Long plane trees*

Sergio Cabello[†1,2], **Michael Hoffmann**[‡3], **Katharina Klost**[4], **Wolfgang Mulzer**[§4], **and Josef Tkadlec**[5]

1 **Faculty of Mathematics and Physics, University of Ljubljana, Slovenia**
sergio.cabello@fmf.uni-lj.si
2 **Institute of Mathematics, Physics and Mechanics, Slovenia**
3 **Department of Computer Science, ETH Zürich, Switzerland**
hoffmann@inf.ethz.ch
4 **Institut für Informatik, Freie Universität Berlin, Germany**
{kathklost,mulzer}@inf.fu-berlin.de
5 **Harvard University, USA**
pepa_tkadlec@fas.harvard.edu

Abstract

Let $\mathcal{P}$ be a finite point set in the plane in general position. For any spanning tree T on $\mathcal{P}$, we denote by $|T|$ the Euclidean length of T. Let T_{OPT} be a plane (noncrossing) spanning tree of maximum length for $\mathcal{P}$. It is not known whether such a tree can be found in polynomial time. Thus, past research has focused on designing polynomial-time approximation algorithms, typically based on trees of small (unweighted) diameter. We extend this line of research and show how to construct in polynomial time a plane tree T_{ALG} on $\mathcal{P}$ such that T_{ALG} has diameter at most four and $|T_{\text{ALG}}| > 0.546 \cdot |T_{\text{OPT}}|$. This improves substantially over the currently best known approximation factor. Furthermore, we consider the special case of a long plane spanning tree with diameter at most three, and we show that it can be found in polynomial time.

Related Version: A full version is available at https://arxiv.org/abs/2101.00445.

1 Introduction

Geometric network design is a common and well-studied task in computational geometry and combinatorial optimization [6–9]. In this family of problems, we are given a set $\mathcal{P}$ of points in general position, and our task is to connect $\mathcal{P}$ into a (geometric) graph that has certain favorable properties. There are many possible objective functions and many different constraints that might be imposed on the resulting graph.

Here, we focus on graphs that achieve a large total edge length while at the same time ensuring that the edges do not cross. In many cases, the objective of maximization and the noncrossing constraint are in conflict. If the goal is to minimize the total edge length, like, e.g., in Euclidean minimum spanning trees or the Euclidean TSP, the noncrossing property is often implied by the triangle inequality. In contrast, when maximizing, say, the total edge length of a spanning tree, the resulting graph will most likely contain many crossings. In this

* This research was started at the 3rd DACH Workshop on Arrangements and Drawings, August 19–23, 2019, in Wergenstein (GR), Switzerland, and continued at the 16th European Research Week on Geometric Graphs, November 18–22, 2019, in Strobl, Austria. We thank all participants of the workshops for valuable discussions and for creating a conducive research atmosphere.
† Supported by the Slovenian Research Agency (P1-0297, J1-9109, J1-8130, J1-8155, J1-1693, J1-2452).
‡ Supported by the Swiss National Science Foundation within the collaborative DACH project *Arrangements and Drawings* as SNSF Project 200021E-171681.
§ Supported in part by ERC StG 757609 and by the German Research Foundation within the collaborative DACH project *Arrangements and Drawings* as DFG Project MU 3501/3-1.

37th European Workshop on Computational Geometry, St. Petersburg, Russia, April 7–9, 2021.
This is an extended abstract of a presentation given at EuroCG'21. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear eventually in more final form at a conference with formal proceedings and/or in a journal.

sense, balancing the noncrossing constraint and the maximization objective is an interesting challenge.

We will consider long plane spanning trees: given a point set $\mathcal{P}$ in general position (i.e., no three points are on a common line), we want to find a longest spanning tree on $\mathcal{P}$ such that no two edges cross. The precise complexity for this problem is unknown, but it is conjectured to be NP-hard. This stands in contrast to the greedy polynomial time algorithms for short (necessarily plane) spanning trees and long (possibly not plane) spanning trees.

As a polynomial-time algorithm for the long plane case has eluded researchers for many years, the focus has shifted to approximation algorithms. The first such algorithm, giving a 0.5-approximation, is due to Alon et al. [1]. This approximation factor was then improved to 0.502 by Dumitrescu and Tóth [5]. This result led to a series of improvements from 0.503 [3] over 0.512 [4] to the most recent result of a 0.519-approximation [2].

We substantially increase the approximation factor to a fixed $f > 0.5467$, and we give a polynomial time algorithm for finding a longest tree of unweighted diameter at most 3.

2 Approximation Algorithm

We describe an algorithm to find a plane spanning tree on a given point set $\mathcal{P}$, and we show that the resulting tree is a good approximation for a longest plane spanning tree on $\mathcal{P}$. The algorithm considers two families of trees. The first family consists of *stars*: for a point $a \in \mathcal{P}$, the star S_a rooted at a is the tree that connects all points $p \in \mathcal{P} \setminus \{a\}$ to a.

The second family of trees $T_{a,b}$ is parameterized by two distinct points $a, b \in \mathcal{P}$. The trees $T_{a,b}$ are defined as follows: let $\mathcal{P}_a$ be the points of $\mathcal{P}$ that are closer to a than to b, and let $\mathcal{P}_b = \mathcal{P} \setminus \mathcal{P}_a$. We connect a to every point in $\mathcal{P}_b$, and then we connect each point of $\mathcal{P}_a \setminus \{a\}$ to some point of $\mathcal{P}_b$ without introducing crossings. We will not go into the details of the second step here, but it is possible to make these connections in a simple deterministic way. See Figure 1 for an example.

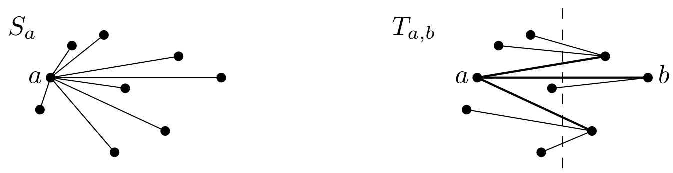

Figure 1 A star S_a and a tree $T_{a,b}$.

The algorithm computes all stars S_a, for $a \in \mathcal{P}$, and all trees $T_{a,b}$, for ordered pairs $(a, b) \in \mathcal{P}^2$ with $a \neq b$. The algorithm returns a longest among all those trees. This process can be implemented in polynomial time, and we call the resulting tree T_{ALG}.

▶ **Theorem 2.1.** *For any finite planar point set $\mathcal{P}$ in general position, the tree T_{ALG} has Euclidean length at least $f \cdot |T_{\text{OPT}}|$, where $|T_{\text{OPT}}|$ denotes the length of a longest plane tree on $\mathcal{P}$ and $f > 0.5467$ is the fourth smallest real root of the polynomial*

$$P(x) = -80 + 128x + 504x^2 - 768x^3 - 845x^4 + 1096x^5 + 256x^6.$$

Proof sketch. *Particularly towards the end of this proof sketch, we omit some details due to space constraints. The detailed calculations can be found in the full version.*

We fix some assumptions on $\mathcal{P}$. First, we assume that $\mathcal{P}$ has diameter exactly 2. Next, fix some optimal tree T_{OPT} on $\mathcal{P}$ and a longest edge ab of T_{OPT}. Write the Euclidean length of ab as $\|ab\| = 2d$. Let $D(a, 2)$ and $D(b, 2)$ be the disks with radius 2 and centers a and b, respectively. As $\mathcal{P}$ has diameter 2, all points lie inside the lens $D(a, 2) \cap D(b, 2)$. Without loss of generality, we assume that $a = (-d, 0)$ and $b = (d, 0)$.

From arguments established in earlier research [4, Lemma 2.1], we can conclude that if $2d \le 1/f$, the result follows. Hence, from now on, we focus on the case $2d > 1/f$.

Let s, s' be the points on the y-axis with $\|sa\| = \|sb\| = \|s'a\| = \|s'b\| = 2df$, where s is the one with the positive y-coordinate. Since $2df > 1$, the circles $\partial D(s, 2df)$ and $\partial D(s', 2df)$ intersect the boundary of the lens. We call the region that is inside the lens but above or below both circles the *far region* (shaded green in Figure 2), and the remaining part of the lens is referred to as the *truncated lens* (shaded yellow in Figure 2).

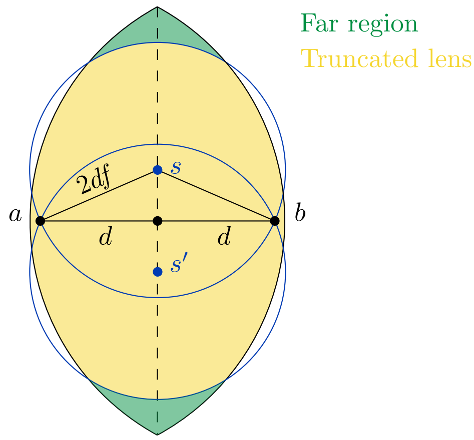

Figure 2 The lens is split into the far region (green) and the truncated lens (yellow).

Suppose that there is a point $c \in \mathcal{P}$ in the far region. Then, we argue that a longest of the stars S_a, S_b, and S_c is a good approximation. Denote by R the circumradius of the triangle $\triangle abc$, and let g be the center of mass of $\mathcal{P} \setminus \{a, b, c\}$. Then, as $\triangle abc$ is acute-angled, there is one point q among a, b, c such that $\|gq\| \ge R$. Having fixed q, the triangle inequality gives $\sum_{p \in \mathcal{P} \setminus \{a,b,c\}} \|pq\| \ge (n-3) \cdot R$. Together with $\|qa\| + \|qb\| + \|qc\| \ge 2R$, this yields $|S_q| = \sum_{p \in \mathcal{P}} \|pq\| \ge (n-1) \cdot R \ge f \cdot (n-1) \cdot 2d \ge f \cdot |T_{\text{OPT}}|$.

Hence, from now on, we can assume that $2d > 1/f$ and that the far region contains no point from $\mathcal{P}$. Consider the five trees $T_{a,b}$, $T_{b,a}$, S_a, S_b, and T_{OPT}. They all contain the edge ab. We conceptually direct the other edges towards ab and, given a point $p \in \mathcal{P}$ and a tree T, denote the length of the outgoing edge from p in T by $\ell_T(p)$. Given a real parameter $\beta \in (0, 1/2)$ we define the weighted average of the lengths of the edges assigned to a point p over the first four trees as:

$$\text{avg}(p, \beta) = (1/2 - \beta) \cdot \ell_{S_a}(p) + \beta \cdot \ell_{T_{a,b}}(p) + \beta \cdot \ell_{T_{b,a}}(p) + (1/2 - \beta) \cdot \ell_{S_b}(p)$$

Summing up over all $p \in \mathcal{P} \setminus \{a, b\}$ and adding the length of the edge ab yields the weighted average of the length of all four trees. Note that a longest of the four trees is guaranteed to be longer than the weighted average. This reduces our problem to that of finding β such that for each point in $\mathcal{P} \setminus \{a, b\}$ we have:

$$\text{avg}(p, \beta) \ge f \cdot \ell_{T_{\text{OPT}}}(p). \tag{1}$$

In the following we assume without loss of generality that $p = (x, y)$ with $x, y \ge 0$. Let p_a be the point with x-coordinate $-(2-d)$ on the ray pa. Furthermore, if the x-coordinate of p is less than d, let p_b be the point with x-coordinate $2-d$ on the ray pb. When the x-coordinate of p is at least d, then the ray pb does not intersect the vertical line with x-coordinate $2-d$ and we set $p_b = b$. Additionally, define p_u to be the furthest point from p on the portion of the boundary of the far region that is contained in the circle $k = \partial D(s, 2df)$. See Figure 3 for a visualization. We claim that

$$\ell_{T_{\mathrm{OPT}}}(p) \leq \min\{2d, \max\{\|pp_a\|, \|pp_u\|\}\}. \tag{2}$$

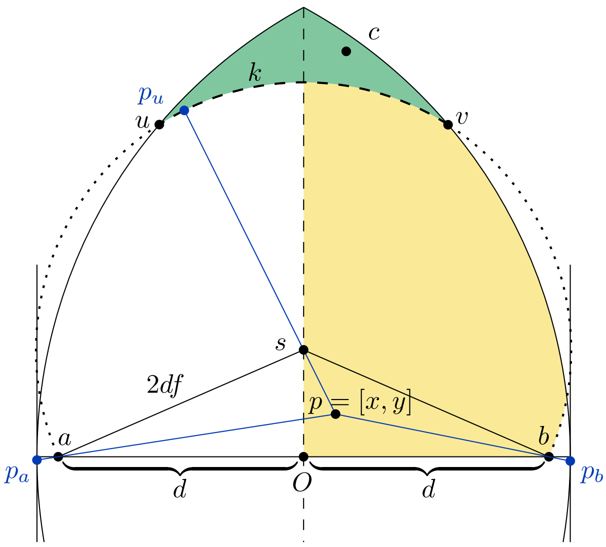

Figure 3 The definition of the special points

To show (2), first note that if $\|pp_a\| \geq 2d$ then we are done as the right-hand side becomes $2d$ and the left-hand side is at most $2d$ by assumption. Next, tedious algebra shows that if $\|pp_a\| \leq 2d$ then $\|pp_b\| \leq \|pp_a\|$. The rest follows by denoting by p_f the point in the truncated lens furthest from p and doing a case distinction over the quadrant containing p_f:

1. If p_f lies in the third or fourth quadrant then $\|pp_a\|$ is larger than $\|pp_f\|$.
2. If p_f lies in the first quadrant, mirroring p_f along the y-axis gives a point further away from p than p_f.
3. If p_f lies in the second quadrant then it lies on the boundary of the truncated lens. Let t be the topmost point in the truncated lens and let u be the to higher intersection point of $\partial D(s, 2df)$ and $\partial D(b, 2)$, see Figure 4 for an illustration. If p_f lies on the arc tu and $p \in N$ we are done, as $p_u = u$. If $p \in S$ we are done by triangle inequality. Finally if p_f lies on the arc ul we have $\|pp_f\| \leq \max\{\|pl\|, \|pu\|\}$ and are again done, as $\|pl\| \leq \|pp_a\|$ and $\|pu\| \leq \|pp_u\|$.

Combining (1) and (2), it suffices to find $\beta \in [0, 1/2]$ such that both $\mathrm{avg}(p, \beta) \geq f \cdot \min\{2d, \|pp_a\|\}$ and $\mathrm{avg}(p, \beta) \geq f \cdot \min\{2d, \|pp_u\|\}$. To find such β, our main technical insight is the following lower bound on $\mathrm{avg}(p, \beta)$ that holds for any $\beta \in [0, 1/2]$:

$$\mathrm{avg}(p, \beta) \geq \frac{d \cdot (1-\beta) + x \cdot 2\beta}{d + x}. \tag{3}$$

To prove (3), we start with unpacking the definition:

$$\mathrm{avg}(p, \beta) \geq (1/2 - \beta) \cdot \|pa\| + \beta \cdot \|pa\| + \beta \cdot x + (1/2 - \beta) \cdot \|pb\|.$$

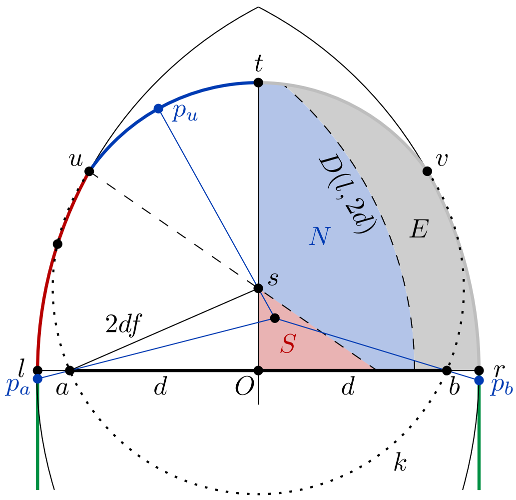

Figure 4 For $p \in E$ we have $\ell_{T_{\mathrm{OPT}}}(p) \le 2d$, for $p \in N$ we have $\ell_{T_{\mathrm{OPT}}}(p) \le \max\{\|pp_a\|, \|pu\|\}$ and for $p \in S$ we have $\ell_{T_{\mathrm{OPT}}}(p) \le \max\{\|pp_a\|, \|pp_u\|\}$.

Let $p' = (-x, y)$ be the reflection of p about the y axis (see Figure 5). Triangle inequality $\|p'p\| + \|pb\| \ge \|p'b\| = \|pa\|$ yields $\beta \cdot x = \frac{1}{2}\beta \cdot \|p'p\| \ge \frac{1}{2}\beta \cdot (\|pa\| - \|pb\|)$ and we get

$$\mathrm{avg}(p,\beta) \ge \frac{1}{2} \cdot \|pa\| + \frac{1}{2}\beta \cdot \|pa\| + \left(\frac{1}{2} - \frac{3}{2}\beta\right) \cdot \|pb\|.$$

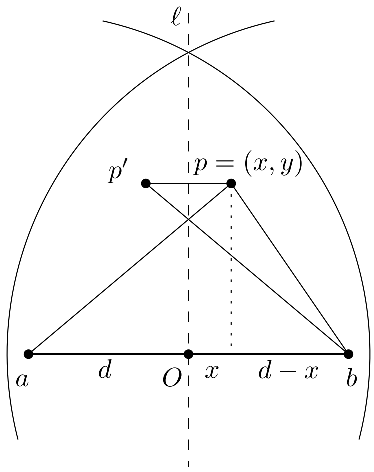

Figure 5 Mirroring p along the y-axis.

Next we claim that $\|pb\| \ge \frac{d-x}{d+x} \cdot \|pa\|$: Indeed, upon squaring, using the Pythagorean theorem and clearing the denominators this becomes $y^2 \cdot 4dx \ge 0$ which is true. Using this bound on the term containing $\|pb\|$, the inequality (3) is proved as follows:

$$\mathrm{avg}(p,\beta) \ge \frac{(1+\beta)(d+x) + (1-3\beta)(d-x)}{2(d+x)} \cdot \|pa\| = \frac{(1-\beta)\cdot d + 2\beta \cdot x}{d+x} \cdot \|pa\|.$$

Using (3) and some elementary geometric considerations (briefly sketched below) we can prove the following claims for any point $p = (x, y)$ with $x, y \ge 0$:

1. We have $\text{avg}(p, \beta) \geq f \cdot \min\{2d, \|pp_a\|\}$, provided that

$$\frac{2f-1}{5-8f} \leq \beta \leq \frac{1}{2} \cdot f. \tag{4}$$

2. We have $\text{avg}(p, \beta) \geq f \cdot \min\{2d, \|pp_u\|\}$, provided that $\beta < \frac{151}{304} \cdot f$ and $\frac{1}{2} \leq f \leq \frac{19}{32}$ and

$$\frac{2f-1}{2\sqrt{5-8f}-1} \leq \beta \leq 1 - f\sqrt{4f^2-1} - 2f^2. \tag{5}$$

To prove 1., we distinguish the cases $x \geq 3d-2$ and $x < 3d-2$ and show $\text{avg}(p, \beta) \geq f \cdot 2d$ and $\text{avg}(p, \beta) \geq f\|pp_a\|$, respectively. To prove 2., we distinguish the cases $y \leq y(u)$ and $y > y(u)$. In the first case we use that the left-hand side of (3) is increasing in y, while $\|pp_u\|$ is decreasing. The second case essentially follows from basic algebra and the Pythagorean theorem.

Finally, it is straightforward (although again tedious) to check that for our definition of f the left-hand side and the right-hand side of the constraint (5) are equal. To be precise, after setting the left-hand side and the right-hand site equal and applying some algebra, which includes two times the squaring of both sites, yields the polynomial given in the statement of the theorem. Our choice of f is not only a solution to the origin polynomial but also yields a value value $\beta^\star \doteq 0.1604$ such that setting $\beta = \beta^\star$ satisfies also (4) and the other constraints in 2. Note that for any $f' > f$, the two constraints on β from (5) are contradictory, except for $f' = \frac{5}{8}$ when they reduce to $-\frac{1}{4} \leq \beta \leq -\frac{1}{4}$ which is not a permissible value of β. ◀

3 Polynomial-time algorithm for diameter 3

In this section we describe a polynomial time algorithm for finding a longest plane spanning tree of diameter at most 3 for a given set of points in general position. Note that all trees of diameter at most 3 have a cut edge ab whose removal decomposes the tree into at most 2 stars, one rooted at a and one rooted at b. We also call such trees *bistars* rooted at a and b.

▶ **Theorem 3.1.** *We can find a longest bistar on a point set $\mathcal{P}$ in polynomial time.*

Proof. We argue how to find the best bistar with fixed roots a, b in polynomial time. The statement of the theorem then follows by iterating over all possible pairs of roots.

Without loss of generality we assume that ab is a horizontal edge and that all points lie above ab. We use dynamic programming. We call a pair $p, q \in \mathcal{P}$ a *valid pair*, if ap and bq do not cross. Let $Q(p, q)$ be the quadrilateral defined by ab, ap, bq and the horizontal line through $\min\{y(p), y(q)\}$ as shown in Figure 6.

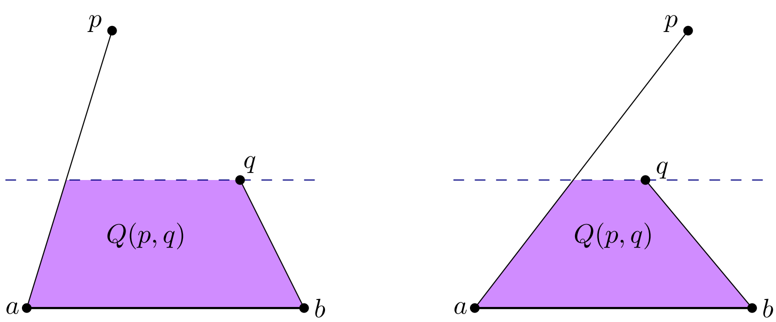

Figure 6 Two examples of valid pairs p, q with their quadrilaterals $Q(p, q)$ shaded.

We denote by $Z(p,q)$ the length of the best bistar rooted at a and b on the points in the interior of $Q(p,q)$. Let $k_{p,q}$ be the highest point in $Q(p,q)$ and let $L_{p,q}$ and $R_{p,q}$ be the regions in $Q(p,q)$ to the left of $ak_{p,q}$ and to the right of $bk_{p,q}$, respectively (see Figure 7).

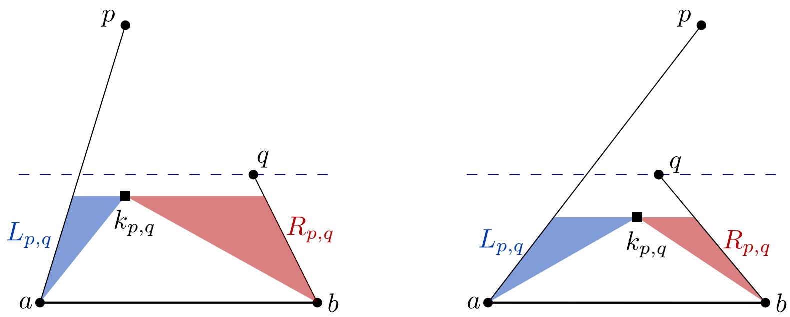

Figure 7 Fixing $k_{p,k}$ gives two possible triangular regions where edges are fixed.

As we aim to compute a bistar, $k_{p,q}$ is either connected to a or to b. In the first case, all points in $L_{p,q}$ have to be connected to a in order to prevent crossings. The points in $Q(k_{p,q},q)$ can now be connected to either a or b, so the best bistar can be found recursively in this region. A symmetric argument holds if $k_{p,q}$ is connected to b. The resulting tree is a bistar rooted at a and b by induction. Formally, for each valid pair p,q we have the following recurrence:

$$Z(p,q) = \begin{cases} 0 & \text{if no point of } \mathcal{P} \text{ is in } Q(p,q), \\ \max \begin{cases} Z(k_{p,q},q) + \|ak_{p,q}\| + \sum_{l \in L_{p,q}} \|al\| \\ Z(p,k_{p,q}) + \|bk_{p,q}\| + \sum_{r \in R_{p,q}} \|br\| \end{cases} & \text{otherwise.} \end{cases}$$

Let p,q be a valid pair, let L_p be all points to the left of the line through ap and let R_q be all points to the right of the line through bq. Then all values of the form

$$\left(\sum_{l \in L_p} \|al\|\right) + \left(\sum_{r \in R_q} \|br\|\right) + \|ap\| + \|bq\| + \|ab\| + Z(p,q), \tag{6}$$

describe the length of a plane, not necessarily spanning, bistar rooted at a and b.

Now assume that a is connected to the point with maximum y-coordinate. Furthermore let q^* be the highest point connected to b and p^* be the point with smallest angle at a above q^*. Then p^*, q^* is a valid pair and the length of the optimal bistar is obtained by plugging p^* and q^* into (6). Now by taking the maximum of (6) for all valid pairs, together with the maximum of S_a and S_b, we obtain a longest bistar rooted at a and b.

The recurrence can be implemented in polynomial time using a standard dynamic programming approach. ◀

In the full version of the paper, we give details on the implementation and show that the dynamic program can be implemented in $\mathcal{O}(n^2)$ time.

4 Conclusion

We showed an $f \geq 0.5467$-approximation algorithm for the longest plane spanning tree problem. Furthermore we gave a polynomial time algorithm to solve the problem for fixed diameter at most 3.

There are some related open questions. First, is the factor f we obtained tight for our algorithm? While all the steps in the analysis are tight, the overall analysis might not be. Second, how well does the optimal tree of diameter 3 approximate the optimal tree of arbitrary diameter? This is especially interesting, as we know of point sets with longest trees of diameter $\Theta(n)$. Finally, there are the broader questions of finding a PTAS and settling the complexity.

References

1. Noga Alon, Sridhar Rajagopalan, and Subhash Suri. Long non-crossing configurations in the plane. *Fundam. Inform.*, 22(4):385–394, 1995. doi:10.3233/FI-1995-2245.
2. Ahmad Biniaz. Euclidean bottleneck bounded-degree spanning tree ratios. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 826–836. SIAM, 2020. doi:10.1137/1.9781611975994.50.
3. Ahmad Biniaz, Prosenjit Bose, Kimberly Crosbie, Jean-Lou De Carufel, David Eppstein, Anil Maheshwari, and Michiel Smid. Maximum plane trees in multipartite geometric graphs. *Algorithmica*, 81(4):1512–1534, 2019. doi:10.1007/s00453-018-0482-x.
4. Sergio Cabello, Aruni Choudhary, Michael Hoffmann, Katharina Klost, Meghana M Reddy, Wolfgang Mulzer, Felix Schröder, and Josef Tkadlec. A better approximation for longest noncrossing spanning trees. In *36th European Workshop on Computational Geometry (EuroCG)*, 2020.
5. Adrian Dumitrescu and Csaba D. Tóth. Long non-crossing configurations in the plane. *Discrete Comput. Geom.*, 44(4):727–752, 2010. doi:10.1007/s00454-010-9277-9.
6. David Eppstein. Spanning trees and spanners. In Jörg-Rüdiger Sack and Jorge Urrutia, editors, *Handbook of Computational Geometry*, pages 425–461. North Holland / Elsevier, 2000. doi:10.1016/b978-044482537-7/50010-3.
7. Sariel Har-Peled. *Geometric approximation algorithms*, volume 173 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011. doi:10.1090/surv/173.
8. Joseph S. B. Mitchell. Shortest paths and networks. In Jacob E. Goodman, Joseph O'Rourke, and Csaba D. Tóth, editors, *Handbook of Discrete and Computational Geometry*, chapter 31, pages 811–848. CRC Press, Boca Raton, 3rd edition, 2017. doi:10.1201/9781315119601.
9. Joseph S. B. Mitchell and Wolfgang Mulzer. Proximity algorithms. In Jacob E. Goodman, Joseph O'Rourke, and Csaba D. Tóth, editors, *Handbook of Discrete and Computational Geometry*, chapter 32, pages 849–874. CRC Press, Boca Raton, 3rd edition, 2017. doi:10.1201/9781315119601.