Discussion of "Location-Scale Depth" by I. Mizera and C. H. Müller

David Eppstein*

May 21, 2004

Mizera's previous work [13] introduced *tangent depth*, a powerful method for defining robust statistics. In this framework, one measures the quality of a putative fit ϑ to a sample point z_i by a *criterial function* $F_i(\vartheta)$. We say that ϑ is a *nonfit* if there is another fit φ with $F_i(\varphi) < F_i(\vartheta)$ for all i. The *depth* of ϑ is then defined as the smallest value d such that d of the sample points can be removed causing ϑ to be a nonfit for the remaining data. This method actually leads to two notions of depth: *global depth* is as defined above, while *tangent depth* is the limiting case of global depth as φ is restricted to lie within a small neighborhood of ϑ. However, when the criterial functions are strictly quasiconvex (in an appropriate parametrization of ϑ) the two notions of depth coincide. For location fitting problems, using as the criterial function the distance from z_i (or any monotonic function of the distance) recovers the classical notion of Tukey depth. For linear regression, using as the criterial function the residual distance recovers the notion of *regression depth* introduced by Rousseeuw and Hubert [14].

This interesting new paper by Mizera and Müller applies the tangent depth technique to problems of fitting a set of sample points with a value that encodes both location and scale. To do this, the authors define fits and associated criterial functions that arise naturally from statistics. The fits are chosen to be scaled translates of a symmetric probability distribution (e.g. Student, Gaussian, logistic, slash, or Laplace distributions) and the criterial functions are the likelihoods according to these distributions. That is, this method posits a signal in the form of samples from an unknown distribution of known type, hidden by noise in the form of a relatively small number of arbitrary outliers, and uses depth-based methods to sift the signal from the noise. The results are especially elegant for the Gaussian and Student distributions: the depth of a fit in $\mathbb{R}^d$ can be viewed as equivalent to Tukey depth in hyperbolic space

*School of Information & Computer Science, Univ. of California, Irvine, CA 92697-3425, USA, eppstein@ics.uci.edu. Work supported in part by NSF grant CCR-9912338.

$\mathbb{H}^{d+1}$. By standard techniques for modeling hyperbolic space in Euclidean spaces, all the previous machinery of Tukey depth can be brought to bear on this new depth problem.

In this discussion we examine Mizera and Müller's results from the viewpoint of discrete and computational geometry: how can their definitions be reformulated in a combinatorial way that would be more familiar to computational geometers, and what is known and unknown about the algorithmic complexity of location-scale depth problems? Along the way we visit some related topics from the computational geometry literature: parabolic lifting, optimal Möbius transformation, and finite element mesh partitioning.

Throughout, we define depth in terms of cardinalities of sets of sample points, whereas Mizera and Müller define depth in terms of the relative proportions of such sets when viewed as subsets of the whole sample set; however this difference is of no consequence. More significantly, Mizera and Müller treat only the case of univariate (one-dimensional) sample data, while we take their equivalence to hyperbolic Tukey depth as a given and consider its implications in higher dimensions.

1 Combinatorial Interpretation

Part of the reason for the continued interest from the computational geometry community in Tukey and regression depths, and more generally in depth-based statistical inference, is that these depths are easily understood without much statistical background knowledge. The objects being sought are themselves simple (points and lines or hyperplanes respectively) and their depths can be defined very simply in terms of counting incidences between simple geometric objects such as points and halfplanes. The Tukey depth of a point μ among a set of sample points in $\mathbb{R}^d$, for instance, is the minimum number of sample points contained in any closed halfspace the boundary of which contains μ. The regression depth of a line l among sample points in the plane is the minimum number of sample points contained in any *double wedge* bounded by l and any vertical line; such double wedges are familiar in computational geometry as the projective duals of rays.

Can we come up with a similar definition of location-scale depth, where the object we seek is a simple geometric figure carrying location and scale information (such as a circle) and the depth is defined as the minimum number of sample points in a shape drawn from some simple family? For tangent depth, in general, the answer is yes: the depth of ϑ is the minimum number of sample points in a subset $S_\varphi = \{z \mid F_z(\varphi) < F_z(\vartheta)\}$, for φ arbitrarily close to ϑ. The question is how to describe in

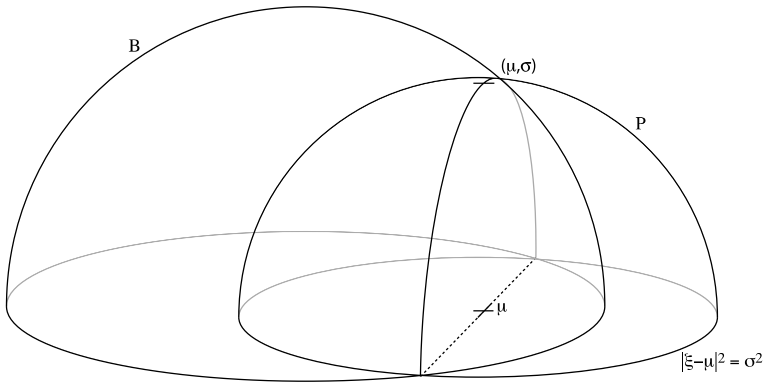

Figure 1: Any hyperbolic plane B through μ, σ meets the plane P with μ, σ as apex in diametrically opposed infinite points.

a more explicit way these subsets S_φ.

A fit for Student or Gaussian depth can be described as a pair (μ, σ) where μ belongs to the sample space and σ can be interpreted as the standard deviation of the Gaussian. As Mizera and Müller show, these pairs (μ, σ) can also interpreted as coordinates for a Poincaré halfspace model of the hyperbolic plane, and in this model the Student depth becomes the hyperbolic Tukey depth of the sample points $(x_i, 0)$. Interpreted hyperbolically, the sample points represent infinite points on the boundary of the halfspace model. One can therefore define location-scale depth in terms of numbers of incidences between samples and hyperbolic halfspaces, but we would like a more intrinsic description involving geometric figures in the original sample space. Such a description is provided for multivariate data by the following characterization:

Theorem 1 *Let $\mu \in \mathbb{R}^2$ and $\sigma \in \mathbb{R}$ be the center and radius of a circle $C = \{\xi : |\xi - \mu|^2 = \sigma^2\}$, and let there be given a set of samples in $\mathbb{R}^2$. Then the hyperbolic Tukey depth of (μ, σ) for the given sample set equals the minimum number of samples in a closed disk or disk complement in $\mathbb{R}^2$ bounded by a circle passing through two diametrically opposed points of C.*

Proof If we view $\mathbb{R}^2$ as the boundary of a halfspace model of a hyperbolic space $\mathbb{H}^3$, C is the set of infinite limit points of a hyperbolic plane P, which is modeled by a hemisphere with C as its boundary and with the point μ, σ as its apex (Figure 1). The hyperbolic Tukey depth of μ, σ is the minimum number of sample points in a subset of $\mathbb{H}^3$ bounded by a hemisphere B containing μ, σ. The intersection of this subset with $\mathbb{R}^2$ is a disk or complement of a disk bounded by a circle $\mathcal{O}$. Since P and B intersect in a hyperbolic line l, modeled as a semicircle that contains the apex μ, σ, C and $\mathcal{O}$ intersect in the endpoints of l.

Projecting l vertically onto $\mathbb{R}^d$ produces a line segment, which contains the point μ at the center of C, so the two intersection points are diametrically opposed.

Conversely, if $\mathcal{O}$ is the boundary of any disk or disk complement that contains few sample points, where $\mathcal{O}$ passes through diametrically opposed points of C, then the hemisphere bounded by $\mathcal{O}$ provides a hyperbolic plane P containing μ, σ, and the disk or disk complement bounded by $\mathcal{O}$ is the set of limit points of a hyperbolic halfspace bounded by P. Thus any circle of the form provided by the lemma leads to a halfspace lower bounding the Tukey depth of μ, σ, and vice versa, so the result follows. □

More generally, for sample points in $\mathbb{R}^d$, $d > 2$, the same technique shows that the hyperbolic Tukey depth of (μ, σ) is the minimum number of points in a ball or ball complement the boundary of which passes through an equatorial $(d-1)$-sphere of the sphere $\{\xi : |\xi - \mu|^2 = \sigma^2\}$.

2 Parabolic Klein Model

Mizera and Müller use hyperbolic geometry to convert a problem involving round objects (circles or radially symmetric distributions) to one involving more mathematically tractable flat objects (hyperbolic planes). However, it is possible to achieve the same linearization effect while remaining in a Euclidean instead of hyperbolic geometry. This can be done e.g. by stereographic projection to a sphere [3] but the technique more commonly used in the computational geometry literature is the *parabolic lifting map* [6] that maps points μ in $\mathbb{R}^d$ to $p(\mu) = (\mu, |\mu|^2)$ on a paraboloid in $\mathbb{R}^{d+1}$. If C is a sphere in $\mathbb{R}^d$, then $p(C)$ is the intersection of the paraboloid with a hyperplane in $\mathbb{R}^{d+1}$, and conversely every hyperplane that has a nonempty intersection with the paraboloid intersects it in the lifted image of a sphere.

Mizera and Müller allude to this possibility briefly but dismiss it, writing

> It is tempting to think that what we deal with here is just the bivariate location depth with respect to the datapoints lifted on a parabola—but one has to keep in mind that τ_i depend on μ and σ, so when the parameters change, the position of lifted points changes too.

However, the set of points in $\mathbb{R}^{d+1}$ above the paraboloid can be viewed as a parabolic version of the Klein model of hyperbolic geometry, in which hyperbolic planes are modeled by Euclidean planes that intersect with the paraboloid. So,

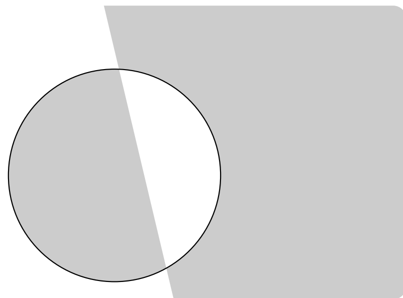

Figure 2: Depth-defining regions for lifted regression depth are symmetric differences of a halfplane and a disk bounded by the given circle.

combinatorially, there is little difference between the Poincaré halfspace model of hyperbolic geometry used by Mizera and Müller, and the Euclidean geometry of the points on the lifted paraboloid.

In more detail, the correspondence between the hyperbolic and parabolic interpretations of location-scale depth is as follows. For any point $c = (\mu, h)$ above the paraboloid (that is, satisfying the inequality $h > |\mu|^2$) let $\sigma = \sqrt{h - |\mu|^2}$, and let C be the sphere in $\mathbb{R}^d$ with center μ and radius σ. Then, C is the unique sphere centered at μ such that the hyperplane in $\mathbb{R}^{d+1}$ containing $p(C)$ also contains the point c. Conversely, any sphere C in $\mathbb{R}^d$ corresponds in this way to a point c at the intersection of the hyperplane through $p(C)$ and the vertical line through the center of C. The Tukey depth of c among the set of lifted sample points in $\mathbb{R}^{d+1}$ is defined to be the minimum number of samples in any closed halfspace bounded by c. The boundary of such a closed halfspace intersects the paraboloid in the lifted image of a sphere, which can be shown to intersect C equatorially. That is, the equivalent of Theorem 1 holds for parabolic lifting, so we get the same depth for C when we view $\mathbb{R}^d$ as the boundary of a hyperbolic Poincaré model and measure the hyperbolic Tukey depth of (μ, σ) as we get when we lift $\mathbb{R}^d$ to the paraboloid and measure the Euclidean Tukey depth of c.

This gives us a way of interpreting location-scale depth as Tukey depth without resorting to hyperbolic geometry. The penalty for using this interpretation, however, is that the coordinate h of the Euclidean points (μ, h) cannot be directly interpreted as a scale. Instead, we must recover the scale using the simple formula $\sigma = \sqrt{h - |\mu|^2}$.

If we use parabolic lifting to compute a depth value of a circle C in $\mathbb{R}^2$ (or sphere in higher dimensions), it is natural to consider an alternative definition of the depth of C, in which we consider the regression depth of the plane (or hyperplane) through $p(C)$ among the lifted samples rather than the Tukey depth of the point c associated with C. If depth were defined in this way, it would equal the minimum number of samples in any set formed as the closure of a symmetric difference of a halfspace with the disk bounded by C (Figure 2). Thus, this definition of depth is quite difference from location-scale depth, which as we have seen can be defined as the minimum number of samples in a ball or ball complement intersecting C equatorially. It is unclear to us what statistical significance this lifted regression depth might have for multivariate data, but in the univariate case it appears to give as the deepest fit merely the pair of $\frac{1}{3}$- and $\frac{2}{3}$-quantiles.

3 Optimal Möbius Transformation

The problem of Tukey depth for points on the boundary of a model of hyperbolic space, used by Mizera and Müller to visualize location-scale depth, has also arisen in non-statistical contexts.

The halfspace hyperbolic model used by Mizera and Müller can be converted by stereographic projection into a spherical Klein or Poincaré model of hyperbolic space, but this projection can be performed in different ways, leaving different hyperbolic points at the center of the spherical model. These different spherical models are related to each other by *Möbius transformations* of the sphere, and the problem of finding a Tukey median can be viewed as one of *optimal Möbius transformation* [2] in which one seeks a Möbius transformation of a spherical point set that minimizes some function of the transformed input. If we find the hyperbolic Tukey median of a spherical point set S, and transform the points to place that median at the center of the sphere, the resulting transformation minimizes the *hemisphere discrepancy*

$$\max_H \left| \frac{|H \cap S|}{|S|} - \frac{1}{2} \right|,$$

where the maximum is taken over all hemispheres H of the sphere. This interpretation makes clear the Möbius equivariance observed by Mizera and Müller.

The idea of using Tukey depth for a stereographic projection of the input was used by Teng et al. [7, 10, 11] as part of a scheme for partitioning finite element meshes and other geometric graphs. The authors of these papers use stereographic projection to map mesh vertices from $\mathbb{R}^d$ to the surface of a sphere in $\mathbb{R}^{d+1}$, compute an approximate Tukey median of the resulting spherical point set [5], transform the sphere so the computed median is at its center, and partition the sphere by a plane through the center. This partition plane is chosen either randomly [10] or by a deterministic technique that mimics the properties of randomly chosen planes [7]. The stereographic projection can then be reversed to map the partition plane back to a circle or sphere in the original d-dimensional space, which partitions the space into two pieces. Alternatively, this procedure can be viewed as treating $\mathbb{R}^d$ as the boundary of a halfspace model of $\mathbb{H}^{d+1}$, finding a high Tukey depth point in $\mathbb{H}^{d+1}$, choosing a random hyperbolic plane through that point, and partitioning $\mathbb{R}^d$ on the limit points of that plane.

The low hemisphere discrepancy of the transformed spherical point set found by this method (or equivalently, the high Tukey depth of the chosen point in $\mathbb{H}^{d+1}$) implies that any plane through the center point leads to a well-balanced partition in which both the inside and the outside of the partition sphere contain a constant fraction of the original mesh's vertices. Teng et al. showed that their choice of a random plane causes only $\mathcal{O}(n^{1-1/d})$ of the mesh elements (in expectation) to be cut by the partition sphere, so these ideas result in an efficient algorithm for partitioning meshes into large pieces with few elements on the boundaries of the pieces.

4 Algorithmic Implications

For practitioners interested in using algorithms for location-scale depth, it is fortunate, and for theoreticians interested in developing algorithms for location-scale depth, it is unfortunate, that this new depth notion can be transformed so easily into standard concepts of Tukey depth. As the authors note, the existence of Klein models for hyperbolic geometry allows one to use any algorithm developed for Tukey depth to compute efficiently the corresponding properties of location-scale depth. In this way we can immediately find algorithms for computation of location-scale depth [4], finding Student medians [4] or approximate medians [5], visualizing depth contours [8, 12], and maintaining guaranteed-accuracy approximations of depth for streaming data [1]. In each case the computational complexity of the location-scale depth problem in dimension d is equal to that for the corresponding Tukey depth problem in dimension $d+1$. Known bounds on the breakdown point of the Tukey median can also be transferred to location-scale depth. Mizera and Müller suggest that the time complexity of location-scale depth problems may in some cases be better than that for the higher dimensional Tukey depth, citing computation of depth contours as one example, but the Tukey depth contour algorithms they cite in this comparison are not the best known. However, for the problem of computing a single depth contour,

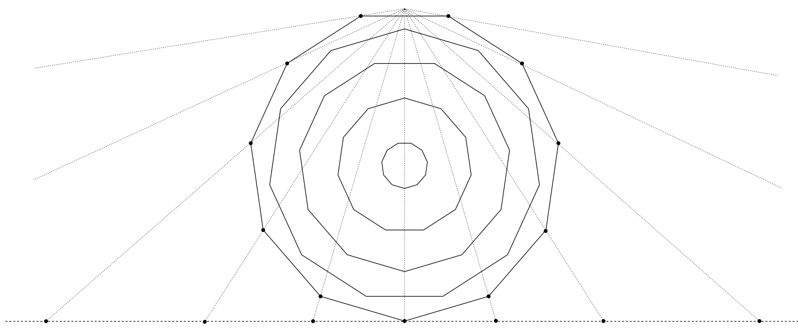

Figure 3: Tukey depth contours for regular n-gon, n odd, consist of $(n-1)/2$ nested regular n-gons with a total of $n(n-1)/2$ edges; stereographic projection produces a univariate sample set with combinatorially equivalent location-scale depth contours.

the $\mathcal{O}(n \log n)$ bound they give may indeed be an improvement on the best known time bound for computing a single bivariate Tukey depth contour, $\mathcal{O}(n \log^4 n)$ [9].

Rather than incurring a further transformation to a Euclidean problem, Mizera and Müller suggest implementing their algorithms by staying within the Poincaré halfplane model of hyperbolic geometry:

> While the transformation principle provides a theoretical argument, it is better in practical computations to perform all necessary operations directly in the original Poincaré plane, to avoid rounding errors arising in transforming to and from the Klein disk.

However, we feel that from the point of view of code simplicity and reuse, it may be better to perform computations using the parabolic lifting transformation. That is, a location-scale computation could be performed by replacing each sample z_i with the point $(z_i, |z_i|^2)$ in one higher dimension, using a Euclidean Tukey depth algorithm for the lifted points, and then transforming the results back to the hyperbolic model with the simple formula for the scale coordinate $\sigma = \sqrt{h - |\mu|^2}$. It is not clear to us, and would take a more careful numerical analysis to determine, whether these two transformational steps result in larger rounding errors than would be incurred by the increased complexity of the formulas needed for working directly within the Poincaré model.

Mizera and Müller ask whether the $\mathcal{O}(n^2)$ complexity of the algorithm for computing all depth contours [12] can be improved in the case of univariate location-scale depth, as the algorithm was developed to find contours for a more general problem, bivariate Tukey depth. The answer, at least in the worst case, is no. If an odd number of sample points are spaced regularly on a circle in the plane, then every pair of samples contributes a line segment to one of the depth contours (Figure 3). Stereographically projecting this regularly spaced point set onto a line produces a univariate sample with combinatorially equivalent location-scale depth contours, so quadratic time complexity is unavoidable.

It is also not possible to improve the $\mathcal{O}(n^{d+1} + n \log n)$ complexity of computing d-variate location-scale depth, relative to that for $(d+1)$-variate Tukey depth. To see this, suppose that we have a set of $(d+1)$-dimensional samples z_i, and wish to compute the Tukey depth of a point ξ. Replacing each sample z_i by the point $\xi + (z_i - \xi)/|z_i - \xi|$ leaves the Tukey depth of ξ unchanged, but modifies the sample set so that each sample lies on a unit sphere centered at ξ. Stereographic projection then produces a d-variate sample set, and a fit (μ, σ), that has the same location-scale depth as the Tukey depth of ξ in the original sample set. So, computation of d-variate location-scale depth can be no faster than $(d+1)$-variate Tukey depth.

The logistic and slash depths (if those turn out not to be rescaled versions of Student depth) and the Laplace version of location-scale depth, all discussed briefly by Mizera and Müller, may be more promising directions for algorithmic research, as little seems to be known already about these quantities. The suggestion of using other depth measures such as simplicial depth in the hyperbolic setting may also lead to new algorithmic problems, but seems less well motivated statistically than the likelihood-based tangent depth approach.

References

[1] A. Bagchi, A. Chaudhary, D. Eppstein, and M. T. Goodrich. Deterministic sampling and range counting in geometric data streams. To appear in *Proc. 20th Symp. Computational Geometry (SoCG 2004)*. ACM, 2004, arXiv:cs.CG/0307027.

[2] M. W. Bern and D. Eppstein. Optimal Möbius transformations for information visualization and meshing. *Proc. 7th Worksh. Algorithms and Data Structures (WADS 2001)*, pp. 14–25. Springer-Verlag, Lecture Notes in Computer Science 2125, August 2001, arXiv:cs.CG/0101006.

[3] K. Q. Brown. Voronoi diagrams from convex hulls. *Information Processing Letters* 9(5):223–228, 1979.

[4] T. M.-Y. Chan. An optimal randomized algorithm for maximum Tukey depth. *Proc. 15th Symp. Discrete Algorithms (SODA 2004)*, pp. 423–429. ACM and SIAM, January 2004.

[5] K. L. Clarkson, D. Eppstein, G. L. Miller, C. Sturtivant, and S.-H. Teng. Approximating center points with iterated Radon points. *Int. J. Computational Geometry & Applications* 6(3):357–377, September 1996.

[6] H. Edelsbrunner and R. Seidel. Voronoi diagrams and arrangements. *Discrete & Computational Geometry* 1:25–44, 1986.

[7] D. Eppstein, G. L. Miller, and S.-H. Teng. A deterministic linear time algorithm for geometric separators and its applications. *Fundamenta Informaticae* 22(4):309–331, April 1995.

[8] S. Krishnan, N. H. Mustafa, and S. Venkatasubramanian. Hardware-assisted computation of depth contours. *Proc. 13th Symp. Discrete Algorithms (SODA 2002)*, pp. 558–567. ACM and SIAM, 2002.

[9] J. Matoušek. Computing the center of planar point sets. *Discrete and Computational Geometry: Papers from the DIMACS Special Year*, pp. 221–230. Amer. Math. Soc., DIMACS Ser. Discrete Mathematics 6, 1991.

[10] G. L. Miller, S.-H. Teng, W. P. Thurston, and S. A. Vavasis. Separators for sphere-packings and nearest neighbor graphs. *J. ACM* 44(1):1–29, January 1997.

[11] G. L. Miller, S.-H. Teng, W. P. Thurston, and S. A. Vavasis. Geometric separators for finite-element meshes. *SIAM J. Scientific Computing* 19(2):364–386, March 1998.

[12] K. Miller, S. Ramaswami, P. J. Rousseeuw, J. A. Sellarès, D. Souvaine, I. Streinu, and A. Struyf. Efficient computation of location depth contours by methods of computational geometry. *Statistics and Computing* 13(2):153–162, April 2003.

[13] I. Mizera. On depth and deep points: A calculus. *Annals of Statistics* 30(6):1681–1736, December 2002.

[14] P. J. Rousseeuw and M. Hubert. Regression depth. *J. American Statistical Assoc.* 94:388–402, 1999.