Volume in General Metric Spaces

Ittai Abraham[1], Yair Bartal[2], Ofer Neiman[3], and Leonard J. Schulman[4]

[1] Microsoft Research. ⋆
[2] Hebrew University ⋆⋆
[3] New York University.⋆⋆⋆
[4] Caltech. †

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 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 $S = \{v_1, \ldots, v_k\}$ of points in a general metric space: the product of the distances from v_i to $\{v_1, \ldots, v_{i-1}\}$, 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$^{1/(k-1)}$ 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 embeddings.

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.

⋆ Email: ittaia@microsoft.com.
⋆⋆ 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 (195/02) and in part by a grant from the National Science Foundation (NSF CCF-0652536).
⋆⋆⋆ Email: neiman@cims.nyu.edu. partially funded by NSF Expeditions in Computing award, 2009-2010.
† Email: schulman@caltech.edu. Supported in part by NSF CCF-0515342 and NSA H98230-06-1-0074.

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 [10].

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 a 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, showing that the cost of greedy is within an additive term of the minimum spanning tree, implying a constant competitive ratio. Our main result 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 ℓ_q-distortion), while the embedding still maintains the best possible worst case distortion bound, simultaneously. Hence our embedding generalizes both [17] and [2] (see Related Work below).

Volume in general metric spaces.

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

$$\mathrm{vol}_E(X) = \frac{1}{(n-1)!} \prod_{i=2}^{n} d_E(v_i, \mathrm{affspan}(v_1, \ldots, v_{i-1})).$$

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, $X = \{v_1, \ldots, v_n\}$. Let S_n be the symmetric group on n symbols, and let $\pi_P \in S_n$ be the 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 $(n-1)!$):

$$\mathrm{vol}_F(X) = \frac{1}{(n-1)!} \prod_{i=2}^{n} d_X(v_{\pi_P(i)}, \{v_{\pi_P(1)}, \ldots, v_{\pi_P(i-1)}\}). \tag{1}$$

π_P minimizes the above expression[5].

It should be noted that even if X is a subset of Euclidean space, vol_E and vol_F do not agree. (The latter can be arbitrarily larger than the former.) 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):*

$$1 \le \left[\frac{\mathrm{vol}_F(X)}{\sup_{f \in \mathcal{L}_2(X)} \mathrm{vol}_E(f(X))}\right]^{1/(n-1)} \le 2.$$

[5] This is because the (sorted) vector of edge lengths created by Prim's algorithm is smaller or equal in each coordinate than any sorted vector of a spanning tree's edge lengths

Thus, remarkably, $\mathrm{vol}_F(X)$ 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 π_P defined as above, and for every $\pi \in S_n$:*

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

An alternative interpretation of this result can be presented as the analysis of the following online problem. Consider the following variant of the online Steiner tree problem [14]. Given a 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 $E_1, \ldots, E_i$ must induce a connected graph among the current set of vertices $v_1, \ldots, v_i$. 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 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. Theorem 2 can be interpreted as a competitive analysis of the greedy algorithm in this model, showing that the cost of the greedy algorithm is within $O(n)$ additive term of the minimum spanning tree, which implies an $O(1)$ competitive ratio for this problem.

Our work, part II: Volume Preserving Embeddings We use Theorem 2 and recent results on metric embeddings [2] to show that their algorithm 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 $f : X \to L_2$ is called an *embedding*. An embedding is $(k-1)$*-dimensional non-contractive* if for any $S \in \binom{X}{k}$: $\mathrm{vol}_E(f(S)) \geq \mathrm{vol}_F(S)$.

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:

$$\mathrm{dist}_f(S) = \left[\frac{\mathrm{vol}_E(f(S))}{\mathrm{vol}_F(S)} \right]^{1/(k-1)}.$$

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

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

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

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

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 $\mathrm{dist}_\infty^{(k-1)}(f)$ and the average $(k-1)$-dimensional distortion is expressed by the $\mathrm{dist}_1^{(k-1)}(f)$-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 ℓ_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.[6]

Our main theorem on volume preserving embeddings is:

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

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 [2] (in the context of pairwise distortion) along with combinatorial arguments that enable the stated bounds on the average and ℓ_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 4.

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 [5] 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 ℓ_q-distortion for all $q \le \infty$.

Volume preserving embeddings. Feige [10] introduced volume preserving embeddings. He showed that Bourgain's embedding provides an embedding into Euclidean space with $(k-1)$-dimensional distortion of $O(\sqrt{\log n} \cdot \sqrt{\log n + k \log k})$.

[6] 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.

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 [20] studies the case where X is planar or is an excluded-minor metric showing constant $(k-1)$-dimensional distortions. Gupta [12] showed an improved approximation of the bandwidth for trees and chordal graphs. As the Feige volume does not coincide with the standard volume of Euclidean set it is also interesting to study this special case when the metric space is given in Euclidean space. This case was studied by Rao [20], Dunagan and Vempala [8] and by Lee [19]. We note that our work provides the first average distortion and ℓ_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 [20]. 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 $k \le n$. This bound has been further improved to $O(\log n)$ in work of Krauthgamer et al. [17]. Krauthgamer, Linial and Magen [18] show a matching $\Omega(\log n)$ lower bound on the $(k-1)$-dimensional distortion for all $k < n^{1/3}$.

In this paper we provide embedding with guarantees on the $(k-1)$-dimensional ℓ_q-distortion for all $q \le \infty$ simultaneously. As a special case, our bounds imply the best possible worst case $(k-1)$-dimensional distortion of $O(\log n)$ (matching the result of [17]).

Average and ℓ_q Distortion. The notions of average distortion and ℓ_q-distortion is tightly related to the notions of partial embeddings and scaling embedding[7], which demand strong guarantees for a $(1-\epsilon)$ fraction of the pairwise distances. These notions were introduced by Kleinberg, Slivkins and Wexler [15], largely motivated by the study of distances in computer networks.

In [1] partial embedding into L_p with tight $O(\log 1/\epsilon)$ partial distortion were given. The embedding method of [2] 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 ℓ_q-distortion bounded by $O(q)$, while maintaining the best worse 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 [2], and in addition have been used in solving graph theoretic problems [9]. Following [15,1,2] related notions have been studied in various contexts [6,16,3,7].

2 Robustness of the Metric Volume

Proof of Theorem 2.

For a tree T on n vertices $\{v_1, \ldots, v_n\}$ let $\overline{\text{vol}}(T)$ be the product of the edge lengths. Because of the matroid exchange property, this product is minimized by an MST. Thus for any metric space on points $\{v_1, \ldots, v_n\}$ and any spanning tree T, $\text{vol}_F(v_1, \ldots, v_n) \le \overline{\text{vol}}(T)/(n-1)!$; the inequality is saturated by any (and only a) minimum spanning tree.

---
[7] alternatively known as embeddings with slack and embeddings with gracefully degrading distortion.

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 $\Delta(F) = \max_{1 \le i \le m} \Delta(T_i)$. For a metric space (X, d) let $\Delta_k(X) = \min\{\Delta(F) \mid F \text{ is a spanning forest of } X \text{ with } k \text{ connected components}\}$.*

Lemma 1. *Let (X, d) be a metric space. Let $k \ge 1$. 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 $k \ge 2$, 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 $k - 1$.)

Corollary 1. *For any n-point metric space (X, d) and any FST T' for X, $\overline{\text{vol}}(T') \le \prod_{k=1}^{n-1} \Delta_k(X)$.*

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

Using Corollary 1, our proof of Theorem 2 reduces to showing that for any MST T of X, $\prod_{k=1}^{n-1} \Delta_k(X) \le e^{O(n-1)}\overline{\text{vol}}(T)$. Specifically we shall show that for any spanning tree T,

$$\prod_{k=1}^{n-1} \Delta_k(X) \le \frac{1}{n^2} \left(\frac{4\pi^2}{3}\right)^{n-1} \overline{\text{vol}}(T).$$

(Observe incidentally that the FST created by the Gonzalez [11] and Hochbaum-Shmoys [13] process has $\overline{\text{vol}}$ 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: $\sum_{j \ge 1} 1/j^2 = \pi^2/6$.

Edge selection. Find a diametric path γ of T, i.e., a simple path whose length $|\gamma|$ equals the diameter $\Delta(T)$. For appropriate $\ell \ge 2$ let $u_1, \dots, u_\ell$ be the weights of the edges of γ in the order they appear on the path. Select the j'th edge on the path, for a $1 \le j \le \ell$ for which $u_j/|\gamma| > 1/(2(\pi^2/6)\min\{j, \ell+1-j\}^2)$. Such an edge exists, as otherwise $\sum_1^\ell u_j \le (6/\pi^2)|\gamma| \sum_1^\ell j^{-2} < |\gamma|$. Without loss of generality $j \le \ell + 1 - j$ (otherwise flip the indexing on γ), 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

$$\prod_{1}^{n-1} \Delta(F_k(T)) \leq \frac{1}{n^2}\left(\frac{4\pi^2}{3}\right)^{n-1} \overline{\mathrm{vol}}(T).$$

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

$$\prod_{1}^{n-1} \Delta(F_k(T)) \leq e^{c_1(n-1)-c_2\log n}\overline{\mathrm{vol}}(T),$$

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

$$\prod_{2}^{n-1} \Delta(F_k(T)) = (\prod_{1}^{m} \Delta(F_k(T_1)))(\prod_{1}^{n-2-m} \Delta(F_k(T_2))).$$

Therefore:

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

Now by induction:

$$\prod_{1}^{n-1} \Delta(F_k(T)) \leq \Delta(T)\cdot e^{c_1 m - c_2\log(m+1)}\cdot\overline{\mathrm{vol}}(T_1)\cdot e^{c_1(n-2-m)-c_2\log(n-1-m)}\cdot\overline{\mathrm{vol}}(T_2).$$

As $\overline{\mathrm{vol}}(T) = u_j\cdot\overline{\mathrm{vol}}(T_1)\overline{\mathrm{vol}}(T_2)$ we get

$$\begin{aligned}\prod_{1}^{n-1} \Delta(F_k(T)) &\leq (\Delta(T)/u_j)\cdot\exp\{c_1(n-2) - c_2(\log(m+1) + \log(n-1-m))\}\,\overline{\mathrm{vol}}(T)\\ &\leq \exp\left\{\log(2(\pi^2/6)j^2) + c_1(n-2) - c_2(\log(m+1)+\log(n-1-m))\right\}\overline{\mathrm{vol}}(T)\\ &\leq \exp\left\{\log(\pi^2 j^2/3) + c_1(n-2) - c_2(\log j + \log(n/2))\right\}\overline{\mathrm{vol}}(T)\\ &\leq \exp\left\{c_1(n-1) - c_2\log n - (c_2-2)\log j - (c_1 - c_2\log 2 - \log(\pi^2/3))\right\}\overline{\mathrm{vol}}(T)\end{aligned}$$

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

$$\ldots \leq \exp\{c_1(n-1) - c_2 \log n\} \overline{\mathrm{vol}}(T).$$

□

3 Volume Preserving Embeddings

In this section we prove Theorem 3. In [2] a general framework for embedding metrics into normed spaces was introduced. In particular we define an embedding $\hat{f} : X \to L_2$ in $O(\log n)$ dimensions and show that the (pairwise) distortion is $O(\log n)$ and the ℓ_q-distortion is $O(q)$. Here, we extend this work to apply to sets of points of higher cardinality. We use the same map of [2] while taking more dimensions: $O(k \log n)$, so the map has the stronger property of being volume preserving. In Section 3.2 to Section 3.4 we give Lemma 2 which states a property of this embedding which allows us to prove the $(k-1)$-dimensional distortion bounds of the embedding, followed by the analysis of these bounds for the worst (ℓ_∞-distortion) case and average (ℓ_1-distortion) case (the general ℓ_q-distortion case is deferred to the full version). The content of Section 3.2 to Section 3.4 does not require knowledge of the definition of the embedding beyond its properties given in Lemma 2. However the proof of this lemma is based on the definition of the embedding given in Section 3.1 and its proof is deferred to the full version.

3.1 The Embedding

The embedding of [2] is partition-based [4,20]. It is constructed by concatenating $O(k \log n)$ random maps $X \to \mathbb{R}$ where each such map is formed by summing terms over all scales, where each scale is a an embedding created using an approach similar to [20], using the uniform probabilistic partition techniques of [2]. See the full version for the required definitions and notations for uniform probabilistic partitions and their properties. In a nut-shell, a partition of X is a pair-wise disjoint collection of clusters covering X. In a Δ bounded partition the diameter of every cluster is at most Δ, and the (η, δ)-padding property of a distribution over Δ-bounded partitions is that for all $x \in X$, $\Pr[B(x, \eta\Delta) \subseteq P(x)] \geq \delta$ (where $P(x)$ denotes the cluster containing x).

Let $D = c \cdot k \log n$, $\Delta_0 = \mathrm{diam}(X)$, $I = [\lceil \log_4(\Delta_0) \rceil]$, where c is a constant to be determined later. For all $j \in I$, $\Delta_j = \Delta_0/4^j$. Fix some $h \in [D]$. For all $j \in I$ create a Δ_j-bounded $(\eta_j, 1/2)$-padded probabilistic partition $P_j^{(h)}$ sampled from a certain distribution $\hat{\mathcal{P}}_j$ over a set of partitions $\mathcal{P}_j$ (for details see [2]). This distribution $\hat{\mathcal{P}}_j$ is accompanied by a collection of uniform functions[8] $\{\xi_P : X \to \{0,1\} \mid P \in \mathcal{P}_j\}$ and $\{\eta_P : X \to (0,1] \mid P \in \mathcal{P}_j\}$. Roughly speaking, $\eta_P(x)$ is the inverse logarithm of the local growth rate of the space in the cluster containing x, and $\xi_P(x)$ is an indicator for sufficient local growth rate. Define for $x \in X$, $0 < j \in I$, $\phi_j^{(h)} : X \to \mathbb{R}^+$, by

$$\phi_j^{(h)}(x) = \xi_{P_j^{(h)}}(x) / \eta_{P_j^{(h)}}(x) \ .$$

[8] A function f is uniform with respect to a partition P if for any $x, y \in X$, $P(x) = P(y)$ implies that $f(x) = f(y)$.

Let $\{\sigma_j^{(h)}(A) | A \in P_j, 0 < j \in I\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. Define the embedding $f : X \to L_2^D$ by defining for all $h \in [D]$ a function $f^{(h)} : X \to \mathbb{R}_+$ and let $f = D^{-1/2} \bigoplus_{h \in D} f^{(h)}$. For all $j \in I$ define $f_j^{(h)} : X \to \mathbb{R}_+$ and let $f^{(h)} = \sum_{j>0} f_j^{(h)}$. For $x \in X$ define

$$f_j^{(h)}(x) = \sigma_j^{(h)}(P_j^{(h)}(x)) \cdot \min\{\phi_j^{(h)}(x) \cdot d(x, X \setminus P_j^{(h)}(x)), \Delta_j\} ,$$

Finally we let $\hat{f} = C \cdot f$ be scaled version of f, where $C > 1$ is a universal constant.

3.2 Analyzing the $(k-1)$-dimensional distortion

In what follows we give the necessary definitions and state the main technical lemma (Lemma 2) that summarizes the distortion properties of the embedding needed to prove Theorem 3. We start with some definitions.

Definition 3. *For a point $x \in X$ and radius $r \geq 0$ let $B(x,r) = \{y \in X | d(x,y) \leq r\}$. For $x \in X$, and $\epsilon > 0$ let $r_\epsilon(x)$ be the minimal radius r such that $|B(x,r)| \geq \epsilon n$.*

In [2] it was shown that $1 \leq \frac{\|\hat{f}(x)-\hat{f}(y)\|_2}{d(x,y)} \leq O(\log(1/\epsilon))$, for any $0 < \epsilon < 1/2$ such that $\min\{r_{\epsilon/2}(x), r_{\epsilon/2}(y)\} < d(x,y)$. In this section we generalize the analysis to sets of size k: First we define the ϵ values for a set, then in Lemma 2 we show an analogue for pair distortion on some pairs in the set (even a stronger bound is given, with respect to the affine span), then we show in Lemma 4 that the volume distortion of a set S is bounded and finally conclude the appropriate bounds on the various ℓ_q distortions.

For any sequence $S = (s_0, s_1, \ldots s_{k-1})$, define a sequence $\epsilon^{(S)} = (\epsilon_1^{(S)}, \ldots, \epsilon_{k-1}^{(S)})$ as follows. For any $i \in \{1, \ldots, k-1\}$ let $t(i) \in \{0, \ldots, i-1\}$ be the index of the point satisfying $d(s_i, \{s_0, \ldots, s_{i-1}\}) = d(s_i, s_{t(i)})$, then $\epsilon_i^{(S)} = 2^{-j}$ where j is the minimal integer such that $\min\left\{r_{\epsilon_i^{(S)}/2}(s_i), r_{\epsilon_i^{(S)}/2}(s_{t(i)})\right\} < d(s_i, s_{t(i)})$. In other words, if $\epsilon_i^{(S)} = 2^{-j}$ then either the radius of B_1 or the radius of B_2 is smaller than $d(s_i, s_{t(i)})$ where B_1 is the ball around s_i that contains $n2^{-(j+1)}$ points and B_2 is the ball around $s_{t(i)}$ that contains $n2^{-(j+1)}$ points.

Lemma 2. *Let (X,d) be an n point metric space, $2 \leq k \leq n$, and let $\hat{f} : X \to L_2$ be the embedding defined in Section 3.1. Then with high probability, for any $S \in \binom{X}{k}$ there exists an ordering $S = (s_0, s_1, \ldots s_{k-1})$ such that for all $1 \leq i < k$:*

$$1 \leq \frac{d_E(\hat{f}(s_i), \text{affspan}(\hat{f}(s_0), \ldots, \hat{f}(s_{i-1})))}{d(s_i, \{s_0, \ldots, s_{i-1}\})} \leq O(\log(1/\epsilon_i^{(S)})),$$

where d_E denotes the Euclidean distance.

We defer the proof of Lemma 2 to the full version. In what follows we show that this lemma implies the desired distortion bounds for the embedding.

From now on fix some $2 \leq k \leq n$. We first make use of Theorem 2 to bound the $(k-1)$-dimensional distortion of each set $S \in \binom{X}{k}$ as a function of $\epsilon^{(S)}$ implied by Lemma 2. We start by showing that the embedding is (volume) non-contractive.

Lemma 3. *The embedding $\hat{f}$ is (volume) non-contractive.*

Proof. Fix some $S = (s_0, \ldots, s_{k-1})$. Let $q_i = d_E(\hat{f}(s_i), \text{affspan}(\hat{f}(s_0), \ldots, \hat{f}(s_{i-1}))$. By definition, $\text{vol}_E(\hat{f}(S)) = \frac{\prod_{i=1}^{k-1} q_i}{(k-1)!}$. By the lower bound part of Lemma 2:

$$\text{vol}_F(S) \leq \frac{1}{(k-1)!} \prod_{i=1}^{k-1} d(s_i, \{s_0, \ldots, s_{i-1}\}) \leq \frac{\prod_{i=1}^{k-1} q_i}{(k-1)!} = \text{vol}_E(\hat{f}(S))$$

Lemma 4. *For any* $S \in \binom{X}{k}$ *:*

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

Proof. Let $S = (s_0, \ldots, s_{k-1})$ be the sequence determined by Lemma 2. By definition, $\text{vol}_E(\hat{f}(S)) = \frac{\prod_{i=1}^{k-1} q_i}{(k-1)!}$, where $q_i = d_E(\hat{f}(s_i), \text{affspan}(\hat{f}(s_0), \ldots, \hat{f}(s_{i-1}))$. By the upper bound part of Lemma 2:

$$\text{vol}_E(\hat{f}(S)) = \frac{\prod_{i=1}^{k-1} q_i}{(k-1)!} \leq \frac{1}{(k-1)!} \prod_{i=1}^{k-1} \left(c_1 \log(1/\epsilon_i^{(S)}) \cdot d(s_i, \{s_0, \ldots, s_{i-1}\})\right),$$

where c_1 is an appropriate constant. Now Theorem 2 guarantees that:

$$\frac{1}{(k-1)!} \prod_{i=1}^{k-1} d(s_i, \{s_0, \ldots, s_{i-1}\}) \leq c_2^{k-1} \cdot \text{vol}_F(S),$$

for an appropriate constant c_2, implying that:

$$\text{vol}_E(\hat{f}(S)) \leq (c_1 c_2)^{k-1} \cdot \text{vol}_F(S) \prod_{i=1}^{k-1} \log(1/\epsilon_i^{(S)}).$$

3.3 Analyzing the worst case (ℓ_∞) volume distortion

Lemma 5. *The* $(k-1)$*-dimensional distortion of* $\hat{f}$ *is* $O(\log n)$ i.e. $\text{dist}_\infty^{(k-1)}(\hat{f}) = O(\log n)$.

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

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

3.4 Analyzing the average (ℓ_1) volume distortion

Lemma 6. *The average* $(k-1)$*-dimensional distortion of* $\hat{f}$ *is* $O(\log k)$ i.e. $\text{dist}_1^{(k-1)}(\hat{f}) = O(\log k)$.

Proof. Define for every $S \in \binom{X}{k}$, and for any $s_i \in S$, $\hat{\epsilon}_i^{(S)}$ as a power of $1/2$ and the maximal such that $d(s_i, S \setminus \{s_i\}) > r_{\hat{\epsilon}_i^{(S)}/2}(s_i)$. By definition $r_{\hat{\epsilon}_i^{(S)}/2}(s_i) \leq r_{\epsilon_i^{(S)}/2}(s_i)$

and hence, $\hat{\epsilon}_i^{(S)} \leq \epsilon_i^{(S)}$. Let C be an appropriate constant. The average $(k-1)$-dimensional distortion can be bounded as follows

$$\begin{aligned}\frac{\mathrm{dist}_1^{(k-1)}(\hat{f})}{C} &\leq \mathbb{E}_{S\in\binom{X}{k}}\left[\left(\prod_{i=1}^{k-1}\log(1/\epsilon_i^{(S)})\right)^{1/(k-1)}\right] \\ &\leq \mathbb{E}\left[\left(\prod_{i=1}^{k-1}\log(1/\hat{\epsilon}_i^{(S)})\right)^{1/(k-1)}\right] \\ &\leq \mathbb{E}\left[\frac{1}{k-1}\sum_{i=1}^{k-1}\log(1/\hat{\epsilon}_i^{(S)})\right] \\ &= \frac{1}{k-1}\sum_{i=1}^{k-1}\mathbb{E}\left[\log(1/\hat{\epsilon}_i^{(S)})\right]\end{aligned}$$

using the arithmetic-geometric mean inequality and the linearity of expectation.

For every set $S \in \binom{X}{k}$ let $m = m(S)$ be the maximal integer such that for all $i \in \{0,1,\ldots,k-1\}$, $B(s_i, r_{m/n}(s_i)) \cap S = \{s_i\}$. That is, for *every* point $s \in S$ the first $m-1$ nearest neighbors (in X) of s, are not in S. Since $\hat{\epsilon}_i^{(S)} \geq m/(2n)$ then $\mathbb{E}\left[\log(1/\hat{\epsilon}_i^{(S)})\right] \leq \mathbb{E}\left[\log(n/m)\right] + 1$, for all $S \in \binom{X}{k}$ and $i \in \{0,1,\ldots,k-1\}$ (note that here the range for i includes the first index 0). We now proceed to bound $\mathbb{E}\left[\log(n/m)\right]$. Let $A(s,t)$ be the event that the t-th nearest neighbor of a point $s \in S$ is also in S (using a consistent lexicographic order on the points so that the t-th nearest neighbor is unique). The probability that $A(s,t)$ occurs is exactly $(k-1)/(n-1)$, since given that $s \in S$ there are $k-1$ additional points to choose for S uniformly at random. Hence by union bound

$$\Pr[m(S)=t] \leq \Pr[\bigcup_{s\in S} A(s,t)] \leq \sum_{s\in S}\Pr[A(s,t)] \leq \frac{k(k-1)}{n-1}.$$

Let $h = \lceil \frac{n}{k^2} \rceil$, it follows that $\Pr[m(S)=t] \leq 2/h$. Hence,

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

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

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

References

1. 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.

2. I. Abraham, Y. Bartal, and O. 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.
3. 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.
4. 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.
5. J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.
6. 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.
7. 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.
8. J. Dunagan and S. Vempala. On Euclidean embeddings and bandwidth minimization. *Lecture Notes in Computer Science*, 2129, 2001.
9. Michael Elkin, Christian Liebchen, and Romeo Rizzi. New length bounds for cycle bases. *Information Processesing Letters*, 104(5):186–193, 2007.
10. 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).
11. T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.
12. A. Gupta. Improved bandwidth approximation for trees and chordal graphs. *J. Algorithms*, 40(1):91–36, 2001.
13. D. S. Hochbaum and D. B. Shmoys. A best possible heuristic for the k-center problem. *Math. Oper. Res.*, 10:180–184, 1985.
14. Makoto Imase and Bernard M. Waxman. Dynamic steiner tree problem. *SIAM J. Discrete Math.*, 4(3):369–384, 1991.
15. J. Kleinberg, A. Slivkins, and T. Wexler. Triangulation and embedding using small sets of beacons. In *J. ACM*, volume 56, pages 1–37, New York, NY, USA, 2009. ACM.
16. 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.
17. 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.
18. R. Krauthgamer, N. Linial, and A. Magen. Metric embeddings–beyond one-dimensional distortion. *Discrete Comput. Geom.*, 31(3):339–356, 2004.
19. 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.
20. 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.