Distance-based Depths for Directional Data

Giuseppe Pandolfo[*], Davy Paindaveine[†] and Giovanni Porzio[‡]

University of Naples Federico II[*], Université libre de Bruxelles[†],

and University of Cassino and Southern Lazio[‡]

Abstract

Directional data are constrained to lie on the unit sphere of $\mathbb{R}^q$ for some $q \geq 2$. To address the lack of a natural ordering for such data, depth functions have been defined on spheres. However, the depths available either lack flexibility or are so computationally expensive that they can only be used for very small dimensions q. In this work, we improve on this by introducing a class of distance-based depths for directional data. Irrespective of the distance adopted, these depths can easily be computed in high dimensions too. We derive the main structural properties of the proposed depths and study how they depend on the distance used. We discuss the asymptotic and robustness properties of the corresponding deepest points. We show the practical relevance of the proposed depths in two applications, related to (i) spherical location estimation and (ii) supervised classification. For both problems, we show through simulation studies that distance-based depths have strong advantages over their competitors.

1 Introduction

Directional data analysis is relevant when the sample space is the unit hypersphere $\mathscr{S}^{q-1}$
$:= \left\{ x \in \mathbb{R}^q : x^T x = 1 \right\}$ in $\mathbb{R}^q$, which occurs when observations are directions, axes, rotations, or cyclic events. Applications arise in numerous fields, including astronomy, earth sciences, biology, meteorology and political science; see Gill and Hangartner (2010) for an exemple in the latter field. Directional data analysis can also be exploited to study patterns of unit vectors in $\mathbb{R}^q$, such as those encountered in text mining (Hornik et al., 2012).

Statistically, analyzing and describing directional data requires tackling some interesting problems associated with the lack of a reference direction and with a sense of rotation not uniquely defined. Another important issue when dealing with such data is the lack of a natural ordering, which generates a special interest in depth functions on the sphere. Parallel to their role in the usual Euclidean

case, directional depths are to measure the degree of centrality of a given spherical location with respect to a distribution on the sphere and to provide a center-outward ordering of spherical locations; see Agostinelli and Romanazzi (2013b).

Depth concepts for directional data were first considered by Small (1987) and Liu and Singh (1992). Following the pioneering work of Small (1987), Liu and Singh (1992) popularized the concept of angular Tukey depth (ATD), which is the directional analog of the celebrated halfspace depth (Tukey, 1975). The same paper introduced two further depths for directional data, namely the angular simplicial depth (ASD), which is the directional version of the simplicial depth from Liu (1990), and the arc distance depth (ADD), which is based on the concept of arc length distance.

Unlike the ADD, the ATD and ASD have been studied and used in the literature. For instance, Rousseeuw and Struyf (2004) investigated some of the properties of the ATD, while Agostinelli and Romanazzi (2013b) considered some of the possible applications of the ASD and ATD. R packages are also available for these depths: the package depth (Genest et al., 2012) allows to compute ATD values for $q = 2$ or 3, whereas the package localdepth (Agostinelli and Romanazzi, 2013a) implements specific functions for the evaluation of the ATD for $q = 2$, and of the ASD for an arbitrary $q \geq 2$.

The main drawback of both the ASD and ATD is the computational effort they require, especially for higher dimensions q. The angular Mahalanobis depth of Ley et al. (2014), that is based on a concept of directional quantiles, is computationally much more affordable, but suffers from other disadvantages: it requires the preliminary choice of a spherical location functional and it is less flexible than the ASD/ATD in the sense that it produces rotationally symmetric depth contours, even if the underlying distribution is not rotationally symmetric.

On the one hand, depth functions for directional data are useful, yet on the other hand, they lack flexibility (and depend on some user's choice) or are computationally too demanding. In order to improve on this, this work introduces a new class of directional depth functions that is based on spherical distances and contains the ADD as a particular case. These depth functions are computationally feasible even in high dimensions and are generally more flexible. Distance-based directional depths show several other advantages over their ASD/ATD competitors: they take positive values everywhere on $\mathscr{S}^{q-1}$ (but in the uninteresting case of a point mass distribution), whereas the ASD/ATD can take zero values (which is undesirable when performing supervised classification). Further advantages of the proposed distance-based depths is that they typically do not provide ties in the sample case (whereas ties are unavoidable for the ASD/ATD, due to their step function nature) and that they do not require any assumption on the underlying distribution (unlike the angular Mahalanobis depth that, when based on the spherical mean, is not defined for zero-mean distributions).

The paper is organized as follows. In Section 2, we introduce the proposed class of distance-based depth functions for directional data, and we consider three particular cases, namely the arc distance depth (ADD), the cosine distance depth (CDD) and the chord distance depth (ChDD). In Section 3, we derive the main structural properties of the proposed depths and study how they depend on the distance used. In Section 4, we compare the various depths considered for several empirical distributions on the circle ($q = 2$), which also allows us to illustrate the theoretical results of Section 3. In Section 5, we discuss the asymptotic and robustness properties of the proposed concepts. In Section 6, we show the practical relevance of the distance-based depths in two applications, related to (i) spherical location estimation (Section 6.1) and (ii) supervised classification (Section 6.2). For both problems, we perform simulations that show the advantages of the proposed depths over their competitors. Final comments are provided in Section 7. Finally, an appendix collects technical proofs.

2 Distance-based depths for directional data

In Definition 1 below, we introduce a class of depths on the unit sphere $\mathscr{S}^{q-1}$. A particular member of this class will be obtained by fixing a particular (bounded) distance $d(\cdot,\cdot)$ on $\mathscr{S}^{q-1}$. For such a distance, $d^{\sup} := \sup\{d(\boldsymbol{\theta},\boldsymbol{\psi}) : \boldsymbol{\theta},\boldsymbol{\psi} \in \mathscr{S}^{q-1}\}$ will throughout denote the upper bound of the distance between any two points on $\mathscr{S}^{q-1}$.

Definition 1 (Directional distance-based depths) Let $d(\cdot,\cdot)$ be a bounded distance on $\mathscr{S}^{q-1}$ and H be a distribution on $\mathscr{S}^{q-1}$. Then the directional d-depth of $\boldsymbol{\theta}(\in \mathscr{S}^{q-1})$ with respect to H is

$$D_d(\boldsymbol{\theta}, H) := d^{\sup} - E_H[d(\boldsymbol{\theta}, W)], \tag{1}$$

where E_H is the expectation under the assumption that W has distribution H.

While, in principle, any distance d can be used in this definition, it is natural to consider distances that are rotation-invariant in the sense that $d(O\boldsymbol{\theta}, O\boldsymbol{\psi}) = d(\boldsymbol{\theta},\boldsymbol{\psi})$ for any $\boldsymbol{\theta}, \boldsymbol{\psi} \in \mathscr{S}^{q-1}$ and any $q \times q$ orthogonal matrix O. As we show for the sake of completeness in the appendix (see Proposition 1), any rotation-invariant distance d is of the form

$$d(\boldsymbol{\theta}, \boldsymbol{\psi}) = d_{\delta}(\boldsymbol{\theta}, \boldsymbol{\psi}) = \delta(\boldsymbol{\theta}'\boldsymbol{\psi})$$

for some function $\delta : [-1, 1] \to \mathbb{R}^+$. The standard distance axioms impose that $\delta(1) = 0$ but do not impose that δ is monotone non-increasing (unexpectedly, the triangle inequality may hold without this monotonicity condition). All classical choices, however, are monotone non-increasing; these include the arc length distance d_{arc} and the cosine distance $d_{\cos}$, that are associated with $\delta(t) = \delta_{\text{arc}}(t) = \arccos t$ and $\delta(t) = \delta_{\cos}(t) = 1 - t$, respectively. Another rotation-invariant distance

3

for which this monotonicity condition holds is the chord distance d_{chord} defined through $d_{\mathrm{chord}}(\theta,\psi) = \|\theta - \psi\| = \sqrt{2(1-\theta'\psi)}$ $=: \delta_{\mathrm{chord}}(\theta'\psi)$. Throughout, we will denote the corresponding arc distance depth (ADD), cosine distance depth (CDD) and chord distance depth (ChDD) as $D_{\mathrm{arc}}, D_{\mathrm{cos}}$ and D_{chord}, respectively.

The ADD is the arc distance depth introduced by Liu and Singh (1992). For the CDD, a direct computation yields

$$D_{\mathrm{cos}}(\theta,H) = 2 - E_H[1-\theta'W] = 1 + \theta'E_H[W]. \qquad (2)$$

Under the assumption that $E_H[W]$ is non-zero, this rewrites $D_{\mathrm{cos}}(\theta,H) = 1 + \|E_H[W]\|$ $(\theta'\mu_H)$, where $\mu_H := E_H[W]/\|E_H[W]\|$ is the spherical mean of H. This shows that the CDD is then in a one-to-one relationship with the angular Mahalanobis depth of Ley et al. (2014), provided that the location functional needed in the latter is chosen as the spherical mean. We stress, however, that, unlike the angular Mahalanobis depth, the CDD does not require choosing a location functional on the sphere and is defined also in cases where $\mu_H = 0$. To the best of our knowledge, the ChDD has not been considered in the literature.

3 Structural properties

In this section, we derive the main properties of a generic directional d-depth. We start with the following invariance result.

Theorem 1 (Rotational invariance) Let $d = d_\delta$ be a rotation-invariant distance and H be a distribution on $\mathscr{S}^{q-1}$. Then $D_{d_\delta}(\theta,H)$ is a rotation-invariant depth, in the sense that $D_{d_\delta}(O\theta,H_O) = D_{d_\delta}(\theta,H)$ for any $q \times q$ orthogonal matrix O, where H_O denotes the image of H by the transformation $x \mapsto Ox$, that is, H_O is the distribution of OW when W has distribution H.

A corollary is that if H is rotationally symmetric about θ_0 in the sense that $H_O = H$ for any $q \times q$ orthogonal matrix O fixing θ_0, then $d_\delta(O\theta,H) = d_\delta(\theta,H)$ for any such O. In particular, for any α, the α-depth region — that, as usual, is defined as the collection of θ values with a depth larger than or equal to α — is invariant under rotations fixing θ_0, hence reflects the symmetry of the distribution H about θ_0.

In contrast, parallel to the angular Mahalanobis depth of Ley et al. (2014), the CDD provides symmetric depth regions of this form for any H, i.e, irrespectively of the fact that H is rotationally symmetric or not. This follows from the comments at the end of Section 2.

Theorem 2 (Continuity) Assume that the distance d is continuous; if $d = d_\delta$, then this is equivalent to assuming that $\delta : [-1,1] \to \mathbb{R}^+$ is continuous. Let H be a

distribution on $\mathscr{S}^{q-1}$. Then, (i) the mapping $\boldsymbol{\theta} \mapsto D_d(\boldsymbol{\theta}, H)$ is continuous on $\mathscr{S}^{q-1}$; (ii) there exists $\boldsymbol{\theta}_d(H) \in \mathscr{S}^{q-1}$ such that $D_d(\boldsymbol{\theta}_d(H), H) = \sup_{\boldsymbol{\theta} \in \mathscr{S}^{q-1}} D_d(\boldsymbol{\theta}, H)$.

Note that the continuity result in Theorem 2(i) holds without any assumption on H, hence will also hold in the empirical case. Theorem 2(ii) guarantees the existence of a D_d-deepest point $\boldsymbol{\theta}_d(H)$. The deepest point (or collection of deepest points) typically depends on the distance d adopted. For the CDD, the deepest point is the spherical mean, provided that $\mathrm{E}_H[W] \neq 0$, whereas the deepest point for the ADD is the spherical median of Fisher (1985), which reduces to the circular median (Mardia and Jupp, 2000, p. 30) in dimension $q = 2$. This is in line with the Euclidean case where deepest points typically depend on the depth considered and may be multivariate medians (e.g., Tukey's halfspace or Liu's simplicial deepest points) or mean vectors (e.g., the zonoid of Koshevoy and Mosler (1997) or the moment-based Mahalanobis deepest points).

The deepest point may not be unique; for the uniform distribution on $\mathscr{S}^{q-1}$, for instance, any rotation-invariant distance-based depth will be constant over the sphere (this readily follows from Theorem 1). This lack of unicity also holds in the Euclidean case, where the barycentre of the collection $\mathscr{C}$ of deepest points is often taken as its unique representative; for most depths, it then follows from the convexity of the depth regions (which guarantees convexity of $\mathscr{C}$) that this barycentre indeed has maximal depth. It is interesting to note that directional depths are fundamentally different in this respect, as no such convexity arguments can be used. The particular nature of the sample space may induce depth regions that are even disconnected. This may occur for some multimodal distributions H; an example is given in Section 4. In contrast, note that, for $D_{\cos}$, the collection of deepest points is either $\{\boldsymbol{\mu}_H\}$, when $\mathrm{E}_H[W] \neq 0$, or $\mathscr{S}^{q-1}$, when $\mathrm{E}_H[W] = 0$, and hence it is always spherically convex.

It is desirable that if the distribution H on $\mathscr{S}^{q-1}$ has an "indisputable" location centre $\boldsymbol{\theta}_0$, then the deepest point $\boldsymbol{\theta}_d(H)$ is unique and coincides with $\boldsymbol{\theta}_0$. The following theorem provides such a Fisher consistency result.

Theorem 3 (Fisher consistency under monotone rotational symmetry) Assume that the rotation-invariant distance $d = d_\delta$ is based on a monotone strictly decreasing function $\delta : [-1, 1] \to \mathbb{R}^+$. Assume that the distribution H on $\mathscr{S}^{q-1}$ admits a density of the form $x \mapsto c_{q,h} h(x' \boldsymbol{\theta}_0)$ for some $\boldsymbol{\theta}_0 \in \mathscr{S}^{q-1}$ and some monotone strictly increasing function $h : [-1, 1]$ $\to \mathbb{R}^+$. Then, $\boldsymbol{\theta} \mapsto D_{d_\delta}(\boldsymbol{\theta}, H)$ is a monotone strictly increasing function of $\boldsymbol{\theta}' \boldsymbol{\theta}_0$, so that $\boldsymbol{\theta} \mapsto D_{d_\delta}(\boldsymbol{\theta}, H)$ is uniquely maximized at $\boldsymbol{\theta}_0$.

Theorem 3 ensures that the ADD-, CDD-, and ChDD-deepest points are equal and coincide with the modal location $\boldsymbol{\theta}_0$ of H in case the latter admits a density of the form given in the theorem. The monotonicity result entails that, irrespective of the distance d_δ used, the depth regions are of the form $\{\boldsymbol{\theta} \in \mathscr{S}^{q-1} : \boldsymbol{\theta}' \boldsymbol{\theta}_0 \geq c\}$.

In this setup, the maximal depth, $\max_{\theta \in \mathscr{S}^{q-1}} D_{d_\delta}(\theta, H)$, measures the concentration of H, as showed in the following theorem.

Theorem 4 (Maximal depth as a concentration measure) Assume that the rotation-invariant distance $d = d_\delta$ is based on a monotone strictly decreasing function $\delta : [-1, 1] \to \mathbb{R}^+$. Assume that the distribution H_κ on $\mathscr{S}^{q-1}$ admits the density $x \mapsto c_{q,\kappa,h} h(\kappa x' \theta_0)$ for some $\theta_0 \in \mathscr{S}^{q-1}$ and some monotone strictly increasing and differentiable function $h : \mathbb{R} \to \mathbb{R}^+$ such that $t \mapsto t \frac{d}{dt} \log h(t)$ is monotone strictly increasing. Then the maximal depth $D_{d_\delta}(\theta_0, H_\kappa)$ is a strictly increasing function of κ.

In Theorem 4, κ plays the role of a concentration parameter; typically, the larger κ, the more concentrated the probability mass is about the modal location θ_0. Since the maximal depth is a strictly increasing function of κ, it is itself a concentration (or spread) measure. Note that the assumption that $t \mapsto t \frac{d}{dt} \log h(t)$ is monotone strictly increasing holds in particular if h is log-convex, so that the result applies for von Mises–Fisher (vMF) distributions that are obtained for $h(u) = \exp(u)$. While Theorem 4 restricts to rotationally symmetric distributions, the maximal cosine distance depth $\max_{\theta \in \mathscr{S}^{q-1}} D_{\cos}(\theta, H) = 1 + \|E_H[W]\|$ is, irrespective of H, related to the "spherical variance" (Mardia and Jupp, 2000, p. 164), that is, to the mean resultant length $\|E_H[W]\|$ of W.

We conclude this section by stating a property showing that the proposed depths may inherit anti-symmetry properties of the distances on which they are based. More precisely, we have the following result which is restricted to rotationally-invariant distances, although a similar result can be stated for an arbitrary distance d.

Theorem 5 (Anti-symmetry) Assume that the rotation-invariant distance $d = d_\delta$ is based on a function $\delta : [-1, 1] \to \mathbb{R}^+$ that is anti-symmetric about 0, i.e., $\delta(-t) + \delta(t) = \delta(-1)$. Let H be a distribution on $\mathscr{S}^{q-1}$. Then,

(i) $\theta \mapsto D_{d_\delta}(\theta, H)$ is anti-symmetric on $\mathscr{S}^{q-1}$ in the sense that

$$D_{d_\delta}(-\theta, H) = d_\delta^{\text{sup}} - D_{d_\delta}(\theta, H);$$

(ii) If θ_0 has maximal depth, then $-\theta_0$ has minimal depth.

The arc length and cosine distances are based on anti-symmetric functions δ, but the chord distance is not. If δ is anti-symmetric, then an antipodally symmetric distribution $H \in \mathscr{S}^{q-1}$, under which $H(-B) = H(B)$ for any measurable set B on $\mathscr{S}^{q-1}$, leads to a depth function $\theta \mapsto D_{d_\delta}(\theta, H)$ that is constant. This is another property contrasting sharply with the Euclidean case, where no distribution will provide a constant depth function. To show why the claim on the constancy holds

true, consider an arbitrary measurable set $B \subset \mathscr{S}^{q-1}$ such that $\mathscr{S}^{q-1} = (-B) \cup B$ and $(-B) \cap B = \emptyset$. Then, using the antipodal symmetry of H and the antisymmetry of δ, we obtain

$$
\begin{aligned}
D_{d_\delta}(\theta, H) &= \delta(-1) - \int_{-B} \delta(\theta'w)\, dH(w) - \int_B \delta(\theta'w)\, dH(w) \\
&= \delta(-1) - \int_B \delta(-\theta'w)\, dH(w) - \int_B \delta(\theta'w)\, dH(w) \\
&= \delta(-1) - \int_B \delta(-1)\, dH(w) = \frac{\delta(-1)}{2}.
\end{aligned}
$$

An interesting question is whether or not antipodal symmetry of H is also a necessary condition for the constancy of $\theta \mapsto D_{d_\delta}(\theta, H)$ with an anti-symmetric function δ. While Liu and Singh (1992) proved that this is indeed the case for the ADD in dimension $q = 2$ under the assumption that H admits a density, it is not the case for any δ function. For instance, for the CDD, it directly follows from (2) that $\theta \mapsto D_{\cos}(\theta, H)$ is constant if and only if $E_H[W] = 0$, which shows that antipodal symmetry is not a necessary condition for the constancy of $D_{\cos}$.

4   Illustrations

This short section illustrates the theoretical results of the previous section for three empirical distributions on the circle $\mathscr{S}^1$; we restrict to the circle to allow for a visual comparison of the various depths. Denoting as $H_{\alpha,\kappa}^{\mathrm{vMF}}$ the vMF distribution on $\mathscr{S}^1$ with modal location $\theta = (\cos\alpha, \sin\alpha)'$ and concentration κ, the three empirical distributions considered are associated with a random sample of size $n = 500$ from each of the following distributions: $H_1 = H_{\pi,2}^{\mathrm{vMF}}$ (unimodal case), $H_2 = \frac{1}{2} H_{\frac{3\pi}{4},5}^{\mathrm{vMF}} + \frac{1}{2} H_{\frac{5\pi}{4},5}^{\mathrm{vMF}}$ (bimodal symmetric case), $H_3 = \frac{1}{2} H_{\frac{5\pi}{9},7}^{\mathrm{vMF}} + \frac{1}{2} H_{\frac{13\pi}{9},17}^{\mathrm{vMF}}$ (bimodal asymmetric case).

For each of the resulting empirical distributions $H_{\ell n}$, $\ell = 1, 2, 3$, Figure 1 provides plots of the distance-based depths ADD, CDD and ChDD, as well as the competing angular simplicial depth (ASD) and angular Tukey depth (ATD). The ASD and ATD were computed through the packages localdepth and depth, respectively. The distance-based depths were computed by means of R functions written by the authors. Simulated data and their graphical representations were obtained through the R package circular (Lund and Agostinelli, 2013), which is a standard reference to work with data on the unit circle.

For H_{1n}, all distance-based depth functions are monotonically strictly decreasing from their deepest point ($\approx \pi$) and do so in a symmetric way, which is in accordance with Theorems 1 and 3. These depths functions are also continuous; see Theorem 2. In contrast, the ATD is constant outside the interval of length π centered at its deepest point, which holds for any distribution on the circle (Liu and
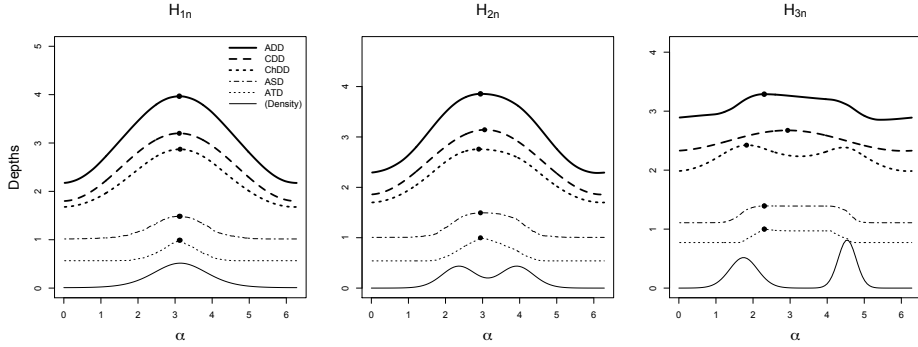
Figure 1 Plots of the depth mapping $\alpha \mapsto D(\left(\begin{smallmatrix}\cos\alpha\\\sin\alpha\end{smallmatrix}\right), H_{\ell n})$, for the distance-based depths ADD, CDD and ChDD, as well as the angular simplicial depth (ASD) and angular Tukey depth (ATD), and the empirical distributions $H_{\ell n}$, $\ell = 1,2,3$ described in Section 4 (for easier visualization, depth values were actually multiplied by 1.5 for distance-based depths, by 1 for the ASD, and by 0.5 for the ATD). Deepest points are maked by a black dot. The parent density is also plotted in each case.

Singh, 1992, Proposition 4.6.), and both the ASD and ATD are piecewise constant functions. The center-outward rankings provided by the ASD and ATD therefore yield many ties and are more rough than those given by distance-based depths. For the symmetric bimodal distribution H_{2n}, all depth functions are unimodal, hence fail to capture the bimodality of the distribution, which is not a problem since depths are not density measures but rather centrality measures. In contrast with the Euclidean case, some directional depths may exhibit multimodality, as it is the case for the ChDD for the distribution H_{3n}, where modes are more separated than in H_{2n}; (2) entails that the CDD will never exhibit such a multimodal pattern. In this last example, the depth functions reflect the asymmetry of the distribution and do not identify the same deepest point; in particular, the CDD is maximized at the spherical mean, whereas the ADD is maximized at the circular median (Mardia and Jupp, 2000, p. 20), and so are the ASD and ATD.

5   Asymptotic and robustness properties

In this section, we present asymptotic results for the distance-based depths introduced in Definition 1 and for the corresponding deepest points, as well as a robustness result regarding the breakdown point of these. We start with a Glivenko-Cantelli-type result.

Theorem 6 (Uniform almost sure consistency) Let d be a bounded and continuous distance on $\mathscr{S}^{q-1}$ and H be a distribution on $\mathscr{S}^{q-1}$. Denote as H_n the empirical distribution associated with a random sample of size n from H. Then

$$\sup_{\theta \in \mathscr{S}^{q-1}} \left| D_d(\theta, H_n) - D_d(\theta, H) \right| \to 0$$

8

almost surely as $n \to \infty$.

This result implies that we may explore empirically the properties of $D_d(\theta, H)$ by considering the corresponding sample depth function $D_d(\theta, H_n)$ for a large n. This justifies a posteriori the illustration of Theorem 3 in the previous section. The following asymptotic normality is a direct result of the central limit theorem.

Theorem 7 (Asymptotic normality of sample depth) Let d be a bounded distance on $\mathscr{S}^{q-1}$ and H be a distribution on $\mathscr{S}^{q-1}$. Denote as H_n the empirical distribution associated with a random sample of size n from H. Then as $n \to \infty$, $\sqrt{n}(D_d(\theta, H_n) - D_d(\theta, H))$ converges weakly to the normal distribution with mean zero and variance $\mathrm{Var}_H[d(\theta, W)]$.

We turn to asymptotic and robustness results for deepest points. The following strong consistency result requires that the deepest point is uniquely defined, as it is in Theorem 3.

Theorem 8 (Almost sure consistency of the deepest point) Let d be a bounded and continuous distance on $\mathscr{S}^{q-1}$ and H be a distribution on $\mathscr{S}^{q-1}$. Assume that the deepest point $\theta_d(H)$ is unique. Denote as H_n the empirical distribution associated with a random sample of size n from H, and let $\theta_d(H_n)$ be an arbitrary deepest point with respect to H_n. Then

$$\theta_d(H_n) \to \theta_d(H)$$

almost surely as $n \to \infty$.

Constructing confidence zones for $\theta_d(H)$ requires the availability of the asymptotic distribution of $\theta_d(H_n)$. Since $\theta_d(H_n)$ is an M-estimator for a location parameter on $\mathscr{S}^{q-1}$, its asymptotic distribution can easily be obtained from the results of Ko and Chang (1993), at least under rotationally symmetric distributions. We do not pursue in this direction here.

Since deepest points are commonly used as robust location estimators, it is natural to investigate their robustness, and we therefore end this section by deriving a result on their breakdown point (BDP). In the directional setup considered, the classical BDP concept (Hampel et al. (1986), pp. 97-98) is not suitable, and we adopt the directional concept of Liu and Singh (1992) defining the BDP of the (more generally, of a) deepest point $\theta_d(H)$ as the infimum of ε such that, for some contaminating distribution G on $\mathscr{S}^{q-1}$, $-\theta_d(H)$ is a deepest point of $D_d(\theta, H_\varepsilon)$ with $H_\varepsilon := (1 - \varepsilon)H + \varepsilon G$. The following result extends to an arbitrary distance d the lower bound result obtained in Liu and Singh (1992) for the arc length distance.

Theorem 9 (Breakdown point of deepest points) Let d be a bounded distance on $\mathscr{S}^{q-1}$ and H be a distribution on $\mathscr{S}^{q-1}$. Let $\theta_d(H)$ be a deepest point of $D_d(\theta, H)$. Then the breakdown point of $\theta_d(H)$ is larger than or equal to $(D_d(\theta_d(H), H) - D_d(-\theta_d(H), H))$
$/(2d^{\text{sup}})$.

To investigate how the distance d affects the lower bound, we consider the important case of vMF distributions. If $H_{q,\theta_0,\kappa}^{\text{vMF}}$ denotes the vMF(θ_0, κ) distribution on $\mathscr{S}^{q-1}$, then, for a rotation-invariant distance d_δ that is decreasing in the sense of Theorem 3, we have $\theta_{d_\delta}(H_{q,\theta_0,\kappa}^{\text{vMF}}) = \theta_0$ and

$$D_{d_\delta}(\pm\theta_0, H_{q,\theta_0,\kappa}^{\text{vMF}}) = d_\delta^{\text{sup}} - \frac{\int_{-1}^{1} \delta(\pm v)(1-v^2)^{(q-3)/2} \exp(\kappa v)\, dv}{\int_{-1}^{1}(1-v^2)^{(q-3)/2}\exp(\kappa v)\, dv},$$

which allows us to evaluate the lower bound from Theorem 9.

Figure 2 plots this lower bound as a function of κ for various dimensions q and for the ADD, CDD and ChDD. Clearly, irrespective of the dimension and the distance, the lower bound is arbitrarily small for arbitrarily small values of κ and goes to 50% as κ goes to infinity. The lower bound decreases as the dimension q increases. More importantly, for vMF distributions, the CDD-deepest point, namely the spherical mean, provides a larger lower bound than the ADD- and CHDD-deepest ones do.
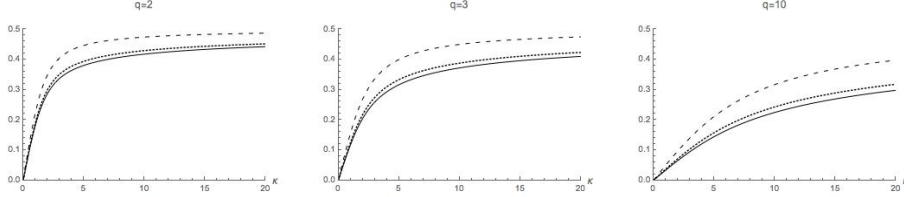

Figure 2 Plots of the lower bound in Theorem 9, for various dimensions q and for the ADD (——), CDD ( ), and ChDD ($\cdots$), as a function of the concentration κ of the underlying vMF distribution on $\mathscr{S}^{q-1}$.

6 Applications

We present two applications, which are related to spherical location estimation and supervised classification.

6.1 Spherical location estimation

Depth functions find applications in robust statistics, with the deepest point considered as a robust location estimator.

For this reason, we conducted a simulation study to investigate the efficiency and robustness properties of the deepest points associated with the proposed distance-based depths, and to compare them with those of the competing ASD- and ATD-deepest points We start with efficiency properties. For any combination of a dimension $q \in \{3, 5\}$, a sample size $n \in \{25, 50, 100\}$ and a concentration $\kappa \in \{5, 10\}$, we generated $M = 500$ independent random samples of size n from the distribution $H_{q,\theta,\kappa}^{\mathrm{vMF}}$, where $\theta = e_q$ is the last vector of the canonical basis of $\mathbb{R}^q$. For each estimator $\hat{\theta}$ of θ considered, this leads to estimates $\hat{\theta}_1, \ldots, \hat{\theta}_M$. Figure 3 provides boxplots of the resulting squared errors

$$\mathrm{SE}_m = \|\hat{\theta}_m - \theta\|^2 = 2(1 - \hat{\theta}_m' \theta), \qquad m = 1, \ldots, M, \tag{3}$$

and indicates the resulting mean square errors $\mathrm{MSE} = (1/M) \sum_{m=1}^{M} \mathrm{SE}_m$. The computational burden for the ASD- and ATD-deepest points is so prohibitive that these were considered for dimension $q = 3$ only.

Results indicate that, in dimension $q = 3$, the estimators associated with distance-based depths slightly dominate their ATD competitor and outperform their ASD one. As expected, the CDD-deepest point, that is the maximum likelihood estimator in the distributional setup considered, is in most cases the most efficient estimator. In dimension $q = 5$, where the ASD/ATD estimators could not be computed, the distance-based depths perform similarly. On the other hand, while the CDD estimator slightly dominates at all sample sizes in dimension $q = 3$, it dominates only at the largest considered sample size in dimension $q = 5$.

We now turn to the investigation of robustness properties for which we restricted to dimension $q = 3$. For any combination of a contamination level $\varepsilon \in \{0, 0.05, 0.10\}$ and a concentration $\kappa \in \{5, 10\}$, we generated $M = 500$ independent random samples of size $n = 100$ from the contaminated distributions $(1 - \varepsilon) H_{q,\theta,\kappa}^{\mathrm{vMF}} + \varepsilon \Delta_{\theta_r}$, $r = 1, 2$, where θ is set as e_q, $\theta_1 = e_{q-1}$, $\theta_2 = -\theta$, Δ_ψ denotes the point mass distribution at ψ. Hence, $r = 1, 2$ refers to contamination at an orthogonal point to θ and at the antipodal point to θ, respectively. In each sample, the deepest points of the same five depths as in Figure 3 were computed. The resulting boxplots of squared errors SE_m for $m = 1, \ldots, M$ and the mean squared errors (MSE) are provided in Figure 4.

The results show that the estimators associated with distance-based depths enjoy good robustness properties. In particular, irrespective of the contamination level ε and the type of contamination, the ADD, CDD and ChDD estimators outperform the ASD one in terms of robustness. The domination over the ATD estimator is less significant.
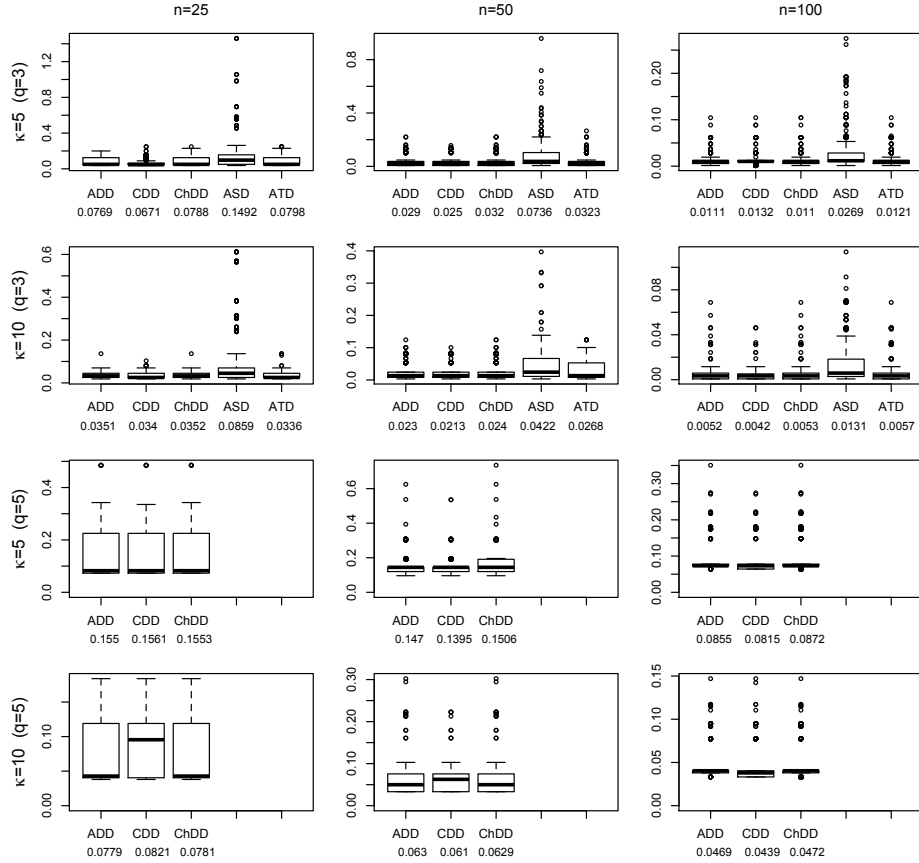
Figure 3 Boxplots, for $q \in \{3, 5\}$, $n \in \{25, 50, 100\}$ and $\kappa \in \{5, 10\}$, of the squared errors SE_m, $m = 1, \ldots, M$ (see (3)) of various depth-based estimators of θ obtained from $M = 500$ independent random samples of size n from the vMF distribution $H_{q,\theta,\kappa}^{\mathrm{vMF}}$ with location $\theta = e_q$ (the last vector of the canonical basis of $\mathbb{R}^q$). The estimators considered are the ADD-, CDD- and ChDD-deepest points, as well as (due to computational issues, for dimension $q = 3$ only) the deepest points associated with the ASD and ATD. In each case, the corresponding mean square error $\mathrm{MSE} = (1/M) \sum_{m=1}^M \mathrm{SE}_m$ is provided.

6.2 Supervised classification

Classification has been one of the most successful applications of statistical depth in the last decade, both for multivariate and functional data. While some proposals were based on the use of local depth concepts (Paindaveine and Van Bever, 2013) or a depth-based version of kNN classification (Paindaveine and Van Bever, 2015), the dominant solution finds its source in the max-depth approach of Ghosh and Chaudhuri (2005) that has later been refined by Li et al. (2012). To the best of our knowledge, depth-based classification for directional data has not been considered in the literature. In this section, we show that the max-depth approach also applies for directional data and that, in conjunction with the proposed distance-based depths, it provides classifiers on the hypersphere that dominate ASD/ATD-based
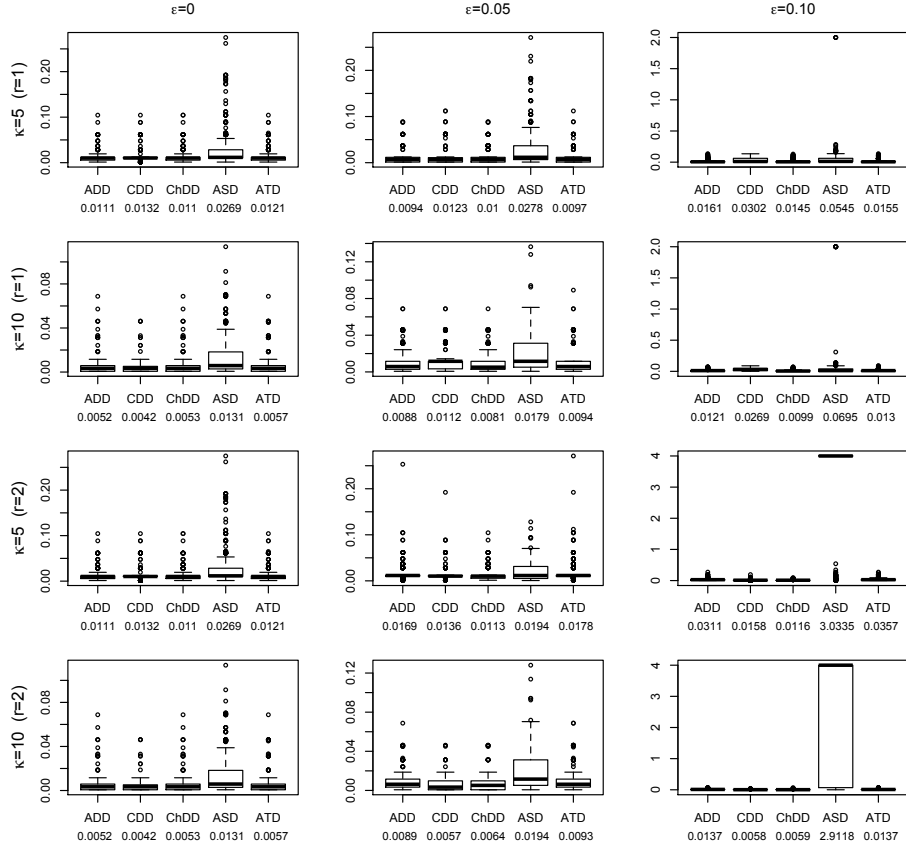
Figure 4 Boxplots, for $q = 3$, $\varepsilon \in \{0, 0.05, 0.10\}$ and $\kappa \in \{5, 10\}$, of the squared errors SE_m, $m = 1, \ldots, M$ (see (3)) of various depth-based estimators of θ obtained from $M = 500$ independent random samples of size $n = 100$ from the contaminated distribution $(1 - \varepsilon)H_{q,\theta,\kappa}^{\text{vMF}} + \varepsilon\Delta_{\theta_r}$, where θ is the last vector of the canonical basis of $\mathbb{R}^q$, Δ_ψ denotes the point mass distribution at ψ, and where θ_1 (resp., θ_2) is an orthogonal point to θ: $r = 1$ (resp., the antipodal point to θ: $r = 2$). The estimators considered are the ADD-, CDD- and ChDD-deepest points, as well as the deepest points associated with the ASD and ATD. In each case, the corresponding mean square error $\text{MSE} = (1/M)\sum_{m=1}^{M}\text{SE}_m$ is provided.

ones and that can be applied in higher dimensions as well.

Consider the spherical classification problem where independent random samples W_{1i}, $i = 1, \ldots, n_1$ and W_{2i}, $i = 1, \ldots, n_2$, respectively, come from distributions H_1 and H_2 on $\mathscr{S}^{q-1}$, and one is given the task to classify a point $w(\in \mathscr{S}^{q-1})$ as arising from H_1 ("population 1") or from H_2 ("population 2"). Denoting as $H_{\ell n_\ell}$ the empirical distribution associated with $W_{\ell i}$, $i = 1, \ldots, n_\ell$ ($\ell = 1, 2$), the max-depth classifier associated with a depth D classifies w into population 1 if $D(w, H_{1n_1}) > D(w, H_{2n_2})$, and population 2 otherwise; if $D(w, H_{1n_1}) = D(w, H_{2n_2})$, then the classification decision is based on the flip of a fair coin.

To investigate the finite-sample performances of such classifiers, we consider the Monte Carlo algorithm that was performed for dimensions $q = 2$ and $q = 10$. Denoting as e_j the jth vector in the canonical basis of $\mathbb{R}^q$ and using the

notations $H_{\alpha,\kappa}^{\mathrm{vMF}}$ and $H_{q,\theta_1,\kappa}^{\mathrm{vMF}}$ from Sections 4 and 5, respectively, we considered the following three distributional setups:

- Setup A involves the vMF distributions $H_1 = H_{\frac{\pi}{4},5}^{\mathrm{vMF}}$ and $H_2 = H_{\frac{3\pi}{4},5}^{\mathrm{vMF}}$ for $q = 2$, and $H_1 = H_{q,e_1,5}^{\mathrm{vMF}}$ and $H_2 = H_{q,e_q,5}^{\mathrm{vMF}}$ for $q = 10$; Setup A therefore involves distributions differing through the modal location only.

- In Setup B, $H_1 = H_{\frac{\pi}{3},2}^{\mathrm{vMF}}$ and $H_2 = H_{\frac{2\pi}{3},5}^{\mathrm{vMF}}$ for $q = 2$, and $H_1 = H_{q,e_q,2}^{\mathrm{vMF}}$ and $H_2 = H_{q,(\cos\frac{\pi}{6})e_{q-1}+(\sin\frac{\pi}{6})e_q,5}^{\mathrm{vMF}}$ for $q = 10$; in this setup, distributions differ through location and concentration.

- Setup C involves discrimination between the vMF distribution $H_1 = H_{\frac{3\pi}{4},4}^{\mathrm{vMF}}$ and the mixture distribution $H_2 = \frac{1}{2}H_{0,4}^{\mathrm{vMF}} + \frac{1}{2}H_{\frac{\pi}{2},4}^{\mathrm{vMF}}$ for $q = 2$, and $H_1 = H_{q,(\cos\frac{7\pi}{4})e_{q-1}+(\sin\frac{7\pi}{4})e_q,4}^{\mathrm{vMF}}$ and $H_2 = \frac{1}{2}H_{q,e_{q-1},4}^{\mathrm{vMF}} + \frac{1}{2}H_{q,e_q,4}^{\mathrm{vMF}}$ for $q = 10$.

For each setup and each q, we generated $M = 250$ independent training samples of size $n_{\mathrm{train}} = 200$ and test samples of size $n_{\mathrm{test}} = 100$ by sampling randomly from $\frac{1}{2}H_1 + \frac{1}{2}H_2$. In replication $m \in \{1,\ldots,250\}$, this associates with any depth D on $\mathscr{S}^{q-1}$ the misclassification rate $p_m(D) = N_m(D)/n_{\mathrm{test}}$, where $N_m(D)$ is the number of observations in the mth test sample that were misclassified by the max-depth classifier associated with D when based on the mth training sample. Figure 5 provides the boxplots, for several depths D, of the resulting $M = 250$ misclassification rates. As in Section 6.1, the depths considered are the ADD, CDD, ChDD, ASD and ATD; again, computational issues prevented to consider the ASD and ATD in dimension $q = 10$.

Results indicate that distance-based depth classifiers dominate in most cases their counterparts based on the ASD/ATD. It is only in Setup C that the ASD/ATD classifiers seem to slightly improve over the ADD and CDD classifiers. In all cases, the classifier based on the ChDD is the best classifier. Most importantly, in higher dimensions, the computational burden for the ASD/ATD is such that only the distance-based depth classifiers can be used.

7 Discussion

In the Euclidean multivariate setup, statistical depth has allowed to tackle in a nonparametric and robust way diverse problems, including location/scatter estimation, two-sample hypothesis testing, supervised classification, etc. While depths in the spherical setup, such as the ASD and ATD, were proposed more than two decades ago, the concept has not made its way to applications. Arguably, the reasons are that these depths are, even for moderate dimensions, very computationally intensive and that it is challenging to derive their asymptotic properties.
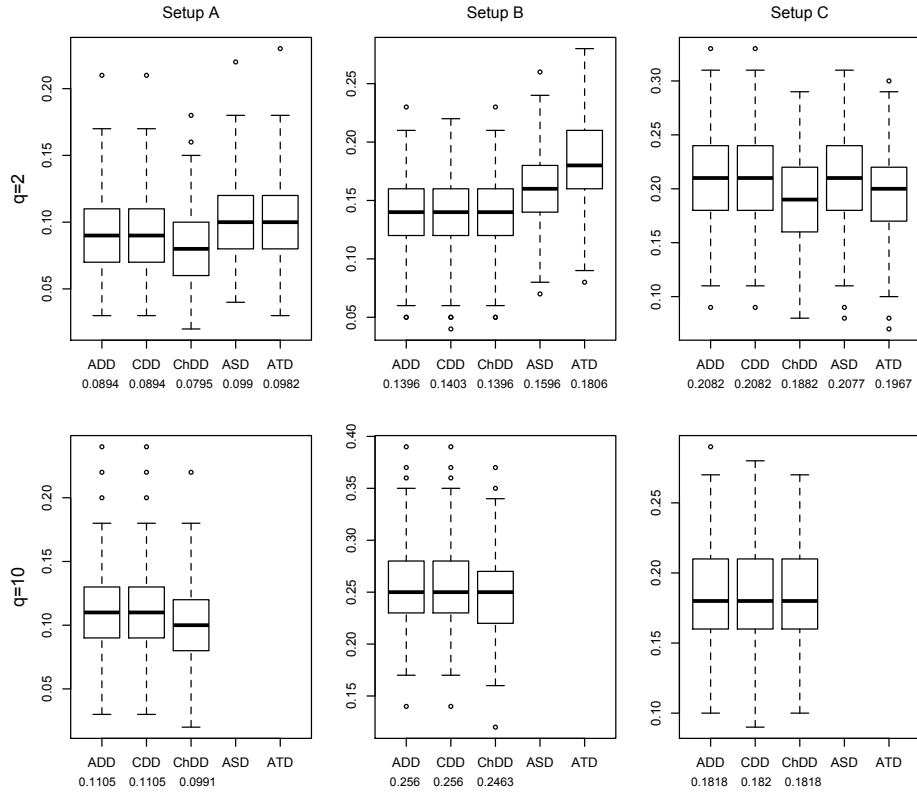
Figure 5 Boxplots, for $q \in \{2, 10\}$, of the misclassification rates $p_m(D)$, $m = 1, \ldots, M$, obtained from $M = 250$ independent replications in three different distributional setups (see Section 6.2 for details), for the max-depth classifiers associated with the ADD, CDD, ChDD, ASD and ATD (due to computational issues, the ASD and ATD were considered for dimension $q = 2$ only). In each case, the corresponding mean misclassification rate $p(D) = (1/M) \sum_{m=1}^{M} p_m(D)$ is provided.

The class of distance-based depths for directional data defined in this work clearly improve on this. These depths were showed to be computable in higher dimensions, and asymptotic results can be obtained by using standard *M*-estimation techniques. For small dimensions, where distance-based depths as well as the ASD/ATD can be evaluated, we showed through simulations that inference procedures based on the former compete equally or even dominate those based on the latter. In high dimensions, only distance-based depths can be used for directional data, which makes them of potential interest for applications involving high-dimensional spherical problems, such as those encountered in magnetic resonance, gene expression, or text mining; see, among others, Dryden (2005), Banerjee et al. (2003), and Banerjee et al. (2005).

Perspectives for future research are rich and diverse. Obviously, it would be of interest to investigate how distance-based depths can tackle the problems considered in the aforementioned high-dimensional applications. More generally, irre-

spective of the dimension, it would be desirable to develop depth-based inference procedures in various setups, including two-sample hypothesis testing and supervised classification. Finally, the present work also raised some theoretical questions of interest. For instance, in dimension $q = 2$, the arc distance depth is constant if and only if the underlying distribution H is antipodal, whereas the cosine distance depth is constant if and only if H has zero mean. In view of this, it is natural to wonder what property of H is characterized by constancy of the chord distance depth. The question can be raised on the circle with $q = 2$ or for a general dimension $q > 2$. Such characterization results are of interest since they obviously provide the basis for universally consistent tests of the corresponding properties.

As announced in Section 3, we prove the following result for the sake of completeness.

Proposition 1 Let d be a rotation-invariant distance on $\mathscr{S}^{q-1}$. Then there exists a function $\delta : [-1,1] \to \mathbb{R}^+$ such that $d(\boldsymbol{\theta}, \boldsymbol{\psi}) = \delta(\boldsymbol{\theta}'\boldsymbol{\psi})$.

Proof of Proposition 1. For any $\boldsymbol{\theta}, \boldsymbol{\psi} \in \mathscr{S}^{p-1}$, let $\boldsymbol{\psi}_{\boldsymbol{\theta}} = (\boldsymbol{\psi} - (\boldsymbol{\psi}'\boldsymbol{\theta})\boldsymbol{\theta})/\|\boldsymbol{\psi} - (\boldsymbol{\psi}'\boldsymbol{\theta})\boldsymbol{\theta}\|$ and denote as $\boldsymbol{\Gamma}_{\boldsymbol{\theta},\boldsymbol{\psi}}$ an arbitrary $q \times (q-2)$ matrix such that $O_{\boldsymbol{\theta},\boldsymbol{\psi}} = (\boldsymbol{\theta} \vdots \boldsymbol{\psi}_{\boldsymbol{\theta}} \vdots \boldsymbol{\Gamma}_{\boldsymbol{\theta},\boldsymbol{\psi}})$ is orthogonal (if $q = 2$, then we simply consider $O_{\boldsymbol{\theta},\boldsymbol{\psi}} = (\boldsymbol{\theta} \vdots \boldsymbol{\psi}_{\boldsymbol{\theta}})$). Since d is rotation-invariant, we have $d(\boldsymbol{\theta}, \boldsymbol{\psi}) = d(O'_{\boldsymbol{\theta},\boldsymbol{\psi}}\boldsymbol{\theta}, O'_{\boldsymbol{\theta},\boldsymbol{\psi}}\boldsymbol{\psi}) = d(e_1, O'_{\boldsymbol{\theta},\boldsymbol{\psi}}\boldsymbol{\psi})$, where e_1 stands for the first vector of the canonical basis of $\mathbb{R}^q$. The result then follows from the fact that $O'_{\boldsymbol{\theta},\boldsymbol{\psi}}\boldsymbol{\psi} = (\boldsymbol{\theta}'\boldsymbol{\psi}, (1 - (\boldsymbol{\theta}'\boldsymbol{\psi})^2)^{1/2}, 0, \dots, 0)'$ depends on $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ through $\boldsymbol{\theta}'\boldsymbol{\psi}$ only. $\qquad\square$

Proof of Theorem 1. Using the notation introduced in the theorem, we have that $D_{d_\delta}(O\boldsymbol{\theta}, H_O) = \delta(-1) - E_{H_O}[\delta((O\boldsymbol{\theta})'W)] = \delta(-1) - E_H[\delta((O\boldsymbol{\theta})'OW)] = \delta(-1) - E_H[\delta(\boldsymbol{\theta}'W)] = D_{d_\delta}(\boldsymbol{\theta}, H)$. $\qquad\square$

Proof of Theorem 2. (i) Since the function $w \mapsto d(\boldsymbol{\theta}, w)$ is continuous in w for any $\boldsymbol{\theta} \in \mathscr{S}^{q-1}$ and is bounded, uniformly in $\boldsymbol{\theta}$, by the integrable function $w \mapsto d^{\mathrm{sup}}$, the continuity of

$$\boldsymbol{\theta} \mapsto D_d(\boldsymbol{\theta}, H) = d^{\mathrm{sup}} - \int_{\mathscr{S}^{q-1}} d(\boldsymbol{\theta}, w) \, dH(w)$$

results from Corollary 2.8.7(i) in Bogashev (2007). (ii) The result follows from the fact that a continuous function on a compact domain attains its maximal value. $\square$

Proof of Theorem 3. Since the distribution H is rotationally symmetric about $\boldsymbol{\theta}_0$, Theorem 1 implies that $D_{d_\delta}(\boldsymbol{\theta}, H)$ depends on $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}'\boldsymbol{\theta}_0$. Consider then an arbitrary geodesic path $t \mapsto \boldsymbol{\theta}_t$ from $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}_1 = -\boldsymbol{\theta}_0$. The monotonicity assumption on h readily implies that, for any $s \in [-1,1]$, the function $t \mapsto P_H[\boldsymbol{\theta}_t'W \geq s]$ is

monotone strictly decreasing. Since

$$
\begin{aligned}
E_H[\delta(\theta_t'W)] &= \int_0^{\delta(-1)} z \frac{d}{dz} P_H[\delta(\theta_t'W) \leq z]\, dz \\
&= \delta(-1) - \int_0^{\delta(-1)} P_H[\delta(\theta_t'W) \leq z]\, dz \\
&= \delta(-1) - \int_0^{\delta(-1)} P_H[\theta_t'W \geq \delta^{-1}(z)]\, dz,
\end{aligned}
$$

it follows that

$$
D_{d_\delta}(\theta_t, H) = \delta(-1) - E_H[\delta(\theta_t'W)] = \int_0^{\delta(-1)} P_H[\theta_t'W \geq \delta^{-1}(z)]\, dz \tag{4}
$$

is strictly decreasing in t. This establishes the result. $\qquad\square$

Proof of Theorem 4. First note that for any s,

$$
P_{H_\kappa}[\theta_0'W \geq s] = \frac{\int_s^1 (1-v^2)^{(q-3)/2} h(\kappa v)\, dv}{\int_{-1}^1 (1-v^2)^{(q-3)/2} h(\kappa v)\, dv} \tag{5}
$$

(see, e.g., Paindaveine and Verdebout, 2017), which provides

$$
\frac{P_{H_\kappa}[\theta_0'W \geq s]}{1 - P_{H_\kappa}[\theta_0'W \geq s]} = \frac{\int_s^1 (1-v^2)^{(q-3)/2} h(\kappa v)\, dv}{\int_{-1}^s (1-v^2)^{(q-3)/2} h(\kappa v)\, dv}. \tag{6}
$$

Differentiation with respect to κ yields

$$
\frac{d}{ds} \frac{P_{H_\kappa}[\theta_0'W \geq s]}{1 - P_{H_\kappa}[\theta_0'W \geq s]}
$$

$$
= \frac{\int_s^1 \int_{-1}^s [v \dot{h}(\kappa v) h(\kappa u) - u \dot{h}(\kappa u) h(\kappa v)]((1-u^2)(1-v^2))^{(q-3)/2}\, du\, dv}{(\int_{-1}^s (1-v^2)^{(q-3)/2} h(\kappa v)\, dv)^2}.
$$

Since $t \mapsto t \frac{d}{dt} \log h(t) = t\dot{h}(t)/h(t)$ is strictly increasing, this derivative is strictly positive at any κ, so that the lefthand side of (6), hence also that of (5), is a monotone strictly increasing function of κ. The result then follows from the identity $D_{d_\delta}(\theta_0, H_\kappa) = \int_0^{\delta(-1)} P_{H_\kappa}[\theta_0'W \geq \delta^{-1}(z)]\, dz$; see (4). $\qquad\square$

Proof of Theorem 5. (i) The anti-symmetry of $\delta(\cdot)$ readily yields $D_{d_\delta}(-\theta, H) + D_{d_\delta}(\theta, H) = 2\delta(-1) - E_H[d_\delta(-\theta, W) + d_\delta(\theta, W)] = 2\delta(-1) - E_H[\delta(-\theta'W) + \delta(\theta'W)] = \delta(-1)$, which establishes the result. (ii) Ad absurdum, assume that $-\theta_0$ does not have minimal depth, so that there exists $\theta_1 \in \mathscr{S}^{q-1}$ with $D_{d_\delta}(\theta_1, H) < D_{d_\delta}(-\theta_0, H)$. Then Part (i) of the result implies that $D_{d_\delta}(-\theta_1, H) > D_{d_\delta}(\theta_0, H)$, which contradicts the fact that θ_0 has maximal depth. $\qquad\square$

Proof of Theorem 6. The result directly follows from Theorem 16(a) in Ferguson (1996). $\qquad\square$

Proof of Theorem 7. The result trivially follows from applying the central limit theorem to the expression $\sqrt{n}(D_d(\boldsymbol{\theta}, H_n) - D_d(\boldsymbol{\theta}, H)) = -n^{-1/2} \sum_{i=1}^{n} (d(\boldsymbol{\theta}, W_i) - \mathrm{E}_H[d(\boldsymbol{\theta}, W)])$. $\qquad\square$

Proof of Theorem 8. In view of Theorem 6, the result is a corollary of Theorem 2.12 and Lemma 14.3 in Kosorok (2008). $\qquad\square$

Proof of Theorem 9. From Lemma 2.3 in Strasser (1985), we obtain that, for any $\boldsymbol{\theta} \in \mathscr{S}^{q-1}$, $|D_d(\boldsymbol{\theta}, H_\varepsilon) - D_d(\boldsymbol{\theta}, H)| = \varepsilon(E_G[d(\boldsymbol{\theta}, W)] - E_H[d(\boldsymbol{\theta}, W)]) \leq \varepsilon d^{\sup} d_1(H, G)$, where $d_1(H, G)$ denotes the variational distance between H and G. Lemmas 2.4 and 2.5(i) in Strasser (1985) then yield that, still for any $\boldsymbol{\theta} \in \mathscr{S}^{q-1}$, $|D_d(\boldsymbol{\theta}, H_\varepsilon) - D_d(\boldsymbol{\theta}, H)| \leq \varepsilon d^{\sup}$. The result readily follows. $\qquad\square$

References

Agostinelli, C. and M. Romanazzi (2013a). Localdepth. R package version 0.5-7.

Agostinelli, C. and M. Romanazzi (2013b). Nonparametric analysis of directional data based on data depth. Environmental and Ecological Statistics 20, 253–270.

Banerjee, A., I. Dhillon, J. Ghosh, and S. Sra (2003). Generative model-based clustering of directional data. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 19–28.

Banerjee, A., I. Dhillon, J. Ghosh, and S. Sra (2005). Clustering on the unit hypersphere using von mises-fisher distributions. Journal of Machine Learning research 6, 1345–1382.

Bogashev, V. I. (2007). Measure Theory, Volume I. Berlin Heidelberg: Springer.

Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. Annals of Statistics 33, 1643–1665.

Ferguson, T. (1996). A Course in Large Sample Theory. Chapman and Hall/CRC.

Fisher, N. I. (1985). Spherical medians. Journal of the Royal Statistical Society Series B 47, 342–348.

Genest, M., J. Masse, and J. Plante (2012). Depth. R package version 2.0-0.

Ghosh, A. K. and P. Chaudhuri (2005). On maximum depth and related classifiers. Scandinavian Journal of Statistics 32(2), 327–350.

Gill, J. and D. Hangartner (2010). Circular data in political science and how to handle it. Political Analysis 18, 316–336.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). Robust Statistics: The Approach Based on Influence Functions. New York: Wiley.

Hornik, K., I. Feinerer, M. Kober, and C. Buchta (2012). Spherical k-means clustering. Journal of Statistical Software 50, 1–22.

Ko, D. and T. Chang (1993). Robust *m*-estimators on spheres. Journal of Multivariate Analysis 45, 104–136.

Koshevoy, G. and K. Mosler (1997). Zonoid trimming for multivariate distributions. Annals of Statistics 25, 1998–2017.

Kosorok, M. R. (2008). Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics. New York: Springer.

Ley, C., C. Sabbah, and T. Verdebout (2014). A new concept of quantiles for directional data. Electronic Journal of Statistics 8, 795–816.

Li, J., J. Cuesta-Albertos, and R. Y. Liu (2012). Dd-classifier: Nonparametric classification procedures based on dd-plots. Journal of the American Statistical Association 107(498), 737–753.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. Annals of Statistics 18(1), 405–414.

Liu, R. Y. and K. Singh (1992). Ordering directional data: concepts of data depth on circles and spheres. Annals of Statistics 20, 1468–1484.

Lund, U. and C. Agostinelli (2013). Circular. R package version 0.4-7.

Mardia, K. V. and P. E. Jupp (2000). Directional Statistics. John Wiley & Sons.

Paindaveine, D. and G. Van Bever (2013). From depth to local depth : a focus on centrality. Journal of the American Statistical Association 105, 1105–1119.

Paindaveine, D. and G. Van Bever (2015). Nonparametrically consistent depth-based classifiers. Bernoulli 21, 62–82.

Paindaveine, D. and T. Verdebout (2017). Inference on the mode of weak directional signals: a le cam perspective on hypothesis testing near singularities. Annals of Statistics 45, 800–832.

Rousseeuw, P. J. and A. Struyf (2004). Characterizing angular symmetry and regression symmetry. Journal of Statistical Planning and Inference 122(1-2), 161–173.

Small, C. (1987). Measures of centrality for multivariate and directional distributions. Canadian Journal of Statistics 39, 31–39.

Strasser, H. (1985). Mathematical Theory of Statistics. Statistical Experiments and Asymptotic Decision Theory. Berlin, New York: De Gruyter.

Tukey, J. W. (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.