

# Volume in General Metric Spaces

Ittai Abraham\*    Yair Bartal<sup>†</sup>    Ofer Neiman<sup>‡</sup>    Leonard J. Schulman<sup>§</sup>

April 23, 2014

## Abstract

A central question in the geometry of finite metric spaces is how well can an arbitrary metric space be “faithfully preserved” by a mapping into Euclidean space. In this paper<sup>1</sup> we present an algorithmic embedding which obtains a new strong measure of faithful preservation: not only does it (approximately) preserve distances between pairs of points, but also the volume of any set of  $k$  points. Such embeddings are known as volume preserving embeddings. We provide the first volume preserving embedding that obtains *constant* average volume distortion for sets of any fixed size. Moreover, our embedding provides constant bounds on all bounded moments of the volume distortion while maintaining the best possible worst-case volume distortion.

Feige, in his seminal work on volume preserving embeddings defined the volume of a set  $S = \{v_1, \dots, v_k\}$  of points in a general metric space: the product of the distances from  $v_i$  to  $\{v_1, \dots, v_{i-1}\}$ , normalized by  $\frac{1}{(k-1)!}$ , where the ordering of the points is that given by Prim’s minimum spanning tree algorithm. Feige also related this notion to the maximal Euclidean volume that a Lipschitz embedding of  $S$  into Euclidean space can achieve. Syntactically this definition is similar to the computation of volume in Euclidean spaces, which however is invariant to the order in which the points are taken. We show that a similar robustness property holds for Feige’s definition: the use of any other order in the product affects volume <sup>$1/(k-1)$</sup>  by only a constant factor. Our robustness result is of independent interest as it presents a new competitive analysis for the greedy algorithm on a variant of the online Steiner tree problem where the cost of buying an edge is logarithmic in its length. This robustness property allows us to obtain our results on volume preserving embedding.

## 1 Introduction

Recent years have seen a large outpouring of work in analysis, geometry and theoretical computer science on metric space embeddings guaranteed to introduce only small distortion into the distances between pairs of points.

Euclidean space is not only a metric space, it is also equipped with higher dimensional volumes. General metrics do not carry such structure. However, a general definition for the volume of a set of points in an arbitrary metric was developed by Feige [Fei00].

---

\*Microsoft Research. Email: ittaia@microsoft.com.

<sup>†</sup>Hebrew University. Email: yair@cs.huji.ac.il. Part of the research was done while the author was at the Center of the Mathematics of Information, Caltech, CA, USA, and was supported in part by a grant from the Israeli Science Foundation (1609/11) and in part by a grant from the National Science Foundation (NSF CCF-0652536).

<sup>‡</sup>Ben-Gurion University of the Negev. E-mail: neimano@cs.bgu.ac.il. Supported in part by ISF grant No. (523/12) and by the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement  $n^\circ$ 303809.

<sup>§</sup>Caltech. Email: schulman@caltech.edu. Supported in part by NSF CCF-0515342 and NSA H98230-06-1-0074.

<sup>1</sup>This paper is a full version based on the conference paper [ABNS10].

In this paper we extend the study of metric embeddings into Euclidean space by first, showing a robustness property of the general volume definition. Using this robustness property, together with existing metric embedding methods, to show an embedding that guarantees small distortion not only on pairs, but also on the volumes of sets of points. The robustness property (see [Theorem 2](#)) is that the minimization over permutations in the volume definition affects it by only a constant. This result is of independent interest as it provides an analysis for the greedy algorithm on a variant of the online Steiner tree problem, where the cost of buying an edge is logarithmic in its length. We show that the greedy algorithm has a constant competitive ratio to the optimum. Our main application of [Theorem 2](#) is an algorithmic embedding (see [Theorem 3](#)) with *constant* average distortion for sets of any fixed size. In fact, our bound on the average distortion scales logarithmically with the size of the set. Moreover this bound holds even for higher moments of the distortion (the  $\ell_q$ -distortion), while the embedding still maintains the best possible worst case distortion bound, simultaneously. Hence our embedding generalizes both [[KLMN04](#)] and [[ABN11](#)] (see [Related Work](#) below).

### Volume in general metric spaces.

Let  $d_E$  denote Euclidean distance, and let  $\text{affspan}$  denote the affine span of a point set. The  $(n - 1)$ -dimensional Euclidean volume of the convex hull of points  $X = \{v_1, \dots, v_n\} \subseteq \mathbb{R}^d$  is

$$\phi_E(X) = \frac{1}{(n - 1)!} \prod_{i=2}^n d_E(v_i, \text{affspan}(v_1, \dots, v_{i-1})).$$

This definition is, of course, independent of the order of the points.

**Feige's notion of volume.** Let  $(X, d_X)$  be a finite metric space,  $X = \{v_1, \dots, v_n\}$ . Let  $S_n$  be the symmetric group on  $n$  symbols, and let  $\pi_P \in S_n$  be an order in which the points of  $X$  may be adjoined to a minimum spanning tree by Prim's algorithm. (Thus  $v_{\pi_P(1)}$  is an arbitrary point,  $v_{\pi_P(2)}$  is the closest point to it, etc.) Feige's notion of the volume of  $X$  is (we have normalized by a factor of  $(n - 1)!$ ):

$$\phi_F(X) = \frac{1}{(n - 1)!} \prod_{i=2}^n d_X(v_{\pi_P(i)}, \{v_{\pi_P(1)}, \dots, v_{\pi_P(i-1)}\}). \quad (1)$$

$\pi_P$  minimizes the above expression (1) (see [Section 2](#)).

It should be noted that even if  $X$  is a subset of Euclidean space,  $\phi_E$  and  $\phi_F$  do not agree. (The latter can be arbitrarily larger than the former, for instance, a very thin triangle.) The actual relationship that Feige found between these notions is nontrivial. Let  $\mathcal{L}_2(X)$  be the set of non-expansive embeddings from  $X$  into Euclidean space. Feige proved the following:

**Theorem 1** (Feige). *For any  $n$  point metric space  $(X, d)$ :*

$$1 \leq \left[ \frac{\phi_F(X)}{\sup_{f \in \mathcal{L}_2(X)} \phi_E(f(X))} \right]^{1/(n-1)} \leq 2.$$

Thus, remarkably,  $\phi_F(X)$  is characterized to within a factor of 2 (after normalizing for dimension) by the Euclidean embeddings of  $X$ .

**Our work, part I: Robustness of the metric volume.** What we show first is that Feige’s definition is insensitive to the minimization over permutations implicit in Equation (1), and so also a generalized version of [Theorem 1](#) can be obtained.

**Theorem 2.** *There is a constant  $C$  such that for any  $n$ -point metric space  $(X, d)$ , and with  $\pi_P$  defined as above, and for every  $\pi \in S_n$ :*

$$1 \leq \left( \frac{\prod_{i=2}^n d_X(v_{\pi(i)}, \{v_{\pi(1)}, \dots, v_{\pi(i-1)}\})}{\prod_{i=2}^n d_X(v_{\pi_P(i)}, \{v_{\pi_P(1)}, \dots, v_{\pi_P(i-1)}\})} \right)^{1/(n-1)} \leq C.$$

An alternative interpretation of this result can be presented as the analysis of the following online problem. Consider the following variant of the online metric Steiner tree problem [[IW91](#)]. Given a complete weighted graph  $(V, E)$ , at each time unit  $i$ , the adversary outputs a vertex  $v_i \in V$  and an online algorithm can buy edges  $E_i \subseteq E$ . At each time unit  $i$ , the edges bought  $E_1, \dots, E_i$  must induce a connected graph among the current set of vertices  $v_1, \dots, v_i$ . The competitive ratio of an online algorithm is the worst ratio between the cost of the edges bought and the cost of the edges bought by the optimal offline algorithm. This problem has been well-studied when the cost of buying an edge is proportional to its length. Imase and Waxman prove that the greedy algorithm is  $O(\log n)$  competitive, and shown that this bound is asymptotically tight. It is natural to consider a variant where the cost of buying is a concave function of the edge length. In this case a better result may be possible. In particular we analyze the case where this cost function is logarithmic in edge length. Such a logarithmic cost function may capture the economy-of-scale effects where buying multiplicatively longer edges costs only additively more. In [Section 2.1](#), we prove the following corollary of [Theorem 2](#),

**Corollary 1.** *Given a complete weighted graph with arbitrary weights which are at least 2, the greedy algorithm is  $O(1)$ -competitive for the Online Metric Steiner Tree with logarithmic edge costs.*

**Our work, part II: Volume Preserving Embeddings** We use [Theorem 2](#) and recent results on metric embeddings [[ABN11](#)] to show an algorithm that provides a non-contractive embedding into Euclidean space that faithfully preserves volume in the following sense: the embedding obtains simultaneously both  $O(\log k)$  average volume distortion and  $O(\log n)$  worst case volume distortion for sets of size  $k$ .

Given an  $n$  point metric space  $(X, d)$  an *injective* mapping  $f : X \rightarrow L_2$  is called an *embedding*. An embedding is  $(k - 1)$ -dimensional non-contractive if for any  $S \in \binom{X}{k}$ :  $\phi_E(f(S)) \geq \phi_F(S)$ .

Let  $f$  be a  $(k - 1)$ -dimensional non-contractive embedding. For a set  $S \subseteq \binom{X}{k}$  define the  $(k - 1)$ -dimensional distortion of  $S$  under  $f$  as:

$$\text{dist}_f(S) = \left[ \frac{\phi_E(f(S))}{\phi_F(S)} \right]^{1/(k-1)}.$$

For  $2 \leq k \leq n$  define the  $(k - 1)$ -dimensional distortion of  $f$  as

$$\text{dist}^{(k-1)}(f) = \max_{S \in \binom{X}{k}} \text{dist}_f(S)$$

More generally, for  $2 \leq k \leq n$  and  $1 \leq q \leq \infty$ , define the  $(k - 1)$ -dimensional  $\ell_q$ -distortion of  $f$  as:

$$\text{dist}_q^{(k-1)}(f) = \mathbb{E}_{S \sim \binom{X}{k}} [\text{dist}_f(S)^q]^{1/q}$$

where the expectation is taken according to the uniform distribution over  $\binom{X}{k}$ . Observe that the notion of  $(k - 1)$ -dimensional distortion is expressed by  $\text{dist}_\infty^{(k-1)}(f)$  and the average  $(k - 1)$ -dimensional distortion is expressed by the  $\text{dist}_1^{(k-1)}(f)$ -distortion.

It is worth noting that Feige’s definition of volume is related to the maximum volume obtained by non-expansive embeddings, while the definition of average distortion and  $\ell_q$ -distortion are using non-contractive embeddings. We note that these definitions are crucial in order to capture the coarse geometric notion described above and achieve results that significantly beat the usual worst case lower bounds (which depend on the size of the metric). It is clear that one can modify the definition to allow arbitrary embeddings (in particular non-contractive) by defining distortions normalized by taking their ratio with respect to the largest contraction.<sup>2</sup>

Our main theorem on volume preserving embeddings is:

**Theorem 3.** *For any metric space  $(X, d)$  on  $n$  points and any  $2 \leq k \leq n$ , there exists a map  $f : X \rightarrow L_2$  such that for any  $1 \leq q \leq \infty$ ,  $\text{dist}_q^{(k-1)}(f) \in O(\min\{\lceil q/(k - 1) \rceil \cdot \log k, \log n\})$ . In particular,  $\text{dist}_\infty^{(k-1)}(f) \in O(\log n)$  and  $\text{dist}_1^{(k-1)}(f) \in O(\log k)$ .*

On top of the robustness property of the general volume definition of [Theorem 2](#) the proof of [Theorem 3](#) builds on the embedding techniques developed in [\[ABN11\]](#) (in the context of pairwise distortion) along with combinatorial arguments that enable the stated bounds on the average and  $\ell_q$ -volume distortions.

Our embedding preserves well sets with typically large distances and can be viewed within the context of coarse geometry where we desire a “*high level*” geometric representation of the space. This follows from a special property formally stated in [Lemma 5](#).

## 1.1 Related Work

Embeddings of metric spaces have been a central field of research in theoretical computer science in recent years, due to the fact the metric spaces are important objects in representation of data. A fundamental theorem of Bourgain [\[Bou85\]](#) states that every  $n$  point metric space  $(X, d)$  can be embedded in  $L_2$  with distortion  $O(\log n)$ , where the distortion is defined as the worst-case multiplicative factor by which a pair of distances change. Our work extends this result in two aspects: (1) bounding the distortion of sets of arbitrary size, and (2) providing bounds for the  $\ell_q$ -distortion for all  $q \leq \infty$ .

**Volume preserving embeddings.** Feige [\[Fei00\]](#) introduced volume preserving embeddings while developing an approximation algorithm for the bandwidth problem. He showed that Bourgain’s embedding provides an embedding into Euclidean space with  $(k - 1)$ -dimensional distortion of  $O(\sqrt{\log n} \cdot \sqrt{\log n + k \log k})$ .

Following Feige’s work some special cases of volume preserving embeddings were studied, where the metric space  $X$  is restricted to a certain class of metric spaces. Rao [\[Rao99\]](#) studies the case where  $X$  is planar or is an excluded-minor metric showing constant  $(k - 1)$ -dimensional distortions. Gupta [\[Gup01\]](#) showed an improved approximation of the bandwidth for trees and chordal graphs. As the Feige volume does not coincide with the standard Euclidean volume, it is also interesting to study this special case when the metric space is given in Euclidean space. This case was studied by Rao [\[Rao99\]](#), Dunagan and Vempala

---

<sup>2</sup>There are other notions of average distortion that may be of interest, in particular such notions which normalize with respect to the maximum distortion have been considered. While these have advantages of their own, they take a very different geometric perspective which puts emphasis on small distance scales (as opposed to the coarse geometric perspective in this paper) and the worst case lower bounds hold for these notions. See [\[Rab03\]](#).

[DV01] and by Lee [Lee06]. We note that our work provides the first average distortion and  $\ell_q$ -distortion analysis also in the context of this special case.

The first improvement on Feige’s volume distortion bounds comes from the work of Rao [Rao99]. As observed by many researchers Rao’s embedding gives more general results depending on a certain decomposability parameter of the space. This provides a bound on the  $(k - 1)$ -dimensional distortion of  $O((\log n)^{3/2})$  for all  $k \leq n$ . This bound has been further improved to  $O(\log n)$  in work of Krauthgamer et al. [KLMN04]. Krauthgamer, Linial and Magen [KLM04] show a matching  $\Omega(\log n)$  lower bound on the  $(k - 1)$ -dimensional distortion for all  $k < n^{1/3}$ .

**Average and  $\ell_q$  Distortion.** The notions of average distortion and  $\ell_q$ -distortion is tightly related to the notions of partial embeddings and scaling embedding<sup>3</sup>. A  $(1 - \epsilon)$  partial embedding requires distortion at most  $\alpha$  for at least  $(1 - \epsilon)$  fraction of the pairs. In a scaling embedding we have a function  $\alpha : (0, 1) \rightarrow \mathbb{R}$ , and it demands  $(1 - \epsilon)$  fraction of the pairs to have distortion at most  $\alpha(\epsilon)$ , for all  $\epsilon \in (0, 1)$  *simultaneously*. These notions were introduced by Kleinberg, Slivkins and Wexler [KSW09], largely motivated by the study of distances in computer networks.

In [ABC<sup>+</sup>05] partial embedding into  $L_p$  with tight  $O(\log 1/\epsilon)$  partial distortion were given. The embedding method of [ABN11] provides a scaling embedding with  $O(\log 1/\epsilon)$  distortion for all values of  $\epsilon > 0$  simultaneously. As a consequence of having scaling embedding, they show that any metric space can be embedded into  $L_p$  with constant average distortion, and more generally that the  $\ell_q$ -distortion bounded by  $O(q)$ , while maintaining the best worst case distortion possible of  $O(\log n)$ , simultaneously.

Previous results on average distortion have applications for a variety of approximation problems, including uncapacitated quadratic assignment [ABN11], and in addition have been used in solving graph theoretic problems [ELR07]. Following [KSW09, ABC<sup>+</sup>05, ABN11] related notions have been studied in various contexts [CDG06, KRXY07, ABN07, Din07].

## 2 Robustness of the Metric Volume

### Proof of Theorem 2.

For a tree  $T$  on  $n$  vertices  $\{v_1, \dots, v_n\}$  let  $\bar{\phi}(T)$  be the product of the edge lengths. Because of the matroid exchange property, this product is minimized by an MST.<sup>4</sup> Thus for any metric space on points  $\{v_1, \dots, v_n\}$  and any spanning tree  $T$ ,  $\phi_F(v_1, \dots, v_n) \leq \bar{\phi}(T)/(n - 1)!$ ; the inequality is saturated by any (and only a) minimum spanning tree.

**Definition 1.** A *forced spanning tree (FST)* for a finite metric space is a spanning tree whose vertices can be ordered  $v_1, \dots, v_n$  so that for every  $i > 1$ ,  $v_i$  is connected to a vertex that is closest among  $v_1, \dots, v_{i-1}$ , and to no other among these. (We call such an ordering *admissible* for the tree.)

An MST is an FST with the additional property that in an admissible ordering  $v_i$  is a closest vertex to  $v_1, \dots, v_{i-1}$  among  $v_i, \dots, v_n$ .

**Definition 2.** For a tree  $T$  let  $\Delta(T)$  denote its diameter (the largest distance between any two points in the tree). Let the diameter  $\Delta(F)$  of a forest  $F$  with components  $T_1, T_2, \dots, T_m$  be  $\Delta(F) = \max_{1 \leq i \leq m} \Delta(T_i)$ . For a metric space  $(X, d)$  let  $\Delta_k(X) = \min\{\Delta(F) \mid F \text{ is a spanning forest of } X \text{ with } k \text{ connected components}\}$ .

<sup>3</sup>alternatively known as embeddings with slack and embeddings with gracefully degrading distortion.

<sup>4</sup>The exchange property guarantees that the greedy algorithm for a minimum weight maximal independent set in a weighted matroid (with positive weights), gives a set whose (sorted) weight vector is smaller or equal in every coordinate than that of any other maximal independent set. This implies that the product of weights is also minimized.

**Lemma 2.** Let  $(X, d)$  be a metric space. Let  $k \geq 1$ . An FST for  $X$  has at most  $k - 1$  edges of length greater than  $\Delta_k(X)$ .

*Proof.* Let  $v_1, \dots, v_n$  be an admissible ordering of the vertices of the FST. Assign each edge to its higher-indexed vertex. Since the ordering is admissible, this assignment is injective. The lemma is trivial for  $k = 1$ . For  $k \geq 2$ , cover  $X$  by the union of  $k$  trees each of diameter at most  $\Delta_k(X)$ . Only the lowest-indexed vertex in a tree can be assigned an edge longer than  $\Delta_k(X)$ . (Note that  $v_1$  is assigned no edge, hence the bound of  $k - 1$ .)  $\square$

**Corollary 3.** For any  $n$ -point metric space  $(X, d)$  and any FST  $T'$  for  $X$ ,  $\bar{\phi}(T') \leq \prod_{k=1}^{n-1} \Delta_k(X)$ .

*Proof.* Order the edges from 1 to  $n - 1$  by decreasing length. The  $k$ 'th edge is no longer than  $\Delta_k(X)$ .  $\square$

Using [Corollary 3](#), our proof of [Theorem 2](#) reduces to showing that for any MST  $T$  of  $X$ ,  $\prod_{k=1}^{n-1} \Delta_k(X) \leq e^{O(n-1)} \bar{\phi}(T)$ . Specifically we shall show that for any spanning tree  $T$ ,

$$\prod_{k=1}^{n-1} \Delta_k(X) \leq \frac{1}{n^2} \left( \frac{4\pi^2}{3} \right)^{n-1} \bar{\phi}(T).$$

(Observe incidentally that the FST created by the Gonzalez [[Gon85](#)] and Hochbaum-Shmoys [[HS85](#)] process has  $\bar{\phi}$  at least  $2^{1-n} \prod_{k=1}^{n-1} \Delta_k(X)$ .)

The idea is to recursively decompose  $T$  by cutting an edge; letting the two remaining trees be  $T_1$  (with some  $m$  edges) and  $T_2$  (with  $n - 2 - m$  edges), we shall upper bound  $\prod_1^{n-1} \Delta_k(T)$  in terms of  $\prod_1^m \Delta_k(T_1)$  and  $\prod_1^{n-2-m} \Delta_k(T_2)$ . More on this after we show how to pick an edge to cut. Recall:  $\sum_{j \geq 1} 1/j^2 = \pi^2/6$ .

**Edge selection.** Find a diametric path  $\gamma$  of  $T$ , i.e., a simple path whose length  $|\gamma|$  equals the diameter  $\Delta(T)$ . For appropriate  $\ell \geq 2$  let  $u_1, \dots, u_\ell$  be the weights of the edges of  $\gamma$  in the order they appear on the path. Select the  $j$ 'th edge on the path, for a  $1 \leq j \leq \ell$  for which  $u_j/|\gamma| > 1/(2(\pi^2/6) \min\{j, \ell + 1 - j\}^2)$ . Such an edge exists, as otherwise  $\sum_1^\ell u_j \leq (6/\pi^2)|\gamma| \sum_1^\ell j^{-2} < |\gamma|$ . Without loss of generality  $j \leq \ell + 1 - j$  (otherwise flip the indexing on  $\gamma$ ), hence cutting  $u_j$  contributes overhead  $|\gamma|/u_j < 2(\pi^2/6)j^2$  to the product  $\prod_1^{n-1} \Delta_k$ , and yields subtrees  $T_1$  and  $T_2$  each containing at least  $j - 1$  edges.

Think of this recursive process as successively breaking the spanning tree into a finer and finer forest. Note that we haven't yet specified which tree of the forest is cut, but we have specified which edge in that tree is cut. The order in which trees are chosen to be cut is:  $F_k(T)$  (which has  $k$  components) is defined by (a)  $F_1(T) = T$ ; (b) For  $1 < k < n$ ,  $F_k(T)$  is obtained from  $F_{k-1}(T)$  by cutting an edge in the tree of greatest diameter. Note that by definition  $\Delta_k(X) \leq \Delta(F_k(T))$ .

**Induction.** Now we show that

$$\prod_1^{n-1} \Delta(F_k(T)) \leq \frac{1}{n^2} \left( \frac{4\pi^2}{3} \right)^{n-1} \bar{\phi}(T).$$

It will be convenient to do this by an induction showing that there are constants  $c_1, c_2 > 0$  such that

$$\prod_1^{n-1} \Delta(F_k(T)) \leq e^{c_1(n-1) - c_2 \log n} \bar{\phi}(T),$$

and finally justify the choices  $c_1 = \log(4\pi^2/3)$  and  $c_2 = 2$ . As to base-cases,  $n = 1$  is trivial, and  $n = 2$  is assured for any  $c_1 \geq 0$ .

For  $n > 2$  let the children of  $T$  be  $T_1$  and  $T_2$ , that is to say,  $F_2(T) = \{T_1, T_2\}$ . Let  $m$  and  $n - 2 - m$  be the numbers of edges in  $T_1$  and  $T_2$  respectively. Observe that with  $j$  as defined above,  $\min\{m, n - 2 - m\} \geq j - 1 \geq 0$ .

Examine three sequences of forests: the  $T$  sequence,  $F_1(T), \dots, F_{n-1}(T)$ ; the  $T_1$  sequence,  $F_1(T_1), \dots, F_m(T_1)$ ; the  $T_2$  sequence,  $F_1(T_2), \dots, F_{n-2-m}(T_2)$ .

As indicated earlier, in each forest  $f$  in the  $T$  sequence other than  $F_1(T)$ , choose a component  $t$  of greatest diameter, i.e., one for which  $\Delta(t) = \Delta(f)$ . (In case of ties some consistent choice must be made within the  $T, T_1$  and  $T_2$  sequences.)

If  $t$  lies within  $T_1$ , assign  $f$  to the forest in the  $T_1$  sequence that agrees with  $f$  within  $T_1$ . Similarly if  $t$  lies within  $T_2$ , assign  $f$  to the appropriate forest in the  $T_2$  sequence. Due to the process defining the forests  $F_k(T)$ , this assignment is injective. Moreover, a forest in the  $T$  sequence, and the forest it is assigned to in the  $T_1$  or  $T_2$  sequence, share a common diameter. Hence

$$\prod_2^{n-1} \Delta(F_k(T)) = \left( \prod_1^m \Delta(F_k(T_1)) \right) \left( \prod_1^{n-2-m} \Delta(F_k(T_2)) \right).$$

Therefore:

$$\prod_1^{n-1} \Delta(F_k(T)) = \Delta(T) \cdot \prod_2^{n-1} \Delta(F_k(T)) = \Delta(T) \cdot \left( \prod_1^m \Delta(F_k(T_1)) \right) \left( \prod_1^{n-2-m} \Delta(F_k(T_2)) \right).$$

Now by induction:

$$\prod_1^{n-1} \Delta(F_k(T)) \leq \Delta(T) \cdot e^{c_1 m - c_2 \log(m+1)} \cdot \bar{\phi}(T_1) \cdot e^{c_1(n-2-m) - c_2 \log(n-1-m)} \cdot \bar{\phi}(T_2).$$

As  $\bar{\phi}(T) = u_j \cdot \bar{\phi}(T_1) \bar{\phi}(T_2)$  we get

$$\begin{aligned} \frac{\prod_1^{n-1} \Delta(F_k(T))}{\bar{\phi}(T)} &\leq (\Delta(T)/u_j) \cdot \exp \{c_1(n-2) - c_2(\log(m+1) + \log(n-1-m))\} \\ &\leq \exp \{ \log(2(\pi^2/6)j^2) + c_1(n-2) - c_2(\log(m+1) + \log(n-1-m)) \} \\ &\leq \exp \{ \log(\pi^2 j^2/3) + c_1(n-2) - c_2(\log j + \log(n/2)) \} \\ &\leq \exp \{ c_1(n-1) - c_2 \log n - (c_2 - 2) \log j - (c_1 - c_2 \log 2 - \log(\pi^2/3)) \} \end{aligned}$$

Choose  $c_2 \geq 2$  to take care of the third term in the exponent, and choose  $c_1 \geq \log(\pi^2/3) + c_2 \log 2$  to take care of the fourth term in the exponent. (In the theorem statement, both of these choices have been made with equality.) So

$$\dots \leq \exp \{c_1(n-1) - c_2 \log n\}.$$

□

## 2.1 Online Metric Steiner Tree

Here we prove [Corollary 1](#). Recall that in the online metric Steiner Tree problem we are given a complete weighted graph  $G = (V, E, w)$ , with  $d_G$  the shortest path metric on  $G$  with respect to the weights, and the

cost of each edge is the logarithm of its weight (we shall assume all weights are at least 2, so the cost of every edge is at least 1). Given a sequence  $v_1, \dots, v_n$  of vertices from  $V$ , we should output at every step  $1 \leq i \leq n$  a subgraph  $C_i$  such that  $v_1, \dots, v_i$  are connected in  $C_i$ , and such that  $C_{i-1} \subseteq C_i$  for all  $i$ . The cost of the subgraph  $C_i$  is the summation over the costs of edges in  $C_i$ . The greedy algorithm does the following: at every step  $i \geq 2$ , add the edge  $\{v_i, v_j\}$  where  $v_j \in \{v_1, \dots, v_{i-1}\}$  is the closest to  $v_i$  among the previous vertices.

First we lower bound the cost of the optimal (offline) algorithm. For each  $i$ , the contribution of connecting  $v_i$  to the minimum Steiner tree is at least  $\frac{1}{2}d_G(v_i, V \setminus \{v_i\})$ . Since the weights are at least 2, for any path  $u_1, \dots, u_k$  of length  $\ell$  we have that

$$\sum_{j=2}^k \log(d_G(u_{j-1}, u_j)) \geq \log \left( \sum_{j=2}^k d_G(u_{j-1}, u_j) \right) \geq \log \ell.$$

This implies that the cost of connecting each  $v_i$  is at least  $\log \left( \frac{1}{2}d_G(v_i, V \setminus \{v_i\}) \right)$ , and thus

$$\text{cost}(OPT) \geq \sum_{i=2}^n \log \left( \frac{1}{2}d_G(v_i, V \setminus \{v_i\}) \right) = \log \left( \frac{1}{2^{n-1}} \prod_{i=2}^n d_G(v_i, V \setminus \{v_i\}) \right).$$

Next we upper bound the cost of the greedy algorithm, which is

$$\text{cost}(ALG) \leq \sum_{i=2}^n \log (d_G(v_i, \{v_1, \dots, v_{i-1}\})) = \log \left( \prod_{i=2}^n d_G(v_i, \{v_1, \dots, v_{i-1}\}) \right).$$

Using [Theorem 2](#) we have that

$$\text{cost}(ALG) \leq \text{cost}(OPT) + O(n),$$

and as  $\text{cost}(OPT) \geq n - 1$  we have that the greedy is also a  $O(1)$  multiplicative approximation algorithm.

### 3 Volume Preserving Embeddings

In this section we prove [Theorem 3](#). The construction will be based on the embedding of [\[ABN11\]](#), who gave a general framework for embedding metrics into normed spaces. It was shown in [\[ABN11\]](#) that for every metric space  $(X, d)$  on  $n$  points there exists a distribution over maps  $f : X \rightarrow \mathbb{R}$  with the following properties: Every map in the support has expansion  $O(\log n)$ , and for every pair of points  $x, y \in X$ , with probability  $1/2$  the map  $f$  does not contract  $x, y$ . Moreover, it was shown that distortion is *scaling*: for every  $0 < \epsilon < 1$ , at least  $1 - \epsilon$  fraction of the pairs of  $X$  have expansion only  $O(\log(1/\epsilon))$ . Using this, one can construct an embedding of  $X$  into  $\mathbb{R}^{O(\log n)}$  by taking  $O(\log n)$  independent copies of  $f$ , and applying concentration bounds. Having such a scaling distortion implies  $O(1)$  average distortion and an  $O(q)$  bound on the  $\ell_q$  distortion.

Here we extend this framework for embedding that preserve the volume of subsets of  $X$  of cardinality  $k$ . First we strengthen the analysis of the line embedding of [\[ABN11\]](#), so that the bound on the contraction of the embedding holds for any  $x \in X$  and any affine combination of the images of points in a *subset*  $S \subset X$  (with constant probability). We then define an appropriate analogue notion of scaling distortion for sets of size  $k$ , and show that taking  $O(k \log n)$  independent copies of the random line embedding yields an embedding with the appropriate bounds on the worst case, average and in general the  $(k - 1)$ -dimensional  $\ell_q$  distortion.



### 3.1 The Embedding

The following is a variation on a Lemma from [ABN11], where the bound on contraction is strengthened to hold for subsets  $S = \{s_0, \dots, s_{k-1}\}$ , rather than just for pairs. More precisely, instead of simply lower bounding the distance between two images of points in  $X$ , we will need to lower bound the distance of the image of some point  $s_i$  from any affine combination of the images of  $s_0, \dots, s_{i-1}$  (conditioned on the values of these images). We can only prove this for some very specific ordering of the points in  $S$ , so from now on we shall enforce an ordering on every subset  $S \subseteq X$  that complies with the requirements of the Lemma.

**Lemma 4.** *There exists a universal constant  $\hat{C}$  such that for every finite metric space  $(X, d)$  on  $n$  points, there exists a distribution  $\mathcal{D}$  over functions  $f : X \rightarrow \mathbb{R}$  such that the following holds.<sup>5</sup>*

- For all  $u, v \in X$  and all  $f \in \text{supp}(\mathcal{D})$ ,

$$|f(u) - f(v)| \leq \hat{C} \cdot \log \left( \frac{n}{|B(u, d(u, v))|} \right) \cdot d(u, v).$$

- For every subset  $S \subseteq X$  of size  $k$ , there exists an ordering  $S = (s_0, \dots, s_{k-1})$ , such that for any  $1 \leq i \leq k-1$ , values  $x_0, \dots, x_{i-1} \in \mathbb{R}$  and coefficients  $\alpha_0, \dots, \alpha_{i-1} \in \mathbb{R}$  with  $\sum_{j=0}^{i-1} \alpha_j = 1$ :

$$\Pr_{f \sim \mathcal{D}} \left[ \left| f(s_i) - \sum_{j=0}^{i-1} \alpha_j x_j \right| \geq d(s_i, \{s_0, \dots, s_{i-1}\}) / \hat{C} \mid f(s_j) = x_j \forall 0 \leq j \leq i-1 \right] \geq 1/2.$$

Let  $D = c \cdot k \ln n$  where  $c$  is a constant to be determined later. Define the embedding  $g : X \rightarrow \mathbb{R}^D$  by

$$g = \frac{4\hat{C}}{\sqrt{D}} \bigoplus_{t=1}^D f_t,$$

where each  $f_t$  is sampled independently according to Lemma 4.

Next, we generalize the notion of scaling distortion for subsets of size  $k$ . To this end, for each subset  $S \in \binom{X}{k}$  with its ordering  $S = (s_0, \dots, s_{k-1})$  (the ordering enforced by Lemma 4), define a sequence  $(\epsilon_1, \dots, \epsilon_{k-1})$  as follows. For each  $1 \leq i \leq k-1$  let  $0 \leq j(i) < i$  be such that  $d(s_i, \{s_0, \dots, s_{i-1}\}) = d(s_i, s_{j(i)})$ . Let  $\epsilon_i$  be the value such that  $|B(s_{j(i)}, d(s_i, s_{j(i)}))| = \epsilon_i n$ . In other words,  $s_i$  is the  $\epsilon_i n$  nearest neighbor of the closest point to it in  $\{s_0, \dots, s_{i-1}\}$ .

**Lemma 5.** *For any embedding  $g$  in the support of the distribution and any  $S \in \binom{X}{k}$ ,*

$$\frac{\phi_E(g(S))}{\phi_F(S)} \leq \prod_{i=1}^{k-1} O(\log(1/\epsilon_i)).$$

*Proof.* Fix any  $1 \leq i \leq k-1$ , and let  $0 \leq j(i) \leq i-1$  be such that  $d(s_i, s_{j(i)}) = d(s_i, \{s_0, \dots, s_{i-1}\})$ . By the definition of  $\epsilon_i$ ,  $|B(s_{j(i)}, d(s_i, s_{j(i)}))| = \epsilon_i n$ . Using the first property of Lemma 4 we have for any  $t \in [D]$ ,

$$\begin{aligned} |f_t(s_i) - f_t(s_{j(i)})| &\leq \hat{C} \cdot \log \left( \frac{n}{|B(s_{j(i)}, d(s_i, s_{j(i)}))|} \right) \cdot d(s_i, s_{j(i)}) \\ &= \hat{C} \cdot \log(1/\epsilon_i) \cdot d(s_i, s_{j(i)}), \end{aligned}$$

<sup>5</sup>Moreover, there is an efficient algorithm to sample from  $\mathcal{D}$ .

thus also

$$d_E(g(s_i), g(s_{j(i)})) \leq \left( \frac{(4\hat{C})^2}{D} \sum_{t=1}^D (\hat{C} \cdot \log(1/\epsilon_i) \cdot d(s_i, s_{j(i)}))^2 \right)^{1/2} \leq 4\hat{C}^2 \log(1/\epsilon_i) \cdot d(s_i, s_{j(i)}) . \quad (2)$$

Now [Theorem 2](#) suggests that

$$\prod_{i=1}^{k-1} d(s_i, s_{j(i)}) \leq C^k \prod_{i=1}^{k-1} d(s_i, S \setminus \{s_i\}) = C^k (k-1)! \cdot \phi_F(S),$$

and we conclude the proof by

$$\begin{aligned} \phi_E(g(S)) &= \frac{1}{(k-1)!} \prod_{i=1}^{k-1} d_E(g(s_i), \text{affspan}(g(s_0), \dots, g(s_{i-1}))) \\ &\leq \frac{1}{(k-1)!} \prod_{i=1}^{k-1} d_E(g(s_i), g(s_{j(i)})) \\ &\leq \frac{1}{(k-1)!} \prod_{i=1}^{k-1} 4\hat{C}^2 \log(1/\epsilon_i) \cdot d(s_i, s_{j(i)}) \\ &\leq \prod_{i=1}^{k-1} O(\log(1/\epsilon_i)) \cdot \phi_F(S) . \end{aligned}$$

□

**Lemma 6.** *With probability at least  $1 - 1/n^k$ , the embedding  $g$  is  $(k-1)$ -dimensional non-contractive.*

*Proof.* Fix some  $S = (s_0, \dots, s_{k-1})$  and  $1 \leq i \leq k-1$ . Let  $\delta_i = d(s_i, \{s_0, \dots, s_{i-1}\})$  and  $A_i = \text{affspan}(g(s_0), \dots, g(s_{i-1}))$ . We would like to give a lower bound on the distance from  $g(s_i)$  to  $A_i$  in terms of  $\delta_i$ . The main difficulty is that the nearest point to  $g(s_i)$  in  $A_i$  naturally depends on the value of  $g(s_i)$ , thus we cannot use the second property of [Lemma 4](#) directly on the nearest point (we may condition only on  $g(s_j)$  for  $j < i$ ). The solution is as follows: rather than showing a lower bound on the distance from  $g(s_i)$  to the closest point in  $A_i$ , we will show a lower bound on the distance from  $g(s_i)$  to *all the points* in a suitable net of  $A_i$ .

To this end, let  $0 \leq j \leq i-1$  be such that  $\delta_i = d(s_i, s_j)$ , and let  $N_i$  be a  $\delta_i$ -net of  $B(g(s_j), 8\hat{C}^2 \delta_i \log n) \cap A_i$ . As  $A_i$  is an  $i-1$  dimensional space, and in such low dimensional space a ball of radius  $2r$  can be covered by  $2^{O(i)}$  balls of radius  $r$ . Applying this covering repeatedly we conclude that  $B(g(s_j), 8\hat{C}^2 \delta_i \log n) \cap A_i$  can be covered by  $2^{O(i \log \log n)}$  balls of radius  $\delta_i/2$ , and as each net point is contained in at most one of these balls, it follows that  $|N_i| = 2^{O(i \log \log n)} < n^k$  (for sufficiently large  $n$ ).

Now, if  $b_i \in A_i$  is the closest point to  $g(s_i)$ , then by using (2) and the fact that  $\epsilon_i \geq 1/n$ .

$$\begin{aligned} d_E(g(s_j), b_i) &\leq d_E(g(s_i), g(s_j)) + d_E(g(s_i), b_i) \\ &\leq 2d_E(g(s_i), g(s_j)) \\ &\leq 2(4\hat{C}^2) \log n \cdot d(s_i, s_j) \\ &= 8\hat{C}^2 \log n \cdot \delta_i . \end{aligned}$$

This suggests that indeed  $b_i \in B(g(s_j), 8\hat{C}^2\delta_i \log n)$ , so that there exists  $a'_i \in N_i$  with

$$d_E(a'_i, b_i) \leq \delta_i. \quad (3)$$

Next we prove that there is a high probability that  $g(s_i)$  is sufficiently far from *all* net points. Let  $a_i \in N_i$  be an arbitrary point of the net, and let  $\alpha_0, \dots, \alpha_{i-1}$  be such that  $\sum_{j=0}^{i-1} \alpha_j = 1$  and  $a_i = \sum_{j=0}^{i-1} \alpha_j g(s_j)$ . Observe that

$$d_E(g(s_i), a_i)^2 = \frac{(4\hat{C})^2}{D} \sum_{t=1}^D \left( f_t(s_i) - \sum_{j=0}^{i-1} \alpha_j f_t(s_j) \right)^2$$

For each  $t \in [D]$  let  $Z_t$  be an indicator random variable for the event  $|f_t(s_i) - \sum_{j=0}^{i-1} \alpha_j f_t(s_j)| \geq \delta_i/\hat{C}$ , and let  $Z = Z(S, i, a_i) = \sum_{t=1}^D Z_t$ . Observe that if it is the case that  $Z \geq D/4$ , then

$$\begin{aligned} d_E(g(s_i), a_i) &\geq \left( \frac{(4\hat{C})^2}{D} \sum_{t: Z_t=1} \left( f_t(s_i) - \sum_{j=0}^{i-1} \alpha_j f_t(s_j) \right)^2 \right)^{1/2} \\ &\geq \left( \frac{(4\hat{C})^2}{D} \cdot \frac{D}{4} (\delta_i/\hat{C})^2 \right)^{1/2} \\ &= 2\delta_i. \end{aligned} \quad (4)$$

By the triangle inequality, (3) and (4) used on  $a'_i$ ,

$$d_E(g(s_i), b_i) \geq d_E(g(s_i), a_i) - d_E(a'_i, b_i) \geq 2\delta_i - \delta_i = \delta_i.$$

We conclude that

$$\begin{aligned} \phi_E(g(S)) &= \frac{1}{(k-1)!} \prod_{i=1}^{k-1} d_E(g(s_i), b_i) \\ &\geq \frac{1}{(k-1)!} \prod_{i=1}^{k-1} \delta_i \\ &\geq \phi_F(S). \end{aligned}$$

It remains to show that with probability at least  $1 - 1/n^k$ , all of the bad events  $\{Z(S, i, a_i) < D/4\}_{S, i, a_i}$  do not happen. For a given  $Z$ , by [Lemma 4](#)  $\Pr[Z_t \mid g(s_0), \dots, g(s_{i-1})] \geq 1/2$  (because the different coordinates are independently chosen, so for each  $Z_t$ , conditioning on  $g$  is the same as conditioning just on  $f_t$ ), so that  $\mathbb{E}[Z] \geq D/2$ . The crucial observation is that in the definition of the bad events we fixed only  $g(s_0), \dots, g(s_{i-1})$  (to determine  $A_i$  and the net), but not  $g(s_i)$ . Using a standard Chernoff bound

$$\Pr[Z < D/4] \leq \Pr[Z < \mathbb{E}[Z]/2] \leq e^{-\mathbb{E}[Z]/8} \leq e^{-D/(16)} \leq n^{-3k},$$

where the last inequality holds when  $c = 48$ , say. By applying the union bound on all possible  $\binom{n}{k}$  sets  $S$ , all  $k$  possible indices  $i$ , and all the different points  $a_i \in N_i$  (recall that  $|N_i| < n^k$ ) we have that with probability at most  $k \cdot n^k \cdot \binom{n}{k} / n^{3k} \leq 1/n^k$ , some bad event happened.  $\square$

By [Lemma 5](#) and [Lemma 6](#) we have that there is high probability to obtain an embedding  $g : X \rightarrow \mathbb{R}^D$ , such that for every subset  $S \in \binom{X}{k}$  with its sequence  $(\epsilon_1, \dots, \epsilon_{k-1})$ ,

$$\text{dist}_g(S) \leq O \left( \left( \prod_{i=1}^{k-1} \log(1/\epsilon_i) \right)^{1/(k-1)} \right). \quad (5)$$

In the following sections we analyze the  $\ell_q$  volume distortion of such an embedding. For the sake of simplicity we start by the  $\ell_\infty$  and then the  $\ell_1$  volume distortions before handling the general  $(k-1)$ -dimensional  $\ell_q$ -distortion.

### 3.1.1 Bounding the $(k-1)$ -dimensional distortion

**Lemma 7.** *The (worst case)  $(k-1)$ -dimensional distortion of  $g$  is  $O(\log n)$  i.e.  $\text{dist}_\infty^{(k-1)}(g) = O(\log n)$ .*

*Proof.* For any set  $S \in \binom{X}{k}$  and  $i \in [k-1]$ ,  $\epsilon_i \geq 1/n$ . So by (5)

$$\text{dist}_g(S) \leq O \left( \left( \prod_{i=1}^{k-1} \log(1/\epsilon_i) \right)^{1/(k-1)} \right) \leq O(\log n).$$

□

### 3.1.2 Bounding the average $(k-1)$ -dimensional distortion

**Lemma 8.** *The average  $(k-1)$ -dimensional distortion of  $g$  is  $O(\log k)$  i.e.  $\text{dist}_1^{(k-1)}(g) = O(\log k)$ .*

*Proof.* For every set  $S \in \binom{X}{k}$  let  $m = m(S) = \min_i \{\epsilon_i n\}$ . By (5) there is a universal constant  $C'$  such that the average distortion over all possible  $S \in \binom{X}{k}$  can be bounded as follows.

$$\begin{aligned} \frac{\text{dist}_1^{(k-1)}(g)}{C'} &\leq \mathbb{E}_{S \in \binom{X}{k}} \left[ \left( \prod_{i=1}^{k-1} \log(1/\epsilon_i) \right)^{1/(k-1)} \right] \\ &\leq \mathbb{E} \left[ \left( \prod_{i=1}^{k-1} \log(n/m) \right)^{1/(k-1)} \right] \\ &= \mathbb{E} [\log(n/m)] \end{aligned}$$

In what follows we attempt to bound  $\mathbb{E} [\log(n/m)]$ . First we show that for every  $i$  and every  $t \in [n]$ , we have that  $\Pr[\epsilon_i = t/n] \leq 2k/n$  (recall that  $g$  is fixed, and the probability is over a uniform choice of a subset  $S$ ). This is because conditioning on any  $s_0, \dots, s_{i-1}$ , for every  $0 \leq j \leq i-1$  the probability that  $s_i$  is the  $t$ -th nearest neighbor of  $s_j$  is at most  $1/(n-i)$ , and by the union bound, the probability that there exists such a  $j$  is at most  $i/(n-i) \leq k/(n-k) \leq 2k/n$  (assuming  $k < n/2$ , as otherwise the lemma is trivial). It follows by the union bound that for all  $t \in [n]$ ,

$$\Pr[m = t] \leq \sum_{i=1}^{k-1} \Pr[\epsilon_i = t/n] \leq 2k^2/n.$$

Let  $h = \lceil \frac{n}{k^2} \rceil$ , so that  $\Pr[m = t] \leq 2/h$ . Now,

$$\begin{aligned} \mathbb{E} [\log(n/m)] &\leq \sum_{t=1}^h \Pr[m = t] \cdot \log(n/t) + \Pr[m > h] \cdot \log(n/h) \\ &\leq \frac{2}{h} \left( h \log n - \sum_{t=1}^h \log t \right) + \log(k^2) + 2 \end{aligned}$$

Note that  $\sum_{t=1}^h \log t = \log(h!) \geq h \log(h/e)$ , hence

$$\mathbb{E} [\log(n/m)] \leq 2(\log n - \log(n/(ek^2))) + 2 \log k + 2 = O(\log k).$$

□

### 3.1.3 Bounding the $(k - 1)$ -dimensional $\ell_q$ -distortion

Here we generalize the bounds on the worst-case and average  $(k - 1)$ -dimensional distortion, to the  $\ell_q$  norm of the  $(k - 1)$ -dimensional distortion for arbitrary  $1 \leq q \leq \infty$ , and thus prove [Theorem 3](#). Taking higher norms of the distortion suggests that we have to be more careful in estimating the probability of a sequence  $(\epsilon_1, \dots, \epsilon_{k-1})$  for a random set  $S$  (unlike the  $q = 1$  case, where we could use only the minimal  $\epsilon_i$ ).

**Lemma 9.** *For any  $1 \leq q \leq \infty$ ,  $\text{dist}_q^{(k-1)}(\hat{f}) = O(\lceil q/(k - 1) \rceil \cdot \log k)$ .*

*Proof.* For  $\ell \in \{0, 1, \dots, k - 1\}$ , let  $\mathcal{S}^{(\ell)} \subseteq \binom{X}{k}$  contain all the sets  $S$  that have exactly  $\ell$  values of  $\epsilon_i$  which are bigger than  $1/k^6$ . In what follows we attempt to bound  $|\mathcal{S}^{(\ell)}|$ . There are at most  $\binom{k-1}{\ell} \leq 2^k$  possibilities to choose the  $\ell$  locations in the sequence  $(\epsilon_1, \dots, \epsilon_{k-1})$  that will have values larger than  $1/k^6$ . Assume that these locations are fixed to be the first  $\ell$  elements of the sequence, by sorting the sequence. How many sets correspond to such sequences? There are at most  $\binom{n}{\ell+1}$  possibilities to choose  $s_0$  and the other  $\ell$  points which induce the first  $\ell$  values  $\epsilon_1, \dots, \epsilon_\ell$ . As for the other values, observe that for  $i > \ell$  we have  $\epsilon_i < 1/k^6$ , which suggests that  $s_i$  is one of the  $\epsilon_i n$  nearest neighbor of at least one of the other  $k - 1$  points, and thus there are at most  $k \cdot \epsilon_i n$  choices for  $s_i$ . Let  $K_\ell = \{6 \log k, 6 \log k + 1, \dots, \log n\}^{k-\ell-1}$ , for ease of notation assume  $K_\ell$  is indexed by integers  $\ell < i < k$  (that is, for  $x \in K_\ell$  denote by  $x_{\ell+1}$  the first element of  $x$ , and by  $x_{k-1}$  the last one), and fix some  $x \in K_\ell$ . Denote by  $\mathcal{S}_x^{(\ell)}$  the collection of sets in  $\mathcal{S}^{(\ell)}$  satisfying that  $2^{-x_i} < \epsilon_i \leq 2^{-x_i+1}$  for all  $\ell < i < k$ . Then

$$|\mathcal{S}_x^{(\ell)}| \leq \binom{n}{\ell+1} \cdot 2^k \cdot \prod_{i=\ell+1}^{k-1} k \cdot n/2^{x_i}. \quad (6)$$

Let  $C'$  be the constant in [Lemma 5](#), and let  $m = \lceil q/(k - 1) \rceil$ . First we use [\(5\)](#) and the monotonicity of

the normalized  $\ell_p$  norm to argue that

$$\begin{aligned}
\frac{\text{dist}_q^{(k-1)}(g)}{C'} &\leq \mathbb{E}_{S \in \binom{[k]}{k}} \left[ \left( \prod_{i=1}^{k-1} \log(1/\epsilon_i) \right)^{q/(k-1)} \right]^{1/q} \\
&\leq \mathbb{E}_{S \in \binom{[k]}{k}} \left[ \left( \prod_{i=1}^{k-1} \log(1/\epsilon_i) \right)^m \right]^{\frac{1}{m(k-1)}} \\
&= \left[ \binom{n}{k}^{-1} \sum_{\ell=0}^{k-1} \sum_{S \in \mathcal{S}^{(\ell)}} \left( \prod_{i=1}^{k-1} \log(1/\epsilon_i) \right)^m \right]^{\frac{1}{m(k-1)}} \\
&\leq \left[ \binom{n}{k}^{-1} \sum_{\ell=0}^{k-1} \sum_{x \in K_\ell} \sum_{S \in \mathcal{S}_x^{(\ell)}} \left( (6 \log k)^\ell \prod_{i=\ell+1}^{k-1} \log(1/\epsilon_i) \right)^m \right]^{\frac{1}{m(k-1)}} \\
&\leq \left[ \binom{n}{k}^{-1} \sum_{\ell=0}^{k-1} (6 \log k)^{\ell m} \sum_{x \in K_\ell} \sum_{S \in \mathcal{S}_x^{(\ell)}} \left( \prod_{i=\ell+1}^{k-1} x_i \right)^m \right]^{\frac{1}{m(k-1)}} \\
&\stackrel{(6)}{\leq} \left[ \binom{n}{k}^{-1} \sum_{\ell=0}^{k-1} (6 \log k)^{\ell m} \sum_{x \in K_\ell} \binom{n}{\ell+1} \cdot 2^k \cdot \prod_{i=\ell+1}^{k-1} k \cdot n/2^{x_i} \prod_{i=\ell+1}^{k-1} x_i^m \right]^{\frac{1}{m(k-1)}} \\
&= \left[ \binom{n}{k}^{-1} \cdot 2^k \sum_{\ell=0}^{k-1} \binom{n}{\ell+1} (6 \log k)^{\ell m} \cdot (kn)^{k-\ell-1} \sum_{x \in K_\ell} \prod_{i=\ell+1}^{k-1} x_i^m / 2^{x_i} \right]^{\frac{1}{m(k-1)}}. \quad (7)
\end{aligned}$$

Next, we focus on the expression  $\sum_{x \in K_\ell} \prod_{i=\ell+1}^{k-1} x_i^m / 2^{x_i}$ . Recall that  $K_\ell$  is a  $k - \ell - 1$  tuple of elements in  $\{6 \log k, \dots, \log n\}$ , so that each  $x_i \geq 6 \log k$ , and we may bound  $1/2^{x_i} \leq 1/2^{3 \log k} \cdot 1/2^{x_i/2} = 1/k^3 \cdot 1/2^{x_i/2}$ . Now, rather than a summation of products, we will take a product of summations. That is, we write

$$\begin{aligned}
\sum_{x \in K_\ell} \prod_{i=\ell+1}^{k-1} x_i^m / 2^{x_i} &\leq \frac{1}{k^3} \sum_{x \in K_\ell} \prod_{i=\ell+1}^{k-1} x_i^m / 2^{x_i/2} \\
&= \prod_{i=\ell+1}^{k-1} \left( \sum_{z=6 \log k}^{\log n} z^m / 2^{z/2} \right), \\
&= \left( \sum_{z=6 \log k}^{\log n} z^m / 2^{z/2} \right)^{k-\ell-1} \quad (8)
\end{aligned}$$

where the first equation holds because for any choice of a vector  $x \in K_\ell$  with its corresponding product, we can associate a unique sequence of  $k - \ell - 1$  choices of numbers  $z$ . Next we bound the summation, with the

variable change  $y = z - 6 \log k$ ,

$$\begin{aligned}
\sum_{z=6 \log k}^{\log n} z^m / 2^{z/2} &\leq \sum_{z=6 \log k}^{\infty} z^m / 2^{z/2} \\
&= \sum_{y=0}^{\infty} (y + 6 \log k)^m / 2^{(y+6 \log k)/2} \\
&\leq \frac{1}{k^3} \sum_{y=0}^{\infty} (2y^m + 2(6 \log k)^m) / 2^{y/2} \\
&\leq \frac{8(6 \log k)^m}{k^3} + \frac{2}{k^3} \sum_{y=0}^{\infty} y^m / 2^{y/2},
\end{aligned}$$

where the last inequality is using that  $\sum_{y \geq 0} 2^{-y/2} \leq 4$ . We replace the sum by an integral and calculate

$$\sum_{y=0}^{\infty} y^m / 2^{y/2} \leq \sqrt{2} \int_0^{\infty} y^m / 2^{y/2} dy \leq (16m)^m,$$

which yields the following bound on (8)

$$\left( \sum_{z=6 \log k}^{\log n} z^m / 2^{z/2} \right)^{k-\ell-1} \leq \left( \frac{8(6 \log k)^m}{k^3} + \frac{4(16m)^m}{k^3} \right)^{k-\ell-1} \quad (9)$$

Plugging (9) into (7) we get that

$$\begin{aligned}
&\frac{\text{dist}_q^{(k-1)}(g)}{C'} \\
&\leq \left[ \binom{n}{k}^{-1} 2^k \sum_{\ell=0}^{k-1} \binom{n}{\ell+1} (6 \log k)^{\ell m} \cdot (kn)^{k-\ell-1} \left( \frac{8(6 \log k)^m}{k^3} + \frac{4(16m)^m}{k^3} \right)^{k-\ell-1} \right]^{\frac{1}{m(k-1)}} \\
&\leq \left[ \binom{n}{k}^{-1} 2^k \sum_{\ell=0}^{k-1} \binom{n}{\ell+1} (n/k^2)^{k-\ell-1} \left( (100 \log k)^{m(k-1)} \cdot (100m)^{m(k-1)} \right) \right]^{\frac{1}{m(k-1)}} \\
&\leq \left[ k \cdot 2^k \left( (100 \log k)^{m(k-1)} \cdot (100m)^{m(k-1)} \right) \right]^{\frac{1}{m(k-1)}},
\end{aligned}$$

where the last inequality is using that (for  $k \geq 2$  and) for every  $0 \leq \ell \leq k-1$ ,  $\binom{n}{\ell+1} (n/k^2)^{k-\ell-1} \leq \binom{n}{k}$ . Note that the above expression is at most  $O(m \log k) = O(\lceil q/(k-1) \rceil \cdot \log k)$  as required.  $\square$

## References

- [ABC<sup>+</sup>05] I. Abraham, Y. Bartal, T-H. H. Chan, K. Dhamdhere Dhamdhere, A. Gupta, J. Kleinberg, O. Neiman, and A. Slivkins. Metric embeddings with relaxed guarantees. In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 83–100, Washington, DC, USA, 2005. IEEE Computer Society.

- [ABN07] I. Abraham, Y. Bartal, and O. Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007.
- [ABN11] I. Abraham, Y. Bartal, and O. Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026–3126, 2011.
- [ABNS10] I. Abraham, Y. Bartal, O. Neiman, and L. J. Schulman. Volume in general metric spaces. In *Proceedings of the 18th annual European conference on Algorithms: Part II, ESA'10*, pages 87–99, Berlin, Heidelberg, 2010. Springer-Verlag.
- [Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.
- [CDG06] T.-H. H. Chan, M. Dinitz, and A. Gupta. Spanners with slack. In *ESA'06: Proceedings of the 14th conference on Annual European Symposium*, pages 196–207, London, UK, 2006. Springer-Verlag.
- [Din07] M. Dinitz. Compact routing with slack. In *PODC '07: Proceedings of the twenty-sixth annual ACM symposium on Principles of distributed computing*, pages 81–88, New York, NY, USA, 2007. ACM.
- [DV01] J. Dunagan and S. Vempala. On Euclidean embeddings and bandwidth minimization. *Lecture Notes in Computer Science*, 2129, 2001.
- [ELR07] M. Elkin, C. Liebchen, and R. Rizzi. New length bounds for cycle bases. *Information Processing Letters*, 104(5):186–193, 2007.
- [Fei00] U. Feige. Approximating the bandwidth via volume respecting embeddings. *J. Comput. System Sci.*, 60(3):510–539, 2000. 30th Annual ACM Symposium on Theory of Computing (Dallas, TX, 1998).
- [Gon85] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.
- [Gup01] A. Gupta. Improved bandwidth approximation for trees and chordal graphs. *J. Algorithms*, 40(1):91–36, 2001.
- [HS85] D. S. Hochbaum and D. B. Shmoys. A best possible heuristic for the  $k$ -center problem. *Math. Oper. Res.*, 10:180–184, 1985.
- [IW91] M. Imase and B. M. Waxman. Dynamic steiner tree problem. *SIAM J. Discrete Math.*, 4(3):369–384, 1991.
- [KLM04] R. Krauthgamer, N. Linial, and A. Magen. Metric embeddings—beyond one-dimensional distortion. *Discrete Comput. Geom.*, 31(3):339–356, 2004.
- [KLMN04] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 434–443. IEEE, October 2004.



- [KRXY07] G. Konjevod, A. W. Richa, D. Xia, and H. Yu. Compact routing with slack in low doubling dimension. In *PODC '07: Proceedings of the twenty-sixth annual ACM symposium on Principles of distributed computing*, pages 71–80, New York, NY, USA, 2007. ACM.
- [KSW09] J. Kleinberg, A. Slivkins, and T. Wexler. Triangulation and embedding using small sets of beacons. *J. ACM*, 56(6):1–37, 2009.
- [Lee06] J. R. Lee. Volume distortion for subsets of euclidean spaces. In *SCG '06: Proceedings of the twenty-second annual symposium on Computational geometry*, pages 207–216, New York, NY, USA, 2006. ACM.
- [Rab03] Yuri Rabinovich. On average distortion of embedding metrics into the line and into  $\ell_1$ . In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 456–462. ACM Press, 2003.
- [Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.

## A Proof of Lemma 4

In this section we prove the properties of the line embedding. We begin by several definitions. As usual  $(X, d)$  is a metric space on  $n$  points.

**Definition 3.** The local growth rate of  $x \in X$  at radius  $r > 0$  for given scales  $\gamma_1, \gamma_2 > 0$  is defined as

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

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

The following simple claim was shown in [ABN11]

**Claim 10.** Let  $x, y \in X$ , let  $\gamma_1, \gamma_2 > 0$  and let  $r$  be such that  $2(1 + \gamma_2)r < d(x, y) \leq (\gamma_1 - \gamma_2 - 2)r$ , then

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

**Definition 4** (Partition). A partition  $P$  of  $X$  is a collection of pairwise disjoint sets  $\mathcal{A}(P) = \{A_1, A_2, \dots, A_t\}$  for some integer  $t$ , such that  $X = \cup_j A_j$ . The sets  $A_j$  are called clusters. For  $x \in X$  denote by  $P(x)$  the cluster containing  $x$ . Given  $\Delta > 0$ , a partition is  $\Delta$ -bounded if for all  $j \in [t]$ ,  $\text{diam}(A_j) \leq \Delta$ . For  $Z \subseteq X$  we denote by  $P[Z]$  the restriction of  $P$  to points in  $Z$ .

**Definition 5** (Probabilistic Partition). A probabilistic partition  $\hat{P}$  of a metric space  $(X, d)$  is a distribution over a set  $\mathcal{P}$  of partitions of  $X$ . Given  $\Delta > 0$ ,  $\hat{P}$  is  $\Delta$ -bounded if each  $P \in \mathcal{P}$  is  $\Delta$ -bounded. Let  $\text{supp}(\hat{P}) \subseteq \mathcal{P}$  be the set of partitions with non-zero probability under  $\hat{P}$ .

**Definition 6** (Uniform Function). Given a partition  $P$  of a metric space  $(X, d)$ , a function  $f$  defined on  $X$  is called uniform with respect to  $P$  if for any  $x, y \in X$  such that  $P(x) = P(y)$  we have  $f(x) = f(y)$ .

Let  $\hat{P}$  be a probabilistic partition. A collection of functions defined on  $X$ ,  $f = \{f_P | P \in \mathcal{P}\}$  is uniform with respect to  $\hat{P}$  if for every  $P \in \mathcal{P}$ ,  $f_P$  is uniform with respect to  $P$ .

**Definition 7** (Uniformly Padded Local PP). Given  $\Delta > 0$  and  $0 < \delta \leq 1$ , let  $\hat{\mathcal{P}}$  be a  $\Delta$ -bounded probabilistic partition of  $(X, d)$ . Given collection of functions  $\eta = \{\eta_P : X \rightarrow [0, 1] \mid P \in \mathcal{P}\}$ , we say that  $\hat{\mathcal{P}}$  is  $(\eta, \delta)$ -locally padded if the event  $B(x, \eta_P(x)\Delta) \subseteq P(x)$  occurs with probability at least  $\delta$  regardless of the structure of the partition outside  $B(x, 2\Delta)$ . Formally, for all  $x \in X$ , for all  $A \subseteq X \setminus B(x, 2\Delta)$  and all partitions  $P'$  of  $A$ ,

$$\Pr[B(x, \eta_P(x)\Delta) \subseteq P(x) \mid P[A] = P'] \geq \delta$$

Let  $0 < \hat{\delta} \leq 1$ . We say that  $\hat{\mathcal{P}}$  is strong  $(\eta, \hat{\delta})$ -locally padded if for any  $\hat{\delta} \leq \delta \leq 1$ ,  $\hat{\mathcal{P}}$  is  $(\eta \cdot \ln(1/\delta), \delta)$ -padded.

We say that  $\hat{\mathcal{P}}$  is  $(\eta, \delta)$ -uniformly locally padded if  $\eta$  is uniform with respect to  $\mathcal{P}$ .

The following Lemma was shown in [ABN11]

**Lemma 11.** Let  $(Z, d)$  be a finite metric space. Let  $0 < \Delta \leq \text{diam}(Z)$ . Let  $\hat{\delta} \in (0, 1/2]$ ,  $\gamma_1 \geq 2$ ,  $\gamma_2 \leq 1/16$ . There exists a  $\Delta$ -bounded probabilistic partition  $\hat{\mathcal{P}}$  of  $(Z, d)$  and a collection of uniform functions  $\{\xi_P : Z \rightarrow \{0, 1\} \mid P \in \mathcal{P}\}$  and  $\{\eta_P : Z \rightarrow (0, 1] \mid P \in \mathcal{P}\}$  such that the probabilistic partition  $\hat{\mathcal{P}}$  is a strong  $(\eta, \hat{\delta})$ -uniformly locally padded probabilistic partition; and the following conditions hold for any  $P \in \text{supp}(\hat{\mathcal{P}})$  and any  $x \in Z$ :

- If  $\xi_P(x) = 1$  then:  $2^{-6}/\ln \rho(x, 2\Delta, \gamma_1, \gamma_2) \leq \eta_P(x) \leq 2^{-6}/\ln(1/\hat{\delta})$ .
- If  $\xi_P(x) = 0$  then:  $\eta_P(x) = 2^{-6}/\ln(1/\hat{\delta})$  and  $\bar{\rho}(x, 2\Delta, \gamma_1, \gamma_2) < 1/\hat{\delta}$ .

In what follows we prove Lemma 4. Let  $\Delta_0 = \text{diam}(X)$ . For  $l \in \mathbb{N}$  let  $\Delta_l = (1/8)^l \Delta_0$  and let  $P_l$  be a  $\Delta_l$ -bounded partition. For all  $l \in \mathbb{N}$  let  $\sigma_l : X \rightarrow [0, 1]$ ,  $\xi_l : X \rightarrow \{0, 1\}$ ,  $\eta_l : X \rightarrow \mathbb{R}^+$  be uniform functions with respect to  $P_l$ , the functions  $\eta_l$  and  $\xi_l$  will be randomly generated by the probabilistic partition. For every scale  $l \in \mathbb{N}$  define  $\varphi_l : X \rightarrow \mathbb{R}^+$  as

$$\varphi_l(x) = \min \left\{ \frac{\xi_l(x)}{\eta_l(x)} d(x, X \setminus P_l(x)), \Delta_l/4 \right\}, \quad (10)$$

and for  $l \in \mathbb{N}$  define  $\psi_l : X \rightarrow \mathbb{R}^+$  as

$$\psi_l(x) = \sigma_l(x) \cdot \varphi_l(x).$$

Finally let  $f : X \rightarrow \mathbb{R}^+$  be defined as  $f = \sum_{l \in \mathbb{N}} \psi_l$ . Note that  $f$  is well defined because  $f(x) = \sum_{l \in \mathbb{N}} \psi_l(x) \leq \sum_{l \in \mathbb{N}} \Delta_l$ , and this is a geometric progression.

The distribution  $\mathcal{D}$  on embeddings  $f$  is obtained by choosing each  $P_l$  from the distribution  $\hat{\mathcal{P}}_l$  as in Lemma 11 with parameters  $Z = X$ ,  $\Delta = \Delta_l$ ,  $\hat{\delta} = 1/2$ ,  $\gamma_1 = 32$  and  $\gamma_2 = 1/16$ . For each  $l \in \mathbb{N}$  set  $\xi_l = \xi_{P_l}$  and  $\eta_l = \eta_{P_l}$  as defined in the lemma, and let  $\sigma_l$  be a uniform function with respect to  $P_l$  defined by letting  $\{\sigma_l(A) \mid A \in P_l, l \in \mathbb{N}\}$  be i.i.d random variables chosen uniformly in the interval  $[0, 1]$ , and setting  $\sigma_l(x) = \sigma_l(P_l(x))$ . We begin by proving the first property of Lemma 4.

**Lemma 12.** For all  $u, v \in X$ ,  $f \in \text{supp}(\mathcal{D})$ ,

$$|f(u) - f(v)| \leq \hat{C} \left[ \log \left( \frac{n}{|B(u, d(u, v))|} \right) \right] d(u, v).$$

where  $\hat{C}$  is a universal constant.

*Proof.* Fix some  $u, v \in X$  and  $f \in \text{supp}(\mathcal{D})$ . Hence  $\{P_l\}_{l \in \mathbb{N}}, \{\sigma_l\}_{l \in \mathbb{N}}$  are fixed. Let  $h \in \mathbb{N}$  be the maximum index such that  $\Delta_h \geq 2d(u, v)$ , if no such  $h$  exists then let  $h = 0$ . We bound  $|f(u) - f(v)|$  by separating the sum into two parts  $0 \leq l < h$ , and  $l \geq h$ :

$$|f(u) - f(v)| \leq \sum_{0 \leq l < h} |\psi_l(u) - \psi_l(v)| + \sum_{l \geq h} |\psi_l(u)| + \sum_{l \geq h} |\psi_l(v)|. \quad (11)$$

We begin by bounding the first summation. Note that for any set  $U \subseteq X$  and numbers  $a, r \geq 0$ , by the triangle inequality

$$\min\{a \cdot d(u, U), r\} - \min\{a \cdot d(v, U), r\} \leq a \cdot d(u, v). \quad (12)$$

We will show that this implies the following

$$\psi_l(u) - \psi_l(v) \leq \frac{\xi_l(u)}{\eta_l(u)} d(u, v). \quad (13)$$

To see (13), we use the fact that  $\sigma_l, \xi_l, \eta_l$  are uniform functions. If it is the case that  $P_l(u) = P_l(v)$  then by (12) and as  $\sigma_l(u) = \sigma_l(v) \leq 1$  we have that  $\psi_l(u) - \psi_l(v) \leq \frac{\xi_l(u)}{\eta_l(u)} d(u, v)$ . Otherwise, if  $P_l(u) \neq P_l(v)$ , then  $d(u, X \setminus P_l(u)) \leq d(u, v)$  and hence  $\psi_l(u) - \psi_l(v) \leq \psi_l(u) \leq \frac{\xi_l(u)}{\eta_l(u)} d(u, v)$ . By symmetry we have that

$$|\psi_l(u) - \psi_l(v)| \leq \frac{\xi_l(u)}{\eta_l(u)} d(u, v) + \frac{\xi_l(v)}{\eta_l(v)} d(u, v). \quad (14)$$

For any  $x \in X$ ,

$$\begin{aligned} \sum_{0 \leq l < h} \frac{\xi_l(x)}{\eta_l(x)} &= \sum_{0 \leq l < h: \xi_l(x)=1} \eta_l(x)^{-1} \\ &\leq \sum_{0 \leq l < h: \xi_l(x)=1} 2^6 \ln \rho(x, 2\Delta_l, \gamma_1, \gamma_2) \\ &\leq 2^6 \sum_{0 \leq l < h} \ln \left( \frac{|B(x, 2\gamma_1 \Delta_l)|}{|B(x, 2\gamma_2 \Delta_l)|} \right) \\ &\leq 2^6 \cdot 3 \ln \left( \frac{|X|}{|B(x, \Delta_{h-1}/8)|} \right) \\ &\leq 2^9 \ln \left( \frac{|X|}{|B(x, \Delta_h)|} \right). \end{aligned} \quad (15)$$

The first inequality follows from the first property of [Lemma 11](#), and the third inequality holds as  $2\gamma_1 \Delta_l = 8^2 \Delta_l = 2 \cdot 8^3 \gamma_2 \Delta_l = 2\gamma_2 \Delta_{l-3}$ , this suggests that the sum is telescopic and is bounded accordingly. And now, noticing that  $|\psi_l(u)| \leq \Delta_l/4$  for all  $l \in \mathbb{N}$ ,

$$\begin{aligned} |f(u) - f(v)| &\leq \sum_{0 \leq l < h} |\psi_l(u) - \psi_l(v)| + \sum_{l \geq h} |\psi_l(u)| + \sum_{l \geq h} |\psi_l(v)| \\ &\stackrel{(14)}{\leq} \sum_{0 \leq l < h} \left( \frac{\xi_l(u)}{\eta_l(u)} + \frac{\xi_l(v)}{\eta_l(v)} \right) d(u, v) + \sum_{l \geq h} \Delta_l/2 \\ &\stackrel{(15)}{\leq} 2^9 \left( \ln \left( \frac{|X|}{|B(u, \Delta_h)|} \right) + \ln \left( \frac{|X|}{|B(v, \Delta_h)|} \right) \right) d(u, v) + \Delta_h \\ &\leq \hat{C} \left[ \ln \left( \frac{|X|}{|B(u, d(u, v))|} \right) \right] d(u, v), \end{aligned}$$

where  $\hat{C} > 2^{10}$  is a constant. The last inequality uses the fact that  $B(u, d(u, v)) \subseteq B(u, \Delta_h) \cap B(v, \Delta_h)$  and that the maximality of  $h$  suggests that  $\Delta_h \leq 16d(u, v)$ .  $\square$

We now proceed to prove the second property of [Lemma 4](#). Let us first introduce the desired ordering of a set  $S \subseteq X$ .

**Claim 13.** *For any  $S \in \binom{X}{k}$  there exists an ordering  $(s_0, s_1, \dots, s_{k-1})$  of the elements of  $S$  that satisfy the following property: For all  $i \in \{0, \dots, k-1\}$ , if  $h \in \mathbb{N}$  is the unique integer such that  $6\Delta_h \leq d(s_i, \{s_0, \dots, s_{i-1}\}) < 6\Delta_{h-1}$ , then*

$$\bar{\rho}(s_i, 2\Delta_h, \gamma_1, \gamma_2) \geq 2.$$

*Proof.* Let  $(s_0, s_1, \dots, s_{k-1})$  be an ordering according to the following iterative process. Start with  $W = S$ ,  $i = k-1$ .

1. Let  $\{u, v\} = \min_{s, t \in W} \{d(s, t)\}$ . Let  $h$  be the unique integer such that  $6\Delta_h \leq d(u, v) < 6\Delta_{h-1}$ .
2. By [Claim 10](#) with parameters  $r = 2\Delta_h$ ,  $\gamma_1 = 32$  and  $\gamma_2 = 1/16$  (as defined above), at least one of  $u, v$ , w.l.o.g  $u$ , has  $\bar{\rho}(u, 2\Delta_h, \gamma_1, \gamma_2) \geq 2$ . Then set  $s_i = u$  and  $W = W \setminus \{u\}$ .
3. If  $i > 1$ , set  $i = i-1$  and goto 1. Otherwise (when  $i = 1$ ) set  $s_0$  to be the last element of  $W$ .

The claim follows by the very definition of the process.  $\square$

We will also use the following simple claim.

**Claim 14.** *Let  $A \in \mathbb{R}^+$  and let  $\alpha$  be random variable uniformly distributed in  $[0, 1]$ . Then for any  $C \in \mathbb{R}$  and  $\gamma > 0$ :*

$$\Pr[|C + A\alpha| < \gamma \cdot A] < 2\gamma.$$

**Lemma 15.** *For every subset  $S \subseteq X$  of size  $k$ , there exists an ordering  $S = (s_0, \dots, s_{k-1})$ , such that for any  $1 \leq i \leq k-1$ , values  $x_0, \dots, x_{i-1} \in \mathbb{R}$  and coefficients  $\alpha_0, \dots, \alpha_{i-1} \in \mathbb{R}$  with  $\sum_{j=0}^{i-1} \alpha_j = 1$ :*

$$\Pr_{f \sim \mathcal{D}} \left[ \left| f(s_i) - \sum_{j=0}^{i-1} \alpha_j x_j \right| \geq d(s_i, \{s_0, \dots, s_{i-1}\}) / \hat{C} \mid \forall 0 \leq j \leq i-1, f(s_j) = x_j \right] \geq 1/2.$$

*Proof.* Fix  $S \subseteq X$  of size  $|S| = k$  and let  $(s_0, \dots, s_{k-1})$  be the ordering induced by [Claim 13](#). Fix some  $1 \leq i \leq k-1$ , values  $x_0, \dots, x_{i-1} \in \mathbb{R}$  and coefficients  $\alpha_0, \dots, \alpha_{i-1} \in \mathbb{R}$  with  $\sum_{j=0}^{i-1} \alpha_j = 1$  as in the statement of the Lemma. Let  $h \in \mathbb{N}$  be such that  $6\Delta_h \leq d(s_i, \{s_0, \dots, s_{i-1}\}) < 6\Delta_{h-1}$ . Note that the different scales  $l \in \mathbb{N}$  are sampled independently, and that for the scale  $h$ , by [Lemma 11](#) even when conditioning on the values  $f(s_j)$  for  $0 \leq j \leq i-1$ , there is probability  $\delta = e^{-1/4} > 3/4$  that the following event holds

$$\mathcal{E}_{s_i\text{-pad}} = \{B(s_i, \eta_h(u) \cdot \Delta_h/4) \subseteq P_h(s_i)\}.$$

Also note that by [Claim 13](#) we have  $\bar{\rho}(s_i, 2\Delta_h, \gamma_1, \gamma_2) \geq 2$  and also  $1/\hat{\delta} = 2$ , thus the properties of [Lemma 11](#) imply that it must be that  $\xi_h(s_i) = 1$ . Given that event  $\mathcal{E}_{s_i\text{-pad}}$  holds, we obtain

$$\varphi_h(s_i) = \min \left\{ \frac{d(s_i, X \setminus P_h(s_i))}{\eta_h(s_i)}, \Delta_h/4 \right\} \geq \Delta_h/4. \quad (16)$$

Next consider the value  $a = \sum_{j=0}^{i-1} \alpha_j x_j$ , and condition on the event that for all  $0 \leq j \leq i-1$ ,  $f(s_j) = x_j$ . Let  $C = f(s_i) - a - \psi_h(s_i)$ ,  $A = \varphi_h(s_i)$  and  $\alpha = \sigma_h(s_i)$ , and observe that  $|f(s_i) - a| = |C + \alpha A|$ . By [Claim 14](#) with  $\gamma = 1/8$  it follows that

$$\Pr_{\alpha}[|C + \alpha A| < A/8] \leq 1/4. \quad (17)$$

Since the partition  $P_h$  is  $\Delta_h$  bounded and  $d(s_j, s_i) > \Delta_h$  for all  $0 \leq j \leq i-1$ , we have that  $P_h(s_j) \neq P_h(s_i)$ . This suggests that  $\alpha = \sigma_h(s_i)$  is indeed chosen independently of the  $f(s_j)$ . Since both [\(16\)](#) and [\(17\)](#) each holds with probability at least  $3/4$ , there is probability at least  $1/2$  that both of them hold (again, even when conditioning on the  $f(s_j)$ ), in which case

$$|f(s_i) - a| \geq A/8 \geq \Delta_h/32 = d(s_i, \{s_0, \dots, s_{i-1}\})/\hat{C},$$

for sufficiently large constant  $\hat{C}$ .

□