Ordinal Embedding: Approximation Algorithms and Dimensionality Reduction

Mihai Bădoiu[1], Erik D. Demaine[2⋆], MohammadTaghi Hajiaghayi[3], Anastasios Sidiropoulos[2], and Morteza Zadimoghaddam[4]

[1] Google Inc., mihai@theory.csail.mit.edu
[2] MIT Computer Science and Artificial Intelligence Laboratory, {edemaine,tasos}@mit.edu
[3] AT&T Labs — Research, hajiagha@research.att.com
[4] Department of Computer Engineering, Sharif University of Technology, zadimoghaddam@ce.sharif.edu

Abstract. This paper studies how to optimally embed a general metric, represented by a graph, into a target space while preserving the relative magnitudes of most distances. More precisely, in an ordinal embedding, we must preserve the relative order between pairs of distances (which pairs are larger or smaller), and not necessarily the values of the distances themselves. The relaxation of an ordinal embedding is the maximum ratio between two distances whose relative order is inverted by the embedding. We develop polynomial-time constant-factor approximation algorithms for minimizing the relaxation in an embedding of an unweighted graph into a line metric and into a tree metric. These two basic target metrics are particularly important for representing a graph by a structure that is easy to understand, with applications to visualization, compression, clustering, and nearest-neighbor searching. Along the way, we improve the best known approximation factor for ordinally embedding unweighted trees into the line down to 2. Our results illustrate an important contrast to optimal-distortion metric embeddings, where the best approximation factor for unweighted graphs into the line is $O(n^{1/2})$, and even for unweighted trees into the line the best is $\tilde{O}(n^{1/3})$. We also show that Johnson-Lindenstrauss-type dimensionality reduction is possible with ordinal relaxation and ℓ_1 metrics (and ℓ_p metrics with $1 \le p \le 2$), unlike metric embedding of ℓ_1 metrics.

1 Introduction

The maturing field of *metric embeddings* (see, e.g., [IM04]) originally grew out of the more classic field of *multidimensional scaling (MDS)*. In MDS, we are given a finite set of points and measured pairwise distances between them, and our goal is to embed the points into some target metric space while (approximately) preserving the distances. Originally, the MDS community considered embeddings into an ℓ_p space, with the goal of aiding in visualization, compression, clustering,

⋆ Supported in part by NSF under grant number ITR ANI-0205445.

or nearest-neighbor searching; thus, the focus was on low-dimensional embeddings. An *isometric embedding* preserves all distances, while more generally, *metric embeddings* trade-off the dimension with the fidelity of the embeddings.

But the distances themselves are not essential in nearest-neighbor searching and many contexts of visualization, compression, and clustering. Rather, the *order* of the distances captures enough information; in order words, we only need an embedding of a monotone mapping of the distances into the target metric space. The early MDS literature considered such embeddings heavily under the terms *ordinal embeddings*, *nonmetric MDS*, or *monotone maps* [CS74,Kru64a,Kru64b,She62,Tor52].

While the early work on ordinal embeddings was largely heuristic, there has been some work with provable guarantees since then. Define a *distance matrix* to be any matrix of pairwise distances, not necessarily describing a metric. Shah and Farach-Colton [SFC04] have shown that it is NP-hard to decide whether a distance matrix can be ordinally embedded into an additive metric, i.e., the shortest-path metric in a tree. Define the *ordinal dimension* of a distance matrix to be the smallest dimension of a Euclidean space into which the matrix can be ordinally embedded. Bilu and Linial [BL04] have shown that every matrix has ordinal dimension at most $n-1$. They also applied the methods of [AFR85] to show that (in a certain well-defined sense) almost every n-point metric space has ordinal dimension $\Omega(n)$. It is also known that ultrametrics have ordinal dimension exactly $n-1$ [ABD$^+$].

While ordinal embeddings and ordinal dimension provide analogs of exact isometric embedding with monotone mapping, Alon et al. [ABD$^+$] introduced an ordinal analog of distortion to enable a broader range of embeddings. Specifically, a metric M' is an *ordinal embedding with relaxation* $\alpha \geq 1$ of a distance matrix M if $\alpha M[i,j] < M[k,l]$ implies $M'[i,j] < M'[k,l]$. In other words, the embedding must preserve the relative order of significantly different distances. Note that in an ordinary ordinal embedding, we must respect distance equality, while in an ordinal embedding with relaxation 1, we may break ties. The goal of the *ordinal relaxation problem* is to find an embedding of a given distance matrix into a target family of metric spaces while minimizing the relaxation. Here we optimize the confidence with which ordinal relations are preserved, rather than the number of ordinal constraints satisfied (as in [Opa79,CS98,SFC04]).

Our results. We develop polynomial-time constant-factor approximation algorithms for minimizing the relaxation in an embedding of an unweighted graph into a line metric and into a tree (additive) metric. These two basic target metrics are particularly important for representing a graph by a structure that is easy for humans to understand, with applications to visualization, compression, clustering, and nearest-neighbor searching.

Our 10/3-approximation for unweighted graphs into the line (Section 3) illustrates an important contrast to optimal-distortion metric embeddings, where the best approximation factor for unweighted graphs into the line is $O(n^{1/2})$, and even for unweighted trees into the line the best is $\tilde{O}(n^{1/3})$ [BDG$^+$05]. This result significantly generalizes the previously known 3-approximation for minimum-relaxation ordinal embedding of unweighted trees into the line [ABD+]. Along the way, we also improve this result to a 2-approximation. The main approach of our algorithm is to embed the given graph G into the line with additive distortion at most 4α (2α from expansion and 2α from contraction), where α is the minimum relaxation of an ordinal embedding of G into a tree. We show that this embedding has (multiplicative) ordinal relaxation at most 4α, a property not necessarily true of multiplicative distortion. When G is a tree, we show that the embedding is contractive, and thus we obtain a 2-approximation. For general graphs G, we modify the embedding by contracting certain distances to improve the (asymptotic) approximation factor to 10/3.

Our 27-approximation for unweighted graphs into trees (Section 4) is in fact an approximation algorithm for both minimum-relaxation ordinal embedding and minimum-distortion metric embedding. We show that lower bounds on the ordinal relaxation (which are also lower bounds on metric distortion) provide new insight into the structure of both problems. Our result improves the best previous 100-approximation for metric distortion, and is also the first illustration that relaxation and distortion are within constant factors of each other in this context. The main approach of our algorithm is to construct a supergraph H of the given graph G such that (1) G can be embedded into H with distortion at most 9α, where α is the minimum relaxation of an ordinal embedding of G into a tree, and (2) H can be embedded into a spanning tree of H with distortion at most 3. The resulting embedding of distortion $27\,\alpha$ is a 27-approximation for both distortion and relaxation.

In each context where we obtain constant-factor approximations, e.g., ordinally embedding unweighted graphs into the line, it remains open to prove NP-hardness or inapproximability of minimizing relaxation.

Another topic of recent interest is dimensionality reduction. The famous Johnson-Lindenstrauss Theorem [JL84] guarantees low-distortion reduction to logarithmic dimension for arbitrary ℓ_2 metrics, but recently it was shown that the same is impossible without significant distortion for ℓ_1 metrics [BC05,LN04] (despite their usefulness and flexibility for representation). In contrast, we show in Section 5 that arbitrary ℓ_1 metrics can be ordinally embedded into logarithmic-dimension ℓ_1 space with relaxation $1+\varepsilon$ for any $\varepsilon > 0$. More generally, our analog of the Johnson-Lindenstrauss Theorem applies to ℓ_p metrics with $1 \le p \le 2$. We show that this result in fact follows easily from a combination of known results: the monotone property of ordinal embeddings, power transformations for making metrics Euclidean, the Johnson-Lindenstrauss Theorem, and Dvoretzky-type results to return to the desired ℓ_p space [FLM77,Ind07].

2 Definitions

In this section, we formally define ordinal embeddings and relaxation (as in [ABD+]) as well as the contrasting notions of metric embeddings and distortion.

Consider a finite metric $D : P \times P \to [0, \infty)$ on a finite point set P—the *source metric*—and a class $\mathcal{T}$ of metric spaces $(T, d) \in \mathcal{T}$ where d is the distance function for space T—the *target metrics*. An *ordinal embedding (with no relaxation)* of D into $\mathcal{T}$ is a choice $(T, d) \in \mathcal{T}$ of a target metric and a mapping $\phi : P \to T$ of the points into the target metric such that every comparison between pairs of distances has the same outcome: for all $p, q, r, s \in P$, $D(p, q) \leq D(r, s)$ if and only if $d(\phi(p), \phi(q)) \leq d(\phi(r), \phi(s))$. Equivalently, ϕ induces a monotone function $D(p, q) \mapsto d(\phi(p), \phi(q))$. An *ordinal embedding with relaxation* α of D into $\mathcal{T}$ is a choice $(T, d) \in \mathcal{T}$ and a mapping $\phi : P \to T$ such that every comparison between pairs of distances not within a factor of α has the same outcome: for all $p, q, r, s \in P$ with $D(p, q) > \alpha\, D(r, s)$, $d(\phi(p), \phi(q)) > d(\phi(r), \phi(s))$. Equivalently, we can view a relaxation α as defining a partial order on distances $D(p, q)$, where two distances $D(p, q)$ and $D(r, s)$ are comparable if and only if they are not within a factor of α of each other, and the ordinal embedding must preserve this partial order on distances.

We pay particular attention to contrasts between relaxation in ordinal embedding relaxation and distortion in "standard" embedding, which we call "metric embedding" for distinction. A *contractive metric embedding with distortion* c of a source metric D into a class $\mathcal{T}$ of target metrics is a choice $(T, d) \in \mathcal{T}$ and a mapping $\phi : P \to T$ such that no distance increases and every distance is preserved up to a factor of c: for all $p, q \in P$, $1 \leq D(p, q)/d(\phi(p), \phi(q)) \leq c$. Similarly, we can define an *expansive metric embedding with distortion* c with the inequality $1 \leq d(\phi(p), \phi(q))/D(p, q) \leq c$. When $c = 1$, these two notions coincide to require exact preservation of all distances; such an embedding is called a *metric embedding with no distortion* or an *isometric embedding*. In general, $c^* = c^*(D, \mathcal{T})$ denotes the minimum possible distortion of a metric embedding of D into $\mathcal{T}$. (This definition is equivalent for both contractive and expansive metric embeddings, by scaling.)

3 Constant-Factor Approximations for Embedding Unweighted Graphs and Trees into the Line

In this section we give an asymptotically 10/3-approximation algorithm for minimum-relaxation ordinal embedding of the shortest-path metric of an unweighted graph into the line. This result shows a sharp contrast from metric embedding, where the best known polynomial-time approximation algorithm for unweighted graphs into the line achieves an approximation ratio of just $O(n^{1/2})$, and even for unweighted trees into the line the best is $\tilde{O}(n^{1/3})$ [BDG+05]. Along the way, we give a 2-approximation algorithm for minimum-relaxation ordinal embedding of unweighted trees into the line, improving on the 3-approximation of [ABD+].

Let $G = (V, E)$ be the input unweighted graph. Suppose that there exists an ordinal embedding h of G into the line $\mathbb{R}$ with relaxation α. Let u and v be the vertices in G that h maps onto the leftmost and rightmost points, respectively, in the line. In other words, $h(u)$ and $h(v)$ are the minimum and maximum values taken by h. The algorithm guesses the vertices u and v, i.e., repeats the following procedure for all possible pairs of vertices u and v.

For a given guess of u and v, the algorithm computes an (arbitrary) shortest path P from u to v in G, say visiting vertices $u = v_0, v_1, v_2, \ldots, v_\delta = v$. Then it computes the Voronoi partition of the vertices $V = V_0 \cup V_1 \cup \cdots \cup V_\delta$ where the sites are the vertices $v_0, v_1, \ldots, v_\delta$ of the path P, i.e., for each $i \in \{0, 1, \ldots, \delta\}$ and for each $x \in V_i$, $D_G(x, v_i) = \min\{D_G(x, v_j) \mid v_j \in P\}$. In particular, $v_i \in V_i$. This partition defines a function $f : V \to \mathbb{R}$ by $f(x) = i$ for $x \in V_i$. This function will turn out to be a good embedding if G is a tree, but it will need further refinement for general graphs. We begin by deriving some properties of f.

Lemma 1. *For any $i \in \{0, 1, \ldots, \delta\}$ and any $x \in V_i$, we have $\alpha \geq D_G(x, v_i)$, and if G is a tree, we have $\alpha \geq D_G(x, v_i) + 1$.*

Proof. Suppose for contradiction that some vertex $x \in V_i$ has $\alpha < D_G(x, v_i)$. Consider the ordinal embedding h of G into $\mathbb{R}$ with relaxation α. By construction, $h(v_0) \leq h(x) \leq h(v_\delta)$, so some j with $0 \leq j < \delta$ has $h(v_j) \leq h(x) \leq h(v_{j+1})$. By assumption, $D_G(x, v_j) \geq D_G(x, v_i) > \alpha = \alpha\, D_G(v_j, v_{j+1}) = \alpha$. By definition of relaxation, $|h(x) - h(v_j)| > |h(v_j) - h(v_{j+1})|$, contradicting that $h(v_j) \leq h(x) \leq h(v_{j+1})$.

If G is a tree, we have the property that $|D_G(x, v_j) - D_G(x, v_{j+1})| = 1$. By construction, both $D_G(x, v_j)$ and $D_G(x, v_{j+1})$ are at least $D_G(x, v_i)$, and hence the larger is at least $D_G(x, v_i) + 1 > \alpha + 1$. The rest of the proof for trees is as above. $\square$

Lemma 2. *For any two vertices x_1 and x_2 in G, we have*

$$D_G(x_1, x_2) - 2\alpha \leq |f(x_1) - f(x_2)| \leq D_G(x_1, x_2) + 2\alpha,$$

and if G is a tree, we have

$$D_G(x_1, x_2) - 2(\alpha - 1) \leq |f(x_1) - f(x_2)| \leq D_G(x_1, x_2) + 2(\alpha - 1).$$

Proof. Suppose x_1 and x_2 are in V_{i_1} and V_{i_2}, respectively. By Lemma 1, $D_G(x_1, v_{i_1}) \leq \alpha$ and $D_G(x_2, v_{i_2}) \leq \alpha$. By the triangle inequality, $D_G(x_1, x_2) \leq D_G(x_1, v_{i_1}) + D_G(v_{i_1}, v_{i_2}) + D_G(v_{i_2}, x_2) \leq \alpha + |f(x_1) - f(x_2)| + \alpha$. We also have $|f(x_1) - f(x_2)| = D_G(v_{i_1}, v_{i_2}) \leq D_G(v_{i_1}, x_1) + D_G(x_1, x_2) + D_G(x_2, v_{i_2}) \leq \alpha + D_G(x_1, x_2) + \alpha$. If G is a tree, we can replace each α with $\alpha - 1$ by Lemma 1 and obtain the stronger inequalities. $\square$

Next we show the efficiency of f as an ordinal embedding for trees, improving on the 3-approximation of [ABD+]:

Theorem 1. *There is a polynomial-time algorithm which, given an unweighted tree T that ordinally embeds into the line with relaxation α, computes an ordinal embedding with relaxation $2\alpha - 1$.*

Proof. We prove that the function f defined above is an ordinal embedding with relaxation $2\alpha - 1$.

First we claim that, for any two vertices x and y, we have $|f(x) - f(y)| \leq D_T(x, y)$. Because T is a tree, there is a unique simple path Q from x to y.

Suppose x and y belong to V_i and V_j, respectively. If $i = j$, then $f(x) = f(y)$, and the claim is trivial. Otherwise, Q must be the simple path from x to v_i to v_j (along P) to y. Therefore the length of Q is at least $|i - j| = |f(x) - f(y)|$. In other words, the embedding f does not increase the distance between x and y.

Next let x_1, x_2, x_3, x_4 be vertices of T with $D_T(x_1, x_2)/D_T(x_3, x_4) > 2\alpha - 1$. It remains to show that $|f(x_1) - f(x_2)| > |f(x_3) - f(x_4)|$. Because $\alpha \geq 1$ and $D_T(x_3, x_4) \geq 1$, we have $D_T(x_1, x_2) > (2\alpha - 1)D_T(x_3, x_4) \geq 2\alpha - 2 + D_T(x_3, x_4)$. By Lemma 2, we have $|f(x_1) - f(x_2)| \geq D_T(x_1, x_2) - 2\alpha + 2$, which is greater than $D_T(x_3, x_4)$. Above we proved that $D_T(x_3, x_4) \geq |f(x_3) - f(x_4)|$. Therefore $|f(x_1) - f(x_2)| > |f(x_3) - f(x_4)|$. $\square$

Before we define our embedding for general unweighted graphs, we prove a final property of f:

Lemma 3. *For any $\varepsilon > 1/\alpha$, any vertex x, and any vertices y_1 and y_2 adjacent to x, we have either $\min\{f(y_1), f(y_2)\} > f(x) - \alpha(1+\varepsilon)$ or $\max\{f(y_1), f(y_2)\} < f(x) + \alpha(1 + \varepsilon)$.*

Proof. Suppose for contradiction that there is a vertex x with neighbors y_1 and y_2 for which $f(y_1) \leq f(x) - \alpha(1 + \varepsilon)$ and $f(x) \leq f(y_2) - \alpha(1 + \varepsilon)$. Thus $|f(y_1) - f(y_2)| \geq 2\alpha(1 + \varepsilon)$. But $D_G(y_1, y_2) \leq 2$, so by Lemma 2 we conclude $|f(y_1) - f(y_2)| \leq 2 + 2\alpha$, which is a contradiction for $\varepsilon > 1/\alpha$. $\square$

Finally we can define our ordinal embedding $g : V \to \mathbb{R}$ for a general unweighted graph $G = (V, E)$, for any $\varepsilon > 0$:

$$g(x) = \begin{cases} f(x) - \alpha/3 & \text{if } x \text{ has a neighbor } y \text{ in } G \text{ with } f(y) \leq f(x) - \alpha(1 + \varepsilon), \\ f(x) + \alpha/3 & \text{if } x \text{ has a neighbor } y \text{ in } G \text{ with } f(y) \geq f(x) + \alpha(1 + \varepsilon), \\ f(x) & \text{otherwise.} \end{cases}$$

By Lemma 3, the embedding g is well-defined. It remains to bound its relaxation.

Lemma 4. *For any two vertices x_1 and x_2 in G, we have*

$$D_G(x_1, x_2) - 8\alpha/3 \leq |g(x_1) - g(x_2)| \leq D_G(x_1, x_2) + 8\alpha/3.$$

Proof. By construction, $|g(x) - f(x)| \leq \alpha/3$ for any vertex x. By Lemma 2,

$$D_G(x_1, x_2) - 2\alpha - 2\alpha/3 \leq |g(x_1) - g(x_2)| \leq D_G(x_1, x_2) + 2\alpha + 2\alpha/3.$$

$\square$

Lemma 5. *For any $\varepsilon > 3/(2\alpha)$ and any edge $e = (x, y)$ in G, we have $|g(x) - g(y)| \leq (4/3 + \varepsilon)\alpha$.*

Proof. Without loss of generality, suppose that $f(x) \leq f(y)$. By Lemma 2, $|f(x) - f(y)| \leq 1 + 2\alpha$. If $f(x) < f(y) - \alpha(1 + \varepsilon)$, then $g(x) = f(x) + \alpha/3$ and $g(y) = f(y) - \alpha/3$. In this case, we have

$$|g(x) - g(y)| = |f(x) - f(y)| - 2\alpha/3 \leq 2\alpha + 1 - 2\alpha/3 \leq (4/3 + \varepsilon)\alpha$$

for $\alpha \geq 1/\varepsilon$.

It remains to consider the case $f(x) \leq f(y) + (1+\varepsilon)\alpha$. Observe that $g(x)$ is equal to one of the values $f(x) - \alpha/3$, $f(x)$, and $f(x) + \alpha/3$. There are also three cases for $g(y)$. So there are nine cases to consider. But the claim is clearly true for eight of them. The only case for which the claim is nontrivial is when $g(x) = f(x) - \alpha/3$ and $g(y) = f(y) + \alpha/3$.

In this case, we have $|g(x) - g(y)| = |f(x) - f(y)| + 2\alpha/3$. By definition of g, we conclude that there is a vertex x' adjacent to x in G such that $f(x') \leq f(x) - (1+\varepsilon)\alpha$. Similarly, there is a vertex y' adjacent to y for which we have $f(y') \geq f(y) + (1+\varepsilon)\alpha$. Thus $f(y') - f(x') \geq (2+2\varepsilon)\alpha$. But we know that $D_G(x', y') \leq 3$, and $|f(x') - f(y')|$ must be at most $3 + 2\alpha$, which is a contradiction for $\varepsilon > 3/(2\alpha)$. Therefore this case does not occur, and the claim is true for all nine cases. $\square$

Lemma 6. *The ordinal embedding g has relaxation* $(10/3 + \varepsilon)\alpha + 1$ *for* $\varepsilon > 3/(2\alpha)$.

Proof. Consider $x_1, x_2, x_3, x_4 \in V$ for which $D_G(x_1, x_2)/D_G(x_3, x_4) > (10/3 + \varepsilon)\alpha + 1$. It suffices to show that $|g(x_1) - g(x_2)| > |g(x_3) - g(x_4)|$. We consider two cases.

First suppose that $D_G(x_3, x_4) > 1$. Then

$$D_G(x_1, x_2) - D_G(x_3, x_4) > [(10/3 + \varepsilon)\alpha + 1 - 1] D_G(x_3, x_4) > 20\alpha/3.$$

By Lemma 4, $|g(x_1) - g(x_2)| \geq D_G(x_1, x_2) - 8\alpha/3$ and $|g(x_3) - g(x_4)| \leq D_G(x_3, x_4) + 8\alpha/3$. Thus

$$|g(x_1) - g(x_2)| - |g(x_3) - g(x_4)| \geq [D_G(x_1, x_2) - 8\alpha/3] - [D_G(x_3, x_4) + 8\alpha/3] \geq 1.$$

Therefore $|g(x_1) - g(x_2)| > |g(x_3) - g(x_4)|$.

In the second case, there is an edge between vertices x_3 and x_4. We also know that $D_G(x_1, x_2) > (10/3 + \varepsilon)\alpha + 1$. By Lemma 5, $|g(x_3) - g(x_4)| \leq (4/3 + \varepsilon)\alpha$. It suffices to prove that $|g(x_1) - g(x_2)| > (4/3 + \varepsilon)\alpha$. By Lemma 2, $|f(x_1) - f(x_2)| \geq D_G(x_1, x_2) - 2\alpha > (4/3 + \varepsilon)\alpha$. If $|g(x_1) - g(x_2)| \geq |f(x_1) - f(x_2)|$, the claim is true. On the other hand, if $|f(x_1) - f(x_2)| > (2 + \varepsilon)\alpha$, then because $|g(x_1) - g(x_2)| \geq |f(x_1) - f(x_2)| - 2\alpha/3$, we have $|g(x_1) - g(x_2)| > (4/3 + \varepsilon)\alpha$. So we can suppose that $|g(x_1) - g(x_2)| < |f(x_1) - f(x_2)|$ and that $|f(x_1) - f(x_2)| \in [(4/3 + \varepsilon)\alpha, (2 + \varepsilon)\alpha]$. Without loss of generality, we can suppose that $f(x_1) < f(x_2)$, and consequently $f(x_2) \in [f(x_1) + (4/3 + \varepsilon)\alpha, f(x_1) + (2 + \varepsilon)\alpha]$. Because $|g(x_1) - g(x_2)| < |f(x_1) - f(x_2)|$, and by the symmetry between x_1 and x_2, we can suppose that $g(x_1) = f(x_1) + \alpha/3$ and $g(x_2) \leq f(x_2)$.

We conclude that there exists a vertex x_5 for which $e = (x_1, x_5) \in E(G)$ and $f(x_1) + (1 + \varepsilon)\alpha \leq f(x_5) \leq f(x_1) + 2\alpha$. As a consequence, $D_G(x_5, x_2) \geq D_G(x_1, x_2) - 1 > (10/3 + \varepsilon)\alpha$ and $f(x_5) \in [f(x_1) + (1 + \varepsilon)\alpha, f(x_1) + 2\alpha]$. Therefore $|f(x_5) - f(x_2)| \leq \alpha$. But this inequality contradicts that $|f(x_5) - f(x_2)| \geq D_G(x_5, x_2) - 2\alpha \geq (4/3 + \varepsilon)\alpha$. We conclude that $|g(x_1) - g(x_2)| > (4/3 + \varepsilon)\alpha$, which completes the proof. $\square$

Substituting $\varepsilon = 3/(2\alpha) + \delta/\alpha$ in Lemma 6, we obtain the following result:

Theorem 2. *For any $\delta > 0$, there is a polynomial-time algorithm which, given an unweighted graph that ordinally embeds into the line with relaxation α, computes an ordinal embedding with relaxation $(10/3)\alpha + 5/2 + \delta$*

4 Constant-Factor Approximation for Embedding Unweighted Graphs into Trees

In this section, we develop a 27-approximation for the minimum-relaxation ordinal embedding of an arbitrary unweighted graph into a tree metric. Specifically, we give a polynomial-time algorithm that embeds a given unweighted graph G into a tree with (metric) distortion at most $27\,\alpha_G$, where α_G is the minimum relaxation needed to ordinally embed G into a tree. Because the relaxation of an embedding is always at most its distortion [ABD$^+$, Proposition 1], we obtain the desired 27-approximation for minimum relaxation. Furthermore, because the optimal relaxation is also at most the optimal distortion, the same algorithm is a 27-approximation for minimum distortion. This result improves substantially on the 100-approximation for minimum-distortion metric embedding of an unweighted graph into a tree [BIS07]. Furthermore, we obtain that the minimum possible distortion c_G is $\Theta(\alpha_G)$ for any graph G, a property which is not true in many other cases [ABD$^+$].

4.1 Lower Bound for Ordinal Embedding of Graphs into Trees

We start with a lower bound on the minimum relaxation needed to embed a graph with a special structure into any tree.

Theorem 3. *Any graph G has $\alpha_G \geq 2l/3$ if there are two vertices u and v and two paths P_1 and P_2 between them with the following properties:*

1. *P_1 is a shortest path between u and v; and*
2. *there is a vertex w on P_1 whose distance to any vertex on P_2 is at least l.*

Proof. Suppose that G can be ordinally embedded into a tree T with relaxation less than $2l/3$. Let $u = v_1, v_2, \ldots, v_m = v$ be the vertices of the path P_1 in G. By Property 2, we have $m \geq 2l$ because u and v are also two vertices on P_2. Note that in addition to u and v, P_1 and P_2 may have more vertices in common. Let v_i be mapped onto v'_i in this embedding, $v'_i \in V(T)$. Let P' be the unique path between v'_1 and v'_m in T. Also suppose that x_i is the first vertex on path P' that we meet when we are moving from v'_i to v'_m. Note that such a vertex necessarily exists because v'_m is a vertex on P' which we meet during our path in T, and there might be more vertices like v'_m. According to this definition, x_i is a vertex on P', and the vertices $v'_1 = x_1, x_2, \ldots, x_m = v'_m$ are not necessarily distinct. Let k be the maximum distance between two vertices x and y in T over all pairs (x, y) with the property that their representatives in G are adjacent.

Because there is exactly one path between any pair of vertices in T, we know that, if $x_i \neq x_{i+1}$, then the vertex x_i lies in the (shortest) path between v'_i and v'_{i+1} in T. Consequently, we have $d_T(v'_i, v'_{i+1}) = d_T(v'_i, x_i) + d_T(x_i, v'_{i+1})$ where $d_T(a, b)$ is the distance between a and b in T. Note that by definition of k, for any i where $x_i \neq x_{i+1}$, the sum of these two terms is at most k. This means that either $d_T(v'_i, x_i)$ or $d_T(x_i, v'_{i+1})$ is at most $k/2$. We use this fact frequently in the rest of proof.

Let w be the ith vertex on P_1. Equivalently, let w be v_i. In order to complete our proof, we consider two cases. At first, suppose that $x_{i-l/3} = x_{i-l/3+1} = \cdots = x_i = x_{i+1} = \cdots = x_{i+l/3}$. In this case, let i_1 and i_2 be respectively the minimum and maximum numbers for which we have $x_{i_1} = x_i = x_{i_2}$. We prove that either $d_T(v'_{i_1}, x_{i_1})$ or $d_T(x_{i_1}, v'_{i_1-1})$ is at most $k/2$. If $i_1 = 1$, then we have $x_{i_1} = v'_{i_1}$ and consequently $d_T(x_{i_1}, v'_{i_1}) = 0$. Otherwise, we have $x_{i_1} \neq x_{i_1-1}$ and therefore we deduce that either $d_T(v'_{i_1}, x_{i_1})$ or $d_T(x_{i_1}, v'_{i_1-1})$ is at most $k/2$. According to the symmetry of the case, we also have that either $d_T(v'_{i_2}, x_{i_2})$ or $d_T(x_{i_2}, v'_{i_2+1})$ is at most $k/2$. Note that $x_{i_1} = x_{i_2}$. Finally we conclude that there exist $j_1 \in \{i_1-1, i_1\}$ and $j_2 \in \{i_2, i_2+1\}$ such that $d_T(v'_{j_1}, v'_{j_2}) \leq k/2+k/2 = k$. Note that the distance between v_{j_1} and v_{j_2} is at least $2l/3$ in G. Because there are two adjacent vertices in G such that their distance in T is k, we can say that the relaxation is at least $\frac{2l/3}{1} = 2l/3$.

Now we consider the second and final case. In this case, There exists a vertex $j_1 \in \{i+1-l/3, i+2-l/3, \ldots, i-1+l/3\}$ such that we have either $x_{j_1} \neq x_{j_1-1}$ or $x_{j_1} \neq x_{j_1+1}$. Using each of these inequalities, we reach the fact that there exists $j_2 \in \{j_1 - 1, j_1, j_1 + 1\}$ for which we have $d_T(v'_{j_2}, x_{j_1}) \leq k/2$. We define some similar terms for path P_2. Let $u = u_1, u_2, \ldots, u_{m'} = v$ be the vertices of the path P_2 in graph G. Let u_i is mapped onto u'_i in this embedding, $u'_i \in V(T)$. Suppose that y_i is the first vertex on path P' that we meet when we are moving from u'_i to u'_m. We know that either $x_{j_1} \neq v'_1$ or $x_{j_1} \neq v'_m$. Without loss of generality, suppose that $x_{j_1} \neq v'_1$. Now we know that $y_1 = v'_1$ lies before x_{j_1} on path P', and $y_{m'} = v'_m$ does not lie before x_{j_1} on this path. Therefore there exists a number j_3 for which y_{j_3} lies before x_{j_1} on P', and y_{j_3+1} does not lie before x_{j_1} on the path. Therefore x_{j_1} occurs in the (shortest) path between u'_{j_3} and u'_{j_3+1} in T. In the other words, we have $d_T(u'_{j_3}, u'_{j_3+1}) = d_T(u'_{j_3}, x_{j_1}) + d_T(x_{j_1}, u'_{j_3+1}) \leq k$. We can say that either $d_T(u'_{j_3}, x_{j_1})$ or $d_T(x_{j_1}, u'_{j_3+1})$ is at most $k/2$. Suppose that $d_T(u'_{j_3}, x_{j_1})$ is at most $k/2$. The proof in the other case is exactly the same.

Finally we reach the inequality $d_T(v'_{j_2}, u'_{j_3}) \leq d_T(v'_{j_2}, x_{j_1}) + d_T(x_{j_1}, u'_{j_3}) \leq k/2 + k/2 = k$. Note that the distance between v_{j_2} and $w = v_i$ is at most $l/3$ in G, and therefore the distance between v_{j_2} and u_{j_3} which is a vertex on path P_2 is at least $l - l/3 = 2l/3$ in G. Again we can say that there are two adjacent vertices in G such that their distance in T is k, and therefore the relaxation is at least $(2l/3)/1 = 2l/3$. □

4.2 27-Approximation Algorithm

In this section we embed a given graph G into a tree with distortion (and hence relaxation) at most $27\,\alpha_G$. We find the embedding in two phases. At first, we construct graph H from the given graph G only by adding some edges to G. Then we propose an algorithm which finds a spanning tree of H like T. Next, we prove that the distortion of embedding G into H is at most $O(\alpha_G)$. We also prove that the embedding H into T has distortion at most 3. Therefore the distortion of embedding G into T is $O(\alpha_G)$.

Let G be the given graph. We construct H as follows. Choose an arbitrary vertex v, and run a breadth-first search to find a tree T_v rooted at v in which the distance between each vertex and v is equal to their distance in G. The vertices of G occur in different levels of T_v. The ith level of this tree, L_i, consists of vertices whose distance to v is i. We have $L_0 = \{v\}$ and $V(G) = \bigcup_{i=0}^{n-1} L_i$. In constructing H from G, we add an edge between two vertices u_1 and u_2 if and only if u_1 and u_2 are in the same level such as L_i or in two consecutive levels such as L_i and L_{i+1}, and there is a path between u_1 and u_2 that does not use the vertices of levels $L_0, L_1, \ldots, L_{i-1}$. In the other words, there exists a path between u_1 and u_2 in graph $G[V - \bigcup_{j=0}^{i-1} L_j]$ where $G[X]$ is the subgraph of G induced by vertex set X. Using Lemma 3, we prove the following lemma.

Lemma 7. *The distortion of embedding G into H is at most $9\,\alpha_G$.*

Proof. Because we only add edges to G to form H, the distance between vertices does not increase. Therefore this metric embedding is contractive. The distortion of the embedding is thus $\max_{u,v \in V(G)=V(H)} d_G(u,v)/d_H(u,v)$. We also know that this maximum is equal to $\max_{(u,v)\in E(H)} d_G(u,v)/d_H(u,v)$ because, if we know that the distance between two vertices adjacent in H is at most k in G, then the distance between every pair of vertices in G is at most k times their distance in H. Therefore we just need to prove that, for each edge (u_1, u_2) that we add, the distance between u_1 and u_2 in G is at most $9\,\alpha_G$. In the rest of proof, when we talk about the distance between two vertices or a path between them, we consider all of them in graph G. Note that u_1 and u_2 are either in the same level such as L_i or in two consecutive levels L_i and L_{i+1}, and there is a path P_1 between them which uses only vertices in levels $L_i, L_{i+1}, \ldots$. Consider a shortest path P_2 between u_1 and u_2. There is also a unique path P_3 between u_1 and u_2 in the breadth-first-search tree rooted at v. Note that these paths are not necessarily disjoint. Let l be the length of P_2. We prove that $l \le 9\,\alpha_G$. We consider two cases.

First suppose that there is a vertex in P_2 like w that is in $\bigcup_{j=0}^{i-l/6} L_j$. For $i < l/6$, $\bigcup_{j=0}^{i-l/6} L_j$ is empty. The distance between w and any vertex in P_1 is at least $l/6$ because the distance between v and w is at most $i - l/6$, and the distance between v and any vertex in P_1 is at least i. Applying Lemma 3 to P_2 as the shortest path, P_1 as the other path, and vertex w, G cannot be ordinally embedded into any tree with relaxation less than $\frac{2}{3} \cdot \frac{l}{6} = l/9$. Therefore $9\,\alpha_G \ge l$.

In the second case, all vertices of the path P_2 are in $\bigcup_{j=i+1-l/6}^{n-1} L_j$, including the vertex in the middle of P_2. Let w be the vertex in the middle of the P_2. Because P_2 is a shortest path, the distance between w and u_1 and u_2 is at least $\frac{l-1}{2}$. We assert that the distance between w and any vertex in the path P_3 is at least $l/6$. Consider a vertex in P_3 like x. If x is in $\bigcup_{j=0}^{i+1-l/3} L_j$, the distance between w and x is at least $(i+1-l/6)-(i+1-l/3)=l/6$. Otherwise because of the special structure of path P_3, the distance between x and at least one of the vertices u_1 and u_2 is at most $i+1-(i+1-l/3+1)=l/3-1$. Because the distance between w and both u_1 and u_2 is at least $\frac{l-1}{2}$, we can say that the distance between w and x is at least $\frac{l-1}{2}-(l/3-1)\geq l/6$. Again applying Lemma 3 to P_2 as the shortest path, P_3 as the other path, and vertex w, G cannot be ordinally embedded into any tree with relaxation less than $\frac{2}{3}\cdot\frac{l}{6}=l/9$. Therefore $9\,\alpha_G\geq l$. □

Now we are going to find a spanning tree T of H with distortion at most 3. Before proposing the algorithm, we mention some important properties of H.

The subgraph $G[L_i]$ of H induced by vertices in level L_i is a union of some cliques. In fact, if there are two edges (a,b) and (b,c) in $G[L_i]$, then there must be a path between a and b in G that uses only vertices in $\bigcup_{j=i}^{n-1} L_j$, and also a path between b and c in G which uses only vertices in $\bigcup_{j=i}^{n-1} L_j$. Therefore there exists a path between a and c in G that uses only vertices in $\bigcup_{j=i}^{n-1} L_j$. Consequently we must have added an edge between a and c in constructing H from G. Because the connectivity relation in each level is transitive, each level is a union of some cliques. There is another important property of H. For any $a,b\in L_{i+1}$ and $c\in L_i$, if b is adjacent to both a and c in H, then there must be an edge between a and c in H. The claim is true because of the special definition of edges in H. Therefore, for each clique in level L_{i+1}, there exists a vertex in L_i that is adjacent to all vertices of that clique.

Now we find the tree T as follows. For any $i>0$ and any clique C in level L_i, we just need to find a vertex v_C in L_{i-1} that is adjacent to all vertices in C, and then add all edges between vertex v_C and the vertices in C into the tree. Actually this tree is a breadth-first-search tree in graph H.

Lemma 8. *The distortion of embedding H into T is at most* 3.

Proof. It is clear that we obtain a spanning tree T. The embedding is expansive because T is a subgraph of H. In order to bound the distortion of this embedding, we must prove that, for each edge (x,y) in H, the distance between x and y is at most 3 in T. There are two kinds of edges in H: the edges between vertices in the same level and edges between vertices in two consecutive levels. If x and y are in the same level L_i, then they are connected to a vertex z in L_{i-1} in tree T. Therefore their distance in tree T is 2. Otherwise, suppose that x is in L_i and y is in L_{i-1}. Vertex x is connected to a vertex z in L_{i-1} in tree T. If $z=y$, then the claim is clear. If $y\neq z$, then by definition, there is an edge between y and z in H, and they are also in the same level L_{i-1}. Therefore the distance between y and z in T is 2, and consequently the distance between x and y is 3 in T. □

Combining Lemmas 7 and 8, we obtain the following result:

Theorem 4. *There is a polynomial-time algorithm that embeds a given graph G into a tree with distortion at most $27\,\alpha_G$.*

5 Dimensionality Reduction in ℓ_1

In this section, we prove that dimensionality reduction in ℓ_1, and indeed any ℓ_p space with $1 \leq p \leq 2$, is possible with ordinal embeddings of logarithmic dimension and relaxation $1+\varepsilon$. This result sharply contrasts metric embedding distortion, where any embedding of an ℓ_1 metric of distortion c requires $n^{\Omega(1/c^2)}$ dimensions in ℓ_1 [BC05,LN04].

Theorem 5. *Any ℓ_p metric with $1 \leq p \leq 2$ can be embedded into $O(\varepsilon^{-4} \lg n)$-dimensional ℓ_p space with ordinal relaxation $1+\varepsilon$, for any $\varepsilon > 0$ and positive integer p.*

Proof. First we take the $(p/2)$th power of the pairwise distances in the given ℓ_p metric D. The resulting metric D' is an ℓ_2 metric [Sch38,WW75]; see also [MN04]. Also, because $x \mapsto x^{p/2}$ is a monotone function, D' is an ordinal embedding of D (without relaxation). Next we apply Johnson-Lindenstrauss ℓ_2 dimensionality reduction [JL84] to obtain an $d = O((\log n)/\delta^2)$-dimensional ℓ_2 metric D'' with $1+\delta$ distortion relative to D'. Finally, we can embed this d-dimensional ℓ_2 metric into $O(d/\delta^2)$-dimensional ℓ_p space D''' with distortion $1+\delta$ relative to D'' [FLM77]; see also [Ind07,JS03]. Thus D''' is an $O((\log n)/\delta^4)$-dimensional ℓ_1 metric with distortion $(1+\delta)^2$ relative to D'.

We claim that D''' is an ordinal embedding of D with relaxation at most $1+\varepsilon$ for any desired $\varepsilon > 0$ and a suitable choice of δ. Suppose we have two distances $D[p,q]$ and $D[r,s]$ with $D[p,q]/D[r,s] \geq 1+\varepsilon$ for a desired $\varepsilon > 0$. Then $D'[p,q]/D'[r,s] = D'[p,q]^{2/p}/D'[r,s]^{2/p} = (D'[p,q]/D'[r,s])^{2/p} \geq (1+\varepsilon)^{2/p} \geq 1+(2/p)\varepsilon$. Thus, if we choose $\delta < \min\{\frac{2}{3}\varepsilon/p, 1\}$, then the distortion of D''' relative to D' is $(1+\delta)^2 \leq 1+3\delta < 1+(2/p)\varepsilon \leq D'[p,q]/D'[r,s]$, so the embedding preserves the order of distances $D'''[p,q] > D'''[r,s]$. Therefore the relaxation of D''' relative to D is at most $1+\varepsilon$ as desired. The dimension of the D''' embedding is $O((\log n)/\delta^4) = O((\log n)/\varepsilon^4)$. □

This approach is pleasingly simple in its use of prior results as black boxes. By more involved arguments, it may be possible to improve the dependence on ε in the number of dimensions by directly analyzing with a modification of Johnson-Lindenstrauss [JL84] and avoiding the use of [FLM77].

Acknowledgments

We thank Bo Brinkman for suggesting the approach for the proof in Section 5 and Noga Alon, Martin Farach-Colton, Piotr Indyk, and Assaf Naor for helpful discussions.

References

ABD$^+$. Noga Alon, Mihai Bădoiu, Erik D. Demaine, Martin Farach-Colton, MohammadTaghi Hajiaghayi, and Anastasios Sidiropoulos. Ordinal embeddings of minimum relaxation: General properties, trees, and ultrametrics. *ACM Transactions on Algorithms*. To appear.

AFR85. Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 277–280, Portland, Oregon, 1985.

BC05. Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in l_1. *Journal of the ACM*, 52(5):766–788 (electronic), 2005.

BDG+05. Mihai Bădoiu, Kedar Dhamdhere, Anupam Gupta, Yuri Rabinovich, Harald Raecke, R. Ravi, and Anastasios Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Vancouver, British Columbia, Canada, January 2005.

BIS07. Mihai Bădoiu, Piotr Indyk, and Anastasios Sidiropoulos. Approximation algorithms for embedding general metrics into trees. In *Proceedings of the 18th Symposium on Discrete Algorithms*, pages 512–521, January 2007.

BL04. Y. Bilu and N. Linial. Monotone maps, sphericity and bounded second eigenvalue. arXiv:math.CO/0401293, January 2004.

CS74. J. P. Cunningham and R. N. Shepard. Monotone mapping of similarities into a general metric space. *Journal of Mathematical Psychology*, 11:335–364, 1974.

CS98. B. Chor and M. Sudan. A geometric approach to betweennes. *SIAM Journal on Discrete Mathematics*, 11(4):511–523, 1998.

FLM77. T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139(1-2):53–94, 1977.

IM04. Piotr Indyk and Jiři Matoušek. Low-distortion embeddings of finite metric spaces. In J. E. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, chapter 8, pages 177–196. CRC Press, second edition, 2004.

Ind07. Piotr Indyk. Uncertainty principles, extractors, and explicit embeddings of l2 into l1. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 615–620, 2007.

JL84. W. B. Johnson and J. Lindenstrauss. Extensions of lipshitz mapping into hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

JS03. W. B. Johnson and G. Schechtman. Very tight embeddings of subspaces of L_p, $1 \leq p < 2$, into ℓ_p^n. *Geometric and Functional Analysis*, 13(4):845–851, 2003.

Kru64a. J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1–28, 1964.

Kru64b. J. B. Kruskal. Non-metric multidimensional scaling. *Psychometrika*, 29:115–129, 1964.

LN04. James R. Lee and Assaf Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geometric and Functional Analysis*, 14(4):745–747, 2004.

MN04. Manor Mendel and Assaf Naor. Euclidean quotients of finite metric spaces. *Advances in Mathematics*, 189(2):451–494, 2004.

Opa79. J. Opatrny. Total ordering problem. *SIAM J. Computing*, 8:111–114, 1979.

Sch38. I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536, 1938.

SFC04. R. Shah and M. Farach-Colton. On the complexity of ordinal clustering. *Journal of Classification*, 2004. To appear.

She62. R. N. Shepard. Multidimensional scaling with unknown distance function I. *Psychometrika*, 27:125–140, 1962.

Tor52. W. S. Torgerson. Multidimensional scaling I: Theory and method. *Psychometrika*, 17(4):401–414, 1952.

WW75. J. H. Wells and L. R. Williams. *Embeddings and extensions in analysis*. Springer-Verlag, New York, 1975. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84.