Learning Lines with Ordinal Constraints

Bohan Fan
Department of Computer Science, University of Illinois at Chicago, IL, USA
bfan4@uic.edu

Diego Ihara
Department of Computer Science, University of Illinois at Chicago, IL, USA
https://dihara2.people.uic.edu/
dihara2@uic.edu

Neshat Mohammadi
Department of Computer Science, University of Illinois at Chicago, IL, USA
nmoham24@uic.edu

Francesco Sgherzi
Department of Computer Science, University of Illinois at Chicago, IL, USA
https://fsgher2.people.uic.edu/
fsgher2@uic.edu

Anastasios Sidiropoulos
Department of Computer Science, University of Illinois at Chicago, IL, USA
https://sidiropo.people.uic.edu/
sidiropo@uic.edu

Mina Valizadeh
Department of Computer Science, University of Illinois at Chicago, IL, USA
mvaliz2@uic.edu

Abstract

We study the problem of finding a mapping f from a set of points into the real line, under ordinal triple constraints. An ordinal constraint for a triple of points (u, v, w) asserts that $|f(u) - f(v)| < |f(u) - f(w)|$. We present an approximation algorithm for the dense case of this problem. Given an instance that admits a solution that satisfies $(1-\varepsilon)$-fraction of all constraints, our algorithm computes a solution that satisfies $(1 - O(\varepsilon^{1/8}))$-fraction of all constraints, in time $O(n^7) + (1/\varepsilon)^{O(1/\varepsilon^{1/8})}n$.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis

Keywords and phrases metric learning, embedding into the line, ordinal constraints, approximation algorithms

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2020.45

Category APPROX

Related Version https://arxiv.org/abs/2004.13202

Supplementary Material https://github.com/lnghrdntcr/lloc

Funding Supported by NSF grants CCF-1815145, CCF-1934915, and by NSF CAREER award 1453472.

1 Introduction

Geometric methods provide several tools for the analysis of complicated data sets, such as nearest-neighbor search, clustering, and dimensionality reduction. The key abstraction is to encode a set of objects by mapping each object to a point in some metric space, such that the distance between points quantifies the pairwise dissimilarity between the corresponding

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020).
Editors: Jarosław Byrka and Raghu Meka; Article No. 45; pp. 45:1–45:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

objects. The success of this paradigm crucially depends on the metrical representation used to encode the data. Motivated by this fact, metric learning aims at developing methods for discovering an underlying metric space from proximity information (we refer the reader to [23, 17] for a detailed exposition). There are several different formulations of the metric learning problem that have been considered in the literature. Here, we focus on the popular case of *ordinal* constraints. In this case, the input consists of a set of points $X = [n]$, together with a set $\mathcal{T}$ of ordered triples (u, v, w) of points, representing the fact that u is *more similar* to v than to w. The goal is to find a mapping $f : X \to Y$, for some host metric space (Y, ρ), such that for all $(u, v, w) \in \mathcal{T}$, we have

$$\rho(f(u), f(v)) < \rho(f(u), f(w)). \tag{1}$$

In general, there might be no mapping f that satisfies all constraints of the form (1), so we are interested in the algorithmic problem of computing a mapping that minimizes the fraction of violated constraints. We focus on the case where the host space is the real line, so the objective can be formulated as computing a mapping $f : [n] \to \mathbb{R}$, where for each $(u, v, w) \in \mathcal{T}$ we have the constraint

$$|f(u) - f(v)| < |f(u) - f(w)|. \tag{2}$$

We refer to this problem as Line Learning with Ordinal Constraints (LLOC).

1.1 Our contribution

We present an approximation algorithm for learning a line metric space under ordinal constraints, for the case of dense instances. Here, the density condition means that all ordinal information is given, i.e. for any distinct $u, v, w \in [n]$, we have either $(u, v, w) \in \mathcal{T}$, or $(u, w, v) \in \mathcal{T}$. Our main result is summarized in the following.

▶ **Theorem 1.1.** *There exists an algorithm that given an instance of LLOC that admits a solution satisfying $(1 - \varepsilon)$-fraction of all constraints, outputs a solution that satisfies $(1 - O(\varepsilon^{1/8}))$-fraction of all constraints, in time $O(n^7) + (1/\varepsilon)^{O(1/\varepsilon^{1/8})}n$.*

Brief overview of our approach

The main idea used to obtain Theorem 1.1 is to first compute an ordering that is close to the ordering of the points in the optimal solution. This is done by "guessing" a point p^* that lies within the few left-most points in an optimal solution, and such that p^* is not involved in many violated constraints. We show that the ordinal constraints involving p^* can be used to order the points by first solving an instance of the Minimum Feedback Arc Set problem on a tournament, and then computing a topological ordering of the remaining acyclic graph. We use this ordering to partition the points into "buckets", and we show that for almost all buckets, almost all their points must be mapped inside an interval that does not contain many other points. This property allows us to define a smaller instance of the problem by contracting each bucket into a single point. This new smaller instance can be solved exactly, and its solution can be pulled back to the original problem.

1.2 Related work

Metric learning

Another popular formulation of the metric learning problem uses *contrastive* constraints. In this case, the input consists of a set of points $X = [n]$, together with sets $\mathcal{S}, \mathcal{D} \subseteq \binom{X}{2}$, where $\mathcal{S}$ contains pairs labeled as *similar*, and $\mathcal{D}$ contains pairs labeled as *dissimilar*. The goal is to find a mapping $f : X \to Y$, for some host metric space (Y, ρ), such that for all $\{u, v\} \in \mathcal{S}$,

$$\rho(f(u), f(v)) \leq \ell,$$

and for all $\{u, v\} \in \mathcal{D}$,

$$\rho(f(u), f(v)) \geq h,$$

for some given threshold values $\ell, h > 0$. This problem is easily seen to be a generalization of Correlation Clustering. It has been studied for the case dense instances, when the host metric space is either Euclidean or a tree [13]. The main result of [13] is a FPTAS for the case where there exists a mapping that satisfies all constraints, that is allowed to violate the constraints by a small multiplicative factor which is referred to as *contrastive distortion*. In contrast, in the present work, we do not introduce any distortion, and we do not need to assume that there exists a mapping satisfying all the constraints.

We also note that the case of arbitrary instances (i.e., not necessarily dense) under contrastive constraints has been studied for the setting of learning Mahalanobis metric spaces (i.e., when X is a set of points in d-dimensional Euclidean space, and f is required to be linear) [12]. This version of the problem is related to the theory of LP-type problems.

Embedding into the line

The problem of computing a geometric representation of a data set into the real line has been studied extensively in various forms. This is arguably the simplest instance of dimensionality reduction, which is also a prototypical unsupervised metric learning task. Various objectives have been studied, including multiplicative [19, 20, 6, 5, 7, 11], additive [3], and average [10, 22] distortion. We refer the read to [14] for a detailed exposition. A related notion is ordinal embeddings, where one seeks to obtain mappings that approximately preserve the relative ordering of pairwise distances [2, 4]. We remark that a key difference between these works and our result is that they seek to minimize the *ordinal distortion*, which is a multiplicative factor of violation of the ordinal constraints, while we are interested in minimizing the number of violated ordinal constraints (without introducing ordinal distortion).

Betweenness

In the Betweenness problem we are given some set $X = [n]$ and a set $\mathcal{T}$ of ordered triples $(a, b, c) \in [n]^3$. The goal is to find a bijection $g : [n] \to [n]$ such that for any $(a, b, c) \in \mathcal{T}$, $g(b)$ appears between $g(a)$ and $g(c)$. This problem has been studied extensively in the literature. It is known to be MAXSNP-hard [9] (see also [21]), and remains hard to approximate even on dense instances [1]. The case of tournaments has been shown to admit a PTAS [15], while the best approximation algorithm for general instances is the 1/3-approximation obtained by taking a uniformly random ordering, assuming the Unique Games conjecture [8] (see also [18]).

The Betweenness problem is conceptually similar to the Line Learning with Ordinal Constraints problem studied here. However, as we now explain, the two problems have some important differences. A first difference is that the ordinal constraint (2) does not imply any ordering constraint[1]. A second difference is that the solution space to the Line Learning with Ordinal Constraints problem that we study is larger. In other words, the ordering of the points is not always enough to recover a nearly-optimal constraint. For example, consider the instance on $X = \{0, 2, 4, \ldots, 2k, 2k+1, \ldots, 3k\}$, with all constraints $(u, v, w) \in X^3$, such that $|u - v| < |u - w|$. Clearly, setting f to be the identity results in a solution that satisfies all constraints. However, just the ordering of the points in f is not enough to obtain a good solution: setting $g(u_i) = i$, where $g(u_1) < g(u_2) < \ldots < g(u_n)$ results in a solution g that violates a constant fraction of all constraints.

1.3 Organization

The rest of the paper is organized as follows. Section 2 presents, as a warm up, an exact polynomial-time algorithm for the case where there exists a solution that satisfies all constraints. Section 3 presents the algorithm for the general case. Section 4 presents the analysis. Section 5 gives the proof of a technical Lemma which is used in the proof of the main result.

2 Warm up: An exact algorithm with no violations

We now describe an exact polynomial-time algorithm for the case where there exists an optimal solution that satisfies all constraints. This algorithm is significantly simpler than the one used to prove our main result. However, it illustrates the main idea of using the constraints involving some point p to deduce an ordering of all points, and then using this ordering to obtain an embedding into the line. The algorithm is summarized in the following.

▶ **Theorem 2.1.** *Let ε^* be the fraction of violated constraints, then there exists a polynomial-time algorithm which given an instance $([n], \mathcal{T})$ of the LLOC problem, either computes a mapping $f : [n] \to \mathbb{R}$ that satisfies all the constraints, or correctly decides that no such mapping exists.*

Proof. Fix some optimal mapping $f^* : [n] \to \mathbb{R}$, that satisfies all constraints in $\mathcal{T}$. We guess $p = \arg\min_{x \in [n]} f^*(x)$. For all $i, j \in [n]$, let $d_{i,j} = |f^*(x_j) - f^*(x_i)|$. We first determine the ordering of all the points on the real line, and then we compute the mapping using their distance constraints and solving some LP.

Suppose that $[n] = \{x_1, \ldots, x_n\}$, such that

$$f^*(p) = f^*(x_1) < f^*(x_2) < \ldots < f^*(x_n).$$

Since $\varepsilon^* = 0$, it follows that for all $i < j \in [n]$, we have $d_{1,i} < d_{1,j}$, and $(1, i, j) \in \mathcal{T}$. Therefore, for any $q, q' \in [n]$, we can decide whether $f^*(q) < f^*(q')$ or $f^*(q') < f^*(q)$ based on whether $(p, q, q') \in \mathcal{T}$ or $(p, q', q) \in \mathcal{T}$. Therefore, we can compute the ordering $x_1, \ldots, x_n$ of $[n]$ by running a sorting algorithm using pairwise comparisons.

We now compute a mapping using an LP. For any $i < j \in \{1, \ldots, n\}$, we have $|f^*(x_i) - f^*(x_j)| = \sum_{t=i}^{j-1} d_{t,t+1}$. Therefore for each $(x_i, x_j, x_k) \in \mathcal{T}$, the constraint $|f^*(x_i) - f^*(x_j)| < |f^*(x_i) - f^*(x_k)|$ can be written as $\sum_{t=i}^{j-1} d_{t,t+1} < \sum_{t=i}^{k-1} d_{t,t+1}$. Thus computing the desired mapping f can be done by computing a feasible solution to the following LP:

[1] For example, the constraint (u, v, w) is satisfied by both solutions $f(u) = 1, f(v) = 2, f(w) = 3$, and $f(u) = 2, f(v) = 1, f(w) = 4$, however the former solution implies the ordering $f(u) < f(v) < f(w)$, while the latter implies $f(v) < f(u) < f(w)$

$$d_{i,j} \geq 0 \text{ for all } i < j \in [n]$$
$$\sum_{t=i}^{j-1} d_{t,t+1} < \sum_{t=i}^{k-1} d_{t,t+1} \text{ for all } (x_i, x_j, x_k) \in \mathcal{T}$$

This concludes the proof. ◀

3 The algorithm for the general case

In this Section we present the algorithm for the general case of the problem. The algorithm uses as a subroutine an exact algorithm for a generalized weighted version of the problem. This exact algorithm is used on small instances that are constructed via a process which we refer to as a *retraction*.

3.1 Retractions

We now define a weighted version of the metric learning problem, where each constraint is associated with some weight, and the goal is to maximize the total weight of all satisfied constraints. Formally, an input to the Weighted Line Learning with Ordinal Constraints (WLLOC) problem is defined by a tuple $([b], \mathcal{T}, w)$, where $b \in \mathbb{N}$, and $\mathcal{T}$ are as before, and $w : \mathcal{T} \to \mathbb{R}$ is a weight function. The goal is to find a solution $f : [b] \to [0,1]$ that minimizes the total weight of violated constraints.

▶ **Theorem 3.1.** *There exists an exact algorithm for the WLLOC problem with running time $O(n^{3n})$.*

Proof. We identify the space of possible solutions with $[0,1]^n$, by mapping each solution $f : [b] \to [0,1]$ to the vector $x_f = (f(1), \ldots, f(n)) \in [0,1]^n$. For any $(i,j,k) \in \mathcal{T}$, we have the constraint

$$|f(i) - f(j)| < |f(i) - f(k)|.$$

The feasible region for this constraint is thus defined as a union of certain cells in an arrangement $A_{(i,j,k)}$ of a constant number of open halfspaces in $\mathbb{R}^n$. Let A be the arrangement obtained as the union of all halfspaces for all $(i,j,k) \in \mathcal{T}$. It is known that any arrangement of a halfspaces in $\mathbb{R}^b$ has complexity $O(a^b)$ (see [24] and references therein), and thus A has complexity $O(|\mathcal{T}|^n) = O(n^{3n})$. By enumerating all the cells in this arrangement, we find a solution that satisfies a set of constraints of maximum total weight, which results in an algorithm with running time $O(n^{3n})$. ◀

As mentioned earlier, the exact algorithm from Theorem 3.1 will be used as a subroutine on smaller instances. The following Definition describes a process for mapping large unweighted instances to smaller weighted ones.

▶ **Definition 3.2** (Retraction). *Given an instance $\phi = ([n], \mathcal{T})$ of the LLOC problem, and some partition $\mathcal{B} = \{B_1, \ldots, B_b\}$ of $[n]$, we define the $\mathcal{B}$-retraction of ϕ to be the instance $\phi' = ([b], \mathcal{T}', w)$ of the WLLOC problem where for any $(i,j,k) \in \mathcal{T}'$, we have*

$$w((i,j,k)) = |\mathcal{T} \cap (B_i \times B_k \times B_j)|.$$

3.2 The algorithm

The last ingredient we need is an approximation algorithm for the Minimum Feedback Arc Set problem on tournaments, which is summarized in the following.

▶ **Theorem 3.3** (Kenyon-Mathieu & Schudy [16]). *There exists a randomized algorithm for the Minimum Feedback Arc Set problem on weighted tournaments. Given $\epsilon > 0$, it outputs a solution with expected cost at most $(1+\epsilon)\mathsf{OPT}$. The expected running time is $O(1/\varepsilon)n^6 + 2^{\tilde{O}(1/\varepsilon)}n^2 + 2^{2^{\tilde{O}(1/\varepsilon)}}n$.*

We are now ready to describe the general algorithm. Let $\mathcal{T}_n$ denote the set of all ordered triples of distinct elements in $[n]$. Recall that the input consists of a set $\mathcal{T} \subseteq \mathcal{T}_n$, such that for any set of distinct $i, j, k \in [n]$, we have that exactly one of the triples (i, j, k) and (i, k, j) is contained in $\mathcal{T}$.

The algorithm proceeds in the following steps:

Step 1: Exhaustively computing a left-most point. Iterate Steps 2–5 for all values $p \in [n]$.

Step 2: Cycle removal. Construct a tournament $G^{(p)} = ([n], A^{(p)})$, where

$$A^{(p)} = \{(i, j) : (p, i, j) \in \mathcal{T}\}.$$

Compute an $O(1)$-approximate minimum feedback arc set, $F^{(p)} \subset A^{(p)}$, in $G^{(p)}$, using the algorithm in Theorem 3.3.

Step 3: Ordering. Compute a topological ordering $z_1^{(p)}, \dots, z_n^{(p)}$ of $G^{(p)} \setminus F^{(p)}$.

Step 4: Retraction. Let $b = O(\varepsilon^{-1/8})$. For any $i \in [b]$, let

$$\mathcal{B}_i^{(p)} = \bigcup_{j=(i-1)n/b+1}^{in/b} \{z_j^{(p)}\}.$$

Let $\psi^{(p)}$ be the $\mathcal{B}^{(p)}$-retraction of $\phi^{(p)}$.

Step 5: Extension. Using the algorithm from Theorem 3.1, we compute an optimal solution $g^{(p)} : [b] \to [0, 1]$ for the instance $\psi^{(p)}$ of WLLOC. We define $f^{(p)} : [n] \to [0, 1]$ by setting for any $i \in [n]$, $f^{(p)}(i) = g^{(p)}(j)$, where $j \in [b]$ such that $i \in B_j^{(p)}$. The algorithm outputs the solution $f^{(p)}$.

Step 6: Return the best solution found among $f^{(1)}, \dots, f^{(n)}$.

This completes the description of the algorithm.

4 Analysis of the algorithm

This Section presents the analysis of the algorithm, which is the proof of Theorem 1.1.

For the remainder of the analysis, let us fix some optimal solution $f_{\mathsf{OPT}} : [n] \to [0, 1]$ for the instance $([n], \mathcal{T})$ of the LLOC problem. Fix a numbering $\{x_1, \dots, x_n\} = [n]$, such that

$$f_{\mathsf{OPT}}(x_1) \leq f_{\mathsf{OPT}}(x_2) \leq \ldots \leq f_{\mathsf{OPT}}(x_n).$$

For any $f : [n] \to [0, 1]$, for any $i \in [n]$, and for any $\alpha \in [0, 1]$, we say that i is *α-good in f*, if at least α-fraction of the constraints of the form $(i, j, k) \in \mathcal{T}$ are satisfied; i.e.:

$$|\{(i, j, k) \in \mathcal{T} : |f(i) - f(j)| < |f(i) - f(k)|\}| \geq \alpha \binom{n-1}{2}.$$

We first argue that there exists a $(1 - \varepsilon^{1/2})$-good point that is close to the left-most point in the optimal solution:

▶ **Lemma 4.1.** *There exists $i^* \in [2\varepsilon^{1/2}n]$, such that x_{i^*} is $(1-\varepsilon^{1/2})$-good in f_{OPT}.*

Proof. Let ξ be the total number of constraints violated by f_{OPT}. We have $\xi \leq \varepsilon \cdot |\mathcal{T}| = \varepsilon n\binom{n-1}{2}$. Suppose that there exists no $i \in [2\varepsilon^{1/2}n]$ such that x_i is $(1-\varepsilon^{1/2})$-good. Therefore every $i \in [2\varepsilon^{1/2}n]$ participates in at least $\varepsilon^{1/2}\binom{n-1}{2}$ violated constraints of the form (i,j,k), for some $j,k \in [n]$. Thus the total number of violated constraints is at least $\xi \geq 2n\varepsilon\binom{n-1}{2}$, which is a contradiction, concluding the proof. ◀

For the remainder of this section, fix some $i^* \in [2\varepsilon^{1/2}n]$, such that x_{i^*} is $(1-\varepsilon^{1/2})$-good, as in Lemma 4.1. Let f' be the embedding obtained from f_{OPT} by exchanging the images of x_1 and x_{i^*}, that is for all $i \in [n]$,

$$f'(x_i) = \begin{cases} f_{\mathsf{OPT}}(x_{i^*}) & \text{if } i = 1 \\ f_{\mathsf{OPT}}(x_1) & \text{if } i = i^* \\ f_{\mathsf{OPT}}(x_i) & \text{otherwise} \end{cases}$$

We next show that f' is near-optimal.

▶ **Lemma 4.2.** *The total number of violated constraints in f' is at most $(\varepsilon + O(1/n))n\binom{n-1}{2}$.*

Proof. Let $\mathcal{T}_1 \subseteq \mathcal{T}$ be the set of constraints that are violated in f' and in f_{OPT}. Let $\mathcal{T}_2 \subseteq \mathcal{T}$ be the set of constraints that are violated in f' but not in f_{OPT}. We have $|\mathcal{T}_1| \leq \varepsilon n\binom{n-1}{2}$. Since f_{OPT} and f' differ only on x_1 and x_{i^*}, it follows that every constraint $(i,j,k) \in \mathcal{T}_2$ must contain at least one of 1 and i^*. There are at most $6n^2$ such constraints. Thus $|\mathcal{T}_2| \leq 6n^2$. We conclude that the total number of constraints violated in f' is at most $|\mathcal{T}_1| + |\mathcal{T}_2| \leq (\varepsilon + O(1/n))n\binom{n-1}{2}$, which concludes the proof. ◀

The next Lemma shows that x_{i^*} remains $(1-O(\varepsilon^{1/2}))$-good in f'.

▶ **Lemma 4.3.** *We have that x_{i^*} is $(1-O(\varepsilon^{1/2}))$-good in f'.*

Proof. Let $\gamma = (x_{i^*}, j, k) \in \mathcal{T}$, and suppose that γ is satisfied in f_{OPT}. If

$$f_{\mathsf{OPT}}(x_{i^*}) \leq f_{\mathsf{OPT}}(j) \leq f_{\mathsf{OPT}}(k),$$

then, since $f'(j) = f_{\mathsf{OPT}}(j)$, and $f'(k) = f_{\mathsf{OPT}}(k)$, it follows that

$$f'(x_{i^*}) \leq f'(j) \leq f'(k),$$

and thus γ is also satisfied in f'.

Thus, the only possible constraints of the form $(x_{i^*}, j, k) \in \mathcal{T}$, that are not violated in f_{OPT}, but are violated in f', must satisfy either $f_{\mathsf{OPT}}(j) < f_{\mathsf{OPT}}(x_{i^*})$, or $f_{\mathsf{OPT}}(k) < f_{\mathsf{OPT}}(x_{i^*})$. In other words, we must have $\{j,k\} \cap \{x_1, \ldots, x_{i^*-1}\} \neq \emptyset$. Therefore, there are at most $2\varepsilon^{1/2}n^2$ such constraints. Since x_{i^*} is $(1-\varepsilon^{1/2})$-good in f_{OPT}, it follows that x_{i^*} is $(1-O(\varepsilon^{1/2}))$-good in f', which concludes the proof. ◀

Let $F' = \{(j,k) \in A^{(i^*)} : (x_{i^*}, j, k) \in \mathcal{T} \text{ and } f' \text{ violates } (x_{i^*}, j, k)\}$. The next Lemma shows F' is a valid feedback arc set for $G^{(i^*)}$.

▶ **Lemma 4.4.** *F' is a feedback arc set for $G^{(i^*)}$, with $|F'| \leq (O(\varepsilon^{1/2}))\binom{n-1}{2}$.*

Proof. By Lemma 4.3, x_{i^*} is $(1 - O(\varepsilon^{1/2}))$-good, and thus $|F'| \leq (O(\varepsilon^{1/2}))\binom{n-1}{2}$. Thus, it suffices to show that F' is a feedback vertex set. For any $(j, k) \in A^{(i^*)} \setminus F'$, we have that (x_{i^*}, j, k) is satisfied in f'. Since x_{i^*} is mapped to the left-most point in f', it follows that $f'(j) < f'(k)$. Thus,

$$x_{i^*}, x_2, x_3, \ldots, x_{i^*-1}, x_1, x_{i^*+1}, x_{i^*+2}, \ldots, x_n$$

is a topological ordering of $G^{(i^*)} \setminus F'$, and thus F' is a feedback arc set, which concludes the proof. ◀

If the instance admits a solution with no violations, then it can be shown that the bucketing $\mathcal{B}^{(i^*)}$ computed by the algorithm agrees with a partition of the optimal solution to contiguous disjoint intervals. In the following, we show that, in the general case, the bucketing is "close" to such a partition. First, we introduce a notion of "stability" which formalizes what it means for a bucket to be close to an optimal interval.

▶ **Definition 4.5** (Stability). *Let $i \in [b]$. We say that i is* stable *if there exists some interval $I \subset \mathbb{R}$, such that*

$$\left|I \cap f'\left(B_i^{(i^*)}\right)\right| \geq (1 - \varepsilon^{1/8}) \cdot n/b,$$

and

$$\left|I \cap f'\left([n] \setminus B_i^{(i^*)}\right)\right| \leq \varepsilon^{1/8} \cdot n/b,$$

We also say that i is I-stable. We say that i is unstable *(I-unstable) if it is not stable (I-stable).*

The following Lemma gives a characterization of unstable buckets.

▶ **Lemma 4.6.** *Suppose that $i \in [b]$ is unstable. Then there exist pairwise disjoint intervals $I_1, I_2, I_3 \subset \mathbb{R}$, that appear in this order from left to right in the line, such that*

$$\left|I_1 \cap f'\left(B_i^{(i^*)}\right)\right| \geq n\varepsilon^{1/8}/(2b),$$

$$\left|I_3 \cap f'\left(B_i^{(i^*)}\right)\right| \geq n\varepsilon^{1/8}/(2b),$$

and

$$\left|I_2 \cap f'\left([n] \setminus B_i^{(i^*)}\right)\right| > \varepsilon^{1/8} \cdot n/b,$$

Proof. Let $I_1 \subset \mathbb{R}$ be the minimal interval that contains the $n\varepsilon^{1/8}/(2b)$ left-most points in $f'(B_i^{(i^*)})$, and let $I_3 \subset \mathbb{R}$ be the minimal interval that contains the $n\varepsilon^{1/8}/(2b)$ right-most points in $f'(B_i^{(i^*)})$. Let $I_2 \subset \mathbb{R}$ be the maximal interval that is contained between I_1 and I_3. Since $\varepsilon^{1/8} < 1$, we have that $I_1 \cap I_3 = \emptyset$, and therefore, all intervals I_1, I_2, I_3 are well-defined and pairwise disjoint. By construction, I_1 and I_3 each contains exactly $n\varepsilon^{1/8}/(2b)$ points in $f'(B_i^{(i^*)})$. Therefore, it remains to show that I_2 contains more than $\varepsilon^{1/8}n/b$ points in $f'([n] \setminus B_i^{(i^*)})$. Suppose, for the sake of contradiction, that I_2 contains at most $\varepsilon^{1/8}n/b$ in $f'([n] \setminus B_i^{(i^*)})$. Then, I_2 contains exactly $(1 - \varepsilon^{1/8})n/b$ points in $f'(B_i^{(i^*)})$, and at most $\varepsilon^{1/8}n/b$ points in $f'([n] \setminus B_i^{(i^*)})$, implying that B_i is stable, which is a contradiction. This concludes the proof. ◀

We next show that for each unstable bucket, the feedback arc set must contain many edges incident to vertices in the bucket.

▶ **Lemma 4.7.** *Let $i \in [b]$ be unstable. Then, $F^{(i^*)} \cup F'$ contains at least $\varepsilon^{1/4} n^2/(2b^2)$ arcs having exactly one endpoint in $B_i^{(i^*)}$.*

Proof. Let $I_1, I_2, I_3 \subset \mathbb{R}$ be the intervals given by Lemma 4.6. Let $v \in [n] \setminus B_i^{(i^*)}$, such that $f'(v) \in I_2$. Pick $j \in [b]$, such that $v \in B_j^{(i^*)}$. We consider two cases:

Case 1: Suppose that $j < i$. Let $u \in B_i^{(i^*)}$, such that $f'(u) \in I_1$. If $(v,u) \in A^{(i^*)}$, then it follows that f' violates (x_{i^*}, v, u), and thus $(v,u) \in F'$. Otherwise, we have $(u,v) \in A^{(i^*)}$. Since u appears after v in the topological sort of $G^{(i^*)} \setminus F^{(i^*)}$, it follows that $(u,v) \in F^{(i^*)}$. Thus, in either case, $F^{(i^*)} \cup F'$ contains either (u,v) or (v,u). Therefore, $F^{(i^*)} \cup F'$ contains at least $n\varepsilon^{1/8}/(2b)$ arcs having u as an endpoint.

Case 2: Suppose that $j > i$. This case is similar to Case 1, and is included for completeness. Let $u \in B_i^{(i^*)}$, such that $f'(u) \in I_3$. If $(u,v) \in A^{(i^*)}$, then it follows that f' violates (x_{i^*}, u, v), and thus $(u,v) \in F'$. Otherwise, we have $(v,u) \in A^{(i^*)}$. Since u appears before v in the topological sort of $G^{(i^*)} \setminus F^{(i^*)}$, it follows that $(v,u) \in F^{(i^*)}$. Thus, in either case, $F^{(i^*)} \cup F'$ contains either (u,v) or (v,u). Therefore, $F^{(i^*)} \cup F'$ contains at least $n\varepsilon^{1/8}/(2b)$ arcs having u as an endpoint.

We conclude that, in either case, for any $u \in B_i^{(i^*)}$, $F^{(i^*)} \cup F'$ contains at least $n\varepsilon^{1/8}/(2b)$ arcs having u as an endpoint. Summing over all $u \in B_i^{(i^*)}$, we obtain that $F^{(i^*)} \cup F'$ contains at least $\varepsilon^{1/4} n^2/(2b^2)$ arcs having an endpoint in $B_i^{(i^*)}$. This concludes the proof. ◀

Next, we bound the number of unstable buckets.

▶ **Lemma 4.8.** *Let $J = \{i \in [b] : i \text{ is unstable}\}$, we have $|J| \le O(\varepsilon^{1/4}) 2b^2$.*

Proof. By Lemma 4.3 we have that x_{i^*} is $(1 - O(\varepsilon^{1/2}))$-good in f', and by Lemma 4.4 we have that $G^{(i^*)}$ admits a feedback arc set of size at most $(O(\varepsilon^{1/2}))\binom{n-1}{2}$. Thus, by Theorem 3.3, the algorithm computes some feedback arc set $F^{(i^*)} \subset A^{(i^*)}$, with $|F^{(i^*)}| = O(\varepsilon^{1/2} n^2)$. We note that here we only use Theorem 3.3 to obtain a $O(1)$-approximation. By Lemma 4.7,

$$\begin{aligned} |J| &\le |F^{(i^*)} \cup F'|/(\varepsilon^{1/4} n^2/(2b^2)) \\ &\le O(\varepsilon^{1/4}) 2b^2, \end{aligned}$$

which concludes the proof. ◀

For any stable $i \in [b]$, let $\mathcal{I}_i \subset \mathbb{R}$ be the interval that contains at least $(1 - \varepsilon^{1/8})n/b$ points in $f'(B_i^{(i^*)})$, and at most $\varepsilon^{1/8} n/b$ other points. Let also $\mathcal{J}_i \subset \mathcal{I}_i$ be an open interval that contains all but the $\varepsilon^{1/8} n/b$ leftmost points in $f'(B_i^{(i^*)}) \cap \mathcal{I}_i$, and the $\varepsilon^{1/8} n/b$ rightmost points in $f'(B_i^{(i^*)}) \cap \mathcal{I}_i$. Thus, $|\mathcal{J}_i \cap f'(B_i^{(i^*)})| \ge (1 - 3\varepsilon^{1/8})n/b$. It follows that for any $i \neq j \in [b]$, such that both i and j are stable, we have $\mathcal{J}_i \cap \mathcal{J}_j = \emptyset$.

Intuitively, we intend to find a solution that satisfies a nearly-optimal fraction of constraints, while ignoring all constraints that involve points that are mapped outside the intervals $\mathcal{J}_i$, where $i \in [b]$ is stable. To that end, we define a small set of points that the analysis can safely "ignore":

$$X_{\mathsf{Noise}} = \bigcup_{i \in [b] : i \text{ stable}} \left\{ v \in B_i^{(i^*)} : f'(v) \notin \mathcal{J}_i \right\}.$$

Since $|\mathcal{J}_i \cap f'(B_i^{(i^*)})| \geq (1 - 3\varepsilon^{1/8})n/b$, it follows that

$$|X_{\mathsf{Noise}}| \leq 3\varepsilon^{1/8} n \tag{3}$$

Let also, for any $i \in [b]$,

$$\bar{B}_i^{(i^*)} = B_i^{(i^*)} \setminus X_{\mathsf{Noise}}.$$

We identify a set of triples $(i, j, k) \in [b]^3$ for which, intuitively, it is difficult to satisfy at least some significant fraction of all constraints with one point from each of the clusters $B_i^{(i^*)}$, $B_j^{(i^*)}$, and $B_k^{(i^*)}$. Formally, we say that some $(i, j, k) \in [b]^3$ is *brittle* if there exist $u, u' \in \mathcal{J}_i$, $v, v' \in \mathcal{J}_j$, and $w, w' \in \mathcal{J}_k$, such that

$$|u - v| < |u - w|,$$

and

$$|u' - v'| > |u' - w'|.$$

Intuitively, the above property implies that if for all $t \in [b]$, all points in $\bar{B}_t^{(i^*)}$ get mapped to the same point $p_t \in \mathcal{J}_t$, then there exist choices for the points $\{p_t\}_t$, such that some constraint in $\bar{B}_i^{(i^*)} \times \bar{B}_j^{(i^*)} \times \bar{B}_k^{(i^*)}$ is violated; in other words, if a triple (i, j, k) is not brittle, then the choice of the points p_t does not affect the satisfiability of the constraints in $\bar{B}_i^{(i^*)} \times \bar{B}_j^{(i^*)} \times \bar{B}_k^{(i^*)}$.

We are now ready to show that the retraction computed by the algorithm admits a solution of low total cost.

▶ **Lemma 4.9.** *The instance $\psi^{(i^*)}$ of WLLOC constructed in Step 4 admits a solution that satisfies constraints of total weight at least $|\mathcal{T}| \cdot (1 - O(\varepsilon^{1/8}))$.*

Proof. We define a mapping $g : [b] \to [0, 1]$, and $g' : [b] \to [0, 1]$, as follows. For each $i \in [b]$, pick $v_i \in \mathcal{B}_i^{(i^*)}$, arbitrarily, and set

$$g'(i) = f'(v_i).$$

For any $j \in [b]$, we set

$$g(j) = g'(i),$$

where $i \in [b]$ is the unique integer such that $i \in B_i^{(i^*)}$. By the definition of the WLLOC instance $\psi^{(i^*)}$, the total weight of the constraints violated by g' equals the total number of constraints violated by g. It therefore suffices to upper bound the number of constraints in $\mathcal{T}$ that are violated by g.

We define a partition $\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3 \cup \mathcal{T}_4 \cup \mathcal{T}_5$, where

$$\mathcal{T}_0 = \{(u, v, w) \in \mathcal{T} : f' \text{ violates } (u, v, w)\},$$
$$\mathcal{T}_1 = \{(u, v, w) \in \mathcal{T} : \text{ at least two of } u, v, w \text{ are in the same cluster in } \mathcal{B}^{(i^*)}\},$$
$$\mathcal{T}_2 = \{(u, v, w) \in \mathcal{T} : u \in B_i^{(i^*)}, v \in B_j^{(i^*)}, w \in B_k^{(i^*)}, \text{ at least one of } i, j, k \text{ is unstable}\},$$
$$\mathcal{T}_3 = \{(u, v, w) \in \mathcal{T} : u \in B_i^{(i^*)}, v \in B_j^{(i^*)}, w \in B_k^{(i^*)}, \text{ and } (i, j, k) \text{ is brittle}\},$$
$$\mathcal{T}_4 = \{(u, v, w) \in \mathcal{T} : \{u, v, w\} \cap X_{\mathsf{Noise}} \neq \emptyset\},$$
$$\mathcal{T}_5 = \mathcal{T} \setminus (\mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3 \cup \mathcal{T}_4).$$

By Lemma 4.2 we have

$$|\mathcal{T}_1| \le (\varepsilon + O(1/n))n\binom{n-1}{2}.$$

Since every cluster in $\mathcal{B}^{(i^*)}$ has n/b points, we have

$$|\mathcal{T}_1| \le 3n^3/b^2. \tag{4}$$

In order to bound $|\mathcal{T}_3|$ we need a bound on the number of brittle triples. This is done in Lemma 4.8, which appears in Section 5. We thus have

$$|\mathcal{T}_2| \le O(\varepsilon^{1/4})2n^3 b. \tag{5}$$

By Lemma 5.5 we have

$$|\mathcal{T}_3| \le n^3/b. \tag{6}$$

By (3) we have

$$|\mathcal{T}_4| \le O(\varepsilon^{1/8})n^3 \tag{7}$$

Let $(u,v,w) \in \mathcal{T}_5$. By the definition of $\mathcal{T}_5$, we have that $u \in \bar{B}_i^{(i^*)}$, $v \in \bar{B}_j^{(i^*)}$, and $w \in \bar{B}_k^{(i^*)}$, for some distinct $i,j,k \in [b]$, such that (i,j,k) is not brittle, and f' satisfies (u,v,w), that is

$$|f'(u) - f'(v)| < |f'(u) - f'(w)|.$$

By the definition of a brittle triple we get

$$|g(u) - g(v)| < |g(u) - g(w)|,$$

and thus g satisfies (u,v,w). We obtain that g satisfies all constraints in $\mathcal{T}_5$. Thus, by (4)–(7), the number of constraints violated by g is at most $|\mathcal{T}_0| + \ldots + |\mathcal{T}_4| \le n^3 O(\varepsilon^{1/8})$, which concludes the proof. ◀

We are now ready to prove our main result.

Proof of Theorem 1.1. By Lemma 4.9 we have that WLLOC instance $\psi^{(i^*)} = ([b], \mathcal{T}', w)$ constructed at Step 4 of the algorithm, admits a mapping $g' : [b] \to \mathbb{R}$, such that the total weight of the constraints in $\mathcal{T}'$ violated by g' is at most $O(\varepsilon^{1/8}n^3)$. Therefore, in Step 5, using the exact algorithm from Theorem 3.1, we compute a mapping $g : [b] \to \mathbb{R}$, violating the same total weight as g'. By the definition of retraction, it follows that the mapping $f^{(i^*)}$ computed in Step 5 violates at most $O(\varepsilon^{1/8}n^3)$ constraints in $\mathcal{T}$, as required.

It remains to bound the running time. Step 2 uses the algorithm from Theorem 3.3 to obtain a $O(1)$-approximate minimum feedback arc set, and thus takes time $O(n^6)$. Step 3 takes time $O(n)$ and Step 4 takes time $O(n^2)$. Step 5 runs the algorithm from Theorem 3.1 on an input of size b, and thus takes time $O(b^{3b}) + O(n)$. Step 6 requires computing the number of violated constraints in each of the n solutions, and thus takes total time $O(n^4)$. Due to Step 1, the Steps 2–5 are repeated n times, and thus the total running time is at most $O(n^7 + b^{3b}n) = O(n^7) + (1/\varepsilon)^{O(1/\varepsilon^{1/8})}n$, which concludes the proof. ◀

5 Bounding the number of brittle triples

This Section is devoted to proving an upper bound on the number of brittle triples. We begin by deriving a simple condition that is a consequence of brittleness.

▶ **Lemma 5.1.** *Let $j < i < k \in [b]$. We have that if (i, j, k) is brittle, then there exist $p_i \in \mathcal{J}_i$, $p_j \in \mathcal{J}_j$, $p_k \in \mathcal{J}_k$, such that*

$$p_i - p_j = p_k - p_i.$$

Proof. If (i, j, k) is brittle, it is easy to see that $\mathcal{J}_i$ must be located between $\mathcal{J}_j$ and $\mathcal{J}_k$; otherwise, any representative point chosen in $\mathcal{J}_i$ must be closer to all the points in $\mathcal{J}_j$ than those in $\mathcal{J}_k$, or vice versa. By definition, there exist $p_i \in \mathcal{J}_i$, $p'_j \in \mathcal{J}_j$, $p'_k \in \mathcal{J}_k$, such that

$$p_i - p'_j \geq p'_k - p_i,$$

and $p''_j \in \mathcal{J}_j$, $p''_k \in \mathcal{J}_k$, such that

$$p_i - p''_j < p''_k - p_i.$$

Without loss of generality, assume $p'_j < p''_j$ and $p'_k < p''_k$, and define $\delta \in [0, 1]$. Comparing $d_{ij}(\delta) = p_i - (p'_j + \delta(p''_j - p'_j))$ and $d_{ik}(\delta) = (p'_k + \delta(p''_k - p'_k)) - p_i$, we have $d_{ij}(0) - d_{ik}(0) \geq 0$ and $d_{ij}(1) - d_{ik}(1) < 0$. There exist $\delta' \in [0, 1]$, s.t. $d_{ij}(\delta') - d_{ik}(\delta') = 0$.

Define $p_j = (p'_j + \delta'(p''_j - p'_j)) \in \mathcal{J}_i$ and $p_k = (p'_k + \delta'(p''_k - p'_k)) \in \mathcal{J}_j$, we have

$$p_i - p_j = p_k - p_i,$$

which concludes the proof. ◀

▶ **Lemma 5.2.** *Let $i_1, i_2, i_3, j_1, j_2, j_3, k_1, k_2, k_3 \in \mathbb{R}$, with $i_1 < i_2 < i_3$, $j_1 < j_2 < j_3$, $k_1 < k_2 < k_3$. For any $\alpha, \beta, \gamma \in \{1, 2\}$, let $H_{\alpha,\beta,\gamma}$ be the axis-parallel parallelepiped defined by*

$$H_{\alpha,\beta,\gamma} := \mathcal{CH}(\{(i_{\alpha+\alpha'}, j_{\beta+\beta'}, k_{\gamma+\gamma'}) : \alpha', \beta', \gamma' \in \{0, 1\}\}).$$

Let h be any plane in $\mathbb{R}^3$. Then, there exist $\alpha^*, \beta^*, \gamma^* \in \{0, 1\}$, such that h does not intersect the interior of $H_{\alpha^*,\beta^*,\gamma^*}$.

Proof. For any $d \geq 2$, any d-dimensional halfspace containing the origin must also contain at least one d-orthant. The assertion follows immediately from the case $d = 3$. ◀

▶ **Lemma 5.3.** *Let $i, j, k \in [b]$, with $j+1 < i$, and $i+1 < k$. Then, there exist $i', j', k' \in \{0, 1\}$ such that $(i + i', j + j', k + k')$ is not brittle.*

Proof. Define the plane

$$h = \{(x_I, x_J, x_K) \in \mathbb{R}^3 : x_I - x_J = x_K - x_I\}.$$

By Lemma 5.1, we have that if $(i + i', j + j', k + k')$ is brittle, then h must intersect the hyperrectangle $\mathcal{J}_{i+i'} \times \mathcal{J}_{j+j'} \times \mathcal{J}_{k+k'}$. However, by Lemma 5.2, it follows that there exist $i', j', k' \in \{0, 1\}$, such that h does not intersect $\mathcal{J}_{i+i'} \times \mathcal{J}_{j+j'} \times \mathcal{J}_{k+k'}$, and thus $(i + i', j + j', k + k')$ is not brittle, which concludes the proof. ◀

▶ **Lemma 5.4** (Brittle convexity). *Let $\{e_1, e_2, e_3\}$ be the standard orthonormal basis in $\mathbb{R}^3$. Let $v \in [b - 2]^3$, and let $w \in \{e_1, e_2, e_3\}$, such that v and $v + 2w$ are both brittle. Then, $v + w$ is also brittle.*

Proof. By Lemma 5.1, there exist $p_i \in \mathcal{J}_i$, $p_j \in \mathcal{J}_j$, $p_k \in \mathcal{J}_k$, such that

$$p_i - p_j = p_k - p_i. \tag{8}$$

Let $w = (i', j', k')$. Similarly, there exist $q_i \in \mathcal{J}_{i+2i'}$, $q_j \in \mathcal{J}_{j+2j'}$, $q_k \in \mathcal{J}_{k+2k'}$, such that

$$q_i - q_j = q_k - q_i. \tag{9}$$

For any $\alpha \in [0, 1]$, let

$$z_i^{(\alpha)} = (1 - \alpha)p_i + \alpha q_i,$$

$$z_j^{(\alpha)} = (1 - \alpha)p_j + \alpha q_j,$$

$$z_k^{(\alpha)} = (1 - \alpha)p_k + \alpha q_k.$$

Let us assume that $w = e_1$. The cases $w = e_2$ and $w = e_3$ can be handled in a similar manner. We have that for all $\alpha \in [0, 1]$, $z_j^{(\alpha)} \in \mathcal{J}_j$, and $z_k^{(\alpha)} \in \mathcal{J}_k$. Moreover, $z_i^{(0)} \in \mathcal{J}_i$, and $z_i^{(1)} \in \mathcal{J}_{i+2}$, which implies that there exists some $\alpha^* \in [0, 1]$, such that $z_i^{(\alpha^*)} \in \mathcal{J}_{i+1}$. We have

$$\begin{aligned} z_i^{(\alpha^*)} - z_j^{(\alpha^*)} &= (1 - \alpha^*)p_i + \alpha^* q_i - (1 - \alpha^*)p_j - \alpha^* q_j \\ &= (1 - \alpha^*)(p_i - p_j) + \alpha^*(q_i - q_j) \\ &= (1 - \alpha^*)(p_k - p_i) + \alpha^*(q_k - q_i) \\ &= (1 - \alpha^*)p_k + \alpha^* q_k - (1 - \alpha^*)p_i - \alpha^* q_i \\ &= z_k^{(\alpha^*)} - z_i^{(\alpha^*)}, \end{aligned}$$

which by Lemma 5.1 implies that $v + w$ is brittle, and concludes the proof. ◀

We are now ready to bound the number of brittle triples, which is the main result of this Section.

▶ **Lemma 5.5.** *The number of brittle triples is at most $O(b^2)$.*

Proof. Let $B \subseteq [b]^3$ be the set of all brittle triples, and let $B' = [b]^3 \setminus B$. For any $s \in \{0, 1\}^3$, let

$$U_s = s \cdot b/2 + [b/2]^3,$$

and $B_s = B \cap U_s$. Since $B = \bigcup_s B_s$, and there are only 8 different values for s, it suffices to show that for any $s \in \{0, 1\}^3$, $|B_s| = O(b^2)$. We shall prove this for the case $s = (0, 0, 0)$. All remaining cases can be handled in a similar manner.

For the remainder for the proof, let $s = (0, 0, 0)$. By Lemma 5.3, it follows that for any $v \in B_s^3$, there exists $v' \in B'$, with $v' - v \in \{0, 1\}^3$. This implies that there exists $u \in B$, and $u' \in B'$, with $u - v \in \{0, 1\}^3$, $u' - v \in \{0, 1\}^3$, and $u' - u \in \{e_1, e_2, e_3\}$, where $\{e_1, e_2, e_3\}$ is the standard orthonormal basis in $\mathbb{R}^3$. Let $t = u' - u$. By Lemma 5.4, it follows by induction that for any $i \in \{1, \ldots, b/2\}$, the triple $u + i \cdot t$ is brittle. Let

$$R_v = \bigcup_{i=1}^{b/2} \{u + c \cdot i\}.$$

Thus $R_v \subseteq B'$. Note that, since $s = (0, 0, 0)$, we have

$$|R_v| \geq b/2. \tag{10}$$

For any $j \in \{1,2,3\}$, we say that v is *type-j*, if $t = e_j$.

Let

$$B_{s,j} = \{v \in B_s : v \text{ is type-}j\}.$$

Let $j^* \in \{1,2,3\}$, such that $|B_{s,j^*}| \geq |B_s|/3$.

By the above construction, it follows that for any $v, w \in B_{s,j^*}$, with $\|v - w\|_\infty \geq 2$, we have $R_v \cap R_w = \emptyset$. We greedily construct some $C \subseteq B_{s,j^*}$ as follows. We start with $C := \emptyset$, and $D := B_{s,j^*}$. While $D \neq \emptyset$, we pick any $v \in D$, and we set $C := C \cup \{v\}$, and $D := D \setminus \mathsf{Ball}_\infty(v,1)$, where $\mathsf{Ball}_\infty(v,r)$ denotes the ℓ_∞-ball of radius r centered at v. For every v added to C, we delete at most 9 elements from D, and thus

$$|C| \geq |B_{s,j^*}|/9 \geq |B_s|/27.$$

Since for any $v, w \in C$, we have $\|v - w\|_\infty$, it follows that $R_v \cap R_w = \emptyset$. Combining with (10), we get

$$b^3 \geq |B'| \geq \left|\bigcup_{v \in C} R_v\right| = \sum_{v \in C} |R_v| \geq |C| \cdot b/2 \geq |B_s| \cdot b/54,$$

and thus $|B_s| \leq 54b^2$, which concludes the proof. ◀

References

1 Nir Ailon and Noga Alon. Hardness of fully dense problems. *Information and Computation*, 205(8):1117–1129, 2007.

2 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):1–21, 2008.

3 Mihai Badoiu. 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.

4 Mihai Bădoiu, Erik D Demaine, MohammadTaghi Hajiaghayi, Anastasios Sidiropoulos, and Morteza Zadimoghaddam. Ordinal embedding: Approximation algorithms and dimensionality reduction. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 21–34. Springer, 2008.

5 Mihai Badoiu, Kedar Dhamdhere, Anupam Gupta, Yuri Rabinovich, Harald Räcke, Ramamoorthi Ravi, and Anastasios Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. In *SODA*, volume 5, pages 119–128. Citeseer, 2005.

6 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.

7 Timothy Carpenter, Fedor V Fomin, Daniel Lokshtanov, Saket Saurabh, and Anastasios Sidiropoulos. Algorithms for low-distortion embeddings into arbitrary 1-dimensional spaces. In *34th International Symposium on Computational Geometry (SoCG 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

8 Moses Charikar, Venkatesan Guruswami, and Rajsekar Manokaran. Every permutation csp of arity 3 is approximation resistant. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 62–73. IEEE, 2009.

9 Benny Chor and Madhu Sudan. A geometric approach to betweenness. *SIAM Journal on Discrete Mathematics*, 11(4):511–523, 1998.

10 Kedar Dhamdhere, Anupam Gupta, and R. Ravi. Approximation algorithms for minimizing average distortion. *Theory Comput. Syst.*, 39(1):93–111, 2006. doi:10.1007/s00224-005-1259-6.

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

12 Diego Ihara, Neshat Mohammadi, Francesco Sgherzi, and Anastasios Sidiropoulos. Robust mahalanobis metric learning via geometric approximation algorithms. *CoRR*, abs/1905.09989, 2019. arXiv:1905.09989.

13 Diego Ihara, Neshat Mohammadi, and Anastasios Sidiropoulos. Algorithms for metric learning via contrastive embeddings. In *35th International Symposium on Computational Geometry (SoCG 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

14 Piotr Indyk, Jiří Matoušek, and Anastasios Sidiropoulos. Low-distortion embeddings of finite metric spaces. In Jacob E. Goodman, Joseph O'Rourke, and Csaba D. Toth, editors, *Handbook of Discrete and Computational Geometry, Second Edition.* Chapman and Hall/CRC, 2017.

15 Marek Karpinski and Warren Schudy. Approximation schemes for the betweenness problem in tournaments and related ranking problems. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 277–288. Springer, 2011.

16 Claire Kenyon-Mathieu and Warren Schudy. How to rank with few errors. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 95–103, 2007.

17 Brian Kulis et al. Metric learning: A survey. *Foundations and Trends® in Machine Learning*, 5(4):287–364, 2013.

18 Yury Makarychev. Simple linear time approximation algorithm for betweenness. *Operations research letters*, 40(6):450–452, 2012.

19 Amir Nayyeri and Benjamin Raichel. Reality distortion: Exact and approximate algorithms for embedding into the line. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 729–747. IEEE, 2015.

20 Amir Nayyeri and Benjamin Raichel. A treehouse with custom windows: Minimum distortion embeddings into bounded treewidth graphs. In Philip N. Klein, editor, *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 724–736. SIAM, 2017. doi:10.1137/1.9781611974782.46.

21 Jaroslav Opatrny. Total ordering problem. *SIAM Journal on Computing*, 8(1):111–114, 1979.

22 Yuri Rabinovich. On average distortion of embedding metrics into the line and into l1. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 456–462, 2003.

23 Gregory Shakhnarovich. *Learning task-specific similarity.* PhD thesis, Massachusetts Institute of Technology, 2005.

24 Csaba D Toth, Joseph O'Rourke, and Jacob E Goodman. *Handbook of discrete and computational geometry.* Chapman and Hall/CRC, 2017.