Contributions to nonparametric and semiparametric inference based on statistical depth

VAN BEVER, GERMAIN

Link to publication

Faculté des Sciences

Département de Mathématique

Contributions to Nonparametric and Semiparametric Inference based on Statistical Depth

Germain VAN BEVER

Thèse présentée en vue de l'obtention du grade de **Docteur en Sciences**, orientation Statistique

Promoteur: Davy Paindaveine

Membres du jury: Catherine Dehon, Marc Hallin, Siegfried Hörmann, Mia Hubert, Maarten Jansen, Karl Mosler

Septembre 2013

Contributions to Nonparametric and Semiparametric Inference based on Statistical Depth

Germain VAN BEVER

Thèse présentée en vue de l'obtention du grade de **Docteur en Sciences**, orientation Statistique

"What do I know, or think I know, from my own experience and not by literary osmosis? An honest answer would be: 'Not much, and I am not too sure of most of it'. "

Dean Acheson

in *J. Cornfield's "A statistician's apology"*.

Remerciements

Malgré la richesse et la diversité qu'il existe dans la recherche et l'écriture d'une thèse, il est une constante: l'urgence dans laquelle un doctorant écrit ses remerciements. C'est donc sans déroger à la règle que je m'attelle (à quelques heures du dépôt, donc) à la rédaction de ces quelques lignes.

Pour commencer, mes plus chaleureux et sincères remerciements vont à la personne sans laquelle cette aventure n'aurait pas été possible, le Professeur Davy Paindaveine. Je tiens à le remercier d'avoir accepté d'encadrer cette thèse et lui suis redevable de m'avoir fait découvrir dès la fin de mon master ce sujet passionnant qu'est la profondeur. Sa passion pour la recherche, sa capacité à sans cesse trouver de nouvelles idées (pour sortir des nombreuses voies de garage que compte la profondeur) et, surtout, son extraordinaire optimisme (mathématique, académique et personnel) sont autant de facteurs qui m'auront marqué au cours de ces quatre années. Aucune séance de travail (que ce soit en face d'un tableau, autour d'un thé/café ou en courant dans le bois de la Cambre) n'a manqué de me redonner la confiance (un peu trop souvent manquante) nécessaire afin de terminer ce travail et me donner l'envie d'un "One more thing". Merci. Sincèrement.

Je voudrais également exprimer ma reconnaissance envers tous les membres du Jury, Mesdames les Professeurs Catherine Dehon et Mia Hubert, Messieurs les Professeurs Marc Hallin, Siegfried Hörmann, Maarten Jansen, Karl Mosler et Davy Paindaveine, et en particulier à Siegfried Hörmann d'en avoir accepté la présidence.

Mes remerciements vont aussi au *Fonds National pour la Recherche Scientifique – F.N.R.S.* (Belgique) pour le soutien financier apporté via un mandat d'*Aspirant* durant les quatre années de mon doctorat.

Je remercie mes collègues de bureau (et amis) qui n'ont pas manqué d'agrémenter ce doctorat de très bons moments. Merci Carine d'avoir supporté un doctorant parfois grognon (mais au final aux horaires un peu décalés) dans notre bureau, Christophe et Yvik (pour les "petits problèmes" et la proposition de m'embarquer dans l'aventure BSSM), et, dans le désordre, toute l'équipe Gauss'Oh Fast (qui renaîtra de ses cendres), Naïm, Laurent, Julian, etc...

Merci Jean, pour les pauses forcées et prolongées (mais nécessaires!) au département et pour la brillante idée de me proposer d'être ton témoin seulement 10 jours après ma date de dépôt. A charge de revanche...

Pauliina, thank you for bringing your craziness in my PhD. I'll certainly miss our pre-or-post work meeting discussions.

Merci à ma famille, pour ne pas trop avoir demandé si tout se passait bien, ou surtout pour avoir compris que le tableau que je dressais n'était pas aussi sombre. Merci d'avoir supporté l'ermite que je suis devenu sur la fin et pour le soutien, de près ou de loin.

Merci à tous mes amis, pour l'entrain, la folie et l'ambiance qu'ils ont apportés aux différentes journées et soirées passées en leur compagnie, pour leur présence et soutien dans les moments cruciaux. Timothée, Camille, Laurent, Virginie, Florian, Aline, Catherine, Christophe, Vincent, Jérôme, Isabelle, toute la petite troupe de 2004 à aujourd'hui, qui a fait que ces études et les liens développés me sont chers. Malgré la distance et le temps qui passe, c'est un vrai bonheur de vous connaître tous.

Contents

Introduction

1 The notion of centrality — Depth

1.1 The origins – Multivariate medians

The notion of center of an object, be it a set of observations, a physical object or a random variable X, is difficult to define. Whether definitions refer to *a point, pivot or axis around which anything rotates or revolves*[1], therefore being inseparable from the notion of symmetry itself, or to *the middle point, as the point or part that is equally distant from all points, sides, ends or surfaces of something*[2] there is no canonical way to define it. All definitions, however, agree on the importance of distance or geometry in the construction of such notion.

From a mathematical point of view, many such notions of center—of a random variable X having distribution P, say—were defined in the univariate case and an abundant literature of so-called *univariate location measures* exists. The most canonical notion, of course, is the mean or expectation $\mathrm{E}[X] = \int x \, dP(x)$. However, due to the high sensitivity of this particular location functional, it is often advocated that, should the focus be put on broader and more robust applicability, the competing notion of median presents much more appeal. Indeed, it is a well-known fact that, although it suffices to have a single point contaminating a data set and going to infinity to force the mean to do the same, the median will require, by contrast, 50% of the data to be moved to infinity before it does as well.

The geometry of the quadratic distances that underline the definition of the mean (that, alternatively, can be defined as the location minimizing the functional $x \mapsto \mathrm{E}[(X - x)^2]$) makes it particularly amenable to generalization in higher dimensions. This is the reason why theory based upon the multivariate Gaussian distribution has been dominating multivariate analysis for a long time.

The median of a random variable X, denoted $\mathrm{Med}(X)$, is defined through the cumulative distribution as the point m_P such that $P\big[X \leq m_P\big] \geq 1/2$ and $P\big[X \geq m_P\big] \leq 1/2$. Accordingly, the median $\mathrm{Med}(X^{(n)})$ of a dataset $X^{(n)} = \{X_1, \ldots, X_n\} \subset \mathbb{R}$ will be defined by substituting the empirical distribution on $\mathbf{X}^{(n)}$, $P^{(n)}$ say, to P.

The lack of natural ordering in $\mathbb{R}^d$ prevents a straightforward extension of the latter definitions, so that one may wonder what the appropriate analogues in two or more dimensions are. Regardless of the notion employed, there are, however, certain properties these notions should definitely have, the first of which being, as in the univariate case, robustness[3]. Another such condition is that, under symmetry, the (multivariate) median should coincide with the symmetry center. In the univariate case, the notion of symmetry presents no ambiguity (a random vector X is symmetric about μ if $X - \mu \overset{d}{=} \mu - X$, where $\overset{d}{=}$ denotes equality in distribution) and univariate location measures typically coincide under symmetry. This is not necessarily so in higher dimensions as symmetry can be generalized in many ways, see Section 1.3 for details.

Extending the concept of median to the multivariate setup (or a similar approach that consists in ordering multivariate observations) has generated numerous publications over the past few years;

---

[1]Harrap's Dictionnary of Contemporary English.

[2]*ibidem.*

[3]Many tools for measuring robustness exist. A classical way to compare robustness of location measures is through their breakdown point (see Hodges (1967) for the univariate definition and Hampel (1971) more generally). See also Section 2.4.

see, for example, Barnett (1976) or Hettmansperger *et al.* (1992) for a review about how to order multivariate data, and Donoho & Gasko (1987) or Small (1990) for a comprehensive summary of existing multivariate analogues of the median at that time.

The naive[4] attempt to take the componentwise median (first used by Hayford (1902) for geographical considerations) showed that, despite some encouraging robustness properties, very poor performances were to be expected, particularly in the case of highly correlated univariate components. To add on this drawback, this vector of medians is not equivariant under rotation or arbitrary affine transformation of the data (see, for example, Bickel, 1964; Barnett, 1976), so that the way of measuring the data will have a strong impact on the outcome of the procedure (which is of course to be avoided).

To improve on this definition, many authors independently considered the spatial median[5] as a natural generalization of the univariate median in different situations, see, e.g. Gini & Galvani (1929), Scates (1933) and Haldane (1948).

Definition 1.1. Let $\mathbf{X}$ be a random vector having distribution P on $\mathbb{R}^d$. The *spatial median* of $\mathbf{X}$ is the location $\hat{\boldsymbol{\mu}}_S(P) \in \mathbb{R}^d$ that minimizes $\mathrm{E}_P[||\mathbf{X} - \boldsymbol{\mu}||]$, where $||.||$ denotes the standard Euclidian norm.

Note that the distribution P may be that of the empirical distribution $P^{(n)}$ of n i.i.d. data points $\mathbf{X}_1, \ldots, \mathbf{X}_n$ sharing the same distribution P. Locating the spatial median, in that case, amounts to finding the solution of

$$\hat{\boldsymbol{\mu}}^{(n)} = \mathrm{arginf}_{\boldsymbol{\mu} \in \mathbb{R}^d} \sum_{i=1}^n ||\mathbf{X}_i - \boldsymbol{\mu}||.$$

This problem, for which there now exist plenty of algorithmic solutions (see, for example, Vardi & Zhang, 2001), is actually far much older than the introduction of the multivariate median. Indeed, minimizing a weighted sum of the Euclidian distances from m points in $\mathbb{R}^d$ was already known, in industrial applications, as the optimal location problem of Weber (1909). The problem actually goes back to Fermat in the seventeenth century (for $m = 3$ and equal weights) but was only generalized to its actual form by Simpson (1750) (see Kuhn, 1973). It is interesting to note that Kemperman (1987), following the same idea, discussed the median of a finite measure on an arbitrary Banach space and proved, under strict convexity of the underlying space and provided the distribution is not supported on a straight line, uniqueness of the resulting location functional[6].

A different celebrated and closely related instance, replacing the expected absolute deviation with expected volume of a simplex, is the simplicial volume median from Oja (1983). Let $S(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1})$ denote the (closed) convex hull of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1} \in \mathbb{R}^d$ and ΔS its volume.

Definition 1.2. Let P be a distribution on $\mathbb{R}^d$ and $\alpha > 0$. The simplicial volume location functional of $\mathbf{X}$ of order α is the location $\hat{\boldsymbol{\mu}}_{SV}^\alpha(P) \in \mathbb{R}^d$ that minimizes $\boldsymbol{\mu} \mapsto \mathrm{E}[\{\Delta S(\mathbf{X}_1, \ldots, \mathbf{X}_d, \boldsymbol{\mu})\}^\alpha]$, where $\mathbf{X}_1, \ldots, \mathbf{X}_d$ are i.i.d. P. The *simplicial volume median* is $\hat{\boldsymbol{\mu}}_{SV}^1(P)$.

---

[4]Actually, the second-in-order naive approach, if one considers the poorly defined tentative extension through the cdf, defining "a" median as a location $\boldsymbol{m}$ such that $P[\mathbf{X} \leq \boldsymbol{m}] \geq 1/2$ and $P[\mathbf{X} \geq \boldsymbol{m}] \leq 1/2$.

[5]The denomination *mediancenter* is used in Gower (1974) and the first reference as a "spatial median" can be found in Brown (1983).

[6]Uniqueness in the Euclidian case of $\mathbb{R}^d$ was treated by Milasevic & Ducharme (1987). This is in strict contrast with the univariate case where uniqueness does not hold in general.

Another technique was introduced in Barnett (1976) (see also the comment by Plackett, 1976) and Green (1981), where the authors suggested "peeling" the distribution. A somewhat similar approach based on nested sets can be found in Eddy (1982, 1985). A multivariate median of a dataset is obtained by sequentially suppressing the observations lying on the boundary of the convex hull of the data and taking the mean of the innermost layer. Note, however, that this construction does not allow to define a multivariate median for a generic distribution P as it does not have any population equivalent.

Based on univariate measures of outlyingness, Stahel (1981) and, independently, Donoho (1982) used projection pursuit ideas to generalize the univariate weighted location estimator of Mosteller & Tukey (1977) and defined the robust location estimator (that we describe in the sample case for simplicity) $\hat{\boldsymbol{\mu}}_W(\mathbf{X}^{(n)}) = \sum w_i \mathbf{X}_i / \sum w_i$ where the weights $w_i = w(r_d(\mathbf{X}_i; \mathbf{X}^{(n)}))$ are decreasing as a function of the outlyingness measure

$$r_d(\boldsymbol{x}; \mathbf{X}^{(n)}) = \max_{||\boldsymbol{u}||=1} \frac{|\boldsymbol{u}'\boldsymbol{x} - \mathrm{Med}(\boldsymbol{u}'\mathbf{X}^{(n)})|}{\mathrm{MAD}(\boldsymbol{u}'\mathbf{X}^{(n)})}. \tag{1.1}$$

Here, $\mathrm{MAD}(\mathbf{X}^{(n)}) = \mathrm{Med}(|\mathbf{X}^{(n)} - \mathrm{Med}(\mathbf{X}^{(n)})|)$ denotes the *median absolute deviation* of the dataset $\mathbf{X}^{(n)} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$. The related projection median was studied in Tyler (1994).

Definition 1.3. Let $\mathbf{X}$ be a random vector having distribution P on $\mathbb{R}^d$. The *projection median* of $\mathbf{X}$ is

$$\hat{\boldsymbol{\mu}}_{Pr}(P) = \mathrm{arginf}_{\boldsymbol{\mu} \in \mathbb{R}^d} \mathrm{argsup}_{\boldsymbol{u} \in \mathcal{S}^{d-1}} \frac{|\boldsymbol{u}'\boldsymbol{\mu} - \mathrm{Med}(\boldsymbol{u}'\mathbf{X})|}{\mathrm{MAD}(\boldsymbol{u}'\mathbf{X})},$$

where $\mathcal{S}^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x}'\boldsymbol{x} = 1\}$ denotes the unit sphere in $\mathbb{R}^d$.

A classical requirement for location estimators/functionals is the *affine-equivariance property*.

Property 1.4. Let $\mathbf{X}$ be a random vector on $\mathbb{R}^d$ having distribution P and $\hat{\boldsymbol{\mu}}(P)$ be a multivariate location functional. Then $\hat{\boldsymbol{\mu}}(.) : \mathcal{P} \to \mathbb{R}^d$ is said to be *affine-equivariant* if $\hat{\boldsymbol{\mu}}(P_{\mathbf{A}\mathbf{X}+\boldsymbol{b}}) = \mathbf{A}\hat{\boldsymbol{\mu}}(P) + \boldsymbol{b}$, where $P_{\mathbf{A}\mathbf{X}+\boldsymbol{b}}$ denotes the distribution of $\mathbf{A}\mathbf{X} + \boldsymbol{b}$ for the $d \times d$ invertible matrix $\mathbf{A}$ and $\boldsymbol{b} \in \mathbb{R}^d$.

As it turns out, both Definitions 1.2 and 1.3 satisfy Property 1.4. This is not the case for Definition 1.1 for a non-orthogonal matrix $\mathbf{A}$ in general. This is the reason why Chakraborty & Chaudhuri (1996, 1998) and Chakraborty *et al.* (1998) proposed a data-driven transformation-retransformation technique turning the spatial median into an affine-equivariant location functional. A similar approach was adopted in Hettmansperger & Randles (2002), where the initial data is first standardized using Tyler's M-estimator of scatter (Tyler, 1987).

1.2   Existing notions – Depth functions

Many of the definitions introduced in the previous section share a common construction. Indeed, most of them define a multivariate median (in $\mathbb{R}^d$) as a location optimising some criterion, that, in some sense, reflects the centrality of a point $\boldsymbol{x}$ with respect to the underlying distribution. This motivated the development of general ways to measure centrality via *depth functions*. Such mappings provide, in turn, new multivariate medians. They also—and contrary to the naive approach to multivariate location that looks only for the most central point—allow for (i) comparing relative centrality of two locations and, consequently, (ii) providing a center-outward ordering (that would, in turn, make possible the definition of multivariate quantiles, see Serfling, 2002b).

More precisely, letting $\mathcal{P}$ denote the class of distributions over the Borel sets $B \in \mathcal{B}^d$ of $\mathbb{R}^d$, a depth function is a mapping $D(.,.) : \mathbb{R}^d \times \mathcal{P} \to \mathbb{R} : (\boldsymbol{x}, P) \mapsto D(\boldsymbol{x}, P)^7$ that, intuitively, associates with any location $\boldsymbol{x}$ a value reflecting its centrality with respect to distribution P. Several recent reviews on data depth include Liu *et al.* (2006) (and in particular the introductive chapter by Serfling, 2006a), Cascos (2009), Romanazzi (2009), Mosler (2012) or the theoretical approach from Zuo & Serfling (2000a).

Many such mappings have been introduced in the literature and are now described in their population version. We present here these functions in historical order of introduction, starting with the two seminal examples of *halfspace depth* and *simplicial depth*.

- **Halfspace Depth**: The earliest notion of depth dates back to Tukey (1975) (see also Tukey, 1977). Initially introduced as a tool to picture the data, the halfspace depth became increasingly popular and was quickly widely used in many procedures, due to its numerous useful properties and its intuitive interpretation.

  In the univariate case, the median (of some distribution with cdf F) is univocally characterized as the location maximizing $D(x, F) = \min\big(F(x), 1 - F(x^-)\big)$, where $F(x^-)$ denotes the left-sided limit of F at x. Generalizing this last quantity to the multivariate case, the halfspace depth of $\boldsymbol{x} \in \mathbb{R}^d$ is defined as the "minimal" probability of any closed halfspace containing $\boldsymbol{x}^8$.

  Definition 1.5. Let $\boldsymbol{x} \in \mathbb{R}^d$. Let $\mathbf{X}$ be a random vector on $\mathbb{R}^d$ with distribution $P \in \mathcal{P}$. The *halfspace depth* of $\boldsymbol{x}$ with respect to P is

  $$D_H(\boldsymbol{x}, P) = \inf_{\boldsymbol{u} \in S^{d-1}} P\big[\boldsymbol{u}'(\mathbf{X} - \boldsymbol{x}) \geq 0\big].$$

  Germ of this definition, in the bivariate case and only interested in the associated multivariate median, can be traced back to Hotelling (1929). The halfspace depth is actually a special case of particular applications used in economic game theory called "index functions"; see Small (1987). Rousseeuw & Ruts (1999) cover many of the properties of halfspace depth.

- **Simplicial Depth**: Another characterization of the median serves as foundation for this depth, first introduced in Liu (1987, 1988) and thoroughly developed in Liu (1990). For X_1 and X_2 two i.i.d. observations with common cdf F, the median $\mathrm{Med}(F)$ is the location with the highest probability to be covered by the random segment $\overline{X_1 X_2}$. More precisely, $\mathrm{Med}(F) = \arg\max_x D(x, F)$, where $D(x, F) = P\big[x \in \overline{X_1 X_2}\big] = 2F(x)(1 - F(x^-))$. Seeing $\overline{X_1 X_2}$ as the convex hull of the set $\{X_1, X_2\}$, it is therefore natural to introduce the following definition.

  Definition 1.6. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $P \in \mathcal{P}$. The *simplicial depth* of $\boldsymbol{x}$ with respect to P is

  $$D_S(\boldsymbol{x}, P) = P\big[\boldsymbol{x} \in S(\mathbf{X}_1, \ldots, \mathbf{X}_{d+1})\big],$$

  where $S(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1})$ still stands for the simplex with vertices $\boldsymbol{x}_1, \ldots \boldsymbol{x}_{d+1}$ and $\mathbf{X}_1, \ldots, \mathbf{X}_{d+1}$ are i.i.d. random vectors with common distribution P.

- **Majority Depth**: This restricted depth notion (probably due to its high computational costs and its lack of sound theoretical properties) was introduced in Singh (1991) and further studied

---

[7]Some rare depth functions will only be defined for a subset of $\mathcal{P}$.

[8]Note that the bivariate halfspace depth of a point is equivalent to the sign test statistic of Hodges (1955).

and used in Liu & Singh (1993). Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d$ in $\mathbb{R}^d$ be in general position, so that they determine a unique hyperplane $H_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d}$ containing them. Let $H_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d}^P$ denote the closed halfspace with boundary $H_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d}$ that carries a P-probability $\geq 1/2$.

Definition 1.7. The *majority depth* of $\boldsymbol{x} \in \mathbb{R}^d$ with respect to the continuous distribution $P \in \mathcal{P}$ is defined by
$$D_{\mathrm{Maj}}(\boldsymbol{x}, P) = P\big[\boldsymbol{x} \in H_{\mathbf{X}_1, \ldots, \mathbf{X}_d}^P\big],$$
where $\mathbf{X}_1, \ldots, \mathbf{X}_d$ is a random sample from P.

In the univariate case, this last expression boils down to $D_{\mathrm{Maj}}(x, F) = 1/2 + \min(F(x), (1 - F(x^-)))$. This last expression, again, is maximized at the median.

- **Projection Depth**: Liu (1992) turned the outlyingness measure (1.1) used in the Stahel/Donoho location estimator into a proper depth function. The general version of projection depth we now present was introduced in Zuo & Serfling (2000a).

  For $\mathbf{X}$ a random vector on $\mathbb{R}^d$ having distribution P and $\boldsymbol{u} \in \mathcal{S}^{d-1}$, let $P_{[\boldsymbol{u}]}$ denote the distribution of $\boldsymbol{u}'\mathbf{X}$. Let $\mu(P)$ and $\sigma(P)$ be univariate location and scale functionals.

  Definition 1.8. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $\mathbf{X}$ be a random vector with distribution $P \in \mathcal{P}$. The *projection depth* of $\boldsymbol{x}$ with respect to P is defined by
  $$D_P(\boldsymbol{x}, P) = \Big(1 + \sup_{\boldsymbol{u} \in \mathcal{S}^{d-1}} \frac{|\boldsymbol{u}'\mathbf{X} - \mu(P_{[\boldsymbol{u}]})|}{\sigma(P_{[\boldsymbol{u}]})}\Big)^{-1}.$$

  Classical choices of location and scale include $\mu(P_{[\boldsymbol{u}]}) = \mathrm{E}[\boldsymbol{u}'\mathbf{X}]$ and $\sigma(P_{[\boldsymbol{u}]}) = \mathrm{Var}[\boldsymbol{u}'\mathbf{X}]$ or $\mu(P_{[\boldsymbol{u}]}) = \mathrm{Med}(P_{[\boldsymbol{u}]})$ and $\sigma(P_{[\boldsymbol{u}]}) = \mathrm{MAD}(P_{[\boldsymbol{u}]})$. The latter choices, as already noticed in Donoho & Gasko (1992), of course lead to more robust procedures (see Section 2.4).

  A more general version of projection depth, defined for observations that are no longer vectors in $\mathbb{R}^d$ but rather tensors, was introduced in Hu *et al.* (2011).

- **Mahalanobis Depth**: In the same work, Liu (1992) suggested to use the Mahalanobis distance to the mean (Mahalanobis, 1936) as a measure of outlyingness to develop the corresponding depth function (in the same spirit as projection depth above). The Mahalanobis distance between two points $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^d$ with respect to the positive definite matrix $\mathbf{M}$ (chosen to be the standard covariance matrix $\mathrm{Cov}(P)$ in Liu, 1992) is $d_{\mathbf{M}}(\boldsymbol{x}, \boldsymbol{y}) = \big[(\boldsymbol{x} - \boldsymbol{y})'\mathbf{M}(\boldsymbol{x} - \boldsymbol{y})\big]^{1/2}$. Liu & Singh (1993) pointed out the lack of robustness of the resulting function and the fact that it may fail to achieve maximality at the center of certain symmetric distributions. A more general version goes as follows.

  Let $\boldsymbol{\mu}(P)$ and $\boldsymbol{\Sigma}(P)$ denote some affine-equivariant location and scatter functionals. Recall that a scatter functional $\boldsymbol{\Sigma}(.)$ is affine-equivariant whenever $\boldsymbol{\Sigma}(P_{\mathbf{A}\mathbf{X}+b}) = \mathbf{A}\boldsymbol{\Sigma}(P_{\mathbf{X}})\mathbf{A}'$ for any invertible $d \times d$ matrix $\mathbf{A}$ and d-vector $\boldsymbol{b}$.

  Definition 1.9. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $P \in \mathcal{P}$. The *Mahalanobis depth* of $\boldsymbol{x}$ with respect to P is defined by
  $$D_{\mathrm{Mah}}(\boldsymbol{x}, P) = \Big(1 + d_{\boldsymbol{\Sigma}(P)}^2(\boldsymbol{x}, \boldsymbol{\mu}(P))\Big)^{-1}.$$

  The required affine-equivariance of $\boldsymbol{\mu}(P)$ and $\boldsymbol{\Sigma}(P)$ will allow, in turn, the depth function to be affine-invariant (see Section 1.3). Plugging in robust estimators of location and scatter provides a robust depth measure.

6

- **Zonoid Depth**: This slightly different notion of depth (of an L^2, rather than L^1, nature) was introduced in Koshevoy & Mosler (1997). For $\mathbf{X}$ a random vector with distribution P on $\mathbb{R}^d$ having finite expectation, define the zonoid α-trimmed regions ($\alpha \in (0, 1]$) as

$$D_Z^\alpha(P) = \left\{ \int_{\mathbb{R}^d} \boldsymbol{x} g(\boldsymbol{x}) dP(\boldsymbol{x}) \Big| g : \mathbb{R}^d \to [0, \frac{1}{\alpha}] \text{ measurable and } \int_{\mathbb{R}^d} g(\boldsymbol{x}) dP(\boldsymbol{x}) = 1 \right\}.$$

While this definition may seem obscure, it is interesting to note that, for P a continuous distribution, $D_Z^\alpha(P) = \left\{ \mathrm{E}_P[\mathbf{X} \mathbf{1}_U(\mathbf{X})] : U \in \mathcal{B}^d, P[U] = \alpha \right\}$, where $\mathbf{1}_\mathbf{A}(\boldsymbol{x})$ denotes the indicator function on the set $\mathbf{A}$. In the univariate case, it therefore holds $D_Z^\alpha(F) = [Q_\alpha^-, Q_\alpha^+]$, where Q_α^- (resp., Q_α^+) is the (P-)gravity center of the lower (resp., upper) tail with probability α.

Definition 1.10. Let $\boldsymbol{x} \in \mathbb{R}^d$ and let $P \in \mathcal{P}$ have finite expectation. The *zonoid depth* of $\boldsymbol{x}$ with respect to P is

$$D_Z(\boldsymbol{x}, P) = \sup\{\alpha : \boldsymbol{x} \in D_Z^\alpha(P)\},$$

should $\boldsymbol{x} \in D_Z^\alpha(P)$ for some $\alpha \in (0, 1]$, 0 otherwise.

It is clear that $0 \le D_Z^\alpha(\boldsymbol{x}, P) \le 1$. Furthermore, it holds that $D_Z^1(P) = \{\mathrm{E}[\mathbf{X}]\}$, so that the zonoid depth function is uniquely maximized at the expectation.

- **Simplicial volume Depth**: Zuo & Serfling (2000a) generalized the multivariate median from Oja (1983) into a depth function. Let $\Delta S(\mathbf{X}_1, \ldots, \mathbf{X}_{d+1})$ denote the volume of the d-dimensional simplex.

Definition 1.11. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $P \in \mathcal{P}$. The *simplicial volume depth* of $\boldsymbol{x}$ with respect to P of order $\alpha \ge 1$ is

$$D_{SV}^\alpha(\boldsymbol{x}, P) = \left(1 + E\left[(\Delta S(\boldsymbol{x}, \mathbf{X}_1, \ldots, \mathbf{X}_d))^\alpha \right] \right)^{-1},$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_{d+1}$ are i.i.d. random vectors with common distribution P.

Standardizing the simplicial volume with $(\det(\boldsymbol{\Sigma}(P)))^{1/2}$, for some affine-equivariant scatter functional $\boldsymbol{\Sigma}(P)$, ensures affine-invariance of the resulting depth function.

- L^p **Depth**: Introduced in Zuo & Serfling (2000a) in its general version, this notion uses general ways to measure distance via the L^p norm (recall that, for vectors $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id})' \in \mathbb{R}^d$, $i = 1, 2$, $||\boldsymbol{x}_1 - \boldsymbol{x}_2||_p = \left(\sum_{j=1}^d (x_{1j} - x_{2j})^p \right)^{1/p}$).

Definition 1.12. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $\mathbf{X}$ be a random vector with distribution $P \in \mathcal{P}$. The L^p *depth* of $\boldsymbol{x}$ in $\mathbf{X}$ is

$$D_{L^p}(\boldsymbol{x}, P) = \left(1 + E\left[||\boldsymbol{x} - \mathbf{X}||_p \right] \right)^{-1}.$$

Despite its name, this depth function remains of a "spatial" nature (as only the p-distance—and not a power of it—is used in the expectation). In particular, the deepest point of the L^2 depth is the spatial median from Definition 1.1. It is also common to use the *standardized L^2 depth*, defined as $D_{\widetilde{L^2}}(\boldsymbol{x}, P) = (1 + E[d_{\mathrm{Cov}(P)}(\boldsymbol{x}, \mathbf{X})])^{-1}$.

- **Spatial Depth**: There exists some confusion between the L^2 depth introduced above and the depth function proposed in Vardi & Zhang (2000), where the authors actually proposed a

canonical way to associate to any multivariate median $\hat{\boldsymbol{\mu}}(.) : \mathcal{P} \to \mathbb{R}^d$ a depth function (quite the reverse direction from the usual one that defines a multivariate median as the point with maximum depth).

Definition 1.13. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $P \in \mathcal{P}$. Let $\hat{\boldsymbol{\mu}}(.)$ be a multivariate median. The associated depth $D_{\hat{\boldsymbol{\mu}}}(\boldsymbol{x}, P)$ of $\boldsymbol{x}$ with respect to P is $1 - w(\boldsymbol{x})$ where

$$w(\boldsymbol{x}) = \inf \left\{ \eta \mid \hat{\boldsymbol{\mu}}(P_{\boldsymbol{x}}^{\eta}) = \boldsymbol{x} \right\}$$

is the smallest incremental mass η at location $\boldsymbol{x}$ needed for $\boldsymbol{x}$ to become the median of the resulting mixture $P_{\boldsymbol{x}}^{\eta} = (\eta \delta_{\boldsymbol{x}} + P)/(1 + \eta)$, where $\delta_{\boldsymbol{x}}$ denotes the point mass at $\boldsymbol{x}$.

Provided that the multivariate median is such that $\hat{\boldsymbol{\mu}}(P) = \boldsymbol{y}$ as soon as $P[\{\boldsymbol{y}\}] \geq 1/2$, the resulting depth function is nonnegative and well defined for all $\boldsymbol{x} \in \mathbb{R}^d$. Taking $\hat{\boldsymbol{\mu}}$ as in Definition 1.1 gives rise to the *spatial depth* $D_{Sp}(\boldsymbol{x}, P)$, the properties of which were only partially explored.

- **Spatial rank Depth**: Gao (2003) suggested the use of spatial ranks to define a depth notion. Let $S(\boldsymbol{x}) = (\boldsymbol{x}/||\boldsymbol{x}||)\mathbf{1}_{\mathbb{R}_0^d}(\boldsymbol{x})$ be the multivariate sign function. The spatial rank of $\boldsymbol{x}$ with respect to the random vector $\mathbf{X}$ having distribution $P \in \mathcal{P}$ is $R(\boldsymbol{x}, P) = \mathrm{E}\big[S(\mathbf{X} - \boldsymbol{x})\big]$.

Definition 1.14. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $P \in \mathcal{P}$. The *spatial rank depth* of $\boldsymbol{x}$ with respect to P is $D_{SR}(\boldsymbol{x}, P) = 1 - ||R(\boldsymbol{x}, P)||^2$.

In the same spirit, Serfling (2002a) used the spatial quantiles from Chaudhuri (1996) to define an associate spatial (quantile) depth function.

- **Spherical Depth**: The proposed depth concept from Elmore *et al.* (2006) provides a good balance between computational tractability and sound statistical properties in any dimension.

Definition 1.15. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $P \in \mathcal{P}$. The *spherical depth* of $\boldsymbol{x}$ with respect to P is

$$D_{\mathrm{Spher}}(\boldsymbol{x}, P) = P\big[\boldsymbol{x} \in S(\mathbf{X}, \mathbf{Y})\big],$$

where $\mathbf{X}$ and $\mathbf{Y}$ are independent and P-distributed and $S(\mathbf{X}, \mathbf{Y}) = \{\boldsymbol{x} : ||\boldsymbol{x} - (\mathbf{X} + \mathbf{Y})/2|| \leq ||\mathbf{X} - \mathbf{Y}||/2\}$ denotes the unique, closed random hypersphere for which the segment $\overline{\mathbf{XY}}$ forms a diameter.

A transformation-retransformation method (in the spirit of Chakraborty *et al.*, 1998) producing affine-invariant version of this depth is also proposed in the same paper. Note that this concept (on the contrary to what its denomination may imply) does not apply to directional data (see Section 2.10 for more details about "directional depths").

- **Lens Depth**: Related to the spherical depth, the proposal from Liu & Modarres (2011) has been recently introduced and thoroughly studied. Let $L(\mathbf{X}, \mathbf{Y})$ denote the intersection of the two closed balls with radius $||\mathbf{X} - \mathbf{Y}||$, centered at $\mathbf{X}$ and $\mathbf{Y}$, respectively. Again, substituting Mahalanobis distances to the Euclidian ones will allow for an affine-invariant concept.

Definition 1.16. Let $\boldsymbol{x} \in \mathbb{R}^d$ and $P \in \mathcal{P}$. The *lens depth* of $\boldsymbol{x}$ with respect to P is

$$D_L(\boldsymbol{x}, P) = P\big[\boldsymbol{x} \in L(\mathbf{X}, \mathbf{Y})\big],$$

where $\mathbf{X}$ and $\mathbf{Y}$ are independent and P-distributed.

A few other depth functions were defined elsewhere in the literature but were not considered in the list above because they do not meet one of the following natural requirements: (i) Although locating the center of a data set is important, the notion of depth should be defined as generally as possible and should, in particular, be able to deal with continuous distributions P. (ii) The argument is actually valid the other way around as depth should not be only limited to the latter type of distributions. (iii) Finally, depth functions that, even under the strongest hypothesis of symmetry on the distribution (that is, in circumstances where the center can be unequivocally defined) may fail to assign maximal value to the symmetry center should not be considered.

The incriminated depth functions were the likelihood/probing depth from Fraiman *et al.* (1997) and Fraiman & Meloche (1999), the interpoint distance depth from Lok & Lee (2011) (see also Bartoszyński *et al.*, 1997) or the convex hull peeling depth and the proximity depth (also known as Delaunay depth) from Hugg *et al.* (2006).

1.3 Statistical depth functions – A paradigmatic approach

Each definition in the previous section has its own advantages and drawbacks (depending, also, on the objectives at hand), so that one might find it difficult to know which depth function to use. To discriminate between the many depth definitions, Zuo & Serfling (2000a) stated four desirable properties that depth functions should ideally satisfy. Without loss of generality, only non-negative and bounded functions are considered. The four properties are

(P1) *affine-invariance*: The depth of a point $\boldsymbol{x} \in \mathbb{R}^d$ should not depend on the underlying coordinate system nor on the scales used;

(P2) *Maximality at center*: For a symmetric distribution, the depth function should attain its maximum value at the center of symmetry;

(P3) *Monotonicity relative to deepest point*: For a distribution possessing a unique deepest point, the depth of a point $\boldsymbol{x} \in \mathbb{R}^d$ should be decreasing as $\boldsymbol{x}$ moves away along any ray from that point;

(P4) *Vanishing at infinity*: The depth of a point $\boldsymbol{x} \in \mathbb{R}^d$ should converge to zero as $||\boldsymbol{x}||$ approaches infinity.

The notion of symmetry used in Property (P2), although defined unambiguously in the univariate case, may differ from one concept to another. They include, in decreasing order of generality,
-*Halfspace symmetry*: A random vector $\mathbf{X}$ is halfspace symmetric about $\boldsymbol{\mu}$ if $P[H] \geq 1/2$ for any closed halfspace containing $\boldsymbol{\mu}$,
-*Angular symmetry*: A random vector $\mathbf{X}$ is angularly symmetric about $\boldsymbol{\mu}$ if $(\mathbf{X} - \boldsymbol{\mu})/||\mathbf{X} - \boldsymbol{\mu}|| \stackrel{d}{=}$ $(\boldsymbol{\mu} - \mathbf{X})/||\mathbf{X} - \boldsymbol{\mu}||$, where $\stackrel{d}{=}$ denotes equality in distribution,
-*Central symmetry*: A random vector $\mathbf{X}$ is centrally symmetric about $\boldsymbol{\mu}$ if $\mathbf{X} - \boldsymbol{\mu} \stackrel{d}{=} \boldsymbol{\mu} - \mathbf{X}$, and
-*Spherical symmetry*: A random vector $\mathbf{X}$ is spherically symmetric about $\boldsymbol{\mu}$ if $(\mathbf{X} - \boldsymbol{\mu}) \stackrel{d}{=} \mathbf{O}(\mathbf{X} - \boldsymbol{\mu})$ for any orthogonal matrix $\mathbf{O}$.

Let $\mathcal{P}$ denote the set of all distributions on $\mathbb{R}^d$ and $P_{\mathbf{X}}$ the distribution of the random vector $\mathbf{X}$. In view of the previous requirements, Zuo & Serfling (2000a) adopted the following definition of *statistical depth function*.

Definition 1.17. The bounded mapping $D(.,.) : \mathbb{R}^d \times \mathcal{P} \to \mathbb{R}^+$ is called a *statistical depth function* if it satisfies the four following properties :

(P1) for any $d \times d$ invertible matrix $\mathbf{A}$, any d-vector $\mathbf{b}$, any $\boldsymbol{x} \in \mathbb{R}^d$ and any random vector $\mathbf{X} \in \mathbb{R}^d$,
$D(\mathbf{A}\boldsymbol{x} + \mathbf{b}, P_{\mathbf{A}\mathbf{X}+\mathbf{b}}) = D(\boldsymbol{x}, P_{\mathbf{X}})$;

(P2) if $\boldsymbol{\mu}$ is a center of (central, angular or halfspace) symmetry of $P \in \mathcal{P}$, then it holds that
$D(\boldsymbol{\mu}, P) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}, P)$;

(P3) for any $P \in \mathcal{P}$ having deepest point $\boldsymbol{\mu}$, $D(\boldsymbol{x}, P) \leq D((1-\lambda)\boldsymbol{\theta} + \lambda\boldsymbol{x}, P)$ for any $\boldsymbol{x}$ in $\mathbb{R}^d$ and any $\lambda \in [0, 1]$;

(P4) for any P, $D(\boldsymbol{x}, P) \to 0$ as $||\boldsymbol{x}|| \to \infty$.

Other proposals of such paradigmatic approach to depth functions have been introduced elsewhere in the literature. Comparison of depth functions based on different criterions, among which the "stochastic order preservation" was provided in Zuo (2003). Also, Dyckerhoff (2002) (see also Mosler, 2012) did not use property (P2) but also added the technical property

(P5) *upper semicontinuity*: For any $P \in \mathcal{P}$, the upper level sets $D_\alpha(P) = \{\boldsymbol{x} \in \mathbb{R}^d | D(\boldsymbol{x}, P) \geq \alpha\}$ are closed for all $\alpha > 0$.

Under (P3), the sets $D_\alpha(P)$ (commonly known as the *depth regions*) are nested and star-shaped about the deepest point, should it exist. They are of particular interest, as they bring much information about the spread, shape and symmetry of the underlying distribution (see Serfling, 2004). The *depth contours* (boundary of the depth regions) even characterize, under very mild conditions, the underlying distribution[9] (see Kong & Zuo (2010) and references therein). When the depth function $D(.,.)$ satisfies the more stringent assumption

(P3') *Quasiconcavity*: For any $P \in \mathcal{P}$, $D(., P)$ is a quasiconcave function, that is, its upper level sets $D_\alpha(P)$ are convex for all $\alpha > 0$,

the resulting depth function is often called a *convex statistical depth function*. Some depth functions might fail to satisfy some properties for all distributions $P \in \mathcal{P}$ or, in Property (P1), for all invertible $d \times d$ matrix $\mathbf{A}$. Definition 1.17 being the most widely used in the literature, we will restrict to Properties (P1)-(P4) to describe a statistical depth function. All depth functions (from Definition 1.5 to 1.16) introduced in the previous section are statistical depth functions in that sense, although some restrictions may be required.

(i) **Restrictions on $\mathcal{P}$**: Some depths (see Definitions 1.8, 1.9 and 1.10) require distribution P to have finite (first- or second-order) moments. Furthermore, a few depth functions may fail to satisfy one or more properties under *discrete* distributions. In particular, the simplicial depth D_S will only be a statistical depth function when considered as a mapping $D : \mathbb{R}^d \times \mathcal{P}_c \to \mathbb{R}^+$, where $\mathcal{P}_c$ denotes the set of continuous distributions on $\mathcal{B}^d$. Note also that Definition 1.7 already required P to be continuous as the halfspace $H^P_{\mathbf{X}_1,...,\mathbf{X}_d}$ may fail to be properly defined otherwise.

(ii) **The symmetry used**: Only few depth functions (that typically are not based on distances but rather on halfspaces) satisfy (P2) under the broadest assumption of halfspace symmetry: D_H, D_P and D_{Maj}. The same property holds for $D_{\tilde{L}^2}$, D_S and D_{Sp} under the slightly stronger assumption of angular symmetry, while central symmetry is required for D_{SV}, D_Z and D_L to fulfil (P2). Maximality at center for D_{Spher} and D_{SR} has only be proved under the assumption of spherical symmetry. Note also that property (P2) will be fulfilled for D_{Mah} under a symmetric distribution P as soon as $\boldsymbol{\mu}(P)$ coincides with the center of (halfspace, angular or central) symmetry of P.

---

[9]Partial results on the question whether the depth function uniquely determines the underlying distribution are available in the literature: see Struyf & Rousseeuw (1999); Koshevoy (2002, 2003); Mosler & Hoberg (2006); Hassairi & Regaieg (2008).

(iii) **Orthogonal statistical depth functions**: Some depth functions, namely D_{SV}, D_{L^2}, D_{Sp}, D_{Spher} and D_L, only satisfy (P1) for orthogonal matrices. As already discussed, affine-invariant versions are typically obtained by substituting affine-invariant distances to the Euclidian ones used in their definition. This, however, may affect their robustness properties.

Upper semicontinuity (P5) was proved to actually hold for all the depth functions introduced here (see Liu & Singh, 1993; Mizera & Volauf, 2002; Mosler, 2012; Gao, 2003; Elmore *et al.*, 2006; Liu & Modarres, 2011). Results about convexity are more sparse. (P3') holds for D_H (Rousseeuw & Ruts, 1999) as well as for D_{Mah}, D_P, D_{SV} or D_Z (see, e.g., Mosler, 2012) but does not for simplicial depth D_S.

Interestingly, Zuo & Serfling (2000a) also identified four general structures of depth functions and derived the properties that such general functions should satisfy.

(A) Let $h(\boldsymbol{x}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ be a *bounded* and non-negative function measuring, in some sense, the "closeness" of $\boldsymbol{x} \in \mathbb{R}^d$ to the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. The corresponding *Type A depth function* measures the average proximity of $\boldsymbol{x}$ to a random sample of size k and is defined by $D(\boldsymbol{x}, P) = \mathrm{E}\big[h(\boldsymbol{x}; \mathbf{X}_1, \ldots, \mathbf{X}_k)\big]$, for $\mathbf{X}_1, \ldots, \mathbf{X}_k$ a random sample from P.

(B) Let $h(\boldsymbol{x}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ be an *unbounded* non-negative function measuring the "distance" of $\boldsymbol{x} \in \mathbb{R}^d$ from the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. The corresponding *Type B depth function* is $D(\boldsymbol{x}, P) = \big(1 + \mathrm{E}\big[h(\boldsymbol{x}; \mathbf{X}_1, \ldots, \mathbf{X}_k)\big]\big)^{-1}$.

(C) Let $O(\boldsymbol{x}, P)$ be a measure of *outlyingness* of $\boldsymbol{x} \in \mathbb{R}^d$ with respect to the distribution P. If $O(\boldsymbol{x}, P)$ is unbounded, then the corresponding bounded *Type C depth function* is $D(\boldsymbol{x}, P) = \big(1 + O(\boldsymbol{x}, P)\big)^{-1}$.

(D) Let $\mathcal{C}$ be a class of closed subsets of $\mathbb{R}^d$ satisfying the two conditions that (i) if $C \in \mathcal{C}$, then $\bar{C}^c \in \mathcal{C}$ and (ii) for $C \in \mathcal{C}$ and $\boldsymbol{x} \in C^\circ$, there exists $C_1 \in \mathcal{C}$ with $\boldsymbol{x} \in \partial C_1$, $C_1 \subset C^\circ$, where ∂C, C^c, C° and $\bar{C}$ denote, respectively, the boundary, complement, interior and closure of C. The corresponding *Type D depth function* is $D(\boldsymbol{x}, P; \mathcal{C}) = \inf_{C \in \mathcal{C}} \big\{ P[C] : \boldsymbol{x} \in C \big\}$.

2   Around location depth and beyond

Many studies and different directions of possible generalizations of the concept of (location) depth have been provided in the past decades. From extending depth functions to other parametric or nonparametric setups (including regression) to using depth as a tool for classification (to name but a few), this section provides an extended look at the many fields depth has been applied to, as well as the state-of-the-art about, among others, robustness, asymptotics or computational aspects. Nonetheless, the first sections are devoted to a brief discussion on sample depth and the depth-based location functionals.

2.1   Sample depth

Sample versions of any depth function $D(.,.)$ from Section 1.2 can be obtained by replacing the distribution of the random vector $\mathbf{X} \sim P$ by the empirical distribution $P^{(n)}$ of the i.i.d. (having common distribution P) random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$. This naturally leads to the finite-sample counterpart $D^{(n)}(\boldsymbol{x}) = D(\boldsymbol{x}, P^{(n)})$ of $D(\boldsymbol{x}, P)$.

An important and standard requirement (although, somewhat surprisingly, not embedded in the paradigmatic foundations of statistical depth functions in Section 1.3) is that sample depth converges to its population counterpart. More precisely, it is desirable that, for a depth function D and for a fixed distribution P, almost surely as $n \to \infty$,

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} \left| D^{(n)}(\boldsymbol{x}) - D(\boldsymbol{x}, P) \right| \to 0. \tag{2.1}$$

Strong uniform consistency of the sample depth function is of natural interest but also plays a crucial role for other purposes, see Section 2.2. Results concerning (2.1) are available for several depth notions. Consistency of sample halfspace depth D_H was proved in Donoho & Gasko (1992), while the same property for simplicial depth D_S has been established in Liu (1990), Dümbgen (1992) and Arcones & Giné (1993). Actually, Dümbgen (1992) proved the stronger property

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} \left| D_S(\boldsymbol{x}, P') - D_S(\boldsymbol{x}, P) \right| \le (d-1) ||P' - P||_{\mathcal{H}},$$

where $||.||_{\mathcal{H}}$ denotes the Kolmogorov-Smirnov norm with respect to the set of all intersections of d open halfspaces in $\mathbb{R}^d$.

Under suitable conditions on P, Liu & Singh (1993) showed that (2.1) holds for the sample majority depth D_{Maj} as well as for the Mahalanobis depth D_{Mah}. The same property holds true for the sample projection depth (where the location and scatter functional used are the median and the MAD, respectively), as well as for all Type D depth functions, see Zuo & Serfling (2000c). Consistency of the zonoid depth is proved in Mosler (2002).

2.2 Depth-based estimators and asymptotics

As described earlier, location depth provides a measure of centrality, and is therefore an appropriate tool that will allow to define new estimates of location. These are essentially of two types.

First, the location in $\mathbb{R}^d$ with maximal depth will be naturally called *depth-based median*. The multivariate medians from Section 1.1 are obtained in that manner based on the depth functions from Section 1.2 (see Definitions 1.1, 1.2 and 1.3 and their related depth in Definitions 1.12, 1.11 and 1.8). Secondly, due to the natural ordering depth provides, authors also introduced depth-weighted L statistics, that often consists in averaging a given proportion of the "most central" points and are therefore known as *depth-based trimmed means*.

Studying the asymptotics of such estimators (or these of the depth regions they are founded on) is proving difficult, due to the complex nature of the depth function on which they are defined and typically requires U-process theory as well as strong results on empirical processes. A brief overview of the existing results is now provided.

Nolan (1992) established asymptotic properties of some univariate trimmed means based on halfspace depth. Root-n consistency of the bivariate halfspace median under suitable conditions on the underlying distribution is considered in Nolan (1999). Extension of this result to higher dimensions (together with the asymptotic distribution—characterized through a max-min operation of a continuous process—of the maximal regression depth estimator) was provided in Bai & He (1999). Root-n consistency (to a functional of a Brownian motion) of the regression maximum depth estimator was already considered in He & Portnoy (1998).

The simplicial median and simplicial volume median, as locations maximizing U-processes, are proved to be asymptotically normal in Arcones *et al.* (1994) (central limit theorems for U-processes are available in Arcones & Giné, 1993). Similar results for the simplicial volume median were actually already available in Oja & Niimimaa (1985). Dümbgen (1992) derived the asymptotic normality of simplicial depth-based trimmed means after establishing a central limit theorem for the associated empirical depth process. Asymptotics of a halfspace-based trimmed mean were considered in Massé & Theodorescu (1994).

The asymptotic behavior of the corresponding halfspace process was studied in Massé (2004) where it was proved that it may fail to converge weakly. A necessary and sufficient condition for the asymptotic normality of a special class of depth-based trimmed means is provided. Two other types of L statistics were considered and studied in Massé (2009). Asymptotics of trimmed mean estimators based on projection depth can be found in Zuo *et al.* (2004b).

Results for depth regions can also be found in the literature. Uniform consistency of the contours (in the elliptical setup) under conditions on the underlying depth measure is proved in He & Wang (1997). Root-n consistency of these contours is proved using empirical process theory and U-process theory in Kim (2000). As a corollary, root-n consistency of a trimmed mean based on Oja's depth (already defined in Kim, 1992) is obtained.

A seminal paper about convergence of the depth contours and regions, related to the consistency properties of the underlying depth functions is Zuo & Serfling (2000c).

2.3 Regression and parametric depths

The successful story of depth in location has motivated extending the concept to other parametric setups. Several proposals exist in the regression model and a full parametric approach to the notion of centrality has been developed in the early 2000's.

Regression depths Parallel to the extension from the univariate to the multivariate median, where a structural property of the one-dimensional median serves as ground to define a multidimensional counterpart, alternative characterizations of (halfspace) depth are required for proper generalisation. A first equivalence result can be found in Carrizosa (1996), where it is proved that

$$D_H(\boldsymbol{x}, P) = \inf_{\boldsymbol{y} \in \mathbb{R}^d} P\Big(\big\{\boldsymbol{a} : |\boldsymbol{y} - \boldsymbol{a}| \geq |\boldsymbol{x} - \boldsymbol{a}|\big\}\Big),$$

that is, the halfspace depth of $\boldsymbol{x}$ is the smallest probability (among all fixed choices of $\boldsymbol{y}$) of the set of points that are closer to $\boldsymbol{x}$ than to $\boldsymbol{y}$. The latter equality allows extension to problems with non-Euclidian metrics or dissimilarity measures $\delta(\boldsymbol{x}, \boldsymbol{y})$ by defining the depth of an element $\boldsymbol{x}$ as

$$D_\delta(\boldsymbol{x}, P) = \inf_{\boldsymbol{y}} P\big[\big\{\boldsymbol{a} : \delta(\boldsymbol{y}, \boldsymbol{a}) \geq \delta(\boldsymbol{x}, \boldsymbol{a})\big\}\big].$$

Parallel extension to the (single-output) regression setup goes as follows. Given a probability measure P on $\mathbb{R}^d \times \mathbb{R}$, corresponding to a multivariate random variable $(\mathbf{X}, Y)'$, the depth of the hyperplane $H_{\boldsymbol{a},b} \equiv \boldsymbol{y} = \boldsymbol{a}'\boldsymbol{x} + b$ is defined as

$$D_{\mathrm{Reg}}(H_{\boldsymbol{a},b}, P) = \inf_{(\boldsymbol{c},d) \in \mathbb{R}^d \times \mathbb{R}} P\Big(\big\{(\boldsymbol{x}, y) \in \mathbb{R}^d \times \mathbb{R} : |y - \boldsymbol{a}'\boldsymbol{x} - b| \geq |y - \boldsymbol{c}'\boldsymbol{x} - d|\big\}\Big).$$

Both depths above received very little attention in the literature as only few properties were explored.

Equivalently, in the sample case, Tukey depth of $\boldsymbol{x}$ can also be seen as the minimal relative number of points that need to be removed before $\boldsymbol{x}$ ceases to be a Pareto optimum (for the distance function $f(\boldsymbol{x},.) = ||\boldsymbol{x} - .||$) with respect to the remaining dataset[10]. This motivates the celebrated regression depth introduced in Rousseeuw & Hubert (1999), admitting the following definition in the sample case.

Definition 2.1. The *regression depth* $D_R(H_{\boldsymbol{a},b}, P^{(n)})$ of an hyperplane $H_{\boldsymbol{a},b}$ with respect to the dataset $\{(\boldsymbol{x}_i, y_i), 1 \leq i \leq n\} \subset \mathbb{R}^d \times \mathbb{R}$ (whose empirical distribution is denoted $P^{(n)}$) is the minimal relative number of points that need to be removed to make $(\boldsymbol{a}, b)$ a non-Pareto optimum of the residual function $f : (\mathbb{R}^d \times \mathbb{R})^2 \to \mathbb{R}^+ : ((\boldsymbol{a}, b), (\boldsymbol{x}, y)) \to |y - \boldsymbol{a}'\boldsymbol{x} - b|$ with respect to the remaining dataset.

As usual, maximizing the depth function will provide a median-type estimate of regression. Bounds on the minimal depth of this estimator are provided, for the bivariate case, in Rousseeuw & Hubert (1999), through the construction of the "catline", an hyperplane with minimal depth $1/3$. Although the population version of the latter definition does not seem easy to define, it will be given as a by-product of the general tangent depth introduced below.

Parametric depth Mizera (2002) based on the same ideas of Pareto optimality a concept of *global depth*, that extends location and regression depths to an arbitrary parametric model. To describe this, consider a random d-vector $\mathbf{X}$ with a distribution $P = P_{\boldsymbol{\theta}_0}$ in the parametric family $\mathcal{P} = \{P_{\boldsymbol{\theta}} | \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k\}$ (k may differ from d). Let $(\boldsymbol{\theta}, \boldsymbol{x}) \mapsto F_{\boldsymbol{\theta}}(\boldsymbol{x})$ be a mapping that measures the "quality" of the parameter value $\boldsymbol{\theta}$ for the observation $\boldsymbol{x}$. A natural definition of depth is

Definition 2.2. Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The *global depth* of $\boldsymbol{\theta}$, $D_G(\boldsymbol{\theta}, P^{(n)})$, with respect to the empirical distribution of the i.i.d random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with common distribution $P \in \mathcal{P}$ is the minimal relative number of points that need to be removed before $\boldsymbol{\theta}$ is no longer a Pareto optimum of $f(\boldsymbol{\theta}, \boldsymbol{x}) = F_{\boldsymbol{\theta}}(\boldsymbol{x})$ with respect to the remaining dataset.

Now, while this definition still remains uninspiringly limited to the sample case, the following restriction will allow a full treatment of parametric depth. Assuming that the objective function $F_{\boldsymbol{\theta}}(\boldsymbol{x})$ is differentiable and convex with respect to $\boldsymbol{\theta}$, it is easy to show (see Mizera (2002) for details) that $D_G(\boldsymbol{\theta}, P^{(n)}) = \min_{||\boldsymbol{u}||=1} \sharp\{i : \boldsymbol{u}'\nabla_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(\mathbf{X}_i) \geq 0\} = D_H(\mathbf{0}_k, P^{(n)}_{\nabla_{\boldsymbol{\theta}}F})$, where $P^{(n)}_{\nabla_{\boldsymbol{\theta}}F}$ denotes the empirical distribution of $\nabla_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(\mathbf{X}_i)$, $1 \leq i \leq n$ and $\mathbf{0}_k = (0, \ldots, 0)' \in \mathbb{R}^k$. This amounts to looking at the depth of $\mathbf{0}_k$ among the directions $\nabla_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(\mathbf{X}_i)$, $i = 1, \ldots, n$, of maximal increase of $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}(\boldsymbol{x})$.

The following concept then typically attributes large depth to "good" parameter values, that is, to parameter values $\boldsymbol{\theta}$ that are close to $\boldsymbol{\theta}_0$.

Definition 2.3. The *tangent depth* of $\boldsymbol{\theta}$ with respect to $P \in \mathcal{P}$ is $D_T(\boldsymbol{\theta}, P) = D_H(\mathbf{0}, P_{\nabla_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(\mathbf{X})})$, where $P_{\nabla_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(\mathbf{X})}$ denotes the distribution of $\nabla_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(\mathbf{X})$ under $\mathbf{X} \sim P$.

Tangent depth reduces to the particular cases of classical (location) halfspace depth for $\boldsymbol{\theta} = \boldsymbol{x}$ and $f(\boldsymbol{x}, \boldsymbol{y}) = ||\boldsymbol{x} - \boldsymbol{y}||$, and of regression depth, for which $\boldsymbol{\theta} = (\boldsymbol{a}, b)$ and $f((\boldsymbol{a}, b), (\boldsymbol{x}, y))$ is as in

---

[10]Recall that a point $\boldsymbol{x}$ is a Pareto optimum for the function $f(\cdot, \cdot)$ with respect to some dataset $\mathbf{A}$ if there exists no $\boldsymbol{y}$ such that $f(\boldsymbol{y}, \mathbf{a}) \leq f(\boldsymbol{x}, \mathbf{a})$ for all $\mathbf{a} \in \mathbf{A}$, with a strict inequality for at least one element of $\mathbf{A}$.

Definition 2.1. Actually, any modification of the objective function F to $h(F)$, with $h : \mathbb{R}^+ \to \mathbb{R}^+$ smooth and monotone increasing would lead to the same definition.

A particular example of application of tangent depth can be found in Mizera & Müller (2004), where *location-scale* depth of $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+$ with respect to the univariate distribution P is developed, based on Definition 3.2, and explored. Interestingly, the proposed depth can be seen as the bivariate halfspace depth of projected observations in the Poincaré plane model embedded with the Lobachevski geometry.

Choosing appropriately the objective function $(\boldsymbol{\theta}, \boldsymbol{x}) \mapsto F_{\boldsymbol{\theta}}(\boldsymbol{x})$ may be difficult in some setups. Denoting $L_{\boldsymbol{\theta}}(\boldsymbol{x})$ for the likelihood function, the general, *likelihood-based*, approach consists in taking $F_{\boldsymbol{\theta}}(\boldsymbol{x}) = -\log L_{\boldsymbol{\theta}}(\boldsymbol{x})$; see, e.g., Mizera & Müller (2004); Müller (2005). In the location and regression cases considered above, it can be seen that Gaussian or t_ν-likelihoods lead to the halfspace and regression depth, respectively. The same result holds true for the location-scale depth of Mizera & Müller (2004).

2.4 Robustness

It is commonly accepted in the literature that "depth functions are robust". Although this might in fact be a hasty shortcut, this saying holds true for many depth-based inference procedures.

Robustness of halfspace depth is well studied and many results about the breakdown point, influence function or maximum bias are available.

Donoho & Gasko (1992) (following the seminal work in Donoho (1982)) proved that the (finite-sample enlargement) breakdown point of the halfspace median was at least $1/(d+1)$ for general distributions, while this result could be refined to an asymptotic value as high as $1/3$ for i.i.d. random vectors with a common centrosymmetric distribution. These results were extended in Chen (1995b), where the author provides sharp lower and upper bounds for the limiting (finite-sample as well as population) breakdown point for general (in particular, possibly assymetric) distributions in $\mathbb{R}^d$. The results yield, in particular, that, for $d = 2$, the exact breakdown point of the halfspace median is $1/3$, whatever the underlying distribution may be.

Influence function, maximum bias (hence also breakdown point) and contamination sensitivity[11] of the halfspace median are provided in Chen & Tyler (2002), for absolutely continuous and halfspace symmetric distributions. Notably, the influence function is showed to be bounded (resulting in a finite gross-error sensitivity) and, similar to the univariate median, constant along rays originating from the center of the distribution. Interestingly, the maximum bias is showed to have a relatively simple form, as the greatest distance between the deepest point and some depth contour (the order of which depends on the contamination).

Influence function (and related concepts, together with other results about the shape of the contours, among others) of halfspace depth under multivariate symmetric stable distributions are explored in Chen & Tyler (2004) while both sample and population influence function of halfspace depth, at any location $\boldsymbol{x} \in \mathbb{R}^d$, is derived in Romanazzi (2001). In both cases, the influence functions are proved to be bounded (see also Wang & Serfling (2006) for the influence function in a more general setting). Additional sensitivity analysis showed that, as intuited, inner regions are more stable.

---

[11]Defined as the limiting relative (with respect to the ϵ-contamination) maximum contamination bias as ϵ goes to zero; see Hampel *et al.* (1986).

Zhang (2002) slightly extended the concept of halfspace depth so as to use the information about dispersion of the data and proved that the corresponding estimates keep their asymptotic breakdown point of 1/3 (together with a consistency rate of $n^{-1/2}$).

Similar results that, for most of them, extend the halfspace case, are available for other depth functions. Results similar to those in Donoho & Gasko (1992) were obtained in Chen (1995a) for the simplicial median, which is proved to possess a strictly positive (again, finite-sample) breakdown point but to be less robust (at least from a breakdown point of view) than Tukey's median. Note, however, that Oja's median is proved to have breakdown point 0 (at least in the bivariate case, see Oja *et al.*, 1990). Note also that, together with its full analysis of the halfspace depth influence function, Romanazzi (2001) also provides the sample influence of the simplicial depth. Influence function and maximum bias of projection depth are derived in Zuo (2003) and Zuo *et al.* (2004a) (establishing asymptotic normality by the mean of the influence function for spatial, simplicial and halfspace depth was provided in Dang *et al.*, 2009). Estimators of multivariate location based on the projection depth and generalizing the univariate trimmed means are introduced in Zuo (2006). Influence function and finite-sample breakdown point are investigated, together with asymptotics of the depth trimmed means. Breakdown point of the lens depth is derived in Liu & Modarres (2011).

A more particular study of robustness was conducted recently in Denecke & Müller (2011, 2012), where consistency and robustness of *tests* (rather than estimators) based on depth are considered, by extending to test procedures a new characterization of consistency and breakdown via the *concentration parameter*.

2.5 Classification

The use of depth in classification problems arises naturally from the fact that most classical procedures compare centrality to discriminate between several populations.

The first proposed approach consisted in assigning the point that needs to be classified to the population with respect to which it has the largest depth. This *maxdepth* classification procedure was first proposed in Liu *et al.* (1999) and was then investigated thoroughly in Ghosh & Chaudhuri (2005b). The same construction, using the spatial depth from Vardi & Zhang (2000), was used in Hartikainen & Oja (2006) in their simulation study, comparing various parametric and nonparametric discrimination rules. Dutta & Ghosh (2012a,b) considered maxdepth classifiers based on projection depth or (an affine-equivariant version of) the L^p depth, respectively. Another modification of the maxdepth approach, also based on projection depth and coping better with skewed data has been introduced in Hubert & Van der Veeken (2010).

More recently, refinement of the maxdepth method was proposed in Li *et al.* (2012). The "Depth vs Depth" (DD) classifier is introduced, that consists in constructing appropriate (polynomial) separating curves in the scatter plot of $\{(D_0^{(n)}(\mathbf{X}_i), D_1^{(n)}(\mathbf{X}_i)), 1 \le i \le n\}$[12], rather than using solely the main bisector (as it is the case for the maxdepth approach). Separating curves are chosen so as to minimize the empirical misclassification rate on the training sample and the order of the polynomial defining those curves is chosen by cross-validation. Further modification of the DD-classifiers, that are computationally efficient and applicable in higher dimensions (up to $d = 20$), were introduced in Lange *et al.* (2012). Discriminating between $k \ge 2$ populations is achieved by applying an efficient discrimination algorithm (the $DD\alpha$ procedure) to the k-dimensional zonoid depth plots.

---

[12]$D_j^{(n)}(\mathbf{X}_i)$ denotes here the depth of $\mathbf{X}_i$ with respect to the data points coming from Population j.

Other depth-based classifiers were introduced in Mosler & Hoberg (2006); Cui *et al.* (2008) and Billor *et al.* (2008), where the authors used, respectively, the zonoid depth, projection depth ideas and transvariation probabilities of data depth to develop new classification procedures. Robustification (using the depth contours) of the classification techniques based on convex hulls from Hardy & Rasson (1982) were developed in Ruts & Rousseeuw (1996).

Link between regression depth and linear discriminant analysis was also explored in Christmann & Rousseeuw (2001) and Christmann *et al.* (2002), without giving rise to a proper "depth-based" classifier.

Results about unsupervised classification are seldom available in the literature. Using data depth as a visualization tool, Jörnsten *et al.* (2002) proposed a new clustering procedure. Jörnsten (2004), in order to analyze microarray gene expression data, introduced appropriate clustering techniques based on the L^2 depth and the Relative Data Depth plot.

2.6 Functional depth

In the past few years, the use of real time monitoring in many different fields such as stock markets, quality control, or medicine and the increase in quantity of available data justified the development of new statistical methods, better suited to tackle these "large dimensional" problems. In practice, it has long been advocated (see, for example, Ramsay & Silverman, 2005; Ferraty & Vieu, 2006) that the use of *functional data* is preferable to that of finite-but-large-dimensional vectors that often would lead to computationally intensive (when possible) procedures. It is therefore quite naturally that various concepts of depth for such data were introduced.

Attempt to generalize the median to the functional setup (very much in the spirit of the L^2 depth) was already found in Kemperman (1987). However, the first proper instance of *functional depth*, aiming at providing an equivalent notion of α-trimmed mean for functional observations, goes back to Fraiman & Muniz (2001). A univariate functional counterpart $I_D(x)$ of any (one-dimensional) depth $D(.)$ is proposed, which aggregates the depth values across the domain of the function x (assumed, without loss of generality, to be in $C([0,1])$, the set of continuous functions on the interval $[0,1]$). This yields the simple definition, denoting as P_t the distribution of the function x at time t, $I_D(x) = \int_0^1 D(x(t), P_t)dt$.

Shortly after, a more graphical and "simplicial in spirit" functional depth was proposed in López-Pintado & Romo (2005). The (partial) *band depth* $S^{(j)}(x)$ is defined as the probability that a random band $V(X_1, \ldots, X_j) = \{(t,y)|t \in [0,1], \min_i X_i(t) \le y \le \max_i X_i(t)\}$ based on j observations entirely contains $G(x)$, the graph of x. The *band depth* is then defined as $S_J(x) = \sum_{j=2}^J S^{(j)}(x)$. A generalized version was introduced in López-Pintado & Romo (2009) where the (generalized) band depth now uses the proportion of the domain for which the band $V(X_1, \ldots, X_j)$ contains $G(x)$ rather than the indicator function $I[G(x) \subset V(X_1, \ldots, X_j))]$. A further extension, obtained by refining the definition of the band used, was provided in López-Pintado & Jörnsten (2007).

In López-Pintado & Romo (2011), the *half-region depth* is introduced, based on the hypograph $hyp(x) = \{(t,y) \in [0,1] \times \mathbb{R} : y \le x(t)\}$ and the epigraph $epi(x) = \{(t,y) \in [0,1] \times \mathbb{R} : y \ge x(t)\}$ and defined as $S_H(x) = \min(P[G(x) \subset hyp(x)], P[G(x) \subset epi(x)])$. Modification of the half-region depth along the same lines of the generalized band depth is also provided.

Both band and half-region depths are shown to provide new, computationally feasible, multivariate depths. The multivariate analogue of the former functional depth is also proved to be a statistical depth function meeting the four key properties stated in Section 1.3.

Cuevas *et al.* (2007) also proposed four new definitions of functional depth, based on some expected kernelized distance to the random curve X (hMode depth) or on random projections of X (random projection depth, double random projection depth and h-modal projection depth). A more detailed study of the functional projection depth can be found in Cuevas & Fraiman (2009). A similar notion, also based on random projections, has been developed in Cuesta-Albertos & Nieto-Reyes (2010).

All depths mentioned above are defined for univariate curves $x \in C([0,1])$ only. Extensions to multivariate functional depth, assessing centrality of multivariate curves $\boldsymbol{x} = (x_1, \ldots, x_d)$, $x_j \in C([0,1])$, $1 \leq j \leq d$ are recent. Ieva & Paganoni (2013) aggregate univariate (modified) band depths, weighting adequately the different components to take into account the possible correlation among the curves. A more appropriate multivariate functional version of halfspace depth was developed in Claeskens *et al.* (2012) as a tool for detecting outlying curves and was later thoroughly studied in Hubert *et al.* (2012). They also studied the benefits from applying functional depth to multivariate functions obtained from univariate ones by adding information about, among other, derivatives, integrals or warping functions.

A comprehensive approach, extending any depth to the functional setup (or, actually to any Banach space E) was introduced in Mosler & Polyakova (2012). The authors provided a general class of functional depths, named Φ-depths, that, very much in the spirit of halfspace depth, define the depth of a curve z as the smallest d-variate depth of $\phi(z)$ with respect to some projected measure on $\mathbb{R}^d$, where $\phi \in \Phi$, a subset of all continuous mappings from E to $\mathbb{R}^d$. Properties inherited from that of the underlying multivariate depth are also studied and a general paradigm, in the same spirit as Zuo & Serfling (2000a) or Dyckerhoff (2004), is developed.

Depth-based classifiers have also been extended to the functional context. López-Pintado & Romo (2006) used distances to the trimmed mean and weighted average distance to provide two new functional classification procedures. Cuevas *et al.* (2007) extended the Ghosh & Chaudhuri (2005a,b) classifiers and compared the performances of five different such generalizations. Cuesta-Albertos & Nieto-Reyes (2008) also studied the performances of their projection-based classifier. Recently Sguera *et al.* (2012) introduced the functional spatial depth and the kernelized functional spatial depth (direct extension of the classical spatial depth) and used them to develop new classification methods. Hlubinka & Nagy (2012) introduced the K-band depth and showed how this new notion can be applied to discriminate between two samples. Inference for functional depth through functional band depth regions, able to analyze the structure of a collection of curves, is considered in López-Pintado & Romo (2007).

Finally, functional outlier detection based on depth was considered in Febrero *et al.* (2008) to identify abnormal nitrogen oxides emission levels. Based on a trimming technique, an iterative procedure is used to account for possible masking effects and a bootstrap construction is applied to determine the threshold under which a depth value of a curve would qualify it as "potential outlier".

2.7 Local depth

The first purpose of depth was to provide a center-outward ordering from the deepest point (the center of the distribution) towards exterior points. Adding the classical property that depth functions decrease along any ray issuing from the center, this generally implies that the depth regions form convex and nested sets, which results in depth being suitable for unimodal and convexly supported distributions only. This restriction has been noticed by various authors (see, among others, Zuo & Serfling, 2000a; Izem *et al.*, 2008; Lok & Lee, 2011).

In order to provide a concept flexible enough to deal with more general distributions, a few extensions became available in the literature, under the name of *local depths*.

Hlubinka *et al.* (2010) proposed a generalised version of halfspace depth, relying on weighting non uniformly the halfspaces appearing in Tukey's definition, that may be considered more appropriate for mixture of distributions or nonsymmetric distributions. Under some restrictions on the weight function, uniform strong consistency (2.1) is proved to hold for weighted halfspace depth in Kotík (2009). A similar extension, rather based on interpoint distances, was provided in Lok & Lee (2011) and showed to respect multimodality in various data configurations. This extension has been proved to be useful in many inference problems, including classification or confidence region construction. Aiming first at outlier detection, Chen *et al.* (2009) introduced a kernelized version of spatial depth that allows to study nonconvex distributions.

Other notions of local depth have been recently introduced in Agostinelli & Romanazzi (2011), where the authors propose localization of both halfspace and simplicial depths. For the former, locality is achieved by substituting finite-width slabs for halfspaces, while, in the latter case, it is obtained by imposing a restriction on the volumes of the simplices considered.

2.8 Testing and Diagnostics

Not only does depth provide many different estimators, it has also been used extensively in testing procedures (see, for example, Zuo & Cui, 2004). The induced center-outward ordering, together with the natural link between depth and multivariate quantiles often provides the necessary tool to carry out rank procedures in the multivariate context. A perfect illustration resides in the analogues of the Wilcoxon's rank-sum and signed-rank tests, based on the simplicial volume median (Oja, 1983) that were introduced and thoroughly studied in Brown & Hettmansperger (1987, 1989), Hettmansperger *et al.* (1992) and Hettmansperger & Oja (1994). Other rank tests based on data depth were also suggested in Liu (1992). Initially introduced as a tool to detect overall discrepancies between two populations, other multivariate versions of Wilcoxon's tests were proposed in Liu & Singh (1993) (see also Zuo & He, 2006). Similarly, Li & Liu (2004) described several nonparametric tests of location and scale differences using the idea of the permutation tests and derived from the DD-plot. The latter tests were, actually, already suggested in Liu & Singh (2003). Kruskal-Wallis-type tests for multivariate multisample procedures (based on several different depths) are developed in Chenouri & Small (2012).

Inference, based on depth, in general parametric models (not only restricted to location) has been made possible by the notion of tangent depth (see Section 2.3). Extending the notion of likelihood depth to that of simplicial likelihood depth, Müller (2005) derived tests for regression in generalized linear models, while Wellmann *et al.* (2009) and Wellmann & Müller (2010b) concentrated

on polynomial regression and multiple regression through the origin. Both procedures were proved to be outlier robust. Simplicial tangent depth was also used in the orthogonal regression context (Wellmann & Müller, 2010a). Consistency and robustness of simplicial likelihood depth-based test was studied in Denecke & Müller (2012).

Using bootstrap methods, Liu & Singh (1997) designed a general methodology to determine P-values in testing hypotheses. Also, confidence regions using depth have been developed, based on depth regions. Yeh & Singh (1997) studied bootstrap confidence regions based on halfspace depth. Lee (2012) used data depth as a tool to aggregate different approaches to confidence region construction in order to provide robust confidence sets.

Rousseeuw & Struyf (2002, 2004) noticed that the halfspace depth of a location $\boldsymbol{\theta}_0$ with respect to a distribution P[13] will attain upper bound $1/2$ if and only if P is angularly symmetric about $\boldsymbol{\theta}_0$. This result holds as well for regression, for a particular concept of regression symmetry. This allowed the introduction of depth-based tests for symmetry in both contexts (see also Dutta *et al.* (2011)). In the same vein, Ley & Paindaveine (2012) extended McWilliams' runs test to define bivariate central symmetry tests based on multivariate simplicial runs. Different methods for testing various symmetry hypotheses, together with approaches for measuring the direction and magnitude of skewness in distributions are reviewed in Serfling (2006b). Scatter measures based on depth were also developed in Zuo & Serfling (2000b).

2.9 Computational aspects

Computing depth (except for more parametric instances such as Mahalanobis depth) has always proved difficult. Providing computationally efficient algorithms is therefore crucial. One example of such efficient method was introduced in Rousseeuw & Ruts (1996), where the authors developed an algorithm computing, in the bivariate case, both halfspace and simplicial depths in $O(n \log n)$ time, rather than the $O(n^2)$ and $O(n^3)$ steps needed for naive algorithms. Still in dimension two, when the interest lies solely on finding the halfspace median (and not on the complete depth field), one could for example use the methods described in Rousseeuw & Ruts (1998). One of the most recent and efficient algorithms computing the halfspace deepest point can be found in Chan (2004). An approximation of Tukey depth, that only takes into account a finite number of random one-dimensional projections, has been studied in Cuesta-Albertos & Nieto-Reyes (2008). Recently, Hallin *et al.* (2010) introduced efficient algorithms to compute halfspace depth contours by defining and studying a new concept of directional multivariate quantiles based on L^1 optimization ideas.

Ruts & Rousseeuw (1996), as well as Miller *et al.* (2003), described exact procedures computing the (halfspace) depth contours. Rousseeuw & Struyf (1998) introduced new algorithms for computing halfspace depth (and, actually, regression depth) in higher dimensions (exactly up to $d = 4$, approximately for $d > 4$). New proposals, using cuts of convex cones with hyperplanes, were developed in Liu & Zuo (2011b). Fukuda & Rosta (2004) provided an algorithm able to compute halfspace depth contours in arbitrary dimensions. There also exist many studies in the field of *computational geometry* (see, e.a., Rafalin & Souvaine (2004); Aloupis *et al.* (2002)). A noticeable proposal can be found in Bremner *et al.* (2008), where algorithms having running times increasing with the depth value (and therefore well suited for outlier detection) are introduced.

---

[13]Note that P may not necessarily possess a density, but cannot put mass on the symmetry center.

The same ideas were used again to compute exactly the projection depth at any location (Liu & Zuo, 2011a; Liu *et al.*, 2011). The latter algorithms can actually be used to compute efficiently depth contours and median in the bivariate case, extending the pointwise exact computation of Zuo & Lai (2011).

Computation of the spatial median (sometimes known as the "Fermat-Weber location problem") has generated a huge literature. A first proposal dates back to Weiszfeld (1937), and nontrivial modifications were proposed in Vardi & Zhang (2001).

Computation of the zonoid depth at a given location was developed in Dyckerhoff *et al.* (1996). Dyckerhoff (2000) provided the algorithm allowing for the computation of the zonoid depth regions in dimension $d = 2$. Extending the computations to dimensions $d > 2$ was done in Mosler *et al.* (2009).

Algorithms for majority depth have only been developed recently. The bivariate case is treated in Chen & Morin (2011), while an approximation in higher dimension is provided via the procedure introduced in Chen & Morin (2012).

2.10 Directional data

Directional data appears in several diverse domains such as ecology, meteorology, earth sciences, or astronomy. They naturally arise in multivariate problems for which (i) the observations are intrinsically unit vectors or in which (ii) the magnitude of the observed vector is irrelevant. Relevance of directional data was emphasised in Mardia (1972, 1975). Absence of well-defined zero-direction, together with the lack of natural ordering, explains why the depth approach provides a coherent and unified framework to study such objects.

Fisher (1985) first pointed out the interest to develop a notion of spherical median as robust alternative to existing location functionals and studied, under unimodality assumption, two spherical analogues of the standard univariate and spatial medians. More details are available in Fisher (1993). Similarly, Ducharme & Milasevic (1987) introduced an estimator of location for rotationally symmetric distributions on the hyperspheres and used it to construct confidence regions for the modal direction of a distribution on an hypersphere.

Three new definitions, that are the directional equivalents to the simplicial, halfspace and L^2 depths, are provided in Liu & Singh (1992). Their interest in classification and in building confidence region is discussed. Angular Tukey depth was already defined in Small (1987).

A more recent notion of depth has been obtained as a by-product of a new definition of quantiles for spherical data by Ley *et al.* (2013). An existing spherical median $\boldsymbol{\theta}_M$ (such as Fisher's) is used to define directional quantiles based on the projected distribution $\mathbf{X}'\boldsymbol{\theta}_M$. The associated depth function satisfies the four properties of Definition 1.17. Note, however, that only rotational invariance is required in this particular setup.

Agostinelli & Romanazzi (2013) introduced new depth functions based on halfspace and simplicial ideas and studied their behavior. Particular attention is paid on depth regions and dispersion measures.

3   Objectives and structure of the thesis

This thesis consists (besides this general Introduction) in three chapters that tackle different problems related to statistical depth. Each chapter constitutes a paper that has been accepted by, is in revision or is in preparation for publication in an international journal.

The **first chapter** proposes a new classification procedure based on depth. More precisely, a k-nearest neighbors type classifier is introduced, based on the center-outward ordering of symmetrized versions of the dataset. Like the many procedures described in Section 2.5, and in contrast to the standard kNN classifier, the proposed method is affine-invariant. Unlike the existing depth-based procedures, for which Bayes consistency is achieved only for elliptical distributions, we show that our proposal is consistent for virtually any absolutely continuous distributions (actually, under the sole assumption that the set of discontinuity points of the associated density has Lebesgue measure zero). Convergence of the misclassification probability to that of the Bayes classifier under this broad class of distributions will be called *nonparametric consistency*, to stress the difference with the stronger property of *universal consistency* of the standard (yet inappropriate) kNN procedure. Finite-sample performances are investigated through simulations and the proposed method is showed to outperform a natural affine-invariant version of kNN, as well as the other competing depth-based procedures from Li *et al.* (2012). Illustration on two real-data examples is also provided and applicability of the depth-based neighborhoods to density estimation and nonparametric regression is shortly discussed. This chapter has been accepted for publication in *Bernoulli*.

The **second chapter** extends the concept of (location) depth to the more general framework of *local depth*, a notion able to get rid of the limitations of its global counterpart. Our construction (in contrast, surprisingly, to the existing notions mentioned in Section 2.7) is achieved by conditioning the distribution to some neighborhoods. We will, actually, make use of the depth-based neighborhoods introduced in the first chapter. The construction applies in a generic way to any statistical depth function[14]. Also, we show that the proposed local depth concepts remain of a genuine depth nature and measure (local) centrality at any locality level. This is in contrast with the other notions that, for extreme localization, converge to either the density or a constant value. Two inferential applications are proposed: first, a more general version of the maxdepth classifiers from Ghosh & Chaudhuri (2005b) based on local depth is introduced and compared to other classical procedures. Also, a new test for (central) symmetry is proposed (see Section 2.8 for (angular) symmetry testing methods). The local depth construction is also extended to the regression and functional depth contexts (see Sections 2.3 and 2.6, respectively). Many properties of the proposed depths, such as affine-invariance, consistency of their sample version, and limit behavior under extreme locality, are derived. Throughout the chapter, we illustrate those results and the interest of local depth on univariate and multivariate, artificial, and real data sets. This chapter has been accepted for publication in the *Journal of the American Statistical Association*.

Finally, the **third chapter** introduces a depth function for the *shape parameter* of elliptical distributions. This parameter, a normalized version of the corresponding scatter matrix, is of interest in many inference problems of multivariate analysis, such as principal component analysis, canonical correlation analysis, tests of sphericity, etc. If the shape parameter is normalized so as to have determinant one, the resulting concept results from the parametric depth construction described in Section 2.3. However, defining a reasonable shape depth concept for other normalizations requires

---

[14]Actually, only Properties (P2) and (P3) from Definition 1.17 are required to properly define the neighborhoods.

developing a semiparametric version of this construction. In particular, we show that the resulting depth is affine-invariant and does not depend (in contrast to its tangent depth version) on the standardization adopted. We also prove that the proposed depth, under elliptical distributions[15], is maximized at the true shape value. We close this chapter by considering depth-based tests for shape and by conducting simulations in order to investigate their finite-sample performances, both in terms of power and robustness. This chapter is in preparation for publication.

---

[15]Actually, shape depth naturally extends to the shape parameter of distributions having elliptical directions.

Chapter I

Nonparametrically Consistent Depth-based Classifiers

Nonparametrically Consistent Depth-based Classifiers[1]

Davy PAINDAVEINE[a,b,*] and Germain VAN BEVER[a,c].

[a] *ECARES & Département de Mathématique, Université libre de Bruxelles (ULB)*
[b] *Email: dpaindav@ulb.ac.be*
[c] *Email: gvbever@ulb.ac.be*
[*] *Corresponding Author. Address: Université libre de Bruxelles, ECARES, Avenue F.D. Roosevelt, 50, CP 114/04, B-1050 Brussels, Belgium.*

Abstract We introduce a class of depth-based classification procedures that are of a nearest-neighbor nature. Depth, after symmetrization, indeed provides the center-outward ordering that is necessary and sufficient to define nearest neighbors. Like all their depth-based competitors, the resulting classifiers are affine-invariant, hence in particular are insensitive to unit changes. Unlike the former, however, the latter achieve Bayes consistency under virtually any absolutely continuous distributions—a concept we call *nonparametric consistency*, to stress the difference with the stronger *universal consistency* of the standard kNN classifiers. We investigate the finite-sample performances of the proposed classifiers through simulations and show that they outperform affine-invariant nearest-neighbor classifiers obtained through an obvious standardization construction. We illustrate the practical value of our classifiers on two real data examples. Finally, we shortly discuss the possible uses of our depth-based neighbors in other inference problems.

Keywords: Affine-invariance · Classification procedures · Nearest Neighbors · Statistical depth functions · Symmetrization

---

1 Introduction

The main focus of this work is on the standard classification setup in which the observation, of the form $(\mathbf{X}, Y)$, is a random vector taking values in $\mathbb{R}^d \times \{0, 1\}$. A classifier is a function $m : \mathbb{R}^d \to \{0, 1\}$ that associates with any value $\boldsymbol{x}$ a predictor for the corresponding "class" Y. Denoting by $\mathbb{I}_A$ the indicator function of the set A, the so-called Bayes classifier, defined through

$$m_{\text{Bayes}}(\boldsymbol{x}) = \mathbb{I}\left[\eta(\boldsymbol{x}) > 1/2\right], \quad \text{with} \quad \eta(\boldsymbol{x}) = P[Y = 1 \,|\, \mathbf{X} = \boldsymbol{x}], \tag{1.1}$$

is optimal in the sense that it minimizes the probability of misclassification $P[m(\mathbf{X}) \neq Y]$. Under absolute continuity assumptions, the Bayes rule rewrites

$$m_{\text{Bayes}}(\boldsymbol{x}) = \mathbb{I}\left[\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} > \frac{\pi_0}{\pi_1}\right], \tag{1.2}$$

where $\pi_j = P[Y = j]$ and f_j denotes the pdf of $\mathbf{X}$ conditional on $[Y = j]$. Of course, empirical classifiers $\hat{m}^{(n)}$ are obtained from i.i.d. copies $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$, of $(\mathbf{X}, Y)$, and it is desirable that such classifiers are consistent, in the sense that, as $n \to \infty$, the probability of misclassification of $\hat{m}^{(n)}$, conditional on $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$, converges in probability to the probability of misclassification of the Bayes rule. If this convergence holds irrespective of the distribution of $(\mathbf{X}, Y)$, the consistency is said to be *universal*.

Classically, parametric approaches assume that the conditional distribution of $\mathbf{X}$ given $[Y = j]$ is multinormal with mean $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$ $(j = 0, 1)$. This gives rise to the so-called *quadratic discriminant analysis (QDA)*—or to *linear discriminant analysis (LDA)* if it is further assumed that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1$. It is standard to estimate the parameters $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ $(j = 0, 1)$ by the corresponding sample means and empirical covariance matrices, but the use of more robust estimators was recommended in many works; see, for example, Randles *et al.* (1978), He & Fung (2000), Dehon & Croux (2001), or Hartikainen & Oja (2006). Irrespective of the estimators used, however, these classifiers fail to be consistent away from the multinormal case.

Denoting by $d_{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{\mu}) = ((\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}))^{1/2}$ the Mahalanobis distance between $\boldsymbol{x}$ and $\boldsymbol{\mu}$ in the metric associated with the symmetric and positive definite matrix $\boldsymbol{\Sigma}$, it is well known that the QDA classifier rewrites

$$m_{\text{QDA}}(\boldsymbol{x}) = \mathbb{I}\left[d_{\boldsymbol{\Sigma}_1}(\boldsymbol{x}, \boldsymbol{\mu}_1) < d_{\boldsymbol{\Sigma}_0}(\boldsymbol{x}, \boldsymbol{\mu}_0) + C\right], \tag{1.3}$$

where the constant C depends on $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_1$, and π_0, hence classifies $\boldsymbol{x}$ into Population 1 if it is sufficiently more central in Population 1 than in Population 0 (centrality, in elliptical setups, being therefore measured with respect to the geometry of the underlying equidensity contours). This suggests that *statistical depth functions*, that are mappings of the form $\boldsymbol{x} \mapsto D(\boldsymbol{x}, P)$ indicating how central $\boldsymbol{x}$ is with respect to a probability measure P (see Section 2.1 for a more precise definition), are appropriate tools to perform nonparametric classification. Indeed, denoting by P_j the probability measure associated with Population j $(j = 0, 1)$, (1.3) makes it natural to consider classifiers of the form

$$m_D(\boldsymbol{x}) = \mathbb{I}\left[D(\boldsymbol{x}, P_1) > D(\boldsymbol{x}, P_0)\right],$$

based on some fixed statistical depth function D. This *max-depth approach* was first proposed in Liu *et al.* (1999) and was then investigated in Ghosh & Chaudhuri (2005b). Dutta & Ghosh (2012a,b) considered max-depth classifiers based on the projection depth and on (an affine-invariant version of) the L^p depth, respectively. Hubert & Van der Veeken (2010) modified the max-depth approach based on projection depth to better cope with possibly skewed data.

Recently, Li *et al.* (2012) proposed the "Depth vs Depth" (DD) classifiers that extend the max-depth ones by constructing appropriate polynomial separating curves in the DD-plot, that is, in the scatter plot of the points $(D_0^{(n)}(\mathbf{X}_i), D_1^{(n)}(\mathbf{X}_i))$, $i = 1, \ldots, n$, where $D_j^{(n)}(\mathbf{X}_i)$ refers to the depth of $\mathbf{X}_i$ with respect to the data points coming from Population j. Those separating curves are chosen to minimize the empirical misclassification rate on the training sample and their polynomial degree m is chosen through cross-validation. Lange *et al.* (2012) defined modified DD-classifiers that are computationally efficient and apply in higher dimensions (up to $d = 20$). Other depth-based classifiers were proposed in Jörnsten (2004), Ghosh & Chaudhuri (2005a), and Cui *et al.* (2008).

Being based on depth, these classifiers are clearly of a nonparametric nature. An important requirement in nonparametric classification, however, is that consistency holds as broadly as possible and, in particular, does not require "structural" distributional assumptions. In that respect, the depth-based classifiers available in the literature are not so satisfactory, since they are at best consistent under elliptical distributions only[2]. This restricted-to-ellipticity consistency implies that, as far as consistency is concerned, the Mahalanobis depth is perfectly sufficient and is by no means inferior to the "more nonparametric" (Tukey, 1975) halfspace depth or (Liu, 1990) simplicial depth, despite the fact that it uninspiringly leads to LDA through the max-depth approach. Also, even this restricted consistency often requires estimating densities; see, e.g., Dutta & Ghosh (2012a,b). This is somewhat undesirable since density and depth are quite antinomic in spirit (a deepest point may very well be a point where the density vanishes). Actually, if densities are to be estimated in the procedure anyway, then it would be more natural to go for density estimation all the way, that is, to plug density estimators in (1.2).

The poor consistency of the available depth-based classifiers actually follows from their *global* nature. Zakai & Ritov (2009) indeed proved that any universally consistent classifier needs to be of a *local* nature. In this paper, we therefore introduce local depth-based classifiers, that rely on nearest-neighbor ideas (kernel density techniques should be avoided, since, as mentioned above, depth and densities are somewhat incompatible). From their nearest-neighbor nature, they will inherit consistency under very mild conditions, while from their depth nature, they will inherit affine-invariance and robustness, two important features in multivariate statistics and in classification in particular. Identifying nearest neighbors through depth will be achieved via an original symmetrization construction. The corresponding depth-based neighborhoods are of a nonparametric nature and the good finite-sample behavior of the resulting classifiers most likely results from their data-driven adaptive nature.

The outline of the paper is as follows. In Section 2, we first recall the concept of statistical depth functions (Section 2.1) and then describe our symmetrization construction that allows to define the depth-based neighbors to be used later for classification purposes (Section 2.2). In Section 3, we define the proposed depth-based nearest-neighbor classifiers and present some of their basic properties (Section 3.1) before providing consistency results (Section 3.3). In Section 4, Monte

---

[2]The classifiers from Dutta & Ghosh (2012b) are an exception that slightly extends consistency to (a subset of) the class of L_p-elliptical distributions.

Carlo simulations are used to compare the finite-sample performances of our classifiers with those of their competitors. In Section 5, we show the practical value of the proposed classifiers on two real-data examples. We then discuss in Section 6 some further applications of our depth-based neighborhoods. Finally, the Appendix collects the technical proofs.

2 Depth-based neighbors

In this section, we review the concept of statistical depth functions and define the depth-based neighborhoods on which the proposed nearest-neighbor classifiers will be based.

2.1 Statistical depth functions

Statistical depth functions allow to measure *centrality* of any $\boldsymbol{x} \in \mathbb{R}^d$ with respect to a probability measure P over $\mathbb{R}^d$ (the larger the depth of $\boldsymbol{x}$, the more central $\boldsymbol{x}$ is with respect to P). Following Zuo & Serfling (2000a), we define a statistical depth function as a bounded mapping $D(\cdot, P)$ from $\mathbb{R}^d$ to $\mathbb{R}^+$ that satisfies the following four properties:

(P1) *affine-invariance:* for any $d \times d$ invertible matrix $\mathbf{A}$, any d-vector $\mathbf{b}$ and any distribution P over $\mathbb{R}^d$, $D(\mathbf{A}\boldsymbol{x}+\mathbf{b}, P^{\mathbf{A},\mathbf{b}}) = D(\boldsymbol{x}, P)$, where $P^{\mathbf{A},\mathbf{b}}$ is defined through $P^{\mathbf{A},\mathbf{b}}[B] = P[\mathbf{A}^{-1}(B-\mathbf{b})]$ for any d-dimensional Borel set B;

(P2) *maximality at center:* for any P that is symmetric about $\boldsymbol{\theta}$ (in the sense[3] that $P[\boldsymbol{\theta} + B] = P[\boldsymbol{\theta} - B]$ for any d-dimensional Borel set B), $D(\boldsymbol{\theta}, P) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}, P)$;

(P3) *monotonicity relative to the deepest point:* for any P having deepest point $\boldsymbol{\theta}$, for any $\boldsymbol{x} \in \mathbb{R}^d$ and any $\lambda \in [0, 1]$, $D(\boldsymbol{x}, P) \leq D((1 - \lambda)\boldsymbol{\theta} + \lambda\boldsymbol{x}, P)$;

(P4) *vanishing at infinity:* for any P, $D(\boldsymbol{x}, P) \to 0$ as $\|\boldsymbol{x}\| \to \infty$.

For any statistical depth function and any $\alpha > 0$, the set $R_\alpha(P) = \{\boldsymbol{x} \in \mathbb{R}^d : D(\boldsymbol{x}, P) \geq \alpha\}$ is called *the depth region of order* α. These regions are nested, and, clearly, inner regions collect points with larger depth. Below, it will often be convenient to rather index these regions by their probability content : for any $\beta \in [0, 1)$, we will denote by $R^\beta(P)$ the smallest $R_\alpha(P)$ that has P-probability larger than or equal to β. Throughout, subscripts and superscripts for depth regions are used for depth levels and probability contents, respectively.

Celebrated instances of statistical depth functions include

(i) the Tukey (1975) halfspace depth $D_H(\boldsymbol{x}, P) = \inf_{\boldsymbol{u} \in \mathcal{S}^{d-1}} P[\boldsymbol{u}'(\mathbf{X} - \boldsymbol{x}) \geq 0]$, where $\mathcal{S}^{d-1} = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| = 1\}$ is the unit sphere in $\mathbb{R}^d$;

(ii) the Liu (1990) simplicial depth $D_S(\boldsymbol{x}, P) = P[\boldsymbol{x} \in S(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{d+1})]$, where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{d+1}$ are i.i.d. P and where $S(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{d+1})$ denotes the closed simplex with vertices $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1}$;

(iii) the Mahalanobis depth $D_M(\boldsymbol{x}, P) = 1/(1+d^2_{\boldsymbol{\Sigma}(P)}(\boldsymbol{x}, \boldsymbol{\mu}(P)))$, for some affine-equivariant location and scatter functionals $\boldsymbol{\mu}(P)$ and $\boldsymbol{\Sigma}(P)$;

---

[3]Zuo & Serfling (2000a) also considers more general symmetry concepts; however, we restrict in the sequel to central symmetry, that will be the right concept for our purposes.

(iv) the projection depth $D_{Pr}(\boldsymbol{x}, P) = 1/(1+\sup_{\boldsymbol{u}\in\mathcal{S}^{d-1}} |\boldsymbol{u}'\boldsymbol{x}-\mu(P_{[\boldsymbol{u}]})|/\sigma(P_{[\boldsymbol{u}]}))$, where $P_{[\boldsymbol{u}]}$ denotes the probability distribution of $\boldsymbol{u}'\mathbf{X}$ when $\mathbf{X} \sim P$ and where $\mu(P)$ and $\sigma(P)$ are univariate location and scale functionals, respectively.

Other depth functions are the simplicial volume depth, the spatial depth, the L_p depth, etc. Of course, not all such depths fulfill Properties (P1)-(P4) for any distribution P; see Zuo & Serfling (2000a). A further concept of depth, of a slightly different (L_2) nature, is the so-called *zonoid depth*; see Koshevoy & Mosler (1997).

Of course, if d-variate observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are available, then sample versions of the depths above are simply obtained by replacing P with the corresponding empirical distribution $P^{(n)}$ (the sample simplicial depth then has a U-statistic structure).

A crucial fact for our purposes is that a sample depth provides a *center-outward ordering* of the observations with respect to the corresponding deepest point $\hat{\boldsymbol{\theta}}^{(n)}$: one may indeed order the $\mathbf{X}_i$'s in such a way that

$$D(\mathbf{X}_{(1)}, P^{(n)}) \geq D(\mathbf{X}_{(2)}, P^{(n)}) \geq \ldots \geq D(\mathbf{X}_{(n)}, P^{(n)}). \tag{2.1}$$

Neglecting possible ties, this states that, in the depth sense, $\mathbf{X}_{(1)}$ is the observation closest to $\hat{\boldsymbol{\theta}}^{(n)}$, $\mathbf{X}_{(2)}$ the second closest, $\ldots$, and $\mathbf{X}_{(n)}$ the one farthest away.

For most classical depths, there may be infinitely many deepest points, that form a convex region in $\mathbb{R}^d$. This will not be an issue in this work, since the symmetrization construction we will introduce, jointly with Properties (Q2)-(Q3) below, asymptotically guarantee unicity of the deepest point. For some particular depth functions, unicity may even hold for finite samples. For instance, in the case of halfspace depth, it follows from Rousseeuw & Struyf (2004) and results on the uniqueness of the symmetry center (Serfling, 2006) that, under the assumption that the parent distribution admits a density, symmetrization implies almost sure unicity of the deepest point.

2.2 Depth-based neighborhoods

A statistical depth function, through (2.1), can be used to define neighbors of the deepest point $\hat{\boldsymbol{\theta}}^{(n)}$. Implementing a nearest-neighbor classifier, however, requires defining neighbors of any point $\boldsymbol{x} \in \mathbb{R}^d$. Property (P2) provides the key to the construction of an $\boldsymbol{x}$-outward ordering of the observations, hence to the definition of depth-based neighbors of $\boldsymbol{x}$: symmetrization with respect to $\boldsymbol{x}$.

More precisely, we propose to consider depth with respect to the empirical distribution $P_{\boldsymbol{x}}^{(n)}$ associated with the sample obtained by adding to the original observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ their reflections $2\boldsymbol{x} - \mathbf{X}_1, \ldots, 2\boldsymbol{x} - \mathbf{X}_n$ with respect to $\boldsymbol{x}$. Property (P2) implies that $\boldsymbol{x}$ is the—unique (at least asymptotically; see above)—deepest point with respect to $P_{\boldsymbol{x}}^{(n)}$. Consequently, this symmetrization construction, parallel to (2.1), leads to an ($\boldsymbol{x}$-outward) ordering of the form

$$D(\mathbf{X}_{\boldsymbol{x},(1)}, P_{\boldsymbol{x}}^{(n)}) \geq D(\mathbf{X}_{\boldsymbol{x},(2)}, P_{\boldsymbol{x}}^{(n)}) \geq \ldots \geq D(\mathbf{X}_{\boldsymbol{x},(n)}, P_{\boldsymbol{x}}^{(n)}).$$

Note that the reflected observations are only used to define the ordering but are not ordered themselves. For any $k \in \{1, \ldots, n\}$, this allows to identify—up to possible ties—the k nearest neighbors $\mathbf{X}_{\boldsymbol{x},(i)}$, $i = 1, \ldots, k$, of $\boldsymbol{x}$. In the univariate case ($d = 1$), these k neighbors coincide—irrespective

of the statistical depth function D—with the k data points minimizing the usual distances $|X_i - x|$, $i = 1, \ldots, n$.

In the sequel, the corresponding *depth-based neighborhoods*—that is, the sample depth regions $R_{\boldsymbol{x},\alpha}^{(n)} = R_\alpha(P_{\boldsymbol{x}}^{(n)})$—will play an important role. In accordance with the notation from the previous section, we will write $R_{\boldsymbol{x}}^{\beta(n)}$ for the smallest depth region $R_{\boldsymbol{x},\alpha}^{(n)}$ that contains at least a proportion β of the data points $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. For $\beta = k/n$, $R_{\boldsymbol{x}}^{\beta(n)}$ is therefore the smallest depth-based neighborhood that contains k of the $\mathbf{X}_i$'s; ties may imply that the number of data points in this neigborhood, $K_{\boldsymbol{x}}^{\beta(n)}$ say, is strictly larger than k.

Note that a distance (or pseudo-distance) $(\boldsymbol{x}, \boldsymbol{y}) \mapsto d(\boldsymbol{x}, \boldsymbol{y})$ that is symmetric in its arguments is not needed to identify nearest neighbors of $\boldsymbol{x}$. For that purpose, a collection of "distances" $\boldsymbol{y} \mapsto d_{\boldsymbol{x}}(\boldsymbol{y})$ from a fixed point is indeed sufficient (in particular, it is irrelevant that this distance satisfies or not the triangular inequality). In that sense, the (data-driven) symmetric distance associated with the Oja & Paindaveine (2005) *lift-interdirections*, that was recently used to build nearest-neighbor regression estimators in Biau *et al.* (2012), is unnecessarily strong. Also, only an ordering of the "distances" is needed to identify nearest neighbors. This *ordering* of distances *from a fixed point* $\boldsymbol{x}$ is exactly what the depth-based $\boldsymbol{x}$-outward ordering above is providing.

3   Depth-based kNN classifiers

In this section, we first define the proposed depth-based classifiers and present some of their basic properties (Section 3.1). We then state the main result of this paper, related to their consistency properties (Section 3.3).

3.1   Definition and basic properties

The standard k-nearest-neighbor (kNN) procedure classifies the point $\boldsymbol{x}$ into Population 1 iff there are more observations from Population 1 than from Population 0 in the smallest Euclidean ball centered at $\boldsymbol{x}$ that contains k data points. Depth-based kNN classifiers are naturally obtained by replacing these Euclidean neighborhoods with the depth-based neighborhoods introduced above, that is, the proposed kNN procedure classifies $\boldsymbol{x}$ into Population 1 iff there are more observations from Population 1 than from Population 0 in the smallest depth-based neighborhood of $\boldsymbol{x}$ that contains k observations—i.e., in $R_{\boldsymbol{x}}^{\beta(n)}$, $\beta = k/n$. In other words, the proposed depth-based classifier is defined as

$$\hat{m}_D^{(n)}(\boldsymbol{x}) = \mathbb{I}\left[\textstyle\sum_{i=1}^n \mathbb{I}[Y_i = 1] W_i^{\beta(n)}(\boldsymbol{x}) > \sum_{i=1}^n \mathbb{I}[Y_i = 0] W_i^{\beta(n)}(\boldsymbol{x})\right], \tag{3.1}$$

with $W_i^{\beta(n)}(\boldsymbol{x}) = \frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \mathbb{I}[\mathbf{X}_i \in R_{\boldsymbol{x}}^{\beta(n)}]$, where $K_{\boldsymbol{x}}^{\beta(n)} = \sum_{j=1}^n \mathbb{I}[\mathbf{X}_j \in R_{\boldsymbol{x}}^{\beta(n)}]$ still denotes the number of observations in the depth-based neighborhood $R_{\boldsymbol{x}}^{\beta(n)}$. Since

$$\hat{m}_D^{(n)}(\boldsymbol{x}) = \mathbb{I}\left[\hat{\eta}_D^{(n)}(\boldsymbol{x}) > 1/2\right], \quad \text{with } \hat{\eta}_D^{(n)}(\boldsymbol{x}) = \textstyle\sum_{i=1}^n \mathbb{I}[Y_i = 1] W_i^{\beta(n)}(\boldsymbol{x}), \tag{3.2}$$

the proposed classifier is actually the one obtained by plugging, in (1.1), the depth-based estimator $\hat{\eta}_D^{(n)}(\boldsymbol{x})$ of the conditional expectation $\eta(\boldsymbol{x})$. This will be used in the proof of Theorem 3.1 below. Note that in the univariate case ($d = 1$), $\hat{m}_D^{(n)}$, irrespective of the statistical depth function D, reduces to the standard (Euclidean) kNN classifier.

It directly follows from Property (P1) that the proposed classifier is affine-invariant, in the sense that the outcome of the classification will not be affected if $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $\boldsymbol{x}$ are subject to a common (arbitrary) affine transformation. This clearly improves over the standard kNN procedure that, for instance, is sensitive to unit changes. Of course, one natural way to define an affine-invariant kNN classifier is to apply the original kNN procedure on the standardized data points $\hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{X}_i$, $i = 1, \ldots, n$, where $\hat{\boldsymbol{\Sigma}}$ is an affine-equivariant estimator of shape—in the sense that

$$\hat{\boldsymbol{\Sigma}}(\mathbf{A}\mathbf{X}_1 + \mathbf{b}, \ldots, \mathbf{A}\mathbf{X}_n + \mathbf{b}) \propto \mathbf{A}\hat{\boldsymbol{\Sigma}}(\mathbf{X}_1, \ldots, \mathbf{X}_n)\mathbf{A}'$$

for any invertible $d \times d$ matrix $\mathbf{A}$ and any d-vector $\mathbf{b}$. A natural choice for $\hat{\boldsymbol{\Sigma}}$ is the regular covariance matrix, but more robust choices, such as, e.g., the shape estimators from Tyler (1987), Dümbgen (1998), or Hettmansperger & Randles (2002) would allow to get rid of any moment assumption. Here, we stress that, unlike our *adaptive* depth-based methodology, such a transformation approach leads to neighborhoods that do not exploit the geometry of the distribution in the vicinity of the point $\boldsymbol{x}$ to be classified (these neighborhoods indeed all are ellipsoids with $\boldsymbol{x}$-independent orientation and shape); as we show through simulations below, this results into significantly worse performances.

The main depth-based classifiers available—among which those relying on the max-depth approach of Liu *et al.* (1999) and Ghosh & Chaudhuri (2005b), as well as the more efficient ones from Li *et al.* (2012)—suffer from the "outsider problem[4]" : if the point $\boldsymbol{x}$ to be classified does not sit in the convex hull of any of the two populations, then most statistical depth functions will give $\boldsymbol{x}$ zero depth with respect to each population, so that $\boldsymbol{x}$ cannot be classified through depth. This is of course undesirable, all the more so that such a point $\boldsymbol{x}$ may very well be easy to classify. To improve on this, Hoberg & Mosler (2006) proposed extending the original depth fields by using the Mahalanobis depth outside the supports of both populations, a solution that quite unnaturally requires combining two depth functions. Quite interestingly, our symmetrization construction implies that the depth-based kNN classifier (that involves one depth function only) does not suffer from the outsider problem; this is an important advantage over competing depth-based classifiers.

While our depth-based classifiers in (3.1) are perfectly well-defined and enjoy, as we will show in Section 3.3 below, excellent consistency properties, practitioners might find quite arbitrary that a point $\boldsymbol{x}$ such that $\sum_{i=1}^{n} \mathbb{I}[Y_i = 1] W_i^{\beta(n)}(\boldsymbol{x}) = \sum_{i=1}^{n} \mathbb{I}[Y_i = 0] W_i^{\beta(n)}(\boldsymbol{x})$ is assigned to Population 0. Parallel to the standard kNN classifier, the classification may alternatively be based on the population of the next neighbor. Since ties are likely to occur when using depth, it is natural to rather base classification on the proportion of data points from each population in the next depth region. Of course, if the next depth region still leads to an ex-aequo, the outcome of the classification is to be determined on the subsequent depth regions, until a decision is reached (in the unlikely case that an ex-aequo occurs for all depth regions to be considered, classification should then be done by flipping a coin). This treatment of ties is used whenever real or simulated data are considered below.

Finally, practitioners have to choose some value for the smoothing parameter k_n. This may be done, for example, through cross-validation (as we will do in the real data example of Section 5). The value of k_n is likely to have a strong impact on finite-sample performances, as confirmed in the simulations we conduct in Section 4.

---

[4]The term "outsider" was recently introduced in Lange *et al.* (2012).

3.2   Consistency results

As expected, the local (nearest-neighbor) nature of the proposed classifiers makes them consistent under very mild conditions. This, however, requires that the statistical depth function D satisfies the following further properties:

(Q1) *continuity:* if P is symmetric about $\boldsymbol{\theta}$ and admits a density that is positive at $\boldsymbol{\theta}$, then $\boldsymbol{x} \mapsto D(\boldsymbol{x}, P)$ is continuous in a neighborhood of $\boldsymbol{\theta}$.

(Q2) *unique maximization at the symmetry center:* if P is symmetric about $\boldsymbol{\theta}$ and admits a density that is positive at $\boldsymbol{\theta}$, then $D(\boldsymbol{\theta}, P) > D(\boldsymbol{x}, P)$ for all $\boldsymbol{x} \neq \boldsymbol{\theta}$.

(Q3) *consistency:* for any bounded d-dimensional Borel set B, $\sup_{\boldsymbol{x} \in B} |D(\boldsymbol{x}, P^{(n)}) - D(\boldsymbol{x}, P)| = o(1)$ almost surely as $n \to \infty$, where $P^{(n)}$ denotes the empirical distribution associated with n random vectors that are i.i.d. P.

Property (Q2) complements Property (P2), and, in view of Property (P3), only further requires that $\boldsymbol{\theta}$ is a strict local maximizer of $\boldsymbol{x} \mapsto D(\boldsymbol{x}, P)$. Note that Properties (Q1)-(Q2) jointly ensure that the depth-based neighborhoods of $\boldsymbol{x}$ from Section 2.2 collapse to the singleton $\{\boldsymbol{x}\}$ when the depth level increases to its maximal value. Finally, since our goal is to prove that our classifier satisfies an asymptotic property (namely, consistency), it is not surprising that we need to control the asymptotic behavior of the sample depth itself (Property (Q3)). As shown by Theorem 1 in the Appendix, Properties (Q1)-(Q3) are satisfied for many classical depth functions.

We can then state the main result of this paper.

Theorem 3.1. Let D be a depth function satisfying (P2), (P3) and (Q1)-(Q3). Let k_n be a sequence of positive integers such that $k_n \to \infty$ and $k_n = o(n)$ as $n \to \infty$. Assume that, for $j = 0, 1$, $\mathbf{X}|[Y = j]$ admits a density f_j whose collection of discontinuity points has Lebesgue measure zero. Then the depth-based k_nNN classifier $m_D^{(n)}$ in (3.1) is consistent in the sense that

$$P[m_D^{(n)}(\mathbf{X}) \neq Y \,|\, \mathcal{D}_n] - P[m_{\text{Bayes}}(\mathbf{X}) \neq Y] = o_P(1) \quad \text{as } n \to \infty,$$

where $\mathcal{D}_n$ is the sigma-algebra associated with $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$.

Classically, consistency results for classification are based on a famous theorem from Stone (1977); see, e.g., Theorem 6.3 in Devroye *et al.* (1996). However, it is an open question whether Condition (i) of this theorem holds or not for the proposed classifiers, at least for some particular statistical depth functions. A sufficient condition for Condition (i) is actually that there exists a partition of $\mathbb{R}^d$ into cones $C_1, \ldots, C_{\gamma_d}$ with vertex at the origin of $\mathbb{R}^d$ (γ_d not depending on n) such that, for any $\mathbf{X}_i$ and any j, there exist (with probability one) at most k data points $\mathbf{X}_\ell \in \mathbf{X}_i + C_j$ that have $\mathbf{X}_i$ among their k depth-based nearest neighbors. Would this be established for some statistical depth function D, it would prove that the corresponding depth-based k_nNN classifier $\hat{m}_D^{(n)}$ is *universally consistent*, in the sense that consistency holds without *any* assumption on the distribution of $(\mathbf{X}, Y)$.

Now, it is clear from the proof of Stone's theorem that this condition (i) may be dropped if one further assumes that $\mathbf{X}$ admits a uniformly continuous density. This is however a high price to pay, and that is the reason why the proof of Theorem 3.1 rather relies on an argument recently used in Biau *et al.* (2012); see the Appendix.

4   Simulations

We performed simulations in order to evaluate the finite-sample performances of the proposed depth-based kNN classifiers. We considered six setups, focusing on bivariate $\mathbf{X}_i$'s ($d = 2$) with equal a priori probabilities ($\pi_0 = \pi_1 = 1/2$), and involving the following densities f_0 and f_1:

Setup 1 (multinormality): f_j, $j = 0, 1$, is the pdf of the bivariate normal distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, where

$$\boldsymbol{\mu}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1 = 4\boldsymbol{\Sigma}_0;$$

Setup 2 (bivariate Cauchy): f_j, $j = 0, 1$, is the pdf of the bivariate Cauchy distribution with location center $\boldsymbol{\mu}_j$ and scatter matrix $\boldsymbol{\Sigma}_j$, with the same values of $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ as in Setup 1;

Setup 3 (flat covariance structure): f_j, $j = 0, 1$, is the pdf of the bivariate normal distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, where

$$\boldsymbol{\mu}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 5^2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_0;$$

Setup 4 (uniform distributions on half-moons): f_0 and f_1 are the densities of

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} U \\ V \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} -0.5 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix},$$

respectively, where $U \sim \text{Unif}(-1, 1)$ and $V|[U = u] \sim \text{Unif}(1 - u^2, 2(1 - u^2))$;

Setup 5 (uniform distributions on rings): f_0 and f_1 are the uniform distributions on the concentric rings $\{\boldsymbol{x} \in \mathbb{R}^2 : 1 \leq \|\boldsymbol{x}\| \leq 2\}$ and $\{\boldsymbol{x} \in \mathbb{R}^2 : 1.75 \leq \|\boldsymbol{x}\| \leq 2.5\}$, respectively;

Setup 6 (bimodal populations): f_j, $j = 0, 1$, is the pdf of the multinormal mixture $\frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_j^I, \boldsymbol{\Sigma}_j^I) + \frac{1}{2}\mathcal{N}(\boldsymbol{\mu}_j^{II}, \boldsymbol{\Sigma}_j^{II})$, where

$$\boldsymbol{\mu}_0^I = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\mu}_0^{II} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \quad \boldsymbol{\Sigma}_0^I = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}, \quad \boldsymbol{\Sigma}_0^{II} = 4\boldsymbol{\Sigma}_0^I,$$

$$\boldsymbol{\mu}_1^I = \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}, \quad \boldsymbol{\mu}_1^{II} = \begin{pmatrix} 4.5 \\ 4.5 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1^I = \begin{pmatrix} 4 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma}_1^{II} = \begin{pmatrix} 0.75 & 0 \\ 0 & 5 \end{pmatrix}.$$

For each of these six setups, we generated 250 training and test samples of size $n = n_{\text{train}} = 200$ and $n_{\text{test}} = 100$, respectively, and evaluated the misclassification frequencies of the following classifiers:

1. the usual LDA and QDA classifiers (LDA/QDA);

2. the standard Euclidean kNN classifiers (kNN), with $\beta = k/n = 0.01$, 0.05, 0.10 and 0.40, and the corresponding "Mahalanobis" kNN classifiers (kNNaff) obtained by performing the Euclidean kNN classifiers on standardized data, where standardization is based on the regular covariance matrix estimate of the pooled training sample;

3. the proposed depth-based kNN classifiers (D-kNN) for each combination of the k used in kNN/kNNaff and a statistical depth function (we focused on halfspace depth, simplicial depth, or Mahalanobis depth);

4. the depth vs depth (DD) classifiers from Li *et al.* (2012), for each combination of a polynomial curve of degree m ($m = 1, 2,$ or 3) and a statistical depth function (halfspace depth, simplicial depth, or Mahalanobis depth). Exact DD-classifiers (DD) as well as smoothed versions (DDsm) were actually implemented —although, for computational reasons, only the smoothed version was considered for $m = 3$. Exact classifiers search for the best separating polynomial curve $(d, r(d))$ of order m passing through the origin and m "DD-points" $(D_0^{(n)}(\mathbf{X}_i), D_1^{(n)}(\mathbf{X}_i))$ (see the Introduction) in the sense that it minimizes the misclassification error

$$\sum_{i=1}^{n} \left(\mathbb{I}[Y_i = 1]\mathbb{I}[d_i^{(n)} > 0] + \mathbb{I}[Y_i = 0]\mathbb{I}[-d_i^{(n)} > 0] \right), \tag{4.1}$$

with $d_i^{(n)} := r(D_0^{(n)}(\mathbf{X}_i)) - D_1^{(n)}(\mathbf{X}_i)$. Smoothed versions use derivative-based methods to find a polynomial minimizing (4.1), where the indicator $\mathbb{I}[d > 0]$ is replaced by the logistic function $1/(1 + e^{-td})$ for a suitable t. As suggested in Li *et al.* (2012), value $t = 100$ was chosen in these simulations. 100 randomly chosen polynomials were used as starting points for the minimization algorithm, the classifier using the resulting polynomial with minimal misclassification (note that this time-consuming scheme always results into better performances than the one adopted in Li *et al.* (2012), where only one minimization is performed, starting from the best random polynomial considered).

Since the DD classification procedure is a refinement of the max-depth procedures of Ghosh & Chaudhuri (2005b) that leads to better misclassification rates (see Li *et al.* (2012)), the original max-depth procedures were omitted in this study.

Boxplots of misclassification frequencies (in percentages) are reported in Figures 3 and 2. It is seen that in most setups, the proposed depth-based kNN classifiers compete well with the Euclidean kNN classifiers. The latter, however, should be avoided since (i) their outcome may unpleasantly depend on measurement units, and since (ii) the spherical nature of the neighborhoods used lead to performances that are severely affected by the—notoriously delicate—choice of k; see the "flat" setup 3. This motivates restricting to affine-invariant classifiers, that (i) are totally insensitive to any unit changes and that (ii) can adapt to the flat structure of Setup 3 as they show there performances that are much more stable in k.

Now, regarding the comparisons between affine-invariant classifiers, the simulations results lead to the following conclusions: (i) the proposed affine-invariant depth-based classifiers outperform the natural affine-invariant versions of kNN classifiers. In other words, the natural way to make the standard kNN classifier affine-invariant results into a dramatic cost in terms of finite-sample performances. (ii) The proposed depth-based kNN classifiers also compete well with DD-classifiers both in elliptical and non-elliptical setups. Away from ellipticity (Setups 4 to 6), in particular, they perform at least as well—and sometimes outperform (Setup 4)—DD-classifiers; a single exception is associated with the use of Mahalanobis depth in Setup 5, where the DD-classifiers based on $m = 2, 3$ perform better. Apparently, another advantage of depth-based kNN classifiers over DD-classifiers is that their finite-sample performances depend much less on the statistical depth function D used.
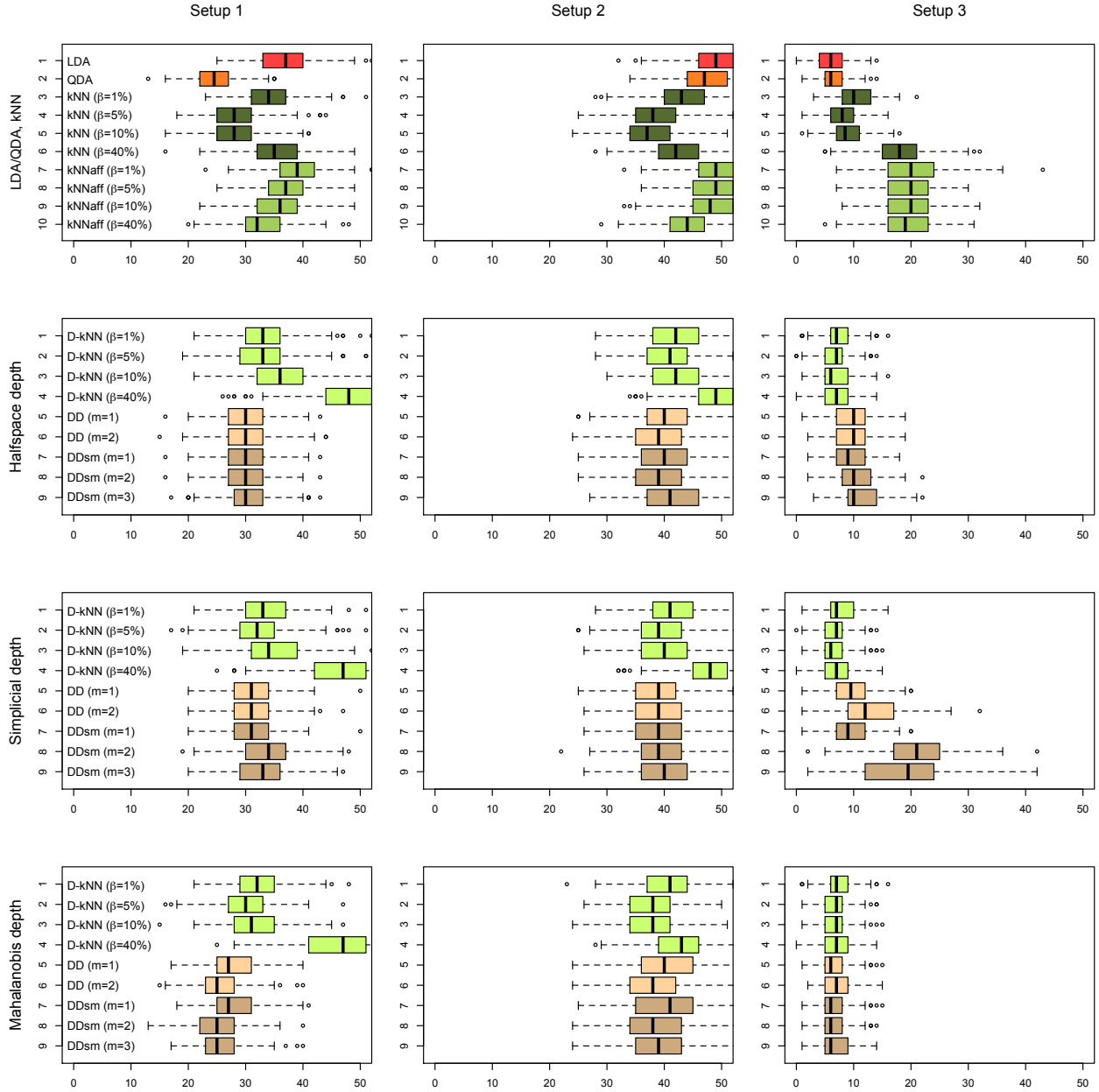
Figure 1: *Boxplots of misclassification frequencies (in percentages), from 250 replications of Setups 1 to 3 described in Section 4, with training sample size $n = n_{\text{train}} = 200$ and test sample size $n_{\text{test}} = 100$, of the LDA/QDA classifiers, the Euclidean kNN classifiers (kNN) and their Mahalanobis (affine-invariant) counterparts (KNNaff), the proposed depth-based kNN classifiers (D-kNN), and some exact and smoothed version of the DD-classifiers (DD and DDsm); see Section 4 for details.*
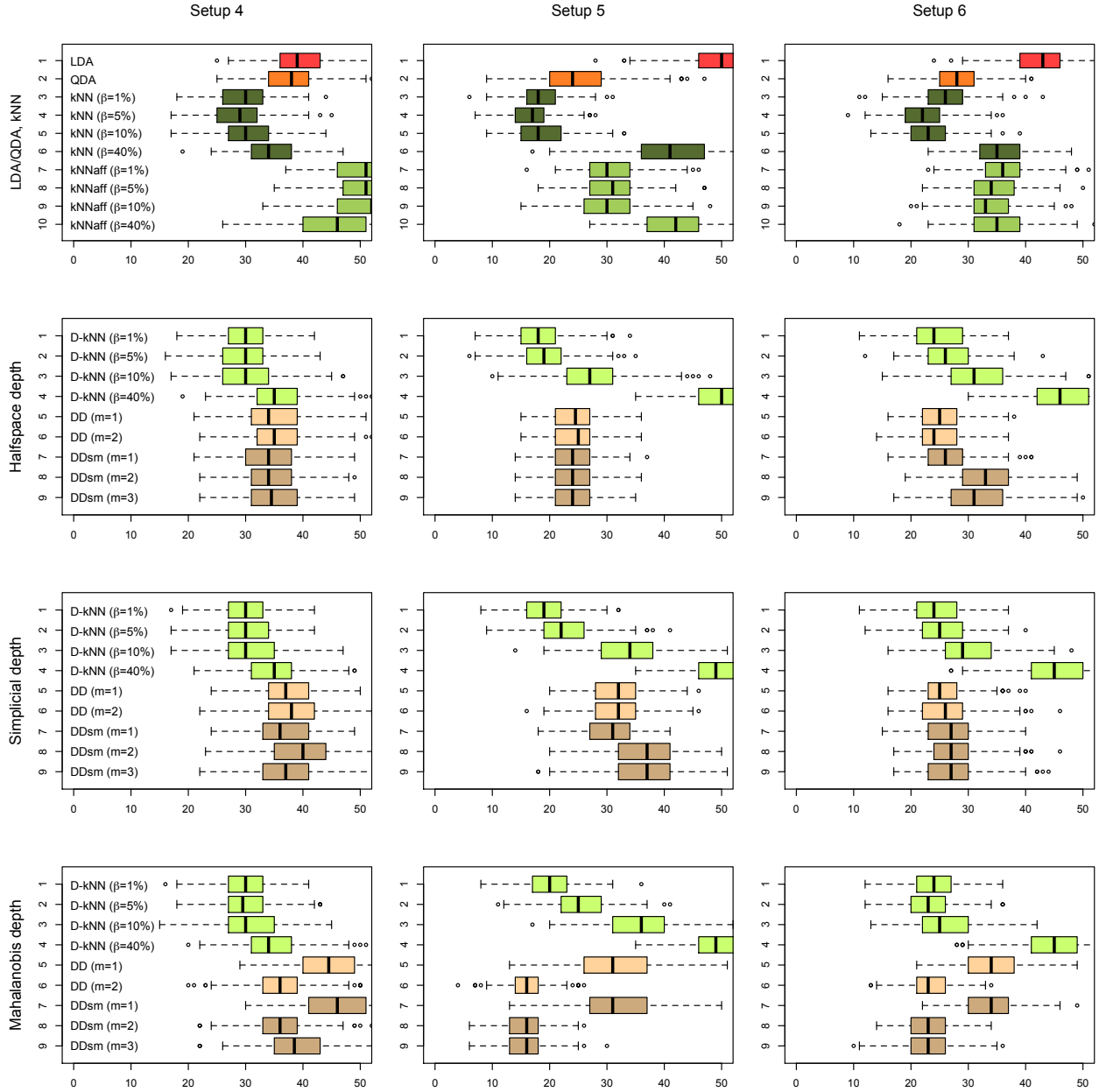
Figure 2: *Boxplots of misclassification frequencies (in percentages), from 250 replications of Setups 4 to 6 described in Section 4, with training sample size $n = n_{\text{train}} = 200$ and test sample size $n_{\text{test}} = 100$, of the LDA/QDA classifiers, the Euclidean kNN classifiers (kNN) and their Mahalanobis (affine-invariant) counterparts (KNNaff), the proposed depth-based kNN classifiers (D-kNN), and some exact and smoothed version of the DD-classifiers (DD and DDsm); see Section 4 for details.*

5   Real-data examples

In this section, we investigate the performances of our depth-based kNN classifiers on two well known benchmark datasets. The first example is taken from Ripley (1996) and can be found on the book's website (`http://www.stats.ox.ac.uk/pub/PRNN`). This data set involves well-specified training and test samples, and we therefore simply report the test set misclassification rates of the different classifiers included in the study. The second example, blood transfusion data, is available at `http://archive.ics.uci.edu/ml/index.html`. Unlike the first data set, no clear partition into a training sample and a test sample is provided here. As suggested in Li *et al.* (2012), we randomly performed such a partition 100 times (see the details below) and computed the average test set misclassification rates, together with standard deviations. A brief description of each dataset is as follows:

Synthetic data was introduced and studied in Ripley (1996). The dataset is made of observations from two populations, each of them being actually a mixture of two bivariate normal distributions differing only in location. As mentioned above, a partition into a training sample and a test sample is provided: the training and test samples contain 250 and 1000 observations, respectively, and both samples are divided equally between the two populations.

Transfusion data contains the information on 748 blood donors selected from the blood donor database of the Blood Transfusion Service Center in Hsin-Chu City, Taiwan. It was studied in Yeh *et al.* (2009). The classification problem at hand is to know whether or not the donor gave blood in March 2007. In this dataset, prior probabilities are not equal; out of 748 donors, 178 gave blood in March 2007, when 570 did not. Following Li *et al.* (2012), one out of two linearly correlated variables was removed and three measurements were available for each donor: Recency (number of months since the last donation), Frequency (total number of donations) and Time (time since the first donation). The training set consists in 100 donors from the first class and 400 donors from the second, while the rest is assigned to the test sample (therefore containing 248 individuals).

Table 1 reports the—exact (synthetic) or averaged (transfusion)—misclassification rates of the following classifiers: the linear (LDA) and quadratic (QDA) discriminant rules, the standard kNN classifier (kNN) and its Mahalanobis affine-invariant version (kNNaff), the depth-based kNN classifiers using halfspace depth (D_H-kNN) and Mahalanobis depth (D_M-kNN), and the exact DD-classifiers for any combination of a polynomial order $m \in \{1, 2\}$ and a statistical depth function among the two considered for depth-based kNN classifiers, namely the halfspace depth (DD_H) and the Mahalanobis depth (DD_M)—smoothed DD-classifiers were excluded from this study, as their performances, which can only be worse than those of exact versions, showed much sensitivity to the smoothing parameter t; see Section 4. For all nearest-neighbor classifiers, leave-one-out cross-validation was used to determine k.

The results from Table 1 indicate that depth-based kNN classifiers perform very well in both examples. For synthetic data, the halfspace depth-based kNN classifier (10.1%) is only dominated by the standard (Euclidian) kNN procedure (8.7%). The latter, however, has to be discarded as it is dependent on scale and shape changes—in line with this, note that the "kNN classifier" applied in Dutta & Ghosh (2012b) is actually the kNNaff classifier (11.7%), as classification in that paper is performed on standardized data. The Mahalanobis depth-based kNN classifiers (14.4%) does not perform as well as its halfspace counterpart. For transfusion data, however, both depth-based kNN classifiers dominate their competitors.

Table 1: *Misclassification rates (in %) on the two benchmark datasets from Section 5.*

	Synthetic	Transfusion
LDA	10.8	29.60 (0.9)
QDA	10.2	29.21 (1.5)
kNN	8.7	29.74 (2.0)
kNNaff	11.7	30.11 (2.1)
D_H-kNN	10.1	27.75 (1.6)
D_M-kNN	14.4	27.36 (1.5)
DD_H ($m=1$)	13.4	28.26 (1.7)
DD_H ($m=2$)	12.9	28.33 (1.6)
DD_M ($m=1$)	17.5	31.44 (0.1)
DD_M ($m=2$)	12.0	31.54 (0.6)

6 Final comments

The depth-based neighborhoods we introduced are of interest in other inference problems as well. As an illustration, consider the regression problem where the conditional mean function $\boldsymbol{x} \mapsto m(\boldsymbol{x}) = E[Y \mid \mathbf{X} = \boldsymbol{x}]$ is to be estimated on the basis of mutually independent copies $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$ of a random vector $(\mathbf{X}, Y)$ with values in $\mathbb{R}^d \times \mathbb{R}$, or the problem of estimating the common density f of i.i.d. random d-vectors $\mathbf{X}_i$, $i = 1, \ldots, n$. The classical k_nNN estimators for these problems are

$$\hat{f}^{(n)}(\boldsymbol{x}) = \frac{k_n}{n \, \mu_d(B_{\boldsymbol{x}}^{\beta_n})} \text{ and } \hat{m}^{(n)}(\boldsymbol{x}) = \sum_{i=1}^{n} W_i^{\beta_n(n)}(\boldsymbol{x}) Y_i = \frac{1}{k_n} \sum_{i=1}^{n} \mathbb{I}\big[\mathbf{X}_i \in B_{\boldsymbol{x}}^{\beta_n}\big] Y_i, \qquad (6.1)$$

where $\beta_n = k_n/n$, $B_{\boldsymbol{x}}^{\beta}$ is the smallest Euclidean ball centered at $\boldsymbol{x}$ that contains a proportion β of the $\mathbf{X}_i$'s, and μ_d stands for the Lebesgue measure on $\mathbb{R}^d$. Our construction naturally leads to considering the depth-based k_nNN estimators $\hat{f}_D^{(n)}(\boldsymbol{x})$ and $\hat{m}_D^{(n)}(\boldsymbol{x})$ obtained by replacing in (6.1) the Euclidean neighborhoods $B_{\boldsymbol{x}}^{\beta_n}$ with their depth-based counterparts $R_{\boldsymbol{x}}^{\beta_n}$ and $k_n = \sum_{i=1}^{n} \mathbb{I}\big[\mathbf{X}_i \in B_{\boldsymbol{x}}^{\beta_n}\big]$ with $K_{\boldsymbol{x}}^{\beta_n(n)} = \sum_{i=1}^{n} \mathbb{I}\big[\mathbf{X}_i \in R_{\boldsymbol{x}}^{\beta_n}\big]$.

A thorough investigation of the properties of these depth-based procedures is of course beyond the scope of the present paper. It is, however, extremely likely that the excellent consistency properties obtained in the classification problem extend to these nonparametric regression and density estimation setups. Now, recent works in density estimation indicate that using non-spherical (actually, ellipsoidal) neighborhoods may lead to better finite-sample properties; see, e.g., Chacón (2009) or Chacón *et al.* (2011). In that respect, the depth-based kNN estimators above are very promising since they involve non-spherical (and for most classical depth, even non-ellipsoidal) neighborhoods whose shape is determined by the local geometry of the sample. Note also that depth-based neighborhoods only require choosing a single scalar bandwidth parameter (namely, k_n), whereas general d-dimensional ellipsoidal neighborhoods impose selecting $d(d+1)/2$ bandwidth parameters.

Acknowledgments

Appendix — Proofs

The main goal of this Appendix is to prove Theorem 3.1. We will need the following lemmas.

Lemma 1. *Assume that the depth function D satisfies (P2), (P3), (Q1), and (Q2). Let P be a probability measure that is symmetric about $\boldsymbol{\theta}$ and admits a density that is positive at $\boldsymbol{\theta}$. Then, (i) for all $a > 0$, there exists $\alpha < \alpha_* = \max_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}, P)$ such that $R_\alpha(P) \subset B_{\boldsymbol{\theta}}(a) := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{\theta}\| \leq a\}$; (ii) for all $\alpha < \alpha_*$, there exists $\xi > 0$ such that $B_{\boldsymbol{\theta}}(\xi) \subset R_\alpha(P)$.*

Proof of Lemma 1. (i) First note that the existence of α_* follows from Property (P2). Fix then $\delta > 0$ such that $\boldsymbol{x} \mapsto D(\boldsymbol{x}, P)$ is continuous over $B_{\boldsymbol{\theta}}(\delta)$; existence of δ is guaranteed by Property (Q1). Continuity implies that $\boldsymbol{x} \mapsto D(\boldsymbol{x}, P)$ reaches a minimum in $B_{\boldsymbol{\theta}}(\delta)$, and Property (Q2) entails that this minimal value, α_δ say, is strictly smaller than α_*. Using Property (Q1) again, we obtain that, for each $\alpha \in [\alpha_\delta, \alpha_*]$,

$$r_\alpha : \quad \mathcal{S}^{d-1} \quad \rightarrow \quad \mathbb{R}^+$$
$$\boldsymbol{u} \quad \mapsto \quad \sup\{r \in \mathbb{R}^+ : \boldsymbol{\theta} + r\boldsymbol{u} \in R_\alpha(P)\}$$

is a continuous function that converges pointwise to $r_{\alpha_*}(\boldsymbol{u}) \equiv 0$ as $\alpha \to \alpha_*$. Since $\mathcal{S}^{d-1}$ is compact, this convergence is actually uniform, i.e., $\sup_{\boldsymbol{u} \in \mathcal{S}^{d-1}} |r_\alpha(\boldsymbol{u})| = o(1)$ as $\alpha \to \alpha_*$. Part (i) of the result follows.

(ii) Property (Q2) implies that, for any $\alpha \in [\alpha_\delta, \alpha_*)$, the mapping r_α takes values in $\mathbb{R}_0^+$. Therefore there exists $\boldsymbol{u}_0(\alpha) \in \mathcal{S}^{d-1}$ such that $r_\alpha(\boldsymbol{u}) \geq r_\alpha(\boldsymbol{u}_0(\alpha)) = \xi_\alpha > 0$. This implies that, for all $\alpha \in [\alpha_\delta, \alpha_*)$, we have $B_{\boldsymbol{\theta}}(\xi_\alpha) \subset R_\alpha(P)$, which proves the result for these values of α. Nestedness of the $R_\alpha(P)$'s, which follows from Property (P3), then establishes the result for an arbitrary $\alpha < \alpha_*$. $\square$

Lemma 2. *Assume that the depth function D satisfies (P2), (P3), and (Q1)-(Q3). Let P be a probability measure that is symmetric about $\boldsymbol{\theta}$ and admits a density that is positive at $\boldsymbol{\theta}$. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. P and denote by $\mathbf{X}_{\boldsymbol{\theta},(i)}$ the ith depth-based nearest neighbor of $\boldsymbol{\theta}$. Let $K_{\boldsymbol{\theta}}^{\beta_n(n)}$ be the number of depth-based nearest neighbors in $R_{\boldsymbol{\theta}}^{\beta_n}(P^{(n)})$, where $\beta_n = k_n/n$ is based on a sequence k_n that is as in Theorem 3.1 and $P^{(n)}$ stands for the empirical distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Then, for any $a > 0$, there exists $n = n(a)$ such that $\sum_{i=1}^{K_{\boldsymbol{\theta}}^{\beta_n(n)}} \mathbb{I}[\|\mathbf{X}_{\boldsymbol{\theta},(i)} - \boldsymbol{\theta}\| > a] = 0$ almost surely for all $n \geq n(a)$.*

Note that, while $\mathbf{X}_{\boldsymbol{\theta},(i)}$ may not be properly defined (because of ties), the quantity $\sum_{i=1}^{K_{\boldsymbol{\theta}}^{\beta_n(n)}} \mathbb{I}[\|\mathbf{X}_{\boldsymbol{\theta},(i)} - \boldsymbol{\theta}\| > a] = 0$ always is.

Proof of Lemma 2. Fix $a > 0$. By Lemma 4.1, there exists $\alpha < \alpha_*$ such that $R_\alpha(P) \subset B_{\boldsymbol{\theta}}(a)$. Fix then $\bar{\alpha}$ and $\varepsilon > 0$ such that $\alpha < \bar{\alpha} - \varepsilon < \bar{\alpha} + \varepsilon < \alpha_*$. Theorem 4.1 in Zuo & Serfling (2000b) and the fact that $P_{\boldsymbol{\theta}}^{(n)} \to P_{\boldsymbol{\theta}} = P$ weakly as $n \to \infty$ (where $P_{\boldsymbol{\theta}}^{(n)}$ and $P_{\boldsymbol{\theta}}$ are the $\boldsymbol{\theta}$-symmetrized versions of $P^{(n)}$ and P, respectively) then entail that there exists an integer n_0 such that

$$R_{\bar{\alpha}+\varepsilon}(P) \subset R_{\bar{\alpha}}(P_{\boldsymbol{\theta}}^{(n)}) \subset R_{\bar{\alpha}-\varepsilon}(P) \subset R_\alpha(P)$$

almost surely for all $n \geq n_0$. From Lemma 4.1 again, there exists $\xi > 0$ such that $B_{\boldsymbol{\theta}}(\xi) \subset R_{\bar{\alpha}+\varepsilon}(P)$. Hence, for any $n \geq n_0$, one has that

$$B_{\boldsymbol{\theta}}(\xi) \subset R_{\bar{\alpha}}(P_{\boldsymbol{\theta}}^{(n)}) \subset B_{\boldsymbol{\theta}}(a)$$

almost surely.

Putting $N_n = \sum_{i=1}^n \mathbb{I}[\mathbf{X}_i \in B_{\boldsymbol{\theta}}(\xi)]$, the SLLN yields that $N_n/n \to P[B_{\boldsymbol{\theta}}(\xi)] = P[B_{\boldsymbol{\theta}}(\xi)] > 0$ as $n \to \infty$, since $X \sim P$ admits a density that, from continuity, is positive over a neighborhood of $\boldsymbol{\theta}$. Since $k_n = o(n)$ as $n \to \infty$, this implies that, for all $n \geq \tilde{n}_0 (\geq n_0)$,

$$\sum_{i=1}^n \mathbb{I}[\mathbf{X}_i \in R_{\bar{\alpha}}(P_{\boldsymbol{\theta}}^{(n)})] \geq N_n \geq k_n$$

almost surely. It follows that, for such values of n,

$$R_{\boldsymbol{\theta}}^{\beta_n}(P^{(n)}) = R^{\beta_n}(P_{\boldsymbol{\theta}}^{(n)}) \subset R_{\bar{\alpha}}(P_{\boldsymbol{\theta}}^{(n)}) \subset B_{\boldsymbol{\theta}}(a)$$

almost surely, with $\beta_n = k_n/n$. Therefore, $\max_{i=1,\ldots,K_{\boldsymbol{\theta}}^{\beta_n(n)}} \|\mathbf{X}_{\boldsymbol{\theta},(i)} - \boldsymbol{\theta}\| \leq a$ almost surely for large n, which yields the result. $\square$

Lemma 3. *For a "plug-in" classification rule $\tilde{m}^{(n)}(\boldsymbol{x}) = \mathbb{I}[\tilde{\eta}^{(n)}(\boldsymbol{x}) > 1/2]$ obtained from a regression estimator $\tilde{\eta}^{(n)}(\boldsymbol{x})$ of $\eta(\boldsymbol{x}) = E[\mathbb{I}[Y = 1] \,|\, \mathbf{X} = \boldsymbol{x}]$, one has that $P[\tilde{m}^{(n)}(\mathbf{X}) \neq Y] - L_{\mathrm{opt}} \leq 2 \left(E[(\tilde{\eta}^{(n)}(\mathbf{X}) - \eta(\mathbf{X}))^2] \right)^{1/2}$, where $L_{\mathrm{opt}} = P[m_{\mathrm{Bayes}}(\mathbf{X}) \neq Y]$ is the probability of misclassification of the Bayes rule.*

Proof of Lemma 3. Corollary 6.1 in Devroye *et al.* (1996) states that

$$P[\tilde{m}^{(n)}(\mathbf{X}) \neq Y \,|\, \mathcal{D}_n] - L_{\mathrm{opt}} \leq 2 \, E[|\tilde{\eta}^{(n)}(\mathbf{X}) - \eta(\mathbf{X})| \,|\, \mathcal{D}_n],$$

where $\mathcal{D}_n$ stands for the sigma-algebra associated with the training sample $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$. Taking expectations in both sides of this inequality and applying Jensen's inequality readily yields the result. $\square$

Proof of Theorem 3.1. From Bayes' theorem, $\mathbf{X}$ admits the density $\boldsymbol{x} \mapsto f(\boldsymbol{x}) = \pi_0 f_0(\boldsymbol{x}) + \pi_1 f_1(\boldsymbol{x})$. Letting $\mathrm{Supp}_+(f) = \{\boldsymbol{x} \in \mathbb{R}^d : f(\boldsymbol{x}) > 0\}$ and writing $C(f_j)$ for the collection of continuity points of f_j, $j = 0, 1$, put $N = \mathrm{Supp}_+(f) \cap C(f_0) \cap C(f_1)$. Since, by assumption, $\mathbb{R}^d \setminus C(f_j)$ $(j = 0, 1)$

has Lebesgue measure zero, we have that

$$P[\mathbf{X} \in \mathbb{R}^d \setminus N] \quad \leq \quad P[\mathbf{X} \in \mathbb{R}^d \setminus \text{Supp}_+(f)] + \sum_{j \in \{0,1\}} P[\mathbf{X} \in \mathbb{R}^d \setminus C(f_j)]$$

$$= \quad \int_{\mathbb{R}^d \setminus \text{Supp}_+(f)} f(\boldsymbol{x}) \, dx = 0,$$

so that $P[\mathbf{X} \in N] = 1$. Note also that $\boldsymbol{x} \mapsto \eta(\boldsymbol{x}) = \pi_1 f_1(\boldsymbol{x})/(\pi_0 f_0(\boldsymbol{x}) + \pi_1 f_1(\boldsymbol{x}))$ is continuous over N.

Fix $\boldsymbol{x} \in N$ and let $Y_{\boldsymbol{x},(i)} = Y_{j(\boldsymbol{x})}$ with $j(\boldsymbol{x})$ such that $\mathbf{X}_{\boldsymbol{x},(i)} = \mathbf{X}_{j(\boldsymbol{x})}$. With this notation, the estimator $\hat{\eta}_D^{(n)}(\boldsymbol{x})$ from Section 3.1 rewrites

$$\hat{\eta}_D^{(n)}(\boldsymbol{x}) = \sum_{i=1}^n Y_i W_i^{\beta(n)}(\boldsymbol{x}) = \frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} Y_{\boldsymbol{x},(i)}.$$

Proceeding as in Biau *et al.* (2012), we therefore have that (writing for simplicity β instead of β_n in the rest of the proof)

$$T^{(n)}(\boldsymbol{x}) := E[(\hat{\eta}_D^{(n)}(\boldsymbol{x}) - \eta(\boldsymbol{x}))^2] \leq 2T_1^{(n)}(\boldsymbol{x}) + 2T_2^{(n)}(\boldsymbol{x}),$$

with

$$T_1^{(n)}(\boldsymbol{x}) = E\left[\left|\frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} \left(Y_{\boldsymbol{x},(i)} - \eta(\mathbf{X}_{\boldsymbol{x},(i)})\right)\right|^2\right]$$

and

$$T_2^{(n)}(\boldsymbol{x}) = E\left[\left|\frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} \left(\eta(\mathbf{X}_{\boldsymbol{x},(i)}) - \eta(\boldsymbol{x})\right)\right|^2\right].$$

Writing $\mathcal{D}_X^{(n)}$ for the sigma-algebra generated by $\mathbf{X}_i$, $i = 1, \ldots, n$, note that, conditional on $\mathcal{D}_X^{(n)}$, the $Y_{\boldsymbol{x},(i)} - \eta(\mathbf{X}_{\boldsymbol{x},(i)})$'s, $i = 1, \ldots, n$, are zero mean mutually independent random variables. Consequently,

$$T_1^{(n)}(\boldsymbol{x}) \quad = \quad E\left[\frac{1}{(K_{\boldsymbol{x}}^{\beta(n)})^2} \sum_{i,j=1}^{K_{\boldsymbol{x}}^{\beta(n)}} E\left[(Y_{\boldsymbol{x},(i)} - \eta(\mathbf{X}_{\boldsymbol{x},(i)}))(Y_{\boldsymbol{x},(j)} - \eta(\mathbf{X}_{\boldsymbol{x},(j)})) \,\middle|\, \mathcal{D}_X^{(n)}\right]\right]$$

$$= \quad E\left[\frac{1}{(K_{\boldsymbol{x}}^{\beta(n)})^2} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} E\left[(Y_{\boldsymbol{x},(i)} - \eta(\mathbf{X}_{\boldsymbol{x},(i)}))^2 \,\middle|\, \mathcal{D}_X^{(n)}\right]\right]$$

$$\leq \quad E\left[\frac{4}{K_{\boldsymbol{x}}^{\beta(n)}}\right] \leq \frac{4}{k_n} = o(1),$$

as $n \to \infty$, where we used the fact that $K_{\boldsymbol{x}}^{\beta(n)} \geq k_n$ almost surely. As for $T_2^{(n)}(\boldsymbol{x})$, the Cauchy-

Schwarz inequality yields (for an arbitrary $a > 0$)

$$
\begin{aligned}
T_2^{(n)}(\boldsymbol{x}) &\leq E\left[\frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} \left(\eta(\mathbf{X}_{\boldsymbol{x},(i)}) - \eta(\boldsymbol{x})\right)^2\right] \\
&= E\left[\frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} \left(\eta(\mathbf{X}_{\boldsymbol{x},(i)}) - \eta(\boldsymbol{x})\right)^2 \mathbb{I}[\|\mathbf{X}_{\boldsymbol{x},(i)} - \boldsymbol{x}\| \leq a]\right] \\
&\quad + E\left[\frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} \left(\eta(\mathbf{X}_{\boldsymbol{x},(i)}) - \eta(\boldsymbol{x})\right)^2 \mathbb{I}[\|\mathbf{X}_{\boldsymbol{x},(i)} - \boldsymbol{x}\| > a]\right] \\
&\leq \sup_{y \in B_{\boldsymbol{x}}(a)} |\eta(y) - \eta(\boldsymbol{x})|^2 + 4\, E\left[\frac{1}{K_{\boldsymbol{x}}^{\beta(n)}} \sum_{i=1}^{K_{\boldsymbol{x}}^{\beta(n)}} \mathbb{I}[\|\mathbf{X}_{\boldsymbol{x},(i)} - \boldsymbol{x}\| > a]\right] \\
&=: \tilde{T}_2(\boldsymbol{x}; a) + \bar{T}_2^{(n)}(\boldsymbol{x}; a).
\end{aligned}
$$

Continuity of η at $\boldsymbol{x}$ implies that, for any $\varepsilon > 0$, one may choose $a = a(\varepsilon) > 0$ so that $\tilde{T}_2(\boldsymbol{x}; a(\varepsilon)) < \varepsilon$. Since Lemma 2 readily yields that $\bar{T}_2^{(n)}(\boldsymbol{x}; a(\varepsilon)) = 0$ for large n, we conclude that $T_2^{(n)}(\boldsymbol{x})$—hence also $T^{(n)}(\boldsymbol{x})$—is $o(1)$. The Lebesgue dominated convergence theorem then yields that $E[(\hat{\eta}_D^{(n)}(\mathbf{X}) - \eta(\mathbf{X}))^2]$ is $o(1)$. Therefore, using the fact that $P[\hat{m}_D^{(n)}(\mathbf{X}) \neq Y \,|\, \mathcal{D}_n] \geq L_{\text{opt}}$ almost surely and applying Lemma 3, we obtain

$$
E\left[|P[\hat{m}_D^{(n)}(\mathbf{X}) \neq Y \,|\, \mathcal{D}_n] - L_{\text{opt}}|\right] = E[P[\hat{m}_D^{(n)}(\mathbf{X}) \neq Y \,|\, \mathcal{D}_n] - L_{\text{opt}}]
$$

$$
= P[\hat{m}_D^{(n)}(\mathbf{X}) \neq Y] - L_{\text{opt}} \leq 2\left(E[(\hat{\eta}_D^{(n)}(\mathbf{X}) - \eta(\mathbf{X}))^2]\right)^{1/2} = o(1),
$$

as $n \to \infty$, which establishes the result. $\qquad\square$

Finally, we show that Properties (Q1)-(Q3) hold for several classical statistical depth functions.

Theorem 1. Properties (Q1)-(Q3) hold for (i) the halfspace depth and (ii) the simplicial depth. (iii) If the location and scatter functionals $\boldsymbol{\mu}(P)$ and $\boldsymbol{\Sigma}(P)$ are such that (a) $\boldsymbol{\mu}(P) = \boldsymbol{\theta}$ as soon as the probability measure P is symmetric about $\boldsymbol{\theta}$ and such that (b) the empirical versions $\boldsymbol{\mu}(P^{(n)})$ and $\boldsymbol{\Sigma}(P^{(n)})$ associated with an i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from P are strongly consistent for $\boldsymbol{\mu}(P)$ and $\boldsymbol{\Sigma}(P)$, then Properties (Q1)-(Q3) also hold for the Mahalanobis depth.

Proof of Theorem 1. (i) The continuity of D in Property (Q1) actually holds under the only assumption that P admits a density with respect to the Lebesgue measure; see Proposition 4 in Rousseeuw & Ruts (1999). Property (Q2) is a consequence of Theorems 1 and 2 in Rousseeuw & Struyf (2004) and the fact that the angular symmetry center is unique for absolutely continuous distributions; see Serfling (2006). For halfspace depth, Property (Q3) follows from (6.2) and (6.6) in Donoho & Gasko (1992).

(ii) The continuity of D in Property (Q1) actually holds under the only assumption that P admits a density with respect to the Lebesgue measure; see Theorem 2 in Liu (1990). Remark C in Liu (1990) shows that, for an angularly symmetric probability measure (hence also for a centrally symmetric probability measure) admitting a density, the symmetry center is the unique point maximizing simplicial depth provided that the density remains positive in a neighborhood of the symmetry

center; Property (Q2) trivially follows. Property (Q3) for simplicial depth is stated in Corollary 1 of Dümbgen (1992).

(iii) This is trivial. □

Finally, note that Properties (Q1)-(Q3) also hold for projection depth under very mild assumptions on the univariate location and scale functionals used in the definition of projection depth; see Zuo (2003).

References

Biau, G., Devroye, L., Dujmovic, V., & Krzyzak, A. (2012). An affine invariant k-Nearest Neighbor regression estimate. *J. Multivariate Anal.*, *112*, 24–34.

Chacón, J. E. (2009). Data-driven choice of the smoothing parametrization for kernel density estimators. *Canad. J. Statist.*, *37*, 249–265.

Chacón, J. E., Duong, T., & Wand, M. P. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Statist. Sinica*, *21*, 807–840.

Cui, X., Lin, L., & Yang, G. (2008). An extended projection data depth and its applications to discrimination. *Comm. Statist. Theory Methods*, *37*, 2276–2290.

Dehon, C. & Croux, C. (2001). Robust linear discriminant analysis using S-estimators. *Canad. J. Statist.*, *29*, 473–492.

Devroye, L., Györfi, L., & Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition (Stochastic Modelling and Applied Probability)*. Springer, New York.

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

Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statist. Probab. Lett.*, *14*, 119–128.

Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, *50*, 471–491.

Dutta, S. & Ghosh, A. K. (2012a). On robust classification using projection depth. *Ann. Inst. Statist. Math.*, *64*, 657–676.

Dutta, S. & Ghosh, A. K. (2012b). On classification based on L_p depth with an adaptive choice of p. submitted. Tech. rep., Technical Report Number R5/2011, Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata, India.

Ghosh, A. K. & Chaudhuri, P. (2005a). On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli*, *11*, 1–27.

Ghosh, A. K. & Chaudhuri, P. (2005b). On maximum depth and related classifiers. *Scand. J. Statist.*, *32*, 327–350.

Hartikainen, A. & Oja, H. (2006). On some parametric, nonparametric and semiparametric discrimination rules. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, *72*, 61–70.

He, X. & Fung, W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *J. Multivariate Anal.*, *72*, 151–162.

Hettmansperger, T. P. & Randles, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika*, *89*, 851–860.

Hoberg, A. & Mosler, K. (2006). Data analysis and classification with the zonoid depth. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, *72*, 49–59.

Hubert, M. & Van der Veeken, S. (2010). Robust classification for skewed data. *Adv. Data Anal. Classif.*, *4*, 239–254.

Jörnsten, R. (2004). Clustering and classification based on the L_1 data depth. *J. Multivariate Anal.*, *90*, 67–89.

Koshevoy, G. & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.*, *25*, 1998–2017.

Lange, T., Mosler, K., & Mozharovskyi, P. (2012). Fast nonparametric classification based on data depth. *Stat. Papers*. To appear.

Li, J., Cuesta-Albertos, J., & Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedures based on dd-plots. *J. Amer. Statist. Assoc.*, *107*(498), 737–753.

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

Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.*, *27*, 783–840.

Oja, H. & Paindaveine, D. (2005). Optimal signed-rank tests based on hyperplanes. *J. Statist. Plann. Inference*, *135*, 300–323.

Randles, R. H., Broffitt, J. D., Ramberg, J. S., & Hogg, R. V. (1978). Generalized linear and quadratic discriminant functions using robust estimates. *J. Amer. Statist. Assoc.*, *73*, 564–568.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

Rousseeuw, P. J. & Ruts, I. (1999). The depth function of a population distribution. *Metrika*, *49*, 213–244.

Rousseeuw, P. J. & Struyf, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference*, *122*, 161–173.

Serfling, R. J. (2006). Multivariate symmetry and asymmetry. *Encyclopedia statist. sci.*, *8*, 5338–5345.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, *5*, 595–620.

Tukey, J. W. (1975). Mathematics and the picturing of data. *Proc. Internat. Cong. Math.*, *2*, 523–531.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, *15*, 234–251.

Yeh, I., Yang, K., & Ting, T. (2009). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Syst. Appl.*, *36*, 5866–5871.

Zakai, A. & Ritov, Y. (2009). Consistency and localizability. *J. Mach. Learn. Res.*, *10*, 827–856.

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

Zuo, Y. & Serfling, R. (2000a). General notions of statistical depth function. *Ann. Statist.*, *28*, 461–482.

Zuo, Y. & Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.*, *28*, 483–499.

Chapter II

From Depth to Local Depth : A Focus on Centrality

From Depth to Local Depth : A Focus on Centrality[1]

Davy PAINDAVEINE[a,b,*] and Germain VAN BEVER[a,c].

[a] *ECARES & Département de Mathématique, Université libre de Bruxelles (ULB)*
[b] *Email: dpaindav@ulb.ac.be*
[c] *Email: gvbever@ulb.ac.be*
[*] *Corresponding Author. Address: Université libre de Bruxelles, ECARES, Avenue F.D. Roosevelt, 50, CP 114/04, B-1050 Brussels, Belgium.*

Abstract Aiming at analysing multimodal or non-convexly supported distributions through data depth, we introduce a local extension of depth. Our construction is obtained by conditioning the distribution to appropriate depth-based neighborhoods, and has the advantages, among others, to maintain affine-invariance and to apply to all depths in a generic way. Most importantly, unlike their competitors, that (for extreme localization) rather measure probability mass, the resulting *local depths* focus on centrality and remain of a genuine depth nature at any locality level. We derive their main properties, establish consistency of their sample versions, and study their behavior under extreme localization. We present two applications of the proposed local depth (for classification and for symmetry testing), and we extend our construction to the regression and functional depth contexts. Throughout, we illustrate the results on some, artificial and real, univariate and multivariate data sets.

Keywords: Classification · functional depth · multimodality · non-convex support · regression depth · statistical depth functions · symmetry testing.

---

1 Introduction

Data depth was originally introduced as a way to generalize the concept of median to the multivariate setup but has long been known in the statistical literature as a powerful data analytic tool able to reveal very diverse features of the underlying distribution. Indeed, not only does depth provide a robust multivariate location functional (through the deepest point), it also yields information about spread, shape, and symmetry (through depth regions, Serfling, 2004), and even characterizes the underlying distribution under very mild conditions (see Kong & Zuo, 2010 and the references therein). Celebrated instances of such depths include Tukey's halfspace depth (Tukey, 1975), Liu's simplicial depth (Liu, 1990), the projection depth (Zuo, 2003), or the Mahalanobis depth (see, e.g., Zuo & Serfling, 2000a). Depth methods allow to address several inference problems, including, e.g., testing for location and scale differences based on the DD-plot (first introduced as a graphical display for data exploration; Liu *et al.* (1999), Li & Liu (2004)), multivariate extensions of Wilcoxon rank sum tests (Liu & Singh, 1993), diagnostics of non-normality (Liu *et al.*, 1999), and outlier detection (Chen *et al.*, 2009). More recently, depth was used extensively in a classification context (see, among many others, Ghosh & Chaudhuri, 2005; Li *et al.*, 2012; Dutta & Ghosh, 2012; Paindaveine & Van Bever, 2012).

Depth deals with centrality. Its first purpose is to provide a *center-outward ordering* from the deepest point—the center of the distribution[2]—towards less deep, exterior points. Classical depth functions indeed associate with any center of symmetry (should it exist) a maximal depth value. Together with the fact that depth decreases along any halfline originating from any deepest point, this leads to nested star-shaped (in most cases, convex) depth regions, whatever the underlying distribution may be (depth/quantile regions that may be non-convex are defined in Wei (2008)). That is the reason why it is often reported that depth is suitable for unimodal convexly-supported distributions only; see, e.g., Zuo & Serfling (2000a); Lok & Lee (2011); Izem *et al.* (2008); Hlubinka *et al.* (2010). Distributions that are multimodal or have a non-convex support, however, are met in many fields of applications, among which, obviously, those involving mixture models or clustering problems (see, e.g., McLachlan & Basford, 1988, or McLachlan & Peel, 2000). This motivates extending the concept of depth to make it flexible enough to deal with such distributions.

A few such extensions are available in the literature, under the name of *local depths*. In particular, Agostinelli & Romanazzi (2011) introduced local versions of the halfspace and simplicial depths. For halfspace depth, locality is achieved by replacing halfspaces with finite-width slabs, while, for simplicial depth, it is obtained by restricting to simplices with a volume smaller than some fixed threshold. When considering all possible values of the locality parameter involved, these local depths—after adequate standardization—provide a continuum between (global) halfspace/simplicial depths and the density of the underlying distribution. Density and depth, however, are antinomic in spirit : for instance, the symmetry center of a centrally symmetric bimodal distribution always assumes maximal depth while the density may very well be zero there; also, uniform distributions have non-trivial depth contours but do not show proper equidensity contours.

Similarly, other proposals for local depth—or, more generally, other extensions of depth aiming at distributions with possibly non-convex supports—converge, as locality becomes extreme, to either a density measure (Hlubinka *et al.*, 2010) or a constant value (Chen *et al.*, 2009), hence lose their nature of a centrality measure. The purpose of this paper is to introduce a new concept of local

---

[2]Uniqueness may fail to hold; however, the maximizers typically form a convex region.

depth that, at any locality level, remains of a genuine depth nature and provides a measure of *local* centrality. Our construction will actually allow to turn, in a common generic way, *any* (global) depth into a corresponding local depth. This is another advantage over the competing local depths, that focus on a specific depth (Hlubinka *et al.*, 2010; Chen *et al.*, 2009) or require a specific definition for each global depth considered (Agostinelli & Romanazzi, 2011). The proposed local depth is defined as global depth conditional on some neighborhood of the point of interest. To make this local concept purely based on depth, we use the neighborhoods that were recently introduced (for classification purposes) in Paindaveine & Van Bever (2012). As we will show, the resulting local depths allow for interesting inferential applications.

The outline of the paper is as follows. In Section 2, we illustrate our local depth concept on two real data sets, that highlight the need for this extension from global to local centrality and allow for a comparison with the local depths from Agostinelli & Romanazzi (2011). In Section 3, we first review the basics of depth (Section 3.1). We then describe the depth-based neighborhoods from Paindaveine & Van Bever (2012) and show how they allow to define local depth (Section 3.2). We also establish consistency of the corresponding sample local depth (Section 3.3). Section 4 is dedicated to the limit behavior of the proposed local depth as locality becomes extreme. Section 5 illustrates the results of the previous sections on several univariate and multivariate examples. Section 6 presents two inferential applications of the proposed local depth concept. In Section 7, we show that our construction extends to the regression and functional depth contexts. Computational aspects are discussed in Section 8. Finally, the Appendix collects technical proofs.

2   Motivating examples

As mentioned above, we introduce a concept of local depth that can cope with multimodal and/or non-convexly supported distributions. Here we illustrate this on the basis of two real data sets, that are freely available in the well-known R package *MASS* (the first one provides a univariate bimodal example, whereas the second one involves a bivariate distribution with a non-convex support). Inferential applications based on the proposed local depth are deferred to Section 6.

2.1   Geyser data

The *Geyser* data set is related to eruption data from the *Old Faithful geyser* in the Yellowstone National Park, Wyoming, USA (see Härdle, 1991). It contains $n = 299$ measurements of two variables : "duration" (duration, in minutes, of the eruption) and "waiting" (waiting time, still in minutes, between two eruptions). As we want to start with a univariate data set, we focus here on the bimodal variable "waiting".

Figure 1 starts with reporting a histogram of the waiting times (upper left), together with the halfspace and simplicial depths of 100 equispaced values in the range of interest (upper right). The lower subplots draw the proposed local halfspace and simplicial depths at locality levels $\beta = .7$ (intermediate localization) and .3 (more extreme localization) ; in the present univariate setup, we simply define the local depth of a waiting time x, at locality level β, as the (global) depth of x with respect to the $\lceil n\beta \rceil$ observed waiting times that are closest to x. For the sake of comparison, we also report the local halfspace and simplicial depths from Agostinelli & Romanazzi (2011), at locality levels $\tau = 23$ and 7 (for proper comparison, these τ-values, as in Agostinelli & Romanazzi (2011), were selected as the .7- and .3-quantiles of the $\binom{n}{2}$ distances between observed waiting times; in

order to avoid these local depth functions collapsing to zero as $\tau \to 0$, they were scaled so that the deepest waiting time receives depth $1/2$ in each case).

With the exception of the Agostinelli & Romanazzi (2011) ($\tau = 23$)-local simplicial depth, all local depths clearly show the obvious multimodality that is missed by global depth. For more extreme localization, all local depths reveal both local modes about 55 and 80. Unlike our local depths, that attribute comparable depth values to both local modes, the Agostinelli & Romanazzi (2011) local depths, that, at such locality levels, are not local centrality measures but rather density measures, clearly reflect the heterogeneous probability masses around the two local modes.

For $\beta = 0.3$, the proposed local depths show a third local center (about $x_0 = 65$ minutes), which is in line with the fact that, at this locality level, the distribution is nearly symmetric about x_0, so that it should receive a large (local) centrality measure. If needed, discriminating between the two "true" local modes and this "artificial" mode about x_0 may e.g. be based on the corresponding depth-based neighborhoods involved (See Section 3.2 below), that are much wider at x_0 than at both "true" modes. Detecting modes, however, is not one of the primary applications of the proposed local depth concept ; see Section 6 for such applications.
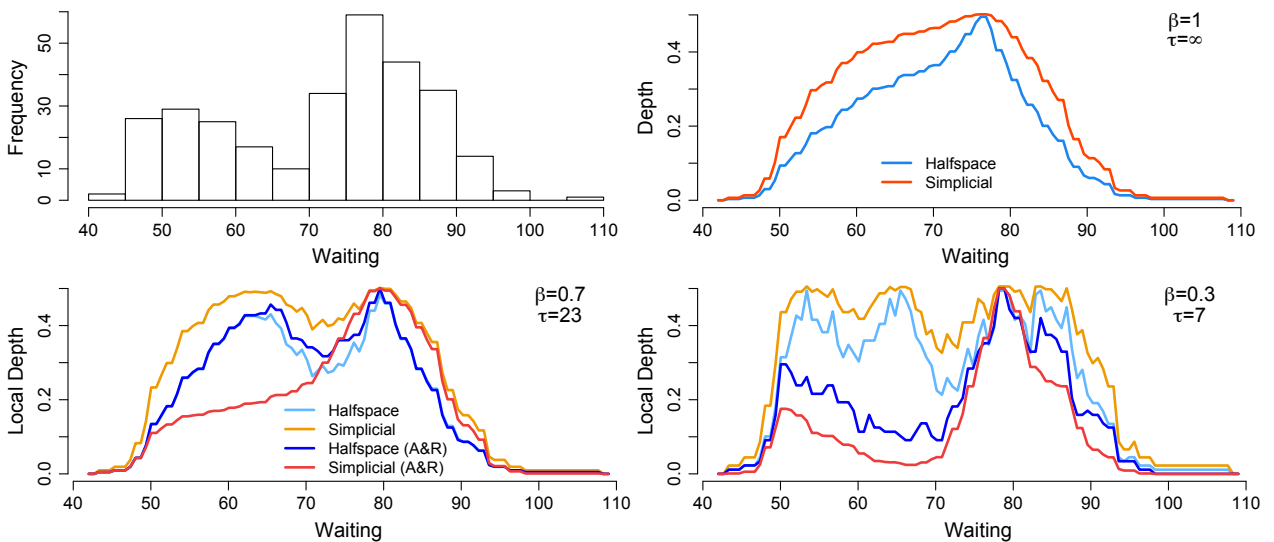

Figure 1: *(Upper left:) Histogram of the variable "Waiting" from the* Geyser *data set. (Upper right): Plots of halfspace (blue) and simplicial (orange-red) depths over 100 equispaced points. (Lower:) the proposed local halfspace (light blue) and simplicial (orange) depths at locality levels $\beta \in \{.7, .3\}$, along with their halfspace (dark blue) and simplicial (red) counterparts from Agostinelli & Romanazzi (2011) for locality levels $\tau \in \{27, 3\}$.*

2.2   Boston data

The *Boston* data set was first introduced in Harrison & Rubinfeld (1978). It contains 506 observations related to housing and was first used to estimate the "need for clean air" in the Boston area. The data set originally contains 14 different variables. For the sake of illustration, we restrict here to two variables, namely "NOX" (annual average of nitrogen oxide concentration, in parts per ten million) and "DIS" (the weighted mean of distances to five Boston employment centers, in miles). The upper left panel of Figure 2 shows a scatter plot of the resulting 506 bivariate data points.

This scatter plot shows that the data set has a non-convex support; this entails that there may be points whose respective depth values do not reflect properly what one would naturally consider the relative centrality of these points in the data set. To illustrate this, we consider four particular locations, marked in orange, blue, red, and green in the scatter plot. Both for halfspace and simplicial depths, the green location is considered more central than the blue one, which is somehow paradoxical since the green location is much closer to the boundary of the support. Similarly, the red location—that actually is the halfspace deepest one—is about twice as (halfspace or simplicial) deep as the blue location, while visual inspection suggests that the latter is more central than the former (or at least is of comparable centrality).

Parallel to the univariate case, the β-local depth of a point $\boldsymbol{x} \in \mathbb{R}^2$ is still obtained as the global depth of $\boldsymbol{x}$ with respect to the data points sitting in a neighborhood of $\boldsymbol{x}$ containing a proportion β of the observations (the exact definition of this neighborhood, that is actually of a depth-based nature, will be provided in Section 3.2). The upper right panel of Figure 2 shows the plots of the proposed local (halfspace and simplicial) depths for the four locations above, as a function of the locality level β. As β moves away from one (that still corresponds to going from global depth to more and more local depth), the paradoxes above vanish : both the green location (that is close to the boundary of the support) and the red location show decreasing local depths that eventually fall below the local depth of the blue one. Note that, except for very small β-values[3], the orange location has uniformly low local depth, which is expected since it is close to the boundary of the convex hull of the data (would this point be outside the convex hull, its local depth would be zero for any β).

The lower left panel of Figure 2 plots scaled versions of the Agostinelli & Romanazzi (2011) local halfspace and simplicial depths of the four locations, as a function of the locality level τ in some appropriate range[4]; scaling was performed in such a way that, at any fixed τ, the largest τ-local (halfspace and simplicial) depths considered are equal to one (this still allows to investigate, for any fixed τ, the corresponding (halfspace and simplicial) depth rankings, and was done because those local depths are hardly comparable for different τ-values). It is seen that the (local) depth rankings depend much more on the choice of (halfspace or simplicial) depth than for the proposed local depths (particularly so for the green location). For halfspace depth, the red point remains the deepest for most τ-values; it actually is so for all τ-values in the lower right panel of Figure 2, that reports the corresponding local depths after a unit change expressing the DIS variable in yards (this consists in multiplying DIS by 1760, but the results are similar for much smaller factors). The particular τ-indexing used for local simplicial depth makes it affine-invariant, but local halfspace depth fails to be so, irrespective of the τ-indexing used; the unit change considered, unpleasantly, has a strong impact on the local halfspace depth from Agostinelli & Romanazzi (2011) (the τ-local halfspace depth of the green location now dramatically decreases for small τ-values, and, as mentioned above, the red location remains, for all τ, the halfspace deepest point among all locations). In contrast, our local depths are affine-invariant, hence are not affected by any unit change.

---

[3]Actually, little attention should be paid to small β-values, as the corresponding local depths are computed on the basis of very few observations in each neighborhood. When investigating extreme locality, it is important to choose β as a function of the sample size n; in some sense, β is a smoothing parameter, pretty much as the bandwidth in a density estimation context.

[4]For halfspace depth, the maximum value $\tau_{\max}$ of τ was chosen as the minimal τ-value for which the τ-local depths of the four locations all coincide with the corresponding global halfspace depths ; for simplicial depth, the τ-values at which the τ-local depths were evaluated are, as suggested in Agostinelli & Romanazzi (2011), the percentiles of the volumes of the $\binom{n}{3}$ data-based simplices, which also ensures that global depth is obtained for the largest τ considered.
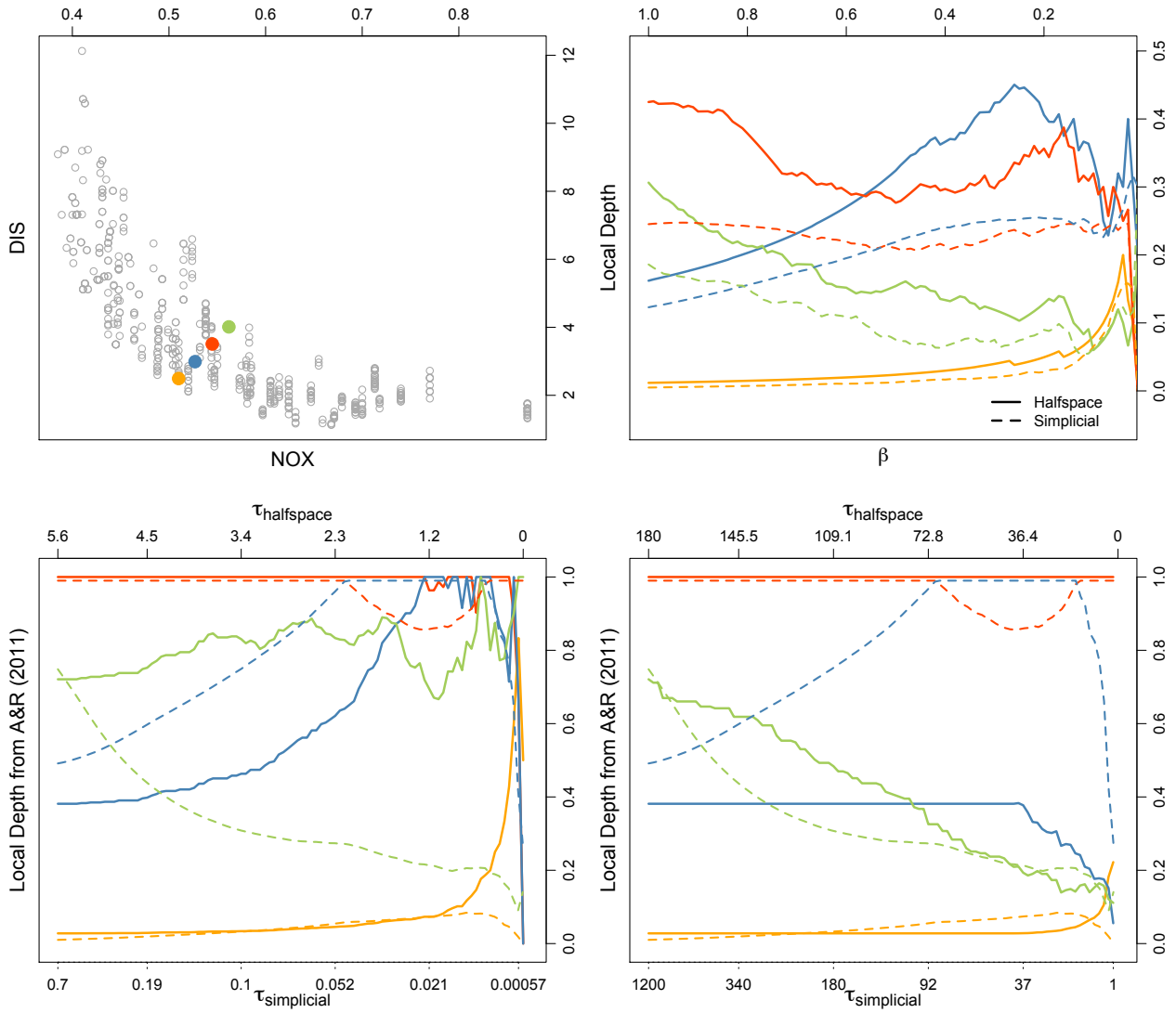
Figure 2: *(Upper left:) Scatterplot of the NOX and DIS variables from the Boston data set, with four particular locations. (Upper right:) Plots, as a function of the locality level β, of the proposed local halfspace (solid curves) and simplicial (dashed curves) depths of these locations. (Lower left:) scaled versions of the corresponding Agostinelli & Romanazzi (2011) local depths ; see Section 2.2 for details. (Lower right:) The same curves as in the lower left panel, when expressing the DIS variable in yards (such unit change does not affect the local depths in the upper right panel).*

3 From global to local depth

In this section, we first review the concept of depth (Section 3.1). We then explain how it can be used to construct neighborhoods of any $\boldsymbol{x} \in \mathbb{R}^d$, and propose a local version of any depth (Section 3.2). Finally, we define the sample local depths that were already put at work in Section 2, and establish their consistency (Section 3.3).

3.1 Depth functions

A depth function $D(\,\cdot\,, P)$ associates with any $\boldsymbol{x} \in \mathbb{R}^d$ a measure $D(\boldsymbol{x}, P)(\geq 0)$ of its *centrality* with respect to the probability measure P over $\mathbb{R}^d$ (the more *central* $\boldsymbol{x}$ is, the *deeper* it is). The two most celebrated depths are the following.

Definition 3.1 (Tukey, 1975)**.** Denoting by $\mathcal{S}^{d-1}$ the set of unit vectors in $\mathbb{R}^d$, the *halfspace depth* of $\boldsymbol{x}$ with respect to P is the "minimal" probability of all halfspaces containing $\boldsymbol{x}$, i.e., $D_H(\boldsymbol{x}, P) = \inf_{\boldsymbol{u} \in \mathcal{S}^{d-1}} P\left[\boldsymbol{u}'(\mathbf{X} - \boldsymbol{x}) \geq 0\right]$, where $\mathbf{X} \sim P$.

Definition 3.2 (Liu, 1990)**.** Letting $S(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1})$ be the convex hull of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1}$, the *simplicial depth* of $\boldsymbol{x}$ with respect to P is $D_S(\boldsymbol{x}, P) = P\left[\boldsymbol{x} \in S(\mathbf{X}_1, \ldots, \mathbf{X}_{d+1})\right]$, where $\mathbf{X}, \ldots, \mathbf{X}_{d+1}$ are i.i.d. from P.

There are numerous other concepts of depth, including the spatial depth (Chaudhuri, 1996), the standardized spatial depth (Serfling, 2010), the projection depth (Zuo, 2003), the Mahalanobis depth (Zuo & Serfling, 2000a), the zonoid depth (Koshevoy & Mosler (1997)), the simplicial volume depth (Oja, 1983; Zuo & Serfling, 2000a), etc. The halfspace depth and—under absolute continuity—the simplicial depth are *statistical depth functions*, in the following sense.

Definition 3.3 (Zuo & Serfling, 2000a)**.** A bounded mapping $D(\,\cdot\,, P)$ from $\mathbb{R}^d$ to $\mathbb{R}^+$ is a *statistical depth function* if it satisfies the four following properties :

(P1) *affine-invariance:* for any $d \times d$ invertible matrix $\mathbf{A}$, any d-vector $\mathbf{b}$, and any distribution P on $\mathbb{R}^d$, $D(\mathbf{A}\boldsymbol{x} + \mathbf{b}, P^{\mathbf{A},\mathbf{b}}) = D(\boldsymbol{x}, P)$, where $P^{\mathbf{A},\mathbf{b}}$ stands for the distribution of $\mathbf{A}\mathbf{X} + \mathbf{b}$ when $\mathbf{X}$ has distribution P;

(P2) *maximality at center:* if $\boldsymbol{\theta}$ is a center of (central, angular or halfspace) symmetry of P, then it holds that $D(\boldsymbol{\theta}, P) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}, P)$;

(P3) *monotonicity relative to deepest point:* for any P having deepest point $\boldsymbol{\theta}$, $D(\boldsymbol{x}, P) \leq D((1-\lambda)\boldsymbol{\theta} + \lambda\boldsymbol{x})$ for any $\boldsymbol{x}$ in $\mathbb{R}^d$ and any $\lambda \in [0, 1]$;

(P4) *vanishing at infinity:* for any P, $D(\boldsymbol{x}, P) \to 0$ as $||\boldsymbol{x}|| \to \infty$.

For any depth function, the *depth regions* $R_\alpha(P) = \{\boldsymbol{x} \in \mathbb{R}^d \,|\, D(\boldsymbol{x}, P) \geq \alpha\}$ (of *order* $\alpha > 0$) are of paramount importance as they reveal very diverse characteristics from P: location, dispersion, dependence structure, etc. (see, e.g., Liu *et al.*, 1999). Clearly, these regions are nested, and inner regions contain points with larger depth. When defining local depth below, it will be more appropriate to index the family $\{R_\alpha(P)\}$ by means of probability contents: for any $\beta \in [0, 1]$, we define

$$R^\beta(P) = \bigcap_{\alpha \in A(\beta)} R_\alpha(P), \quad \text{with } A(\beta) = \left\{\alpha \geq 0 : P\left[R_\alpha(P)\right] \geq \beta\right\}, \tag{3.1}$$

the smallest depth region with P-probability larger than or equal to β; we use subscripts and superscripts to denote depth regions associated with some fixed order (α) and some fixed probability content (β), respectively.

3.2 Depth-based neighborhoods and local depth

Below, we will naturally base the definition of local depth of a point $\boldsymbol{x} \in \mathbb{R}^d$ on some *neighborhoods* of $\boldsymbol{x}$ (this may seem quite natural, but is actually in contrast with all concepts of local depth available in the literature). To ensure that the resulting local depth is of a purely depth nature, we will make use of the depth-based neighborhoods from Paindaveine & Van Bever (2012), which we now describe.

Let $D(\,\cdot\,, P)$ be a depth function satisfying Properties (P2)-(P3) in Definition 3.3. For any P, the depth regions $R_\alpha(P)$ or $R^\beta(P)$ provide neighborhoods of the[5] deepest point $\boldsymbol{\theta}_P$, say. As mentioned above, we need depth-based neighborhoods of any $\boldsymbol{x} \in \mathbb{R}^d$. This may be achieved by symmetrizing the distribution P with respect to $\boldsymbol{x}$, that is by replacing $P = P^{\mathbf{X}}$ with $P_{\boldsymbol{x}} = \frac{1}{2}P^{\mathbf{X}} + \frac{1}{2}P^{2\boldsymbol{x}-\mathbf{X}}$. Properties (P2)-(P3) indeed readily imply that the resulting depth regions $R_\alpha(P_{\boldsymbol{x}})$ or $R^\beta(P_{\boldsymbol{x}})$ provide nested neighborhoods of $\boldsymbol{x}$. In line with most definitions of depth functions, the construction of these (depth-based) neighborhoods is done in a completely nonparametric way. The parameter α (resp., β) will play the role of the locality parameter, smaller neighborhoods corresponding to larger values of α (resp., to smaller values of β).

Definition 3.4 (Depth-based neighborhoods)**.** The order-α (resp., probability-β) depth-based neighborhood of $\boldsymbol{x}$ with respect to the distribution P is $R_{\boldsymbol{x},\alpha}(P) = R_\alpha(P_{\boldsymbol{x}})$ (resp., $R_{\boldsymbol{x}}^\beta(P) = R^\beta(P_{\boldsymbol{x}})$).

Before proceeding to local depth, we note that there are other symmetrization processes mapping the distribution P to a distribution $P_{\boldsymbol{x}}$ that is centro-symmetric about $\boldsymbol{x}$, such as, e.g., the one that maps $P = P^{\mathbf{X}}$ to $g(P^{\mathbf{X}}) = \frac{1}{2}P^{\boldsymbol{x}-\mathbf{X}} + \frac{1}{2}P^{\boldsymbol{x}+\mathbf{X}}$. The symmetrization process $g(\,\cdot\,)$, however, leads to less natural depth-based neighborhoods of $\boldsymbol{x}$; in particular, if $P^{\mathbf{X}}$ is spherically symmetric about $\boldsymbol{x}(\neq \boldsymbol{0})$, then the depth-based neighborhoods obtained from the symmetrization we propose will be spherically symmetric about $\boldsymbol{x}$, whereas those resulting from the symmetrization $g(\,\cdot\,)$ will not. Other possible symmetrization processes, such as $g(P^{\mathbf{X}}) = \frac{1}{4}P^{\mathbf{X}} + \frac{1}{4}P^{\mathrm{Rot}_{\boldsymbol{x}}^{\pi/2}(\mathbf{X})} + \frac{1}{4}P^{\mathrm{Rot}_{\boldsymbol{x}}^{\pi}(\mathbf{X})} + \frac{1}{4}P^{\mathrm{Rot}_{\boldsymbol{x}}^{3\pi/2}(\mathbf{X})}$, where $\mathrm{Rot}_{\boldsymbol{x}}^{\omega}$ stands for the rotation about $\boldsymbol{x}$ with angle ω (in radians), would lead to depth-based neighborhoods that are not affine-equivariant and would require more computational efforts in the sample case. These considerations motivate the proposed symmetrization process $P^{\mathbf{X}} \mapsto \frac{1}{2}P^{\mathbf{X}} + \frac{1}{2}P^{2\boldsymbol{x}-\mathbf{X}}$.

Now, conditioning on the depth-based neighborhoods from Definition 3.4 provides a local version of any depth D. More precisely, we adopt the following definition.

Definition 3.5 (Local depth)**.** Let $D(\,\cdot\,, P)$ be a depth function. The corresponding local depth function at locality level $\beta(\in (0,1])$—or simply, β-*local depth function*—is

$$LD^\beta(\,\cdot\,, P) \,:\, \mathbb{R}^d \to \mathbb{R}^+ : \boldsymbol{x} \mapsto LD^\beta(\boldsymbol{x}, P) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^\beta),$$

where $P_{\boldsymbol{x}}^\beta : B \mapsto P_{\boldsymbol{x}}^\beta[B] = P[B \cap R_{\boldsymbol{x}}^\beta(P)]/P[R_{\boldsymbol{x}}^\beta(P)]$ is the conditional (on $R_{\boldsymbol{x}}^\beta(P)$) distribution of P.

---

[5]Uniqueness of $\boldsymbol{\theta}_P$ is not guaranteed in general, so that the depth regions will rather define a neighborhood of the (convex) collection of deepest points. However, note that, for halfspace depth, uniqueness holds under the assumption of angular symmetry (Rousseeuw & Struyf, 2004). This will be sufficient for our purposes.

As announced, we favored the β-parametrization over the α-parametrization when defining local depth. The reason is twofold. First, the maximal depth order $\alpha_*(P) = \max_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}, P)$, hence also the range of relevant α-values, depends on P. Second, and more importantly, the neighborhood $R_{\boldsymbol{x},\alpha}(P)$ may have P-probability zero for α close to $\alpha_*(P)$[6], in which case the conditional distribution $P_{\boldsymbol{x},\alpha}[\,\cdot\,] = P[\,\cdot\,|R_{\boldsymbol{x},\alpha}(P)]$, hence also the local depth function $\boldsymbol{x} \mapsto LD_\alpha(\boldsymbol{x}, P) = D(\boldsymbol{x}, P_{\boldsymbol{x},\alpha})$, is not properly defined. In contrast, β-local depth is always well-defined, and the range of β-values does not depend on P, nor on the particular depth function used : β will always assume values between 0 (extreme localization) and 1 (no localization).

Unlike its competitors, this construction of local depth applies in a generic way to any depth function $D(\,\cdot\,, P)$, and it ensures affine-invariance at any fixed locality level β (which trivially follows from Property (P1)). For $\beta = 1$, the local depth clearly reduces to its global antecedent $D(\,\cdot\,, P)$, which shows that the proposed concept provides an extension of usual (global) depth. The properties of $LD^\beta(\,\cdot\,, P)$ for extreme locality—that is, as $\beta \to 0$—will be considered in Section 4.

3.3   Sample local depth and consistency

We now turn to the sample case. To do so, consider d-variate mutually independent observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with common distribution P, and denote by $P^{(n)}$ the corresponding empirical distribution. Classically, sample (global) depths are obtained by substituting $P^{(n)}$ for P in $D(\,\cdot\,, P)$, which leads, e.g., to the sample halfspace depth

$$D_H(\boldsymbol{x}, P^{(n)}) = \frac{1}{n} \inf_{\boldsymbol{u} \in \mathcal{S}^{d-1}} \#\big\{i = 1, \ldots, n : \boldsymbol{u}'(\mathbf{X}_i - \boldsymbol{x}) \geq 0\big\},$$

and the sample simplicial depth

$$D_S(\boldsymbol{x}, P^{(n)}) = \binom{n}{d+1}^{-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_{d+1} \leq n} \mathbb{I}\big[\boldsymbol{x} \in S(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_{d+1}})\big],$$

where $\mathbb{I}[B]$ stands for the indicator function of B. Sample depth regions are defined accordingly: $R_\alpha(P^{(n)})$ is defined as the collection of $\boldsymbol{x}$'s with $D(\boldsymbol{x}, P^{(n)})$ larger than or equal to α, and $R^\beta(P^{(n)})$ as the intersection of all $R_\alpha(P^{(n)})$ with $P^{(n)}$-probability larger than or equal to β. In this sample case, $R^\beta(P^{(n)})$ is thus the smallest sample depth region that contains at least a proportion β of the $\mathbf{X}_i$'s. We refer to He & Wang (1997) and Zuo & Serfling (2000c) for results on sample depth regions.

As in the population case, our sample local depth concept will require considering, for any $\boldsymbol{x} \in \mathbb{R}^d$, the symmetrized distribution $P_{\boldsymbol{x}}^{(n)}$, that is the empirical distribution associated with the $2n$ random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n, 2\boldsymbol{x} - \mathbf{X}_1, \ldots, 2\boldsymbol{x} - \mathbf{X}_n$. We are then able to define the sample version of the local concept introduced in Section 3.2.

Definition 3.6 (Sample local depth). Let $D(\,\cdot\,, P)$ be a depth function. The corresponding sample local depth function at locality level $\beta(\in (0, 1])$—or simply, *sample β-local depth function*—is

$$LD^\beta(\,\cdot\,, P^{(n)}) \,:\, \mathbb{R}^d \to \mathbb{R}^+ \,:\, \boldsymbol{x} \,\mapsto LD^\beta(\boldsymbol{x}, P^{(n)}) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\beta,(n)}),$$

where $P_{\boldsymbol{x}}^{\beta,(n)}$ denotes the empirical measure associated with those data points among $\mathbf{X}_i$, $i = 1, \ldots, n$ that sit in $R_{\boldsymbol{x}}^\beta(P^{(n)})(= R^\beta(P_{\boldsymbol{x}}^{(n)}))$.

---

[6]An example is obtained for $\boldsymbol{x} = \mathbf{0} \in \mathbb{R}^d$ and P being the distribution of $\mathbf{X}$ conditional on $[\|\mathbf{X}\| > 1]$, where $\mathbf{X}$ is standard d-variate normal.

By definition, $R_{\boldsymbol{x}}^{\beta}(P^{(n)})$ is the smallest sample depth region that contains at least a proportion β of the $2n$ random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n, 2\boldsymbol{x} - \mathbf{X}_1, \ldots, 2\boldsymbol{x} - \mathbf{X}_n$, or equivalently (symmetrization indeed implies that these depth regions are centro-symmetric about $\boldsymbol{x}$), a proportion β of the n original data points $\mathbf{X}_i$. Note that, for $k \in \{1, 2, \ldots, n-1\}$, ties may imply that $R_{\boldsymbol{x}}^{k/n}(P^{(n)})$ contains more than k of the $\mathbf{X}_i$'s.

Some applications of local depth will make use of all β-values, while others will be based on a single β-value or on a small collection of them. In the latter case, the choice of β crucially depends on the application at hand, and there is of course no hope to identify a universal rule to select the appropriate β-value(s). Instead, it is desirable, in every specific application, to define data-driven β-selection procedures, at least whenever the results strongly depend on β. This will be illustrated in Section 6 below, where we present two applications of the proposed local depth concept.

Theorem 3.1 below provides consistency of sample local depth under absolute continuity assumptions. Of course, we need assuming consistency for the original global depth $D(\,\cdot\,, P)$: for any absolutely continuous P and any $\boldsymbol{x} \in \mathbb{R}^d$, $|D(\boldsymbol{x}, P^{(n)}) - D(\boldsymbol{x}, P)| \overset{a.s.}{\to} 0$ as $n \to \infty$. Actually, we will need the following reinforcement.

(Q1) *weak continuity*: for any absolutely continuous P, any sequence of probability measures (P_n) that converges weakly to P as $n \to \infty$, and any $\boldsymbol{x} \in \mathbb{R}^d$, we have that $|D(\boldsymbol{x}, P_n) - D(\boldsymbol{x}, P)| \to 0$ as $n \to \infty$.

This reinforcement is needed to cope with the complex dependence of the sample local depth $LD^{\beta}(\boldsymbol{x}, P^{(n)}) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\beta,(n)})$ on $P^{(n)}$. Note indeed that the dependence of $P_{\boldsymbol{x}}^{\beta,(n)}[\,\cdot\,] = P^{(n)}[\,\cdot\,|R^{\beta}(P_{\boldsymbol{x}}^{(n)})]$ on empirical measures is twofold.

Theorem 3.1 (Consistency)**.** Fix $\boldsymbol{x} \in \mathbb{R}^d$ and let $D(\,\cdot\,, P)$ satisfy Property (P2), (P3), and (Q1). Then, for any absolutely continuous P and any sequence $\beta_n \to \beta$, we have that $LD^{\beta_n}(\boldsymbol{x}, P^{(n)}) \overset{a.s.}{\to} LD^{\beta}(\boldsymbol{x}, P)$ as $n \to \infty$.

Property (Q1) actually holds for many depths. In particular, Proposition 1 of Mizera & Volauf (2002) and Theorem 2.2 (ii) of Zuo (2003) establish (Q1) for the halfspace and projection depths, respectively. For simplicial depth, Dümbgen (1992) proved the stronger property $\sup_{\boldsymbol{x} \in \mathbb{R}^d} |D_S(\boldsymbol{x}, P_n) - D_S(\boldsymbol{x}, P)| \to 0$ as $P_n \to P$ weakly.

4 Extreme localization

As we pointed out in the Introduction, all available extensions of depth that aim to deal with non-convexly supported distributions converge, as locality becomes extreme, to either a density measure or to a constant value, hence lose their nature of a centrality measure. We now show that the proposed local depths improve on this.

4.1 Assumptions and extreme local regions

For the sake of convenience, we are listing here the assumptions—all on the original depth D—we will need in this section.

(Q1$^+$) *uniform weak continuity*: for any two sequences (P_n) and (P'_n) of absolutely continuous distributions for which $|P_n[B] - P'_n[B]| \to 0$ as $n \to \infty$ for any Borel set B, $|D(\boldsymbol{x}, P_n) - D(\boldsymbol{x}, P'_n)| \to 0$ as $n \to \infty$ for any $\boldsymbol{x} \in \mathbb{R}^d$;

(Q2) *unique maximization at the symmetry center*: if P is absolutely continuous (with density f, say) and is centrally symmetric about $\boldsymbol{\theta}$ in the closure $\mathrm{Supp}(f)$ of $\mathrm{Supp}_+(f) = \{\boldsymbol{x} \in \mathbb{R}^d \,|\, f(\boldsymbol{x}) > 0\}$, then $D(\boldsymbol{\theta}, P) > D(\boldsymbol{x}, P)$ for all $\boldsymbol{x}$;

(Q3) *P-independent depth at the symmetry center*: if P is absolutely continuous and centrally symmetric about $\boldsymbol{\theta}$, then $c_D = D(\boldsymbol{\theta}, P)$ (that, under (P2), is equal to $\max_{\boldsymbol{x} \in \mathbb{R}^d} D(\boldsymbol{x}, P)$) is independent of P (which justifies the notation c_D).

We defined local depth above through $LD^\beta(\boldsymbol{x}, P) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^\beta)$, where $P_{\boldsymbol{x}}^\beta$ is obtained from P by conditioning it on $R_{\boldsymbol{x}}^\beta(P)$. Provided that the depth D satisfies Property (Q1), it then seems natural to expect that

$$\lim_{\beta \to 0} LD^\beta(\boldsymbol{x}, P) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^0), \tag{4.1}$$

where $P_{\boldsymbol{x}}^0$ denotes the possible weak limit of $P_{\boldsymbol{x}}^\beta$. Unfortunately, the situation is not so simple, as we show in Section 4.2 below. We start with a result concerning the support $R_{\boldsymbol{x}}^0(P) := \bigcap_{\beta > 0} R_{\boldsymbol{x}}^\beta(P)$ of $P_{\boldsymbol{x}}^0$.

Lemma 4.1. *Let $D(\,\cdot\,, P)$ satisfy (P2), (P3), (Q1), and (Q2). Fix an absolutely continuous P (with density f, say). Then, (i) for any $\boldsymbol{x} \in \mathrm{Supp}(f)$, for all $\varepsilon > 0$, there exists $\beta > 0$ such that $R_{\boldsymbol{x}}^\beta(P) \subset B_{\boldsymbol{x}}(\varepsilon) := \{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{y} - \boldsymbol{x}\| \leq \varepsilon\}$, so that $R_{\boldsymbol{x}}^0(P) = \{\boldsymbol{x}\}$; (ii) if one further assumes that (Q2) also holds for symmetry centers $\boldsymbol{\theta} \notin \mathrm{Supp}(f)$, then, for any $\boldsymbol{x} \notin \mathrm{Supp}(f)$, $\boldsymbol{x}$ belongs to the interior of $R_{\boldsymbol{x}}^0(P)$.*

This result shows that the support $R_{\boldsymbol{x}}^0(P)$ of the limiting distribution $P_{\boldsymbol{x}}^0$, hence also $P_{\boldsymbol{x}}^0$ itself, is of a different nature depending on whether $\boldsymbol{x}$ belongs to $\mathrm{Supp}(f)$ or not. This motivates treating these two cases separately, in the next two subsections.

4.2 Extreme behavior in the support of the distribution

We start with the case $\boldsymbol{x} \in \mathrm{Supp}(f)$, for which $R_{\boldsymbol{x}}^0(P) = \{\boldsymbol{x}\}$ (Lemma 4.1(i)). For such $\boldsymbol{x}$, the result expected in (4.1) does not hold because $P_{\boldsymbol{x}}^0$ does not exist. To prove this, let us reach a contradiction by assuming that it does exist. Note first that, from Lemma 4.1(i), it is clear that any open halfspace that does not contain $\boldsymbol{x}$ needs to have $P_{\boldsymbol{x}}^0$-probability zero, which implies that an open halfspace H having $\boldsymbol{x}$ on its boundary should also receive $P_{\boldsymbol{x}}^0$-probability zero. However, a direct computation—along the same lines as in the proof of Theorem 4.1 below—rather provides that

$$P_{\boldsymbol{x}}^0[H] = \lim_{\beta \to 0} P_{\boldsymbol{x}}^\beta[H] = \lim_{\beta \to 0} \frac{\mathrm{Vol}(R_{\boldsymbol{x}}^\beta(P) \cap H)}{\mathrm{Vol}(R_{\boldsymbol{x}}^\beta(P))} = \frac{1}{2},$$

a contradiction. The non-existence of the weak limit $P_{\boldsymbol{x}}^0$ explains why we have to reinforce (Q1) into (Q1$^+$), under which we can show the following result (see the Appendix for the Proof).

Theorem 4.1. *Let $D(\,\cdot\,, P)$ satisfy (P2), (P3), (Q1$^+$), (Q2), and (Q3). Fix an absolutely continuous P (with density f, say). Let $\boldsymbol{x} \in \mathrm{Supp}_+(f)$ be a continuity point of f. Then $LD^\beta(\boldsymbol{x}, P) \to c_D$ as $\beta \to 0$, where c_D is the constant in (Q3).*

This confirms that, unlike most of its competitors, the proposed local depth concept is not of a density nature under extreme localization; irrespective of the density at $\boldsymbol{x} \in \mathrm{Supp}_+(f)$, the

limiting local depth at $\boldsymbol{x}$ takes a constant value c_D, which supports the intuition that, for extreme locality, points inside the support of the distribution get most central (clearly, under (Q3), $\sup_{\boldsymbol{x} \in \mathbb{R}^d} LD^\beta(\boldsymbol{x}, P) \leq c_D$ for any β).

On the contrary, a point $\boldsymbol{x}$ at the boundary of the support may assume, as $\beta \to 0$, any limiting local depth value between the minimal possible value 0 and the maximal possible value c_D. To illustrate this, we consider the following bivariate example. For any $\eta \in (0, \pi)$, let $P = P_\eta$ be the uniform distribution on the unit disk $B_{\boldsymbol{0}}(1)$ (centered at the origin $\boldsymbol{0} = \binom{0}{0}$) deprived from a sector with radius $1/2$ and angle η, that is, more precisely, the uniform distribution on the set

$$C_\eta := B_{\boldsymbol{0}}(1) \setminus \left\{ \boldsymbol{x} = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix} \in B_{\boldsymbol{0}}(1/2) : \varphi \in [-\pi, \pi] \setminus [-\eta/2, \eta/2] \right\};$$

see Figure 3. For any $\eta \in (0, \pi)$, the origin lies on the boundary of the support of P_η, and one can show that $\ell_\eta := \lim_{\beta \to 0} LD^\beta(\boldsymbol{0}, P_\eta)$ ranges from $\ell_0 = \lim_{\eta \overset{>}{\to} 0} \ell_\eta = c_D$ to $\ell_\pi = \lim_{\eta \overset{<}{\to} \pi} \ell_\eta = 0$. This confirms that points on the boundary of the support, for extreme localization, may receive arbitrarily small or arbitrarily large local depths. This is far from being undesirable, though, and, in this example, is perfectly translating the obvious fact that (extreme) local centrality of the origin is a decreasing function of η. Note that the global depths $D(\boldsymbol{0}, P_\eta)$, $\eta \in (0, \pi)$, remain bounded away from zero.
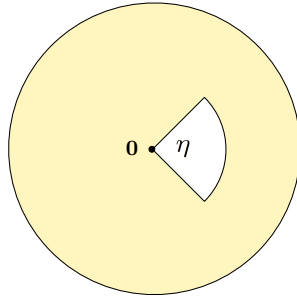

Figure 3: *Support C_η of the uniform distribution considered above.*

Before turning to points $\boldsymbol{x}$ outside the support, we focus on the univariate halfspace and simplicial depths, for which more precise results can be derived. For $x \in \mathrm{Supp}_+(f)$, we have the following result (see the Appendix for a proof).

Theorem 4.2. Fix $x \in \mathrm{Supp}_+(f)$. Then, (i) provided that f admits a continuous derivative f' in a neighborhood of x, we have that, as $\beta \to 0$, $LD_H^\beta(x, P) = \frac{1}{2} - \frac{|f'(x)|}{8f^2(x)}\beta + o(\beta)$; (ii) provided that f admits a continuous second derivative f'' in a neighborhood of x, we have that, as $\beta \to 0$, $LD_S^\beta(x, P) = \frac{1}{2} - \frac{(f'(x))^2}{16f^4(x)}\beta^2 + o(\beta^2)$.

This result shows that, for small values of β, the behavior of the local depth is not characterized by $f(x)$, but rather by $|f'(x)|/f^2(x)$. The latter quantity is a measure of *local asymmetry* at x, which further indicates that, as desired, our local depth provides a centrality measure for x, and not a density measure at x. From Theorem 4.2, LD_S^β is seen to converge to $1/2 (= c_{D_S} = c_{D_H}$ for $d = 1)$ faster than LD_H^β does. As a consequence, one may expect having to consider larger β-values for simplicial depth than for halfspace depth to find out about the above local asymmetry features; this may actually be seen in Figure 3 below.

For points on the boundary of the support, too, the picture is clearer for the univariate halfspace and simplicial depths than in the general multivariate case. Indeed, it can easily be shown that, in

the univariate case, $\lim_{\beta \to 0} LD_H^\beta(x, P) = 0 = \lim_{\beta \to 0} LD_S^\beta(x, P)$ as soon as, for some $\varepsilon > 0$, f is continuous in $(x - \varepsilon, x + \varepsilon) \setminus \{x\}$.

4.3 Extreme behavior outside the support of the distribution

Finally, we turn to the case $\boldsymbol{x} \notin \mathrm{Supp}(f)$. When it does exist, the weak limit $P_{\boldsymbol{x}}^0 = \lim_{\beta \to 0} P_{\boldsymbol{x}}^\beta$ then coincides with the probability measure obtained by conditioning P on $R_{\boldsymbol{x}}^0$ (which, according to Lemma 4.1(ii), is a neighborhood of $\boldsymbol{x}$). Since the interior of $R_{\boldsymbol{x}}^0$ has zero P-probability, the support of the conditional distribution $P_{\boldsymbol{x}}^0$ is contained in the boundary $\partial R_{\boldsymbol{x}}^0$ of $R_{\boldsymbol{x}}^0$, so that $P_{\boldsymbol{x}}^0$ may *not* be absolutely continuous.

Quite fortunately, for the halfspace and simplicial depths, Property (Q1) extends to P's that are not absolutely continuous; see Remark 2.5 in Zuo (2003). For these depths, we may therefore conclude that $\lim_{\beta \to 0} LD^\beta(\boldsymbol{x}, P) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^0)$ as in (4.1). For most $\boldsymbol{x} \notin \mathrm{Supp}(f)$, the support of $P_{\boldsymbol{x}}^0$ will be contained in an open halfspace having $\boldsymbol{x}$ on its boundary hyperplane, in which case $\lim_{\beta \to 0} LD^\beta(\boldsymbol{x}, P) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^0) = 0$ for both halfspace and simplicial depths. It is only in some very specific points $\boldsymbol{x} \notin \mathrm{Supp}(f)$, that typically are symmetry centers of the corresponding limiting region $R_{\boldsymbol{x}}^0$, that $\lim_{\beta \to 0} LD^\beta(\boldsymbol{x}, P) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^0)$ will be non-zero. Quite interestingly, the resulting value needs not be the maximal value c_D, but is obtained from $P_{\boldsymbol{x}}^0$ in a natural way.

We illustrate this in the univariate case $d = 1$, where we can again go further than in the multivariate case. If $d = 1$, the limiting region R_x^0 is always an interval of the form $[x - h_x^0, x + h_x^0]$. From the general discussion above, we know that the support of the limiting distribution P_x^0 is included in $\partial R_x^0 = \{x - h_x^0, x + h_x^0\}$. Denoting by p_x^- and p_x^+ the respective probabilities P_x^0 assigns to $x - h_x^0$ and $x + h_x^0$, we obtain that

$$\lim_{\beta \to 0} LD_H^\beta(x, P) = D_H(x, P_x^0) = \min(p_x^-, p_x^+), \tag{4.2}$$

$$\lim_{\beta \to 0} LD_S^\beta(x, P) = D_S(x, P_x^0) = 2p_x^- p_x^+. \tag{4.3}$$

The probabilities (p_x^-, p_x^+) can be computed from the identities

$$p_x^- + p_x^+ = 1 \quad \text{and} \quad \frac{p_x^+}{p_x^-} = \lim_{\varepsilon \overset{>}{\to} 0} \frac{P[X \in (x + h_x^0, x + h_x^0 + \varepsilon)]}{P[X \in (x - h_x^0 - \varepsilon, x - h_x^0)]} \quad (\in [0, \infty]).$$

An explicit example is provided in Section 5.

5 Examples

We first consider two univariate examples. We restrict our attention to local halfspace and simplicial depths, that admit the following explicit expressions in the univariate case (these expressions readily follow from the well-known formulae $D_H(x, P) = \min(F(x), 1 - F(x))$ and $D_S(x, P) = 2F(x)(1 - F(x))$, where $F(x) = P[(-\infty, x)]$ is the cumulative distribution function associated with P).

Proposition 5.1. Let $x_\beta := x - \inf\{h > 0 : F(x + h) - F(x - h) \geq \beta\}$, where F denotes the cumulative distribution function associated with the absolutely continuous distribution P. Then the local halfspace and simplicial depths of x with respect to P are given by $LD_H^\beta(x, P) = \frac{1}{\beta} \min\left[F(x) - F(x_\beta), F(2x - x_\beta) - F(x)\right]$ and $LD_S^\beta(x, P) = \frac{2}{\beta^2}\left(F(x) - F(x_\beta)\right)\left(F(2x - x_\beta) - F(x)\right)$, respectively.

Both univariate examples involve mixture probability measures $P = P^X$; more precisely, we considered the Gaussian and uniform mixtures, obtained with $X \sim \frac{1}{2}\mathcal{N}(\mu_a = -2, 2) + \frac{1}{2}\mathcal{N}(\mu_b = 2, 1)$ and $X \sim \frac{1}{2}\mathrm{Unif}(-5, -1) + \frac{1}{2}\mathrm{Unif}(1, 3)$, respectively. Figures 3 and 5 report the plots of the corresponding β-local halfspace and simplicial depth functions for various β ranging from $\beta = 1$ (global depth) to a small β-value (extreme locality), along with the plot of the density f of X.

We start by commenting results for the Gaussian mixture. As expected, global depth functions are unimodal, while local depth functions allow for local maxima. In line with the univariate example from Section 2, small β-values give raise to three local maxima : two located about the modes μ_a and μ_b, and a third one (also for simplicial depth, although it is less visible than for halfspace depth) at $\mu \in (\mu_a, \mu_b)$, say. The large local centrality measure μ gets for small enough β is associated with the fact that, in the corresponding β-neighborhoods, the mixture distribution is almost symmetric about μ ; however, the large volume of $R_\mu^\beta(P)$, compared to $R_{\mu_a}^\beta(P)$ and $R_{\mu_b}^\beta(P)$, allows to discriminate between both types of local maxima. Finally, the plot associated with $\beta = 0.01$ illustrates Theorem 4.1 (pointwise convergence of local depth functions to the constant function $1/2$).
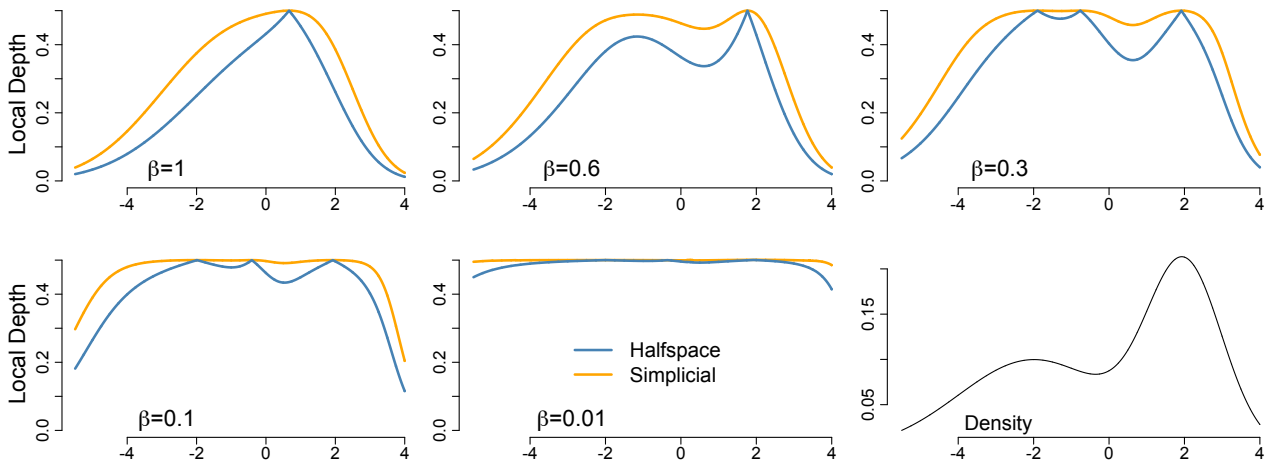

Figure 4: *Plots of several β-local halfspace (blue) and simplicial (orange) depth functions for a mixture of Gaussian distributions ($X \sim \frac{1}{2}\mathcal{N}(-2, 2) + \frac{1}{2}\mathcal{N}(2, 1)$), along with a plot of the corresponding density.*

Regarding the uniform mixture, comments similar to those made for the Gaussian mixture can be repeated, and we therefore rather focus on what is specific to this second example. In line with the general univariate results from Section 4.2, $\lim_{\beta \to 0} LD_i^\beta(x, P) = 1/2$ $(i = H, S)$ for all $x \in \mathrm{Supp}_+(f)$ (Theorem 4.1), and $\lim_{\beta \to 0} LD_i^\beta(x, P) = 0$ $(i = H, S)$ for $x \in \mathrm{Supp}(f) \setminus \mathrm{Supp}_+(f) = \{-5, -1, 1, 3\}$. For points $x \in \mathbb{R} \setminus \mathrm{Supp}(f) = (-\infty, -5) \cup (-1, 1) \cup (3, \infty)$, it is easy to check that $(p_x^-, p_x^+) = (0, 1)$ if $x \in (-\infty, -5) \cup (0, 1)$, $(p_x^-, p_x^+) = (1, 0)$ if $x \in (-1, 0) \cup (3, \infty)$ and $(p_x^-, p_x^+) = (1/3, 2/3)$ if $x = 0$, which according to (4.2)-(4.3), results into $\lim_{\beta \to 0} LD_i^\beta(x, P) = 0$ $(i = H, S)$ for all non-zero such values of x, and into $\lim_{\beta \to 0} LD_H^\beta(x, P) = D_H(x, P_x^0) = 1/3$ and $\lim_{\beta \to 0} LD_S^\beta(x, P) = D_S(x, P_x^0) = 4/9$ for $x = 0$. This thoroughly explains the plot corresponding to $\beta = 10^{-4}$ in Figure 5.

We now turn to two multivariate (simulated) examples : for the first example, that involves a bimodal distribution, we generated $n = 1,000$ independent observations of the form $\mathbf{X}_i = \sqrt{0.3}\, h(\mathbf{Z}_i)\mathbf{Z}_i + T_i \boldsymbol{\mu}_a + (1 - T_i)\boldsymbol{\mu}_b$, where $\boldsymbol{\mu}_a = \binom{0}{0}$, $\boldsymbol{\mu}_b = \binom{2}{0}$, the $\mathbf{Z}_i$'s are i.i.d. standard bivariate normal, $h(\mathbf{z})$ is the indicator that the Euclidean norm of $\mathbf{z}$ is smaller than 0.6, and the T_i's are
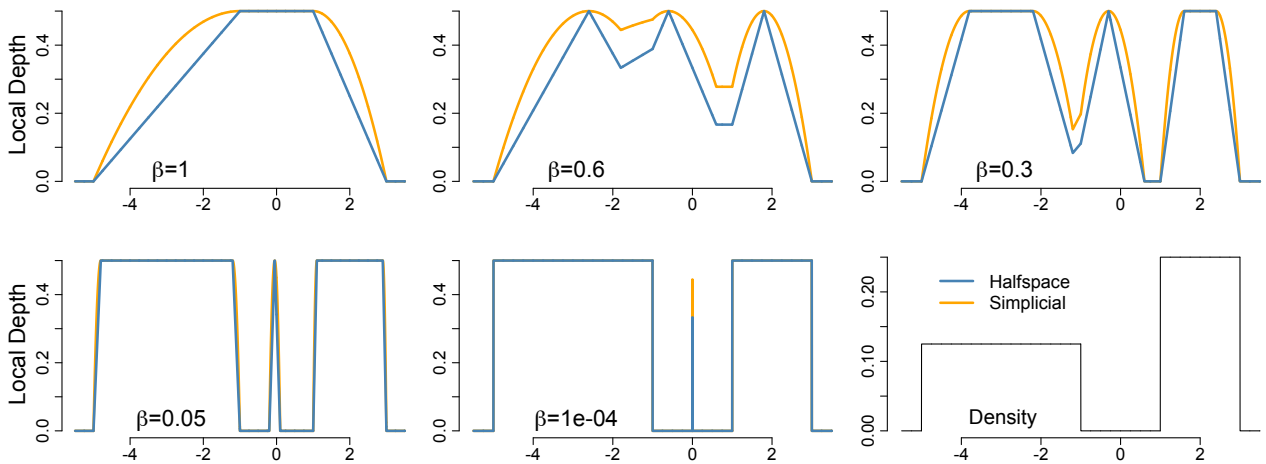
Figure 5: *Plots of several β-local halfspace (blue) and simplicial (orange) depth functions for a mixture of uniform distributions ($X \sim \frac{1}{2}\text{Unif}(-5, -1) + \frac{1}{2}\text{Unif}(1, 3)$), along with a plot of the corresponding density.*

i.i.d $\text{Bin}(0, 1/2)$, independent from the $\mathbf{Z}_i$'s. The second example, that focuses on a non-convexly supported distribution, is based on $n = 500$ independent observations of the form $\binom{X_i}{Y_i}$, where $X_i \sim \text{Unif}(-1, 1)$ and $Y_i|[X_i = x] \sim \text{Unif}(1.5(1 - x^2), 2(1 - x^2))$. Figures 6 and 7 show heatplots of the corresponding local halfspace depth functions at locality levels $\beta = 1$ (global halfspace depth), 0.7, 0.5, 0.3, 0.2, and 0.1, along with observations in the upper left panels.

In Figure 6, one can see that, as β moves away from one, the multimodal nature of the distribution is revealed (a task in which global halfspace depth clearly fails). At any β, a third local maximum is present around $\boldsymbol{\mu} = (\boldsymbol{\mu}_a + \boldsymbol{\mu}_b)/2$, resulting from the symmetry of the distribution about $\boldsymbol{\mu}$ at any locality level β (i.e., $P_{\boldsymbol{\mu}}^{\beta}$ is centrally symmetric about $\boldsymbol{\mu}$ for any β). This is in line with the fact that our local depth is a centrality measure, and not a density measure. As in the univariate case, discriminating between the true modes around $\boldsymbol{\mu}_a$, $\boldsymbol{\mu}_b$, and this "artificial" mode in $\boldsymbol{\mu}$ may be based on a comparison of the volumes of the neighborhoods $R_{\boldsymbol{x}}^{\beta}(P^{(n)})$, for $\boldsymbol{x} = \boldsymbol{\mu}_a, \boldsymbol{\mu}_b, \boldsymbol{\mu}$. Incidentally, we point out that, in some applications (including, in particular, classification; see Section 6.1), such artificial modes, due to the zero (or small) probability mass there, will have no (or low) impact in practice.

Parallel to the Boston example in Section 2, Figure 7 illustrates that global depth cannot deal with non-convexly supported distributions, since in particular the global deepest point is very close the boundary on the support. As β decreases, it is seen that local depth much better reflects centrality in the present setup. Small β-values clearly illustrate Theorem 4.1, since local depth is then almost constant in the support. We point out that this would hold irrespective of the (non-vanishing) density over the same support. In sharp contrast, the local depths from Agostinelli & Romanazzi (2011), for $\beta \to 0$, would, after appropriate normalization, converge to the density, which, for many densities, would clearly fail to reflect centrality.

Figure 6: *Heatplots of local halfspace depth functions at locality levels β = 1 (global halfspace depth), 0.7, 0.5, 0.3, 0.2, and 0.1, for n = 1,000 independent observations from the bivariate mixture distribution described in Section 5.*
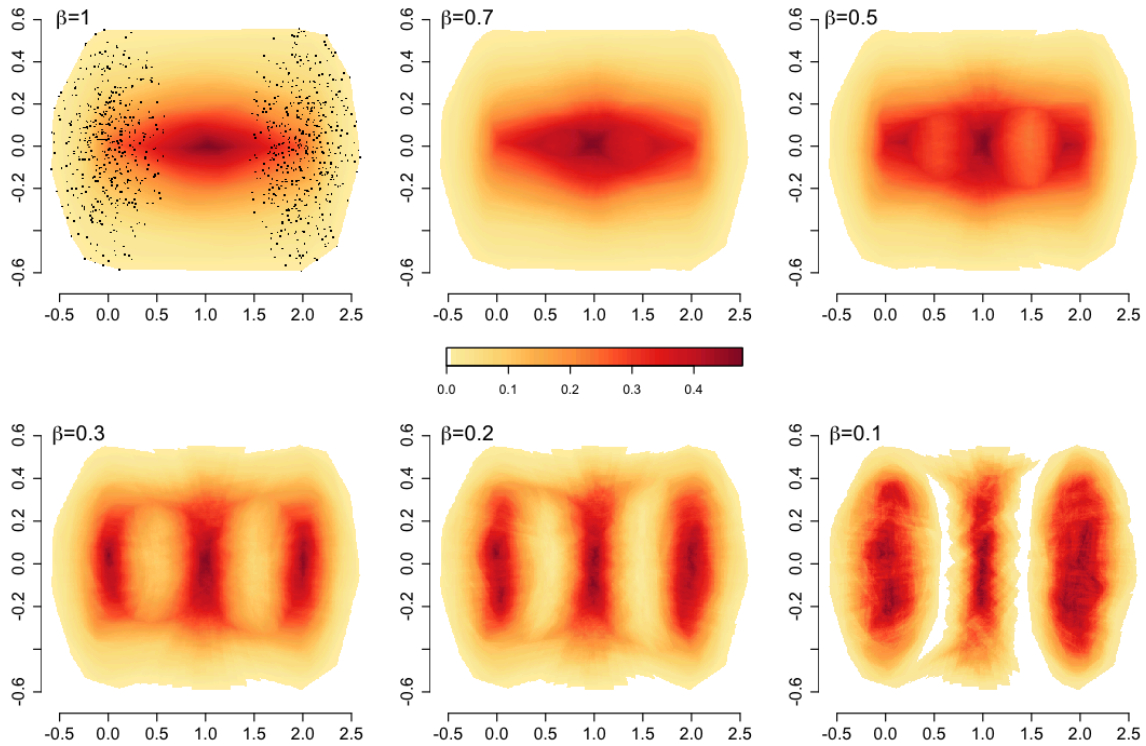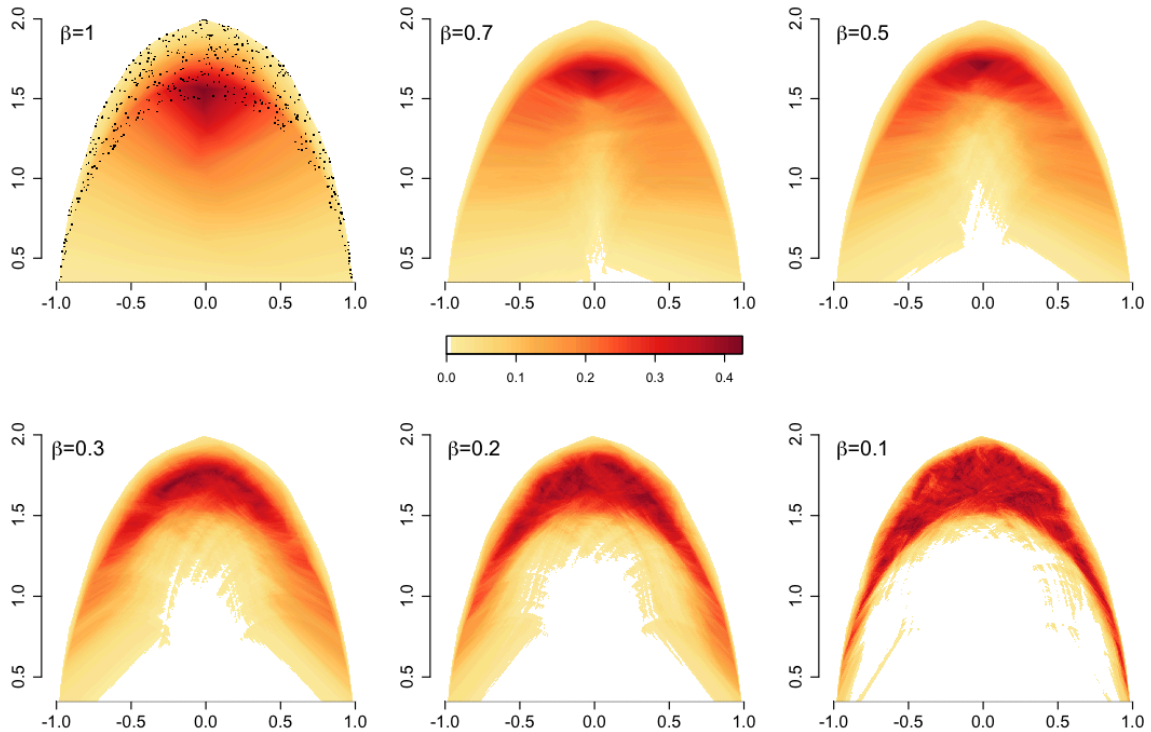

Figure 7: *Heatplots of local halfspace depth functions at locality levels β = 1 (global halfspace depth), 0.7, 0.5, 0.3, 0.2, and 0.1, for n = 500 independent observations from the distribution with a non-convex ("moon-shaped") support described in Section 5.*

6 Inferential applications

In this section, we describe two applications of the proposed local depth concept. The first one is related to classification, while the second one deals with symmetry testing.

6.1 Max-depth classification

Consider the classical two-population problem in which a random d-vector is to be classified as arising from any of two probability measures P_0 or P_1, on the basis of the value $\boldsymbol{x}$ it assumes. This is to be achieved on the basis of a "training sample", made of two mutually independent random samples $(\mathbf{X}_{01}, \ldots, \mathbf{X}_{0n_0})$ and $(\mathbf{X}_{11}, \ldots, \mathbf{X}_{1n_1})$ from P_0 and P_1, respectively. Depth-based classifiers typically match $\boldsymbol{x}$ to the population with respect to which it is most central : denoting by $P_j^{(n)}$, $j = 0, 1$, the empirical distribution associated with $(\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jn_j})$, $\boldsymbol{x}$ is classified into Population 0 (resp., Population 1) if $D(\boldsymbol{x}, P_0^{(n)}) > D(\boldsymbol{x}, P_1^{(n)})$ (resp., $D(\boldsymbol{x}, P_0^{(n)}) < D(\boldsymbol{x}, P_1^{(n)})$), while ties are decided at random.

This *max-depth* approach was first proposed in Liu *et al.* (1999), and was then investigated in Ghosh & Chaudhuri (2005). In the same vein, Li *et al.* (2012) recently proposed the "Depth vs Depth" (DD) classifiers that improve on the max-depth ones by constructing appropriate polynomial separating curves in the DD-plot, that is, in the scatter plot of $(D(\mathbf{X}_i, P_0^{(n)}), D(\mathbf{X}_i, P_1^{(n)}))$, $i = 1, \ldots, n$ (the original max-depth classifiers simply use the main bisector in the DD-plot as a separating curve).

As we showed in Section 2, global depth may fail to properly measure centrality for non-convexly supported distributions. Consequently, max-depth classifiers may perform poorly when P_0 and/or P_1 have a non-convex support (which is confirmed in our simulations below). Since the proposed local depths can deal with such non-convexity, one may think of defining max-*local*-depth classifiers obtained by substituting, in max-depth classifiers, β-local depth (for some β) for (global) depth. In practice, β may be chosen through cross-validation, that is, by minimizing in $\beta \in (0, 1]$, the resulting empirical misclassification rate evaluated on the training sample.

We conducted the following simulation exercise both to show that max-depth classifiers may indeed behave poorly under non-convexly supported distributions and to investigate the performances of the proposed max-local-depth classifiers. Three bivariate distributional setups were investigated :

Setup 1 (multinormality): P_j, $j = 0, 1$, is bivariate normal with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, with $\boldsymbol{\mu}_0 = \binom{0}{0}$, $\boldsymbol{\mu}_1 = \binom{2}{2}$, $\boldsymbol{\Sigma}_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and $\boldsymbol{\Sigma}_1 = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$;

Setup 2 (moon- and ball-supported distributions): P_0 is the distribution of $\binom{X}{Y}$, where $X \sim \text{Unif}(-1, 1)$ and $Y|[X = x] \sim \text{Unif}(1.5(1 - x^2), 2(1 - x^2))$, whereas P_1 is the uniform distribution on the ball with center $\binom{0}{1.3}$ and radius 0.7;

Setup 3 (ring- and rectangle-supported distributions): P_0 is the distribution of $R\mathbf{U}$, where $R \sim \text{Unif}(1,2)$ and $\mathbf{U} = \binom{\cos \Theta}{\sin \Theta}$, with $\Theta \sim \text{Unif}(0, 2\pi)$, are independent, while P_1 is the uniform distribution on the rectangle $(-1.5, 1.5) \times (-2.5, 2.5)$.

Exactly as in Li *et al.* (2012), we generated, for each setup, 100 training samples of size $n_0 = n_1 = 200$, and recorded, on corresponding test samples of size $n_{test} = 1,000$ (500 observations from each

population), the misclassification frequencies of the following classifiers (all depth-based classifiers below are based on halfspace depth) : (i) the Linear and Quadratic Discriminant Analysis classifiers (LDA/QDA); (ii) the standard kNN classifier, where k is chosen through cross-validation (kNN); (iii) the max-depth classifier from Ghosh & Chaudhuri (2005) (max-D); (iv) its (exact) linear and quadratic exact DD-refinements from Li *et al.*, 2012 (DD1 and DD2). These classifiers actually search for the separating linear (resp., quadratic) curve $\{(d, f(d)) : d \in (0,1)\}$ passing through the origin and one (resp., two) DD-points $(D(\mathbf{X}_i, P_0^{(n)}), D(\mathbf{X}_i, P_1^{(n)}))$, $i = 1, \ldots, n$, that minimizes the misclassification rate

$$\sum_{i=1}^{n_0} \mathbb{I}[f(D(\mathbf{X}_{0i}, P_0^{(n)})) > D(\mathbf{X}_{0i}, P_1^{(n)})] + \sum_{i=1}^{n_1} \mathbb{I}[f(D(\mathbf{X}_{1i}, P_0^{(n)})) < D(\mathbf{X}_{1i}, P_1^{(n)})]$$
$$+ \frac{1}{2}\sum_{i=1}^{n_0} \mathbb{I}[f(D(\mathbf{X}_{0i}, P_0^{(n)})) = D(\mathbf{X}_{0i}, P_1^{(n)})] + \frac{1}{2}\sum_{i=1}^{n_1} \mathbb{I}[f(D(\mathbf{X}_{1i}, P_0^{(n)})) = D(\mathbf{X}_{1i}, P_1^{(n)})];$$

(iv) our cross-validated max-local-depth classifier (max-LD ($\beta = \beta_{\mathrm{CV}}$)); (v) various max-local-depth classifiers based on a fixed β, with $\beta = .8, .6, .4$ and $.2$ (max-LD).

Figure 8 shows boxplots of the resulting misclassification frequencies, and further reports, in each setup, the median of the 100 β-values selected through cross-validation.
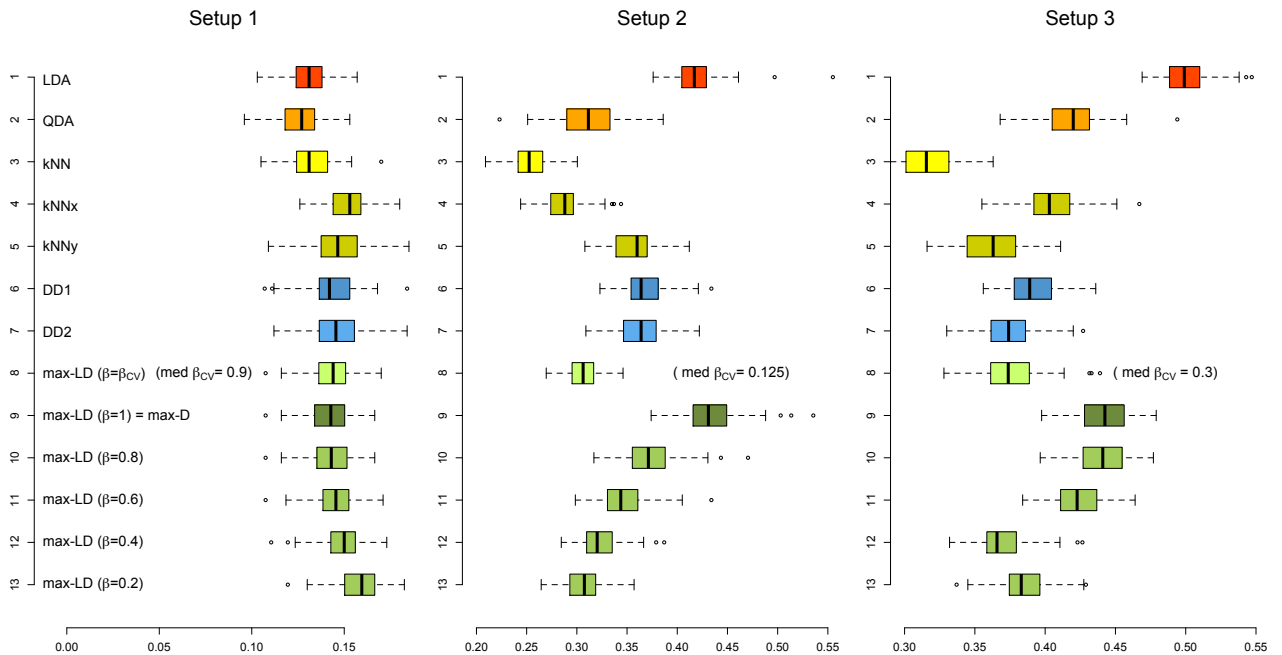

Figure 8: *Boxplots of missclassification frequencies from* 100 *replications, in Setups* 1 *to* 3 *described in Section 6.1, with training sample sizes $n_0 = n_1 = 200$ and test sample size $n_{\mathrm{test}} = 1,000$ (500 observations from each population), of the LDA/QDA classifiers, the exact linear (DD1) and quadratic (DD2) DD-classifiers, the proposed cross-validated max-local-depth classifiers (max-LD ($\beta = \beta_{\mathrm{CV}}$)), as well as some max-local-depth classifiers with fixed β, for $\beta = 1$ (max-depth classifier) and $\beta = 0.8, 0.6, 0.4, 0.2$.*

Our cross-validated max-local-depth classifier shows similar performances as its depth-based competitors under ellipticity (Setup 1), but clearly outperforms its competitors under non-convex populations (Setups 2 and 3), with the only exception of the classifier DD2 in Setup 3 with whom it

competes equally. The β-values selected through cross-validation nicely reflect the non-convexity of the underlying setup, hence the need to restrict to observations that are "close" to the point to be classified (β small) or the allowance to base classification on all observations (β close to 1). This is seen in the three setups, where the medians of the 100 selected β-values are, respectively, .9 (convex setup), .125 and .3 (non-convex setups).

Comparison with classical benchmarks is also of interest. As expected, our cross-validated max-local-depth classifier dominates LDA/QDA classifiers under non-convexity. On the contrary, the (universally consistent) kNN classifiers seem to dominate the proposed classifiers, hence also our depth-based competitors from Li *et al.* (2012) (which may seem unexpected in view of the Monte Carlo comparisons conducted there). Unlike depth-based classifiers, however, kNN classifiers fail to be affine-invariant, hence may show significantly poorer performances under unit changes. This is illustrated in our simulations where it is seen that, in all setups, misclassification rates of kNN classifiers suffer from multiplying one of both coordinates by a factor 10.

6.2   Testing for central symmetry

There are many graphical methods based on depth—or on the companion concept of multivariate quantiles—to assess departures from angular symmetry, central symmetry, or other types of multivariate symmetry; see Liu *et al.* (1999) and Serfling (2004). There are, however, few genuine tests of symmetry based on depth. To the best of our knowledge, the only such tests, available in any dimension d, are

• the test from Rousseeuw & Struyf (2002), that is a test for angular symmetry about a specified center $\boldsymbol{x}_0$ rejecting the null for large values of $T_{\boldsymbol{x}_0}^{(n)}$, with $T_{\boldsymbol{x}}^{(n)} = \frac{1}{2} - D_H(\boldsymbol{x}, P^{(n)})$. Quite remarkably, $T_{\boldsymbol{x}_0}^{(n)}$ is distribution-free under the null, which allows to approximate arbitrary well the exact fixed-n critical values through simulations;

• the test from Dutta *et al.* (2011), that may be seen as the companion test for the null of angular symmetry about an unspecified center, as it rejects this null for large values of $T^{(n)} = T_{\hat{\boldsymbol{\theta}}}^{(n)}$, where $\hat{\boldsymbol{\theta}}$ denotes the halfspace deepest point of $P^{(n)}$ (or, if unicity fails, the barycenter of the collection of deepest points). Critical values are obtained from bootstrap-type samples (as in Dutta *et al.* (2011), we will use the term "bootstrap", although the corresponding tests are rather of a permutation nature).

The motivation for both tests comes from the following characterization result : for an absolutely continuous P, $D_H(\boldsymbol{x}_0, P) \leq 1/2$, and equality holds iff P is angularly symmetric about $\boldsymbol{x}_0$; see Zuo (1998), Zuo & Serfling (2000b), Rousseeuw & Struyf (2004), and Dutta *et al.* (2011).

Since the null of central symmetry is at least as relevant for applications as the null of angular symmetry, it is unfortunate that there is no depth-based tests of central symmetry available in any dimension d. As we show now, the proposed local depth concept allows to define (universally consistent) tests of central symmetry. This relies on the following result, that characterizes central symmetry through local depth (see the Appendix for the proof).

Theorem 6.1 (Characterization of central symmetry through local depth)**.** Let P be an absolutely continuous distribution over $\mathbb{R}^d$. Then P is centrally symmetric about $\boldsymbol{x}_0(\in \mathbb{R}^d)$ if and only if $LD_H^\beta(\boldsymbol{x}_0, P) = 1/2$ for all $\beta \in (0, 1]$.

Testing central symmetry about $\boldsymbol{x}_0$ may then be based on the Cramèr-Von Mises (CV) or Kolmogorov-Smirnov (KS) statistics

$$CV^{(n)}_{\boldsymbol{x}_0;\beta_n} = \int_{\beta_n}^1 \left(LD_H^\beta(\boldsymbol{x}_0, P^{(n)}) - 1/2 \right)^2 d\beta \qquad (6.1)$$

$$KS^{(n)}_{\boldsymbol{x}_0;\beta_n} = \sup_{\beta \in [\beta_n, 1]} \left| LD_H^\beta(\boldsymbol{x}_0, P^{(n)}) - 1/2 \right|, \qquad (6.2)$$

where the sequence (β_n) is such that $\beta_n \to 0$ and $n\beta_n \to \infty$ (such a sequence typically allows to achieve universal consistency while discarding, at any given sample size n, the levels at which local depth can only be poorly estimated ; see Footnote 2 in Page 52). Critical values are obtained as in Dutta *et al.* (2011). More precisely, one first generates "bootstrap" samples of the form $\mathbf{X}^*_{(m)} = (\boldsymbol{x}_0 + s_{(m)1}(\mathbf{X}_1 - \boldsymbol{x}_0), \ldots, \boldsymbol{x}_0 + s_{(m)n}(\mathbf{X}_n - \boldsymbol{x}_0))$, $m = 1, \ldots, M$, where $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ denotes the original sample and the $s_{(m)i}$'s are mutually independent variables taking values ± 1 with equal probability $1/2$. The α-level critical value for $CV^{(n)}_{\boldsymbol{x}_0;\beta_n}$ is then simply the order-α quantile in the series $CV^{(n)}_{\boldsymbol{x}_0;\beta_n}(\mathbf{X}^*_{(m)})$, $m = 1, \ldots, M$ (discreteness may require randomization to achieve null size α). Critical values for $KS^{(n)}_{\boldsymbol{x}_0;\beta_n}$ are computed in the exact same way.

We conducted a simulation study in order to investigate the finite-sample behavior of these tests. For any of the following setups and any corresponding value of a, we generated $1,000$ independent random samples $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of size $n = 400$ from the same distribution as the generic random vector $\mathbf{X}$:

Setup 1: $\mathbf{X} = R\binom{\cos\Theta}{\sin\Theta}$, where $\Theta \sim \text{Unif}(0, 2\pi)$ and $R|[\Theta = \theta] \sim \text{Unif}(0, \theta^a)$, for $a = 0$ (central symmetry) and $a = .125, .250, .375, .500$ (angular symmetry);

Setup 2: $\mathbf{X} = R\binom{\cos\Theta}{\sin\Theta}$, where $R \sim \text{Unif}(0, 1)$ and $(\Theta/2\pi)^{1/(1+a)} \sim \text{Unif}(0, 1)$, for $a = 0$ (central symmetry) and $a = .15, .30, .45, .60$ (no angular symmetry);

Setup 3: $\mathbf{X} = R\binom{\cos\Theta}{\sin\Theta} + \binom{a}{a}$, where $R \sim \text{Unif}(0, 1)$ and $\Theta \sim \text{Unif}(0, 2\pi)$, for $a = 0$ (central symmetry) and $a = .125, .250, .375, .500$ (no angular symmetry).

Figure 9 plots the resulting rejection frequencies (at nominal level 5%) of the angular symmetry test based on $T^{(n)}_{\boldsymbol{x}_0}$ and of the central symmetry tests based on $CV^{(n)}_{\boldsymbol{x}_0;\beta_n}$ and $KS^{(n)}_{\boldsymbol{x}_0;\beta_n}$, for $\beta_n = .15, .16, \ldots, .30$; exact critical values were used for $T^{(n)}_{\boldsymbol{x}_0}$ (see Rousseeuw & Struyf, 2002), while critical values for $CV^{(n)}_{\boldsymbol{x}_0;\beta_n}$, and $KS^{(n)}_{\boldsymbol{x}_0;\beta_n}$ were obtained as described above from $M = 1,000$ bootstrap samples.

The results show that the bootstrap procedure indeed leads to central symmetry tests that have the correct size under the null. As expected, these tests succeed in detecting central asymmetry in all setups, while the angular symmetry test, of course, shows no power in Setup 1 (which confirms that it is inappropriate as a test for central symmetry). The angular symmetry test seems to dominate the central symmetry ones in Setup 2, and the opposite holds in Setup 3. Most importantly, the proposed Cramér Von Mises local-depth-based tests, that dominate their Kolmogorov-Smirnov counterparts, show empirical powers that barely depend on β_n; consequently, in contrast with classification in Section 6.1, it is not needed here to design a β-selection procedure (one just needs using a β_n-value that is small, but large enough to make it so that the actual sample size $(n\beta_n)$ used in the most extreme local depth involved (level β_n) does not fall below 50, say).
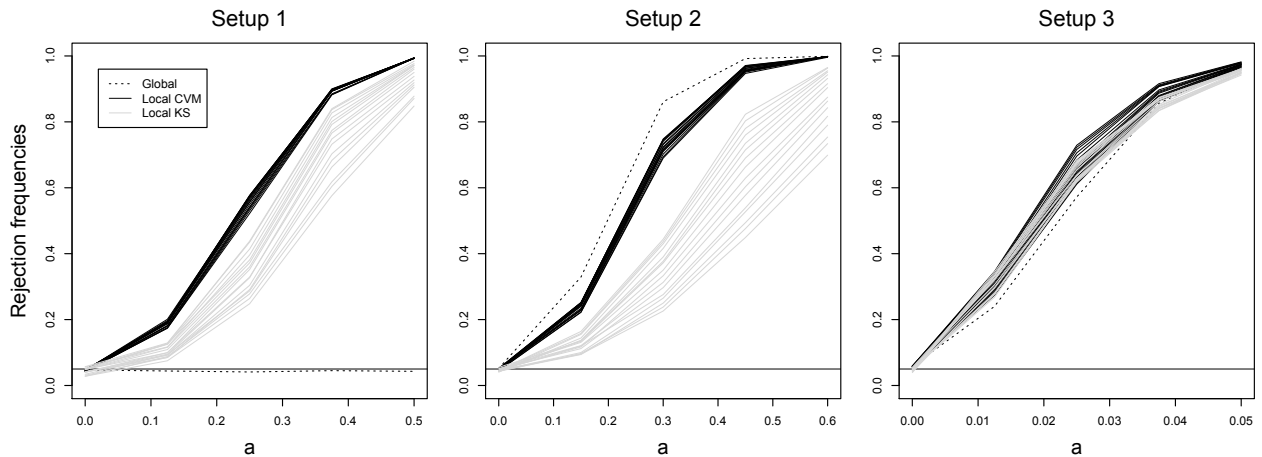
Figure 9: *Rejection frequencies, in each of the three setups described in Section 6.2, of the angular symmetry test from Rousseeuw & Struyf (2002), and of the proposed Cramèr-Von Mises and Kolmogorov-Smirnov central symmetry tests, for $\beta_n = .15, .16, \ldots, .30$; results are based on 1,000 replications and the sample size is $n = 400$.*

Of course, tests for central symmetry about an unspecified center may be obtained, as in Dutta *et al.* (2011), by rejecting the null for large values of $CV^{(n)}_{\hat{\boldsymbol{\theta}};\beta_n}$ and $KS^{(n)}_{\hat{\boldsymbol{\theta}};\beta_n}$.

7 Extension to other setups

We focused so far on *location* depths, which are the most well-known ones (and the first to have been introduced). In the last fifteen years, depth has however been extended to more general contexts, that cover regression models (Rousseeuw & Hubert, 1999), location-scale models (Mizera & Müller, 2004), or generic parametric models (Mizera, 2002). More recently, many extensions of depth to the functional data context were also proposed ; See Fraiman & Muniz (2001); López-Pintado & Romo (2009, 2011); Cuevas *et al.* (2007), etc.

Quite nicely, our construction of local depth extends naturally to these other depths. We now explain this in the empirical case, to which we restrict for the sake of exposure (a thorough investigation of the resulting local depths would go beyond the scope of the present paper and is therefore left for future research).

7.1 Tangent depth and regression depth

Let the random k-vector $\mathbf{Z}$ have a distribution in the parametric family $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d\}$ and assume that independent copies $\mathbf{Z}_i$, $i = 1, \ldots, n$, of $\mathbf{Z}$ are available. Let $F_i(\boldsymbol{\theta}) = F(\mathbf{Z}_i, \boldsymbol{\theta})$ be a measure of how well the parameter value $\boldsymbol{\theta}$ fits observation $\mathbf{Z}_i$. In such parametric context, the following depth function gives large depth to parameter values $\boldsymbol{\theta}$ that provide a good overall fit for the sample $\mathbf{Z}_i$, $i = 1, \ldots, n$.

Definition 7.1 (Mizera, 2002)**.** The *tangent depth* of $\boldsymbol{\theta}$ with respect to the empirical distribution $P^{(n)}$ of $\mathbf{Z}_i$, $i = 1, \ldots, n$ is $TD(\boldsymbol{\theta}, P^{(n)}) = D_H(\mathbf{0}, P^{(n)}_{\nabla}(\boldsymbol{\theta}))$, where $\mathbf{0} = (0, \ldots, 0)' \in \mathbb{R}^d$ and $P^{(n)}_{\nabla}(\boldsymbol{\theta})$ denotes the empirical distribution of $\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta})$, $i = 1, \ldots, n$.

An example is simple linear regression, where observations are of the form $\mathbf{Z}_i = \binom{X_i}{Y_i}$ and the parameter value $\boldsymbol{\theta} = \binom{\theta_1}{\theta_2}$ is associated with the regression line $y = \theta_1 x + \theta_2$. For $F_i(\boldsymbol{\theta}) = (Y_i - \theta_1 X_i - \theta_2)^2$, $i = 1, \ldots, n$, the tangent depth in Definition 3.2 reduces to the so-called *regression depth* from Rousseeuw & Hubert (1999).

Now, since tangent depth is defined through *location* (halfspace) depth, a local tangent depth concept may readily be obtained from our local *location* depth.

Definition 7.2. Using the same notation as in Definition 3.2, the *local tangent depth* of $\boldsymbol{\theta}$ with respect to $P^{(n)}$, at locality level $\beta(\in (0,1])$—or simply, β-*local tangent depth*—is $LTD^{\beta}(\boldsymbol{\theta}, P^{(n)}) = LD_H^{\beta}(\mathbf{0}, P_{\nabla}^{(n)}(\boldsymbol{\theta}))$.

We now present an example illustrating the resulting *local regression depth*. We generated $n = 500$ independent regression observations $\mathbf{Z}_i = \binom{X_i}{Y_i}$ from a balanced mixture of simple linear regression models, according to $Y = \theta_1 X + \theta_2 + \varepsilon$, where $X \sim \text{Unif}(0,5)$, $\varepsilon \sim \mathcal{N}(0,.1)$, and $\boldsymbol{\theta} = \binom{\theta_1}{\theta_2}$ uniformly distributed over $\{\boldsymbol{\theta}_a = \binom{.75}{0}, \boldsymbol{\theta}_b = \binom{-.25}{1}\}$, are mutually independent. Figure 10 shows the heatplots of the β-*local regression depth* from Definition 7.2, for $\beta = 1$ (classical regression depth), 0.8, 0.6, 0.4 and 0.2, along with a scatter plot of the bivariate data.
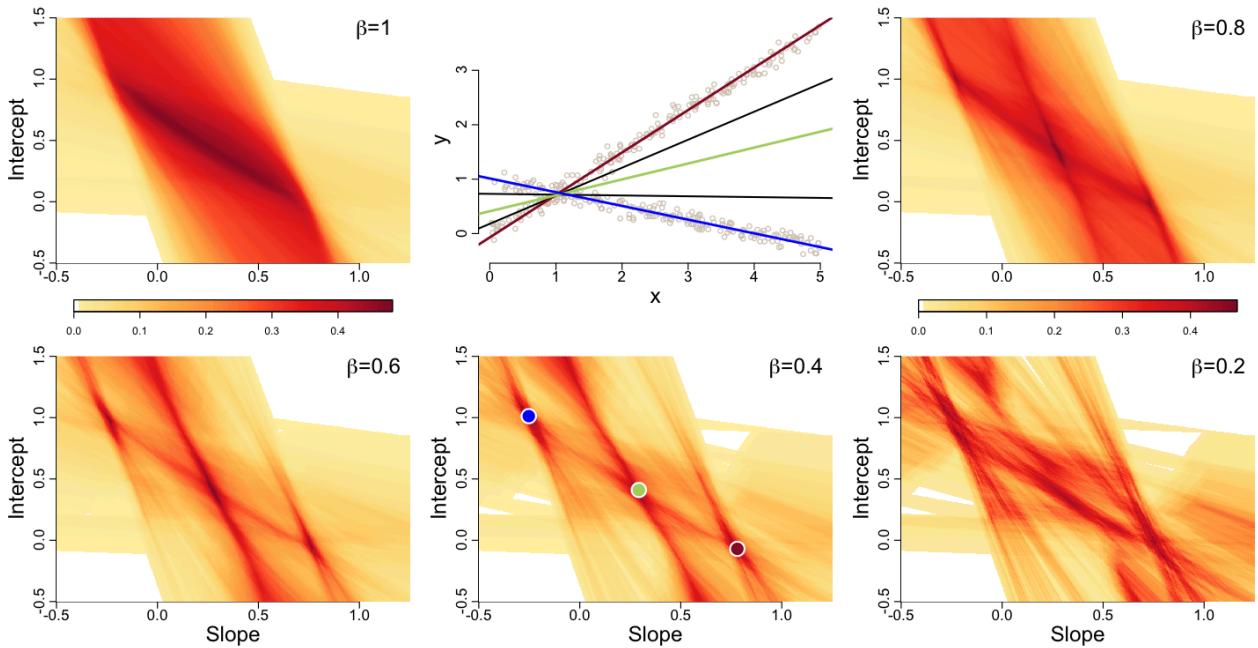

Figure 10: *(Upper center:) Scatter plot of the 500 data points generated from the mixture of linear regression models described in Section 7.1. Maxima of global regression depth (black lines) and local maxima of $\beta = 0.4$-local regression depth (brown, green, and blue lines) are pictured. (Others:) Heatplots of local regression depth functions at locality levels $\beta = 1$ (global regression depth), 0.8, 0.6, 0.4, and 0.2. Local maxima are highlighted in the plot for $\beta = 0.4$.*

All maximizers of global regression depth lie approximately on a segment in the slope-intercept space, which corresponds to a collection of regression lines passing through a fixed point $\binom{x}{y}$; we plotted in the observation space the regression lines associated with the maximizers with smallest and largest slopes (in solid lines). Clearly, this shows that, as in the location case, global regression depth misses the mixture or "bimodal" structure of the model. In contrast, β-local regression depths

clearly show local maxima about $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$, and, parallel to the location examples from the previous sections, also in a third intermediate parameter value $\boldsymbol{\theta}$, between $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$, that corresponds to a symmetry center. The regression lines associated with $\boldsymbol{\theta}_a$, $\boldsymbol{\theta}_b$, and $\boldsymbol{\theta}$ are plotted in the observation space; the corresponding parameter values are reported in the heatplot for $\beta = 0.4$.

7.2 Local functional depth

As stated above, many concepts of depth are available for functional data. Here, we describe a local version of one of the most successful functional depths, namely the *modified band depth* introduced in López-Pintado & Romo (2009). We start with a short description of this depth.

We consider functional observations $t \mapsto f_i(t)$, $i = 1, \ldots, n$, all defined on $[0, 1]$. The J-band depth of a given function $f : [0, 1] \to \mathbb{R}$ with respect to $\{f_1, \ldots, f_n\}$ is the proportion of J-tuples $(i_1, \ldots, i_J)$, $1 \le i_1 < \ldots < i_J \le n$ for which

$$\min_{j=1,\ldots,J} f_{i_j}(t) \le f(t) \le \max_{j=1,\ldots,J} f_{i_j}(t) \quad \forall t \in [0, 1].$$

Since one single extreme value of $f(t)$ is enough to give zero J-band depth to f, the following modified version was proposed.

Definition 7.3 (López-Pintado & Romo, 2009)**.** The *modified band depth* of f with respect to the empirical distribution $P^{(n)}$ of f_i, $i = 1, \ldots, n$ is

$$MBD(f, P^{(n)}) = \int_0^1 \binom{n}{J}^{-1} \sum_{1 \le i_1 < \ldots < i_J \le n} \mathbb{I}[\min_{j=1,\ldots,J} f_{i_j}(t) \le f(t) \le \max_{j=1,\ldots,J} f_{i_j}(t)] \, dt.$$

Our construction of local depth therefore suggests the following local extension.

Definition 7.4. Denote by $P_f^{(n)}$ the empirical measure associated with the $2n$ functions $f_1, \ldots f_n$, $2f - f_1, \ldots, 2f - f_n$. Then the *local modified band depth* of $\boldsymbol{\theta}$ with respect to $P^{(n)}$, at locality level $\beta (\in (0, 1])$—or simply, β-*local modified band depth*—is $LMBD^\beta(f, P^{(n)}) = MBD(f, P_f^{\beta,(n)})$, where $P_f^{\beta,(n)}$ denotes the empirical measure associated with the $\lceil n\beta \rceil$ functions that have largest $MBD(\,\cdot\,, P_f^{(n)})$ among $f_1, \ldots, f_n$.

We consider an illustration where we generated functions $f_1, \ldots, f_n$ by repeating independently $n = 400$ times the following procedure : (i) selecting randomly (with equal probability) $m(t) = \max(0, (t - 0.2))$ or $m(t) = \min(0, 0.2 - t)$; (ii) generating points of the form $(t_\ell, y_\ell) = (\frac{\ell}{100}, m(\frac{\ell}{100}) + \varepsilon_\ell)$, $\ell = 0, \ldots, 100$, where the ε_ℓ's are i.i.d. $\mathcal{N}(0, 0.2)$; (iii) performing a spline regression of order 3, with 10 basis functions, over these (t_ℓ, y_ℓ)'s.

The left panel of Figure 11 plots those f_i's, along with an artificial function f_{artif} that was randomly generated in the same fashion as the f_i's but from a trend $m(t) = \max(0, (0.2 - t)/2)$ and ε_ℓ's that are i.i.d. $\mathcal{N}(0, 0.05)$; the figure also emphasizes two particular observations, namely the (global) MBD-deepest observation $f_{\max}^1 = \arg\max_i MBD(f_i, P^{(n)})$ and the ($\beta = 0.5$)-local MBD-deepest observation $f_{\max}^{0.5} = \arg\max_i LMBD^{0.5}(f_i, P^{(n)})$ (the modified band depths of these functions—with $J = 2$, as suggested in López-Pintado & Romo (2009)—were estimated on the basis of 101 equispaced values of t). The middle panel of the figure plots the β-local modified band depth of f_{artif}, $f_{\max}^1$, and $f_{\max}^{0.5}$, as a function of the locality level β.

As expected, f_{artif} is (globally) deeper than $f_{\max}^1$ and $f_{\max}^{0.5}$, but its local depth decreases much with β. The depth of $f_{\max}^1$ exhibits a similar behavior, but remains quite deep at all locality levels β. Finally, the large β-local depth of $f_{\max}^{0.5}$ for small to moderate values of β translates its visible higher local centrality. Again, β-local depth for small β should considered with care, as it is typically computed from very few observations.

This last remark is illustrated by the dramatic increase, for β close to zero, of the β-local depths of $f_{\max}^1$ and $f_{\max}^{0.5}$, compared to that of f_{artif}. This corresponds to a bias arising from the fact that the former functions, unlike the latter, are part of the sample; the bias of β-local depth for observed functions is about $1/\lceil\beta n\rceil$, which is large for small β-values. In order to compare $f_{\max}^1$, $f_{\max}^{0.5}$, and f_{artif} on a common basis, we recomputed the β-local depths of these functions with respect to the sample of 398 functions obtained by removing $f_{\max}^1$ and $f_{\max}^{0.5}$ from the original sample. The right panel of Figure 11 shows that this indeed eliminates the above bias. Of course, such a bias potentially affects the other local depths introduced in this paper (the reason why the previous illustrations did not show any bias is that local depth was evaluated there at locations that do not bear observations).
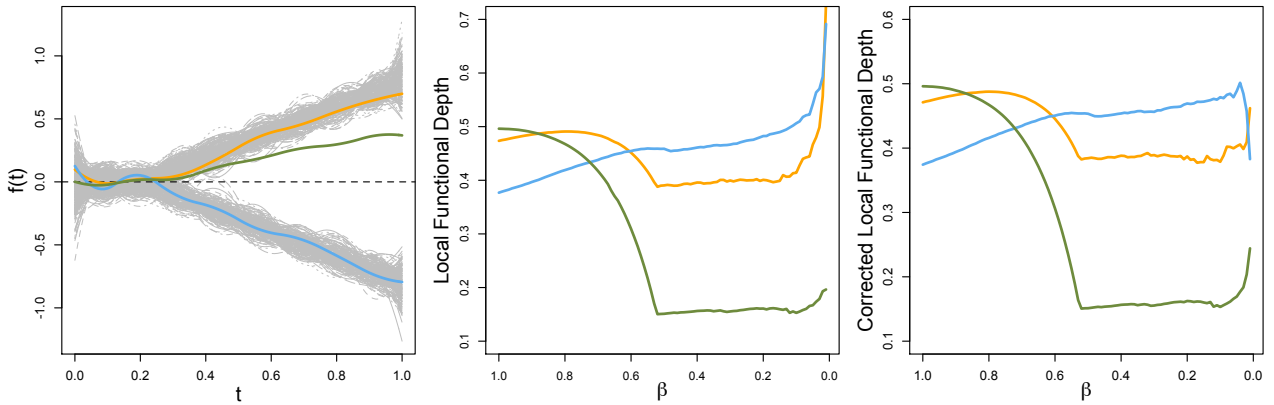

Figure 11: *(Left:) Plots of the $n = 400$ observed functions, with three particular functions highlighted : the (global) MBD-deepest observation $f_{\max}^1$ (orange curve), the $\beta = 0.5$-local MBD-deepest observation $f_{\max}^{0.5}$ (blue curve), and the artificial function f_{artif} (green curve); (Middle:) β-local modified band depths of $f_{\max}^1$, $f_{\max}^{0.5}$, and f_{artif}. (Right:) The corresponding bias-corrected β-local depths; see Section 7.2 for details.*

8 Computational aspects

In the (possibly multivariate) location case, the evaluation of $LD^\beta(\boldsymbol{x}, P^{(n)})$ at a fixed point $\boldsymbol{x} \in \mathbb{R}^d$ with respect to the empirical distribution $P^{(n)}$ associated with observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ proceeds along the following few simple steps :

1. Evaluate $D(\mathbf{X}_i, P_{\boldsymbol{x}}^{(n)})$, $i = 1, \ldots, n$, where $P_{\boldsymbol{x}}^{(n)}$ is the empirical distribution associated with the symmetrized observations $\mathbf{X}_1, \ldots, \mathbf{X}_n, 2\boldsymbol{x} - \mathbf{X}_1, \ldots, 2\boldsymbol{x} - \mathbf{X}_n$;

2. Rank the (original) observations according to the ordered depths $D(\mathbf{X}_{(1)}, P_{\boldsymbol{x}}^{(n)}) \geq D(\mathbf{X}_{(2)}, P_{\boldsymbol{x}}^{(n)})$ $\geq \ldots \geq D(\mathbf{X}_{(n)}, P_{\boldsymbol{x}}^{(n)})$ (this ranking is not unique in case of ties, but this will not affect the final value of local depth);

3. Determine $n_\beta(P_{\boldsymbol{x}}^{(n)}) = \max\left\{\ell = \lceil n\beta\rceil, \ldots, n : D(\mathbf{X}_{(\ell)}, P_{\boldsymbol{x}}^{(n)}) = D(\mathbf{X}_{(\lceil n\beta\rceil)}, P_{\boldsymbol{x}}^{(n)})\right\}$;

4. Compute $LD^\beta(\boldsymbol{x}, P^{(n)}) = D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\beta,(n)})$, where $P_{\boldsymbol{x}}^{\beta,(n)}$ is the empirical measure associated with $\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n_\beta(P_{\boldsymbol{x}}^{(n)}))}$.

The computation of the local tangent depth from Section 7.1 (of which local regression depth is a particular case) is obtained by substituting in these four steps, $\mathbf{0}(\in \mathbb{R}^d)$ for $\boldsymbol{x}$ and $\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta})$ for $\mathbf{X}_i$, $i = 1, \ldots, n$, and by restricting to halfspace depth $D = D_H$. Similarly, the computation of local functional depth of f with respect to $f_1, \ldots, f_n$ is obtained by substituting there f for $\boldsymbol{x}$, f_i for $\mathbf{X}_i$, $i = 1, \ldots, n$, and the modified band depth MBD for the depth D.

The procedure in Steps 1-4 above makes clear that the proposed sample local depths can be computed from global depth routines only (all illustrations in this paper were simply obtained from the R package *depth*). This is another advantage over the competing local depths, that do require developing specific routines or packages ; see, e.g., the R package *localdepth*, from Agostinelli & Romanazzi (2011).

Note that the evaluation of $LD^\beta(\boldsymbol{x}, P^{(n)})$ may be time consuming since it requires computing $(n+1)$ depth values (n depth values, in a sample of $2n$ data points, in Step 1, and one depth value, in a sample of $n_\beta(P_{\boldsymbol{x}}^{(n)})(\le n)$ data points, in Step 4). Quite fortunately, there has been much progress in the computation of depth in the recent years ; see in particular Hallin *et al.* (2010) for halfspace depth, and Liu & Zuo (2011a,b) and Liu *et al.* (2011) for projection depth.

Of course, computing "the whole local depth field" $\{LD^\beta(\boldsymbol{x}, P^{(n)}) : \boldsymbol{x} \in \mathbb{R}^d\}$ — in practice, computing local depth on a fine grid in a compact set — may still be very demanding. Generating the heat plots in Figures 6, 7, and 10 relied on a trivial method, where evaluation of $LD^\beta(\boldsymbol{x}, P^{(n)})$ started from scratch at any newly considered $\boldsymbol{x}$, which, indeed, may be slow for moderate to large sample sizes n. However, the value of $LD^\beta(\boldsymbol{x} + \boldsymbol{\Delta}, P^{(n)})$, with $\boldsymbol{\Delta}$ small, might be computed from the previous evaluation of $LD^\beta(\boldsymbol{x}, P^{(n)})$, by exploiting the fact that the distributions $P_{\boldsymbol{x}}^{(n)}$ and $P_{\boldsymbol{x}+\boldsymbol{\Delta}}^{(n)}$, hence also the empirical measures $P_{\boldsymbol{x}}^{\beta,(n)}$ and $P_{\boldsymbol{x}+\boldsymbol{\Delta}}^{\beta,(n)}$ (leading to the corresponding local depth values in Step 4 above), are close to each other. How to turn this into a practical algorithm allowing to compute efficiently the local depth field clearly remains a non-trivial question, that is beyond the scope of this methodological paper.

Now, most importantly, practical applications of local depth typically do not require evaluating the whole local depth field, but rather requires computing local depth at one or a reasonably small number of locations $\boldsymbol{x}$ only. This is the case for both applications considered in Section 6 : classification indeed requires evaluating local depth only at points to be classified (and at data points if β is selected through cross-validation), whereas symmetry testing only involves the local depth of the null symmetry center. Incidentally, we stress that, for symmetry testing, (i) the discrete nature of halfspace depth implies that (6.1)-(6.2) can be obtained from a finite number of β-values only; (ii) the bootstrap procedure there can be implemented in practice, since the M bootstrap samples, by symmetry, lead to the same results in Steps 1-3, that therefore need to be performed only once (only Step 4, in which a single depth value is computed, needs to be performed for each bootstrap sample).

Finally, we point out that computing local depth of a fixed point for ℓ distinct β-values typically requires much less time than computing ℓ times local depth for one fixed β-value. One can indeed take advantage of the fact that Step 1 above is common to the various computations of β-local depths (there is some analogy with quantile regression, where the information used to compute a

fixed regression quantile may be exploited when computing regression quantiles at other quantile levels). Also, if the computational effort is an important issue, one always may use a (global) depth concept that is not computationally intensive, such as, e.g., the Mahalanobis depth. If it is felt that this depth is too "parametric", it can then be used in Steps 1-3 only, while a more nonparametric depth (halfspace depth, simplicial depth, projection depth, etc.) is used in Step 4. This possibility to base local depth on two different global depths has not been considered in the paper, but leads to a local depth concept enjoying all nice properties of the one we introduced.

Acknowledgments

Appendix — Proofs

This appendix collects proofs of technical results. We start with the proof of Theorem 3.1, which requires the following preliminary result. Throughout this section, $R_{\boldsymbol{x}}^{\beta}$ will denote $R_{\boldsymbol{x}}^{\beta}(P)$, when no ambiguity is possible.

Lemma 1. *Let $D(\,\cdot\,,P)$ be a depth function satisfying Property (Q1). Then, for any $\boldsymbol{x} \in \mathbb{R}^d$, any Borel set $B \subset \mathbb{R}^d$, and any absolutely continuous distribution P, the mapping $\beta \mapsto P[R_{\boldsymbol{x}}^{\beta} \cap B]$ is continuous over $(0, 1]$.*

Proof of Lemma 1. Note first that Property (Q1) implies that, for any absolutely continuous P, $\boldsymbol{x} \mapsto D(\boldsymbol{x}, P)$ is a continuous function : indeed, if $\mathbf{X}$ is a random d-vector with distribution $P = P^{\mathbf{X}}$, then Property (P1) entails that, for any sequence $\boldsymbol{x}_n$ converging to $\boldsymbol{x}$, $|D(\boldsymbol{x}_n, P) - D(\boldsymbol{x}, P)| = |D(\boldsymbol{x}, P^{\mathbf{X}+(\boldsymbol{x}-\boldsymbol{x}_n)}) - D(\boldsymbol{x}, P)| \to 0$ as $n \to \infty$, since $P^{\mathbf{X}+(\boldsymbol{x}-\boldsymbol{x}_n)}$ converges weakly to P. Together with the fact that P is absolutely continuous, this implies that $P[R_{\boldsymbol{x}}^{\beta}(P)] = \beta$ for any $\beta \in (0, 1]$.

Now, fix $\beta_0 \in (0, 1]$ and a Borel set B. Consider a decreasing sequence (β_n) converging to β_0. The numbers $\gamma_n = P[R_{\boldsymbol{x}}^{\beta_n} \cap B]$ form a monotone decreasing sequence that is lower bounded by $\gamma_0 = P[R_{\boldsymbol{x}}^{\beta_0} \cap B]$. Hence they admit a limit $\lim_{n\to\infty} \gamma_n \geq \gamma_0$. Letting $\bar{\gamma}_n = P[R_{\boldsymbol{x}}^{\beta_n} \cap B^c]$, with $B^c = \mathbb{R}^d \setminus B$, we similarly obtain that $\lim_{n\to\infty} \bar{\gamma}_n \geq \bar{\gamma}_0 = P[R_{\boldsymbol{x}}^{\beta_0} \cap B^c]$. If $\lim_{n\to\infty} \gamma_n > \gamma_0$, then we have $\lim_{n\to\infty} \beta_n = \lim_{n\to\infty}(\gamma_n + \bar{\gamma}_n) > \gamma_0 + \bar{\gamma}_0 = \beta_0$, a contradiction. Hence, we must have that $\lim_{n\to\infty} \gamma_n = \gamma_0$, i.e., that $\beta \mapsto P[R_{\boldsymbol{x}}^{\beta} \cap B]$ is right continuous at β_0. The result then follows since left continuity can be established along the same lines. $\square$

Proof of Theorem 3.1. In view of (Q1), it is sufficient, in order to show that

$$\left| LD^{\beta_n}(\boldsymbol{x}, P^{(n)}) - LD^{\beta}(\boldsymbol{x}, P) \right| = \left| D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\beta_n,(n)}) - D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\beta}) \right| \overset{a.s.}{\to} 0 \quad \text{as } n \to \infty,$$

to prove that $P_{\boldsymbol{x}}^{\beta_n,(n)}[B] \overset{a.s.}{\to} P_{\boldsymbol{x}}^{\beta}[B]$ for any Borel set B.

Fix then such a B and $\varepsilon > 0$. Lemma 1 implies that there exist $\delta, \eta > 0$ such that

$$[P[R_{\boldsymbol{x}}^{\beta-\delta} \cap B] - \eta, P[R_{\boldsymbol{x}}^{\beta+\delta} \cap B] + \eta] \subset [P[R_{\boldsymbol{x}}^{\beta} \cap B] - \beta\varepsilon, P[R_{\boldsymbol{x}}^{\beta} \cap B] + \beta\varepsilon]. \tag{1}$$

Now, Theorem 3 in Zuo & Serfling (2000c) implies that there exists n_0 such that $R_{\boldsymbol{x}}^{\beta-\delta} \subset R_{\boldsymbol{x}}^{\beta_n,(n)} \subset R_{\boldsymbol{x}}^{\beta+\delta}$ a.s. for all $n \geq n_0$ (throughout the proof, $R_{\boldsymbol{x}}^{\beta_n,(n)}$ stands for $R^{\beta_n}(P_{\boldsymbol{x}}^{(n)})$), which of course yields that, a.s. for all $n \geq n_0$,

$$P^{(n)}[R_{\boldsymbol{x}}^{\beta-\delta} \cap B] \leq P^{(n)}[R_{\boldsymbol{x}}^{\beta_n,(n)} \cap B] \leq P^{(n)}[R_{\boldsymbol{x}}^{\beta+\delta} \cap B], \tag{2}$$

The SLLN entails that $P^{(n)}[R_{\boldsymbol{x}}^{\beta\pm\delta} \cap B] \overset{a.s.}{\to} P[R_{\boldsymbol{x}}^{\beta\pm\delta} \cap B]$ as $n \to \infty$; consequently, there exists n_1 such that, a.s. for all $n \geq n_1$,

$$\left[P^{(n)}[R_{\boldsymbol{x}}^{\beta-\delta} \cap B], P^{(n)}[R_{\boldsymbol{x}}^{\beta+\delta} \cap B]\right] \subset \left[P[R_{\boldsymbol{x}}^{\beta-\delta} \cap B] - \eta, P[R_{\boldsymbol{x}}^{\beta+\delta} \cap B] + \eta\right]. \tag{3}$$

Combining (1)-(3), we proved that, a.s. for all $n \geq \max(n_0, n_1)$,

$$P[R_{\boldsymbol{x}}^{\beta} \cap B] - \beta\varepsilon \leq P^{(n)}[R_{\boldsymbol{x}}^{\beta_n,(n)} \cap B] \leq P[R_{\boldsymbol{x}}^{\beta} \cap B] + \beta\varepsilon,$$

or equivalently, $P_{\boldsymbol{x}}^{\beta}[B] - \varepsilon \leq \frac{1}{\beta} P^{(n)}[R_{\boldsymbol{x}}^{\beta_n,(n)} \cap B] \leq P_{\boldsymbol{x}}^{\beta}[B] + \varepsilon$. In other words, we have proved that, as $n \to \infty$,

$$\frac{1}{\beta} P^{(n)}[R_{\boldsymbol{x}}^{\beta_n,(n)} \cap B] \overset{a.s.}{\to} P_{\boldsymbol{x}}^{\beta}[B]. \tag{4}$$

Taking $B = \mathbb{R}^d$ in (4) yields $P^{(n)}[R_{\boldsymbol{x}}^{\beta_n,(n)}] \overset{a.s.}{\to} \beta$, which, jointly with (4), establishes that $P_{\boldsymbol{x}}^{\beta_n,(n)}[B] = P^{(n)}[B|R_{\boldsymbol{x}}^{\beta_n,(n)}] \overset{a.s.}{\to} P_{\boldsymbol{x}}^{\beta}[B]$, as was to be proved. $\qquad\square$

Proof of Lemma 4.1. (i) Fix $\boldsymbol{x} \in \mathrm{Supp}(f)$ and $\varepsilon > 0$. By Lemma A.1 in Paindaveine & Van Bever (2012) (whose proof, under the properties (Q1)-(Q2) introduced in the present paper, trivially extends to the case where the symmetry center $\boldsymbol{\theta}$ belongs to $\mathrm{Supp}(f) \setminus \mathrm{Supp}_+(f)$), there exist $\delta > 0$ and $\alpha < \alpha_{\boldsymbol{x}}^* := \max_{\boldsymbol{y} \in \mathbb{R}^d} D(\boldsymbol{y}, P_{\boldsymbol{x}})$ such that $B_{\boldsymbol{x}}(\delta) \subset R_{\boldsymbol{x},\alpha} \subset B_{\boldsymbol{x}}(\varepsilon)$. Since $\boldsymbol{x} \in \mathrm{Supp}(f)$, we then have that $\beta_0 := P[R_{\boldsymbol{x},\alpha}] \geq P[B_{\boldsymbol{x}}(\delta)] > 0$. From the definition of $R_{\boldsymbol{x}}^{\beta_0}$, it follows that $R_{\boldsymbol{x}}^{\beta_0} \subset R_{\boldsymbol{x},\alpha} \subset B_{\boldsymbol{x}}(\varepsilon)$.

(ii) Fix $\boldsymbol{x} \notin \mathrm{Supp}(f)$ and let $\varepsilon > 0$ be such that $P[B_{\boldsymbol{x}}(\varepsilon)] = 0$. If one assumes that (Q2) also holds for $\boldsymbol{\theta} \notin \mathrm{Supp}(f)$, then it is easy to check that the proof of Lemma A.1(i) in Paindaveine & Van Bever (2012) further extends to the case where the symmetry center does not belong to $\mathrm{Supp}(f)$. Therefore there still exist $\delta > 0$ and $\alpha < \alpha_{\boldsymbol{x}}^*$ such that $B_{\boldsymbol{x}}(\delta) \subset R_{\boldsymbol{x},\alpha} \subset B_{\boldsymbol{x}}(\varepsilon)$. The definition of $R_{\boldsymbol{x}}^{\beta}$ implies that $R_{\boldsymbol{x},\alpha} \subset R_{\boldsymbol{x}}^{\beta}$ for any $\beta > 0$. It follows that $\boldsymbol{x} \in B_{\boldsymbol{x}}(\delta) \subset R_{\boldsymbol{x},\alpha} \subset R_{\boldsymbol{x}}^0 = \cap_{\beta>0} R_{\boldsymbol{x}}^{\beta}$, hence that $\boldsymbol{x}$ is an interior point of $R_{\boldsymbol{x}}^0$. $\qquad\square$

Lemma 2. *Under the assumptions of Theorem 4.1, $\beta/\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta}) \to f(\boldsymbol{x})$ as $\beta \to 0$.*

Proof of Lemma 2. Fix $\varepsilon > 0$ and let $r = r(\varepsilon)$ be such that $f(\boldsymbol{x}) - \varepsilon \leq f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \varepsilon$ for any $\boldsymbol{y} \in B_{\boldsymbol{x}}(r) = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z} - \boldsymbol{x}\| < r\}$. Lemma A.1 in Paindaveine & Van Bever (2012) ensures that there exists $\beta_0 > 0$ such that $R_{\boldsymbol{x}}^{\beta_0} \subset B_{\boldsymbol{x}}(r)$. Therefore, for any $\beta \in (0, \beta_0)$, one has $(f(\boldsymbol{x}) - \varepsilon)\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta}) \leq \beta = \int_{R_{\boldsymbol{x}}^{\beta}} f(\boldsymbol{y})d\boldsymbol{y} \leq (f(\boldsymbol{x}) + \varepsilon)\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta})$, or equivalently $f(\boldsymbol{x}) - \varepsilon \leq \beta/\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta}) \leq f(\boldsymbol{x}) + \varepsilon$. The result follows. $\qquad\square$

Proof of Theorem 4.1. Fix $\boldsymbol{x} \in \mathbb{R}^d$ such that f is positive and continuous at $\boldsymbol{x}$. For any β, let $B \mapsto P_{\boldsymbol{x}}^{\mathrm{sym},\beta}[B] = P_{\boldsymbol{x}}[B|R_{\boldsymbol{x}}^{\beta}]$ be the symmetrized (about $\boldsymbol{x}$) version of P, conditional to $R_{\boldsymbol{x}}^{\beta}$—recall that we let $P_{\boldsymbol{x}} = \frac{1}{2}P^{\mathbf{X}} + \frac{1}{2}P^{2\boldsymbol{x}-\mathbf{X}}$. We have

$$|LD^{\beta}(\boldsymbol{x}, P) - c_D| = |D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\beta}) - c_D| = |D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\beta}) - D(\boldsymbol{x}, P_{\boldsymbol{x}}^{\mathrm{sym},\beta})|,$$

where we used the fact that $D(\,\cdot\,, P)$ satisfies (Q3). In view of (Q1$^+$), it is therefore sufficient to prove that, for any Borel set B, $P_{\boldsymbol{x}}^{\beta}[B] - P_{\boldsymbol{x}}^{\mathrm{sym},\beta}[B] \to 0$ as $\beta \to 0$. To do so, fix such a B and, denoting by $f_{\boldsymbol{x}}^{\mathrm{sym}}$ the density of $P_{\boldsymbol{x}}^{\mathrm{sym}}$, write

$$P_{\boldsymbol{x}}^{\beta}[B] - P_{\boldsymbol{x}}^{\mathrm{sym},\beta}[B] = \frac{1}{\beta}\int_{R_{\boldsymbol{x}}^{\beta} \cap B}(f(\boldsymbol{y}) - f_{\boldsymbol{x}}(\boldsymbol{y}))\,d\boldsymbol{y} = \frac{1}{2\beta}\int_{R_{\boldsymbol{x}}^{\beta} \cap B}(f(\boldsymbol{y}) - f(2\boldsymbol{x} - \boldsymbol{y}))\,d\boldsymbol{y}.$$

If $\boldsymbol{x}$ lies in the interior of B, Lemma 4.1(i) shows that there exists $\beta_0 > 0$ such that, for all $\beta \leq \beta_0$, we have $R_{\boldsymbol{x}}^{\beta} \cap B = R_{\boldsymbol{x}}^{\beta}$. Clearly, this implies that for all $\beta \leq \beta_0$, the integral above, hence also $P_{\boldsymbol{x}}^{\beta}[B] - P_{\boldsymbol{x}}^{\mathrm{sym},\beta}[B]$, is equal to zero. If $\boldsymbol{x}$ does not belong to the closure of B, then the same lemma implies that $R_{\boldsymbol{x}}^{\beta} \cap B$ is empty for β small enough, which leads to the same conclusion. It remains to consider the case where $\boldsymbol{x}$ belongs to the boundary of B. For such an $\boldsymbol{x}$, we may write

$$P_{\boldsymbol{x}}^{\beta}[B] - P_{\boldsymbol{x}}^{\mathrm{sym},\beta}[B] = \frac{\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta} \cap B)}{2\beta}(\mathcal{I}_{\beta} - \mathcal{I}_{\beta}^{\mathrm{refl}}), \quad \text{where}$$

$$\mathcal{I}_{\beta} = \frac{1}{\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta} \cap B)}\int_{R_{\boldsymbol{x}}^{\beta} \cap B}f(\boldsymbol{y})\,d\boldsymbol{y} \quad \text{and} \quad \mathcal{I}_{\beta}^{\mathrm{refl}} = \frac{1}{\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta} \cap B^{\mathrm{refl}})}\int_{R_{\boldsymbol{x}}^{\beta} \cap B^{\mathrm{refl}}}f(\boldsymbol{y})\,d\boldsymbol{y},$$

and where $B^{\mathrm{refl}} = 2\boldsymbol{x} - B$ denotes the reflection of B about $\boldsymbol{x}$. The same reasoning as in the proof of Lemma 2 allows to show that both $\mathcal{I}_{\beta}$ and $\mathcal{I}_{\beta}^{\mathrm{refl}}$ converge to $f(\boldsymbol{x})$ as $\beta \to 0$. The result then follows from the fact that $\mathrm{Vol}(R_{\boldsymbol{x}}^{\beta} \cap B)/\beta \leq \mathrm{Vol}(R_{\boldsymbol{x}}^{\beta})/\beta$ remains bounded as $\beta \to 0$ (Lemma 2). $\square$

The next lemma is needed to prove Theorem 4.2. Recall that X is an absolutely continuous distribution with cdf F and pdf f, and put $g(\beta) := (F(x) - F(x_{\beta}))/\beta$ and $h(\beta) := (F(2x - x_{\beta}) - F(x))/\beta$, where x_{β} was defined in the statement of Proposition 5.1.

Lemma 3. *Fix $x \in \mathrm{Supp}_+(f)$. (i) If f is continuous in a neighborhood of x, then $\lim_{\beta \to 0} g(\beta) = \frac{1}{2} = \lim_{\beta \to 0} h(\beta)$; (ii) if f admits a continuous derivative f' in a neighborhood of x, then $\lim_{\beta \to 0} g'(\beta) = -\frac{f'(x)}{8f^2(x)}$ and $\lim_{\beta \to 0} h'(\beta) = \frac{f'(x)}{8f^2(x)}$; if f admits a continuous second derivative f'' in a neighborhood of x, then (iii) $\lim_{\beta \to 0} h''(\beta) = 0$ and $\lim_{\beta \to 0} g''(\beta) = 0$.*

Proof of Lemma 3. First note that x_{β} is the $(1-\beta)/2$-quantile of the symmetrized distribution of X about x, that is, $x_{\beta} = (F^Y)^{-1}(\frac{1-\beta}{2})$, where F^Y is the cdf $\frac{1}{2}F + \frac{1}{2}F^{2x-X}$. Below, the corresponding pdf will be denoted f^Y.

(i) Absolute continuity implies that $\lim_{\beta \to 0} x_{\beta} = x$. Therefore, L'Hôpital's rule can be applied and, together with the continuity of f in a neighborhood of x and the expression for x_{β} given above, we obtain

$$\lim_{\beta \to 0} g(\beta) = \lim_{\beta \to 0}\left(-f(x_{\beta})\frac{dx_{\beta}}{d\beta}\right) = \lim_{\beta \to 0}\frac{f(x_{\beta})}{2f^Y(x_{\beta})}.$$

Since $f^Y(x) = f(x)$, the result follows for g. Computations for $h(\beta)$ are extremely similar, hence will be omitted here (as well as in the proof of (ii)-(iii) below).

(ii) Straightforward calculus shows that $g'(\beta) = (\frac{f(x_\beta)}{2f^Y(x_\beta)} - g(\beta))/\beta$. Taking the limit of $g'(\beta)$ and applying L'Hôpital's rule gives

$$\lim_{\beta \to 0} g'(\beta) = \lim_{\beta \to 0} \frac{d}{d\beta}\left[\frac{f(x_\beta)}{2f^Y(x_\beta)}\right] - \lim_{\beta \to 0} g'(\beta).$$

The result then follows after some calculations using that $(f^Y)'$ is continuous in a neighborhood of x and takes value zero at x.

(iii) Tedious calculations show that

$$g''(\beta) = \frac{1}{\beta}\left\{\frac{-f'(x_\beta) + f(x_\beta)(f^Y)'(x_\beta)\left(f^Y(x_\beta)\right)^{-1}}{4\left(f^Y(x_\beta)\right)^2} - 2g'(\beta)\right\}.$$

L'Hôpital's rule then establishes the result, after some derivations using that $(f^Y)''$ is continuous in a neighborhood of x and takes value $f''(x)$ at x. □

Proof of Theorem 4.2. Under the assumptions considered, it is clearly sufficient to prove that, as $\beta \to 0$,

(a) $LD_H^\beta(x, P) \to \frac{1}{2}$ and $\frac{\partial}{\partial\beta} LD_H^\beta(x, P) \to -\frac{|f'(x)|}{8f^2(x)}$;

(b) $LD_S^\beta(x, P) \to \frac{1}{2}$, $\frac{\partial}{\partial\beta} LD_S^\beta(x, P) \to 0$, and $\frac{\partial^2}{\partial\beta^2} LD_S^\beta(x, P) \to -\frac{(f'(x))^2}{16f^4(x)}$.

Proposition 5.1 shows that $LD_H^\beta = \min(g(\beta), h(\beta))$. A simple Taylor expansion then yields

$$\lim_{\beta \to 0} LD_H^\beta(x, P) = \begin{cases} \lim_{\beta \to 0} g'(\beta) & \text{if } f'(x) > 0 \\ \lim_{\beta \to 0} h'(\beta) & \text{if } f'(x) < 0 \\ \lim_{\beta \to 0} \min(g'(\beta), h'(\beta)) & \text{if } f'(x) = 0 \end{cases}$$

Lemma 3 then directly establishes (a). In order to prove (b), note that Proposition 5.1 states that $LD_S^\beta(x, P) = 2g(\beta)h(\beta)$, which yields $\frac{\partial}{\partial\beta} LD_S^\beta(x, P) = 2g'(\beta)h(\beta) + 2g(\beta)h'(\beta)$ and $\frac{\partial^2}{\partial\beta^2} LD_S^\beta(x, P) = 2g''(\beta)h(\beta) + 4g'(\beta)h'(\beta) + 2g(\beta)h''(\beta)$. The limits in (b) then follow from Lemma 3. □

Proof of Theorem 6.1. (Necessity:) For any β, the region $R_{\boldsymbol{x}_0}^\beta(P)$ is centrally symmetric about $\boldsymbol{x}_0$: $R_{\boldsymbol{x}_0}^\beta(P) = 2\boldsymbol{x}_0 - R_{\boldsymbol{x}_0}^\beta(P)$. Hence the central symmetry of P about $\boldsymbol{x}_0$ implies that $P_{\boldsymbol{x}_0}^\beta[\cdot] = P[\cdot | R_{\boldsymbol{x}_0}^\beta(P)]$ is also centrally symmetric about $\boldsymbol{x}_0$. This implies that $LD_H^\beta(\boldsymbol{x}_0, P) = D_H(\boldsymbol{x}_0, P_{\boldsymbol{x}_0}^\beta) = 1/2$ for any β.

(Sufficiency:) For any β, $LD_H^\beta(\boldsymbol{x}_0, P) = D_H(\boldsymbol{x}_0, P_{\boldsymbol{x}_0}^\beta) = 1/2$ implies that $P_{\boldsymbol{x}_0}^\beta$ is angularly symmetric about $\boldsymbol{x}_0$. In other words, for any β, $P_{\boldsymbol{x}_0}^\beta[C] = P_{\boldsymbol{x}_0}^\beta[2\boldsymbol{x}_0 - C]$, for any C in the collection $\mathcal{C}_{\boldsymbol{x}_0}$ of cones originating from $\boldsymbol{x}_0$. This of course rewrites $P[C \cap R_{\boldsymbol{x}_0}^\beta(P)] = P[(2\boldsymbol{x}_0 - C) \cap R_{\boldsymbol{x}_0}^\beta(P)]$, $\forall C \in \mathcal{C}_{\boldsymbol{x}_0}$, $\forall \beta \in (0, 1]$. Since the regions $R_{\boldsymbol{x}_0}^\beta(P)$ are symmetric with respect to $\boldsymbol{x}_0$, this implies that

$$P[C \cap (R_{\boldsymbol{x}_0}^{\beta_2}(P) \setminus R_{\boldsymbol{x}_0}^{\beta_1}(P))]$$
$$= P[2\boldsymbol{x}_0 - (C \cap (R_{\boldsymbol{x}_0}^{\beta_2}(P) \setminus R_{\boldsymbol{x}_0}^{\beta_1}(P)))] \quad \forall C \in \mathcal{C}_{\boldsymbol{x}_0}, \ \forall \beta_1 < \beta_2 \in (0, 1].$$

This proves the result since the sigma-algebra generated by the subsets $C \cap (R_{\boldsymbol{x}_0}^{\beta_2} \setminus R_{\boldsymbol{x}_0}^{\beta_1})$, $C \in \mathcal{C}_{\boldsymbol{x}_0}$, $0 < \beta_1 < \beta_2 \leq 1$, coincides with the Borel sigma-algebra on $\mathbb{R}^d$. □

References

Agostinelli, C. & Romanazzi, M. (2011). Local depth. *J. Statist. Plann. Inference*, *141*(2), 817–830.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, *91*(434), 862–872.

Chen, Y., Dang, X., Peng, H., & Bart, H. L. J. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Trans. Patt. Anal Mach. Int.*, *31*(2), 288–305.

Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.*, *22*, 481–496.

Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statist. Probab. Lett.*, *14*(2), 119–128.

Dutta, S. & Ghosh, A. K. (2012). On robust classification using projection depth. *Ann. Inst. Statist. Math.*, *64*(3), 657–676.

Dutta, S., Ghosh, A. K., & Chaudhuri, P. (2011). Some intriguing properties of Tukey's halfspace depth. *Bernoulli*, *17*(4), 1420–1434.

Fraiman, R. & Muniz, G. (2001). Trimmed means for functional data. *Test*, *10*(2), 419–440.

Ghosh, A. K. & Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scand. J. Statist.*, *32*(2), 327–350.

Hallin, M., Paindaveine, D., & Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth (with discussion). *Ann. Statist.*, *38*, 635–669.

Härdle, W. (1991). *Smoothing techniques*. Springer Series in Statistics. New York: Springer-Verlag.

Harrison, D. & Rubinfeld, D. (1978). Hedonic prices and the demand for clean air. *J. Environ. Econ. Manage.*, *5*, 81–102.

He, X. & Wang, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.*, *25*(2), 495–504.

Hlubinka, D., Kotík, L., & Vencálek, O. (2010). Weighted halfspace depth. *Kybernetika*, *46*(1), 125–148.

Izem, R., Rafalin, E., & Souvaine, D. L. (2008). Describing multivariate distributions with nonlinear variation using data depth. Tech. rep., Technical Report TR-2006-2, Tufts University, Department of Computer Science.

Kong, L. & Zuo, Y. (2010). Smooth depth contours characterize the underlying distribution. *J. Multivariate Anal.*, *101*(9), 2222–2226.

Koshevoy, G. & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.*, *25*(5), 1998–2017.

Li, J., Cuesta-Albertos, J., & Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedures based on dd-plots. *J. Amer. Statist. Assoc.*, *107*(498), 737–753.

Li, J. & Liu, R. Y. (2004). New nonparametric tests of multivariate locations and scales using data depth. *Statist. Sci.*, *19*(4), 686–696.

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

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

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

Liu, X. & Zuo, Y. (2011a). Computing halfspace depth and regression depth. *Commun. Stat. Simulat.*.

Liu, X. & Zuo, Y. (2011b). Computing projection depth and its associated estimators. *Stat. Comput.*.

Liu, X., Zuo, Y., & Wang, Z. (2011). Exactly computing bivariate projection depth contours and median. *Comput. Statist. Data Anal.*.

Lok, W. & Lee, S. M. S. (2011). A new statistical depth function with applications to multimodal data. *J. Nonparametr. Stat.*, *23*(3), 617–631.

López-Pintado, S. & Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, *104*(486), 718–734.

López-Pintado, S. & Romo, J. (2011). A half-region depth for functional data. *Comput. Statist. Data Anal.*, *55*(4), 1679–1695.

McLachlan, G. & Peel, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.

McLachlan, G. J. & Basford, K. E. (1988). *Mixture models*, vol. 84 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker Inc.

Mizera, I. (2002). On depth and deep points: a calculus. *Ann. Statist.*, *30*(6), 1681–1736.

Mizera, I. & Müller, C. H. (2004). Location-scale depth (with discussion). *J. Amer. Statist. Assoc.*, *99*(468), 949–989.

Mizera, I. & Volauf, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *J. Multivariate Anal.*, *83*(2), 365–388.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.*, *1*(6), 327–332.

Paindaveine, D. & Van Bever, G. (2012). Nonparametrically consistent depth-based classifiers. Under revision.

Rousseeuw, P. J. & Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.*, *94*(446), 388–433.

Rousseeuw, P. J. & Struyf, A. (2002). A depth test for symmetry. In *Goodness-of-fit tests and model validity (Paris, 2000)*, Boston, MA: Birkhäuser Boston, Stat. Ind. Technol. 401–412.

Rousseeuw, P. J. & Struyf, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference*, *122*(1-2), 161–173.

Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *J. Statist. Plann. Inference*, *123*(2), 259–278.

Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *J. Nonparametr. Stat.*, *22*, 915–926.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*. Canad. Math. Congress, Montreal, Que., 523–531.

Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Amer. Statist. Assoc.*, *103*(397-409).

Zuo, Y. (1998). *Contributions to the theory and applications of statistical depth functions*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–The University of Texas at Dallas.

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

Zuo, Y. & Serfling, R. (2000a). General notions of statistical depth function. *Ann. Statist.*, *28*(2), 461–482.

Zuo, Y. & Serfling, R. (2000b). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Statist. Plann. Inference*, *84*(1-2), 55–79.

Zuo, Y. & Serfling, R. (2000c). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.*, *28*(2), 483–499.

Chapter III

Shape Depth

Shape Depth[1]

Davy PAINDAVEINE[a,b,*] and Germain VAN BEVER[a,c].

[a] *ECARES & Département de Mathématique, Université libre de Bruxelles (ULB)*
[b] *Email:* `dpaindav@ulb.ac.be`
[c] *Email:* `gvbever@ulb.ac.be`
[*] *Corresponding Author. Address: Université libre de Bruxelles, ECARES, Avenue F.D. Roosevelt, 50, CP 114/04, B-1050 Brussels, Belgium.*

Abstract In many problems from multivariate analysis (principal component analysis, testing for sphericity, etc.), the parameter of interest is the so-called *shape* matrix, that is a normalized version of the corresponding scatter or dispersion matrices. In this paper, we propose, under elliptical assumptions, a depth concept for shape. If shape matrices are normalized to have determinant one, our shape depth results from the parametric depth construction in Mizera (2002). For other normalizations, however, defining a proper shape depth requires a semiparametric extension of this construction, that is likely to have applications in other contexts. We show that the proposed shape depth is affine-invariant and does not depend on the normalization adopted. We also establish consistency, in the sense that shape depth is maximized at the true shape value. Finally, we consider depth-based tests for shape, and investigate their finite-sample performances through simulations.

Keywords: Elliptical distributions · Shape matrix · Statistical depth functions · Tangent depth · Tests for sphericity

---

[1]Manuscript is in preparation for publication.

1 Introduction

Elliptical distributions play a crucial role in many fields of statistics. They form a quite flexible extension of the multinormal model, on which most textbook statistical procedures in multivariate analysis are based. A random d-vector $\mathbf{X}$ is said to be elliptically distributed if its characteristic function is of the form

$$
\begin{aligned}
\mathbb{R}^d &\rightarrow \mathbb{C} \\
\boldsymbol{t} &\mapsto \exp(i\boldsymbol{t}'\boldsymbol{\mu})\psi(\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}),
\end{aligned}
\tag{1.1}
$$

where the d-vector $\boldsymbol{\mu}$ is a *location* parameter and the symmetric and positive definite $d \times d$ matrix $\boldsymbol{\Sigma}$ is a *scatter* parameter. The *characteristic generator* $\psi(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}$ then fully determines the distribution of the Mahalanobis distance $d_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{X}) = \left((\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right)^{1/2}$, hence in particular determines whether $\mathbf{X}$ has lighter-than-normal or heavier-than-normal tails. To make $\boldsymbol{\Sigma}$ and ψ identifiable without imposing any moment assumption, one may e.g. require that $d_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{X})$ has median one. Under absolute continuity assumptions, the level sets of the corresponding density are *hyper-ellipsoids* that are centered at $\boldsymbol{\mu}$ and whose shape and orientation are determined by $\boldsymbol{\Sigma}$, which justifies the terminology.

The scatter parameter $\boldsymbol{\Sigma}$ is of paramount importance in many inference procedures, including principal component analysis (PCA), canonical correlation analysis (CCA), testing for sphericity, etc. As it is often the case, however, these three applications do only require to know or to estimate the scatter matrix up to a positive scalar factor. In other words, factorizing $\boldsymbol{\Sigma}$ into $\sigma^2 \mathbf{V}$, where $\sigma^2 = (\det \boldsymbol{\Sigma})^{1/d}$ is a *scale* parameter and $\mathbf{V} = \boldsymbol{\Sigma}/(\det \boldsymbol{\Sigma})^{1/d}$ is a *shape* parameter, it is often so that the parameter of interest is $\mathbf{V}$, while σ^2 plays the role of a nuisance. In PCA, for instance, principal directions may be interchangeably computed from $\boldsymbol{\Sigma}$ or from $\mathbf{V}$, and both scatter and shape matrices will similarly lead to the same proportions of explained variances. Other factorizations of scatter into scale $\times$ shape are possible, such as those leading to shape matrices with fixed trace d or upper-left entry equal to one. The determinant-based normalization above was shown to be canonical in Paindaveine (2008), and it still plays a very particular role in the present work.

Many recent works focused on developing specific inference procedures for shape, and proposed, among others, tests of sphericity—or, more generally, tests that the underlying shape is equal to a given value. Most of these tests are based on estimators of shape—that are typically obtained by normalizing existing robust estimators of scatter. A quite systematic analysis of the properties of M-, S-, and R-estimators of shape has been performed in Frahm (2009). The shape estimator based on the celebrated MCD estimator of scatter was recently investigated in Paindaveine & Van Bever (2013). Shape estimators based on multivariate signs and ranks were proposed in Tyler (1987), Hallin *et al.* (2006) and Taskinen *et al.* (2010). Tests for sphericity relying explicitly on the concept of shape were proposed, e.g., in Hallin & Paindaveine (2006b), Sirkiä *et al.* (2009), and Paindaveine & Van Bever (2013).

The procedures above are of a *likelihood* nature. In particular, assuming that i.i.d. observations $\mathbf{X}_i$, $i = 1, \ldots, n$, are available, the corresponding estimators $\hat{\mathbf{V}}$ are M-estimators, in the sense that they are defined through

$$
\hat{\mathbf{V}} = \arg\min_{\mathbf{V}} \sum_{i=1}^{n} F_i(\mathbf{V}) \quad \text{or} \quad \sum_{i=1}^{n} \boldsymbol{a}_i(\hat{\mathbf{V}}) = \mathbf{0},
\tag{1.2}
$$

where the scalar $F_i(\mathbf{V})$ typically measures how well the shape parameter value $\mathbf{V}$ fits the observation $\mathbf{X}_i$ and where the vector $\boldsymbol{a}_i(\mathbf{V})$ is, e.g., taken as $\boldsymbol{a}_i(\mathbf{V}) = \nabla_{\mathbf{V}} F_i(\mathbf{V})$. In the important particular case where $F_i(\mathbf{V})$ is the ith term in the (Gaussian) log-likelihood, $\hat{\mathbf{V}}$ is the (Gaussian) maximum likelihood estimator of $\mathbf{V}$; see Section 2.2 below. The purpose of this paper is to introduce a depth notion for shape, *à la* Mizera (2002), that looks for a (possibly non-unique) Pareto-optimal value of $\mathbf{V}$, essentially minimizing individually as many $F_i(\mathbf{V})$ as possible, instead of minimizing a global measure of fit as in (1.2). More precisely, the depth of $\mathbf{V}$ will be defined as the minimal probability mass—in the empirical case, the minimal proportion of the sample—that needs to be removed for the shape value $\mathbf{V}$ not to be Pareto-optimal anymore. Instead of using this formulation of parametric depth in terms of $F_i(\mathbf{V})$, we rather use below the essentially equivalent Mizera (2002) "tangent depth" formulation that is based on the gradients $\nabla_{\mathbf{V}} F_i(\mathbf{V})$; see Section 3.

The outline of the paper is as follows. Section 2 fixes the notations we will use in the present elliptical setup, and discusses M-estimation for shape, which gives the opportunity to provide the score functions for shape that will be needed in the sequel. Section 3 first reviews the concept of (location) halfspace depth and its extension to an arbitrary parametric setup proposed by Mizera (2002) (Section 3.1). Then it defines the proposed concept of shape depth and establishes its affine-invariance (Section 3.2). Section 4 shows that shape depth is uniquely maximized at the true depth value. Section 5 introduces depth-based tests for shape, and investigates their finite-sample properties through simulations. While the previous sections actually focus on the determinant-based definition of shape, Section 6 extends the construction to other scale functionals, which requires to extend the parametric Mizera (2002) scheme into a semiparametric one. Section 7 briefly discusses the unspecified location case, as well as the application of the proposed concept to point estimation. Finally, the Appendix collects technical proofs.

2 Shape and M-estimation of shape in elliptical families

In this section, we first define the concept of *shape* in elliptical families and introduce the notation that will be needed in the sequel (Section 2.1). Then we discuss M-estimation of shape, and provide the scores that enter classical likelihood-based inference procedures (Section 2.2).

2.1 Shape

Consider a random d-vector $\mathbf{X}$ that has an elliptical distribution described by (1.1), for some location $\boldsymbol{\mu} \in \mathbb{R}^k$ and some scatter $\boldsymbol{\Sigma} \in \mathcal{S}_d$, where $\mathcal{S}_d$ denotes the set of all $d \times d$ symmetric positive definite matrices. Identifiability of $\boldsymbol{\Sigma}$ is ensured by imposing that the Mahalanobis distance $d_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{X})$ has median one. The unit vector $\mathbf{U}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{X}) = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})/d_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{X})$ is uniform over the unit sphere $\mathcal{S}^{d-1}$ in $\mathbb{R}^d$, and is independent of $d_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{X})$ (throughout, $\mathbf{A}^{1/2}$, for a symmetric and positive definite matrix $\mathbf{A}$, stands for the symmetric and positive definite square root of $\mathbf{A}$).

As in Paindaveine (2008), scale and shape parameters may be obtained from the scatter parameter $\boldsymbol{\Sigma}$ by using an arbitrary scale functional S.

Definition 2.1. A mapping $S : \mathcal{S}_d \to \mathbb{R}_0^+$ is a *scale functional* iff (i) S is *1-homogeneous* (i.e. $S(\lambda\boldsymbol{\Sigma}) = \lambda S(\boldsymbol{\Sigma}) \; \forall \lambda > 0$), (ii) S is differentiable with $\frac{\partial S}{\partial \boldsymbol{\Sigma}_{11}}(\boldsymbol{\Sigma}) \neq 0$ for all $\boldsymbol{\Sigma} \in \mathcal{S}_d$, and (iii) $S(\mathbf{I}_d) = 1$. The shape and scale parameters associated with $\boldsymbol{\Sigma} \in \mathcal{S}_d$ are then $\mathbf{V}_S(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}/S(\boldsymbol{\Sigma})$ and $\sigma_S^2(\boldsymbol{\Sigma}) = S(\boldsymbol{\Sigma})$, respectively.

Scale functionals therefore allow to factorize the scatter matrix $\boldsymbol{\Sigma}$ into $\sigma^2 \mathbf{V}$, where the scale σ^2 belongs to $\mathbb{R}_0^+$ and the shape $\mathbf{V}$ is in the set $\mathcal{V}_d^S = \left\{ \mathbf{V} \in \mathcal{S}_d : S(\mathbf{V}) = 1 \right\}$. Below, superscript and subscript S will only be used when the dependence on the underlying scale functional is needed. Classical scale functionals include

(i) $S_{\text{one}}(\boldsymbol{\Sigma}) = \Sigma_{11}$ (Hallin & Paindaveine, 2006a; Hettmansperger & Randles, 2002; Randles, 2000),

(ii) $S_{\text{trace}}(\boldsymbol{\Sigma}) = (\operatorname{tr} \boldsymbol{\Sigma})/d$ (Dümbgen, 1998; Tyler, 1987),

(iii) $S_{\text{det}}(\boldsymbol{\Sigma}) = (\det \boldsymbol{\Sigma})^{1/d}$ (Dümbgen & Tyler, 2005; Hallin & Paindaveine, 2008; Taskinen *et al.*, 2006; Tatsuoka & Tyler, 2000), and

(iv) $S_{\text{trace}^{-1}}(\boldsymbol{\Sigma}) = d/(\operatorname{tr} \boldsymbol{\Sigma}^{-1})$ (Frahm, 2009).

As advocated in Paindaveine (2008), the determinant-based standardisation S_{det} may be considered canonical since it is the only one for which the scale and shape parameters are orthogonal (meaning that the corresponding Fisher information matrix is block-diagonal); we refer to Frahm (2009) for other appealing properties of the scale functional S_{det}.

The following notation will be needed in the sequel. For any $d \times d$ matrix $\mathbf{A}$, let $\operatorname{vec}(\mathbf{A})$ be the d^2-dimensional vector resulting from stacking the columns of $\mathbf{A}$ one over each other. The Kronecker product $\mathbf{A} \otimes \mathbf{A}$ will be denoted as $\mathbf{A}^{\otimes 2}$. For a d-vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)'$, $\operatorname{diag}(\boldsymbol{\lambda})$ will be the $d \times d$ diagonal matrix with diagonal elements $\lambda_1, \ldots, \lambda_d$. If $\boldsymbol{\Sigma}$ is a symmetric matrix, denote by $\operatorname{vech}(\boldsymbol{\Sigma}) = (\Sigma_{11}, (\overset{\circ}{\operatorname{vech}} \boldsymbol{\Sigma})')'$ the $d(d+1)/2$-vector resulting from stacking the elements of the upper-triangular part of $\boldsymbol{\Sigma}$. The vector $\overset{\circ}{\operatorname{vech}} \boldsymbol{\Sigma}$ (with dimension $D := d(d+1)/2 - 1$) is therefore the vector $\operatorname{vech} \boldsymbol{\Sigma}$ deprived from its first component. Finally, for a given scale functional S and a shape matrix $\mathbf{V} \in \mathcal{V}_d^S$, $\mathbf{M}_S^{\mathbf{V}}$ will denote the $D \times d^2$ matrix such that $(\mathbf{M}_S^{\mathbf{V}})'(\overset{\circ}{\operatorname{vech}} \mathbf{v}) = \operatorname{vec}(\mathbf{v})$ for all matrices $\mathbf{v} \in \mathcal{S}_d$ satisfying $\left(\nabla S(\operatorname{vech} \mathbf{V}) \right)'(\operatorname{vech} \boldsymbol{v}) = 0$.

For a shape matrix $\mathbf{V} \in \mathcal{V}_d^S$, there is a one-to-one relationship between $\mathbf{V}$ (or $\operatorname{vech} \mathbf{V}$) and $\overset{\circ}{\operatorname{vech}} \mathbf{V}$, since V_{11} may be obtained from $\overset{\circ}{\operatorname{vech}} \mathbf{V}$ by imposing the constraint $S(\mathbf{V}) = 1$. The shape parameter is therefore $\overset{\circ}{\operatorname{vech}} \mathbf{V}$, and, in the sequel, $\nabla_{\mathbf{V}}$ will actually denote the gradient with respect to $\overset{\circ}{\operatorname{vech}} \mathbf{V}$.

2.2 M-estimation of shape

Consider the problem of conducting inference on the shape parameter on the basis of n mutually independent copies $\mathbf{X}_i$, $i = 1, \ldots, n$, of the random d-vector $\mathbf{X}$ considered in Section 2.1. In the absolutely continuous case, the common distribution of the $\mathbf{X}_i$'s admits a Lebesgue density of the form

$$\boldsymbol{x} \mapsto L_{\boldsymbol{\mu}, \sigma, \mathbf{V}; g}(\boldsymbol{x}) := \frac{1}{\sigma^d (\det \mathbf{V})^{1/2}} g\left(\frac{d_{\boldsymbol{\mu}, \mathbf{V}}(\boldsymbol{x})}{\sigma} \right), \tag{2.1}$$

where $\boldsymbol{\mu}(\in \mathbb{R}^d)$ is the location parameter, $\sigma(>0)$ and $\mathbf{V}(\in \mathcal{V}_d^S)$ are the scale and shape associated with the scatter $\boldsymbol{\Sigma}$, and where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is the *radial density*. As usual, maximum likelihood estimators of shape are simply obtained by solving the system of likelihood equations $\sum_{i=1}^n \nabla_{\mathbf{V}} \log L_{\boldsymbol{\mu}, \sigma, \mathbf{V}; g}(\mathbf{X}_i) = \mathbf{0}$, that rewrites

$$\frac{1}{2} \mathbf{M}_S^{\mathbf{V}} (\mathbf{V}^{\otimes 2})^{-1/2} \sum_{i=1}^n \operatorname{vec}\left(\varphi_g\left(\frac{d_{i; \boldsymbol{\mu}, \mathbf{V}}}{\sigma} \right) \frac{d_{i; \boldsymbol{\mu}, \mathbf{V}}}{\sigma} \mathbf{U}_{i; \boldsymbol{\mu}, \mathbf{V}} \mathbf{U}_{i; \boldsymbol{\mu}, \mathbf{V}}' - \mathbf{I}_d \right) = \mathbf{0}, \tag{2.2}$$

where we let $d_{i;\boldsymbol{\mu},\mathbf{V}} = d_{\boldsymbol{\mu},\mathbf{V}}(\mathbf{X}_i)$, $\mathbf{U}_{i;\boldsymbol{\mu},\mathbf{V}} = \mathbf{U}_{\boldsymbol{\mu},\mathbf{V}}(\mathbf{X}_i)$, and $\varphi_g = -g'/g$; see Hallin & Paindaveine (2006b). Particularizing to the determinant-based scale functional $S = S_{\det}$, (2.2) simplifies to

$$\frac{1}{2}\mathbf{M}_S^{\mathbf{V}}(\mathbf{V}^{\otimes 2})^{-1/2}\sum_{i=1}^{n}\varphi_g\Big(\frac{d_{i;\boldsymbol{\mu},\mathbf{V}}}{\sigma_S}\Big)\frac{d_{i;\boldsymbol{\mu},\mathbf{V}}}{\sigma_S}\,\mathrm{vec}\Big(\mathbf{U}_{i;\boldsymbol{\mu},\mathbf{V}}\mathbf{U}_{i;\boldsymbol{\mu},\mathbf{V}}' - \frac{1}{d}\mathbf{I}_d\Big) = \mathbf{0}, \qquad (2.3)$$

as we then have $\mathbf{M}_S^{\mathbf{V}}(\mathrm{vec}\,\mathbf{V}^{-1}) = \mathbf{0}$; see, e.g., Paindaveine (2008). Three examples of such likelihood estimators—or more generally, M-estimators—of shape are the following.

(i) Replacing g with the radial density associated with the multinormal case and solving (2.3) for $\mathbf{V}$ gives $\hat{\mathbf{V}}_{\mathcal{N},\boldsymbol{\mu}} = \mathbf{S}_{\boldsymbol{\mu}}/(\det \mathbf{S}_{\boldsymbol{\mu}})^{1/d}$, where $\mathbf{S}_{\boldsymbol{\mu}} = \frac{1}{n}\sum(\mathbf{X}_i-\boldsymbol{\mu})(\mathbf{X}_i-\boldsymbol{\mu})'$ is the classical covariance matrix with known location $\boldsymbol{\mu}$. Of course, when $\boldsymbol{\mu}$ is unspecified, jointly solving (2.3) and the corresponding likelihood equations for location would provide the Gaussian maximum likelihood shape estimator $\hat{\mathbf{V}}_{\mathcal{N}} = \mathbf{S}/(\det \mathbf{S})^{1/d}$, with $\mathbf{S} = \frac{1}{n}\sum(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$, that is simply the appropriately normalized regular covariance matrix.

(ii) For a fixed location $\boldsymbol{\mu}$, the well-known estimator of shape from Tyler (1987)—namely the unique solution $\mathbf{V}$ (with determinant one), of $\frac{1}{n}\sum_{i=1}^{n}\mathbf{U}_{i;\boldsymbol{\mu},\mathbf{V}}\mathbf{U}_{i;\boldsymbol{\mu},\mathbf{V}}' = \frac{1}{d}\mathbf{I}_d$—is obtained as the limiting solution (as $\nu \to 0$) of (2.3) when g is taken as the radial density g_ν of the elliptical t distribution with ν degrees of freedom (note indeed that $\varphi_{g_\nu}(z) \propto (d+\nu)z/(\nu+z^2)$, so that $\varphi_{g_\nu}(z)z$ goes to a constant as $\nu \to 0$). For the unspecified location case, the corresponding estimator of $(\boldsymbol{\mu},\mathbf{V})$ is the one proposed in Hettmansperger & Randles (2002).

(iii) Alternatively, R-estimators of shape may be obtained by replacing, in (2.3), $\varphi_g(d_i/\sigma)d_i/\sigma$ with a function $K(R_i)$ involving the rank R_i of d_i among $d_1,\ldots,d_n$. The resulting rank-based estimators of shape exhibit very good robustness and efficiency properties; see Hallin *et al.* (2006).

Very classically, these examples all involve minimization of an aggregate, global, objective function; see (1.2) in the Introduction. Minimizing, in a "Pareto" fashion, individual measures of fit (as it is the case with maxdepth estimation) will allow the definition of an alternative inferential approach for shape, described in the next section.

3 Shape depth

Inference on location $\boldsymbol{\mu}$ and shape $\mathbf{V}$ is of obvious interest. The literature provides abundant solutions to the location problem. One classical way to estimate the parameter $\boldsymbol{\mu}$ is through depth functions, measuring the *centrality* of any location with respect to the underlying population, therefore providing a (so-called *maxdepth*) estimate of the location parameter as the point in space with maximal depth. Interestingly, the notion of depth has been extended to, first, the regression setup (Rousseeuw & Hubert, 1999), and, later, to any parametric setup (Mizera, 2002; Mizera & Müller, 2004). Estimation based on depth in such setups can be achieved in the same fashion, by finding the parameter value with largest *tangent depth*. In this section, we first review (Section 3.1) the classical concept of halfspace (location) depth (Tukey, 1975) as well as that of tangent depth (Mizera, 2002). Then we introduce (Section 3.2) the proposed shape depth concept, in the particular case in which shapes are normalized to have determinant one.

3.1 Halfspace depth and tangent depth

Depth functions, of the form $D : \mathbb{R}^d \to \mathbb{R}^+ : \boldsymbol{\mu} \mapsto D(\boldsymbol{\mu}, P)$, measure the centrality of an arbitrary location $\boldsymbol{\mu}$ with respect to a probability measure P on $\mathbb{R}^d$. Many depths are available in the literature, among which the simplicial depth (Liu, 1990), the spatial depth (Vardi & Zhang, 2000), the standardized spatial depth (Zuo & Serfling, 2000), the projection depth (Zuo, 2003), the Mahalanobis depth (Zuo & Serfling, 2000), the simplicial volume depth (Oja, 1983; Zuo & Serfling, 2000), or the zonoid depth (Koshevoy & Mosler, 1997). Here, we focus on the following celebrated depth.

Definition 3.1 (Tukey, 1975)**.** The *halfspace depth* of $\boldsymbol{\mu}$ with respect to the random d-vector $\mathbf{X}$ having distribution P is $D_H(\boldsymbol{\mu}, P) = \inf_{\boldsymbol{u} \in \mathcal{S}^{d-1}} P[\boldsymbol{u}'(\mathbf{X} - \boldsymbol{\mu}) \geq 0]$, the smallest probability mass of any halfspace whose boundary hyperplane contains $\boldsymbol{\mu}$.

Interestingly, an equivalent definition of halfspace depth, allowing the extension of depth to other setups, is the following. Let the objective function $\boldsymbol{\mu} \mapsto F_{\boldsymbol{\mu}}(\boldsymbol{x})$ measure how well (actually, how poorly) the parameter value $\boldsymbol{\mu}$ "fits" the point $\boldsymbol{x}$ in the sample space; one may, typically, take $F_{\boldsymbol{\mu}}(\boldsymbol{x}) = h(\|\boldsymbol{x} - \boldsymbol{\mu}\|)$, where $h : \mathbb{R}^+ \to \mathbb{R}^+$ is smooth and monotone increasing. Letting $\mathbf{0}_d = (0, \ldots, 0)' \in \mathbb{R}^d$, one may then directly check that

$$D_H(\boldsymbol{\mu}, P) = D_H(\mathbf{0}_d, P_{\nabla_{\boldsymbol{\mu}} F_{\boldsymbol{\mu}}(\mathbf{X})}) \tag{3.1}$$

(throughout, $P_{g(\mathbf{X})}$ denotes the distribution of the random vector $g(\mathbf{X})$ when $\mathbf{X}$ has distribution P). In the empirical case, this amounts to looking at the depth of $\mathbf{0}_d$ among the directions $\nabla_{\boldsymbol{\mu}} F_{\boldsymbol{\mu}}(\mathbf{X}_i)$, $i = 1, \ldots, n$, of maximal increase of $F_{\boldsymbol{\mu}}(\cdot)$. As mentioned in the Introduction, depth somehow individually minimizes such objective functions rather than minimizing a global objective function of the form $\sum_{i=1}^{n} F_{\boldsymbol{\mu}}(\mathbf{X}_i)$.

Mizera (2002) based on (3.1) a concept of *tangent depth*, that extends location depth to an arbitrary parametric model. In order to describe this concept, consider a random d-vector $\mathbf{X}$ with a distribution $P = P_{\boldsymbol{\theta}_0}$ in the parametric family $\mathcal{P} = \{P_{\boldsymbol{\theta}} | \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k\}$ (here, k may be different from d). As above, let $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}(\boldsymbol{x})$ be a measure of fit of the parameter value $\boldsymbol{\theta}$ for the observation $\boldsymbol{x}$. The following concept then typically attributes large depth to "good" parameter values, that is, to parameter values $\boldsymbol{\theta}$ that are close to $\boldsymbol{\theta}_0$.

Definition 3.2 (Mizera, 2002)**.** The *tangent depth* of $\boldsymbol{\theta}$ with respect to $P = P_{\mathbf{X}}$ is $TD(\boldsymbol{\theta}, P) = D_H(\mathbf{0}, P_{\nabla_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(\mathbf{X})})$.

An important particular case is the linear regression setup, where the observation takes the form $(\mathbf{X}, Y)'$, with values in $\mathbb{R}^{p-1} \times \mathbb{R}$, and involves a $(p-1)$-dimensional covariate $\mathbf{X}$ and a scalar response Y, while the parameter value $\boldsymbol{\theta}(\in \boldsymbol{\Theta} = \mathbb{R}^p)$ is associated with the regression hyperplane $y = \boldsymbol{\theta}'\binom{1}{\boldsymbol{x}}$. For $F_{\boldsymbol{\theta}}(\boldsymbol{x}, y) = h(|y - \boldsymbol{\theta}'\binom{1}{\boldsymbol{x}}|)$, where $h : \mathbb{R}^+ \to \mathbb{R}^+$ is still smooth and monotone increasing, tangent depth reduces to the well-known *regression depth* from Rousseeuw & Hubert (1999).

In some setups, it may be difficult to choose an appropriate objective function $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}(\boldsymbol{x})$. A general, *likelihood-based*, approach consists in taking $F_{\boldsymbol{\theta}}(\boldsymbol{x}) = -\log L_{\boldsymbol{\theta}}(\boldsymbol{x})$, where $L_{\boldsymbol{\theta}}(\boldsymbol{x})$ stands for the likelihood function; see, e.g., Mizera & Müller (2004); Müller (2005). It is easily checked that, in the location and regression cases considered above, Gaussian or t_{ν}-likelihoods lead to the Tukey (1975) location depth and Rousseeuw & Hubert (1999) regression depth, respectively. However, the

resulting depth concepts may severely depend on the likelihood function used; an example is given by the location-scale depth from Mizera & Müller (2004), where different unimodal densities may give rise to heterogeneous depth functionals.

3.2   Shape depth

For the determinant-based scale functional S_{det} (to which, unless otherwise stated, we restrict up to Section 6), the discussion above makes it natural to define the *shape depth* of an arbitrary shape value $\mathbf{V}$, with respect to a distribution $P = P_{\mathbf{X}}$ on $\mathbb{R}^d$, as $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) = D_H\big(\mathbf{0}_D, P_{\nabla_{\mathbf{V}} \log L_{\boldsymbol{\mu},\sigma,\mathbf{V};g}(\mathbf{X})}\big)$, where

$$\nabla_{\mathbf{V}} \log L_{\boldsymbol{\mu},\sigma,\mathbf{V};g}(\mathbf{X}) = \frac{1}{2}\, \varphi_g\Big(\frac{d_{\boldsymbol{\mu},\mathbf{V}}}{\sigma}\Big)\frac{d_{\boldsymbol{\mu},\mathbf{V}}}{\sigma}\, \mathbf{M}_S^{\mathbf{V}}(\mathbf{V}^{\otimes 2})^{-1/2} \text{vec}\Big(\mathbf{U}_{\boldsymbol{\mu},\mathbf{V}}\mathbf{U}'_{\boldsymbol{\mu},\mathbf{V}} - \frac{1}{d}\mathbf{I}_d\Big) \qquad (3.2)$$

is the corresponding score for $\mathbf{V}$; see (2.3). Assuming that g is monotone strictly decreasing, the scalar quantity $\varphi_g(d_{\boldsymbol{\mu},\mathbf{V}}/\sigma)d_{\boldsymbol{\mu},\mathbf{V}}/\sigma$ is positive, hence may be dropped in (3.2) without affecting the depth, that therefore does not depend on σ nor g. This leads to the following definition.

Definition 3.3. For any $\mathbf{V} \in \mathcal{V}_d^S$, the *shape depth* of $\mathbf{V}$, with respect to the distribution $P = P_{\mathbf{X}}$ on $\mathbb{R}^d$ having known deepest point $\boldsymbol{\mu}$, is $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) = D_H\big(\mathbf{0}_D, P_{\mathbf{W}^{\boldsymbol{\mu},\mathbf{V}}}\big)$, where $\mathbf{W}^{\boldsymbol{\mu},\mathbf{V}} = \mathbf{M}_S^{\mathbf{V}}(\mathbf{V}^{\otimes 2})^{-1/2} \text{vec}\big(\mathbf{U}_{\boldsymbol{\mu},\mathbf{V}}\mathbf{U}'_{\boldsymbol{\mu},\mathbf{V}} - \frac{1}{d}\mathbf{I}_d\big)$.

While shape depth does not depend on σ and g (in contrast with other instances of tangent depth, such as, e.g., the location-scale depth from Mizera & Müller (2004)), it does explicitly involve the deepest location $\boldsymbol{\mu}$[2]—which, in the elliptical case, coincides with the location parameter $\boldsymbol{\mu}$ from the previous sections. Definition 3.3 therefore applies in the specified location case; extension to the unspecified location case will be briefly discussed in Section 7.

As mentioned above, we restricted here to the determinant-based scale functional S_{det}. Extending the concept of shape depth to an arbitrary scale functional is non-trivial, and, in particular, building tangent depth on the generic score in (2.2) instead of the determinant-based score in (2.3) would not allow to get rid of σ nor g. A proper extension of shape depth to an arbitrary scale functional will actually require modifying the Mizera (2002) tangent depth construction, which will be achieved in Section 6.

Coming back to Definition 3.3, it is interesting to note that the shape depth there involves the underlying $\mathbf{X}$ only through the direction—or multivariate sign—$\mathbf{U}_{\boldsymbol{\mu},\mathbf{V}}$. Consequently, shape depth intrinsically is a *sign* concept, and the resulting inference procedures, parallel to Tyler's estimator of shape introduced in Section 2.2, will be *multivariate sign procedures*, hence will enjoy natural robustness properties. Since the Mahalanobis distance $d_{\boldsymbol{\mu},\mathbf{V}}$ is not involved in shape depth, any result we will prove under the assumption of ellipticity will *de facto* also hold under the much weaker assumption that the underlying distribution $P = P_{\mathbf{X}}$ has *elliptical directions*—in the sense that $\mathbf{X}$ is distributed as $R\mathbf{A}\mathbf{U}$, where $\mathbf{U}$ is uniform over the unit sphere $\mathcal{S}^{d-1}$ in $\mathbb{R}^d$ and where the nonnegative random variable R may be stochastically dependent of $\mathbf{U}$ (and may fail to be absolutely continuous).

In the location setup, affine-invariance is one of the classical requirements for depth functions; see Property (P1) in Zuo & Serfling (2000). The following theorem states that the proposed shape depth inherits the affine-invariance properties of its location antecedent, namely halfspace depth.

---

[2]If uniqueness does not hold, we follow the standard practice that consists in defining $\boldsymbol{\mu}$ as the barycenter of the deepest region.

Theorem 3.1. Fix $\mathbf{V} \in \mathcal{V}_d^S$ and an arbitrary distribution $P = P_{\mathbf{X}}$ on $\mathbb{R}^d$. Then

$$ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) = ShD_{\mathbf{A}\boldsymbol{\mu}+\mathbf{b}}\Big(\frac{\mathbf{A}\mathbf{V}\mathbf{A}'}{S(\mathbf{A}\mathbf{V}\mathbf{A}')}, P_{\mathbf{A}\mathbf{X}+\mathbf{b}}\Big),$$

for any invertible $d \times d$ matrix $\mathbf{A}$ and any d-vector $\mathbf{b}$.

The proof requires extra insight on the random vector $\mathbf{W}^{\boldsymbol{\mu},\mathbf{V}}$ and is therefore deferred to the Appendix.

4 Consistency

In this section, we provide a "consistency" result, stating that, under ellipticity, the shape depth from the previous section is uniquely maximized at the true shape value. The proof relies on a preliminary result, that in turn requires introducing the following notations. We will write $\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}$, $\overline{\mathbf{W}}_{D+1}^{\boldsymbol{\mu},\mathbf{V}}$, and $\overline{\mathbf{W}}_D^{\boldsymbol{\mu},\mathbf{V}}$, for the vec, the vech, and the vech forms of the random matrix $\mathbf{U}_{\boldsymbol{\mu},\mathbf{V}}\mathbf{U}_{\boldsymbol{\mu},\mathbf{V}}' - \frac{1}{d}\mathbf{I}_d$, respectively. We then have the following result.

Lemma 4.1. *Fix* $\mathbf{V} \in \mathcal{V}_d^S$ *and an arbitrary distribution* $P = P_{\mathbf{X}}$ *on* $\mathbb{R}^d$. *Then* $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) \overset{(i)}{=} D_H\big(\mathbf{0}_{d^2}, P_{\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}}\big) \overset{(ii)}{=} D_H\big(\mathbf{0}_{D+1}, P_{\overline{\mathbf{W}}_{D+1}^{\boldsymbol{\mu},\mathbf{V}}}\big) \overset{(iii)}{=} D_H\big(\mathbf{0}_D, P_{\overline{\mathbf{W}}_D^{\boldsymbol{\mu},\mathbf{V}}}\big).$

This lemma is the key result to many properties of shape depth. In particular, it is the main step in the proof of the affine-invariance property in Theorem 3.1, and, as announced, it also plays a crucial role in the proof of the following consistency result.

Theorem 4.1. Fix an arbitrary elliptical distribution $P = P_{\mathbf{X}}$ on $\mathbb{R}^d$ ($d \in \{2,3\}$), with location $\boldsymbol{\mu}$ and shape $\mathbf{V}_0$. Then, for all $\mathbf{V} \in \mathcal{V}_d^S$, the (known-location) shape depth of $\mathbf{V}$ satisfies $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P) \leq ShD_{\boldsymbol{\mu}}(\mathbf{V}_0, P)$, with equality if and only if $\mathbf{V} = \mathbf{V}_0$.

The proof is long and deferred to the Appendix. However, we summarize some intermediary steps below, as they bring much information about the structure of the proposed shape depth.

First note that it is sufficient to prove Theorem 4.1 for $\mathbf{V}_0 = \mathbf{I}_d$: if the elliptical random d-vector $\mathbf{X}$ has shape $\mathbf{V}_0$, then affine-invariance yields $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) = ShD_{\boldsymbol{\mu}}(\mathbf{V}_0^{-1/2}\mathbf{V}\mathbf{V}_0^{-1/2}, P_{\mathbf{Y}})$, where $\mathbf{Y} = \mathbf{V}_0^{-1/2}\mathbf{X}$ is spherically distributed, i.e., is elliptically distributed with shape matrix $\mathbf{I}_d$. Applying the $\mathbf{V}_0 = \mathbf{I}_d$ result in Theorem 4.1 therefore implies that

$$ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) = ShD_{\boldsymbol{\mu}}(\mathbf{V}_0^{-1/2}\mathbf{V}\mathbf{V}_0^{-1/2}, P_{\mathbf{Y}}) \leq ShD_{\boldsymbol{\mu}}(\mathbf{I}_d, P_{\mathbf{Y}}),$$

with equality if and only if $\mathbf{V}_0^{-1/2}\mathbf{V}\mathbf{V}_0^{-1/2} = \mathbf{I}_d$, that is, if $\mathbf{V} = \mathbf{V}_0$. This shows that the general statement in Theorem 4.1 indeed follows from the $\mathbf{V}_0 = \mathbf{I}_d$ subresult.

Now, this subresult is a direct corollary of the following three lemmas.

Lemma 4.2. *Fix an arbitrary elliptical distribution* $P = P_{\mathbf{X}}$ *on* $\mathbb{R}^d$ ($d \in \{2,3\}$), *with location* $\boldsymbol{\mu}$ *and shape* $\mathbf{V}_0 = \mathbf{I}_d$. *Then, letting* $\mathcal{D}_d := \{\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)' \in \mathbb{R}^d : \sum_{i=1}^d \lambda_i = 1\}$, *we have*

$$ShD_{\boldsymbol{\mu}}(\mathbf{I}_d, P_{\mathbf{X}}) = \inf_{\boldsymbol{\lambda} \in \mathcal{D}_d} P\Big[\sum_{i=1}^d \lambda_i U_i^2 \geq 1/d\Big],$$

where $\mathbf{U} = (U_1, \ldots, U_d)'$ *is uniformly distributed over the unit sphere* $\mathcal{S}^{d-1}$ *in* $\mathbb{R}^d$.

Lemma 4.3. *Let* $\mathbf{U} = (U_1, \ldots, U_d)'$ *be uniformly distributed over* $\mathcal{S}^{d-1}$ *(* $d \in \{2, 3\}$ *). Then,* $\inf_{\boldsymbol{\lambda} \in \mathcal{D}_d} P[\sum_{i=1}^d \lambda_i U_i^2 \geq 1/d] = P[U_1^2 > 1/d] = 1 - F^{\mathrm{Beta}}\big(1/d; 1/2, (d-1)/2\big)$, *where* $F^{\mathrm{Beta}}(\,\cdot\,; a, b)$ *denotes the cumulative distribution function of the* $\mathrm{Beta}(a, b)$ *distribution.*

Lemma 4.4. *Fix an arbitrary elliptical distribution* $P = P_{\mathbf{X}}$ *on* $\mathbb{R}^d$ *(* $d \in \{2, 3\}$ *), with location* $\boldsymbol{\mu}$ *and shape* $\mathbf{V}_0 = \mathbf{I}_d$, *and let again* $\mathbf{U} = (U_1, \ldots, U_d)'$ *be uniformly distributed over* $\mathcal{S}^{d-1}$. *Then, for any* $\mathbf{V} \in \mathcal{V}_d^S \setminus \{\mathbf{I}_d\}$, $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) < P[U_1^2 > 1/d]$.

Lemma 4.2 states that the (shape) depth of the "true" shape matrix coincides with the halfspace depth, with respect to the Dirichlet distribution of $(U_1^2, \ldots, U_d^2)'$, of the mean $\frac{1}{d}\mathbf{1}_d$ of this Dirichlet (here, $\mathbf{1}_d$ denotes the d-variate vector of ones). Lemma 4.3 provides one of the minimal halfspaces of this depth problem. Finally, Lemma 4.4 obviously is the key result to establish unicity of the maximizer in (the $\mathbf{V}_0 = \mathbf{I}_d$ version of) Theorem 4.1. All details can be found in the Appendix.

5 Inferential applications

In this section, we turn to depth-based inference for shape, which of course requires considering the *sample* version of the population concept introduced in Definition 3.3. Assuming i.i.d. d-variate observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are available (with corresponding empirical distribution $P^{(n)}$, say), we define the *sample shape depth* of $\mathbf{V}$ with respect to $P^{(n)}$ as

$$ShD_{\boldsymbol{\mu}}(\mathbf{V}, P^{(n)}) = D_H(\mathbf{0}, P^{(n)}_{\mathbf{W}^{\boldsymbol{\mu}}, \mathbf{V}}) = \frac{1}{n} \min_{\boldsymbol{u} \in \mathcal{S}^{D-1}} \sharp\Big\{ i = 1, \ldots, n \,\big|\, \boldsymbol{u}' \mathbf{W}_i^{\boldsymbol{\mu}, \mathbf{V}} \geq 0 \Big\},$$

where $P^{(n)}_{\mathbf{W}^{\boldsymbol{\mu}}, \mathbf{V}}$ stands for the empirical distribution of

$$\mathbf{W}_i^{\boldsymbol{\mu}, \mathbf{V}} = \mathbf{M}_S^{\mathbf{V}} \big(\mathbf{V}^{\otimes 2}\big)^{-1/2} \mathrm{vec}\Big(\mathbf{U}_{i; \boldsymbol{\mu}, \mathbf{V}} \mathbf{U}'_{i; \boldsymbol{\mu}, \mathbf{V}} - \frac{1}{d}\mathbf{I}_d \Big), \qquad i = 1, \ldots, n. \tag{5.1}$$

As the notation suggests, this defines sample shape depth for specified location $\boldsymbol{\mu}$; more precisely, in the same spirit as in Definition 3.3, this covers the case where the common deepest point of the $\mathbf{X}_i$'s is known to be equal to $\boldsymbol{\mu}$.

We here restrict to hypothesis testing (point estimation will be briefly discussed in Section 7). More specifically, we consider the case where, on the basis of i.i.d. d-variate observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ that have a common elliptical distribution with known location center $\boldsymbol{\mu}$ and unknown shape matrix $\mathbf{V}$ (normalized to have determinant one), we want to test

$$\mathcal{H}_{0, \boldsymbol{\mu}} : \mathbf{V} = \mathbf{V}_0 \qquad \text{vs} \qquad \mathcal{H}_{1, \boldsymbol{\mu}} : \mathbf{V} \neq \mathbf{V}_0$$

at level $\alpha \in (0, 1)$, where $\mathbf{V}_0 \in \mathcal{V}_d^{S_{\mathrm{det}}}$ is fixed. The important particular case obtained with $\mathbf{V}_0 = \mathbf{I}_d$ corresponds to the problem of testing sphericity.

Theorem 4.1 suggests to consider a depth-based test rejecting the null for small values of $T_{\boldsymbol{\mu}}^{(n)} = ShD_{\boldsymbol{\mu}}(\mathbf{V}_0, P^{(n)})$, where $P^{(n)}$ denotes the empirical distribution of the $\mathbf{X}_i$'s. Clearly, the random vectors $\mathbf{W}_i^{\boldsymbol{\mu}, \mathbf{V}_0}$ (see (5.1)) are distribution-free under the null $\mathcal{H}_{0, \boldsymbol{\mu}}$. In particular, the α-quantile $t_\alpha^{(n)}$ of $T_{\boldsymbol{\mu}}^{(n)}$ under $\mathcal{H}_{0, \boldsymbol{\mu}}$ does not depend on the underlying radial density g; the notation $t_\alpha^{(n)}$ is justified by the fact that this quantile also does not depend on $\boldsymbol{\mu}$ nor—in view of the affine-invariance of shape depth—on the null shape value $\mathbf{V}_0$. Consequently, evaluating the sample α-quantile in a (large) collection of m values of the ($\mathbf{V}_0 = \mathbf{I}_d$) test statistic, obtained from m mutually independent

standard normal random samples, will provide a valid[3] estimate $\hat{t}_\alpha^{(n;m)}$ of $t_\alpha^{(n)}$ (estimated values of $\hat{t}_\alpha^{(n;5,000)}$ for various values of α and n are provided in Table 1 for $d = 2$ and for $d = 3$). The resulting depth-based test then rejects the null $\mathcal{H}_{0,\boldsymbol{\mu}}$ at level α whenever $T^{(n)} < \hat{t}_\alpha^{(n;m)}$.

Table 1: *Estimated critical values $\hat{t}_\alpha^{(n,m)}$ from $m = 5,000$ independent d-dimensional standard normal random samples ($d = 2, 3$), for various values of the nominal level α and sample size n.*

	$d = 2$ $(d = 3)$				
$\alpha \setminus n$	50	200	500	1,000	10,000
0.01	0.26 (0.14)	0.38 (0.28)	0.422 (0.334)	0.443 (0.360)	0.4824 (0.4031)
0.025	0.28 (0.16)	0.385 (0.29)	0.428 (0.340)	0.448 (0.364)	0.4837 (0.4044)
0.05	0.30 (0.18)	0.395 (0.295)	0.434 (0.344)	0.452 (0.368)	0.4848 (0.4056)
0.1	0.32 (0.18)	0.405 (0.305)	0.438 (0.350)	0.457 (0.371)	0.4862 (0.4070)
0.2	0.34 (0.20)	0.415 (0.315)	0.446 (0.356)	0.462 (0.376)	0.4879 (0.4084)

A simulation study was conducted in order to assess the finite-sample behavior of the proposed depth-based test. Four competitors were considered :

(i) The $\boldsymbol{\mu}$-specified version of the Gaussian test from John (1972)—more precisely, the $\boldsymbol{\mu}$-specified version of its robustification to any elliptical distributions with finite fourth-order moments defined in Hallin & Paindaveine (2006b). This test is based on the Gaussian maximum likelihood estimator of shape $\hat{\mathbf{V}}_{\mathcal{N},\boldsymbol{\mu}} = \mathbf{S}_{\boldsymbol{\mu}}/(\det \mathbf{S}_{\boldsymbol{\mu}})^{1/d}$, where $\mathbf{S}_{\boldsymbol{\mu}} = \frac{1}{n}\sum(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'$ is the classical covariance matrix with known location $\boldsymbol{\mu}$; see Section 2.2. The test statistic is

$$Q_{\mathcal{N},\boldsymbol{\mu}}^{(n)} = \frac{n\hat{c}_d}{2}\left(\text{tr}\left[(\mathbf{V}_0^{-1}\hat{\mathbf{V}}_{\mathcal{N},\boldsymbol{\mu}})^2\right] - \frac{1}{d}\text{tr}^2\left[\mathbf{V}_0^{-1}\hat{\mathbf{V}}_{\mathcal{N},\boldsymbol{\mu}}\right]\right),$$

with $\hat{c}_d := (d+2)(\frac{1}{n}\sum_{i=1}^n d_{i;\boldsymbol{\mu},\mathbf{S}_{\boldsymbol{\mu}}}^2)^2/[d(\frac{1}{n}\sum_{i=1}^n d_{i;\boldsymbol{\mu},\mathbf{S}_{\boldsymbol{\mu}}}^4)]$.

(ii)-(iii) Two multivariate signed-rank tests from Hallin & Paindaveine (2006b), that are based on test statistics of the form

$$Q_{K,\boldsymbol{\mu}}^{(n)} = \frac{nd(d+2)}{2(\int_0^1 K^2(u)\,du)}\left(\text{tr}\left[(\mathbf{V}_0^{-1}\hat{\mathbf{V}}_{K,\boldsymbol{\mu}})^2\right] - \frac{1}{d}\text{tr}^2\left[\mathbf{V}_0^{-1}\hat{\mathbf{V}}_{K,\boldsymbol{\mu}}\right]\right),$$

where $\hat{\mathbf{V}}_{K,\boldsymbol{\mu}} = \mathbf{S}_{K,\boldsymbol{\mu}}/(\det \mathbf{S}_{K,\boldsymbol{\mu}})^{1/d}$ is the shape associated with the (null) signed-rank scatter matrix $\mathbf{S}_{K,\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^n K(R_{i;\boldsymbol{\mu},\mathbf{V}_0}/(n+1))\mathbf{U}_{i;\boldsymbol{\mu},\mathbf{V}_0}\mathbf{U}_{i;\boldsymbol{\mu},\mathbf{V}_0}'$; here, $R_{i;\boldsymbol{\mu},\mathbf{V}_0}$ stands for the rank of $d_{i;\boldsymbol{\mu},\mathbf{V}_0}$ among $d_{1;\boldsymbol{\mu},\mathbf{V}_0}, \ldots, d_{n;\boldsymbol{\mu},\mathbf{V}_0}$. The test (ii) is a pure sign test based on $K(u) \equiv 1$, whereas the test (iii) is a "van der Waerden" signed-rank test based on the (Gaussian) score function $K(u) = \Psi_d^{-1}(u)$, where Ψ_d denotes the cumulative distribution function of the chi-square distribution with d degrees of freedom.

(iv) The test based on the MCD_γ shape estimator $\hat{\mathbf{V}}_{\text{MCD}_\gamma,\boldsymbol{\mu}} = \mathbf{S}_{\text{MCD}_\gamma,\boldsymbol{\mu}}/(\det \mathbf{S}_{\text{MCD}_\gamma,\boldsymbol{\mu}})^{1/d}$, where $\mathbf{S}_{\text{MCD}_\gamma,\boldsymbol{\mu}}$ is the celebrated MCD_γ estimator of scatter; γ, that determines the proportion of the sample retained to compute the final covariance matrix estimate, was fixed to 0.8 to achieve a good balance between efficiency and robustness. The corresponding test statistic is

$$Q_{\text{MCD}_\gamma,\boldsymbol{\mu}}^{(n)} = \frac{n\hat{c}_\gamma}{2}\left(\text{tr}\left[(\mathbf{V}_0^{-1}\hat{\mathbf{V}}_{\text{MCD},\boldsymbol{\mu}})^2\right] - \frac{1}{d}\text{tr}^2\left[\mathbf{V}_0^{-1}\hat{\mathbf{V}}_{\text{MCD},\boldsymbol{\mu}}\right]\right).$$

---

[3]Due to the discreteness of the distribution of $T_{\boldsymbol{\mu}}^{(n)}$, randomization may be needed to achieve exact null size α.

We refer to Paindaveine & Van Bever (2013) for an expression of $\hat{c}_\gamma$.

The tests in (i)-(iv) all reject the null for large values of the corresponding test statistics, whose common null asymptotic distribution is chi-square with $D = d(d+1)/2 - 1$ degrees of freedom. Consequently, these four competing tests reject the null at asymptotic level α whenever their test statistic exceeds the upper α-quantile $\Psi_D^{-1}(1-\alpha)$ of the χ_D^2 distribution.

We performed a first simulation to investigate the finite-sample performances of our depth-based tests in terms of power. We restricted to the problem of testing bivariate sphericity about the origin ($\boldsymbol{\mu} = \mathbf{0}_d$ and $\mathbf{V}_0 = \mathbf{I}_d$, with $d = 2$), at level $\alpha = 5\%$. We considered three types of bivariate elliptical distributions, namely standard normal, t_5, and Cauchy ones. For each type of distribution and each value of ℓ, we generated $M = 3,000$ independent random samples $\mathbf{X}_i$, $i = 1, \ldots, n = 500$, with location center $\boldsymbol{\mu} = \mathbf{0}_2$ and shape matrix

$$V(\ell, \xi) = \frac{\mathbf{I}_2 + \ell\xi\left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & -1 \end{smallmatrix}\right)}{\left(\det\left[\mathbf{I}_2 + \ell\xi\left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & -1 \end{smallmatrix}\right)\right]\right)^{1/2}}, \qquad \ell = 0, 1, \ldots, 6;$$

the value $\ell = 0$ corresponds to the null, whereas $\ell = 1, \ldots, 6$ are associated with increasingly severe alternatives. The value of ξ was chosen as $\xi = 0.035$, 0.04, and 0.045 for bivariate normal, t_5, and Cauchy samples, respectively, to ensure that the most extreme alternatives (corresponding to $\ell = 6$) for each type of distribution lead to approximately identical (and close to one) rejection frequencies.

For each sample, we performed the proposed depth-based test (associated with the estimated critical value $t_\alpha^{(n,m)} = 0.434$ from Table 1, with $\alpha = 0.05$, $n = 500$, and $m = 5,000$) and its four competitors (i)-(iv) above (based on the asymptotic critical value $\Psi_D^{-1}(1-\alpha) = \Psi_2^{-1}(0.95) \approx 5.99$). Plots of the resulting rejection frequencies (as a function of ℓ) are reported in Figure 1. Clearly, the empirical power curves of the proposed depth-based test are very similar to (although they may seem slightly lower than) those of the sign test in (ii), which is in line with the fact that the depth-based test is also of a sign nature. Consequently, the proposed depth-based test performs very well under heavy tails (it can indeed be checked that the sign test in (ii) coincides with the a.e. limit, as $\nu \to 0$, of the tests—ϕ_ν, say— achieving parametric optimality under t_ν elliptical densities), hence beats all other tests there. As expected, the MCD$_\gamma$-based test shows low empirical powers (although the proportion γ of observations used to estimate the shape parameter is quite large; see Paindaveine & Van Bever, 2013), and the Gaussian test collapses under infinite fourth-order moments.

We conducted another simulation in order to compare the various tests in terms of robustness. There is no general agreement on how to study resistance of tests subject to contamination; here, we focused on the concept of "level robustness" as described in He *et al.* (1990). More precisely, we investigated the impact on the null size of each test above under various contaminations of the null hypothesis $\mathcal{H}_{0,\boldsymbol{\mu}} : \mathbf{V} = \mathbf{V}_0$, with $\boldsymbol{\mu} = \mathbf{0}_2$ and $\mathbf{V}_0 = \text{diag}(2, 1/2)$. To do so, we considered mixture distributions of the form $P^{\mathbf{X}^{(\eta)}} = (1 - B_\eta)P^{\mathbf{X}} + B_\eta P^{\mathbf{Y}}$, where B is a Bern(η) random variable independent from $\mathbf{X}$ and $\mathbf{Y}$, with $\eta = 0$ (zero contamination), .025, .05, .1 and .2 (increasingly severe contamination). Here, $\mathbf{X}$ is a bivariate normal, t_5, or Cauchy elliptically symmetric random vector, with center $\boldsymbol{\mu}$ and shape $\mathbf{V}_0$ as above, hence is compatible with the null hypothesis. The distribution of the bivariate random vector $\mathbf{Y}$ determines the contamination pattern considered, and was chosen as follows:
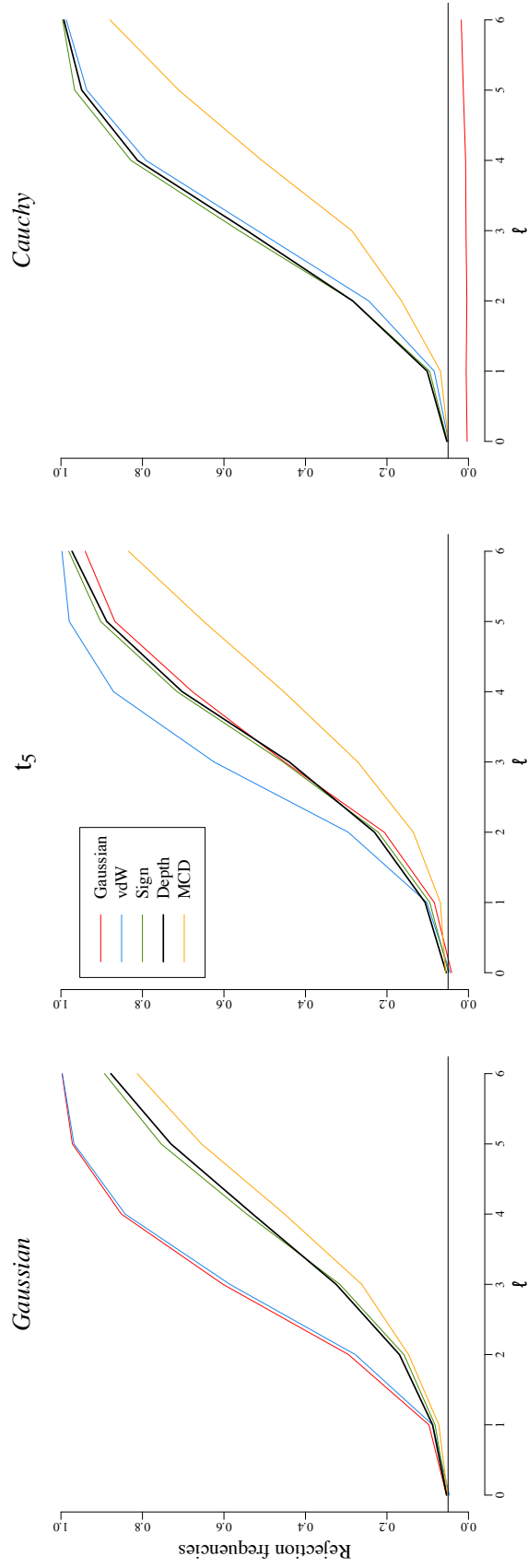
Figure 1: *Rejection frequencies, under bivariate normal (left), t_5 (middle), and Cauchy (right) elliptical densities, of the following tests: the proposed depth-based test of sphericity (in black), the Gaussian test (in red), the sign test (in green), the van der Waerden signed-rank test (in blue), and the MCD_γ-based test (in orange). Results are based on 3,000 replications and the sample size is $n = 500$. See Section 5 for details on the alternatives considered.*

(i) **(Non-uniform directional contamination)**: $\mathbf{Y}$ has the same distribution as the vector obtained by rotating $\mathbf{X}$ about the origin by an angle $\pi/4$ radiants.

(ii) **(Uniform directional contamination)**: $\mathbf{Y}$ has the same elliptical distribution as $\mathbf{X}$ but for the fact that its shape is $\mathbf{V} = \mathbf{I}_2$.

(iii) **(Uniform directional and radial contamination)**: $\mathbf{Y}$ is obtained by multiplying by four the random vector $\mathbf{Y}$ in the contamination pattern (ii).

In (i), the contamination is directional and typically shows up in the first eigendirection of $\mathbf{Y}$, that is, in the direction of the main bisector, whereas the original distribution $P^{\mathbf{X}}$ rather puts mass along the horizontal axis. The contamination pattern (ii) rather provides a directional contamination that is uniformly distributed over the unit circle. As for the last contamination pattern (iii), it combines the directional outlying feature of (ii) with a radial outlyingness.

For each combination of an elliptical density type (Gaussian, t_5, or Cauchy), a contamination pattern ((i)-(iii)), and an η-value among those given above, we generated $3,000$ corresponding independent random samples $\mathbf{X}_i^{(\eta)}$, $i = 1, \ldots, n = 200$. The resulting rejection frequencies of the five tests considered in the previous simulation are plotted in Figure 2 as functions of the contamination level η.

The results show the very good (level) robustness of the proposed depth-based test. In particular, it always dominates its sign-based competitor. Only the MCD-based test seems to dominate the proposed test in terms of robustness. At the sample size considered, the MCD-based test, however, is very liberal under heavy tails (see also Paindaveine & Van Bever, 2013) and, as we have seen in the first simulation, exhibits very low finite-sample powers. Finally, note that radial outliers appear to have a strong impact on both the Gaussian and van der Waerden tests.

6 Extensions to other scale functionals

As we explained in Section 2.1, different scale functionals S may be used to normalize scatter matrices $\boldsymbol{\Sigma}$ into shape matrices $\mathbf{V} = \boldsymbol{\Sigma}/S(\boldsymbol{\Sigma})$. Throughout the previous sections, we restricted to the determinant-based scale functional $S_{\det}(\boldsymbol{\Sigma}) = (\det \boldsymbol{\Sigma})^{1/d}$. This scale functional plays an important role in semiparametric inference on shape, since it is the only one that provides parameter-orthogonality between the resulting shape parameter $\mathbf{V}$ and scale parameter $\sigma^2 = S(\boldsymbol{\Sigma})$; see Paindaveine (2008). In some setups, however, it may be more suitable to work with scale functionals—such as, e.g., $S_{\text{one}}(\boldsymbol{\Sigma}) = \Sigma_{11}$ or $S_{\text{trace}}(\boldsymbol{\Sigma}) = (\operatorname{tr} \boldsymbol{\Sigma})/d$—for which the resulting shape matrices form an affine space. In this section, we discuss the construction of shape depth for an arbitrary scale functional satisfying Definition 2.1.

Of course, it is tempting to adopt, as we did in the particular case of $S_{\det}$, the tangent depth scheme from Definition 3.2, that is, to define the S-shape depth of $\mathbf{V}_S$, relative to a distribution $P_{\mathbf{X}}$, as $ShD_{\boldsymbol{\mu},\sigma_S,g,S}(\mathbf{V}_S, P_{\mathbf{X}}) = D_H\big(\mathbf{0}_D, P_{\nabla_{\mathbf{V}_S} \log L_{\boldsymbol{\mu},\sigma_S,\mathbf{V}_S;g}(\mathbf{X})}\big)$, where

$$\nabla_{\mathbf{V}_S} \log L_{\boldsymbol{\mu},\sigma_S,\mathbf{V}_S;g}(\mathbf{X}) = \frac{1}{2} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \operatorname{vec}\Big(\varphi_g\Big(\frac{d_{\boldsymbol{\mu},\mathbf{V}_S}}{\sigma_S}\Big) \frac{d_{\boldsymbol{\mu},\mathbf{V}_S}}{\sigma_S} \mathbf{U}_{\boldsymbol{\mu},\mathbf{V}_S} \mathbf{U}'_{\boldsymbol{\mu},\mathbf{V}_S} - \mathbf{I}_d\Big) \qquad (6.1)$$

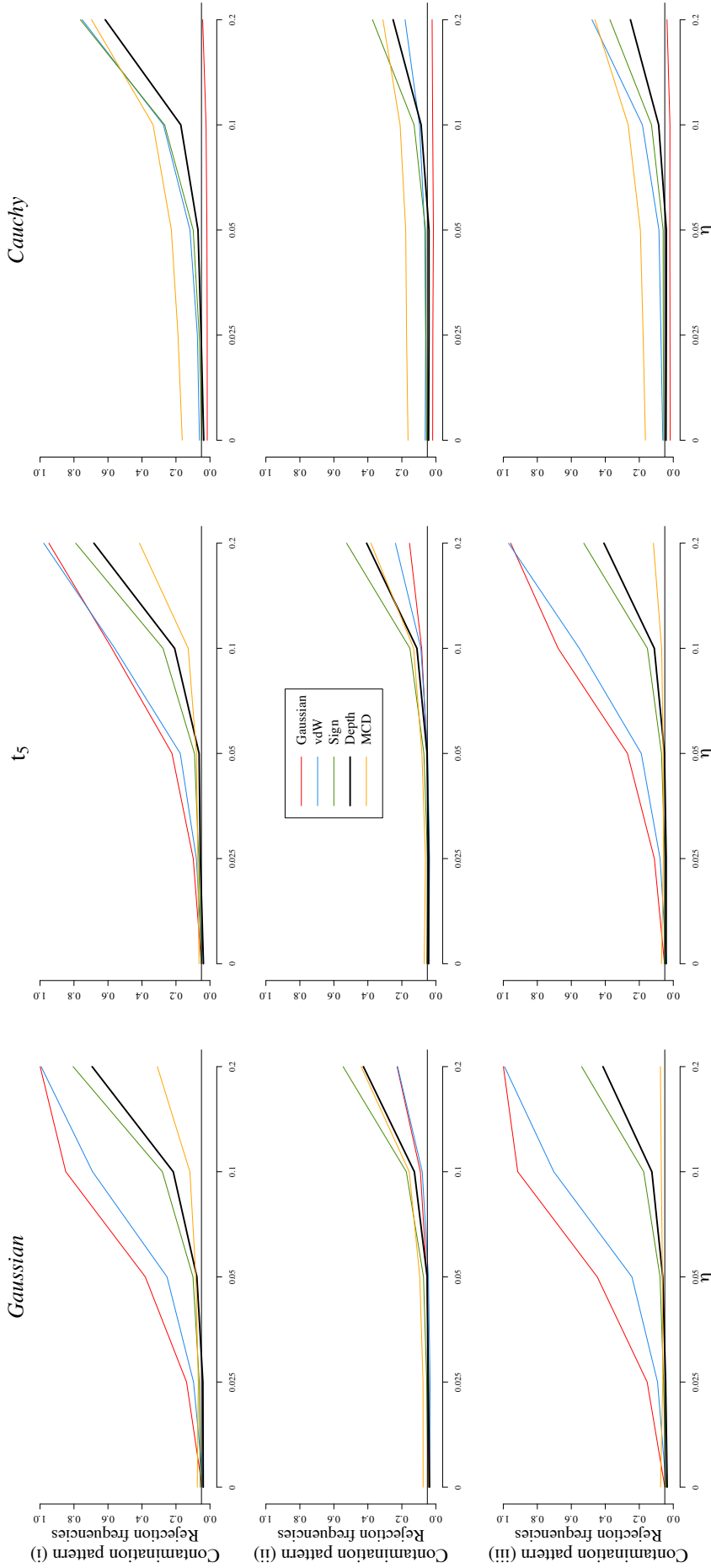is the generic score function for S-shape; see (2.2).

Figure 2: *Rejection frequencies, as a function of the level η at which the null hypothesis is contaminated, of the same five tests as in Figure 1, under bivariate normal (left), t_5 (middle), and Cauchy (right) elliptical densities. Three types of contaminations are considered : (i) non-uniform directional contamination (first row), (ii) uniform directional contamination (second row), and (iii) uniform directional and radial contamination (third row); see Section 5 for details. Results are based on 3,000 replications and the sample size is n = 200.*

As natural as this may be, the resulting S-shape depth concept is far to be as satisfactory as for the particular case $S_{\det}$ considered earlier. The reason is twofold. First, as already pointed out in Section 3, the shape depth $ShD_{\boldsymbol{\mu},\sigma_S,g,S}(\mathbf{V}_S, P_{\mathbf{X}})$ above not only depends on the location $\boldsymbol{\mu}$ but also on the scale σ_S and the radial density g. Second, and more importantly, the consistency property in Theorem 4.1 does not hold for an arbitrary scale functional, *even if one restricts to the collection of elliptical densities with the location $\boldsymbol{\mu}$, scale σ_S, and radial density g used to evaluate shape depth.*

We illustrate this inconsistency through the following bivariate example that involves the scale functional $S_{\text{trace}}(\boldsymbol{\Sigma}) = (\operatorname{tr}\boldsymbol{\Sigma})/2$. Let $\mathbf{X}$ be a bivariate normal random vector with mean $\boldsymbol{\mu}_0 = \mathbf{0}_2$, scale $\sigma_{0,S_{\text{trace}}} = \sigma_0 = 1$, Gaussian radial density $g = \phi$ and shape $\mathbf{V}_{0,S_{\text{trace}}} = \mathbf{V}_{a_0} = \mathbf{V}_{3/4}$, where we put $\mathbf{V}_a = \operatorname{diag}(a, 2 - a)$ for any $a \in (0,2)$. Consistency would imply that, for any $a \neq a_0$, $ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_a, P_{\mathbf{X}}) < ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_{a_0}, P_{\mathbf{X}})$. However, we have the following result (see the Appendix for the proof).

Lemma 6.1. *In the setup described above,*

$$ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_{a_0}, P_{\mathbf{X}}) < ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_1, P_{\mathbf{X}}).$$

For the sake of illustration, Figure 3 plots (estimated versions of) $ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_a, P_{\mathbf{X}})$ and $ShD_{\boldsymbol{\mu}_0,S_{\det}}(\mathbf{V}_a/S_{\det}(\mathbf{V}_a), P_{\mathbf{X}})$, as functions of a. Estimations were obtained as follows: we generated $M = 100$ mutually independent random samples from the distribution $P_{\mathbf{X}}$ considered in Lemma 6.1. Then, for every value $a_i = i/100$, with $i = 1, \ldots, 199$, we estimated $ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_a, P_{\mathbf{X}})$ and $ShD_{\boldsymbol{\mu}_0,S_{\det}}(\mathbf{V}_a/S_{\det}(\mathbf{V}_a), P_{\mathbf{X}})$ by averaging over the $M = 100$ samples available the respective sample depths (sample S_{trace}-shape depth is obtained from his population version in the exact same way as for the $S_{\det}$-shape depth in Section 5). Figure 3 shows that S-shape depth crucially depends on the scale functional used, and confirms that the consistency result from Theorem 4.1 does not hold for the scale functional S_{trace}.
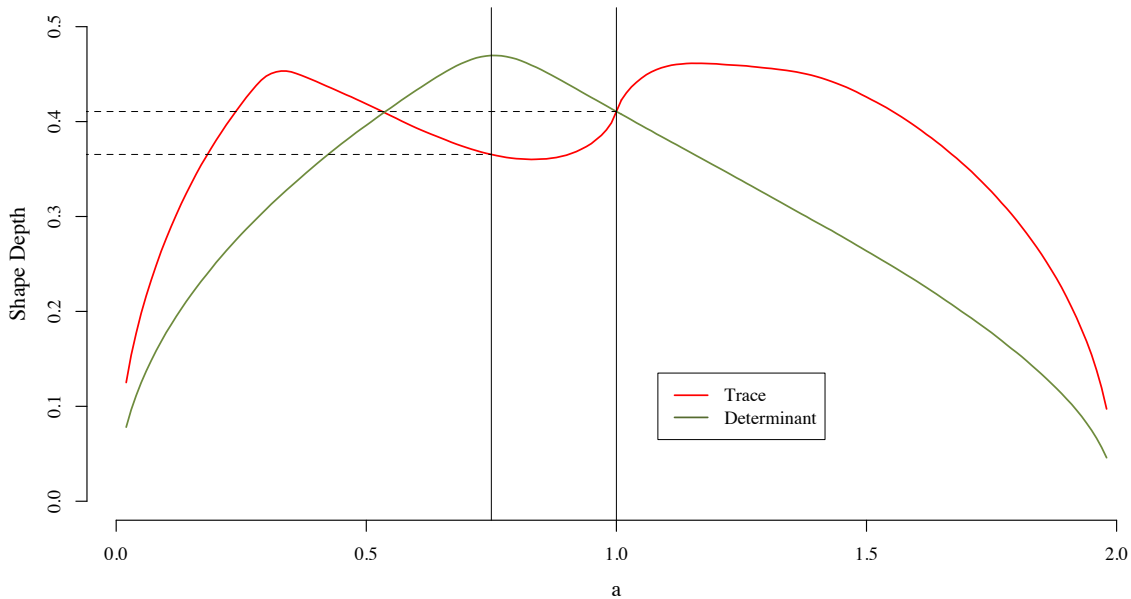

Figure 3: *Plots of (estimated versions of) $ShD_{S_{\text{trace}}}(\mathbf{V}_a, P_{\mathbf{X}})$ (in red) and $ShD_{S_{\det}}(\mathbf{V}_a/S_{\det}(\mathbf{V}_a), P_{\mathbf{X}})$ (in green) as functions of $a \in (0,2)$, where $\mathbf{X}$ is elliptical with S_{trace}-shape $\mathbf{V}_{a_0}$, $a_0 = 3/4$ (hence also $S_{\det}$-shape $\mathbf{V}_{a_0}/S_{\det}(\mathbf{V}_{a_0})$).*

This dependence of S-shape depth on the scale functional S—but also its dependence on the underlying scale and radial density—makes it desirable to define an alternative shape depth concept, that would be as satisfactory for an arbitrary scale functional S as the one we introduced in Section 3.2 for the scale functional $S_{\det}$. Quite interestingly, the appropriate S-shape depth concept may be obtained by replacing scores for S-shape in (6.2) by their efficient (in the semiparametric sense) versions

$$\nabla^*_{\mathbf{V}_S} \log L_{\boldsymbol{\mu},\sigma_S,\mathbf{V}_S;g}(\mathbf{X}) = \frac{1}{2}\, \varphi_g\Big(\frac{d_{\boldsymbol{\mu},\mathbf{V}_S}}{\sigma_S}\Big)\frac{d_{\boldsymbol{\mu},\mathbf{V}_S}}{\sigma_S}\,\mathbf{M}^{\mathbf{V}_S}_S(\mathbf{V}^{\otimes 2}_S)^{-1/2}\,\mathrm{vec}\Big(\mathbf{U}_{\boldsymbol{\mu},\mathbf{V}_S}\mathbf{U}'_{\boldsymbol{\mu},\mathbf{V}_S} - \frac{1}{d}\mathbf{I}_d\Big); \quad (6.2)$$

see (9) in Paindaveine (2008). The same arguments as above Definition 3.3 then allow to get rid of the scalar factor involving Mahalanobis distance $d_{\boldsymbol{\mu},\mathbf{V}_S}/\sigma_S$, which finally leads to the following definition.

Definition 6.1. For any $\mathbf{V}_S \in \mathcal{V}^S_d$, the *efficient shape depth* of $\mathbf{V}_S$, with respect to the distribution $P = P_{\mathbf{X}}$ on $\mathbb{R}^d$ having known deepest point $\boldsymbol{\mu}$, is $ShD^*_{\boldsymbol{\mu},S}(\mathbf{V}_S, P_{\mathbf{X}}) = D_H\big(\mathbf{0}_D, P_{\mathbf{W}^{*\boldsymbol{\mu}},\mathbf{V}_S}\big)$, where $\mathbf{W}^{*\boldsymbol{\mu},\mathbf{V}_S} = \mathbf{M}^{\mathbf{V}_S}_S(\mathbf{V}^{\otimes 2}_S)^{-1/2}\,\mathrm{vec}\big(\mathbf{U}_{\boldsymbol{\mu},\mathbf{V}_S}\mathbf{U}'_{\boldsymbol{\mu},\mathbf{V}_S} - \frac{1}{d}\mathbf{I}_d\big)$.

Parallel to $S_{\det}$-shape depth, efficient S-shape depth, for any given scale functional S, does not require knowing the scale σ_S nor the radial density g, but only the location $\boldsymbol{\mu}$. Actually, Definitions 3.3 and 6.1 appear to be strictly the same definitions. Note however that the matrix $\mathbf{M}^{\mathbf{V}_S}_S$ depends on the scale functional S, so that it might be so that, despite this strong similarity between both definitions, efficient S-shape depth still crucially depends on the scale functional S. The following result shows that efficient S-shape depth, on the contrary, is a concept that does not depend on S (see the Appendix for the proof).

Theorem 6.1. Let S_1 and S_2 be two scale functionals. Then, for any probability distribution $P = P^{\mathbf{X}}$ (with known location $\boldsymbol{\mu}$ and any $\boldsymbol{\Sigma} \in \mathcal{S}_d$),

$$ShD^*_{\boldsymbol{\mu},S_1}\Big(\frac{\boldsymbol{\Sigma}}{S_1(\boldsymbol{\Sigma})}, P\Big) = ShD^*_{\boldsymbol{\mu},S_2}\Big(\frac{\boldsymbol{\Sigma}}{S_2(\boldsymbol{\Sigma})}, P\Big).$$

As a corollary, our results on affine-invariance (Theorem 3.1) and consistency (Theorem 4.1)—but also more minor results such as Lemmas 4.2, 4.3 and 4.4—extend immediately to efficient S-shape depth for an arbitrary scale functional S. This is in sharp contrast with the original—parametric, Mizera (2002)-type—S-shape depth concept introduced earlier in this section. It directly follows from Paindaveine (2008) that $S_{\det}$ is the only scale functional for which the original S-shape $ShD_S(\mathbf{V}, P_{\mathbf{X}})$ coincides with its efficient version $ShD^*_S(\mathbf{V}, P_{\mathbf{X}})$. This explains why we started by defining shape depth for this particular scale functional.

As we showed above, replacing parametric scores by semiparametric (efficient) ones is needed—unless the scale functional $S_{\det}$ is adopted—to obtain a shape depth concept that achieves consistency (in the sense of Theorem 4.1). This provides a *semiparametric construction of depth* that extends the parametric one from Mizera (2002) and, to the best of our knowledge, is original. This extension is needed as soon as the semiparametric model at hand is not adaptive. All semiparametric models where the parametric Mizera (2002) depth had been used so far—namely, regression (Rousseeuw & Hubert, 1999), location-scale (Mizera & Müller, 2004), etc.—are adaptive, which may partly explain why the semiparametric depth we are introducing here was not considered before. Of course, it would be of interest to consider other instances of semiparametric depth, and to develop a general theory.

7   Final comments

In the previous sections, shape depth was defined in the $\boldsymbol{\mu}$-specified case only, and inferential applications focused on hypothesis testing. Here, we shortly comment on point estimation and on the extension to the $\boldsymbol{\mu}$-unspecified case.

Regarding point estimation, Theorem 4.1—or, more precisely, its extension[4] to an arbitrary scale functional S—suggests defining the depth-based estimator as

$$\hat{\mathbf{V}}_S^{(n)} = \underset{\mathbf{V} \in \mathcal{V}_d^S}{\arg \max} \; ShD_{\boldsymbol{\mu},S}^*(\mathbf{V}, P^{(n)}), \tag{7.1}$$

where $P^{(n)}$ stands, as usual, for the empirical distribution of the d-variate observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ at hand. The so-called *maxdepth* approach in (7.1) is quite classical; see, e.g., Donoho (1982), Rousseeuw & Hubert (1999), and Mizera & Müller (2004), for location, linear regression, and location-scale, respectively.

The exact properties of $\hat{\mathbf{V}}_S^{(n)}$ (consistency, asymptotic distribution, robustness, etc.) remain to be explored. Also, computational aspects are non-trivial, even for $d = 2$ (a case to which we now restrict for the sake of illustration). If the optimization in (7.1) is to be performed (in an approximate way) by running over a fine grid of the parameter space, then it may seem more convenient to work with the trace-based normalization of shape than with the determinant-based one; indeed, the former leads to a parametric space of the form

$$\begin{aligned}
\mathcal{V}_2^{S_{\text{trace}}} &= \left\{ \begin{pmatrix} a & b \\ b & 2-a \end{pmatrix} \in \mathbb{R}^{2 \times 2} \,\middle|\, a \in (0, 2), a(2 - a) - b^2 > 0 \right\} \\
&\cong \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, (a - 1)^2 + b^2 < 1 \right\} = \underline{\mathcal{V}}_2^{S_{\text{trace}}},
\end{aligned}$$

which is bounded, unlike the parametric space

$$\begin{aligned}
\mathcal{V}_2^{S_{\text{det}}} &= \left\{ \begin{pmatrix} a & c \\ c & b \end{pmatrix} \in \mathbb{R}^{2 \times 2} \,\middle|\, ab - c^2 = 1, a > 0, b > 0 \right\} \\
&\cong \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, a > 0, b \in \mathbb{R} \right\} = \underline{\mathcal{V}}_2^{S_{\text{det}}}
\end{aligned}$$

obtained for the latter. Now, it may be so that running over a grid may be avoided by exploiting the possible quasi-concavity of shape depth, that would result into convex shape depth regions. Figure 4 illustrates this possible quasi-concavity, both for the determinant- and trace-based normalizations, by providing heatplots of the corresponding sample depth functions computed from a random sample of $n = 1,000$ bivariate standard normal observations (note that for non-linear scale functionals such as S_{det}, one needs to define what is exactly meant with quasi-concavity).

Turning to the $\boldsymbol{\mu}$-unspecified case, a naive approach consists in replacing in Definition 3.3 the unknown value of $\boldsymbol{\mu}$ with its maxdepth estimator $\hat{\boldsymbol{\mu}}^{(n)} = \arg \max_{\boldsymbol{\mu}} D_H(\boldsymbol{\mu}, P^{(n)})$. This is current practice in semiparametric inference for shape, as it is well-known that parameter-orthogonality between $\boldsymbol{\mu}$ and $\mathbf{V}$ implies that this plug-in strategy does not affect the asymptotic behavior of *likelihood-type* inference procedures for shape. Since there is no guarantee that this also holds for *depth-based* inference procedures for shape, it seems safer to directly go for a joint estimation of $(\boldsymbol{\mu}, \mathbf{V})$ through *location-shape* depth, very much in the spirit of the location-scale depth approach from Mizera & Müller (2004).

---

[4]In the previous section, we showed that this extension holds provided that efficient shape depth is considered.
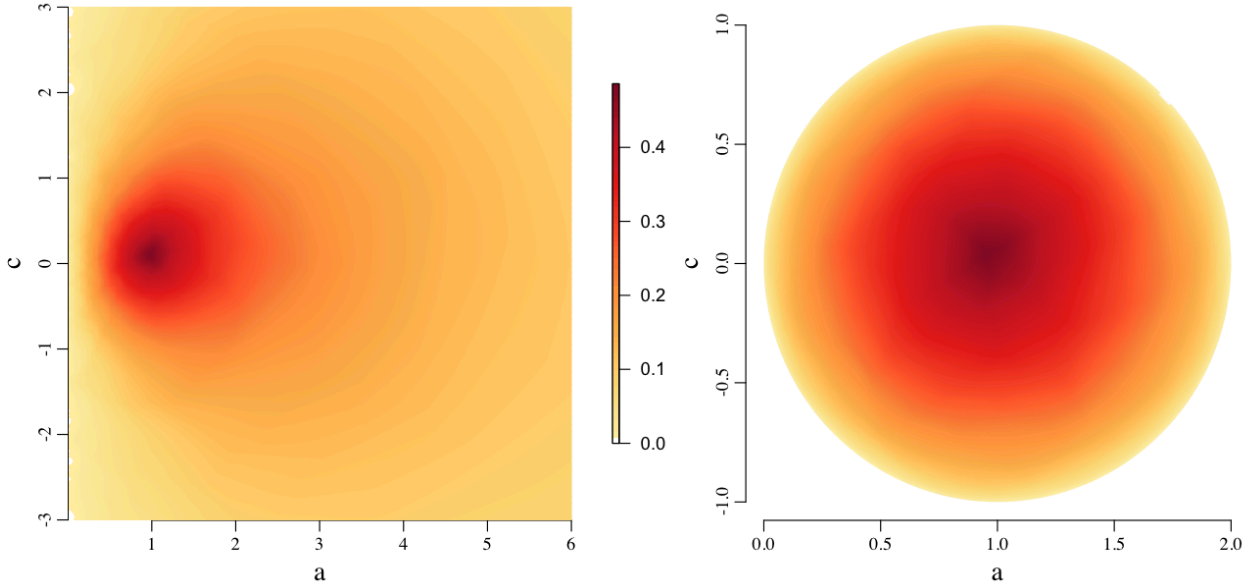
Figure 4: *Heatplots of the shape depth functions, computed from a random sample of $n = 1,000$ bivariate standard normal observations, for the parameter spaces $\underline{\mathcal{V}}_2^{S_{\text{trace}}}$ (left) and $\underline{\mathcal{V}}_2^{S_{\text{det}}}$ (right).*

The discussion related to semiparametric depth in the previous section makes it natural to define location-shape depth on the basis of the efficient score function for $(\boldsymbol{\mu}, \mathbf{V}_S)$, that is of the form

$$
\nabla^*_{\boldsymbol{\mu}, \mathbf{V}_S} \log L_{\boldsymbol{\mu}, \sigma_S, \mathbf{V}_S; g}(\mathbf{X}) = \begin{pmatrix} \frac{1}{\sigma_S} \varphi_g\left(\frac{d_{\boldsymbol{\mu}, \mathbf{V}_S}}{\sigma_S}\right) \mathbf{V}_S^{-1/2} \mathbf{U}_{\boldsymbol{\mu}, \mathbf{V}_S} \\ \frac{1}{2} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \varphi_g\left(\frac{d_{\boldsymbol{\mu}, \mathbf{V}_S}}{\sigma_S}\right) \frac{d_{\boldsymbol{\mu}, \mathbf{V}_S}}{\sigma_S} \text{vec}\left(\mathbf{U}_{\boldsymbol{\mu}, \mathbf{V}_S} \mathbf{U}'_{\boldsymbol{\mu}, \mathbf{V}_S} - \frac{1}{d} \mathbf{I}_d\right) \end{pmatrix};
$$
(7.2)

see Paindaveine (2008). This leads to the following definition.

Definition 7.1. For any $\boldsymbol{\mu} \in \mathbb{R}^d$ and any $\mathbf{V}_S \in \mathcal{V}_d^S$, the *location-shape depth* of $(\boldsymbol{\mu}, \mathbf{V}_S)$, with respect to the distribution $P = P_{\mathbf{X}}$ on $\mathbb{R}^d$, is

$$
LShD\big((\boldsymbol{\mu}, \mathbf{V}_S), P\big) = D_H\Big(\mathbf{0}_{d+D}, P_{\nabla^*_{\boldsymbol{\mu}, \mathbf{V}_S} \log L_{\boldsymbol{\mu}, \sigma_S, \mathbf{V}_S; g}(\mathbf{X})}\Big),
$$

where $\nabla^*_{\boldsymbol{\mu}, \mathbf{V}_S} \log L_{\boldsymbol{\mu}, \sigma_S, \mathbf{V}_S; g}(\mathbf{X})$ is the efficient score in (7.2).

Provided that g is monotone strictly decreasing, location-scale depth does not depend on σ_S nor on g, since one may then get rid of the positive scalar factor $(1/\sigma_S)\varphi_g(d_{\boldsymbol{\mu}, \mathbf{V}_S}/\sigma_S)$. Removal of this factor, however, does not completely eliminate the Mahalanobis distance $d_{\boldsymbol{\mu}, \mathbf{V}_S}$ from the above efficient score, so that, in contrast with shape depth, location-schape depth is not a sign concept. Of course, properties of location-shape depth remain to be explored.

Acknowledgments

Appendix — Proofs

This appendix collects the proofs of all theorems, propositions, and lemmas stated in the manuscript. We start with the key lemma from Section 4.

Proof of Lemma 4.1. (i) Letting $\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} = (\mathbf{V}^{\otimes 2})^{-1/2}\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}$, we first show that

$$ShD_{\boldsymbol{\mu}}(\mathbf{V},P) \quad = \quad \inf_{\boldsymbol{w}\in\mathbb{R}^D} P\big[(\mathbf{M}_S^{\mathbf{V}}\boldsymbol{w})'\,\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} \geq 0\big] \tag{3}$$

$$= \quad \inf_{\tilde{\boldsymbol{w}}\in\mathbb{R}^{d^2}} P\big[\tilde{\boldsymbol{w}}'\,\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} \geq 0\big] \tag{4}$$

$$= \quad D_H(\mathbf{0}_{d^2},P_{\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}}). \tag{5}$$

Since (3) and (5) merely hold by definition, it is sufficient to prove (4). Fix then $\tilde{\boldsymbol{w}} \in \mathbb{R}^{d^2}$ and let us show that there exists a D-vector $\boldsymbol{w}$ such that

$$P\big[(\mathbf{M}_S^{\mathbf{V}}\boldsymbol{w})'\,\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} \geq 0\big] = P\big[\tilde{\boldsymbol{w}}'\,\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} \geq 0\big]. \tag{6}$$

Let $\mathbf{A}_{\tilde{\boldsymbol{w}}}$ be the $d \times d$ matrix defined through $\tilde{\boldsymbol{w}} = (\text{vec}\,\mathbf{A}_{\tilde{\boldsymbol{w}}})$. Without loss of generality, we may assume that $\mathbf{A}_{\tilde{\boldsymbol{w}}}$ is symmetric : since $\mathbf{W}^{\boldsymbol{\mu},\mathbf{V}} = \text{vec}\big(\mathbf{V}^{-1}\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}}(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}})'\mathbf{V}^{-1} - \frac{1}{d}\mathbf{V}^{-1}\big)$ is the vec of a symmetric matrix, we indeed have $(\text{vec}\,\mathbf{A}_{\tilde{\boldsymbol{w}}})'\,\mathbf{W}^{\boldsymbol{\mu},\mathbf{V}} = \big(\text{vec}(\mathbf{A}_{\tilde{\boldsymbol{w}}} + \mathbf{A}_{\tilde{\boldsymbol{w}}}')/2\big)'\,\mathbf{W}^{\boldsymbol{\mu},\mathbf{V}}$.

Since $(\text{vec}\,\mathbf{V})'\mathbf{W}^{\boldsymbol{\mu},\mathbf{V}} = 0$, we may then rewrite $\tilde{\boldsymbol{w}}'\,\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}$ in (6) as

$$\tilde{\boldsymbol{w}}'\,\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} = (\text{vec}\,\mathbf{A}_{\tilde{\boldsymbol{w}}})'\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} = (\text{vec}\,(\mathbf{A}_{\tilde{\boldsymbol{w}}} - c\mathbf{V}))'\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}, \tag{7}$$

where we put $c := (\nabla S(\text{vech}\,\mathbf{V}))'(\text{vech}\,\mathbf{A}_{\tilde{\boldsymbol{w}}})$. Now, differentiating with respect to λ both sides of the identity $S(\lambda(\text{vech}\,\mathbf{V})) = \lambda$ (which follows from homogeneity of the scale functional[5] S), we obtain $(\text{vech}\,\mathbf{V})'\nabla S(\text{vech}\,\mathbf{V}) = 0$, which yields

$$(\nabla S(\text{vech}\,\mathbf{V}))'\,\text{vech}(\mathbf{A}_{\tilde{\boldsymbol{w}}} - c\mathbf{V}) = 0.$$

The definition of $\mathbf{M}_S^{\mathbf{V}}$ then implies that $\text{vec}(\mathbf{A}_{\tilde{\boldsymbol{w}}} - c\mathbf{V}) = (\mathbf{M}_S^{\mathbf{V}})'\,\overset{\circ}{\text{vech}}(\mathbf{A}_{\tilde{\boldsymbol{w}}} - c\mathbf{V})$, so that $\boldsymbol{w} = \overset{\circ}{\text{vech}}(\mathbf{A}_{\tilde{\boldsymbol{w}}} - c\mathbf{V})$ is a D-vector that satisfies (6). This shows that the infimum in (3) is smaller than or equal to the infimum in (6). Since the reverse equality trivially holds, this establishes (6). Finally, affine-invariance of halfspace depth directly yields that $ShD_{\boldsymbol{\mu}}(\mathbf{V},P) = D_H(\mathbf{0}_{d^2},P_{\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}}) = D_H(\mathbf{0}_{d^2},P_{\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}})$.

(ii) Since $\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}$ and $\overline{\mathbf{W}}_{D+1}^{\boldsymbol{\mu},\mathbf{V}}$ are the vec and vech of a common $d \times d$ symmetric matrix, there exists a full rank $d^2 \times (D+1)$ matrix $\mathbf{P}$ such that $\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} = \mathbf{P}\overline{\mathbf{W}}_{D+1}^{\boldsymbol{\mu},\mathbf{V}}$. As $\mathbf{P}'$ has full column rank, we then obtain

$$D_H(\mathbf{0}_{d^2},P_{\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}}) \quad = \quad \inf_{\boldsymbol{w}\in\mathbb{R}^{d^2}} P\big[\boldsymbol{w}'\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} \geq 0\big]$$

$$= \quad \inf_{\boldsymbol{w}\in\mathbb{R}^{d^2}} P\big[(\mathbf{P}'\boldsymbol{w})'\overline{\mathbf{W}}_{D+1}^{\boldsymbol{\mu},\mathbf{V}} \geq 0\big]$$

$$= \quad D_H(\mathbf{0}_{D+1},P_{\overline{\mathbf{W}}_{D+1}^{\boldsymbol{\mu},\mathbf{V}}}).$$

---

[5] Recall that, whenever partial are computed S is considered as a function of $\text{vech}(\mathbf{V})$ rather that $\mathbf{V}$.

(iii) The equality to be proved similarly follows from the existence of a full rank $d^2 \times D$ matrix $\mathbf{Q}$ such that $\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} = \mathbf{Q}\overline{\mathbf{W}}_D^{\boldsymbol{\mu},\mathbf{V}}$. This comes from the fact that $(\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}})_1 = U_1^2 - (1/d)$ may be written as $(\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}})_1 = -\sum_{j=2}^{d}(U_j^2 - (1/d)) = -\sum_{j=2}^{d}(\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}})_j$. $\qquad\square$

Proofs of theorems 3.1 and 6.1 are now mere consequences of the Lemma 4.1.

Proof of Theorem 3.1. Let $\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}(\mathbf{A}, \boldsymbol{b})$ denote the random vector $\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} = (\mathbf{V}^{\otimes 2})^{-1/2}\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}$ computed from the shape matrix $\mathbf{AVA}'/S(\mathbf{AVA}')$ and the random observation $\mathbf{AX} + \boldsymbol{b}$. Since standard properties of the vec operator and the Kronecker product yield

$$
\begin{aligned}
\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} &= (\mathbf{V}^{\otimes 2})^{-1/2}\mathrm{vec}\Big(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}}(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}})' - \frac{1}{d}\mathbf{I}_d\Big) \\
&= \mathrm{vec}\Big(\mathbf{V}^{-1/2}\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}}(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}})'\mathbf{V}^{-1/2} - \frac{1}{d}\mathbf{V}^{-1}\Big) \\
&= \mathrm{vec}\Big(\frac{\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\mathbf{V}^{-1}}{(\mathbf{X}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})} - \frac{1}{d}\mathbf{V}^{-1}\Big),
\end{aligned}
$$

we readily obtain

$$
\begin{aligned}
\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}(\mathbf{A}, \boldsymbol{b}) &= S(\mathbf{AVA}')\,\mathrm{vec}\Big(\mathbf{A}'^{-1}\Big[\frac{\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\mathbf{V}^{-1}}{(\mathbf{X}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})} - \frac{1}{d}\mathbf{V}^{-1}\Big]\mathbf{A}^{-1}\Big) \\
&= S(\mathbf{AVA}')\,(\mathbf{A}\otimes\mathbf{A})^{-1}\,\mathrm{vec}\Big(\frac{\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\mathbf{V}^{-1}}{(\mathbf{X}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})} - \frac{1}{d}\mathbf{V}^{-1}\Big) \\
&= S(\mathbf{AVA}')\,(\mathbf{A}\otimes\mathbf{A})^{-1}\,\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}.
\end{aligned}
$$

The result then follows from the identity $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P) = D_H(\mathbf{0}_{d^2}, P_{\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}})$ (Lemma 4.1) and from affine-invariance of halfspace depth $\qquad\square$

Proof of Theorem 6.1. Let $\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}_{S_i}}$, $i = 1, 2$, denote the random vectors $\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}} = (\mathbf{V}^{\otimes 2})^{-1/2}\overline{\mathbf{W}}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}$ evaluated at $\mathbf{V}_{S_i} = \boldsymbol{\Sigma}/S_i(\boldsymbol{\Sigma})$, $i = 1, 2$. Clearly, $\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}_{S_1}} = \mathbf{U}^{\boldsymbol{\mu},\mathbf{V}_{S_2}}$ and $\mathbf{V}_{S_1} = (S_2(\boldsymbol{\Sigma})/S_1(\boldsymbol{\Sigma}))\mathbf{V}_{S_2}$, which implies that

$$
\begin{aligned}
\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}_{S_1}} &= (\mathbf{V}_{S_1}^{\otimes 2})^{-1/2}\,\mathrm{vec}\Big(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}_{S_1}}(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}_{S_1}})' - \frac{1}{d}\mathbf{I}_d\Big) \\
&= \frac{S_1(\boldsymbol{\Sigma})}{S_2(\boldsymbol{\Sigma})}(\mathbf{V}_{S_1}^{\otimes 2})^{-1/2}\,\mathrm{vec}\Big(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}_{S_2}}(\mathbf{U}^{\boldsymbol{\mu},\mathbf{V}_{S_2}})' - \frac{1}{d}\mathbf{I}_d\Big) = \frac{S_1(\boldsymbol{\Sigma})}{S_2(\boldsymbol{\Sigma})}\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}_{S_2}}.
\end{aligned}
$$

As in the proof of Theorem 3.1, the result then follows from the identity $ShD_{\boldsymbol{\mu}}(\mathbf{V}, P) = D_H(\mathbf{0}_{d^2}, P_{\mathbf{W}_{d^2}^{\boldsymbol{\mu},\mathbf{V}}})$ (Lemma 4.1) and from the affine-invariance of halfspace depth. $\qquad\square$

We now establish Theorem 4.1 by proving Lemmas 4.2, 4.3, and 4.4.

Proof of Lemma 4.2. In this proof, we write $\mathbf{U}$ and $\overline{\mathbf{W}}_{D+1}$ for $\mathbf{U}^{\boldsymbol{\mu},\mathbf{I}_d}$ and $\overline{\mathbf{W}}_{D+1}^{\boldsymbol{\mu},\mathbf{I}_d}$, respectively. Actually, instead of the original $\overline{\mathbf{W}}_{D+1}$, we will work with

$$
\widetilde{\mathbf{W}}_{D+1} = (U_1^2 - 1/2, U_2^2 - 1/2, U_1 U_2)',
$$

for $d = 2$, and with

$$\widetilde{\mathbf{W}}_{D+1} = (U_1^2 - 1/3, U_2^2 - 1/3, U_3^2 - 1/3, U_1 U_2, U_2 U_3, U_1 U_3)',$$

for $d = 3$. The new vectors $\widetilde{\mathbf{W}}_{D+1}$ are obtained from the original ones by permuting their columns, hence lead to the same halfspace depth of $\mathbf{0}_{D+1}$.

Both for $d = 2$ and $d = 3$, there exist a $(D + 1) \times d$ matrix $\mathbf{A}$, and a $(D + 1) \times (D + 1)$ invertible matrix $\mathbf{B}$ such that $\mathbf{B}\widetilde{\mathbf{W}}_{D+1} = (\mathbf{A}\mathbf{U}) \odot (\mathbf{A}\mathbf{U}) - \frac{1}{d}\mathbf{1}_{D+1}$, where $\odot$ denotes the Hadamard (i.e., entrywise) vector product ; for $d = 2$, the $\mathbf{A}$ and $\mathbf{B}$ matrices are given by

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \mathbf{I}_3,$$

while, for $d = 3$, one has

$$\mathbf{A} = \begin{pmatrix} & \mathbf{I}_3 & \\ -1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{3} & -1/\sqrt{3} & -1/\sqrt{3} \end{pmatrix} \quad \text{and } \mathbf{B} = \begin{pmatrix} \mathbf{I}_3 & & \mathbf{0}_{3\times 3} \\ & 1/3 & -1/3 & -1/3 \\ \mathbf{0}_{3\times 3} & -1/3 & 1/3 & -1/3 \\ & -1/3 & -1/3 & 1/3 \end{pmatrix}.$$

Now, for any $\boldsymbol{w} = (w_1, \ldots, w_{D+1})' \in \mathcal{D}_{D+1}$,

$$
\begin{aligned}
P[\boldsymbol{w}'\mathbf{B}\widetilde{\mathbf{W}}_{D+1} \geq 0] &= P\Big[\boldsymbol{w}'\big((\mathbf{A}\mathbf{U}) \odot (\mathbf{A}\mathbf{U}) - (1/d)\mathbf{1}_{D+1}\big) \geq 0\Big] \\
&= P\Big[\sum_{i=1}^{D+1} w_i((\mathbf{A}\mathbf{U})_i)^2 \geq (1/d)\boldsymbol{w}'\mathbf{1}_{D+1}\Big] \\
&= P\big[\mathbf{U}'(\mathbf{A}'\mathrm{diag}(\boldsymbol{w})\mathbf{A})\mathbf{U} \geq 1/d\big],
\end{aligned}
$$

where $\mathrm{diag}(\boldsymbol{w})$ denotes the $(D+1) \times (D+1)$ diagonal matrix with diagonal entries $w_1, \ldots, w_{D+1}$. Factorizing $\mathbf{A}'\mathrm{diag}(\boldsymbol{w})\mathbf{A}$ into $\mathbf{O}\mathrm{diag}(\boldsymbol{\lambda}^{\boldsymbol{w}})\mathbf{O}'$, where $\mathbf{O}$ is orthogonal and $\boldsymbol{\lambda}^{\boldsymbol{w}} = (\lambda_1^{\boldsymbol{w}}, \ldots, \lambda_d^{\boldsymbol{w}})'$, then yields (recall that $\mathbf{U}$ is spherically symmetric)

$$P[\boldsymbol{w}'\mathbf{B}\widetilde{\mathbf{W}}_{D+1} \geq 0] = P\big[\mathbf{U}'\mathrm{diag}(\boldsymbol{\lambda}^{\boldsymbol{w}})\mathbf{U} \geq 1/d\big] = P\Big[\sum_{i=1}^{d} \lambda_i^{\boldsymbol{w}} U_i^2 \geq 1/d\Big].$$

It follows that

$$
\begin{aligned}
ShD_{\boldsymbol{\mu}}(\mathbf{I}_d, P) &= \inf_{\boldsymbol{w} \in \mathbb{R}^{D+1}} P[\boldsymbol{w}'\widetilde{\mathbf{W}}_{D+1} \geq 0] = \inf_{\boldsymbol{w} \in \mathbb{R}^{D+1}} P[\boldsymbol{w}'\mathbf{B}\widetilde{\mathbf{W}}_{D+1} \geq 0] \\
&= \inf_{\boldsymbol{w} \in \mathcal{D}_{D+1}} P[\boldsymbol{w}'\mathbf{B}\widetilde{\mathbf{W}}_{D+1} \geq 0] = \inf_{\boldsymbol{w} \in \mathcal{D}_{D+1}} P\Big[\sum_{i=1}^{d} \lambda_i^{\boldsymbol{w}} U_i^2 \geq 1/d\Big]. \quad (8)
\end{aligned}
$$

A direct computation shows that $\mathbf{A}'\mathrm{diag}(\boldsymbol{w})\mathbf{A}$ has trace one, implying that $\boldsymbol{\lambda}^{\boldsymbol{w}} \in \mathcal{D}_d$, which in turn

shows that

$$ShD_{\boldsymbol{\mu}}(\mathbf{I}_d, P) = \inf_{\boldsymbol{w} \in \mathcal{D}_{D+1}} P\Big[\sum_{i=1}^{d} \lambda_i^{\boldsymbol{w}} U_i^2 \geq 1/d\Big] \geq \inf_{\boldsymbol{\lambda} \in \mathcal{D}_d} P\Big[\sum_{i=1}^{d} \lambda_i U_i^2 \geq 1/d\Big]. \tag{9}$$

Since, for any $\boldsymbol{\lambda} \in \mathcal{D}_d$, taking $\boldsymbol{w} = (\boldsymbol{\lambda}', \mathbf{0}'_{D+1-d})'$ yields $\boldsymbol{\lambda}^{\boldsymbol{w}} = \boldsymbol{\lambda}$, the reverse inequality in (9) also holds. This establishes the result. □

In order to prove Lemma 4.3, we will need the following lemma about depth.

Lemma 1. *Let* $\mathbf{X}$ *be a random vector supported on an hyperplane* $\Pi_1 \subset \mathbb{R}^d$. *Fix* $\boldsymbol{\vartheta} \in \Pi_1$ *and let* $T : \mathbb{R}^d \to \mathbb{R}^d$ *denote the projection on the hyperplane* Π_2, $\Pi_2 \not\perp \Pi_1$. *Then* $D_H(\boldsymbol{\vartheta}, P_{\mathbf{X}}) = D_H(T(\boldsymbol{\vartheta}), P_{T(\mathbf{X})})$.

Proof of Lemma 1. Let $\boldsymbol{\pi}_\ell$, $\ell = 1, 2$, be unit d-vectors orthogonal to Π_ℓ, $\ell = 1, 2$, respectively. For any $\boldsymbol{u} \in \mathcal{S}^{d-1}$, let $H_{\boldsymbol{\vartheta}, \boldsymbol{u}} = \{\boldsymbol{x} | \boldsymbol{u}'(\boldsymbol{x} - \boldsymbol{\vartheta}) \geq 0\}$. Then

$$D_H(\boldsymbol{\vartheta}, P_{\mathbf{X}}) = \inf_{\boldsymbol{u} \in \mathcal{S}^{d-1}} P_{\mathbf{X}}[H_{\boldsymbol{\vartheta}, \boldsymbol{u}}] = \inf_{\boldsymbol{u} \in \mathcal{S}^{d-1} \setminus \{\boldsymbol{\pi}_1\}} P_{\mathbf{X}}[H_{\boldsymbol{\vartheta}, \boldsymbol{u}}],$$

since $P_{\mathbf{X}}[H_{\boldsymbol{\vartheta}, \boldsymbol{\pi}_1}] = 1$. Would there exist a bijective (hence invertible) function $\boldsymbol{v} : \mathcal{S}^{d-1} \setminus \{\boldsymbol{\pi}_1\} \to \mathcal{S}^{d-1} \setminus \{\boldsymbol{\pi}_2\} : \boldsymbol{u} \mapsto \boldsymbol{v}(\boldsymbol{u})$, such that $P_{\mathbf{X}}[H_{\boldsymbol{\vartheta}, \boldsymbol{u}}] = P_{T(\mathbf{X})}[H_{T(\boldsymbol{\vartheta}), \boldsymbol{v}(\boldsymbol{u})}]$ for all $\boldsymbol{u} \in \mathcal{S}^{d-1} \setminus \{\boldsymbol{\pi}_1\}$, the result would be proved, since we would then have

$$\inf_{\boldsymbol{u} \in \mathcal{S}^{d-1} \setminus \{\boldsymbol{\pi}_1\}} P_{\mathbf{X}}[H_{\boldsymbol{\vartheta}, \boldsymbol{u}}] = \inf_{\boldsymbol{v} \in \mathcal{S}^{d-1} \setminus \{\boldsymbol{\pi}_2\}} P_{T(\mathbf{X})}[H_{T(\boldsymbol{\vartheta}), \boldsymbol{v}}] = D_H(T(\boldsymbol{\vartheta}), P_{T(\mathbf{X})}).$$

Let us then show that such a mapping $\boldsymbol{v}$ does exist. For any $\boldsymbol{u} \in \mathcal{S}^{d-1} \setminus \{\boldsymbol{\pi}_1\}$, $(\partial H_{\boldsymbol{\vartheta}, \boldsymbol{u}}) \cap \Pi_1$ is of dimension $d-2$, hence of codimension 1 in Π_1. Given that T is linear and bijective (due to the fact that Π_1 and Π_2 are not orthogonal), it holds that $T((\partial H_{\boldsymbol{\vartheta}, \boldsymbol{u}}) \cap \Pi_1)$ remains of codimension 1 in Π_2, so that there exists $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{u}) \in \mathcal{S}^{d-1}$ (different from $\boldsymbol{\pi}_2$) such that $T(H_{\boldsymbol{\vartheta}, \boldsymbol{u}} \cap \Pi_1) = H_{T(\boldsymbol{\vartheta}), \boldsymbol{v}(\boldsymbol{u})} \cap \Pi_2$. Therefore, $P_{\mathbf{X}}[H_{\boldsymbol{\vartheta}, \boldsymbol{u}}] = P_{\mathbf{X}}[H_{\boldsymbol{\vartheta}, \boldsymbol{u}} \cap \Pi_1] = P_{T(\mathbf{X})}[T(H_{\boldsymbol{\vartheta}, \boldsymbol{u}} \cap \Pi_1)] = P_{T(\mathbf{X})}[H_{T(\boldsymbol{\vartheta}), \boldsymbol{v}} \cap \Pi_2] = P_{T(\mathbf{X})}[H_{T(\boldsymbol{\vartheta}), \boldsymbol{v}}]$. Note also that the mapping $\boldsymbol{u} \mapsto \boldsymbol{v}(\boldsymbol{u})$ is bijective, since T is invertible. The result therefore follows. □

Proof of Lemma 4.3. Applying Lemma 1 with $\Pi_1 \equiv x_1 + \cdots + x_d = 1$ and $\Pi_2 \equiv x_d = 0$, we obtain

$$\begin{aligned} \inf_{\boldsymbol{\lambda} \in \mathcal{D}_d} P\Big[\sum_{i=1}^{d} \lambda_i U_i^2 \geq 1/d\Big] &= \inf_{\boldsymbol{\lambda} \in \mathcal{S}^{d-1}} P\Big[\sum_{i=1}^{d} \lambda_i (U_i^2 - 1/d) \geq 0\Big] \\ &= D_H\big((1/d)\mathbf{1}_d, P_{(U_1^2, \cdots, U_d^2)'}\big) \\ &= D_H\big((1/d)(\mathbf{1}_{d-1}, 0)', P_{(U_1^2, \cdots, U_{d-1}^2, 0)'}\big), \end{aligned} \tag{10}$$

where $\mathbf{1}_\ell$ denotes the ℓ-dimensional vector of ones. For $d = 2$, (10) is equal to

$$\inf_{(u_1, u_2)' \in \mathcal{S}^1} P\big[u_1(U_1^2 - 1/2) \geq 0\big] = \min\{P[U_1^2 \leq 1/2], P[U_1^2 \geq 1/2]\} = P[U_1^2 \geq 1/2].$$

We may therefore focus on the case $d = 3$ in the sequel. For $d = 3$, (10) rewrites

$$
\begin{aligned}
D_H\big((1/3)(\mathbf{1}_2, 0)', P_{(U_1^2, U_2^2, 0)'}\big) &= \inf_{\boldsymbol{\lambda} \in \mathcal{S}^2} P\left[\lambda_1(U_1^2 - 1/3) + \lambda_2(U_2^2 - 1/3) \geq 0\right] \qquad (11) \\
&= \inf_{\boldsymbol{\lambda} \in \mathcal{D}_2} P\left[\lambda_1 U_1^2 + \lambda_2 U_2^2 \geq 1/3\right].
\end{aligned}
$$

Note indeed that a minimizer of (11) cannot be of the form $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)' \in \mathcal{S}^2$ with $\lambda_1 + \lambda_2 = 0$.

Clearly, $(U_1^2, U_2^2, U_3^2)'$ is equal in distribution to $(X_1^2, X_2^2, X_3^2)'/(X_1^2 + X_2^2 + X_3^2)$, where $\mathbf{X} = (X_1, X_2, X_3)'$ is standard multinormal. Assuming, without loss of generality, that $\lambda_1 \geq \lambda_2$ (so that $\lambda_1 \geq 1/2$), the probability in (11) can then be rewritten

$$
P\left[\lambda_1 U_1^2 + \lambda_2 U_2^2 \geq 1/3\right] = P\left[(\lambda_1 - 1/3)X_1^2 + (\lambda_2 - 1/3)X_2^2 \geq (1/3)X_3^2\right]. \qquad (12)
$$

Now, two cases arise, namely (i) $\lambda_2 < 1/3$ or (ii) $\lambda_2 \geq 1/3$.

(i) For $\lambda_2 < 1/3$, rewrite (12) as

$$
P\left[(\lambda_1 - 1/3)X_1^2 \geq (1/3 - \lambda_2)X_2^2 + (1/3)X_3^2\right] = P[Y_{\mathbf{c}(\lambda_1, \lambda_2)} \geq 1],
$$

where $Y_{\mathbf{c}(\lambda_1, \lambda_2)} = X_1^2/\big(c_1(\lambda_1, \lambda_2)X_2^2 + c_2(\lambda_1, \lambda_2)X_3^2\big)$, with $c_1(\lambda_1, \lambda_2) = (1/3 - \lambda_2)/(\lambda_1 - 1/3)$ and $c_2(\lambda_1, \lambda_2) = (1/3)/(\lambda_1 - 1/3)$.

Since the $c_j(\boldsymbol{\lambda})$'s are non-negative and sum up to one, we have $(1/2, 1/2) =: \bar{\mathbf{c}} \prec \mathbf{c}(\boldsymbol{\lambda})$ (where "$\prec$" denotes majorization; see Marshall *et al.* (2011)), so that, in view of (1.4) in Eaton & Olshen (1972), $Y_{\bar{\mathbf{c}}}$ is stochastically smaller than $Y_{\mathbf{c}(\boldsymbol{\lambda})}$; see also the main result in Hájek (1962) or (1) in Lawton (1968). In particular, $P[Y_{\bar{\mathbf{c}}} \geq 1] \leq P[Y_{\mathbf{c}(\boldsymbol{\lambda})} \geq 1]$ for any $\boldsymbol{\lambda}$, which implies that $P[Y_{\mathbf{c}(\lambda_1, \lambda_2)} \geq 1]$ is minimized at $\mathbf{c}(\boldsymbol{\lambda}) = \bar{\mathbf{c}}$, which corresponds to $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) = (1, 0)$.

(ii) For $\lambda_2 \geq 1/3$, the quantity to be minimized is $P\big[Z_{\mathbf{c}(\lambda_1, \lambda_2)} \leq 1\big]$, where we denote $Z_{\mathbf{c}(\lambda_1, \lambda_2)} = X_3^2/\big(c_1(\lambda_1, \lambda_2)X_1^2 + c_2(\lambda_1, \lambda_2)X_2^2\big)$, with $c_1(\lambda_1, \lambda_2) = 3\lambda_1 - 1$ and $c_2(\lambda_1, \lambda_2) = 3\lambda_2 - 1$. Following the same argument as above, a majorization of $\mathbf{c}(\boldsymbol{\lambda})$ via $\mathbf{c}(\boldsymbol{\lambda}) \prec (1, 0)$ ensures that $Z_{\mathbf{c}(\boldsymbol{\lambda})}$ is stochastically smaller than $Z_{(1,0)}$, hence $P[Z_{\mathbf{c}(\boldsymbol{\lambda})} \leq 1] \leq P[Z_{(1,0)} \leq 1]$ which therefore gives a minimizer (for the case $\lambda_2 \geq 1/3$) at $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) = (2/3, 1/3)$.

Comparing both minimal values yields the result since $P[U_1^2 > \frac{1}{3}] = 1 - \frac{1}{\sqrt{3}}$ and $P[\frac{2}{3}U_1^2 + \frac{1}{3}U_2^2 > \frac{1}{3}] = 1/2$. $\qquad \square$

Proof of Lemma 4.4. Fix $\mathbf{V} \in \mathcal{V}_d^S$. There exists a $d \times d$ orthogonal matrix $\mathbf{O}$ and a diagonal matrix $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_d)$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$, such that $\mathbf{V} = \mathbf{O}\boldsymbol{\Lambda}\mathbf{O}'$. Affine-invariance entails that

$$
ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) = ShD_{\boldsymbol{\mu}}(\bar{\boldsymbol{\Lambda}}, P_{\mathbf{O}'\mathbf{X}}) = ShD_{\boldsymbol{\mu}}(\bar{\boldsymbol{\Lambda}}, P_{\mathbf{X}}).
$$

where $\bar{\boldsymbol{\Lambda}} := \boldsymbol{\Lambda}/S(\boldsymbol{\Lambda})$ is a $d \times d$ diagonal matrix with diagonal entries $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \ldots \geq \bar{\lambda}_d$, say. We

then have

$$ShD_{\boldsymbol{\mu}}(\mathbf{V}, P_{\mathbf{X}}) = ShD_{\boldsymbol{\mu}}(\bar{\boldsymbol{\Lambda}}, P_{\mathbf{X}}) \leq P\big[(U_1^{\bar{\boldsymbol{\Lambda}}})^2 \geq 1/d\big] = P\Big[X_1^2 \geq (1/d)\sum_{i=1}^{d} \frac{\bar{\lambda}_1}{\bar{\lambda}_i}X_i^2\Big]$$

$$\leq P\Big[X_1^2 \geq (1/d)\sum_{i=1}^{d} X_i^2\Big] = P\big[U_1^2 \geq 1/d\big], \tag{13}$$

where $\mathbf{U} = (U_1, \ldots, U_d)'$ is uniformly distributed on the unit sphere $\mathcal{S}^{d-1}$. Equality in (13) occurs if and only if $\bar{\lambda}_i = \bar{\lambda}_1$ for $i = 2, \ldots, d$, that is, if and only if $\bar{\boldsymbol{\Lambda}} = \mathbf{I}_d$, i.e., if and only if $\boldsymbol{\Lambda} = \mathbf{I}_d$, hence if and only if $\mathbf{V} = \mathbf{I}_d$. $\qquad\square$

Proof of Lemma 6.1. For the trace-based scale functional considered, it follows from the definition of $M_{S_{\text{trace}}}^{\mathbf{V}}$ that, for $d = 2$,

$$M_{S_{\text{trace}}}^{\mathbf{V}} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}.$$

Denoting $\overline{\mathbf{W}}_{D;a} = \Big(2X_1 X_2, \frac{X_2^2}{(2-a)^2} - \frac{X_1^2}{a^2} - \big(\frac{1}{2-a} - \frac{1}{a}\big)\Big)'$ and using $\varphi_g(z) = z$ and $\sigma_S = 1$, one then readily obtains $ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_a, P_{\mathbf{X}}) = D_H(\mathbf{0}_2, P_{\overline{\mathbf{W}}_{D;a}})$.

We start by computing the depth of $\mathbf{V}_1 = \mathbf{I}_2$, on the basis of $\overline{\mathbf{W}}_{D;1}$. The spherical symmetry of the random vector $\mathbf{Z} = (Z_1, Z_2)' = (X_1/\sqrt{a_0}, X_2/\sqrt{2-a_0})'$ yields

$$ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_1, P_{\mathbf{X}}) = D_H(\mathbf{0}_2, P_{\overline{\mathbf{W}}_{D;1}}) = \inf_{\boldsymbol{w}\in\mathbb{R}^2} P[\boldsymbol{w}'\overline{\mathbf{W}}_{D;1} \geq 0]$$

$$= \inf_{\boldsymbol{w}\in\mathbb{R}^2} P[\mathbf{Z}'\mathbf{Q}_{\boldsymbol{w}}\mathbf{Z} \geq 0] = \inf_{\boldsymbol{w}\in\mathbb{R}^2} P[\lambda_+^{\boldsymbol{w}} Z_1^2 + \lambda_-^{\boldsymbol{w}} Z_2^2 \geq 0], \tag{14}$$

where $\lambda_{\pm}^{\boldsymbol{w}} = (1 - a_0)w_2 \pm \sqrt{(2a_0 - a_0^2)w_1^2 + w_2^2}$ are the eigenvalues of

$$\mathbf{Q}_{\boldsymbol{w}} := \begin{pmatrix} -a_0 w_2 & \sqrt{a_0(2-a_0)}w_1 \\ \sqrt{a_0(2-a_0)}w_1 & (2-a_0)w_2 \end{pmatrix}.$$

The last infimum in (14) can be taken over all $\boldsymbol{w}$ vectors that are of the form $\boldsymbol{w} = \boldsymbol{w}(\theta) = (\cos(\theta)/\sqrt{2a_0 - a_0^2}, \sin(\theta))'$, for $\theta \in [0, 2\pi]$. For these $\boldsymbol{w}(\theta)$, one has $\lambda_{\pm}^{\boldsymbol{w}} = (1 - a_0)\sin(\theta) \pm 1$, and the infimum is obtained for $\sin\theta = -1$, which corresponds to $\boldsymbol{w} = (0, -1)'$. The depth of $\mathbf{V}_1 = \mathbf{I}_2$ is therefore given by

$$ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_1, P_{\mathbf{X}}) = P\big[a_0 Z_1^2 + (a_0 - 2)Z_2^2 \geq 0\big] =: c_{a_0}(1). \tag{15}$$

Turning to the depth of V_{a_0}, we of course have that

$$ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_{a_0}, P_{\mathbf{X}}) \leq P[(\overline{\mathbf{W}}_{D;a_0})_2 \leq 0] = P\big[a_0 Z_2^2 + (a_0 - 2)Z_1^2 \leq (2a_0 - 2)\big] =: c_{a_0}(a_0),$$

where $\mathbf{Z} = (Z_1, Z_2)'$ is as above. By conditioning on Z_2^2, one can show that

$$c_{a_0}(1) = 1 - \int_0^{\infty} F\Big(\frac{(2-a_0)z}{a_0}\Big)f(z)\,dz$$

and

$$c_{a_0}(a_0) = \int_{\frac{2-2a_0}{2-a_0}}^{\infty} F\left(\frac{(2a_0 - 2) + (2 - a_0)z}{a_0}\right) f(z)\, dz,$$

where $F(\cdot)$ and $f(\cdot)$ are the cdf and pdf of the χ_1^2 distribution, respectively. For $a_0 = .75$, this yields that

$$ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_{a_0}, P_{\mathbf{X}}) \le c_{a_0}(a_0) \approx .3732$$

is indeed strictly smaller than

$$ShD_{\boldsymbol{\mu}_0,\sigma_0,\phi,S_{\text{trace}}}(\mathbf{V}_1, P_{\mathbf{X}}) = c_{a_0}(1) \approx .4196. \qquad \square$$

References

Donoho, D. (1982). *Breakdown properties of multivariate location estimators*. Ph.D. thesis, Department of Statistics, Harvard University.

Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, *50*(3), 471–491.

Dümbgen, L. & Tyler, D. E. (2005). On the breakdown properties of some multivariate M-functionals. *Scand. J. Statist.*, *32*(2), 247–264.

Eaton, M. L. & Olshen, R. A. (1972). Random quotients and the Behrens-Fisher problem. *The Annals of Mathematical Statistics*, *43*, 1852–1860.

Frahm, G. (2009). Asymptotic distributions of robust shape matrices and scales. *J. Multivariate Anal.*, *100*, 1329–1337.

Hájek, J. (1962). Inequalities for the generalized Student's distribution and their applications. In *Select. Transl. Math. Statist. and Probability, Vol. 2*, Providence, R.I.: American Mathematical Society. 63–74.

Hallin, M., Oja, H., & Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *Ann. Statist.*, *34*(6), 2757–2789.

Hallin, M. & Paindaveine, D. (2006a). Parametric and semiparametric inference for shape: the role of the scale functional. *Statist. Decisions*, *24*(3), 327–350.

Hallin, M. & Paindaveine, D. (2006b). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.*, *34*(6), 2707–2756.

Hallin, M. & Paindaveine, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.*, *36*(3), 1261–1298.

He, X., Simpson, D., & Portnoy, S. (1990). Breakdown robustness of tests. *J. Amer. Statist. Assoc.*, *85*, 446–452.

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

John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, *59*, 169–173.

Koshevoy, G. & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.*, *25*(5), 1998–2017.

Lawton, W. H. (1968). Concentration of random quotients. *Ann. Math. Statist.*, *39*, 466–480.

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

Marshall, A. W., Olkin, I., & Arnold, B. C. (2011). *Inequalities: theory of majorization and its applications*. Springer Series in Statistics. New York: Springer, second ed.

Mizera, I. (2002). On depth and deep points: a calculus. *Ann. Statist.*, *30*(6), 1681–1736.

Mizera, I. & Müller, C. H. (2004). Location-scale depth (with discussion). *J. Amer. Statist. Assoc.*, *99*(468), 949–989.

Müller, C. H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *J. Multivariate Anal.*, *95*(1), 153–181.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.*, *1*(6), 327–332.

Paindaveine, D. (2008). A canonical definition of shape. *Statist. Probab. Lett.*, *78*(14), 2240–2247.

Paindaveine, D. & Van Bever, G. (2013). Inference on the shape of elliptical distributions based on the MCD. Tech. rep., ULB, ECARES Working Paper 2013-27.

Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.*, *95*(452), 1263–1268.

Rousseeuw, P. J. & Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.*, *94*(446), 388–433.

Sirkiä, S., Taskinen, S., Oja, H., & Tyler, D. E. (2009). Multivariate tests and estimates for shape based on spatial signs and ranks. *J. Nonparametr. Stat.*, *21*, 155–176.

Taskinen, S., Croux, C., Kankainen, A., Ollila, E., & Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *J. Multivariate Anal.*, *97*(2), 359–384.

Taskinen, S., Sirkiä, S., & Oja, H. (2010). k-step shape estimators based on spatial signs and ranks. *J. Statist. Plann. Inference*, *140*, 3376–3388.

Tatsuoka, K. S. & Tyler, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *Ann. Statist.*, *28*(4), 1219–1243.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*. Canad. Math. Congress, Montreal, Que., 523–531.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, *15*(1), 234–251.

Vardi, Y. & Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA*, *97*(4), 1423–1426.

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

Zuo, Y. & Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, *28*(2), 461–482.

General Bibliography

Bibliography

Agostinelli, C. & Romanazzi, M. (2011). Local depth. *J. Statist. Plann. Inference*, *141*(2), 817–830.

Agostinelli, C. & Romanazzi, M. (2013). Nonparametric analysis of directional data based on data depth. *Environ. Ecol. Stat.*, *20*(2), 253–270.

Aloupis, G., Cortes, F., Gomez, F., Soss, M., & Toussaint, G. (2002). Lower bounds for computing statistical depth. *Comput. Statist. Data Anal.*, *40*(2), 223–229.

Arcones, M. A., Chen, Z., & Giné, E. (1994). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.*, *22*(3), 1460–1477.

Arcones, M. A. & Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, *21*(3), 1494–1542.

Bai, Z.-D. & He, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Ann. Statist.*, *27*(5), 1616–1637.

Barnett, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A*, *139*(3), 318–355. With a discussion by R. L. Plackett, K. V. Mardia, R. M. Loynes, A. Huitson, G. M. Paddle, T. Lewis, G. A. Barnard, A. M. Walker, F. Downton, P. J. Green, M. Kendall, A. Robinson, A. Seheult and D. H. Young.

Bartoszyński, R., Pearl, D., & Lawrence, J. (1997). A multidimensional goodness-of-fit test based on interpoint distances. *J. Amer. Statist. Assoc.*, *92*, 577–586.

Bickel, P. J. (1964). On some alternative estimates for shift in the p-variate one sample problem. *Ann. Math. Statist.*, *35*, 1079–1090.

Billor, N., Abebe, A., Turkmen, A., & Nudurupati, S. V. (2008). Classification based on depth transvariations. *J. Classification*, *25*(2), 249–260.

Bremner, D., Chen, D., Iacono, J., Langerman, S., & Morin, P. (2008). Output-sensitive algorithms for Tukey depth and related problems. *Statist. Comput.*, *18*, 259–266.

Brown, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B*, *45*(1), 25–30.

Brown, B. M. & Hettmansperger, T. P. (1987). Affine invariant rank methods in the bivariate location model. *J. Roy. Statist. Soc. Ser. B*, *49*(3), 301–310.

Brown, B. M. & Hettmansperger, T. P. (1989). An affine invariant bivariate version of the sign test. *J. Roy. Statist. Soc. Ser. B*, *51*(1), 117–125.

Carrizosa, E. (1996). A characterization of halfspace depth. *J. Multivariate Anal.*, *58*, 21–26.

Cascos, I. (2009). Data depth: Multivariate statistics and geometry. In W. Kendall & I. Molchanov (Eds.), *New perspectives in stochastic geometry*. Clarendon Press, Oxford University Press, Oxford.

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

Chakraborty, B. & Chaudhuri, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *60*(1), 145–157.

Chakraborty, B., Chaudhuri, P., & Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statist. Sinica*, *8*(3), 767–784.

Chan, T. (2004). An optimal randomized algorithm for maximum Tukey depth. In *Proceedings of the 15th ACM-SIAM Symposium on discrete algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, 423–429.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, *91*(434), 862–872.

Chen, D. & Morin, P. (2011). Algorithms for bivariate majority depth. In *Proceedings of the 23rd Canadian Conference on Computational Geometry (CCCG'11)*. 425–430.

Chen, D. & Morin, P. (2012). Approximating majority depth. In *Proceedings of the 24th Canadian Conference on Computational Geometry (CCCG'12)*. 229–234.

Chen, Y., Dang, X., Peng, H., & Bart, H. L. J. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Trans. Patt. Anal Mach. Int.*, *31*(2), 288–305.

Chen, Z. (1995a). Bounds for the breakdown point of the simplicial median. *J. Multivariate Anal.*, *55*(1), 1–13.

Chen, Z. (1995b). Robustness of the half-space median. *J. Statist. Plann. Inference*, *46*(2), 175–181.

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

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

Chenouri, S. & Small, C. G. (2012). A nonparametric multivariate multisample test based on data depth. *Electron. J. Stat.*, *6*, 760–782.

Christmann, A., Fischer, P., & Joachims, T. (2002). Comparison between various regression depth methods and the support vector machine to approximate the number of misclassifications. *Comput. Statist.*, *17*, 273–287.

Christmann, A. & Rousseeuw, P. J. (2001). Measuring overlap in binary regresion. *Comput. Statist. Data Anal.*, *37*, 65–75.

Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2012). Multivariate functional halfspace depth. Tech. Rep. FEB Research Report KBI-1229, KU Leuven - Faculty of Economics and Business.

Cornfield, J. (1975). A statistician's apology. *J. Amer. Statist. Assoc.*, *70*(349), 7–14.

Cuesta-Albertos, J. & Nieto-Reyes, A. (2008). Functional classification and the random Tukey depth. Pratical issues. In C. Borgelt, G. Gonzalez-Rodriguez, W. Trutschnig, M. Lubiano, M. Gil, P. Grzegorzewski, & O. Hryniewicz (Eds.), *Combining Soft Computing and Statistical Methods in Data Analysis*. Springer, vol. 77, 123–130.

Cuesta-Albertos, J. & Nieto-Reyes, A. (2010). A random functional depth. In S. Dabo-Niang & F. Ferraty (Eds.), *Funct. Op. Stat.*. Springer, 121–126.

Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.*, *22*, 481–496.

Cuevas, A. & Fraiman, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *J. Multivariate Anal.*, *100*, 753–766.

Cui, X., Lin, L., & Yang, G. (2008). An extended projection data depth and its applications to discrimination. *Comm. Statist. Theory Methods*, *37*(13-15), 2276–2290.

Dang, X., Serfling, R., & Zhou, W. (2009). Influence functions of some depth functions, and application to depth-weighted L-statistics. *J. Nonparametr. Stat.*, *21*(1), 49–66.

Denecke, L. & Müller, C. H. (2011). Robust estimators and tests for copulas based on likelihood depth. *Comput. Statist. Data Anal.*, *55*, 2724–2738.

Denecke, L. & Müller, C. H. (2012). Consistency and robustness of tests and estimators based on depth. *J. Statist. Plann. Inference*, *142*(9), 2501–2517.

Donoho, D. (1982). *Breakdown properties of multivariate location estimators*. Ph.D. thesis, Department of Statistics, Harvard University.

Donoho, D. & Gasko, M. (1987). Multivariate generalizations of the median and trimmed mean, i. Tech. rep., Department of Statistics, University of California.

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

Ducharme, G. R. & Milasevic, P. (1987). Spatial median and directional data. *Biometrika*, 212–215.

Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statist. Probab. Lett.*, *14*(2), 119–128.

Dutta, S. & Ghosh, A. K. (2012a). On classification based on L_p depth with an adaptive choice of p. Tech. rep., Theoretical Statistics and Mathematics Unit, Indian Statistical Institute.

Dutta, S. & Ghosh, A. K. (2012b). On robust classification using projection depth. *Ann. Inst. Statist. Math.*, *64*(3), 657–676.

Dutta, S., Ghosh, A. K., & Chaudhuri, P. (2011). Some intriguing properties of Tukey's halfspace depth. *Bernoulli*, *17*(4), 1420–1434.

Dyckerhoff, R. (2000). Computing zonoid trimmed regions of bivariate data sets. In J. Bethlehem & P. van der Heijden (Eds.), *Proceedings in Computational Statistics. COMPSTAT 2000*. Physica, Heidelberg, 295–300.

Dyckerhoff, R. (2002). Inference based on data depth. In *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*, Springer, New York. 133–163.

Dyckerhoff, R. (2004). Data depths satisfying the projection property. *Allgemeines Statistisches Archiv*, *88*, 163–190.

Dyckerhoff, R., Koshevoy, G. A., & Mosler, K. (1996). Zonoid data depth: Theory and computation. In A. Pratt (Ed.), *Proceedings in Computational Statistics. COMPSTAT 1996*. Physica, Heidelberg, 235–240.

Eddy, W. (1982). Convex hull peeling. In H. Caussinus (Ed.), *Proceedings in Computational Statistics. COMPSTAT 1982*. Physica, Vienna, 42–47.

Eddy, W. (1985). Ordering multivariate data. In L. Billard (Ed.), *Computer science and Statistics: The interface*, North-Holland, Amsterdam. 25–30.

Elmore, R. T., Hettmansperger, T. P., & Xuan, F. (2006). Spherical data depth and a multivariate median. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 87–101.

Febrero, M., Galeano, P., & González-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels. *Environmetrics*, *19*, 331–345.

Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York.

Fisher, N. (1985). Spherical medians. *J. Roy. Statist. Soc. Ser. B*, *47*, 342–348.

Fisher, N. (1993). *Statistical analysis of circular data*. Cambridge University Press.

Fraiman, R., Liu, R. Y., & Meloche, J. (1997). Multivariate density estimation by probing depth. In L_1-*statistical procedures and related topics (Neuchâtel, 1997)*, Hayward, CA: Inst. Math. Statist., vol. 31 of *IMS Lecture Notes Monogr. Ser.*. 415–430.

Fraiman, R. & Meloche, J. (1999). Multivariate L-estimation (with discussion). *Test*, *8*, 255–317.

Fraiman, R. & Muniz, G. (2001). Trimmed means for functional data. *Test*, *10*(2), 419–440.

Fukuda, K. & Rosta, V. (2004). Exact parallel algorithms for the location depth and the maximum feasible subsystem problems. In C. Floudas & P. Pardalos (Eds.), *Frontiers in Global Optimization*. Kluwer Acad Publ., Boston, MA, 123–133.

Gao, Y. (2003). Data depth based on spatial ranks. *Statist. Probab. Lett.*, *65*, 217–225.

Ghosh, A. K. & Chaudhuri, P. (2005a). On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli*, *11*(1), 1–27.

Ghosh, A. K. & Chaudhuri, P. (2005b). On maximum depth and related classifiers. *Scand. J. Statist.*, *32*(2), 327–350.

Gini, C. & Galvani, L. (1929). Di talune estensioni, dei concetti di media ai caratteri qualitativi. *Metron*, *8*, 3–209.

Gower, J. (1974). Algorithm AS78. The mediancenter. *Appl. Stat.*, *23*, 466–470.

Green, P. (1981). Peeling bivariate data. In V. Barnett (Ed.), *Interpreting Multivariate Data*, Wiley, New York. 3–20.

Haldane, J. (1948). Note on the median of a multivariate distribution. *Biometrika*, *35*(3/4), 414–415.

Hallin, M., Paindaveine, D., & Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth (with discussion). *Ann. Statist.*, *38*, 635–669.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, *42*, 1887–1896.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc.

Hardy, A. & Rasson, J.-P. (1982). Une nouvelle approche des problèmes de classification automatique. *Statist. Analyse Données*, *7*, 41–56.

Hartikainen, A. & Oja, H. (2006). On some parametric, nonparametric and semiparametric discrimination rules. In *Data depth: robust multivariate analysis, computational geometry and applications*, Providence, RI: Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 61–70.

Hassairi, A. & Regaieg, O. (2008). On the tukey depth of a continuous probability distribution. *Statist. Probab. Lett.*, *78*, 2308–2313.

Hayford, J. (1902). What is the center of an area, or the center of a population? *J. Amer. Statist. Assoc.*, *8*, 47–58.

He, X. & Portnoy, S. (1998). Asymptotics of the deepest line. In *Applied statistical science, III*, Commack, NY: Nova Sci. Publ. 71–81.

He, X. & Wang, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.*, *25*(2), 495–504.

Hettmansperger, T. P., Nyblom, J., & Oja, H. (1992). On multivariate notions of sign and rank. In Y. Dodge (Ed.), L_1 *Statistical and Related Methods*. North-Holland, Amsterdam, 267–278.

Hettmansperger, T. P. & Oja, H. (1994). Affine invariant multivariate multisample sign tests. *J. Roy. Statist. Soc. Ser. B*, *56*, 235–249.

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

Hlubinka, D., Kotík, L., & Vencálek, O. (2010). Weighted halfspace depth. *Kybernetika*, *46*(1), 125–148.

Hlubinka, D. & Nagy, S. (2012). Depth-based classification of functional data. Report.

Hodges, J. (1955). A bivariate sign test. *Ann. Math. Statist.*, *26*, 523–527.

Hodges, J. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proc. 5th Berkeley Symp.*. vol. 1, 163–178.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *39*, 41–57.

Hu, Y., Wang, Y., & Wu, Y. (2011). Tensor-based projection depth. *Bernoulli*, *17*(4), 1386–1399.

Hubert, M., Claeskens, G., De Ketelaere, B., & Vakili, K. (2012). A new depth-based approach for detecting outlying curves. In *Proceedings of Compstat 2012 - 20th International Conference on Computational Statistics*. 329–340.

Hubert, M. & Van der Veeken, S. (2010). Robust classification for skewed data. *Adv. Data Anal. Classif.*, *4*(4), 239–254.

Hugg, J., Rafalin, E., Seyboth, K., & Souvaine, D. (2006). An experimental study of old and new depth measures. In *In Proc. Workshop on Algorithm Engineering and Experiments , Lecture Notes in Computer Science*. Springer, 51–64.

Ieva, F. & Paganoni, A. (2013). Depth measures for multivariate functional data. *Comm. Statist. Theory Methods*, *42*, 1265–1276.

Izem, R., Rafalin, E., & Souvaine, D. L. (2008). Describing multivariate distributions with nonlinear variation using data depth. Tech. rep., Technical Report TR-2006-2, Tufts University, Department of Computer Science.

Jörnsten, R. (2004). Clustering and classification based on the L_1 data depth. *J. Multivariate Anal.*, *90*(1), 67–89.

Jörnsten, R., Vardi, Y., & Zhang, C.-H. (2002). A robust clustering method and visualization tool based on data depth. In Y. Dodge (Ed.), *Statistical data analysis*. Birkhäuser-Verlag, Basel, 353–366.

Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical data analysis based on the L_1-norm and related methods (Neuchâtel, 1987)*, Amsterdam: North-Holland. 217–230.

Kim, J. (2000). Rate of convergence of depth contours: with application to a multivariate metrically trimmed mean. *Statist. Probab. Lett.*, *49*(4), 393–400.

Kim, S. (1992). The metrically trimmed mean as a robust estimator of location. *Ann. Statist.*, *20*, 1534–1547.

Kong, L. & Zuo, Y. (2010). Smooth depth contours characterize the underlying distribution. *J. Multivariate Anal.*, *101*(9), 2222–2226.

Koshevoy, G. A. (2002). The Tukey depth characterizes the atomic measure. *J. Multivariate Anal.*, *83*(2), 360–364.

Koshevoy, G. A. (2003). Lift-zonoid and multivariate depths. In *Developments in Robust Statistics. International Conference on Robust Statistics 2001 (ICORS)*.

Koshevoy, G. A. & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.*, *25*(5), 1998–2017.

Kotík, L. (2009). Uniform strong consistency of halfspace-like depths. *WDS'09 Proceedings of Contributed Papers*, *1*, 89–94.

Kuhn, H. (1973). A note on Fermat's problem. *Math. Program.*, *4*, 98–107.

Lange, T., Mosler, K., & Mozharovskyi, P. (2012). Fast nonparametric classification based on data depth. *Stat. Papers*. To appear.

Lee, S. M. S. (2012). Hybrid confidence regions based on data depth. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, *74*(1), 91–109.

Ley, C. & Paindaveine, D. (2012). Depth-based Runs tests for bivariate central symmetry. Under revision.

Ley, C., Sabbah, C., & Verdebout, T. (2013). A new concept of quantiles for directional data. Submitted.

Li, J., Cuesta-Albertos, J., & Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedures based on dd-plots. *J. Amer. Statist. Assoc.*, *107*(498), 737–753.

Li, J. & Liu, R. Y. (2004). New nonparametric tests of multivariate locations and scales using data depth. *Statist. Sci.*, *19*(4), 686–696.

Liu, R. Y. (1987). Simplicial depth and the related location estimators. Tech. rep., Dept. Statist. Rutgers Univ.

Liu, R. Y. (1988). On a notion of simplicial depth. *Proc. Nat. Acad. Sci. U.S.A.*, *85*(6), 1732–1734.

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

Liu, R. Y. (1992). Data depth and multivariate rank tests. In Y. Dodge (Ed.), *L$_1$ Statistics and Related Methods*. North-Holland, Amsterdam, 279–294.

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

Liu, R. Y., Serfling, R. J., & Souvaine, D. L. (Eds.) (2006). *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*. Amer. Math. Soc.

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

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

Liu, R. Y. & Singh, K. (1997). Notions of limiting P-values based on data depth and bootstrap. *J. Amer. Statist. Assoc.*, *91*, 266–277.

Liu, R. Y. & Singh, K. (2003). Rank tests for comparing multivariate scale using data depth: Testing for expansion or contraction. Unpublished manuscript.

Liu, X. & Zuo, Y. (2011a). Computing halfspace depth and regression depth. *Commun. Stat. Simulat.*.

Liu, X. & Zuo, Y. (2011b). Computing projection depth and its associated estimators. *Stat. Comput.*.

Liu, X., Zuo, Y., & Wang, Z. (2011). Exactly computing bivariate projection depth contours and median. *Comput. Statist. Data Anal.*.

Liu, Z. & Modarres, R. (2011). Lens data depth and median. *J. Nonparametr. Stat.*, *23*, 1063–1074.

Lok, W. & Lee, S. M. S. (2011). A new statistical depth function with applications to multimodal data. *J. Nonparametr. Stat.*, *23*(3), 617–631.

López-Pintado, S. & Jörnsten, R., Rebecka (2007). Functional analysis via extensions of the band depth. In *Complex datasets and inverse problems*, Beachwood, OH: Inst. Math. Statist., vol. 54 of *IMS Lecture Notes Monogr. Ser.*. 103–120.

López-Pintado, S. & Romo, J. (2005). On the concept of depth for functional data. Preprint.

López-Pintado, S. & Romo, J. (2006). Depth-based classification for functional data. In *Data depth: robust multivariate analysis, computational geometry and applications*, Providence, RI: Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 103–119.

López-Pintado, S. & Romo, J. (2007). Depth-based inference for functional data. *Comput. Statist. Data Anal.*, *51*(10), 4957–4968.

López-Pintado, S. & Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, *104*(486), 718–734.

López-Pintado, S. & Romo, J. (2011). A half-region depth for functional data. *Comput. Statist. Data Anal.*, *55*(4), 1679–1695.

Mahalanobis, P. (1936). On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India*, *12*, 49–55.

Mardia, K. (1972). *The Statistics of directional data*. Academic Press, New York.

Mardia, K. (1975). Statistics of directional data (with discussion). *J. Roy. Statist. Soc. Ser. B*, *37*, 349–393.

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

Massé, J.-C. (2009). Multivariate trimmed means based on the Tukey depth. *J. Statist. Plann. Inference*, *139*(2), 366–384.

Massé, J.-C. & Theodorescu, R. (1994). Halfplane trimming for bivariate distributions. *J. Multivariate Anal.*, *48*(2), 188–202.

Milasevic, P. & Ducharme, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.*, *15*(3), 1332–1333.

Miller, K., Ramaswami, S., Rousseeuw, P., Sellarès, T., Souvaine, D., Streinu, I., & Struyf, A. (2003). Efficient computation of location depth contours by methods of computational geometry. *Statist. Comput.*, *13*, 153–162.

Mizera, I. (2002). On depth and deep points: a calculus. *Ann. Statist.*, *30*(6), 1681–1736.

Mizera, I. & Müller, C. H. (2004). Location-scale depth (with discussion). *J. Amer. Statist. Assoc.*, *99*(468), 949–989.

Mizera, I. & Volauf, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *J. Multivariate Anal.*, *83*(2), 365–388.

Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*. Springer, New York.

Mosler, K. (2012). Data depth. Arxiv:1207-4988.

Mosler, K. & Hoberg, R. (2006). Data analysis and classification with the zonoid depth. In *Data depth: robust multivariate analysis, computational geometry and applications*, Providence, RI: Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 49–59.

Mosler, K., Lange, T., & Bazovskin, P. (2009). Computing zonoid trimmed regions in dimension $d > 2$. *Comput. Statist. Data Anal.*, *53*, 2500–2510.

Mosler, K. & Polyakova, Y. (2012). General notions of depth for functional data. Report.

Mosteller, C. & Tukey, J. W. (1977). *Data analysis and regression*. Addison-Wesley.

Müller, C. H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *J. Multivariate Anal.*, *95*(1), 153–181.

Nolan, D. (1992). Asymptotics for multivariate trimming. *Stochastic Process. Appl.*, *42*(1), 157–169.

Nolan, D. (1999). On min-max majority and deepest points. *Statist. Probab. Lett.*, *43*(4), 325–333.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.*, *1*(6), 327–332.

Oja, H. & Niimimaa, A. (1985). Asymptotic properties of the generalized median in the case of multivariate normality. *J. Roy. Statist. Soc. Ser. B*, *47*, 372–377.

Oja, H., Niimimaa, A., & Tableman, M. (1990). The finite-sample breakdown point of the oja bivariate median and the corresponding half samples version. *Statist. Probab. Lett.*, *10*, 325–328.

Plackett, R. (1976). Comment on "the ordering of multivariate data" by v. barnett. *J. Roy. Statist. Soc. Ser. A*, *139*, 344–346.

Rafalin, E. & Souvaine, D. L. (2004). Computational geometry and statistical depth measures. In *Theory and applications of recent robust methods*, Basel: Birkhäuser, Stat. Ind. Technol. 283–295.

Ramsay, J. & Silverman, B. (2005). *Functional data analysis*. Springer-Verlag. 2nd Edition.

Romanazzi, M. (2001). Influence function of halfspace depth. *J. Multivariate Anal.*, *77*(1), 138–161.

Romanazzi, M. (2009). Data depth, random simplices and multivariate dispersion. *Statist. Probab. Lett.*, *79*(12), 1473–1479.

Rousseeuw, P. & Ruts, I. (1996). Algorithm as 307: Bivariate location depth. *J. Roy. Statist. Soc. Ser. C*, *45*(4), 516–526.

Rousseeuw, P. J. & Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.*, *94*(446), 388–433.

Rousseeuw, P. J. & Ruts, I. (1998). Constructing the bivariate Tukey median. *Statist. Sinica*, *8*(3), 827–839.

Rousseeuw, P. J. & Ruts, I. (1999). The depth function of a population distribution. *Metrika*, *49*(3), 213–244.

Rousseeuw, P. J. & Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statist. Comput.*, *8*, 193–203.

Rousseeuw, P. J. & Struyf, A. (2002). A depth test for symmetry. In *Goodness-of-fit tests and model validity (Paris, 2000)*, Boston, MA: Birkhäuser Boston, Stat. Ind. Technol. 401–412.

Rousseeuw, P. J. & Struyf, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference*, *122*(1-2), 161–173.

Ruts, I. & Rousseeuw, P. J. (1996). Computing depth contours of bivariate point clouds. *Comput. Statist. Data Anal.*, *23*, 153–168.

Scates, D. (1933). Locating the median of the population in the united states. *Metron*, *11*, 49–66.

Serfling, R. (2002a). A depth function and a scale curve based on spatial quantiles. In Y. Dodge (Ed.), *Statistical data analysis based on the L_1-norm and Related Methods*, Birkhäuser. 25–28.

Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *J. Statist. Plann. Inference*, *123*(2), 259–278.

Serfling, R. (2006a). Depth functions in nonparametric multivariate inference. In *Data depth: robust multivariate analysis, computational geometry and applications*, Providence, RI: Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 1–16.

Serfling, R. (2006b). Multivariate symmetry and asymmetry. In S. Kotz, N. Balakrishnan, C. Read, & B. Vidakovic (Eds.), *Encyclopedia of Statistical Sciences*, Wiley, vol. 8. second ed., 5338–5345.

Serfling, R. J. (2002b). Quantile functions for multivariate analysis: Approaches and applications. *Stat. Neerl.*, *56*(2), 214–232. Special issue: Frontier research in theoretical statistics, 2000 (Eindhoven).

Sguera, C., Galeano, P., & Lillo, R. (2012). Spatial depth-based classification for functional data. Report.

Simpson, R. (1750). *Doctrine and Application of Fluxions*. Printed for John Nourse, London.

Singh, K. (1991). A notion of majority depth. Unpublished document.

Small, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.*, *15*, 31–39.

Small, C. G. (1990). A survey of multidimensional medians. *Internat. Statist. Rev.*, *58*(3), 263–277.

Stahel, W. A. (1981). *Robuste Schaätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen*. Ph.D. thesis, Eidgenossiche Technische Hochschule, Zürich.

Struyf, A. & Rousseeuw, P. J. (1999). Halfspace depth and regression depth characterizes the empirical distribution. *J. Multivariate Anal.*, *69*, 135–153.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*. Canad. Math. Congress, Montreal, Que., 523–531.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, *15*(1), 234–251.

Tyler, D. E. (1994). Finite-sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.*, *22*, 1024–1044.

Vardi, Y. & Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Nat. Acad. Sci. U.S.A.*, *97*(4), 1423–1426.

Vardi, Y. & Zhang, C.-H. (2001). A modified Weiszfeld algorithm for the Fermat-Weber location problem. *Math. Program., Ser. A*, *90*, 559–566.

Wang, J. & Serfling, R. (2006). Influence functions for a general class of depth-based generalized quantile functions. *J. Multivariate Anal.*, *97*(4), 810–826.

Weber, A. (1909). Über den Standort der Industrien, Tubingen. In *Alfred Weber's theory of location of industries*, University of Chicago Press. English translation by Friedriech C.J. (1929).

Weiszfeld, E. (1937). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Math. J.*, *43*, 355–386.

Wellmann, R., Harmand, P., & Müller, C. H. (2009). Distribution-free tests for polynomial regression based on simplicial depth. *J. Multivariate Anal.*, *100*(4), 622–635.

Wellmann, R. & Müller, C. H. (2010a). Depth notions for orthogonal regression. *J. Multivariate Anal.*, *101*(10), 2358–2371.

Wellmann, R. & Müller, C. H. (2010b). Tests for multiple regression based on simplicial depth. *J. Multivariate Anal.*, *101*(4), 824–838.

Yeh, A. & Singh, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *J. Roy. Statist. Soc. Ser. B*, *59*, 639–652.

Zhang, J. (2002). Some extensions of Tukey's depth function. *J. Multivariate Anal.*, *82*(1), 134–165.

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

Zuo, Y. (2006). Multidimensional trimming based on projection depth. *Ann. Statist.*, *34*(5), 2211–2251.

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

Zuo, Y. & He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.*, *34*(6), 2879–2896.

Zuo, Y. & Lai, S. (2011). Exact computation of bivariate projection depth and the Stahel–Donoho estimator. *Comput. Statist. Data Anal.*, *55*, 1173–1179.

Zuo, Y. & Serfling, R. (2000a). General notions of statistical depth function. *Ann. Statist.*, *28*(2), 461–482.

Zuo, Y. & Serfling, R. (2000b). Nonparametric notions of multivariate "scatter measure" and "more scattered" based on statistical depth functions. *J. Multivariate Anal.*, *75*(1), 62–78.

Zuo, Y. & Serfling, R. (2000c). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.*, *28*(2), 483–499.

Zuo, Y.-J. & Cui, H.-J. (2004). Statistical depth functions and some applications. *Adv. Math. (China)*, *33*(1), 1–25.

Zuo, Y.-J., Cui, H.-J., & He, X. (2004b). On the Stahel-Donoho estimator and depth-weighted means for multivariate data. *Ann. Statist.*, *32*, 169–190.