Distortion lower bounds for line embeddings

Claire Mathieu, Charalampos Papamanthou*

Department of Computer Science, Brown University, P.O. Box 1910, Providence, RI, USA

**Abstract**

In this paper, we show how we can derive lower bounds and also compute the exact distortion for the line embeddings of some special metrics, especially trees and graphs with certain structure. Using linear programming to formulate a simpler version of the problem gives an interesting intuition and direction concerning the computation of general lower bounds for distortion into the line. We also show that our lower bounds on special cases of metrics are a lot better than previous lower bounds.

**1. Introduction**

Embedding a metric induced by an unweighted graph into the line with minimum distortion is a well-known NP-hard problem [4]. Approximation algorithms and hardness results are discussed in [3,4,8]. Improved algorithms for optimal embeddings are presented in [6], where, however, the embedding is required to be a bijection. In [1], embeddings of series-parallel graphs into $\ell_1$ and embeddings between two line metrics are discussed. In this paper, we show how we can derive lower bounds for the line distortion of some special metrics, especially trees and graphs with certain structure. To do this, we use linear programming to formulate another simpler version of the problem, where we know the permutation of the vertices of the graph on the line in advance. This approach gives an interesting intuition and direction concerning the computation of general lower bounds for distortion into the line.

**Problem.** Let $G = (V, E)$ be an unweighted graph with the implied shortest path metric $d(u, v)$. A map $f$ of $V$ into the real line is non-contracting if $|f(u) - f(v)| \geq d(u, v)$ for all $u, v$. The minimum distortion problem is to find a map $f$ with minimum distortion, where the distortion of a non-contracting map $f$ is defined as $\max_{u, v \in V} |f(u) - f(v)|/d(u, v)$.

**Previous work.** There are some results on lower bounds concerning embeddings into higher dimensions. For example, you can embed the complete binary tree in the Euclidean space with distortion $\Omega(\sqrt{\log \log n})$ [5], where $n$ is the size of the tree. A short proof for that is also presented in [11]. Also upper bounds for embedding the binary tree into Euclidean space are presented in [5]. Additionally, in [2], lower bounds of embeddings of trees into $\mathbb{R}^2$ are discussed. Lower bounds of low distortion embeddings to higher dimensional spaces are also presented in [5].

However, there are few results for the computation of lower bounds concerning embeddings into the line. Actually, in [4,9], it is proved that the least distortion required to embed the star into the line is $\Omega(n)$. Also, in [3] a lower bound method that applies to all unweighted graphs is presented.

In [3], an $O(\sqrt{n})$ approximation algorithm for computing embeddings of unweighted graphs ($n$ is the size of the graph) into the line is described. This algorithm partitions the nodes of the graph into distinct clusters and explicitly computes the intervals between the nodes on the line.

Define the local density of an unweighted graph $G = (V, E)$ to be the quantity $\Delta$ such that

$$\Delta = \max_{v \in V, r \in \mathbb{N}, 0} \left\{ \frac{|B(v, r)| - 1}{2r} \right\}$$
Theorem 2. Given an unweighted graph \( G = (V, E) \) then the minimum distortion embedding of \( G \) into the line has distortion equal to

\[
c^* = \min_{\pi:V\rightarrow\{1,\ldots,n\}} \left\{ \max_{(u,v)\in E} \left\{ \frac{\pi(v)-1}{\sum_{k=\pi(u)}^{\pi(v)-1} r_k^\pi} \right\} \right\}
\]

where \( r_k^\pi = d(\pi^{-1}(i), \pi^{-1}(i+1)) \).

We start the proof with two simple observations.

Lemma 3. The distortion of a map \( f \) that maps the vertices of an unweighted graph \( G = (V, E) \) into the line equals \( \max_{(u,v)\in E} |f(u) - f(v)| \).

Proof. For \( u, v \in V \), let \( u = y_1, y_2, \ldots, y_k = v \) be a shortest path from \( u \) to \( v \). Then

\[
|f(u) - f(v)| \leq \sum_{j=1}^{k-1} |f(y_j) - f(y_{j+1})| \\
\leq c(k - 1) = cd(u, v)
\]

where \( c = \max_{(u,v)\in E} |f(u) - f(v)| \). Since for every edge \( (u,v) \in E \) it is \( d(u,v) = 1 \), the result follows. \( \square \)

Lemma 4. The map \( f \) is non-contracting if and only if \( f(y_{i+1}) - f(y_i) \geq d(y_{i+1}, y_i) \) for all \( i \), where \( f(y_i) \) is a labelling of the vertices in order of their embedding on the line by \( f \), i.e., \( f(y_1) < f(y_2) < \cdots < f(y_n) \).

Proof. For \( u, v \in V \), let \( i, j \) be such that \( u = y_j \) and \( v = y_i \), and without loss of generality \( i < j \). Then

\[
f(u) - f(v) = \sum_{k=1}^{j-1} (f(y_{k+1}) - f(y_k)) \\
\geq \sum_{k=1}^{j-1} d(y_{k+1}, y_k) \geq d(y_j, y_i) = d(u, v).
\]

The inverse holds trivially by the definition of a non-contracting embedding and the hypothesis. \( \square \)

Combining the two lemmas, we can write a linear program to compute the minimum distortion \( c \) of a non-contracting map, given the permutation \( \pi \) such that \( \pi(u) \) is the rank of \( u \) in the embedding. Here \( x_i \) is the distance on the real line between the \( i \)th point and the \((i + 1)\)th point of the embedding.

\[
\text{min}(c) \quad \text{s.t.} \quad \left\{ \begin{array}{l}
\sum_{k=\pi(u)}^{\pi(v)-1} r_k^\pi \leq c & \forall (u, v) \in E, \\
x_i \geq d(\pi^{-1}(i), \pi^{-1}(i+1)) & \forall i, \\
x_i, c \geq 0.
\end{array} \right.
\]

Recall that \( r_k^\pi = d(\pi^{-1}(i), \pi^{-1}(i+1)) \). It is easy to see that the minimum is reached when \( x_i = r_i^\pi \), and given that, that the minimum is obtained for \( c = \max_{(u,v)\in E} \left\{ \sum_{k=\pi(u)}^{\pi(v)-1} r_k^\pi \right\} \).

Minimizing over all permutations \( \pi \) yields the proof.

\[ \square \]

Theorem 5. Let \( G \) be an unweighted graph. Let \( v_1, v_2, v_3 \) be three vertices, and \( p, p', p'' \) be three paths from \( v_1 \) to \( v_3 \), from \( v_3 \) to \( v_2 \) and from \( v_2 \) to \( v_1 \) respectively. Then

\[
c^* \geq 2 \min \left\{ \min_{u \in p} \left| d(v_2, u) \right|, \min_{u \in p'} \left| d(v_1, u) \right|, \min_{u \in p''} \left| d(v_3, u) \right| \right\}
\]

where \( c^* \) is given in Theorem 2.

Proof. Let \( t \) denote the minimum of the three quantities. Let \( \pi \) be an arbitrary permutation and \( f \) be the optimal embedding respecting \( \pi \). First assume that in \( \pi \), \( v_2 \) is between \( v_1 \) and \( v_3 \), and so, \( f(v_2) \) is between \( f(v_1) \) and \( f(v_3) \). Since \( p \) is a path from \( v_1 \) to \( v_3 \) in \( G \), there must exist an edge \( (x, x') \) of \( p \) such that \( f(v_2) \) is between \( f(x) \) and \( f(x') \). Then, since \( f \) is non-contracting:

\[
d(x', v_2) + d(v_2, x) \leq (f(x') - f(v_2)) + (f(v_2) - f(x)) = f(x') - f(x) = \sum_{k=\pi(x)}^{\pi(x')-1} r_k^\pi.
\]

Here the last equality comes from the fact that for \( f \) optimal, the second set of inequalities of the linear program are tight. Therefore for edge \( (x, x') \)

\[
\sum_{k=\pi(x)}^{\pi(x')-1} r_k^\pi \geq d(x', v_2) + d(v_2, x)
\]

which can be more simplified into (decreasing the lower bound a little bit)

\[
\sum_{k=\pi(x)}^{\pi(x')-1} r_k^\pi \geq 2t.
\]

The other cases are similar. Thus, for every permutation \( \pi \), there exists an edge \( (x, x') \) such that \( \sum_{k=\pi(x)}^{\pi(x')-1} r_k^\pi \geq 2t \). By Theorem 2 this implies \( c^* \geq 2t \). \( \square \)
3. Examples

3.1. The two-path tree

**Theorem 6.** A two-path tree $T(n, j)$ is an unweighted graph which is the tree consisting of two paths $V = v_1, v_2, \ldots, v_n$ and $U = u_1, u_2, \ldots, u_n$ and an edge $[v_j, u_j]$ such that $j \leq (n + 1)/2$. Let $c^*(T(n, j))$ denote the optimal distortion into the line of a 2-path tree $T(n, j)$. Then:

$$2j - 2 \leq c^*(T(n, j)) \leq 2j - 1.$$  

**Proof.** To prove the lower bound, use Theorem 5 for the nodes $v_1, u_1, v_n$ to yield $c^* \geq 2 \min\{j - 1, j, n - j\} = 2(j - 1)$. Also we can get an even better lower bound equal to $2j - 1$, by using directly Eq. (1). To prove the upper bound, embed the nodes in the following order (see Fig. 1):

$$P(n, j) = v_n(1)v_{n-1}(1)\ldots(1)v_1(j)u_1(1)\ldots u_{j-1}(1)(1)u_1(1)u_{j+1}(1)u_{j+2}(1)\ldots(1)u_n$$

where $x(d)y$ means that node $y$ is embedded after node $x$ at distance $d$. Note that the embedding is non-contracting and also

$$\frac{d(f(u_j), f(v_j))}{d(u_j, v_j)} = \frac{2j - 1}{1} = 2j - 1.$$

Hence $c^*(T(n, j)) \leq 2j - 1$ and the presented embedding is optimal. □

Note that Theorem 1 would give a lower bound of $\Omega(1)$ only, whereas our lower bound is essentially tight.

3.2. The rectangular grid

**Theorem 7.** Let $S(n, j)$, $n \geq j$ denote the unweighted graph with $nj$ nodes consisting of the $n \times j$ rectangular grid. Let $c^*(S(n, j))$ denote the optimal distortion. Then:

$$2j - 2 \leq c^*(S(n, j)) \leq 2j - 1.$$  

**Proof.** Each node of the grid is denoted $s_{kl}$, $i = 1, \ldots, n$, $l = 1, \ldots, j$. For the lower bound, use Theorem 5 with nodes $s_{11}, s_{n1}, s_{nj}$. The paths we are using are $s_{11}, s_{21}, \ldots, s_{n1}$, $s_{n1}, s_{n2}, \ldots, s_{nj}$, $s_{nj}, s_{n(n-1)}, \ldots, s_{1j}, s_{1(j-1)}, \ldots, s_{11}$.

By using again Eq. (1) we can easily get the best lower bound equal to $2j - 1$. For the upper bound, consider the embedding (see Fig. 1) represented with the following permutation (without loss of generality we assume $n$ is odd):

$$P = s_{11}, s_{12}, \ldots, s_{1j}, s_{2j}, s_{2(j-1)}, \ldots, s_{21}, \ldots, s_{nj}, s_{n(n-1)}, \ldots, s_{1j}, s_{1(j-1)}, \ldots, s_{n1}.$$

For this permutation $\pi$, by the linear programming formulation we have that

$$c^*(S(n, j)) = \max_{[u, v] \in E} \left\{ \sum_{k = \pi(u)}^{\pi(v) - 1} r_k^\pi \right\} \leq \sum_{k = \pi(s_{11})}^{\pi(s_{1j}) - 1} r_k^\pi = \sum_{k = \pi(s_{11})}^{\pi(s_{1j}) - 1} r_k^\pi = 2j - 1.$$

Hence the optimal distortion is $2j - 1$. □

3.3. Star metrics

Consider a star metric that consists of a star with $O(1)$ branches of length $O(n)$ with the following properties: Then, our lower bound is $\Omega(n)$, hence tight, whereas the lower bound by [3] is $\Omega(1)$ (there is no ball of constant radius that contains $O(n)$ nodes).

3.4. The complete binary tree

Next, we give some thoughts that may help in order to prove some bounds for the optimal distortion of the complete binary tree.

**Corollary 8.** The complete binary tree of $n$ nodes and height $\ell$ cannot be embedded in the line with a distortion less than

$$\frac{n - 1}{2^{\ell/\ell - 1} = \Omega(\frac{n}{\log n}) = \Omega(\frac{2^\ell}{\ell})}.$$

**Proof.** Follows from Theorem 1 if we choose the root as the center of the ball. □

![Diagram](https://via.placeholder.com/150)
Also, we can easily prove the following:

**Theorem 9.** Let $T$ be a complete binary tree of height $\ell$. Then, an inorder traversal of the tree gives a non-contracting embedding of distortion $2^{\ell} - 1$.

Finally we believe that the following conjecture holds:

**Conjecture 10.** The complete binary tree can be embedded into the line with distortion $\Theta(2^\ell)$.

Optimal embeddings of a complete binary tree under a different setting though (where non-contractness is not an issue and also expansion is allowed) are presented in [10].

4. **Concluding remarks**

In this paper we propose a technique to compute good lower bounds for the optimal distortion into the line of certain metrics. We give better lower bounds than the ones derived when using an already proposed general lower bound. To do this we use a linear programming formulation that gives the optimal solution when the optimal permutation of the embedding is known. We give explicit constructions for various metrics, such as trees, grids and star metrics. Also, we propose an interesting conjecture. Finally we note that it might be possible to resolve the proposed conjecture by taking some ideas from [10] in order to compute an optimal upper bound.

### Acknowledgements

The authors would like to thank Aris Anagnostopoulos and Eli Upfal for their useful feedback at the early stages of this work.

### References