Electronic Colloquium on Computational Complexity, Report No. 110 (2006)

Improved Algorithms for Optimal Embeddings

Nishanth Chandran   Ryan Moriarty   Rafail Ostrovsky   Omkant Pandey   Amit Sahai

Department of Computer Science, UCLA
{nishanth,ryan,rafail,omkant,sahai}@cs.ucla.edu

Abstract

In the last decade, the notion of metric embeddings with small distortion received wide attention in the literature, with applications in combinatorial optimization, discrete mathematics, functional analysis and bio-informatics. The notion of embedding is, given two metric spaces on the same number of points, to find a bijection that minimizes maximum Lipschitz and bi-Lipschitz constants. One reason for the popularity of the notion is that algorithms designed for one metric space can be applied to a different metric space, given an embedding with small distortion. The better the distortion, the better is the effectiveness of the original algorithm applied to a new metric space. The goal that was recently studied by Kenyon, Rabani, and Sinclair [14] is to consider all possible embeddings and to try to find the best possible one, i.e., consider a single objective function over the space of all possible embeddings that minimizes the distortion. In this paper we continue this important direction. In particular, we resolve an open problem stated by the previous paper, improve their distortion lower bound for the line, generalize their method and show its inherent limitation. While the improved distortion differs only by a constant factor for the line (i.e., we show an improvement from $3+2\sqrt{2}$ to $5+2\sqrt{6}$), it requires novel techniques and insight into high-distortion patterns. Furthermore, we show an inherent limitation of this method to be at most $7+4\sqrt{3}$. We also show that previous techniques on general embeddings apply to a more general class of metrics.

1 Introduction

For a bijection $\sigma : U \to V$ between two n-point metric spaces (U, d) and (V, d'), expansion of σ is defined as

$$\text{expansion}(\sigma) = \max_{x,y\in U, x\neq y} \frac{d'(\sigma(x), \sigma(y))}{d(x, y)}$$

Distortion σ is defined as follows: $\text{dist}(\sigma) = \text{expansion}(\sigma) \times \text{expansion}(\sigma^{-1})$. The MINIMUM DISTORTION problem is to find a bijection σ between two equal-sized finite metric spaces (U, d) and (V, d') such that $\text{dist}(\sigma)$ is minimum over all possible bijections.

The MINIMUM DISTORTION problem is interesting to study for both theoretical as well as practical reasons. From complexity theoretic point of view, it has interesting connections to GRAPH ISOMORPHISM [10]. In particular, graph isomorphism on two input graphs G and H is trivially reduced to deciding if there exists an isometric (i.e., distortion 1) bijection between $\mathcal{M}_G$ and $\mathcal{M}_H$, where $\mathcal{M}_X$ denotes the shortest path metrics of a graph X.

On the practical side, we note that applications dealing with shape matching and object recognition (e.g., signature matching, character recognition, matching facial features, pattern matching in complicated protein structures, and so on) require good measures of similarity. Distortion is an attractive measure of similarity between two point sets [2, 12]. In fact, among currently known measures of similarity [5, 7], it seems to be the most sensitive measure. Thus, from the point of view of aforementioned applications, good algorithms for finding minimum distortion (interchangeably, optimal) bijection are highly desirable.

Kenyon, Rabani, and Sinclair [14] show that the MINIMUM DISTORTION problem is NP-hard even to approximate (within a factor of 2), and provide two positive results:

ISSN 1433-8092

- A polynomial time algorithm for *exactly* finding the minimum distortion bijection between two line metrics if the optimal bijection has distortion strictly less than $3 + 2\sqrt{2}$.

- A parameterized polynomial time algorithm for *exactly* finding optimal bijection between bounded-degree tree metric and an arbitrary unweighted graph metric.

In this paper, we improve and generalize the results of Kenyon, Rabani, and Sinclair.

- In particular, we first provide a polynomial time algorithm for *exactly* finding an optimal bijection between two line metrics if the optimal bijection has distortion strictly less than $5 + 2\sqrt{6}$.

  Achieving this improvement requires new insights into the problem. In particular, [14] look at a single "pattern" (partial bijection of size 4) and its inverse. They call this pattern a forbidden pattern. Optimizing over those bijections that do not contain this forbidden pattern gives rise to the algorithm of [14] and its analysis. We develop a generalized definition of forbidden patterns. Our definition classifies patterns of increasing size (depending on a parameter k) with high distortion. We call these patterns (k, k)-FP. We show that one must exclude *all* forbidden patterns of size k to guarantee the correctness of the algorithm for higher distortion. We use this generalized methodology for constructing parameterized algorithms and analyzing them. Interestingly, the forbidden pattern (and its inverse) considered by [14] comprises the entire $(4, 4)$-FP family.

- Next, based on the idea of families of forbidden patterns, we are able to design a dynamic programming algorithm which finds a minimum distortion bijection on more instances than [14]. Thus our work answers a direct open question posed in [14].

- We also show a limitation of the "forbidden pattern" approach, by showing that there exists arbitrarily large families of forbidden patterns with bounded distortion. This lower bound shows the extent to which this approach will be useful and indicates a new approach must be taken to pass this bound.

We remark that after the work of [14], most research has focussed on either approximating the distortion [4, 3] or proving the hardness of approximating it [11, 18, 6]. Hall and Papadimitriou in [11], show that line embeddings are hard to approximate even within large factors when the distortion is high. There have been no positive results for polynomial time algorithms that *exactly* find the minimum distortion bijection; we are the first such improvement.

We also consider the case of embedding a bounded degree unweighted tree metric into an arbitrary unweighted graph metric. We prove that the algorithm of [14] actually works for a larger class of graphs - unweighted bounded degree graphs with maximum cycle length 3. That is, we show that their algorithm finds optimal bijection between a bounded degree graph with maximum cycle length 3 and an arbitrary unweighted graph metric.

1.1 Related Work

The problem of embedding distance metrics into geometric spaces has been studied extensively [15, 16, 19, 20, 1, 13, 17]. The MINIMUM DISTORTION problem is a natural variant of bi-Lipschitz embeddings questions that were initially motivated by the study of Banach spaces.

A problem closely related to MINIMUM DISTORTION is MINIMUM BANDWIDTH. MINIMUM DISTORTION can be viewed as a variation and generalization of the MINIMUM BANDWIDTH problem [8, 9]. Good solutions for the MINIMUM BANDWIDTH problem, however, typically incur very large contraction and hence do not seem useful for solving MINIMUM DISTORTION.

After its introduction, MINIMUM DISTORTION problem has received considerable attention in the research community. Most of the results, however, have been negative showing that the problem is hard even to approximate. Among such results are the following: Hall and Papadimitriou show that the line embeddings problem is hard to approximate even within large factors if distortion is high [11], Papadimitriou and Safra show that the general embeddings problem is hard to approximate within a factor of 3 in three-dimension [18], and some results of Cary, Rudra, and Sabharwal [6].

Due to such results, some of the research work focusses on approximating MINIMUM DISTORTION under certain circumstances (e.g., see the work in [3, 4]) and considering new definitions of distortion (e.g. additive distortion [11]). To the best of our knowledge, there have been no positive results on exactly solving the embeddings problem with multiplicative distortion as the measure of similarity. After [14], our results are the only positive results for finding optimal embeddings.

2 Line Embeddings

This section proves the following main result.

Theorem 1 *Let $U = \{u_1, ..., u_n\}$ and $V = \{v_1, ..., v_n\}$ be two subsets of the real numbers. Let $\alpha < \sqrt{5 + 2\sqrt{6}}$. There is a polynomial time algorithm to decide whether there exists a bijection between U and V with expansion and inverse expansion at most α.*

From this, we obtain a polynomial time algorithm for the minimum distortion problem in one dimension provided the distortion is below the threshold value $5 + 2\sqrt{6}$.

2.1 Definitions

Let $U = \{u_1, u_2, \ldots, u_n\}, V = \{v_1, v_2, \ldots, v_n\}$ be the sets of n points on the real line, and let π be a fixed permutation on $\{1, 2, \ldots, k\}$ (a "pattern"). Following [14], we say that a bijection σ from U to V *contains* π if there exists a subset $u_{a_1} < u_{a_2} < \ldots < u_{a_k}$ such that $\sigma(u_{a_i}) < \sigma(u_{a_j})$ iff $\pi(i) < \pi(j)$, for $1 \leq i, j \leq k$. Otherwise, we say that σ *avoids* π.

Line. The image of a point $x \in U$ is denoted by x'. A *line* is defined as a point x along with its image x'. It is denoted by x-x'.

Neighbor. A point that is immediately to the left (right) of $x \in U$ is called the *left (right) neighbor* of x.

Consecutive Mapping. It is a mapping from $m \geq 2$ consecutive points in U to m consecutive points in V.

By σ_K , we denote the mapping of the points in the set K. By $\sigma_{K+K'}$, we denote the mapping of the points in the set $K \cup K'$. If $K' = \{a\}$, we write σ_{K+a}. Similarly by $\sigma_{K-K'}$, we denote the mapping of the points in the set $K - K'$. The size of σ_K (denoted by $|\sigma_K|$) is defined as the number of lines in σ_K .

Encompass and Include. Consider the mapping σ_K . Let $K = \{u_1, u_2, \ldots, u_k\}$ such that $u_1 < u_2 < \ldots < u_k$. Let the leftmost image of all points in K be u'_ℓ and the rightmost image be u'_r. Then, the mapping σ_K *encompasses* any line x-x', $x \notin K$, such that $u_1 \leq x \leq u_k$ or $u'_l \leq x' \leq u'_r$. If $x \in K$ then we say that σ_K includes x-x'.

If σ_K encompasses (includes) all lines x-$x' \in \sigma_{K'}$, we say that σ_K encompasses (includes) $\sigma_{K'}$. We shall sometimes abuse notation by saying σ_K encompasses (includes) x, when we mean σ_K encompasses (includes) x-x'.

Border. We say a line $x - x'$ from U to V borders a mapping σ_K if x and x' are both neighbors of points in σ_K , but are not encompassed by σ_K .

Forbidden Pattern. Consider a bijection $\sigma : U \to V$. Pick the smallest subsequence S in U of size $i > 1$, such that $\sigma(S)$ is a subsequence of V. Replace subsequences S and $\sigma(S)$ with single points in their respective sets U and V.

Repeat the above *reduction* process until there is only one point in U. If k is the size of the longest subsequence S in this process, then bijection σ is a (n, k)*-forbidden-pattern*, denoted by (n, k)-FP, and a (n, k_0)*-non-forbidden-pattern* for all $k_0 > k$, denoted by (n, k_0)-NFP.

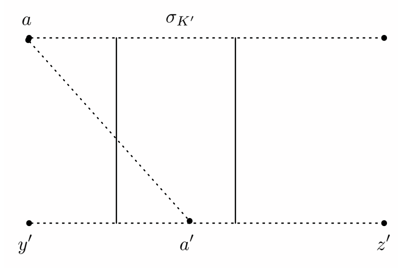


Figure 1:

2.2 Structural Properties

Lemma 1 *For every (k,k)-FP σ_K , which is a mapping from U to V with $k \geq 5$, with endpoints y' and z' in V, there exists a (k_0,k_0)-FP with $k_0 \geq k-4$ that is contained in σ_K and includes $z-z'$ and $y-y'$.*

Proof. Base case: It is easy to see that for every $(5,5)$-FP there exists a $(4,4)$-FP pattern that is contained in the $(5,5)$-FP which contains both endpoints in V (see appendix A.1).

Assume for induction that the lemma holds up to $k-1$. We will show with this inductive hypothesis that we can prove the lemma for k.

Suppose we are given a (k,k)-FP σ_K from U to V. Let y' and z' be the left and right end points of V respectively. Also let a be the left most point in U.

Look at the mapping σ_{K-a} . Let $\sigma_{K'}$ be a consecutive mapping in σ_K such that $|\sigma_{K'}| \geq 2$ and $|\sigma_{K'}|$ is minimal (see figure 1). $\sigma_{K'}$ will obviously always exist and may equal σ_{K-a} .

Any consecutive mapping in σ_{K-a} must encompass a' in the original mapping σ_K . If these consecutive mappings did not encompass a', the original pattern would have $< k$ consecutive points mapping to $< k$ consecutive points. This is not the case by the definition of a (k,k)-FP. Thus any consecutive mapping must encompass a'.

Any consecutive mapping that encompasses a' will include some points in $\sigma_{K'}$. A mapping that includes any of $\sigma_{K'}$ must include all of $\sigma_{K'}$ because there is no consecutive mapping in $\sigma_{K'}$ of size $m < |\sigma_{K'}|$. Thus any consecutive mapping in σ_{K-a} that includes a point in $\sigma_{K'}$ will include all of $\sigma_{K'}$. Since all mappings must include $\sigma_{K'}$ and $\sigma_{K'}$ is the smallest such mapping, this implies $\sigma_{K'}$ is unique.

Note that in the previous argument we implied that there are *no* consecutive mappings in $\sigma_{K-K'-a}$ that do not encompass $\sigma_{K'}$.

Suppose $|\sigma_{K'}| \geq 5$, then by inductive hypothesis $\sigma_{K'}$ contains at least a $(|\sigma_{K'}|-4, |\sigma_{K'}|-4)$-FP that contains both endpoints of $\sigma_{K'}$ in V; call this mapping σ_{K_r} . Remove all lines in $\sigma_{K'}$ that are not in σ_{K_r} . Note that since we have kept the endpoints of $\sigma_{K'}$ we have certainly kept the endpoints of σ_K . It is also obvious that a' is encompassed by σ_{K_r} since we have kept both endpoints of $\sigma_{K'}$.

Now look at the mapping $\sigma_{K-K'+K_r}$. We claim that this is the (k_0,k_0)-FP. Assume it is not a (k_0,k_0)-FP. Then there is some consecutive mapping of size less than k_0, call this mapping $\sigma_{K''}$. We have shown that there are no consecutive mappings in $\sigma_{K-K'-a}$ that do not encompass $\sigma_{K'}$, $\sigma_{K''}$ must include either a or $\sigma_{K'}$ or both.

If $\sigma_{K''}$ includes $a-a'$ it must include a neighbor of a', but both neighbors of a' are in σ_{K_r} , since σ_{K_r} is a forbidden pattern we must include all points in σ_{K_r} Similarly if we include any point in σ_{K_r} in $\sigma_{K''}$ we must include all points in σ_{K_r} and thus must include $a-a'$. Thus $\sigma_{K''}$ must include both $a-a'$ and σ_{K_r} .

If we now replace σ_{K_r} with $\sigma_{K'}$ in $\sigma_{K''}$, the size of $\sigma_{K''}$ has only increased by at most 4, since all removed points are encompassed by $\sigma_{K''}$. This consecutive mapping is in σ_K and is of size less than k which is a contradiction. Thus $\sigma_{K-K'+K_r}$ is a (k_0,k_0)-FP and includes points y' and z'.

Now we must handle all cases where $|\sigma_{K'}| < 5$.

Size of $\sigma_{K'}$ = 3 or 4. We will construct a mapping that is a (k_0,k_0)-FP as follows. Start with σ_K .

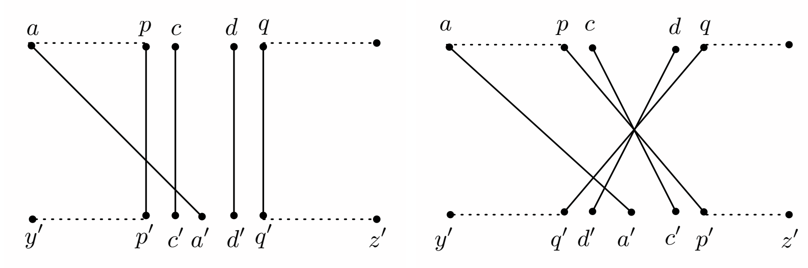


Figure 2: $|\sigma_{K'}| = 3$ or 4

Remove all points in $\sigma_{K'}$ keeping the two end points of $\sigma_{K'}$ in V. Call these endpoints c' and d' such that c is to the left of d in U. Note that since we have kept the end points of $\sigma_{K'}$ we have obviously kept the endpoints of σ_K .

There might exist a line $p - p'$ such that p is the left neighbor of c and p' is a neighbor of c', if so remove $c - c'$ and rename $p - p'$ as $c - c'$ for convenience. Similarly there might exist a line $q - q'$ such that q is the right neighbor of d and q' is the neighbor of d', if so remove $d - d'$ and rename $q - q'$ as $d - d'$ (see figure 2). Note that we still maintain the endpoints of σ_K in V.

Note that c does not form a consecutive mapping with its new left neighbor. If it could have then p would have formed a consecutive mapping with this new neighbor. This would have been a consecutive mapping in $\sigma_{K-K'-a}$ that did not encompass $\sigma_{K'}$, which we have shown can not be the case. Similarly d can not form a consecutive mapping with its new right neighbor.

We now claim that the resulting mapping is a (k_0, k_0)-FP that includes both end points in V. Clearly the pattern contains both endpoints in V. Suppose it is not a (k_0, k_0)-FP then there exists a consecutive mapping $\sigma_{K''}$ of size less than k_0.

- Suppose $\sigma_{K''}$ does not contain both c-c' and d-d'. Then $\sigma_{K''}$ must contain two points from $\sigma_{K-K'-a}$ and hence the mapping must include $c-c'$ and $d-d'$ since there is no consecutive mapping in $\sigma_{K-K'-a}$ that does not encompass $\sigma_{K'}$. If this is the case then we can put back all the points which we removed. The lines we removed do not encompass any points not in $\sigma_{K''}$, because they are all in $\sigma_{K'}$ and $\sigma_{K''}$ encompasses $\sigma_{K'}$. Thus when we put back the removed lines we increase the size of the consecutive mapping by at most 4. If $|\sigma_{K''}| < k_0$ then this is a consecutive mapping in σ_K of size less than k which is a contradiction. Thus if $\sigma_{K''}$ contains two points from $\sigma_{K-K'-a}$ we will have reached a contradiction.

- Suppose $\sigma_{K''}$ contains both $c - c'$ and $d - d'$. Then we are done by the same argument as above.

- Suppose $\sigma_{K''}$ contains either $c - c'$ or $d - d'$. If $\sigma_{K''}$ contains $c - c'$ and not $d - d'$ it must include two points outside of $\sigma_{K'}$. This is because $c - c'$ can not merge with its left neighbor. If one of these lines is $a - a'$, note that we can now include $d - d'$ without including any more points and we still have a consecutive mapping. Thus $c - c'$ and $d - d'$ are part of a consecutive mapping and we are done. If one of these lines is not $a - a'$ then $\sigma_{K''}$ contains two points from $\sigma_{K-K'-a}$ and we are done. If $\sigma_{K''}$ contains $d - d'$ and not $c - c'$ a similar argument holds.

Size of $\sigma_{K'} = 2$. We will construct a mapping that is a (k_0, k_0)-FP as follows. Let the lines in $\sigma_{K'}$ be $c - c'$ and $d - d'$ respectively from left to right in U. We show that the removal of either line $c - c'$ or line $d - d'$ will result in a mapping that will contain a (k_0, k_0)-FP.

Let e be the left neighbor of c if one exists. If c does not have a left neighbor then $a = c$. However if $a = c$ or $a = e$ then $\sigma_{K'+a}$ is a consecutive mapping of σ_K . This can not be the case since σ_K is a (k, k)-FP and $|\sigma_{K'}| = 2$. Hence $a \neq c$ and $a \neq e$. Thus e must exist and not equal a.

Note for later use that $c - c'$ and $e - e'$ can not form a consecutive mapping since c and e are neighbors in σ_K .

Let f be the right neighbor of d if one exists.

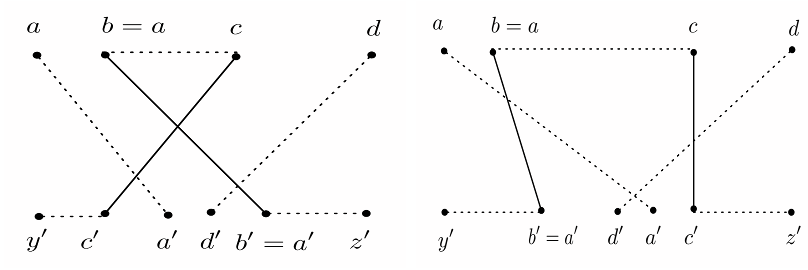


Figure 3: $|\sigma_{K'}|=2$

Suppose f does not exist. If this is the case then d is an endpoint in U (see figure 3). d' can not be an endpoint in V, since if this was the case then σ_{K-d} would have been a consecutive mapping in σ_K of size $k-1$.

Thus we will remove $d-d'$. Let b be the right neighbor of a. If the right neighbor of b is c, $|\sigma_K|$ would be at most 4, but $|\sigma_K|>4$. b can not form a consecutive mapping with its right neighbor since in that case b and its right neighbor form a consecutive mapping in $\sigma_{K-K'-a}$. It may be the case that b' was a neighbor of d'. If this were the case then after the removal of $d-d'$, $a-a'$ and $b-b'$ will form a consecutive mapping. In this case remove $a-a'$ and rename $b-b'$ as $a-a'$. Note that a' was not an endpoint in V and thus we still have both endpoints of V. Now $a-a'$ can not form a consecutive mapping with its right neighbor.

Call the remaining mapping σ_{K_N} we claim that σ_{K_N} is a (k_0,k_0)-FP that includes both endpoints of σ_K in V. Clearly both endpoints in V are included. Suppose σ_{K_N} is not a (k_0,k_0)-FP then there exists a consecutive mapping, $\sigma_{K''}$ in σ_{K_N} of size less than k_0.

Suppose $\sigma_{K''}$ contains two points in $\sigma_{K-K'-a}$. Then $\sigma_{K''}$ encompasses $\sigma_{K'}$ and hence it must include $c-c'$. This is because there were no points to the right of d and thus the mapping must encompass $\sigma_{K'}$ in V. This also means that the mapping must include $a-a'$ because either $a-a'$ is encompassed by $\sigma_{K'}$ or is a neighbor of $\sigma_{K'}$ in V. Thus adding $a-a'$ to the mapping if it was removed will increase the mapping size by one. Also adding back $d-d'$ will increase the mapping by at most one since it is part of $\sigma_{K'}$. By a similar argument as we used earlier we will reach a contradiction.

Suppose $\sigma_{K''}$ includes $c-c'$. Thus it must include two points not in $\sigma_{K'}$ since c has no right neighbor and it can not merge with its left neighbor. If one of these points is a then we can again add back all removed lines without increasing the size of this mapping by more than two. And we can again reach a contradiction. If neither of the points is a then we have included two points in $\sigma_{K-K'-a}$ and we have a contradiction.

Suppose $\sigma_{K''}$ includes $a-a'$. Since a can not merge with its neighbor then we must include two points to the right of a. If one of these points is c we are done. If neither of the points is c then we have included two points in $\sigma_{K-K'-a}$ and we are done.

Now the only case left is when f (the right neighbor of d) exists. Due to space limitations, the details of this case are given in appendix A.2. ■

Now we can prove the main theorem of this section:

Theorem 2 *The minimum distortion over all* (n,k)-FP $\geq$ *the minimum distortion over all* $(k-4,k-4)$-FP.

PROOF. Suppose we are given a (n,k)-FP. By definition, this contains a (k_0,k_0)-FP ($k_0\geq k$). This pattern contains at least a (k_0-4,k_0-4)-FP. Call this forbidden pattern σ_R. If we add back lines to the forbidden pattern we only increase the distortion and will never lower it. Thus the minimum distortion of the original (n,k)-FP $\geq$ the distortion of σ_R, which is in turn $\geq$ the minimum distortion over all (k_0-4,k_0-4)-FP. ■

2.3 The Algorithm

2.3.1 Algorithm Intuition

Our algorithm will guarantee that we solve all inputs whose optimal bijection does not contain a (k_0, k_0)-FP for any $k_0 > k$, where k is the parameter. We know that the minimum distortion of all (n, k_0)-FP is $\geq$ the minimum distortion of all $(k+1, k+1)$-FP, $(k+2, k+2)$-FP, $(k+3, k+3)$-FP, $(k+4, k+4)$-FP. Thus, if the distortion of the optimal bijection is less than the minimum distortion of all $(k+1, k+1)$-FP, $(k+2, k+2)$-FP, $(k+3, k+3)$-FP, $(k+4, k+4)$-FP, we know that it does not contain a (n, k_0)-FP and we can solve it. So setting $k = 5$ (see appendix B) would give us a guarantee that we could solve any pattern with an optimal embedding with distortion less than $\frac{\sqrt{\frac{\sqrt{5}+1}{2}+1}}{\sqrt{\frac{\sqrt{5}+1}{2}-1}} \approx 8.352$. Our main result comes from setting $k = 8$, in which case we get that we can solve any embedding with distortion less than $5+2\sqrt{6}$.

On an intuitive level our algorithm will solve any $(n, k+1)$-NFP as follows. It looks at every possible subinterval of the points in U against every possible subinterval of the points in V starting with size 2 and working up to size n. It will break the subintervals into every possible k subsubintervals (including the empty sets). It will then try match these k subsubintervals by trying all $k!$ possible bijections of the subsubintervals. If a match is found with low enough distortion the match will be saved for future reference. How the subintervals are mapped is no longer important; the only things we need to know about the subinterval to continue the process is whether there was a bijection with distortion less than α, and the image and the preimages respectively of the first and last point of U and the first and last point in V. The reason we need to keep the mappings of the first and last points in U and V is because when we try to combine two subintervals we need to check the expansion and inverse expansion between them. We store this information in a table. When the subinterval is U and V, if we can map U to V by the same process with distortion less than α we output "yes".

Another way to think about the algorithm is that the algorithm is looking for mappings that contain a pattern size k_1 for some $k_1 \leq k$. If it finds such a pattern it now thinks of that entire set as one mapping that could be part of another pattern of size $\leq k$ and looks for such a pattern.

2.3.2 Algorithm

The algorithm gets as input, two line metrics (U, d) and (V, d'). It also gets as parameters, α the maximum expansion and inverse expansion allowed, as well as a parameter k which is related to the bijections that the algorithm tries.

The algorithm proceeds by building a dynamic programming boolean table T which is indexed by the following parameters:

- a subinterval $I = \{u_m, u_{m+1}, \ldots, u_{m+c-1}\}$ of U and a subinterval $J = \{v_{m'}, v_{m'+1}, \ldots, v_{m'+c-1}\}$ of V of the same size $c \geq 1$;

- four elements $v, v' \in J$ and $u, u' \in I$.[1] v is the image of the first point in I. v' is the image of the last point in I. Similarly u is the preimage of the first point in J. And u' is the preimage of the last point in J.

We set the table entry $T[I, J, v, v', u, u']$ to true if there is a bijection $\sigma : I \to J$ such that $\sigma(u_m) = v, \sigma(u_{m+c-1}) = v', \sigma^{-1}(v_{m'}) = u$ and $\sigma^{-1}(v_{m'+c-1}) = u'$, and with expansion and inverse expansion at most α.

The algorithm runs from $c = 1$ to n. The base case $c = 1$ is trivial, with all entries set to true. For $c > 1$, compute every entry $T[I, J, v, v', u, u']$ with $|I| = c$ and $|J| = c$ as follows: consider all partitions of I and J into $2 \leq k_0 \leq k$ subintervals $I = \bigcup_{a=1}^{k_0} I_a$, $J = \bigcup_{b=1}^{k_0} J_b$. Try all possible combinations of pairs of I_a, J_b $(\sigma(I_a) = J_b)$ over all a,b and set $T[I, J, v, v', u, u']$ to true if and only if in at least one of the combinations, the following conditions hold:

[1]Here, v' and u' do not denote images. They are just normal points. The same will hold throughout this subsection and we will specifically mention the images.

- $\forall a, b \;\; T[I_a, J_b, v_b, v'_b, u_a, u'_a]$ is true, where $\sigma(I_a) = J_b$.
- Let $J_{b_1} = \sigma(I_a), J_{b_2} = \sigma(I_{a+1}), I_{a_1} = \sigma^{-1}(J_b), I_{a_2} = \sigma^{-1}(J_{b+1})$. Then,

$$\begin{aligned} d(v_{b_2}, v'_{b_1}) &\leq \alpha \cdot d(\min(I_{a+1}), \max(I_a)) \\ d(u_{a_2}, u'_{a_1}) &\leq \alpha \cdot d(\min(J_{b+1}), \max(J_b)) \end{aligned}$$

These inequalities ensure that the edges connecting the subintervals have expansion and inversion expansion at most α.[2]

Once the table is prepared, the algorithm just checks that $T[U, V, v, v', u, u']$ is true for some (v, v', u, u').

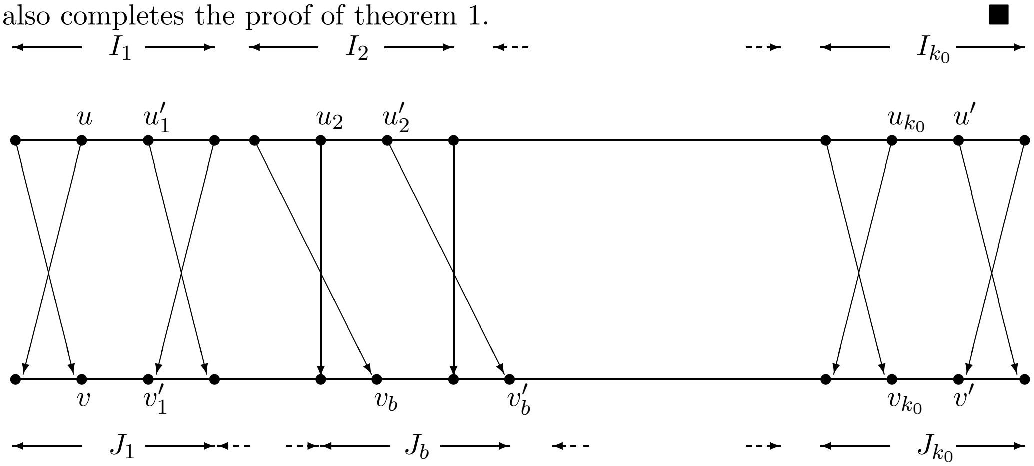
2.3.3 Analysis

Correctness For the correctness of this algorithm we must show that we can solve any bijection whose optimal mapping is a $(n, k+1)$-NFP. By definition the $(n, k+1)$-NFP can be reduced iteratively to a single point, where at each iteration the maximum size of the subsequence will be at most k. Thus the algorithm will try each of these partial mappings and return a value of true for them.

Now that we have shown that our algorithm can solve any $(n, k+1)$-NFP, it will only fail when the optimal bijection contains a (k_0, k_0)-FP for any $k_0 > k$. It follows directly from Theorem 2 that the algorithm is guaranteed to solve any instance with optimal distortion less than the minimum distortion of all $(k+1, k+1)$-FP, $(k+2, k+2)$-FP, $(k+3, k+3)$-FP, $(k+4, k+4)$-FP, since if the distortion is less than the minimum distortion of all $(k+1, k+1)$-FP, $(k+2, k+2)$-FP, $(k+3, k+3)$-FP, $(k+4, k+4)$-FP, it does not contain a (n, k_0)-FP. Note that with $k = 8$, we can solve any embedding with distortion less than $5 + 2\sqrt{6}$.

Running Time The running time of the algorithm is easy to bound. Notice that the table size is just $O(n^7)$. Computing each entry $T[I, J, v, v', u, u']$ of the table is polynomial in n: the sets I and J can be split into $k_0 \leq k$ sets in $O(n^{k-1})$ ways and for each such possible splitting we store $4(k_0 - 2) + 2 + 2 \leq 4(k-1)$ mappings, which can be done in $O(n^{4k-4})$; and finally there are $k!$ possible ways of mapping various I_a to various J_b. Thus computing each entry takes $O(n^{4k-4} \cdot n^{k-1} \cdot k!) = O(n^{5k-5})$ time. So, computing the whole table takes $O(n^{5k-5} \cdot n^7) = O(n^{5k+2})$.

This also completes the proof of theorem 1. ■



Filling in the table: a possible case

2.4 Limitations of the Line Embeddings Algorithm

Let β be such that the distortions of all forbidden patterns of sizes $k+1, k+2, k+3$, and $k+4$ are all at least β. Then, given the guarantee that the optimal bijection between two line metrics is less than β, we

[2] Note that we only need to consider the expansion and inverse expansion of edges [14].

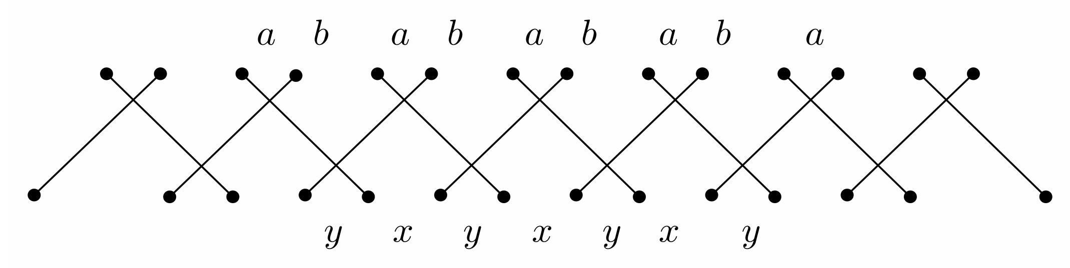


Figure 4: An arbitrarily large pattern with distortion approaching $7 + 4\sqrt{3}$

can find it. One might hope that the value of β can be made an arbitrarily large constant. However, this is not true. Figure 4 shows a forbidden pattern of arbitrarily large size that has distortion approaching $7 + 4\sqrt{3}$. This shows that using our algorithm, we cannot achieve a better guarantee than $7 + 4\sqrt{3}$.

Using the tightness property of edges in this pattern, we get the following equations[3]

$$\begin{array}{l} \alpha a = 2x + 3y \ ; \ \alpha b = x + 2y \\ \alpha x = 2a + 3b \ ; \ \alpha y = a + 2b \end{array} \Rightarrow \begin{array}{c} \alpha(2b - a) = y \ ; \ \alpha(2a - 3b) = x \\ \alpha^2(2a - 3b) = 2a + 3b \ ; \ \alpha^2(2b - a) = a + 2b \end{array}$$

From which we get $\alpha^2 = 7 + 4\sqrt{3} \approx 13.928$

3 Embeddings for Bounded Degree Graphs with Short Cycles

Theorem 3 *Let (U, d) be the shortest-path metric of an unweighted graph G of maximum degree b. Let (V, d') be the shortest-path metric of an arbitrary unweighted graph G'. Then, the problem of finding an optimal bijection between U and V is NP-Hard.*

Proof. See appendix C.1.

In this section, we prove the following in a very similar manner to the algorithm presented in [14].

Theorem 4 *Let (U, d) be the shortest-path metric of an unweighted graph G of maximum degree b and largest cycle length 3. Let (V, d') be the shortest-path metric of an arbitrary unweighted graph G'. Then, for any fixed constants b and α, there is an $O(n^2)$ algorithm that decides whether there exists a bijection between U and V with expansion and inverse expansion at most α.*

3.1 Structural properties

The following lemma is shown (in [14]) in the case where (U, d) is the shortest-path metric of an unweighted tree T of maximum degree b.

Let $B(u, l)$ (resp., $B'(u, l)$) denote the closed ball of radius l around any vertex u in T (resp., in G'). For a subset of vertices $A \subseteq T$ (resp., in G'), let $\Gamma(A)$ (resp., $\Gamma'(A)$) denote the set of neighbors of A that lie outside A. Assume that T is rooted at an arbitrary vertex r_0. The subtree rooted at any vertex r of T is denoted by T_r.

Lemma 2 ([14]) *Let $\sigma : U \to V$ be a bijection with expansion and inverse expansion at most α. Then*

1. *G' has maximum vertex degree at most b^α.*
2. *For any vertex $r \in T$, each connected component of $G \backslash B'(\sigma(r), \alpha^2)$ lies either entirely in $\sigma(T_r)$ or entirely in $G' \backslash \sigma(T_r)$.*

[3] These equations hold only when the shown pattern extends to infinity. In the case of a finite pattern of this type, since this pattern will be contained in the infinite pattern, its distortion is at most the distortion of the infinite pattern.

3. *For any $r \in T$, for any adjacent pair (u', v') in G' with $u' \in \sigma(T_r)$ and $v' \notin \sigma(T_r)$, both $\sigma^{-1}(u')$ and $\sigma^{-1}(v')$ are in $B(r, \alpha - 1)$*

We claim that this lemma is true even in the case where (U, d) is the shortest-path metric of an unweighted graph G of maximum degree b and largest cycle length 3. In other words, T is replaced by G with bounded degree b and maximum cycle length 3 and T_r is replaced by G_r (defined below).

Definition 1 *We say that a graph is* graphrooted *at vertex r_0 by the following definition:*

1. *Vertex r_0 is at level 0 of the graph. In other words $level(0) = r_0$.*
2. *Level i of the graph is defined recursively as $level(i) = \bigcup_{v \in level(i-1)} \Gamma(v) \setminus \bigcup_{j<i} level(j)$*

Since the graph is connected, every vertex is present at some level of the graph. Hence, the definition covers all nodes in the graph. It is also clear that $level(i) \cap level(j) = \emptyset$ for $i \neq j$. This follows from the iterative definition of $level(i)$.

Definition 2 *G_r is the subgraph rooted at vertex r according to the following definition:*

1. *Let l_r be such that $r \in level(l_r)$. If $\exists$ a path from r to v such that for every vertex v' (including v) in this path, $v' \in$ level(k) where $k > l_r$, then v' is a vertex of G_r*
2. *If (v_1, v_2) is an edge in G, and both v_1 and v_2 are $\in G_r$, then the edge (v_1, v_2) is an edge in G_r*

Now we continue with the proof.

Proof.

We note that the following statement is true. If $u \in G_r$ and $v \notin G_r$, then the shortest path from u to v goes through r. Suppose the shortest path from u to v does not go through r. In this case, this path has to go through a node (r') that is at least as low as the level in which r is present (otherwise, v is a vertex of G_r). Note that there is a path from r to r' that goes through nodes only at a level less than that of r. Hence, there is a path from r to r' of length at least 2 that does not overlap with the paths from u to r and u to r'. Now, consider the non-overlapping part of the paths from u to r and u to r'. The lengths of these parts are at least 1 each and hence we get a cycle of length at least 4 (by joining the path from r to r' completely at lower levels and the path from r to r' completely at higher levels). This is a contradiction to the maximum cycle length restriction of 3 on G. Hence, the shortest path from u to v goes through r.

Using this, the proof of the above lemma follows as in [14]. For completeness sake, we have provided the proof in appendix C.2 ■

For the algorithm, its analysis, and proof of theorem 4, see appendix C.3.

References

[1] Working group on algrorithms and multidimensional scaling. http://dimacs.rutgers.edu /Workshops/Algorithms.

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

[3] M. Badoiu, J. Chuzhoy, P. Indyk, and A. Sidiropoulos. Low-distortion embeddings of general metrics into the line. In *STOC'05.* ACM, 2005.

[4] M. Badoiu, K. Dhamdhere, A. Gupta, Y. Rabinovich, H. Racke, R. Ravi, and A. Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. In *SODA'05.* ACM, 2005.

[5] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[6] M. Cary, A. Rudra, and A. Sabharwal. On the hardness of embeddings between two finite metrics. In *ICALP'05*, volume 3580 of *LNCS*, pages 1412–1423. Springer, 2005.

[7] B. Chazelle, D. Liu, and A. Magen. Sublinear geometric algorithms. In *STOC'03*. ACM, 2003.

[8] P.Z. Chinn, J. Chvatalova, A.K. Dewdney, and N.E. Gibbs. The bandwidth problem for graphs and matrices – a survey. *Journal of Graph Theory*, 6:223–254, 1982.

[9] J. Diaz, J. Petit, and M. Serna. A survey on graph layout problems. *ACM Computing Surveys*, 34(3):313–356, 2002.

[10] S. Fortin. The graph isomorphism problem. Technical Report 96-20, University of Alberta, Edomonton, Alberta, Canada, 1996.

[11] A. Hall and C. Papadimitriou. Approximating the distoriton. In *APPROX and RANDOM*, volume 3624 of *LNCS*, pages 111–122. Springer, 2005.

[12] F. Hoffmann, K. Kriegel, and C. Wenk. Matching 2d patterns of protein spots. In *ACM Symposium on Computational Geometry*, pages 231–239. ACM, 1998.

[13] W.B. Johnson and J. Lindenstrauss. *Handbook of Geometry of Banach Spaces*. North-Holland, 2003.

[14] C. Kenyon, Y. Rabani, and A. Sinclair. Low Distortion Maps Between Point Sets. In *STOC'04*. ACM, 2004.

[15] J.B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Pyschometrika*, 29:1–27, 1964.

[16] J.B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Pyschometrika*, 29:115–129, 1964.

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

[18] C. Papadimitriou and S. Safra. The complexity of low-distortion embeddings between point sets. In *SODA'05*. ACM, 2005.

[19] R.N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function 1. *Pyschometrika*, 27:125–140, 1962.

[20] R.N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function 2. *Pyschometrika*, 27:216–246, 1962.

Appendix

A Proofs of Section 2

A.1 Proof of base case for Lemma 1

Figure 5 is the list of $(5,5)$-FP. By removing the dotted line in each of the forbidden patterns, we obtain a $(4,4)$-FP that contains both end points in V. This proves the base case of Lemma 1.

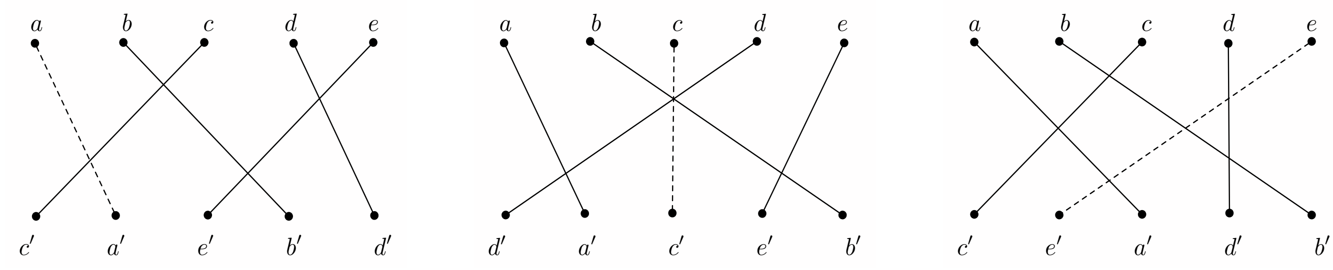


Figure 5: Base case: $(5,5)$-FP

A.2 Case $|\sigma_{K'}| = 2$ continued

We will now show the existence of the (k_0, k_0)-FP by separating two cases of c',d' containing four cases of e',f'.

The argument in all cases will go as follows. We will remove either $c - c'$ or $d - d'$ from the mapping and obtain a new mapping σ_{K_N} . Assume that σ_{K_N} is a $(|\sigma_{K_N}|, |\sigma_{K_N}|)$-NFP. Then there exists a consecutive mapping $\sigma_{K''}$ with $|\sigma_{K''}| < |\sigma_{K_N}|$. We will show that such a $\sigma_{K''}$ must encompass $\sigma_{K'}$, $a - a'$ and all the removed lines or that the removed lines will border $\sigma_{K''}$. At this point we will be done, due to the following observation.

Adding back all of $\sigma_{K'}$ and the removed lines to $\sigma_{K''}$ would give us a consecutive mapping in σ_K that will increase the size by at most the number of removed lines since all lines border or encompass $\sigma_{K''}$. If this map is of size less than k then this contradicts σ_K being a (k, k)-FP. If this map is of size k then $\sigma_{K''}$ is all of σ_{K_N} and hence σ_{K_N} is a forbidden pattern. $|\sigma_{K_N}|$ is equal to $|\sigma_K|$ minus the total number of lines removed. The total number of lines we removed is at most 4. Thus we get that $|\sigma_{K_N}| \geq k - 4$.

To obtain σ_{K_N} we will first remove either $c - c'$ or $d - d'$ (depending on the case). Call the remaining line $r - r'$.

Let b be the right neighbor of a. If $a - a'$ and $b - b'$ form a consecutive mapping rename $a - a'$ as $t - t'$ and rename $b - b'$ as $a - a'$ and remove $t - t'$. Now $a - a'$ will not be a consecutive mapping with its neighbor. Note that V still contains both its endpoints since a' is not an endpoint in V.

The remaining mapping is now σ_{K_N} . Suppose there is a consecutive mapping in σ_{K_N} call it $\sigma_{K''}$. We will show that if $\sigma_{K''}$ does exist then it must contain $r - r'$. Suppose $\sigma_{K''}$ lies in $\sigma_{K-K'-a}$ then it must encompass $\sigma_{K'}$, thus it must include $r - r'$. Suppose that $\sigma_{K''}$ includes $a - a'$. Since a can not merge with it right neighbor it must include two points to the right of a. If one of these points is r then $\sigma_{K''}$ includes r. If neither of these points is r then we have two points in $\sigma_{K-K'-a}$ that are in $\sigma_{K''}$ and by the above argument we must include $r - r'$.

Thus in all cases any mapping $\sigma_{K''}$ must include $r - r'$ (see figure 6 for all subcases).

Case 1: c' is to the left of a'.

Subcase i: e' and f' lie to the right of a'. Let $r - r'$ equal $c - c'$ and remove and rename as discussed above; the remaining mapping is σ_{K_N} . Thus $\sigma_{K''}$ must include $c - c'$. Now we need to show that $\sigma_{K''}$ encompasses $\sigma_{K'}$, $a - a'$ and all the removed lines or that the removed lines border $\sigma_{K''}$. Since $c - c'$ is in $\sigma_{K''}$, one of $e - e'$ or $f - f'$ must be in $\sigma_{K''}$ (since e and f are the neighbors of c).

If $e - e'$ is in $\sigma_{K''}$ then $a - a'$ must be in $\sigma_{K''}$. Since e' is to the right of d', $\sigma_{K''}$ encompasses $d - d'$ and therefore $\sigma_{K'}$. We can include $t - t'$ in this set without adding any more lines because $t - t'$ borders $\sigma_{K''}$. Thus we are done.

If $f - f'$ is in $\sigma_{K''}$, then $a - a'$ must be in $\sigma_{K''}$, thus $\sigma_{K'}$ and any removed lines are encompassed by (or border) $\sigma_{K''}$ by a similar argument as above.

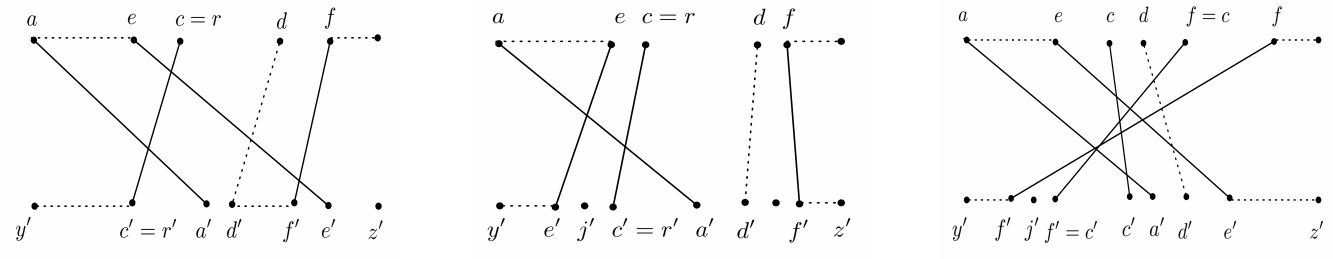


Figure 6: Subcases i, iii, and iv

Subcase ii: e' and f' lie to the left of a'. Let $r - r'$ equal $d - d'$ and remove and rename as discussed above; the remaining mapping is σ_{K_N} . The remaining mapping is σ_{K_N} and the proof is analogous to subcase i.

Subcase iii: e' lies to the left of a' and f' lies to the right of a'. Let $r - r'$ equal $c - c'$ and remove and rename as discussed above; the remaining mapping is σ_{K_N} . Thus $\sigma_{K''}$ must include $c - c'$. $\sigma_{K''}$ must also include one of $e - e'$ or $f - f'$ as discussed in subcase i.

If $f - f'$ is in $\sigma_{K''}$, then $a - a'$ must be in $\sigma_{K''}$. Thus $\sigma_{K'}$ and any removed lines are encompassed by (or border) $\sigma_{K''}$ by a similar argument to subcase i.

If $e - e'$ is in $\sigma_{K''}$, recall that $e - e'$ and $c - c'$ do not form a consecutive mapping. Since e' and c' are not neighbors there exists a point j' in between that must be part of any mapping including $e - e'$ and $c - c'$. j' is not in $\sigma_{K'}$. Thus $\sigma_{K''}$ must include at least two points in $\sigma_{K-K'-a}$. But every consecutive mapping in $\sigma_{K-K'-a}$ encompasses $\sigma_{K'}$. Thus $\sigma_{K''}$ has a point to the right of f in U or to the right of a' in V. In either case $\sigma_{K''}$ must include $a - a'$, $t - t'$ and any other line removed is in $\sigma_{K''}$ or borders $\sigma_{K''}$. Thus we have shown that $\sigma_{K''}$ encompasses or is bordered by $\sigma_{K'}$, $a - a'$ and any lines removed.

Subcase iv: e' lies to the right of a' and f' lies to the left of a'. Let $r - r'$ equal $c - c'$ and remove and rename as discussed above; the remaining mapping is σ_{K_N} . Thus $\sigma_{K''}$ must include $c - c'$. $f - f'$ and $c - c'$ may be a consecutive mapping, if so rename $c - c'$ as $m - m'$ and remove $m - m'$ and rename $f - f'$ as $c - c'$. Notice that m' was not an endpoint in V because it had two neighbors thus we still have both endpoints in V in the current mapping. Note that the new $c - c'$ will be included in any $\sigma_{K''}$. Let the right neighbor of the current c be f (if it exists). If f does not exist, then we are exactly in a case discussed earlier (where the right neighbor of d, f does not exist and the proof technique used there holds). Note that $c - c'$ and $f - f'$ do not form a consecutive mapping, since they either were not a consecutive mapping and were never renamed or they were renamed and are both in $\sigma_{K-K'-a}$.

Notice that if f' and e' are both to the right of a', we are in the same position as subcase i. Thus we must only consider the case where f' is to the left of a'. If f' is to the left of a' there must be a point in between f' and c'; call it j'.

Let the current mapping be σ_{K_N} . $\sigma_{K''}$ will include $c - c'$. Since $c - c'$ is in $\sigma_{K''}$ one of $e - e'$ or $f - f'$ must be in $\sigma_{K''}$ (since e and f are the neighbors of c).

If $e - e'$ is in $\sigma_{K''}$, then $a - a'$ must be in $\sigma_{K''}$. Thus $\sigma_{K'}$ and any removed lines are encompassed by (or border) $\sigma_{K''}$ by a similar argument to subcase i.

If $f - f'$ is in $\sigma_{K''}$, then j' is also in $\sigma_{K''}$ (since j' is between f' and c'). We now have two points (f' and j') that are in $\sigma_{K-K'-a}$ that must be in $\sigma_{K''}$. Thus any consecutive mapping that includes these two points must encompass $\sigma_{K'}$. If $\sigma_{K''}$ encompasses $\sigma_{K'}$ then it must contain a point either to the left of e in U or to the right of a' in V. In either case we must include $a - a'$. Hence $\sigma_{K''}$ encompasses (or borders) $\sigma_{K'}$, $a - a'$ and any removed lines.

Case 2: c' is to the right of a'. Each subcase can be shown using the similar arguments as before. ■

B Minimum Distortion of certain (k,k)-FP

We can now build on Theorem 2 and give the minimum distortions over all (k,k)-FP for some low values of k.

Following [14] we calculated the distortion where every edge was *tight*. A tight edge is an edge that achieves expansion or inverse expansion. When calculating the minimum distortion of a pattern we may assume that the edges on both lines are tight. This is due to the simple observation that if some edge is not tight, we could decrease the distance between the endpoints of the edge until it became tight and this would never increase the distortion.

Thus we can assume that every edge gives expansion or inverse expansion α and solve the system of equations for the distortion α^2.

Solving the system of equation for every (k,k)-FP, we get the following results for $k \leq 12$.

k	4	5	6	7	8	9	10	11	12
Minimum Distortion (upto 3 decimal pts.)	5.828	5.828	8.352	8.352	8.352	9.899	9.899	9.899	10.896

C Proofs of Section 3

C.1 Proof of Theorem 3

PROOF. This proof is based on the proof that it is NP-hard to approximate the minimum distortion problem within a factor better than 2 given in [14]. Let G' be an unweighted, undirected graph on n vertices. Construct a metric (V,d') by setting $d'(u,v) = 1$ if u,v is an edge of G', and $d'(u,v) = 2$ otherwise. Let the bounded degree graph G be the unweighted cycle on n vertices, C. Clearly C is of bounded degree $b = 2$ and construct the metric (U,d) in the same manner as (V,d'). It is easy to check that, if G' contains a Hamilton cycle, then an optimal bijection between (U,d) and (V,d') has distortion exactly 2. If G' does not contain a Hamilton cycle, then any bijection must have distortion at least 4. Hence the problem of finding an optimal bijection between (U,d) and (V,d') as described above is NP-Hard. Since the given instance is a particular case of the metrics in the lemma, the lemma is true.

C.2 Proof of Lemma 2

PROOF.

For the first statement, for any $v \in G'$, the expansion of σ^{-1} implies that $\sigma^{-1}(B'(v,1)) \subseteq B(\sigma^{-1}(v),\alpha)$, and the cardinality of this ball is at most b^α by the degree bound on G.

For the second statement, let $v' = \sigma(v)$ be a vertex of $\Gamma'(\sigma(G_r))$ graphrooted at r. By the definition of Γ', v' is adjacent to some vertex $u' = \sigma(u)$ of $\sigma(G_r)$. From the inverse expansion bound, we have $d(u,v) \leq \alpha$. We shall show that $d(r,v)$ is also $\leq \alpha$. Since the shortest path from u to v goes through r, clearly $d(r,v) \leq \alpha$. Thus, we have $d'(\sigma(r),v') \leq \alpha^2$. From this we get

$$\Gamma'(\sigma(G_r)) \subseteq B'(\sigma(r),\alpha^2)$$

from which we get the second statement.

For the third statement, note that by the expansion of σ^{-1}, we get that $d(\sigma^{-1}(u'),\sigma^{-1}(v')) \leq \alpha$. Now, the shortest path from u to v goes through r, which in turn implies that $d(r,\sigma^{-1}(u')) \leq \alpha - 1$ and $d(r,\sigma^{-1}(u')) \leq \alpha - 1$. ■

C.3 Algorithm and Proof of Theorem 4

The algorithm is a dynamic programming algorithm in the same way as given in [14]. The graph G is graphrooted arbitrarily at a node r_0. The dynamic programming table T is indexed by the following parameters

1. $r \in \{u_1, ..., u_n\}$, the root of the subgraph G_r (with respect to the graphrooting G)
2. $r' \in \{v_1, ..., v_n\}$
3. An injection τ from $B(r, \alpha) \cap G_r$ into $B'(r', \alpha^2)$
4. A subset S of the vertices of G' with the property that each connected component of $G' \backslash B'(r', \alpha^2)$ lies entirely within S or entirely outside S.

An entry of the table is true if and only if there exists an injection $\sigma : G_r \to G'$ such that $\sigma(r) = r'$, σ coincides with r on $B(r, \alpha) \cap G_r$, $\sigma(G_r) = S$, and expansion of every edge of G_r and inverse expansion of every edge of $\sigma(G_r)$ are each at most α.

To compute $T(r, r', \tau, S)$, we run through all combination of entries $T(r_i, r'_i, \tau_i, S_i)_i$ all of which have value true. r_i are the children of a given root r. We set the result to be true if at least one of these combinations satisfies the conditions below and false otherwise.

1. The map τ is consistent with all the maps τ_i, the τ_is are consistent among themselves, the S_i do not include r', and S is the union of the S_i plus the vertex r'.
2. For each r'_i, we have $d'(r', r'_i) \leq \alpha d(r, r_i)$.
3. For each adjacent pair v', w' in G', that belong to different sets S_i (or with $v' = r'$), both v' and w' are in the image of τ and satisfy $d(\tau^{-1}(v'), \tau^{-1}(v')) \leq \alpha$.

After all entries of the dynamic programming table are computed, the algorithm checks if some table entry $T(r_0, ., ., .)$ is true.

Running time and Correctness.

The degree bound on G' implies that $B'(v, \alpha^2)$ has size at most b^{α^3} for any v. We claim that the size of the table T is at most

$$n \times n \times \left(b^{\alpha^3}\right)^{b^\alpha} \times 2^{2b^{\alpha^3}} = O(n^2)$$

The two n terms come from the r and r' in the table. The third factor bounds the number of maps from $B(r, \alpha)$ to $B'(r', \alpha^2)$. From the second part of the lemma, we get the number of possibilities for the set S as the fourth factor. Filling the table entries takes constant time and thus the overall running time is $O(n^2)$.

The correctness of the algorithm follows in the same way as in [14] by an induction (bottom-up the levels in G). This also completes the proof of theorem 4. ■

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/