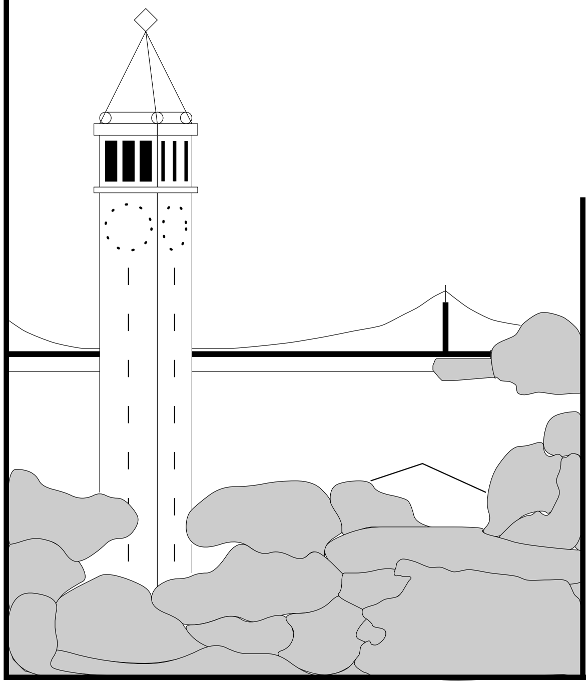
Metric embeddings - beyond one-dimensional distortion

Robert Krauthgamer[1] *Nathan Linial*[2] *Avner Magen*[3]



Report No. UCB/CSD-02-1181

May 15, 2002

Computer Science Division (EECS)
University of California
Berkeley, California 94720

Metric embeddings - beyond one-dimensional distortion

Robert Krauthgamer[§] Nathan Linial[¶] Avner Magen[‖]

May 15, 2002

Abstract

Metric spaces and their embeddings have recently played a prominent role in the development of new algorithms. So far, most of the emphasis was on embeddings that preserve pairwise distances. A very intriguing concept introduced by Feige [Fei00], allows us to quantify the extent to which higher-dimensional structures are preserved by a given embedding. We investigate this concept for several basic graph families such as paths, trees, cubes and expanders.

1 Introduction

Finite metric spaces and their embeddings have received much attention in recent years. Ideas from this area have led to a number of beautiful algorithmic applications, see e.g. the survey by Indyk [Ind01] and Matoušek's book [Mat01, Chapter 15]. Much of the activity in this area revolves around the following general problem: A given finite metric space (X, d) is to be "approximated" as well as possible by a metric space from a fixed class $\mathcal{C}$ of "simple" metric spaces. Typical classes $\mathcal{C}$ include Euclidean, other l_p, and tree metrics. Traditionally, similarity among metric spaces has been investigated in terms of low-distortion embeddings. Namely, a mapping $\varphi : X \to Y$ with $(Y, \rho) \in \mathcal{C}$, where for every two points $x_1, x_2 \in X$, the distance $\rho(\varphi(x_1), \varphi(x_2))$ is very close to the original distance $d(x_1, x_2)$. Since a metric is a bivariate function $d : X \times X \to \mathbb{R}$, it seemed almost self-evident that "approximation" could not mean anything beyond similar pairwise distances.

In a remarkable paper, Uri Feige [Fei00] woke us from our dogmatic slumber and pointed out the possibility of considering higher-order structures in metric spaces. This depends on a comparison with Euclidean metrics, where volume can be defined. To this end, every k-subset of (X, d) is associated with a $(k-1)$-*dimensional volume*, where the one-dimensional volume of $\{x_1, x_2\} \subseteq X$ is simply $d(x_1, x_2)$. A proper definition of higher-dimensional volumes allows us to quantify the extent

[*]ICSI and Computer Science Division, University of California, Berkeley, CA 94720. Part of this work done while the author was at the Hebrew University of Jerusalem. Supported in part by NSF grants CCR-9820951 and CCR-0121555 and DARPA cooperative agreement F30602-00-2-0601. Email: robi@cs.berkeley.edu

[†]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email: nati@cs.huji.ac.il

[‡]NEC Research Institute, Princeton, NJ. Part of this work was done while at the Hebrew University in Jerusalem. E-mail: avner@research.nj.nec.com

[§]ICSI and Computer Science Division, University of California, Berkeley, CA 94720. Part of this work done while the author was at the Hebrew University of Jerusalem. Supported in part by NSF grants CCR-9820951 and CCR-0121555 and DARPA cooperative agreement F30602-00-2-0601. Email: robi@cs.berkeley.edu

[¶]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email: nati@cs.huji.ac.il

[‖]NEC Research Institute, Princeton, NJ. Part of this work was done while at the Hebrew University in Jerusalem. E-mail: avner@research.nj.nec.com

to which such volumes are distorted by a given embedding $\varphi : X \to l_2$. Let us recall the definition: Let S be a set of k points in a metric space (X, d). The *volume* $\mathrm{Vol}(S)$ of S is defined as the maximum of the Euclidean volume $\mathrm{Evol}(\Psi(S))$ over all nonexpansive embeddings $\Psi : S \to l_2$. (If K is a set of k points in l_2, then $\mathrm{Evol}(K)$ denotes the $(k-1)$-dimensional Lebesgue measure of K's convex hull. An embedding $\varphi : X \to l_2$ is *nonexpansive* if $\|\varphi(x_1) - \varphi(x_2)\| \le d(x_1, x_2)$ for all $x_1, x_2 \in X$.) We say that a nonexpansive $\Phi : X \to l_2$ *distorts the volume of* S by $\left(\frac{\mathrm{Vol}(S)}{\mathrm{Evol}(\Phi(S))}\right)^{\frac{1}{k-1}}$. The maximum over all k-subsets S, is called the $(k-1)$-*dimensional distortion* of Φ. One-dimensional distortion coincides with what is simply called distortion in the literature.

Some of the basic phenomena known from the study of one-dimensional distortions have similar higher-order counterparts, though usually new ideas are required to prove them in the more general case. For example, a fundamental theorem of Bourgain [Bou85] says that every n-point metric space can be embedded in l_2 with distortion $O(\log n)$. Feige [Fei00] shows that a suitable adaptation of Bourgain's embedding yields the same bound also for r-dimensional distortion with any $r = k-1 \le O(\frac{\log n}{\log \log n})$, see Table 1. Bourgain's (one-dimensional) bound is tight, as it is shown in [LLR95, Mat97, LM00] that every embedding of (the metric of) an n-vertex constant-degree expander graph has (one-dimensional) distortion $\Omega(\log n)$. Here we show a similar lower bound for higher-dimensional distortions, namely, for any $r \le n^{1/3}$. For simplicity, we did not optimize the upper bounds on r, and so the constants in the exponent (such as 1/3 or the 1/4 below) are somewhat arbitrary.

However, in some other aspects of the theory the one-dimensional situation differs completely from the general situation. For example, there clearly exist graphs whose metric embeds in l_2 with constant (one-dimensional) distortion, but, perhaps surprisingly, it turns out that a constant higher-order distortion is an extremely stringent requirement. Specifically, for any $2 \le r < \sqrt{n}/2$, a graph can be embedded with bounded r-dimensional distortion if and only if it has a bounded diameter. Another surprise awaits us when we consider trees. Bourgain [Bou86] has determined that the one-dimensional distortion of a complete binary tree of n vertices (and depth roughly $\log n$) is $\Theta(\sqrt{\log \log n})$, and later Matoušek [Mat99] showed that the same upper bound holds for *every* n-vertex tree. For higher-dimensional distortion, the situation is quite different. Every n-vertex tree has a Euclidean embedding with r-dimensional distortion $O((\log n)^{\frac{r-1}{2r}}(\log \log n)^{\frac{1}{2r}})$. This is much larger than the one-dimensional bound and this bound is nearly tight for large constant r. However, this time the bound is nearly attained by an n-vertex *path*, whose r-dimensional distortion is $\Theta(\log^{\frac{r-1}{2r}} n)$ for any $r < \sqrt{n}/2$. Indeed, our work is motivated by the desire to understand how things evolve as we shift our attention from one-dimensional to higher-dimensional distortions. To this end we investigate (the metrics of) several basic graph families: Paths, cubes, trees and expanders. (See Table 1 for a fuller account.) We develop along the way several methods that may, perhaps, become useful in future investigations in this area.

Here are two more findings of this paper. It was known for over thirty years [Enf69] (see also [LM00]) that every embedding of the m-dimensional hypercube has distortion $\ge \sqrt{m}$. It turns out that a similar lower bound holds also for r-dimensional distortions for $r \le 2^{m/4}$. Aside from graph metrics, we also consider Euclidean metrics (the restriction of the l_2 metric to a finite point set). We prove that such a set whose aspect ratio is polynomial in n has r-dimensional distortion at most $O((\log n \log r)^{\frac{r-1}{2r}})$, which offers a slight improvement over a result of Rao [Rao99]. The metric of a path on n vertices shows that this bound is tight for constant $r \ge 2$. Closing the gap for larger r is still an open question.

In the final section we briefly consider an alternative measure for the preservation of large substructure in a metric space under a given embedding.

2 Preliminaries

Volume distortion via tree volume and affine distances. The definition of $\mathrm{Vol}(S)$ is somewhat involved. It is therefore much more convenient to use the following good approximation for it. View an n-point metric space (X,d) as the complete graph K_n where each edge $\{x,y\}$ has weight $d(x,y)$. For every $S \subseteq X$, define $\mathrm{Tvol}(S)$, the *tree volume* of S, as the product of the edge lengths in a minimum spanning tree for S. (Note that this is the same as the least possible product of edge lengths among such trees). It was shown in [Fei00, Theorem 3] that $\mathrm{Vol}(S) \le \frac{\mathrm{Tvol}(S)}{(k-1)!} \le 2^{(k-2)/2} \cdot \mathrm{Vol}(S)$, where $k = |S|$. Since the definition of the $(k-1)$-dimensional volume distortion takes volumes to power $\frac{1}{k-1}$, using $\mathrm{Tvol}(S)$ instead of $\mathrm{Vol}(S)$ in this calculation approximates the true volume distortion within a constant factor.

The $(k-1)$-dimensional volume of the simplex that is the convex hull of $v_1, \ldots, v_k \in l_2$ can be written as $\frac{1}{(k-1)!} \prod_{i=2}^{k} d_{\mathrm{aff}}(v_i; \{v_1, \ldots, v_{i-1}\})$, where $d_{\mathrm{aff}}(v; K)$ denotes the Euclidean distance between the point v and the affine hull of K.

Definition 1. *An ordering $(p_1, \ldots, p_k)$ of the elements in a subset S of (X,d) is said to be* proper *if each initial segment in this order spans a subtree of the minimum spanning tree for S. We note that in this case,* $\mathrm{Tvol}(S) = \prod_{i=2}^{k} d(p_i, \{p_1, \ldots, p_{i-1}\})$.

The above discussion then yields a good estimate for the volume distortion of S under a non-expansive embedding Φ (see also [Fei00, Rao99, DV01]). Namely, the distortion of S under the embedding Φ equals, up to constant factors,

$$\left(\frac{\mathrm{Tvol}(S)}{(k-1)! \cdot \mathrm{Evol}(\Phi(S))}\right)^{\frac{1}{k-1}} = \left(\prod_{i=2}^{k} \frac{d(p_i, \{p_1, \ldots, p_{i-1}\})}{d_{\mathrm{aff}}(\Phi(p_i); \Phi(\{p_1, \ldots, p_{i-1}\}))}\right)^{\frac{1}{k-1}}. \tag{1}$$

This is a geometric mean of expressions that are reminiscent of one-dimensional distortions. We compare the distance from p_i to $\{p_1, \ldots, p_{i-1}\}$ with the distance between the affine hulls of their images under Φ.

Direct sum of embeddings. We need two simple operations on embeddings. First, an embedding into normed space can be multiplied by a real constant. Second, let $\phi_1, \ldots, \phi_t$ be embeddings

Metric family	Lower bound	Upper bound
The path	$\Theta(\log^{\frac{1}{2}-\frac{1}{2(k-1)}} n)$ (Section 3.1)	
The complete tree	$\Theta(\sqrt{\log\log n})$ (Section 3.3)	
Expander graphs	$\Theta(\log n)$ (Section 3.5)	
Bounded diameter graphs	$\Theta(1)$ (Section 3.2)	
The hypercube	$\Omega(\sqrt{\log n})$ (Section 3.6)	$O(\sqrt{\log n}(\log n \log k)^{\frac{1}{2}-\frac{1}{k-1}})$ (Section 3.6)
Trees and chordal graphs	$\Omega(\log^{\frac{1}{2}-\frac{1}{2(k-1)}} n)$ (Path)	$O((\log n)^{\frac{1}{2}-\frac{1}{2(k-1)}}(\log\log n)^{\frac{1}{2(k-1)}})$ (Sect. 3.4)
Planar graphs	$\Omega(\log^{\frac{1}{2}-\frac{1}{2(k-1)}} n)$ (Path)	$O(\sqrt{\log n})$ [Rao99]
Euclidean metrics	$\Omega(\log^{\frac{1}{2}-\frac{1}{2(k-1)}} n)$ (Path)	$O((\log n \log k)^{\frac{1}{2}-\frac{1}{2(k-1)}})$ (Section 4)
General graphs	$\Omega(\log n)$ (Expander)	$\begin{cases} O(\sqrt{\log n}\sqrt{\log n + k \log k}) \text{ [Fei00]} \\ O(\sqrt{\log^3 n \log k}) \text{ [Rao99]} \end{cases}$

Table 1: The $(k-1)$-dimensional distortion of some n-point metric families.

of a metric (X, d) into Euclidean spaces of appropriate dimensions; then the *direct sum* of these embeddings, denoted $\Phi = \oplus_j \phi_j$, is defined by letting $\Phi(x)$ be the concatenation of the vectors $\phi_j(x)$.

Lemma 1. *Let $\phi_1, \ldots, \phi_t$ be nonexpansive embeddings of a finite metric (X, d) into l_2. If $\alpha_1, \ldots, \alpha_t \geq 0$ and $\sum_j {\alpha_j}^2 = 1$, then $\Phi = \oplus_j \alpha_j \phi_j$ is nonexpansive and satisfies:*

(a). For all $S \neq \emptyset$, $x \notin S$, $d^2_{\text{aff}}\Big(\Phi(x); \Phi(S)\Big) \geq \sum_j \alpha_j^2 \cdot d^2_{\text{aff}}\Big(\phi_j(x); \phi_j(S)\Big)$.

(b). For all S of size $k \geq 1$, $\text{Evol}(\Phi(S)) \geq \prod_j (\text{Evol}(\phi_j(S))^{\alpha_j^2}$.

(c). The r-dimensional distortion of Φ is at most $\max_j\{r$-dimensional distortion of $\phi_j\}$.

We remark that [Fei00, Lemma 18] implies $\text{Evol}^{2/(k-1)}(\Phi(S)) \geq \sum_j \alpha_j^2 \cdot \text{Evol}^{2/(k-1)}(\phi_j(S))$, which yields the bound in (b) by the arithmetic-geometric inequality. However, our bounds for general trees and for Euclidean metrics do not follow from either of these two bounds, and we revert to using (a). Note also that (b) gives a simple and intuitive proof that $-\log\det(X)$ is a convex function of symmetric positive semidefinite matrices X. This fact was used in [Fei00] and in [DV01] to determine efficiently whether the constraint $\det(X) \geq c$ holds for symmetric positive semidefinite matrices X and $c > 0$.

Proof of Lemma 1. The embedding Φ is nonexpansive because for every $x, y \in X$ we have $\|\Phi(x) - \Phi(y)\|^2 = \sum_i {\alpha_i}^2 \|\phi_i(x) - \phi_i(y)\|^2 \leq (d(x, y))^2$.

To prove (a), let $S = \{y_1, \ldots, y_k\}$. By definition (of distance between a point and an affine hull) there exist $\lambda_1, \ldots, \lambda_k$ with $\sum_i \lambda_i = 1$ such that

$$\begin{aligned} d^2_{\text{aff}}\Big(\Phi(x); \Phi(S)\Big) &= \Big\|\Phi(x) - \sum_i \lambda_i \Phi(y_i)\Big\|^2 \\ &= \sum_j {\alpha_j}^2 \Big\|\phi_j(x) - \sum_i \lambda_i \phi_j(y_i)\Big\|^2 \\ &\geq \sum_j {\alpha_j}^2 d^2_{\text{aff}}\Big(\phi_j(x); \phi_j(S)\Big). \end{aligned}$$

We prove (b) by induction on $|S|$. The base case $|S| = 1$ holds vacuously because the $\text{Evol}(S) = 0$. To show the inductive step for $|S| \geq 2$, choose some $x \in S$, break S into $(S - x) \cup \{x\}$, use the induction hypothesis on $S - x$, and apply (a) on $S - x$ and x, as follows.

$$\begin{aligned} \text{Evol}^2(\Phi(S)) &= \left(\frac{\text{Evol}(\Phi(S-x)) \cdot d_{\text{aff}}\,(\Phi(x); \Phi(S-x))}{|S| - 1}\right)^2 && (2) \\ &\geq \frac{\Big(\prod_j (\text{Evol}(\phi_j(S-x)))^{2\alpha_j^2}\Big)\Big(\sum_j {\alpha_j}^2 d^2_{\text{aff}}\Big(\phi_j(x); \phi_j(S-x)\Big)\Big)}{(|S| - 1)^2} && (3) \end{aligned}$$

Using the generalized arithmetic mean - geometric mean inequality (namely $\sum_j \beta_j z_j \geq \prod_j {z_j}^{\beta_j}$ whenever $\beta_j, z_j \geq 0$ and $\sum_j \beta_j = 1$) we conclude, as claimed, that

$$\begin{aligned} \text{Evol}^2(\Phi(S)) &\geq \frac{\Big(\prod_j (\text{Evol}(\phi_j(S-x)))^{2\alpha_j^2}\Big)\Big(\prod_j (d_{\text{aff}}\,(\phi_j(x); \phi_j(S-x)))^{2\alpha_j^2}\Big)}{(|S| - 1)^2} \\ &= \prod_j (\text{Evol}(\phi_j(S)))^{2\alpha_j^2}. \end{aligned}$$

Finally, (c) says that $\mathrm{Evol}(\Phi(S)) \geq \min_j \mathrm{Evol}(\phi_j(S))$ which follows easily since (b) bounds $\mathrm{Evol}(\Phi(S))$ from below by some (weighted geometric) average of $\mathrm{Evol}(\phi_j(S))$. $\square$

A comment about symmetry. Let $Aut(G)$ be the automorphism group of a graph G. An embedding f of G into Euclidean space is *transitive* if $\mathrm{Evol}(f(S)) = \mathrm{Evol}(f(\sigma(S)))$ for every $S \subseteq V(G)$ and $\sigma \in Aut(G)$. In particular (and essentially equivalently), it means that all pairwise distances are invariant under automorphisms. Lemma 1 can be used to show that minimum volume distortion can always be obtained by a transitive embedding. This will allows us, for example, to consider only transitive embeddings in the context of proving lower bounds for volume distortion (e.g. in the tightness part of the proof of Theorem 3).

Lemma 2. *For any finite metric and any $r \geq 1$, there is an embedding that achieves minimum distortion of r-dimensional volumes and is also transitive.*

Proof. Let f be an optimal embedding of G with respect to $(k-1)$-dimensional volumes, and let $f^\sigma(v)$ stand for $f(\sigma(v))$. We show below that the direct sum embedding $f^* = \frac{1}{\sqrt{|Aut(G)|}} \bigoplus_{\sigma \in Aut(G)} f^\sigma$ is transitive and has minimum volume distortion.

The fact that the volume distortion of f^* is at most as that of f follows from Lemma 1(c), because $\sum_{\sigma \in Aut(G)} \left(\frac{1}{\sqrt{|Aut(G)|}}\right)^2 = 1$, and because f^σ has the same volume distortion as f.

To show that f^* is transitive we evaluate $(f^*)^\rho$ for arbitrary $\rho \in Aut(G)$.

$$
\begin{aligned}
(f^*)^\rho &= \left(\frac{1}{\sqrt{|Aut(G)|}} \bigoplus_{\sigma \in Aut(G)} f^\sigma \right)^\rho \\
&= \frac{1}{\sqrt{|Aut(G)|}} \bigoplus_{\sigma \in Aut(G)} (f^\sigma)^\rho \\
&= \frac{1}{\sqrt{|Aut(G)|}} \bigoplus_{\sigma \in Aut(G)} f^{\sigma\rho}.
\end{aligned}
$$

As σ ranges over $Aut(G)$, so does $\sigma\rho$. It follows that $(f^*)^\rho$ is attained from f^* by a permutation of coordinates. Therefore, $\mathrm{Evol}(f^*(S)) = \mathrm{Evol}((f^*)^\rho(S))$ for all $\rho \in Aut(G)$, as claimed. $\square$

3 Graph families

3.1 Paths

We now determine the least r-dimensional distortion achievable in an embedding of a path. In Section 3.1.1 we prove the upper bound by showing an embedding, and in Section 3.1.2 we prove a matching lower bound.

Theorem 1. *The metric of a path on n vertices has an embedding whose $(k-1)$-dimensional volume distortion is $O(\log^{\frac{1}{2}-\frac{1}{2(k-1)}} n)$. This bound is tight for any $3 \leq k \leq \sqrt{n}/2$.*

Nearly the same results were recently shown by Dunagan and Vempala [DV01]. They achieve an $O(\sqrt{\log n})$ upper bound using a randomized algorithm. We construct a specific embedding whose distortion is slightly better, and whose geometrical structure is more tractable. Indeed, we subsequently use this structure to embed trees (in Sections 3.3,3.4). Our lower bound slightly improves over the $\Omega(\log^{\frac{1}{2}-\frac{1}{k}} n)$ lower bound sketched in [DV01]. For simplicity, we consider in the lower bound only vertex subsets of size $k \leq \sqrt{n}/2$.

3.1.1 Embedding the path

Our embedding of the path is a modification of an idea due to Bourgain [Bou86], who used it as a building block for his embedding of the complete binary tree. We embed the n-vertex path by taking the last n vertices in Bourgain's embedding of an n^2-vertex path. In accordance with Bourgain's original notation, we denote the vertex set of the path by $V = \{n^2-n+1, \ldots, n^2-1, n^2\}$. The embedding f maps V into $\mathbb{R}^{n^2}$. The s-th coordinate of $f(p)$ is given by

$$f(p)_s = \begin{cases} \sqrt{(p+1-s)/\ln n} & \text{if } 1 \le s \le p, \\ 0 & \text{if } p < s \le n^2. \end{cases}$$

We first bound the one-dimensional distortion of f.

Lemma 3. *The embedding f is nonexpansive and its one-dimensional distortion is $O(1)$.*

Proof. To see that f is nonexpansive it suffices to verify this property for pairs of adjacent vertices. Indeed, for all $1 < q \le n^2$,

$$\begin{aligned} \|f(q) - f(q-1)\| &= (\ln n)^{-1/2} \left(\sum_{s=1}^{q} (\sqrt{q+1-s} - \sqrt{q-s})^2 \right)^{1/2} \\ &= (\ln n)^{-1/2} \left(\sum_{s=1}^{q} \left(\frac{1}{\sqrt{q+1-s} + \sqrt{q-s}} \right)^2 \right)^{1/2} \\ &\le (\ln n)^{-1/2} \left(\sum_{s=1}^{q} \frac{1}{4(q-s)} \right)^{1/2} \\ &\le (\ln n)^{-1/2} (H(q)/4)^{1/2} \\ &< 1. \end{aligned}$$

where $H(m)$ denotes the harmonic sum $1 + \frac{1}{2} + \ldots + \frac{1}{m} = \ln m + \Theta(1)$.

We next bound the contraction of the distance between any two vertices p, q with $n^2 - n < p < q \le n^2$.

$$\begin{aligned} \|f(q) - f(p)\| &\ge (\ln n)^{-1/2} \left(\sum_{s=1}^{p} (\sqrt{q+1-s} - \sqrt{p+1-s})^2 \right)^{1/2} \\ &= (\ln n)^{-1/2} \left(\sum_{s=1}^{p} \left(\frac{q-p}{\sqrt{q+1-s} + \sqrt{p+1-s}} \right)^2 \right)^{1/2} \\ &\ge (\ln n)^{-1/2} (q-p) \left(\sum_{s=1}^{p} \frac{1}{4(q+1-s)} \right)^{1/2} \\ &= \frac{1}{2} (\ln n)^{-1/2} (q-p) \, (H(q) - H(q-p))^{1/2} \\ &= \frac{1}{2} (\ln n)^{-1/2} (q-p) \left(\ln \frac{q}{q-p} + \Theta(1) \right)^{1/2} \\ &\ge \Omega(q-p), \end{aligned}$$

where the last inequality follows from $\frac{q}{q-p} \ge \frac{n^2-n}{n-1} = n$. □

Let $S = \{p_1, \ldots, p_k\}$ be a subset of k vertices of the path, ordered by increasing indices $p_1 < \ldots < p_k$. Clearly, the minimum spanning tree of S consists of the edges (p_{i-1}, p_i) for $i = 2, \ldots, k$. Hence, $\mathrm{Tvol}(S) = \prod_{i=2}^{k}(p_i - p_{i-1})$.

To derive a lower bound on the Euclidean volume of $f(S)$, consider the distance between $f(p_i)$ and the affine hull of $f(p_1), \ldots, f(p_{i-1})$. For $i = 2$ this distance is $\|f(p_2) - f(p_1)\|$ and by Lemma 3 it is at least $\Omega(p_2 - p_1)$. For $i > 2$ we claim that it is $\Omega(\frac{p_i - p_{i-1}}{\sqrt{\log n}})$. Indeed, $f(p_1), \ldots, f(p_{i-1})$ are supported only on coordinates $s \le p_{i-1}$, and so the same applies for their affine hull. Therefore, $d_{\mathrm{aff}}\left(f(i); f(p_1, \ldots, p_{i-1})\right)$ is at least the length of the the projection of $f(p_i)$ onto the coordinates $s > p_{i-1}$, yielding as claimed that

$$d^2_{\mathrm{aff}}\Big(f(p_i); f(p_1, \ldots, p_{i-1})\Big) \ge \sum_{s=p_{i-1}+1}^{p_i} \frac{p_i + 1 - s}{\ln n} \ge \frac{(p_i - p_{i-1})^2}{2 \ln n}.$$

Plugging these into (1) we conclude that the $(k-1)$-dimensional volume distortion of f is $\Big(O(1) \cdot (O(\log n))^{\frac{k-2}{2}}\Big)^{\frac{1}{k-1}} = O(\log^{\frac{k-2}{2k-2}} n)$. This proves the upper bound in Theorem 1.

3.1.2 Lower bound for the path

The key tool for proving the lower bound in Theorem 1 is a geometric inequality, due to Linial and Saks [LS01].

Lemma 4 (Linial and Saks [LS01]). *Let $x_1, \ldots, x_n$ be vectors in $\mathbb{R}^n$, and denote $\Gamma_j = \{(p,q) : 1 \le p < q \le n \ , \ q - p = 2^j\}$ and $\Gamma = \cup_{j \ge 1} \Gamma_j$. Then*

$$\sum_{(p,q)\in\Gamma} \frac{\|x_p - 2x_{(p+q)/2} + x_q\|^2}{(q-p)^2} \le \sum_{p=1}^{n-1} \|x_{p+1} - x_p\|^2. \tag{4}$$

We now denote the vertices of the n-path by $\{1, \ldots, n\}$, as usual. Let ϕ be a nonexpansive embedding and let $x_p = \phi(p)$ be the image of the vertex $p \in \{1, \ldots, n\}$. Since ϕ is nonexpansive, the righthand side of (4) is at most $n - 1$. Hence, there exists $1 \le i \le \frac{1}{2}\log n$, such that

$$\sum_{(p,q)\in\Gamma_i} \frac{\|x_p - 2x_{(p+q)/2} + x_q\|^2}{(2^i)^2} < \frac{2n}{\log n}. \tag{5}$$

Consider a set of $k \le \sqrt{n}/2$ evenly spaced vertices on the path, $S = \{p_1, \ldots, p_k\}$ where $p_{j+1} = p_1 + j \cdot 2^i$ for $j = 1, \ldots, k-1$ (p_1 will be chosen later). Every three consecutive vertices p_{j-2}, p_{j-1}, p_j (for $3 \le j \le k$) are mapped to a triangle $x_{p_{j-2}}, x_{p_{j-1}}, x_{p_j}$ in the Euclidean space. Let $m_j = \|\frac{1}{2}(x_{p_{j-2}} + x_{p_j}) - x_{p_{j-1}}\|$ be the length of this triangle's median from $x_{p_{j-1}}$. We follow an idea from [DV01] and upper bound the volume of $\mathrm{conv}(x_{p_1}, \ldots, x_{p_k})$ by $\mathrm{Evol}(\phi(S)) \le \frac{1}{(k-1)!}\|x_{p_2} - x_{p_1}\| \prod_{j=3}^{k}(2m_j)$. Indeed, denote by h_j the triangle's height from x_{p_j}. Then $h_j = d_{\mathrm{aff}}\left(x_{p_j}; \{x_{p_{j-2}}, x_{p_{j-1}}\}\right) \ge d_{\mathrm{aff}}\left(x_{p_j}; \{x_{p_1}, \ldots, x_{p_{j-1}}\}\right)$ and so $\mathrm{Evol}(\phi(S)) \le \frac{1}{(k-1)!}\prod_{j=3}^{k} h_j$. But similarity of triangles (see Fig. 1) implies that $h_j \le 2m_j$, and the above upper bound on $\mathrm{Evol}(\phi(S))$ follows.

Since ϕ is nonexpansive, $\|x_{p_2} - x_{p_1}\| \le p_2 - p_1 = 2^i$ and we have

$$(k-1)! \cdot \mathrm{Evol}(\phi(S)) \le 2^i \prod_{j=3}^{k}(2m_j) \le 2^i \left(\frac{\sum_{j=3}^{k} 4{m_j}^2}{k-2}\right)^{\frac{k-2}{2}}, \tag{6}$$

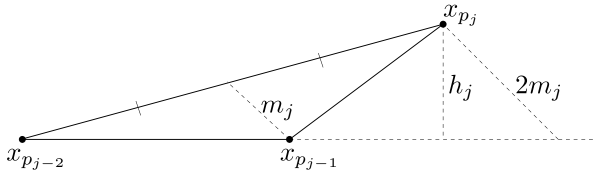


Figure 1: Similarity of triangles.

where the last step uses the arithmetic-geometric inequality.

Suppose that p_1 is chosen uniformly at random from $\{1, 2, \ldots, n/2\}$. (Note that $p_k < n/2 + k\sqrt{n} \le n$.) Inequality (5) yields an upper bound on the expectation of $4{m_j}^2$ (over the choice of p_1), as follows: $\mathbb{E}[4{m_j}^2] = \mathbb{E}\left[\|x_{p_{j-2}} - 2x_{p_{j-1}} + x_{p_j}\|^2\right] \le \frac{(2^i)^2 \cdot 2n}{(n/2)\cdot \log n} = \frac{4\cdot(2^i)^2}{\log n}$. By linearity of expectation, this is also an upper bound on $\mathbb{E}\left[\sum_{j=3}^k 4{m_j}^2/(k-2)\right]$. Therefore there must exist a choice for p_1 for which $\frac{\sum_{j=3}^k 4{m_j}^2}{k-2} \le \frac{4\cdot(2^i)^2}{\log n}$. Together with Equation (6) and the obvious equality $\mathrm{Tvol}(S) = (2^i)^{k-1}$, we lower bound the distortion of S by

$$\left(\frac{\mathrm{Tvol}(S)}{(k-1)! \cdot \mathrm{Evol}(\phi(S))}\right)^{\frac{1}{k-1}} \ge \left(\frac{(2^i)^{k-1}}{(2^i)^{k-1}(4/\log n)^{\frac{k-2}{2}}}\right)^{\frac{1}{k-1}} = \left(\frac{\log n}{4}\right)^{\frac{k-2}{2(k-1)}}.$$

This establishes the tightness claim in Theorem 1.

3.2 Graphs with constant volume distortion

Theorem 2 below allows us to characterize those graphs that can be embedded into l_2 with bounded r-dimensional volume distortion (for any $2 \le r < \sqrt{n}/2$). It follows that the problem of maintaining distances is qualitatively different than that of maintaining higher-dimensional volumes. Interestingly, it also follows that the metric of a graph G has bounded r-dimensional distortion for some $2 \le r < \sqrt{n}/2$ if and only if the same holds for $r = 2$.

Theorem 2. *A graph has a Euclidean embedding whose r-dimensional volume distortion is bounded for some $2 \le r < \sqrt{n}/2$ if and only if the graph has a bounded diameter.*

Proof. Suppose that the diameter of G is $D = O(1)$. We then embed the each vertex i with $\frac{1}{\sqrt{2}}e_i$, where e_i is the ith standard unit vector. This embedding ϕ is obviously nonexpansive. Consider a subset S of $k \ge 3$ vertices. By symmetry of the vectors e_i, $\mathrm{Evol}(\phi(S)) = 2^{-(k-1)/2}\mathrm{Evol}(e_1, \ldots, e_k)$. It can be seen that $(k-1)! \cdot \mathrm{Evol}(\phi(S)) = 2^{-(k-1)/2}\sqrt{k} > 2^{-(k-1)/2}$, e.g. by observing that $d_{\mathrm{aff}}(e_i; \{e_1, \ldots, e_{i-1}\}) = \|e_i - \frac{e_1+\ldots+e_{i-1}}{i-1}\| = \sqrt{i/(i-1)}$. The volume distortion of S is thus $\left(\frac{\mathrm{Tvol}(S)}{(k-1)!\cdot\mathrm{Evol}(\phi(S))}\right)^{\frac{1}{k-1}} < \left(\frac{D^{k-1}}{2^{-(k-1)/2}}\right)^{\frac{1}{k-1}} = \sqrt{2} \cdot D$. The converse follows by applying the lower bound in Theorem 1 to a geodetic path that realizes the diameter of G. ∎

3.3 A complete tree

We next determine the least r-dimensional distortion achievable in an embedding of a complete tree (of any arity $a \ge 2$). Our results below extend the results of Bourgain [Bou86] from one-dimensional to higher-dimensional distortion.

Theorem 3. *The metric of a complete tree of depth D and arity $a \geq 2$ has an embedding whose r-dimensional volume distortion is $O(\sqrt{\log D})$. This bound is tight for any $1 \leq r \leq 2^{\sqrt{D}/4}$.*

Proof. We first prove the upper bound by showing an embedding. To embed a tree T of depth D, we start with Bourgain's embedding [Bou86] of a complete tree of depth D^2 and restrict it to a subtree that is rooted at a vertex of depth $D^2 - D$. The vertex set V of a complete tree of depth D^2 and arity a is identified with the set of strings of length $\leq D^2$ over the alphabet $[a]$. To restrict to the aforementioned subtree, we consider only strings that start with $D^2 - D$ ones. Let $\text{prefix}(p)$ denote the set of prefixes of a string $p \in V$ (including p), i.e. the vertices s on the path from the root to p. Let $\text{depth}(s)$ denote the depth of a vertex s, i.e. the length of the string s. The following map g embeds the tree vertices into $\mathbb{R}^V$ (namely, each a vertex of T is associated with a coordinate in the map g).

$$g(p)_s = \begin{cases} \sqrt{(\text{depth}(p) + 1 - \text{depth}(s))/\log D} & \text{if } s \in \text{prefix}(p), \\ 0 & \text{otherwise.} \end{cases}$$

Note that g is nonexpansive. This has to verified only for adjacent vertices, in which case the calculation in Lemma 3 applies.

To analyze the volume distortion of the embedding g, consider a set S of k vertices. Let $(p_1, \ldots, p_k)$ be the vertices of S ordered by increasing depth, breaking ties arbitrarily. We claim that this ordering is proper. This claim clearly follows if, for every i, there exists a minimum spanning tree of $p_1, \ldots, p_i$ in which p_i is a leaf. To show this, consider a minimum spanning tree for $p_1, \ldots, p_i$ in which p_i has degree ≥ 2, say it's adjacent to p_α and p_β. Let v be the median vertex in T for the triple p_i, p_α, p_β, see Fig 2 for illustration. Since p_i has the largest depth of the three vertices, it follows that $d_T(v, p_i) \geq \min(d_T(v, p_\alpha), d_T(v, p_\beta))$. Consequently, $\max(d_T(p_\alpha, p_i), d_T(p_\beta, p_i)) \geq d_T(p_\alpha, p_\beta)$. Now, add the edge (p_α, p_β) to the spanning tree and omit the longer among (p_i, p_α) and (p_i, p_β). This step does not increase the weight of the spanning tree, and reduces the number of p_i's neighbors. This can be repeated until p_i becomes a leaf in a minimum spanning tree for $p_1, \ldots, p_i$.

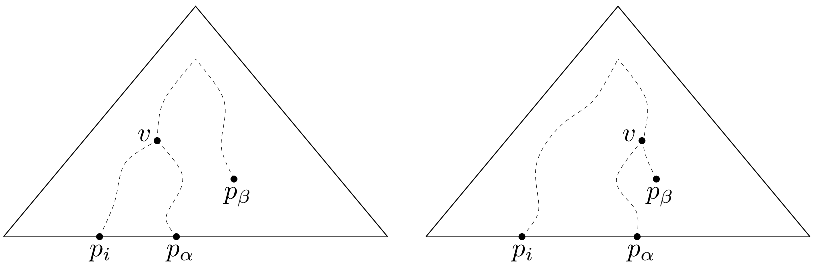


Figure 2: Examples for the median v of the triple p_i, p_α, p_β.

We are now ready to use the estimate in Equation (1) for volume distortion. Fixing the index i for the moment, let y_j be the lowest common ancestor of p_j and p_i, for $j < i$. Let y_ν have the largest depth among $y_1, \ldots, y_{i-1}$. Now p_i is at least as deep as any vertex in $\{p_1, \ldots, p_{i-1}\}$, so $d(p_i, \{p_1, \ldots, p_{i-1}\}) \leq d(p_i, p_\nu) \leq 2d(p_i, y_\nu)$. On the other hand, the vectors $g(p_1), \ldots, g(p_{i-1})$ are not supported on the coordinates that correspond to vertices on the path between p_i and y_ν. We argue as in Section 3.1.1 and conclude that the distance between $g(p_i)$ and the affine hull of $g(p_1), \ldots, g(p_{i-1})$ is at least $\Omega(\frac{\text{depth}(p_i) - \text{depth}(y_\nu)}{\sqrt{\log D}}) = \Omega(\frac{d(p_i, y_\nu)}{\sqrt{\log D}})$. Therefore, $\frac{d(p_i, \{p_1, \ldots, p_{i-1}\})}{d_{\text{aff}}(g(p_i); g(\{p_1, \ldots, p_{i-1}\}))} \leq$

$O(\sqrt{\log D})$. By Equation (1), the $(k-1)$-dimensional volume distortion of g is $O(\sqrt{\log D})$. This proves the upper bound in Theorem 3.

To prove the tightness part of the theorem, we show that for any $1 \le r \le 2^{\sqrt{D}/4}$ the r-dimensional volume distortion of the complete tree of depth D is $\Omega(\sqrt{\log D})$. It clearly suffices to prove the lower bound for arity $a = 2$, so let $T(V,E)$ be a complete binary tree of depth D. Identify the vertices of T with binary strings of length at most D. (The root is the empty string and the two children of vertex α are $\alpha 0$ and $\alpha 1$.) Let $\alpha(j)$ denote the jth prefix of $\alpha \in \{0,1\}^D$, i.e. the jth vertex on the path from the root to the leaf α.

Consider a nonexpansive Euclidean embedding ϕ of T that achieves minimum $(k-1)$-dimensional distortion. By Lemma 2, we may assume that ϕ is transitive. It is known from [Bou85, Mat99, LS01] that there must exist a pair of vertices whose distance in the tree is shrunk by ϕ by a factor of $\Omega(\sqrt{\log D})$. Specifically, suppose that $\alpha \in \{0,1\}^D$ is a leaf of T and p, q are integers with $1 \le p < q \le D$ and $q - p = 2^i$ for $i \ge 1$. We then let β be the leaf obtained from α by flipping the bit indexed by $1 + \frac{p+q}{2}$. Linial and Saks [LS01] show that there must exist α, p, q as above, such that ϕ shrinks the distance between $\alpha(q)$ and $\beta(q)$ by at least $\Omega(\sqrt{\log D})$, i.e.

$$\|\phi(\alpha(q)) - \phi(\beta(q))\| \le O\Big(\frac{d(\alpha(q),\beta(q))}{\sqrt{\log D}}\Big) \le O\Big(\frac{q-p}{\sqrt{\log D}}\Big). \tag{7}$$

It is straightforward from their proof that the stronger requirement $q-p = 2^i$ for $i \ge \frac{1}{2}\log D$ affects only the leading constant in (7). We flip in α any subset of the bits indexed by $1 + \frac{p+q}{2}, \ldots, \frac{p+3q}{4}$, and obtain $2^{\frac{q-p}{4}} \ge 2^{\sqrt{D}/4} \ge k$ leaves of T. Denote $k-1$ of these leaves by $\beta^1, \beta^2, \ldots, \beta^{k-1}$. We show below that the set $S = \{\alpha(q), \beta^1(q), \ldots, \beta^{k-1}(q)\}$ has volume distortion $\Omega(\sqrt{\log D})$.

Let us now analyze the Euclidean volume of $\phi(S)$. Recall that ϕ is transitive and observe that for all $1 \le i \le k-1$, there exists an automorphism of T that fixes $\alpha(q)$ and maps $\beta(q)$ to $\beta^i(q)$. Hence, $\|\phi(\alpha(q)) - \phi(\beta(q))\| = \|\phi(\alpha(q)) - \phi(\beta^i(q))\| \le O(\frac{q-p}{\sqrt{\log D}})$. Considering the vertices of S in the order $(\alpha(q), \beta^1(q), \ldots, \beta^{k-1}(q))$, the distance of each $\phi(\beta^i(q))$ from the affine hull of $\phi(\alpha(q)), \phi(\beta^1(q)), \ldots, \phi(\beta^{i-1}(q))$ is also at most $O(\frac{q-p}{\sqrt{\log D}})$, and hence $(k-1)! \cdot \mathrm{Evol}(S) \le \Big(O(\frac{q-p}{\sqrt{\log D}})\Big)^{k-1}$.

For the tree volume of S, observe that the distance in T between every two vertices of S is at least $\frac{q-p}{4}$, implying $\mathrm{Tvol}(S) \ge (\frac{q-p}{4})^{k-1}$. We therefore conclude that

$$\left(\frac{\mathrm{Tvol}(S)}{(k-1)! \cdot \mathrm{Evol}(\phi(S))}\right)^{\frac{1}{k-1}} \ge \Omega(\sqrt{\log D}).$$

This completes the proof of Theorem 3. □

3.4 General trees and chordal graphs

The embedding of complete trees described in the proof of Theorem 3 yields improved upper bounds on the volume distortion of general trees and of chordal graphs. Matoušek [Mat99] showed that any n-vertex (weighted) tree has an embedding into l_2 with $O(\sqrt{\log\log n})$ one-dimensional distortion, and this bound is tight for the complete binary tree, see [Bou86, Mat99, LS01]. Since every n-vertex tree embeds isometrically into the complete n-ary tree of depth n, Theorem 3 directly implies an $O(\sqrt{\log n})$ bound on higher-dimensional distortion. The direct sum of the latter embedding and the embedding of Matoušek [Mat99] mentioned above, achieves a slightly better bound, described below. This upper bound is nearly attained by an n-vertex path. Its proof appears below.

Theorem 4. *The metric of any tree on n vertices has an embedding whose r-dimensional volume distortion is $O((\log n)^{\frac{r-1}{2r}}(\log\log n)^{\frac{1}{2r}})$.*

Brandstädt et. al. [BCD99] show that for any chordal graph G, there exists a tree $T \subseteq G^2$ such that $d_T(u,v) \le d_G(u,v)$ for every two vertices u, v of G. It follows that $d_T(u,v) = \Theta(d_G(u,v))$, so for every subset S the tree volume of S in G and the tree volume of S in T are within a factor of $\exp(\Theta(k-1))$ from each other. By applying Theorem 4 on T (with appropriate scaling), we obtain the following upper bound on the distortion of chordal graphs.

Corollary 5. *The metric of any chordal graph on n vertices has an embedding whose r-dimensional volume distortion is $O((\log n)^{\frac{r-1}{2r}}(\log\log n)^{\frac{1}{2r}})$.*

Proof of Theorem 4. Let T be a tree on n vertices. Let ϕ_1 be a Euclidean embedding with (one-dimensional) distortion $O(\sqrt{\log\log n})$, as guaranteed by [Mat99]. Now let ϕ_2 be a Euclidean embedding of T whose r-dimensional distortion is $O(\sqrt{\log n})$ for any $r \ge 1$. The existence of ϕ_2 is implied by Theorem 3, since every n-vertex tree embeds isometrically into the complete n-ary tree of depth n. By Lemma 1 we know that the direct sum $\Phi = \phi_1/\sqrt{2} \oplus \phi_2/\sqrt{2}$ is a nonexpansive embedding. Let $(p_1, \ldots, p_k)$ be any proper ordering of the points in a set S. By Lemma 1(a),

$$\begin{aligned}(k-1)! \cdot \mathrm{Evol}(\Phi(S)) &= \|\Phi(p_2) - \Phi(p_1)\| \cdot \prod_{j=3}^{k} d_{\mathrm{aff}}\left(\Phi(p_j); \{\Phi(p_1), \ldots, \Phi(p_{j-1})\}\right) \\ &\ge \frac{\|\phi_1(p_2) - \phi_1(p_1)\|}{\sqrt{2}} \cdot \prod_{j=3}^{k} \frac{d_{\mathrm{aff}}\left(\phi_2(p_j); \{\phi_2(p_1), \ldots, \phi_1(p_{j-1})\}\right)}{\sqrt{2}}.\end{aligned}$$

Since ϕ_1 has a small one-dimensional distortion, $\frac{d_T(p_1,p_2)}{\|\phi_1(p_2)-\phi_1(p_1)\|} \le O(\sqrt{\log\log n})$. Our proof of Theorem 3 shows that there exists a proper ordering $(p_1, \ldots, p_k)$ of the points of S, such that $\frac{d_T(p_i,\{p_1,\ldots,p_{i-1}\})}{d_{\mathrm{aff}}(\phi_2(p_i);\phi_2(\{p_1,\ldots,p_{i-1}\}))} \le O(\sqrt{\log n})$ for all $2 \le i \le k$. For this proper ordering, we thus have from (1) that

$$\begin{aligned}\left(\frac{\mathrm{Tvol}(S)}{(k-1)! \cdot \mathrm{Evol}(\Phi(S))}\right)^{\frac{1}{k-1}} &\le \left(O(\sqrt{\log\log n}) \cdot (O(\sqrt{\log n}))^{k-2}\right)^{\frac{1}{k-1}} \\ &\le O\left((\log n)^{\frac{1}{2}-\frac{1}{2(k-1)}}(\log\log n)^{\frac{1}{2(k-1)}}\right),\end{aligned}$$

as claimed. □

3.5 Expander Graphs

In this section we show a lower bound on the volume distortion of any embedding of an n-vertex expander. Our lower bound matches, up to constant factors, the trivial embedding that maps the vertices of an expander to the vertices of a unit simplex. This bound also shows that Feige's [Fei00] extension of Bourgain's bound [Bou85] is tight for subsets of size $k \le O(\frac{\log n}{\log\log n})$. A graph G is a (d,δ)-*expander* if it is d-regular for $d \ge 3$ and the second eigenvalue of its adjacency matrix satisfies $\lambda_2 \le d-\delta$. Below we do not specify these parameters and simply say that the graph is an expander.

Theorem 5. *Let G be a (d,δ)-expander with n vertices. Then every Euclidean embedding of G has r-dimensional volume distortion at least $\Omega(\log n)$, for any $1 \le r < n^{1/3}$. The implicit coefficient in the Ω depends only on d and δ.*

The key tool for the proof of Theorem 5 is a lower bound on the one-dimensional distortion due to Linial, London and Rabinovich [LLR95] (see also [Mat97, LM00]). Specifically, the next lemma follows from Matoušek's book [Mat01, Equation 15.4].

Lemma 6. *Let ϕ be a Euclidean embedding of an n-vertex expander $G(V, E)$. Then*

$$\sum_{p,q \in V} \|\phi(p) - \phi(q)\|^2 \leq O(n) \cdot \sum_{(p,q) \in E} \|\phi(p) - \phi(q)\|^2. \tag{8}$$

Proof of Theorem 5. Let ϕ be a nonexpansive Euclidean embedding of an n-vertex expander $G(V, E)$. Then by Lemma 6 the lefthand side of (8) is at most $O(n^2)$. It follows that

$$\mathbb{E}\left[\|\phi(p) - \phi(q)\|^2\right] \leq O(1), \tag{9}$$

where p, q are two randomly chosen vertices. Let $p_1, \ldots, p_k$ be k distinct vertices chosen at random from V. We show that with positive probability, the volume distortion of $\{p_1, \ldots, p_k\}$ is $\Omega(\log n)$. To analyze the Euclidean volume of $\phi(p_1), \ldots, \phi(p_k)$, observe that the distance of $\phi(p_i)$ from the affine hull of $\phi(p_1), \ldots, \phi(p_{i-1})$ is at most $\|\phi(p_i) - \phi(p_1)\|$. Therefore, by the arithmetic-geometric inequality, $(k-1)! \cdot \mathrm{Evol}(\phi(p_1), \ldots, \phi(p_k)) \leq \prod_{i=2}^{k} \|\phi(p_i) - \phi(p_1)\| \leq \left(\frac{\sum_{i=2}^{k} \|\phi(p_i) - \phi(p_1)\|^2}{k-1}\right)^{\frac{k-1}{2}}$. By linearity of expectation, $\mathbb{E}\left[\frac{\sum_{i=2}^{k} \|\phi(p_i) - \phi(p_1)\|^2}{k-1}\right] = \mathbb{E}\left[\|\phi(p) - \phi(q)\|^2\right] \leq O(1)$, where the first expectation is over the choice of $p_1, \ldots, p_k$, and the second one is over the choice of p, q. Markov's inequality then implies that

$$\Pr\left[\left((k-1)! \cdot \mathrm{Evol}(\phi(p_1), \ldots, \phi(p_k))\right)^{\frac{1}{k-1}} \leq O(1)\right] \geq 3/4. \tag{10}$$

To analyze the tree volume of $p_1, \ldots, p_k$, observe that if for all $i \neq j$ the distance in G between p_i and p_j is $\geq \Delta$, then $\mathrm{Tvol}(p_1, \ldots, p_k) \geq \Delta^{k-1}$. The neighborhood of radius $\Delta = \log_d \frac{n}{4k^2}$ around any vertex contains at most $\frac{n}{4k^2}$ vertices. It follows that the above lower bound on the tree volume holds with probability at least

$$\prod_{i=1}^{k-1} \frac{n - i\frac{n}{4k^2}}{n - i} \geq (1 - \frac{1}{4k})^{k-1} > 3/4.$$

Combining this with (10), it follows that with probability larger than 1/2,

$$\left(\frac{\mathrm{Tvol}(p_1, \ldots, p_k)}{(k-1)! \cdot \mathrm{Evol}(\phi(p_1), \ldots, \phi(p_k))}\right)^{\frac{1}{k-1}} \geq \Omega(\log_d \frac{n}{4k^2}) \geq \Omega(\log n),$$

which proves the theorem. □

3.6 A hypercube

The following lower bound on the distortion of a hypercube extends the known lower bound for one-dimensional distortion, due to Enflo [Enf69] (see also [LM00, Mat01]). Its proof appears below and is quite similar to that for expanders.

Theorem 6. *The n-vertex hypercube (of dimension $\log n$) has r-volume distortion $\Omega(\sqrt{\log n})$, for any $1 \leq r \leq n^{1/4}/2$.*

Interestingly, the identity embedding of the hypercube is tight with this lower bound for one and two-dimensional distortion, but for higher-dimensional distortion the identity embedding gives no upper bound, since 4 points in a 2-dimensional face (square) are not even in general position. Determining the higher-dimensional volume distortion of hypercubes thus remains an open question. We briefly sketch an embedding of the hypercube whose distortion is better than that of general graphs. Let ϕ_1 be the identity embedding of the hypercube, into $\log n$-dimensional Euclidean space. Rao's algorithm [Rao99] can be applied to this n-point set in Euclidean space. Let ϕ_2 be this (composed) embedding. The direct sum $\phi_1/\sqrt{2} \oplus \phi_2/\sqrt{2}$ is an embedding whose $(k-1)$-dimensional distortion for $k \geq 3$ is $O(\sqrt{\log n}(\log n \log k)^{\frac{1}{2}-\frac{1}{k-1}})$.

The key tool for the proof of Theorem 6 is a lower bound on the distortion of distances (i.e. the case $k = 2$), originally shown by Enflo [Enf69] (see also [LM00]). Specifically, we use a lemma that follows from the proof of Theorem 15.4.1 in Matoušek's book [Mat01].

Lemma 7 (Matoušek [Mat01]). *Let ϕ be a Euclidean embedding of an n-vertex hypercube $H(V, E)$, and let $F^* = \{(p, q) : p < q \; ; \; d(p, q) = \log n \; ; \; p, q \in V\}$ be the collection of ordered antipodal pairs. Then*

$$\sum_{(p,q)\in F^*} \|\phi(p) - \phi(q)\|^2 \leq \sum_{(p,q)\in E} \|\phi(p) - \phi(q)\|^2. \tag{11}$$

Proof of Theorem 6. Let ϕ be a nonexpansive Euclidean embedding of an n-vertex hypercube $H(V, E)$. We first claim that

$$\sum_{(p,q)\in F} \|\phi(p) - \phi(q)\|^2 \leq \frac{n}{2} \sum_{(p,q)\in E} \|\phi(p) - \phi(q)\|^2, \tag{12}$$

where $F = \{(p, q) : p < q \; ; \; p, q \in V\}$. Indeed, apply Lemma 7 to all the possible sub-hypercubes and sum up all the inequalities corresponding to (11). Note that every pair in F is counted once and every edge in E is counted $\sum_{t=1}^{\log n} \binom{\log n-1}{t-1} = \frac{n}{2}$ times.

Since ϕ is nonexpansive, the righthand side of (12) is at most $\frac{1}{4}n^2 \log n$. It follows that

$$\mathbb{E}\left[\|\phi(p) - \phi(q)\|^2\right] \leq \frac{n^2 \log n}{4|F|} = O(\log n), \tag{13}$$

where p, q are chosen uniformly at random.

We now show that with probability $\geq \frac{1}{2}$ a randomly chosen set of vertices $\{p_1, \ldots, p_k\}$ has volume distortion $\Omega(\log n)$. A calculation similar to that in Theorem 5 shows that (13) implies

$$\Pr\left[(k-1)! \cdot \mathrm{Evol}(p_1, \ldots, p_k) \leq (4 \cdot O(\log n))^{\frac{k-1}{2}}\right] \geq 3/4. \tag{14}$$

To analyze the tree volume of $p_1, \ldots, p_k$, observe that if all pairwise distances (in the hypercube H) between p_i and p_j are at least $\frac{1}{10} \log n$, then $\mathrm{Tvol}(p_1, \ldots, p_k) \geq (\frac{1}{10} \log n)^{k-1}$. The size of a ball of radius $\frac{1}{10} \log n$ in the hypercube H is at most $\sum_{i=0}^{\frac{1}{10}\log n} \binom{\log n}{i} \leq \sqrt{n}$, where the last inequality follows from the entropy estimate $\binom{t}{\alpha t} \simeq [\alpha^\alpha(1-\alpha)^{1-\alpha}]^t$. Since the vertices $p_1, \ldots, p_k$ are chosen at random, the above lower bound on the tree volume holds with probability at least

$$\prod_{j=1}^{k-1} \frac{n - j\sqrt{n}}{n - j} \geq \prod_{j=1}^{k-1} (1 - \frac{j}{\sqrt{n}}) \geq 1 - \frac{1 + \ldots + k - 1}{\sqrt{n}} > 3/4.$$

Combining this with (14), we have by the union bound that with probability larger than 1/2,

$$\left(\frac{\mathrm{Tvol}(p_1,\ldots,p_k)}{(k-1)!\cdot\mathrm{Evol}(p_1,\ldots,p_k)}\right)^{\frac{1}{k-1}} \geq \frac{\frac{1}{10}\log n}{(4\cdot O(\log n))^{1/2}} \geq \Omega(\sqrt{\log n}),$$

which proves the theorem. □

4 Euclidean metrics

We provide a small improvement to an upper bound of Rao [Rao99] on the volume distortion of Euclidean metrics. Rao [Rao99] showed how embed an n-point set in l_2 that has polynomial aspect ratio with $(k-1)$-dimensional distortion of $O(\sqrt{\log n \log k})$. We offer here a slight improvement of his result. This is done by taking a direct sum of Rao's embedding and an isometric embedding (recall that the metric is Euclidean). Surprisingly, perhaps, it follows that among all n-point Euclidean metrics with polynomial aspect ratio, paths have the worst asymptotic $k-1$-dimensional distortion for constant k.

Theorem 7. *Every n-point Euclidean metric with polynomial (in n) aspect ratio has an embedding into l_2 whose $k-1$-dimensional volume distortion is $O((\log n \log k)^{\frac{1}{2}-\frac{1}{2(k-1)}})$.*

Proof. Let (X,d) be a Euclidean metric with $n = |X|$ points. Let ϕ_1 be the embedding of X achieved by Rao's (randomized) algorithm [Rao99] for Euclidean metrics. Let ϕ_2 be an isometric embedding of (X,d) into a finite-dimensional Euclidean space. By Lemma 1 we know that the direct sum $\Phi = \phi_1/\sqrt{2} \oplus \phi_2/\sqrt{2}$ is a nonexpansive embedding. Let $(x_1,\ldots,x_k)$ be a proper ordering of the points in a set S. By Lemma 1(a),

$$\begin{aligned}(k-1)!\cdot\mathrm{Evol}(\Phi(S)) &= \|\Phi(x_2)-\Phi(x_1)\|\cdot\prod_{j=3}^{k} d_{\mathrm{aff}}\left(\Phi(x_j);\{\Phi(x_1),\ldots,\Phi(x_{j-1})\}\right)\\ &\geq \frac{\|\phi_2(x_2)-\phi_2(x_1)\|}{\sqrt{2}}\cdot\prod_{j=3}^{k}\frac{d_{\mathrm{aff}}\left(\phi_1(x_j);\{\phi_1(x_1),\ldots,\phi_1(x_{j-1})\}\right)}{\sqrt{2}}.\end{aligned}$$

Since ϕ_2 is an isometry, $\|\phi_2(x_2)-\phi_2(x_1)\| = d(x_1,x_2)$. The analysis of Rao [Rao99, Theorem 9] guarantees that with very high probability, for every $j \geq 2$ (though we only need $j \geq 3$) $\frac{d(x_j,\{x_1,\ldots,x_{j-1}\})}{d_{\mathrm{aff}}(\phi_1(x_j);\{\phi_1(x_1),\ldots,\phi_1(x_{j-1})\})} \leq O(\sqrt{\log n \log k})$. We then obtain from (1) that

$$\left(\frac{\mathrm{Tvol}(S)}{(k-1)!\cdot\mathrm{Evol}(\Phi(S))}\right)^{\frac{1}{k-1}} \leq \left(O(1)\cdot O((\log n\log k)^{\frac{k-2}{2}})\right)^{\frac{1}{k-1}} = O((\log n \log k)^{\frac{1}{2}-\frac{1}{2(k-1)}}),$$

as claimed. □

5 Discussion

Feige [Fei00] demonstrates the existence and applicability of higher-order substructures in finite metric spaces (see Section 1), but there may be alternative notions worthy of investigation that capture such phenomena. For example, it is possible to consider *any* embedding $\Phi : X \to l_2$ (not necessarily nonexpansive) and see how well $\mathrm{Evol}(\Phi(S))$ approximates $\mathrm{Vol}(S)$. Specifically, is there an embedding Φ of the path on $1,2,\ldots,n$ such that $\mathrm{Evol}(\Phi(p),\Phi(q),\Phi(r)) = \Theta((r-q)(q-p))$ for every three path vertices $p < q < r$?

Several problems involving Feige's definition of higher-dimensional distortion remain open. One interesting gap is in the r-dimensional distortion of a hypercube for $r \geq 4$; see Section 3.6. It is also evident from Table 1 that there are gaps in our understanding of the r-dimensional distortion of trees, planar metrics and Euclidean metrics for, say, fixed $r \geq 2$, and that of general metrics for, say, $r \geq \log n$.

References

[BCD99] A. Brandstädt, V. Chepoi, and F. Dragan. Distance approximating trees for chordal and dually chordal graphs. *J. Algorithms*, 30(1):166–184, 1999.

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

[Bou86] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.

[DV01] J. Dunagan and S. Vempala. On Euclidean embeddings and bandwidth minimization. In *Randomization, approximation, and combinatorial optimization*, pages 229–240. Springer, 2001.

[Enf69] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105, 1969.

[Fei00] U. Feige. Approximating the bandwidth via volume respecting embeddings. *J. Comput. System Sci.*, 60(3):510–539, 2000.

[Ind01] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 10–33, October 2001.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[LM00] N. Linial and A. Magen. Least-distortion Euclidean embeddings of graphs: products of cycles and expanders. *J. Combin. Theory Ser. B*, 79(2):157–171, 2000.

[LS01] N. Linial and M. Saks. On the distortion of complete binary trees. Discrete and Computational Geometry, to appear.

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

[Mat99] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.

[Mat01] J. Matoušek. Lectures on Discrete Geometry. Book in preparation. Available from `http://www.ms.mff.cuni.cz/acad/kam/matousek`, 2001.

[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.

[Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306. ACM, 1999.