Cite this article as:
Đ. Baljozović, M. Merkle, ***Spatial medians, depth functions and multivariate Jensen's inequality*****, 2007, http://arxiv.org/abs/math/0701922**

Spatial medians, depth functions and multivariate Jensen's inequality

Đorđe Baljozović and Milan Merkle

Abstract. For any given partial order in a d-dimensional euclidean space, under mild regularity assumptions, we show that the intersection of closed (generalized) intervals containing more than 1/2 of the probability mass, is a non-empty compact interval. This property is shared with common intervals on real line, where the intersection is the median set of the underlying probability distribution. So obtained multivariate medians with respect to a partial order, can be observed as special cases of centers of distribution in the sense of type D depth functions introduced by Y. Zuo and R. Serfling, *Ann. Stat.*, **28** (2000), 461-482. We show that the halfspace depth function can be realized via compact convex sets, or, for example, closed balls, in place of halfspaces, and discuss structural properties of halfspace and related depth functions and their centers. Among other things, we prove that, in general, the maximal guaranteed depth is $\frac{1}{d+1}$. As an application of these results, we provide a Jensen's type inequality for functions of several variables, with medians in place of expectations, which is an extension of the previous work by M. Merkle, *Stat. Prob. Letters*, **71** (2005), 277–281.

2000 Mathematics Subject Classification. 62H05, 60E15

Key words and phrases. Tukey's median, depth function, halfspace depth, Jensen's inequality, multivariate median

Acknowledgements. This work is partially supported by Ministry of Science and environmental protection of Serbia, under the project number 144021. Second author acknowledges his partial affiliation to Računarski fakultet, Beograd, Serbia, and Instituto de Mathematica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

University of Belgrade, Faculty of Electrical Engineering, P.O. Box 35-54, 11120 Belgrade, Serbia
emerkle@etf.bg.ac.yu
djordje.baljozovic@etf.bg.ac.yu

Spatial medians, depth functions and multivariate Jensen's inequality

Đorđe Baljozović and Milan Merkle

1 Introduction.

To attain any median m of a real valued random variable X, we have to pass at least half of the population, coming from either side of the real axis, via the relations

$$\text{Prob}(X \in (-\infty, m]) \geq p, \quad \text{Prob}(X \in [m, +\infty)) \geq 1-p, \qquad p = \frac{1}{2}. \quad (1.1)$$

The proportion $\frac{1}{2}$ is the largest possible, in the sense that there does not exist a $p > \frac{1}{2}$ such that for arbitrary distribution of X there exists $m \in \mathbf{R}$ which satisfies (1.1). We say that a median is the *deepest* point with respect to a given distribution, or a data set. Quantitatively, we may assign a *depth* to each point $x \in \mathbf{R}$, according to the value of a *depth function*

$$D(x; P) = \inf\{P((-\infty, x]), P([x, +\infty))\}, \quad P(S) := \text{Prob}(X \in S), \quad (1.2)$$

from where we can also see that the function D attains its maximum $1/2$ in the set of median points.

The literature devoted to the problem of extension of these observations to higher dimensions contains a variety of different approaches - see [19] or [27] for a comprehensive bibliography. In a high dimensional data set it is desirable to select one point, or a set of points that would correspond to intuitive notions of "deepest point", "most central point", or "the center of a data set", and can serve as a (best) representative or a reference point of a distribution or a data set.

The definition (1.2) of a depth function includes a notion of direction: we can approach a point $x \in \mathbf{R}$ from either left or right. The notion of direction in $\mathbf{R}^d$ is related to partial order. In this paper we show, among other things, that with respect to any partial order that satisfies mild regularity conditions, there is a well defined median set, in the sense that the depth (to be defined in Section 3) of each point of that median set is at least 1/2. The median set is always compact, and can be obtained as an intersection of a certain family of sets.

The median sets defined with respect to a given direction (i.e., a partial order) are not affine invariant. On the other hand, Tukey's median, or halfspace median (first introduced in [20], see also [2] or [27]), which can be defined as the set of maximal depth with respect to the depth function (commonly referred to as the halfspace depth)

$$D(x; P) = \inf\{P(H) \mid H\text{: any halfspace that contains } x\}, \tag{1.3}$$

is affine invariant, but the maximal depth is not guaranteed to be 1/2. As a generalization of (1.3), Zuo and Serfling in [27] offered a general "depth function of type D", which is defined as

$$D(x; P, \mathcal{C}) = \inf\{P(C) \mid x \in C \in \mathcal{C}\}, \tag{1.4}$$

where $\mathcal{C}$ is a given collection of closed sets. A similar concept was introduced in [18]. In this paper we observe the function (1.4) defined with respect to arbitrary class $\mathcal{U}$ of open sets in place of closed sets in $\mathcal{C}$. This slight change enables a considerable reduction of conditions for validity of certain results, and yields a collection of interesting and enlightening examples. We prove several general results for a class $\mathcal{U}$ under two very mild conditions, and we show that for each such class, the set of deepest points can be obtained as the intersection of a family of sets. It turns out that median sets can be also obtained as points having the depth not less than 1/2, with a special choice of $\mathcal{U}$. In this paper, by a multidimensional median set we understand only the set of points with the depth not less than 1/2. It can happen that such a set is empty; in general case we use use the term "center of the distribution" for the set of deepest points. These two cases need to be distinguished, as the corresponding sets have different properties.

We show that in general, with the type D depth in $\mathbf{R}^d$ we can be certain to find only point(s) with $D(x; P) \geq 1/(d+1)$, and we show that this bound is the best possible, thus extending a result from [4] for the halfspace depth.

Finally, we show that each depth function generated by a family $\mathcal{U}$ of complements of compact convex sets, can be defined in terms of a family of halfspaces, and we represent the halfspace center ("Tukey median") via intersection of level sets of depth functions defined with respect to partial orders. In the last section, we discuss an analogue of Jensen's inequality for functions of several variables, with medians (or, generally, points in the center of distribution) in place of expectations.

The results presented in this paper are fully general, and hold without any particular assumptions about the underlying distribution.

For more details regarding Tukey's median see [3], [4] and [27]. The latter paper and [9] give wide-ranging discussion of depth functions in general with numerous examples and a summary of further researches based on the halfspace depth.

In papers [7, 8] the simplical depth function was proposed, primarily important for its corresponding version of multivariate median; also, [8] highlighted some features the (simplical) depth function should fulfill.

Numerous other depth functions have been introduced in the literature - some of the most relevant are presented in [14], [10], [11], [6], [9], [27, 28, 29], [21], [23], [24], [26], [25] and references therein. There are methods and techniques developed especially for the data sets, see for example, [1]. Robustness properties of deepest points based on halfspace depth in context of location statistics are studied in [2].

Other approaches in defining multivariate centers are given in [19] (many of them not related to depth functions), [13], [21], [5] and references therein; as can be seen, this concept is not at all unambiguous like in the univariate case.

Notations. Throughout the paper, $\mathbf{R}$ denotes the set of real numbers, and $\overline{\mathbf{R}}$ is the set $\mathbf{R}$ together with $\pm\infty$. The symbols $\mathbf{R}^d$ and $\overline{\mathbf{R}}^d$ denote corresponding d-dimensional euclidean spaces (Points in $\overline{\mathbf{R}}^d$ are allowed to have $\pm\infty$ as coordinates). In Section 2, and in the material of later sections related to intervals, the points in $\overline{\mathbf{R}}^d$ with $d > 1$ will be denoted by bold letters $\boldsymbol{x}, \boldsymbol{y}, \ldots$ and their coordinates by $x_i, y_i, \ldots$, respectively. We say that the set $S \in \mathbf{R}^d$, $d \geq 1$ is closed if it is closed in euclidean topology. Hence, for $d = 1$, the intervals $[a, +\infty)$ or $(-\infty, a]$ are closed for any $a \in \mathbf{R}$; a similar remark holds for generalized intervals with $d > 1$. We will observe $\mathbf{R}^d$-valued random variables X, considering them as being measurable maps from some abstract probability space $(\Omega, \mathcal{F}, \mathrm{Prob})$ to $(\mathbf{R}^d, \mathcal{B}^d, P)$. Here $\mathcal{B}^d$ is the Borel sigma-field on $\mathbf{R}^d$ and P is a probability measure induced on $\mathbf{R}^d$ by X (the probability distribution of X).

For a random variable X, $\mathrm{Med}\, X$ denotes any of its medians, and $\{\mathrm{Med}\, X\}$ denotes the set of all medians. In the same way, we may talk about medians of a probability distribution P.

2 Multivariate medians.

We start with a characteristic property of univariate median set. Let X be a random variable with a probability distribution P and let J be any closed interval with $P(J) > 1/2$. We will show that J contains every median of X. Indeed, if m is a median of X and $m \notin J$, then one of the intervals $(-\infty, m]$ or $[m, +\infty)$ is disjoint with J, which is not possible, since the sum of probabilities in both cases is greater than 1. Therefore, the intersection

$$\bigcap_{J=[a,b]:\ P(J)>1/2} J$$

is nonempty, and it contains the median set $[u, v]$ of X. Now, observe that for $J_{2n-1} = (-\infty, v + \frac{1}{(2n-1)}]$ and $J_{2n} = [u - \frac{1}{2n}, +\infty)$, $n = 1, 2, \ldots$ we have that $P(J_n) > 1/2$ and so

$$\{\text{Med}\, X\} = [u, v] = \bigcap_{n=1}^{+\infty} J_n \supset \bigcap_{J=[a,b]:\ P(J)>1/2} J,$$

which together with the previous part, shows that

$$\{\text{Med}\, X\} = \bigcap_{J=[a,b]:\ P(J)>1/2} J \tag{2.1}$$

The relation (2.1) can be as well taken as a definition of the median set for a given distribution, and this definition can be extended in a multidimensional environment if we choose one of many possible extensions of the concept of one-dimensional interval. Out of several ones that we may think of (convex sets, star-shaped sets, balls and other special convex sets), only intervals with respect to a partial order can do the work, to ascertain non-emptiness of the intersection at the right hand side of (2.1).

Let $\preceq$ be a partial order in $\overline{\mathbf{R}}^d$ and let $\boldsymbol{a}, \boldsymbol{b}$ be arbitrary points in $\overline{\mathbf{R}}^d$. We define a d-dimensional interval $[\boldsymbol{a}, \boldsymbol{b}]$ as the set of points in $\mathbf{R}^d$ that are between $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbf{R}^d \mid \boldsymbol{a} \preceq \boldsymbol{x} \preceq \boldsymbol{b}\}$$

Note that the interval can be an empty set, or a singleton. For the sake of simplicity, we want all intervals to be topologically closed. The interval can be norm bounded or norm unbounded; it would be reasonable to expect intervals with finite "endpoints" to be norm bounded, hence compact. Further, we would expect that intervals can be "big" as we wish, to contain any ball or any compact set. Finally, we expect that bounded (with respect to partial order) sets posses the least upper bound and greatest lower bound. To summarize, we assume the following three technical conditions:

(I1) Any interval $[\boldsymbol{a}, \boldsymbol{b}]$ is topologically closed, and for any $\boldsymbol{a}, \boldsymbol{b} \in \mathbf{R}^d$ (i.e., with finite coordinates), the interval $[\boldsymbol{a}, \boldsymbol{b}]$ is a compact set.

(I2) For any ball $B \subset \mathbf{R}^d$, there exist $\boldsymbol{a}, \boldsymbol{b} \in \mathbf{R}^d$ such that $B \subset [\boldsymbol{a}, \boldsymbol{b}]$.

(I3) For any set S which is bounded from above with a finite point, there exists a finite $\sup S$. For any set S which is bounded from below with a finite point, there exists a finite $\inf S$.

Example 2.1. Let K be a closed convex cone in $\mathbf{R}^d$, with vertex at origin, and suppose that there exists a closed hyperplane π, such that $\pi \cap K = 0$ (that is, $K \setminus \{0\}$ is a subset of one of open halfspaces determined by π). Define the relation $\preceq$ by $\boldsymbol{x} \preceq \boldsymbol{y} \iff \boldsymbol{y} - \boldsymbol{x} \in K$. The interval is then

$$[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \mid \boldsymbol{x} - \boldsymbol{a} \in K \wedge \boldsymbol{b} - \boldsymbol{x} \in K\} = (\boldsymbol{a} + K) \cap (\boldsymbol{b} - K).$$

If the endpoints have some coordinates infinite, then the interval is either $\boldsymbol{a} + K$ (if $\boldsymbol{b} \notin \mathbf{R}^d$) or $\boldsymbol{b} - K$ (if $\boldsymbol{a} \notin \mathbf{R}^d$) or $\mathbf{R}^d$ (if neither endpoint is in $\mathbf{R}^d$).

It is not difficult to show that $\preceq$ is a partial directed order, and that it satisfies conditions (I1)–(I3).

The simplest, coordinate-wise ordering, can be obtained with K chosen to be the orthant with $x_i \geq 0, i = 1, \ldots, d$. Then

$$\boldsymbol{x} \preceq \boldsymbol{y} \iff x_i \leq y_i, \quad i = 1, \ldots, d. \tag{2.2}$$

For the sake of illustration, let us note that possible kinds of intervals with respect to the relation (2.2) include:

$$[(a_1, a_2), (b_1, b_2)],\ [(a_1, a_2), (b_1, +\infty)],\ [(a_1, a_2), (+\infty, b_2)],$$

$$[(a_1, a_2), (+\infty, +\infty)],\ [(-\infty, -\infty), (b_1, b_2)],\ [(a_1, -\infty), (+\infty, b_2)],$$

where a_1, a_2, b_1, b_2 are real numbers. For infinite endpoints we use strict inequalities, for example the last interval above is the set of $(x, y) \in \mathbf{R}^2$ such that $a_1 \leq x < +\infty$ and $-\infty < y \leq b_2$. The intervals may be empty; for example, the first listed interval is empty if $a_1 > b_1$ or if $a_2 > b_2$.

□

The following theorem extends the one-dimensional property discussed in the beginning of this section.

Theorem 2.1. *Let $\preceq$ be a partial order in $\overline{\mathbf{R}}^d$ such that conditions (I1)–(I3) hold. Let P be a probability measure on $\mathbf{R}^d$ and let $\mathcal{J}$ be a family of intervals with respect to a partial order $\preceq$, with the property that*

$$P(J) > \frac{1}{2}, \quad \text{for each } J \in \mathcal{J}. \tag{2.3}$$

Then the intersection of all intervals from $\mathcal{J}$ is a non-empty compact interval.

The compact interval claimed in the Theorem 2.1 can be, in analogy to (2.1), taken as a definition of the median induced by the partial order $\preceq$:

$$\{\mathrm{Med}\,\boldsymbol{X}\}_{\preceq} \quad = \bigcap_{J=[\boldsymbol{a},\boldsymbol{b}]:\ P_{\boldsymbol{X}}(J)>1/2} J, \tag{2.4}$$

where $\boldsymbol{X}$ is a random variable on $\mathbf{R}^d$ and $P_{\boldsymbol{X}}$ is its probability distribution. In what follows, we will omit the subscript if the underlying relation $\preceq$ is obvious.

It is shown in Appendix (Lemma A.2) that, in the case of the coordinate-wise partial order, the characterization (2.4) is equivalent to a similar characterization given in [12]. It turns out that, in this case, the median set is just the Cartesian product of coordinate-wise median sets, which is the result stated in the form of an example in [18].

Theorem 2.2. *Let the partial order $\preceq$ in $\mathbf{R}^d$, $d > 1$, be defined by (2.2), and let $\{\mathrm{Med}\,\boldsymbol{X}\}$ be the median set of a random vector $\boldsymbol{X} \in \mathbf{R}^d$, defined with respect to the partial order $\preceq$. Then*

$$\{\mathrm{Med}\,\boldsymbol{X}\} = \{\mathrm{Med}\,X_1\} \times \{\mathrm{Med}\,X_2\} \times \cdots \times \{\mathrm{Med}\,X_d\} \tag{2.5}$$

As we already mentioned, for other classes of sets, the intersection in (2.1) is in general empty. However, in the next section we will see that some other classes can also have a non-empty intersection, but then, instead of having the measure $> 1/2$, the sets will generally have to have a greater measure.

3 Depth functions.

We start with a collection of sets $\mathcal{V}$ and the collection $\mathcal{U}$ that contains complements of sets from $\mathcal{V}$, i.e., $\mathcal{U} = \{S^c \mid S \in \mathcal{V}\}$. For each $x \in \mathbf{R}^d$ and any probability measure P on Borel sets of $\mathbf{R}^d$, define a depth function

$$D(x; P, \mathcal{U}) = \inf\{P(U) \mid x \in U \in \mathcal{U}\}. \tag{3.1}$$

The role of $\mathcal{V}$ will be clear later in this section, when we give an alternative description of the depth function in terms of sets in $\mathcal{V}$.

If for some x, there does not exist any $U \in \mathcal{U}$ that contains x, on the right hand side of (3.1) we have empty set, and then $D(x; P, \mathcal{U}) = +\infty$. To avoid this, we assume that

$$(C_1) \quad \text{for every } x \in \mathbf{R}^d \text{ there is a } U \in \mathcal{U} \text{ so that } x \in U.$$

Further, a constant depth function does not serve any purpose; to avoid that situation, we may pose two additional conditions (C_2) :

$$(C_2') \qquad D(x; P, \mathcal{U}) > 0 \text{ for at least one } x \in \mathbf{R}^d \qquad \text{and}$$

$$(C_2'') \qquad \lim_{\|x\| \to +\infty} D(x; P, \mathcal{U}) = 0$$

The condition (C_2'') was also singled out in [27], as a requirement for any reasonable depth function.

Before proceeding further, let us see some examples.

Example 3.1. 1° The simplest family $\mathcal{U}$ that satisfies (C_1) contains only one set - the whole space $\mathbf{R}^d$. Here $D(x; P, \mathcal{U}) = 1$ for all x; condition (C_2'') does not hold. In next three examples, the conditions (C_2') and (C_2'') hold (see Corollary A.1 in Appendix).

2° In $\mathbf{R}^d$, $d \geq 2$, let $\mathcal{V}$ be a family of all closed intervals $[\boldsymbol{a}, \boldsymbol{b}]$ with respect to the partial order induced by a convex cone, as in Example 2.1. It is easy to see that for each point $x \in \mathbf{R}^d$, there exists a closed interval V such that $x \notin V$; hence, the condition (C_1) is satisfied. A particular case of this example is the coordinate-wise partial ordering in $\mathbf{R}^d$, which in $\mathbf{R}^2$ yields $V \in \mathcal{V}$ to be rectangles with sides parallel to axes.

3° Instead of intervals in previous examples, we can take $V \in \mathcal{V}$ to be arbitrary convex and compact sets with a property that the collection $\mathcal{V}$ is closed under translations, and that sets V can be arbitrary "big" (for example, every ball in $\mathbf{R}^d$ should be contained in some $V \in \mathcal{V}$). Because of compactness and the translation property, for each $x \in \mathbf{R}^d$ there exists a $V \in \mathcal{V}$ that does not contain x, and (C_1) follows.

4° Consider now the class $\mathcal{V}$ of all closed halfspaces. The sets $U \in \mathcal{U}$ are then open halfspaces, and then the definition (3.1) of depth function formally differs from Tukey's halfspace depth function that requires closed halfspaces to be in $\mathcal{U}$. However, since $P(U) = \lim P(V_n)$, where $V_n \supset U$ are closed halfspaces with boundaries parallel to the boundary of U at euclidean distance $1/n$, it follows that values of D coincide for these two cases. The condition (C_1) is clearly satisfied. □

We are here interested chiefly in finding the set where the function D attains its global maximum, or, more generally, the sets of the form

$$S_\alpha = S_\alpha(P,\mathcal{U}) := \{x \in \mathbf{R}^d \mid D(x;P,\mathcal{U}) \geq \alpha\}, \tag{3.2}$$

The next Lemma gives a way to find S_α without evaluation of the depth function.

Lemma 3.1. *Let $\mathcal{U}$ be any collection of non-empty sets in $\mathbf{R}^d$, such that the condition (C_1) holds. Then, for any probability distribution P,*

$$S_\alpha(P,\mathcal{U}) = \bigcap_{V\in\mathcal{V}, P(V)>1-\alpha} V, \tag{3.3}$$

for any $\alpha \in (0,1]$ such that there exists a set $U \in \mathcal{U}$ with $P(U) < \alpha$; otherwise $S_\alpha = \mathbf{R}^d$.

It is instructive first to observe S_α in $d = 1$, as in the next example.

Example 3.2. Let X be a real random variable with the distribution P. Take $\mathcal{V}$ to be the family of all closed intervals in $\mathbf{R}$, and $\mathcal{U}$ to be the family of their complements. Then, using similar arguments as in the proof of (2.1) at the beginning of Section 2, it can be derived that $S_\alpha = [q_\alpha, Q_{1-\alpha}]$, where q_α is the smallest quantile of X of order α, and $Q_{1-\alpha}$ is the largest quantile of X of order $1-\alpha$:

$$\begin{aligned} q_\alpha &= \inf\{t \in \mathbf{R} \mid \mathrm{Prob}\,(X \leq t) \geq \alpha\} \quad \text{and} \\ Q_{1-\alpha} &= \sup\{t \in \mathbf{R} \mid \mathrm{Prob}\,(X \geq t) \geq \alpha\}. \end{aligned} \tag{3.4}$$

For $\alpha = \frac{1}{2}$, $[q_{\frac{1}{2}}, Q_{\frac{1}{2}}]$ is the median interval. □

In the case when the family $\mathcal{V}$ is consisted of closed intervals with respect to a partial order that satisfies (I1)–(I3), it follows from Section 2 and Lemma 3.1, that the set $S_{1/2}$ is non-empty, i.e. the corresponding depth function has maximum which is $\geq 1/2$, regardless of distribution P. For other families of $\mathcal{V}$, the guaranteed maximum is smaller.

Example 3.3. Consider the halfspace depth, as in Example 4° of 3.1, in $\mathbf{R}^2$, with the probability measure P which assigns mass 1/3 to points $A(0,1)$, $B(-1,0)$ and $C(1,0)$ in the plane. Each point x in the closed triangle ABC has $D(x) = \frac{1}{3}$; points outside of the triangle have $D(x) = 0$. So, the function D reaches its maximum value $\frac{1}{3}$.

Let us now observe the same distribution, but with depth function defined with the family $\mathcal{V}$ of closed disks. The intersection of *all* closed disks V with $P(V) > 2/3$ is, in fact, the intersection of all disks that contain all three points A, B, C, and that is the closed triangle ABC. For any $\varepsilon > 0$, a disc V with $P(V) > 2/3 - \varepsilon$ may contain only two of points A, B, C, but then it is easy to see that the family of all such discs has the empty intersection. Therefore, S_α is non-empty for $\alpha \leq 1/3$, and again, the function D attains its maximum value $1/3$ at the points of closed triangle ABC.

If the depth function is defined in terms of rectangles with sides parallel to coordinate axes, then the maximum depth is $2/3$ and it is attained at $(0, 0)$. This conclusion follows immediately from Theorem 2.2. □

If α_m is the maximum value of $D(x; P, \mathcal{U})$ for a given distribution P, the set S_{α_m}, i.e., the set of deepest points with respect to P, is called the center of the distribution P, and will be denoted by $C(P, \mathcal{U})$.

In the next theorem, we discuss some properties of the center, in the case when sets in $\mathcal{U}$ are open. A similar result for the family $\mathcal{U}$ of closed sets was obtained in [27, Theorem 2.11], but under more restrictive assumptions.

Theorem 3.1. *Let $\mathcal{V}$ be a collection of closed subsets of $\mathbf{R}^d$, and let $\mathcal{U}$ be the collection of sets V^c, $V \in \mathcal{V}$, such that the condition (C_1) holds. Then, for arbitrary probability measure P, the function $x \mapsto D(x; P, \mathcal{U})$ is upper semicontinuos. Moreover, under conditions (C_2), the set $C(P, \mathcal{U})$ on which D reaches its maximum is equal to the minimal nonempty set S_α, that is,*

$$C(P, \mathcal{U}) = \bigcap_{\alpha : S_\alpha \neq \emptyset} S_\alpha(P, \mathcal{U}).$$

The set $C(P, \mathcal{U})$ is a non-empty compact set and it has the following representation:

$$C(P, \mathcal{U}) = \bigcap_{V \in \mathcal{V}, P(V) > 1 - \alpha_m} V, \quad \text{where } \alpha_m = \max_{x \in \mathbf{R}^d} D(x; P, \mathcal{U}). \tag{3.5}$$

4 Equivalence of depth functions.

We already noticed that the depth functions with $\mathcal{U}$ being all open or all closed halfspaces, have the same value at every point. So, it is possible that two different classes of sets in place of $\mathcal{U}$ generate the same depth function. In the next theorem we give a sufficient condition for the equivalence of two depth functions.

Theorem 4.1. *Let $\mathcal{A}$ and $\mathcal{B}$ be families of subsets of $\mathbf{R}^d$. Suppose that the condition (C_1) holds for at least one of these families, and, in addition, the following condition (E):*

(E') *For each $A \in \mathcal{A}$, $A = \bigcup_{B \in \mathcal{B}, B \subset A} B$, and*

(E'') *For each $B \in \mathcal{B}$, there exists at most countable collection of sets $A_i \in \mathcal{A}$, such that $A_1 \supseteq A_2 \supseteq \dots$ and $B = \bigcap_i A_i$.*

Then the condition (C_1) holds for both $\mathcal{A}$ and $\mathcal{B}$ and depth function with respect to both families are equal, with any probability distribution P:

$$D(x; P, \mathcal{A}) = \inf\{P(A) \mid x \in A \in \mathcal{A}\} = \inf\{P(B) \mid x \in B \in \mathcal{B}\} = D(x; P, \mathcal{B})$$

An important application of Theorem 4.1 is to establish the equivalence of depth functions defined by a family of open sets $A \in \mathcal{A}$ and their topological closures $\bar{A} \in \mathcal{B}$. In this setup, we note that (E) holds in cases when $\mathcal{A}$ is invariant with respect to translations, and consists of (i) open halfspaces or (ii) complements of all closed intervals with any convex cone partial order, as in example 2.1. Another application of Theorem 4.1 will be given in Theorem 5.2.

5 Convex sets and halfspaces.

Suppose that a family $\mathcal{U}$ contains a sequence of nested sets U_n that intersect at one point $x \in \mathbf{R}^d$. Then $D(x; P, \mathcal{U}) = P(x)$ for any P, which is undesirable property. Therefore $\mathcal{U}$ should not contain sets that shrink to a point. A way to avoid that is to choose sets in $\mathcal{U}$ to be unbounded, or to choose sets in $\mathcal{V}$ to be bounded.

Further, it is natural to have a convex center of distribution, which is achieved (via Theorem 3.1) if sets in $\mathcal{V}$ are convex. With non-convex sets in $\mathcal{V}$, and with a discrete distribution P, we can again have that $D(x; P, \mathcal{U}) = P(x)$ for every $x \in \mathbf{R}^d$, as shown in the next example.

Example 5.1. In $\mathbf{R}^2$, let K be the lower half of the first quadrant, bounded by halflines $y = 0$ and $y = x$. Let us consider the family $\mathcal{U}$ of sets that can be obtained by arbitrary rotations and translations of K. Let $\mathcal{V}$ be the family of complements of sets in $\mathcal{U}$: sets in $V \in \mathcal{V}$ are non-convex.

Suppose that a distribution P is concentrated in six points $X_{1,2}(\pm 1, 0)$, $X_{3,4}(\pm 2, 0)$, $X_{5,6}(0, \pm 1)$ (as in the Counterexample 2 in [27]). Then for each point $x \in \mathbf{R}^d$, there is a set $U \in \mathcal{U}$ that does not contain any point of the support different from x; hence the depth of each point is $P(\{x\})$. With a specific discrete distribution P, the center is the point with the greatest probability mass. If P is uniform across the X_i, $i = 1, \ldots, 6$, then the center is the discrete set $\{X_1, \ldots, X_6\}$. This is in a sharp contrast with the halfspace depth function, which in this case yields the single point center at $(0, 0)$, with the maximum depth $\frac{1}{2}$. □

A prototype of depth functions that we discuss in this section is a depth function defined with respect to families $\mathcal{U}$ of complements of compact convex sets. In the light of the arguments given above, these requirements are natural and they are not too restrictive (see also Lemma 5.1). Although it may look that by these requirements we are excluding the halfspace depth from consideration, it is not so, as we will see after the Theorem 5.2.

From the material of Section 2, it follows that the depth function based on a family $\mathcal{V}$ of intervals, attains the maximal value of at least $1/2$, regardless of the dimension d. In general, the maximum depth with a family $\mathcal{V}$ of convex sets, can not be smaller than $\frac{1}{d+1}$. This conclusion follows from the next theorem, which is an extension of results in [4] and [16].

Theorem 5.1. *Let P be any probability measure on Borel sets of $\mathbf{R}^d$. Let $\mathcal{V}$ be any family of closed convex sets in $\mathbf{R}^d$, and let $\mathcal{U}$ be the family of their complements. Assume that conditions (C_1) and (C_2'') hold. Then the condition (C_2') also holds, and there exists a point $x \in \mathbf{R}^d$ with $D(x; P, \mathcal{U}) \geq \frac{1}{d+1}$.*

The lower bound for D in Theorem 5.1 is the greatest generally possible. As the next example shows, for the halfspace depth, in any dimension $d \geq 1$, there exist a probability measure P such that $D(x; P, \mathcal{U}) \leq \frac{1}{d+1}$ for all $x \in \mathbf{R}^d$.

Example 5.2. This is an extension of the example 3.3. Let $A_1, \ldots, A_{d+1}$ be points in $\mathbf{R}^d$ such that they do not belong to the same hyperplane (i.e. to any affine subspace of dimension less than d), and suppose that $P(\{A_i\}) = \frac{1}{d+1}$ for each $i = 1, 2, \ldots, d+1$. Let S be a closed d-dimensional simplex with vertices at $A_1, \ldots, A_{d+1}$, and let $x \in S$. If x is a vertex of S, then there exists a closed halfspace H such that $x \in H$ and other vertices do not belong to H; then $D(x) = P(H) = 1/(d+1)$. Otherwise, let S_x be a d-dimensional simplex with vertices in x and d points among $A_1, \ldots, A_{d+1}$ that make together an affinely independent set. Then for S_x and the remaining vertex, say A_1, there exists a separating hyperplane π such that $\pi \cap S_x = \{x\}$ and $A_1 \notin \pi$ (see [15, Section 11]). Let H be a halfspace with boundary π, that contains A_1. Then also $D(x) = P(H) = 1/(d+1)$. So, all points $x \in S$ have $D(x) = 1/(d+1)$. Points x outside of S have $D(x) = 0$, which is easy to see. So, the maximal depth in this example is exactly $1/(d+1)$. □

In fact, if we have a family of compact convex sets $\mathcal{V}$ that contain arbitrary large sets (in the sense of the following lemma), then it is sufficient to assume only condition (C_1), and then (C_2) will automatically hold. A natural way to choose $\mathcal{V}$ would be then, to choose one compact convex shape, and allow translations (and, possibly, rotations, if we want an affine invariant depth).

Lemma 5.1. *Let $\mathcal{V}$ be a family of compact convex sets in $\mathbf{R}^d$, and let $\mathcal{U}$ be the family of complements of sets in $\mathcal{V}$, such that the condition (C_1) holds. Suppose that for every closed ball $B \in \mathbf{R}^d$ there exist a set $V \in \mathcal{V}$, such that $B \subset V$. Then the family $\mathcal{U}$ and the depth function $D(\cdot\ ; P, \mathcal{U})$ satisfy conditions (C_2') and (C_2''), with any probability measure P on $\mathbf{R}^d$.*

In the next theorem, we use the fact that every closed convex set can be represented as an intersection of closed halfspaces (see, for example, [15, Theorem 11.5]). This representation is not unique (and we do not need uniqueness neither in the statement nor in the proof); however, there is a unique minimal representation of a convex set as the intersection of all its tangent halfspaces [15, Theorem 18.8], which is an intuitive model for the representation (5.1) below.

Theorem 5.2. *Let $\mathcal{V}$ be a collection of closed convex sets and $\mathcal{U}$ the collection of complements of all sets in $\mathcal{V}$. For each $V \in \mathcal{V}$, consider a representation*

$$V = \bigcap_{\alpha \in A_V} H_\alpha, \tag{5.1}$$

where H_α are closed subspaces and A_V is an index set. Let

$$\mathcal{H}^V = \{\overline{H_\alpha^c} + x \mid \alpha \in A_V,\ x \in \mathbf{R}^d\}$$

be the collection of closures of complements of halfspaces H_α and their translations. Further, let

$$\mathcal{H} = \bigcup_{V \in \mathcal{V}} \mathcal{H}^V.$$

If for any $H \in \mathcal{H}$ there exists at most countable collection of sets $V_i \in \mathcal{V}$, such that

$$V_1 \subseteq V_2 \subseteq \cdots \quad \text{and} \quad \overset{\circ}{H} = \bigcup V_i, \tag{5.2}$$

then

$$D(x;P,\mathcal{U}) = D(x;P,\mathcal{H}) = D(x;P,\overset{\circ}{\mathcal{H}}), \qquad \textit{for every } x \in \mathbf{R}^d,$$

where $\overset{\circ}{\mathcal{H}}$ *is the family of open halfspaces from* $\mathcal{H}$.

As a corollary to Theorem 5.2, we can single out two important particular cases. Conditions (5.1) and (5.2) in both cases can be easily proved.

Corollary 5.1. a) *Let* $\mathcal{V}$ *be the family of closed intervals with respect to the partial order defined with a convex cone* K*, as in the Section 2. Then for any probability distribution and any* $x \in \mathbf{R}^d$,

$$D(x;P,\mathcal{U}) = D(x;P,\mathcal{H}),$$

where $\mathcal{U}$ *is the family of complements of sets in* $\mathcal{V}$ *and* H *is the family of all tangent halfspaces to* K*, and their translations.*

In particular, if $\mathcal{V}$ *is the family of intervals with respect to the coordinate-wise partial order, then the corresponding depth function is the same as the depth function generated by halfspaces with borders parallel to the coordinate hyperplanes.*

b) *Let* $\mathcal{H}$ *be the family of all closed halfspaces, let* $\mathcal{U}_k$ *be the family of complements of all compact closed sets, and let* $\mathcal{U}_b$ *be the family of complements of closed balls in* $\mathbf{R}^d$*. Then*

$$D(x;P,\mathcal{H}) = D(x;P,\mathcal{U}_k) = D(x;P,\mathcal{U}_b),$$

That is, the Tukey halfspace depth can be realized via complements of closed convex sets or via complements of closed balls.

The second part of Corollary 5.1 implies, via Lemma 3.1, that for the halfspace depth function D, we have

$$S_\alpha = \{x \in \mathbf{R}^d \mid D(x) \geq \alpha\} = \bigcap_{K:\ P(K)>1-\alpha} K = \bigcap_{B:\ P(B)>1-\alpha} B, \qquad (5.3)$$

where K are compact convex sets, and B are closed balls. The reduction to balls is of the obvious interest in applications, where we have to find deepest points of a high dimensional cloud of data.

6 Affine invariance and another representation of the halfspace center of distribution.

A depth function $D(x; P_X, \mathcal{U})$ in $\mathbf{R}^d$ is said to be affine invariant, if

$$D(Ax+b; P_{AX+b}, \mathcal{U}) = D(x; P_X, \mathcal{U}) \quad \text{for any probability measure } P, \tag{6.1}$$

for any nonsingular $d \times d$ matrix A, any $b \in \mathbf{R}^d$ and $x \in \mathbf{R}^d$, where P_{AX+b} is a probability distribution of a random variable $AX + b$, X being a random variable with the distribution P_X. From the definition (6.1), it follows that one sufficient condition for affine invariance of D is the affine invariance of $\mathcal{U}$: If $U \in \mathcal{U}$, then $AU + b \in \mathcal{U}$, for all A and b. This condition is satisfied with the family $\mathcal{U}$ of all halfspaces; hence the halfspace depth is affine invariant. Due to the fact that the same depth function can be generated by different families $\mathcal{U}$, this condition is not necessary, as the next example shows.

Example 6.1. Let $\mathcal{V}$ be the family of all closed discs in $\mathbf{R}^2$, and let $\mathcal{U}$ be the family of their complements. The family $\mathcal{U}$ is not affine invariant, because the circles transform into ellipses, with a non-orthogonal matrix A. However, the family of all halfplanes $\mathcal{H}$ generates the same depth function as the family $\mathcal{U}$, and so, the depth $D(x; P, \mathcal{U})$ is equivalent to halfspace depth, hence, it is affine invariant. □

For depth functions that can be generated by a family of halfspaces, the conditions of affine invariance can be expressed via translation and rotations, as every halfspace in $\mathbf{R}^d$ can be transformed into another one by one rotation and one translation. That is, for every two halfspaces $H_1, H_2 \in \mathbf{R}^d$, there exists an affine transformation $x \mapsto Ax + b$ with A being an orthogonal matrix, such that $H_2 = AH_1 + b$.

Consider one coordinate system in $\mathbf{R}^d$, with the corresponding set $\mathcal{J}$ of coordinate-wise intervals. Any rotation ρ of the coordinate system will produce another family of intervals $\mathcal{J}_\rho$. According to Theorem 5.2, depth functions based on the family $\cup_\rho \mathcal{J}_\rho$ (where the union goes through all possible rotations) is equivalent to the halfspace depth function. More generally, we may observe any set of partial orders $\{\preceq_\rho\}$ (where ρ belongs to some index set) that satisfy conditions (I1)-(I3) of Section 2 such that the corresponding families of intervals $\mathcal{J}_\rho$ (i.e., families $\mathcal{U}_\rho$ of complements of sets from $\mathcal{J}_\rho$) together generate the halfspace depth function. Let us call such set of partial orders *complete*. For a given probability distribution P, and a complete set of partial orders, let

$$S_{\alpha,\rho} = \{x \in \mathbf{R}^d \mid D(x; P, \mathcal{U}_\rho) \geq \alpha\} = \bigcap_{J \in \mathcal{J}_\rho:\ P(J) > 1-\alpha} J;$$

$$S_\alpha = \{x \in \mathbf{R}^d \mid D(x; P, \mathcal{H}) \geq \alpha\},$$

where $\alpha \leq \alpha_m$, and $\mathcal{H}$ is the family of all open halfspaces. Then by completeness, we have that

$$S_\alpha = \bigcap_\rho \bigcap_{J \in \mathcal{J}_\rho:\ P(J) > 1-\alpha} J = \bigcap_\rho S_{\alpha,\rho}. \tag{6.2}$$

For $\alpha = \alpha_m$, (6.2) gives another representation of the center of a distribution, in terms of sets that are not affine invariant. If we take any finite subset of partial orders, $\rho = 1, \ldots, n$, then we have

$$C(P, \mathcal{H}) \subset \bigcap_{\rho=1}^{n} S_{\alpha_m,\rho}, \tag{6.3}$$

which gives an upper bound for the center of distribution in terms of finitely many partial orders.

Note that, in general, the sets $S_{\alpha_m,\rho}$ are not centers of the distribution with respect to $\preceq_\rho$; we proved in Section 2 that there exist median sets $S_{1/2,\rho}$. In general, we have that $\alpha_m < 1/2$, and $S_{1/2,\rho} \subset S_{\alpha_m,\rho}$. Median sets with respect to different partial orders may have empty intersection. For example, if the distribution is absolutely continuous, then every median set with respect to a coordinate-wise partial order is a singleton; clearly by a rotation of the coordinate system we may obtain different singletons.

7 A version of Jensen's inequality.

Let $\mathcal{V}$ be a family of closed sets, $\mathcal{U}$ the family of complements of sets from $\mathcal{V}$ and let $D(x; P, \mathcal{U})$ be defined as in previous sections. Let $C = C(P, \mathcal{U})$ be the center of a probability measure P in $\mathbf{R}^d$, with α_m being the maximum of the depth function. Assume conditions (C_1) and (C_2).

For a random variable X with the distribution P, the points in the set $C(P, \mathcal{U})$ can be thought of as a kind of mean values of X, in the same sense as univariate medians are being thought of. If f is a real valued function defined on $\mathbf{R}^d$, then the analogous mean value of $f(X)$ are points in the closed interval $[q_{\alpha_m}, Q_{1-\alpha_m}]$, where, by (3.4), q_{α_m} is the smallest quantile of $f(X)$ of order α_m, and $Q_{1-\alpha_m}$ is the largest quantile of order $1 - \alpha_m$. If $\alpha_m = 1/2$, we have the median interval of $f(X)$, and the center $C(P, \mathcal{U})$ becomes $\{\mathrm{Med}\, X\}$. Let $m \in \{\mathrm{Med}\, X\}$ and $M \in \{\mathrm{Med}\, f(X)\}$. With analogy to Jensen's inequality $f(\mathrm{E}\, X) \leq \mathrm{E}\, f(X)$ for convex functions, we may expect that $f(m) \leq M$ for an appropriate class of functions f. Indeed, we prove a result of that kind, for the class of functions that are described in the following definition. The name *C-function* is taken from [12], where it was used in a more particular context.

Definition 7.1. A function $f : \mathbf{R}^d \mapsto \mathbf{R}$ will be called a *C-function* with respect to a given family $\mathcal{V}$ of closed subsets of $\mathbf{R}^d$, if $f^{-1}((-\infty, t]) \in \mathcal{V}$ or is empty set, for every $t \in \mathbf{R}$.

Example 7.1. 1° If $\mathcal{V}$ is the family of all closed convex sets in $\mathbf{R}^d$, then the class of corresponding C-functions is precisely the class of lower continuous quasi-convex functions, i.e., functions f that have the property that $f^{-1}((-\infty, t])$ is a closed set for any $t \in \mathbf{R}$ and

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}, \qquad \lambda \in [0,1], \quad x, y \in \mathbf{R}^d.$$

This is easy to see, starting from the definition 7.1. In particular, every convex function on $\mathbf{R}^d$ is a C-function with respect to the class of all convex sets.

2° Let $D(x)$ be a halfspace depth function. Then it follows from (3.3) that the sets S_α are convex, which implies that the function $x \mapsto 1 - D(x)$ is a C-function with respect to a family of all closed convex sets.

3°A function f is a C-functions with respect to a family of closed intervals (with respect to some partial order), if and only if

$$\{\boldsymbol{x} \in \mathbf{R}^d \mid f(\boldsymbol{x}) \leq t\} = [\boldsymbol{a}, \boldsymbol{b}], \qquad \text{for some } \boldsymbol{a}, \boldsymbol{b} \in \mathbf{R}^d \ , d \geq 1.$$

This condition is not satisfied for all convex functions. For example, in $\mathbf{R}^2$, with the coordinate-wise partial order the function defined by $f(x, y) = x^2 + y^2$ is convex, but the sets $\{(x, y) \mid x^2 + y^2 \leq t\}$ are not intervals.

4° In $\mathbf{R}^2$, with coordinate-wise intervals, the function f defined by

$$f(x, y) = \max\{|x - a_1| - |x - b_1|, \ |y - a_2| - |y - b_2|\}$$

is a C-function, where $\boldsymbol{a}(a_1, a_2)$ and $\boldsymbol{b}(b_1, b_2)$ are given points in $\mathbf{R}^2$.

5° In general, assuming conditions (I1)-(I3), we may define the depth function $D(x)$ with respect to the class $\mathcal{U}$ of complements of the given family of intervals. Since the intersection of closed intervals is again a closed interval, we see that here also the function $x \mapsto 1 - D(x)$ is a C-function.

6° Note that, since we require sets in $\mathcal{V}$ to be closed, every C-function is lower semicontinuous. □

The next two theorems are versions of Jensen's inequality.

Theorem 7.1. *Let $\mathcal{V}$ be a family of closed intervals with respect to a partial order in $\mathbf{R}^d$, such that conditions (I1)–(I3) are satisfied. Let $\{\mathrm{Med}\, X\}$ be the median set of a random variable X with respect to the chosen partial order, and let f be a C-function with respect to the family $\mathcal{V}$. Then for every $M \in \mathrm{Med}\,\{f(X)\}$, there exists an $m \in \{\mathrm{Med}\, X\}$, such that*

$$f(m) \leq M. \tag{7.1}$$

In particular, if m or M are unique, then (7.1) holds for any m, M.

In general case the depth function does not necessarily reach the value of 1/2, and we have only a weaker result:

Theorem 7.2. *Let $\mathcal{V}$ be a family of closed subsets of $\mathbf{R}^d$, and let $\mathcal{U}$ be the family of their complements. Assume that conditions (C_1) and (C_2) hold with a given probability measure P, induced by a random variable X. Let α_m be the maximum of the depth function $D(x; P,\mathcal{U})$, which is achieved in all points of the center $C(P,\mathcal{U})$ and let f be a C-function with respect to $\mathcal{V}$. Then for every $m \in C(P,\mathcal{U})$ we have that*

$$f(m) \leq Q_{1-\alpha_m}, \tag{7.2}$$

where $Q_{1-\alpha_m}$ is the largest quantile of order $1-\alpha_m$ for $f(X)$.

To show that we can not claim anything better in a general case, consider the following example:

Example 7.2. Let A, B, C be non-colinear points in the two dimensional plane, and let $\mathcal{H}$ be the collection of closed halfplanes. Let $l(AB), l(AC), l(BC)$ be the lines determined by two indicated points. Let H_1 be the closed half-space that does not contain the interior of the triangle ABC and has $l(AB)$ for its boundary, and let H_2 be its complement. Define a function f by

$$f(x) = e^{-d(x,l(AB))} \quad \text{if } x \in H_1\ , \qquad f(x) = e^{d(x,l(AB))} \quad \text{if } x \in H_2,$$

where $d(\cdot,\cdot)$ is euclidean distance. Then $f(A) = 1$, $f(B) = 1$ and $f(C) > 1$, and f is clearly a C-function with respect to the class $\mathcal{H}$. Now suppose that P assigns mass 1/3 to each of the points A, B, C. Then, by example 3.3, we know that the center $C(p, \mathcal{H})$ of this distribution is the set of points of the triangle ABC, with $\alpha_m = 1/3$. Hence, for $m \in C(P, \mathcal{H})$, $f(m)$ takes all values in $[1, f(C)]$. On the other hand, quantiles of the order 2/3 are points in the closed interval $[1, f(C)]$; hence the most we can state is that $f(m) \leq f(C)$, with $f(C)$ being the largest quantile of order 2/3.

A APPENDIX: PROOFS AND AUXILIARY RESULTS

In order to prove Theorem 2.1, we need the following lemma.

Lemma A.1. *Let $\preceq$ be a partial order in $\mathbf{R}^d$ such that the conditions (I1) and (I3) hold. Let*

$$\mathcal{J} = \{J_\alpha \mid J_\alpha = [\boldsymbol{a}^\alpha, \boldsymbol{b}^\alpha], \quad \alpha \in A\}$$

be a collection of closed intervals, where A is an index set. Assume that there is at least one α such that $\boldsymbol{a}^\alpha \in \mathbf{R}^d$ (i.e., have all coordinates finite) and at least one β such that $\boldsymbol{b}^\beta \in \mathbf{R}^d$. Suppose that $J_\alpha \cap J_\beta \neq \emptyset$ for all α, β. Then

(i) $\boldsymbol{a}^\alpha \preceq \boldsymbol{b}^\beta$, for any $\alpha, \beta \in A$;

(ii) The intersection of all sets in $\mathcal{J}$ is a non-empty compact interval $[\boldsymbol{a}, \boldsymbol{b}]$, with $\boldsymbol{a}, \boldsymbol{b} \in \mathbf{R}^d$.

Proof. If intervals $[\boldsymbol{a}, \boldsymbol{b}]$ and $[\boldsymbol{c}, \boldsymbol{d}]$ have a common point $\boldsymbol{x}$, then $\boldsymbol{a} \preceq \boldsymbol{x} \preceq \boldsymbol{b}$ and $\boldsymbol{c} \preceq \boldsymbol{x} \preceq \boldsymbol{d}$; hence $\boldsymbol{a} \preceq \boldsymbol{d}$ and $\boldsymbol{c} \preceq \boldsymbol{b}$. This shows (i). Further, to show (ii), note that by assumptions and (i), the set $\{\boldsymbol{a}^\alpha,\ \alpha \in A\}$ is bounded from above with a finite point, and so by (I3), there exists $\boldsymbol{a} = \sup_{\alpha \in A} \boldsymbol{a}^\alpha$. In an analogous way we conclude that there exists $\boldsymbol{b} = \inf_{\beta \in A} \boldsymbol{b}^\beta$. By properties of the infimum and supremum, we have that $\boldsymbol{a}^\alpha \preceq \boldsymbol{a} \preceq \boldsymbol{b} \preceq \boldsymbol{b}^\alpha$, for all $\alpha \in A$, so the interval $[\boldsymbol{a}, \boldsymbol{b}]$ is non-empty; it is compact by assumption (I1), and it is contained in all intervals of the family $\mathcal{J}$. On the other hand, any point $\boldsymbol{c}$ that is common for all intervals J_α must be an upper bound for $\{\boldsymbol{a}^\alpha\}$ and a lower bound for $\boldsymbol{b}^\alpha$; hence $\boldsymbol{a} \preceq \boldsymbol{c} \preceq \boldsymbol{b}$, that is, $\boldsymbol{c} \in [\boldsymbol{a}, \boldsymbol{b}]$, and (ii) is proved. □

Proof of the Theorem 2.1. It is clear that any two intervals in $\mathcal{J}$ have a non-empty intersection; besides, by (I2), at least one of the intervals has finite endpoints. Then the assertion follows by Lemma A.1. □

Proof of the Theorem 2.2. To simplify notations, we give the proof for $d = 2$; the proof in a general case is analogous. We have the sequence of relations

$$\begin{aligned}
\{\mathrm{Med}\,\boldsymbol{X}\} &= \bigcap_{[\boldsymbol{a},\boldsymbol{b}]:\ \mathrm{Prob}(\boldsymbol{X}\in[\boldsymbol{a},\boldsymbol{b}])>\frac{1}{2}} [\boldsymbol{a},\boldsymbol{b}] \\
&= \bigcap_{[\boldsymbol{a},\boldsymbol{b}]:\ \mathrm{Prob}(\boldsymbol{X}\in[\boldsymbol{a},\boldsymbol{b}])>\frac{1}{2}} [a_1,b_1] \times \bigcap_{[\boldsymbol{a},\boldsymbol{b}]:\ \mathrm{Prob}(\boldsymbol{X}\in[\boldsymbol{a},\boldsymbol{b}])>\frac{1}{2}} [a_2,b_2] \\
&\supset \bigcap_{[a_1,b_1]:\ \mathrm{Prob}(X_1\in[a_1,b_1])>\frac{1}{2}} [a_1,b_1] \times \bigcap_{[a_2,b_2]:\ \mathrm{Prob}(X_2\in[a_2,b_2])>\frac{1}{2}} [a_2,b_2]) \\
&= \{\mathrm{Med}\,X_1\} \times \{\mathrm{Med}\,X_2\},
\end{aligned}$$

where we used the fact that $X_1 \in [a_1, b_1]$ whenever $\boldsymbol{X} \in [\boldsymbol{a}, \boldsymbol{b}]$. On the other hand, if $\{\mathrm{Med}\,X_1\} = [a, b]$ and $\{\mathrm{Med}\,X_2\} = [c, d]$, then we have that

$$[a,b]\times[c,d] = [a,+\infty)\times\mathbf{R}\cap(-\infty,b]\times\mathbf{R}\cap\mathbf{R}\times[c,+\infty)\cap\mathbf{R}\times(-\infty,d]$$

and we note that all four two-dimensional intervals on the right hand side of the last identity, can be expressed as intersections of a sequence of intervals J_n with $\mathrm{Prob}(\boldsymbol{X} \in J_n) > 1/2$; for example,

$$[a,+\infty)\times\mathbf{R} = \bigcap_{i=n}^{+\infty}[a-1/n,+\infty)\times\mathbf{R},$$

and $\mathrm{Prob}(\boldsymbol{X} \in [a-1/n,+\infty)\times\mathbf{R}) > 1/2$ because $[a, b]$ is the median set for X_1. From this we conclude that

$$\{\mathrm{Med}\,X_1\}\times\{\mathrm{Med}\,X_2\} \supset \{\mathrm{Med}\,\boldsymbol{X}\},$$

and the theorem is proved. □

Another result related to the coordinate-wise partial order intervals is presented in the next lemma. For a given probability measure P, denote by $\mathcal{J}$ the class of all intervals J with the property that $P(J) > 1/2$.

In [12], the class $\mathcal{I}$ is defined as the family of all closed intervals $I \subset \mathbf{R}^d$ (with respect to coordinate-wise partial order) with the following property: If J is any closed interval (with respect to the same partial order) that contains I as a proper subset, then $J \in \mathcal{J}$.

In the following lemma, we show that the intersection of all intervals in $\mathcal{I}$ coincides with the intersection of all intervals in the class $\mathcal{J}$, i.e., with the median set, as it is defined in the present paper. The purpose of this result is to establish the equivalence between the definition of multivariate medians in the present work and in [12], for the special case of coordinate-wise partial order, which is considered there.

Lemma A.2. *Let $\mathcal{I}$ and $\mathcal{J}$ be families of intervals as defined above. Then*

(i) Each interval $I \in \mathcal{I}$ can be represented as $I = \cap_{J \in \mathcal{J}, J \supset I} J$;

(ii) $\cap_{I \in \mathcal{I}} I = \cap_{J \in \mathcal{J}} J$

Proof. For $I \in \mathcal{I}$, let et $S(I) = \bigcap_{J \in \mathcal{J}, J \supset I} J$. Clearly, $I \subset S(I)$. To show that $S(I) \subset I$, take any $x \notin I$. Since I is a closed interval, there is another closed interval J such that $I \subset J \subset \{x\}^c$, where both inclusions are strict, and thus $x \notin S(I)$. This ends the proof of (i). To show (ii), note that if $J \in \mathcal{J}$, then $J \in \mathcal{I}$, so $\mathcal{J} \subset \mathcal{I}$. Hence,

$$\bigcap_{J \in \mathcal{J}} J \supset \bigcap_{I \in \mathcal{I}} I.$$

Conversely, by the part (i), we have that

$$\bigcap_{I \in \mathcal{I}} I = \bigcap_{I} \bigcap_{J \in \mathcal{J}, J \supset I} J = \bigcap_{J \in \mathcal{J}'} J,$$

where $\mathcal{J}' \subset \mathcal{J}$, and hence, we conclude that $\bigcap_{I \in \mathcal{I}} I \supset \bigcap_{J \in \mathcal{J}} J$. □

Proof of the Lemma 3.1. Evidently, $x \in S_\alpha^c$ if and only if $D(x) < \alpha$, i.e., if and only if there exists a set $U \in \mathcal{U}$ such that $x \in U$ and $P(U) < \alpha$. Therefore, if there are $U \in \mathcal{U}$ with $P(U) < \alpha$, then

$$S_\alpha^c = \bigcup_{U \in \mathcal{U}, P(U) < \alpha} U, \quad \text{and so,} \quad S_\alpha = \bigcap_{U \in \mathcal{U}, P(U) < \alpha} U^c,$$

which is equivalent to the assertion that we wanted to prove. □

Proof of the Theorem 3.1. Under (C_1) and if all sets in $\mathcal{V}$ are closed, the set S_α is closed for every α, via (3.3), and hence, the function D is upper semicontinuous. Under additional conditions (C_2), we will show that there exists at least one α such that S_α is a nonempty compact set. Indeed, by the assumption, there is $x \in \mathbf{R}^d$ so that $D(x) = \alpha_0 > 0$. On the other hand, by assumption of convergence of $D(x)$ to zero as $\|x\| \to +\infty$, there exists an $R > 0$ so that $D(x) < \alpha_0$ for $\|x\| > R$. Therefore, the set S_{α_0} is nonempty and norm bounded, and being closed, it is compact. Then all sets S_α with $\alpha \geq \alpha_0$ are compact, because $S_\alpha \subset S_{\alpha_0}$ for $\alpha \geq \alpha_0$. The intersection of non-empty compact nested sets S_α is a non-empty compact set, and it is clearly the set on which D reaches its maximum. □

Proof of Theorem 4.1. Suppose that the stated conditions hold. If (C_1) holds for $\mathcal{A}$, then (E') implies that it holds for $\mathcal{B}$. If (C_1) holds for $\mathcal{B}$, then it clearly holds for $\mathcal{A}$ by (E'').

Let $x \in \mathbf{R}^d$ be fixed. Then by (E'), for each $A \in \mathcal{A}$ that contains x, there exists a $B_A \in \mathcal{B}$ such that $x \in B_A \subset A$, and, consequently, $P(A) \geq P(B_A)$. Therefore,

$$\begin{aligned} D(x; P, \mathcal{A}) &\geq \inf\{P(B_A) \mid x \in B_A \in \mathcal{B}, A \in \mathcal{A}\} \\ &\geq \inf\{P(B) \mid x \in B \in \mathcal{B}\} = D(x; P, \mathcal{B}) \end{aligned}$$

as the class of all B_A is a subset of the class of all $B \in \mathcal{B}$ that may contain x. On the other hand, by (E''), for each $\varepsilon > 0$ and for each $B \in \mathcal{B}$ that contains x, there exists $A_B \in \mathcal{A}$, such that $P(B) \geq P(A_B) - \varepsilon$. Then

$$\begin{aligned} \inf\{P(B) \mid x \in B \in \mathcal{B}\} &\geq \inf\{P(A_B) \mid x \in A_B, A_B \in \mathcal{A}, B \in \mathcal{B}\} - \varepsilon \\ &\geq \inf\{P(A) \mid x \in A \in \mathcal{A}\} - \varepsilon = D(x; P, \mathcal{A}) - \varepsilon, \end{aligned}$$

and since $\varepsilon > 0$ is arbitrary, we conclude that

$$D(x; P, \mathcal{B}) = \inf\{P(B) \mid x \in B \in \mathcal{B}\} \geq D(x; P, \mathcal{A}),$$

which ends the proof. □

The next Lemma is technical, and we need it for the proof of Theorem 5.1.

Lemma A.3. *Let P be any probability measure on Borel sets of $\mathbf{R}^d$. Let K be a compact set in $\mathbf{R}^d$ and let $\mathcal{A}$ be a family of closed convex subsets of K, with $P(A) > \frac{d}{d+1}$ for every $A \in \mathcal{A}$. Then the intersection of all sets $A \in \mathcal{A}$ is a non-empty compact set.*

Proof. If $P(A_i) > 1-\varepsilon$, $i = 1, 2, \ldots$, then it is easy to prove by induction that $P(A_1 \cdots A_n) > 1-n\varepsilon$ for $n \geq 2$. Therefore, under given assumptions, for any $d+1$ sets $A_1, \ldots, A_n \in \mathcal{A}$, it holds that $P(A_1 \cdots A_{d+1}) > 1-(d+1) \cdot \frac{1}{d+1} = 0$. Hence, every $d+1$ sets of the family $\mathcal{A}$ have a non-empty intersection. By Helly's intersection theorem ([17, 12.12.]), every finite number of convex sets in $\mathcal{A}$ have a non-empty intersection. Since K is compact, then all sets in $\mathcal{A}$ have a non-empty intersection (see e.g. [22, Theorem 17.4]). The intersection is compact since all sets in $\mathcal{A}$ are compact. □

Proof of Theorem 5.1. Let $\delta \in (0, 1)$ be fixed. Assuming that (C_1) holds, we will first prove that every compact convex set $K \subset \mathbf{R}^d$ with $P(K) = 1 - \delta > 0$ contains a point x with $D(x; P, \mathcal{U}) \geq \frac{1-\delta}{d+1}$. Indeed, let

$\varepsilon = \frac{1-\delta}{d+1}$ and suppose, contrary to the statement, that $D(x; P, \mathcal{U}) < \varepsilon$ for every $x \in K$, where K is a compact set with $P(K) = 1 - \delta > 0$. Then (by (C_1)), for every $x \in K$ there exists a $U_x \in \mathcal{U}$, such that $P(U_x) < \varepsilon$. Clearly,

$$\bigcup_{x \in K} U_x \supset K. \tag{A.1}$$

Let $U_x^c = V_x$. Then $V_x \in \mathcal{V}$, and by (A.1) it follows that

$$\bigcap_{x \in K} (V_x \cap K) = \emptyset \tag{A.2}$$

Let us now define a new probability measure P^* on $\mathbf{R}^d$, by

$$P^*(B) = \frac{P(B \cap K)}{1 - \delta}, \qquad \text{where } B \subset \mathbf{R}^d \text{ is a Borel set.}$$

For each $x \in K$, we have that $P(V_x) > 1 - \varepsilon$, and

$$P(V_x \cap K) > P(V_x) + P(K) - 1 > 1 - \varepsilon - \delta = \frac{d(1-\delta)}{d+1},$$

hence $P^*(V_x \cap K) > \frac{d}{d+1}$. Now by Lemma A.3, we conclude that the family of sets $V_x \cap K$ have non-empty intersection, which contradicts (A.2). So, the statement about compact convex sets is proved.

To prove the statement of the Theorem 5.1, note that the statement that we already proved yields the condition (C_2'), and, with additional assumption (C_2''), Theorem 3.1 is applicable. By the first part of the proof, each of the sets

$$S_n = \{x \in \mathbf{R}^d \mid D(x; P, \mathcal{U}) \geq \frac{1 - \frac{1}{n}}{d+1}\}, \qquad n = 1, 2, \ldots$$

is non-empty; then their intersection.

$$\bigcap_{n=1}^{+\infty} S_n = \{x \in \mathbf{R}^d \mid D(x; P, \mathcal{U}) \geq \frac{1}{d+1}\},$$

is also non-empty, by Theorem 3.1. This ends the proof. □

Proof of Lemma 5.1. We first prove that (C_2'') holds. For a fixed $\varepsilon > 0$, and a given probability measure P, let $B_{1-\varepsilon}$ be a closed ball centered at origin, with $P(B_{1-\varepsilon}) > 1 - \varepsilon$. Then, by assumptions, there exists a set $V \in \mathcal{V}$ such that $B_{1-\varepsilon} \subset V$. By compactness, there exists $r > 0$ such that all points $x \in V$ satisfy $\|x\| \leq r$. Therefore, all points x with $\|x\| > r$ are in $U = V^c$, and, since $P(U) = 1 - P(V) < \varepsilon$, we conclude that for a given $\varepsilon > 0$ there exists $r > 0$ so that $D(x; P, \mathcal{U}) < \varepsilon$ for all x with $\|x\| > r$, which proves (C_2''). Then by Theorem 5.1, the condition (C_2') also holds. □

Corollary A.1. *Conditions* (C_2) *hold for examples* $2^\circ - 4^\circ$ *in 3.1, with any probability measure* P.

Proof. For examples 2° and 3° in 3.1 it is straightforward to check that the assumptions of Lemma 5.1 are satisfied; hence both conditions in (C_2) hold. For the halfspace depth in Example 4°, we may apply Theorem 5.2, to conclude that the halfspace depth function is the same as the one based on the family of all compact convex sets, which is the example 3°. □

PROOF OF THEOREM 5.2. Let $\overset{\circ}{H}$ be an open halfspace from $\overset{\circ}{\mathcal{H}}$, and let H be its closure. Given any $x \in \overset{\circ}{H}$, there exists a closed halfspace H_x that can be obtained by translation of H in such a way that the border of H_x contains x. Then $H_x \in \mathcal{H}$ and, clearly,

$$\overset{\circ}{H} = \bigcup_{x \in \overset{\circ}{H}} H_x,$$

which implies condition (E') of Theorem 4.1 with $\mathcal{A} = \overset{\circ}{\mathcal{H}}$ and $\mathcal{B} = \mathcal{H}$. On the other hand, for any given closed halfspace $H \in \mathcal{H}$, there exists a sequence of halfspaces H_i, obtained from H by translation, such that

$$\overset{\circ}{H}_1 \supset \overset{\circ}{H}_2 \supset \cdots \quad \text{and } H = \cap_i \overset{\circ}{H}_i,$$

which is the condition (E''). Therefore, by Theorem 4.1,

$$D(x; P, \mathcal{H}) = D(x; P, \overset{\circ}{\mathcal{H}}). \tag{A.3}$$

Now note that (5.1) gives condition (E') for $\mathcal{A} = \mathcal{U}$ and $\mathcal{B} = \overset{\circ}{\mathcal{H}}$ (by taking complements on both sides); then, as in the proof of Theorem 4.1, we find that

$$D(x; P, \mathcal{U}) \geq D(x; P, \overset{\circ}{\mathcal{H}}), \tag{A.4}$$

for every $x \in \mathbf{R}^d$. In the same way, (5.2) gives condition (E'') for $\mathcal{A} = \mathcal{U}$ and $\mathcal{B} = \mathcal{H}$, and so, again as in the proof of 4.1,

$$D(x; P, \mathcal{U}) \leq D(x; P, \mathcal{H}). \tag{A.5}$$

The statement of the theorem now follows from (A.3), (A.4) and (A.5). □

PROOF OF THEOREM 7.2. Let $Q = Q_{1-\alpha_m}$. Then for every $\varepsilon > 0$, Prob $(f(X) \leq Q + \varepsilon) > 1 - \alpha_m$, and, therefore, the set

$$V_\varepsilon = f^{-1}((-\infty, Q + \varepsilon])$$

contains the center $C(P,\mathcal{U})$. This implies that

$$f(m) \leq Q + \varepsilon, \quad \text{for every } m \in C(P,\mathcal{U}) \text{ and every } \varepsilon > 0.$$

Letting here $\varepsilon \to 0$, we get (7.2). □

PROOF OF THEOREM 7.1. From Sections 2 and 3, we know that, in this case, the depth function reaches its maximum at 1/2; hence, for a given distribution P, and the corresponding random variable X, we have that $C(P,\mathcal{U}) = \{\operatorname{Med} X\}$, where the median set is taken with respect to the given partial order $\preceq$, and $\{\operatorname{Med} X\} = [\boldsymbol{a}_0, \boldsymbol{b}_0]$ for some $\boldsymbol{a}_0, \boldsymbol{b}_0 \in \mathbf{R}^d$.

Then, by Theorem 7.2, $f(m) \leq Q_{1/2}$, for any $m \in \{\operatorname{Med} X\}$. If, besides $Q_{1/2}$, any other median M of $f(X)$ exists, then we have that $P(f(X) \leq M) = 1/2$, hence the set $V_M = \{x \mid f(x) \leq M\}$ has the probability $P(V_M) = 1/2$. Therefore, $V_M = [\boldsymbol{a}, \boldsymbol{b}]$ has a non empty intersection with any interval $V_\alpha = [\boldsymbol{a}^\alpha, \boldsymbol{b}^\alpha]$ with $P(V_\alpha) > \frac{1}{2}$. Then, as in the proof of Lemma A.1, it follows that $\boldsymbol{a}^\alpha \preceq \boldsymbol{b}$ and $\boldsymbol{a} \preceq \boldsymbol{b}^\alpha$ for all α, which implies, via relations $\boldsymbol{a}_0 = \sup_\alpha \boldsymbol{a}^\alpha$ and $\boldsymbol{b}_0 = \inf_\alpha \boldsymbol{b}^\alpha$, that

$$\boldsymbol{a}_0 \preceq \boldsymbol{b} \quad \text{and} \quad \boldsymbol{a} \preceq \boldsymbol{b}_0,$$

hence, $[\boldsymbol{a}, \boldsymbol{b}] \cap [\boldsymbol{a}_0, \boldsymbol{b}_0] = [\boldsymbol{a}_0, \boldsymbol{b}_0] \neq \emptyset$. Then the inequality (7.1) holds with any $m \in [\boldsymbol{a}_0, \boldsymbol{b}_0]$. □

References

[1] B. Chakraborty and P. Chaudhuri, *On a transformation and re-transformation technique for constructing an affine equivariant multivariate median*, Proc. Amer. Math. Soc. **124 (8)** (1996), 2539–2547.

[2] Z. Chen and D. Tyler, *The influence function and maximum bias of Tukey's median*, Ann. Statist. **30 (6)** (2002), 1737–1759.

[3] D.L. Donoho, *Breakdown properties of a multivariate location estimators* PhD thesis, Department of Statistics, Harvard University (1982).

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

[5] T.P. Hettmansperger and R.H. Randles, *A practical affine equivariant multivariate median*, Biometrika **89 (4)** (2002), 851–860.

[6] G. Koshevoy and K. Mosler, *The Lorenz zonoid of a multivariate distribution*, J. Amer. Statist. Assoc. **91** (1996), 873–882.

[7] R.Y. Liu, *On a notion of simplical depth*, Proc. Nat. Acad. Sci. U.S.A. **85** (1988), 1732–1734.

[8] R.Y. Liu, *On a notion of data depth based upon random simplices*, Ann. Statist. **18** (1990), 405–414.

[9] R.Y. Liu, J.M. Parelius, and K. Singh, *Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion)*, Ann. Stat. **27** (1999), 783–858.

[10] R.Y. Liu and K. Singh, *Ordering directional data : concepts of data depth on circles and spheres*, Ann. Statist. **20 (3)** (1992), 1468–1484.

[11] J.-C. Masse and R. Theorodescu, *Halfplane trimming for bivariate distributions*, J. Multivariate Anal. **48** (1994), 188–202.

[12] M. Merkle, *Jensen's inequality for medians*, Stat. Prob. Letters **71** (2005), 277–281.

[13] A. Niinimaa and H. Oja, *Multivariate median*, Encyclopedia of Statistical Sciences (Update Volume 3). Eds. by Kotz, S., Johnson, N.L. and Read, C.P., Wiley, New York (1999), 497–505.

[14] D. Nolan, *Asymptotics for multivariate trimming*, Stochastic Processes And Applications **42** (1992), 157–169.

[15] Tyrrell R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton (1972).

[16] Peter J. Rousseeuw, Ida Ruts, *The depth function of a population distribution*, Metrika **49**(1999), 213–244.

[17] E. Schechter, *Handbook of Analysis and Its Foundations*, Academic Press, New York (1997).

[18] C.G. Small, *Measures of centrality for multivariate and directional distributions*, Canadian J. Statistics, **15** (1987), 31–39.

[19] C.G. Small, *A survey of multidimensional medians*, Internat. Stat. Inst. Rev. **58** (1990), 263–277.

[20] J.W. Tukey, *Mathematics and picturing of data*, Proc. International Congress of Mathematicians, Vancouver,1974 **volume 2** (1975), 523–531, ISBN 0-919558-04-6.

[21] Y. Vardi and C.-H. Zhang, *The multivariate L_1-median and associated data depth*, Nat. Acad. Sci. **97** (2000), 1423–1426.

[22] S. Willard, *General Topology*, Addison-Weseley (1970).

[23] Y. Zuo, *Projection-based depth functions and associated medians*, Ann. Stat. **31** (2003), 1460–1490.

[24] Y. Zuo, *Projection based affinne equivariant multivariate location estimators with the best possible finite sample breakdown point*, Statist. Sinica **14: (4)** (2004), 1199–1208.

[25] Y. Zuo, *Multi-dimensional trimming based on projection depth*, Ann. Stat. **34 (5)** (2006), 42 pages.

[26] Y. Zuo, H. Cui, and D. Young, *Influence function and maximum bias of projection depth based estimators*, Ann. Stat. **32 (1)** (2004), 189–218.

[27] Y. Zuo and R. Serfling, *General notions of statistical depth function*, Ann. Stat. **28** (2000a), 461–482.

[28] Y. Zuo and R. Serfling, *Nonparametric notions of multivariate "scatter measure" and "more scattered" based on statistical depth functions*, J. Multivariate Anal. **75** (2000b), 62–78.

[29] Y. Zuo and R. Serfling, *Structural properties and convergence results for contours of sample statistical depth functions*, Ann. Stat. **28** (2000c), 483–499.