Low distortion Euclidean embeddings of trees

Nathan Linial * Avner Magen † Michael E. Saks ‡

December 30, 1999

1 Introduction

Embeddings of finite metric spaces into normed spaces are of interest in the local theory of Banach spaces ([?, ?, ?]), in combinatorics (e.g. [?] and the references therein) and in the theory of algorithms [?]. In order to study a general metric, one seeks first an approximation in the form of a metric that is induced from a norm. The quality of such an approximation is quantified by the *distortion* of the corresponding embedding, i.e., the bi-Lipschitz constant of the mapping.

Definition 1.1: Let (X, d) be a metric space, $(Y, \|\ \ \|)$ a normed space and $f : X \to Y$ a mapping. The *expansion* of f is $\sup_{x,y \in X} \frac{\|f(x)-f(y)\|}{d(x,y)}$. The *shrinkage* of f is $\sup_{x,y \in X} \frac{d(x,y)}{\|f(x)-f(y)\|}$ and the *distortion* of f is the product of its expansion and its shrinkage. ∎

The following is a basic result in this area. The existential part is due to Bourgain [?] and the tightness is from [?]. If (X, d) is a finite metric space,

*Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel. E-mail: nati@cs.huji.ac.il. Supported in part by grants from the Israeli Academy of Sciences and the Binational Science Foundation Israel-USA.

†Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel. E-mail: avnerma@cs.huji.ac.il.

‡Department of Mathematics, Rutgers University, New Brunswick, NJ 08854, USA. Supported in part by NSF under grants CCR-9215293 and by DIMACS, which is supported by NSF grant STC-91-19999 and by the New Jersey Commission on Science and Technology.

we denote by $c_2(X)$ the least distortion with which (X, d) may be embedded in Euclidean space l_2 (regardless of dimension).

Theorem 1.2: *If (X, d) is a metric space with n points, then $c_2(X) \leq O(\log n)$. The bound is tight.*

In view of this theorem, it is interesting to study the parameter c_2 for various families of finite metric spaces and to understand when it is small or large. In attempting to develop a metric theory of superreflexivity, Bourgain determined c_2 of the complete binary tree T_n with n vertices (with the graph metric).

Theorem 1.3: *(Bourgain)*

$$c_2(T_n) = \Theta(\sqrt{\log \log n}).$$

We deal with general weighted trees. Given a tree T with a weight function w on the edges, a metric on $V(T)$ is induced in the obvious way by interpreting the weights as edge lengths. Formally, let $\pi(x, y)$ be the set of edges on the path that connects the vertices x and y. The distance $d(x, y) = d_{T,w}(x, y)$ is defined as $\sum_{e \in \pi(x,y)} w(e)$. We also introduce the convention that if F is a set of edges, then $\sum_{e \in F} w(e)$ is denoted by $w(F)$. With this notation, $d_{T,w}(x, y) = w(\pi(x, y))$. We write $c_2(T, w)$ for the least distortion of a Euclidean embedding of this metric space. Let $l(T)$ denote the number of leaves of T. Our main result is:

Theorem 1.4: *Let T be a tree and let w be any positive function on its edges. Then*

$$c_2(T, w) \leq O(\log \log l(T)).$$

The proof of this theorem appears in the next section. Our construction essentially generalizes that of Bourgain (although the analysis in the general case gives a slightly worse bound than for his special case). Following the proof we make some remarks about the connection between Bourgain's work and ours.

2 Constructing a good embedding

It will be convenient to designate some (arbitrary) vertex r as the tree's root. Denote $\pi(x,r)$ by $\pi(x)$. The vertices of T are partially ordered "from the root" in the usual way: $x \succ y$ if x lies on the path from y to r. A path that joins two $\succ$-comparable vertices is a *monotone* path.

An embedding of T into $\mathbb{R}^m$ is a map ϕ from $V(T)$ to $\mathbb{R}^m$. Since we can translate an embedding by any fixed vector in $\mathbb{R}^m$ without affecting its distortion, we will restrict attention to embeddings that map the root to the origin. It is convenient to represent such an embedding by a map ψ on edges: for an edge (u,v) with $v \succ u$, $\psi(u,v) = \phi(u) - \phi(v)$. Clearly, ϕ is determined uniquely by ψ via:

$$\phi(x) = \sum_{e \in \pi(x)} \psi(e). \tag{1}$$

To construct a good embedding for T, we will describe a map ψ on edges; our embedding will be the map $\phi = \phi_\psi$ given by (**??**). Now, for any embedding f of a weighted graph G into any metric space M, the maximum of $d_M(f(x),f(y))/d_G(x,y)$ is attained when x and y are adjacent, for if $x = x_0, x_1, \ldots, x_k = y$ is a shortest path from x to y in G then

$$\frac{d_M(f(x),f(y))}{d_G(x,y)} \leq \frac{\sum_i d_M(f(x_i),f(x_{i+1}))}{\sum_i d_G(x_i,x_{i+1})} \leq \mathbf{max}_i \frac{d_M(f(x_i),f(x_{i+1}))}{d_G(x_i,x_{i+1})}.$$

In terms of the map ϕ, we have:

Proposition 2.1: *The expansion of the map ψ_ϕ is the maximum of $\frac{\|\phi(e)\|_2}{|w_e|}$ over all edges e of T.*

To motivate our construction, we describe a sequence of three choices for the function ψ, the last being the one that attains the bounds in the theorem.

2.1 A simple construction.

A simple choice for ψ maps E into the Euclidean space $\mathbb{R}^E$ (whose coordinates correspond to T's edges) under the map $\psi(e) = w(e)\vec{u}^e$ where $\vec{u}^e$ is the unit vector corresponding to e. By Proposition **??**, the expansion of the corresponding map ϕ is 1. To bound the shrinkage of ϕ, note that for any two vertices x, y, $d_{T,w}(x,y) = \sum_{e \in \pi(x,y)} w(e)$, while $\|\phi(x) - \phi(y)\|_2 = \sqrt{\sum_{e \in \pi(x,y)} w(e)^2}$. The ratio $\sum_{e \in \pi(x,y)} w(e) / \sqrt{\sum_{e \in \pi(x,y)} w(e)^2}$ is maximized when all of the $w(e)$ are equal, and this leads to an upper bound on the shrinkage, and also on the distortion, of $\sqrt{diam(T)}$. This is in general much worse than the distortion that will be achieved.

It is easy to show that the above construction is an isometric (distance preserving) embedding of the tree into $\mathbb{R}^E$ under the l_1 norm. On the other hand, paths are the only trees which embed isometrically into Euclidean space: Say that a set S in a metric space is *collinear* if every three points in it satisfy the triangle inequality with equality. In Euclidean space, this definition coincides with the usual meaning of collinearity. If T is not a path, let x be a vertex with three distinct neighbors y_1, y_2, y_3. Suppose ϕ is an isometric embedding of T into some Euclidean space. Since the sets $\{x, y_1, y_2\}$ and $\{x, y_1, y_3\}$ are collinear in T, their images under ϕ are collinear as well. This implies that the whole set $\{\phi(x), \phi(y_1), \phi(y_2), \phi(y_3)\}$ resides on a line, whence the metric on $\{y_1, y_2, y_3\}$ is distorted by ϕ.

What we can do, however, is to employ the next mechanism: take a decomposition of the tree's vertices to relatively few simple parts that intersect in at most one vertex, map each part isometrically, and "glue the parts together" in an efficient way. This is just what we do in the next two constructions.

2.2 An improved construction

We start this construction by partitioning the edges of T into a collection $\mathcal{P}$ of monotone paths. The description of this partition is given below. For an edge e, let $P_e = P_e(\mathcal{P})$ denote the unique path of $\mathcal{P}$ containing e. We map E into $\mathbb{R}^{\mathcal{P}}$, i.e., there is one coordinate corresponding to each member of $P \in \mathcal{P}$ and $\vec{u}^P$ is the unit vector corresponding to this coordinate. We define the mapping ψ via: $\psi(e) = w(e)\vec{u}^{P_e}$. As before, the corresponding embedding ϕ_ψ has expansion 1. To bound the shrinkage, we first introduce some notation:

- The set of edges common to a path $P \in \mathcal{P}$ and to the path $\pi(x,y)$ between x and y is denoted by $\pi_P(x,y)$.

- $\Delta_{x,y} = \Delta_{x,y}(\mathcal{P})$ is the set of paths $P \in \mathcal{P}$ that meet the path between x and y. The cardinality $|\Delta_{x,y}|$ is denoted $\delta_{x,y}$.
- For a vertex x, we index the paths in $\Delta_{x,r}$ as $P^0(x), P^1(x), \ldots, P^{\delta_{x,r}-1}(x)$ according to the order at which they are encountered in traversing from x to r.
- For a vertex x, we write $\pi_P(x)$ for $\pi_P(x,r)$, Δ_x for $\Delta_{x,r}$ and δ_x for $\delta_{x,r}$.
- The maximum of δ_x (over all vertices x) is called $\delta = \delta(\mathcal{P})$.
- If e is an edge, b_e denotes e's vertex that is farthest from r. We write Δ_e for Δ_{b_e} and δ_e for δ_{b_e}.
- The *depth* of $P \in \mathcal{P}$ is the number of $Q \in \mathcal{P}$ that are either P or that lie "above it" (this coincides with δ_e for any $e \in P$).

To bound the shrinkage, consider any two vertices x and y:

$$\phi(y) - \phi(x) = \sum_{P \in \Delta_{x,y}} w(\pi_P(x,y))\vec{u}^P$$

whence

$$\begin{aligned} \|\phi(y) - \phi(x)\| &= \sqrt{\sum_{P \in \Delta_{x,y}} [w(\pi_P(x,y))]^2} \\ &\geq \frac{1}{\sqrt{\delta_{x,y}}} \sum_{P \in \Delta_{x,y}} w(\pi_P(x,y)) \\ &= \frac{d_{T,w}(x,y)}{\sqrt{\delta_{x,y}}} \end{aligned}$$

Thus the distortion of this map is at most $\max\sqrt{\delta_{x,y}}$, where the maximum is over all pairs of vertices x, y. Since $\delta_{x,y} \leq \delta_x + \delta_y$, the distortion does not exceed $\sqrt{2\delta(\mathcal{P})}$.

Thus, in order to minimize the distortion, we seek a partition $\mathcal{P}$ in which $\delta(\mathcal{P})$ is small.

Lemma 2.2: *Every rooted tree T has a partition $\mathcal{P}$ into monotone paths with $\delta(\mathcal{P}) \leq \log_2(2l(T) - 2)$.*

Proof: By induction on $|E(T)|$. If T is a path, the result is trivial. Henceforth we may assume $|E(T)| > 1$ and $l(T) > 2$.

Let s be the vertex closest to the root r (possibly r itself) having at least 2 children, and let $s_1, s_2, \ldots, s_d$ be the children of s. Let T_i be the tree rooted at s consisting of s together with the subtree of T rooted at s_i. By the induction hypothesis, the edges of each T_i have a partition $\mathcal{P}_i$ into monotone paths such that $\delta(\mathcal{P}_i) \leq \log(2l(T_i) - 2)$.

In the case $s = r$ we define the partition $\mathcal{P} = \cup_i \mathcal{P}_i$ of $E(T)$ and we have $\delta(\mathcal{P}) = \max_i \delta(\mathcal{P}_i) \leq \max_i \log(2l(T_i) - 2) \leq \log(2l(T) - 2)$.

When $s \neq r$, say that $l(T_1) = \max_i l(T_i)$. The partition $\mathcal{P}$ of $E(T)$ is obtained by the following modification of $\cup_i \mathcal{P}_i$: Let Q be the path of $\mathcal{P}_1$ that contains the vertex s. Replace Q by the path Q' that is the concatenation of Q and the path from s to r. Now, for any vertex v of T_1, $\delta_v(\mathcal{P}) = \delta_v(\mathcal{P}_1)$ and for $i > 1$ and any vertex v of T_i, $\delta_v(\mathcal{P}) = 1 + \delta_v(\mathcal{P}_i)$. Hence

$$\begin{aligned}
\delta(\mathcal{P}) &= \max\{\delta(\mathcal{P}_1), 1 + \max_{i\geq 2}\delta(\mathcal{P}_i)\} \\
&\leq \max\{\log(2l(T_1) - 2), \max_{i\geq 2}\log(4l(T_i) - 4)\} \\
&\leq \log(2l(T) - 2).
\end{aligned}$$

The last inequality follows from $l(T) - 1 = \sum_i (l(T_i) - 1) \geq 2\max_{i\geq 2} l(T_i) - 2$. ∎

Corollary 2.3: *The construction just described has distortion $O(\sqrt{\log l(T)})$.*

2.3 The final construction.

The construction that achieves the bound of Theorem **??** is a modification of the previous construction. Recall that previously $\psi(e)$ was defined as $w(e)$ times the unit vector $\vec{u}^{P_e}$. In the modified version, $\psi(e)$ will be $w(e)$ times a weighted sum of unit vectors $\vec{u}^P$ where P ranges over all paths in Δ_e. More specifically, we fix positive constants $a_0, a_1, a_2, \ldots, a_{\delta-1}$ (to be specified later) and define:

$$\psi(e) = w(e) \sum_{i=0}^{\delta_e - 1} a_i \vec{u}^{P^i(b_e)}$$

As before, the embedding ϕ is induced from ψ via (**??**).

Note here that if e and e' are in the same member of $\mathcal{P}$, then $\psi(e)/w(e) = \psi(e')/w(e')$, since $P^i(b_e) = P^i(b_{e'})$. Consequently, the restriction of ϕ to any path in $\mathcal{P}$ is indeed an isometry (times some constant).

We proceed to bound the expansion of this map. Let $a = (a_0, a_1, a_2, \ldots, a_{\delta-1})$.

Lemma 2.4: *expansion*$(\phi) \leq \|a\|_2$.

Proof: By Proposition **??**, it suffices to find out the expansion of edges e. But

$$\|\psi(e)\|_2 = w(e)(\sum_{i=0}^{\delta_e-1} a_i^2)^{\frac{1}{2}} \leq w(e)(\sum_{i=0}^{\delta-1} a_i^2)^{\frac{1}{2}} = w(e)\|a\|_2.$$

∎

Next we bound the shrinkage.

Lemma 2.5: *shrinkage*$(\phi) \leq \sqrt{2}\|b\|_2$*, where b is the (unique) vector satisfying*

$$\forall \quad 0 \leq j < \delta \qquad \sum_{i=0}^{j} a_i b_{j-i} = 1 \tag{2}$$

We refer to the above condition as the convolution condition.

Proof:

Let $W = W_{i,j}$ be the following δ by δ matrix

$$W_{i,j} = \begin{cases} a_{j-i} & \text{if } j \geq i \\ 0 & \text{otherwise.} \end{cases}$$

and let $C(W) = \sup\{\frac{\|x\|_1}{\|Wx\|_2} : x \in (\mathbb{R}^+)^\delta \setminus \{\vec{0}\}\}$.

To prove Lemma **??** we first prove:

Claim 2.6: *shrinkage*$(\phi) \leq \sqrt{2}C(W)$.

Proof:

We first check the shrinkage of the distance between $\succ$-comparable vertices, say v and its descendant v'. The supports of both $\phi(v)$ and $\phi(v')$ are contained in $\Delta_{v'}$. Ignoring some zero coordinates, we view $\phi(v') - \phi(v)$ as a δ-dimensional vector, where the coordinate corresponding to a path $P \in \Delta_{v'}$ is enumerated by P'th depth. The remaining highest $\delta - \delta_{v'}$ coordinates are zero.

Let $U^{(v',v)}$ denote the projection onto $\Delta_{v',v}$ (viewed as a linear transformation from $\mathbb{R}^{\mathcal{P}}$ to $\mathbb{R}^{\mathcal{P}}$).

Now, let y and z be the vectors in $\mathbb{R}^{\delta}$ defined by

$$y_j = \sum_{\substack{e \in \pi(v,v') \\ \delta_e = j}} w_e$$

$$z_j = \sum_{\substack{e \in \pi(v,v') \\ \delta_e = j + \delta_{v'} - \delta_{v',v}}} w_e$$

In words, for each path P of depth j in $\Delta_{v',v}$, $y_j = w(\pi_P(v,v'))$, and $\vec{z}$ is just $\vec{y}$ when we shift the indexing of the coordinates by $\delta_{v'} - \delta_{v',v}$.

By definition $d(v',v) = \|\vec{y}\|_1 = \|\vec{z}\|_1$. Also, it is not hard to see that the nonzero coordinates of $\phi(v') - \phi(v)$ are the same as the nonzero coordinates of $W\vec{y}$ and so $\|\phi(v) - \phi(v')\|_2 = \|W\vec{y}\|_2$.

Similarly the nonzero coordinates of $U^{(v',v)}(\phi(v') - \phi(v))$ are the same as the nonzero coordinates of $W\vec{z}$ and so $\|U^{(v',v)}(\phi(v') - \phi(v))\|_2 = \|W\vec{z}\|_2$.

Therefore:

$$\frac{d(v',v)}{\|\phi(v') - \phi(v)\|_2} \leq \frac{d(v',v)}{\|U^{(v',v)}(\phi(v') - \phi(v))\|_2} = \frac{\|z\|_1}{\|Wz\|_2} \leq C(W).$$

Note that the above not only shows $d(v',v)/\|\phi(v') - \phi(v)\|_2 \leq C(W)$, but also that the inequality remains true even when we replace the denominator by the length of the projection of $\phi(v') - \phi(v)$ on the coordinates in $\Delta_{v',v}$. We will use this in proving the claim for a pair of $\succ$-incomparable vertices.

Let v',v'' be two $\succ$-incomparable vertices, and let v be their lowest (farthest from the root) common ancestor.

Since the paths of $\mathcal{P}$ are monotone, the vectors $U^{(v',v)}(\phi(v) - \phi(v'))$ and $U^{(v'',v)}(\phi(v'') - \phi(v))$ have disjoint supports, and are therefore perpendicular. Now

$$
\begin{aligned}
&C(W)^2\|\phi(v')-\phi(v'')\|_2^2 \geq \\
&\geq \; C(W)^2\|U^{(v',v)}(\phi(v')-\phi(v))-U^{(v'',v)}(\phi(v'')-\phi(v))\|_2^2 \\
&= \; C(W)^2\left(\|U^{(v',v)}(\phi(v')-\phi(v))\|_2^2+\|U^{(v'',v)}(\phi(v'')-\phi(v))\|_2^2\right) \\
&\geq \; d^2(v',v)+d^2(v'',v) \geq \frac{1}{2}d^2(v',v'')
\end{aligned}
$$

∎

Thus to prove Lemma **??** it suffices to show $C(W) \leq \|b\|_2$, where b is the vector satisfying the convolution condition (**??**). Let us first note that $C(W)$ is the smallest C for which

$$x^t(W^tW-\frac{1}{C^2}J)x \geq 0 \quad \text{holds for all } x \in (\mathbb{R}^+)^\delta$$

(where J is the all-1 matrix). The matrix $M = W^tW$ is clearly positive definite (W is nonsingular), so the minimal C for which this condition holds, is at most $\theta^{-\frac{1}{2}}$, where θ is the smallest positive number for which $M-\theta J$ is singular.

We thus proceed to consider the equation $det(M-\theta J)=0$. Let A and B be two $n \times n$ matrices, and let S be a subset of $\{1,2,\ldots,n\}$. We denote by $\Delta_S(A,B)$ the $n \times n$ matrix whose i-th column is either the i-th column of A or the i-th column of B according to whether $i \notin S$ or $i \in S$. It is easy to see that

$$det(A+B)=\sum_{S\subseteq\{1,2,\ldots,n\}} det(\Delta_S(A,B)).$$

In the present case $A=M$ and $B=-\theta J$. Since B has rank 1, the only contribution in the sum is due to S such that $|S| \leq 1$, i.e.,

$$det(M-\theta J)=detM+\sum_{i=1}^{\delta} det(\Delta_{\{i\}}(M,-\theta J))$$

Recall Cramer's rule, that if Q is a square nonsingular matrix, and $Qx=y$, then $x_i = detQ^{(i)}(y)/detQ$, where $Q^{(i)}(y)$ is the matrix attained by replacing the i-th column of Q by y.

This implies that $det(\Delta_{\{i\}}(M,-\theta J))$ is just $-\theta \cdot detM \cdot (M^{-1}\ \vec{1})_i$.

Summing it all up, we conclude:

$$det(M-\theta J)=det(M)-\theta\cdot det(M)\cdot \vec{1}^t\ M^{-1}\ \vec{1}=det(M)(1-\theta\cdot\vec{1}^t\ M^{-1}\ \vec{1})$$

and so $\theta = (\vec{1}^t M^{-1}\vec{1})^{-1}$ is the only value for which $M - \theta J$ is singular. It follows that

$$C(W) = (\sum_{i,j} M_{i,j}^{-1})^{\frac{1}{2}}.$$

Now

$$(\sum_{i,j} M_{i,j}^{-1})^{\frac{1}{2}} = (\vec{1}^t M^{-1}\vec{1})^{\frac{1}{2}} = \|\vec{1}^t W^{-1}\|_2$$

But $\vec{1}^t W^{-1}$ is just the (unique) solution to the system

$$b^t W = \vec{1}^t.$$

In other words, b is the vector satisfying the convolution condition (**??**), as claimed. ∎

Lemma 2.7: *The vectors* $a_k = b_k = \binom{2k}{k}2^{-2k}$ *with* $k = 0, 1, \ldots \delta - 1$ *satisfy condition (**??**). Furthermore* $\|a\|_2 = \|b\|_2 = \Theta((\log\delta)^{\frac{1}{2}})$.

Proof: Consider the generating function for the (infinite) series (a_k), i.e., $f(x) = \sum_{k=0}^{\infty}\binom{2k}{k}2^{-2k}x^k$. But $f(x) = (1-x)^{-\frac{1}{2}}$, which can be viewed either as an identity in formal power series, or as the Taylor series of a real function in the range $|x| < 1$. Thus $f^2(x) = (1-x)^{-1} = \sum_{i=0}^{\infty} x^i$ which means

$$\forall j \geq 0 \quad \sum_{i=0}^{j} a_i b_{j-i} = 1$$

and in particular, if we let a and b be the first δ terms of the infinite series $\binom{2k}{k}2^{-2k}$, then condition (**??**) will hold. Now, since $\binom{2k}{k}2^{-2k} = \Theta(k^{-\frac{1}{2}})$, the l_2 norm of a and b is $\Theta((\log\delta)^{\frac{1}{2}})$. ∎

Proof: (of Theorem **??**) Combining Lemmas **??**, **??** and **??**, we conclude that $expansion(\phi) = O((\log\delta)^{\frac{1}{2}})$ and $shrinkage(\phi) = O((\log\delta)^{\frac{1}{2}})$, and so the distortion of ϕ is $O(\log\delta) = O(\log\log l(T))$.
∎

Our mapping essentially reduces to Bourgain's embedding for complete binary trees [**?**]. In that case, the members of $\mathcal{P}$ are the individual edges. Bourgain's construction uses $a_k = k^{-\frac{1}{2}}$ which asymptotically is the same as the present choice. Bourgain's result that the distortion is only $O(\sqrt{\log\log l(T)})$

in this special case, may be attained by noting that the only vectors z that arise in claim **??** are vectors of 1-s followed by 0-s.

How large can $c_2(T)$ get for n-vertex trees? The answer lies between $\sqrt{\log\log n}$ (Bourgain's lower bound for complete binary trees) and our present $\log\log n$ bound. We are still unable to close this gap.

Acknowledgments. We are grateful to Yuri Rabinovich for discussions that led to Lemma **??**. Thanks are also due to Eran London and to Shlomo Hoori for many useful comments on an earlier manuscript.

References

[1] L. Babai and P. Frankl, Linear Algebra Methods in Combinatorics, Preliminary Version 2, Department of Computer Science, The University of Chicago, Chicago, 1992.

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

[3] J. Bourgain, The metrical interpretation of superreflexivity in Banach spaces, Israel J. Math. 56 (1986) 222 - 230.

[4] J. Bourgain, T. Figiel and V. Milman, On Hilbertian subsets of finite metric spaces, Israel J. Math. 55 (1986) 147 - 152.

[5] N. Linial, E. London and Yu. Rabinovich, The geometry of graphs and some of its algorithmic applications, *Combinatorica*, 15(1995) 215 - 245.