An improved approximation ratio for the minimum linear arrangement problem

Uriel Feige * James R. Lee†

Abstract

We observe that combining the techniques of Arora, Rao, and Vazirani, with the rounding algorithm of Rao and Richa yields an $O(\sqrt{\log n}\log\log n)$-approximation for the minimum-linear arrangement. This improves over the previous $O(\log n)$-approximation due to Rao and Richa.

1 Introduction

Given a graph $G = (V, E)$ and positive edge weights $w : E \to \mathbb{R}_+$, a *linear arrangement* is a permutation $\pi : V \to \{1, 2, \ldots, n\}$. The cost of the arrangement is $\sum_{uv \in E} w(u, v) \cdot |\pi(u) - \pi(v)|$. In the *minimum linear arrangement* (MLA) problem, one seeks a linear arrangement of minimum cost. This problem is known to be NP-complete.

Rao and Richa [8] present an algorithm for MLA with an $O(\log n)$ approximation ratio, and another algorithm which achieves a ratio of $O(\log\log n)$ when G is a planar graph. For an account of earlier work on MLA, see [8]. Arora, Rao, and Vazirani [2] introduced new techniques for the rounding of semi-definite programs based on the analysis of finite metric spaces of negative type. In this note, we observe that the techniques of [8] and [2] can be combined to obtain an approximation ratio of $O(\sqrt{\log n}\log\log n)$ for MLA. A similar upper bound was obtained independently by Charikar, Hajiaghayi, Karloff, and Rao [3].

2 The algorithm

The authors of [5] introduce the following "spreading metric" relaxation for MLA. The variables are $d(u, v)$ for $u, v \in V$. We minimize

$$\sum_{uv \in E} w(u, v) \cdot d(u, v)$$

subject to the constraints

1. For every pair $u, v \in V$, $d(u, v) \geq 1$.

   Additionally, for every subset $S \subseteq V$ with $|S| \geq 2$, and every $u \in S$,

$$\sum_{v \in S} d(u, v) \geq \frac{|S|^2}{5}$$

*Microsoft Research, Redmond, Washington, and Department of Computer Science and Applied Mathematics, Weizmann Institute, Rehovot, Israel. urifeige@microsoft.com, uriel.feige@weizmann.ac.il.

†Department of Mathematics, Institute for Advanced Study, Princeton, NJ. jrl@math.ias.edu.

This is a valid constraint because if the vertices of S lie on a path, and $d(\cdot, \cdot)$ is the path distance, then the worst configuration for the above inequality occurs when $|S|$ is odd, the $|S|$ vertices occupy consecutive nodes of the path, and u is the middle node. In this case, the above sum is $2(1 + 2 + \cdots + \frac{|S|-1}{2}) \geq |S|^2/5$.

2. (V, d) is a metric space, i.e. for every triple $u, v, w \in V$,

$$d(u, v) \leq d(u, w) + d(w, v).$$

Observe that the program is optimizing a linear function of the $d(u, v)$ variables subject to linear constraints. The program contains an exponential number of constraints, but it is not difficult to find a separation oracle or to see that the LP is indeed a relaxation (see [5]). We will say that any metric space (V, d) satisfying the first set of constraints (1) is a *spreading metric.* If we require that $d(u, v) = ||x_u - x_v||_2^2$ with $x_u \in \mathbb{R}^n$ for every $u \in V$, then the program can be written naturally as an SDP (see, e.g. [2]), and the metric space (V, d) is said to be of *negative type.* (The program remains a relaxation: Given an optimal arrangement $\pi : V \to \{1, 2, \ldots, n\}$, one sets $x_u = (1, \ldots, 1, 0, \ldots, 0) \in \{0, 1\}^n$ where the number of initial 1's is exactly $\pi(u)$.)

We will say that a metric space (V, d) is *ε-separable* if, for every subset $S \subseteq V$, with $|S| = k \geq 2$, there exist two non-empty subsets $A, B \subseteq S$ with $|A|, |B| = \Omega(k)$, and $d(A, B) \geq \varepsilon k$, where $d(A, B) = \min_{a \in A, b \in B} d(a, b)$. Rao and Richa essentially prove the following theorem whose proof we sketch in the following section.

Theorem 2.1 ([8]). *Let $G = (V, E)$ be an instance of MLA with edge costs $w(u, v)$ and $|V| = n$. Let d be a metric on V which is ε-separable for some $\varepsilon \geq 1/O(\log n)$, and which satisfies $d(u, v) \geq 1$ for every $u, v \in V$. Then there exists an efficient algorithm which outputs a linear arrangement $\pi : V \to \{1, 2, \ldots, n\}$ such that*

$$\sum_{uv \in E} w(u, v) \cdot |\pi(u) - \pi(v)| \leq O(\log \log n/\varepsilon) \cdot \sum_{uv \in E} w(u, v) \cdot d(u, v).$$

In [8], the authors also observe that a theorem of Klein, Plotkin, and Rao [6] shows that if the shortest-path metric on a planar graph is a spreading metric, then it is $\Omega(1)$-separable. They conclude that there is an $O(\log \log n)$-approximation for MLA in planar graphs.

Now suppose that $G = (V, E)$ is an arbitrary graph, and we instead use the SDP solution so that (V, d) is a metric of negative type. The next theorem follows from the techniques of [2].

Theorem 2.2. *Every n-point spreading metric (V, d) which is also of negative type is $1/O(\sqrt{\log n})$-separable, and there exists an efficient algorithm for computing the separated sets.*

Proof. For a node $u \in V$, we denote $B(u, r) = \{v \in V : d(u, v) \leq r\}$. First, we claim that for any $u \in V$, and any $r \geq \frac{1}{5}$, we have $|B(u, r)| \leq 5r$. To see this, let $T = B(u, r)$. If $|T| = 1$, we are done. Otherwise note that $\sum_{v \in T} d(u, v) \leq |T| \cdot r$ on the one hand, and yet this sum must be at least $|T|^2/5$ by the spreading constraints (1). It follows that $|T| \leq 5r$.

Now let $S \subseteq V$ be any subset with $|S| = k \geq 2$. We claim that for at least half the pairs $x \neq y \in S$, we have $d(x, y) \geq k/10$. But this follows easily since for any $x \in S$, we have $|B(x, k/5)| \leq k/2$. Since an $\Omega(1)$ fraction of the pairs $x, y \in S$ satisfy $d(x, y) \geq k/10$, and (S, d) is a metric of negative type, we are in position to apply the techniques of [2]. In particular, in order to refer to a result which appears in the literature, we cite the following stronger theorem [1, Theorem 2.1] which itself follows from the techniques of [2, 7, 4].

Theorem 2.3. *There exist constants $C \geq 1$ and $0 < p < \frac{1}{2}$ such that for every n-point metric space (S, d) of negative type and every $\tau > 0$, the following holds. There exists an efficiently computable distribution μ over subsets $U \subseteq S$ such that for every $x, y \in S$ with $d(x, y) \geq \tau$,*

$$\mu \left\{ U : y \in U \text{ and } d(x, U) \geq \frac{\tau}{C\sqrt{\log n}} \right\} \geq p.$$

In particular, using $\tau = k/10$, there must exist some subset $U \subseteq S$ such that for an $\Omega(1)$ fraction of the pairs $x, y \in S$ which satisfy $d(x, y) \geq k/10$, $x \in U$ and $d(y, U) \geq \varepsilon k$ where $\varepsilon \geq 1/O(\sqrt{\log n})$. In particular, choosing $A = U$ and $B = \{y \in S : d(y, U) \geq \varepsilon k\}$ yields the desired separated sets. □

Combining the preceding theorem with Theorem 2.1 yields an $O(\sqrt{\log n} \log \log n)$-approximation for MLA in general n-vertex graphs.

3 Sketch of Theorem 2.1

We proceed using the ideas of [8]. Define

$$W_S(d) = \sum_{uv \in E : u, v \in S} w(u, v)\, d(u, v) \quad \text{and} \quad W(d) = W_V(d) = \sum_{uv \in E} w(u, v)\, d(u, v).$$

Recall that we have an ε-separable metric space (V, d), hence there exist subsets $A, B \subseteq V$ for which $|A|, |B| = \Omega(n)$, and $d(A, B) \geq \varepsilon n$. We consider the cuts $C_0, \ldots, C_t$ for $t \geq \Omega(\varepsilon n)$, where cut C_i separates the vertices of V into two sets: $A_i = \{v \in V : d(v, A) \leq i\}$, and $B_i = V \setminus A_i$. Note that $A \subset A_i$ for all i, and t is chosen to be not too large, so that $B \subset B_i$ for all i. For a cut C_k, we consider the cost of the edges crossing the cut, namely $W_k = \sum_{uv \in E, u \in A_k, v \in B_k} w(u, v)$.

Proposition 3.1. *There are $\Omega(\varepsilon n)$ values of k for which $W_k \leq O(W(d)/\varepsilon n)$.*

Proof. Since $d(u, v) \geq 1$ for every $u, v \in V$, and d is a metric, we have $W(d) \geq \frac{1}{2} \sum_{k=0}^{t} W_k$. As $t = \Omega(\varepsilon n)$, the average value of W_k is at most $O(W(d)/\varepsilon n)$. At most half the W_k may have value more than twice the expectation. □

Now we present a charging argument broken into two cases.

1. For some $0 \leq k \leq t$, $W_k \leq W(d)/n \log n$. In this case, continue recursively to find a linear arrangement for A_k and a linear arrangement for B_k, and concatenate the results. The cost of the concatenated linear order is composed from the cost of edges within A_k (handled by the recursion), cost of edges with B_k (handled by the recursion), and the *concatenation cost*: that of edges connecting A_k and B_k. Each edge of the latter type is of length at most $n - 1$ in the final solution, whereas it contributes length at least 1 to $W(d)$ (since $d(u, v) \geq 1$ for all u, v). Hence the total cost of these edges is at most $W(d)/\log n$.

   Observe that $W_{A_k}(d) + W_{B_k}(d) \leq W(d)$, and there may be at most $O(\log n)$ levels of recursion, because both A_k and B_k are of size at most $n(1 - \Omega(1))$, and hence the total concatenation cost over all levels is at most $O(W(d))$. This implies that the contribution to $\sum_{uv \in E} w(u, v) \cdot |\pi(u) - \pi(v)|$ from this case is at most $O(W(d))$.

2. For every $0 \leq k \leq t$, $W_k > W(d)/n \log n$. Define buckets $B_0, \ldots, B_l$ with $l = O(\log \log n)$ (so that $2^l > (\log n)^{3/2}$), such that bucket B_q contains all cuts C_k for which $W^q \leq W_k \leq 2W^q$, where $W^q = 2^q \frac{W(d)}{n \log n}$. Proposition 3.1 implies that at least one bucket, say B_q, contains at least $r = \Omega(\varepsilon n / \log \log n)$ cuts. Taking all the cuts in B_q partitions the vertices into sets $V_1, V_2, \ldots$, with a natural linear order among these sets, respecting the order of the cuts. For each set V_i the MLA problem is now solved separately by recursion, and the solutions are concatenated in the natural order.

   Again, let us bound the concatenation cost as a function of $W(d)$. The point (as in [8]) is that even though there are r cuts each of cost at most $2W^q$, their total contribution to $\sum_{uv \in E} w(u,v) \cdot |\pi(u) - \pi(v)|$ is at most $4nW^q$ (the value of r is irrelevant to the bound). This is true because every set of vertices V_i contributes "stretch" $|V_i|$ only to two sets of edges represented in B_q: Those that belong to the cut immediately preceding V_i and those that belong to the cut immediately following V_i (such edges must be stretched over the linear arrangement of V_i).

   On the other hand, every cut contributes to $W(d)$ at least its cost, and these costs are additive because if an edge crosses several cuts (of distance at least 1 apart), then its length is at least as large as the number of cuts that it crosses. If follows that $W(d) \geq \Omega(rW_q)$. The ratio between the concatenation cost and the contribution of the same edges to $W(d)$ is then at most $4n/r \leq O(\log \log n / \varepsilon)$.

Acknowledgements. This research was conducted while the authors were visiting Microsoft Research, and we thank Microsoft for their hospitality. The research of the first author was supported in part by the Israeli Science Foundation.

References

[1] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the Sparsest Cut. In *37th Annual Symposium on the Theory of Computing*, pages 553–562, 2005. Submitted to *Journal of the AMS*. Full version available at `http://www.cs.berkeley.edu/~jrl`.

[2] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *36th Annual Symposium on the Theory of Computing*, pages 222–231, 2004.

[3] M. Charikar, M. T. Hajiaghayi, H. Karloff, and S. B. Rao. ℓ_2^2 spreading metrics for ordering problems. To appear, SODA 2006.

[4] S. Chawla, A. Gupta, and H. Räcke. Embeddings of negative-type metrics and an improved approximation to generalized Sparsest Cut. In *Proceedings of the 16th annual ACM-SIAM symposium on Discrete algorithms*, pages 102—111, 2005.

[5] G. Even, J. Naor, S. Rao, and B. Schieber. Divide-and-conquer approximation algorithms via spreading metrics. *J. ACM*, 47(4):585–616, 2000.

[6] P. N. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 682–690, 1993.

[7] J. R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *Proceedings of the 16th annual ACM-SIAM symposium on Discrete algorithms*, pages 92–101, 2005.

[8] S. Rao and A. W. Richa. New approximation techniques for some linear ordering problems. *SIAM J. Comput.*, 34(2):388–404, 2004.