

# On embedding trees into uniformly convex Banach spaces\*

JIŘÍ MATOUŠEK\*\*

Department of Applied Mathematics  
Charles University  
Malostranské nám. 25, 118 00 Praha 1  
Czech Republic

## Abstract

We investigate the minimum value of  $D = D(n)$  such that any  $n$ -point tree metric space  $(T, \rho)$  can be  $D$ -embedded into a given Banach space  $(X, \|\cdot\|)$ , that is, there exists a mapping  $f : T \rightarrow X$  with  $\frac{1}{D}\rho(x, y) \leq \|f(x) - f(y)\| \leq \rho(x, y)$  for any  $x, y \in T$ . Bourgain showed that  $D(n)$  grows to infinity for any superreflexive  $X$  (and this characterizes superreflexivity), and for  $X = \ell_p$ ,  $1 < p < \infty$ , he proved a quantitative lower bound of  $\text{const} \cdot (\log \log n)^{\min(1/2, 1/p)}$ . We give another, completely elementary proof of this lower bound, and we prove that it is tight (up to the value of the constant). In particular, we show that any  $n$ -point tree metric space can be  $D$ -embedded into a Euclidean space, with no restriction on the dimension, with  $D = O(\sqrt{\log \log n})$ .

## 1 Introduction

Let  $M$  be a metric space with metric  $\rho$ , let  $X$  be a normed space with norm  $\|\cdot\|$ , and let  $f : M \rightarrow X$  be a mapping. We say that  $f$  is a  $D$ -embedding,  $D \geq 1$  a real number, if we have

$$\frac{1}{D} \rho(x, y) \leq \|f(x) - f(y)\| \leq \rho(x, y)$$

for any two points  $x, y \in M$ . We say that  $M$   $D$ -embeds into  $X$  (or that  $M$  embeds into  $X$  with *distortion* at most  $D$ ) if there exists a  $D$ -embedding  $f : M \rightarrow X$ .

The  $D$ -embeddability of finite metric spaces into various normed spaces has been studied in a number of papers. This investigation started in the context of the local Banach space theory, where the general idea was to obtain some analogs for general metric spaces of notions and results dealing with the structure of finite-dimensional subspaces of Banach spaces. Early results in this area are due to Enflo [Enf69a, Enf69b], and some of the more recent references are Bourgain [Bou85, Bou86], Johnson and Lindenstrauss [JL84], Bourgain et al. [BMW86], Johnson et al. [JLS87], Arias-de-Reyna and Rodríguez-Piazza [AR92], and Matoušek [Mat96]. It turns out that  $D$ -embeddings can be of a considerable interest also in theoretical computer science and in some applied areas. They can serve as a useful representation of graphs and other metric spaces helping to visualize their structure, find clusters, small separators, etc.; see Linial et al. [LLR95].

Here we are going to consider the  $D$ -embeddability question for a special class of finite metric spaces, namely for the *tree metric spaces*. A tree metric space can be defined as a

---

\*This paper contains results from my thesis [Mat89] from 1990. Since then, I never got to translating this part to English, but since the subject of bi-Lipschitz embeddings is becoming increasingly popular I finally decided to publish it.

\*\*Supported by Czech Republic Grant GAČR 0194 and by Charles University grants No. 193,194.

metric space  $(T, \rho)$  satisfying the *four-point condition*: For any four points  $x, y, u, v$ , we have  $\rho(x, y) + \rho(u, v) \leq \min(\rho(x, u) + \rho(y, v), \rho(x, v) + \rho(y, u))$  (see e.g., [Dre84] for background and equivalent characterizations). A finite tree metric space  $(T, \rho)$  can be equivalently characterized as follows: There exists a finite (graph-theoretic) tree  $T_0$  with positive real weights on edges such that  $T$  is a subset of the vertex set  $V(T_0)$  and the metric  $\rho$  is the metric induced by the metric on  $T_0$  given by the edge weights. It is easy to see that we may assume  $|V(T_0)| \leq 2|T|$ , and so if we do not care about exact constants of proportionality, we can restrict ourselves to graph-theoretic trees with weighted edges.

Let  $B_m$  denote the *complete binary tree of height  $m$* . This is a graph defined as follows:  $B_0$  is a single vertex (the root), and  $B_{m+1}$  arises by taking one vertex (the root) and connecting it to the roots of two disjoint copies of  $B_m$ . We will regard  $B_m$  as a finite metric space (defined by the graph-theoretic distance on the vertex set, with edges of unit length).

Let  $\ell_p$  denote the space of countable sequences  $x = (x_1, x_2, \dots)$  of real numbers with  $\|x\|_p < \infty$ , where  $\|x\|_p = (\sum_{i=1}^{\infty} |x_i|^p)^{1/p}$ , and let  $\ell_p^n$  be the  $n$ -dimensional subspace spanned by the first  $n$  coordinates. Thus,  $\ell_2^n$  is the usual  $n$ -dimensional Euclidean space. It is well-known that any finite metric space can be isometrically embedded into  $\ell_{\infty}$  (this is an observation due to Fréchet), and it is not too difficult to show that any finite tree metric space is isometrically embeddable into  $\ell_1$ . Hence, the  $\ell_p$  spaces interesting for us are those with  $1 < p < \infty$ .

The embeddability of the complete binary trees into Banach spaces has been investigated by Bourgain [Bou86]. He proved that the distortion of any embedding  $B_m \rightarrow \ell_p$  is at least  $\Omega_p((\log m)^{\min(1/2, 1/p)}) = \Omega_p((\log \log n)^{\min(1/2, 1/p)})$ , where<sup>1</sup>  $n = 2^{m+1} - 1$  is the number of vertices of  $B_m$ . He also characterized the Banach spaces  $X$  such that all the  $B_m$ 's can be  $D$ -embedded into  $X$  for some  $D = D(X)$  independent of  $m$ . These are exactly the spaces that are not *superreflexive*. A superreflexive Banach space can be defined as one that admits an equivalent *uniformly convex* norm (a norm  $\|\cdot\|'$  is *equivalent* to  $\|\cdot\|$  if  $c\|x\| \leq \|x\|' \leq C\|x\|$  holds for all  $x$  and for some constants  $0 < c \leq C$ ). A Banach space  $X$  with norm  $\|\cdot\|$  is *uniformly convex* if for any  $\varepsilon > 0$  there exists  $\delta > 0$  with the following property: For any two points  $x, y \in X$  with  $\|x\| = \|y\| = 1$  and  $\|x - y\| \geq \varepsilon$ , we have  $\|(x + y)/2\| \leq 1 - \delta$ . Let  $\delta_X(\varepsilon)$  denote the infimum of the  $\delta$ 's with this property. The function  $\delta_X$  is called the *modulus of uniform convexity* of the space  $X$ . For the spaces  $\ell_p$ , the modulus of uniform convexity satisfies  $\delta_{\ell_p}(\varepsilon) = \Omega_p(\varepsilon^{\max(2, p)})$ .

Bourgain's proof is short and very elegant. It is formulated in the language of martingales and uses some results about martingales in superreflexive spaces. Here we give another, completely elementary and self-contained proof of his results (I have found this proof without being aware of Bourgain's work, but, not surprisingly, some of the basic ideas in both proofs are similar). Namely, we prove

**Theorem 1** *Let  $X$  be a uniformly convex Banach space whose modulus of convexity satisfies  $\delta_X(\varepsilon) \geq c\varepsilon^p$  for some constants  $p \geq 2$  and  $c \geq 0$ . Then the minimum distortion necessary for embedding  $B_m$  into  $X$  is at least  $c_1(\log m)^{1/p}$  for some  $c_1 = c_1(c, p) > 0$ .*

Bourgain showed that the lower bound is tight up to the value of the constant of proportionality for embedding  $B_m$  into a Euclidean space, i.e. for  $p = 2$ . We prove that *any*  $n$ -point tree metric space can be embedded into  $\ell_p$  with distortion at most  $O((\log \log n)^{\min(1/2, 1/p)})$ , where  $1 < p < \infty$ . In particular, embedding into a Euclidean space is possible with distortion  $O(\sqrt{\log \log n})$ . In fact, we define a certain measure of complexity (dimension) of a tree and we bound the distortion in terms of this dimension.

---

<sup>1</sup>Here the notation  $f = \Omega(g)$  is equivalent to  $g = O(f)$ , and the subscript in  $\Omega_p(\cdot)$  means that the constant of proportionality depends on  $p$ .

For technical reasons, we will work with rooted trees. The root can be chosen arbitrarily. We define the *caterpillar dimension* of a rooted tree  $T$ , denoted by  $\text{cdim}(T)$ , as follows: If  $T$  consists of a single vertex, the root, then we put  $\text{cdim}(T) = 0$ . For a tree  $T$  with at least 2 vertices, we define  $\text{cdim}(T) \leq k+1$  if there exist paths  $P_1, P_2, \dots, P_r$  beginning in the root and otherwise pairwise disjoint such that each component  $T_j$  of  $T - E(P_1) - E(P_2) - \dots - E(P_r)$  satisfies  $\text{cdim}(T_j) \leq k$ . Here  $T - E(P_1) - E(P_2) - \dots - E(P_r)$  denotes the tree  $T$  with the edges of the  $P_i$ 's removed, and the components  $T_j$  are rooted at the single vertex lying on some  $P_i$  (see also Fig. 1 few pages later).

Let us remark that a similar dimension is used in computer science (for unrooted trees), and that  $\text{cdim}(T)$  can be determined in polynomial time for a given rooted tree  $T$ , say by dynamic programming. The name is derived from the case of trees of dimension 1, which are caterpillars with legs. A basic example of a tree with a large caterpillar dimension is the complete binary tree  $B_m$ , for which  $\text{cdim}(B_m) = m$ . We also have the following easy

**Lemma 2** *For any rooted tree  $T$ ,  $\text{cdim}(T) \leq \log_2 |V(T)|$  holds.*

In Section 3, we prove this lemma and the following result.

**Theorem 3** *For any  $p \in (1, \infty)$  and for any finite tree metric space  $T$ , there exists an embedding of  $T$  into  $\ell_p$  with distortion  $O_p((\log \text{cdim}(T))^{\min(1/2, 1/p)})$ .*

Let us remark that the caterpillar dimension can be defined for infinite tree metric spaces as well. The theorem holds also for infinite tree metric spaces with a finite caterpillar dimension.

We will prove the theorem for  $p \geq 2$ . By Dvoretzky's theorem, any infinite-dimensional Banach space contains a  $(1 + \varepsilon)$ -isomorphic copy of  $\ell_2^n$  for each  $n$  and each  $\varepsilon > 0$ , and consequently there is an embedding with distortion  $O(\sqrt{\log \text{cdim}(T)})$  into any infinite-dimensional Banach space. Further, according to a theorem of Maurey and Pisier, an infinite-dimensional Banach space of cotype  $q \geq 2$  contains a  $(1 + \varepsilon)$ -isomorphic copy of  $\ell_q^n$  (again for all  $n$  and all  $\varepsilon > 0$ ), and so tree metric spaces can be embedded into such Banach spaces with distortion  $O((\log \text{cdim}(T))^{1/q})$  (the notions and results used in this paragraph can be found in [MS86], for instance).

Sometimes it is technically convenient with the following definition of the distortion of a mapping. If  $(X, \rho)$  and  $(Y, \sigma)$  are metric spaces and  $f : X \rightarrow Y$  is an injective mapping, we define the *Lipschitz norm* of  $f$  by  $\|f\|_{\text{Lip}} = \sup\{\sigma(f(x), f(y))/\rho(x, y) : x, y \in X, x \neq y\}$ , and the distortion<sup>2</sup> of  $f$  as  $\|f\|_{\text{Lip}} \cdot \|f^{-1}\|_{\text{Lip}}$ . A mapping with distortion 1 thus need not be an isometry but it may re-scale all distances in the same ratio. If a mapping with distortion  $D$  goes into a normed space then it can obviously be re-scaled in such a way that it becomes a  $D$ -embedding as was defined above. Sometimes it is simpler to write out the definition of a mapping with distortion  $D$  when the scaling factor is not required to be 1. A mapping  $f$  is called *non-contracting* if  $\|f^{-1}\|_{\text{Lip}} \leq 1$ .

## 2 Lower bound

In this section, we prove Theorem 1.

Let  $T_{k,m}$  denote the *complete  $k$ -ary tree of height  $m$* . This is an obvious generalization of the complete binary tree; in  $T_{k,m}$ , each non-leaf vertex has  $k$  successors. The *level* of a

---

<sup>2</sup>Also other terminology is used the literature; e.g., a mapping with distortion at most  $D$  is also called a *D-isomorphism*, or a *D-lipeomorphism*, and so on.

vertex of  $T_{k,m}$  is its distance from the root. Let  $P_n$  denote the set  $\{0, 1, 2, \dots, n\}$  regarded as a metric subspace of the real numbers with the usual metric.

The following easy geometric fact will connect the uniform convexity to embedding of trees; the rest of the proof is combinatorial. Let  $(X, \rho)$  be a metric space. Define a  $\delta$ -fork in  $X$  as a subspace  $F = \{x_0, x_1, x_2, x'_2\} \subseteq X$  such that both  $\{x_0, x_1, x_2\}$  and  $\{x_0, x_1, x'_2\}$  are  $(1 + \delta)$ -isomorphic to the space  $P_2$  (with  $x_1$  mapped to the middle point 1). The points  $x_2$  and  $x'_2$  are called the *tips* of  $F$ .

**Lemma 4 (Fork lemma)** *Let  $X$ ,  $\delta_X$ ,  $p$  be as in Theorem 1, and let  $F = \{x_0, x_1, x_2, x'_2\}$  be a  $\delta$ -fork in  $X$ . Then  $\|x_2 - x'_2\| = \|x_2 - x_1\| \cdot O(\delta^{1/p})$ .*

**Proof.** We may assume  $x_0 = 0$ ,  $\|x_1\| = 1$ . Set  $z = x_1 + \frac{x_2 - x_1}{\|x_2 - x_1\|}$ . We have  $\|z - x_1\| = 1$ ,  $\|z - x_2\| \leq 2\delta$ , and  $\|z\| \geq 2 - 4\delta$ . Put  $u = z - 2x_1$ . The vectors  $x = x_1$  and  $y = x_1 + u$  have unit norm, and for their midpoint  $\frac{x+y}{2} = x_1 + \frac{u}{2} = \frac{z}{2}$  we have  $\|\frac{x+y}{2}\| \geq 1 - 2\delta$ . Hence the uniform convexity condition implies  $\|u\| = O(\delta^{1/p})$ , and so  $\|x_2 - 2x_1\| \leq \|x_2 - z\| + \|z - 2x_1\| \leq 2\delta + \|u\| = O(\delta^{1/p})$ . By symmetry, we also have  $\|x'_2 - 2x_1\| = O(\delta^{1/p})$ , and thus  $\|x_2 - x'_2\| = O(\delta^{1/p})$  as claimed.  $\square$

Rather than dealing with the complete binary trees  $B_m$  in the proof of Theorem 1, it will be more convenient to lower-bound the distortion needed for embedding of trees  $T_{k,h}$  with a large  $k$  into  $X$ . We observe that  $T_{k,h}$  can be embedded into the complete binary tree  $B_{2h\lceil\log_2 k\rceil}$  with distortion at most 2. Indeed, if a vertex  $v$  of  $T_{k,h}$  has already been mapped to a vertex  $u$  at some level  $\ell$  of  $B_{2h\lceil\log_2 k\rceil}$ , then the  $k$  successors are mapped to  $k$  vertices above  $u$  at level  $\ell + 2\lceil\log_2 k\rceil$  whose mutual distances are all between  $\lceil\log_2 k\rceil$  and  $2\lceil\log_2 k\rceil$ .

For a rooted tree  $T$ , let  $SP(T)$  stand for the set of all unordered pairs  $\{x, y\}$  of vertices of  $T$  such that  $x$  lies on the path from  $y$  to the root. We need the following simple Ramsey-type result.

**Lemma 5** *Let  $h$  and  $r$  be given natural numbers, and suppose that  $k \geq r^{(h+1)^2}$ . Suppose that each of the pairs from  $SP(T_{k,h})$  is colored by one of  $r$  colors. Then there exists a copy  $T'$  of  $B_h$  in this  $T_{k,h}$  such that the color of any pair  $\{x, y\} \in SP(T')$  only depends on the levels of  $x$  and  $y$ .*

**Proof.** Label each leaf  $z$  of  $T_{k,h}$  by the  $\binom{h+1}{2}$ -element vector consisting of the colors of the pairs  $\{x, y\} \in SP(T_{k,h})$  lying on the path from  $z$  to the root (enumerated in some order common for all leaves  $z$ ). In this way, the leaves of  $T_{k,h}$  are colored by  $r' < r^{(h+1)^2}$  colors. We want to show the existence of a copy of  $B_h$  whose all leaves have the same color. By induction on  $h$ , we prove the following statement: *If the leaves of  $T_{k,h}$  are colored by  $r'$  colors and  $k > r'$  then there exists a copy of  $B_n$  whose all leaves have the same color.* The  $h = 0$  case is trivial. For  $h \geq 1$ , consider all the  $k$  subtrees isomorphic to  $T_{h-1,k}$  attached to the root of  $T_{k,h}$ . In each of them, select a copy of  $B_{h-1}$  with monochromatic leaves. Since  $k > r'$ , two of these copies have the same color of leaves and by connecting them to the root we get the desired copy of  $B_h$ .  $\square$

The next lemma says that if a copy of the metric space  $P_h$  is embedded into any metric space with a constant-bounded distortion, and  $h$  is large enough, then some 3-term arithmetic progression is embedded with distortion close to 1. In order to get asymptotically tight bound for the embedding of trees, we need a slightly complicated quantitative version.

**Lemma 6 (Path embedding lemma)** *For any given constants  $\alpha > 0$  and  $\beta \in (0, 1)$ , a constant  $C = C(\alpha, \beta)$  exists with the following property: Whenever  $f$  is a non-contracting mapping of the metric space  $P_h$  into some metric space  $(X, \rho)$  and  $h \geq 2^{CK^\alpha}$ , where  $K = \|f\|_{Lip}$ , then there exists a subspace  $Z = \{x, x + d, x + 2d\} \subseteq P_h$  such that if we denote by  $f_0$  the restriction of  $f$  on  $Z$ , then  $f_0$  is a  $(1 + \varepsilon)$ -isomorphism with  $\varepsilon = \beta \left( \frac{\rho(f(x), f(x+d))}{d} \right)^{-\alpha}$ . (Qualitatively: The more  $Z$  is expanded by  $f$ , the more precisely must  $f(Z)$  be isomorphic to  $P_2$ .)*

**Proof.** For  $d \in \{1, 2, \dots, h\}$ , define

$$K(d) = \max \left\{ \frac{\rho(f(x), f(y))}{|x - y|}; x \in P_h, |x - y| = d \right\}.$$

It is easy to see that  $K(d) \geq K(2d)$  for all  $d$ .

Next, define a sequence of numbers  $x_0 > x_1 > x_2 > \dots$  by setting  $x_0 = K$  and  $x_{j+1} = x_j / (1 + \frac{\beta}{4x_j^\alpha})$ . Let  $t$  be the first index with  $x_t \leq 1$ ; a simple calculation shows that  $t = O(K^\alpha)$ , and hence we may assume  $2^t \leq h$  (by taking  $C$  large enough). In the sequence  $K(2^0) \geq K(2^1) \geq K(2^2) \geq \dots \geq K(2^t)$ , there exist two consecutive values, say  $K(2^i)$  and  $K(2^{i+1})$ , lying in the same interval  $[x_{j+1}, x_j]$ . Hence  $1 \leq \frac{K(2^i)}{K(2^{i+1})} \leq 1 + \eta$ , where  $\eta = \frac{\beta}{4K(2^i)^\alpha}$ . We choose  $d = 2^i$  and fix points  $x, x + 2d \in P_h$  such that  $K(2d)$  is attained for them, i.e.  $\rho(f(x), f(x + 2d)) = 2dK(2d)$ . We have  $\rho(f(x), f(x + d)) \leq dK(d) \leq d(1 + \eta)K(2d)$ , and analogously  $\rho(f(x + d), f(x + 2d)) \leq d(1 + \eta)K(2d)$ . On the other hand,  $\rho(f(x), f(x + d)) \geq \rho(f(x), f(x + 2d)) - \rho(f(x + d), f(x + 2d)) \geq 2dK(2d) - d(1 + \eta)K(2d) = d(1 - \eta)K(2d)$ . From this, the conditions of the Lemma are straightforward to check.  $\square$

**Proof of Theorem 1.** The plan of the proof is quite simple. We consider a non-contracting embedding  $f : T_{k,h} \rightarrow X$ . Using Lemmas 5 and 6 we show that if  $\|f\|_{Lip}$  is small then there exists a 0-fork in  $T_{k,h}$  mapped to a  $\delta$ -fork in  $X$ , and this contradicts Lemma 4 since the tips of the fork are far apart in  $T_{k,h}$ . A more detailed proof follows.

Let  $\beta > 0$  be a sufficiently small constant (depending on  $p$  and  $c$ ), suppose that  $m$  is large enough, and let  $k, h$  be auxiliary parameters (their dependence on  $m$  will be fixed later). Suppose that  $f : T_{k,h} \rightarrow X$  is a non-contracting mapping with  $\|f\|_{Lip} = K = c_1(\log m)^{1/p}$ ; for  $c_1$  small enough, we derive a contradiction.

Set  $r = \frac{2K^{p+1}}{\beta}$ , and suppose that  $k \geq r^{(h+1)^2}$ . Color the pairs in  $SP(T_{k,h})$  according to the distortion of their distance by  $f$ ; namely, a pair  $\{x, y\} \in SP(T_{k,h})$  gets the color

$$\left\lfloor \frac{K^p}{\beta} \frac{\|f(x) - f(y)\|}{\tau(x, y)} \right\rfloor \in \{0, 1, \dots, r - 1\}$$

where  $\tau$  denotes the metric in  $T_{k,h}$ . By Lemma 5, there exists a copy  $T'$  of  $B_h$  in  $T_{k,h}$  such that the color of pairs  $\{x, y\} \in SP(T')$  only depends on the level of  $x$  and  $y$ . This means that all paths from the root to a leaf of  $T'$  are embedded in the same way by  $f$ , up to a distortion at most  $1 + \beta K^{-p}$ .

Let  $P$  be one such root-leaf path in  $T'$  (isometric to  $P_h$ ). If we set  $h = 2^{CK^p}$ , where  $C = C(p, \beta)$  is as in Lemma 6, we can select three vertices  $x_0, x_1, x_2$  of  $P$  at levels  $\ell, \ell + d, \ell + 2d$  such that  $f$  acts as a  $(1 + \delta)$ -isomorphism on this triple, where

$$\delta = \beta \left( \frac{\|f(x_0) - f(x_1)\|}{d} \right)^{-p}.$$

If we let  $x'_2$  be a vertex of  $T'$  at level  $\ell + 2d$  and at distance  $2d$  from  $x_2$ , we see that the  $f$ -images of  $x_0, x_1, x_2$ , and  $x'_2$  form a  $3\delta$ -fork in  $X$ . By Lemma 4, we get

$$2d \leq \|f(x_2) - f(x'_2)\| = \|f(x_0) - f(x_1)\|O(\delta^{1/p}) = O(\beta^{1/p}d)$$

where the constant of proportionality in the last  $O(\cdot)$  notation doesn't depend on  $\beta$ . In this way, we get a contradiction by choosing  $\beta$  small enough.

It remains to check the choice of the parameters. We have  $h \leq 2^{CK^p}$ , and by setting  $c_1$  in the expression  $K = c_1(\log m)^{1/p}$  sufficiently small, we can guarantee  $h < m^{1/4}$  (say). Then we have  $k = r^{(h+1)^2} = K^{O((p+1)\sqrt{m})} = \exp(O(\log \log m \sqrt{m}))$ , and so  $h \log k = O(m^{3/4}) < m$ . Therefore, by the observation made above Lemma 5, the tree  $T_{k,h}$  used in the above proof can be 2-embedded into  $B_m$ , and Theorem 1 is proved.  $\square$

### 3 Upper bound

**Proof of Lemma 2.** By induction on  $\text{cdim}(T)$ . Suppose that  $\text{cdim}(T) = k + 1$ ; it suffices to show the existence of two disjoint subtrees  $T'$  and  $T''$  of  $T$  with  $\text{cdim}(T'), \text{cdim}(T'') \geq k$ . For contradiction, suppose that no such  $T', T''$  exist. We define a path  $P_1$  inductively, starting in the root. Let  $v$  be the last vertex of  $P_1$  selected so far, and let  $v_1, \dots, v_m$  be its sons in  $T$  (if  $v$  is a leaf then the definition of  $P_1$  is finished). Let  $T_j$  denote the subtree of  $T$  rooted at  $v_j$ . If each of the  $T_j$ 's has caterpillar dimension at most  $k - 1$  we choose an arbitrary  $v_j$  as the next vertex of  $P_1$ . If  $\text{cdim}(T_j) = k$  for some  $j$  (note that such a  $j$  must be unique) we let the next vertex of  $P_1$  be  $v_j$ . This finishes the definition of  $P_1$ .

Each component of  $T - E(P_1)$  has caterpillar dimension at most  $k - 1$ , and this contradicts the assumption  $\text{cdim}(T) = k + 1$ .  $\square$

**Embedding the complete binary tree.** First we show how to embed the tree  $B_m$  into  $\ell_2$  with distortion  $O(\sqrt{\log m})$ ; this is very similar to Bourgain's embedding in [Bou86]. The embedding of an arbitrary weighted tree into  $\ell_p$  is considerably more complicated but it is based on a similar approach.

Let us identify the vertices of  $B_m$  with words of length at most  $m$  over the alphabet  $\{0, 1\}$ . The root of  $B_m$  is the empty word  $\Lambda$ , and the sons of a vertex  $w$  are the vertices  $w0$  and  $w1$ . Now we can define an embedding  $f : V(B_m) \rightarrow \ell_2^{|V(B_m)|-1}$ , where the coordinates in the range of  $f$  are indexed by the vertices of  $B_m$  distinct from the root, i.e. by nonempty words. For a word  $w \in V(B_m)$  on level  $a$  of  $B_m$ , that is, of length  $a$ , we define

$$f(w)_u = \sqrt{a - b + 1}$$

if  $u$  is a nonempty initial segment of  $w$  of length  $b$ , and

$$f(w)_u = 0$$

otherwise. In particular, we have  $f(\Lambda)_u = 0$  for all  $u$ . For example, the indices of the coordinates for embedding of  $B_2$  are 0, 1, 00, 01, 10, 11, and the embedding looks as follows:

$$\begin{array}{cccc} & & 000000 & \\ & & & 010000 \\ & 100000 & & \\ \sqrt{2}01000 & \sqrt{2}00100 & 0\sqrt{2}0010 & 0\sqrt{2}0001 \end{array}$$

It is easy to see that for bounding the distortion for such an embedding, it suffices to consider the distances of vertices  $u, v$  with  $u$  lying on the path from  $v$  to the root. Let  $u, v$  be such

vertices, and let  $u$  have level  $a$  and  $v$  level  $b$  in  $B_m$ ,  $a < b$ . Then we have

$$\begin{aligned} \|f(u) - f(v)\|_2^2 &= \sum_{i=1}^a \left( \sqrt{b-i+1} - \sqrt{a-i+1} \right)^2 + \sum_{i=a+1}^b \left( \sqrt{b-i+1} \right)^2 = \\ &= \sum_{i=1}^a \frac{(b-a)^2}{\left( \sqrt{b-i+1} + \sqrt{a-i+1} \right)^2} + \sum_{i=a+1}^b (b-i+1). \end{aligned}$$

The second sum in this expression is  $\Omega((b-a)^2)$  and hence  $\|f^{-1}\|_{Lip}$  is upper-bounded by a constant. The first sum can be estimated by

$$(b-a)^2 \sum_{i=1}^a \frac{1}{a+b+2-2i} \leq (b-a)^2 \sum_{j=1}^a \frac{1}{2j} \leq (b-a)^2 (\ln a + 1)$$

and so the distortion of  $f$  is bounded by  $O(\sqrt{\log m})$ .

**General weighted trees.** We need some preparatory lemmas.

**Lemma 7** *Let  $y > x > 0$  and  $0 < \alpha < 1$ . Then we have*

$$y^\alpha - x^\alpha \leq \frac{y-x}{y^{1-\alpha}}.$$

**Proof.** We have

$$y^\alpha - x^\alpha = y^\alpha \left( 1 - \left( \frac{x}{y} \right)^\alpha \right) \leq y^\alpha \left( 1 - \frac{x}{y} \right) = \frac{y-x}{y^{1-\alpha}}.$$

□

**Lemma 8** *Let  $x_1, x_2, \dots, x_k$  be real positive numbers. Then we have*

$$\sum_{i=1}^k \frac{x_i}{x_i + x_{i+1} + \dots + x_k + 1} \leq 1 + \ln(x_1 + x_2 + \dots + x_k + 1).$$

**Proof.** Put  $s_i = x_i + x_{i+1} + \dots + x_k + 1$ . Let

$$I_j = \left\{ i \in \{1, 2, \dots, k\} : \frac{1}{j+1} \leq \frac{x_i}{s_i} < \frac{1}{j} \right\},$$

and let  $k_j = |I_j|$ . Then the sum in the lemma is upper-bounded by the sum

$$1 + \sum_{j=1}^{\infty} \frac{k_j}{j}. \tag{1}$$

On the other hand, if  $i \in I_j$ , we get

$$\frac{s_{i+1}}{x_i} = \frac{s_i}{x_i} - 1 \leq j,$$

and so  $s_i = s_{i+1} + x_i = s_{i+1} \left( 1 + \frac{x_i}{s_{i+1}} \right) \geq s_{i+1} \left( 1 + \frac{1}{j} \right)$ . Hence

$$\prod_{j=1}^{\infty} \left( 1 + \frac{1}{j} \right)^{k_j} \leq s_1 = x_1 + x_2 + \dots + x_k + 1,$$

and the lemma follows by taking logarithms on both sides of this last inequality and by using  $\ln(1 + \frac{1}{j}) \leq \frac{1}{j}$  and (1). □

**Lemma 9** For arbitrary positive real numbers  $x, y, z$  and for any  $\alpha \in (0, 1)$ , we have

$$(x + z)^\alpha (y + z)^{1-\alpha} - x^\alpha y^{1-\alpha} \geq z.$$

**Proof.** It suffices to consider the case  $z = 1$ . We have  $(x + 1)^\alpha = x^\alpha (1 + \frac{1}{x})^\alpha \geq x^\alpha (1 + \frac{\alpha}{x})$ , and so

$$(x + 1)^\alpha (y + 1)^\alpha - x^\alpha y^{1-\alpha} \geq x^\alpha y^{1-\alpha} \left( \frac{\alpha}{x} + \frac{1-\alpha}{y} \right) = \alpha t^{1-\alpha} + (1-\alpha)t^{-\alpha},$$

where  $t = \frac{y}{x}$ . It is not difficult to check that for any  $\alpha \in (0, 1)$ , the only zero of the derivative of the function  $f(t) = \alpha t^{1-\alpha} + (1-\alpha)t^{-\alpha}$  is at  $t = 1$ . From this we can verify that  $f(t) \geq 1$  for all  $t > 0$ .  $\square$

**Proof of Theorem 3.** Let  $T$  be a tree metric space with metric  $\rho$ , and let us set  $m = \text{cdim}(T)$ . We define a *path partition* of the tree  $T$ . If  $T$  is a single vertex then the path partition is empty, and otherwise the path partition of  $T$  consists of some paths  $P_1, P_2, \dots, P_r$  as in the definition of  $\text{cdim}(T)$  plus the union of path partitions of the components of  $T - E(P_1) - \dots - E(P_r)$ . We say that the paths  $P_1, \dots, P_r$  have level 1 in such a path partition, and the paths of level  $k \geq 2$  are the paths of level  $k - 1$  in the corresponding path partitions of the components of  $T - E(P_1) - \dots - E(P_r)$ . Note that the paths in a path partition are edge-disjoint and together cover the edge set of  $T$  (see Fig. 1).

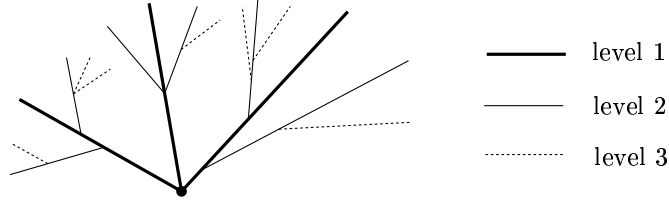


Figure 1: A schematic example of a path partition.

In the sequel, we assume that some path partition  $\Pi$  of  $T$  containing no path of level more than  $m$  is chosen once and for all. Let  $v$  be a vertex of the tree  $T$ , and consider the path from the root to  $v$ . The first segment of this path, of length  $d_1$ , follows some path  $P_1$  of level 1 in the part partition  $\Pi$ , the second segment, of length  $d_2$ , follows a path  $P_2$  of level 2,  $\dots$ , and the last ( $a$ th) segment of length  $d_a$  follows a path  $P_a$  of level  $a$  in  $\Pi$ ,  $a \leq m$ . Let us call the sequence  $(P_1, P_2, \dots, P_a)$  the *path sequence* of  $v$  and the sequence  $(d_1, d_2, \dots, d_a)$  the *path length sequence* of  $v$ .

Let  $q$  denote the exponent dual to  $p$ , that is, the one with  $\frac{1}{p} + \frac{1}{q} = 1$ . Now we can define an embedding  $f : V(T) \rightarrow \ell_p$ . The relevant coordinates in  $\ell_p$  will be indexed by the paths in  $\Pi$ . For  $f(v)$ , only the coordinates corresponding to the paths in the path sequence of  $v$  can be nonzero. If  $(d_1, d_2, \dots, d_a)$  is the path length sequence of  $v$ , we define the numbers

$$s_i(v) = d_i + \sum_{j=i+1}^a \max \left( 0, d_j - \frac{d_i}{2m} \right), \quad (2)$$

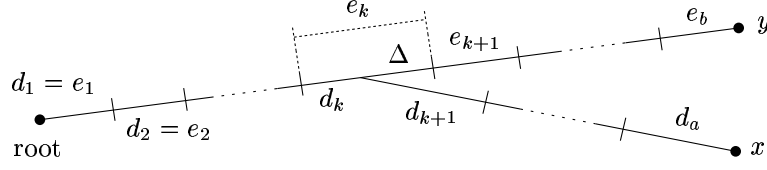
$i = 1, 2, \dots, a$ . Now we let the coordinate indexed by  $P_i$  of the point  $f(v)$  be

$$f(v)_{P_i} = d_i^{1/p} s_i^{1/q}.$$



It remains to estimate the distortion of  $f$ , which is not too difficult but a bit lengthy.

Let  $x$  and  $y$  be two vertices of  $T$  and let  $(d_1, d_2, \dots, d_a)$  and  $(e_1, e_2, \dots, e_b)$  be their respective path length sequences. Let  $k$  be the index such that exactly the first  $k$  terms of the path sequences of  $x$  and of  $y$  coincide, and suppose that the notation  $x, y$  is chosen in such a way that  $d_k \leq e_k$  (note that  $d_i = e_i$  for  $i = 1, 2, \dots, k-1$ ). Let us put  $s_i = s_i(x)$ ,  $i = 1, 2, \dots, a$  and  $t_i = s_i(y)$ ,  $i = 1, 2, \dots, b$ . Here is a schematic picture:



Finally let us write  $\Delta = e_k - d_k$ ,  $d = d_{k+1} + d_{k+2} + \dots + d_a$  and  $e = e_{k+1} + e_{k+2} + \dots + e_b$ . Note that the distance of  $x$  and  $y$  in  $T$  equals  $e + d + \Delta$ .

We now write

$$\|f(x) - f(y)\|_p^p = S + D + A + B,$$

where

$$\begin{aligned} S &= \sum_{i=1}^{k-1} \left| d_i^{1/p} s_i^{1/q} - e_i^{1/p} t_i^{1/q} \right|^p = \sum_{i=1}^{k-1} d_i \left| s_i^{1/q} - t_i^{1/q} \right|^p, \\ D &= \left| d_k^{1/p} s_k^{1/q} - e_k^{1/p} t_k^{1/q} \right|^p, \\ A &= \sum_{i=k+1}^a \left| d_i^{1/p} s_i^{1/q} \right|^p = \sum_{i=k+1}^a d_i s_i^{p-1}, \text{ and} \\ B &= \sum_{i=k+1}^b \left| e_i^{1/p} t_i^{1/q} \right|^p = \sum_{i=k+1}^b e_i t_i^{p-1}. \end{aligned}$$

We will estimate these terms one by one. First we show that

$$\frac{d^p}{4^p} \leq A \leq d^p$$

(and, symmetrically, it follows that  $e^p/4^p \leq B \leq e^p$ ). The upper bound is obtained easily from the inequality  $s_i \leq d$ . As for the lower bound, we first note that (2) gives

$$\begin{aligned} s_i &\geq d_i + (d_{i+1} + \dots + d_a) - \frac{d_i}{2} \geq \frac{d_i + d_{i+1} + \dots + d_a}{2} = \\ &\geq \frac{d - d_{k+1} - d_{k+2} - \dots - d_{i-1}}{2}. \end{aligned} \tag{3}$$

Let us choose an index  $i \geq k+1$  such that  $d_{k+1} + d_{k+2} + \dots + d_i \geq \frac{d}{2}$  and at the same time  $d_{k+1} + d_{k+2} + \dots + d_{i-1} < \frac{d}{2}$ . Then we have

$$\begin{aligned} A &\geq \sum_{j=k+1}^i d_j \left( \frac{d - d_{k+1} - d_{k+2} - \dots - d_{j-1}}{2} \right)^{p-1} \geq \\ &(d_{k+1} + \dots + d_i) \left( \frac{d - d_{k+1} - d_{k+2} - \dots - d_{i-1}}{2} \right)^{p-1} \geq \frac{d}{2} \left( \frac{d}{4} \right)^{p-1} \geq \frac{d^p}{4^p}. \end{aligned}$$

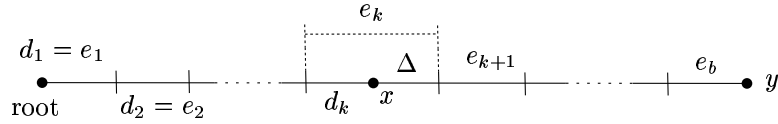
Next, we show that  $\|f^{-1}\|_{Lip}$  is upper-bounded by a constant. If  $\Delta \leq 2(d+e)$ , we have  $\|f(x) - f(y)\|_p \geq (A+B)^{1/p} \geq (d+e)/4 = \Omega(d+e+\Delta) = \Omega(\rho(x,y))$ . On the other hand, for  $\Delta > 2(d+e)$ , we have

$$t_k - s_k \geq e_k - d_k - |d - e| - \frac{d+e}{2} \geq \Delta - \frac{\Delta}{2} - \frac{\Delta}{4} = \frac{\Delta}{4}.$$

Consequently, using Lemma 9, we get

$$\begin{aligned} D^{1/p} &\geq (d_k + \Delta)^{1/p} (s_k + \frac{\Delta}{4})^{1/q} - d_k^{1/p} s_k^{1/q} \geq \\ &(d_k + \frac{\Delta}{4})^{1/p} (s_k + \frac{\Delta}{4})^{1/q} - d_k^{1/p} s_k^{1/q} \geq \frac{\Delta}{4} = \Omega(\rho(x,y)). \end{aligned}$$

It remains to bound  $\|f\|_{Lip}$ , and to this end, we still need to estimate the terms  $D$  and  $S$  from above. Here we can simplify the situation a little by assuming that  $x$  lies on the path from the root to  $y$ . Indeed, if it were not the case, let us consider  $z$ , the last vertex common to the paths from the root to the vertices  $x$  and  $y$ . Then  $\rho(x,y) = \rho(x,z) + \rho(z,y)$ , and we can use the upper bounds for  $\|f(x) - f(z)\|_p$  and  $\|f(z) - f(y)\|_p$  to estimate  $\|f(x) - f(y)\|_p$ . The considered situation can thus be illustrated as follows:



For the  $i$ th term of the sum  $S$ ,  $1 \leq i < k$ , Lemma 7 yields the estimate

$$d_i \left| t_i^{1/q} - s_i^{1/q} \right|^p \leq d_i \frac{|t_i - s_i|^p}{t_i}.$$

Since  $e_1 = d_1$ ,  $e_2 = d_2, \dots$ ,  $e_{k-1} = d_{k-1}$ , from (2) we obtain

$$t_i - s_i = \sum_{j=k}^b \max \left( 0, e_j - \frac{d_i}{2m} \right) - \max \left( 0, d_k - \frac{d_i}{2m} \right). \quad (4)$$

Let us put  $\delta = e_k + e_{k+1} + \dots + e_b$ , and

$$I = \{i \in \{1, 2, \dots, k-1\} : \frac{d_i}{2m} \leq \delta\}.$$

If  $i \notin I$ , we get  $d_k - \frac{d_i}{2m} < 0$  and  $e_j - \frac{d_i}{2m} < 0$  for all  $j = k, k+1, \dots, b$ , and from (4) we see that  $s_i - t_i = 0$ . For  $i \in I$ , we use the estimate  $t_i - s_i \leq \Delta + e_{k+1} + e_{k+2} + \dots + e_b = \rho(x,y)$ . At the same time, as in (3), we have  $t_i \geq (e_i + e_{i+1} + \dots + e_b)/2 = (e_i + e_{i+1} + \dots + e_{k-1} + \delta)/2$ . So we get (recall that  $e_i = d_i$  for  $i < k$ ),

$$\begin{aligned} S &= \sum_{i=1}^{k-1} e_i \left| s_i^{1/q} - t_i^{1/q} \right|^p \leq \sum_{i=1}^{k-1} e_i \frac{|t_i - s_i|^p}{t_i} \leq \\ &O(\rho(x,y)^p) \cdot \sum_{i \in I} \frac{e_i}{e_i + e_{i+1} + \dots + e_{k-1} + \delta} \leq \sum_{i \in I} \frac{e_i}{\delta + \sum_{j \in I} e_j}. \end{aligned}$$

We apply Lemma 8, where the role of the  $x_i$ 's is played by the numbers  $\frac{e_j}{\delta}$  with  $j \in I$ . From this, we get the bound  $S = \rho(x,y)^p O(\log m)$ .

Let us finally consider the term  $D$ . In our situation, we have  $d_k \leq e_k$ ,  $s_k = d_k$  and  $t_k = e_k + \sigma$ , where  $\sigma \leq e_{k+1} + e_{k+2} + \dots + e_b \leq \rho(x, y)$ . We arrive at

$$\begin{aligned} D^{1/p} &= \left| d_k - e_k^{1/p} (e_k + \sigma)^{1/q} \right| = \left| d_k - e_k \left( 1 + \frac{\sigma}{e_k} \right)^{1/q} \right| \leq \\ &= |d_k - e_k| + e_k \left[ \left( 1 + \frac{\sigma}{e_k} \right)^{1/q} - 1 \right] \leq \\ &= \Delta + e_k \left[ 1 + \frac{\sigma}{e_k} - 1 \right] = \Delta + \sigma = O(\rho(x, y)). \end{aligned}$$

Together with the previous estimates for  $B$  and  $S$ , this implies that

$$\|f(x) - f(y)\|_p \leq O(\log^{1/p} m) \cdot \rho(x, y).$$

Theorem 3 is proved.  $\square$

## 4 Concluding remarks

Our results determine tight worst-case asymptotic bounds for the distortion needed to embed an  $n$ -point tree metric space into  $\ell_p$ . For the probably most significant case  $p = 2$ , there is also a polynomial-time algorithm that finds an embedding into  $\ell_2$  with the smallest possible distortion for a given  $n$ -point metric space [LLR95] (I am aware of no result for the analogous question with  $p \neq 2$ ).

Much less is known if we restrict the dimension of the target space, that is, if we ask for the minimum  $D = D(n, d)$  such that all  $n$ -point tree metric spaces  $T$  can be  $D$ -embedded into  $\ell_2^d$  (or into  $\ell_p^d$ ). This setting may be quite interesting for practical applications. For an arbitrary  $n$ -point metric space  $M$ , I have shown [Mat90] that the required distortion for embedding into  $\ell_2^d$ , with  $d$  fixed, is at least  $\Omega(n^{1/\lfloor (d+1)/2 \rfloor})$  and at most  $O(n^{2/d} \log^{3/2} n)$  (with an improvement to  $O(n)$  for  $d = 1, 2$ ). There still remain significant gaps between the lower and upper bounds for odd dimensions  $d \geq 3$ . For instance, for  $d = 3$ , the lower bound is  $\Omega(n^{1/2})$  but the upper bound is only about  $n^{2/3}$ .

For tree metric spaces, it seems that only the  $\Omega(n^{1/d})$  lower bound is available, coming from an easy volume argument (consider embedding of the star with  $n - 1$  leaves). It seems plausible that this bound could be close to the truth for tree metric spaces. Also, it would be interesting to investigate the algorithmic complexity of testing  $D$ -embeddability into  $\ell_2^d$  for tree metric spaces and/or for arbitrary metric spaces.

## References

- [AR92] J. Arias-de-Reyna and L. Rodríguez-Piazza: Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces, *Israel J. Math.* 79(1992) 103–113.
- [Bou85] J. Bourgain: On Lipschitz embedding of finite metric spaces in Hilbert space, *Israel J. Math.* 52(1985) 46–52.
- [Bou86] J. Bourgain: The metrical interpretation of superreflexivity in Banach spaces, *Israel J. Math.* 56,2(1986) 222–230.

- [BMW86] J. Bourgain, V. Milman, H. Wolfson: On type of metric spaces, *Trans. Amer. Math. Soc.* 294(1986), 295–317.
- [Dre84] A. W. M. Dress: Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces, *Adv. Math.* 53(1984) 321–402.
- [Enf69a] P. Enflo: On a problem of Smirnov, *Ark. Mat.* 8(1969), 107–109.
- [Enf69b] P. Enflo: On the nonexistence of uniform homeomorphisms between  $L_p$ -spaces, *Ark. Mat.* 8(1969), 103–105.
- [JL84] W. Johnson and J. Lindenstrauss: Extensions of Lipschitz maps into a Hilbert space, *Contemporary Math.* 26 (Conference in Modern Analysis and Probability) pages 189–206, Amer. Math. Soc. 1984.
- [JLS87] W. Johnson, J. Lindenstrauss, G. Schechtman: On Lipschitz embedding of finite metric spaces in low dimensional normed spaces, in: *Geometrical aspects of functional analysis* (J. Lindenstrauss, V. D. Milman eds.), Lecture Notes in Mathematics 1267, Springer-Verlag, Berlin-Heidelberg 1987.
- [LLR95] N. Linial, E. London and Y. Rabinovich: The geometry of graphs and some its algorithmic applications, *Combinatorica* 15(1995) 215–245.
- [Mat89] J. Matoušek: *Lipschitz distance of metric spaces* (in Czech, with English summary), CSc. degree thesis, Charles University 1990.
- [Mat90] J. Matoušek: Bi-Lipschitz embeddings into low-dimensional Euclidean spaces, *Comment. Math. Univ. Carolinae* 31,3(1990) 589–600.
- [Mat96] J. Matoušek: On the distortion required for embedding finite metric spaces into normed spaces, *Israel J. Math.* 93(1996) 333–344.
- [MS86] V. D. Milman and G. Schechtman: *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics 1200, Springer-Verlag, 1986.