On the Longest Spanning Tree with Neighborhoods

Ke Chen[1] and Adrian Dumitrescu[1]

University of Wisconsin–Milwaukee, Milwaukee, WI, USA
kechen@uwm.edu,dumitres@uwm.edu

Abstract. We study a maximization problem for geometric network design. Given a set of n compact neighborhoods in $\mathbb{R}^d$, select a point in each neighborhood, so that the longest spanning tree on these points (as vertices) has maximum length. Here we give an approximation algorithm with ratio 0.511, which represents the first, albeit small, improvement beyond 1/2. While we suspect that the problem is NP-hard already in the plane, this issue remains open.

Keywords: Maximum (longest) spanning tree, neighborhood, geometric network, metric problem, approximation algorithm.

1 Introduction

In the *Euclidean Maximum Spanning Tree Problem* (EMST), given a set of points in the Euclidean space $\mathbb{R}^d$, $d \geq 2$, one seeks a tree that connects these points (as vertices) and has maximum length. The problem is easily solvable in polynomial time by Prim's algorithm or by Kruskal's algorithm; algorithms that take advantage of the geometry are also available [13]. In the *Longest Spanning Tree with Neighborhoods* (Max-St-N), each point is replaced by a point-set, called *region* or *neighborhood*, and the tree must connect n representative points, one chosen from each region (duplicate representatives are allowed), and the tree has maximum length. The tree edges are straight line segments connecting pairs of points in distinct regions; for obvious reasons we refer to these edges as *bichromatic*. As one would expect, the difficulty lies in choosing the representative points; once these points are selected, the problem is reduced to the graph setting and is thus easily solvable.

The input $\mathcal{N}$ consists of n (possibly disconnected) neighborhoods. For simplicity, it is assumed that each neighborhood is a union of polyhedral regions; the total vertex complexity of the input is N. However, it will be apparent from the context that our methods extend to a broader class of regions, those approximable by unions of polyhedral regions within a prescribed accuracy (for instance unions of balls of arbitrary radii, etc).

Examples. It is worth noting that a greedy algorithm does not necessarily find an optimal tree. Let $\mathcal{N} = \{X_1, X_2, X_3\}$, where $X_1 = \{a, b\}$, $X_2 = \{a, c\}$, $X_3 = \{d\}$, Δabc is a unit equilateral triangle and d is the midpoint of bc; see Figure 1 (left).

Fig. 1: Left: an example on which the greedy algorithm is suboptimal. Right: an example of a long (still suboptimal) spanning tree with 10 regions $\mathcal{N} = \{\mathrm{A}, \mathrm{S} \cup \mathrm{S}, \mathrm{E} \cup \mathrm{E} \cup \mathrm{E}, \mathrm{T} \cup \mathrm{T}, \mathrm{O} \cup \mathrm{O}, \mathrm{F}, \mathrm{N} \cup \mathrm{N}, \mathrm{R}, \mathrm{G}, \mathrm{I}\}$ (some regions are disconnected); the blue segments form a spanning tree on $\mathcal{N}$ and the green dots are the chosen representative points.

A (natural) greedy algorithm chooses two points attaining a maximum inter-point distance with points in distinct regions, and then repeatedly chooses a point in each new region as far as possible from some selected point. Here the selection $b \in X_1$, $c \in X_2$, $d \in X_3$ yields a spanning tree in the form of a star centered at $v_1 = b$ of length $3/2$; on the other hand, selecting vertices $v_i \in X_i$, $i = 1, \ldots, 3$ at a, a, d, respectively, yields a spanning tree in the form of a 2-edge star centered at $v_3 = d$ of length $2 \times \sqrt{3}/2 = \sqrt{3}$ (the edge lengths in the underlying complete graph are $\sqrt{3}/2$, $\sqrt{3}/2$, and 0). Another example appears in Figure 1 (right).

We start by providing a factor $1/2$ approximation to Max-St-N. We then offer two refinement steps achieving a better ratio. The last refinement step proves Theorem 1.

Theorem 1. *Given a set $\mathcal{N}$ of n neighborhoods in $\mathbb{R}^d$ (with total vertex complexity N), a ratio 0.511 approximation for the maximum spanning tree for the regions in $\mathcal{N}$ can be computed in polynomial time.*

Although our improvement in the approximation ratio for spanning trees is very small, it shows that the "barrier" of $1/2$ can be broken. On the other hand, we show that every algorithm that always includes a bichromatic diameter pair in the solution (as the vertices of the corresponding regions) is bound to have an approximation ratio at most $\sqrt{2-\sqrt{3}} = 0.517\ldots$ (via Figure 4 in Section 3).

Definitions and notations. A *geometric graph* G is a graph whose vertices (a finite set) are points in $\mathbb{R}^d$ and whose edges consist of straight line segments. For two points $p, q \in \mathbb{R}^d$, the Euclidean distance between them is denoted by $|pq|$. The *length* of G, denoted $\mathrm{len}(G)$, is the sum of the Euclidean lengths of all edges in G.

For a neighborhood $X \in \mathcal{N}$, let $V(X)$ denote its set of vertices. Let $V = \cup_{X \in \mathcal{N}} V(X)$ denote the union of vertices of all neighborhoods in $\mathcal{N}$; put $N = |V|$.

Given a set $\mathcal{N}$ of n neighborhoods, we define the following parameters. A *monochromatic diameter* pair is a pair of points in the same region attaining a maximum distance. A *bichromatic diameter* pair is a pair of points from two regions attaining a maximum distance, i.e., $p_i \in X_i$, $p_j \in X_j$, where $X_i, X_j \in \mathcal{N}$, $i \neq j$, and $|p_i p_j|$ is maximum. For $X \in \mathcal{N}$ and $p \in X$, let $d_{\max}(p)$ denote the maximum distance between p and any point of a neighborhood $Y \in \mathcal{N} \setminus \{X\}$. It is well known and easy to prove that both a monochromatic diameter and bichromatic diameter pair are attained by pairs of vertices in the input instance. An optimal (longest) Spanning Tree with neighborhoods is denoted by T_{OPT}; it is a geometric graph whose vertices are the representative points of the n regions.

Preliminaries and related work. Computing the minimum or maximum Euclidean spanning trees of a point set are classical problems in a geometric setting [13, 14]. A broad collection of problems in geometric network design, including the classical *Euclidean Traveling Salesman Problem* (ETSP), can be found in the surveys [9, 11, 12]. While past research has primarily focused on minimization problems, the maximization variants usually require different techniques and so they are interesting in their own right and pose many unmet challenges; e.g., see the section devoted to longest subgraph problems in the survey of Bern and Eppstein [5]. The results obtained in this area in the last 20 years are rather sparse; the few articles [4, 8, 10] make a representative sample.

Spanning trees for systems of neighborhoods have also been studied. For instance, given a set of n (possibly disconnected) compact neighborhoods in $\mathbb{R}^d$, select a point in each neighborhood so that the minimum spanning tree on these points has minimum length [7, 18], or maximum length [7], respectively. In the cycle version first studied by Arkin and Hassin [3], called *TSP with neighborhoods* (TSPN), given a set of neighborhoods in $\mathbb{R}^d$, one must find a shortest closed curve (tour) intersecting each region.

2 Approximation Algorithms

For simplicity, we present our algorithms for the plane, namely $d = 2$; the extension to higher dimensions is straightforward, and is briefly discussed at the end.

Let $S = \{p_1, \ldots, p_n\}$, where $p_i = (x_i, y_i)$. Given a point $p \in S$, the *star centered at* p, denoted S_p, is the spanning tree on S whose edges connect p to the other points. Using a technique developed in [8] (in fact a simplification of an earlier approach used in [2]), we first obtain a simple approximation algorithm with ratio 1/2.

Algorithm ***A1***. Compute a bichromatic diameter of the point set V, pick an arbitrary point (vertex) from each of the other $n-2$ neighborhoods, and output the longest of the two stars centered at one of the endpoints of the diameter.

Analysis. Let ab be a bichromatic diameter pair, and assume without loss of generality that ab is a horizontal unit segment, where $a = (0, 0)$ and $b = (1, 0)$. We may assume that $a \in X_1$ and $b \in X_2$; refer to Figure 2. The ratio 1/2 (or

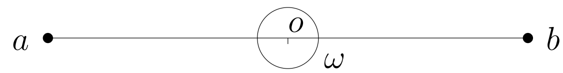

Fig. 2: A bichromatic diameter pair a, b and the disk ω.

$\frac{n}{2n-2}$ which is slightly better) follows from the next lemma in conjunction with the obvious upper bound

$$\mathrm{len}(T_{\mathrm{OPT}}) \leq n - 1. \tag{1}$$

The latter is implied by the fact that each edge of T_{OPT} is bichromatic and thus of length at most 1.

Lemma 1. *Let S_a and S_b be the stars centered at the points a and b, respectively. Then $\mathrm{len}(S_a) + \mathrm{len}(S_b) \geq n$.*

Proof. Assume that $a = p_1$, $b = p_2$. For each $i = 3, \ldots, n$, the triangle inequality for the triple a, b, p_i gives

$$|ap_i| + |bp_i| \geq |ab| = 1.$$

By summing up we have

$$\mathrm{len}(S_a) + \mathrm{len}(S_b) = \sum_{i=3}^{n} (|ap_i| + |bp_i|) + 2|ab| \geq (n-2) + 2 = n,$$

as required. □

We next refine this algorithm to achieve an approximation ratio of 0.511. The technique uses two parameters x and y, introduced below. The smallest value of the ratio obtained over the entire range of admissible x and y is determined and output as the approximation ratio of Algorithm **A2**.

Let o be the midpoint of ab, and ω be the disk centered at o, of minimum radius, say, x, containing at least $\lfloor n/2 \rfloor$ of the neighborhoods $X_3, \ldots, X_n$; in particular, this implies that we can consider $\lfloor n/2 \rfloor$ neighborhoods as contained in ω and $\lceil n/2 \rceil$ neighborhoods having points on the boundary $\partial\omega$ or in the exterior of ω. We may assume that $x \leq 0.2$; if $x \geq 0.2$, the 0.511 approximation ratio easily follows (with room to spare): Since for each of the regions not contained in ω, one of the connections from an arbitrary point of the region to a or b is at least $\sqrt{\frac{1}{4} + x^2}$. If T is the spanning tree consisting of all such longer connections together with ab, then

$$\begin{aligned}
\mathrm{len}(T) &\geq 1 + \frac{1}{2} \left\lfloor \frac{n}{2} \right\rfloor + \left(\left\lceil \frac{n}{2} \right\rceil - 2 \right) \sqrt{\frac{1}{4} + x^2} \\
&\geq \frac{1 + \sqrt{1 + 4x^2}}{4} (n-1) + 1 - \frac{3\sqrt{1 + 4x^2}}{4} \\
&\geq \frac{5 + \sqrt{29}}{20} (n-1) + 1 - \frac{3\sqrt{29}}{20} \geq \frac{5 + \sqrt{29}}{20} (n-1).
\end{aligned}$$

So the approximation ratio is at least $(5+\sqrt{29})/20 = 0.519\ldots$

Let the monochromatic diameter of V be $1+y$, for some $y \in [-1,\infty)$; the next lemma shows that $y \leq 1$, and so the monochromatic diameter of V is $1+y$, for some $y \in [-1,1]$.

Lemma 2. *For every $X \in \mathcal{N}$,* $\mathrm{diam}(X) \leq 2$.

Proof. Let pq be a diameter pair of X. Let r be an arbitrary point of an arbitrary neighborhood $Y \in \mathcal{N} \setminus \{X\}$. By the triangle inequality, we have $|pq| \leq |pr| + |rq| \leq 1+1 = 2$, as required. □

If $y \geq 0.2$, let $a_1, b_1 \in X_1$ be a corresponding diameter pair; choose a point in every other region and connect it to a_1 and b_1. Since $|a_1b_1| = 1+y \geq 1.2$, the longer of the two stars centered at a_1 and b_1 has length at least $(n-1)(1+y)/2 \geq 0.6(n-1)$; this candidate spanning tree offers thereby this ratio of approximation. We will subsequently assume that $y \in [-1, 0.2]$.

As shown above a constant approximation ratio better than 1/2 can be obtained if x or y is sufficiently large. In the complementary case (both x and y are small), an upper bound of cn, for some constant $c < 1$, on the length of T_{OPT} can be derived. We continue with the technical details.

Algorithm ***A2.*** The algorithm computes one or two candidate solutions. The first candidate solution T_1 for the spanning tree is only relevant for the range $y \geq 0$ (if $y < 0$ its length could be smaller than $n/2$). Assume that one of the regions, say, X_1 achieves a diameter pair: $a_1, b_1 \in X_1$; recall that $|a_1b_1| = 1+y$. Choose an arbitrary point in every other region and connect it to a_1 and b_1. Let T_1 be the longer of the two stars centered at a_1 and b_1. By the triangle inequality,

$$\mathrm{len}(T_1) \geq (n-1)\,\frac{1+y}{2}. \tag{2}$$

The second candidate solution T_2 for the spanning tree connects each of the regions contained in ω with either a or b at a cost of at least 1/2 (based on the fact that $\max\{|ap_i|, |bp_i|\} \geq |ab|/2 = 1/2$). For each region X_i, $i \geq 3$, select the vertex of X_i that is farthest from o and connect it with a or b, whichever yields the longer connection. As such, if X_i is not contained in ω, the connection length is at least $\sqrt{\frac{1}{4}+x^2}$. Finally add the unit segment ab. Then,

$$\mathrm{len}(T_2) \geq 1 + \left\lfloor \frac{n}{2} \right\rfloor \frac{1}{2} + \left(\left\lceil \frac{n}{2} \right\rceil - 2 \right) \sqrt{\frac{1}{4} + x^2}. \tag{3}$$

The above expression can be simplified as follows. If n is even, (3) yields

$$\begin{aligned}
\mathrm{len}(T_2) &\geq 1 + \frac{n}{4} + \left(\frac{n}{4} - 1\right)\sqrt{1+4x^2} \\
&= \frac{n-1}{4}\left(1 + \sqrt{1+4x^2}\right) + \left(\frac{5}{4} - \frac{3}{4}\sqrt{1+4x^2}\right) \\
&\geq \frac{n-1}{4}\left(1 + \sqrt{1+4x^2}\right).
\end{aligned}$$

If n is odd, (3) yields

$$\begin{aligned}\mathrm{len}(T_2) &\geq 1 + \frac{n-1}{4} + \left(\frac{n+1}{4} - 1\right)\sqrt{1+4x^2} \\ &= \frac{n-1}{4}\left(1+\sqrt{1+4x^2}\right) + \left(1 - \frac{2}{4}\sqrt{1+4x^2}\right) \\ &\geq \frac{n-1}{4}\left(1+\sqrt{1+4x^2}\right).\end{aligned}$$

Consequently, for every n we have

$$\mathrm{len}(T_2) \geq \frac{n-1}{4}\left(1+\sqrt{1+4x^2}\right). \tag{4}$$

Upper bound on $\mathrm{len}(T_{OPT})$. Let Ω be the disk of radius $R(y)$ centered at o, where

$$R(y) = \begin{cases} \frac{\sqrt{3}}{2} & \text{if } y \leq 0 \\ \frac{\sqrt{3}}{2} + \frac{2}{\sqrt{3}}y & \text{if } y \geq 0 \end{cases}$$

Lemma 3. *V is contained in Ω.*

Proof. Assume for contradiction that there exists a point $p_i \in X_i$ at distance larger than $R(y)$ from o. By symmetry, we may assume that $|ap_i| \leq |bp_i|$ and that p_i lies in the closed halfplane above the line containing ab.

First consider the case $y \leq 0$; it follows that $|bp_i| > \sqrt{\frac{1}{4} + \frac{3}{4}} = 1$. If $i = 2$, then $b, p_i \in X_2$, which contradicts the definition of y; otherwise $b \in X_2$ and $p_i \in X_i$ are points in different neighborhoods at distance larger than 1, in contradiction with the original assumption on the bichromatic diameter of V.

Next consider the case $y \geq 0$; it follows that $|bp_i| \geq \sqrt{\frac{1}{4} + (\frac{\sqrt{3}}{2} + \frac{2}{\sqrt{3}}y)^2} > 1 + y$. If $i = 2$, then $b, p_i \in X_2$, which contradicts the definition of y; otherwise $b \in X_2$ and $p_i \in X_i$ are points in different neighborhoods at distance larger than 1, in contradiction with the original assumption on the bichromatic diameter of V.

In either case (for any y) we have reached a contradiction, and this concludes the proof. □

Recall that for a point $p \in X \in \mathcal{N}$, $d_{\max}(p)$ is the maximum distance between p and any point of a neighborhood $Y \in \mathcal{N} \setminus \{X\}$.

Lemma 4. *Let $\mathcal{N} = \{X_1, \ldots, X_n\}$ be a set of n neighborhoods and T_{OPT} be an optimal spanning tree assumed to connect points (vertices) $p_i \in X_i$ for $i = 1, \ldots, n$. For every $j \in [n]$, we have*

$$\mathrm{len}(T_{OPT}) \leq \sum_{i \neq j} d_{max}(p_i).$$

Proof. Consider T_{OPT} rooted at p_j. Let $\pi(v)$ denote the parent of a (non-root) vertex v. Uniquely assign each edge $\pi(v)v$ of T_{OPT} to vertex v. The inequality $\mathrm{len}(\pi(v)v) \leq d_{\max}(v)$ holds for each edge of the tree. By adding up the above inequalities, the lemma follows. □

Lemma 5. *If $X \in \mathcal{N}$ is contained in ω, and $p \in X$, then $d_{max}(p) \leq \min(1, x + R(y))$.*

Proof. By definition, $d_{\max}(p) \leq 1$. By Lemma 3, the vertex set V is contained in Ω; equivalently, all neighborhoods in $\mathcal{N}$ are contained in Ω. By the triangle inequality, $d_{\max}(p) \leq |po| + R(y) \leq x + R(y)$, as claimed. □

Lemma 6. *The following holds:*

$$\mathrm{len}(T_{OPT}) \leq (n-1) \cdot \min\left(1, \frac{1 + x + R(y)}{2}\right). \tag{5}$$

Proof. Let T_{OPT} be a longest spanning tree of $p_1, \ldots, p_n$, where $p_i \in X_i$, for $i = 1, \ldots, n$. View T_{OPT} as rooted at $p_1 \in X_1$; recall that $a \in X_1$. By Lemma 4,

$$\mathrm{len}(T_{\mathrm{OPT}}) \leq \sum_{i=2}^{n} d_{\max}(p_i).$$

If X_i is not contained in ω, $d_{\max}(p_i) \leq 1$; otherwise, by Lemma 5, $d_{\max}(p_i) \leq \min(1, x + R(y))$. By the setting of x in the definition of ω, we have

$$\mathrm{len}(T_{\mathrm{OPT}}) \leq \left(\left\lceil \frac{n}{2} \right\rceil - 1\right) \cdot 1 + \left\lfloor \frac{n}{2} \right\rfloor \cdot \min(1, x + R(y)).$$

If n is even, the above inequality yields

$$\begin{aligned}
\mathrm{len}(T_{\mathrm{OPT}}) &\leq \left(\frac{n}{2} - 1\right) + \frac{n}{2} \min(1, x + R(y)) \\
&= \frac{n-1}{2} \left(1 + x + R(y)\right) + \frac{\min(1, x + R(y)) - 1}{2} \\
&\leq \frac{n-1}{2} \left(1 + x + R(y)\right),
\end{aligned}$$

while if n is odd, it yields

$$\mathrm{len}(T_{\mathrm{OPT}}) \leq \frac{n-1}{2} + \frac{n-1}{2} \left(x + R(y)\right) = \frac{n-1}{2} \left(1 + x + R(y)\right).$$

Therefore $\mathrm{len}(T_{\mathrm{OPT}}) \leq \frac{n-1}{2} \left(1 + x + R(y)\right)$ in both cases. Then the lemma follows by adjoining the trivial upper bound in equation (1). □

3 Analysis of Algorithm A2

We start with a preliminary argument for ratio 0.506 that comes with a simpler proof. We then give a sharper analysis for ratio 0.511.

A first bound on the approximation ratio of ***A2****.* First consider the case $y < 0$. Then $R(y) = \sqrt{3}/2$, so the ratio of **A2** is at least

$$\min_{\substack{0 \le x \le 0.2 \\ y < 0}} \frac{\mathrm{len}(T_2)}{\mathrm{len}(T_{\mathrm{OPT}})} \ge \min_{0 \le x \le 0.2} \frac{1 + \sqrt{1 + 4x^2}}{\min\left(4, 2 + \sqrt{3} + 2x\right)}.$$

A standard analysis shows that this ratio achieves its minimum $\left(1 + 2\sqrt{2 - \sqrt{3}}\right) \Big/ 4 = 0.508\ldots$ when $x = 1 - \sqrt{3}/2$.

When $y \ge 0$, the ratio of **A2** is at least

$$\min_{0 \le x, y \le 0.2} \max\left(\frac{\mathrm{len}(T_1)}{\mathrm{len}(T_{\mathrm{OPT}})}, \frac{\mathrm{len}(T_2)}{\mathrm{len}(T_{\mathrm{OPT}})}\right).$$

The inequalities (2), (4), (5) imply that this ratio is at least

$$\frac{\max\left(1 + y, (1 + \sqrt{1 + 4x^2})/2\right)}{\min\left(2, 1 + x + R(y)\right)} = \frac{\max\left(1 + y, (1 + \sqrt{1 + 4x^2})/2\right)}{\min\left(2, 1 + \frac{\sqrt{3}}{2} + x + \frac{2}{\sqrt{3}}\, y\right)}.$$

Since the analysis is similar to that for deriving the refined bound we give next, we state without providing details that this piecewise function reaches its minimum value

$$\left(4\sqrt{3} - 1 - 2\sqrt{9 - 3\sqrt{3}}\right) \Big/ 4 = 0.506\ldots$$

when

$$y = \left(4\sqrt{3} - 3 - 2\sqrt{9 - 3\sqrt{3}}\right) \Big/ 2 = 0.0137\ldots$$

and

$$x = \sqrt{3}/2 - 3 + 2\sqrt{3 - \sqrt{3}} = 0.1180\ldots$$

This provides a preliminary ratio 0.506 in Theorem 1.

A refined bound. Let $m = \lfloor n/2 \rfloor$. Assume for convenience that the regions $X_3, \ldots, X_n$ are relabeled so that $X_3, \ldots, X_{m+2}$ are contained in ω and $X_{m+3}, \ldots, X_n$ are not contained in the interior of ω. Recall that $p_i \in X_i$ are the representative points in an optimal solution T_{OPT}. Let $x_i = |op_i|$, for $i = 3, \ldots, m + 2$; as such, $x_3, \ldots, x_{m+2} \le x$. Let the average of $x_3, \ldots, x_{m+2}$ be λx, where $\lambda \in [0, 1]$, i.e., $\sum_{i=3}^{m+2} x_i = m\lambda x$.

Observe that $d_{\max}(p_i) \le |op_i| + R(y) = x_i + R(y)$, for $i = 3, \ldots, m + 2$. Consequently, the upper bound in (5) can be improved to

$$\mathrm{len}(T_{\mathrm{OPT}}) \le \frac{n - 1}{2} \left(1 + \lambda x + R(y)\right). \tag{6}$$

We next obtain an improved lower bound on $\mathrm{len}(T_2)$. Recall that Algorithm **A2** selects the vertex of X_i that is farthest from o for every $i \ge 3$, and connects it with a or b, whichever yields the longer connection. In particular, the length of this connection is at least $\sqrt{\frac{1}{4}+x_i^2}$ for $i=3,\dots,m+2$. Since the function $\sqrt{x}$ is concave, Jensen's inequality yields:

$$\sum_{i=3}^{m+2}\sqrt{1+4x_i^2} \geq m\sqrt{1+4\lambda^2x^2},$$

hence we obtain the following sharpening of the lower bound in (4):

$$\mathrm{len}(T_2) \geq \frac{n-1}{4}\left(\sqrt{1+4\lambda^2x^2}+\sqrt{1+4x^2}\right). \tag{7}$$

In order to handle (6) and (7) we make a key substitution $z=\lambda x$ and simplify the lower bound in (7). Recall that $0 \leq \lambda \leq 1$, and so $0 \leq z \leq x$ and $z \in [0, 0.2]$. We now deduce from (6) and (7) that

$$\mathrm{len}(T_{\mathrm{OPT}}) \leq \frac{n-1}{2}\left(1+z+R(y)\right), \tag{8}$$

and

$$\mathrm{len}(T_2) \geq \frac{n-1}{2}\sqrt{1+4z^2}. \tag{9}$$

To analyze the approximation ratio we distinguish two cases:

Case 1: $y \leq 0$. Then $R(y)=\sqrt{3}/2$, so the ratio of **A2** is at least

$$\min_{0\leq z\leq 0.2}\ \max\left(\frac{\mathrm{len}(T_2)}{\mathrm{len}(T_{\mathrm{OPT}})}\right) \geq \min_{0\leq z\leq 0.2}\frac{2\sqrt{1+4z^2}}{\min\left(4, 2+2z+\sqrt{3}\right)}.$$

When $4 \leq 2+2z+\sqrt{3}$, we have $z \geq 1-\sqrt{3}/2$. Then

$$\frac{\sqrt{1+4z^2}}{2} \geq \frac{\sqrt{8-4\sqrt{3}}}{2} = \sqrt{2-\sqrt{3}} = 0.517\dots.$$

When $4 \geq 2+2z+\sqrt{3}$, or $z \leq 1-\sqrt{3}/2$, let

$$f(z) = \frac{2\sqrt{1+4z^2}}{2+\sqrt{3}+2z}.$$

Then

$$f'(z) = \frac{8\left(2+\sqrt{3}\right)z-4}{\sqrt{1+4z^2}\left(2+\sqrt{3}+2z\right)^2}.$$

Since $8\left(2+\sqrt{3}\right)z-4 \leq 4\left(2+\sqrt{3}\right)\left(2-\sqrt{3}\right)-4=0$, the function is non-increasing on $[0, 1-\sqrt{3}/2]$ and so

$$f(z) \geq f\left(1-\sqrt{3}/2\right) = \sqrt{2-\sqrt{3}} = 0.517\dots.$$

This concludes the proof for the first case.

Case 2: $y \geq 0$. Then the ratio of **A2** is at least

$$\min_{0 \leq y, z \leq 0.2} \max\left(\frac{\text{len}(T_1)}{\text{len}(T_{\text{OPT}})}, \frac{\text{len}(T_2)}{\text{len}(T_{\text{OPT}})}\right).$$

For $0 \leq y, z \leq 0.2$, let

$$g(z, y) = \frac{\max\left(1 + y, \sqrt{1 + 4z^2}\right)}{\min\left(2, 1 + z + R(y)\right)} = \frac{\max\left(1 + y, \sqrt{1 + 4z^2}\right)}{\min\left(2, 1 + \frac{\sqrt{3}}{2} + z + \frac{2}{\sqrt{3}}\, y\right)}.$$

The inequalities (2), (8), (9) imply that the ratio of **A2** is at least

$$\min_{0 \leq y, z \leq 0.2} g(z, y).$$

The curve $\gamma : 1 + y = \sqrt{1 + 4z^2}$ and the line $\ell : 2 = 1 + \frac{\sqrt{3}}{2} + z + \frac{2}{\sqrt{3}}\, y$ split the feasible region $[0, 0.2] \times [0, 0.2]$ into four subregions; see Figure 3. The

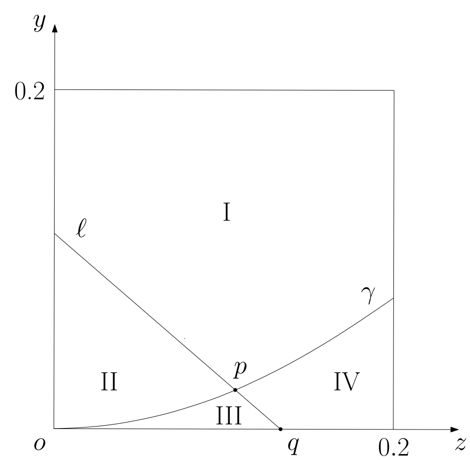

Fig. 3: The feasible region of the function $g(z, y)$.

curve γ intersects line ℓ at point $p = (z_0, y_0)$, where $z_0 = \left(8\sqrt[4]{3} - \sqrt{3} - 6\right)/\,26 = 0.1075\ldots$ and $y_0 = \left(8\sqrt{3} - 2\sqrt[4]{27} - 9\right)/\,13 = 0.0228\ldots$ Set

$$\rho := (1 + y_0)/2 = \left(4\sqrt{3} + 2 - \sqrt[4]{27}\right)\Big/\,13 = 0.5114\ldots \tag{10}$$

In region I, $g(z, y) = (1 + y)/2$. It reaches the minimum value ρ when y is minimized, i.e., $y = y_0$.

In region II, $g(z,y) = \dfrac{1+y}{1+\sqrt{3}/2+z+2y/\sqrt{3}}$. Its partial derivative is positive, i.e.,

$$\frac{\partial g}{\partial y} = \frac{1-\sqrt{3}/6+z}{\left(1+\sqrt{3}/2+z+2y/\sqrt{3}\right)^2} > 0,$$

so $g(z,y)$ reaches its minimum value on the curve γ. On this curve, let

$$G(z) = g\left(z, y(z)\right) = \frac{\sqrt{1+4z^2}}{1-\sqrt{3}/6+z+2\sqrt{1+4z^2}/\sqrt{3}}.$$

Its derivative is

$$G'(z) = \frac{\left(4-2\sqrt{3}/3\right)z-1}{\sqrt{1+4z^2}\left(1-\sqrt{3}/6+z+2\sqrt{1+4z^2}/\sqrt{3}\right)^2}.$$

Note that the numerator of $G'(z)$ is negative, i.e., $\left(4-2\sqrt{3}/3\right)z-1 < 4z-1 < 0$ for $z \in [0, 0.2]$, thus $G'(z) < 0$. So the minimum value is ρ, and is achieved when z is maximized, i.e., $z = z_0$.

In region IV, $g(z,y) = \sqrt{1+4z^2}/2$ which increases monotonically with respect to z. So the minimum value is again ρ and is achieved when z is minimized, i.e., $z = z_0$.

In region III,

$$g(z,y) = \frac{\sqrt{1+4z^2}}{1+\sqrt{3}/2+z+2y/\sqrt{3}}.$$

Its partial derivative is negative, i.e.,

$$\frac{\partial g}{\partial y} = \frac{-2\sqrt{1+4z^2}}{\sqrt{3}\left(1+\sqrt{3}/2+z+2y/\sqrt{3}\right)^2} < 0,$$

so $g(z,y)$ reaches its minimum value on the arc $op \subset \gamma$ or the segment $pq \subset \ell$, where $q = (1-\sqrt{3}/2, 0)$ is the intersection point of ℓ and the z-axis. Since these two curves are shared with region II and IV respectively, by previous analyses, $g(z,y)$ reaches its minimum value ρ at point p.

In summary, we showed that

$$\min_{0 \le y, z \le 0.2} g(z,y) \ge \rho = 0.511\ldots,$$

establishing the approximation ratio in Theorem 1.

Remark. The algorithm can be adapted to work in $\mathbb{R}^d$ for any $d \ge 3$. In the analysis, the disk ω becomes the ball of radius x with the same defining property; the disk Ω becomes the ball of radius $R(y)$. All arguments and relevant bounds still hold since they only rely on the triangle inequality; the verification is left to the reader. Consequently, the approximation guarantee remains the same.

An almost tight example. Let Δabc be an isosceles triangle with $|ca| = |cb| = 1 - \varepsilon$, $|ab| = 1$, for a small $\varepsilon > 0$; e.g., set $\varepsilon = 1/(n-1)$. Let $\mathcal{N} = \{X_1, \ldots, X_n\}$, where $X_1 = ac$, $X_2 = bc$, and $X_3, \ldots, X_n$ are $n-2$ points at distance $1-\varepsilon$ from c, below ab and whose projections onto ab are close to the midpoint of ab (see Figure 4). The spanning tree constructed by **A2** is of length close to

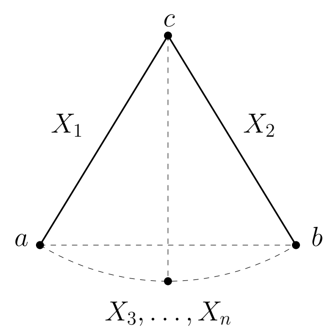

Fig. 4: A tight example.

$\sqrt{2-\sqrt{3}}n = 0.517\ldots n$, while the longest spanning tree has length at least $(1-\varepsilon)(n-1) = n-2$; as such, the approximation ratio of **A2** approaches $\sqrt{2-\sqrt{3}} = 0.517\ldots$ for large n. Note that this is a tight example for the case $y \leq 0$, for which the ratio of **A2** is at least $\sqrt{2-\sqrt{3}}$; and an almost tight example in general, since the overall approximation ratio of **A2** is 0.511. Moreover, the example shows that every algorithm that always includes a bichromatic diameter pair in the solution (as the vertices of the corresponding regions) is bound to have an approximation ratio at most $\sqrt{2-\sqrt{3}}$.

Time complexity of Algorithm ***A2****.* It is straightforward to implement the algorithm to run in quadratic time for any fixed d. All interpoint distances can be easily computed in $O(N^2)$ time. Similarly the farthest point from o in each region (over all regions) can all be computed in $O(N)$ time. Subquadratic algorithms for computing the diameter and farthest bichromatic pairs in higher dimensions can be found in [1, 6, 15–17]; see also the two survey articles [9, 11].

4 Conclusion

We gave two approximation algorithms for Max-St-N: a very simple one with ratio 1/2 and another simple one (with slightly more elaborate analysis but equally simple principles) with ratio 0.511. The following variants represent extensions of the Euclidean maximum TSP for the neighborhood setting.

In the *Euclidean Maximum Traveling Salesman Problem*, given a set of points in the Euclidean space $\mathbb{R}^d$, $d \geq 2$, one seeks a cycle (a.k.a. *tour*) that visits these points (as vertices) and has maximum length; see [4]. In the *Maximum Traveling Salesman Problem with Neighborhoods* (MAX-TSP-N), each point is replaced by a point-set, called *region* or *neighborhood*, and the cycle must connect n representative points, one chosen from each region (duplicate representatives are allowed), and the cycle has maximum length. Since the original variant with points is NP-hard when $d \geq 3$ (as shown in [4]), the variant with neighborhoods is also NP-hard for $d \geq 3$. The complexity of the original problem in the plane is unsettled, although the problem is believed to be NP-hard [10]. In the *path* variant, one seeks a path of maximum length.

The following problems are proposed for future study:

1. What is the computational complexity of MAX-ST-N?
2. What approximations can be obtained for the *cycle* or *path* variants of MAX-TSP-N?

References

1. P. K. Agarwal, J. Matoušek, and S. Suri, Farthest neighbors, maximum spanning trees and related problems in higher dimensions, *Comput. Geom.* **1** (1992), 189–201.
2. N. Alon, S. Rajagopalan, and S. Suri, Long non-crossing configurations in the plane, *Fundamenta Informaticae* **22** (1995), 385–394.
3. E. M. Arkin and R. Hassin, Approximation algorithms for the geometric covering salesman problem, *Discrete Appl. Math.* **55** (1994), 197–218.
4. A. I. Barvinok, S. Fekete, D. S. Johnson, A. Tamir, G. J. Woeginger, and R. Woodroofe, The geometric maximum traveling salesman problem, *Journal of ACM* **50(5)** (2003), 641–664.
5. M. Bern and D. Eppstein, Approximation algorithms for geometric problems, in *Approximation Algorithms for NP-hard Problems (D. S. Hochbaum, ed.)*, PWS Publishing Company, Boston, MA, 1997, pp. 296–345.
6. B. K. Bhattacharya and G. T. Toussaint, Efficient algorithms for computing the maximum distance between two finite planar sets, *Journal of Algorithms* **4** (1983), 121–136.
7. R. Dorrigiv, R. Fraser, M. He, S. Kamali, A. Kawamura, A. López-Ortiz, and D. Seco, On minimum- and maximum-weight minimum spanning trees with neighborhoods, *Theory Comput. Syst.* **56(1)** (2015), 220–250.
8. A. Dumitrescu and Cs. D. Tóth, Long non-crossing configurations in the plane, *Discrete and Computational Geometry* **44(4)** (2010), 727–752.
9. D. Eppstein, Spanning trees and spanners, in J.-R. Sack and J. Urrutia (editors), *Handbook of Computational Geometry*, pages 425–461, Elsevier Science, Amsterdam, 2000.
10. S. Fekete, Simplicity and hardness of the maximum traveling salesman problem under geometric distances, *Proc. 10th Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM/SIAM, 1999, pp. 337–345.
11. J. S. B. Mitchell, Geometric shortest paths and network optimization, in J.-R. Sack and J. Urrutia (editors), *Handbook of Computational Geometry*, pages 633–701, Elsevier Science, Amsterdam, 2000.

12. J. S. B. Mitchell, Shortest paths and networks, in *Handbook of Computational Geometry (J. E. Goodman and J. O'Rourke, eds.)*, Chapman & Hall/CRC, 2nd edition, 2004, pp. 607–641.
13. C. Monma, M. Paterson, S. Suri, and F. Yao, Computing Euclidean maximum spanning trees, *Algorithmica* **5** (1990), 407–419.
14. F. P. Preparata and M. I. Shamos, *Computational Geometry*, Springer-Verlag, New York, 1985.
15. E. A. Ramos, An optimal deterministic algorithm for computing the diameter of a three-dimensional point set, *Discrete and Computational Geometry* **26(2)** (2001), 233–244.
16. J. M. Robert, Maximum distance between two sets of points in $\mathbb{R}^d$, *Pattern Recognition Letters* **14** (1993), 733–735.
17. G. T. Toussaint and M. A. McAlear, A simple $O(n \log n)$ algorithm for finding the maximum distance between two finite planar sets, *Pattern Recognition Letters* **1** (1982), 21–24.
18. Y. Yang, M. Lin, J. Xu, and Y. Xie, Minimum spanning trees with neighborhoods, In *Proc. 3rd Int. Conf. on Algor. Aspects in Information and Management* (AAIM), vol. 4508 of LNCS, Springer, 2007, pp. 306–316.