Cone Depth and the Center Vertex Theorem

Gary L. Miller   Todd Phillips   Don Sheehy

October 11, 2008

Abstract

We generalize the Tukey depth to use cones instead of halfspaces. We prove a generalization of the center point theorem that for $S \subset \mathbb{R}^d$, there is a point $s \in S$, with depth at least $\frac{n}{d+1}$ for cones of half-angle $45°$. This gives a notion of data depth for which an approximate median can always be found among the original set.

1 Cone Depth

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 [GSW92, Alo01, FR05]. One of the most enduring definitions of data depth is the Tukey depth, also known as the half-space depth.

The Tukey depth of a point p relative to a point set S is defined as the minimum number of points on one side of *any* hyperplane through p. The Center Point Theorem states that there exists a point in $\mathbb{R}^d$ with Tukey depth at least $\frac{n}{d+1}$. Because no point can have Tukey depth greater than $\frac{n}{2}$, a center point is a constant-factor approximate median.

One difficulty of traditional measures of data depth is that the median (or even an approximate median) is often not among the points in the set. Popular depth measures such as the Tukey depth and simplicial depth have the property that if the set S is in convex position then the depth of every $s \in S$ is 0. Thus, they may not give much information about the relative depth of the original set.

We introduce the *cone depth*, a natural generalization of Tukey depth in which the halfspace of points is treated as a cone of half-angle $90°$. Let v be a unit vector and define X_v to be the set of points $x \in S$ such that $\frac{x}{|x|} \cdot v \geq 0$. The Tukey depth (of the origin) can be defined formally as $\min_{v \in \mathbb{R}^d} |X_v|$. To generalize this, we replace the 0 in the definition of X_v with a constant $c \in (-1, 1)$, to get $X_{v,c} = \{x \in S | \frac{x}{|x|} \cdot v \geq c\}$. The *$\theta$-cone depth* for a cone with half-angle $\theta/2$ is $\min_{v \in \mathbb{R}^d} |X_{\frac{v}{|v|}, -\cos\frac{\theta}{2}}|$. Equivalently, the cone depth of p is n minus the maximum number of points that may be contained in the interior of a cone with apex at p. In particular, Tukey depth is equivalent to $180°$-cone depth. In this note, we focus on the interesting properties of the $90°$-cone depth.

2 The Main Result

The main result is that among any set of points, there is always one that has linear $90°$-cone depth. We call such a point a *center vertex*.

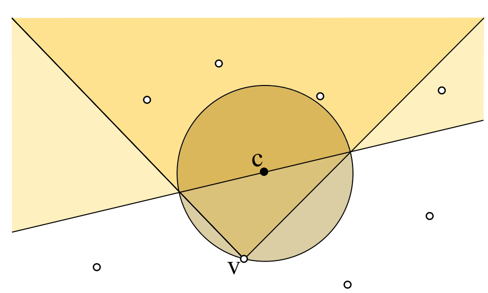

Figure 1: The point c is a center point and v is a center vertex.

Theorem 2.1 *For all n point sets $S \subset \mathbb{R}^d$, there exists a vertex $v \in S$ with $90°$-cone depth $\frac{n}{d+1}$.*

Proof: Let c be a center point for the set X, i.e. c has Tukey depth $\frac{n}{d+1}$. Let v be the nearest point in S to c. Let C be any 90° cone with apex at v. If $c \notin C$ then there exists a plane P through c such that the cone C lies entirely on one side of P. In this case, the theorem follows directly from the center point theorem.

We now consider the case where $c \in C$. We choose a vector $v_c = (c_1, \ldots, c_{d-1}, -c_d)$ to define a plane P through the point c. Because c is a center point, there are at most $\frac{dn}{d+1}$ points above P. To prove the theorem, it will suffice to show that there are no points both in the cone C and below P.

Cone depth is invariant under dilation and rigid transformations so we may assume without loss of generality that v is at the origin, the cone axis is $(0, \ldots, 0, 1)$, and $|c| = 1$. Suppose for contradiction that there exists a point $p \in S$ inside the cone and below P. Since p is below P, we can write it as $p = kc + tq$ where $k \in [0, 1]$, $q \cdot v_c = 0$, and $|q| = 1$. Since $p \in C$, we know that $2p_d^2 > |p|^2$. Substituting $p = kc + tq$ in this inequality yields the following.

$$2(k^2c_d^2 + 2ktc_dq_d + t^2q_d^2) \geq k^2 + 2kt(c \cdot q) + t^2$$

Since $q \cdot v_c = 0$, it follows that $q \cdot c = 2c_dq_d$. Substituting this in the above inequality, we get

$$2(k^2c_d^2 + t^2q_d^2) \geq k^2 + t^2.$$

We can rearrange this to see that $t^2 \leq k^2 \frac{2c_d^2 - 1}{(1-2q_d^2)}$. If θ is the angle that c makes with the cone axis then $c_d = \cos\theta$ and $q_d \leq \sin\theta$. Therefore, $c_d^2 + q_d^2 \leq 1$ and $\frac{2c_d^2-1}{1-2q_d^2} \leq 1$. We can conclude that $t \leq k$. Thus, $|c - p| \leq |c - ck| + t < (1 - k) + k < 1$. This is a contradiction because we assumed that v was the nearest point in S to c and $|c - v| = 1$. ■

3 Expected Depth

Theorem 2.1 says that among S, there is a point with linear 90°-cone depth. The proof of this theorem indicates a way to lower bound the expected depth of any point in S. In the proof, we showed that nearest $s \in S$ to a center point c has depth $\frac{n}{d+1}$. The same arguments can be used to show that the k-th nearest $s \in S$ to c has depth at least $\frac{n}{d+1} - (k-1)$. It follows that at least least $\frac{n}{2(d+1)}$ points in S have depth at least $\frac{n}{2(d+1)}$. Thus, the average depth of a point in S is at least $\frac{n}{4(d+1)^2}$.

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. Moreover, one can find a point with linear depth with high probability in sublinear time. This is accomplished by first finding an approximate center point using methods from [CEM+93]. This will achieve a point c that has Tukey depth $\Omega(\frac{n}{d^2})$ with high probability. Sampling a constant number of points and returning the closest to c will have linear cone depth with high probability.

4 Some Open Problems

We conclude with several interesting problems related to cone depth that remain open. Is 90° the largest value for which the existence of a center vertex is guaranteed? Given a point $p \in \mathbb{R}^d$, how fast can we compute the cone depth of p? Given a set of points $S \subset \mathbb{R}^d$, how fast can we find a center vertex deterministically? How fast can we find a point of maximal cone depth?

References

[Alo01] Greg Aloupis. On computing geometric estimators of location /. Master's thesis, 2001.

[CEM+93] K. L. Clarkson, D. Eppstein, G. L. Miller, C. Sturtivant, and S. Teng. Approximating center points with iterated radon points. In *Proceedings of the 9th acm Symposium on Computational Geometry*, San Diego, CA, 1993.

[FR05] Komei Fukuda and Vera Rosta. Data depth and maximum feasible subsystems. In Avis, Hertz, and Marcotte, editors, *Graph Theory and Combinatorial Optimization, Springer, 2005*. 2005.

[GSW92] Joseph Gil, William L. Steiger, and Avi Wigderson. Geometric medians. *Discrete Mathematics*, 108(1-3):37–51, 1992.