Approximating the Distortion*

Alexander Hall† Christos Papadimitriou‡

Abstract

Kenyon et al. (STOC 04) compute the distortion between one-dimensional finite point sets when the distortion is small; Papadimitriou and Safra (SODA 05) show that the problem is NP-hard to approximate within a factor of 3, albeit in 3 dimensions. We solve an open problem in these two papers by demonstrating that, when the distortion is large, it is hard to approximate within large factors, even for 1-dimensional point sets. We also introduce additive distortion, and show that it can be easily approximated within a factor of two.

1 Introduction

The *distortion problem* is the following: Given two d-dimensional finite points sets $S, T \subseteq \mathbb{R}^d$ with $|S| = |T|$ and a real number $\delta \in \mathbb{R}$ (the *distortion*) is there a bijection $f : S \to T$ such that we have $\mathrm{expansion}(f) \cdot \mathrm{expansion}(f^{-1}) := \max_{x,y \in S} \left(\frac{d(f(x),f(y))}{d(x,y)}\right) \cdot \max_{x,y \in S} \left(\frac{d(x,y)}{d(f(x),f(y))}\right) \leq \delta$? Here $d(x, y)$ denotes the Euclidean distance between two points $x, y \in \mathbb{R}^d$.

The distortion problem was introduced by Kenyon et al. [9], who gave an optimal algorithm for 1-dimensional points sets that are known to have distortion less than $3+2\sqrt{2}$. Their elaborate dynamic programming algorithm crucially relies on the small distortion guarantee to establish and exploit certain restrictions on the bijection between the two point sets. Papadimitriou and Safra [15] present NP-hardness and inapproximability results, which hold for both small and large distortions—albeit for the 3 dimensional case. In both papers the question of whether the distortion of 1-dimensional point sets can be computed or approximated if it is not known to be small was proposed as an open problem.

In this paper we resolve this question by establishing several strong NP-hardness and inapproximability results. In Section 2.1 we show that the distortion problem is NP-hard in the 1-dimensional case when the distortion is at least $|S|^{\varepsilon}$, for any $\varepsilon > 0$. The proof is a surprisingly simple reduction from the (strongly NP-complete) *3-partition* problem. By appropriately modifying the proof we show that, in the same range, even the *logarithm* of the distortion cannot be approximated within a factor better than 2 (Theorem 2). This answers an open question posed in [9]. For larger distortions (growing faster than $|S|$) we show a further inapproximability result: The distortion cannot be approximated within a ratio better than $L_T^{1-\varepsilon}$, where L_T is the ratio of the largest to the smallest distance in point set T; we point out that an approximation ratio of L_T^2 is always trivial, in any metric space. We make more precise the inapproximability bounds given in [15]

*Work partially supported by European Commission - Fet Open project DELIS IST-001907 Dynamically Evolving Large Scale Information Systems.

†ETH Zürich, Switzerland, alex.hall@gmail.com

‡UC Berkeley, USA, christos@cs.berkeley.edu

for 3 dimensions and arbitrary distortion, by making explicit the dependency of the inapproximability ratio on the magnitude of the distortion. An overview of our inapproximability results:

Dimension	Distortion	Inapproximable within	
$d \geq 1$	$\|S\| \geq \delta \geq \|S\|^{\varepsilon}$	$\delta^{1-\varepsilon'}$	Thm. 2
	$\delta \geq \|S\|$	$\sqrt{\delta} \cdot \|S\|^{\frac{1}{2}-\varepsilon'}$	
	$\delta \geq \|S\|^{1+\varepsilon}$	$L_T^{1-\varepsilon'}$	
$d \geq 3$	$\delta > 1$	$\sqrt{9 - 8/\delta^2} - \varepsilon'$	Thm. 4
unbounded d	$\delta > 1$	$\delta - \varepsilon'$	Thm. 5

Motivated by these strong inapproximability results, we introduce a novel variant of the problem that we call the *additive distortion problem:* Given two finite points sets $S, T \subseteq \mathbb{R}^d$ with $|S| = |T|$, find the smallest $\Delta \in \mathbb{R}$ (the *additive distortion*) such that there is a bijection $f : S \to T$ with $d(x, y) - \Delta \leq d(f(x), f(y)) \leq d(x, y) + \Delta$, for all $x, y \in S$. By a modification of the Papadimitriou-Safra construction, it is not hard to see that the additive distortion is NP-hard to approximate by a factor better than 3 in 3 dimensions. In Section 3 we present a 2-approximation algorithm for this problem in the 1-dimensional case and a 5-approximation algorithm for the more general case of embedding an arbitrary metric space onto an 1-dimensional point set. Finally, we conclude by pointing out several open questions raised by this work.

Remark. The first three and the last inapproximability results can be strengthened by a power of 2, if we impose the stronger restriction of $\operatorname{expansion}(f) \leq \sqrt{\delta}$ and $\operatorname{expansion}(f^{-1}) \leq \sqrt{\delta}$ (which implies a distortion of $\leq \delta$). In particular the third bound becomes $L_T^{2-\varepsilon'}$, which is near optimal since a ratio of L_T^2 is trivial also in this more restricted setting.

Related Work. Especially in view of the drastic increase in the number of publications concerning embeddings of metric spaces, it is astonishing that the low distortion problem was only introduced very recently [9]. Most Computer Science related work in this area focuses on the setting where a given finite metric space is to be embedded into an *infinite* host space, usually a low dimensional Euclidean space. Although methods from this area do not seem to apply to our setting of embedding a finite metric onto another finite metric, we give a brief overview of related work.

From a theoretical point of view there has been a large interest in finding worst case bounds for the distortion of embedding a class of metrics, e.g. see the surveys [8, 11, 13]. The problem of finding a good embedding for a given metric ("good" compared to an optimal embedding of this, same metric) is practically more relevant and consequently the vast majority of research in this area—also referred to as multi-dimensional scaling—has been on devising good heuristics. See the web-page of the working group on multi-dimensional scaling [14] for an overview and an extensive list of references. An important theoretical result is Linial et al.'s [12] adaption of Bourgain's construction [2]. They present an approximation algorithm based on semidefinite programming for finding a minimum distortion embedding of a given finite metric. Kleinberg et al. [10] consider approximate embeddings of metrics for which only a small subset of the distances are known. Slivkins [16] recently can improve on the results. Also recently Bădoiu et al. [4] give several approximation algorithms for low distortion embeddings of metrics into $\mathbb{R}^1$ and $\mathbb{R}^2$. The notion of *additive* distortion has also been considered for the case where a finite metric is to be embedded into an infinite host space. Håstad et al. [7] give a 2-approximation for the case of embedding into $\mathbb{R}$ and prove that the problem cannot be approximated within $4/3$, unless P = NP. Later Bădoiu [3] and Bădoiu et al. [5] gave an approximation algorithm and a weakly quasi-polynomial time algorithm, respectively, for 2 dimensions.

Other research loosely related to the distortion problem is on the minimum bandwidth problem (see e.g. [6]) and the maximum similarity problem [1].

2 The Inapproximability of Distortion

2.1 NP-hardness

Theorem 1. *The distortion problem is* NP-*hard for* 1 *dimension and any fixed* $\delta \geq |S|^{\varepsilon}$, *for any constant* $\varepsilon > 0$.

Proof. We reduce the well known *3-partition* problem to it. In this problem we are given a set A of $3n$ items $A = \{1, \ldots, 3n\}$ with associated sizes $a_1, \ldots, a_{3n} \in \mathbb{N}$, and a bound $B \in \mathbb{N}$, with $B/4 < a_i < B/2$, for each i, and $\sum_{i=1}^{3n} a_i = n\,B$, and we must decide whether A can be partitioned into n disjoint sets $I_0, \ldots, I_{n-1}$ such that $\sum_{i \in I_j} a_i = B$, for $j = 0, \ldots, n-1$. Note that due to the bounds for the item sizes a_i, all sets I_j must have cardinality 3.

We now describe how to construct the point sets S and T on the line (see the left part of Figure 1). The point set S consists of $3n$ blobs of points $S_1, \ldots, S_{3n}$, where blob S_i has a_i points. Points in a blob are distributed regularly along the line with distance $x := 1/\sqrt{\delta \cdot B}$ from one point to the next. The blobs themselves are also distributed regularly along the line with distance 1 from one to the next, i.e. two neighboring points in different blobs have distance 1. The point set T is very similar: it consists of n blobs $T_1, \ldots, T_n$ of size B. Here the distance of neighboring points within a blob is $1/\sqrt{B}$ and if they are in different blobs it is again 1. Finally, we add two points to both S and T, far away from the blobs. Their distance in T is 1 and their distance in S is $\sqrt{\delta}$ (clearly the points can be added such that they need to be mapped onto each other in order to obtain distortion $\leq \delta$). This ensures that $\mathrm{expansion}(f^{-1}) = \max_{x,y \in S} \left(\frac{d(x,y)}{d(f(x),f(y))} \right) \geq \sqrt{\delta}$.

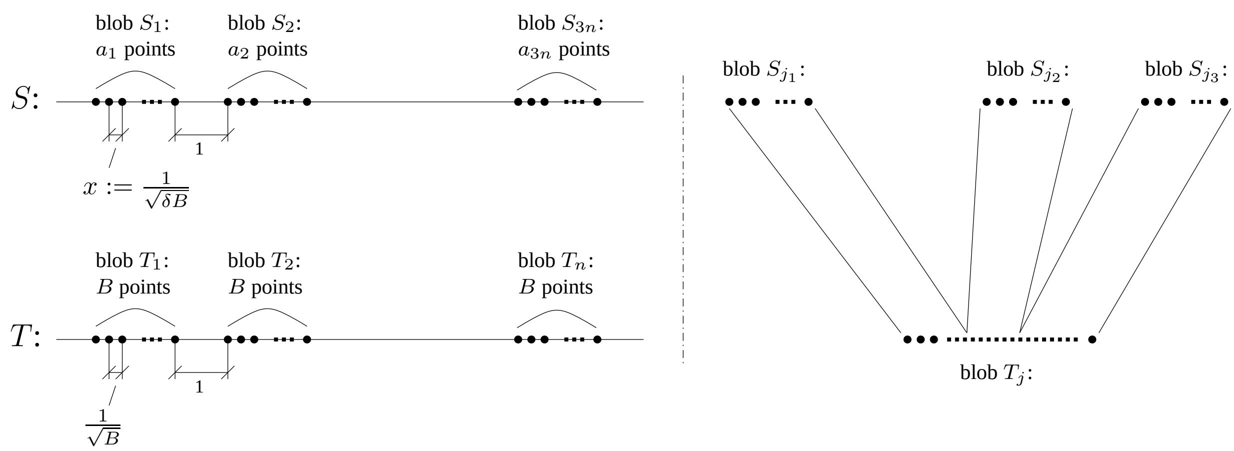

Figure 1: Left: the point sets S and T on a line constructed from a given instance of 3-partition. Right: the mapping of three blobs in S to one blob in T.

Our aim is to show how to derive a low distortion mapping of points from S to T given a solution of the 3-partition instance and vice versa, assuming that $\delta \geq c \cdot n^2 \cdot B$, for an appropriately chosen constant $c > 1$. We start with the forward direction. The mapping is straight forward: we map blobs $S_{j_1}, S_{j_2}, S_{j_3}$ to blob T_j (simply one blob next to the other, as shown in Figure 1 on the right), if $I_j = \{j_1, j_2, j_3\}$. In order to see that the distortion δ is not violated we check the largest possible changes in distance (relatively speaking) and show that $\mathrm{expansion}(f) \leq \sqrt{\delta}$ and $\mathrm{expansion}(f^{-1}) \leq \sqrt{\delta}$ hold, leading to three cases:

1. Two neighboring blobs S_a and S_b are spread as far apart as possible. The distance between two points hereby increases from 1 to less than

$$n \cdot \left(B \cdot \frac{1}{\sqrt{B}} + 1 \right) = n \cdot (\sqrt{B} + 1) \leq \sqrt{\delta},$$

   under the assumption $\delta \geq c \cdot n^2 \cdot B$ with some constant c, which is chosen large enough.

2. The leftmost and the rightmost blob in S, i.e. S_1 and S_{3n}, are mapped next to each other into the same blob in T. The distance decreases from less than $n \cdot (B/\sqrt{\delta \cdot B} + 3)$ to $1/\sqrt{B}$. This leads to an upper bound for the relative distance change of

$$\frac{n \cdot \left(\sqrt{\frac{B}{\delta}} + 3 \right)}{\frac{1}{\sqrt{B}}} \leq n\sqrt{B} \left(\frac{1}{\sqrt{c} \cdot n} + 3 \right) \leq \sqrt{\delta},$$

   again with the assumption $\delta \geq c \cdot n^2 \cdot B$ and some appropriately chosen constant c (which clearly exists).

3. What about the distance increase for two points in the same blob? All distances of points within a blob increase by exactly $(1/\sqrt{B})/(1/\sqrt{\delta \cdot B}) = \sqrt{\delta}$.

For the other direction we only need to check that a blob S_i cannot be split up and mapped to two or more different blobs in T.

4. Two neighboring points in a blob in S cannot be mapped to two different blobs in T. The relative increase in distance would be $1/\frac{1}{\sqrt{\delta \cdot B}} > \sqrt{\delta}$. Together with the two "extra" points that force $\mathrm{expansion}(f^{-1}) \geq \sqrt{\delta}$ this would yield a distortion $> \delta$.

This ensures that in a low distortion mapping from S to T always exactly three blobs from S will be mapped to one blob from T, as wanted.

In order to obtain the nice lower bound for δ we now add a huge blob to both of them which is far away from all other blobs and whose points are very close to each other. Clearly this can be done in such a manner that all points in this huge blob in S will be mapped directly to the corresponding huge blob in T; without interfering with the mapping of all other points. We started with $|S| = n \cdot B$ points and increase the number of points in the huge blob until $|S|^{\varepsilon} > c \cdot n^2 \cdot B$. Since ε is a constant the resulting input size will still be bounded by a polynomial. □

2.2 Inapproximability

We now describe how to modify the proof in order to obtain the strong inapproximability results for large δ. Let $L_T := \frac{\max_{x,y \in T} d(x,y)}{\min_{x,y \in T} d(x,y)}$ be the ratio of maximum to minimum distance in T.

Theorem 2. *For any $\varepsilon, \varepsilon' > 0$, the 1-dimensional distortion problem for $|S| \geq \delta \geq |S|^{\varepsilon}$ is inapproximable within a factor of $\delta^{1-\varepsilon'}$, for $\delta \geq |S|$ is inapproximable within a factor of $\sqrt{\delta} \cdot |S|^{\frac{1}{2}-\varepsilon'}$, and for $\delta \geq |S|^{1+\varepsilon}$ is inapproximable within a factor of $L_T^{1-\varepsilon'}$, unless* P = NP.

Proof. First of all we introduce a new distance g to be defined presently and replace the inter-blob distances (the distance of 1 from one blob to the next) in T by this distance g.

From the list of cases in the proof of Theorem 1, case 2. and case 3. remain untouched by this change. For case 1. (where the "1" in the expression is now a "g") we choose c in our assumption $\delta \geq c \cdot n^2 \cdot B$ such that $\sqrt{\delta}/2 \geq n\sqrt{B}$ and we choose g such that $\sqrt{\delta}/2 \geq n \cdot g$. Choosing $g := \sqrt{\delta}/2n$ will be fine.

For case 4. the relative increase in distance between two neighboring points in a blob in S would be at least $y := g/\frac{1}{\sqrt{\delta \cdot B}}$ if they were split onto two blobs in T. This would amount to a distortion of $y \cdot \sqrt{\delta}$ (note that $\mathrm{expansion}(f^{-1}) \geq \sqrt{\delta}$). The optimum solution is only "allowed" a distortion of δ. Thus, unless P = NP, there can be no approximation algorithm with ratio better than

$$\frac{y \cdot \sqrt{\delta}}{\delta} = \frac{g \cdot \sqrt{\delta \cdot B} \cdot \sqrt{\delta}}{\delta} = \sqrt{\delta} \cdot \frac{\sqrt{B}}{2n} \geq \sqrt{\delta} \cdot B^{\frac{1}{2}(1-\varepsilon')}, \tag{1}$$

with our choice of $g := \sqrt{\delta}/2n$ above. For the inequality we make the assumption of $B \geq n^{4/\varepsilon'}$, which yields $\frac{B}{4n^2} \geq \frac{B}{4B^{\varepsilon'/2}} = \frac{1}{4}B^{1-\varepsilon'/2} \geq B^{1-\varepsilon'}$. Making this assumption poses no problem, since B can easily be increased (i.e. by "blowing it up" similarly to what we do with S), if this should not hold. Let us consider the first statement in the theorem.

$\delta \geq |S|^{\varepsilon}, \delta \leq |S|$. We blow up S and T again, as in the proof of Theorem 1, but we proceed more carefully. Before adding any points to the huge blob, we know that $\delta \leq |S| \leq c \cdot n^2 \cdot B$ (the first by assumption, the second since $|S| = n \cdot B$ holds in the beginning). Now we increase the blob and thereby δ (since by assumption $\delta \geq |S|^{\varepsilon}$) until we have equality $\delta = c'' \cdot n^2 \cdot B$ for some appropriately chosen constant $c'' \geq c$. We obtain $\delta \leq 2 \cdot c'' \cdot B^{\varepsilon'/2} \cdot B \leq B^{1+\varepsilon'}$ and with (1) thus

$$\frac{y \cdot \sqrt{\delta}}{\delta} \geq \sqrt{\delta} \cdot B^{\frac{1}{2}(1-\varepsilon')} \geq \sqrt{\delta} \cdot \delta^{\frac{1}{2} \cdot \frac{1-\varepsilon'}{1+\varepsilon'}} \geq \delta^{1-\varepsilon'}.$$

This gives the first bound for the approximation ratio.

$\delta \geq |S|$. For the second bound in the theorem we lower bound B in terms of $|S|$. To obtain a good bound we will blow up the extra blob of S and T only slightly: we increase the blob until $|S| = B^{1+\varepsilon'}$, which is enough to ensure $\delta \geq B^{1+\varepsilon'} \geq c \cdot B^{1+\varepsilon'/2} \geq c \cdot n^2 \cdot B$ as needed. Note that we assumed $B \geq n^{4/\varepsilon'}$, as before. We insert $|S| = B^{1+\varepsilon'}$ into (1):

$$\frac{y \cdot \sqrt{\delta}}{\delta} \geq \sqrt{\delta} \cdot |S|^{\frac{1}{2} \cdot \frac{1-\varepsilon'}{1+\varepsilon'}} \geq \sqrt{\delta} \cdot |S|^{\frac{1}{2}-\varepsilon'}.$$

$\delta \geq |S|^{1+\varepsilon}$. For the third statement in the theorem we will again search for a lower bound of B in (1), but now by an expression in L_T. In this case we will not blow up S and T with an extra blob, but instead stick to the original construction. The two "extra" points which were added to ensure $\mathrm{expansion}(f^{-1}) \geq \sqrt{\delta}$ can be added at distance 1 from the other blobs in S and at distance g from the other blobs in T. Clearly, if a blob S_i is split apart and partly mapped onto these two points, this again yields the distortion given in (1).

For the maximum ratio of distances we have for the set T:

$$L_T \leq \frac{1+g+n\cdot\left(\frac{B}{\sqrt{B}}+g\right)}{\frac{1}{\sqrt{B}}} = (1+\sqrt{\delta}/2n)\sqrt{B}+n\cdot B+\sqrt{\delta\cdot B}/2 \leq c'\cdot\sqrt{\delta\cdot B}\cdot n$$

holds, with $g := \sqrt{\delta}/2n$, our assumption $\delta \geq c\cdot n^2\cdot B$, and an appropriate constant c'. Note that given L_T we need to adjust the minimum size of the input ($|S|$) accordingly.

We replace the assumption made above for the size of B by $B \geq n^{\max\{8/\varepsilon',1/\varepsilon\}}$. The second term in the "$\max$" expression ensures that $c\cdot n^2\cdot B \leq c\cdot n\cdot B^{\varepsilon}\cdot B \leq (n\cdot B)^{1+\varepsilon} \leq |S|^{1+\varepsilon}$. Since $\delta \geq |S|^{1+\varepsilon}$ holds, we obtain $\delta \geq c\cdot n^2\cdot B$ as needed. With help of the first term in the "$\max$" expression we obtain for (1): $\frac{y\cdot\sqrt{\delta}}{\delta} \geq \sqrt{\delta}\cdot B^{\frac{1}{2}(1-\varepsilon'/2)}$. We plug in $\frac{L_T}{\sqrt{\delta}} \leq c'\cdot\sqrt{B}\cdot n \leq c'\cdot\sqrt{B}\cdot B^{\varepsilon'/8} \leq B^{\frac{1}{2}(1+\varepsilon'/2)}$:

$$\sqrt{\delta}\cdot B^{\frac{1}{2}(1-\varepsilon'/2)} \geq \sqrt{\delta}\cdot\left(\frac{L_T}{\sqrt{\delta}}\right)^{\frac{1-\varepsilon'/2}{1+\varepsilon'/2}} \geq (L_T)^{\frac{1-\varepsilon'/2}{1+\varepsilon'/2}} \geq L_T^{1-\varepsilon'}.$$

This concludes the proof. □

In connection to the last bound, the following observation is interesting.

Observation 3. Any *embedding* $f : S \to T$ *has a distortion of at most* $L_T^2\cdot\delta$*, where* δ *is the optimal distortion, even if* (S,d_S) *and* (T,d_T) *are arbitrary metric spaces.*

Proof. Let the longest and shortest distances in the two metrics be denoted by

$$d_S^{max} := \max_{x,y\in S} d_S(x,y)\,, \qquad d_S^{min} := \min_{x,y\in S} d_S(x,y)\,,$$
$$d_T^{max} := \max_{x,y\in T} d_T(x,y)\,, \quad \text{and} \quad d_T^{min} := \min_{x,y\in T} d_T(x,y)\,.$$

Let us consider which distortion these maximum and minimum distances achieve under any bijection f in the best case: d_S^{max} is mapped to d_T^{max} and d_S^{min} to d_T^{min}; any other mapping would in both cases (min and max) lead to larger changes in distances for at least one of the two corresponding partners. From this fact we can derive a lower bound for the optimal distortion of $\delta \geq \max_{a,b\in\{-1,1\}}\left(\frac{d_S^{max}}{d_T^{max}}\right)^a\cdot\left(\frac{d_S^{min}}{d_T^{min}}\right)^b$. Of the four possibilities, we focus on $a=1$ and $b=-1$. In the worst case an embedding f maps the maximum distance in S onto the minimum distance in T and vice versa. This and our lower bound for δ leads to an upper bound for the distortion of any embedding f of

$$\frac{d_S^{max}}{d_T^{min}}\cdot\frac{d_T^{max}}{d_S^{min}} = L_T\cdot\frac{d_S^{max}}{d_S^{min}} \leq L_T\cdot\frac{d_T^{max}}{d_T^{min}}\cdot\delta \leq L_T^2\cdot\delta,$$

giving the upper bound stated above. □

2.3 Higher Dimensions

For the 3-dimensional problem we can show the following explicit dependence of the inapproximability ratio on the distortion.

Theorem 4. *For any fixed $\delta > 1$ it is* NP*-hard to distinguish whether two given 3-dimensional point sets S and T have distortion $\leq \delta$ or $\geq \sqrt{9\delta^2 - 8}$.*

Proof. By a more detailed analysis of a slight modification of the construction in [15] which we omit here. □

Notice that, in view of the previous theorem, this result is relevant when the distortion is small. Finally, when the dimension is unbounded (that is, for general finite metrics), the reduction of the previous subsection can be adapted to establish that the distortion is even harder to approximate:

Theorem 5. *For any fixed $\delta > 1$ it is* NP*-hard to distinguish whether two given finite metrics S and T have distortion $\leq \delta$ or $\geq \delta^2$.*

Proof. (Sketch.) Repeat the construction with all points in the same blob having distance of 1 from each other, while points belonging to different blobs are at distance δ. This holds for both S and T. If a 3-partition exists, then distances of δ are shrunk to 1, but no distances are dilated, and so the distortion is δ. If a 3-partition does not exist, then certain distances are both shrunk and dilated by δ, and so the distortion is δ^2. □

3 Additive Distortion

For the 1-dimensional additive distortion problem we will show that the simplest possible strategy already yields a 2-approximation. The SWEEP-OR-FLIP algorithm either maps the points in $S := \{s_1, \ldots, s_m\} \subseteq \mathbb{R}$ from left to right onto $T := \{t_1, \ldots, t_m\} \subseteq \mathbb{R}$, or flips the point set and maps them from right to left. In other words, we check the bijections $f(s_i) = t_i$ and $f(s_i) = t_{m-i+1}$ and keep the better one. It is easy to see (Figure 2) that this is not optimal. In fact, by choosing $a + \mu = \Delta/2$ in the figure we get a gap of $\frac{a+\Delta-\mu}{\Delta} = 1.5 - 2\mu/\Delta$ when comparing the optimal to the SWEEP-OR-FLIP embedding, for arbitrarily small $\mu > 0$.

For the setting where we are given an arbitrary metric space (S, d_S) and want to embed onto points $T := \{t_1, \ldots, t_m\} \subseteq \mathbb{R}$ we will present a straightforward $5 + \varepsilon$-approximation algorithm.

3.1 A 2-Approximation

Before proving that SWEEP-OR-FLIP is a 2-approximation, we give two definitions:

Crossing Points. Consider two points $x, y \in S$ with $x < y$ and for which their counterparts are $f(x) > f(y)$. We say x and y *cross* in the mapping f.

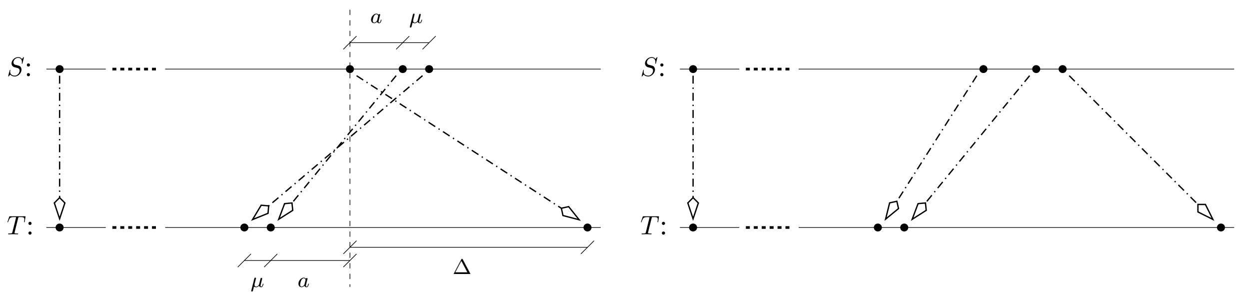

Figure 2: An example showing that the SWEEP-OR-FLIP strategy does not necessarily yield an optimal distortion. The embedding to the left has additive distortion Δ (note that $2(a+\mu) \leq \Delta$). The "left-to-right" embedding to the right has a larger distortion of $a+\Delta-\mu$. Clearly the other direction has a larger distortion as well.

Relative Movement. For fixed S and T, define the *relative movement* of the i-th point to be $\mu_i := t_i - s_i$.

Theorem 6. *The* SWEEP-OR-FLIP *algorithm yields an additive distortion at most* $2 \cdot \Delta$*, where* Δ *is the optimum additive distortion.*

Proof. The proof idea is to fix an optimal embedding f^* and to consider different cases for the relative movements μ_i. We then either show that a distortion of $2 \cdot \Delta$ can be obtained by a "left-to-right" or a "right-to-left" embedding, or arrive at a contradiction by showing that f^* has distortion $> \Delta$. With each of the four steps of the following case analysis we narrow down the situations we need to consider, in terms of the actual relative movements, and also in terms of the mapping of the first and last point in f^*. We start with the latter:

$f^*(s_1) > f^*(s_m)$, i.e. s_1 and s_m cross: If this is the case, we can completely flip T, e.g. by negating all elements of the set without affecting the performance of SWEEP-OR-FLIP. Thus we can assume $f^*(s_1) < f^*(s_m)$.

$|\mu_i| \leq \Delta$, for all $i \in \{1, \ldots, m\}$: Clearly, if we embed left-to-right $f(s_i) = t_i$, the obtained additive distortion is bounded by $2 \cdot \Delta$. To see this take any two points $s_i, s_j \in S$ and note that due to the bounded movement the distance of t_i and t_j can differ by at most $2 \cdot \Delta$ from the distance of s_i and s_j.

$\forall i, j \in \{1, \ldots, m\} : |\mu_i - \mu_j| \leq 2 \cdot \Delta$: Let μ_i be the largest relative movement, and by the previous case assume $\mu_i > \Delta$. Then translate all points in T to the left, until $\mu_i = \Delta$. For all $j \neq i$ and the new relative movements we still have $|\mu_i - \mu_j| \leq 2 \cdot \Delta$. Thus we have $\mu_j \geq -\Delta$ and $\mu_j \leq \mu_i = \Delta$, the former since $\mu_i = \Delta$ and the latter since μ_i is the largest relative movement. In other words, we modified the instance such that the previous case holds. If in the beginning the smallest relative movement is less than $-\Delta$, we proceed analogously translating all points in T to the right.

$\exists i, j \in \{1, \ldots, m\} : |\mu_i - \mu_j| > 2 \cdot \Delta$: Assume $i < j$ and $\mu_i > \mu_j$, otherwise exchange the roles of S and T (the relative movements are negated). Translate all points in T in order to have $\mu_i = \Delta$ and $\mu_j < -\Delta$. Due to a simple counting argument, there must be $k \leq i$ with $f^*(s_k) \geq t_i$ and thus $f^*(s_k) - s_k \geq \mu_i$. Analogously there must be a $l > j$ with $f^*(s_l) \leq t_j$ and thus $f^*(s_l) - s_l \leq \mu_j$. We distinguish the following cases:

s_k and s_l do not cross: We have $d(s_k, s_l) = s_l - s_k \geq s_j - s_i$ and $d(f^*(s_k), f^*(s_l)) \leq t_j - t_i$. This gives a contradiction: $d(s_k, s_l) - d(f^*(s_k), f^*(s_l)) \geq \mu_i - \mu_j > 2 \cdot \Delta$. See the top picture of Figure 3 for an example.

s_k and s_l cross while s_m and s_k do not: See the middle part of Figure 3. Since f^* projects s_k by at least $\mu_i = \Delta$ to the right, we must have $f^*(s_m) \geq s_m$, otherwise the distance $d(f^*(s_k), f^*(s_m))$ would be less than $d(s_k, s_m) - \Delta$. Similarly since f^* projects l by at least $\mu_j < -\Delta$ to the left, we must have $f^*(s_m) < s_m$, which gives the contradiction.

s_k and s_l cross while s_1 and s_l do not: Analogous to the previous case.

$k > 1$ and $l < m$, s_k and s_l cross, s_1 crosses s_l, s_m crosses s_k: See the bottom part of Figure 3. Let us start by making sure that this is the only case left. Due to the very first case we know that s_1 and s_m *do not* cross. Since s_k and s_l *do* cross, either $k > 1$ or $l < m$ must hold. Assume the former, then since s_1 crosses s_l (which it must due to the previous case) we also have the latter (and again due to the last but one case, s_m must cross s_k).

Now we consider the distance $a, b, c, d, e,$ and f as given in Figure 3, bottom. Since we have an additive distortion of Δ,

$$d + e + f \leq b + \Delta \tag{2}$$

must hold. Similarly we have $f \geq b + c - \Delta$, $d \geq a + b - \Delta$, and $e \geq a + b + c - \Delta$, which together gives

$$d + e + f \geq 2a + 3b + 2c - 3\Delta.$$

Subtracting (2) we obtain the contradiction

$$0 \geq 2s + 2b + 2c - 4\Delta \geq 2b - 4\Delta > 0.$$

The last inequality holds since $b \geq \mu_i - \mu_j > 2 \cdot \Delta$. We conclude that there cannot be $i, j \in \{1, \ldots, m\}$ with $|\mu_i - \mu_j| > 2 \cdot \Delta$.

This gives the stated result. □

3.2 Embedding an Arbitrary Metric Space to 1D

We are given an arbitrary finite metric space (S, d_S) and $T \subseteq \mathbb{R}$. The algorithm INTERVALS below finds a mapping of the points in S to the points in T within a factor of 5 of the optimum additive distortion. For ease of exposition we start by assuming that we know the optimum distortion Δ. Below we note why the algorithm also works for the case where the distortion is not given. We also assume that we know the point $x \in S$ mapped to t_1—in fact, we iterate over all $x \in S$.

Feasible Intervals. For $y \in S \setminus \{x\}$ we then define its feasible interval as: $I_y := [t_1 + d_S(x, y) - \Delta, t_1 + d_S(x, y) + \Delta]$. The additive distortion for the pair x, y is $\leq \Delta$ if and only if $f(y) \in I_y$.

ALGORITHM: INTERVALS

---

1. Given $\Delta > 0$ and $x \in S$, compute the feasible intervals I_y for $y \neq x$, and map the remaining nodes of S as follows:

2. Process the feasible intervals by increasing left boundary. For each interval I_y we map y greedily to the leftmost point z in I_y that has not yet been mapped to.

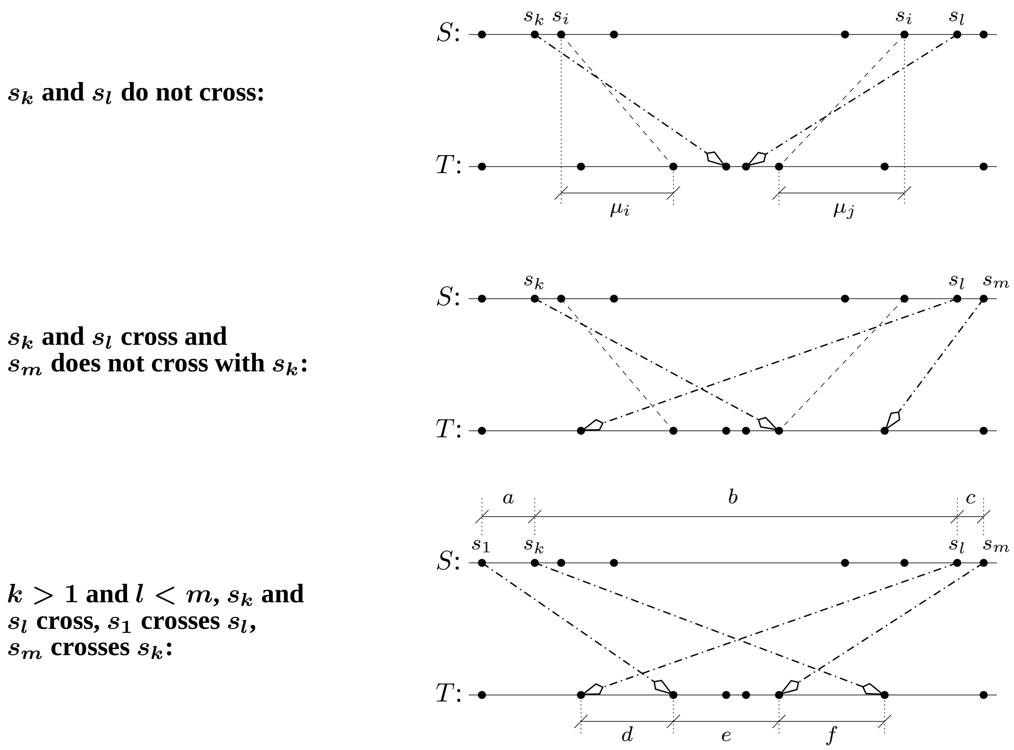

Figure 3: Three examples for the cases concerning that $i, j \in \{1, \ldots, m\}$ exist such that $\mu_i = \Delta$ and $\mu_j < -\Delta$.

Theorem 7. *The* INTERVALS *algorithm yields an additive distortion of* $5 \cdot \Delta$, *where* Δ *is the optimum additive distortion.*

Proof. Consider the mapping f^* that achieves Δ, and assume $f^*(x) = t_1$. Since there is a bijection that maps each $y \neq x$ to a point $z \in I_y$ (f^* is an example of such a bijection), and since all intervals have the same length, it is clear that the algorithm will find such a bijection, call it f. f can increase the distance between any point pair by at most $4 \cdot \Delta$, since the intervals have a width of $2 \cdot \Delta$. Therefore, the additive distortion of f is within a factor of 5 of the additive distortion. □

If INTERVALS succeeds in finding a mapping, it will simply do a left to right mapping of the remaining points $S \setminus \{x\}$ in step 2. Therefore, by iterating over all $x \in S$ and checking for each x the left to right mapping of the rest, we find the same mapping as INTERVALS without knowing Δ in advance.

4 Open Problems

We have made significant progress towards understanding the complexity of computing the distortion of bijections between point sets. But many open problems remain:

- What is the complexity of the distortion problem on the line for large constant values of distortion? The dynamic programming approach seems to exhaust itself after $3 + 2\sqrt{2}$, yet NP-completeness also seems very difficult.

- In view of our results, it seems that, for large distortions, the right quantity to approximate is not δ but $\log \delta$. By Theorem 2 we know that it cannot be approximated by a factor better than 2. Is a constant factor possible? Or is there a generalization of our proof (by some kind of hierarchical 3-partition problem) that shows impossibility?

- In connection to the last open problem, we may want to define the following relaxation of the distortion problem: We are given, besides S, T, and δ, an $\varepsilon > 0$, and we are asked whether there is a bijection between *all but an ε fraction* of S and T such that the distortion of this partial map is δ or less. We conjecture that for any ε there is a polynomial algorithm that approximates $\log \delta$ by a factor of 2.

- Are there better approximation algorithms for the 1-dimensional additive distortion problem? And can one prove the 1-dimensional problem to be NP-complete?

References

[1] T. Akutsu, K. Kanaya, A. Ohyama, and A. Fujiyama, *Point matching under non-uniform distortions*, Discrete Applied Mathematics, 127 (2003), pp. 5–21.

[2] J. Bourgain, *On lipschitz embedding of finite metric spaces into hilbert space*, Isreal Journal of Mathematics, 52 (1985), pp. 46–52.

[3] M. Bădoiu, *Approximation algorithm for embedding metrics into a two-dimensional space*, in Proceedings of the 14th SODA, 2003, pp. 434–443.

[4] M. Bădoiu, K. Dhamdhere, A. Gupta, Y. Rabinovich, H. Rcke, R. Ravi, and A. Sidiropoulos, *Approximation algorithms for low-distortion embeddings into low-dimensional spaces*, in Proceedings of the 16th SODA, 2005, pp. 119–128.

[5] M. Bădoiu, P. Indyk, and Y. Rabinovich, *Approximate algorithms for embedding metrics into lowdimensional spaces*, Manuscript, (2003).

[6] U. Feige, *Approximating the bandwidth via volume respecting embeddings*, Journal of Computer and System Sciences, 60 (2000), pp. 510–539.

[7] J. Håstad, L. Ivansson, and J. Lagergren, *Fitting points on the real line and its application to RH mapping*, Lecture Notes in Computer Science, 1461 (1998), pp. 465–467.

[8] P. Indyk, *Algorithmic applications of low-distortion geometric embeddings*, in Tutorial at the 42nd FOCS, 2001, pp. 10–33.

[9] C. Kenyon, Y. Rabani, and A. Sinclair, *Low distortion maps between point sets*, in Proceedings of the 36th STOC, 2004, pp. 272–280.

[10] J. Kleinberg, A. Slivkins, and T. Wexler, *Triangulation and embedding using small sets of beacons*, in Proceedings of the 45th FOCS, 2004, pp. 444–453.

[11] N. Linial, *Finite metric spaces — combinatorics, geometry and algorithms*, in Proceedings of the International Congress of Mathematicians III, Beijing, 2002, pp. 573–586.

[12] N. Linial, E. London, and Y. Rabinovich, *The geometry of graphs and some of its algorithmic applications*, Combinatorica, 15 (1995), pp. 215–245.

[13] J. Matousek, *Lectures on Discrete Geometry*, Springer-Verlag, Graduate Texts in Mathematics, Vol. 212, 2002.

[14] W. page of the working group on multi-dimensional scaling. http://dimacs.rutgers.edu/ SpecialYears/ 2001_ Data/Algorithms/AlgorithmsMS.htm.

[15] C. Papadimitriou and S. Safra, *The complexity of low-distortion embeddings between point sets*, in Proceedings of the 16th SODA, 2005, pp. 112–118.

[16] A. Slivkins, *Distributed approaches to triangulation and embedding*, in Proceedings of the 16th SODA, 2005, pp. 640–649.