : C \text{ is a cluster in } P_i\}$ be uniform $\{0,1\}$-random variables, and γ_i be a uniform $\{-1,1\}$-random variable. The random objects P_i, σ_i and γ_i are sampled independently of one another. Define $\varphi_i : V \to \mathbb{R}$ by

$$\varphi_i(x) := \gamma_i \cdot \kappa_i(x), \tag{2.1}$$

where $\kappa_i(x) := \sigma_i(P_i(x)) \cdot \min\{d_i(x, V \setminus P_i(x)), D_i/\alpha\}$.

Hence we take the distance from the point x to the closest point outside its cluster, truncate it at D_i/α (recall that α is as defined in Proposition 2.2), and multiply it with the $\{0,1\}$ r.v. associated with its cluster, and the $\{-1,1\}$ r.v. associated with the distance scale i. We shall see that the σ_i's play an important role in bounding the contraction, while the role of γ_i's is to bound the expansion. To summarize, the embedding is defined to be:

$$f := \oplus_{t\in[T]} \Phi^{(t)}; \Phi^{(t)} := \sum_{i\in[I]} \varphi_i^{(t)}. \tag{2.2}$$

We rephrase Theorem 2.1 in terms of the above randomized construction.

THEOREM 2.2. *Suppose the input metric (V, d) has doubling dimension k, and the target dimension T is in the range $\Omega(k \log\log n) \le T \le \ln n$. Then, with non-zero probability, the above procedure produces a mapping $f : V \to \mathbb{R}^T$ such that for all $x, y \in V$, $\Omega\left(\frac{\sqrt{T}}{\dim_D}\right) \cdot d(x,y) \le ||f(x) - f(y)||_2 \le O(\log n) \cdot d(x,y)$. In other words, there exist some realization of the various random objects such that the distortion of the resulting mapping is $O(\frac{\dim_D \log n}{\sqrt{T}})$.*

Note. Before we begin, note that we consider the modified metrics (V, d_i) in order to avoid a dependence on the aspect ratio Δ in the expansion bound for the embedding. Also observe that $|\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y)| \le \min\{d_j(x,y), D_j/\alpha\}$. The proof of the following simple lemma is given in the full version.

LEMMA 2.1. *Suppose $x, y \in V$ and for each $j \in [I]$, define $d_j := \min\{d_j(x,y), D_j/\alpha\}$. Then, the following results hold.*

(a) For each $i \in [I]$, $\sum_{j\ge i} d_j \le O(\log_H n) \cdot d_i(x,y)$.

(b) For each $i \in [I]$, $\sum_{j\ge i} d_j^2 \le O(\log_H n) \cdot d_i(x,y)^2$.

In particular, $\sum_{j\ge i} d_j \le O(\log_H n) \cdot d_i(x,y)$, and $\sum_{i\in[I]} d_i^2 \le O(\log_H n) \cdot d(x,y)^2$.

Proof. We prove statements (a) and (b). The other statements follow from the two in a straight forward manner.

Observe that for $j \ge i$, $d_j \le d_j(x,y) \le d_i(x,y)$, where the second inequality follows from Proposition 2.1(b).

There are three cases to consider depending on the value of j. The first is for very large j's when $d(x,y) \le \frac{D_j}{2n}$: in this case, $d_j(x,y) = 0$. The second case is for moderate values of j when $\frac{D_j}{2n} < d(x,y) \le D_j$: there are at most $O(\log_H n)$ such j's. In (a), adding these up gives a contribution of $O(\log_H n) \cdot d_i(x,y)$; in (b), we have a contribution of $O(\log_H n) \cdot d_i(x,y)^2$.

Finally, the last case is for small values of j, when $d(x,y) > D_j$. Consider the largest j_0 for which this happens. Then, it follows from Proposition 2.1 that $d_i(x,y) \ge d_{j_0}(x,y) > D_{j_0}/2$. Observing that $d_j \le D_j/\alpha$ and $\{D_j\}$ forms a geometric sequence, it follows that $\sum_{i\le j\le j_0} d_j = O(d_i(x,y))$, and $\sum_{i\le j\le j_0} d_j^2 = O(d_i(x,y)^2)$.

Combining the three cases gives the result. □

2.2 Basic Embedding: Bounding Contraction

A natural idea to bound the contraction for a particular pair of points x, y is to use the padding property of the random decomposition: if $d(x,y) \approx H^i$, then at the corresponding scale $i \in [I]$ the two vertices will be in different clusters, and will contribute a large distance. This idea has been extensively used in previous work starting with [42]. However, in these previous works, we have a separate coordinate for each distance scale, which leads to a large number of dimensions. Abraham et al. [2] show that the coordinates for distance scales can actually be combined to form one single coordinate, and with constant probability the contraction is still bounded. Now we want to use a small number of coordinates as well: to do this, we exploit small doubling dimension to use the Lovasz Local Lemma and bound the contraction for all pairs of points.

Fixing the γ's. As noted in the description of the embedding, the γ's do not play any role in bounding the contraction. In fact, we will show something *stronger*: for any realization of the γ's, there exists some realization of the P's and σ's for which the contraction of the embedding f is bounded. For the rest of this section, we assume that the γ's are arbitrarily fixed upfront.

For each $i \in [I]$, let the subset N_i be an arbitrary βD_i-net of (V, d_i), for some $0 < \beta < 1$ to be specified later.

Bounding the Contraction for some Special Points. We first bound the contraction for the pairs in $E_i := \{(x,y) \in N_i \times N_i : 3D_i/2 < d_i(x,y) \le 3HD_i\}$, $i \in [I]$. (Note that from Proposition 2.1(a), it follows that for each $(x,y) \in E_i$, $d(x,y) < 4HD_i$.)

For $t \in [T]$, and $(x,y) \in E_i$, define $A^{(t)}(x,y)$ to be the event that *all the following* happens:

- the vertex x is well-padded: i.e., $B_i(x, \frac{D_i}{\alpha}) \subseteq P_i^{(t)}(x)$;
- the vertex y is mapped to 0: $\sigma_i^{(t)}(P_i^{(t)}(y)) = 0$;
- if $|\sum_{j>i}(\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y))| \leq \frac{D_i}{2\alpha}$, then $\sigma_i^{(t)}(P_i^{(t)}(x)) = 1$, otherwise $\sigma_i^{(t)}(P_i^{(t)}(x)) = 0$.

PROPOSITION 2.3. (CONDITIONING ON HIGHER LEVELS) *Let $(x, y) \in E_i$. Suppose for $j > i$, the random objects $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : t \in [T]\}$ have been arbitrarily fixed. For each $t \in [T]$, sample random partition $P_i^{(t)}$ from Proposition 2.2 and random $\{0,1\}$-variables $\{\sigma_i^{(t)}(C) : C$ is a cluster of $P_i^{(t)}\}$ uniformly, all independently of one another. Then, for each $t \in [T]$, with probability at least $\frac{1}{8}$, the event $A^{(t)}(x, y)$ happens independently over the different t's.*

Moreover, if the event $A^{(t)}(x, y)$ happens, then the inequality $|\sum_{j\geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq \frac{D_i}{2\alpha}$ holds; furthermore, for any realization of the remaining random objects, i.e., $\gamma_i^{(t)}$ and $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : j < i\}$, the inequality $|\sum_{i\in[I]}(\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y))| \geq \frac{D_i}{4\alpha}$ holds, provided $H \geq 8$. (Recall that $D_{i+1} = HD_i$.)

Proof. Given any realization of the random objects of scales larger than i, each of the three defining events for $A^{(t)}(x, y)$ happens independently of one another with probability at least $\frac{1}{2}$, and hence $A^{(t)}(x, y)$ happens with probability at least $\frac{1}{8}$, independently over $t \in [T]$ (since the random objects at scale i are sampled independently over $t \in [T]$).

It follows that if $A^{(t)}(x, y)$ happens, then the partial sum from large scales up to scale i is $|\sum_{j\geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq \frac{D_i}{2\alpha}$. Observe the sum from smaller scales $|\sum_{j<i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))|$ is bounded above by a geometric sum $\sum_{j<i} \frac{D_j}{\alpha}$, which is at most $\frac{D_i}{4\alpha}$, provided that $H \geq 8$. □

In order to show that the contraction for the pair (x, y) is small, we need to show that the event $A^{(t)}(x, y)$ happens for a constant fraction of t's. We define $C(x, y)$ to be the event that for at least $\frac{T}{16}$ values of t, the event $A^{(t)}(x, y)$ happens. We conclude that the event $C(x, y)$ happens with high probability (as a function of T), by using a Chernoff bound: if X is the sum of i.i.d. Bernoulli random variables, then $Pr[X < (1 - \epsilon)E[X]] \leq \exp(-\frac{1}{2}\epsilon^2 E[X])$, for $0 < \epsilon < 1$.

PROPOSITION 2.4. (USING CONCENTRATION) *Under the sampling procedure described in Proposition 2.3, the event $C(x, y)$ fails to happen with probability at most $p := \exp(-\frac{T}{64})$.*

Proof. This follows by applying the Chernoff bound mentioned above with $\epsilon = \frac{1}{2}$. □

Now that each event $C(x, y)$ happens with high enough probability, we use the Lovasz Local Lemma to show that there is some realization of $\{P_i^{(t)}, \sigma_i^{(t))} : t \in [T]\}$ such that for all $(x, y) \in E_i$, the *events $C(x, y)$ happen simultaneously.* In order to use the Local Lemma, we need to analyze the dependence of these events. Recall that N_i is a βD_i-net of (V, d_i).

LEMMA 2.2. (LIMITED DEPENDENCE) *For each $(x, y) \in E_i$, the event $C(x, y)$ is independent of all but $B := (\frac{H}{\beta})^{O(k)}$ of the events $C(u, v)$, where $(u, v) \in E_i$.*

Proof. Observe that the event $C(x, y)$ is determined by the random objects $\{P_i^{(t)}, \sigma_i^{(t)} : t \in [T]\}$. More specifically, it is determined completely by the events $\{B_i(w, \frac{D_i}{\alpha}) \subseteq P_i^{(t)}(w) : t \in [T]\}$ and $\{\sigma_i^{(t)}(P^{(t)}(w)) = 0 : t \in [T]\}$, for $w \in \{x, y\}$. Note that if $d_i(x, w) > 3D_i/2$, then the corresponding events for the points x and w are independent. Note that if $d_i(x, w) \leq 3D_i/2$, then $d(x, w) \leq 2D_i$; moreover, any two net-points in (V, d_i) must be more than βD_i apart in (V, d). Hence, observing that the doubling dimension of the given metric is at most k, for each of x and y, only $(\frac{2D_i}{\beta D_i})^{O(k)}$ net points are relevant. Now, each net point can be incident by at most $(\frac{4H}{\beta})^{O(k)}$ edges in E_i. Hence, it follows that $C(x, y)$ is independent of all but $(\frac{H}{\beta})^{O(k)}$ of the events $C(u, v)$, where $(u, v) \in E_i$. □

Now we can apply the (symmetric form of the) Lovasz Local Lemma.

LEMMA 2.3. (LOVASZ LOCAL LEMMA) *Suppose there is a collection of events such that each event fails with probability at most p. Moreover, each event is independent of all but B other events. Then, if $ep(B + 1) < 1$, then all the events in the collection happen simultaneously with non-zero probability.*

PROPOSITION 2.5. (ONE MORE LEVEL) *Suppose for $j > i$, the random objects $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : t \in [T]\}$ have been arbitrarily fixed. If $T = \Omega(k \log \frac{H}{\beta})$, then there is some realization of $\{P_i^{(t)}, \sigma_i^{(t)} : t \in [T]\}$ such that all the events $\{C(x, y) : (x, y) \in E_i\}$ happen. In particular, such a realization does not depend on the γ's at scale i.*

Define $\mathcal{E}$ to be the event that for all $i \in [I]$, for all $(x, y) \in E_i$, the event $C(x, y)$ happens. By applying Proposition 2.5 repeatedly, we show that the event $\mathcal{E}$ happens with non-zero probability.

Proposition 2.6. (Contraction for Nearby Net Points) *Suppose in the construction the γ's are arbitrarily fixed, and the P's and σ's are still random and independent. Moreover, suppose $T = \Omega(k \log \frac{H}{\beta})$. Then, with non-zero probability, our random construction produces an embedding $f : (V, d) \rightarrow \mathbb{R}^T$ such that the event $\mathcal{E}$ happens; in particular, there exists some realization of the P's and σ's such that $||f(x) - f(y)||_2 \geq \frac{\sqrt{T}}{4} \cdot \frac{D_i}{4\alpha}$.*

Bounding the Contraction for All Points. We next bound the contraction for an arbitrary pair (u, v) of points noting that if all net points do not suffer large contraction (by the above argument), and all pairs do not incur a large expansion (by the argument of Lemma 2.1), then one can extend the contraction result to all pairs of points. Of course, to do so, the net N_i must be sufficiently fine. Recall that N_i is a βD_i-net for (V, d_i).

Lemma 2.4. (Extending to All Pairs) *Suppose the event $\mathcal{E}$ happens, and β is small enough such that $\frac{1}{\beta} = \Theta(\alpha \log_H n)$. Then, for any $x, y \in V$, there exist $T/16$ values of t's for which*
$|\Phi^{(t)}(x) - \Phi^{(t)}(y)| = \Omega(d(x, y))/\alpha H$.

Hence, by setting $H = 16$ and $\frac{1}{\beta} = \Theta(\alpha \log_H n)$, and observing $\alpha = O(k)$ from Proposition 2.2 (where k is the doubling dimension and is at most $\log n$), we have the following result.

Proposition 2.7. (Bounding Contraction) *Suppose the γ's are arbitrarily fixed and β is sufficiently small such that $\frac{1}{\beta} = \Theta(\alpha \log_H n)$ and $H \geq 16$. Then, for $T = \Omega(k \log \log n)$, there exists some realization of P's and σ's that produces an embedding $f : V \rightarrow \mathbb{R}^T$ such that for all $x, y \in V$, $\|f(x) - f(y)\|_2 \geq \Omega(\frac{\sqrt{T}}{k}) \cdot d(x, y)$.*

2.3 Basic Embedding: Bounding Expansion

Recall that $\mathcal{E}$ is the event $\cap_{i \in [I]} \cap_{(x,y) \in E_i} C(x, y)$. We showed in Proposition 2.6 that $Pr[\mathcal{E}] > 0$, and if the event $\mathcal{E}$ happens, the resulting embedding $f : V \rightarrow \mathbb{R}^T$ has bounded contraction. We now bound the expansion of the embedding $f : V \rightarrow \mathbb{R}^T$ for every pair (x, y) of points. In order to bound this expansion, the $\{-1, +1\}$-random variables γ_i will finally be used. Their role is fairly natural: if the contributions from different distance scales are simply summed up, then there would be a factor of $|I|$ (roughly speaking) appearing in the expansion for each coordinate. However, with the random variables γ_i's, the sum starts to behave like a random walk, and the expectation of the sum of the signed contributions would only suffer a factor of $\sqrt{I}$. In order to make this argument formal, we use techniques similar to those used in analyzing the Johnson-Lindenstrauss lemma [3]. The main problem that arises here is that if we condition on the event $\mathcal{E}$, not only the different coordinates of the map but also the γ's are no longer independent, and hence we would not be able to use the "random walk"-like argument. Instead, we sample the γ's first and fix the P's and σ's accordingly in order to apply the large-deviation arguments.

Fixing the P's and σ's. Suppose the γ's are sampled uniformly and independently. From Proposition 2.7, there exists some realization of the P's and the σ's such that the contraction of the embedding f is bounded. Hence, from this point, we can concentrate on bounding the expansion. Since the γ's are randomly drawn, the P's and the σ's are random variables too, and are functions of the γ's. Proposition 2.5 gives a clear idea of the dependency between the random variables: the P's and the σ's at scale i are determined only by the random objects at scales strictly larger than i, and in particular are independent of the γ's at scale i.

Let us fix $x, y \in V$ and define the random variable

$$S := ||f(x) - f(y)||_2^2 = \sum_{t \in [T]} (Q^{(t)})^2,$$

where $Q^{(t)} := \Phi^{(t)}(x) - \Phi^{(t)}(y)$. (The coordinates Φ were defined in (2.1).) We want to show that for large enough T, the r.v. S does not deviate too much from its mean with high probability. Then, a union bound over all pairs (x, y) of points leads to the conclusion that with non-zero probability, the embedding f has bounded expansion.

Observe that $Q^{(t)} := \sum_{i \in [I]} \gamma_i^{(t)} Y_i^{(t)}$, where $Y_i^{(t)} := \kappa_i^{(t)}(x) - \kappa_i^{(t)}(y)$. Define $d_i := \min\{d_i(x, y), D_i/\alpha\}$. Recall that the random variables $\gamma_i^{(t)}$ are uniformly picked from $\{-1, +1\}$, and $|Y_i^{(t)}| \leq d_i$. We can illustrate the dependency between the different random objects in Figure 1.

For i from $I - 1$ down to 0, do:

1. For each $t \in [T]$, the value $Y_i^{(t)}$ is picked adversarially from $[-d_i, d_i]$, hence possibly depending on previously picked values $\{Y_j^{(t)}, \gamma_j^{(t)} : j > i, t \in [T]\}$.
2. For each $t \in [T]$, $\gamma_i^{(t)}$ is picked *uniformly* from $\{-1, +1\}$, and moreover, *independent* of any random objects picked thus far.

Figure 1: Sampling the various random variables.

LEMMA 2.5. (COMPUTING THE M.G.F.) *Suppose the γ's and Y's are picked according to the above description. Moreover, $\nu^2 := \sum_{i\in[I]} d_i^2$. Then for $0 \le h\nu^2 < 1/2$, $E[\exp(hS)] \le (1-2h\nu^2)^{-T/2}$. Moreover, for $\varepsilon > 0$, $Pr[S > (1+\varepsilon)T\nu^2] \le ((1+\varepsilon)\exp(-\varepsilon))^{T/2}$.*

The proof of Lemma 2.5 appears in Section 2.4. Using this lemma, we can bound the expansion of the embedding.

PROPOSITION 2.8. (BOUNDING EXPANSION) *Suppose the target dimension T is at most $\ln n$. Then, for each pair $x, y \in V$, with probability at least $1 - \frac{1}{n^2}$, $||f(x) - f(y)||_2 \le O(\log n) \cdot d(x,y)$.*

Proof. Let $\nu^2 := \sum_{i\in[I]} d_i^2$, and recall that $S = ||f(x) - f(y)||_2^2$. Then, from Lemma 2.5, we have for $\varepsilon > 0$, $Pr[S > (1+\varepsilon)T\nu^2] \le ((1+\varepsilon)\exp(-\varepsilon))^{T/2}$.

Note that for $\varepsilon \ge 8$, $(1+\varepsilon)\exp(-\varepsilon) \le \exp(-\varepsilon/2)$. Hence, for $T \le \ln n$, we set $\varepsilon := \frac{8\ln n}{T}$ and from Lemma 2.1, we have $\nu^2 = \sum_{i\in[I]} d_i^2 \le O(\log n)\cdot d(x,y)^2$. Hence, with failure probability at most $\frac{1}{n^2}$, we have $||f(x) - f(y)||_2^2 \le (1 + \frac{8\ln n}{T}) \cdot T \cdot O(\log n) \cdot d(x,y)^2 \le O(\log^2 n) \cdot d(x,y)^2$. □

Using the union bound over all pairs (x, y) and combining with Proposition 2.7 completes the proof for the low distortion embedding claimed in Theorem 2.2 (modulo the proof of Lemma 2.5, which appears in the following section). In Section 3, we will give an embedding improves the dimension-distortion tradeoff, and proves Theorem 1.2.

2.4 Resolving Dependency among Random Variables Suppose we wish to bound the magnitude of the following sum, whose terms are dependent on one another:

$$S := \sum_{t\in[T]} (Q^{(t)})^2, \tag{2.3}$$

where for each $t \in [T]$, $Q^{(t)} := \sum_{i\in[I]} \gamma_i^{(t)} Y_i^{(t)}$. The $\gamma_i^{(t)}$'s are $\{-1,+1\}$ random variables; for each $i \in [I]$, the $Y_i^{(t)}$'s are random variables taking values in the interval $[-d_i, d_i]$. Figure 1 specifies how the various random variables are being sampled.

A standard technique to analyze the magnitude of S defined in (2.3) is to consider the moment generating function (m.g.f.) $E[\exp(hS)]$, for sufficiently small $h > 0$. This is fairly easy when the terms in the summation S are independent: however, observe that each $Y^{(t)}$ is dependent on the random objects indexed by $j > i$. Moreover, the $Q^{(t)}$'s are not independent either. However, we can get around this and prove the following result, via Lemmas 2.6 and 2.7.

Lemma 2.5 (Computing the m.g.f.) Suppose $\nu^2 := \sum_{i\in[I]} d_i^2$. Then for $0 \le h\nu^2 < 1/2$, $E[\exp(hS)] \le (1 - 2h\nu^2)^{-T/2}$. Moreover, for $\varepsilon > 0$, $Pr[S > (1+\varepsilon)T\nu^2] \le ((1+\varepsilon)\exp(-\varepsilon))^{T/2}$.

Recall that the problem was that each $Y^{(t)}$ is dependent on the random objects indexed by $j > i$. Moreover, the $Q^{(t)}$'s are not independent either. To get around this, we consider random variables related to $Q^{(t)}$. Define $\widehat{Q}^{(t)} := \sum_{i\in[I]} \gamma_i^{(t)} d_i$ and $\overline{Q}^{(t)} := \sum_{i\in[I]} g_i^{(t)} d_i$, where the $g_i^{(t)}$'s are independent normal $N(0,1)$ variables. Define $\widehat{S} := \sum_{t\in[T]} (\widehat{Q}^{(t)})^2$ and $\overline{S} := \sum_{t\in[T]} (\overline{Q}^{(t)})^2$ analogously. Observe that both the $\widehat{Q}^{(t)}$'s and the $\overline{Q}^{(t)}$'s are independent over different t's. Define $\nu^2 := \sum_{i\in[I]} d_i^2$. A standard calculation gives us that $E[\exp(h\overline{S})] \le (1 - 2h\nu^2)^{-T/2}$, for $0 \le h\nu^2 < 1/2$. We show that $E[\exp(hS)]$ is bounded above by the same quantity.

As observed in [3], by the Monotone Convergence Theorem, we have $E[\exp(hS)] = \sum_{r\ge 0} \frac{h^r}{r!} E[S^r]$. Hence, we compare the even powers of Q, $\widehat{Q}$ and $\overline{Q}$.

LEMMA 2.6. *The following inequalities hold.*

1. *For any integer $r \ge 0$, $E[\widehat{Q}^{2r}] \le E[\overline{Q}^{2r}]$.*
2. *For any real number $h > 0$, $E[\exp(h\widehat{S})] \le E[\exp(h\overline{S})]$.*

Proof. The first statement follows from the observation that $E[\gamma_i^{2r}] = 1 \le E[g_i^{2r}]$. The second statement follows from the first statement, observing that the $\widehat{Q}^{(t)}$'s and the $\overline{Q}^{(t)}$'s are independent, and using the identity $E[\exp(hZ)] = \sum_{r\ge 0} \frac{h^r}{r!} E[Z^r]$. □

The next lemma resolves the issue that the $Q^{(t)}$'s are not independent. The idea is to replace each random variable $Y_i^{(t)}$ by a constant d_i and show that this does not decrease the expectation of the relevant random variables.

LEMMA 2.7. *The following properties hold.*

1. *For all $r_t \ge 0$ ($t \in [T]$), $E[\prod_{t\in[T]} (Q^{(t)})^{2r_t}] \le E[\prod_{t\in[T]} (\widehat{Q}^{(t)})^{2r_t}]$.*
2. *For $h > 0$, $E[\exp(hS)] \le E[\exp(h\widehat{S})]$.*

Proof. Note the second statement follows from the first using the identity $E[\exp(hZ)] = \sum_{r\ge 0} \frac{h^r}{r!} E[Z^r]$, and hence it suffices to prove the first statement. Let us define the partial sums $Q_i^{(t)} := \sum_{j\ge i} \gamma_i^{(t)} Y_i^{(t)}$ and $\widehat{Q}_i^{(t)} := \sum_{j\ge i} \gamma_i^{(t)} d_i$. We show the following statement by backward induction on i. The case $i = 1$ gives the required result. We show that for $i \in [I]$, for all $r_t \geq 0$ ($t \in [T]$), $E[\prod_{t\in[T]}(Q_i^{(t)})^{2r_t}] \leq E[\prod_{t\in[T]}(\widehat{Q}_i^{(t)})^{2r_t}]$.

The case $i = I$ follows from the fact that for all $r \geq 0$, for all $t \in [T]$, $|Y_I^{(t)}| \leq d_I$. Hence, for all $r_t \geq 0$ ($t \in [T]$), $E[\prod_{t\in[T]}(Q_I^{(t)})^{2r_t}] = E[\prod_{t\in[T]}(Y_I^{(t)})^{2r_t}] \leq E[\prod_{t\in[T]}(d_I)^{2r_t}] = E[\prod_{t\in[T]}(\widehat{Q}_I^{(t)})^{2r_t}]$.

Assume that for all $l_t \geq 0$ ($t \in [T]$), $E[\prod_{t\in[T]}(Q_{i+1}^{(t)})^{2l_t}] \leq E[\prod_{t\in[T]}(\widehat{Q}_{i+1}^{(t)})^{2l_t}]$, for $i \geq 0$. Fix some $r_t \geq 0$ ($t \in [T]$).

(2.4) $\quad E[\prod_{t\in[T]}(Q_i^{(t)})^{2r_t}]$

(2.5) $= \quad E[\prod_{t\in[T]}(Q_{i+1}^{(t)} + \gamma_i^{(t)} Y_i^{(t)})^{2r_t}]$

(2.6) $= \quad E[\sum_{l_1=0,\ldots,l_t=0}^{r_1,\ldots,\ r_t} \prod_{t\in[T]} \binom{2r_t}{2l_t} (Q_{i+1}^{(t)})^{2r_t-2l_t} (Y_i^{(t)})^{2l_t}]$

(2.7) $\leq \quad E[\sum_{l_1=0,\ldots,l_t=0}^{r_1,\ldots,\ r_t} \prod_{t\in[T]} \binom{2r_t}{2l_t} (Q_{i+1}^{(t)})^{2r_t-2l_t} d_i^{2l_t}]$

(2.8) $\leq \quad E[\sum_{l_1=0,\ldots,l_t=0}^{r_1,\ldots,\ r_t} \prod_{t\in[T]} \binom{2r_t}{2l_t} (\widehat{Q}_{i+1}^{(t)})^{2r_t-2l_t} d_i^{2l_t}]$

(2.9) $= \quad E[\prod_{t\in[T]}(\widehat{Q}_i^{(t)})^{2r_t}]$

The equality (2.6) uses the fact that the r.v.'s $\gamma_i^{(t)}$'s are independent of all other random variables and the expectation of an odd power of $\gamma_i^{(t)}$ is 0. The inequality (2.7) follows from the fact that $|Y_i^{(t)}| \leq d_i$. The inequality (2.8) follows from the linearity of expectation and the induction hypothesis. Finally, equality (2.9) holds for the same reason as that for (2.6). This completes the inductive proof. □

Finally, we are in a position to prove Lemma 2.5:

Proof of Lemma 2.5: From Lemma 2.7, we have $E[\exp(hS)] \leq E[\exp(h\widehat{S})]$, which is at most $E[\exp(h\overline{S})]$, by Lemma 2.6. Finally, from a standard calculation [20], $E[\exp(h\overline{S})] \leq (1 - 2h\nu^2)^{-T/2}$, for $0 \leq h\nu^2 < 1/2$.

To prove the second part of the lemma, let $h\nu^2 = \frac{\varepsilon}{2(1+\varepsilon)} < \frac{1}{2}$. Then, we have

$$\begin{aligned} & Pr[S > (1+\varepsilon)T\nu^2] \\ = \;& Pr[\exp(hS) > \exp((1+\varepsilon)Th\nu^2)] \\ \leq \;& E[\exp(hS)]\exp(-(1+\varepsilon)Th\nu^2) \\ \leq \;& (1-2h\nu^2)^{-T/2} \cdot \exp((1+\varepsilon)Th\nu^2) \\ = \;& ((1+\varepsilon)\exp(-\varepsilon))^{T/2}. \end{aligned}$$

which proves the large-deviation inequality. □

3 A Better Embedding via Uniform Padded Decompositions

Our basic embedding in the previous section uses a simple padded decomposition [17], and serves to illustrate the proof techniques: however, its dependence on $\dim_D$ is sub-optimal. In order to improve the dependence of the distortion on the doubling dimension, we use a more sophisticated decomposition scheme. We modify the uniform padded decomposition in [2], by incorporating the properties of bounded doubling dimension directly within the construction, to achieve both the padding property, as well as independence between distant regions.

3.1 Uniform Padded Decompositions

DEFINITION 3.1. (UNIFORM FUNCTIONS) *Given a partition P of (V, d), a function $\eta : V \to \mathbb{R}$ is* uniform with respect to the partition P *if points in the same cluster take the same value under η, i.e., if $P(x) = P(y)$, then $\eta(x) = \eta(y)$.*

For $r > 0$ and $\gamma > 1$, the "local growth rate" is denoted by $\rho(x, r, \gamma) := \frac{|B(x,r\gamma|)}{|B(x,r/\gamma)|}$, and $\overline{\rho}(x, r, \gamma) := \min_{z\in B(x,r)} \rho(z, r, \gamma)$. All logarithms are based 2 unless otherwise specified.

We show that if (V, d) has bounded doubling dimension, there exists a uniformly padded decomposition. The following lemma is similar to [2, Lemma 4], except that it has additional properties about bounded doubling dimension, and also independence between distant regions. The proof is given in the full version.

LEMMA 3.1. (UNIFORM PADDED DECOMPOSITION) *Suppose (V, d) is a metric space with doubling dimension k, and $D > 0$. Let $\Gamma \geq 8$. Then, there exists a D-bounded α-padded decomposition Π on (V, d), where $\alpha = O(k)$, with the following properties. For each partition P in the support of Π, there exist uniform functions $\xi_P : V \to \{0, 1\}$ and $\eta_P : V \to (0, 1)$ such that $\eta_P \geq \frac{1}{\alpha}$. Moreover, if $\xi_P(x) = 1$, then $2^{-7}/\log\rho(x, D, \Gamma) \leq \eta_P(x) \leq 2^{-7}$; if $\xi_P(x) = 0$, then $\eta_P(x) = 2^{-7}$ and $\overline{\rho}(x, D, \Gamma) < 2$.*

Then, for all $x \in V$, the probability of the event $\{B(x, \eta_P(x)D) \subseteq P(x)\}$ is at least $\frac{1}{2}$. Furthermore, the event $\{B(x, \eta_P(x)D) \subseteq P(x)\}$ is independent of all the events $\{B(z, \eta_P(z)D) \subseteq P(z) : z \notin B(x, 3D/2)\}$.

3.2 The Better Embedding: Defining the Embedding The new embedding is quite similar to the basic embedding of Section 2.1. We use the uniform padded decomposition of Lemma 3.1 to define the new embedding $f : (V, d) \to \mathbb{R}^T$. As before, the metric (V, d) has doubling dimension $\dim_D = k$, and suppose $\alpha = O(k)$ is the padding parameter in Lemma 3.1. Let $D_i := H^i$, and assume that the distances in (V, d) are between 1 and H^{I-1}.

Again, the embedding is of the form $f := \oplus_{t\in[T]}\Phi^{(t)}$, where each $\Phi^{(t)} : V \to \mathbb{R}$ is generated independently according to some distribution; for ease of notation, we drop the superscript t in the following. Also, each Φ is of the form $\Phi := \sum_{i\in[I]} \varphi_i$. We next describe how each $\varphi_i : V \to \mathbb{R}$ is constructed.

For each $i \in [I]$, let P_i be a random partition of (V,d) sampled from the decomposition scheme as described in Lemma 3.1. Suppose $\xi_{P_i} : V \to \{0,1\}$ and $\eta_{P_i} : V \to (0,1)$ are the associated uniform functions with respect to the partition P_i. Let $\{\sigma_i(C) : C \text{ is a cluster of } P_i\}$ be uniform $\{0,1\}$-random variables and γ_i be a uniform $\{-1,+1\}$-random variable. The random objects P_i's, σ_i's and γ_i's are independent of one another. Then φ_i is defined by the realization of the various random objects as:

$$\varphi_i(x) := \gamma_i \cdot \kappa_i(x), \tag{3.10}$$

where $\kappa_i(x) := \sigma_i(P_i(x)) \cdot \min\{\xi_{P_i}(x)\eta_{P_i}(x)^{-1/2} d(x, V \setminus P_i(x)), \frac{D_i}{\sqrt{\alpha}}\}$. Note the difference with (2.1) is in the definition of κ_i.

The proof bounding the distortion will proceed similarly: we show that with non-zero probability, the embedding $f : V \to \mathbb{R}^T$ has low distortion.

3.3 The Better Embedding: Bounding Contraction for Nearby Net Points Again, we assume that the γ's are arbitrarily fixed, and the P's and σ's are random and independent. For each $i \in [I]$, let the subset N_i be an arbitrary βD_i-net of (V,d), for some $0 < \beta < 1$ to be specified later. As in the basic embedding, we first bound the contraction for the pairs in $E_i := \{(x,y) \in N_i \times N_i : 3D_i < d(x,y) \le 4HD_i\}$, $i \in [I]$, and then extend it to all pairs in Section 3.5. The proof of the following proposition appears in the full version.

Proposition 3.1. (Contraction for Nearby Net Points) *Suppose $T = \Omega(k \log \frac{H}{\beta})$. Moreover, the γ's are arbitrarily fixed, and the P's and σ's remain random and independent. Then, there exists some realization of the P's and σ's such that the embedding $f : (V,d) \to \mathbb{R}^T$ satisfies for all $i \in [I]$, for all $(x,y) \in E_i$, $||f(x) - f(y)||_2 \ge \frac{\sqrt{T}}{4} \cdot \frac{D_i}{4\sqrt{\alpha}}$.*

3.4 The Better Embedding: Bounding the Expansion Again, we sample the γ's uniformly and independently, and use Proposition 3.1 to show there exists some realization of the P's and σ's such that the resulting mapping $f : V \to \mathbb{R}^T$ has the guaranteed contraction. Hence, we can focus on analyzing the expansion.

Again, fix $x, y \in V$ and let $S := ||f(x) - f(y)||_2^2 = \sum_{t \in [T]} (Q^{(t)})^2$, where $Q^{(t)} := \sum_{i \in [I]} \gamma_i^{(t)} Y_i^{(t)}$, and $Y_i^{(t)} := \kappa_i^{(t)}(x) - \kappa_i^{(t)}(y)$. Recall that $\gamma_i^{(t)}$ is uniformly picked from $\{-1,+1\}$. Denote $d := \max\{\sqrt{O(\log \rho(x, D_i, \Gamma))}, \sqrt{O(\log \rho(y, D_i, \Gamma))}\} \cdot d(x,y)$, and $\nu^2 := \sum_{i \in [I]} d_i^2$ We next bound the magnitudes of the Y_i's and ν^2 in the following Lemma, whose proof follows the same argument as in [2, Lemma 8].

Lemma 3.2. *Consider a particular $Y_i = \kappa_i(x) - \kappa_i(y)$. Then, $|Y_i| \le d_i$, and $\nu^2 = O(\log_H \Gamma \log n) \cdot d(x,y)^2$.*

The proof now proceeds in the same fashion as in Section 2.3; setting $H := 16$ and $\Gamma := 128$, we have $\nu^2 = O(\log n) \cdot d(x,y)^2$. Hence, applying Lemma 2.5, and setting $\varepsilon := \frac{8 \ln n}{T}$ as before, we have the following result.

Lemma 3.3. (Bounding Expansion) *Suppose $T \le \ln n$. Then, for each pair $x, y \in V$, with probability at least $1 - \frac{1}{n^2}$, $||f(x) - f(y)||_2 \le O(\log n) \cdot d(x,y)$.*

3.5 The Better Embedding: Bounding Contraction for All Pairs Now that we have proved that with non-zero probability, the expansion for every pair of points is at most $O(\log n)$, and the contraction for nearby net points is bounded, we can show that if the βD_i-net N_i for (V,d) is fine enough (i.e., β is small enough), then the contraction bound can be extended to *all pairs.*

Lemma 3.4. (Bounding Contraction for All Pairs) *Suppose the event $\mathcal{E}$ holds and the expansion of the embedding f is bounded in the manner described in Lemma 3.3. Suppose $\beta > 0$ is small enough such that $\beta^{-1} = \Theta(\sqrt{\alpha} \log n)$, where $\alpha = O(k)$. Then, for all $x, y \in V$, $||f(x) - f(y)||_2 \ge \Omega(\sqrt{T/\alpha}) \cdot d(x,y)$.*

Lemmas 3.3 and 3.4 together prove Theorem 1.2.

References

[1] 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 *46th FOCS*, pages 83–100, 2005.

[2] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. In *38th STOC*, pages 271–286, 2006.

[3] Dimitris Achlioptas. Database friendly random projections. In *Proceedings of the Twenteenth ACM Symposium on Principles of Database Systems*, 2000.

[4] Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *38th STOC*, pages 557–563, 2006.

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

[6] Noga Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003.

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

[8] Baruch Awerbuch. Complexity of network synchronization. *J. Assoc. Comput. Mach.*, 32(4):804–823, 1985.

[9] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th FOCS*, pages 184–193, 1996.

[10] József Beck. An algorithmic approach to the Lovász local lemma. I. *Random Structures Algorithms*, 2(4):343–365, 1991.

[11] Alina Beygelzimer, Sham Kakade, and John Langford. Cover trees for nearest neighbor. In *23rd International Conference on Machine Learning*, pages 97–104, 2006.

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

[13] Bo Brinkman and Moses S. Charikar. On the impossibility of dimension reduction in l1. In *44th FOCS*, pages 514–523, 2003.

[14] Mihai Bădoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropou. Embedding ultrametrics into low-dimensional spaces. In *22nd SOCG*, pages 187 – 196, 2006.

[15] Mihai Bădoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Low-distortion embeddings of general metrics into the line. In *37th STOC*, pages 225–233, New York, NY, USA, 2005. ACM Press.

[16] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. *SIAM J. Comput.*, 34(2):358–372, 2004/05.

[17] Hubert T-H. Chan, Anupam Gupta, Bruce M. Maggs, and Shuheng Zhou. On hierarchical routing in DOubling metrics. In *16th SODA*, pages 762–771, 2005.

[18] K. L. Clarkson. Nearest neighbor queries in metric spaces. *Discrete Comput. Geom.*, 22(1):63–93, 1999.

[19] Richard Cole and Lee-Ad Gottlieb. Searching dynamic point sets in spaces with bounded doubling dimension. In *38th STOC*, pages 574–583, 2006.

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

[21] Lars Engebretsen, Piotr Indyk, and R. O'Donnell. Derandomized dimensionality reduction with applications. In *13th SODA*, pages 715–712, 2002.

[22] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004.

[23] Péter Frankl and Hiroshi Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(3):355–362, 1988.

[24] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low–distortion embeddings. In *44th FOCS*, pages 534–543, 2003.

[25] Sariel Har-Peled and Manor Mendel. Fast constructions of nets in low dimensional metrics, and their applications. In *21st SOCG*, pages 150–158, 2005.

[26] Piotr Indyk and Jiri Matoušek. Low-distortion embeddings of finite metric spaces. In Jacob E. Goodman and Joseph O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, chapter 8, pages xviii+1539. Chapman & Hall/CRC, second edition, 2004.

[27] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *30th STOC*, pages 604–613, 1998.

[28] Piotr Indyk and Assaf Naor. Nearest neighbor preserving embeddings. *ACM Transactions on Algorithms*, To appear.

[29] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[30] Philip Klein, Serge A. Plotkin, and Satish B. Rao. Excluded minors, network decomposition, and multicommodity flow. In *25th STOC*, pages 682–690, 1993.

[31] Goran Konjevod, Andra W. Richa, and Donglin Xia. Optimal scale-free compact routing schemes in doubling networks. In *18th SODA*, pages 939–948, 2007.

[32] Goran Konjevod, Andréa W. Richa, and Donglin Xia. Optimal-stretch name-independent compact routing in doubling metrics. In *25th PODC*, pages 198–207, 2006.

[33] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005.

[34] Robert Krauthgamer and James R. Lee. The intrinsic dimensionality of graphs. In *35th STOC*, pages 438–447. ACM Press, 2003.

[35] Robert Krauthgamer and James R. Lee. Navigating nets: simple algorithms for proximity search. In *15th SODA*, pages 798–807, 2004.

[36] D. G. Larman. A new theory of dimension. In *Proc. London Math. Soc., 17*, 1967.

[37] James R. Lee and Assaf Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. In *Geometric and Functional Analysis (GAFA)*. Springer Verlag, 2003.

[38] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. (Preliminary version in *35th FOCS*, 1994).

[39] Nathan Linial and Michael Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993. (Preliminary version in *2nd SODA*, 1991).

[40] Jiří Matoušek. Bi-Lipschitz embeddings into low dimensional Euclidean spaces. *Commentationes Mathematicae Universitatis Carolinae*, 31(3):589–600, 1990.

[41] Michael Molloy and Bruce Reed. Further algorithmic aspects of the local lemma. In *STOC '98 (Dallas, TX)*, pages 524–529. ACM, New York, 1999.

[42] Satish B. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *15th SOCG*, pages 300–306, 1999.

[43] D. Sivakumar. Algorithmic derandomization via complexity theory. In *34th STOC*, pages 619–626. ACM Press, 2002.

[44] Kunal Talwar. Bypassing the embedding: Algorithms for low-dimensional metrics. In *36th STOC*, pages 281–290, 2004.