

# Embedding tree metrics into low dimensional Euclidean spaces\*

Anupam Gupta<sup>†</sup>  
Computer Science Division  
University of California  
Berkeley CA 94720.  
Email: `angup@cs.berkeley.edu`

25th January 2000

## Abstract

We consider embedding metrics induced by trees into Euclidean spaces with a restricted number of dimensions. We show that any weighted tree  $T$  with  $n$  vertices and  $L$  leaves can be embedded into  $d$ -dimensional Euclidean space with  $\tilde{O}(L^{1/(d-1)})$  distortion. Furthermore, we exhibit an embedding with almost the same distortion which can be computed efficiently. This distortion substantially improves the previous best upper bound of  $\tilde{O}(n^{2/d})$  and almost matches the best known lower bound of  $\Omega(L^{1/d})$ .

---

\*A preliminary version of this article previously appeared in the Proceedings of the 31st Annual Symposium on the Theory of Computing [5]

<sup>†</sup>Supported by NSF grant CCR-9505448 and CCR-9820951.

# 1 Introduction

The study of finite metric spaces has been the focus of much attention in recent years due both to its mathematical and to its computational significance. Perhaps the most widely studied class of questions has been that of embeddability of a finite metric space into some target space with the aim of preserving the distances between points in the metric. Often the distances cannot be preserved exactly, and then we aim to keep the distance between any pair of vertices close to its original value: the *distortion* of an embedding is the maximum factor by which any distance is changed by the embedding. (For precise definitions of this and other notions, see section 2.) Not only do these questions help better understand the structure of both the host and the target metrics, but they also have numerous computational applications (e.g., see [8] and the references therein).

Among the most important target metric spaces are the so-called Minkowski  $\ell_p$  spaces, i.e.,  $d$ -dimensional real space  $\mathbb{R}^d$  equipped with one of the  $\ell_p$  norms,  $\|x\|_p = (\sum_i x_i^p)^{1/p}$ . We shall refer to such spaces as  $\ell_p^d$ . (Thus, for example,  $\ell_2^d$  is just standard  $d$ -dimensional Euclidean space.) The topic of embedding tree metrics into Minkowski spaces has been studied by many authors. Bourgain [2] showed that every embedding of the complete binary tree with  $n$  vertices into  $\ell_p$  must incur a distortion of  $\Omega((\log \log n)^{\min\{1/2, 1/p\}})$  for any  $1 < p < \infty$ . (Note that any tree has an isometric embedding into  $\ell_1$ , and it is also well known that *any* metric space is isometrically embeddable into  $\ell_\infty$ .) Subsequently, Matoušek exhibited an  $O((\log \log L)^{\min\{1/2, 1/p\}})$ -distortion embedding for arbitrary trees  $T$  with  $L$  leaves [11], showing that the bound is actually tight. A slightly weaker distortion of  $O(\log \log L)$  for embedding into  $\ell_2$  was obtained independently by Linial et al. [9].

Unfortunately, these techniques typically do not allow the *dimension*  $d$  of the target space to be restricted, and in general, they require  $L$  dimensions. Embedding into a lower dimensional space could cause the distortion of the embedding to increase, and we would like a good tradeoff between dimension and distortion. The question of trading dimensions for distortion has also been previously studied. Johnson and Lindenstrauss [7] showed that one can project any  $n$ -point Euclidean metric in an arbitrary number of dimensions down to a  $O(\log n)$  dimensional space while incurring only a constant distortion, which immediately implies that any tree can be embedded in  $O(\log n)$  dimensions with  $O(\sqrt{\log \log n})$  distortion. Their technique is to project the high-dimensional embedding onto a random  $O(\log n)$ -dimensional subspace and show that, with high probability, the distortion is not more than a constant. This procedure was analyzed for general  $d$  by Matoušek [10], who showed that the distortion incurred by this projection is at most  $O(n^{2/d}(\log n)^{1/2}/\sqrt{d})$ . He also proved that this result is almost optimal by exhibiting metrics which require a distortion  $O(n^{2/d})$  to embed into  $d$ -dimensions.

However, no upper bound better than  $\tilde{O}(n^{2/d})$  is known for the distortion incurred by the projection technique in the special case of trees, and the best lower bound known for tree metrics is only  $O(n^{1/d})$ . Hence a natural question is whether better

embeddings can be obtained for trees into  $\ell_2^d$  [11]. In this paper, we present a new upper bound for the distortion that is very close to the  $\Omega(n^{1/d})$  lower bound. Our main result is the following:

**Theorem 1.1** *Given any tree  $T$  with  $L$  leaves, there exists a mapping  $f : V(T) \rightarrow \ell_2^d$  with distortion*

$$D(f) \leq O(L^{1/(d-1)}(\min\{\log L, d\})^{1/2}). \quad (1)$$

In fact, the  $\log L$  term in equation (1) can be replaced by  $\kappa(T)$ , the caterpillar dimension of  $T$  (defined in section 2). Since  $\kappa(T) \leq \log L$ , this replacement gives (as in [11, 9]) not only a more adaptive bound, but also an embedding for infinite trees with finite caterpillar dimension. However, we note that our embedding technique is not tight for  $O(\log n)$  dimensions, where it incurs a distortion of  $O(\sqrt{\log L})$ , while the optimal embedding of [11] incurs only  $O(\sqrt{\log \log L})$  distortion.

The idea behind our embedding is simple. Our starting point is the naïve embedding scheme for trees, in which we use a new dimension orthogonal to all previous ones for each new edge [11, 9]. However, this requires  $L$  orthogonal vectors, and cannot be used when the dimension  $d \ll L$ , since we can have only  $d$  orthogonal vectors in  $\mathbb{R}^d$ . To circumvent this problem, we relax the orthogonality constraint and find  $L$  vectors that have a small inner product with each other. We then use these vectors in place of the orthogonal vectors used in the naïve embedding of the tree, and show that the price we pay for the non-orthogonality of the vectors is a factor of  $\tilde{O}(L^{1/(d-1)})$  in the distortion.

The rest of the paper is organized as follows. In section 2, we give some definitions and review some elementary geometry which we shall use later. In section 3 we give the actual embedding, which we analyze in section 4. Finally, we show that the lower bound is tight for the special case of star graphs by giving a matching upper bound in section 5. Appendix A gives some more details about the construction of the well-separated vectors, and a simple example is given in Appendix B.

## 2 Definitions and notation

**Trees and Metric Spaces:** Let  $T = (V, E)$  be an undirected weighted tree with  $|V| = n$  and  $|E| = (n - 1)$ . We use  $L$  to denote the number of leaves and  $\Delta(T)$  to denote the weak diameter of  $T$ . Recall that the weak diameter just counts the number of edges on a path, and does not take into account their lengths. The weight (or length) function is  $w : E \rightarrow \mathbb{R}^+$ . We root  $T$  at an arbitrary vertex  $r \in V$ , which defines the usual ancestor-descendant relationship between pairs of vertices. The *subtree rooted at  $v$* ,  $T_v$ , is the subtree induced by the vertices which are descendants of  $v$  (including  $v$  itself). A *monotone path*  $P$  is a path which is a subset of some leaf-root path  $P'$ .

The *caterpillar dimension* [11, 9] of a rooted tree  $T$ , henceforth denoted by  $\kappa(T)$ , is defined thus: For a tree with a single vertex,  $\kappa(T) = 0$ . Else,  $\kappa(T)$  is the maximal value  $\leq k + 1$  if there exist paths  $P_1, P_2, \dots, P_t$  beginning at the root and pairwise edge-disjoint such that each component  $T_j$  of  $T - E(P_1) - E(P_2) - \dots - E(P_t)$  has  $\kappa(T_j) \leq k$ , where  $T - E(P_1) - E(P_2) - \dots - E(P_t)$  denotes the tree  $T$  with the edges of the  $P_i$ 's removed, and the components  $T_j$  are rooted at the unique vertex lying on some  $P_i$ . The collection of edge-disjoint monotone paths in the above recursive definition form a partition of  $E$ , and are called the *caterpillar decomposition* of  $T$ . It is simple to see that the path between any two vertices of  $T$  intersects at most  $2\kappa(T)$  of these paths. It can also be shown that  $\kappa(T)$  is at most  $\log L$ , and is often much smaller. For example, the caterpillar graph has  $\kappa(T) = 2$ , even though it has  $\Omega(n)$  leaves. A decomposition with the minimum value of  $\kappa(T)$  can be computed in polynomial time by dynamic programming (see, e.g., [11]).

**Embeddings and Distortion:** Let  $M = (X, \rho)$  and  $M' = (Y, \mu)$  be two metric spaces, and let  $f : X \rightarrow Y$  be a map. Let us define

$$D_c(f) = \sup_{x, y \in X} \frac{\rho(x, y)}{\mu(f(x), f(y))} \quad \text{and} \quad D_e(f) = \sup_{x, y \in X} \frac{\mu(f(x), f(y))}{\rho(x, y)}.$$

Then we say that  $f$  has *contraction*  $D_c$ , *expansion*  $D_e$  and *distortion*  $D(f) = D_c(f) \times D_e(f)$ . We say that  $M$  *D-embeds* into  $X$  if there exists a map  $f$  with  $D(f) \leq D$ .

Given a tree  $T = (V, E)$  with non-negative edge weights, the distance  $d_T(u, v)$  between vertices  $u, v \in V$  is the length of the unique path in  $T$  between  $u$  and  $v$ . We shall often blur the distinction between the tree  $T$  and the metric  $(V, d_T)$  in the following discussion. It is easy to see that if  $f$  is an embedding of a graph  $G$  into a metric  $(Y, \mu)$ , the expansion is largest for edges of  $G$ .

**Euclidean space and spheres:** Let  $\|\cdot\|_p$  denote the  $\ell_p$  Minkowski norm. For the case  $p = 2$  we shall drop the subscript, and hence  $\|\cdot\|$  denotes the Euclidean norm. Recall that  $\|x\|_q \leq \|x\|_p$  for  $p \leq q$ . Let the vector  $\vec{u}_i$  be the unit vector in the  $i$ -th dimension. Let  $B_d(r)$  denote the  $d$ -dimensional sphere of radius  $r$ . Its  $d$ -dimensional volume is  $V_d(r) = \frac{\pi^{d/2} r^d}{\Gamma(1+d/2)}$  and surface area is  $S_d(r) = \frac{2\pi^{d/2} r^{d-1}}{\Gamma(d/2)}$ . Unless mentioned otherwise, we shall always assume that spheres are centered at the origin. The *angular distance* between any two points  $u$  and  $v$  is the acute angle subtended by them at the origin.

Given a sphere  $S$  (centered at some point  $o$ ), let  $\partial S$  denote the surface of  $S$ . For any  $u, v \in \partial S$ , the *great circle* through  $u$  and  $v$  is the intersection of the plane defined by the three points  $u, v$  and  $o$  with  $\partial S$ . The *spherical distance* is the length of the shortest path between  $u$  and  $v$  that lies in  $\partial S$ . If  $u$  and  $v$  are at angular distance  $\delta \leq \pi/2$ ; their spherical distance is  $r\delta$ , while their Euclidean distance is  $r\sqrt{2(1 - \cos \delta)}$ . We will use the fact that these distances differ by only a constant, i.e.,  $(2/\pi)\delta \leq \sqrt{2(1 - \cos \delta)} \leq \delta$ .

**Spherical Codes and Well Separated Vectors:** A *spherical code* is a set of  $N$  unit vectors (or  $N$  points on the surface of a unit sphere), where the measure of

goodness of the code is the least angle between any two vectors in the code (see [3] for more details). An equivalent formulation is to minimize the radius  $r$  of a sphere such that  $N$  points can be placed on its surface, where any two points are at angular distance at least  $2$  from each other. For our application, it is crucial that the vectors of the code lie in the positive orthant. Standard packing arguments show that we can build such a code  $\mathcal{S}(N)$  with  $N$  vectors, where the radius of the sphere is  $r_0(N) = 8N^{1/(d-1)}$ , and minimum angle between the vectors is  $\delta_0(N) = 2/r_0(N)$ . For completeness, a proof of this fact is given in Appendix A.

### 3 Embedding trees into $d$ dimensions

In this section we describe the actual embedding for trees. We first describe, in section 3.1, a trivial algorithm and then a twist to it (which are also discussed in [9, 11]) both of which use a large number of dimensions. We then build on these ideas using with spherical codes to get our final embedding in section 3.2.

We will always embed the root  $r$  at the origin, i.e.,  $f(r) = \vec{0}$ , and for each edge  $e$ , we will define a vector  $v(e) \in \mathbb{R}^d$ . The image of vertex  $v$  is then given by  $f(v) = \sum_{e \in P} v(e)$ , where  $P$  is the path from  $v$  to the root  $r$ .

#### 3.1 Embeddings which require many dimensions

The first embedding is the naïve embedding in which the  $i$ -th edge  $e_i$  is mapped to a vector of length  $w(e_i)$  orthogonal to all the other vectors, i.e.,  $v(e_i) = w(e_i)\vec{u}_i$ , where  $\vec{u}_i$  is the unit vector in the  $i$ -th dimension. The distortion of this embedding is easily seen to be the maximum of  $\frac{\sum_{e \in P} w(e)}{\sqrt{\sum_{e \in P} w(e)^2}}$  over all paths  $P$ , which is bounded above by  $\sqrt{\Delta(T)}$ . The large contraction above is due to the large number of bends on long paths, since the contraction is proportional to the square root of the number of bends. Using the caterpillar decomposition of section 2, we can remove the dependence of the distortion on the weak diameter. Let  $\mathcal{F} = \{P_1, P_2, \dots, P_t\}$  be a caterpillar decomposition of width  $O(\kappa(T))$ , and associate  $\vec{u}_i$ , the unit vector along the  $i$ -th dimension, with the path  $P_i$ . To embed an edge  $e$ , set  $v(e) = w(e)\vec{u}_j$ , where  $P_j \ni e$ . A simple analysis now shows that the distortion is bounded by  $O(\sqrt{\kappa(T)})$ .

#### 3.2 Embedding into $d$ dimensions

The embeddings given above require  $\Omega(n)$  dimensions in the worst case, and so we cannot hope for constant dimension embeddings using (almost-)orthogonal vectors. Hence we will replace the  $L$  orthogonal vectors in the latter construction by vectors in  $\mathbb{R}^d$  that merely have a large angle with each other. This embedding crucially relies on the fact that the vectors all lie in the positive orthant. By the discussion

above, we know that such a set  $\mathcal{S} = \mathcal{S}(L)$  of vectors exists, with the angle between any two of the vectors being at least  $\delta_0(L)$ .

A little thought will convince the reader that arbitrarily replacing the orthogonal vectors by those in  $\mathcal{S}$  will not work. To handle this problem, for each vertex, we will associate each subtree  $T_v$  with a set of vectors  $\mathcal{S}_v \subseteq \mathcal{S}$ , such that the number of vectors in  $\mathcal{S}_v$  is equal to the number of leaves in  $T_v$ . Further, we will ensure that the cones generated by the vectors associated with two disjoint subtrees are disjoint. (The *cone* generated by a set of vectors is the set of all non-negative linear combinations of those vectors.)

The embedding procedure starts at the root, and a vertex is embedded only after its parent is embedded. As the base case, the root has the entire set  $\mathcal{S}$  of vectors allotted to it, and is mapped to the origin. Inductively, let us consider vertex  $v$  with children  $v_1, v_2, \dots, v_k$ . Further, let the subtree  $T_v$  have the set  $\mathcal{S}_v$  allotted to it and let  $v$  be embedded at position  $f(v)$ . We then partition the set of vectors  $\mathcal{S}_v$  into  $k$  subsets  $\mathcal{S}_{v_i}$ , such that  $|\mathcal{S}_{v_i}|$  is equal to the number of leaves in  $T_{v_i}$ , and further, that the cones of these sets are disjoint. Now, the vertex  $v_i$  is mapped to the position  $f(v) + w_i x_i$ , where  $w_i$  is the length of the edge  $\{v, v_i\}$ , and  $x_i$  is some vector from the set  $\mathcal{S}_{v_i}$ . The reader is urged to look at the example given in Appendix B.

To partition a set of vectors  $\mathcal{S}_v$ , we assume that the ends of the vectors are in general position. We now sweep a hyperplane passing through the origin until we get a set of vectors whose size is equal to the number of leaves in  $T_{v_1}$ . We define this to be the subset  $\mathcal{S}_{v_1}$ , and continue the sweep to get the remaining subsets  $\mathcal{S}_{v_i}$ . Note that this procedure ensures that the cones of the subsets generated will be disjoint from each other, and will also be contained in the cone of the set  $\mathcal{S}_v$ .

To get a bound on the distortion which depends on  $\kappa(T)$ , we have to choose the subsets more carefully. Instead of choosing the vector  $x_i$  arbitrarily in the previous description, we choose it so that vertices lying in any given path in the caterpillar decomposition of  $T$  are embedded along a straight line. W.l.o.g., we can assume that the vertex  $v_1$  is the child of  $v$  that belongs to the same path in the caterpillar decomposition as  $v$ . Hence it suffices to ensure that the vector corresponding to the edge  $\{v, v_1\}$  is the same as the vector corresponding to the edge connecting  $v$  to its parent, and hence it must lie in the set  $\mathcal{S}_{v_1}$ . This can be easily ensured by choosing the direction of the sweeping hyperplane appropriately; the details are routine and left to the reader.

Note that using vectors which lie in the positive orthant is a natural analogue of using positive unit vectors in the naïve embedding, and is crucial to the correctness of the embedding. Also note that the procedure ensures that the angular distance between the images of any two unrelated vertices in the tree is at least  $\delta_0$ .

## 4 Analysis of the embedding

We now bound the distortion of the map  $f$  described in the previous section. Let  $\Lambda = \sqrt{\min\{d, \kappa(T)\}}$ , and as before, let  $r_0 = 2/\delta_0 = 8L^{1/(d-1)}$ .

**Theorem 4.1** *The distortion of the embedding is at most  $4\pi L^{1/(d-1)} \sqrt{\min\{d, \kappa(T)\}}$ .*

**Proof:** Since the expansion of a map is worst for edges of the tree  $T$  and all tree edges have the correct length, the expansion  $D_e = 1$ . Using claims 4.2 and 4.3 below, we get that the contraction  $D_c$  for any pair of vertices is at most  $(\pi/2)r_0\Lambda$ . This, in turn, implies that the distortion is  $D_e \times D_c \leq (\pi/2)r_0\Lambda = 4\pi L^{1/(d-1)} \sqrt{\min\{d, \kappa(T)\}}$ , and completes the proof of the theorem.  $\blacksquare$

**Claim 4.2** *Let  $u$  be an ancestor of  $v$  in  $T$ . Then*

$$\|f(u) - f(v)\| \geq \frac{d_T(u, v)}{\Lambda}.$$

**Proof of Claim 4.2:** Let  $P = \langle u = x_1, x_2, \dots, x_{t+1} = v \rangle$  be the path with  $t$  edges connecting them. Furthermore, let the length of the edge  $e_i = (x_i, x_{i+1})$  be  $w_i$ , which implies that  $\sum_i w_i = d_T(u, v)$ . Let us now consider the vectors  $\vec{v}_i = f(x_{i+1}) - f(x_i)$ . By our construction,  $\|\vec{v}_i\|$  is exactly  $w_i$ , and the projections of  $\vec{v}_i$  on the  $d$  coordinate axes are all positive. Let  $w_{ij}$  denote the length of the projection of  $\vec{v}_i$  on the  $j$ -th axis (for  $1 \leq i \leq t$  and  $1 \leq j \leq d$ ). Note that  $w_{ij} \geq 0$  and  $\sum_j w_{ij}^2 = w_i^2$ .

We will first prove that  $\|f(u) - f(v)\|$  is lower bounded by  $d_T(u, v)/\sqrt{\kappa(T)}$ . To see this, note that having a caterpillar decomposition of width  $\kappa(T)$  ensures that  $P$  cannot have more than  $O(\kappa(T))$  bends, since  $P$  is a monotone path. Our construction also ensures that the angles of the bends cannot be more than  $90^\circ$ . It is intuitively clear that the worst contraction is achieved when all the bends are actually right angles, in which case, the Euclidean distance between  $f(u)$  and  $f(v)$  is at least  $d_T(u, v)/\sqrt{\kappa(T)}$ . Formally, note that the squared distance is

$$\begin{aligned} \|f(v) - f(u)\|^2 &= \sum_{j=1}^d \left( \sum_{i=1}^t w_{ij} \right)^2 \geq \sum_{j=1}^d \sum_{i=1}^t w_{ij}^2 = \sum_{i=1}^t \|\vec{v}_i\|^2 \\ &\geq \frac{1}{t} \left( \sum_i \|\vec{v}_i\| \right)^2 && \text{(Cauchy-Schwarz)} \\ &= \frac{1}{t} \left( \sum_i w_i \right)^2 = \frac{d_T(u, v)^2}{t}, \end{aligned}$$

and thus  $\|f(u) - f(v)\| \geq d_T(u, v)/\sqrt{t} \geq d_T(u, v)/\sqrt{\kappa(T)}$ . A similar analysis (which we omit) shows a lower bound of  $d_T(u, v)/\sqrt{d}$ , which completes the proof.  $\square$

**Claim 4.3** *Let  $u$  and  $v$  be unrelated vertices in  $T$ . Then*

$$\|f(u) - f(v)\| \geq \frac{d_T(u, v)}{(\pi/2)r_0\Lambda}.$$

**Proof of Claim 4.3:** Let  $w$  be the least common ancestor of  $u$  and  $v$ . Let path  $P_u$  connect  $u$  to  $w$ , and the path  $P_v$  connect  $v$  to  $w$ . Note that both these paths monotone and edge disjoint, and that  $|P_u| + |P_v| = d_T(u, v)$ .

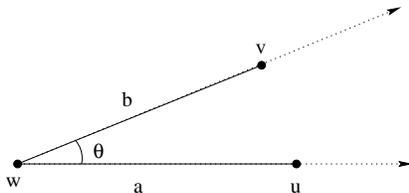


Figure 1: Proof of Claim 4.3

First, consider the simple case when both the paths  $P_u$  and  $P_v$  are embedded without any bends, as in figure 1. Then  $\|f(u) - f(v)\|^2$  is  $|P_u|^2 + |P_v|^2 - 2|P_u||P_v|\cos\theta$ , and the contraction is

$$\frac{(1+x)}{(1+x^2-2x\cos\theta)^{1/2}},$$

where  $x = |P_v|/|P_u|$ . A simple calculation shows that this is maximized when  $x = 1$ , and the maximum value is  $2/\sqrt{2(1-\cos\theta)}$ . However, the angle  $\theta \geq \delta_0$ , and hence this value is at most  $2/\sqrt{2(1-\cos\delta_0)} \leq \pi/\delta_0 \leq (\pi/2)r_0$ .

To handle the general case when the paths are embedded with bends and the Euclidean distance between their endpoints is smaller than their lengths in the tree  $T$ , we use the fact from Claim 4.2 that the lengths of a monotone path in the tree is no more than a factor  $\Lambda$  times longer than the Euclidean distance between its endpoints. Thus the distortion for the case without bends is off by a factor of at most  $\Lambda$  in case there are bends. Formally, if  $\|f(u) - f(w)\| = D_{uw}$  and  $\|f(v) - f(w)\| = D_{vw}$ , then  $D_{uw} \geq |P_u|/\Lambda$  and  $D_{vw} \geq |P_v|/\Lambda$ . Hence the contraction in this case is at most

$$\frac{|P_u| + |P_v|}{(D_{uw}^2 + D_{vw}^2 - 2D_{uw}D_{vw}\cos\theta)^{1/2}} \leq \frac{\Lambda(D_{uw} + D_{vw})}{(D_{uw}^2 + D_{vw}^2 - 2D_{uw}D_{vw}\cos\theta)^{1/2}}.$$

Now mimicking the previous calculation shows that this contraction is at most  $(\pi/2)\Lambda r_0$ .  $\square$

## 5 Embedding the star

In this section, we briefly sketch lower and upper bounds of  $\Theta(L^{1/d})$  for the weighted star graph using simple sphere-packing arguments. We then explore the projec-

tion technique of [7, 10], and show that there is a projection of the natural high-dimensional embedding of the star onto a  $d$ -dimensional space that incurs a distortion of  $\tilde{O}(L^{1/d})$ . This begs the question: for every tree  $T$ , does there exist a projection of its natural high-dimensional embedding into  $d$  dimensions which preserves distances well?

In the sequel, let  $G = (V, E)$  be the star on  $L$  leaves, and let  $0$  denote the center of the star. Thus the edges in  $E$  are  $e_i = \{0, i\}$  for all  $1 \leq i \leq L$ . Let us denote the weight of edge  $e_i$  by  $w_i$ .

## 5.1 Bounds using Sphere packings

It is well known that any embedding of general trees into  $\ell_2$  must incur a distortion of  $\Omega(L^{1/d})$  (see, e.g., [11]) using a very simple volume argument. We sketch a proof of this fact (and a matching upper bound) here for completeness.

**Proposition 5.1** *For each  $L > 2^d$ , any embedding of the (unweighted) star with  $L$  leaves into  $\mathbb{R}^d$  incurs a distortion  $\Omega(L^{1/d})$ . Further, any weighted star  $G = (V, E)$  with  $L$  leaves can be embedded into  $\mathbb{R}^d$  with distortion  $O(L^{1/d})$ .*

**Proof Sketch:** The graph  $G_k = (V_k, E_k)$  will be the star with  $k$  leaves, in which each edge has unit length. Consider any mapping  $f : V_k \rightarrow \mathbb{R}^d$ , and w.l.o.g., let  $f$  be an expansion and let  $f(0)$  be the origin. Now balls of radius  $\frac{1}{2}$  drawn around each leaf must be disjoint. However, there are  $k$  such balls with total volume  $O(k)$ , and hence one of these balls must have its center at distance  $\Omega(k^{1/d})$  away from the origin, and hence some edge must suffer a distortion of  $\Omega(k^{1/d})$ .

This lower bound is, in fact, tight for arbitrarily weighted stars. Let us assume the the edge weights are non-decreasing, i.e.,  $w_1 \leq w_2 \leq \dots \leq w_L$ . Further, let  $B_i$  denote a ball of radius  $w_i$ . We will map the vertex  $0$  to the origin, and then pack the balls  $B_i$  (in increasing order of  $i$ ) so that they are disjoint, and no  $B_i$  is at distance more than  $Dw_i$  from the origin, where  $D = cL^{1/d}$ . Now the map  $f$  sending vertex  $i$  to the center of  $B_i$  will have contraction 1, and thus a distortion of at most  $D$ . To embed  $B_i$ , we claim that there is a point at distance  $2w_i i^{1/d}$  from the origin, which is at distance  $w_i$  from the set  $\bigcup_{j < i} B_j$ . Indeed, if there was no such point, then increasing the radius of balls  $B_j$  (for  $j < i$ ) by  $w_i$  would cover the sphere of radius  $2w_i i^{1/d}$  around the origin. However, these  $(i - 1)$  balls do not have sufficient volume to cover the sphere, which completes the proof. ■

## 5.2 Random Projection arguments

We can also show that if we embed the star into high dimensions, there exists a  $d$ -dimensional subspace  $W$  such that projecting the map onto  $W$  incurs a distortion of only  $\tilde{O}(n^{1/d})$ .

Let us first embed the star  $G$  into  $L$  dimensions: map 0 to the origin, and map the vertex  $i$  to the point  $v_i = w(e_i)\vec{u}_i$ , where  $\vec{u}_i$  is the unit vector in the  $i$ -th dimension. This embedding has a distortion of at most  $\sqrt{2}$ . Now pick  $d$  independent vectors  $\{U_q\}_{q=1}^d$  from the  $L$ -dimensional Gaussian distribution with the identity covariance matrix (i.e. each coordinate  $U_{qi}$  is an independent Gaussian), and set the  $q$ -th coordinate of  $\phi(i)$  to  $\langle U_q, v_i \rangle$  (where  $\langle \cdot \rangle$  is the usual inner product). This is essentially the same as the Johnson-Lindenstrauss projection technique [7, 10], in which a random  $d$ -dimensional subspace is chosen and the high-dimensional map is projected onto this subspace, since  $d$  randomly chosen vectors (for small values of  $d$ ) are nearly orthogonal to each other and define a random  $d$ -dimensional subspace with high probability.

Note that the vertex 0 is again mapped to the origin. For any leaf vertex  $i$ , the values in the  $d$  coordinates of  $\phi(i)$  are independent and distributed according to the normal distribution with mean 0 and variance  $w_i^2$ . Also, the vectors  $\phi(i)$  and  $\phi(j)$  are independent of each other. Define  $\mathcal{E}_{0i}$  to be the event that the edge  $e_i = \{0, i\}$  suffers distortion of more than  $D$  (where the value of  $D$  will be decided later), and  $\mathcal{E}_{ij}$  to be the event that the vertex pair  $\{i, j\}$  suffers a distortion of greater than  $D$ . (Note that this implies that  $\mathcal{E}_{ij} = \mathcal{E}_{ji}$ , though we will often count these as two events to make the presentation simpler.) The observation above shows that the event  $\mathcal{E}_{0i}$  is dependent only on the  $2(n-1)$  events  $\mathcal{E}_{ij}$  and  $\mathcal{E}_{ji}$ , where  $j \neq i$ , and that  $\mathcal{E}_{ij}$  is dependent only on  $\mathcal{E}_{0i}$ ,  $\mathcal{E}_{0j}$ ,  $\mathcal{E}_{ik}$  and  $\mathcal{E}_{kj}$ , where  $k \notin \{i, j\}$  (also at most  $2(n-1)$  events). This means we are now in a situation where the events have only limited dependence and we can apply the Lovász Local Lemma (see, e.g., [1]):

**Theorem 5.2 (Lovász Local Lemma)** *Let  $\{\mathcal{B}_i\}$  be a set of “bad” events such that each  $\mathcal{B}_i$  depends on at most  $t$  other events  $\mathcal{B}_j$ . If  $\Pr[\mathcal{B}_i] \leq p$  and  $ep(t+1) < 1$ , then  $\Pr[\wedge \mathcal{B}_i] > 0$ .*

In our case,  $t = 2(n-1)$ , so it suffices that  $p < 1/2en$  to satisfy the hypothesis of the Local Lemma. Now using Lemma 7 of [6] or Lemma 2.2 of [4], we can show that setting  $D = O(n^{1/d}\sqrt{(\log n)/d})$  ensures that  $\Pr[\mathcal{E}] \leq 1/2en$  for any of the bad events described above, and thus there is a random  $k$ -dimensional subspace onto which we can project with  $\tilde{O}(n^{1/d})$ -distortion.

Unfortunately, this proof breaks down for more complicated graphs, since the limited independence property does not hold in the same fashion. It would be interesting to see if an improved analysis of the random projection technique could be obtained for broader classes of trees.

## Acknowledgements

Many thanks to Marshall Bern, Sanjoy Dasgupta, Ashwin Nayak and Alistair Sinclair for useful discussions and comments. Thanks also to the referees for many useful comments and suggestions.

## References

- [1] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley Interscience, New York, 1992.
- [2] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel Journal of Math.*, 56(2):222–230, 1986.
- [3] J. H. Conway and N. J. Sloane. *Sphere Packings, Lattices and Groups*. Springer-Verlag, 1992.
- [4] S. Dasgupta and A. Gupta. A simple proof of the Johnson-Lindenstrauss lemma. Technical Report 99-006, International Computer Science Institute, 1999.
- [5] A. Gupta. Embedding trees into low dimensional Euclidean spaces. In *31th Annual IEEE Symposium on Theory of Computing*, pages 694–700, 1999.
- [6] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pp. 604–613. 1998.
- [7] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
- [8] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. (Preliminary version in: *35th Annual Symposium on Foundations of Computer Science*, pages 577–591, 1994).
- [9] N. Linial, A. Magen, and M. Saks. Trees and Euclidean metrics. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pp. 169–177. 1998.
- [10] J. Matoušek. Bi-Lipschitz embeddings into low dimensional Euclidean spaces. *Comment. Math. Univ. Carolinae*, 31(3):589–600, 1990.
- [11] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel Journal of Math.*, To appear, 1999. (Czech version in : *Lipschitz distance of metric spaces*, C.Sc. degree thesis, Charles University, 1990).

## A Caps on spheres: Building good spherical codes

In this section, we show how to get good spherical codes with the specifications described in section 2. To build a spherical code with  $N$  vectors, all lying in the positive orthant, we will define  $N$  non-intersecting *caps* of unit radius on the surface of the sphere of radius  $r$  such that all the caps lie in the positive orthant, and

then choose the vectors of the code to be the positions of the centers of these caps. A standard argument, which we sketch below, shows that the radius  $r$  has to be  $\Theta(N^{1/(d-1)})$ .

Let  $S$  be a sphere of radius  $r$  centered at  $o$ . A *cap*  $C$  (of *angular radius*  $\delta$  or *spherical radius*  $\rho = r\delta$ ) on  $S$  is the portion of  $\partial S$  cut out by fixing an axis  $\ell$  passing through  $o$ , and rotating about it a “cutting” line passing through  $o$  making an angle  $\delta$  with  $\ell$ . The area of the  $d$ -dimensional cap of angular radius  $\delta$  is  $\frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)} r^{d-1} \int_0^\delta \sin^{d-2} \phi \, d\phi$ . We can use the simple fact that  $\sin \phi \leq \phi$  to upper bound the integral by  $\int_0^\delta \phi^{d-2} \, d\phi = \frac{\delta^{d-1}}{d-1}$ , and thus upper bound the area of the cap by  $V_{d-1}(r\delta)$ .

To see that  $r = O(N^{1/(d-1)})$ , let  $N'$  be the maximum number of non-overlapping caps possible whose centers lie in the positive orthant. If we increase the spherical radius of the  $N'$  caps to 2, they have to cover the surface of the sphere in the positive orthant. Indeed, if there is a point  $x$  not covered, then it is at distance at least 2 from all the other centers, and so another cap of radius 1 could be placed at  $x$  without intersecting the  $N'$  original caps. Since the volume of a cap of spherical radius 2 is bounded above by  $V_{d-1}(1)$ , we get that  $N'V_{d-1}(2) \geq S_d(r)/2^d$ , and hence  $N' \geq (r/2)^{d-1}/2^d$ . Taking  $r = r_0(N) = 8N^{1/(d-1)}$  will ensure that  $N' \geq N$ . Note that the minimum angle between any two of the chosen vectors is at least  $\delta_0(N) = 2/r_0(N)$ .

To see that  $r = \Omega(N^{1/(d-1)})$ , observe that  $\sin \phi \geq \phi/2$  for  $\phi \leq \pi/2$  and the area of the cap of spherical radius  $\rho$  is at least  $V_{d-1}(\rho/2)$ . Since the caps must be disjoint, simple algebra shows that we can pack at most  $\sqrt{\pi d}(2r)^{d-1}$  caps in  $\partial S$ , and thus  $r = \Theta(N^{1/(d-1)})$ .

**Remark:** The above construction is an existential one, and thus does not give us a polynomial time algorithm for the embedding. However, we can build a spherical code with slightly worse parameters in polynomial time in the following manner: let  $\Sigma_d$  be the surface of a sphere  $S_d \subseteq \mathbb{R}^d$  of radius  $r$ , and suppose we want to find points which are at least spherical distance 2 from each other. Let us consider the family of hyperplanes  $\mathcal{H}_i = \{x_d = 2i\}$  for all  $i \in \mathbb{N}$ , and let  $\Sigma_d(i) = \mathcal{H}_i \cap \Sigma_d$ . Each of the  $\Sigma_d(i)$  are surfaces of spheres in dimension one lower, and we can recursively construct well-separated sets of points (which are distance at least 2 from each other) on each of the  $\Sigma_d(i)$ . (The base case is when we have a circle in 2 dimensions, when the problem is trivial.) Now it can easily be shown that, for any point on the sphere  $\Sigma_d$ , there will be a point we construct which is at distance at most  $2d$  from it. A volume argument identical to the ones above shows that taking  $r$  to be  $O(dN^{1/(d-1)})$  suffices to give us  $N$  well-separated points. Using this spherical code gives us an polynomial-time embedding of trees into  $\mathbb{R}^d$  with distortion at most  $O(dD(f))$ , where  $D(f)$  is the bound on the distortion of the embedding of section 3.

## B An Example

In this section, we give a simple example to illustrate the embedding procedure described in section 3. The tree in figure 2 has all its edges of length 1, except that  $(1,5)$  has length 1.5, and  $(6,9)$  has length 0.5. A caterpillar decomposition of width 2 is indicated by the texture of the lines in the figure. Figure 3 indicates the steps in the embedding, while figure 4 shows the final embedding. Note that all the paths in the caterpillar decomposition have been mapped along straight lines.

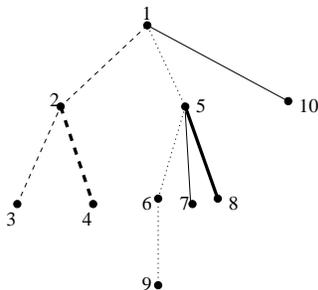


Figure 2: A tree

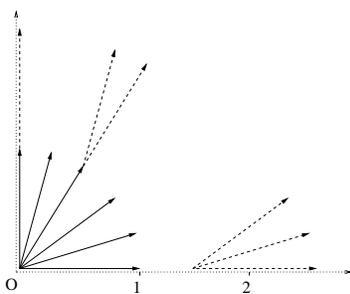


Figure 3: The embedding procedure

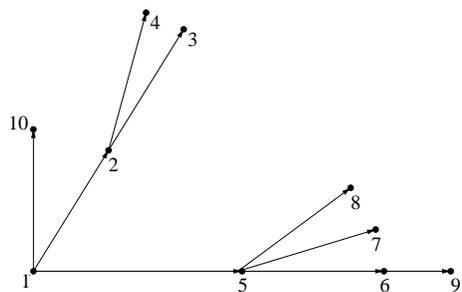


Figure 4: Embedding of the tree