Embedding Ultrametrics Into Low-Dimensional Spaces

Mihai Bădoiu
CSAIL MIT
mihai@mit.edu

Julia Chuzhoy
CSAIL MIT and
Dept. of CIS
U. of Pennsylvania
cjulia@csail.mit.edu

Piotr Indyk
CSAIL MIT
indyk@mit.edu

Anastasios Sidiropoulos
CSAIL MIT
tasos@mit.edu

ABSTRACT

We study the problem of minimum-distortion embedding of ultrametrics into the plane and higher dimensional spaces. Ultrametrics are a natural class of metrics that frequently occur in applications involving hierarchical clustering. Low-distortion embeddings of ultrametrics into the plane help visualizing complex structures they often represent.

Given an ultrametric, a natural question is whether we can efficiently find an optimal-distortion embedding of this ultrametric into the plane, and if not, whether we can design an efficient algorithm that produces embeddings with near-optimal distortion. We show that the problem of finding minimum-distortion embedding of ultrametrics into the plane is NP-hard, and thus approximation algorithms are called for. Given an input ultrametric M, let c denote the minimum distortion achievable by any embedding of M into the plane. Our main result is a linear-time algorithm that produces an $O(c^3)$-distortion embedding. This result can be generalized to embedding ultrametrics into $\Re^d$, for any $d \geq 2$, with distortion $c^{O(d)}$, where c is the minimum distortion achievable for embedding the input ultrametric into $\Re^d$.

Additionally, we show that *any* ultrametric can be embedded into the plane with distortion $O(\sqrt{n})$, and in general, into $\Re^d$ with distortion $d^{O(1)}n^{1/d}$. Combining the two results together, we obtain an $O(n^{1/3})$-approximation algorithm for the problem of minimum-distortion embedding of ultrametrics into the plane.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Non-numerical Algorithms and Problems

General Terms

Algorithms

Keywords

Ultrametrics, Embedding, Approximation Algorithms.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SCG'06, June 5–7, 2006, Sedona, Arizona, USA.
Copyright 2006 ACM 1-59593-340-9/06/0006 ...$5.00.

1. INTRODUCTION

Given two metrics spaces $M = (X, D)$ and $M' = (X', D')$, a non-contracting embedding of M into M' is a mapping $f : X \to X'$, such that for any pair $p, q \in X$, $D(p, q) \leq D'(f(p), f(q))$. Given an embedding $f : X \to X'$, we say that the distortion of f is c, if f is non-contracting, and c is the maximum, over all pairs of points $p, q \in X$, of $D'(f(p), f(q))/D(p, q)$. Low-distortion embeddings have been a subject of extensive mathematical studies, and found numerous applications in computer science (cf. [17, 14]).

Most of the research on embeddings focused on showing *absolute* results, of the form:

> Given a class of metrics $\mathcal{C}$ and a metric M', what is the *smallest distortion* $c \geq 1$ such that any metric $M \in \mathcal{C}$ can be embedded into M' with distortion c ?

However, in many interesting cases, the worst-case distortion over *all* metrics in $\mathcal{C}$ is too large to be interesting or meaningful. This is especially the case for embeddings into low-dimensional spaces, where the worst-case distortion is polynomial in the metric size, even for very simple metrics such as an n-point star. However, in cases where the input metric does have a good embedding into the host metric, it would be useful to have algorithms that produce such embeddings.

Our paper focuses on the *relative* version of the problem, which involves finding embeddings whose distortion is close to optimal. More precisely, assume we are given a fixed class of metrics $\mathcal{C}$ and a host metric M'. We are interested in designing an efficient algorithm, which receives as input a metric $M \in \mathcal{C}$, and produces an embedding of M into M' with distortion at most αc, where c denotes the best possible distortion of an embedding of M into M', and $\alpha \geq 1$ is called the *approximation factor* of the algorithm. We note that α might depend on c. Figure 1 summarizes the previously known results on relative embeddings (see also [10]).

In this paper, we explore relative embeddings into the *plane* (and higher-dimensional spaces). As can be seen from Figure 1, very little is known about this problem. The only relevant result is a 3-approximation algorithm for embedding an n-point subset of a two-dimensional sphere (living in $\Re^3$) into the plane. More specifically, we consider embedding *ultrametrics* into the plane. Ultrametrics are a natural class of metrics, frequently occurring in applications involving hierarchical clustering. They are of particular interest in biology, where they can be used to represent evolutionary trees (cf. [11] or [8], p. 168). Visualizing such trees requires embedding them into the plane, which is exactly the problem we consider in this paper.

Paper	From	Into	Distortion	Comments
[18]	general metrics	l_2	c	uses SDP
[15]	line	line	c	c is constant, embedding is a bijection
	unweighted graphs	bounded degree trees	c	c is constant, embedding is a bijection
[20]	$\Re^3$	$\Re^3$	$> (3-\epsilon)c$	hard to 3-approximate, embedding is a bijection
[9]	unweighted graphs	sub-trees	$O(c \log n)$	
[7]	unweighted graphs	trees	$O(c)$	
[6]	unweighted graphs	line	$O(c^2)$	implies $\sqrt{n}$-approximation
			$> ac$	Hard to a-approximate for some $a > 1$
			c	c is constant
	unweighted trees	line	$O(c^{3/2}\sqrt{\log c})$	
	subsets of a sphere	plane	$3c$	
[1]	general metrics	ultrametrics	c	
[5]	general metrics	line	$O(\Delta^{3/4}c^{11/4})$	
	weighted trees	line	$c^{O(1)}$	
	weighted trees	line	$\Omega(n^{1/12}c)$	Hard to $O(n^{1/12})$-approximate even for $\Delta = n^{O(1)}$

Figure 1: Previous work on relative embedding problems for multiplicative distortion. We use c to denote the optimal distortion, n to denote the number of points in the input metric and Δ to denote the spread of the metric, i.e., the maximum to minimum distance ratio. Note that the table contains only the results that hold for the *multiplicative* definition of the distortion; there is a rich body of work that applies to other definitions of distortion, notably the *additive* or *average* distortion.

Our Results

Our main result is an algorithm which receives as input an ultrametric and outputs its embedding into the plane. If the input ultrametric embeds into the plane with distortion c (under l_p norm for any $1 \leq p \leq \infty$[1]), then the embedding produced by the algorithm has distortion $O(c^3)$. In particular, for the case where the input ultrametric is embeddable into the plane with constant distortion, the distortion of the embedding produced by the algorithm is also constant. The running time of our algorithm is linear in the input size, assuming it is given the value of the optimum distortion c (or its approximation). The algorithm generalizes to embeddings into $\Re^d$ (equipped with the l_2 norm), and the distortion becomes $c^{O(d)}$, where c is the distortion of the optimal embedding of the ultrametric into $\Re^d$.

In our second result we prove that any ultrametric can be embedded into the plane with distortion $O(\sqrt{n})$. More generally, for any $d \geq 2$, we show how to embed any ultrametric into $\Re^d$ with distortion $d^{O(1)}n^{1/d}$. Notice that unlike the first result, this result relates to the absolute version of the distortion minimization problem. The proof is algorithmic - the embedding can be found in polynomial time. Combining the two results together, we obtain an $O(n^{1/3})$-approximation algorithm for embedding ultrametrics into the plane.

Figure 2 summarizes the known results for the *absolute* version of low-distortion embeddings into low-dimensional Euclidean space. Note that the only previous class of *weighted* metrics for which a non-trivial embedding into the plane was known was the class of weighted stars. Our theorem strictly generalizes that result, since any n-point weighted star can be embedded into an ultrametric of size $O(n)$ with constant distortion (an easy proof of this fact is provided at the end of Section 5).

We also remark that for the case of embedding ultrametrics into low-dimensional spaces, it has been shown (cf. [4]) that for any $\epsilon > 0$, any ultrametric can be embedded into $\ell_p^{O(\epsilon^{-2}\log n)}$, with distortion $1+\epsilon$.

Finally, we investigate the hardness of embedding ultrametrics into the plane. We prove that the problem of finding the smallest-distortion embedding is strongly NP-hard, if the distance is measured according to the l_∞ norm. Interestingly, the problem of minimizing the distortion of embedding *into* ultrametrics can be solved exactly in polynomial time [1].

Our techniques

We use the well-known fact that any ultrametric $M = (X, D)$ can be well approximated by hierarchically well-separated trees (HST's) (see Section 2 for definitions). The corresponding HST T has the points of X as its leaves, and each vertex v of T has a label $l(v) \in \Re^+$. The distance of any pair of points $p, q \in X$ is exactly the label of their nearest common ancestor.

The hierarchical structure of HST's naturally enables constructing the embedding in a recursive manner. That is, the mapping is constructed by embedding (recursively and independently) the children of the root node, and then combining the embeddings. Implementing this idea, however, requires overcoming a few obstacles, which we discuss now. For simplicity, we focus on embeddings into the plane.

Distortion lower bound. The first issue is how to obtain a good lower bound for the distortion. It is not difficult to see that the distortion depends on both the number of nodes, and the structure of the ultrametric. For example, the full 2-HST of depth t, where all internal nodes have degree 4, requires $\Omega(t)$ distortion; at the same time, the full 4-HST of depth t, where all internal nodes have degree 4, can be embedded with constant distortion.

Our lower bound is obtained as follows. Consider any node v and its children $u_1 \dots u_k$. Let P_i be the set of leaves in the subtree of the node u_i, $P = P_1 \cup \dots \cup P_k$. By the definition of ultrametrics, the distances between any pair of points $p \in P_i$ and $q \in P_j$ for $i \neq j$, are equal to the same value, namely $l(v)$. Consider any non-contracting embedding $f : P \to \Re^2$. Construct a ball of radius $l(v)/2$ around each point $f(p)$, $p \in P$, and denote this ball by $B(p, l(v)/2)$. It is easy to see that the union of the interiors of the balls around points in P_i and the union of the interiors of the balls around points in P_j must be disjoint if $i \neq j$.

Our lower bound on distortion proceeds by estimating the total volume $C(v)$ of $\cup_{p\in P} B(p, l(v)/2)$. Specifically, by packing argument, one can observe that the distortion of the optimal embedding must be at least $\Omega(\sqrt{C(v)} - O(1))$. Thus, it suffices to have a good lower bound for the volume $C(v)$. It would appear that such

[1]The algorithm is described for the case of the l_2 norm. However, since l_p norms for all $1 \leq p \leq \infty$ in $\Re^2$ are equivalent up to a factor of 2, the algorithm works for any l_p norm as well.

Paper	From	Into	Upper Bound	Lower Bound
[19]	general metrics	l_2^d	$\tilde{O}(n^{2/d})$	$\Omega(n^{1/\lfloor(d+1)/2\rfloor})$
[12]	trees	l_2^d	$\tilde{O}(n^{1/(d-1)})$	$\Omega(n^{1/d})$
[13]	weighted stars	l_2^d	$O(n^{1/d})$	$\Omega(n^{1/d})$
[2]	unweighted trees	l_2^2	$O(n^{1/2})$	$\Omega(n^{1/2})$

Figure 2: Previous work on worst-case embeddings into small dimensional spaces (we are assuming $d = O(1)$).

lower bounds could be obtained by summing $C(u_i)$'s, since the balls around different sets P_i are disjoint. Unfortunately, $C(u_i)$ is the volume of the union of the balls of radius $l(u_i)/2$, not $l(v)/2$, so the above is not strictly true. However, $\cup_{p\in P_i} B(p, l(v)/2)$ can be expressed as a Minkowski sum of $\cup_{p\in P_i} B(p, l(u_i)/2)$ and a ball of radius $[l(v) - l(u_i)]/2$. Then the volume of that set can be bounded from below by using Brunn-Minkowski inequality, by a function of $C(u_i)$ and $l(v) - l(u_i)$. This enables us to obtain a recursive formula for $C(v)$ as a function of $C(u_i)$'s.

Distortion accumulation. The recursive formula for the lower bound suggests a recursive algorithm. Consider some vertex v of the HST, and let $u_1, \dots, u_k$ be its children. For each u_i, $1 \le i \le k$, the leaves in the subtree of u_i are mapped into a square $R(u_i)$ whose volume is at most $C(u_i)$. Then the squares are re-arranged to form a square $R(v)$. The main difficulty with this approach is that the optimal way to pack the squares is difficult to find. In fact, the optimal embedding could, in principle, not pack the points into squares. To overcome this problem, we allow some limited stretching of the squares, to fit them into $R(v)$. However, stretching causes distortion, and thus we need to make sure that stretching done over different levels does not accumulate. In order to avoid such accumulation of distortion, we alternate between the horizontal and vertical stretchings of the squares. Specifically, we assign, for each vertex v of the HST, a bit $g(v)$ that determines whether the squares into which the sub-trees of the children of v are embedded will be stretched in the horizontal or the vertical direction before they are packed into the square $R(v)$. We calculate the values of the bits $g(v)$ in a top-down manner, starting with the leaves of the HST, to ensure that the final stretchings are balanced.

It appears that the need to compute a proper choice of stretching directions (which can also be viewed as rotations) at each level is not just an artifact of our algorithm, but it might be necessary to achieve low distortion. In particular, the only constant distortion embedding of a full 2-HST into the plane that we are aware of uses alternating rotations.

Higher dimensions. We show how to generalize the algorithm for embedding ultrametrics into the plane to higher dimensions. We show an algorithm that produces a $c^{O(d)}$-distortion embedding of the input ultrametric into $\Re^d$ under the l_2 norm, where c denotes the optimal distortion achievable when embedding the input ultrametric into $\Re^d$.

Hardness. We show NP-hardness of the embedding problem for the case of the plane under l_∞ norm. We use a reduction from the square packing problem. Since our algorithm also uses (a variant of) square packing, the packing problem appears to be intimately related to embeddings of ultrametrics.

2. PRELIMINARIES AND DEFINITIONS

A metric $M = (X, D)$ is an *ultrametric*, if it can be represented by a *labeled tree* T whose set of leaves is X, in the following manner. Each non-leaf vertex v of T has a label $l(v) > 0$. If u is a child of v in tree T, then $l(u) \le l(v)$. For any $x, y \in X$, the distance between x and y is defined to be the label of the nearest common ancestor of x and y, and this distance should be equal to $D(x, y)$.

We now proceed to define hierarchically well-separated trees (HST's). For any $\alpha \ge 1$, an α-HST is an ultrametric where for each parent-child pair of vertices (u, v), $l(u) = \alpha l(v)$. It is easy to see that for any $\alpha \ge 1$, any ultrametric can be α-approximated by an α-HST (cf. [3]). Moreover, such an HST can be found in time linear in the input size. Therefore, if the input ultrametric M embeds into $\Re^d$ with distortion c, then the metric M' defined by the corresponding 2-HST embeds into $\Re^d$ with distortion $c' = 2c$. Any non-contracting embedding of M' into $\Re^d$ with distortion c'' represents a non-contracting embedding of M with distortion $O(c'')$. Therefore, from now on we will concentrate on embeddings of HST's into $\Re^d$.

Given a 2-HST T, we will use the following additional notation. Let r denote the root of the tree, and let h denote the tree height. We assume that r belongs to the first level of T, and all the leaves belong to level h. By scaling the underlying metric M, we can assume w.l.o.g., that for each vertex v at level $h-1$, $l(v) = 2$. For any non-leaf vertex v, we denote by X_v the set of leaves of the subtree of T rooted at v, and we denote the number of leaves in the subtree n_v.

We will use the Brunn-Minkowski inequality, defined as follows. Given any two sets $A, B \subseteq \mathbb{R}^d$, let $A \oplus B$ denote the Minkowski sum of A and B, i.e., $A \oplus B = \{a + b \mid a \in A, b \in B\}$.

THEOREM 1 (BRUNN-MINKOWSKI INEQUALITY). *For any pair of sets $A, B \subseteq \mathbb{R}^d$,*

$$Vol(A \oplus B)^{1/d} \ge Vol(A)^{1/d} + Vol(B)^{1/d}.$$

3. A LOWER BOUND ON THE DISTORTION OF OPTIMAL EMBEDDING

In this section we show a lower bound on the distortion of optimal embedding of a metric M' which is defined by a 2-HST denoted by T.

For any $r > 0$, let $B(r)$ denote the ball of radius r in ℓ_2^d centered at the origin. Let $V_d(r)$ denote the volume of a d-dimensional ball of radius r, $V_d(r) = \frac{\pi^{d/2} r^d}{\Gamma(1+d/2)}$. For each vertex v of T, we define a value $C(v)$, which intuitively is a lower-bound on the minimum volume embedding of X_v (the precise statement appears below). The values $C(v)$ are defined recursively, starting from the leaves. For each leaf v, we set $C(v) = V_d(1/2)$.

Consider now vertex v at level $j \in [h-1]$, and let $u_1, \dots, u_k$ be the children of v in T. We define:

$$C(v) = \sum_{i=1}^{k} \left((C(u_i))^{1/d} + (V_d(l(v)/4))^{1/d} \right)^d$$

Given any embedding $\varphi : X \to \ell_2^d$, for any subset $X' \subseteq X$, let $\varphi(X')$ denote the image of points in X' under φ.

LEMMA 1. *Let v be a non-leaf vertex of T, and let φ be any non-contracting embedding of X_v into ℓ_2^d. Then the volume of $\varphi(X_v) \oplus B\left(\frac{l(v)}{2}\right)$ is at least $C(v)$.*

PROOF. Let $u_1, \ldots, u_k$ be the children of v. The proof is by induction. Assume first that v belongs to level $h-1$ of T, and consider $S = \varphi(X_v) \oplus B(l(v)/2)$. Recall that $l(v) = 2$. Since the embedding is non-contracting, for any $1 \leq i < j \leq k$, vertices u_i, u_j are embedded at a distance at least 2 from each other. Therefore, set S consists of k balls of disjoint interiors, of radius 1 each, and thus the volume of S is exactly $kV_d(1) = C(v)$.

Assume now that v belongs to some level $j \in [h-2]$. Let $S = \varphi(X_v) \oplus B(l(v)/2)$. Equivalently, S is the union of $S_i = \varphi(X_{u_i}) \oplus B(l(v)/2)$ for $i \in [k]$. Since the embedding is non-contracting, all the sets S_i have disjoint interiors. For each $i \in [k]$, let us denote $S'_i = \varphi(X_{u_i}) \oplus B(l(u_i)/2)$. Recall that $l(v) = 2l(u_i)$. Therefore, for each $i \in [k]$, $S_i = S'_i \oplus B(l(v)/4)$. Using the induction hypothesis, the volume of S'_i is at least $C(u_i)$. From the Brunn-Minkowski inequality, it follows that:

$$\begin{aligned}(\mathrm{Vol}(S_i))^{1/d} &\geq \left(\mathrm{Vol}(S'_i)\right)^{1/d} + (V_d(l(v)/4))^{1/d} \\ &\geq (C(u_i))^{1/d} + (V_d(l(v)/4))^{1/d}\end{aligned}$$

Therefore, in total,

$$\begin{aligned}\mathrm{Vol}(S) = \sum_{i=1}^{k} \mathrm{Vol}(S_i) &\geq \sum_{i=1}^{k} \left((C(u_i))^{1/d} + (V_d(l(v)/4))^{1/d} \right)^d \\ &= C(v).\end{aligned}$$

□

Suppose we are given some set of points $S \subseteq \Re^d$, that has volume V. We define $\rho_d(V) = \left(\frac{V \cdot \Gamma(1+d/2)}{\pi^{d/2}}\right)^{1/d}$, i.e., $\rho_d(V)$ is the radius of the d-dimensional ball in $\Re^d$ that has volume V. Observe that S has two points at a distance at least $\rho_d(V)$ from each other (otherwise, S is contained in a ball of radius smaller than $\rho_d(V)$, which is impossible).

COROLLARY 1. *Let v be some non-leaf vertex of T, and let φ be any non-contracting embedding of M' into ℓ_2^d, with distortion at most c'. Then $c' \geq \rho_d(C(v))/l(v) - 1$.*

PROOF. Consider $S = \varphi(X_v) \oplus B\,(l(v)/2)$. By Lemma 1, the volume of S is at least $C(v)$, and thus there are two points $x, y \in S$ within a distance at least $\rho = \rho_d(C(v))$ from each other. By the definition of S, it follows that there are two points $a, b \in X_v$, which are embedded at a distance of at least $\rho - l(v)$ from each other. As the distance between a, b in T is at most $l(v)$, the bound on the distortion follows. □

4. APPROXIMATION ALGORITHM FOR EMBEDDING ULTRAMETRICS INTO THE PLANE

4.1 Preliminaries and Intuition

Let $M = (X, D)$ be the input ultrametric that embeds into the plane with distortion c. Let $M' = (X, D')$ be the metric defined by the 2-HST T which 2-approximates M. Then M' embeds into the plane with distortion $c' \leq 2c$, and any non-contracting embedding of M' into the plane with distortion $O(c'^3)$ is also a non-contracting embedding of M with distortion at most $O(c^3)$. Therefore, from now on we concentrate on embedding M' into the plane.

Consider some non-leaf vertex u. We define $a_u = \sqrt{C(u)}$. If $u \neq r$, let v be its father. We define $b_u = a_u + \frac{\sqrt{\pi} l(v)}{4}$.

Our algorithm works in bottom-up fashion. Let v be some vertex. The goal of the algorithm is to embed all the vertices of X_v into a square Q of side a_v, incurring only small distortion. Let $u_1, \ldots, u_k$ be the children of v, and assume that for all $j : 1 \leq j \leq k$, we have already embedded $X(u_j)$ inside a square Q_j of side a_{u_j}. Recall that for any pair of vertices $x \in X_{u_j}$, $y \in X_{u_{j'}}$, where $1 \leq j \neq j' \leq k$, the distance between x and y in T is $l(v)$. Our first step is to ensure non-contraction (or more precisely small contraction), by adding empty strips of width $\frac{b_{u_j} - a_{u_j}}{2} = \frac{\sqrt{\pi} l(v)}{8}$ around the squares. Thus, we obtain a collection $Q'_1, \ldots, Q'_k$ of squares, of sides $b_{u_1}, \ldots, b_{u_k}$, respectively. Our goal now is to pack these squares into one large square Q of side a_v. Observe that from volume view point, $\mathrm{Vol}(Q) = \mathrm{Vol}(Q'_1) + \ldots + \mathrm{Vol}(Q'_k)$, since $a_v^2 = \sum_{j=1}^{k} b_{u_j}^2$, by the definition of C_v. However, it is not always possible to obtain such tight packing of squares. Instead, we convert each square Q'_j to rectangle R_j whose sides are $b_{u_j} s_{u_j}$, b_{u_j}/s_{u_j} for some $s_{u_j} = O(c')$. Observe that the volume of R_j is the same as that of Q'_j. This will enable us to pack all the rectangles $R_1, \ldots, R_k$ into Q. Recall that inside each square Q'_j, vertices of X_{u_j} are embedded. In order to convert square Q'_j into rectangle R_j, we contract all the distances along one axis, and expand all the distances along the other axis, by the same factor s_{u_j}.

Consider now two vertices u, v, and let z be their least common ancestor. The distance between u and v might thus be contracted or expanded when we calculate the embedding of X_z. However, for each vertex z' on the path from z to r, the distance between u and v might be contracted or expanded again, when calculating the embedding of $X_{z'}$. In order to avoid accumulation of distortion, we would like to alternate the contractions and expansions of this distance in an appropriate way. To this end, we calculate, for each vertex v, a value $g(v) \in \{-1, 1\}$. Let $u_1, \ldots, u_k$ be the children of v, and let $Q'_1, \ldots, Q'_k$ be their corresponding squares. If $g(v) = 1$, then when embedding squares $Q'_1, \ldots, Q'_k$ into square Q of side a_v, we expand them along axis x and contract along axis y. If $g(v) = -1$, we do the opposite. The values of $g(v)$ have to be computed in a top-bottom fashion. They are calculated in such a way that the total distortion of distance between any pair of points in X stays below $poly(c')$.

For any non-root vertex u in T, with parent a vertex v, we define $s_u = a_v/b_u$. Also, for the root r of T, let $s_r = 1$.

LEMMA 2. *For each vertex u, $1 \leq s_u \leq 32c'$.*

PROOF. If u is the root, then $s_u = 1$. Otherwise, let $u, v \in T$, such that v is the father of u. We have already observed that a_v^2 is the sum of $b_{u_j}^2$, for all children u_j of v. Thus, $s(u) \geq 1$ holds.

Recall now that by the definition of b_u, its value is at least $\frac{l(v)}{4}$. On the other hand, by Corollary 1, $c' \geq \frac{a_v}{l(v)\sqrt{\pi}} - 1$, and thus $a_v \leq (c'+1)\sqrt{\pi} l(v) \leq 8c' l(v)$. Therefore, $s_u = \frac{a_v}{b_u} \leq 32c'$. □

Let v be some non-leaf vertex, and let $u_1, \ldots, u_k$ be its children. Let $Q'_1, \ldots, Q'_k$ be the squares of side $b_{u_1}, \ldots, b_{u_k}$, respectively, corresponding to the children. In order to pack these squares into a square of side a_v, we transform each square Q'_j into a rectangle with sides $b_{u_j} s_j$, $\frac{b_{u_j}}{s_j}$. The goal of the next lemma is to calculate the values $g(v) \in \{-1, 1\}$ for each $v \in V$, that will determine, along which axis we contract, and along which expand when embedding the subtree of v.

Suppose we have a function $g : V(T) \to \{-1, 1\}$. Consider some vertex $v \in V(T)$, and let $v_1, v_2, \ldots, v_k$ be the vertices on the path from v to r, where $v_1 = r$, $v_k = v$. We define $h(v) = \prod_{j=1}^{k-1} s_{v_{j+1}}^{g(v_j)}$.

LEMMA 3. *We can calculate, in linear time, function $g : V(T) \to \{-1, 1\}$, such that for each $v \in V(T)$, $\frac{1}{32c'} \leq h(v) \leq 32c'$.*

PROOF. Observe first that in order to be able to calculate $h(v)$ for any $v \in V$, it is enough to know the values of $g(v')$ of all the vertices v' on the path from r to v, not including v.

We traverse the tree in the top-bottom fashion. For root r, we set $g(r) = 1$. Since for all the values s_v, $1 \leq s_v \leq 32c'$ holds, we have that for each level-2 vertex v, $\frac{1}{32c'} \leq h(v) \leq 32c'$ holds, as required.

Consider now some vertex $v \in V$ at level k, where $k \geq 2$. Let $v_1, v_2, \ldots, v_k$ be the vertices on the path from r to v, where $v_1 = r$, and $v_k = v$, and assume we have calculated $g(v_1), \ldots, g(v_{k-1})$, such that for each $j : 2 \leq j \leq k$, $\frac{1}{32c'} \leq h(v_j) \leq 32c'$ holds. We set $g(v) = 1$ if $h(v_k) \leq 1$, and we set $g(v) = -1$ otherwise. Let u be a child of v. Since $h(u) = h_v \cdot s_u^{g(v)}$, and $s_u \leq 32c'$, the inequality $\frac{1}{32c'} \leq h(u) \leq 32c'$ holds.

It is easy to see that the running time of the above algorithm is linear, if the values $h(v)$ of the vertices calculated by the algorithm are stored in a table. The algorithm traverses each vertex only once, and for each vertex v the calculation of $h(v)$ and $g(v)$ takes only constant time. $\square$

4.2 Algorithm Description

The algorithm consists of two phases. The first phase is preprocessing, and the second phase is computing the embedding itself.

Phase 1: Preprocessing. In this phase we translate the input ultrametric M into a 2-HST T, and calculate the values $a_v, b_v, s_v, g(v)$ for each vertex $v \in T$. Each one of these operations takes time linear in the input size.

Phase 2: Computing the embedding. The algorithm works in a bottom-up fashion. For any vertex v in tree T, we produce an embedding of vertices X_v inside a square of side a_v. We start from level-h vertices (the leaves). Let v be such vertex. Then $a_v = \sqrt{C(v)} = \sqrt{\pi/4}$. We embed this point in the center of a square with a side of length $\sqrt{\pi/4}$.

Consider some level-i vertex v, for $1 \leq i < h$, and let $u_1, \ldots, u_k$ be its children. We assume that for each $j : 1 \leq j \leq k$, we have calculated the embeddings of u_j into a square Q_j of side a_{u_j}. We convert this square into a rectangle R_j, as follows. First, we add an empty strip of width $\frac{\sqrt{\pi}l(v)}{8}$ along the border of Q_j, so that now we have a new square Q'_j of side b_{u_j}. If $g(v) = 1$, then we expand the square along axis x and contract it along axis y by the factor of s_{u_j}. Otherwise, we expand square Q'_j along axis y and contract it along axis x by the factor of s_{u_j}. Notice that by the definition of s_{u_j}, the length of the longer side of R_j is precisely a_v. As the volume of R_j equals to the volume of Q'_j, and since $a_v^2 = \sum_{j=1}^{k} b_{u_j}^2$, we can pack all the rectangles next to each other inside a square Q of side a_v, with their longer side parallel to the x-axis if $g(v) = 1$, and to y-axis otherwise.

4.3 Analysis

The goal of this section is to bound the distortion produced by the algorithm. We first bound the maximum contraction, and then the maximum expansion of distances.

LEMMA 4. *For any $u, u' \in X$, the distance between the images of u and u', is at least $\Omega(1/c')D(u, u')$.*

PROOF. Let v be the least common ancestor of u, u'.

Let z, z' be the children of v, to whose subtrees vertices u, u' belong, respectively. Let Q, Q' be the squares into which X_z, and $X_{z'}$ are embedded, respectively, and let R, R' be the corresponding rectangles. Recall that we have added a strip of width at least $\frac{\sqrt{\pi}l(v)}{4}$ to squares Q, Q', and then stretched the new squares by a factors of $s(z), s(z')$, respectively. Without loss of generality, we can assume $s(z) \geq s(z')$. Therefore, immediately after computing the embedding for X_v, there is a strip S of width at least $\frac{l(v)}{4s(z)}$ between the rectangles R, R'. The width of strip S in the final embedding is a lower bound on the distance between the images of u and u'. Let $v_1, \ldots, v_k$ be the vertices on the path from r to v, where $v_1 = r, v_k = v$. Let $u_{k+1} = z$. If $g(v) = 1$, then strip S is horizontal, and thus for each $j : 1 \leq j \leq k-1$, if $g(v_j) = 1$ then its width decreases by the factor of $s(v_{j+1})$, and if $g(v_j) = -1$ then its width increases by the same factor. Thus, the final width of S is at least: $\frac{l(v)}{4s(z)^{g(v)}} \prod_{j=1}^{k-1} s(v_{j+1})^{-g(v_j)} = \frac{l(v)}{4} \prod_{j=1}^{k} s(v_{j+1})^{-g(v_j)} \geq \frac{l(v)}{4h(z)} \geq \frac{l(v)}{128c'}$.

If $g(v) = -1$, then strip S is vertical, and thus for each $j : 1 \leq j \leq k-1$, whenever $g(v_j) = 1$, the width of the strip grows by the factor of $s(v_{j+1})$, and whenever $g(v_j) = -1$, this width decreases by the same factor. Thus, in this case, the final width of S is at least: $\frac{l(v)}{4} s(z)^{g(v)} \prod_{j=1}^{k-1} s(v_{j+1})^{g(v_j)} = \frac{l(v)}{4} \prod_{j=1}^{k} s(v_{j+1})^{g(v_j)} \geq \frac{l(v)}{128c'}$.

As $D(u, u') = l(v)$, this concludes the proof of the lemma. $\square$

LEMMA 5. *For any $u, u' \in X$, the distance between the images of u and u', is at most $O(c'^2)D(u, u')$.*

PROOF. Let v be the least common ancestor of u, u'. Then $D(u, u') = l(v)$. Following Corollary 1, $c' \geq \sqrt{\frac{C(v)}{\pi}}/l(v) - 1$, and thus $a_v \leq (c'+1)\sqrt{\pi}l(v) \leq 4c'l(v)$.

When calculating the embedding of X_v, all the vertices in X_v were embedded inside a square A whose side is $a_v \leq 4c'l(v) = O(c'D(u, u'))$.

After computing the final embedding, A is mapped to a rectangle A', which is obtained from A by expanding by a factor of γ along one axis, and by expanding by a factor of $1/\gamma$ along the other axis. If $v_1, \ldots, v_k$ are all the vertices along the path from the root $r = v_1$ to $v = v_k$, then $\gamma = \prod_{j=1}^{k-1} s(v_{j+1})^{g(v_j)} = h(v)$. Thus, by Lemma 3, γ is at least $\Omega(1/c')$, and at most $O(c')$. It follows that the diameter of A' is at most $O(c'^2 D(u, u'))$. Since the images of u and u' in the final embedding are contained inside A', the assertion follows. $\square$

The following result is now immediate:

THEOREM 2. *Given an ultrametric M that c-embeds into the Euclidean plane, we can compute in linear time an embedding of M into the Euclidean plane with distortion $O(c^3)$.*

5. UPPER BOUND ON ABSOLUTE DISTORTION

In this section we show that for any $d \geq 2$, any n-point ultrametric can be embedded into ℓ_2^d with distortion $O(d^{1/2}n^{1/d})$.

Given an ultrametric M, we first compute an α-HST T that α-approximates M, for some constant $\alpha > 16$. Let M' be the metric associated with T. Observe that any embedding of M' into ℓ_2^d with distortion c, is also an embedding of M into ℓ_2^d, with distortion $O(c)$. Thus, it suffices to show that M' can be embedded into ℓ_2^d with distortion $O(d^{1/2}n^{1/d})$.

We will compute an embedding of M' into ℓ_2^d inductively, starting from the leaves of T. For every subtree of T rooted at a vertex u, we compute an embedding f_u of the submetric of M' induced by X_u, into ℓ_2^d. We maintain the following inductive properties of f_u:

- The contraction of f_u is at most 16.
- $f(X_u)$ is contained inside a hypercube of side length $l(u)n_u^{1/d}$.

We assume w.l.o.g. that for each leave v of T, $l(v) = 1$. Thus, we can embed each leave in a center of a hypercube of side 1. The following lemma shows how to compute the recursive embedding of inner vertices of T.

LEMMA 6. *Let v be an internal vertex of T, whose children are $u_1, \ldots, u_k$. Assume that for each $i \in [k]$, we are given an embedding $f_{u_i} : X_{u_i} \to \mathbb{R}^d$, with contraction at most 16, such that $f_{u_i}(X_{u_i})$ is contained inside a d-dimensional hypercube S_{u_i}, with side length $l(u_i)n_{u_i}^{1/d}$. Then we can compute in polynomial time an embedding $f_v : X_v \to \mathbb{R}^d$, with contraction at most 16, such that $f_v(X_v)$ is contained inside a d-dimensional hypercube S_v, with side length $l(v)n_v^{1/d}$.*

PROOF. For each $i \in [k]$, let $r_i = l(u_i)n_{u_i}^{1/d}$ be the length of the side of the hypercube S_{v_i}. Let also S'_{u_i} be a hypercube of side length $r'_i = r_i + l(v)/16$, having the same center as S_{u_i}. We assume w.l.o.g. that $n_1 \geq n_2 \geq \cdots \geq n_k$ and thus $r'_1 \geq \cdots \geq r'_k$. We note that for each $i : 1 \leq i \leq k$, $r'_i \leq l(v)n_v^{1/d}/4$, since $r'_i = r_i + l(v)/16 = l(u_i)n_{u_i}^{1/d} + l(v)/16 \leq l(v)n_v^{1/d}/4$.

We first define a partition $\mathcal{R} = \{R_j\}_{j=1}^{\lambda}$, of the set $[k]$, which we will use to partition the set of hypercubes $\{S_{u_i}\}_{i=1}^{k}$, as follows. We will define $\lambda+1$ integers $t_0, t_1, \ldots, t_\lambda$, where $t_0 = 0$, $t_\lambda = k$, and $t_0 < t_1 < \cdots < t_\lambda$, and then set R_j to contain all the indices $i : t_{j-1} + 1 \leq i \leq t_j$. This defines a partition of the hypercubes into λ sets $\mathcal{S}_1, \ldots, \mathcal{S}_\lambda$, where $\mathcal{S}_j$ contains the hypercubes S_{u_i} with $i \in R_j$. For each $j : 1 \leq j \leq \lambda$, let $\rho_j = r'_{t_{j-1}+1}$ denote the side of the largest hypercube in $\mathcal{S}_j$, and let $\rho'_j = r_{t_j}$ denote the side of the smallest hypercube in $\mathcal{S}_j$.

We now proceed to define the numbers t_j, for $j : 0 \leq j \leq \lambda$. Set $t_0 = 0$, and for each $j \geq 1$, if $t_{j-1} < k$, we inductively define t_j as

$$t_j \quad = \quad \min\{k, t_{j-1} + \lfloor l(v)n_v^{1/d}/r'_{t_{j-1}+1}\rfloor^{d-1}\}.$$

If $t_j = k$ then we set $\lambda = j$.

Note that for any $j \in [\lambda - 1]$,

$$|R_j| = \left\lfloor \frac{l(v)n_v^{1/d}}{\rho_j} \right\rfloor^{d-1}$$

We now define the embedding f_v by placing the hypercubes S'_{u_i} inside a hypercube of side length $l(v)n_v^{1/d}$, such that their interiors do not overlap, using the partition $\mathcal{R}$. For each $j \in [\lambda]$, we place the hypercubes in $\mathcal{S}_j$ inside a parallelepiped W_j having $d-1$ sides of length $l(v)n_v^{1/d}$, and one side of length ρ_j, as follows. It is easy to see that we can pack $|R_j|$ d-dimensional hypercubes of side ρ_j inside W_j. Since each hypercube in $\mathcal{S}_j$ has side at most ρ_j, we can replace each hypercube embedded into W_j by a hypercube from $\mathcal{S}_j$, such that the centers of both hypercubes coincide.

Finally, we place the parallelepipeds W_j inside a parallelepiped W having $d-1$ sides of length $l(v)n_v^{1/d}$, and one side of length $\sum_{j=1}^{\lambda} \rho_j$. Observe first that the contraction of this embedding is at most 16: for any pair of vertices $x, y \in X(v)$, if x, y both belong to a subtree of the same child u_i of v, then by induction hypothesis the distance between them is contracted by at most 16. If $x \in X(u_i), y \in X(u_{i'})$ and $i \neq i'$, then the original distance is $D(x,y) = l(v)$. Since we add emty space of width $l(v)/32$ around the hypercubes $S(u_q)$ when they are transformed into hypercubes $S'(u_q)$, it is clear that the distance between the embeddings of x and y is at least $l(v)/16$.

It now only remains to show that $\sum_{j=1}^{\lambda} \rho_j \leq l(v)n_v^{1/d}$. We partition the parallelepipeds W_j into two types. The first type contains all the parallelepipeds W_j, where $\rho_j/\rho'_j \geq 2$. Additionally, the last parallelepiped W_k is also of the first type, regardless of the ratio ρ_k/ρ'_k. Let $\mathcal{T}_1 \subseteq [k]$ contain all the indices j where W_j is of the first type. All the other parallelepipeds belong to the second type, and let $\mathcal{T}_2 = [k] \setminus \mathcal{T}_1$ contain the indices of the parallelepipeds of the second type. Notice that for $j \in \mathcal{T}_1$, the values ρ_j form a geometric series with ratio $1/2$. Since the sides r'_i of the hypercubes S_{u_i} are bounded by $l(v)n_v^{1/d}/4$, it is easy to see that:

$$\sum_{j \in \mathcal{T}_1} \rho_j \leq \frac{l(v)n_v^{1/d}}{4}\left(1 + \frac{1}{2} + \frac{1}{4} + \cdots\right) \leq \frac{l(v)n_v^{1/d}}{2}$$

It now remains to bound $\sum_{j \in \mathcal{T}_2} \rho_j$. Fix some $j \in \mathcal{T}_2$, and consider some hypercube S'_{u_i} where $i \in R_j$. As W_j is of the second type, we know that $r'_i \geq \rho_j/2$. On the other hand,

$$\begin{aligned} r'_i = r_i + \frac{l(v)}{16} &= l(u_i)n_{u_i}^{1/d} + \frac{l(v)}{16} \\ &\leq \frac{l(v)}{16}\left(1 + n_{u_i}^{1/d}\right) \leq \frac{l(v)}{4} n_{u_i}^{1/d} \end{aligned}$$

Therefore, $n_{u_i} \geq \left(\frac{2\rho_j}{l(v)}\right)^d$. Recall that for $j : 1 \leq j < \lambda$, $|R_j| = \left\lfloor \frac{l(v)n_v^{1/d}}{\rho_j} \right\rfloor^{d-1} \geq \left(\frac{l(v)n_v^{1/d}}{2\rho_j}\right)^{d-1}$. Therefore, we have that

$$\sum_{i \in R_j} n_{u_i} \geq \left(\frac{l(v)n_v^{1/d}}{2\rho_j}\right)^{d-1} \cdot \left(\frac{2\rho_j}{l(v)}\right)^d \geq \frac{2\rho_j}{l(v)} n_v^{1-1/d}$$

Thus, $\rho_j \leq \frac{l(v)\sum_{i \in R_j} n_{u_i}}{2n_v^{1-1/d}}$, and

$$\sum_{j \in \mathcal{T}_2} \rho_j \leq \frac{l(v)n_v}{2n_v^{1-1/d}} \leq \frac{l(v)n_v^{1/d}}{2}$$

We have that in total, $\sum_j \rho_j = \sum_{j \in \mathcal{T}_1} \rho_j + \sum_{j \in \mathcal{T}_2} \rho_j \leq l(v)n_v^{1/d}$. □

We are now ready to prove the main theorem of this section.

THEOREM 3. *For any $d \geq 2$, any n-point ultrametric can be embedded into ℓ_2^d with distortion $O(d^{1/2}n^{1/d})$. Moreover, the embedding can be computed in polynomial time.*

PROOF. Starting from the leaves of T, we inductively compute for each $v \in V(T)$ the embedding f_v as described above. By recursively applying Lemma 6 we can compute in polynomial time the embedding f_v, that also satisfies the inductive properties. Let f be the resulting embedding f_r.

Consider now two points $x, y \in X$, and let v be the nearest common ancestor of x and y. Since $f_v(X_v)$ is contained inside a hypercube of side length $l(v)n_v^{1/d}$, it follows that $\|f(x) - f(y)\|_2 \leq \left(dn_v^{2/d}l^2(v)\right)^{1/2} = d^{1/2}n^{1/d}D(x,y)$. Also, since the contraction of f_v is at most 16, it follows that the distortion of f is $O(d^{1/2}n^{1/d})$.

□

Observe that for $d = 2$, the algorithm provides an $O(\sqrt{n})$-distortion embedding. Combining this with the $O(c^3)$-distortion algorithm from Section 4, we obtain the following result:

THEOREM 4. *There is an efficient $O(n^{1/3})$-approximation algorithm for minimum distortion embedding of ultrametrics into the plane.*

PROOF. Let c be the optimal distortion achievable by any embedding of the input ultrametric into the plane. If $c > n^{1/6}$ then the above algorithm, which produces an $O(\sqrt{n})$-distortion embedding is an $O(n^{1/3})$-approximation. Otherwise, if $c \leq n^{1/6}$, then the algorithm from Section 4 gives $O(c^2) = O(n^{1/3})$-approximation.

□

We remark that Theorem 3 generalizes a result of Gupta [13], who shows that every n-point weighted star metric can be embedded into $\mathbb{R}^d$, with distortion $O(n^{1/d})$. This is a corollary of the following simple observation.

CLAIM 1. *Every n-point weighted star can be embedded into an ultrametric of size $O(n)$ with distortion at most 2.*

PROOF. Consider a star S with root r, and leaves $x_1, \ldots, x_n$, where for each $i \in [n]$, $D_S(r, x_i) = w_i$. Assume w.l.o.g. that $w_1 \leq w_2 \leq \ldots \leq w_n$. We construct a tree T with root r' as follows. T contains a path $z_n, z_{n-1}, \ldots, z_1$, where $z_n = r'$, and for each $i \in [n-1]$, $D_T(r', z_i) = w_n - w_i$. We now embed S into T as follows. For each $i \in [n]$, we add x_i to T, and we connect x_i to z_i with an edge of length w_i. Observe that the shortest-path metric on the leaves of T is an ultrametric, since all the leaves are on the same level. Moreover, for any $i < j \in [n]$, $D_T(x_i, x_j) = 2w_j$, while $D_S(x_i, x_j) = w_i + w_j$, and so the resulting embedding is non-contracting, and has expansion at most 2. □

6. NP-HARDNESS OF EMBEDDING ULTRAMETRICS INTO THE PLANE

In this section we consider embeddings into the plane under the ℓ_∞ norm. We say that a square $S \subset \mathbb{R}^2$ is *orthogonal* if the sides of S are parallel to the axes.

We will show that the problem of computing a minimum distortion embedding of an ultrametric into the plane under the ℓ_∞ norm is NP-hard. We perform a reduction from the following NP-complete problem (see [16]): Given a packing square S and a set of packed squares $L = \{s_1, \ldots, s_n\}$, is there an orthogonal packing of L into S? We call this problem SQUAREPACKING.

For a square s, let $a(s)$ denote the length of its side. Assume w.l.o.g. for each $i \in [n]$, $a(s_i) \in \mathbb{N}$, $a(S) \in \mathbb{N}$, and that $a(s_1) \leq a(s_2) \leq \ldots \leq a(s_n)$. The SQUAREPACKING problem is strongly NP-complete. Thus we can assume w.l.o.g. that there exists $N = \text{poly}(n)$, such that $1 \leq a(s_1) \leq \ldots \leq a(s_n) \leq a(S) < N$.

6.1 The Construction

Consider an instance of the SQUAREPACKING problem, where S is the packing square, and $L = \{s_1, \ldots s_n\}$ is the set of packed squares. We will define an ultrametric $M = (X, D)$ and an integer k, such that M embed into the plane with distortion at most $k-1$ iff there exists an orthogonal packing of L into S. It is convenient to define M by constructing its associated labeled tree T, where each $v \in V(T)$ has a label $l(v) \in \mathbb{Q}$.

Let $k = N^{10}$. For each square $s_i \in L$, we introduce a set of k^2 leaves $y_{i,1}, \ldots y_{i,k^2}$ in T. We connect all of these leaves to a vertex x_i, and we set $l(x_i) = a(s_i) - a(S)/(k-1)$. Note that $l(x_i)$ is very close to $a(s_i)$. Next, we introduce a root vertex $r \in V(T)$, and for each $i \in [n]$, we connect x_i to r. We set $l(r) = a(S)$.

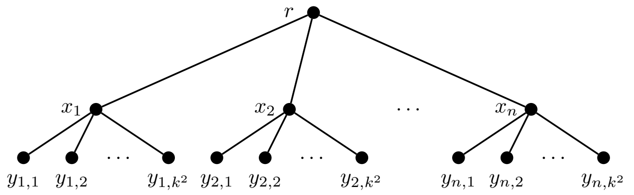

Figure 3: The constructed tree T. The labels of the vertices are: $l(r) = a(S)$ and $l(x_i) = a(s_i) - a(S)/(k-1)$.

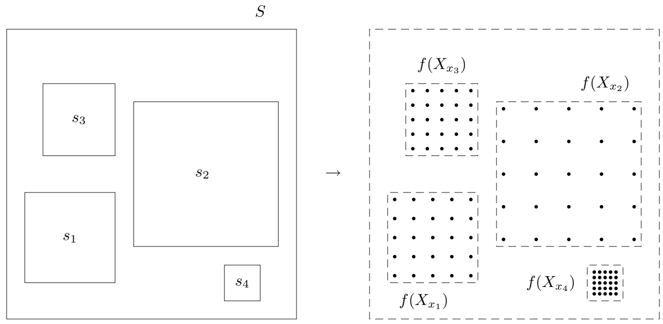

Figure 4: The embedding constructed for the YES instance.

For a vertex $v \in V(T)$, we denote by X_v the set of leaves of T having v as an ancestor. Figure 3 depicts the described construction.

6.2 YES-Instance

Assume that there exists an orthogonal packing of L into S. We will show that there exists an embedding $f : X \to \mathbb{R}^2$ with distortion $k-1$.

As a first step, for each vertex $x_i : 1 \leq i \leq n$, we embed all the vertices of X_{x_i} in a square Q_i of side $(k-1)l(x_i)$. This is done by simply placing a $k \times k$ orthogonal grid with step $l(x_i)$ inside Q_i and embedding the vertices of X_{x_i} on the grid points. Next, we transform the squares Q_i into squares Q'_i by adding empty strips of width $a(S)/2$ around Q_i. Notice that the side of Q'_i is exactly $(k-1)l(x_i) + a(S) = (k-1)a(s_i)$. Finally, we embed the squares Q'_i into a square $\mathcal{S}$ of side $(k-1)a(S)$ according to the packing of the input squares in S.

Figure 4 depicts the resulting embedding f.

We now show that the distortion of the embedding f is at most $k-1$.

Let $u, v \in X$. We have to consider the following cases for u, v:

Case 1: $u, v \in X_{x_i}$ for some $i \in [n]$. Since the vertices of X_{x_i} are embedded on a grid of step $l(x_i)$, it follows that $\|f(u) - f(v)\|_\infty \geq l(x_i) = D(u, v)$. Thus, the contraction is at most 1. Moreover, since all the vertices of X_{x_i} are embedded inside a square Q_i of side $l(x_i)(k-1)$, the expansion is at most $k-1$.

Case 2: $u \in X_{x_i}$ and $v \in X_{x_j}$, for some $i \neq j$. Since we add empty strips of width $a(S)/2$ around the squares Q_i, Q_j, we have that $\|f(u) - f(v)\|_\infty \geq a(S) = l(r) = D(u, v)$. Thus, the contraction is 1. On the other hand, all the vertices are embedded inside a square $\mathcal{S}$ of side $l(r)(k-1) = a(S)(k-1)$, and therefore the expansion is at most $k-1$.

Thus, we have shown that the distortion is at most $k-1$.

6.3 NO-Instance

Assume that there is no orthogonal packing of L inside S. We show that the minimum distortion required to embed M into the plane is greater than $k-1$. Assume that there exists an embedding $f : X \to \mathbb{R}^2$, with distortion at most $k-1$. W.l.o.g. we can assume that f is non-contracting.

The following lemma will be useful in the analysis.

LEMMA 7. *Let $M = (X, D)$ be a uniform metric on k^2 points, for some integer $k > 0$. Then, the minimum distortion for embedding M into the plane is $k-1$. Moreover, an embedding f has distortion $k-1$ iff $f(X)$ is an orthogonal grid.*

PROOF. By scaling M, we can assume w.l.o.g. that for any $u, v \in X$, $D(u, v) = 1$. Consider an non-contracting embedding $f : X \to \mathbb{R}^2$. For any $v \in X$, let A_v be square of side length 1, centered at $f(v)$. Clearly, for any $u, v \in X$, with $u \neq v$, the interiors of squares A_u and A_v are disjoint. Let $A = \bigcup_{v \in X} A_v$. It follows that $\text{Vol}(A) = |X|$. Thus, there exist $p_1, p_2 \in A$, such that $\|p_1 - p_2\|_\infty \geq |X|^{1/2} = k$. Let $v_1, v_2 \in X$ be the centers of the squares A_{v_1}, A_{v_2} to which p_1 and p_2 belong, respectively. Then $\|f(v_1) - p_1\|_\infty \leq 1/2$, and $\|f(v_2) - p_2\|_\infty \leq 1/2$. It follows that $\|f(v_1) - f(v_2)\|_\infty \geq k-1$. Thus the distortion is at least $k-1$.

Clearly, if f maps X onto a $k \times k$ orthogonal grid, the distortion of f is $k-1$. It remains to show that this is the only possible optimal embedding.

Assume that an embedding f has distortion $k-1$, and let f be non-contracting. Observe that since the diameter of $f(X)$ is at most $k-1$, $f(X)$ must be contained inside a square K of side length $k-1$. Let $\{A_v\}_{v \in X}$ be defined as above. It follows that A is contained inside a square K' of side length k. Since $\text{Vol}(A) = \text{Vol}(K')$, it easily follows that $f(X)$ is an orthogonal $k \times k$ grid. ☐

COROLLARY 2. *For each $i \in [n]$, $f(X_{x_i})$ is an orthogonal $k \times k$ grid of side length $(k-1)l(x_i) = (k-1)a(s_i) - a(S)$.*

For each $i \in [n]$, let Q'_i be the square of side length $(k-1)a(s_i)$, that has the same center of mass as $f(X_{x_i})$.

CLAIM 2. *For each $i, j \in [n]$, $i \neq j$, the interiors of the squares Q'_i, Q'_j are disjoint.*

PROOF. Assume that the assertion is not true. That is, there exist $i, j \in [n]$, with $i \neq j$, and $p \in \mathbb{R}^2$, such that p belongs to the interiors of both squares Q'_i, Q'_j. By the definition of Q'_i and Q'_j, there are points $v_1 \in X_{x_i}$, $v_2 \in X_{x_j}$ which are embedded within distance smaller than $a(S)/2$ from p. But then $\|f(v_1) - f(v_2)\|_\infty < a(S)$, contradicting the fact that the embedding is non-contracting. ☐

CLAIM 3. *$\bigcup_{i=1}^n Q'_i$ is contained inside a square of side length $ka(S)$.*

PROOF. Since f has expansion at most $k-1$, $f(X)$ is contained inside an orthogonal square $\mathcal{S}$ of side length $(k-1)l(r) = (k-1)a(S)$. Observe that for each $i \in [n]$, for each point $p \in Q_i$, there exists $v \in X_{x_i}$, such that $\|p - f(v)\|_\infty \leq a(S)/2$. Let $\mathcal{S}'$ be the square of side length $ka(S)$ that has the same center as $\mathcal{S}$. It follows that $\mathcal{S}'$ contains $\bigcup_{i=1}^n Q'_i$. ☐

LEMMA 8. *If M can be embedded into the plane with distortion at most $k-1$, then there exists an orthogonal packing of L inside S.*

PROOF. If there exists an embedding $f : X \to \mathbb{R}^2$ with distortion $k-1$, by Claim 3 we obtain that $\bigcup_{i=1}^n Q_i$ is contained inside a square of side length $ka(S)$. Moreover, by Claim 2, the embeddings of squares Q'_i defines a feasible packing of these squares into the square $\mathcal{S}'$. Note that for each $i : 1 \leq i \leq n$, Q_i has side length $(k-1)a(s_i)$. That is, the squares $Q_1, \ldots, Q_n$ are just scaled copies of the squares $s_1, \ldots, s_n$. Thus, we obtain that there exists an orthogonal packing of L inside a square S' of side length $a(S)\frac{k}{k-1}$. Recall that $k = N^{10} > a(S)^{10}$. Thus, S' has side length less than $a(S) + 1/2$.

Since $a(S)$ and $a(s_i)$ for each $i \in [n]$ are integers, it follows that there is also an orthogonal packing of L into a square of side length $a(S)$. ☐

The following theorem is now immediate.

THEOREM 5. *The problem of minimum-distortion embedding of ultrametrics into the plane under the ℓ_∞ norm is NP-hard.*

7. APPROXIMATION ALGORITHM FOR EMBEDDING ULTRAMETRICS INTO HIGHER DIMENSIONS

In this section we extend the techniques used in Section 4, to obtain an approximation algorithm for embedding ultrametrics into ℓ_2^d.

Given an ultrametric $M = (X, D)$ that embeds into ℓ_2^d with distortion c, we first embed M into a 2-HST $M' = (X, D')$. Let T be the labeled tree associated with M', as in Section 4. Then M' embeds into ℓ_2^d with distortion $c' = O(c)$. We now focus on finding an embedding of M' into the ℓ_2^d with distortion at most $c'^{O(d)}$. The same embedding is an $c^{O(d)}$-distortion embedding of M into ℓ_2^d. We compute an embedding of M' into ℓ_2^d by recursively embedding the subtrees of vertices in a bottom-up fashion.

For any vertex u in the tree, let $a_u = (C(u))^{1/d}$. If u is a non-root vertex, let v be the father of u in T. We set $b_u = a_u + (V_d(l(v)/4))^{1/d}$, and $s_u = a_v/b_u$. If u is the root of the tree, we set $s_u = 1$.

Given a vertex v in the tree, we embed the vertices in X_v into a hypercube of side a_v, recursively. Let $u_1, \ldots, u_k$ be the children of v, and assume that for each $i \in [k]$, we are given an embedding of X_{u_i} into a d-dimensional hypercube Q_{u_i} of side length a_{u_i}. We define an additional hypercube Q'_{u_i} of side length b_{u_i} that has the same center as Q_{u_i} (i.e., Q'_{u_i} is obtained from Q_{u_i} by adding a "shell" of width $(V_d(l(v)/4))^{1/d}/2$ around Q_{u_i}). Let Q_v be a d-dimensional hypercube of side length a_v.

Note that the volume of Q_v equals the sum of volumes of Q'_{u_i}, for $1 \leq i \leq k$. This is since the volume of Q_v is $a_v^d = C(v)$, while the sum of volumes of Q'_{u_i}, $1 \leq i \leq k$ is

$$\sum_{i=1}^k b_{u_i}^d = \sum_{i=1}^k \left((C(u_i))^{1/d} + (V_d(l(v)/4))^{1/d} \right)^d = C(v).$$

Fix one coordinate $j \in [d]$. We now show how to embed the hypercubes $Q'_{u_1}, \ldots, Q'_{u_k}$ into Q_v. Consider some hypercube Q'_{u_i} : $1 \leq i \leq k$. For each dimension $j' \neq j$, we increase the length of the corresponding side of Q'_{u_i} by the factor of s_{u_i}. Additionally, we decrease the length of the side of Q'_{u_i} corresponding to the dimension j by the factor of $s_{u_i}^{d-1}$. Let R_i denote the resulting parallelepiped. Notice that for each dimension $j' \neq j$, the length of the corresponding side of parallelepiped R_i is exactly a_v. Moreover, the volume of R_i equals the volume of Q'_{u_i}. Therefore, we can easily pack the parallelepipeds R_i, $1 \le i \le k$, inside the hypercube Q_v, where the shortest side of R_i is placed along dimension j.

As in the algorithm for embedding ultrametrics into the plane, we need to ensure that these stretchings do not accumulate as we go up the tree. To ensure this, we calculate, for each vertex v a value $g(v) \in [d]$. When calculating the embedding of the hypercubes $Q'_{u_1}, \ldots, Q'_{u_k}$ into the hypercube Q_v, we contract the hypercubes $Q'_{u_1}, \ldots, Q'_{u_k}$ along the dimension $g(v)$ and expand them along all the other dimensions.

Our next goal is to prove an analogue of Lemma 3, that shows how to calculate the values $g(v)$ so that the total distortion is not accumulated.

We start with the following claim:

CLAIM 4. *For each vertex u of the tree,* $1 \le s_u \le 8c'$.

PROOF. If u is the root of the tree, then $s_u = 1$ and the claim is trivially true. Assume now that u is not the root, and let v be its father. We denote the children of v by $u_1, \ldots, u_k$, and we assume that $u = u_i$ for some $i \in [k]$.

Recall that $s_u = a_v/b_u$, and that we have already observed that $a_v^d = \sum_{j=1}^k b_{u_j}^d$, and thus $s_u \ge 1$ clearly holds.

We now prove the second inequality. For the sake of convenience, we denote $V = (V_d(l(v)/4))^{1/d}$. Recall that $b_u = a_u + V \ge V$.

On the other hand, from Corollary 1,

$$c' \ge \rho_d(C(v))/l(v) - 1$$

Therefore, we have that

$$\rho_d(C(v)) = \left(\frac{C(v)\Gamma(1 + d/2)}{\pi^{d/2}}\right)^{1/d} \le 2c'l(v)$$

and thus

$$a_v = C(v)^{1/d} \le 2c'l(v)\left(\frac{\pi^{d/2}}{\Gamma(1 + d/2)}\right)^{1/d} = 8c'V$$

Therefore, $s_u = a_v/b_u \le 8c'V/V \le 8c'$. □

For each vertex u of the tree, for each dimension $j \in [d]$, we recursively define a value $h_j(u)$, as follows. If u is the root, then $h_j(u) = 1$ for all $j \in [d]$. Consider now some vertex u which is not the root, and let v be its father. Then we define $h_j(u) = h_j(v) \cdot s_u^{\alpha_j(v)}$, where $\alpha_j(v)$ is defined to be 1 if $j \neq g(v)$, and it is defined to be $-(d-1)$ if $i = g(v)$. Notice that $\prod_{j \in [d]} h_j(u) = 1$.

Fix any vertex $u \in V(T)$ and any dimension $j \in [d]$. Let Q_u be the hypercube of side a_u into which the vertices of X_u have been embedded when u was processed by the algorithm. Then the value $h_j(u)$ is precisely the stretch along the dimension j of Q_u in the final embedding. In other words, if we take a pair of points $x, y \in Q_u$ such that $x_j = y_j - 1$, and for all the other coordinates j', $x_{j'} = y_{j'}$, then $h_j(u)$ is precisely the distance between x and y in the final embedding. We next prove that we can calculate the values $g(v)$ in a way that ensures that that for each vertex u and for each dimension $j \in [d]$, $h_j(u)$ lies between $(O(1/c'))^d$ and $(O(c'))^d$.

LEMMA 9. *We can compute in polynomial time values $g(u)$ for all $u \in V(T)$, such that for each $u \in V(T)$, for each dimension $j \in [d]$, $(O(1/c'))^d \le h_j(u) \le (O(c'))^d$.*

PROOF. If u is the root, then we arbitrarily set $g(u) = 1$.

Consider now some non-root vertex u, and let v be its parent. Let $j \in [d]$ be the dimension for which $h_j(v)$ is maximized. Then we set $g(u) = j$.

CLAIM 5. *For every vertex u,* $\frac{\max_i\{h_i(u)\}}{\min_i\{h_i(u)\}} \le (8c')^d$.

PROOF. The claim is trivially true for the root r since $\frac{\max_i\{h_i(r)\}}{\min_i\{h_i(r)\}} = 1$. For any non-root vertex u, assume that the claim is true for its parent v. Assume w.l.o.g. that $h_1(v) \ge h_2(v) \ge \cdots \ge h_d(v)$, and $g(u) = 1$. Then $h_1(u) = h_1(v)/s_u^{d-1}$, and for each $i > 1$, $h_i(u) = h_i(v) \cdot s_u$. There are three cases to consider. If $h_1(u)$ equals the maximum value among $\{h_i(u)\}_{i=1}^d$, then clearly $\frac{\max_i\{h_i(u)\}}{\min_i\{h_i(u)\}} \le \frac{\max_i\{h_i(v)\}}{\min_i\{h_i(v)\}} \le (8c')^d$ by the induction hypothesis. If $h_1(u)$ equals the minimum value among $\{h_i(u)\}_{i=1}^d$, then $\frac{\max_i\{h_i(u)\}}{\min_i\{h_i(u)\}} = \frac{h_2(u)}{h_1(u)} = \frac{s_u^d h_2(v)}{h_1(v)} \le s_u^d$. Finally, if neither of the above two cases happens, then $\frac{\max_i\{h_i(u)\}}{\min_i\{h_i(u)\}} = \frac{h_2(u)}{h_d(u)} = \frac{h_2(v)s_u}{h_d(v)s_u} \le (8c')^d$ by the induction hypothesis. □

Since $\prod_{i=1}^d h_i(u) = 1$, we get that $(O(c'))^{-d} \le h_i(u) \le (O(c'))^d$.

It is easy to see that the algorithm for computing the values $g(u)$, runs in polynomial time.

Let $f : X \to \Re^d$ denote the resulting embedding produced by the algorithm. The next two lemmas bound the maximum contraction and the maximum expansion of the distances in this embedding.

LEMMA 10. *For any pair $u, u' \in X$ of points,* $\|f(u) - f(u')\|_\infty \ge (O(c'))^{-d} D'(u, u')$.

PROOF. Fix any pair $u, u' \in X$ of vertices, and let v be their least common ancestor in the tree T. Thus, $D'(u, u') = l(v)$. Let z, z' be the children of v such that $u \in X_z$ and $u' \in X_{z'}$. Assume w.l.o.g. that $s_z > s_{z'}$. Recall that $Q'_z, Q'_{z'}$ contain empty shell of width $(V_d(l(v)/4))^{1/d}/2$ in which no vertices are embedded. When $Q'_z, Q'_{z'}$ are embedded inside Q_v, they are contracted by the factors $s_z, s_{z'}$ respectively along the ith dimension, where $i = g(v)$. Thus, in the embedding of X_v inside Q_v, the distance between the images of u and u' along the ith dimension is at least:

$$\frac{V_d(l(v)/4)}{s_u^{d-1}} = \frac{\sqrt{\pi}l(v)}{4(\Gamma(1 + d/2))^{1/d} s_u^{d-1}} \ge \frac{l(v)}{2^{O(\log d)} s_u^{d-1}}$$

In the final embedding this distance is multiplied by the factor $h_i(v)$. Thus, the final distance is at least

$$\frac{l(v)}{2^{O(\log d)} s_u^{d-1}} h_i(v) = \frac{l(v)}{2^{O(\log d)}} h_i(u) \ge \frac{l(v)}{(O(c'))^d}$$

□

LEMMA 11. *For any pair $u, u' \in X$ of points,* $\|f(u) - f(u')\|_\infty \le (O(c'))^{d+1} D'(u, u')$.

PROOF. Fix any pair $u, u' \in X$ of vertices, and let v be their least common ancestor in the tree T, so that $D'(u, u') = l(v)$.

Recall that Q_v is a hypercube of side a_v, and thus when the embedding of X_v has been computed, the distance between the images of u and u' was at most a_v. In the final embedding this distance increased by the factor of at most $\max_{i \in [d]} \{h_i(v)\} \le$

$(O(c'))^d$, and thus the final distance is at most $a_v\,(O(c'))^d$. From Corollary 1, using the same reasoning as in the proof of Claim 4, we have that

$$a_v \leq 2c'l(v)\frac{\sqrt{\pi}}{(\Gamma(1+d/2))^{1/d}} \leq O(c')l(v)$$

Thus, $\|f(u) - f(u')\|_\infty \leq (O(c'))^{d+1}\,l(v)$. $\square$

Combining the results of Lemma 10 and Lemma 11, we obtain the following theorem.

THEOREM 6. *For any $d > 2$, there is a polynomial time algorithm that embeds any input ultrametric M into ℓ_2^d with distortion $c^{O(d)}$, where c is the optimal distortion of embedding M into ℓ_2^d.*

8. CONCLUSIONS AND OPEN PROBLEMS

In this paper we investigated the problem of embedding ultrametrics into low-dimensional spaces $\Re^d$. In particular, for $d = 2$, we provided two results. The first one was relative: a linear-time algorithm which, given any ultrametric c-embeddable into the plane, produces an embedding with distortion $O(c^3)$. The second result was absolute: any n-point ultrametric can be embedded into the plane with distortion $\sqrt{n}$.

The key question left open by this work is: is it possible to generalize our results to a larger class of (weighted) metrics ? In particular, it would be very interesting to design an algorithm for relative embeddings of (weighted) tree metrics. Such metrics are encountered in many applied areas, such as computational biology. Similarly, it would be interesting to obtain an $o(n)$-distortion embedding of weighted tree metrics into the plane (this problem has been posed already in [2]).

Finally, it remains to determine what is the best possible distortion of relative embeddings of ultrametrics into the plane that can be computed in polynomial time. Our results show that the answer is greater than c but smaller than $O(c^3)$, leaving a wide range of possibilities.

9. REFERENCES

[1] N. Alon, M. Bădoiu, E. Demaine, M. Farach-Colton, M. T. Hajiaghayi, and A. Sidiropoulos. Ordinal embeddings of minimum relaxation: General properties, trees and ultrametrics. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2005.

[2] R. Babilon, J. Matousek, J. Maxovǎ, and Pavel Valtr. Low-distortion embeddings of trees. In *GD '01: Revised Papers from the 9th International Symposium on Graph Drawing*, pages 343–351, London, UK, 2002. Springer-Verlag.

[3] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. *Annual Symposium on Foundations of Computer Science*, 1996.

[4] Yair Bartal and Manor Mendel. Dimension reduction for ultrametrics. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 664–665, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics.

[5] M. Bădoiu, J. Chuzhoy, P. Indyk, and A. Sidiropoulos. Low-distortion embeddings of general metrics into the line. *Annual ACM Symposium on Theory of Computing*, 2005.

[6] M. Bădoiu, K. Dhamdhere, A. Gupta, Y. Rabinovich, H. Raecke, R. Ravi, and A. Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2005.

[7] M. Bădoiu, P. Indyk, and A. Sidiropoulos. A constant-factor approximation algorithm for embedding unweighted graphs into trees. *AI Lab Technical Memo AIM-2004-015*, 2004.

[8] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological sequence analysis*. Cambridge University Press, 1998.

[9] Y. Emek and D. Peleg. Approximating minimum max-stretch spanning trees on unweighted graphs. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2004.

[10] D. Eppstein and K. A. Wortman. Minimum dilation stars. *Proceedings of the ACM Symposium on Computational Geometry*, 2005.

[11] M. Farach and S. Kannan. Efficient algorithms for inverting evolution. *Journal of the ACM*, 46, 1999.

[12] A. Gupta. Embedding trees into low dimensional euclidean spaces. *Annual ACM Symposium on Theory of Computing*, 1999.

[13] A. Gupta. Embedding trees into low dimensional euclidean spaces. *Discrete and Computational Geometry*, 24(1):105–116, 2000.

[14] P. Indyk. Tutorial: Algorithmic applications of low-distortion geometric embeddings. *Annual Symposium on Foundations of Computer Science*, 2001.

[15] C. Kenyon, Y. Rabani, and A. Sinclair. Low distortion maps between point sets. *Annual ACM Symposium on Theory of Computing*, 2004.

[16] Joseph Y.-T. Leung, Tommy W. Tam, C. S. Wong, Gilbert H. Young, and Francis Y.L. Chin. Packing squares into a square. *J. Parallel Distrib. Comput.*, 10(3):271–275, 1990.

[17] N. Linial. Finite metric spaces - combinatorics, geometry and algorithms. *Proceedings of the International Congress of Mathematicians III*, pages 573–586, 2002.

[18] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Proceedings of 35th Annual IEEE Symposium on Foundations of Computer Science*, pages 577–591, 1994.

[19] J. Matoušek. Bi-lipschitz embeddings into low-dimensional euclidean spaces. *Comment. Math. Univ. Carolinae*, 31:589–600, 1990.

[20] C. Papadimitriou and S. Safra. The complexity of low-distortion embeddings between point sets. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 112–118, 2005.