Point set stratification and minimum weight structures

Manuel Abellanas [a,2], Mercè Claverol [*,b,1], Ferran Hurtado [c,1],

[a] *Dpt. Mat. Aplic., FI. UPM. , Boadilla del Monte, 28660 Madrid, Spain*
[b] *Dpt. Mat. Aplic. IV, UPC. EPSEVG. Vilanova i La Geltrú 08800, Spain*
[c] *Dpt. Mat. Aplic. II, UPC. FIB. FME., 08028 Barcelona, Spain*

Abstract

Three different concepts of depth in a point set are considered and compared: Convex depth, location depth and Delaunay depth. As a notion of weight is naturally associated to each depth definition, we also present results on minimum weight structures (like spanning trees, poligonizations and triangulations) with respect to the three variations.

Key words: Tukey depth, halfspace depth, convex depth, Delaunay depth, minimum weight, layers.

1. Introduction

In bivariate data analysis, different ways of partitioning data sets as well as peeling methods (mainly for outlier rejection) have been proposed by different authors according to several definitions of *depth*. Every notion of depth of a point with respect to a point set S gives rise to a partition of the set S into *layers* and also to a partition of the whole plane into *levels*. The layers are the subsets of points of S having the same depth. The level of a point q with respect to S is the depth of q in $S \cup \{q\}$. The *weight* of a segment is the absolute value of the difference of depth between its vertices. There are also several ways to associate weights to geometric structures defined by the points of S and some set of edges (segments with endpoints in S), as described in Section 4.

In [10](pg. 363) Okabe et al. mention the interest in comparing Delaunay depth with respect to other depths. In this paper we do a comparative study of properties of layers and levels associated to finite sets of points in the plane considering three different definitions of depth: *convex depth* [8], *Tukey depth* also know as *halfspace depth* or *location depth* [11] and *Delaunay depth* introduced by Green in [7]. A thorough study is presented in [5].

After introducing basic definitions in Section 2, we study and compare the complexity of layers and levels in Section 3, and we give in Section 4 properties of minimum weight structures (spanning trees, poligonizations and triangulations) related to the three depths considered.

2. Preliminaries

Let S be a set of n points in the plane, $CH(S)$ the convex hull of S and p any point of S. Any generic depth of p with respect to S is denoted by $d_S(p)$.

The *convex depth* of p, is defined recursively as follows: if $p \in CH(S)$, $d_S(p) = 1$, else $d_S(p) = d_{S \setminus CH(S)}(p) + 1$. For values of $j \leq \lfloor n/2 \rfloor$ we say that the *location depth* of p is $d_S(p) = j$ if and only if there is a line through p leaving exactly $j - 1$ points on one side, but no line through p separates a smaller subset. The *Delaunay depth* of p is defined to be $d + 1$ when the graph theoretical distance from p to $CH(S)$ in the Delaunay triangulation $DT(S)$ of S is d. In all three cases we call *depth of* S the depth of its deepest point.

* Corresponding author

Email addresses: mabellanas@fi.upm.es (Manuel Abellanas), merce@mat.upc.es (Mercè Claverol), Ferran.Hurtado@upc.es (Ferran Hurtado).

[1] Partially supported by *DURSY* 2001SGR00224 and *MCYT* BFM2003-0368

[2] Partially supported by *MCYT* TIC2003-08933-C02-01

	$d_S(p)$ (Depth)	$Lay_i(S)$ (Layer i)	$Lev_i(S)$ (Level i)	
Convex	if $p \in CH(S)$, $d_S(p) = 1$ else $d_S(p) = d_{S \setminus CH(S)}(p) + 1$	$Lay_i(S) = CH(S_i)$ $S_i = \{x \in S / d_S(x) = i\}$	Depth of a point relative to a set S $d(p, S) = d_{S \cup \{p\}}(p)$	$Lev_i(S)$= $\{x \in \mathbb{R}^2 / d(x, S) = i\}$
Location depth	$d_S(p) = j, \quad j \leq \lfloor \vert S\vert /2 \rfloor \Leftrightarrow$ some line through p leaves exactly $j-1$ points on one side, none leaves less			
Delaunay	if $p \in CH(S)$, $d_S(p) = 1$ else $d_S(p) =$ distance from p to $CH(S)$ $+1$, in $DT(S)$	$Lay_i(S)$ =subgraph of $DT(S)$ induced by S_i $S_i = \{x \in S \vert d_S(x) = i\}$		

Table 1: Definitions

The i-th *layer* of S, $Lay_i(S)$, is defined for convex depth as well as for location depth by $Lay_i(S) = CH(S_i)$, where $S_i = \{x \in S \mid d_S(x) = i\}$, (Figures 1 and 2). For the Delaunay depth, $Lay_i(S)$ is the subgraph of $DT(S)$ induced by S_i, (Figure 3).

Let p be any point in the plane. For the three depths considered, the depth of p relative to the set S is $d(p, S) = d_{S \cup \{p\}}(p)$ and the i-th *level* for the set S is defined by $Lev_i(S) = \{x \in \mathbb{R}^2 | d(x, S) = i\}$. The concept of k-hull introduced by Cole, Sharir and Yap in [6] corresponds to $\bigcup_{j \geq k} Lev_j(S)$.

Table 1 shows all these definitions together.

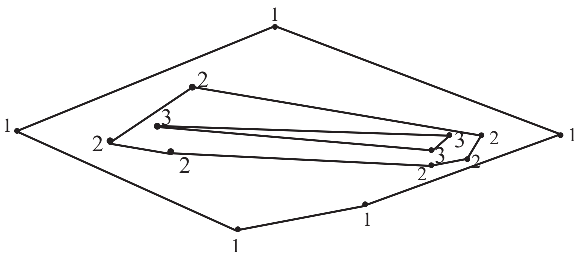

Fig. 1. Convex layers.

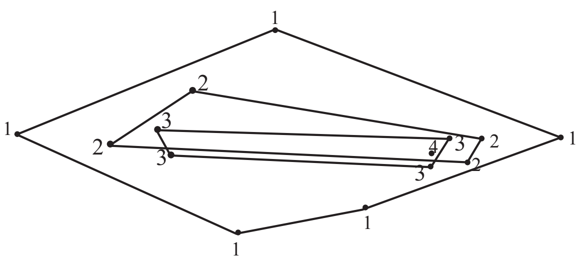

Fig. 2. Location layers.

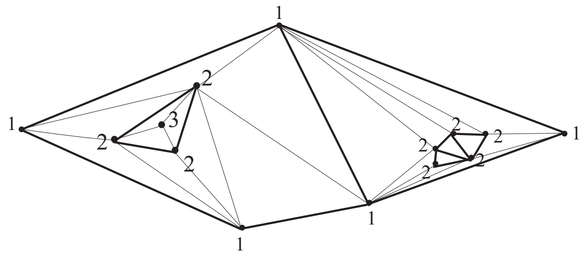

Fig. 3. Delaunay layers.

3. Point set stratification

Given a set S of n points in the plane the convex layers can be constructed with Chazelle's optimal $O(n \log n)$ algorithm [4]. Convex layers form a sequence of nested convex polygons defining a partition of the plane into regions, which coincide with the levels, (Figure 1). Therefore layers and levels have linear complexity in the convex depth case and can be constructed in optimal $O(n \log n)$ time.

As for location depth, a worst case optimal algorithm for computing all $Lev_i(S)$, (where $n/3 \leq i \leq n/2$) in $O(n^2)$ time is obtained by using topological sweep in the dual arrangement of lines (see [5], [9]). The boundaries of the levels, in this case, form a sequence of nested convex polygons. Points of $Lay_i(S)$ are in convex position and belong to the boundary of $Lev_i(S)$, but this boundary can also have other vertices not in S, (Figure 4). Some layers can be empty and different layers can cross each other. While the complexity of levels may reach $O(n^2)$, the size of the layers is $O(n)$. The layers in the location depth case can be computed using the mentioned $O(n^2)$ sweep algorithm yet, to our knowledge, it is an open problem to construct them in less time or to prove a quadratic lower bound for the problem.

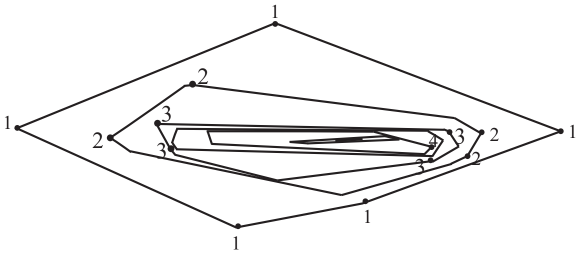

Fig. 4. Location levels.

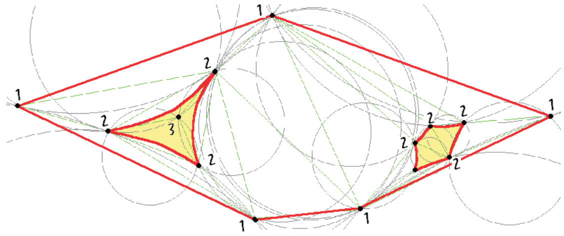

Fig. 5. Delaunay levels.

For Delaunay depth, all the layers $Lay_i(S)$, $i \leq n/3$, can easily be found by visiting $DT(S)$ in linear time once constructed, which requires $O(n \log n)$ time (Figure 3). Notice that one layer can have more than one connected component. The maximum number of connected components is strictly decreasing on the number of layers, (see [5]). This maximum number varies between $\lfloor (n+2)/3 \rfloor$ and $\lfloor n/2 \rfloor$, attained with suitable constructions having $\lfloor n/3 \rfloor$ and 2 layers, respectively.

Delaunay layers are not necessarily polygons, however they form a structure based in nested cycles of points of the same weight.

The weight of a point relative to a set S depends on the Delaunay circles (circumcircles of Delaunay triangles) that contain the point, therefore the arrangement of Delaunay circles contains all the information about Delaunay levels, (Figure 5). As it has size $O(n^2)$ and can be constructed in $O(n^2 \log n)$ time, it is possible to obtain the Delaunay levels within this time. Nevertheless, in the following theorem we prove that in order to obtain all the $Lev_i(S)$ it is not necessary to construct the whole arrangement of circles.

Lemma 1 *Let S be a set of points in the plane. If the Delaunay depth of a point p with respect to S is j, there is a cycle of $Lay_{j-1}(S)$ containing p in its interior.*

As a consequence, the number of levels for Delaunay depth is equal to the number of layers or to the number of layers plus one.

Observation 2 *Let C be a circle having exactly two points u and v of S on its boundary and containing no points of S in its interior. Then any circle crossing the two arcs determined by u and v in the boundary of C contains some interior point from S.*

Theorem 3 *Let S be a set of points in the plane being f its Delaunay depth . The Delaunay levels of S are nested sets. The boundaries between $Lev_j(S)$ and $Lev_{j+1}(S)$, for $2 \leq j \leq f$, are curves composed by arcs of the Delaunay circles determined by two points u, v of $Lay_j(S)$ and one point w of $Lay_{j-1}(S)$.*

Theorem 3 proves that the overall size of the Delaunay levels is $O(n)$ and justifies the steps of the following algorithm.

Algorithm 1 Delaunay levels of S.

1. *Compute $DT(S)$.*
2. *Compute the Delaunay depths for all points in S.*
3. *Compute the boundaries of the levels as follows: $Lev_1(S)$ is the convex hull of S; for every $j \geq 2$, the Delaunay circles C_j, defined by two points u, v of $Lay_j(S)$ and one point w of $Lay_{j-1}(S)$, determine the boundary between $Lev_j(S)$ and $Lev_{j+1}(S)$ which consists of the inner boundary of the union of Delaunay circles C_j (Figure 6).*

The running time of the algorithm is $O(n \log^2 n)$: $DT(S)$ can be computed in $O(n \log n)$ time, Step 2 only takes $O(n)$ time, and every boundary in Step 3 can be computed in $O(n_i \log^2 n_i)$ time where n_i is the number of Delaunay circles C_j considered in the corresponding layer (see [12] pg. 97). Taking into account that the total number of Delaunay circles is $O(n)$, Step 3 takes $O(n \log^2 n)$ time. Note that the expected time for Step 3 is $O(n \log n)$ [12], and therefore, the expected time for the entire algorithm is $O(n \log n)$.

4. Minimum weight structures

A notion of weight related to the concept of depth was associated in [2] to geometric structures consisting of segments (edges) with endpoints in S. The weight of a geometric structure admits two natural definitions.

Definition 4 *The total weight,* t-weight*, of a geometric structure, is the sum of the weights of its edges.*

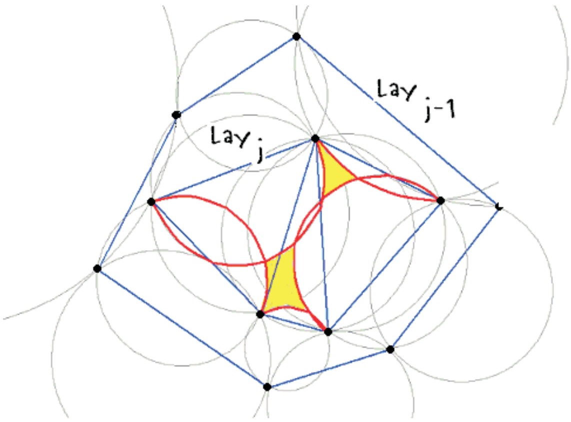

Fig. 6. The shaded region is $Lev_{j+1}(S)$.

Definition 5 *The worse weight,* w-weight, *of a geometric structure, is the weight of its heaviest edge.*

In this paper we focus on minimum t-weight structures.

4.1. *Minimum t-weight spanning trees*

An algorithm for constructing a minimum t-weight spanning tree ($MWST$), can be easily obtained by taking for each weight a spanning tree of the set of points of that weight and connecting whith an edge every pair of trees whose weight differs by one.

In [5] the following combinatorial result on the number of $MWST$ for a set of points is proved for convex depth. The result extends to Delaunay weight but does not extend to the case of location weight, since the vertices of $Lay_k(S)$ may be points not in S.

Proposition 6 *Let S be a set of n points and let C_i be the cardinality of layer i. For convex depth and Delaunay depth, the number of $MWST$ in S is*

$$f(C_1, \cdots, C_k) = \frac{C_1^{C_1} \cdot C_2^{C_2} \cdots C_{k-1}^{C_{k-1}} \cdot C_k^{C_k}}{C_1 \cdot C_k}.$$

The number of $MWST$ is strictly decreasing on the number of layers. The number of $MWST$ varies between n^{n-2}, attainable in the case of one layer, and 3^{n-2}, attainable when the number of layers is $k = (n+2)/3$. For a fixed number of layers, the maximum number of $MWST$ is attained when the last layer has one point and all the others have 3 points except possibly one (this layer cannot be the first except for the case of two layers).

4.2. *Minimum t-weight polygonizations*

Minimum weight polygonizations in the convex case are studied in [2], where next proposition is proved. This result also applies to each of the weights we are considering, the proof is similar in all the cases.

Proposition 7 *Let S be a set of n points with depth f. Every polygonization of S has weight greater or equal to $2f - 2$. The value $2f - 2$ is attained if and only if the sequence of depths of the boundary vertices is (circularly) unimodal.*

In [2] and [3] certain polygons are obtained with good computational properties. In particular, after the removal of any number of consecutive layers, the remaining set can be polygonized in constant time. These polygonizations, called *onion polygonizations*, achieve the minimum weight and furthermore, thanks to the convexity of the layers, are always constructible for any set of points. In the case of the Delaunay weight, we do not always have polygonizations of weight $2f-2$ (see [5]). This does not depend on the number of connected components of the Delaunay layers. Examples can be found of polygonizations with weight $2f - 2$ when the number of components is as big as possible (Figure 7).

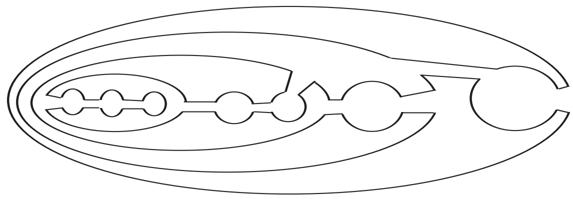

Fig. 7. A polygonization scheme with Delaunay weight $2f - 2$ for $f = 5$. The layers of the point set have 11 connected components.

Proposition 8 *The minimum Delaunay weight of any polygonization of a set of n points is never greater than n. There are examples of sets in which the weight of the minimum weight polygonization is $\Omega(n)$.*

4.3. *Minimum t-weight triangulations*

For the convex depth case we have obtained the following algorithm, which runs in $O(n \log n)$ time and generates triangulations that contain the convex layers.

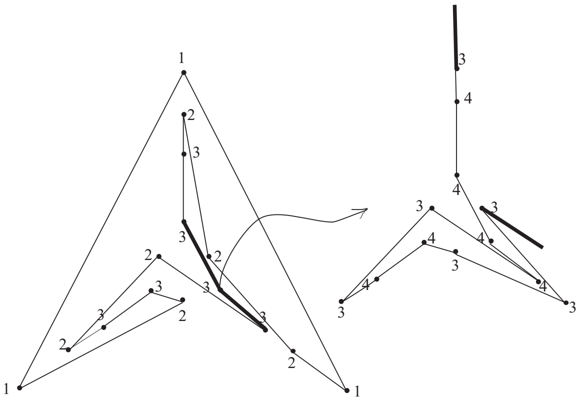

Fig. 8. Every polygonization of this point set has weight $\Omega(n)$.

Algorithm 2 *[5]* SMALL WEIGHT TRIANGULATION T OF A SET S OF n POINTS.

1. *Include in T all the edges of the convex layers.*
2. *For every convex layer find a minimal subpolygon P that contains the next inner layer; add to T the edges of P.*
3. *Complete T in any way by triangulating the remaining holes.*

It is worth noticing that there are other triangulations with the same weight that the triangulations obtained as output of the above algorithm, as shown in Figure 9.

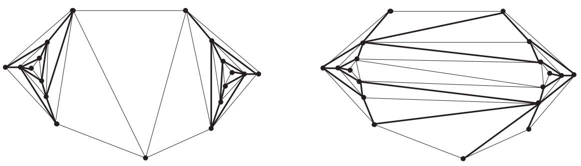

Fig. 9. Two triangulations with the same weight; the right one contains the convex layers. Solid edges have weight 1, light edges have weight 0.

We conjecture that the triangulations obtained using Algorithm 2 are always minimum weight triangulations. We have proved that this is the case for sets with two layers and also when all layers are triangles (except possibly the innermost one), yet a general proof remains elusive to us.

The study of the minimum t-weight triangulations for location depth and Delaunay depth is ongoing work. One would expect the Delaunay triangulation to have minimum Delaunay t-weight, but this is not always the case (see figure 10); however the Delaunay triangulation has minimum Delaunay w-weigt.

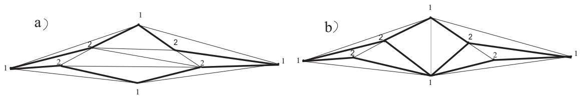

Fig. 10. a) Triangulation of minimum t-weight 8. b) Delaunay triangulation of the same point set as to the left, with t-weight 10, which is not minimum.

References

[1] G. Aloupis, C. Cortés, F. Gómez, M. Soss, G. Toussaint. Lower bounds for computing statistical depth. *Computational Statistics & Data Analysis* **40** (2002) 223–229.

[2] M. Abellanas, J. García, G.Hernández, F. Hurtado, O. Serra, J. Urrutia. Updating Polygonizations, *Computer Graphics Forum* **vol. 12, n.3** (1993) 143–152.

[3] M. Abellanas, J. García, G.Hernández, F. Hurtado, O. Serra, J. Urrutia. Onion Polygonizations, *Information Processing Letters* **vol. 57** (1996) 165–173.

[4] B. Chazelle. On the Convex Layers of a Planar Set, *IEEE Transactions on Information Theory* **vol. IT-31, n.4** (1985) 509–517.

[5] M. Claverol, Geometric Problems on Computational Morphology, *Ph. D Thesis in preparation.*

[6] R. Cole, M. Sharir, C.K. Yap. On K-hulls and related problems, *SIAM Journal of Computing* **vol. 16** (1987) 61–77.

[7] P.J. Green. Peeling Bivariate Data in: *Interpreting multivariate data.* John Wiley y Sons Ltd., (1981) 163–198.

[8] P. J. Hubert. Robust statistics: a review. *Ann. Math. Statistics* **43 n.3** (1972) 1041–1067.

[9] H. K. Miller, S. Ramaswami, P. Rousseeuw, T. Sellares, D. Souvaine, I. Streinu, A. Struyf. Efficient computation of location depth contours by methods of computational geometry, *Statistics and Computing* (2003).

[10] A. Okabe, B. Boots, K. Sugihara. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, John Wiley and Sons Ltd. Chichester (1992). 362–363.

[11] J.W. Tukey, Mathematics and the picturing of data, *Proceedings of the International Congress of Mathematicians* **2** (1975) 523–531.

[12] P. Agarwal, M. Sharir, Arrangements and its applications, in *Handbook of Comput. Geom.*, J.R. Sack and J. Urrutia eds., North-Holland 2000.