Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
European Journal of Combinatorics 25 (2004) 87–92

European Journal of Combinatorics

www.elsevier.com/locate/ejc

Low dimensional embeddings of ultrametrics

Yair Bartal[a], Nathan Linial[a], Manor Mendel[a], Assaf Naor[b]

[a] *School of Computer Science, Hebrew University, Jerusalem 91904, Israel*
[b] *Theory Group, Microsoft Research, One Microsoft Way 113/2131, Redmond, WA 98052-6399, USA*

Received 22 May 2003; received in revised form 29 August 2003; accepted 29 August 2003

Abstract

In this note we show that every n-point ultrametric embeds with constant distortion in $\ell_p^{O(\log n)}$ for every $\infty \geq p \geq 1$. More precisely, we consider a special type of ultrametric with hierarchical structure called a k-hierarchically well-separated tree (k-HST). We show that any k-HST can be embedded with distortion at most $1 + O(1/k)$ in $\ell_p^{O(k^2 \log n)}$. These facts have implications to embeddings of finite metric spaces in low dimensional ℓ_p spaces in the context of metric Ramsey-type theorems.

Keywords: Metric embeddings; Ultrametrics

---

1. Introduction

An *ultrametric* is a metric space (X, d) such that for every $x, y, z \in X$,

$$d(x, z) \leq \max\{d(x, y), d(y, z)\}.$$

Finite ultrametrics have a natural hierarchical description called *dendrogram* (see [1] and references therein). A more restricted class of metrics with potentially stronger hierarchical structure is that of k*-hierarchically well-separated trees*, defined as follows:

Definition 1 ([2]). For $k \geq 1$, a k*-hierarchically well-separated tree* (k-HST) is a metric space whose elements are the leaves of a rooted finite tree T. To each vertex $u \in T$ there is associated a label $\Delta(u) \geq 0$ such that $\Delta(u) = 0$ iff u is a leaf of T. It is required that if a vertex u is a child of a vertex v then $\Delta(u) \leq \Delta(v)/k$. The distance between two leaves $x, y \in T$ is defined as $\Delta(\text{lca}(x, y))$, where $\text{lca}(x, y)$ is the least common ancestor of x and y in T.

E-mail addresses: yair@cs.huji.ac.il (Y. Bartal), nati@cs.huji.ac.il (N. Linial), mendelma@yahoo.com (M. Mendel), anaor@microsoft.com (A. Naor).

0195-6698/$ - see front matter
doi:10.1016/j.ejc.2003.08.003

The notion of 1-HST coincides with that of an ultrametric. Any k-HST is also a 1-HST, i.e., an ultrametric. However, for every $k > 1$ the class of k-HST is a proper subclass of ultrametrics. Ultrametrics and k-HSTs have played a key role in recent work on embeddings of finite metric spaces [3–6].

Let $f : X \to Y$ be an embedding of the metric space (X, d_X) into the metric space (Y, d_Y). We define the *distortion* of f by

$$\operatorname{dist}(f) = \sup_{\substack{x,y\in X\\ x\neq y}} \frac{d_Y(f(x), f(y))}{d_X(x,y)} \cdot \sup_{\substack{x,y\in X\\ x\neq y}} \frac{d_X(x,y)}{d_Y(f(x), f(y))}.$$

We denote by $c_Y(X)$ the least distortion with which X may be embedded in Y. When $c_Y(X) \le \alpha$ we say that X α-embeds into Y. When there is a bijection f between two metric spaces X and Y with $\operatorname{dist}(f) \le \alpha$ we say that X and Y are α-similar.

The following proposition provides a comparison between ultrametrics and k-HSTs.

Proposition 1 ([3]). *For any $k > 1$, any ultrametric is k-similar to a k-HST.*

A basic folklore property of ultrametrics (cf. [7]) is:

Proposition 2. *Any finite ultrametric is isometrically embeddable in ℓ_2.*

Since any finite subset of ℓ_2 isometrically embeds into ℓ_p for every $1 \le p \le \infty$, a similar result follows for embeddings in ℓ_p. Moreover, a careful analysis of the proof of the above proposition yields an isometric embedding of any n-point HST into $\ell_p^{O(n)}$.

Here we consider the *dimension* for which ultrametrics and k-HST spaces embed with a given distortion in ℓ_p, $1 \le p \le \infty$. For ℓ_2 this is answered by the Johnson–Lindenstrauss dimension reduction lemma [8] which states that for every $\epsilon > 0$, any n-point metric space in ℓ_2 can be $(1+\epsilon)$-embedded in $\ell_2^{O(\log n/\epsilon^2)}$. Using [9], it follows that any set of n points in ℓ_2 can be embedded with constant distortion into $\ell_p^{O(\log n)}$ for $1 \le p \le 2$ and into $\ell_p^{O(\sqrt{p}(\log n)^{p/2})}$ for $p > 2$. The main result of this note improves the upper bound on the dimension required to embed n-point ultrametrics into ℓ_p, $p > 2$, and gives additional structural information on the problem for embeddings into low dimensional ℓ_p spaces for $1 \le p \le 2$. Moreover, we show that any n-point k-HST can be embedded in ℓ_p with constant distortion and dimension logarithmic in n. Furthermore, the distortion approaches 1 as k grows.

Proposition 3. *Fix an integer $k > 5$. Then for any $1 \le p < \infty$, any k-HST can be $(k+1)/(k-5)$-embedded in ℓ_p^h with $h = \lceil C(1+k/p)^2 \log D\rceil$, where D is the maximal out-degree of a vertex in the tree defining the k-HST, and $C > 0$ is a universal constant.*

Proposition 3 is proved in Section 2. Combining Propositions 1 and 3 we obtain the following:

Corollary 4. *For any* $1 \le p \le \infty$, *any* n *point ultrametric can be* $O(1)$*-embedded into* $\ell_p^{O(\log n)}$.

We also show how to apply this lemma to the metric Ramsey-type problems. A metric Ramsey-type theorem states that a given metric space contains a large subset which can be embedded with small distortion in some "well-structured" family of metric spaces (e.g. Euclidean metrics). This can be formulated using the following notion.

Definition 2. Let $\mathcal{M}$ be some class of metric spaces. Denote by $R_{\mathcal{M}}(\alpha, n)$ the largest integer m such that any n-point metric space has a subset of size m that α-embeds into a member of $\mathcal{M}$. When $\mathcal{M} = \{\ell_p\}$, we use R_p rather than R_{ℓ_p}.

In [5] it is shown that for every $1 \le p \le \infty$ and $\alpha > 2$, $R_p(\alpha, n) \ge n^{1-O\left(\frac{\log \alpha}{\alpha}\right)}$ and for every $0 < \epsilon < 1$, $R_p(2+\epsilon, n) \ge n^{\Omega\left(\frac{\epsilon}{\log(2/\epsilon)}\right)}$. We refer to [5] and the references therein for a comprehensive description of metric Ramsey problems and their history. Using Proposition 3, we prove the following variant of the result of [5] in which there is control on the dimension in the metric Ramsey problem for ℓ_p, $p \ge 1$. This application was our original motivation for studying low-dimensional embeddings of ultrametrics.

Theorem 1. *The following assertions hold:*

(1) *There exist absolute constants* $c, C > 0$ *such that for all* $1 \le p \le \infty$ *and for every* $\alpha > 2$,

$$R_{\ell_p^d}(\alpha, n) \ge n^{1-C\frac{\log\alpha}{\alpha}}, \qquad \text{where } d = \lceil c \log n \rceil.$$

(2) *There are absolute constants* $C, c > 0$ *such that for every* $0 < \epsilon < 1$, $1 \le p < \infty$ *and every integer* n,

$$R_{\ell_p^d}(2+\epsilon, n) \ge n^{\frac{c\epsilon}{\log(2/\epsilon)}}, \qquad \text{where } d = \left\lceil C \frac{\epsilon \lceil (\epsilon p)^{-2} \rceil}{\log(2/\epsilon)} \log n \right\rceil.$$

2. Embedding HSTs in low dimensional ℓ_p spaces

We follow Definition 1, and associate with any k-HST, the tree T *defining the HST*. An internal vertex in T with out-degree 1 is said to be *degenerate*. If u is non-degenerate, then $\Delta(u)$ is the diameter of the sub-space induced on the subtree rooted by u. Degenerate nodes do not influence the metric on T's leaves, hence we may assume that all internal nodes are non-degenerate. In particular for an HST X, $\operatorname{diam}(X) = \Delta(\operatorname{root}(T))$, where T is the tree defining X.

We make use of the following standard construction of codes, the proof of which is included for the sake of completeness. In what follows, for $w, v \in \{0,1\}^h$, $w \oplus v$ denotes the point-wise addition modulo 2 of v and w.

Lemma 5. *For any* $h \in \mathbb{N}$, *and* $\tau \in (0,1)$, *there exists* $K \subset \{0,1\}^h$ *such that the Hamming distance between any two distinct elements of* K *is in the range* $[(1-\tau)h/2, (1+\tau)h/2]$ *and* $|K| \geq \lfloor e^{h\tau^2/8} \rfloor$.

Proof. Let $w, v \in \{0,1\}^h$ be independent and equidistributed random elements. Then by the Chernoff bound, the probability that $w \oplus v$ has less than $(1-\delta)h/2$ 1's is at most $e^{-\delta^2 h/4}$. Similarly, the probability it has more than $(1+\delta)h/2$ 1's is also at most $e^{-\delta^2 h/4}$. Given m random elements $w_1, \ldots, w_m \in \{0,1\}^h$, the probability that the distance between any two of them isn't in the range $[(1-\delta)h/2, (1+\delta)h/2]$ is at most $\binom{m}{2} 2e^{-\delta^2 h/4} < m^2 e^{-\delta^2 h/4}$. Thus, choosing $m = \lfloor e^{\delta^2 h/8} \rfloor$ implies that with a positive probability the subset $K = \{w_1, \ldots, w_m\}$ has the required properties. □

Proof of Proposition 3. Let u be the root of the tree defining X and $X_1, \ldots, X_s$ be the leaf sets of subtrees rooted at the children of u. Note that $s \leq D$. For $p < \infty$, let $\tau = (1 + k/p)^{-1}/6$. Set $h = \lceil 8\tau^{-2} \log D \rceil$, so that $e^{h\tau^2/8} \geq s$. By Lemma 5 there exists $K \subset \{0,1\}^h$ with all Hamming distances in the range $[(1-\tau)h/2, (1+\tau)h/2]$ and $|K| \geq s$. Choose s distinct $c_1, \ldots, c_s \in K$. By switching to $c_1 \oplus c_1, c_2 \oplus c_1, \ldots, c_s \oplus c_1$ we may assume that $c_1 = 0$, in which case for $1 \leq i \leq s$, $\|c_i\|_1 \leq \frac{1+\tau}{2} h$.

Assume inductively that for each i we have an embedding $\phi_i : X_i \to \ell_p^h$, such that:

- For all $x, y \in X_i$, $\frac{k-5}{k+1} d_{X_i}(x,y) \leq \|\phi_i(x) - \phi_i(y)\|_p \leq d_{X_i}(x,y)$.
- For every $x \in X_i$, $\|\phi_i(x)\|_p \leq \operatorname{diam}(X_i)$.

Let $\lambda = (\frac{1+\tau}{2} h)^{-1/p} \frac{k-2}{k}$, and let $\Delta = \operatorname{diam}(X)$. Define an embedding $\phi : X \to \ell_p^h$ of X as follows: for $x \in X_i$,

$$\phi(x) = \phi_i(x) + \lambda \Delta c_i.$$

Then

$$\begin{aligned} \|\phi(x)\|_p &\leq \|\phi_i(x)\|_p + \lambda\Delta\|c_i\|_p \leq \operatorname{diam}(X_i) + \left(\frac{1+\tau}{2}h\right)^{-1/p} \frac{k-2}{k} \Delta \|c_i\|_1^{1/p} \\ &\leq \frac{\Delta}{k} + \frac{k-2}{k}\Delta < \Delta. \end{aligned}$$

For $x, y \in X_i$, $\|\phi(x) - \phi(y)\|_p = \|\phi_i(x) - \phi_i(y)\|_p$, so by the induction hypothesis

$$\frac{k-5}{k+1} d_X(x,y) \leq \|\phi(x) - \phi(y)\|_p \leq d_X(x,y).$$

For $x \in X_i$, $y \in X_j$ and $i \neq j$, we have $d_X(x,y) = \Delta$. Now

$$\begin{aligned} \|\phi(x) - \phi(y)\|_p &\leq \lambda\Delta\|c_i - c_j\|_p + \|\phi_i(x)\|_p + \|\phi_j(x)\|_p \\ &\leq \lambda\Delta\|c_i - c_j\|_1^{1/p} + \operatorname{diam}(X_i) + \operatorname{diam}(X_j) \\ &\leq \left(\frac{1+\tau}{2}h\right)^{-1/p} \frac{k-2}{k}\Delta \left(\frac{1+\tau}{2}h\right)^{1/p} \\ &\quad + \frac{2}{k}\Delta = \Delta = d_X(x,y), \end{aligned}$$

and

$$\begin{aligned}\|\phi(x)-\phi(y)\|_p &\geq \lambda\Delta\|c_i-c_j\|_p - \|\phi_i(x)\|_p - \|\phi_j(x)\|_p\\ &\geq \lambda\Delta\|c_i-c_j\|_1^{1/p} - \mathrm{diam}(X_i) - \mathrm{diam}(X_j)\\ &\geq \left(\frac{1+\tau}{2}h\right)^{-1/p}\frac{k-2}{k}\Delta\left(\frac{1-\tau}{2}h\right)^{1/p} - \frac{2}{k}\Delta\\ &\geq \left(\left(\frac{1-\tau}{1+\tau}\right)^{1/p}\frac{k-2}{k} - \frac{2}{k}\right)\Delta\\ &\geq \left(\frac{k}{k+1}\cdot\frac{k-2}{k} - \frac{2}{k}\right)\Delta \geq \frac{k-5}{k+1}d_X(x,y).\end{aligned}$$

The last inequality holds for $k > 5$ and the preceding derivation follows from the definition of τ:

$$\begin{aligned}\left(\frac{1-\tau}{1+\tau}\right)^{1/p} &\geq (1+3\tau)^{-1/p} \geq (1+6\tau/p)^{-1}\\ &= (1+(1+k/p)^{-1}/p)^{-1}\\ &\geq (1+1/k)^{-1}. \quad \square\end{aligned}$$

3. Implications

Denote by UM the class of all ultrametrics. We will need the following theorem:

Theorem 2 ([5]). *The following assertions hold for every integer n:*

(1) *There exists an absolute constant $C' > 0$ such that for every $\alpha > 2$,*

$$R_{\mathrm{UM}}(\alpha, n) \geq n^{1-C'\frac{\log\alpha}{\alpha}}.$$

(2) *There is an absolute constant $c > 0$ such that for any $k \geq 1$ and $0 < \epsilon < 1$, for any integer n*

$$R_{k-\mathrm{HST}}(2+\epsilon, n) \geq n^{\frac{c\epsilon}{\log(2k/\epsilon)}}.$$

Proposition 2 implies similar bounds for $R_2(\alpha, n)$. We next show how to extend those results for embedding into $\ell_p^{O(\log n)}$ by using Proposition 3.

Proof of Theorem 1. We begin with the first claim of the theorem. Let $C' > 0$ be the constant at the first assertion in Theorem 2, and let β be a universal constant such that any n-point ultrametric β embeds in $\ell_p^{O(\log n)}$ (Corollary 4). We choose $C = \beta C'$, so that $C\frac{\log\alpha}{\alpha} \geq C'\frac{\log(\alpha/\beta)}{\alpha/\beta}$. From Theorem 2 we deduce that

$$R_{\mathrm{UM}}(\alpha/\beta, n) \geq n^{1-C'\frac{\log(\alpha/\beta)}{\alpha/\beta}} \geq n^{1-C\frac{\log\alpha}{\alpha}}.$$

The subset described by this statement is (α/β)-similar to an ultrametric and so, by Corollary 4, it is α-embeddable in $\ell_p^{O(\log n)}$.

We next consider the second statement in the theorem. Let $\delta = \epsilon/4$ and $k = \lfloor 5 + 6/\delta \rfloor$, then by Theorem 2, there exists $c' > 0$ such that $R_{k-\mathrm{HST}}(2+\delta, n) \geq n^{\frac{c'\delta}{\log(2/\delta)}}$. Let M be an arbitrary metric space. For an appropriate choice of c this means that M contains a subset Y of size $m = \lceil n^{\frac{c\epsilon}{\log(2/\epsilon)}} \rceil$ that is $(2+\delta)$-similar to some k-HST X. By Proposition 3 and our choice of k, there exists some constant $C' > 0$ such that X can be $(1+\delta)$-embedded in ℓ_p^d, where

$$d = \lceil C' \lceil (\delta p)^{-2} \rceil \log m \rceil = \left\lceil C \frac{\epsilon \lceil (\epsilon p)^{-2} \rceil}{\log(2/\epsilon)} \log n \right\rceil,$$

for an appropriate choice of C. Therefore Y is $(2+\delta)(1+\delta) \leq (2+\epsilon)$-embedded in ℓ_p^d. □

Acknowledgements

Y. Bartal and N. Linial are supported in part by a grant from the Israeli National Science Foundation. M. Mendel is supported in part by the Landau Center.

References

[1] A.D. Gordon, A review of hierarchical classification, J. Roy. Statist. Soc. Ser. A 150 (2) (1987) 119–137.

[2] Y. Bartal, Probabilistic approximation of metric spaces and its algorithmic applications, in: 37th Annual Symposium on Foundations of Computer Science, Burlington, Vermont, IEEE Computer Society, 1996, pp. 184–193.

[3] Y. Bartal, On approximating arbitrary metrics by tree metrics, in: Proceedings of the 30th Annual ACM Symposium on Theory of Computing, Dallas, TX, ACM, 1998, pp. 183–193.

[4] Y. Bartal, B. Bollobás, M. Mendel, A Ramsey-type theorem for metric spaces and its applications for metrical task systems and related problems, in: 42nd IEEE Symposium on Foundations of Computer Science, Las Vegas, NV, 2001, IEEE Computer Soc., Los Alamitos, CA, 2001, pp. 396–405.

[5] Y. Bartal, N. Linial, M. Mendel, A. Naor, On metric Ramsey-type phenomena, in: Proceedings of the 35th Annual ACM Symposium on Theory of Computing, San-Diego, CA, ACM, 2003, pp. 463–472.

[6] J. Fakcharoenphol, S. Rao, K. Talwar, A tight bound on approximating arbitrary metrics by tree metrics, in: Proceedings of the 35th Annual ACM Symposium on Theory of Computing, San-Diego, CA, ACM, 2003, pp. 448–455.

[7] A.J. Lemin, Isometric embedding of ultrametric (non-Archimedean) spaces in Hilbert space and Lebesgue space, in: p-Adic Functional Analysis, Ioannina, 2000, Lecture Notes in Pure and Applied Mathematics, vol. 222. Dekker, New York, 2001, pp. 203–218.

[8] W.B. Johnson, J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, in: Conference in Modern Analysis and Probability, New Haven, Conn., 1982, American Mathematical Society, Providence, RI, 1984, pp. 189–206.

[9] T. Figiel, J. Lindenstrauss, V.D. Milman, The dimension of almost spherical sections of convex bodies, Acta Math. 139 (1–2) (1977) 53–94.