



# Theory of angular depth for classification of directional data

Stanislav Nagy<sup>1</sup> · Houyem Demni<sup>2</sup> · Davide Buttarazzi<sup>2</sup> · Giovanni C. Porzio<sup>2</sup>

Received: 30 November 2022 / Revised: 17 June 2023 / Accepted: 8 August 2023 /  
Published online: 23 September 2023  
© Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

Depth functions offer an array of tools that enable the introduction of quantile- and ranking-like approaches to multivariate and non-Euclidean datasets. We investigate the potential of using depths in the problem of nonparametric supervised classification of directional data, that is classification of data that naturally live on the unit sphere of a Euclidean space. In this paper, we address the problem mainly from a theoretical side, with the final goal of offering guidelines on which angular depth function should be adopted in classifying directional data. A set of desirable properties of an angular depth is put forward. With respect to these properties, we compare and contrast the most widely used angular depth functions. Simulated and real data are eventually exploited to showcase the main implications of the discussed theoretical results, with an emphasis on potentials and limits of the often disregarded angular halfspace depth.

**Keywords** Angular depth · Directional data · Hyperspheres · Statistical depth · Supervised classification

**Mathematics Subject Classification** 62H11 · 62H30 · 62G05

---

✉ Stanislav Nagy  
nagy@karlin.mff.cuni.cz

Houyem Demni  
houyem.demni@unicas.it

Davide Buttarazzi  
davide.buttarazzi@unicas.it

Giovanni C. Porzio  
porzio@unicas.it

<sup>1</sup> Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

<sup>2</sup> Department of Economics and Law, University of Cassino and Southern Lazio, Cassino, Italy

## 1 Introduction: classification of directional data

In many research areas, data living in nonlinear subdomains of  $\mathbb{R}^d$  are common. That data can be described by random variables whose support is a nonlinear manifold. The complicated nature of such datasets calls for methods going beyond the standard statistical techniques, and the need for specific tools applicable to nonlinear datasets arises.

Our work considers *directional* (or *spherical*) data whose support is the surface of the unit hyper-sphere  $\mathbb{S}^{d-1}$  of  $\mathbb{R}^d$ , or a subset of it. Directional data describe values typically recorded as a set of angles (for  $d = 2$ , also known as *circular* data) or as unit vectors in  $\mathbb{R}^d$ . A large amount of literature exists on the analysis of this kind of data, from the seminal book of Mardia (1972) to many recent works on clustering (Salah and Nadif 2019; Bry and Cucala 2022), and on nonparametric data analysis (Arnone et al. 2022; Saavedra-Nieves and Crujeiras 2022), to cite a few.

We focus on the use of angular depth functions in supervised classification. That is, on which angular depth should be adopted when classifying newly observed directional values into groups, given a preliminary knowledge of directional data whose group labels are available. Also known as supervised learning of directional objects, this issue has become quite an active area of research in recent years (see, e.g., Figueiredo (2009); Tsagris and Alenazi (2019); Jana and Dey (2021)). We pursue a nonparametric approach, and within the many, we consider angular depth functions and the associated DD-classifiers (depth-depth classifiers). Statistical depth functions (Liu et al. 1999; Zuo and Serfling 2000a) present a set of tools applicable to multivariate and non-Euclidean data that generalize notions of ranks, orderings, and quantiles to more complex datasets. Exploiting the nonparametric nature of statistical depth functions and their robustness properties, the depths have already been successfully applied also in the context of directional data analysis (Liu and Singh 1992; Agostinelli and Romanazzi 2013; Ley et al. 2014; Demni et al. 2021).

DD-classifiers were first introduced for multivariate data living in  $\mathbb{R}^d$  (Li et al. 2012). Because of their simplicity, efficacy, and optimality properties, they have been soon extended to data defined over different domains, such as functional data (Cuesta-Albertos et al. 2017; Mosler and Mozharovskiy 2017) and networks (Tian and Gel 2019).

Depth functions evaluate how much “inner” a point is with respect to a probability distribution. DD-classifiers assign to any point in the training set its depth value with respect to each of the competing distributions/groups. Next, these values are mapped into a Euclidean space spanned by the depths themselves (DD-space), and a discriminating function within such a DD-space is constructed. In the second step, new data points are classified by evaluating their depth with respect to each labeled training set, and assigned to a class by means of the discriminating function previously constructed in the DD-space. The depths thus serve as a tool offering elaborate nonlinear dimension reduction methods applicable to classification.

In principle, any depth function can be used to construct a DD-classifier. However, its performance is related to the depth which is adopted. A crucial issue is therefore the choice of a depth among the many available. In  $\mathbb{R}^d$ , related problems have been addressed from a different perspective, even recently (Mosler and Mozharovskiy

2022). For directional data and angular depths, though, apart from a few remarks in particular simulation studies, no discussion on that pressing issue is found in the literature.

The aim of this work is twofold. First, we propose a set of desired conditions that angular depths should satisfy. The focus will be on the properties exploitable in a classification setting. Second, we compare the available angular depths in view of those conditions, and we offer a series of theoretical results regarding the general behavior and the shapes of the depth contours for the most established angular depths.

The paper is organized as follows. After introducing notations (Sect. 1.1), Sect. 2 lists a set of desirable properties for an angular depth function. Several observations regarding these conditions are given; in particular, the conditions are contrasted to those available for depths in linear spaces (Liu 1990; Zuo and Serfling 2000a; Serfling 2006) and the special traits of directional data are stressed out. Section 3 introduces several prominent angular depth functions and discusses which of them enjoy the properties laid out in Sect. 2. Our theoretical analysis pinpoints the often disregarded angular halfspace depth (Small 1987) as a candidate suitable to be used for classification purposes. Section 4 uses simulated and real datasets to outline the main implications of the obtained theoretical results, focusing on the power of the angular halfspace depth and its limitations within a supervised classification setting. Final remarks are offered in Sect. 5. Technical details and proofs of our theoretical results are postponed to the Appendix.

## 1.1 Notations

We write  $\mathbb{R}^d$  for the Euclidean space of points  $x \in \mathbb{R}^d$  equipped with the Euclidean norm  $\|x\|$  and a scalar product  $\langle \cdot, \cdot \rangle$ . A vector  $x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$  is meant to be a column vector; when not necessary, the transposition in  $(x_1, \dots, x_d)^\top$  will be dropped to ease the notations. The unit sphere in  $\mathbb{R}^d$  is  $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ ; elements of  $\mathbb{S}^{d-1}$  are frequently called directions. For a topological space  $\mathcal{S}$ ,  $\mathcal{P}(\mathcal{S})$  stands for the collection of all Borel probability measures on  $\mathcal{S}$ .

## 2 Desirable properties of angular depth functions

We begin by discussing the main properties that an angular depth function should enjoy in a classification framework. A depth  $D$  on a topological space  $\mathcal{S}$  is, formally speaking, a bounded mapping<sup>1</sup>

$$D : \mathcal{S} \times \mathcal{P}(\mathcal{S}) \rightarrow [0, \infty) \quad (1)$$

that to a point  $x \in \mathcal{S}$  and a probability measure  $P \in \mathcal{P}(\mathcal{S})$  assigns  $D(x; P)$ , the depth of  $x$  with respect to (w.r.t.)  $P$ . The depth is intended to quantify how much “inner”, or

<sup>1</sup> Some depth functions, e.g., the angular Mahalanobis depth treated in Sect. 3.1 below, are not defined for all  $P \in \mathcal{P}(\mathcal{S})$ , but rather only for  $x \in \mathcal{S}$  and  $P$  in a subset  $\mathcal{D} \subseteq \mathcal{P}(\mathcal{S})$  of probability measures. We slightly abuse the notation and write (1) for simplicity.

“centrally located” the point  $x$  is w.r.t. the distribution  $P$ . The set of points maximizing the depth function  $x \mapsto D(x; P)$  is called the set of *depth medians*. As one moves away from the median set, the depth w.r.t.  $P$  is supposed to decrease until it reaches values close to zero when  $x$  is taken far from the main bulk of the mass of  $P$ . Any depth  $D$  is naturally characterized by the collection of all its *central regions*

$$D_\alpha(P) = \{x \in \mathcal{S} : D(x; P) \geq \alpha\} \quad \text{with } \alpha \geq 0, \text{ for any } P \in \mathcal{P}(\mathcal{S}). \quad (2)$$

The shape of the regions  $D_\alpha(P)$  should reflect the geometry of the distribution  $P$ . We are naturally concerned by describing the shapes and the structural properties of these central regions, since they play a crucial role in depth-based classification.

## 2.1 Depth functions in linear spaces $\mathbb{R}^d$

The methodology of statistical depth functions in linear spaces  $\mathbb{R}^d$  has been studied thoroughly by Zuo and Serfling (2000a). According to Zuo and Serfling (2000a), a (*statistical*) *depth*  $D$  in the linear space  $\mathcal{S} = \mathbb{R}^d$  is a function (1) that fulfills the following properties for all  $P \in \mathcal{P}(\mathbb{R}^d)$ :

- (L<sub>1</sub>) *Affine invariance*:  $D(x; P) = D(Ax + b; P_{AX+b})$  for all  $x \in \mathbb{R}^d$ ,  $A \in \mathbb{R}^{d \times d}$  non-singular, and  $b \in \mathbb{R}^d$ , where  $P_{AX+b} \in \mathcal{P}(\mathbb{R}^d)$  stands for the distribution of the transformed random vector  $AX + b$  with  $X \sim P$ ;
- (L<sub>2</sub>) *Maximality at center*: If  $\mu \in \mathbb{R}^d$  is a “center” of  $P$  then

$$D(\mu; P) = \sup_{x \in \mathbb{R}^d} D(x; P); \quad (3)$$

- (L<sub>3</sub>) *Monotonicity along rays*:  $D(x; P) \leq D(\mu + \alpha(x - \mu); P)$  holds for all  $x \in \mathbb{R}^d$  and  $\alpha \in [0, 1]$ , where  $\mu \in \mathbb{R}^d$  is any point that satisfies (3);
- (L<sub>4</sub>) *Vanishing at infinity*:  $\lim_{\|x\| \rightarrow \infty} D(x; P) = 0$ .

The notion of a “center”  $\mu$  in condition (L<sub>2</sub>) is ambiguous. It is meant to be applied to a specified notion of symmetry of  $P$ , with  $\mu$  being the unique center of that symmetry. Often,  $P$  is considered to be symmetric around  $\mu$  if each closed halfspace that contains  $\mu$  carries  $P$ -mass at least 1/2, which corresponds to the so-called halfspace symmetry of  $P$ , the arguably weakest notion of symmetry considered in  $\mathcal{P}(\mathbb{R}^d)$ . Indeed, as shown in Zuo and Serfling (2000b), many other familiar notions of multivariate symmetry (spherical, elliptical, central, or angular) imply halfspace symmetry of a measure in  $\mathbb{R}^d$ .

In addition to conditions (L<sub>1</sub>)–(L<sub>4</sub>), several other desiderata of depths have been proposed. For classification, interesting are those concerning the shapes of the central regions (2). In Serfling (2006), the following conditions can be found.

- (L<sub>5</sub>) *Upper semi-continuity*:  $D(\cdot; P) : \mathbb{R}^d \rightarrow [0, \infty) : x \mapsto D(x; P)$  is upper semi-continuous;
- (L<sub>6</sub>) *Quasi-concavity*: The central regions (2) are convex for all  $\alpha \geq 0$ .

For the central regions  $D_\alpha(P)$ , condition (L<sub>5</sub>) guarantees that each set  $D_\alpha(P)$  is closed in  $\mathbb{R}^d$ , which together with the boundedness condition from (L<sub>4</sub>) implies that the central regions (2) must be compact sets for all  $\alpha > 0$ . The additional condition (L<sub>6</sub>) refines (L<sub>3</sub>). In (L<sub>3</sub>), we require that each straight line segment between a maximizer  $\mu \in \mathbb{R}^d$  of  $D(\cdot; P)$  and  $x \in \mathbb{R}^d$  lies inside the central region  $D_\alpha(P)$  with  $\alpha = D(x; P)$ . That property is called the *star convexity* of the central regions around  $\mu$ . The stronger convexity condition (L<sub>6</sub>) refines this and requires that for any two points  $x, y \in D_\alpha(P)$  the whole line segment between  $x$  and  $y$  is contained in  $D_\alpha(P)$ , for any  $\alpha \geq 0$ .

A prototypical depth function in  $\mathbb{R}^d$  that satisfies all conditions (L<sub>1</sub>)–(L<sub>6</sub>) is the Tukey’s halfspace depth (Tukey 1975; Donoho and Gasko 1992), defined for  $x \in \mathbb{R}^d$  and  $P \in \mathcal{P}(\mathbb{R}^d)$  by

$$hD(x; P) = \inf_{u \in \mathbb{S}^{d-1}} P(H_{x,u}), \tag{4}$$

where  $H_{x,u} = \{y \in \mathbb{R}^d : \langle y, u \rangle \geq \langle x, u \rangle\}$  is a closed halfspace whose boundary hyperplane passes through  $x$  with inner normal  $u$ . An angular version of  $hD$  will be presented and studied in Sect. 3.4 below.

### 2.2 Angular depth functions in $\mathbb{S}^{d-1}$

In the setup of directional data, we design depths on the unit sphere  $\mathbb{S}^{d-1}$  w.r.t. a distribution  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ . An *angular* (or *directional*) (*statistical*) *depth function*  $aD$  should thus naturally be a bounded function

$$aD : \mathbb{S}^{d-1} \times \mathcal{P}(\mathbb{S}^{d-1}) \rightarrow [0, \infty) \tag{5}$$

that fulfills the following properties for all  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ :

- (D<sub>1</sub>) *Rotational invariance:*  $aD(x; P) = aD(Ox; P_{OX})$  for all  $x \in \mathbb{S}^{d-1}$  and any orthogonal matrix  $O \in \mathbb{R}^{d \times d}$ , where  $P_{OX} \in \mathcal{P}(\mathbb{S}^{d-1})$  stands for the distribution of the transformed random vector  $OX$  with  $X \sim P$ ;

- (D<sub>2</sub>) *Maximality at center:*

$$aD(\mu; P) = \sup_{x \in \mathbb{S}^{d-1}} aD(x; P), \tag{6}$$

for any  $P$  with “center” at  $\mu \in \mathbb{S}^{d-1}$ ;

- (D<sub>3</sub>) *Monotonicity along great circles:*

$$aD(x; P) \leq aD((\mu + \alpha(x - \mu)) / \|\mu + \alpha(x - \mu)\|; P) \tag{7}$$

holds for all  $x \in \mathbb{S}^{d-1} \setminus \{-\mu\}$  and  $\alpha \in [0, 1]$ , where  $\mu \in \mathbb{S}^{d-1}$  is any point that satisfies (6);

- (D<sub>4</sub>) *Minimality at the anti-median:*  $aD(-\mu; P) = \inf_{x \in \mathbb{S}^{d-1}} aD(x; P)$ , for any  $\mu \in \mathbb{S}^{d-1}$  that satisfies (6).

Conditions (D<sub>1</sub>)–(D<sub>4</sub>) are direct translations of the classical requirements (L<sub>1</sub>)–(L<sub>4</sub>) from  $\mathbb{R}^d$  to  $\mathbb{S}^{d-1}$ . Indeed, since no sensible notion of an affine transform exists in  $\mathbb{S}^{d-1}$ , a natural family replacing the set of affine maps in (L<sub>1</sub>) is the set of orthogonal transforms in the ambient space  $\mathbb{R}^d$ , that leave the unit sphere intact. Condition (D<sub>3</sub>) interprets straight lines in  $\mathbb{R}^d$  as geodesics, which are the shortest curves joining any couple of points. The geodesics in  $\mathbb{S}^{d-1}$  are the great circles passing through pairs of directions; naturally, the expression  $(\mu + \alpha(x - \mu))/\|\mu + \alpha(x - \mu)\|$  on the right hand side of (7) encodes the points of the shortest arc between  $\mu$  and  $x \neq -\mu$  parameterized by  $\alpha \in [0, 1]$ . The case  $x = -\mu$  is excluded in (D<sub>3</sub>), because there are infinitely many great circles joining  $\mu$  and  $-\mu$ . The latter situation is therefore specifically treated in (D<sub>4</sub>), which can be seen as an analog of the vanishing property (L<sub>4</sub>) from  $\mathbb{R}^d$ . Indeed, interpreting the meridians of great circles joining  $\mu$  with  $-\mu$  as “straight lines centered at  $\mu$ ” in  $\mathbb{S}^{d-1}$ , the antipodal point  $-\mu$  plays the role of all “points at infinity” in  $\mathbb{S}^{d-1}$ .

The vague notion of a “center” from (D<sub>2</sub>) is the most difficult to translate to  $\mathbb{S}^{d-1}$ . A simple solution is to consider only rotationally symmetric distributions<sup>2</sup> in  $\mathbb{S}^{d-1}$  (Ley and Verdebout 2017), and to declare (D<sub>2</sub>) to be desired only for  $\mu$  being (one of the two antipodally symmetric) directions of rotational symmetry of  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ . There are two issues to be addressed when applying (D<sub>2</sub>) with this definition of a center:

- The center of rotational symmetry is never unique. For example, whenever  $O$  fixes  $\mu$ , it fixes also its antipodal direction  $-\mu \in \mathbb{S}^{d-1}$ .
- The uniform distribution on  $\mathbb{S}^{d-1}$  is rotationally symmetric around each direction.

The first problem may be resolved by defining the “center” in (D<sub>2</sub>) to be exactly one of the points  $\{\mu, -\mu\}$ . Its antipodal point is then used in (D<sub>4</sub>). The second problem with the uniformly distributed random variable  $U \sim P_U \in \mathcal{P}(\mathbb{S}^{d-1})$  on  $\mathbb{S}^{d-1}$  is of a more fundamental nature. The rotational invariance (D<sub>1</sub>) immediately gives that for any  $aD$  satisfying (D<sub>1</sub>),  $aD(\cdot; P_U)$  must be a constant on  $\mathbb{S}^{d-1}$ , and thus no ordering of points is possible. This is, however, understandable since the symmetry of  $P_U$  makes it impossible to distinguish points of  $\mathbb{S}^{d-1}$  w.r.t. their location; see also our Theorem 1 below.

We now consider analogues of conditions (L<sub>5</sub>) and (L<sub>6</sub>) in  $\mathbb{S}^{d-1}$ . The upper semi-continuity from (L<sub>5</sub>) is defined in any topological space, and in particular, it is simple to state also in  $\mathbb{S}^{d-1}$ . We say that a function  $g : \mathbb{S}^{d-1} \rightarrow \mathbb{R}$  is upper semi-continuous if

$$\limsup_{x_n \rightarrow x} g(x_n) \leq g(x) \quad \text{for all } x \in \mathbb{S}^{d-1}$$

for any sequence  $\{x_n\}_{n=1}^\infty \subset \mathbb{S}^{d-1}$  that converges to  $x \in \mathbb{S}^{d-1}$ . A set  $A \subseteq \mathbb{S}^{d-1}$  is called *spherical convex* (see, e.g., Besau and Werner (2016)) if its radial extension defined as

$$\text{rad}(A) = \left\{ \lambda a \in \mathbb{R}^d : a \in A \text{ and } \lambda \geq 0 \right\} \subseteq \mathbb{R}^d \tag{8}$$

<sup>2</sup> Recall that a distribution  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  is called *rotationally symmetric* around a direction  $\mu \in \mathbb{S}^{d-1}$  if  $X \sim P$  has the same distribution as  $OX$  for any orthogonal matrix  $O \in \mathbb{R}^{d \times d}$  that fixes  $\mu$ , that is  $O\mu = \mu$ .

is convex in  $\mathbb{R}^d$ . Observe that if a spherical convex set  $A$  is not contained in a closed hemisphere in  $\mathbb{S}^{d-1}$ , necessarily  $A = \mathbb{S}^{d-1}$ . This particular property of spherical convexity will be important later in our discussion.

We are ready to state the directional versions of conditions (L<sub>5</sub>) and (L<sub>6</sub>):

- (D<sub>5</sub>) *Upper semi-continuity*:  $aD(\cdot; P) : \mathbb{S}^{d-1} \rightarrow [0, \infty) : x \mapsto aD(x; P)$  is upper semi-continuous;
- (D<sub>6</sub>) *Quasi-concavity*: The central regions (2) are spherical convex for all  $\alpha \geq 0$ .

As in the linear case, condition (D<sub>5</sub>) guarantees that the central regions (2) for depth  $aD$  are closed in  $\mathbb{S}^{d-1}$ . Since  $\mathbb{S}^{d-1}$  is a bounded space, it guarantees that a *directional median*  $\mu \in \mathbb{S}^{d-1}$  of  $P$  induced by  $aD$  in (6) always exists, since the supremum must be attained. Condition (D<sub>6</sub>) is stronger than (D<sub>3</sub>) in the same sense as in  $\mathbb{R}^d$ . While (D<sub>3</sub>) can be seen as a spherical version of star convexity of the central regions of  $aD(\cdot; P)$ , condition (D<sub>6</sub>) imposes spherical convexity of the central regions.

We already argued that for the very special uniform distribution on  $\mathbb{S}^{d-1}$ , any sensible angular depth  $aD$  must take a constant value over  $\mathbb{S}^{d-1}$ . Our first observation is that, as a consequence of our conditions, a much larger set of directional distributions must possess angular depth that is constant over  $\mathbb{S}^{d-1}$ . The proof of the following theorem, as well as the proofs of all our other theoretical results, are deferred to Appendix B.1.

**Theorem 1** Suppose that  $aD$  is an angular depth (5) that satisfies (D<sub>1</sub>), (D<sub>4</sub>), and (D<sub>5</sub>). Let  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  be centrally symmetric, meaning that  $X \sim P$  has the same distribution as  $-X$ . Then  $aD(\cdot; P)$  must be constant over  $\mathbb{S}^{d-1}$ .

Observe that the condition of central symmetry in Theorem 1 is a weak one. As proved by Rousseeuw and Struyf (2004, Theorem 2),  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  is centrally symmetric as in Theorem 1 if and only if it is halfspace symmetric around the origin  $0 \in \mathbb{R}^d$ . Equivalently,  $P$  is centrally symmetric if and only if each closed hemisphere of  $\mathbb{S}^{d-1}$  is of  $P$ -mass at least 1/2. For an additional discussion on these symmetry considerations see Zuo and Serfling (2000b) and Nagy et al. (2019, Sect. 4.2). Another direct consequence of our conditions follows for rotationally symmetric distributions. Its proof is in Appendix B.2.

**Theorem 2** Let  $aD$  be an angular depth (5) that satisfies (D<sub>1</sub>), and let  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  be rotationally symmetric around  $\mu \in \mathbb{S}^{d-1}$ . Then  $aD(\cdot; P)$  is a function of the inner product  $\langle \cdot, \mu \rangle$  only, i.e. there exists a function  $\xi : [-1, 1] \rightarrow [0, \infty)$  such that  $aD(x; P) = \xi(\langle x, \mu \rangle)$  for all  $x \in \mathbb{S}^{d-1}$ . If, in addition, also (D<sub>3</sub>) is true, then  $\xi$  can be taken monotone on  $[-1, 1]$ ; under the additional assumption (D<sub>5</sub>), the function  $\xi$  is upper semi-continuous on  $[-1, 1]$ .

Theorem 2 gives that for  $P$  rotationally symmetric, also the angular depth contours of any  $aD$  that verifies (D<sub>1</sub>) must be invariant w.r.t. rotations fixing the center of symmetry  $\mu$  of  $P$ . If also (D<sub>3</sub>) is true, all central regions (2) must be spherical caps orthogonal to  $\mu$ .<sup>3</sup> Note, however, that Theorem 2 does not claim that only rotationally symmetric

<sup>3</sup> A spherical cap  $S$  is defined as an intersection of  $\mathbb{S}^{d-1}$  with a (closed) halfspace  $H$  in  $\mathbb{R}^d$ . The cap  $S$  is said to be orthogonal to the direction  $u \in \mathbb{S}^{d-1}$  of the unit normal of  $H$ .

measures can have rotationally symmetric angular depth contours. An example is the angular Mahalanobis depth that will be discussed in Sect. 3.1.

At this point, it is also important to realize that even in the situation when  $P$  is rotationally symmetric (and thus all central regions  $aD_\alpha(P)$  are spherical caps by Theorem 2), the spherical convexity  $(D_6)$  does not have to be satisfied for  $aD$ . Remarkably, spherical caps of the form  $\{x \in \mathbb{S}^{d-1} : \langle x, \mu \rangle \geq \beta\}$  are spherical convex sets only for  $\beta \geq 0$ ; the only spherical convex set larger than a closed hemisphere is  $\mathbb{S}^{d-1}$  itself. We formalize our observation in the following result proved in Appendix B.3.

**Theorem 3** Suppose that  $aD$  is an angular depth (5) that satisfies  $(D_6)$ . Then for any  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  there must exist a direction  $u \in \mathbb{S}^{d-1}$  such that all points in the open hemisphere  $S = \{x \in \mathbb{S}^{d-1} : \langle x, u \rangle < 0\}$  attain the same depth, i.e. there exist  $u \in \mathbb{S}^{d-1}$  and  $c \geq 0$  such that  $aD(x; P) = c$  for all  $x \in \mathbb{S}^{d-1}$  such that  $\langle x, u \rangle < 0$ . If, in addition, the depth  $aD$  satisfies  $(D_4)$ , then  $c = \inf_{x \in \mathbb{S}^{d-1}} aD(x; P)$ .

We see that contrary to the linear case where the quasi-concavity assumption  $(L_6)$  is rather standard, for directional data the requirement  $(D_6)$  of convexity of central regions is questionable. As we will see in Sect. 3, the only common angular depth that satisfies condition  $(D_6)$  in full is the angular halfspace depth.

While full-blown quasi-concavity of angular depth functions does not appear to be completely desirable, it is certainly beneficial if the contours of the angular depth reflect the geometric properties of the distribution  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ . In the linear space  $\mathbb{R}^d$ , the affine invariance  $(L_1)$  tacitly imposes this requirement. Its directional analogue  $(D_1)$ , however, does not allow for transformations other than orthogonal, which would in  $\mathbb{R}^d$  correspond to mere rotation invariance of a depth — a condition that is known to be too weak for many practical applications. Several procedures have therefore been proposed to make statistical depths in  $\mathbb{R}^d$  satisfy the stronger affine invariance  $(L_1)$ ; we refer to Serfling (2010). The desire for an angular depth to reflect the shape properties and the geometry of the distribution  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  brings about our final requirement on the contours of an angular depth function. It can be phrased as a requirement of non-rigidity of the central regions  $(2)$  of  $aD$ .

$(D_7)$  *Non-rigidity of central regions:* There exists a measure  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  such that for some  $\alpha > 0$  the central region  $aD_\alpha(P)$  from (2) is not a spherical cap.

### 3 Angular depth functions and their properties

We now revise several angular depth functions known from the literature<sup>4</sup> (i) the Mahalanobis depth (Ley et al. 2014); (ii) the angular distance-based depths (Pandolfo et al. 2018) including the arc distance depth, the cosine distance depth, and the chord distance depth; (iii) the angular simplicial depth (Liu and Singh 1992); and (iv) the angular halfspace depth (Small 1987). Our main interest is in the study of the desired properties  $(D_1)$ – $(D_7)$  from Sect. 2. A summary of our main results is in Table 1.

<sup>4</sup> The very recently introduced *spatial depth* for directional data from Konen (2022) is not discussed here; for a comprehensive treatment of that depth we refer to the PhD thesis Konen (2022).



**Table 1** A table of results summarizing which of the theoretical properties (D<sub>1</sub>)–(D<sub>7</sub>) are satisfied (indicated by ✓) or not (indicated by ×) for the considered angular depths

Sec.	Angular Depth		Properties						
			(D <sub>1</sub> )	(D <sub>2</sub> )	(D <sub>3</sub> )	(D <sub>4</sub> )	(D <sub>5</sub> )	(D <sub>6</sub> )	(D <sub>7</sub> )
3.1	Angular Mahalanobis <sup>1</sup>	$aMD$	✓	✓	✓	✓	✓	×	×
3.2.1	Cosine distance	$aD_{cos}$	✓	✓	✓	✓	✓	×	×
3.2.2	Arc distance	$aD_{arc}$	✓	×	×	✓	✓	×	✓
3.2.3	Chord	$aD_{chord}$	✓	×	×	×	✓	×	✓
3.3	Angular simplicial	$asD$	✓	×	×	×	✓	×	✓
3.4	Angular halfspace	$ahD$	✓	✓	✓	✓	✓	✓	✓

[1] Defined only if the Fréchet median of  $X \sim P$  is unique

### 3.1 Angular Mahalanobis depth

Arguably one of the simplest depths for directional data is the *angular Mahalanobis depth* from Ley et al. (2014). It depends crucially on the notion of a median direction, which is typically taken to be the Fréchet median (Fisher 1985) defined as (any) direction  $\mu \in \mathbb{S}^{d-1}$  that minimizes the objective function

$$\varphi : \mathbb{S}^{d-1} \rightarrow [0, \pi] : x \mapsto \int_{\mathbb{S}^{d-1}} \arccos(\langle x, y \rangle) dP(y). \tag{9}$$

The value  $\arccos(\langle x, y \rangle)$  is the arc length distance between  $x$  and  $y$  on  $\mathbb{S}^{d-1}$ . Thus, in analogy with the median from  $\mathbb{R}$ , we search for  $\mu$  minimizing the expected distance of  $x$  from  $X \sim P$ . The objective function (9) does not have to be minimized at a singleton; in case there are multiple points of minima of (9), the whole set of those  $\mu \in \mathbb{S}^{d-1}$  that minimize (9) is called the set of *Fréchet medians* of  $P$ . One situation when the Fréchet median is not unique is the case when the distribution of  $X \sim P \in \mathcal{P}(\mathbb{S}^{d-1})$  is centrally (sometimes also called antipodally) symmetric, i.e.  $X$  has the same distribution as  $-X$ , in which case the set of Fréchet medians is antipodally symmetric (for each Fréchet median  $\mu \in \mathbb{S}^{d-1}$  of  $P$  is also  $-\mu$  a Fréchet median of  $P$ ).

The angular Mahalanobis depth is defined only for  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  that admit a unique Fréchet median  $\mu \in \mathbb{S}^{d-1}$  (Ley et al. (2014), Assumption A). For such distributions  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ , the *angular Mahalanobis depth* of  $x \in \mathbb{S}^{d-1}$  w.r.t.  $P$  is

$$aMD(x; P) = P(\langle X, \mu \rangle \leq \langle x, \mu \rangle).$$

Here,  $X \sim P$  is a random vector distributed as  $P$  that is defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Clearly, the angular Mahalanobis depth depends only on the angle of  $x$  and

$\mu$ , or equivalently on the quantity  $\langle x, \mu \rangle$ . Its central regions (2) are thus always of the form

$$aMD_\alpha(P) = \left\{ x \in \mathbb{S}^{d-1} : \langle x, \mu \rangle \geq c_\alpha \right\}, \tag{10}$$

where  $c_\alpha \in [-1, 1]$  is the  $\alpha$ -quantile of the univariate distribution of  $\langle X, \mu \rangle$ , with  $X \sim P$ .

The Fréchet median of  $P$  is certainly rotationally equivariant. Thus, as remarked already in Ley et al. (2014), under the condition that  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  possesses a unique Fréchet median, the angular Mahalanobis depth trivially satisfies all conditions (D<sub>1</sub>)–(D<sub>5</sub>). It does not satisfy the directional quasi-concavity condition (D<sub>6</sub>) because of Theorem 3. The angular Mahalanobis depth is well suited especially for rotationally symmetric distributions with unique Fréchet median, particularly due to the very rigid form of its central regions (10). That collection of distributions contains some of the most important families, such as the Fisher-von Mises-Langevin distributions (Ley and Verdebout (2017), Sect. 2.3.1), for which the Fréchet (and thus also *aMD*-induced) median coincides with their location parameter. Nevertheless, a downside of *aMD* is that it does not satisfy condition (D<sub>7</sub>).

### 3.2 Angular distance-based depths

We now proceed to elaborate on the properties of the class of distance-based angular depths (Pandolfo et al. 2018), based on the seminal work of Liu and Singh (1992). For  $\delta : [-1, 1] \rightarrow [0, \infty)$  bounded and non-increasing, a general depth of this type is, for  $x \in \mathbb{S}^{d-1}$  and  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ , given by

$$aD_\delta(x; P) = \delta(-1) - E\delta(\langle x, X \rangle), \tag{11}$$

where  $X \sim P$  is a random variable distributed as  $P$ . Note that because  $\langle x, X \rangle \in [-1, 1]$  and  $\delta$  is bounded, these depths are well defined for any  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ . It is also easy to see that for any  $O \in \mathbb{R}^{d \times d}$  orthogonal we have

$$\langle Ox, Oy \rangle = x^\top O^\top Oy = x^\top O^{-1}Oy = x^\top y = \langle x, y \rangle \quad \text{for all } x, y \in \mathbb{S}^{d-1},$$

meaning that each distance-based depth trivially satisfies condition (D<sub>1</sub>). Provided that  $\delta$  is continuous on  $[-1, 1]$ , it is easy to see that using the Lebesgue dominated convergence theorem (Dudley (2002), Theorem 4.3.5) for the function

$$\varphi_\delta : \mathbb{S}^{d-1} \rightarrow [0, \infty) : x \mapsto E\delta(\langle x, X \rangle) = \int_{\mathbb{S}^{d-1}} \delta(\langle x, y \rangle) dP(y),$$

we obtain that as  $x_n \rightarrow x$  in  $\mathbb{S}^{d-1}$ , we have that  $\delta(-1) - aD_\delta(x_n; P) = \varphi_\delta(x_n) \rightarrow \varphi_\delta(x) = \delta(-1) - aD_\delta(x; P)$  as  $n \rightarrow \infty$ . Thus, for any  $\delta$  continuous, condition (D<sub>5</sub>) is satisfied, even with continuity in  $x \in \mathbb{S}^{d-1}$  for any  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ .

Three different choices of the distance function  $\delta$  are considered in Pandolfo et al. (2018):

$$\begin{aligned}\delta_{\cos}(t) &= 1 - t && \text{(cosine distance) ,} \\ \delta_{\text{arc}}(t) &= \arccos(t) && \text{(arc distance) ,} \\ \delta_{\text{chord}}(t) &= \sqrt{2(1 - t)} && \text{(chord distance) .}\end{aligned}$$

The function  $\delta_{\cos}$  leads to the so-called *cosine distance depth*,  $\delta_{\text{arc}}$  to the *arc distance depth*, and  $\delta_{\text{chord}}$  to the *chord distance depth*. Following Pandolfo et al. (2018), we denote the cosine distance depth by  $aD_{\cos}$ , the arc distance depth by  $aD_{\text{arc}}$ , and the chord distance depth by  $aD_{\text{chord}}$ . We elaborate on the properties of these depths case by case.

### 3.2.1 Cosine distance depth

The cosine distance depth of  $x \in \mathbb{S}^{d-1}$  simplifies to

$$aD_{\cos}(x; P) = 2 - \mathbb{E}(1 - \langle x, X \rangle) = 1 + \langle x, \mathbb{E}X \rangle.$$

This depth can be seen as a variant of the angular Mahalanobis depth discussed in Sect. 3.1. In particular, just as  $aMD$ , also  $aD_{\cos}$  depends only on the angle of  $\mathbb{E}X \in \mathbb{R}^d$  and  $x$  represented via the inner product  $\langle x, \mathbb{E}X \rangle$ , and thus all its central regions are spherical caps oriented in the direction of  $\mathbb{E}X$ , given that  $\mathbb{E}X \neq 0 \in \mathbb{R}^d$ . A distinctive feature of this depth is that it may happen that  $\mathbb{E}X = 0 \in \mathbb{R}^d$ , which is the only case when  $aD_{\cos}$  is constant on  $\mathbb{S}^{d-1}$ . Thus, strictly speaking, the cosine distance depth satisfies all (D<sub>1</sub>)–(D<sub>5</sub>) (with condition (D<sub>5</sub>) satisfied even in the sense of continuity rather than semi-continuity), but the depth itself is based solely on the expectation  $\mathbb{E}X$  and shares all the shortcomings of  $aMD$ . Of course,  $aD_{\cos}$  fails to obey both (D<sub>6</sub>) and (D<sub>7</sub>).

### 3.2.2 Arc distance depth

Already for  $d = 2$ , the arc distance depth

$$aD_{\text{arc}}(x; P) = \pi - \mathbb{E} \arccos(\langle x, X \rangle)$$

fails to be unimodal, in the sense that the set of (local, or global) maxima of the depth may fail to form a connected subset of the unit circle  $\mathbb{S}^1$ . In fact, a stronger statement can be shown. In the proof of the following theorem, given in Appendix B.4, we derive an explicit expression for the arc distance depth of an arbitrary distribution  $P \in \mathcal{P}(\mathbb{S}^1)$ .

**Theorem 4** There exists a distribution  $P \in \mathcal{P}(\mathbb{S}^1)$  such that the depth  $aD_{\text{arc}}(\cdot; P)$  attains infinitely many local extremes on  $\mathbb{S}^1$ .

Our Theorem 4 immediately shows that conditions (D<sub>3</sub>) and (D<sub>6</sub>) cannot be valid for  $aD_{\text{arc}}$ . We now illustrate the statement of Theorem 4 in a concrete example.

**Example 1** We first define distribution  $Q \in \mathcal{P}(\mathbb{R}^2)$  that is supported in two parallel straight lines  $L_+ = \{(x, 1) \in \mathbb{R}^2 : x \in \mathbb{R}\}$  and  $L_- = \{(x, -1) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ . We fix two constants:  $K > 0$  and a positive integer  $m \geq 1$ . Inside  $L_+$ , the measure  $Q$  is uniform on the interval  $A = [-K, K] \times \{1\}$ , with  $Q(A) = 1/2$ . Inside  $L_-$ , we take the finite set of  $m$  points in the interval  $[-K, K] \times \{-1\}$  given by the sequence  $\{y_j\}_{j=1}^m$ . We take  $y_j = (-K - \ell/2 + j\ell, -1) \in L_-$ , where  $\ell = 2K/m$  is the distance between two consecutive points of the sequence, putting  $Q(\{y_j\}) = 1/(2m)$  for each  $j = 1, \dots, m$ . The circular distribution  $P \in \mathcal{P}(\mathbb{S}^1)$  is defined as the image of  $Q$  under the projection  $\xi : \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{S}^1 : x \mapsto x/\|x\|$ . According to the detailed discussion given in Appendix B.5, the arc distance depth of  $P$  possesses at least  $m$  local extremes in  $\mathbb{S}^1$ .

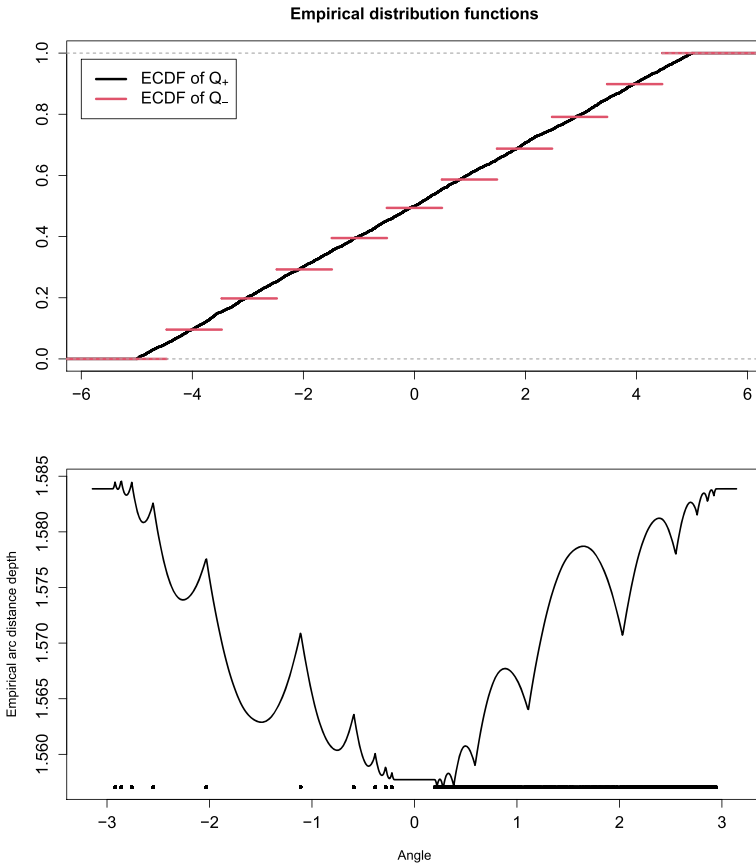
To illustrate this result, we took  $K = 5$  and  $m = 10$  in Fig. 1. We sampled  $n = 10\,000$  independent random points from  $P \in \mathcal{P}(\mathbb{S}^1)$  and computed the sample arc distance depth of all the points in a dense grid in  $\mathbb{S}^1$ . The resulting sample arc distance depth is displayed in the bottom panel of Fig. 1. As we see, the depth exhibits a rather wild behavior. On both the upper and lower half-circle, it possesses  $m = 10$  clearly distinguished local extremes (local maxima in the upper half-circle, and local minima in the lower half-circle). In between these local extremes, the depth changes drastically. Denote by  $F_+$  ( $F_-$ ) the (one-dimensional) distribution function of  $Q$  restricted to the line  $L_+$  ( $L_-$ ). As shown in Appendix B.5, the local extremes correspond exactly to points where the two distribution functions  $F_+$  and  $F_-$  intersect (upper panel of Fig. 1). This corroborates empirically our finding from Theorem 4.

Note that the fact that the measure  $P$  has atoms is not important in our example; instead of the  $m$  atoms on the lower half-circle, one could equally well take for  $Q_-$  a mixture of  $m$  Gaussian distributions with means  $y_j$ ,  $j = 1, \dots, m$ , respectively, and variances that are chosen small enough. In that case, the resulting distribution  $P$  can be taken smooth w.r.t. the Lebesgue measure on the unit circle, yet its arc distance depth will be quite similar to that in Fig. 1. For comparison, in Fig. 2 we display also the chord distance depth, and the cosine distance depth of the same dataset. While the chord distance depth does possess several local extremes, the cosine distance depth has a single (both local and global) maximum at the sample version of the directional mean  $E X / \|E X\|$  of the dataset. That is in accordance with our analysis from Sect. 3.2.1.

It is interesting to observe that the set of arc distance depth medians is exactly the set of all Fréchet medians obtained by minimizing (9). Indeed, the objective function in (9) is directly  $\varphi(x) = E \arccos(\langle x, X \rangle) = \pi - aD_{\text{arc}}(x; P)$ . Condition (D<sub>2</sub>) therefore translates into a natural question whether any distribution that is rotationally symmetric around  $\mu \in \mathbb{S}^{d-1}$  must contain  $\mu$  in the set of its Fréchet medians. In the following example, we show that already for  $d = 2$  this is not necessarily true.

**Example 2** Take  $P \in \mathcal{P}(\mathbb{S}^1)$  supported in four atoms

$$\begin{aligned} P(\{(\cos(\eta), \sin(\eta))\}) &= p/2, \\ P(\{(\cos(\pi - \eta), \sin(\pi - \eta))\}) &= p/2, \\ P(\{(\cos(-\varepsilon), \sin(-\varepsilon))\}) &= (1 - p)/2, \\ P(\{(\cos(-\pi + \varepsilon), \sin(\pi + \varepsilon))\}) &= (1 - p)/2, \end{aligned}$$

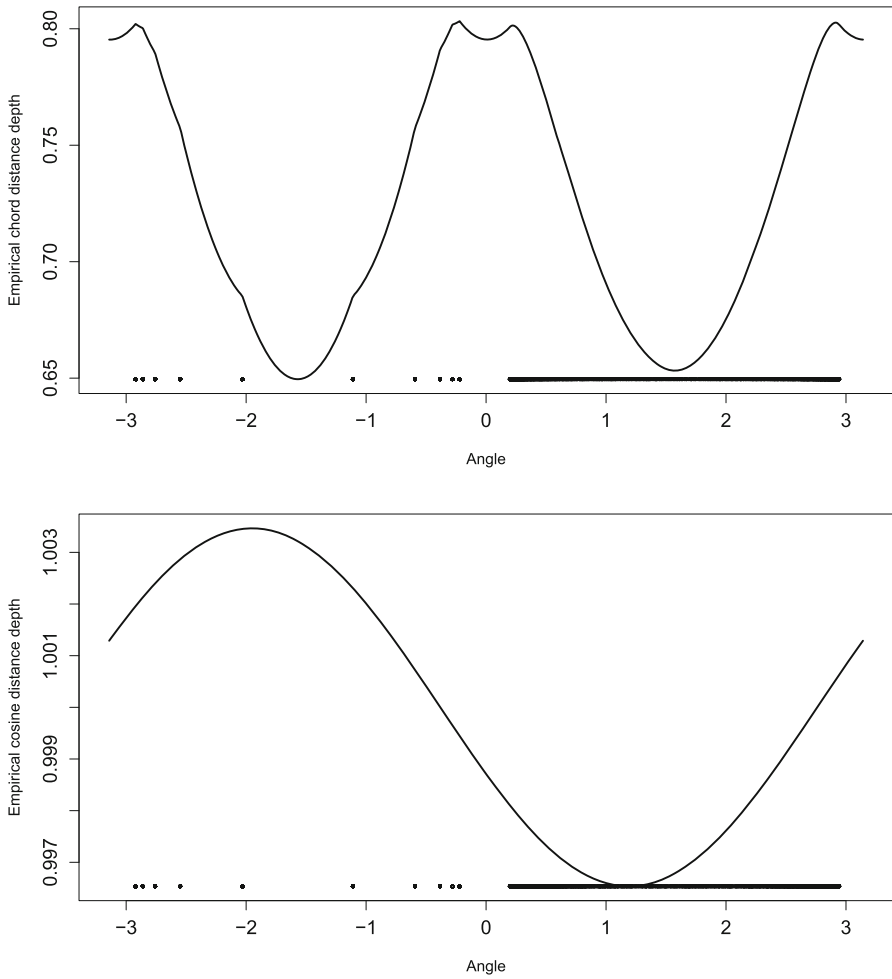


**Fig. 1** Arc distance depth and Example 1. Upper panel: The empirical versions of the distribution functions  $F_+$  (black line) and  $F_-$  (red line) of the components of  $Q$ , respectively, with  $K = 5$  and  $m = 10$ . Lower panel: The sample arc distance depth of all points  $x(\theta) = (\cos(\theta), \sin(\theta))$  of the unit circle  $\mathbb{S}^1$  as a function of their angle  $\theta \in [-\pi, \pi]$ . As can be seen, the arc distance depth behaves quite wildly, with several local extremes. The sample points are indicated as the small points at the bottom of the lower panel (color figure online)

for some  $p \in (0, 1)$  and  $\varepsilon, \eta \in [0, \pi/2]$ . Any such distribution is rotationally symmetric around the axis given by  $\mu = (0, 1) \in \mathbb{S}^1$ . By Theorem 2, we thus know that  $aD_{\cos}(\cdot; P)$  is also symmetric around the axis given by  $\mu$ . Take  $x(\theta) = (\cos(\theta), \sin(\theta)) \in \mathbb{S}^1$  to be a point given by its angle  $\theta \in [-\pi/2, \pi/2]$ . It is easy to evaluate the objective function (9) at  $x(\theta)$  directly. The function takes the form

$$\varphi(x(\theta)) = \frac{p}{2}(\rho(\theta - \eta) + \rho(\pi - \theta - \eta)) + \frac{1 - p}{2}(\rho(\theta + \varepsilon) + \rho(\pi + \theta - \varepsilon)),$$

where  $\rho(t) = \min\{|t|, 2\pi - |t|\}$  is the length of the shorter arc from the point  $x(t)$  to  $x(0) = (1, 0) \in \mathbb{S}^1$ . Direct computation gives that for, e.g.,  $p = 7/10, \eta = \pi/5$  and



**Fig. 2** The chord distance depth (upper panel) and the cosine distance depth (lower panel) for the dataset from Example 1. For a description see the caption of Fig. 1. For the chord distance depth we observe several local extremes, similarly as in Fig. 3 below. For the cosine distance depth we observe only a single maximum at the angle corresponding to the circular sample mean of the dataset

$\varepsilon = \pi/4$  we get

$$\pi - aD_{arc}(x(\theta); P) = \varphi(x(\theta)) = \begin{cases} \frac{113\pi}{200} & \text{for } \theta \in [-\pi/2, -\pi/4), \\ \frac{1}{50}(15\theta + 32\pi) & \text{for } \theta \in [-\pi/4, -\pi/5), \\ \frac{1}{10}(5\pi - 4\theta) & \text{for } \theta \in [-\pi/5, \pi/5), \\ \frac{3}{50}(5\theta + 6\pi) & \text{for } \theta \in [\pi/5, \pi/4), \\ \frac{87\pi}{200} & \text{for } \theta \in [\pi/4, \pi/2]. \end{cases}$$

In particular, it is directly seen that the set of Fréchet medians of  $P$  is exactly the pair of points  $\{x(\eta), x(\pi/2 - \eta)\}$ , and this set contains neither  $\mu = x(\pi/2)$ , nor  $-\mu = x(-\pi/2)$ . Consequently, condition  $(D_2)$  is not in general satisfied for the arc distance depth.

The arc distance depth verifies  $(D_4)$  thanks to the following more general result. It is proved in Appendix B.6.

**Theorem 5** Consider a distance-based depth  $aD_\delta$  from (11) with a function  $\delta$  that satisfies for some constant  $c > 0$  the condition

$$\delta(-t) = c - \delta(t) \quad \text{for all } t \in [-1, 1].$$

Then for all  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  and  $x \in \mathbb{S}^{d-1}$  we have that

$$aD_\delta(x; P) + aD_\delta(-x; P) = 2\delta(-1) - c. \tag{12}$$

In particular, condition  $(D_4)$  is satisfied for  $aD_\delta$ .

Applying Theorem 5 to the arc distance depth with  $c = \pi$  we obtain

$$aD_{arc}(x; P) + aD_{arc}(-x; P) = \pi \quad \text{for all } x \in \mathbb{S}^{d-1} \text{ and } P \in \mathcal{P}(\mathbb{S}^{d-1}).$$

Note, however, that this is a much stronger condition than  $(D_4)$ . It is in fact hardly desirable for an angular depth that (12) is true; it says that knowing the angular depth on any hemisphere of  $\mathbb{S}^{d-1}$ , we can determine the exact depth for the remaining points of  $\mathbb{S}^{d-1}$ . This is a property not unlike the problem with the constancy of the angular depths satisfying condition  $(D_6)$  at a hemisphere, as discussed in Theorem 3.

### 3.2.3 Chord distance depth

Consider again the measure  $P \in \mathcal{P}(\mathbb{S}^1)$  from Example 2 and apply  $aD_{chord}$ . A computation completely analogous to that performed in Example 2 shows that the chord distance depth median set of  $P$  is given by  $\{x(\eta), x(\pi/2 - \eta)\}$ , just as for the arc distance depth. At the same time, the antipodal points of the median set do not minimize the chord distance depth of  $P$ ; the unique minimizer of  $aD_{chord}(\cdot; P)$  is, in fact, the direction  $-\mu = (0, -1) \in \mathbb{S}^1$ . Thus, the chord distance depth verifies neither  $(D_2)$  nor  $(D_4)$ .

Further, we show that  $aD_{chord}$  and its central regions (2) violate also  $(D_3)$  and  $(D_6)$ . Even worse, already for quite simple distributions, the corresponding chord distance depth median set does not have to be a connected set in  $\mathbb{S}^{d-1}$ . To see this, consider a dataset of four points

$$x_1 = \frac{1}{7}(6, -2, 3), \quad x_2 = \frac{1}{\sqrt{19}}(3, 1, 3),$$

$$x_3 = \frac{1}{7}(-6, -2, 3), \quad \text{and } x_4 = \frac{1}{\sqrt{19}}(-3, 1, 3).$$

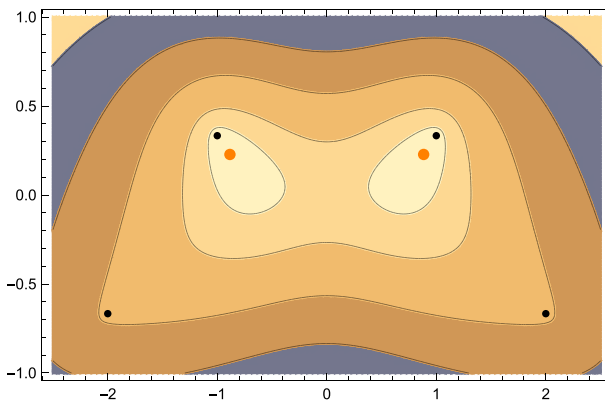
This dataset is contained in the northern hemisphere  $\mathbb{S}_+^2$ . A direct calculation gives the exact expression for the chord distance depth of each point  $x \in \mathbb{S}_+^2$ , w.r.t. the atomic measure  $P \in \mathcal{P}(\mathbb{S}^2)$  which assigns mass 1/4 to each of the four points  $x_i$ . This expression is maximized at two distinct points in  $\mathbb{S}_+^2$ , located at

$$y_1 \approx (0.653, 0.168, 0.739) \quad \text{and } y_2 \approx (-0.653, 0.168, 0.739).$$

We visualize the present setup using the projection

$$\Pi : \mathbb{S}_+^2 \rightarrow H : x = (x_1, x_2, x_3) \mapsto x/x_3 \tag{13}$$

of the northern hemisphere of  $\mathbb{S}^2$  into its tangent plane  $H = \{(x_1, x_2, 1) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$ . The map (13) is a special case of the gnomonic projection  $\Pi$  described in Appendix A. It is not difficult to see that (13) maps great circles in  $\mathbb{S}^2$  onto straight lines in  $H$ . In particular,  $\Pi$  preserves convexity and connectedness of sets. Several chord distance depth contours and the two different median points are displayed in Fig. 3. From the figure, we see that the central regions of the chord distance depth have to be neither convex nor connected. Also, several distinct median points can occur, even for very simple datasets. Consequently, both conditions (D<sub>3</sub>) and (D<sub>6</sub>) are violated for the chord distance depth.



**Fig. 3** Several contours of the chord distance depth  $aD_{chord}(\cdot; P)$  for  $P$  the uniform distribution in the four atoms displayed as black points. In the figure, the whole setup was projected using (13) from the open northern hemisphere of  $\mathbb{S}^2$  into the plane  $H$  tangent to  $\mathbb{S}^2$  at the northern pole  $(0, 0, 1) \in \mathbb{S}^2$ ; for details see Appendix A.2. The two median points of the chord distance depth are displayed in orange. As can be seen, for the chord distance depth, the central regions are neither spherical (star) convex nor connected, in general (color figure online)



### 3.3 Angular simplicial depth

The angular simplicial depth of  $x \in \mathbb{S}^{d-1}$  w.r.t.  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  is defined as

$$asD(x; P) = P(x \in \text{sconv}(\{X_1, \dots, X_d\})), \tag{14}$$

where  $X_1, \dots, X_d$  are independent random variables distributed as  $P$  defined on  $(\Omega, \mathcal{F}, P)$ , and  $\text{sconv}(A)$  is the spherical closed convex hull of the set  $A \subseteq \mathbb{S}^{d-1}$  defined as

$$\text{sconv}(A) = \mathbb{S}^{d-1} \cap \text{conv}(\text{rad}(A))$$

for  $\text{rad}(A)$  the radial extension of the set  $A$  defined in (8). In words,  $\text{sconv}(A)$  is the intersection of  $\mathbb{S}^{d-1}$  with the closed convex hull of the positive cone of  $A$ . In typical situations, e.g. if all the variables  $X_1, \dots, X_d$  are contained in a hemisphere, the set  $\text{sconv}(\{X_1, \dots, X_d\})$  constitutes a spherical simplex, e.g. the shorter arc between  $X_1$  and  $X_2$  for  $d = 2$ , or a spherical triangle determined by  $X_1, X_2, X_3$  for  $d = 3$ . This also motivates the name angular simplicial depth, in analogy with the simplicial depth in linear spaces (Liu 1990). This parallel can be used to show that just like the simplicial depth, also its angular variant satisfies (D<sub>5</sub>). The proof can be found in Appendix B.7. It uses the fact that (D<sub>1</sub>) is true for  $asD$ , which has been observed already by Liu and Singh (1992).

**Theorem 6** The function  $\mathbb{S}^{d-1} \rightarrow [0, 1] : x \mapsto asD(x; P)$  is upper semi-continuous for any  $P \in \mathcal{P}(\mathbb{S}^{d-1})$ .

It is, however, important to realize that in (14) it can happen that all  $X_1, \dots, X_d \in \mathbb{S}^{d-1}$  lie inside a hyperplane passing through the origin in  $\mathbb{R}^d$ , e.g. in a single great circle of  $\mathbb{S}^{d-1}$ . That occurs if there are hyperplanes  $H$  passing through the origin such that  $H \cap \mathbb{S}^{d-1}$  is of positive  $P$ -mass. In those situations, the spherical convex hull of  $X_1, \dots, X_d$  is a subset of  $H \cap \mathbb{S}^{d-1}$ , which bears the interpretation of a degenerate spherical simplex. In the extreme case when all  $X_j, j = 1, \dots, d$ , lie in a single line in direction  $u \in \mathbb{S}^{d-1}$ , that is  $X_j \in \{u, -u\}$  for all  $j = 1, \dots, d$  and not all  $X_j$  are the same direction, the set  $\text{sconv}(\{X_1, \dots, X_d\})$  reduces to only the two-point set  $\{u, -u\}$ . We use this simple observation to construct our next example. It demonstrates that even for rotationally symmetric distributions, the angular simplicial depth may exhibit undesirable behavior.

**Example 3** Consider the uniform distribution  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  that is supported on the equator of the unit sphere  $\mathbb{S}_0^{d-1} = \mathbb{S}^{d-1} \cap H_0$ , where  $H_0 = \{u \in \mathbb{S}^{d-1} : \langle u, (0, \dots, 0, 1) \rangle = 0\}$ . For  $d = 2$ , this amounts to the uniform distribution on the pair of points  $(1, 0), (-1, 0) \in \mathbb{S}^1$ . For  $d = 3$  we have the uniform probability measure on the equator of  $\mathbb{S}^2$ , which is the great circle  $\{(u_1, u_2, 0) \in \mathbb{S}^2 : u_1^2 + u_2^2 = 1\}$ . Since any simplex formed by independent random points  $X_1, \dots, X_d$  sampled from  $P$  lies in the hyperplane  $H_0$ , we have that also  $\text{sconv}(\{X_1, \dots, X_d\}) \subset H_0$  almost surely. Thus, the only points in  $\mathbb{S}^{d-1}$  that attain

positive angular simplicial depth are the points on the equator  $\mathbb{S}_0^{d-1}$ . On the other hand, the distribution  $P$  is clearly rotationally symmetric around the axis given by  $\mu = (0, \dots, 0, 1) \in \mathbb{S}^{d-1}$ , yet since  $\mu \notin \mathbb{S}_0^{d-1}$ , we have  $asD(\mu; P) = asD(-\mu; P) = 0$ , and condition (D<sub>2</sub>) is not satisfied. In addition, the rotational invariance of the simplicial depth (D<sub>1</sub>) applied to the  $(d - 2)$ -sphere  $\mathbb{S}_0^{d-1}$  gives that the value of the angular simplicial depth of  $P$  on  $\mathbb{S}_0^{d-1}$  is a constant denoted by  $c_d > 0$ . But,  $\mathbb{S}_0^{d-1}$  is antipodally symmetric, meaning that neither condition (D<sub>3</sub>) nor (D<sub>4</sub>) can be satisfied for  $asD$  in general. To see that, it is enough to take any geodesic joining a point  $x_0 \in \mathbb{S}_0^{d-1}$  with its antipodal reflection  $x_1 = -x_0 \in \mathbb{S}_0^{d-1}$  that does not lie entirely inside  $\mathbb{S}_0^{d-1}$ . At the endpoints  $x_0, x_1 \in \mathbb{S}_0^{d-1}$  we have  $asD(x_0; P) = asD(x_1; P) = c_d > 0$ , while  $asD(x; P) = 0$  for each  $x$  on the rest of the geodesic.

For more contrived examples of distributions  $P$  in  $\mathbb{S}^{d-1}$  that make  $asD$  violate conditions (D<sub>2</sub>), (D<sub>3</sub>), or (D<sub>6</sub>), one could design  $P$  to lie in a hemisphere and use the gnomonic projection from Appendix A.2 to transfer  $asD$  from  $\mathbb{S}^{d-1}$  to the usual simplicial depth in  $\mathbb{R}^{d-1}$ . Several examples of distributions in  $\mathbb{R}^{d-1}$  for which the simplicial depth violates (L<sub>2</sub>), (L<sub>3</sub>), or (L<sub>6</sub>) are given in Nagy (2023). All these can be naturally used also as counterexamples for  $asD$ .

### 3.4 Angular halfspace depth

Expanding the classical Tukey’s halfspace depth  $hD$  (Tukey 1975; Donoho and Gasko 1992) from (4) to directional data, Small (1987) defined the *angular halfspace depth* of  $x \in \mathbb{S}^{d-1}$  w.r.t.  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  to be

$$ahD(x; P) = \inf_{H \in \mathcal{H}_0} \{P(H) : x \in H\}, \tag{15}$$

where  $\mathcal{H}_0 = \{H_{0,u} : u \in \mathbb{S}^{d-1}\}$  is the set of those closed halfspaces

$$H_{0,u} = \{y \in \mathbb{R}^d : \langle y, u \rangle \geq 0\}$$

in  $\mathbb{R}^d$  that contain the origin  $0 \in \mathbb{R}^d$  on their boundary and possess inner normal vector  $u \in \mathbb{S}^{d-1}$ . The angular halfspace depth has been studied in Liu and Singh (1992), where several of its properties have been discussed.

Since that pioneering work, however, the angular halfspace depth has not received much attention in the literature, mainly for two reasons: (i) due to the problem of  $ahD$  being constant on a hemisphere of minimal  $P$ -mass (Liu and Singh 1992, Proposition 4.6, see also our Theorem 3), and (ii) because of the perceived high computational cost of  $ahD$  for random samples (Pandolfo et al. 2018). The latter difficulty was recently fully resolved by making substantial progress on the computational front (Dyckerhoff and Nagy 2023). The problem with the constancy of  $ahD$  on a hemisphere is tightly connected with the quasi-concavity of  $ahD$  and our condition (D<sub>6</sub>); directly from the definition (15) it is easy to see that the hemisphere of constant  $ahD$  will be (any) open hemisphere with smallest  $P$ -mass. The latter drawback can be an

issue in classification studies. As will be illustrated shortly in Sect. 4, in applications where each training set belongs to a hemisphere, the angular halfspace depth can be successfully adopted. If the data does not satisfy this condition, strategies specifically designed to cope with the problem of constancy of the  $ahD$  are required. Possible approaches will be mentioned in Sect. 5 below, and investigated in a follow-up work on the practice of depth-based classification of directional data.

The theoretical properties of the angular halfspace depth have been recently studied in Nagy and Laketa (2023). They can be shown to align closely with the properties of the standard halfspace depth in linear spaces (Donoho and Gasko 1992; Rousseeuw and Ruts 1999). In particular, in Nagy and Laketa (2023) it is proved that  $ahD$  satisfies all conditions (D<sub>1</sub>)–(D<sub>7</sub>), including the debatable condition of full quasi-concavity (D<sub>6</sub>). This reinforces the idea that the angular halfspace depth is also interesting in the context of the classification of directional data (at least, provided that the constancy problem is properly addressed).

## 4 Supervised classification using angular depths

This section aims at illustrating the main implications of the theoretical results derived above when the goal is to classify directional data. More specifically, the focus will be on the desired properties of non-rigidity of central regions (D<sub>7</sub>), and of quasi-concavity (D<sub>6</sub>) entailing spherical convexity of the depth regions. Simulated and real data are exploited to highlight these implications.

As discussed throughout this paper, a large body of research exists on using statistical depth functions in the supervised classification task. The most frequently used depth-based tools are the max-depth and the DD-classifiers (Li et al. 2012). As the latter generalizes the former, we focus on DD-classification, and start by briefly recalling its use within the directional setting in Sect. 4.1. The performance of DD-classifiers depends on the particular depth function used. For that reason, Sects. 4.2–4.4 investigate the classifier performances in terms of misclassification rates when associated with different angular depth functions. Three main scenarios are investigated. First, the importance of the desired property (D<sub>7</sub>) is underlined by considering a case where the optimal discriminating boundary on the sphere is not a great circle (Sect. 4.2). Then, the advantages and disadvantages of property (D<sub>6</sub>) are contrasted in Sects. 4.3 and 4.4. For each scenario, performances of the angular distance-based, the angular halfspace, and the angular simplicial depths are compared. For the angular simplicial depth, however, it must be mentioned that severe computational issues hamper its use in practice (see e.g. the discussion in Pandolfo and D’Ambrosio 2021); its computation is practically feasible only for training sets with sizes less than 400 observations. Hence, the results for the angular simplicial depth are provided only for smaller training set sizes. In addition, the angular Mahalanobis depth is not considered in our simulations because of its similarity with the angular cosine depth, see Sect. 3.2.1.

In a general classification problem in a measurable space  $\mathcal{S}$ , we aim to distinguish between points sampled from one of  $G \geq 2$  distinct probability measures  $\{P_i\}_{i=1}^G \subset \mathcal{P}(\mathcal{S})$ . A generic classifier in  $\mathcal{S}$  can be defined as any function class  $\gamma : \mathcal{S} \rightarrow \{1, 2,$

$\dots, G\}$ , which associates a point  $x \in \mathcal{S}$  with one of the distributions  $P_i \in \mathcal{P}(\mathcal{S})$  for  $i = \text{class}(x)$ .

Our depth-based classification results will be benchmarked against the empirical Bayes classifier assuming data comes from the family of Kent distributions. Assuming equal priors, the Bayes classifier is generally defined as

$$\text{class}_B(x) = \underset{i=1, \dots, G}{\text{argmax}} f_i(x), \tag{16}$$

where  $f_i$  is the density which, by assumption, corresponds to the distribution  $P_i$ . Note that the classifier (16) is optimal as it reaches the minimum achievable average misclassification rate, if the densities  $f_i$  are known.

Kent’s family is a special case of the Fisher-Bingham distribution, whose probability density function on  $\mathbb{S}^{d-1}$  is given by

$$f_{FB}(x; \mu, \kappa, A) = \frac{1}{a(\kappa, A)} \exp(\kappa \mu^\top x - x^\top A x), \tag{17}$$

where  $\mu \in \mathbb{S}^{d-1}$  and  $\kappa > 0$  are the location and concentration parameters, respectively,  $A$  is a  $d \times d$  symmetric matrix, and  $a(\kappa, A) > 0$  is a normalizing constant. To obtain a Kent distribution, the eigenvalues of  $A$  are constrained to be  $\lambda_1 = 0$  and  $\lambda_2 = -\lambda_3$ , and we fix  $|\lambda_2| < \kappa/2$  to have a unimodal distribution.

Kent’s distribution is a quite general distribution on the sphere. Partially, it mimics the contours of the bivariate normal distribution. Plugging (17) into (16), we obtain the empirical Bayes classifier under Kent. It takes the form

$$\text{class}_{Bk}(x) = \underset{i=1, \dots, G}{\text{argmax}} f_{FB}(x; \hat{\mu}_i, \hat{\kappa}_i, \hat{A}_i),$$

where  $\hat{\mu}_i, \hat{\kappa}_i, \hat{A}_i$  are the maximum likelihood estimates of the parameters  $\mu, \kappa$  and  $A$ , respectively, obtained from the  $i$ -th training set (the set of observations labeled empirical as belonging to  $P_i$ ).

### 4.1 The DD-classifier

A directional depth-based classifier is a particular classifier that takes the form

$$\text{class}_{aD}(x) = r(aD(x; P_1), aD(x; P_2), \dots, aD(x; P_G)), \tag{18}$$

where  $x \in \mathbb{S}^{d-1}$  is a new observation to be classified,  $aD(x; P_i)$  is an angular depth of  $x$  w.r.t. the distribution  $P_i \in \mathcal{P}(\mathbb{S}^{d-1})$ , and  $r : [0, \infty)^G \rightarrow \{1, 2, \dots, G\}$  is a discriminating function in the depth space. In practice, of course, the true distributions  $P_i \in \mathcal{P}(\mathbb{S}^{d-1})$  are not known, and each  $P_i$  is in (18) replaced by the corresponding empirical measure of a random sample of data labeled as coming from  $P_i$ . We denote such an empirical measure by  $\hat{P}_i \in \mathcal{P}(\mathbb{S}^{d-1})$ .

In (18), the choice of the depth function  $aD$  and of the discriminating function  $r$  matters. As for the latter, three functions have been considered in the literature when the data are directional (Demni et al. 2021; Pandolfo and D'Ambrosio 2021):

- *The linear discriminant (LDA)*;
- *The quadratic discriminant (QDA)*; and
- *The  $k$ -nearest neighbors discriminant function (KNN)*.

Each of these methods discriminates between groups directly in the DD-space of depths, which is a subset of  $[0, \infty)^G$ , irrespective of the dimension  $(d - 1)$  of the directional data in  $\mathbb{S}^{d-1}$ . This gives rise to three groups of DD-classifiers based on angular depths that we explore in what follows.

As for the choice of the angular depth function in (18), in the literature we find only classifiers using the collection of distance-based angular depths (Demni et al. 2021; Pandolfo and D'Ambrosio 2021). The adoption of distance-based angular depths in the situation when the underlying distributions in  $\mathbb{S}^{d-1}$  fail to be rotationally symmetric is, however, questionable (Demni and Porzio 2021). For the sake of illustration, in what follows, directional DD-classifiers are investigated for data in  $\mathbb{S}^2$  and in the case of two classes, i.e.,  $G = 2$ .

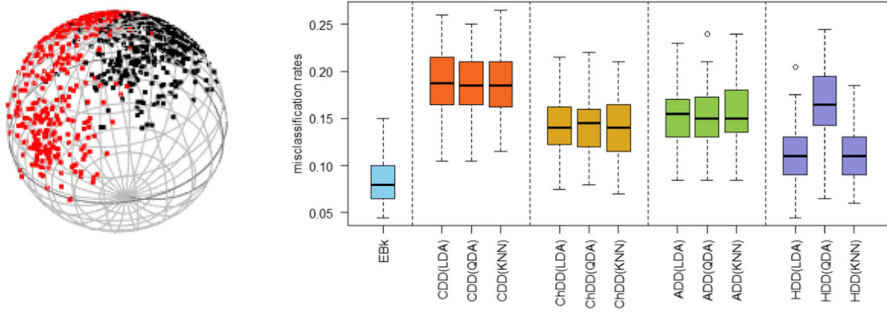
## 4.2 Discriminating boundary is not a great circle: property (D<sub>7</sub>) matters

The non-rigidity of central regions allows depth functions to consider the geometry of the data. In particular, if a depth does not satisfy condition (D<sub>7</sub>), it will perform poorly whenever the optimal discriminating boundary is not a great circle. To illustrate this, data were generated according to two specific settings for which the optimal Bayes discriminating boundary does not simplify to a great circle (Kent and Mardia 2013):

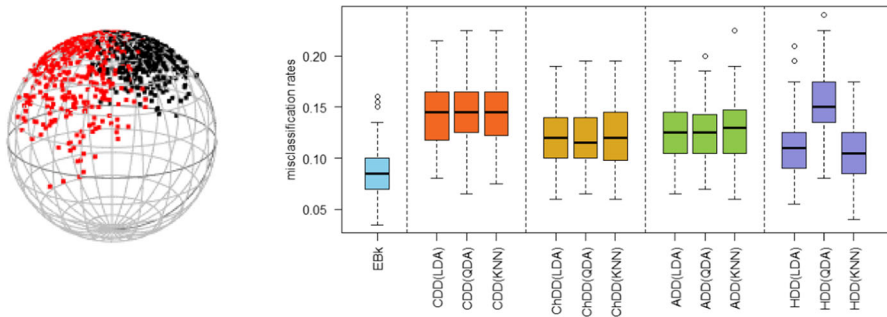
- **Setup 1:**  $P_1$  and  $P_2$  from Kent, difference only in location. The two groups come from two different rotations of the same Kent distribution  $P_0$ , for which we set  $\mu_0 = (0, 0, 1) \in \mathbb{S}^2$ ,  $\kappa_0 = 25$ , and  $A_0 = \text{diag}(0, 8, -8)$  is a diagonal matrix in  $\mathbb{R}^{3 \times 3}$ . The distribution  $P_1$  is obtained by rotating  $P_0$  by a rotation matrix  $R$  such that  $\mu_1 = R\mu_0$ , with  $\mu_1 = (\sin(\pi/6)\cos(\pi/4), \sin(\pi/6)\sin(\pi/4), \cos(\pi/6)) \in \mathbb{S}^2$ . The distribution  $P_2$  is obtained by rotating the  $P_0$  through the rotation matrix  $R^T$ . Random samples from  $P_1$  and  $P_2$  are depicted in black and in red in Fig. 4 (left panel).
- **Setup 2:**  $P_1$  and  $P_2$  from Kent, difference in location and concentration. Data are simulated by adopting the same procedure as in **Setup 1**, using the same rotation matrix  $R$ . However, the second group comes from a rotation (by  $R^T$ ) of a Kent distribution  $P_0$  as in **Setup 1** but with concentration  $\kappa = 20$ . A training set from these distributions is plotted in Fig. 5 (left panel).

Within each setup, 100 training sets of size  $n_{tr} \in \{100, 200, 400\}$  were generated (with each group size given by  $n_{tr}/2$ ). Then, for each of the 100 training sets, 100 testing sets of size  $n_{test} \in \{50, 100, 200\}$  ( $n_{test}/2$  per each group) were generated and classified.

The performance of the classifiers is evaluated in terms of empirical misclassification rates, i.e. the proportion of misclassified observations in each replication. For each



**Fig. 4** Section 4.2: Discriminating boundary is not a great circle, **Setup 1**. Kent distributions with a difference in location. A spherical plot of a single training set (left panel) and boxplots of empirical misclassification rates (right panel). Each boxplot reports the performance of a different classifier: the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD) and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and *KNN*, respectively) (color figure online)



**Fig. 5** Section 4.2: Discriminating boundary is not a great circle, **Setup 2**. Kent distributions with a difference in location and in concentration. A spherical plot of a single training set (left panel) and boxplots of empirical misclassification rates (right panel). Each boxplot reports the performance of a different classifier: the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD) and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and *KNN*, respectively) (color figure online)

simulation setting and each classifier, the average misclassification rates for different training set sizes are given in Tables 2 and 3. The distribution of the misclassification rates for training set size  $n_{tr} = 400$  is also summarized through the boxplots in Figs. 4 and 5 (right panels). Given their much higher computational cost, the angular simplicial DD-classifiers were not included in the comparison with  $n_{tr} = 400$ .

Generally speaking, our results highlight the importance of property (D<sub>7</sub>): the DD-classifier based on the cosine depth performs relatively poorly when compared to the others. The same would have happened if the angular Mahalanobis depth had been considered. We also note that, for  $n_{tr} = 400$ , the angular halfspace depth DD-classifiers associated with LDA and *KNN* perform quite well and outperform all distance-based DD-classifiers in both **Setups 1** and **2** (Figs. 4 and 5, right panels).

**Table 2** Section 4.2: Discriminating boundary is not a great circle, **Setup 1**. Kent distributions with a difference in location. Average misclassification rates for training set sizes  $n_{tr} \in \{100, 200, 400\}$  of the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD), the angular simplicial DD-classifiers (SDD) and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and KNN, respectively). Best performance for each training size is highlighted in bold

Training size	EBk	Separating function	CDD	ChDD	ADD	SDD	HDD
$n_{tr} = 100$	<b>0.084</b>	LDA	0.183	0.145	0.154	0.169	0.154
		QDA	0.180	<b>0.139</b>	0.147	0.187	0.187
		KNN	0.189	0.141	0.150	0.167	0.151
$n_{tr} = 200$	<b>0.081</b>	LDA	0.181	0.140	0.150	0.134	0.126
		QDA	0.179	0.139	0.148	0.178	0.176
		KNN	0.182	0.141	0.150	<b>0.121</b>	0.124
$n_{tr} = 400$	<b>0.084</b>	LDA	0.187	0.141	0.151	—	0.111
		QDA	0.185	0.141	0.151	—	0.167
		KNN	0.187	0.140	0.153	—	<b>0.110</b>

**Table 3** Section 4.2: Discriminating boundary is not a great circle, **Setup 2**. Kent distributions with a difference in location and in concentration. Average misclassification rates for training set sizes  $n_{tr} \in \{100, 200, 400\}$  of the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD), the angular simplicial DD-classifiers (SDD) and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and KNN, respectively). Best performance for each training size is highlighted in bold

Training size	EBk	Separating function	CDD	ChDD	ADD	SDD	HDD
$n_{tr} = 100$	<b>0.070</b>	LDA	0.132	0.108	0.112	0.145	0.135
		QDA	0.126	<b>0.107</b>	0.112	0.167	0.174
		KNN	0.137	0.113	0.116	0.146	0.129
$n_{tr} = 200$	<b>0.083</b>	LDA	0.139	0.116	0.121	0.129	<b>0.111</b>
		QDA	0.140	0.116	0.122	0.171	0.164
		KNN	0.141	0.119	0.124	0.121	0.112
$n_{tr} = 400$	<b>0.088</b>	LDA	0.144	0.120	0.126	—	0.110
		QDA	0.143	0.120	0.124	—	0.154
		KNN	0.143	0.121	0.127	—	<b>0.104</b>

The average misclassification rates recorded for different values of the training set size (Tables 2, 3, 4) confirm our results described for  $n_{tr} = 400$ . Worth noting is that the angular simplicial depth, which is computationally prohibitive, does not fare better overall than the angular halfspace depth within this scenario.

**Table 4** Section 4.3: Ellipse-like density contours, **Setup 3**. Kent distributions with a difference in location, concentration, and shape. Average misclassification rates for training set sizes  $n_{tr} \in \{100, 200, 400\}$  of the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD), the angular simplicial DD-classifiers (SDD) and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and KNN, respectively). Best performance for each training size is highlighted in bold

Training size	EBk	Separating function	CDD	ChDD	ADD	SDD	HDD
$n_{tr} = 100$	<b>0.179</b>	LDA	0.321	0.275	0.276	0.252	0.258
		QDA	0.262	0.275	0.276	0.251	0.290
		KNN	0.305	0.286	0.280	0.265	<b>0.242</b>
$n_{tr} = 200$	<b>0.169</b>	LDA	0.303	0.257	0.259	0.238	0.227
		QDA	0.237	0.253	0.255	0.245	0.250
		KNN	0.247	0.260	0.263	0.232	<b>0.215</b>
$n_{tr} = 400$	<b>0.176</b>	LDA	0.291	0.249	0.250	–	0.217
		QDA	0.218	0.249	0.250	–	0.249
		KNN	0.221	0.251	0.254	–	<b>0.206</b>

### 4.3 Ellipse-like density contours: property (D<sub>6</sub>) matters

Spherical convexity of the angular depth central regions ensures the ability to capture ellipse-like data structure on (hemi)spheres. As a consequence, depth functions fulfilling condition (D<sub>6</sub>) should provide good performances if the underlying density has ellipse-like density contours. In our next setup, data were drawn from the Kent and the scaled von Mises-Fisher distribution. While both of them possess ellipse-like contours, the latter exhibits ellipses with a higher level of eccentricity.

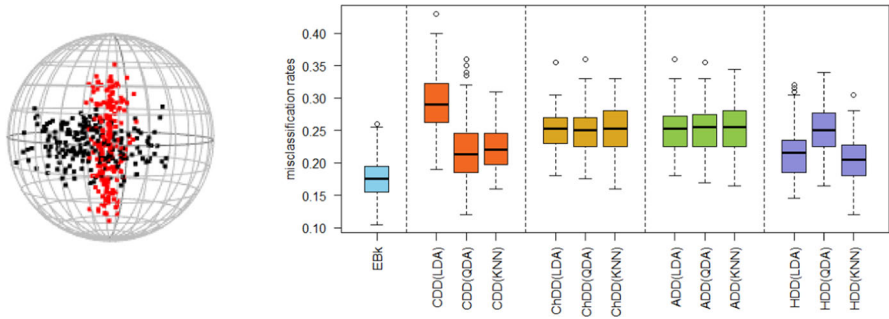
The family of scaled von Mises-Fisher distributions (Sealy and Wood 2019) on  $\mathbb{S}^{d-1}$  is generated by applying a bijective transformation of the sphere  $\mathbb{S}^{d-1}$  to itself. The probability density function of the scaled von Mises-Fisher distribution w.r.t. the spherical Lebesgue measure on  $\mathbb{S}^{d-1}$  takes the form

$$f(x) = \frac{1}{c_d(\kappa)a_1} \left\{ (x^\top \mu/a_1)^2 + \sum_{j=2}^d (x^\top \gamma_j/a_j)^2 \right\}^{-(d-1)/2} \times \exp \left\{ \frac{\kappa x^\top \mu/a_1}{\left\{ (x^\top \mu/a_1)^2 + \sum_{j=2}^d (x^\top \gamma_j/a_j)^2 \right\}^{1/2}} \right\},$$

where  $c_d(\kappa) > 0$  is a normalizing constant,  $\{\mu, \gamma_2, \dots, \gamma_d\} \subset \mathbb{R}^d$  are location parameters,  $a_1, a_2, \dots, a_d > 0$  are shape parameters satisfying  $\prod_{j=2}^d a_j = 1$ , and  $\kappa > 0$  is a concentration parameter.

The following two simulation settings were considered.



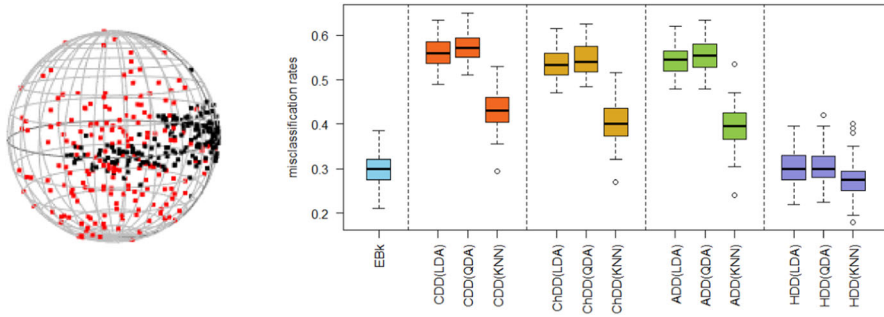


**Fig. 6** Section 4.3: Ellipse-like density contours, **Setup 3**. Kent distributions with a difference in location, concentration, and orientation. A spherical plot of a single training set (left panel) and boxplots of empirical misclassification rates (right panel). Each boxplot reports the performance of a different classifier: the empirical Bayes under Kent (EBK), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD) and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and KNN, respectively) (color figure online)

- **Setup 3:**  $P_1$  and  $P_2$  from Kent, difference in location, concentration, and orientation. For  $P_1$  we have  $\mu_1 = (1, 0, 0) \in \mathbb{S}^2$ ,  $\kappa_1 = 25$  and  $A_1 = \text{diag}(0, 8, -8)$ . Data from  $P_2$  are obtained by transforming data sampled from a Kent distribution with  $\mu = (-1, 0, 0) \in \mathbb{S}^2$ ,  $\kappa = 80$  and  $A = \text{diag}(0, -40, 40)$  to polar coordinates, and slightly perturbing the first polar coordinate direction (by adding a constant corresponding to an angle  $\pi/18$ ). Random samples from these distributions can be seen in Fig. 6 (left panel).
- **Setup 4:**  $P_1$  and  $P_2$  from scaled von Mises-Fisher, difference in location and shape. Distribution  $P_1$  has mean direction  $\mu_1 = (0, 0, 1) \in \mathbb{S}^2$ , and  $P_2$  has  $\mu_2 = (0.43, -0.9, 0) \in \mathbb{S}^2$ , respectively, and concentration levels  $\kappa_1 = \kappa_2 = 5$ . The shape parameters are set to be  $a_1 = 1, a_2 = 10, a_3 = 0.1$  for  $P_1$  and  $a_1 = a_2 = a_3 = 1$  for  $P_2$ . Random samples from the two distributions are given in Fig. 7 (left panel).

Results for both setups, for  $n_{tr} = 400$ , are summarized through boxplots (Figs. 6 and 7, right panels). Misclassification rates are obtained over 100 independent runs. Tables 4 and 5 also report the average misclassification rates for  $n_{tr} = 100$ , and  $n_{tr} = 200$ , with values obtained when also the angular simplicial depth is adopted for these latter cases.

Overall, we found the angular halfspace depth working quite, if not extremely well. Worth noting is the case of the scaled von Mises-Fisher distributions (**Setup 4**). There, the classifier based on the angular halfspace depth is able to outperform even the empirical Bayes under Kent. Interestingly, the Kent distribution is known to share quite similar properties to the scaled von Mises-Fisher (Scealy and Wood 2019). Adopting the angular halfspace depth thus seems to be quite beneficial if minor departures from the assumed distribution are present (which can be well the case for real data applications).



**Fig. 7** Section 4.3: Ellipse-like density contours, **Setup 4**. Scaled von Mises distributions with differences in location and shape. A spherical plot of a single training set (left panel) and boxplots of empirical misclassification rates (right panel). Each boxplot reports the performance of a different classifier: the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD), and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and KNN, respectively) (color figure online)

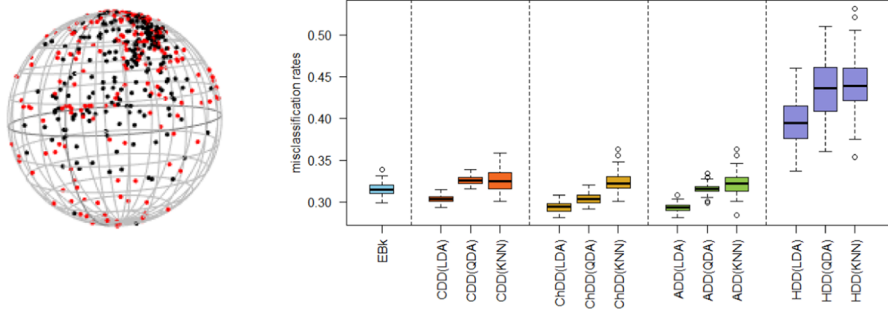
**Table 5** Section 4.3: Ellipse-like density contours, **Setup 4**. Scaled von Mises distributions with differences in location and shape. Average misclassification rates for training set sizes  $n_{tr} \in \{100, 200, 400\}$  of the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD), the angular simplicial DD-classifiers (SDD) and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and KNN, respectively). Best performance for each training size is highlighted in bold

Training size	EBk	Separating function	CDD	ChDD	ADD	SDD	HDD
$n_{tr} = 100$	<b>0.315</b>	LDA	0.562	0.537	0.545	0.346	0.329
		QDA	0.571	0.543	0.554	0.332	0.322
		KNN	0.524	0.490	0.493	0.326	<b>0.316</b>
$n_{tr} = 200$	0.304	LDA	0.561	0.539	0.547	0.324	0.313
		QDA	0.577	0.552	0.559	0.313	0.301
		KNN	0.484	0.445	0.435	<b>0.284</b>	0.289
$n_{tr} = 400$	0.299	LDA	0.559	0.534	0.544	—	0.304
		QDA	0.572	0.544	0.555	—	0.303
		KNN	0.433	0.404	0.394	—	<b>0.274</b>

**4.4 Data not confined to a hemisphere: property (D<sub>6</sub>) hampers**

Spherical convexity of central regions (D<sub>6</sub>) helps a lot when rotational symmetry cannot be assumed, and when the geometry of P is ellipse-like. On the other hand, this same blessed property brings damnation — by Theorem 3, (D<sub>6</sub>) implies constancy over a hemisphere. That is clearly an undesirable feature in classification studies of directional data.

To illustrate, we present a small real data example. The problem of detecting cardiac arrhythmia from Electrocardiogram (ECG) waves is considered. More specifically,



**Fig. 8** Section 4.4: Data not confined to a hemisphere, real data example. A spherical plot of the angular variables QRST- and T-waves for healthy patients (in black) and patients with arrhythmia (in red, left panel), and boxplots of empirical misclassification rates (right panel). Each boxplot reports the performance of a different classifier: the empirical Bayes under Kent (EBk), the cosine DD-classifiers (CDD), the chord DD-classifiers (ChDD), the arc distance DD-classifiers (ADD), and the angular halfspace DD-classifiers (HDD). For each angular depth, three classification rules were applied in the DD-space (LDA, QDA, and KNN, respectively) (color figure online)

waves are dealt with as angular variables, and we select two of the waves used in Demni (2021), recorded for 430 patients: the QRST-wave and the T-wave. The first describes the global ventricular repolarization while the second corresponds to the rapid repolarization of the contractile cells.

The analysis aims to distinguish normal vs. arrhythmia cases. A visualization of the two angular variables on  $\mathbb{S}^2$  for 245 normal patients (depicted in black) and for 185 patients having arrhythmia (depicted in red) is given in Fig. 8 (left panel).

A simulation study was then carried on, and — to evaluate the performance of the different classifiers — a threefold stratified cross-validation method was adopted (the percentage of samples from each class was preserved). The experiment was repeated 100 times. The obtained distribution of the misclassification rates of the empirical Bayes under Kent, the distance-based DD-classifiers, and the angular halfspace DD-classifiers are provided through box-plots in Fig. 8 (right panel). The angular simplicial depth was excluded because of the size of the training set.

The distance-based DD-classifiers achieved the best overall performance, especially the chord DD-classifiers and the arc distance DD-classifiers when associated with the linear discriminant rule (LDA). Those methods performed even better than the empirical Bayes under Kent. The reason for that is that the data do not follow an ellipse-like geometry.

On the other hand, the DD-classifiers based on the angular halfspace depth, which is the only depth satisfying condition  $(D_6)$ , provide the worst performance. This happens because the two groups have most of the points quite overlapping, and also the corresponding hemispheres with the smallest empirical probability do overlap substantially. As a result, points lying in the intersection of these hemispheres cannot be well classified.

## 5 Final remarks

We have discussed the use of angular depth functions within a classification setting. Our viewpoint was mainly theoretical, but with a perspective of potential applications in classification and data analysis. With that aim, desirable properties that angular depth functions should enjoy were first provided. The most commonly used angular depths (angular Mahalanobis, distance-based, simplicial, halfspace) were compared w.r.t. these properties. It turned out that the angular halfspace depth is the only function that enjoys all the listed properties, including the condition of its central regions being spherical convex.

To illustrate our main findings, we used several small simulation exercises and a real dataset. We highlighted that (i) if the optimal discriminating boundary is not a great circle, depths not satisfying the non-rigidity property are not apt to be used for classification purposes; and (ii) if the data structure is ellipse-like, depths satisfying the quasi-concavity property should be adopted; but (iii) for data that lie in the whole sphere, depths satisfying the quasi-concavity property may perform poorly.

From a data analysis point of view, the following guidelines arise. First, avoid using the cosine distance and the Mahalanobis angular depths, unless the data comes from a rotationally symmetric distribution with a unique Fréchet median. Second, choose the angular halfspace depth if the geometric structure of the distribution is ellipse-like. Third, do not use the plain angular halfspace depth if the data are distributed sparsely on the whole surface of the sphere.

Our findings suggest that the angular halfspace depth deserves further attention. Certainly, in classification tasks, the angular halfspace depth is a powerful tool if data from each group lie in a hemisphere. For general directional data, however, the issue of constancy of the angular halfspace depth on hemispheres of minimum probability must be addressed. In analogy with the issues related to the halfspace depth in linear spaces, solutions that borrow ideas from the bagdistance (Hubert et al. 2015, 2017) or the illumination depth (Nagy and Dvořák 2021) can be investigated. We aim to report on those advances elsewhere.

**Acknowledgements** The work of S. Nagy was supported by Czech Science Foundation (EXPRO project n. 19-28231X). The work of H. Demni and G. C. Porzio was partially funded by the BiBiNet project (grant H35F21000430002) within the POR-Lazio FESR 2014–2020.

## Appendix A: Gnomonic projection of the sphere and angular depths

### A.1. Gnomonic projection and its basic properties

A crucial mathematical tool that allows us to connect the angular depths in the unit sphere  $\mathbb{S}^{d-1}$  with usual depths in the linear space  $\mathbb{R}^{d-1}$  is the *gnomonic projection* and its inverse. The gnomonic projection takes points from the unit sphere  $\mathbb{S}^{d-1}$ , and maps them into the hyperplane  $H = \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_d = 1\}$  that is tangent to  $\mathbb{S}^{d-1}$  at its “northern pole”  $e_d = (0, \dots, 0, 1)$ . Formally speaking, writing  $\mathbb{S}_0^{d-1} = \{(x_1, \dots, x_d) \in \mathbb{S}^{d-1} : x_d = 0\}$  for the “equator” of the unit sphere, the gnomonic projection is defined by

$$\Pi : \mathbb{S}^{d-1} \setminus \mathbb{S}_0^{d-1} \rightarrow H : x = (x_1, \dots, x_d) \mapsto x/x_d.$$

The map  $\Pi$  is not defined for the points on the equator  $\mathbb{S}_0^{d-1}$  of  $\mathbb{S}^{d-1}$ . If necessary, one could identify those points with their unit vectors, and expand the mapping  $\Pi$  by projecting  $x \in \mathbb{S}_0^{d-1}$  into a “point in infinity” of  $H$  in the direction of the vector  $x$ . Mathematically, that construction corresponds to extending the mapping  $\Pi$  from  $\mathbb{S}^{d-1}$  to the projective space associated with  $H$  (Berger (2010), Sect. I.5). For our purposes, this however turns out not to be necessary.

The gnomonic projection is a bijection, that is an invertible, one-to-one mapping, between the “northern” hemisphere  $\mathbb{S}_+^{d-1} = \{(x_1, \dots, x_d) \in \mathbb{S}^{d-1} : x_d > 0\}$  and  $H$ . Also, it is a bijection between the “southern” hemisphere  $\mathbb{S}_-^{d-1} = -\mathbb{S}_+^{d-1}$  and  $H$ , meaning that altogether it is a double cover of  $H$ . When restricted to  $\mathbb{S}_+^{d-1}$  only, the inverse of this map taking any point in  $H$  back into  $\mathbb{S}_+^{d-1}$  is denoted by  $\Pi^{-1}$ . Of course, it is defined by

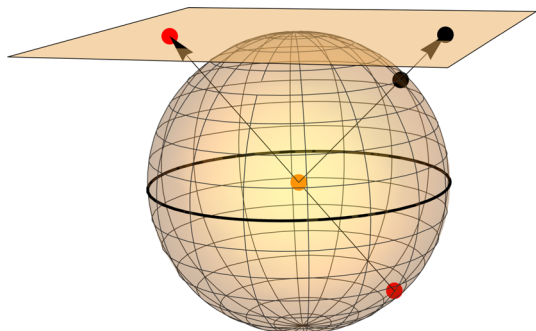
$$\Pi^{-1} : H \rightarrow \mathbb{S}_+^{d-1} : x \mapsto x/\|x\|.$$

An interesting feature of the map  $\Pi$  is that

$$\Pi \text{ maps great circles in } \mathbb{S}^{d-1} \setminus \mathbb{S}_0^{d-1} \text{ into straight lines in } H. \tag{19}$$

This can be seen directly by considering a great circle as an intersection of a (two-dimensional) plane  $L$  passing through the origin in  $\mathbb{R}^d$  with the sphere  $\mathbb{S}^{d-1}$ . Since  $L$  contains the origin, its gnomonic projection is exactly the intersection  $L \cap H$ , which is a straight line in  $H$ . This setup is visualized in Fig. 9. Another important consequence of (19) is that halfspaces in  $\mathbb{R}^d$  whose boundary passes through the origin in  $\mathbb{R}^d$  are mapped via  $\Pi$  into halfspaces in  $H$ , when identified with the linear space  $\mathbb{R}^{d-1}$ , and that spherical convex sets in  $\mathbb{S}_+^{d-1}$  map into convex subsets of  $H$ . All these results are direct consequences of (19).

**Fig. 9** The gnomonic projection of the sphere  $\mathbb{S}^2$ . Points from the northern hemisphere  $x \in \mathbb{S}_+^2$  (black point) map to  $H$  along the direction of the unit vector  $x$  from the origin (orange point); points from the southern hemisphere  $y \in \mathbb{S}_-^2$  (red point) project in the direction of the antipodal point  $-y \in \mathbb{S}_+^2$  (color figure online)



## A.2. Gnomonic projection and angular depths

In Dyckerhoff and Nagy (2023), the gnomonic projection is used to devise a fast computation algorithm for the angular halfspace depth. It turns out that most angular depths relate very closely with usual depths of points in the linear space  $H$  (identified with  $\mathbb{R}^{d-1}$ ), when the measure  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  is properly mapped into  $H$  via the gnomonic projection. For instance, if  $P$  is concentrated only in the open hemisphere  $\mathbb{S}_+^{d-1}$  (where  $\Pi$  is invertible), the angular halfspace (or simplicial) depth of any  $x \in \mathbb{S}_+^{d-1}$  is equivalent with the usual halfspace (or simplicial) depth of  $\Pi(x)$  w.r.t. the measure of  $P$  projected into  $H$  via  $\Pi$ . This can be seen directly from the property (19), since halfspaces passing through the origin in  $\mathbb{R}^d$  map via  $\Pi$  into halfspaces in  $H$ , and spherical convex hulls on  $\mathbb{S}_+^{d-1}$  map via  $\Pi$  into (linear) convex hulls in  $H$ . Thus, for e.g. the angular simplicial depth of any  $x \in \mathbb{S}_+^{d-1}$  w.r.t.  $P \in \mathcal{P}(\mathbb{S}_+^{d-1})$  concentrated in the northern hemisphere we can write

$$asD(x; P) = sD(\Pi(x); P_\Pi),$$

where  $P_\Pi$  stands for the distribution  $P$  projected into  $H$  via  $\Pi$ , and  $sD(y; Q)$  is the usual simplicial depth of  $y \in \mathbb{R}^{d-1}$  w.r.t.  $Q \in \mathcal{P}(\mathbb{R}^{d-1})$  (Liu 1990).

More generally, observe that thanks to the property (19), it is very useful to visualize angular depths  $aD$  for distributions  $P \in \mathcal{P}(\mathbb{S}_+^2)$  supported in the northern hemisphere of  $\mathbb{S}^2$  by displaying

$$\mathbb{R}^2 \rightarrow [0, \infty) : y \mapsto aD(y; P_\Pi) \quad (20)$$

instead of the more complicated function  $x \mapsto aD(x; P)$ . In (20), of course, we interpret  $y$  as  $\Pi(x)$  for  $x \in \mathbb{S}_+^2$ . Using the contour plot of (20) we can, for example, detect departures of the depth  $aD(\cdot; P)$  from (star) convexity (in  $\mathbb{S}^{d-1}$ ), observe multimodal features of the angular depth surface, or assess rotational symmetry around the direction given by the northern pole. This tool is used in Sect. 3 to infer the contours of angular depths in  $\mathbb{S}^2$ .

## Appendix B: Proofs of theoretical results

### B.1. Proof of Theorem 1

Take any directional median  $\mu$  that satisfies (3); the existence of such a point  $\mu$  follows from (D<sub>5</sub>). By (D<sub>4</sub>) we then have  $aD(-\mu; P) = \inf_{x \in \mathbb{S}^{d-1}} aD(x; P)$ . Consider the orthogonal transform given by the matrix  $O = -I_d \in \mathbb{R}^{d \times d}$ , for  $I_d$  the identity matrix, and apply the induced rotation in (D<sub>1</sub>). We obtain

$$aD(x; P) = aD(-x; P_{-X}) = aD(-x; P) \quad \text{for all } x \in \mathbb{S}^{d-1},$$

where we used the fact that  $P_{-X} = P$  by the assumption of central symmetry of  $P$ . In particular, applying the previous formula to  $x = \mu$ , we get

$$\sup_{x \in \mathbb{S}^{d-1}} aD(x; P) = aD(\mu; P) = aD(-\mu; P) = \inf_{x \in \mathbb{S}^{d-1}} aD(x; P),$$

and necessarily  $aD(\cdot; P)$  must be constant over  $\mathbb{S}^{d-1}$ .

### B.2. Proof of Theorem 2

By rotational symmetry of  $P$  we know that for any  $O \in \mathbb{R}^{d \times d}$  orthogonal that fixes  $\mu$  we have  $P = P_{OX}$  in the notation from (D<sub>1</sub>). Condition (D<sub>1</sub>) then gives

$$aD(x; P) = aD(Ox; P_{OX}) = aD(Ox; P) \quad \text{for all } x \in \mathbb{S}^{d-1},$$

and necessarily  $aD(\cdot; P)$  is constant on each set

$$\left\{ Ox : O \in \mathbb{R}^{d \times d} \text{ orthogonal such that } O\mu = \mu \right\} \subset \mathbb{S}^{d-1}.$$

This set is precisely the collection of those  $y \in \mathbb{S}^{d-1}$  such that  $\langle y, \mu \rangle = \langle x, \mu \rangle$ . Indeed, we have

$$\langle Ox, \mu \rangle = (Ox)^\top \mu = x^\top O^\top \mu = x^\top O^{-1} \mu = \langle x, O^{-1} \mu \rangle = \langle x, \mu \rangle,$$

where we used that for  $O$  orthogonal we have  $O^\top = O^{-1}$  and that  $O\mu = \mu$  implies  $O^{-1}\mu = \mu$ . For the other inclusion, we can assume without loss of generality that  $\mu = (1, 0, \dots, 0) \in \mathbb{S}^{d-1}$  is the first canonical vector; otherwise, we first rotate the sphere so that  $\mu$  maps to that vector. Suppose that  $\langle y, \mu \rangle = \langle x, \mu \rangle$ . Then both  $x, y$  are contained in the  $(d - 2)$ -dimensional sphere  $S$  obtained as an intersection of  $\mathbb{S}^{d-1}$  and the hyperplane  $H = \{z \in \mathbb{R}^d : \langle z, \mu \rangle = \langle x, \mu \rangle\}$  orthogonal to  $\mu$ . Restricting to the  $(d - 1)$ -dimensional affine space  $H$  and canonically identifying  $H$  with  $\mathbb{R}^{d-1}$ , inside the sphere  $S$  we can find an orthogonal matrix  $O' \in \mathbb{R}^{(d-1) \times (d-1)}$  mapping  $x$  to  $y$ . Extending  $O'$  to a  $d \times d$  matrix  $O$  by appending it in its first row and first column with the vector  $(1, 0, \dots, 0) \in \mathbb{R}^d$  gives the desired orthogonal transform  $O$  mapping  $x$  to  $y$  that fixes  $\mu$ .

To conclude the proof, it remains to define  $\xi(t) = aD(x; P)$  for any  $x \in \mathbb{S}^{d-1}$  such that  $\langle x, \mu \rangle = t \in [-1, 1]$ . By our argumentation above, this is a correct definition and  $aD(x; P) = \xi(\langle x, \mu \rangle)$  for all  $x \in \mathbb{S}^{d-1}$ .

### B.3. Proof of Theorem 3

The central regions (2) corresponding to  $aD$  are clearly nested in the sense that if  $0 \leq \alpha_1 \leq \alpha_2$  then for any  $P \in \mathcal{P}(\mathbb{S}^{d-1})$  we have  $aD_{\alpha_2}(P) \subseteq aD_{\alpha_1}(P)$ . At the same time,  $aD_0(P) = \mathbb{S}^{d-1}$ . By assumption (D<sub>6</sub>) we have that each  $aD_\alpha(P)$  is spherical

convex, which must be either the whole sphere  $\mathbb{S}^{d-1}$ , or a set contained in some closed hemisphere in  $\mathbb{S}^{d-1}$ . By the nestedness of the central regions, we can define  $\alpha_0 = \inf\{\alpha \geq 0 : aD_\alpha(P) \text{ is contained in a hemisphere}\}$ . We consider two cases: either (C<sub>1</sub>)  $aD_{\alpha_0}(P)$  is contained in a hemisphere  $S$ ; or (C<sub>2</sub>)  $aD_{\alpha_0}(P)$  is not contained in any hemisphere in  $\mathbb{S}^{d-1}$ .

In case (C<sub>1</sub>) choose  $u \in \mathbb{S}^{d-1}$  to be the direction that gives

$$S = \{y \in \mathbb{S}^{d-1} : \langle y, u \rangle \geq 0\}. \tag{21}$$

In case (C<sub>2</sub>) necessarily  $aD_{\alpha_0}(P) = \mathbb{S}^{d-1}$ . Then we take  $\alpha_n = \alpha_0 + 1/n$ , and observe that for each  $n = 1, 2, \dots$  we have that  $aD_{\alpha_n}(P)$  is contained in a hemisphere  $S_n \subset \mathbb{S}^{d-1}$ . The hemispheres  $S_n$  do not have to be the same, but if  $m \geq n$ , then  $aD_{\alpha_n}(P) \subseteq aD_{\alpha_m}(P) \subseteq S_m$ , meaning that each  $S_m$  contains all central regions  $aD_{\alpha_n}(P)$  at levels  $n \leq m$ . To each  $S_n$  find a direction  $u_n \in \mathbb{S}^{d-1}$  such that (21) holds true with  $S$  replaced by  $S_n$  and  $u$  replaced by  $u_n$ . The sequence  $\{u_n\}_{n=1}^\infty$  must have a limit point  $u \in \mathbb{S}^{d-1}$ ; define then  $S$  directly by (21) using this direction  $u$ . Necessarily, we obtain that all  $aD_\alpha(P)$  that are contained in a hemisphere must be contained in  $S$  in both cases (C<sub>1</sub>) and (C<sub>2</sub>) above.

Take now any two points  $x, y \in \mathbb{S}^{d-1} \setminus S$ , and suppose that  $\alpha' = aD(x; P) > aD(y; P)$ . Consider the central region  $aD_{\alpha'}(P)$  that contains  $x$ . Since  $aD_{\alpha'}(P)$  does not contain  $y$ , this region is not the whole sphere  $\mathbb{S}^{d-1}$ . But,  $aD_{\alpha'}(P)$  must be spherical convex, so if it is not  $\mathbb{S}^{d-1}$ , it must be contained in a hemisphere, in which case  $\alpha' \geq \alpha_0$  by the definition of  $\alpha_0$ . We get  $\alpha' = \alpha_0$ . Using  $x \notin S$  again, necessarily case (C<sub>2</sub>) above must hold true (that is,  $aD_{\alpha_0}(P)$  is not contained in  $S$ ). But then, there must exist a hemisphere  $S' \subset \mathbb{S}^{d-1}$  such that  $aD_{\alpha_0}(P)$  is contained in  $S'$ , which directly contradicts case (C<sub>2</sub>) as in that situation  $aD_{\alpha_0}(P)$  is contained in  $S'$ . Altogether, we have obtained that for all  $x \notin S$ , we must have  $aD(x; P) = c$  for some constant  $c \in [0, \alpha_0]$ .

Under the additional condition (D<sub>4</sub>), clearly  $c = \inf_{x \in \mathbb{S}^{d-1}} aD(x; P)$ .

### B.4. Proof of Theorem 4

We use the gnomonic projection, described in detail in Appendix A, to express the exact arc distance depth of  $P$  as a depth on the tangent line  $H$  of  $\mathbb{S}^1$  that touches the unit circle at its northern pole  $(0, 1) \in \mathbb{S}^1$ . For a precise definition of  $H$  and the properties of the gnomonic projection please refer to Appendix A. Let  $Q_+$  be the pushforward measure of  $P$  restricted to  $\mathbb{S}_+^1$  via the gnomonic projection  $\Pi$ , and analogously let  $Q_-$  be the pushforward of  $P$  restricted to  $\mathbb{S}_-^1$  (notations such as  $\mathbb{S}_+^1$  and  $\mathbb{S}_-^1$  come from Appendix A). We suppose that the equator  $\mathbb{S}_0^1$  has null  $P$ -mass. This is made without loss of generality, as by condition (D<sub>1</sub>) we are allowed to rotate  $P$  to a position where  $P(\mathbb{S}_0^1) = 0$ . Denote by  $F_+$  and  $F_-$  the cumulative distribution functions of  $Q_+$  and  $Q_-$ , respectively, when  $H$  is identified with  $\mathbb{R}$ . For  $x \in \mathbb{S}_+^1$  we can write



$$\begin{aligned}
 \pi - aD_{arc}(x; P) &= \int_{\mathbb{S}^1} \arccos(\langle x, y \rangle) dP(y) \\
 &= \int_{\mathbb{S}_+^1} \arccos(\langle x, y \rangle) dP(y) + \int_{\mathbb{S}_-^1} \arccos(\langle x, y \rangle) dP(y) \\
 &= \int_{\mathbb{S}_+^1} \arccos(\langle x, y \rangle) dP(y) + \int_{\mathbb{S}_-^1} (\pi - \arccos(-\langle x, y \rangle)) dP(y) \\
 &= \pi P(\mathbb{S}_-^1) + \int_{\mathbb{S}_+^1} \arccos(\langle x, y \rangle) dP(y) - \int_{\mathbb{S}_+^1} \arccos(\langle x, y \rangle) dP(-y) \\
 &= \pi P(\mathbb{S}_-^1) + \int_{\mathbb{S}_+^1} \arccos(\langle x, y \rangle) d(P(y) - P(-y)) \\
 &= \pi P(\mathbb{S}_-^1) + \int_H \arccos(\langle \Pi^{-1}(\Pi(x)), \Pi^{-1}(\tilde{y}) \rangle) d(Q_+(\tilde{y}) - Q_-(\tilde{y})) \\
 &= \pi P(\mathbb{S}_-^1) + \int_H \arccos\left(\frac{\langle \tilde{x}, \tilde{y} \rangle}{\|\tilde{x}\| \|\tilde{y}\|}\right) d(Q_+(\tilde{y}) - Q_-(\tilde{y})) \\
 &= \pi P(\mathbb{S}_-^1) + \int_{\mathbb{R}} \arccos\left(\frac{1 + xy}{\sqrt{(1+x^2)(1+y^2)}}\right) d(F_+(y) - F_-(y)).
 \end{aligned}$$

We used that  $\arccos(t) + \arccos(-t) = \pi$ , a change of variables  $y \mapsto -y$  in the integral over  $\mathbb{S}_-^1$ , and finally the change of variables for the integral with the pushforward measures  $Q_+$  and  $Q_-$ . In the final equalities we denoted  $\tilde{x} = \Pi(x) = (x, 1) \in H$  and  $\tilde{y} = (y, 1) \in H$ . We obtain an explicit form for the arc distance depth in terms of the density of the mass of  $P$  projected onto  $H$ .

To find local extremes of the arc distance depth, we take a derivative of the final expression above w.r.t.  $x \in \mathbb{R}$ . Given that the function  $\arccos$  is bounded and differentiable, we can exchange the derivative and the integral, and after a simplification and rearrangement we obtain

$$\frac{\partial aD_{arc}(x; P)}{\partial x} = -\frac{1}{1+x^2} \int_{\mathbb{R}} \text{sign}(x-y) d(F_+(y) - F_-(y)).$$

In particular, the depth  $aD_{arc}(\cdot; P)$  has a vanishing derivative at  $x$  if and only if

$$\int_{\mathbb{R}} \text{sign}(x-y) dF_+(y) = \int_{\mathbb{R}} \text{sign}(x-y) dF_-(y),$$

or equivalently if

$$\int_{(-\infty, x)} dF_+(y) - \int_{(x, \infty)} dF_+(y) = \int_{(-\infty, x)} dF_-(y) - \int_{(x, \infty)} dF_-(y).$$

Suppose now that  $Q_+(H) = Q_-(H) = 1/2$ , and that  $P$  does not have atoms in  $\mathbb{S}^1$ . Then the preceding formula simplifies further to

$$F_+(x) = F_-(x).$$

In particular, our computation shows that under our assumptions on  $P$ , whenever the distribution functions  $F_+$  and  $F_-$  cross in  $\mathbb{R}$ , the arc distance depth  $aD_{arc}(\cdot; P)$  possesses a critical point.

**B.5. Analysis of Example 1**

In the notations from the proof of Theorem 4, we see that  $P \in \mathcal{P}(\mathbb{S}^1)$  is the inverse gnomonic projection of  $Q$ . The corresponding distribution  $Q_+$  from the proof of Theorem 4 is uniform on the interval  $[-K, K]$ , and  $Q_-$  is uniform on the  $m$  points  $\{-K - \ell/2 + j\ell\}_{j=1}^m$ . The distribution functions  $F_+$  and  $F_-$  of  $Q_+$  and  $Q_-$ , respectively, are

$$F_+(y) = (y + K)/(2K) \quad \text{for } y \in [-K, K],$$

$$F_-(y) = j/m \quad \text{for } y \in [y_j, y_{j+1}),$$

with the notation  $y_{m+1} = \infty$ . These two distribution functions intersect in  $m$  points  $\{y_j\}_{j=1}^m$  inside the interval  $[-K, K]$ , see the upper panel of Fig. 1.

The circular distribution  $P \in \mathcal{P}(\mathbb{S}^1)$  is defined on the upper half-circle as the inverse gnomonic projection of  $Q_+$  when considered inside the straight line  $L_+$ ; on the lower half-circle it is the inverse gnomonic projection of  $Q_-$  inside the antipodally reflected line  $L_-$ . Both  $Q_+$  and  $Q_-$  contribute by mass  $1/2$  to the final measure  $P$ . According to our discussion from the proof of Theorem 4, the arc distance depth of such circular distribution  $P$  possesses at least  $m$  local extremes.

**B.6. Proof of Theorem 5**

For  $X \sim P$  and  $x \in \mathbb{S}^{d-1}$  we have

$$aD_\delta(-x; P) = \delta(-1) - E\delta(-\langle x, X \rangle) = \delta(-1) - c + E\delta(\langle x, X \rangle)$$

$$= 2\delta(-1) - c - aD_\delta(x; P).$$

Necessarily,  $aD_\delta(x; P)$  is maximized if and only if  $aD_\delta(-x; P)$  takes the minimum value over  $\mathbb{S}^{d-1}$ , and (D<sub>4</sub>) is satisfied.

**B.7. Proof of Theorem 6**

Take any sequence  $\{x_n\}_{n=1}^\infty \subset \mathbb{S}^{d-1}$  that converges to  $x \in \mathbb{S}^{d-1}$ . Since  $x_n \rightarrow x$  as  $n \rightarrow \infty$ , we can define a sequence of orthogonal matrices  $\{O_n\}_{n=1}^\infty \subset \mathbb{R}^{d \times d}$  such that  $O_n x_n = x$  for all  $n$ , and at the same time  $O_n$  converges to the identity matrix  $I_d \in \mathbb{R}^{d \times d}$  in, say, the Frobenius matrix norm (the choice of the matrix norm is irrelevant in this case, as in finite-dimensional spaces all metrics are equivalent). Defining  $P_n \in \mathcal{P}(\mathbb{S}^{d-1})$  as the pushforward measure of  $P$  under the mapping  $\mathbb{S}^{d-1} \rightarrow \mathbb{S}^{d-1} : y \mapsto O_n y$ , we obtain that the sequence  $\{P_n\}_{n=1}^\infty \subset \mathcal{P}(\mathbb{S}^{d-1})$  converges weakly to  $P$ . As observed in Liu and Singh (1992), the depth  $aSD$  satisfies (D<sub>1</sub>). Due to that, we

have that  $asD(x_n; P) = asD(x; P_n)$  for all  $n = 1, 2, \dots$ . The angular simplicial depth of  $x \in \mathbb{S}^{d-1}$  w.r.t. any measure  $Q \in \mathcal{P}(\mathbb{S}^{d-1})$  can be written as  $asD(x; Q) = Q^d(S_x)$ , where  $Q^d \in \mathcal{P}((\mathbb{S}^{d-1})^d)$  is the product of  $d$  copies of measure  $Q$ , and  $S_x \subset (\mathbb{S}^{d-1})^d$  is a closed set (e.g., Nagy et al. (2016), proof of Theorem A.6). The portmanteau theorem (Dudley (2002), Theorem 11.1.1) applied to the sequence  $\{P_n\}_{n=1}^\infty$  and the closed set  $S_x$  then guarantees that

$$\begin{aligned} \limsup_{n \rightarrow \infty} ahD(x_n; P) &= \limsup_{n \rightarrow \infty} ahD(x; P_n) \\ &= \limsup_{n \rightarrow \infty} P_n^d(S_x) \leq P^d(S_x) = ahD(x; P), \end{aligned}$$

as we wanted to show.

## References

- Agostinelli C, Romanazzi M (2013) Nonparametric analysis of directional data based on data depth. *Environ Ecol Stat* 20(2):253–270
- Arnone E, Ferraccioli F, Pigolotti C, Sangalli LM (2022) A roughness penalty approach to estimate densities over two-dimensional manifolds. *Comput Stat Data Anal* 174:107527
- Berger M (2010) *Geometry revealed*. Springer, Heidelberg
- Besau F, Werner EM (2016) The spherical convex floating body. *Adv Math* 301:867–901
- Bry X, Cucala L (2022) A von Mises-Fisher mixture model for clustering numerical and categorical variables. *Adv Data Anal Classif* 16(2):429–455
- Cuesta-Albertos JA, Febrero-Bande M, Oviedo de la Fuente M (2017) The  $DD^G$ -classifier in the functional setting. *Test* 26(1):119–142
- Demni H (2021) Directional supervised learning through depth functions: an application to ECG waves analysis. In: Balzano S, Porzio GC, Salvatore R, Vistocco D, Vichi M (eds) *Stat Learn Model Data Anal*. Springer International Publishing, Cham, pp 79–87
- Demni H, Messaoud A, Porzio GC (2021) Distance-based directional depth classifiers: a robustness study. *Commun Stat Simul Comput*. <https://doi.org/10.1080/03610918.2021.1996603>
- Demni H, Porzio GC (2021) Directional  $DD$ -classifiers under non-rotational symmetry. In 2021 IEEE International conference on multisensor fusion and integration for intelligent systems (MFI), pp 1–6. IEEE
- Donoho DL, Gasko M (1992) Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann Stat* 20(4):1803–1827
- Dudley RM (2002) *Real analysis and probability*, volume 74 of *Cambridge studies in advanced mathematics*. Cambridge University Press, Cambridge. Revised reprint of the 1989 original
- Dyckerhoff R, Nagy S (2023) Exact computation of angular halfspace depth. Under review
- Figueiredo A (2009) Discriminant analysis for the von Mises-Fisher distribution. *Commun Stat Simulation Comput* 38(9):1991–2003
- Fisher NI (1985) Spherical medians. *J Roy Stat Soc Ser B* 47(2):342–348
- Hubert M, Rousseeuw P, Segaert P (2017) Multivariate and functional classification using depth and distance. *Adv Data Anal Classif* 11(3):445–466
- Hubert M, Rousseeuw PJ, Segaert P (2015) Multivariate functional outlier detection. *Stat Methods Appl* 24(2):177–202
- Jana N, Dey S (2021) Classification of observations into von Mises-Fisher populations with unknown parameters. *Commun Stat Simul Comput*. <https://doi.org/10.1080/03610918.2021.1962347>
- Kent JT, Mardia KV (2013) Discrimination for spherical data. In: Mardia KV, Gusnanto A, Riley AD, Voss J (eds) *LASR 2013—statistical models and methods for non-Euclidean data with current scientific applications*, pp 71–74
- Konen D (2022) Topics in multivariate spatial quantiles. PhD thesis, ULB
- Ley C, Sabbah C, Verdebout T (2014) A new concept of quantiles for directional data and the angular Mahalanobis depth. *Electron J Stat* 8(1):795–816

- Ley C, Verdebout T (2017) Modern directional statistics. CRC Press, Boca Raton, FL, Chapman & Hall/CRC Interdisciplinary statistics series
- Li J, Cuesta-Albertos JA, Liu RY (2012) *DD*-classifier: nonparametric classification procedure based on *DD*-plot. *J Am Stat Assoc* 107(498):737–753
- Liu RY (1990) On a notion of data depth based on random simplices. *Ann Stat* 18(1):405–414
- Liu RY, Parelius JM, Singh K (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann Stat* 27(3):783–858
- Liu RY, Singh K (1992) Ordering directional data: concepts of data depth on circles and spheres. *Ann Stat* 20(3):1468–1484
- Mardia KV (1972) Statistics of directional data. *Probability and mathematical statistics*, no 13. Academic Press, London, New York
- Mosler K, Mozharovskyi P (2017) Fast *DD*-classification of functional data. *Stat Papers* 58(4):1055–1089
- Mosler K, Mozharovskyi P (2022) Choosing among notions of multivariate depth statistics. *Stat Sci* 37(3):348–368
- Nagy S (2023) Simplicial depth and its median: selected properties and limitations. *Stat Anal Data Min* 16(4):374–390
- Nagy S, Dvořák J (2021) Illumination depth. *J Comput Graph Statist* 30(1):78–90
- Nagy S, Gijbels I, Omelka M, Hlubinka D (2016) Integrated depth for functional data: statistical properties and consistency. *ESAIM Prob Stat* 20:95–130
- Nagy S, Laketa P (2023) Angular halfspace depth: Theoretical properties. Under review
- Nagy S, Schütt C, Werner EM (2019) Halfspace depth and floating body. *Stat Surv* 13:52–118
- Pandolfo G, D'Ambrosio A (2021) Depth-based classification of directional data. *Exp Syst Appl* 169:1144433
- Pandolfo G, Paindaveine D, Porzio GC (2018) Distance-based depths for directional data. *Can J Stat* 46(4):593–609
- Rousseeuw PJ, Ruts I (1999) The depth function of a population distribution. *Metrika* 49(3):213–244
- Rousseeuw PJ, Struyf A (2004) Characterizing angular symmetry and regression symmetry. *J Stat Plan Inference* 122(1–2):161–173
- Saavedra-Nieves P, Crujeiras RM (2022) Nonparametric estimation of directional highest density regions. *Adv Data Anal Classif* 16(3):761–796
- Salah A, Nadif M (2019) Directional co-clustering. *Adv Data Anal Classif* 13(3):591–620
- Scealy JL, Wood ATA (2019) Scaled von Mises-Fisher distributions and regression models for paleomagnetic directional data. *J Am Stat Assoc* 114(528):1547–1560
- Serfling R (2006) Depth functions in nonparametric multivariate inference. In: *Data depth: robust multivariate analysis, computational geometry and applications*, vol. 72 of *DIMACS Ser Discrete Math Theoret Comput Sci*, pages 1–16. Amer. Math. Soc., Providence, RI
- Serfling R (2010) Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J Nonparametr Stat* 22(7):915–936
- Small CG (1987) Measures of centrality for multivariate and directional distributions. *Can J Stat* 15(1):31–39
- Tian Y, Gel YR (2019) Fusing data depth with complex networks: community detection with prior information. *Comput Stat Data Anal* 139:99–116
- Tsagris M, Alenazi A (2019) Comparison of discriminant analysis methods on the sphere. *Commun Stat Case Stud Data Anal Appl* 5(4):467–491
- Tukey JW (1975) Mathematics and the picturing of data. In: *Proceedings of the international congress of mathematicians (Vancouver, B. C., 1974)*, vol 2, pp 523–531. *Canad. Math. Congress*, Montreal, Que
- Zuo Y, Serfling R (2000) General notions of statistical depth function. *Ann Stat* 28(2):461–482
- Zuo Y, Serfling R (2000) On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J Stat Plan Inference* 84(1–2):55–79

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.