

n° 2017-71

**Fast computation of Tukey trimmed regions
and median in dimension $p > 2$**

X. LIU¹

K. MOSLER²

P. MOZHAROVSKIY³

Les documents de travail ne reflètent pas la position du CREST et n'engagent que leurs auteurs.
Working papers do not reflect the position of CREST but only the views of the authors.

¹ School of Statistics, Research Center of Applied Statistics of Jiangxi University of Finance and Economics, E.mail : X.Liu2@lboro.ac.uk

² Institute of Econometrics and Statistics, University of Cologne, E.mail : mosler@statistik.uni-koeln.de

³ CREST-ENSAI, Université Bretagne Loire, E.mail : pavlo.mozharovskyi@ensai.fr

Fast computation of Tukey trimmed regions and median in dimension $p > 2$

Xiaohui Liu*

School of Statistics, Research Center of Applied Statistics of
Jiangxi University of Finance and Economics,

Karl Mosler

Institute of Econometrics and Statistics, University of Cologne
and

Pavlo Mozharovskyi

CREST, Ensai, Université Bretagne Loire

December 6, 2017

Abstract

Given data in \mathbb{R}^p , a Tukey κ -trimmed region, shortly Tukey κ -region or just Tukey region, is the set of all points that have at least Tukey depth κ w.r.t. the data. As they are visual, affine equivariant and robust, Tukey regions are useful tools in nonparametric multivariate analysis. While these regions are easily defined and interpreted, their practical application is impeded by the lack of efficient computational procedures in dimension $p > 2$. We derive a strict bound on the number of facets of a Tukey region and construct a new efficient algorithm to compute the region, which runs much faster than existing ones. The new algorithm is compared with a slower exact algorithm, yielding always the same correct results. Finally, the approach is extended to an algorithm that efficiently calculates the innermost Tukey region and its barycenter, the Tukey median.

Keywords: Tukey depth, Tukey median, halfspace depth, location depth, depth contours, depth regions, computational geometry, efficient implementation.

*Xiaohui Liu's research was supported by *NNSF of China (Grant No.11601197, 11461029)*, *China Postdoctoral Science Foundation funded project (2016M600511, 2017T100475)*, *NSF of Jiangxi Province (No.20171ACB21030, 20161BAB201024)*, and the *Key Science Fund Project of Jiangxi provincial education department (No.GJJ150439)*.

1 Introduction

To describe the centrality of a point in a multivariate set of data, based on an idea of Tukey (1975), Donoho (1982) made a celebrated proposal: Given data $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^p$ and an additional point \mathbf{x} ,

$$d(\mathbf{x}) = d(\mathbf{x}|P_{\mathcal{X}}) = \inf_{\mathbf{u} \in S^{p-1}} \frac{1}{n} \# \{i \in \{1, \dots, n\} : \mathbf{u}^\top \mathbf{x} \leq \mathbf{u}^\top \mathbf{x}_i\} \quad (1)$$

measures how central \mathbf{x} is situated w.r.t. the data. Here $S^{p-1} = \{\mathbf{v} \in \mathbb{R}^p : \|\mathbf{v}\| = 1\}$ denotes the unit sphere in \mathbb{R}^p , $\#(B)$ the cardinal number of a set B , and $P_{\mathcal{X}}$ is the empirical distribution of the data. As a function of \mathbf{x} , (1) is referred to in the literature as *location depth* or *halfspace depth*, and also, to recognize the seminal work of Tukey (1975), as *Tukey depth*. The Tukey depth attains its maximum at the *Tukey median set*. The depth decreases when \mathbf{x} moves on a ray originating from any point of the Tukey median set; it vanishes outside the convex hull, $\text{conv}(\mathbf{x}_1, \dots, \mathbf{x}_n)$, of the data. Thus, the depth provides a center-outward ordering of points in \mathbb{R}^p .

The Tukey depth is invariant against a simultaneous affine transformation of \mathbf{x} and the data. Moreover, and most important, its value does not change as long as any of the points $\mathbf{x}, \mathbf{x}_1, \dots, \mathbf{x}_n$ is moved without crossing a hyperplane that is generated by p elements of $\{\mathbf{x}, \mathbf{x}_1, \dots, \mathbf{x}_n\}$. By the last property, the Tukey depth is rather robust against outlying observations; see Donoho (1982).

A *Tukey κ -region* $\mathcal{D}(\kappa)$ is the set of points in \mathbb{R}^p that have at least Tukey depth κ , $0 < \kappa \leq 1$. It is a closed convex polyhedron, included in $\text{conv}(\mathbf{x}_1, \dots, \mathbf{x}_n)$, and hence compact. Tukey regions are nested, they shrink with increasing κ . An empirical distribution is fully characterized by its Tukey regions (Struyf and Rousseeuw, 1999).

Similarly, given a probability distribution P in \mathbb{R}^p , the centrality of a point \mathbf{x} regarding P in \mathbb{R}^p is measured by the *population version* of the Tukey depth as follows:

$$d(\mathbf{x}) = d(\mathbf{x}|P) = \inf_{\mathbf{u} \in S^{p-1}} P\{\mathbf{z} \in \mathbb{R}^p : \mathbf{u}^\top \mathbf{x} \leq \mathbf{u}^\top \mathbf{z}\} \quad (2)$$

Note that (1) is a special case of (2) with $P = P_{\mathcal{X}}$. The population version of Tukey regions determines the underlying distribution if either the distribution is discrete (Koshevoy, 2002) or the regions' boundaries are smooth (Kong and Zuo, 2010).

In multivariate analysis, a broad nonparametric methodology has been based on Tukey depth and Tukey regions; including multivariate procedures of signs and ranks, order statistics, quantiles, bagplots, and measures of outlyingness and risk. As the Tukey depth is affine invariant and robust, so is any inference based on it. The same holds for Tukey regions as set-valued statistics. For such procedures, see for example Yeh and Singh (1997), Serfling (2006), Mozharovskiy et al. (2015), Hubert et al. (2015) and references therein. Many other depth notions have been proposed and used in the literature for descriptive as well as inferential procedures. Among them are the simplicial depth (Liu, 1990), the zonoid depth (Koshevoy and Mosler, 1997) and the projection depth (Liu, 1992; Zuo, 2003). For a recent survey, see, e.g., Mosler (2013). General definitions of a depth function can be found in Zuo and Serfling (2000) and Dyckerhoff (2004).

The various notions of data depth differ in their theoretical properties: especially, regarding invariance and robustness, convergence to their population version, and whether they fully characterize an underlying distribution. By this, from an applied view, different depths fit to

different applications. However, most decisive in possible applications is, whether the depth can be efficiently computed for realistic sample sizes and in higher dimensions. To calculate the Tukey depth, feasible algorithms have been developed by Ruts and Rousseeuw (1996) for dimension $p = 2$, Rousseeuw and Struyf (1998) for $p = 3$, and most recently by Dyckerhoff and Mozharovskiy (2016) and Liu (2017) for general p ; see also Miller et al. (2003).

To compute a Tukey region of given data $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ appears as an even more challenging combinatorial task, as it involves a very large number of observational hyperplanes to be possibly inspected. By an *observational hyperplane* we mean a hyperplane that passes through p elements of $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. In dimension two the task has been solved by use of a circular sequence, which enumerates all intersections of observational hyperplanes (Ruts and Rousseeuw, 1996). Kong and Mizera (2012) demonstrate that a Tukey κ -region is bordered by hyperplanes $\mathcal{H}_{\text{KM}}(\kappa, \mathbf{u})$, which correspond to κ -quantiles of projections on their normals \mathbf{u} . Moreover, the Tukey κ -region $\mathcal{D}(\kappa)$ is the *infinite* intersection over all directions \mathbf{u} of the inner halfspaces bordered by $\mathcal{H}_{\text{KM}}(\kappa, \mathbf{u})$. Hereafter,

$$\mathcal{H}_{\text{KM}}(\kappa, \mathbf{u}) = \{\mathbf{z} \in \mathbb{R}^p : \mathbf{u}^\top \mathbf{z} \geq q_1(\kappa, \mathbf{u})\},$$

where $q_1(\kappa, \mathbf{u})$ denotes the sample κ -quantile of the projections of $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ onto \mathbf{u} .

But, as a convex polytope, a Tukey region is the intersection of a *finite* number of these halfspaces. The facets of the polytope lie on the hyperplanes that border the halfspaces. Clearly, by the definition of Tukey depth (1), each of these hyperplanes must be an observational hyperplane. Consequently, the Tukey region is completely determined by a finite number of observational hyperplanes. Hence, to calculate it, the key step is to identify the proper observational hyperplanes. A naïve procedure consists in passing through all $\binom{n}{p}$ observational hyperplanes and checking their depth. For a more efficient procedure, we need a strategy to identify those observational hyperplanes that contain the facets.

Hallin et al. (2010) and Paindaveine and Šiman (2011), hereafter HPS, point out a direct connection between a Tukey region and multivariate regression quantiles. Each such quantile consists of, in general more than one, parallel hyperplanes, one of which may contain a facet of the Tukey region. In their pioneering work, Hallin et al. (2010) show that those directions giving the same set of hyperplanes form a polyhedral cone, and that a finite number of these cones fills \mathbb{R}^p . Each cone is represented by the directions generating its edges, which, again, are finitely many. HPS propose an algorithm that calculates Tukey regions in dimension $p > 2$ via quantile regression and parametric programming. To guarantee all cones to be addressed, a breadth-first search is used. For details of the implementations, see Paindaveine and Šiman (2012a,b). However, also these procedures are rather slow and inefficient.

In the sequel we present two new algorithms, a naïve and an efficient one, that calculate Tukey regions in arbitrary dimension $p > 2$. In building the efficient algorithm, certain combinatorial properties of Tukey regions are derived and exploited that substantially reduce the computational load. Consequently this algorithm runs much faster and requires much less RAM than the algorithms by HPS.

Throughout the paper we assume that the data is *in general position*, i.e., that any observational hyperplane contains exactly p data points (Donoho, 1982). Otherwise, the data may be slightly perturbed to meet the assumption.

Specifically, as a first main result, an upper bound is derived on the number of *non-redundant hyperplanes* of a Tukey region, that is, those observational hyperplanes that contain a facet of the region. The bound is sharp and turns out to be very useful in assessing the

computational complexity and performance of the algorithms. Up to our knowledge of the literature, the bound is new.

The HPS procedures are inefficient for two reasons. First, it appears that both their implementations yield a great number of *redundant directions*, which are normal to an observational hyperplane but provide no facet of the trimmed region. However, all these directions are considered in HPS and used to calculate regions. Given κ , the HPS procedures *actually calculate* p successive regions instead of one by breadth-first; but many of these regions have depth $\neq \kappa$ (see Paindaveine and Šiman (2011), remark after Theorem 4.2). Second, the *cone-by-cone* search strategy is both RAM- and time-consuming. This is because, in the HPS procedures, each cone is characterized by its facets and vertices, and facets are identified by p -variate vectors. A rather large RAM is required to store these identifiers with sufficient precision.

First, we present a naïve combinatorial algorithm (Algorithm 0). It serves as a benchmark for our main algorithm (Algorithm 1), which is efficient and fast. The naïve procedure, in searching for facets' candidates, simply passes through all combinations of $p - 1$ observations as the case may be. No memory-consuming structure has to be created, and the computational time is independent of κ . In contrast, our fast and efficient approach (Algorithm 1) uses a breadth-first search strategy. However, instead of covering \mathbb{R}^p *cone-by-cone*, as it is done by HPS, it searches the directions *ridge-by-ridge*, where a *ridge* corresponds to a combination of $p - 1$ observations in \mathbb{R}^p . This strategy yields only relevant hyperplanes (that cut exactly the required number of observations off from $\mathcal{D}(\kappa)$), and thus examines much fewer cases. Additionally, we store each ridge by the subscripts of its $p - 1$ corresponding observations and use some novel tricks to substantially save both RAM and computational time. Obviously, Algorithm 0 is exact. For Algorithm 1, we have no theoretical proof of its exactness though. But we have broad numerical evidence that it computes the exact region. In all our experiments Algorithm 1 yields precisely the same Tukey region as the exact Algorithm 0 does.

Similar to HPS, our approach exploits the connection between Tukey trimmed regions and quantile regions pointed out by Kong and Mizera (2012), *viz.* that the Tukey κ -region is the intersection of directional κ -quantile halfspaces, taken over all directions of the unit sphere. Additionally, we make use of further combinatorial properties of a Tukey region, which are basic to the algorithms of this paper.

As they involve only simple operations and no optimisation techniques, both our algorithms are easy to program; they also show high numerical precision even in larger dimension. Particularly, Algorithm 1, by its speed and storage efficiency, enables the use of statistical methodology based on Tukey-region statistics. To investigate the performance of the algorithms, a simulation study as well as real data calculations (up to dimension 9) are provided below. Our procedures have been implemented in C++ and visualized in **R**. They are available in the R-package **TukeyRegion** and can be downloaded from CRAN.

The *Tukey median* (Tukey, 1975) is one of the most famous generalizations of the ordinary median in Euclidean spaces of dimension $p > 1$. It is usually defined (Donoho, 1982) as the average of all points in the Tukey median set. Hence its computation depends essentially on the computation of this innermost Tukey region. As an extension of our approach, we provide an algorithm for efficiently computing the Tukey median.

The rest of this paper is organized as follows. Section 2 presents an upper bound on the number of non-redundant hyperplanes of a Tukey region, together with some results that are useful for our algorithms. Section 3 describes the naïve algorithm as well as the efficient one. Section 4 studies the computational performance. Section 5 is about computing the Tukey median, and Section 6 concludes.

2 A bound on the number of facets

In this section, as a first principal result, a bound on the number of facets of a Tukey region is derived, that is, on the number of non-redundant halfspaces defining the region.

Assume that the observed data set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subset \mathbb{R}^p$, $n > p \geq 2$, is in general position. Consequently, each facet of a Tukey region $\mathcal{D}(\kappa)$ lies on an observational hyperplane containing exactly p observations. Consider an observational hyperplane Π and denote $m_\kappa = \lceil n\kappa \rceil$ with $\lceil \cdot \rceil$ being the ceiling function. Then the following proposition holds.

Proposition 1. *If Π passes through observations $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_p}$ and cuts at most $m_\kappa - 1$ other observations off, then*

$$\Pi \cap \mathcal{D}(\kappa) \subset \text{conv}(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_p}).$$

Proof: Denote $\mathcal{F} = \Pi \cap \mathcal{D}(\kappa)$. If Π cuts exactly $m_\kappa - 1$ observations off and $\mathcal{F} \neq \emptyset$, then \mathcal{F} must be included in the boundary of $\mathcal{D}(\kappa)$. Now suppose that there exists $\mathbf{x}_0 \in \mathcal{F} \setminus \text{conv}(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_p})$. One may use the same technique as Dyckerhoff and Mozharovskiy (2016) in moving Π around \mathbf{x}_0 to exclude all points $\mathbf{x}_{i_j}, j = 1, \dots, p$, from it. This results in depth $d(\mathbf{x}_0) < \kappa$, contradicting with $\mathbf{x}_0 \in \mathcal{D}(\kappa)$. If $\Pi(\mathbf{Z}_p)$ cuts away less than $m_\kappa - 1$ observations, then $\mathcal{F} = \emptyset$, and the claim trivially holds. \square

By a κ -outside ridge we mean the intersection of two observational hyperplanes that, on their lower side, cut away at most $m_\kappa - 1$ observations. Any κ -outside ridge is a $(p - 2)$ -dimensional affine space passing through a set of $p - 1$ observations, say $\{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}\}$. Let $\mathbf{V} = \text{span}(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}})$ be the $(p - 2)$ -dimensional vector space spanned by $\{\mathbf{x}_{i_1} - \mathbf{x}_{i_{p-1}}, \mathbf{x}_{i_2} - \mathbf{x}_{i_{p-1}}, \dots, \mathbf{x}_{i_{p-2}} - \mathbf{x}_{i_{p-1}}\}$, and \mathbf{V}^\perp be its orthogonal complement. (Observe that $\mathbf{V} = \{\mathbf{0}\}$ if $p = 2$.) Consider the projection of $\mathcal{D}(\kappa)$ onto the two-dimensional vector space \mathbf{V}^\perp , which is $\text{proj}_{\mathbf{V}^\perp}(\mathcal{D}(\kappa)) = \{\mathbf{x}' \in \mathbf{V}^\perp : \mathbf{x}' + \mathbf{x}'' \in \mathcal{D}(\kappa) \text{ with } \mathbf{x}'' \in \mathbf{V}\}$. Clearly, this projection is a polygone. Relying on Proposition 1, one obtains the following result, which will be useful in constructing our efficient algorithm (Section 3).

Proposition 2. *Consider a κ -outside ridge with corresponding observations $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}$, and the vector space \mathbf{V}^\perp as above. Then the projection of $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}$ onto \mathbf{V}^\perp is either a vertex or an outside point of $\text{proj}_{\mathbf{V}^\perp}(\mathcal{D}(\kappa))$.*

A hyperplane that cuts exactly $m_\kappa - 1$ observations off will be mentioned as a *relevant hyperplane*, its inner halfspace as a *relevant halfspace*. Obviously, the Tukey region $\mathcal{D}(\kappa)$ is the intersection of its relevant halfspaces. For any κ -outside ridge, $\kappa \in \{1/n, 2/n, \dots, \kappa^*\}$ with $\kappa^* = \sup_{\mathbf{x} \in \mathbb{R}^p} d(\mathbf{x} | P_{\mathcal{X}})$, we see from Proposition 2 and the convexity of $\mathcal{D}(\kappa)$ that there exist at least two relevant hyperplanes that contain the observations $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}$ corresponding to this ridge (together with another observation) and cut exactly $m_\kappa - 1$ observations off. Moreover, the intersection of the respective halfspaces contains the corresponding $(p - 2)$ -dimensional affine space. It is easy to see that among them two halfspaces exist, say \mathcal{H}_1 and \mathcal{H}_2 , such that $\mathcal{D}(\kappa) \subset \mathcal{H}_1 \cap \mathcal{H}_2$. Figure 1 illustrates this; it also demonstrates that a relevant hyperplane can be redundant. Hence, we have the following result.

Theorem 1. *Let the data be in general position and $\kappa \in \{1/n, 2/n, \dots, \kappa^*\}$. If the Tukey region $\mathcal{D}(\kappa)$ is not a singleton, the number of its non-redundant bordering hyperplanes (= number of facets) is bounded above by $2 \binom{n}{p-1} / p$.*

Proof: As discussed above, if $\mathcal{D}(\kappa)$ is not a singleton, each κ -outside ridge can participate in the construction of no more than two facets. Combinatorially, the number of such ridges is bounded from above by $\binom{n}{p-1}$. On the other hand, each $(p-1)$ -dimensional facet lies on an observational hyperplane which corresponds to exactly p κ -outside ridges. From this follows that at most $2\binom{n}{p-1}/p$ facets exist. \square

By the convexity of the Tukey region, a direction vector yields at most one facet of the corresponding trimmed region. In this sense, Theorem 1 actually provides also an upper bound for the number of *non-redundant directions*. It is useful in assessing the performance of an algorithm, and will be used in Step 4 of Algorithm 1 below.

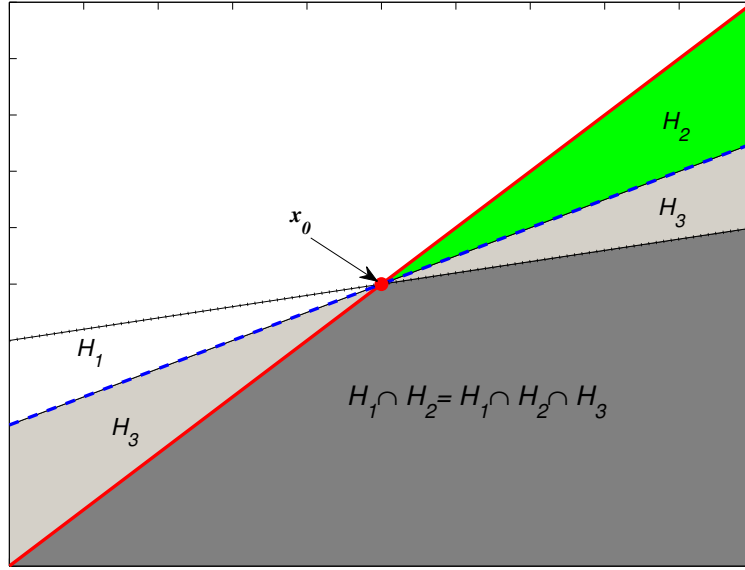


Figure 1: An illustration of the intersection of three relevant halfspaces. Here \mathbf{x}_0 denotes the projection of $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}$ onto \mathbf{V}^\perp . Clearly, $\bigcap_{k=1}^3 \mathcal{H}_k = \mathcal{H}_1 \cap \mathcal{H}_2$, and \mathcal{H}_3 is redundant.

It is worth mentioning that the upper bound $2\binom{n}{p-1}/p$ is *attainable* and thus cannot be further improved. For instance, the Tukey region at depth $\kappa = \frac{1}{p+1}$ of a simplex contains $p+1$ facets, which is equal to the upper bound: $2\binom{p+1}{p-1}/p = p+1$.

Proposition 2 and Theorem 1 with their proofs reveal further important properties of κ -outside ridges, which are useful in constructing an efficient algorithm to calculate $\mathcal{D}(\kappa)$ for $\kappa \in \{1/n, 2/n, \dots, \kappa^*\}$:

- Every κ -outside ridge can be utilized to compute at least two κ -critical directions. These are related to the observational hyperplanes that cut *exactly* $m_\kappa - 1$ observations off.
- The ridges are connected with each other in the following sense: Given a κ -outside ridge, we may replace one of its defining observations $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}$ with another observation \mathbf{x}^* , so that the hyperplane through $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}, \mathbf{x}^*$ cuts *exactly* $m_\kappa - 1$ observations off. Then another κ -outside ridge is obtained, which may also be utilized in the computation of $\mathcal{D}(\kappa)$.

- Observe that a κ -outside ridge is also a λ -outside for all $\lambda \in \{\frac{m_\kappa+1}{n}, \frac{m_\kappa+2}{n}, \dots, \kappa^*\}$. Hence, if one has to calculate *more than one region*, one may store all κ -outside ridges during the computation of $\mathcal{D}(\kappa)$ and recycle them when computing a more central trimmed region $\mathcal{D}(\lambda)$ with $\lambda > \kappa$.

3 Algorithms

This section presents two new algorithms to compute a Tukey region of given depth. We start by introducing Algorithm 0 in Subsection 3.1, which is a naïve application of Proposition 2. The Algorithm 0 is simple and intuitive; it will later serve to verify the correctness of Algorithm 1. After this, in Subsection 3.2 we describe the efficient algorithm, Algorithm 1.

3.1 The naïve combinatorial algorithm

Algorithm 0 simply passes through all combinations $\{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}}\}$ of $p-1$ out of n points and searches for hyperplanes passing through p points and cutting exactly $m_\kappa - 1$ observations off. To do this, first for each choice of observations $\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}}$, the $(p-2)$ -dimensional vector space spanned by them is calculated (Step 2a), and the sample is projected onto its orthogonal complement, which is a two-dimensional vector space (Step 2b). Then the search narrows down to finding two lines in this plane passing through the point to which $\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}}$ are projected and another point from the sample and cutting $m_\kappa - 1$ observations off (Step 2c). Figure 2 (right) visualizes the set of hyperplanes found during one iteration in Steps 2a to 2c. Each found hyperplane is stored as a number to the basis n . This requires $\lceil p \log_2(n) \rceil$ binary digits (bits) of memory space (Step 2d). Note that, under the assumption of general position made above, ties cannot occur, as any hyperplane contains at most p observations. Since the number of those combinations is $\binom{n}{p-1}$ and search in each of them is algorithmically dominated by the angle-sorting procedure having time complexity $O(n \log(n))$, the time complexity of the algorithm amounts to $O(n^p \log(n))$. Clearly, this presumes that the time complexity of Step 2d(iii) is not larger than that of Step 2d(i). This can be achieved by using appropriate store-search structures such as, *e.g.*, search trees (access time complexity $O(\log(n))$) or binary hypermatrices (access time complexity $O(1)$). As Algorithm 0 does not take account of any space ordering it requires minimum memory and saves computational time, which is otherwise needed for multiple access of search structures and may grow substantially with n and p . In addition, due to the same reason its execution time only negligibly depends on the geometry of the data cloud and the depth value κ .

Algorithm 0 (Naïve combinatorial algorithm).

Input: $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$, κ .

Step 1. Set $\mathcal{H}_\kappa = \emptyset$.

Step 2. For each subset $\{i_1, \dots, i_{p-1}\} = I \subset \{1, \dots, n\}$ do:

- Consider the plane normal to $\text{span}(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}})$ and find a basis of it. Let \mathbf{B}_I denote the basis matrix, that is the $(p \times 2)$ matrix \mathbf{B}_I containing the two basis vectors as columns.
- Compute $\mathbf{y}_i = \mathbf{B}_I^\top \mathbf{x}_i$ for $i = 1, \dots, n$.

- (c) Find a subset $I_\kappa \subset \{1, \dots, n\} \setminus I$ such that for each $i_0 \in I_\kappa$ holds $\#\{j : \mathbf{u}^\top \mathbf{y}_j > \mathbf{u}^\top \mathbf{y}_{i_0}, j = 1, \dots, n\} = m_\kappa - 1$ whenever $\mathbf{u}^\top (\mathbf{y}_{i_1} - \mathbf{y}_{i_0}) = 0$.
- (d) For each $i_0 \in I_\kappa$ do:
 - i. $(k_1, \dots, k_p) = \text{sort}(i_0, i_1, \dots, i_{p-1})$.
 - ii. Compute $h = k_1 + k_2 n + k_3 n^2 + \dots + k_p n^{p-1}$.
 - iii. If $h \notin \mathcal{H}_\kappa$ then add h to \mathcal{H}_κ .

Output: \mathcal{H}_κ .

3.2 The ridge-by-ridge breadth-first search strategy

In this subsection we present our main procedure, Algorithm 1, that computes the Tukey κ -region for given κ in an efficient way. The algorithm has been implemented in function `TukeyRegion` of the R-package `TukeyRegion` and can be downloaded from CRAN. Its output depends on the user's request, depending on which the algorithm terminates after all required elements have been computed. A detailed description of the algorithm with reference to user-accessible options is given right below.

To find all relevant hyperplanes of a Tukey κ -region, Algorithm 1 follows the breadth-first search idea. Briefly, this idea can be explained as follows: push an initial set of ridges into the queue, pop a ridge from the queue and push all those neighbors into the queue that have not been seen before; continue until the queue is empty (Steps 1 to 5). After this has been done, the region is constructed as an intersection of halfspaces, given by their hyperplanes. For this, firstly, an inner point of the region has to be found, which is used for a dual transformation of the hyperplanes (Step 6); then the QHULL (Barber et al., 1996) algorithm is run to eliminate the redundant hyperplanes (Step 7). Vertices, facets and the barycenter of the Tukey region are computed from the remaining hyperplanes (Step 8). The input consists of the observations, the depth level κ and a precision parameter ϵ .

Algorithm 1 (Algorithm for computing the Tukey region).

Input: $\mathbf{x}_1, \dots, \mathbf{x}_n \subset \mathbb{R}^p$, κ , ϵ .

Step 1. Initialization:

Set $\mathcal{A} = (\text{false}_n)^{p-1}$, $\mathcal{H}_\kappa = \emptyset$, an empty queue \mathcal{Q} .

Step 2. Construct initial set of ridges:

- (a) Find a subset $\{i_1, \dots, i_{p-1}\} = I \subset \{1, \dots, n\}$ such that $(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}})$ define a ridge of $\text{conv}(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Set $A(\text{sort}(I)) = \text{true}$ and push $\text{sort}(I)$ into \mathcal{Q} .
- (b) Compute a basis matrix \mathbf{B}_I of the plane normal to $\text{span}(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}})$.
- (c) Compute $\mathbf{y}_i = \mathbf{B}_I^\top \mathbf{x}_i$ for $i = 1, \dots, n$.
- (d) Find two indices l_1, l_2 such that holds $\#\{j : \mathbf{u}_k^\top \mathbf{y}_j > \mathbf{u}_k^\top \mathbf{y}_{l_k}, j = 1, \dots, n\} = m_\kappa - 1$ whenever $\mathbf{u}_k^\top (\mathbf{y}_{l_k} - \mathbf{y}_{i_1}) = 0$ for $k = 1, 2$.
For $k = 1, 2$ and for each subset $J \subset I$ of order $p - 2$, set $A(\text{sort}(J \cup \{l_k\})) = \text{true}$ and push $\text{sort}(J \cup \{l_k\})$ into \mathcal{Q} .

- (e) For $k = 1, 2$ and for each l such that holds $\mathbf{u}_k^\top \mathbf{y}_l > \mathbf{u}_k^\top \mathbf{y}_{l_k}$ and for each subset $J \subset I$ of order $p - 2$, set $A(\text{sort}(J \cup \{l\})) = \text{true}$ and push $\text{sort}(J \cup \{l\})$ into \mathcal{Q} .

Step 3. Pop a ridge $I = \{i_1, \dots, i_{p-1}\}$ from \mathcal{Q} .

Step 4. Spread to neighboring ridges:

- (a) Compute a basis matrix \mathbf{B}_I of the plane normal to $\text{span}(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}})$.
- (b) Compute $\mathbf{y}_i = \mathbf{B}_I^\top \mathbf{x}_i$ for $i = 1, \dots, n$.
- (c) Find a subset $I_\kappa \subset \{1, \dots, n\} \setminus I$ such that for each $i_0 \in I_\kappa$ holds $\#\{j : \mathbf{u}^\top \mathbf{y}_j > \mathbf{u}^\top \mathbf{y}_{i_0}, j = 1, \dots, n\} = m_\kappa - 1$ whenever $\mathbf{u}^\top (\mathbf{y}_{i_1} - \mathbf{y}_{i_0}) = 0$.
- (d) For each $i_0 \in I_\kappa$ do:
 - i. If $(I \cup \{i_0\}) \notin \mathcal{H}_\kappa$ then add $(I \cup \{i_0\})$ to \mathcal{H}_κ .
 - ii. For each subset $J \subset I_\kappa$ of order $p - 2$ do:

If $A(\text{sort}(J \cup \{i_0\})) = \text{false}$ then set $A(\text{sort}(J \cup \{i_0\})) = \text{true}$ and push $\text{sort}(J \cup \{i_0\})$ into \mathcal{Q} .

Step 5. If \mathcal{Q} is not empty then go to **Step 3**, else go to the following step.

(So far, all p -tuples of observations that define relevant halfspaces are stored in \mathcal{H}_κ . The intersection of these halfspaces is the Tukey κ -region. We give details below.)

Step 6. Find an inner point of the region:

- (a) For each $(i_1, \dots, i_p) \in \mathcal{H}_\kappa$ ($l = 1, \dots, n_\kappa = \#\mathcal{H}_\kappa$) do:
 - i. Find $\mathbf{u}_l \perp \text{span}(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_p})$ such that $\#\{j : \mathbf{u}_l^\top \mathbf{x}_j > \mathbf{u}_l^\top \mathbf{x}_{i_1}, j = 1, \dots, n\} = m_\kappa - 1$ and $\|\mathbf{u}_l\|_2 = 1$.
 - ii. Compute $b_l = \mathbf{u}_l^\top \mathbf{x}_{i_1}$.
- (b) Compute $\mathbf{x}_0 = \arg\max_{\mathbf{x} \in \mathbb{R}^p} \{\mathbf{x}^\top (1, 0, \dots, 0)^\top : \mathbf{u}_l^\top \mathbf{x} \leq b_l - \epsilon, l = 1, \dots, n_\kappa\}$.
- (c) If \mathbf{x}_0 cannot be found then **stop**.

Step 7. Eliminate redundant halfspaces:

- (a) For $j = 1, \dots, n_\kappa$ do:

$$\mathbf{w}_j = \frac{1}{b_j - \mathbf{u}_j^\top \mathbf{x}_0} \mathbf{u}_j.$$
- (b) $(\mathbf{i}_1, \dots, \mathbf{i}_{n_\kappa}^v)^\top = QHULL(\mathbf{w}_1, \dots, \mathbf{w}_{n_\kappa})$ with $\mathbf{i}_j = (i_{j1}, \dots, i_{jp})$.

Step 8. Compute elements that define the region:

- (a) For $j = 1, \dots, n_\kappa^v$ do:

$$\mathbf{v}_j = ((\mathbf{u}_{i_{j1}}, \dots, \mathbf{u}_{i_{jp}})^\top)^{-1} (b_{i_{j1}}, \dots, b_{i_{jp}})^\top.$$
- (b) For each $i \in \text{unique}(\mathbf{i}_1, \dots, \mathbf{i}_{n_\kappa^v})$ ($j = 1, \dots, n_\kappa^f$) do:
 - i. $\mathbf{d}_j = \mathbf{u}_i$.
 - ii. $t_j = b_i$.
- (c) $\mathbf{c} = \text{ave}(\text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{n_\kappa^v}))$, the barycenter of $\text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{n_\kappa^v})$.

Output:

- (a) Vertices: $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_{n_\kappa^v}\}$.
- (b) Facets' (non-redundant) hyperplanes: $\mathcal{F} = \{\mathbf{d}_1, \dots, \mathbf{d}_{n_\kappa^f}\}$ (outside-pointing normals) and $\mathcal{T} = \{t_1, \dots, t_{n_\kappa^f}\}$ (thresholds on these normals).
- (c) Barycenter: \mathbf{c} .

In **Step 1**, $(\text{false}_n)^{p-1}$ is a $(p-1)$ -dimensional logical matrix having format $n \times \dots \times n$, that is, in the beginning an n -dimensional vector of logical zeros, brought to power $p-1$ as a Cartesian product. (E.g., if $p=4$ this is a cube of size $n \times n \times n$.) Indeed, only one upper corner of this matrix is used, which includes those cells having strictly decreasing subscripts. Further, a single bit of RAM suffices to store a logical value, which amounts to eight values per byte. However, this matrix is memory demanding when n and p are large. Fortunately, in this last case the matrix is very probable to be sparse, and thus some dynamic storing structure may be used, e.g. a search tree. \mathcal{H}_κ is a set for storing hyperplanes as p -tuples of integer numbers, and \mathcal{Q} is literally a queue supporting operations of pushing an element on one side and popping it from another one.

Step 2 aims at finding a set of $(p-1)$ -tuples defining ridges to be used as starting points for the algorithm. First, a ridge of the convex hull of $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ defined by, say, $\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_{p-1}}\}$ is found (Step 2a). This is implemented in the QHULL algorithm. Further, using the logic of the Steps 2a to 2c of Algorithm 0, two points \mathbf{x}_{l_1} and \mathbf{x}_{l_2} are found (Steps 2b to 2d). Thus, each of the two hyperplanes defined by $\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_{p-1}}, \mathbf{x}_{l_1}\}$ and $\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_{p-1}}, \mathbf{x}_{l_2}\}$ cuts exactly $m_\kappa - 1$ observations off from the sample, see Figure 2 (left) for an illustration. This is always guaranteed as long as the sample is in general position and $m_\kappa \leq \lfloor \frac{n-(p-1)}{2} \rfloor$; here $\lfloor \cdot \rfloor$ denotes the floor function. Then, all $(p-1)$ -tuples containing $(p-2)$ points out of $\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_{p-1}}\}$ and one of the points cut off by or lying in one of these hyperplanes are chosen. Together with the ridge on the convex hull of the data set, this gives exactly $1 + 2m_\kappa(p-1)$ initial ridges (each corresponding to a $(p-1)$ -tuple of points from the sample).

Steps 3 and **5** wrap the search step procedure of Step 4 by implementing the queue. Step 3 pops a ridge to be processed from the head of the queue while Step 5 returns to Step 3 if the queue contains at least one element.

Step 4 provides identification of the neighboring ridges, by that implementing the ridge-by-ridge search strategy. First, a set I_κ of indices is found that defines, together with the current ridge $\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_{p-1}}\}$, a hyperplane cutting $m_\kappa - 1$ observations off from the data. (Here we follow the logic of Steps 2a to 2c of Algorithm 0 as before.) The information gained by each $i_0 \in I_\kappa$ is twofold. First, an i_0 defines a hyperplane determined by a p -tuple $\{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{p-1}}, \mathbf{x}_{i_0}\}$ which is critical for the Tukey region. We check whether this hyperplane is visited for the first time and add it to \mathcal{H}_κ if this is the case (Step 4d(i)). Second, each such hyperplane contains p ridges defined by $(p-1)$ -tuples of points. $p-1$ of these ridges – those containing i_0 – can potentially lead to a hyperplane not visited before. Thus we check each of these $(p-1)$ ridges whether it has been visited before or not. If a ridge has not been visited we add it to the queue \mathcal{Q} and mark it as visited (Step 4d(ii)). The set of ridges found in Step 4 of a single iteration is visualized in Figure 2 (right).

As the total number of these hyperplanes (after the first five steps of Algorithm 1 have been performed) can potentially be as large as their maximum number, the complexity of the algorithm is the same as that of Algorithm 0, i.e. $O(n^p \log(n))$. On the other hand, this worst

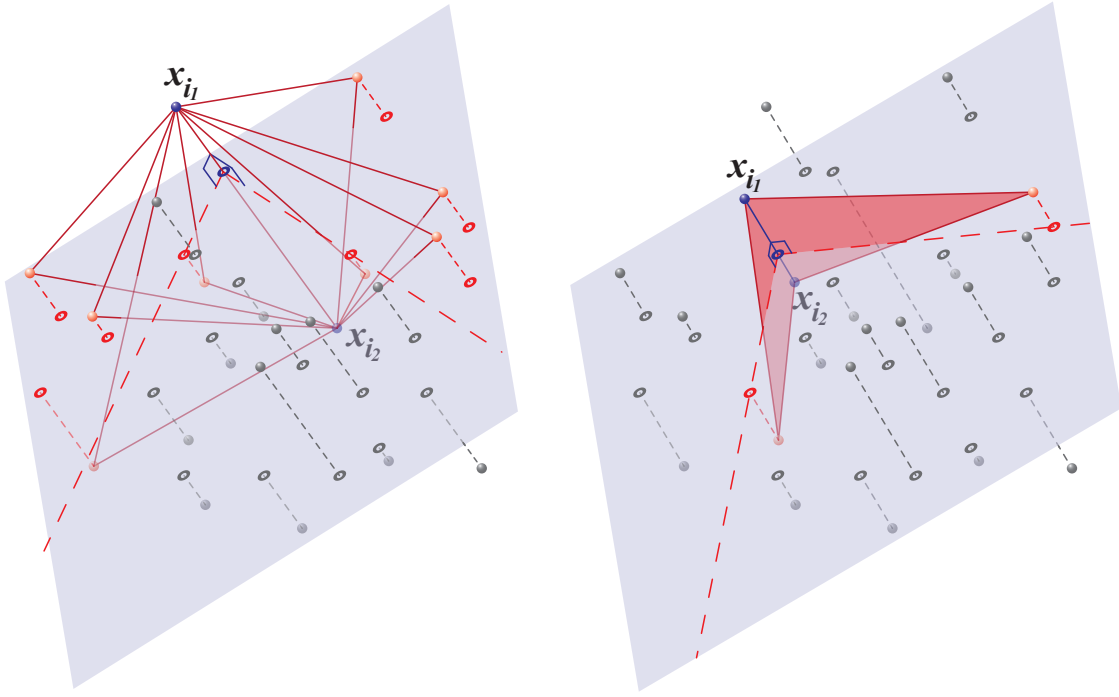


Figure 2: Left: Step 2 of Algorithm 1, solid red lines indicate initial ridges. Right: The generic Step 2 (a) – (c) of Algorithm 0 and Step 4 (a) – (c) of Algorithm 1, solid red lines indicate new ridges added to the queue.

case happens only for degenerate data sets. Thus it is reasonable to expect that in most cases the number of facets is substantially lower than the upper bound, and this number actually defines the computational speed of the algorithm. Some empirical insights to this question will follow in Section 4.

Steps 1 to 5 aim to compute all necessary hyperplanes defining halfspaces that determine the Tukey region by their intersection. We will use this set to check the correctness of the algorithm when comparing it empirically with the output of Algorithm 0 in Section 4.1. The implementation of Algorithm 1 in the R-package `TukeyRegion` performs these five first steps always independently of the chosen options. The following steps are optional. If the algorithm terminates after Step 5, only the hyperplanes are output, each as a p -tuple $\{j_1, j_2, \dots, j_p\}$ of indices of points from the sample.

Step 6 searches a point strictly belonging to the interior of the Tukey region. If such a point cannot be found then the algorithm terminates. See also Section 5. Step 6 is performed either if the interior point is directly requested (flag `retInnerPoint` is set) or if any of the options for the following steps is chosen.

Each element of \mathcal{H}_κ is a p -tuple of indices (j_1, j_2, \dots, j_p) defining a halfspace by the hyperplane containing $\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \dots, \mathbf{x}_{j_p}\}$. First, for each of these hyperplanes the normal pointing outside the Tukey region and the threshold defining the hyperplane's position are calculated (Step 6a); each of them defines a condition for the interior point. Then, the inner point \mathbf{x}_0 of the region is searched by means of linear programming as a point satisfying these conditions (Step 4b). If the inner point is not found this means that the Tukey region of depth κ has (numerically, with *precision* ϵ) zero volume or does not exist. In this case the algorithm terminates.

Step 7 eliminates redundant hyperplanes. It is performed either if these are requested (flag `retHyperplanesNR`) or if any of the options for the following step is chosen. First, the duality transformation is applied. It represents each halfspace by a vector \mathbf{w}_j , which is its outer normal multiplied by the inverse distance of the hyperplane to the inner point (Step 7a). Second, the QHULL algorithm is applied to the set of all \mathbf{w}_j 's. It returns the convex hull of the \mathbf{w}_j 's, an $n_\kappa^v \times p$ matrix where each row defines a facet by indices of p points (Step 7b). Since the data is in general position, each facet of this convex hull is defined by exactly p points.

Step 8 computes the elements of the region; it is the step that provides the practically important output. This step depends whether region's vertices (flag `retVertices` is set), facets (flag `retFacets` is set) and/or barycenter (flag `retBarycenter` is set) are requested and requires the successful execution of all preceding steps. Each facet of the convex hull of \mathbf{w}_j 's in the dual space (corresponding to a row of the matrix returned by the QHULL algorithm in Step 7b) defines a vertex of the Tukey region, *i.e.* each such vertex is computed as an intersection of p non-redundant hyperplanes (Step 8a).

The facets of the Tukey region are contained in the hyperplanes defined by pairs (\mathbf{d}_j, t_j) , $j = 1, \dots, n_\kappa^f$ obtained as non-repeating entries of the matrix $(\mathbf{i}_1, \dots, \mathbf{i}_{n_\kappa^v})^\top$ (Step 8b). To obtain facets as polygons one can apply QHULL to the set of regions' vertices $\{\mathbf{v}_1, \dots, \mathbf{v}_{n_\kappa^v}\}$. In addition, the barycenter of a Tukey region can be computed as the weighted average of the triangulated (doable by the QHULL algorithm as well) region, where points are the means of vertices of simplices (p -dimensional triangles) and weights are the volumes of these simplices (Step 8c).

To illustrate the output of the algorithm, Figure 3¹ exhibits $\mathcal{D}(\kappa)$ for $\kappa = 0.025, 0.1, 0.25$ for 50 observations drawn from a standard normal and a skewed normal distribution in \mathbb{R}^3 . The latter is according to Azzalini and Capitanio (1999) having skewness parameter equal to 5 in the first coordinate (see Section 4.1).

4 Numerical study

This section presents the results of a simulation study that demonstrates the validity of Algorithm 1 under diverse data generating distributions and explores its computational performance. All results on execution times are obtained using statistical software R on a Macbook Pro laptop possessing processor Intel(R) Core(TM) i7-4980HQ (2.8 GHz) having 16 GB of physical memory and macOS Sierra (Version 10.12.6) operating system. The validation experiment has been conducted on the computing cluster of École Nationale de la Statistique et de l'Analyse de l'Information.

4.1 Validation

The main idea here is to compare results of Algorithm 1 with those of the naïve Algorithm 0, that is known to be correct. We restrict the calculations to those which are feasible regarding Algorithm 0, that is, to pairs (n, p) for which Algorithm 0 computes a single Tukey region within less than one hour. As noted above, computational times of Algorithm 0 are only slightly sensible to the depth level of the Tukey region. Computational times which are less than one hour, averaged over 10 runs, are given in Table 1. The nonempty cells indicate those

¹Figures 3, 5 and 6 were generated with the QHULL software written at the Geometry Center, University of Minnesota.

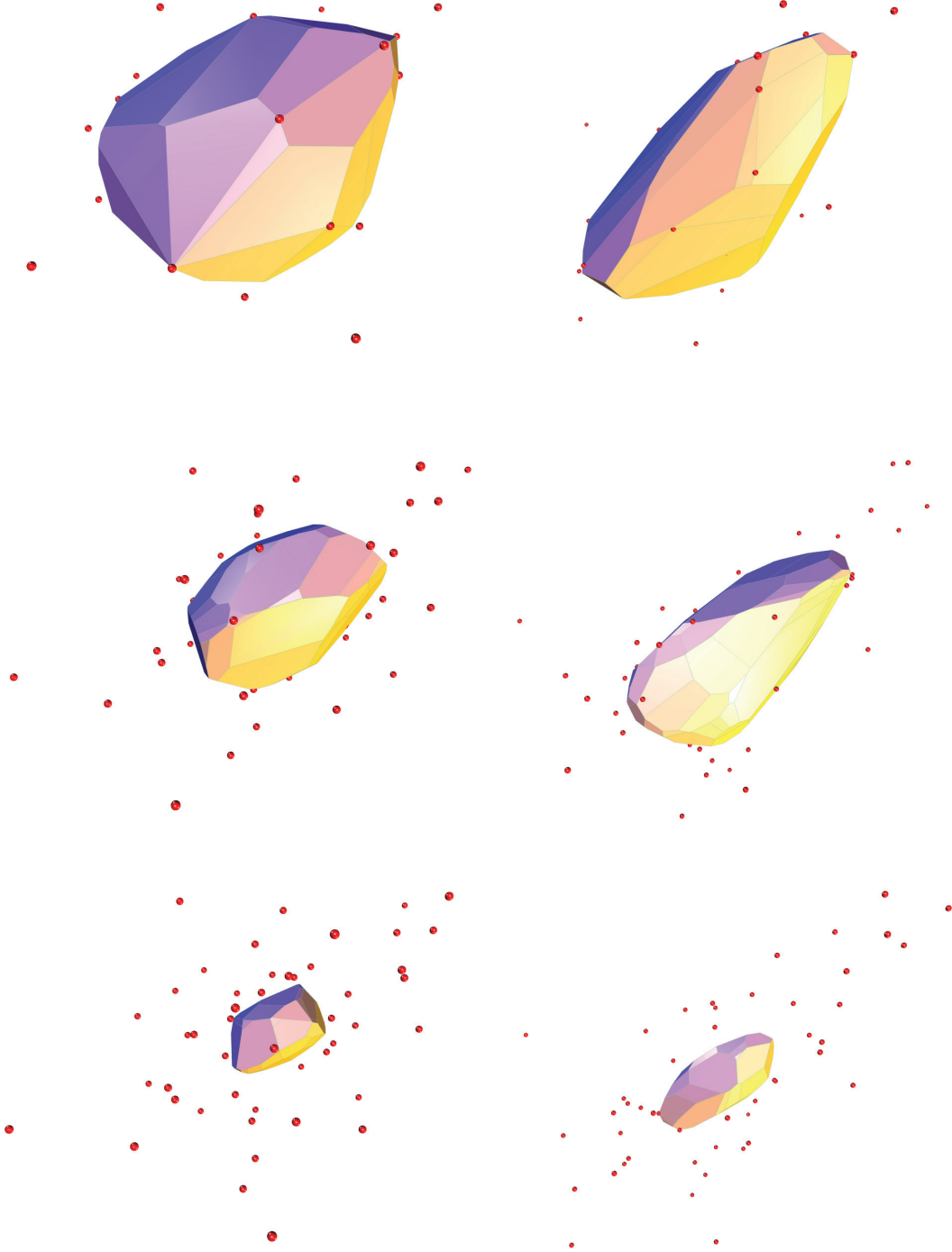


Figure 3: Tukey κ -regions of 50 observations from a standard normal distribution (left) and a skewed normal distribution (right), for $\kappa = 0.025$ (top), 0.1 (middle), 0.25 (bottom); see Section 4.1.

pairs (n, p) for which the results of the two algorithms regarding non-redundant hyperplanes and facets will be compared in order to verify the correctness of Algorithm 1.

For each of these 21 pairs (n, p) , selected by their computational feasibility, the following experiment is performed. We consider six distributions:

Table 1: Average computational times (in seconds, over 10 runs) of Algorithm 0, in case they are less than one hour.

$p \backslash$	n						
	40	80	160	320	640	1280	2560
3	0.00347	0.0264	0.184	1.44	11.7	95.2	792
4	0.0385	0.582	9.28	152	2480	—	—
5	0.395	12	387	—	—	—	—
6	3.05	196	—	—	—	—	—
7	18.8	2660	—	—	—	—	—
8	100	—	—	—	—	—	—
9	459	—	—	—	—	—	—

- multivariate standard normal distribution;
- elliptical Student- t distribution with five degrees of freedom without scaling;
- elliptical Cauchy distribution without scaling;
- uniform distribution on $[-1, 1]^p$;
- multivariate skewed-normal distribution with skewness parameter equal to 5 in the first coordinate according to Azzalini and Capitanio (1999);
- product of independent univariate exponential distributions having parameter 1.

For $p = 2$ and $n = 320$, a random samples generated from each of them is pictured in Figure 4.

For each of the six distributions we draw 100 random samples and compute the hyperplanes of Tukey regions (stopping after Step 5) with Algorithm 1 and with Algorithm 0. We do this for depth levels selected uniformly from $\{\frac{1}{n}, \dots, \frac{\lfloor 0.35n \rfloor}{n}\}$. In all $21 \times 6 \times 100 = 12600$ cases, the sets of the obtained hyperplanes coincided exactly. (The comparison is facilitated by the fact that each hyperplane is stored as a p -tuple of indices of points from the sample and thus can be matched exactly.)

4.2 Computational performance

Further, we measure the time taken by the execution of Algorithm 1. Computational times for couples (p, n) and several depth levels are indicated in Table 2. As above we restrict the calculations to those of no more than one hour for obtaining hyperplanes and facets. First, Table 2 demonstrates the superiority of Algorithm 1 over existing algorithms (*cf.* the times indicated by Paindaveine and Šiman, 2012a,b). Further, it demonstrates the applicability of the algorithm even for a substantial number of points (up to 5000) in dimension three (the still visualizable case) and its capability of computing bounding hyperplanes of Tukey regions even in dimension nine as well as calculating the exact shape (facets and vertices) of Tukey regions up to dimension six. Nevertheless, one observes an exponential increase of computational time, which indicates an eventual computational intractability of Algorithm 1 when the number of observations or the dimension of the data become even larger. One can also see that in higher dimensions, while bounding hyperplanes can be found, the QHULL algorithm fails in filtering them in reasonable time. Note that in all indicated cases an interior point has been rapidly found in Step 6 of Algorithm 1.

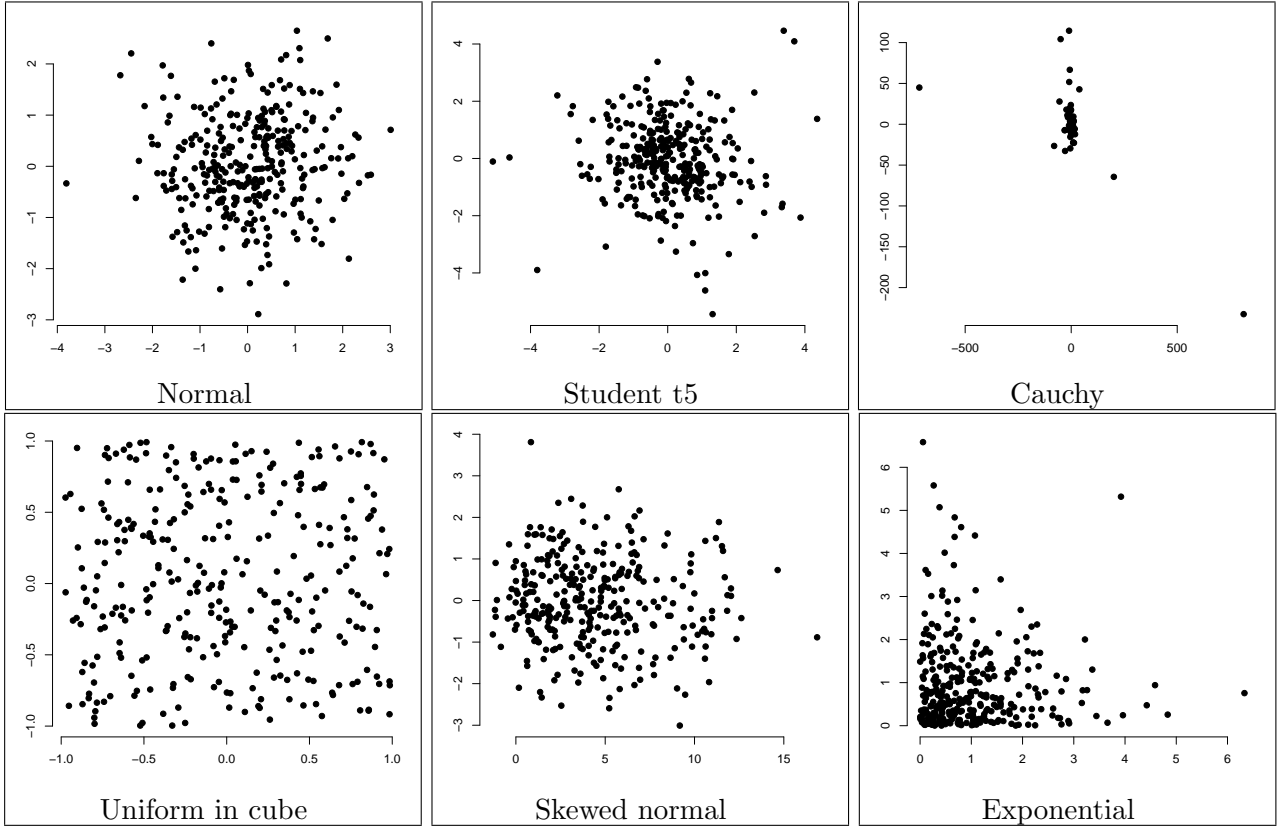


Figure 4: Data from six distributions; $p = 2$, $n = 320$.

Next, we take a closer look at the numbers of hyperplanes and Tukey region’s facets found by Algorithm 1. These results are summarized in Table 3. One observes that in most cases the number of facets of the Tukey region is substantially smaller than the number of all relevant hyperplanes (see Figure 5); *i.e.* most of the computing time is still utilized unnecessarily. (Note that this refers to the general result by Hallin et al. (2010) and the successive algorithms by Paidaveine and Šiman (2012a,b). Exceptional cases are limited to relatively small n , and thus do not constitute a computational problem. The results for $\kappa = 0.025 = 1/40$, $n = 40$, and $p = 6$ attract attention as there some of the found hyperplanes – though all belonging to the convex hull – are considered redundant. This happens due to precision merging of the QHULL algorithm in higher dimensions. The number of the hyperplanes itself, on the other hand, shows exponential growth in n and p , which limits the computational feasibility of the Algorithm 1. We further observe that in our experimental settings the upper bound on the facets is never achieved, ranging from close to 0.134 (in a few rare cases) to 0.0000258.

5 Computing the Tukey median

The *Tukey median* is the gravity center of the Tukey median set, that is, the Tukey region having maximum depth. With the above algorithms we are able to compute Tukey regions at

Table 2: Computational times (in seconds) of Algorithm 1, taken by finding all relevant hyperplanes (measured after Step 5) and by finding those non-redundant (measured after Step 7, in parentheses). For intractable cases, “t” indicates reaching the time limit (of one hour) and “m” reaching the memory limit (of 16 GB).

p	$\kappa \setminus$	Computational times							
		40	80	160	320	640	1280	2560	5120
3	0.025	0.000933	0.00149	0.00688	0.0424	0.307	2.24	17.8	143
		(0.000541)	(0.000969)	(0.00219)	(0.00662)	(0.0264)	(0.144)	(0.936)	(6.84)
	0.1	0.00125	0.00714	0.0436	0.308	2.38	19.1	154	1270
		(0.000903)	(0.00299)	(0.00966)	(0.041)	(0.219)	(1.29)	(8.45)	(61.7)
	0.2	0.00267	0.0155	0.105	0.753	5.92	48.4	389	3230
		(0.00156)	(0.00559)	(0.0231)	(0.102)	(0.55)	(3.36)	(22)	(158)
4	0.3	0.0039	0.0228	0.15	1.14	9.15	75	602	t
		(0.00242)	(0.0074)	(0.0316)	(0.155)	(0.855)	(5.16)	(33.3)	—
	0.025	0.00134	0.00809	0.0835	0.913	11.1	162	2440	t
		(0.00196)	(0.0139)	(0.0776)	(0.384)	(2.75)	(17.6)	(173)	—
	0.1	0.00966	0.0937	1.28	16.7	249	3850	t	—
		(0.017)	(0.0696)	(0.419)	(3.57)	(33.4)	(422)	—	—
5	0.2	0.0271	0.335	4.36	64.7	988	t	—	—
		(0.0253)	(0.16)	(1.29)	(12.7)	(132)	—	—	—
	0.3	0.0465	0.615	8.38	121	1880	—	—	—
		(0.056)	(0.233)	(2.28)	(23)	(294)	—	—	—
	0.025	0.0052	0.0473	0.717	13.7	322	t	—	—
		(0.0235)	(7.23)	(28.6)	(340)	(1990)	—	—	—
6	0.1	0.0627	1.16	26.2	695	t	—	—	—
		(1.05)	(8.25)	(81.3)	(590)	—	—	—	—
	0.2	0.24	5.55	149	t	—	—	—	—
		(0.512)	(5.87)	(88.5)	—	—	—	—	—
	0.3	0.448	11.9	340	—	—	—	—	—
		(2.39)	(6.97)	(154)	—	—	—	—	—
7	0.025	0.0181	0.307	6.19	201	m	—	—	—
		(1.85)	(t)	(t)	(t)	—	—	—	—
	0.1	0.318	12	497	m	—	—	—	—
		(1630)	(t)	(t)	—	—	—	—	—
	0.2	1.89	83.2	t	—	—	—	—	—
		(35.4)	(t)	—	—	—	—	—	—
8	0.3	4.76	214	—	—	—	—	—	—
		(219)	(1480)	—	—	—	—	—	—
	0.025	0.0636	1.35	50.1	m	—	—	—	—
		(t)	(t)	(t)	—	—	—	—	—
	0.1	1.63	101	m	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
9	0.2	10.9	1040	—	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
	0.025	0.202	6.51	m	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
	0.1	6.41	789	—	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
10	0.2	55.3	m	—	—	—	—	—	—
		(t)	—	—	—	—	—	—	—
	0.025	0.636	30.4	m	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
	0.1	25.7	m	—	—	—	—	—	—
		(t)	—	—	—	—	—	—	—
11	0.2	259	—	—	—	—	—	—	—
		(t)	—	—	—	—	—	—	—

Table 3: Number of relevant hyperplanes and those non-redundant (in parentheses) delivered by Algorithm 1. For intractable cases, “t” indicates reaching the time limit (of one hour) and “m” reaching the memory limit (of 16 GB).

p	$\kappa \setminus$	Number of directions							
		40	80	160	320	640	1280	2560	5120
3	0.025	28.6	94.6	369	1380	5280	20400	82000	326000
		(28.6)	(69.1)	(160)	(318)	(571)	(965)	(1650)	(2710)
	0.1	185	753	2890	11500	46000	184000	732000	2930000
		(69.9)	(143)	(268)	(441)	(765)	(1250)	(2060)	(3350)
	0.2	449	1830	7340	29100	116000	467000	1860000	7460000
		(66.4)	(124)	(227)	(362)	(606)	(1010)	(1650)	(2670)
	0.3	677	2720	11200	44700	181000	719000	2880000	t
		(42.3)	(70)	(132)	(234)	(388)	(637)	(1000)	—
	0.025	100	510	3250	21900	153000	1160000	9020000	t
		(100)	(360)	(1180)	(3250)	(7620)	(17100)	(36000)	—
4	0.1	1130	8180	61600	482000	3780000	30100000	t	—
		(343)	(966)	(2250)	(5170)	(11200)	(23400)	—	—
	0.2	3810	30000	236000	1890000	15100000	t	—	—
		(301)	(700)	(1640)	(3550)	(7680)	—	—	—
	0.3	6750	55500	448000	3580000	28800000	—	—	—
		(112)	(286)	(705)	(1610)	(3470)	—	—	—
	0.025	328	2550	24400	272000	3550000	t	—	—
		(328)	(1570)	(7920)	(29100)	(91700)	—	—	—
	0.1	5850	76100	1060000	15700000	t	—	—	—
		(1530)	(5780)	(18200)	(52800)	—	—	—	—
5	0.2	26100	396000	6140000	t	—	—	—	—
		(1020)	(3430)	(10600)	—	—	—	—	—
	0.3	52200	863000	13900000	—	—	—	—	—
		(143)	(758)	(2950)	—	—	—	—	—
	0.025	968	11900	169000	3110000	m	—	—	—
		(864)	(t)	(t)	(t)	—	—	—	—
	0.1	25200	603000	15200000	m	—	—	—	—
		(5650)	(t)	(t)	—	—	—	—	—
	0.2	143000	4320000	t	—	—	—	—	—
		(2930)	(t)	—	—	—	—	—	—
6	0.3	327000	11000000	—	—	—	—	—	—
		(84.6)	(1430)	—	—	—	—	—	—
	0.025	2630	48400	1070000	m	—	—	—	—
		(t)	(t)	(t)	—	—	—	—	—
	0.1	95300	3980000	m	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
	0.2	678000	39200000	—	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
	0.025	6930	189000	m	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
7	0.1	320000	24400000	—	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
	0.2	2790000	m	—	—	—	—	—	—
		(t)	—	—	—	—	—	—	—
	0.025	18800	670000	m	—	—	—	—	—
		(t)	(t)	—	—	—	—	—	—
	0.1	1040000	m	—	—	—	—	—	—
		(t)	—	—	—	—	—	—	—
	0.2	10300000	—	—	—	—	—	—	—
		(t)	—	—	—	—	—	—	—

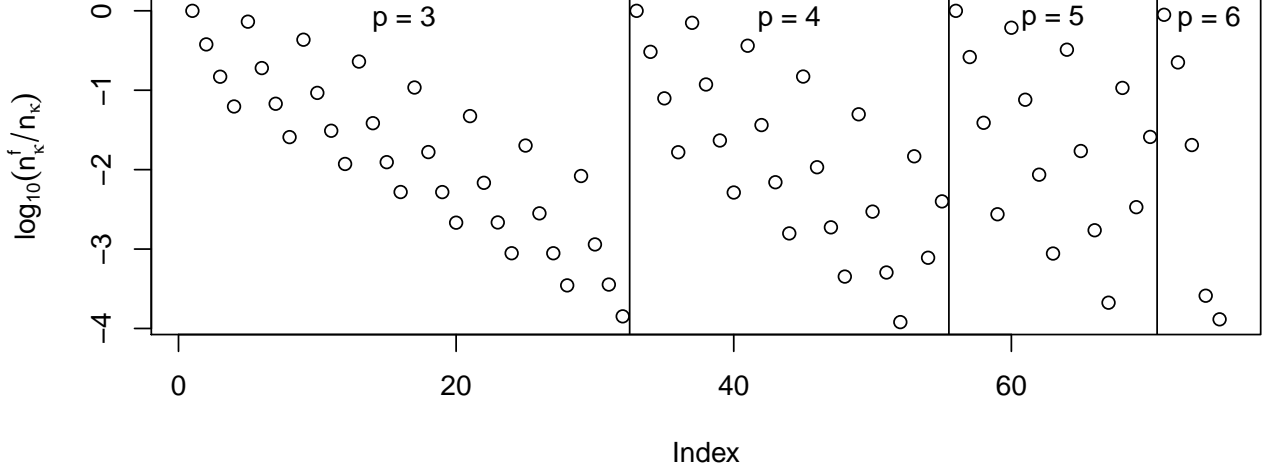


Figure 5: Logarithmized ratio of the number of facets of the Tukey region to the number of relevant hyperplanes, $\log_{10} \frac{n_{\kappa}^f}{n_{\kappa}}$. The index of the points is row-wise (κ increasing)/column-wise (n increasing) for each dimension $p = 3, 4, 5, 6$ of Table 3; *e.g.* sixth point stands for $(n, \kappa) = (80, 0.1)$ for $p = 3$.

any given depth level. However the maximum Tukey depth, κ^* , depends on the data and has to be determined from them. For data in general position it is known (Liu et al., 2017) that

$$\frac{1}{n} \left\lceil \frac{n}{p+1} \right\rceil \leq \kappa^* \leq \frac{1}{n} \left\lfloor \frac{n-p+2}{2} \right\rfloor. \quad (3)$$

In two-dimensional space, Rousseeuw and Ruts (1998), hereafter RR, have developed an efficient algorithm, which is of time complexity $O(n^2 \log^2(n))$. Their algorithm employs a bisection strategy that starts with a lower and an upper bound on κ^* and updates these bounds until they coincide. In updating, RR calculate the mean $\bar{\kappa}$ of the two active bounds and check whether $\mathcal{D}(\bar{\kappa})$ exists. If yes, the lower bound is changed to $\bar{\kappa}$; if no, the upper bound becomes $\bar{\kappa}$. The RR bisection approach extends easily to data of dimension greater than 2. With our new Algorithm 1 at hand, it is natural to investigate its possible use in computing higher-dimensional Tukey medians.

In searching for the maximum level of Tukey depth, we employ a modified bisection strategy. Given a Tukey region $\mathcal{D}(\kappa)$ at some level κ , we consider its barycenter. If the region is nonempty, its barycenter is more central than its boundary points, and thus may have a larger depth value. Observe that computing the depth of a single point is computationally cheaper (*i.e.* has lower time complexity) than computing a trimmed region. Hence, after having checked that at a given depth level κ the region $\mathcal{D}(\kappa)$ is nonempty, we may further compute the barycenter of $\mathcal{D}(\kappa)$ to possibly come closer to the depth maximum. This motivates us to construct the following algorithm. In the algorithm, $\text{Tukey_depth}(\mathbf{x}; \mathbf{x}_1, \dots, \mathbf{x}_n)$ stands for any procedure that computes the Tukey depth of a point $\mathbf{x} \in \mathbb{R}^p$ w.r.t. data $\mathbf{x}_1, \dots, \mathbf{x}_n \subset \mathbb{R}^p$, like those in Dyckerhoff and Mozharovskiy (2016) or Liu (2017). Further, $\text{Alg1}(\mathbf{x}; \mathbf{x}_1, \dots, \mathbf{x}_n; \kappa; \epsilon)$ signifies a Tukey region resulting from Algorithm 1.

Algorithm 2 (Algorithm for computing the Tukey median).

Input: $\mathbf{x}_1, \dots, \mathbf{x}_n \subset \mathbb{R}^p$, ϵ .

Step 1. Initialize bounds on κ^* :

- (a) Compute $\mathbf{x}_0 = (\text{med}(\mathbf{x}_{11}, \dots, \mathbf{x}_{n1}), \text{med}(\mathbf{x}_{12}, \dots, \mathbf{x}_{n2}), \dots, \text{med}(\mathbf{x}_{1d}, \dots, \mathbf{x}_{nd}))^\top$.
- (b) Compute $d_0 = \text{Tukey_depth}(\mathbf{x}_0; \mathbf{x}_1, \dots, \mathbf{x}_n)$.
- (c) Set $\kappa_{low} = \max \left\{ \frac{1}{n} \lceil \frac{n}{p+1} \rceil, d_0 \right\}$, $\kappa_{up} = \frac{1}{n} \lfloor \frac{n-p+2}{2} \rfloor + \frac{1}{n}$.

Step 2. Update bounds:

- Let $\bar{\kappa} = \frac{1}{n} \lfloor \frac{n(\kappa_{low} + \kappa_{up})}{2} \rfloor$, and compute the region $\mathcal{D}(\bar{\kappa}) = \text{Alg1}(\mathbf{x}_1, \dots, \mathbf{x}_n; \bar{\kappa}; \epsilon)$.
- (a) If $\mathcal{D}(\bar{\kappa})$ does not exist (that is, Algorithm 1 stops at its Step 6), then set $\kappa_{up} = \bar{\kappa}$,
 - (b) If $\mathcal{D}(\bar{\kappa})$ exists, then calculate the barycenter \mathbf{c} of $\mathcal{D}(\bar{\kappa})$ and set $\kappa_{low} = \text{Tukey_depth}(\mathbf{c}; \mathbf{x}_1, \dots, \mathbf{x}_n)$.
 - (c) If $\kappa_{low} < \kappa_{up} - \frac{1}{n}$, then repeat **Step 2**, else stop.

Output: $\text{Alg1}(\mathbf{x}_1, \dots, \mathbf{x}_n; \kappa_{low}; \epsilon)$.

Step 1c initializes the lower bound either according to (3) or with the depth of the coordinate-wise median, which is cheaply computed and often has a rather high depth value. In Step 2b, the depth of the barycenter of the Tukey region is computed (in case it exists) to obtain a higher value for the lower bound, which substantially contributes to the speed of Algorithm 2.

We compare our Algorithm 2 with the bisection approach of RR in a small simulation study based on symmetrical as well as non-symmetrical data. Only those cases are considered which are computed in less than one hour. The computing times of the two algorithms are reported in Tables 4 and 5. It is seen that Algorithm 2 runs slower than RR bisection when the sample size is relatively small. But it substantially outperforms the RR approach as the sample size increases. In fact, when n is small, κ^* is rather easily found as its upper and lower bounds given in (3) are tight, while Algorithm 2 uses much time in Step 2b to compute barycenters. Therefore, we recommend the new Algorithm 2 for larger sample sizes n , while the RR bisection approach is to be preferred for smaller n .

Table 4: Average time (in seconds, over 10 runs) to compute the Tukey median with the RR bisection approach and with Algorithm 2, given p -variate standard normal data.

Algorithm	p	n				
		40	80	160	320	640
RR bisection	3	0.0314	0.324	2.35	47.9	679
Algorithm 2	3	0.0339	0.242	1.93	41	537
RR bisection	4	0.71	33.1	723	—	—
Algorithm 2	4	0.872	35	892	—	—
RR bisection	5	16.5	—	—	—	—
Algorithm 2	5	25.5	—	—	—	—

The positions of the sample mean, the coordinate-wise median and the Tukey median can be quite different. To illustrate this, we provide a three-dimensional artificial data set in Table 6, which is inspired by (Rousseeuw and Leroy, 1987, chapter 7.1a). The data is constructed in

Table 5: Average time (in seconds, over 10 runs) to compute the Tukey median with the RR bisection approach and with Algorithm 2, given p -variate skewed normal data as in Section 4.1.

Algorithm	p	n				
		40	80	160	320	640
RR bisection	3	0.058	0.988	17.3	398	2570
Algorithm 2	3	0.0606	0.763	12.9	284	1060
RR bisection	4	2.51	347	2000	—	—
Algorithm 2	4	3.05	295	553	—	—
RR bisection	5	108	—	—	—	—
Algorithm 2	5	142	—	—	—	—

Table 6: An artificial data set for illustrating the different location of the sample mean, the coordinate-wise median and the Tukey median.

#	x_1	x_2	x_3
1	1	0	0
2	0	1	0
3	0	0	1
4	1.5	1.5	1.5
5	0.309151	0.286697	0.653584
6	0.733359	0.040291	0.316318
7	0.15937	0.304677	0.558091
8	0.056376	0.19044	0.912733
9	0.517479	0.533977	0.19188
10	1.011993	0.058608	0.099067
11	0.117582	0.164475	0.92203
12	0.175112	0.918897	0.221602
13	0.240206	0.454373	0.1701
14	0.906328	0.056292	0.11981

location estimators	x_1	x_2	x_3	depth
mean	0.480497	0.39348	0.476087	$\frac{1}{14}$
coordinate-wise median	0.274678	0.238568	0.26896	0
Tukey median	0.453501	0.270065	0.413392	$\frac{2}{7}$

the following way: The first three points correspond to the three canonical unit vectors. The fourth point has coordinates $(1.5, 1.5, 1.5)^\top$; it represents an outlier. Further ten points have coordinates $Z + U \cdot (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^\top$, where Z is a random vector having a Dirichlet distribution with parameters $(1, 1, 1)$ and U is a random variable uniformly distributed on $[-\frac{1}{4}, \frac{1}{4}]$. The sample mean, the coordinate-wise median and the Tukey median are given in Table 6, together with their depths in the data cloud. Table 6 indicates that the coordinate-wise median lies outside the convex hull of the data set, while the sample mean – being sensitive to the outlier – lies outside the convex hull of the main data (= data without the outlier at $(1.5, 1.5, 1.5)^\top$), which is visualized in Figure 6. Moreover, the mean and the component-wise median are found on opposite sides. On the other hand, the Tukey median lies in the middle of the convex hull of the main data.

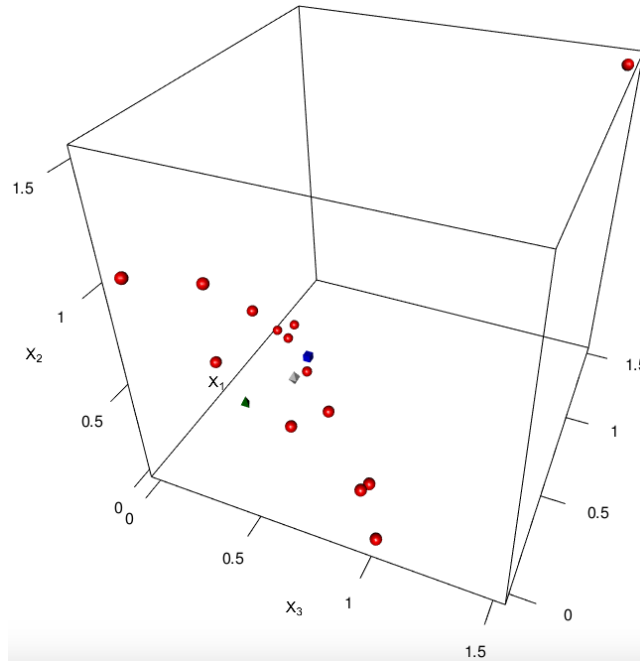


Figure 6: Data (red spheres), with mean (blue cube), coordinate-wise median (green tetrahedron) and Tukey median (gray rhombus); see Table 6.

6 Concluding remarks

Two new algorithms have been constructed for computing a κ -trimmed Tukey region when $p > 2$. While the first algorithm (Algorithm 0) comes without a particular search strategy, special search rules have been implemented in the second algorithm (Algorithm 1). Different to the approaches suggested by Hallin et al. (2010) and Kong and Mizera (2012), Algorithm 1 exploits neither cone segmentation of \mathbb{R}^p nor univariate projections of the data. Employing the breadth-first ridge-by-ridge search strategy it focuses purely on the location of the data and saves both computational time on creating and searching vast structures as well as memory on storing them. Especially, implementations employing linear programming and post-optimization (e.g., Paindaveine and Šiman (2012a,b)) occupy memory with their structures and need time for their creation and processing, which is completely avoided in Algorithm 1. Storing a facet of a direction cone as an average of vertices in \mathbb{R}^p needs $8p^2$ bytes of physical memory (see Paindaveine and Šiman (2012a,b)), while in Algorithm 1 storing a ridge can be efficiently reduced to $\lceil \frac{p-1}{8} \log_2 n \rceil$ bytes, which is substantially smaller as long as approximately (in the order of magnitude) $n < 2^{64p}$. This explains the high computational time efficiency of Algorithm 1, which has been numerically demonstrated in Section 4.2. In addition, Algorithm 1 finds the region's polytope and not just the relevant hyperplanes. It includes the search for an inner point, the computation of the halfspaces' intersection and the determination of the region's facets and vertices. The latter tasks can even take a major part of its computation time; see, for example, $p \geq 5$ in Table 2.

Throughout the article, the data is assumed to be in general position. Violation of this assumption can be compensated by a slight perturbation of the data, which – due to the fact that the Tukey trimmed region is defined by hyperplanes through data points – influences location and shape of the Tukey regions just negligibly. Note that in Step 6c of Algorithm 1

only regions with positive (subject to the precision constant ϵ) volume are identified, otherwise they are output as empty sets.

The simulation study of Section 4 reveals important aspects of Algorithm 1. Firstly, its broad numerical comparison with Algorithm 0, which is obviously correct, substantiates that Algorithm 1 is correct as well. Secondly, Table 2 demonstrates its high speed. Thirdly, Table 3 and Figure 5 show that, on an average, facets of a Tukey region are contained in much fewer hyperplanes than those found in Steps 1–5 of Algorithm 1. Algorithm 1 shares this feature with the developments of HPS. This still leaves room for potential improvements.

Also an algorithm (Algorithm 2) to calculate the Tukey median, which is the gravity center of the innermost Tukey region, has been constructed. To our knowledge this is the first algorithm that solves the task for dimension $p > 2$. Algorithm 2 is an enriched bisection procedure that employs Algorithm 1 and an algorithm calculating the Tukey depth of a point to speed up the search for maximum Tukey depth. Its computation time is compared with a straightforward extension of the bisection approach by Rousseeuw and Ruts (1998). It turns out that the latter is strongly outperformed by Algorithm 2 when the sample size is large.

The literature contains many other depth notions, such as the projection depth and others, which, similar to the Tukey depth, satisfy the projection property, that is, are equal to the minimum of univariate depths in any direction (Dyckerhoff, 2004). It turns out that several of them can be computed by cutting convex polytopes with hyperplanes; see Mosler et al. (2009) and Liu and Zuo (2014) for details. By this, similar algorithms may be constructed to calculate the respective central regions in dimensions $p > 2$.

Algorithms 0, 1 and 2, including proper visualization procedures, are implemented in the R-package `TukeyRegion`.

References

- AZZALINI, A. AND CAPITANIO, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B* **61**(3) 579–602.
- BARBER, C. B., DOBKIN, D. P., AND HUHDANPAA, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software* **22**(4) 469–483.
- DONOHU, D. L. (1982). *Breakdown Properties of Multivariate Location Estimators*. PhD thesis, Harvard University.
- DYCKERHOFF, R. (2004). Data depths satisfying the projection property. *AStA - Advances in Statistical Analysis* **88**(2) 163–190.
- DYCKERHOFF, R. AND MOZHAROVSKIY, P. (2016). Exact computation of the halfspace depth. *Computational Statistics and Data Analysis* **98** 19–30.
- HALLIN, M., PAINDAVEINE, D., AND ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *The Annals of Statistics* **38**(2) 635–669.
- HUBERT, M., ROUSSEEUW, P. J., AND SEGHAERT, P. (2015). Multivariate functional outlier detection. *Statistical Methods and Applications* **24**(2) 177–202.

- KONG, L. AND MIZERA, I. (2012). Quantile tomography: using quantiles with multivariate data. *Statistica Sinica* **22**(4) 1589–1610.
- KONG, L. AND ZUO, Y. (2010). Smooth depth contours characterize the underlying distribution. *Journal of Multivariate Analysis* **101**(9) 2222–2226.
- KOSHEVOY, G. A. (2002). The Tukey depth characterizes the atomic measure. *Journal of Multivariate Analysis* **83** 360–364.
- KOSHEVOY, G. AND MOSLER, K. (1997). Zonoid trimming for multivariate distributions. *The Annals of Statistics* **25** 1998–2017.
- LIU, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics* **18** 405–414.
- LIU, R. Y. (1992). Data depth and multivariate rank tests. In: Dodge, Y. (ed.) *L₁-Statistical Analysis and Related Methods* 279–294, Elsevier, Amsterdam.
- LIU, X. (2017). Fast implementation of the Tukey depth. *Computational Statistics* **32** 1395–1410.
- LIU, X. AND ZUO, Y. (2014). Computing projection depth and its associated estimators. *Statistics and Computing* **24** 51–63.
- LIU, X., LUO, S., AND ZUO, Y. (2017). Some results on the computing of Tukey’s halfspace median. *Statistical Papers*, in print.
- MILLER, K., RAMASWAMI, S., ROUSSEEUW, P. J., SELLARÈS, J. A., SOUVAINE, D., STREINU, I., AND STRUYF, A. (2003). Efficient computation of location depth contours by methods of computational geometry. *Statistics and Computing* **13**(2) 153–162.
- MOSLER, K. (2013). Depth statistics. In: Becker, C., Fried, R., and Kuhnt, S. (eds.) *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather* 17–34, Springer, Berlin.
- MOSLER, K., LANGE, T., AND BAZOVKIN, P. (2009). Computing zonoid trimmed regions of dimension $d > 2$. *Computational Statistics and Data Analysis* **53**(7) 2500–2510.
- MOZHAROVSKIY, P., MOSLER, K., AND LANGE, T. (2015). Classifying real-world data with the $DD\alpha$ -procedure. *Advances in Data Analysis and Classification* **9** 287–314.
- PAINDAVEINE, D. AND ŠIMAN, M. (2011). On directional multiple-output quantile regression. *Journal of Multivariate Analysis* **102**(2) 193–212.
- PAINDAVEINE, D. AND ŠIMAN, M. (2012a). Computing multiple-output regression quantile regions. *Computational Statistics and Data Analysis* **56**(4) 840–853.
- PAINDAVEINE, D. AND ŠIMAN, M. (2012b). Computing multiple-output regression quantile regions from projection quantiles. *Computational Statistics* **27**(1) 29–49.
- ROUSSEEUW, P. J. AND LEROY, A. M. (1987). *Robust regression and outlier detection*. Wiley, New York.

- ROUSSEEUW, P. J. AND RUTS, I. (1998). Constructing the bivariate Tukey median. *Statistica Sinica* **8**(3) 827–839.
- ROUSSEEUW, P. J. AND STRUYF, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing* **8**(3) 193–203.
- RUTS, I. AND ROUSSEEUW, P. J. (1996). Computing depth contours of bivariate point clouds. *Computational Statistics and Data Analysis* **23**(1) 153–168.
- SERFLING, R. (2006). Depth functions in nonparametric multivariate inference. In: Liu, R. Y., Serfling, R., and Souvaine, D. L. (eds.) *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, 1–16. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **72**, American Mathematical Society, Providence.
- STRUYF, A. J. AND ROUSSEEUW, P. J. (1999). Halfspace depth and regression depth characterize the empirical distribution. *Journal of Multivariate Analysis* **69** 135–153.
- TUKEY, J. W. (1975). Mathematics and the picturing of data. In: James, R. D. (ed.) *Proceedings of the International Congress of Mathematicians (Volume 2)* 523–531, Canadian Mathematical Congress.
- YEH, A. B. AND SINGH, K. (1997). Balanced confidence regions based on Tukey’s depth and the bootstrap. *Journal of the Royal Statistical Society: Series B* **59**(3) 639–652.
- ZUO, Y. (2003). Projection-based depth functions and associated medians. *The Annals of Statistics* **31**(5) 1460–1490.
- ZUO, Y. AND SERFLING, R. (2000). General notions of statistical depth function. *The Annals of Statistics* **28** 461–482.