# The Median Line Segment Problem: Computational Complexity and Constrained Variants

Ovidiu Daescu[*] Ka Yaw Teo[†]


## Abstract

In the *median line segment problem*, we are given a set $P$ of $n$ points in $\mathbb{R}^d$ and a real number $\ell > 0$ with the objective to find a line segment of length $\ell$ such that the sum of the Euclidean distances from $P$ to the line segment is minimized. We prove that, in general, it is impossible to construct a median line segment for $n \geq 3$ non-collinear points in the plane by using only ruler and compass. We then consider two constrained variants of the median line segment problem in $\mathbb{R}^2$ – i) point-anchored and ii) constant-slope. In the point-anchored variant, an endpoint of the median line segment is given as input, whereas in the constant-slope variant, the orientation of the median line segment is fixed. We present a deterministic $(1+\varepsilon)$-approximation algorithm for solving each constrained variant. For approximating a point-anchored median line segment, we give a space-subdivision method with a time complexity of $O(n\varepsilon^{-2}\alpha_\theta^{-1})$, where $\alpha_\theta$ is a parameter dependent on the coordinates of $P$. For approximating a constant-slope median line segment, a prune-and-search approach is used, and its time complexity is $O(kn \log n)$, where $k$ is inversely proportional to $\varepsilon$.


## 1 Introduction

The *median line segment problem* is formally defined as follows.

*Given a set $P$ of $n$ points in $\mathbb{R}^d$ and a positive real number $\ell$, locate a line segment of length $\ell$ such that the sum of the Euclidean distances from $P$ to the located line segment is minimized.*

The problem applies to any real-world scenario that involves finding a best location for any object that could be modeled as a line segment. The problem could arise in many industries and sectors, where we wish to find the optimal placement of various facilities to maximize their efficiency, impact, and profit. These facilities may include highways, railroads, pipelines, telecommunication lines, electronic circuit connectors, and electrodes. In addition to location science, the median line segment problem could have potential applications in other subject areas with less obvious connections such as cluster analysis in data science and pattern recognition in computer vision.

The median line segment problem is closely related to one of the oldest non-trivial problems in facility location theory – the (generalized) *Fermat-Torricelli problem*, which asks to find a point with the minimal sum of distances to a given set of $n$ points. The optimal point is referred to as the Fermat-Torricelli point or simply the (geometric) median. For $n \geq 5$ points in general position, it has been proven that the Fermat-Torricelli point cannot be constructed by strict usage of ruler and compass [1, 7]. In other words, the Fermat-Torricelli problem is unsolvable by radicals over the field of rationals. Consequently, no exact algorithm exists for solving the problem under computational models with basic arithmetic operations and the extraction of $k$-th roots. This leaves us with only numerical or symbolic approximation methods for $n \geq 5$ points (e.g., see [2, 3, 4]). Furthermore, it remains unclear whether the problem is in $\mathcal{NP}$.

Another problem related to ours is the *median line problem*, which asks to locate a line minimizing the sum of the distances between a given set of $n$ points and the located line. When considering the median line problem in two dimensions, the optimal line has been shown to exhibit the following properties. The median line must divide the given points into two equal halves, and must pass through at least two of the given points [8]. As a consequence, the median line problem could be solved exactly in $O(n^2)$ time (by mainly exploiting the property that the optimal line must contain a pair of given points) [6]. The optimal solution could also be found in $O(h \log n)$ time, where $h$ is the number of halving lines [11, 12]. Currently, the best upper bound for $h$ is $O(n^{4/3})$. However, no exact algorithm is known to solve the median line problem in higher dimensions.

Unlike the Fermat-Torricelli problem and the median line problem, which have been extensively studied over the years, the median line segment problem has not thus far received any proper attention in the literature.

[*]Department of Computer Science, University of Texas at Dallas, ovidiu.daescu@utdallas.edu

[†]Department of Computer Science, University of Texas at Dallas, ka.teo@utdallas.edu

## 2 Our results

We prove that it is impossible to construct a median line segment for $n \geq 3$ non-collinear points in $\mathbb{R}^2$ by using only ruler and compass (Section 4). We then consider the median line segment problem under different geometric constraints. Particularly, we derive a $(1+\varepsilon)$-approximation algorithm for solving the point-anchored median line segment problem in the plane (Section 5). In this constrained problem, an endpoint of the median line segment is given as part of the input. By essentially dividing the space around the anchor point into $O(n)$ intervals with certain geometric properties, our algorithm finds an approximate solution in $O(n\varepsilon^{-2}\alpha_\theta^{-1})$ time, where $\alpha_\theta$ is a parameter dependent on the coordinates of $P$. Furthermore, we provide an algorithm for computing a $(1+\varepsilon)$-approximate constant-slope median line segment in $\mathbb{R}^2$, where the slope of the median line segment is fixed at input (Section 6). Our algorithm is a tailored extension of the prune-and-search approach given by Bose et al. [2], and its running time is $O(kn \log n)$, where $k = \frac{2\pi}{\cos^{-1}(1+\varepsilon)^{-2}}$.

## 3 Preliminaries

For any two points $a$ and $b$ in $\mathbb{R}^d$, let $ab$ denote the line segment bounded by $a$ and $b$, and let $\|ab\| = \|b-a\|$ be the Euclidean distance between $a$ and $b$.

For any line segment $ab$ in $\mathbb{R}^d$, let $H_a$ (resp. $H_b$) be the hyperplane containing $a$ (resp. $b$) and orthogonal to $ab$. Let $S_a$ (resp. $S_b$) be the closed half-space bounded by $H_a$ (resp. $H_b$) and not containing $ab$. Define $S_{ab} = \mathbb{R}^d \setminus (S_a \cup S_b)$.

For a line segment $ab$ in $\mathbb{R}^2$, let $L_{ab}$ be the line containing $ab$. Let $H^+$ denote a closed half-plane bounded by $L_{ab}$, and let $H^- = \mathbb{R}^2 \setminus H^+$. Define $S_a^+ = S_a \cap H^+$, $S_a^- = S_a \cap H^-$, $S_b^+ = S_b \cap H^+$, $S_b^- = S_b \cap H^-$, $S_{ab}^+ = S_{ab} \cap H^+$, and $S_{ab}^- = S_{ab} \cap H^-$ (Figure 1).

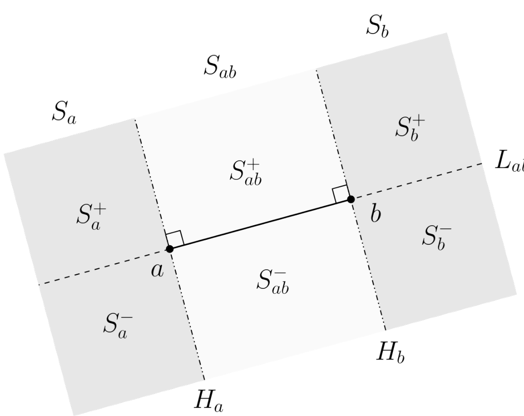


Figure 1: Regions defined with respect to a line segment $ab$.

We assume, without loss of generality, that the points of $P$ have been uniformly scaled such that the length of the median line segment is $\ell = 1$. Let $\mathcal{D}$ denote the diameter of point set $P$. Note that if $\ell \geq \mathcal{D}$, then our problem effectively becomes the median line problem. Thus, in this paper, we assume that $\ell < \mathcal{D}$, unless specified otherwise.

A line segment $s$ is said to be a $(1+\varepsilon)$-approximate solution if the sum of the distances from $P$ to $s$ is at most $(1+\varepsilon)$ times that of the optimal line segment.

## 4 Inconstructibility of the median line segment

**Theorem 1** *The construction of a median line segment is, in general, impossible for $n = 3$ and more points in the plane by strict usage of ruler and compass.*

**Proof.** In order to prove the theorem, we require the following lemma.

**Lemma 2** *Let $p^*$ denote the Fermat-Torricelli point for a point set $\{p_1, p_2, p_3\}$. Let $\beta = \arg\max_i \|p^* p_i\|$. For $i \neq \beta$, let $\eta_i$ be the distance from $p_\beta$ to the foot of the altitude from $p_i$ in triangle $\triangle p_1 p_2 p_3$.*

*A. If $\ell \leq \|p^* p_\beta\|$, then there exists a median line segment $s^* = a^* b^*$ such that its endpoint $a^*$ coincides with $p^*$, and $s^*$ lies in $p^* p_\beta$ (Figure 2A).*

*B. If $\ell > \|p^* p_\beta\|$, then there is a median line segment $s^* = a^* b^*$ such that its endpoint $a^*$ coincides with $p_\beta$.*

*(1) $l \leq \min\{\eta_i : i \neq \beta\}$. For $i \neq \beta$, let $\phi_i$ be the acute angle formed by $b^* p_i$ and the line supporting $s^*$. Endpoint $b^*$ must be located such that $\phi_i = \phi_j$, where $i, j \neq \beta$ and $i \neq j$ (Figure 2B).*

*(2) $l > \min\{\eta_i : i \neq \beta\}$. For $i \neq \beta$, let $q_i$ be the closest point on $s^*$ to $p_i$, and let $w_i$ denote the distance from $a^*$ to $q_i$. Note that $w_i \in [0,1]$. Let $\bar{d}_i$ denote the vector from $q_i$ to $p_i$, and let $\bar{h}_i$ be the component of $\bar{d}_i$ normal to $s^*$ multiplied by $w_i$. For $i, j \neq \beta$ and $i \neq j$, endpoint $b^*$ must be located such that $\|\bar{h}_i\| / \|\bar{d}_i\| = \|\bar{h}_j\| / \|\bar{d}_j\|$.*

**Proof.** We refer to the full paper for the proof. □

Part A of Lemma 2 essentially implies that if $\ell \leq \|p^* p_\beta\|$, then a median line segment $s^*$ can be constructed by using ruler and compass, since the exact Euclidean construction of the Fermat-Torricelli point for $n = 3$ points is possible. However, in part B of Lemma 2 ($\ell > \|p^* p_\beta\|$) – case 1 in particular – in order to construct a median line segment $s^*$, we have to look

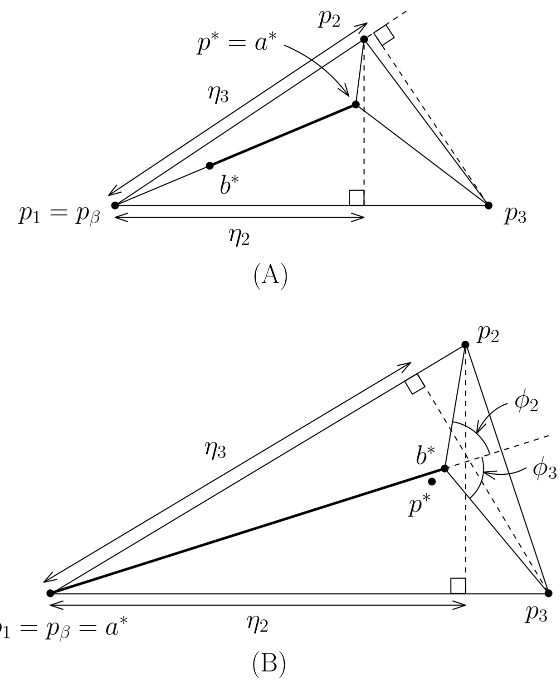


Figure 2: Illustrations for Lemma 2. (A) Part A. (B) Case 1 of part B.

for a point $b^*$ on the circumference of a circle of radius $\ell$ centered at $a^* = p_\beta$ such that the rays emanating from $p_i$ and $p_j$, where $i \neq j$ and $i, j \neq \beta$, meeting at $b^*$ make equal angles with the normal at $b^*$. This is known as (and equivalent to) the Alhazen's billiard problem, to which the general solution has been proven to be inconstructible using only ruler and compass [9]. Briefly, the problem requires solving a quartic equation that is irreducible over $\mathbb{Q}$ (and so does not have constructible solutions). Hence, we conclude that the ruler-and-compass construction of a median line segment is, in general, impossible for $n = 3$ (and more) points. □

## 5 Approximating the point-anchored median line segment

In this section, we consider the following restricted variant of the median line segment problem.

*Given a set $P$ of $n$ points in $\mathbb{R}^2$, a point $q \in \mathbb{R}^2$, and a real number $\ell > 0$, find a line segment of length $\ell$ with an endpoint at $q$ such that the sum of the Euclidean distances from $P$ to the line segment is minimized.*

**Remark 1** *It follows from the proof of Theorem 1 that the point-anchored median line segment problem is, in general, not solvable by radicals over $\mathbb{Q}$ for $n \geq 2$ points.*

**Theorem 3** *For the point-anchored median line segment problem in $\mathbb{R}^2$, given any $\varepsilon > 0$, one can compute a $(1+\varepsilon)$-approximate solution in time $O(n\varepsilon^{-2}\alpha_\theta^{-1})$, where $\alpha_\theta$ is a function dependent on the coordinates of $P$.*

**Proof.** Let $s$ denote any line segment of length $\ell$ with an endpoint fixed at $q$. Assume, without loss of generality, that the fixed endpoint of line segment $s$ is $a = q = (0, 0)$ (by a translation of $P$), and the length of line segment $s$ is $\ell = 1$ (through a uniform scaling of $P$). Let $\theta$ be the counterclockwise angle of line segment $s$ with respect to the positive $x$-axis rooted at $a$. The sum of the distances from $P = \{p_1, p_2, ..., p_n\}$ to line segment $s$ is given by the following objective function:

$$f(\theta) = \sum_{\substack{1 \leq i \leq n \\ p_i \in S_a}} \sqrt{{x_i}^2 + {y_i}^2} + \sum_{\substack{1 \leq i \leq n \\ p_i \in S_{ab}^+}} \left[x_i(-\sin\theta) + y_i \cos\theta\right] + \sum_{\substack{1 \leq i \leq n \\ p_i \in S_{ab}^-}} \left[-x_i(-\sin\theta) - y_i \cos\theta\right] + \sum_{\substack{1 \leq i \leq n \\ p_i \in S_b}} \sqrt{(x_i - \cos\theta)^2 + (y_i - \sin\theta)^2}$$

where $x_i$ and $y_i$ are the $x$- and $y$-coordinates of $p_i \in P$, respectively. We consider $\theta \in [0, \pi/2)$ only, and each subsequent quadrant can be handled analogously. The quadrant $[0, \pi/2)$ can be divided into a set $T$ of at most $\Theta(n)$ contiguous intervals, in each of which the subsets of points of $P$ in $S_a$, $S_{ab}^+$, $S_{ab}^-$, $S_b^+$, and $S_b^-$, respectively, remain constant. We partition each interval of $T$ into a number of small sub-intervals such that the relative error in computing the sum of the distances from $P$ to a line segment $s$, whose angle $\theta$ is given by a boundary of a sub-interval, does not exceed $\varepsilon$.

To evaluate the number of sub-intervals, we perform the following analysis. Let $I$ denote a sub-interval. Suppose that the optimal line segment $s^*$ lies within $I$. First, we note that the distance from any given point $p_i \in S_a$ to endpoint $a$ of line segment $s$ remains constant within sub-interval $I$. For simplicity of notation, the subscript $i$ is dropped, and $p$ is equivalent to $p_i$ hereafter.

For a point $p \in S_{ab}^+$, let $d_p = d(p, s)$ denote its orthogonal distance to a line segment $s$ whose location is defined by a boundary of interval $I$ (Figure 3A). Suppose that $d_p^* = d(p, s^*)$ is the distance from $p$ to the optimal line segment $s^*$. We rotate the coordinate system such that the positive $x$-axis contains $s$, and the first quadrant of the defined $xy$-plane contains sub-interval $I$ (and thus $s^*$). Specifically, consider the worst-case scenario where $s$ and $s^*$ are located at the two ends of sub-interval $I$. Let $\Delta\theta$ be the size of sub-interval $I$. In addition, let $x_p$ and $y_p$ denote the $x$- and $y$-coordinates,

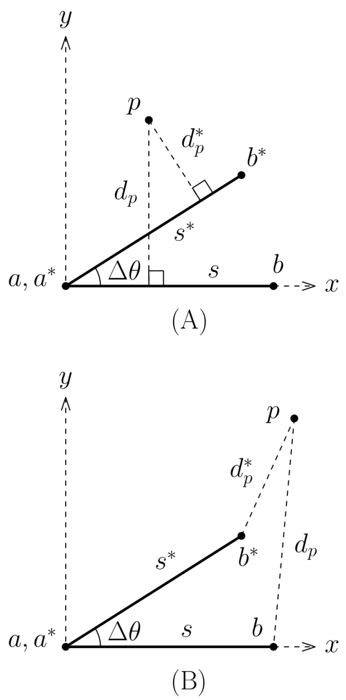


Figure 3: A point $p \in P$ located in (A) $S^+_{ab}$ or (B) $S^+_b$.

respectively, of point $p$. In order to have $d_p \leq (1+\varepsilon)d^*_p$, the following must hold:

$$
\begin{aligned}
d_p &\leq (1+\varepsilon)\, d^*_p \\
y_p &\leq (1+\varepsilon)\,(-x_p \sin\Delta\theta + y_p \cos\Delta\theta) \\
\frac{1}{1+\varepsilon} &\leq -\frac{x_p}{y_p}\sin\Delta\theta + \cos\Delta\theta \\
&= \sqrt{1+\left(\frac{x_p}{y_p}\right)^2}\cos\left(\Delta\theta + \tan^{-1}\frac{x_p}{y_p}\right) \\
\Delta\theta &\leq \cos^{-1}\left(\frac{1}{(1+\varepsilon)\sqrt{1+\left(\frac{x_p}{y_p}\right)^2}}\right) - \tan^{-1}\frac{x_p}{y_p}
\end{aligned}
$$

Let $A^+_{ab,p}$ denote the right-hand term of the last inequality above. Given that

$$
\begin{aligned}
A^+_{ab,p} &\geq \varepsilon\left(\cos^{-1}\left(\frac{1}{2\sqrt{1+\left(\frac{x_p}{y_p}\right)^2}}\right) - \tan^{-1}\frac{x_p}{y_p}\right) \\
&= \varepsilon\alpha^+_{ab,p}
\end{aligned}
$$

for $0 < \varepsilon < 1$, if we have $\Delta\theta = \varepsilon\alpha^+_{ab,p}$, then the desired condition $d_p \leq (1+\varepsilon)d^*_p$ is fulfilled. Note that $\alpha^+_{ab,p}$ is a trigonometric function in terms of the coordinates of point $p$. We can satisfy $d_p \leq (1+\varepsilon)d^*_p$ for all points $p \in S^+_{ab}$ if we set $\Delta\theta = \varepsilon \cdot \min\{\alpha^+_{ab,p} : p \in {S_{ab}}^+\}$.

The analysis for $S^-_{ab}$ is similar to that for $S^+_{ab}$ due to symmetry, and we obtain $\{\alpha^-_{ab,p} : p \in S^-_{ab}\}$ accordingly.

We can also perform a similar analysis for each point $p \in S^+_b$. Let $d_p = d(p, s)$ denote the distance from $p$ to endpoint $b$ of a line segment $s$ located at a boundary of sub-interval $I$ (Figure 3B). Let $d^*_p = d(p, s^*)$ be the shortest distance from $p$ to the optimal line segment $s^*$. As before, we define a coordinate system on $s$ such that the positive $x$-axis contains $s$, and the first quadrant of the $xy$-plane contains sub-interval $I$, at whose boundaries $s$ and $s^*$ are positioned. Let $\Delta\theta$ be the size of sub-interval $I$. If $d_p \leq (1+\varepsilon)d^*_p$, then we have

$$
\begin{aligned}
&d_p \leq (1+\varepsilon)\, d^*_p \\
&\sqrt{(x_p-1)^2 + {y_p}^2} \\
&\qquad \leq (1+\varepsilon)\sqrt{(x_p - \cos\Delta\theta)^2 + (y_p - \sin\Delta\theta)^2} \\
&\frac{(x_p-1)^2 + {y_p}^2}{(1+\varepsilon)^2} \\
&\qquad \leq (x_p - \cos\Delta\theta)^2 + (y_p - \sin\Delta\theta)^2 \\
&\qquad = {x_p}^2 - 2x_p\cos\Delta\theta + {y_p}^2 - 2y_p\sin\Delta\theta + 1 \\
&-\frac{1}{2}\left(\frac{(x_p-1)^2 + {y_p}^2}{(1+\varepsilon)^2} - {x_p}^2 - {y_p}^2 - 1\right) \\
&\qquad \geq x_p\cos\Delta\theta + y_p\sin\Delta\theta \\
&\qquad = \sqrt{{x_p}^2 + {y_p}^2}\cos\left(\Delta\theta + \tan^{-1}\left(-\frac{x_p}{y_p}\right)\right) \\
&\Delta\theta \leq \tan^{-1}\left(\frac{x_p}{y_p}\right) - \cos^{-1}\left(-\frac{1}{2\sqrt{{x_p}^2 + {y_p}^2}}\right. \\
&\qquad \left.\left(\frac{(x_p-1)^2 + {y_p}^2}{(1+\varepsilon)^2} - {x_p}^2 - {y_p}^2 - 1\right)\right)
\end{aligned}
$$

Let $A^+_{b,p}$ denote the right-hand side of the last inequality above. Since

$$
\begin{aligned}
A^+_{b,p} &\geq \varepsilon^2\left[\tan^{-1}\left(\frac{x_p}{y_p}\right) - \cos^{-1}\left(-\frac{1}{2\sqrt{{x_p}^2 + {y_p}^2}}\right.\right. \\
&\qquad \left.\left.\left(\frac{(x_p-1)^2 + {y_p}^2}{(1+\varepsilon')^2} - {x_p}^2 - {y_p}^2 - 1\right)\right)\right] \\
&= \varepsilon^2\alpha^+_{b,p}
\end{aligned}
$$

where $\varepsilon' = \min(1, \varepsilon_p)$,

$$
\varepsilon_p = \frac{\sqrt{(x_p-1)^2 + {y_p}^2}}{\sqrt{\left(x_p - \frac{x_p}{\sqrt{{x_p}^2+{y_p}^2}}\right)^2 + \left(y_p - \frac{y_p}{\sqrt{{x_p}^2+{y_p}^2}}\right)^2}} - 1
$$

and $0 < \varepsilon \leq \varepsilon' < 1$, if we set $\Delta\theta = \varepsilon^2\alpha^+_{b,p}$, then $d_p \leq (1+\varepsilon)d^*_p$ is satisfied. Note that $\alpha^+_{b,p}$ is a trigonometric function dependent on the coordinates of point $p$. In order to uphold $d_p \le (1+\varepsilon)d_p^*$ for all points $p \in S_b^+$, we can simply set $\Delta\theta = \varepsilon^2 \cdot \min\{\alpha_{b,p}^+ : p \in S_b^+\}$.

Points $p \in S_b^-$ can be handled analogously as those in $S_b^+$, and we obtain $\{\alpha_{b,p}^- : p \in S_b^-\}$ as the result.

In summary, for each given interval $\tau \in T$, we compute $\alpha_{ab}^+ = \min\{\alpha_{ab,p}^+ : p \in S_{ab}^+\}$, $\alpha_{ab}^- = \min\{\alpha_{ab,p}^- : p \in S_{ab}^-\}$, $\alpha_b^+ = \min\{\alpha_{b,p}^+ : p \in S_b^+\}$, and $\alpha_b^- = \min\{\alpha_{b,p}^- : p \in S_b^-\}$. We then use $\Delta\theta = \min\left\{\varepsilon\alpha_{ab}^+, \varepsilon\alpha_{ab}^-, \varepsilon^2\alpha_b^+, \varepsilon^2\alpha_b^-\right\}$ for partitioning the given interval $\tau$ into sub-intervals of size at most $\Delta\theta$.

We now derive an upper bound on the number of sub-intervals as follows. Let $s(\tau)$ denote the set $\{\alpha_{ab}^+, \alpha_{ab}^-, \alpha_b^+, \alpha_b^-\}$ computed for each interval $\tau$ of $T$. Define $\alpha_\theta = \min\{\alpha \in s(\tau) : \tau \in T\}$. Then, we have a total of $2\pi/(\varepsilon^2\alpha_\theta)$ sub-intervals in the worst case. Since it takes $O(n)$ algebraic operations to compute the sum of distances for each candidate line segment (defined by the boundaries of the sub-intervals), we can obtain a solution, whose sum of distances to $P$ is at most $(1+\varepsilon)$ times that of the optimal solution, in $2\pi n/(\varepsilon^2\alpha_\theta) = O(n\varepsilon^{-2}\alpha_\theta^{-1})$ time. □

## 6 Approximating the constant-slope median line segment

In this section, we address a constrained variant of the median line segment problem stated as follows.

*Given a set $P$ of $n$ points in $\mathbb{R}^2$, an angle $\theta$, and a real number $\ell > 0$, find a line segment of length $\ell$ making angle $\theta$ with the abscissa axis such that the sum of the Euclidean distances from $P$ to the line segment is minimized.*

**Theorem 4** *For the constant-slope median line segment problem in $\mathbb{R}^2$, given any $\varepsilon > 0$, one can find a line segment, whose sum of distances to $P$ is at most $(1+\varepsilon)$ times that of the optimal line segment, in time $O(kn\log n)$, where $k = \frac{2\pi}{\cos^{-1}(1+\varepsilon)^{-2}}$.*

**Proof.** We denote by $s = ab$ any line segment of length $\ell$ making angle $\theta$ with the positive $x$-axis. Assume, without loss of generality, that $\theta = 0$ and $\ell = 1$. Let $x_a$ and $y_a$ be the $x$- and $y$-coordinates of the endpoint $a$ of line segment $s$, respectively. Then, the sum of the distances from $P = \{p_1, p_2, ..., p_n\}$ to line segment $s$ can be written as the following objective function:

$$\begin{aligned} f(s) &= f(x_a, y_a) \\ &= \sum_{\substack{1\le i\le n \\ p_i \in S_a}} \sqrt{(x_i - x_a)^2 + (y_i - y_a)^2} \\ &\quad + \sum_{\substack{1\le i\le n \\ p_i \in S_{ab}^+}} (y_i - y_a) + \sum_{\substack{1\le i\le n \\ p_i \in S_{ab}^-}} (y_a - y_i) \\ &\quad + \sum_{\substack{1\le i\le n \\ p_i \in S_b}} \sqrt{(x_i - x_a - 1)^2 + (y_i - y_a)^2} \end{aligned}$$

where $x_i$ and $y_i$ are the $x$- and $y$-coordinates of $p_i \in P$, respectively.

**Remark 2** *$f$ is a piecewise convex function, where each piece consists of a sum of two convex functions and two linear functions, and the transition between any two consecutive pieces corresponds to a point of $P$ moving between $S_a$, $S_{ab}^+$, $S_{ab}^-$, and $S_b$. Since the number of such transitions is bounded by $\Theta(n^2)$, the minimum of function $f$ can be obtained by solving $\Theta(n^2)$ two-variable convex optimization problems.*

We begin by defining the so-called $k$-oriented distance function $d_k$ [5, 10] to approximate the Euclidean distance as follows.

**$k$-oriented distance.** A cone in $\mathbb{R}^2$ is defined as the intersection of two half-planes, each of whose supporting lines contains the origin $O$. A simplicial cone $c$ has a diameter bounded by an angle $\gamma$ if, for any two points $p$ and $q$ in $c$, we have $\angle pOq \le \gamma$. Let $C = \{c_1, ..., c_k\}$ be a set of $k$ cones, each of which has a diameter bounded by $\gamma$, and $C$ forms a partition of $\mathbb{R}^2$. Note that $k$ is a function of $\gamma$. Thus, $C$ could be a set of cones defined by the rays originating at $O$ making angles $\{(i-1)2\pi/k : 1 \le i \le k\}$ with respect to the abscissa axis. The two rays that bound a cone $c$ are called the axes of $c$. For a point $p \in \mathbb{R}^2$, let $t_i(p)$ denote $p$ represented in the coordinate system whose axes are those of $c_i$. For a point $p$ in a cone $c_i$, $d_k(O,p) = \|t_i(p)\|$ is called the $k$-oriented distance from $O$ to $p$, and is defined as the length of the shortest path from $O$ to $p$ traveling only in the directions parallel to the axes of $c_i$. For any two points $p$ and $q$ in $c_i$, we have $d_k(p,q) = d_k(O, q-p)$. Notice that, if $\gamma = \pi/2$, then the corresponding $d_k$ is known as the rectilinear (Manhattan) distance function. For any two points $p, q \in \mathbb{R}^2$, $\|pq\| \le d_k(p,q) \le (1+\varepsilon)\|pq\|$, where $\varepsilon$ is a positive constant that decreases as $k$ increases.

We now derive an explicit expression for $k$ in terms of $\varepsilon$. Assume, without loss of generality, that point $p$ is located at the origin $O$ (i.e., $p = O$). Let $\rho_1$ and $\rho_2$ be the two rays originating at $O$ and defining the cone that contains point $q$. Recall that the cone has a diameter bounded by angle $\gamma$. Consider the case that $\gamma$ is less than $\pi/2$. Define $m$ to be the line with the same slope as ray $\rho_1$ and passing through $q$. Let $r$ be the intersection of $m$ and $\rho_2$. Note that $d_k(p,q) = \|pr\| + \|rq\|$. Furthermore, according to the law of cosines, we have

$$\begin{aligned} \|pr\|^2 + \|rq\|^2 - 2\|pr\|\|rq\|\cos(\pi-\gamma) &= \|pq\|^2 \\ \|pr\|^2 + \|rq\|^2 + 2\|pr\|\|rq\|\cos\gamma &= \|pq\|^2 \end{aligned}$$

Given that $0 < \gamma < \pi/2$,

$$\left(\|pr\|^2+\|rq\|^2+2\|pr\|\|rq\|\right)\cos\gamma \le \|pq\|^2$$
$$(\|pr\|+\|rq\|)^2 \le \frac{\|pq\|^2}{\cos\gamma}$$
$$\|pr\|+\|rq\| \le \frac{\|pq\|}{\sqrt{\cos\gamma}}$$

Thus, we have $d_k(p,q) \le (1+\varepsilon)\|pq\|$, where $\varepsilon = \frac{1}{\sqrt{\cos\gamma}} - 1$. Since $\gamma = 2\pi/k$, we obtain $k = \frac{2\pi}{\cos^{-1}(1+\varepsilon)^{-2}}$ for $0 < \varepsilon < 1$.

Recall that the objective function $f(s)$ denotes the sum of the Euclidean distances from $P$ to $s$. We can approximate $f(s)$ using

$$f_k(s) = \sum_{\substack{1\le i\le n \\ p_i \in S_a}} d_k(p_i, a) + \sum_{\substack{1\le i\le n \\ p_i \in S_b}} d_k(p_i, b) + \sum_{\substack{1\le i\le n \\ p_i \in S_{ab}^+}} y_i - y_a + \sum_{\substack{1\le i\le n \\ p_i \in S_{ab}^-}} y_a - y_i$$

Observe that function $f_k(s)$ is convex and piecewise linear. Hence, we can find the minimum of $f_k(s)$ using the prune-and-search approach described by Bose et al. [2] after some careful modifications.

**Prune and search.** Consider the set of cones $C$ used in evaluating $d_k$. Recall that each cone $c \in C$ is defined by two lines. Let $L$ be the set of lines defining $C$. For each point $p \in P$, we create a point at a distance $\ell$ to the right of $p$. Let $P'$ denote the newly created set of points. For each point $p \in P \cup P'$, we construct a copy of $L$ such that each of the lines in $L$ passes through $p$. The result is an arrangement of lines $A$. Observe that each cell of $A$ corresponds to a linear piece of the surface $f_k$. Consequently, $f_k$ reaches a minimum when the endpoint $a$ of line segment $s$ coincides with a vertex of $A$.

We now describe a prune-and-search algorithm to find the lowest point on the surface $f_k$. Note that $A$ consists of $k$ sets of parallel lines. Let $H_i$ denote a given set of parallel lines in $A$, where $1 \le i \le k$. We begin by finding a median line $h \in H_i$ that divides $H_i$ into two nearly equal sets. Line $h$ partitions $\mathbb{R}^2$ into two half-planes, $h_1$ and $h_2$, one of which contains a minimum of $f_k$. Suppose, without loss of generality, that $h_1$ contains the minimum. Then, we can simply ignore all the lines in $h_2$, and continue to recurse on $h_1$. This recursive process takes $O(\log n)$ rounds for each set $H_i$.

In each aforesaid round, we first find a point $p_h$ on $h$ that minimizes $f_k$. We can then, based on $p_h$, determine if the minimum lies in $h_1$ or $h_2$.

The problem of finding $p_h$ is a one-dimensional instance of our problem (i.e., constrained to line $h$). Since $f_k$ is piecewise linear, $p_h$ lies on an intersection of $h$ with some other line in $H = \{H_1, ..., H_k\} \setminus h$. Hence, we i) compute all the intersections of $h$ with $H$, ii) find the median intersection point $q_m$ and the two intersection points $q_1$ and $q_2$ that are adjacent to $q_m$ on $h$, and iii) determine if $p_h$ lies to the left of $q_m$, right of $q_m$, or is $q_m$ by evaluating $f_k(q_m)$, $f_k(q_1)$, and $f_k(q_2)$.

Let $u$ be the size of $H$. The time complexity of finding $p_h$ is given by the recurrence relation $T(u) = T(u/2) + O(u + Q(n))$, where $Q(n)$ denotes the query time to evaluate $f_k$. This recurrence solves to $O(u+Q(n)\log u)$.

After finding $p_h$, we determine whether the minimum lies in $h_1$ or $h_2$ as follows. Consider two opposite rays $r_1$ and $r_2$, which are i) originating at $p_h$, ii) orthogonal to $h$, and iii) contained in $h_1$ and $h_2$, respectively. We identify the first lines $h_{r_1}$ and $h_{r_2}$ intersected by $r_1$ and $r_2$, respectively. Let $v_1$ (resp. $v_2$) be the intersection point of $h_{r_1}$ and $r_1$ (resp. $h_{r_2}$ and $r_2$). There are three possible cases to be considered: (1) If $f_k(v_1) \le f_k(p_h) \le f_k(v_2)$, then a minimum of $f_k$ lies in $h_1$. (2) If $f_k(v_1) \ge f_k(p_h) \ge f_k(v_2)$, then a minimum of $f_k$ lies in $h_2$. (3) If $f_k(v_1) > f_k(p_h)$ and $f_k(v_2) > f_k(p_h)$, then $p_h$ is a minimum of $f_k$. Verifying these cases require the computation of all the intersections of $H$ with $r_1$ and $r_2$, and the evaluation of $f_k$ at $v_1$ and $v_2$. So, the time complexity of determining whether a minimum lies in $h_1$ or $h_2$ is $O(u + Q(n))$.

Observe that $u = O(kn)$. Thus, the time taken by the recursive procedure for each set $H_i$ is given by the recurrence relation $T(n) = T(kn/2) + O(kn + Q(n)\log kn)$, which solves to $O(kn + Q(n)\log kn)$. Given that we have $k$ sets $H_i$, the overall time taken by the prune-and-search algorithm to compute the point that minimizes $f_k$ is $O(P(n) + k(kn + Q(n)\log kn))$, where $P(n)$ is the preprocessing time to construct the data structure for evaluating $f_k$, and $Q(n)$ is the query time to evaluate $f_k$.

We claim that a data structure with a preprocessing time $P(n) = O(kn\log n)$ exists for evaluating $f_k$ in query time $Q(n) = O(k\log n)$ (refer to the full paper for details). As a result, the overall running time of our algorithm is $O(kn\log n)$. □

**Remark 3** *Alternatively, the space-subdivision procedure previously used in approximating a point-anchored median line segment could be extended to address the constant-slope variant. The resulting $(1+\varepsilon)$-approximation algorithm would have a time complexity of $O(n^2 + n\varepsilon^{-4}\alpha_{xy})$, where $\alpha_{xy}$ is a function dependent on the coordinates of $P$.*

## 7 Conclusion

We have proven that a median line segment is not constructible for $n \geq 3$ non-collinear points in the plane by using only ruler and compass. We have presented a $(1+\varepsilon)$-approximation algorithm for solving the constrained median line segment problem in $\mathbb{R}^2$ where an endpoint or the slope of the median line segment is given at input. These algorithms are space-subdivision and prune-and-search approaches, and their time complexities are near-linear in $n$. At last, we leave open the question of whether our approximation algorithms for solving the constrained variants can be extended to obtain a $(1+\varepsilon)$-approximate solution to the unconstrained median line segment problem.

## References

[1] C. Bajaj. The algebraic degree of geometric optimization problems. *Discrete & Computational Geometry*, 3(2):177–191, 1988.

[2] P. Bose, A. Maheshwari, and P. Morin. Fast approximations for sums of distances, clustering and the Fermat-Weber problem. *Computational Geometry*, 24(3):135–146, 2003.

[3] R. Chandrasekaran and A. Tamir. Algebraic optimization: the Fermat-Weber location problem. *Mathematical Programming*, 46(1):219–224, 1990.

[4] M. B. Cohen, Y. T. Lee, G. Miller, J. Pachocki, and A. Sidford. Geometric median in nearly linear time. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 9–21, 2016.

[5] J. M. Keil. Approximating the complete Euclidean graph. In *Scandinavian Workshop on Algorithm Theory*, pages 208–213, 1988.

[6] N. M. Korneenko and H. Martini. Approximating finite weighted point sets by hyperplanes. In *Scandinavian Workshop on Algorithm Theory*, pages 276–286, 1990.

[7] S. Mehlhos. Simple counter examples for the unsolvability of the Fermat- and Steiner-Weber-problem by compass and ruler. *Beiträge zur Algebra und Geometrie*, 41(1):151–158, 2000.

[8] J. G. Morris and J. P. Norback. A simple approach to linear facility location. *Transportation Science*, 14(1):1–8, 1980.

[9] P. M. Neumann. Reflections on reflection in a spherical mirror. *The American Mathematical Monthly*, 105(6):523–528, 1998.

[10] J. Ruppert and R. Seidel. Approximating the $d$-dimensional complete Euclidean graph. In *Proceedings of the 3rd Canadian Conference on Computational Geometry*, pages 207–210, 1991.

[11] R. Schieweck and A. Schöbel. Properties and algorithms for line location with extensions. In *Proceedings of the 28th European Workshop on Computational Geometry*, pages 185–188, 2012.

[12] P. Yamamoto, K. Kato, K. Imai, and H. Imai. Algorithms for vertical and orthogonal $L_1$ linear approximation of points. In *Proceedings of the 4th Annual Symposium on Computational Geometry*, pages 352–361, 1988.