A Proof of the Oja-depth Conjecture in the Plane

Nabil H. Mustafa* Hans Raj Tiwary† Daniel Werner‡

Abstract

Given a set P of n points in the plane, the *Oja-depth* of a point $x \in \mathbb{R}^2$ is defined to be the sum of the areas of all triangles defined by x and two points from P, normalized by the area of convex-hull of P. The Oja-depth of P is the minimum Oja-depth of any point in $\mathbb{R}^2$. The Oja-depth conjecture states that any set P of n points in the plane has Oja-depth at most $n^2/9$ (this would be optimal as there are examples where it is not possible to do better). We present a proof of this conjecture.

We also improve the previously best bounds for all $\mathbb{R}^d$, $d \geq 3$, via a different, more combinatorial technique.

1 Introduction

We first present some examples of the several different versions of data-depth that have been studied.

The *location-depth* of a point x is the minimum number of points of P lying in any halfspace containing x [11, 20, 19]. The Center-point Theorem [9] asserts that there is always a point of location-depth at least $n/(d+1)$, and that this is the best possible. The point with the highest location-depth w.r.t. to a point-set P is called the *Tukey-median* of P. The corresponding computational question of finding the Tukey-median of a point-set has been studied extensively, and an optimal algorithm with running time $O(n \log n)$ is known in $\mathbb{R}^2$ [7].

The *simplicial-depth* [13] of a point x and a set P is the number of simplices spanned by P that contain x. The First Selection Lemma [14] asserts that there always exists a point with simplicial-depth at least $c_d \cdot n^{d+1}$, where $c > 0$ is a constant depending only d. The optimal value of c_d is known only for $d = 2$, where $c_2 = 1/27$ [5]. For c_3 is still open, though it has been the subject of a flurry of work recently [3, 6, 10]. The current-best algorithm computes the point with maximum simplicial-depth in time $O(n^4 \log n)$ [1].

The L_1 depth, proposed by Weber in 1909, is defined to be the sum of the distances of x to the n input points. It is known that the point with the lowest such depth is unique in $\mathbb{R}^2$.

Oja-depth. In this paper, we study another well-known measure called the *Oja depth* of a point-set. Given a set P of n points in $\mathbb{R}^d$, the *Oja-depth* (first proposed by Oja [16] in 1983) of a point $x \in \mathbb{R}^d$ w.r.t. P is defined to be the sum of the volumes of all d-simplices spanned by x and d other points of P. Formally, given a set $Q \subset \mathbb{R}^d$, let $\mathrm{conv}(Q)$ denote the convex-hull of Q, and let $\mathrm{vol}(Q)$ denote its d-dimensional volume. Then,

$$\text{Oja-depth}(x) = \sum_{y_1,\ldots,y_d \in \binom{P}{d}} \frac{\mathrm{vol}(\mathrm{conv}(x, y_1, \ldots, y_d))}{\mathrm{vol}(\mathrm{conv}(P))}$$

The Oja-depth of P is the minimum Oja-depth over all $x \in \mathbb{R}^d$. From now onwards, w.l.o.g., assume that $\mathrm{vol}(\mathrm{conv}(P)) = 1$.

Known bounds. First we note that

$$\left(\frac{n}{d+1}\right)^d \leq \text{Oja-depth}(P) \leq \binom{n}{d}.$$

For the upper-bound, observe that any d-simplex spanned by points inside the convex-hull of P can have volume at most 1, and so a trivial upper-bound for Oja-depth of any $P \subset \mathbb{R}^d$ is $\binom{n}{d}$, achieved by picking any $x \in \mathrm{conv}(P)$. For the lower-bound, construct P by placing $n/(d+1)$ points at each of the $d+1$ vertices of a unit-volume simplex in $\mathbb{R}^d$.

The conjecture [8] states that this lower bound is tight:

Conjecture 1 *Oja-depth*$(P) \leq (\frac{n}{d+1})^d$ *for any* $P \subset \mathbb{R}^d$ *of* n *points.*

The current-best upper-bound [8] is that the Oja-depth of any set of n points is at most $\binom{n}{d}/(d+1)$. In particular, for $d = 2$, this gives $n^2/6$.

The Oja-depth conjecture states the existence of a low-depth point, but given P, computing the *lowest*-depth point is also an interesting problem. In $\mathbb{R}^2$, Rousseeuw and Ruts [18] presented a straightforward $O(n^5 \log n)$ time algorithm for computing the lowest-depth point, which was improved to the current-best

*Dept. of Computer Science, LUMS, Pakistan. nabil@lums.edu.pk

†Département de Mathématique, Université Libre de Bruxelles, Belgium, hans.raj.tiwary@ulb.ac.be

‡Institut für Informatik, Freie Univ. Berlin, Germany. daniel.werner@fu-berlin.de - This research was funded by Deutsche Forschungsgemeinschaft within the Research Training Group (Graduiertenkolleg) “Methods for Discrete Structures”

algorithm with running time $O(n \log^3 n)$ [1]. An approximate algorithm utilizing fast rendering systems on current graphics hardware was presented in [12, 15]. For general d, various heuristics for computing points with low Oja-depth were given by Ronkainen, Oja and Orponen [17].

Our results. In Section 2, we present our main theorem, which completely resolves the conjecture for the planar case.

Theorem 1 *Every set P of n points in $\mathbb{R}^2$ has Oja-depth at most $\frac{n^2}{9}$. Furthermore, such a point can be computed in $O(n \log n)$ time.*

In Section 3, using completely different (and more combinatorial) techniques for higher dimensions, we also prove the following:

Theorem 2 *Every set P of n points in $\mathbb{R}^d$, $d \geq 3$, has Oja-depth at most $\frac{2n^d}{2^d d!} - \frac{2d}{(d+1)^2(d+1)!}\binom{n}{d} + O(n^{d-1})$.*

This improves the previously best bounds by an order of magnitude.

2 The optimal bound for the plane

We now come to prove the optimal bound for $\mathbb{R}^2$. First, let us give some basic definitions. The *center of mass* or *centroid* of a convex set X is defined as

$$c(X) = \frac{\int_{x \in X} x \; dx}{\text{area}\,(X)}.$$

For a discrete point set P, the center of mass is simply defined as the center of mass of the convex hull of P. When we talk about the *centroid of P*, we refer to the center of mass of the convex hull and hope the reader does not confuse this with the discrete centroid $\sum p/|P|$. In what follows, we will bound the Oja-depth of the centroid of a set, and show that it is worst-case optimal. Our proof will rely on the following two Lemmas.

Lemma 3 *[Winternitz [4]] Every line through the centroid of a convex object has at most $\frac{5}{9}$ of the total area on either side.*

Lemma 4 *[8] Let P be a convex object with unit area and let c be its center of mass. Then every simplex inside P which has c as a vertex has area at most $\frac{1}{3}$.*

To simplify matters, we will use the following proposition.

Proposition 5 *If we project an interior point $p \in P$ radially outwards from the centroid c to the boundary of the convex hull, the Oja-depth of the point c does not decrease.*

Proof. First, observe that the center of mass does not change. It suffices to show that every triangle that has p as one of its vertices increases its area. Let $T := \Delta(c, p, q)$ be any triangle. The area of T is $\frac{1}{2}\|c - p\| \cdot h$, where h is the height of T with respect to $p - c$. If we move p radially outwards to a point p', h does not change, but $\|c - p'\| > \|c - p\|$. □

This implies that in order to prove an upper bound, we can assume that all points lie on the convex hull.

From now on, let P be a set of points, and let $c := c(\text{conv}(P))$ denote its center of mass as defined above. Further, let $p_1, \ldots, p_n$ denote the points sorted clockwise by angle from c. We define the *distance* of two points as the difference of their position in this order (modulo n). A triangle that is formed by c and two points at distance i is called an *i-triangle*, or *triangle of type i*. Observe that for each i, $1 \leq i < \lfloor n/2 \rfloor$, there are exactly n triangles of type i. Further, if n is even, then there are $n/2$ triangles of type $\lfloor n/2 \rfloor$, otherwise there are n. These constitute all possible triangles.

Let $C \subseteq P$, and let $\mathcal{C}$ be they boundary of the convex hull of C. This will be called a *cycle*. The length of a cycle is simply the number of elements in C. A cycle $\mathcal{C}$ of length i *induces i triangles* that arise by taking all the triangle formed by an edge in $\mathcal{C}$ and the center of mass c (of conv(P)). The area induced by $\mathcal{C}$ is the sum of areas of these i triangles.

The triangles induced by the entire set P form a partition of conv(P). Thus, Lemma 5 implies the following:

Corollary 6 *The total area of all triangles of type* 1 *is exactly* 1.

The following shows that we can generalize this Lemma, i.e., that we can bound the total area induced by *any* cycle.

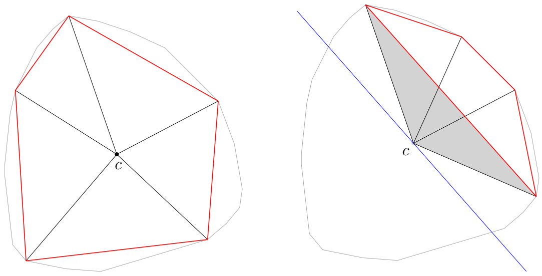
Lemma 7 *Let $\mathcal{C}$ be a cycle. Then $\mathcal{C}$ induces a total area of at most* 1.

Proof. We distinguish two cases.

Case 1: The centroid lies in the convex hull of $\mathcal{C}$. In this case, all triangles are disjoint, so the area is at most 1. See Fig. 1(a).

Case 2: The centroid does not lie in the convex hull of $\mathcal{C}$. By the Separation Theorem [14], there is a line through c that contains all the triangles. Then we can remove one triangle to get a set of disjoint triangles, namely the one induced by the pair $\{p_{i_j}, p_{i_{j+1}}\}$ that has c on the left side. By Lemma 3, the area of the remaining triangles can thus be at most 5/9. By Lemma 4, the removed triangle has an area of at most 1/3. Thus, the total area is at most 8/9. See Fig. 1(b). Here, the gray triangle can be removed to get a set of disjoint triangles.

□



(a) Case 1 (b) Case 2

Figure 1: The two cases

We now prove the crucial lemma, which is a general version of Corollary 6.

Lemma 8 *The total area of all triangles of type i is at most i.*

Proof. We will proceed as follows: For fixed i, we will create n cycles. Each cycle will consist of one triangle of type i, and $n-i$ triangles of type 1, multiplicities counted. We then determine the total area of these cycles and subtract the area of all 1-triangles. This will give the desired result.

Let $p_1, \dots, p_n$ be the points ordered by angles from the centroid c. Let $\mathcal{C}_j$ be the cycle consisting of the $n-i+1$ points $P - \{p_{i+1 \mod n}, \dots, p_{i+j-1 \mod n}\}$. This is a cycle that consists of one triangle of type i, namely the one starting a p_j, and $n-i$ triangles of type 1.

By Lemma 7, every cycle $\mathcal{C}_j$ induces an area of at most 1. If we sum up the areas of all n cycles $\mathcal{C}_j$, $1 \le j \le n$, we thus get an area of at most n.

We now determine how often we have counted each triangle. Each i-triangle is counted exactly once. Further, for every cycle we count $n-i$ triangles of type 1. For reasons of symmetry, each 1-triangle is counted equally often. Thus, each is counted *exactly* $n-i$ times over all the cycles. By Corollary 6, their area is *exactly* $n-i$, which we can subtract from n to get the total area of the i-triangles:

$$\begin{aligned} \sum_{i-\Delta\ T} \operatorname{area}(T) &\le n - \left(\sum_{1-\Delta\ T} (n-i) \operatorname{area}(T) \right) \\ &= n-(n-i) = i. \end{aligned}$$

This completes the proof. □

Theorem 9 *Let P be any set of points in the plane and c be the centroid of its convex hull. Then the Oja-depth of c is at most $\frac{n^2}{9}$.*

Proof. We will bound the area of the triangles depending on their type. For i-triangles with $1 \le i \le \lfloor n/3 \rfloor$, we will use Lemma 8. For i-triangles with $\lfloor n/3 \rfloor < i \le \lfloor n/2 \rfloor$, this would give us a bound worse than $n/3$, so we will use Lemma 4 for each of these.

By Lemma 8, the sum of the areas of all triangles of type at most $\lfloor n/3 \rfloor$ is at most

$$\sum_{i=1}^{\lfloor n/3 \rfloor} i = \frac{\lfloor n/3 \rfloor (\lfloor n/3 \rfloor + 1)}{2} \le \frac{n^2}{18} + \frac{1}{2} \lfloor n/3 \rfloor .$$

For the remaining triangles, we use Lemma 4 to bound the size of each by $1/3$. Thus, in total we get

$$\text{Oja-depth}(P) \le \frac{n^2}{18} + \frac{n\left(\lfloor n/2 \rfloor - \lfloor n/3 \rfloor\right)}{3} + \frac{n}{6}.$$

By a simple case distinction, it is easy to see that the lower order term disappears. This finishes the proof. □

3 Higher Dimensions

We now present improved bounds for the Oja-depth problem in dimensions greater than two. Before the main theorem, we need the following two lemmas.

Lemma 10 *Let P be a set of n points in $\mathbb{R}^d$. Let $q \in \mathbb{R}^d$. Then any line l through q intersects at most $f(n,d)$ $(d-1)$-simplices spanned by P, where $f(n,d) = \frac{2n^d}{2^d d!} + O(n^{d-1})$.*

Proof. Project P onto the hyperplane H orthogonal to l to get the point-set P' in $\mathbb{R}^{d-1}$. The line l becomes a point on H, say point p_l. Then l intersects the $(d-1)$-simplex spanned by $\{p_1, \dots, p_d\}$ if and only if the convex hull of the corresponding points in P' contain the point p_l.

By a result of Barany [2], any point in $\mathbb{R}^d$ is contained in at most

$$\frac{2(n-d)}{n+d+2} \binom{(n+d+2)/2}{d+1} + O(n^d)$$

simplices induced by a point set.

Applying this lemma to P' in $\mathbb{R}^{d-1}$ and simplifying the expression, we get the desired result. □

Lemma 11 *Given any set P of n points in $\mathbb{R}^d$, there exists a point q such that any half-infinite ray from q intersects at least $\frac{2d}{(d+1)^2(d+1)!}\binom{n}{d}$ $(d-1)$-simplices spanned by P.*

Proof. Gromov [10] showed that, given any set P, there exists a point q contained in at least $\frac{2d}{(d+1)(d+1)!}\binom{n}{d+1}$ simplices spanned by P. Now any half-infinite ray from q must intersect exactly one $(d-1)$-dimensional face (which is a $(d-1)$-simplex) of each d-simplex containing q, and each such $(d-1)$-simplex can be counted at most $n-d$ times. □

Theorem 12 *Given any set P of n points in $\mathbb{R}^d$, there exists a point q with Oja-depth at most*

$$B := \frac{2n^d}{2^d d!} - \frac{2d}{(d+1)^2(d+1)!}\binom{n}{d} + O(n^{d-1}).$$

Proof. Let q be the point from Lemma 11. Let $w(r)$ denote the number of simplices spanned by q and d points from P that contain r. In what follows, we will give a bound on $w(r)$, and thus on the Oja-depth of q.

If r is contained in a simplex, then any half-infinite ray $\overrightarrow{qr}$ intersects a $(d-1)$-facet of that simplex. Therefore, $w(r)$ is upper-bounded by the number of $(d-1)$-simplices spanned by P that are intersected by the ray $\overrightarrow{qr}$.

To upper-bound this, note that the ray starting from q but in the opposite direction to the ray $\overrightarrow{qr}$, intersects at least $\frac{2d}{(d+1)^2(d+1)!}\binom{n}{d}$ $(d-1)$-simplices (by Lemma 11). On the other hand, by Lemma 10, the entire line passing through q and r intersects at most $\frac{2n^d}{2^d d!} + O(n^{d-1})$ $(d-1)$-simplices. These two together imply that the ray $\overrightarrow{qr}$ intersects at most B $(d-1)$-simplices, and this is also an upper-bound on $w(r)$. Finally, we have

$$\begin{aligned} \text{Oja-depth}(q,P) &= \int_{\text{conv}(P)} w(x)\,dx \\ &\leq \int_{\text{conv}(P)} B\,dx = B \end{aligned}$$

finishing the proof. □

Acknowledgments. This research was done during the DCG Special Semester in Lausanne. We thank the EPFL and the organizers János Pach and Emo Welzl.

References

[1] G. Aloupis. On computing geometric estimators of location, 2001. M.Sc. Thesis, McGill University.

[2] I. Barany. A generalization of caratheodory's theorem. *Discrete Mathematics*, 40:141–152, 1982.

[3] A. Basit, N. H. Mustafa, S. Ray, and S. Raza. Hitting simplices with points in 3d. *Discrete & Computational Geometry*, 44(3):637–644, 2010.

[4] W. Blaschke. *Vorlesungen uber Differentialgeometrie. II, Affine Differentialgeometrie.* Springer, 1923.

[5] E. Boros and Z. Furedi. The maximal number of covers by the triangles of a given vertex set on the plane. *Geom. Dedicata*, 17:69–77, 1984.

[6] B. Bukh, J. Matousek, and G. Nivasch. Stabbing simplices by points and flats. *Discrete & Computational Geometry*, 43(2):321–338, 2010.

[7] T. M. Chan. An optimal randomized algorithm for maximum tukey depth. In *SODA*, pages 430–436, 2004.

[8] D. Chen, O. Devillers, J. Iacono, S. Langerman, and P. Morin. Oja Medians and Centers of Mass. In *Proceedings of the 22nd CCCG*, pages 147–150, 2010.

[9] J. Eckhoff. Helly, Radon and Carathéodory type theorems. In *Handbook of Convex Geometry*, pages 389–448. North-Holland, 1993.

[10] M. Gromov. Singularities, expanders and topology of maps. part 2: From combinatorics to topology via algebraic isoperimetry. *Geometric and Functional Analysis*, 20:416–526, 2010.

[11] J. Hodges. A bivariate sign test. *Annals of Mathematical Statistics*, 26:523–527, 1955.

[12] S. Krishnan, N. Mustafa, and S. Venkatasubramanian. Statistical data depth and the graphics hardware. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, pages 223–250. DIMACS Series, 2006.

[13] R. Liu. A notion of data depth based upon random simplices. *The Annals of Statistics*, 18:405–414, 1990.

[14] J. Matoušek. *Lectures in Discrete Geometry.* Springer-Verlag, New York, NY, 2002.

[15] N. H. Mustafa. *Simplification, Estimation and Classification of Geometric Objects.* PhD thesis, Duke University, Durham, North Carolina, 2004.

[16] H. Oja. Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, 1:327–332, 1983.

[17] T. Ronkainen, H. Oja, and P. Orponen. Computation of the multivariate oja median. In *R. Dutter and P. Filzmoser.* International Conference on Robust Statistics, 2003.

[18] P. Rousseeuw and I. Ruts. Bivariate location depth. *Applied Statistics*, 45:516–526, 1996.

[19] P. J. Rousseeuw, I. Ruts, and J. W. Tukey. The bagplot: A bivariate boxplot. *The American Statistician*, 53(4):382–387, 1999.

[20] J. Tukey. Mathematics and the picturing of data. In *Proc. of the international congress of mathematicians*, pages 523–531, 1975.