Counting the Onion

Ketan Dalal
School of Computer Science
McGill University, Montreal, Canada
kdalal@cs.mcgill.ca

May 3, 2004

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Masters of Science.

Your file Votre référence
ISBN: 0-612-98616-0
Our file Notre référence
ISBN: 0-612-98616-0

Abstract

Iteratively computing and discarding a set of convex hulls creates a structure known as an "onion". In this thesis, we show that the expected number of layers of a convex hull onion for n uniformly and independently distributed points in a disk is $\Theta\left(n^{\frac{2}{3}}\right)$. Additionally, we show that in general the bound is $\Theta\left(n^{\frac{2}{d+1}}\right)$ for points distributed in a d-dimensional ball. Further, we show that this bound holds more generally for any fixed, bounded, full-dimensional shape with a non-empty interior. The results of this thesis were published in *Random Structures and Algorithms* (2004) [1].

Résumé

Un oignon est l'ensemble des enveloppes convexes qui partitionnent un nuage de points, couche par couche. Nous prouvons dans cette thèse que le nombre moyen de couches de l'oignon de n points indépendents et distribués uniformément dans un disque est $\Theta\left(n^{\frac{2}{3}}\right)$. Nous montrons aussi que ce résultat s'étand en dimension á quelconque: pour n points indépendents et distribués uniformément dans une boule d-dimensionnelle, le nombre moyen de couches est $\Theta\left(n^{\frac{2}{d+1}}\right)$. Cette limite reste valide si on remplace la boule par un ensemble fixe de plein-dimensionnelle et d'intérieur non vide. Les résultats de prouvés ici ont été publiés dans *Random Structures and Algorithms* (2004) [1].

Acknowledgement

I wish to thank Bernard Chazelle and Luc Devroye for their help in this research and the development of this thesis. Additionally, I am very grateful to the anonymous reviewer who provided the construction for higher dimensions.

This thesis is dedicated to Allison Klein.

Contents

List of Tables

List of Figures

1 Introduction

A problem in statistical estimation is to determine the "depth" of a point with respect to a sample set. For example, in one dimension, it is useful to know how close a particular sample is to the median of the sample set. See [2] for a survey of statistical approaches for estimating the median of a set of points. One of these approaches is to iteratively compute and discard the convex hull of the samples [3] choosing the centroid of the last convex set as the median. The collection of hulls is called an *onion*, and the *depth* of a sample is defined to be the number of hulls that need to be discarded until the sample lies on the convex hull of the remaining points. For the planar case, an optimal deterministic algorithm with worst-case running time of $O(n \log n)$ was given in [4]. However, it is still unclear whether this can be improved in the 'average' case or even what the expected depth of the average case is. This paper addresses the latter question.

There are many results regarding the expected number of points on the (outermost) convex hull of a set of points that are distributed in various ways e.g., [5]. In particular, the expected number of hull points for n points uniformly and independently distributed in a disk is $\Theta\left(n^{\frac{1}{3}}\right)$ [6]; however for a square, this value is $\Theta(\log n)$[7]. Thus the geometric shape of the region where the points are distributed has a strong effect on the properties of the outermost hull and it might be supposed that the same would be true for the number of layers.

The main result of this paper is to show that for n uniformly and independently distributed points over a disk, the expected number of layers is $\Theta\left(n^{\frac{2}{3}}\right)$. More generally, we show that n such points in d-dimensional ball, the expected number of layers is $\Theta\left(n^{\frac{2}{d+1}}\right)$. Furthermore, we show that this bound holds for *every* fixed, bounded d-dimensional shape that contains a d-dimensional ball! Thus, this bound applies even to shapes that are neither convex nor contiguous.

2 Previous Work

Several of the results in this paper are known to the community but have not been published. In particular, Lemma 2 which proves a lower bound for the number of layers is easy as is the extension to general shapes. (However, the fact that the lower bound holds for non-convex shapes seems less well known.) Nonetheless, we present these results together because they are simple and they show the tightness and generality of our main result, the upper bounds proven in Theorems 1 and 2.

The results presented in this thesis have been published in [1].

3 Layout of the Thesis

We will start with a simple but powerful fact about onions in Section 4. This will be followed by upper and lower bounds for the the planar case in Section 5. In Section 6, we extend the work to higher dimensions following the same general structure as the planar case. Next, we consider shapes other than d-dimensional balls in Section 7.

4 Basics

Let S be a set of n points. We define the *depth* of S to be the number of layers in the onion generated from S. For any point $p \in S$, we define the *depth* of p to be the number of hulls that needs to be removed before p lies on the hull of the remaining points.

All of the proofs in this paper will rely on a convenient property of onions that we prove now.

Lemma 1. *Given a set S of points in $\mathbb{R}^d$, adding a point p will either leave the depth of S*

unchanged or increase it by one.

Proof. If p is inside any of the layers of S's onion, it can not affect the composition of the layers that contain p. Hence, we can assume, without loss of generality, that p lies outside of the convex hull of S. Adding p to S, p must lie on the convex hull of $p \cup S$. This may cause points that were formerly on the hull of S to be ejected from this outermost hull when p is added to S. These ejected points are necessarily in convex position and must replace a contiguous subset of the next hull (otherwise, it would imply that some point on an inner hull was not completely contained in an outer hull). Iteratively applying this argument, we continue to replace contiguous sections of hulls until we reach the center, where the current set of ejected points forms one additional hull. At any point along the iteration, the ejected subset may just merge with the next inner hull creating no hull subsection that needs to be propagated. In this case, the depth of the onion stays constant. □

5 Upper and Lower Bounds for the Disk

First, we prove a lower bound for the depth using a result from [8].

Lemma 2. *Let S be a set of n points that are distributed uniformly and independently over a d-dimensional ball. The expected depth is $\Omega\left(n^{\frac{2}{d+1}}\right)$.*

Proof. Previous work [8] has shown that the expected number of points on the convex hull of a uniform distribution of points over a general convex body, $K \subset \mathbb{R}^d$, is less than $c_d n^{(d-1)/(d+1)}$ for some constant c_d that only depends on d.

Let h_i be the number of points on the i-th hull and let k be the number of hulls. Define h_i

to be zero if $i > k$. Let t be a constant (depending only on n and d) that we will choose later.

$$\begin{aligned}
\Pr[k < t] &\leq \Pr[k \leq t] \\
&= \Pr\left[\sum_{i=1}^{\lfloor t \rfloor} h_i \geq n\right] \\
&\leq \frac{1}{n}\mathbf{E}\left[\sum_{i=1}^{\lfloor t \rfloor} h_i\right] \text{ (Markov's inequality)} \\
&= \frac{1}{n}\sum_{i=1}^{\lfloor t \rfloor} \mathbf{E}[h_i].
\end{aligned}$$

Since the (at most n) points inside any of the convex hulls are uniformly distributed in that hull, $\mathbf{E}[h_i] \leq c_d n^{(d-1)/(d+1)}$ for all i. Thus, $\Pr[k < t] \leq \frac{t}{n} c_d n^{(d-1)/(d+1)}$.

$$\begin{aligned}
\mathbf{E}[k] &\geq t \Pr[k \geq t] \\
&\geq t\left(1 - \frac{t}{n} c_d n^{(d-1)/(d+1)}\right).
\end{aligned}$$

Choosing $t = n/2\left(c_d n^{(d-1)/(d+1)}\right)$ yields,

$$\mathbf{E}[k] \geq t/2 = \frac{n}{4\left(c_d n^{(d-1)/(d+1)}\right)} = \frac{1}{4c_d} n^{\frac{2}{d+1}},$$

which completes the proof. $\square$

We prove that this bound is tight for the planar case first as it contains all of the main ideas of the general case and is more intuitive. In Theorem 2, we prove the bound for higher dimensions.

Theorem 1. *The expected depth for n points distributed uniformly and independently in a disk is $\Theta\left(n^{\frac{2}{3}}\right)$.*

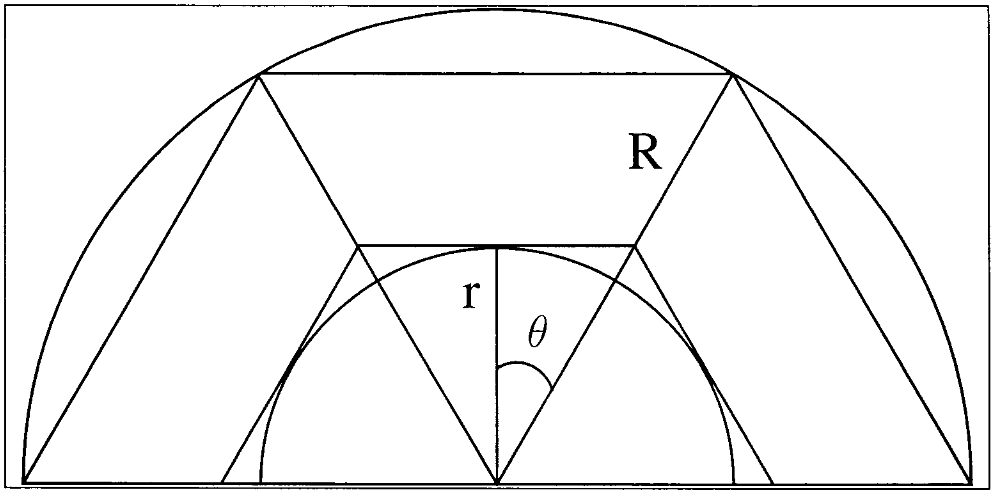

Figure 1: Calculating the area of the polygonal ring which lies between the inscribing and circumscribing m-gons.

Proof. The lower bound is shown in Lemma 2 so we need only consider the upper bound. Let $m^3 = n$. Since n can be rounded up to the nearest cube with only $O\left(n^{\frac{2}{3}}\right)$ additional points, we assume without loss of generality that m is an integer. Also we assume that n is large enough. Let K_1 be a disk with radius $R = \pi^{-\frac{1}{2}}$, thus the area of K_1 is 1. Let K_2 be a disk (concentric with K_1) with radius $r = R\left(1 - \frac{\pi^2}{m^2}\right)$.

Inscribe K_1 with a regular m-gon and then circumscribe K_2 with a homotopic copy of the same polygon, as depicted in Figure 1. Let A be the area of the polygonal ring (between the inscribing and circumscribing m-gons).

Rewriting r as $R\left(1-\theta^2\right)$ with $\theta = \frac{\pi}{m}$, we get:

$$\begin{aligned}
A &= mR^2 \sin\theta\cos\theta - mr^2\tan\theta \\
&= mR^2\left[\left(\theta - \frac{\theta^3}{6} + O\left(\theta^5\right)\right)\left(1 - \frac{\theta^2}{2} + O\left(\theta^4\right)\right)\right] - mr^2\tan\theta \\
&= mR^2\left[\left(\theta - \frac{2}{3}\theta^3 + O\left(\theta^5\right)\right) - \left(1 - 2\theta^2 + O\left(\theta^4\right)\right)\left(\theta + \frac{\theta^3}{3} + O\left(\theta^5\right)\right)\right] \\
&= \frac{m}{\pi}\left[\left(\theta - \frac{2}{3}\theta^3 + O\left(\theta^5\right)\right) - \left(\theta + \frac{\theta^3}{3} - 2\theta^3 + O\left(\theta^5\right)\right)\right] \\
&= \frac{m}{\pi}\left[\theta^3 + O\left(\theta^5\right)\right] \\
&= \frac{\pi^2}{m^2} + O\left(\frac{1}{m^4}\right) \geq \frac{9}{m^2} \text{ for } n \text{ large enough.}
\end{aligned}$$

Generate a sequence of concentric and homotopic regular m-gons starting with K_1 and K_2 keeping the ratio between consecutive copies fixed. Continue generating copies as long as the innermost polygon has area $\geq 1/4$. Because the polygons are all similar, the area of any polygonal ring is at least $A/4$ and hence the number of rings, k, is at most m^2.

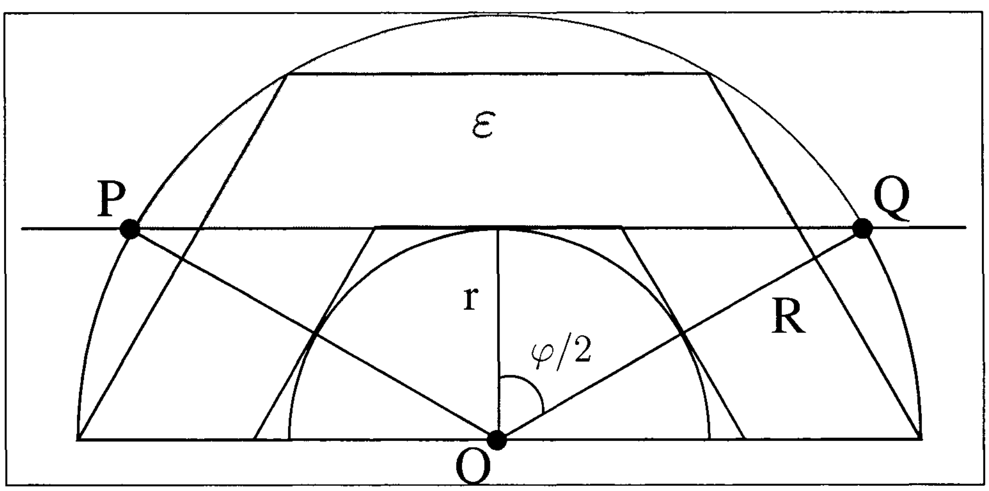

Figure 2: Calculating ε and φ.

Next, we compute two more numbers essential to this construction, ε and φ. Let ε be the area of $\tau \cap K_1$ where τ is a half-plane that is tangent to K_2 and does not include the origin, O. Let P and Q be the intersections of τ with K_1 such that $\angle POQ \leq \pi$. Let φ be the angle

$\angle POQ$. These variables are depicted in Figure 2 with ε as the area of the shaded section.

Clearly $\cos(\varphi/2) = r/R$ and $\varepsilon = \frac{\varphi}{2\pi}\left(\pi R^2\right) - Area(\triangle OPQ)$. Since $r/R \to 1$ as $n \to \infty$, we use the appropriate power series for arccos as follows:

$$\begin{aligned}
\varphi &= 2\arccos\left(\frac{r}{R}\right) \\
&= 2\sqrt{2}\sqrt{1-\frac{r}{R}}\left(1+O\left(1-\frac{r}{R}\right)\right) \\
&\leq \left(\frac{3\pi}{m}\right) \text{ for large enough } m.
\end{aligned}$$

Thus, for large enough m, the number of sides of the outermost polygon that lie within $\tau \cap K_1$ is at most $\left\lfloor \frac{3\pi}{m} \cdot \frac{m}{2\pi} \right\rfloor + 2 = 3$. (In fact, the number of adjacent sides is always odd, but this inequality is sufficient for our needs.) We call these outer sides *visible* if they intersect τ. We will use this fact soon but first we need to calculate ε.

Recalling that ε is the area of $\tau \cap K_1$, we note that ε is upper bounded by the fraction of the outer ring that is swept by φ. Thus,

$$\begin{aligned}
\varepsilon &\leq \frac{\varphi}{2\pi}\left(\pi R^2 - \pi r^2\right) \\
&\leq \frac{3}{2m}\left(1-\left(1-\frac{\pi^2}{m^2}\right)^2\right) \\
&\leq \frac{3}{2m} \cdot \frac{2\pi^2}{m^2} \\
&= \frac{3\pi^2}{m^3} \\
&\leq \frac{30}{n}.
\end{aligned}$$

Let K_i and K_{i+1} be two adjacent polygons and s a side of K_{i+1}. If τ_s is the halfplane through s that does not include the origin, then the visible sides of s are sides of K_i that intersect τ_s. Further, ε is the area of $\tau_s \cap K_i$. Using similarity, we note that φ stays the same

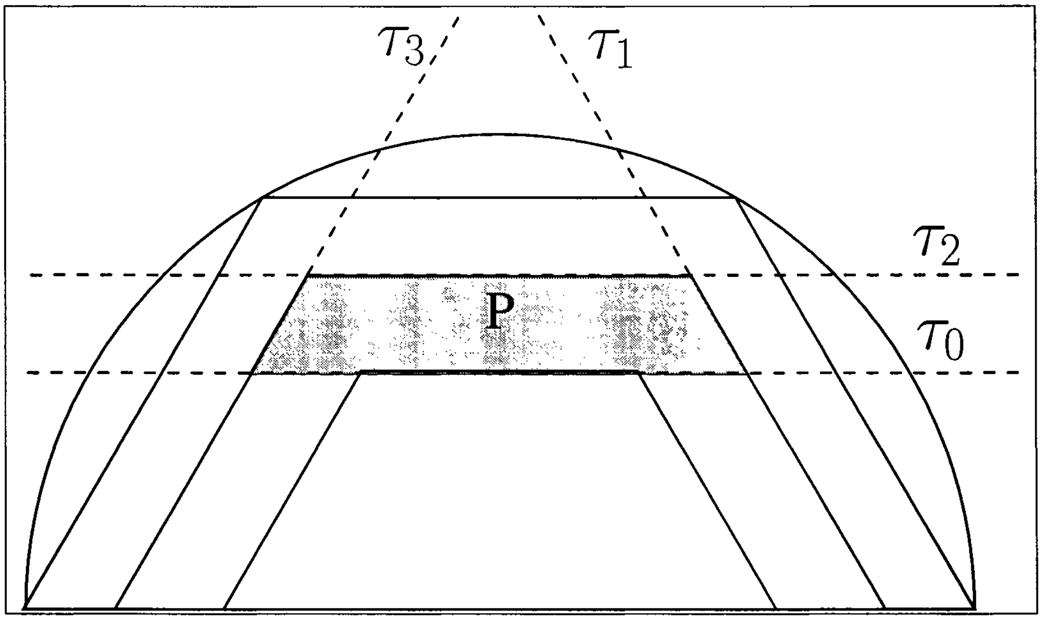

Figure 3: Calculating the depth of a visible region.

for all of the m-gons while ε can only decrease. Similarly, the number of sides that are visible to any other side is fixed at 3 for the entire construction.

The construction is now complete. The remainder of this proof will focus on imposing a combinatorial structure that will bound the depth of any point in these k rings. Observe that for any point p to have depth d where $d > 0$, every half-plane containing p must also contain a point with depth $d - 1$. (Otherwise, we could remove the first $d - 2$ hulls and then p would be alone in a half-plane and thus must lie on the $(d-1)$-th convex hull.)

We now construct a network of inequalities relating the depth of any point to its location in the disk. As in Figure 3, let P be a quadrilateral with sides, s_i, such that half-planes, τ_i, lie along each s_i and do not include the origin. Let p_{max} be the point with the largest depth that lies in τ_0 but outside of P i.e., $p_{max} \in \tau_0 \setminus P$. Then every point in P must have depth $\leq d + \|P\|$ where d is the depth of p_{max} and $\|P\|$ is the number of points in P. (It is not necessary for this proof but may be instructive to note that equality is only possible when the points in P have depths $\{d+1, d+2, \ldots, d+\|P\|\}$). Assign a value to each τ_i corresponding to the maximum depth of any point that lies in τ_i where each half-plane faces away from the center of the disk. Thus, the value for τ_0 is at most $\|P\|$ plus the maximum of the values for the other three half-planes. This is true because the union of the areas covered by τ_1, τ_2 and τ_3 will completely cover $\tau_0 \setminus P$.

Associate with every side of the m-gons a visible region defined by $K_1 \cap \tau$ where τ is a half-plane that is tangent to that side and does not include the origin. We can think of each visible region as a node in a directed acyclic graph i.e., a *DAG*. Each node (except the leaves) has three children corresponding to the three visible sides of the next largest polygon. The DAG has m roots corresponding to the m sides of the innermost polygon and has m leaves corresponding to the visible regions of the polygon circumscribing K_2. Each node performs a 'max' operation on its children and adds to that value the number of points that are in its visible region but not in its children. The maximum of the values generated by this process at each of the m roots gives an upper bound on the maximum depth of any point that lies outside the innermost m-gon.

Everything shown so far applies to any configuration of points in K_1. We now use the fact that the points in K_1 are independently and uniformly distributed to complete the proof. The maximum value of the DAG corresponds to a particular path from a root to the leaves. The number of paths is $m3^k \leq \exp(2k + \ln m) \leq \exp\left(2m^2 + \ln m\right)$ where k is the number of rings.

Observing that all paths cover the same area, let C be this area. Thus, $C \leq k\varepsilon \leq \frac{30}{m}$. Since the points in K_1 (which has unit area) are independently and uniformly distributed, the number of points that lie in any path is described by $B(n, C)$ where $B(n, p)$ is the binomial distribution. Using a version of Chernoff's Bound [9, 10] for $\delta > 1$,

$$\begin{aligned}\Pr\left[B(n, C) \geq (1+\delta)\frac{30n}{m}\right] &\leq \Pr\left[B\left(n, \frac{30}{m}\right) \geq (1+\delta)\frac{30n}{m}\right] \\ &\leq \exp\left(-30\delta m^2\right) .\end{aligned}$$

Thus for a good (constant) choice of δ, we know that, with probability greater than $1-1/n$, no path has value greater than $(1+\delta)30m^2$. Since the maximum depth is less than n, the expected maximum is at most $O\left(m^2\right) = O\left(n^{\frac{2}{3}}\right)$.

To finish the proof, let D_k be the smallest disk that encloses the innermost polygon. We note that with probability greater than $1 - 1/n$, the number of points in D_k is less than $n/2$. (The area of D_k is only $1/4 + O\left(\frac{1}{m^2}\right)$.) Let d be the maximum depth for points outside the inner polygon. Removing the outer d rings would exactly reduce the total number of layers by d. Thus by Lemma 1, removing a subset of the outer d rings can only reduce the number of layers by at most d. Specifically, we remove the points that lie outside D_k. Let $D(n)$ be the expected depth of n points in K_1. Using $D(c) = 1$ for $c \leq 3$ as the base case, we have the following difference equation:

$$\begin{aligned}D(n) &\leq O\left(n^{\frac{2}{3}}\right) + D(\lfloor n/2 \rfloor) + 1 \\ &= O\left(n^{\frac{2}{3}}\right) . \quad \square\end{aligned}$$

6 Higher Dimensions

In this section, we extend the result to higher dimensions. The essential step is to create a higher-dimensional equivalent of the regular m-gons used in the previous theorem.

Theorem 2. *The expected depth for n points distributed uniformly and independently in a d-dimensional euclidean ball is $\Theta\left(n^{\frac{2}{d+1}}\right)$.*

Proof. The lower bound is shown in Lemma 2 so we need only consider the upper bound. Let R be the radius of a d-dimensional euclidean ball with unit volume. Let $m = n^{1/(d+1)}$ which we can assume to be an integer. Let $\phi = 1/m$ and $\lambda = 1 - h$ where $h = 1/m^2$. K is the polytope to be constructed, $B(r)$ is a ball centered at the origin with radius r and $S(r)$ is its boundary. $C_\phi(a)$ is a cap centered at $a \in S(r)$ and angle ϕ. Let $b_1, b_2, \ldots$ be positive constants that depend only on d.

Choose a set of points $a_1, a_2, \ldots, a_N \in S(R)$, maximal with respect to the property such that for each i, j

$$C_\phi(a_i) \cap C_\phi(a_j) \;=\; \emptyset.$$

Since the caps are pair-wise disjoint, N times the surface are of a single cap is at most $\omega_{d-1}R^{d-1}$ where ω_{d-1} is the surface area of a unit d-dimensional ball (i.e., radius $= 1$). Thus, $N \leq b_1 m^{d-1}$. (This estimate is accurate to within a constant factor because the set of caps $C_{2\phi}(a_i)$ cover $S(R)$. Otherwise, there would be room to place another a_i in an uncovered portion of $S(R)$.)

Using $\{a_i\}$ and using $a_i a_j = a_i \cdot a_j$, we can define our polytope K by

$$K \;=\; \bigcap_{i=1}^{N} \left\{ x \in \mathbb{R}^d : a_i(x - a_i) \leq 0 \right\}$$

This polytope K has the following properties:

Claim 1. $K \subset B(R/\cos 2\phi)$.

Recall that $\lambda = 1 - h$, and $h = 1/m^2$. Let P_i be the polytope $K \cap \{x : \lambda a_i(x - \lambda a_i) \geq 0\}$.

Claim 2. For each i, Vol $P_i \le b_2/n$.

Claim 3. Again for each i, the polytope P_i has at most b_3 facets.

Proof of Claim 1. Assume $z \in K \setminus B(R)$ and let $z^* = Rz/\|z\| \in S(R)$. Then, because of the maximality of the $\{a_i\}$, we know that there exists an a_i such that

$$C_\phi(a_i) \cap C_\phi(z^*) \quad \neq \quad \emptyset$$

which shows that the maximum angle between a_i and z is at most 2ϕ. Then,

$$R^2 = a_i^2 \ge a_i z \ge \|a_i\| \|z\| \cos 2\phi = R\|z\| \cos 2\phi.$$

Thus, $\|z\| \le R/\cos 2\phi$.

Proof of Claim 2. Let C be a cap of $B\left(R/\cos 2\phi\right)$ that is cut off by a hyperplane H that is a distance λR from the origin. The volume of this cap is an upper bound for the volume of P_i. The radius, ρ of the disk $H \cap B\left(R/\cos 2\phi\right)$, can be estimated with:

$$\rho = O\left(\frac{R}{m}\right).$$

Then,

$$\begin{aligned}
\text{Vol}\, C &\le \omega_{d-1}\rho^{d-1}\left(\frac{R}{\cos 2\phi} - \lambda R\right) \\
&\le \omega_{d-1}\rho^{d-1}R\left(\frac{1}{1-2\phi^2} - (1-h)\right) \\
&\le \omega_{d-1}\rho^{d-1}R\left(1 + \frac{2\phi^2}{1-2\phi^2} - 1 + h\right) \\
&\le \omega_{d-1}\rho^{d-1}R\left(\frac{2}{m^2-2} + \frac{1}{m^2}\right) \\
&\le b_2/m^{d+1} = b_2/n,
\end{aligned}$$

which proves the claim.

Proof of Claim 3. Let H_i be the hyperplane with the equation $\lambda a_i\,(x-\lambda a_i)=0$. The facets of K are given by the inequalities $a_j\,(x-a_j)\leq 0$. If such an inequality defines a facet of P_i, then the facet has to intersect the cap C (see above) centered at a_i. Then the angle between a_i and a_j is at most $\phi+\psi$ where ψ is defined by $\sin\psi=\rho/\left(R/\cos 2\phi\right)$. Then the cap $C_\phi(a_j)$ is contained in $C_{2\phi+\psi}(a_i)$. As the small caps are pairwise disjoint, comparing the surface areas of these two caps tells that at most a constant number, b_3, of the small caps can 'fit' inside $C_{2\phi+\psi}(a_i)$ which proves the claim.

To complete the construction, set $K_0=K$ and $K_t=\lambda^t K_0=\left\{x:\lambda^{-t}x\in K_0\right\}$ for $t=0,1,\ldots T$ where $T=m^2$. Then,

$$\begin{aligned}
\mathrm{Vol}\;K_T &= \lambda^{dT}\mathrm{Vol}\;K_0\\
&\leq (1-h)^{dT}(\cos 2\phi)^{-d}\\
&\leq \exp\{-hdT\}\exp\left\{db_4/m^2\right\}\\
&\leq \exp\left\{-d+db_4/m^2\right\}\\
&\leq 2^{-d}\text{ (if } m \text{ is large enough).}
\end{aligned}$$

We can now write $P(i,t)=\lambda^t P_i$, and let X_n be the random sample of n uniform and independent points from $B(R)$. The cells $P(i,t)$ form a DAG with max degree $D\leq b_3$ (from Claim 2), with the arrows directed from cell $P(i,t)$ to $P(j,t-1)$ if these two cells touch along facets.

The depth of any point in $B(R)\setminus B(R/2)$ is at most the maximum of the number of points taken along any path through this DAG. The number of paths is at most N (the number of facets of K) times D^T. Thus,

$$ND^T \leq \exp(b_5 m^2).$$

And the sum of the volumes along any path is at most

$$T \text{ Vol } P_i \leq b_6 m^2/n,$$

and thus the expected number of points in a path is $b_6 m^2$. Using the same Chernoff bounding technique as in Theorem 1, we find that for an appropriate choice of δ, the probability that no path has length greater than $(1+\delta)b_6 m^2$ is at least $1 - 1/n$. Since the longest path is bounded by n, this means that the expected maximum is less than $O\left(m^2\right) = O\left(n^{2/d+1}\right)$. Since, with high probability, there are only a constant fraction of points remaining inside K_T, the rest of the argument holds. □

7 Beyond The Disk

All well and good, but does the expected number of layers for n points distributed uniformly and independently in a square vary significantly from that of a disk? In fact, it was to answer this question that this research was initiated. In particular, as mentioned earlier, the number of points on the convex hull of a square is $\Theta(\log n)$[7] as opposed to the $\Theta(n^{1/3})$[6] for the disk, so one might suppose that the depth of the square should be much greater than that of the disk. In this section, we show that the same asymptotic upper and lower bounds hold for both cases, as well as for a broad family of other shapes.

Recalling that the *depth* of the set S is defined as the number of layers of S's onion, we note a direct consequence of Lemma 1.

Corollary 1. *Let S, T be sets of points such that $T \subseteq S$. The depth of S is at least the depth of T.*

Let $\mathbf{E}R(n)$ be the expected depth for n points distributed independently and uniformly in some region $R \subset \mathbb{R}^d$. Specifically, let $\mathbf{E}D(n)$ be the expected depth for n points distributed independently and uniformly in a d-dimensional ball. Also let $||R||$ and $||D||$ be the number of points contained in R and D respectively.

Lemma 3. *For any bounded region $R \subset \mathbb{R}^d$ that has a non-empty interior, the expected depth of n points distributed independently and uniformly in R has the same asymptotic lower bound (within a constant factor) as for the d-dimensional ball. Formally,*

$$\mathbf{E}R(n) = \Omega\left(n^{\frac{2}{d+1}}\right).$$

Proof. Let D be a d-dimensional ball of non-zero volume entirely contained in the region R. Since R has a non-empty interior, this is always possible. Since R is bounded, we can (uniformly) scale R until it has unit volume without changing any of the convex hulls. Let c be the volume of D after such a rescaling. By Corollary 1, the depth of the points in R is at least the depth of the points in D. Since this is always true for any particular configuration, it is also true in expectation. The number of points in D is described by the distribution $B(n, c)$ and the probability that this is less than $nc/2$ is much less than $1/n$ for n large enough. Thus, using Lemma 2:

$$
\begin{aligned}
\mathbf{E}R(n) &\geq \mathbf{E}\left[\text{depth of points in } D\right] \\
&\geq \mathbf{E}\left[D\left(\frac{nc}{2}\right)\right] \Pr\left[\|D\| \geq \frac{nc}{2}\right] \\
&\geq \Omega\left(\left(\frac{nc}{2}\right)^{\frac{2}{d+1}}\right)\left(1 - \frac{1}{n}\right) \\
&= \Omega\left(n^{\frac{2}{d+1}}\right). \quad \square
\end{aligned}
$$

Lemma 4. *Let $R \subset \mathbb{R}^d$ be a bounded region such that Vol $(R) > 0$. Then the expected depth of n points distributed independently and uniformly in R has the same asymptotic upper bound (within a constant) as for the d-dimensional ball. Formally,*

$$
\mathbf{E}R(n) = O\left(n^{\frac{2}{d+1}}\right)
$$

.

Proof. Let D be a finite d-dimensional ball entirely containing the region R. Since R is bounded this is always possible. Since D is finite, we can (uniformly) scale D until it has unit volume without changing any of the convex hulls. Let c be the volume of R after such a rescaling. By Corollary 1, the depth of the points in R is at most the depth of the points in D. Since this is true for any particular configuration, this must be true for the expectation. We uniformly and independently distribute $2n/c$ points in D.

Since the number of points in R is described by $B(2n/c, c)$, the probability that R will have fewer than n points is much less than $1/n$ for n large enough. If R has too few points by this construction, we can use an upper bound of n for the depth. Otherwise, we use the depth of D. Thus, using Theorem 2:

$$\begin{aligned}
\mathbf{E}R(n) &\leq \mathbf{E}\,[\text{depth of points in } D] + n \Pr[\|R\| < n] \\
&\leq \mathbf{E}\left[D\left(\frac{2n}{c}\right)\right] \Pr[\|R\| \geq n] + n \Pr[\|R\| < n] \\
&\leq O\left((2n/c)^{\frac{2}{d+1}}\right)\left(1 - \frac{1}{n}\right) + O(1) \\
&= O\left(n^{\frac{2}{d+1}}\right). \quad \square
\end{aligned}$$

Hence, upper and lower bounds for the expected depth of a uniform distribution over a d-dimensional ball will apply to many interesting regions, R. The only requirements are for R to be fixed, bounded and have a non-empty interior.

References

[1] K. Dalal. Counting the onion. *Random Structures And Algorithms*, 24:155–165, 2004.

[2] C. G. Small. A survey of multidimensional medians. *International Statistical Review*, 58:263–277, 1990.

[3] F. Preparata and M. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, 1988.

[4] B. Chazelle. Optimal algorithms for computing depths and layers. *Proceedings 21st Allerton Conference on Comm. Control and Comput.*, pages 427–436, 1983.

[5] B. Efron. The convex hull of a random set of points. *Biometrika*, 52:331–343, 1965.

[6] H. Raynaud. Sur l'enveloppe convexe des nuages de points aléatoires dans $\mathcal{R}^n$. *J. of Appl. Probab.*, 7:35–48, 1970.

[7] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten I. *Z. Wahrsch. verw. Gebiete*, 2:75–84, 1963.

[8] I. Bárány. Intrinsic volumes and f-vectors of random polytopes. *Math. Ann*, 285:671–699, 1989.

[9] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[10] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.