Uniform convergence rates for the approximated halfspace and projection depth

Stanislav Nagy[*]

Charles University,
Faculty of Mathematics and Physics,
Department of Probability and Math. Statistics,
Prague, Czech Republic
e-mail: nagy@karlin.mff.cuni.cz

Rainer Dyckerhoff

University of Cologne,
Institute of Econometrics and Statistics,
Cologne, Germany
e-mail: rainer.dyckerhoff@statistik.uni-koeln.de

and

Pavlo Mozharovskyi

LTCI, Télécom Paris,
Institut polytechnique de Paris, France
e-mail: pavlo.mozharovskyi@telecom-paris.fr

Abstract: The computational complexity of some depths that satisfy the projection property, such as the halfspace depth or the projection depth, is known to be high, especially for data of higher dimensionality. In such scenarios, the exact depth is frequently approximated using a randomized approach: The data are projected into a finite number of directions uniformly distributed on the unit sphere, and the minimal depth of these univariate projections is used to approximate the true depth. We provide a theoretical background for this approximation procedure. Several uniform consistency results are established, and the corresponding uniform convergence rates are provided. For elliptically symmetric distributions and the halfspace depth it is shown that the obtained uniform convergence rates are sharp. In particular, guidelines for the choice of the number of random projections in order to achieve a given precision of the depths are stated.

---

1. Introduction

Data depth is a general concept that intends to provide a base for nonparametric statistical analysis of multivariate data. The idea is to quantify the centrality of each point x in $\mathbb{R}^d$, with respect to a given dataset. Points that tend to lie in the central bulk of the data are assigned high depth values; points that fail to follow the prevailing pattern of the observations are flagged by their low depth. One obtains a data-dependent ranking of points in $\mathbb{R}^d$, that enables constructions of analogues of quantiles, ranks, and orderings, applicable to multivariate datasets. Many definitions of depths appeared in the literature; we refer to [45, 25, 30] and references therein. Here we focus on two important depths that share common traits — the halfspace depth of Tukey [39], and the (generalized) projection depth whose history traces back to Stahel [37] and Donoho [12].

While theoretically the depths are appealing, it is in practice often difficult to evaluate them exactly. For instance, the computation of the halfspace depth of a single point in arbitrary dimension is known to be NP-hard [22]. Therefore, a great deal of research has focused on procedures that approximate the true depth [15, 8, 6, 35, 4, 36]. A particularly simple upper bound on the halfspace and projection depth can be devised if one uses their so-called *projection property* [15], which means that the overall (multivariate) depth of a point x is expressed as the infimum of (univariate) depths of projections of x with respect to the projected dataset. This suggests the following approximation procedure: (i) draw a random sample of n directions U_i, $i = 1, \ldots, n$, uniformly from the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$; (ii) evaluate the (univariate) depths of $\langle U_i, x \rangle$ with respect to the dataset projected onto U_i for each i; and (iii) approximate the depth of x by the minimum of these numbers. This approximation was first proposed in [15]; for the halfspace depth, it leads to what is sometimes called the random Tukey (or halfspace) depth [8].

Here we study the statistical properties of this approximation procedure. We are interested in conditions under which uniform convergence of the approximated depth to its theoretical counterpart can be guaranteed, and its uniform convergence rates. It turns out that for finite datasets, uniform convergence is never achieved. Thus, we focus on general probability measures on $\mathbb{R}^d$, and show that under appropriate regularity conditions, uniform approximations of the depths are valid, and their sharp convergence rates are possible to be devised. These results lend valuable insights into the behaviour of the approximated depths. They allow us to state general guidelines for the number of directions n that need to be taken in order to achieve a desired precision. Nevertheless, even in very regular models, n grows fast with the dimension d. Above $d = 5$, for reasonable n the considered randomization scheme does not attain precision sufficient for practical applications, and more elaborate approximation methods should be preferred.

In Sections 2, 3 and 4 we deal with the halfspace depth. Section 2 introduces notations and fixes the basic ideas of the paper. Our two main theorems, of a rather technical nature, are given in Section 3. Their applications to different classes of probability distributions are discussed in Section 4. Explicit, and

exact, rates of convergence are derived for (i) general unimodal elliptically symmetric distributions; (ii) multivariate Gaussian distributions; (iii) multivariate t-distributions; (iv) uniform distributions on balls; and (v) the general collection of multivariate p-symmetric measures [18]. We give explicit guidelines for the choice of the parameter n that allow to achieve a pre-specified quality of the approximation. In Section 4.5 we discuss situations when uniform approximation cannot be achieved; in particular, we deal with empirical measures. Extensions to the (generalized, or asymmetric) projection depth are treated in Section 5. Some concluding remarks can be found in Section 6. All the proofs are deferred to the appendix — Appendix A for the halfspace depth, and Appendix B for the generalized projection depths.

2. Halfspace depth and its approximations

Let $(\Omega, \mathcal{A}, \mathsf{P})$ be the probability space on which all random elements are defined. Denote by $\mathcal{P}\left(\mathbb{R}^d\right)$ the space of all Borel probability measures on $\mathbb{R}^d$ equipped with the Euclidean norm $\|\cdot\|$ and the scalar product $\langle \cdot, \cdot \rangle$, $d = 1, 2, \ldots$. We write $\mathbb{S}^{d-1} = \left\{ x \in \mathbb{R}^d \colon \|x\| = 1 \right\}$ for the unit sphere in $\mathbb{R}^d$. Notation $X \sim P$ stands for a random variable X whose distribution is $P \in \mathcal{P}\left(\mathbb{R}^d\right)$.

For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ and $x \in \mathbb{R}^d$, the *halfspace depth* (or *Tukey depth*) of x with respect to $X \sim P$ is defined as

$$HD(x; P) = \inf_{u \in \mathbb{S}^{d-1}} \mathsf{P}\left(\langle u, X \rangle \leq \langle u, x \rangle\right). \tag{1}$$

Where no mistake can be made, $HD(x; P)$ is shortened to $HD(x)$. The halfspace depth was proposed in [39], see also [12, 13] and [32].

For $n = 1, 2, \ldots$ consider $U_1, \ldots, U_n$ a random sample from the uniform distribution on $\mathbb{S}^{d-1}$. We are interested in the quality of the approximation of the depth (1) by its randomized counterpart [15, Section 6]

$$HD_n(x; P) = \min_{i=1,\ldots,n} \mathsf{P}\left(\langle U_i, X \rangle \leq \langle U_i, x \rangle\right). \tag{2}$$

Again, $HD_n(x; P)$ will be shortened to $HD_n(x)$. The quantity $HD_n(x; P)$ is a random variable that depends on the sample of directions $U_1, \ldots, U_n \in \mathbb{S}^{d-1}$, and thus also on $\omega \in \Omega$. For each $\omega \in \Omega$ fixed, it is evident that

$$HD_n(x) \geq HD_{n+1}(x) \geq HD(x) \quad \text{for all } x \in \mathbb{R}^d \text{ and } n = 1, 2, \ldots$$

Furthermore, for $d = 1$ the random depth HD_n with high probability reduces to the true depth HD

$$\mathsf{P}\left(\sup_{x \in \mathbb{R}} |HD_n(x) - HD(x)| = 0\right) = 1 - 2\left(\frac{1}{2}\right)^n \quad \text{for all } n = 1, 2, \ldots$$

In the interesting case $d > 1$, for any $x \in \mathbb{R}^d$ the random $HD_n(x)$ approximates $HD(x)$. In [15, Proposition 11] it was shown that for any $x \in \mathbb{R}^d$ and $P \in \mathcal{P}\left(\mathbb{R}^d\right)$

the convergence $HD_n(x) \xrightarrow[n\to\infty]{\text{a.s.}} HD(x)$ holds true. Here we are interested in establishing uniform extensions of that convergence result, and deriving the corresponding rates of convergence; for an array of applications of these results to the computation of the depth see [31, Section 2.3].

Define the *halfspace function* of $X \sim P$, given for $x \in \mathbb{R}^d$ by

$$\varphi_x \colon \mathbb{S}^{d-1} \to [0,1] \colon u \mapsto \mathsf{P}\left(\langle u, X\rangle \le \langle u, x\rangle\right). \tag{3}$$

Both $HD(x)$ and its approximation $HD_n(x)$ can be expressed in terms of φ_x as

$$HD(x) = \inf_{u \in \mathbb{S}^{d-1}} \varphi_x(u), \quad \text{and} \quad HD_n(x) = \min_{i=1,\dots,n} \varphi_x(U_i). \tag{4}$$

If the minimum value of φ_x is attained in a single direction in $\mathbb{S}^{d-1}$, we denote this direction by $\widetilde{u}(x) \in \mathbb{S}^{d-1}$, that is $HD(x) = \varphi_x\left(\widetilde{u}(x)\right)$.

When deriving the rates of convergence of the depth approximations that are uniform on a set $C \subset \mathbb{R}^d$, it is necessary to assume a certain form of the equicontinuity of the halfspace functions

$$\{\varphi_x \colon x \in C\}. \tag{5}$$

In the first step, in Section 3.1 we derive our main result under the assumption of uniformly Lipschitz continuous functions φ_x. In the second step in Section 3.2 we extend the previous theorem, and deal with a general equicontinuous class of functions (5).

3. Approximation of the halfspace depth: Main results

We are interested in the rate of convergence of the random sequence

$$\Delta_n(C) = \sup_{x \in C} |HD_n(x) - HD(x)|$$

as $n \to \infty$. We start by two technical results that will be of great importance in the sequel. For $\Gamma\left(\cdot\right)$ the gamma function, denote by

$$a_d(\varphi) = \frac{\Gamma\left(d/2\right)}{\Gamma\left((d-1)/2\right)\sqrt{\pi}} \int_0^\varphi \left(\sin(\theta)\right)^{d-2} \,\mathrm{d}\theta, \quad \text{for } \varphi \in [0,\pi] \tag{6}$$

the $(d-1)$-dimensional Hausdorff measure of a cap of a polar angle[1] φ of a sphere in $\mathbb{R}^d$ with unit surface (hyper-)area. In other words, $a_d\left(\varphi\right)$ is the ratio of the surface area of the spherical cap of polar angle φ in $\mathbb{S}^{d-1}$, and the surface area of the whole $(d-1)$-sphere $\mathbb{S}^{d-1}$. The function a_d takes particularly simple forms for $d=2$ and $d=3$

$$a_2\left(\varphi\right) = \frac{\varphi}{\pi} \quad \text{and} \quad a_3\left(\varphi\right) = \left(\sin\left(\frac{\varphi}{2}\right)\right)^2.$$

For higher dimensions d, it is a trigonometric polynomial. Note that all functions a_d are strictly increasing and continuous on their domain. Denote the inverse of a_d by $a_d^{-1} \colon [0,1] \to [0,\pi]$.

---

[1] The angle between the rays from the center of the hypersphere to the apex of the cap, and the edge of the $(d-1)$-dimensional disk that forms the base of the cap.

3.1. Lipschitz continuous halfspace functions

For starters, assume that the set of functions (5) is uniformly Lipschitz continuous with constant $L \geq 0$, i.e. that

$$\sup_{x \in C} |\varphi_x(u) - \varphi_x(v)| \leq L \|u - v\|_g \quad \text{for all } u, v \in \mathbb{S}^{d-1}, \tag{7}$$

where $\|u - v\|_g = \arccos(\langle u, v \rangle)$ is the great-circle distance, i.e. the geodesic distance between u and v. Note that the great-circle distance and the Euclidean distance are related via

$$\|u - v\| = \sqrt{2 \left(1 - \cos\left(\|u - v\|_g\right)\right)} = 2 \sin\left(\|u - v\|_g / 2\right).$$

This relation follows easily from the cosine formula and the half-angle formula. A condition similar to (7) was used in [5] when deriving the uniform rates of convergence for the ordinary empirical halfspace depth process. Examples of distributions that satisfy property (7) will be listed in Section 4.

Theorem 1. *Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be such that (7) holds true for a set $C \subset \mathbb{R}^d$. Then*

$$\limsup_{n \to \infty} \frac{n\, a_d\left(\Delta_n(C)/L\right) - \log n}{\log \log n} \leq d \quad \text{a.s.}$$

The detailed proof of Theorem 1 can be found in Appendix A.1. Its main idea rests in the introduction of the quantity

$$S_n = \sup_{u \in \mathbb{S}^{d-1}} \min_{i=1,\ldots,n} \|u - U_i\|_g,$$

the largest distance of a point on the sphere to the closest random direction $U_1, \ldots, U_n$. Suppose that the halfspace functions (5) satisfy certain uniform equicontinuity conditions (at least in neighbourhoods of their arguments of minima at $x \in C$), such as (7) considered in this section, or (9) introduced below. Formula (4) then gives that it may be expected that $\Delta_n(C)$ behaves similarly as S_n with $n \to \infty$. We formalize this argument, and combine it with the law of the iterated logarithm for S_n known from the literature [20, 21]

$$\limsup_{n \to \infty} \frac{n\, a_d(S_n) - \log n}{\log \log n} = d \quad \text{a.s.,}$$

to obtain rates such as those from Theorem 1.

3.2. General uniformly continuous projections

Now, assume that the class of functions (5) is uniformly equicontinuous, but not necessarily uniformly Lipschitz. Consider the minimal modulus of continuity of this class of functions

$$\delta \colon [0, \pi] \to [0, 1] \colon t \mapsto \sup_{x \in C} \sup_{u, v \in \mathbb{S}^{d-1}, \, \|u - v\|_g \leq t} |\varphi_x(u) - \varphi_x(v)|. \tag{8}$$

Using the minimal modulus of continuity we can bound the distance between $\varphi_x(u)$ and $\varphi_x(v)$ uniformly for all $x \in C$,

$$\sup_{x \in C} |\varphi_x(u) - \varphi_x(v)| \leq \delta \left(\|u - v\|_g \right) \quad \text{for all } u, v \in \mathbb{S}^{d-1}. \tag{9}$$

Under the condition of uniform equicontinuity of (5), the function $\delta(t)$ is continuous, non-negative and non-decreasing, with $\delta(0) = 0$, see [9, Chapter 2, § 6]. With no loss of generality, in the sequel we assume that δ is increasing, and denote its inverse function by δ^{-1}. If δ is constant on some interval, proper use of a generalized inverse function of δ in place of δ^{-1} yields the same conclusions [17]. The proof of the next result is in Appendix A.2.

Theorem 2. *Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ and $C \subset \mathbb{R}^d$ be such that the function δ from (8) is continuous at 0 with an inverse function δ^{-1}. Then*

$$\limsup_{n \to \infty} \frac{n\, a_d \left(\delta^{-1}\left(\Delta_n(C) \right) \right) - \log n}{\log \log n} \leq d \quad \text{a.s.} \tag{10}$$

If (5) is uniformly Hölder continuous, i.e. for some $K > 0$ and $\alpha \in (0, 1]$ we may choose $\delta(t) \leq Kt^\alpha$, the last formula gives

$$\limsup_{n \to \infty} \frac{n\, a_d \left(\left(\Delta_n(C)/K \right)^{1/\alpha} \right) - \log n}{\log \log n} \leq d \quad \text{a.s.} \tag{11}$$

This expression is possible to be rewritten in terms of $\Delta_n(C)$ directly. To do this, first note that using L'Hôpital's rule we know that

$$\lim_{\varphi \to 0+} \frac{a_d(\varphi)}{\varphi^{d-1}} = \lim_{\varphi \to 0+} \frac{\Gamma\left(d/2\right)}{\Gamma\left((d-1)/2\right)\sqrt{\pi}(d-1)} \left(\frac{\sin(\varphi)}{\varphi} \right)^{d-2}$$
$$= \frac{\Gamma\left(d/2\right)}{\Gamma\left((d-1)/2\right)\sqrt{\pi}(d-1)}.$$

Denote the constant on the right hand side by $c_d \in (0, \infty)$. Substitution $s = a_d(\varphi)$ provides

$$c_d^{-1/(d-1)} = \left(\lim_{\varphi \to 0+} \frac{a_d(\varphi)}{\varphi^{d-1}} \right)^{-1/(d-1)} = \lim_{s \to 0+} \left(\left(\frac{s^{1/(d-1)}}{a_d^{-1}(s)} \right)^{d-1} \right)^{-1/(d-1)}$$
$$= \lim_{s \to 0+} \frac{a_d^{-1}(s)}{s^{1/(d-1)}},$$

meaning that $a_d^{-1}(s) = \mathcal{O}(s^{1/(d-1)})$ as $s \to 0$ from the right. That yields, in an inversion of formula (10), that for the dimension d fixed, we have under the same conditions as (10) was derived,

$$\Delta_n(C) = \mathcal{O} \left(\delta \left(\left(\frac{\log n}{n} \right)^{1/(d-1)} \right) \right) \quad \text{a.s.} \tag{12}$$

Note that in (12) the function δ from (9) appears. The rate at which this function approaches zero as $t \to 0$ from the right directly affects the error rate $\Delta_n(C)$ — the faster δ decreases to zero, the faster convergence of $\Delta_n(C)$ can be expected.

In the special case of Hölder continuous functions (5) satisfying (11), formula (12) gives

$$\Delta_n(C) = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{\alpha/(d-1)}\right) \quad \text{a.s.}$$

With $\alpha = 1$ we obtain the corresponding rate for uniformly Lipschitz continuous functions (5).

4. Applications to probability distributions

4.1. Approximations and affine invariance

A depth D is said to be affine invariant if for any $X \sim P \equiv P_X \in \mathcal{P}\left(\mathbb{R}^d\right)$, $A \in \mathbb{R}^{d \times d}$ non-singular, and $b \in \mathbb{R}^d$

$$D(x; P_X) = D(Ax + b; P_{AX+b}) \qquad \text{for all } x \in \mathbb{R}^d.$$

Affine invariance is a desired property of a depth function. Both the halfspace depth and the (generalized) projection depth satisfy it.

The following theorem asserts that for the halfspace depth, computation of approximations with respect to affine images of a measure affects the resulting rates of convergence only by a multiplicative constant. This allows us to focus in some examples in the sequel only on the isotropic situation, which typically means that the measure is considered to be centred to have the expected value in the origin, and the scatter matrix a multiple of the identity matrix. The proof of the next result can be found in Appendix A.3.

Theorem 3. *Let Theorem 2 hold true. Then for $A \in \mathbb{R}^{d \times d}$ non-singular, $b \in \mathbb{R}^d$, and $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$ there exists a constant $K > 0$ that depends only on A such that for*

$$\widetilde{\Delta}_n(C) = \sup_{x \in AC+b} |HD_n(x; P_{AX+b}) - HD(x; P_{AX+b})|$$

it holds true that

$$\limsup_{n \to \infty} \frac{n \, a_d \left(\delta^{-1}\left(\widetilde{\Delta}_n(C)\right)/K\right) - \log n}{\log \log n} \leq d \quad \text{a.s.}$$

4.2. General distributions with densities

Let us now explore when the equicontinuity conditions (7) or (9) hold true. The first interesting case is that of a bounded set $C \subset \mathbb{R}^d$.

Theorem 4. *Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be a distribution such that*

$$P(u^\mathsf{T} X = \alpha) = 0 \text{ for all } u \in \mathbb{S}^{d-1} \text{ and } \alpha \in \mathbb{R} \tag{13}$$

that is concentrated in a bounded subset C of $\mathbb{R}^d$, that is $P(C) = 1$. Then $\Delta_n\left(\mathbb{R}^d\right) \xrightarrow[n\to\infty]{\text{a.s.}} 0$.

Condition (13) is satisfied if P admits a density in $\mathbb{R}^d$. If the density is bounded, an explicit rate of convergence of $\Delta_n(C)$ can be provided.

Theorem 5. *Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be a distribution with a density bounded from above by a constant $M > 0$, that vanishes outside a bounded set $C \subset \mathbb{R}^d$. Let further be $\operatorname{diam}(C) = \sup\{\|x - y\| : x, y \in C\}$ the diameter of C. Then the rate from Theorem 1 holds true with*

$$L = \frac{M\pi^{d/2-1}\left(\operatorname{diam}(C)\right)^d}{\Gamma\left(d/2+1\right)}.$$

For the proofs of these two theorems see Appendix A.4 and A.5. Note that the rate from Theorem 5 is quite general, yet usually weak.

4.3. Elliptically symmetric distributions

A much finer result can be shown for elliptically symmetric distributions with unimodal densities on $\mathbb{R}^d$. By a unimodal elliptically symmetric density we understand a density

$$f(x) = |\Sigma|^{-1/2}\, g\left((x-\mu)^\mathsf{T} \Sigma^{-1}\, (x-\mu)\right) \quad \text{for } x \in \mathbb{R}^d,$$

with $\Sigma \in \mathbb{R}^{d\times d}$ non-singular, $\mu \in \mathbb{R}^d$, and $g\colon [0,\infty) \to [0,\infty)$ a non-increasing scalar function. This important collection of distributions covers all Gaussian distributions, or multivariate extensions of the t-distributions. In this situation we are allowed to consider $C = \mathbb{R}^d$, and study the uniform convergence of the approximated depth over the whole sample space. In accordance with the discussion in Section 4.1 we primarily focus on distributions with Σ the identity matrix. Such distributions are frequently called spherically symmetric distributions.

Theorem 6. *Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be a distribution with an elliptically symmetric density and Σ the identity matrix. Then $\Delta_n(\mathbb{R}^d) \xrightarrow[n\to\infty]{\text{a.s.}} 0$. If, moreover, the density of P is unimodal, then the rate in Theorem 2 holds true with both function δ taken as $\delta_1(t) = (1 - \cos(t))/2$, and $\delta_2(t) = t^2/4$, with $C = \mathbb{R}^d$.*

Note that in the proof of the previous theorem, given in Appendix A.6, for F_p the distribution function of the random variable $\langle u, X\rangle$, $u \in \mathbb{S}^{d-1}$, the better bound

$$\sup_{t\geq 0}\left(F_p(t) - F_p(t\cos(\varepsilon))\right) \leq (1 - \cos(\varepsilon))/2$$

is still not the tightest possible one. Thus, Theorem 6 is suboptimal for particular choices of P. For elliptically symmetric distributions, define the *tight modulus of continuity* of the halfspace functions (the modulus of continuity of F_p evaluated near the argument of the minimum of φ_x) by

$$\delta(\varepsilon) = \sup_{t \geq 0} \left(F_p(t) - F_p(t \cos(\varepsilon)) \right) \quad \text{for } \varepsilon \in [0, \pi]. \tag{14}$$

Using the same proof as before with this choice of the modulus δ, it can be asserted that for all n large enough the rate from Theorem 2 holds true for P also with δ given in (14) and $C = \mathbb{R}^d$.

Interestingly, the latter bound using the modulus (14) can be shown to be optimal for spherically symmetric distributions. For this, recall that for two sequences of real numbers $a = \{a(n)\}_{n=1}^{\infty}$ and $b = \{b(n)\}_{n=1}^{\infty}$ we say that a is asymptotically bounded both from above and below by b, written $a(n) = \Theta(b(n))$, if

$$0 < \liminf_{n \to \infty} \frac{a(n)}{b(n)} \leq \limsup_{n \to \infty} \frac{a(n)}{b(n)} < \infty.$$

The proof of the next theorem is given in Appendix A.7.

Theorem 7. *Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be a distribution with a unimodal elliptically symmetric density with Σ the identity matrix, and let $\delta(\varepsilon)$ be defined as in (14). Then*

$$\liminf_{n \to \infty} \frac{n \, a_d \left(\delta^{-1} \left(\Delta_n(\mathbb{R}^d) \right) \right) - \log n}{\log \log n} \geq d - 2 \quad \text{a.s.}$$

In particular, for $d > 2$

$$\mathsf{P}\left(n \, a_d \left(\delta^{-1} \left(\Delta_n(\mathbb{R}^d) \right) \right) - \log n = \Theta\left(\log \log n \right) \right) = 1.$$

Specific rates of convergence can be explored with particular models at hand. Here we study three important special cases — multivariate Gaussian distributions, multivariate elliptically symmetric t-distributions, and uniform distributions on ellipsoids. In accordance with the discussion from Section 4.1, we treat only standardized distributions.

Example (Multivariate Gaussian distribution). For a standard multivariate Gaussian distribution X, $\langle u, X \rangle$ is standard univariate Gaussian for any $u \in \mathbb{S}^{d-1}$, i.e.

$$F_p(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-s^2/2) \, \mathrm{d}\, s \quad \text{for } t \in \mathbb{R}.$$

To get tight convergence rates, we must evaluate the modulus (14) for $\varepsilon > 0$ fixed. It is not difficult to find that the maximal value of the latter function is attained at

$$t^* = \frac{\sqrt{-2 \log\left(\cos(\varepsilon) \right)}}{\sin(\varepsilon)} \quad \text{for } \varepsilon \in (0, \pi/2),$$

which means that for $\|\widetilde{u}(x) - u\|_g \leq \varepsilon$ with $\varepsilon \in (0, \pi/2)$

$$\sup_{x \in \mathbb{R}^d} |\varphi_x(\widetilde{u}(x)) - \varphi_x(u)| \leq \delta(\varepsilon) = F_p(t^*) - F_p(t^* \cos(\varepsilon)) \leq (1 - \cos(\varepsilon))/2 \leq \varepsilon^2/4.$$
(15)

The first bound above is the optimal one from Theorem 7; the other two follow from Theorem 6. All of them can be used for numerical evaluation of the exact convergence rates. For that, it is enough to invert the inequality in Theorem 2 and obtain an error bound of the form $\Delta_n(C) \leq \delta\left(a_d^{-1}\left((d \log \log n + \log n)/n\right)\right)$ which will approximately hold true almost surely for n large enough. A comparison of the optimal uniform approximation errors for Gaussian distributions can be found in Table 1. Graphically, these errors are compared with the general bounds from Theorem 6 in Figure 1, see also the function $\delta(\varepsilon)$ from (15) displayed in Figure 3. The tight bound presents a substantial improvement over the general bounds from Theorem 6.

TABLE 1

Uniform error bounds for the approximation of the halfspace depth for elliptically symmetric unimodal distributions. Ellipt. 1 and 2 stand for the general bounds applicable to unimodal elliptically symmetric distributions from Theorem 6 and δ_1, δ_2, respectively. 2-sym. is the general bound from Theorem 8 for $p = 2$ and δ_1.

	n	$d = 2$	$d = 3$	$d = 5$	$d = 10$	$d = 20$
Ellipt. 1	10^2	0.01441	0.09187	0.21855	0.35805	0.45576
	10^3	0.00029	0.01271	0.07629	0.20148	0.31099
	10^4	0.00000	0.00159	0.02625	0.11841	0.22779
	10^5	0.00000	0.00019	0.00892	0.07080	0.17139
Ellipt. 2	10^2	0.01448	0.09483	0.23663	0.41149	0.54924
	10^3	0.00029	0.01276	0.07831	0.21669	0.34995
	10^4	0.00000	0.00159	0.02648	0.12340	0.24756
	10^5	0.00000	0.00019	0.00895	0.07254	0.18219
2-sym.	10^2	0.24005	0.60619	0.93498	1.19675	1.35021
	10^3	0.03384	0.22544	0.55240	0.89774	1.11532
	10^4	0.00429	0.07968	0.32404	0.68821	0.95456
	10^5	0.00052	0.02745	0.18890	0.53218	0.82798
Gaussian	10^2	0.00707	0.04896	0.13535	0.27027	0.40894
	10^3	0.00014	0.00623	0.03997	0.12211	0.21854
	10^4	0.00000	0.00077	0.01305	0.06500	0.14277
	10^5	0.00000	0.00009	0.00436	0.03688	0.10010
Cauchy	10^2	0.00465	0.03226	0.09022	0.18834	0.31595
	10^3	0.00009	0.00410	0.02632	0.08119	0.14906
	10^4	0.00000	0.00051	0.00858	0.04288	0.09532
	10^5	0.00000	0.00006	0.00287	0.02428	0.06632
Uniform	10^2	0.00930	0.05831	0.14912	—	—
	10^3	0.00018	0.00743	0.04430	—	—
	10^4	0.00000	0.00092	0.01447	—	—
	10^5	0.00000	0.00011	0.00483	—	—

Example (Multivariate t-distribution). Consider the multivariate t-distribution $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ with ν degrees of freedom, whose density is given by

$$f(x) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma(\nu/2)(\nu\pi)^{d/2}|\Sigma|^{1/2}}\left(1 + \frac{1}{\nu}(x-\mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)\right)^{-(\nu+d)/2} \quad \text{for } x \in \mathbb{R}^d,$$
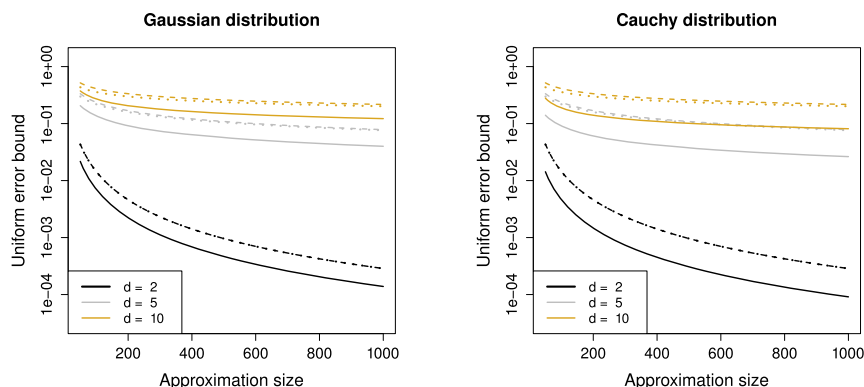
FIG 1. *Uniform error bounds for $\Delta_n(\mathbb{R}^d)$ as a function of n: the (standard) multivariate Gaussian distribution (left panel) and the (standard) multivariate Cauchy distribution (right panel). The solid curves represent the tight error bounds from Theorem 7; the dashed and the dotted lines represent the two respective bounds from Theorem 6, valid for any unimodal elliptically symmetric distribution. The dashed and the dotted lines are nearly identical in the plot. Note the logarithmic scale of the vertical axes.*

where $\Sigma \in \mathbb{R}^{d \times d}$ is non-singular, $\mu \in \mathbb{R}^d$, and $\nu = 1, 2, \ldots$. For $\nu = 1$ it is also known as the multivariate Cauchy distribution. Obviously, P satisfies the assumptions of Theorem 6.

We again focus on the case $\mu = 0$, and Σ the identity matrix. The multivariate t-distribution is then spherically symmetric about the origin, with

$$F_p(t) = \int_{-\infty}^{t} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\nu\pi}} \left(1 + \frac{s^2}{\nu}\right)^{-(\nu+1)/2} \mathrm{d}\,s \quad \text{for } t \in \mathbb{R}$$

the distribution function of the univariate standard t-distribution with ν degrees of freedom, see [24, Theorem 1].

Let us evaluate the tight modulus of continuity δ from (14) as in the previous example. For general ν, the argument of the maximum t^* of the function $t \mapsto (F_p(t) - F_p(t\cos(\varepsilon)))$ is not possible to be expressed in a closed form. For some special values of ν it can be computed that

$$t^* = \begin{cases} (\cos(\varepsilon))^{-1/2} & \text{for } \nu = 1, \\ \sqrt{\dfrac{3\left(-(\sin(\varepsilon))^2 - (\cos(\varepsilon))^{3/2} + \sqrt{\cos(\varepsilon)}\right)}{\sqrt{\cos(\varepsilon)}\left((\cos(\varepsilon))^3 - 1\right)}} & \text{for } \nu = 3, \end{cases} \quad \text{for } \varepsilon \in (0, \pi/2).$$

For other small values of ν the formulas for t^* can still be obtained explicitly, yet their expressions are already rather complicated. With t^* known, the tight bound on the approximation errors (14) can be obtained as in (15). In Figure 1 and Table 1 we see the numerically computed bounds on the approximation error in the present setup for several values of n and d. The function δ defined in (14) can be found in Figure 3. It can be seen that the approximation for

heavy-tailed distributions P is in fact more precise than for light-tailed ones. Indeed, if the tails of P are rather flat and decrease slowly, the halfspace function $\varphi_x(u)$ will never change drastically for small deviations from its argument of the minimum $\widetilde{u}(x)$.

Example (Uniform distribution on a ball). Consider the uniform distribution P on the unit ball in $\mathbb{R}^d$. The marginal distribution of $\langle u, X \rangle$ for any $u \in \mathbb{S}^{d-1}$ is given by the distribution function

$$F_p(t) = \int_{-1}^{t} \frac{\Gamma\left(d/2+1\right)\left(1-s^2\right)^{\frac{d-1}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{d+1}{2}\right)} \,\mathrm{d}\,s \quad \text{for } t \in (-1,1).$$

Note that unlike in our previous examples, F_p depends on the dimension d. For small values of d, it is possible to express the point that realises the tight bound in (15) explicitly

$$t^* = \begin{cases} \sqrt{\frac{2}{\cos(2\varepsilon)+3}} & \text{for } d = 2, \\ \sqrt{\frac{\cos(\varepsilon)-1}{(\cos(\varepsilon))^3-1}} & \text{for } d = 3. \end{cases}$$

In higher dimensions, approximations are easily obtained. Though, it is well known that as $d \to \infty$, the distribution of $\langle u, X \rangle$ is approximately Gaussian [11], which reduces the problem for higher values of d approximately to the case of the multivariate Gaussian distribution. In Table 1 and Figure 2 the error bounds are compared for $d = 2$, 3, and 5. As can be seen, the approximation is slower than for both Gaussian and t-distributions. This is explained by the discontinuous nature of the density of P — for x near the boundary of the unit ball and d small, the halfspace function φ_x deviates substantially from its minimum value already for u rather close to its minimizer $\widetilde{u}(x)$. Though, the general bound from Theorem 6 still performs somewhat worse.

4.4. p-symmetric distributions

A further important family of probability distributions whose halfspace depth can be evaluated explicitly are the p-symmetric distributions with $p \in (0, 2]$, see [18]. That family of measures extends the spherically symmetric distributions treated in Section 4.3 that can be written as p-symmetric distributions with $p = 2$, as well as the multivariate extensions of stable distributions. Recall that for $p \in (0, 2]$ a random vector $X = (X_1, \ldots, X_d)^\mathsf{T} \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is said to have a p-symmetric distribution if for any $u \in \mathbb{R}^d$ the random variable $\langle X, u \rangle$ has the same distribution as $\|u\|_p X_1$, where $\|u\|_p = \left(\sum_{i=1}^d |u_i|^p\right)^{1/p}$ for $u = (u_1, \ldots, u_d)^\mathsf{T}$, see [18, Theorem 7.1]. We shall also write $\|u\|_\infty = \max_{i=1,\ldots,d} |u_i|$. Observe that in this notation $\|\cdot\|_2 \equiv \|\cdot\|$.

For $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ a p-symmetric distribution, the exact depth $HD(x)$ was computed in [29, Example (C)] and [7, Theorem 3.1]. Let F denote the (univariate)
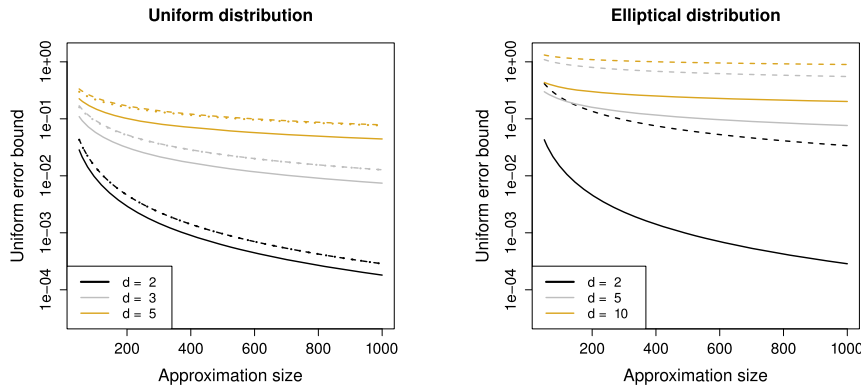
FIG 2. *Uniform error bounds for $\Delta_n(\mathbb{R}^d)$ as a function of n for the uniform distribution on the unit ball in $\mathbb{R}^d$ (left panel), and for general unimodal elliptically symmetric distributions (right panel). On the left panel, the solid curves represent the tight error bounds from Theorem 7; the dashed and the dotted lines represent the two respective bounds from Theorem 6, valid for any unimodal elliptically symmetric distribution. The dashed and the dotted lines are nearly identical in the plot. On the right panel we compare bounds from Theorem 6 with δ_1 valid for unimodal elliptical distributions (solid lines), and the more general bounds from Theorem 8 for $p = 2$ and δ_1 (dashed lines). Note the logarithmic scale of the vertical axes.*
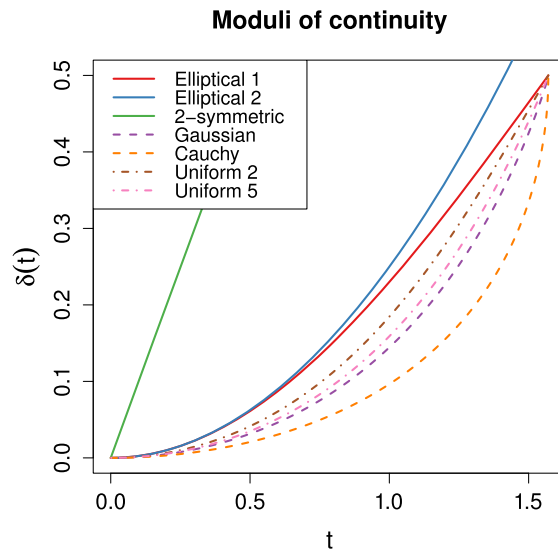

FIG 3. *The moduli of continuity $\delta(t)$ of the halfspace functions φ_x evaluated near their minima for various models. Elliptical 1 and 2 stand for the moduli from Theorem 6 with δ_1 and δ_2, respectively; 2-symmetric is the modulus δ_1 from Theorem 8 for $p = 2$; uniform 2 and 5 are the tight moduli for the uniform distribution with $d = 2$ and 5, respectively. The faster this function is decreases as $t \to 0$, the better approximation rate can be achieved.*

distribution function of the marginal X_1 of X, and set

$$q = \begin{cases} \infty & \text{for } p \in (0, 1], \\ p/(p-1) & \text{for } p \in (1, 2] \end{cases} \tag{16}$$

to be the conjugate index to p. As demonstrated in [7, Lemma A.1 and Theorem 3.1], we can write

$$\varphi_x(u) = F\left(\frac{\langle x, u \rangle}{\|u\|_p}\right), \quad \text{and} \quad HD(x) = 1 - F\left(\|x\|_q\right) = F\left(-\|x\|_q\right). \tag{17}$$

The following result generalizes Theorem 6. Its proof is in Appendix A.8.

Theorem 8. *For $X = (X_1, \ldots, X_d)^\mathsf{T} \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$ with a p-symmetric distribution such that the density of X_1 is unimodal we have that $\Delta_n(\mathbb{R}^d) \xrightarrow[n \to \infty]{\text{a.s.}} 0$. Furthermore, the rate of convergence from Theorem 2 holds true with $C = \mathbb{R}^d$ and either*

$$\delta_1(t) = \begin{cases} d^{1/p-1/2}\left(d^{1/2-1/q} + 1\right)\sin\left(t/2\right) & \text{for } p \in [1, 2], \\ d^{1/p}\sin\left(t/2\right) + \frac{d^{1/p-1/2}}{p}2^{p-1}\sin^p\left(t/2\right) & \text{for } p \in (0, 1), \end{cases}$$

or the weaker

$$\delta_2(t) = \begin{cases} d^{1/p-1/2}\left(d^{1/2-1/q} + 1\right)t/2 & \text{for } p \in [1, 2], \\ d^{1/p}t/2 + \frac{d^{1/p-1/2}}{p}t^p/2 & \text{for } p \in (0, 1), \end{cases}$$

where in both expressions $t \in [0, \pi]$ and q is given by (16).

For $p = 2$ we can compare the obtained rate with that from Theorem 6 and see that here the bounds are somewhat weaker. That is not surprising, since Theorem 8 holds under much more general conditions. The unimodality assumption is satisfied for most commonly used p-symmetric distribution. For instance, it holds true for symmetric p-stable distributions [42, Theorem 2.7.4].

4.5. Non-uniform approximation

If the measure P is not regular enough, uniform approximations of the halfspace depth may not be attainable. We illustrate our claim in two examples. In the first one we construct distributions that are atomic, yet their approximated depth fails to converge uniformly to its true counterpart. In the second example we demonstrate that if the density of P is unbounded, uniform rates of convergence of the halfspace depth approximation are not possible to be derived. The latter result should be compared with Theorems 4 and 5.

Example. If P does not admit a density, the halfspace functions φ_x may contain discontinuities. In particular, this occurs when the depth is computed with respect to the empirical measure of a random sample of observations, i.e. for the
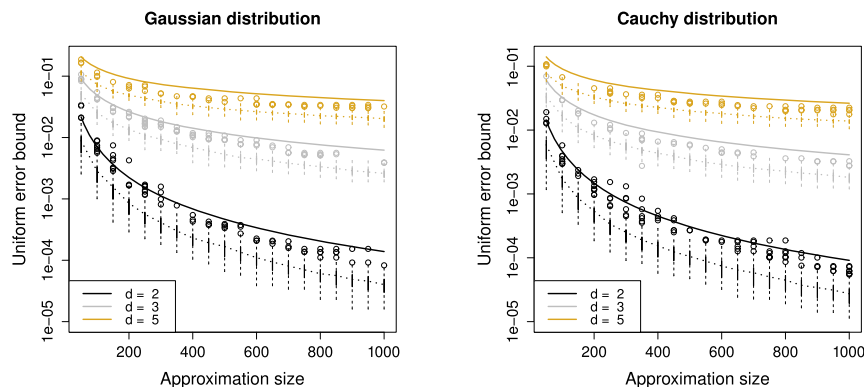
FIG 4. *The tight uniform bounds (solid lines) for $\Delta_n(\mathbb{R}^d)$ from Figure 1 for the (standard) multivariate Gaussian distribution (left panel) and the (standard) multivariate Cauchy distribution (right panel). The bounds are complemented with boxplots of simulated data, and the dotted lines that interpolate the corresponding simulated means. In each of the 100 independent runs, the supremum in $\Delta_n(\mathbb{R}^d)$ was approximated by the maximum value of $|HD(x_i) - HD_n(x_i)|$ for 500 points x_i sampled independently from the true distribution. The theoretical upper bounds match very well with the true approximation error.*

sample halfspace depth. In that case, it can be inferred that the convergence of the approximations is not uniform, even on bounded sets C.

Consider the example of P atomic, supported in a set $\{x_i\}_{i=1}^m$ in $\mathbb{R}^d$ such that $P(\{x_i\}) = p_i$ with $\sum_{i=1}^m p_i = 1$. Let $0 < p_1 \leq p_2 \leq \cdots \leq p_m$. Without loss of generality, assume that the convex hull $K \subset \mathbb{R}^d$ of these points is d-dimensional, i.e. that K is not fully contained in an affine subspace of $\mathbb{R}^d$ of dimension lower than d. Let x be a point on a facet of K, and let $x_\varepsilon = x + \varepsilon v$ for $\varepsilon > 0$, where $v \in \mathbb{S}^{d-1}$ is the outward normal of the facet of K on which x lies. Since $x_\varepsilon \notin K$ by definition, $HD(x_\varepsilon) = 0$ for any $\varepsilon > 0$. For its approximation clearly $HD_n(x_\varepsilon) \geq p_1$ if and only if all halfspaces

$$H(x_\varepsilon, U_i) = \left\{ y \in \mathbb{R}^d \colon \langle U_i, y \rangle \leq \langle U_i, x_\varepsilon \rangle \right\}$$

whose boundary passes through x_ε with outer normal $U_i \in \mathbb{S}^{d-1}$, $i = 1, \ldots, n$, intersect K. For ε small, this will occur with high probability if the directions U_i are sampled uniformly on $\mathbb{S}^{d-1}$ — as $\varepsilon \to 0$ from the right, the condition $H(x_\varepsilon, U_i) \cap K = \emptyset$ effectively reduces to $U_i = -v$, an event of null probability. Thus, the convergence of the approximations cannot be uniform on the line segment $x + \varepsilon v$ with $\varepsilon \in (0, 1)$, and

$$\limsup_{n \to \infty} \sup_{\varepsilon \in (0,1)} (HD_n(x_\varepsilon) - HD(x_\varepsilon)) \geq p_1 > 0 \quad \text{a.s.}$$

Example. Assume that P is supported in a bounded subset of $\mathbb{R}^d$ and has a density, but that density is unbounded. By Theorem 4 the uniform convergence of the approximations holds. But, no uniformly valid rate can be derived. To see this, we construct a distribution with an arbitrarily slow convergence rate.

Let $\{x_i\}_{i=0}^{\infty}$ be distinct points on the upper half of the unit circle in $\mathbb{R}^2$, indexed in a clock-wise sense, denote $l_i = \|x_i - x_{i-1}\|/2$ for $i \geq 1$, and set $l_0 = l_1$. Then, $\{l_i\}_{i=0}^{\infty}$ is a sequence of positive numbers converging to zero. For $i \geq 0$, let B_i denote a ball centred at x_i with radius $e^{-i}l_i$, and let C_i be the intersection of B_i with the convex hull of the points $\{x_i\}_{i=0}^{\infty}$. Each C_i is a convex subset of the unit ball in $\mathbb{R}^2$, with $x_i \in C_i$, and the sets $\{C_i\}_{i=0}^{\infty}$ are pairwise disjoint. Let P be the mixture of uniform distributions on the sets C_i with mixing proportions $\{p_i\}_{i=0}^{\infty}$, where $\sum_{i=0}^{\infty} p_i = 1$ and $p_0 > p_1 > \cdots > 0$. Consider the sequence of mid-points $y_j = (x_j + x_{j-1})/2$, $j = 1, 2, \ldots$. Because each y_j lies on the boundary of the convex hull of the support of P, the depth $HD(y_j)$ with respect to P is zero for all j. Though, obviously, the single minimizer $\widetilde{u}(y_j)$ of the halfspace function φ_{y_j} is the inwards normal of the facet corresponding to x_{j-1} and x_j. If $\|u - \widetilde{u}(y_j)\|_g \geq \varepsilon$, then by the construction of P for all $j \geq j(\varepsilon) = \lceil -\log\sin\varepsilon + 1 \rceil$, the smallest integer larger than $-\log\sin\varepsilon + 1$, it can be seen that either C_j, or C_{j-1} lies inside the halfspace whose normal is u and passes through y_j. Thus,

$$\varphi_{y_j}(u) = \mathsf{P}\left(\langle u, X\rangle \leq \langle u, y_j\rangle\right) \geq p_j.$$

Using a result on maximal spacings in $\mathbb{R}$ from [10, Theorem 5.2] it can be seen that for $U_1, \ldots, U_n$ uniformly distributed on the circumference of the unit circle, it almost surely holds true that for all n large enough and any j we have $\min_{i=1,\ldots,n} \|\widetilde{u}(y_j) - U_i\|_g \geq a(n)$, for a fixed positive sequence $\{a(n)\}_{n=1}^{\infty}$ that converges to zero. This means that almost surely for all n large enough, for $j = \lceil -\log\sin a(n) + 1 \rceil$ we have

$$\sup_{x \in \mathbb{R}^2} |HD_n(x) - HD(x)| \geq HD_n(y_j) \geq p_j.$$

Because $\{p_j\}_{j=0}^{\infty}$ can be made to converge to zero slowly, this means that no universal rate of convergence can be found, if the density of P is allowed to be unbounded.

5. Extensions to generalized projection depths

Let us now focus on another example of a depth that satisfies the projection property, yet is difficult to compute exactly — the (generalized) projection depth. To define the depth, consider first the mappings

$$m\colon \mathcal{P}(\mathbb{R}) \to \mathbb{R} \quad \text{and} \quad s\colon \mathcal{P}(\mathbb{R}) \to [0, \infty],$$

that satisfy the following conditions:

- $m(aX + b) = a\,m(X) + b$ for all $a, b \in \mathbb{R}$ and $X \sim P \in \mathcal{P}(\mathbb{R})$;
- $s(aX + b) = a\,s(X)$ for all $a > 0$ and $X \sim P \in \mathcal{P}(\mathbb{R})$.

Functionals m and s are called the *location* and the *scale parameter* of X, respectively. Using m and s for univariate distributions, it is possible to follow

the ideas from [37, 12], and define the outlyingness function in a multivariate space. For $x \in \mathbb{R}^d$ and $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$, the *(projection) outlyingness* of x with respect to P is given by

$$O(x; P) = \sup_{u \in \mathbb{S}^{d-1}} \frac{\langle u, x \rangle - m(\langle u, X \rangle)}{s(\langle u, X \rangle)}. \tag{18}$$

Function $O(\cdot; P)$ measures the largest deviation of a projection of a point from the location parameter of the corresponding projected distribution.

The outlyingness function (18) is closely related to depth — high outlyingness indicates low centrality. Therefore, to construct a depth it suffices to transform the outlyingness index (18) by a non-increasing function c. A family of depth functions based on this idea is called the family of projection depths. It was studied in detail in [45, 43], and [15].

Consider a continuous function $c \colon [-\infty, \infty] \to [0, 1]$ such that $c(x) = 1$ for $x < 0$, and the restriction c to $[0, \infty]$ is bijective, and strictly decreasing. The *generalized projection depth* of $x \in \mathbb{R}^d$ with respect to $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is defined by

$$PD(x; P) = \inf_{u \in \mathbb{S}^{d-1}} c\left(\frac{\langle u, x \rangle - m(\langle u, X \rangle)}{s(\langle u, X \rangle)} \right). \tag{19}$$

The family of depths (19) was proposed in [15] as a generalization of the class of projection depths from [43]. The original projection depths are obtained by considering a scale functional s invariant for reflections

$$s(-X) = s(X) \quad \text{for all } X \sim P \in \mathcal{P}\left(\mathbb{R}\right).$$

Just as the halfspace depth, also the projection depth is difficult to calculate exactly, especially in high dimensions [44, 27]. In the same way as for the halfspace depth in (2), define the approximated projection depth of $x \in \mathbb{R}^d$ with respect to $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ based on the independent randomly sampled directions $U_1, \ldots, U_n \in \mathbb{S}^{d-1}$ distributed uniformly on $\mathbb{S}^{d-1}$ by

$$PD_n(x; P) = PD_n(x) = \min_{i=1,\ldots,n} c\left(\frac{\langle U_i, x \rangle - m(\langle U_i, X \rangle)}{s(\langle U_i, X \rangle)} \right).$$

For a given set $C \subset \mathbb{R}^d$, we are interested in the uniform approximation of PD by PD_n

$$\Delta_n^P(C) = \sup_{x \in C} |PD_n(x) - PD(x)|.$$

Note that for the projection depth, the following function plays a role similar to that of the halfspace function from (3) for the halfspace depth

$$\varphi_x^P(u) = \frac{\langle u, x \rangle - m\left(\langle u, X \rangle\right)}{s\left(\langle u, X \rangle\right)}.$$

A distinctive role will be now played by the argument(s) of the maxima of the function $\varphi_x^P(\cdot)$ on $\mathbb{S}^{d-1}$. Any representative of this set will be denoted by $\widetilde{u}^P(x)$.

The next theorem provides simple conditions that guarantee the almost sure convergence of $\Delta_n^P(\mathbb{R}^d)$ to zero. In the proof of that result, the minimal multiplicative modulus of continuity of the function c given by

$$\zeta(\tau) = \sup_{t>0} \left(c(\tau\, t) - c(t)\right), \qquad \text{for } \tau \in (0,1), \tag{20}$$

plays a crucial role. Note that for a continuous function c the modulus ζ is a well-defined function continuous from the left at $\tau = 1$. To see this, pick a sequence $\tau_n \to 1$ from the left, and apply Dini's theorem [14, Theorem 2.4.10] to the sequence of continuous functions $\psi_n(t) = c(\tau_n t)$ defined on a compact set $[0, \infty]$ (with $\psi_n(\infty) = 0$) that converge monotonically on $t \in [0, \infty]$ to $\psi_0(t) = c(t)$ (again, $\psi_0(\infty) = 0$). Dini's theorem asserts that this convergence must be uniform on the domain of ψ_n, which can be rewritten to $\zeta(\tau_n) = \sup_{t \geq 0} \left(\psi_n(t) - \psi_0(t)\right) = \sup_{t \geq 0} \left(c(\tau_n\, t) - c(t)\right) \xrightarrow[n \to \infty]{} 0$.

Theorem 9. *Let $X \sim P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be such that the functions*

$$u \mapsto m_u = m(\langle u, X \rangle), \quad \text{and} \quad u \mapsto s_u = s(\langle u, X \rangle),$$

are Lipschitz continuous on $\mathbb{S}^{d-1}$ with Lipschitz constants C_m and C_s, respectively. Let $\sup_{u \in \mathbb{S}^{d-1}} |m_u| = \lambda_m < \infty$, $\inf_{u \in \mathbb{S}^{d-1}} s_u = \lambda_{s,1} > 0$, and $\sup_{u \in \mathbb{S}^{d-1}} s_u = \lambda_{s,2} < \infty$. Then $\Delta_n^P\left(\mathbb{R}^d\right) \xrightarrow[n \to \infty]{\text{a.s.}} 0$. Furthermore, an explicit upper bound on the rate of convergence of $\Delta_n^P(\mathbb{R}^d)$ may be devised in terms of the constants above, the minimal modulus of continuity of c given by

$$\delta_c(t) = \sup_{u,v>0,\, |u-v| \leq t} |c(u) - c(v)|,$$

and the minimal multiplicative modulus of continuity of c defined in (20).

Theorem 9 stands as an analogue of Theorem 1 stated for the halfspace depth. Further minor modifications to the proof of Theorem 9, stated in Appendix B.1, allow to formulate analogous results for continuous (thus uniformly continuous) location and scale parameters, in the spirit of Theorem 2. We omit those technical details for brevity.

For the important case of spherically symmetric distributions, the rate from Theorem 9 can be improved. The proof of this statement can be found in Appendix B.2.

Theorem 10. *Let $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ be a spherically symmetric distribution, and let ζ be a strictly increasing function whose inverse is denoted by ζ^{-1}. Then*

$$\limsup_{n \to \infty} \frac{n\, a_d \left(1 - \zeta^{-1}\left(\Delta_n^P(\mathbb{R}^d)\right)\right) - \log n}{\log \log n} \leq d \quad \text{a.s.}$$

Furthermore, for $c(x) = \frac{1}{1+(x)_+^k}$ with $k = 1, 2, \ldots$ it holds true that

$$\zeta(\tau) = \frac{1 - \tau^{k/2}}{1 + \tau^{k/2}} \qquad \text{for } \tau \in (0,1). \tag{21}$$

Because the projection depths are affine invariant, analogously as in Theorem 3 this result extends to elliptically symmetric distributions.

5.1. *Approximation of projection outlyingness*

Contrary to the projection depths (19), the outlyingness (18) is not uniformly approximated by its randomizations

$$O_n(x; P) = \max_{i=1,\dots,n} \frac{\langle U_i, x \rangle - m(\langle U_i, X \rangle)}{s(\langle U_i, X \rangle)},$$

even if the conditions of Theorems 9 and 10 are satisfied. We illustrate this in a simple example.

Example. Let $X \sim P \in \mathcal{P}\left(\mathbb{R}^2\right)$ have the uniform distribution on the unit circle. Because each projection $\langle u, X \rangle$ has the same distribution symmetric about the origin, $m_u = 0$ and $s_u = S$ for all $u \in \mathbb{S}^1$ and some constant $S \geq 0$. For any reasonable scale functional s the constant S is positive. Consider a point $x = (x_1, 0)^\mathsf{T} \in \mathbb{R}^2$ with $x_1 > 0$. Then

$$\varphi_x^P(u) = \frac{\langle u, x \rangle - m_u}{s_u} = \frac{u_1 x_1}{S} \quad \text{for all } u = (u_1, u_2)^\mathsf{T} \in \mathbb{S}^1.$$

This function is maximized at $\widetilde{u}^P(x) = (1, 0)^\mathsf{T} \in \mathbb{S}^{d-1}$ with $O(x; P) = x_1/S$. For any $u \in \mathbb{S}^1$ with $u_1 = \langle u, \widetilde{u}^P(x) \rangle < \cos(\varepsilon)$, on the other hand, $\varphi_x^P(\widetilde{u}^P(x)) - \varphi_x^P(u) > (1 - \cos(\varepsilon)) x_1/S$. Thus,

$$\lim_{n \to \infty} \sup_{x \in \mathbb{R}^2} |O(x; P) - O_n(x; P)|$$

$$\geq \lim_{n \to \infty} \sup_{x_1 > 0} \max_{i=1,\dots,n} \left(1 - \cos(\|\widetilde{u}^P(x) - U_i\|_g)\right) x_1 = \infty$$

almost surely, because for any degree of approximation of the target direction $\widetilde{u}^P(x)$ there exists a point $x = (x_1, 0)^\mathsf{T}$ with a large norm so that its outlyingness is approximated poorly.

6. Discussion

In the present paper we discussed the uniformity aspects of the depth approximation task. We demonstrated that for a distribution P that is regular enough, the approximated depth does converge uniformly to its exact counterpart. This result justifies the simple approximation procedure from a theoretical point of view, as the almost sure uniform convergence of the depth approximations carries over to the approximated depth-statistics such as the depth median (i.e. an argument of the maximum of the depth on $\mathbb{R}^d$), or the level sets of the depth.

In addition, we have presented and compared several almost sure upper bounds on the uniform discrepancies between the true halfspace and projection depth, and their approximated counterparts. Depending on the degree of regularity assumed about P, guidelines for the choice of the number of ap-

proximating directions n in order to achieve a desired precision can be devised from the theory presented. For the halfspace depth, in Table 1 we saw that the random approximation scheme is feasible in lower dimensions, especially for distributions whose densities are continuous and rather flat. In dimensions $d > 5$, hundreds of thousands of random directions may not be enough for sufficiently close approximations, and more elaborate algorithms appear to be needed.
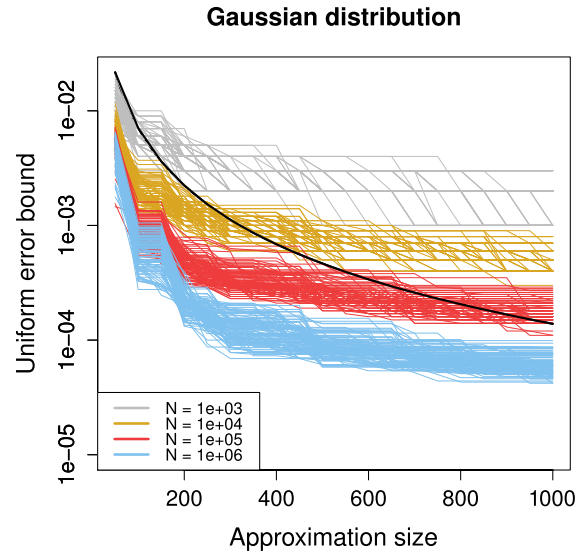

Gaussian distribution

FIG 5. *For a bivariate Gaussian distribution P, the tight bound from Theorem 7 on $\Delta_n(\mathbb{R}^2)$ (black solid line) is compared with* 100 *independent realisations of simulated trajectories of* $\max_{i=1,\ldots,100} \left| HD_n\left(x_i, \widehat{P}_N\right) - HD\left(x_i, \widehat{P}_N\right) \right|$ *as a function of the number of directions n, for $\{x_i\}_{i=1}^N$ a random sample from P. Considered sample sizes are $N = 10^3, 10^4, 10^5$, and 10^6.*

In Figure 4 we saw that for particular distributions, the theoretical bounds match the simulated results already for intermediate n closely. Nevertheless, in practice, one does not observe P directly, but rather only an empirical measure $\widehat{P}_N$ of a random sample $X_1, \ldots, X_N$ of size $N = 1, 2, \ldots$ generated from P, and computes the depth with respect to $\widehat{P}_N$ as a surrogate for $HD(x; P)$ or $PD(x; P)$. For practical considerations, it would therefore be desirable to obtain also upper bounds on the difference between the true depth and its approximations with respect to the empirical measure $\widehat{P}_N$. This can be done. For the halfspace depth, for instance, one may still use the theory presented, and devise a simple bound valid almost surely, for any fixed $\varepsilon > 0$ and all N and n

large enough, that takes the form

$$\sup_{x \in C} \left| HD_n(x; \widehat{P}_N) - HD(x; P) \right|$$

$$\leq \sup_{x \in C} \left| HD_n(x; \widehat{P}_N) - HD_n(x; P) \right| + \sup_{x \in C} \left| HD_n(x; P) - HD(x; P) \right|$$

$$\leq \sup_{x \in C} \max_{i=1,\ldots,n} \left| \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\left[\langle U_i, X_j \rangle \leq \langle U_i, x \rangle \right] - \varphi_x(U_i) \right| + \Delta_n(C) \qquad (22)$$

$$\leq \sup_{x \in \mathbb{R}^d} \sup_{u \in \mathbb{S}^{d-1}} \left| \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}\left[\langle u, X_j \rangle \leq \langle u, x \rangle \right] - \varphi_x(u) \right| + \Delta_n(C)$$

$$\leq \sqrt{\frac{\log \log N}{2N}} + \varepsilon + \Delta_n(C).$$

We used the fact that the sample halfspace depth process can be bounded by an empirical process given by the collection of closed halfspaces in $\mathbb{R}^d$. For the latter process, the bound follows from the law of the iterated logarithm devised in [23, Corollary 2.4], see also [2, 3, 41]. Alternatively, deviation inequalities for empirical processes such as those from [40, 2] or [38] may be also employed in (22). The third summand in the last expression of (22) can be handled using the results provided in this paper. Thus, the approximation of the empirical halfspace depth can be, at least for the number of observations N large enough, approached by the approximation results for the true sampling distribution P, and an upper bound on the deviation of P from $\widehat{P}_N$. An empirical argument supporting this finding is presented in Figure 5, where for $d = 2$ and the standard Gaussian distribution, the theoretical bound from Theorem 7 is compared with simulated trajectories of $\Delta_{n,N} = \max_{i=1,\ldots,100} \left| HD_n\left(x_i, \widehat{P}_N\right) - HD\left(x_i, \widehat{P}_N\right) \right|$ for several choices of $n \in [50, 1000]$ and four sample sizes $N = 10^3, 10^4, 10^5$ and 10^6. Note that for sample size N, the quantity $\Delta_{n,N}$ is a multiple of $1/N$, and therefore unless $\Delta_{n,N} = 0$ (which did not happen in our simulation study), the simulated trajectory never decreases below $1/N$ (i.e. the tick $1e{-}3$ for $N = 1\,000$ etc.). In addition, because the supremum $\Delta_n(\mathbb{R}^2)$ is compared with a maximum of only 100 points, also for N large the simulated trajectories do not follow the theoretical bound as closely as in Figure 4. Nevertheless, the obtained rates of convergence do appear to couple, even though the true distribution is replaced by an empirical one. Therefore, the general guidelines for the choice of n are relevant also in the situation when the sample depth is approximated. For instance, if the theoretical bound in Table 1 is already too high for the practical application in mind, the simple approximation of the halfspace depth is certainly not a good idea, and more sophisticated methodologies must be employed; for an empirical comparison of several of such approximation procedures we refer to [16]. In any case, for the sample depth it must be kept in mind that according to the first example of Section 4.5, for empirical distributions the depth approximations are inherently non-uniform.

Appendix A: Proofs of the theoretical results: Halfspace depth

A.1. Proof of Theorem 1

Denote by

$$S_n = \sup_{u \in \mathbb{S}^{d-1}} \min_{i=1,\ldots,n} \|u - U_i\|_g$$

the largest distance of a point on the sphere to the closest sample point $U_1, \ldots, U_n$. The quantity S_n closely relates to the so-called maximal spacing problem, extensively studied in the literature. By Janson [21, the main theorem and the second remark on page 276] and [20, Theorem 1.2] we know that in our setting it holds true that

$$\limsup_{n\to\infty} \frac{n\, a_d(S_n) - \log n}{\log\log n} = d \quad \text{a.s.},$$

where a_d is the function defined in (6). Take $\varepsilon > 0$ fixed, and for all $x \in C$ let $\widetilde{u}(x) \in \mathbb{S}^{d-1}$ be such that

$$\varphi_x(\widetilde{u}(x)) - \varepsilon < HD(x).$$

By $j(x)$ denote any index such that

$$\left\|\widetilde{u}(x) - U_{j(x)}\right\|_g = \min_{i=1,\ldots,n} \|\widetilde{u}(x) - U_i\|_g \le S_n.$$

For the halfspace depth (1) and its approximation (2) this implies that by (7)

$$
\begin{aligned}
0 \le \Delta_n(C) = \sup_{x \in C} & \left(HD_n(x) - HD(x) \right) \\
< \sup_{x \in C} & \left(\min_{i=1,\ldots,n} \varphi_x(U_i) - \varphi_x(\widetilde{u}(x)) \right) + \varepsilon \\
\le \sup_{x \in C} & \left(\varphi_x(U_{j(x)}) - \varphi_x(\widetilde{u}(x)) \right) + \varepsilon \\
\le L \sup_{x \in C} & \min_{i=1,\ldots,n} \|\widetilde{u}(x) - U_i\|_g + \varepsilon \\
\le L \sup_{u \in \mathbb{S}^{d-1}} & \min_{i=1,\ldots,n} \|u - U_i\|_g + \varepsilon \\
= L\, S_n + \varepsilon,
\end{aligned}
\tag{23}
$$

where the last inequality holds for any $\varepsilon > 0$. Therefore, $S_n \ge \Delta_n(C)/L$ and we may conclude that

$$\limsup_{n\to\infty} \frac{n\, a_d\left(\Delta_n(C)/L\right) - \log n}{\log\log n} \le d \quad \text{a.s.}$$

A.2. Proof of Theorem 2

The proof follows analogously to that of Theorem 1. The only difference is that in (23) we use a bound defined by means of the modulus of continuity δ, instead of the Lipschitz property (7) of the class (5). Since $\|U_{j(x)} - \widetilde{u}(x)\|_g \leq S_n$ and because of (9), we obtain

$$0 \leq \Delta_n(C) \leq \sup_{x \in C} \left(\varphi_x(U_{j(x)}) - \varphi_x(\widetilde{u}(x)) \right) + \varepsilon \leq \delta\left(S_n \right) + \varepsilon \qquad (24)$$

for any $\varepsilon > 0$, and the conclusion follows.

Remark. Note that the key step in the proofs of Theorems 1 and 2 is a conceptually simple application of a continuity argument to an asymptotic result on maximal spacings in $\mathbb{S}^{d-1}$. Analogously, using the equicontinuity of the halfspace functions (or their analogues for the projection depth) only, any other asymptotic expression for the maximal spacings could be translated to an appropriate formula for the depth approximations using very similar proof techniques. We have opted for an asymptotic result of Janson [20, 21] that takes the form of the law of the iterated logarithm for maximal spacings. This leads to upper bounds on the convergence rates that are quite strong in their formulation, yet may be rather conservative for some practical applications.

Remark. If the random sample of approximating directions $U_1, \ldots, U_n$ is taken from a non-uniform distribution on $\mathbb{S}^{d-1}$, results analogous to those described here can be obtained using the recent study of Aaron et al. [1], which complements the theory of Janson [21].

A.3. Proof of Theorem 3

For the approximated depth of the affinely transformed measure we can write

$$
\begin{aligned}
HD_n\left(Ax + b; P_{AX+b}\right) &= \min_{i=1,\ldots,n} \mathsf{P}\left(\langle U_i, AX + b \rangle \leq \langle U_i, Ax + b \rangle \right) \\
&= \min_{i=1,\ldots,n} \mathsf{P}\left(\langle A^{\mathsf{T}}U_i, X \rangle \leq \langle A^{\mathsf{T}}U_i, x \rangle \right) \\
&= \min_{i=1,\ldots,n} \varphi_x\left(\frac{A^{\mathsf{T}}U_i}{\|A^{\mathsf{T}}U_i\|} \right)
\end{aligned}
$$

where on the right hand side we see the approximated depth of x with respect to the untransformed random vector X, with the random directions U_i sampled from a possibly non-uniform distribution on $\mathbb{S}^{d-1}$. As in (23) and (24) it follows that for any $\varepsilon > 0$

$$
\begin{aligned}
0 \leq \widetilde{\Delta}_n(C) &= \sup_{x \in AC+b} |HD_n(x; P_{AX+b}) - HD(x; P_{AX+b})| \\
&= \sup_{x \in C} |HD_n(Ax + b; P_{AX+b}) - HD(Ax + b; P_{AX+b})| \\
&\leq \sup_{x \in C} \left| \min_{i=1,\ldots,n} \varphi_x\left(\frac{A^{\mathsf{T}}U_i}{\|A^{\mathsf{T}}U_i\|} \right) - \varphi_x\left(\widetilde{u}(x) \right) \right| + \varepsilon.
\end{aligned}
$$

Thus, we need to bound the spherical maximal spacing that corresponds to a random sample from $A^\mathsf{T} U / \left\| A^\mathsf{T} U \right\|$ for U uniform on $\mathbb{S}^{d-1}$. Let $u, v \in \mathbb{S}^{d-1}$. We bound the distance between the transforms of u and v

$$
\begin{aligned}
\left\| \frac{A^\mathsf{T} u}{\|A^\mathsf{T} u\|} - \frac{A^\mathsf{T} v}{\|A^\mathsf{T} v\|} \right\|_g &\leq \frac{\pi}{2} \left\| \frac{A^\mathsf{T} u}{\|A^\mathsf{T} u\|} - \frac{A^\mathsf{T} v}{\|A^\mathsf{T} v\|} \right\| \\
&\leq \frac{\pi}{2} \left(\left\| \frac{A^\mathsf{T} u}{\|A^\mathsf{T} u\|} - \frac{A^\mathsf{T} v}{\|A^\mathsf{T} u\|} \right\| + \left\| \frac{A^\mathsf{T} v}{\|A^\mathsf{T} u\|} - \frac{A^\mathsf{T} v}{\|A^\mathsf{T} v\|} \right\| \right) \\
&= \frac{\pi}{2} \left(\frac{1}{\|A^\mathsf{T} u\|} \left\| A^\mathsf{T} (u - v) \right\| + \left\| A^\mathsf{T} v \right\| \left| \frac{1}{\|A^\mathsf{T} u\|} - \frac{1}{\|A^\mathsf{T} v\|} \right| \right) \\
&\leq \frac{\pi}{2} \left(\frac{\lambda_1 \|u - v\|}{\lambda_d} + \frac{\lambda_1 \left\| A^\mathsf{T} u - A^\mathsf{T} v \right\|}{\lambda_d^2} \right) \\
&\leq \frac{\pi}{2} \left(\frac{\lambda_1 \|u - v\|}{\lambda_d} \left(1 + \frac{\lambda_1}{\lambda_d} \right) \right) \leq K \|u - v\|_g \qquad (25)
\end{aligned}
$$

where λ_1 and λ_d are the largest and the smallest singular value, respectively, of the matrix A^T and $K = \frac{\pi}{2} \frac{\lambda_1}{\lambda_d} \left(1 + \frac{\lambda_1}{\lambda_d} \right)$. The third inequality above is justified by [19, Theorem 3.1.2], formula $\inf_{u \in \mathbb{S}^{d-1}} \left\| A^\mathsf{T} u \right\| = \lambda_d$ that follows from the same theorem, and the mean value theorem applied to function $g(t) = 1/t$. The first and the last inequality in (25) hold since for $u, v \in \mathbb{S}^{d-1}$ we have $\|u - v\|_g \leq \frac{\pi}{2} \|u - v\| \leq \frac{\pi}{2} \|u - v\|_g$. We can now bound the spherical maximal spacing $\widetilde{S}_n$ of the non-uniformly distributed directions $A^\mathsf{T} U_i / \left\| A^\mathsf{T} U_i \right\|$ by

$$
\begin{aligned}
\widetilde{S}_n &= \sup_{u \in \mathbb{S}^{d-1}} \min_{i=1,\ldots,n} \left\| u - \frac{A^\mathsf{T} U_i}{\|A^\mathsf{T} U_i\|} \right\|_g = \sup_{v \in \mathbb{S}^{d-1}} \min_{i=1,\ldots,n} \left\| \frac{A^\mathsf{T} v}{\|A^\mathsf{T} v\|} - \frac{A^\mathsf{T} U_i}{\|A^\mathsf{T} U_i\|} \right\|_g \\
&\leq \sup_{v \in \mathbb{S}^{d-1}} \min_{i=1,\ldots,n} K \|v - U_i\|_g = K S_n.
\end{aligned}
$$

Now it follows that

$$
0 \leq \widetilde{\Delta}_n(C) \leq \delta \left(\widetilde{S}_n \right) + \varepsilon \leq \delta \left(K S_n \right) + \varepsilon
$$

for any $\varepsilon > 0$, and the proof can be concluded as that of Theorem 2.

A.4. Proof of Theorem 4

With no loss of generality, assume that $C \subset \mathbb{R}^d$ is compact with $P(C) = 1$. Define the function

$$
\varphi \colon C \times \mathbb{S}^{d-1} \to [0, 1] \colon (x, u)^\mathsf{T} \mapsto \varphi_x(u).
$$

Because P satisfies (13), φ is continuous on its domain [28, Proposition 4.5], which is a compact subset of $\mathbb{R}^{2d}$ [14, Theorem 2.2.8]. Therefore, φ must be

uniformly continuous on its domain [14, Corollary 2.4.6] and for δ_φ the modulus of continuity of φ defined by

$$\delta_\varphi \colon [0, \infty) \to [0, \infty) \colon t \mapsto \sup_{\|x-y\|+\|u-v\|_g \leq t} |\varphi(x, u) - \varphi(y, v)|$$

it holds true that for all $t \geq 0$

$$\sup_{x \in C} \sup_{\|u-v\|_g \leq t} |\varphi_x(u) - \varphi_x(v)| = \sup_{x \in C} \sup_{\|u-v\|_g \leq t} |\varphi(x, u) - \varphi(x, v)| \leq \delta_\varphi(t).$$

Thus, we can apply Theorem 2 with the modulus δ_φ, and the approximated depth approaches the theoretical one uniformly in x.

A.5. Proof of Theorem 5

For any $x \in C$, $u, v \in \mathbb{S}^{d-1}$ we can bound

$$|\varphi_x(u) - \varphi_x(v)| \leq 2 \int_{W(r,u,v)} M \, \mathrm{d}t,$$

where on the right hand side $W(r, u, v)$ is a wedge of a ball in $\mathbb{R}^d$ of radius $r = \mathrm{diam}(C)$ and center x (that therefore contains C), whose bounding hyperplanes have normals $u, v \in \mathbb{S}^{d-1}$. Since the volume of this wedge is

$$\|u - v\|_g \frac{\pi^{d/2-1} r^d}{2\Gamma(d/2 + 1)}$$

we obtain

$$\sup_{x \in C} |\varphi_x(u) - \varphi_x(v)| \leq \frac{M \pi^{d/2-1} r^d}{\Gamma(d/2 + 1)} \|u - v\|_g,$$

and Theorem 1 provides the result.

A.6. Proof of Theorem 6

Due to the translation invariance of HD, we can without loss of generality assume that $\mu = 0$. We thus assume that P admits a density

$$f(x) = g\left(\|x\|^2\right) \quad \text{for } x \in \mathbb{R}^d, \tag{26}$$

with $g \colon [0, \infty) \to [0, \infty)$. For such P, it is known that the true depth takes the form $HD(x) = h\left(\|x\|^2\right)$ for some non-increasing function $h \colon [0, \infty) \to [0, \infty)$ [26, Lemma 3.1]. All contours of the depth HD are therefore centred spheres. Consequently, for any $x \in \mathbb{R}^d$, $x \neq 0$, the single halfspace that realises the depth $HD(x)$ at x is the one whose inner normal is $x/\|x\|$. In other words, for each $x \in \mathbb{R}^d$

$$HD(x) = \begin{cases} \varphi_x(u) & \text{for any } u \in \mathbb{S}^{d-1} \text{ if } x = 0, \\ \varphi_x(-x/\|x\|) & \text{if } x \neq 0. \end{cases}$$

For $x = 0$, $\varphi_x(u) = 1/2$ for each $u \in \mathbb{S}^{d-1}$, and for any approximation we obtain the exact depth, i.e. $HD_n(x) - HD(x) = 0$ almost surely for all n. To assess the quality of the approximation at $x \neq 0$, it is necessary to control only the values $\varphi_x(u) - \varphi_x(\widetilde{u}(x))$ for u close to $\widetilde{u}(x) = -x/\|x\|$. This makes the problem easier than the general case described in Theorems 1 and 2 when no information about the position of $\widetilde{u}(x)$ is available.

Since $X \sim P$ is assumed to be spherically symmetric about 0, the random vector AX has the same distribution as X for any orthogonal matrix $A \in \mathbb{R}^{d \times d}$ [34]. Thus, any projection $\langle u, X \rangle$ for $u \in \mathbb{S}^{d-1}$ has the same univariate distribution. Denote the cumulative distribution function of this projected random vector by

$$F_p(t) = \mathsf{P}\left(\langle u, X \rangle \leq t\right).$$

Since P was assumed to admit a density, F_p is continuous. The halfspace functions φ_x can be expressed in terms of F_p

$$\varphi_x(u) = \mathsf{P}\left(\langle u, X \rangle \leq \langle u, x \rangle\right) = F_p\left(\langle u, x \rangle\right) \quad \text{for all } x \in \mathbb{R}^d \text{ and } u \in \mathbb{S}^{d-1}. \quad (27)$$

For the depth, we can write

$$HD(x) = F_p\left(\langle \widetilde{u}(x), x \rangle\right) = F_p\left(-\|x\|\right) \quad \text{for all } x \in \mathbb{R}^d. \quad (28)$$

Writing again $U_{j(x)}$ for the closest element of the random sample $U_1, \ldots, U_n$ from $\widetilde{u}(x)$, now we can bound

$$\begin{aligned}
\sup_{x \in \mathbb{R}^d} & |HD_n(x) - HD(x)| \\
&\leq \sup_{x \in \mathbb{R}^d} \left|\varphi_x(U_{j(x)}) - \varphi_x(\widetilde{u}(x))\right| \\
&= \sup_{x \in \mathbb{R}^d} \left|F_p\left(\langle U_{j(x)}, x \rangle\right) - F_p\left(-\|x\|\right)\right| \\
&= \sup_{x \in \mathbb{R}^d} \left|F_p\left(-\|x\|\left\langle U_{j(x)}, \widetilde{u}(x) \right\rangle\right) - F_p\left(-\|x\|\right)\right| \\
&= \sup_{x \in \mathbb{R}^d} \left|F_p\left(-\|x\|\cos\left(\left\|U_{j(x)} - \widetilde{u}(x)\right\|_g\right)\right) - F_p\left(-\|x\|\right)\right| \\
&\leq \sup_{t \geq 0} \left|F_p\left(-t\cos\left(S_n\right)\right) - F_p(-t)\right| \\
&= \sup_{t \geq 0} \left(F_p(t) - F_p(t\cos(S_n))\right), \quad (29)
\end{aligned}$$

where we used the fact that $\left\|U_{j(x)} - \widetilde{u}(x)\right\|_g \leq S_n$, further that F_p is non-decreasing, and that the distribution of $\langle u, X \rangle$ is symmetric about zero, i.e. $F_p(t) = 1 - F_p(-t)$. Take an arbitrary decreasing sequence $\{\varepsilon_\nu\}_{\nu=1}^\infty$ such that $\varepsilon_\nu \xrightarrow[\nu \to \infty]{} 0$, and define a sequence of functions

$$\left\{F_\nu(t) = \begin{cases} F_p\left(t\cos\left(\varepsilon_\nu\right)\right) & \text{for } t \in [0, \infty), \\ 1 & \text{for } t = \infty. \end{cases}\right\}_{\nu=1}^\infty. \quad (30)$$

This sequence consists of continuous functions defined on a compact set $[0, \infty]$. As F_p is continuous and non-decreasing,

$$
\begin{aligned}
F_\nu(t) &\xrightarrow[\nu \to \infty]{} F_p(t) \\
F_\nu(t) &\leq F_{\nu+1}(t)
\end{aligned}
\qquad \text{for all } t \in [0, \infty] \text{ and } \nu = 1, 2, \ldots
$$

Therefore, it is possible to use Dini's theorem [14, Theorem 2.4.10] to obtain that the sequence of functions (30) converges to F_p, uniformly in $[0, \infty]$. Thus, the limit of the right hand side of (29) is zero and finally, using a result on maximal spacings of Janson [21] we may conclude that the uniform convergence of the approximated depth holds true.

Let us now obtain the rate of convergence of the approximation if the density f is unimodal. First, we show that for $d > 1$ and $t \geq 0$ the distribution function $F_p(t)$ must be concave. To see this, note that for the density of the random variable $\langle u, X \rangle$ with $u = (0, \ldots, 0, 1) \in \mathbb{S}^{d-1}$ we have

$$
f_p(t) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g\left(\|(x_1, \ldots, x_{d-1}, t)\|^2 \right) \, d x_1 \ldots d x_{d-1} \quad \text{for } t \in \mathbb{R}.
$$

For f unimodal, function g in (26) must be non-increasing. For $0 \leq t_1 \leq t_2$ this means that $g\left(\|(x_1, \ldots, x_{d-1}, t_1)\| \right) \geq g\left(\|(x_1, \ldots, x_{d-1}, t_2)\| \right)$ for any $x_1, \ldots, x_{d-1} \in \mathbb{R}$, and that

$$
f_p(t_1) \geq f_p(t_2) \quad \text{for any } 0 \leq t_1 \leq t_2.
$$

Since $f_p(t)$ is non-increasing for $t \geq 0$, the distribution function $F_p(t)$ must be concave for $t \geq 0$. Thus, using the obvious fact that $F_p(0) = 1/2$, for $S_n \leq \pi/2$ and $t \geq 0$ we have

$$
\begin{aligned}
0 &\leq F_p(t) - F_p(\cos(S_n)t) \leq F_p(t) - (1 - \cos(S_n))F_p(0) - \cos(S_n)F_p(t) \\
&= (1 - \cos(S_n))\, F_p(t) - (1 - \cos(S_n))/2 \leq (1 - \cos(S_n))/2.
\end{aligned}
\tag{31}
$$

Combining the last formula with (29) gives

$$
\sup_{x \in \mathbb{R}^d} |HD(x) - HD_n(x)| \leq (1 - \cos(S_n))/2.
\tag{32}
$$

This inequality holds true under the condition that for any $u \in \mathbb{S}^{d-1}$ there exists a random direction U_i such that $\|u - U_i\|_g \leq S_n$. In other words, the inequality remains valid if the polar angle that corresponds to the maximal spacing given by the random sample of directions $U_1, \ldots, U_n$ does not exceed S_n. This leads to an inequality of the same type as (24) from Theorem 2, and the conclusion follows by the same argument as that used in the proof of Theorem 2. The simpler bound follows by application of the inequality $1 - \cos(t) \leq t^2/2$ to (32). This way one obtains

$$
\sup_{x \in \mathbb{R}^d} |HD(x) - HD_n(x)| \leq S_n^2/4,
$$

for S_n the maximal spacing from the proofs of Theorems 1 and 2. Again, the same technique as that from the proof of Theorem 2 gives the final rate of convergence.

A.7. Proof of Theorem 7

Again, without loss of generality $\mu = 0$. It is enough to show the first inequality — the second assertion then follows from Theorem 6. Consider the main theorem of Janson [21] once more. By the first part of that result, for S_n the spacings defined in the proof of Theorem 1 we have

$$\liminf_{n \to \infty} \frac{n \, a_d(S_n) - \log n}{\log \log n} = d - 2 \quad \text{a.s.,} \tag{33}$$

which means that almost surely for n large there exists a vector $v_n \in \mathbb{S}^{d-1}$ such that each sampled direction U_i, $i = 1, \ldots, n$, is far enough from v_n, i.e. for any $\varepsilon > 0$ almost surely for all n large enough there is v_n such that

$$\min_{i=1,\ldots,n} \|v_n - U_i\|_g \geq a_d^{-1} \left(\frac{(d - 2 - \varepsilon) \log \log n + \log n}{n} \right) - \varepsilon. \tag{34}$$

Denote the right hand side of the previous inequality by r_n.

By the definition of δ from (14) for any $\eta > 0$ there exists $K_n > 0$ such that

$$F_p(K_n) - F_p(K_n \cos(r_n)) = F_p(-K_n \cos(r_n)) - F_p(-K_n) \geq \delta(r_n) - \eta.$$

Consider the point $x_n = -K_n v_n$. Then $\|x_n\| = K_n$, and v_n is the minimizer of φ_{x_n}, i.e. $v_n = \widetilde{u}(x_n)$. The latter follows from the proof of Theorem 6, where it was argued that for a spherically symmetric P, $\widetilde{u}(x_n)$ minimizes φ_{x_n} for all x_n on the ray $\lambda \widetilde{u}(x_n)$ with $\lambda < 0$. Let $u \in \mathbb{S}^{d-1}$ be any direction with $\|u - \widetilde{u}(x_n)\|_g \geq r_n$. By formulas (27) and (28) from the proof of Theorem 6 and the monotonicity of F_p we have

$$\varphi_{x_n}(u) - \varphi_{x_n}(\widetilde{u}(x_n)) = F_p\left(-\|x_n\| \cos\left(\|u - \widetilde{u}(x_n)\|_g \right) \right) - F_p(-\|x_n\|)$$
$$\geq F_p(-\|x_n\| \cos(r_n)) - F_p(-\|x_n\|) \geq \delta(r_n) - \eta.$$

We have that if there exists $v_n \in \mathbb{S}^{d-1}$ that satisfies (34), then for any $\eta > 0$ it is possible to find x_n such that for random directions U_i the above inequality holds true with u replaced by U_i for all $i = 1, \ldots, n$. In other words, we have that in this situation

$$\Delta_n(\mathbb{R}^d) \geq HD_n(x_n) - HD(x_n) = \min_{i=1,\ldots,n} \left(\varphi_{x_n}(U_i) - HD(x_n) \right) \geq \delta(r_n) - \eta. \tag{35}$$

By (33) and (34), however, we know that almost surely for all n large such v_n exists. Therefore, it is possible to invert (35), which allows us to write that almost surely for any $\varepsilon > 0$ and $\eta > 0$

$$\liminf_{n \to \infty} \frac{n \, a_d \left(\delta^{-1} \left(\Delta_n(\mathbb{R}^d) + \eta \right) + \varepsilon \right) - \log n}{\log \log n} \geq d - 2 - \varepsilon,$$

and the conclusion follows.

A.8. Proof of Theorem 8

The proof is led in a spirit similar to that of Theorem 6. Define

$$r \colon \mathbb{R}^d \setminus \{0\} \times \mathbb{S}^{d-1} \to \mathbb{R} \colon \left(x^\mathsf{T}, u^\mathsf{T}\right)^\mathsf{T} \mapsto \frac{\langle x, u \rangle}{\|u\|_p} \bigg/ \frac{\langle x, \widetilde{u}(x) \rangle}{\|\widetilde{u}(x)\|_p} \ .$$

For $p \leq 1$ the minimizing direction $\widetilde{u}(x) \in \mathbb{S}^{d-1}$ is not uniquely defined. Nevertheless, from (17) we see that for any choice of the minimal direction

$$\frac{\langle x, \widetilde{u}(x) \rangle}{\|\widetilde{u}(x)\|_p} = -\|x\|_q, \tag{36}$$

and the function r is well defined with

$$r(x, u) = -\frac{\langle x, u \rangle}{\|x\|_q \|u\|_p}.$$

The essential part of this proof is to show that the function r converges to 1 uniformly in $x \in \mathbb{R}^d \setminus \{0\}$ when $u \in \mathbb{S}^{d-1}$ comes from a small neighbourhood of $\widetilde{u}(x)$. More precisely, we show that for any $\varepsilon > 0$ there exists $t > 0$ such that

$$\sup_{x \in \mathbb{R}^d \setminus \{0\}} \sup_{\substack{u \in \mathbb{S}^{d-1} \\ \|u - \widetilde{u}(x)\|_2 \leq t}} |r(x, u) - 1| \leq \varepsilon. \tag{37}$$

Note that for the special case $p = 2$ it is possible to proceed as in the proof of Theorem 6 and write (see formula (29))

$$|r(x, u) - 1| = \left| \frac{\langle x, u \rangle}{\langle x, \widetilde{u}(x) \rangle} - 1 \right| = |\langle u, \widetilde{u}(x) \rangle - 1| = \left| 1 - \cos\left(\|u - \widetilde{u}(x)\|_g \right) \right|.$$

For general $p \in (0, 2]$ the derivation below is somewhat similar, yet more involved.

From formula (37) it follows that for any $x \in \mathbb{R}^d \setminus \{0\}$ and $u \in \mathbb{S}^{d-1}$ with $\|u - \widetilde{u}(x)\|_2 \leq t$ we can write

$$\frac{\langle x, u \rangle}{\|u\|_p} \in \left[-(1 + \varepsilon) \|x\|_q, -(1 - \varepsilon) \|x\|_q \right],$$

and

$$\varphi_x(u) = F\left(\frac{\langle x, u \rangle}{\|u\|_p} \right) \in \left[F\left(-(1 + \varepsilon) \|x\|_q \right), F\left(-(1 - \varepsilon) \|x\|_q \right) \right].$$

By the unimodality assumption, the distribution function F must be concave on $[0, \infty)$. Thus, using derivation analogous to (31) we can write

$$0 \leq \varphi_x(u) - \varphi_x\left(\widetilde{u}(x)\right) = \varphi_x(u) - HD(x) = F\left(\frac{\langle x, u \rangle}{\|u\|_p} \right) - F\left(-\|x\|_q \right)$$

$$\leq F\left(-(1 - \varepsilon) \|x\|_q \right) - F\left(-\|x\|_q \right) = F\left(\|x\|_q \right) - F\left((1 - \varepsilon) \|x\|_q \right) \leq \frac{\varepsilon}{2}.$$

The last inequality holds for all $x \neq 0$ and $u \in \mathbb{S}^{d-1}$ with

$$\|u - \widetilde{u}(x)\|_2 = 2\sin\left(\|u - \widetilde{u}(x)\|_g / 2\right) \leq t. \tag{38}$$

Therefore, using the same argumentation as in the proof of Theorem 2 we can conclude that the uniform convergence of the approximations holds true as claimed.

It remains to show (37). First, observe that for any $\alpha > 0$ it holds that $r(x, u) = r(\alpha x, u)$. Thus, in (37) we may focus only on $x \in \mathbb{S}^{d-1}$ in the first supremum. Assume that $u \in \mathbb{S}^{d-1}$ is such that $\|u - \widetilde{u}(x)\|_2 \leq t$. Let us rewrite r as a product of two terms

$$r(x, u) = \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} \frac{\langle x, u \rangle}{\langle x, \widetilde{u}(x) \rangle} \tag{39}$$

and inspect them separately. For the first factor we have to distinguish the cases $p \in (0, 1)$ and $p \in [1, 2]$.

The following relations hold true for $0 < p < q \leq \infty$ and $x \in \mathbb{R}^d$ [see, e.g., 7, Lemma A.1]

$$\|x\|_q \leq \|x\|_p \leq d^{1/p - 1/q} \|x\|_q. \tag{40}$$

With $q = 2$ and since $\|u\|_2 = \|\widetilde{u}(x)\|_2 = 1$ it follows that

$$\frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} \leq \frac{d^{1/p - 1/2} \|\widetilde{u}(x)\|_2}{\|u\|_2} = d^{1/p - 1/2}.$$

For $p \geq 1$ the function $\|\cdot\|_p$ is a norm. Thus, it obeys the (reverse) triangle inequality, and for the first factor in (39) we can bound for $p \in [1, 2]$

$$\left| \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} - 1 \right| = \frac{1}{\|u\|_p} \left| \|\widetilde{u}(x)\|_p - \|u\|_p \right| \leq \frac{\|u - \widetilde{u}(x)\|_p}{\|u\|_p} \leq d^{1/p - 1/2} \|u - \widetilde{u}(x)\|_2. \tag{41}$$

Now, we discuss the case $p \in (0, 1)$. For $p \in (0, 1)$ it is known that the following variant of the reverse triangle inequality holds for all $x, y \in \mathbb{R}^d$ [33, Exercise 24, Chapter 3]

$$\left| \|x\|_p^p - \|y\|_p^p \right| \leq \|x - y\|_p^p.$$

Using this result, for $p \in (0, 1)$ we can proceed analogously to (41) and write

$$\left| \frac{\|\widetilde{u}(x)\|_p^p}{\|u\|_p^p} - 1 \right| \leq \frac{\|u - \widetilde{u}(x)\|_p^p}{\|u\|_p^p} \leq d^{1 - p/2} \|u - \widetilde{u}(x)\|_2^p, \tag{42}$$

as by (40) we have that

$$\|u\|_p^p \geq \|u\|_2^p = 1 \tag{43}$$

that holds true for any $p \in (0, 2]$. By (40) again and the previous inequality we know that

$$\frac{\|\widetilde{u}(x)\|_p^p}{\|u\|_p^p} \leq d^{1 - p/2}. \tag{44}$$

For $0 \leq x \leq d^{1-p/2}$ and $p \in (0,1)$ we have by the mean value theorem that

$$\left| x^{1/p} - 1 \right| \leq \frac{d^{(1-p/2)(1/p-1)}}{p} \left| x - 1 \right|.$$

It means that by (42) and (44) we can write

$$\left| \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} - 1 \right| \leq \frac{d^{(1-p/2)(1/p-1)}}{p} \left| \frac{\|\widetilde{u}(x)\|_p^p}{\|u\|_p^p} - 1 \right| \leq \frac{d^{1/p-1/2}}{p} \|u - \widetilde{u}(x)\|_2^p, \quad (45)$$

which gives the final inequality for the first term of (39) also for $p \in (0,1)$.

The second term in (39) can be treated at once for all $p \in (0,2]$. One can write

$$\frac{\langle x, u \rangle}{\langle x, \widetilde{u}(x) \rangle} = 1 + \frac{\langle x, u - \widetilde{u}(x) \rangle}{\langle x, \widetilde{u}(x) \rangle}$$

and bound the second summand using (36), (40), and (43) by

$$\left| \frac{\langle x, u - \widetilde{u}(x) \rangle}{\langle x, \widetilde{u}(x) \rangle} \right| \leq \frac{\|x\|_2 \|u - \widetilde{u}(x)\|_2}{|\langle x, \widetilde{u}(x) \rangle|} \leq \frac{\|x\|_2 \|u - \widetilde{u}(x)\|_2}{\|x\|_q \|\widetilde{u}(x)\|_p} \leq d^{1/2-1/q} \|u - \widetilde{u}(x)\|_2.$$
(46)

Joining (41) and (46) together, for $p \in [1,2]$ we obtain for any $x \in \mathbb{R}^d$

$$\begin{aligned}
|r(x,u) - 1| &= \left| \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} \left(\frac{\langle x, u \rangle}{\langle x, \widetilde{u}(x) \rangle} - 1 + 1 \right) - 1 \right| \\
&\leq \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} \left| \frac{\langle x, u \rangle}{\langle x, \widetilde{u}(x) \rangle} - 1 \right| + \left| \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} - 1 \right| \\
&\leq d^{1/p-1/2} \left(d^{1/2-1/q} + 1 \right) \|u - \widetilde{u}(x)\|_2 \\
&\leq 2d^{1/p-1/2} \left(d^{1/2-1/q} + 1 \right) \sin \left(\|u - \widetilde{u}(x)\|_g /2 \right)
\end{aligned}$$
(47)

and the desired (37) is verified. For $p \in (0,1)$ we combine (45) and (46) and get

$$\begin{aligned}
|r(x,u) - 1| &\leq \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} \left| \frac{\langle x, u \rangle}{\langle x, \widetilde{u}(x) \rangle} - 1 \right| + \left| \frac{\|\widetilde{u}(x)\|_p}{\|u\|_p} - 1 \right| \\
&\leq d^{1/p-1/q} \|u - \widetilde{u}(x)\|_2 + \frac{d^{1/p-1/2}}{p} \|u - \widetilde{u}(x)\|_2^p \\
&\leq 2d^{1/p-1/q} \sin \left(\|u - \widetilde{u}(x)\|_g /2 \right) + 2^p \frac{d^{1/p-1/2}}{p} \sin^p \left(\|u - \widetilde{u}(x)\|_g /2 \right).
\end{aligned}$$
(48)

The statement about the rates of convergence follows by application of the proof technique from Theorem 2, bounds (47) and (48), and the relation (38).

Appendix B: Proofs of the theoretical results: Projection depth

B.1. Proof of Theorem 9

Take $M > \lambda_m$ and denote $B_M = \{x \in \mathbb{R}^d \colon \|x\| < M\}$. Consider $x \in B_M$ first. For such x we can write

$$
\begin{aligned}
&\left|\varphi_x^P(u) - \varphi_x^P(v)\right| \\
&\leq \left|\frac{\langle u, x\rangle - m_u}{s_u} - \frac{\langle u, x\rangle - m_u}{s_v}\right| + \left|\frac{\langle u, x\rangle - m_u}{s_v} - \frac{\langle v, x\rangle - m_v}{s_v}\right| \\
&\leq (|\langle u, x\rangle| + |m_u|)\left|\frac{1}{s_u} - \frac{1}{s_v}\right| + \frac{1}{s_v}\left(|\langle u - v, x\rangle| + |m_v - m_u|\right) \\
&\leq (\|x\| + |m_u|)\frac{1}{\lambda_{s,1}^2}|s_u - s_v| + \frac{1}{\lambda_{s,1}}\left(\|u - v\|\,\|x\| + |m_v - m_u|\right) \\
&\leq (\|x\| + \lambda_m)\frac{1}{\lambda_{s,1}^2}C_s\|u - v\| + \frac{1}{\lambda_{s,1}}\left(\|u - v\|\,\|x\| + C_m\|u - v\|\right) \\
&= \|u - v\|\left(\|x\|\left(\frac{C_s}{\lambda_{s,1}^2} + \frac{1}{\lambda_{s,1}}\right) + \frac{\lambda_m C_s}{\lambda_{s,1}^2} + \frac{C_m}{\lambda_{s,1}}\right) \\
&= \|u - v\|\left(\|x\| A + B\right) \leq \|u - v\|\left(MA + B\right),
\end{aligned}
$$

where the constants $A, B > 0$ are given by the last equality. From this it follows that

$$
\Delta_n^P(B_M) \leq \delta_c\left(\|u - v\|\left(MA + B\right)\right) = \delta_c\left(2\sin\left(\|u - v\|_g /2\right)(MA + B)\right)
$$
$$
\leq \delta_c\left(\|u - v\|_g (MA + B)\right), \tag{49}
$$

where δ_c is the minimal modulus of continuity of c.

Now suppose that $x \notin B_M$. In what follows we bound, with $\widetilde{u}^P(x)$ abbreviated to $\widetilde{u}$ for simplicity, the expression

$$
\left|\frac{\varphi_x^P(u)}{\varphi_x^P(\widetilde{u})} - 1\right| \leq \left|\frac{s_{\widetilde{u}}}{s_u}\right|\left|\frac{\langle x, u\rangle - m_u}{\langle x, \widetilde{u}\rangle - m_{\widetilde{u}}} - 1\right| + \left|\frac{s_{\widetilde{u}}}{s_u} - 1\right| \tag{50}
$$

in terms of $\|\widetilde{u} - u\|_g$. First of all, note that $|s_{\widetilde{u}}/s_u| \leq \lambda_{s,2}/\lambda_{s,1}$. Secondly, by the definition of the maximizer $\widetilde{u} = \widetilde{u}^P(x)$, we have that

$$
\varphi_x^P(\widetilde{u}) = \frac{\langle x, \widetilde{u}\rangle - m_{\widetilde{u}}}{s_{\widetilde{u}}} \geq \frac{\langle x, x/\|x\|\rangle - m_{x/\|x\|}}{s_{x/\|x\|}} \geq \frac{\|x\| - \lambda_m}{\lambda_{s,2}} \geq \frac{M - \lambda_m}{\lambda_{s,2}} \geq 0,
$$

and thus

$$
|\langle x, \widetilde{u}\rangle - m_{\widetilde{u}}| = \langle x, \widetilde{u}\rangle - m_{\widetilde{u}} \geq \frac{\lambda_{s,1}}{\lambda_{s,2}}\left(\|x\| - \lambda_m\right) \geq \frac{\lambda_{s,1}}{\lambda_{s,2}}\left(M - \lambda_m\right).
$$

We can therefore write

$$\left| \frac{\langle x, u \rangle - m_u}{\langle x, \widetilde{u} \rangle - m_{\widetilde{u}}} - 1 \right| = \left| \frac{\langle x, u - \widetilde{u} \rangle + m_{\widetilde{u}} - m_u}{\langle x, \widetilde{u} \rangle - m_{\widetilde{u}}} \right| \leq \frac{C_m \left\| \widetilde{u} - u \right\|_g + \left\| x \right\| \left\| \widetilde{u} - u \right\|}{\left| \langle x, \widetilde{u} \rangle - m_{\widetilde{u}} \right|}$$

$$\leq \frac{\lambda_{s,2}}{\lambda_{s,1}} \left(\frac{C_m \left\| \widetilde{u} - u \right\|_g}{M - \lambda_m} + \frac{\left\| x \right\| \left\| \widetilde{u} - u \right\|}{\left\| x \right\| - \lambda_m} \right)$$

$$= \frac{\lambda_{s,2}}{\lambda_{s,1}} \left(\frac{C_m \left\| \widetilde{u} - u \right\|_g}{M - \lambda_m} + \left\| \widetilde{u} - u \right\| + \frac{\lambda_m \left\| \widetilde{u} - u \right\|}{\left\| x \right\| - \lambda_m} \right)$$

$$\leq \frac{\lambda_{s,2}}{\lambda_{s,1}} \left(\frac{C_m \left\| \widetilde{u} - u \right\|_g}{M - \lambda_m} + \left\| \widetilde{u} - u \right\| + \frac{\lambda_m \left\| \widetilde{u} - u \right\|}{M - \lambda_m} \right).$$

The second summand in (50) can be bounded by

$$\left| \frac{s_{\widetilde{u}}}{s_u} - 1 \right| = \left| \frac{s_{\widetilde{u}} - s_u}{s_u} \right| \leq \frac{C_s \left\| \widetilde{u} - u \right\|_g}{\lambda_{s,1}}.$$

Altogether, we can write

$$\left| \frac{\varphi_x^P(u)}{\varphi_x^P(\widetilde{u})} - 1 \right| \leq \left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{C_m \left\| \widetilde{u} - u \right\|_g + M \left\| \widetilde{u} - u \right\|}{M - \lambda_m} + \frac{C_s \left\| \widetilde{u} - u \right\|_g}{\lambda_{s,1}}.$$

This means that for $u \in \mathbb{S}^{d-1}$ such that $\left\| \widetilde{u} - u \right\|_g \leq \varepsilon$, i.e. $\left\| \widetilde{u} - u \right\| \leq 2 \sin(\varepsilon/2) \leq \varepsilon$, we have

$$\left| \varphi_x^P(u) - \varphi_x^P(\widetilde{u}) \right| \leq \varphi_x^P(\widetilde{u}) \left(\left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{\varepsilon \, C_m + 2M \sin(\varepsilon/2)}{M - \lambda_m} + \frac{\varepsilon \, C_s}{\lambda_{s,1}} \right)$$

$$\leq \varepsilon \, \varphi_x^P(\widetilde{u}) \left(\left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{C_m + M}{M - \lambda_m} + \frac{C_s}{\lambda_{s,1}} \right)$$

or, because $\varphi_x^P(\widetilde{u}) \geq \varphi_x^P(u)$ for all $u \in \mathbb{S}^{d-1}$ by the definition of $\widetilde{u}$,

$$\varphi_x^P(u) \geq \varphi_x^P(\widetilde{u}) \left(1 - \left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{\varepsilon \, C_m + 2M \sin(\varepsilon/2)}{M - \lambda_m} - \frac{\varepsilon \, C_s}{\lambda_{s,1}} \right)$$

$$\geq \varphi_x^P(\widetilde{u}) \left(1 - \varepsilon \left(\left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{C_m + M}{M - \lambda_m} + \frac{C_s}{\lambda_{s,1}} \right) \right).$$

For any $n \in 1, 2, \ldots$ such that $\min_{i=1,\ldots,n} \left\| U_i - \widetilde{u} \right\|_g \leq \varepsilon$ for all $x \notin B_M$ the last bounds result in

$$PD(x; P) = c \left(\varphi_x^P(\widetilde{u}) \right),$$

$$PD_n(x; P) \leq c(\varphi_x^P(u))$$

$$\leq c \left(\varphi_x^P(\widetilde{u}) \left(1 - \left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{\varepsilon \, C_m + 2M \sin(\varepsilon/2)}{M - \lambda_m} - \frac{\varepsilon \, C_s}{\lambda_{s,1}} \right) \right)$$

$$\leq c \left(\varphi_x^P(\widetilde{u}) \left(1 - \varepsilon \left(\left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{C_m + M}{M - \lambda_m} + \frac{C_s}{\lambda_{s,1}} \right) \right) \right).$$

This holds true for any $x \notin B_M$. Consequently, if $\min_{i=1,\ldots,n} \|U_i - \widetilde{u}\|_g \leq \varepsilon$, we can write

$$\Delta_n^P \left(\mathbb{R}^d \setminus B_M \right) \leq c \left(\varphi_x^P(\widetilde{u})\tau \right) - c(\varphi_x^P(\widetilde{u})) \leq \zeta(\tau_1) \leq \zeta(\tau_2) \tag{51}$$

for

$$\tau_1 = 1 - \left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{\varepsilon\, C_m + 2M \sin(\varepsilon/2)}{M - \lambda_m} - \frac{\varepsilon\, C_s}{\lambda_{s,1}},$$

$$\tau_2 = 1 - \varepsilon \left(\left(\frac{\lambda_{s,2}}{\lambda_{s,1}} \right)^2 \frac{C_m + M}{M - \lambda_m} + \frac{C_s}{\lambda_{s,1}} \right).$$

Finally, the rate of convergence of the approximations of the projection depth can be obtained as a combination of (49) and (51)

$$\begin{aligned} \Delta_n^P(\mathbb{R}^d) &\leq \inf_{M > \lambda_m} \max \left\{ \delta_c \left(2\sin(\varepsilon/2) \left(MA + B \right) \right), \zeta(\tau_1) \right\} \\ &\leq \inf_{M > \lambda_m} \max \left\{ \delta_c \left(\varepsilon \left(MA + B \right) \right), \zeta(\tau_2) \right\}, \end{aligned} \tag{52}$$

where ε is substituted by the maximal spacing S_n introduced in the proof of Theorem 1. The final rates of convergence then follow by application of the same strategy as in the proof of Theorem 2, where the result of Janson [21] is employed to devise an almost sure upper bound on the rate of convergence of the depth approximation. Note that to obtain explicit, almost sure bounds such as those from Theorems 1 and 2 it is necessary to invert (an upper bound to) one of the functions on the right hand side of (52) considered as a function of ε.

B.2. Proof of Theorem 10

Due to the translation invariance of PD, the distribution P can be assumed to be spherically symmetric about the origin. Then, it can be seen that $m_u = 0$ and $s_u = S$ for some constant $S \geq 0$. The latter fact follows because for such spherically symmetric distributions, all projections of P onto lines passing through the origin have the same univariate distribution [34]. If $S = 0$, the assertion of the theorem is trivially satisfied, as in this case $PD(x; P) = PD_n(x; P) = 0$ for all $x \in \mathbb{R}^d$. For $S > 0$ for any $x \in \mathbb{R}^d$

$$PD(x; P) = c \left(\sup_{u \in \mathbb{S}^{d-1}} \frac{\langle x, u \rangle}{S} \right) = c \left(\|x\| / S \right),$$

and the depth is realised in $\widetilde{u}^P(x) = x / \|x\|$. Likewise, for function φ_x^P from the approximating projection depth it holds true that

$$\varphi_x^P(u) = \frac{\langle x, u \rangle}{S}.$$

Similarly as in the proof of Theorem 9 let us bound the expression bounding (50). Here we can for $x \neq 0$ write

$$\left| \frac{\varphi_x^P(u)}{\varphi_x^P(\widetilde{u}^P(x))} - 1 \right| = \left| \frac{\langle x, u \rangle}{\langle x, \widetilde{u}^P(x) \rangle} - 1 \right| = \frac{\left| \langle x, u - \widetilde{u}^P(x) \rangle \right|}{\|x\|}$$
$$\leq \left\| u - \widetilde{u}^P(x) \right\| \leq \left\| u - \widetilde{u}^P(x) \right\|_g.$$

Thus, in the same way as in (51) in the proof of Theorem 9 we obtain that under the maximal spacing condition $\min_{i=1,\dots,n} \|U_i - \widetilde{u}\|_g \leq \varepsilon$ we can write

$$\Delta_n^P\left(\mathbb{R}^d\right) \leq c\left(\varphi_x^P(\widetilde{u}^P(x))\tau\right) - c(\varphi_x^P(\widetilde{u}^P(x))) \leq \zeta(\tau_1) \leq \zeta(\tau_2),$$

where

$$\tau_1 = 1 - 2\sin(\varepsilon/2),$$
$$\tau_2 = 1 - \varepsilon.$$

For the final rates we apply the same technique as in the proof of Theorem 2; note that in the statement of the theorem only the second, simpler rate is stated.

The formula for ζ in (21) can be obtained by direct maximization in the expression for the multiplicative modulus of continuity.

References

[1] Aaron, C., Cholaquidis, A., and Fraiman, R. (2017). A generalization of the maximal-spacings in several dimensions and a convexity test. *Extremes*, 20(3):605–634. MR3679984

[2] Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.*, 12(4):1041–1067. MR0757769

[3] Bingham, N. H. (1986). Variants on the law of the iterated logarithm. *Bull. London Math. Soc.*, 18(5):433–467. MR0847984

[4] Bogićević, M. and Merkle, M. (2018). Approximate calculation of Tukey's depth and median with high-dimensional data. *Yugosl. J. Oper. Res.*, 28(4):475–499. MR3884451

[5] Burr, M. A. and Fabrizio, R. J. (2017). Uniform convergence rates for halfspace depth. *Statist. Probab. Lett.*, 124:33–40. MR3608209

[6] Chen, D., Morin, P., and Wagner, U. (2013). Absolute approximation of Tukey depth: theory and experiments. *Comput. Geom.*, 46(5):566–573. MR3015942

[7] Chen, Z. and Tyler, D. E. (2004). On the behavior of Tukey's depth and median under symmetric stable distributions. *J. Statist. Plann. Inference*, 122(1-2):111–124. MR2057917

[8] Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.*, 52(11):4979–4988. MR2526207

[9] DeVore, R. A. and Lorentz, G. G. (1993). *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin. MR1261635

[10] Devroye, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Ann. Probab.*, 9(5):860–867. MR0628878

[11] Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.*, 12(3):793–815. MR0751274

[12] Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Qualifying paper, Harvard University.

[13] Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20(4):1803–1827. MR1193313

[14] Dudley, R. M. (2002). *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Revised reprint of the 1989 original. MR1932358

[15] Dyckerhoff, R. (2004). Data depths satisfying the projection property. *Allg. Stat. Arch.*, 88(2):163–190. MR2074729

[16] Dyckerhoff, R., Mozharovskyi, P., and Nagy, S. (2020). Approximate computation of projection depths. arXiv:2007.08016.

[17] Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Math. Methods Oper. Res.*, 77(3):423–432. MR3072795

[18] Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London. MR1071174

[19] Horn, R. A. and Johnson, C. R. (1994). *Topics in matrix analysis*. Cambridge University Press, Cambridge. Corrected reprint of the 1991 original. MR1288752

[20] Janson, S. (1986). Random coverings in several dimensions. *Acta Math.*, 156(1-2):83–118. MR0822331

[21] Janson, S. (1987). Maximal spacings in several dimensions. *Ann. Probab.*, 15(1):274–280. MR0877603

[22] Johnson, D. S. and Preparata, F. P. (1978). The densest hemisphere problem. *Theoret. Comput. Sci.*, 6(1):93–107. MR0464674

[23] Kuelbs, J. and Dudley, R. M. (1980). Log log laws for empirical measures. *Ann. Probab.*, 8(3):405–418. MR0573282

[24] Lin, P. E. (1972). Some characterizations of the multivariate t distribution. *J. Multivariate Anal.*, 2:339–344. MR0331633

[25] Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.*, 27(3):783–858. MR1724033

[26] Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, 88(421):252–260. MR1212489

[27] Liu, X. and Zuo, Y. (2014). Computing projection depth and its associated estimators. *Stat. Comput.*, 24(1):51–63. MR3147697

[28] Massé, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 10(3):397–419. MR2061438

[29] Massé, J.-C. and Theodorescu, R. (1994). Halfplane trimming for bivariate

distributions. *J. Multivariate Anal.*, 48(2):188–202. MR1261798

[30] Mosler, K. (2002). *Multivariate dispersion, central regions and depth: The lift zonoid approach*, volume 165 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin. MR1913862

[31] Mozharovskyi, P., Mosler, K., and Lange, T. (2015). Classifying real-world data with the $DD\alpha$-procedure. *Adv. Data Anal. Classif.*, 9(3):287–314. MR3392037

[32] Rousseeuw, P. J. and Ruts, I. (1999). The depth function of a population distribution. *Metrika*, 49(3):213–244. MR1731769

[33] Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition. MR0924157

[34] Serfling, R. (2006). Multivariate symmetry and asymmetry. *Encyclopedia of Statistical Sciences, Second Edition*, 8:5338–5345.

[35] Shao, W. and Zuo, Y. (2012). Simulated annealing for higher dimensional projection depth. *Comput. Statist. Data Anal.*, 56(12):4026–4036. MR2957851

[36] Shao, W. and Zuo, Y. (2020). Computing the halfspace depth with multiple try algorithm and simulated annealing algorithm. *Comput. Stat.*, 35:203–226.

[37] Stahel, W. A. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen. Ph.D. thesis, ETH Zürich.

[38] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1):28–76. MR1258865

[39] Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, BC, 1974), Vol. 2*, pages 523–531. Canad. Math. Congress, Montreal, Que. MR0426989

[40] Vapnik, V. N. and Chervonenkis, A. Y. (2015). On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of complexity*, pages 11–30. Springer, Cham. Reprint of Theor. Probability Appl. 16 (1971), 264–280. MR3408730

[41] Yukich, J. E. (1990). The law of the iterated logarithm for empirical processes. In *Probability in Banach spaces 6 (Sandbjerg, 1986)*, volume 20 of *Progr. Probab.*, pages 265–282. Birkhäuser Boston, Boston, MA. MR1056715

[42] Zolotarev, V. M. (1986). *One-dimensional stable distributions*, volume 65 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver. MR0854867

[43] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.*, 31(5):1460–1490. MR2012822

[44] Zuo, Y. and Lai, S. (2011). Exact computation of bivariate projection depth and the Stahel-Donoho estimator. *Comput. Statist. Data Anal.*, 55(3):1173–1179. MR2741406

[45] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, 28(2):461–482. MR1790005