

The Centervortex Theorem for Wedge Depth*

Gary L. Miller[†]

Todd Phillips[‡]

Donald R. Sheehy[§]

July 3, 2009

Abstract

There are many depth measures on point sets that yield centerpoint theorems. These theorems guarantee the existence of points of a specified depth, a kind of geometric median. However, the deep point guaranteed to exist is not guaranteed to be among the input, and often, it is not. The α -wedge depth of a point with respect to a point set is a natural generalization of halfspace depth that replaces halfspaces with wedges (cones or cocones) of angle α . We introduce the notion of a *centervortex*, a point with depth at least $\frac{n}{d+1}$ among the set S . We prove that for any finite set $S \subset \mathbb{R}^d$, a centervortex exists. We also present a simple algorithm for computing an approximate centervortex.

1 Introduction

Many different notions of data depth have been proposed as ways to generalize the rank and the median of an ordered list to the case of higher dimensional point sets. Several nice surveys are available on different depth measures and how to compute them [12, 2, 11]. Aloupis et al give some lower bounds for computing depth [3]. One of the most enduring definitions of data depth is the halfspace depth, also known as the Tukey depth or location depth. The halfspace depth of a point p relative to a point set S is defined as the minimum number of points halfspace containing p .

Definition 1.1 *Given a set $S \subset \mathbb{R}^d$, the **halfspace depth** of a point $x \in \mathbb{R}^d$ is*

$$D_\pi(x) = \min \{|H \cap S| : H \text{ is a closed halfspace, and } x \in H\}$$

An α -wedge is defined by an apex t and an axis r and is the set of all points $p \in \mathbb{R}^d$ such that the line segment \overline{pt} makes an angle of at most $\alpha/2$ with r . The *alpha-wedge depth* is a natural generalization of halfspace depth in which the halfspace is interpreted as a π -wedge. It was first introduced by Erickson et al.[10]

Definition 1.2 *Given a set $S \subset \mathbb{R}^d$, the **α -wedge depth** of a point $x \in \mathbb{R}^d$ is*

$$D_\alpha(x) = \min \{|W \cap S| : W \text{ is an } \alpha\text{-wedge with apex } x\}$$

The definitions are equivalent when $\alpha = \pi$, so the given notation is not ambiguous. In this extended abstract, we focus on the interesting properties of the $D_{3\pi/2}$.

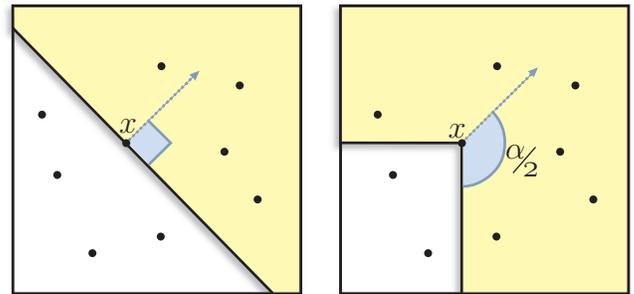


Figure 1: The α -wedge depth counts points in wedges rather than halfspaces.

The Centerpoint Theorem states that there exists a point in \mathbb{R}^d with halfspace depth at least $\frac{n}{d+1}$ [14, 9]. Because no point can have halfspace depth greater than $\frac{n}{2}$, a centerpoint is a constant-factor approximate median.

We distinguish between input *vertices* and *points* in space which may not be in the input. One difficulty of traditional measures of data depth is that the median (or even an approximate median) is often not among the input vertices. We show that for $\alpha > 3\pi/2$, one to always find a point with α -wedge depth at least $\left\lceil \frac{n}{d+1} \right\rceil$ among the input vertices (Section 3). We call such a point, a *centervortex*. This is a major advantage because for many inputs, in particular sets in convex position, the halfspace depth may be $O(1)$ for all $s \in S$.

The strength of the α -wedge depth to always identify vertices of nontrivial depth among the input is also interesting from the perspective of algorithm design. Many geometric divide and conquer algorithms can benefit from pivoting on a centerpoint. However, if the circumstances demands that the pivot be a vertex in the original set, a centervortex may be a better option.

*This work was supported in part by the National Science Foundation under grants CCF-0635257

[†]Department of Computer Science, Carnegie Mellon University, glmiller@cs.cmu.edu

[‡]Department of Computer Science, Carnegie Mellon University, tp517@cs.cmu.edu

[§]Department of Computer Science, Carnegie Mellon University, dsheehy@cs.cmu.edu

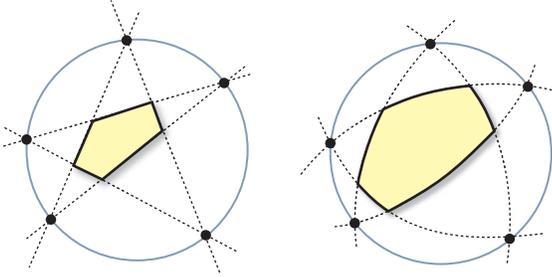


Figure 2: On the left, the center region for $\alpha = \pi$ is the set of all centerpoints. On the right, the center region for $\alpha > \pi$ replaces lines with circular arcs, expanding the set of centerpoints.

In Section 4 we extend our analysis of α -wedge depth at the input vertices. We also address the question of when it is possible to use values of α less than $3\pi/2$ and still guarantee a linear depth vertex. Moreover, we show that the expected depth at the vertices is linear. Section 5 addresses the computation of center vertices.

2 Related Work

The notion of α -wedge depth was introduced by Erickson et al., who proved several guarantees on the min-max depth point as α is allowed to vary [10]. The level sets of the wedge depth in the plane were studied by Avis et al. as way to describe the “shape” of a point set [5]. Abellenas et al. studied subsets of S called (α, k) -sets. These are the sets that can be separated by an α -wedge from the rest of S [1].

Of the many definitions of data depth used in practice, a couple arise quite naturally out of discrete convex geometry: the halfspace depth, also known as Tukey depth or location depth; the regression depth; and the simplicial depth. As previously noted, the halfspace depth of a point x is the minimum number of vertices of S contained in any closed halfspace containing x [18]. The regression depth of a hyperplane P relative to S is the minimum number of vertices that P passes through in a continuous rotation to vertical [15]. The simplicial depth of a point x is the number of simplices with vertices in S that contain x .

In each case, non-trivial lower bounds are known for the deepest point, or median. For the Tukey depth, The Centerpoint Theorem guarantees that the median has depth at least $\frac{n}{d+1}$ [9]. Bárány showed a non-trivial lower bound for the simplicial depth median [6] (see also [17] and [16]). Wagner shows that in fact a centerpoint of S is bounded by a constant fraction of all simplices with vertices in S [19]. The use of a centerpoint as a good candidate for other depth measures is a recurring theme in the literature. Amenta et al. showed that among any hyperplane arrangement, there is always one

with regression depth at least $\frac{n}{d+1}$ [4]. Their proof relies on a clever use of topological fixed point theorems and centerpoints. In this paper, we also use the Centerpoint Theorem as a starting point for finding points of high α -wedge depth.

3 The Centervertex Theorem

Any notion of data depth naturally gives rise to a median, or a point of maximum depth. The classic Centerpoint Theorem of Danzer et al. establishes a lower bound on the depth of the median for halfspace depth [9]. For any set $S \subset \mathbb{R}^d$ and $\alpha < \pi$, the D_α -median is at least the halfspace depth so the Centerpoint Theorem implies an immediate lower bound on the D_α -median. However, we are interested in picking a median from among the vertex set S . For the D_π -median and many other depth measures, the median may not (and most likely is not) among S . In fact, for the simple case of vertices in convex position, $D_\pi(s) = 1$ for every vertex $s \in S$.

The main result of this section is that among any set of vertices, there exists a vertex s such that $D_{3\pi/2}(s) \geq \frac{n}{d+1}$. We call such a point a *centervertex*. The proof of the Centervertex Theorem (Theorem 2 below) depends on the following Lemma relating $D_{3\pi/2}$ at vertices and D_π at arbitrary points.

Lemma 1 *Given a set $S \subset \mathbb{R}^d$, a point $x \in \mathbb{R}^d$ and a vertex $s \in S$, if s is the k th nearest vertex to x then $D_{3\pi/2}(s) \geq D_\pi(x) - k + 1$.*

Proof. It suffices to show that for any $3\pi/2$ -wedge W with apex at s , $|W \cap S| \geq D_\pi(x) - k + 1$.

First consider the case where $x \in W$. Then there exists a closed halfspace H containing x that is contained entirely in W . So, $H \cap S \subseteq W \cap S$ and thus

$$|W \cap S| \geq |H \cap S| \geq D_\pi(x) \geq D_\pi(x) - k + 1.$$

We now consider the case where $x \notin W$. Then x is in the cone $W' = \mathbb{R}^d \setminus W$. Wedge depth is invariant under dilation and rigid transformation so we may assume without loss of generality that s is at the origin, the wedge axis is $(0, \dots, 0, -1)$, and $|x| = 1$. We choose a vector $v_x = (x_1, \dots, x_{d-1}, -x_d)$ to define a hyperplane H through the point x that partition the space into (H^+, H^-) the halfspaces above and below H respectively (H^+ is closed and H^- is open). There are at most $n - D_\pi(x)$ vertices of S in H^+ .

Because $W' \subset H^+ \cup (H^- \setminus W)$, it will suffice to show that there are at most $k - 1$ vertices of S in $H^- \setminus W$.

Let p be any point in $S \cap (H^- \setminus W)$. Since $p \in H^-$, we can write it as $p = rx + tq$ where $r \in [0, 1]$, $q \cdot v_x = 0$, and $|q| = 1$. Since $p \in W'$, we know that $2p_d^2 > |p|^2$. Substituting $p = rx + tq$ in this inequality yields

$$2(r^2 x_d^2 + 2rt x_d q_d + t^2 q_d^2) \geq r^2 + 2rt(x \cdot q) + t^2.$$

Since $q \cdot v_x = 0$, it follows that $q \cdot x = 2x_d q_d$. Substituting this in the above inequality, we get

$$2(r^2 x_d^2 + t^2 q_d^2) \geq r^2 + t^2.$$

We can rearrange this to see that $t^2 \leq r^2 \frac{2x_d^2 - 1}{(1 - 2q_d^2)}$. If θ is the angle that x makes with the cone axis then $x_d = \cos \theta$ and $q_d \leq \sin \theta$. Therefore, $x_d^2 + q_d^2 \leq 1$ and $\frac{2x_d^2 - 1}{1 - 2q_d^2} \leq 1$. We can conclude that $t \leq r$, and thus,

$$|x - p| \leq |x - xr| + t < (1 - r) + r < 1 = |x - s|.$$

This means that every such p is a nearer neighbor to x than s . Because s is the k th nearest neighbor to x , there can be at most $k - 1$ such points $p \in S \cap (H^- \setminus W)$. \square

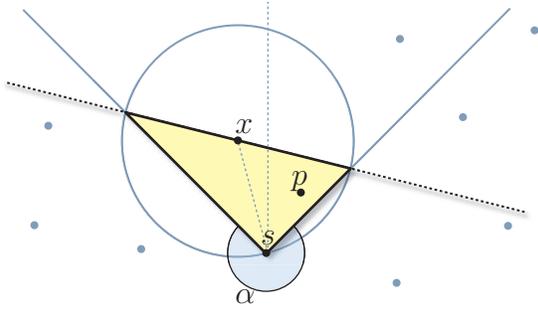


Figure 3: The triangle is $H^- \cap W'$ and is completely contained in the ball centered at x .

Theorem 2 (The Centervertex Theorem) *For all n point sets $S \subset \mathbb{R}^d$, there exists a vertex $s \in S$ with $D_{3\pi/2} \geq \frac{n}{d+1}$.*

Proof. Let c be a centerpoint of S . By the Centerpoint Theorem, $D_\pi(c) \geq \frac{n}{d+1}$. Let s be the nearest vertex in S to c . The result follows directly from Lemma 1 applied to the point c and vertex s . \square

4 Bounding the wedge depth at vertices

We are interested in the cone depth at vertices in S . The following corollary to Lemma 1 gives a simple way to bound the cone depth at the vertices.

Lemma 3 *Given $S \subset \mathbb{R}^d$ and $s \in S$,*

$$D_{3\pi/2}(s) \geq \max_{x \in \text{Vor}(s)} D_\pi(x),$$

where $\text{Vor}(s)$ is the Voronoi cell of s , i.e the set of all points in \mathbb{R}^d whose nearest vertex in S is s .

Proof. The proof is immediate from Lemma 1. \square

If we assume that the dimension d of the point set is a constant, Theorem 2 says that for some $s \in S$, $D_{3\pi/2}(s)$ is $\Theta(n)$. In fact, the average depth of the vertices in S is also linear as shown in the following Theorem.

Theorem 4 *Given $S \subset \mathbb{R}^d$, if $s \in S$ is sampled uniformly at random $\mathbb{E}[D_{3\pi/2}(s)] \geq \frac{n}{2(d+1)^2}$.*

Proof. Let c be a centerpoint of S . There are $\lceil \frac{n}{d+1} \rceil$ vertices $s_1, \dots, s_{\lceil \frac{n}{d+1} \rceil} \in S$ such that $D_{3\pi/2}(s_i) \geq i$, namely the $\lceil \frac{n}{d+1} \rceil$ closest vertices to c . This follows from Lemma 1. Thus we can bound the average $3\pi/2$ -wedge depth as follows.

$$\frac{1}{n} \sum_{t \in S} D_{3\pi/2}(t) \geq \frac{1}{n} \sum_{i=1}^{\lceil \frac{n}{d+1} \rceil} D_{3\pi/2}(s_i) \quad (1)$$

$$\geq \frac{1}{n} \sum_{i=1}^{\lceil \frac{n}{d+1} \rceil} i \quad (2)$$

$$\geq \frac{n}{2(d+1)^2}. \quad (3)$$

\square

The linear expected depth of the points in S may have ramifications for randomized algorithms, as it implies that a randomly chosen point can be used to generate a roughly balanced geometric partition of S . Theorem 4 also leads to a randomized algorithm for computing approximate centervertex in sub-linear time (see Section 5).

5 Computing Center Vertices

The proof of the Centervertex Theorem implies an algorithm for computing a centervertex. First compute a centerpoint c and then return the nearest vertex to c . The only difficulty with such an algorithm is that it is not known how to compute a centerpoint in time polynomial in n and d . The best known method is due to Chan and will produce a Tukey median in time $O(n^{d-1})$ randomized time [7]. Thus, we can easily find centerpoints in $O(n^d)$ time if we use the Chan algorithm as a black box.

A more time efficient approach is to use an approximate centerpoint. The Iterated-Tverberg algorithm of Miller and Sheehy can compute a point x guaranteed to have $D_\pi(x) \geq \frac{n}{2(d+1)^2}$ in $O\left(\frac{n^{1+\lg(d+2)}}{(d+1)^{2\lg(d+2)-1}}\right)$ time [13]. Their algorithm is a derandomization of an algorithm of Clarkson et al. that computes a point of similar depth with high probability in time $O(n(\lg n)^{\lg d})$ [8]. The randomized algorithm works by sampling the input and can be made to run in sub-linear time.

To extend the algorithm for an approximate centerpoint to an algorithm for an approximate centervertex, we sample $O(d^2)$ points and return the nearest point in the sample to the approximate centerpoint c . With high probability, the point returned is among the $O(n/d^2)$ nearest neighbors of c . Thus Lemma 1 implies that the depth is linear. Alternatively, we can search for the exact nearest neighbor of the approximate centerpoint in linear time.

6 Conclusion

We have explored the problem of finding points of linear depth among the input set S . Although most traditional depth measures do not make any guarantees about the depth of the points in S , we have shown that the $3\pi/2$ -wedge depth guarantees that some $s \in S$ will be a so-called centervertex. Moreover, we have shown how to bound the α -wedge depth in terms of the halfspace depth of nearby points and used this to bound $D_{3\pi/2}$ at input points. This led to some straightforward algorithms for computing deep vertices, including a randomized approximate centervertex algorithm that can be made to run in sublinear time.

References

- [1] M. Abellenas, M. Claverol, F. Hurtado, and C. Seara. (α, k) -sets en el plano. In *Proc. III Jornadas de Matemática Discreta y Algorítmica*, 2002.
- [2] G. Aloupis. On computing geometric estimators of location. Master's thesis, School of Computer Science, McGill University, 2001.
- [3] G. Aloupis, C. Cortes, F. Gomez, and G. Toussaint. Lower bounds for computing statistical depth. *Computational Statistics and Data Analysis*, 40:223–229, 2002.
- [4] N. Amenta, M. Bern, D. Eppstein, and S. Teng. Regression depth and center points. *Discrete and Computational Geometry*, 23:305–323, 2000.
- [5] D. Avis, B. Beresford-Smith, L. Devroye, H. Elgindy, E. Guévremont, F. Hurtado, and B. Zhu. Unoriented Θ -maxima in the plane: Complexity and algorithms. *SIAM J. Comput.*, 28(1):278–296, 1998.
- [6] I. Bárány. A generalization of Charathéodory's theorem. *Discrete Mathematics*, 40:141–152, 1982.
- [7] T. Chan. An optimal randomized algorithm for maximum Tukey depth. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, 2004.
- [8] K. Clarkson, D. Eppstein, G. Miller, C. Sturivant, and S.-H. Teng. Approximating center points with iterated Radon points. *International Journal of Computational Geometry and Applications*, 6(3):357–377, Sep 1996. invited submission.
- [9] L. Danzer, B. Grünbaum, and V. Klee. Helly's theorem and its relatives. In *Proc. Symp. Pure Math*, volume VII, 1963.
- [10] J. Erickson, F. Hurtado, and P. Morin. Centerpoint theorems for wedges. *Discrete Mathematics and Theoretical Computer Science*, 11(1):45–54, 2009.
- [11] K. Fukuda and V. Rosta. Data depth and maximum feasible subsystems. In Avis, Hertz, and Marcotte, editors, *Graph Theory and Combinatorial Optimization, Springer, 2005*. 2005.
- [12] J. Gil, W. L. Steiger, and A. Wigderson. Geometric medians. *Discrete Mathematics*, 108(1-3):37–51, 1992.
- [13] G. L. Miller and D. R. Sheehy. Approximate center points with proofs. In *Symposium on Computational Geometry*, 2009. manuscript.
- [14] R. Rado. A theorem on general measure. *J. Lond. Math. Soc.*, 21:291–300, 1947.
- [15] P. Rousseeuw and M. Hubert. Regression depth. *Journal of American Statistical Association*, 94:388–402, 1999.
- [16] D. Souvaine, M. Burr, and E. Rafalin. Simplicial depth: An improved definition, analysis, and efficiency for the finite sample case. In R. Liu, R. Serfling, and D. Souvaine, editors, *Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications*. AMS DIMACS, 2006.
- [17] D. Souvaine and E. Rafalin. Computational geometry and statistical depth measures. In M. Hubert, G. Pison, A. Struyf, and S. V. Aelst, editors, *Theory and Applications of Recent Robust Methods*, Statistics for Industry and Technology, pages 283–296. AMS DIMACS, 2004.
- [18] J. W. Tukey. Mathematics and the picturing of data. In R. D. James, editor, *Proceedings of the International Congress of Mathematicians*, pages 523–531, 1975.
- [19] U. Wagner. *On k -sets and applications*. PhD thesis, ETH Zurich, 2003.