Exact algorithms for computing the location depth and the k-th depth regions based on parallel arrangement constructions

Komei Fukuda
School of Computer Science
McGill University

Vera Rosta
Department of Mathematics and Statistics
McGill University

October 11, 2002

Abstract

The *location depth* of a point P relative to a data set S of n points in R^d, is the smallest number of points of S in any closed halfspace with boundary through P. The *k-th depth region* D_k, is the set of points whose location depth is at least k. A point with maximum location depth or of deepest location is called Tukey median. For any dimension d, as a consequence of Helly's theorem, the deepest location is at least $\lceil \frac{n}{d+1} \rceil$ and the set of points with at least this depth is called the *center*.

Johnson and Preparata proved [4], that the computation of the location depth of a given point is NP hard, even if the data set consists of points on a d-dimensional sphere. Amaldi and Kann point out that the location depth problem for a data point set on the sphere is equivalent to the homogeneous maximum feasibility subsystem problem MFS, thus MFS is also NP hard.

Although there are numerous papers on the location depth and the k-th depth regions in high dimensions, e.g. [2, 3, 4, 6], there has been no attempt to develop implementable exact algorithms for computing them. In this paper, we propose exact, memory efficient and highly parallelizable algorithms to find the location depth of a point, and the boundary of the location depth regions in any dimension. Note that in dimension 2, the paper [5] presents an algorithm and implementation to compute D_k based on the topological sweep technique. Unfortunately the topological sweep does not generalize to dimension higher than 2 because of the existence of non-Euclidean oriented matroids and such operation can get "stuck."

To compute the location depth of a point we consider the equivalent problem of finding the sign vector in a hyperplane arrangement with maximal difference in the number of positive and negative signs. We use a newly developed, hyperplane arrangement construction algorithm based on reverse search and linear programming which is memory efficient, highly parallelizable and output sensitive. i.e. polynomial in the size of input and output. A preliminary parallel implementation of the arrangement construction algorithm based on ZRAM and CDDLIB is available from the first author's homepage. This code can be easily adaptable to the location depth computation using our transformation.

The depth regions computation is closely related to efficient k-set enumeration and to (i, j)-partitions. In the k-set problem a hyperplane cleanly separates k-points from the remaining $n - k$ points for an n-point set. Andrzejak and Fukuda [1] developed algorithms for enumerating all k-sets of a point set in R^d. The (i, j)-partitions generalize the notion of k-sets, where an oriented hyperplane contains i points and the positive halfspace contains j points, thus the k-sets correspond to the $(0, k)$-partitions.

We propose an algorithm to construct the location depth regions D_k based on the enumeration of k-sets and the vertex enumeration of a convex polytope. The algorithm is memory efficient and parallelizable. Again, the algorithm can be implemented by using ZRAM and the basic functions of CDDLIB.

References

[1] A. Andrzejak and K. Fukuda, Optimization over k-set polytopes and efficient k-set enumeration, *Proc. 6th International Workshop on Algorithms And Data Structures (WADS'99)*, LNCS 1663, 1–12. Springer-Verlag, 1999.

[2] K.L. Clarkson, D. Eppstein, G.L. Miller, C. Sturtivant and S-H. Teng, Approximating center points with iterative Radon points, *International Journal of Computational Geometry and Applications*, (1993), 1–21.

[3] L. Danzer, J. Fonlupt and V. Klee, Helly's theorem and its relatives, *Proceedings of Symposia in Pure Mathematics 7, AMS* (1963), 101–180.

[4] D.S. Johnson and F.P. Preparata, The densest hemisphere problem, *Theoretical Computer Science* **6** (1978), 93–107.

[5] K. Miller, S. Ramaswami, P. Rousseeuw, T. Sellares, D. Souvaine, I. Streinu and A. Struyf, Fast Implementation of Depth Contours using Topological Sweep, *Proc. 12th Symposium on Discrete Algorithms (SODA), Washington DC, Jan 7-9* (2001), 690–699, and *Statistics and Computing* (2002), to appear.

[6] A. Struyf and P. Rousseeuw, High-dimensional computation of the deepest location, *Computational Statistics and Data Analysis*, **34** (2000), 415-426.