Metric embeddings with outliers

Anastasios Sidiropoulos∗ Yusu Wang†

Abstract

We initiate the study of metric embeddings with outliers. Given some metric space \((X, \rho)\) we wish to find a small set of outlier points \(K \subset X\) and either an isometric or a low-distortion embedding of \((X \setminus K, \rho)\) into some target metric space. This is a natural problem that captures scenarios where a small fraction of points in the input corresponds to noise.

For the case of isometric embeddings we derive polynomial-time approximation algorithms for minimizing the number of outliers when the target space is an ultrametric, a tree metric, or some constant-dimensional Euclidean space. The approximation factors are 3, 4 and 2, respectively. For the case of embedding into an ultrametric or tree metric, we further improve the running time to \(O(n^2)\) for an \(n\)-point input metric space, which is optimal. We complement these upper bounds by showing that outlier embedding into ultrametrics, trees, and \(d\)-dimensional Euclidean space for any \(d \geq 2\) are all NP-hard, as well as NP-hard to approximate within a factor better than 2 assuming the Unique Game Conjecture.

For the case of non-isometries we consider embeddings with small \(\ell_\infty\) distortion. We present polynomial-time bi-criteria approximation algorithms. Specifically, given some \(\varepsilon > 0\), let \(k_\varepsilon\) denote the minimum number of outliers required to obtain an embedding with distortion \(\varepsilon\). For the case of embedding into ultrametrics we obtain a polynomial-time algorithm which computes a set of at most \(3k_\varepsilon\) outliers and an embedding of the remaining points into an ultrametric with distortion \(O(\varepsilon \log n)\). Finally, for embedding a metric of unit diameter into constant-dimensional Euclidean space we present a polynomial-time algorithm which computes a set of at most \(2k_\varepsilon\) outliers and an embedding of the remaining points with distortion \(O(\sqrt{\varepsilon})\).

∗Dept. of Computer Science and Engineering and Dept. of Mathematics, The Ohio State University. Columbus, OH, USA. Supported by NSF grants CCF 1423230 and CAREER 1453472.

†Dept. of Computer Science and Engineering, The Ohio State University. Columbus, OH, USA. The work is partially supported by NSF under grant CCF-1319406.
1 Introduction

Metric embeddings provide a framework for addressing in a unified manner a variety of data-analytic tasks. Let $\mathcal{X} = (X, \rho), \mathcal{Y} = (Y, \rho')$ be metric spaces. At the high level, a metric embedding is a mapping $f : X \to Y$ that either is isometric or preserves the pairwise distances up to some small error called the distortion\(^1\). The corresponding computational problem is to decide whether an isometry $f$ exists or, more generally, to find a mapping $f$ with minimum distortion. The space $\mathcal{Y}$ might either be given or it might be constrained to be a member of a collection of spaces, such as trees, ultrametrics, and so on. The problems that can be geometrically abstracted using this language include phylogenetic reconstruction (e.g. via embeddings into trees \([1, 2, 11]\) or ultrametrics \([20, 4]\)), visualization (e.g. via embeddings into constant-dimensional Euclidean space \([8, 21, 10, 32, 19, 9, 17, 22]\)), and many more (for a more detailed exposition we refer the reader to \([26, 25]\)).

Despite extensive research on the above metric embedding paradigm, essentially nothing is known when the input space $\mathcal{X}$ can contain outliers. This scenario is of interest for example in applications where outliers can arise from measurement errors. Another example is when real-world data does not perfectly fit a model due to mathematical simplifications of physical processes.

We propose a generalization of the above high-level metric embedding problem which seeks to address such scenarios: Given $\mathcal{X}$ and $\mathcal{Y}$ we wish to find some small $K \subseteq X$ and either an isometric or low-distortion mapping $f : X \setminus K \to Y$. We refer to the points in $K$ as outliers.

We remark that it is easy to construct examples of spaces $\mathcal{X}, \mathcal{Y}$ where any embedding $f : X \to Y$ has arbitrarily large distortion (for any “reasonable” notion of distortion), yet there exists $x \in X$ and an isometry $f : X \setminus \{x\} \to Y$. Thus new ideas are needed to tackle the more general metric embedding problem in the presence of outliers.

1.1 Our contribution

Approximation algorithms. We focus on embeddings into ultrametrics, trees, and constant-dimensional Euclidean space. We first consider the problem of computing a minimum size set of outliers such that the remaining point-set admits an isometry into some target space. We refer to this task as the minimum outlier embedding problem.

Outlier embeddings into ultrametrics. It is well-known that a metric space is an ultrametric if and only if any 3-point subset is an ultrametric. We may therefore obtain a 3-approximation as follows: For all $(x, y, z) \in X^3$, if the triple $(x, y, z)$ is not an ultrametric then remove $x$, $y$, and $z$ from $X$. It is rather easy to see that this gives a 3-approximation for the minimum outlier embedding problem into ultrametrics (as for every triple of points that we remove, at least one of them must be an outlier in any optimal solution), with running time $O(n^3)$. By exploiting further structural properties of ultrametrics, we obtain a 3-approximation with running time $O(n^2)$. We remark that this running time is optimal since the input has size $\Theta(n^2)$ and it is straightforward to show that any 3-approximation has to read all the input (e.g. even to determine whether $\mathcal{X}$ is an ultrametric, which corresponds to the case where the minimum number of outliers is zero).

Outlier embeddings into trees. Similarly to the case of ultrametrics, it is known that a space is a tree metric if and only if any 4-point subset is a tree metric. This similarly leads to a 4-approximation algorithm in $O(n^4)$ time. We further improve the running time to $O(n^2)$, which is

---

\(^1\)Various definitions of distortion have been extensively considered, including multiplicative, additive, average, and $\ell_p$ distortion, as well as expected distortion when the map $f$ is random \([9]\).
also optimal. However, obtaining this improvement is significantly more complicated than the case of ultrametrics.

**Outlier embeddings into $$\mathbb{R}^d$$**. It is known that for any $$d \geq 1$$ any metric space admits an isometric embedding into $$d$$-dimensional Euclidean space if and only if any subset of size $$d + 3$$ does [33]. This immediately implies a $$(d + 3)$$-approximation algorithm for outlier embedding into $$d$$-dimensional Euclidean space with running time $$O(n^{d+3})$$, for any $$d \geq 1$$. Using additional rigidity properties of Euclidean space we obtain a 2-approximation with the same running time.

**Hardness of approximation.** We show that, assuming the Unique Games Conjecture [28], the problems of computing a minimum outlier embedding into ultrametrics, trees, and $$d$$-dimensional Euclidean space for any $$d \geq 2$$, are all NP-hard to approximate within a factor of $$2 - \nu$$, for any $$\nu > 0$$. These inapproximability results are obtained by combining reductions from Vertex Cover to minimum outlier embedding and the known hardness result for the former problem [29]. Note that for the case of embedding into $$d$$-dimensional Euclidean space for any $$d \geq 2$$ this inapproximability result matches our upper bound.

**Bi-criteria approximation algorithms.** We also consider non-isometric embeddings. All our results concern $$\ell_\infty$$ distortion. For some outlier set $$K \subseteq X$$, the $$\ell_\infty$$ distortion of some map $$f : X \setminus K \to Y$$ is defined to be

$$\sup_{x,y \in X \setminus K} |\rho(x,y) - \rho'(f(x), f(y))|.$$  

In this context there are two different objectives that we wish to minimize: the number of outliers and the distortion. For a compact metric space $$Z = (Z, \rho_Z)$$, denote by $$\Delta(Z) = \sup_{z,z' \in Z} \rho_Z(z,z')$$ the diameter of $$Z$$.

**Definition 1.1 (($$\varepsilon,k$$)-Outlier embedding).** We say that $$\mathcal{X}$$ admits a $$(\varepsilon,k)$$-outlier embedding into $$\mathcal{Y}$$ if there exists $$K \subseteq X$$ with $$|K| \leq k$$ and some $$f : X \setminus K \to Y$$ with $$\ell_\infty$$ distortion at most $$\varepsilon \Delta(\mathcal{X})$$. We refer to $$K$$ as the outlier set that witnesses a $$(\varepsilon,k)$$-outlier embedding of $$X$$.

Note that the multiplication of the distortion by $$\Delta(X)$$ is to make the parameter $$\varepsilon$$ scale-free. Since an isometry can be trivially achieved by removing all but one points; thus the above notion is well-defined for all $$\varepsilon > 0$$. We now state our main results concerning bi-criteria approximation:

**Bi-criteria outlier embeddings into ultrametrics**: We obtain a polynomial-time algorithm which given an $$n$$-point metric space $$\mathcal{X}$$ and some $$\varepsilon > 0$$ such that $$\mathcal{X}$$ admits a $$(\varepsilon,k)$$-outlier embedding into an ultrametric, outputs a $$O(\varepsilon \log n, 3k)$$-outlier embedding into an ultrametric.

**Bi-criteria outlier embeddings into $$\mathbb{R}^d$$**: We present an algorithm which given an $$n$$-point metric space $$\mathcal{X}$$ and some $$\varepsilon > 0$$ such that $$\mathcal{X}$$ admits a $$(\varepsilon,k)$$-outlier embedding in $$\mathbb{R}^d$$, outputs a $$(O(\sqrt{\varepsilon}), 2k)$$-outlier embedding of $$\mathcal{X}$$ into $$\mathbb{R}^d$$. The algorithm runs in time $$O(n^{d+3})$$.

**Bi-criteria outlier embeddings into trees**: Finally we mention that one can easily derive a bi-criteria approximation for outlier embedding into trees by the work of Gromov on $$\delta$$-hyperbolicity [23] (see also [14]). Formally, there exists a polynomial-time algorithm which given a metric space $$\mathcal{X}$$ and some $$\varepsilon > 0$$ such that $$\mathcal{X}$$ admits a $$(\varepsilon,k)$$-outlier embedding into a tree, outputs a $$(O(\varepsilon \log n), 4k)$$-outlier embedding into a tree. Let us briefly outline the proof of this result: $$\delta$$-hyperbolicity is a four-point condition such that any $$\delta$$-hyperbolic space admits an embedding into a tree with $$\ell_\infty$$ distortion $$O(\delta \log n)$$, and such an embedding can be computed in polynomial time. Any metric that admits an embedding into a tree with $$\ell_\infty$$ distortion $$\varepsilon$$ is $$O(\varepsilon)$$-hyperbolic. Thus by removing all
4-tuples of points that violate the $O(\varepsilon)$-hyperbolicity condition and applying the embedding from [23] we immediately obtain an $(O(\varepsilon \log n), 4k)$-outlier embedding into a tree. We omit the details.

1.2 Previous work

Over the recent years there has been a lot work on approximation algorithms for minimum distortion embeddings into several host spaces and under various notions of distortion. Perhaps the most well-studied case is that of multiplicative distortion. For this case, approximation algorithms and inapproximability results have been obtained for embedding into the line [34, 10], constant-dimensional Euclidean space [9, 17, 19, 32, 10], trees [11, 15], ultrametrics [4], and other graph-induced metrics [15]. We also mention that similar questions have been considered for the case of bijective embeddings [35, 24, 27, 19, 30]. Analogous questions have also been investigated for average [18], additive [5], $\ell_p$ [2], and $\ell_\infty$ distortion [20, 1].

Similar in spirit with the outlier embeddings introduced in this work is the notion of embeddings with slack [12, 13, 31]. In this scenario we are given a parameter $\varepsilon > 0$ and we wish to find an embedding that preserves $(1 - \varepsilon)$-fraction of all pairwise distances up to a certain distortion. We remark however that these mappings cannot in general be used to obtain outlier embeddings. This is because typically in an embedding with slack the pairwise distances that are distorted arbitrarily involve a large fraction of all points.

1.3 Discussion

Our work naturally leads to several directions for further research. Let us briefly discuss the most prominent ones.

An obvious direction is closing the gap between the approximation factors and the inapproximability results for embedding into ultrametrics and trees. Similarly, it is important to understand whether the running time of the 2-approximation for embedding into Euclidean space can be improved. More generally, an important direction is understanding the approximability of outlier embeddings into other host spaces, such as planar graphs and other graph-induced metrics.

In the context of bi-criteria outlier embeddings, another direction is to investigate different notions of distortion. The case of $\ell_\infty$ distortion studied here is a natural starting point since it is very sensitive to outliers. It seems promising to try to adapt existing approximation algorithms for $\ell_p$, multiplicative, and average distortion to the outlier case.

Finally, it is important to understand whether improved guarantees or matching hardness results for bi-criteria approximations are possible.

2 Definitions

A metric space is a pair $\mathcal{X} = (X, \rho)$ where $X$ is a set and $\rho : X \times X \to \mathbb{R}_{\geq 0}$ such that (i) for any $x, y \in X$, $\rho(x, y) = \rho(y, x) \geq 0$, (ii) $\rho(x, y) = 0$ if and only if $x = y$, and (iii) for any $x, y, z \in X$, $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$. Given two metric spaces $\mathcal{X} = (X, \rho_X)$ and $\mathcal{Y} = (Y, \rho_Y)$, an embedding of $\mathcal{X}$ into $\mathcal{Y}$ is simply a map $\varphi : X \to Y$, and $\varphi$ is an isometric embedding if for any $x, x' \in X$, $\rho_X(x, x') = \rho_Y(\varphi(x), \varphi(x'))$.

In this paper our input is an $n$-point metric $(X, \rho)$, meaning that $X$ is a discrete set of cardinality $n$. Given an $n$-point metric space $\mathcal{X} = (X, \rho)$ and a value $\alpha \geq 0$, we denote by $\mathcal{X} + \alpha$ the metric space $(X, \rho')$ where for any $x \neq y \in X$ we have $\rho'(x, y) = \rho(x, y) + \alpha$. 
**Definition 2.1** (Ultrametric space). A metric space \((X, \rho)\) is an ultrametric (tree) space if and only if the following three-point condition holds for any \(x, y, z \in X\):

\[
\rho(x, y) \leq \max\{\rho(x, z), \rho(z, y)\}.
\]  

**Definition 2.2** (Tree metric). A metric space \((X, \rho)\) is a tree metric if and only if the following four-point condition holds for any \(x, y, z, w \in X\):

\[
\rho(x, y) + \rho(z, w) \leq \max\{\rho(x, z) + \rho(y, w), \rho(x, w) + \rho(y, z)\}.
\]  

An equivalent formulation of the four-point condition is that for all \(x, y, z, w \in X\), the largest two quantities of the following three terms are equal:

\[
\rho(x, y), \ \rho(x, z) + \rho(y, w), \ \rho(x, w) + \rho(y, z).
\]  

In particular, an \(n\)-point tree metric \((X, \rho)\) can be realized by a weighted tree \(T\) such that there is a map \(\varphi : X \to V(T)\) into the set of nodes \(V(T)\) of \(T\), and that for any \(x, y \in X\), the shortest path distance \(d_T(\varphi(x), \varphi(y))\) in \(T\) equals \(\rho(x, y)\). In other words, \(\varphi\) is an isometric embedding of \((X, \rho)\) into the graphic tree metric \((T, d_T)\). An ultrametric \((X, \rho)\) is in fact a special case of tree metric, where there is an isometric embedding \(\varphi : X \to V(T)\) to a rooted tree \((T, d_T)\) such that \(\varphi(X)\) are leaves of \(T\) and all leaves are at equal distance from the root of \(T\).

### 3 Approximation algorithms for outlier embeddings

In this section we present approximation algorithms for the minimum outlier embedding problem for three types of target metric spaces: ultrametrics, tree metrics, and Euclidean metric spaces. We show in Appendix 3 that finding optimal solutions for each of these problems is NP-hard (and roughly speaking hard to approximate within a factor of 2 as well). In the cases of ultrametric and tree metrics, it is easy to approximate the minimum outlier embedding within constant factor in \(O(n^3)\) and \(O(n^4)\) time, respectively. The key challenge (especially for embedding into tree metric) is to improve the time complexity of the approximation algorithm to \(O(n^2)\), which is optimal.

#### 3.1 Approximating outlier embeddings into ultrametrics

**Theorem 3.1.** Given an \(n\)-point metric space \((X, \rho)\), there exists a 3-approximation algorithm for minimum outlier embedding into ultrametrics, with running time \(O(n^2)\).

**Proof.** We can obtain a polynomial-time 3-approximation algorithm as follows: For each triple of points \((x, y, z) \in X^3\), considered in some arbitrary order, check whether it satisfies (1). If not, then remove \(x, y, \) and \(z\) from \(X\) and continue with the remaining triples. Let \(K\) be the set of removed points. For every triple of points removed, at least one must be in any optimal solution; therefore the resulting solution \(K\) is a 3-approximation. The running time of this method is \(O(n^3)\). We next show how to improve the running time to \(O(n^2)\).

Let \(X = \{x_1, \ldots, x_n\}\). We inductively compute a sequence \(X_0, \ldots, X_n \subseteq X\), where \(X_0\) is set to be \(X_0 = \emptyset\). Given \(X_{i-1}\) for some \(i \in \{1, \ldots, n\}\), assuming the invariance that \(X_{i-1}\) is an ultrametric, we compute \(X_i\) as follows. We check whether \((X_{i-1} \cup \{x_i\}, \rho)\) is an ultrametric. If it is, then we set \(X_i = X_{i-1} \cup \{x_i\}\). Otherwise, there must exist \((x, y, z) \in (X_{i-1} \cup \{x_i\})^3\) that violates (1). Since
\((X_{i-1}, \rho)\) is an ultrametric, it follows that every such triple must contain \(x_i\). Therefore it suffices to show how to quickly find \(y, z \in X_{i-1}\) such that \((x_i, y, z)\) violates (I), if they exist. To this end, let \(x_i^*\) be a nearest neighbor of \(x_i\) in \(X_{i-1}\), that is \(x_i^* = \arg\min_{x \in X_{i-1}} \{\rho(x, x_i)\}\), where we break ties arbitrarily. Instead of checking \(x_i\) against all possible \(y, z\) from \(X_{i-1}\), we claim that (I) holds for all \((x_i, y, z)\) with \(y, z \in X_{i-1}\) if and only if for all \(w \in X_{i-1}\) we have

\[
\begin{align*}
(i) \ & \rho(x_i, x_i^*) \leq \max\{\rho(x_i, w), \rho(x_i^*, w)\} \quad \text{and} \quad (ii) \ & \rho(x_i, w) = \rho(x_i^*, w). \quad (4)
\end{align*}
\]

Indeed, assume that (i) and (ii) above hold for all \(w \in X_{i-1}\), yet there exist some \(y, z \in X_{i-1}\) such that \((x_i, y, z)\) violates (I), say w.l.o.g., \(\rho(x_i, y) > \max\{\rho(x_i, z), \rho(y, z)\}\). Then by (ii) above, we have \(\rho(x_i, y) = \rho(x_i^*, y)\) and \(\rho(x_i, z) = \rho(x_i^*, z)\), implying that \(\rho(x_i^*, y) > \max\{\rho(x_i^*, z), \rho(y, z)\}\). Hence \((x_i^*, y, z)\) also violates (I), contradicting the fact that \((X_{i-1}, \rho)\) is an ultrametric. Hence no such \(y, z\) can exist, and (I) is sufficient to check whether \(X_{i-1} \cup \{x_i\}\) induces an ultrametric or not.

Finally, we can clearly check in time \(O(n)\) whether both conditions in (10) hold for all \(w \in X_{i-1}\). If either (i) or (ii) in (10) fails then \((x_i, x_i^*, w)\) violates (I), which concludes the proof. \(\square\)

### 3.2 Approximating outlier embeddings into trees

We now present a 4-approximation algorithm for embedding a given \(n\)-point metric space into a tree metric with a minimum number of outliers. Using the four-point condition (2) in Definition 2.2, it is fairly simple to obtain a 4-approximation algorithm for the problem with running time \(O(n^2)\) as follows: Check all 4-tuples of points \(x, y, z, w \in X\). If the 4-tuple violates the four-point condition, then remove \(x, y, z, w\) from \(X\). It is immediate that for any such 4-tuple, at least one of its points much be an outlier in any optimal solution. It follows that the result is a 4-approximation.

We next show how to implement this approach in time \(O(n^2)\). The main technical difficult is in finding a set of violating 4-tuples quickly. The high-level description of the algorithm is rather simple, and is as follows. Let \((X, \rho)\) be the input metric space where \(X = \{x_1, \ldots, x_n\}\). Set \(X_1 = \{x_1\}\). For any \(i = 2, \ldots, n\), we inductively define \(X_i \subseteq X\). At the beginning of the \(i\)-th iteration, we maintain the invariance that \((X_{i-1}, \rho)\) is a tree metric. If \((X_{i-1} \cup \{x_i\}, \rho)\) is a tree metric, then we set \(X_i = X_{i-1} \cup \{x_i\}\). Otherwise there must exist \(y_i, z_i, w_i \in X_{i-1}\) such that the 4-tuple \((x_i, y_i, z_i, w_i)\) violates the four-point condition; we set \(X_i = X_{i-1} \setminus \{y_i, z_i, w_i\}\).

To implement this idea in \(O(n^2)\) time, it suffices to show that for any \(i = 2, \ldots, n\), given \(X_{i-1}\), we can compute \(X_i\) in time \(O(n)\). The algorithm will inductively compute a collection of edge-weighted trees \(T_1, \ldots, T_n\), with \(T_1\) simply being the graph with \(V(T_1) = \{x_1\}\), and maintain the following invariants for each \(i \in [1, n]\):

\[(I-1) \quad X_i \subseteq V(T_i)\) and all leaves of \(T_i\) are in \(X_i\). \((X_i, \rho)\) embeds isometrically into \((T_i, d_{T_i})\); that is, the shortest-path metric of \(T_i\) agrees with \(\rho\) on \(X_i\); for any \(x, y \in X_i\), \(d_{T_i}(x, y) = \rho(x, y)\).

\[(I-2) \quad \text{At the \(i\)-th iteration either } X_i = X_{i-1} \cup \{x_i\} \text{ or } X_i = X_{i-1} \setminus \{y_i, z_i, w_i\} \text{ where the 4-tuple } \{x_i, y_i, z_i, w_i\} \text{ violates the four-point condition under metric } \rho.\]

**Definition 3.2 (Leaf augmentation).** Given \(X' \subseteq X\), let \(T\) be a tree with \(V(T) \cap X = X'\). Given \(a \in X \setminus X'\) and \(u, v \in X'\), the \(a\)-leaf augmentation of \(T\) at \(\{u, v\}\) is the tree \(T'\) obtained as follows. Let \(P\) be the path in \(T\) between \(u\) and \(v\) (which may contain a single vertex if \(u = v\)). Set \(r = (\rho(a, u) + \rho(a, v) - \rho(u, v))/2\). Let \(s\) be a vertex in \(P\) with \(d_P(s, u) = \rho(a, u) - r\); if no such vertex exists then we introduce a new such vertex \(s\) by subdividing the appropriate edge in \(P\)
and update the edge lengths accordingly. In the resulting tree $T'$ we add the vertex $a$ and the edge \( \{a, s\} \) if they do not already exist, and we set the length of \( \{a, s\} \) to be $r$. We call $s$ the stem of $a$ (w.r.t. the leaf augmentation). When $u = v$ we say that $T'$ is the $a$-leaf augmentation of $T$ at $u$, in which case $T'$ is obtained from $T$ simply by adding $a$ as a leaf attached to $u$, and $u$ is the stem of $a$.

In what follows, we set $x_i^*$ to be the nearest neighbor of $x_i$ in $X_{i-1}$, that is

$$x_i^* = \arg\min_{x \in X_{i-1}} \rho(x, x_i),$$

where we break ties arbitrarily. Intuitively, if we can build a new tree $T'$ from $T_{i-1}$ so that \( (X_{i-1} \cup \{x_i\}, \rho) \) can be isometrically embedded in $T'$, then $T'$ is a $x_i$-leaf augmentation of $T_{i-1}$ at some pair \( \{x_i^*, u\} \). Our approach will first compute an auxiliary structure, called $x_i$-orientation on $T_{i-1}$, to help us identify a potential leaf augmentation. We next check for the validity of this leaf augmentation. The key is to produce this candidate leaf augmentation such that if it is not valid, then we will be able to find a 4-tuple violating the four-point condition from it quickly.

**Definition 3.3 ((a, u, v)-orientation).** Let $X' \subseteq X$ and let $T$ be a tree with $V(T) \cap X = X'$. Let $a \in X \setminus X'$ and $u, v \in X'$. The (a, u, v)-orientation of $T$ is a partially oriented tree $\tilde{T}$ obtained as follows: Let $T'$ be the $a$-leaf augmentation of $T$ at \( \{u, v\} \), and let $a'$ be the stem of $a$. We orient every edge in $P_{uv} \cap E(T')$ towards $a'$, where $P_{uv}$ is the unique path in $T$ between $u$ and $v$. All other edges in $E(T') \setminus P_{uv}$ remain unoriented.

If $E(T) \neq E(T')$ then there exists a unique edge in $e \in P_{uv} \setminus E(T')$ (which is subdivided in $T'$); this edge $e$ remains undirected in $\tilde{T}$. We call this edge the sink edge w.r.t. \( \{u, v\} \). If there is no sink edge, then there is a unique vertex in $P_{uv}$ with no outgoing edges in $P_{uv}$, which we call the sink vertex w.r.t. \( \{u, v\} \). Note that the sink is the simplex of smallest dimension that contains the stem of a w.r.t. the leaf augmentation at \( \{u, v\} \).

See the right figure for an example: where $s$ is stem of $a$ in the leaf augmentation at \( \{u, v\} \). The thick path $P_{uv}$ is oriented, other than the sink edge (the one that contains the stem $s$).

**Definition 3.4 ($x_i$-orientation).** An $x_i$-orientation of $T_{i-1}$ is any partial orientation $\tilde{T}_{i-1}$ of $T_{i-1}$ obtained via the following procedure: Consider any ordering of $X_{i-1}$, say \( \{v_1, v_2, \ldots, v_{\ell}\} = X_{i-1} \). Start with $\tilde{T}_{i-1} = T_{i-1}$, i.e. all edges in $\tilde{T}_{i-1}$ are initialized as undirected and we will iteratively modify their orientation. Process vertices $v_1, \ldots, v_{\ell}$ in this order. For each $v_j$, denote by $P_j$ the path in $T_{i-1}$ between $v_j$ and $x_i^*$. Traverse $P_j$ starting from $v_j$ until we reach either $x_i^*$ or an edge which is already visited. For each unoriented edge we visit, we set its orientation to be the one in the ($x_i, v_j, x_i^*$)-orientation of $T_{i-1}$. An edge that is visited in the above process is called masked.

Since the above procedure is performed for all leaves of $T_{i-1}$, an $x_i$-orientation will mask all edges. However, a masked edge may not be oriented, in which case this edge must be the sink edge w.r.t. \( \{v_j, x_i^*\} \) for some $v_j \in X_{i-1}$.

**Definition 3.5 (Sinks).** Given an $x_i$-orientation $\tilde{T}$ of tree $T$, a sink is either an un-oriented edge, or a vertex $v \in V(T)$ such that all incident edges have an orientation toward $v$. The former is also called a sink edge w.r.t. $\tilde{T}$ and the latter a sink vertex w.r.t. $\tilde{T}$.

It can be shown that each sink edge/vertex must be a sink edge/vertex w.r.t. \( \{v_j, x_i^*\} \) for some $v_j \in X_{i-1}$, and we call $v_j$ a generating vertex for this sink.
An \(x_i\)-orientation may have multiple sinks. We further augment the \(x_i\)-orientation to record a generating vertex \(v_j\) for every sink (there may be multiple choices of \(v_j\) for a single sink, and we can take an arbitrary one). We also remark that a sink w.r.t. some \(\{v_j, x^*_i\}\) may not ultimately be a sink for the global \(x_i\)-orientation: see the right figure for an example, where \(s\) is a sink vertex w.r.t. \(\{v_j, x^*_i\}\), but not a sink vertex for the global \(x_i\)-orientation.

The proofs of the following two results can be found in Appendix A.

**Lemma 3.6.** An \(x_i\)-orientation of \(T_{i-1}\) (together with a generating vertex for each sink) can be computed in \(O(i)\) time.

**Lemma 3.7.** Any \(x_i\)-orientation \(T_{i-1}'\) of \(T_{i-1}\) has at least one sink.

![Figure 1](image)

Figure 1: (a) An illustration for Case 1 in Lemma 3.8. Solid edges are from \(T_{i-1}\) and the dotted edge connecting \(x_i\) to its stem \(s\) is in \(L\). (b) Case 2 when \(y = s\). (c) If \(\rho(x_i, w) > d_L(x_i, w)\), then we can find \(z\) the processing of which during the construction of \(T_{i-1}'\) assigns the orientation of edge \((y, y')\).

**Lemma 3.8.** For any \(i \in \{2, \ldots, n\}\), given \(T_{i-1}\), we can compute \(T_i\) and \(X_i\) satisfying invariants (I-1) and (I-2) in \(O(i)\) time.

**Proof.** It suffices to show that in \(O(i)\) time we can either find a 4-tuple of points in \(X_{i-1} \cup \{x_i\}\), that violates the four-point condition, or we can compute a tree having a shortest-path metric that agrees with \(\rho\) on \(X_{i-1} \cup \{x_i\}\). By Lemma 3.6 we can compute an \(x_i\)-orientation \(T_{i-1}'\) of \(T_{i-1}\) in \(O(i)\) time. Consider any sink of \(T_{i-1}'\) (whose existence is guaranteed by Lemma 3.7), and let \(u\) be its associated generating vertex; \(u\) must be in \(X_{i-1}\). Let \(L\) be the \(x_i\)-leaf augmentation of \(T_{i-1}\) at \(\{x_i^*, u\}\), and let \(d_L\) denote the shortest path metric on the tree \(L\).

Since \(L\) is the \(x_i\)-leaf augmentation of \(T\), we have for all \(w, w' \in X_{i-1}\), \(d_L(w, w') = d_{T_{i-1}}(w, w') = \rho(w, w')\) (the last quality is because \((X_{i-1}, \rho)\) embeds isometrically into \(T_{i-1}\)). Thus \(d_L\) may only disagree with \(\rho\) on pairs of points \(x_i, v\), for some \(v \in X_{i-1}\). We check in \(O(i)\) time if, for all \(v \in X_{i-1}\), we have \(d_L(x_i, v) = \rho(x_i, v)\) via a traversal of \(L\) starting from the stem of \(x_i\) in \(L\). If the above holds, then obviously \((X_{i-1} \cup \{x_i\}, \rho)\) embeds isometrically into \(L\). We then set \(X_i = X_{i-1} \cup \{x_i\}\) and output \(T_i = L\). Otherwise, let \(w \in X_{i-1}\) be such that \(d_L(x_i, w) \neq \rho(x_i, w)\). We now show that we can find a 4-tuple including \(x_i\) that violates the four-point condition in constant time.

Let \(s\) be the stem of \(x_i\) in \(L\). Consider \(L\) as rooted at \(x_i^*\) and let \(y \in V(L)\) be the lowest common ancestor of \(u\) and \(w\). Note that \(y\) must be a vertex from \(V(T_{i-1})\) too. Let \(P_{xy}\) denote the unique path in \(L\) between any two \(v\) and \(v'\). The vertex \(y\) must be in the path \(P_{ux_i^*}\).
Case 1: \( y \neq s \). In this case, \( y \) is either in the interior of path \( P_{su} \) or of path \( P_{sx_i^*} \). Assume w.o.l.g. that \( y \) is in the interior of \( P_{su} \); the handling of the other case is completely symmetric. See Figure 1(a) for an illustration. Since \( d_L \) is a tree metric, we know that the 4-tuple \( \{x_i, x_i^*, u, w\} \) should satisfy the four-point condition under the metric \( d_L \). Using the alternative formulation of four-point condition in Definition 2.2, we have that the largest two quantities of the following three terms should be equal:

\[
d_L(x_i, x_i^*) + d_L(u, w), \quad d_L(x_i, u) + d_L(x_i^*, w), \quad d_L(x_i, w) + d_L(x_i^*, u).
\]

(5)

For this specific configuration of \( y \), we further have:

\[
d_L(x_i, x_i^*) + d_L(u, w) < d_L(x_i, u) + d_L(x_i^*, w) = d_L(x_i, w) + d_L(x_i^*, u).
\]

(6)

On the other hand, by construction, we know that \( d_L \) agrees with \( \rho \) on \( X_{i-1} \). Furthermore, since \( L \) is the \( x_i \)-leaf augmentation of \( T_{i-1} \) at \( \{u, x_i^*\} \), we have that \( d_L(x_i, u) = \rho(x_i, u) \) and \( d_L(x_i, x_i^*) = \rho(x_i, x_i^*) \). Hence [6] can be rewritten as

\[
\rho(x_i, x_i^*) + \rho(u, w) < \rho(x_i, u) + \rho(x_i^*, w) = d_L(x_i, w) + \rho(x_i^*, u).
\]

(7)

If \( \rho(x_i, w) \neq d_L(x_i, w) \), then the largest two quantities of

\[
\rho(x_i, x_i^*) + \rho(u, w), \rho(x_i, u) + \rho(x_i^*, w), \rho(x_i, w) + \rho(x_i^*, u)
\]

can no longer be equal as \( \rho(x_i, x_i^*) + \rho(u, w) < \rho(x_i, u) + \rho(x_i^*, w) \). Hence the 4-tuple \( \{x_i, x_i^*, u, w\} \) violates the four-point condition under the metric \( \rho \) (by using [5]).

Case 2: \( y = s \), in which case \( s \) must be a sink vertex: see Figure 1(b) for an illustration. For this configuration of \( y \), it is necessary that

\[
\rho(x_i, x_i^*) + \rho(u, w) = \rho(x_i, u) + \rho(x_i^*, w) = d_L(x_i, w) + \rho(x_i^*, u).
\]

(8)

Hence if \( \rho(x_i, w) > d_L(x_i, w) \), then the 4-tuple \( \{x_i, x_i^*, u, w\} \) violates the four-point condition under the metric \( \rho \) because

\[
\rho(x_i, x_i^*) + \rho(u, w) = \rho(x_i, u) + \rho(x_i^*, w) < \rho(x_i, w) + \rho(x_i^*, u).
\]

What remains is to find an violating 4-tuple for the case when \( \rho(x_i, w) < d_L(x_i, w) \).

Now imagine performing the \( x_i \)-leaf augmentation of \( T_{i-1} \) at \( \{w, x_i^*\} \). We first argue that the stem \( s' \) of \( x_i \) w.r.t. \( \{w, x_i^*\} \) necessarily lies in \( P_{wy} \) in \( T_{i-1} \). Let \( r_w = (\rho(x_i, x_i^*) + \rho(x_i, w) - \rho(x_i^*, w))/2 \). In the augmented tree \( L \), \( d_L(x_i, y) = (\rho(x_i, x_i^*) + \rho(x_i, u) - \rho(x_i^*, u))/2 \). Combing [8] and \( \rho(x_i, w) < d_L(x_i, w) \) we have that \( r_w < d_L(x_i, y) \). On the other hand, following Definition 3.2 the position of \( s' \) is such that \( d_{T_{i-1}}(x_i^*, s') = \rho(x_i^*, x_i) - r_w \), while the position of \( y = s \) was that \( d_{T_{i-1}}(x_i^*, s) = \rho(x_i^*, x_i) - d_L(x_i, y) \). It then follows that \( s' \) must lie in the interior of path \( P_{wy} \).

Since the stem of \( x_i \) w.r.t. \( \{w, x_i^*\} \) is in \( P_{wy} \), it means that before we process \( w \) in the construction of the \( x_i \)-orientation \( T_{i-1} \), there must exist some other leaf \( z \in V(T_{i-1}) \) such that the process of \( z \) assigns the orientation of the edge \( (y, y') \in P_{wy} \) to be towards \( y \); See Figure 1(c). This is because if no such \( z \) exists, then while processing \( w \), we would have
Theorem 3.11. The metric space $d_{\mathbb{R}^k}$ is simply an isometric embedding

|Section| Theorem 3.9. There exists a 4-approximation algorithm for minimum outlier embedding into trees, with running time $O(n^2)$. |
|---|---|
|Proof. By Lemma 3.8 and induction on $i = 1, \ldots, n$, it follows immediately that we can compute $T_n$ in time $O(n^2)$. By invariant (I-1), the output $(X_n, \rho)$ is a tree metric as it can be isometrically embedded into $T_n$. Furthermore, by invariant (I-2), each 4-tuple of points we removed forms a violation of the four-point condition, and thus must contain at least one point from any optimal outlier set. As such, the total number of points we removed can be at most four times the size of the optimal solution. Hence our algorithm is a 4-approximation as claimed. |
Furthermore, given an \( m \)-point metric space \((X, \rho)\), it is known that one can decide whether \((X, \rho)\) is embeddable in some Euclidean space by checking whether a certain \( m \times m \) matrix derived from the distance matrix \( \rho \) is positive semi-definite, and the rank of this matrix gives the embedding dimension of \( X \); see e.g. \cite{36}.

Following Theorem 3.11, one can easily come up with a \((d+3)\)-approximation algorithm for minimum outlier embedding into \( \mathbb{R}^d \), by simply checking whether each \((d+3)\)-tuple of points is \( d \)-embeddable, and if not, removing all these \( d+3 \) points. Our main result below is an 2-approximation algorithm within the same running time. In particular, Algorithm 1 satisfies the requirements of Theorem 3.12 and the proof is in Appendix A.

**Theorem 3.12.** Given an \( n \)-point metric space \((X, \rho)\), for any \( d \geq 1 \), there exists 2-approximation algorithm for minimum outlier embedding into \( \mathbb{R}^d \), with running time \( O(n d^3) \).

**Algorithm 1** 2-approximation outlier embedding in \( \mathbb{R}^d \).

**Input:** An \( n \)-point metric space \((X, \rho)\)

**Output:** A set of outliers \( \hat{K} \)

1. Initialize the set of candidate outlier sets \( C \) to be empty.

   **(Step-0)** For each \( d + 1 \) number of distinct points \( Y_d = \{y_0, \ldots, y_d\} \subset X \), perform the following:
   - Initialize sets \( Z \) and \( K \) to be the empty set.
   - **(Step-1)** Check whether \((Y_d, \rho)\) is \( d \)-embeddable in \( \mathbb{R}^d \): If not, return to (Step-0). Otherwise, compute its embedding dimension \( d' \); note, \( d' \leq d \).
   - **(Step-2)** For each remaining point \( x \in X \setminus Y_d \), check whether \((Y_d \cup \{s\}, \rho)\) is \( d' \)-embeddable: If yes, insert \( x \) to \( Z \); otherwise, insert \( x \) to the outlier set \( K \).
   - **(Step-3)** Construct a graph \( G = (Z, E) \), where \((z, z') \in E\) if \((Y_d \cup \{z, z\}', \rho)\) is not \( d' \)-embeddable. Compute a 2-approximation \( Z' \subset Z \) of the vertex cover of \( G \). Set \( K = K \cup Z' \), and we add the set \( K \) to the collection of candidate outlier sets \( C \).
   - **(Step-4)** Let \( \hat{K} \) be the set from \( C \) with smallest cardinality. We return \( \hat{K} \) as the outlier set.

**Hardness results.** In Appendix \[B\] we show that the minimum outlier embedding problems into ultrametrics, trees and Euclidean space are all NP-hard, by reducing the Vertex Cover problem to them in each case. In fact, assuming the unique game conjecture, it is NP-hard to approximate each of them within \( 2 - \nu \), for any positive \( \nu \). For the case of minimum outlier embedding into Euclidean space, we note that our 2-approximation algorithm above matches the hardness result.

## 4 Bi-criteria approximation algorithms

### 4.1 Bi-criteria approximation for embedding into ultrametrics

Let \( T = (X, E) \) be a tree with non-negative edge weights. The ultrametric induced by \( T \) is the ultrametric \( U = (X, \delta) \) where for every \( x, y \in X \) we have that \( \delta(x, y) \) is equal to the maximum weight of the edges in the unique \( x \)-\( y \) path in \( T \) (it is easy to verify that the metric constructed as such is indeed an ultrametric \[20\]). Given an metric space \( X = (X, \rho) \), we can view it as a weighted graph and talk about its minimum spanning tree (MST). The following result is from \[20\].
**Theorem 4.2.** Concludes the proof.

**Theorem 4.3.** Given an n-point metric space $X = (X, \rho)$, let $X'$ be an ultrametric minimizing $\|X - U^\star\|_\infty$. Let $\hat{U}$ be ultrametric induced by a MST of $X$. Then there exists $\alpha \geq 0$, such that $U^\star = \hat{U} + \alpha$.

In particular, let $\beta = \|\hat{U} - X\|_\infty$. Then $U^\star = \hat{U} + \beta/2$. This further implies that $\|U^\star - X\|_\infty = \|\hat{U} - X\|_\infty/2 = \beta/2$.

**Theorem 4.2.** There exists a polynomial-time algorithm which given an n-point metric space $X = (X, \rho)$, $\varepsilon \geq 0$, and $k \geq 0$, such that $X$ admits a $(\varepsilon, k)$-outlier embedding into an ultrametric, outputs an $(O(\varepsilon \log(n)), 3k)$-outlier embedding into an ultrametric.

**Proof.** For simplicity, assume that the diameter $\Delta(X) = 1$. The algorithm is as follows. We first enumerate all triples $\{x, y, z\} \in \binom{X}{3}$. For any such triple, if $\rho(x, y) > \max\{\rho(x, z), \rho(z, y)\} + 2\varepsilon$, then we remove $x$, $y$, and $z$ from $X$. Let $X'$ be the resulting point set. We output the ultrametric $U'$ induced by an MST of the metric space $(X', \rho)$. This completes the description of the algorithm.

It suffices to prove that the output is indeed an $(O(\varepsilon \log(n)), 3k)$-outlier embedding. Let $K^* \subseteq X$, with $|K^*| \leq k$, be such that $(X \setminus K^*, \rho)$ admits a $(\varepsilon, 0)$-outlier embedding into an ultrametric. Let $\{x, y, z\} \in \binom{X}{3}$ be such that $\rho(x, y) > \max\{\rho(x, z), \rho(z, y)\} + 2\varepsilon$. It follows by Lemma 4.1 that $(\{x, y, z\}, \rho)$ does not admit a $(\varepsilon, 0)$-outlier embedding into an ultrametric. Thus, $K^* \cap \{x, y, z\} \neq \emptyset$. It follows that $|X \setminus X'| \leq 3 \cdot |K^*| \leq 3k$. In other words, the algorithm removes at most $3k$ points.

It remains to bound the distortion between $U'$ and $(X', \rho)$. Let $T$ be the MST of $(X', \rho)$ such that the algorithm outputs the ultrametric $U' = (X', \delta)$ induced by $T$. We will prove by induction on $i \geq 0$, that for all $x, y \in X'$, if the $x$-$y$ path in $T$ contains at most $2^i$ edges, then $\rho(x, y) \leq \delta(x, y) + 2i \cdot \varepsilon$.

For the base case $i = 0$ we have that $\{x, y\} \in E(T)$. Since $T$ is the minimum spanning tree of $(X', \rho)$, it follows that $\delta(x, y) = \rho(x, y)$, proving the base case. For the inductive step, let $x, y \in X'$ such that the $x$-$y$ path contains at most $2^i$ edges, for some $i \geq 1$. Let $w \in X'$ be such that $w$ is in the $x$-$y$ path in $T$, and moreover the $x$-$w$ and $w$-$z$ paths in $T$ have at most $2^{i-1}$ edges each. Since $\{x, y, w\} \in X'$, it follows that the triple $\{x, y, w\}$ was not removed by the algorithm, and thus

$$\rho(x, y) \leq \max\{\rho(x, w), \rho(w, y)\} + 2\varepsilon. \tag{9}$$

By the inductive hypothesis we have

$$\rho(x, w) \leq \delta(x, w) + 2(i - 1) \cdot \varepsilon, \quad \text{and} \quad \rho(w, y) \leq \delta(w, y) + 2(i - 1) \cdot \varepsilon. \tag{10}$$

By (9) and (10) we get

$$\rho(x, y) \leq \max\{\delta(x, w), \delta(w, y)\} + 2(i - 1) \cdot \varepsilon + 2\varepsilon \leq \delta(x, y) + 2i \cdot \varepsilon.$$

Hence $U'$ is an $(2\varepsilon \log(n), 0)$-embedding of $(X', \rho)$. This, together with the bound on $|X \setminus X'|$ concludes the proof.

### 4.2 Bi-criteria approximation for embedding into $\mathbb{R}^d$

**Theorem 4.3.** Given an n-point metric space $X = (X, \rho)$, if $X$ admits a $(\varepsilon, k^\star)$-outlier embedding in $\mathbb{R}^d$, then we have an algorithm that outputs an $(O(\sqrt{\varepsilon}), 2k^\star)$-outlier embedding of $X$ in $\mathbb{R}^d$ in $O(n^{d+3})$ time. Here the big-O notation hides constants depending on the dimension $d$.

We only state the main result here, and the details can be found in Appendix C. In particular, the high level structure of our algorithm parallels that of Algorithm 1 which intuitively is an algorithm for the special case when $\varepsilon = 0$. However, the technical details here are much more involved so as to tackle several issues caused by the near-isometric embedding.
References

[1] Richa Agarwala, Vineet Bafna, Martin Farach, Mike Paterson, and Mikkel Thorup. On the approximability of numerical taxonomy (fitting distances by tree metrics). *SIAM Journal on Computing*, 28(3):1073–1085, 1998.

[2] Nir Ailon and Moses Charikar. Fitting tree metrics: Hierarchical clustering and phylogeny. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 73–82. IEEE, 2005.

[3] P. Alestalo, D. A. Trotsenko, and J. Väisälä. Isometric approximation. *Israel Journal of Mathematics*, 125(1):61–82, 2001.

[4] 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 (TALG)*, 4(4):46, 2008.

[5] Mihai Bădoiu. Approximation algorithm for embedding metrics into a two-dimensional space. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 434–443. Society for Industrial and Applied Mathematics, 2003.

[6] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 184–193. IEEE, 1996.

[7] Leonard M. Blumenthal. *Theory and applications of distance geometry*. Chelsea Publishing Company, 2 edition.

[8] Mihai Bădoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Low-distortion embeddings of general metrics into the line. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 225–233. ACM, 2005.

[9] Mihai Bădoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Embedding ultrametrics into low-dimensional spaces. In *Proceedings of the 22nd ACM Symposium on Computational Geometry, Sedona, Arizona, USA, June 5-7, 2006*, pages 187–196, 2006.

[10] Mihai Bădoiu, Kedar Dhamdhere, Anupam Gupta, Yuri Rabinovich, Harald Räcke, R Ravi, and Anastasios Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 119–128. Society for Industrial and Applied Mathematics, 2005.

[11] Mihai Bădoiu, Piotr Indyk, and Anastasios Sidiropoulos. Approximation algorithms for embedding general metrics into trees. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 512–521. Society for Industrial and Applied Mathematics, 2007.

[12] T-H Hubert Chan, Kedar Dhamdhere, Anupam Gupta, Jon Kleinberg, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. *SIAM Journal on Computing*, 38(6):2303–2329, 2009.
[13] T-H Hubert Chan, Michael Dinitz, and Anupam Gupta. Spanners with slack. In *Algorithms–ESA 2006*, pages 196–207. Springer, 2006.

[14] Victor Chepoi, Feodor Dragan, Bertrand Estellon, Michel Habib, and Yann Vaxès. Diameters, centers, and approximating trees of delta-hyperbolic geodesic spaces and graphs. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 59–68. ACM, 2008.

[15] Victor Chepoi, Feodor F Dragan, Ilan Newman, Yuri Rabinovich, and Yann Vaxes. Constant approximation algorithms for embedding graph metrics into trees and outerplanar graphs. *Discrete & Computational Geometry*, 47(1):187–214, 2012.

[16] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. 3 edition.

[17] Mark de Berg, Krzysztof Onak, and Anastasios Sidiropoulos. Fat polygonal partitions with applications to visualization and embeddings. *JoCG*, 4(1):212–239, 2013.

[18] Kedar Dhamdhere, Anupam Gupta, and R Ravi. Approximation algorithms for minimizing average distortion. *Theory of Computing Systems*, 39(1):93–111, 2006.

[19] Jeff Edmonds, Anastasios Sidiropoulos, and Anastasios Zouzias. Inapproximability for planar embedding problems. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 222–235. Society for Industrial and Applied Mathematics, 2010.

[20] Martin Farach, Sampath Kannan, and Tandy Warnow. A robust model for finding optimal evolutionary trees. *Algorithmica*, 13(1-2):155–179, 1995.

[21] Michael Fellows, Fedor Fomin, Daniel Lokshtanov, Elena Losievskaja, Frances A Rosamond, and Saket Saurabh. Parameterized low-distortion embeddings-graph metrics into lines and trees. *arXiv preprint arXiv:0804.3028*, 2008.

[22] Michael R Fellows, Fedor V Fomin, Daniel Lokshtanov, Elena Losievskaja, Frances A Rosamond, and Saket Saurabh. Distortion is fixed parameter tractable. In *Automata, Languages and Programming*, pages 463–474. Springer, 2009.

[23] Mikhael Gromov. *Hyperbolic groups*. Springer, 1987.

[24] Alexander Hall and Christos Papadimitriou. Approximating the distortion. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 111–122. Springer, 2005.

[25] Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *focs*, page 10. IEEE, 2001.

[26] Piotr Indyk and Jiri Matoušek. Low-distortion embeddings of finite metric spaces. *Handbook of Discrete and Computational Geometry*, page 177, 2004.

[27] Claire Kenyon, Yuval Rabani, and Alistair Sinclair. Low distortion maps between point sets. *SIAM Journal on Computing*, 39(4):1617–1636, 2009.
[28] Subhash Khot. On the power of unique 2-prover 1-round games. In Proceedings of the thiry-fourth annual ACM symposium on Theory of computing, pages 767–775. ACM, 2002.

[29] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2 - \epsilon$. Journal of Computer and System Sciences, 74(3):335 – 349, 2008. Conference version appeared in Computational Complexity 2003.

[30] Subhash Khot and Rishi Saket. Hardness of embedding metric spaces of equal size. In Approximation, randomization, and combinatorial optimization. Algorithms and techniques, pages 218–227. Springer, 2007.

[31] Christiane Lammersen, Anastasios Sidiropoulos, and Christian Sohler. Streaming embeddings with slack. In Algorithms and Data Structures, 11th International Symposium, WADS 2009, Banff, Canada, August 21–23, 2009. Proceedings, pages 483–494, 2009.

[32] Jiří Matoušek and Anastasios Sidiropoulos. Inapproximability for metric embeddings into $\mathbb{R}^d$. Transactions of the American Mathematical Society, 362(12):6341–6365, 2010.

[33] Karl Menger. New foundation of euclidean geometry. American Journal of Mathematics, pages 721–745, 1931.

[34] Amir Nayyeri and Benjamin Raichel. Reality distortion: Exact and approximate algorithms for embedding into the line. In Foundations of Computer Science, 2015. FOCS 2015, to appear. IEEE, 2015.

[35] Christos Papadimitriou and Shmuel Safra. The complexity of low-distortion embeddings between point sets. In SODA, volume 5, pages 112–118, 2005.

[36] Michal Parnas and Dana Ron. Testing metric properties. Information and Computation, 187:155–195, 2003. An earlier conference version appeared in STOC 2001.

[37] Manfred J. Sippl and Harold A. Scheraga. Cayley-Menger coordinates. Proc. Natl. Acad. Sci. USA, 83:2283–2287, 1986.
A Missing details from Section 3

Proof of Lemma 3.6. Start with an arbitrary ordering of $X_{i-1}$, and follow the procedure laid-out in Definition 3.4. Specifically, for each $v_j$, we can compute the distance $r_j = (\rho(v_j, x_i) + \rho(x_i^*, v_j) - \rho(x_i, x_i^*)) / 2$ in constant time. Hence we know the distance $l_j = \rho(v_j, x_i) - r_j$ from $v_j$ to the stem of $x_i$ w.r.t. $\{v_j, x_i^*\}$. Now, starting from $v_j$, we follow the unique path towards the root $x_i^*$, till we meet an masked edge or we meet the root $x_i^*$. (In other words, we do not traverse the entire path $P_j$ from $v_j$ to $x_i^*$.) For any edge we traverse, first we orient it towards $x_i^*$ till we meet an edge $(w, w')$ such that $\rho(w, v_j) < l_j$ and $\rho(w', v_j) > l_j$, or a vertex $w$ such that $\rho(w, v_j) = l_j$. In this case, we set this edge or vertex to be a sink (as it would contain the stem of $x_i$ at $\{v_j, x_i^*\}$) and associate $v_j$ as its generating vertex. After the sink, we orient each edge to be towards the sink, till we meet an masked edge. The entire procedure to process all $v_j$’s takes $O(|V(T_{i-1})|)$ time – since all leaves of $T_{i-1}$ are from $X_{i-1}$, we have that $O(|V(T_{i-1})|) = O(|X_{i-1}|) = O(i)$.

Note that in this procedure, some sink vertex w.r.t. certain $\{v_j, x_i^*\}$ may ultimately not be a sink vertex for the final $x_i$-orientation; recall the figure above this lemma. So we may associate generating vertex information for some non-sink vertices.

Proof of Lemma 3.7. Consider the ordering of leaf vertices of $T_{i-1}$, denoted by $\{v_1, \ldots, v_k\}$, that gives rise to the $x_i$-orientation $\tilde{T}_{i-1}$. Following the procedure laid out in the proof of Lemma 3.6, after processing $v_1$, we obtained the $(x_i, v_1, x_i^*)$-orientation along the path $P_{v_1x_i^*}$ (from $v_1$ to $x_i^*$ in $T_{i-1}$). Hence at this point, there must be a sink w.r.t. $\{v_1, x_i^*\}$. If this is a sink edge, since this edge is masked, it remain unoriented throughout the entire procedure; that is, this edge will still be a sink edge for the final $x_i$-orientation $\tilde{T}_{i-1}$. Hence $\tilde{T}$ has at least one sink, and we are done.

Otherwise, consider the sink vertex $w_1$ w.r.t. $\{v_1, x_i^*\}$. Since $w_1$ is not a sink vertex for $\tilde{T}_{i-1}$, there must exist an outgoing edge $(w_1, u)$ oriented towards $u$. In other words, the orientation is away from the root $x_i^*$. Follow any outgoing edge from $u$ and continue in this manner: we have to either meet a sink edge / vertex (at which point we cannot find another outgoing edge to continue), or we reach a leaf $v_j$ of $T_{i-1}$. All edges we visited during this process are oriented away from the root $x_i^*$. Since when we start with each $v_j$, we always first orient edges towards the root $x_i^*$, this means such a path cannot lead us to a leaf $v_j$, and thus we have to stop at a sink of $\tilde{T}_{i-1}$. Putting everything together, $\tilde{T}_{i-1}$ has at least one sink.

Proof of Theorem 3.12. We claim that the output outlier set $\tilde{K}$ of Algorithm 1 stated in the main text 2-approximates the minimum outlier $d$-embedding of $(X, \rho)$. Specifically, let $K^*$ be an optimal outlier set. We will show: (a) $(X \setminus \tilde{K}, \rho)$ is $d$-embeddable; and (b) $|K^*| \leq |\tilde{K}| \leq 2|K^*|$.

To prove (a), let $\tilde{Y}_d$ be the $d+1$ points in (Step-0) that gives rise to $\tilde{K}$, and $\tilde{d}(\leq d)$ the embedding dimension of $(\tilde{Y}_d, \rho)$. Set $Z_d = X \setminus (K \cup \tilde{Y}_d)$. (Step-3) ensures that for any two points $z, z' \in Z_d$, $(\tilde{Y}_d \cup \{z, z'\}, \rho)$ must be $\tilde{d}$-embeddable. This is because any pair $z, z'$ that violates this condition will form an edge in the graph $G$ constructed in Step-3, and subsequently, the vertex cover will contain at least one point from such a pair. Statement (a) then follows from Theorem 3.11.

We now prove statement (b). The left inequality is evident. Let $Z^* = X \setminus K^*$. Since $(Z^*, \rho)$ is $d$-embeddable, by Theorem 3.11 there exists $d^* + 1$ points, say $Y^* = \{y_0, \ldots, y_{d^*}\}$ with $d^* \leq d$ being the embedding dimension of $Z^*$, such that $(Y^*, \rho)$ is strongly $d^*$-embeddable.

Now consider the time when the algorithm enumerates some $Y_d$ such that $Y_d \subseteq Z^*$ and $Y^* \subseteq Y_d$ in (Step-0); $Y_d = Y^*$ if $d^* = d$. In the subsequent (Step-2), any point inserted into $K$ violates
condition (ii) of Theorem 3.11 and thus must belong to the optimal outlier set \( K^* \) as well. In (Step-3), again by Theorem 3.11 (ii), for any edge \((z, z') \in E\), at least one of its endpoint is necessarily in \( K^* \). Hence a smallest set of outliers consistent with \( Y_d \) would consist of the points removed in (Step-2) and the vertex cover of the graph \( G \) constructed in (Step-3). Since the algorithm computes a 2-approximation of the vertex cover, it follows that the set \( K \) at the end of (Step-3) contains at most \( 2|K^*| \) points. Since \(|\hat{K}| \leq |K|\), we then have that \(|\hat{K}| \leq 2|K^*|\).

Hence the output outlier set \( \hat{K} \) is a 2-approximation for the minimum outlier embedding of \((X, \rho)\) to \( \mathbb{R}^d \) as claimed.

Finally, we analyze the time complexity of our algorithm. Steps 1 – 3 will be executed \( O(n^{d+1}) \) number of times. For each \( Y_d \), (Step-1) takes time polynomial in \( d \) due to the discussion below Theorem 3.11 (Step-2) performs \( O(n) \) number of embedding test, each for \( O(d) \) points. Hence it takes \( O(n^{dO(1)}) \) total time. In (Step-3), it takes \( O(n^2d^{O(1)}) \) time to construct the graph \( G = (Z, E) \). A 2-approximation of the vertex cover can easily be computed in \( O(|V| + |E|) \) time \[16\]. Putting everything together, the running time is \( O(n^{d+3}d^{O(1)}) = O(n^{d+3}) \), where the big-O notation hides terms polynomial in \( d \).

### B Hardness results

In this section, we show that the minimum outlier embedding problems into ultrametrics, trees, and Euclidean space are all NP-hard, by reducing the vertex cover problem to them in each case. In fact, it is NP-hard to approximate each of them within \( 2 - \varepsilon \), for any \( \varepsilon > 0 \), unless the unique game conjecture is true. For the case of minimum outlier embedding into Euclidean space, we note that our 2-approximation algorithm from the previous section almost matches the hardness result here.

All the hardness results are obtained via reduction from vertex cover. The high-level approach is very similar, although the specific reduction is different in each case. Below we present the result for tree metric first.

**Theorem B.1.** The problem of minimum outlier embedding into a tree metric is NP-hard. Furthermore, assuming the Unique Games Conjecture, it is NP-hard to approximate within a factor of \( 2 - \varepsilon \), for any \( \varepsilon > 0 \).

**Proof.** We use a reduction from Vertex Cover. This problem is known to be NP-hard. Furthermore, assuming the Unique Games Conjecture, it is also hard to approximate within a multiplicative factor of \( 2 - \varepsilon \) for any \( \varepsilon > 0 \) \[29\].

Let \( G = (V, E) \) be an instance to Vertex Cover. The goal is to find the smallest subset of vertices of \( V \) covering all edges in \( E \). We construct a discrete metric space \((X, d_X)\) as follows: Let \( \nu \) be an arbitrarily small positive number.

- For any \( v_i \in V \), add two nodes \( x_i, y_i \in X \). Set \( X = \{x_i, y_i\}_{i \in [1,n]} \cup \{o\} \).
- Set \( d_X(x_i, y_i) = 1 \), \( d_X(o, x_i) = 2 \), and \( d_X(o, y_i) = 1 \) for each \( i \in [1,n] \). Set \( d_X(y_i, y_j) = 2 \) and \( d_X(x_i, y_j) = 3 \) for \( i \neq j \).
- Set \( d_X(x_i, x_j) = 4 - \nu \) if \((v_i, v_j) \in E\) is an edge in \( E \); otherwise, set \( d_X(x_i, x_j) = 4 \) if \((v_i, v_j) \notin E\).

It is easy to verify that the above description indeed specifies a metric space. Intuitively, \((X, d_X)\) is close to the tree metric induced by a star-shaped tree: \( o \) is the root, tree edges are \((o, x_i) \) and
Theorem B.2. The problem of minimum outlier embedding into an ultrametric is NP-hard. Furthermore, assuming the Unique Games Conjecture, it is NP-hard to approximate within a factor of \(2 - \varepsilon\), for all \(\varepsilon > 0\).

Proof. As for the case of embedding into trees, we give a reduction from Vertex Cover. Specifically, let \(G = (V, E)\) be an instance of Vertex Cover. We construct a discrete metric space \((U, d_U)\) as follows: Let \(\nu\) be a sufficiently small positive number.

- For each \(v_i \in V, i \in [1, n]\), add a pair of nodes \(u_i, w_i \in U\). Hence \(U = \{u_i, w_i\}_{i \in [1, n]}\).
- Set \(d_U(u_i, w_i) = 2\nu\), and \(d_U(u_i, w_j) = 1\), \(d_U(w_i, w_j) = 1\) for any \(i \neq j\). For \(i \neq j\), set \(d_U(u_i, u_j) = 1 - \nu\) if \((v_i, v_j) \in E\); otherwise, set \(d_U(u_i, u_j) = 1\).

It is easy to verify that the above description indeed specifies a metric. Let \(K^*\) denote the optimal set of outliers so that \(d_U\) restricted to \(U \setminus K^*\) gives rise to an ultrametric. Let \(C^*\) denote the optimal vertex cover. Similar to the proof of Theorem B.1, we will show that \(|C^*| = |K^*|\). First, it is easy to see that by removing those vertices \(u_i\) such that \(v_i \in C^*\), the resulting metric is an ultrametric. Hence \(|K^*| \leq |C^*|\). It remains to show that \(|C^*| \leq |K^*|\).

To this end, construct \(\hat{V} = \{v_i \in V \mid u_i \in K^* \text{ or } w_i \in K^*\}\); obviously, \(|\hat{V}| \leq |K^*|\). We claim that \(\hat{V}\) forms a vertex cover for \(G = (V, E)\). Specifically, assume this is not the case and there exists an edge \((v_i, v_j) \in E\) such that \(v_i \notin \hat{V}\) and \(v_j \notin \hat{V}\). Then \(u_i, w_i, u_j \in U \setminus K^*\). However, these three points violate the three-point condition for ultrametrics as in [1], as \(d_U(w_i, u_j) = 1\), which is strictly larger than \(\max\{d_U(w_i, u_i) = 2\nu, d_U(u_j, u_i) = 1 - \nu\}\), which is a contradiction. Hence \(\hat{V}\) is a vertex cover and thus \(|C^*| \leq |\hat{V}| \leq |K^*|\). Overall, \(|C^*| = |K^*|\) and the theorem then follows.

We next present a hardness result for outlier embedding into constant-dimensional Euclidean space.
Proof. We first present the construction for \( d = 2 \), and then discuss how to extend it to the case \( d > 2 \). As before, we give a reduction from Vertex Cover. Specifically, given a graph \( G = (V,E) \), we construct a discrete metric space \((X,d_X)\) as follows: Let \( \nu \) be a sufficiently small positive real number. Set \( X = \{x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n\} \).

- We define:
  1. for any \( i \in [1,n] \), \( x_i = (i, -1), y_i = (i, 1) \), and \( z_i = (i, 0) \).

- We assign distances \( d_X \) as follows:
  1. \( d_X(x_i, z_i) = 1 \), \( d_X(y_i, z_i) = 1 \) and \( d_X(x_i, y_i) = 2 \), for any \( i \in [1,n] \).
  2. For any \( i,j \in [1,n] \), for \( a,b \in \{x,y\} \), \( d_X(a_i, b_j) = \|a_i - b_j\| \).
  3. For any \( i,j \in [1,n] \), \( d_X(z_i, x_j) = \|z_i - x_j\| \), \( d_X(z_i, y_j) = \|z_i - y_j\| \).
  4. For \( i \neq j \), if \( (v_i, v_j) \in E \), then \( d_X(z_i, z_j) = |j - i| - \nu \); otherwise, set \( d_X(z_i, z_j) = |j - i| \).

In other words, other than \( z_i \) and \( z_j \) potentially deviates from the distance given by their coordinates, the distance between all other pairs of points are consistent with their coordinates. It is easy to verify that \((X,d_X)\) as described above satisfies triangle inequality and is a valid metric. Let \( K^* \) denote the optimal set of outliers so that \( d_X \) restricted to \( X \setminus K^* \) can be embedded in \( \mathbb{R}^2 \) isometrically. Let \( C^* \) denote the optimal vertex cover for graph \( G = (V,E) \). We now show \( |K^*| = |C^*| \).

Specifically, first note that given \( C^* \), we can construct the set \( K \) containing all \( y_i \) such that \( v_i \in C^* \). It is easy to see that the coordinates for the remaining points \( X \setminus K \) are consistent with their pairwise distances. Hence \( |K^*| \leq |K| = |C^*| \). We next show that \( |C^*| \leq |K^*| \).

Consider \( K^* \). We construct the set \( \hat{V} = \{v_i \mid x_i \in K^* \text{ or } y_i \in K^* \text{ or } z_i \in K^*\} \). We now argue that \( \hat{V} \) is a vertex cover for the graph \( G = (V,E) \). Assume that this is not true, and that there exists an edge \((v_i, v_j) \in E \) such that \( v_i \notin \hat{V} \) and \( v_j \notin \hat{V} \). This means that the 6 points in \( \{x_i, y_i, z_i, x_j, y_j, z_j\} \) must all belong to \( X \setminus K^* \). However, we argue that their pairwise distances are not consistent. In particular, consider the 3 points \( x_i, y_i, x_j \). They are isometrically embeddable into \( \mathbb{R}^2 \) (but not into \( \mathbb{R}^1 \)). Hence once their coordinates are fixed, there is only a unique position possible for any other point \( p \) given its distance to these three points. In other words, we can uniquely embed \( z_i \), as well as \( z_j \) into \( \mathbb{R}^2 \) w.r.t. the current coordinates of \( \{x_i, y_i, x_j\} \) (which form a valid isometry consistent with \( d_X \)). However, the Euclidean distance between \( z_i \) and \( z_j \) is not consistent with \( d_X \). Hence the points \( \{x_i, y_i, x_j, z_i, z_j\} \) are not isometrically embeddable into \( \mathbb{R}^2 \), which contradicts our assumption that all of them are in \( X \setminus K^* \). It then follows that \( \hat{V} \) has to be a vertex cover, and \( |C^*| \leq |\hat{V}| \leq |K^*| \). This completed the proof for \( d = 2 \).

We next discuss how the above construction can be generalized to the case \( d > 2 \). In the construction of \((X,d_X)\) we map each \( v_i \in V \) to a sequence of \( d + 1 \) points \( x_i^{(0)}, \ldots, x_i^{(d)} \) in \( X \) contained in the subspace \( \mathbb{R}^{d-1} \times \{i\} \), where the first \( d \) points form a canonical \((d-1)\)-simplex, while the last point \( x_i^{(d)} \) is the centroid of this simplex. The rest of the argument remains essentially identical so we omit the details.
C Bi-criteria approximation for embedding into $\mathbb{R}^d$

As before, let $d_E$ denote the Euclidean distance in $\mathbb{R}^d$. Given a point $x \in \mathbb{R}^d$ and a set $A \subset \mathbb{R}^d$, the distance from $x$ to $A$ is defined as $d(x, A) = \inf_{y \in A} d_E(x, y)$.

**Definition C.1** (Near-isometric $d$-embedding). Given a metric space $\mathcal{X} = (X, \rho_X)$, we say that a map $\varphi : X \rightarrow \mathbb{R}^d$ is a $\delta$-near-isometric embedding of $\mathcal{X}$ into $\mathbb{R}^d$ if for any $x, x' \in X$, we have $|\rho(x, x') - d_E(\varphi(x), \varphi(x'))| \leq \delta$. In this case, we also say that $\mathcal{X}$ admits a $\delta$-near-isometric $d$-embedding, and $(\varphi(\mathcal{X}), d_E)$ is a $\delta$-approximation of $\mathcal{X}$.

We note that the parameter $\delta$-above is not normalized w.r.t. the diameter of the set; this is for the convenience of the latter argument in this section. Before we describe our algorithm, we need some more definitions (following the notations of [3]).

**Definition C.2** (Normalized sequence). Given an ordered sequence of points $U = \{u_0, u_1, \ldots, u_d\} \subset \mathbb{R}^d$, we say that $U$ is normalized if $u_0 = 0 = (0, \ldots, 0)$, and if $u_i \in \mathbb{R}_+$ for all $1 \leq i \leq d$.

Given $U \subseteq A \subset \mathbb{R}^d$, if $f : A \rightarrow \mathbb{R}^d$ is a map such that $f(U)$ is normalized, then we say that the map $f$ is normalized at $U$.

**Definition C.3** (Maximal sequence). Let $k \leq d$. Given an ordered sequence of points $U = \{u_0, \ldots, u_k\}$ from a set $A \subset \mathbb{R}^d$, we say that $U$ is maximal in $A$ if: (i) $d_E(u_0, u_1) = \Delta(A)$, and (ii) for any $2 \leq i \leq k$, the distance $d(u_i, \text{aff}\{u_0, \ldots, u_{i-1}\})$ is maximal in $A$. Here, $\text{aff}B$ is the affine subspace of $\mathbb{R}^d$ spanned by points in set $B \subset \mathbb{R}^d$.

**Description of the algorithm.** Assume that the input $n$-point metric space $\mathcal{X} = (X, \rho)$ admits a $(\varepsilon, k^*)$-outlier embedding into $\mathbb{R}^d$. We now present an algorithm, described in Algorithm 2, to find an approximation of such an embedding. The high level structure of this algorithm parallels that of Algorithm 1, which intuitively is an algorithm for the special case when $\varepsilon = 0$. However, the technical details here are more involved so as to tackle several issues caused by the near-isometric embedding. In this algorithm, $C_d$ is a positive constant depending only on the dimension $d$.

For the time being, for the sake of simplicity of presenting the main ideas, let us assume that we can implement (Step-1,2) (i.e., the computation of $P_d$ and $p_\varepsilon$) in polynomial time. We will address how to modify Algorithm 2 to achieve this later in this section. Our main result is the following:

**Theorem C.4.** Given an $n$-point metric space $\mathcal{X} = (X, \rho)$, if $\mathcal{X}$ admits a $(\varepsilon, k^*)$-outlier embedding in $\mathbb{R}^d$, then Algorithm 2 returns an outlier set $\hat{K}$ that witnesses an $(O(\sqrt{\varepsilon}), |\hat{K}|)$-outlier embedding of $\mathcal{X}$ in $\mathbb{R}^d$ with $|\hat{K}| \leq 2k^*$. Here $O(\sqrt{\varepsilon})$ hides constants depending on the dimension $d$.

**Proof of Theorem C.4.** Let $\hat{Z} = X \setminus \hat{K}$. By (Step-2) and (Step-3), for any two points $z, z' \in \hat{Z}$, we computed $p_\varepsilon$ and $P_d$ in (Step-2) such that $|\rho(z, z') - d_E(p_\varepsilon, P_d)| \leq C_d\sqrt{\varepsilon}\Delta(X) + \varepsilon\Delta(X)$. Hence $(\hat{Z}, \rho)$ indeed has an $O(\sqrt{\varepsilon}\Delta(X))$-near-isometric embedding in $\mathbb{R}^d$.

It remains to show that $|\hat{K}| \leq 2k^*$. Let $K^*$ be an optimal set of outliers such that $|K^*| = k^*$ and $(X \setminus K^*, \rho)$ admits a $\varepsilon\Delta(X)$-near-isometric $d$-embedding, say $\varphi : X \setminus K^* \rightarrow \mathbb{R}^d$. Set $Z^* = X \setminus K^*$. For simplicity, we assume that there does not exist $d' < d$ such that $(X \setminus K^*, \rho)$ admits a $\varepsilon\Delta(X)$-near-isometric $d'$-embedding; the case where such $d' < d$ exists can be handled similarly.

Let $Y^* = \{y_0, \ldots, y_d\}$ be the ordered sequence such that $\varphi(Y^*)$ is maximal in $\varphi(Z^*)$. We assume w.l.o.g. that $\varphi(Z^*)$ is also normalized in $\mathbb{R}^d$; if not, there exists an isometry $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ so that $T(\varphi(Y^*))$ is normalized, and we can simply take $\varphi$ as $T \circ \varphi$. 

19
Proof. For simplicity of presentation, set $y_{d+1} = x$, and consider the set $A = \{q_0, \ldots, q_d, q_{d+1} = \varphi(x)\} \subset Q^*$, and the map $g_x : A \rightarrow \mathbb{R}^d$ defined as $g_x(q_i) = h_x(y_i)$ for any $i \in [0, d+1]$. Since both $\varphi$ and $h_x$ are $\varepsilon \Delta(X)$-near-isometric maps, we have by triangle inequality that: for any $i, j \in [0, d+1]$, 

$$|d_E(q_i, q_j) - d_E(h_x(y_i), h_x(y_j))| \leq |d_E(q_i, q_j) - \rho(y_i, y_j)| + |\rho(y_i, y_j) - d_E(h_x(y_i), h_x(y_j))| \leq 2\varepsilon \Delta(X).$$

Consider the time when our algorithm enumerates $Y_d = Y^*$ in (Step-0). In the subsequent (Step-1), the algorithm will be able to find an normalized sequence $P_d$ as claimed, since $\varphi(Y_d)$ satisfies the requirements. Following our earlier assumption on $Z^*$, $d' = d$ for $P_d$. In (Step-2), for a point $x \in X \setminus Z^*$, we claim that there exists $p_x \in \mathbb{R}^d$ such that $(P_d \cup \{p_x\}, d_E)$ is an $2\varepsilon \Delta(X)$-approximation of $(Y_d \cup \{x\}, \rho)$. To prove this, we need to use the following result proved in [3].

Proposition C.5 (Section 2.6, [3]). For any fixed $d > 1$, let $A = \{o, u_1, \ldots, u_d, x\} \subset \mathbb{R}^d$ be such that the $\tilde{d}$-sequence $U = (o, u_1, \ldots, u_d)$ is normalized and maximal in $A$. Let $f : A \rightarrow \mathbb{R}^d$ be a map such that $(f(A), d_E)$ is an $\delta \Delta(A)$-approximation of $(A, d_E)$, and $f$ is normalized at $U$. Then there exists a constant $c_d$ depending only on the dimension $d$ such that $d_E(f(x), f(x)) \leq c_d \sqrt{\delta \Delta(A)}$.

Consider the image $Q^* = \varphi(Z^*)$ of an optimal embedding $\varphi : Z^* = X \setminus K^* \rightarrow \mathbb{R}^d$. Denote by $Q_d = \{q_0 = \varphi(y_0), q_1 = \varphi(y_1), \ldots, q_d = \varphi(y_d)\}$ and $g : Q_d \rightarrow P_d \subset \mathbb{R}^d$ the map that sends $q_i$ to $p_i$, for each $i \in [0, d]$. We also use $h : Y^* \rightarrow P_d$ to denote the $\varepsilon \Delta(X)$-near-isometric embedding computed in (Step-1).

For any $x \in Z^*$, by the same argument as for the existence of $P_d$, we can show that there must exist a $\varepsilon \Delta(X)$-near-isometric embedding $h_x : Y_d \cup \{x\} \rightarrow \mathbb{R}^d$ of $(Y_d \cup \{x\}, \rho)$ such that the ordered sequence $\{h_x(y_0), \ldots, h_x(y_d)\}$ is normalized. Furthermore, Proposition C.5 implies the following:

Claim C.6. For any $x \in Z^*$ and any $h_x$ as described above, $d_E(\varphi(x), h_x(x)) \leq 2c_d \sqrt{\varepsilon \Delta(X)}$. 

Proof. For simplicity of presentation, set $y_{d+1} = x$, and consider the set $A = \{q_0, \ldots, q_d, q_{d+1} = \varphi(x)\} \subset Q^*$, and the map $g_x : A \rightarrow \mathbb{R}^d$ defined as $g_x(q_i) = h_x(y_i)$ for any $i \in [0, d+1]$. Since both $\varphi$ and $h_x$ are $\varepsilon \Delta(X)$-near-isometric maps, we have by triangle inequality that: for any $i, j \in [0, d+1]$, 

$$|d_E(q_i, q_j) - d_E(h_x(y_i), h_x(y_j))| \leq |d_E(q_i, q_j) - \rho(y_i, y_j)| + |\rho(y_i, y_j) - d_E(h_x(y_i), h_x(y_j))| \leq 2\varepsilon \Delta(X).$$
In other words, this means that the map $g_x$ is a $2\varepsilon \Delta(X)$-nearisometry for $(A, d_E)$. Setting $\delta = 2\varepsilon \Delta(X)/\Delta(A)$, it then follows from Proposition C.5 that

$$d_E(g_{d+1} = \varphi(x), h_x(x)) \leq c_d \sqrt{\delta} \Delta(A) = c_d \sqrt{2\varepsilon \Delta(X) \Delta(A)} \leq 2c_d \sqrt{\varepsilon} \Delta(X).$$

The last inequality follows that since $\varphi$ is $\varepsilon \Delta(X)$-nearisometry, we have that $\Delta(A) \leq \Delta(Q^*) \leq (1 + \varepsilon) \Delta(Z^*) \leq 2\Delta(X)$ for $\varepsilon \leq 1$. □

On the other hand, note that the above statement is generic for any $x \in Z^*$ and any $\varepsilon \Delta(X)$-near-isometric map $h_x$ such that the ordered sequence $\{h_x(y_0), \ldots, h_x(y_d)\}$ is normalized. Hence we can choose $x = y_i$ for any $i \in [0, d]$ and choose $h_x = h$ computed in (Step-1) $h : Y_d \to P_d$ that gives rise to $P_d$. It then follows from Claim C.6 that:

$$\text{for any } i \in [0, d], \ d_E(\varphi(y_i), p_i = h(y_i)) \leq 2c_d \sqrt{\varepsilon} \Delta(X). \tag{11}$$

Finally for any $x \in X \setminus Y_d$ inspected in (Step-2), set $p_x = h_x(x)$ for any $\varepsilon \Delta(X)$-near-isometry $h_x : Y^* \cup \{x\} \to \mathbb{R}^d$ (which must exist as we argued earlier). Combining Claim C.6, Eqn (11), the triangle inequality and the fact that $\varphi$ is an $\varepsilon \Delta(X)$-nearisometry, we thus have:

$$|\rho(y_i, x) - d_E(p_i, p_x)| \leq |\rho(y_i, x) - d_E(\varphi(y_i), \varphi(x))| + |d_E(\varphi(y_i), \varphi(x)) - d_E(p_i, p_x)| \leq \varepsilon \Delta(X) + d_E(\varphi(y_i), p_i) + d_E(\varphi(x), p_x) \leq \varepsilon \Delta(X) + 4c_d \sqrt{\varepsilon} \Delta(X). \tag{12}$$

Setting $C_d = 4c_d$, we obtain the following.

**Corollary C.7.** For any $x \in Z^*$, there exists $p_x$ satisfying the requirements in (Step-2).

By a similar argument as in Eqn (12), we can also obtain:

**Corollary C.8.** For any $z, z' \in Z^*$, let $p_z$ and $p_{z'}$ be their corresponding points in $\mathbb{R}^d$ computed in (Step-2). Then $|\rho(z, z') - d_E(p_z, p_{z'})| \leq C_d \sqrt{\varepsilon} \Delta(X) + \varepsilon \Delta(X)$.

**Lemma C.9.** $|\hat{K}| \leq 2 |K^*|$.

*Proof.* First, as argued earlier above Corollary C.7, for any $x \in Z^*$, (Step-2) will be able to compute the required $p_x \in \mathbb{R}^d$. Let $K_1$ denote the set $O$ at the end of (Step-2). It then follows that $K_1 \subset K^*$. Furthermore, by Corollary C.8 for any edge $((z, p), (z', p')) \in E$ in the graph constructed in Step-3, at least one of $z$ and $z'$ is necessarily in $K^*$. Hence $|K^* \setminus K_1|$ is at least the size of the vertex cover of $G$. Since we compute a 2-approximation of the vertex cover, it follows that the set $K$ at the end of Step-3 satisfies that $|K| \leq 2 |K^*|$. This proves the claim as $|\hat{K}| \leq |K|$. □

This completes the proof of Theorem C.4.

**Computational issues.** The computation of $P_d$ and $p_x$ in (Step-1) and (Step-2) require solving a system of $O(d)$ quadratic equations and inequalities, and the solution would require infinite precision. This can be addressed using real algebraic geometry theory for semi-algebraic sets which can output the solutions in arbitrary precision in time exponential in $O(d)$. We can also use the following “gridding” strategy: After (Step-1), we discretize the $d$-dimensional cube $[-3\Delta(X), 3\Delta(X)]^d$ where the grid length is $\tau = \frac{\varepsilon}{27} \Delta(X)$. In (Step-1) and (Step-2), we look for solutions for $P_d$ and $p_x$ where coordinates fall on grid points. This requires us to relax the distance distortion bound by another factor of $\varepsilon \Delta(X)$. However it will not change the final statement in Theorem C.4. The time to compute each near-isometry embedding is $(\frac{d}{\varepsilon})^{O(d)}$. Overall, following a similar analysis as for Algorithm 1, we conclude with Theorem 4.3.