# Computing a Maximal Depth Point in the Plane

Stefan Langerman* William Steiger†

## Abstract

Given a set $S = \{P_1, \ldots, P_n\}$ of $n$ points in $R^d$, the *depth* of a point $Q \in R^d$ is the minimum number of points of $S$ that must be in a closed halfspace containing $Q$. A median is a point of maximal depth. For dimension $d = 2$ we give an $O(n(\log n)^4)$ deterministic algorithm to compute a median $\mu$ and its depth $\delta^*$, and we give an $\Omega(n \log n)$ lower bound for the height of an algebraic decision tree that can find the depth of a median.

## 1 Introduction and Summary

Given a set $S = \{P_1, \ldots, P_n\}$ of $n$ real numbers, the depth of $x \in R$ is defined to be

$$d(x) = \min(|\{P_i \leq x\}|, |\{P_i \geq x\}|), \tag{1}$$

and may be thought of as the "size" of the smallest halfspace containing $x$. A median is a point of maximal depth ($\lceil (n+1)/2 \rceil$) and may be found in $O(n)$ time.

There are many situations - for example in Statistics - where multivariate generalizations of ranks and order statistics are useful, and several proposals for depth in $R^d$ have been made [1], [7],[10],[11]. One of the more familiar ones was proposed by John Tukey [12], a natural extension to $d > 1$ of the notion in (1). Given a set $S = \{P_1, \ldots, P_n\}$ of $n$ points in $R^d$, the Tukey depth, or simply depth of $x \in R^d$ is defined by

$$\delta(x) = \min_{u:\|u\|=1} (d_u(x)) \tag{2}$$

where $d_u(x)$ is the directional depth of $x$ in direction $u$ and is defined as the depth of the $P_i$ projected orthogonally onto the line $\{x + tu, t \in R\}$. As $d_u(x)$ counts the size of the smaller halfspace of the hyperplane containing $x$ and having normal vector $u$, $\delta$ measures the smallest number of points in any halfspace containing $x$. A median is a point $\mu$ of maximal depth, and we write $\delta^* = \delta(S)$ for the Tukey depth of a median.

For integer $k > 0$ let $C_k(S) \subset R^d$ denote the set of points of depth at least $k$. Then

$$\delta^* = \max (k : C_k(S) \neq \phi). \tag{3}$$

It follows from (2) that $C_k(S)$ is convex. A well known consequence of Helly's Theorem (e.g.,[6]) is that there is a point $x \in R^d$ (not necessarily in $S$) of Tukey depth at least $\lceil n/(d+1) \rceil$. Such a point is called a *centerpoint*. The *center* $C(S)$ is the set of all centerpoints.

The main question addressed in this paper concerns the computational complexity of selection; that is, finding points of a given depth $k$, or finding points of maximal depth $\delta^*$, along with the value of $\delta(S)$. Clearly

$$\lceil n/(d+1) \rceil \leq \delta(S) \leq \lceil (n+d)/2 \rceil.$$

We use the unit cost RAM model of computation. A centerpoint in $R^d$ can be found by solving a set of $\Theta(n^d)$ linear inequalities, using linear programming [2]. When $d = 2$ the situation improves. Cole, Sharir and Yap described an $O(n(\log n)^5)$ algorithm [5] to construct a centerpoint, and subsequent ideas of Cole [3] could be used to lower the complexity to $O(n(\log n)^3)$. Recently, Jadhav and Mukhopadhyay [8] described a linear time algorithm to find a Tukey centerpoint.

Matoušek [9] attacked the harder problem of computing a Tukey median. He first described an $O(n(\log n)^4)$ algorithm that gives a description of $C_k(S)$, $k \geq \lceil n/3 \rceil$. Then, using this in a binary search on $k \in [\lceil n/(d+1) \rceil, \lceil (n+d)/2 \rceil]$, the largest $k$ for which $C_k(S) \neq \phi$ is found in $O(n \log n)^5)$.

The present paper gives two results that narrow down the range for the complexity of finding a median of $n$ points in the plane.

**Theorem 1** *Given $S = \{P_1, \ldots, P_n\} \subset R^2$ a point $\mu$ of depth $\delta^*$ can be found in $O(n(\log n)^4)$ time while an algebraic decision tree to compute $\delta(S)$ must have height $\Omega(n \log n)$.*

The upper bound is given via an algorithm with the stated complexity. In fact within the same time bounds, the algorithm can actually construct a description of $C_k(S)$ for any $k \leq \delta^*$. The proof introduces several new algorithmic tools for arangements, and is of independent interest. The lower bound construction is sketched in Section 3.

## 2 The Upper Bound

Given the set $S$, and $k > 0$, the algorithm decides in time $O(n(\log n)^3)$ whether $k \leq \delta^* = \delta(S)$. If the answer is YES it returns a witness point of depth at least $k$. Using this algorithm in a binary search on $k$, we get the value of $\delta^*$ in $O(n(\log n)^4)$ time. Within the same time bound, the actual description of $C_{\delta^*}$ can then be computed using Matoušek's algorithm.

Of independent interest, the proof has algorithms which, given a set $L$ of $n$ lines forming the arrangement $A(L)$, can

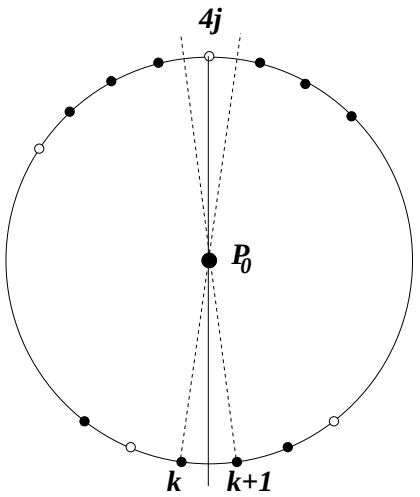
- count the number of vertices of $A(L)$ interior to a given polygon $\mathcal{P}$ with $s$ sides, and which meets $k$ lines of $L$, in time $O(s + k \log k)$
- compute a ham-sandwich cut for the vertices of $A(L)$ that are on opposite sides of a given line $\ell$, in time $O(n(\log n)^2)$

*Computer Science, Rutgers University
†Computer Science, Rutgers University

- find lines $L_1, L_2$ forming quadrants, each "missing" $n/4$ of the lines (an equitable partition of lines) in time $O(n)$

Details appear in the full paper.

# 3 The Lower Bound




Let $C$ denote the circumference of the unit circle in $R^2$ and let $T$ denote an algebraic decision tree that can decide whether $\delta(S) \geq \lceil (m+1)/2 \rceil$ for a set $S$ of $m$ points in $R^2$. We restrict attention to sets $S$ with $P_0 = (0,0)$ and $P_i = (\cos z_i, \sin z_i) \in C,\ i = 1, \ldots, 4n+3,\ z_i \in [0, 2\pi]$. These restricted inputs can be encoded by points $(z_1, \ldots, z_{m-1}) \in [0, 2\pi]^{m-1}$. We argue that the subset $Y \subseteq [0, 2\pi]^m$ of restricted inputs where $T$ returns YES has $\Omega(n!)$ path connected components.

Let

$$\theta_i = \left(\frac{i}{4n+3}\right) 2\pi,\ i = 1, \ldots, 4n+3. \tag{4}$$

This encodes a set $S$ with $m = 4n+4$ points, $4n+3$ of them equally spaced on $C$, and no two of them diametric in $C$. The diameter containing $P_i$, $i > 0$, is a halving line and both its closed halfspaces contain $2n+3$ points of $S$. A diameter containing no $P_i$, $i > 0$, is a non-halving diameter and has one closed halfspace with $2n+2$ points and the other with $2n+3$ points, so $\delta(P_0)$ is $2n+2$. This is maximal because for any $x \neq P_0$ the line through $x$ perpendicular to $\overline{xP_0}$ has a closed halfspace with less than $2n+2$ points of $S$.

For inputs to $T$, we take $P_0$ and $4+3$ other points encoded via the set $I$, where $z = (z_1, \ldots, z_{m-1}) \in I$ if all $z_i \in [0, 2\pi]$ and $z_j = \theta_j,\ j \equiv 1, 2, 3 \bmod 4\}$. Let $\pi$ and $\rho$ be two different permutations of the integers $4, 8, \ldots, 4n$, and define inputs $z_\pi \in I$ by

$$(\theta_1, \theta_2, \theta_3, \theta_{\pi_1}, \ldots, \theta_{4n-1}, \theta_{\pi_n}, \theta_{4n+1}, \theta_{4n+2}, \theta_{4n+3})$$

and input $z_\rho \in I$ by

$$(\theta_1, \theta_2, \theta_3, \theta_{\rho_1}, \ldots, \theta_{4n-1}, \theta_{\rho_n}, \theta_{4n+1}, \theta_{4n+2}, \theta_{4n+3}).$$

Both $z_\pi$ and $z_\rho$ are in $Y \cap I$, as they describe the $m$ points in $S$, but in different order. They are in different components because on a continuous path from $z_\pi$ to $z_\rho$ in $I$, some $z_{4j}$ must "cross" (i.e., reverse its radial ordering) a neighbor, $\theta_{4j-1}$ or $\theta_{4j+1}$ (see figure). Write $k = 4j + 2n + 1 \bmod (m+1)$; $\theta_k$ and $\theta_{k+1}$ are the neighbors of the diameter through $\theta_{4j}$. If $\theta_{4j}$ crosses a neighbor, it first will cross the diameter through $\theta_k$ or the diameter through $\theta_{k+1}$. After this occurs, that diameter will have closed halfspaces with $2n+2$ points on one side and $2n+4$ on the other, and the depth of $P_0$ described by this input $z \in I$ will be $\delta(S) = 2n+1$, so $z \notin Y$. Therefore $Y \cap I$ has at least $n!$ different components. As $m = 4n+3$, the height of $T$ is $\Omega(m \log m)$.

The above argument also establishes an $\Omega(m \log m)$ lower bound for computing the Tukey depth of a point $x$. It would be interesting to know if the same lower bound applies to the seemingly easier questions of reporting a point $x$, of maximal Tukey depth.

# References


[1] B. Chazelle. Geometric Order statistics. *Proc ACM Geometry*, 1988.

[2] K. L. Clarkson and D. Eppstein and G. L. Miller and C. Sturtivant and S.-H. Teng. "Approximating center points with iterative Radon points" *Internat. J. Comput. Geom. Appl. 6*, 357-377 (1996)

[3] R. Cole. Slowing Down Sorting Networks to Obtain Faster Sorting Algorithms. *J. ACM 34(1)*, 200-208, (1987).

[4] R. Cole, J. Salowe, W. Steiger, and E. Szemerédi. An Optimal Time Algorithm for Slope Selection, *SIAM J. Comp. 18*, 792-810 (1989).

[5] R. Cole, M. Sharir, and K. Yap. On $k$-Hulls and Related Problems. *Siam J. Comput. 16(1)*,61-77, (1987).

[6] H. Edelsbrunner. *Algorithms in Combinatorial Geometry.* Springer-Verlag, Berlin, 1987.

[7] J. Gill, W. Steiger, and A. Wigderson. Geometric Medians, *Discrete Math. 108*, 37-51. (1992),

[8] S. Jadhav and A. Mukhopadhyay. Computing a Centerpoint of a Finite Planar Set of Points in Linear Time. *Discrete and Comp. Geom. 12*, 291-312. (1994).

[9] J. Matoušek. Computing the Center of a Planar Point Set. *Discrete and Computational Geometry: Papers from the DIMACS Special Year* Amer. Math. Soc., J.E. Goodman, R. Pollack, W. Steiger, Eds, (1992), 221-230.

[10] P. Rousseeuw and M. Hubert. Depth in an Arangement of Hyperplanes. *Discrete and Comp. Geom.* (1999)

[11] P. Rousseeuw and M. Hubert. Regression Depth. *J. Amer. Statist. Assoc. 94*, 388-402 (1999).

[12] John Tukey. "Mathematics and the Picturing of Data". *International Conf. of Mathematicians, Vancouver*, 1971.