On Low Dimensional Local Embeddings

Ittai Abraham[*] Yair Bartal[†] Ofer Neiman[‡]

January 13, 2010

Abstract

We study the problem of embedding metric spaces into low dimensional L_p spaces while faithfully preserving distances from each point to its k nearest neighbors. We show that any metric space can be embedded into $L_p^{O(e^p \log^2 k)}$ with k-local distortion of $O((\log k)/p)$. We also show that any ultrametric can be embedded into $L_p^{O(\log k)/\epsilon^3}$ with k-local distortion $1+\epsilon$.

Our embedding results have immediate applications to *local Distance Oracles*. We show how to preprocess a graph in polynomial time to obtain a data structure of $O(nk^{1/t}\log^2 k)$ bits, such that distance queries from any node to its k nearest neighbors can be answered with stretch $O(t)$.

[*]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).

1 Introduction

In [ABN07], we initiated the study of *local embeddings*, embeddings that preserve the local structure of the original space. Indeed in many important applications of embedding, preserving the distances of nearby points is much more important than preserving all distances. An embedding with k-local distortion of α is a map from a metric space to host metric space such that the distances from each point to its k nearest neighbors are faithfully preserved in the host space with distortion $\leq \alpha$. For k-local embedding into a normed space, say L_2, the challenge is to obtain an embedding whose *distortion* and *dimension* depend solely on k. Examining the metric of constant degree expander graphs shows that the best one can hope for is a k-local distortion of $\Omega(\log k)$ using $\Omega(\log k)$ dimensions. A partial answer to the problem was provided in [ABN07] (see Theorem 8), it was shown that under a *constant weak growth bound*[1] assumption on the metric, such embeddings exist. Unfortunately, even metrics arising from simple graphs like an un-weighted star do not have constant growth bound. In fact even a one dimensional subset of L_2 can have an unbound growth rate.

Many models and measurements for the Internet network (for example, "power law" models) predict that the Internet network has a very high, non-constant, growth rate. Moreover, it seems that the growth bound assumption was essential for the type of embedding in [ABN07], since the Local Lemma argument had to depend on events which are slightly farther away than the k nearest neighbor.

In this paper we show that no matter how large the metric is, if one is interested in constant distortion for the distances of the nearest neighbors of each point then the metric can be "folded" into a constant dimensional space. We show that any metric space can be embedded into L_p with k-local distortion $O((\log k)/p)$ and dimension that depends only on p and k.

Theorem 1. *For any n point metric space (X, d) and parameters $k \leq n$, $p \leq \log k$ there exists an embedding into L_p with k-local distortion $O((\log k)/p)$ and dimension $O(e^p \log^2 k)$.*

The celebrated Johnson Lindenstrauss dimension reduction Lemma [JL84] states that for any n points in L_2^n and any $\epsilon > 0$ there exists an embedding into $L_2^{O(\log n/\epsilon^2)}$ with distortion $1+\epsilon$. This result has numerous applications in many practical areas like Learning, Artificial Intelligence and Databases. In [ABN07] it was asked if similar k local dimension reduction results exist where the dimension is $O(\log k)$ and the k-local distortion is constant. On the negative side, Adi Shraibman and Gideon Schechtman [SS09] recently showed that obtaining such embedding with $1+\epsilon$ distortion for all metric spaces is impossible. In fact they show the nearly tight bound known for the general case, of $\Omega((\log n)/(\epsilon^2 \log(1/\epsilon)))$ on the number of dimensions required to embed n point subset of L_2 with 2-local distortion $1+\epsilon$. On the positive side, we study k local dimension reduction embeddings for the family of ultrametrics. Bartal and Mendel [BM04] show that an ultrametric (X, d) can be embedded with distortion $1+\epsilon$ into L_2 with dimension $O(\log |X| \log(1/\epsilon)\epsilon^{-2})$. We give a *local* analogue of this result:

Theorem 2. *Let (X, d) be an ultrametric, then for any $p \geq 1$, $0 < \epsilon \leq 1$ and $k \leq |X|$ there is an embedding of X into L_p with k-local distortion $1+\epsilon$ and dimension $O((\log k)/\epsilon^3)$.*

[1] a metric X, d has a $\chi > 1$ *weak growth bound* if $|B(u, \log |B(u,r)| r)| \leq |B(u,r)|^\chi$ for all $u, r > 0$ such that $|B(u,r)| > 1$

The main new ingredient in the proof is Lemma 12 which states that any m-bounded out-degree HST can be embedded with $1+\epsilon$ distortion using only $O((\log m)/\epsilon^3)$ dimensions.

1.1 Local Distance Oracles

Consider the following well known problem. Given a description of a large network, such as the Internet, or a large road network, such as the US road network. We wish to preprocess the network, so that subsequent distance queries could be answered quickly and accurately. Solutions to this problem are known as *distance oracles* (see Peleg [Pel00], Thorup and Zwick [TZ05]). For an n node network, a well known asymptotically tight trade off obtains distance oracles with $\tilde{\Theta}(n^{1+1/t})$ bits that answers distance queries with $\Theta(t)$ stretch. *i.e.* if the distance between u and v is $d(u,v)$ then the distance oracle returns an approximation $h(u,v)$ such that $d(u,v) \leq h(u,v) \leq t \cdot d(u,v)$.

In this paper we show that better trade offs exists if one is mainly interested in local distances. For example, suppose that we are interested in obtaining approximate distances on a large road network, but we are mainly interested to give distance queries between endpoints whose distance can be driven by vehicle in one day, or suppose we are interested in approximating Internet latencies, but we are only interested in distances within our local network neighborhood. Following [ABN07], a k local distance oracle, is a data structure that faithfully preserves the distances between each node and its k nearest neighbors. For k local distance oracle with stretch 1, the obvious solution is to store a table with $k \cdot n$ entries, each containing the required distance. Our main results give k local distance oracles, for any parameter $t \leq \log k$, using only $O(nk^{1/t}\log^2 k)$ bits and answers distance queries with $O(t)$ stretch.

Distance oracles and metric embeddings are closely related. Theorem 1 immediately translates into the following distance oracle result.

Corollary 1. *Given an undirected graph with non-negative edge weights on n nodes, and parameters $1 \leq k \leq n$, $1 \leq t \leq \log k$. The graph can be preprocessed in polynomial time to produce a data structure of $O(nk^{1/t}\log^2 k)$ bits, such that distance queries from any node to its k nearest neighbors can be answered with stretch $O(t)$.*

This follows by choosing $p = \frac{\ln k}{t}$, and for every point storing its $O(k^{1/t}\log^2 k)$ coordinates. Notice that for any fixed k, the data structure size is linear in n.

1.2 Local embeddings

Given a metric space (X,d), let $B(u,r) = \{v \mid d(u,v) \leq r\}$. For any point x let $<_x$ be an order relation on the points in $X \setminus \{x\}$ such that for any $u,v \in X \setminus \{x\}$ if $d(u,x) \leq d(v,x)$ then $u <_x v$ (breaking ties arbitrarily). For any $k \in \mathbb{N}$ let $N_k(x)$ be the set of first k elements of $X \setminus \{x\}$ according to $<_x$, i.e., $N_k(x)$ is the set of k nearest neighbors of x. Let $r_k(x)$ be minimal radius such that $N_k(x) \subseteq B(x, r_k(x))$. For any $x \in X$ let $\bar{N}_k(x) = \{y \mid y \in N_k(x) \wedge x \in N_k(y)\}$.

Definition 2. Let (X, d_X) be a metric space on n points, (Y, d_Y) a target metric space and $k \in \mathbb{N}$, let $f : X \to Y$ be an embedding.

- f is *non-expansive* if for any $u,v \in X$, $d_Y(f(u), f(v)) \leq d_X(u,v)$.

- f is an embedding with *k-local distortion* α if f is non-expansive and for any $u, v \in X$ such that $v \in N_k(u)$,
$$d_Y(f(u), f(v)) \geq \frac{d_X(u,v)}{\alpha}.$$

2 Local Embedding into L_p with Low Dimension

Theorem 1. *For any n point metric space (X, d) and parameters $k \leq n$, $p \leq \log k$ there exists an embedding into L_p with k-local distortion $O((\log k)/p)$ and dimension $O(e^p \log^2 k)$.*

Let $s = e^p$. The proof of this theorem will require a composition of two functions $f : X \to L_p^D$ and $g : X \to L_p^{D'}$ with the following properties:

1. Both f and g embed into $D = cs \ln^2 k$ dimensions for a universal constant c.
2. The functions f, g are non-expansive, *i.e.* for all $x, y \in X$
$$\|f(x) - f(y)\|_p \leq d(x,y) \ , \ \|g(x) - g(y)\|_p \leq d(x,y)$$
3. For any pair $x, y \in X$ such that $y \in N_k(x)$ and $d(x,y) < r_k(x)/24$,
$$\|f(x) - f(y)\|_p > Cp \cdot d(x,y)/\log k$$
for a universal constant C.
4. For any pair $x, y \in X$ such that $y \in N_k(x)$ and $r_k(x)/24 \leq d(x,y) \leq r_k(x)$,
$$\|g(x) - g(y)\|_p > C'p \cdot d(x,y)/\log k$$
for a universal constant C'.

The embedding is defined as $f \oplus g$, and it follows directly from property 1 that the dimension is $O(e^p \log^2 k)$ and from properties $2, 3, 4$ that the k-local distortion is $O((\log k)/p)$.

The map f is a simplification of the map of [ABN07] Theorem 8. While f gives a lower bound on the contraction of pairs for which $d(x,y) < r_k(x)/24$, we could not extend the Local Lemma argument to work for arbitrary metrics (non-growth bounded) for the case $r_k(x)/24 \leq d(x,y) \leq r_k(x)$. The reason is that there are dependencies between x, y and other k-nearest neighbor pairs that are in the local neighborhood of x or y. When $d(x,y) \leq r_k(x)/24$ these pairs are fully contained in $B(x, r_k(x))$ so there are at most $\approx k^2$ such pairs, but when $d(x,y) > r_k(x)/24$ there could be $\approx n^2$ of these, and the Local Lemma argument fails. So for this case we need a new map g tailored for "far away" pairs. From a high level the map g takes the approach of the maps of [ABN06, ABN07], however there are several subtle differences whose combination yields the desired result.

We highlight two of the new ideas here:

1. In order to define the map g, we use a new type of probabilistic partition, where clusters are bounded not by their diameter but by the number of points they contain. Since we need to apply the Local Lemma, the padding probability must depend only on local events. A related

partitioning notion was suggested by Charikar, Makarychev and Makarychev in [?], however their partition algorithm was based on the probabilistic partitions of Calinescu , Karloff and Rabani [CKR01], Fakcharoenphol, Rao and Talwar [FRT03], which are inherently non-local and hence cannot be used for our application. The construction of our bounded cardinality probabilistic partition uses the truncated exponential distribution approach of [ABN06]. The proof requires some technical modifications to adapt to the bounded cardinality case (see Lemma 4).

2. A common use of probabilistic partitions for embeddings is to randomly color each cluster by 0 or 1 (see [ABN06]). This typically means that the distortion of a pair depends on the color event of the clusters of both vertices. Even in the local setting it could be that some nodes participates in many pairs (for example the center node in a star metric), then this may create dependencies among many pairs and hence prohibit the use of the Local Lemma. The way that [ABN07] handled this was to assume some growth bound on the metric. To overcome this issue without any assumptions, we deterministically color each cluster into a $\bar{D} = \Theta(\log k)$ dimensional vector in such a way that if y is among the k nearest neighbors of x and x, y belong to different clusters A, B then the hamming distance between the colors of A and B is at least $\bar{D}/8$. This allows to define the success event for the pair x, y for the map g only as a function of the probabilistic partition around x *independent* of the events around y.

We will show a proof of the "large" distance map g, and defer the proof for the "small" distance map f to Appendix A.

2.1 Bounded cardinality probabilistic partitions

Definition 3 (Partition)**.** Let (X, d) be a finite metric space. A partition P of X is a collection of disjoint set of non-empty clusters $\mathcal{C}(P) = \{C_1, C_2, \ldots, C_t\}$ such that $X = \cup_j C_j$. The sets C_i are called clusters. For $x \in X$ we denote by $P(x)$ the cluster containing x.

In order to define the map g, we use a new type of probabilistic partitions, where each cluster contains at most k points. That is, instead of the usual notion of bounded diameter partitions, we require a bound on the *cardinality* of the clusters, as captured in the following definition.

Definition 4. Let $1/k \leq \delta \leq 1$. A distribution on partitions $\hat{\mathcal{P}}$ of a metric space (X, d) is k-bounded and locally padded with parameter δ if

1. For any $P \in \text{supp}(\hat{\mathcal{P}})$ and $x, y \in X$, if $r_k(x)/24 \leq d(x, y)$ then $y \notin P(x)$.
2. Denote by $\mathcal{L}(x)$ the event that $B(x, 2^{-11} r_k(x) \log(1/\delta)/\log k) \subseteq P(x)$. For any $Z \subset X \setminus \bar{N}_k(x)$:

$$\Pr[\neg\mathcal{L}(x) \mid \bigwedge_{z \in Z} \mathcal{L}(z)] \leq 1 - \delta$$

The first property bounds the number of points in each cluster by k. The second property states that for any point x, with probability at least δ, the ball around x with radius proportional to $r_k(x)$ is contained in the cluster that contains x, and that the probability of this event depends only on local events.

Lemma 5. *For any metric space (X, d) on n points, any $k \leq n$ and any $1/k \leq \delta \leq 1$, there exists a k-bounded and locally padded probabilistic partition with parameter δ.*

Proof. See Appendix C. □

2.2 The "large" distances embedding

We now detail the map g, that takes care of pairs such that $r_k(x)/24 \leq d(x, y) \leq r_k(x)$.

Recall that $s = e^p$, and let $\delta = 1/s$ (assuming that $s \leq k$). Let $D' = \hat{D} \cdot \bar{D}$ where $\hat{D} = 16s \ln 4 \cdot \ln k$ and $\bar{D} = 16 \log k$. Let $\hat{\mathcal{P}}$ be a locally k-padded probabilistic partition as in Lemma 5. For each $t \in [\hat{D}]$ fix some $P = P^{(t)} \in \mathcal{P}$ (the particular choice of P will be detailed later in Lemma 9). Define a directed graph $G = (V, E)$, which will be the k-neighborhood graph between the clusters of the partition P. Let the vertex set V be the clusters of P. Draw a directed edge (A, B) between clusters A and B iff there exists points $a \in A$, $b \in B$ such that $b \in N_k(a)$. As every cluster contains at most k points, the out-degree of G is at most k^2.

We use the following property of directed graphs with bounded out-degree.

Lemma 6. *Any directed graph $G' = (V, E)$ with maximal out-degree k can be properly colored*[2] *using $2k + 1$ colors.*

Proof. The proof is by induction. Assume we can color with $2k+1$ colors any graph on less that $|V|$ vertices, whose out-degree is bounded by k. Consider G, the un-directed version of G' (connecting two vertices iff there was an edge between them in G' in either direction). Since there are at most $k \cdot |V|$ edges in the graph, and each edge touches two vertices, there must be a vertex x with $deg(x) \leq 2k$. Remove x and all the edges touching x from G, and note that the resulting graph's degree is still bounded by k. Using the induction hypothesis, properly color the remaining vertices with $2k + 1$ colors. Now we add x back to the graph, since it has at most $2k$ neighbors we can properly color it with a color none of its neighbors has. □

We also use a set S of vectors in $\{-1, 1\}^{O(\log k)}$ such that any two points in S are "far" from each other.

Lemma 7. *For any integer $\bar{D} > 1$ and $\Omega(1/\bar{D}) < \delta \leq 1/2$ there exists a set $S \subseteq \{-1, 1\}^{\bar{D}}$, $|S| \geq 2^{\bar{D}(1-H(\delta))/2}$ (H is the entropy function), such that for any $u, v \in S$, the Hamming distance between u and v is at least $\delta\bar{D}$.*

In particular, fixing $\delta = 1/8$ and recalling that $\bar{D} = 16 \log k$ we get a set S of $2k^2 + 1$ vectors in $\{-1, 1\}^{\bar{D}}$ such that the Hamming distance between each two vectors is at least $\bar{D}/8$. Using Lemma 6 we can properly color G with $m = 2k^2 + 1$ colors, and define $\sigma = \sigma^{(t)} : V \to S$, such that if $(A, B) \in E$ then $\sigma(A) \neq \sigma(B)$, by giving each color class of V a distinct vector in S. For any $t \in [\hat{D}]$ define $g^{(t)} : X \to L_p^{\bar{D}}$ by

$$g^{(t)}(x) = \bar{D}^{-1/p} \cdot d(x, X \setminus P^{(t)}(x)) \cdot \sigma(P^{(t)}(x)).$$

[2] Properly colored means that the end points of every directed edge are colored by different colors.

The embedding $g : X \to L_p^{D'}$ is the normalized concatenation of the $g^{(t)}$s,

$$g(x) = \hat{D}^{-1/p} \bigoplus_{t=1}^{\hat{D}} g^{(t)}(x)$$

Observe that for any cluster $A \in P$, $\sigma(A)$ is a $\bar{D} = O(\log k)$ dimensional vector hence $g^{(t)}(x)$ is a mapping into $\bar{D}$ dimensions and $g(x)$ is a mapping into $D' = \hat{D} \cdot \bar{D} = O(e^p \log^2 k)$ dimensions.

Lemma 8. *There exists a universal constant C_1 such that for any $x, y \in X$, $\|g^{(t)}(x) - g^{(t)}(y)\|_p \leq C_1 \cdot d(x, y)$.*

Proof. We distinguish between two cases

Case 1: $P(x) = P(y)$. Denote by $(a_1, \ldots a_{\bar{D}}) = \sigma(P(x)) = \sigma(P(y))$, then as $\big|d(x, X \setminus P(x)) - d(y, X \setminus P(x))\big| \leq d(x, y)$:

$$\|g^{(t)}(x) - g^{(t)}(y)\|_p^p = \bar{D}^{-1}\big|d(x, X \setminus P(x)) - d(y, X \setminus P(x))\big|^p \sum_{i=1}^{\bar{D}} |a_i|^p \leq d(x, y)^p.$$

Case 2: $P(x) \neq P(y)$, then $d(x, X \setminus P(x)) \leq d(x, y)$, hence

$$\|g^{(t)}(x) - g^{(t)}(y)\|_p^p \leq \|g^{(t)}(x)\|_p^p + \|g^{(t)}(y)\|_p^p \leq 2\bar{D}^{-1}\big|d(x, y)\big|^p \sum_{i=1}^{\bar{D}} 1^p \leq (2d(x, y))^p.$$

□

Lemma 9. *There exists a universal constant C_2 and partitions $P^{(t)} \in \mathrm{supp}(\hat{\mathcal{P}})$ for each $t \in [\hat{D}]$ such that for any $x, y \in X$ with $y \in N_k(x)$ and $r_k(x)/24 < d(x, y) \leq r_k(x)$,*

$$\|g(x) - g(y)\|_p \geq C_2 p \cdot d(x, y)/\log k$$

Proof. Fix any $t \in [\hat{D}]$ and let $P = P^{(t)}$. From the first property of Definition 4, $y \notin P(x)$. Since $y \in N_k(x)$ by the proper coloring $\sigma(P(x)) \neq \sigma(P(y))$. Let $I_{xy} = I_{xy}(t) \subseteq [\bar{D}]$ be the subset of at least $\bar{D}/8$ coordinates such that for any $i \in I_{xy}$ we have $\sigma(P(x))_i \neq \sigma(P(y))_i$, and note that for any two positive numbers a, b we have that $|a \cdot \sigma(P(x))_i - b \cdot \sigma(P(y))_i| = a + b$. By the second property of Definition 4 with probability $1/s$ we have that x is padded, if it holds then $d(x, X \setminus P(x)) \geq 2^{-11} r_k(x) \cdot \log s/\log k \geq 2^{-11} p \cdot d(x, y)/\log k$. So

$$\begin{aligned} \|g^{(t)}(x) - g^{(t)}(y)\|_p^p &\geq \bar{D}^{-1}(d(x, X \setminus P(x)) + d(y, X \setminus P(y)))^p \sum_{i \in I_{xy}} 1^p \\ &\geq |I_{xy}|/\bar{D} \cdot d(x, X \setminus P(x))^p \\ &\geq (1/8) \cdot (2^{-11} p \cdot d(x, y)/\log k)^p. \end{aligned}$$

Let $Z_t(x)$ be an indicator for the event that x is padded in $P^{(t)}$. Note that this is the only requirement for getting sufficient contribution in the t-th coordinate. For any $x, y \in X$ with

$y \in N_k(x)$ and $r_k(x)/24 < d(x,y) \le r_k(x)$ define a success event $\mathcal{E}_{x,y}$, as the existence of a subset $T \subseteq [\hat{D}]$ of size $\hat{D}/(2s)$ such that for all $t \in T$: $Z_t(x)$ holds. Note that if $\mathcal{E}_{x,y}$ holds then

$$\|g(x) - g(y)\|_p^p \ge \hat{D}^{-1} \sum_{t \in T} \|g^{(t)}(x) - g^{(t)}(y)\|_p^p \ge \Omega((1/s) \cdot (p \cdot d(x,y)/\log k)^p).$$

As required, so it remains to show that there exists some choice of randomness such that all events $\mathcal{E}_{x,y}$ for pairs such that $y \in N_k(x)$ hold simultaneously.

Let $Z(x) = \sum_{t \in [\hat{D}]} Z_t(x)$, then $\mathbb{E}[Z(x)] \ge \hat{D}/s$. In order for $\mathcal{E}_{x,y}$ to hold, we need that $Z(x) \ge \hat{D}/(2s)$. Using Chernoff bound,

$$\Pr[Z(x) \le \hat{D}/(2s)] = \Pr[Z(x) \le \mathbb{E}[Z(x)]/2] \le e^{-\hat{D}/(8s)} \le 1/(4k^2).$$

Define a dependency graph whose vertices are events $\mathcal{E}_{x,y}$, and draw an edge $(\mathcal{E}_{x,y}, \mathcal{E}_{x',y'})$ iff $x' \in \bar{N}_k(x)$ (note that this is a symmetric definition). It can be seen that the out-degree of the graph is at most k^2, and the second property of Definition 4 states that given any outcome for events which $\mathcal{E}_{x,y}$ is not connected to by an edge in the dependency graph, the padding probability is bounded accordingly, hence there is probability at most $1/(4k^2)$ that the event $\mathcal{E}_{x,y}$ does not hold. By the Local Lemma (see Lemma 15) there is a choice of randomness for which all good events hold simultaneously. □

3 Local Dimension Reduction

3.1 Local Dimension Reduction for the Equilateral Metric

The "usual suspect" for high dimensionality is the equilateral metric[3]. Alon [?] shows that this it is the best known lower bound example for dimension reduction - an n point equilateral requires dimension at least $\Omega(\log n/(\log(1/\epsilon) \cdot \epsilon^2))$, for $1+\epsilon$ distortion. However, this is *not* the case for local embedding.

To embed an equilateral metric, first consider the neighborhood graph $G = (X, E)$, where $(u,v) \in E$ iff $v \in N_k(u)$ (note that we allow adversarial choice of neighbors). By Lemma 6 there exists a proper coloring of G with $2k+1$ colors, using Lemma 16 with $m = 2k+1$, we can embed every color class to a point in L_p^D where $D = O((\log m)/\epsilon^2)$ and obtain k-local distortion of $1+\epsilon$: For any point $u \in X$, all the points in $N_k(x)$ have different colors from the color of x, so the distance between them is maintained up to $1+\epsilon$ distortion. So for any $\epsilon > 0$, any finite equilateral metric embeds into L_p with k-local distortion $1+\epsilon$ and dimension $O((\log k)/\epsilon^2)$.

3.2 Local Dimension Reduction for Ultrametrics

Even though local dimension reduction is impossible in general, we show that it is also possible for the class of ultrametrics. The local dimension reduction can be done in any L_p space, in contrast to the JL[JL84] (non-local) dimension reduction that can be done in L_2, is impossible in L_1 and unknown for other p. The proof has three main steps: first embed the ultrametric to a $\theta - e$HST (Hierarchically Separated Tree, see definition below) with distortion θ, then embed the HST with

[3] A metric (X,d) is equilateral if $d(u,v) = 1$ for all $u \ne v \in X$

k-local distortion 1 into a bounded degree HST, where the bound is polynomial in k. Finally we extend the general framework of [BM04], who showed dimension reduction for ultrametrics, and give lower dimension for bounded degree HSTs: showing that such HSTs can be embedded, preserving all distances up to $1+\epsilon$, into a dimension that is logarithmic in the degree of the HST. For simplicity of presentation we show here the proof for $p=1$. Recall,

- **Ultrametric:** An ultrametric (X,d) is a metric space satisfying a strong form of the triangle inequality, for all $x,y,z \in X$, $d(x,z) \leq \max\{d(x,y), d(y,z)\}$.
- **HST:** For $\theta \geq 1$, a $\theta - e$HST is a finite metric space defined on the branches of a rooted infinite tree, having a finite number of branches. For branches x,y denote by $\text{lca}(x,y)$ the least common ancestor of x and y in the tree, *i.e.* , the deepest node in $x \cap y$, and by $\text{dlca}(x,y)$ its depth. The distance between branches is defined as $d(x,y) = \theta^{-\text{dlca}(x,y)}$. Denote by x_i the i-th node in the branch x.

Theorem 2. *Let (X,d) be an ultrametric, then for any $p \geq 1$, $0 < \epsilon \leq 1$ and $k \leq |X|$ there is an embedding of X into L_p with k-local distortion $1+\epsilon$ and dimension $O((\log k)/\epsilon^3)$.*

To prove this theorem we first introduce the following lemmata:

Lemma 10 ([Bar96])**.** *For any $\theta > 1$, any ultrametric embeds in a $\theta - e$HST with distortion θ.*

Lemma 11. *For all $\theta > 1$, any $\theta - e$HST T' can be embedded into a $\theta - e$HST T, where every internal node in the tree representation of T has degree at most $2k^2+1$, with k-local distortion 1.*

Proof. For a $\theta - e$HST T' let $r(T')$ denote the root of T' and for a node $u \in T'$ let $c(u)$ the set of all children of u. The intuition behind the construction of T from T' is by defining a neighborhood graph on the children of the root, and unite those children which are not connected by an edge in this graph, thus obtaining a small number of children. Once we have few children continue recursively on each of them. Formally, perform the following recursive process on T' creating T:

1. Let $r = r(T')$. Define a neighborhood graph with vertices $c(r) = \{v_1, \ldots, v_\ell\}$ by adding a directed edge (v_i, v_j) iff one of the branches x in the subtree rooted at v_i has $y \in N_k(x)$ where y is a branch in the subtree rooted in v_j. It can be seen that only children with at most k branches have out-going edges, hence the out-degree of this graph is bounded by k^2.
2. Using Lemma 6 properly color the graph with $m = 2k^2+1$ colors. For any $1 \leq i \leq m$ let $v_{i_1}, \ldots, v_{i_s}$ be the children colored by color i, replace them by a single node r_i and set $c(r_i) = \bigcup_{j=1}^{s} c(v_{i_j})$.
3. For each $1 \leq i \leq m$ continue recursively on the subtree rooted at r_i.

Let x,y be two branches such that $y \in N_k(x)$. Then for any level i of the recursive process, let v be the current root - if it is the case that $v \notin x$ or $v \notin y$ then the unions done to the children of v cannot affect $d(x,y)$. Otherwise, let v_j, v_ℓ be the children of v which lie on branches x,y respectively (it could be that $j=\ell$). Since the graph we define on the children of v contains the (directed) edge (v_j, v_ℓ), the vertices v_j, v_ℓ will be colored by distinct colors and will not be united. It follows that the distance between x,y will never change in the process.

Note that the construction of the tree in the deeper recursion levels is done with respect to the *original* set $N_k(x)$, which guarantees that distances between k-nearest neighbors are preserved. □

Lemma 12. *Let $0 < \epsilon \le 1/2$, $m \in \mathbb{N}$ and $\theta = e^{\epsilon}$. Let T be a $\theta-\mathrm{eHST}$ with branches H, such that the out-degree of every node in T is at most m. Then T can be embedded into $L_p^{\bar{D}}$ with distortion $1+\epsilon$ where $\bar{D} = O((\log m) \cdot (\ln(2\min\{p, 1/\epsilon\})/\epsilon)^2 \cdot \max\{1, 1/(\epsilon p)\}$.*

Proof. Let $\theta = e^{\epsilon}$, $\alpha = \ln(2\min\{p, 1/\epsilon\})$, $d = \alpha/\epsilon$ (note that $\theta^d = \min\{2p, 2/\epsilon\}$). Let $D = c\ln m \cdot \alpha/\epsilon \cdot \max\{1, 1/(\epsilon p)\}$ for some constant c to be determined later and $\bar{D} = D \cdot d$. Let $(e_i)_{i\in\{0,\ldots,d-1\}}$ be the standard orthonormal basis of $\mathbb{R}^d$, and $(e_i)_{i\in\mathbb{N}}$ its extension to a periodic sequence modulo d. For each node $a \in T$ let $b_a \in \{0,1\}$ be a random symmetric i.i.d bit. Define for all $t \in [D]$, $i > 0$ $f_i^{(t)} : H \to L_1^d$ as

$$f_i^{(t)}(x) = \theta^{-i} b_{x_i} e_i \ ,$$

and define $f^{(t)} : H \to L_1^d$ as $f^{(t)}(x) = \sum_{i=1}^{\infty} f_i^{(t)}(x)$. Finally define $f : H \to L_1^{\bar{D}}$ by

$$f(x) = \bigoplus_{t=1}^{D} f^{(t)}(x) \ .$$

Fix some $x, y \in H$ and $t \in [D]$. For any $j \in \{0,\ldots,d-1\}$ let $Z_j^{(t)} = |(f^{(t)}(x) - f^{(t)}(y))_j|^p$. Let $i_j = \min\{i > \mathrm{dlca}(x,y) \mid i = j \mod (d)\}$ and let $I_j = \{i \ge i_j \mid i = j \mod d\}$. Note that since $x_i = y_i$ for any $i \le \mathrm{dlca}(x,y)$, we have that $Z_j^{(t)} = |\sum_{i\in I_j}(f_i^{(t)}(x) - f_i^{(t)}(y))_i|^p$. Then we have the following

$$0 \le Z_j^{(t)} \le \left(\sum_{i\in I_j} \theta^{-i}\right)^p = \left(\theta^{-i_j}\sum_{i=0}^{\infty}\theta^{-id}\right)^p = \left(\frac{\theta^{-i_j}}{1-\theta^{-d}}\right)^p \qquad (1)$$

Claim 13. *For any $j \in \{0,\ldots,d-1\}$ and $t \in [D]$, $\Pr[Z_j^{(t)} \ge \theta^{-i_j p}] \ge 1/16$. Moreover, this bound depends only on random variables b_{x_i}, b_{y_i} where $i \in \{i_j, i_j + d\}$.*

Proof. There is probability of $1/16$ that the random bits $b_{x_{i_j}} = 1$, $b_{y_{i_j}} = 0$, $b_{x_{i_j+d}} = 1$ and $b_{y_{i_j+d}} = 0$. In such a case $(f_{i_j}^{(t)}(x) - f_{i_j}^{(t)}(y))_{i_j} = \theta^{-i_j}$ and $(f_{i_j+d}^{(t)}(x) - f_{i_j+d}^{(t)}(y))_{i_j+d} = \theta^{-i_j-d}$. Note that $\frac{\theta^{-d}}{1-\theta^{-d}} = \frac{1}{e^{\alpha}-1} \le 1$ (because $\alpha \ge \ln 2$), which suggests that

$$\Big|\sum_{i\in I_j\setminus\{i_j,i_j+d\}}(f_i^{(t)}(x) - f_i^{(t)}(y))_i\Big| \le \theta^{-i_j-2d}\sum_{i=0}^{\infty}\theta^{-di} = \frac{\theta^{-i_j-2d}}{1-\theta^{-d}} \le \theta^{-i_j-d}.$$

It follows that with probability at least $1/16$

$$\begin{aligned} Z_j^{(t)} &= \left|\sum_{i\in I_j}(f_i^{(t)}(x) - f_i^{(t)}(y))_i\right|^p \\ &\ge \left|\theta^{-i_j} + \theta^{-i_j-d} - \Big|\sum_{i\in I_j\setminus\{i_j,i_j+d\}}(f_i^{(t)}(x) - f_i^{(t)}(y))_i\Big|\right|^p \\ &\ge \theta^{-i_j p} \end{aligned}$$

□

For any $1 \le t \le D$ let $Z^{(t)}(x,y) = Z^{(t)} = \sum_{j=0}^{d-1} Z_j^{(t)} = \|f^{(t)}(x) - f^{(t)}(y)\|_p^p$. Also let $Z(x,y) = Z = \sum_{t=1}^{D} Z^{(t)}(x,y)$ and $\mu = \mathbb{E}[Z]$. Note that $\mu/d(x,y)^p$ is a constant independent of $d(x,y)$, because we can write

$$\begin{aligned}
\mu &= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}[Z_j^{(t)}] \\
&= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\sum_{i\in I_j}(f_i^{(t)}(x) - f_i^{(t)}(y))_i\right|^p\right] \\
&= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\sum_{i\in I_j}\theta^{-i}(b_{x_i} - b_{y_i})\right|^p\right] \\
&= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\theta^{-\mathrm{dlca}(x,y)-1}\sum_{i=0}^{\infty}\theta^{-id-j}(b_{x_{\mathrm{dlca}(x,y)+id+j}} - b_{y_{\mathrm{dlca}(x,y)+id+j}})\right|^p\right] \\
&= d(x,y)^p\theta^{-p}\cdot\sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\sum_{i=0}^{\infty}\theta^{-id-j}(b_{x_i} - b_{y_i})\right|^p\right]
\end{aligned}$$

The last equality holds as for calculating the expectation over the random bits it does not matter from which level of the tree they are taken - they are all independent with the same distribution. We conclude that we can scale the embedding f by this constant. It follows that it is enough to prove that there exists a choice of the random bits such that $|Z^{1/p} - \mu^{1/p}| < \epsilon\mu^{1/p}$ for all pairs, then the embedding will have distortion $1 + O(\epsilon)$. Now the analysis is different for various values of p, first we prove for the case that $p \le 4/\epsilon$, note that in this case $D \ge (c/4)\ln m \cdot \alpha/(\epsilon^2 p)$. Let $M_j = \left(\frac{\theta^{-i_j}}{1-\theta^{-d}}\right)^p \le e\cdot\theta^{-i_j p}$, using that $(1-\theta^{-d})^p \ge (1-1/(2p))^p \ge e^{-p/p} = 1/e$, and note that (1) suggests that $0 \le Z_j^{(t)} \le M_j$.

By linearity of expectation and by Claim 13 it follows that

$$\begin{aligned}
\mu &\ge \frac{D}{16}\sum_{j=0}^{d-1}\theta^{-i_j p} \\
&= \frac{D\cdot\theta^{-p(\mathrm{dlca}(x,y)+1)}}{16}\sum_{i=0}^{d-1}\theta^{-ip} \\
&= \frac{D\cdot\theta^{-p}\cdot d(x,y)^p}{16}\cdot\frac{1-\theta^{-dp}}{1-\theta^{-p}} \\
&\ge \frac{D\cdot d(x,y)^p}{2000\epsilon p},
\end{aligned}$$

where in the last inequality we used that $1-\theta^{-dp} \ge 1 - e^{-p} \ge 1 - 1/e \ge 1/2$, that $1-\theta^{-p} = 1 - e^{-\epsilon p} \le \epsilon p$ and that $\theta^{-p} \ge 1/e^4$.

Now we want to show that with high enough probability it will be the case that $|Z - \mu| < \epsilon p\mu$, and since $\epsilon p \le e^{\epsilon p} - 1 \le (1+2\epsilon)^p - 1$ (using that $e^\epsilon \le 1 + 2\epsilon$ for $0 < \epsilon \le 1/2$) it will imply that $(1-2\epsilon)\mu^{1/p} < Z^{1/p} < (1+2\epsilon)\mu^{1/p}$ which gives the desired distortion $1 + O(\epsilon)$. By

Hoeffding's inequality (Lemma 14) with $r = \frac{\epsilon p\mu}{dD} \geq \frac{d(x,y)^p}{2000d}$ and using that $M = \sum_{t=1}^{D}\sum_{j=1}^{d} M_j^2 \leq e^2 D \sum_{j=0}^{d-1} \theta^{-2i_j p} \leq \frac{e^2 D \cdot \theta^{-2(\mathrm{dlca}(x,y)+1)p} \cdot (1-\theta^{-2pd})}{1-\theta^{-2p}} \leq \frac{e^2 D \cdot d(x,y)^{2p}}{e^{2\epsilon p}-1} \leq \frac{4D \cdot d(x,y)^{2p}}{\epsilon p}$,

$$\Pr[|Z-\mu| \geq \epsilon p\mu] = \Pr[|Z-\mu| \geq r \cdot dD] \leq 2e^{-2(dDr)^2/M} \leq 2e^{-\epsilon p D/8000000} \leq e^{-10\ln m \cdot \alpha/\epsilon},$$

for a large enough constant c.

Consider the case $p > 4/\epsilon$, where $D = c\ln m \cdot \alpha/\epsilon$. First observe that $1-\theta^{-d} = 1-\epsilon/2$ and $\theta^{-p} \leq (1-\epsilon+\epsilon^2/2)^p \leq (1-3\epsilon/4)^p$ (the last inequality holds since $\epsilon \leq 1/2$) hence $\frac{\theta^{-p}}{(1-\theta^{-d})^p} \leq \frac{(1-3\epsilon/4)^p}{(1-\epsilon/2)^p} \leq (1-\epsilon/4)^p \leq 1/e$. Also note that $1-\theta^{-p} \geq 1-e^{-4}$. Assume w.l.o.g that i_0 is the minimal among $i_0, i_1, \ldots, i_{d-1}$, then by (1) for any $0 \leq \ell \leq d-1$,

$$\begin{aligned} \sum_{j=\ell}^{d-1} Z_j^{(t)} &\leq \frac{1}{(1-\theta^{-d})^p}\sum_{j=\ell}^{d}\theta^{-i_j p} \qquad (2)\\ &= \frac{\theta^{-(i_\ell-1)p}\cdot\theta^{-p}}{(1-\theta^{-d})^p}\cdot\frac{1-\theta^{-dp}}{1-\theta^{-p}}\\ &\leq \theta^{-(i_\ell-1)p}\cdot\frac{1/e}{1-1/e^4}\\ &\leq \theta^{-(i_\ell-1)p}/2, \end{aligned}$$

In particular for $\ell = 0$ we have that for any $1 \leq t \leq D$

$$Z^{(t)} \leq \theta^{-(i_0-1)p}/2 = \theta^{-\mathrm{dlca}(x,y)\cdot p}/2 = d(x,y)^p/2,$$

hence with probability 1

$$Z = \sum_{t=1}^{D} Z^{(t)} \leq D \cdot d(x,y)^p \ .$$

By Claim 13 with probability at least $1/16$ we have that $Z_0^{(t)} \geq \theta^{-i_0 p}$, and (2) suggests that $\sum_{j=1}^{d-1} Z_j^{(t)} \leq \theta^{-i_0 p}/2$. It follows that with probability $1/16$

$$Z^{(t)} \geq \theta^{-i_0 p} - |\sum_{j=1}^{d} Z_j^{(t)}| \geq \theta^{-i_0 p}/2 \geq (1-\epsilon)^p \cdot d(x,y)^p/2 \ . \qquad (3)$$

Let K_t be an indicator random variable for the event that $Z^{(t)} \geq \theta^{-i_0 p}/2$, and $K = \sum_{t=1}^{D} K_t$. Note that $\mathbb{E}[K] \geq D/16$, denote $\mathbb{E}[K] = \gamma D/16$ for some $\gamma \geq 1$, then by Chernoff bound

$$\Pr[K < D/32] \leq \Pr[\mathbb{E}[K] - K > \gamma D/32] \leq e^{-\gamma D/8} \leq e^{-D/8} = e^{-10\ln m\cdot\alpha/\epsilon}$$

for c large enough. So with probability at least $1-e^{-10\ln m\cdot\alpha/\epsilon}$ we have that at least $1/32$ fraction of the $Z^{(t)}$ are lower bounded as in (3), hence

$$Z = \sum_{t=1}^{D} Z^{(t)} \geq D/32 \cdot (1-\epsilon)^p \cdot d(x,y)^p/2 = D(1-\epsilon)^p \cdot d(x,y)^p/64 \ ,$$

as required, since the distortion is $\left(\frac{64}{(1-\epsilon)^p}\right)^{1/p} = 1 + O(\epsilon)$.

Finally we need to argue that there is some probabilities for all pairs to have the desired distortion. Define equivalence relation on unordered pairs of branches such that $\{x, y\} \sim \{x', y'\}$ iff $x_{a+2d} = x'_{a+2d}$, $y_{a+2d} = y'_{a+2d}$ where $a = \text{dlca}(x, y) + 1$. Denote by $[x, y]$ be the equivalence class of $\sim$ that contains the pair x, y. Claim 13 implies that for all the pairs in $[x, y]$ the success event (in either of the above cases) for each one of the pairs is defined by exactly the same random variables - the first $2d$ nodes on the branches x, y just after $\text{lca}(x, y)$. Moreover, all the pairs require exactly the same events to occur. Let $Y_{[x,y]}$ be an indicator variable for the event that $|Z(x, y) - \mathbb{E}[Z(x, y)]| < (1 + \epsilon)^p \mathbb{E}[Z(x, y)]$. As noted above, if this holds it will also mean that $|Z(x', y') - \mathbb{E}[Z(x', y')]| < (1 + \epsilon)^p \mathbb{E}[Z(x', y')]$ for any $(x, y) \in [x, y]$.

By Claim 13, the success of the event $Y_{[x,y]}$ depends only on the choice of random bits for the first $2d$ levels of the tree after $\text{lca}(x, y)$. It follows that events $Y_{[x,y]}$ and $Y_{[x',y']}$ depend on each other iff $\text{lca}(x, y)$ and $\text{lca}(x', y')$ are on the same branch in T and their tree distance is at most $2d$. Since the out-degree of T is bounded by m, for each $u \in T$ there are at most m^{4d} different equivalence classes $[x, y]$ for which the node u is $\text{lca}(x, y)$. In addition there are at most $m^{2d} + 2d$ other possible nodes $u' \in T$ at tree distance at most $2d$ from u such that both u and u' are on the same branch of T. It follows that the number of dependencies for each event $Y_{[x,y]}$ is at most $m^{8d} \leq e^{9 \ln m \cdot \alpha/\epsilon}$.

We conclude that number of dependencies is smaller than four times the inverse success probability of the "bad" events that we bounded above, hence according to the Local Lemma (Lemma 15) there is some positive probability that none of the bad events $\{|Z(x, y)^{1/p} - \mu^{1/p}| \geq \epsilon\mu^{1/p}\}$ occur for any $x, y \in X$.

□

Proof of Theorem 2. Fix any $p \geq 1$ and $0 < \epsilon \leq 1$. Let $\hat{\epsilon} = \epsilon/4$ and $\theta = e^{\hat{\epsilon}}$. Using Lemma 10 embed the ultrametric (X, d) into a $\theta - e$HST T' with distortion θ, and using Lemma 11 embed T' into a $\theta - e$HST T such that the degree of any internal node of T is at most $m = 2k^2 + 1$, with k-local distortion 1. Finally using Lemma 12 with parameter $\hat{\epsilon} < 1/2$, embed T into D dimensional L_p space with distortion $1 + \hat{\epsilon}$, where $D = O((\log k)/\epsilon^3)$. The total k-local distortion is at most $\theta(1 + \hat{\epsilon}) \leq 1 + \epsilon$.

□

References

[ABN06] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 271–286, New York, NY, USA, 2006. ACM Press.

[ABN07] Ittai Abraham, Yair Bartal, and Ofer Neiman. Local embeddings of metric spaces. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 631–640, New York, NY, USA, 2007. ACM.

[Bar96] 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.

[BM04] Yair Bartal and Manor Mendel. Dimension reduction for ultrametrics. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 664–665. Society for Industrial and Applied Mathematics, 2004.

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[CKR01] Gruia Calinescu, Howard J. Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Symposium on Discrete Algorithms*, pages 8–16, 2001.

[FRT03] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 448–455. ACM Press, 2003.

[JL84] 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.

[Mat97] J. Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[Pel00] David Peleg. *Distributed Computing: A Locality-Sensitive Approach.* SIAM Monographs on Discrete Mathematics and Applications, 2000.

[SS09] Gideon Schechtman and Adi Shraibman. Lower bounds for local versions of dimension reductions. *Discrete Comput. Geom.*, 41(2):273–283, 2009.

[TZ05] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24, 2005.

A The "small" distances

In this section we prove the properties of the map f that shows a lower bound for the small distances in Theorem 1. It is a local version of Bourgain's embedding method [Bou85], with Matoušek's modifications for large p [Mat97]

Let $s = e^p$, $t = \lceil \log_s k \rceil$, $q = cs \ln k$ for some constant c to be determined later, $D = t \cdot q$, $T = \{i \mid 1 \le i \le t\}$ and $Q = \{j \mid 1 \le j \le q\}$. Choose random subsets A_{ij} for every $i \in T$, $j \in Q$, such that each point is independently included in A_{ij} with probability s^{-i}. We now define the embedding $f : X \to L_p^D$ by defining for each $i \in T$, $j \in Q$ a function $f_{i,j} : X \to \mathbb{R}_+$ by $f_{i,j}(u) = d(u, A_{ij})$, and

$$f(u) = D^{-1/p} \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{q} f_{i,j}(u)$$

Let $u, v \in X$ be such that $v \in B(u, r_k(u)/24)$. For all $i \in T$ let $r'_{s^i} = \max\{r_{s^i}(u), r_{s^i}(v)\}$, let $w_i \in \{u, v\}$ be the point obtaining the maximum and $z_i \in \{u, v\}$ the other point. Let $t' \in T$ be the minimal such that $r'_{s^{t'}} \ge d(u,v)/2$ and let $r_{s^i} = \min\{r'_{s^i}, d(u,v)/2\}$ for all $i \le t'$. Set $\delta_i = r_{s^i} - r_{s^{i-1}}$. Since $d(u,v) \le r_k(u) \le r'_{s^t}$ it follows that $\sum_{i=1}^{t'} \delta_i = r_{s^{t'}} = d(u,v)/2$.

For any $j \in Q$, and $i \le t'$ let $G(u,v,i,j)$ be the event that $A_{ij} \cap B(w_i, r_{s^i}) = \emptyset$ and $A_{ij} \cap B(z_i, r_{s^{i-1}}) \neq \emptyset$. In such a case $|f_{i,j}(u) - f_{i,j}(v)| \ge r_{s^i} - r_{s^{i-1}} = \delta_i$. By standard arguments it can be shown that $\Pr[G(u,v,i,j)] \ge 1/(8s)$. Let $G(u,v)$ be the event that for all $i \le t'$ there exists $Q'(i) \subseteq Q$ of cardinality $|Q'(i)| \ge q/(16s)$ such that for all $j \in Q'(i)$ event $G(u,v,i,j)$ holds. First we show that if $G(u,v)$ holds then the distortion of the pair u,v is small, the upper bound:

$$\|f(u) - f(v)\|_p^p \le D^{-1} \sum_{i \in T} \sum_{j \in Q} d(u,v)^p \le d(u,v),$$

and lower bound:

$$\begin{aligned}
\|f(u) - f(v)\|_p^p &= D^{-1} \sum_{i \in T} \sum_{j \in Q} |f_{i,j}(u) - f_{i,j}(v)|^p \\
&\ge D^{-1} \sum_{i \le t'} \sum_{j \in Q'(i)} |f_{i,j}(u) - f_{i,j}(v)|^p \\
&\ge D^{-1} \frac{q}{16s} \sum_{i \le t'} \delta_i^p \\
&\ge D^{-1} \frac{q}{16s \cdot t^{p-1}} \left(\sum_{i \le t'} \delta_i \right)^p \\
&\ge \frac{(d(u,v)/2)^p}{16s \cdot t^p}.
\end{aligned}$$

Hence $\|f(u) - f(v)\|_p \ge \Omega(d(u,v)/t) = \Omega(p \cdot d(u,v)/\log k)$.

Define a dependency graph on the events where two events $G(u,v)$, $G(u',v')$ are connected by an edge iff $u' \in \bar{N}_k(u)$ (note that this is symmetric relation), the degree of the graph is at most k^2. Notice that event $G(u,v)$ depends only on the choice of points in the ball $B(u, 2d(u,v))$. Assume that events $G(u,v)$ and $G(u',v')$ are not connected by an edge, *i.e.* $u \notin N_k(u')$ or $u' \notin N_k(u)$. Since by the assumption $2d(u,v) \le r_k(u)/12$ and also $2d(u',v') \le r_k(u')/12 \le (d(u,u') + r_k(u))/12$, it follows that if $d(u,u') \ge r_k(u)$ then $d(u,u') - 2d(u,v) - 2d(u',v') \ge 11d(u,u')/12 - r_k(u)/6 > 0$ (there is a symmetric calculation for the case that $d(u,u') \ge r_k(u')$), hence the balls $B(u, 2d(u,v))$ and $B(u', 2d(u',v'))$ are disjoint.

Let $G(u,v,i) = \sum_j G(u,v,i,j)$, then $\mathbb{E}[G(u,v,i)] \ge q/(8s)$ hence by Chernoff bound

$$\Pr[G(u,v,i) \le q/(16s)] \le e^{-q/(64s)} \le k^{-4},$$

for a large enough constant c, so

$$\Pr[\neg G(u,v)] = \Pr[\exists i \le t', G(u,v,i) \le q/(16s)] \le t \cdot k^{-4} \le k^{-3}.$$

Now by Lemma 15 there is some positive probability that all the good events $G(u,v)$ hold simultaneously.

B Some Basic Tools

Lemma 14 (Hoeffding)**.** *Let Z_i be independent random variables for $i = 1, \ldots, d$, let $\mathbb{E}[Z_i] = \mu_i$ and $0 \le Z_i \le M_i$. Let $Z = \sum_{i=1}^d Z_i$, $\mu = \sum_{i=1}^d \mu_i$ and $M = \sum_{i=1}^d M_i^2$. Then for $r > 0$*

$$\Pr[|Z - \mu| \ge rd] \le 2e^{-2(rd)^2/M}.$$

Lemma 15 (Local Lemma)**.** *Let $\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_n$ be events in some probability space. Let $G(V, E)$ be a graph on n vertices with degree at most d, each vertex corresponding to an event. Assume that for any $i = 1, \ldots, n$*

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

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

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

Lemma 16 ([JL84])**.** *For any $\epsilon > 0$ and integer $m > 1$, there exist $x_1, \ldots x_m \in L_p^D$ where $D = O((\log m)/\epsilon^2)$, such that for any $1 \leq i < j \leq m$:*

$$1 - \epsilon/3 \leq \|x_i - x_j\|_p \leq 1 + \epsilon/3 \ .$$

C Proof of Lemma 5

Let $\eta = 2^{-11} \log(1/\delta)/\log k$. Define the partition P of X into clusters by generating a sequence of clusters: $C_1, C_2, \ldots C_s$, for some fixed $s \in [n]$. Notice that we are generating a distribution over partitions and therefore the generated clusters are random variables. First we deterministically assign centers $v_1, v_2, \ldots, v_s$ by the following iterative process: Let $W_1 = X$ and $j = 1$.

1. Let $v_j \in W_j$ be the point maximizing $r_k(x)$ over all $x \in W_j$.
2. Let $W_{j+1} = W_j \setminus B(v_j, r_k(v_j)/256)$.
3. Set $j = j + 1$. If $W_j \neq \emptyset$ return to 1.

Now the algorithm for the partition is as follows: Let $Z_1 = X$. For $j = 1, 2, 3 \ldots s$:

- Let $C_j = B_{Z_j}(v_j, r)$ and $Z_{j+1} = Z_j \backslash S_{v_j}$ where r is chosen according to a truncated exponential distribution with parameter $\lambda = 8(\ln k)/\Delta$ where $\Delta = r_k(v_j)/64$, *i.e.*

$$f(r; \lambda) = \begin{cases} \frac{k^2}{1-k^{-2}} \lambda e^{-\lambda r} & r \in [\Delta/4, \Delta/2] \\ 0 & \text{otherwise} \end{cases}$$

Observe that some clusters may be empty, it is not necessarily the case that $v_m \in C_m$, and every cluster contains at most k points.

Let $x \in X$ be in cluster C with center v, then we have the following

Claim 17. $r_k(v) \leq 2 r_k(x)$.

Proof. First note that since $d(x, v) \leq r_k(v)/128$ it follows that $|B(x, r_k(v)/2)| \leq |B(v, (1/2 + 1/128) r_k(v))| < k$, hence $r_k(x) \geq r_k(v)/2$. Now $d(x, v) \leq r_k(v)/128 \leq r_k(x)/2$, hence

$$r_k(v) \leq d(v, x) + r_k(x) \leq r_k(x)/2 + r_k(x) \leq 2 r_k(x) \ .$$

□

Now we are ready to show the first property, that if $y \in X$ is such that $d(x,y) \geq r_k(x)/24$ then $y \notin C$: As $r_k(x) \leq r_k(v)$ and $C \subseteq B(v, r_k(v)/128)$ we get that $d(v,y) \geq d(y,x) - d(x,v) \geq r_k(x)/24 - r_k(v)/128 > r_k(v)/64 - r_k(v)/128 = r_k(v)/128$ (using Claim 17). It follows that $y \notin C$.

Next we will prove the locality of the second property of the partition. For any $x \in X$ let $T_x = B(x, r_k(x)/32)$, and note that any center v that can cut the ball of radius $\eta \cdot r_k(x)$ around x must have $v \in T_x$. Let v be such a center. Since the choice of radius is the only randomness in the process of creating P, the event of padding for $x \in X$ is determined by the choice of radiuses for centers $v_j \in T_x$. Let $z \notin \bar{N}_k(x)$ and we will show that any center that can cut the ball around z will not be in T_x. There are two possibilities: either $z \notin N_k(x)$ and hence $d(x,z) \geq r_k(x)$, or that $x \notin N_k(z)$, therefore $d(x,z) \geq r_k(x)/2$ (assume by contradiction that it is not so, then $B(z, d(x,z)) \subseteq B(x, 2d(x,z)) \subsetneq B(x, r_k(x))$, so $x \in N_k(z)$). Now assume by contradiction that the center v can cut $B(z, \eta \cdot r_k(z))$, *i.e.* that $v \in T_z$ as well. By Claim 17 $r_k(v) \leq 2r_k(x)$, then since $r_k(z) \leq d(z,v) + r_k(v) \leq r_k(z)/4 + 2r_k(x)$ we get that $r_k(z) < 3r_k(x)$. Now $d(x,z) \leq d(x,v) + d(z,v) \leq r_k(x)/8 + r_k(z)/8 < r_k(x)/2$, contradiction.

We conclude by proving the bound on the padding probability. Consider the distribution over the clusters $C_1, C_2, \ldots C_s$ as defined above. For $1 \leq m \leq s$, define the events:

$$\begin{aligned}
\mathcal{Z}_m &= \{\forall j, 1 \leq j < m,\ B(x, \eta \cdot r_k(x)) \subseteq Z_{j+1}\}, \\
\mathcal{E}_m &= \{\exists j,\ m \leq j < s \text{ s.t. } B(x, \eta \cdot r_k(x)) \bowtie (S_{v_j}, \bar{S}_{v_j}) | \mathcal{Z}_m\}.
\end{aligned}$$

Also let $T = T_x$ and $\theta = \sqrt{\delta}$. We prove the following inductive claim: For every $1 \leq m \leq s$:

$$\Pr[\mathcal{E}_m] \leq (1-\theta)(1 + \theta \sum_{j \geq m, v_j \in T} k^{-1}). \tag{4}$$

Note that $\Pr[\mathcal{E}_s] = 0$. Assume the claim holds for $m+1$ and we will prove for m. Define the events:

$$\begin{aligned}
\mathcal{F}_m &= \{B(x, \eta \cdot r_k(x)) \bowtie (S_{v_m}, \bar{S}_{v_m}) | \mathcal{Z}_m\}, \\
\mathcal{G}_m &= \{B(x, \eta \cdot r_k(x)) \subseteq \bar{S}_{v_m} | \mathcal{Z}_m\} = \{\mathcal{Z}_{m+1} | \mathcal{Z}_m\}, \\
\bar{\mathcal{G}}_m &= \{B(x, \eta \cdot r_k(x)) \nsubseteq \bar{S}_{v_m} | \mathcal{Z}_m\} = \{\bar{\mathcal{Z}}_{m+1} | \mathcal{Z}_m\}.
\end{aligned}$$

First we bound $\Pr[\mathcal{F}_m]$. Recall that the center v_m of C_m is determined deterministically. The radius r_m is chosen from the interval $[r_k(v_m)/256, r_k(v_m)/128]$. We claim that if $B(x, \eta \cdot r_k(x)) \bowtie (S_{v_m}, \bar{S}_{v_m})$ then $v_m \in T$. First observe that $\eta \cdot r_k(x) \leq r_k(x)/128$, therefore $d(v_m, x) \leq (r_k(v_m) + r_k(x))/128$. Note that $r_k(v_m) \leq d(v_m, x) + r_k(x) \leq (r_k(v_m) + r_k(x))/128 + r_k(x)$, hence $r_k(v_m) \leq 2r_k(x)$, which imply that $d(v_m, x) \leq (r_k(v_m) + r_k(x))/128 \leq r_k(x)/32$. Therefore if $v_m \notin T$ then $\Pr[\mathcal{F}_m] = 0$. Otherwise, using the maximality of $r_k(v_m)$ we get that $\eta \cdot r_k(x) \leq \eta \cdot r_k(v_m) = \frac{1}{16} \ln(1/\theta)/\ln k \cdot \Delta$, then by Lemma 18

$$\begin{aligned}
&\Pr[\mathcal{F}_m] \\
&= \Pr[B(x, \eta \cdot r_k(x)) \bowtie (S_{v_m}, \bar{S}_{v_m}) | \mathcal{Z}_m] \\
&\leq (1-\theta)(\Pr[B(x, \eta \cdot r_k(x)) \nsubseteq \bar{S}_{v_m} | \mathcal{Z}_m] + \theta k^{-1}) \\
&= (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta k^{-1}).
\end{aligned} \tag{5}$$

Using the induction hypothesis we prove the inductive claim:

$$\begin{aligned}
\Pr[\mathcal{E}_m] &\leq \Pr[\mathcal{F}_m] + \Pr[\mathcal{G}_m]\Pr[\mathcal{E}_{m+1}] \\
&\leq (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta \mathbf{1}_{\{v_m \in T\}} k^{-1}) + \\
&\quad \Pr[\mathcal{G}_m] \cdot (1-\theta)(1+\theta \sum_{j \geq m+1, v_j \in T} k^{-1}) \\
&\leq (1-\theta)(1+\theta \sum_{j \geq m, v_j \in T} k^{-1}),
\end{aligned}$$

The second inequality follows from 5 and the induction hypothesis. Note that for any $x \in X$, $|T_x| \leq k$ we get that $\sum_{j \geq 1, v_j \in T_x} k^{-1} \leq 1$. We conclude from the claim (4) for $m = 1$ that:

$$\Pr[B(x, \eta \cdot r_k(x)) \nsubseteq P(x)] \leq \Pr[\mathcal{E}_1] \leq$$
$$(1-\theta)(1+\theta \cdot \sum_{j \geq 1, v_j \in T} k^{-1}) \leq (1-\theta)(1+\theta) = \delta.$$

The following was shown in [ABN06]

Lemma 18 (Probabilistic Decomposition)**.** *Let (X, d) be a metric space and $Z \subseteq X$. let $\chi \geq 2$ be a parameter. Given $0 < \Delta < \mathrm{diam}(Z)$ and a center point $v \in Z$, there exists a probability distribution over partitions $(S, \bar{S})$ of Z such that $S = B_Z(v, r)$, and r is chosen from a probability distribution in the interval $[\Delta/4, \Delta/2]$, such that for any $\theta \in (0,1)$ satisfying $\theta \geq \chi^{-1}$, let $\eta = \frac{1}{16}\ln(1/\theta)/\ln\chi$ then for any $x \in Z$, the following holds:*

$$\Pr[B_Z(x, \eta\Delta) \bowtie (S, \bar{S})] \leq$$
$$(1-\theta)\left[\Pr[B_Z(x, \eta\Delta) \nsubseteq \bar{S}] + 2\chi^{-2}\right].$$