DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Geometric Measures of Data Depth

Greg Aloupis

Abstract. Several measures of data depth have been proposed, each attempting to maintain certain robustness properties. This paper lists the main approaches known to the computer science community. Properties and algorithms are mentioned, for computing the depth of a point and the location of the deepest point.

1. Introduction

Given a set of data points whose underlying probability distribution is unknown, or non-parametric and unimodal, the task is to estimate a point (location) which best describes the set. When dealing with such a problem, it is important to consider the issue of *robustness*: how much does the answer change if some of the data is perturbed? How easy is it to create a data set for which an estimator yields an absurd answer?

The mean is a classic example of a non-robust estimator: if one point is moved far from the rest, the mean will follow. To be literally extreme, if we move the "corrupt" point to infinity, the mean will also go to infinity. This example indicates that robustness is particularly important when there is a possibility that the data set contains "contaminated" or "corrupt" data. In contrast to the mean, at least half of the points in a set in $\mathbb{R}^1$ must be moved to infinity in order to force the *median* to do the same. This suggests a measure of robustness for *estimators of location*:

- The *breakdown point* is the proportion of data which must be moved to infinity so that the estimator will do the same.

Rousseeuw and Lopuhaa [**RL91**] discussed the breakdown point of various estimators. They gave credit to Hodges [**Hod67**] and Hampel [**Ham68**] for introducing the concept. Lopuhaa [**Lop92**] stated that it was Donoho and Huber [**DH83**] who suggested the definition given above, which is intended for finite data sets. Donoho and Gasko [**DG92**] also provided many results concerning the breakdown point, some of which are mentioned further on.

In R^1 the median has a breakdown point of $\frac{1}{2}$ and the mean has a breakdown point of $\frac{1}{n}$, where n is the number of data points. It has been shown [**RL91**] that the maximum breakdown point for any estimator is $\frac{1}{2}$, so in R^1 the median excels

1991 *Mathematics Subject Classification.* Primary 68W99, 62G35; Secondary 52C30, 68R05.

according to this robustness criterion. In fact, Bassett [**Bas91**] proved that in R^1 the median is the only estimator which possesses the properties of equivariance, maximum breakdown, and monotonicity (the median never moves in the opposite direction of a perturbation). Each of these desirable properties can be well defined in higher dimensions. A more difficult task is to find an estimator that excels, as does the median in R^1.

One of the first suggestions for generalizing the median to higher dimensions was to combine various one-dimensional medians taken in different directions. In 1902, Hayford suggested the *vector-of-medians* of orthogonal coordinates [**Hay02, Sma90**]. This is similar to finding the multivariate mean, and works well for other problems such as finding gradients of functions. Unfortunately, as Hayford was aware, the vector-of-medians depends on the choice of orthogonal directions. In fact, as Rousseeuw [**Rou85**] pointed out, this method may yield a median which is outside the convex hull of the data. Mood [**Moo41**] also proposed a joint distribution of univariate medians and used integration to find the multivariate median.

Soon after Hayford's median definition, Weber [**Web09**] proposed a new location estimator. Weber's estimator is the point which minimizes the sum of distances to all data points. Such a point defines the univariate median, so Weber's estimator may be considered to be a generalized multivariate median definition. This is discussed further in section 2.1.

In 1929 Hotelling [**Hot29**] described the univariate median as the point which minimizes the maximum number of data points on one of its sides. This notion was generalized to higher dimensions many years later by Tukey [**Tuk75**]. The Tukey median, or *halfspace* median, is perhaps the most widely studied and used multivariate median in recent years. Properties and algorithms for this median are given in section 2.2

A very intuitive definition for the univariate median is to continuously remove pairs of extreme data points. A generalization of this notion appeared by Shamos [**Sha76**] and by Barnett [**Bar76**], although Shamos stated that the idea originally belongs to Tukey. *Convex hull peeling* iteratively removes convex hull layers of points until a convex set remains. Section 2.3 contains more information.

In 1983, Oja [**Oja83**] introduced a definition for the multivariate median which generalizes the notion that the univariate median is the point with minimum sum of distances to all data points. However, Oja measured distance as one-dimensional volume. Thus the Oja median is a point for which the total volume of simplices formed by the point and appropriate subsets of the data set is minimum (see section 2.4).

In 1990, Liu [**Liu90**] proposed another multivariate definition, generalizing the fact that the univariate median is the point contained in the most intervals between pairs of data points. In higher dimensions, intervals (segments) are replaced by simplices. This median is described in section 2.5.

Gil, Steiger and Wigderson [**GSW92**] compared robustness and computational aspects of certain medians, although they imposed the restriction that the median must be one of the data points. They proposed a new definition for the multivariate median: for every data point, take the vector sum of all unit vectors to other data points. The median is any data point for which the length of the vector sum is less than or equal to one. The authors noted that the univariate median satisfies this condition, and claimed that their median is unique. It seems that this definition may lead to undefined cases in the restricted version, so perhaps it is more appropriate to allow the median to be located anywhere, not only at a data point.

Several other estimators of location exist, which are not necessarily generalizations of the univariate median. In *minimal volume ellipsoid covering*, by Rousseeuw [**Rou85**], the estimator is the center of a minimum volume ellipsoid which contains roughly half of the data. It is invariant to affine transformations and has a 50% breakdown point for a data set in general position. Rousseeuw also described another estimator with the same properties, introduced independently by Stahel [**Sta81**] and Donoho [**Don82**]. Their method involves finding a projection for each data point x in which x is most outlying. They then compute a weighted mean based on the results. Dyckerhoff, Koshevoy and Mosler [**DKM96**] defined *Zonoid depth*. Their measure locates the deepest point by considering nested convex regions that are defined by increasingly more restrictive linear combinations of the data. Gopala and Morin [**GM04**] give efficient algorithms for zonoid depth computation. Toussaint and Poulsen [**TP79**] proposed successive pruning of the *minimum spanning tree (MST)* of n points as a method of determining their center. To construct a MST, we connect certain data points with edges so that there exists a path between any two points and the sum of all edge lengths is minimized. This method seems to work well in general, although certain patterns of points may cause the location of the center to seem unnatural. Another graph-based approach was suggested by Green [**Gre81**]. Green proposed constructing a graph joining points adjacent in the *Delaunay triangulation* of the data set. Two data points $\{a, b\}$ in R^2 are adjacent if there exists a point in R^2 for which the closest data points are a and b. The Delaunay depth of a data point is the number of edges on the shortest path from the point to the convex hull. In general, the center of a graph has also been considered to be the set of points for which the maximum path length to reach another point is minimized [**Har69**].

Several other estimators which are based more on statistical methods exist. For example, *M-Estimators, L-Estimators* and *R-Estimators* were discussed by Huber [**Hub72**].

Another reason for using the median in R^1 is that it provides a method of ranking data. This concept can be generalized for the Oja, Liu and Tukey medians, as well as for convex hull peeling. Each of these medians maximizes/minimizes a certain *depth* function, and any point in R^d can be assigned a depth according to the particular function.

Robust estimators of location have been used for data description, multivariate confidence regions, p-values, quality indices, and control charts (see [**RR96**]). Applications of depth include hypothesis testing, graphical display [**MRR$^+$01**] and even voting theory [**RR99**]. Halfspace, hyperplane and simplicial depth are also closely related to regression [**RR99**]. For a recent account of statistical uses of depth, see [**LPS99**]. For a detailed introduction to robust estimators of location, the classic paper by Small [**Sma90**] is recommended. Finally, more links between computational geometry and statistics have been made by Shamos [**Sha76**].

The following sections list properties and algorithms that have been given, for some main depth measures that can be considered to be generalizations of the univariate median.

2. Computing Bivariate Medians

2.1. The L1 Median. Note: The median definition described below is known under several names in the literature. Here, the "neutral" term *L1 median* is used.

In 1909 Alfred Weber published his *theory on the location of industries* [**Web09**]. As a solution to a transportation cost minimization problem, Weber considered the point which minimizes the sum of Euclidean distances to all points in a given data set. In R^1 the median is such a point, so Weber's solution can be used as a generalization of the univariate median. The solution to Weber's problem, will be referred to in this section as the $L1$ median. Both Weber and Georg Pick, who wrote the mathematical appendix in Weber's book, were not able to find a solution for a set of more than three points in the plane. In the 1920's, the same definition was rediscovered independently by other researchers, who were primarily interested in finding the centers of population groups. Eells [**Eel30**] pointed out that for years the U.S. Census Bureau had been using the mean to compute the center of the U.S. population. Apparently they thought that the mean is the location which minimizes the sum of distances to all points in a set. In 1927, Griffin [**Gri27**] independently made the same observation, based on some earlier work by Eells. According to Ross [**Ros30**], Eells had discovered the error in 1926, but publication was delayed for bureaucratic reasons. Ross also printed a section of a paper by Gini and Galvani [**GG29**], translated from Italian to English. Gini and Galvani proposed the use of the $L1$ median and argued its superiority over the mean and the vector-of-medians approach (such as Hayford's).

None of the authors mentioned above were able to propose a method for computing the $L1$ median for sets of more than three points, even though the solution for three points had been known since the 17th century. According to Ross [**Ros30**] and Groß and Strempel [**GS98**], Pierre de Fermat first posed the problem of computing the point with minimum sum of distances to the vertices of a given triangle. Fermat discussed this problem with Evangelista Torricelli, who later computed the solution (see also [**dF34**] and [**LV19**]). Hence the $L1$ median for $n > 3$ is often referred to as the *generalized Fermat-Torricelli point*.

Scates [**Sca33**] suggested that the location of the $L1$ median for $n > 3$ cannot be computed exactly. At the same time, Galvani [**Gal33**] proved that the solution is a unique point in R^2 and higher dimensions. All algorithms to this date involve gradients or iterations and only find an approximate solution. One of the first such algorithms was by Gower [**Gow74**], who referred to the $L1$ median as the *mediancentre*. Groß and Strempel [**GS98**] discussed iteration and gradient methods for computing the $L1$ median.

The $L1$ median is known to be invariant to rotations of the data, but not to changes in scale. Another interesting property of the $L1$ median is that any measurement X of the data set can be moved along the vector from the median to X without changing the location of the median [**GG29**]. The breakdown point of the $L1$ median has been found to be $\frac{1}{2}$ [**Rou85**].

2.2. The Halfspace Median. In 1929, Hotelling [**Hot29**](see also [**Sma90**]) introduced his interpretation of the univariate median, while considering the problem of two competing ice-cream vendors on a one-dimensional beach. Hotelling claimed that the optimal location for the first vendor to arrive would be the location which minimized the maximum number of people on one side. This location happens to coincide with the univariate median. Tukey [**Tuk75**] is credited for generalizing this notion. The multivariate median based on this concept is usually referred to as the *Tukey* or *halfspace* median. Donoho and Gasko [**DG92**] were actually the first to define the multivariate halfspace median, since Tukey was primarily concerned with depth.

In order to define the multivariate halfspace median it is easier in fact to first consider the notion of *halfspace depth* for a query point θ with respect to a set S of n points in R^d. For each closed halfspace that contains θ, count the number of points in S which are in the halfspace. Take the minimum number found over all halfspaces to be the depth of θ. For example in R^2, place a line through θ so that the number of points on one side of the line is minimized. The halfspace median of a data set is any point in R^d which has maximum halfspace depth.

The halfspace median is generally not a unique point. However the set of points of maximal depth is guaranteed to be a closed, bounded convex set. The median is invariant to affine transformations, and can have a breakdown point between $\frac{1}{d+1}$ and $\frac{1}{3}$. If the data set is in general position, maximum depth is bounded below by $\lceil \frac{n}{d+1} \rceil$ and above by $\lceil \frac{n}{2} \rceil$ (see [**DG92**]).

Notice that any point outside of the convex hull of S has depth zero. To find the region of maximum depth in R^2, we can use the fact that its boundaries are segments of lines passing through pairs of data points. In other words, the vertices of the desired region must be intersection points of lines through pairs of data points. There are $O(n^4)$ intersection points, and it is a straightforward task to find the depth of a query point in $O(n^2)$ time (for each line defined by the query point and a data point, count the number of data points above and below the line). Thus in $O(n^6)$ time, we can find the intersection points of maximum depth. An improvement upon this was made by Rousseeuw and Ruts [**RR96**], who showed how to compute the halfspace depth of a point in $O(n \log n)$ time (for a matching lower bound on depth computation, see [**LS00**] or [**ACG$^+$02**]). Their algorithm therefore uses $O(n^5 \log n)$ time to compute the deepest point. Later they also gave a more complicated $O(n^2 \log n)$ version and provided implementations [**RR98**]. They seemed to be unaware that before this, Matoušek [**Mat91**] had presented an $O(n \log^5 n)$ algorithm for computing the halfspace median. Matoušek showed how to compute any point with depth greater than some constant k in $O(n \log^4 n)$ time and then used a binary search on k to find the median. Matoušek's algorithm was improved by Langerman and Steiger [**LS00**], whose algorithm computes the median in $O(n \log^4 n)$ time. A further improvement to $O(n \log^3 n)$ appeared in Langerman's Ph.D. thesis [**Lan01**] (see also [**LS03**]) . A recent implementation by Miller et al [**MRR$^+$01**] uses $O(n^2)$ time and space to find all depth contours, after which it is easy to compute the median. They claim that Matoušek's algorithm is too complicated for practical purposes. Recently, Chan [**Cha04**] gave a randomized algorithm to compute the halfspace median, with time complexity $O(n \log n)$.

Another interesting location based on halfspace depth is the *centerpoint*, which is a point with depth at least $\lceil \frac{n}{d+1} \rceil$. Gil, Steiger and Wigderson [**GSW92**] stated that the centerpoint could be used as a multivariate median since it coincides with the median in R^1 . Edelsbrunner [**Ede87**] showed that a centerpoint can be found for any data set. The number $\lceil \frac{n}{d+1} \rceil$ arises from Helly's theorem (see [**RH99a**] for a more detailed account). Cole, Sharir and Yap [**CSY87**] were able to compute the centerpoint in $O(n \log^5 n)$ time. Matoušek improved this result by setting k

equal to $\lceil \frac{n}{d+1} \rceil$ in his algorithm, thus obtaining the centerpoint in $O(n \log^4 n)$ time. Finally, Jadhav and Mukhopadhyay [**JM94**] gave an $O(n)$ algorithm to compute the centerpoint.

2.3. Convex Hull Peeling and Related Methods. Perhaps the most intuitive visual interpretation of the univariate median is the idea of peeling away outlying data: we throw out the smallest and largest values recursively, until we are left with one or two. We can easily generalize this idea to higher dimensions. One way is to iteratively peel away convex hull layers of the data set, until a convex set remains. As in the univariate case, if more than one point remains, we take the mean. Most of the credit for introducing this concept is given to Shamos [**Sha76**] and Barnett [**Bar76**]. However, Shamos stated that the idea originally belongs to Tukey.

A brute-force algorithm for convex hull peeling takes $O(n^2 \log n)$ time in R^2: The convex hull may be computed in $O(n \log n)$ time [**Gra72**] (for matching lower bounds see [**Avi82, Yao81**]). In the worst case it is possible that only three data points will be removed in each iteration, which leads to $O(n)$ convex hull calculations. The task can be done easily in $O(n^2)$ time by slightly modifying the Jarvis "gift-wrapping" convex hull algorithm [**Jar73**]. This algorithm takes $O(hn)$ time to compute the convex hull, where h is the number of vertices on the hull. Once the hull is computed, the modified algorithm continues with the remaining points. This modification was first proposed by Shamos [**Sha76**]. Later Overmars and van Leeuwen [**OvL81**] designed a data structure which maintains the convex hull of a set of points after the insertion/deletion of arbitrary points, with a cost of $O(\log^2 n)$ time per insertion/deletion. This provides an $O(n \log^2 n)$ time method for convex hull peeling. Finally, Chazelle [**Cha85**] improved this result by ignoring insertions and taking advantage of the structure in the sequence of deletions in convex hull peeling. Chazelle's algorithm uses $O(n \log n)$ time to compute all convex layers and the depth of any point.

A technique similar to convex hull peeling was proposed by Titterington [**Tit78**]. He proposed iteratively peeling minimum volume ellipsoids containing the data set. Both methods of peeling data can have very low breakdown points. Donoho and Gasko [**DG92**] proved that the breakdown point of these methods cannot exceed $\frac{1}{d+1}$ in R^d and stated that the breakdown point seems to always approach 0 as n approaches infinity.

Another ellipsoid method was proposed by Rousseeuw [**Rou85**]. His median is the center of the minimum volume ellipsoid that covers approximately half of the data points. Rousseeuw proves that this estimator is invariant under affine transformations and has a 50% breakdown point for data in general position.

2.4. The Oja Simplex Median. Consider *d+1* points in R^d. These points form a simplex, which has a d-dimensional volume. For example, in R^3 four points form a tetrahedron, and in R^2 three points form a triangle whose area is "2-dimensional volume". Now consider a data set in R^d for which we seek the median. Oja proposed the following measure for a query point θ in R^d [**Oja83**].

- for every subset of d points from the data set, form a simplex with θ.
- sum together the volumes of all such simplices.

We can call this sum *Oja depth*. The Oja simplex median is any point $\hat{\mu}$ in R^d with minimum Oja depth.

Since one-dimensional volume is length, the Oja median reduces to the standard univariate median. In R^1, the Oja median minimizes the sum of distances to all data points, as does the $L1$ median. Unlike the $L1$ median, the Oja median is not guaranteed to be a unique point in higher dimensions. However Oja mentions that the points of maximum depth form a convex set and that to compute such a point in R^2 it suffices to consider only intersection points of lines formed by pairs of data points.

An important feature of the Oja median is that it is invariant to affine transformations. However, data sets may be constructed for which the breakdown point approaches zero. Notice that if the data does not "span" the dimension of the space that it is in (ex: data on a line in R^2 or data on a plane in R^3, then it is possible to find simplices with zero volume even at infinity. This seems like an unrealistic case for any application though. A nicer example for R^2 is given by Niinimaa, Oja and Tableman [**NOT90**], although the example used resembles a bimodal distribution and the corrupt points must all be moved to infinity in specific directions. This raises a question about forming a proper, practical definition of the breakdown point in higher dimensions. It seems that a strong definition should give the same answer regardless of *where* at infinity the corrupt points are placed.

The straightforward method of calculating the Oja median is to compute the depth of each intersection point. In R^2 this can be done in $O(n^6)$ time: for each of the $O(n^4)$ intersection points there are $O(n^2)$ triangles for which the area must be computed, and the area of a triangle can be computed in constant time. The same upper bound holds for an algorithm of Niinimaa, Oja and Nyblom [**NON92**]. The algorithm selects a line between two data points and computes the gradient of Oja depth at each intersection point along the line until one such point is found to be a minimum. A new line from this point is then selected, and the procedure is repeated. The gradient is computed in $O(n^2)$ time and it is possible that all intersection points will be visited. Rousseeuw and Ruts [**RR96**] provided a technique for computing the Oja gradient in $O(n \log n)$ time and stated that the same algorithm can be used to find the median in $O(n^5 \log n)$ time. Aloupis, Soss and Toussaint used the convexity of the depth function in order to perform a type of binary search that computed the median, first in $O(n^3 \log^2 n)$ time [**AST01**], and then in $O(n^3 \log n)$ time [**Alo01**]. With the incorporation of a powerful optimization technique for arrangements of lines, developed by Langerman and Steiger [**LS03**], the time complexity of computing the Oja median was improved to $O(n \log^3 n)$ time [**ALST03**]. An $\Omega(n \log n)$ lower bound for computing the Oja depth of a point is to appear soon [**AM04**].

2.5. The Simplicial Median. Another interpretation of the univariate median is that it is the point which lies inside the greatest number of intervals constructed from the data points. Liu generalized this idea as follows [**Liu90**]: the simplicial median in R^d is a point in R^d which is contained in the most simplices formed by subsets of *d+1* data points. The simplicial depth of a point in R^d is the number of simplices which contain the point. Liu's original definition involves closed simplices, although later in [**Liu95**] she repeats the definition using open simplices. Unless mentioned otherwise, when we refer to simplicial depth or the simplicial median, we will use the original definition (a point on the boundary of a simplex is inside the simplex).

Liu showed that the simplicial median is invariant to affine transformations. However, not much information is known about the breakdown point. Gil, Steiger and Wigderson [**GSW92**] constructed a set of points for which the *data point* of maximum simplicial depth could be moved arbitrarily far away with only a few corrupting points. This does not necessarily imply anything for the simplicial median of Liu.

Although Liu was the first to define simplicial depth and introduce the notion to the statistical community, this concept had already been considered before. Boros and Füredi [**BF84**] proved that for a set of n points in general position in R^2 there always exists a point contained in at least $\frac{n^3}{27} + O(n^2)$ open triangles formed by the points. This implies that the depth of the simplicial median in R^2 is $\Theta(n^3)$. Bárány [**Bár82**] showed that in R^d there always exists a point contained in

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

simplices.

A straightforward method of finding the simplical median in R^2 is to partition the plane into cells which have segments between points as boundaries. First, notice that every point within a given cell has equal depth. Furthermore, a point on a boundary between two cells must have depth at least as much as any adjacent interior point. Similarly, an intersection point (where more than two cells meet) must have depth at least as much as any adjacent boundary point. Therefore by determining how many triangles contain each intersection point we can find the simplicial median.

If a set of n line segments has k intersection points, they can be reported in $O(n \log n + k)$ time and $O(n)$ space with a line sweeping technique of Balaban [**Bal95**]. In the case of simplicial depth, we have $O(n^2)$ line segments formed between pairs of data points, and unfortunately k is $\Theta(n^4)$ [**SW94**]. Thus the algorithm of Balaban takes $O(n^4)$ time and $O(n^2)$ space for our purposes, so it is better to use brute-force to compute each intersection point. The total time for this is $O(n^4)$ and the space used is $O(n)$. Since there are $O(n^3)$ triangles formed by n points, a brute-force calculation of the simplicial median uses $O(n^7)$ time and $O(n)$ space. Khuller and Mitchell [**KM89**] proposed an $O(n \log n)$ time algorithm to compute the number of triangles formed by triples of a data set which contain a query point in R^2. Their algorithm came just before Liu's definition. Gil, Steiger and Wigderson [**GSW92**] independently proposed the same algorithm and considered the simplicial median to be the *data point* with maximum depth. A third version of this algorithm appeared later, but also independently, by Rousseeuw and Ruts [**RR96**]. They were the first to point out that by computing the depth of each intersection point, the simplicial median may be found in $O(n^5 \log n)$ time. A matching lower bound for depth computation exists [**ACG$^+$02**]. However, Aloupis, Soss, Langerman and Toussaint produced a faster algorithm for computing the median, by processing each intersection point along a segment in constant time, after some preprocessing. The total running time is $O(n^4)$. Clearly there is still much room for improvement, though the simplicial depth function has much less structure than other depth functions (such as Oja and halfspace). It is not difficult to see that the regions of maximum depth may be disconnected.

For R^3 Gil, Steiger and Wigderson [**GSW92**] proposed an algorithm to compute the simplicial depth of a point in $O(n^2)$. Cheng and Ouyang [**CO98**] discovered a slight flaw in this algorithm and provided a corrected version. They also provided a $O(n^4)$ time algorithm for R^4 and commented that for higher dimensions the brute-force algorithm becomes better. They mention that an algorithm suggested by Rousseeuw and Ruts [**RR96**] for higher dimensions seems to have some discrepancies.

2.6. Hyperplane Depth. The most recent notion of depth for a point in R^d is by Rousseeuw and Hubert [**RH99a**], although their definition of depth is not with respect to a set of points. They defined the *hyperplane depth* of a point θ with respect to a set of n hyperplanes to be the minimum number of hyperplanes that a ray emanating from θ must cross.

The authors remarked that the point with maximum hyperplane depth (which they call the *hyperplane median*) can be seen as the "deepest" or "most central" point in the arrangement of hyperplanes. They proved that in R^1 such a point is in fact the median. They conjectured that in R^d there always exists a point with depth greater than $\lceil \frac{n}{d+1} \rceil$ and proved this for $d = 1$ and $d = 2$. They also provided an $O(n)$ time algorithm to compute such a point in R^2. Furthermore they showed that maximum hyperplane depth cannot be greater than $\lfloor \frac{n+d}{2} \rfloor$ if the hyperplanes are in general position (no two hyperplanes parallel and no $d+1$ hyperplanes concurrent), and that the hyperplane median is invariant to affine transformations.

Langerman and Steiger [**LS00**] provided an $O(n \log n)$ time algorithm to compute the hyperplane median of n hyperplanes in R^2. If a query point is on a hyperplane, Langerman and Steiger do not count this plane as crossing a ray from the query point. They also provided a matching lower bound and mentioned previous results such as an $O(n^3)$ time algorithm by Rousseeuw and Hubert [**RH99b**] and an $O(n \log^2 n)$ time algorithm by van Kreveld et al [**vKMR$^+$99**]. Amenta et al [**ABET00**] proposed an $O(n^d)$ time algorithm which constructs the arrangement of the n hyperplanes and uses a breadth first search to locate the hyperplane median in R^d. They also proved the conjecture of Rousseeuw and Hubert.

Although hyperplane depth is equivalent to the median in R^1, no generalization has been made for the multivariate median of a set of points in R^d. The concept of a ray intersecting a minimum number of hyperplanes may be used as follows: given a set S of n points in R^d, construct the set H of hyperplanes formed by subsets of d points in S. Find the point μ with maximum hyperplane depth in H.

References

[ABET00] N. Amenta, M. Bern, D. Eppstein, and S. Teng, *Regression depth and center points*, Discrete and Computational Geometry **123** (2000), no. 3, 305–323.

[ACG$^+$02] G. Aloupis, C. Cortes, F. Gomez, M. Soss, and G. Toussaint, *Lower bounds for computing statistical depth*, Computational Statistics and Data Analysis **40** (2002), 223–229.

[Alo01] G. Aloupis, *On computing geometric estimators of location*, M.Sc. thesis, School of Computer Science, McGill University, Montreal, Canada, March 2001.

[ALST03] G. Aloupis, S. Langerman, M. Soss, and G. Toussaint, *Algorithms for bivariate medians and a Fermat-Torricelli problem for lines*, Computational Geometry: Theory and Applications **26** (2003), no. 1, 69–79.

[AM04] G. Aloupis and E. McLeish, *A lower bound for computing Oja depth*, manuscript, School of Computer Science, McGill University, 2004.

[AST01] G. Aloupis, M. Soss, and G. Toussaint, *On the computation of the bivariate median and a Fermat-Torricelli problem*, Technical Report SOCS-01.2, School of Computer Science, McGill University, February 2001.

[Avi82] D. Avis, *On the complexity of finding the convex hull of a set of points*, Discrete Applied Math **4** (1982), 81–86.

[Bal95] I. Balaban, *An optimal algorithm for finding segment intersections*, Proc. 11th Annu. ACM Sympos. Comput. Geom., 1995, pp. 211–219.

[Bar76] V. Barnett, *The ordering of multivariate data*, Journal of the Royal Statistical Society ser.A **139** (1976), 318–355.

[Bár82] I. Bárány, *A generalization of Carathéodory's theorem*, Discrete Math **40** (1982), 141–150.

[Bas91] G. Bassett, *Equivariant monotonic, 50% breakdown estimators*, The American Statistician **45** (1991), no. 2, 135–137.

[BF84] E. Boros and Z. Füredi, *The maximal number of covers by the triangles of a given vertex set on the plane*, Geom. Dedicata **17** (1984), 69–77.

[Cha85] B. Chazelle, *On the convex layers of a planar set*, IEEE Transactions on Information Theory **IT-31** (1985), no. 4, 509–517.

[Cha04] T. Chan, *An optimal randomized algorithm for maximum Tukey depth*, Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms, 2004, pp. 430–436.

[CO98] A. Cheng and M. Ouyang, *On algorithms for simplicial depth*, Technical Report dcs-tr-368, Department of Computer Science, Rutgers University, 1998.

[CSY87] R. Cole, M. Sharir, and K. Yap, *On k-hulls and related problems*, SIAM J. Comput. **16** (1987), no. 1, 61–77.

[dF34] P. de Fermat, *Abhandlungen über maxima und minima*, Ostwalds Klassiker der exacten Wissenschaften, M. Miller (ed) **238** (1934).

[DG92] D. Donoho and M. Gasko, *Breakdown properties of location estimates based on halfspace depth and projected outlyingness*, Annals of Statistics **20** (1992), 1803–1827.

[DH83] D. Donoho and P. Huber, *The notion of breakdown point*, A Festschrift for Erich L. Lehmann (Belmont, California) (P. Bickel, K. Doksum, and J. Hodges, eds.), Wadsworth International Group, 1983, pp. 157–184.

[DKM96] R. Dyckerhoff, G. Koshevoy, and K. Mosler, *Zonoid data depth: Theory and computation*, COMPSTAT '96 - Proceedings in Computational Statistics (A. Prat, ed.), Physica, Heidelberg, 1996, pp. 235–240.

[Don82] D. Donoho, *Breakdown properties of multivariate location estimators.*, Ph.D. thesis, Harvard University, 1982.

[Ede87] H. Edelsbrunner, *Algorithms in combinatorial geometry*, Springer-Verlag, Berlin, 1987.

[Eel30] W. Eells, *A mistaken conception of the center of population*, Journal of the American Statistical Association **25** (1930), 33–40.

[Gal33] L. Galvani, *Sulla determinazione del centro di gravità e del centro mediano di una popolazione*, Metron **11** (1933), no. 1, 17–47.

[GG29] C. Gini and L. Galvani, *Di talune estensioni dei concetti di media a caratteri qualitivi*, Metron **8** (1929), 3–209.

[GM04] H. Gopala and P. Morin, *Algorithms for bivariate zonoid depth*, submitted for publication (2004).

[Gow74] J. Gower, *The mediancentre*, Applied Statistics **23** (1974), no. 3, 466–470.

[Gra72] R. Graham, *An efficient algorithm for determining the convex hull of a finite planar set*, Information Processing Letters **7** (1972), 175–180.

[Gre81] P. J. Green, *Peeling bivariate data*, Interpreting Multivariate Data (New York) (V. Barnett, ed.), Wiley, 1981.

[Gri27] F. Griffin, *Abstract: Points of minimum travel for a distributed population*, Bulletin of the American Mathematical Society **33** (1927), 516.

[GS98] C. Groß and T. Strempel, *On generalizations of conics and on a generalization of the Fermat-Torricelli problem*, American Mathematical Monthly **105** (1998), no. 8, 732–743.

[GSW92] J. Gil, W. Steiger, and A. Wigderson, *Geometric medians*, Discrete Mathematics **108** (1992), 37–51.

[Ham68] R. Hampel, *Contributions to the theory of robust estimation.*, Ph.D. thesis, University of California, Berkeley, 1968.

[Har69] F. Harary, *Graph theory*, Addison-Wesley, Reading, Mass., 1969.

[Hay02] J. Hayford, *What is the center of an area, or the center of a population?*, Journal of the American Statistical Association **8** (1902), 47–58.

[Hod67] J. Hodges, *Efficiency in normal samples and tolerance of extreme values for some estimates of location*, 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1967, pp. 163–186.

[Hot29] H. Hotelling, *Stability in competition*, Economic Journal **39** (1929), 41–57.

[Hub72] P. Huber, *Robust statistics: A review*, The Annals of Mathematical Statistics **43** (1972), no. 3, 1041–1067.

[Jar73] R. Jarvis, *On the identification of the convex hull of a finite set of points in the plane*, Information Processing Letters **2** (1973), 18–21.

[JM94] S. Jadhav and A. Mukhopadhyay, *Computing a centerpoint of a finite planar set of points in linear time*, Discrete and Computational Geometry **12** (1994), 291–312.

[KM89] S. Khuller and J. Mitchell, *On a triangle counting problem*, Information Processing Letters **33** (1989), 319–321.

[Lan01] S. Langerman, *Algorithms and data structures in computational geometry*, Ph.D. thesis, Department of Computer Science, Rutgers University, New Brunswick, 2001.

[Liu90] R. Liu, *On a notion of data depth based upon random simplices*, The Annals of Statistics **18** (1990), 405–414.

[Liu95] ———, *Control charts for multivariate processes*, Journal of the American Statistical Association **90** (1995), no. 432, 1380–1387.

[Lop92] H. Lopuhaa, *Highly efficient estimator of multivariate location*, The Annals of Statistics **20** (1992), no. 1, 398–413.

[LPS99] R. Liu, J. Parelius, and K. Singh, *Multivariate analysis of data depth: descriptive statistics and inference*, Annals of Statistics **27** (1999), no. 3, 783–858.

[LS00] S. Langerman and W. Steiger, *The complexity of hyperplane depth in the plane*, Japan Conference on Discrete and Computational Geometry, November 2000.

[LS03] ———, *Optimization in arrangements*, 20th International Symposium on Theoretical Aspects of Computer Science, 2003, pp. 50–61.

[LV19] G. Loria and G. Vassura (eds.), *Opere di Evangelista Torricelli*, vol. 1, Faenza, 1919.

[Mat91] J. Matoušek, *Computing the center of planar point sets*, Computational Geometry: Papers from the DIMACS special year (J. Goodman, R. Pollack, and W. Steiger, eds.), vol. 6, American Mathematical Society, 1991, pp. 221–230.

[Moo41] A. Mood, *On the joint distribution of the medians in samples from a multivariate population*, The Annals of Mathematical Statistics **12** (1941), 268–278.

[MRR+01] K. Miller, S. Ramaswami, P. Rousseeuw, T. Sellarès, D. Souvaine, I. Streinu, and A. Struyf, *Fast implementation of depth contours using topological sweep*, Proc. 12th Symposium on Discrete Algorithms (SODA) (Washington D.C.), 2001.

[NON92] A. Niinimaa, H. Oja, and J. Nyblom, *Algorithm AS 277: The Oja bivariate median*, Applied Statistics **41** (1992), 611–617.

[NOT90] A. Niinimaa, H. Oja, and M. Tableman, *The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version*, Statistics and Probability Letters **10** (1990), 325–328.

[Oja83] H. Oja, *Descriptive statistics for multivariate distributions*, Statistics and Probability Letters **1** (1983), 327–332.

[OvL81] M. Overmars and J. van Leeuwen, *Maintenance of configurations in the plane*, J. Comput. and Syst. Sci. **23** (1981), 166–204.

[RH99a] P. Rousseeuw and M. Hubert, *Depth in an arrangement of hyperplanes*, Discrete Computational Geometry **22** (1999), 167–176.

[RH99b] ———, *Regression depth*, Journal of the American Statistical Association **94** (1999), 388–402.

[RL91] P. Rousseeuw and H. Lopuhaa, *Breakdown points of affine equivariant estimators of multivariate location and covariance matrices*, The Annals of Statistics **119** (1991), no. 1, 229–248.

[Ros30] F. Ross, *Editor's note on the center of population and point of minimum travel*, Journal of the American Statistical Association **25** (1930), 447–452.

[Rou85] P. Rousseeuw, *Multivariate estimation with high breakdown point*, Mathematical Statistics and Applications (Dordrecht) (W. Grossman, G. Pflug, I. Vincze, and W. Wertz, eds.), vol. B, Reidel Publishing Company, 1985, pp. 283–297.

[RR96] P. Rousseeuw and I. Ruts, *Bivariate location depth*, Applied Statistics **45** (1996), 516–526.

[RR98] ———, *Constructing the bivariate Tukey median*, Statistica Sinica **8** (1998), 828–839.

[RR99] ———, *The depth function of a population distribution*, Metrika **49** (1999), no. 3, 213–244.

[Sca33] D. Scates, *Locating the median of the population in the United States*, Metron **11** (1933), no. 1, 49–65.

[Sha76] M. Shamos, *Geometry and statistics: Problems at the interface.*, Recent Results and New Directions in Algorithms and Complexity (J. Traub, ed.), Academic Press, 1976, pp. 251–280.

[Sma90] C. Small, *A survey of multidimensional medians*, International Statistical Review **58** (1990), 263–277.

[Sta81] W. Stahel, *Robust estimation: Infinitesimal optimality and covariance matrix estimators*, Ph.D. thesis, ETH, Zurich, 1981.

[SW94] E. Scheinerman and H. Wilf, *The rectilinear crossing number of a complete graph and Sylvester's "four point problem" of geometric probability*, American Mathematical Monthly **101** (1994), no. 10, 939–943.

[Tit78] D. Titterington, *Estimation of correlation coefficients by ellipsoidal trimming*, Applied Statistics **27** (1978), 227–234.

[TP79] G. T. Toussaint and R. S. Poulsen, *Some new algorithms and software implementation methods for pattern recognition research*, Proc. IEEE Computer Society Conf. on Computer Software, 1979, pp. 55–63.

[Tuk75] J. Tukey, *Mathematics and the picturing of data*, Proceedings of the International Congress of Mathematicians (Vancouver), 1975, pp. 523–531.

[vKMR+99] M. van Kreveld, J. Mitchell, P. Rousseeuw, M. Sharir, J. Snoeyink, and B. Speckmann, *Efficient algorithms for maximum regression depth*, Proceedings of the 15th Symposium on Computational Geometry, ACM, 1999, pp. 31–40.

[Web09] A. Weber, *Über den standort der industrien, Tubingen. English translation by C. Friedrich (1929): Alfred Weber's theory of location of industries*, University of Chicago Press, 1909.

[Yao81] A. Yao, *A lower bound to finding convex hulls*, Journal of the ACM **28** (1981), no. 4, 780–787.

School of Computer Science, McGill University
E-mail address: athens@cgm.cs.mcgill.ca